Li He · Dingjiang Yang
Guoqiang Ni

Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN

國防工業出版社
National Defense Industry Press
Springer

Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN

Li He · Dingjiang Yang · Guoqiang Ni

Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN

With Contributions by:
Lu Chen, Xiaoshuang Chen, Yong Deng, Ruijun Ding, Qi Feng, Haimei Gong, Liwei Guo, Ping Han, Li He, Weida Hu, Xiaoning Hu, Xiangyang Li, Wu Liu, Wei Lu, Ji Luo, Guoqiang Ni, Jun Shao, Xiumei Shao, Hao Sun, Yanfeng Wei, Jintong Xu, Dingjiang Yang, Jianrong Yang, Zhenhua Ye, Songlin Yu, Dongqing Yue, Yan Zhang, Degang Zhao, Liqing Zhou, Xian Zhu

國防工業出版社
National Defense Industry Press

Springer

Li He
Shanghai Institute of Technical Physics
Chinese Academy of Sciences
Shanghai
China

Guoqiang Ni
Opto-electronics School
Beijing Institute of Technology
Beijing
China

Dingjiang Yang
Tianjin Lishen Battery Joint-Stock Co. Ltd.
Tianjin
China

ISBN 978-3-662-52716-0 ISBN 978-3-662-52718-4 (eBook)
DOI 10.1007/978-3-662-52718-4

Jointly published with National Defence Industry Press, China

Library of Congress Control Number: 2016942799

Translation from the Chinese language edition: *Introduction to Advanced Focal Plane Arrays* by Li He,

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer-Verlag GmbH Berlin Heidelberg

Preface

Electro-optical (EO) infrared (IR) imaging is distinguished by such characteristics as operating passively and possessing high spatial resolution as compared to active microwave imaging technology. IR detectors for military usage were first developed in WW2, but most significant recent development occurred after the demonstration of focal plane arrays (FPAs) in the 1970s. Nowadays, the technology of IR FPAs has become one of the most essential elements of information-based military systems. As driven by military and civil needs, FPA technology is continuously evolving to ever stronger capabilities in long range sensitivity, high speed, environmental and weather compatibility, compactness in volume, low-power consumption, and cost.

In general, advanced FPAs can be a comprehensive concept. They may work in spectral ranges of either broad or very narrow spectral bands from thermal IR to UV wavelengths. They may be sensitive to polarization or phase of incident radiation, and may also work actively with laser beams for 3D imaging. The fundamental semiconducting material for IR FPAs can be HgCdTe or AlGaAs/GaAs quantum wells (or dots), InAs/GaSb superlattices as well as VO_x or amorphous Si, etc. Since EO imaging device technology is progressing very rapidly, it is difficult to treat the state-of-the-art technologies in a single book. Instead, this book intends to provide readers with a fundamental guide for understanding advanced FPAs of HgCdTe or AlGaN based on third-generation IR fabrication technology. Emphasis will be on features of multipixel arrays for very large-scale and/or multiband use, pixel (column)-level analog to digital conversion, digital signal multiplexing and integrated processing. Some recent results obtained by the authors on device design and fundamental epitaxial fabrication processes are also presented.

The Chinese version of this book was published in 2011. Some updates and modifications for the present English version are made to reflect recent developments. The chapter dealing with optical links and data processing in the original version is removed due to space limitations in the English version which contains seven chapters. Chapter 1 briefly reviews the history and trends of IR FPAs. Advanced FPAs of HgCdTe or AlGaN are outlined to provide readers with

background for the subsequent chapters. Chapter 2 presents numerical methods for designing HgCdTe multiband pixels. Chapters 3 and 4 present epitaxial techniques for multilayered HgCdTe devices on Si substrates and AlGaN, respectively. Device processing for pixel arrays of HgCdTe and AlGaN are discussed in Chaps. 5 and 6, respectively. Chapter 7 introduces methods for designing and testing CMOS readout circuits for dual-band preamps and analog to digital conversions.

The authors greatly benefitted from the work achieved by scientists and engineers worldwide in developing advanced FPAs. Those remarkable achievements have established theoretical and technical bases for this study. The authors would like to acknowledge academician Junhong Su, and Zailong Sun, Suisheng Mei, Xiaochi Zhu, Shuping Zhang, Bangkui Fan, Yi Cai, Yadong Jiang, Yingrui Wang, and Xin Lyu for encouragement, critical reviews, and instructive advice. Academician Jiaxiong Fang, academician Junhao Chu, and Ning Dai, Yanjin Li, Zhifeng Li, Xiaohao Zhou, Yunzhi Ni, Zili Xie, Ruolian Jiang, Ming Du, Yefang Zhou, Shurong Dai, Ling Wang, Jianzhong Jiang, Huimin Hou, and Jianjun Yin all made important contributions to this work. Names of students and staff involved in the work can be found in the references. Academician Lianghui Chen and others provided helpful suggestions in completing the book. Jianxin Chen, Gangyi Xu, Chun Lin, Kaihui Chu, Honglei Chen, Changzhi Shi, Yan Huang, Shiwei Xue, Hui Qiao, Quanzhi Sun, Xing Chen, Xintian Chen, Weicheng Qiu, Jian Liang, Jiao Xu et al. assisted the authors with translation and proofreading. James Torley, University of Colorado–Colorado Springs, USA, kindly assisted in language revising. The authors also thank Jianzhen Pan for her extensive efforts on the organization and coordination of this work.

Shanghai, China Li He
May 2015

Contents

Chapter 1
Fundamentals of Focal Plane Arrays

1.1 History and Trends of Infrared Imaging Detectors

Infrared (IR) detector technology has attracted attention since IR light was identified by Sir Frederick William Herschel more than 200 years ago. IR detectors for military use were developed in World War 2, but most significant development occurred after the demonstration of IR focal plane arrays (FPAs) in the 1970s. Driven by the progress of Si CMOS technology in the same period, IR FPAs of linear and staring formats were realized through hybrids of photovoltaic photodiode arrays of HgCdTe or InSb with CMOS integrated circuits (ICs). Subsequently, several fabrication milestones such as HgCdTe epitaxial growth and CdTe surface passivation for HgCdTe photodiode arrays were soon established. IR imaging with FPAs overcomes several drawbacks of the traditional single- or multiple-element detector imaging method using mechanical scanning, such as low sensitivity, slow response, high power consumption, and bulkiness. From then on, various IR FPAs of both cooled and uncooled types were developed. These arrays sensed a great portion of the IR spectral bands from the near-IR, short wavelength (SW), and mid-wavelength (MW) to long wavelength (LW).

A linear FPA is a one-dimensional device, and it accomplishes imaging relying on the motion of the carrier platform (known as the "push-broom" mode) or the mechanical scanning to produce another dimension of the two-dimensional image. In the linear FPA, pixels are linearly aligned either to one or N rows (N usually is 4–8 for LW), the latter is termed as time delay integration (TDI). In a TDI array, the pixels in different rows of same column sense the same scene (with the same detector-angular-subtense, DAS) at a defined delay in time relative to each other as the uniform movement of the platform or scanning progresses. Then, the signals of these pixels, imaging the same scene at different times, are accumulated by a hybrid CMOS readout integrated circuit (ROIC) to enhance the signal-to-noise ratio (SNR). In this way, the integration time which is often restricted by the reduced dwell time for fast scanning is extended to N times without slowing down the

L. He et al., *Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN*,
DOI 10.1007/978-3-662-52718-4_1

scanning speed. Therefore, the detectivity can be theoretically increased by a factor of $\sqrt{N}$. For satellite-based push-broom imaging of moderate resolutions, linear FPAs with one row of pixels are usually employed due to sufficient dwell time.

A staring FPA images using a two-dimensional pixel array and a ROIC multiplexing the signals of each pixel (electronic scanning). The integration time of such an array can be made quite long by eliminating mechanical scanning, usually several to tens of ms for FPAs of MW and tens to hundreds of μs for that of LW. A long integration time greatly enhances the SNR. In this case the array is focused on a particular scene and is not moved during the exposure much like a photographic camera.

MW FPAs were the most impressive in the development of the second generation (2G) technology. "Staring" with ROICs broke the limit on integration time, and ultimately brought a revolutionary change in the traditional opinion that a detector had to work at LW to obtain high sensitivity. Technical maturity in the fabrication of MW HgCdTe directly pushed MW FPAs rapidly into large-scale arrays. IR systems with small optical apertures (large f-number, denoted as f/#) became possible as the sensitivity of FPAs could be made high enough, and therefore greatly reduced the volume and cost of IR systems.

Meanwhile, new concepts and new materials for FPAs continuously emerged. FPAs of AlGaAs/GaAs quantum-well IR photo-detectors (QWIPs) and uncooled FPAs based on thermal electrical effects were developed in the period from the end of 1980s to the beginning of the 1990s. In the mid 1990s, cooled and uncooled 2G FPAs became production mature. Nowadays, there are several kinds of FPAs incorporating different operating mechanisms. FPAs of HgCdTe, InSb, and InAs/GaSb superlattices (SLs) are photovoltaic with the IR absorption through inter-band transition. QWIPs can be either photovoltaic or photoconductive for inter-subband transitions. PtSi is a photo-emissive type device and uncooled FPAs are thermal-type devices incorporating VO_x or amorphous Si. Those constitute a complete family of IR imaging devices for tactical and strategic applications of different detecting ranges, spatial resolution, and cost. The detection range and the cost of HgCdTe FPAs are the highest. Roughly speaking, taking a typical detection range for nominal HgCdTe FPA as 10 km, it reduced detection to 6 and 2 km for QW and uncooled FPAs, respectively.

Figure 1.1 summarizes the history and trend of IR detectors. The emergence of FPAs triggered a revolutionary change in IR imaging. The important wavelength of IR detectors extended from only LW in the past to MW, which quickly opened a door for extensive applications.

In 1999, Reago et al. [1] introduced the concept of third generation (3G) IR image detectors for improving performance in long-range target detection, recognition, and identification (DRI) with affordable cost. The high performance in temperature resolution and response speed of these devices is now addressed. The noise equivalent temperature difference (NETD) of 3G FPAs is expected to approach 1 mK with resolutions of higher than one million pixels and frame rates faster than 1 kHz. For reducing device cost, developing uncooled FPAs of high

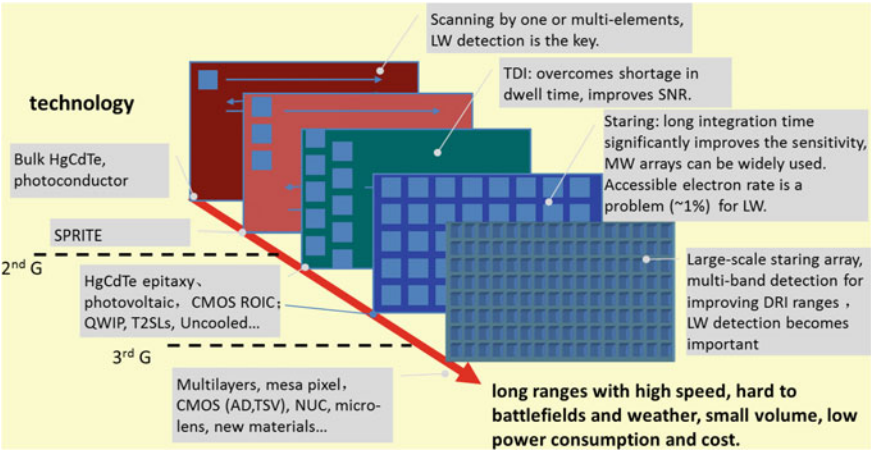


Fig. 1.1 History and trend of IR imaging devices

performance is favorable. In the past decade, significant progress has been achieved, with the technical trend as follows.

FPAs for LWIR (spectral band of 8–12 μm) and large-format arrays have been developed for increasing DRI range. LW FPAs are favorable due to more photons in LW region through the atmospheric transparency window of 8–12 μm. As shown in Fig. 1.2, according to Planck's Radiation Law the photon irradiance in the wavelength band of 8–12 μm is roughly 46 times greater than that in the 3–5 μm band for a target at 300 K. Therefore, fabricating FPAs which operate over the spectral band included in the whole atmosphere transparent window is of great importance for obtaining high sensitivity and response speed.

Very large arrays (VLA) of HgCdTe, QWIPs, or SLs were demonstrated for obtaining high spatial resolution at long range. Dual or multiband large-scale arrays

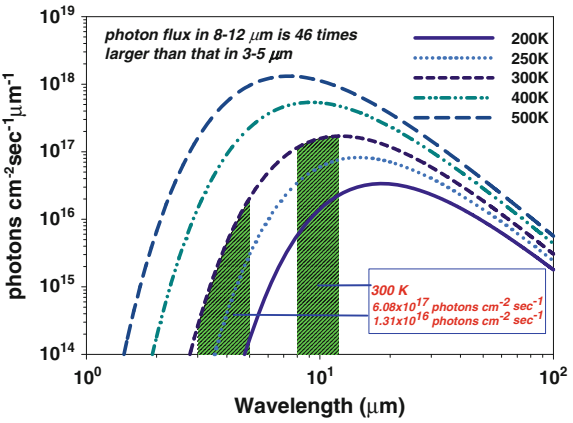


Fig. 1.2 Dependence of photon irradiance with wavelength at different temperatures

(MLA) are progressing most impressively in 3G FPA development for increasing target contrast to backgrounds. FPAs for spectrally resolved applications, and of active mode (or active/passive dual-modes) or polarization sensitive have been attracting more and more attention for increasing the detection probability for camouflaged or smoke-obscured targets in complicated environments of meteorology and the battlefield.

FPAs of affordable cost have been continuously developed. In parallel with efforts to ensure fabrication processes suitable for scaled production, low cost, large-area epitaxial HgCdTe IR FPAs of small pixel size and high-operating-temperature (HOT) have become more and more important. The latter two characteristics significantly impact the size, weight, and power consumption of IR systems.

Although the performance of uncooled FPAs is not as good as those of cooled HgCdTe arrays, the advantage of working at room temperature with no need for cooling brings to IR systems great benefits in low power consumption, reduced volume, and weight. They are widely applied in night vision, battlefield reconnaissance, seeker and earth observation from satellites. Current development efforts are to reduce pixel size (currently a pitch size of 15–17 μm is available) and to increase the uniformity in NETD. For a detector-limited system, the DAS is proportional to the ratio of pitch size to focal length. At a given focal length, the DAS or spatial resolution of an imaging system is determined by the pitch size of the FPA employed, a smaller pitch corresponds to a higher resolution. If the DAS and the f/# are maintained unchanged, a decrease in pitch size by half corresponds to reductions in focal length and aperture by 1/2 respectively. Additionally, the volume of the system shrinks to 1/8, thus overall system weight and cost are substantially reduced. Challenges for uncooled FPAs with small pixels are to increase pixel sensitivity by improving the thermal efficiency and suppressing shot, thermal, and 1/f noise. Unlike cooled FPAs, uncooled FPAs lack cold shields, which makes them much more sensitive to the local environment. For an f/1 system, the irradiation received by the FPA from environment is about 4 times that from a target. Variation in environment temperature can be different depending on scenarios. The environment temperature change may be slow for reconnaissance or fast with some abrupt changes in fire search or missile guidance, causing complication in nonuniformity calibration (NUC) for NETD. Currently commercially available uncooled FPAs are mainly based on VO_x and amorphous Si. However, new devices of different materials and structures are continuously emerging.

As applications are being continuously exploited, the response of FPAs has been extended into the ultraviolet (UV) regime. In addition to traditional UV detectors of photon multiplier tubes (PMT) or Si CCDs, AlGaN FPAs have attracted much attention which encouraged rapid progresses of GaN LEDs stimulated by the huge market for illumination sources since 1993. The large volume and high-voltage requirements of PMT are significant drawbacks to this technology. Quantum efficiency and broad spectrum response are problematic for Si diodes. Therefore, for this material UV filters are usually necessary. However, these filters are difficult to make with high-band pass transmission and off-band suppression. For FPAs of

AlGaN, the response wavelength can be adjusted by changing the fractional Al composition to cover the solar-blind (C-band: 200–280 nm) or visible–blind (A-band: 315–400 nm) UV bands with "band pass filters" embedded in multilayered structures.

As mentioned, FPAs have been developed very quickly and a variety of new approaches and concepts have been continuously updated and demonstrated. Technologies of HgCdTe, QWIP have shown potential for developing VLA and MLA. New FPAs of type-II SLs are appearing to be a most hopeful alternative technology. Looking back at the history of IR imaging detectors, it can be found that the driving force comes from ever stronger requirements for long range with high speed, environmental and weather compatibility, low in volume, low in power consumption, and low cost both in fabrication and for application. These requirements keep pushing the IR imaging technology forward.

1.2 Introduction to Advanced FPAs of HgCdTe and AlGaN

1.2.1 Outline

In general, the term of advanced FPAs is comprehensive, as mentioned in the previous section. An advanced FPA may work in spectral ranges of either broad or narrow bands from thermal IR to UV wavelengths. They may be sensitive to polarization or phase of the radiation, and may also work actively with laser beams for 3D imaging. The fundamental material can be traditional HgCdTe or AlGaAs/GaAs quantum wells, InAs/GaSb superlattices as well as VO_x or amorphous Si etc. Instead of trying to cover all of these topics, this section uses HgCdTe and AlGaN FPAs to emphasize their most technical fundamental characteristics for achieving high performance.

Consistent with the concept of 3G FPAs, advanced FPAs of HgCdTe or AlGaN can be preliminarily outlined as shown in Figs. 1.3 and 1.4. They have the features of VLA and/or MLA with pixel (column)-level ADCs, digitalized signal multiplexing and processing, as to be described in the following:

(1) VLA or MLA chips for long-range target acquisition. In the IR band, photovoltaic photodiode arrays of HgCdTe grown on Si composite substrates are emphasized for capabilities in large-area and thermal consistency. For the FPA in tactical applications or in the situation of high background signal levels, the technology of growing HgCdTe on alternative substrates becomes a technical trend. Arrays of AlGaN pin diodes are used for UV detection in solar-blind and visible-blind spectral bands as a complementary means of IR detection for target identification.

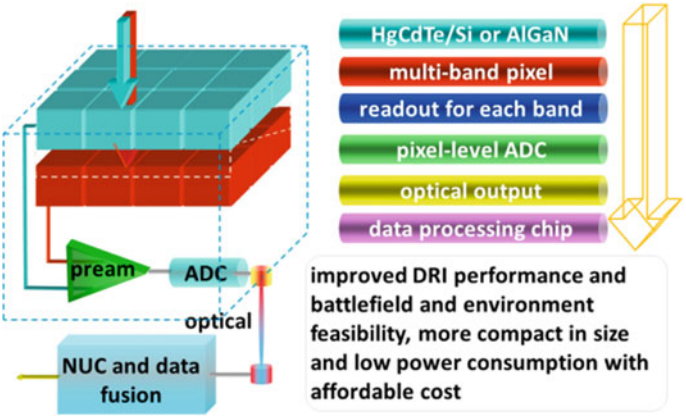


Fig. 1.3 A preliminary outline of advanced FPAs of HgCdTe or AlGaN

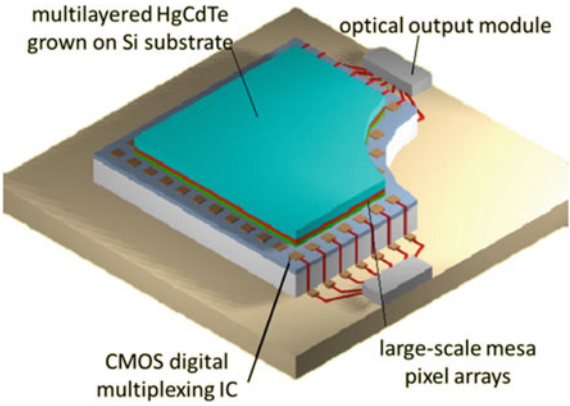


Fig. 1.4 An illustration of advanced FPA chip of HgCdTe

(2) CMOS ROICs capable of signal pre-amplification and ADCs at pixel (column)-level, high-speed digital multiplexing. The rate of signal transmission can be significantly increased for large amounts of data from VLA/MLA, providing capabilities for high-temperature resolution and signal processing.
(3) Optical output is proposed for reducing the thermal load from an increased number of signal wires. An integrated chip structure is employed for signal processing, data fusing, and autonomous target recognition (ATR).

In the next subsections, background for those features and implementations are discussed in terms of improving DRI ranges and environment tolerance.

1.2.2 Improving DRI Range by High Spatial and Temperature Resolutions

1.2.2.1 Johnson's Criteria

In 1958, Johnson [2] proposed criteria for target DRI. To detect a target with confidence of greater than 50 %, the image of a target projected on a pixel array should be larger in area than at least one pixel, and it should occupy at least 4 and 6.5 pixels for both recognition and identification, respectively. If the confidence is to be greater than 95 %, the numbers of pixels that an image comprises for DRI should exceed 2, 8, and 16 respectively. Therefore, geometrically, a larger pixel array and a smaller optical DAS will result in greater DRI ranges.

In addition, DRI ranges also depend on the target contrast, IR atmospheric transmission, and detector sensitivity. The relation between them can be described by the minimum resolvable temperature difference (MRTD), which gives the sensitivity of an IR system in terms of spatial response (modulation transfer function, MTF) and of temperature response (noise equivalent temperature difference, NETD). Figure 1.5 illustrates how the resolvable temperature difference of an IR system decreases with increasing spatial frequency. The solid and dot-dashed curves correspond to FPA #1 and #2, respectively, where the pixel scale of FPA #2 is assumed to be larger than that of FPA #1. DRI ranges can be estimated according to Johnson's criteria. For example, neglecting atmospheric attenuation and assuming a target of 3 m in height with a temperature contrast of 2 K, a spatial frequency of 17 cycles/mrad or a DAS of 1/17 rad (FPA #1) can be obtained (Fig. 1.5). Maximum DRI ranges with FPA #1 are $R_{\mathrm{d}} = 3 \cdot 17 \times 10^3 = 51(\mathrm{km})$, $R_{\mathrm{r}} = 3 \cdot \frac{17 \times 10^3}{4} = 13(\mathrm{km})$, and $R_{\mathrm{i}} = 3 \cdot \frac{17 \times 10^3}{6.5} = 8(\mathrm{km})$, respectively. When atmospheric attenuation is considered, detection range reduces to 27 km. It is obvious

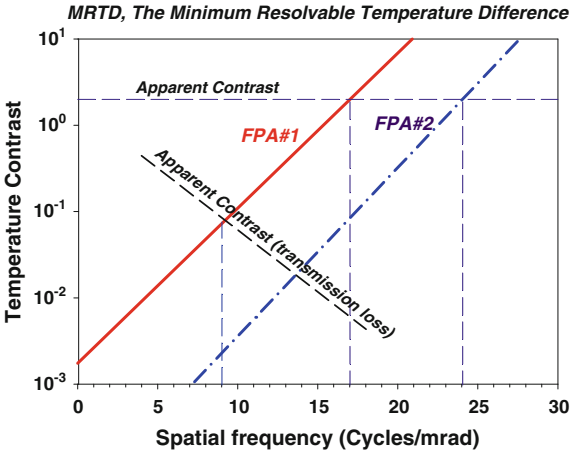


Fig. 1.5 Representative chart showing the resolvable temperature difference of an IR system decreases with the increasing spatial frequency. *Solid* and *dot-dashed curves* correspond to FPAs of different sensitivities and formats

that to improve the DRI range a low NETD is also of importance in addition to a large format.

1.2.2.2 Approach for Large-Scale (High Spatial Resolution) HgCdTe FPAs

An obstacle for large-scale FPAs comes from the size limit of the CMOS fabrication process and difficulties in HgCdTe fabrication with affordable cost.

Figure 1.6 shows typical sizes for different formats. The array is about 30 × 30 mm in size for an FPA of 2048 × 2048. Unlike typical integrated circuits, the pixel size of an FPA cannot shrink very much as it is limited by the optical diffraction or the size of Airy disk which is given by $d = 2.44 \cdot \lambda \cdot f/\#$, λ being wavelength. A pixel size of 15 μm is usual for a MW FPA for under-sampling optics of a small f/#.

The size of ROIC is limited by the standard CMOS process of lithography. To overcome this limitation, a technique called "stitching" was developed, by which a ROIC chip is divided into several subblocks to fit the CMOS standard size, and the blocks are then "stitched up" together by specially developed processes.

For the size of HgCdTe FPAs, however, there is no fundamental limit. The area of lattice-matched ZnCdTe substrates for HgCdTe epitaxy can be larger than 70 × 70 mm^2, enough for accommodating a 4096 × 4096 (with pixel size of 15 μm) FPA. Although it is technically possible for even larger substrates, however, the problem could be the cost. Material properties of ZnCdTe, such as low thermal conductivity and low defect formation energy, make it difficult to obtain a high ingot yield. [211]-orientated substrates for molecular-beam epitaxy (MBE) are more demanding because boundaries of twins and stacking faults intersect the surfaces.

The advanced HgCdTe FPA employs HgCdTe epitaxial wafers grown on Si substrates (HgCdTe/Si) to overcome the problems of substrate area and thermal mismatch between HgCdTe arrays with CMOS ROIC. HgCdTe/Si also helps to drive down the cost. Like CMOS technology, the number of sensor chips per wafer

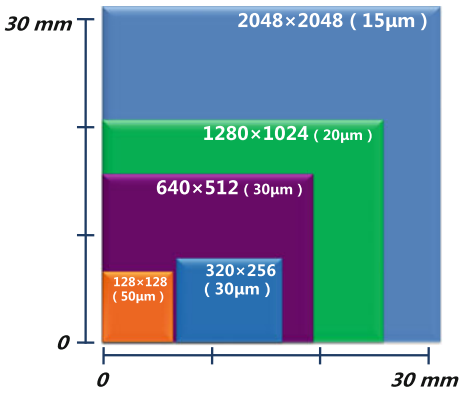


Fig. 1.6 Typical sizes of different FPA format

drives the total cost of fabrication. A 6 in. HgCdTe/Si wafer accommodates 120 chips of 640 × 480 or 37 chips of 1280 × 720 at a pixel size of 20 μm. HgCdTe FPA cost is also remarkably impacted by substrate prices. Taking the price of a Si substrate as 1, GaAs will be 5, Ge will be 8, and traditional ZnCdTe will be as high as 200. InSb and InGaAs are widely used for MW and SW FPAs, respectively. Commercially available sizes of InSb and InGaAs wafers are about 4 in. (6 in. will be soon available) and 3 in., respectively. Taking the price of InSb as 1, MW HgCdTe/Si will be 0.6. The price of HgCdTe/Si is about 40 % that of InGaAs. It is obvious that HgCdTe/Si represents a remarkable potential in realizing low cost FPAs.

Due to large differences in crystal properties between Si and HgCdTe, a high density of defects is exhibited at the interface of the HgCdTe epilayer and Si substrate. A carefully designed buffer layer is critical to localize these defects to prevent their extending into the subsequently grown HgCdTe. For LW applications, a process of removing substrates and defected interfaces may be necessary.

1.2.2.3 Approach for High Sensitivity (Temperature Resolution) HgCdTe FPAs

The detectivity D* or charge-handling capacity is another key issue for temperature resolution. The crucial limiting factor for D* is shot noise due to dark current, background photocurrent, and thermal noise (Johnson noise).

Figure 1.7 shows the relation between the cutoff wavelength and R_0A (a product of dynamic resistance at zero-bias and junction area with a unit of Ω cm^2) of HgCdTe photodiodes at 80 K for the case of diffusion current limiting high performance. η_{BLIP} is a ratio of photon noise to total noise. η_{BLIP} = 100 % means that the detector works at a condition limited by background radiation. At a zero-bias condition, there is no dark current present and the main noise source is Johnson noise i_{noise}, which is given by

$$i_{\mathrm{noise}}^2(V=0)=\frac{4kTA_{\mathrm{d}}\Delta f}{R_0A},$$

where A_{d} is the area of photodiode, Δf and T are the bandwidth and FPA temperature, respectively. It can be seen in Fig. 1.7 that the diffusion current limited performance of FPAs at different spectral bands of 3–5, 8–10, and 8–12 μm approach the background noise level, which suggests that no more margin is available for reducing noise. Therefore, methods for improving the performance must turn to increasing the integration time or charge-handling capacity. Taking a 640 × 512 FPA of 30 μm sized pixels as an example, a maximum allowable integration time is 1.2 ms for a typical pixel rate of 40 MHz and a frame rate of 50 Hz. A maximum number of detectable photons by the FPA with *f*/2 optics can be calculated as being 46 M, 1G and 2G for spectral bands of 3–5, 8–10, and 8–12 μm, respectively. Assuming a charge-handling capacity of 30 M electrons (50 % full well) and

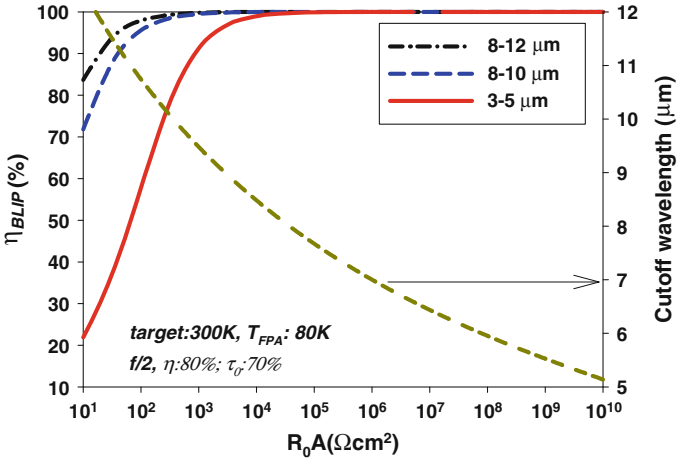


Fig. 1.7 Relations of background limit ratio and FPA cutoff wavelength with R_0A of diffusion current limited case (n-on-p architecture) for a *f*/2 optics, where target temperature, detector quantum efficiency, optical efficiency, and FPA working temperature of 300 K, 80, 70 % and 80 K were assumed for calculation, respectively

neglecting dark current, ratios between the signal photons to maximum accessible electrons (accessible ratio) are 65.2, 3.1, and 1.4 % for the above mentioned spectral bands, respectively. It becomes obvious that a low ratio of signal photons to accessible electrons for this LW case creates a limit to sensitivity.

Figure 1.8 shows the relation between charge-handling capacity of FPAs as a function of NETD for a 300 K target at different spectral bands. Taking an *f*/2 optic as an example, to achieve a goal of a NETD of 1 mK, the charge-handling capacity should be as high as 1.6G electrons for MW of 3–5 μm, and 11Ge for LW of 8–12 μm, respectively. For the same NETD in the LW band the charge-handling

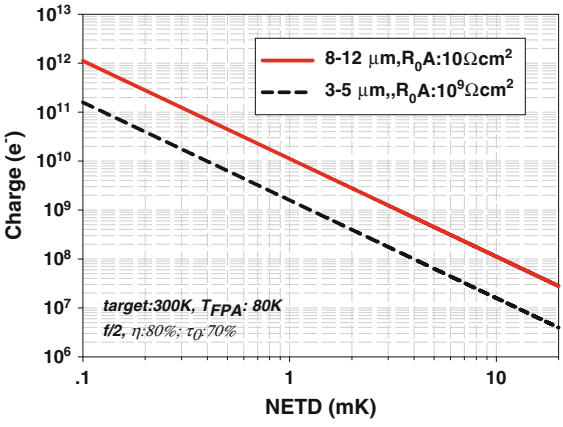


Fig. 1.8 Relation of charge-handling capacity with NETD for target of 300 K for MW and LW bands, respectively

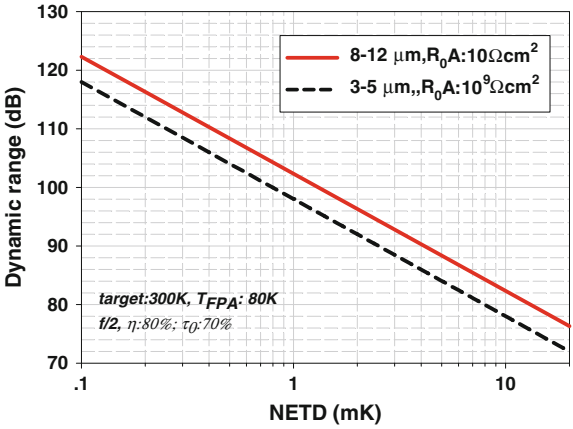


Fig. 1.9 Dynamic range of FPAs with dependence on NETD for *f*/2 optics, MW and LW bands, respectively

capacity must be larger than that for the MW band due to a reduced thermal contrast in the LW band. It should be noted that in the calculation shown in Fig. 1.8 junction resistances of R_0A of 1G Ω cm^2 and R_0A of 10 Ω cm^2 for MW and LW FPAs were assumed, respectively. In the extreme case of being dark current free, a charge-handling capacity of at least 3G electrons is required for obtaining a 1 mK NETD.

A capacity of 1G electron is almost 20 times greater than that which can be provided by the standard CMOS process for a pitch size of 30 × 30 μm. An approach using through Si vias (TSV) becomes a technical concern for increasing the charge handle capability of ROIC. With the continuous shrinking of CMOS line width fabrication techniques, gate oxide is becoming much thinner than before allowing larger CMOS capacitors. However, the working voltage is decreasing also, and noise related to large capacitors may arise to cancel out the gain achieved by increasing capacitor size.

Figure 1.9 shows the dynamic range of FPAs as related to NETD in different spectral bands for *f*/2 optics. A target temperature of 300 K, 70 % optical efficiency, and 80 % quantum efficiency were assumed in the calculation. As can be seen in the figure, to obtain a NETD of 1 mK a dynamic range of ∼100 dB is necessary for both MW and LW bands. Assuming a noise floor of 0.3 mV for subsequent processing electronics, with the current CMOS voltage of 3 V, a maximum dynamic range of 80 dB is obtained. Therefore, the dynamic range is another limitation to high sensitivity.

A mitigation technique for the charge-handling capacity limitation and high dynamic range is to integrate the charge outside of the pixels, specifically, to transfer the charge outside the FPAs and then accumulate it. In this way, the problems in charge-handling capacity and dynamic range are then changed to the problem of processing speed within ROIC for multiplexing for low noise in a given frame time.

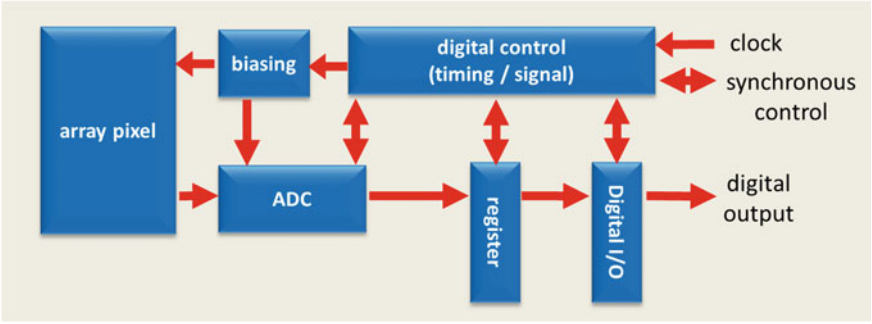


Fig. 1.10 A block diagram of a digital multiplexing ROIC

It is known that analog signal transmission is limited by noise and bandwidth proportionally increases with transmission rate. It takes 16 ms to transfer data from a 640 × 512 frame serially at a pixel rate of 5–20 MHz, occupying about 80 % of a 50 Hz frame time for transmission. Although the transmission time can be effectively reduced by means of parallel outputs of subframes as commonly employed, it will complicate applications by increasing the number of signal processing stages which may become more serious for multiband FPAs. An effective way to solve the problem is by employing ADCs at pixel (or column)-level to surpass the analog transmission limit.

Figure 1.10 shows a block diagram of a digital ROIC. The signals are multiplexed digitally to break the bottleneck of speed and noise of analog transmission for signal transmission of FPAs of VLA and MLA. It is obvious that ADCs in pixel (or column)-level are important components of advanced FPA technology.

CMOS image sensors, many uncooled and some cooled FPAs, already employ single or multiple ADCs. A digital function of nonuniformity calibration is also integrated in some of them. However, most of those devices work at room temperature, thus cooling power consumption is not a major concern. While for HgCdTe FPAs which operate at low temperatures of typically 80 K, the cooling power required is an important factor, being directly related to weight, volume, and lifetime of FPA assembles. Therefore, for cooled FPA applications, ADC design should be optimized in terms of performance and power consumption, and some trade-offs have to be made. Both high speed and high resolution (with high power) are required to justify an ADC at chip-level. Although for in-pixel level ADCs the conversion speed can be much slower and the resolution can be low too, however, the simplified sampling logic and suppression of power consumption become important issues.

1.2.3 Improving DRI Range by Multiband Imaging

An IR imager identifies the target from background by geographic and spectral features. Using a single IR band, geographic features obtained are not always reliable. For example, it is hard for a single-band system to distinguish between IR signal from a bird nearby and the atmospherically attenuated signal from a missile far away. In fact, imaging within a single band is often bothered by problems such as gray-body, low-contrast, and IR camouflage. With dual-band imaging, spectral information can be used in distinguishing targets against a background and, in some sense, relieves the requirement for scale size.

Dual-band imaging provides more freedom. MW FPAs nominally work in the atmospherically transparent window of 3–5 μm. In a spectral band shorter than 3.8 μm, aerosol-scattered sunlight forms a strong background making targets (300 K) to be negatively contrasted. For this reason, the response wavelength of MW FPAs is usually longer than 3.8 μm. Inside the transparent atmospheric window for MW, a region between 4.2 and 4.6 μm is opaque due to CO_2 absorption. By Planck's Radiation Law, radiance of a 300 K target in 4.6–4.8 μm band is 1.5 times that in the 3.8–4.2 μm transparent band. Therefore, a detection band can be optimized to operate in the 4.6–4.8 μm band in daytime to avoid sunlight scattering and in the 3.8–4.8 μm band or 3.8–4.2/4.6–4.8 μm dual-bands at night to obtain the maximum SNR. Traditionally, a dual-band or multiband IR system relies on two or more single-band FPAs using optical splitters. It is obvious that by employing advanced FPAs capable of dual-band or multiband simultaneous or sequential imaging, system complexity, weight, and power consumption can be significantly reduced, and reliability can be enhanced.

IR transmission through the atmosphere closely depends on wavelength. Generally speaking, the atmospheric transmission is poorer for shorter wavelength. Rayleigh scattering probability is inversely proportional to the fourth power of wavelength and Mie scattering is inversely proportional to the 1.3–1.6th power of wavelength for a moderate to fine weather conditions. The DAS for diffraction-limited optics is proportional to wavelength, and inversely proportional to optical aperture. Shorter wavelengths correspond to a higher optical resolution. For a diffraction-limited dual-band system of common aperture, a MW band is preferred to obtain high spatial resolution as its DAS is roughly one half of that for the LW band. If a dual-band FPA is used, an f/# for the MW band is twice that for the LW band (since dual-band detection is performed with the same pixel size), suggesting that optics of variable f/# are required to achieve the best resolution in both bands.

In addition to the attenuation by atmospheric scattering and absorption, IR transmission in the transparent windows is known to be influenced by the atmospheric turbulence along transmission path. The turbulence causes degradation in spatial resolution of an IR system. The fluctuations or scintillations in transmission path distort the phase of the wave front. These effects can be significantly enhanced as the path length is increased, resulting in image dancing, and blurring. The optical

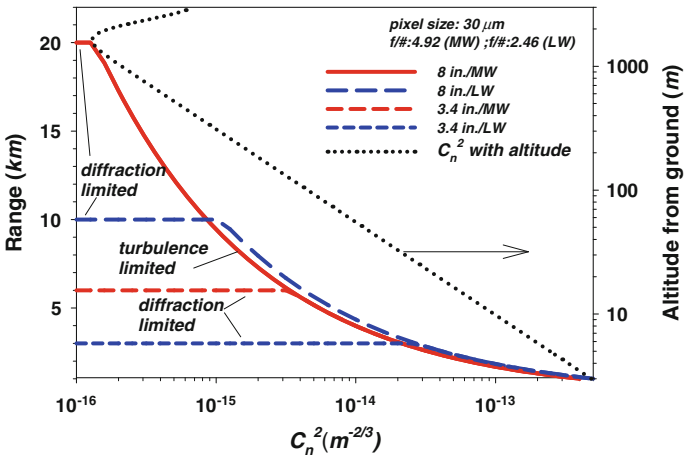


Fig. 1.11 Detection ranges as a function of C_n^2 with different optical apertures and wavelengths according to the method proposed by ref. [3]. Correlation between C_n^2 and the altitude from the ground is also shown in the figure by using the modified calculation model

turbulence is described by the refractive-index structure parameter C_n^2, a measure of the strength of the fluctuations in the refractive index, a function of pressure, temperature, and wavelength.

Therefore, especially for long-range/large-aperture IR systems, effects of atmospheric turbulence for long path-lengths must be considered for applications at an altitude bellow 1–2 km. When an optical aperture becomes larger than the atmosphere coherent length r_0, the resolution does not increase as the aperture being increased further. As shown in Fig. 1.11 for diffraction-limited optics of large common aperture, at the low turbulence region ($C_n^2 \sim 10^{-16}$ $m^{-2/3}$), the range for detection is limited by the optics and the spatial resolution of MW band is superior to that of LW. As C_n^2 increases to $\sim 10^{-15}$ $m^{-2/3}$, the superiority of MW begins to decrease while the resolution of LW band is unchanged at the diffraction limit. As C_n^2 increases further, the turbulence restricts spatial resolutions of both MW and LW band. For a small aperture system, the influence of atmospheric turbulence is less significant due to shorter transmission paths. The example above shows that dual-band FPAs can provide a way of making optimal choices in obtaining both higher spatial and temperature resolutions depending on different atmosphere conditions.

In the above discussion, advantages of dual or multiband FPAs in IR imaging were mentioned in terms of the requirements for target DRI. For developing IR multiband FPAs, the technical driving forces are (1) MBE growth of HgCdTe multi layers, (2) pixels of mesa structure with high fill factors, and (3) digital ROIC. A supporting technique may be variable f/# Dewar assemblies. For LW a small f/#

is used for searching and surveillance, whereas for MW a larger f/# optic is used for identification.

UV imaging is also considered as a part of advanced FPAs. IR imaging assisted with UV for a missile seeker is attracting more and more attention as IR countermeasures are being rapidly developed. UV radiation can be absorbed by the ozonosphere in a spectral range termed "solar-blind," in which the UV radiation from the sun is very weak below the ozonosphere. In other words, in this spectral region, the UV background is "dark" even in daytime which helps sensors capture artificial UV signals produced by missile plumes. Because UV light is seriously scattered in the atmosphere, a UV detector can find radiation from the solid propellant plume of missile over a wide viewing angle. Thus, UV imaging is of great interest in providing important detection data in addition to IR warning systems for missile approach warning (MAW) by reducing the false alarm probability.

Oil pollution at sea can be monitored by the fusion of IR and UV imagery. UV imaging senses a floating oil film as thin as 0.1 mm by reflected UV radiation. IR imaging senses emissivity differences between the oil film and the surrounding water when the film is thin and senses heating effects on the water for an oil film thicker than 0.5 mm.

The AlGaN UV FPA is a type of the photonic detector of high SNR. It has advantages of no cooling requirements, no optical scanning, small volume, and light weight, and is emerging as an alternative technology for PMTs and CCDs as mentioned previously.

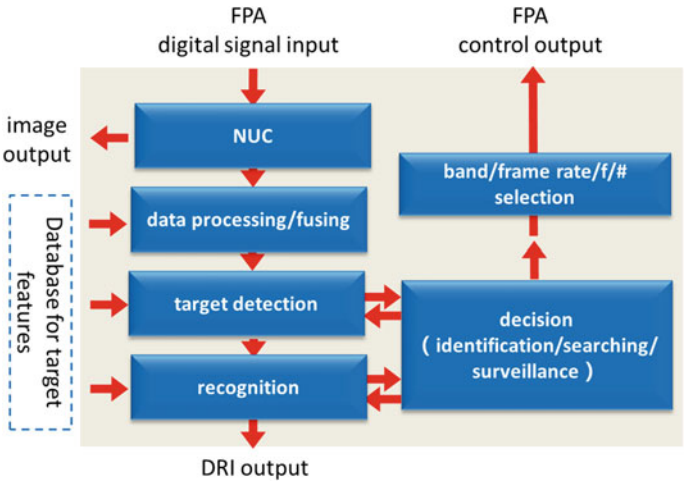


Fig. 1.12 A block diagram for a data processing chip

1.2.4 Improving Compactness and Intelligence by Integrated Processing Chip

The integrated chip for signal processing of advanced FPAs is schematically shown in Fig. 1.12. It is capable of fusing pixel allocations and attributes, complementing data managing and corresponding, and controlling bandwidth, processing resolution, synchronization, optimal synthesis etc.

As mentioned, because the residual non-uniformity (RNU) of response can bring spatial noise to images, it is another limiting factor for achieving very low NETD. For example, for obtaining a NETD of 2.5 mK the RNU should be less than 0.004 %. Thus, together with efforts in improving uniformities of FPAs, optimized algorithms with scenarios, algorithms for image fusion and adaptive tools in the decision level for ATR are also important aspects of achieving a very low NETD.

1.3 Summary

The history and trends of IR imaging devices were briefly reviewed. Considering the evolution of 3G FPAs, some distinguished features of advanced FPAs of HgCdTe or AlGaN were discussed for large-scale pixel arrays of multiband sensitivity, and capability for high-speed multiplexing. Image data can be optically transmitted, processed and fused. Fundamental imaging device characteristics were introduced in order to understand the methods required for improving IR FPA range performance for target DRI and feasibility for battlefield environments, as well as for developing more compact and intelligent FPAs. Some limiting factors for HgCdTe VLAs and charge handling were analyzed, solutions of employing HgCdTe on alternative Si composite substrates and pixel (or column)-level ADCs for digital multiplexing were discussed. The necessity of using multiband detection was described in conjunction with features of target characteristics and atmospheric transmission. The basics of integrated chip design for data processing and data fusion were outlined.

References

1. Reago DA, Horn SB, Campbell J, et. al. Third-generation imaging sensor system concepts. SPIE, 1999;3701:108–17.
2. Johnson J. Analysis of image forming systems, in image intensifier symposium, AD 220160 (Warfare Electrical Engineering Department, U.S. Army Research and Development Laboratories, Ft. Belvoir, Va.) 1958;244–73.
3. Accetta JS, Shumaker DL, editors. The infrared and electro-optical systems handbook, Vol.2. Atmospheric propagation of radiation, Infrared Information Analysis Center and SPIE Optical Engineering Press, 1993.

Chapter 2
Design Methods for HgCdTe Infrared Detectors

2.1 Introduction

HgCdTe infrared detectors have been available for half a century. Many physical models of HgCdTe infrared detectors, which are based on the basic principles of semiconductor physics, have been developed. These models have been widely used in the design and simulation of HgCdTe infrared detectors for the development of infrared detector technology. However, the accuracy of device and material parameters used in the physics models, which is very important for design and simulation, is strongly dependent on the processing conditions of device and material. Obtaining these parameters precisely is a problem, which should be solved urgently by the development of device design technology. This chapter covers some methods and results for the extraction of device and material parameters. Another key point is the effectiveness of the models used in specific design and simulation software. Because of the complexity of actual device structure, three-dimensional (or two-dimensional equivalent structure) modeling software is often required for an accurate device simulation. Simplifying the calculation model is usually applied to reduce the probability of divergence in finite element mathematical problems and computation time. However, the validity of the simplified model is only effective to certain specialized devices. At the same time, one has to acquire a set of characteristic device and material parameters based on the basic model used for the actual design method. In this chapter, methods of characteristic parameter extraction and model optimization are introduced in detail.

First, this chapter presents a basic theoretical framework based on semiconductor *p-n* junction theoretical models for simulation of HgCdTe infrared detectors, and some precision optimization results of the relationship between carrier's concentration and Fermi level for HgCdTe materials. Second, the spectral response, cross talk and other important properties of HgCdTe infrared detectors, including two-color devices, are calculated and compared with experimental results. Lastly, the effective extraction method of the characteristic parameters from fabricated

L. He et al., *Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN*,
DOI 10.1007/978-3-662-52718-4_2

HgCdTe infrared detectors is introduced, including extractions of material parameters from device current–voltage characteristics, extraction of minority carrier lifetime in *pn* junction structure from photoelectric response, and so on.

2.2 Simulation and Design of HgCdTe Infrared Detectors

Hg is usually unstable in HgCdTe materials due to the fragility of chemical bonds Hg–Cd, which results in the formation of Hg vacancy and Hg clusters. These formations cause some physical parameters of the current–voltage characteristics of HgCdTe *p-n* junctions be randomly discrete, and together with the influence of other defects, result in the nonuniformity of device performance. Therefore, it is necessary to establish a device design platform, which is closely related to the nominal manufacturing process. This section outlines the basic assumptions of the conventional analytical model, and establishes the fitting method of characteristic parameter extraction. Effects of process conditions on the device performance are analyzed statistically, providing a theoretical basis for design and performance optimization of HgCdTe infrared detectors.

2.2.1 Foundation for HgCdTe Infrared Detector Designs

Traditional simulation of semiconductor devices is mainly to solve the combination of the Poisson equation (Eq. 2.1), steady-state continuity equation (Eq. 2.2), and carrier current density equation (Eq. 2.3) within specific boundary conditions of the device:

$$\nabla^2\psi = -\frac{q}{\varepsilon_0\varepsilon}(\Gamma + p - n) - \frac{1}{\varepsilon}\nabla\psi\nabla\varepsilon \tag{2.1}$$

$$\frac{1}{q}\nabla\vec{J}_{\mathrm{n}} + (G - R) = 0; \quad \frac{1}{q}\nabla\vec{J}_{\mathrm{p}} - (G - R) = 0 \tag{2.2}$$

$$\vec{J}_{\mathrm{n}} = qn\mu_{\mathrm{n}}\vec{E}_{\mathrm{n}} + qD_{\mathrm{n}}\nabla n; \quad \vec{J}_{\mathrm{p}} = qp\mu_{\mathrm{p}}\vec{E}_{\mathrm{p}} - qD_{\mathrm{p}}\nabla p \tag{2.3}$$

where Ψ, q, and ε are the electrostatic potential, electron charge, and permittivity of the semiconductors respectively. R is the carrier recombination rate, and G is the carrier generation rate. n and p are the concentrations of electrons and holes, where J_{n} and J_{p} are the electron and hole current densities respectively. D_{n} and D_{p} are the electron and hole diffusion coefficients, E_{n} and E_{p} are the electron and hole effective electric fields, and μ_{n} and μ_{p} are the electron and hole mobility respectively. These equations evolve into nonlinear equation groups with the finite element method (FEM), then are solved by a method such as the Newton iteration process. For the

analysis of the photoelectric effect for semiconductor devices, the photogeneration rate should be added to Eq. 2.2.

The photogeneration is simulated by ray tracing:

$$G^{\text{opt}}(z) = J(x, y, z_0) \cdot \alpha(\lambda, z) \cdot \exp\left[-\left|\int_{z_0}^{z} \alpha(\lambda, z)\mathrm{d}z\right|\right] \tag{2.4}$$

where λ is the wavelength, $\alpha(\lambda, z)$ is the absorption coefficient of HgCdTe material, $J(x, y, z_0)$ is the optical beam spatial variation of intensity over the window where rays enter the device, and z_0 is the position along the ray where absorption begins.

Narrow gap semiconductors, such as HgCdTe, have the following characteristics: (1) Conduction band $E(k)$ of HgCdTe, which can be described by Kane's $\boldsymbol{k}.\boldsymbol{p}$ model, is a non-parabolic belt, so the electron concentration in the conduction band should be fixed [1–4]. (2) The effective electron mass of the HgCdTe conduction band is very small, which causes the semiconductor to degenerate due to the low effective state density of the conduction band under low temperatures. Therefore, the distribution of carrier concentration obeys Fermi–Dirac statistics [5–7]. Fermi–Dirac integrals must be employed in the calculation of these two additional characteristics, which results in many difficulties for theoretical simulation. Only HgCdTe material parameters were joined into the parameter library when the numerical model is applied for our theoretical analysis [8–11]. However, the influence of these two characteristics has never been considered in analytical modeling [12–14]. In order to perfect the existing analytical and numerical models, this section presents a simple carrier approximation model, and applies these two characteristics of HgCdTe to device simulations. Finally, the necessity and adaptability of carrier approximate models are assessed by the analytical model.

2.2.1.1 Model of Carrier Density Approximation

The effects of carrier degeneracy and conduction band non-parabolicity are introduced into the device simulation by the expression that describes the relation between the Fermi energy and the carrier density [14]. Until the present time, there have been no detailed reports on the effects of carrier degeneracy and conduction band non-parabolicity on the simulation of the HgCdTe photovoltaic devices [14–20]. Some papers mention that neglecting the two factors leads to enormous errors in the calculation of the electron density and causes an overestimated dark current in the devices [20]. Therefore, practically understanding the carrier degeneracy and conduction band non-parabolicity effects on the simulation of HgCdTe photovoltaic devices and establishing a simple carrier density approximation that takes account of the two factors in simulation will be greatly beneficial to the design, analysis, and characterization of HgCdTe devices.

The general carrier density can be expressed by [15]:

$$n = \int \frac{\mathrm{d}N}{\mathrm{d}E} \cdot f(E) \cdot \mathrm{d}E \tag{2.5}$$

where $\mathrm{d}N/\mathrm{d}E$ is the density of the energy states, and $f(E)$ the distribution function.

By considering both carrier degeneracy and conduction band non-parabolicity, the distribution function $f(E)$ should be a Fermi function because of the carrier degeneracy. And the carrier density is given by the following integral expression [16–18]:

$$n = \frac{2N_c}{\sqrt{\pi}} \int_0^\infty \frac{\varepsilon^{1/2}(1+\alpha\varepsilon)^{1/2}(1+2\alpha\varepsilon)}{\exp(\varepsilon-\phi)+1}\mathrm{d}\varepsilon \tag{2.6}$$

Here we take the zero energy to be at the bottom of the conduction band, $\varepsilon = E/kT$ is the normalized electron kinetic energy, $\phi = E_f/kT$ the reduced Fermi energy, N_c the effective density of states. The coefficient $\alpha = \frac{1}{\varepsilon_g}\left(1 - \frac{m_e^*}{m_0}\right)^2$ is the non-parabolicity factor, here $\varepsilon_g = E_g/kT$ is the normalized band gap, and m_e^* and m_0 are electron effective mass and free electron mass, respectively.

The numerical calculation of Eq. (2.6) is time-consuming and hence is not suitable for efficient device simulation. Therefore, a simplified analytic approximation of Eq. (2.6) is often needed. Ariel et al. have proposed a relationship between Fermi energy ϕ and carrier density n, where carrier degeneracy and conduction band non-parabolicity have been taken into account [18]. Ariel's model, however, is only applicable for weak degeneracy. Additionally, they treated the effective mass m_e as constant. It is well known that m_e has a dependence on the band gap ε_g. When we take this dependence into consideration, it is shown in Fig. 2.1 that Ariel model deviates from the numerical solution of Eq. (2.6) in the range of about $\phi > 7$. For n-on-p devices whose junction is formed by ion implantation, the carrier concentration on n-side could be up to the order of 10^{18} cm^{-3} [16, 19], and at the liquid nitrogen cooled temperature, the Fermi energy ϕ is in the order of 20–30. Hence, it is necessary to have the solution of Eq. (2.6) in the range of $\phi < 30$ for practical HgCdTe device simulations.

In this work, we propose a new simple carrier approximation derived from the Ariel model, adding three parameters α_1, α_2, and B_2 as follows:

$$\phi = \ln\frac{n}{B_0 N_c} + B_1\left(\frac{n}{N_c}\right)^{\alpha_1} + B_2\left(\frac{n}{N_c}\right)^{\alpha_2} \tag{2.7a}$$

Here, the expression of parameter α_1, α_2, B_1, B_2 is shown in reference [14]. Figure 2.1 shows a comparison between the actual numerical solution of Eq. 2.6 in open dots and our approximation of Eq. (2.7a) in solid lines. Also shown is the Ariel model in open triangles. As can be seen, our new approximation is in good agreement with the numerical solution in the range of $\phi < 30$ for both LWIR (Fig. 2.1a) and MWIR (Fig. 2.1b) devices. This approximation appears in the same

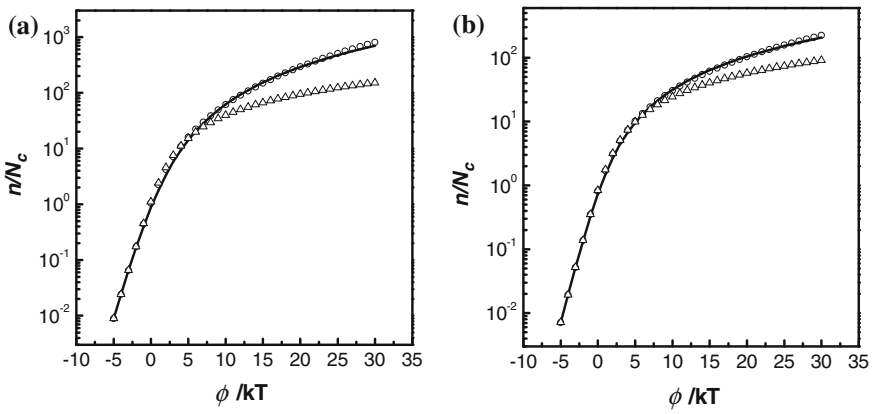


Fig. 2.1 Comparisons of carrier concentration versus reduced Fermi level in the range of $\phi < 30$ for $Hg_{0.8}Cd_{0.2}Te$ (**a**), and $Hg_{0.6}Cd_{0.4}Te$ (**b**) at 80 K calculated by actual numerical solution (*open circle*), Ariel model (*open triangle*), and the approximation Eq. (2.7a) (*solid lines*). Reprinted with permission from Ref. [14] © 2006, Elsevier

form as the Boltzmann's approximation, and is applicable to both analytic and numerical device simulation models.

The hole density p, can be written as an expression similar to Eq. (2.7a) but with generally different coefficients

$$-\phi - \varepsilon_g = \ln\frac{p}{B_{0v}N_v} + B_{1v}\left(\frac{p}{N_v}\right)^{\alpha_{1v}} + B_2 v\left(\frac{p}{N_v}\right)^{\alpha_{2v}} \tag{2.7b}$$

The valence band is parabolic which is equivalent to assuming that the non-parabolicity factor $\alpha = 0$ in Eq. (2.7a). Therefore, we get the valence band coefficient $B_{0v} = 1$, $B_{1v} = 0.944$, $B_{2v} = -0.577$, $\alpha_{1v} = 0.745$, and $\alpha_{2v} = 0.624$ from the expressions for B_0, B_1, B_2, α_1, and α_2 by putting $\alpha = 0$.

We now consider an intrinsic semiconductor where electrical neutrality requires $n = p = n_i$. Using this condition in Eqs. (2.7a) and (2.7b), an implicit expression for n_i can be obtained:

$$\frac{n_i^2}{B_0 B_{0v}}\exp\left[B_1\left(\frac{n_i}{N_c}\right)^{\alpha_1} + B_2\left(\frac{n_i}{N_c}\right)^{\alpha_2} + B_{1v}\left(\frac{n_i}{N_v}\right)^{\alpha_{1v}} + B_{2v}\left(\frac{n_i}{N_v}\right)^{\alpha_{2v}}\right] = N_c N_v \exp(-\varepsilon_g) \tag{2.8}$$

This equation can be solved iteratively for n_i if the other quantities are known.

Subtracting Eq. (2.7b) from Eq. (2.7a) and using the charge neutrality condition, we obtain the following expression for the intrinsic Fermi level ϕ_i of a non-parabolic semiconductor:

$$\phi_i = -\frac{\varepsilon_g}{2} + \frac{1}{2}\ln\left(\frac{B_{0v}N_v}{B_0N_c}\right) + \frac{1}{2}\left[B_1\left(\frac{n_i}{N_c}\right)^{\alpha_1} + B_2\left(\frac{n_i}{N_c}\right)^{\alpha_2} - B_{1v}\left(\frac{n_i}{N_v}\right)^{\alpha_{1v}} - B_{2v}\left(\frac{n_i}{N_v}\right)^{\alpha_{2v}}\right] \tag{2.9}$$

With known n_i, ϕ_i, and with the help of Eqs. (2.7a), (2.7b), (2.8), (2.9), we derive new relations that describe the Fermi level in terms of carrier densities and intrinsic properties:

$$\phi_n - \phi_i = \ln\left(\frac{n}{n_i}\right) + B_1\left[\left(\frac{n}{N_c}\right)^{\alpha_1} - \left(\frac{n_i}{N_c}\right)^{\alpha_1}\right] + B_2\left[\left(\frac{n}{N_c}\right)^{\alpha_2} - \left(\frac{n_i}{N_c}\right)^{\alpha_2}\right] \tag{2.10a}$$

$$\phi_i - \phi_p = \ln\left(\frac{p}{n_i}\right) + B_{1v}\left[\left(\frac{p}{N_v}\right)^{\alpha_{1v}} - \left(\frac{n_i}{N_v}\right)^{\alpha_{1v}}\right] + B_{2v}\left[\left(\frac{p}{N_v}\right)^{\alpha_{2v}} - \left(\frac{n_i}{N_v}\right)^{\alpha_{2v}}\right] \tag{2.10b}$$

The expressions (2.10a, 2.10b) can be solved iteratively for Fermi level E_{fn} and E_{fp}, which are the Fermi energy level of the n-side and the p-side without contact with each other. So we can obtain the built-in potential $V_{bi} = E_{fn} - E_{fp}$ of the p-n junction.

Therefore, with the help of Eqs. (2.7a, 2.7b), non-parabolicity and carrier degeneracy can be adequately modeled by the calculation of built-in potential V_{bi}, intrinsic carrier concentration n_i and intrinsic Fermi level E_i.

Different material compositions can therefore be described by using different carrier density models. In this section, we analyze four carrier density models by considering (I) both the carrier degeneracy and the conduction band non-parabolicity, (II) only the carrier degeneracy, (III) only the conduction band non-parabolicity, and (IV) neither of these two factors.

In Model II, only carrier degeneracy is considered, assuming a parabolic conduction band. The carrier density is given by the following integral expression:

$$n = \frac{2N_c}{\sqrt{\pi}}\int_0^\infty \frac{\sqrt{\varepsilon}}{1 + \exp(\varepsilon - \phi)}\mathrm{d}\varepsilon \tag{2.11}$$

Equation (2.6) will be the same as Eq. (2.11) if the non-parabolicity factor α equals to zero. According to Eq. (2.7a), we can obtain the carrier approximation of Model II:

$$\phi = \ln\frac{n}{N_c} + 0.994\left(\frac{n}{N_c}\right)^{0.745} - 0.577\left(\frac{n}{N_c}\right)^{0.624} \tag{2.12a}$$

The hole density approximation is given by

$$-\phi - \varepsilon_g = \ln\frac{p}{N_v} + 0.994\left(\frac{p}{N_v}\right)^{0.745} - 0.577\left(\frac{p}{N_v}\right)^{0.624} \tag{2.12b}$$

In Model III, the distribution function $f(E)$ can be a Boltzmann function because of carrier non-degeneracy, and only conduction band non-parabolicity is considered. The carrier density is given by the following integral expression:

$$n = \frac{2N_c}{\sqrt{\pi}}\int_0^\infty \varepsilon^{1/2}(1+\alpha\varepsilon)^{1/2}(1+2\alpha\varepsilon)\exp(\phi-\varepsilon)\mathrm{d}\varepsilon \tag{2.13}$$

The approximation of Eq. (2.13) can be given by a modified Boltzmann approximation [18]:

$$n = B_0 N_c \exp(\phi) \tag{2.14a}$$

Likewise, the hole density approximation is given by

$$p = B_{0v} N_v \exp(-\phi - \varepsilon_g) \tag{2.14b}$$

In Model IV, neither carrier degeneracy nor conduction band non-parabolicity is considered. The carrier density is given by the following integral expression:

$$n = \frac{2N_c}{\sqrt{\pi}}\int_0^\infty \sqrt{\varepsilon}\cdot\exp(\phi-\varepsilon)\cdot\mathrm{d}\varepsilon \tag{2.15}$$

The carrier density can be given by a conventional Boltzmann approximation

$$n = N_c \exp(\phi) \tag{2.16a}$$

The hole density approximation is given by

$$p = N_v \exp(-\phi - \varepsilon_g) \tag{2.16b}$$

We present comparisons among the four approximations for carrier density versus Fermi energy in Fig. 2.2. As can be seen from the figure, for both $Hg_{0.8}Cd_{0.2}Te$ (a) and $Hg_{0.6}Cd_{0.4}Te$ (b), comparing with Model I (considering both carrier degeneracy and conduction band non-parabolicity), Model II (considering only carrier degeneracy) underestimates the carrier density, while Model III (considering only conduction band non-parabolicity) and Model IV (considering neither of them) yield significantly larger carrier density at the range of $\phi > 0$ [14]. When Cd composition x changes from 0.2 to 0.4, the difference between Model I and

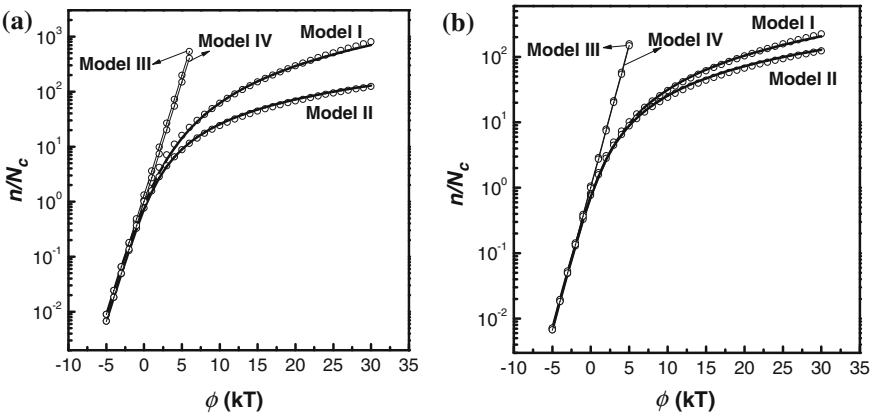


Fig. 2.2 Calculated carrier concentration versus reduced Fermi level with Model I (considering both carrier degeneracy and conduction band non-parabolicity), Model II (considering only carrier degeneracy), Model III (considering only conduction band non-parabolicity), and Model IV (considering neither of them) for $Hg_{0.8}Cd_{0.2}Te$ (**a**), and $Hg_{0.6}Cd_{0.4}Te$ (**b**) at 80 K (*solid lines*). The *open circles* are respective actual numerical solutions. Reprinted with permission from Ref. [14] © 2006, Elsevier

Model II decreases because the increased band gap reduces the non-parabolicity factor α. In addition, one can also see in Fig. 2.2 that the calculations using the four approximations (solid lines) accurately fit the corresponding numerical solutions (open circles) obtained by solving Eqs. (2.6), (2.11), (2.13), and (2.15), respectively, in the range of $\phi < 30$. This indicates that these models can be used in place of a numerical solution to investigate the effects of carrier degeneracy and conduction band non-parabolicity on the simulation of HgCdTe photovoltaic devices.

To further verify the applicability of the above approximations, the measured R_d–V curves from real HgCdTe devices are fitted with the analytic model for device simulation (see Appendix) by using the four models. The samples are LWIR ($x = 0.2323$) and MWIR ($x = 0.2927$) n-on-p diodes fabricated using ion implantation [21]. Their material and device parameters, which are not used as fitting parameters, are summarized in Table 2.1, where x is the Cd composition, T is the measuring temperature, A is the junction area, N_a is the dopant densities of the p-side, and μ_p is the mobility of excess holes.

The fitting parameters are listed in Tables 2.2 and 2.3 for Sample I and Sample II [14], respectively. For Sample II ($x = 0.2927$), the fitting parameter μ_n/τ_n is absent because the diffusion current is insignificant in the dark current for MWIR devices

Table 2.1 Material and device parameters of the two HgCdTe photovoltaic samples

	x	T (*K*)	A (cm^2)	N_a (cm^{-3})	μ_p (cm^2/V s)
Sample I	0.2323	80	2.5×10^{-5}	5.65×10^{15}	813
Sample II	0.2927	80	2.5×10^{-5}	4.66×10^{15}	579

Reprinted with permission from Ref. [14] © 2006, Elsevier

Table 2.2 Fitting parameters obtained by fitting the four models to the measured R_d–V curves of Sample I

Model	μ_p/τ_n ($cm^2/V\ s^2$)	τ_0 (ns)	N_d (cm^{-3})	E_t/E_g	N_t (cm^{-3})
I	1.613×10^{14}	0.154	1.493×10^{16}	0.486	1.659×10^{12}
II	2.128×10^{14}	0.121	1.478×10^{16}	0.488	1.624×10^{12}
III	1.096×10^{14}	0.142	1.564×10^{16}	0.488	1.721×10^{12}
IV	1.248×10^{14}	0.119	1.543×10^{16}	0.488	1.722×10^{12}

Reprinted with permission from Ref. [14] © 2006, Elsevier

Table 2.3 Fitting parameters obtained by fitting the four models to the measured R_d–V curves of Sample II

Model	τ_0 (ns)	N_d (cm^{-3})	E_t/E_g	N_t (cm^{-3})
I	1.488	8.649×10^{17}	0.586	1.302×10^{13}
II	1.389	4.759×10^{17}	0.589	1.174×10^{13}
III	1.492	1.0×10^{19}	0.581	4.398×10^{13}
IV	1.492	1.0×10^{19}	0.581	4.398×10^{13}

Reprinted with permission from Ref. [14] © 2006, Elsevier

at 80 K. It is evident from these tables that these parameters from the four models are different from each other, and the sensitivity of the fitted values of device parameters to the used model is obvious. This implies that the omission of carrier degeneracy and conduction band non-parabolicity brings deviations into the simulation of the HgCdTe device, and the magnitude of the deviations is dependent on the parameters of device and material.

In Fig. 2.3, the solid line denotes the theoretical calculation by using Model I, and the open circles denote the measured data for $x = 0.2323$ (a) and $x = 0.2927$ (b). It can be seen that the theoretical calculation fits the experimental data well. In

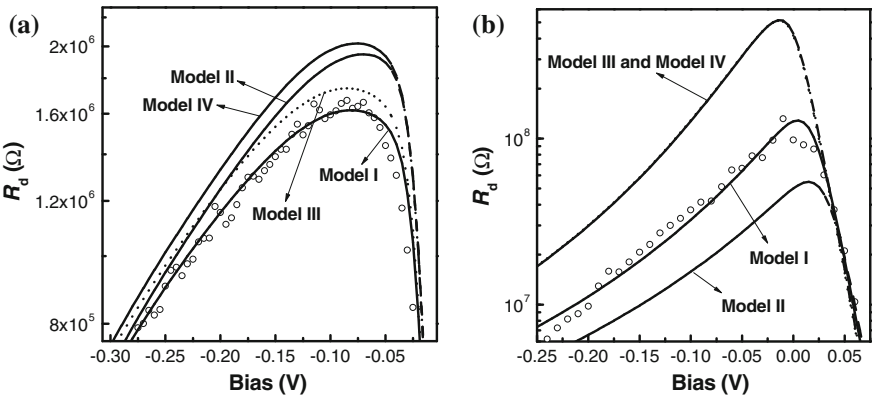


Fig. 2.3 Fitting curves of Model I (*solid lines*) with experimental R_d-V characteristics (*open circles*) and the calculations of Model II (*dashed lines*), III (*dotted lines*) and IV (*dash-dotted lines*) using the parameters in Table 2.1 and the parameters from Model I in Tables 2.2 and 2.3 for Sample I (**a**) and Sample II (**b**) at 80 K. Reprinted with permission from Ref. [14] © 2006, Elsevier

this work, we consider that Model I is a perfect model for HgCdTe, and the fitting parameters from Model I are the authentic parameters for HgCdTe devices. To estimate the deviation of R_d–V characteristics calculated by Model II, III, and IV from those calculated using Model I, all the parameters in Table 2.1 and the parameters for Model I in Tables 2.2 and 2.3 are also employed in the further calculations using Model II, III, and IV. The calculated results are also shown in Fig. 2.3. It is evident from the figure that, compared to Model I, Model IIII predicts larger R_d for LWIR ($x = 0.2323$) and smaller R_d for MWIR ($x = 0.2927$), while Model III and Model IV yield larger R_d for both LWIR and MWIR.

In the above calculations, the differences of the built-in potential $V_{bi,}$ the intrinsic carrier concentration n_i and the intrinsic Fermi level E_i among the three approximations are responsible for the differences in the above three R_d–V curves. It is via these three parameters that the carrier degeneracy and the conduction band non-parabolicity influence the simulation calculations. Therefore, we have calculated V_{bi}, n_i and E_i with various device parameters such as composition and doping level in Model II, III, and IV to estimate their deviation from Model I so that the dark current deviation of Model II, III, and IV from Model I can be further evaluated.

For the purpose of brevity and clarity, the relative deviations of V_{bi}, n_i, and E_i are hereafter referred to simply as A_j/A_1 (A denotes V_{bi}, n_i and E_i, and subscript $j = 1$–4 stands for the four models, respectively). n_i and E_i are dependent on the composition and temperature of the device. Since HgCdTe devices usually work at liquid nitrogen cooled temperatures, we consider a temperature of 80 K in this work. We carried out the calculations for E_{ij}/E_{i1} and n_{ij}/n_{i1} in the composition range of $x = 0.2$–0.4, and the results are shown in Fig. 2.4. Figure 2.5 shows the plots of V_{bij}/V_{bi1} versus the composition x (a) and the doping density of the n-side, N_d (b). The deviations of V_{bij}/V_{bi1} from unity increase with the decreasing x and increasing N_d. The detailed analysis of Figs. 2.4 and 2.5 is shown in the paper of Ref. [14].

Tables 2.4, 2.5, and 2.6 summarize the relative deviations, comparing with Model I, of dark currents in various mechanisms obtained from Model II, III and IV, respectively. For SWIR and MWIR with light doping, the effects of carrier degeneracy and conduction band non-parabolicity on the simulation may be omitted. For LWIR with light doping or those devices with heavy doping, the omission of carrier degeneracy and conduction band non-parabolicity will lead to enormous error in the calculation of the dark current. By using Model I, an accurate simulation can be obtained.

2.2.1.2 Heterostructure Models

The HgCdTe heterojunction has greatly improved the performance of HgCdTe photovoltaic detectors. However, the potential barrier of minority carriers in the heterojunction, prevents the transport of photogenerated minority carriers through the junction, which leads to a decrease in device performance [22–26]. By adjusting the design parameters of the heterojunction, the potential barrier could be decreased

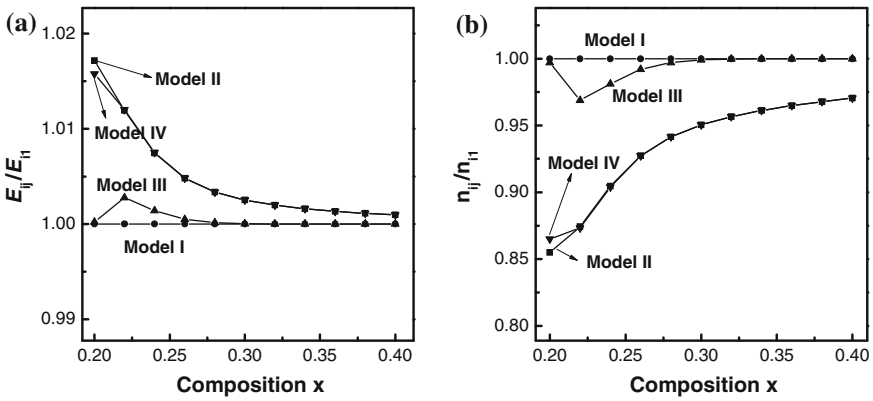


Fig. 2.4 Relative deviations of Model II, III and IV against Model I in intrinsic Fermi level E_{ij}/E_{i1} (**a**) and intrinsic carrier concentration n_{ij}/n_{i1} (**b**) at various composition x (subscript j = 1–4 denotes four models, respectively). Reprinted with permission from Ref. [14] © 2006, Elsevier

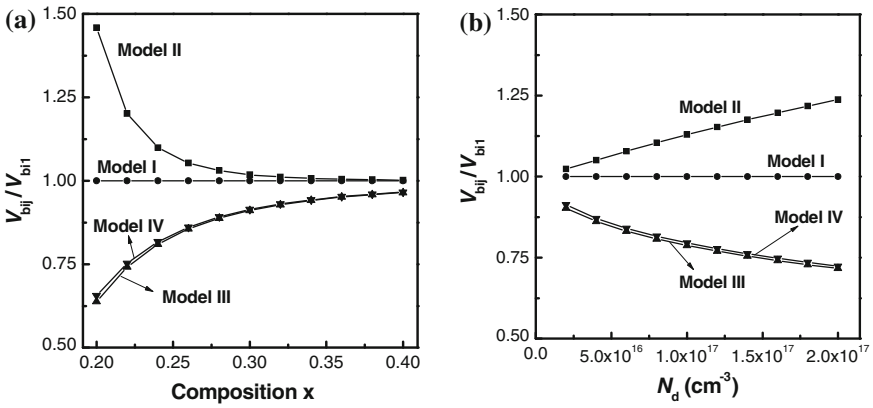


Fig. 2.5 Relative deviations of Model II, III and IV against Model I in built-in potential V_{bij}/V_{bi1} at various composition x (**a**) and various n-side doping density N_d (**b**) (subscript j = 1–4 denotes the four models, respectively). Reprinted with permission from Ref. [14] © 2006, Elsevier

Table 2.4 Relative deviations in various dark currents of Model II (considering only carrier degeneracy) from Model I (considering both carrier degeneracy and conduction band non-parabolicity)

	Light doping		Heavy doping	
	SWIR	LWIR	SWIR	LWIR
I_{diff}	U (<10 %)	U (<30 %)	U (<10 %)	U (<30%)
$I_{g\text{-}r}$	U (<5 %)	U (<20 %)	U (<8 %)	U (<50%)
I_{TAT}	Zero	O (<10 %)	O (<100 %)	O (<350%)
I_{BBT}	Zero	O (<30 %)	O (<1000 %)	O (<5000%)

(*U* denotes 'underestimate', and *O* denotes 'overestimate'.) Reprinted with permission from Ref. [14] © 2006, Elsevier

Table 2.5 Relative deviations in n_i, V_{bi} and various dark currents of Model III (considering only conduction band non-parabolicity) from Model I (considering both carrier degeneracy and conduction band non-parabolicity)

	Light doping		Heavy doping	
	SWIR	LWIR	SWIR	LWIR
I_{diff}	Zero	U (<6 %)	Zero	U (<6 %)
$I_{g\text{-}r}$	Zero	U (<3 %) or O (<10 %)	O (<10 %)	U (<3 %) or O (<60 %)
I_{TAT}	Zero	U (<18 %)	U (<90 %)	U (<99 %)
I_{BBT}	Zero	U (<45 %)	U (<99 %)	U (<99 %)

(*U* denotes 'underestimate', and *O* denotes 'overestimate'). Reprinted with permission from Ref. [14] © 2006, Elsevier

Table 2.6 Relative deviations in n_i, V_{bi} and various dark currents of Model IV (considering neither degeneracy nor non-parabolicity) from Model I (considering both carrier degeneracy and conduction band non-parabolicity)

	Light doping		Heavy doping	
	SWIR	LWIR	SWIR	LWIR
I_{diff}	U (< 10 %)	U (<30 %)	U (<10 %)	U (<30 %)
$I_{g\text{-}r}$	U (< 5 %)	U (<15 %) or O (<8 %)	U (<5 %) or O (<10 %)	U (<15 %) or O (<60 %)
I_{TAT}	Zero	U (<15 %)	U (<95 %)	U (<99 %)
I_{BBT}	Zero	U (<35 %)	U (<99 %)	U (<99 %)

(*U* denotes 'underestimate', and *O* denotes 'overestimate'). Reprinted with permission from Ref. [14] © 2006, Elsevier

and even be eliminated. However, there are many parameters for HgCdTe and the relationships between them are very complicated. Therefore, a reliable theoretical method is urgently needed for the design of the band structure of heterojunctions [27].

Much research has been done in this field. Migliorato and White [28] considered the interdiffusion effect of both composition and doping, and calculated the band structure of HgCdTe/CdTe heterojunction. Bratt and Casselman [29] developed the MW model, which can be suitable for calculating the heterojunction with different composition of $Hg_xCd_{1-x}Te$. However, these two models are both based on common anion rule, namely assuming that there was no discontinuity of the valence band, and their calculations both neglected the degeneracy of the carrier and the non-parabolic conduction band of HgCdTe. Then, after Madrasz and Szmulowiz [30–32] introduced the Fermi–Dirac integral based on these two models, the following characteristics of HgCdTe can be included in the calculation of band structure: (1) the carrier degeneracy; (2) the non-parabolic conduction band; (3) the valence band discontinuity. However, the Fermi–Dirac integral without any approximation greatly increases the total computing time, which limits the use of the model and makes it only suitable for the solution of the Poisson equation under

equilibrium. The research of Bratt and Casselman [29] demonstrated that the band structure of heterojunctions can be optimized by controlling the positions of the *p-n* junction and the Cd composition junction, but the work of Madrasz and Szmulowiz [30–32] assumed that the position of the *p-n* junction coincides with that of the Cd composition junction and ignored the effect of relative position between the *p-n* junction and the Cd composition junction on the band structure. Moreover, recent work indicated that there was charge at the interface of a heterojunction [33, 34]. It was stated that the interface charge was usually caused by a discontinuity within the lattice and the process technology. Though the lattice discontinuity between HgCd and CdTe is less than 0.3 %, the fixed charge also exists at the interface because of the different polar intensities of HgCd and CdTe. Therefore, the density of interface charge, which can be either positive or negative, is very hard to be controlled and extracted in general cases [33]. The density of charge at the interface between HgCd and CdTe is about $\pm 1 \times 10^{12}$ cm^{-2}. However, the density of charge at the interface of HgCdTe with different compositions varies linearly with the gradient of the Cd composition [35]. The influence of this magnitude of interface charge on band structure is not ignorable, so it is very necessary to study the influence of interface charge in the design of a heterojunction structure. In this section, the following three aspects are considered in the physical model: (1) the carrier degeneracy; (2) the non-parabolic conduction band; (3) the valence band discontinuity. Based on the approximation model of carriers in Sect. 2.2.1.1, the Fermi–Dirac integral can be calculated rapidly. Therefore, it not only satisfies the demand for computational speed, but also enhances the accuracy of the physical model. As a result, a feasible method can be proposed for weakening the effect of interface charge.

In the Fermi–Dirac integral calculation, z is set for the location, $0 < z < L$, L is the length along the growth orientation of the device, and $L = 4$ μm. Assuming that the valance band at $z = 0$ is a zero energy point, then the Poisson equation can be expressed as:

$$\nabla^2\psi = -\frac{q}{\varepsilon}\left[p - n + N_{\mathrm{d}}^{+} - N_{\mathrm{a}}^{-} + Q_{\mathrm{e}}\right] - \frac{1}{\varepsilon}\nabla\psi\nabla\varepsilon \tag{2.17}$$

where Ψ is electrostatic potential, ε is the permittivity, p is the hole density, n is the electron density, N_{d}^{+} is the ionized donor density, N_{a}^{-} is the ionized acceptor density, Q_{e} is the interface charge profile. It is assumed that the donor and acceptor are both fully ionized during the calculation.

The boundary conditions can be expressed as:

$$\psi(0) = 0 \tag{2.18}$$

$$\psi(L) = \phi(L) - \phi(0) + \Delta E_{\mathrm{v}}(L) \tag{2.19}$$

where ϕ is counted by iteration according to formula (2.3) and charge neutrality condition, ΔE_{v} is the valence band discontinuity.

Assumed that ΔE_v^0 is the valence band discontinuity of the HgTe/CdTe heterojunction [36, 37], and $\Delta E_v^0 = 0.35$ eV. Since the valence band discontinuity of the HgCdTe/HgCdTe heterojunction is proportional to the difference of the Cd composition (Δx), $\Delta E_v = \Delta x \cdot \Delta E_v^0$. Because of the interdiffusion effect of both composition and doping, there is a composition-graded region around the composition junction and a doping-graded region around the *p-n* junction respectively.

For a graded composition junction, assuming that material A layer is followed by material B layer along the z direction, then the profile of Cd composition can be described by an error function of z:

$$x(z) = x_B + 0.5 \times (x_A - x_B) \times \mathrm{erfc}\left(\frac{z - z_c}{0.5W_c}\right) \tag{2.20}$$

where x_A and x_B are the Cd composition of A and B HgCdTe material respectively, W_c is the width of graded composition, z_c is the location of composition junction, in the later calculations, z_c is fixed at $z = 2.0$ μm. $z = 0$ is the reference point, $\Delta E_v = \Delta x \cdot \Delta E_v^0$, then, the discontinuity of valence band along z direction $\Delta E_v(z) = [x(0) - x(z)] \cdot \Delta E_v^0$, put formula (2.20) into it:

$$\Delta E_v(z) = (x_A - x_B) \cdot \Delta E_v^0 \left[1 - \frac{1}{2}\mathrm{erfc}\left(\frac{z - z_c}{0.5W_c}\right)\right] \tag{2.21}$$

The doping profile along the z direction can be described by an error function of the coordinate z:

$$N(z) = N(L) + \frac{N(0) - N(L)}{2} \cdot \mathrm{erfc}\left(\frac{z - z_B}{0.5W_B}\right) \tag{2.22}$$

where $N = N_d - N_a$, W_B is the width of graded doping, and z_B is the location of *p-n* junction. In the design of a HgCdTe heterojunction, W_B is 0.01 μm.

The profile of the interface charge along the z direction is described by a Gauss function:

$$Q_e(z) = Q \cdot \exp\left[-\frac{(z - z_Q)^2}{2\sigma^2}\right] \tag{2.23}$$

In the design of a HgCdTe heterojunction, set the maximum density (Q) of interface charge as a parameter for calculation. The interface charge is mainly produced from the lattice discontinuity between two materials on both sides of the heterojunction. Therefore, the profile of the interface charge is related to the location and width of the composition junction. Assuming that the center position of the function (z_Q) coincides with the position of the composition junction (z_c), then the full width half maximum (σ) is equal to $0.25W_c$.

Since the p and n in Eq. (2.17) are a function of the variable Ψ, the Poisson equation is nonlinear. However, the Poisson equation can be disposed discretely and linearly by the finite difference method, and then it can be solved by Newton iteration.

The valance band and conduction band can be expressed as:

$$E_{\mathrm{v}}(z) = \Delta E_{\mathrm{v}}(z) - q \cdot \psi(z) \tag{2.24}$$

$$E_{\mathrm{c}}(z) = E_{\mathrm{v}}(z) + E_{\mathrm{g}}(z) \tag{2.25}$$

For convenience, the *p-n* junction of wider (narrower) bandgap p-type HgCdTe material on narrower (wider) bandgap n-type HgCdTe material is set as "Pn" and "pN" structure respectively.

2.2.2 *Design of Heterojunctions HgCdTe Infrared Detectors*

The mainstream of HgCdTe detectors nowadays are homojunctions. However, HgCdTe devices tend to be manufactured by heterojunctions with the evolution of fabrication processes. Wide band gap material can decrease the tunneling rate of carriers through the junction, which results in the decrease of the dark current and noise of devices [23]. Therefore, the performance of a device with a heterojunction structure can be greatly improved. However, the structure of heterojunction is more complicated than homojunction, which makes the analysis of heterojunction devices become more difficult. For heterojunction devices, it is difficult for us to analyze the device's performance by using solely an analytical model. Therefore, a numerical model is needed to be proposed for design of heterojunction devices.

2.2.2.1 The Calculation of Band Structure in Heterojunction Devices

The influence of the *p-n* junction location (z_{B}) on the band structure is shown in Figs. 2.6 and 2.7. The calculated results are in good agreement with Figs. 1 and 5 in Ref. [28], which demonstrates the accuracy of the approximation Eqs. (2.7a, 2.7b). The comparison of the band structures (with no valance discontinuity) with different z_{B} is shown in Fig. 2.8a, it is shown that the potential barrier decreases and finally disappears when the *p-n* junction enters into narrow band gap material. This conclusion is in good agreement with BC model. However, the effect of valance discontinuity on the band structure is shown in Fig. 2.8b. It is shown that a potential well appears at the same time, and the potential well increases when the *p-n* junction enters into narrow band gap material. The appearance of the potential well also degrades the detector performance, so the entrance of the *p-n* junction into narrow band gap material should be avoided in this situation.

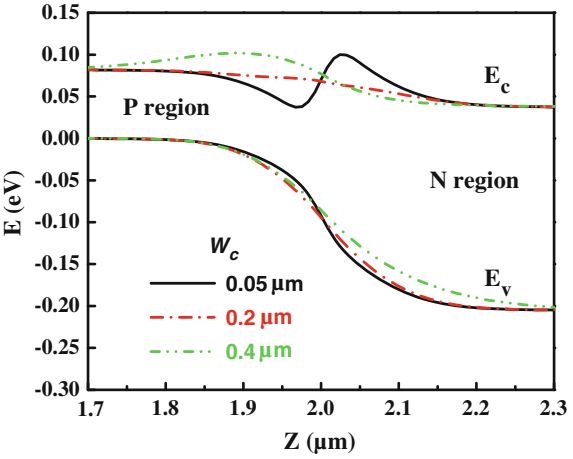


Fig. 2.6 The band structure profile of pn structure with different composition graded width W_c at 77 K. Where $x_n = 0.3$, $x_p = 0.2$, $N_a = N_d = 5 \times 10^{15}$ cm^{-3}. Redrawn from Ref. [27], Copyright 2006, The Chinese Physical Society

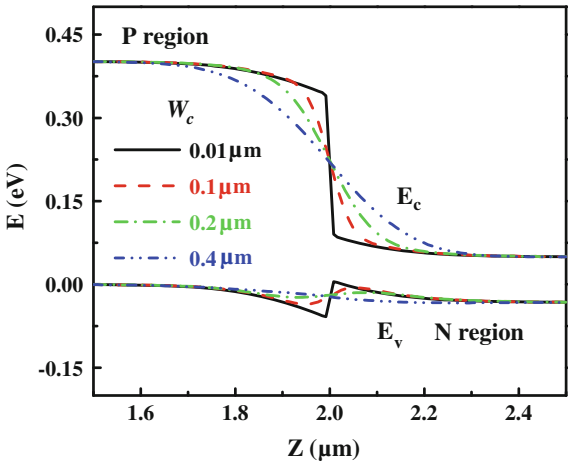


Fig. 2.7 The band structure profile of Pn structure with different composition graded width W_c at 77 K. Where $x_n = 0.2$, $x_p = 0.4$, $N_a = N_d = 1 \times 10^{15}$ cm^{-3}. Redrawn from Ref. [27], Copyright 2006, The Chinese Physical Society

For the Pn structure, Figs. 2.8 and 2.9 illustrate that the potential barrier of valance band will decrease, and a potential well will appear when the *p-n* junction enters into narrow band gap region. When the *p-n* junction enters into a wide band gap region, the potential well of valance band becomes shallow, and the height of potential barrier increases. For the pN structure, when the *p-n* junction enters into narrow band gap region, the potential barrier of conduction band decreases, and a potential well appears. When the *p-n* junction enters into a wide band gap region, the potential well of the conduction band becomes shallow, and the height of potential barrier increases.

(1) The influence of the interface charge density (Q) on the band structure in pn device is shown in Fig. 2.10. When the interface charge is positive, the potential barrier of conduction band electrons decreases. When interface

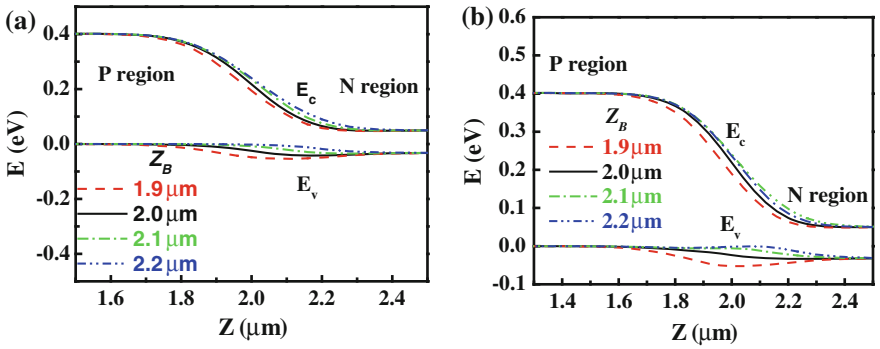


Fig. 2.8 The band structure profile of Pn structure with different *p-n* junction location (z_B) at 77 K, Where $x_n = 0.2$, $x_p = 0.4$, $N_a = N_d = 1 \times 10^{15}$ cm^{-3}, $W_c = 0.4$ μm. **a** The valance discontinuity $\Delta E_v^0 = 0$ eV, **b** The valance discontinuity $\Delta E_v^0 = 0.35$ eV. Redrawn from Ref. [27], Copyright 2006, The Chinese Physical Society

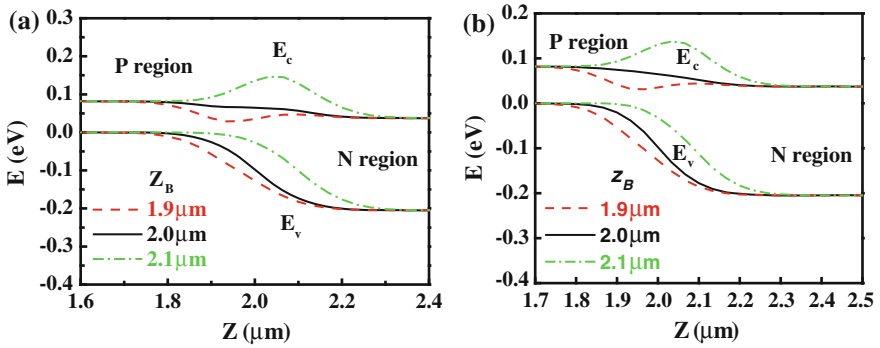


Fig. 2.9 The band structure profile of pN structure with different *p-n* junction location (z_B) at 77 K, Where $x_n = 0.3$, $x_p = 0.2$, $N_a = N_d = 5 \times 10^{15}$ cm^{-3}, $W_c = 0.2$ μm. **a** The valance discontinuity $\Delta E_v^0 = 0$ eV, **b** The valance discontinuity $\Delta E_v^0 = 0.35$ eV. Redrawn from Ref. [27], Copyright 2006, The Chinese Physical Society

charge is negative, the potential barrier of electrons in the conduction band will increase. For a Pn device, the appearance of the interface charge can influence the shape of valance band. When the interface charge is positive, the potential barrier of holes in the valance band will increase. When the interface charge is negative, the potential barrier of holes in the valance band will decrease, and the potential well of holes will increase.

(2) The optimization of parameter z_B and Q. Figure 2.11 shows the band profile of pN and Pn structures with different position (z_B). For the pn structure, when the interface charge is negative, the potential barrier of electron in conduction band can be decreased by adjusting the position of *p-n* junction. However, a

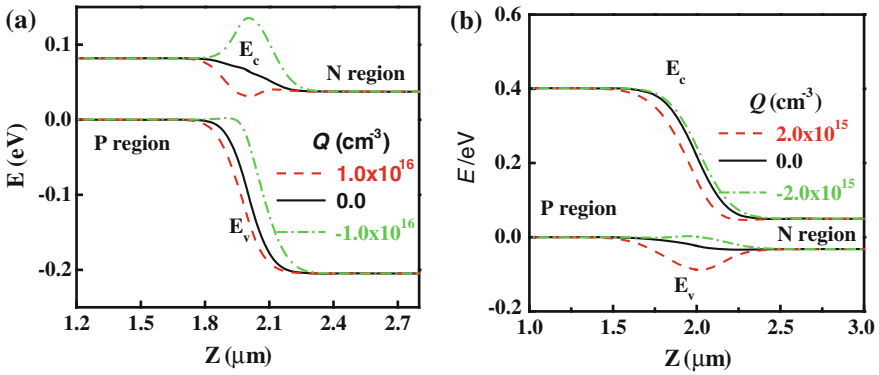


Fig. 2.10 The band profiles with different Q at 77 K. **a** pN structure, where $x_n = 0.3$, $x_p = 0.2$, $N_a = N_d = 5 \times 10^{15}$ cm^{-3}, and $W_c = 0.2$ μm, **b** Pn structure, where $x_n = 0.2$, $x_p = 0.4$, $N_a = N_d = 1 \times 10^{15}$ cm^{-3}, and $W_c = 0.4$ μm. Redrawn from Ref. [27], Copyright 2006, The Chinese Physical Society

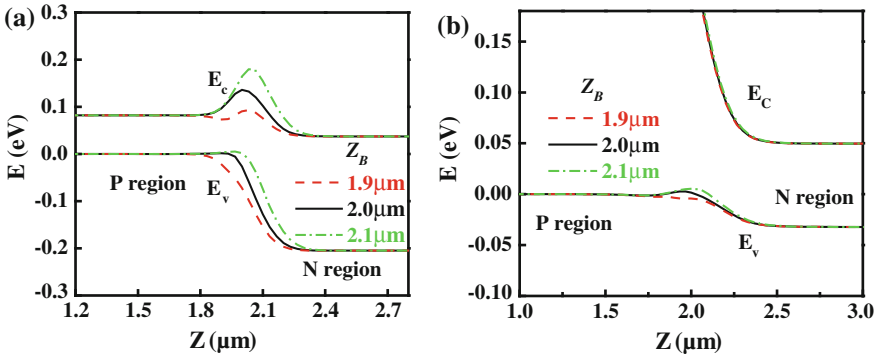


Fig. 2.11 The band profiles with different z_B at 77 K when interface charge is negative. **a** pN structure, where $x_n = 0.3$, $x_p = 0.2$, $N_a = N_d = 5 \times 10^{15}$ cm^{-3}, $W_c = 0.2$ μm, and $Q = -1 \times 10^{16}$ cm^{-3}, **b** Pn structure, where $x_n = 0.2$, $x_p = 0.4$, $N_a = N_d = 1 \times 10^{15}$ cm^{-3}, $W_c = 0.4$ μm, and $Q = -2 \times 10^{15}$ cm^{-3}. Redrawn from Ref. [27], Copyright 2006, The Chinese Physical Society

potential well will appear when the potential barrier decreases to a certain value. For the Pn structure, the potential well of holes in the valance band can also be decreased by adjusting the position of *p-n* junction.

When the interface charge is positive, the band profile of pN and Pn structures with different position (z_B) is shown in Fig. 2.12. For the pN structure, the potential well of electrons in the conduction band can be decreased by adjusting the position of *p-n* junction. For the Pn structure, the potential barrier of hole in valance band can also be decreased by adjusting the position of *p-n* junction. However, when the potential barrier decreases to a certain value, a potential well will appear.

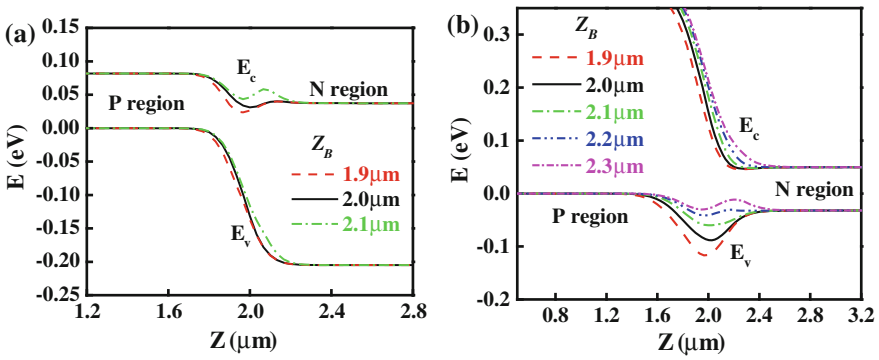


Fig. 2.12 The band profiles with different z_B at 77 K when interface charge is positive. **a** pN structure, where $x_n = 0.3$, $x_p = 0.2$, $N_a = N_d = 5 \times 10^{15}$ cm^{-3}, $W_c = 0.2$ μm, and $Q = 1 \times 10^{16}$ cm^{-3}, **b** Pn structure, where $x_n = 0.2$, $x_p = 0.4$, $N_a = N_d = 1 \times 10^{15}$ cm^{-3}, $W_c = 0.4$ μm, and $Q = 2 \times 10^{15}$ cm^{-3}. Redrawn from Ref. [27], Copyright 2006, The Chinese Physical Society

Therefore, the effect of z_B on the band structure is likely to be relatively limited, and is determined by the density of the interface charge. When the density of the interface charge is low enough, the potential barrier or well can be eliminated by optimizing the position of *p-n* junction. However, when the interface charge density is greater than a certain value, fine-tuning the position of *p-n* junction can only decrease the potential barrier or potential well. Additionally, a further change in the position of the composition junction will generate other potential barriers or potential wells. The optimal design of parameter z_B and Q can be obtained from Figs. 2.11 and 2.12. For the pN structure, when the interface charge is negative, the position of the *p-n* junction should be near the narrow band gap material. For the *Pn* junction, when the interface charge is positive, the position of the *p-n* junction should be near the wide band gap material.

2.2.2.2 The Influence of Potential Barrier on the Device Performance

As above mentioned, the potential barrier of the heterojunction will degrade the performance of the device. The influence of the potential barrier on the transport of minority carriers in P^+-on-*n* heterojunctions was studied by experimental and theoretical methods [24]. It was discovered that when the height of potential barrier was $2kT$, the quantum efficiency decreased to 95 %, when the height of potential barrier was $4.5kT$, the quantum efficiency decreased to 50 %. However, only the optical characteristic of the device was studied for the potential barrier of the heterojunction. The electrical characteristics of the device were neglected. The influence of the position of the *pn* junction in relation to the composition junction on band structure was studied [26, 27]. In the above section, the change of graded length of composition in heterojunction can produce the potential barrier.

Therefore, this section will analyze quantitatively the influence of the graded length of the composition on the device performance.

In the previous section, Eqs. (2.7a, 2.7b) for carrier density has been established, which is very suitable for HgCdTe material. The form of this equation is similar to the Boltzmann approximation equation. Therefore, it is suitable for the analytical model, as well as the numerical model. However, as this equation is a transcendental equation, there are too many iterations in numerical calculation, which make the convergence of the solution be worse and the computing time become longer. In order to solve these problems, the approximation model of carrier density should be proposed as a form of non-transcendental equation.

For a given ϕ, many analytical approximation models have been reported to solve the density of electrons [27, 38]. The revised Boltzmann approximation model only considered the non-parabolic approximation, which is not suitable for the degenerate case. The Ehrenberg model is not suitable for the highly degenerate case, and the Sommerfeld model is only suitable for the highly degenerate case. The results of these models and the accurate numerical Eq. (2.6) are shown in Fig. 2.13. It illustrates that when $\phi < -2$, the result of the Ehrenberg model is in agreement with the numerical result, and when $\phi > 5$, the result of the Sommerfeld model is in agreement with the numerical result. In order to fit the result of the carrier approximation model with that of numerical integral in the range of $\phi < 30$, it is very necessary to modify the Sommerfeld model and add it into Ehrenberg model:

$$n = \frac{B_0^* N_c \exp(\phi)}{1 + 0.75 B_0^* B_1^* \exp(\phi)} + \frac{\sqrt{2}}{3\sqrt{\pi}} N_c [(\phi + |\phi|) \cdot (1 + \alpha\phi)]^{3/2} \tag{2.26}$$

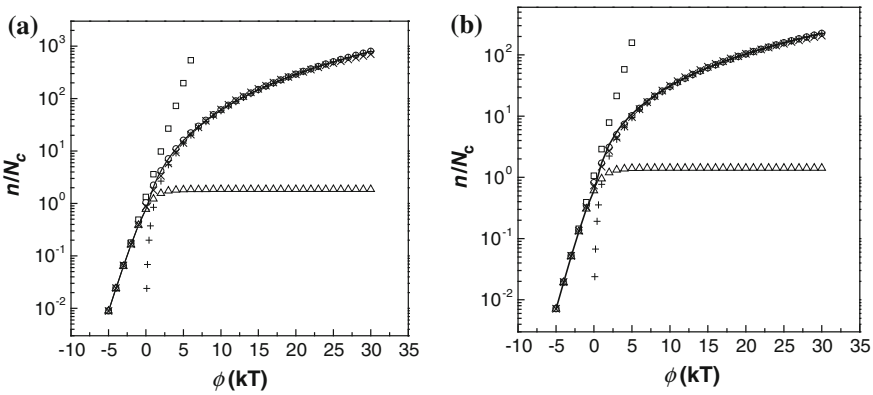


Fig. 2.13 The curve of carrier density versus Fermi level calculated by several approximation model of carrier density and numerical integral in **a** $Hg_{0.8}Cd_{0.2}Te$ and **b** $Hg_{0.6}Cd_{0.4}Te$ at 80 K. *Open circle* is solved by accurate numerical integral, *open square* is solved by the revised Boltzmann approximation, *open triangle* is solved by Ehrenberg model, *plus* is solved by Sommerfeld model, *cross* is the calculation of Eq. (2.3), *solid line* is the approximation of carrier density solved by Eqs. (2.10a, 2.10b)

The results of Eqs. (2.2), (2.3), and (2.26) are shown in Fig. 2.13. It illustrates that Eq. (2.26) is fitted well with Eqs. (2.2) and (2.3) in the range of $\phi < 30$. Therefore, Eq. (2.3) can be replaced by Eq. (2.26).

The hole density of HgCdTe can be expressed as:

$$p = \frac{2N_v}{\sqrt{\pi}} F_{1/2}(-\phi - \varepsilon_g) \tag{2.27}$$

where $F_{1/2}(\eta)$ is 1/2 of Fermi level [39]:

$$F_{1/2}(\eta) \approx \frac{2\sqrt{\pi}}{3\sqrt{\pi}a^{-3/8} + 4\exp(-\eta)} \tag{2.28}$$

where $a = \eta^4 + 33.6\eta\{1-0.68\exp[-0.17(\eta + 1)^2]\} + 50$. When η varies from $-\infty$ to $+\infty$, the relative error between the above analytical model and accurate calculation is not less than 0.4 %.

Since the form of Eqs. (2.26) and (2.27) are not similar with that of Boltzmann approximation equations, the equations are rewritten as [40]:

$$n = n_{i,\text{eff}} \cdot \gamma_n \cdot \exp\left(\frac{E_{fn} - E_i}{kT}\right) \tag{2.29}$$

$$p = n_{i,\text{eff}} \cdot \gamma_p \cdot \exp\left(\frac{E_i - E_{fp}}{kT}\right) \tag{2.30}$$

where γ_n and γ_p is the function of η_n and η_p, respectively:

$$\gamma_n = \frac{n}{N_c}\exp(-\eta_n) \tag{2.31}$$

$$\gamma_p = \frac{p}{N_v}\exp(-\eta_p) \tag{2.32}$$

$$\eta_n = \frac{E_{fn} - E_c}{kT} \tag{2.33}$$

$$\eta_p = \frac{E_v - E_{fp}}{kT} \tag{2.34}$$

According to Eqs. (2.29) and (2.30) of carrier density, a solution of the one-dimensional HgCdTe *pn* junction is obtained by finite difference method. The detailed process and method of one-dimensional simulation refers to the literature [41–43]. In order to verify the validity of program, the calculated result was compared with the solution obtained using the commercial software DESSIS. The one-dimensional schematic of HgCdTe device is shown in Fig. 2.14, and the relevant material and structural parameters are listed in Table 2.7.

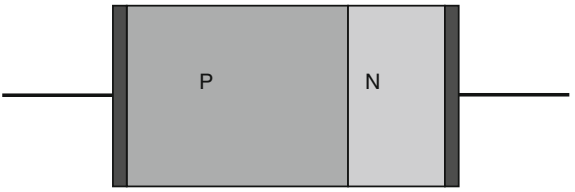


Fig. 2.14 One-dimensional simulated schematic of HgCdTe device

Table 2.7 Relevant material and structure parameters of one-dimensional HgCdTe device

Composition x	0.22
Temperature	77 (K)
SRH lifetime	10 (ns)
Doping concentration of N region	1×10^{16} (cm^{-3})
Doping concentration of P region	8×10^{15} (cm^{-3})
Thickness of N region	1 (μm)
Thickness of P region	9 (μm)
Junction area	50×50 (μm^2)

The result of locally produced software is in good agreement with that of DESSIS software, which demonstrates the validity of the program (Fig. 2.15). However, only the thermal effects, namely diffusion current and generation–recombination current are considered in the dark current aspects of this program. In narrow band gap semiconductors such as HgCdTe, the device performance is usually limited by the tunneling mechanism. Therefore, to exactly calculate the electrical characteristics of the HgCdTe heterojunction, the effects of trap-assisted tunneling and band-to-band tunneling should be added into the numerical model, where the two models are integrated into the continuity equation as generation–recombination terms. This revised recombination rate in the continuity equation of the numerical model is:

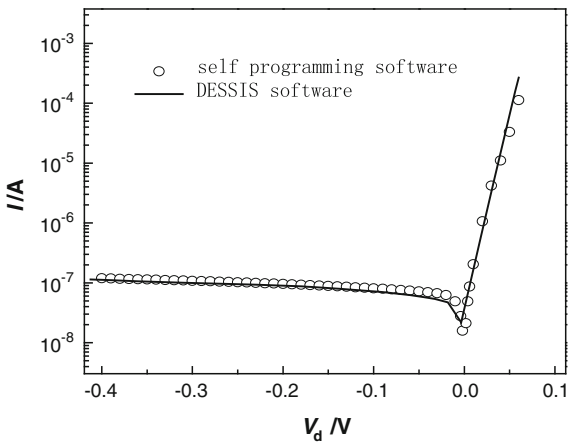


Fig. 2.15 Comparable results of locally developed software and commercial DESSIS software

$$U = U_{\text{rad}} + U_{\text{aug}} + U_{\text{SRH}} + U_{\text{tat}} + U_{\text{bbt}} \tag{2.35}$$

where U_{rad} is the Radiation recombination, U_{aug} is the Auger recombination, and U_{SRH} is the Shockley–Read–Hall (SRH) recombination [44]. U_{tat} and U_{bbt} is the recombination rate of trap-assisted tunneling and band-to-band tunneling respectively. The analytical expressions of U_{tat} and U_{bbt} are expressed as follows:

$$U_{\text{tat}} = A_{\text{tat}} \cdot E \cdot \exp\left(-\frac{B_{\text{tat}}}{E}\right) \tag{2.36}$$

$$U_{\text{bbt}} = A_{\text{bbt}} \cdot E^2 \cdot \exp\left(-\frac{B_{\text{bbt}}}{E}\right) \tag{2.37}$$

$$A_{\text{tat}} = -\frac{\pi^2 q N_{\text{t}} m_{\text{e}}^* M^2}{1.62 \cdot h^3 (E_{\text{g}} - E_{\text{t}})} \tag{2.38}$$

$$B_{\text{tat}} = \frac{1.62 \cdot \sqrt{3} E_{\text{g}}^2 F(a)}{8\sqrt{2} q P} \tag{2.39}$$

$$A_{\text{bbt}} = -\frac{q^2 \sqrt{2 m_{\text{e}}^*}}{1.62^2 \cdot 4\pi^3 \hbar^2 \sqrt{E_{\text{g}}}} \tag{2.40}$$

$$B_{\text{bbt}} = \frac{1.62 \cdot \pi \sqrt{m_{\text{e}}^*/2} E_{\text{g}}^{3/2}}{2q\hbar} \tag{2.41}$$

In the above equations, the electric field (E) of depletion region is the average electric field because the depletion region approximation is considered in the analytical model. However, the depletion region approximation is not suitable for the numerical model. Therefore, a factor of 1.62 [45] is introduced for equivalent calculating.

The R–V curve calculated by the analytical model compared with that calculated by the numerical model is shown in Fig. 2.16, where R_{sz} is the dynamic resistance of the numerical model. It is shown that the R–V curve of the analytical model is in good agreement with that of the numerical model. However, the extracted parameters of these models are listed in Table 2.8 which illustrate that there is quite a difference between them, mainly caused by the depletion region approximation.

In order to study the influence of the graded length of composition on the performance of devices, a one-dimensional program is applied for the characterization of the pN and Pn structures (see Fig. 2.7). The parameters are listed in Table 2.9. The R–V curves of pN and Pn structures with different graded lengths of composition are calculated by 1D simulation program, as shown in Fig. 2.17. The corresponding R–V curve of the homojunction is marked with the composition of 0.

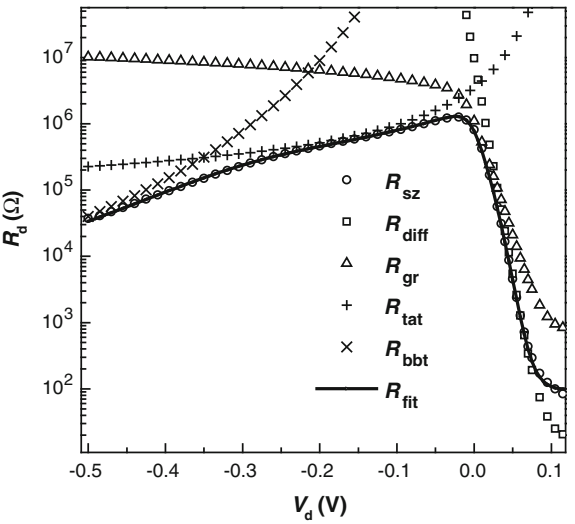


Fig. 2.16 Fitted results of *R–V* curve of self programming software with that of analytical model

Table 2.8 Parameters of HgCdTe device in analytical model and numerical model

	Analytical model	Numerical model
Composition x	0.2323	0.2323
Temperature (K)	77.3	77.3
Doping concentration of P region (cm^{-3})	9×10^{15}	9×10^{15}
Electron mobility (cm^2/V s)	20,000	20,000
Hole mobility (cm^2/V s)	600	600
Hole effective mass	0.55	0.55
Thickness of N region	1 (μm)	–
Thickness of P region	9 (μm)	–
Junction area (μm^2)	28×28	28×28
Doping concentration of N region (cm^{-3})	2×10^{16}	2.044×10^{16}
SRH lifetime in neutral region (ns)	0.5	0.606
SRH lifetime in depletion region (ns)	0.5	0.602
Trap level (E_g)	0.645	0.602
Trap concentration (cm^{-3})	2×10^{12}	1.426×10^{12}
Series resistance (Ω)	–	79.3

It demonstrates that the dynamic resistances of a pN structure with different graded lengths are almost the same under a forward bias. However, the dynamic resistance of a *Pn* heterojunction is bigger than that of homojunction. This is mainly because that the minority lifetime of the *p* region is very short, and the dark current is dominated by diffusion current under a forward bias. Diffusion current mainly is produced in the *p* region, and has almost no dependence on the junction region. Therefore, the length of graded composition has little effect on the dynamic

Table 2.9 Parameters used in one-dimensional simulation program

	Pn structure	Pn structure
Temperature (K)	77	77
Junction area (μm^2)	28 × 28	28 × 28
SRH lifetime (ns)	5	500
Trap level (E_g)	E_i	E_i
Trap concentration (cm^{-3})	2×10^{12}	2×10^{12}
Hole effective mass	0.55	0.55
Composition of *N* region *x*	0.3	0.2
Composition of *P* region *x*	0.2	0.4
Doping concentration of *N* region (cm^{-3})	5×10^{15}	1×10^{15}
Doping concentration of *P* region (cm^{-3})	5×10^{15}	1×10^{15}
Thickness of *N* region (μm)	1	9
Thickness of *P* region (μm)	9	1
Electron mobility (cm^2/V s)	1×10^5	1×10^5
Hole mobility (cm^2/V s)	600	600

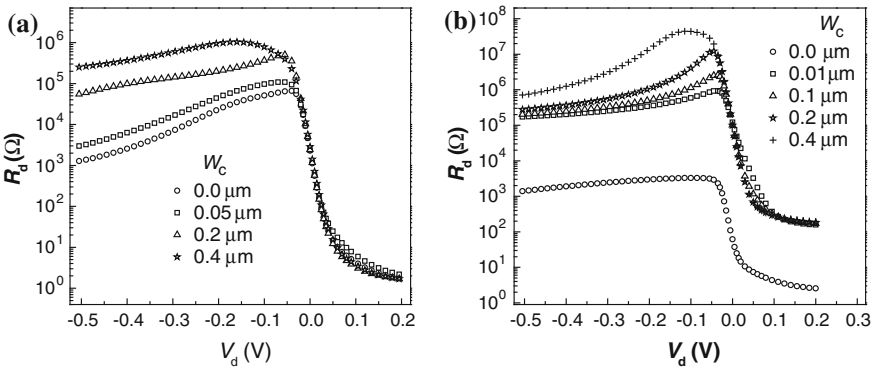


Fig. 2.17 *R–V* curve with different W_c. **a** pN structure, **b** Pn structure. $W_c = 0$ represents the result of homojunction

resistance under a forward bias. However, for a pN structure, the minority lifetime of the *p* region is long, and the dark current is dominated by recombination current under a forward bias. Therefore, the length of the graded composition has a great effect on the dynamic resistance under a forward bias. Moreover, the dynamic resistance of these two structures both increase with the length of graded composition under a reverse bias. At present, infrared detectors usually work under a reverse bias. Therefore, the longer the length of graded composition, the better is the electrical performance. However, the detectivity of the detector is not only proportional to dynamic resistance (*R*), but also to the quantum efficiency (η). In

order to study the influence of the potential barrier on device performance, quantum efficiency η, namely the optical characteristics, should be considered and calculated.

In the numerical model, the generation–recombination rate in the continuity equation can be expressed as:

$$U = U_{\text{rad}} + U_{\text{aug}} + U_{\text{SRH}} + U_{\text{tat}} + U_{\text{bbt}} - G_{\text{photo}} \tag{2.42}$$

where the photogenerated term G_{photo} is:

$$G_{\text{photo}}(z) = Q \cdot (1 - r) \cdot \alpha(z) \cdot \exp\left[- \int_0^z \alpha(t) \mathrm{d}t \right] \tag{2.43}$$

where Q is the power density of the incident light, and r is the reflectance coefficient, $\alpha(z)$ is the absorption coefficient.

The photoresponse curve of two heterojunction structures with different lengths of grade composition is shown in Fig. 2.18, where the incident light power is 0.01 W/m^2 [27]. According to the electrical characteristics shown in Fig. 2.17 and the optical characteristics shown in Fig. 2.18, the influence of the potential barrier on the device performance can be analyzed quantitatively. It is known that the detectivity of device is proportional to $\eta \cdot \sqrt{RA}$, so $\eta \cdot \sqrt{RA}$ can be used to characterize the performance of detectors. The normalized quantum efficiency (η) and dynamic resistance (R) with the length of grade composition in two heterojunction structures are shown in Fig. 2.19, where the bias voltage is −0.05 V. The normalized $\eta \cdot \sqrt{RA}$ in two heterojunction structures with the length of grade composition is shown in Fig. 2.20. For a pN structure, it demonstrates that the performance of the heterojunction device is the best when the length of grade composition is 0.2 μm. For a Pn structure, the performance of the heterojunction device is the best when the length of grade composition is 0.4 μm. This conclusion is consistent with the optimal design of band structure shown in Fig. 2.7.

2.2.3 Design of Long Wavelength HgCdTe Detectors

Currently, the study of $Hg_{1-x}Cd_xTe$ detectors is mainly based on various one-dimensional analytical models. Though the physical meaning of each variable is clear, the description of the devices is not very accurate. The real device is always a three-dimensional *pn* junction, where the lateral current and electric field cannot be analyzed by a one-dimensional model. For $Hg_{1-x}Cd_xTe$ devices with two contacts on the same side, this problem is more complex, and is beyond the theoretical framework of a one-dimensional analytical model. However, it is very complicated to establish the integrated three-dimensional model. Lateral current and the electric

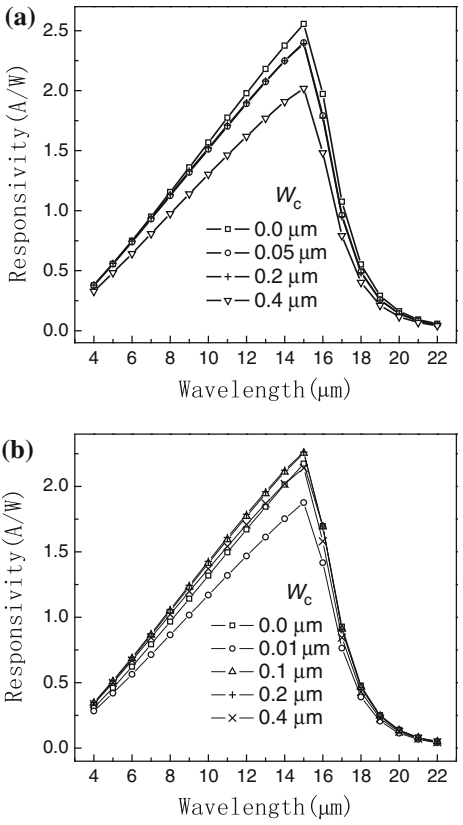


Fig. 2.18 Photo responsivity with different width of graded composition W_c. **a** Pn structure, **b** Pn structure

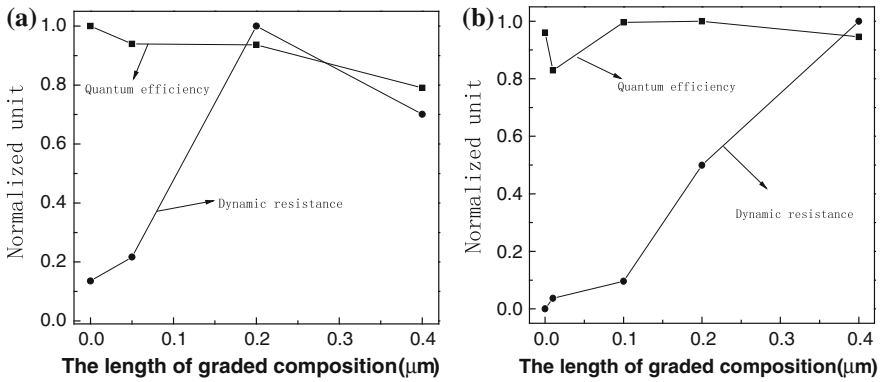


Fig. 2.19 Normalized quantum efficiency and dynamic resistance with the length of graded composition in two heterojunction structures. **a** pN structure, **b** Pn structure

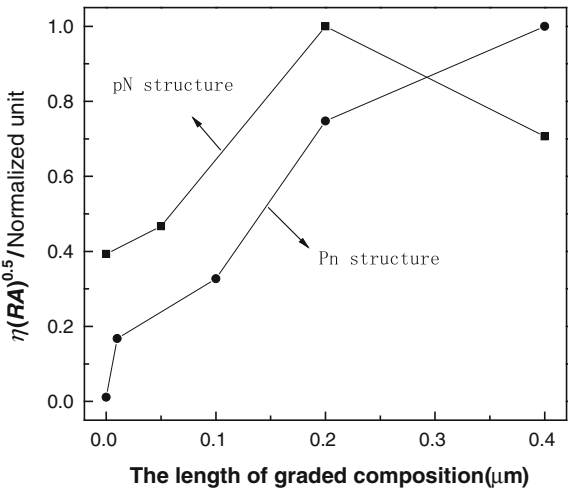


Fig. 2.20 Normalized $\eta \cdot \sqrt{RA}$ with the length of grade composition in two heterojunction structures

field can be handled by using a two-dimensional model, which overcomes the shortage of the one-dimensional model.

In this section, numerical simulation is used to study the relationship of R_0A and photocurrent with the thickness of p region, distance of contacts, doping concentration, minority lifetime and trap-assisted tunneling effect, and the influence of electron recombine lifetime and surface recombine rate on the spectral response. The spectral response is investigated experimentally and theoretically, which provides a basic reference for the optimization of device design.

The related material parameters of HgCdTe in this calculation are listed in Table 2.10, the composition of the p-type epitaxial layer is $x = 0.224$, the power density of infrared radiation is 10 W/m^2. In this calculation, the typical device parameters are listed as follows: the acceptor concentration of the p region is 10^{15} cm^{-3}, the thickness (junction depth) of the n^+ region is 1 μm, and the operating temperature is 80 K. The definite mesh of device is carefully created, which is shown in Fig. 2.21b.

Table 2.10 Related material parameters of $Hg_{1-x}Cd_xTe$ ($x = 0.224$) in calculation ($T = 80$ K)

Parameter	Value	Parameter	Value
Band gap E_g	0.122 eV	Electron mobility μ_e	5.0×10^4 cm^2/(V s)
Absorption coefficient α	2.15×10^3 cm^{-1}	Hole mobility μ_h	400 cm^2/(V s)
Effective electron mass m_e	9.15×10^{-3} m_0	Auger recombination rate G_{A1}	3.54×10^{-25} cm^6/s
Effective hole mass m_h	0.50 m_0	Radiative recombination rate G_R	2.00×10^{-10} cm^3/s
Refractive index n	3.54		

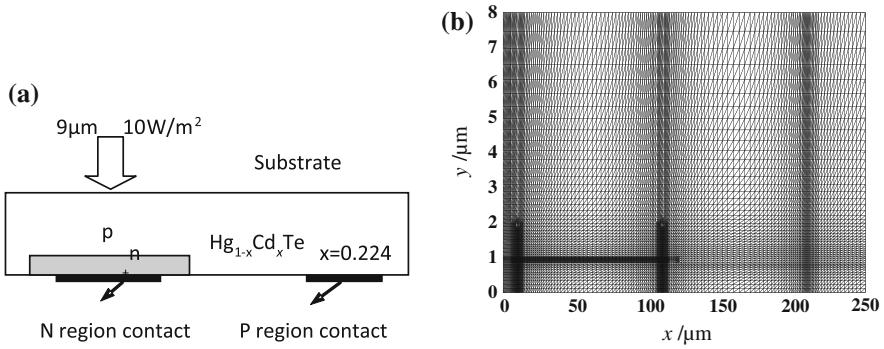


Fig. 2.21 **a** N-on-p structure of HgCdTe infrared detector, **b** The finite mesh of device

2.2.3.1 R_0A and Photocurrent

In the design of long wavelength devices, thickness of *p* region, distance of contacts, minority lifetime, doping concentration, and trap-assisted tunneling effect in a deep level have important influences on the R_0A, photocurrent, and quantum efficiency of device.

The relationship between the thickness of the *p* region and R_0A is shown in Fig. 2.22. Radiation recombination, Auger recombination, and band-to-band tunneling effects are included in the calculation. The thickness of the n^+ region is about 1 μm. The n^+ region is the heavy doping region, where the resistance is very small. In contrast, the *p* region is relatively thick and is the main absorption region under back illumination, and the resistance of it is relative large. The determination of thickness of the *p* region should consider both the light absorption and R_0A. When the thickness of the *p* region increases, the efficiency of absorption will also increase. However, the photogenerated carriers are more difficult to diffuse into the junction region, so the thickness cannot be too large. Additionally, the increase in

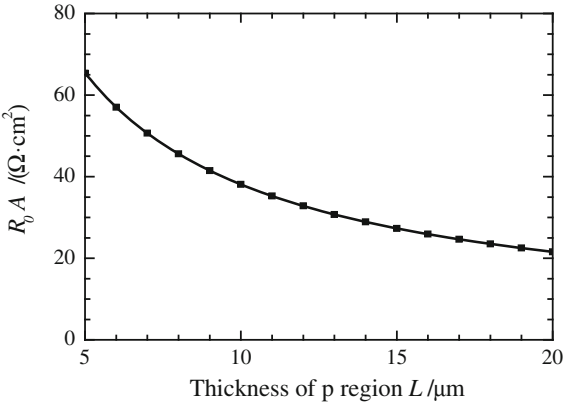


Fig. 2.22 Relationship between R_0A and the thickness of *p* region

the thickness of the p region will decrease the parallel resistance. Therefore, R_0A decreases with the thickness of p region. As known from the calculation, when the thickness of the p region varies from 7 to 20 μm, the photocurrent will change little. Therefore, the proper thickness of p region is about 10 μm. It should be noted that in the 1D model, R_0A will be nearly proportional to the thickness of the p region, which is very different than the 2D model and is not in accord with the real device. Actually, both vertical and lateral current exist in the device (see Fig. 2.1), and the distribution of current density is not uniform (Figure 2.23a, b shows the profile of current density in the device with thicknesses of 8 and 20 μm respectively). In the following calculation, the thickness of the p region is 10 μm.

The contact size of the n^+ region with heavy doping has no effect on the device performance. The resistance of p region is large. The calculated results show that the device performance with different width of contacts in p region is the same. However, the location of the contact has an important influence on both R_0A and photocurrent, as shown in Figs. 2.24 and 2.25.

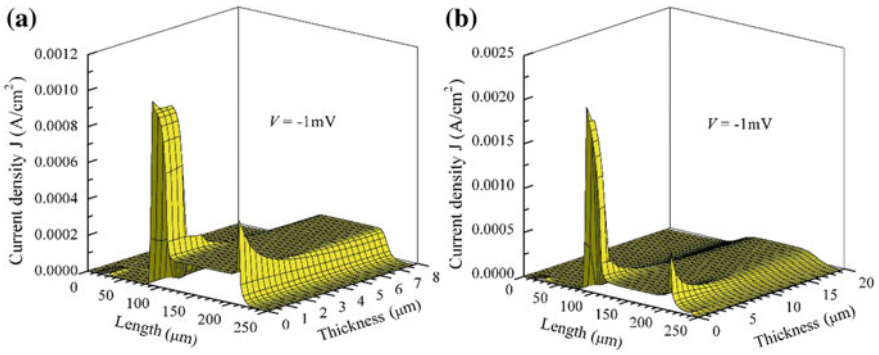


Fig. 2.23 Profile of current density with **a** the p region thickness of 8 μm and **b** the p region thickness of 20 μm

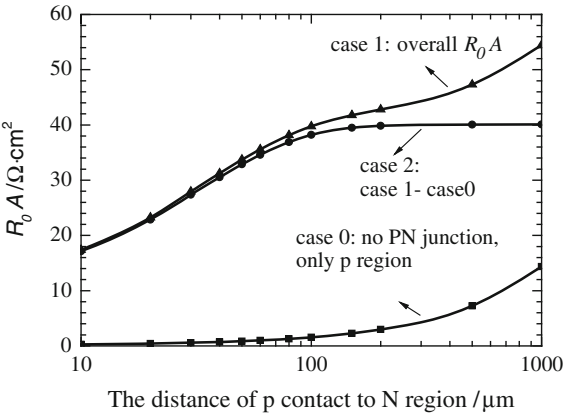


Fig. 2.24 Relationship between R_0A and the location of contact in p region

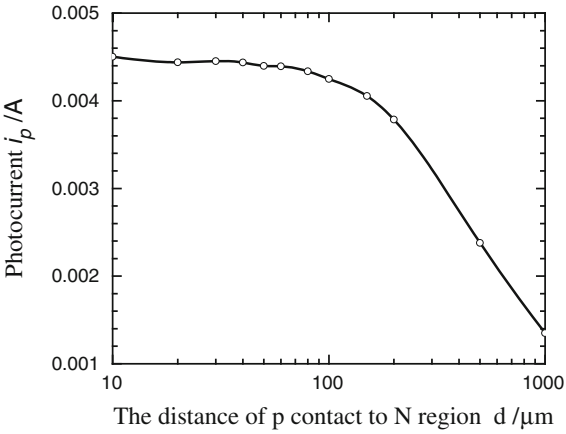


Fig. 2.25 Relationship between photocurrent (i_p) under zero bias and the location of contact in *p* region

Considering all the resistances, R_0A obviously increases with the distance of contacts (Marked as case 1). One reason is the influence of series resistance in the *p* region. If the whole device is set to a uniform p-type structure, series resistance is calculated by the same way (Marked as case 0). This shows that the resistance is approximately linear with the distance of contacts (Semi-log coordinate in Fig. 2.24). The difference between them (case2 = case1 − case0) can be treated as the R_0A of the *pn* junction without series resistance. When the distance of contacts is 100 μm, R_0A increases with the distance of contacts. And when the distance is more than 100 μm, R_0A tends to be stable. This is mainly due to the profile of the electric field and current density and is related to the location of contact in the *p* region. The current density with a contact distance of 10 μm (see Fig. 2.26) is different from that with a contact distance of 100 μm (see Fig. 2.23). The influence of the contact distance on the electric field (in the junction area) and current density is gradually weakened when the distance of contacts is great enough. Therefore, R_0A of *pn* junction tends to be stable.

The situation of photocurrent under zero bias is different from R_0A. The photocurrent under zero bias decreases with the distance of contacts. As shown in Fig. 2.25, when the distance of contacts is 100 μm, the photocurrent under zero bias changes little, and when the distance is more than 100 μm, the photocurrent under zero bias is decreased notably. Photocurrent reflects quantum efficiency. Therefore, the decrease in photocurrent illustrates the decrease of quantum efficiency. Apart from the influence of electric field and current density due to the change of the contact location, the main reason for the decrease of quantum efficiency is that the collecting efficiency of photogenerated carriers to the contact decreases with the distance of contacts. Assume that the gain is 1, quantum efficiency (η) can be extracted from the photocurrent. As shown in Fig. 2.27, the trend of quantum efficiency is similar to that of the photocurrent. The detectivity D^*, η, and R_0A of device have the following relationship [46]:

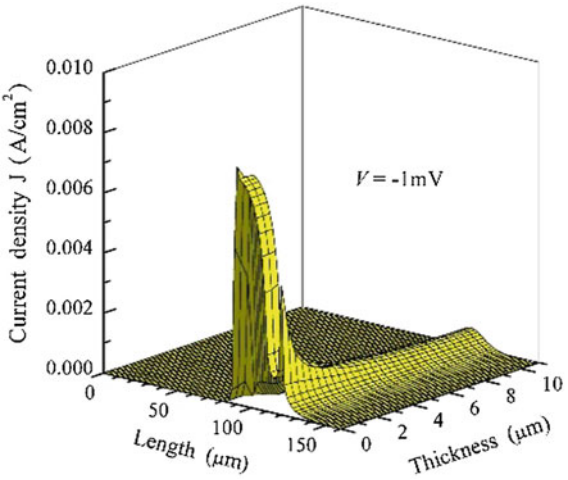


Fig. 2.26 Distribution of current density in device when the distance of contacts in p region is 10 μm

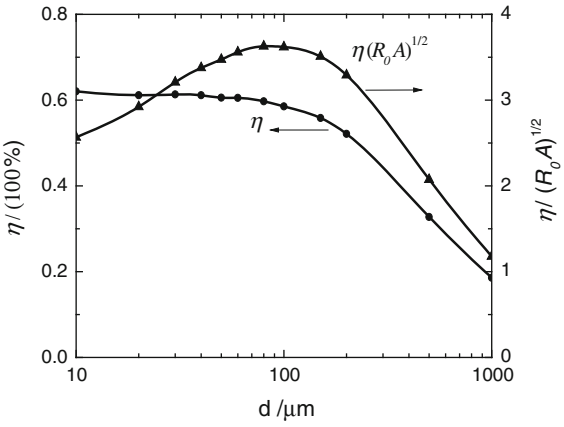


Fig. 2.27 Relationship between quantum efficiency (η, η $(R_0A)^{1/2}$) and the distance from p layer contact to n^+ area (d)

$$D^* \propto \eta(R_0A)^{1/2} \tag{2.44}$$

The relationship between $\eta(R_0A)^{1/2}$ and the location of contact is shown in Fig. 2.27. When the contact is about 100 μm from n^+-type region, $\eta(R_0A)^{1/2}$ will reaches the maximum size. It is the optimal location of contact in p region.

The structure of a linear HgCdTe device is shown in Fig. 2.28. Two lines of pn junction are staggered, the first line of pn junction is about 60 μm away from the common contact, and the second line is about 150 μm. Based on the above discussion, different distances of contacts will produce the different dynamic resistance and photocurrent, which leads to the difference of overall signal. A simple and feasible solution is to set a symmetrical common contact in the p region, which can eliminate the difference of signal caused by the geometrical structure of device. In

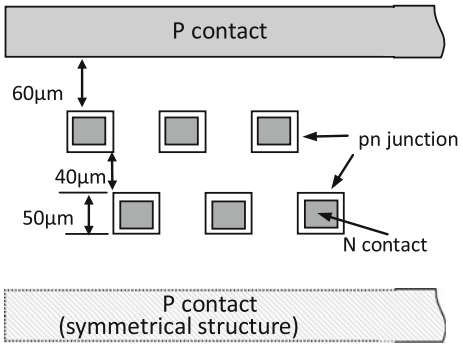


Fig. 2.28 Planar geometrical structure of linear HgCdTe detector

addition, if the distance of two contacts is too large, the photocurrent will not increase linearly and distinctly with the light power density. Therefore, the distance of contacts should not be too large.

In the above calculation, it is assumed that an ideal abrupt *pn* junction is formed at the interface between the n^+ region and the *p* region. In fact, there are different distributions of donor impurity, which are dependent on the process technology. HgCdTe-based *pN* junctions are usually formed by ion implantation. A p-type epitaxial layer is changed into an n type in the ion implantation region. The profile of donor impurity in the ion implantation region varies with the energy of ion implantation. Moreover, the ion implantation damage defects can influence the mobility and lifetime of HgCdTe material. In this section, the effect of ion implantation damage defects is not considered. Only the distribution of donor impurity in p-type region is discussed here.

For a *pN* junction formed by ion implantation, the distribution of donor impurity is approximated by Gaussian functions with appropriate parameters [47]. It is assumed that the distribution of donor impurity is described by Gaussian functions with variance δ, as shown in Fig. 2.29.

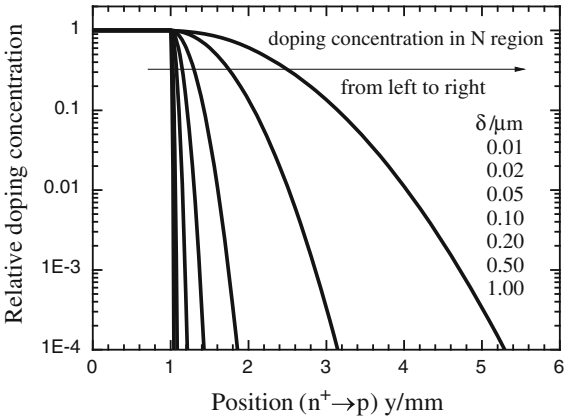


Fig. 2.29 Distribution of donor impurity in junction

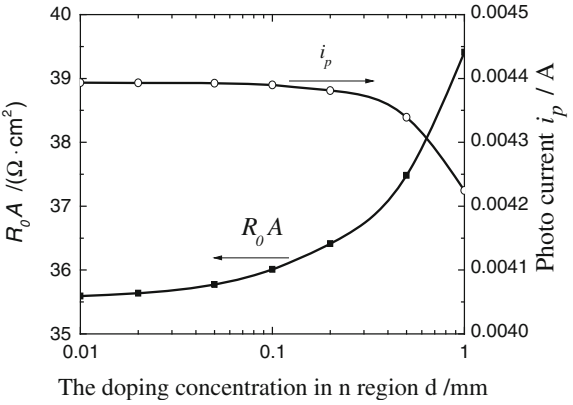


Fig. 2.30 Relationship of R_0A and photocurrent under zero bias with the different distribution of donor impurity in junction

The relationships of R_0A and photocurrent under zero bias with the distribution of donor impurity are shown in Fig. 2.30. R_0A increases slowly with the broadening of the donor impurity distribution. On the other hand, the photocurrent under zero bias decreased slightly with the broadening of the donor impurity distribution. Therefore, it is obvious that there is little influence of the distribution of donor impurity on R_0A and photocurrent. It illustrates that the conclusion drawn from the abrupt junction can be applied in the graded junction.

For an n-on-p structure, light absorption mainly occurs in the *p* region, thus the minority (electron) lifetime in the *p* region is very important for the device. In contrast, the majority (hole) lifetime has little effect on device. The electron lifetime of HgCdTe material is determined by three recombination processes: Radiative recombination, Auger recombination, and SRH recombination. The prior two recombination processes belong to the inherent properties of HgCdTe material regardless of the fabrication processes. However, SRH recombination is closely related to the fabrication processes. The defects, dislocations, and impurities of material are likely to form the recombination center, which results in increase of electron lifetime.

The relationship of R_0A and photocurrent (i_p) under zero bias with the electron lifetime of SRH recombination is shown in Fig. 2.31. It demonstrates that device performance decreases rapidly with the electron lifetime of SRH recombination. The electron lifetime is short, which reflects a large density of recombination centers formed by material defects, dislocations, impurities, etc. Therefore, R_0A and photocurrent are decreased. If the electron lifetime is long, photogenerated carriers will have enough time to diffuse into the junction region. In order to obtain the excellent device performance, the electron lifetime of SRH recombination should be longer than 100 ns. R_0A and photocurrent tend to be saturation when electron lifetime is about 1 μs. For high-quality HgCdTe material, the density of recombination centers is very small, and the electron lifetime of SRH recombination is long. The minority lifetime is an important parameter of device design.

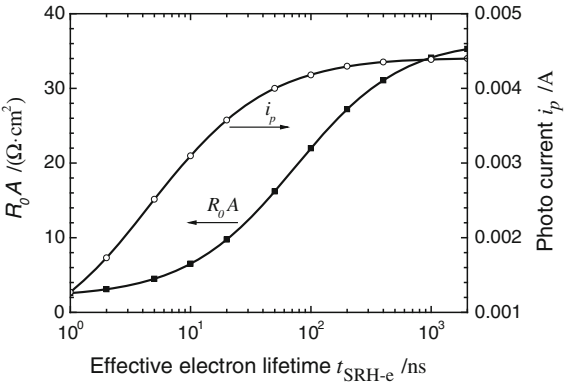


Fig. 2.31 R_0A and photocurrent (i_p) under zero bias with the electron lifetime of SRH recombination

The influence of the donor concentration in the n^+ region on device performance is seldom mentioned in literatures. In the above calculation, the donor concentration in the n^+ region is 10^{17} cm^{-3} (the effective doping concentration is about 1.6×10^{16} cm^{-3}) R_0A and photocurrent with different donor concentration in the n^+ region are discussed in detail. The acceptor concentration in p region is still 10^{15} cm^{-3} in the calculations. In general, the mobility of carriers and concentration of recombination center vary with the doping concentration. However, there is no reported qualitative explanation of this problem. Therefore, assume that other parameters like carrier mobility and concentration of recombination center do not vary with the doping concentration in calculation.

The calculated results are shown in Fig. 2.32, where the abscissa axis is the effective doping concentration. The effective mass of electrons in HgCdTe is very small. Therefore, the effective state density of conduction band is low (approximately 3.0×10^{15} cm^{-3} at 80 K). When the doping concentration is more than 1×10^{16} cm^{-3}, the conduction band will become degenerate, and impurity ionization rate decreases rapidly, which results in the failure of high doping. Figure 2.32 shows that both R_0A and photocurrent increase with the donor concentration in n^+ region. R_0A and photocurrent tend to be saturated when the donor concentration is 2×10^{16} cm^{-3}. The increase of donor concentration in n^+ region

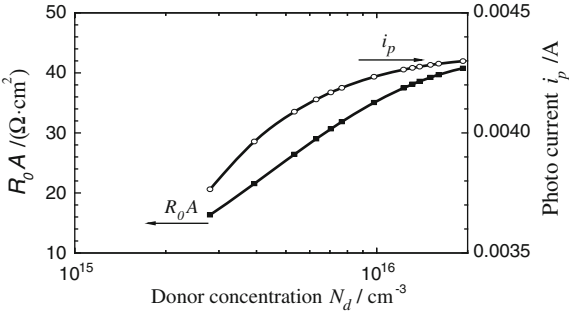


Fig. 2.32 The relationship of R_0A and photocurrent with effective donor concentration

can enhance the built-in electric field and widen the junction, which results in the increase of R_0A and photocurrent. The results show that the effective concentration of n^+ region should be more than 1×10^{16} cm^{-3}.

The effective hole mass of HgCdTe is relatively large, the effective state density of the conduction band is high, about 1.4×10^{18} cm^{-3} at 80 K. Therefore, apart from a very high doping concentration, the conduction is usually in non-degenerate state, and impurity ionization rate is very high. The acceptor concentration of the *p* region in infrared HgCdTe photovoltaic the *n* region, mobility and the concentration of recombination centers will vary with the acceptor concentration. Here the device performance is discussed in a simple condition. Assume that other parameters like carrier mobility and concentration of recombination center do not vary with the acceptor concentration in calculation.

The abscissa axis in Fig. 2.33 is the effective acceptor concentration. It shows that R_0A and photocurrent will increase with the acceptor concentration. R_0A and photocurrent tend to be saturated when the acceptor concentration is about 10^{16} cm^{-3}. The acceptor concentration in p-type material is about 10^{15}–10^{16} cm^{-3}, this acceptor concentration can make photocurrent and R_0A reach a maximum value. The *I*–*V* curve illustrates that the band-to-band tunneling rate also increases with the acceptor concentration in *p* region, and results in the increase of differential resistance (R_d) at low reverse bias. The results show that the effective acceptor concentration of *p* region is at the range of 10^{15}–10^{16} cm^{-3}.

When radiative recombination, Auger recombination, SRH recombination, and band-to-band tunneling (BBT) effect are considered in the calculation, the dark current is still smaller than in the real HgCdTe detector. Therefore, there must be other physical effects, which contribute to the dark current of the device. Many researchers have proposed different models to explain dark current. Nemirovsky et al. [48, 49] proposed a trap-assisted tunneling model (TAT), Elliot et al. [50] proposed an impact ionization model, Dosenfeld et al. [51] proposed a TAT described by segmented function. There are several free parameters that can be used to adjust these models. Therefore, the measured dark current can be explained to some extent. At present, there still is not an acknowledged model that can explain the total dark current in the experiment. Therefore, Hurkx et al. proposed a revised

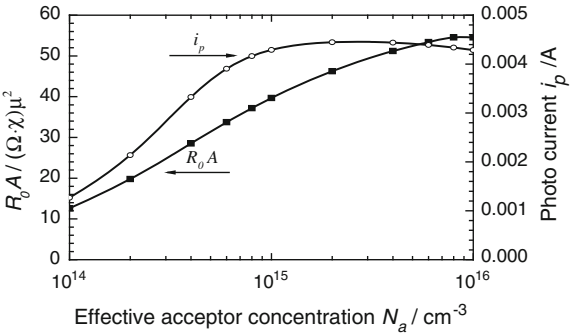


Fig. 2.33 Relationship of R_0A and photocurrent with effective acceptor concentration of *p* region

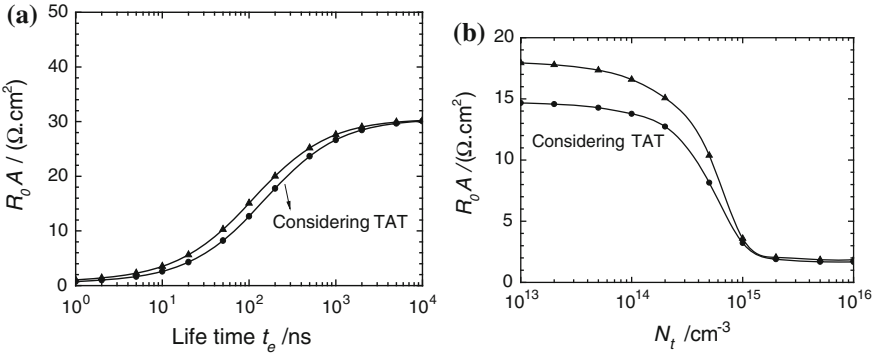


Fig. 2.34 R_0A with and without trap-assisted tunneling effect

model [52], which combines the TAT effect with the SRH recombination effect. The recombination rate can be expressed as

$$R_{\mathrm{TAT+SRH}} = \frac{pn - n_{\mathrm{ie}}}{\frac{\tau_{\mathrm{p}}}{1+\Gamma_{\mathrm{p}}}\left[n + n_{\mathrm{ie}}\exp\left(\frac{E_{\mathrm{T}}-E_{\mathrm{i}}}{kT}\right)\right] + \frac{\tau_{\mathrm{n}}}{1+\Gamma_{\mathrm{n}}}\left[p + n_{\mathrm{ie}}\exp\left(\frac{E_{\mathrm{T}}-E_{\mathrm{i}}}{kT}\right)\right]} \tag{2.45}$$

where E_{T} is the deep level of the trap, E_{i} and n_{ie} are the Fermi levels of intrinsic material and carrier concentration respectively, τ_{p} and τ_{n} are the recombination lifetime of holes and electrons respectively. Parameters Γ_{p} and Γ_{n} reflect the trap-assisted effect. For a weak electric field, this model can be simplified as SRH recombination. Lui et al. proposed a new model based on this current model, and the expression is similar to Eq. (2.45).

R_0A calculated by the Hurkx model varies with recombination lifetime and trap density is shown in Fig. 2.34. In the calculation, the effective doping concentration in the n^+ region is 10^{16} cm^{-3}. It shows that R_0A will decrease when the TAT effect is added into the calculation, especially when the electron recombination lifetime is very short. When the electron recombination lifetime is more than 100 ns, the TAT effect is not significant. When the trap concentration is more than 10^{14} cm^{-3}, R_0A will decrease rapidly. Figure 2.35 shows the relationship of quantum efficiency with trap concentration. It shows that when the concentration is more than 10^{15} cm^{-3}, quantum efficiency will decrease rapidly. Considering both R_0A and quantum efficiency, the trap concentration should be less than 10^{14} cm^{-3}.

2.2.3.2 Spectral Response

Figure 2.36 shows the spectral response (quantum efficiency) under different values of electron SRH lifetime. When the electron lifetime varies from 10 ns to 100 ns, quantum efficiency is increased by 50 %, and when the electron lifetime varies from 100 ns to 10 μs, the increase of quantum efficiency is limited, which is the same

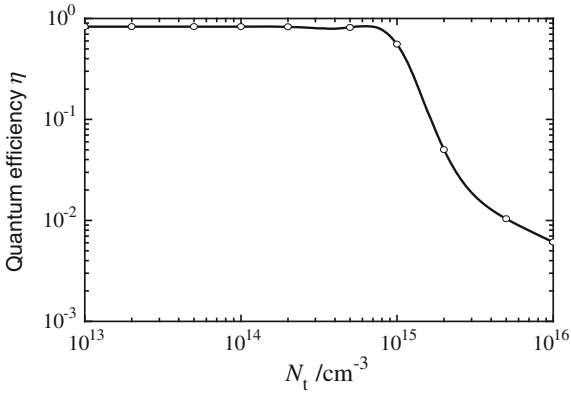


Fig. 2.35 Relationship with quantum efficiency and deep level concentration (λ = 9.5 mm)

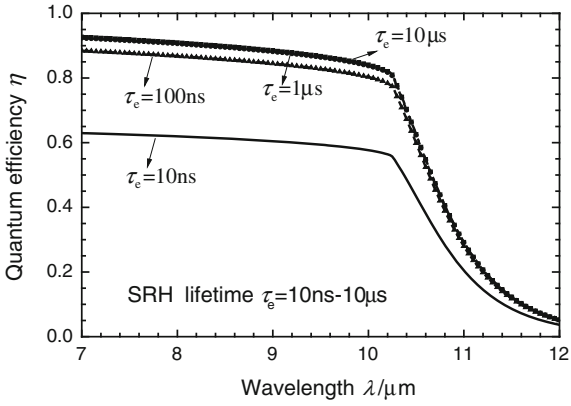


Fig. 2.36 Relationship of quantum efficiency and electron recombination lifetime

trend as Fig. 2.31. The device has relatively high quantum efficiency when the recombination lifetime of electron is more than 100 ns. The cutoff wavelength of the device is about 10.8 μm. Figure 2.37 shows the energy profile along the thickness direction (the direction of incident light). It shows that the shorter the wavelength, the radiation absorption is closer to the surface. Besides, the photon whose energy is only below the band gap tends to be uniform absorbed.

The material discontinuity at surface or interface of semiconductor devices, leads to a certain density of interface state, which results in the increase of the carrier recombination rate. If surface tarnish, oxidation and adsorption exist, surface recombination centers are possibly induced. In order to reduce the interface state, the surface of a HgCdTe device is generally passivated, including anodizing or prepared passivation film, such as ZnS and CdTe film. After the surface passivation, the density of the interface state is greatly reduced. However, the surface recombination effect still exists, which influences the dark current and photocurrent. Research into the influence of surface recombination on device performance has practical significance.

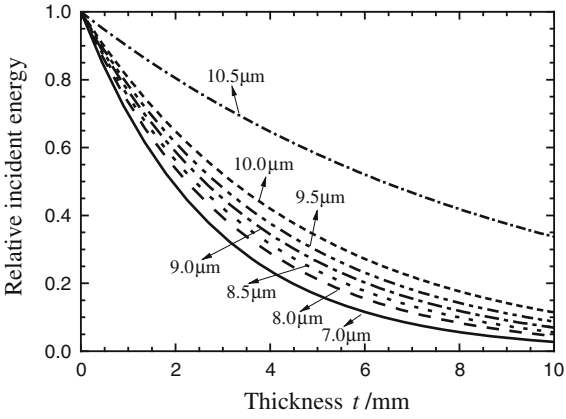


Fig. 2.37 Distribution of incident radiation energy in the device along the direction of thickness

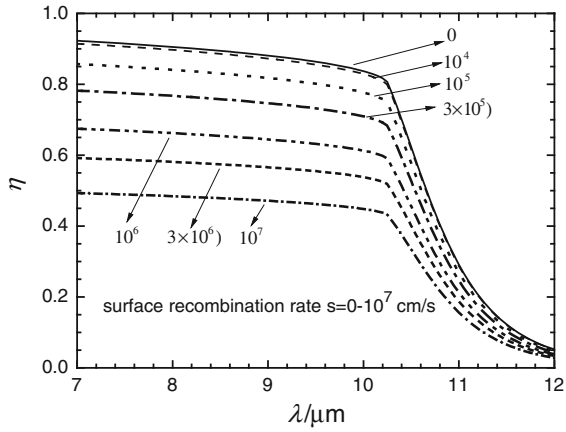


Fig. 2.38 Relationship of quantum efficiency and surface recombination rate

The relationship between quantum efficiency with surface recombination rate is shown in Fig. 2.38. The surface recombination rate is very low, and has little effect on the quantum efficiency. When the surface recombination rate increases to 10^5 cm/s, quantum efficiency begins to decrease. When it is more than 10^5 cm/s, quantum efficiency decreases rapidly. In order to obtain high quantum efficiency, the surface recombination rate for a long wavelength n^+-on-p device should be controlled below 10^5 cm/s, and preferably less than 10^4 cm/s.

Rosenfeld et al. have studied the responsivity of long wavelength devices with an n-type absorber layer. The results show that the responsivity decreases rapidly when the surface recombination is about 2×10^3 cm/s (about two orders of magnitude different from the above conclusion 10^5 cm/s). The reason can be explained as follows:

The influence of surface recombination on responsivity is related to the factor γ

$$\gamma = \frac{sL}{D} = s\sqrt{\frac{\tau}{D}} = \sqrt{\frac{q}{kT}}s\sqrt{\frac{\tau}{\mu}} \tag{2.46}$$

where μ and τ are the (bulk) minority mobility and lifetime respectively. The absorption region mainly contributes to the responsivity. The factor γ represents capacity of surface recombination. For the device with a p-type absorber layer, electrons acting as the minority carrier have a short lifetime and large mobility. Therefore, γ is small under the same surface recombination rate, and has little impact on the responsivity. For the device with an n-type absorber layer, holes acting as minority carrier have a long lifetime and small mobility. Therefore, γ is large under the same recombination rate at surface, and has large impact on the responsivity. Therefore, the device with the n-type absorber layer is a better choice under the same surface recombination rate.

Figure 2.39 shows the experiment measurements (solid line) and theoretical calculation (dot line) of spectral response of two n^+-on-p detectors. The results are in agreement with the experimental measurements near the cutoff wave band and the short wave band. When the wavelength is shorter than the 1–2 μm wave band (corresponding to the material band gap), the experimental data are lower than that of theoretical calculation. This phenomenon happens in many long wavelength devices. One possible reason is the nonuniform composition of material in the device.

There are lateral and vertical profiles of components in HgCdTe material. For a large material region, the lateral nonuniformity of composition may be obvious. A detector grown by MBE with an active detection area of only 50 × 50 μm^2 is shown in Fig. 2.39. Therefore, the lateral nonuniformity of composition can be neglected. The vertical nonuniformity of the composition is the main factor for the spectral response.

The device is a GaAs/CdTe/HgCdTe multilayer structure. The diffusion of composition is inevitable in the process of material growth and fabrication, which

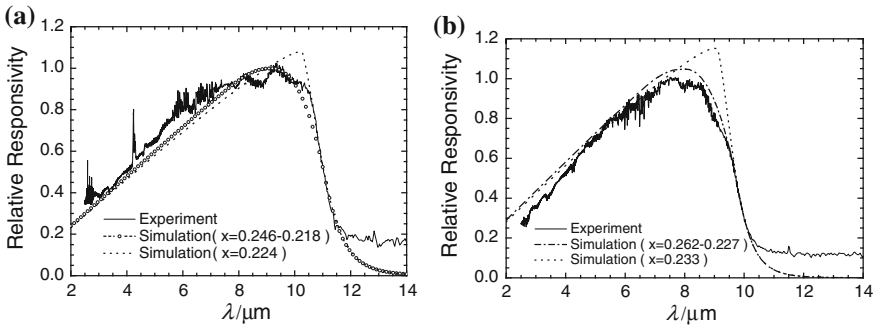


Fig. 2.39 Spectral response of long wavelength. **a** Detector 1, **b** detector 2

results in the vertical nonuniformity of HgCdTe material. Based on the theory of graded components, the numerical simulations of detectors 1 and 2 are shown in Fig. 2.39. From the interface to the surface of device, the components of detector 1 vary from 0.246 to 0.218, and the composition of detector 2 varies from 0.262 to 0.227. It shows that the results of the graded composition theory are more accurate compared to those of the uniform composition theory.

2.2.3.3 Thickness of Absorption Layer and Interface Charge Density

As HgCdTe materials and device technology matures, HgCdTe detector structure tends to become more complicated, such as heterojunctions, two-color detectors [53–59], etc. The minority diffusion length in the base region of long wavelength HgCdTe detectors is relatively small, but the absorption layer must be large, so it is necessary to optimize the thickness of the base region and, due to bandgap narrowing, the influence of the interface charge is more obvious so it is very necessary to analyze the impact of interface charge.

For a typical n-on-p HgCdTe detector, when the incident light is back illuminated from substrate into the detector, the light is absorbed in base region. The photogenerated carriers then diffuse to junctions, and are separated by the built-in electric field in the junction to form a photoelectromotive force. Therefore, under the same conditions, the thicker the base absorption layer, the more the light is completely absorbed thus creating more photogenerated carriers, and higher responsivity. On the other hand, the absorption of light is not uniform, but varies along the length of the exponential decay path, i.e., absorption of light mainly occurs at the beginning of the absorption path. The thicker absorption layer, resulting in the main absorption region is further from the junction. Therefore, the probability of the photogenerated carriers participating in the diffusion process is greater, resulting in a lower generated responsivity. It is acknowledged that the responsivity of the device, as a function of the light absorption layer thickness, changes according to the two competing mechanisms. When the absorption layer is thinner, the former mechanism dominates, therefore an increase in responsivity is due to an increase in the absorption layer thickness. When the thickness absorption of layer is large, the latter mechanism is dominant and therefore the responsivity decreases with an increase in absorption layer thickness. So there is an optimal absorption layer thickness in the middle.

The following is an example of a long wavelength detector in which the optimal thickness of the absorption layer is obtained by simulation. Detailed material and structural parameters are listed in Table 2.11. *P* region is the base region, so the thickness of the *p* region is the absorption layer. According to the material parameters, the minority diffusion length can be solved:

Table 2.11 Material and structural parameters of long wavelength HgCdTe photodetector

Parameters	Value	Units
Cd composition (x)	0.211	–
Temperature	77	K
Electron SRH lifetime	5	ns
Hole SRH lifetime	5	ns
Doping density of N region	1×10^{17}	cm^{-3}
Doping density of P region	8×10^{15}	cm^{-3}
Thickness of N region	1	mm
Thickness of P region	d	mm
Width of N region	30	mm
Distance of unit center	40	mm

Reprinted from Ref. [22], with kind permission from Springer Science+Business Media

$$L_{\mathrm{d}} = \sqrt{D_{\mathrm{n}} \cdot \tau} = \sqrt{\frac{kT}{q} \mu_{\mathrm{n}} \cdot \tau} = 21\ \mu\mathrm{m} \tag{2.47}$$

The longer the diffusion length, the easier minority carriers diffuse to a junction. So diffusion length is a very important parameter that directly influences the responsivity.

Figure 2.40 shows the responsivity as function of the wavelength of incident light. The peak responsivity occurs when the wavelength of incident light is 12.4 μm and the thickness of absorption layer changes from 5 to 21 μm. Figure 2.40a shows, at the short wavelength range, the responsivity monotonously decreases with the increase in the thickness of the absorption layer. The absorption coefficient at the short wavelength range is larger than that of the long wavelength range, thus the light is completely absorbed over a small fraction of the absorption

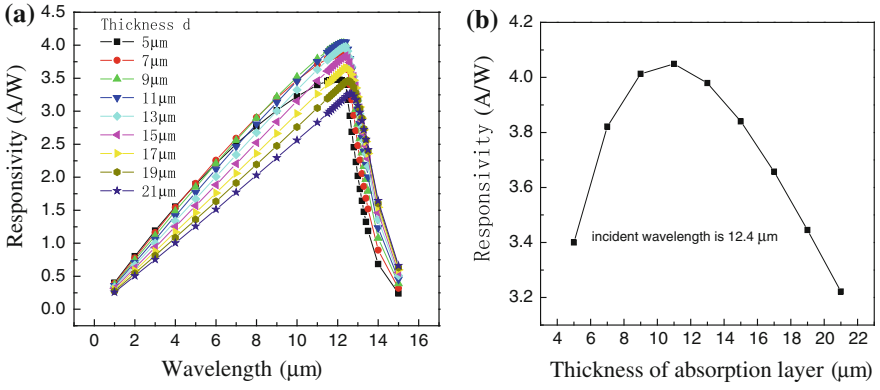


Fig. 2.40 **a** Photo response as function of the wavelength of incident light with the thickness of absorption layer changing from 5 to 21 μm, **b** when the wavelength of incident light is 12.4 μm, photo response as function of the thickness of absorption layer. Reprinted from Ref. [22], with kind permission from Springer Science+Business Media

layer. Here, diffusion is the dominant mechanism of the responsivity causing the monotonous decrease. It is also shown in Fig. 2.40a that the responsivity monotonously increases with the increase of the thickness of the absorption layer, and finally saturates, when the incident light wavelength is longer than the cut-off wavelength. This is because the absorption coefficient in the long wavelength range is very small, so the absorption length is equal to the thickness of the absorption layer. Here, absorption is the dominant mechanism of the responsivity causing the monotonous increase. When the thickness of absorption layer is close to the diffusion length of the minority carrier, the absorption at one diffusion length distance from the *p-n* junction can not affect the responsivity, consequently causing the gradual saturation. It can be seen in Fig. 2.40b that, near the peak responsivity wavelength, the responsivity increases with the increase of the thickness of absorption layer first, and then decreases. The maximum responsivity, which is the consequence of competing effects of the absorption and diffusion, is reached when thickness of absorption layer is 11 μm.

Figure 2.41 compares the photoresponse for different minority carrier lifetimes when the incident light wavelength is 12.4 μm as a function of minority carrier lifetime for 5, 10, 20, and 50 ns. The corresponding minority diffusion length for each minority carrier lifetime is 21, 29.7, 42, and 66.4 μm, respectively. The absorption length, which is determined by the absorption coefficient, is 5.46 μm. As shown in Fig. 2.41, responsivity increases with an increase of the minority carrier lifetimes. Therefore, the optimal d_{abs} increase with increasing of the diffusion lengths.

Figure 2.42 compares the responsivity for the different wavelengths of incident light when the minority carrier lifetime is 5 ns. The corresponding absorption lengths are 0.77, 1.97, 3.33, and 5.46 μm, respectively, and the minority diffusion length is 21 μm. It is found that the responsivity increases with the increase in the wavelength of the incident light. And the optimal thickness of the absorption layer

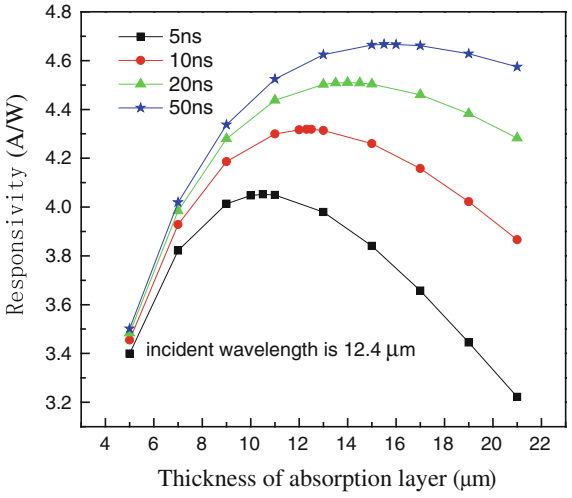


Fig. 2.41 When incident light wavelength is 12.4 μm, responsivity versus the thickness of absorption layer under different minority carrier lifetime. Reprinted from Ref. [22], with kind permission from Springer Science+Business Media

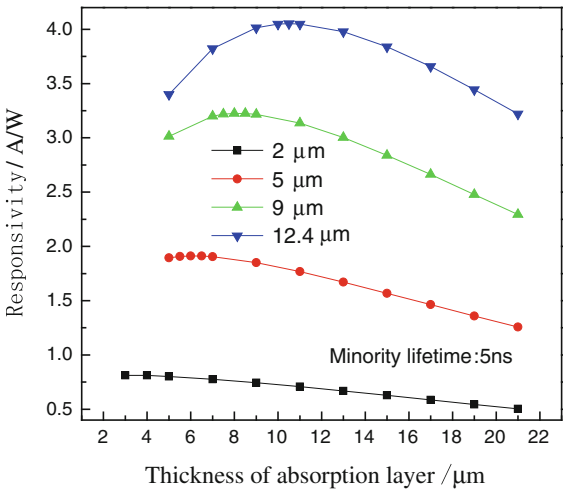


Fig. 2.42 Responsivity versus the thickness of absorption layer under different incident light wavelengths with the minority lifetime of 5 ns. Reprinted from Ref. [22], with kind permission from Springer Science+Business Media

increases as well. The reason for the absorption layer increase is that the absorption coefficient decreases with increasing wavelength of incident light with constant Cd mole fraction of HgCdTe material. Specifically, the absorption length increases with increasing wavelength of the incident light. Therefore, the optimal thickness of the absorption layer increases with increasing absorption length. When the wavelength of incident light remains constant, the intrinsic absorption coefficient of HgCdTe material decreases with increasing of the mole fraction. Consequently, the optimal thickness of the absorption layer of mid and short wavelength devices is smaller than that for long wavelength devices.

By discussion, the optimal absorption layer thickness when the minority carrier lifetime is 5, 10, 20 and 50 ns, and incident light wavelength is 2, 5, 9, and 12.4 μm, a fitted empirical formula for predicting the optimal absorption layer thickness as function of diffusion length and absorption length can be obtained:

$$d = 1.72 \times \sqrt{L_a} \cdot \ln L_d - 1.31 \tag{2.48}$$

In Eq. (2.48), d is the optimal absorption layer, L_a is absorption length, L_d is diffusion length, and the unit are all μm, as shown in Fig. 2.43.

Due to the lattice discontinuity and complex fabrication process, many traps will be formed in the substrate and the passivation layer of HgCdTe device causing a significant impact to device performance. Traps can be usually divided into two categories: donor trap and acceptor trap. A full state of the trap is defined as the state that is occupied by electrons, and an empty state of the trap is the state that is occupied by holes. For a donor trap, it is neutral in full state, which can capture holes or release electrons. Additionally, a donor trap is positively charged in the empty state, and can capture electrons or release holes. Therefore, the interface of donor traps is usually positively charged. An acceptor trap is negatively charged in the full state and can capture holes or release electrons. Additionally, it is neutral in

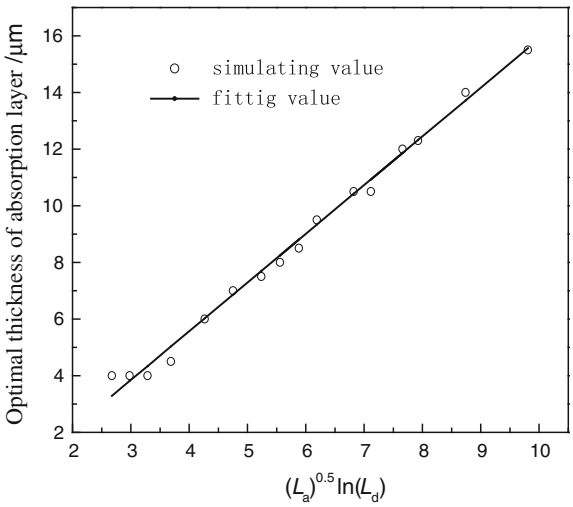


Fig. 2.43 Fitting curve of the optimal thickness of absorption layer as a function of absorption length L_a and diffusion length L_d. Reprinted from Ref. [22], with kind permission from Springer Science+Business Media

the empty state and can capture electrons or release holes. Therefore, the interface of acceptor traps is usually negatively charged. In addition, the fixed charge and moveable charge usually exist in the passivation layer. Here, they are all equivalent as interface charges in calculations.

The structural and material parameters are the same with that of the device listed in Table 2.11, and the thickness of the p region is 11 μm. The spectral response of the device without interface charge and with interface charge is shown in Fig. 2.44. The interface charge in Fig. 2.44a is 1×10^{12} cm^{-2}, and in Fig. 2.44b is -1×10^{12} cm^{-2}.

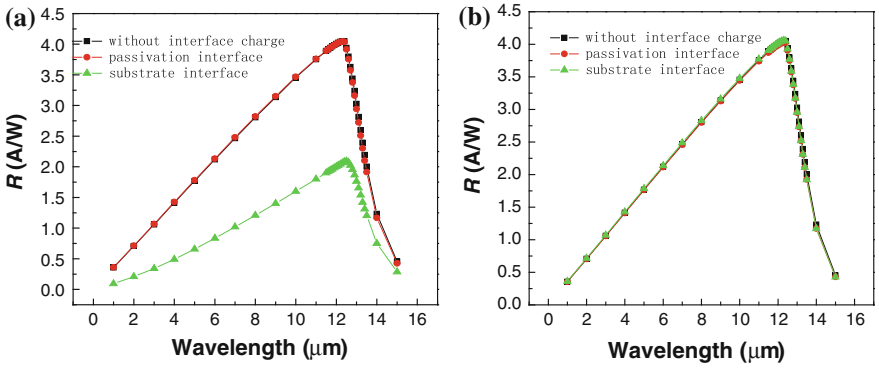


Fig. 2.44 Spectral response of the device without interface charge and with interface charge. **a** The density of interface charge is 1×10^{12} cm^{-2}, **b** the density of interface charge is -1×10^{12} cm^{-2}

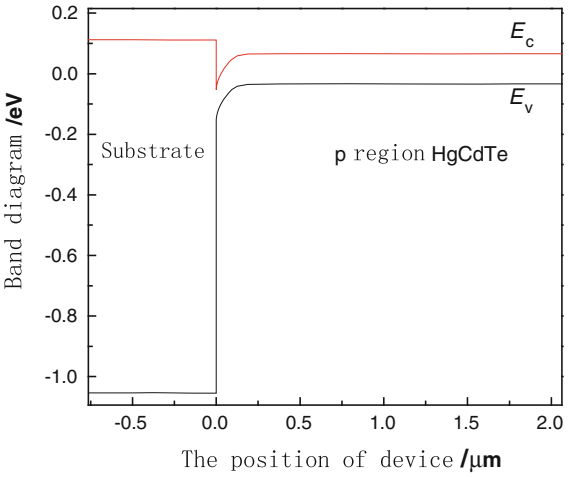


Fig. 2.45 The band diagram of device when the substrate interface charge is 1×10^{12} cm^{-2}

As shown in Fig. 2.44, regardless if the charge is positive or negative, the interface charge in the passivation layer has little effect on the device responsivity. However, the positive charge at the substrate interface has a great effect on the responsivity of device. The reason is that the energy band is changed by the charge at the substrate interface. As illustrated in Fig. 2.45, the energy band is bent down due to the positive interface charge, which results in the formation of an induced *pn* junction at the interface. When the device is back illuminated, photogenerated carriers are formed at the interface and some will be separated by the induced junction, which ultimately reduces the responsivity of the device. When the interface charge is negative, the energy band is bent upward, which results in the formation of an accumulation layer at the substrate interface. This accumulation layer has little effect on the responsivity of the device.

The analysis of the passivation layer is similar to the above situations. When the interface charge is positive, the direction of the built-in electric field generated by the induced junction is the same as that of the device. However, it is far away from the light absorption region, so it has little effect on the device responsivity. If a strong inversion occurs in the passivation layer, an n-type channel will be formed, which results in an increase of dark current and a decrease in device performance. When the interface charge is negative, a p-type accumulation layer at the interface will be formed, which can also lead to the increase of tunnel leakage currents at the surface.

The responsivity as a function of the positive substrate interface charge is shown in Fig. 2.46, where the wavelength of incident light is 12.4 μm. It shows that when the positive interface charge is more than 2×10^{10} cm^{-2}, the device responsivity will begin to decrease; when the positive interface charge is more than 5×10^{11} cm^{-2}, photo response will reach the minimum value. This phenomenon can be explained by the theory of MOS devices. The energy band is bent downward due to the positive interface charge, which results in the depletion of the p-type region at

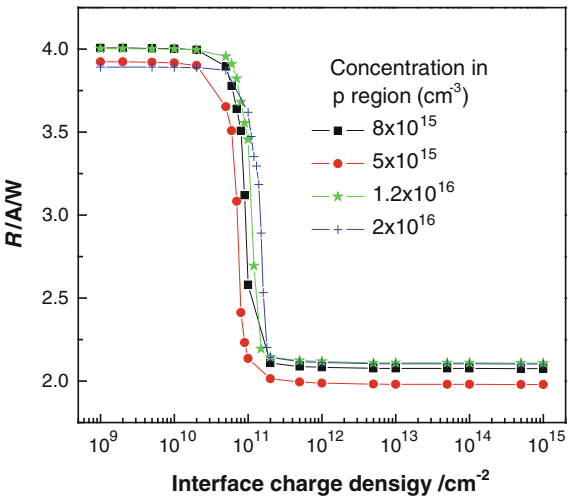


Fig. 2.46 Responsivity of device with the substrate interface charge, where the doping density in *p* region is 8 $\times$ 10^{15} cm, the wavelength of incident light is 12.4 μm

interface. When the density of the interface charge is 2 $\times$ 10^{10} cm^{-2}, an inverted *n* region will be formed at the interface, which results in the formation of an induced junction. The depletion width of the induced *pn* junction increases with the charge density. Photo-generated carriers generated in this area will be separated by the induced junction, which drastically reduces the device responsivity. When the density of the interface charge is 5 $\times$ 10^{11} cm^{-2}, a strong inversion layer will be formed at the substrate interface. The depletion width of the induced *pn* junction will stop increasing. Therefore, the responsivity of the device remains stable with the density of the interface charge.

Based on the theory of MOS devices [15], the interface charge density of a device with nominal inversion and strong inversion is expressed as follow:

$$N_s = N_a \cdot W = \sqrt{\frac{2\varepsilon_s \varepsilon_0 N_a V_s}{q}} \tag{2.49}$$

where W is the width of depletion region, V_s is the interface potential, ε_s is the relative permittivity of HgCdTe. When the composition of HgCdTe is 0.211, the relative permittivity of HgCdTe is 17.5. The surface potential of a device with the nominal inversion and strong inversion are expressed as follows:

$$V_{s-\text{inversion}} = \frac{kT}{q} \cdot \ln\left(\frac{N_a}{n_i}\right) \tag{2.50}$$

$$V_{s-\text{strong inversion}} = 2 \cdot \frac{kT}{q} \cdot \ln\left(\frac{N_a}{n_i}\right) \tag{2.51}$$

The concentration of p region is 8×10^{15} cm^{-3}. Integrating Eqs. (2.50) and (2.51) with Eq. (2.49), shows the calculated charge density of device with the inversion and strong inversion to be 8.05×10^{10} cm^{-2} and 1.14×10^{11} cm^{-2} respectively. The magnitude of the calculated results is the same as that in simulation. The difference between them is due to the depletion layer approximation in MOS theory.

Equation (2.49) shows that the critical density of the interface charge is related to the concentration the p region. The greater the concentration, the greater the critical density of interface charge. Therefore, the influence of the interface charge on the responsivity of device can be reduced by increasing the concentration of the p region. Figure 2.46 also shows the responsivity of the device varies with the interface charge, when the concentration of p region is 5×10^{15} cm^{-3}, 1.2×10^{16} cm^{-3} and 2×10^{16} cm^{-3}. The simulated results are very consistent with that of formula (2.49). In Fig. 2.46, the responsivity of a device with a concentration of p region 5×10^{15} cm^{-3} is less than that of a device when the concentration of p region is 8×10^{15} cm^{-3}. This is mainly because series resistance decreases with an increase of the p region concentration. When interface charge density is relatively small, the responsivity of a device with the p region concentration of 5×10^{16} cm^{-3} is smaller than that of device with the of p region concentration of 8×10^{15} cm^{-3}. This is mainly because an increase of p region concentration will reduce the minority lifetime. When the strong inversion of device occurs, the responsivity of a device with high of p region concentration will be slightly decreased. Instead, there is a relatively high responsivity in a device with a high density of interface charge. It is noted that an increase of the p region concentration can reduce the influence of the substrate interface charge on the responsivity of device, however, as concentration increases, the built-in electric field will also increase which results in an increase in tunneling current. At the same time, the decrease of carrier lifetime will lead to an increase of diffusion current, which also degrades the device performance.

2.2.4 *Design of Two-Color HgCdTe Detector*

With the development of infrared imaging technology, single wave band (monochromatic) imaging systems are increasingly being perfected. There is now an urgent demand for multiband (or multicolor, including two-color) imaging systems. Multicolor systems are designed to provide multiband data to reduce the complex background of a target. In terms of target recognition, signal identification and antijamming performance, a multicolor system is superior to monochromatic detectors. Research into multicolor systems has important application to many applications, such as earth observation, environment sensing, target discrimination and identification [60]. At present, many agencies in the world have carried out research work into multicolor infrared detectors [61–67]. HgCdTe two-color detectors have two working patterns: “sequential” pattern and “uniform” pattern

[68]. For a "sequential" pattern [69, 70], each unit has only one output electrode, the diode detecting band is determined by selecting an offset voltage, each band works in time-sharing and the device structure is relatively simple. Although two bands cannot be synchronously detected in space and time by varying the bias voltage to select the working band, it is useful for removing false signals and improving detection efficiency. For the "uniform" pattern [71, 72], each unit has two output electrodes, the two bands operate at the same time, but the device structure is more complicated than that of the device with "sequential" pattern. However, the two working bands are fully synchronous in space and time, which greatly expands functionality over monochromatic devices. Compared to monochrome devices, two-color devices are more difficult in the aspect of structure, process, and performance control [73–77]. Simulation calculations have become an effective means to predict the performance of two-color devices and reduce the device cost. Simulation can provide an important reference value for research into device manufacturing techniques.

2.2.4.1 Typical Two-Color HgCdTe Detectors

At present, there are mainly four types of typical two-color HgCdTe detector structures [61–77]: ① The n-p^+-n structure originally reported by the Hughes Research Laboratory (HRL); ② The p-n-n-n-p structure reported by the Lockheed Martin company (now called BAE Systems); ③ The n-p-p-p-n structure reported by the French LETI/LIR; ④ The p^+-on-n structure reported by the French LETI/LIR. This section will compare the photoelectric properties of four MW1/MW2 two-color detectors, and investigate the optimal structures that meet the demand of applications and the existing technology platform.

① HRL Structure. As shown in Fig. 2.47, this structure overcomes the shortcomings of the two-color detector with "sequential" pattern does not permit two photodiodes to operate at the same optimum bias thus causing an increase in spectral cross talk. However, each unit consists of two electrodes, and readout circuit to separate and read out two band signals at the same time. The theoretical spectral curve of this structure is shown in Fig. 2.48, and its structural parameters are listed in Table 2.12.

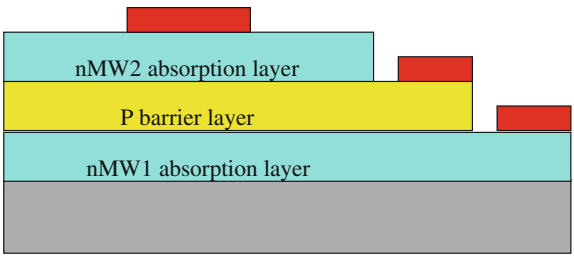


Fig. 2.47 The n-p^+-n schematic structure of two-color detector reported by HRL

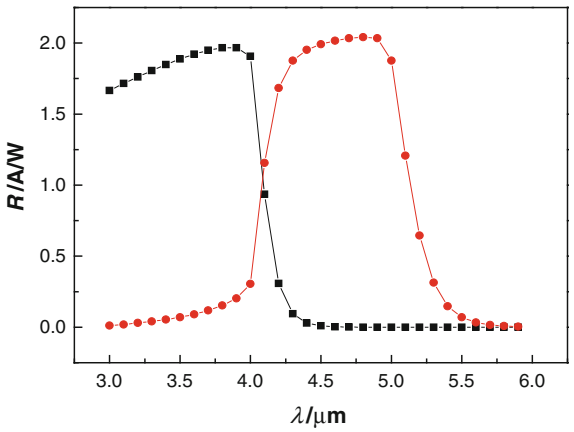


Fig. 2.48 The theoretical spectral curve of n-P^+-N two-color detector reported by HRL

Table 2.12 The structural parameters of n-p^+-n two-color detector reported by HRL

	Component	Doping (cm^{-3})	Thickness (μm)	Bandgap wavelength (μm)	SRH lifetime
nMW1 absorption layer	0.342	n-type 1×10^{15}	6	4.0	1 μs
p-barrier	0.4	p-type 1×10^{17}	2	3	
nMW2 absorption layer	0.304	n-type 1×10^{15}	7	5.0	

② BAE Structure. As shown in Fig. 2.49, an n-n-n barrier is introduced to this kind of structure between two photoelectric diodes, which can stop the diffusion of minority carrier hole between two wave bands, thus reducing the spectral cross talk. The theoretical spectral curve of this structure is shown in Fig. 2.50, and its structural parameters are listed in Table 2.13.

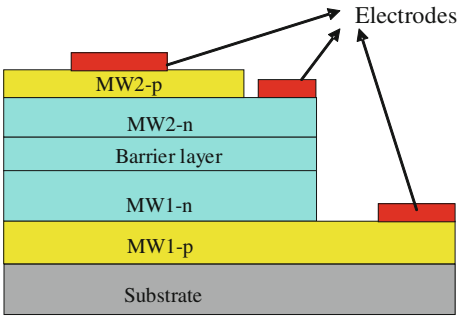


Fig. 2.49 The p-n-n-n-p schematic structure of two-color detector reported by BAE

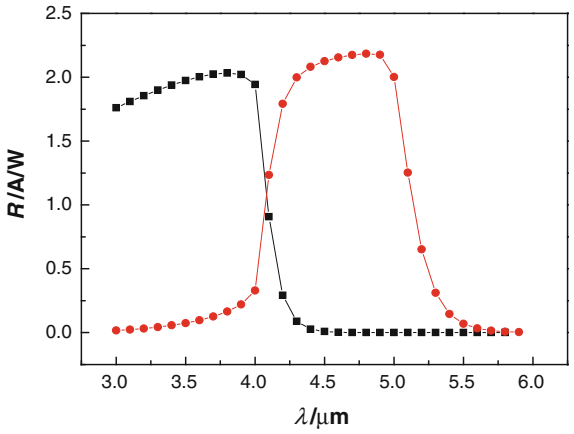


Fig. 2.50 The theoretical spectral curve of p-n-N-N-P two-color detector reported by BAE

Table 2.13 The structural parameters of p-n-n-n-p two-color detector reported by BAE

	Component	Doping (cm^{-3})	Thickness (μm)	Bandgap wavelength (μm)	SRH lifetime
MW1-p	0.42	p-type 2×10^{17}	1.2	2.86	1 μs
MW1-n	0.342	n-type 1×10^{15}	6	4.0	
Barrier	0.4	n-type 5×10^{15}	0.2	3	
MW2-n	0.304	n-type 1×10^{15}	6.0	5.0	
MW2-p	0.4	p-type 2×10^{17}	1.0	3.0	

③ LTEI Structure. As shown in Fig. 2.51, the structure of MW2 photoelectric diode is designed as an injected planar junction, and the MW1 diode is in situ a mesa junction. There is a high duty cycle in the LTEI structure. The theoretical spectral curve of LTEI structure is shown in Fig. 2.52, and its structural parameters are listed in Table 2.14.

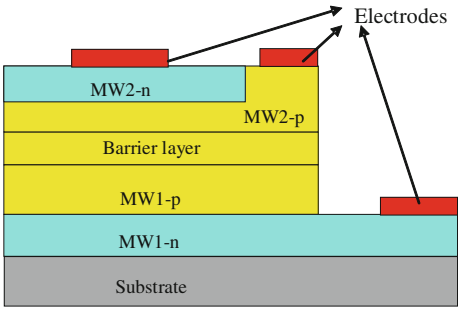


Fig. 2.51 The n-p-p-p-n schematic structure of two-color detector reported by LETI

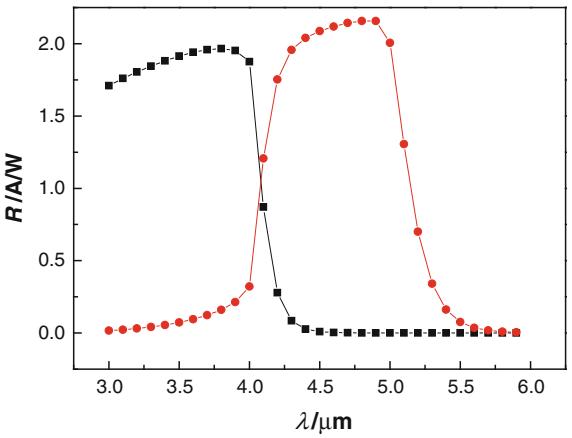


Fig. 2.52 The theoretical spectral curve of n-p-p-p-n two-color detector reported by LETI

Table 2.14 The structural parameters of n-p-p-p-n two-color detector reported by LETI

	Component	Doping (cm^{-3})	Thickness (μm)	Bandgap wavelength (μm)	SRH lifetime
MW1-n	0.42	n-type 2×10^{17}	1.2	2.86	10 μs
MW1-p	0.342	p-type 1×10^{15}	6	4.0	
Barrier	0.4	p-type 5×10^{15}	0.2	3	
MW2-p	0.304	p-type 1×10^{15}	7	5.0	
MW2-n	0.304	n-type 2×10^{17}	1.2	5.0	

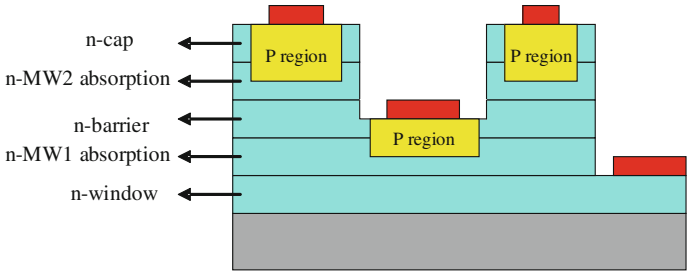


Fig. 2.53 The schematic structure of two-color detector based on p^+-on-n reported by Rockwell

④ Rockwell Structure. As shown in Fig. 2.53, this structure is actually made up of two vertical integration p^+-on-n planar diodes, which is formed by twice ion implantation in n-n-n heterogeneous material. Due to the directions of two wavelengths photoelectric diodes are the same, two light signals does not need to be separated. Therefore, its readout circuit structure is relatively simple. The theoretical spectral curve of Rockwell structure is shown in Fig. 2.54, and its structural parameters are listed in Table 2.15.

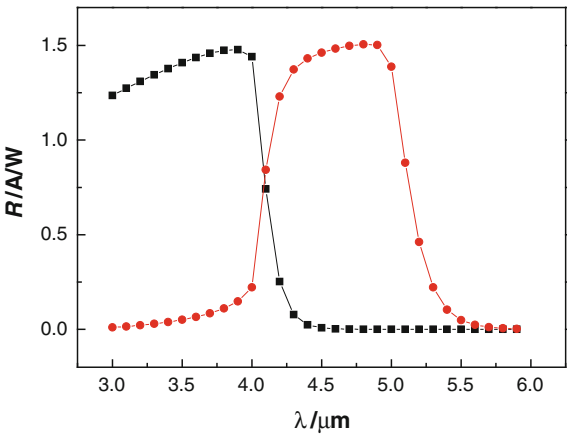


Fig. 2.54 The theoretical spectral curve of two-color detector based on p^+-on-n reported by Rockwell

Table 2.15 The structural parameters of two-color detector based on p^+-on-n reported by Rockwell

	Component	Doping (cm^{-3})	Thickness (μm)	Bandgap wavelength (μm)	SRH lifetime
n-window	0.42	n-type 1×10^{15}	1.2	2.86	1 μs
nMW1 absorption	0.342	n-type 1×10^{15}	6	4.0	
barrier	0.4	n-type 1×10^{15}	1	3	
nMW2 absorption	0.304	n-type 1×10^{15}	6.2	5.0	
n-cap	0.4	n-type 1×10^{15}	0.8	3.0	
p-implant	–	p-type 1×10^{17}	1.0	–	

Table 2.16 The quantum efficiency and spectral cross talk of four structures

	Quantum efficiency η		Spectral cross talk	
	MW1	MW2	C_{12} (%)	C_{21} (%)
HRL structure	64.2	52.7	0.02	7.76
BAE structure	66.3	56.4	0.02	8.12
LETI structure	64.2	55.7	0.02	8.14
Rockwell structure	48.1	38.9	0.02	7.5

In conclusion, the quantum efficiency and spectral cross talk of these various structures are listed in Table 2.16. C_{12} is defined for the cross talk of MW1 to MW2 (incident light wavelength is 4.8 μm). C_{21} is defined for the cross talk of MW2 to MW1 (incident light wavelength is 3.8 μm). It is obvious that the cross talk of MW2 to MW1 can be ignored, but MW1 to MW2 is too large. The main reason is that the absorption layer of MW1 is not thick enough, so that too much light is

allowed to enter MW2 through MW1. A possible solution is increasing the absorption layer thickness of MW1 to reduce the optical cross talk. However, this may reduce the quantum efficiency of MW1 and R_0A at the same time. Therefore, it is very necessary to design the best absorption layer thickness of MW1by theoretical calculation.

The performance of BAE structure is proved to be the best, according to the calculation results of the quantum efficiency, but this kind of structure is strongly dependent on the growth of the material. Therefore, the LETI structure may be most feasible, considering the requirements of device fabrication. If there are no specific notations, the two-color HgCdTe infrared detectors discussed in this section refer to the LETI structure.

2.2.4.2 Numerical Simulation of Spectral Photoresponse

Integrated as a single pixel in a focal plane array, a two-color detector has the ability to detect two spectral bands separately and independently, with high temporal and spatial coherence. When used to acquire infrared images, this type of detector provides an additional dimension of contrast that is available for signal processing and serves as a visual aid in scene interpretation, allowing determination of both absolute temperature and unique signatures of the objects from the background [78]. Several research groups have already presented interesting results on two-color focal plane arrays over the entire 2–12 μm wavelength region [67, 78–80].

HgCdTe two-color n–p–p–p–n photodetector arrays have recently been reported in Ref. [72, 81]. At 78 K, the measured SWIR and MWIR photodiode cut-off wavelengths are consistent with the expected configuration of the device. The calculated full width half maximum (FWHM) of the short wavelength (SW) spectrum is quite small, with the shape of the SW photoresponse being narrow and sharp. This phenomenon is also known as the 'short wavelength narrow effect'.

Figure 2.55 is the schematic of the n^+-p-p-p-n two-color infrared detector. Under back-illumination, the bottom *pn* junction ($x = 0.4$) absorbs SW radiation for wavelengths up to its cut-off value. This *pn* junction also acts as a transparent window for mid-wavelength light to pass through and be absorbed by the top n^+-p junction with the MW cutoff. The specific details of the detector, such as the doping densities, composition, thickness of each epilayer, etc., are listed in Table 2.17.

The spectral responsivity of the two-color infrared detector was measured with the samples held at 78 K using an NEXUS 670 FTIR spectral measurement system with a specific external optical path setup. Figure 2.56 shows the spectral photoresponse for the integrated HgCdTe two-color infrared detector illustrated in Fig. 2.55 where the corresponding two spectral bands with cut-off wavelengths of 3.04 and 5.74 μm are evident. Firstly, it can be observed that the cut-off wavelength of the SW photovoltaic diode is close to the cut-on wavelength of the MW photovoltaic diode. This effectively enables the SW photovoltaic diode to not only

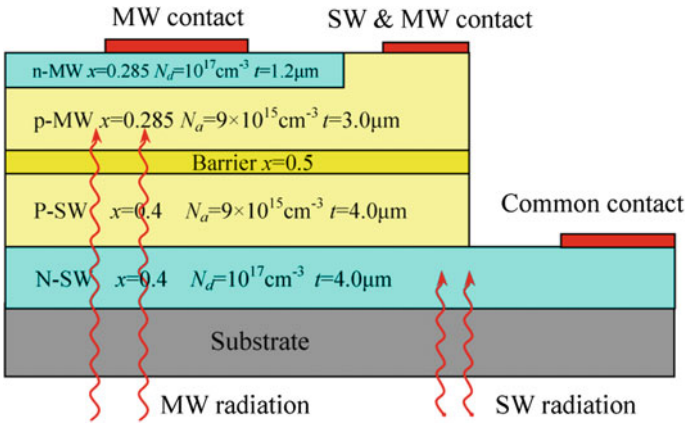


Fig. 2.55 Schematic of n^+-p-p-p-n HgCdTe two-color infrared detector

Table 2.17 Material and structural parameters of the two-color $Hg_{1-x}Cd_xTe$ photodetector

Parameters (SW)	Value (Units)	Parameters (MW)	Value (Units)
Cd molar fraction (x) of N region	0.4	Cd molar fraction (x) of N region	0.285
Cd molar fraction (x) of P region	0.4	Cd molar fraction (x) of P region	0.285
Temperature	77 (K)	Temperature	77 (K)
SRH lifetime for electron and hole	τ (ns)	SRH lifetime for electron and hole	τ (ns)
Doping density of N region	1×10^{17} (cm^{-3})	Doping density of N region	1×10^{17} (cm^{-3})
Doping density of P region	9×10^{15} (cm^{-3})	Doping density of P region	9×10^{15} (cm^{-3})
Thickness of N region	4 (μm)	Thickness of N region	1.2 (μm)
Thickness of P region	4 (μm)	Thickness of P region	3 (μm)

absorb and respond to the SW radiation, but also acts as a window for any MW radiation.

To further study the physical mechanism of the spectral response, two-dimensional numerical simulations using Sentaurus Device was carried out. For plain drift–diffusion simulation, the well-known Poisson equation and continuity equations are used. The carrier generation–recombination process consists of Shockley–Read–Hall, Auger and optical generation–recombination terms. Additionally, tunneling effects such as band-to-band and TATs are included in the continuity equations by incorporating them as additional generation–recombination processes. Comparison of the simulation results with different interface generation–recombination velocities and carrier lifetimes in the highly doped n region are presented in Fig. 2.57. The value of minority carrier lifetime in the bottom n-SW

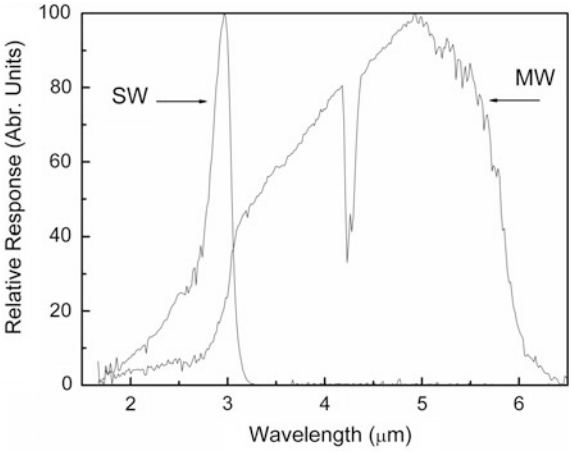


Fig. 2.56 Experimentally measured spectral photoresponse of preliminary integrated HgCdTe two-color infrared detector

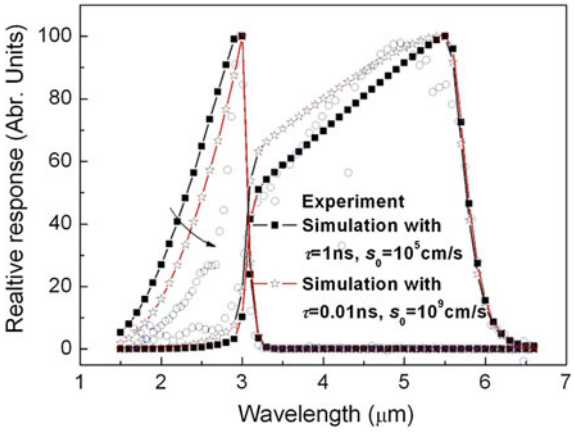


Fig. 2.57 Spectral photoresponse of a newly proposed integrated HgCdTe two-color infrared detector employing a heterostructure SW *pn* junction in place of the homostructure SW *pn* junction. The unfilled circles represent the experimental data and the solid lines with symbols represent the simulation results. τ is the maximum carrier lifetime, and s_0 is the surface recombination velocity in the bottom n layer

layer is assumed as 0.1 ns in the simulation. The simulated photoresponse with high interface generation–recombination velocity and short carrier lifetime is in good agreement with the experimental data, reaffirming the validity of the above analysis as well as providing an additional avenue for further device performance improvement through simulations.

Due to the high generation–recombination velocity at the buffer interface and short carrier lifetime of the highly doped *n* region, photogenerated carriers undergo significant recombination at the interface region which induces the short wavelength narrow effect of the integrated HgCdTe two-color infrared detector. A reasonable solution is to change the bottom *pn* junction from the homostructure to a heterostructure. This should enable SW radiation to transmit through the highly doped bottom *n* region and be absorbed in the low-doped *P* region which is further away from the buffer interface, hence reducing undesirable interface effects on the photo-generated carriers. Figure 2.58 illustrates a schematic of the proposed

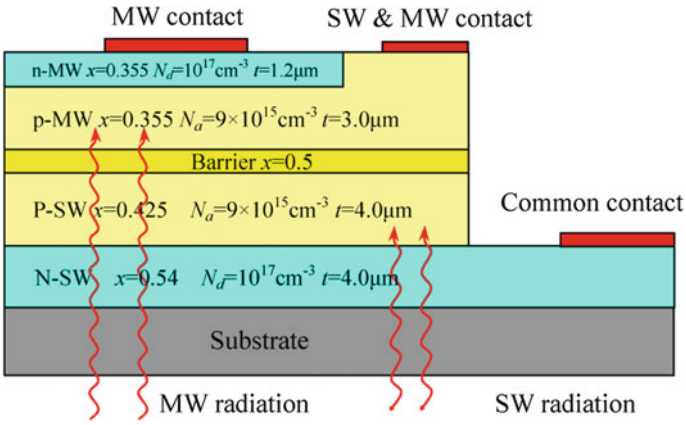


Fig. 2.58 Schematic of proposed heterostructure n^+-p-p-p-n HgCdTe two-color infrared detector incorporating a SW pn heterostructure in place of the SW pn homostructure

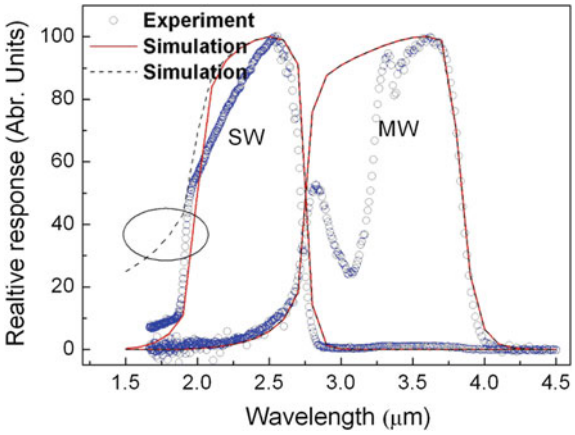


Fig. 2.59 Spectral photoresponse of a newly proposed integrated HgCdTe two-color infrared detector employing a heterostructure SW pn junction in place of the homostructure SW pn junction. The unfilled circles represent the experimental data while the *solid* and *dashed lines* represent the simulation results. The maximum carrier lifetime in the simulation for the solid line is 0.1 ns for the bottom *n* layer and 10 ns for the other layers. The maximum carrier lifetime in the simulation for the dashed line is 10 ns for the bottom *n* layer and the other layers

structure where the heterostructure of the bottom *pn* junction consists of a bigger energy gap *n* region with high doping density and a smaller energy gap *P* region with low doping density. Figure 2.59 shows the spectral photoresponse of the newly proposed integrated HgCdTe two-color infrared detector with a heterostructure SW *pn* junction with device parameters given in Fig. 2.58. The proposed integrated HgCdTe two-color infrared detector with a heterostructure SW *pn* junction has a much larger FWHM than that of the initial integrated HgCdTe two-color infrared detector, resulting in an increase in device performance.

High temporal and spatial coherent simultaneous long wavelength/mid-wavelength (LW/MW) two-color focal plane array (FPA) infrared detection is the cutting-edge technique for third-generation infrared remote sensing [82–84]. HgCdTe LW/MW two-color infrared detectors were designed and fabricated [82]. The top long wavelength and bottom mid-wavelength infrared planar photodiodes are processed by selective B^+-implantation after etching the long wavelength epi-layer into a curvature and exposing the mid-wavelength layers for the implantation of the *n* region of the MW photodiode by a micro-mesa array technique. Non-planar boron ion implantation and metallization are developed by using photoresist spray coating technology for the two-color HgCdTe IR detector. Figure 2.60a shows the schematic of two-color LW/MW HgCdTe photovoltaic detector with an n_1^+-p_1-p-p_2-n_2^+ architecture. Under back-illumination, the bottom p_2-n_2^+ junction absorbs MW

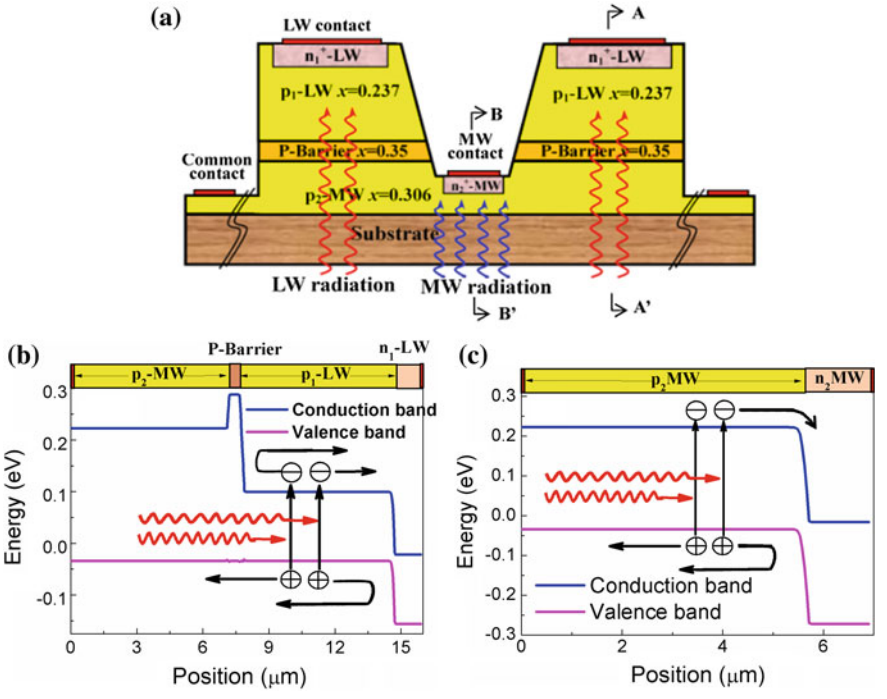


Fig. 2.60 **a** Schematic of grooved HgCdTe two-color infrared detector. **b** Equilibrium energy band diagram cut at A-A′. **c** Equilibrium energy band diagram cut at B-B′

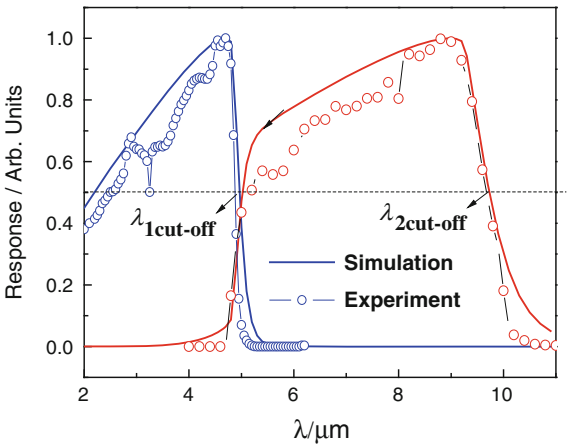


Fig. 2.61 Experimental (*dotted*) and simulated (*solid*) spectral photoresponse of LW/MW HgCdTe two-color infrared detector with cut-off wavelengths of $\lambda_{1\text{cut-off}} = 4.8$ μm and $\lambda_{2\text{cut-off}} = 9.7$ μm for the LW and MW diodes respectively. A voltage of 0.01 V is used in simulation and experiment

radiation for wavelengths up to its cut-off value. This p-type layer junction for MW light also acts as a transparent window for LW light to pass through and be absorbed by the horizontal n_1^+-p_1-P-p_2 junction with LW cutoff, as shown in Figs. 2.60b, c.

Figure 2.61 shows the spectral photoresponse for the HgCdTe two-color infrared detector with the proposed structure from the numerical simulations. The cut-off wavelengths of the LW and MW diode are 4.8 and 9.7 μm, respectively. It can be observed that the cut-off wavelength of the MW diode is close to the cut-on wavelength of the LW diode. This effectively enables the MW diode to not only absorb and respond to the MW radiation, but also acts as a window for any LW radiation. The simulated spectral photoresponse is also compared to that of experimental results showing that the simulation and experiment are self-consistently in good agreement. The calculated peak detectivity is 3.2×10^{11} and 4.3×10^{10} $\text{cmHz}^{1/2}$ for the MW and LW photodiodes, respectively. The proposed LW/MW HgCdTe two-color structure is very promising for third-generation intelligent infrared imaging.

2.2.4.3 The Relationship of Spectral Response with Minority Carrier Lifetimes

Integrated HgCdTe two-color infrared device, with multilayer heterojunction structure, is developed based on monochromatic infrared device. Composition, doping concentration distribution and thickness of each layer are needed to be designed precisely. This type of device technology is more complicated than for a monochromatic device, greatly increasing the difficulty of device performance control. Ballet [80], Ferret [85], Baylet [86], etc. studied the fabrication process and performance characterization of two-color devices. Jozwikowski and Rogalski [87] calculated the photoelectric gain and the noise of two-color devices. The spectral responses of a two-color device as a function of temperature are simulated by

Bellotti et al. [88]. However, in addition to the device response and noise characteristics of the specific structure, the other key material factors and structure parameters on the devices also require in-depth research. This research can provide a theoretical guide for optimizing device performance and fabrication control parameters such as the relationship between the performance of a HgCdTe two-color device and the lifetime of minority carriers, the components of the barrier layer for inhibiting electrical cross talk, etc. The basic structure of HgCdTe two-color device is shown in Fig. 2.60.

The spectral response and quantum efficiency characteristics of a two-color device with different SRH lifetimes are shown in Fig. 2.62a, b. Under the same SRH recombination lifetime, the quantum efficiency of the MW2 band is significantly below that of the MW1 band The main reason for this is that the duty ratio of MW2 (75 %) is lower than that of MW1 (nearly 100 %). Figure 2.62c illustrates the relationship of quantum efficiency to SRH recombination lifetime under the typical wavelengths of MW1, MW2 band. It is shown that the quantum efficiency is strongly dependent on the SRH recombination lifetime. When the SRH lifetime is less than 10 ns, the quantum efficiency drops quickly with the SRH lifetime, the quantum efficiency of MW1 band is reduced by more than 30 % when the SRH lifetime drops from 10 to 1 ns, while the quantum efficiency of MW2 band is

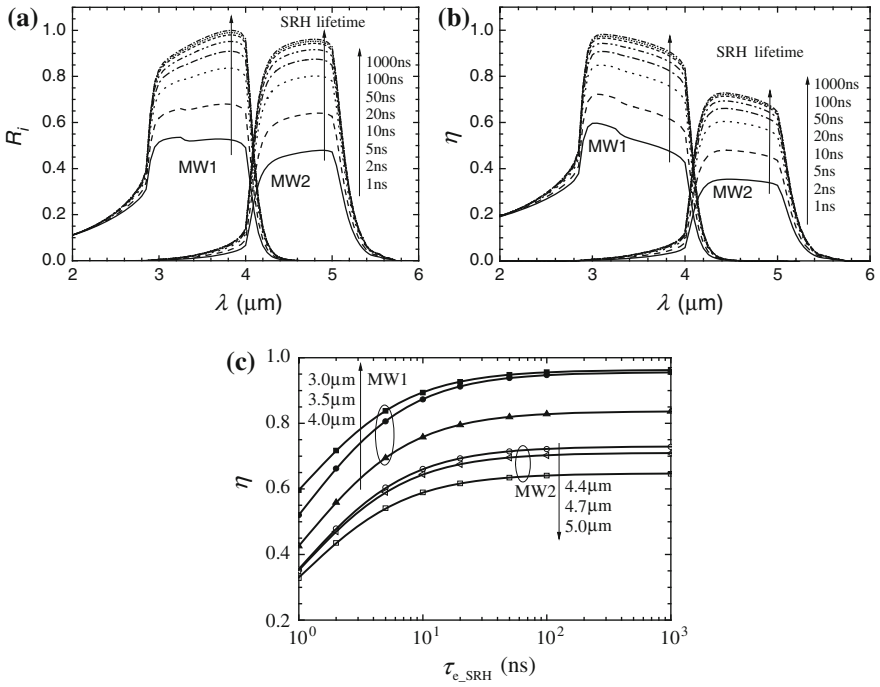


Fig. 2.62 The spectral response of two-color detector varies with the SRH electronic lifetime in p zone. **a** Spectral response R_i, **b** Quantum efficiency η, **c** The relationship between the quantum efficiency and the SRH electronic lifetime (*Dotted lines* represent the results of simulation calculation, *solid lines* represent the fitting results using formula (2.52))

reduced by almost 50 %; when the SRH lifetime varies between 100 and 10 ns, the quantum efficiency of MW1 and MW2 bands are both little changed.

According to the SRH recombination lifetime, the total life τ_e and diffusion length L_e of the electron can be calculated as:

$$\frac{1}{\tau_e} = \frac{1}{\tau_{e\_R}} + \frac{1}{\tau_{e\_Auger}} + \frac{1}{\tau_{e\_SRH}} \tag{2.52}$$

$$L_e = \sqrt{\frac{kT}{q}\mu_e\tau_e} \tag{2.53}$$

where τ_{e_R}, τ_{e_Auger} are the Radiation recombination lifetime and Auger recombination lifetime, μ_e is the electron mobility, and K, T, q are the Boltzmann's constant, temperature, and basic charge respectively. The Radiative and Auger recombination lifetime of HgCdTe material are the function of temperature and components, and have nothing to do with the technology process. However, the recombination lifetime of SRH is closely related to the technology process.

2.2.4.4 The Relationship of Cross Talk with the Barrier Layer

As can be seen from the response in Fig. 2.62, the cross talk of MW1-to-MW2 is greater than that of MW2-to-MW1 under the different carrier lifetimes. There are two formation mechanisms of the cross talk: optical cross talk and electrical cross talk. Optical cross talk is caused by the absorption of MW1 radiation in the MW2 area. The optical cross talk has been discussed in detail in Chap. 1, therefore the focus here is to study the electrical cross talk problem.

The spectral responses of the two-color device (see Fig. 2.63) with different barrier layer components (Δx_b) are shown in Fig. 2.64. Δx_b is defined as the

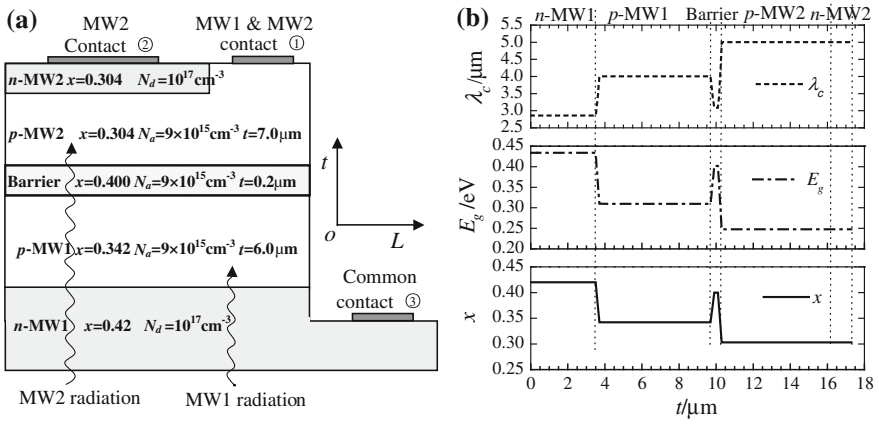


Fig. 2.63 The structure of two-color detector. **a** The structure of n-p-p-p-n $Hg_{1-x}Cd_xTe$ two-color detector, **b** Device component x, forbidden band width E_g and the cutoff wavelength λ_c varies with the direction of thickness

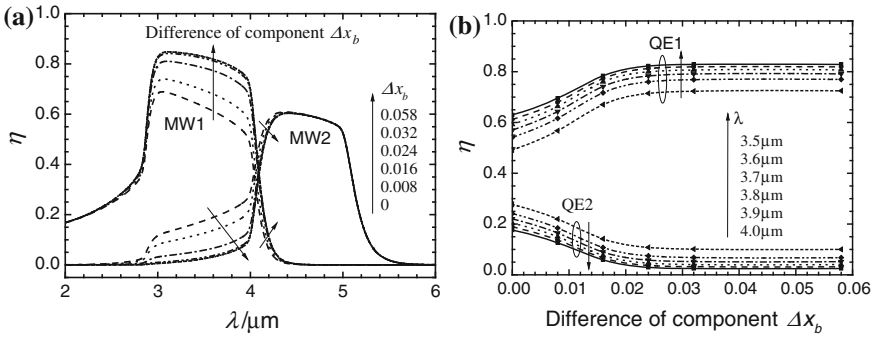


Fig. 2.64 The spectral response (quantum efficiency) of two-color device with different difference of component. **a** Spectral response; **b** Quantum efficiency varies with the difference of barrier layer components Δx_b

difference between the highest composition x_b of barrier layer and that of p-MW1 area. It is evident from Fig. 2.64 that the component of the barrier layer has an important effect on the quantum efficiency of MW1 band (3.0–4.0 μm): the quantum efficiency of MW1 increases with the component of barrier layer, while the quantum efficiency of MW2 decreases with the component of barrier layer. The quantum efficiency changes when Δx_b varies from 0 to 0.02 although it remains unchanged when Δx_b exceeds 0.03; when Δx_b varies from 0 to 0.032, the quantum efficiency of MW1 is increased from 49 to 72 % at the radiation wavelength of 4.0 μm, while for MW2 it is decreased from 28 to 10 %. The output ratio of MW2 to MW1 is known as the cross talk of MW1 to MW2. For incident light in MW2 band (4.4–5.0 μm), the spectral response has nothing to do with the component of barrier layer. Therefore, the design of the barrier layer is only related to the incident light in the MW1 band.

As mentioned previously, incident light in MW1 band is absorbed in the p-MW1 area and the photogenerated electrons may diffuse into either the MW1 junction or MW2 junction. If the photogenerated electrons diffuse into the MW2 junction, this results in the electrical cross talk of the MW1 to MW2 signal. However, if there are any barrier layers, the diffusion of electrons from p-MW1 to p-MW2 area will be restrained, and cross talk is reduced. That is why the barrier layer is introduced in the middle of a two-color device (see Fig. 2.62). The different components of the conduction band barrier layer are shown in Fig. 2.65 under the equilibrium state. The barrier height is calculated:

$$\Delta V_b \approx \Delta E_g = E_{g\_b} - E_{g\_pMW1} \tag{2.54}$$

where ΔV_b is defined as the difference of the conduction band between barrier layer and p-MW1 diffusion zone, and ΔE_g is the difference of the band gap between barrier layer (E_{g_b}) and p-MW1 diffusion zone ($E_{g\_p\text{-}MW1}$). Therefore, ΔV_b and ΔE_g

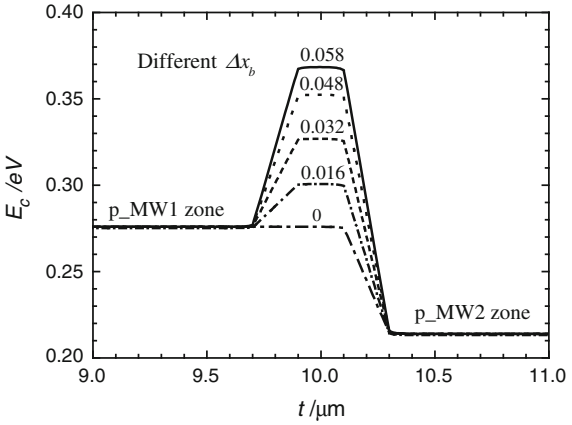


Fig. 2.65 The formation of the barrier with different components of barrier layer in the conduction band under equilibrium

are obviously dependent on the different components between the barrier layer and p-MW1 area (Δx_b). For HgCdTe material, ΔV_b and ΔE_g are approximately linear to Δx_b. In the device simulation, the transition of components is introduced at the interface between the barrier layer and p-MW1, p-MW2 area, which is consistent with the actual device. The thickness in the barrier layer is 0.2 μm in the simulation.

In order to accurately study the relationship of cross talk with the barrier layer's components (x_b), the cross talk with different barrier layer's components is calculated in detailed under laser illumination with the representative wavelength of 3.5, 4.0 μm in MW1 band. As shown in Fig. 2.66, the barrier height ΔV_b with each component in the barrier layer is obtained. Figure 2.65 shows the formation of the barrier in the conduction band with different components of barrier layer under equilibrium state.

To ensure the cross talk caused by carrier diffusion is maximally suppressed, theoretical results show that the component of barrier layer x_b should be at least 0.03 higher than that of p-MW1 area for the two-color device (see Fig. 2.63). For

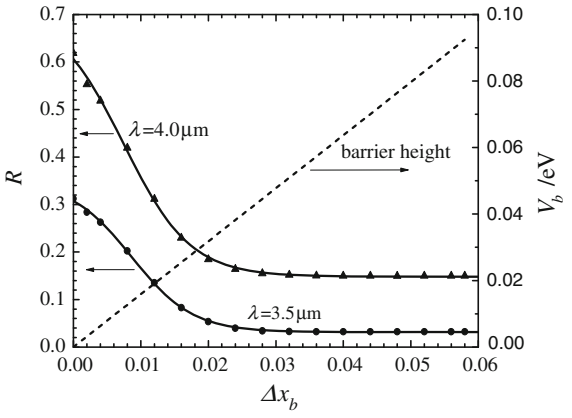


Fig. 2.66 The cross talk of MW1-to-MW2 (R) with the barrier height (ΔV_b) and the difference of component (Δx_b)

the design of the barrier layer's components in the actual device, the only applied standard is that the electrical cross talk should be far less than the optical cross talk.

2.2.4.5 The Optimum Thickness of the Absorption Layers

Compared with monochromatic devices, multicolor detectors are unique from the performance indicators to the device structure. Absorption layers of p-MW1 and p-MW2 are the locations where the incident light is absorbed and translated into electrical signals. The parameters of the absorption layers directly affect the performance of the device, including the material and device parameters such as components, minority carrier lifetime, the thickness of absorption zone, etc.

According to the previous discussion, as long as the components of barrier layer are appropriately designed, the electrical cross talk can be greatly reduced, even negligible. In this case, the optical cross talk becomes the component of the total cross talk. The thickness of the absorption zone has an important influence on the optical cross talk and the quantum efficiency of two-color device. This section mainly studies the optimization of the absorption area from two aspects: the quantum efficiency and cross talk (optical).

The optimization of the absorption area is analyzed under an ideal condition: the lifetime of SRH recombination is assumed to be very long, so that the main process of recombination is determined by the Radiative recombination, and Auger recombination. In this simulation, the barrier layer is high enough ($\Delta x_b = 0.058$) so that the electrical cross talk is completely suppressed. The spectral efficiency of the two-color device with different thickness of p-MW1 absorption area is shown in Fig. 2.67. For the back-illuminated n-p-p-pn two-color structure, shortwave radiation passes from p-MW1 to p-MW2 area. It is first absorbed in the p-MW1 area, and then the remainder is absorbed in p-MW2 area.

It is observed that with an increase in the thickness of p-MW1, more light is absorbed and a larger the signal from the MW1 is obtained. However, less infrared radiation reaches p-MW2, and a smaller signal of MW2 is obtained. The quantum efficiency of MW1 gradually moves closer to 100 % with the increase of the thickness of p-MW1 under the laser wavelength of 3–4 μm in the MW1 band, as shown in Fig. 2.67c. When the thickness of p-MW1 is about 8 μm, the rate of increase in quantum efficiency of MW1 slows until it becomes static. Therefore, in order to ensure that the internal quantum efficiency of MW1 is above 90 %, the thickness of p-MW1 should be 8–10 μm.

The spectral efficiency of the two-color device with different thickness of p-MW2 absorption area is shown in Fig. 2.68. According to the previous theoretical analysis, the thickness of MW1 is chosen as 10 μm in the calculation. Compared with the case of p-MW1 zone, the thickness of p-MW2 zone only affects the quantum efficiency of MW2, and makes little impact on the quantum efficiency of MW1. It is found that the greater the thickness of p-MW2, the larger the quantum efficiency of MW2. When the thickness of p-MW2 zone varies from 4 to 7 μm, the quantum efficiency of MW2 will increase quickly; when the thickness of p-MW2 is

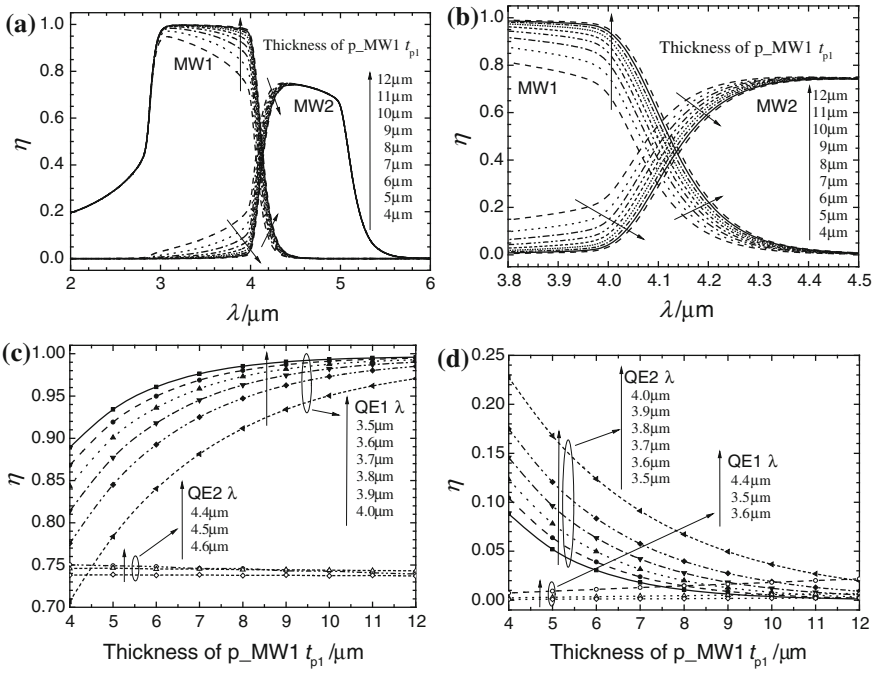


Fig. 2.67 The quantum efficiency with different thickness of p-MW1 zones. **a** Spectral response, **b** A magnified view of the middle of Spectral response, **c** Quantum efficiency along with the thickness of p-MW1 zone under the key radiation wavelength, **d** The quantum efficiency with the thickness of p-MW1 zone under the specific radiation wavelength, which has an effect on the cross talk

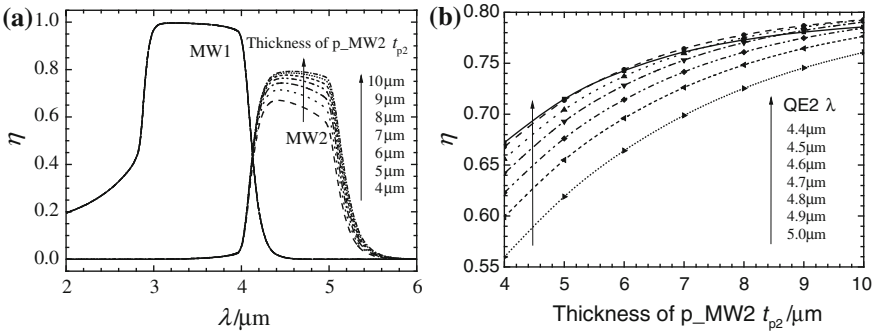


Fig. 2.68 The quantum efficiency with different thickness of p-MW2 zones. **a** Spectral response, **b** Quantum efficiency along with the thickness of p-MW2 zone under the key radiation wavelength in MW2 band

about 8 μm, the increasing trend of quantum efficiency will slow down. However, in addition to part of the transverse junction, the MW2 zone also owns part of longitudinal junction (see Fig. 2.63). Actually, the illumination area is greater than the *pn* junction area. Therefore, the right side of diode in the MW2 zone also has a contribution to quantum efficiency, and the quantum efficiency of MW2 may finally be more than 75 %.

On the other hand, for the MW2 diode, medium wave radiation is mainly absorbed in the p-MW2 area near the substrate region. The greater the thickness of the p-MW2 area, the more the absorption center is further away from the junction area, and the minority carrier (electrons) are easier to be lost due to the recombination, which results in the decrease of quantum efficiency of MW2 with the thickness of p-MW2 area.

The relationship of cross talk to the thickness of p-MW2 zone is shown in Fig. 2.69. The cross talk of MW1-to-MW2 increases slightly with the thickness of p-MW2 zone, the reason is that the quantum efficiency of MW2 increases slightly with the thickness of p-MW2 under the laser wavelength of MW1 band. On the other hand, the cross talk of MW2-to-MW1 decreases with the thickness of p-MW2 zone, the reason is that the quantum efficiency of MW1 has nothing to do with the thickness of p-MW2 under the laser wavelength of MW2 band, but, the quantum efficiency of MW2 increases with the thickness of p-MW2 zone. Therefore, the ratio of the former to the latter (cross talk) decreases with the thickness of p-MW2 zone.

Taken as a whole, the cross talk of MW1-to-MW2 and MW2-to-MW1 both changes little with the thickness of p-MW2 zone. In conclusion, the optimization of p-MW2 thickness can be based on the quantum efficiency of MW2. According to the above analysis, together with the principle of reducing device thickness as far as possible, p-MW2 with a thickness of 8 μm is a reasonable choice. The quantum efficiency of MW2 is over 70 %, the monochromatic cross talk of MW1-to-MW2 and MW1-to-MW2 are less than 3.9 and 2.4 %, as shown in Fig. 2.70.

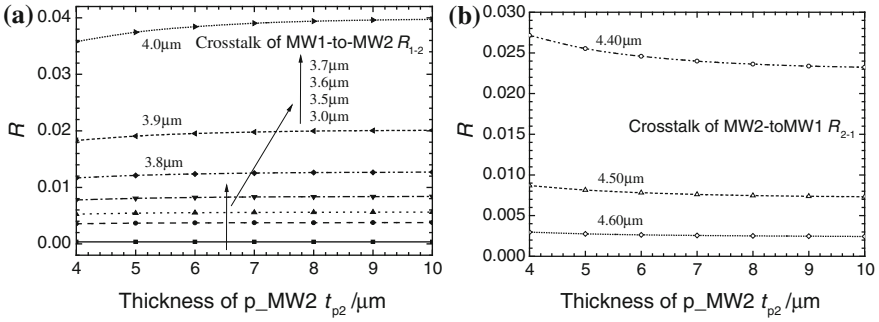


Fig. 2.69 The cross talk with different thickness of p-MW2 zone. **a** The cross talk of MW1-to-MW2; **b** The cross talk of MW2-to-MW

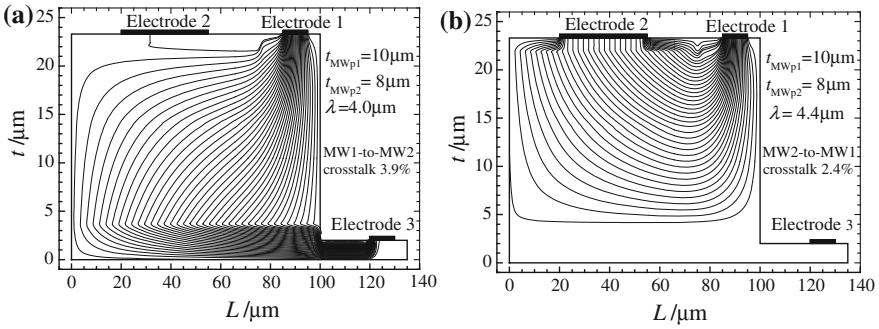


Fig. 2.70 The distribution of photocurrent in device under the optimized thickness of absorption zone. **a** λ = 4.0 μm, **b** λ = 4.4 μm

2.3 Methods of Extracting Parameters from HgCdTe Materials and Chips

2.3.1 Extracting Device Parameters by Electrical Method

Either a current–voltage (*I–V*) curve or a resistance–voltage(*R–V*) curve is a standard characterization to quantify the dark current performance of a HgCdTe photovoltaic detector. One can obtain *R–V* characteristics by mathematical differential calculations from *I–V* characteristics which are measured by experiments, and also can obtain *R–V* characteristic directly by implement differential signal from the experiment. As is known, the properties of HgCdTe photodiodes can be improved by minimizing dark current. Therefore, it is important to quantitatively analyze *I–V* or *R–V* curves to obtain device parameters and reveal roles of the different physical mechanisms of the parameters in HgCdTe photodiodes.

2.3.1.1 Extract Device Parameters for Long Wavelength HgCdTe Photodiodes

At present, the typical fitting process for the *I–V* or *R–V* curves are empirically made to extract parameters in a sequential mode [89–91]. Various dark current mechanisms are used in different bias regions in which they have the dominant effect on the fitted curve. However, many HgCdTe photodiodes have more than one mechanism in most bias regions having comparable effects on the fitted curve. Then the extract parameters by the sequential mode will have large error. A method that could fit the *R–V* curves simultaneously in all bias regions will be valuable to device designers. Few studies on the simultaneous fitting approach have been reported on *R–V* curves analysis, owing to its time-consuming procedure and the multi-minimum problem in mathematics. Moreover, the errors of the extracted

parameters have not been analyzed in the previously published works, although it is very important for the estimation of the quality of HgCdTe photodiodes. In this chapter, a data processing technique in the simultaneous mode [92] has been developed. It is shown that this method can be used to fit the *R–V* curves of long wavelength HgCdTe n-on-p photodiodes for the determination of the device parameters and their errors.

The measured *I–V* curves generally include photocurrent due to the background illumination. Based on the assumption of low photon injection, the photocurrent can be regarded as bias-independent [93]. The *R–V* curves of illuminated photodiodes are the same as those of unilluminated ones. Therefore, the *R–V* curves are taken to be the fitting object in our fitting process. Then the effects of the current offsets in the device will be eliminated in our fitting procedure. The measured *R–V* curves of HgCdTe photodiodes are fitted by a dark current model including diffusion current (I_{diff}), generation–recombination current (I_{gr}), trap-assisted tunneling current (I_{tat}), and band-to-band tunneling current (I_{bbt}).The bias voltage is an effective bias $V_e = V_d - I \times R_s$ corrected by the series resistance R_s. Here, V_d is the applied voltage, I is the total dark current, and R_s is the series resistance. The total dynamic resistance is given as:

$$R_{\text{fit}} = \left(\frac{1}{R_{\text{diff}}} + \frac{1}{R_{\text{gr}}} + \frac{1}{R_{\text{tat}}} + \frac{1}{R_{\text{bbt}}} \right)^{-1} + R_s \tag{2.55}$$

At present, the mainstream of HgCdTe focal plane detector design is n-on-p diodes. These diodes are fabricated using n-type boron ion implantation on mercury vacancy doped HgCdTe material. For planar n-on-p HgCdTe photodiodes, there are six fitting parameters to be extracted from *R–V* curves as follows: the dopant density N_d in *n*-region, the ratio of mobility to lifetime of electrons μ_n/τ_n in *p*-region, the effective lifetime τ_0 in the depletion region, the relative energy position of trap level E_t/E_g and its density N_t in the depletion region, and the series resistance R_s. A theoretical *R–V* curve can be obtained from substituting a set of the six parameters into Eq. (2.55). In our fitting procedure, the algorithm is to minimize the function value of $F = \sum_{i=1}^{N} \left[\log(R_{\text{fit}}(V_{d\,i})) - \log\left(R_{\text{exp}}(V_{d\,i})\right) \right]^2$, where R_{exp} is the experimental data, and N is the number of data. An initial value should be input first, and then the parameters vary in the their corresponding range to minimize the fitting variance F by using a standard nonlinear gradient search method combined with the algorithm of reconstruction of the whole-region-minimum for *N*-dimensional function [94]. Figure 2.71 is the flowchart of the fitting procedure. Through such a process, we can obtain a theoretical *R–V* curves which correspond to the experiment as well as a set of fitting parameters. The range of error is difficult to determine due to the fact that F is a multi-valued minimum function consisting of six parameters in six dimensional space and the fitting parameters will change when the initial value is varied through this method. Moreover, the range of these parameters is large, and the fitting time also increases accordingly, so looking for a quick and accurate fitting approach is necessary.

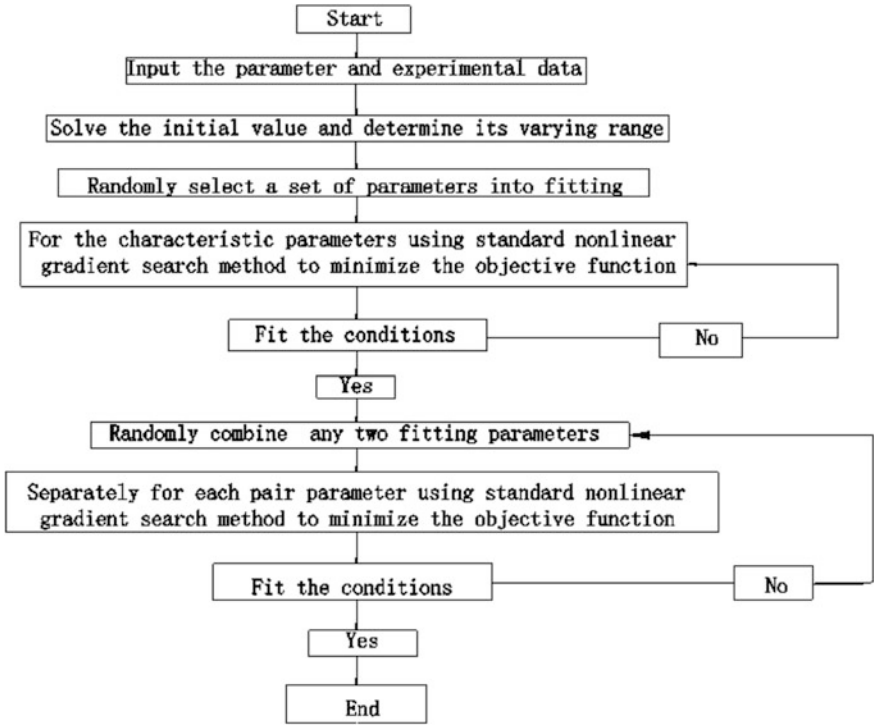


Fig. 2.71 Flow chart of the fitting procedure

For the first step in the fitting process, one must determine the initial values of the fitting parameters carefully. According to Eq. (2.55), the following equation comes into existence for each bias value:

$$R_{\text{exp}} - \left(\frac{1}{R_{\text{diff}}} + \frac{1}{R_{\text{gr}}} + \frac{1}{R_{\text{tat}}} + \frac{1}{R_{\text{bbt}}}\right)^{-1} - R_{\text{s}} = 0 \tag{2.56}$$

There are six unknown variables in this equation, namely six fitting parameters. A set of six variable equations is obtained by choosing six characteristic points on the measured R–V curve. The initial values of the six fitting parameters can be determined by solving these equations set with the method of iteration based on the Taylor series expanded nonlinear terms. A given set of initial values is still needed for the iteration. A fault-tolerant problem in the iteration may lead to a divergent solution once these initial values deviate from the solution to a certain extent. Therefore, it is impractical to directly solve the equations set.

In order to solve Eq. (2.56) for initial values, we use the sequential mode algorithm because it is well known that the four dark current mechanisms have different contributions to the dark current under different biases. Generally, for long

wavelength HgCdTe diodes, I_{diff} dominates the dark current under a large forward bias; I_{gr} becomes dominant near zero forward bias; I_{tat} has the largest contribution under middle reverse bias; and I_{bbt} dominates under large reverse bias. The series resistance R_s, which is comparable to the junction resistance only under a large forward bias, can be omitted in the initial value determination. Then, Eq. (2.56) can be solved as follows.

First, I_{bbt} is only related to the dopant density N_d in n-region. In the large reverse bias region, I_{bbt} dominates the dark current. Equation (2.56) can be replaced by

$$R_{exp} - R_{bbt} = 0 \tag{2.57}$$

Substituting a measured point in the large reverse bias region into Eq. (2.57) with one variable, the initial value of N_d is obtained.

Likewise, for the large forward bias region, I_{diff}, which depends on N_d and μ_n/τ_n, is the dominant dark current. Equation (2.56) can be simplified as

$$R_{exp} - R_{diff} = 0 \tag{2.58}$$

Substituting N_d and a measured point in the large forward bias region into Eq. (2.58) with one variable, the initial value of μ_n/τ_n is obtained.

For the small forward bias region, the combination of I_{diff} and I_{gr} dominates the dark current. Equation (2.56) is approximated as

$$R_{exp} - \left(\frac{1}{R_{diff}} + \frac{1}{R_{gr}}\right)^{-1} = 0 \tag{2.59}$$

I_{gr} is related to N_d and the effective lifetime τ_0 in the depletion region. Substituting N_d, μ_n/τ_n and a measured point in the small forward bias region into Eq. (2.59) with one variable, the initial value of τ_0 is obtained.

Finally, for the middle reverse bias region, the combination of I_{gr}, I_{tat} and I_{bbt} dominates the dark current. From Eq. (2.56) one has

$$R_{exp} - \left(\frac{1}{R_{gr}} + \frac{1}{R_{tat}} + \frac{1}{R_{bbt}}\right)^{-1} = 0 \tag{2.60}$$

I_{tat} is related to N_d, the relative energy position of trap level E_t/E_g, and the trap density N_t. So Eq. (2.60) is a two-variable equation, which can be solved using a joint method of search and iteration. In the search process, for each value of E_t/E_g in the reasonable range (0.2–0.8) with the interval of 0.01, two values of N_t are obtained from Eq. (2.60) for two of the measured points in the middle reverse bias region. The searched value of E_t/E_g, for which the difference between the two values of N_t is the smallest, is assigned to be the initial value for the iteration method. Then in the iteration process, at step-1, N_t is obtained by substituting N_d, τ_0, E_t/E_g and one measured point in the middle reverse bias region into Eq. (2.60). At step-2, a new E_t/E_g is obtained by substituting N_d, τ_0, N_t and another measured

point in the middle reverse bias region into Eq. (2.60). Then the step-1 is repeated by taking the new E_t/E_g obtained in the previous step-2 as a new initial value. This iteration loop does not stop until the solution comes to convergence. By so doing, the initial values of both E_t/E_g and N_t are obtained. Now we have obtained all the initial values for the six fitting parameters. Accordingly, the variation range for each parameter can be assigned. Typically, the upper limit of the variation range is twice of the initial value, and the lower limit is half of it.

To evaluate the fitting accuracy, the errors of the fitting parameters should be analyzed. The error of the dynamic resistance can be expressed by $\Delta R = |R_{fit}(V_d) - R_{exp}(V_d)|$. Assuming that this error is effectively resulted from one of the six fitting parameters, this parameter's error can be given as,

$$\sigma_{x_j} = \left|\frac{\partial R_{fit}}{\partial x_j}\right|^{-1} \cdot \Delta R \tag{2.61}$$

where, x_j is the six fitting parameters, and σ_{x_j} is the error of the parameter x_j. This value reflects the upper limit of the error for the fitting parameters to a certain extent.

The R–V curves of HgCdTe photodiodes, however, have the feature that the four types of dark current mechanisms have different contributions to the dark current under different biases. The error of μ_n/τ_n, for example, will be magnified if it is analyzed when I_{gr} dominates the dark current. This error will be even infinitely enlarged under large reverse bias. Therefore, the simple averaging of σ_{x_j} over the whole bias range is meaningless for evaluating the accuracy of the fitting parameters. The different bias regions at which each fitting parameter has the largest influence to the R–V curve should be ascertained. We define $\eta = \left|\frac{\partial R_{fit}(V_d)}{\partial x_j} \cdot \frac{1}{R_{fit}(V_d)}\right|$ as a criterion of the sensitivity. The larger the value of η is, the greater the influence on ΔR from the corresponding fitting parameter is. The average of σ_{x_j} for those biases is taken as the theoretical estimation error for the specific parameter x_j, where η is larger than half of its maximum.

To demonstrate the determination process of the fitting parameters errors and to verify its applicability, some artificial R–V curves, derived with the combination of the generated noise current and the calculated current with preassigned fitting parameters, are fitted as the experimental data by our fitting program. The differences between the fitting parameters obtained from the fittings and the preassigned parameters are defined as the real errors. The fitting results given in this work are the averaged results from ten fits in different fitting paths. Figure 2.72 shows the fitting results of the artificial R–V curves whose noise level is 1, 2, and 3 %, respectively. The material and device parameters, which are not used as fitting parameters, are summarized in Table 2.18, where x is the Cd composition, T is the measurement temperature, A is the junction area, N_a is the dopant densities at p-region, and μ_p is the mobility of excess holes. The voltage range for the fitting

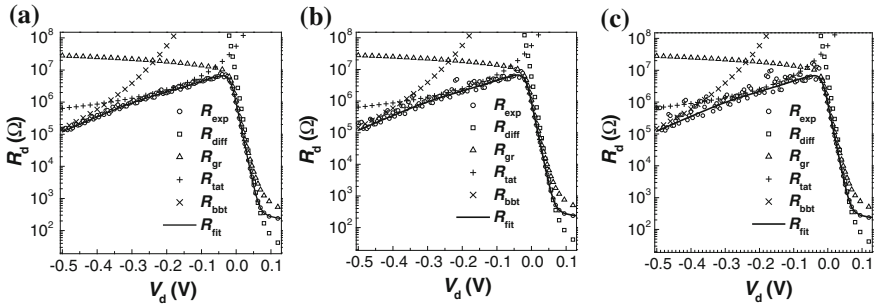


Fig. 2.72 Fitting results of the artificial *R–V* curves with the generated noise level of 1 % (**a**), 2 % (**b**), and 3 % (**c**). Reprinted with permission from Ref. [92] © 2006, American Institute of Physics

Table 2.18 Input parameters for the fitting program

X	N_a (cm^{-3})	μ_p (cm^2/Vs)	A (μm^2)	T (K)
0.233	8×10^{15}	633	28×28	77.4

Reprinted with permission from Ref. [92] © 2006, American Institute of Physics

Table 2.19 Comparisons of the assigned values of the six fitting parameters and the fitting results of the theoretically generated *R–V* curves whose noise level is 1, 2, and 3, respectively

	N_d (cm^{-3})	μ_n/τ_n ($cm^2/V\ s^2$)	τ_0 (ns)	E_t/E_g	N_t (cm^{-3})	R_s (Ω)
Assigned	1.676×10^{16}	9.51×10^{13}	0.207	0.491	1.968×10^{12}	200
Noise 1 %	1.683×10^{16}	1.05×10^{14}	0.192	0.4889	1.988×10^{12}	205.2
	(6×10^{12})	(1×10^{12})	(0.001)	(0.0004)	(1×10^{10})	(0.2)
Noise 2 %	1.676×10^{16}	9.92×10^{13}	0.1834	0.4866	2.028×10^{12}	204.2
	(5×10^{12})	(9×10^{11})	(0.001)	(0.0003)	(9×10^{9})	(0.2)
Noise 3 %	1.667×10^{16}	1.008×10^{14}	0.1843	0.4869	1.962×10^{12}	204.6
	(3×10^{12})	(3×10^{11})	(0.0003)	(0.0001)	(3×10^{9})	(0.1)

Data in brackets are the standard errors of the results from ten fits in different fitting paths) Reprinted with permission from Ref. [92] © 2006, American Institute of Physics

is –0.48–0.12 V. The assigned values and the fitting results of the fitting parameters are listed in Table 2.19.

Figure 2.72 shows the fitting results of the artificial *R–V* curves. It can be seen from Table 2.19 that when the noise level is smaller than 3 %, both the deviation of the fitting results from the assigned ones and the discreteness among the results from the ten fits in different fitting paths are very small. In this noise level, our data processing approach can give a good fitting precision for quantitative analysis of the *R–V* curves. Since the noise level is generally smaller than 3 % under regular experimental conditions, our fitting method is feasible in practice.

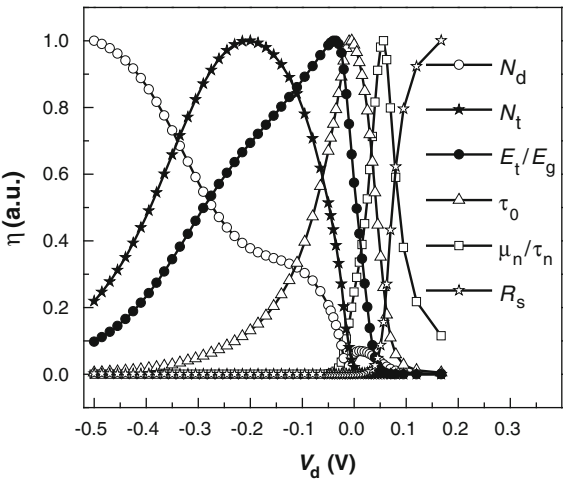


Fig. 2.73 Normalized sensitivity criterion (η) versus the bias for the six fitting parameters. Reprinted with permission from Ref. [92] © 2006, American Institute of Physics

Figure 2.73 shows the dependence of the normalized sensitivity criterion η on the bias. As one can see the largest η for each fitting parameter is located at a different bias region where its related dark current mechanism dominates the dark current. Accordingly, the estimation error for each fitting parameter can be obtained. The error range of the fitting parameters can be defined as follows: the standard error of the ten fits in different fitting paths (in Table 2.19) is regarded as the minimum possible error in the data processing procedures. Therefore, it is taken as the lower limit. The estimation error by Eq. (2.61) is regarded as the maximum possible error—the upper limit. The error ranges and the real errors for the fitting results of the artificial R–V curves, whose noise level are 1, 2, and 3 %, respectively, are listed in Table 2.20. All the real errors drop between the upper and lower limits. This indicates that the estimation errors calculated using this method can be used to represent the accuracy of the fitting parameters.

Table 2.20 Error ranges of the six fitting parameters of the theoretical R–V curves whose noise level is 1, 2, and 3 %, respectively

Error ranges	ΔN_d (cm^{-3})	$\Delta\mu_n/\tau_n$ ($cm^2/V\ s^2$)	$\Delta\tau_0$ (ns)	$\Delta E_t/E_g$	ΔN_t (cm^{-3})	ΔR_s (Ω)
Noise 1 %	(7×10^{13})	(1×10^{13})	(0.015)	(0.002)	(2×10^{10})	(5)
	6×10^{12} ~ 9×10^{14}	1×10^{12} ~ 3×10^{13}	0.001 ~0.022	0.0004 ~0.0051	1×10^{10} ~ 2.4×10^{11}	0.2 ~5.8
Noise 2 %	(< 1×10^{13})	(4×10^{12})	(0.024)	(0.0044)	(6×10^{10})	(4)
	5×10^{12} ~ 2.6×10^{15}	9×10^{11} ~ 3×10^{13}	0.001 ~0.038	0.0003 ~0.014	9×10^{9} ~ 7.5×10^{11}	0.2 ~5.4
Noise 3 %	(9×10^{13})	(6×10^{12})	(0.023)	(0.0041)	(6×10^{9})	(5)
	3×10^{12} ~ 7.2×10^{15}	3×10^{11} ~ 3×10^{13}	0.0003 ~0.081	0.0001 ~0.034	3×10^{9} ~ 2.0×10^{12}	0.1 ~5.3

Data in brackets are the real errors). Reprinted with permission from Ref. [92] © 2006, American Institute of Physics

Table 2.21 Input materials and device parameters of the R_d–V fitting procedure for the fabricated devices A, B, and C

Sample	X	N_a (cm^{-3})	μ_p (cm^2/V s)	A (μm^2)	T (K)
A	0.233	7.69×10^{15}	1290	784	77.4
B	0.2323	9.03×10^{15}	622	784	77.3
C	0.224	8.92×10^{15}	828	784	77.0

Reprinted with permission from Ref. [92] © 2006, American Institute of Physics

To verify the applicability of the above developed fitting procedure, the R–V curves of three long wavelength devices A, B, and C with different Cd composition have been studied. The material and device parameters, which are not used as fitting parameters, are summarized in Table 2.21. Their experimental measurement process is the same as that described in Ref. [95, 96]. The fitting voltage range is −0.48 to 0.12 V. Figure 2.74 shows the fitting results. The fitting parameters and their error ranges are listed in Table 2.22.

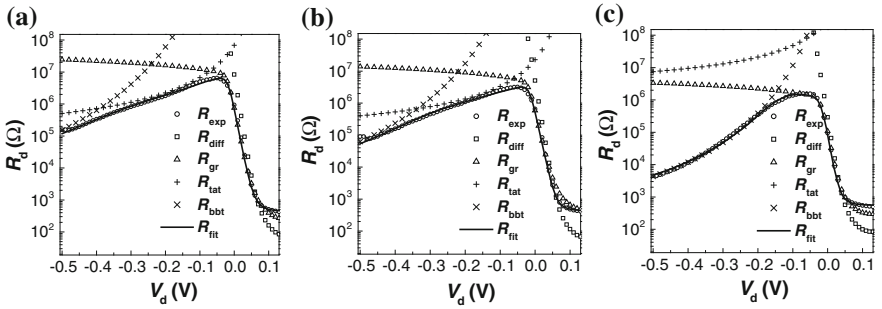


Fig. 2.74 Measured R–V curves of three real devices and their fitting results for Samples A (**a**), B (**b**), and C (**c**). Reprinted with permission from Ref. [92] © 2006, American Institute of Physics

Table 2.22 Six fitting parameters of N_d, μ_n/τ_n, τ_0, E_t/E_g, N_t, and R_s and their error ranges extracted from the measured R–V curves of the real devices A, B, and C

Sample	N_d (cm^{-3})	μ_n/τ_n (cm^2/V s^2)	τ_0 (ns)	E_t/E_g	N_t (cm^{-3})	R_s (Ω)
A	1.79×10^{16}	4.258×10^{13}	0.0796	0.4524	3.68×10^{12}	395.5
	(2×10^{13} ~ 3×10^{14})	(3×10^{12} ~ 6×10^{13})	(0.001 ~ 0.02)	(0.0007 ~ 0.006)	(4×10^{10} ~ 3×10^{11})	(1 ~ 14.8)
B	1.701×10^{16}	1.238×10^{14}	0.1248	0.5008	2.31×10^{12}	388.2
	(5×10^{12} ~ 4.5×10^{14})	(2×10^{12} ~ 1.2×10^{14})	(0.0007 ~ 0.04)	(0.0003 ~ 0.005)	(1×10^{10} ~ 2.3×10^{11})	(0.4 ~ 18.1)
C	2.109×10^{16}	1.024×10^{15}	0.0325	0.4601	8.49×10^{10}	498.7
	(3×10^{12} ~ 9×10^{14})	(2×10^{14} ~ 1.3×10^{15})	(0.003 ~ 0.005)	(0.005 ~ 0.006)	(5×10^{9} ~ 1.9×10^{11})	(2.1 ~ 41.6)

Data in brackets are the error ranges. Reprinted with permission from Ref. [92] © 2006, American Institute of Physics

In Table 2.22, compared to the value of the corresponding parameter, the lower limits of the error ranges are very small for all the samples. The maximum lower limit in Sample A is less than 10 % of the value of the corresponding parameter (μ_n/τ_n); and that of Sample B < 2 % (for N_t); and that of Sample C < 20 % (for μ_n/τ_n). These data indicate that the discreteness among the results from the ten fits in different fitting paths is very small. The upper limit of the error ranges is less than 40 % of the value of the corresponding parameter, except for the parameter N_t (in Sample C) and μ_n/τ_n (in all three samples). The upper error limits of μ_n/τ_n in all the three samples are almost equal to the value of μ_n/τ_n itself because the diffusion current mechanism, which is strongly correlated with μ_n/τ_n, fails to become dominant under forward biases. As shown in Fig. 2.74, when I_{diff} has only a small amount larger contribution to the dark current than I_{gr} after $V_d > 0.05$ V, the effect of series resistance increases and enlarges the estimation error (the upper limits of the error ranges). Likewise, the upper limits of the error ranges of N_t for Sample C is twice as much as the value of N_t, since the trap-assisted tunneling current mechanism, which is correlated with the parameter N_t, fails to become dominant under any biases. Nevertheless, one can see in Fig. 2.74 that the calculated *R–V* curves accurately fit the measured curves for all the three samples. Moreover, our fitting method can also correctly fit the specific case like Sample C, for which both the diffusion current mechanism and the trap-assisted tunneling current mechanism fail to be the dominant dark current at any bias in measured *R–V* curve. For Sample C, the parameters' multi-valued phenomena in the ten fits with different fitting paths do not exist. However, the upper limits of the error ranges are magnified. This only effects correctly estimating the accuracy of the fitting parameters. Therefore, we conclude that the fitting method developed in this work has highly fault-tolerant capability and could be expected to be an effective tool to analyze the *R–V* curves of long wavelength HgCdTe *n*-on-*p* photodiodes.

2.3.1.2 Temperature Dependence for Long Wavelength HgCdTe Photodiodes

The working temperature of a HgCdTe detector is generally around 80 K, this is mainly in order to allow the system to work in the scope of background limitations, so that the signal from background radiation is greater than the thermal noise. Background radiation signals are associated with the quantum efficiency and carrier lifetime, the thermal noise is associated with the working temperature and doping concentration of the base area. Therefore, determining the detector's carrier lifetime, carrier concentration and the working temperature limit becomes very important [97]. In addition, the nonuniformity of the temperature from the detector array can also produce different performance between different samples. These problems can be analyzed through research of the temperature dependence of the device's dark current. Analysis of the dark current characteristics under different operating temperatures is rather useful for determining the dark current mechanism. Many researchers on the analysis of HgCdTe devices have used the temperature

dependence of dark current [98–101] to obtain detailed operational information. In this chapter, we measured the variable temperature data of the dark current of HgCdTe detectors, and then used the fitting procedure that was introduced in the previous chapter to analyze the data [101].

The experiment set up 31 measurement points between 30 K and 300 K. We measured the *I–V* curve of the HgCdTe long wave detectors for each temperature in order to obtain the temperature dependence characteristics. Additionally, the experiment tests four samples on the same chip in order to ensure the repeatability of experimental results. The first step for the initial value in the analysis method is to find the initial carrier concentration of the *n* region under a large reverse bias. The carrier concentration of the *n* region is associated with a direct tunneling. Therefore, we want to measure the *I–V* curve in the bias region where the direct tunneling mechanism dominated the dark current, so that the reverse bias voltage should be appropriately large. Here, the range of measurement bias is −0.8 to 0.3 V, the bias interval is 5 mV. Sample C1, C2, C3, and C4 are tested from one chip C. The parameters of the chip C are shown in Table 2.21.

Figure 2.75a shows the temperature dependence of the R_0A product of n^+-on-p LW HgCdTe photodiodes for the four selected samples from 30 to 300 k. R_0A products of the four samples in the same chip show very good repeatability indicating the stability of the process. Figure 2.75b gives the property of the *R–V* curves for the sample C1 from 40 to 150 k. As the Fig. 2.75b shown: (1) the dynamic resistance decreases monotonously with the increasing of the temperature at the near zero-bias region; (2) the peak value of the dynamic resistance moves to the reverse bias region with increasing temperature. The peak value decreased with increasing temperature when the temperature is lower than 70 K. As the temperature is increased from 70 to 110 K, the peak value increased with increasing temperature. For temperatures above 110 K, the value again decreased with increasing temperature; (3) in the reverse bias region, the dynamic resistance increased with increasing temperature until 120 K, and then decreased.

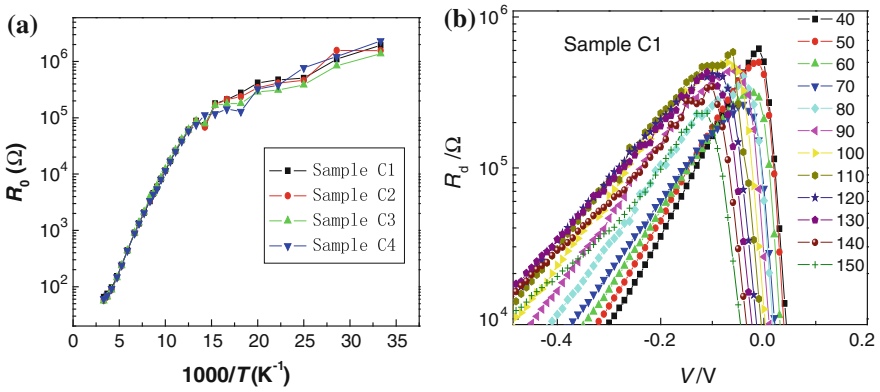


Fig. 2.75 **a** Temperature dependence of the R_0A for the four selected samples in the same chip. **b** Diode dynamic resistance versus bias for the sample C1 at different temperatures. Reprinted with permission from Ref. [101] © 2009, American Institute of Physics

To explain the experimental phenomena, we analyzed the experimental results with the fitting procedure. Figure 2.76 gives the fitting results of the *R–V* curves at some typical temperatures. As the fitting results show: (1) under a small forward

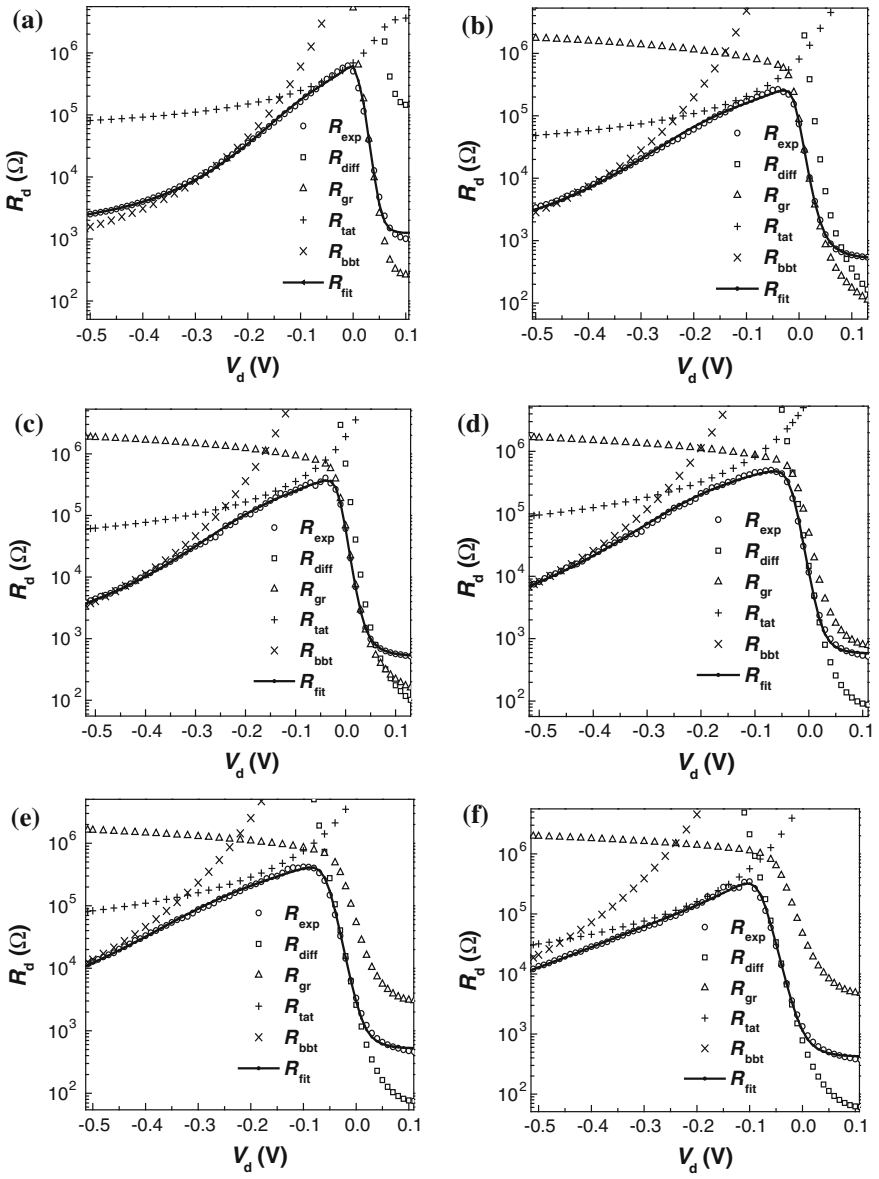


Fig. 2.76 Measured *R–V* curves and their fitted current components for the sample C1 at the temperatures of 40 K (**a**), 70 K (**b**), 80 K (**c**), 100 K (**d**), 120 K (**e**), and 140 K (**f**). Reprinted with permission from Ref. [101] © 2009, American Institute of Physics

bias, diffusion current is the dominant mechanism and the region dominated by diffusion will expand as temperature increases. For the forward bias region, I_{gr} dominates the dark current at low temperatures, while I_{diff} dominates the dark current in high temperatures. Both I_{diff} and I_{gr}, which linearly increase with the increasing temperature, are the thermal current components. This result explains the first phenomenon observed in the experiment. (2) The peak value of the *R–V* curves is decided by the TAT mechanism in low temperatures. As the temperature is increased to 70 K, the I_{gr} begins to jointly dominate the peak value. For temperatures above 100 K, the I_{gr} completely becomes the dominant mechanism. However, as the temperature rises further, I_{diff} dominates the peak value to replace the I_{gr}. When the temperature is increased to 140 K, the TAT mechanism again dominates the peak value of the *R–V* curves. In this process, RTAT decreased, then increased, and then decreased with the increasing of the temperatures. The resistance of the thermal mechanism decreased and moved to the left, so the peak value of the *R–V* curve tends to move to the left with the increasing temperature, it corresponds to the second point of experimental phenomenon (3). In the reverse bias region, the resistance of the TAT mechanism increased with increasing temperatures at low temperature, and was unchanged when the temperature is larger than 100 K. As the temperature increases, the region which is dominated by the TAT mechanism is moved to the large reverse bias side. So with increasing temperature, the dynamic resistance increased, but then the resistance began to decline when the temperature was higher than 120 K.

In the process of fitting, we also obtained a set of fitting parameters corresponding to each temperature. Figure 2.77 gives the characteristic parameters as function of the temperature for the four samples. It is important to note that the concentration of holes in the sample is measured from the Hall effects measurements at liquid nitrogen temperature, and the frozen effect [102] of acceptors in p-type HgCdTe material must be taken into account in the low temperature. Thus, according to the research results from Scott [103], shallow acceptor ionization energy and hole concentration at the temperature of 77 k have a simple relationship:

$$E_A = E_0 - \alpha P_0^{1/3} \tag{2.62}$$

where E_0 = 17 meV, α = 3 × 10-8 eV·cm, and P_0 is the hole concentration in the temperature of 77 k.

Figure 2.77a shows the dopant density in the *n* region (N_d) as a function of the temperature. Note that N_d is not monotonously increased with increasing temperature, but decreased with increasing temperature less than 60 K and increased with increasing temperature at a temperature lager than 60 k. These phenomena can be explained as follows. The ion implantation damaged n-type donors are completely ionized. Figure 2.77b shows the electron lifetime in the *p* region (τ_n) as a function of the temperature. Because fitting procedures can only fit the ratio of electron mobility to lifetime (μ_n/τ_n), we must use an empirical formula to calculate the electron mobility μ_n, then obtain τ_n. Figure 2.77c shows the temperature

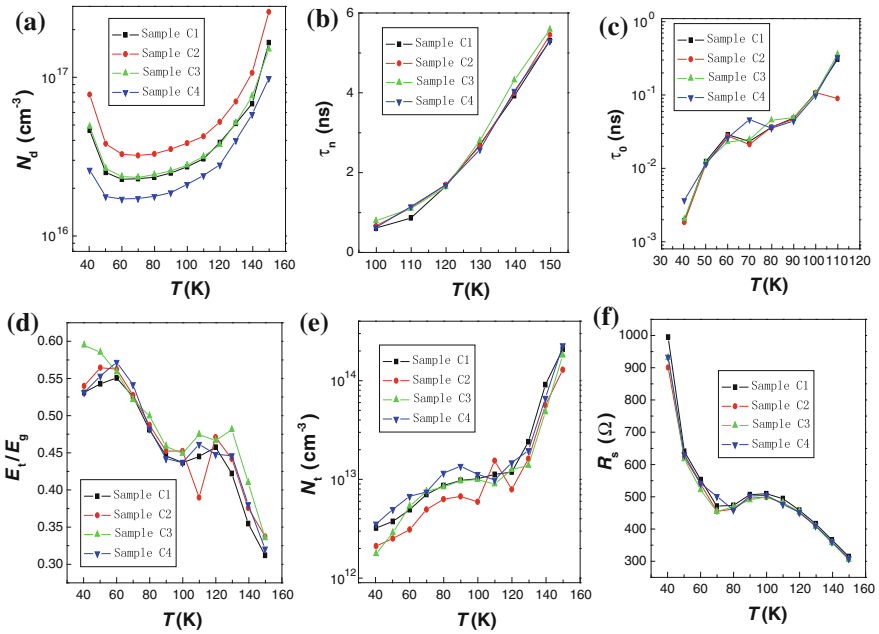


Fig. 2.77 Temperature dependence of the fitting parameters for the four samples. **a** the dopant density in the *n* region (N_d), **b** electron lifetime in the *n* region (τ_n), **c** effective lifetime in the depletion region (τ_o), **d** relative energy position of trap level (E_t/E_g) and (e) trap density (N_t) in the depletion region, and **f** series resistance (R_s). Reprinted with permission from Ref. [101] © 2009, American Institute of Physics

dependence of effective lifetime in the depletion region (τ_0). As the g-r current dominates the dark current for the forward bias at the low temperature, the fitting results of τ_n below 100 K, where the fitting error of τ_n is very large, are not shown in Fig. 2.77b. Similarly, the diffusion current dominates the dark current for forward bias at high temperature, magnifying the fitting error of τ_0. Therefore, the fitting results of τ_0 above 110 K are not shown in Fig. 2.77c. As the two figures show, minority carrier lifetime increased with increasing temperature, the qualitative temperature dependency of τ_n is in agreement with the reported in literatures [104–105].

Figure 2.77b shows the temperature dependence of τ_n. The value of τ_n is comparable to that reported in [106], about 3–150 ns. Note that the fitting value of the effective lifetime in the depletion region (in Fig. 2.77c) is smaller than that of normal one. We attribute parts of the deviation to the screening effect reduction and resonance scattering enhancement from deep level trap and charged trap in the depletion region in the high voltage field.

Figure 2.77d shows the relative energy position of the trap level (E_t/E_g) as function of temperature. From the figure, there are two stable stages: $0.45E_g$ for 90–130 K and $0.55E_g$ for temperature lower than 60 K. Therefore, we can consider that

HgCdTes have two stable characteristic trap levels. The trap level of $0.55E_g$ dominates the TAT mechanism when the temperature is lower than 60 k. When the temperature is increased to 60 K, the $0.45E_g$ begin to jointly dominate the TAT mechanism. For temperature above 90 K, the $0.45E_g$ completely becomes the dominant trap. Thus, we can also speculate that there is a trap level in $0.25E_g$ at high temperature. Figure 2.77e shows the temperature dependence of the trap density (E_t), trap concentration increased exponentially with the increasing of the temperature. Figure 2.77f shows the temperature dependence of the series resistance (R_s), the series resistance decreased with increasing temperature.

2.3.1.3 Statistical Analysis of Long Wave HgCdTe Photodiodes

At present, the processing of HgCdTe material and devices is much less mature than that of Si and GaAs due to the sensitive nature of HgCdTe material. Therefore, HgCdTe IRFPAs still suffer severely from temperature and photo related defects and nonuniformities in the performance of individual elements. In order to thoroughly understand the performance of HgCdTe material, a feasible method is to obtain the basic parameters of HgCdTe material from statistical analysis. However, the characteristics research of a single device can't meet the needs of such statistical requirement. For example: (1) the physical parameters obtained by the analysis of *I–V* characteristics from one unit cannot be applied to the whole chip; (2) the physical parameters obtained by the analysis of single chip cannot be applied to another chip even under the same process technology. An efficient approach is to study the *I–V* characteristics of many units between different chips under the same process technology. Then, the statistical results of basic physical parameters can be extracted by the proposed physical model. These statistical results of HgCdTe devices associated with the process condition have a great significance for the design and optimization for HgCdTe devices, especially for the next generation of focal plane devices such as HgCdTe two-color detectors.

In order to obtain the statistical parameters of a long wave HgCdTe device, the *R–V* curves of 392 HgCdTe diodes are analyzed and fitted. The Cd components of these devices are different from 0.223 to 0.238, which are tested at liquid nitrogen temperature (77.0–80.9 K). According to the fitting results, the diffuse current cannt be the dominant current all the time at liquid nitrogen temperature so that its contribution to the dark current can be ignored. Therefore, the fitting parameter μ_n/τ_n related to the diffuse current is unbelievable, and the size of μ_n/τ_n is not listed in this section.

The statistical range of each parameter obtained by fitting method is as follows: the dopant density in the n-type region is varied from 3×10^{15} to 3×10^{16} cm^{-3}; the effective lifetime (τ_0) of the depletion region is varied from 0.01 to 0.2 ns; the energy level (E_t) of the trap is varied from $0.4Eg$ to $0.6Eg$; the concentration of the trap (N_t) is varied from 1×10^{12} to 1×10^{13} cm^{-3}; the size of series resistance (R_s) is about 500 Ω. τ_0 is smaller than the lifetime of the neutral region due to the enhanced scattering from trap in high field. The other parameters are consistent with

the typical experiment results, which demonstrate the accuracy of the statistical method.

In order to obtain the distribution of parameters associated with the different process technologies, the performance parameters of HgCdTe devices manufactured by MBE and LPE process is studied for comparison. Figures 2.78, 2.79, 2.80, 2.81 and 2.82 show the distribution of the parameters from these two technologies (Black lines represent the MBE device, gray lines represent the LPE device, R_I is the responsivity of current, D^* is the detectivity, R_0 is the average value of the current responsivity, and D_0^* is the average value of detectivity). In the figure, the responsivity and detectivity of current are the experimental measurement data. The results show that the parameters obtained from MBE device are close to that of LPE device.

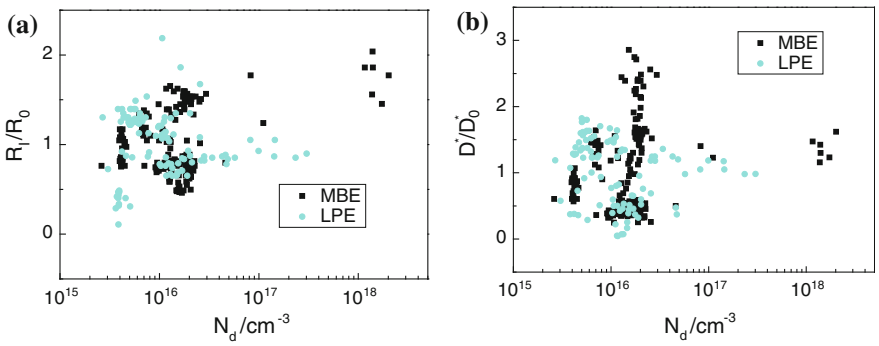


Fig. 2.78 **a** The statistical distribution of the responsivity with the dopant density in n-type region. **b** The statistical distribution of the detectivity with the dopant density in n-type region

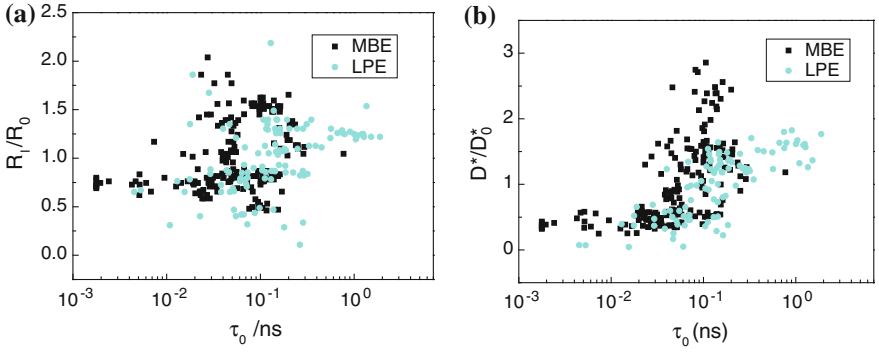


Fig. 2.79 **a** The statistical distribution of the responsivity with the effective lifetime of the depletion region. **b** The statistical distribution of the detectivity with the effective lifetime of the depletion region

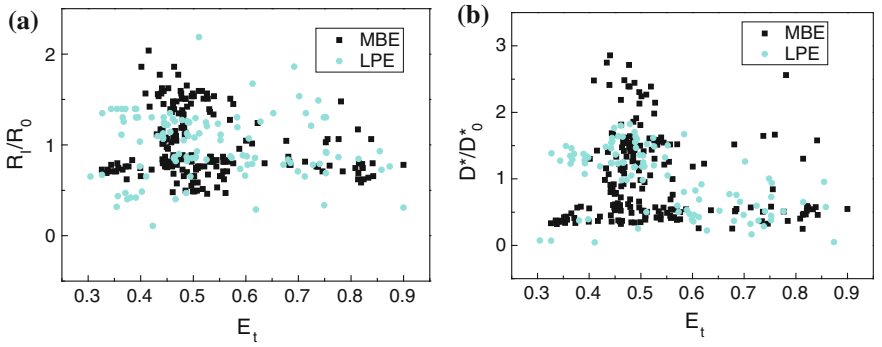


Fig. 2.80 a The statistical distribution of the responsivity with the relative energy level of the trap. **b** The statistical distribution of the detectivity with the relative energy level of the trap

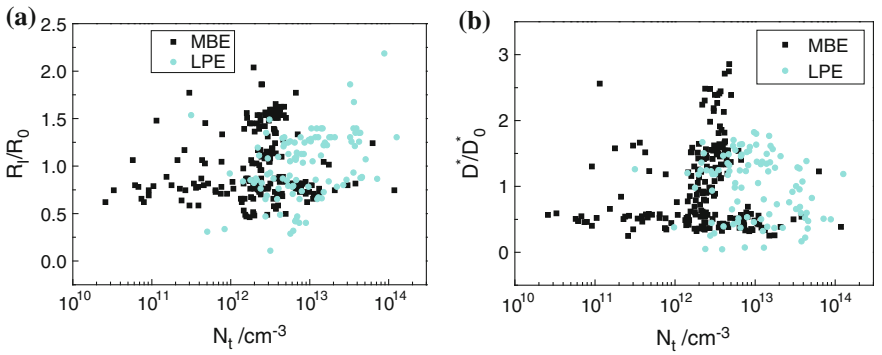


Fig. 2.81 **a** The statistical distribution of the responsivity with the concentration of the trap. **b** The statistical distribution of the detectivity with the concentration of the trap

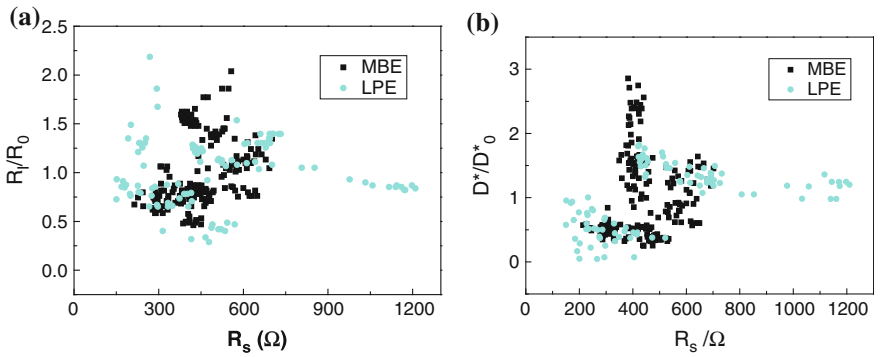


Fig. 2.82 **a** The statistical distribution of the responsivity with the series resistance. **b** The statistical distribution of the detectivity with the series resistance

2.3.1.4 Statistical Analysis of Mid-Wavelength HgCdTe Photodiodes

In order to obtain statistical parameters of medium wave HgCdTe device, the *R*–*V* curves of 206 HgCdTe diodes are analyzed and fitted. The Cd components of these devices are different from 0.2927 to 0.314, which are tested at liquid nitrogen temperature (77.0–78.3 K). The parameters of HgCdTe devices manufactured by MBE process are studied. According to the fitting results, the diffuse current cannot be the dominant current all the time at liquid nitrogen temperature so that its contribution to the dark current can be ignored. In addition, the direct tunneling cannot be the dominant current in a large reverse bias region all the time at liquid nitrogen temperature so that its contribution to the dark current also can be ignored. Therefore, the fitting parameter μ_n/τ_n and N_d related to the diffuse current and the direct tunneling current respectively are unreasonable, and the values of μ_n/τ_n and N_d are not listed in this section.

The statistical range of each parameter obtained by the by fitting method is as follows: the effective lifetime (τ_0) of the depletion region is varied from 0.01 to 1 ns; the energy level (E_t) of the trap is varied from 0.45*E*g to 0.7*E*g; the concentration of the trap (N_t) is varied from 1×10^{12} to 1×10^{15} cm^{-3}; the size of series resistance (R_s) is about 1000 Ω. The parallel resistance (*R*Sh) is varied from 108 to 109 Ω. τ_0 is smaller than the lifetime of the neutral region due to the enhanced scattering from trap in high field. The other parameters are consistent with the typical experiment results, which demonstrate the accuracy of the statistical method.

2.3.2 Extracting Device Parameters by Photoelectric Method

2.3.2.1 Photon-Generated Minority Carrier Lifetime

Minority carrier lifetime is one of the basic parameters of performance for semiconductor devices. For a conventional semiconductor, measuring the lifetime of minority carriers has become a routine technology. Even though HgCdTe has been studied extensively for infrared detectors, there is still a great deal of ambiguity in some issues, such as the minority carrier lifetime and its dominating recombination mechanisms. This is because of the instability of HgCdTe, in which the material property may be changed during the formation process of *pn* junction. Therefore, the parameter of the raw material can not be applied to estimate the properties of *pn* junction devices. Moreover, there are great differences between the actual parameters and the design parameters such as trap concentration, carrier concentration, the junction depth, the junction width in conventional techniques. These factors have a lot of unpredictable effects on the minority carrier lifetime in a *pn* junction device. In order to determine the minority carrier lifetime and the dominating recombination mechanisms of electron in a HgCdTe photodiode, measurements must be

carried out on the actual devices then the extracted parameters can be applied in devices design and simulate.

Many measurements have been developed to determine the minority carrier lifetime for other material devices such as: short-circuit current, open-circuit voltage decay (OCVD), pulse recovery technique, etc. However, the minority carrier lifetimes of HgCdTe material are in the nanosecond range, and these methods are not suitable to measure such short lifetime. In this chapter, we measure the minority carrier lifetime using an improved photo-induced OCVD measurement technique which compensates the effects of the junction equivalent capacitor and the trap center on the measurements. In order to minimize the effect of the junction equivalent capacitor and series resistance on the measurements, we used an Oriel QTH lamp as the steady-state bias light source. By recording the OCVD and fitting to the exponential decay curve, the minority carrier lifetime can be extracted.

The incident pulse laser having wavelength tuning range 2.3–10 μm was provided by a commercial optical parametric oscillator and a difference frequency generator which were pumped by a picosecond Nd: YAG laser. The structure of the machine is shown in Fig. 2.83. The laser delivered a pulse of 30 ps in duration at a frequency 10 Hz. Therefore, the influence of the laser pulse on the falling time was avoided.

As shown in Fig. 2.84, all HgCdTe samples were grown by MBE on GaAs substrates with CdTe buffer layers and an abrupt n^+-on-p structure were formed by

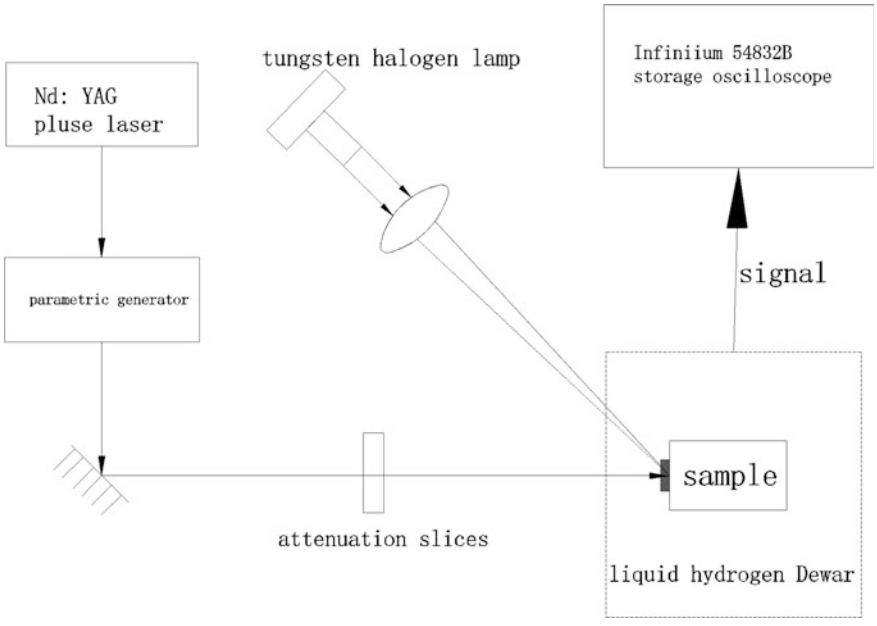


Fig. 2.83 The diagrammatic map of the experimental facility for measuring the lifetime of HgCdTe *p-n* junction use pulse laser open-circuit voltage decayed method

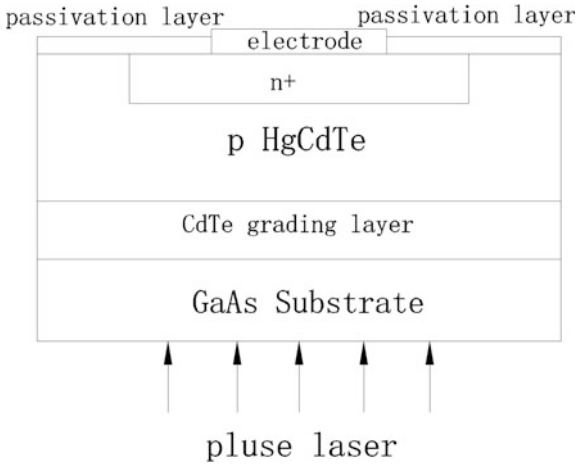


Fig. 2.84 Schematic of HgCdTe diode in the experiment

the ion implantation of B^+ in p-type $Hg_{1-x}Cd_xTe$. As ZnS films were formed on the $Hg_{1-x}Cd_xTe$ surface for passivation, the measured lifetime values were not influenced by the surface treatment. The detectors were processed into 50 × 50 μm^2 area mesa structures. The photogenerated voltage has been recorded by a storage oscilloscope.

In the experiments, according to the intensity of the injected carrier, the decay curve of photovoltage in the *p-n* junction can be divided into three different areas: strong injection, medium injection, and the low injection. If the bias light source is strong enough, the influence of n^+p junction and impurity deep energy level to the photovoltage decay behavior can be ignored. Then the lifetime can be determined by analyzing the result of the photovoltage decay curve. The lifetime of minority carriers in strong injection is given by: $\tau = \frac{2kT}{q}\left|\frac{1}{dV_{oc}/dt}\right|$. The concentration of excess minority carriers in p region is much higher than the equilibrium minority carrier but less than the equilibrium majority carrier at the condition of medium injection. The lifetime of minority carrier in medium injection is given by: $\tau = \frac{kT}{q}\left|\frac{1}{dV_{oc}/dt}\right|$. Here, k is the Boltzmann constant, T is the absolute temperature, q is the electron charge, t is the time. We can see, in strong and medium injection, the photovoltage (V_{oc}) linearly decays. The concentration of excess minority carriers in the p region is less than the equilibrium minority carrier in low rejection, and the curve of the photovoltage exponentially decays.

$$V_{oc} = \frac{kT}{q}\left[\exp\left(\frac{qV(0)}{kT}\right) - 1\right]\exp(-\frac{t}{\tau}) \tag{2.63}$$

where $V(0)$ is the open-circuit voltage when the light stops. In order to make the measured minority carriers reflect the actual working state of the device, we must take the non-equilibrium carriers in a state of low injection.

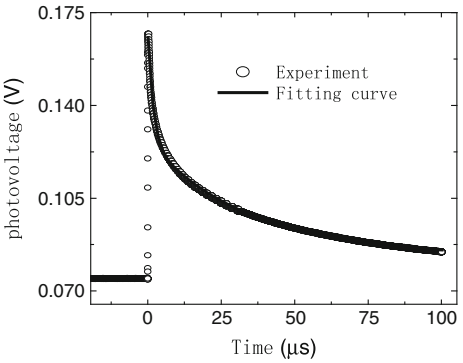


Fig. 2.85 Pulsed photo-response profiles from photodiode illuminated with laser pulses at zero bias light intensity. Reproduced from Ref. [109] with kind permission from Springer Science and Business Media

When the HgCdTe photodiode is excited by the pulsed laser, the photo-response shows a rapid increase and slow decay process. Depending on the intensity of the excitation source, the device is in a low injection condition. Figure 2.85 gives the pulsed photoresponse profiles from a photodiode illuminated with laser pulses at zero bias light intensity. The curve is an exponential decay to visual inspection. However, when the decay tail of the photoresponse was analyzed, we were not able to fit the measurement data with a suitable first-order exponential decay. Instead, a good fit is realized with a second-order exponential decay function with two different time constants: $V_{oc} = A\exp(-\frac{t}{\tau_1}) + B\exp(-\frac{t}{\tau_2})$. Here, $\tau_1 \sim 2$ μs and $\tau_2 \sim 35$ μs. Neglecting the detailed photo-generation and recombination mechanisms, we describe the decay curve profiles as dominated by the *RC* time constant and trap energy level effects on excess carrier's relaxation. Therefore, neither of the two time constants is the lifetime of minority carriers.

Figure 2.86 gives the OCVD curves at different intensities of bias light in the condition of the photodiode illuminated with a background. The steady-state photovoltage of the photodiode increases with the rising of the bias incident intensity. The decay time constants of photovoltaic response induced by the pulsed

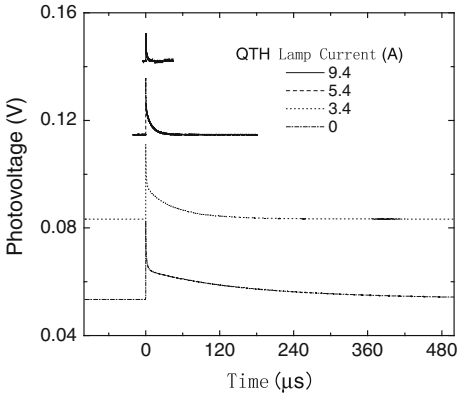


Fig. 2.86 OCVD curves at different intensity of bias light. Reproduced from Ref. [109] with kind permission from Springer Science and Business Media

laser are becoming shorter and shorter and the peak amplitude decreases with the increase of incident intensity, which is given in Fig. 2.86. This phenomenon can be attributed to the compensating of the junction equivalent capacitor and the trap center energy level effects under the bias light condition.

When the steady-state photovoltages do not increase with an increase in the background intensity, the photogenerated carriers recombination will dominate the decay time constant which is related to the minority carrier lifetime. Since the values of the resistance and the carrier lifetime are much larger in the p region than in the n^+ region. In addition, the photogenerated carriers in the emitter are about one percent of the carriers generated in the base; therefore, we can assume that the carriers stored in the base play a dominant role in the OCVD process. The compensated photoexcited OCVD decay curve is fitted with the exponential decay function, and the lifetime magnitude of the minority carrier in p region is determined to be 190 ns, which is shown in Fig. 2.87.

Using the method mentioned above, the minority carrier lifetime of the HgCdTe photodiode with different compositions can be obtained and is shown in Fig. 2.88. The results show that the carrier lifetimes are in the range of 18–407 ns at 77 K for the measured detectors of four compositions. With increasing composition, the minority carrier lifetimes have an increasing tendency, and the lifetime of the shortest wavelength infrared detectors are the longest compared to other detectors. The results of the same composition came from different units of one array. The basis of distinction between the different units within the same array is that the HgCdTe raw material is non-uniform or the growing process can not be mastered. The electron lifetimes extracted from the experiments are reasonable because of the lifetime magnitude consist with others results. Generally, there is a certain difficulty to measuring minority carrier lifetime precisely. Even for the silicon material, the accuracy scope of the minority carrier lifetime is ±135 % in different laboratory in American Society for Testing and Materials (ASTM). Therefore, it is acceptable that there is some difference in the lifetime experiment measuring of HgCdTe photodiode.

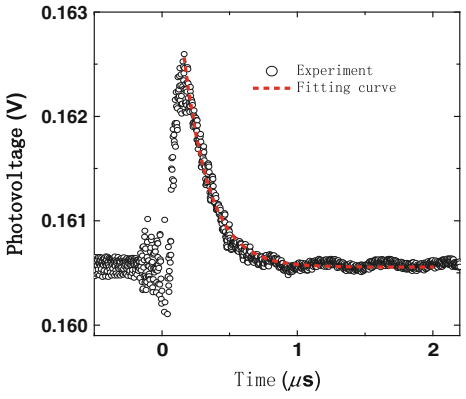


Fig. 2.87 Pulsed photo-response profiles from photodiode illuminated with laser pulses at suit bias light intensity. Reproduced from Ref. [109] with kind permission from Springer Science and Business Media

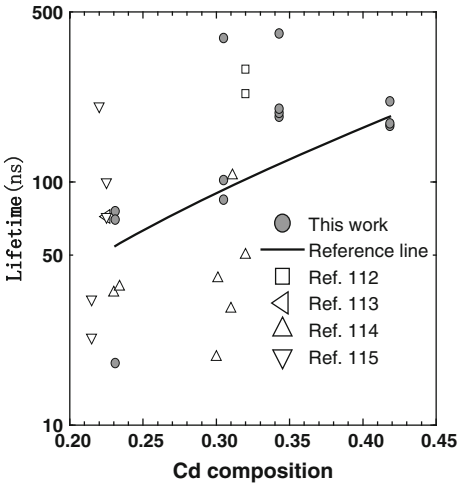


Fig. 2.88 Variation of the extra carrier lifetime with Cd composition [106–109], the line is for guiding. Reproduced from Ref. [109] with kind permission from Springer Science and Business Media

2.3.2.2 Laser Beam-Induced Current Microscopy for HgCdTe Photodiodes

Laser beam-induced current (LBIC) microscopy has proven useful for quick and non-destructive measurement of the junction performance of each pixel in large arrays at an early stage in the fabrication process. Therefore, significant cost savings and processing improvement would be expected in FPAs. Two-dimensional LBIC microscopy, also generally called photocurrent mapping (PC mapping), can provide spatially resolved information about local electrical properties and *p-n* junction formation in photovoltaic infrared photodetectors from which it is possible to extract material and device parameters such as junction area, junction depth, diffusion length, leakage current position and minority carrier diffusion length, etc. To date, LBIC as a nondestructive method has been widely used for infrared photodiode array characterization. The big challenge is to quantify the relationship between LBIC mapping and parameters that influence the electrical performance of devices. The main issue of this section is the application with respect to extracting material and device parameters [110].

(1) **LBIC setup**

In preparation for the LBIC experiment, only two shorted Ohmic contacts would be constructed at remote positions on either side of the device(s). The principle of *p-n* junction array LBIC testing is schematically represented in Fig. 2.89. He–Ne laser irradiance at a wavelength of 632.8 nm is focused onto the surface of the semiconductor, and stepped incrementally across the sample in the horizontal direction. When the laser spot is more than a few carrier diffusion lengths from a built-in electric field, the photogenerated electron–hole pairs recombine without reaching a junction region and therefore no current signal will be induced. In contrast, if the

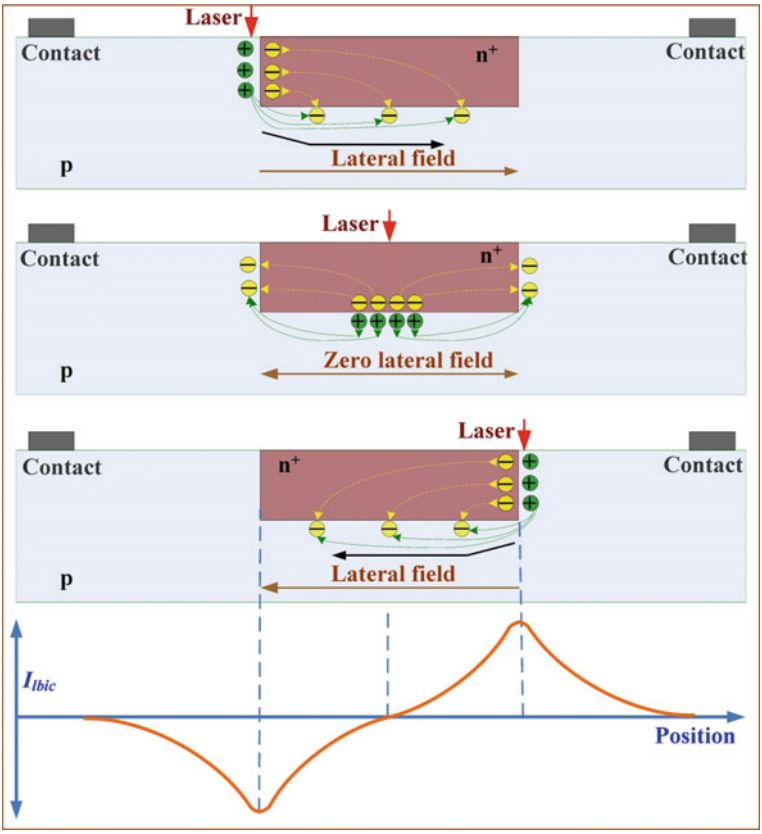


Fig. 2.89 The principle of a single n^+-on-p pixel LBIC testing. Reprinted with permission from Ref. [111] © 2014, American Institute of Physics

photogenerated electron–hole pairs can diffuse and be separated by the built-in electric field, the separated carriers seek to recombine via returning current paths, which could be both the internal and external closed circuits; resulting in an induced current that can be measured.

The LBIC profile presents a positive and negative bimodal distribution according to the three typical positions of a laser spot on the surface of the sample. This is explained by assuming the direction from right to left of device as being the direction of negative current flow.

Case 1: Illumination on the left of p-type region close to the junction

If the laser is focused onto the surface of the p-type region close to the junction, the photogenerated carriers that are separated by the built-in electric field would provide excess electrons (denoted by "-") into the n-type region while excess holes (denoted by "+") remain in the p-type region. Since the conductivity of the n-type

region is nearly two orders of magnitude more than that in the p-type region, the excess electrons can quickly redistribute themselves over the region uniformly. A reinjection of electrons can occur in the underlying p-type region. Finally, the presence of these excess holes and electrons constitutes a localized forward bias of the junction (direction from left to right), that drives the excess carriers to seek any possible returning current paths for recombination. Among these paths, the returning current via the external circuit contribute to the LBIC's negative signal.

Case 2: Illumination on the right of p-type region close to the junction

This situation is similar to the previous analysis; the photogenerated carriers diffuse and are separated by the built-in electric field, resulting in excess electrons passing into the n-type region while excess holes pass into the p-type region. However, a localized forward bias of the junction in the opposite direction (from right to left) is formed. Therefore, a positive LBIC signal can be obtained.

Case 3: Illumination on the surface of n-type region

A unique feature in the situation where the laser is focused onto the surface of the n-type region was the appearance of two types of returning current flows in the opposite direction. If the laser spot is exactly located in the middle of n-type region, two opposite return current flows will cancel each other, producing a zero net LBIC signal.

Two different structures of the laser beam induced current (LBIC) test system are shown in Fig. 2.90. The test system typically consists of a laser light source, CCD camera, computer, temperature controller, and an induced current measuring system with a lock-in amplifier. The sample is placed in a temperature controlled Dewar of liquid nitrogen. The He–Ne laser is focused to a spot of 1–1.5 μm in diameter which is stepped across the sample. The induced current is recorded by a SR830 DSP lock-in amplifier as a function of *x*-*y* scanning coordinates to provide a spatial LBIC map. The main difference between the two LBIC test systems is the method used to scan the samples. In Fig. 2.90a, the sample is scanned on a two-dimensional computer controlled mobile platform. Another method is the use of a two-dimensional scanning galvo system to precisely control the direction of the laser beam by the computer, as shown in Fig. 2.90b.

(2) **Junction depth and length extraction**

The temperature dependence of the peak-to-peak LBIC measurements on a *p*-*n* junction was first examined by Redfern et al. [112], indicating that a saturation state can be reached under low temperatures. Under the saturation conditions, the junction resistance may be enough to dominate the resistance of each of the possible current paths. Therefore, the peak-to-peak magnitude of LBIC is independent of the bulk material properties. Other parameters, including the junction geometry, can be examined by the LBIC peak-to-peak magnitude, greatly reducing the complexity of the analysis. Simulated LBIC peak-to-peak magnitudes as a function of temperature in HgCdTe photodiodes with different components are shown in Fig. 2.91 [113].

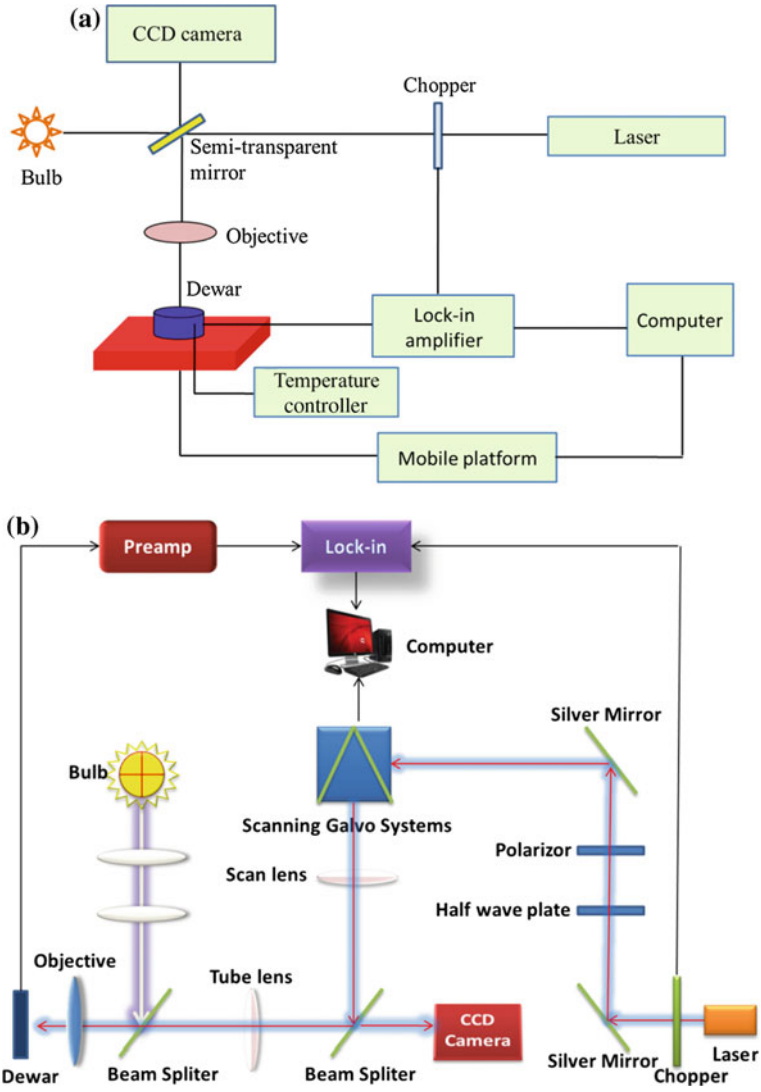


Fig. 2.90 Two different structure frames of laser beam induced current (LBIC) test system. **a** Scanning is performed by changing position of sample under control of a two-dimensional mobile platform. **b** Scanning is performed by changing the direction of the laser beam under control of a scanning galvo system

The condition of threshold temperature for $Hg_{0.69}Cd_{0.31}Te$ is below 200 K. The effects of junction depth and length on LBIC profiles in $Hg_{0.69}Cd_{0.31}Te$ photodiodes at 170 K below the threshold temperature are shown in Fig. 2.92. It has been found that there is a linear relationship between the magnitude of the LBIC peak and the value of the junction depth and length. For photodiodes with greater depth,

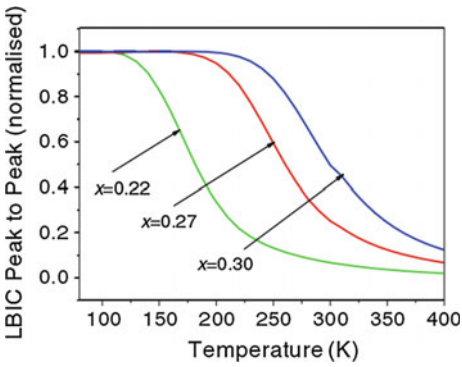


Fig. 2.91 Simulated LBIC peak-to-peak magnitudes as a function of temperature in $Hg_{1-x}Cd_xTe$ photodiodes. Reprinted from Ref. [113], with kind permission from Springer Science+Business Media

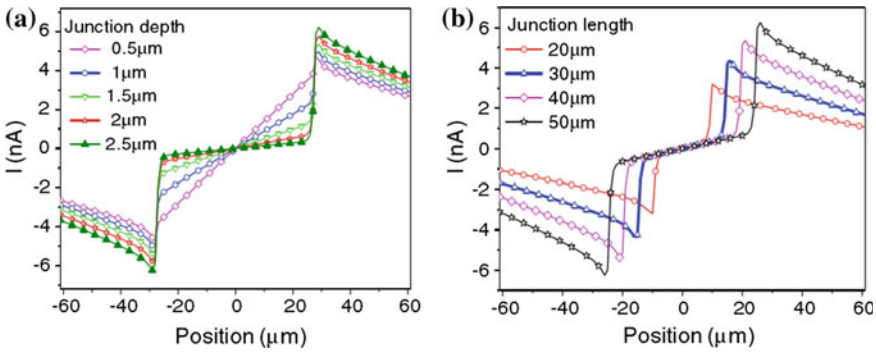


Fig. 2.92 **a** LBIC profiles with different depths of $Hg_{0.69}Cd_{0.31}Te$ p-n junction and **b** LBIC profiles with different lengths of $Hg_{0.69}Cd_{0.31}Te$ *p-n* junction. Reprinted from Ref. [113], with kind permission from Springer Science+Business Media

more photogenerated electrons in the p-type region can be swept across the junction, which results in a stronger LBIC signal. On the other hand, loop circuit resistance is reduced by shortening of the distance between the contact and junction in the photodiodes with longer length. Therefore, the LBIC current also becomes larger when the junction length becomes larger [113].

(3) **Minority carrier diffusion length extraction**

The minority carrier diffusion length is a key indicator of material quality and device performance. The dependence of laser beam induced current (LBIC) on minority carrier diffusion length of n-on-p HgCdTe photodiode has been investigated earlier [114, 115]. In the standard diffusion length (L_p) test method, the procedure may bring about damage to the *p-n* junction. The test structure is also

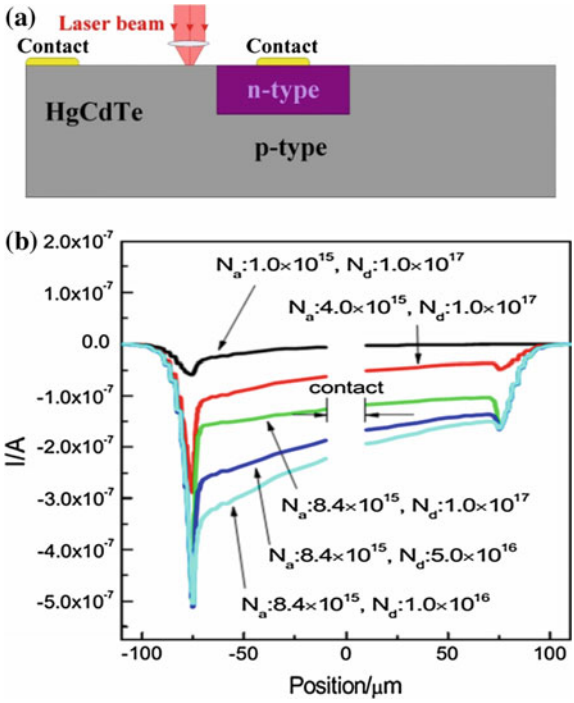


Fig. 2.93 **a** The standard diffusion length (L_p) test structure **b** the standard diffusion length (L_p) test in HgCdTe photodiodes with different doping concentrations. Reprinted with permission from Ref. [115] © 2009, Chinese Physical Society

difficult to fabricate because of the need for electrical contacts on the *p-n* junction unit as shown in Fig. 2.93a. In contrast, the LBIC test structure consists of only two Ohmic contacts at remote positions on either side of the device(s). The decay of the LBIC as the laser spot is scanned away from the edge of *p-n* junction is related to the diffusion length in p-type region. The characteristic diffusion length (L) can be obtained by fitting a simple exponential function to the LBIC curve. The exponential formula for the attenuation curve in p-type region is given by [116]:

$$|I_{\mathrm{LBIC}}(d)| = k \cdot \mathrm{e}^{-\frac{d}{L}} \tag{2.64}$$

where k is the proportional coefficient, d is the distance from the laser spot location to the boundary of *p-n* junction, and L is the characteristic diffusion length.

In two characteristic ways of diffusion length, the attenuation curves in p-type region are both related to the minority carrier diffusion length. When the laser spot is near the carrier diffusion length from a built-in electric field, the photogenerated electron–hole pairs can diffuse and be separated by the built-in field. The only difference is that the separated carriers recombine via return current paths, i.e., the internal and external closed circuits (two electrodes are both in *p* region). A LBIC signal can be measured in an external closed circuit of LBIC in this phenomenon. However, the separated carriers are directly collected by the electrodes in *p* region and n region of the standard protocol. Therefore, there may be a negligible

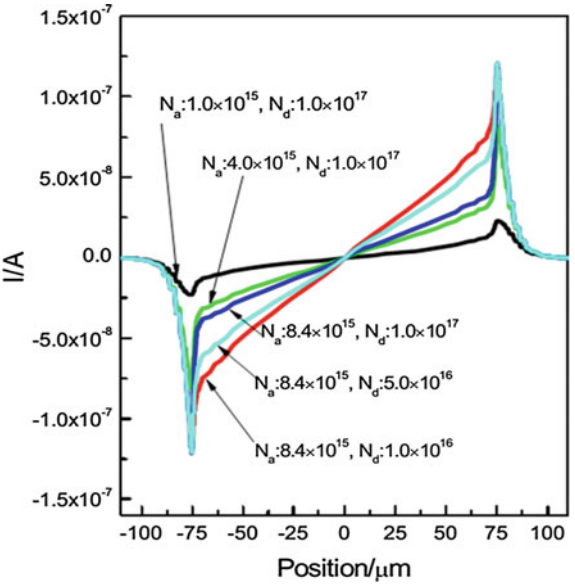


Fig. 2.94 The LBIC test for characteristic diffusion length (L) in HgCdTe photodiodes with different doping concentrations. Reprinted with permission from Ref. [115] © 2009, Chinese Physical Society

difference between the size of L and L_p. One can extract the ratio of L/L_p from the experimental results using two different methods.

The standard diffusion length (L_p) and the characteristic diffusion length (L) tested by LBIC in HgCdTe photodiodes with different doping concentrations are shown in Fig. 2.93 and Fig. 2.94, respectively [115].

The diffusion lengths extracted by both these methods are listed in Table 2.23 [115]. It is found that the L/L_p ratio is close to 1 and is independent of the doping concentration distribution. At the same time, the ratio of L/L_p remains the same, irrespective of the size of SRH carrier lifetime and mobility. Therefore, the characteristic diffusion length by LBIC can be considered equivalent with the standard diffusion length.

Table 2.23 Extracted L/L_p under different N_a and N_b

$N_a/10^{15}$ cm^{-3}	$N_d/10^{17}$ cm^{-3}	Characteristic diffusion length L (μm)	Standard diffusion length L_p (μm)	L/L_p
1.0	1.0	8.34	7.81	1.07
4.0	1.0	6.09	5.73	1.06
8.4	1.0	5.34	4.97	1.07
8.4	5.0	5.41	5.02	1.08
8.4	1.0	5.40	5.06	1.07

Reprinted with permission from Ref. [115] © 2009, Chinese Physical Society

(4) Localized junction leakage characterization

The effect of localized defects is one of the main limitations in the performance of HgCdTe infrared focal plane arrays. Such defects, including voids; line dislocations; and triangles, may influence the overall integrity of the *p-n* junction and significantly degrade the performance of the photodiodes due to the localized junction leakage. LBIC, as an efficient and nondestructive tool, is used for the localized junction leakage characterization in the photodiodes. Redfern et al. [117, 118] studied the LBIC profiles when a small localized leakage region was introduced at various positions along the horizontal portion of the junction. The leakage current was simulated by including a small piece of metal that was Ohmic to both sides of the junction [119]. The model structure of the *p-n* junction with localized junction leakage path is shown in Fig. 2.95.

It was shown that when the leakage point is asymmetric inside the device, an asymmetry in the LBIC line profile induced by the small metallic region can be obtained by this leakage model. However, the assumption of bringing in metal with large conductance in the HgCdTe photodiodes is not very practical.

On the basis of this data, an improved leakage model including a small HgCdTe region with extremely short carrier lifetime instead of the metallic region is proposed [120]. Many factors, such as trap-assisted tunneling, generation–recombination (g-r) and diffusion current, can influence the junction leakage current. Dark current dominant mechanisms change with the operating temperature. The temperature dependence of the LBIC profiles is shown in Fig. 2.96.

The asymmetric LBIC profile can indicate that there is localized leakage somewhere inside the sample structure. The localized defects reduce the resistance of the return current path crossing the junction via localized leakage, and most current flows through the leakage current path and fails to contribute to the external circuit current. In addition, the temperature dependence of symmetry in the LBIC profile reflects the change in the dominating mechanism of dark current from a certain degree.

It can be observed that the LBIC profiles have different distributions for different temperatures. Below 170 K, the diffusion and generation–recombination dark currents are very small. However, with increasing the temperature, the deep level traps are activated that lead to a lower SRH lifetime. The LBIC profile tends to be more asymmetric with increasing temperature. When temperature is above 170 K, the diffusion current component becomes dominant. Then, the leakage current is relatively reduced with the diffusion current being dominant at higher temperatures.

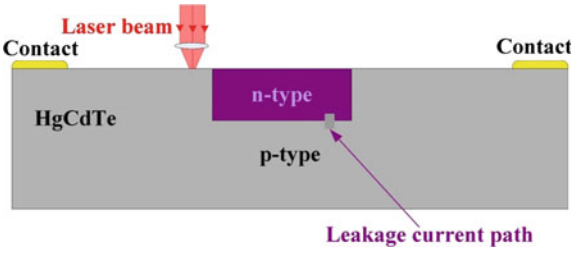


Fig. 2.95 The model structure of the *p-n* junction with localized junction leakage path along the horizontal portion of the junction

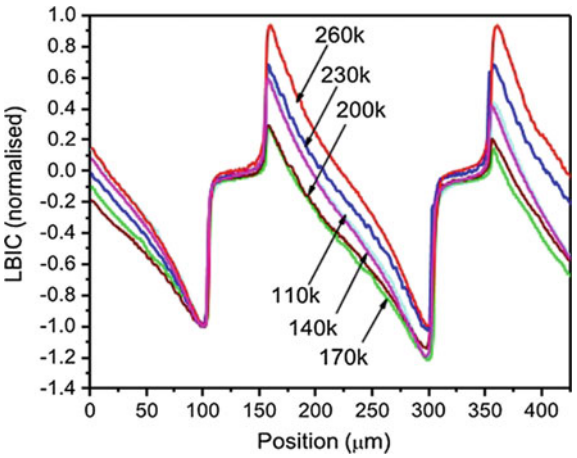


Fig. 2.96 The LBIC measurements are taken at a range of temperatures from 110 to 260 K. Reprinted with permission from Ref. [120] © 2013, American Institute of Physics

The LBIC profile becomes more symmetric with increasing temperature above 170 K. However, when the localized leakage point is situated at the center in the device, it is very difficult to confirm whether there is leakage in the diode because the distribution of LBIC profile is symmetrical. Therefore, in most situations, the asymmetric LBIC profile may indicate that there is localized leakage somewhere inside the sample structure.

(5) **Electrically active defects-related junction transformation**

Recent study [111, 121–125] shows that the phenomenon of extended defects induced by semiconductor manufacturing process always exists in HgCdTe infrared arrays, such as in B^+ ion implantation and pulsed laser drilling for the formation of a *p-n* junction. Those defects may be sensitive to temperature and laser excitation intensity and play a very significant role in the junction transformation. Hu et al. [121, 126] first observed that B ion implantation damage-induced defects can potentially produce a deformation of the LBIC in As-doped long wavelength HgCdTe infrared detector pixel arrays. This discovery unveiled an application of LBIC for the corresponding implant-induced defects analysis and characterization. Correlated theoretical models for trap-related *p-n* junction transformation have been proposed to analyze the deformation of LBIC curve induced by the extended defects.

The model structure of the *p-n* junction transformation in As-doped long wavelength $Hg_{1-x}Cd_xTe$ ($x \approx 0.224$) infrared arrays under different temperatures is shown in Fig. 2.97, When the temperature is relatively high, the deep levels (acceptor-type) induced by the ion implantation damage are fully activated and can trap significant numbers of free electrons, which are temperature sensitive. It makes the B^+ ion implantation region become n^--type or p-type as shown in Fig. 2.97b, c. Furthermore, the laser beam intensity is another key factor that determines the reversion. The photo-generated carriers are comparable to the temperature induced intrinsic carriers, and create an n^+-type to n^--type reversion [121]. At cryogenic

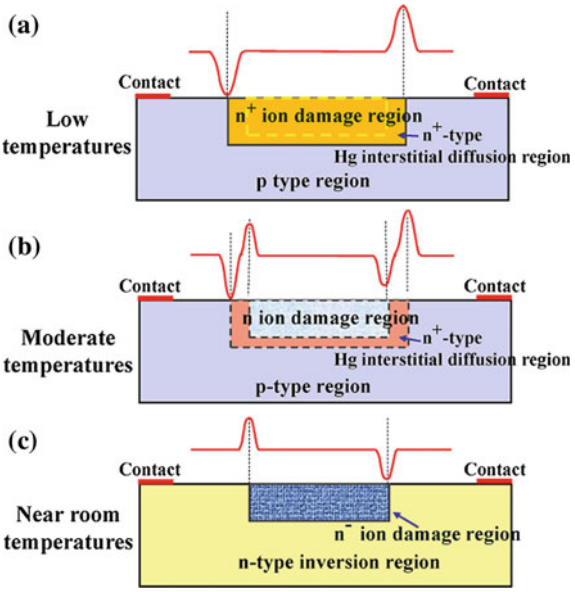


Fig. 2.97 Proposed *p-n* junction transformation models **a** at low temperature where the typical n^+-*on-p* junction is formed, **b** at moderate temperature where the *n*-n^+-*on-p* junction is formed, and **c** at near room temperature where the n^--on-*n* junction are formed. Reprinted with permission from Ref. [126] © 2012, American Institute of Physics

temperatures the implantation damage-induced traps in the n^+ region are almost inactivated, therefore, a typical n^+-on-p junction is formed.

In addition, the mixed conduction effect for p-type narrow band gap HgCdTe materials must be taken into account. This effect is mainly caused by the higher electron mobility compared with that of holes and the temperature increase generated large number of intrinsic carriers. Therefore, the narrower the forbidden bandgap, the more obvious is the mixed conduction effect. The mixed conduction, together with As-doping amphoteric behavior, makes the p-type absorption layer transform to an n-type layer at near room temperature. At moderate temperature, the coupling *n*-n^+-on-p junction is formed.

Ion implantation traps, As-doping amphoteric behavior, and the mixed conduction effect, are the main reasons for inducing the polarity reversion coupling of LBIC at different temperatures. Figure 2.98 shows the experimental results of the polarity inversion and coupling of the LBIC in As-doped long wavelength HgCdTe infrared detector pixel arrays.

In summary, accompanied by the technological development and extensive applications of infrared focal plane arrays, position-dependent LBIC or PC mapping has triggered a wave of research interest for device characterization in the early stages of the fabrication process. LBIC or PC mapping technologies rely on a scanning laser system that is close to the diffraction limit, making them highly sensitive to spatially resolved electric fields and localized nonuniformities in infrared materials. The spatially resolved information about electrical properties makes the extraction of performance parameters in infrared materials easier and more efficient through the establishment of accurate simulation models. The high

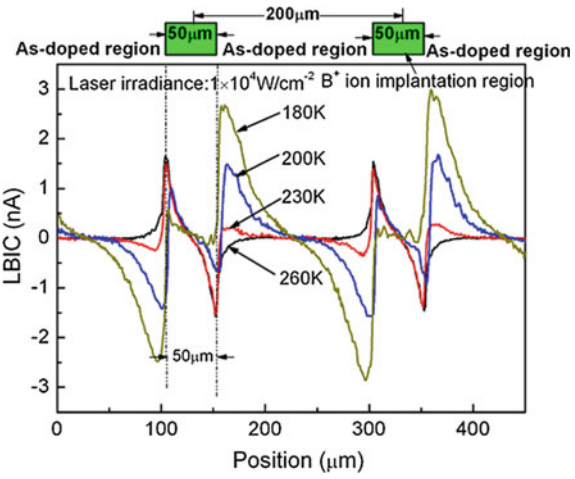


Fig. 2.98 Experimental results of temperature-dependent LBIC signal profiles with a laser power density of 1×10^4 W/cm^2. Reprinted with permission from Ref. [126] © 2012, American Institute of Physics

flexibility, low operational costs, and excellent spatial resolution properties of LBIC enable this method to be a highly useful tool for the characterization and optimization of future photovoltaic studies.

2.4 Summary

During the rapid development of HgCdTe infrared detector technology, many new models and methods have been proposed and developed. In this chapter, those models and methods, which are associated with the process of materials and devices, have been introduced. First, to improve the accuracy of device modeling, the influences of non-parabolic conduction bands and carrier degeneracy on dark currents are studied systematically. The analysis method of the HgCdTe dark current is established. At the same time, statistical results of characteristic parameters are extracted. Second, a feasible numerical method is achieved to optimize the thickness of the absorption layer and reduce the influence of the interface charge on device performance. In addition, the optimized structure of HgCdTe infrared detectors with low cross talk is proposed. Thirdly, the dependences of material and device parameters on actual HgCdTe devices are discussed comprehensively. Simulation methods for the structure characteristics, temperature effects, and cross talk of HgCdTe two-color detection are proposed. Finally, the lifetime of minority carriers in actual HgCdTe photodetector is extracted. The LBIC analytical method, which is accepted as a nondestructive tool and has been widely used for infrared photodiode array characterization, is introduced in detail. The application of LBIC with respect to extracting material and device parameters is systematically summarized.

References

1. Schmit JL. Intrinsic carrier concentration of $Hg_{1-x}Cd_xTe$ as a function of x and T using k.p calculations. J Appl Phys. 1970;41:2876–9.
2. Leonard WF, Michael ME. Carrier concentration of $Hg_{1-x}Cd_xTe$. J Appl Phys. 1974;45:958–60.
3. Hansen GL, Schmit JL. Calculation of intrinsic carrier concentration in $Hg_{1-x}Cd_xTe$. J Appl Phys. 1983;54:1639–40.
4. Lowney JR, Seiler DG, Littler CL, et al. Intrinsic carrier concentration of narrow-gap mercury cadmium telluride based on the nonlinear temperature dependence of the band gap. J Appl Phys. 1992;71:1253–8.
5. Nemirovsky Y, Finkman E. Intrinsic carrier concentration of $Hg_{1-x}Cd_xTe$. J Appl Phys. 1979;50:8107–11.
6. Chu J. The physics of narrow band gap semiconductors. 1th ed. Beijing: Science press; 2005. p. 283.
7. Madarasz FL, Szmulowicz F. Intrinsic carrier concentration in $Hg_{1-x}Cd_xTe$ with the use of Fermi-Dirac statistics. J Appl Phys. 1985;58:2770–2.
8. Kosai K. Status and application of HgCdTe device modeling. J Electron Mater. 1995;24:635–40.
9. Williams GM, De Wames RE. Numerical simulation of HgCdTe detector characteristics. J Electron Mater. 1995;24:1239–48.
10. Wenus J, Rutkowski J, Rogalski A. Two-dimensional analysis of double-layer heterojunction HgCdTe photodiodes. IEEE Trans Electron Devices. 2001;48:1326.
11. Hess GT, Thomas JS. HgCdTe double layer heterojunction detector device. SPIE. 2000;4028:353–63.
12. Gopal V, Singh SK, Mehra RM. Excess dark currents in HgCdTe p^+-n junction diodes. Semicond Sci Technol. 2001;16:372–6.
13. Velicu S, Ashokan R, Sivananthan S. A model for dark current and multiplication in HgCdTe avalanche photodiodes. J Electron Mater. 2000;29:823–7.
14. Quan ZJ, Chen GB, Sun LZ, Ye ZH, Li ZF, Lu W. Effects of carrier degeneracy and conduction band non-parabolicity on the simulation of HgCdTe photovoltaic devices. Infrared Phys Technol. 2006;50:1–8.
15. Liu E, Zhu B, Luo J. The physics of semiconductors. 6th ed. Beijing: Publishing House of Electronics Industry; 2003. p. 71.
16. Bhan RK, Dhar V. Carrier density approximation for non-parabolic and highly degenerate HgCdTe semiconductors. Semicond Sci Technol. 2004;19:413–6.
17. Bhan RK, Dhar V. Improved model for surface shunt resistance due to passivant for HgCdTe photoconductive detectors. Semicond Sci Technol. 2003;18:1043–54.
18. Wang J, Chen XS, Wang ZQ, Hu WD, Lu W, Xu FQ. The mechanism of the photoresponse blueshift for the n-type conversion region of n^+-on-p $Hg_{0.722}Cd_{0.278}Te$ infrared photodiode. J Appl Phys. 2010;107:044513.
19. Gopal V, Gupta S, Bhan RK, et al. Modeling of dark characteristics of mercury cadmium telluride n^+-p junctions. Infrared Phys. Technol. 2003;44:143–52.
20. Jo NH, Yoo SD, Ko BG, et al. Two-dimensional numerical simulation of HgCdTe infrared detectors. SPIE. 1998;3436:50–60.
21. Ye ZH, Hu XN, Zhang HY, Liao QJ, Li YJ, He L. Study of dark current for mercury cadmium telluride long-wavelength photodiode detector with different structures. J Infrared Millim Waves. 2004;23(2):86–90.
22. Hu WD, Chen XS, Yin F, Ye ZH, Lin C, Hu XN, Quan ZJ, Li ZF, Lu W. Simulation and design consideration of photoresponse for HgCdTe infrared photodiodes. Opt Quant Electron. 2009;40:1255–60.
23. Madarasz FL, Szmulowicz F. Equilibrium barrier formation in p-on-N and P-on-n graded HgCdTe heterojunctions. SPIE. 1989;1106:117–32.

24. Weiler MH, Reine MB. Effect of a valence-band barrier on the quantum efficiency and background-limited dynamic resistance of compositionally graded HgCdTe P-on-n heterojunction photodiodes. J Electron Mater. 1995;24:1329–39.
25. Wenus J, Rutkowski J. Influence of valence-band barriers in VLWIR HgCdTe P-on-n heterojunctions on photodiodes parameters. Phys Stat Sol (b). 2002;229:1093–6.
26. Vasilyev VV, Predein AV. Influence of graded p-P heterojunction's potential barrier on characteristics of three-dimensional HgCdTe photodiode. SPIE. 2005;5834:83–91.
27. Quan ZJ, Sun LZ, Ye ZH, Li ZF, Lu W. Optimization design of the band profiles of HgCdTe heterojunctions. Acta Physica Sinica. 2006;55:3611–6.
28. Migliorato P, White AM. Common anion heterojunctions: CdTe-CdHgTe. Solid-State Electron. 1983;26:65–9.
29. Bratt PR, Casselman TN. Potential barriers in HgCdTe heterojunctions. J Vac Sci Technol. 1985;A3:238–45.
30. Madarasz FL, Szmulowicz F. Barrier formation in graded $Hg_{1-x}Cd_xTe$ heterojunctions. J Appl Phys. 1987;62:3267–77.
31. Madarasz FL, Szmulowicz F. The effect of a valence-band offset on barrier formation in graded $Hg_{1-x}Cd_xTe$ heterojunctions. J Appl Phys. 1988;64:6373–8.
32. Madarasz FL, Szmulowicz F. Valence-band barrier formation in graded $Hg_{1-x}Cd_xTe$ heterojunctions with a valence-band offset included. J Appl Phys. 1989;66:3082–7.
33. Djaloshinski L, Nemirovsky Y. Methodology of abrupt heterostructures: band diagram calculations. Solid-State Electronics. 1996;39:1385–90.
34. Nemirovsky Y, Gordon G, Goren D. Measurement of band offsets and interface charges by the C-V matching method. J Appl Phys. 1998;84:1113–20.
35. Goren D, Asa G, Nemirovsky Y. Barrier formation at graded HgTe/CdTe heterojunctions. J Appl Phys. 1996;80:5083–8.
36. Johnson NF, Hui PM, Ehrenreich H. Valence-band-offset controversy in HgTe/CdTe superlattices: a possible resolution. Phys Rev Lett. 1988;61:1993–5.
37. Shih CK, Spicer WE. Determination of a natural valence-band offset: the case of HgTe-CdTe. Phys Rev Lett. 1987;58:2594–7.
38. Altschul VA, Finkman E, Bahir G. Approximations for Carrier density in non-parabolic semiconductors. IEEE Trans Elec Dev. 1992;39:1312–6.
39. Fu Y, Lu W. Physical models of semiconductor quantum devices. 1th ed. Beijing: Science Press; 2005. p. 308.
40. DESSIS ISE TCAD Manual, Release 10.06 (ISE Integrated Systems Engineering AG, Zurich, 2005), 15.137.
41. He Y, Wei T. The computer simulation method of semiconductor devices. 1st ed. Beijing: Science Press; 1989.
42. Zhao H. The computer simulation of semiconductor devices.1th ed. Tianjin: Tianjin University Press; 1989.
43. Cang T. The numerical analysis of semiconductor devices. 1st ed. Beijing: Publishing House of Electronics Industry; 1985.
44. Lopes VC, Syllaios AJ, Chen MC. Minority carrier lifetime in mercury cadmium telluride. Semicond Sci Technol. 1993;8:824–41.
45. Hu W-D, Chen X-S, Ye Z-H, Zhang J, Yin F, Lin C, Li Z, Lu W. Accurate simulation of temperature-dependence of dark current in HgCdTe infrared detector assisted by analytical modeling. J Electron Mater. 2010;39:981–5.
46. Rogalski A. Infrared detectors. Amsterdam: Gordon and Breach Science Publishers; 2000. p. 200–2.
47. Talipov NK, Ovsyuk VN, Remesnik VG, Vasilyew W. Electrical activation of boron implanted in p-HgCdTe ($x = 0.22$) by low-temperature annealing under an anodic oxide. Mater Sci Eng B. 1997;44:266–9.
48. Nemirovsky Y, Rosenfeld D, Adar R, Kornfeld A. Tunneling and dark currents in HgCdTe photodiodes. J Vac Sci Technol, A. 1989;7:528–35.

49. Nemirovsky Y, Fastow R, Meyassed M, et al. Trapping effects in HgCdTe. J Vac Sci Technol, B. 1991;9:1829–39.
50. Ellion CT, Gordon NT, Hall RS. Reverse breakdown in long wavelength lateral collection $Cd_xHg_{1-x}Te$ diodes. J Vac Sci Technol, A. 1990;8:1251–3.
51. Rosenfeld D, Bahir G. A model for the trap-assisted tunneling mechanism in diffused n-p and implanted n^+-p HgCdTe photodiodes. IEEE Trans Elec Dev. 1992;39:1638–45.
52. Hurkx GAM, Klaassen DBM, Knuvers MPG. A new recombination model for device simulation including tunneling. IEEE Trans Elec Dev. 1992;39(2):331–8.
53. Ye ZH, Wu J, Hu XN, et al. Study of HgCdTe p(+)-on-n long-wavelength heterojunction detector. J Infrared Millim Waves. 2004;23(6):423–6.
54. Kozlowski L, Vural K, Luo J, et al. Low-noise infrared and visible focal plane arrays. Opto-Electron Rev. 1999;7:259–69.
55. Quan ZJ, Chen XS, Hu WD, Ye ZH, Hu XN, Li ZF, Lu W. Modeling of dark characteristics for long-wavelength HgCdTe photodiode. Opt Quant Electron. 2006;38:1107–13.
56. Gopal Vishnu, Qiu Weicheng, Weida Hu. Modelling of illuminated current–voltage characteristics to evaluate leakage currents in long wavelength infrared mercury cadmium telluride photovoltaic detectors. J Appl Phys. 2014;116:184503.
57. Qiao H, Liao Y, Hu W D, et al. "Real-time study of gamma irradiation on $Hg_{1-x}Cd_{-x}Te$ focal plane photodiodes. Acta Physica Sinica. 2008:7088–93.
58. Quan Z-J, Z-F LI, W-D HU, et al. Parameter s extraction from the dark current characteristics of long-wavelength HgCdTe photodiode. J Infrared Millim Waves. 2007;2:003.
59. Quan Z-J, Ye Z-H, Hu W-D, et al. Study on structural optimization of decreasing the crosstalk of planar HgCdTe focal plane arrays. J Infrared Millim Waves. 2006;25(5).
60. Rogalski A. Infrared detectors: status and trends. Prog Quantum Electron. 2003;27:59–210.
61. Becker LSR. Multicolor LWIR focal plane array technology for space and ground based applications. SPIE. 2004;5564:1–14.
62. Mitra P, Barnes SL, Case FC. MOCVD of bandgap-engineered HgCdTe p-n-N-P dual-band infrared detector arrays. J Electron Mater. 1997;26(6):482–7.
63. Baylet J, Zanatta J P, Chance D, et al. Recent development in infrared FPAs with multispectral 1282 IRCMOS. SPIE. 2002;4650:128–37.
64. Rajavel RD, Brewer PD, Jamba DM, et al. Status of HgCdTe-MBE technology for producing dual-band infrared detectors. J Crystal Growth. 2000;214:1100–5.
65. Zanatta JP, Ferret P, Loyer R, et al. Single and two colour infrared focal plane arrays made by MBE in HgCdTe. Proc SPIE. 2000;4130:441–51.
66. Tennant WE, Thomas M, Kozlowski LJ, et al. A novel simultaneous unipolar multispectral integrated technology approach for HgCdTe IR detectors and focal plane arrays. J Electron Mater. 2001;30(6):590–4.
67. Smith EPG, Pham LT, Venzor GM, et al. HgCdTe focal plane arrays for dual-color mid- and long-wavelength infrared detection. J Electron Mater. 2004;33(6):509–16.
68. Ye Z. Integrated two-color HgCdTe photodetector. Ph.D. thesis; 2005.
69. Rajavel RD, Jamba DM, Wa OK, et al. High performance HgCdTe two-color infrared detectors grown by molecular beam epitaxy. J Crystal Growth. 1997;175:653–8.
70. Rajavel RD, Jamba DM, Jensen JE, Hu OK, Wilson JA, Johnson JL, Patten EA, Kosai K, Goetz P, Johnson SM. Molecular beam epitaxial growth and performance of integrated multispectral HgCdTe photodiodes for the detection of mid-wave infrared radiation. J. Crystal Growth. 1998;184:1272–8.
71. Rajavel RD, Jamba DM, Jensen JE, Wu OK, Brewer PD, Wilson JA, Johnson JL, Patten EA, Kosai K, Caulfield JT, Goetz PM. Molecular beam epitaxial growth and performance of HgCdTe-Based simultaneous-mode two-color detectors. J Electron Mater. 1998;27:747–51.
72. Ye ZH, Zhou WH, Hu WD, Hu XN, Ding RJ, He L. Spectral study on response of HgCdTe IR two-color detector arrays. J Infrared Millims Waves. 2009;28(1):4–7.
73. Hu W-D, Chen X-S, Ye Z-H, Lu W. An improvement on short-wavelength photoresponse for heterostructure HgCdTe two-color infrared detector. Semicond Sci Technol. 2010;25:045028.

74. Borniol ED, Baylet J, Zanatta JP, Mibord S, Gravrand O, Rothan F, Castelein P, Chamonal JP, Ravetto M, Destefanis G. Dual-band infrared HgCdTe focal plane array. Proc SPIE. 2003;4820:491–9.
75. Baylet J, Zanatta J P, Chance D. et al. Recent advances in development of infrared multispectral 1282 FPAs. Proc SPIE. 2002;4721:134–43.
76. Sood A, Egerton J, Puri Y, et al. Design and development of multi-color detector arrays. Proc SPIE. 2004;5564:27–33.
77. Tribolet Philippe and Destefanis. Gerard Third generation and multi-color IRFPA developments: a unique approach based on DEFIR. Proc SPIE. 2005;5783:350–65.
78. Rogalski A. HgCdTe infrared detector material: history, status and outlook. Rep Prog Phys. 2005;68:2267.
79. Cabanski W, Breiter R, Mauk KH, et al. High-resolution focal plane array IR detection modules and digital signal processing technologies at AIM. Proc SPIE. 2003;5074:72–5.
80. Ballet P, Noel F, Pottier F, Pottier F, Plissard S, Zanatta JP, Baylet J, Gravrand O, De Borniol E, Martin S, Castelein P. Dual-band infrared detectors made on high-quality HgCdTe epilayers grown by molecular beam epitaxy on CdZnTe or CdTe/Ge substrates. J Electron Mater. 2004;33(6):667–72.
81. Ye ZH, Wu J, Hu XN, Wu Y, Liao QJ, Zhang HY, Wang JX, Ding RJ, He L. SPIE. 2004;5640:66.
82. Hu WD, Ye ZH, Liao L, Chen HL, Chen L, Ding RJ, He L, Chen XS, Lu W. A 128 × 128 long-wavelength/mid-wavelength two-color HgCdTe infrared focal plane array detector with ultra-low spectral crosstalk. Opt Lett. 2014;39:5130–3.
83. Liang J, Hu WD, Ye ZH, Liao L, Li ZF, Chen XS, Lu W. Improved performance of HgCdTe infrared detector focal plane arrays by modulating light field based on photonic crystal structure. J Appl Phys. 2014;115:184504.
84. Li Y, Ye ZH, Hu WD, Lei W, Gao YL, He K, Hua H, Zhang P, Chen YY, Lin C, Hu XN, Ding RJ, He L. Numerical simulation of refractive-microlensed HgCdTe infrared focal plane arrays operating in optical systems. J Electron Mater. 2014;43:2879–87.
85. Ferret P, Zanatta JP, Hamelin R, et al. Status of the MBE technology at Leti LIR for the manufacturing of HgCdTe focal plane arrays. J Electron Mater. 2000;29(6):641–7.
86. Baylet J, Gravrand O, Laffosse E, Vergnaud C, Ballerand S, Aventurier B, Deplanche JC, Ballet P, Castelein P, Chamonal JP, Million A, Destefanis G. Study of the pixel-pitch reduction for HgCdTe infrared dual-band detectors. J Electron Mater. 2004;33(6):690–700.
87. Jozwikowski K, Rogalski A. Computer modeling of dual-band HgCdTe photovoltaic detectors. J Appl Phys. 2001;90(3):1286–91.
88. Bellotti E, D'Orsogna D. Numerical analysis of HgCdTe simultaneous two-color photovoltaic infrared detectors. IEEE J Quantum Electron. 2006;42(4):418–26.
89. Akira Ajisawa, Naoki Oda. Improvement in HgCdTe diode characteristics by low temperature post-implantation annealing. J Electr Mater. 1995;24:1105–11.
90. Cao G, Gong H, Qiu H, Kong L, Hu SH, Dai N. Bias-dependent photocurrent of $Hg_{1-x}Cd_xTe$ photodiodes. J Appl Phys. 2005;98:064504.
91. Gilmore AS, Bangs J, Gerrish A. Current voltage modeling of current limiting mechanisms in HgCdTe Focal Plane Array photodetector. J Electr Mater. 2005;34:913–21.
92. Quan ZJ, Li ZF, Hu WD, Ye ZH, Hu XN, Lu W. Parameter determination from resistance-voltage curve for long-wavelength HgCdTe photodiode. J Appl Phys. 2007;100:084503.
93. Graft A, Fischer T, Gray A, et al. Illumination-dependent dynamic resistance of $Hg_{1-x}Cd_xTe$ heterojunction photodiodes. J Appl Phys. 1993;74:5705–12.
94. Wei L, Hong L, et al. The application of solid crystallization process in optimization algorithm. Chin J Comput Phys. 1999;216:141–4.
95. Hu WD, Chen XS, Ye ZH, Lu W. A hybrid surface passivation on HgCdTe long wave infrared detector with in-situ CdTe deposition and high-density Hydrogen plasma modification. Appl Phys Lett. 2011;99:091101.

96. Chen GB, Lu W, Chen XS, Li ZF, Cai WY, He L, Hu XN, Li YJ, Shen SC. Study on optimizing the performance of infrared detectors using material chip technology. Semicond Sci Technol. 2003;18:887–90.
97. Sizov FF, Gumenjuk-Sichevska JV, Lysiuk IO, et al. Temperature dependence of the dark current in HgCdTe photodiode arrayes. SPIE. 1998;5957(59571L):1–8.
98. Wenus J, Rutkowski J, Rogalski A. Analysis of VLWIR HgCdTe photodiode performance. Opto-Electron Rev. 2003;11:143–9.
99. Juang FS, Su YK, Chang SM, et al. Analysis of the dark current of focal-plane-array $Hg_{1-x}Cd_xTe$ diode. Mater Chem Phys. 2000;64:131–6.
100. Zemel A, Lukomsky I, Weiss E. Mechanism of carrier transport across the junction of narrow band-gap planar np HgCdTe photodiodes grown by liquid-phase epitaxy. J Appl Phys. 2005;98(1–7):054504.
101. Hu WD, Chen XS, Yin F, Quan ZJ, Ye ZH, Hu XN, Li ZF, Lu W. Analysis of temperature dependence of dark current mechanisms for long-wavelength HgCdTe photovoltaic infrared detectors. J Appl Phys. 2009;105:104502.
102. Chu J. The physics of narrow band gap semiconductors. 1th ed. Beijing: Science Press;2005. p. 283.
103. Scott W, Stelzer EL, Hager RG. Electrical and far-infrared optical properties of p-type $Hg_{1-x}Cd_xTe$. J Appl Phys. 1974;47:1408–14.
104. Nishino H, Ozaki K, Tanaka M, et al. Acceptor level related Shockley-Read-Hall centers in p-HgCdTe. J. Crys. Growth. 2000;214(215):275–9.
105. Chen MC, Colombo L, Dodge JA et al. The minority carrier lifetime in doped and undoped p-type $Hg_{0.78}Cd_{0.22}Te$ liquid phase epitaxy films. J Electr Mater. 1995;24:539–44.
106. Lanir M, Vanderwyck AHB, Wang CC. Minority-carrier-lifetime determination in $Hg_{0.68}Cd_{0.32}Te$. J Appl Phys. 1978;49:6182–4.
107. Fastow R, Nemirovsky Y. The excess carrier lifetime in p-type HgCdTe measured by photoconductive decay. J Appl Phys. 1989;66:1705–10.
108. De Souza ME, Boukerche M, Faurie JP. Minority-carrier lifetime in p-type (111) B HgCdTe grown by molecular-beam epitaxy. J Appl Phys. 1990;68:5195–9.
109. Cui HY, Zeng JD, Tang NY, Tang Z. Analysis of the mechanisms of electron recombination in HgCdTe infrared photodiode. Opt. Quantum Electron. 2013;45(7):629–34.
110. Qiu Weicheng, Weida Hu. Laser beam induced current microscopy and photocurrent mapping for junction characterization of infrared photodetectors. Sci China-Phys Mech Astron. 2015;58(2):027001.
111. Qiu WC, Cheng XA, Wang R, et al. Novel signal inversion of laser beam induced current for femtosecond-laser-drilling induced junction on vacancy-doped p-type HgCdTe. J Appl Phys. 2014;115:204506.
112. Redfern DA, Fang W, Ito K, et al. Low temperature saturation of p-n junction laser beam induced current signals. Solid-State Electron. 2004;48:409.
113. Feng AL, Li G, He G, et al. Dependence of laser beam induced current on geometrical sizes of the junction for HgCdTe photodiodes. Opt Quantum Electron. 2013.
114. Redfern DA, Thomas JA, Musca CA, et al. Diffusion length measurements in p-HgCdTe using laser beam induced current. J Electron Mater. 2001;30:696–703.
115. Yin F, Hu WD, Quan ZJ, Zhang B, Hu XN, Li ZF, Chen XS, Lu W. Determination of electron diffusion length in HgCdTe photodiodes using laser beam induced current. Acta Physica Sinica. 2009;58:7885–9.
116. Ong VKS, Wu D. Determination of diffusion length from within a confined region with the use of EBIC. IEEE Trans Electron Devices. 2001;48(2):332.
117. Redfern DA, Smith E, Musca CA, Dell JM, Faraone L. Interpretation of current flow in photodiode structures using laser beam-induced current for characterization and diagnostics. IEEE Trans Electron Devices. 2006;53:23–31.
118. Redfern DA, Musca CA, Dell JM, Faraone L. Characterization of electrically active defects in photovoltaic detector arrays using laser beam-induced current. IEEE Trans Electron Devices. 2005;52:2163–74.

119. Yin F, Hu WD, Zhang B, Li ZF, Hu XN, Chen XS, Lu W. Simulation of laser beam induced current for HgCdTe photodiodes with leakage current. Opt Quant Electron. 2009;41:805–10.
120. Feng AL, Li G, He G, et al. The role of localized junction leakage in the temperature-dependent laser-beam-induced current spectra for HgCdTe infrared focal plane array photodiodes. J Appl Phys. 2013;114:173107.
121. Hu WD, Chen XS, Ye ZH, Feng AL, Yin F, Zhang B, Liao L, Lu W. Dependence of ion-implant-induced LBIC novel characteristic on excitation intensity for long-wavelength HgCdTe-based photovoltaic infrared detector pixel arrays. IEEE J Sel Top Quant. 2013;19:4100107.
122. Bubulac LO, Tennant WE. Role of Hg in junction formation in ion-implanted HgCdTe. Appl Phys Lett. 1987;51:355–7.
123. Bubulac LO. Dependence of junction formation on substrate in implanted HgCdTe. Appl Phys Lett. 1985;46:976–8.
124. Arias JM, Zandian M, Bajaj J, Pasko JG, Bubulac LO, Shin SH, Dewames RE. Molecular beam epitaxy HgCdTe growth-induced void defects and their effect on infrared photodiodes. J Electron Mater. 1995;24:521–4.
125. Buell AA, Pham LT, Newton MD, et al. Physical structure of molecularbeam epitaxy growth defects in HgCdTe and their impact on two-color detector performance. J Electron Mater. 2004;33:662–6.
126. Hu WD, Chen XS, Ye ZH, Chen YG, Yin F, Zhang B, Lu W. Polarity inversion and coupling of laser beam induced current in As-doped long-wavelength HgCdTe infrared detector pixel arrays: Experiment and simulation. Appl Phys Lett. 2012;101:181108.

Chapter 3
CdTe/Si Composite Substrate and HgCdTe Epitaxy

3.1 Introduction

Advanced HgCdTe infrared focal plane array (IRFPA) technology extends the ability of optoelectronic detection from single-band two-dimensional target imaging to multiple spectral three-dimensional imaging. It is developing continuously toward the direction to achieve higher resolution and more accurate identification. In terms of material technologies, the IRFPAs are requiring larger size wafers with higher performances and more complex structures in order to better detect multiple information from multiple objectives.

For epitaxy growth of mercury cadmium telluride, the size of the material is mainly limited by the size of a substrate. In convention, cadmium zinc telluride (ZnCdTe) and gallium arsenide (GaAs) are the two mature substrates. The advantages of these substrates are (1) they are lattice-matched or relatively matched to HgCdTe, (2) It is relatively easy to grow high-quality HgCdTe films on these substrates. But the thermal matching between Si ROICs and these substrates are not good enough to support large-scale IRFPAs. In principle, only silicon substrates can eventually overcome such technical constraints. In the meantime, Si substrates have the advantages of large size and low costs. Therefore, Si-based HgCdTe epitaxy has become one of the core technologies in manufacturing advanced HgCdTe IRFPA detectors.

However, the lattice mismatch between Si and HgCdTe is as high as 19 % which is the major obstacle in developing Si-based HgCdTe epitaxy technology. Early exploratory works followed the way of growing GaAs epilayer on Si substrates. A milestone achievement was the success of ZnTe epitaxy on Si substrates using a surface modification technique with arsenic, which laid the foundation of today's Si-based HgCdTe MBE technology.

The success of Si-based HgCdTe epitaxy proves the feasibility of very-large-scale IRFPAs although it is a quite hard work to obtain high-quality HgCdTe epitaxial materials. There are many factors that restrict the quality of materials which include the choice of Si crystal orientation, the surface treatments of Si substrate, the buffer

L. He et al., *Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN*,
DOI 10.1007/978-3-662-52718-4_3

layer growth, and the HgCdTe epitaxy process. The Si is a crystal of nonpolar diamond structure, while the II–VI compound semiconductor is a crystal of zinc-blend structure. The researchers have encountered considerable difficulties in understanding the structure change in the growing process. In the past 10 years, these difficulties and challenges have inspired researchers to undertake intensive studies in all of the fields, and some significant progress has been achieved.

Besides Si-based HgCdTe, epitaxial growth of doped multilayer compositional heterostructures is also an important part of advanced HgCdTe IRFPAs. The structure of the multilayer meets the requirements to respond to signals in multiple bands and to achieve a variety of response modes. Doping is an essential technology in manufacturing advanced multidimensional signal detectors. The low adhesion coefficient of arsenic (As) in HgCdTe is a major issue affecting the doping efficiency. Meanwhile, the amphoteric behavior of As in HgCdTe also brings challenges when activating As in HgCdTe as a p-type dopant. At present, a lack of understanding the mechanisms of As-doping and As-activation hinders improving the technology level.

This chapter focuses on the fundamentals related to the Si-based HgCdTe epitaxy and the preparation of doped multilayer materials as well as their recent progress. The main contents include: the transition process from Si surface structure to ZnTe buffer layer through As modification and its characteristics using the first-principle method, the microconfiguration of As in HgCdTe materials and the related defect energy levels as well as their effects on the HgCdTe materials, Si-based HgCdTe MBE technology, liquid phase epitaxy (LPE) technology using Si-based substrate prepared by MBE, heat treatment technology related to As-doping and As-activation process, and theoretical analysis and experimental results of the thermal stress characteristics of the materials.

3.2 Basic Models on Si-Based HgCdTe Epitaxy

Due to the alloying effects, the weak Te–Hg bond in HgCdTe materials results in a high density of Hg vacancy defects. The applications of HgCdTe materials mainly depend on whether they can be doped with a donor or acceptor. There are extremely complicated defect structures and impurity behaviors. These defects and impurities restrict the performances of HgCdTe infrared detectors.

With the development of materials preparation technology and material processes, n-type doping technology is relatively mature in HgCdTe. In recent years, attentions have been focused on the group V elements, especially arsenic, as the p-type dopant. However, the incorporation of As will induce distortion of the crystal structure and changes the electronic energy states. The amphoteric behavior of As is also complicated and difficult to be clarified. Using a semi-empirical analysis method, an activation model has been suggested by Berding et al. in 1999. This model partly explains the dependence of As p-type activation on the low-temperature annealing process conditions. However, the mechanism is still not very clear. Arsenic p-type doping has been one of the major challenges in further

improvements of HgCdTe performances. Hence, the understanding of the amphoteric behavior of As is important to optimize the properties of HgCdTe.

On the other hand, the thermal expansion mismatch between the Si readout chip and the HgCdTe-based infrared focal plane arrays (IRFPAs) constrains the format of IRFPAs. One of the possible solutions to this problem is to use silicon substrates which can match the thermal expansion of Si readout chip. With the developments of material theory, numerical methods, and computer capability, materials design and simulations have been gradually turned from semi-empirical qualitative description into a quantitative prediction and an accurate control stage. First-Principle calculations help to predict and guide the epitaxial growth of HgCdTe materials on Si substrates. Combined with Miller Index direction of the substrate surface, reconstruction form, epitaxial growth mode etc., formation energy and stability of dislocation in HgCdTe can be obtained by the First-Principle method. The mechanism of inhibition of dislocation density can be revealed through different types of dislocation structures built theoretically and taking into account the methods of inhibiting formation of dislocations and reducing the dislocation density on experiments. These results help the development of large-scale focal plane devices.

In summary, the dependence of As-doped HgCdTe epitaxial materials technology and Si-based HgCdTe epitaxial technology on basic researches has become stronger. Using the first-principles methods based on quantum analysis, this chapter will focus on the defect forms of impurities, doping activation, and p-type activation model with multi-quantum coupling effects in HgCdTe. Through theoretical investigations, we will have an insight into physical basis about quantum characteristic structure of impurities and defects and selective growth of HgCdTe-on-Si substrates.

3.2.1 Physical Model of Selective Growth on Si Surface (Mechanism of as Passivation on Surface)

The {111} plane has been widely used in LPE growth of HgCdTe, while {211} plane has been the preferential orientation in MBE growth. Compared to other crystal planes, CdTe(211)B with a high sticking coefficient of Hg [1, 2] favors suppression of the growth of twins and has been demonstrated to be very good substrates for the fabrication of infrared photo detectors. Large-area high-quality CdTe(211)B layers have been successfully grown on the As-passivated Si(211) surface for subsequent growth of HgCdTe experimentally [3–7]. Devices fabricated by these HgCdTe materials have also shown better performances [8, 9].

Studies have shown that pretreatment of the substrate on Si(211) is a very critical step. The pretreatment cleaning procedure of Si(211) is similar to Si(100) [10]. Before being loaded into the vacuum chamber, the Si(211) wafer has been RCA cleaned. The wafer is then placed in an ultra-high vacuum degassed at 400 °C and finally deoxidized at 850–900 °C. For epitaxial growth, the wafer should be cooled to an appropriate temperature. During the cooling process, As_4-flux is first introduced to passivate the Si surface. The passivation can improve the quality of

epitaxial layer and inhibit the A-face epitaxial growth [3, 11, 12, 13, 14]. According to the RHEED experimental results, both the clean and As-passivated Si(211) surfaces are (2 × 1) reconstructions [13, 15]. The coverage of the arsenic is about 77 %, not 100 %, as in XPS experiments [11, 12, 16]. That means arsenic cannot passivate all the surface atoms.

There are, so far, few theoretical reports in the literature. Most of the available studies focus on the Si(100) and Si(111) surfaces. Si(100) surface shows $p(2 \times 2)$, or $c(4 \times 2)$ reconstruction predominantly [17]. (2 × 1) reconstruction is a metal stable structure on Si(111) surface [18]. After annealing, Si(111) surface shows (7 × 7) reconstruction [19]. For (100) surface, arsenic is adsorbed to form As-As dimers in most cases [20]. Some studies indicate that the topmost Si atoms of the Si (111) surface can be replaced by As atoms [21]. Most of the theoretical studies are about low-Miller-Index surfaces considering its simplicity and these results are under debate [22]. With epitaxial large-area HgCdTe wafers, many experiments use the {211} plane as the substrate. It is necessary to reconsider the reconstruction configuration and passive mode of Si(211). For the Si(211) clean surface, Mankefors found that dangling bonds of the atoms at step edges can cause a metallic characteristic [23]. Total energy calculation results show 1 × reconstruction is more stable than 2 × reconstructions but it has larger twists [24].

Some theoretical results have shown the surface reconstruction of Si(211) [25]. However, to our knowledge, that there are few studies concerning the understanding of the nucleation process of II–VI materials on the clean and arsenic-passivated Si (211) surfaces [26, 27]. The adsorptions of Te, As, Ga, H, Ge, Br, Sb, and Al atoms on Si(211) surface have been studied by Dhar et al., Sen et al. and Gupta et al. [28–32].

1. Method

Based on the density functional theory, the first-principle calculations were used [33]. All the atoms in the cell were represented by Vanderbilt ultrasoft pseudopotentials [34]. The electronic wave functions were expanded in a plane-wave set with a cutoff of 270 eV and the Schrodinger equation was solved within the framework of the density function theory combined with the generalized-gradient correction GGA-PW91 of the exchange-correlation potential [35]. The geometry optimization was carried out by using the Broyden–Fletcher–Goldfrab–Shanno routine [36].

The Si(211) surface was represented in a repeated slab geometry. Each slab contained nine Si layers and a vacuum region of 8 Å. Each layer contained four Si atoms —two along $[01\bar{1}]$ and two along $[\bar{1}11]$. The dangling Si bonds at the bottom of the slab were saturated with hydrogen atoms to simulate bulk-like situation. The z axes was taken perpendicular to the Si(211) surface, while x and y axes were along $[\bar{1}11]$ and $[01\bar{1}]$, respectively. Here 2 × 2 × 1 k-points were used for the Monkhorst–Pack k-point sample [37]. Two lowermost Si layers and the additional H layer were fixed.

2. Bulk-terminated Si(211) surface

Vicinal (111) with tilts up to 19.5° toward (001) is considered as the “ideal” bulk-terminated (211) Si surface. The unit cell of the 1 × 1 surface contains two

threefold-coordinated terrace atoms and one twofold-coordinated edge atom. The surface atoms are named after Gupta's notation [26]. Each threefold-coordinated atom has one dangling bond on the first terrace, called "the terrace atom" (marked as "T"), and on the second terrace, called "the trench atom" (marked as "Tr"); each twofold-coordinated atoms on the step edge is denoted as "E" (also called "the step edge"), having two dangling bond. All of the bond lengths are 2.34 Å and the angles are 109.5°, which are consistent with the bulk silicon material. The detailed structure is shown in Fig. 3.1. The space between adjacent layers is 1.10 Å. The vertical distance is 3.90 Å between "E" and "Tr" and 3.12 Å between "Tr" and "T". In the latter discussion, quotation marks are used to indicate the location of an atom. By using the same computational method, the DOS of the simulated structures is calculated and the results are shown in Fig. 3.2. The bulk-terminated surface shows typical metallic properties because of the high dangling-bond density. It is the same as the symmetric-(2 × 1)Si(100) surface.

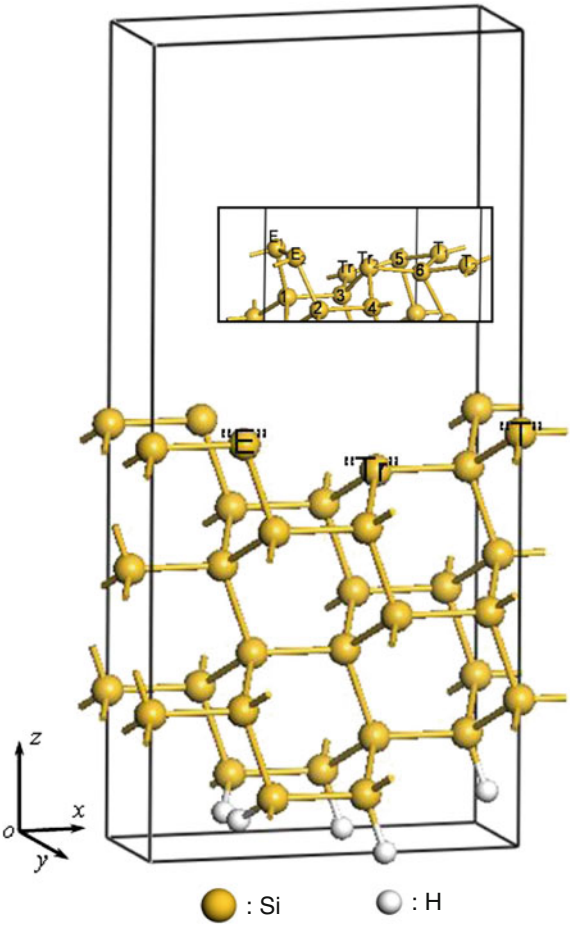


Fig. 3.1 Ideal geometry of the Si(211) surface. The inset shows the reconstructed structure of the top three layers. Reproduced from Ref. [38] with kind permission from World Scientific Publishing

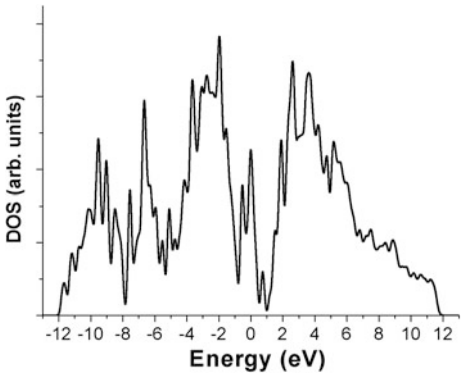


Fig. 3.2 Calculated density of states (DOS) for the ideal Si(211)

3. Si(211)(2 × 1) reconstructed surface

The bulk-terminated surface is an unstable structure because of its high dangling-bond density. In other words, the bulk-terminated surface does not exist in reality. The edge Si atoms, which have two dangling bonds, are extraordinarily unstable. Hence they are expected to dimerize two by two, leading to a doubling of the lattice vector along the [$[01\bar{1}]$] direction. The reconstructed (2 × 1) model is shown as an inset in Fig. 3.1. The bond of two edge atoms is 2.39, 0.89 Å shorter than that on the ideal surface. The two terrace atoms move closer to 3.47 Å for the convenience of the edge atoms to rebond. The average length of the surface bonds is 2.23 Å, smaller than the bond length in the bulk material. The largest shrinkage is 5 %. The reconstruction reduces the total energy 2.6 eV, i.e., 0.04 eV/Å^2. Besides the dimerization, the edge atoms move upward and the terrace atoms move downward under the force of the *sp*-hybridization. The surface layers slant to be a line. Mankerfors's calculation results have shown the same trend [23]. The bond angle of T_1–E_1–6, T_2–E_2–6, T_1–E_1–5 and T_2–E_2–5 is 117.6°, 114.5°, 129.2°, and 132.1° separately. The six surface atoms (T_1, T_2, E_1, E_2, 5, 6) are in the same inclined plane. The bond distortion directions of Tr_1–5 and Tr_2–6 are slightly different. "Tr_2" is higher than the entire plane. "Tr_1" is in this plane exactly. "Tr_2" is 0.79 Å higher than "Tr_1" in *z* direction. Figure 3.3 describes the charge density difference of bonds Tr_1–5 and Tr_2–6. All these facts indicate that the surface functions are not of the sp^3 characteristic anymore and have been converted to sp^2 hybridization. The results consist with that obtained by Grein [24].

In the simulation below, the reconstructed surface was used as a clean growth substrate. Although the unreconstructed surface was adopted in the calculation of Ref. [27], the (2 × 1) reconstruction indeed happened, which can be seen from their resulting figures [27].

The (2 × 1) reconstruction results in a semiconducting surface, which can be verified from Fig. 3.4. The occupied (unoccupied) surface state corresponds to filled (empty) dangling-bond surface states. The detailed results can be obtained by analyzing the localized-DOS of each atom. Primarily, the peak on the top of the valance band is caused by all the surface atoms except for the lower "Tr" atoms.

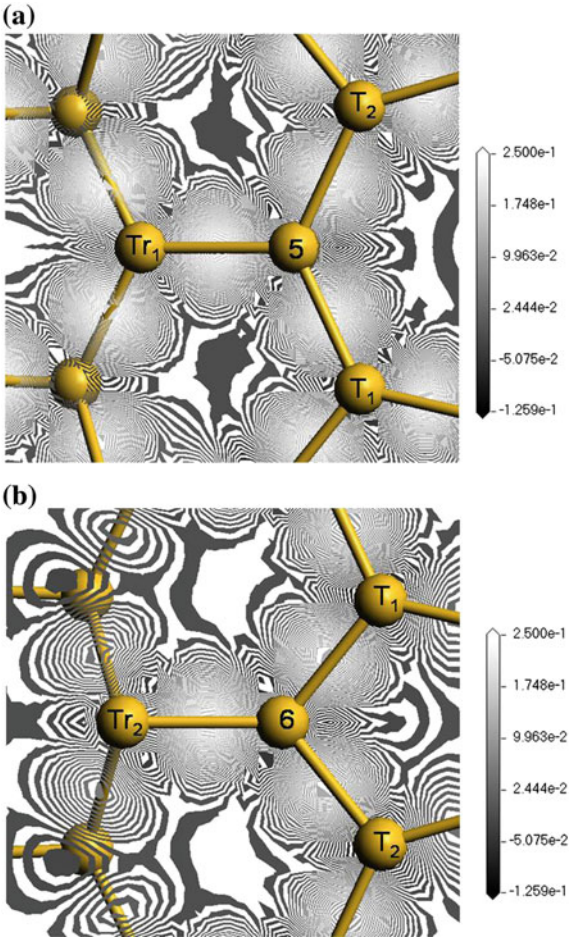


Fig. 3.3 Bond charge density of silicon reconstructed surface. **a** bond Tr_1-5, **b** bond Tr_2-6

The peak at the bottom of the conductive band comes from the "T" and the lower "Tr" atoms. Because the surface states appear between the valance and conductive bands, the band gap is half of that in the bulk material.

Table 3.1 shows the overlap population of ideal surface and reconstructed surface. Because of reconstruction, the overlap population of surface bonds increased by 21.4 %. The overlap population is 0.74 in bulk Si. This means that the surface reconstruction enhances the covalent bonds. Bond lengths and angles of the reconstructed surface have been listed in Table 3.1. We will discuss this data in a later section about As-passivated surfaces.

4. As-exposed Si(211) reconstructed surface

To ensure a single-domain B surface polarity in the process of silicon epitaxial CdTe, As-passivation is the key step in experiments. However, the mechanism is still unclear. It has been observed by experiments that there is no difference in the

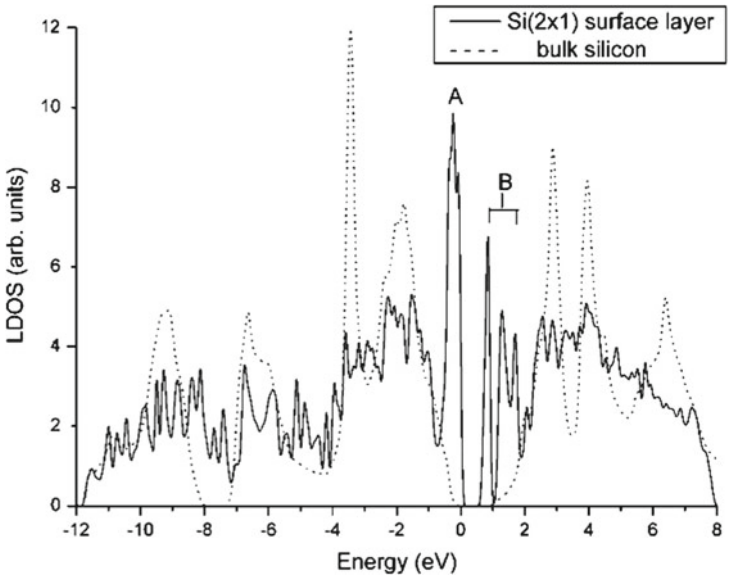


Fig. 3.4 Calculated local density of states (LDOS) for the surface layer of (2×1) reconstructed Si (211) in solid line. The dotted lines indicates the calculated DOS for the crystal silicon

Table 3.1 Bond length and bond population of ideal and (2 × 1) reconstructed Si(211) surface

Bond	Bond length (Å)	Bond population		
		Ideal surface	(2 × 1) reconstructed surface	Change rate (%)
E_1–E_2	2.39 (2.39[a])	Unbonded	0.72	–
E_1–T_1	2.23 (2.33[a])	0.76	0.89	17.1
E_2–T_2	2.22	0.76	0.92	21.1
Tr_1–5	2.31 (2.33[a])	0.69	0.79	14.5
Tr_2–6	2.29	0.69	0.70	1.4

[a]Reference [24]

passivated form of Si(211) surface before and after reconstruction [13]. Nano facets on the clean surface disappear after passivation [13]. The experimental results demonstrate that the passivation of the arsenic not only maintains the reconstruction form of the original clean Si(211) surface, but also stabilizes it further. There are two popular mechanisms for arsenic passivation on silicon surface: adsorption and replacement. Many study results indicate the topmost Si atoms of the Si(111) surface are replaced by As atoms during adsorption [21]. For (100) surface, arsenic is generally considered to be adsorbed on it to form As-As dimmers [20]. But it is still under debate. It is relatively more complicated on Si(211) surface because both (111)-like and (100)-like atoms exist in this case. So, it is necessary for us to study both of the passivation situations.

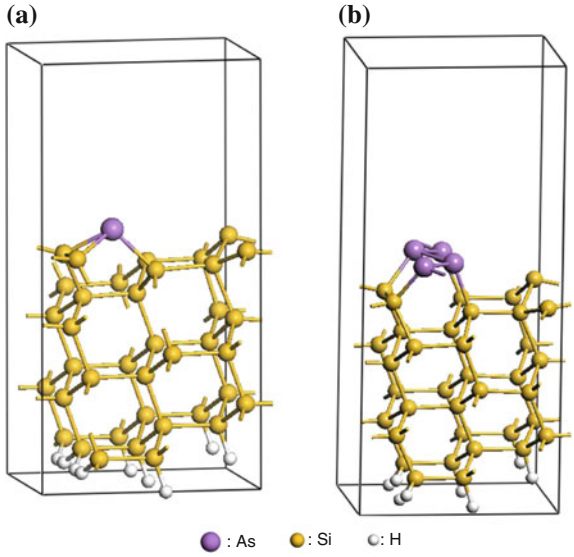


Fig. 3.5 Adsorption structures of **a** single and **b** multiple As atoms on Si(211) reconstructed surfaces. Reproduced from Ref. [38] with kind permission from World Scientific Publishing

(1) Adsorption model

The most suitable adsorption site for an individual As atom is in the trench (Fig. 3.5a). The As atom forms three bonds with one edge Si atom and two trench Si atoms, respectively. Arsenic belongs to group VA elements. Its s^2p^3 configuration of valence electrons lacks three electrons to complete an eight-electron closed-shell, which can be achieved by forming three covalent single bonds with other atoms. After several As atoms are adsorbed on the surface, they form zigzag structures (Fig. 3.5b). One bond contacts to the edge or trench Si atoms, the other two bonds contact the other two As atoms. This reconstruction form is actually not consistent with the original (2×1) form.

The above results are consistent with the results obtained by Sen et al. [27]. However, RHEED results demonstrated that the reconstruction form does not change. There is a contradiction between adsorption model and experimental results.

(2) Replacement model

In a replacement mechanism, when arsenic passivates the substrate, the surface reconstruction form is not changed. In the bulk material, the silicon atom must form four sp^3 valence bonds with the nearest neighbor atoms, and arsenic only needs three p^3 valence bonds. So arsenic is more appropriate for the surface, on which every atom has three bonds with the neighboring atoms. The substitution of arsenic for silicon annihilates the dangling-bonds of the surface atoms and stabilizes the surface geometry further.

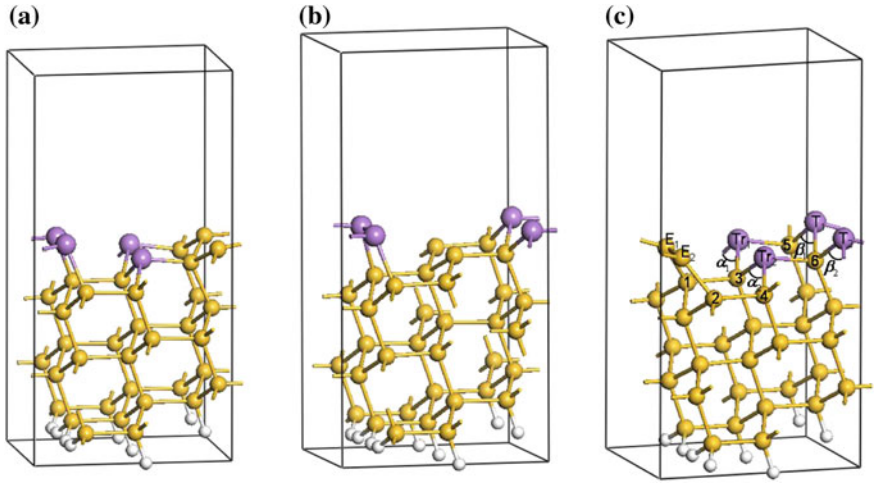


Fig. 3.6 Several geometries for substitutional As adsorption on the reconstructed (2 × 1)Si(211) surface: **a** As atoms substitute Si atoms at E and Tr sites. **b** As atoms substitute Si atoms at E and T sites. **c** As atoms Substitute Si atoms at Tr and T sites. Reproduced from Ref. [40] with kind permission from Springer Science and Business Media

It was observed by RHEED experiments that the form of surface reconstruction does not change before and after As-passivation [13]. Nano facets on the clean surfaces disappear after passivation [39]. The experimental results demonstrate that the passivation of the arsenic not only maintains the reconstruction form of the original clean Si(211) surface, but also stabilizes it further. Therefore, combined with the previous calculation results, studies show Si(211) surface passivation is not a process of atomic adsorption, but a process of atomic displacement.

According to the XPS experimental results, the coverage of arsenic is about 70 %, not 100 % [11, 12, 16]. That is to say, arsenic cannot passivate all the surface atoms. By adopting the replacement mechanism, we calculate and compare several partial As-passivation configurations with the arsenic coverage around 67 % (Fig. 3.6). The energy difference between these structures is 0.48, 1.07, and 0.0 eV separately. The most stable structure is illustrated in Fig. 3.6c, where the terrace and trench silicon atoms are replaced by arsenic atoms. The bond between the two rebounded edge atoms is 7 % shorter than the one on the bulk-terminated surface, and the other bonds are 2 % larger. Compared with the clean surface, the bond between the two rebounded edge atoms is 8 % shorter than the one on the surface, and the other bonds are 3 % larger. Hence, the As-passivation enhances the reconstruction of the substrate surface and weakens the distortion difference of the surface atoms. The total energy of this structure is 1.46 eV lower than that of adsorption model. Experimental results and simulation results indicate that As-passivation is a process of surface Si atoms partly replaced by As atoms. In the following studies, we take this structure

(Fig. 3.6c) as the As-passivated substrate. Each unit cell is divided into two parts: the As-passivated part (this part include As atoms at terrace and trench sites) and the clean part (this part include Si atoms at edge sites).

The bond length and bond population of the surface bonds are listed in Table 3.2. Aresenic passivation enhances the reconstruction of the substrate surface. The E_1–E_2 bond shrinks 8.4 % from 2.39 to 2.19 Å. The population of the E_1–E_2 bond is up to 1.23, 0.51 higher than the one before passivation. The covalence characteristic of this bond has been enhanced greatly. The passivation weakens the distortion difference of the surface atoms as well. The edge atoms move downward to form *sp* hybridization. So the six atoms (T_1, T_2, E_1, E_2, 5, 6) are not in the same plane after passivation. The high difference between the Tr_1 and Tr_2 arsenic atom in the perpendicular direction is fading away. In our calculation, except the E_1–E_2 bond, other surface bonds in the parallel $[\bar{1}11]$ direction are increased by 4 % on average compared with the clean surface. From the perspective of bond population, the covalence characteristic of the surface bonds connecting As with Si atoms is much weaker than the E_1–E_2 (Si–Si) bond.

Using VASP, the large bond length contract proportion is up to 8 % in the calculated case of the in situ impurity As_{Te} in HgCdTe [41]. The large lattice mismatch between Si and HgCdTe is the main cause of the low-quality interface. These results indicate that the enlargement of silicon substrate surface and the shrinkage of the HgCdTe material caused by arsenic can reduce the mismatch between the substrate and the adsorption materials to some extent. Comparing the DOS of partially As-passivated Si(211) substrate with the entirely As-passivated one (Fig. 3.7), we find that the surface states are annihilated totally in the entirely-passivated model, but they are preserved in partially passivated model even though the strength is weakened.

(3) Replacement process

Now the replacement process for substitutional As adsorption are investigated. Several possible geometries that may appear during this process have been calculated. We extract the most possible exchange way from these structures. The initial geometry is the As atoms deposit on the Si(211) surface (Fig. 3.8a). First, two of the As adatoms exchange with the Si atoms at the "E" sites to form the structure showed in Fig. 3.8b. Although the total energy is 0.37 eV/adatom higher than the initial one, the total energy will be 0.53 eV/adatom lower after the As atoms at the "E" sites exchange with the Si atoms at the "T" sites to form the structure showed in Fig. 3.8c, that is, 0.16 eV/adatom lower than geometry in Fig. 3.8b. Secondly, the remaining two As adatoms exchange with the Si atoms at the "Tr" sites and the exchanged Si atoms will fly out from the surface to form Fig. 4d. Only by this dissociation and the subsequent diffusion does the As sites recover their reactivity. The energy evolutions of the process are shown in Fig. 3.9. Energy barrier during this process is 0.37 eV/adatom.

Table 3.2 Bond length, bond angle, and population of the Si(211) surface

Bond	Bond length (Å)			Bond population			Bond angle (°)		
	(2 × 1) reconstruction	After passivation	Change rate (%)	(2 × 1) reconstruction	After passivation	Change rate (%)		(2 × 1) reconstruction	After passivation
E_1–T_1	2.23	2.34	5.2	0.89	0.50	−43.8	α_1	122.7	103.5
E_2–T_2	2.22	2.32	4.6	0.92	0.49	−46.7	α_2	102.4	107.5
Tr_1–5	2.31	2.38	3.0	0.79	0.57	−27.8	β_1	125.4	105.6
Tr_2–6	2.29	2.35	2.9	0.70	0.45	−35.7	β_2	99.5	104.9
E_1–E_2	2.39	2.19	−8.4	0.72	1.23	70.8			

Reproduced from Ref. [40] with kind permission from Springer Science and Business Media

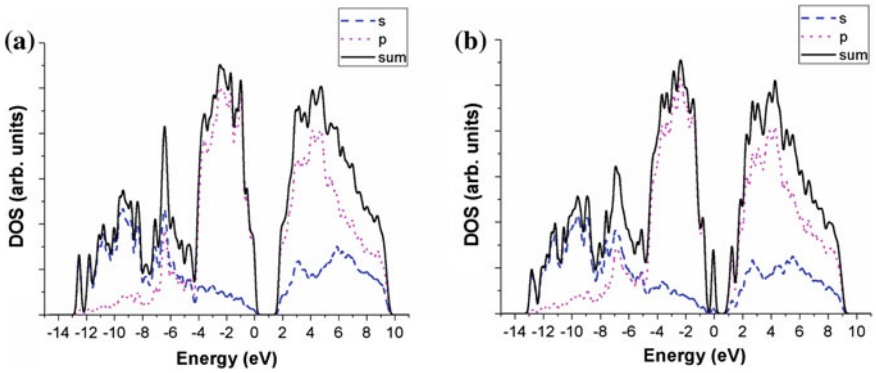


Fig. 3.7 The calculated DOS for **a** entirely and **b** partial As-passivation silicon substrates

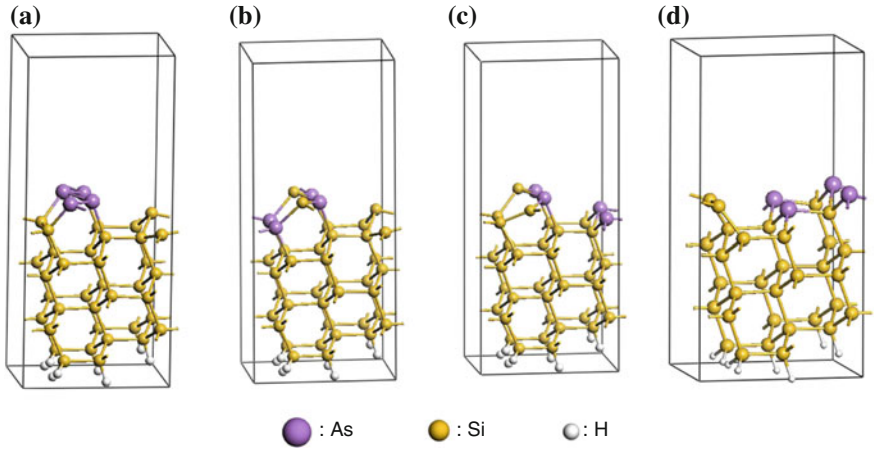


Fig. 3.8 Schematic diagram of the substitution process

3.2.2 Atomic Distribution Model of Si Substrate ZnTe/CdTe

Adsorption of independent Cd or Te atoms has been simulated using clean and As-passivated Si (211) surface as an epitaxially grown substrate. An independent Cd or Te atom can be adsorbed on the clean Si(211) surface stably, but they are hard to be deposited on the passivated region of the As-passivated surface especially for Cd atom.

Consistent with the experimental results, As-passivation not only plays a key role in the selective B-face type growth, but it can also improve the quality of the epitaxial layer to some extent.

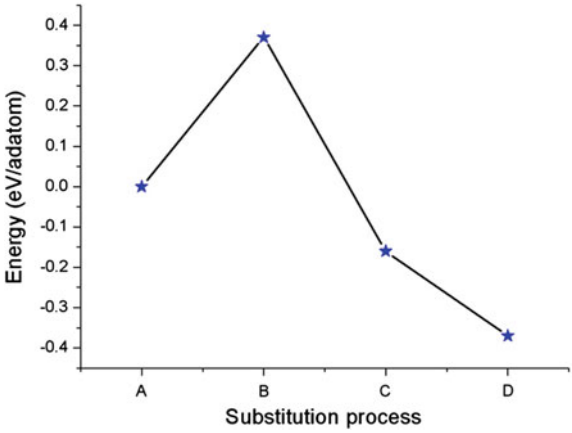


Fig. 3.9 Energy evolutions of the substitution process

1. Adsorption of an independent Te atom

Here, a clean substrate refers to Si(211)(2 × 1) reconstructed surface. The As-passivated Si (211) substrate refers to the clean surface in which terrace and trench silicon atoms are replaced by arsenic atoms. An independent Te atom is relatively easy to grow on the clean and As-passivated Si (211) surface, but the most stable adsorption position is slightly different.

On the clean substrate, Te atom prefers to form bonds with Si "T" and "E" Si atoms on the highest platform (Fig. 3.10a). The total energy is 0.5 eV lower than other stable deposit structures.

On the As-passivated Si(211) substrate, Te atom prefers to form bonds with two reconstructed "E" Si atoms. The average bond length between Te and Si atom is 2.52 Å. These bonds tilting to step edges saturate dangling bonds of surface Si atoms, without influence on the rebond between "E–E" atoms (Fig. 3.10b). Although a Te atom can be deposited at a passivated region, the adsorption energy is 2.5 eV higher.

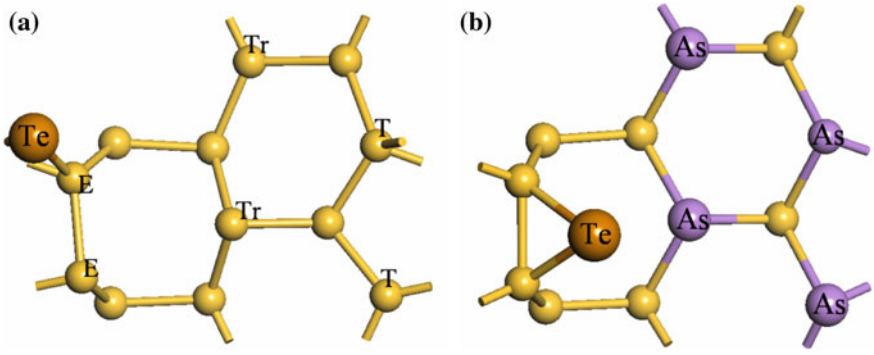


Fig. 3.10 Top view of single Te atom adsorbed on a Si(211) surface. Reproduced from Ref. [38] with kind permission from World Scientific Publishing

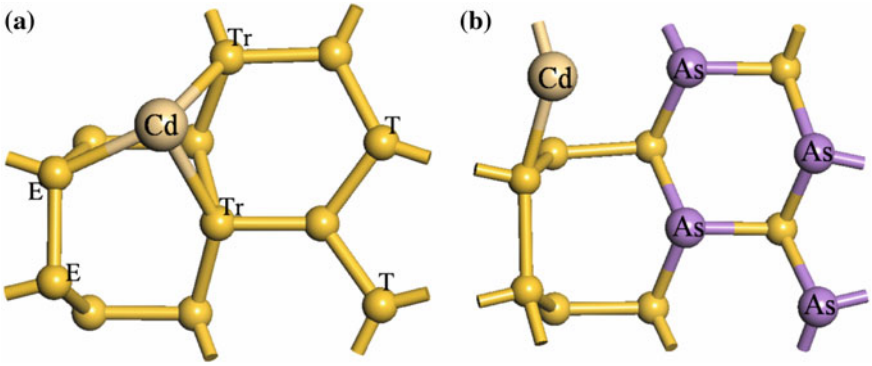


Fig. 3.11 Top view of single Cd atom adsorbed on Si(211) surface. Reproduced from Ref. [38] with kind permission from World Scientific Publishing

2. Adsorption of an independent Cd atom

A Cd atom can form three bonds with "E" and "Tr" Si atoms at the trench of the clean substrate (Fig. 3.11a). The average bond length between Cd and Si atom is 2.75 Å. The adsorption energy is 1.1 eV lower than other structures. On the As-passivated Si (211) substrate, a Cd atom prefers to form bonds with two reconstructed "E" Si atoms at unpassivated region (Fig. 3.11b). The reconstruction of "E-E" has been destroyed. At the passivated region, no stable deposit structure forms because of the strong repulsive force between As and Cd atoms.

3. Atomic model of CdTe epitaxy growth

An independent Cd or Te atom can be adsorbed on the clean Si(211) surface stably, but they are hard to be deposited on the passivated region of the As-passivated surface. CdTe is a polar semiconductor. During the epitaxy growth of CdTe, the surface polarity problems will be involved. The adsorption of an independent Cd or Te atom cannot reveal the selective polarity mechanism clearly. In order to understand the selective mechanism, 1ML CdTe is deposited on the surface. HgCdTe(211) B surface structure has been described in Refs. [42, 43]. For clarity, Fig. 3.12 shows the (211) A, B polar face structures. Cd atoms deposited above "E" positions, and Te atoms deposited above the "T" and "Tr" positions in the A-face type growth. The arrangement of Te and Cd atoms in the B-face type growth is exactly the opposite. Te atoms deposited above "E" positions, and Cd atoms deposited above the "T" and "Tr" positions.

After geometry optimizations, we obtain the stable structures which absorb 1ML CdTe. On the clean substrate, the atoms at the "T," "E," or "Tr" sites can form bonds with cadmium or tellurium easily (Fig. 3.13). So the A- or B-face type epitaxial growth takes place. After the passivation with arsenic, the dangling bonds of substrate surface are saturated partially, which can cause the surface states to become weakened. Although Cd can bond with edge Si atoms, the strong repulsion

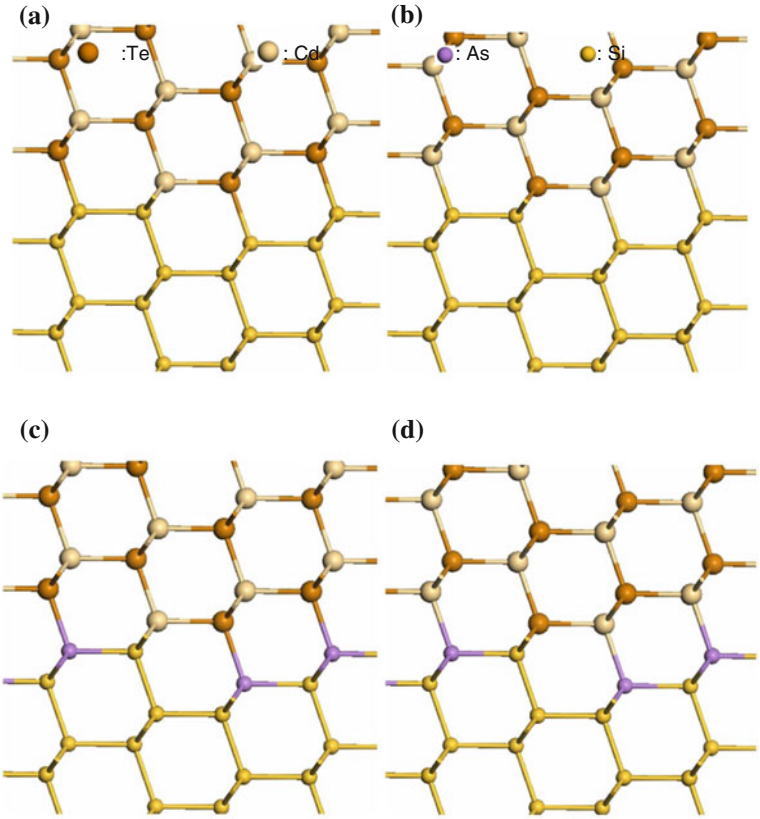


Fig. 3.12 Schematic diagram of the (211) A, B polar face. **a** A-CdTe/Si, **b** B-CdTe/Si, **c** A-CdTe/As/Si, **d** B-CdTe/As/Si

between arsenic and tellurium prevents them from bonding. So the A-face epitaxy growth is inaccessible (Fig. 3.14a). According to our calculated DOS (Fig. 3.14c), the Te atoms and Si atoms at E sites can cause a peak in the gap. In the B-face epitaxy growth (Fig. 3.14b), Te can be adsorbed by the edge Si atoms easily. Cadmium can also overcome the repulsion coming from arsenic. No peak has been observed in the gap of the calculated DOS (Fig. 3.14d) and the bonds between As and Cd make the B-face epitaxy growth possible. The partial As-passivation plays an important selective role in the B-face epitaxial growth. The bond lengths of these structures are shown in Table 3.3.

When the substrate absorbed 2ML CdTe, there was a greater distortion even significant void in A-face type growth. In B-face type growth there has been a large mess but in terms of less distortion compared to the A-face type growth (Fig. 3.15). Due to the large lattice mismatch, the results cannot be used for analysis. However, it shows that partial As-passivated substrate is in favor of B-face type growth of CdTe epitaxial layers.

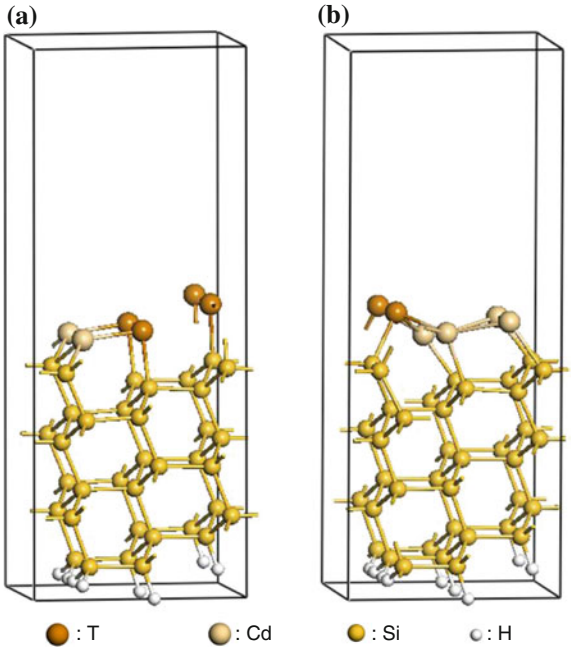


Fig. 3.13 Simulated 1ML CdTe deposited on the clean Si(211) substrates by **a** A-face and **b** B-face epitaxial growth types

By using the model of valence electron counting [13], Chen has discussed the interface between the substrate and the epitaxy growth layer. A or B-face epitaxy growth of CdTe on clean Si(211) surface can make the number of electrons match. On the As-passivated surface, they consider the tilt surface which contains the T and Tr atoms only, not including E atoms. However, in their discussions, the replacement ratio of As is 50 % [13]. We use this method to reconsider the number of electrons at the interface between the As-passivated surface and CdTe(A) or (B) epitaxy layers. The total electron counting at the interface parts are zoomed out in Fig. 3.16. Each Si atom can contribute one electron to each bond; and each As, Cd, and Te atom can contribute 5/4, 1/2, and 3/2 electrons, respectively. The total electrons of each bond should be 2. If the epitaxial face is A type, the superfluous 3 (3/2 times 2) electrons of the two As–Te bonds overcome the lack of 1/2 electrons of the Si–Cd bond. The electron count at the interface mismatches. If it is B type, the superfluous 1/2 electrons of the Si–Te bond at E site can be compensated by the two As–Cd bonds at T and Tr sites. This makes the electron count match naturally.

4. Atomic structure model of ZnTe epitaxy growth

To avoid the large lattice mismatch between Si substrate and CdTe epitaxial layers, ZnTe buffer layers will be grown before epitaxial growth of CdTe layers in experiments [14–18]. We investigate ZnTe epitaxial growth on different As-passivated Si (211) substrates theoretically. Calculation results are very close to the epitaxial growth of CdTe. On clean substrates, A- or B-face type growth can occur.

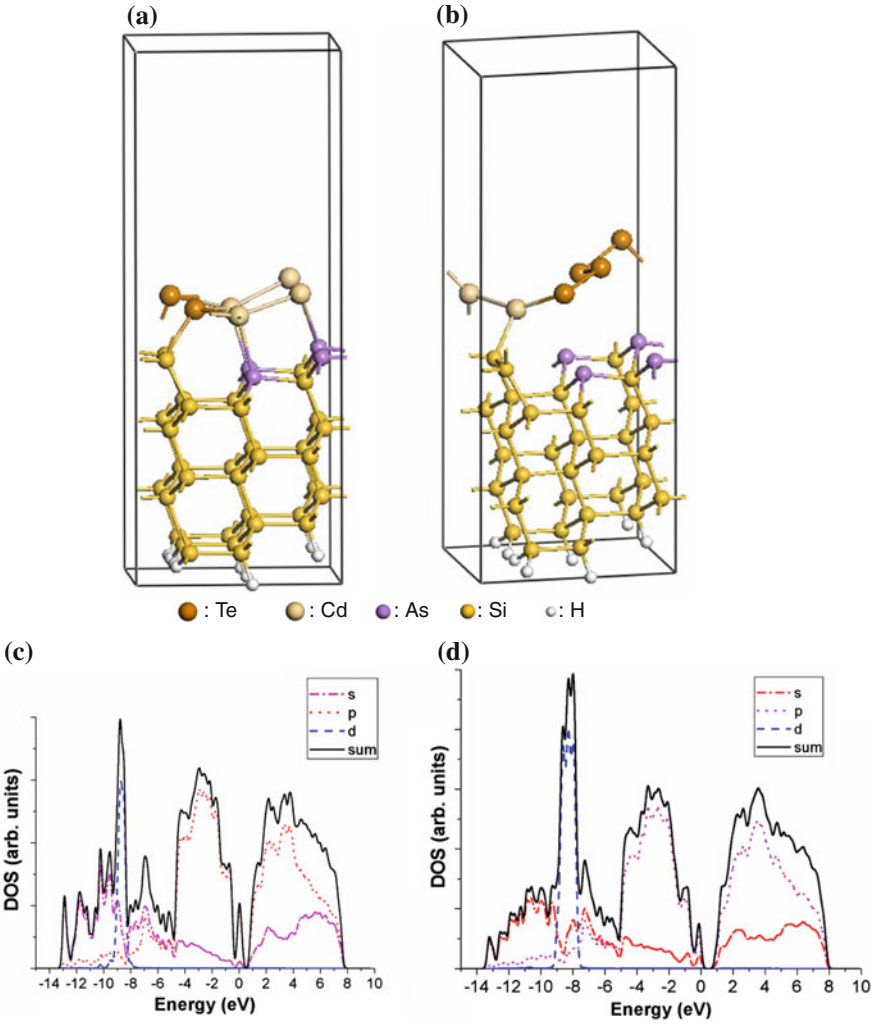


Fig. 3.14 Simulated 1ML CdTe deposited on the As-passivated Si(211) substrates by **a** A-face and **b** B-face epitaxial growth types. Calculated DOS of **c** A-face and **d** B-face epitaxial growth structures. Reproduced from Ref. [40] with kind permission from Springer Science and Business Media

Furthermore, A-face type epitaxial layers can only grow on clean substrates (Fig. 3.17a), but cannot grow on partially or entirely As-passivated substrates (Fig. 3.17b, c). B-face type epitaxial layer can grow both on clean substrates and As-passivated substrates (Fig. 3.17d, e, f). Figure 3.17a shows the stable structure of 2ML ZnTe absorbed on clean substrate. The substrate is entirely As-passivated (Fig. 3.17b) and partially As-passivated (Fig. 3.17c). There are great distortions in B-face growth on clean or entirely As-passivated substrates, while less distortion on

Table 3.3 Bond lengths of 1ML CdTe adsorbed on clean and As-passivated Si(211) surface (*Unit* Å)

Bond	A-face growth		B-face growth	
	Clean Si(211) surface	As-passivated Si (211) surface	Clean Si(211) surface	As-passivated Si (211) surface
1	2.39	2.26	2.40	2.34
2	2.38	2.26	2.34	2.32
3	2.38	2.31	2.40	2.33
4	2.32	2.31	2.33	2.32
5	2.45	Unbonded	2.46	2.34
	–	–	Te-Si: 2.51	Te-Si: 2.46
	–	–	Cd-Si: 2.58	Cd-As: 3.02
	Te-Si: 2.53	Te-As: unbonded	–	–
	Cd-Si: 2.52	Cd-Si :2.64	–	–

Reproduced from Ref. [40] with kind permission from Springer Science and Business Media

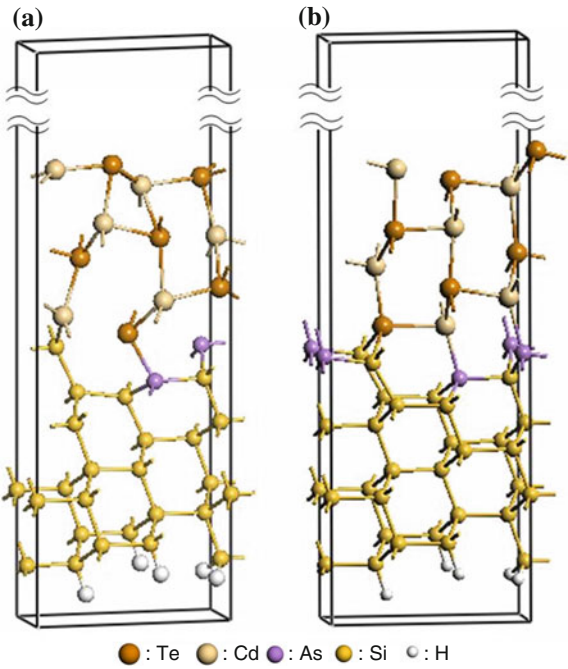


Fig. 3.15 Simulated 2ML CdTe deposited on the As-passivated Si(211) substrates by **a** A-face and **b** B-face epitaxial growth types

partially As-passivated substrate. It is evident that the quality of the epitaxial layer on partially As-passivated substrate is the best (Fig. 3.17e) and the lattice mismatch is less than double epitaxial CdTe layers. From the point of view of valence electron counting, the electron count matches naturally during B-face type epitaxial growth on partially As-passivated substrate. Residual surface states also may play a role in the selective epitaxial growth. It is the same as the situation of CdTe epitaxial growth.

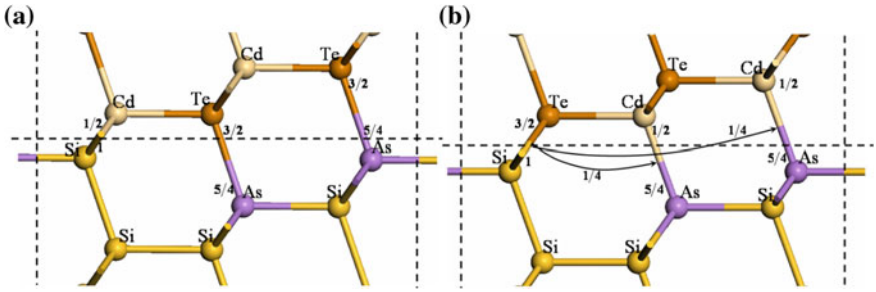


Fig. 3.16 Total electron counting at interfaces of the **a** A-face and **b** B-face CdTe(211) epitaxial layers on the partially As-passivated Si(211) substrates. Reproduced from Ref. [40] with kind permission from Springer Science and Business Media

3.2.3 Arsenic Impurity in MCT

It is well known that structural defects are introduced inevitably in semiconductor materials either in materials growth or in device fabrication. Point defects are important in controlling optical and electronic properties. The most important native defects in $Hg_{1-x}Cd_xTe$ are the mercury vacancy (V_{Hg}) [44, 45]. V_{Hg} will cause Jahn–Teller effect either in certain alloy compositions or in disordered configurations. The mercury vacancy defect in MCT introduces not only a shallow acceptor level but also a deep level at 40 % of the band gap [46, 47]. The activated V_{Hg} impurities are still observed to incorporate as donors, self-compensating the p-type doping [48]. Due to the low electron mobility and a large number of complex deep centers in p-type $Hg_{1-x}Cd_xTe$ with the V_{Hg} [49, 50], attentions have focused on the group V elements, especially arsenic, as the p-type dopant. In the following section, we will study As-doping in detail.

1. Structural Relaxation of Arsenic impurity

One of the most important issues that may limit the applications of semiconductor materials is the difficulty of n- or p-type doping via incorporation of suitable impurities. Using In as-doping imputes, n-type doping is relatively easy to achieve [51, 52]. The doping levels can reach to 10^{18} cm^{-3}. Because it is easy for I group elements, such as Li, Cu, Ag, Au and etc., to substitute Hg element in HgCdTe to achieve p-type doping, I group elements have been widely used as p-type dopants at early stage. However, I group elements are unable to meet the requirements of the growth process for the development of focal plane arrays [53]. Because high diffusivity of I group elements, the doped material is not stable [54, 55]. Arsenic is a group V element that has received great attention in recent years [56]. Due to its low diffusivity [57] in $Hg_{1-x}Cd_xTe$ compared with the native acceptor and the group I elements, attentions have focused on the group V elements, especially arsenic, as the p-type dopant [58]. The experimental results have indicated that arsenic can provide a shallow acceptor level in $Hg_{1-x}Cd_xTe$ grown by bulk method, LPE,

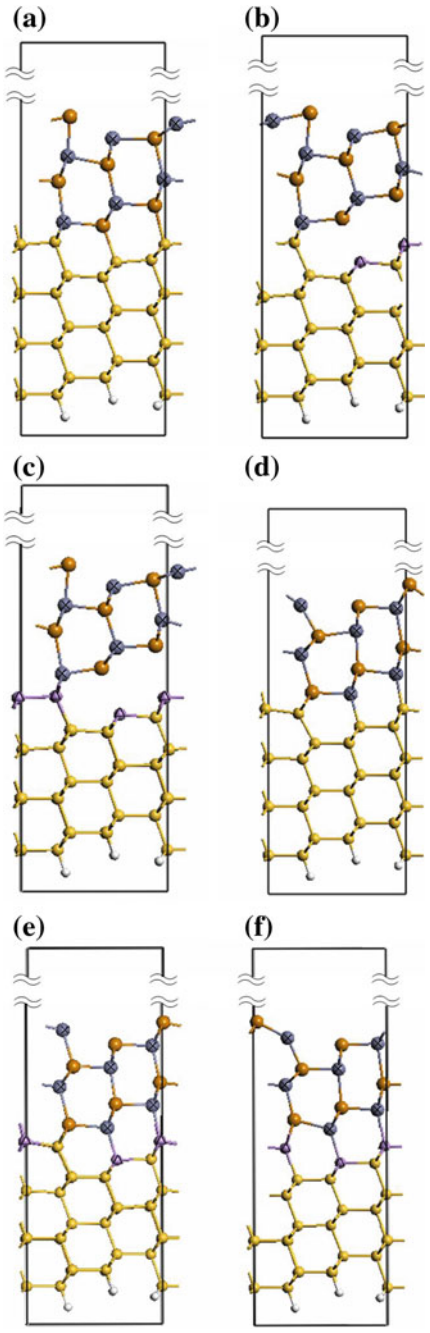


Fig. 3.17 ZnTe epitaxial growth structures on different As-passivated Si(211) substrates. Simulated 2ML ZnTe deposited on the **a** clean, **b** partially As-passivated, and **c** entirely As-passivated Si(211) substrates by A-face type epitaxial growth. Simulated 2ML ZnTe deposited on the **d** clean, **e** partially As-passivated, and **f** entirely As-passivated Si(211) substrates by B-face type epitaxial growth

molecular beam epitaxy, and metal organic chemical vapor deposition. At the same time, the amphoteric behavior of arsenic in HgCdTe has been determined [59, 60]. Arsenic behaves as a p-type dopant under Hg-rich growth conditions and behaves as an n-type or inactive dopant under Te-rich growth conditions [61]. Although MBE is done at much lower temperatures than LPE, and thus offers some distinct advantages, the group V impurities are still observed to incorporate as donors, self-compensating the p-type doping [62, 63]. Research has shown that there is a significant fraction of the arsenic residing on the cation sublattice even under cation-saturated conditions. According to the arsenic dopant incorporation model and the quasi-chemical predictions, some of the doped arsenics in MCT grown by MBE are incorporated on the cation sublattice as isolated defects, As_{Hg}, unbound to vacancies, as donors. So, an annealing process is required for its electrical activation [68–71]. After activation, there still exist As_{Hg} donor impurities in HgCdTe [68]. However, as the doping levels reach 5×10^{18} cm^{-3}, the active efficiency of the p-type arsenic doping drastically drops due to the strong self-compensation effect [69]. The amphoteric behavior of the arsenic in situ impurities and bonding mechanism in MCT is a typical prototype to systematically study the technology requirements of infrared focal planes.

We considered a $2 \times 2 \times 2$ supercell (SC) with a total of 64 atoms as the calculation model. The doped system was modeled by putting one As atom at the center of a periodic SC. A satisfactory self-consistent convergence has been achieved by considering a number of FP-LAPW basis functions [70] up to $R_{MT}K_{max} = 7.0$, where the muffin-tin (MT) radius of Te, Hg, Cd, and As is 2.70, 2.65, 2.65, and 2.60 atomic unit (a.u.) separately. We adopted the Perdew–Burke–Ernzerhof function [71] to describe the exchange–correlation interaction. The relativistic effect of spin-orbit (SO) coupling was also included. Self-consistent iteration was considered to be converged when both the total energy are stable within 10^{-4} eV per unit cell. The relaxation procedures were conducted following the damped Newton dynamics schemes. The criterion of the force convergence for all atoms was 0.05 eV/Å. The plane-wave cutoff energy was chosen as 300 eV in the calculation by PPW method [72]. Standard ultrasoft pseudopotentials [73] are employed for all the atoms, and the Brillouin zone was sampled by using $8 \times 8 \times 8$ Monkhorst-Pack grid [74]. The energy convergent criterion was 10^{-5} eV per unit cell, and forces on all relaxed atoms are less than 0.03 eV/Å.

Both theoretical and experimental results show that As may reside on either the metallic sublattice behaving as a donor or the nonmetallic sublattice acting as an acceptor. It is well known that impurities induce the atomic structural relaxation in the host and modify the electronic structure of the system. The relaxations of As_{Hg} impurity calculated by using the FP-LAPW and VASP, respectively, are shown in Figs. 3.18 and 3.19. The relaxing results are listed in Tables 3.4 and 3.5. From the numerical results using two different methods, we can find that:

A: The As impurity, substituting Te, lead to inward relaxations of the nearest neighbor (NN) host atoms around the impurity. In contrast, the NNN cations show outward relaxation.

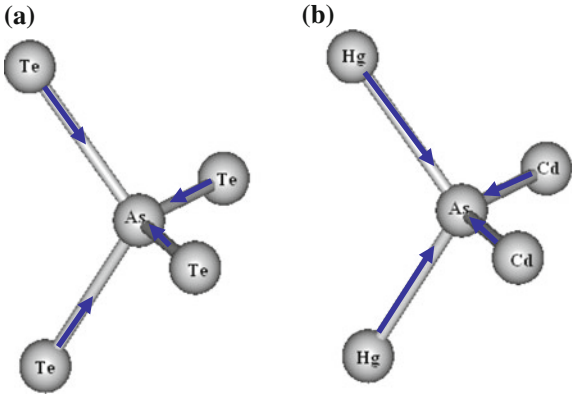


Fig. 3.18 The diagram of the atomic relaxation of the NN tellurium atoms around the **a** As_{Hg} and **b** As_{Te}. Reprinted with permission from Ref. [75]. Copyright 2005 by American Physical Society

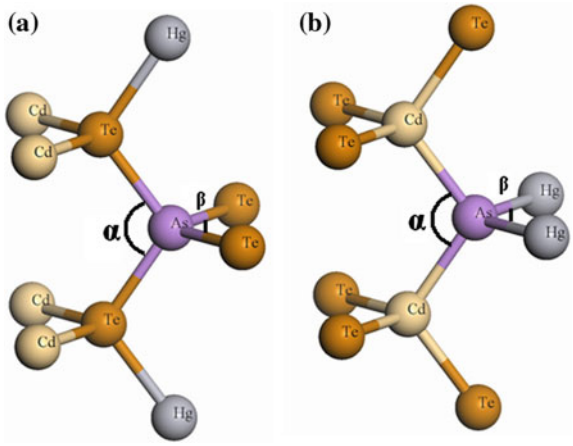


Fig. 3.19 The diagram of the bond angle relaxation of the NN tellurium atoms around the **a** As_{Hg} and **b** As_{Te}. Reprinted with permission from Ref. [75]. Copyright 2005 by American Physical Society

B: The relaxation of the bond angles around the impurity indicates that the inhomogenous cation configuration around the NN tellurium makes it relax along the arrow as shown in Fig. 3.18a, which produces an inverse change of the bond angles α and β ($\alpha > \beta$).
C: the bond angle of the NN Cd atoms and the NN Te atoms becomes smaller with respect to the ideal tetrahedral bond angle.

There are two reasons for the inward relaxation of the NN Te atoms in As_{Hg} impurity. The smaller radius of As than that of Hg (As radius is 1.18 Å and that of Hg is 1.48 Å [76]) causes the NN Te to relax inwardly, and the stronger covalent bonding also results in NN Te inward relaxation. The results of bond relaxation of As_{Hg} show asymmetrical distributing of the NNN cations of As impurity which cause tetragonal distortion of As_{Hg}. The atomic symmetry of As and the NNN cation is reduced from T_d to D_{2d}, namely the Jahn–Teller effect [77]. The Jahn–Teller effect will lead to the removal of degeneracy of t_2^v energy level in conduction band and will reduce energy of t_x state to form a shallow donor level.

Table 3.4 The relaxing results of As_{Hg} impurity in $Hg_{0.5}Cd_{0.5}Te$

	As–Te^1	As–Cd^2	As–Hg^2	Te^1–Hg^2	Te^1–Cd^2	Te^3–Cd^2	Te^3–Hg^2
Before relax	2.831	4.624	4.624	2.831	2.831	2.831	2.831
After relax[a]	2.826	4.638	4.649	2.856	2.863	2.818	2.835
After relax[b]	2.722	4.605	4.694	2.861	2.896	2.826	2.832
Varity[a]	−0.010	0.013	0.021	0.025	0.028	−0.013	0.007
Varity[b]	−0.108	−0.019	0.07	0.03	0.065	−0.005	0.001
Change rate[a] (%)	−0.37	0.28	0.47	0.89	0.99	−0.48	0.28
Change rate[b] (%)	−3.8	−0.41	1.5	1.05	2.29	−0.17	0.04

The superscript denotes nearest neighbor number of the doping atoms. Negative relaxation results indicated contraction relaxation to the dopant atoms

[a, b]WIEN2K and VASP relaxation results separately (Å)

Table 3.5 The bond angle relaxing results of As_{Hg} impurity in $Hg_{0.5}Cd_{0.5}Te$

	Cd^2–Te^1–Cd^2	Cd^2–Te^1–Hg^2	α(Te^1–As–Te^1)	β(Te^1–As–Te^1)
Before relax	109.471°	109.471°	109.471°	109.471°
After relax[a]	109.307°	109.632°	109.889°	109.261°
After relax[b]	108.249°	110.461°	111.350°	108.539°
Varity[a]	−0.164°	0.161°	0.418°	−0.21°
Varity[b]	−1.222°	0.99°	1.879°	−0.932°
Change rate[a] (%)	−0.15	−0.14	0.38	−0.19
Change rate[b] (%)	−1.11	0.9	1.71	−0.85

[a, b]WIEN2K and VASP relaxation results separately

Negative relaxation results indicated decreases of bond angles

It is worth noting that relaxations of As_{Hg} have been calculated by both the FP-LAPW method and the conjugate gradient method. The trend of the results obtained from the two different methods is consistent with each other, especially the inward relaxation and Jahn–Teller distortion of the NN Te atoms in As_{Hg}. However, the extent of relaxation obtained by the VASP method is larger than the FP-LAPW method. The main reason for this phenomenon is the different algorithmical methods in these two methods, different convergence accuracy and different constraints of symmetry in the calculation systems. Taking into account the large amount of computation, we used coarser convergence accuracy and keep the P222 symmetry in the FP-LAPW calculations. From the above results, the same relaxation trend is the most important aspect in the material simulation. This also shows the applicability of both methods for simulation of infrared optoelectronic materials. By comparing these calculations, we can get more reliable results. The most important is that one can have a reasonable choice of different ways to complete the

calculation and analysis according to the computational requirements and the actual situation.

The relaxations of As_{Te} impurity calculated by using the VASP are shown in Figs. 3.10b and 3.11b. Th relaxation of AsTe impurity in Hg0.5Cd0.5Te is listed in, respectively, Tables 3.6 and 3.7. From the numerical results, we can find that:

(1) Large bond relaxations of the NNN cation are caused by As_{Te}. As–Hg^1 is 8 % shorter than unrelaxed structure. As–Cd^1 is 7 % shorter than unrelaxed structure.
(2) The relaxations of As_{Te} show Jahn–Teller distortion. The atomic symmetry of As and the NNN cation is reduced from T_d to D_{2d}.

Using large-size Si substrates for growth of HgCdTe has many advantages, such as large available size, lower cost, superior mechanical strength, and increased thermal conductivity [78–81]. However, the main problem of using Si substrates is the $\sim$19 % lattice mismatch between HgCdTe and Si, which will result in dislocation densities $\sim 10^6$ cm^{-2}. It has been demonstrated experimentally that the interface between HgCdTe and Si can be greatly improved when the Si surface is passivated with As. From the above relaxation calculations of As_{Te}, the bond length between As and the NNN cation is shortened which will reduce the mismatch between Si and HgCdTe material. This result also gives us an idea to solve the lattice mismatch problem in Si/HgCdTe materials. Looking for an element which can reduce the HgCdTe bond length and increase the bond length between Si and impurities may solve the problem of lattice mismatch. The calculation results by CASTEP finds that the bond length between Si and impurities are increased on certain Si surface, such as Si(211) surface [82].

Table 3.6 Relaxation of As_{Te} impurity in $Hg_{0.5}Cd_{0.5}Te$

	As–Hg^1	As–Cd^1	Te^2–Hg^1	Te^2–Hg^1	Te^2–Cd^1	Te^2–Cd^1
Before relax	2.831	2.831	2.831	2.831	2.831	2.831
After relax	2.601	2.641	2.926	2.921	2.951	2.953
Difference	−0.230	−0.190	0.095	0.09	0.120	0.122
Change rate (%)	−8.12	−6.71	3.35	3.17	4.23	4.30

The superscript denotes nearest neighbor number of the doping atoms
Negative relaxation results indicated contraction relaxation to the dopant atoms

Table 3.7 The bond angle relaxing results of As_{Hg} impurity in $Hg_{0.5}Cd_{0.5}Te$

	Hg^1–As–Hg^1	Cd^1–As–Cd^1	Hg^1–As–Cd^1	Te^2–Hg^1–Te^2	Te^2–Cd^1–Te^2
Before relax	109.471°	109.471°	109.471°	109.471°	109.471°
After relax	101.131°	116.628°	109.490°	102.044°	107.031°
Varity	−0.164°	0.161°	0.418°	−7.427°	−2.44°
Change rate (%)	−1.11	0.9	1.71	−6.78	−2.22

Negative relaxation results indicated decreases of bond angles

2. As impurity clusters

Arsenic (As), due to its low diffusivity, is employed as a most desired p-type dopant [83, 84]. When the As-doping concentration is about 10^{16}–10^{18} cm^{-3}, the activation efficiency can reach to 100 %. Success in achieving p-type doping of HgCdTe using arsenic has been reported by several groups [85–88]. However, arsenic is observed to be incorporated as donor (which is strongly compensated by unknown defects) rather than acceptor as desired, therefore necessitating post-growth annealing to activate the arsenic acceptor. Experimental evidence is not yet available as to the chemical nature of the arsenic impurity.

Major efforts in the last few years have been made experimentally to search for effective activation scheme; nevertheless, the nature and origin of the As impurities in as-grown materials remains the least understood, which is a crucial step toward controlling the conversion of arsenic from a donor to an acceptor. Berding et al. originally proposed that As_{Hg}–V_{Hg} pair is the starting defect for the activation of As acceptor [89, 90]. Because MBE growth of HgCdTe is normally restricted to the Te-rich region, there is a high density of V_{Hg} in HgCdTe and most As atoms (99 %) are likely to be incorporated into Hg vacancies. They also stated that half of As_{Hg} prefer to bind to V_{Hg} as As_{Hg}–V_{Hg} pairs at the same time. As_{Hg} behaves as a strongly compensated donor, rather than an acceptor as desired because of V_{Hg} nearby. It is thus suggested that the shallow acceptors V_{Hg} and complexes associated with As_{Hg} and V_{Hg} are likely to be responsible for the observed compensation in As-doped HgCdTe. This model allows consistent interpretation of experimental observations on arsenic doping of HgCdTe but is not comprehensive. First, the probability of being interstitial atoms is much larger than that of being substitutional atoms for As because of smaller atomic radius and lower surfactivity. Second, electrical measurements have concluded the n-type carrier compensation in as-grown As-doped HgCdTe. Among the proposed candidates are mercury vacancies and arsenic clusters. Third, complete As activation can be achieved in Hg-saturated ambient even with a low Hg pressure, but the activation efficiency is less than 3 % under vacuum [91]. This suggests that for the activation of As, Hg plays a vital role during annealing, while the role of V_{Hg} is very limited. Furthermore, it has been found by Selamet et al. that the low carrier mobility of the native sample is caused by As_4 cluster scattering centers [92]. The double-deep-level spectra found in PL measurements was also attribute to As_4 cluster [93]. Taking into account the limitations of the (As_{Hg}–V_{Hg}) single-atom model, Sivananthan group proposed an As molecular adsorption model. In this model, an arsenic source in a large number of As_4 or As_2 molecules and small quantity of singular arsenic atoms has been proposed. It is also suggested that the arsenic flux is predominantly in the form of As_4 or As_2 molecules [94] under the typical low-temperature conditions of MBE growth.

The doping characteristics of V_{Hg}, As_{Hg}, and As_{Te} have been studied systematically [75, 95]. In the next section, the structures of As_4 and As_2 clusters will be investigated based on the As molecular adsorption model.

(1) Structure models of As impurity clusters and computational methods

We chose HgTe as the host system. The random distributions of cation atoms in HgCdTe system are not considered. The choice of HgTe is feasible because of the equivalent properties between HgTe and technologically important HgCdTe with low Cd fraction. The defect system was modeled by putting a complex defect at the center of a periodic SC of HgTe, and all the defect structures considered here are electronically neutral.

The total energy calculations were performed within the generalized-gradient approximation (GGA) [96] implemented by the plane-wave total energy VASP code with the Vanderbilt ultrasoft pseudopotentials [97]. The cutoff energy for the wave functions was 280 eV. The calculated ground lattice constant for bulk HgTe was 6.66 Å, in agreement with the experimental value 6.46 Å [98]. We used a 64-atom SC and a $9 \times 9 \times 9$ k-points Monkhorst-pack mesh in the Brillouin zone of the SC. All internal structural parameters were fully relaxed until the Hellmann–Feynman (HF) forces were converged to 0.1 meV/Å.

(2) Chemical Potential Dependence of the Defect Formation Energy

Generally speaking, the defect formation energy includes the contribution from three parts. One is free energy variation between the defective system and a perfect system. The other is the chemical variation caused by the partial (include atomic and electron) exchange between the host and the chemical reservoirs when defects were introduced.

During the formation of As_{Hg} impurity in HgTe system (Fig. 3.20), one Hg atom will be removed from the host to the chemical reservoirs of Hg atoms, and at the same time, one As atom will be added into HgTe system from the chemical reservoirs of As atom. If the defective system is in charge state +q, one electron should be

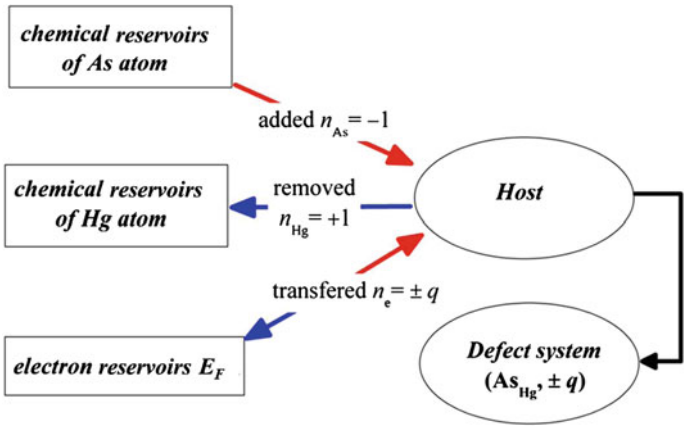


Fig. 3.20 Schematic diagram of chemical potential variation due to the formation of AsHg impurity in HgTe system

transferred from the neutral defective system to electron reservoirs (correspond to Fermi level). In this process, the chemical potential of the system will be changed.

The formation energies of the defect system were calculated from the total energies based on the standard Eq. (3.1) [99]. For neutral defects $\alpha(q = 0)$ in HgTe, the formation energy $\Delta H_f(\alpha, q = 0)$ depends on the chemical potential μ_i (i refers to Hg, Te, and As).

$$\Delta H_f(\alpha, q = 0) = E_{\text{tot}}(\alpha, q = 0) - E_{\text{tot}}(\text{HgTe}) + \sum_i n_i(\mu_i^0 + \mu_i) \tag{3.1}$$

Here, $E_{\text{tot}}(\alpha, q = 0)$ is the total energy for the defective SC, E_{tot}(HgTe) is the total energy for the same SC in the absence of α, the n_i is the number of Hg, Te, and As atoms transferred from the SC to the reservoirs in forming the defect, respectively, μ_i is the chemical potential referenced to elemental liquid/solid with energy μ_i^0. However, the achievable values of μ_i are limited thermodynamically under equilibrium growth conditions. First, to avoid precipitation of the constituent i, μ_i is bound by

$$\mu_i < 0 \tag{3.2}$$

Second, to maintain a stable HgTe compound, μ_i is bound by

$$\mu_{\text{Hg}} + \mu_{\text{Te}} = \Delta H_f(\text{HgTe}) \tag{3.3}$$

The calculated ΔH_f (HgTe) −3.65 eV is in agreement with the experimental value −3.3 eV. Finally, to avoid the formation of secondary phases As_2Te_3, μ_i is bound by

$$2\mu_{\text{As}} + 3\mu_{\text{Te}} < \Delta H_f(\text{As}_2\text{Te}_3) \tag{3.4}$$

Figure 3.21 plots the accessible chemical potential region for HgTe:As, as defined by Eqs. (3.2)–(3.4). It is shown that because arsenic and tellurium form a very stable compound As_2Te_3 with a rather low formation energy −3.65 eV, the highest possible μ_{As}, at the Hg-rich condition ($\mu_{\text{Hg}} = 0$, $\mu_{\text{Te}} = -3.65$ eV), is −3.85 eV. Under the Te-rich condition ($\mu_{\text{Hg}} = -3.65$ eV, $\mu_{\text{Te}} = 0$), μ_{As} is further reduced to less than −9.33 eV. Above these chemical potential limits, secondary As_2Te_3 compound will be formed, thus stopping the arsenic doping process. Such low-accessible μ_{As} is one of the limiting factors for arsenic doping in HgCdTe.

(3) Structures of As_n (n = 1, 2, 4) clusters

① Arsenic Tetramers

Under Te-saturated condition, the density of V_{Hg} is much higher than that of tellurium vacancy (V_{Te}), so that the As_4 defects with two of the four arsenic atoms occupying mercury sites are expected to be the most reasonable configuration. In our calculations, we typified such As_4 defects into arsenic tetramers (As_t) and isolated arsenic atoms (4 As), respectively (see Fig. 4.2).

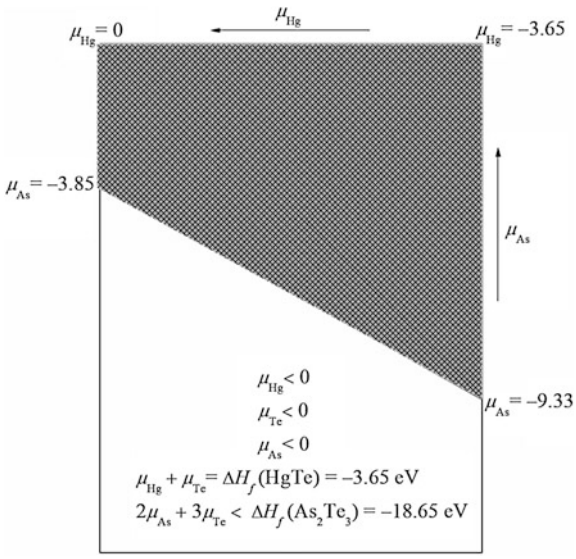


Fig. 3.21 Calculated available equilibrium chemical potential region for HgTe: As in the two-dimensional (μ_{Hg} and μ_{As}) plane. The shaded and diagonal area is forbidden under equilibrium growth condition. Reproduced from Ref. [100] with kind permission from Springer Science and Business Media

As_t can be regarded as the distorted pyramidal structure of As_4 molecules. During typical low-temperature MBE growth, there are difficulties in cracking As_4 into As_2 or As_1 on the growing surface, so that As_4 molecules initially absorbed on the surface are subjected to the stress exerted by the HgTe lattice owing to the different properties of the dopant and host atoms. The stress can be released at the cost of breaking the partial As–As bonds. In this way, the pyramidal structure distorts into the tetramer configuration.

For As_t shown in Fig. 3.22, the optimized As–As split changes from 2.45 Å for configuration (c) to 2.555 Å for configuration (a), approximately conserving the experimental value 2.435 Å for the pyramidal As_4 molecular [74] (Fig. 3.23).

One can also see that different occupation patterns for As_t produce dissimilar numbers of the broken As–As bonds (designated as δ), and the values of δ for all the possible As_t defects do not exceed 3 ($\delta \leq 3$). This can be understood from the bond length of 2.435 Å for the pyramidal structure [101], slightly less than the Hg–Te distance of 2.798 Å but far less than the spacing of the nearest neighboring Hg atoms in HgTe (4.57 Å). As a result, the As_{Hg}–As_{Hg} bond is to be broken up while the As_{Hg}–As_{Te} bond is left to adjust itself to accommodate the host lattice. In the same way, it is conjectured that more As_{Hg} units in the As_4 defects leads to the larger value of δ. Figure 3.22 also gives the calculated formation energies of the As_4 defects at the Te-rich limit. In the case of As_t, the configuration (a) and (b) are metastable states from the formation energies, and the configuration (c) is the most probable one. Therefore, under the equilibrium stage the former two configurations are to be converted into the latter one by means of extra V_{Hg} to release the lattice stress. The geometry of the configuration (c), optimized in a manner of $[(As_{Hg}–As_i)_d]_3$ copolymer (dimer denoted as "d"), is not only more favorable in energy by

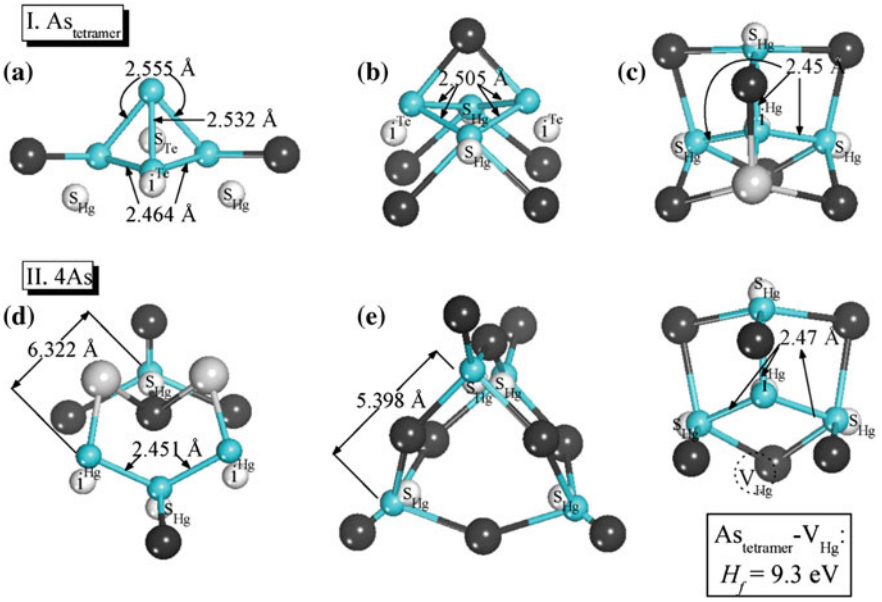


Fig. 3.22 Calculated ground geometries for the As_4 defects in HgTe. In this figure, the Hg atoms are of the largest size (in *light grey*), the Te atoms of intermediate size (in *black*), and the As atoms are the smallest (in *cyan*). To analyze the atomic displacements, the As atoms in the pre-relaxed structures are also shown by the smallest atoms (in *white*). For the labels on the As atoms, "s" and "i" denotes the substitutional site and the interstitial site, respectively. Reproduced from Ref. [100] with kind permission from Springer Science and Business Media

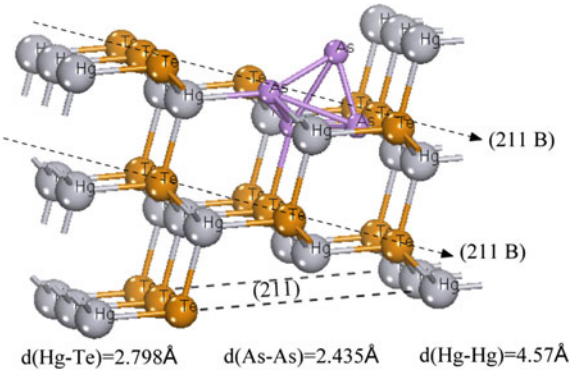


Fig. 3.23 Adsorption of As_4 cluster on Si(211) substrate

2 eV lower than the strongest configuration (b), but also stronger with 25 % gain in the binding energy with respect to the quasi pyramidal configuration (a).

When the As–As bonds for As_4 molecular are broken completely ($\delta = 6$) or partially ($3 < \delta < 6$), the perfect pyramidal dissociates into isolated atoms and/or dimers, as demonstrated by the configurations of type II in Fig. 3.22. Such a

situation occurs only under high-temperature growth conditions, which is not applied for the state-of-the-art MBE. Focusing on the configuration (c), one can see that the atomic relaxation makes the three As_{Hg} atoms to move toward the common As_i atom by an amplitude of 0.348 Å, which leads to the final optimized geometry in $[(As_{Hg}–As_i)_d]_3$ copolymer configuration. However, such a configuration fails to explain the experimental observation of the slight n-type conduction in as-grown materials because of the three As_{Hg} atoms with strong donor characteristics. The previous study [90] has given a reasonable explanation of the $As_{Hg}–V_{Hg}$ pairs, suggesting that the behavior of the compensation conduction in as-grown samples is possibly associated with V_{Hg} which is bound to the donor-like arsenic defects. To augment our understanding of such a suggestion, we computed the complex defect of $As_t–V_{Hg}$ (see the last section in Fig. 3.22), which was modeled by placing an additional V_{Hg} to the given As_t in the configuration (c). The calculated ground geometry of $As_t–V_{Hg}$ displays that the configuration of $[(As_{Hg}–As_i)_d]_3$ is stable, without experiencing the relaxation to other configurations with the perturbation of V_{Hg}. However, the complex defect of $[(As_{Hg}–As_i)_d]_3–V_{Hg}$ is found to substantially lower the formation energy $[(As_{Hg}–As_i)_d]_3$ by about 3 eV. This reduction of the formation energy is attributed to the removing of Hg atoms from the defect system, thus effectively compensating the energy lost by the introduction of As atoms into the defect system. Also, the configuration information of the complex defect of $[(As_{Hg}–As_i)_d]_3–V_{Hg}$ is deemed to be in analogy to that of the clusters of $As_{Hg}–V_{Hg}$ pairs with one common V_{Hg}, so that both of them are of compensating donor nature, as supported by our calculated transition energies (unpublished).

② Structures of As_2 and As atom

The calculations of As_d and As_s are based on the fact that these defects in a small quantity are present in a noncracking arsenic source but predominant in a cracking arsenic source.

In Fig. 3.24a, we show the probable configuration of As_d with the lower formation energy. The As–As spacing is 2.228 Å, in agreement with the experimental value 2.10 Å for As_2 molecule [101]. We note that the As_{Hg} atom deviates significantly from the ideal lattice sites toward the interstitial site by the amplitude of 0.783 Å, which suggests that the interstitial sites are favorable in energy to the absorption of As_2 molecules. However, our calculations indicate that two arsenic atoms, both residing in the interstitial sites, can never be configured in As_d unless a number of energy barriers are overcome when an extra V_{Hg} is available; otherwise, they are in the configuration of two isolated arsenic atoms (2As), as shown in Fig. 3.24b, c. So the configuration with two arsenic atoms of split interstitial is expected to be the most probable for As_d, but this assumption needs to be examined in our further work. At the same time, we have modeled the complex defect ($As_{Hg}–As_{i)d}–V_{Hg}$), as demonstrated by configuration (f) in Fig. 3.25. An obvious shift of As_i toward V_{Hg} can be observed, and the amplitude of the As_i relaxation is more significant relative to the $(As_{Hg}–As_i)_d$ defect.

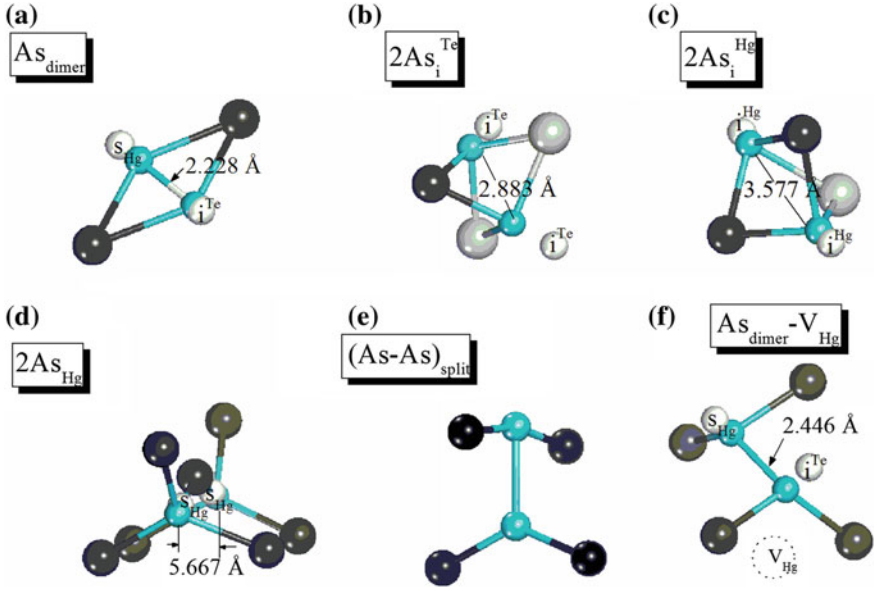


Fig. 3.24 Calculated ground geometries for As_2 defects in HgTe. The configuration of (As–As) split is assumed to be more plausible than other ones for As_d. The atomic notations are the same as denoted in Fig. 3.23. Reproduced from Ref. [11] with kind permission from Springer Science and Business Media

There are four probable configurations of As impurities when an As atom is in different substituted sites [structure (a) and (d)] or interstitial sites [structure (b) and (c)] as shown in Fig. 3.25. The displacements of As are negligible during the atomic relaxations. The lattice relaxation energies are of the order of meV.

During the short-displacement and successive steps of the relaxations, the work done by the HF force can be approximated to be that by the first (1) and second (2) nearest neighbors of impurity while the contribution from the more distant neighbors is not taken into account.

$$W^{\mathrm{HF}} = W_1^{\mathrm{HF}} + W_2^{\mathrm{HF}} = \left(\sum F_{i0}^{\mathrm{HF}} \cdot \Delta d_{i0} + \sum F_{ij}^{\mathrm{HF}} \cdot \Delta d_{ij}\right) \Big/ 2 \tag{3.5}$$

Table 3.8 displays the evaluation of the work done by the HF force together with the relaxation energies $\Delta E_{\mathrm{relax}}$. One can see that the sum of W_1^{HF} and W_2^{HF} is comparable to the value of $\Delta E_{\mathrm{relax}}$, and the main contributions of the work done by the HF forces come from the second-nearest neighbors. In addition, the value of $\Delta E_{\mathrm{relax}}$ in As_{Hg} system is much less than that in other systems because the matchable electronegativities between As atoms and the NN Te atoms lead them to bond easily.

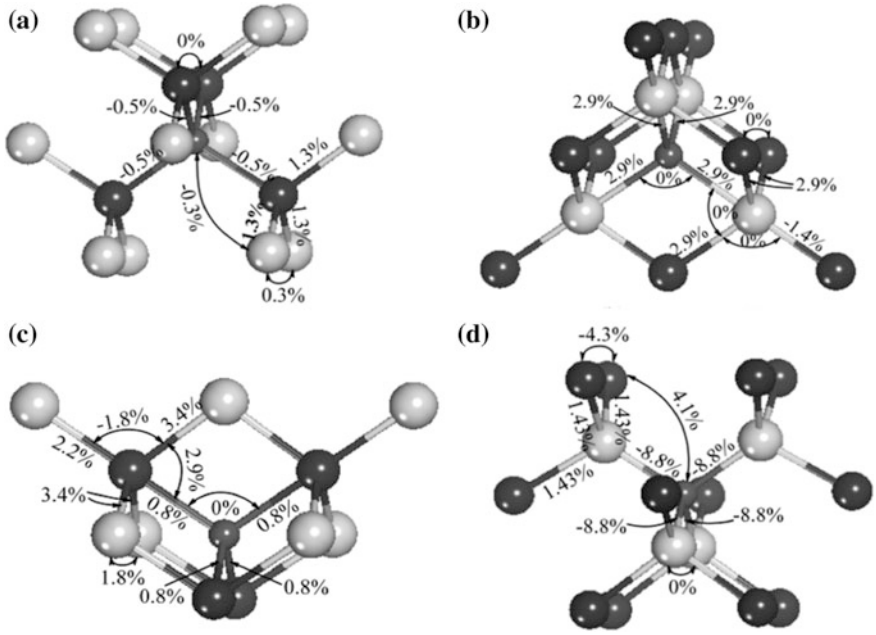


Fig. 3.25 Calculated ground geometries for the As atom defects in HgTe

Table 3.8 Work of the Hellmann–Feynman forces (HF) (in meV) for the first (1) and second (2) nearest neighbors as well the lattice relaxation energies ΔE_{relax} (in meV)

(α, $q = 0$)	W_1^{HF}	W_2^{HF}	ΔE_{relax}
As_{Hg}	1.68	15.84	19.58
As_{Te}	143.96	530.24	709.67
AS_i^{Hg}	27.76	92.13	339.97
AS_i^{Te}	3.54	90.76	366.81

Reproduced from Ref. [100] with kind permission from Springer Science and Business Media

(3) Comparison of the Formation Energies for As_n (n = 1, 2, and 4)

In Fig. 3.26 we compare the formation energies per arsenic atom for various types of the arsenic defects As_n (n = 1, 2, and 4). It can be seen that As_t is more favorable in energy than As_d and As_s, indicating that As_t is more abundant than As_d and As_s in as-grown materials. The order of the formation energies is relevant to the atomic site configurations in arsenic defects, for example, the energy levels of the arsenic defects containing the As_{Te} atoms are much higher than those containing the As_{Hg} atoms. This is a direct effect of the chemical potentials on the formation energies. Under Te-rich condition ($\mu_{Te} = 0$ and $\mu_{Hg} < 0$), more Hg atoms involved in the given defect results in the lower energy levels. From the formation energies of As_n, the values of the binding energies are also obtained for As_t and As_d. We find that the

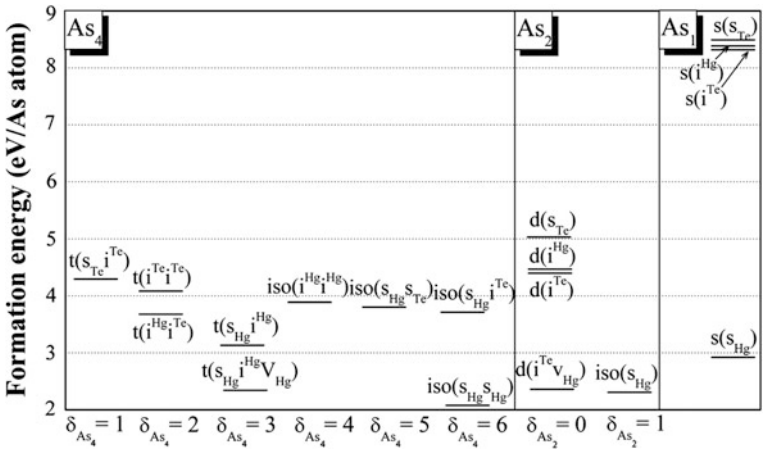


Fig. 3.26 Formation energy of As_n (n = 1, 2, 4) impurities

more stable configuration corresponds to the stronger binding of the arsenic atoms in As_t and As. Since the distorted pyramidal structure of As_4 molecular relieves the lattice stress during MBE growth, the heavier degree of the distortion yields more amount of the lattice stress to be released and thus more steady configurations as well. Moreover, it can be concluded that under thermal equilibrium isolated arsenic atoms prefer to be bound into two or four nearest neighboring atoms, because the binding energies of 2As and 4As defects are at least 0.61 eV/As atom and 0.84 eV/As atom higher than the substitutional singlet (As_{Hg}), respectively.

3.2.4 Amphoteric Doping Behavior of as in MCT

1. Activation model of p-type doping

(1) Electronic properties of As_{Hg} in MCT

In order to understand the bonding mechanism of arsenic doping in $Hg_{0.5}Cd_{0.5}Te$, the valence charge density and the bonding-charge density have been calculated. Figures 3.27 and 3.28 show the valence charge density and the bonding-charge density on the (110) plane, respectively. From these two figures, we can see that the charge density distribution between the arsenic impurity and NN tellurium atoms shows the covalent characteristic. The valence charge density along the As–Te bond line has been picked up as shown in Fig. 3.29, which indicates that it is a nearly perfect covalent bond. Moreover, the smaller atomic radius of arsenic than that of mercury causes the NN tellurium to relax inwardly; the stronger covalent bonding also results in NN Te inward relaxation.

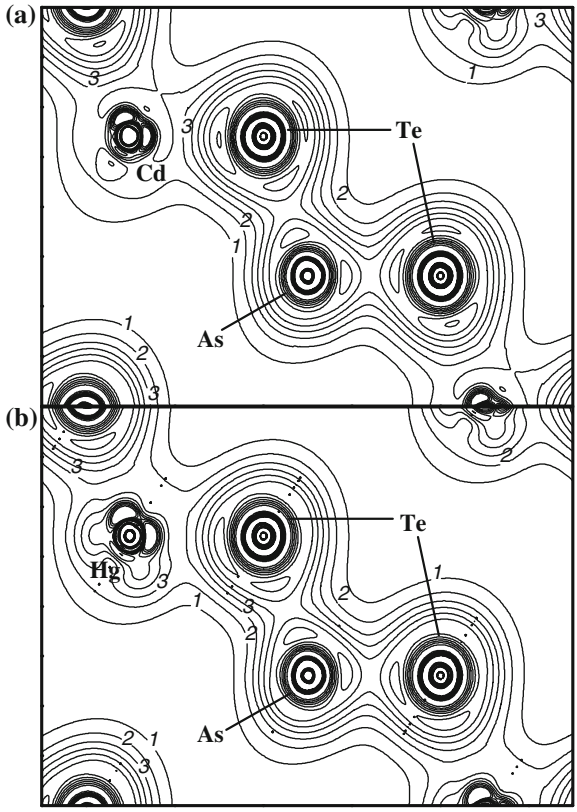


Fig. 3.27 The total charge density in the (110) plane of $Hg_{0.5}Cd_{0.5}Te$ with As_{Hg}, where the contour step size is 6×10^{-3}e/(a.u.)^3. Reprinted with permission from Ref. [75]. Copyright 2005 by American Physical Society

It is instructive to compare the total valence charge of the doped materials with that of the undoped one for the corresponding site ions with the same size MT spheres which is the basis for the comparison in a different compound. The charge-transfer results, listed in Table 3.9, indicate that the impurity influences not only the charge redistribution of the nearest neighbors, but also that of the NNN cations.

The DOS of undoped and doped $Hg_{0.5}Cd_{0.5}Te_{0.5}$ are shown in Fig. 3.30a, b, respectively. In comparison with the undoped materials, the valence bandwidth of the As-doped materials is broadened about 176 meV due to the two extra As 4*p* bonding states coupling into the valence band. Substitution of mercury by arsenic makes the As-doped MCT metallic because an electron of the As-*p* state fills up to the conduction band. The Fermi level does not lie within the band gap, but it extends to the conduction band, as shown in Fig. 3.30b. Following the principle discussed in Ref. [24], for simple extrinsic impurities, in principle, one can predict whether a dopant is a donor or an acceptor by simply counting the number of the valence electrons of the dopant and the host elements. Here, the integral of the density of states from CBM to the Fermi level accommodates one electron

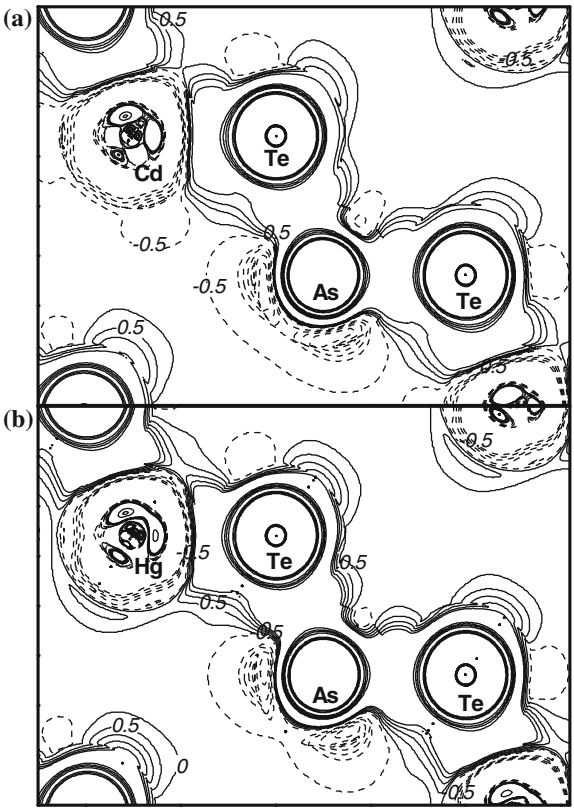


Fig. 3.28 The bonding-charge density in the (110) plane of $Hg_{0.5}Cd_{0.5}Te$ with As_{Hg}, where the contour step size is 5×10^{-3}e/(a.u.)^3. Reprinted with permission from Ref. [75]. Copyright 2005 by American Physical Society

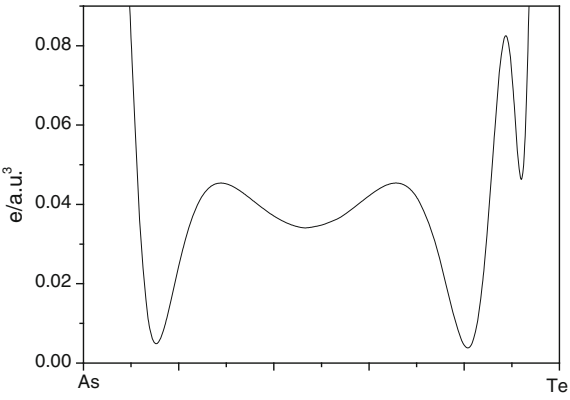


Fig. 3.29 The total charge density along the bond line of As–Te in the (110) plane of $Hg_{0.5}Cd_{0.5}Te$ with As_{Hg}. Reprinted with permission from Ref. [75]. Copyright 2005 by American Physical Society

occupying the t_2^c state. The energy state behaves as a single donor. The results agree with the theoretical and experimental findings, where the impurity As_{Hg} in $Hg_{0.5}Cd_{0.5}Te_{0.5}$ behaves as a single donor.

Table 3.9 Charge transfer for the As_{Hg} case and average charge transfer in the atomic sphere for the undoped case

	$Hg_{0.5}Cd_{0.5}Te$–As_{Hg}			$Hg_{0.5}Cd_{0.5}Te$		
	$Q_{crystal}$	Q_{atomic}	ΔQ	$Q_{crystal}$	Q_{atomic}	ΔQ
As	31.699	31.469	0.230	–	–	–
Te	49.958	49.730	0.228	49.974	49.730	0.217
Cd	46.628	46.541	0.087	46.576	46.541	0.035
Hg	78.466	78.398	0.068	78.424	78.398	0.026

Subscripts "crystal" and "atomic" mean the crystal charge and the superposition of atomic charge, respectively. $\Delta Q = Q_{crystal} - Q_{atomic}$

Reprinted with permission from Ref. [75]. Copyright 2005 by American Physical Society

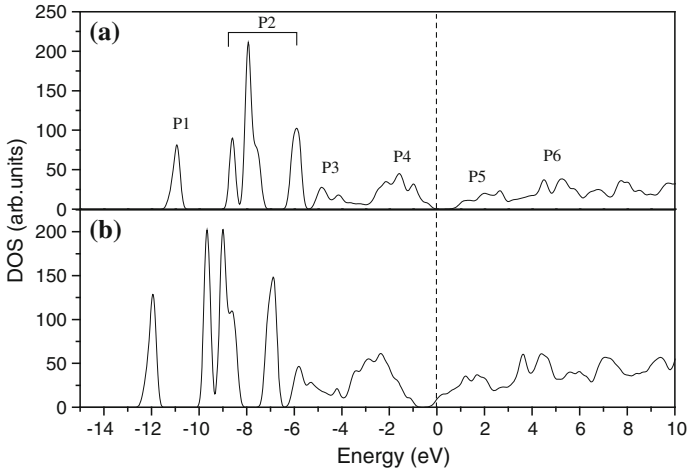


Fig. 3.30 DOS in $Hg_{0.5}Cd_{0.5}Te$ **a** without and **b** with As_{Hg}. Reprinted with permission from Ref. [75]. Copyright 2005 by American Physical Society

It is well known that the calculation in terms of the LDA approximation cannot give the accurate band gap, as does the absolute position of the donor level for the present case. But the absolute energy positions calculated by the ab initio method are less significant than the relative ones. The results of the calculated band gaps for undoped and doped cases (Fig. 3.30), respectively, show that the doping causes the band gap to be 19 meV smaller than that of the undoped one. If we neglect the effect of the valence-band broadening on the band gap, the change can be ascribed to the donor state. The donor level can be qualitatively located in the region of less than 19 meV below the CBM. The quantitatively calculated results from the single-particle eigenvalues reveal that the donor level is 9 meV below the CBM, as a shallow donor (Fig. 3.31). The results are in agreement with the experimental findings [102] for the As_{Hg}-doped MCT with stoichiometric $x = 0.39$. In the meantime, the donor level has the *s*-like characteristics as mentioned above, so the

relative position between the donor level and CBM, which is also derived from the *s* states of group II, is accurate to some extent. Based on the relation between the ionization of the donor level and the mole fraction of MCT predicated by experiments [103–105], we can predict that the As_{Hg} donor level will be shallower as the mole fraction becomes smaller. Especially, according to the most technologically important composition x around 0.2, the donor level is about 5 meV.

The coupling strength is determined by the distance between the impurity and the host atoms (Fig. 3.33). Figure 3.32 shows that the effect of the coupling strength between the impurity and host atoms on the conduction-band edge is reduced as the distance becomes larger. By comparing Fig. 3.31a with Fig. 3.31b, we can find that the band edge of As-doped MCT shows more apparent shoulders than that of the perfect MCT. As shown in Fig. 3.33, the difference mainly comes from the coupling of As-*s* and NN Te-*p* at the VBM, and the coupling of As-*p* and the NNN

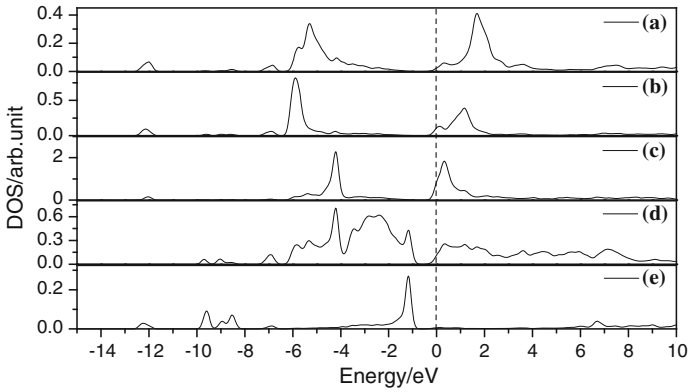


Fig. 3.31 The partial DOS of **a** Cd-s, **b** Hg-s, **c** As-p, **d** Te-p, and **e** As-s states

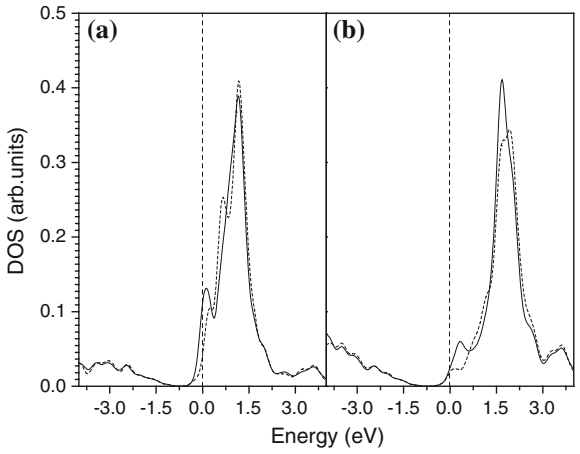


Fig. 3.32 The partial DOS of cation-s states with the different distance from the impurity. **a**, **b** are the Hg-s and Cd-s states, respectively. The *solid* and *dashed lines* are the NNN Hg (Cd) around the impurity and the Hg (Cd) furthest from the impurity in the supercell. Reprinted with permission from Ref. [75]. Copyright 2005 by American Physical Society

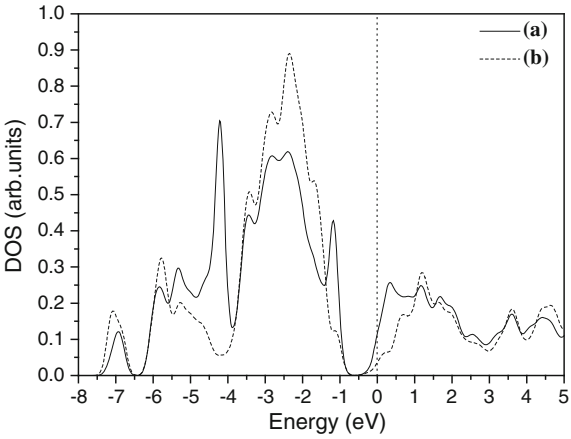


Fig. 3.33 The partial DOS of Te-p states with the different distance from the impurity. **a**, **b** are the Te-p states, respectively. The *solid lines* are the NN Te-p around the impurity and *dashed lines* are the furthest Te-p from the impurity in the supercell

cation-*s* at the CBM. From Fig. 3.34, we can predict that the single donor of As_{Hg}-doped $Hg_{0.5}Cd_{0.5}Te_{0.5}$ mainly comes from the coupling of the As-*p* and the NNN cation-*d* states.

As shown in energy level diagram for As, Cd, Te, and Hg atoms (Fig. 3.35), As-*p* state energy level is nearby Te-*p* state energy level, while cation-*d* state energy level is nearby Te-*p* state energy level. Therefore, the coupling between As-*p* state energy level and cation-*d* state energy level is due to the NN Te-*p* state energy level forming a peak around 4.22 eV below the Fermi level in DOS (Fig. 3.36).

(2) Electronic properties of As_{Te} in MCT

To determine the reason for the NN atom inward relaxation in MCT caused by the As impurity, the valence and the bonding-charge density have been calculated

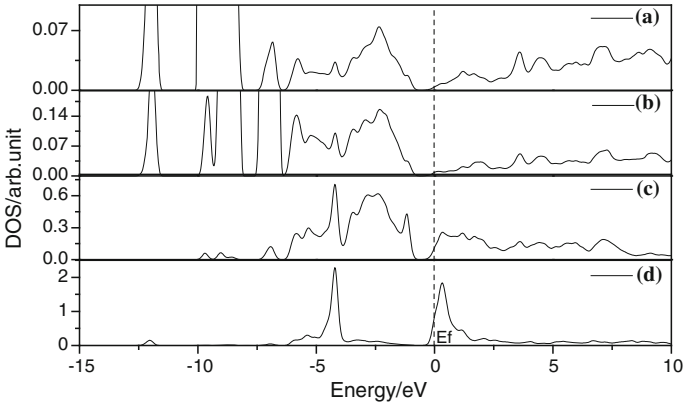


Fig. 3.34 **a**, **b** The partial DOS of the NNN Cd(Hg)-d state. **c**, **d** The partial DOS of the NN Te (As)-d state

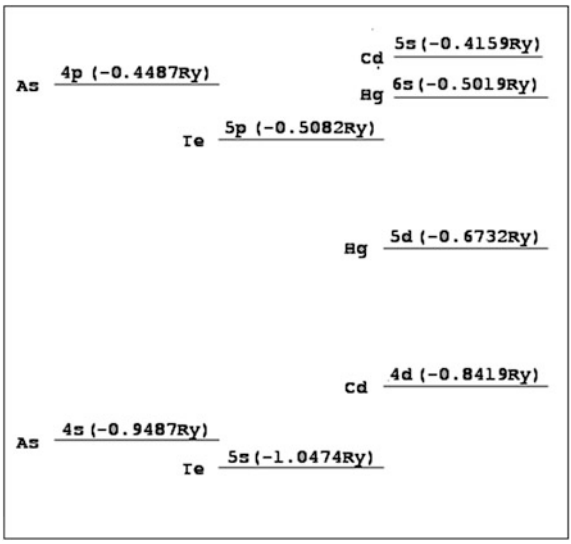


Fig. 3.35 Energy level diagram for As, Cd, Te, and Hg atoms

(Figs. 3.37 and 3.38). The charge density distributions between the arsenic impurity and the NN host atoms show the covalent characteristic. The bonding characteristic between the impurity and the host atoms is similar to the situation in the case of As_{Hg}. The valence charge density along the line of As/Te–Hg/Cd bond in the case of As_{Te} is shown in Figs. 3.39 and 3.40. The results indicate that As–Hg/Cd bonds for the As_{Te} case are covalent with stronger ionic characteristics than that of Te–Hg/Cd bond in the pure system. Together with the bonding mechanism analysis in the case of As_{Hg}, we can see that, in addition to the smaller Arsenic radius, the strong bonding also results in the NN host atom inward relaxation. Moreover, the

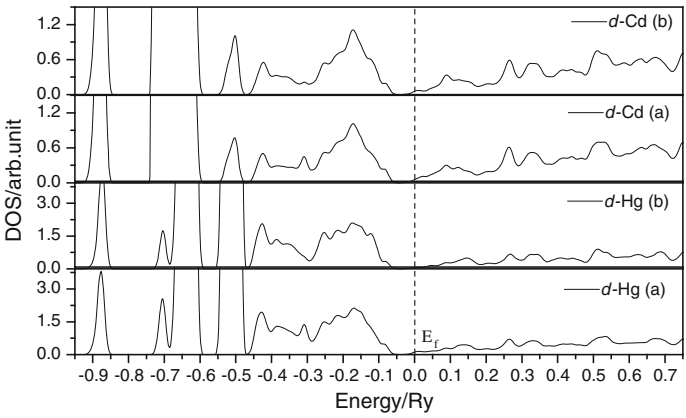


Fig. 3.36 The coupling between As-p state energy level and cation-d state energy level

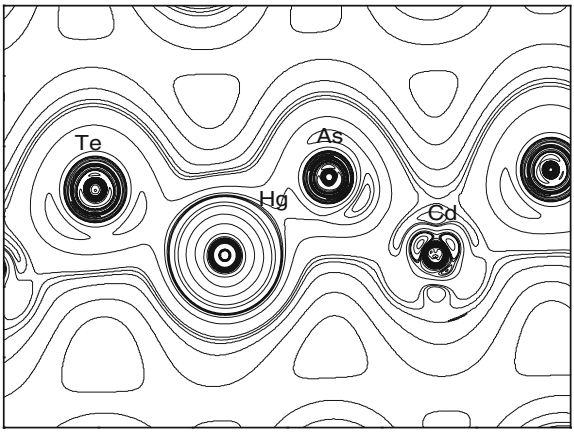


Fig. 3.37 Valence charge density in the (110) plane for AsTe. The contour step size is 8×10^{-3}e/(a.u)^3 for total charge density. Reprinted with permission from Ref. [95]. Copyright 2007 by American Physical Society

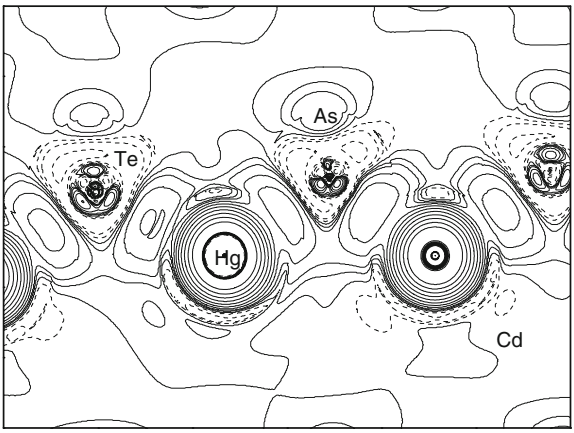


Fig. 3.38 Bonding-charge density in the (110) plane for As_{Te}. The contour step size is 5×10^{-3}e/(a.u)^3 for bonding charge density. Reprinted with permission from Ref. [95]. Copyright 2007 by American Physical Society

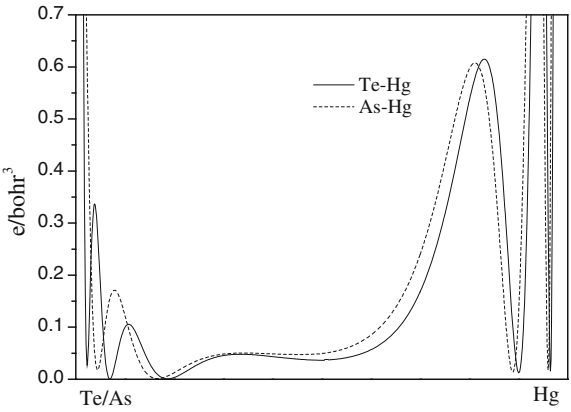


Fig. 3.39 Valence charge density along the bond line of As–Hg for the As_{Te} case and the total charge density along the bond line of Te–Hg for the pure system

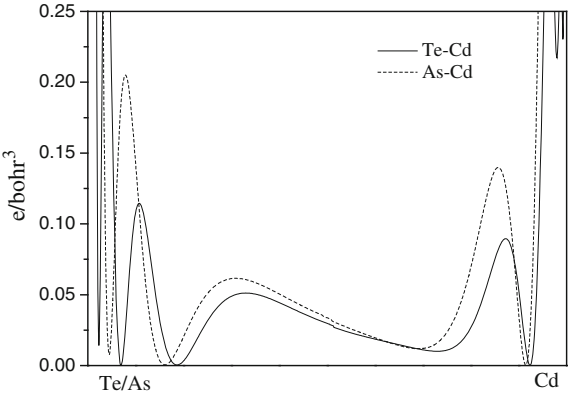


Fig. 3.40 Valence charge density along the bond line of As-Cd for the As_{Te} case and the total charge density along the bond line of Te–Cd for the pure system

Table 3.10 Charge transfer in the atomic sphere for As_{Te} case

	$Hg_{0.5}Cd_{0.5}Te–As_{Te}$		
	$Q_{crystal}$	Q_{atomic}	ΔQ
As	31.508	31.103	0.405
Te[2]	50.000	49.730	0.270
Cd[1]	46.223	46.199	0.024
Hg[1]	77.919	77.937	−0.018

Subscripts "atomic" mean the superposition of atomic charge. $\Delta Q = Q_{crystal} - Q_{atomic}$

strong bonding characteristics account for the stable doping behavior of As in MCT, which produces low diffusivity of the As-doping (Table 3.10).

The DOS of As_{Te}-doped $Hg_{0.5}Cd_{0.5}Te_{0.5}$ is shown in Fig. 3.41. Substitution of a tellurium atom by an arsenic atom produces the resulting metallic system. Evidently, the Fermi level of the AsTe-doped system (see Fig. for that of the As_{Hg}) does not lie within the band gap which extends to the valence band, as shown in

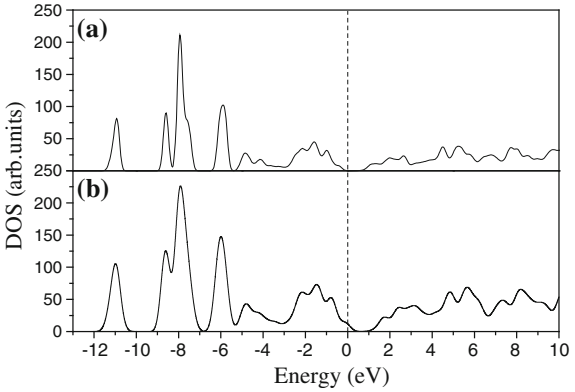


Fig. 3.41 Total DOS of **a** undoped and **b** As_{Te}-doped MCT-0.5. Reprinted with permission from Ref. [95]. Copyright 2007 by American Physical Society

Fig. 3.41. The DOS integrated from the Fermi level to the midgap accommodates one electron. The results show that the impurity As_{Te} in MCT behaves as a single acceptor. The theoretical predications are in good agreement with the experimental results. For the case of charged neutral As_{Te}, the defect center has a total of seven electrons. Two of them occupy the a_1^v states. Five of them occupy the t_2^v states. One of the unoccupied t_2^v states just above the valence band maximum (VBM) behaves as a single acceptor. The results of the calculated band gaps for undoped and doped HgCdTe (Fig. 3.41), respectively, show that the doping causes the band gap to be 13 meV smaller than that of the undoped one. If we neglect the effect of the valence-band broadening on the band gap, the change can be ascribed to the acceptor state. The acceptor level can be qualitatively located in the region of more than 13 meV above the VBM. The quantitatively calculated results from the single particle eigenvalues reveal that the acceptor level is 14 meV above the VBM, which means AsTe is a shallow acceptor. Based on the relation between the ionization of the acceptor level and the mole fraction of MCT from experiments, we can predict that the AsTe acceptor level will be shallower as the mole fraction becomes smaller. Especially, according to the most technologically important composition x around 0.2, the acceptor level is about 5–8 meV above the VBM [106, 101].

(3) Formation energy of As_{Te} and $As_{Hg/Cd}$ in MCT

Figure 3.42 shows the calculated formation energies of As-doped MCT for $x = 0.25$, 0.5, 0.75, and 1 as a function of the Fermi energy under cation- and Te-rich growth conditions. The slope of the line gives the charge state of the doping at the Fermi energy. The transition energy level is the Fermi energy, at which the slope changes value. Under $\mu_{Te} = 0$ condition, as shown in Fig. 3.42, the formation energy of As_{Hg} for $x = 0.25$, 0.5, and 0.75 is smaller than that of As_{Te} as the Fermi energy shifts inside the band gap. These results indicate that arsenic behaves as the dominant n-type dopant for MCT when $x = 0.25$, 0.5, and 0.75 under Te-rich condition. The experiments of MBE growth of HgCdTe [10, 41, 44] (for which the typical growth condition is Te saturated) and LPE growth of HgCdTe incorporate arsenic as dopant under Te-saturated condition, result in n-type doping behavior. The results shown in Fig. 3.42 indicate that the formation energies of As_{Hg} are about 1.0–1.5 eV smaller than that of As_{Te} as the Fermi level at VBM. It indicates that As easily incorporates into the cation site under the Te-saturated growth condition, behaving as n-type dopant. However, under the Te-rich limit, the formation energy of As_{Cd} for CdTe is smaller than that of As_{Te} as E_F below 0.3 eV. As the Fermi energy shifts up toward the midgap, the formation energy of As_{Te} is smaller than that of As_{Cd}. Namely, the As_{Te} and As_{Cd} compensate each other at the Te-rich limit and the Fermi energy is pinned at a level closer to the VBM; consequently, the arsenic impurity in CdTe, even at Te-rich limit, tends to be slightly p-type dopant. Although there are few experimental reports under this extreme growth condition, arsenic-doped CdTe layers grown by organometallic vapor-phase epitaxy produce p-type doping even when the partial pressures of Cd and Te are equal. Under $\mu_{Cd} = 0$ condition, as shown in Fig. 3.42, the formation energy of

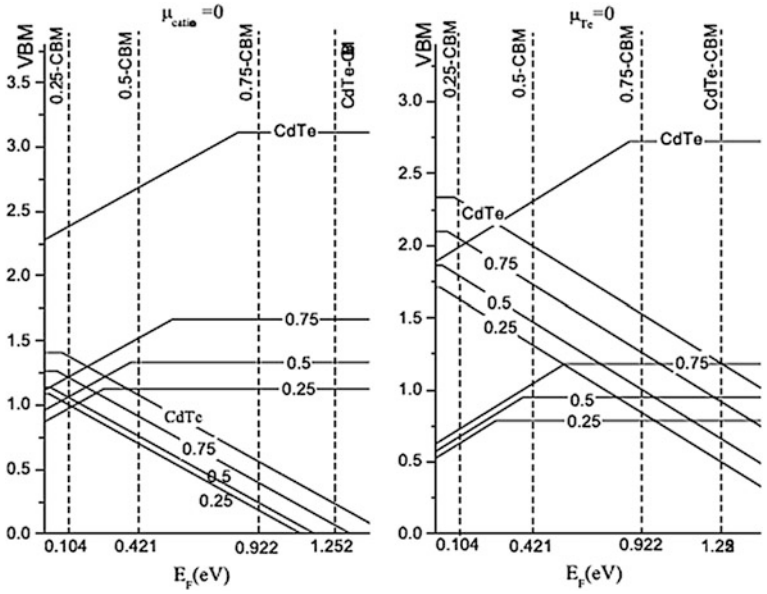


Fig. 3.42 Calculated formation energy of As_{Te} and $As_{Hg/Cd}$ according to the mole fraction of MCT as a function of the Fermi levels. The VBM of the entire four mole fraction is set to zero. Reprinted with permission from Ref. [95]. Copyright 2007 by American Physical Society

As_{Te} is nearly 1 eV smaller than that of As_{Cd} for CdTe. The arsenic dominantly incorporates into the Te site, behaving as p-type dopant under the Cd-saturated condition. The result is in good agreement with the MBE growth CdTe doped by arsenic under Cd-saturated condition as efficient *p* type. The difference of the formation energy of As_{Hg} and As_{Te} for $x = 0.25$, 0.5, and 0.75 is around 0.5 eV as the Fermi energy at VBM under Hg-rich limit. As for $x = 0.25$, the formation energy of As_{Hg} is smaller than that of As_{Te} as the Fermi energy shifts inside the band gap. These results indicate that the arsenic in MCT-0.25 dominantly behaves as n-type dopant, even under Hg-saturated growth condition. The compensation effects between As_{Hg} and As_{Te} occur for $x = 0.5$ and 0.75 and the Fermi energy is pinned at about 0.1 eV above VBM. Namely, arsenic doping in MCT-0.5 and MCT-0.75 tends to be slight p-type dopant under Hg-rich limit.

2. Mechanism of compensated p-type doping

The current technology requires well-controlled p-type doping of HgCdTe with very well defined junction formation. Arsenic is employed as a most desired p-type dopant. However, arsenic is observed to be incorporated as donor (which is strongly compensated by unknown defects) rather than acceptor [101, 105], therefore necessitating postgrowth annealing to activate arsenic acceptor. On the other hand, arsenic doping is related to the adsorption of As molecules under the typical

low-temperature conditions of MBE growth, which results in the complication of As-doping.

In situ incorporation of arsenic is routine in HgCdTe MBE growth, but there are still controversies over the doping characteristics. For example, the observed highly compensated n-type conductivity in the as-grown samples has never been satisfactorily explained although different growth models have been suggested [107–109]. The 100 % activation efficiency is limited to high arsenic concentrations in the As-doping levels about 10^{16}–10^{18} cm^{-3} range. However, As-activation efficiency drastically decreases at the doping levels above 10^{18} cm^{-3} [110, 111]. The As activation seems to be masked by some unidentified n-type defects. Moreover, details of shallow defect levels can be obtained by electrical or optical experimental measurements, but there are controversies on the deep level characteristics, such as the deep level at 60–110 meV in the annealing sample [111]. Therefore, the nature of the doping characteristic is a subject of ongoing research on the As-doping of HgCdTe.

As mentioned previously, As clusters combined with V_{Hg} are believed to account for the highly compensated n-type characteristic in the as-grown HgCdTe. To better understand the microscopic nature of the As dopants, we further analyzed the electronic structures and doping characteristic of As_2 clusters combined with V_{Hg} by the first-principle calculations.

(1) Modeling and calculation method

The technologically important alloy $Hg_{0.75}Cd_{0.25}Te$ was chosen as host system in this section. According to the growth conditions, As_2 clusters (As_{dimer}) are classified as follows:

(a) When the density of V_{Hg} is high, structure (a) (Fig. 3.43) denoted as $s_{Hg}i_{Te}$ with one of the two arsenic atoms occupying the mercury site and the other arsenic atom occupying the interstitial site. These two arsenic atoms form a dimer bond. After structure relaxation, As_{Hg} deviates from the mercury site indicating that the interstitial site is more suitable for As_2 cluster to combine with V_{Hg} than lattice site. Then three possible structures with the two As atoms at the interstitial sites to form As_{dimer} around V_{Hg} are denoted as $(i_{Hg}i_{Te})_{spl}$, $(i_{Te}i_{Te})_{spl}$ and $(i_{Hg}i_{Hg})_{spl}$ [structure (b), (c) and (d)], respectively. As_2 cluster bounded by V_{Hg}can be divided into two categories. One is two As atoms are bounded at the same side of V_{Hg}, such as structure $s_{Hg}i_{Te}$ and $(i_{Hg}i_{Te})_{spl}$ [structure (a) and (b)]. The other is two As atoms are bounded at the split sides of V_{Hg}, such as structure $(i_{Hg}i_{Hg})_{spl}$ and $(i_{Te}i_{Te})_{spl}$ [structure (c) and (d)]. As has been discussed before, the As_{dimer} related to V_{Hg} is a possible compensating donor in as-grown materials.

(b) When there is a certain density of V_{Te}, we put one As atom at V_{Te} site, the other As atom at the interstitial site to form As_{dimer} denoted as $s_{Te}i_{Te}$ [structure (f)]. After structure relaxation, As_{Te} atom is still at Te site, which indicate that this structure does not need too much energy in the activation annealing process.

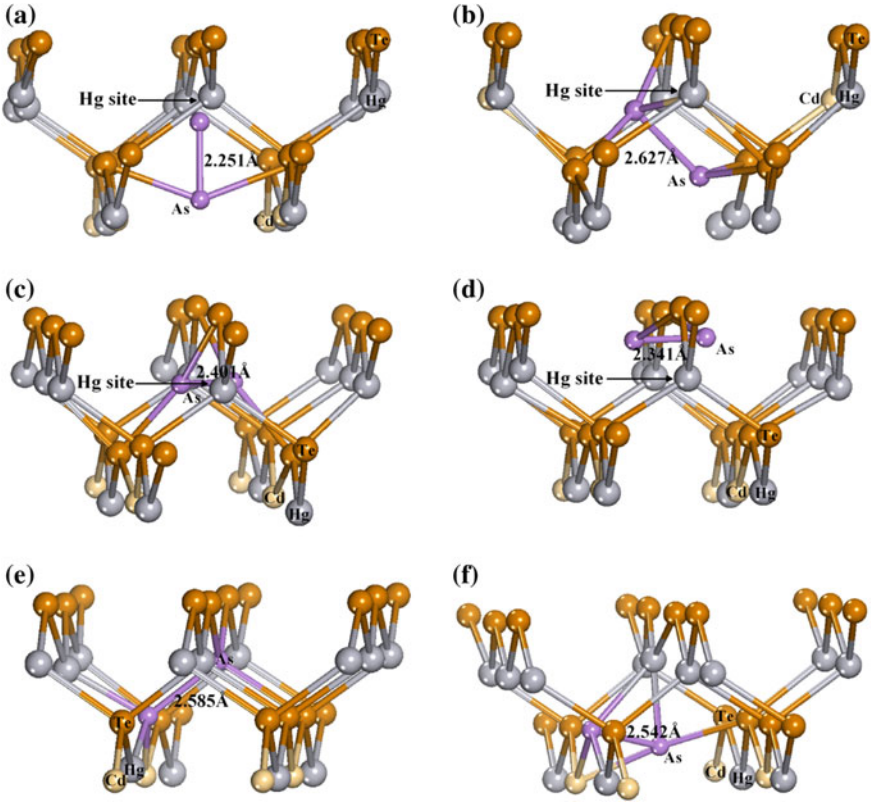


Fig. 3.43 Calculated As_2 cluster structures related to V_{Hg} and V_{Te}

(c) When the density of V_{Hg} and V_{Te} is similar, we put one As atom at V_{Hg} site, the other As atom at the V_{Te} site to form As_{dimer} denoted as $s_{Hg}s_{Te}$ [structure (e)]. After structure relaxation, As_{Hg} atom is still at Hg site, but As_{Te} atom deviated slightly from Te site along As–As bond direction due to the smaller bond length of As_2 than that of Hg–Te bond. This indicates that As_{Hg} is stable and requires certain energy to be activated in the annealing process.

Besides As_{2dimer}, there is a small quantity of singular arsenic atoms in a cracking arsenic source. The singular arsenic atom occupied V_{Hg} denoted as s_{Hg}. The complex impurity combined s_{Hg} and V_{Hg} is denoted as (s_{Hg}–V_{Hg}).

Simulations were performed in a $3 \times 3 \times 2$ SC containing 144 atoms. For charge defects, a uniform background charge (i.e., jellium charge) was added to keep the global charge neutrality when electrons are exchanged between the SC and the electron reservoir (i.e., Fermi level). Integrations over the Brillouin zone were performed using an energy cutoff of 150 eV and a $3 \times 3 \times 3$ special k-point mesh.

Using the general-gradient approximation (GGA) and pseudopotentials as implemented in the VASP code, all the defect calculations started from equilibrium configurations. The energy convergent criterion was 10^{-5} eV per unit cell. During the structural optimization procedure, all the atoms within the defect SC were allowed to relax until the Hellmann–Feynman forces acting on them were below 0.1 meV/Å.

The formation energy of a defect α in charge state q is defined as

$$\Delta H_f(\alpha, q) = \Delta H_f(\alpha, q = 0) + \delta E(\alpha, q) + \delta E(\text{host}, -q) + q(\varepsilon_i + E_\text{F}) \tag{3.6}$$

The first term in Eq. 3.6 is the formation energy for the defects in a neutral SC.

$$\Delta H_f(\alpha, q = 0) = E^N_\text{tot}(\alpha, q = 0) - E^M_\text{tot}(\text{host}) + \sum_i n_i(\mu_i^0 + \mu_i) \tag{3.7}$$

N, M denotes the total number of electrons in a defective SC and a perfect SC separately. The second term in Eq. (3.6) is the total energy difference between the host with $|q|$ holes at the VBM (or with q electrons at the CBM) and the (defect-free) M electron host.

$$\delta E(\alpha, q) = E^{(N-q)}_\text{tot}(\alpha, q) - E^N_\text{tot}(\alpha, q = 0) \tag{3.8}$$

These two terms correspond to the change of free energy and the change of chemical potential caused by atomic transfer. The third term in Eq. (3.6) is associated with jellium neutralization.

$$\begin{aligned} \delta E(\text{host}, -q) &= E^{(M+q)}_\text{tot}(\text{host}) - E^M_\text{tot}(\text{host}) \\ &= -q(\varepsilon_i - E_i) \end{aligned} \tag{3.9}$$

The last term in Eq. (3.6) is the change of chemical potential caused by electronic transfer. The last two terms correspond to the change of chemical potential caused by electronic transfer.

Substituting Eqs. (3.7)–(3.9) into Eq. (3.6), ΔH_f (α, q) can be expressed as

$$\begin{aligned} \Delta H_f(\alpha, q) &= \left\{ E^{(N-q)}_\text{tot}(\alpha, q) - E^M_\text{tot}(\text{host}) \right\} + \sum_i n_i(\mu_i^0 + \mu_i) + q(E_i + E_\text{F}) \\ &= \Delta E(\alpha, q) + \sum_i n_i \mu_i + q E_\text{F} \end{aligned} \tag{3.10}$$

(2) Analysis of the DOS

Figure 3.44 gives the DOS of structure (a)–(g) (except (f)) together with that of $Hg_{0.75}Cd_{0.25}Te$ for comparison. The Fermi level is set to zero.

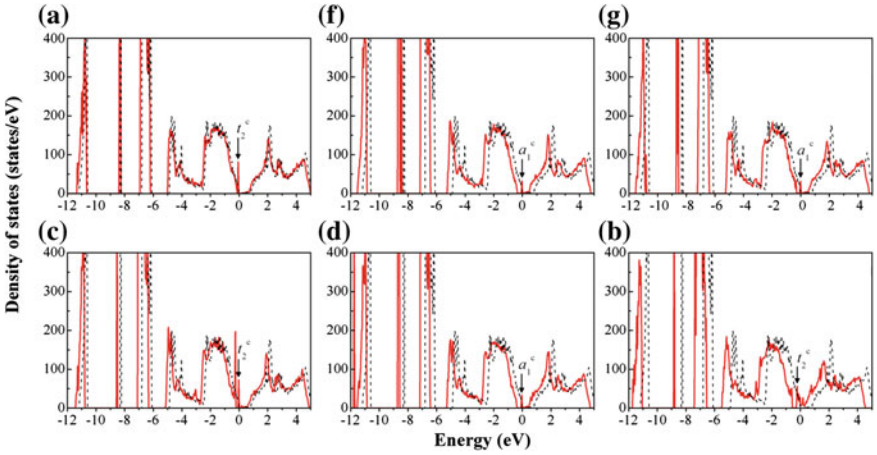


Fig. 3.44 Calculated DOS of As_2 clusters in HgCdTe. The dashed lines are DOS of $Hg_{0.75}Cd_{0.25}Te$ as reference

For all neutral defects, the main DOS features are similar to that of perfect HgCdTe, except for extra levels are around the band gap. The defective levels of structure (a) and (c) are located near the top of valence band. The defective levels of structure (a) and (c) appear in the vicinity of a band gap of about 1/3. Similar features around the band gap reflect the similarity of doping characteristics. Through the integration from the top of the valence band to the Fermi level, one can determine the filling states of electrons in the highest defective levels. The results demonstrate that the number of electrons that can be filled into the highest defective level is 3 for structure (a) and (c), 1 for structure (d) and (f), 8 for structure (b), 2 for structure (f), and 0 for structure (e), respectively.

(3) Single-electron energy levels of defective states

We will further investigate the characteristics of defective states in structure (a)–(g) and the SC containing s_{Hg}, V_{Hg}, or s_{Hg}–V_{Hg}. First, the energy levels corresponding to defective states should be determined by analysis of the energy eigenvalues of Gamma points in these SCs. Then, according to the degeneration of energy eigenvalues and filling situation of electronic states at Gamma points, the number of electrons corresponding to the defect states will be determined. Finally, based on the law of electronic filling for zinc-blend structures with T_d symmetry, the filling characteristics of defective states in the structures mentioned above can be determined.

Figure 3.45 describes the law of electronic filling for zinc-blend structure compound. Γ_{15v} state with t_2 symmetry at the top of valance band can accommodate up to six electrons. This state has *p*-like state characteristic. Γ_{1c} state with a_1 symmetry at the bottom of conduction band has *s*-like state characteristic. This state can accommodate up to two electrons.

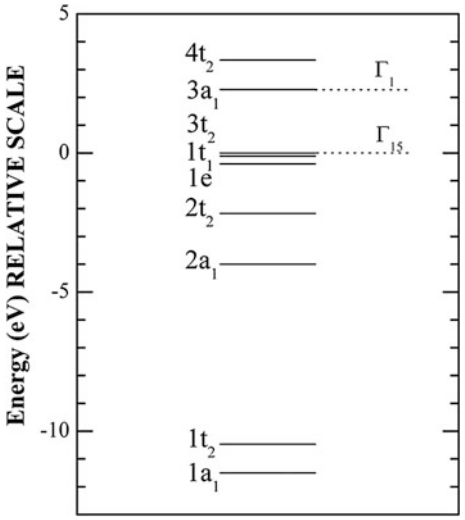


Fig. 3.45 Illustration Electronic filling law for zinc-blend structure compound with Td symmetry

Using the average potential field of the farthest atom away from the defect as a benchmark for the valence band top of defective systems, Fig. 3.46 shows the corresponding energy level and electronic structures of the defective state with highest energy. The figure also shows the photoluminescence (PL) spectrum results [93] of as-grown As-doped $Hg_{1-x}Cd_xTe$ ($x \approx 0.3$) as a comparison. According to the characteristics of defective states, the defective SCs are classified as one of the following:

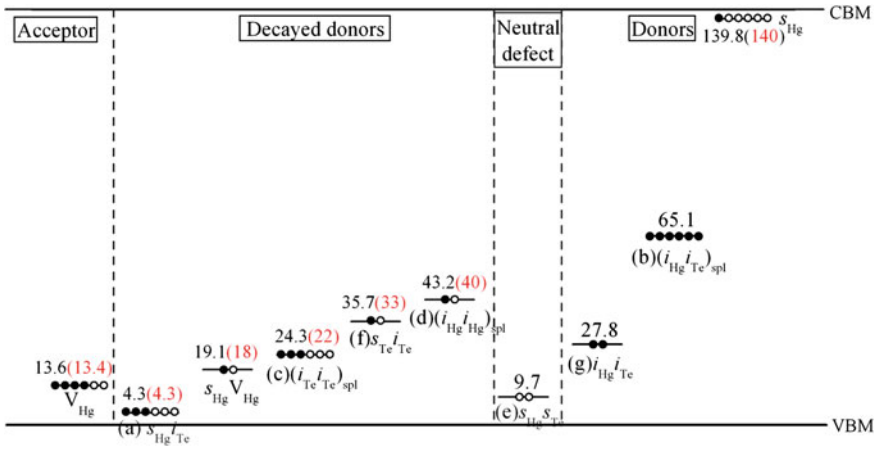


Fig. 3.46 Corresponding energy level and electronic structures of the defective state. PL spectrum results are listed in brackets

(a) Acceptor impurity:

There is one t_2^v unoccupied state in V_{Hg} respect to a perfect system. The lack of two electrons removed from t_2^v state. The unoccupied state moves 13.6 meV above the top of valance band forming a defective level. This defective state can capture electrons from valance band through thermal excitation. Therefore, V_{Hg} is a kind of typical shallow acceptor impurity. The peak at 13.4 meV above the top of valance band in PL spectrum is attributed to luminescence from valance band to shallow acceptor V_{Hg} level. This peak with strong intensity shows a large amount of V_{Hg} defects in the as-grown samples.

(b) Compensated donor impurity:

Both structure (a) and (c) provide a complete system with five electrons. Two electrons fill into lower a_1^c state making this level move into the valance band. The remaining three electrons fill into higher t_2^c state making this level a half-filled state. This level moves into band gap nearby the top of valance band. Both structures (d) and (f) provide a complete system with one electron. This electron occupies the a_1^v state making this level a half-filled state in the vicinity of band gap of about 1/3. In *s*Hg–V_{Hg} system, structure As_{Hg} provide with 3 electrons. Two of the electrons fill into a_1^c state, the remaining one electron fills into t_2^c state. The missing two electrons in structure V_{Hg} move from t_2^c and a_1^c state. In this system, the a_1^v state is a half-filled state and the defective level moves into the band gap in the vicinity of VBM.

The defective levels of these impurities have three characteristics. First, the highest defective level is in the band gap nearby the VBM. This defective state can capture electrons from the valance band through thermal excitation and form a deep level. Second, the number of filling electrons is no more than half-full. Therefore, when the electron capture cross section is large enough, the deep level can act as DX centers. Lastly, these impurities are very active chemically due to the unpaired electrons. They are nucleation centers to form larger As clusters.

These characteristics have been observed in the experiments. Electrical measurements show that there are two possible deep levels in band gap related to V_{Hg} [112]. In the PL spectrum, there are two peaks in the vicinity of about 1/3 band gap with energy difference 18 meV and are attributed to the deep center formed by As_4 or As_2 clusters. As shown in Fig. 3.46, the defective levels of structure (c) and (d) coincide with these findings. The two deep-acceptor levels are corresponding to these As_2 clusters isomers. There are peaks in PL spectrum corresponding to structures (a), (f), and S_{Hg}–V_{Hg}. The strongest peak corresponds to structure S_{Hg}–V_{Hg}, followed by structures (a), (c), and (d). The weakest peak corresponds to structure (f). This indicates that the dominant As impurities are related to V_{Hg} in as-grown samples. The density of the As impurities related to V_{Te} is less than that of V_{Hg}. Moreover, the binding ability to As impurity of V_{Hg} is stronger than that of As_2 cluster. Therefore, the density of S_{Hg}–V_{Hg} is higher than that of As_2 cluster binding with V_{Hg}.

(c) Neutral impurity:

S_{Hg} in Structure (e) has three extra electrons compared with a complete system and hence acts as a donor, while S_{Hg} acts as an acceptor due to missing three electrons. Therefore, structure (e) is a neutral impurity. This result is consistent with the following result of ionization energy analysis.

(d) Donor impurity:

S_{Hg} provides complete system with three electrons. Two of the electrons fill into a_1^c state making this full level move from conduction band down to valence band. The remaining one electron fills into t_2^c state. This state is not full and moves down to the bottom of conduction band about 10 meV acting as a defective level. So S_{Hg} is a typical shallow donor impurity. Structures (b) and (g) provide complete system with eight and two electrons separately. These electrons can fill full a_1^c state or a_1^c and t_2^c state. These two structures can only act as donor impurities.

The above analysis results show the common compensated donor characteristic of structure (a), (c), and (d), except for structure (b). So the As_2 clusters binding with V_{Hg} are very important for the highly compensated as-grown HgCdTe materials. Structures (e) and (f) binding with V_{Te} ($S_{Hg}i_{Te}$ and $S_{Hg}S_{Te}$) have compensated donor and neutral characteristics separately. Considering the low density of V_{Te}, the density of As_2 clusters being with V_{Te} is not high.

(4) Ionization energy analysis

According to Eqs. (3.2–3.11), ΔH_f (α, q) depends not only on the atomic chemical potential μ_i, butis also closely related to the electron chemical potential (i.e., Fermi level E_F). That is to say, the same defective system can have different states of charge, and the ΔH_f (α, q) corresponding to same charge state q is a linear function of E_F. When an electron is excited from a defect state with lower energy level in charge state q to a higher energy level in charge state q', the Fermi level will be in a position within the band gap range ε (q/q') to form defective states in charge q and q' at the same time. The defective formation energy of these two defects states is equal,

$$\Delta H_f(\alpha, q)\big|_{E_f=\varepsilon(q/q')} = \Delta H_f(\alpha, q')\big|_{E_f=\varepsilon(q/q')} \tag{3.11}$$

where ε (q/q') has been defined as a defect states ionization energy,

$$\varepsilon(q/q') = [\Delta E(\alpha, q) - \Delta E(\alpha, q^p rime)]/(q' - q) \tag{3.12}$$

ε (q/q') reflects the position of donor or acceptor level in the band gap.

According to Eq. 3.12, we have the analyzed ionization energies of the above four different doping impurities. Figure 3.47 shows the relationship between defective formation energy and E_f. The intersection of the corresponding straight

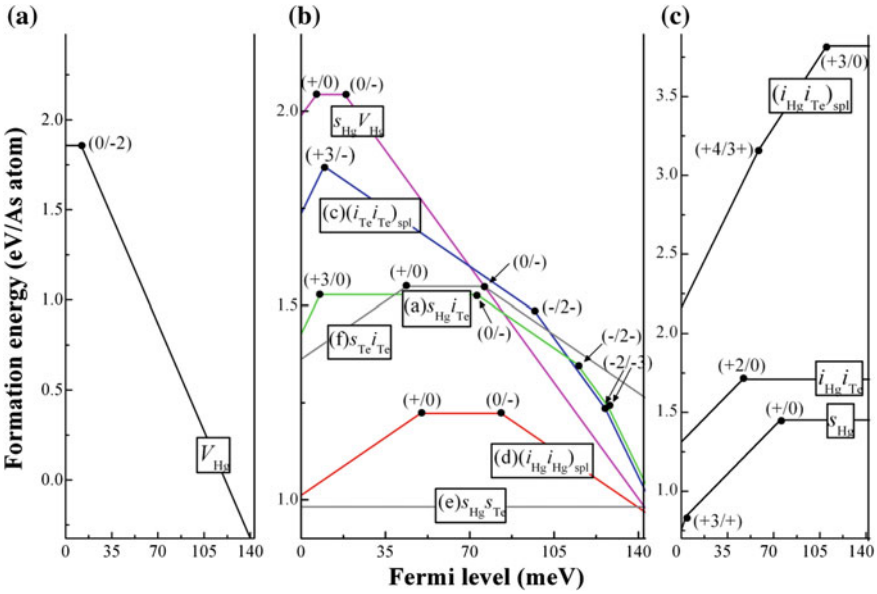


Fig. 3.47 The relationship between defective formation energy and E_f. Intersection of the corresponding *straight line* is the ionization energy of each ionization process

line with different charge states in the band gap (that is the inflection point in the figure) is the ionization energy of each ionization process. As can be seen above, ionization energy analysis results are entirely consistent with the single-electron energy levels analysis results. (1) V_{Hg} can accept two electrons thermally excited from the valence band and forming a shallow acceptor ionization level ε (0/−2) on above the top of valence band about 12 meV. Hall test results show that fully activation energy level of V_{Hg} is 10–12 meV [113]. This is in good agreement with the calculated results. (2) Compensated donor impurities can form both donor ionization level and acceptor ionization energy level at the same time. In these impurities, structure (d) has the minimum defect formation energy and is the most important factor in arsenic compensated n-type doping. As shown in Fig. 3.41, structure (d) can adopt one electron to make a_1^c state a full state. At the same time, one electron of the a_1^c state can be excited to the conduction band. Therefor, deep donor level ε (+/0) and deep-acceptor level ε(0/−) will appear 49 and 83 meV above the top of valance band. Electrical measurement results [114] also show the existence of fully ionized deep levels in the band gap and the donor behavior are highly compensated. (3) Because the formation energy of S_{Hg} is lower than that of As_2 cluster, S_{Hg} is the main donor impurities in as-grown HgCdTe materials. The electron of t_2^c state can be excited to conduction band in S_{Hg}, and form shallow donor level ε (+/0) above the top of valence band 76 meV.

3.3 HgCdTe Growth on Si by MBE

The study of HgCdTe growth on Si by MBE was focused on the growth of composited ZnTe/CdTe buffer and extrinsic doping on HgCdTe in this chapter. The technical challenges of heteroepitaxy on Si included polarity control, suppression of twins on large mismatch interface, as well as the large amount of misfit dislocation in hetero-epilayer. The main issues of extrinsic doping were the doping efficient, the activation of doping elements, and As incorporation.

The theoretical models and calculations were discussed detailedly in Sect. 3.2. The experimental methods, growth processing and results were described in this section. The Si deoxide temperature was decreased to lower than 800 °C through chemical treatments based on Akitoshi Ishizaka solutions, which was optimized for surface morphology. Arsenic passivation was introduced on Si wafer after deoxide, which was proved to guarantee B-face growth on diamond substrates. A ZnTe/CdTe lattice-gradient buffer layer was used to suppress the twins and misfit dislocations caused by high interface energy due to large lattice misfits. A variety of methods on this complex buffer layer were studied, including tilting angle of substrates, growth condition of ZnTe nucleation, and CdTe epitaxy, to improve the crystal quality. The uniformity of epilayer and the formation mechanism of surface defects were studied on HgCdTe growth on large-size substrates. In intrinsic doping, it has been recognized that the difficulties in the As-doping in MBE-grown HgCdTe mainly come from the amphoteric behavior of As due to the growth mode of Te-rich, and the low-temperature growth condition as required by the low surface-sticking coefficient (SSC) of As, which further complicates the understanding about the primarily possible modes of As incorporation. In this section, the experiments on As-doping in growth and activation method by annealing were described.

3.3.1 ZnTe/CdTe Grading Buffer on Si by MBE

TheMCT FPA reliability issue can be ultimately resolved by epitaxial growth of HgCdTe-on-Si substrates, which completely matches the thermal expansion of a Si readout chip. Additionally, Si substrates offer the advantages of large available size, lower cost, superior mechanical strength, and increased thermal conductivity. The HgCdTe-on-Si technology was applied for fabrication of the third-generation, large-format HgCdTe IRFPAs. However, the technology of HgCdTe on Si by MBE faces dramatic challenges on polarity control, lattic misfit, and growth processing. The studies and results on the above issues are described below.

1. ZnTe/CdTe complex epitaxy on Si

The key technologies are as follows:

- Si is an active atom, which is easy to form carbon or oxide compounds. The contaminants on Si substrates, such as oxide and carbon, are the main cause of crystal defects in the epilayers. The cleaning temperature for carbon contaminants is about 1200 °C, which causes undesirable impurity diffusion and exceeds the heating capacity of MBE UHV system. The etching methods and chemical treatments on Si wafer were reported by many papers, but they do not meet the surface requirements of MBE on ultimate cleaning environment. So the chemical treatment of Si substrates is a crucial step, which significantly influences crystalline quality of subsequent epilayer.
- The main difficulty of HgCdTe epitaxy on Si is polarity control, which exceeds the concepts of well-known heteroepitaxy. The diamond crystal structure of Si has no polarity, which is different from the zinc-blend structure of CdTe. The polycrystalline is easily formed when CdTe is directly grown on Si without polarity control. The elemental issue is the transfer in crystal structure from Si diamond to CdTe zinc-blend.
- The major technical challenges in the growth of HgCdTe epilayer on Si substrates are large lattice mismatch of 19.3 % between Si and HgCdTe, which results in the twin formation and poor crystalline quality. Some previous studies showed that the twins formed on the interface of CdTe/Si if without any buffer layer. How to obtain the twin-free epilayer was reported and discussed in the later section.

(1) Chemical treatment on Si substrates

The perfect crystal film could not be obtained on the impurified substrates, which is also the main cause of defects. Usually, the impurities on substrates were removed by thermal dioxide, before film growth. While the carbon and oxide film are easily formed on a Si surface, the fact that the oxide film on Si is removed at temperature between 800 and 1000 °C in a vacuum was first reported by Lander and Morrison [115]. The removal of carbon is more difficult than oxide, and the CKLL peak disappears only after heating above 1100 °C [116, 117]. However, this technique causes undesirable impurity diffusion and changes the designed impurity concentration profile within the Si substrates. Furthermore, crystal defects, such as dislocations and sticking faults, tend to increase and slip lines are often generated across the Si substrate during high-temperature treatment. Therefore, if a carbon-free oxidized Si surface can be prepared, low-temperature thermal etching for surface cleaning is possible. Several studies on low-temperature cleaning techniques have been reported, such as ion sputtering [118], laser annealing [119], and Si flux deposition under Ga atmosphere [120]. However, the above cleaning techniques have their drawbacks of causing the points defects on the Si surface. Ion sputtering, however, produces undesirable radiation damage at the surface and point

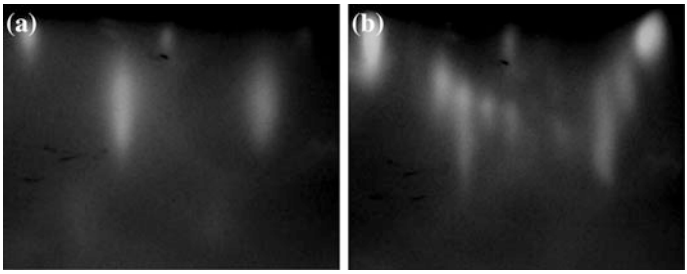


Fig. 3.48 RHEED patterns on Si(211) during deoxide. **a** RHEED patterns in [0–11] direction before deoxide, **b** RHEED patterns in [0–11] direction after deoxide

defects often remain even after annealing. Galliantion is effective in removing silicon oxide, but complete removal of Ga atoms from the Si surface with no Ga diffusion into the Si substrate has not yet been fully confirmed. Also, for the Si beam cleaning method, complete removal of carbon contaminants has not been certified (Fig. 3.48).

The chemical treatment in this work based on the Ishizaka cleaning method [121]. Si (211) substrates were prepared by the following procedures: ① organic cleaning, ② oxide formation and removal of the oxide layers, based on Ishizaka method, including boiling in HNO_3 or NH_4OH:H_2O_2:H_2O etching bath and 2.5 % HF solution etching, ③ thin protective oxide formation by Cl-passivation on Si surface, HCL:H_2O_2:H_2O = 3:1:1, which can be removed easily by heating at lower temperature. In procedure ②, two methods for oxide formation were experimented separately. (In Fig. 3.49 Method 1: Si was etched in HNO_3 for 10 min. After HF etching, Si was etched in NH_4OH:H_2O_2:H_2O for 10 min. In Fig. 3.49 Method 2: HNO_3 etching for 10 min, and then HF etching, the above procedures were repeated for 3 times.) After the chemical treatments, the Si substrate was flushed by DI water for 10 min and dried by N2 purge, then it was loaded to MBE chamber and outgassed at about 400 °C. The thin oxide formed by Cl-passivation on wafer was removed by heating the wafer to 740–830 °C.

The transformation of RHEED patterns on Si(211) during deoxide at 780 °C is shown in Fig. 3.49. When the wafer was heated to the proper temperature, it is clearly seen that the thin oxide cap by Cl-passivation was removed, corresponding to the RHEED pattern of (2 × 1) reconstruction.

Figure 3.49 shows the relation between different substrate etching methods and deoxide temperature. The Si substrates were, respectively, treated by method 1 and method 2 as above, with different HF etching time from 15 to 30 s. All the wafers could be deoxidized at temperatures below 830 °C. The deoxidation temperature could be reduced by repeated cycles of oxide formation and deeper HF etching for extending time. The deoxidation process can be stably controlled at a temperature below 800 °C with HF deep etching for $\geq$25 s. The optimal HF etching time was founded to be 25 s, as overetching may cause point defects on the surface.

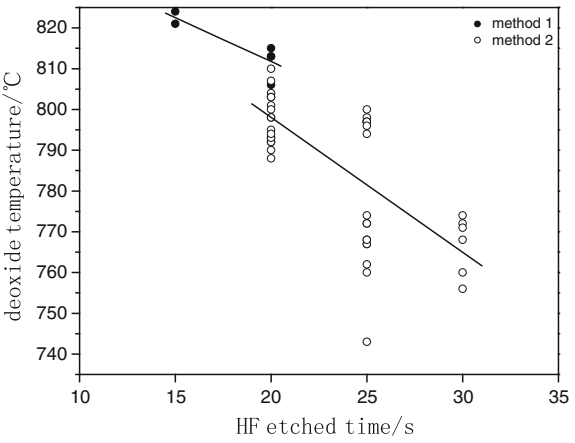


Fig. 3.49 The relation between different substrate etching methods and deoxide temperature

Figure 3.50 shows the run-to-run deoxidation temperatures of Si, by means of the optimized chemical treatment. The thermal clean temperature could be controlled up to 780–790 °C, within the UHV heating capacity.

(2) Surface polarity control

The most important issue is polarity control of CdTe on nonpolarity Si substrates. Polycrystal and two domains could be formed [122], due to the different structures at interface. Both A and B-face can be grown on Si. It depends on growth conditions and flux initial passivation on Si surface. Furthermore, concerned with the continuous HgCdTe growth requirement, the sticking coefficient of Hg atom on B-face (Te) is much larger than it on A face (Cd). Before CdTe growth on Si, polarity control is the most essential step, which would guarantee B-face mode on CdTe.

Several studies on surface polarity control have been reported. The methods used include Te passivation on cleaned Si, as well as the CdTe growth on Si with

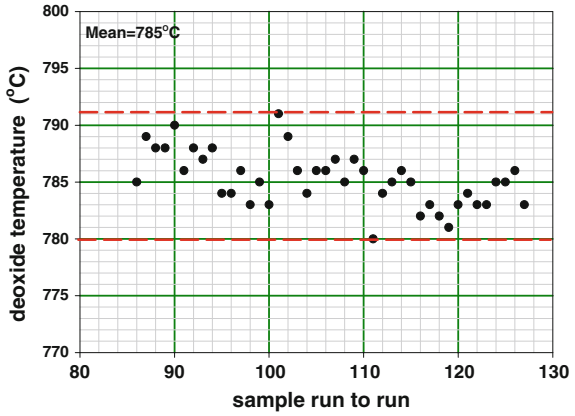


Fig. 3.50 The dioxide temperature of Si by optimized chemical treatment, based on run-to-run stastics

over-Cd flux [123, 124]. The CdTe epilayer was demonstrated to be A face by the above process, and the surface was extremely rough. Compared with Si-Cd bonding, SiTe formation was reasonable since Si was easier to be bonded with Te, due to the Si-Te bonding energy being lower by about 3.46 eV [125]. Furthermore, the sticking coefficient of Cd atom on Si surface is less than on the Te. So after Te passivation on Si, the Te atom overlapped on the edge and terrace, which enhanced the Cd immigration on Te face and prevented Cd atom to bonding directly with Si. Thus, the CdTe A face was formed under 3D growth mode. As the calculation above analyzed, the polarity control was greatly related with the stick coefficient of each atom and bonding energy. If As flux was introduced after Si dioxide, Si–Te bonding was blocked, which greatly improved CdTe/Si surface morphology. So far, As-passivation on Si was widely seemed as the most effective way to control polarity and obtain B-face growth.

In 3.2.1, the reconstruction of Si(211) during deoxide and the mechanism of As-passivation on Si were discussed in details by calculation and models. In this section, the experimental data was studied. As-passivation was employed after deoxide and it is a crucial step for obtaining single-domain B-face CdTe(211) on Si (211). The Si surface by As-passivation was analysed by RHEED, which is consistent with the theoretical calculation. Compared with Fig. 3.48b, the diffraction intensity turned to be slightly weak in Fig. 3.51, without any noticeable change on (2×1) reconstruction. This indicated that the Si atoms on terrace were replaced by As atoms, not absorbed. Figure 3.52 shows the intensity transformation of RHEED pattern during As-passivation. The feature has no noticeable change, while the diffraction intensity decreased after As deposited on Si at initial stage, then it reached saturation level after 20 s, which demonstrated the replacement between As and Si was finished. Decorated etching of CdTe epilayer also confirmed that As-passivation could ensure the B-face epitaxy of CdTe on Si.

(3) Twin suppression in large lattice mismatched heterostructure

A. The formation of twin

In heteroepitaxial systems, the difference between lattice parameters of the substrates and the overgrowth is substantial [126]. The larger lattice mismatch

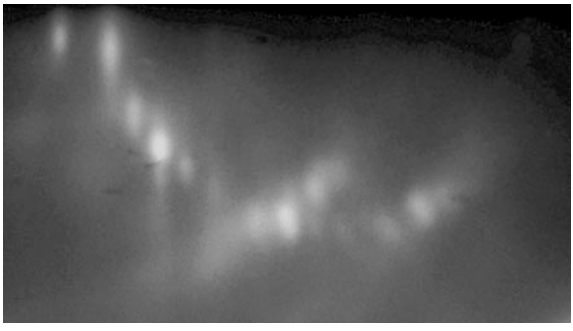


Fig. 3.51 RHEED patterns in [0–11] direction of Si (211) after As-passivation

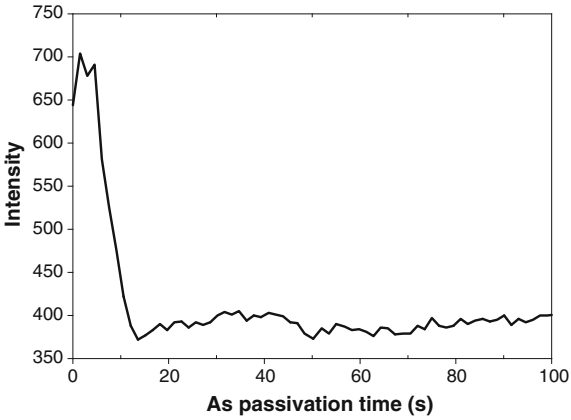


Fig. 3.52 Intensity trend line of RHEED patterns during As-passivation

accommodates the strain energy by forming a high density of islands during the initial stages of layer growth. Furthermore, the surface kinetic barriers are higher thus generating disoriented and randomly distributed nuclei at the initial stages of the growth [127, 128]. The island-coalescence is thus delayed which leads to 3D growth and produces a columnar-like surface morphology. It is obvious that large lattice mismatch is the main cause of twin formation.

The lattice mismatch between GaAs and CdTe is 14.6 %。 The twin was easily formed if CdTe is directly deposited on GaAs at normal temperature, due to the high surface energy. In Fig. 3.53, the RHEED patterns were compared on CdTe (211)B/GaAs grown at same temperature. In Fig. 3.53a, the twin diffractions due to (133) were clearly identified on the CdTe directly growth on GaAs. In Fig. 3.53b, the intensive streaked pattern demonstrated that ZnTe grading buffer made improvement in twin suppression. In Fig. 3.53c, the twin also could be suppressed by lower 25 °C growth temperature, which affected a decreasing surface energy. After the good nucleation of CdTe on GaAs, the temperature should be increased to the normal point, which guarantees high crystal quality. The 2D growth features were both indicated in Fig. 3.53b, c, which demonstrated that the lower surface energy is effective to suppress the twin.

B. Twin-free CdTe(211)B on Si

The studies on GaAs/CdTe shown that the twin could be suppressed by lowered the surface energy. One way is ZnTe grading buffer growth between substrate and CdTe. And it is also quite effective if CdTe is nucleated at a lower temperature.

Due to the larger lattice mismatch (19.3 %) between Si and CdTe, the twin and 3D growth mode are more easily formed than on CdTe/GaAs. The twin in CdTe/Si could not be ultimately suppressed by the directly growth mode. ZnTe is an ideal grading buffer, because its lattice constant is between Si and CdTe, which mismatch 12.3 and 6.2 % respectively. To obtain twin-free epilayer, the growth conditions for sequential epitaxy processes were studied, including initial nucleation of ZnTe on Si, high-temperature annealing, and CdTe or ZnTe growth.

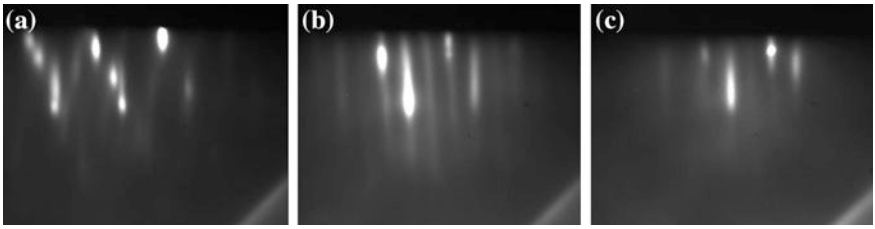


Fig. 3.53 RHEED patterns in [0–11] direction of CdTe(211)B/GaAs. **a** CdTe directly growth on GaAs; **b** ZnTe buffer growth before CdTe epitaxy on Si; **c** CdTe directly growth on GaAs at temperature 25 °C lower than **a**

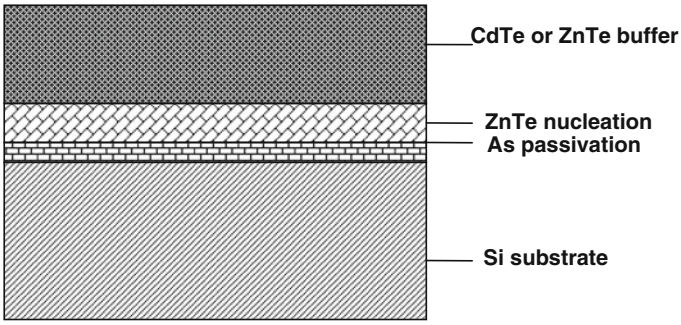


Fig. 3.54 The composite structure of Si/ZnTe/CdTe

The heteroepitaxial structure is illustrated in Fig. 3.54. After deoxide, As_4 flux was first deposited on the Si wafer during cool down to a proper growth temperature. In this work, the effect of initial ZnTe nucleated layers on (133) twin suppression was studied at different temperatures of 156, 170, 200, 220, 250, 280 °C. The thickness of initial ZnTe layers was 25–100 Å. The continuous high-temperature annealing was performed and CdTe or ZnTe layers grown at normal temperature followed. The normal ZnTe layers (about 2 μm) were grown at 320 °C and CdTe layers (3–12 μm) were grown at 280 °C.

C. ZnTe nucleation at low temperature

It was found that the proper nucleation temperature is one of the essential issues to suppress the twin. Figure 3.55 shows RHEED patterns of initial ZnTe layer (about 100 Å) nucleated at 156 and 200 °C. The diffractions in Fig. 3.55a, c were almost smeared out and polycrystalline structure was observed at such a lower nucleation temperature. The polycrystalline structure is easier to form, as the migration of surface atoms is limited at a lower nucleation temperature. The crystalline structure could be improved by high-temperature annealing at 380 °C for 15 min, as shown in Fig. 3.55b, d. Nevertheless, it should be noted that the

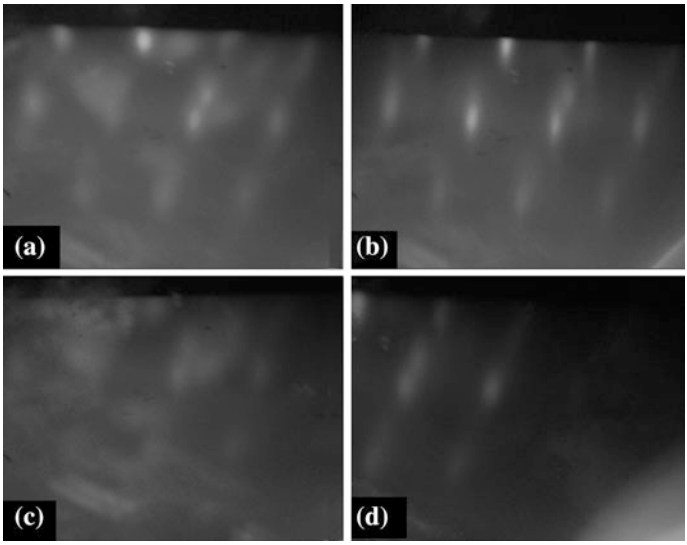


Fig. 3.55 RHEED patterns in $[0\bar{1}1]$ direction of ZnTe/Si(211)B. **a** Nucleation at 200 °C (No. SZT011), **b** After in situ annealing at 390 °C (No. SZT011), **c** Nucleation at 156 °C (No. SZT010), **d** After in situ annealing at 390 °C (No. SZT010)

effect of high-temperature annealing on crystal quality is limited. The twin phase could not be ultimately eliminated by high-temperature annealing. If the initial nucleation temperature was too low or the initial ZnTe layer was too thick, the polycrystalline structure was excessively formed, and then it could not be improved by high-temperature annealing.

It should be pointed that, when nucleated at a higher temperature (above 230 °C), the twin is found to occur much easier, because the surface atoms migrate faster to the proper location to reduce the surface energy by large lattice mismatch. Furthermore, the twin could not be eliminated by high-temperature annealing, which was shown in Figs. 3.55b and 3.56. As Fig. 3.58 shows, when a ZnTe initial layer was nucleated above 230 °C, XRD results showed that (133) twin reflection was intensified, which also evaluated the content of twins in film.

D. ZnTe and CdTe epitaxy at proper temperature

Due to the high surface energy by large lattice mismatch, (133) twins could not be totally suppressed by a thin ZnTe nucleation and high-temperature annealing. The later buffer layer growth was performed, to eliminate the twin formation.

Figure 3.57 shows the RHEED patterns in [0–11] direction of Si/ZnTe(211) layer. The twin spotty pattern of ZnTe nucleation is obvious, which could be greatly improved by high-temperature annealing. The following 1.71 μm ZnTe growth at proper temperature was found to effectively suppress (133) twin, which contributed

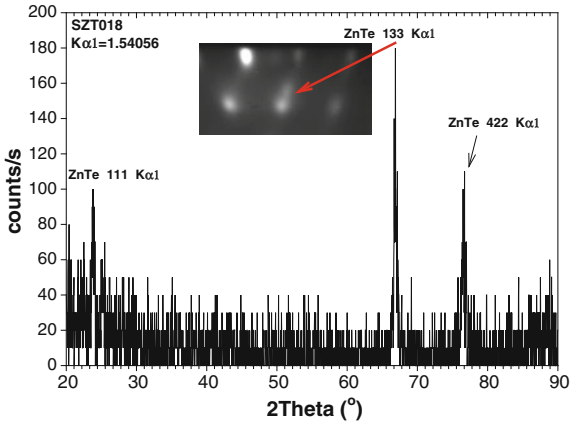


Fig. 3.56 θ–2θ *x*-ray scan for CdTe(211)B/Si(211) with different growth condition

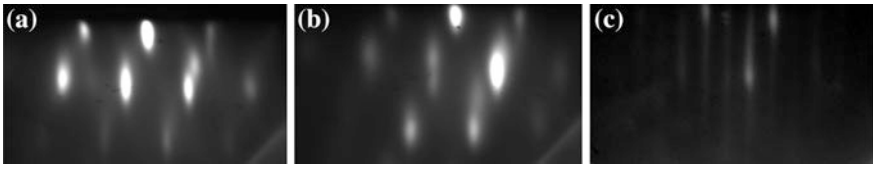
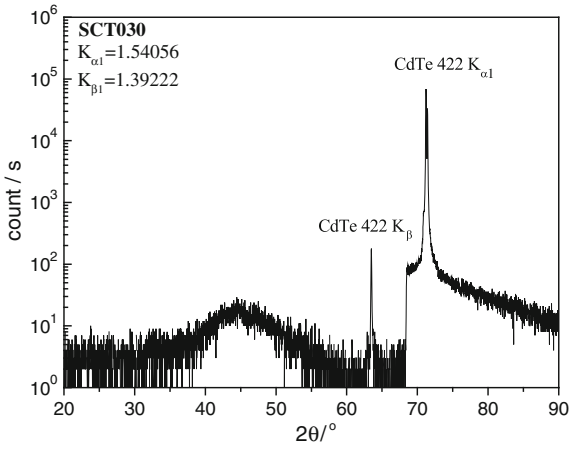


Fig. 3.57 RHEED patterns in $[0\bar{1}1]$ direction of ZnTe/Si(211)B. **a** ZnTe nucleation, **b** after high-temperature annealing, **c** ZnTe grown at 320 °C

Fig. 3.58 XRD θ–2θ scan on Si/CdTe(211)B



to the improvement of crystal quality. The stacky and intensified pattern indicated the good 2D growth mode and single-crystal diffraction, which was also demonstrated by XRD measurement.

As well as the ZnTe epitaxy, twin-free CdTe on Si could also be obtained at about 280 °C, which was performed on the proper ZnTe initial layer. XRD analysis has been used to evaluate the crystalline quality of Si/ZnTe/CdTe(211)B (sample No. SCT030). The CdTe thickness is 5.5 μm and ZnTe initial layer is about 100 Å. In Fig. 3.58, the single domain of CdTe(422) was examined by θ–2θ x-ray scan, as well as the real-time monitoring result by RHEED. The peak at $2\theta = 63.488°$ is the typical diffraction of CdTe(224) K_{β}. The wide peak at $2\theta = 45°$ is the background diffraction of $\lambda = 1$ Å, which is no related to any domain.

After growth conditions are optimized during the ZnTe initial nucleation, high temperature, and thick CdTe epitaxy, a twin-free CdTe(211)B layer was reliably obtained on 3 inch Si wafer. In Fig. 3.59, x-ray high-resolution (10 μm) Berg-Barrett reflection topography (BeDe Bescan) on 4.65 μm Si/CdTe(211) demonstrated the good crystalline quality across the whole 3 inch wafer, without any other domain or stacky defects.

(4) Improvement of crystal quality

The heteroepitaxial growth of CdTe on Si poses some extreme difficulties due to the large lattice mismatch (about 19.3 %), thermal expansion coefficient mismatch and the valence mismatch between the substrate and epilayer. These lead to large dislocation densities at the interface and a drastic change in the nature of the bonding at the interface. Double domains and twins are the most commonly observed defects that severely degrade the quality of heteroepitaxial film. Here, the way to obtain twin-free film was discussed. In this section, the experiments on dislocation suppression were carried out, such as the growth condition optimized of ZnTe grading layer and sequential CdTe layer, as well as the misorientation of Si substrates.

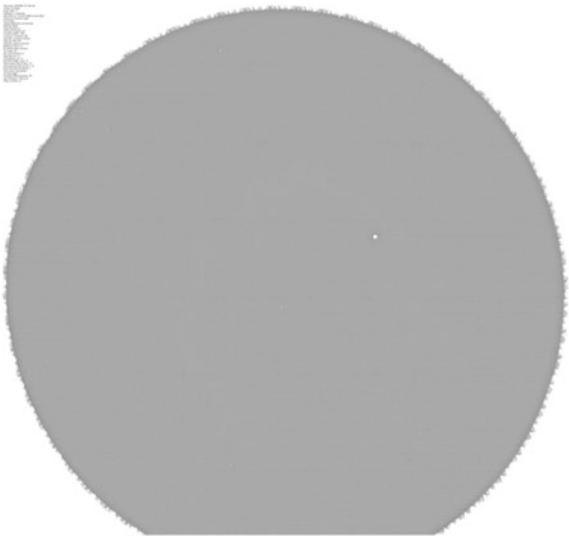


Fig. 3.59 Berg–Barrett reflection topography on CdTe/Si(211)B (sample No. SCT078)

A. ZnTe initial nucleation

The grading layer is commonly introduced to large lattice mismatch heteroepitaxy. A thin ZnTe was initially employed on Si, not only to ensure twin-free growth, but also to greatly affect the dislocation suppression. Many experiments have shown that the crystalline quality was very sensitive to the growth condition of ZnTe nucleation.

(1) Relation of crystal quality and Zn/Te ratio

The substrate temperature greatly impacts the atom sticking coefficient, as well as their migration onto the substrate surface. So the substrate temperature and atoms ratio are the key elements to the crystal quality. For ZnTe nucleation at 200 °C, the effect of Zn/Te ratio on crystal quality was compared. The subsequential 1–2.3 μm ZnTe was grown at 320 °C. It was demonstrated by RHEED on real-time monitor during growth, that the twin ion was greatly suppressed by increasing Zn/Te ratio, which is also corresponded to DCRD results. In Fig. 3.60, crystal quality is sensitively related to Zn/Te ratio at 2–6. When Zn/Te is 7–20, the trend of improvement becomes slowed down. When Zn/Te = 30–32, FWHM result of layer turns to be a little worse.

(2) Relation of crystal quality and nucleation temperature

We have systematically studied the growth of CdTe/ZnTe/Si(211) with different ZnTe initial nucleation temperature at 180–200 °C. For each sample, Zn/Te ratio is 20 and the thickness is about 25 Å, as well as CdTe was grown under the same conditions. In Fig. 3.61, *x*-ray double-crystal rocking curve analysis indicated that crystal quality was obviously improved by decreasing the nucleation temperature. When the nucleation temperature is lower than 180 °C, the improvement effect becomes less. If one is concerned about the 3D growth and twin formation at lower temperature as above mentioned, the most proper nucleation temperature of ZnTe is about 190 °C.

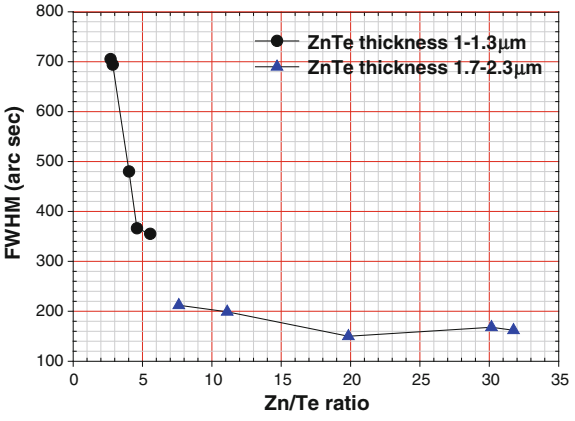


Fig. 3.60 The relationship between crystal quality and Zn/Te ratio

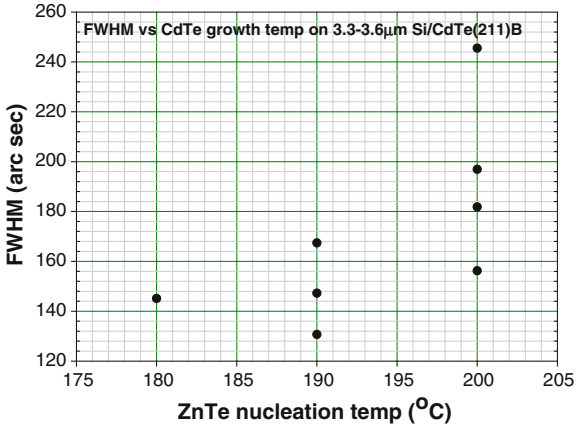


Fig. 3.61 FWHM of XDRC of Si/CdTe(211)B with 3.0–3.6 μm, ZnTe nucleated at the temperature in range of 180–200 °C

(3) Relation of crystal quality and nucleation thickness

It was found that (133) twin fraction could be suppressed by increasing the thickness of the initial ZnTe layer. After all, this thin layer was grown under relatively low temperature, which crucially contributed to lower the interface energy and ensured 2D growth mode. It will apparently impact on the crystal quality. ZnTe nucleation thickness should be compromised, not only for twin suppression, but also for dislocation reduction. The experiments on CdTe/Si(211) or ZnTe/Si(211) with different thickness of 15–200 Å were launched. For each sample, ZnTe nucleated at 190 °C by Zn/Te ratio = 20, as well as CdTe was grown at 280 °C and ZnTe was grown at 320 °C. The crystal quality was evaluated by means of RHEED real-time monitor in Fig. 3.62 and XRD after growth In Fig. 3.63. As Fig. 3.62a, c, e indicated, (133) twin fraction was related to the thickness of ZnTe layer at the range of 15–80 Å. The twin content is much more in the thinner ZnTe, which could compared by the intensity of two domains, as shown in Fig. 3.62a, b. For the sample of SZT044 with about 15 Å ZnTe nucleation, the twin diffraction is still stronger than (211) diffraction, even after annealing at 390 °C after ZnTe nucleation. The XRD result in Fig. 3.63 also demonstrated crystal quality of composited layers with different thickness of initial ZnTe layer. With a thick ZnTe nucleation layer, the intensity of (211) diffraction becomes stronger than the one of (133) twin, and the effect of twin suppression by high-temperature annealing is much more evident. In Fig. 3.62f, twin diffraction almost faded away, in sample SCT049 with 80 Å ZnTe nucleation. Nevertheless, concerning the comparison in Fig. 3.63, the threading dislocations were multiplicated by the thicker ZnTe nucleation above 50 Å, which was examined by FWHM of *x*-ray DCRC. After these systematical experiments and analysis, the optimized thickness of ZnTe nucleation is 25–50 Å.

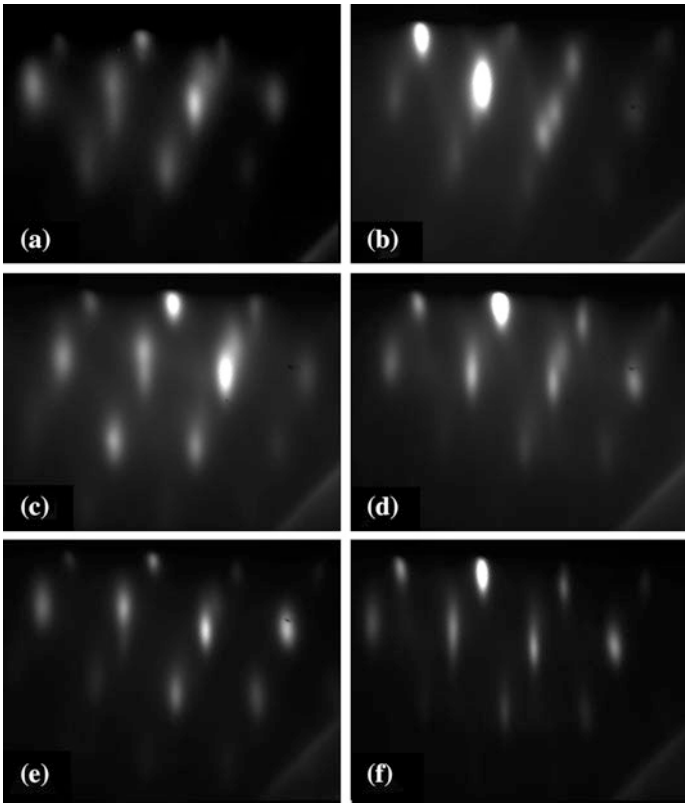
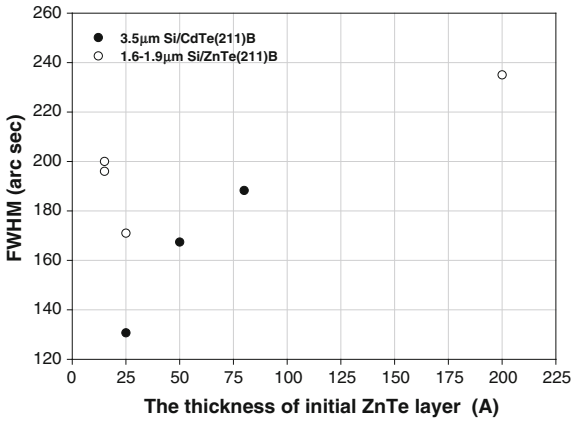


Fig. 3.62 RHEED patterns in $[0\bar{1}1]$ direction of CdTe/ZnTe/Si(211)B. **a** ZnTe nucleation is 15 Å (SZT044), **b** After annealed at 390 °C (SZT044), **c** ZnTe nucleation is 25 Å (SCT041), **d** After annealed at 390 °C (SCT041), **e** ZnTe nucleation is 80 Å (SCT049), **f** After annealed at 390 °C (SCT049)

Fig. 3.63 FWHM XDRC of CdTe/ZnTe/Si(211)B with different thickness



B. Suppression of dislocation by optimization of the tilt angle between Si and CdTe

The density of threading dislocations in CdTe buffer layers is one of the main factors impacting the quality of HgCdTe. In a highly lattice mismatched structure, the threading dislocations are promoted by the misfit strain at the "coincidence interface" (a two-layer composite between the substrate and the epilayer) for releasing the strain energy by extending to the interface to form misfit dislocations. The lattice tilting was observed in highly lattice mismatched systems modifying the strain. Accordingly, the relation of epilayer tilting with lattice mismatch deserves a clarification [129–132].

As shown in Fig. 3.64a, the epilayers grown on lattice mismatched (211) substrates always show some lattice tilting, the surface index of the epilayers deviates from (211) toward to (311) around CdTe[0-11]//Sub[0-11] proportionally to the misfit strain. The result suggests that the tilting is inherently driven by misfit strain to decrease the interface energy, acting as a good scale for the degree of misfit. In an attempt to facilitate strain relaxation by modifying the coincidence interface, a series of CdTe layers were grown on (211) Si and (211)B GaAs substrates misorientated by 1–10° toward to [108] direction. The results are summarized in Fig. 3.64b. It was found that the FWHM of the *x*-ray double-crystal rocking curves (XRDCRC) decreased to saturation as the tilted angle between the (211) planes of CdTe epilayer and Si decreased from 4.2° to 2.75°. No remarkable change in the

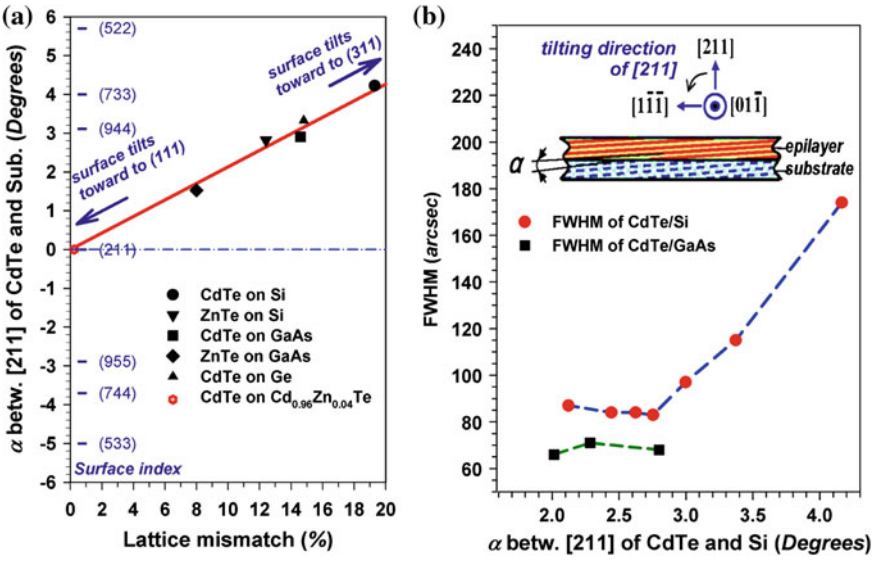


Fig. 3.64 Lattice tilting of epilayers, where **a** dependence of the tilted angle between the (211) planes of CdTe and standard (211) substrates with the lattice mismatch, **b** dependence of the FWHM of (422) XRDCRC with the tilted angle between the (211) planes of CdTe epilayers and the (211) misorientated substrates. The insert in **b** shows the tilting configuration

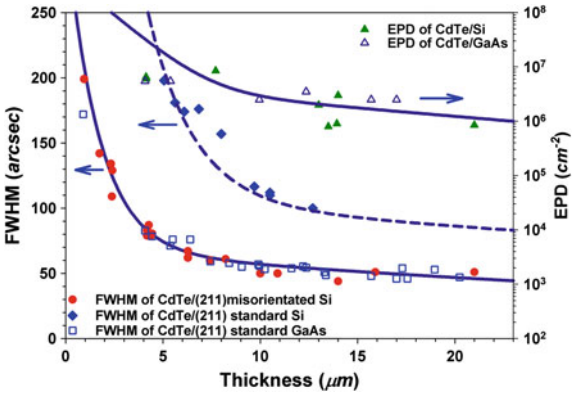


Fig. 3.65 Dependences of the FWHM of (422) XRDCRC and EPD of CdTe epilayers grown on GaAs and Si with the epilayer thickness

FWHM was observed in the CdTe grown on (211) misorientated GaAs, showing that the modification by substrate misorientation was not effective or significant for the less-mismatched structures like CdTe/GaAs. This also explains the saturation in FWHM as observed in CdTe/Si, where the strain was reduced to a certain saturation below which the modification became less effective. The result indicates that growth on misorientated Si substrates reduces the misfit strain as manifested by the lattice tilting, and consequently the requirement for strain relaxation by dislocations is released as reflected by the decreased FWHM of XRDCRC. This was further confirmed by comparing the crystal quality of the CdTe epilayers grown on (211) misorientated Si with those on (211) standard GaAs as shown in Fig. 3.65, where the FWHM of XRDCRC or EPD profile obtained on CdTe grown on misorientated Si was reduced to an identical level with that on CdTe/GaAs. In Fig. 3.65, the (211) planes of CdTe on Si tilted toward to [108] at an angle of 2.75° having the narrowest FWHM (Fig. 3.64b), which roughly corresponded to a reduced lattice misfit of $\sim$13 % as scaled by Fig. 3.64. The reduction in the threading dislocations with increasing CdTe thickness is due to the dislocation cancelation driven by the residual stress. On the other hand, the CdTe grown on standard (211) Si showed higher EPD values (a factor of about 5) or broader FWHM. In this case, more dislocations are required to relax the misfit strain at the interface of CdTe/Si.

By refining the optimal substrate misorientation and growth conditions, a value in FWHM of XRDCRC in a range of 50–60 arcsec (corresponding to an EPD of 9–30 $\times$ 10^5 cm^{-2}) was routinely obtained on CdTe grown on Si and on GaAs, which is the best result reported to our knowledge for CdTe grown on Si. HgCdTe of different compositions were subsequently grown on CdTe buffer layers via a graded composition region of HgCdTe to reduce the effect of the lattice mismatch (0.2 %) between CdTe and HgCdTe. The FWHM of XRDCRC varied in a range of 55–75 arcsec with the EPD of 1–5 $\times$ 10^5 cm^{-2}, as statistically observed in many runs. The quality of HgCdTe grown on Si was found to be identical with that grown on GaAs.

C. The optimal CdTe growth condition

The proceeding results on dislocation reduction related to substrate tilt assume that an optimized growth sequence should be established. By real-time RHEED monitoring, CdTe growth condition also has an influence on the density of threading dislocation. To determine the appropriate CdTe growth temperature, a series of CdTe layers were grown at 274, 295, 303, 308, 313, and 328 °C. At about 25 Å ZnTe, nucleation was grown at 190 °C under the optimized conditions as previous process, on Si (211) tilted toward to [108] at an angle of 2.75°. Then, CdTe was grown at the temperature of 270–330 °C.

Figure 3.66 records the FWHM DCRC of the series samples at different growth temperatures, which were measured for CdTe (211)B on Si of various epilayer thickness. It is obvious that the better quality film was obtained under the higher temperature. If with concerning of films morphology, the best quality epilayer was grown at 295 °C.

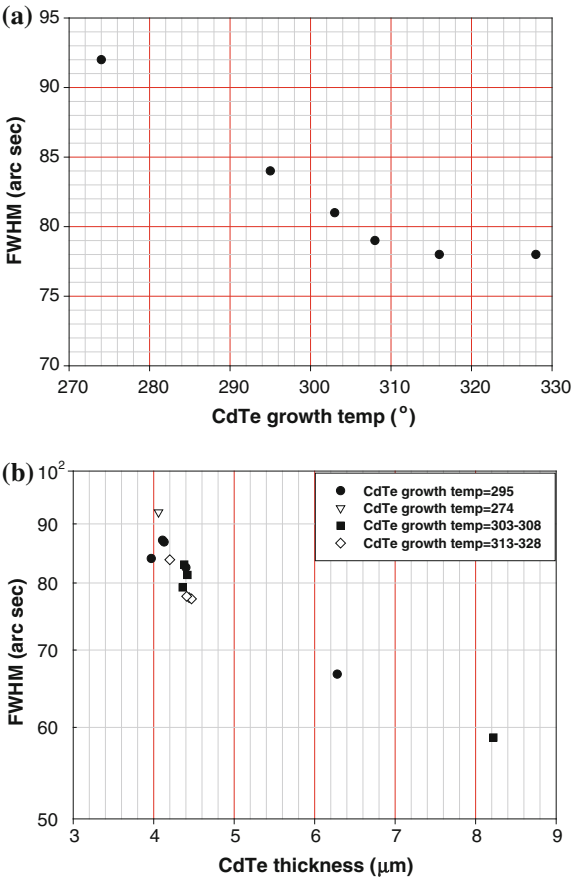


Fig. 3.66 The FWHM XDRC of CdTe/ZnTe/Si(211)B grown at different temperature. **a** 4–4.4 μm CdTe/Si (211)B grown at different temperature, **b** the relationship between crystal quality and thickness of CdTe layer, with the series samples grown on different temperature

3.3.2 HgCdTe Growth on Large-size Alternative Substrates

1. Precise control of growth temperature

The wafer size increasing from 2 inches to 3 inches for the HgCdTe epitaxy brings a great challenge for HgCdTe MBE technology. The key issue focused on the precise control of growth temperature and uniformity across the whole wafer. Compared with other III–V or II–VI wide band gap epitaxy, the optimized growth window is very narrow, due to the stacking coefficient of Hg being very low and very sensitive to growth temperature, which has to be in the range of less than ±2 °C. It is observed that even quite small fluctuations of growth temperature will greatly influence the morphology of the epilayer, such as the density and size of surface defects [133], which would seriously affect the application in the focal plane arrays.

Usually, the substrate, less than 2 inches, was indium-bonded on a molybdenum holder. Lateral uniformity of the substrate temperature depended on the thermal conduction through indium and the holder. It is easy to employ the infrared radiation thermometer as a real-time monitor on growth temperature. Unfortunately, the wafer mounting mode should change to indium-free with wafer size larger than 3 inches. Indium-bonding is not proper any more to ensure the adherent and temperature uniformity of large-size substrates. In addition, due to the difference in thermal expansion coefficient between the substrate material and the moly holder, it will result in a large thermal stress inside the epitaxial material. Furthermore, indium as a doping impurity was observed by SIMS inside films. The main challenge of HgCdTe growth on substrates larger than 3 inches is the mode of wafer mounting and temperature monitoring.

To avoid thermal stress problems, and to ensure the lateral uniformity of the substrate temperature, 3-inch indium-free mounting mode is performed. The substrate in the Mo holder is under a stress-free state, and is heated by radiation from the backside heater directly. Because the alternative substrate is infrared transparent (from 1.1 to 10 μm), the real temperature of the substrate or HgCdTe material surface cannot be measured accurately by an infrared thermometer due to thermal radiation interference from the substrate heater.

Currently, a noncontact thermocouple was used on manipulator, on alternative substrates of 3 inches or area plus. The issue of substrate temperature measurement is more complex. During the beginning of HgCdTe growth, the emissivity and infrared absorption factor of epilayer can be constantly changing, leading to thermal loss and absorption at the same time. The temperature value obtained from the noncontact thermocouple is a result of a combined effect, consisting of substrate heater radiation, infrared absorption, and radiation of epilayer. The composition, the roughness of epilayer surface, as well as the thickness of layer, are the factors which mostly influence the T/C reading.

There are many studies on experience in temperature control of HgCdTe growth using indium-free Mo holder. Bajaj et al. [134] used a thermocouple contacting

with the Mo block mechanically, while it comes with complexity for MBE heating system, greatly limits the rotation of the substrate during the growth, thus restricting the uniformity improvement of large-area material. At the same time, mechanical contact between the thermocouple and the substrate holder makes the thermocouple short-lived, consequently shortening the actual running time of MBE equipment.

Ferret et al. [135] in order to solve the problem of temperature control, used a special graphite substrate holder mechanically fixed to the substrate. However, this method is difficult to ensure the lateral temperature uniformity of a large-area substrate. To avoid this problem, Brill et al. [136] used a real-time observation of high-energy electron diffraction, summing up the empirical temperature curve of the noncontact thermocouple during growth. However, the problem with this method is that the high-energy electron diffraction image is not very sensitive to small fluctuations of temperature, thus difficult to guarantee reliable growth temperature measurement or stable feedback control. Almeida et al. [137] used an ellipsometer for real-time monitoring. The differences in the growth temperature will result in changes of Hg sticking coefficient, and a small deviation from the proper HgCdTe growth window will directly affect combining Hg atoms with Te atoms, eventually cause on Cd composition, and therefore the temperature control can be achieved by real-time composition measurement. However, the accuracy of this temperature control method is not high, small temperature fluctuations in the growth window cannot be exactly obtained by composition variation.

In our studies [138], adding the infrared absorption layer is an effective method to solve the above problems. The substrate temperature could be accurately measured by IR pyrometer, which is the most effective way to ensure the precise control on the real-time. The thermocouple is in noncontact mode, while avoiding the problems of complex relationship between the T/C reading and the real temperature of HgCdTe surface. It is not affected by different composition and growth rate of HgCdTe, suitable for HgCdTe growth with cutoff wavelength from shortwave to long wave. The maximal deviation of temperature control is less than ± 1 %, well meeting the requirements of the high-quality HgCdTe epitaxy. In addition, this method is applicable regardless of substrate size or type.

2. Improvement of lateral composition uniformity on large-area HgCdTe

A stringent requirement from the FPAs of large formats is the lateral uniformity of material properties concerning crystal quality, surface defects and composition. The composition uniformity has become an important issue for large-area HgCdTe. It was found that the uniformity across the wafer was sensitive to the geometrical configurations of flux cones projected onto the wafers as well as the level of crucible charging and its evolution, and was a tradeoff between the growth rates and flux consumptions. To improve the composition uniformity, some demanding factors were quantified by many growth runs. Examples of radical distributions of x values and the projecting cutoff wavelengths were obtained on epiwafers grown on Si and GaAs. By refining the processes of both flux measurements and the amount of charging material in crucibles, a significant improvement in composition

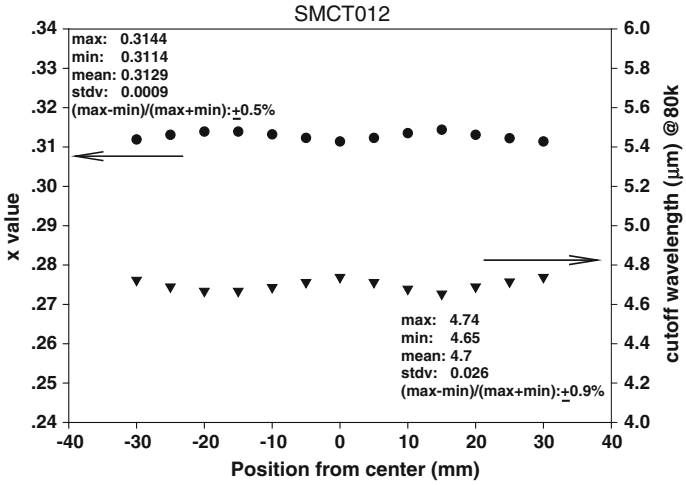


Fig. 3.67 The radial distributions of both x values and the converted cutoff wavelengths at 80 K over the 3-inch epiwafers of HgCdTe on Silicon

reproducibility was achieved. For long wavelength (LW) applications, (a mean x value of 0.229), a standard deviation (STDDEV) in x value of less than 0.0017 was obtained in a run-to-run base. As shown in Fig. 3.67, a yield of cutoff wavelength of 73 % was obtained as screened by a deviation of less than 0.2 μm from the targets. The radial distributions of both x values and the converted cutoff wavelengths at 80 K over the 3-inch epiwafers of HgCdTe on Silicon are shown in Fig. 3.67. The result was obtained by The Fourier Infrared Transmission Spectroscopy (BRO-RAD STF-65A), and the size of the light spot used is 50×50 μm^2. The thickness of each layer is obtained by comparing the transmission curve with the theoretical curve getting from computer simulation, and the HgCdTe composition determined according to 500 cm^{-1} position of absorption coefficient. The HgCdTe epilayers showed good lateral uniformity in both composition and thickness. It shows an example of composition or cutoff uniformity for a 3-in. HgCdTe wafer grown on Si, the maximum deviation in cutoff wavelength at 80 K is less than 0.1 μm over the central area of a diameter of 70 mm, satisfying the requirement by the FPAs fabrication.

3. Surface defects

The surface defects are responsible for the degradation in the operability of imaging devices. Because of the very low sticking coefficient of Hg, the nucleation of surface defects is a sensitive function of the growth condition employed. In order to minimize the density of surface defects, the origin of defect nucleation as well as its relation with growth condition should be clarified. The surface quality of HgCdTe is also affected by the propagation of the defects in CdTe buffer layers.

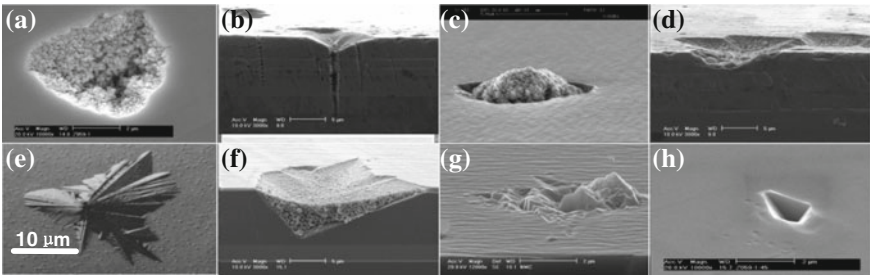


Fig. 3.68 SEM and Normaski micrographs of surface defects. **a** top view of type3 defects, **b** cross-section view of type3 defects, which formed on the surface of the substrates, **c** the hill-locks defects originated from the impurities on the surface, **d** SEM of type 1 defects, which related to insufficient Hg conditions, **e** OM of type1 defects, **f** cross-section view of type1 defects, nucleation occurred during the growth process, **g** type4 defects, related to the Hg-rich conditions, **h** type5 defect, diamond-shaped voids

The features of various surface defects were studied by using scanning electron microscopy (SEM), energy dispersive *x*-ray fluorescence spectroscopy (EDX) as well as atomic force microscopy (AFM) on HgCdTe epilayers grown at different conditions. The EDX analyses were performed by using standard-less ZAF corrections (atomic number, absorption and fluorescence coefficients) for qualification. The nucleation and evolution mechanism of different surface defects in HgCdTe thin films were analyzed. The relationship between growth conditions, surface treatment, and other factors with different surface defects was summarized and analyzed from a series of experiments. The relation of defects to the growth temperature and the flux ratio was obtained in experimental curve.

Some typical defects in CdTe were systematically identified in this work, and are not being described here as limited by pages. As shown in Fig. 3.68, various kinds of surface defects were observed on the HgCdTe surfaces. Some of these defects were voids, some were hillocks, and some were a mixture of voids and hillocks. The origin of these defects is either substrate-related or growth-related or both substrate and growth-related. Figure 3.68a–c show substrate-related defects (type 3); (d) shows voids related to Hg-deficiency (type 2); (e)–(f) show the voids when Hg is further deficient (type 1); (g) shows defects related to Hg enrichment (type 4), and (h) shows the void sometimes observed when the growth temperature was lowered from the optimal condition by ~ 2 °C (type 5). For FPA applications, obviously type 1–3 defects are more problematic because they are large in size and may penetrate deep into the film. Even one such a defect may completely kill the pixel. By efforts devoted to optimize each process starting from the substrate preparation to the setting up and control of growth parameters for CdTe and HgCdTe, our accumulated data showed that densities of the surface defects (≥ 2 μm) on HgCdTe were reduced to below 300 cm^{-2}.

4. Overall performances of HgCdTe on silicon

(1) Crystal quality

The growth condition was optimized by the efforts on dislocation reduction of CdTe/ZnTe/Si, including tilted toward [108] Si substrate, lattice grading layer epitaxy, and high-temperature annealing. The serial experiments of CdTe/Si growth were carried out on a run-to-run basis. The crystal quality was evaluated by full width at half maximum (FWHM) of the *x*-ray double-crystal rocking curve measurements (XDRC). Figure 3.69 shows FWHM of CdTe epilayers grown on GaAs and Si versus epilayer thickness. The epilayers grown under optimized process showed narrower FWHM than that before optimization.

Figure 3.69 also indicates the dislocation density in the epilayer under different thickness. Generally, XDRC FWHM will decrease with the increasing thickness of the epilayer. The curve can be divided into two zones, which is discussed for misfit dislocation at the interface and threading dislocation inside the layer, respectively. Considering the *X*-ray penetrating depth is about 3–4 μm, FWHM values can effectively reveal the misfit dislocation density at the interface of CdTe/Si epilayer with 3 μm thickness. The effect of misfit dislocation reduction at the interface was demonstrated by the slope of FWHM value decreased. In Fig. 3.69, it is illuminated that large amount of misfit dislocations at the interface have been greatly reduced after process optimized. When the thickness is more than 4 μm, threading dislocation under the certain growth condition could be clarified clearly.

For heteroepitaxy with a large lattice mismatch, relatively higher interface energy can be released by producing high-density misfit dislocations. When the thickness of the epitaxial layer is less than the penetrating depth of *X*-rays, the high-density dislocations at the interface are reflected through XDRC FWHM values. With increasing thickness, misfit dislocations in the material tend to slip and form penetrating dislocations. Since the penetrating dislocation density within the thickness range is high, dislocations are easy to meet each other and form closed loops through the moving process in the material, so that the dislocation density is

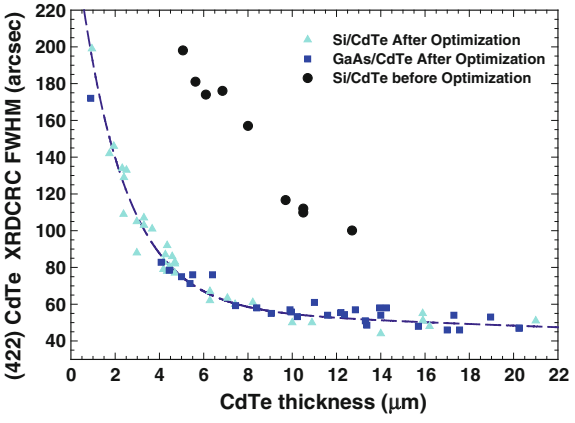


Fig. 3.69 Dependence of the FWHM value of (422) XDRC of CdTe epilayers grown on GaAs and Si with the epilayer thickness

decreased, and therefore produce a sharp drop of XDRC FWHM value in the smaller range of epilayer thickness.

When the epilayer thickness exceeds the *X*-ray penetrating depth, the information of the high-density misfit dislocations at the interface gradually decreases in the diffraction proportion. With further increase in the epilayer thickness, XDRC FWHM values decrease constantly but slowly. This is because the dislocation spacing increases in the material and the chances of forming closed loops through dislocations intersecting during the moving process is significantly reduced after the dislocation density decreases rapidly at the abrupt drop zone. So when the thickness reaches a certain level, XDRC FWHM values have a very slow decreasing trend and become nearly constant.

Figure 3.70 shows the XDRC FWHM values have good correlation with the etch pits density (EPD) values of HgCdTe grown on GaAs or Si. When XDRC FWHM values are over 70 arcsec, the EPD value reduces with the decreasing XDRC FWHM value; while XDRC FWHM values drop to about 40–60 arcsec, the decreasing trend of EPD values become slowly. It is well known that EPD obtained by chemical etching method can be a most direct evaluation of the dislocation density in the material, but this evaluating method is destructive comparing with the XRD measuring method being nondestructive. Therefore, obtaining the experimental curve like in Fig. 3.70 is positive to the establishing criterion of the nondestructive evaluation.

To increase the available area of the HgCdTe epilayer on an alternative substrate for device applications, attention should be paid to the lateral distribution of the crystal quality of the large-area epilayer and the effective improvement of the overall crystal quality level should be done. It is necessary to focus on crystal quality uniformity in order to improve the performance and uniformity of the device. Using the mapping of XDRC FWHM to evaluate the crystal quality uniformity of 3-inch epiwafers is a relatively effective mean.

Figure 3.71a, b show the distribution of the XDRC FWHM values of two 3-inch CdTe(422) epilayers noted SCT068 (central thickness 4.4 μm) with SCT097

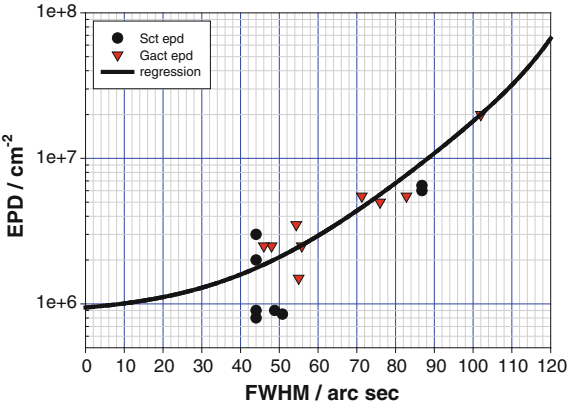


Fig. 3.70 The relationship between the FWHM and the etch pits density (*EPD*) of HgCdTe on GaAs and Si

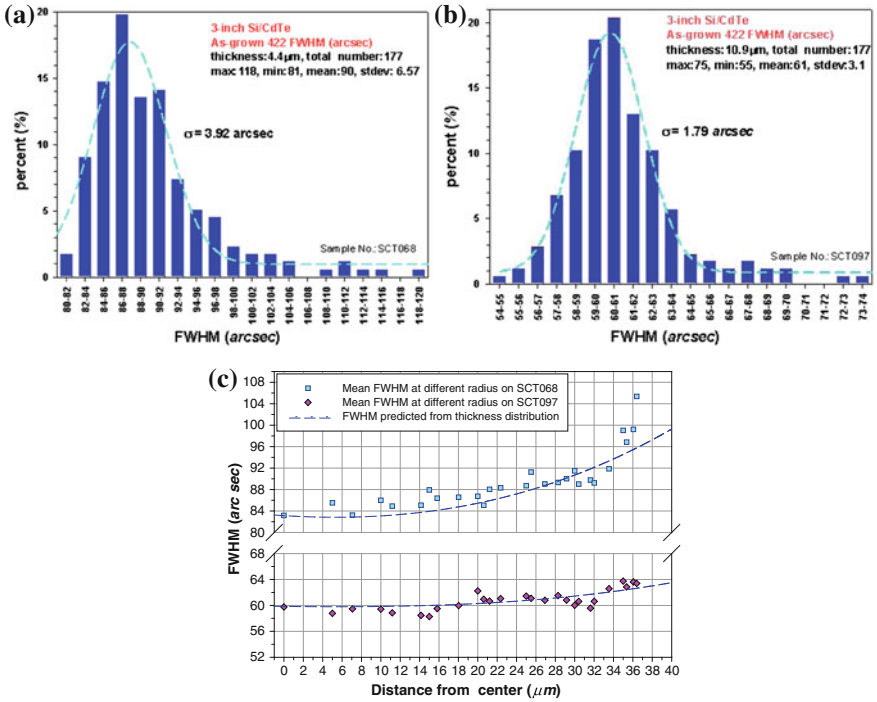


Fig. 3.71 Distributions of the FWHM values of (422) XDRC of 3-inch CdTe epilayer on Si. **a** 4.4 μm CdTe, **b** 10.9 μm CdTe, **c** radial distributions across wafer

(central thickness 10.9 μm). By Gaussian function fitting, the Gaussian standard deviation σ can be obtained, and the standard deviation (stdev short for standard deviation) of the FWHM values is also given in the figure. Figure 3.71c is the radial distribution of FWHM values, which come from averaging the FWHM values at the same radius. Due to the shielding effect from the Mo ring of the indium-free substrate holder, about 2–3 mm ring area at the edge of the 3-inch epitaxial layer is not covered by material or is covered by poor-quality material, resulting in FWHM values at the edge being abnormally larger and relatively random. The abnormal values have a great impact on the stdev but little effect on the Gaussian standard deviation σ. Hence, Gaussian standard deviation σ is more valuable for evaluating crystal quality uniformity of a 3-inch epilayer than stdev.

It is observed from Fig. 3.71a, b, the σ value of 3-inch Si/CdTe with the 4.4 μm central thickness is approximately twice that with the 10.9 μm central thickness. Furthermore, the FWHM values of both samples increase in the radial direction and the former is significantly higher than the latter, though the statistical FWHM data in Fig. 3.71c are relatively dispersed. The possible reasons for greater σ value and the FWHM values increasing radially, just include ZnTe nucleation quality determined by the rigorous conditions, the lateral uniformity of the substrate temperature

during the growth process, and the CdTe thickness uniformity decided by the beam source geometrical broadening state, etc. In particular, the radial thickness uniformity of the CdTe has the most dominant and direct effect on the radial uniformity of the FWHM values, shown by the dashed line in Fig. 3.71c, which is the expected FWHM values obtained by fitting the correlation of the CdTe FWHM and CdTe thickness according to line shown in Fig. 3.69.

Figure 3.71c shows the expected FWHM values changing with the radial distance, which basically have the same trends in the radial direction with the experimental FWHM values, indicating, in a relatively smaller CdTe thickness case, the radial distribution of the CdTe thickness plays a dominant role on the radial distribution of the FWHM value and the same for the Gaussian standard deviation σ value.

Mapping of XDRC FWHM values of HgCdTe epilayer (SHGCDTE011, $x = 0.31$) grown on Si is shown in Fig. 3.72. The thickness of CdTe buffer for this sample is 8 μm and HgCdTe layer is 10 μm having the best result for XDRC FWHM values of 51 arcsec. Currently, silicon-based HgCdTe materials have already been fabricated in focal plane device application.

2. Distribution mapping of Surface defect on large-size wafer

The surface defect distribution statistics is considered as conventional evaluation of surface defects for 3-inch GaAs/HgCdTe and Si/HgCdTe epilayers. Furthermore, the analysis of surface defects distribution of GaAs/CdTe and Si/CdTe epilayers is also presented in order to monitor surface defects in process and study their evolution from CdTe layer to HgCdTe layer. Detailed data for surface defects can be obtained by scanning and analyzing the full-spatial 3-inch surface with the macro-programs of the computer software. Different types of diagrams and tables can be sorted and achieved.

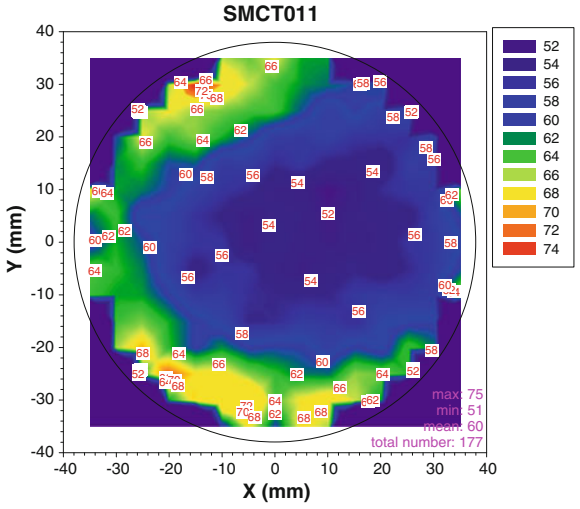


Fig. 3.72 Mapping of XDRC FWHM values of HgCdTe epilayer ($x = 0.31$) grown on Si

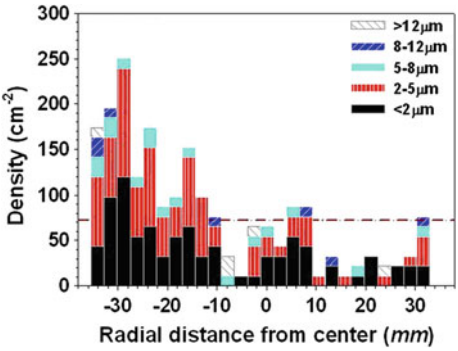


Fig. 3.73 The surface defect density of 3-inch CdTe epilayer on Si with $x = 0.31$

The surface defect distribution of a 4.35 μm CdTe epilayer grown on 3-inch Si is shown in Fig. 3.73. The CdTe surface defects exhibit the regionally distributing characteristic. The defects density whose sizes are less than 8 μm is highest in radius from 0 to 10 and 15 to 25 mm, being classified to the particular defects for Si-based material. There are three defect density anomalies at the sample's edge, corresponding to manual damage. The figure indicates 80 % of the defects have a size smaller than 8 μm. The number of the defects with size larger than 20 μm is 36, which is remarkably larger than that of GaAs-based material. This is mainly related to the special and complex Si substrate pretreatment, in which dust is more easily to be introduced.

3.3.3 Extrinsic Doping in HgCdTe by MBE

The third-generation infrared imaging systems require the $Hg_{1-x}Cd_xTe$-based IRFPAs of increased formats and with multicolor sensing capabilities. A major reason for requiring doping control involves the development of graded heterojunctions and heterostructure homojunctions. It is important to properly design and grow the heterostructures. The capabilities of in situ formation of pn junctions are a crucial technology and are extensively studied by many groups worldwide. The main factors of this technology include extrinsic doping and impurity activation. The requirement here is the simultaneous achievement of the optimum heterostructure and in situ doping profiles. MBE offers distinct advantages for achieving these requirements. The research of doping technology in MBE HgCdTe was studied systematically in this section, which involved evaluation of impurity and activation ratio, in situ N-type Indium doping, P-type As-doping, and the ex situ activation annealing of As. In addition, the experiment of ex situ As activation was carried out under Te-rich vapor in open-tube, which was an effective way for the fabrication of large-format IRFPAs application.

1. Impurity and activation ratio measurement

Secondary Ion Mass Spectrometry (SIMS) is normally used for doping concentration measurement, which is an important method for surface analysis. Surface atoms are ionized and sputtered into vacuum with the kinetic energy of the incident ion. The surface element can be identified by the charge-mass ratio of the secondary ion, collected by mass spectrometer.

Compared to other analysis measurement, SIMS has many advantages, such as: 1. High analysis sensitivity of ppm content. 2. Good finger-print identification property, even slight difference of element charge-mass ratio. 3. Accurate surface analysis, for those secondary ions come from the very surface layer. 4. Beam spot up to several μm, which is used to obtain three-dimensional distribution of the surface element.

The main disadvantage of SIMS is that the primary ion will cause surface defects, such as: 1. crystal plane disorder. 2. Most of the primary ion being implanted into the sample. 3. Leading to the adsorption of residual gas and increasing the degree of contamination. and 4. Changing the distribution of the original elements. Although there is some weakness above, it still occupies a prominent position in the surface analysis due to its high sensitivity and good surface properties.

SIMS is divided in two modes: static and dynamic. The main difference is the distinction of the sputtering rate. Dynamic SIMS is used for depth analysis of the surface composition. A primary ion is used for etching source, without considering preferential sputtering effect.

Standard sample method is usually used for concentration calibration by SIMS. Research results of Deline [139], Willians [140], Yu [141] show that: the relative ratio of secondary ion intensity of different elements sputtered from the same substrate is generally the same. Standard sample for calibration of elements is to take advantage of this. A sample whose composition is known is used as a standard in order to obtain the relative sensitivity factor (RSF) of the element to be analyzed. After that, each element of the sample to be measured can be calibrated.

When using the standard sample method to determine the concentration, RSF of certain element should be measured first.

$$\mathrm{RSF} = (I_\mathrm{m}/I_\mathrm{i}) \times C_\mathrm{i} \tag{3.13}$$

where

I_i is the secondary ion intensity of impurity elements;
I_m is the secondary ion intensity of matrix elements;
C_i is Impurity concentration of standard sample.

After RSF was obtained, the impurity concentration can be expressed as below:

$$C_\mathrm{i} = (I_\mathrm{i}/I_\mathrm{m}) \times \mathrm{RSF} \tag{3.14}$$

I of the equations above was the secondary ion intensity of impurity elements and matrix elements of sample to be analyzed.

Because Te content of HgCdTe is always 50 %, not varying with the changes of composition, it is considered as the reference value of RSF data.

Preparation of standard sample is important in SIMS test. Besides the concentration requirement of unknown element, matrix composition and structure of the standard sample should be similar to that of the test sample, so ion-implanted sample is generally used as a standard sample.

For ion implantation sample, implanting ions in the sample shows a Gaussian distribution. Two ways are used to fit the peak concentration. One is fitting the implantation dose and concentration which is measured by SIMS test. The other is using simulation implantation procedures to obtain the injection depth and FWHM and calculating the peak concentration.

These two methods are introduced below:

Dose Method

Under certain experimental conditions, the relationship between ion current $I(t)$ at time t and implantation concentration $N(x)$ can be described as $I(t) = kN(x)$. Assuming the etching rate is a constant, which is expressed as $dx/dt = X/T$, ion flux intensity $Is(T)$ in the total bombardment time (T) equal to $Is\ (t)$ of the integral, which can be expressed as below:

$$Is(T) = \int_0^T Is(t)\mathrm{d}t = k\frac{T}{X}\int_0^X N(x)\mathrm{d}x$$

Implantation dose D is defined as:

$$D = \int_0^X N(x)\mathrm{d}x$$

Then,

$$I(t) = \frac{X}{T}\frac{Is(T)}{D}N(x)$$

Translate integration into summation:

$$N(x) = \frac{1}{X/T} \times \frac{D}{\sum I(t) \times \Delta t} I(t) \tag{3.15}$$

Implantation Fitting Method

According to Gauss equation and mathematical derivation, the formula can be expressed as below:

$$N_{\max} = \frac{N_0}{\sqrt{2\pi} \times \overline{\Delta R_P}} \tag{3.16}$$

where

$N_{\max}$ is peak implantation concentration;
N_0 is implantation dose;
$\overline{\Delta R_P}$ is standard deviation of implantation depth.

Jack [142] studied the SIMS detection limit of various elements in HgCdTe and reported that the detection limit of As is 5×10^{15} cm^{-3} when using O^{2+} detection as primary ion source, which means that As-doping concentration is only credible in SIMS test when the doping concentration is higher than this limit.

In order to calculate the activation ratio of As-doping, the acceptor concentration is also needed, which means that the carrier concentration of P-type material should be confirmed by the Hall test. Impurity activation ratio equals doping concentration to the carrier concentration.

2. Indium doping and activation

Generally, the electrical properties of HgCdTe material change by adjusting the stoichiometric ratio (intrinsic doping). Hg vacancy in HgCdTe is considered as an electrical acceptor and hole concentration can be controlled by adjusting Hg vacancy concentration through different annealing methods [143]. After the Hg vacancies are eliminated by heating treatment under Hg-rich atmosphere, the donor concentration of intrinsic-doped HgCdTe is determined by the remaining impurity acted as donor. Therefore, intrinsic doping by a background donor of HgCdTe faces the technology limits, in which n-type carriers concentration could not be obtained over a wide range. Moreover, HgCdTe focal plane devices require vertical integration of different wavelength response HgCdTe p–n junctions, which means that electrical parameters of each layer should be determined layer by layer during HgCdTe growth. It is obvious that intrinsic doping cannot meet the requirement of flexible multilayer growth any more.

Over the past decade, N-type HgCdTe doping technology has been studied extensively. Indium is considered as an ideal source of N-type doping and has good thermal stability in HgCdTe. The diffusion coefficient of In in HgCdTe is shown in Fig. 3.74 [144]. In the following section, In-doping process in HgCdTe are described including activation, Hall analysis, and In diffusion behavior in HgCdTe as well.

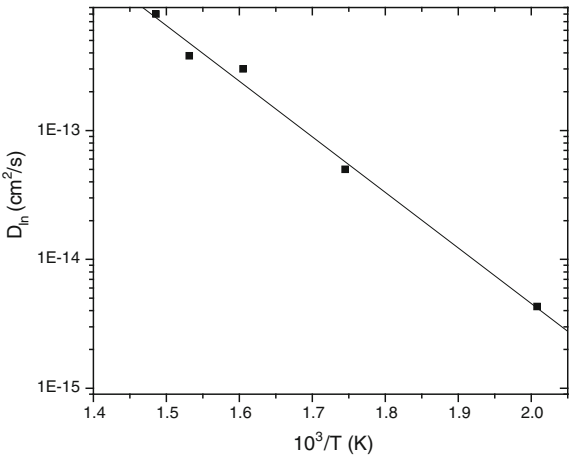


Fig. 3.74 Relationship between diffusion coefficient and temperature of In in HgCdTe

(1) Experimental study of In doping in HgCdTe

HgCdTe films are grown in a Riber 32P system on GaAs (211)B substrate. High-purity elemental In (7 N) is used as an In-dopant source. Before HgCdTe growth, a CdTe buffer layer thicker than 3 μm is introduced to reduce the mismatch dislocations and prevent substrate impurity from concentrating into HgCdTe layer during postgrowth annealing.

HgCdTe of Cd x = 0.21–0.24 was grown at about 190 °C, then annealed under 240 °C for 8 h with saturated Hg vapor pressure to eliminate Hg vacancies. After growth, the samples were diced to 10 mm × 10 mm pieces, and annealed by closed-tube annealing treatment. The mercury source was placed at one end of the quartz tube and the sample at the other end. Hg pressure can be controlled by adjusting the temperature of Hg source. To avoid other impurities, before the experiment, the annealing quartz tube, stopper, and sample holder were cleaned thoroughly. The Cleaning procedure was sequentially carried out by organic cleaning, etching, and flushing by DI water. Then, dried by N2. The samples were annealed at 240 °C for 48 under Hg-rich vapor pressure. Electrical parameters of the annealed sample are measured by Van Der Pauw four electrodes method.

Boukerche et al. [145] reported that the electrical activation ratio of In is related to the In-doping concentration through doping experiments of HgCdTe (111) film on the (100) GaAs substrates by MBE method. The maximum activation rate is about 60 % when the In-doping concentration is around 1×10^{18} cm^{-3}. When the doping concentration decreased or increased, activation rate dropped to 10–40 %. Activation ratio can reach 100 % by optimizing growth conditions [146, 147].

Our results were verified in Fig. 3.75, which indicated a good linear profile of relationship between the estimated In concentration by SIMS and carrier concentration (N_D–N_A) by Hall measurements at 77 K. The activation ratio of In is close to 100 % when the doping concentration range is from 2×10^{14} to 1×10^{17} cm^{-3} (FIG diagonal line is the 100 % activation ratio line). Indium can be considered as

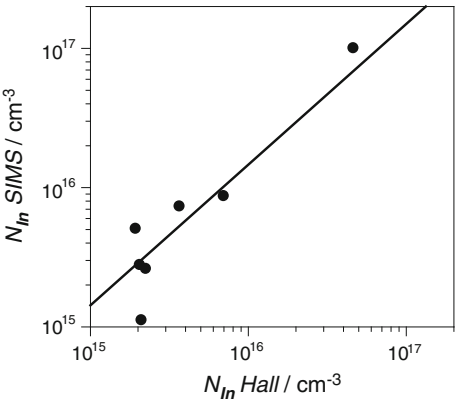


Fig. 3.75 Relationship between concentration of In atom by SIMS and carrier concentration ($N_D - N_A$) by Hall measurement

donor impurity by replacing group II metal atom in HgCdTe. Due to many Hg vacancies in HgCdTe, it is easily for In to enter a Hg vacancy and such a high In electrical activation ratio is understandable.

(3) Hall measurements and activation energy

Typical electrical properties of In-doped HgCdTe films are shown in Fig. 3.76. Relationship between the carrier concentration $n = N_D–N_A$ or mobility and temperature is represented, which reveals that there is no anomalous Hall effect caused by multi-carrier composition [148, 149]. At around room temperature, the intrinsic carrier is dominant. Carrier concentration decreased rapidly when temperature drops. At around 100 K, the extrinsic carrier begins to be dominant. The carrier concentration is temperature independent less than 100 K and does not decrease when temperature is further reduced, indicating that the donor activation energy of In atoms is small and its donor level is close to the bottom of the conduction band of HgCdTe. Even in the very low temperature of 7 K, carrier do not freeze out, demonstrating that In atomic activation energy should be at least less than 7 KT, particularly 0.6 meV. In the high-temperature region, the carrier mobility decreased monotonically when temperature increased. When the temperature is lower than 30 K, mobility becomes saturated and temperature independent. For the high-doping concentration (5×10^{16} cm^{-3}) sample, carrier mobility is 7×10^4 cm^2/V s. When the concentration decreased to 2.3×10^{15} cm^{-3}, the long wave ($x = 0.21$–0.24) HgCdTe electron mobility is 7×10^4 to 2×10^5 cm^2/V s. In addition to the low carrier concentration sample of g146 and g150, mobility at low temperatures decreased when doping concentration increased, which indicates that In^+ ionized impurity scattering dominates. For binary semiconductor materials, a mobility peak can be found in mobility–temperature curve which is caused by the switch from high-temperature phonon-dominant scattering to low-temperature impurity-dominant scattering. As Fig. 3.76b shows, the mobility peak does not occur in In-doped HgCdTe material. Mobility "platform region" in the low-temperature is a common feature in high-purity HgCdTe material [150]. This

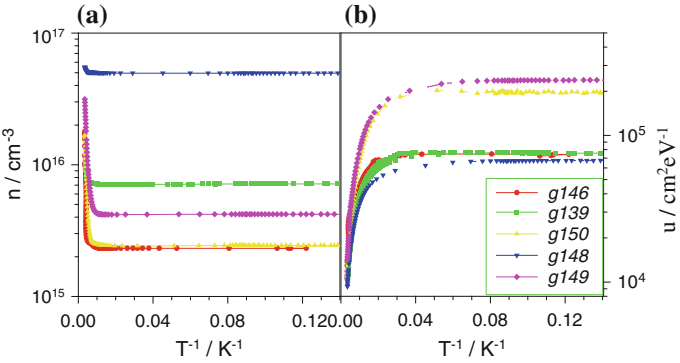


Fig. 3.76 Electronical property of In-doped HgCdTe by Hall temperature-dependent measurement. **a** carrier concentration $n = (N_D - N_A)$, **b** mobility

may be due to alloy scattering or disordered scattering at low temperature that could not be ignored. Mobility–temperature relationship caused by this scattering is in contrast with that caused by ion impurity scattering, which shows that mobility decreases when temperature increases [151].

Contradicting the explanation that ionized impurity scattering dominates at low temperature, although doping concentration of sample g146 ($x = 0.237$) is low (Fig. 3.76b, 2.3×10^{15} cm^{-3}), the low-temperature mobility does not increase accordingly (7×10^4 cm^2/V s).

In some experiments, intrinsic-doped HgCdTe material with lower concentration has a lower mobility. Edwall, Temofonte et al. [152] also reported that the electron mobility is at a low level, although the carrier concentration is only $1–2 \times 10^{15}$ cm^{-3}.

Figure 3.77 shows the relationship between electron concentration or mobility and the temperature of intrinsic-doped long-wave HgCdTe after Hg vacancy eliminating annealing. The sample composition x is 0.239. The figure clearly shows that the maximum value of mobility is at around 40 K. The theoretical analysis of

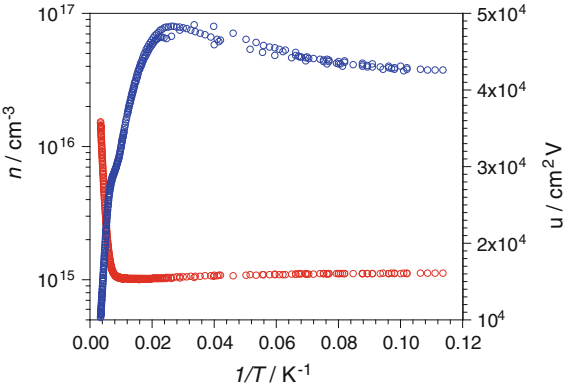


Fig. 3.77 Electronical property of intrinsic-doped HgCdTe by Hall temperature-dependent measurement, after n-type annealing

Meyer et al. [153] indicated that the mobility peak occurs due to the compensation effect of impurities. When electron doping concentration is low, acceptor impurity in material is in an ionized state. Since it is not fully charge compensated by donor electron, the scattering effect is more obvious, which results in a lower mobility. Because donor impurities at low temperature are in a fully ionized state, compensation level and acceptor impurity concentration cannot be obtained from the experimental data. In addition, Rafol et al. [154] found that In doping could improve the uniformity of the material in a Shubikov-de Haas oscillation experiment in order to obtain high electron mobility. Thus, in the preparation of the IRFPA detectors, it is necessary to control the doping concentration of $\sim 3\times10^{15}$ cm^{-3} level in order to obtain high electron mobility.

(4) Indium diffusion coefficient

Stability of In atoms in HgCdTe and diffusion behavior under annealing temperature determines whether it is suitable for device application. It has been reported that an In atom at a lower temperature still has a large diffusion coefficient. Diffusion behavior of In in HgCdTe can be studied by designing and preparing HgCdTe film with mutation distribution and gradient distribution structure of In concentration.

Sample structure is shown in Fig. 3.78. Thickness of CdTe buffer layer is about 3 μm and HgCdTe thickness is about 11 μm. In order to study the diffusion behavior of In at the substrate interface, a 6 μm undoped HgCdTe layer is first grown on the CdTe buffer layer. Then, opening the shutter of In effusion cell, a mutation interface layer of In-doping concentration is formed, whose thickness is 0.3 μm and doping temperature is 480 °C. Then raise In-doping temperature to 510 °C linearly, and a gradient distribution layer of In-doping concentration is

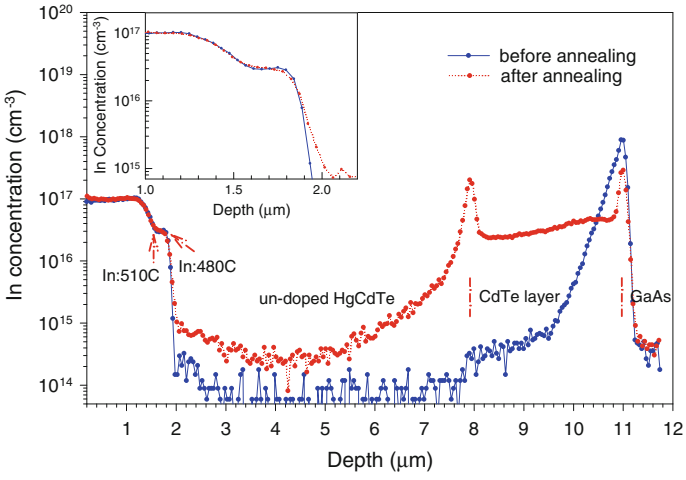


Fig. 3.78 In concentration distribution before and after annealing

formed, whose thickness is about 0.4 μm. Then maintain the In-doping temperature, and a 1.3 μm high In-doping layer was prepared. After being cut as growth completed, some part of HgCdTe material was annealed under 400 °C/30 min and 240 °C/48 h. Annealed and non-annealed pieces were tested by SIMS under the same conditions in order to obtain the original distribution and annealing diffusion distribution of impurity.

The diffusion coefficient can be calculated according to a diffusion equation. Impurity diffusion process under high-temperature annealing is a quasi-static process, which is ruled by diffusion equation:

$$\frac{\partial n}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial n}{\partial x}\right) \tag{3.17}$$

Converting into a numerical equation:

$$C_j^{n+1} = r_j^n\left(C_{j-1}^n + C_{j+1}^n - 2C_j^n\right) + C_j^n + \frac{\Delta t}{4\Delta^2}\left(C_{j+1}^n - C_{j-1}^n\right)\left(D_{j+1}^n - D_{j-1}^n\right) \tag{3.18}$$

In this formula,

$$r_j^n = D_j^n\frac{\Delta t}{\Delta^2}$$
$$j = 1, 2, 3, \ldots, n+1 \quad n = 0, 1, 2, \ldots, N$$

where:

N represents the number of integration. The total annealing time is $N\Delta t$. Epitaxial layer is divided into $n + 1$ layer. j represents the position of the jth HgCdTe layer. Δ is the distance between the jth layer and the $j + 1$th layer. D_j^n represents the diffusion coefficient of HgCdTe layer at time n and position j.

Diffusion coefficient is obtained by program fitting. The result shows that the diffusion coefficient of In at 400 °C in HgCdTe layer is about $\sim 1 \times 10^{-14}$ cm^2/s, which is in accordance with that reported in the literature [155] of $\sim 1 \times 10^{-14}$ cm^2/s. For the In-doped graded layer, no significant change in distribution is observed by annealing. The above analysis shows that In is stable in HgCdTe, which is able to meet the requirements for device fabrication and can be considered as a good dopant.

3 As-doping and annealing technology

During the growth of (211)B HgCdTe, p-type material (Hg vacancies as acceptance) can be obtained by postgrowth annealing method on the intrinsic materials [156]. Carrier concentration could be achieved in the range of 1–20 $\times$ 10^{15} cm^{-3} due to annealing temperature and time. But this has several undesirable traits including Shockley-read-Hall lifetimes associated with vacancies

states mediating recombination. This points to the need for an extrinsic p-type dopant.

The n-type extrinsic doping(10^{15}–10^{18} cm^{-3}) of HgCdTe using indium has been achieved in as-grown MBE layers and can be optimized by a postgrowth isothermal annealing at 240 °C for 48 h [157]. However, the situation is quite different for p-type extrinsic doping, which is much more complex for HgCdTe [158, 159]. For II–VI compounds, both group I and group V elements are p-type dopant candidates. Group I elements, such as Cu, Ag, Au, and Li, serve as p-type dopants if they are incorporated on metallic sites. Unfortunately, group I elements diffuse too fast even at growth temperature and during annealing. It is not the best option for reliable doping. Group V elements, such as As, serve as p-type dopants if they are incorporated on nonmetallic sites, which are stable and well-controlled. However, As-doping is not straightforward. Because MBE growth of HgCdTe occurs under Te-rich conditions, as-grown, undoped HgCdTe layers usually contain a high density of Hg vacancies (metallic-site vacancies). Thus, As is predicted to be predominately incorporated on Hg sites and serve as active donors or neutral structures in the case of As complexes. Therefore, research on As-doping activation for HgCdTe MBE growth is highly important. The activation of As as an acceptor in HgCdTe was studied in this section, the mechanism of activation also was discussed by experiments as well.

(1) As-doping in HgCdTe

It has been recognized that the difficulties in the As-doping in MBE-grown HgCdTe mainly come from the amphoteric behavior of As due to the growth mode of Te-rich, and the low-temperature growth condition as required by the low SSC of As, which further complicates the understanding about the primarily possible modes of As incorporation.

(2) Sticking coefficient of As atom in HgCdTe

The relationship of As sticking coefficient and growth temperature was experimented on the epitaxy of As in poly CdTe, in crystalline CdTe and in HgCdTe between RT to 200 °C. To understand the sticking behavior of As atoms and its relation with the presence of Hg, the SSC of As during HgCdTe growth was measured as shown in Fig. 3.79. Compared with CdTe, the As sticking in HgCdTe shows surprisingly different behavior, decreases much more rapidly with increasing temperature, more than an order of magnitude within a temperature increment of only 10 °C. The temperature sensitive behavior of As sticking suggests that the ability in the growth temperature control is critical in terms of doping stability as well. A deviation in growth temperature of 1 °C would cause a variation of more than 10 % in doping concentration when growing at 160–170 °C range. As shown in Fig. 3.79, the fact that the SSC of As in HgCdTe is much lower than that in CdTe at the temperatures higher than 160 °C implies that it may has some correlation with the sticking behavior of Hg.

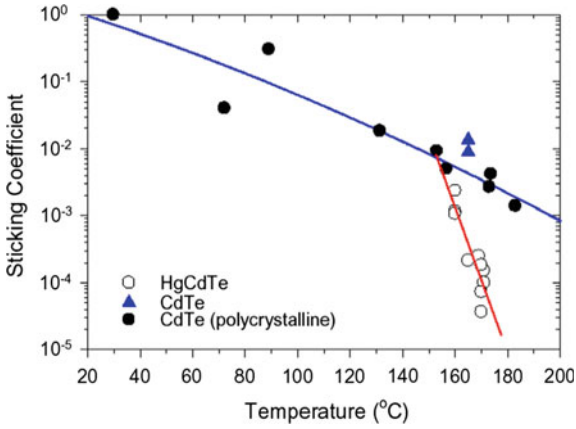


Fig. 3.79 The SSC of As in CdTe and HgCdTe as functions of the growth temperature

To clarify the effect of Hg on As sticking coefficient during growth, SIMS results of As-doping were examined on both poly HgCdTe and CdTe under Hg flux, which is compared in Fig. 3.80. As-doping CdTe was deposited on CdZnTe substrate and then HgCdTe was deposited on CdTe in sequence. It is difficult to adjust Hg flux rapidly during deposition. Since the Hg sticking coefficient is sensitive to growth temperature, the growth temperature is changed intentionally to adjust the Hg flux exposed on the wafer. During the deposition process, As cell temperature is 320 °C while the CdTe and Hg flux are kept stabile. From Fig. 3.80, SIMS result indicated that As concentration is higher in poly CdTe. After opening the Hg shutter, the concentration of As drops rapidly. During continuous poly HgCdTe epitaxy, the apparent decline of Hg concentration occurred as growth temperature increased. It is noticeable that the concentration of As increased with the rise of depositing temperature, according to the reduction trend of Hg concentration at the same time. Obviously, it is an exact evidence to illuminate that the Hg plays a major role in As sticking coefficient on the surface. The Hg presents much more great impact on the reduction of sticking coefficient, than the effect from growth temperature. Polymaterials cannot maintain a fixed composition, which is different from single-crystal materials. Different Hg concentrations will alter the ratio of HgTe and CdTe in polydeposit HgCdTe with a fixed Te flux. With the increase of growth temperature, the portion of HgTe decreases so that the material is more CdTe like and As concentration increases. The experiment clearly demonstrated the effect of Hg to As sticking coefficient. As dopant activities are closely related to Hg and Cd or the bond of HgTe and CdTe. AHgCdTe/CdTe superlattice was used to solve the As-doping problem in HgCdTe by atomic layer doping technology in CdTe (He, Wijewarnasuriya etc.). The experiment explained why the method is easy to obtain As-doped HgCdTe well.

However, it is very difficult to obtain good crystal quality materials if HgCdTe is grown at 170 °C. According to the experiment, the doping efficient of As enhanced

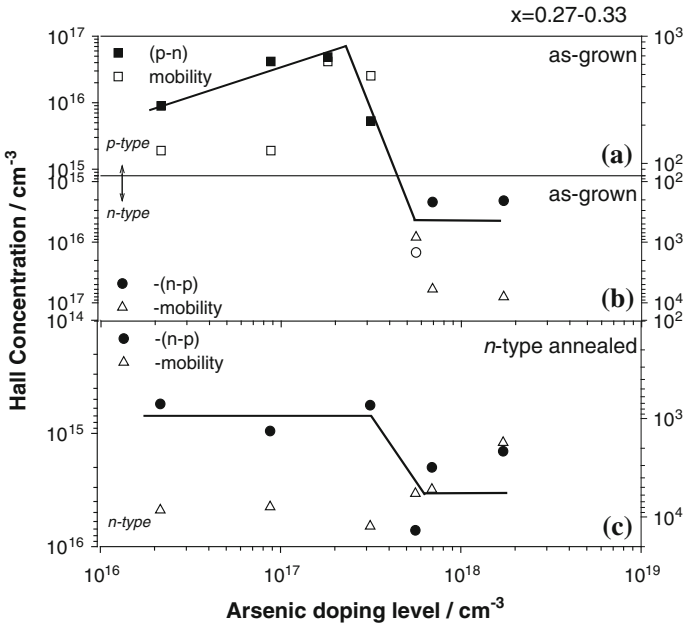


Fig. 3.80 The relationship of carrier concentration/mobility and doping level in extrinsic As-doping HgCdTe, Hall measurement results of 77 K. **a**, **b** as-grown HgCdTe, **c** after n-type annealed

a lot by using As cracker cell. After increasing the growth temperature of HgCdTe from 170 to 187 °C and keeping the As bulk at the same temperature as effusion cell, the same level of As-doping concentration was achieved.

(3) As incorporation in HgCdTe

In order to understand the behavior of As incorporation in the epilayers, electrical properties were analyzed by Hall measurement on As extrinsic doped samples. In Fig. 3.80, Hall concentration and mobility were compared between the as-grown samples and the one after n-type annealing. The n-type annealing was performed at 240 °C for 48 h under a Hg-saturated pressure to annihilate the metallic vacancies created either during the growth or during other annealing processes.

The temperature-dependent Hall measurements for undoped as-grown samples grown at 170 °C showed a compensated n-type characteristic, which indicates the concentration of Vhg grown at 170 °C is lower than that grown at normal temperature (185–190 °C). In the normal growth case, the typical donor concentration in intrinsic-doped HgCdTe(x = 0.3) is about 1×10^{15} $cm^{-3,}$ with the mobility of 7×10^4 cm^2/V s.

Figure 3.80 shows the relationship of carrier concentration/mobility and doping level in as-grown doped sample and n-type annealed sample at 77 K. For as-grown samples, the electrical properties are examined as p-type when the As-doping level was less than 3×10^{17} cm^{-3}, while donor acted as major role with As-doping up to 5×10^{17} cm^{-3}. The Hall concentration is fully converted to n-type and no changes depend on the increasing of doping level after n-type annealing in Fig. 3.80c.

From Fig. 3.80a, p-type concentration increased in the range of $1–8 \times 10^{16}$ cm^{-3} by the As-doping level from 2 to 30×10^{16} cm^{-3}, which is higher than the usual result of intrinsic-doped HgCdTe. It could be concluded that doping As into HgCdTe generates additional Hg vacancies. With a doping level up to 5×10^{17} cm^{-3}, donors dominated the electrical property, which was probably generated by heavy doping with As atoms, whose effect exceeded the additional Hg vacancies by large amount of As atom joining. However, as shown in Fig. 3.80c, the donor mobility is lower than the result of intrinsic-doped HgCdTe after n-type annealing.

Berding [160] proposed the single-atom doping model in 1998. Grein [161] reported the possibility of As_2 and As_4 existence in theoretical calculations in 1999. Lee [162] presented the relationship between N-type electrical properties and As_5 atom cluster after researching on low-temperature N-type annealing at 200 °C.

The result of the experiments was consistent with the single-atom doping model in two aspects: 1. Great amount of isolated AsHg make as-grown samples N-type, but at low doping level the as-grown sample appears to be P type. After n-type annealing, the doping level should be directly proportional to AsHg, but the N-type carrier concentration has no obvious change after increasing the doping level. This illustrated that the possibility that a single-atom As existing in the sample is quite low. As As_2 is a natural structure, it cannot explain the above experimental result, which also indicates a low possibility.

However, the calculation model for the As_4 cluster in Sect. 3.2.3 can explain the above experimental result. It shows that the formation energy for the structure model in Fig. 3.22c is the lowest which is the most possible form of As_4. AsHg–Vhg is similar to AsTc which shows acceptor properties. So this model can explain the reason why as-grown sample appears to be P type when the doping level is low. This structure consists of three AsHg atoms. The calculation shows that when they gather to As, the position relaxation rate is low and most of them keep in Hg site. This indicates the structure is a strong donor impurity. So the single As_4 cluster model cannot explain why, when the As-doping level is above 5×10^{17} cm^{-3}, the sample appears to be weak N type and does not increase with the increasing doping level.

After research on As tetramer-Vhg based on the cluster structure (3.22C) and AsHg–Vhg model, the As_4 cluster which related to Vhg is a typical donor impurity after compensation. This structure explains the experiment result in Fig. 3.80. The cluster structure may act as a deep energy impurity which compensates the intrinsic N type carrier concentration and mobility and makes them lower than that in undoped sample.

(4) Ex situ annealing for As activation

In order to understand the behavior of As incorporation in the epilayers, the annealing experiments were carried out either under a Hg pressure in closed quartz ampoules in a dual-zone oven, or in the vacuum environment at different temperatures. Before annealing, the samples were etched with bromine methanol solution to remove CdTe cap layers. The experiments were performed with different combinations of three basic anneals: the n-type anneal, the activation anneal, and the vacuum-anneal. The n-type anneal here means annealing at 240 °C for 48 h under a Hg-saturated pressure to annihilate the metallic vacancies created either during the growth or during other annealing processes. All the annealing experiments in this study were terminated by the n-type anneal, to ensure Hg vacancies were eliminated completely and only As-activated as donors. Some samples were pre-annealed by this procedure prior to the other anneals. The activation anneal or p-type anneal refers the annealing in a temperature range of 285–440 °C for 0.5–30 h at different Hg pressures. This process also creates metallic vacancies. The vacuum-anneal is performed in vacuum without participation of the outcoming Hg, and is aimed to understand the role of Hg in As activation. It was performed at 360 °C for 30 h which is known to produce metallic vacancies. The electrical properties were measured by temperature-dependent Hall measurements in the van der Pauw configuration in a temperature range of 300–12 K at a magnetic field strength of 0.2 T. Before the measurements, the samples were etched to remove the possible n-type top layers.

Figure 3.81 shows the hole concentration as the function of the doping level of the epilayers annealed at different conditions for activation. The general trend in As activation is consistent with those reported [163, 164], the electrical activation approaches 100 % when the doping level is low, and saturates or drops as the doping level up to 1×10^{18} cm^{-3}. High-temperature annealing will eliminate the

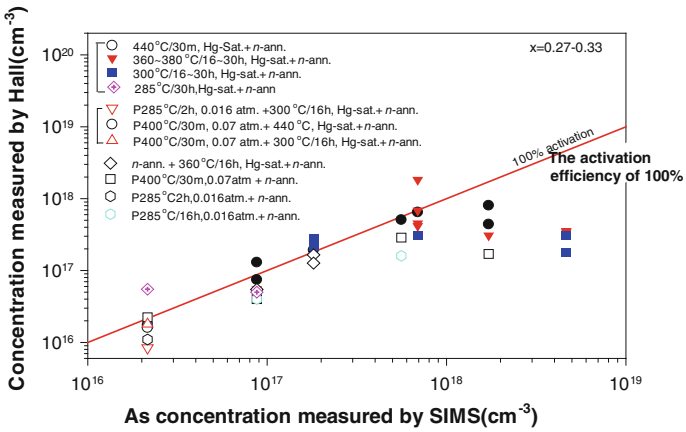


Fig. 3.81 The hole concentration as the function of the doping level of the epilayers annealed at different conditions for activation. The insert shows the annealing history for each sample

advantage of low-temperature growing for MBE, so lower the annealing temperature can keep the heterojunction structure grown at low temperature and restrain the diffusion of impurity in the material, which has an important significance.

To verify the effects of Hg on the activation, vacuum-anneals were performed on some of samples (doping levels between 1 and 2 × 10^{17} cm^{-3}). The Hall measurements on those samples (not shown in the figure) indicated that only 2–4 % of As atoms were converted to acceptors without the help of Hg, and confirmed the effect of Hg in As activation. In an attempt to understand the roles of metallic vacancies, two kinds of opposite pre-anneals were employed. One was *n*-type pre-anneal to annihilate the metallic vacancies created either during growth or by As incorporation as mentioned above, the other one was *p*-type pre-anneal at 285 and 400 °C at Hg partial pressures of 0.016 and 0.07 atm (Te-saturated limit), respectively, trying to introduce more metallic vacancies at the beginning. The pretreated samples were then subsequently annealed at 300, 360, and 440 °C under Hg-saturated pressure. As it shown in Fig. 3.81, the results did not produce any remarkable change in As activation as compared to those without the pretreatment. The result can be explained by the fact that metallic vacancies were also created during the activation anneals regardless of the initial concentration of the metallic vacancies. Another attempt was to vary the Hg partial pressure during the activation anneals. As shown in Fig. 3.81, a certain amount of As could be activated during low-temperature anneals at 285 °C even under a Hg partial pressure of 0.016 atm. For samples annealed at 400 °C at the Te-saturated limit, a noticeable reduction in activation was observed for highly doped samples (>1 × 10^{18} cm^{-3}) as compared with those annealed under the Hg-saturation.

(5) Diffusion characteristic of As in HgCdTe

Annealing processes cause redistribution of impurities, resulting in a substantial change in the doping concentration of the samples. In the previous reports, As was expected to lower the diffusion coefficient in HgCdTe, which was identified as the most desirable p-type dopant of hetero-homojunction fabrications. However, these studies were mostly based on the ion implantation samples. The studies based on in situ As-doped HgCdTe have not yet been reported. Therefore, the diffusion characteristic of As-activation process which were annealed under several temperature of 240, 380, and 440 °C was studied in extrinsic As-doped HgCdTe.

All the samples were annealed in a closed tube. Generally, As-activation anneal was performed under the annealing conditions of 440 °C/30 min (high-temperature annealing) + 240 °C/48 h (low-temperature annealing) or 380 °C/16 h (high-temperature annealing) + 240 °C/48 h (low-temperature annealing). Thus, the annealing runs under 240, 380 and 440 °C were performed in the As diffusion experiments. The annealing time was also chosen based on the conditions of As-activation annealing of HgCdTe. A 240 °C anneal was performed for 24 and 60 h, while a 380 °C anneal was performed for 16 h, and a 440 anneal for 30 min. Meanwhile, in order to study the effect of the mercury pressure on the As diffusion

process, the experiments were also performed separately with mercury pressure and without mercury pressure.

Under a saturating mercury pressure, the diffusion coefficients of As in HgCdTe at the annealing temperature of 240, 380 and 440 °C were $(1.0 \pm 0.9) \times 10^{-16}\,\mathrm{cm}^2/\mathrm{s}$, $(8 \pm 3) \times 10^{-15}\mathrm{cm}^2/\mathrm{s}$ and $(1.5 \pm 0.9) \times 10^{-13}\,\mathrm{cm}^2/\mathrm{s}$. The diffusion analysis of these three temperatures after annealing is shown in Figs. 3.82, 3.83 and 3.84. Since the diffusion coefficient of As at a 240 °C anneal is small, the mercury pressure takes little effect on the As diffusion. So the diffusion coefficient of As which was obtained by the simulation at that temperature without mercury pressure was similar to that with saturated mercury pressure. The annealing condition of 380 °C/16 h caused some effects on the distribution of As in HgCdTe. So the original steep doping interface becomes smooth. The annealing under these conditions will make a pn abrupt junction into a graded junction. Since the diffusion coefficient of As at a 380 °C

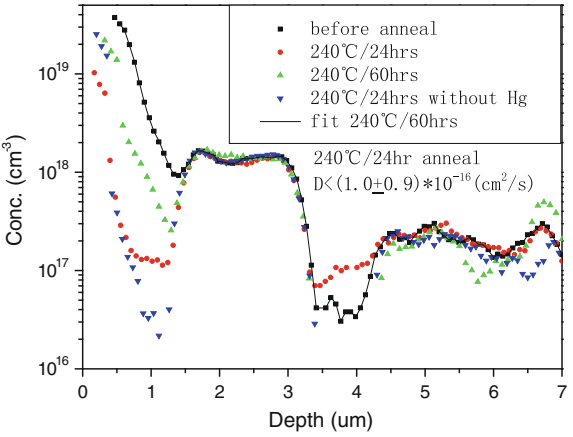


Fig. 3.82 Diffusion characteristic of As in HgCdTe after 240 °C annealing

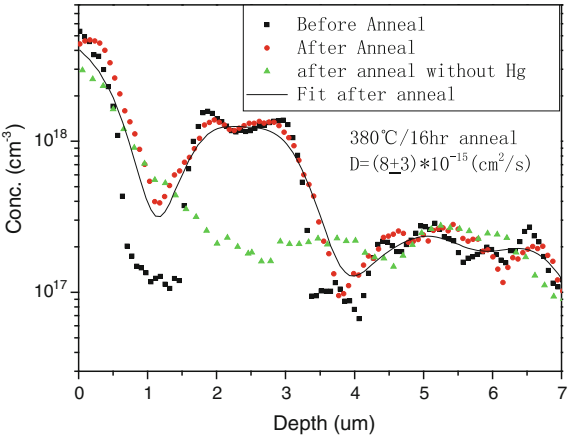


Fig. 3.83 Diffusion characteristic of As in HgCdTe after 380 °C annealing

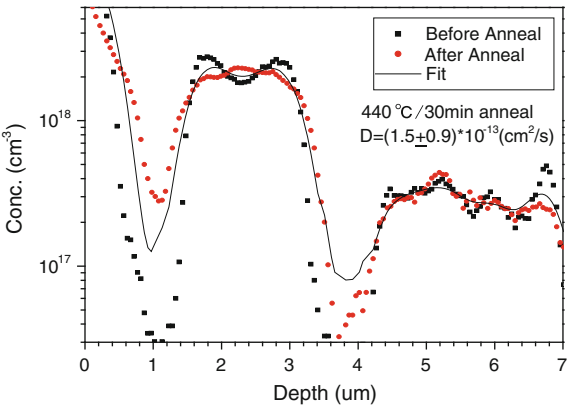


Fig. 3.84 Diffusion characteristic of As in HgCdTe after 440 °C annealing

anneal is large, the mercury pressure has a significant effect on the As diffusion. After simulation, it was found that the diffusion coefficient of As without mercury pressure was approximately $10^{-11}\mathrm{cm}^2/\mathrm{s}$. This was because a large amount of Hg vacancies was produced due to lack of Hg in the samples when the annealing was performed without mercury pressure. And in the high-temperature annealing, the impurities As can easily enter into the Hg vacancies, so a large amount of Hg vacancies in the samples makes the diffusion of As greatly accelerated. This phenomenon was also discussed in the literature [165]. It is shown in Fig. 3.84 that the As-doped structure in the sample was much different after a annealing under this condition for 16 h. It changed the distribution of As in HgCdTe. The diffusion coefficient of As at a 440 °C anneal is large, but the anneal took less effect on the distribution of As because of the short annealing time.

Figure 3.85 shows the relationship between the diffusion coefficient of As with the annealing temperature. The diffusion activation energy can be obtained from the slopes of the two lines in the figure: $E = -1000\mathrm{a} \times K_\mathrm{B}$, in which is the slope obtained from the simulation in Fig. 3.85 and K_B is the Boltzmann's constant.

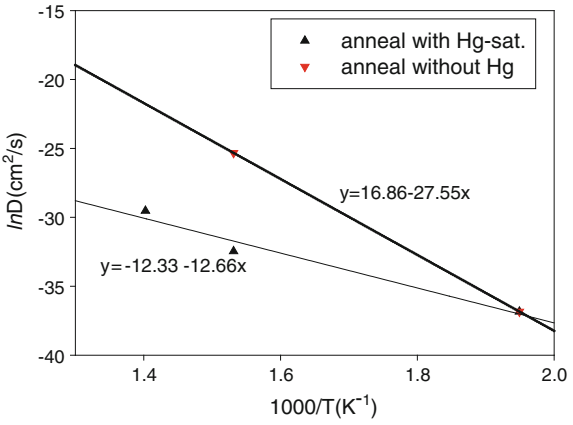


Fig. 3.85 The relationship between the diffusion coefficient of As with the annealing temperature

Table 3.11 The diffusion rate D_0 and diffusion activation energy E

	D_0 (cm^2/s)	E(eV)
Saturated mercury pressure	4.42e−6	1.091
Without mercury pressure	2.10e−7	2.375

Table 3.11 shows that the diffusion activation energy of As in HgCdTe annealed with saturated mercury pressure is not the same as that annealed without mercury pressure. Under saturated mercury pressure, the activation energy is small but the diffusion rate is also small; under vacuum pressure, the activation energy is larger while the diffusion rate is also large. This shows that As diffusion is carried out with different mechanisms under those two conditions.

Annealing under vacuum causes the sample to produce a large number of Hg vacancies, while annealing under saturated mercury pressure actually is an activation anneal. The results show, annealed under vacuum pressure, As atoms mainly exchange location through the Hg vacancies to complete the diffusion process. Because of the large number of Hg vacancies, the activation energy is large; however, the diffusion coefficient is also high.

2. Hg-rich high-temperature As activation with an open-tube technique

In the previous experimental results, the As-activation process is based on annealing technology in a closed tube. The HgCdTe samples and mercury source are enclosed in the ends of a vacuum quartz tube. The heat treatment temperature is controlled by adjusting the HgCdTe material temperature, and Hg pressure in quartz tube is controlled by adjusting Hg source temperature. The advantage of this technique is that the equipment is simple, the mercury pressure and the sample temperature are easy to be controlled. However, the disadvantage is that the cleaning and sealing processes of the quartz tube are complicated. With sample size increase, the difficulty of using close-tube technique increases greatly. At the same time, the high cost of the quartz tube also makes that the process difficult to meet the requirements of practical technology. Therefore, a better way is to anneal in the open-tube, which is to achieve a stable mercury pressure by controlling mercury reflux in the vertical cavity. The specific process includes the following three aspects.

(1) Open-tube equipment for Hg-rich high-temperature As-activation technology

Figure 3.86 shows a device which is suitable for the batch annealing process. The apparatus is mainly composed of a sample holder driving unit, the pressure control unit and the temperature control unit. In order to ensure the samples are not in the low-temperature region when being transmitting and being heated and cooled, the transfer device can be accurately positioned. The pressure control unit is to maintain gas pressure during the annealing, and high-pressure gas can make mercury reflux and suppress the Hg within a certain range. The annealing is

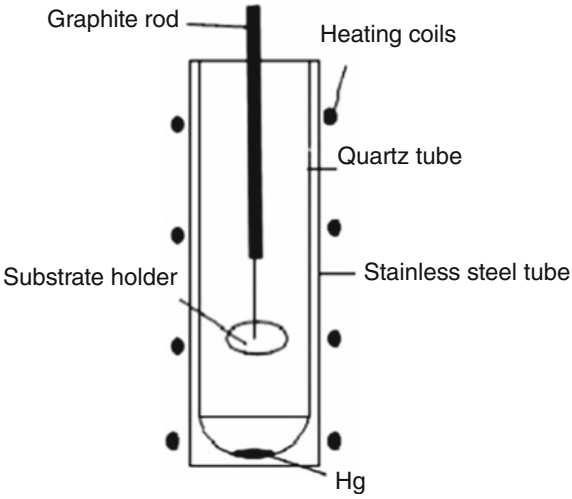


Fig. 3.86 Open-tube equipment for Hg-rich high-temperature As-activation technology

performed under a hydrogen atmosphere. The device can simultaneously hold three 4-inch samples which are annealed under high-pressure mercury atmosphere. In order to control the temperature of the sample and mercury source, the furnace includes four heating segments. One segment is used to control mercury source temperature so the samples can be annealed at different temperatures and under different mercury pressure conditions. Meanwhile, in order to make the temperature of the samples and the mercury source controllable, the furnace is supported by eurotherm with precise PID regulator which accuracy is up to ±0.1 °C.

2) Controlling the surface morphology of the samples annealed in opened tube equipment

In annealing process, the surface of HgCdTe is easily contaminated by mercury if the mercury droplets contact the sample surface. Therefore, it is necessary to adjust the profile of the temperature of the device and to adjust the position of the sample holder during the annealing process.

The surface morphology and the electrical property of the samples after annealing are obtained by subsequent tests. The results showed that, after annealing, the surface did not change significantly (Fig. 3.87). The defects on the sample surface were introduced during HgCdTe growth or during the later storage. To prove this point, one 3-inch Si wafer after polishing and cleaning was annealed under the same conditions (300 °C/16 h + 240 °C/48 h). The result is shown in Fig. 3.88. Except for some small loops, which with microscopes are difficult to observe, the sample surface is almost not affected. No defects are observed on the surface, which indicates that in such annealing device and process, the surface of HgCdTe materials can be well protected.

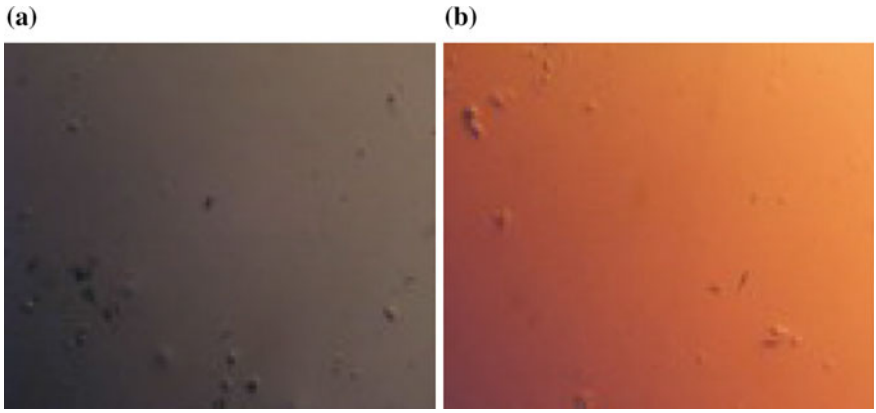


Fig. 3.87 The surface morphology of sample g145-2 before and after annealing, view area 110 × 100 μm^2. **a** before annealing, **b** after annealing

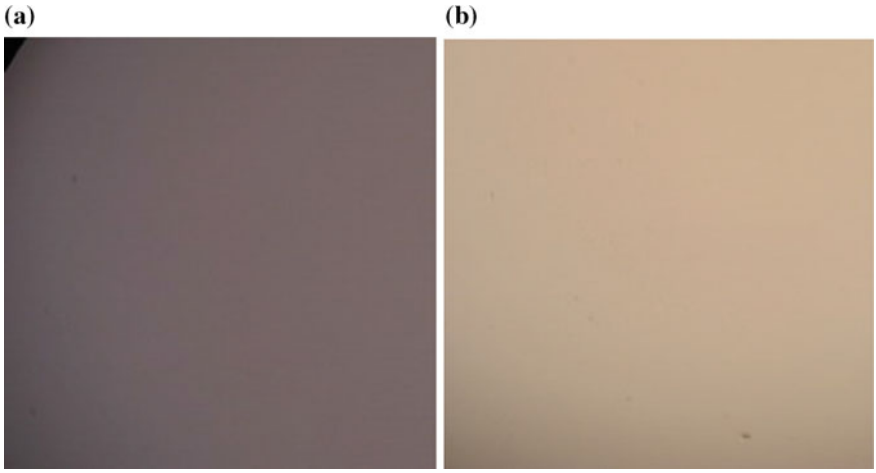


Fig. 3.88 The surface morphology of sample g145-2 before and after annealing, view area 2.56 × 1.76 mm^2. **a** before annealing; **b** after annealing

(3) Diffusion characteristics of As in HgCdTe after annealed in open-tube equipment

To study the diffusion characteristic of As in HgCdTe after annealing in open-tube equipment, HgCdTe grown by MBE on GaAs substrates were chosen for test. Meanwhile, some HgCdTe grown by LPE onCdZnTe substrates with thickness of 10–15 μm, were chosen for comparison. The growth temperature of MBE is about 170 °C. Without any annealing, HgCdTe grown by MBE are weak P-type or

N-type materials. HgCdTe grown by LPE at about 460 °C are strong P-type materials without any annealing.

The closed-tube annealing technology was also performed in order to prove the effect of the open-tube annealing technology for As activation.

In reference [166], it has been reported that the lowest temperature for As activation of HgCdTe grown by MBE was 280 °C. Meanwhile, the lower the annealing temperature, the less effect on the material structure. Therefore, the annealing condition we chose is 300 °C/16 h + 240 °C/48 h. The first step of 300 °C/16 h annealing is to activate As atoms in HgCdTe (to make As atoms move to Te position), while the second step of 240 °C/48 h anneal is to eliminate the Hg vacancy in HgCdTe. However, it was found that the As-activation annealing temperature of HgCdTe grown by LPE, which was 390 °C/10 h, was higher than that of grown by MBE. The second step of annealing was the same as that of MBE material.

Tables 3.12 and 3.13 show the electrical parameters of some samples which were measured at 77 K after annealing. In the experiments, we chose 15 As-doped HgCdTe grown by MBE with 4 labels, as well as 4 As-doped HgCdTe grown by LPE. These samples had good surface morphologies after annealing (300 °C/16 h + 240 °C/48 or 390 °C/10 h + 240 °C/48 h). Hall results showed that these samples were all P-type. In order to verify whether the As_{Te} acceptor or Hg vacancy acceptor, which was not eliminated, make the materials show p type, in each annealing process, we put in some undoped MW HgCdTe samples. Because the remaining impurities in HgCdTe are donor impurities, the undoped materials, after annealing to eliminate Hg vacancies should be n-type and the carrier concentration is generally $1–10 \times 10^{14}$ cm^{-3}. The Hall results showed that those accompany samples after annealing turned into N type, and the N-type carrier concentration and mobility were about -5.03×10^{14} cm^{-3}, -4.19×10^{4} cm^2/V s. This means that the Hg vacancies in HgCdTe had been eliminated after annealing. It was proved that a certain amount of As atoms in HgCdTe had been activated after annealing in the As-doped materials, so they showed p type. Wu [167] had studied the activation ratio in HgCdTe by using close-tube annealing technology. In order to verify the activation ratio after the open-tube annealing, some samples of the same series were chosen to be annealed in open-tube annealing equipment. The results were also shown in Table 3.12. Compared with the results of the two kinds of annealing method, the activation rate of Sample Gamt067 after open-tube annealing is twice more than that of the closed-tube annealing. Meanwhile, the activation rate of the other series of samples after open-tube annealing is similar to that of the closed-tube annealing. This result shows that the effect of open-tube annealing is better than that of closed-tube annealing. It is shown in that the samples of As-doped HgCdTe grown by LPE were also As-activated after open-tube annealing. However, it did not succeed after closed-tube annealing with the same annealing as shown in Table 3.13. This result also shows that the effect of opened tube annealing is better than that of closed-tube annealing.

Generally, for HgCdTe, the Hall concentration dose not always corresponds to the concentration of hole in the material, only if it dose not change with the

Table 3.12 Electrical parameters of HgCdTe grown by MBE by different annealing conditions

Samples	Annealing condition	77 K hall concentration (cm^{-3})	77 K carrier mobility (cm^2/V s)
Gamt253-14	Opened tube annealing 300 °C16 h + 240 °C 48 h	2.46E+17	2.00E+02
Gamt253-7	Closed-tube annealing 300 °C16 h + 240 °C 48 h	2.81E+17	1.63E+02
Gamt072-2	Opened tube annealing 300 °C16 h + 240 °C 48 h	1.84E+17	1.50E+02
Gamt072-3	Closed-tube annealing 300 °C16 h + 240 °C 48 h	-1.55E+16	-1.43E+03
Gamt067-8	Opened tube annealing 300 °C16 h + 240 °C 48 h	4.56E+17	2.90E+02
Gamt067-2	Closed-tube annealing 300 °C16 h + 240 °C 48 h	1.59E+17	2.86E+02
Gamt067-6	Opened tube annealing 300 °C16 h +240 °C 48 h	5.06E+17	2.94E+02
Gamt067-2	Closed-tube annealing 300 °C16 h + 240 °C 48 h	1.59E+17	2.86E+02

Table 3.13 Electrical parameters of HgCdTe grown by LPE after annealing

Samples	Annealing condition	77 K hall concentration (cm^{-3})	77 K carrier mobility (cm^2/V s)
lpel0256	Opened tube annealing 390 °C10 h + 240 °C 48 h	2.345E+16	458.5
lpel0257		4.37E+15	542.9
lpel0264		3.918E+16	144
lpel0267		−5.2949E+14	−4823.713
lpel0272		−2.479E+15	−79046.95
lpel0273		1.081E+16	564.9
lpel0274		5.103E+15	367.1
lpel0276		2.412E+16	72.33

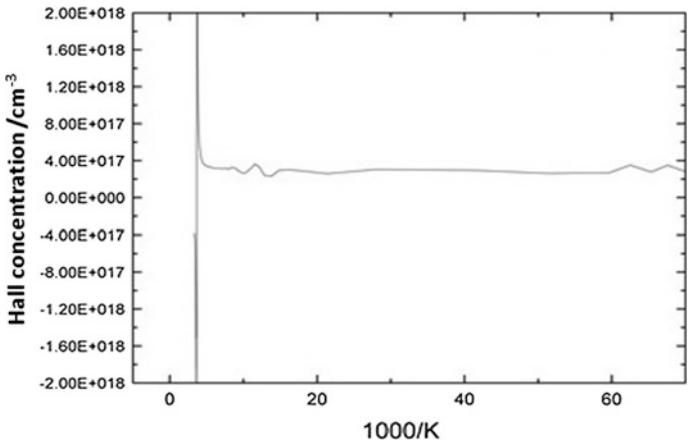


Fig. 3.89 The results of the variable temperature hall measurement of gamctub067-6

temperature. Figure 3.89 shows the experimental curve of the Hall concentration of sample gamctub067-6 with the temperature. The result shows no change for the Hall concentration below the temperature of 77 K.

The research shows that the open-tube annealing technology with saturated Hg vapor can be used to activate As atoms in HgCdTe. The technology is superior to closed-tube annealing. Meanwhile, it has the advantages of quick, simple and it also can hold multiple samples for annealing at the same time.

3.4 Si-Based HgCdTe LPE Technology

The MBE can be performed on substrates with a mismatch as high as 20 %. However, the larger the lattice mismatch between the substrate and the epitaxial layer, the worse the crystal quality of the epitaxial material is. The lattice mismatch between Si and HgCdTe is about 19 %. Although the quality of the epilayer could be improved by growing a buffer layer or superlattice buffer layer, the dislocation density of Si-based HgCdTe MBE material is now still higher than 1×10^6 cm^{-2}, which is 1–2 orders of magnitude higher compared to the HgCdTe material grown on CdZnTe substrate. Among the various HgCdTe epitaxy technologies, LPE is considered to be the best way to obtain high-quality epilayer because LPE can be performed at high temperature. This makes the performance of infrared detector prepared by LPE material still in the lead. Companies which are using the LPE technology to achieve mass production of IRFPAs include DRS, BAE, and Sofradir. Naturally, it is a research direction to grow materials with low dislocation density by LPE on Si/CdTe substrate prepared by MBE technology.

The results published by Mitsuo Yoshikaw and TW James have clearly indicated the ability of LPE to reduce the dislocation density [168, 169]. In the early 1990s, there have been a number of studies which focused on HgCdTe LPE with Si-based substrates. The technical difficulties related to the HgCdTe LPE on Si-based substrate include:

(1) The surface treatment after cutting the Si-based substrate

The surface of the Si/CdTe composite substrate is clean in atomic scale. The substrate must be cut and further processed if we adopt LPE as the subsequent growing technique. During the cutting, the surface contamination cannot be entirely avoided although the surface is usually protected by some coating layer, e.g. photoresist. The diced wafer will be cleaned chemically again after cutting. Since the thickness of the CdTe buffer layer on the composite substrate is only a few microns, it cannot be treated through the chemical-mechanical polishing (CMP) process but slightly etched in bromine solution. So, it is quite difficult to remove the contamination and impurities on the substrate surface.

(2) The protection of the Si/CdTe composite substrate during epitaxy

At high temperature, Si/CdTe composite substrates can easily dissolve in HgCdTe growth solution and the solution may be contaminated. This will seriously affect the surface quality of the epitaxial material. So, it is necessary to protect the surface of the Si/CdTe composite substrate and reduce the chance for direct contact between Si and the HgCdTe solution.

(3) The melt-back control of CdTe bufferlayer

A melt-back technique is usually adopted during the LPE on CdZnTe substrate in order to improve the quality of the substrate surface. However, for a CdTe/Si composite substrate, it is difficult to adopt meltback to remove the surface contamination and the damaged layer usually caused by polishing because the CdTe buffer layer is only several microns thick.

The Si/CdTe composite substrate is usually prepared by MBE on (111) Si wafers. The quality of the LPE layer grown on such composite substrates is not satisfied due to the poor quality of the CdTe/Si in the early studies. Besides that, the (111) crystal orientation is not suitable for the HgCdTe MBE. Thereafter, MBE technology focused on the silicon substrate with (211) crystal orientation. The Si/CdTe composite substrate also achieved a major breakthrough by using low-temperature nucleation of a ZnTe buffer layer and high-temperature annealing technology, and the substrate quality has also significantly improved [170]. This section will introduce the HgCdTe LPE technology on (211) Si/CdTe composite substrates in detail and give the feasibility of further development and evaluation about this technology.

3.4.1 The Surface Treatment of CdTe/Si Composite Substrate

The surface treatment of the substrate is quite important for the LPE technique. Without the appropriate treatment, the defects on the substrate will extend to the epilayer above and eventually deteriorate the film.

The CdTe/Si composite substrate is a wafer of 3 inch or larger. It needs to be diced into squares before LPE. It is a key issue to protect the substrate surface during the cutting and cleaning process.

For LPE, the shape of the substrate is determined by the graphite boat adopted. If the substrate groove of the graphite boat is a rectangle of 2×3 cm^2, the 3″ Si/CdTe wafer will be diced into the same shape. Damage or contamination cannot be avoided during dicing because a series of operations, such as patching with wax, chemical cutting, are involved. Impurities in the substrate surface are the main cause of the defects in epitaxial layer. The impurities on the substrate surface must be removed completely before epitaxial growth. As mentioned above, it is difficult to remove the surface defects by using the melt-back technique because the CdTe buffer layer thickness is only about 10 microns. A CMP technique is also not suitable for such a thin layer. In this case, the photoresist can be used to protect the substrate surface. Photoresist is a commonly used material in the semiconductor industry. The coating, removing, and cleaning process with photoresist can be carried out easily.

As shown in Fig. 3.90, the original surface of CdTe/Si composite substrate is Fig. 3.90a, b is the substrate surface after cutting without the protection of photoresist. It can be seen the surface is slightly contaminated. Fig. 3.90c is the surface after cutting with the photoresist protection. The edge of the substrate in Fig. 3.90c shows some cutting damages, but no contamination is observed. In order to obtain clean surface, a SiO_2 protective layer may also be grown on the top of CdTe by magnetron sputtering technology. The cutting is then carried out with the photoresist protection. The result of this method is shown in Fig. 3.91.

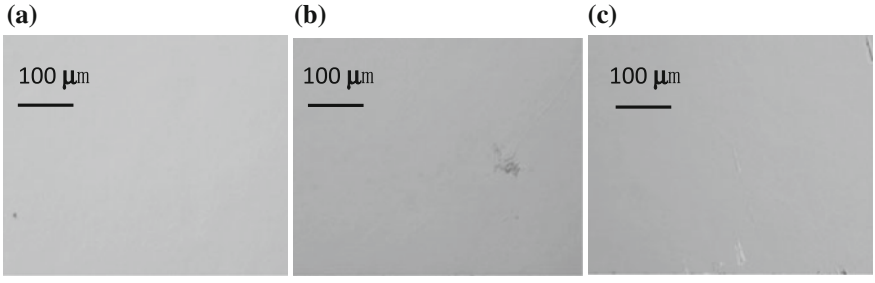


Fig. 3.90 The micrograph of the CdTe/Si composite substrate. **a** original surface, **b** the surface diced without photoresist protection, **c** the surface diced with photoresist protection

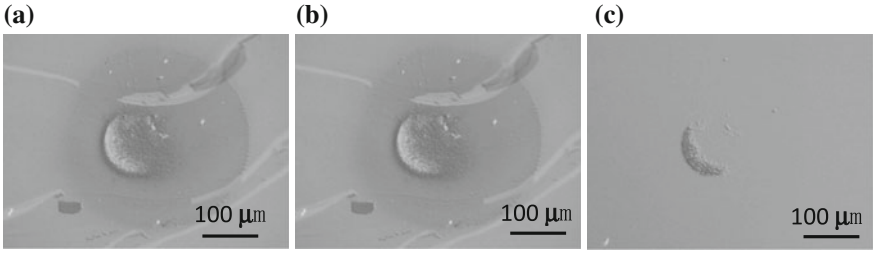


Fig. 3.91 The micrograph of the surface of the composite substrate after cutting with the protection of SiO_2 and photoresist. **a** After removing the photoresist, **b** After removing the photoresist and the SiO_2, **c** Removing the photoresist again after step **a**, **b**

In Fig. 3.91a, the substrate with SiO_2 and photoresist protections is cut and then the photoresist is removed by acetone. There are obvious residual photoresist on the surface. In Fig. 3.92b, the SiO_2 protective layer is then removed by hydrofluoric acid solution ($HF:NH_4F:H_2O$ = 3 ml:6 g:10 ml). The residual photoresist can also be seen on the surface. In order to remove the photoresist completely, the substrate is dipped into acetone for further cleaning. As shown in Fig. 3.91c, there is still a small quantity of photoresist on the surface after this additional step of acetone cleaning. The above results say more protective layers are not always effective because they may offer some new contaminations. Through the experiments above, an effective method to protect the surface can be derived. That is, covering the original surface with photoresist to prevent damage and containments caused during the cutting. Using this method, a clean and residue-free substrate can be obtained after cleaning with acetone.

Another problem in using Si/CdTe composite substrate is the Si on the edge of the substrate could dissolve into the melt, which will affect the purity of the melt. Coating methods may be used to grow a thin layer of SiO_2 at the edges and back of the substrate which could protect the silicon substrate from the erosion by melt. The substrate protected by SiO_2 is shown in Fig. 3.92. This is equivalent to opening a window on the surface of the substrate. The distance from the window to the edge of the substrate is 1.5–2 mm. The etchant used to open the window is also the hydrofluoric acid buffer solution($HF:NH_4F:H_2O$ = 3 ml:6 g:10 ml). The back and

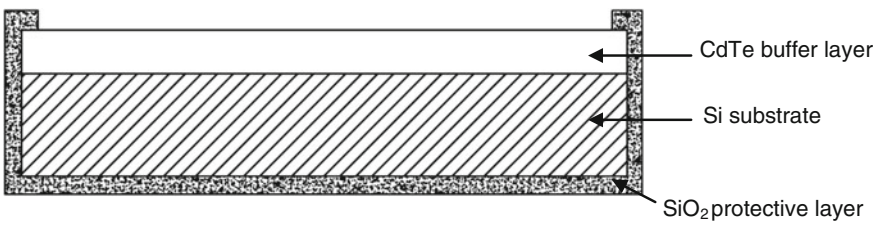


Fig. 3.92 The Si/CdTe composite substrate with SiO_2 protective layer

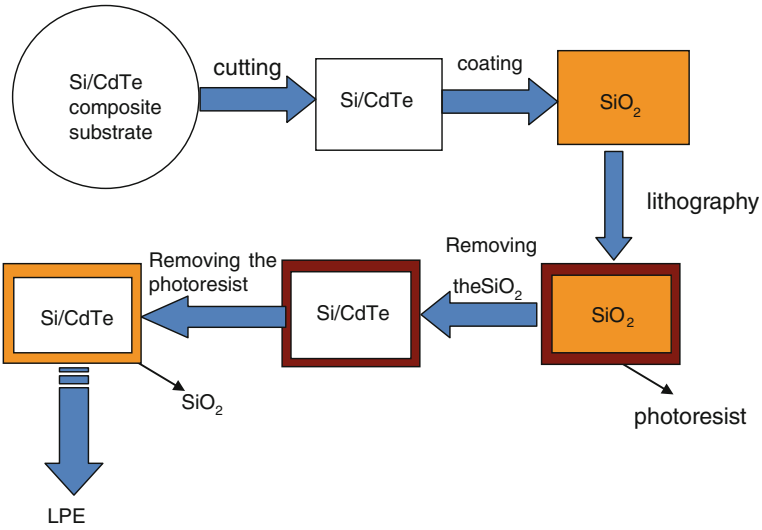


Fig. 3.93 The pretreatment process of the Si/CdTe composite substrate

edge are all covered with SiO_2 protective layers which thickness is about 0.3 μm. The protective layer can effectively prevent the silicon from the erosion by the melt.

Summing up the above steps, the entire pretreatment process of the silicon substrate is shown in Fig. 3.93. We first grow a CdTe buffer layer on (211) Si substrate of 3 inches by MBE technology. The thickness of the buffer layer is about 8 μm. Then the Si substrate is coated by photoresist and diced into 20 × 30 mm^2 squares. After the photoresist is removed clearly, a SiO_2 film is grown on the rectangular substrate with thickness of 0.2 μm. Then using the photolithography process, a window is opened on substrate while the edges and backside of the substrate are coated with photoresist. The substrate is then dipped in hydrofluoric acid buffer solution to remove the SiO_2 on the window. After that, the photoresist on the back and edge of the substrate are removed by acetone. We eventually get the Si/CdTe composite substrate with a window on the center and the protective layer on the back and edge.

The dissolving effect of the Si substrate could be greatly depressed by the above technique as illustrated in Fig. 3.94. The micrograph of the edge of the LPE film without SiO_2 protection is shown in Fig. 3.94a. The magnification is 50 times. It can be seen that the edge has been seriously dissolved. As a comparison, in Fig. 3.94b which is protected by SiO_2, the dividing line between protective layer and the epilayer is quite clear and there are no dissolving traces observed.

The composite substrate should be cleaned thoroughly before epitaxy. The cleaning process is similar to that of CdZnTe substrates. The process is divided into two steps: organic cleaning to remove the oil and the etching by bromine methanol to remove the thin oxide layer on the surface.

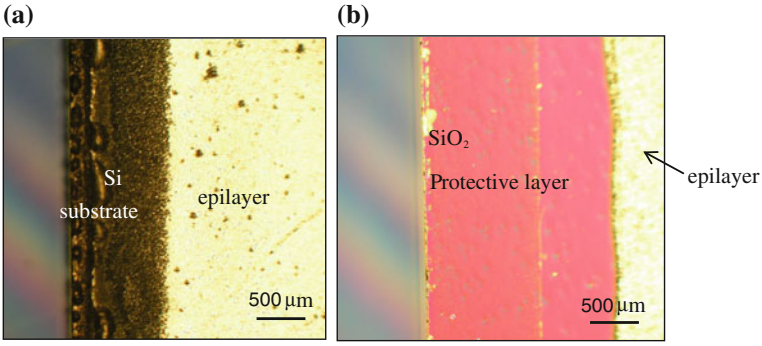


Fig. 3.94 Protective effect of SiO_2 to the edge of substrate. **a** The surface morphology of LPE film without SiO_2 protective layer; **b** The surface morphology of LPE film with SiO_2 protective layer

By using organic reagents (e.g. Trichlorethylene and methanol), the oil or wax can be removed efficiently. In accordance with the normal routine of ZnCdTe substrate treatment, the next step should be bromine polished in bromine solution. Because the CdTe buffer layer of the Si/CdTe composite substrate is as thin as several microns, the surface polishing process is not suitable for the composite substrate. In order to remove the oxide layer on the substrate surface, a slight etching in bromine methanol solution with 0.2 % concentration is adopted. The etching time can be determined by observing the surface morphology.

From the results shown in Fig. 3.95, it can be seen the more rough the more etching time. The optimized etching time is about 10 s according to the experiments.

3.4.2 LPE Process Optimization

The LPE is a method of crystal film preparation. The growth solution is saturated near the growth temperature. Because the solubility of the solute will decrease with

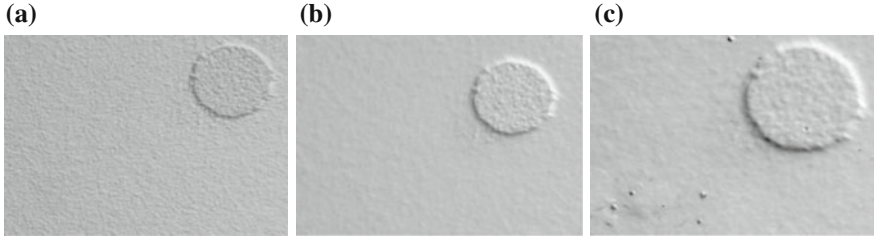


Fig. 3.95 The surface of the composite substrate with different etching time, the magnification is 1000X

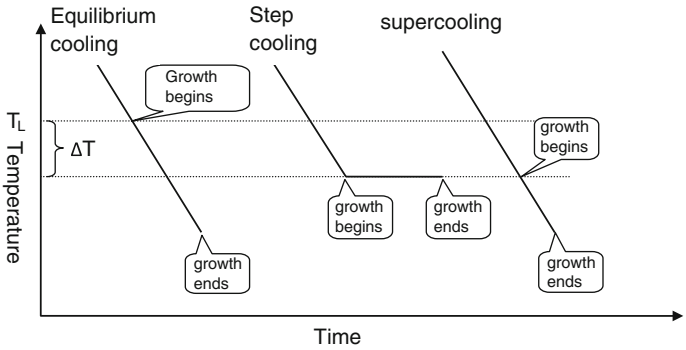


Fig. 3.96 Cooling processes of the melt for the three kinds of LPE techniques

decreasing the temperature, the solute will precipitate from the solution to form a film on the substrate when temperature decreases. There are usually three kinds of LPE techniques, equilibrium-cooling, step-cooling, and supercooling as show in Fig. 3.96.

In Fig. 3.96, T_L is the liquid temperature of the melt. For the equilibrium-cooling technique, the substrate and the solution are cooled with a uniform rate to a temperature which is quite close to T_L, then the solution and the substrate are brought into contact and the growth process begins. In the whole growth process, the temperature decreases with a uniform rate. At the end of the growth process, the substrate and the solution are separated. For the step-cooling technique, the substrate and the solution are cooled to a temperature below the T_L, (not low enough for spontaneous precipitation), then the temperature is kept constant until the growth process ends. For the supercooling technique, the substrate and the solution are cooled to a temperature $T_L - \Delta T$, and then the substrate and the solution are brought into contact. In the growth process, the temperature continuously decreases at a uniform rate until the process ends and the substrate and the solution are separated. In the frame of the diffusion-limited physical model and the self-infinity boundary condition, the relation between the film thickness and the growth parameters can be deduced. For the equilibrium process, $d = 2/3\ kRt^{3/2}$, where d is the films thickness, R is the cooling rate, t is the growth time, and k is a parameters related to the melt which can be regards as a constant. For the step-cooling process, $d = k\Delta Tt^{1/2}$, ΔT is the degree of the melt's supercooling. For the supercooling process, $d = k\Delta Tt^{1/2} + 2/3\,kRt^{3/2}$.

There are usually three kinds of LPE apparatus, which are tipping, dipping, and sliding apparatuses. The tipping method was used in the early period of LPE development. The advantage is that the device is simple and easy to operate. The disadvantage is that sometimes the solution and the film surface cannot be completely separated and the layer thickness and composition uniformity is difficult to control. The advantages of dipping method are that it can hold several substrates in a single run. At the end of growth, the substrate and melt can be separated

completely due to gravity and there is no melt left on the film surface. The disadvantage is that it cannot achieve multilayer growth and it is also difficult to control the layer thickness. For the sliding method, it can achieve single-layer, multilayer, single substrate, or multi-substrate growth. But its disadvantage is mainly the residue of the melt on the film surface. For the HgCdTe LPE, the sliding method is now widely used. The repeatability of the process is quite good for the sliding method because it uses anew melt for each run. Meanwhile, the sliding method can grow double-layer heterostructure.

Both Te and Hg can be used as solvents. Accordingly, $Hg_{1-x}Cd_xTe$ can be grown in Hg-rich or Te-rich melt. Since the solubility of Cd or Hg in Te solvent is quite high and the partial pressure of Hg in Te solvent is small, it is a commonly used in the LPE method to grow in a Te-rich melt. Several issues must be concerned about the Te-rich LPE growth. The composition uniformity of the epilayer is a key issue for the fabrication of a high performance device. The Hg partial pressure should keep constant in order to obtain films with good composition uniformity. An acceptable method of maintaining constant Hg pressure is the tube-sealed style. However the growth efficiency is low using this method. Wang [170], Harman [171], Chiang [172] adopted three different methods to control the Hg pressure in the open-tube LPE process. Wang etc. used high-pressure gas (Ar, H_2) to control the Hg loss in the LPE growth. Harman et al. used an additional Hg source to compensate for the Hg lost due to evaporation. The temperature of the Hg source depended on the growth-melt composition as well as the growth temperature. The advantage of this method is the good effect of controlling Hg pressure, but the amount of Hg in each growth is large. On the other hand, since the fluctuation of the furnace temperature will affect the Hg pressure it is difficult to maintain the balance of Hg loss in the whole LPE process. Chiang etc. used a ball-shaped tube (0.5 g Hg in it) with two arms fixed on the slider-boat to balance the Hg pressure. Solid HgTe can also be used to compensate for the Hg loss in the melt besides liquid Hg. In a horizontal slider-boat configuration, HgTe can be placed near the growth melt. This is a simple method, but the leakage of Hg from the gap between the boat and its cover cannot be avoided completely. In the Te-rich LPE process, the melt adhesion would happen if the melt and the substrate can't separate completely. The melt drop will cause dislocation or other defects when it cools down. The key point to reduce the melt adhesion is to control the thickness of the substrate. The residual melt can be greatly reduced if the gap between the substrate and the graphite slider-block can be controlled to about 20 μm.

In the Hg-rich LPE process with limited melt volume, the quality of the film is not satisfied because of the high Hg pressure ($\sim$10 atm) and the low solubility of Cd in Hg melt ($\leq 10^{-3}$ mol%). Tung etc. solved this problem by using a vertical LPE (VLPE) method with nearly unlimited melt volume and obtained films with good quality. In the Hg-rich LPE process, there is no melt adhesion problem and multilayer growth can be achieved because the meltback of CdTe substrate is very slow and is easy to control.

The step-cooling technique is suitable for the Si-based LPE because the CdTe buffer layer in the CdTe/Si composite substrate is thin. By adopting this technique,

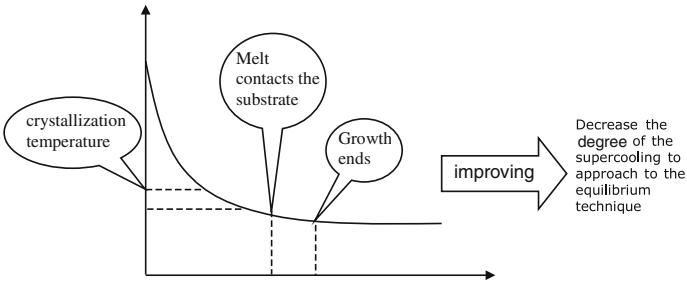


Fig. 3.97 Schematic diagram of LPE process on Si-base substrate

the dissolving of Si can be depressed as long as an enough degree of supercooling can be maintained. But there are also some advantages involved in this technique. At the beginning of the step-cooling growth, the spontaneous nucleation will deteriorate the films severely if there is a large degree of supercooling (e.g. 15 °C). So, we should decrease the degree of supercooling as much as possible to approach the equilibrium-cooling growth process. However, as light fluctuation of the temperature will destroy the CdTe buffer layer and cause epitaxial growth failure if a whole equilibrium-cooling technique is adopted. It is necessary to explore the favorable temperature to decrease the degree of supercooling so that the epitaxy can conduct in a nearly equilibrium-cooling manner. The process adopted is shown in Fig. 3.97. The melt temperature is slightly lower than the crystallization temperature before growth. Then, the substrate contacts the melt and the growth process begins. The temperature of the growth system decreases gradually and slowly in the whole growth process. Experiments show this growth method can efficiently suppress the spontaneous nucleation and improve the quality of the films.

3.4.3 Basic Properties of HgCdTe LPE Materials

Evaluations of the basic properties of HgCdTe LPE material include surface morphology, epilayer thickness, epilayer component, uniformity of thickness and component, *x*-ray diffraction FWHM, *x*-ray diffraction morphology, electrical parameters (carrier type, carrier concentration, carrier mobility), dislocation density and so on. Figure 3.98 shows an epilayer sample grown on (211) Si-based composite substrate by Te-rich horizontal sliding boat LPE technology. Usually there is some residual melt on the edge of LPE film material. The residual melt of the material shown in Fig. 3.98 has been cut off. The useful size is larger than 16×16 mm^2.

Figure 3.99 shows a microscopic picture of the Si-based HgCdTe material. The magnification is 50 times. It can be seen that epilayer surface is not flat and different from that of grown on normal (111) CdZnTe substrate. As shown in the figure, on (211) substrate, the surface show rugged pattern morphology. This morphology will

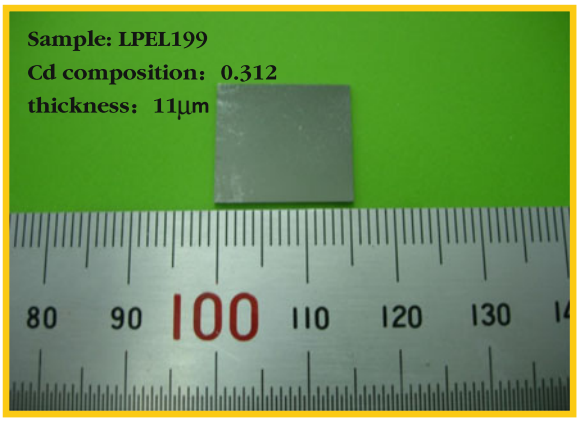


Fig. 3.98 Si-based LPE film

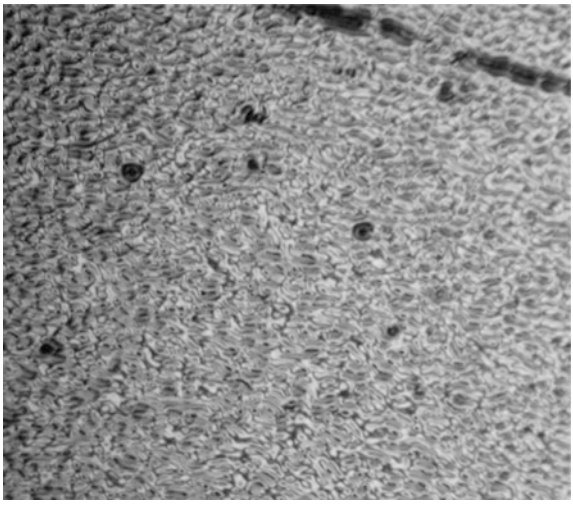


Fig. 3.99 Si-based LPE film micrograph with the 50× magnification

cause incident light to diffuse and hence it cannot present the effect of mirror reflection. Besides the rough effect caused by crystal orientation, another reason is contamination introduced in the cleaning process of the substrate. The contaminations will result in growth defects which make the surface rugged, as the black dots shown in Fig. 3.99.

Because the Si on the edge of the composite substrate can partly dissolve in the growth melt, the substrate edge must be protected by a particular technique. An effective method is using SiO_2 to protect the material edge. Figure 3.100 shows the micrograph of the substrate edge, the magnification is 100 times. In the figure, the left region is the SiO_2 protecting layer and the right region is the epilayer. The boundary of the two regions is quite clear. The edge of Si substrate is covered well and the dissolution of Si by growth melt is depressed.

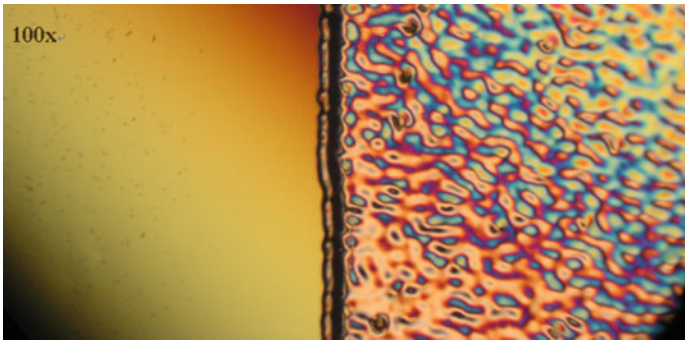


Fig. 3.100 Si-based LPE film micrograph

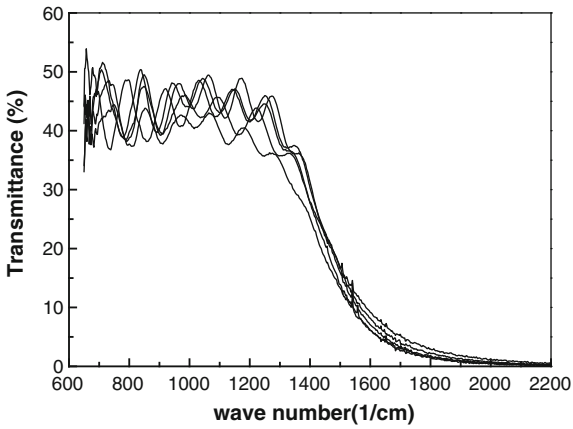


Fig. 3.101 Infrared transmitted spectrum of Si-based LPE film

The composition, thickness, and transverse uniformity of the epilayer are important. Figure 3.101 shows the infrared transmitted spectra of an HgCdTe LPE material on Si -based substrate. In Fig. 3.101, the five transmitted spectrum curves correspond to five points on the sample. There are clear absorption edge and interference peaks. The Cd component of the Epilayer can be calculated from the position of the absorption edge, and epilayer thickness can be calculated from the space of the interference peaks. In the calculations, the position on which the absorption coefficient equals to 500 cm^{-1} is regarded as the absorption edge, and the Cd component can be deduced according to Hansen's formula. It can be seen from the measurements that Si-based material's transmittance is 48 % which is lower than that of CdZnTe-based material. Usually, the CdZnTe-based material's transmittance can exceed 55 %. Calculated from the transmitted spectrum, the Cd component is 0.216 and the mean square deviation is 4.28×10^{-4}. The component of 0.216 corresponds to a cutoff wavelength of 11.38 μm at 77 K. Calculated from the interference fringe, the average thickness is 13.2 μm, and thickness mean square

deviation is 1.21 μm. Although the component uniformity is close to the CdZnTe-based material, the thickness uniformity is not good.

High-resolution *X*-ray diffraction is an important method to characterize crystal quality. In the situation of a given *X*-ray wavelength and fixed diffraction crystal plane, the double-crystal rocking curve FWHM reflects the film's quality. For CdZnTe-based LPE HgCdTe film, (333) plane's FWHM is usually less than 40 arcsec. Real crystals always contain some structural defects. Especially for heteroepitaxial material, large misfits between the substrate and the epilayer will cause lattice deformation and misfit dislocation in epilayer which is the one of main reasons of the broaden FWHM.

Lattice misfit between Si and CdTe is as high as 19 %. The thickness of CdTe buffer layer on Si-based substrate is only about 5–8 μm. So there are many misfit dislocations in Si-based substrate. As reported, for (211) Si/CdTe substrate with CdTethickness of 4–4.4 μm, the average FWHM is 83 arcsec. And the best results are 60 arcsec with CdTe of 7.4μm. The Rocking curve of LPE HgCdTe on above-mentioned substrate is shown in Fig. 3.102. The FWHM is 66 arcsec which is close to the result of Si/CdTe substrate.

X-ray diffraction morphology can be used to characterize film's uniformity. Figure 3.103 shows the reflective *X*-ray diffraction morphology of epilayers on Si-based composite substrate and CdZnTe substrate, respectively. For the epilayer on CdZnTe substrate, its morphology shows a"crosshatch" pattern. Because of small lattice mismatch between substrate and epilayer, mismatch stress which extends to epilayer from boundary forms a stress field effect to result in the crosshatch. For the epilayer on Si-based substrate, the large lattice mismatch between the substrate and the epilayer forms the "mosaic" pattern.

The dislocation of the material is an important factor to restrict the device performance. Dislocation will induce tunnel assistant leakage current which plays a leading role in the leakage current mechanisms for LW HgCdTe infrared focal plane device. It causes many difficulties for the development of LW devices.

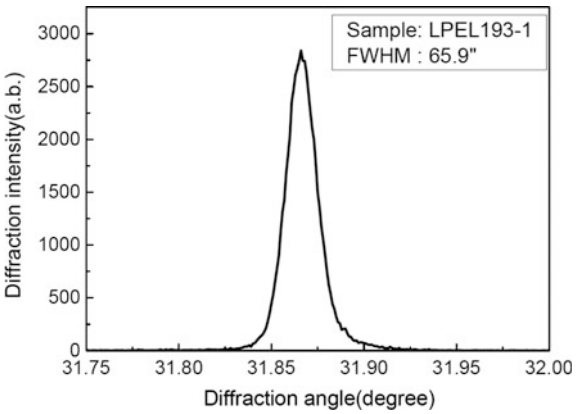


Fig. 3.102 High-resolution *x*-ray diffraction of Si-based LPE film

Fig. 3.103 *X*-ray diffraction morphology of Si-based LPE film

There are only a few reports on Si-based HgCdTe LPE materials. Smith et al. reported that the dislocation etching pits density(EPD) of Si-based HgCdTe LPE material is about 6×10^5 cm^{-2} [173]. But there are hardly any reports on surface defect morphology and the origin and suppression methods of LPE HgCdTe on (211) Si/CdTe. The results reported about LPE HgCdTe surface morphology are only that of (100) Si. These results indicate that it is possible for LPE growth of HgCdTe on Si, but it is hard to say Si-based LPE has the ability to grow material with low dislocation density. To prove that LPE has the ability to suppress dislocation, further experimental comparison should be applied for MBE and LPE HgCdTe growth on the same Si/CdTe composite substrate.

Schaake and Chen etchants are used to reveal defects on (111)B and (211)B surface, respectively. They can display the defects' type and shape quite well. For HgCdTe material on (211)B Si/CdTe composite substrate, Chen dislocation etchant is selected. The chemical ratio of Chen etchant is H_2O:HCl:HNO_3:$K_2Cr_2O_7$ = 80 ml:10 ml:20 ml:8 g, and the etching rate for epilayer is about 1.5 μm/min. The dislocation etching pits are observed and calculated by Olympus BX51 differential interference microscope.

Figure 3.104a is the surface dislocation etching pits of LPE HgCdTe material with thickness about 10 μm on (111)B CdZnTe substrate. Figure 3.104b shows the etching pits of LPE HgCdTe on (211)B CdZnTe, which etched by Chen etchant for 210 s. From the results, it can be seen that the triangle characteristic on (211) substrate is not obvious compared to that of on (211) substrate although most of them can be recognized. The EPD of film on (211) substrate is slightly higher than that of on (111) substrate, both materials' EPD are less than 1×10^5 cm^{-2}.

The observed etching pits in the figure were defined as C2 type etching pits [174, 175]. C2 etching pits are regular equilateral triangles. C2 etching pit density is different in different samples or different positions of the same sample, and its

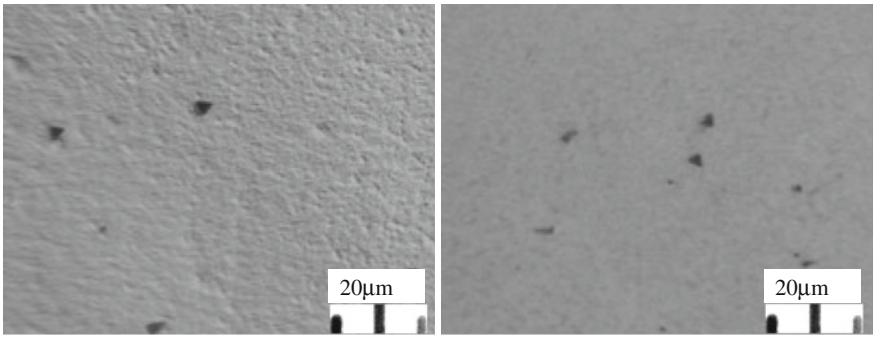


Fig. 3.104 Dislocation etching pits of HgCdTe films on CdZnTe substrate

distribution range is wide. C2 etching pit density will increase rapidly when etching to about 2–3 μm near a boundary. When the etching process continues etching pits will become fainter, and disappear at last. But, at the same time, new C2 etching pits will appear. C2 etching pits do not increase around the macro-defects. James observed the dislocation defects around an interface by transmission electron microscope [169]. It is found that because many dislocations were hindered during pass-through process to epilayer, the dislocations kinked and moved toward the ⟨110⟩ direction which parallels the boundary. Because dislocations were hindered during the pass-through process, etching pits showed strong space-located characteristic in the longitudinal direction. Local random component fluctuation may appear during epilaxy process which results in random distribution of the misfit dislocations. The origin of C2 etching pits are attributed to the misfit dislocations in material. Because the lattice of CdZnTe substrate is matched to the HgCdTe epilayer, the dislocation density is low, less than 10^5 cm^{-2}. There is another etching pit defined as C1 which has the pass-through property. C1 etching pits disappear when approaching about 2–3 μm from the interface. The reason is that it is the component-transition region in which Cd component is high and the Chen etchant did not work. C1 and C2 etching pits are different in size. The area of a C2 etching pit is larger than that of C1. But the shape of both C1 or C2 pits is close to an equilateral triangle.

Figure 3.105a shows the dislocation etching pits of HgCdTe LPE material grown on Si/CdTe composite substrate, Fig. 3.105b shows the dislocation etching pits of HgCdTe MBE material grown on Si/CdTe composite substrate. The HgCdTe film thickness is both 10 μm. The scale is 10 μm per grid in figure. For both LPE and MBE materials, etching pits grown on Si-based substrates are smaller than that on CdZnTe substrates. From the etching pit shape, it can be found that most of the etching pits observed are C1 type. After deeply etching the sample as shown in Fig. 3.105a (etching time 3 min, etching depth 4.5 μm), the etching pit deforms and the amount increases as shown in Fig. 3.106. Most of the dislocation etching pits observed are C2 type. With the increasing of etching depth and being close to CdTe/HgCdTe interface, the misfit dislocations increase obviously.

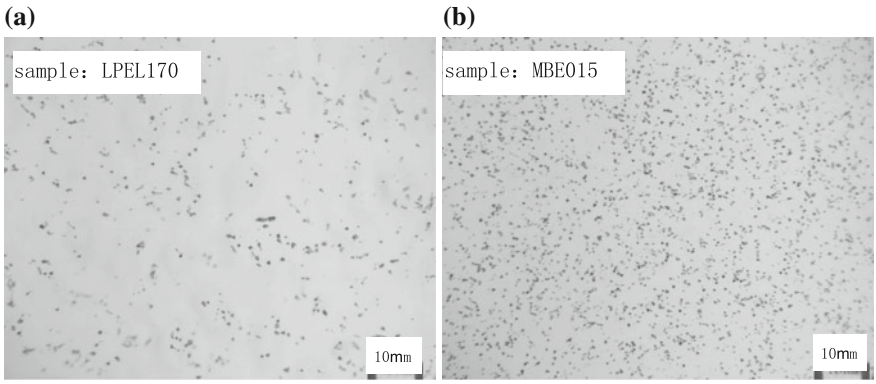


Fig. 3.105 Dislocation etching pits revealed by Chen etchant with etching time of 90 s. **a** Si - based LPE HgCdTe. **b** Si-based MBE HgCdTe

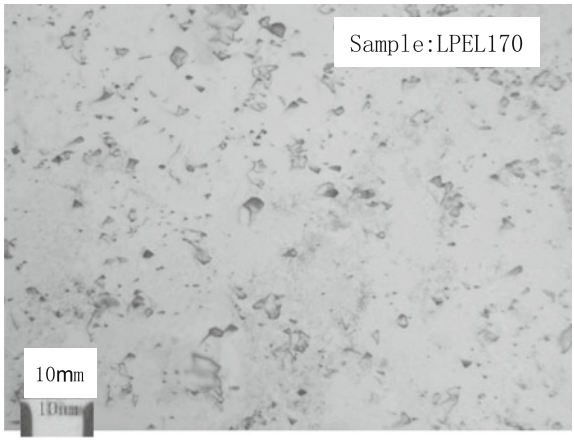


Fig. 3.106 The EPD increases after etching 3 min on HgCdTe/CdTe/Si LPE material

In addition, it can be seen from the figure that etching pits density of LPE material is obviously less than that of MBE material. The dislocation etching pits density of Si-based HgCdTe LPE material is 5–8 $\times$ 10^5 cm^{-2} while for MBE material it is 5–8 $\times$ 10^6 cm^{-2}. The surface dislocation density of LPE material is lower by about one order of magnitude. The ability for suppressing the dislocation prolongation for LPE is obvious. Compared to MBE, LPE on Si/CdTe has no advantage on surface morphology and surface defects. But the high-temperature growth characteristic for LPE can make misfit dislocation kink and close near the interface. This process is helpful to decrease the dislocation density and improve crystal quality. But, the dislocation density ($\sim 5 \times 10^4$ cm^{-2}) of HgCdTe LPE material on Si/CdTe composite substrate is still one order of magnitude higher than that of the material on CdZnTe substrate. The main reason is the 19 % lattice misfit between Si and HgCdTe epilayer.

Hall measurement is a typical method in semiconductor material evaluation. By measurement and analysis of magnetic resistance voltage $V_p(I, T)$ and Hall voltage $V_H(I, T)$, some basic physical parameters can be attained, such as carrier concentration n, mobility μ and so on. Further research for material's physical properties can be carried outby studying the variation of these physical parameters with external conditions such as temperature, magnetic field, and so on.

In 1958, Van der Pauw indicated that sample shape is not necessary to be very regular in real measurement. In fact, if sample thickness is uniform, no voids exist, for electrodes on the edge of the sample with size far less than the sample, Van der Pauw method can be used to measure the sample's resistivity and Hall coefficient.

For narrow band gap semiconductor material, such as $Hg_{1-x}Cd_xTe$, the mobility of the electrons and holes are very different because of the different effective masses. This results in the strong electron and hole mix conducting behavior for P-type material. The variation of Hall coefficient R_H and conductivity σ with the temperature is more complex than normal semiconductor.

The electric performance of as-grown LPE $Hg_{1-x}Cd_xTe$ material is hard to control. It often shows mixed conducting behavior. P type $Hg_{1-x}Cd_xTe$ material with stable hole concentrations can be obtained by P-type annealing technology.

Figure 3.107 shows the electric properties of P type HgCdTe material grown by LPE and MBE. Different dots indicate different growth methods and different substrates. Hole mobility of LPE material is higher than that of MBE material in the range of high concentration. It maybe resulting from the lower dislocation density of LPE material compared to MBE material. The effects of dislocation on mobility become more obvious as the carrier concentration increases. It also can be seen from the figure that LPE on Si/CdTed substrate also generates many defects related to lattice misfit, and the mobility of Si/CdTe LPE material is lower than that of CdZnTe LPE material.

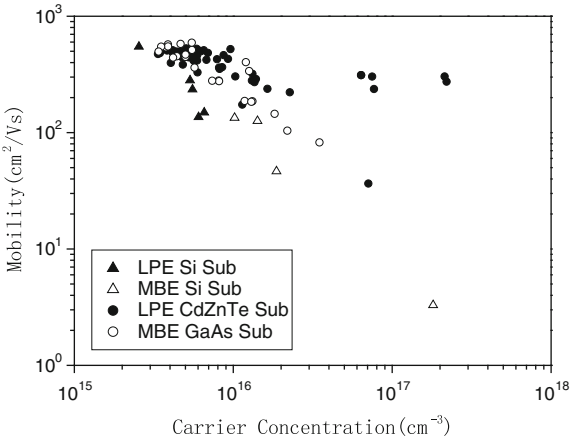


Fig. 3.107 Electric properties of P type HgCdTe material by LPE and MBE

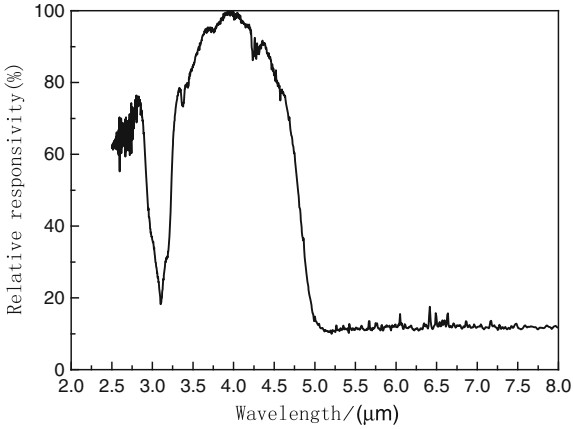


Fig. 3.108 The spectral response of device made of Si-based LPE material

A mid-wavelength 256 × 1 linear device is fabricated to verify Si-based HgCdTe LPE material. Its photosensitive area is 28 × 28 μm^2. Figure 3.108 shows the device spectral response at liquid nitrogen temperature. The cutoff wavelength is 4.8 μm.

Figure 3.109 shows the I–V and R–V curves of the devices at liquid nitrogen temperature. There is no cold shield applied when it is measured. The solid lines indicate the current while the circle lines indicate the dynamic impedance. The average zero bias dynamic impedance R_0 is 4.58 × 10^8 Ω. The product of zero bias impedance and the pixel area, R_0A, is 3500 Ω cm^2.

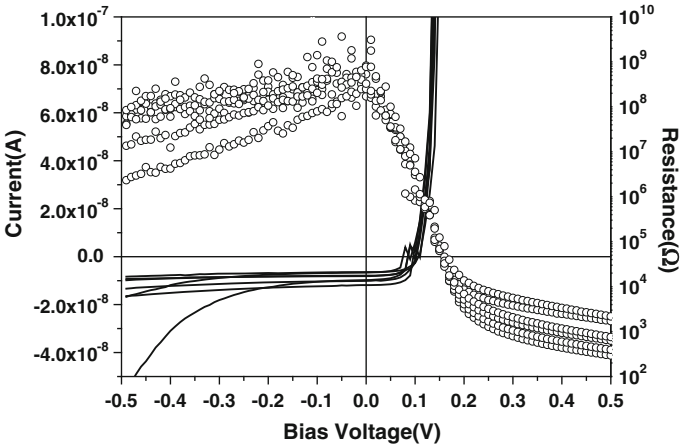


Fig. 3.109 I–V curves of device made of Si-based LPE material

3.4.4 Remaining Issues and Analysis

The most ideal technology for Si-based HgCdTe epilaxy is to use MBE or MOCVD to grow the buffer layer and the epilayer in the same growth chamber. But the dislocation density of Si-based HgCdTe grown by the vapor-phase epitaxial method cannot be decreased to the level of CdZnTe-based HgCdTe material. Dislocation is an important factor to restrict the improvements of the device performance. Tunnel-assisted current induced by the dislocation plays the leading role in leakage current mechanisms, which is a big challenge to developing HgCdTe infrared focal plane devices with LW sensitivity.

The LPE process at a relatively high temperature can make misfit dislocations kink and close at the interface. This mechanism is helpful to decrease the dislocation density and improve the crystal quality. SMITH et al. results of HgCdTe LPE material grown on (100)Si-based substrates(dislocation EPD is about 6×10^5 cm^{-2}) also confirmed that LPE can decrease material dislocation density [176]. But there is still an order of magnitude gap of the dislocations between Si-based HgCdTe LPE material and CdZnTe-based HgCdTe LPE material [177]. That means LPE can suppress the pass-through of dislocations, but the effect still does not reach the ideal requirement.

In the present results, the biggest problems which affect Si-based HgCdTe LPE material are bad surface morphology and high surface defect density [178]. For back-illuminated devices, the microscopic rippled feature induced by the (211) crystal plane will give rise to diffuse scattering which will increase the cross talk between the different pixels of the detector array. However, it has little relationship to the technical level and it can be eliminated by the postprocessing. The effect of surface defects on the device performance is very serious, and the emergence of such defects has something to do with the LPE technology level.

Figure 3.110 shows a typical surface morphology of Si-based HgCdTe LPE material. There are defects with the size of more than 10 μm and density of about

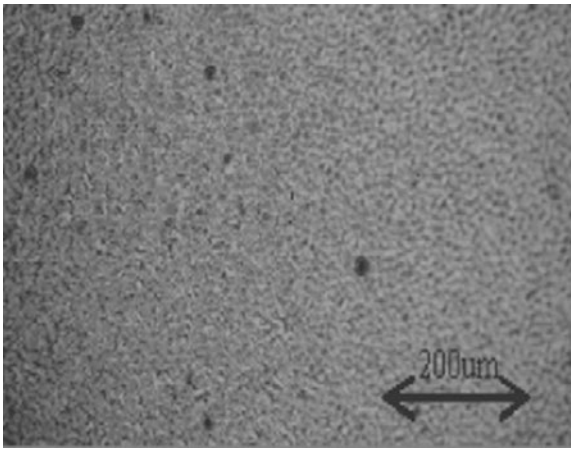


Fig. 3.110 Surface morphology of Si-based HgCdTe LPE material

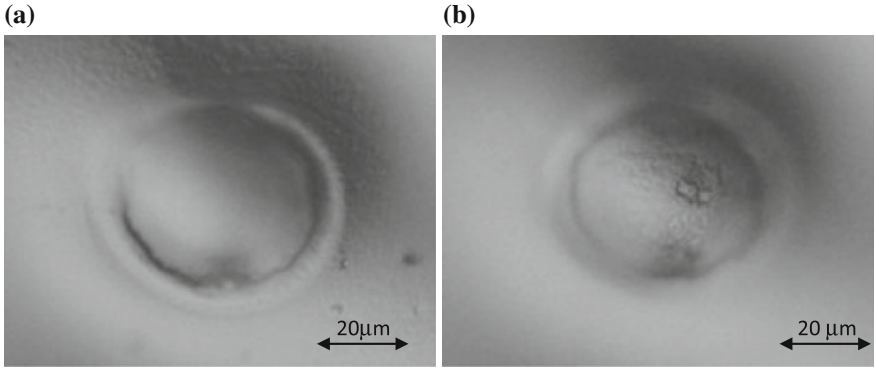


Fig. 3.111 Surface defect of Si-based LPE HgCdTe. **a** *Top* of the Si-based HgCdTe surface defect. **b** *Bottom* of the Si-based HgCdTe surface defect

500 cm^{-2}. If observed at high magnification, the defects are pits. Figure 3.111 shows the morphology of the defect under a 1000× optical microscope. Figures 3.111a, b are the morphologies focusing on the top and the bottom of the pit. The distance of the focused point is 15 μm which indicates that the depth of the pit is about 15 μm. The thickness of the HgCdTe epilayer is 10 μm and the thickness of the buffer layer is about 5 μm. So, the 15 μm depth means that the origin of the defects is related to the defects in the substrate and CdTe buffer layer. When checking the surface of MBE CdTe/Si composite substrate, no such large defects with the same density and size are observed. Therefore, these defects may come from the processing of the substrate or the processing enlarges the defects in CdTe buffer layer.

When observing the material surface under a 1000× microscope, it is still found there are round surface defects with the size of 2–3 μm on Si-based HgCdTe LPE material surface(see in Fig. 3.112) The defect density can be up to 10^5 cm^{-2}. It is also observed that nonuniformly distributed surface defects with the size less than 1 μm exist on the MBE Si/CdTe composite substrate surface (see in Fig. 3.113).

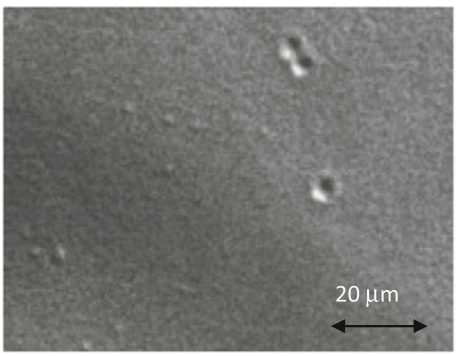


Fig. 3.112 Surface morphology of Si-based HgCdTe LPE material

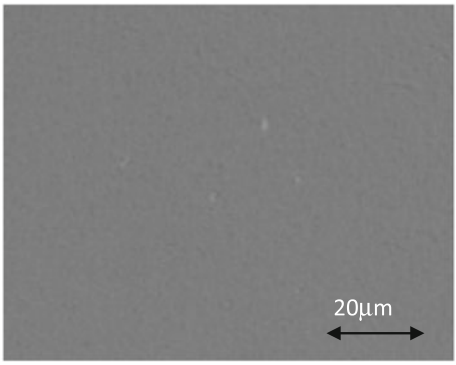


Fig. 3.113 Surface morphology of Si-based CdTe buffer layer

These defects are common on MBE material surface, so it can be deduced that small size defects on Si-based HgCdTe LPE material surface come from surface defects of CdTe layer in composite substrate.

The above results show Si-based

HgCdTe LPE indeed can suppress the dislocation. Therefore, this technology has a certain development prospect. But the quality of the Si-based HgCdTe LPE still has a large gap in performance compared to that of the traditional HgCdTe LPE material grown on CdZnTe. The gap manifests itself in three respects, first in the relatively high dislocation density, second in the bad surface microscopic flatness, and third in the high surface defects density. The development prospect of Si-based HgCdTe LPE technology also depends on the feasibility of solving these problems.

The high dislocation density results from two factors. One factor is that the dislocation density of CdTe buffer layer on Si-based composite substrate is relatively high. The other is that the misfit between CdTe and HgCdTe is larger compared to that of ZnCdTe substrate. There are two ways to solve these problems. One is that MBE will continue to improve the quality of the composite substrate. The Raytheon company had achieved good results on the surface quality of composite substrate and the lattice-match technology. They used CdTeSe instead of CdTe to meet the lattice-matching requirement of HgCdTe as well as improve the epilayer surface quality greatly. The other way is to use heat treatment at high temperature to further decrease the dislocation density in HgCdTe epitaxial material. The results obtained by Yu Meifang et al. confirmed the ability of heat treatment to decrease the dislocation density in HgCdTe epitaxial materials [179]. If this technology is applied to CdTe or CdTeSe buffer layer, it may achieve the purpose of decreasing the dislocation density.

Regarding microscopic flatness, although its fluctuation is too small to have serious effect on the technique of device fabrication, it has an obvious effect on the transmittance of material. So this is a problem to overcome. For LPE, it results from the (211) crystal orientation. Experiments indicate the fluctuation of the surface microscopic ripples is related to the epitaxial growth rate. The fluctuation can be decreased by decreasing the epitaxial growth rate. The optimization of the parameters, such as growth temperature, growth time, and growth cooling rates may

further improve material surface quality. Surface defects mainly come from the substrate defects and the processing technology of the substrate. According to the level of HgCdTe MBE technology at present, the surface defects density of the epitaxial material can be controlled at 1×10^3 cm^{-3}, and the best can be less than 300 cm^{-2}. Additionally, the defects are small in size, which has little effect on the infrared focal plane devices. So it is not a big problem now. However, there are still many fields to improve in the processing technology of substrates. For example, decreasing the time between removing the Si-based composite substrate from the cavity and protecting the fresh surface, increasing the thickness of CdTe buffer layer or growing a thicker polycrystalline CdTe layer on the surface and then removing it by Br-methanol polishing technology could help improve the process. Another method is to abandon the substrate processing, namely use the whole wafer for HgCdTe LPE. A SiO_2 protective layer can be prepared before the CdTe buffer layer growth. So, the Si/CdTe substrate prepared by MBE can be directly used for LPE growth without further processing which could completely eliminate the effect of substrate processing on the material surface quality.

3.5 Thermal Stress of Si-Based HgCdTe Materials

The epitaxial growth of HgCdTe on Si substrates solves the thermal mismatch issue between the substrate and the readout circuit. However, the thermal mismatch between the silicon substrate and the HgCdTe epilayer still exists, which is more serious compared with ZnCdTe or GaAs substrate [180–183].

The origins of the stress mainly include: stress due to lattice mismatch of two kinds of materials; thermal stress caused by two kinds of materials with different coefficients of thermal expansion after the temperature change. The existence of stress not only has dramatic impacts on the growth mode (two-dimensional layer growth, three-dimensional island growth, etc.), but also causes misfit dislocations at the interface. Threading dislocations and other crystal defects extend from the substrate to the epitaxial film and will cause new crystal defects in the film. Excessive crystal defects degrade the performance of the semiconductor film significantly. For example, the dislocations that exist in HgCdTe materials can act as recombination and scattering centers which affect the carrier lifetime and mobility directly and eventually degrade the performance of the device [184]. Therefore, in order to reduce the defects in HgCdTe materials, especially to reduce the dislocation density, the studies on dislocations caused by stress (distribution and its evolution) are very important. Only with a good grasp of stress-related information in silicon-based CdTe system can one well understand the mechanism of proliferation and expansion of dislocations in this particular system, and thus inhibit dislocations and reduce the dislocation density, improve the performance of silicon-based CdTe materials, and achieve the purpose of protecting the materials from fracture due to the increase of stress in low-temperature environment eventually.

Usually, heterogeneous interface states can be divided into three types: pure strain, pure released, and partially released [185, 186]. Pure strain refers to the lattice structure of the epitaxial layer being parallel to the interface after deformation and the lattice structure of the substrate is also parallel to the interface and they are exactly the same, which is the so-called coherent growth. When the lattice constant of the epitaxial layer is larger than the lattice constant of the substrate, the lattice parallel to the interface of epitaxial layer is compressed without considering the deformation of the substrate (substrate thickness is much greater than the thickness of the epitaxial layer). According to elastic theory [187], the epitaxial layer will expand along the direction perpendicular to the interface to compensate for the lattice shrinkage of the parallel interface. Pure strain can only exist in a relatively thin epitaxial layer, which is smaller than the critical thickness. With the increase of epitaxial layer thickness, the strain energy increases gradually, which leads to the generation of misfit dislocations. At this point, the completely strained state of the interface is broken. The epitaxial layer will transit to the partially released state, and eventually transit to pure released state. In addition, when molecular beam epitaxial growth is performed on heterogeneous substrate, the material of epitaxial growth is at high temperature (185 °C). At the end of the growth, the material temperature decreases to room temperature. Infrared detectors usually work at cryogenic temperature, therefore the temperature will further decrease to −193 °C. The temperature span is as large as about 378 °C. The lattice mismatch between Si (expansion coefficient 2.6×10^{-6} K^{-1}) and HgCdTe epitaxial layer (expansion coefficient 4.1×10^{-6} K^{-1}) is about 0.01 % for the temperature difference in this process. Due to the difference of thermal expansion coefficient between the substrate and the epitaxial layer, thermal stress will generate in the epitaxial layer, which makes the epitaxial layer suffer severe strain. The existence of stress not only causes the proliferation of defects in HgCdTe material which results in degraded device performances, but also can cause cracking of HgCdTe films, which leads to the increase in bad pixels in focal plane arrays, or even the whole device failure.

To solve the stress problem of silicon-based HgCdTe materials and meet the application requirements of silicon-based HgCdTe materials, the theoretical analysis of material stress, measurements of material characteristic parameters under stress, and influences of thermal shock stress on material defect are introduced respectively in this section.

3.5.1 Spectral Characteristics of Si-Based HgCdTe Materials

1. Variable temperature PL

Measurements of variable temperature PL spectra of Si-based CdTe epitaxial film materials with different thickness are used to study the thermal stress

dependence on temperature and film thickness through the change of luminescence emission intensity and peak position in the cooling process. Experiments use a UV laser as the excitation source to measure the PL spectrum of samples under variable temperature. Cooling is implemented by closed mechanical cryocooler. The excitation of the sample is by focused Ar ion laser, the PL is detected by a photomultiplier tube and finally data is recorded by using a photon counting system and a computer.

Figures 3.114 and 3.115 are the variable temperature PL spectra of Si-based CdTe epitaxial film with thickness of 4 and 12 μm, respectively. Both of these figures show the same trend, namely, the luminescence intensity is strengthening, and the main peak position is moving to the direction of high energy with the decreasing temperature. The main peak of PL spectra corresponds to the band edge transition of the material, that is, the valence band transits to the conduction band. The band renormalization occurs due to thermal expansion and electron-phonon interaction [188, 189] and the band gap of the material appears to drift when the temperature decreases. The stress will be another factor to be considered in this section.

Figure 3.116 shows the change of the band gap of Si-based CdTe film with temperature. The thicknesses are 4 and 12 μm, respectively. The band gap of the CdTe crystal materials without any stress is shown in the figure. The following main conclusions can be drawn according to Fig. 3.116:

(1) In the cooling process of Si-based CdTe epitaxial film, both from the growth temperature down to room temperature and from room temperature to liquid nitrogen temperature, thermal stress will generate in the epitaxial films. The stress is tensile stress, which is consistent with the early results of high-resolution *X*-ray diffraction measurements.
(2) The thermal stress of epitaxial films depends on temperature, namely the thermal stress decreases with the reduction of temperature, which is mainly due to the difference of thermal expansion coefficient between the CdTe epitaxial thin film and the Si substrate. Thermal stress of the CdTe epitaxial materials depends on the film thickness. Thermal stress in thicker CdTe film is smaller than that in thinner CdTe film, which is consistent with precalculated results. However, the stress reduction rate in thicker CdTe film is faster than that in thinner CdTe film during the cooling process, which leads the thicker CdTe film more likely to fracture under thermal cycling. This result is also consistent with the thermal shock experiments.

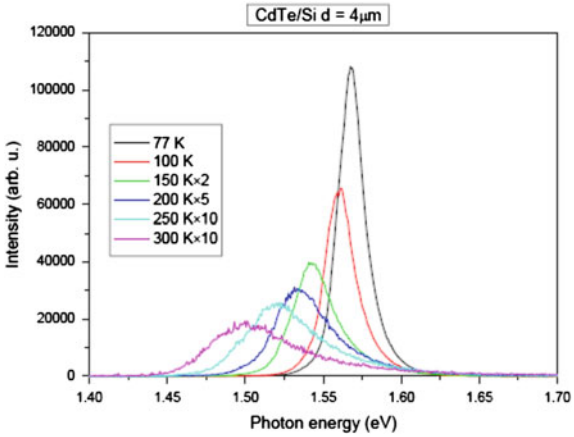


Fig. 3.114 Variable temperature photoluminescence spectra of Si-based CdTe film with thickness 4 μm

Fig. 3.115 Variable temperature photoluminescence spectra of Si-based CdTe film with thickness 12 μm

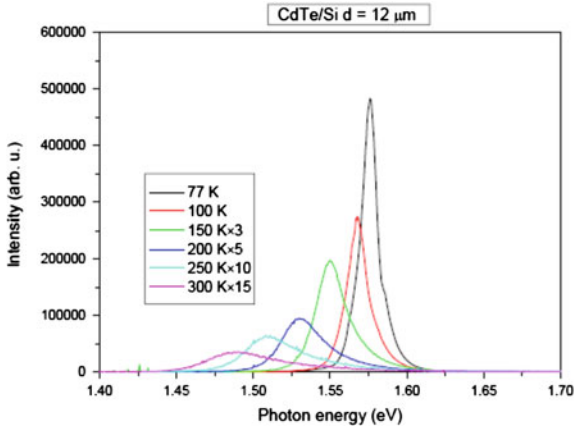
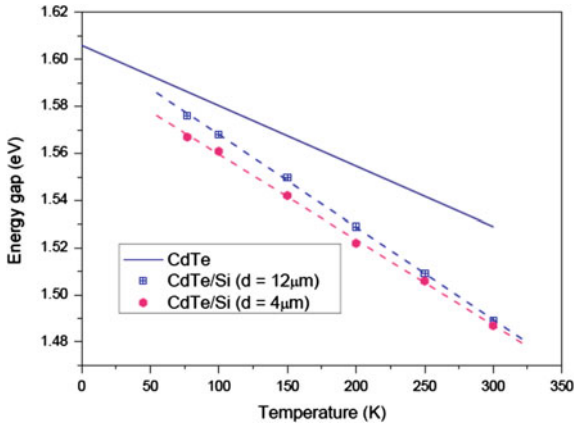


Fig. 3.116 Band gaps of Si-based CdTe films as a function of temperature and the comparison to unstrained CdTe crystal materials



2. Room temperature Raman spectrum

Stress can change the crystal atoms distance, and thereby change the interatomic bonding force constant. This will result in change of Raman frequency of phonons in crystals. Therefore, Raman scattering spectra can be used to study the stress of the film. The frequency, intensity, and FWHM of the characteristic peaks of Raman scattering are related to material composition, grain integrity, grain size, lattice constant of the substrate, thermal expansion coefficient, and other factors. CdTe forms a tetrahedral crystal. When the film is subjected to compressive stress, Raman scattering peaks moving toward a high wave number. On the contrary, when it is subjected to tensile stress the Raman scattering peaks shifted toward a low wave number. As shown in Fig. 3.117, the Raman peaks at 123 and 141 cm^{-1} are the basic characteristic peak frequencies of CdTe film Raman scattering. They correspond to the energies generated by the vibration of A1 (Te) phonons and E (Te) phonons in CdTe [190]. Because the A1 (Te) and E (Te) phonon vibration Raman peaks of CdTe material without any stress is located in 121 and

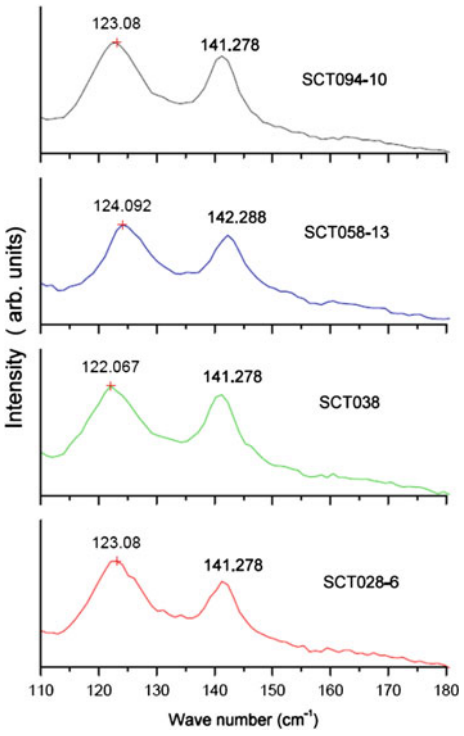


Fig. 3.117 Raman scattering spectra of Si-based CdTe film

Table 3.14 Si-based CdTe film stress

Sample	A1 phonon peak (cm^{-1})	E phonon peak (cm^{-1})	A1 phonon frequency offset (cm^{-1})	E phonon frequency offset (cm^{-1})	Tensile stress σ (MPa)
SCT028-6	123.08	141.278	2.08	0.278	0.312
SCT038	122.067	141.278	1.067	0.278	0.160
SCT058-13	124.09	142.288	3.09	1.288	0.464
SCT094-10	123.08	141.278	2.08	0.278	0.312

141 cm^{-1}, Si-based CdTe epitaxial film materials series of the samples are subject to certain amount of stress. Studies show that the relation of Raman peak position shift and stress in CdTe can be expressed as $\Delta\omega$ (cm^{-1}) = $k\,\sigma$, wherein the pressure coefficient k is 0.7 cm^{-1}/kbar. Therefore, the quantitative estimation results of stress in the Si-based CdTe epitaxial film are shown in Table 3.14.

3. Photoluminescence spectrum

According to the relationship between the energy band and stress in semiconductor physics, the materials band gap E_{1g} under tensile stress will be less than the unstrained band gap E_g and E_{2g} under compressive stress will be greater than E_g [191]. The main peaks of PL spectra correspond to band edge transitions of materials, namely the transition from the valence band to the conduction band, and accordingly, its peak energy corresponds with the band gap of a semiconductor material. PL spectra for a series of Si-based CdTe epitaxial films at room temperature and liquid nitrogen temperature are shown in Figs. 3.118 and 3.119, respectively.

Based on the experimental measurement, and the relationship between the changes of band gap and stress [192],

$$\bar{E} - E_0 = \frac{\sigma}{c_{11} + 2c_{12}} \left(D_d^c - D_d^v\right)$$

The stress of Si-based CdTe epitaxial film at liquid nitrogen temperature can be calculated as follows:

$$\sigma_{SCT028-8} = -1.43 \times 10^7\ \text{Pa}$$
$$\sigma_{SCT038} = -7.58 \times 10^6\ \text{Pa}$$
$$\sigma_{SCT058-13} = -1.07 \times 10^7\ \text{Pa}$$
$$\sigma_{SCT094-10} = -1.38 \times 10^7\ \text{Pa}$$

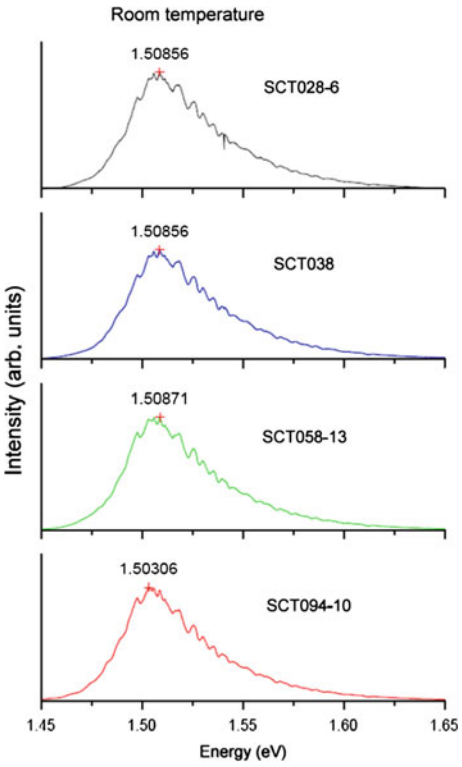


Fig. 3.118 Photoluminescence spectrum of Si-based CdTe films at room temperature

3.5.2 Theoretical Analysis of Stress of Multilayer Structure Materials

1. Modified theoretical model of bimetal strip

A bimetal strip model [193, 194] is mainly used to analyze the stress issues of the multilayer semiconductor structure for one-dimensional situation. The thermal stress comes from the difference of thermal expansion coefficients in different layers. According to the boundary conditions, the sum of average stress of each layer is zero. After bending deformation, the sum of the total bending moment and the moment of each layer is zero. Due to the shear strain at the interface of each layer being equal, the relationship between curvature R, stress E, and number of layers, thickness, Young's modulus, temperature and coefficient of thermal expansion could be deduced. However, the model does not consider the mismatch strain caused by lattice mismatch [195]. Dislocation caused by lattice mismatch must be taken into account to modify the model. In practice, the energy of the system can be divided into two parts [196–198], namely

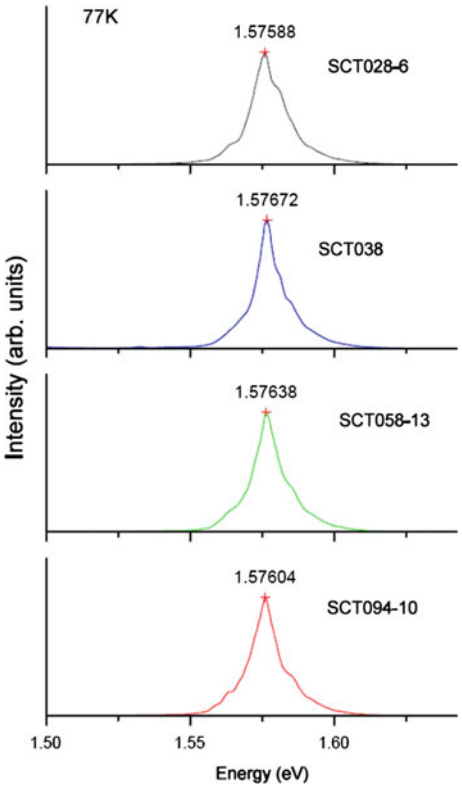


Fig. 3.119 Photoluminescence spectrum of Si-based CdTe films at liquid nitrogen temperature

$$E = E_{\text{strain}} + E_{\text{dislocation}} \tag{3.18}$$

wherein

$$E_{\text{strain}} = Bh\left(\varepsilon_1 - \frac{b}{s}\right)^2 \tag{3.19}$$

$$E_{\text{dislocation}} = \frac{\mu b^2}{4\pi(1-\nu)}\frac{2}{S}\ln\left(\frac{\beta h}{b}\right) \tag{3.20}$$

In Eq. (3.19), the elastic energy factor $B = 2\mu\ (1 + \upsilon)/(1 - \upsilon)$, μ and υ denote the shear modulus and Poisson's ratio of the material, h the film thickness, $\varepsilon 1$ the initial crystal lattice misfit strain, b the Burgers vector, S the distance between dislocations, and β in Eq. (3.20) is a parameter for measurement of energy in the β-core central region, generally 4 for semiconductor. In fact, the energy of the system is a function of dislocation density. Dislocation density can be obtained from the derivation of the energy. Figure 3.120 shows the relationship between the

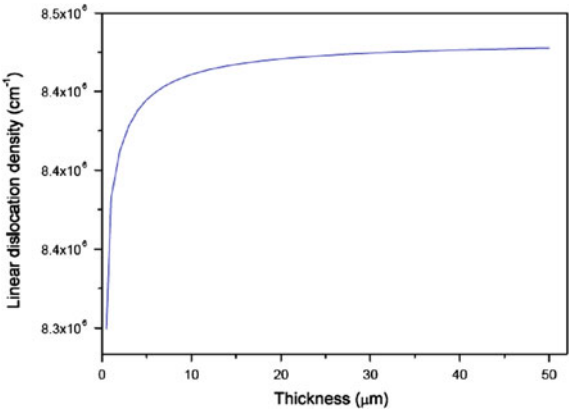


Fig. 3.120 Dislocation density due to lattice mismatch in Si-based CdTe films as a function of film thickness

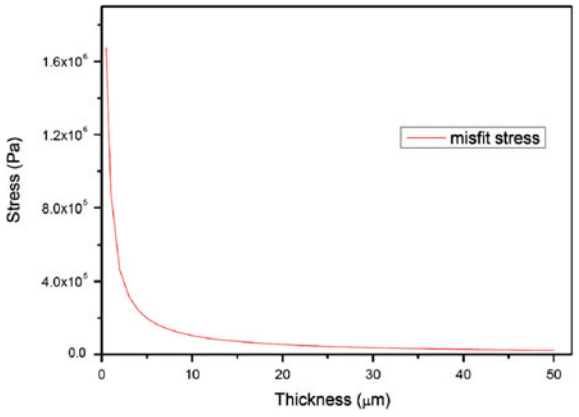


Fig. 3.121 Mismatch stress in Si-based CdTe films as a function of film thickness

dislocation density due to lattice mismatch in Si-based CdTe films and the film thickness. The corresponding change of mismatch stress is shown in Fig. 3.121.

Figure 3.122 indicates the stress state of Si-based CdTe film at liquid nitrogen temperature which is calculated based on the bimetal strip model, and the influence of the substrate thickness on the calculation results is also considered.

After taking both the thermal stress and the lattice mismatch strain into account, the stress of Si-based CdTe film is shown in Fig. 3.123, where the parameter J represents the weight between two kinds of stress in the system. The comparison with the experimental results of PL is shown in Fig. 3.124.

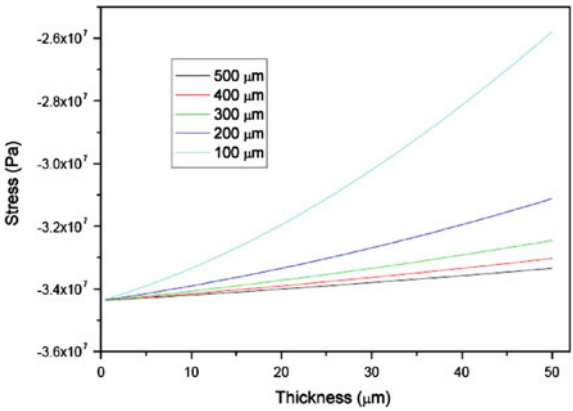


Fig. 3.122 Stress of Si-based CdTe film at liquid nitrogen temperature

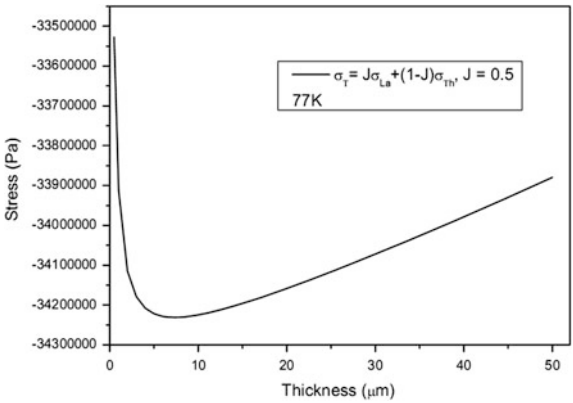


Fig. 3.123 Stress of Si-based CdTe film by taking both the thermal stress and the lattice mismatch strains into account

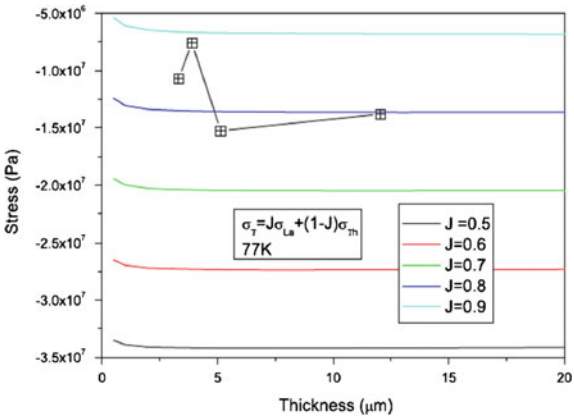


Fig. 3.124 The comparison of theoretical calculation with the photoluminescence experiments of stress in Si-based CdTe film at liquid nitrogen temperature

2. Modified theoretical model of multilayer strip

Because the mismatch stress(lattice mismatch stress and thermal mismatch stress) between epitaxial films and substrate as well as between the epitaxial films and epitaxial films form a force couple, the heterogeneous structure must bend to balance the torque that is generated by the mismatch stress between the epitaxial films and substrate. Therefore, the strain can be divided into the misfit strain and bending strain. Correspondingly the stress can be divided into the mismatch stress and the bending stress, respectively. According to the balance principle of multilayer strip, the film/substrate heterogeneous structures must meet the three equilibrium conditions of a multilayer strip [199–203]. First of all, the total mismatch stresses of the heterostructures are zero. At the same time, the total bending stresses of heterostructures also are zero. In addition, the heterostructure of bending moment must also meet the equilibrium conditions. According to the multilayer strip model, the stress distribution of the substrate and epitaxial film can be expressed as the following equation,

$$\sigma_s = \frac{\Delta T E_s E_f t_f \left\{ E_f t_f^3 + E_s t_s \left[6\gamma t_f - 3(t_f - 2\gamma)t_s - 2t_s^2 \right] \right\} (\alpha_f - \alpha_s)}{E_f^2 t_f^4 + E_s^2 t_s^4 + 2E_f E_s t_f t_s \left(2t_f^2 + 3t_f t_s + 2t_s^2 \right)} \quad (0 \le \gamma \le t_s) \tag{3.21}$$

$$\sigma_f = \frac{\Delta T E_s E_f t_s \left\{ E_s t_s^3 + E_f t_f \left[4t_f + 6(t_s - \gamma)t_s + t_f(9t_s - 6\gamma) \right] \right\} (\alpha_s - \alpha_f)}{E_f^2 t_f^4 + E_s^2 t_s^4 + 2E_f E_s t_f t_s (2t_f^2 + 3t_f t_s + 2t_s^2)} \quad (t_s \le \gamma \le t_s + t_f) \tag{3.22}$$

The stress of substrate and epitaxial film in the Si-based CdTe film at liquid nitrogen temperature can be calculated by using the above equation, and the results are shown in Figs. 3.125 and 3.126 respectively.

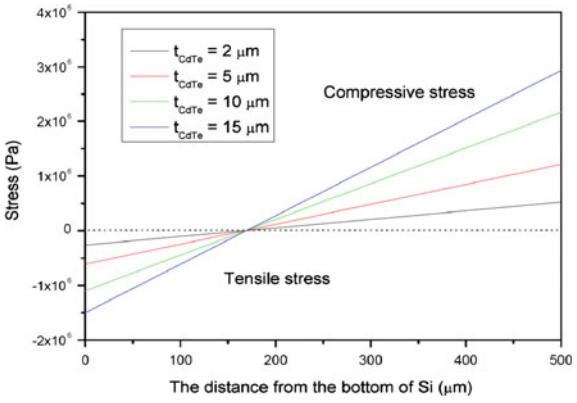


Fig. 3.125 Stress of Si-based CdTe substrate at liquid nitrogen temperature

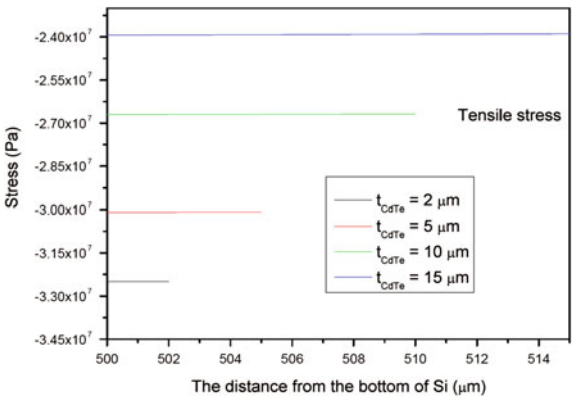


Fig. 3.126 Stress of Si-based CdTe film with thickness 2, 5, 10, 15 μm at liquid nitrogen temperature

However, the multilayer strip model is based on elastic approximation [204, 205]. In fact, in the Si-based CdTe system the epitaxial films in general undergo plastic deformation due to the large lattice and thermal mismatch between the substrate and the epitaxial layer. The bimetal strip model should be modified in order to describe the system better. Generally, the deformation state of epitaxial films includes pure elastic deformation, partial plastic deformation and pure plastic deformation, which can be described by Fig. 3.127. The type of the deformation depends on the relative stress state and the yield strength of the material. The yield strength ε represents the relationship between stress and strain in the stage of plastic deformation and it is conceptually equivalent to the elastic modulus of the elastic deformation stage, as illustrated in Fig. 3.128. E, H_f and σ_γ in the figure represent Young's modulus, strain hardening and initial yield stress. The experimental values of the yield strength of the common semiconductor materials can be found in Ref. [206].

According to the bimetal strip model,

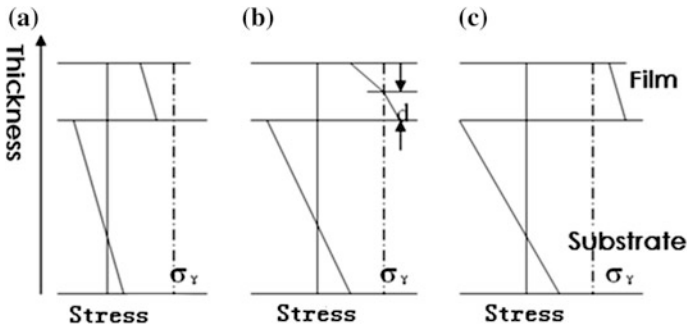


Fig. 3.127 Film stress. **a** Pure elastic; **b** partially in plastic; **c** pure plastic

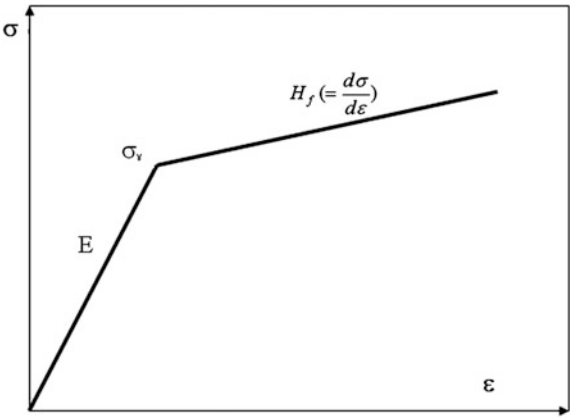


Fig. 3.128 Diagram of yield strength

$$\sigma_f^{\max} = \sigma_f|_{y=t_s} = \frac{\Delta T E_s E_f t_s \left(E_s t_s^3 + 4E_f t_f^3 + 3E_f t_f^2 t_s\right)(\alpha_s - \alpha_f)}{E_f^2 t_f^4 + E_s^2 t_s^4 + 2E_f E_s t_f t_s \left(2t_f^2 + 3t_f t_s + 2t_s^2\right)} \tag{3.23}$$

$$\sigma_f^{\text{top}} = \sigma_f|_{y=t_f+t_s} = \frac{\Delta T E_s E_f t_s \left(E_s t_s^3 - 2E_f t_f^3 - 3E_f t_f^2 t_s\right)(\alpha_s - \alpha_f)}{E_f^2 t_f^4 + E_s^2 t_s^4 + 2E_f E_s t_f t_s \left(2t_f^2 + 3t_f t_s + 2t_s^2\right)} \tag{3.24}$$

The calculated results for Si-based CdTe epitaxial film system show that the minimum and maximum stresses are located in the interface and the surface of the film.

When the stress in the epitaxial film is greater than the yield strength, it can be considered that partial plastic deformation occurs. When the minimum stress in the film is greater than the yield strength, it can be considered that pure plastic deformation occurs. The critical temperature at which the partial and pure plastic deformation occur can be estimated by the following equations:

$$\frac{1}{\Delta T_1} = \frac{E_s E_f t_s \left(E_s t_s^3 + 4E_f t_f^3 + 3E_f t_f^2 t_s\right)(\alpha_s - \alpha_f)}{\sigma_\gamma \left[E_f^2 t_f^4 + E_s^2 t_s^4 + 2E_f E_s t_f t_s \left(2t_f^2 + 3t_f t_s + 2t_s^2\right)\right]} \tag{3.25}$$

$$\frac{1}{\Delta T_2} = \frac{E_s E_f t_s (E_s t_s^3 - 4H_f t_f^2 (2t_f + 3t_s)(\alpha_s - \alpha_f))}{\sigma_\gamma \left\{E_s t_s \left[E_s t_s^3 - H_f t_f^2 \left(2t_f^2 + 3t_s\right)\right] + E_f t_f \left(H_f t_f^3 + E_s t_s \left(6t_f^2 + 9t_f t_s + 4t_s^2\right)\right)\right\}} \tag{3.26}$$

The results are shown in Fig. 3.129. The system is actually in the fully plastic deformation stage when the temperature decreases to 70 k. If the elastic approximation used to describe the stress state is not accurate enough, the bimetal strip

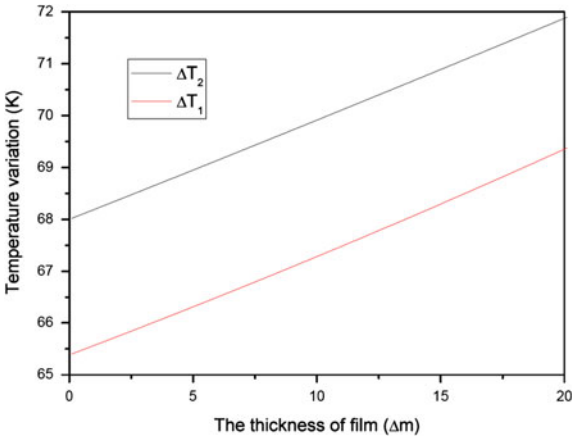


Fig. 3.129 Critical temperature that make partially and pure plastic deformation occur in Si-based CdTe epitaxial film

model can be extended to the plastic deformation range by introducing the yield strength of material. Then, the Eq. (3.27) should be used to calculate the stress in epitaxial films.

$$\sigma_f^{\text{mean}} = \frac{-E_s t_s \left(E_s t_s^3 + H_f t_f^3\right)\left\{E_f[\Delta T H_f(\alpha_f - \alpha_s) - \sigma_\gamma] + H_f \Delta\gamma\right\}}{E_f\left[H_f^2 t_f^4 + E_s^2 t_s^4 + 2E_f E_s t_f t_s\left(2t_f^2 + 3t_f t_s + 2t_s^2\right)\right]} \tag{3.27}$$

Figure 3.130 shows the difference of calculated results of Si-based CdTe epitaxial film before and after correction of the multilayer strip model. After correction, the stress value is smaller than the initial model which is consistent with the previous results from PL experiment, as shown in Fig. 3.131.

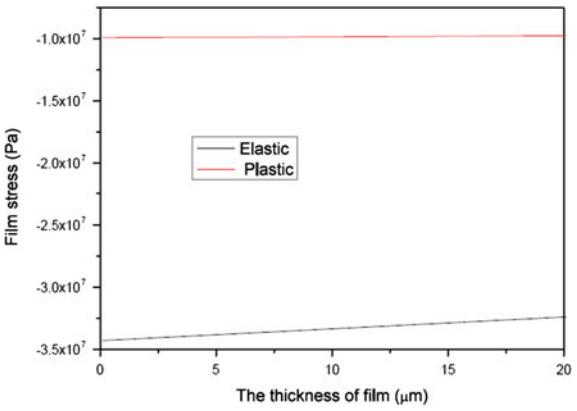


Fig. 3.130 Comparison of calculation results of Si-based CdTe epitaxial film before and after correction of the multilayer strip model

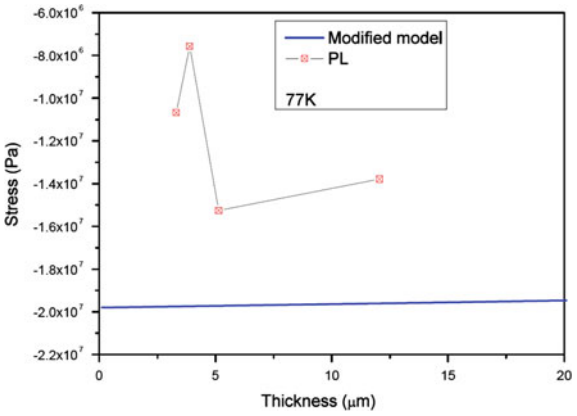


Fig. 3.131 The influence of modified model on calculation results

3. Fracture mechanics analysis

According to the theory of fracture mechanics [207, 208], the cracks preexist in the material before break, which will constantly expand and result in fracture eventually under the influence of continuing stress. In fact, material fracture is the process of stress release. Establishing criteria to avoid fractures is of great significance for practical applications. From the energy point of view, the cracks can be considered the release of strain energies which continues to expand. When the release rate of strain energy is greater than fracture energy of the material itself, fracture occurs. Therefore, the fracture will not occur only when the following expression (3.28) is satisfied [209],

$$\beta \frac{\sigma^2 h}{M} < \Gamma \tag{3.28}$$

Wherein β is a parameter related to the fracture mode, σ stress, h the film thickness, M the modulus of elasticity and Γ the fracture energy. For a certain material, the fracture is determined by both stress and film thickness. This explains why the CdTe film is prone to fracture when the layer thickness exceeds a certain value in the Si-based CdTe epitaxial film system although the stress decreases with the increase of the film thickness. According to the results calculated from the corrected bimetal strip model, the stress decreases with the increase of the film thickness. But the magnitude of change is relatively small. Compared to the amount of change of the epitaxial film thickness, the change of stress is relatively small. Thus, according to the expression (3.28), the thin epitaxial film is less prone to fracture.

3.6 Summary

Using the first-principles methods, we have an insight into physical basis about As-passivation on Si(211) surface. Simulation results have shown that a zigzag structure will form on the Si surface after adsorption of the As atoms. These As atoms will substitute for Si atoms on the surface with an energy barrier 0.37 eV/adatom. Arsenic can saturate part of the dangling bonds and weaken the surface states. The enlargements of silicon substrate surface caused by arsenic can reduce the mismatch between the substrate and the adsorption materials, and consequently improve the quality of crystal. The theoretical simulation results further show that the sticking coefficient of Cd in the interface has been increased because of As atoms. The larger sticking coefficient will avoid replacement of Te at the step edge and will make Si(211) B-polarity surface the most stable interface in MBE growth of HgCdTe/Si.

Several technologies have been improved, such as the surface deoxygenation process, adjusting the crystal orientation of the substrate, lowering the temperature of the buffer layer nucleation, increasing growth temperature of the buffer layer and etc. Qualities of CdTe/Si buffer layers and epitaxial HgCdTe materials have been greatly improved. The twinning FWHM of 3 inch MBE HgCdTe/Si materials has reached 52 arcsecs. The half width of the in-plane distribution uniformity has reached 4.29 arcsecs. The composition uniformity of the materials (component standard deviation) has reached 0.0009. The success of HgCdTe/Si MBE technology laid the foundation for the development of large-scale HgCdTe infrared focal plane technology.

With the aid of CdTe/Si surface treatment technology and CdTe thin epitaxial substrate technology, HgCdTe/Si(211)B LPE material has been obtained. By using high growth temperatures, the dislocation density of materials can be reduced by one order of magnitude. As shown in measured materials properties, the dislocation density of HgCdTe has been reduced to $(5–8)\times10^5$ cm^3.

First-principle calculation results show that As impurity atoms, substituting for Hg or Te atoms, will form covalent bonds in HgCdTe. In aHg-rich condition, the formation energy difference between As_{Te} and As_{Hg} is small. This is conducive to migration of As impurities from Hg site to Te site forming donor As_{Te}. In a Te-rich condition, As tetramer structure associated with the Hg vacancy is more conducive for As-doping than As dimer structure. Complex impurity As tetramer-V_{Hg} has the same amphoteric behavior of As Singlet-V_{Hg} and inhibits the probability of becoming donors. The theoretical model well explained the electrical properties of As-doped HgCdTe epitaxial material with the n-type and p-type heat treatment.

Through the analysis of the density of states, single-electron energy levels, and ionization energy, the dominant form of defects in as-grown As-doped HgCdTe are V_{Hg}, As_{Hg}, As_2 combined by V_{Hg}, and As atoms. The structure that two As atoms bounded at the same side of V_{Hg} will form a deep level near the top of the valence band. There are two deep levels in the vicinity of the band gap about 1/3 with energy difference of 18 meV when the two As atoms are bounded at the split side of V_{Hg}.

These two deep levels are half-filled states. Arsenic is the most important factor for the highly compensated as-grown HgCdTe materials. This gives a reasonable explanation for weak p-type conductivity behavior in low doping concentration as-grown HgCdTe. Both theoretical and experimental results show that increase in the mercury pressure favors transformation from various types of arsenic defects to acceptor As_{Te}. Based on these results, the process conditions of 285 °C heat treatment to achieve 100 % activation of the dopant atoms has been realized in the experiments. By using a beam source furnace, the issues of consistency in heavily doped epitaxial HgCdTe and the temperature windows this material has been solved.

The stress state in Si-based CdTe epitaxial film can be reasonably described by the bimetal strip model theory which contains the effects of misfit dislocation and material yield strength. It also can be nondestructively examined by observing the peak position and the strength variation of the PL and Raman spectrum.

References

1. Faurie JP, Sporken R, Sivananthan S, Lange MD. New development on the control of homoepitaxial and heteroepitaxial growth of CdTe and HgCdTe by MBE. J Cryst Growth. 1991;111(1–4):698–710.
2. Badano G, Chang Y, Garland JW, Sivananthan S. In-situ ellipsometry studies of adsorption of Hg on CdTe(211)B/Si(211) and molecular beam epitaxy growth of HgCdTe(211)B. J Electron Mater. 2004;33(6):583–9.
3. Niraula M, Yasuda K, Ohnishi H, Takahashi H, Eguchi K, Noda K, Agata Y. Direct growth of high-quality thick CdTe epilayers on Si(211) substrates by metalorganic vapor phase epitaxy for nuclear radiation detection and imaging. J Electron Mater. 2006;35(6):1257–61.
4. Wang YZ, Chen L, Wu Y, et al. Heteroepitaxy of CdTe on tilting Si(211) substrates by molecular beam epitaxy. J Cryst Growth. 2006;290(2):436–40.
5. Million A, Dhar NK, Dinan JH. Heteroepitaxy of CdTe on {211}Si substrates by molecular beam epitaxy. J Cryst Growth. 1996;159(1–4):76–80.
6. Rujirawat S, Almeida LA, Chen YP, Sivananthan S, Smith DJ. High quality large-area CdTe (211)B on Si(211) grown by molecular beam epitaxy. Appl Phys Lett. 1997;71(13):1810–2.
7. Delyon TJ, Rajavel D, Johnson SM, Cockrum CA. Molecular-beam epitaxial-growth of CdTe(112) on Si(112) substrates. Appl Phys Lett. 1995;66(16):2119–21.
8. Bornfreund R, Rosbeck JP, Thai YN, Smith EP, Lofgreen DD, Vilela MF, Buel AA, Newton MD, Kosai K, Johnson SM, de Lyon TJ, Jensen JE, Tidrow MZ. High-performance LWIR MBE-grown HgCdTe/Si focal plane arrays. J Electron Mater. 2007;36(8):1085–91.
9. Varesi JB, Buell AA, Peterson JM, Bornfreund RE, Vilela MF, Radford WA, Johnson SM. Performance of molecular-beam epitaxy-grown midwave infrared HgCdTe detectors on four-inch Si substrates and the impact of defects. J Electron Mater. 2003;32(7):661–6.
10. Xiangliang Fu. MBE HgCdTe on Si and Ge substrates. Infrared. 2005;9:19–24.
11. Zavitz DH, Evstigneeva A, Singh R, et al. Influence of arsenic on the atomic structure of the Si(112) surface. J Electron Mater. 2005;34(6):839–45.
12. Jaime-Vasquez M, Martinka M, Jacobs RN, et al. In-Situ spectroscopic study of the As and Te on the Si(112) surface for high-quality epitaxial layers. J Electron Mater. 2006;35 (6):1455–60.
13. Chen L. Doctoral thesis. Research on MBE on large scale Si based CdTe composite substrate. Shanghai: Shanghai Institute of Technical Physics, CAS; 2005.
14. Brill G, Chen Y, Dhar NK, et al. Nucleation of ZnTe/CdTe epitaxy on high-Miller-index Si surfaces. J Electron Mater. 2003;32(7):717–22.

15. Fulk C, Sivananthan S, Zavitz D, et al. The structure of the Si(211) surface. J Electron Mater. 2006;35(6):1449–54.
16. Fulk C, Sporken R, Dumont J, et al. Arsenic deposition as a precursor layer on Silicon (211) and (311) surfaces. J Electron Mater. 2005;34(6):846–50.
17. Dabrowski J, MusSig HJ. Silicon surfaces and formation of interfaces. Singapore: World Scientific; 2000.
18. Lander JJ, Gobeli GW, Morrison J. Structural properties of cleaved Silicon and germanium surfaces. J Appl Phys. 1963;34(8):2298–306.
19. Schlier RE, Farnsworth HE. Structure and adsorption characteristics of clean surfaces of germanium and Silicon. J Chem Phys. 1959;30(4):917–25.
20. Krüger P, Pollmann J. Ab initio calculations of Si, As, S, Se, and Cl adsorption on Si(001) surfaces. Phys Rev B. 1993;47(4):1898–910.
21. Olmstead MA, Bringans RD, Uhrberg RIG, et al. Arsenic overlayer on Si(111): removal of surface reconstruction. Phys Rev B. 1986;34(8):6041–4.
22. Centoni SA, Sadigh B, Gilmer GH, et al. First-principles calculation of free Si(100) surface impurity enrichment. Appl Phys Lett. 2005;87(23):232101.
23. Mankefors S. Ab initio study of the Ge(211) and Si(211) bulk terminated surfaces. Surf Sci. 1999;443(1–2):99–104.
24. Grein CH. First principles calculations of Si(211) surface reconstructions. J Cryst Growth. 1997;180(1):54–60.
25. Dhar NK, Goldsman N, Wood CEC. Tellurium desorption kinetics from (112) Si: Si-Te binding energy. Phys Rev B. 2000;61(12):8256–61.
26. Gupta BC, Batra IP, Sivananthan S. Growth of Te on As-exposed Si(211): electronic structure calculations. Phys Rev B. 2005;71(7):075328.
27. Sen P, Batra IP, Sivananthan S, et al. Electronic structure of Te- and As-covered Si(211). Phys Rev B. 2003;68(4):045314.
28. Prokes SM, Glembocki OJ. Studies of the formation of Ga and Al wires on Si(112) facet surfaces. J Mater Sci Mater Electron. 2001;12(4–6):277–83.
29. Baski AA, Whitman LJ. A scanning-tunneling-microscopy study of hydrogen adsorption on Si(112). J Vacu Sci Tech A. 1995;13(3):1469–72.
30. Erwin SC, Baski AA, Whitman LJ, et al. Frenkel-Kontorova model of vacancy-line interactions on Ga/Si(112). Phys Rev Lett. 1999;83(9):1818–21.
31. Michel EG, Etelaniemi V, Materlik G. Adsorption Sites of Br on Si(211) investigated with X-ray standing wave fields. Surf Sci. 1992;270:89–93.
32. Cho ES, Kim MK, Park JW, et al. Photoemission study on the Sb-induced reconstruction of the Si(112) surface. Surf Sci. 2005;591(1–3):38–44.
33. Segall MD, Lindan PJD, Probert MJ et al. First-principles Simulation: ideas, illustrations and the CASTEP code. J Phys Cond Matt. 2002;14(11):2717–44.
34. Vanderbilt D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. Phys Rev B. 1990;41(11):7892–5.
35. Perdew JP, Wang Y. Accurate and simple analytic representation of the electron-gas correlation energy. Phys Rev B. 1992;45(23):13244–9.
36. Fischer TH, Almlof J. General methods for geometry and wave function optimization. J Phys Chem. 1992;96(24):9768–74.
37. Pack JD, Monkhorst HJ. Special points for Brillouin-zone integrations—a reply. Phys Rev B. 1977;16(4):1748–9.
38. Huang Y, Chen XS, Zhu XY, et al. Theoretical studies about absorption on silicon surface. Int J Mod Phys B. 2007;21(15):2577–614.
39. Baski AA, Whitman LJ. Quasiperiodic nanoscale faceting of high-index Si surfaces. Phys Rev Lett. 1995;74(6):956–9.
40. Huang Y, Chen XS, Duan H, et al. Selective growth of CdTe on Si(211): first-principle calculations. J Electron Mater. 2007;36(8):925–30.
41. Sun L. Doctoral thesis. First principle research on infrared photoelectron material HgCdTe functional doping and defects. Shanghai: Shanghai Institute of Technical Physics, CAS;2005.

42. Benson JD, Stoltz AJ, Varesi JB, et al. Surface structure of plasma-etched (211)B HgCdTe. J Electron Mater. 2005;34(6):726–32.
43. Rujirawat S, Smith DJ, Faurie JP, et al. Microstructural and optical characterization of CdTe (211)B/ZnTe/Si(211) grown by molecular beam epitaxy. J Electron Mater. 1998;27(9): 1047–52.
44. Kroemer H. Nobel Lecture: QuaSielectric fields and band offsets: teaching electrons new tricks. Rev Mod Phys. 2001;73(3):783–93.
45. Brill G, Chen Y, Dhar NK, et al. Nucleation of ZnTe/CdTe epitaxy on high-Miller-index Si surfaces. J Electron Mater. 2003;32(7):717–22.
46. Berding MA, Sher A, Chen AB. Vacancy formation and extraction energies in semiconductor compounds and alloys. J Appl Phys. 1990;68(10):5064–77.
47. Berding MA, van Schilfgaarde M, Sher A. First-principles calculation of native defect densities in Hg0.8Cd0.2Te. Phys Rev B. 1994;50(3):1519–34.
48. Wei SH, Zhang SB. Chemical trends of defect formation and doping limit in II–VI semiconductors: the case of CdTe. Phys Rev B. 2002;66(15):155211(1–10).
49. Dornhaus R, Nimtz G, Schlicht B. Narrow-gap semiconductors: the properties and applications of the HgCdTe alloy system. Berlin Heidelberg: Springer; 1983.
50. Rogalski A, Adamiec K, Rutkowski J. Narrow-gap semiconductor photodiodes. Bellingham: SPIE Press;2000.
51. Wu Owen K, Kamath GS, Radford WA, et al. Chemical Doping of HgCdTe by Molecular-beam Epitaxy. J Vac Sci Technol A. 1990;8(2):1034–8.
52. Jones CE, James K, Merz J, et al. Status of point defects in HgCdTe. J Vac Sci Technol A. 1985;3(1):131–7.
53. Bubulac LO, Edwall DD, McConnell D, et al. P-on-n Arsenic-activated Junctions in MOCVD LWIR HgCdTe/GaAs. Semicond Sci Technol. 1990;5(3s):45–8.
54. Skauli T, Steen H, Colin T, et al. Auger suppression in CdHgTe heterostructure diodes grown by molecular beam epitaxy using silver as acceptor dopant. Appl Phys Lett. 1996;68 (9):1235–7.
55. Giles NC, Lee Jaesun, Myers TH, et al. Optical properties of undoped and iodine doped CdTe. J Electron Mater. 1995;24(5):691–6.
56. Vydyanath HR, Abbott RC, Nelson DA. Mode of incorporation of phosphorus in Hg0.8Cd0.2Te. J Appl Phys. 1983;54(3):1323–31.
57. Berding MA, Sher A. Amphoteric behavior of arsenic in HgCdTe. Appl Phys Lett. 1999;74 (5):685–7.
58. McGill TC, Collins DA. Prospercts for the future of narrow bandgap materials. Semicond Sci Technol. 1993;8(1S):S1–5.
59. Vydyanath HR, Ellsworth JA, Devaney CM. Electrical activity. Mode of incorporation and distribution coefficient of group V elements in $Hg_{1-x}Cd_xTe$ grown from tellurium rich liquid phase epitxial growth solutions. J Electron Mater. 1986;16(1):13–25.
60. Vydyanath HR. Amphoteric behaviour of group V dopants in (Hg, Cd)Te. Semicond Sci Technol. 1990;5(3S):S213–6.
61. Capper PA. Review of impurity behavior in bulk and epitaxial $Hg_{1-x}Cd_xTe$. J Vac Sci Technol B. 1991;9(3):1667–81.
62. Shi XH, Rujirawat S, Ashokan R, et al. Ionization energy of acceptors in As-doped HgCdTe grown by molecular beam epitaxy. Appl Phys Lett. 1998;73(5):638–40.
63. Shin SH, Arias JM, Zandian M, et al. Enhanced arsenic diffusion and activation in HgCdTe. J Electron Mater. 1995;24(5):609–15.
64. Dai N, Chang Y, Wang XG, et al. Photo-electronic phenomena in narrow gap $Hg_{1-x}Cd_xTe$. Current Appl. Phys. 2002;2(5):365–71.
65. Shin SH, Arias JM, Zandian M, et al. Annealing effect on the p-type carrier concentration in low temperature processed arsenic doped HgCdTe. J Electron Mater. 1993;22(8):1039–47.
66. Zandian M, Chen AC, Edwall DD, et al. P-type arsenic doping of $Hg_{1-x}Cd_xTe$ by molecular beam epitaxy. Appl Phys Lett. 1997;71(19):2815–7.

67. Berding MA, Sher A. Arsenic incorporation during MBE growth of HgCdTe. J Electron Mater. 1999;28(6):799–803.
68. Berding MA, Sher A, Van Schilfgaarde M, et al. Modeling of arsenic activation in HgCdTe. J Electron Mater. 1998;27(6):605–9.
69. Wijewarnasuriya PS, Sivananthan S. Arsenic incorporation in HgCdTe grown by molecular beam epitaxy. Appl Phys Lett. 1998;72(14):1694–6.
70. Kohlera B, Wilkea S, Scheffler M, et al. Force calculation and atomic-structure optimization for the full-potential linearized augmented plane-wave code WIEN. Comp Phys Commun. 1996;94(1):31–48.
71. Payne MC, Teter MP, Allan DC, et al. Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. Rev Mod Phys. 1992;64(4):1045–97.
72. Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. Phys Rev Lett. 1996;77(18):3865–8.
73. Perdew JP, Wang Y. Accurate and simple analytic representation of the electron-gas correlation energy. Phys Rev B. 1992;45(23):13244–9.
74. Monkhorst HJ, Pack JD. Special points for Brillouin-zone integrations. Phys Rev B. 1976;13 (12):5188–92.
75. Sun LZ, Chen XS, Sun YL, et al. Structural and electronic properties of the in Situ impurity AsHg in Hg0.5Cd0.5Te: first-principles study. Phys Rev B. 2005;71(19):193203.
76. Kittel C. Introduction to solid state physics. New York: Wiley;1996.
77. Sturge MD. Solid state physics. In: Seitz F, Turnbull D, Ehrenreich H, editors. New York: Academic;1967. vol. 20, p. 91.
78. Dhar NK, Zandian M, Pasko JG, et al. Planar p-on-n HgCdTe heterostructure infrared photodiodes on Si substrates by molecular beam epitaxy. Appl Phys Lett. 1997;70(13):1730.
79. Gorbach TY, Kuzma M, Smertenko PS et al. Anisotropically etched Si surface and the electrical properties of Si/HgCdTe heterostructures. Thin Solid Films 2003;428(1–2):165–169.
80. Sen P, Batra IP, Sivananthan S, et al. Electronic structure of Te- and As-covered Si(211). Phys Rev B. 2003;68(4):045314.
81. Gupta BC, Batra IP, Sivananthan S. Growth of Te on As-exposed Si(211): electronic structure calculations. Phys Rev B. 2005;71(7):075328.
82. Huang Y. Doctoral thesis. First principle research on GaAs and CdTe epitaxial on Si surface. Shanghai: Shanghai Institute of Technical Physics, CAS;2007.
83. Schi M, Harman TC. Optically pumped LPE-grown $Hg_{1-x}Cd_xTe$ laser. J Electron Mater. 1989;8:191–200.
84. Wijewarnasuriya PS, Sivananthan S. Arsenic incorporation in HgCdTe grown by molecular beam epitaxy. Appl Phys Lett. 1998;72(14):1694–7.
85. Bevan MJ, Chen MC, Shih HD. High-quality p-type HgCdTe prepared by metalorganic chemical vapor depoSition. Appl Phys Lett. 1995;67(23):3450.
86. Aqariden F, Wijewarnasuriya PS, Sivananthan S. Arsenic incorporation in HgCdTe grown by molecular beam epitaxy. J Vac Sci Technol B. 1998;16(3):1309–11.
87. Selamet Y, Badano G, Grein CH, et al. Electrical activation and electrical properties of arsenic-doped Hg 1-xCdxTe epilayers grown by MBE. Proc SPIE. 2001;71:4454.
88. Aqariden F, Shih HD, Kinch MA, et al. Electrical properties of low-arsenic-doped HgCdTe grown by molecular beam epitaxy. Appl Phys Lett. 2001;78(22):3481–4.
89. Berdinbg M, Sher A, Vanschilfgaarde M, et al. Modeling of arsenic activation in HgCdTe. J Electron Mater. 1998;27(6):605–9.
90. Berding MA, Sher A. Amphoteric behavior of arsenic in HgCdTe. Appl Phys Lett. 1999;74 (5):685–8.
91. Wu J, Xu FF, Wu Y, Chen L, Yu MF, He L. P-type activation research of As-doping in MBE HgCdTe films. J Infrared Millim W. 2005;24(2):81–83.

92. Selamet Y, Grein CH, Lee TS, et al. Electrical properties of in Situ As doped $Hg_{1-x}Cd_xTe$ epilayers grown by molecular beam epitaxy. J Vac Sci Technol B. 2001;19(4):1488–91.
93. Fangyu Y. Research on doped HgCdTe middle wavelength and long wavelength infrared luminescence and absorption spectrum. Shanghai: Shanghai Institute of Technical Physics, CAS; 2007.
94. Garland JW, Grein CH, Yang B, et al. Evidence that arsenic is incorporated as As4 molecules in the molecular beam epitaxial growth of HgCdTe:As. Appl Phys Lett. 1999;74(14):1975.
95. Sun LZ, Chen XS, Zhao JJ, et al. Electronic properties and chemical trends of the arsenic in situ impurities in $Hg_{1-x}Cd_xTe$: first-principles study. Phys Rev B. 2007;76(4):045219.
96. Perdew JP, Wang Y. Accurate and simple analytic representation of the electron-gas correlation energy. Phys Rev B. 1992;45(23):13244.
97. Vanderbilt D. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. Phys Rev B. 1990;41(11):7892–5.
98. Chen XJ, Hua XL, Hu JS, et al. Band structures of II–VI semiconductors using gaussian basis functions with separable ab initio pseudopotentials: application to prediction of band offsets. Phys Rev B. 1996;53(3):1377–87.
99. Zhang SB, Northrup JE. Chemical potential dependence of defect formation energies in GaAs: application to Ga self-diffusion. Phys Rev Lett. 1991;67(17):2339–42.
100. Duan H, Chen XS, Huang Y, et al. Microscopic origin of electrical compensation in arsenic-doped HgCdTe by molecular beam epitaxy: density functional study. J Electron Mater. 2007;36(8):890–4.
101. Hedberg K. Trans Am crystallogr Assoc. 1966;2:79.
102. Chang Y, Chu JH, Tang WG, et. al. Acta Phys Sincia. 1995;4–66.
103. Hunter AT, McGill TC. Luminescence from HgCdTe alloys. J Appl Phys. 1981;52(9): 5779–86.
104. Hunter AT, McGill TC. Luminescence studies of HgCdTe alloys. J Vac Sci Technol. 1982;21(1):205–7.
105. Zandian M, Chen AC, Edwall DD, et al. p-type arsenic doping of $Hg_{1-x}Cd_xTe$ by molecular beam epitaxy. Appl Phys Lett. 1997;71(19):2815–8.
106. Berding MA, Vanschilfgaarde M, Sher A. Hg0.8Cd0.2Te native defects: densities and dopant properties. J Electron Mater. 1993;22(8):1005–10.
107. Berding MA, Sher A. Amphoteric behavior of arsenic in HgCdTe. Appl Phys Lett. 1999;74 (5):685–8.
108. Garland JW, Grein CH, Yang B, et al. Evidence that arsenic is incorporated as As4 molecules in the molecular beam epitaxial growth of $Hg_{1-x}Cd_xTe$:As. Appl Phys Lett. 1999;74 (14):1975.
109. Grein CH, Garland JW, Sivananthan S, et al. Arsenic incorporation in MBE grown HgCdTe. J Electron Mater. 1999;28(6):789–92.
110. Boieriu P, Grein CH, Jung HS, et al. Arsenic activation in molecular beam epitaxy grown, in situ doped HgCdTe(211). Appl Phys Lett. 2005;86(21):212106.
111. Selamet Y, Grein CH, Lee TS, et al. Electrical properties of in situ As doped $Hg_{1-x}Cd_xTe$ epilayers grown by molecular beam epitaxy. J Vac Sci Technol B. 2001;19(4):1488–91.
112. Li X, Hu X, Zhao J, et al. Infrared Technology. 1997;2.
113. Li B, Gui YS, Chen ZH, et al. Study of impurity states in p-type $Hg_{1-x}Cd_xTe$ uSing far-infrared spectroscopy. Appl Phys Lett. 1998;73(11):1538–41.
114. Cooper DE, Harrison WA. PosSible negative-U properties of the cation vacancy in HgCdTe. J Vac Sci Technol A. 1990;8(2):1112–5.
115. Lander TJ, Morrsion J. Low voltage electron diffraction study of the oxidation and reduction of silicon. J Appl Phys. 1962;33(6):2089–92.
116. Becker GE, Bean JC. Acceptor dopants in silicon molecular-beam epitaxy. J Appl Phys. 1977;48(8):3395–3399.

117. Shiraki Y, Katayama Y, et al. Molecular beam and solid-phase epitaxies of silicon under ultra-high vacuum. J Crystal Growth. 1978;45:287–91.
118. Bean JC, Becker GE, et al. Dependence of residual damage on temperature during Ar+ sputter cleaning of silicon. J Appl Phys. 1977;48(3):907–13.
119. Zehner DM, White CW, Ownby GW. Preparation of atomically clean silicon surfaces by pulsed laser irradiation. Appl Phys Lett. 1980;36(1):56–9.
120. Wright S, Kroemer H. Reduction of oxides on silicon by heating in a gallium molecular beam at 800 °C. Appl Phys Lett. 1980;36(3):210–1.
121. Ishizaka A, Shiraki Y. Low temperature surface cleaning of silicon and its application to silicon MBE. J. Electrochemical Soc. 1986;133(4):666–71.
122. Chen YP, Sivananthan S, Faurie JP. Structure of CdTe(111)B grown by MBE on misoriented Si(001). J Electron Mater. 1993;22(8):951–957.
123. Dhar NK, Boyd PR, et al. CdZnTe heteroepitaxy on 3' (112) Si interface, surface and layer characteristics. J Electron Mater. 2000;29(6):748–53.
124. Million A, Dhar NK, Dinan JH. Heteroepitaxy of CdTe on (211) Si substrates by molecular beam epitaxy. J Crystal Growth. 1996;159(1–4):76.
125. Dhar NK, Goldsman N. Tellurium desorption kinetics from (112)Si:Si–Te binding energy. Phys Rev B. 2000;61(12):8256–61.
126. Chadi DJ. Theoretical study of the atomic structure of Silicon(211), (111), (331) surfaces. Phys Rev B. 1984;29(2):785–92.
127. Dhar NK, Wood CEC, et al. Heteroepitaxy of CdTe on {211} Si using crystallized amorphous ZnTe templates. J Vac Sci Technol B. 1996;14(3):2366–70.
128. Brill G, Chen Y, Dhar NK, et al. Nucleation of ZnTe/CdTe epitaxy on high-miller-index Si surface. J Electron Mater. 2003;32(7):717–22.
129. Smith DJ, Tsec SCY, et al. Growth and characterization of CdTe/Si heterostructures-effect of substrate orientation. Mater Sci Eng B. 2000;77(1):93–100.
130. de Lyon TJ, Johnson SM, et al. Direct MBE growth of CdZnTe on Si(001) and Si(112) substrates for large-area HgCdTe IRFPAs. SPIE. 1993;2021:114–24.
131. Kawano M, Ajisawa A, Oda N. HgCdTe and CdTe(-1-1-3)B growth on Si(112) 5° off by MBE. Appl Phys Lett. 1996;69(19):2876–8.
132. Chen YP, Sivananthan S, Faurie JP. Structure of CdTe(111)B grown by MBE on misoriented Si(001). J Electron Mater. 1993;22(8):951–7.
133. He L, Wu Y, Chen L, et al. Composition control and surface defects of MBE-grown HgCdTe. J Cryst Growth. 2001;227–228:677–82.
134. Bajaj J, Arias JM, Zandian M, et al. Uniform low defect density molecular beam epitaxial HgCdTe. J Electron Mater. 1996;25(8):1394–401.
135. Ferret P, Zanatt JP, Hamelin R, et al. Status of the MBE technology at Leti LIR for the manufacturing of HgCdTe focal plane arrays. J Electron Mater. 2000;29(6):641–7.
136. Brill G, Velicu S, Boieriu P, et al. MBE growth and device processing of MWIR HgCdTe on large area Si substrates. J. Electron Mater. 2001;30(6):717–22.
137. Almeida LA, Dhar NK, Martinka M, et al. HgCdTe heteroepitaxy on three-inch (112) CdZnTe/Si: Ellipsometric control of substrate temperature. J Electron Mater. 2000;29(6):754–9.
138. Chen L, Wu Y, He L, et al. Surface defects on MBE grown HgCdTe. J Infrared Millim W. 2001;20(6):406–10.
139. Deline VR, et al. A unified explanation for secondary ion yields. Appl Phys Lett. 1978;33 (7):578–80.
140. Wittmaack K. Implications in the use of reactive ion bombardment for secondary ion yield enhancement. Appl Surf Sci. 1981;9(1–4):315–34.
141. Yu ML. Chemical enhancement effects in SIMS analysis. Nucl Instrum Methods Phys Res Sect B. 1986;15(1–6):151–8.
142. Sheng J, Wang L, Lux GE. SIMS characterization of HgCdTe and related II–VI compounds. J Electron Mater. 1996;25(8):1165–71.

143. He L, Wang SL, Yang JR, et al. MBE in situ high temperature annealing of HgCdTe. J Cryst Growth. 1999;201–202:524–9.
144. Capper P. Narrow gap cadmium-based compounds. England: Short Run Press Ltd; 1994.
145. Boukerche M, Reno J, Sou IK, et al. Indium doping of HgCdTe layers during growth by molecular beam epitaxy. Appl Phys Lett. 1986;48(25):1733–5.
146. Boukerche M, Sivananthan S, Wijewarnasuriya PS, et al. Electrical properties of intrinsic p-type shallow levels in HgCdTe grown by molecular-beam epitaxy in the (111)B orientation. J Vac Sci Technol A. 1989;7(2):311–3.
147. Wu OK, Jamba DN, Kamath GS. Growth and properties of In- and As-doped HgCdTe by MBE. J Crystal Growth. 1993;127(1–4):365–70.
148. Chen MC, Parker SG, Weirauch DF. Inhomogeneity model for anomalous hall effects in n-type Hg0.8Cd0.2Te Liquid-phase-epitaxy films. J Appl Phys. 1985;58(8):3150–3.
149. Lou LF, Frye WH. Hall effect and resistivity in liquid-phase-epitaxial layers of HgCdTe. J Appl Phys. 1984;56(8):2253–67.
150. Nimtz G, Bauer G, Dornhaus R, Mueller KH. Transient carrier decay and transport properties in $Hg_{1-x}Cd_xTe$. Phys Rev B. 1974;10(8):3302–10.
151. Sasaki T, Oda N, Kawano M, et al. Mercury annealing effect on the electrical properties of HgCdTe grown by molecular beam epitaxy. J Crystal Growth. 1992;117(1–4):222–6.
152. Temofonte TA, Noreika AJ, Bevan MJ, et al. Low-level extrinsic doping for p- and n-type (100) HgCdTe grown by molecular-beam epitaxy. J Vac Sci Technol A. 1989;7(2):440–444.
153. Meyer JR, Bartoli FJ. Theory for electron mobilities in n-type HgCdTe and CdTe at low temperatures. J Vac Sci Technol. 1982;21(1):237–40.
154. Rafol SB, Wijewarnasuriya PS, Sou IK, et al. Shubnikov-de Haas oscillations in as-grown and annealed molecular-beam-epitaxy-grown HgCdTe alloys doped with indium. J Appl Phys. 1993;73(1):216–25.
155. Myers TH, Harris KA, Anka RW, et al. J Vac Sci Technol B. 1992;10:1438.
156. Reine MB. Review of HgCdTe photodiodes for IR detection. Proc SPIE. 2000;4028:320.
157. Rogalski A. Dual-band infrared detectors. Proc SPIE. 2000;3948:17.
158. He L, Becker CR, Bicknell-Tassius RN, et al. Molecular beam epitaxial growth and evaluation of intrinsic and extrinsically doped Hg0.8Cd0.2Te on (100) Cd0. 96Zn0.04Te. J Appl Phys. 1993;73(7):3305–12.
159. Wijewarnasuriya PS, Yoo SS, Sivananthan S, et al. P-type doping with arsenic in (211)B HgCdTe grown by MBE. J Electron Mater. 1996;25(8):1300–05.
160. Berding MA, et al. Modeling of arsenic activation in HgCdTe. J Electron Mater. 1998;27 (6):605–9.
161. Grein CH, Garland JW, Sivananthan S, et al. Arsenic incorporation in MBE grown $Hg_{1-x}Cd_xTe$. J Electron Mater 1999;28(6):789–792.
162. Lee TS, Garland J, Gren CH, et al. Correlation of arsenic incorporation and its electrical activation in MBE HgCdTe. J Electron Mater 2000;29(6):869–872.
163. Sivananthan S, Wijewarnasriya PS, Aqariden F. Mode of arsenic incorporation in HgCdTe grown by MBE. J Electron Mater 1997;26(6):621–24.
164. Chen AC, Zandian M, Edwall DD, et al. MBE growth and characterization of In situ arsenic doped HgCdTe. J Electron Mater. 1998;27(6):595–99.
165. Chand AD, Goodwin MW, Chen MC, et al. Variation of arsenic diffusion, coefficients in HgCdTe alloys with temperature and Hg pressure: tuning of p on n double layer heterojunction diode properties. J Electron Mater. 1995;24:599–8.
166. Boieriu P, Chen Y, Nathan V. Low-temperature activation of As in $Hg_{1-x}Cd_xTe$(211) grown on Si by molecular beam epitaxy. J Electron Mater. 2002;37:694–8.
167. Wu J. Doctoral thesis. Research on MBE HgCdTe P type As doping. Shanghai: Shanghai Institute of Technical Physics, CAS;2004.
168. Yoshikawa M. Dislocations in Hg1–xCdxTe/Cd1–zZnzTe epilayers grown by liquid-phase epitaxy. J Appl Physics. 1988;63(5):1533–40.
169. James TW, Stoller RE. Blocking of threading dislocations by $Hg_{1-x}Cd_xTe$ expitaxial layers. Appl Phys Lett. 1984;44(1):56–8.

170. Wang CC, Shin SH, Chu M, Lanir M, Vanderwyck AHB. Liquid phase growth of HgCdTe epitaxial layers. J Electrochem Soc Solid-State Sci Technol. 1980;127(1):175–9.
171. Harman TC. Slider LPE of $Hg_{1-x}Cd_xTe$ using mercury pressure controlled growth solutions. J Electron Mater. 1981;10(6):1069–84.
172. Chiang CD, Wu TB, Chung WC, Yang SJ, Pang YM. A new attachment for stable control of mercury pressure in the slider LPE of $Hg_{1-x}Cd_xTe$. J Cryst Growth. 1988;87:161–8.
173. Smith FT, et al. Te-rich liquid phase epitaxial growth of HgCdTe on Si-based substrates. J Electron Mater. 1995;24(9):1287–92.
174. Xiuliang C. Research on HgCdTe epitaxial material defects characteristics. Semiconductor. 2006;27(8):1401–5.
175. Xiuliang C, Jianrong Y. Research on HgCdTe epitaxial material surface etching pits characteristics. Laser Infrared. 2005;35(11):845–8.
176. Smith FT, et al. Te-rich liquid phase epitaxial growth of HgCdTe on Si-based substrates. J Electron Mater. 1995;24(9):1287–92.
177. Qingqing X, Xinqiang C, Yanfeng W, et al. HgCdTe LPE film prepared on (211) CdZnTe substrate. Laser Infrared. 2005;35(11):842–4.
178. Qingqing X, Xinqiang C, Yanfeng W, et al. Growth and properties analysis of HgCdTe LPE material on Si/CdTe composite substrate. Semiconductor. 2007;28(7):1078–82.
179. Meifang Y, Jianrong Y, Li H, et al. Research on MBE HgCdTe film dislocation density. Semiconductor. 1999;20(5):378.
180. Varesi JB, Bornfreund RE, Childs AC, et al. Fabrication of high-performance large-format MWIR focal plane arrays from MBE-grown HgCdTe on 4″ silicon substrates. J Electron Mater. 2001;30:566–73.
181. Maranowski KD, Peterson JM, Johnson SM, et al. MBE growth of HgCdTe on silicon substrates for large format MWIR focal plane arrays. J Electron Mater. 2001;30:619–22.
182. Jain SC, Maes HE, Pinardi K, De Wolf I. Stresses and strains in lattice-mismatched stripes, quantum wires, quantum dots, and substrates in Si technology. J Appl Phys. 1996;79: 8145–65.
183. Steegen A, Maex K. Silicide-induced stress in Si: origin and consequences for MOS technologies. Mater Sci Eng R. 2002;38:1–53.
184. Carmody M, Lee D, Zandian M, Phillips J, et al. Threading and misfit-dislocation motion in molecular-beam epitaxy-grown HgCdTe epilayers. J Electron Mater. 2003;32:710–6.
185. Hu SM. Stress-related problem in silicon technology. J Appl Phys. 1991;70:53–80.
186. Hwang D-MD. Strain relaxation in lattice-mismatched epitaxy. Mater Chem Phys. 1995;40:291–7.
187. Pinardi K, Jain U, Jain SC, Maes HE, et al. Critical thickness and strain relaxation in lattice mismatched II–VI semiconductor layers. J Appl Phys. 1998;83:4724–33.
188. Allen PB, Cardona M. Theory of the temperature dependence of the direct gap of germanium. Phys Rev B. 1983;21:1495–505.
189. Allen PB, Cardona M. Temperature dependence of the direct gap of Si and Ge. Phys Rev B. 1983;27:4760–9.
190. Amirtharaj PM, Pollak FH. Raman scattering study of the properties and removal of excess Te on CdTe surfaces. Appl Phys Lett. 1984;45:789–91.
191. Sugo M, Uchida N, Yamamoto A, et al. Residual strains in heteroepitaxial III–V semiconductor films on Si(100) substrates. J Appl Phys. 1989;65:591–5.
192. Thomas DG. Excitons and band splitting produced by uniaxial stress in CdTe. J Appl Phys. 1961;32:2298–304.
193. Feng Z, Liu H. Generalized formula for curvature radius and layer stresses caused by thermal strain in semiconductor multilayer structures. J Appl Phys. 1983;54:83–5.
194. People R, Bean JC. Calculation of critical layer thickness versus lattice mismatch for $GexSi_{1-x}$/Si strained-layer heterostructures. Appl Phys Lett. 1985;47:322–4.
195. Nishino H, Sugiyama I, Nishijima Y. Misfit stress relaxation mechanism in CdTe(100) and CdTe/ZnTe(100) on a GaAs(100) highly mismatched heteroepitaxial layer. J Appl Phys. 1996;80:3238–43.

196. Payne AP, Lairson BM, Clemens BM. Strain relaxation in ultrathin films: a modified theory of misfit-dislocation energetics. Phys Rev B. 1993;47:13730–6.
197. Huang FY. Theory of strain relaxation for epitaxial layers grown on substrate of a finite dimension. Phys Rev Lett. 2000;85:784–7.
198. Wiesauer K, Springholz G. Critical thickness and strain relaxation in high-misfit heteroepitaxial systems: PbTe1-xSex on PbSe (001). Phys Rev B. 2004;69:245313.
199. Olsen GH, Ettenberg M. Calculated stresses in multilayered heteroepitaxial structures. J Appl Phys. 1977;48:2453–7.
200. Nakajima K. Calculation of stresses in $InxGa_{1-x}As$/InP strained multilayer heterostructures. J Appl Phys. 1992;72:5213–9.
201. Suhir E. Predicted thermally induced stresses in, and the bow of, a circular substrate/thin-film structure. J Appl Phys. 2000;88:2363–70.
202. Hsueh C-H. Modeling of elastic deformation of multilayers due to residual stresses and external bending. J Appl Phys. 2002;91:9652–6.
203. Malzbender J. Mechanical and thermal stresses in multilayered materials. J Appl Phys. 2004;95:1780–2.
204. Hu YY, Huang WM. Elastic and elastic-plastic analysis of multilayer thin films: closed-form solutions. J Appl Phys. 2004;96:4154–60.
205. Zhang Y, Zhao Y. Applicability range of Stoney's formula and modified formulas for a film/substrate bilayer. J Appl Phys. 2006;99:053513.
206. Yonenaga I, et al. Yield strength and dislocation mobility in plastically deformed ZnSe. Phys B. 2006;376–377:771–4.
207. Weertman J. Zener–Stroh crack, Zener–Hollomon parameter, and other topics. J Appl Phys. 1986;60:1877–87.
208. Suo Z. Cracking and debonding of microlaminates. J Vac Sci Technol A. 1993;11:1367–72.
209. Liu XH, Suo Z, Ma Q, Fujimoto H. Developin design rules to avert cracking and debonding in the integrated circuit structures. Eng Fract Mechnan. 2000;66:387–402.

Chapter 4
AlGaN Epitaxial Technology

4.1 Introduction

During the development of semiconductor science and technology, Si and Ge were the first-generation semiconductor material, while GaAs, InP are known as the second-generation semiconductor material. In recent years, rapid developments have occurred in GaN and related materials (including AlN, AlGaN, GaN, InGaN and InN), thus becoming the third-generation of semiconductor material. The band gaps of InN, GaN, AlN are 0.7, 3.4, 6.2 eV, respectively, thus enabling the fabrication of sensors with these materials to cover near infrared, visible light, and ultraviolet wavebands. Therefore, these materials will play important roles in optical displays, memory devices, illumination sources, and detection devices. These recent years have seen the commercialization of GaN-based blue light-emitting diodes and great breakthroughs have been achieved in long-life GaN laser devices [1, 2]. Another important photoelectronic device of the III-group nitride, the GaN-based ultraviolet detector has found an important application in missile warning, ultraviolet communication, fire monitoring, vehicle guidance, and other fields, and got an increasing attraction from the researchers all over the world. Significant international research into this material is currently underway.

The fabrication of large-scale and high-performance GaN-based ultraviolet focal plane arrays is an important step for the development of GaN-based detect or to fulfill the needs mentioned above, while achieving high-quality GaN and AlGaN material is an indispensable foundation, and AlGaN epitaxial technology thus plays a crucial role in the whole development. To master the epitaxial technology of AlGaN material, it is necessary to fully understand the basic principles of epitaxial growth, the methods of material growth monitoring and the basic physical properties of AlGaN. There are several major scientific issues to be resolved, such as how we shall grow high-quality material on a mismatched substrate heteroepitaxially and what we should do to reduce the pre-reaction of Al atom and control the component of Al, as well as the way to increase the surface mobility of Al atoms to

L. He et al., *Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN*,
DOI 10.1007/978-3-662-52718-4_4

improve the quality of the material. These issues must be overcome in order to grow appropriate material.

In this chapter, the main issues of AlGaN epitaxial technology will be introduced in detail, especifically the two-step epitaxial technology which has been a great contribution to the use of GaN-based material. In order to make a thorough explanation of the AlGaN epitaxial technology and mechanism, GaN epitaxial technology will be introduced first, including the basic characteristics of GaN material, the basic principles of metalorganic chemical vapor deposition (MOCVD) epitaxial technology and the in situ monitoring system. Then the AlGaN epitaxial technology will be introduced accordingly, including AlGaN's growth on GaN and the AlN buffer layer, the P doping of GaN, and the analysis of the Alga combination property. Specific growth parameters will also be discussed, but it should be noted that these parameters vary with different fabrication tools. So, it is strongly recommended that attention should be paid to the basic mechanisms, which will help to give an insight into the material growth, characterization, and analysis.

4.2 Basic Properties of GaN-Based Material and Preparation Techniques

This section will begin with the introduction of development progress of GaN-based material, the physical properties of GaN crystal structure, optimum substrate standards, and the use of GaN in ultraviolet detection. Then the basic principle and the equipment structure of MOCVD epitaxial technology will be introduced, as well as the methods of growth monitoring them. A good apprehension of these related topics will be a strong groundwork for the understanding of the epitaxial growth technology of AlGaN material.

4.2.1 The Basic Properties of GaN-Based Material and Its Use in Ultraviolet Detectors

The earliest report of obtaining GaN material from reaction of Ga and ammonia can be dated back to 1932 by Johnson [3]. Later, researchers obtained small nitride crystals through heating GaN or AlN in an ammonia atmosphere [4–7], however, subsequent progress was slow until the 1960s [8–17]. The development was soon accelerated by the dawning technology of MOCVD epitaxial GaN [18] and MBE GaN in 1981 [19]. It was thanks to the use of two-step-growth method that there has been great progress in the research on nitrides since 1980. The milestone work was accomplished by Amanoin 1986 [20], who first found that the use of an AlN buffer layer in low temperature may increase the quality of crystal greatly. Nakamura then found similar results using a low-temperature GaN buffer [21]. P doping has long

been another obstacle toward the mature material growth of GaN device. Amano unintentionally found that Mg acceptor could be activated by a low-energy electron beam [22]. Nakamura also found that Mg acceptor could be activated by thermal treatment at 700–800 °C [23]. From then on, the road to GaN device fabrication has been opened widely, and different kinds of GaN devices have flourished quickly including high-performance blue light-emitting diode, laser detectors, ultraviolet detection, etc., [24] with a following commercialization of light-emitting diodes and laser detectors in particular. Great progress has been attained in solar-blind ultraviolet detectors as a result of the breakthrough on GaN-based material growth. As the band gaps of GaN and AlN are 3.4, 6.2 eV, respectively, so the band width of alloy AlGaN can be tuned continuously between 3.4 and 6.2 eV with corresponding wavelengths of 200–365 nm, which covers the solar-blind area. There are great differences between the material growth of GaN and AlGaN, for the low surface migration ability of atomic Al could cause a poor quality of AlGaN material with respect to both surface topography and dislocation density. After the adoption of the atomic layer epitaxy (ALE) technique, material quality of AlN and AlGaN has been improved significantly [25]. Based on these results, Razeghi from Northwestern University has developed 320 × 256 solar-blind focal plane arrays [26], and M. Asif Khan and Taniyasu have fabricated AlGaN and AlN deep ultraviolet light-emitting diodes [27, 28]. Thus, it can be seen that epitaxial growth technology has greatly pushed the development of high-performance solar-blind ultraviolet detectors.

GaN is a binary compound semiconductor, which is the most studied material among III-group nitrides. It has three crystal structures: wurtzite (hexagonal phase), zinc blende (cubic phase), and halite (NaCl structure) [29]. Only two structures can be observed in single-crystal GaN: wurtzite (a = 3.189 Å, c = 5.185 Å), zinc blende (a = 4.52 Å, $c \approx$ 4.54 Å). The wurtzite structure is more stable than zinc blende for GaN according to thermodynamics. The crystal structure of GaN is dependent on the material of the substrate and the symmetry of the substrate surface. Wurtzite GaN can be grown on substrates such as (0001) Al_2O_3, (111) Si, (111) GaAs, (0001) 6H–SiC, and (0001) ZnO. Zinc blende GaN can be grown on substrates such as (001) GaAs, (001) 3C–SiC, (001) Mg and (001) Si, which are crystalline substrates with a cubic structure [30]. Figure 4.1 is the schematic diagram of these two crystal structures.

Wurtzite and zinc blende structures have the same number of the nearest atoms, but a different number of the second nearest atoms. So, the two crystal structures have the similar physical properties. Since zinc blende GaN is difficult to grow and is metastable, GaN-based material with wurtzite structure is now widely used. Polish scientists have successfully grown bulk GaN with metal Ga and ammonia (NH_3) in high-temperature (1600 °C) and high-pressure (15–20 kbar) [31] environments, although problems still exist to provide large-scale products in a short time. GaN epitaxial deposition has always been realized on heterosubstrates with the absence of single-crystal bulk GaN. Sapphire shares the same symmetry with group-III nitrides and with the wurtzite structure, therefore it is the most popular substrate up to now, and the C surface (0001) of sapphire is used as the substrate,

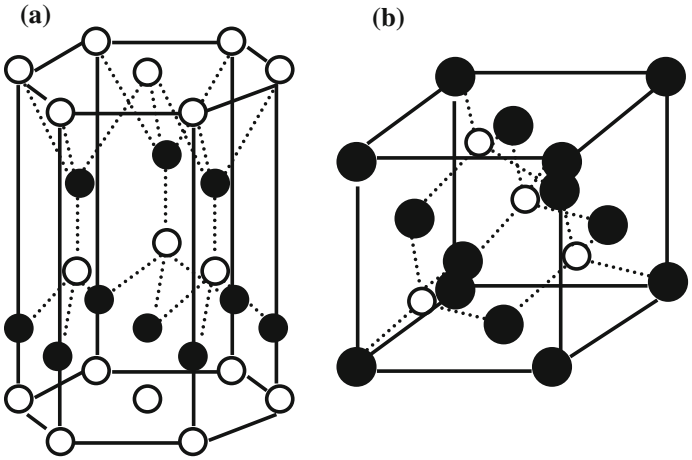


Fig. 4.1 Schematic diagrams of wurtzite (**a**) and zinc blende (**b**)

although it has a large lattice mismatch (16 %) and a different thermal expansion coefficient from the hexagonal GaN epitaxial layer. Owing to a continuous advance in the process technology, the crystal quality of GaN epitaxial layer on sapphire has been greatly improved. At present, most GaN-based detectors are fabricated on sapphire substrates. GaN and AlGaN material with a wurtzite structure can also be achieved on sapphire substrates [24].

4.2.2 MOCVD Epitaxial Deposition System and In Situ Monitoring Method

1. MOCVD system introduction and fundamental reaction processes of GaN-based material

MOCVD is a nonequilibrium growth technology, which originated from the research of single-crystal GaAs growth by Manaservil in 1986. It depends on the transportation of the source gases and the thermal cracking reaction of group-III alkyl compound and group-V hydrogenide [32]. Subsequently, this technology was introduced into the growth of GaAsPy, GaAsSb, and Al compounds by Manaservil and Simpson in 1969 and Manaservil and Hess in 1970 [33, 34]. The group-III organic sources could be liquid such as trimethyl gallium (TMG) or could be solid such as trimethyl indium (TMIn), which were transported into the reaction room by carrier gas through adjusting the temperature of the source bottle and accurate control of the pressure of the source. Group-V sources are generally gaseous hydrogenide (such as NH_3 used for the material growth of GaN). MOCVD technology has now become the most important epitaxial technology in the research of

mesoscopic physics, semiconductor materials and devices, as well as in the field of massive production.

As a general gas phase epitaxial growth equipment, MOCVD can be divided into two types: vertical and horizontal. Low/high pressure and hot/cold wall of the reaction room are other options for MOCVD growth, although high frequency and radiation heating of cold wall is the most popular method. Due to the facts of flammable, explosive, and highly toxic sources and the requirement of multicomposition, large-area, thin, or ultrathin heterogeneous films for MOCVD growth, the design concept for equipment must contain good system sealing, accurately controlled gas flow and temperature, and rapid composition transition. Generally speaking, MOCVD equipment is composed of source supply system, off-gas treatment system, security alarm system, as well as an automatic operation and electric control system [35].

The dynamics of the reaction during MOCVD growth is determined by the source, substrate, growth pressure, carrier gas, and the geometry of the reaction room. Possible chemical reactions and reaction velocity restriction process during the MOCVD growth of GaN with TMG [Ga $(CH_3)_3$] and NH_3 are demonstrated in Fig. 4.2 [36]. As shown in the figure, in spite of the complexity of the growth process, MOCVD can still be divided into the following steps:

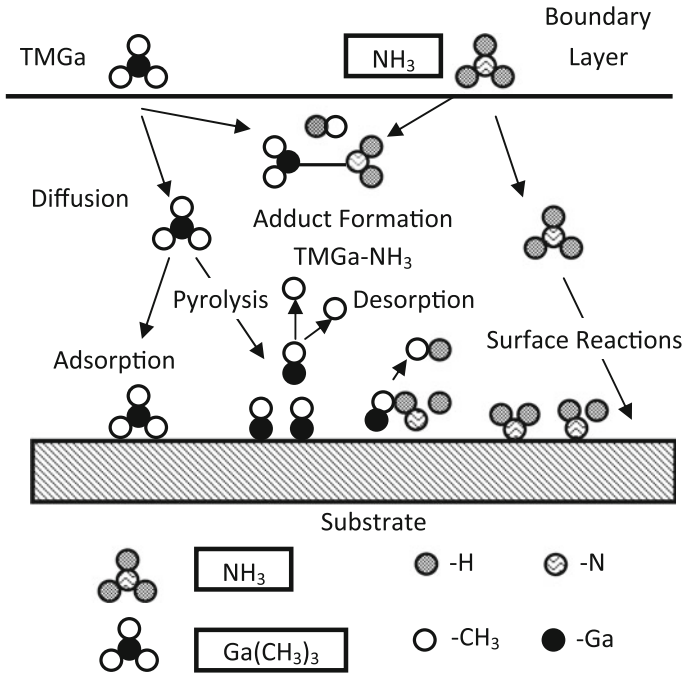


Fig. 4.2 The possible chemical reaction between TMGa and NH_3 and velocity restriction stage during MOCVD growth

(1) Transportation of the source mixture to the epitaxial growth area.
(2) The diffusion of the gas source: material participated in the reaction diffuse through the boundary layer to the growth surface (substrate surface).
(3) The physical adsorption process: the physical adsorption of the reactants on the high-temperature substrate along the surface steps.
(4) Heat decomposition and chemical reaction process: crystal atoms and gas by-products generated by the chemical reaction between the adsorption molecules or between the adsorbent and gas molecules.
(5) Chemical adsorption and the incorporation into the lattice process: the atoms that form the crystal diffuse along the substrate surface to a certain angle of the lattice or enter the crystal lattice.
(6) Desorption process of the by-products: the by-products gather on the surface continuously diffuse through boundary layer to the main flow by desorption.

These above steps occur in order. As the rate of each step is different, the overall growth rate is determined by the slowest step which is named as the rate restriction step. At low temperatures, the chemical reaction of the adsorption on the surface is slowest, and it will determine the speed of the whole epitaxial growth. This process is called the surface control process or dynamic control process. At high temperature, the reaction at the surface is rapid, and the diffusion of the reactants to the surface carried by the main airflow is the slowest, so the growth rate will be determined by the mass transportation, and this process is called mass control process. Epitaxial growth is governed by these steps, depending on the growing conditions.

2. Configuration of in situ monitoring system and its application in MOCVD used to grow GaN material

The growth of GaN material by MOCVD is a complex dynamics process which always contains the high-temperature treatment of sapphire substrate, the nitriding of the substrate, the growth of a low-temperature buffer layer, and the growth of the GaN epitaxial layer [37]. In situ monitoring tool was first introduced by Nakamura into the MOCVD growth of GaN, and it has played a very important role in understanding the growth mechanism of GaN material [38–40]. An in situ "near normal incidence" laser reflectance (LR) monitoring system has been established by the semiconductor Institute of Chinese Academy of Sciences as shown in Fig. 4.3, and a 650 nm laser was used as the light source in this system with the power <1 mW. The incident probe laser beam first will pass through a high-frequency heating coil with an incidence angle of about 1° and then through the inner tube window. The light reflected by the grown sample surface is detected by a Si detector combined with a phase lock amplifier, and the output signal will be sent to an analysis software program.

It is a simple physical principle behind the LR in situ monitoring in the material growth The reflectivity of the substrate with a thin film is given by the following formula [41]:

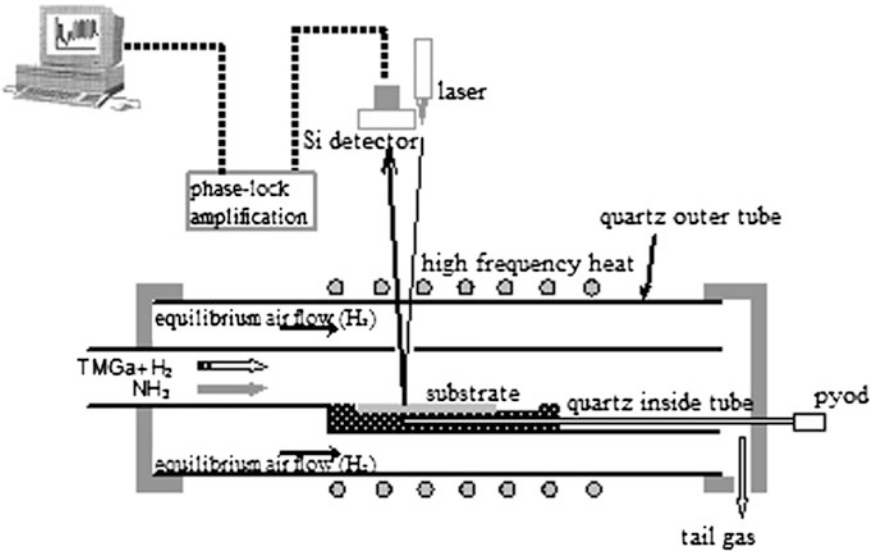


Fig. 4.3 General view of GaN in situ monitoring system

$$R = \frac{r_2 + r_1 \exp(-i\delta)}{1 + r_1 r_2 \exp(-i\delta)}, \quad \delta = 4\pi N_2 d/\lambda \tag{4.1}$$

$$r_1 = \frac{N_1 - N_2}{N_1 + N_2}, \quad r_2 = \frac{N_2 - N_3}{N_2 + N_3} \tag{4.2}$$

$$N_j = n_j - ik_j, \quad k_j = \alpha_j \lambda / 4\pi \tag{4.3}$$

where α is the absorption coefficient of material at the wavelength λ. N_1, N_2, and N_3 are the complex refractive index of the substrate, GaN and the reaction chamber and d is the thickness of the GaN epitaxial layer.

It can be seen from (4.1) that the reflected intensity varies with the GaN film thickness for a monochronic incident light. The peak position of the reflected light is determined by the formula:

$$n_2 d = 2m\lambda/4 \quad m = 1, 2\ldots \tag{4.4}$$

And the minimum position of the reflectance is determined by the formula:

$$n_2 d = (2m - 1)\lambda/4 \quad m = 1, 2\ldots \tag{4.5}$$

For the air/h-GaN/Sapphire configuration, no light is absorbed in the blue sapphire substrate, $k = 0$, $N_3 = n_3 = 1.77$ [42]. The extinction coefficient k of GaN is 0.001 at 650 nm [43]. According to structure shown in Fig. 4.4, the influence of

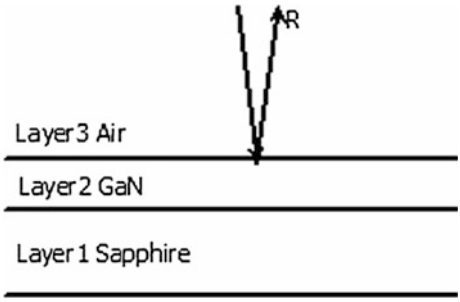


Fig. 4.4 Calculation mode of in situ monitoring reflectivity

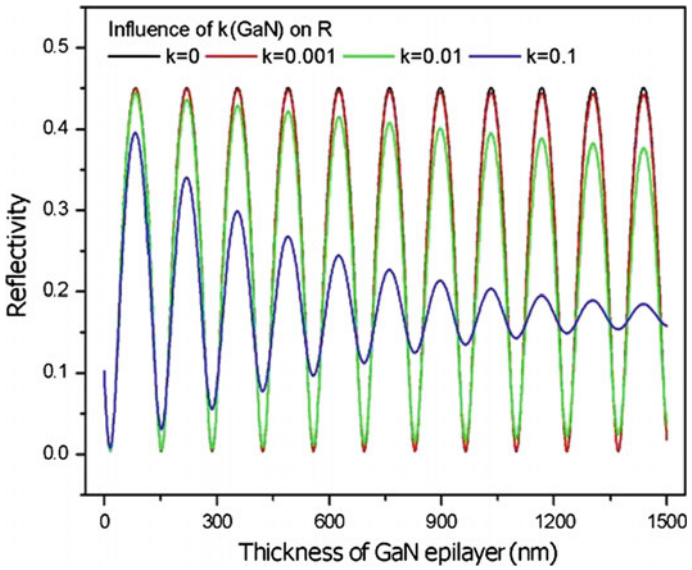


Fig. 4.5 Influence of k(GaN) to reflectivity

extinction coefficient k of GaN on the variation of the reflectivity during GaN growth is calculated with a GaN thickness of 1.5 μm. It is found from Fig. 4.5 that the attenuation is negligible for the reflected light caused by the extinction coefficient variation with film thickness. Therefore, the refractive index of GaN is $N_2 = n_2 = 2.4$. It should be noticed that the contribution of the thin low-temperature GaN buffer layer (usually $\sim$25 nm) to the reflectivity has been ignored in the calculation.

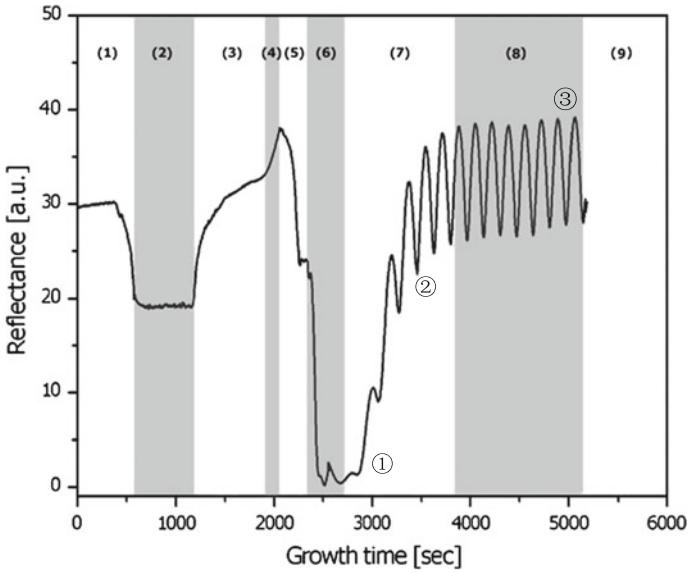


Fig. 4.6 In situ monitoring curves for the epitaxial growth of intrinsic GaN

3. Analysis of in situ monitoring curves in MOCVD

A representative in situ monitoring curve for the whole process of intrinsic GaN epitaxial growth is shown in Fig. 4.6 after parameter optimization, and it can be obviously divided into nine stages in accordance with the growing process [37, 44]

(1) The temperature of sapphire substrate rises after a heat treatment at a low temperature (300 °C).
(2) Heat treatment of the sapphire substrate at a high temperature in hydrogen atmosphere.
(3) The temperature of the reaction chamber is adjusted to 530–550 °C and the low-temperature buffer layer will be grown when the temperature stabilizes.
(4) The growth of the low-temperature GaN buffer layer. As the refractive index of GaN is higher than the sapphire substrate, the refractive index of the buffer layer will increase after the growth of GaN. As the thickness of the GaN increases, the reflected signal will be enhanced gradually.
(5) Close TMGa and the temperature of the reaction chamber will increase for the preparation of high-temperature GaN growth. After a stable temperature period of about 90 s, the growth will begin. Because of the poor crystal quality of the buffer layer deposited at a low temperature, during the temperature annealing process the surface will become rough due to the recrystallization of the buffer layer. Meanwhile, the buffer layer will become thinner because the decomposition of GaN at temperatures above 830 °C. So the reflectivity in the monitoring curve decreases quickly.

(6) Using the optimized conditions, the high-temperature GaN growth begins in a three-dimensional island model, that is, a nucleating island grows in both horizontal and vertical directions at the same time. Due to the growth of the nucleating island, surface roughening increases, and the reflectivity decreases, even below the reflectivity of the sapphire substrate.
(7) Owing to the gradual merge of nucleating islands formed at the beginning of the high-temperature process, the surface tends to become flat which results in the restoration of surface reflectivity.
(8) Smooth surface of the epitaxial layer is formed as a result of the ending of merge of the nucleating islands. The high-temperature GaN growth proceeds in a two-dimensional mode, and the interference fringe appears with a surface reflectivity oscillation in a stable amplitude and cycle.
(9) Reduce the temperature of the reaction chamber and take out the samples.

Much information important to the material growth can be drawn from the LR in situ monitoring curves, such as the thickness of the GaN epitaxial layer, growth rate, the optical constants (n, k), the roughness of the surface, etc. In the following discussion, the methods of extracting useful information from the monitoring curves will be discussed.

According to Eqs. (4.4) and (4.5), thickness interference fringes in the monitoring curve contain the information of the epitaxial layer thickness therefore the epitaxial layer thickness as a function of growth rate can be obtained from the in situ monitoring data. The cyclic D of the thickness interference fringe in the monitoring curve is

$$D = \frac{\lambda \cos(\theta_{\text{incidence}})}{2n} \approx \frac{\lambda}{2n} \tag{4.6}$$

where θ is the incident angle of the laser, here $\cos(\theta_{\text{incidence}} = \cos(1^\circ) \approx 1)$; λ is the wavelength of the incident laser light, 650 nm, n is the refractive index of the growth temperature, 2.12. It can be seen from Eq. (4.6), that in the experimental configuration of near-vertical incident light, the thickness cycle of the interference oscillation stripe in the monitoring curve only relates to the wavelength of the incident laser light and the refractive index of the epitaxial layer. The thickness of the epitaxial layer can be obtained after the determination of the thickness of GaN responding to interference stripe oscillation period multiplied by the cycles of the oscillation.

$$d = \frac{\lambda}{2n} m = \frac{650\ (\text{nm})}{2 \times 2.12} \times m = 153.3 \times m\ (\text{nm}) \tag{4.7}$$

Figure 4.7 shows an in situ monitoring curve of the GaN growth process. The shadow portion of the diagram corresponds to the high-temperature growth of GaN. If the parameters of intrinsic GaN are changing during the growth process, that is, the time for growing one cycle thickness is changing. The numbers of the thickness cycle can be determined quickly by counting the cycle of the oscillation in the

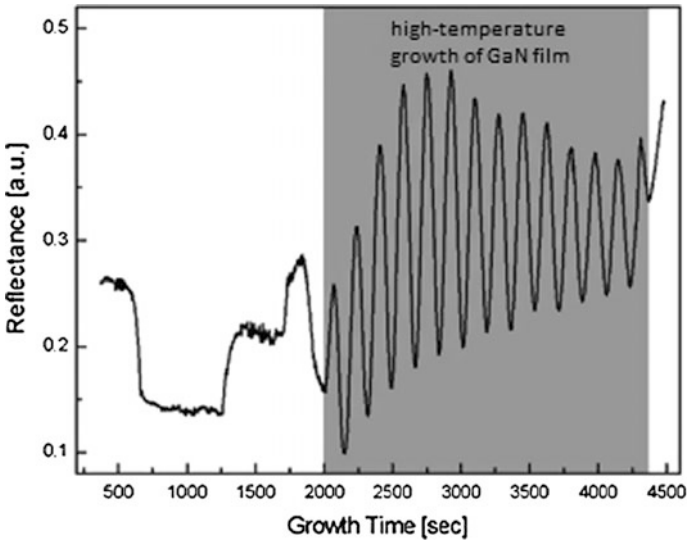


Fig. 4.7 The epitaxial layer thickness determined by the monitoring curve

monitoring curve. If the parameters of the growth process are constant and the growth rate is constant, the numbers of the thickness cyclic can be determined quickly by the growth time. Combined with Fig. 4.4, the air/GaN/Sapphire reflectance curve of an ideal model, it can be concluded that the oscillation cycle of the high-temperature growing process is 14. Therefore, it can be calculated from Eq. (4.7) that the epitaxial layer thickness $d = 2.146$ μm. The average growth rate of the epitaxial layer can be calculated according to $V = d/T$.

The surface roughness can also be obtained from the monitoring curve. In groups II–VI material growth processes, the roughness factor of the epitaxial layer has been taken into account. Stafford et al. has carried out quantitative analysis on the surface roughness factor in the growth of CdTe on the sapphire substrate [43]. This work also compared the repeatability of the roughness factor in different experiments with the material quality. As to MOCVD growth GaN, Stafford used a semi-quantitative method, such as introducing the coarse factor $F_{rs} = R_t/R_m - 1$, where R_m is the reflectivity of the inflection point (peak/valley) of the monitoring curve, R_t is the corresponding theoretical reflection. It is found that the change rule of F_{rs} is close to the annealing process before high-temperature growth and the flow of ammonia. As research on such quantitative analysis is few, there is no common analytical agreement yet.

The wavelength of the monitoring light source is 650 nm in this experiment. And GaN is transparent to the incident laser illumination (k is small). Light interference in the GaN layer forms monitoring system oscillation curves which contain information of film thickness and optical constants. When the epitaxial layer surface

is smooth; the monitoring curve should have constant amplitude. So the observed attenuation of the oscillation in the monitoring curve is due to the uneven surface of the epitaxial layer.

When light is incident upon a surface with root-mean-square roughness of σ, the reflectivity is [45]:

$$R_\sigma = R_0 \exp\left[-(4\pi\sigma)^2 \Big/ \lambda^2\right] \tag{4.8}$$

where R_σ is the practical reflectivity of the surface, R_0 is the reflectivity of the surface when it is flat, σ is the roughness of the sample surface, and λ is the wavelength of the incident light ($\lambda \gg \sigma$).

In the following discussion, a typical GaN growth process is used to illustrate obtaining the surface roughness characteristics from an in situ monitoring curve.

(1) It can be seen from Eq. (4.8) that σ is related to $R(d)$. So the horizontal axis in the monitoring curve should change from time to thickness, which is $R(t) \rightarrow R(d)$

The original in situ monitoring curve is shown in Fig. 4.8a. The starting point of high-temperature growth in the monitoring curve is t_g = 2114 s, and a conversion of $t \leftarrow (t - t_g)$ will be needed for the abscissa. From Eq. (4.6), the growth rate of the epitaxial layer is $v = \lambda/2nT$ = 0.892 nm/s. Then another conversion of $d \leftarrow t \cdot v$ can transfer $R(t)$ to $R(d)$ in the monitoring curve.

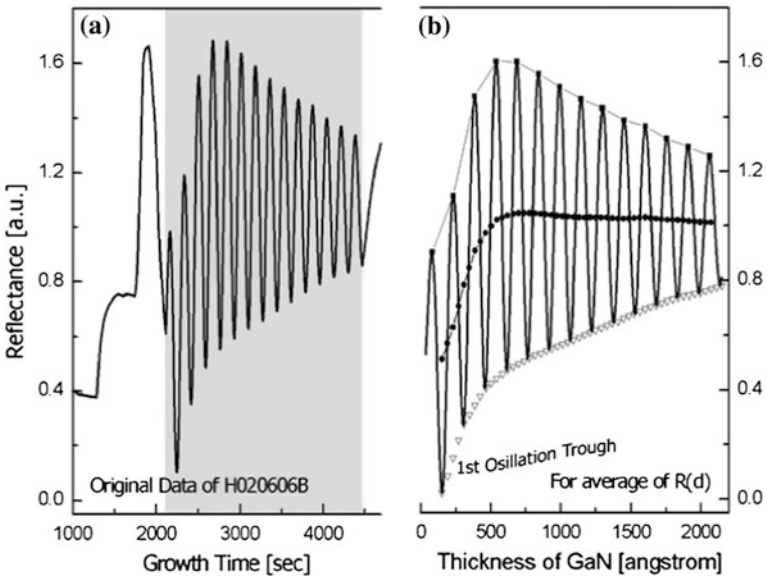


Fig. 4.8 Fit of the monitoring curve **a** original data, **b** the calculation of average reflectivity

(2) Calculation of the average reflected intensity

After the calculation of $R(d)$, position the first oscillation minimum of the curve at $\lambda/2n$ = 153.3 nm, which is the theoretical first minimum point of reflectivity. Calculate the average of $R(d)$ at this time using relevant tools and the average intensity of reflection curve, as shown in Fig. 4.8b.

(3) Normalize $R(d)$ according to its average reflection intensity

Divide $R(d)$ by average reflected intensity and the curve $R(d)\Big/R_{(\mathrm{d})}^{\mathrm{mean}}$ can be obtained.

From Eq. (4.1):

$$R = \frac{(n_0 - n_2)^2 \cos^2(2\pi n_1 d/\lambda) + (n_0 n_2/n_1 - n_1)^2 \sin^2(2\pi n_1 d/\lambda)}{(n_0 + n_2)^2 \cos^2(2\pi n_1 d/\lambda) + (n_0 n_2/n_1 + n_1)^2 \sin^2(2\pi n_1 d/\lambda)} \tag{4.9}$$

while $n_{0,}$ $n_{1,}$ n_2 are refractive index for air, GaN, sapphire respectively, d is the layer thickness of GaN. To fit $R(d)\Big/R_{(\mathrm{d})}^{\mathrm{mean}}$, the following function can be got from Eq. (4.9):

$$R_{\mathrm{sim}} \leftarrow \left[\left(R(d) - R_{(\mathrm{d})}^{\mathrm{mean}}\right)\exp(-d/\alpha) + R_{(\mathrm{d})}^{\mathrm{mean}}\right] \cdot \beta \tag{4.10}$$

where $R(d)$ is a function of n_1, n_2; α is the attenuation factor for the oscillation curve; β is the product factor, and $R_{(\mathrm{d})}^{\mathrm{mean}} \cdot \beta$ is the average reflectivity in theoretical reflectance curve.

The fitting result of R_{sim} is shown in Fig. 4.9, and the fitting results are in reasonable agreement with the experimental curve, n_1 = 2.1237, n_2 = 1.41, α = 1690, β = 6.196. The fitting result of refracting index for high-temperature GaN matches closely to the literature reported in 2.12 [43], which also illustrates the feasibility and accuracy of the experiment data processing method.

(4) Curve fitting to $\sigma(d)$

The theoretical reflected intensity curve can be calculated with the fitting results R_{sim} of step (3). And according to high-temperature fitting result n_2, the value of $R_{\mathrm{HT}}^{\mathrm{Sapp}}$ is $R_{\mathrm{HT}}^{\mathrm{Sapp}} = (1 - n_2)^2/(1 + n_2)^2 = 0.0289$. The reflectivity of sapphire at high-temperature in the original monitoring curve can be calibrated using this $R_{\mathrm{HT}}^{\mathrm{Sapp}}$, and then the reflectivity of the actual sample $R_\sigma(d)$ with surface roughness information can be calculated.

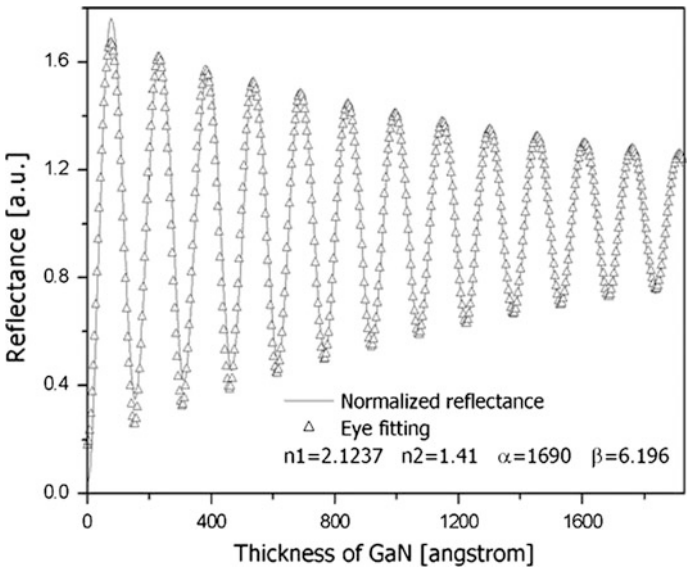


Fig. 4.9 Fit result of curve $R(d)/R^{\text{mean}}_{(\text{d})}$

According to Eq. (4.8):

$$\sigma(d) = \frac{\lambda}{4\pi}\sqrt{\ln\frac{Ro(d)}{R\sigma(d)}} \tag{4.11}$$

Just as Fig. 4.10 shows, the changing information of surface roughness can be obtained from Eq. (4.11). As can be seen from the changing of $\sigma(d)$, when the temperature begins to increase, the surface has a coarsening process, but this stage is very short. Then the GaN surface roughness falls sharply, this process corresponds to the merging of islands during epitaxial growth. When the thickness of GaN is about 80 nm, the surface roughness decreases slowly and steadily, this means the end of the island merging and the beginning of the two-dimensional growth model.

In order to analyze the high-temperature growth phase, three samples were grown with different thickness of the GaN epitaxial layer: 128, 500, 2000 nm. Sample surfaces were observed by a NanoScopeIIIa type atomic force microscope (AFM). The corresponding surface RMS roughness were 49.1, 44.1, 18.7 nm. In Fig. 4.10, the surface RMS roughness tested by AFM and the surface roughness fitted by in situ monitoring curve are compared and the two groups of data matches well. Therefore the accuracy of the analysis processing method on the monitoring curve is further verified.

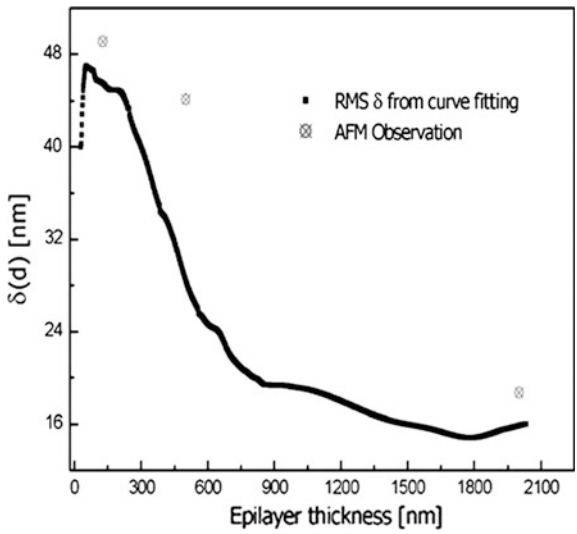


Fig. 4.10 Surface roughness changing course during GaN epitaxial growth quantitatively processed by monitoring curve

4.3 MOCVD Epitaxial Growth Technique of AlGaN Material

Because of the difficulty of growing bulk single-crystal GaN and the lack of suitable substrate material, GaN epitaxial film is generally grown on a large mismatched substrate. But it is difficult to realize hetero-epitaxy. Early GaN growth process showed that it was difficult to get a smooth surface without film cracks. Additionally, the film usually has a high background concentration. In 1986, Amano et al. first improved film quality by using a two-step GaN growth technique which first grows an AlN buffer layer in a low temperature and then grows GaN in a high temperature [20]. Further study found that when the growth of GaN was directly on (0001) face sapphire, the growth mechanism was typically three dimensional, which experienced the process of isolated islands, three-dimensional growth and the formation of an uneven surface. But the growth with a low-temperature buffer layer is similar to nuclear growth forms which contain the process of isolated islands, lateral growth, island merger and quasi-two-dimensional growth. Nakumura reported a two-step growth method of firstly growing GaN buffer layer in low temperature and then growing GaN in high temperature. Nakumura also found that the optimal low-temperature GaN buffer layer thickness is about 200–250 Å.

Because of the invention and application of the buffer layer and two-step technology, breakthroughs in GaN epitaxial technique were achieved. GaN devices such as blue light-emitting diode, lasers, and detectors were prepared, and some devices have been commercialized. This section first introduces the basic process of the two-step method, subsequently the GaN and AlGaN epitaxial technique on a GaN buffer layer will be introduced; then growth technology and mechanisms for

GaN, AlGaN, AlN epitaxial layers on an AlN buffer layer are introduced. Considering that the growth mechanism between GaN and AlGaN materials have both uniqueness and commonality, GaN epitaxial technology is first introduced, then AlGaN epitaxial technology is introduced. Finally, P-type doping techniques of GaN material are introduced. MOCVD material growth in this section is always on substrates of (0001) face sapphire. Ammonia (NH_3), trimethyl aluminum (TMAl), trimethyl gallium (TMGa) is respectively, sources of N, Al, Ga, and hydrogen (H_2) is a carrier gas.

4.3.1 AlGaN Epitaxial Technology on GaN Buffer Layer

The invention of the buffer layer and two-step technology acted as a milestone in the development history of GaN-based materials. An understanding of the growth mechanism of the epitaxial layer on GaN enables the study of the growth mechanism of an epitaxial layer on an AlN buffer layer. This, in turn, enables the understanding of the fabrication of back-lighted solar-blind AlGaN detectors which must use an AlN buffer layer. In the following discussion, the two-step growth technology is first introduced, then GaN and AlGaN epitaxial technology on low-temperature GaN buffer layer is introduced.

1. Two-step growth technology

Due to the large lattice and thermal mismatch between GaN and sapphire being, if high-temperature GaN is directly grown on a sapphire substrate, uneven hexagon pyramid-shaped patterns will appear with a rough surface and poor crystal quality. In order to prepare smooth and high-quality GaN material on sapphire substrates, a special process must be used. The two-step MOCVD material growth method is a typical process for GaN material growth. Figure 4.11 gives the flow chart of time for a typical two-step process for GaN MOCVD growth.

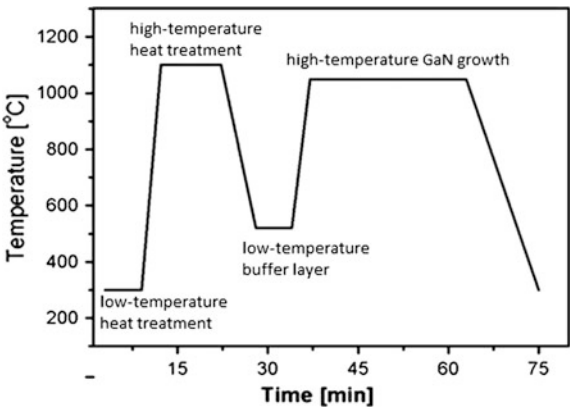


Fig. 4.11 Flowchart of time for MOCVD growth GaN

The basic process for two-step method is described in [20, 21, 45] and usually adopts a low-temperature deposited AlN or GaN buffer layer. Then the process finishes the growth of GaN on the buffer layer at a high temperature of 1000 °C. By this method, the dislocation density is greatly reduced and the quality of the epitaxial layer is improved, however the dislocation density of the GaN epitaxial layer may remain high, usually with an average dislocation density of 10^{8-10} cm^{-2} [46].

Akasaki et al. [47] first put forward a growth model for GaN: the low-temperature deposition of an AlN buffer layer, and nucleation and crystallization of GaN grown at high-temperature, then the merging of nucleation islands by lateral growth, finally the growth of the high-temperature GaN epitaxial layer. Hiramatsu et al. [48] observed the epitaxial layer interface and the crystallization of nucleation and pillar structure of high-temperature GaN at the early stage of high-temperature growth with transmission electron microscopy (TEM). The orientation of the crystal column is slightly different between the two structures, and dislocation is produced between the crystal boundaries. There are two situations of producing the non-conforming orientation of crystal columns: first, the column crystal growth direction tilts (Tilt), deviated from the *c* axis, which is perpendicular to the direction of the substrate surface (0001); second, the crystal column in the *c* axis torsion (Twist) [49]. Tilt will introduce screw dislocation parallel to the direction of growth, and twist will introduce edge dislocation parallel to the growth direction. Most dislocations propagating along the growth direction are pure edge dislocation, Burgers vector is 1/3 ⟨11-20⟩, and these dislocations will form low angle grain boundary in the epitaxial layer. Another cause of the dislocation in the epitaxial layer is thermal stress caused by the different thermal expansion coefficient between GaN and the sapphire substrate. For the growth of GaN on Si substrate, edge dislocation was observed during the cooling process after growth. The thermal stress of GaN at room temperature can be as high as 10^9 dyn/cm^2 [50]. So at the cooling process after growth, the thermal stress could promote further dislocation.

In the two-step growth process, the technique of creating the buffer layer at a low temperature is crucial to the crystal quality of the GaN high-temperature epitaxial layer. The optimization of the growth condition of the low-temperature buffer layer can effectively reduce the dislocation density in the epitaxial layer. The introduction of the buffer layer may play roles such as: (1) It provides a nucleation center with the same orientation as the substrate. (2) It releases the stress coming from the lattice mismatch between the substrate and GaN and the thermal stress caused by the thermal expansion coefficient mismatch between the substrate and GaN. (3) It provides a flat surface of nucleation for further growth, reduces the contact angle of the nuclear growth, makes the island growth of GaN crystal grain become a plane in smaller thickness and become a two-dimensional epitaxial growth. It is because of these effects that a GaN buffer layer significantly improves the material quality of GaN.

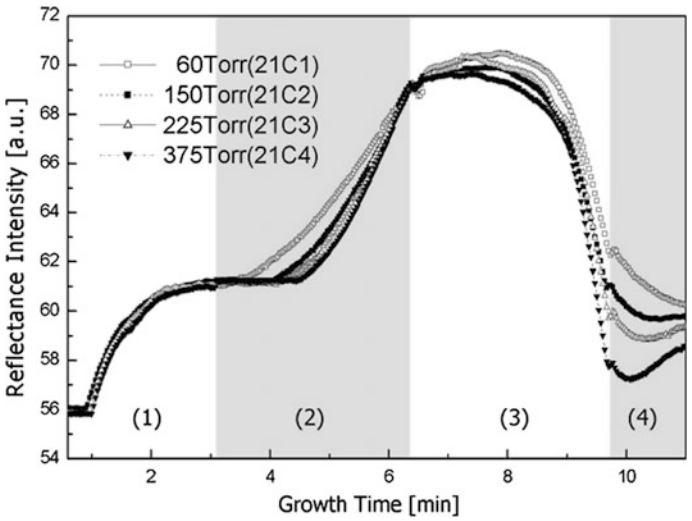


Fig. 4.12 Monitoring curve of buffer layer grown at different pressure

2. GaN epitaxial layer growth with a GaN buffer layer

In a standard two-step growth method, a low-temperature buffer layer is an important factor to improve the integrity of GaN epitaxial layer. In the following, the influence of changing the growth pressure of the low-temperature GaN buffer layer to the process of GaN MOCVD growth is introduced. In order to study the effect of the pressure in buffer layer growth stage, four buffer layer samples are prepared; 21C1, 21C2, 21C3, 21C4 with the reaction chamber pressure in the growth process being 60, 150, 225, 375 Torr [51, 52], respectively. Figure 4.12 shows the monitoring curves of 21C1, 21C2, 21C3, 21C4 in the growth process. The stages (1)–(4) in the figure correspond to process of high-temperature treatment and cooling of the substrate to the temperature of the buffer layer growth, buffer layer growth, the annealing process of rising to the epitaxial growth temperature, and finally the 90 s high-temperature annealing process.

As can be seen from the monitoring curve, in the process of high-temperature annealing of 90 s at 1100 °C, the reflectivity in the monitoring curve first continues to decrease, and then is gradually restored, and tends to be stable in the end. As shown in Fig. 4.12(4), when the pressure of the buffer layer growth is small, the decrease and recovery of the reflectivity is not obvious, and with the increase of growth pressure, reflectivity of the change process is more and more obvious.

Figure 4.13 shows the AFM images of sample 21C1 and 21C4's native buffer layer (as-grown buffer). It can be seen clearly from the table that the buffer layer of the island is even over the surface of the sapphire substrate. But there are obvious difference between the images of 21C1 and 21C4. The surface of 21C1 is relatively rough; the RMS roughness is 0.83 nm. But 21C4 has a relatively flat surface, and its RMS roughness is 0.67 nm. In addition, it can also be noted that the grain size of

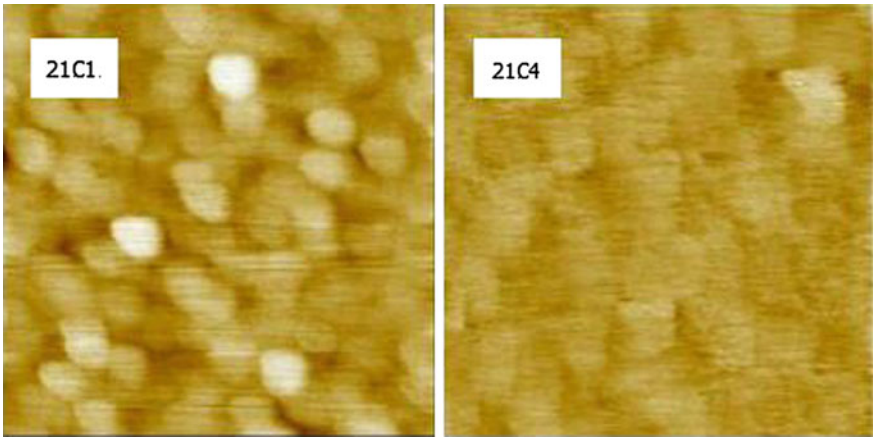


Fig. 4.13 The surface images of GaN as-grown buffer layer grown at different pressure analyzed by AFM (1 μm × 1 μm)

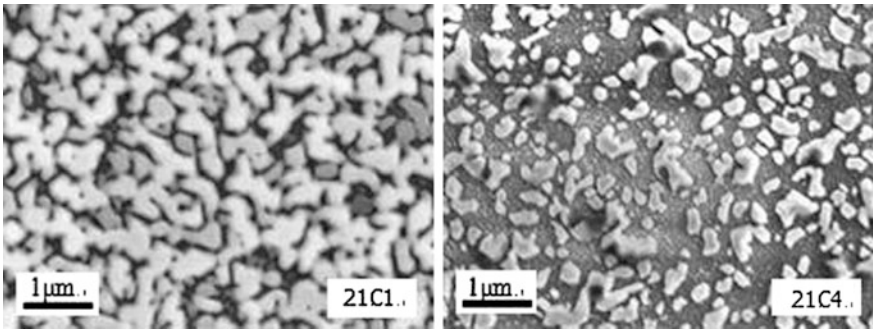


Fig. 4.14 The surface images of GaN buffer layer after annealing grown at different pressure

21C1 is less than the grain size of 21C4. That is, the grain size of as-grown buffer layer increases with the increase of the pressure of the reaction chamber. Then, scan electron microscopy (SEM) was used to analyze the buffer layer surface. Figure 4.14 is the surface morphology of sample 21C1 and 21C4 after annealing. In Fig. 4.14a, the islands formed after annealing are large volume and low density ($\sim$2.2 μm^{-2}), and the islands spacing is so small that many are connected together. But for 21C4, the volume of the islands formed after annealing is relatively small, the distance between the islands is relatively large, and the density of the islands is increased ($\sim$7.3 μm^{-2}).

It is generally believed that when the buffer layer is grown under low pressure, the nucleation selectivity of GaN on the buffer layer is poor, and the surface is full of GaN grains. The height of the grain will decrease because of the decomposition of GaN during the annealing process, which causes the low reflectivity in the monitoring curve. Due to the spacing of the island being very small, the lateral size

of the island is constant, and the decomposition of the buffer layer is layer by layer, as shown by the gradual reduction of the reflectivity in Fig. 4.12 stage (3). As the growth pressure increases, the nucleation selectivity is better and the grain spacing is larger. Due to the decomposition of GaN during high-temperature annealing, the height of the grain decreases and the reflectivity of the surface is reduced. As the lateral size of the islands decreases and spacing between islands increases, the area of the island and the increase of the platform area decrease. That is, when the pressure is high, there is a surface coarsening process during the annealing of the buffer layer. When the ratio of the area of the island to the area of the island platform is close to a certain value (for example 1:1), the surface roughness of the buffer layer is the largest, and the reflectivity is a minimum. With continued annealing, the lateral sizes of the islands continue to decrease, and the platform area increases, and the surface tends to level off with the reflectivity gradually recovering. So it is observed in Fig. 4.12 that when the pressure of buffer layer growth is high (21C4), the reflectivity in phase (4) first decreases and then recovers.

In order to study the influence of the buffer layer growth pressure to the quality of the GaN epitaxial layer, two samples marked as #LP and #HP are grown under the pressure of 60 and 375 Torr. The two samples have a low-temperature buffer layer of 25 nm and a 2 μm thick unintentional doping of GaN. Figure 4.15 is the monitoring curves of samples #LP and #HP. The vertical solid line shows the high-temperature starting point. In the early stage of high-temperature growth, the model is three dimensional. The nuclear island grows in both horizontal and vertical directions at the same time, and the volume becomes larger. When the neighboring islands begin to merge, the reflectivity recovers gradually and increases to the

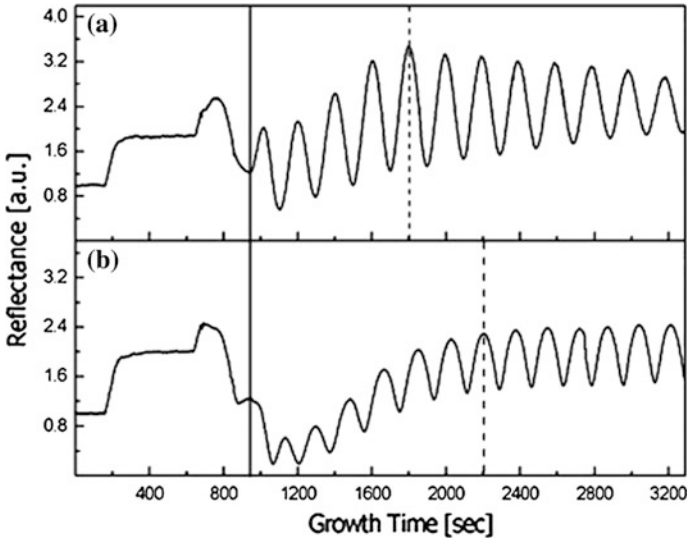


Fig. 4.15 Monitoring curves of samples with different buffer layer growth pressure. **a** #LP, **b** #HP

maximum. The corresponding stage in Fig. 4.15 is between the vertical solid line and dotted line. When the islands completely merge, GaN grows in a two-dimensional mode, and the reflectivity oscillates with stable amplitude, and forms a thickness interference fringe. And the corresponding stage in Fig. 4.15 is after the dotted line.

With a different growth pressure at the buffer layer growth stage, the low-temperature monitoring curve of the buffer layer growth stage is different, and high-temperature stage monitoring curve is obviously different. For #LP specifically, the buffer layer was grown under low pressure and the amplitude of the monitoring curve reaches a peak value after 14 min of high-temperature growth of GaN. That is to say, the GaN epitaxial growth process soon realizes the transformation of growth pattern from three dimensional to two dimensional. The monitoring curve begins to achieve maximum amplitude after 21 min high-temperature growth after the pressure of the buffer layer growth increases to 375 Torr. The transformation from three dimensional to two dimensional is rather slow. It can be seen from Fig. 4.15b that at the beginning of the high-temperature growth, the reflectance of #HP in the monitoring curve does not directly increase as for #LP, but first shows a sharp drop, then gradually restores to the saturated amplitude.

The difference between #LP and #HP monitoring curves at the high-temperature stage can be preliminarily explained by the SEM results of 21C1 and 21C4 in Fig. 4.14. The buffer layer grown under low pressure formed low density, large volume, and small spacing nucleation islands. After the start of the high-temperature growth, there is a rapid surface smoothing process, which quickly fills the gap between the islands. Due to the low-temperature buffer layer growth, the spacing between the islands is small, and this process quickly fills the gap between islands, forming a compact GaN epitaxial layer, and achieves a two-dimensional growth. So, at the early stage of high-temperature growth in Fig. 4.15, the reflectivity in the monitoring curve rises immediately and soon reaches the maximum value, realizing the change of three-dimensional growth to two dimensional. Grains with small size, large spacing and high density will be formed on the surface after annealing, after the pressure for the buffer layer growth increases to 375 Torr. The surface filling process after the high-temperature growth makes the nucleation islands grow in both horizontal and vertical directions, and the size of the islands increases in both horizontal and vertical directions. An increase of the lateral size of the island makes the area of the islands increase, due to the small size of the islands; the area of the islands is far less than the area of the mesa between the islands. Thus the surface roughness of the buffer layer increases, and the reflectivity of the surface decreases quickly, and becomes even less than the buffer layer. The lateral growth of the islands makes the volume of the islands increase further. The roughness is the largest and the reflectivity reaches the minimum value when the area of the islands and area of the mesas between the islands are equal. This shows when the pressure at the early stage of high-temperature growth is large, there is a spontaneous coarsening process on the buffer layer surface. As the lateral size of the islands increases, the surface begins to level off, and the reflectivity increases. The nearby islands begin to merge, and the gap

Table 4.1 Half high width (FWHM) of the X-ray rocking curve for #LP and #HP (arcsec)

Sample number	(0002) face	(10 – 12) face
#LP	482	986
#HP	360	820

between the islands levels off quickly. So a dense and smooth GaN epitaxial layer forms and reflectivity quickly recovers, corresponding to the curve of vertical dotted line in Fig. 4.15b. The process of coarsening first and then recovery at the high-temperature growth stage explains the phenomenon that #HP needs more time to recover after the changing from 3 d to 2 d growth than #LP.

#LP and the (0002) and (10-12) faces of #HP were measured by double-crystal X-ray diffraction (DCXRD). As shown in Table 4.1, when the buffer layer growth pressure increases from 60 to 375 Torr, half high width (FWHM) of the X-ray rocking curve for (0002) and (10-12) faces lowers significantly. The full width at half maximum (FWHM) of the X-ray rocking curve for (0002) and (10-12) faces indirectly reflects the screw dislocation and edge dislocation density of the epitaxial layer [53–55]. So it can be determined that the GaN epitaxial layer grown at higher chamber pressure will have the better structure performance. Because the lattice mismatch between GaN and the sapphire substrate is large; many penetrating dislocations will be generated at the interface. At the stage of lateral growth, the vertical penetrating dislocations distort to the horizontal direction, which reduces the density of the dislocations penetrating to the epitaxial layer. Buffer layers grown under low pressure will result in grains with high volume and small spacing. The lateral growth of the grain will make the grain merge quickly, and a large number of dislocations continue to extend to the epitaxial layer by the normal direction. As the pressure of buffer layer growth increases, the grain spacing is large; and horizontal grain growth prolongs the time needed for the grains to merge. A large number of dislocations reverse, which reduces the dislocations extending to GaN epitaxial layer. Thus the FWHM of the X-ray rocking curve for (0002) and (10-12) faces narrows and the crystal quality of the epitaxial layer improves.

The photoluminescence spectrum of sample #LP and #HP were measured at room temperature in the experiments. The photoluminescence spectrum has the merits of simple sample preparation, high sensitivity, and nondestructive, and is widely used in the research of the optical properties of semiconductor materials. The excitation light source used in the PL test is He–Cd laser with 20 mW power and a 325 nm center wavelength. PL measurements results are consistent with the results of X-ray measurements. The PL spectrum of #LP and #HP at room temperature is given in Fig. 4.16. There is a strong edge peak at the wavelength of 362 nm (3.43 eV). #HP not only has a significantly higher edge peak luminous intensity than #LP, but also has a less luminous intensity at deep level. X-ray and PL spectrum measurement results show that, with the increase of buffer layer growth pressure, GaN crystal quality was improved at the same time, as well as the optical quality.

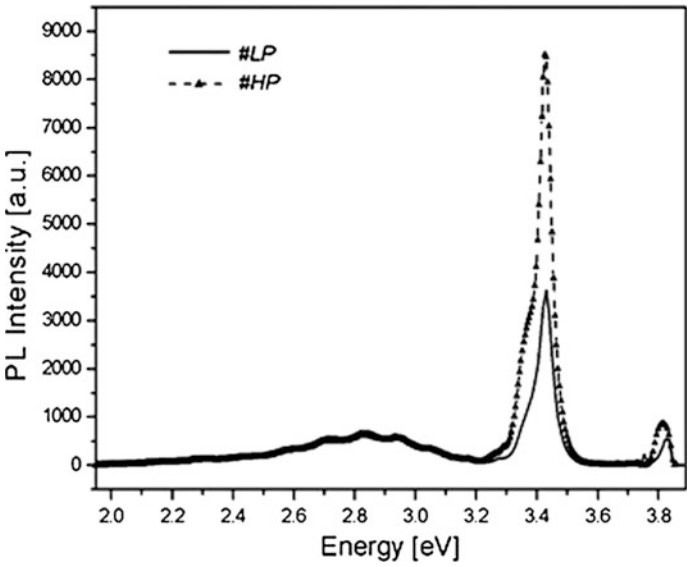


Fig. 4.16 PL spectrum measurement results of #LP and #HP at room temperature

Pressure has a great influence on the GaN epitaxial layer growth mode. To understand the influence of chamber pressure on the subsequent growth of high-temperature GaN, low-temperature buffer layers with thickness of 25 nm were respectively deposited under 60 and 375 Torr, and both of the buffer layers were annealed at high temperature and for 90 s at 1100 °C. These two samples were used as templates, with GaN epitaxial layers of different thickness obtained by changing the high-temperature growth time. The conditions of high-temperature growth are the same with sample #LP and #HP in this chapter [56]. When the pressure of buffer layer growth is 60 Torr, the high-temperature GaN deposition time are 0, 2.7, 10.4, 39 min, respectively, and the corresponding thickness are 0, 128, 500, and 2000 nm (low-temperature series). Low pressure series was marked as #LP1–#LP4. The high-temperature growth time for high pressure series are 0, 4, 15, 22.5, and 38 min, and the corresponding thickness are 210, 780, 1170, and 2000 nm. High pressure series was marked as #HP1–#HP5.

Figure 4.17 shows the corresponding monitoring curves for the high and low pressure series samples. The monitoring curves are normalized by the reflectance of the sapphire buffer layer after high-temperature heat treatment. Samples #LP1–4 and #HP1–5 are marked in the Figure for each stage. Some research teams have also observed similar transition delay of growth mode from 2D to 3D in high-temperature growth stage [57–59]. It is found that the surface topography of GaN before high-temperature growth is strongly influenced by the carrier gas flow rate for the low-temperature buffer layer and the pressure of the chamber at the high-temperature stage. And it is also observed that the decline in reflectivity at the early stage of high-temperature growth is due to the formation of sparse and large

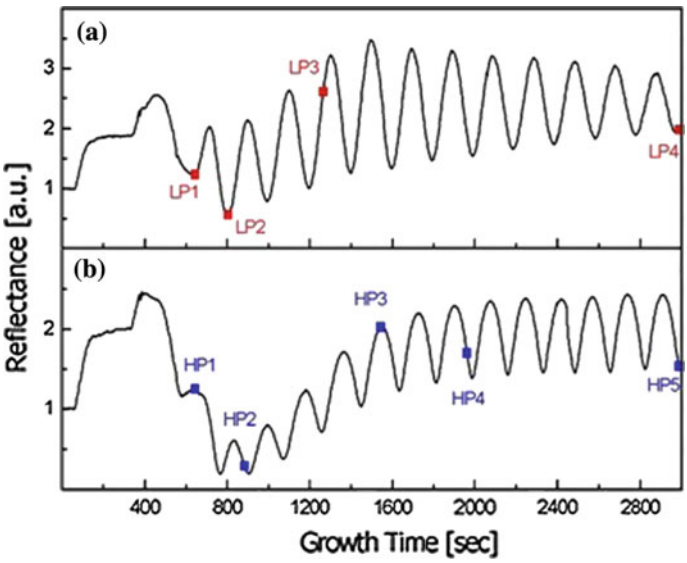


Fig. 4.17 Different thickness samples of high and low pressure series in the corresponding stage of the monitoring curve. **a** LP series, **b** HP series

grain size of nucleation center on the buffer layer surface. The only growth parameter change here is the pressure of the buffer layer growth. So the difference in reflectivity changes at the early stage of high-temperature growth may be caused by the different surface condition of the buffer layer before high-temperature growth.

Figure 4.18 shows the scanning electron microscopy (SEM) of high-temperature GaN at different growth stages with buffer layers grown at low pressure. In the annealing process from buffer layer growth temperature to high-temperature GaN growth temperature, decomposition, and recrystallization of the buffer layer make the surface coarser. The island volume on the surface of #LP1 is large, and the top of the island is flat with (0001) face. It is similar to a pyramid without a top. As shown in Fig. 4.18a, the nuclear island densely distributed on the surface of the substrate, and the spacing between the islands is small. After the start of high-temperature growth, the islands will quickly merge. With the increase of the reaction chamber temperature, migration rate of the atoms participating in the reaction will be very high and the small gaps between the islands will soon be filled, and a flat top with (0001) face will be formed. When the thickness of high-temperature GaN gets to 500 nm, as shown in Fig. 4.18c, most nuclear islands have merged, although there are still many pits on the surface. These growing hexagonal pits correspond to the unincorporated areas between the merging islands. The sides of the pits correspond to the {10-11} sides of pyramid nuclear islands, which are formed in the process of lateral propulsion. As the high-temperature growth of GaN continues, the islands merge completely, a flat (0001) surface forms,

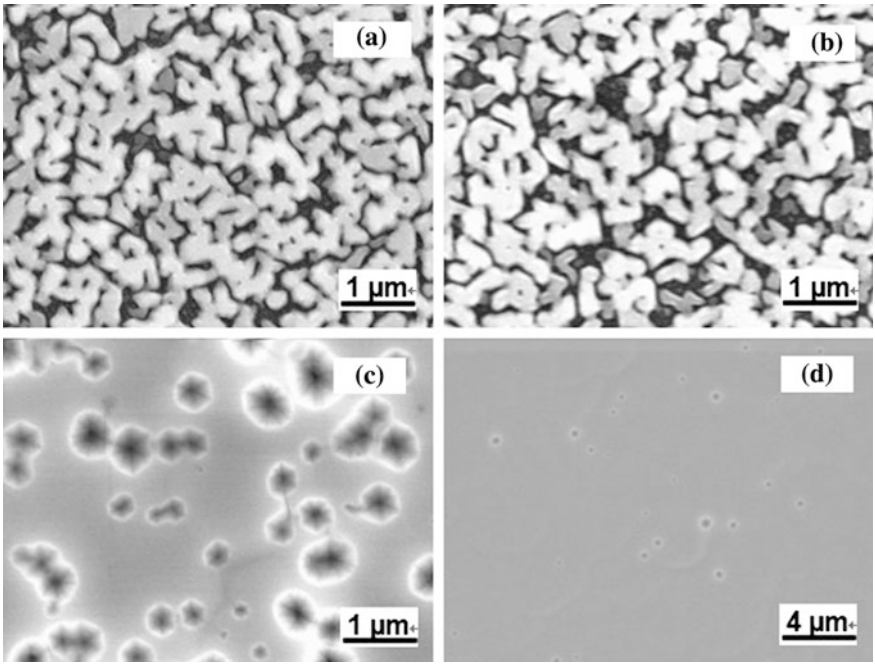


Fig. 4.18 SEM images of low pressure series samples: **a** #LP1, **b** #LP2, **c** #LP3, **d** #LP4

and the growth pattern transforms from two dimensional to three dimensional. As shown in Fig. 4.18d, when the thickness of GaN is 2000 nm, the surface is smooth and flat.

There is marked contrast between the surface evolution processes for high pressure and low pressure. Figure 4.19 is the surface evolution process of high-pressure samples (#HP1–#HP5). It is found from Figs. 4.18a and 4.19a that the nucleation density of #HP1 is higher than that of #LP1. The nucleation islands are distributed on the surface sparsely, and the spacing between the islands is large. These islands are mostly pyramidal in shape although a few of them are pyramidal with a flat top. The growth of high-temperature GaN first crystallizes on the single-crystal island. It can be seen from Fig. 4.19b that the area between islands may be a buffer layer after annealing or it is formed because of the nucleation of GaN at high-temperature growth. After the start of the growth, the growth rate of {10-11} side is fast, and the adjacent islands began to merge to form a larger island. Thus the island in Fig. 4.19b is bigger than those in Fig. 4.19a. In the process of lateral growth, the (0001) surface growth rate of the top is slow, so a flat (0001) surface can be observed in Fig. 4.19b. When the GaN epitaxial thickness is 780 nm, the nucleated islands almost cover the surface of the buffer layer, as shown in Fig. 4.19c. As for the integration of the islands, the surface coverage of the nuclear island becomes larger and larger, and the pit between the islands has been

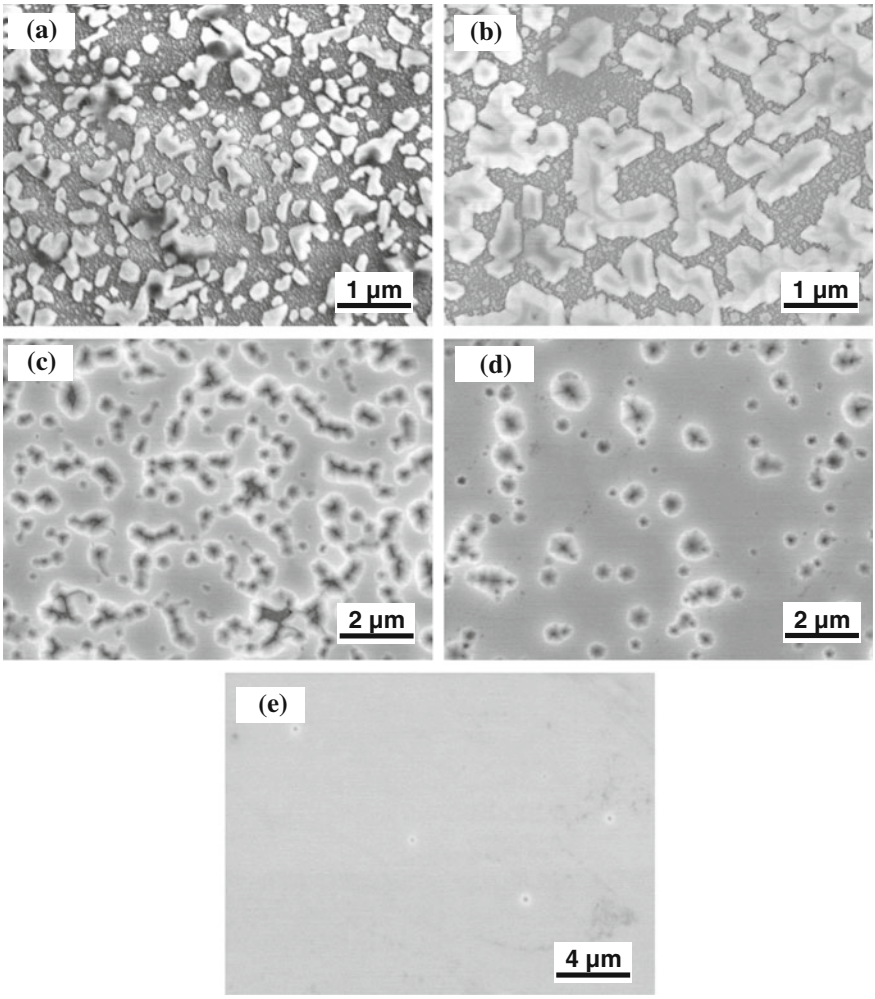


Fig. 4.19 SEM images of high pressure series samples **a** #HP1, **b** #HP2, **c** #HP3, **d** #HP4, **e** #HP5

reduced, as shown in Fig. 4.19d. When the thickness is 2000 nm, the islands have been fully merged.

A possible mechanism for reducing the dislocation density in the epitaxial layer was proposed according to the AFM and SEM images of the low- and high-pressure sample surface, as shown in Fig. 4.20. The vertical line in the Figure represents the edge dislocation and the oblique line represents the hybrid penetrating dislocation. In the actual physical process, the merger of the islands will form a new edge dislocation, most of which will penetrate the GaN epitaxial layer. As a result of the existence of mixed dislocations, when they meet with an edge dislocation, dislocation loops can possibly be formed, thus reducing the dislocation density

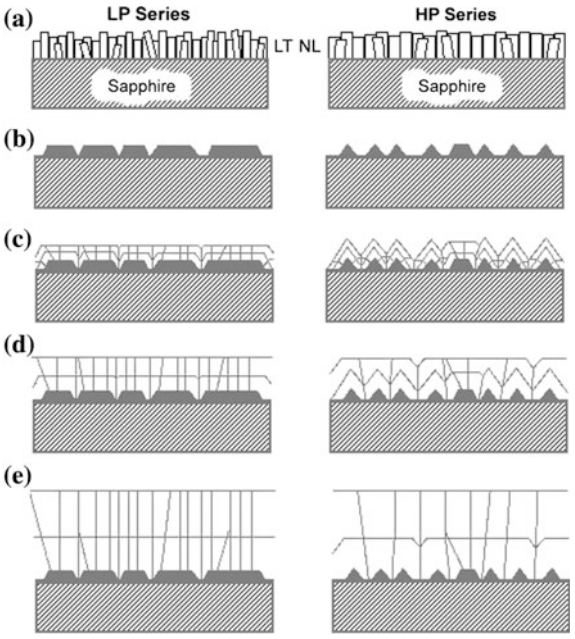


Fig. 4.20 Schematic of GaN growth process and the possible dislocation mechanism

penetration to the epitaxial layer, as shown in Fig. 4.20d, e. It is worth noting that the initial nucleation density in #HP series is much larger than that in #LP series samples. So a large number of new edge dislocations will be formed in the nuclear merging process, and the effects of reducing the dislocation density caused by the lateral growth in the early high-temperature process is inhibited. So a further optimization of the growth parameters is necessary to reduce the nucleation density in the annealed buffer layer.

3. AlGaN epitaxial technology with a buffer layer of GaN

AlGaN is an important part of GaN-based UV detectors and its material quality directly affects the device performance. Compared with Ga atoms, Al atoms show many different properties, such as: the pre-reaction of Al atom is more serious, and the surface diffusion length is very small, and so on [60–63]. Because of these special properties, there is significant difference between the AlGaN growth mechanism and the GaN growth mechanism. If an AlGaN epitaxial layer is directly grown on an AlN buffer layer or epitaxial layer, many difficulties will be encountered due to the relatively weak understanding of the growth mechanism. But if the growth of AlGaN material on a GaN buffer layer is studied, the AlGaN material growth mechanism will be understood and the differences of AlGaN material growth and GaN material growth will be evident. There is great significance for improving the AlGaN material quality and related device performance on an AlN buffer layer.

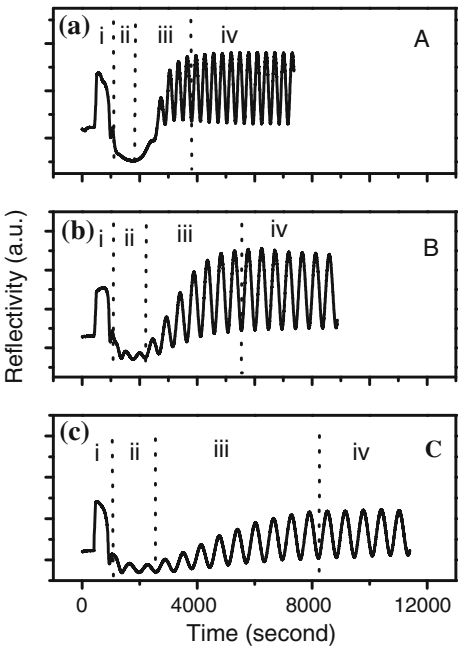


Fig. 4.21 The monitoring curves of sample A, B, C grown by MOCVD, while the TMAl flow are respectively 0, 80, 160 sccm for sample A, B, C

The following research is on GaN and AlGaN materials with a low-temperature GaN buffer layer using the MOCVD system [64]. Five samples are grown with TMAl flow of 0, 40, 80, 120, 80 sccm (ml/min) in this experiment respectively. Figures 4.21a–c are the monitoring curves of samples A, B, and C, respectively, while sample A is GaN material (with a TMAl flow 0), samples B and C are both AlGaN material (the TMAl flow is 80 sccm for B and 160 sccm for C during growth process). The growth of AlGaN can be divided into four stages [58, 65]: (i) the growth of a low-temperature GaN buffer layer and high-temperature annealing, (ii) the initial stage of high-temperature growth of epitaxial layer, (iii) the merging of the islands and lateral growth, (iv) quasi-two-dimensional growth of epitaxial layer. It can be seen that there are great differences among the monitoring curves of the three samples: (1) for sample A, in the initial epitaxial layer growth of stage (ii), there are no oscillation interference fringes in the curves. This indicates that the surface is very rough. But the reflectivity of initial growth of sample B was increased and weak interference fringes appear. With a further increase of TMAl flow of sample C, the reflectivity of the initial growth is stronger, and the interference fringes are more obvious. (2) In the following growth process of stage (iii), stable oscillation interference fringes formed quickly, while in sample B they can be achieved only after a long time. Sample C needed a longer time to realize stable oscillation interference fringe. (3) For the GaN monitoring curve, the spacing of the oscillation interference fringes is small with a short cycle. While spacing of the interference fringe of oscillation is larger with a long cycle for AlGaN. With the further increase of TMAl flow, the spacing of the oscillation interference fringe is

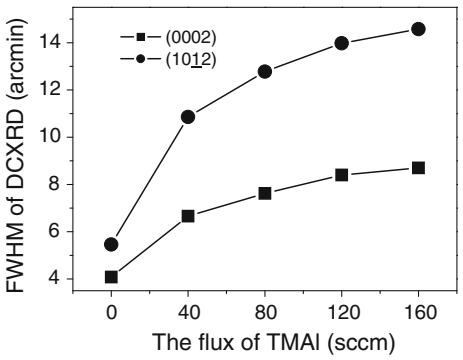


Fig. 4.22 The relationship between the (00002) and (10-12) plane half high width of the DCXRD rocking curve and the change of the TMAl flow

wider than in sample B and the cycle is also longer. The obvious difference between the monitoring curves of GaN and AlGaN also means the growth mechanisms of the two materials are quite different.

Material quality of GaN and AlGaN are also characterized by X-ray diffraction technology [66]. Figure 4.22 shows the relationship between the (00002) and (10-12) plane half high width of the DCXRD rocking curve and the change of the TMAl flow. It shows that the half high width of the DCXRD rocking curve of plane (0002) and (10-12) reflects a screw dislocation and edge dislocation information [54, 55]. When the half high width is wider, the dislocation density is higher, and the sample quality is poor. As can be seen from Fig. 4.22, with the increase of TMAl flow, the half high width of both the (0002) and (10-12) plane increases, which means the density of the screw dislocation and edge dislocation is increased and the quality of the sample becomes poor and the quality of samples A, B, and C is also worse. Generally speaking, the material growth process of GaN is a transformation of three-dimensional to two-dimensional growths. In the early growth, the surface is rough, and the growth process is three-dimensional island model. As the epitaxial process continues, the islands begin to merge and the dislocations begin to steer. And the quality of GaN becomes better [58, 67]. The monitoring curves in Fig. 4.21a clearly show the growth process. In the early growth stages, the reflectivity is low and the surface is rough, and there are no interference fringes in the curve. As the islands began to merge, oscillation interference fringes gradually formed, and the growth process changes from three dimensional to two dimensional. The quality of GaN grown in this manner is better. But as shown in Fig. 4.21b, in the early growth stage, there are interference fringes in the monitoring curve, and the reflectivity is relatively high. This indicates that the surface is smooth and there is small amount of island growth. In subsequent growth stages, many of the dislocations penetrating to the epitaxial layer cannot achieve steering so the quality of sample B is poor. With the increase in TMAl flow, as shown in Fig. 4.21c, the surface is smoother and the early interference fringes are also more obvious. This makes the transition from 3D to 2D growth difficult, and sample quality is poor. Compared to GaN growth, AlGaN may nucleate between islands easily, and it is not easy to realize island growth at the beginning of the growth.

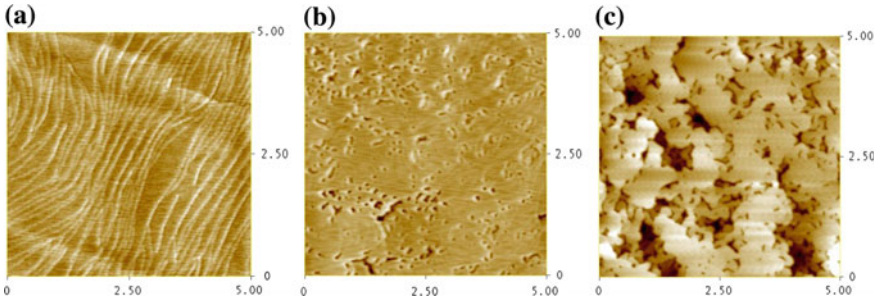


Fig. 4.23 Morphology of sample A, B, C (5 μm × 5 μm), the surface roughness is 0.514, 1.385, 5.505 nm

It can also be seen from Fig. 4.21 that there is a great difference among the intensity of the oscillation interference fringes. This means the surface morphology is different. Figure 4.23a–c is the surface morphology of sample A (GaN), sample B (AlGaN), and sample C (AlGaN). There is obviously a great difference between the surface morphologies. In the morphology of GaN, there are obvious step streams, and the surface is very smooth with a roughness of only 0.514 nm. This means that there is an obvious two-dimensional growth. In sample B (AlGaN), the surface is rough with a roughness of 1.385 nm. And there are many holes in the surface, which means many islands have not been merged. With a further increase of TMAl flow, sample C (AlGaN) has an even rougher surface, and the roughness is 5.505 nm and the size of the holes is larger. Combined with Fig. 4.21, it can be seen that GaN can soon realize the two-dimensional growth. But the AlGaN sample needs a long time to achieve two-dimensional growth. As the TMAl flow increases, it takes even longer for sample C to achieve two-dimensional growth. Obviously, as the TMAl flow increases, it is more difficult to realize two-dimensional growth and for the islands to merge. Because the bond energy of Al–N is higher than Ga–N, the surface diffusion length of Al is less than that of Ga. In the material growth process, it is difficult for Al atoms to achieve horizontal diffusion and so two-dimensional growth is difficult to be realized and the surface becomes rough. With the increase of TMAl flow, more and more Al atoms become involved in growth, so it is less likely to achieve two-dimensional growth. AlGaN samples have an even rougher surface.

It can be seen from the monitoring curves in Fig. 4.21 that the cycle of the interference fringes are obviously different. As the TMAl flow increases, the cycle time is longer. The cycle of the interference fringe reflects the change of the growth rate. The longer the cycle, the slower the growth rate will be. The growth rate of GaN and AlGaN are obtained by calculation. The solid line in Fig. 4.24 shows the relationship between growth rate and TMAl flow. There is an interesting phenomenon in this Figure, namely with the increase of TMAl flow, the growth rate of AlGaN decreases. In order to understand this phenomenon further, TMAl flow is reduced to 0 to study the relation of AlN growth rate and TMAl flow. As shown in

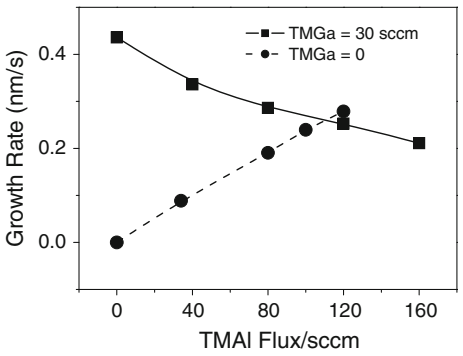


Fig. 4.24 The relationship between growth rate and TMAl flow. *Solid line* shows growth rate with a TAGa of 30 sccm, while *broken line* shows the growth rate with TAGa of 0 sccm

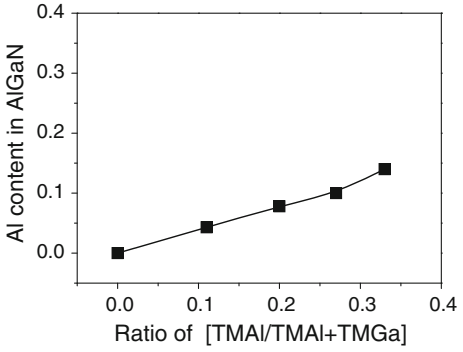


Fig. 4.25 The relationship between the actual measurements of Al components and Al components in gas phase

the broken curve in Fig. (4.24), the growth rate increases with the TMAl flow. This phenomenon can be explained as follows: during the growth of AlGaN material, there is a competition process involving Al and GaN atoms. As the bond energy of Al–N is stronger, Al atoms are more easily bonded to N and this process prevents the surface adsorption of Ga atoms. With the increase of TMAl flow, more and more Al atoms will bond with N atoms, but more and more Ga atoms are prevented to participate in material growth resulting in the decline of the material growth rate.

Al components in AlGaN material were also measured and studied. Figure 4.25 is the relationship between the actual measurements of Al components and Al components in gas phase (the mole ratio of Al in the source). It can be seen that the actual Al components in the material are far fewer than the Al components in the gas phase. The results in Fig. 4.24 show that Al atoms are more competitive to participate in the growth process, and the actual Al components should be more than the Al components in the gas phase. MOCVD growth is roughly divided into two processes: one is mass transportation and the second is surface reaction. Because of the activity of Al atoms, they will pre-react (also called parasitic reaction) with N atoms during the mass process. As this process consumes a large number of Al atoms, Al atoms that migrate to the surface to participate in the

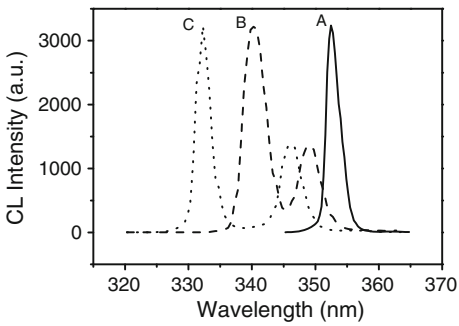


Fig. 4.26 CL spectrum for (Al)GaN samples at 6 K. The TMAl flow for sample A, B, C are 0, 90, and 160 sccm respectively

material growth are greatly reduced in number resulting in Al components in AlGaN material being fewer than the Al components in the gas phase.

The above experimental results also show that AlGaN material growth is greatly affected by Al atoms. There are many differences between the growth mechanism of GaN and AlGaN. The material quality of AlGaN is hard to improve. Mainly because the surface diffusion length of Al atom is very small, and the migration ability of Al is low, thus it is not easy to achieve two-dimensional growth. If the growth temperature is appropriately increased, Al atoms will acquire higher energy and the surface migration ability of Al atoms can also be increased so that AlGaN crystal quality and surface topography should be improved. In addition, as the pre-reaction of Al atoms in MOCVD system influence the components of AlGaN strongly, this pre-reaction process must be reduced.

Cathode luminescence (CL) tests have been carried out on the above-mentioned samples A, B, and C. Sample A is GaN material. Sample B and C are AlGaN materials with different Al components, while the Al component in sample B is higher in sample C. Figure 4.26 is the CL spectrum of sample A, B, C normalized by the peak value at 6 K. As can be seen from the chart, there is a simple light-emitting peak at 352 nm in sample A, which corresponds to the belt edge peak [68]. While sample B and sample C are completely different, both of them have two photoluminescence peaks at different positions. For sample B, the position of the photoluminescence peaks are at 340 and 349 nm, respectively. While photoluminescence peak position is located at 332 and 346 nm for sample C. The energy differences of the two peaks are 94 and 151 meV for sample B and C, respectively. With the increase in the Al component, the division of the photoluminescence peak is more obvious.

In order to study the photoluminescence peak splitting phenomenon in AlGaN material, cathode fluorescent imagery was used in the experiments. Figure 4.27a is the CL spectrum of sample C. It can clearly be seen that the two photoluminescence peaks are at 332 and 346 nm. Figure 4.27b is the surface topography for sample C. The surface of sample C is rather rough. Cathode fluorescent imaging is used for the two photoluminescence peaks in Fig. 4.27a. Figure 4.27c, d are the cathode fluorescent images for wavelength of 332 and 346 nm. As can be seen from the two figures, the distribution of the light intensity is uneven. And it should be noted that

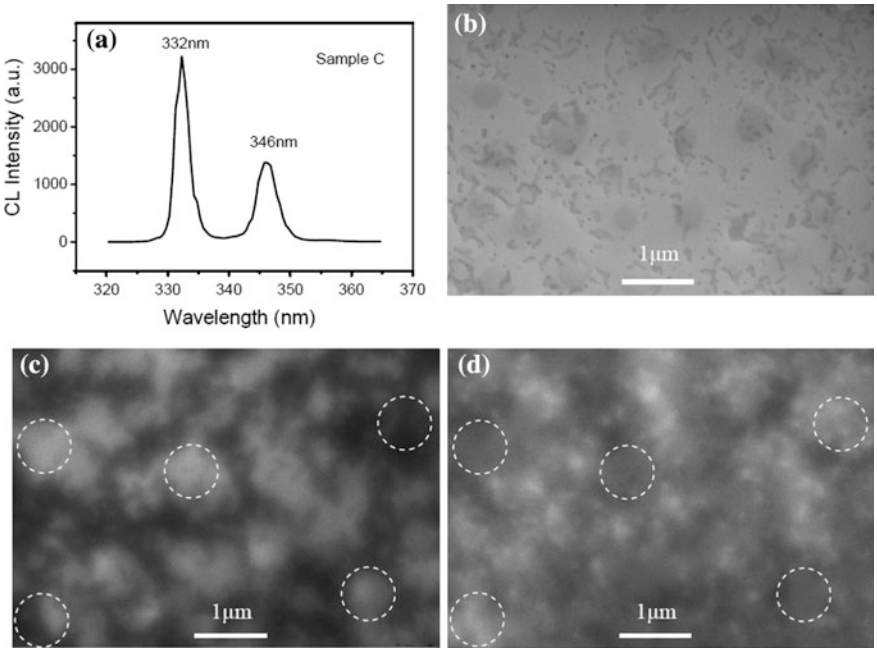


Fig. 4.27 **a** CL spectra of AlGaN sample C, **b** surface topography of sample C tested by SEM, **c** cathode fluorescent image at wavelength of 332 nm, **d** cathode fluorescent image at wavelength of 346 nm

the intensity distribution in the Figures has a certain degree of complementation. A few circles are traced in Fig. 4.27c, d. For these traced areas, if it is bright in (c), it will be dark in (d). But if it is dark in (c), it will be bright in (d). It can be understood that the two photoluminescence peaks are from the different areas of the AlGaN material. The main reason for this effect may be inhomogeneity of Al components during the material growth process. Both transverse luminescent properties and longitudinal optical properties of AlGaN material are studied. Figure 4.28 is the CL spectra under different acceleration voltages. It can be seen that the CL spectrum appears differently under different acceleration voltages. When the acceleration voltage is 5 kV, only two photoluminescence peaks were observed. When accelerating voltage was increased to 10 kV, three photoluminescence peaks can be observed. But when the accelerating voltage further increases to 15, 20 kV, five photoluminescence peaks were observed. As the penetration depth of the cathode ray is closely related to the accelerating voltage, the higher the accelerating voltage and deeper the penetration depth. So it can be believed that the fluctuation of the Al component is more significant in AlGaN material near the low-temperature GaN buffer layer.

Generally speaking, the surface migration ability of Ga atoms is far stronger than that of Al atoms [63, 69]. After the high-temperature annealing of the

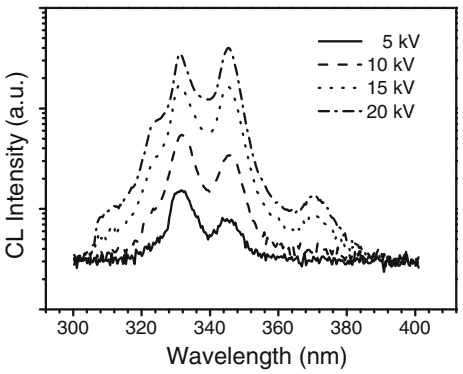


Fig. 4.28 CL spectrum for AlGaN sample at 6 K, with acceleration voltage of 5, 10, 15, and 20 kV

low-temperature buffer layer, the surface is very rough with many islands of nucleation center. At the beginning of the subsequent epitaxial process, the islands will grow with many Ga and AL atoms diffusing to the nucleation center and becoming involved in growth [70]. Because of the strong surface migration ability of Ga atoms, it is easy for them to spread to the nucleation center and make a large contribution to the growth of the islands. But for Al atoms with weak surface migration ability, it is too difficult for them to spread to nucleation centers to make contributions to the growth of the islands. So there is an uneven distribution of Ga atoms at the beginning of the growth. Ga content is very high around the islands, but the content of Al is relatively low. At other region, Ga content is low, but Al content is relatively high. At this time, the Al component is unevenly distributed in the material. In addition, at the beginning of the growth, the surface is rough and many crystal faces are exposed. As different crystal faces have different growth rate, this also leads to unevenly distributed Al components [71]. So that at the beginning of the growth, volatility of Al component is very high in the AlGaN material near low-temperature GaN buffer layer. Many photoluminescence peaks were observed in the CL spectrum. In the subsequent epitaxial growth process, the surface roughness will also increase gradually. There are still many hill-shape structures, and the component inhomogeneity will still exist. But the volatility of Al component will be better [72]. After the transition to the quasi-two-dimensional growth stage, the surface is relatively smooth. The crystal faces exposed are all (0001) faces and the growth rate is uniform. The difference in surface migration ability of Al and Ga will still have a weak influence on the volatility of Al component, but the uniformity of distribution for Al components at this time would be much better [73, 74]. So, the closer to the surface of the sample, the more evenly distributed are the Al components, and the fewer luminous summits in the CL spectrum. It follows that for the preparation of high-performance GaN-based detectors, it is better to prepare AlGaN material on buffers such as smooth GaN or AlN epitaxial layer to reduce the inhomogeneity of the Al component.

4.3.2 AlN Buffer Layer and AlGaN Epitaxial Technique

To fabricate large-scale GaN ultraviolet detector focal plane arrays, it is best to adopt a back-illuminated structure which enables the chips of the device array and the readout circuits to be interconnected by the flip chip technology. If a back-illuminated GaN detector is to be prepared, AlN but not GaN must be used as a buffer layer [75, 76]. As a direct band gap semiconductor material, GaN has a high absorption coefficient, which will greatly reduce the number of the photons exposed to the active region, thus greatly reduce the response rate of the device. As a foundation for the preparation of high-quality AlGaN material, GaN and AlGaN epitaxial technology on an AlN buffer layer is introduced in this section based on the GaN and AlGaN epitaxial technology employed on the previously discussed GaN buffer layer.

1. GaN epitaxial growth with AlN buffer layer

The study of the growth of GaN on AlN buffer layers can enable a more thorough understanding of the influence of the AlN buffer layer on the epitaxial layer. It provides a foundation for the further study of the growth of AlN and AlGaN epitaxial layers on AlN buffer layer [20, 22, 47, 77–80]. The growth of GaN material on AlN buffer layer is introduced as follows [67]. The growth process is: first, the low-temperature growth of AlN buffer layer at 540 °C, second, the growth of a GaN epitaxial layer after the temperature-rising annealing to 1040 °C. The time of temperature-rising annealing and the effect of low-temperature AlN thickness on the growth of the GaN epitaxial layer are now discussed.

Figure 4.29a, b are the in situ monitoring curves of sample A and B grown by MOCVD. The thickness of the AlN buffer layer for both samples is 20 nm, but the time of the temperature-rising annealing is different with 1000 s for sample A and 300 s for sample B. The monitoring curves can be divided into three parts: (i) the growth of the low-temperature AlN buffer layer, (ii) the temperature-rising annealing of the AlN buffer layer, (iii) the GaN epitaxial growth at high temperature. The growth process and growth mechanism can be studied because the monitoring curve is sensitive to the change of surface roughness. As can be seen from the two growth curves, the different temperature-rising annealing times caused a great difference in the growth process. In the temperature-rising annealing of AlN buffer layer of sample A, there is no obvious change in the curve. But at the beginning of the GaN epitaxial growth, the reflectivity is very low. Then the reflectivity of the curve gradually increases and faint oscillation of the interference fringes began to appear. Finally, reflectivity reached a stable value and there are obvious interference fringes in the curve. The growth process of sample B is different. Although in temperature-rising annealing process (ii), there is no obvious difference between sample A and B the GaN epitaxial growth process (iii) is completely different. At this point, a relatively high reflectivity and obvious interference fringes began to appear in the curve, and in the following growth process, the reflectivity soon reaches a stable value, and interference fringes are more obvious. There are commonly three

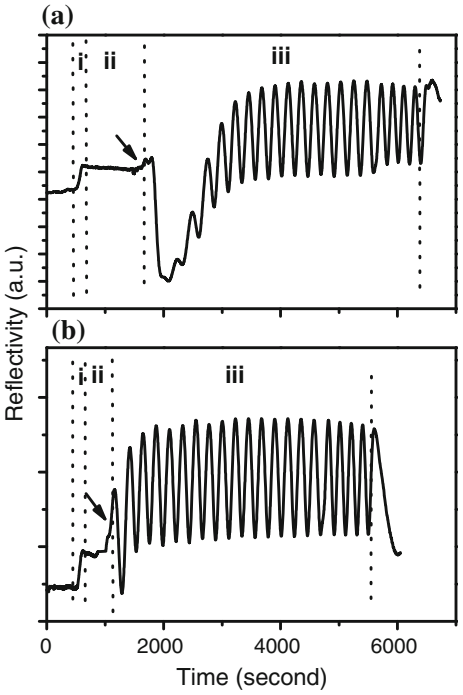


Fig. 4.29 **a** and **b** are respectively the in situ monitoring curves of sample A and B. Time for temperature-rising annealing of sample A and B are 1000 and 300 s. The thickness of the AlN buffer is 20 nm for both samples

processes in the growth of GaN: three-dimensional growth, island merger and quasi-two-dimensional growth. In the early stage of growth, the longitudinal growth rate is greater than the lateral growth rate, and the growth is three dimensional. In this stage, the epitaxial layer is an island-shaped structure and the surface is very rough. As the growth continues, the lateral growth began to be dominant, the islands gradually merged, and the surface began to flatten gradually. In the end, the islands were almost entirely merged, the epitaxial layer started quasi-two-dimensional growth, and the surface is very smooth. In the epitaxial growth of sample A, the reflectivity is low, which means the surface is very rough and the growth process is a three-dimensional pattern. As the growth process continues, the reflectivity increased and weak interference fringes began to appear, which means the surface gradually flattened, lateral growth began to be dominant and the islands began to merge. Finally, the reflectivity of the curve reaches a stable value and there are obvious interference fringes. Now the island merging process was basically completed and the growth process transformed to quasi-two-dimensional pattern. The surface is smooth. But for the epitaxial growth of sample B, the reflectivity in the curve is not as low as that of sample A in the initial stage of the growth. And there are obvious interference fringes in the curves. This means the time for three-dimensional island growth is very short, and the lateral growth quickly became dominant and the island merging process soon appeared and quickly finished. The growth process became two dimensional. It can be seen from the experiment results that lateral

Table 4.2 Test results and growth parameters of GaN epitaxial layer on AlN buffer layer

Sample	AlN buffer layer		XRD FWHM (arc-min)		Carrier mobility (cm^2/V s)
	Annealing time (s)	(0002) (nm)	(0002)	(10-12)	
A, E	1000	20	6.9	11.2	360
B	300	20	8.1	19.1	142
C	1000	45	10.2	28.6	73
D	1000	30	7.2	13.9	217
F	1000	16	–	–	–

epitaxial growth of GaN can be improved by temperature-rising annealing of AlN for a long time. For sample A and sample B, both the growth process and the crystal quality are significantly different. As shown in Table 4.2, for the XRD test results of half high width, the symmetric diffraction on (0002) face and skew symmetric diffraction on (10-12) face for sample A are narrower than that of sample B. Generally speaking, for GaN material, half high width of the symmetric diffraction on (0002) face and skew symmetric diffraction on (10-12) face reflect screw dislocation and edge dislocation density, respectively. The greater the half high width, the higher the dislocation density will be. XRD results show that the quality of sample A is better than that of sample B. On the other hand, carrier mobility of sample A is also higher than that of sample B, which again illustrates that the quality of sample A is better than that of sample B.

In addition to the time for temperature-rising annealing, the thickness of the low-temperature AlN buffer layer also has great influence on the growth of the GaN epitaxial layer. In order to study the influencing mechanism of AlN buffer layer thickness on the GaN epitaxial layer, AlN buffer layers with thickness of 45, 30, 20, and 16 nm numbered as sample C, D, E, F were prepared. Figure 4.30a–d are the in situ monitoring curves of sample C, D, E, and F. The time of temperature-rising annealing is 1000 s for these samples. Observing the curves, it can be seen that there are great differences between the growths of the four samples. In the growth process of sample C, surface reflectivity soon reaches a stable value and significant interference fringes appear. This means the island merging process finished quickly and the growth process soon change to two-dimensional. In the growth process of sample D, the surface reflectivity in the early stage of growth is small. As the growth continues, the surface reflectivity soon reaches a stable value and significant interference fringes appear. It means the growth process is three dimensional in the beginning. But soon the islands merge and quasi-two-dimensional soon begins. While in the growth process of sample E, the surface reflectivity is low and there are only weak interference fringes. With the extension of time, the reflectivity gradually increased and significant interference fringes appeared which illustrated the transition from three-dimensional to quasi-two-dimensional growth emerged in sample E. In the growth process of sample F, the surface reflectivity became lower and lower, which means the surface is more and more rough. As there were no signs of leveling of the surface, the growth process was interrupted. That is to say, it was

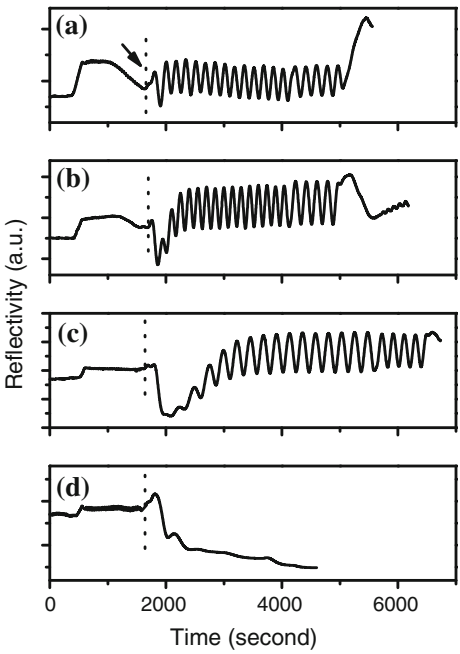


Fig. 4.30 **a–d** are respectively the in situ monitoring curves of sample C, D, E and F. The thickness of the AlN buffer is 45, 30, 20 and 16 nm for sample C, D, E and F. Time for temperature-rising annealing of these samples is 1000 s

always three-dimensional growth in the growth process of sample F. Since the difference between the growth processes were so significant, so were the difference of qualities between the samples. As shown in Table 4.2, among C, D, E, and F, sample E has the best quality. It has not only the minimal dislocation density, but also the highest carrier mobility. Thus, the quality of the GaN epitaxial layer will be adversely affected by an AlN buffer layer with a thickness too thick or too thin. Appropriate thickness is beneficial to the growth of high-quality GaN epitaxial layer.

The surface roughness of GaN buffer layer is very important for the high-temperature growth of GaN epitaxial on low-temperature GaN buffer layer. High-quality GaN epitaxial layer can only be prepared on GaN buffer layer with a rough surface [38, 39, 58]. But it can be seen from 4.30 (a) and (c), after the temperature-rising annealing, reflectivity of sample C reduced and the surface was rough, while reflectivity of sample E changes slightly and the surface topography is smooth although the quality of the epitaxial layer varies widely. Thus, high-quality GaN epitaxial layer can be prepared not only on low-temperature AlN buffer layers with rough surface. This may be the different influencing mechanism for quality between low-temperature AlN and GaN buffer layers.

In order to further understand the influencing mechanism of an AlN buffer layer on GaN epitaxial layers, surface morphology of AlN buffer layer after temperature-rising annealing was studied. Figure 4.31a–c is the surface morphology of a low-temperature AlN buffer layer for sample A, B, and C. The samples are

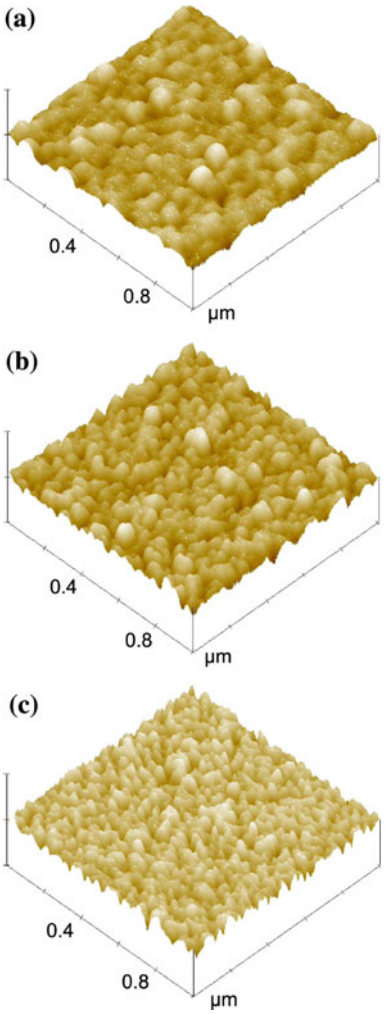


Fig. 4.31 **a–d** are respectively the surface morphology of low-temperature AlN buffer layer after annealing. The sample numbers are expressed as **a–c**

numbered as a (1000 s annealing and with a thickness of 20 nm), b (300 s annealing and with a thickness of 20 nm) and c (1000 s annealing with a thickness of 45 nm). It can be seen from the chart that sample A has the largest grain size and the minimum nucleation density, while sample C has the smallest grain size and the largest nucleation density. In the growth process of sample A on buffer layer, there will be growth processes such as three-dimensional growth, island merge, and a quasi-two-dimensional. During the island merging process, many dislocations will steer and annihilation. So dislocation density in the epitaxial layer will decrease and the epitaxial layer quality will be improved. For the growth process of sample C on buffer layer c, because of the low nucleation density, the islands will be merged soon to complete the transition to the quasi-two-dimensional growth and a large

number of dislocations will penetrate the epitaxial layer. Of course, if the nucleation density is too small, the islands cannot merge. The growth process will remain three dimensional and the quality of the sample will be poor. Therefore, large grain size and a small nucleation density in a low-temperature AlN buffer layer are very important for the high-temperature GaN epitaxial layer growth. A long-time temperature-rising annealing of AlN buffer layer with appropriate thickness can achieve this. Of course, there are many differences between the growth of GaN epitaxial on AlN buffer layer and GaN buffer layer, therefore further research on various configurations should be done.

2. The pre-reaction of Al atoms and the influence on AlN growth rate

To fabricate the high-performance solar-blind GaN UV detector, it is very important to achieve the accurate control of Al components in AlGaN materials. The pre-reaction of Al atoms is serious in the MOCVD growth process [81–86]. As many Al atoms were consumed before spreading to the surface to participate in the reaction, it is difficult to achieve high component values for AlGaN material. In general, if the pre-reaction is weak, more atoms will involve in the reaction and the growth rate of AlN material will be higher. On the other hand, if the pre-reaction is strong, more and more Al atoms will participate in the pre-reaction before getting to the surface and the growth rate of AlN will be slow. In order to study the pre-reaction mechanism of Al atoms, the relationship between AlN growth rate and growth parameters in MOCVD system is studied [87].

Figure 4.32 is the relationship between AlN growth rate and growth temperature, at the same time, the relationship under different pressure of the reaction chamber is shown. It can be noticed that at a chamber pressure of 50 Torr, AlN growth rate slightly increased with the increase of growth temperature. But when the reaction chamber pressure is 200 Torr, the growth rate of AlN obviously decreased with the increase of growth temperature. When the reaction chamber pressure is 500 Torr, the growth rate of AlN dropped rapidly. Generally speaking, because of the strong

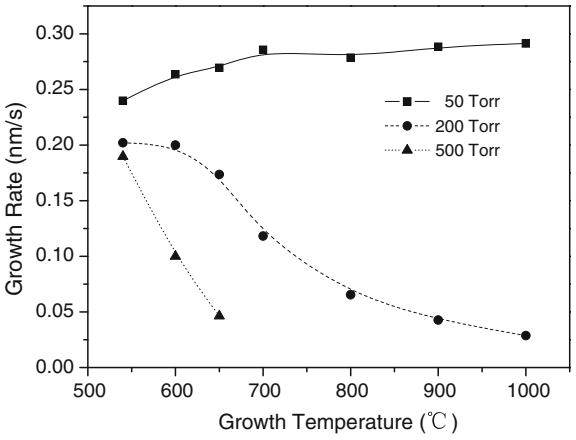


Fig. 4.32 The relationship between AlN growth rate and temperature

bonding energy, it is not easy for NH_3 to crack at a low temperature. According to the normal rules, with the increase of growth temperature, the pyrolysis efficiency of NH_3 increases, and the growth rate should be increased. But the results of the experiments are opposite at a rather high pressure of the reaction chamber. Even when the reaction chamber pressure is only 50 Torr, with the increase of the growth temperature, there is only a weak increase in the growth rate. In the MOCVD system, the growth of material can be divided into two processes: mass transport and surface reaction. It is likely that Al and N atoms pre-react during the process of mass transport, and part of the source is consumed. Consequently, fewer of the Al and N atoms can spread to the surface to participate in the reaction, which leads to a loss of material growth rate.

In the MOCVD system, the pre-reaction of Al atoms are as follows:

$$Al(CH_3)_3 + NH_3 \rightarrow Al(CH_3)_3{\cdot}NH_3 \tag{4.12}$$

$$Al(CH_3)_3{\cdot}NH_3 \rightarrow Al(CH_3)_2{\cdot}NH_2 + CH_4 \tag{4.13}$$

$$n Al(CH_3)_2{\cdot}NH_2 \rightarrow \left[Al(CH_3)_2{\cdot}NH_2\right]n, \quad n \geq 3 \tag{4.14}$$

$$2Al(CH_3)_2{\cdot}NH_2 \rightarrow \left[Al(CH_3)_2{\cdot}NH_2\right]_2 \rightarrow \text{AlN particles} \tag{4.15}$$

$$\left[Al(CH_3)_2{\cdot}NH_2\right]_3 \rightarrow \text{AlN particles} \tag{4.16}$$

During the MOCVD growth process, AlN particles grown by dimer $[Al(CH_3)_2{\cdot}NH_2]_2$ according to (4.15) will be taken away by the air flow and will have no contribution to the growth. The same condition occurs with trimer $[Al(CH_3)_2{\cdot}NH_2]_3$ [60]. As the temperature increases, the possibility for dimer and trimer to form AlN particles also increases. As a result, when the pre-reaction is dominant, the growth rate of AlN decreases with the increase in temperature. The experiment results show that when the pressure of the reaction chamber is 50 Torr, the growth rate of AlN increases slightly with an increase of temperature. This means the surface reaction process is dominant at 50 Torr. When the pressure of the reaction chamber increases to 200 Torr, the pre-reaction process is dominant, and the growth rate decreases with an increase in the temperature. As the pressure of the reaction chamber further increases to 500 Torr, the pre-reaction process is more dominant, and the growth rate of AlN falls even faster with an increase in the temperature.

From the experiment results in Fig. 4.32, there is a certain relationship between the growth rate and the pressure of the reaction chamber. As shown in Fig. 4.33, when the pressure of the reaction chamber increases, the growth rate of AlN drops. Especially when the growth temperature is higher, the growth rate of AlN drops quickly. In the chemical reaction process of (4.14), whether the trimer can produce AlN particles or not, they will have no contribution to the growth. For the polymer, it is carried away by airflow and has no contribution to the growth [88]. But with the increase of the pressure of the reaction chamber, the chemical reaction will be

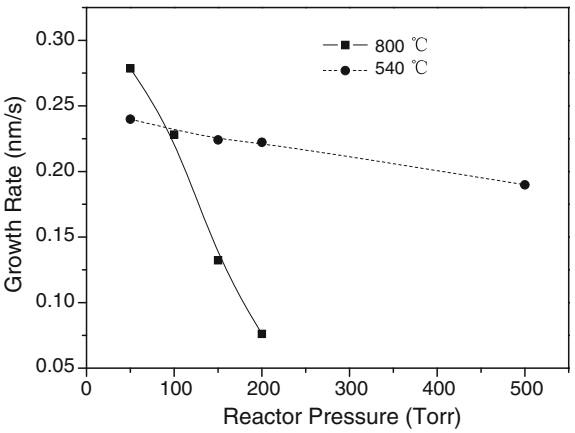


Fig. 4.33 The relationship between the growth rate of AlN and the pressure of the reaction chamber

strengthened, and more and more trimer and polymer will be formed. Strengthening of the pre-reaction will cause the growth rate of AlN to reduce with an increase in the chamber pressure. In addition, the reduction of the pressure in the reaction chamber will increase the velocity of the airflow. Thus the opportunity for Al and N atoms to meet each other will be weak, and the pre-reaction also becomes weak. If the growth temperature was increased at the same time, the reaction velocity of (4.14) will be faster. That is to say, at the same pressure of the reaction chamber, the pre-reaction will be strengthened by a high temperature, and the growth rate of AlN will be lower. Thus in the case of growing at a high temperature, the growth rate of AlN decreases fastly with an increase of the pressure in the reaction chamber. It can be seen from Figs. 4.32 and 4.33 that, growth temperature and the pressure of the reaction chamber significantly influence the pre-reaction.

Figure 4.34 is the relationship between the growth rate of AlN and the flux of NH_3. It can be noticed that whatever other growth conditions, with the increase of flux of NH_3, the growth rate of AlN will decrease. The difference of the decrease is shown in the curve slope. According to the normal rules, the growth rate will fall with the increase of the flux of source. This also suggests the pre-reaction process exists in the MOCVD system. The research results show that the energy needed to form trimer and polymer in chemical reaction Eq. (4.14) can be reduced by NH_3. That is, when the flow rate of NH_3 increases, the reaction rate of the pre-reaction (4.14) will increase and the pre-reaction will be strengthened, and there will be more and more trimer and polymer. The influence of NH_3 flow on the growth rate relates with other growth conditions. As shown in Fig. 4.34, when the pressure of the reaction chamber is 50 Torr and the temperature is rather high, the growth rate of AlN falls faster with an increase of NH_3 flux. The pre-reaction is more serious now. When the growth temperature is 540 °C and the pressure of the reaction chamber is rather high, the growth rate of AlN will decrease faster. From the experimental results in Fig. 4.33, with an increase of the pressure in the reaction chamber, the pre-reaction will be stronger. So the growth rate of AlN will be

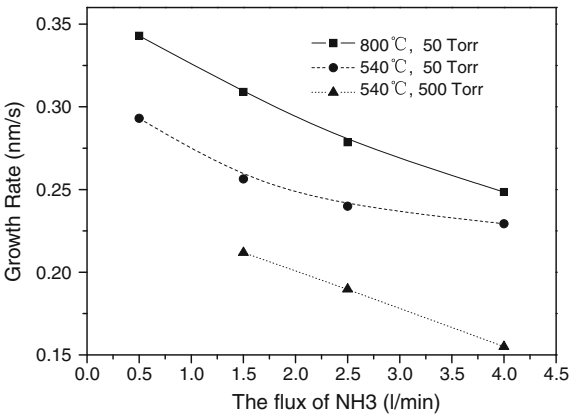


Fig. 4.34 The relationship between the growth rate of AlN and the flux of NH_3

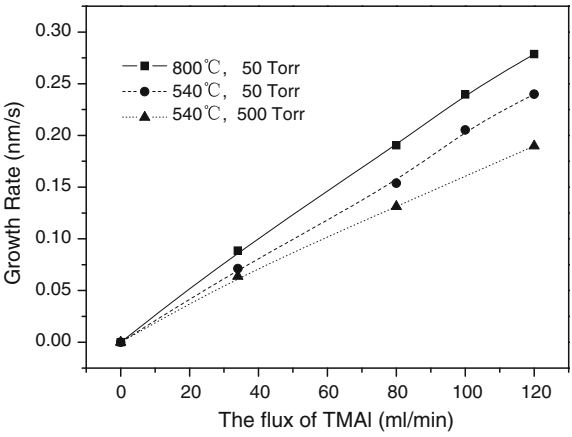


Fig. 4.35 The relationship between the growth rate of AlN and the flux of TMAl

reduced more quickly with an increase in the flux of NH_3. Of course, if the flux of NH_3 is too small and enough N atoms cannot be provided, though the pre-reaction is weak, the growth rate of AlN will decline.

The relationship between the growth rate of AlN and TMAl flux is also studied. As shown in Fig. 4.35, with the increase of TMAl flow, regardless of other growth conditions, the growth rate increases according to normal rules. However, it should be noticed that the increase slope of AlN growth rate is associated with specific growth conditions. By the experimental results in Fig. 4.33, when the pressure of the reaction chamber is reduced to 50 Torr, pre-reaction is relatively weak and the normal material growth process is dominant. So when TMAl flow increases, the surface reaction is promoted by the rather high temperature and the growth rate of AlN increases rapidly. When the growth temperature is fixed at 540 °C and the pressure of the reaction chamber is low, the growth rate of AlN increases fast with the increase of the TMAl flux. This further illustrates that a rather high-reaction

chamber pressure will aggravate the pre-reaction. The experimental results in Fig. 4.35 further prove that the pre-reaction strongly depends on the reaction chamber pressure and the growth temperature.

It is found in the material growth that the effect of pre-reaction is weak on the growth process of GaN. It can be inferred that the Al components in AlGaN is determined not only by the number of moles of Al/Ga, but also by the extent of the pre-reaction. From the experimental results shown in Figs. 4.32, 4.33, 4.34, and 4.35, a rather low temperature, a rather low pressure of the reaction chamber and a rather low flux of NH_3 are conducive to reduce the pre-reaction. In the actual material growth of AlGaN, it is found the Al component can be improved by reducing the pressure of the reaction chamber appropriately. It also illustrates Al components are greatly influenced by the pre-reaction. Of course, besides the components of Al in AlGaN material, material quality of AlGaN has a very large impact on the performance of the device. The reduction of pre-reaction does not imply an improvement in the material quality. In the material growth process, it is necessary to combine the control of Al components and the quality increase of the material to prepare a high-performance device structure.

3. The influence of low-temperature AlN buffer layer thickness on the quality of AlGaN epitaxial

The following discussion focuses on the influence of different thickness of an AlN buffer layer on the AlGaN epitaxial layer crystal quality and the surface morphology, etc. This section will further study the mechanism of AlGaN epitaxial layer growth [89]. Figure 4.36 is the in situ monitoring curves of three AlGaN samples grown on the low-temperature (540 °C) AlN buffer layer [90, 91]. Figure 4.36a–c in the figure corresponds to the samples with a buffer layer thickness of 10, 20, and 45 nm. The growth process includes including the pretreatment of the substrate, the growth of the AlN buffer layer, annealing process, and the growth of AlGaN epitaxial layer. As the thickness of the low-temperature AlN buffer layer increases, significant changes have taken place on the crystal quality and surface morphology. As shown in Fig. 4.36a, when the thickness of the buffer layer is as thin as 10 nm, the grain size is large and the nucleation density is small after the annealing process [67]. In the subsequent merging process, the merge is incomplete because the nucleation density is too small and the lateral surface migration of Al atoms is too slow. Finally the size of the islands is large. The surface is very rough when it gets to a certain thickness and the reflectivity will be attenuated nearly to zero. The grain size decreases and the nucleation density increases after annealing when the thickness of the buffer layer increases as shown in Fig. 4.36b. This leads to a rapid merging process and quasi-two-dimensional growth mode begins after the island merging is completed. As shown in Fig. 4.36c, when the thickness of the buffer layer increases to 45 nm, the grain size decreases continuously and the nucleation density increases continuously after annealing. This leads to an even faster merging process. A large number of dislocations were produced in this process, which makes the quality of the epitaxial poor and

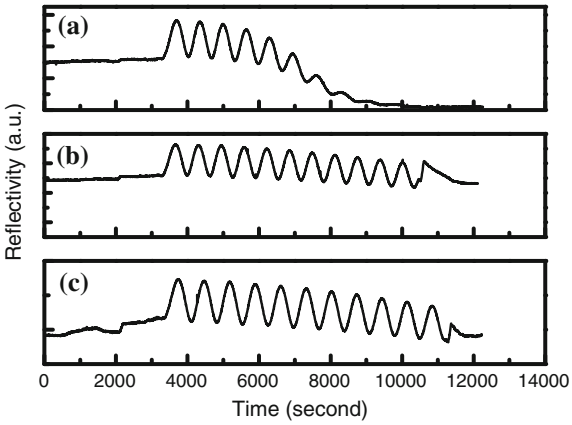


Fig. 4.36 The in situ monitoring curves of three AlGaN samples. The thickness of the low-temperature AlN buffer layer for **a–c** is respectively 10, 20 and 45 nm

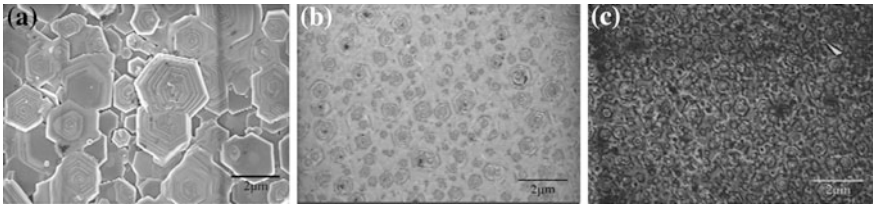


Fig. 4.37 The change of surface topography of AlGaN epitaxial layer with buffer layer thickness. The thickness of the low-temperature AlN buffer layer for **a–c** is respectively 10, 20, and 45 nm

quasi-two-dimensional growth mode begins after the complete merging. Surface topographies of the three samples are shown in Fig. 4.37. When the buffer layer is thin, the growth process tends to be three dimensional. The surface is rough because it consists of large hexagonal islands in micron size and these islands have not merged together to form a film. With the increase of the thickness of the buffer layer, the growth process of the AlGaN epitaxial layer tends to be quasi-two-dimensional and a film has been formed in the surface. A large number of hexagonal islands still appear on the surface because of the small lateral migration of Al atoms. But the size is far less than the islands on the surface of the thin AlGaN buffer layer.

The crystal quality of the AlGaN epitaxial layer changes with the thickness of AlN buffer layer, as shown in Fig. 4.38. The thickness of the low-temperature AlN buffer layer is 10–45 nm. At first, the crystal quality improves with the increase of the buffer layer thickness. But as the thickness increases beyond a certain value, the crystal quality begins to decrease. The main reason for the change of crystal quality lies in the merging process of material growth. When the buffer layer is thin, it is

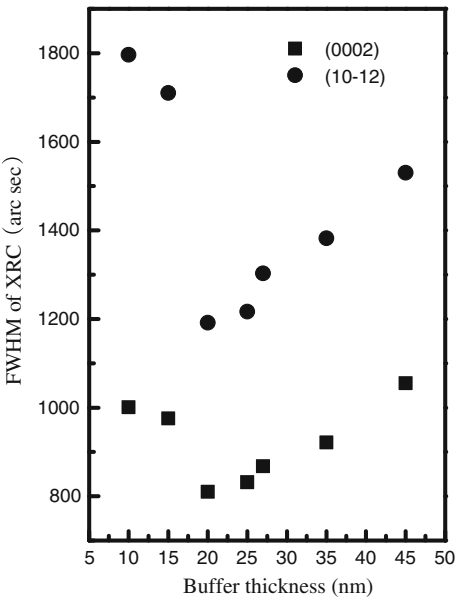


Fig. 4.38 DCXRD of AlGaN epitaxial layers, the relationship between the half-width of the rocking curve for (0002), (10-12) and the thickness of buffer layer

difficult for nucleation to merge. Dislocations are usually initiated from the interface between the sapphire and buffer layer and they directly penetrate the epitaxial layer surface almost without bending or merging, so there are large numbers of dislocations. When the buffer layer thickness increases to between 20 and 30 nm, the dislocation density is decreased during the islands merging process because of the small nucleation density and relatively slow merging process and partially bending and tilting of the dislocations. Thus the crystal quality is improved. As the thickness of the buffer layer further increases, the density of nucleation center is higher, and the merging process is very fast. Many dislocations cannot steer and annihilate, so the quality of crystal decreases.

The cathode fluorescent (CL) technique is used to study the effect of buffer layer thickness on the optical properties of the AlGaN material [92]. Figure 4.39a is the CL spectrum for AlGaN epitaxial with buffer layer thicknesses of 10, 20, and 45 nm. The test is finished at a temperature of 6 K. It can be seen that the intensity of band-edge photoluminescence peak is the highest and the half-width is the smallest for the sample with a buffer layer thickness of 20 nm while the intensity of band-edge photoluminescence peaks is rather weak for the other two samples. For sample A (with the thinnest buffer layer), there is only one peak near the band edge while there is a weak peak at the low-energy side of the band-edge photoluminescence peak for the other two samples (Perhaps the components are uneven in the two samples). Figure 4.39b is the spectrum near the low-energy limit, and the luminescence within 3.2–3.7 eV is labeled as “blue”, the luminescence within 2.2–2.5 eV is labeled as “yellow”. Table 4.3 is the intensity of the photoluminescence peak normalized by the intensity of band-edge photoluminescence peak and

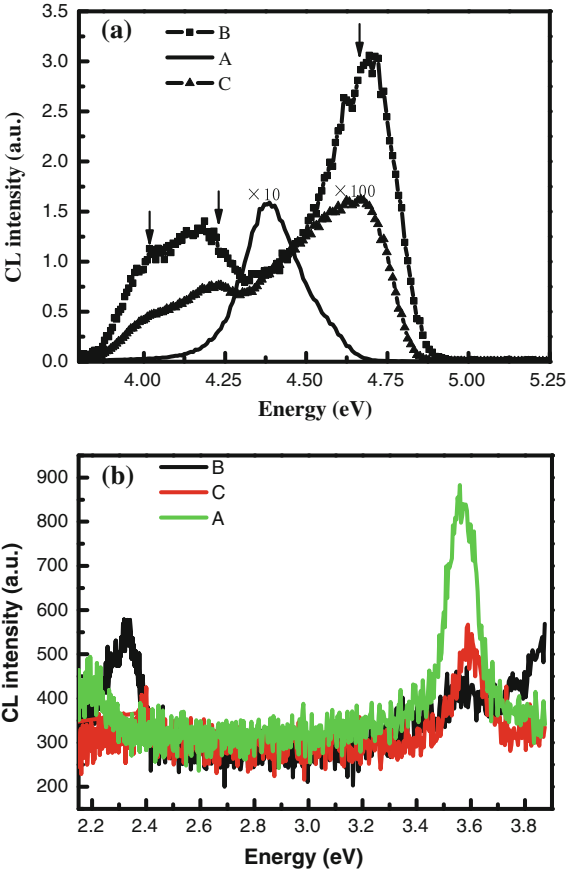


Fig. 4.39 CL spectrum of AlGaN epitaxial layer with different thickness of AlN buffer layer. The thickness of the buffer layer for sample A, B, and C is 10, 20, and 45 nm

Table 4.3 Crystal quality and optical quality of AlGaN samples with different buffer layer thickness

Sample	AlN buffer thickness (nm)	X-ray FWHM (arc-min)		$\frac{I_{BL}}{I_{NBE(1)}}$	$\frac{I_{BL}}{I_{NBE(2)}}$	$\frac{I_{YL}}{I_{NBE(1)}}$	$\frac{I_{YL}}{I_{NBE(2)}}$
		(0002)	(10-12)				
A	10	16.68	29.94	0.019		<0.009	
B	20	13.5	19.86	≈0	≈0	0.008	0.007
C	45	17.58	25.5	0.056	0.033	0.009	0.005

their ratio. It can be found that when the thickness of the buffer layer is thin, the intensity of the blue and yellow peak is strong. But when the thickness of the buffer layer is too thick, the blue peak is the strongest. Only when the buffer layer thickness is 20 nm, the blue peak is too small and the intensity of yellow peak is in the middle. The origin of the photoluminescence peaks is different, but it relates to impurities, defects, or deep levels together. It is generally believed that the higher

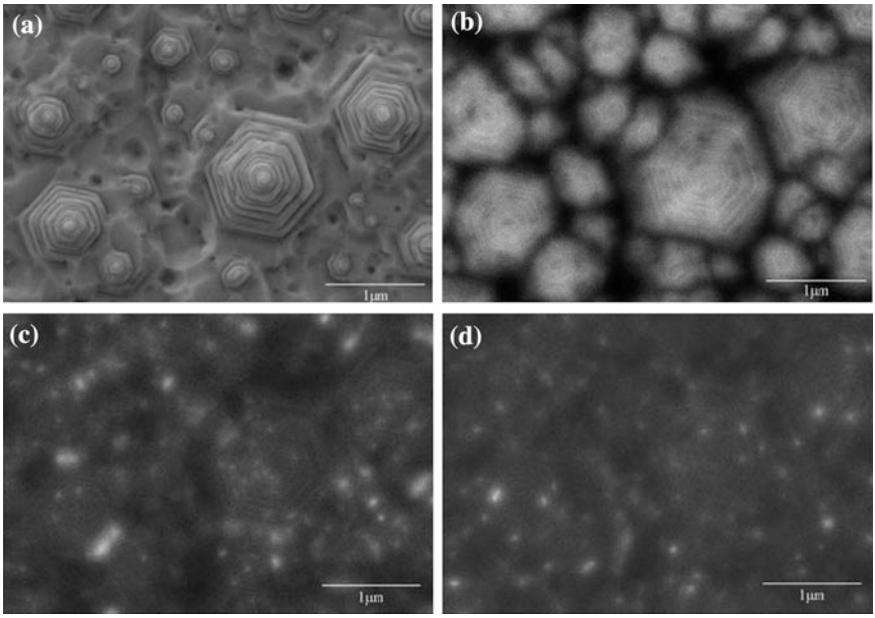


Fig. 4.40 **a** The secondary electron image of the sample and CL images with different energies, **b** 4.66 eV, **c** 4.22 eV, **d** 4.02 eV

the intensity of the peak, the poorer the quality of the material. It is shown by the optical test results that when the thickness of the buffer layer is 20 nm, the crystal quality is best. This is consistent with the result of X-ray diffraction (XRD).

The weak peaks at the low-energy side of the band-edge photoluminescence are now discussed. For the sample with a buffer layer thickness of 20 nm, at the position of Fig. 4.40(a), energies of 4.66, 4.22, and 4.02 eV were used to get the CL images in Fig. 4.40b–d. The three energy levels correspond to the three arrows in Fig. 4.39. There are some black spots in the middle and boundary of the domain in Fig. 4.40b, which corresponds to the bright spots in Fig. 4.40c, d. This indicates that these black spots are caused by the luminescence of the area with low Al components. It can be seen that there are some black spots in the three CL images at the same position, it maybe relate to the non-radioactive luminescence of the defects and impurities.

In order to determine the surface component of the sample, the CL line scan is used to scan along the white line in Fig. 4.41a from left to right with a step of 200 nm. As shown in Fig. 4.41b, the spectra 1, 2 and 6, 7, 8 recorded the luminescence of the left and right of the island. Spectra 3, 4, and 5 record the luminescence of the island. It is found that spectra 3, 4, 5 at a highest energy have the highest luminous intensity although the spectra for 7 and 8 at the lowest energy have the weakest luminous intensity. This fully shows that the component of the island is higher than that of the platform. While the component is the lowest in the pit, that is,

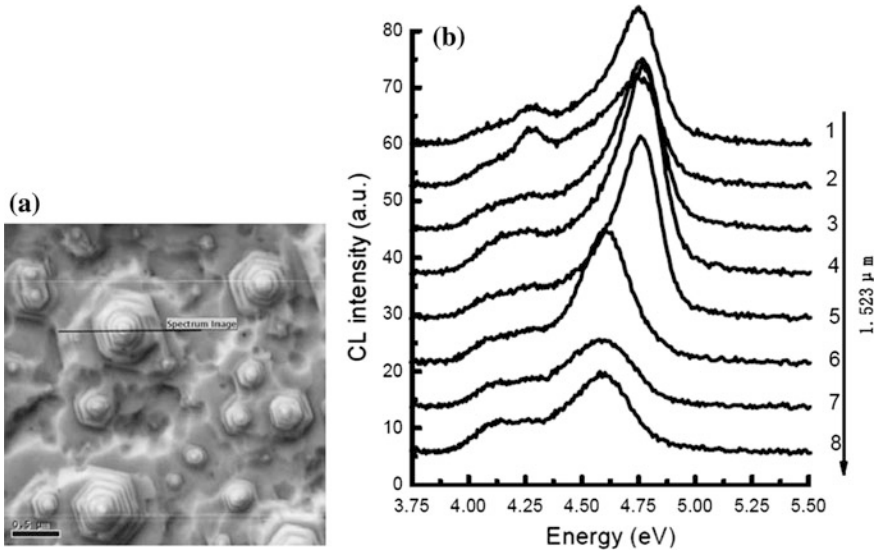


Fig. 4.41 Line scan of the sample

lower than that of the platform. Generally, other luminous intensity is related to material quality and surface topography. It can be seen from Fig. 4.40b–d, most black spots are located at the boundary of the domain, which explains the luminous intensity of the platform and sunken place is very low. In general, the pits on the surface of the sample are connected with the dislocations; this will also reduce the luminous intensity. Another possibility is that the dislocations may absorb the impurities, and dislocations with high content of impurities also affect the luminous intensity.

4. Influence of AlGaN growth parameters to material quality

There are two important factors that influence the quality of the AlGaN epitaxial layer. One is the pre-reaction between TMAl and NH_3. The other is Al atoms have a lower lateral migration rate at the growth surface. Usually the parameters associated with pre-reaction are V/III ratio (including NH_3 flow, TMAl flow), the pressure of the reaction chamber, growth temperature, and the structure design of the reaction chamber, etc. The specific reaction mechanism has been introduced in detail in the second section of this chapter. The following discussion will study the influence of the growth parameters on the quality of the epitaxial layer [89].

First, the influence of V/III ratio on AlGaN epitaxial layer is studied. The value of V/III ratio has an important impact on the AlGaN epitaxial growth. It affects both the merging and growth rates of the Al atoms, and also affects the surface migration rate of the Al atoms thus affecting the growth rate of the material, surface morphology, and the crystal quality. NH_3 flow is changed here to make the growth of AlGaN epitaxial layer V/III ratio of 783, 1175, 1958, and 1175, respectively.

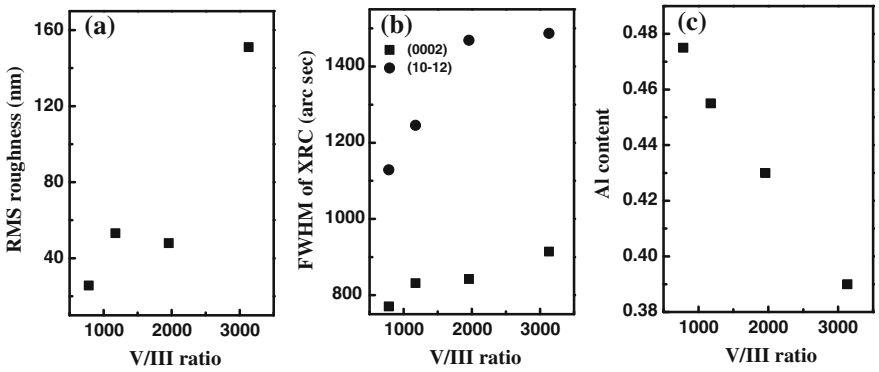


Fig. 4.42 Influence of V/III ratio to surface morphology, crystal quality, and the merge rate of Al atoms

Figure 4.42a is the change trend of the surface roughness with V/III ratio. It is found that with the increase of V/III ratio, the surface roughness will follow. It is also found by AFM images that the surface is full of hexagonal islands and the diameter and height of these components also increases with the increase of V/III ratio. This is different from the growth of a GaN epitaxial layer. For GaN, the growth process tends to be lateral and the surface tends to be smoother with a high V/III ratio. So in the process of AlGaN material growth, besides the fact that pre-reaction will worsen the surface morphology, a high V/III ratio will restrain the diffusion rate of Al atoms on the growth surface perhaps because under the condition of a high V/III ratio, more NH_3 can reach the growth surface and more N atoms can be cracked. But as Al atoms that can reach the growth surface are few and the bond energy of Al–N is high, Al atoms will form AlN with N without spreading. And this process will lead to a decrease of the lateral growth rate and the surface will be smoother. Other results show that the surface migration ability of Al atoms decreases under a high V/III ratio [93]. In addition, AlN powder generated by the pre-reaction of NH_3 and TMAl is not completely taken away by the carrier and some will fall on the growth surface which will lead to growth and nucleation, and this will also affect the surface morphology of the material.

Figure 4.42b is the change of crystal quality with V/III ratio. Its rule is consistent with the relationship between crystal quality and surface morphology. As the V/III ratio increases, the crystal quality will be poor. Especially for the (10-12) face, the half-width change is larger. This is still attributed to the influence of two aspects: one is the pre-reaction, and another is the V/III ratio. Because the surface diffusion velocity of Al atoms will be influenced by the V/III ratio, Al atoms cannot get to an energy optimal position, which leads to a large stress. But a means to release stress is through the generation of dislocations. Because AlN powder will be generated by the pre-reaction and will fall on the growth surface, it will introduce a large number of dislocations due to the lattice mismatch. Additionally, with the increase in the V/III ratio, the reaction will be stronger, and crystal quality will be poorer. At the

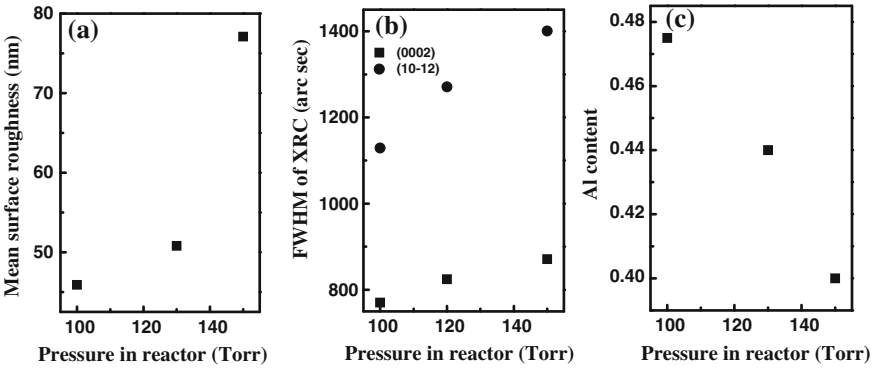


Fig. 4.43 Influence of pressure of the reaction chamber to the surface morphology of AlGaN epitaxial, crystal quality and the merge rate of Al atoms

same time, a strong pre-reaction will lead to exhaustion of TMAl source, and few Al atoms can reach the growth surface. This will affect the merging rate of Al atoms; and make the Al component in AlGaN lower than that in gas phase, as shown in Fig. 4.42c. It can be concluded that with a high V/III ratio, the low merging rate of Al atoms will result in bad crystal quality and a rough surface.

For the growth of AlGaN, the pressure of the reaction chamber is another important factor besides V/III ratio. Here the effect of pressure of the reaction chamber on AlGaN material is studied. In this experiment, the V/III ratio for the growth of AlGaN is 783, the pressure of the reaction chamber is 100, 130, and 150 Torr. Figure 4.43 is the influence of pressure of the reaction chamber upon the surface morphology of AlGaN epitaxial, crystal quality, and the merge rate of Al atoms.

With an increase in the pressure of the reaction chamber, the surface of the epitaxial layer will be rougher, and RMS grows, as shown in Fig. 4.43a. The quality of the epitaxial layer is decreased as the pressure of the reaction chamber increases. The main reason for this is probably the pre-reaction of TMAl and NH_3. As the pressure of the reaction chamber increases, the pre-reaction of TMAl and NH_3 is strengthened and will result in a reduction of Al atoms reaching to the growth surface. The merging rate of Al atoms is also affected and Al content in material will be reduced. In addition, as the pre-reaction is strengthened, there will be more by-products, which cannot be taken away by the carrier gas. Some of the by-products will fall on the growth surface and affect the surface morphology and the crystal quality. The increase in pressure of the reaction chamber will decrease the diffusion ability of Al atoms, and so will affect the surface morphology and the crystal quality. In addition, at a condition of low V/III ratio and low pressure in the reaction chamber, the pre-reaction of TMAl and NH_3 will be suppressed and the surface diffusion ability will be increased, thus material quality and surface topography will be improved.

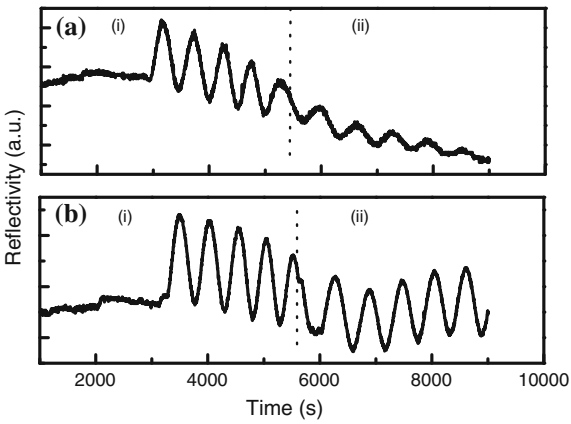


Fig. 4.44 The in situ monitoring curves of AlGaN samples

The influences of the inserted layers of trim ethyl indium (TMIn) and superlattice on AlGaN epitaxial are discussed in the following section.

During the growth process of Al(Ga) N, TMIn was usually used as surfactant to increase the surface migration rate of Al and Ga atoms and to release the stress. Material quality and surface morphology are improved in this way [94–99]. In this AlGaN growth experiment, TMIn was used as surfactant to increase the surface migration rate of Al. The material growth process is as follows: first, a low-temperature AlN buffer layer is grown on the (0001) face of the sapphire substrate. Then a high-temperature AlN buffer layer is grown. And then an AlN/AlGaN superlattice layer is grown. The AlGaN epitaxial layer is finally grown. The superlattice grown here is to decrease the dislocation density and improve the material quality [100, 101]. In order to study the influence of doping TMIn on the AlGaN layer, one of the samples has been doped with TMIn, for comparison, another sample was not doped. The two samples are set as A and B, respectively.

Figure 4.44 is the in situ monitoring curves for the two samples, while (a) is sample A without any doping, (b) is sample B doped with TMIn. The monitoring curve is divided into two processes as (i) and (ii), (i) for the complex buffer layer, including all layers below AlGaN and (ii) the growth process of the AlGaN epitaxial layer. It is found in the experiment that the two samples are different in the growth process (ii). The reflectivity in the doped TMIn curves increased, but for the sample without doping, the reflectivity attenuated totally. It shows that using TMIn as a dopant can improve the surface migration rate of Al atoms and improve the surface morphology. As shown in Fig. 4.45, there are also many small hexagonal islands on the surface of doped sample. But for the undoped sample, the surface is rough and the size of the hexagonal islands is large.

The difference of crystal quality is very large between the two samples. X-ray double-crystal diffraction is used to measure the crystal quality of two samples. It is found that for the sample without doping, half-width of (0002) is 16 arc-min and half-width of (10-12) is 18.7 arc-min. For the doped sample, half-width of (0002) is 10 arc-min and half-width of (10-12) is 16 arc-min. XRD data also shows that

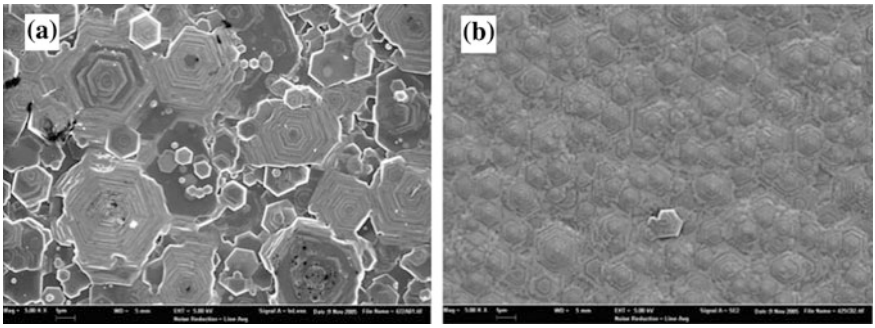


Fig. 4.45 SEM images of two AlGaN samples

material quality of the doped sample is better than that of the un-doped sample. It can be seen from surface morphology of the sample and the crystal quality that the TMIn doping will accelerate the lateral migration of Al atoms on the growth surface. When AlGaN is grown on a multilevel buffer layer, the surface morphology is improved by promoting the merging of the islands through an increase in the migration rate of Al atoms which was evidenced by increased surface reflectivity as observed in the monitoring curve. On the other hand, the dislocations will tilt or bend in the process of merging, which can reduce the dislocation density and improve the crystal quality of material accordingly.

In the experiment, TEM was used to study the filtering effect of superlattice to the dislocations. Figure 4.46 is the TEM images of sample A doped with TMIn, (a) and (b) are for diffraction vectors of (00002) and (10-10) respectively. Many dislocations were generated at the interface of sapphire and buffer layer. But the dislocation density falls sharply at 100 nm from the interface. During the dislocation penetration process from high-temperature AlN layer to superlattice, some dislocations disappeared at the interface, and some dislocations would bend and merge with other dislocations and be annihilated in the superlattice. And still some dislocations will extend from the superlattice to the AlGaN layer. It can be seen in Fig. 4.46b that edge dislocations disappeared in the superlattice. These two figures indicate the effect of superlattice for filtering dislocations.

The results show that an AlGaN/AlN superlattice can be used to filter the dislocations and improve the quality of the material. In the growth process of the epitaxial layer, the migration rate of Al atoms was effectively improved by the TMIn, thus surface morphology of the epitaxial and the material quality were improved.

It can be seen from the research results mentioned above that, the quality of AlN buffer layer must be further improved in order to grow high-quality AlGaN material. There is a similar situation in the growth of an AlN and AlGaN epitaxial layer. So it is necessary to find a new way to improve the quality of the material. Atomic Layer Epitaxy (ALE) was first used by Suntola and Antson for the deposition of ZnSe on glass [102]. Now this method is widely used in the growth of different compounds by different equipments [103]. Compared with ordinary MOCVD, many types of reaction sources are transferred into the reaction chamber

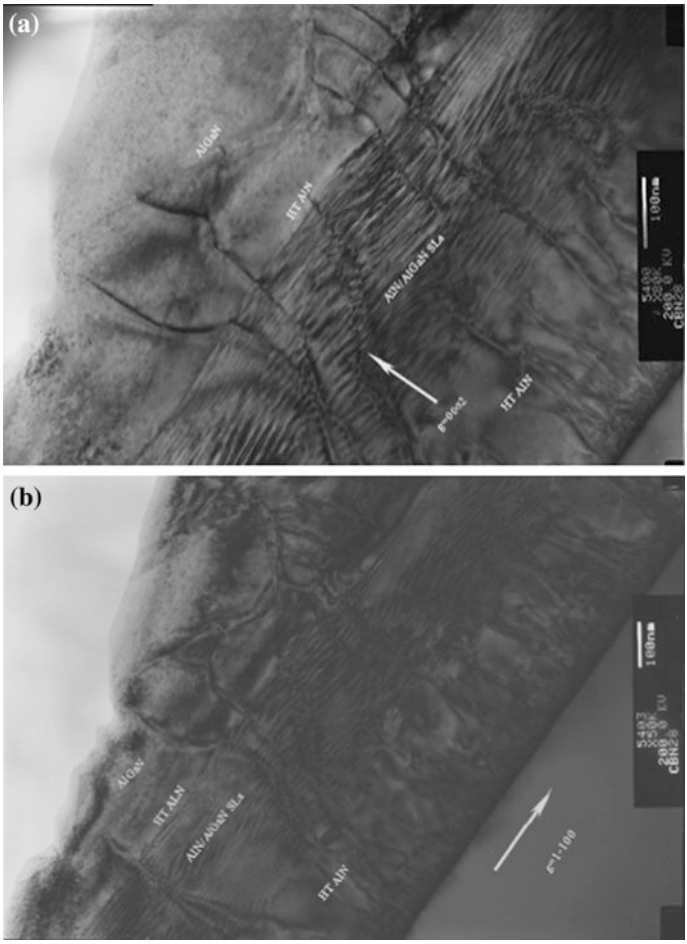


Fig. 4.46 TEM image of AlGaN sample doped with TMIn

in turn for ALE, but not at the same time. Atomic layer epitaxy reduces the rate of the reaction between the sources and increases the diffusion rate of the metal atoms on the growth surface. It is suitable for the growth of AlN and high Al components AlGaN [25, 103–105]. The quality of AlGaN can be improved by the growth of AlGaN on AlN which was finished by ALE first [89].

4.3.3 The P-Type Doping Technique of GaN Material

One of the problems that limit the development of GaN devices is the doping of P-type material. Mg is the most popular acceptor in P-GaN. But in MOCVD

growth, it is easy for Mg to form an Mg–H complex [106, 107], and Mg acceptors need to be motivated. Low-energy electron beam illumination (LEEBI) was first used to motivate Mg acceptors by Amano [22, 108], and achieved holes densities in GaN up to 10^{17} cm^{-3}, and a resistivity of 12 Ω cm. Then Nakamura used the method of fast annealing to motivate Mg acceptors [23, 109]. Nakamura also obtained the low-resistivity P-GaN. Chung used the multistep annealing method [110], where the thermal load was only half. This method can avoid the generation of N vacancies (V_N), thus avoiding the self-compensating effect. The hole density he achieved is $1–2\times10^{18}$ cm^{-3}, which was 50 % greater than the traditional single step method. The higher ionization energy ($\sim$200 meV) of Mg acceptors determines that there are only about 1 % that can be ionized [111] at room temperature, which limits the efficiency of Mg doping and causes the high contact resistance between P-GaN and metal. Another reason that limits the p doping is the compensation of background. As the generation energy of V_N in P-GaN is low, V_N can act as donors [112, 113] and cause the compensation of background. Furthermore, when the density of the doping Mg is high enough, V_N also can generate the complex ($Mg_{Ga}V_N$) with some of the Mg acceptors. The complex shows donors features and causes self-compensation [114, 115]. Yoshida also promoted another method, which is using co-doping to improve the doping efficiency of Mg in GaN [116, 117], but it was found that the growth window of efficient co-doping was too thin.

To understand the mechanism of doping and activation in P-GaN better, the changes of defects/impurities annealed in different temperature in GaN: Mg is now discussed. Figure 4.47a shows the PL spectra of GaN: Mg measured in room temperature before and after annealing for 20 min. The peak near 3.5 eV is the sideband photoluminescence peak of GaN. The wide ultraviolet peak near 3.2 eV (UVL) is considered to be the luminescence from the low acceptor–acceptor (Mg_{Ga}) to (DAP) [118]. After annealing in different temperatures, the intensity of luminescence declines by at least one order of magnitude and the peak moves to 3.27 eV gradually. The reason for peak's move is the density of carriers increases and their shield makes potential fluctuation weaker [119]. The luminescence intensity changes with the anneal temperature.

To analyze the changes of these peaks' intensity, Fig. 4.47b shows the respective intensity of the blue light peak to ultraviolet peak (BL/UVL) changes with different annealing temperatures. The BL/UVL ratio rises after annealing at 550°C then decreases with a rise (650–850 °C) of annealing temperature. Continued increasing of the annealing temperature to 950 °C causes the ratio of BL/UVL to increase rapidly. The changes of GaN: Mg PL peak with annealing temperature can be concluded to be due to the differences of defects/impurities that correspond with these illumination peaks in GaN: Mg. Before annealing, the density of shallow donors is high, resulting in an ultraviolet peak of a specific value. After annealing in N_2 at 550 °C for 20 min, the shallow donors are eliminated so the intensity of the ultraviolet peak decreases sharply. The fact that these shallow donors can be removed at such a low temperature as 550 °C means that, this type of shallow donor is easily moved and maybe interstitial H [106, 107, 114].

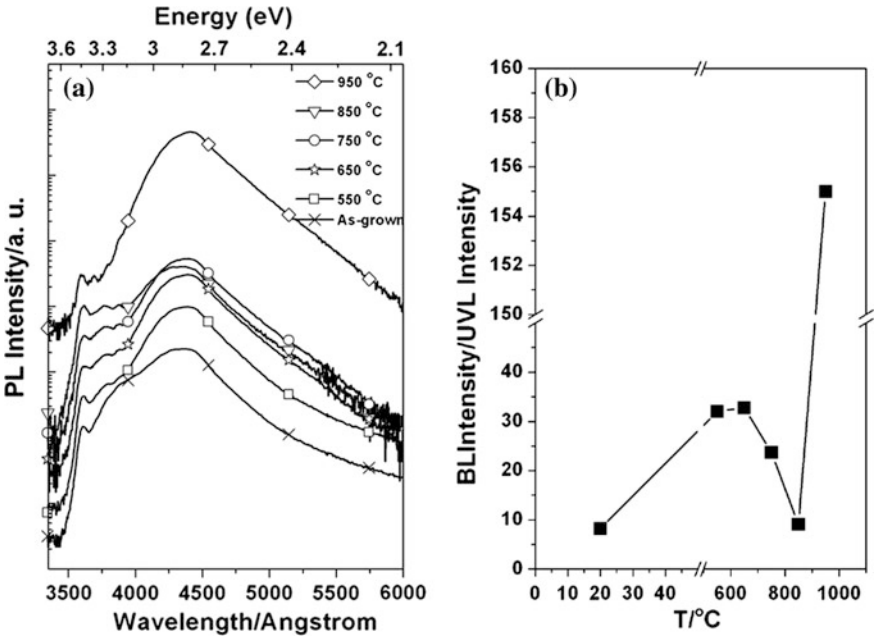


Fig. 4.47 The PL spectral measured of GaN: Mg before and after annealing (**a**) BL/UVL changes with anneal temperature (**b**)

Figure 4.48 shows the Hall testing result before and after annealing at different temperatures for 20 min. Before annealing, the GaN: Mg is P type, the annealing at 550 °C has little impact on the hole density and resistivity. As the annealing temperature changes from 550 to 850 °C, the density of holes rises from 10^{16} to 8×10^{17} cm^{-3} and the resistivity decreases from 9 to 0.8 Ω cm. But as the temperature goes higher to 950 °C, the resistivity rises to 4 Ω cm again. It can be seen that a suitable annealing temperature can motivate Mg acceptors effectively, although too high annealing temperature lowers the P doping level. Longer annealing time acts like higher annealing temperatures.

Figure 4.49 shows the PL spectra (a) and hole density and resistivity (b) of GaN: Mg grown with different flow rates of Cp_2Mg and annealed in 750 °C for 20 min. It can be seen from Fig. 4.49b that the density of holes rises with the increase of Mg doping density but continuing to increase the Mg doping results in the decrease of hole density. This is mainly because the self-compensation effects of Mg [114]. It can be seen from Fig. 4.49a that there are two peaks—UVL (strong) and BL (weak)-in the PL spectra when the sample is lightly doped although for heavy doping, the UVL peak disappears and the BL peak enhances. With the increase of Mg doping density, the BL peak increases continuously. Comparing (a) and (b) of Fig. 4.49, we also can find that the intensity of the blue peak has nothing to do with the density of holes. It only increases monotonically with the density of doping Mg. The blue peak can be considered as the recombination illumination from Mg's

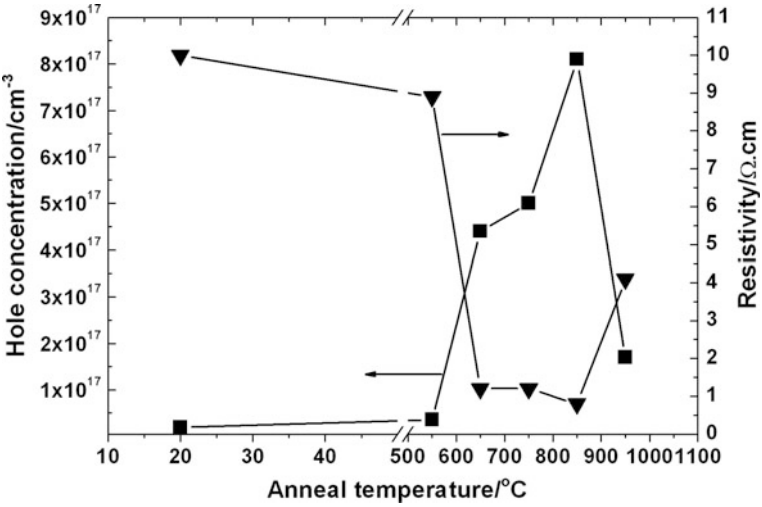


Fig. 4.48 Hole density and resistivity change with anneal temperature

self-compensation deep donors to Mg acceptors. These deep donors may be the complex of Mg_{Ga} and V_N. Because there are many V_N in P-GaN, and they are easy to move at the growth temperature about 1000 °C, and they also have the opposite charge with Mg_{Ga}. So they tend to combine with Mg_{Ga} and form Mg_{Ga} V_N.

After annealing at 550–850 °C, Mg_{Ga} V_N decomposes and the self-compensation becomes weaker, so the BL intensity becomes weaker with the increase of hole density, as shown in Figs. 4.47 and 4.48. During research, it is found that the Mg_{Ga} V_N often appears in the GaN: Mg grown by MOCVD, but it rarely appears in the GaN: Mg grown by MBE [114]. The difference between them is whether there is H in the growth environment. The GaN: Mg grown by MOCVD must be motivated to get a higher hole density. But the GaN: Mg grown by MBE, as there is no H, has better p-type conductivity without motivation. The result from theory [119] indicated that Mg_{Ga} V_N is stable only when its Fermi energy band is near the middle of the band gap, or it will decompose. This can explain the rest of photoelectric results: the original GaN: Mg grown by MOCVD has a very low hole density, and Fermi levels are close to the middle of the band gap therefore Mg_{Ga} V_N is stable. After annealing at 550–850 °C, the hole density rises, the Fermi level approaches the top of the valance band, and Mg_{Ga} V_N decomposes. When the annealing temperature rises from 850 to 950 °C, the high temperature leads to a large increase in the number of V_N. V_N compensations, the acceptors and the hole density declines and Mg_{Ga} V_N was formed again.

According to the results discussed above, the luminous mechanism of UVL for shallow donors–acceptors pair (DAP) and the mechanism of BL for deep donors-shallow acceptors pair (DDAP) can be expressed by energy band graph. It is shown in Fig. 4.50. According to the analysis above, the shallow donors

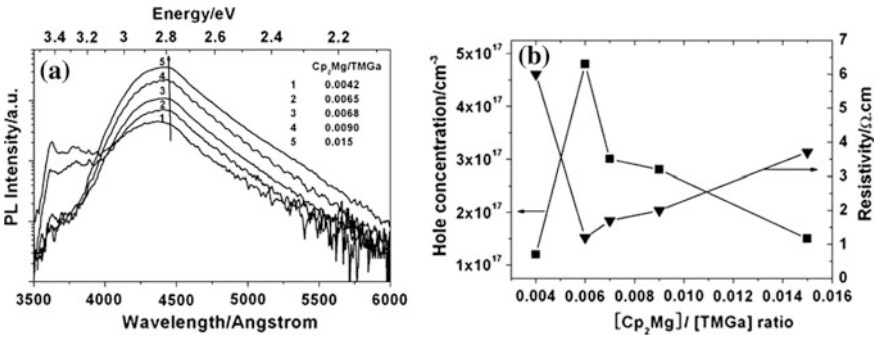


Fig. 4.49 The PL spectral of GaN: Mg with different Mg doping (**a**); hole density and resistivity (**b**)

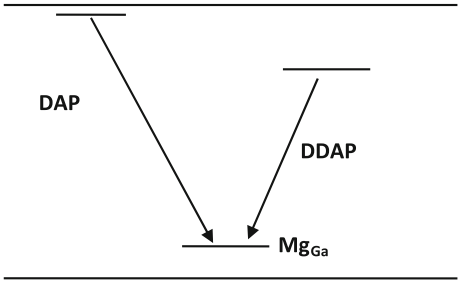


Fig. 4.50 The energy band of GaN: Mg, D is shallow acceptors-interstitial H, DD is deep donors-the complex of Mg_{Ga} and V_N

contributing to UVL can be considered as the interstitial H, and the deep donors contributing to blue peak can be considered as the complex of MgGa and V_N. Using this model, we can reasonably understand the experimental results. The results also indicated that, the hole density is the highest by annealing at 850 °C for 20 min and can reach 8×10^{17} cm^{-3}. But the resistivity is lowest of 0.8 Ω cm. The P doping level declines as the annealing temperature keeps rising.

The P-type doping of AlGaN is more difficult. The main reason is the ionization energy in AlGaN is much higher than that in the GaN, and the ionization energy increases with the component of Al rapidly. To improve the motivation efficiency of Mg, Nakamura found that replacing the uniform doping p-type AlGaN by a modulation-doped P-type AlGaN/GaN superlattice can improve the Mg ionization efficiency, thus improving the hole density [120]. The reason for this is that, under a polarization effect, there is strong built-in electric field in AlGaN/GaN, which makes the top of valance band looks zigzag. So some Mg impurity energy levels are below the Fermi level and then ionize. In spite of this, the experiment results of P doping in GaN are also helpful to the P doping of AlGaN. Furthermore, the new testing method of P-type density may have advantages to the P-type doping to GaN and AlGaN [121].

4.4 Overall Performance Analysis of AlGaN Material

To fabricate high-performance semiconductor devices, high-quality material is the base requirement. Especially for GaN-based devices, the material quality is particularly important. To grow high-quality material, the basic properties of the material must be understood in-depth, and research into the growth mechanism must be accomplished. This section addresses the material used to prepare GaN ultraviolet detectors, and introduces the testing and analysis of optical and electrical properties of GaN, AlGaN. This can be helpful to understand the MOVCD epitaxial growth of GaN, AlGaN.

4.4.1 Effects on Optical and Electrical Properties of GaN Material from Dislocations

Even though GaN material has been researched for many years, and GaN-based high-performance blue lasers, LEDs, ultraviolet detectors have been produced, but many fabrication problems have not been solved completely. In the fluorescence spectra of GaN material, there is a wide peak at 2.2–2.3 eV. No matter how good the quality of GaN material is, the spectral peak always exists. Because the spectrum is in the range of yellow light, it is known as "yellow peak". There are many views about the origin of the yellow peak [122–127]. For example, C impurity and Ga vacancies were once considered the sources of yellow peak. However, there is a common view that yellow peak comes from the transition glow from donor–acceptor pair [88, 128, 129]. In the following section, the source of the GaN yellow peak enhancement and the relationship between electrical properties and structural properties of the material are discussed.

Two series of samples are prepared to research the sources of the GaN yellow peak [130]. Figure 4.51 shows the PL spectral, which is normalized by sideband peak intensity, of three typical GaN samples in the first series near 3.42 eV. During the MOCVD growth of these samples, the samples were all lightly doped with the same density. In the PL spectra, a wider illumination band with a peak at 2.3 eV can be seen, and there are apparent interference fringes in the band. This illumination band is the yellow peak. The relative intensities (the ratio of yellow peak integral intensity to sideband peak illumination integral intensity, I_{YL}/I_{BE}) of samples A, B, and C are 2.14, 1.20, and 0.27, respectively. Obviously, the yellow peak of A is the highest, B takes second place and C is the lowest. The testing results of X-ray double-crystal diffraction (XRD) and Hall examination are shown in Table 4.4. It can be seen that, there are no differences between the half-height width of symmetric diffraction at (002) surface in the three samples. However, there are many differences between the half-height widths of skew symmetric diffraction at (102) surface. The half-height width of A is the biggest and that of C is the smallest. The (002) and (102) surfaces also can be considered as (0002), (1012). What can be

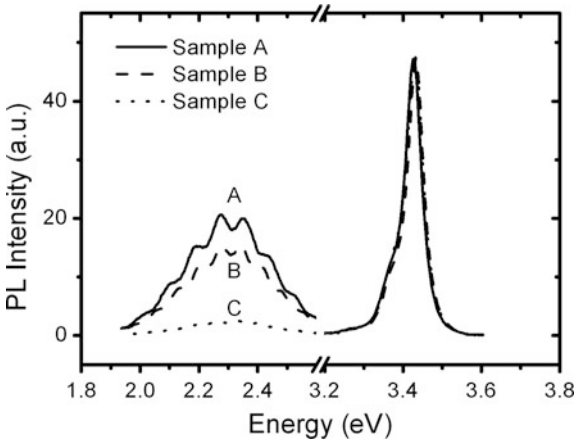


Fig. 4.51 The PL spectral being sideband normalized of three samples of A, B, C in room temperature. The tree simples have the same doping level of Si. The *yellow* peak intensity I_{YL}/I_{BE} of A, B, C is 2.14, 1.20, 0.27, respectively

concluded from these testing results of the samples is that, the intensity of yellow peak seems to correspond with the half-height width of (102), but has nothing to do with the half-height of (002). Because the XRD's half-height width of (002) and (102) represent the helical dislocation density and edge dislocation density of GaN, the bigger the half-height width is, the higher the dislocation density [54, 55]. In other words, the edge dislocation density in A is the highest and is the lowest in C, but the helical dislocation density of three samples has little difference. Combining the yellow peak intensity of three samples; it can be seen that the higher the edge dislocation density is, the higher the intensity of the yellow peak and vice versa. Therefore it appears that the intensity of yellow peak has nothing to do with the helical dislocations.

The rectangular solid points in Fig. 4.52 show the relationship between relative intensity of yellow peak and the XDR half-height width of (102) surface of the first series GaN material. It can be seen that as XRD half-height width of (102) surface increases, the relative yellow peak of GaN increases. We also find in the experiment that the intensity of yellow peak in this series of samples has nothing to do with the XRD half-height width of (002) surface. In other words, the intensity of yellow peak of GaN material is closely related to the edge dislocations, and has nothing to do with the helical locations. As the density of the edge dislocations increase, the yellow peak also increases.

To take a deeper look into the specific action of edge dislocations in the GaN material, the Hall testing results were combined. As shown in the Table 4.4, the carrier density of A, B, C are 8.8×10^{16} cm^{-3}, 2.1×10^{17} cm^{-3}, 2.6×10^{17} cm^{-3} respectively. The carrier density of A is the lowest, and the C is the highest. During the growth, these samples are Si-doped with the same density. It can be considered that the differences of carrier density are caused by the different compensation processes. The density of edge dislocations in samples A, B, C reduces successively. The density of edge dislocations may have large effects on the electrical properties. In Fig. 4.52, the hollow rectangular points represent the relationship

Table 4.4 Characterization results and Si doping condition of the first series n-type GaN sample

Sample	XRD FWHM (arcsec)		SiH_4 flux (nmol/min)	n (cm^{-3})	Mobility (cm^2/V s)	I_{YL}/I_{BE} (a.u.)
	(002)	(102)				
A	428	835	0.72	8.8×10^{16}	217	2.14
B	382	605	0.72	2.1×10^{17}	412	1.20
C	425	490	0.72	2.6×10^{17}	453	0.27

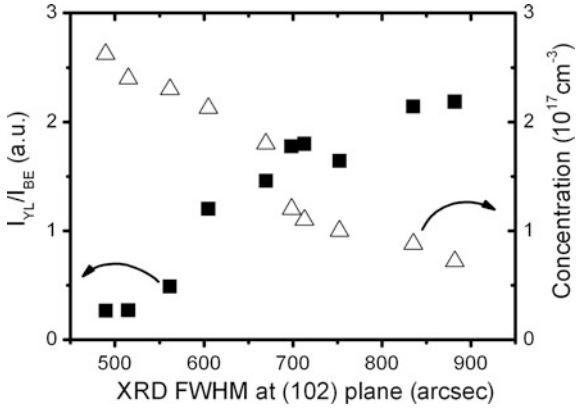


Fig. 4.52 The relationship of the relative yellow peak intensity (I_{YL}/I_{BE}) with the net carrier density and XRD half-height width of the first group GaN simples

between carrier density and the XRD half-height of (102) surface in this series of samples. It can be seen that as XRD half-height width increases, the density of carriers in GaN material decreases. In other words, as the density of edge dislocation increases, the density of carriers in GaN material decreases. Generally, there are many dangling bonds at the dislocation line of edge dislocations and these bonds may capture electrons from n-type material and produce acceptors [131, 132]. Intensive discussions have occurred about the action of edge dislocations. Some researchers even believe that the edge dislocations have no charge and were electric neutral [132, 133]. But there were experiments conducted which proved that the penetrating edge location line in GaN material has negative charges that might be caused by the environment around the dislocation line [134]. In fact, many researchers have proved that the dislocation lines act as scattering centers because they have a negative charge [135–137]. It can be shown that, as the edge dislocations can bring acceptors, so they can also bring compensation centers. So in the condition of different dislocation densities, the carrier density of GaN is different. The higher the edge dislocation density, the greater the compensation and the lower the carrier density of n-GaN.

So far, there is a common point of view that the yellow peak of GaN material comes from the transition of electrons from donors to acceptors. The average distance between donors and acceptors is important to the illumination efficiency. In n-type GaN, there are high numbers of positive charges gathering near the dislocation lines, which have negative charges [134]. As the distance between donors

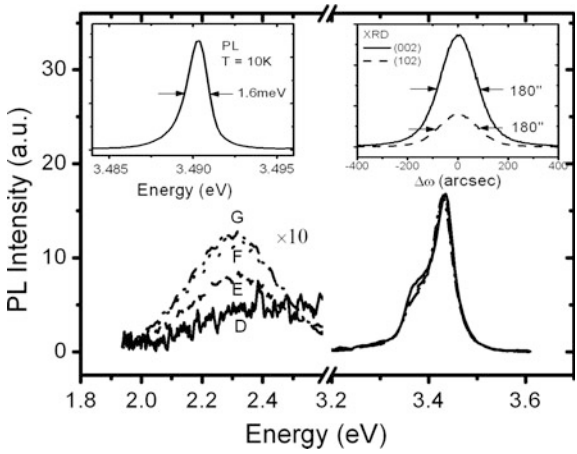


Fig. 4.53 The specific PL spectral of simple D, E, F, G at room temperature

and acceptors around the dislocation lines decreases, the wave function of electrons and holes overlap significantly, which increases the probability of transition. It is further indicated that the edge dislocations can enforce the yellow peak intensity of GaN material.

To further investigate the source of the yellow peak, a second series of samples was prepared. In these samples, the density of dislocation is almost the same, but the density of Si doping is different. The quality in series 2 is perfect and the XRD half-height width of (002) and (120) surfaces are only $\sim$180 arc-sec, which indicated that the dislocation of material is very small. Figure 4.53 shows the specific PL spectral in room temperature. Among them, the SiH_4 flows are 0, 0.05, 0.22, 0.45 nmol/min for sample D, E, F, G, respectively. The corresponding carrier density is 5.0×10^{15}, 1.5×10^{16}, 8.8×10^{16} and 1.8×10^{17} cm^{-3}. To see the relationship of yellow peak intensity and the Si doping density clearly, the spectra of yellow peak is amplified 10 times. The relative intensity of yellow peak (I_{YL}/I_{BE}) of D, E, F, and G is 0.15, 0.23, 0.32, and 0.35. Clearly, the yellow peak increases with the increase of Si density. It is known that Si acts as a donor in GaN material and can increase the electron density. So it can be inferred that Si may be one of the sources of yellow peak. The research results are in agreement with other reports [138–142]. It should be emphasized that there are many sources of donors that can lead to a yellow peak, such as O impurity, N impurity, etc. The cause of the yellow peak is a combined action of all these factors [143]. As the material quality of this group is very high, the XRD half-height width is narrow and the background density is very low. The background density is 5.0×10^{15} cm^{-3} without doping. The illustrations in Fig. 4.53 include the PL spectra of G at 10 K, the ω scanned rocking curve of XRD at (002) and (102) surfaces. These prove that these groups of samples are of high quality. The high quality of the samples makes it easy to observe the enhancement caused by Si doping to yellow peak. It can be seen from the above results that even though there are many sources for the yellow peak in

GaN, the transition from Si to the acceptors introduced by edge dislocations can enhance the intensity of yellow peak.

Among them, the SiH_4 flows are 0, 0.05, 0.22, 0.45 mol/min for sample D, E, F, G respectively. To study the relationship of yellow peak intensity and the Si doping density clearly, the spectra of yellow peak are amplified for 10 times. The illustrations indicate the PL spectral of G at 10 K, the ω scanned rocking curve of XRD at (002) and (102) surfaces.

See Table 4.4 again, where the mobility of A, B, and C is 212, 412, and 453 cm^2/V s, respectively, but the XRD half-height width of A, B, C at (102) surface is 835, 605, 490 arc-sec. It can be seen that the mobility of A is the lowest; the XRD half-height is widest. The mobility of C is the biggest, and the XRD half-height is narrowest. As the testing results of the three samples show, it seems the mobility of GaN is related to the XRD half-height width at (102) surface. The bigger the XRD half-height width the sample, the less the mobility and it has nothing to do with the XRD half-height width at (002) surface. As the XRD half-height width at (002) surface and (102) surface represent the helical dislocation density and edge dislocation density, it is found that the edge dislocation density of A is the highest but its mobility is the least. The edge dislocation density of C is the least, and its mobility is the highest. It seems the carrier mobility is related to the edge dislocation density. Of course, it is also noticed that the mobility seems to have nothing to do with the helical dislocation.

Figure 4.54 shows the relationship between XRD half-height width at (102) surface and the carrier mobility of the first group of GaN samples. It can be seen that as the XRD half-height width at the (102) surface increases, the mobility of GaN decreases. We also find that, the XRD half-height width at (002) surface has nothing to do with the mobility of this series of samples. In other words, as seen in the testing results of this series of samples, carrier mobility is related to the edge dislocations in GaN material, but is irrelevant to the helical dislocations. As the density of the edge dislocation increases, the mobility of carriers declines. We know from the results obtained from Fig. 4.52 that, there are many dangling bands at the dislocation line of edge dislocations. The dangling bands will introduce acceptors and then have negative charges. At the same time, they introduce the scattering

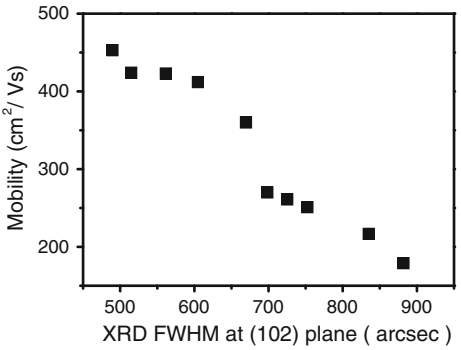


Fig. 4.54 The relationship between carrier mobility and XDR half-height width at (102) surface of the first group of GaN samples

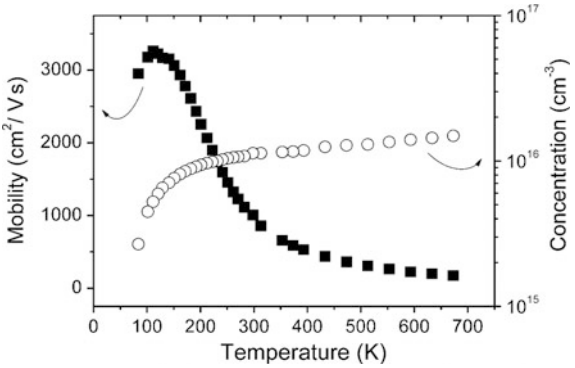


Fig. 4.55 The Hall testing results of high-quality GaN samples at different temperature, the *solid points* represent the carrier mobility and the *hollow points* represent the carrier density. It can be seen from the testing results that, the mobility in temperature is as high as 1005 cm^2/V s, the mobility of carriers is $1.1 \times 10^{16} cm^{-3}$

centers. This process shows that GaN with different mobility will have different density of dislocations. In other words, the higher the number of edge dislocations, the more scattering centers, and the lower the mobility is of the carriers. The edge dislocations make an importance influence on the electrical properties of the GaN material. A further study found that the edge dislocations also affect the blue peak (another deep energy level illumination peak) of GaN [144]. Furthermore, the point dislocations may make effects to the carrier density of GaN [145].

Due to the discovery of the important roles the edge dislocations play in the GaN materials, especially the scattering effects of the edge locations, intensive research was conducted into the MOCVD growing mechanism of GaN and the special epitaxial methods. This research found that as the density of edge dislocations is sharply decreased, the XRD half-height width at (002) and (102) surface of the samples are all 180 arc-sec. Figure 4.55 shows the Hall curve at different temperature of high-quality GaN samples. The testing results in the figure shows that the mobility of that sample at room temperature is 1005 cm^2/V s, and the background carrier density is about 1.1×10^{16} cm^{-3}. It is one of the best results reported in the current literature [146]. It is also found that Ga vacancies dislocations can reduce the minority diffusion distance in GaN [147], and then affect the performance of GaN ultraviolet detectors [148, 149]. As a result of this research, it is proposed that, during the growth of material, the generation of Ga vacancies should be reduced.

4.4.2 *Measurement of Al Components in AlGaN and Determination of Its Strain State*

Measuring Al components and the strain state accurately is extremely important to the application and understanding of its physical properties. The PL spectral is one of the shortcut methods, but it cannot eliminate the influences caused by strain [150]. However, X-ray diffraction can measure the component and strain together

[145]. Generally speaking, the Vegard theorem can be used to calculate the component regardless of the relaxation or complete elastic states [151, 152] in epitaxial layer. If there are partial relaxations in the AlGaN epitaxial layer, the epitaxial index must be specified and then the Al component can be calculated. Using Rutherford backscattering (RBS) to measure the Al component can avoid the influence caused by strain and stain relaxation [153]. But measuring every sample with RBS is not practical. So if the relaxation is known, then the component can be quickly calculated from the X-ray diffraction results. On the other hand, the relationship between relaxation and component is helpful for understanding material growth. In the following section, the Al component was measured with RBS, and the lattice parameters in $Al_xGa_{1-x}N$ plate were measured accurately with a tricrystal X-ray diffraction method. The relaxation index T was calculated, and the relationship between relaxation state and component of AlGaN with a thickness of 570 nm was confirmed [154].

To measure the strain of three-element complex with tricrystal X-ray diffraction, the direct way is to measure the lattice parameters. But in the actual epitaxial films, the lattice parameter at complete relaxation a_0 must be constrained to confirm the strain state. So the component should be calculated with the Vegard theorem. But the Vegard theorem is limited to the strain. If the sample is in the double-axis strain state, the component calculated by Vegard theorem using an axis lattice parameter is larger and the component calculated is smaller when using the c axis lattice parameter. If the relaxation of samples is an elastic process, then it can be solved using the equations below.

Let the lattice parameters of the complete relax state $Al_xGa_{1-x}N$ be a_0 and c_0, and the measured value is a and c, then:

$$\begin{cases} \left(\dfrac{a_0 - a}{a_0}\right) \times \left(\dfrac{c_0}{c_0 - c}\right) = \dfrac{C_{33}}{2C_{13}} \\ a_{\mathrm{AlN}} + (1-x)a_{\mathrm{GaN}} = a_0 \\ c_{\mathrm{AlN}} + (1-x)c_{\mathrm{GaN}} = c_0 \end{cases} \tag{4.17}$$

Among them, C_{33} and C_{13} are the elastic constants of $Al_xGa_{1-x}N$, and can be calculated from the elasticity of GaN and AlN by linear interpolation. As there are great differences between the elastic constants of GaN and AlN [155], the component calculated in this way also have some fluctuation.

The $Al_xGa_{1-x}N$ material investigated is grown by MOCVD on a GaN epitaxial layer. The component of Al changes from 0.22 to 1. At first, the component of each sample was measured by RBS. Sample D is AlN grown on GaN directly, so it is not measured by RBS. Figure 4.56a–c show the RBS random spectra and channel spectra of samples A, B, and C. The labels of Ga, Al, and N represent the He^+ energy scatted by Ga, Al, and N atoms at the surface. The Al component of A, B, and C is simulated by random spectra of 0.22, 0.27, and 0.42, respectively. The thickness of $Al_xGa_{1-x}N$ and GaN is 570 and 600 nm. The minimum yields obtained

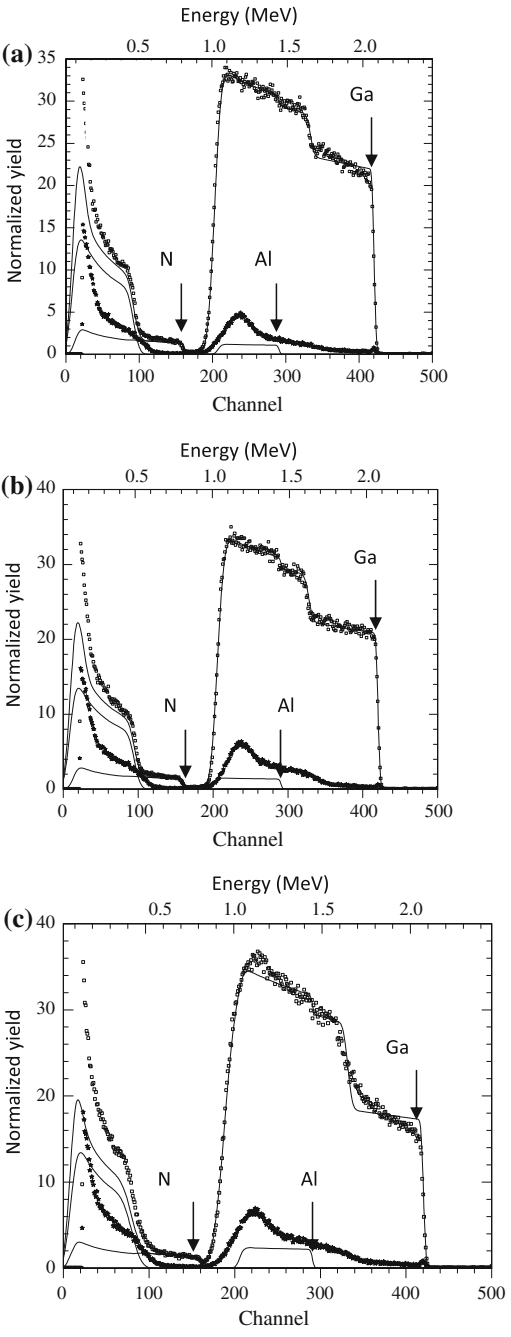


Fig. 4.56 **a–c** represent the RBS random spectral and channel spectral of A, B and C

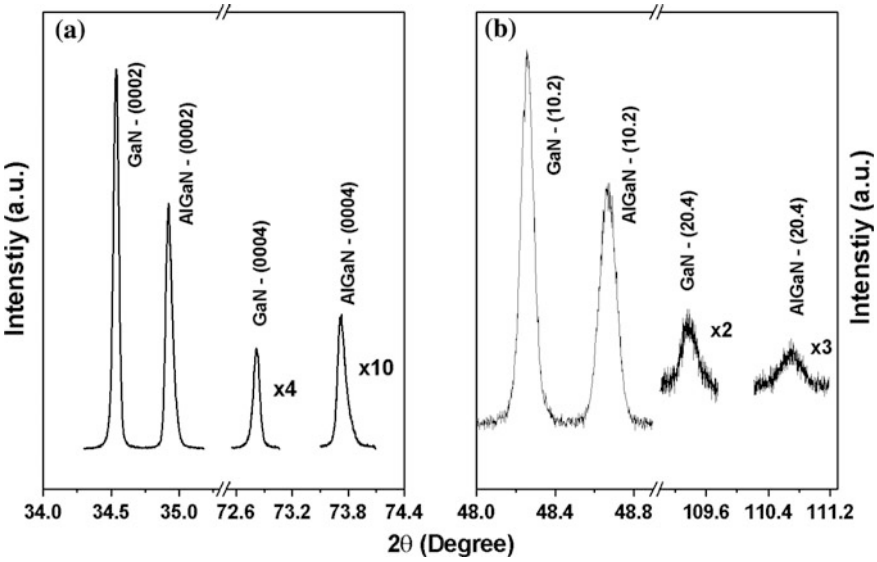


Fig. 4.57 **a** and **b**, A typical TAXRD curve (simple A) of measuring lattice parameters

from the Ga atoms' ⟨0001⟩ channels are listed at Table 4.5. The results shows that, the crystal quality of A and B is better than C.

Here $\omega/2\theta$ scanning of (0002) and (0004) were used to calculate the lattice parameter c, and the diffraction of (10-12) and (20-24) face were used to calculate the lattice parameter a [156]. Figure 4.57a, b show a typical TAXRD curve (simple A) of the measured lattice parameter. The diffractions of GaN and AlGaN separate clearly with each other. Every low level diffraction peak is narrow and sharp (the peaks of other samples are similar). This fact indicates that the quality of the epitaxial layer is good. The lattice parameters of $Al_xGa_{1-x}N$ and GaN are calculated and listed in Table 4.5. It can be seen that all the GaN layers are in the state of strain compared with the body material. Also shown in the table are the density of edge dislocations and helical dislocations calculated from the half-height width of bicrystals scanned at (0002) and (10-12). Figure 4.58 shows the (0004) triaxial crystal $\omega/2\theta$ scan. The curve of D is not listed because D had many surface cracks which may influence the shape of diffraction. It can be seen from the figure that all the diffraction peaks are asymmetric, and they are different from each other. The tails of A and B are at the left side of the diffraction peak, while the tail of C is at the right side. Furthermore, the tail at A is higher and longer that the tail of B.

The relaxation of extension R can be calculated with the formula (4.18) below:

$$R = \frac{a_s - a}{a_s - a_0} \tag{4.18}$$

Table 4.5 The structure features of $Al_xGa_{1-x}N$epitaxial layer and the lattice parameters of GaN substrate

Samples	a_{AlGaN} (Å)	c_{AlGaN} (Å)	a_{GaN} (Å)	χ_{min} (%)	Helical dislocations (10^8 cm^{-2})	Edge dislocations ($10^8 cm^{-2}$)
A	3.179	5.141	3.180	1.7	2.3	73.2
B	3.179	5.134	3.183	1.6	2.9	78.0
C	3.165	5.138	3.182	2.7	4.2	104.8
D	3.113	4.983	3.182	–	3.5	125.2

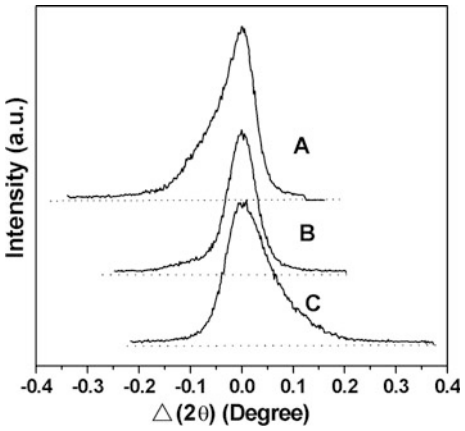


Fig. 4.58 The (0040) triaxle crystal $\omega/2\theta$ scan of simple A, B and C

In the formula, a_s and a_0 are the inner plane lattice parameters of substrate GaN and complete relaxed $Al_xGa_{1-x}N$ epitaxial layer and "a" is the measured value of $Al_xGa_{1-x}N$. Figure 4.59 shows the changing of R with the Al component. When the component is less than 0.42, R increases linearly with the Al component and reaches 70 % when the component is 0.42 and then changes slowly. When the component is 1 (AlN), the epitaxial layer is totally relaxed.

It can be seen from Table 4.5 that, all the GaN substrates are at strain state. As the lattice parameters become small, the lattice mismatch between GaN and AlGaN decreases, and the critical thickness of AlGaN increases. It is known by extrapolation that the $Al_xGa_{1-x}N$ epitaxial layer with Al component of 0.16 and with a thickness of 570 nm can be coherently grown on the GaN layer. Combining Fig. 4.59 and Table 4.5, it can be seen that R increases with the increase intensity of edge dislocations. It indicates that the edge dislocations (including the edge component in mixed dislocations) play an important role in the strain relaxation of $Al_xGa_{1-x}N$. In sample D, the cracks at the surface also have great importance on releasing the strain in addition to the dislocations.

During the growth of $Al_xGa_{1-x}N$, there may be a very thin component AlGaN transition layer between GaN and $Al_xGa_{1-x}N$. The heterogeneity of the Al

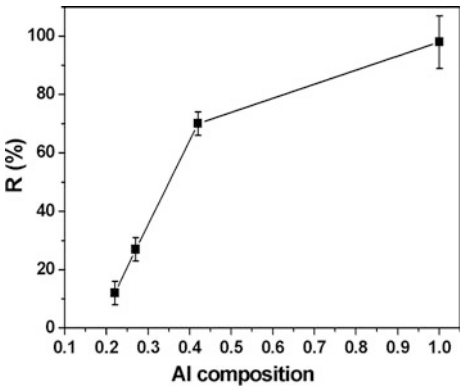


Fig. 4.59 The relaxation *R* changes with the Al component

component may cause the asymmetry of $\omega/2\theta$ diffraction peak and form the tail. But in this experiment, as the Al component increases, the tail goes from the left to right of the diffraction gradually, so perhaps there is another mechanism that should be taken into account. It can be seen from Fig. 4.59 that the differences of relaxation is one reason that causes asymmetrical diffraction. When the $Al_xGa_{1-x}N$ coherency grows on GaN, it is in a state of tensile strain. The lattice parameter c is less than the relaxed material. So the distance between the GaN and $Al_xGa_{1-x}N$ (0004) diffraction peaks is greater than that of relaxed material. As the relaxation increases, the lattice parameter c of the relaxed $Al_xGa_{1-x}N$ increases, which causes the (0004) diffraction peak asymmetry and the tail is at left side of the peak. As the relaxation increases, the proportion of relaxed AlGaN increases. That causes the tail to move from left to right, as shown in Fig. 4.58. So we can consider that, when $R = 50$ %, (0004) the peak is symmetric. The cracks in sample D are caused by the greater lattice mismatch between GaN and AlN.

Even if the relaxation is gradual, most of the relaxation is at a very thin layer near the GaN and $Al_xGa_{1-x}N$ interface. In other locations, the relaxation index *R* changes little. This can be seen from Fig. 4.57. Even though there is a small tail at (0002) and (0004) diffraction peaks, the main peak is narrow and strong, indicating that most of the epitaxial layer is uniform. So the relaxation index we determined still reflects the state of relaxation state of the $Al_xGa_{1-x}N$ epitaxial layer.

The results above show that, when the component is less than 0.42, the relaxation of $Al_xGa_{1-x}N$ increases linearly with the increase of component, and reached 70 % when the component is 0.42. The relaxation increases slowly when the component is greater than 0.42 and is totally relaxed when the component is 1. The changes of relaxation cause the asymmetry and shape changes of (0004) $\omega/2\theta$ scan peak. In this experiment, since GaN is in a state of stress strain, the mismatches between the $Al_xGa_{1-x}N$ and GaN substrate decrease. This allows the $Al_xGa_{1-x}N$ with same thickness and high Al component (or same component and greater thickness) to grow coherently.

4.4.3 *Uniformity of AlGaN with High Al Component*

Uniformly varying alloy semiconductor material can lead to nonuniform spatial distribution of band gap, and thus will affect the optical and electrical properties of materials, so research of the uniformity of ternary nitride material is of great significance. There have been some achievements on AlGaN with high Al component [100, 101], and here investigation of uniformity of AlGaN with high Al component is discussed as follows [102].

An AlGaN sample is prepared by MOCVD growth, and the process is as follows: First, an AIN buffer layer is grown at a low temperature, and then an AlN epitaxial layer is grown at a high temperature, finally AlGaN epitaxial layers with Al constituent about 40 % are formed. Figure 4.60a is a double-crystal ω scan of the sample. It can be seen that there is a stretch on the right side of the diffraction peak, which could be Gauss fitted for two peaks. There are two possible reasons: one is probably that there is a two-layer AlGaN with different strain states [242], and another is that the component of AlGaN is nonuniform. So as to distinguish between these two cases, we use low temperature (6 K) spatially resolved CL to analysis the sample. First, we made a wide range spectrum of the sample as shown in Fig. 4.60b. There is also a stretch in the spectral line on the low-energy side, which could be fitted by two Gaussian peaks. In order to offer an explanation to the origin of these two peaks, we made a CL image using the two peak energy as shown in Fig. 4.61.

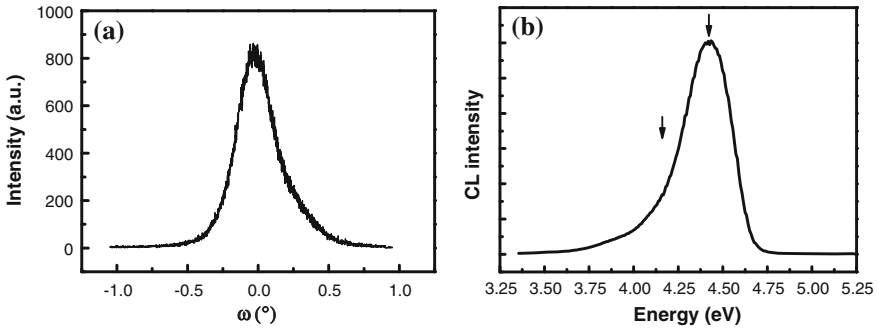


Fig. 4.60 XRD rocking curve of AlGaN (**a**) and the CL spectrum (**b**)

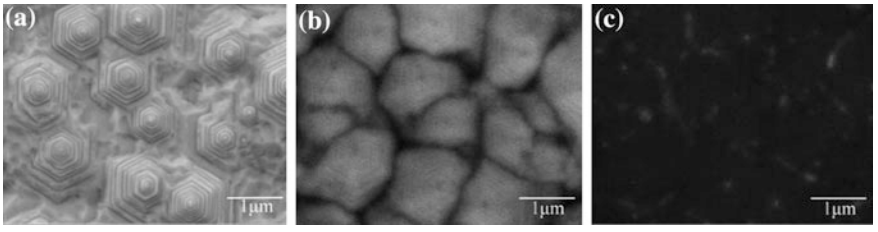


Fig. 4.61 SEM image of the sample (**a**) and CL image of photon energy corresponding to different regions (**b**) 4.42 eV; (**c**) 4.12 eV

Figure 4.61a is the SEM image of the AlGaN sample with a hexagonal-island surface. Figure 4.61b, c are CL images of 4.42 and 4.12 eV respectively at the same position of Fig. 4.61a. Figure 4.61b is composed of pieces of bright regions which are separated by a black border. If the region surrounded by the black border is considered a domain and the black border considered to be a domain boundary, we could find that every domain corresponds to a hexagonal island in Fig. 4.61a, which is just slightly larger than the size of the island. A hexagonal step that is barely able to be seen within the domain is shown in the SEM image. But there are a large number of black dots in the middle and above the steps of the domain, which could be seen in the same position corresponding to Fig. 4.61c. These black dots could be due to dislocations or defects. In addition, the other places are complemental in light and shade except for the black dots shown in Fig. 4.61b, c, which indicates that the AlGaN material used in the experiment is composed of two different Al components except the region of defects.

Furthermore, we use a CL line scan to analyze the distribution of the components in the horizontal direction. As the white line shows in Fig. 4.62a, we scan every 200 nm along the line from left to right and obtain the spectral data in Fig. 4.62b. It can be concluded from the series of measured spectral curves that the peak energy emitted from the island is larger than that between the islands, and the luminous intensity of the island is stronger than the mesa. Luminous intensity links directly to the component, which indicates that Al component of the islands is higher than that of between the islands. While luminous intensity is closely related to defects, a stronger luminous intensity of an island is because of the presence of defects between the islands leading to non-luminescence and decreasing of the emission intensity. Usually in the process of merging, many dislocations will generate at the boundary of the island to release a large amount of stress.

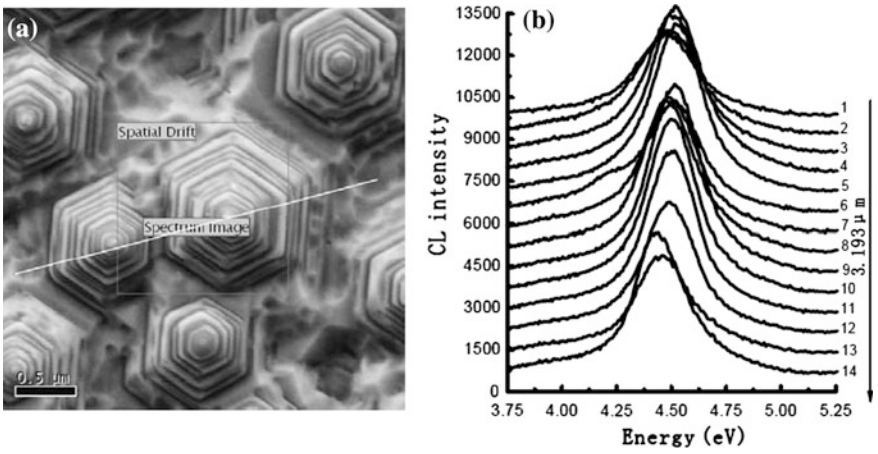


Fig. 4.62 **a** SEM image of the sample, **b** CL Spectrum along the white line of Fig. 4.63a from *left* to *right*

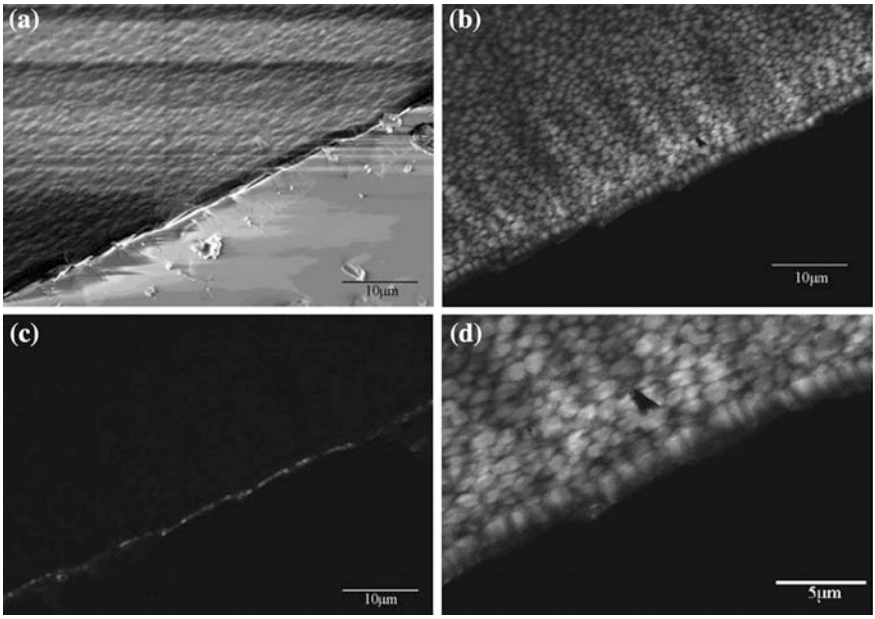


Fig. 4.63 Samples section of energy-selection CL images **a** Cross-sectional view of SEM, **b**, **c** CL images of 4.42 and 3.75 eV respectively **d** Enlarged view of (**b**)

Nonuniformity of the components can be divided into horizontal and vertical directions. Figures 4.61 and 4.62 have shown component unevenness of AlGaN materials in the horizontal direction. In order to study uniformity of the constituent in the vertical direction (the material growth direction), a cross-sectional image was taken of the AlGaN samples with different energy and at an oblique angle between the cross section and electron beam to collect the information from both the surface and the cross section. Figure 4.63a is a cross-sectional SEM photograph of the sample, while Fig. 4.63b, c are respectively at 4.42 and 3.75 eV, and Fig. 4.63d is an enlarged partial area of Fig. 4.63b. Corresponding to the image of the structure of the material and the thickness of each layer, we find that Fig. 4.63b is a view corresponding to the AlGaN epitaxial layer, and Fig. 4.63c is the interface between the AlN layer and AlGaN. First, we can see from Fig. 4.63b that the luminous intensity of AlGaN in the longitudinal layer is relatively uniform, indicating that the Al components are uniform in the longitudinal direction. From a further lateral view, it shows that the light-emitting region of the AlGaN layer is composed of configurations with columnar structure. While the columnar boundaries are black, indicating a relatively low component or more defects in the columnar boundaries.

Considering the component unevenness of AlGaN from the growth mechanism, it is clear to see from Fig. 4.63d a completely columnar growth of AlGaN, with the columnar size about 1–2 μm. From the surface topography of Fig. 4.63a, the sample surface is composed of hexagonal layered islands as well as some border

strip pits. In conjunction with Fig. 4.61, the size of domain boundary in the CL image is larger than that of the islands in the SEM image, indicating an island growth process of AlGaN. As the growth continues, the size of the islands reduces but the components remain unchanged. As we can see from Fig. 4.64, a possible growth of AlGaN can be inferred from these patterns. A low-temperature AlN buffer layer is grown on the sapphire substrate and nucleation begins (Fig. 4.64a). Following the annealing process, the nucleation centers begin to grow and the nucleation density decreases, and gradually they begin to merge (Fig. 4.64b). Then the growth begins to be quasi-two-dimensional (Fig. 4.64c). However, due to the low migration rate of Al atoms in the surface, the subsequent growth of AlN layer changes to a three-dimensional model gradually. And the surface begins to show a large number of islands with a relatively large size (Fig. 4.64d). When an AlN layer is subsequently grown on the AlGaN epitaxial layer, the reflectivity of in situ monitoring curve increases due to the consolidation process of the large islands (Fig. 4.64e), and then columnar growth begins. The non-uniformity of the Al component should begin with the merger of the big islands. Prior to the growth of AlGaN, the surface is composed of many islands, and when Al atoms and Ga atoms fall on the surface, due to a larger ratio of surface migration of the Ga atom than Al atom, Ga atoms migrate more easily than Al atoms to the island border. Due to the lower rate of surface migration, most of the Al atoms could not find a location with suitable energy and piled on top of the island, showing columnar growth (Fig. 4.64f), thus causing uneven components and columnar of the light-emitting region of AlGaN layer.

A low-energy light-emitting layer appears in Fig. 4.63c, for several reasons: one is that this is part of the AlGaN and AlN interface, due to the large lattice mismatch between the two, misfit dislocations also occur at the interface. Dislocations in the lower buffer layer penetrate to the top. These dislocations will not only reduce the emission intensity, but may also affect the incorporation of Al atoms. Another reason may be that before the growth of AlGaN epitaxial layer, due to a very low surface migration rate of Al atoms, the high-temperature AlN layer surface is very rough, leaving many hexagonal islands. When growth begins, Ga atoms fill faster to the island border, leading to a lower Al content of AlGaN layer on the interface. It was found that this phenomenon may be the result of interaction between the two. What can also be seen from the above results is that component uniformity of AlGaN material grown on the AlN layer is much better and more suitable for the preparation of the device, compared to the AlGaN material on a low-temperature GaN buffer layer [157].

4.4.4 Oxidation of AlGaN Materials

Oxidation of nitride on the surface has a very important influence for surface treatment and preparation of the device. It will cause the carrier recombination via surface states and will create anomalies in the device caused by the Fermi level

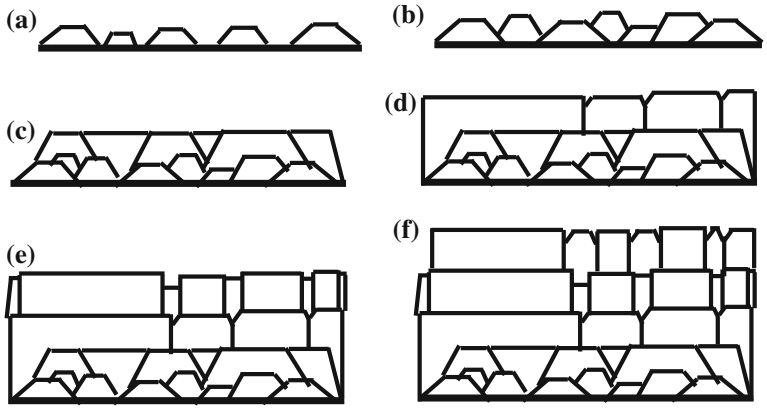


Fig. 4.64 A possible AlGaN growth pattern

pinning at the surface. There are many reports for the oxidation of GaN materials [158, 159], but few for AlGaN ternary alloy material [160]. Easily oxidized Al has a greater influence on the surface characteristics of the AlGaN material and the subsequent device performance compared to GaN. The oxidation properties of AlGaN materials are described in the following. [161].

Two AlGaN samples were prepared with different Al components on the GaN layer, and the thickness is about 400 nm. The Al components of samples A and B were 8 and 12 %, respectively. First, a small piece of AlGaN sample was cut to test for surface O content using XPS, then photoluminescence testing for the two samples and atomic force microscopy (AFM) was used to observe the surface. After that, the samples were placed in air for photoluminescence test and atomic force microscopy observation every ten days. Finally, XPS, photoluminescence and atomic force microscopy were used to test the samples after forty days.

First, the optical properties of AlGaN materials were investigated. The results of the photoluminescence of the samples obtained have been listed in Fig. 4.65. The illustration in Fig. 4.65 is one PL spectra with three peaks. They are the GaN emission peak, the AlGaN emission peak and the laser peak respectively from the long wave to short wave. The measured AlGaN peak is normalized by the peak intensity of GaN, and then it was found by linear fitting that the AlGaN sample emission intensity decreases with time in the air, and luminous intensity reduces for samples with a high Al component.

AFM surface topography was used to test the sample, and determine the trend for surface topography of the two samples as a function of air exposure time. Only the morphology of sample A is selected to be explained. Figure 4.66a–d are photos of the surface topography test as a function of the time when sample A was grown and placed immediately in air and every ten days, respectively. The surface topography of the grown sample is smooth with a step flow. While there are many irregular spots in the stepped surface of the table placed in the air for ten days and

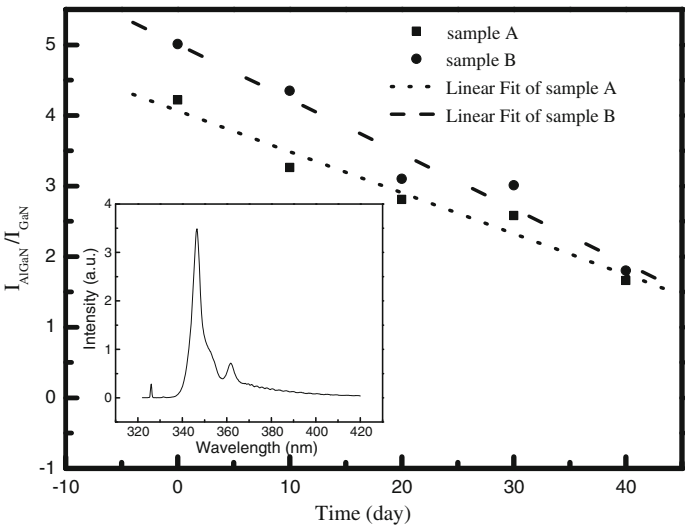


Fig. 4.65 The tendency of optical properties of AlGaN sample placed in the air as a function of time

the surface becoming rougher, no phenomenon was observed at the step gully. After twenty days placed in the air, irregular spots of the surface disappear almost entirely, and step flow on the surface topography was also blurred, although the surface roughness is smaller. Figure 4.66d is surface topography for the sample placed in the air after thirty days. The surface undertook a relatively large change, as is characterized in the steps of the flow, but the surface roughness becomes large. The results of the surface roughness of the samples are listed in Fig. 4.67. The trend for surface roughness of the two samples is the same, with both becoming rougher when placed in the air at the beginning, then a decrease of roughness, then rougher again to a certain extent.

The chemical property for the surface of AlGaN is also studied. Photoelectron spectroscopy (XPS) is a convenient means of surface measurements and can detect the surface chemical component, constituents, band bending, etc. For better comparison, an AlGaN sample C (identical to the growth conditions of A) was placed in a vacuum cavity of XPS test with samples A, B. There is a time of 10 h for sample C to finish native oxiding. The main consideration is the sample surface oxidation, and therefore sample surface were tested with energy level of Al 2p, O 1s, Ga 2p, N 1s. The specific data is shown in Fig. 4.68. For the energy level of Al 2p, binding energy of sample B is 73.15, and 73.4 eV for sample A, however 73.6 eV for sample C. Normal binding energy of Al–O is 72.9 and 74.4 eV for Al–N. It can be clearly seen in photoelectron spectroscopy that the Al–N moves to Al–O. The same phenomenon was observed in the curve of Ga 2p and N 1s energy level. Moreover, when it comes to the curves of the O 1s energy level, the peak intensity of sample C is very weak, while those of the other two samples are relatively strong.

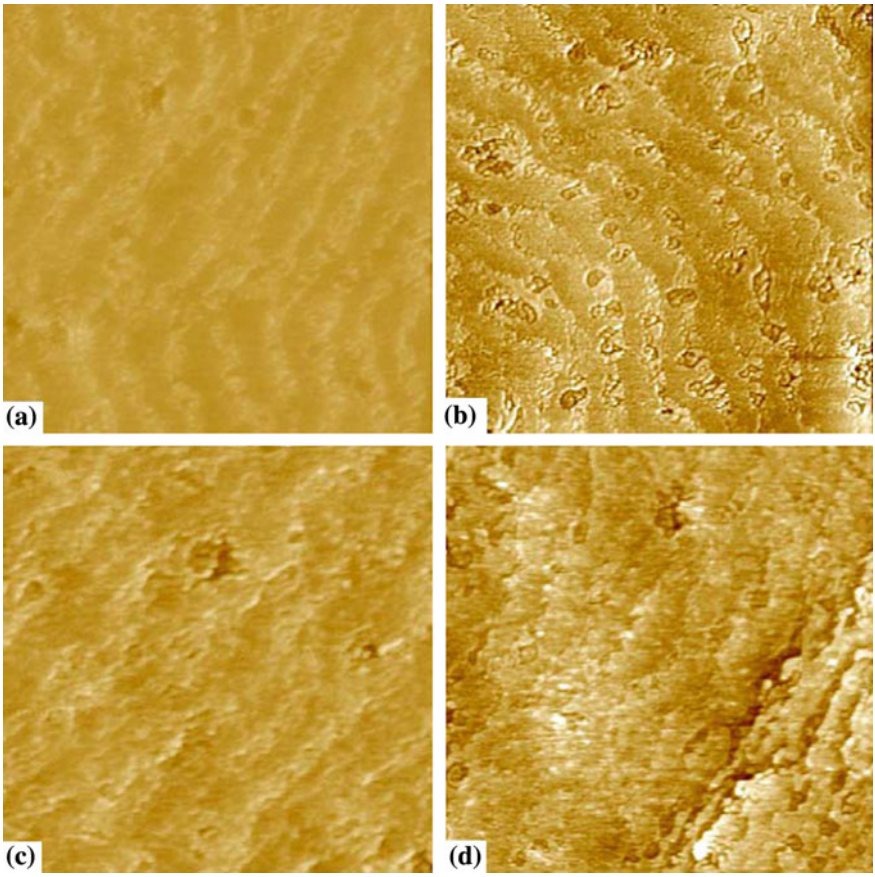


Fig. 4.66 AFM photographs of AlGaN sample placed in air with different times

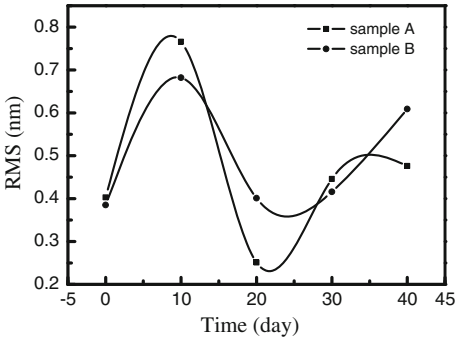


Fig. 4.67 Surface roughness of the AlGaN samples placed in the air with time

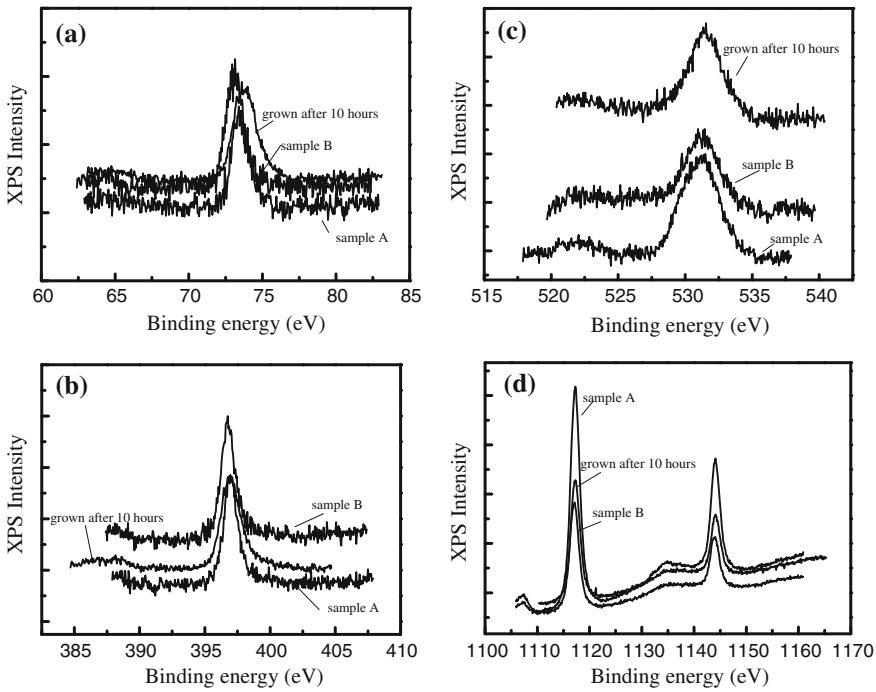


Fig. 4.68 XPS test of AlGaN sample before and after placed in the air. **a** Al 2p, **b** N 1s, **c** O 1s, **d** Ga 2p

A conclusion that can be drawn from these phenomena is that the surface of AlGaN material placed in the air is covered with a layer of oxide film, and oxidation of AlGaN will be severe with a longer air exposure time.

After a closer observation of samples A and B, it is found that they have different binding energy positions. Chemical shift of the sample with high Al content is more obvious, which means a more serious degree of oxidation of the sample with high Al content. According to formula (4.19), the approximate relative O content of the sample surface could be calculated.

$$C_i = \frac{I_i/\alpha_i}{\sum_i^n I_i/\alpha_i} \tag{4.19}$$

Here C_i is the percentage of element i, and I_i is the photoelectron spectroscopy intensity of element i, and α_i is the sensitivity factor of elements. Calculated by the above formula, the O content of sample A surface is about 10.9 %, and 11.4 % for the surface of sample B. This value explains that the degree of oxidation of sample B is more serious than that of sample A.

Combined with the XPS results, the time-dependent nature of the PL results in air is clearly understood. The degree of surface oxidation of the sample placed in

the air increases with the length of time and increases the intensity of the emission peak of the impurity, and the intensity of the band-edge emission peaks decrease. Due to the Al component of sample B being higher than that of sample A, the degree of oxidation becomes more severe and band-edge emission intensity decreases faster. This also gives an explanation of changes in surface topography and roughness. Because Al atoms have a lower surface migration rate than that of Ga atoms, the Al content of the surface step flow at the stairs is higher than that of the gully between steps. At the beginning of exposure, the degree of oxidation above steps is faster than that of the gully when placed in the air. The Fig. 4.66b shows that irregular top spots appear on the top of stairs, while the gully department does not change significantly and the surface roughness degree will increase at the same time and, when it has been placed in air for a long time, the ravines department also begins to be oxidized. When the degree of oxidation of the top has reached saturation, the typical surface topography will be smeared out. The surface roughness becomes slightly smooth, and finally when the oxidation rate at the top of the steps and gully is almost same, the surface morphology becomes what is shown in Fig. 4.66d.

The above results show that AlGaN material is easily oxidized in the air, and oxidation has a significant impact on the optical properties and surface chemistry of material surface. This is not conducive to the subsequent device fabrication, so necessary measures to prevent oxidation must be taken in the process of preparing AlGaN devices.

4.5 Summary

An investigation of the device structures of back-illuminated solar-blind GaN-based ultraviolet detectors and MOCVD epitaxial growth techniques of GaN, AlGaN materials were introduced in this chapter. As a breakthrough for GaN-based materials, the detailed explanations of the two-step epitaxial techniques were specially introduced. This chapter adopted the method of explaining epitaxial growth technique gradually from the GaN to AlGaN. First, the basics of GaN materials, the basic principles of MOCVD epitaxial growth technology and in situ monitoring systems were introduced. Then AlGaN epitaxial technology, P-type doping of GaN, integrated performance analysis of AlGaN material and so on were introduced. All the contents can be summarized as follows:

The characteristics of MOCVD equipment and the basic process of MOCVD epitaxial growth for GaN materials were briefly introduced. The optical monitoring methods of MOCVD growth process were elaborated. A reasonable explanation was given for the curve of in situ monitoring of MOCVD growth combined with the specific growth process. Through analysis of the in situ monitoring curve of GaN, growth rate, thickness of the epitaxial layer, the refractive index at high-temperature growth of GaN and surface roughness of the growth process were all confirmed.

MOCVD in situ monitoring system is a useful tool for the understanding of the process and mechanism of epitaxial growth in-depth.

The two-step technique of MOCVD growth of GaN material was investigated. Using a near-vertical laser reflecting in situ monitoring system, the impact of growth parameters of a low-temperature buffer layer on the traditional two-step method was systematically studied; also a microscopic analysis of GaN growth process was conducted. It is found that the rough and low nucleation density surface after the buffer layer annealing is most suitable for growing a high-quality GaN epitaxial layer. Reasonable explanations were made for improving the crystal quality of GaN as the chamber pressure for the buffer layer growth increases. GaN and AlGaN growth mechanism on the GaN buffer layer were compared.

Analysis of the growth process and mechanisms for GaN epitaxial layers grown on low-temperature AlN buffer layer was made. It was found that a larger grain size and smaller nucleation density in the low-temperature AlN buffer layer is of great importance for the growth of high-temperature GaN epitaxial layer. It is also found that reductions of the pressure in the reaction chamber and the flow rate of NH_3 tend to reduce pre-reaction and achieve a high Al component AlGaN material. There are further studies for epitaxial growth techniques of AlGaN material on the AlN buffer layer. It is noticed that a small amount of In-doping and a superlattice structure for the inserted layer in growth process will help to improve the quality of AlGaN material. Also GaN P-type doping techniques were introduced.

Synthesized performance of the GaN, AlGaN material was studied. It was found that the introduction of acceptor by edge dislocations in GaN material not only enhances the luminous intensity of yellow defects, but also affects mobility. Al component and strain state of AlGaN materials were measured, and there is a larger relaxation in the case of a relatively high Al component. Possible mechanisms were explained for the nonuniformity of the Al component in AlGaN material, also a significant impact on the optical properties and surface chemistry of AlGaN surface oxidation was found, so that necessary measures to prevent oxidation in the preparation of AlGaN device process have to be taken.

References

1. Nakamura S, Pearton S, Fasol G. The blue laser diode. Berlin: Springer; 2000.
2. Nakamura S. The role of structural imperfections in InGaN-based blue light-emitting diodes and laser diodes. Science. 1998;281(5379):956–61.
3. Johnson WC, Parsons JB, Crew MC. Nitrogen compounds of gallium III Gallic nitride. J Phys Chem. 1932;36:2561–72.
4. Tiede E, Thimann M, Sensse K. Phosphorescence-capable by silicon activated aluminum nitride. Chem Berichte. 1928;61:1568–73.
5. Juza R, Hahn J. Crystal structures of Cu_3N, GaN and InN—metallic amides and metallic nitrides V announcement. Zeitschr Anorgan Allgem Chem. 1938;239:282–7.
6. Zetterstrom RB. Synthesis and growth of single crystals of gallium nitride. J Mater Sci. 1970;12(5):1102–4.

7. Dingle R, Shakee KL, Leheny EF, Zetterstrom RB. Stimulated emission and laser action in gallium nitride. Appl Phys Lett. 1971;1(19):5–7.
8. Hovel HJ, Cuomo JJ. Electrical and optical propertied of RF-sputtered GaN and InN. Appl Phys Lett. 1972;2(20):71–3.
9. Puychevrier N, Menoret M. Synthesis of 3–5 semiconductor nitride by reactive cathodic sputtering. Thin Solid Films. 1976;1(36):141–5.
10. Maruska HP, Tietjen JJ. Preparation and propertied of vapor-deposited single-crystalline GaN. Appl Phys Lett. 1969;10(15):327–9.
11. Ban VS. Mass spectrometric studies of vapor-phase crystal growth GaN. J Electrochem Soc. 1972;6(119):761–5.
12. Llegems M, Dingle T, Logan TA. Luminescence of Zn- and Cd-doped GaN. J Appl Phys. 1972;9(43):3797–800.
13. Pankove JI, Berkeuheiser JE, Maruska SP, Wittke J. Luminescent properties of GaN. Solid State Commun. 1970;13(8):1051–3.
14. Ilegems M. Vapour epitaxy of gallium nitride. J Cryst Growth. 1972;13–14:360-4-364.
15. Matsumoto T, Sano M, Aoki M. Pair luminescence from Zn-doped GaN. Jpn J Appl Phys. 1974;13(2):373–4.
16. Seifert W, Fitzl G, Butter E. Study on the growth-rate in VPE of GaN. J Crystal Growth. 1981;52:257–62.
17. Naniwae K, Itoh S, Amano H, Itoh K, Hiramatsu K, Akasaki I. Growth of single-crystal GaN substrate using hydride vapor-phase epitaxy. J Crystal Growth. 1990;99(1–4):381–4.
18. Manasevit HM, Simpson WI. Use of metal-organics in preparation of semiconductor materials epitaxial gallium-V compounds. Appl Phys Lett. 1969;116(12):1725–32.
19. Gotoh H, Suga T, Suzuki H, Kimata M. Low-temperature growth of Gallium nitride. Jpn J Appl Phys. 1981;20(7):L545–8.
20. Amano H, Sawaki N, Akasaki I, Toyoda Y. Metalorganic wapor-phase epitaxial-growth of a high-quality GaN film using an AIN buffer layer. Appl Phys Lett. 1986;48(5):353–5.
21. Nakamura S. GaN growth using GaN buffer layer. Jpn J Appl Phys. 1991;30(10A):L1705–7.
22. Amano H, Akasaki I, Hiramatsu K, Koide N, Sawaki N. Effects of the buffer layer in metalorganic vapor-phase epitaxy of GaN on sapphire substrate. Thin Solid films. 1988;163:415–420.
23. Nakamura S, Mukai T, Senoh M, Iwasa N. Thermal annealing effects on p-type Mg-doped GaN films. Jpn J Appl Phys. 1992;Part 2 31(2B):L139–L142.
24. Fen J. MOVPE lateral eptaxial growth of GaN and research of its structure and properties. Doctoral dissertation of Institute of semiconductor, Chinese Academy of Sciences; 2003.
25. Asif Khan M, Kuznia JN, Skogman RA, Olson DT, Millan M, Choyke WJ. Low pressure metalorganic chemical vapor deposition of AlN over sapphire substrates. Appl Phys Lett. 1992;61(21):2539–41.
26. McClintock R, Mayes K, Yasan A, Shiell D, Kung P, Razeghi M. 320 × 256 solar-blind focal plane arrays based on $Al_xGa_{1-x}N$. Appl Phys Lett. 2005;86(1):011117.
27. Zhang JP, Hu X, Bilenko Y, Deng J, Lunev A, Shur MS, Gaska R, Shatalov M, Yang JW, Khan MA. AlGaN-based 280 nm light-emitting diodes with continuous-wave power exceeding 1 mW at 25 mA. Appl Phys Lett. 2004;85(23):5532–4.
28. Taniyasu Y, Kasu M, Makimoto T. An aluminium nitride light-emitting diode with a wavelength of 210 nm. Nature. 2006;441(7091):325–8.
29. Pankove JI, Moustakas TD. Gallium Nitride (GaN) I. CA: Academic Press; 1998.
30. Yang H, Zheng LX, Li JB, Wang XJ, Xu DP, Wang YT, Hu XW, Han PD. Cubic-phase GaN light-emitting diodes. Appl Phys Lett. 1999;74(17):2498–500.
31. Poroswki S, Grzegory I. Chapter 9. In: Pearton SJ editor. GaN and related materials. Gordon and Breach Science Publishers; 1997.
32. Manasevit HM. Single-crystal gallium arsenide on insulating substrates. Appl Phys Lett. 1968;12(4):156–9.
33. Manasevit HM, Simpson WI. Use of metal-organic in preparation of semiconductor materials epitaxial Gallium-Vcompounds. J Electrochem Soc. 1969;116(12):1725–32.

34. Manasevit HM, Hess KL. Use of metal-organic in preparation of semiconductor materials gallium antimonide. J Electrochem Soc. 1979;126(11):2031–3.
35. Yang SR, Ding MY. Chapter 3. Epitaxial growth technique. Beijing: National defence industry press.
36. Denbaars SP, MaaYB Dapkus PD, Lee HC. Homogeneous and heretogeneous thermal-decomposition rates of trimethyl gallium and arsine and their relevance to the growth of GaAs by MOCVD. J Crystal Growth. 1986;77(1–3):188–93.
37. Ng TB, Han J, Biefeld RM, Weckwerth MV. In-situ reflectance monitoring during MOCVD of AlGaN. J Electron Mater. 1998;27(4):190–5.
38. Nakamura S. Analysis of real-time monitoring using interference effects. Jpn J Appl Phys. 1991;30(7):1348–53.
39. Nakamura S. In-situ monitoeing of GaN growth using interference effects. Jpn J Appl Phys. 1991;30:1620–7.
40. Nakamura S, Mukai T, Senoh M. In-situ monitoring and hall measurement of GaN growth with GaN buffer layers. J Appl Phys. 1992;71(11):5543–9.
41. Azzam RMA, Bashara NM. North Holland New York: Ellipsometry and Polarized Light; 1988. p. 332.
42. Mohammad SN, Salvador AA, Morkoc H. Emerging gallium nitride based devices. Proc IEEE. 1995;83(10):1306–55.
43. Stafford A, Irvine SJC, Hess KL, Bajaj J. The use of in situ laser interferometry for MOCVD process control. Semicond Sci Technol. 1998;13(12):1407–11.
44. Figge S, Bottcher T, Einfeldt S, Hommel D. In situ and ex situ evaluation of the film coalescence for GaN growth on GaN nucleation layers. J Crystal Growth. 2000;221:262–6.
45. Nakamura S, Harada Y, Seno M. Novel metalorganic chemical vapor-deposition system for GaN growth. Appl Phys Lett. 1991;58(18):2021–3.
46. Lester SD, Ponce FA, Craford MG, Steigerwald DA. High dislocation densities in high-sfficiency GaN-based light-emitting-diodes. Appl Phys Lett. 1995;66(10):1249–51.
47. Akasaki I, Amano H, Koide Y, Hiramatsu K, Sawaki N. Effects of AlN buffer layer on crystallographic structure and on electrical and optical-properties of GaN and AlGaN films grown on sapphire substrate by MOVPE. J Crystal Growth. 1989;98(1–2):209–19.
48. Hiramatsu K, Itoh S, Amano H, Akasaki I, Kuwano N, Shiraishi T, Oki K. Growth-mechanism of GaN growth on sapphire with AlN buffer layer by MOVPE. J Cryst Growth. 1991;115(1–4):628–33.
49. Ponce FA. Defects and interfaces in GaN epitaxy. MRS Bull. 1997;22(2):51–7.
50. Hiramatsu K, Detchprohm T, Akasaki I. Relaxation mechanism of thermal-stresses in the heterostructure of GaN grown on sapphire by vapor-phase epitaxy. Jpn J Appl Phys. 1993; Part 1 32(4):1528–1533.
51. Chen J. MOCVD epitaxial growth of hexagonal Ga. Master's thesis of Sichuan University; 2003.
52. Chen J, Zhang SM, Zhang BS, Zhu JJ, Feng G, Shen XM, Wang YT, Yang H, Zheng WC. Effects of reactor pressure on GaN nucleation layers and subsequent GaN epilayers grown on sapphire substrate. J Crystal Growth. 2003;254(3–4):348–52.
53. Zhang JC, Zhao DG, Wang JF, Wang YT, Chen J, Liu JP, Yang H. The influence of AlN buffer layer thickness on the properties of GaN epilayer. J Crystal Growth. 2004;268(1–2):24–29.
54. Heinke H, Kirchner V, Einfeldt S, Hommel D. X-ray diffraction analysis of the defect structure in epitaxial GaN. Appl Phys Lett. 2000;77(22):2145–7.
55. Heying B, Wu XH, Keller S, Li Y, Kapolnek D, Keller BP, DenBaars SP, Speck JS. Role of threading dislocation structure on the x-ray diffraction peak width in epitaxial GaN films. Appl Phys Lett. 1996;68(5):643–5.
56. Chen J, Zhang SM, Zhang BS, Zhu JJ, Shen XM, Feng G, Liu JP, Wang YT, Yang H, Zheng WC. Influence of reactor pressure of GaN buffer layers on moephological evolution of GaN grown by MOCVD. J Crystal Growth. 2003;256(3–4):248–53.

57. Yang T, Uchida K, Mishima T, Kasai J, Gotoh J. Control of initial nucleation by reducing theV/III ratio duting the early stages of GaN growtn. Phys Stat Sol A. 2000;180(1):45–50.
58. Han J, Ng TB, Biefeld RM, Crawford MH, Follstaedt DM. The effect of H-2 on morphology evolution during metalorganic chemical vapor deposition. Appl Phys Lett. 1997;71 (20):3114–6.
59. Koleske DD, Fischer AJ, Allerman AA, Mitchell CC, Cross KC, Kurtz SR, Figiel JJ, Fullmer KW, Breiland WG. Improved brightness of 380 nm GaN light emitting diodes through intentional delay of the nucleation island coalescence. Appl Phys Lett. 2002;81 (11):1940–2.
60. Mihopoulos TG, Gupta V, Jensen KF. A reaction-transport model for AlGaN MOVPE growth. J Crystal Growth. 1998;195(1–4):733–9.
61. Makino O, Nakamura K, Tachibana A, Tokunaga H, Akustu N, Matsumoto K. Quantum chemical mechanism in parasitic reaction of AlGaN alloys formation. Appl Surf Sci. 2000;159:374–9.
62. Shih CF, Chen NC, Lin SY, Liu KS. AlGaN films grown on (0001) sapphire by a two-step method. Appl Phys Lett. 2005;86(21):211103.
63. Bremser MD, Perry WG, Zheleva T, Edwards NV, Nam OH, Parikh N, Aspnes DE, Davis RF. Growth, doping and characterization of AlGaN thin film alloys on 6H–SiC(0001) structures. MRS Internet J Nitride Semicond Res. 1996;1(1–46):U59–71.
64. Zhao DG, Liu ZS, Zhu JJ, Zhang SM, Jiang DS, Yang H, Liang JW, Li XY, Gong HM. Effects of Al incorporation on the AlGaN grown by metalorganic chemical vapor deposition. Appl Surf Sci. 2006;253(5):2452–5.
65. Zhao DG, JiangDS, ZhuJJ, Liu ZS, Zhang SM, Yang H, Liang JW. The influence of V/III ratio in the initial growth stage on the properties of GaN epitalyer deposited on low temperature AlN buffer layer. J Crystal Growth. 2007;303(2):414–418.
66. Xu ZJ. Test and analysis of semiconductor. The second edition: Science press; 2007.
67. Zhao DG, Zhu JJ, Liu ZS, Zhang SM, Yang H, Jiang DS, Zhao DG, Zhu JJ, Liu ZS, Zhang SM, Yang H, Jiang DS. Surface morphology of AlN buffer layer and its effect on GaN grown by metalorganic chemical vapor deposition. Appl Phys Lett. 2004;85(9):1499–501.
68. Reshchikov MA, Morkoc H. Luminescence properties of defects in GaN. J Appl Phys. 2005;97(6):061301.
69. Khan MA, Shatalov M, Maruska HP, Wang HM, Kuokstis E. III-nitride devices. Jpn J Appl Phys. 2005;44(10):7191–206.
70. Lada M, Cullis AG, Parbrook PJ. Effect of annealing temperature on GaN nucleation layer transformation. J Crystal Growth. 2003;258(1–2):89–99.
71. Zhou HL, Chua SJ, Peng C. Al incorppration in AlGaN on (11(2) over-bar-2) and (0001) surface orientation. J Crystal Growth. 2006;292(1):5–9.
72. Ke WC, Ku CS, Huang HY, Chen WC, Lee L, Chen WK, Chou WC, Chen WH, Lee MC, Lin WJ, Cheng YC, Cheng YT. Microphoto-luminescence spectra of hillocks in $Al_{0.11}Ga_{0.89}N$ films. Appl Phys Lett. 2004;85(15):3047–9.
73. Chen P, Chua SJ, Miao ZL. Phase separation in AlGaN/GaN heterojunction grown by matalorganic chemical vapor deposition. J Crystal Growth. 2004;273(1–2):74–8.
74. Jahn U, Jiang DS, Ploog KH, Wang XL, Zhao DG, Yang H. Correlation between optical and structural properties of (Al, Ga)N layers grown by MOCVD. Phys Stat Sol (a). 2007;204 (1):294–8.
75. Brown JD, Yu ZH, Matthews J, Harney S, Boney J, Schetzina JF, Benson JD, Dang KW, Terrill C, Nohava T, Yang W. Visible-blind UV digital cameta based on a 32 × 32 array of GaN/AlGaN p–i–n photodiodes. MRS Internet J Nitride Semicond Res. 1999;4(9):art. no. 9.
76. Brown JD, Boney J, Matthews J, Srinivasan P, Schetzina JF, Nohava T, Yang W. Solar-blind AlGaN heterostructure photodiodes. MRS Internet J Nitride Semicond Res. 2000;5(9):1–7.
77. Ito T, Ohtsuka K, Kuwahara K, Sumiya M, Takano Y, Fuke S. Effect of AlN buffer layer deposition conditions on the properties of GaN layer. J Crystal Growth. 1999;205(1–2):20–4.
78. Sasaki T, Matsuoka T. Analysis of 2-step-growth conditions for GaN on an AlN buffer layer. J Appl Phys. 1995;77(1):192–200.

79. Kuznia JN, Asif Khan M, Olson DT. Influence of buffer layers on the deposition of high-quality single-crystal GaN over sapphire substrate. J Appl Phys. 1993;73(9):47004702.
80. Kobayashi Y, Akasaka T, Kobayashi N. Thermal stability of low-temperature GaN and AlN buffer layers during metalorganic vapor phase epitaxy monitored by in situ shallow angle reflectance using ultraviolet light. Jpn J Appl Phys. 1998;37(10B):L1208–10.
81. Nakamura F, Hashimoto S, Hara M, Imanaga S, Ikeda M, Kawai H. AlN and AlGaN growth using low-pressure metalorganic chemical vapor deposition. J Crystal Growth. 1998;195(1–4):280–5.
82. Kim S, Seo J, Lee K, Lee H, Park K, Kim Y, Kim CS. Growth of AlGaN epilayers related gas-phase reactions using TPIS-MOCVD. J Crystal Growth. 2002;245(3–4):247–53.
83. Creighton JR, Wang GT, Breiland WG, Coltrin ME. Nature of the parasitic chemistry during AlGaInN OMVPE. J Crystal Growth. 2004;261(2–3):204–13.
84. Ruffenach-Clur S, Briot O, Rouviere JL, Gil B, Aulombard RL. MOVPE growth and characterization of $Al_xGa_{1-x}N$. Mater Sci Eng B. 1997;50(1–3):219–22.
85. Chen CH, Liu H, Steigerwald D, Imler W, Kuo CP, Craford MG, Ludowise M, Lester S, Amano J. A study of parasitic reactions between NH_3 and TMGa or TMAI. J Electronic Materials. 1996;25(6):1004–8.
86. Nakamura K, Makino O, Tachibana A, Matsumoto K. Quantum chemical study of parasitic reaction in III-V nitride semiconductor crystal growth. J Organomet Chem. 2000;611(1–2):514–24.
87. Zhao DG, Zhu JJ, Jiang DS, Yang H, Liang JW, Li XY, Gong HM. Parasitic reaction and its effect on the growth rate of AlN by metalorganic chemical vapor deposition. J Crystal Growth. 2006;289(1):72–5.
88. Reshchikov MA, Yi GC, Wessels BW. Behavior of 2.8- and 3.2-eV photoluminescence bands in Mg-doped GaN at different temperatures and excitation densities. Phys Rev B. 1999;59(20):13176–83.
89. Wang XL. MOCVD growth and properties research of high Al content AlGaN film. Doctoral dissertation of Institute of Semiconductor, Chinese Academy of Science; 2007.
90. Wang XL, Zhao DG, Li XY, Gong HM, Yang H, Liang JW. The effects of LT AlN buffer thickness on the properties of high Al composition AlGaN epilayers. Mater Lett. 2006;60(29–30):3693–6.
91. Wang XL, Zhao DG, Yang H, Liang JW. Growth of AlGaN epitaxial film with high Al content by metalorganic chemical vapour deposition. Chin Phys Lett. 2007;24(3):774–7.
92. Wang XL, Zhao DG, Jahn U, Ploog K, Jiang DS, Yang H, Liang JW. The effects of LT AlN buffer thickness on the optical properties of AlGaN grown by MOCVD and Al composition inhomogeneity analysis. J Phys D Appl Phys. 2007;40(4):1113–7.
93. Yu HB, Strupinski W, Butun S. Mg-doped AlGaN grown on an AlN/sapphire template by metalorganic chemical vapour deposition. Phys Status Solidi A. 2006;203(5):868–73.
94. Adivarahan V, Simin G, Tamulaitis G, Srinivasan R, Yang J, Asif Khan M, Shur MS, Gaska R. Indium–silicon co-doping of high-aluminum-content AlGaN for solar blind photodetectors. Appl Phys Lett. 2001;79(12):1903–5.
95. Keller S, Heikman S, Ben-Yaacov I, Shen L, Denbaars SP, Mishra UK. Indium surfactant assisted growth of AlN/GaN heterostructures by metal-organic chemical vapor deposition. Phys Stat Sol (a). 2001;188(2):775–8.
96. Widmann F, Daudin B, Feuillet G, Pelekanos N, Rouviére JL. Improved quality GaN grown by molecular beam epitaxy using In as a surfactant. Appl Phys Lett. 1998;73(18):2642–4.
97. Shen XQ, Ramvall P, Riblet P, Aoyagi Y. Improvements of the optical and electrical properties of GaN films by using in-doping method during growth. Jpn J Appl Phys. 1999;38(4B):L411–3.
98. Mochizuki S, Detchprohm T, Sano S, Nakamura T, Amano H, Akasaki I. Reduction of threading dislocation density in Al(X)Ga(1-X)N grown on periodically grooved substrates. J Cryst Growth. 2002;237:1065–9.

99. Hirayama H, Ainoya M, Kinoshita A, Hirata A, Aoyagi Y. Fabrication of a low-threading-dislocation-density $Al_xGa_{1-x}N$ buffer on SiC using highly Si-doped $Al_xGa_{1-x}N$ superlattices. Appl Phys Lett. 2002;80(12):2057–9.
100. Tsai YL, Wang CL, Lin PH, Liao WT, Gong JL. Observation of compositional pulling phenomenon in $Al_xGa_{1-x}N$ ($0.4 < x < 1.0$) films grown on (0001) sapphire substrates. Appl Phys Lett. 2003;82(1):31–3.
101. Chang L, Lai SK, Chen FR, Kai JJ. Observation of Al segregation around dislocation in AlGaN. Appl Phys Lett. 2001;79(7):928–30.
102. Wang XL, Zhao DG, Jiang DS, Yang H, Liang JW, Jahn U, Ploog K. Al composition inhomogeneity of AlGaN epilayer with a high Al composition grown by matal-organic chemical vapour deposition. J Phys Conden Matter 2007;19(17):176005.
103. Yan F, Tsukihara M, Nakamura A, Yadani T, Fukumoto T, Naoi Y, Sakai S. Surface smoothing mechanism of AlN film by initially alternating supply of ammonia. Jpn J Appl Phys. 2004;43(8B):L1057–9.
104. Zhang JP, Asif Khan M, Sun WH, Wang HM, Chen CQ, Fareed Q, Kuokstis E, Yang JW. Pulsed atomic-layer epitaxy of ultrahigh-quality $Al_xGa_{1-x}N$ structures for deep ultraviolet emissions below 230 nm. Appl Phys Lett. 2002;81(23):4392–4.
105. Zhang JP, Wang HM, Sun WH, Adivarahan V, Wu S, Chitnis A, Chen CQ, Shatalov M, Kuokstis E, Yang JW, Asif Khan M. High-quality AlGaN layers over pulsed atomic-layer epitaxially grown AlN templates for deep ultraviolet light-emitting diodes. J Electron Mater. 2003;32(5):364–70.
106. Götz W, Johnson NM, Bour DP, McCluskey MD, Haller EE. Local vibrational modes of the Mg–H acceptor complex in GaN. Appl Phys Lett. 1996;69(24):3725–7.
107. Reboredo FA, Pantelides ST. Novel defect complexes and their role in the p-type doping of GaN. Appl Phys Lett. 1999;82(9):1887–90.
108. Amano H, Kito M, Hiramatsu K, Akasaki I. P-type conduction in Mg-doped GaN and treated with low-energy electron-beam irradiation. Jpn J Appl Phys. 1989;Part 2 28(12): L2112–L2114.
109. Nakamura S, Iwasa N, Senoh M, Mukai T. Hole compensation mechanism of P-type GaN films. Jpn J Appl Phys. 1992;Part1 31(5A):1258–1266.
110. Chung HYA, Pelzmann A, Drechsler M, Scherer M, Schwegler V, Seyboth M, Kirchner C, Kamp M. Multiple-step annealing for 50 % enhanced p-conductivity of GaN. J Crystal Growth. 2001;230(3–4):549–53.
111. Götz W, Johnson NM, Walker J, Bour DP, RA Street. Activation of acceptors in Mg-doped GaN grown by metalorganic chemical vapor deposition. Appl Phys Lett. 1996;68(5):667–9.
112. Look DC, Reynolds DC, Hemsky JW, Sizelove JR, Jones RL, Molnar RJ. Defect donor and acceptor in GaN. Phys Rev Lett. 1997;79(12):2273–6.
113. Seifert W, Franzheld R, Butter E, Sobotta H, Riede V. On the opigin of free-carriers in high-conducting normal GaN. Cryst Res Technol. 1983;18(3):383–90.
114. Hautakangas S, Oila J, Alatalo M, Saarinen K, Liszkay L, Seghier D, Gislason HP. Vacancy defects as compensating centers in Mg-doped GaN. Phys Rev Lett. 2003;90(13):137402.
115. Obloh H, Bachem KH, Kaufmann U, Kunzer M, Maier M, Ramakrishnan A, Schlotter P. Self-compensation in Mg doped p-type GaN grown by MOCVD. J Crystal Growth. 1998;195(1–4):270–3.
116. Katayama-Yoshida H, Nishimatsu T, Yamamoto T, Orita N. Codoping method for the fabrication of low-resistivity wide band-gap semiconductors in p-type GaN, p-type AlN and n-type diamond: prediction versus experiment. J Phys Condens Matter. 2001;13 (40):89018914.
117. Yamamoto T, Katayamo YH. Materials design for the fabrication of low-resistivity p-type GaN using a codoping method. Jpn J Appl Phys. 1997;36(2B):L180–3.
118. Viswanath AK, Shin EJ, Lee JI, Yu S, Kim D, Kim B, Choi Y, Hong CH. J Appl Phys. 1998;83(4):2272–5.
119. Van deWalle CG. Interactions of hydrogen with native defects in GaN. Phys Rev B. 1997;56: R10020–3.

120. Nakamura S, Senoh M, Nagahama S, Iwasa N, Yamada T, Matsushita T, Kiyoku H, Sugimoto Y, Kozaki T, Umemoto H, Sano M, Chocho K. InGaN/GaN/AlGaN-based laser diodes with modulation-doped strained-layer superlattices grown on an epitaxially laterally overgrown GaN substrate. Appl Phys Lett. 1998;72(2):211–3.
121. Zhao DG, Jiang DS, Zhu JJ, Wang H, Liu ZS, Zhang SM, Yang H. Hole concentration test of p-type GaN by analyzing the spectral response of p–n(+) structure GaN ultraviolet photodetector. J Alloy Compd. 2010;492(1–2):300–2.
122. Chen Y, Gulino DA. Residual stress in GaN films grown by metalorganic chemical vapor deposition. J Vac Sci Technol A. 1999;17(5):3029–32.
123. Ponce FA, Bour DP, Gotz W, Wright PJ. Spatial distribution of the luminescence in GaN thin films. Appl Phys Lett. 1996;68(1):57–9.
124. Li G, Chua SJ, Xu SJ, Wang W. Nature and elimination of yellow-band luminescence and donor-acceptor emission of undoped GaN. Appl Phys Lett. 1999;74(19):2821–3.
125. Saarinen K, Seppala P, Olila J, Hautojarvi P, Corbel C, Briot O, Aulombard RL. Gallium vacancies and the growth stoichiometry of GaN studied by positron annihilation spectroscopy. Appl Phys Lett. 1998;73(22):3253–5.
126. Basak D, Lachab M, Nakanishi T, Sakai S. Effect of reactive ion etching on the yellow luminescence of GaN. Appl Phys Lett. 1999;75(23):3710–2.
127. Saarinen K, Laine T, Kuisma S, Nissilä J, HautoJärvi P, Dobrzynski L, Baranowski J, Pakula K, Stepniewski R, Wojdak M, Wysmolek A, Suski T, Leszczynski M, Grzegory I, Porowski S. Observation of native Ga vacancies in GaN by positron annihilation. Phys Rev Lett. 1997;79(16):3030–3.
128. Korotkov RY, Reshchikov MA, Wessels BW. Acceptors in undoped GaN studied by transient photo luminescence. Phys B. 2003;325(1–4):1–7.
129. Kwon YH, Shee SK, Park GH, Hwang SJ, Song JJ. Time-resolved study of yellow and blue luminescence in Si- and Mg-doped GaN. Appl Phys Lett. 2000;76(7):840–2.
130. Zhao DG, Jiang DS, Yang H, Zhu JJ, Liu ZS, Zhang SM, Liang JW, Li X, Li XY, Gong HM. Role of edge dislocations in enhancing the yellow luminescence of n-type GaN. Appl Phys Lett. 2006;88(25):241917.
131. Read WT. Theory of dislocations in germanium. Philos Mag. 1954;45(367):775–96.
132. Podor B. Electron mobility in plastically defoemed germanium. Phys Status Solidi. 1966;16 (2):K167–70.
133. Shiojima K, Suemitsu T, Ogura M. Correlation between current-voltage characteristics and dislocations for n-GaN Schottky contacts. Appl Phys Lett. 2001;78(23):3636–8.
134. Krtschil A, Dadgar A, Krost A. Decoration effects as origin of dislocation-related charges in gallium nitride layers investigated by scanning surface potential microscopy. Appl Phys Lett. 2003;82(14):2263–5.
135. Ng HM, Doppalapudi D, Moustakas TD, Weimann NG, Eastman LF. The role of dislocation scattering in n-type GaN films. Appl Phys Lett. 1998;73(6):821–3.
136. Weimann NG, Eastman LF. Scattering of electrons at threading dislocations in GaN. J Appl Phys. 1998;83(7):3656–9.
137. Look DC, Sizelove JR. Dislocation scattering in GaN. Phys Rev Lett. 1999;82(6):1237–40.
138. Ilegems M, Montgomery HC. Electrical properties of n-type vapor-grown gallium nitride. J Phys Chem Solids. 1973;34(5):885–95.
139. Soh CB, Chua SJ, Lim HF, Chi DZ, Tripathy S, Liu W. Assignment of deep levels causing yellow luminescence in GaN. J Appl Phys. 2004;96(3):1341–7.
140. Kaufmann U, Kunzer M, Obloh H, Maier M, Manz C, Ramakrishnan A, Santic B. Origin of defect-related photoluminescence bands in doped and nominally undoped GaN. Phys Rev B. 1999;59(8):5561–7.
141. Lee IH, Choi IH, Lee CR, Noh SK. Evolution of stress relaxation and yellow luminescence in GaN/sapphire by Si incorporation. Appl Phys Lett. 1997;71(10):1359–61.
142. Neugebauer J, Van De Walle CG. Atomic geometry and electronic-structure of native defects in GaN. Phys Rev B. 1994;50(11):8067–70.

143. Reshchikov MA, Morkoc H. Luminescence properties of defects in GaN. J Appl Phys. 2005;97(6):061301.
144. Zhao DG, Jiang DS, Zhu JJ, Liu ZS, Wang H, Zhang SM, Wang YT, Yang H. Role of edge dislocation and Si impurity in linking the blue luminescence and yellow luminescence in n-type GaN films. Appl Phys Lett. 2009;95(4):041901.
145. Zhao DG, Jiang DS, Zhu JJ, Guo X, Liu ZS, Zhang SM, Wang YT, Yang H. Defect evolution and accompanied change of electrical properties during the GaN growth by metalorganic chemical vapor deposition. J Alloy Compd. 2009;487(1–2):400–3.
146. Zhao DG, Yang H, Zhu JJ, Jiang DS, Liu ZS, Zhang SM, Wang YT, Liang JW. Effects of edge dislocations and intentional Si doping on the electron mobility of n-type GaN films. Appl Phys Lett. 2006;89(11):112106.
147. Zhao DG, Jiang DS, Yang H, Zhu JJ, Liu ZS, Zhang SM, Liang JW, Hao XP, Wei L, Li X, Li XY, Gong HM. Effect of lightly Si doping on the minority carrier diffusion length in n-type GaN films. Appl Phys Lett. 2006;88(25):252101.
148. Zhao DG, Jiang DS, Zhu JJ, Liu ZS, Zhang SM, Liang JW, Yang H, Li X, Li XY, Gong HM. Influence of defects in n(-)-GaN layer on the responsivity of Schottky barrier ultraviolet photodetectors. Appl Phys Lett. 2007;90(6):062106.
149. Zhao DG, Zhang S, Liu WB, Hao XP, Jiang DS, Zhu JJ, Liu ZS, Wang H, Zhang SM, Yang H, Wei L. Role of Ga vacancies in enhancing the leakage current of GaN Schottky barrier ultraviolet photodetectors. Chin Phys B. 2010;19(5):057802.
150. Zhao DG, Xu SJ, Xie MH, Tong SY, Yang H. Stress and its effect on optical properties of GaN epilayers grown on Si(111), 6H–SiC(0001), and c-plane sapphire. Appl Phys Lett. 2003;83(4):677–9.
151. Singh R, Doppalapudi D, Moustakas TD, Romano LT. Phase separation in InGaN thick films and formation of InGaN/GaN double heterostructures in the entire alloy composition. Appl Phys Lett. 1997;70(9):1089–91.
152. El-Masty NA, Piner EL, Liu SX, Bedair SM. Phase separation in InGaN grown by matalorganic chemical vapor deposition. Appl Phys Lett. 1998;72(1):40–2.
153. Wu MF, Vantomme A, Hogg S, Langouche G, Van der Strich W, Jacobs K, Moerman I. Rutherford backscattering/channeling study of a thin AlGaN layer on $Al_{(2)}O_{(0)3}$(0001). Nucl Instrum Methods Phys Res B. 2001;174(1–2):181–6.
154. Zhang JC. Research of structure and optical properties of group III nitride. Doctoral dissertation of Institute of semiconductor. Dissertation, Chinese Academy of Science; 2005.
155. Wright AF. Elastic properties of zinc-blend and wurtzite AlN, GaN, and AlN. J Appl Phys. 1997;82(6):2833–9.
156. Zheng XH, Wang YT, Feng ZH, Yang H, Chen H, Zhou JM, Liang JW. Method for measurement of lattice parameter of cubic GaN layers on GaAs(001). J Cryst Growth. 2003;250(3–4):345–8.
157. Zhao DG, Jiang DS, Zhu JJ, Liu ZS, Zhang SM, Yang H, Jahn U, Ploog KH. Al composition variations in AlGaN films grown on low-temperature GaN buffer layer by metalorganic chemical vapor deposition. J Crystal Growth. 2008;310(24):5266–9.
158. Watkins NJ, Wicks GW, Gao YL. Oxidation study of GaN using x-ray photoemission spectroscopy. Appl Phys Lett. 1999;75(17):2602–4.
159. Wolter SD, Luther BP, Waltemyer DL, Önneby C, Mohney S, Molnar RJ. Appl Phys Lett. 1997;70(16):2156–8.
160. Hashizume T, Ootomo S, Nakasaki R, Oyama S, Kihara M. X-ray photoelectron spectroscopy characterization of AlGaN surfaces exposed to air and treated in NH_4OH solution. Appl Phys Lett. 2000;76(20):2880–2.
161. Wang XL, Zhao DG, Chen J, Li XY, Gong HM, Yang H. Effect of oxidation on the optical and surface properties of AlGaN. Appl Surf Sci. 2006;252(24):8706–9.

Chapter 5
HgCdTe Detector Chip Technology

5.1 Introduction

As a pseudobinary system material with the characteristics of adjustable band gap, high optical absorption coefficient, long carrier lifetime, high electron mobility, etc. [1], HgCdTe has always been the preferred material for fabricating infrared detectors for many years. As the requirements of the infrared systems on the detector performances increase continuously, the HgCdTe detector technology has become significant. From single-element, multielement photoconductive/photovoltaic devices to long linear array scanning focal plane arrays (FPAs) and large-scale staring FPAs, the detector scale has been enlarged constantly (at present, the largest scale of medium wave infrared (MWIR) detectors is 4 K × 4 K) [2] with pixel size decreasing (currently, the minimum size is 15 μm [1]), and the corresponding chip-making technology has been improved greatly. The classical chip architectures of HgCdTe detectors include ion-implanted planar junction detector [1] and ion-etching loophole junction detector [2]. The wet chemistry treatment technique, photolithography technique, passivation film growth technique, ohmic contact technique, indium bump preparation technique, and flip chip bonding technique have been developed and applied successfully in the development of HgCdTe detectors.

As mentioned in Chap. 1, the advanced imaging FPA detection technology requires higher spatial resolution and better capability of target recognition. Therefore, the advanced infrared FPA system chips are the intelligent imaging devices that can detect spectral signals on multiple bands simultaneously. The multiple-band detection of infrared spectra is achieved structurally through the vertical integration of multilayer HgCdTe films [2–5]. It can realize the full spatial synchronization of dual-band detection, while a novel direction is applied for the precise acquisition of target signals, thereby improving the capability of target recognition. Furthermore, these highly integrated two-color FPAs can greatly reduce the difficulties associated with the optics and cryogenics of infrared/photoelectric

L. He et al., *Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN*,
DOI 10.1007/978-3-662-52718-4_5

two-color detector systems, reduce the size, weight, and cost of system, and improve the system reliability. In the chip architecture, the photoelectric conversion of incident infrared signals from different bands should be accomplished in corresponding to HgCdTe films, and the electrical signals could output separately. Thus, the chip architectures are quite different from the planar structure of classical IR chips [2, 3]. The fabrication of advanced IR FPA system chips raises new challenges to IR chip processing technologies [2–5]. Additionally, for advanced IR FPA system chips, Si-based HgCdTe materials are applied to achieve very-large-scale IR imaging detection, whereas the chip processing problems, caused by the lattice and thermal mismatches between silicon and HgCdTe, present new challenges for the chip fabrication process.

In this chapter, according to the technological features of advanced IR FPA system chips, the processing principles and methods of HgCdTe multiband chips with deep mesa structure and vertical integration are introduced. Also, the features and processing methods of Si-based HgCdTe chips are presented.

In Sect. 5.2, the chip techniques of micro-mesa detectors fabricated by HgCdTe multilayer materials are presented, including dry etching, passivation, electrode lead-out, photolithography, and indium bump interconnection for a HgCdTe deep micro-mesa array.

In the first part of Sect. 5.2.1, the purpose of deep micro-mesa isolation is introduced, and the main characterization parameters of etching HgCdTe are demonstrated. Also, different methods of etching HgCdTe are discussed comparing advantages and disadvantages among them. The micro-mesa array of HgCdTe multilayer materials is achieved by low-damage etching that accurately transfers the mask patterns to HgCdTe materials. The key to this process is to form the deep micro-mesa array with fine surface morphology by the anisotropic etching process with low or no damage on HgCdTe utilizing low-energy plasma. The inductively coupled plasma (ICP) etching technique, which can avoid the limitation that the etching energy and plasma density cannot be controlled independently in traditional reactive ion etch (RIE), is one of the high density plasma (HDP) techniques. The HDP-ICP technique is quite suitable for the low-damage etching of HgCdTe materials [2]. In the second part of Sect. 5.2.1, the ICP etching process of HgCdTe and the influences of different etching conditions on chips are presented. Moreover, the optimal methods of etching mask are discussed, and the mask technique with the etching selection ratio higher than 20:1 is introduced [2–4]; The microcosmic processes under different chamber pressures and gas proportions are further analyzed [6–10] by X-ray photoelectron spectroscopy (XPS) analysis and scanning electron microscopy (SEM) observation of the etching surface microzone under different technological conditions and the reaction mechanism of ICP etching HgCdTe is explained preliminarily [2–5, 7]. For loading effect and aspect ratio dependence effect (ARDE) in deep micro-mesa etching and the resulting uniformity problems [2, 3], the critical dimension that the etching delay effect causes is investigated according to the loading effect curves of ICP etching HgCdTe [2, 3, 6, 7, 11]. Based on the experimental results, such as the process gas proportions, RF power, ICP power, and operating pressure, the control methods of the

interconversion between physical etching and chemical etching are analyzed [2]. The etching techniques used to provide a clean and smooth surface, good profile, high etching rate, and low induced damage are presented [2, 3, 6, 7] by investigating the reasonable balance between etching rate, etching profile (including etching surface), and etching damage The suppression and recovery of etching damage in a HgCdTe micro-mesa array has become a research hotspot and challenge [2–5, 11]. In the third part of Sect. 5.2.1, the evaluation methods of sample damage are discussed, and the corresponding experimental results of LBIC, *I–V* test, differential Hall test, etc., are analyzed. In view of the experimental results based on the relativity between the technological parameters (e.g., RF power, etching time, process gas and mask) and etching induced damage, the mechanism of induced damage by ICP dry etching in HgCdTe is demonstrated. Additionally, for the problems that dry etching causes weak damage layer, the recovery methods and technologies of etching damage are introduced. The photolithography, surface passivation, and climbing metallization should be carried out [2–4] after zero-damage formation of deep micro-mesa array. However, the concave–convex surface of mesa arrays brings new challenges to whirl coating, exposure, and developing of photolithography. In Sect. 5.2.2, the photolithographic process for fabricating a deep micro-mesa array is introduced combining the selection of photoresist, redundancy design of device size, and optimization of photolithographic condition.

Surface passivation is an important technological method for reducing surface leakage current and improving device stability [2–5]. On the formation of deep micro-mesa array, it is necessary to passivate the sidewall of *p-n* junction. Therefore, in order to satisfy the demands of preparing high-performance chips, sidewall passivation techniques must be improved, and the sidewall passivation technologies are presented in this section. In Sects. 5.2.4 and 5.2.5, the climbing metallization and hybridized interconnection are introduced, and the compatibility of these technologies is analyzed in detail.

In Sect. 5.3, the two-color micro-mesa chips are introduced, and the practical chip architectures for multiband IR detection are analyzed. Based on the experimental results of deep micro-mesa isolation and dry damage suppression/recovery, the processing techniques, including photolithography, sidewall passivation, metallization, indium bump growth, and hybridized interconnection, are given for a HgCdTe IR micro-mesa array chip. Accordingly, the experimental results of prototype chip are provided for verifying these key technologies.

Utilizing silicon as an alternative substrate in the HgCdTe chip technology is one of the key technologies for the third-generation FPAs. Silicon substrate has many advantages, such as large dimension, high mechanical strength, low cost, and good thermal matching with ROIC. It makes Si-based HgCdTe epitaxial material become an effective way to break through the limits of size and cost of CdZnTe substrates and improve the reliability of large-scale FPAs. The research results indicate that the HgCdTe FPA techniques based on silicon substrates are imperative to satisfy the demands of large-scale FPAs and potential IR imaging devices in the future.

In Sect. 5.4, the structural features and stress analysis of Si-based HgCdTe FPAs are demonstrated in detail, and the low-damage processing technologies of 3-in. Si-based HgCdTe chips are presented.

5.2 HgCdTe Detector Chip Processing Technologies

The advanced FPA chips are achieved by the indium bump hybridized interconnection of HgCdTe multiband photoelectric detector array and matched multiband ROIC [5, 6, 12–15]. The multiband IR detector chips are realized by employing micro-mesa processing technologies for small pixels, utilizing HgCdTe two-color materials grown by molecular beam epitaxy (MBE) or metalorganic vapor phase epitaxy (MOVPE) and in situ doping technique. Here, the micro-mesa processing technology is defined as the etching formation using high depth-to-width ratio trenches to isolate a small micro-mesa pixel array, as well as the subsequent dielectric passivation and climbing metallization on the concave–convex surface of detector chip [6, 14–18].

The multiband HgCdTe micro-mesa processing technologies, similar to the mature and popular micromechanical machining technologies, are involved in photolithography, ion implantation, etching, film growth, metallization, etc. However, due to the disadvantages of weak Hg–Te bonds, worse stability, and low mechanical strength in HgCdTe materials, the micro-mesa processing of HgCdTe is obviously different from the conventional microelectronic processing and micromechanical machining. Thus, in the presentation of small pixel isolation on micro-mesa array chip, photolithography on concave–convex surface, sidewall passivation and climbing metallization of deep trenches involved in micro-mesa processing technologies, the specificities of these technologies will be demonstrated emphatically, and their differences from conventional semiconductor technologies will be analyzed.

5.2.1 Isolation Technology of HgCdTe Micro-Mesa Array

HgCdTe micro-mesa array isolation technology is employed to separate HgCdTe multilayer heterojunction materials into individual micro-mesa structures, in order to isolate the in situ doped *p-n* junction and keep the connection of a common zone. For the new generation of HgCdTe IR FPA chips, e.g., two-color, multicolor, heterojunction long wave devices and avalanche photodiode (APD), an individual pixel can contain one, two, or more in situ doped *p-n* heterojunctions, which should be isolated by micro-mesa array isolation technology. Figure 5.1 shows the profile structure diagram of n^+-p3-p2-p1-n^+ HgCdTe two-color detector array chip [6, 12, 13, 19, 20]. As shown in Fig. 5.1, the in situ doped p1-n^+ junctions within micro-mesa array are isolated completely by deep trenches. Meanwhile, the common zones of n^+ layer are still connected to ensure the electrical contact between isolated pixel array and common electrode.

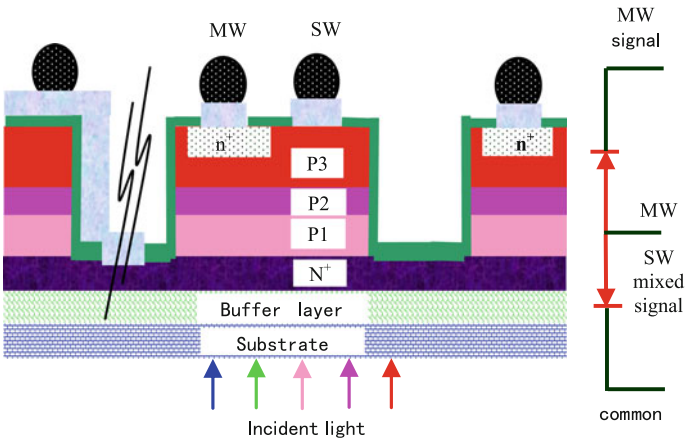


Fig. 5.1 Profile structure diagram of HgCdTe two-color detector array chip

1. Main Characterization Parameters and Methods of Etching HgCdTe

(1) Characterization parameters of etching HgCdTe [2, 6, 7, 11]

A technology for precise control of isolation trench depth for micro-mesa array isolation is required to satisfy the requirement of in situ doped *p-n* junction isolation and common zone connection. Furthermore, to ensure high duty cycle of the micro-mesa array after isolation, it is also necessary that the isolation trenches have high depth-to-width ratio. On the other hand, the characteristics of weak Hg–Te bonds and low technological damage threshold in HgCdTe materials should be taken into consideration Physical or electrical damages should not be introduced into the detailed technological process, so as to ensure that the pixels have normal photoelectric characteristics after isolation [6, 7, 11, 12, 15].

Considering semiconductor fabricating techniques, the etching process isolating micro-mesa pixels involves many technical parameters, such as etching rate, etching profile, etching deviation, selectivity ratio, uniformity, residues, polymers, damage induced by plasma, surface roughness, etc. These parameters are directly related to the quality of the etching process of isolation trenches. So the technical parameters of the etching process will be demonstrated as follows, especially the most important technical parameters [2, 7, 21].

a. Etching rate

The etching rate is defined as the rate of removing HgCdTe material from the surface in the etching process, using nm/min as a unit [21–23]. The depth of the etching opening is called mesa height. Considering the demands of high yield, short fabrication period, and equipment efficiency, the etching rate is expected normally to be high, so it is an important parameter in the monolithic technology. The etching rate is determined by technologies and equipments, e.g., etching material, structure

configuration of etching machine, etching gas, and technological parameter setting. The etching rate can be calculated as

$$\text{Etching rate} = \Delta T/t \ (\text{nm/min}),$$

where ΔT is the thickness of removed material (nm), t is the elapsed time of etching (min).

Usually, the etching rate is proportional to the concentration of etchant. In addition, the etching rates between different zones and batches could be influenced by total area of etching zone, dimension of etching opening, surface status of etching zone, Cd component, etc., even if the identical etching processes are carried out on the same etching machine. Table 5.1 shows the factors influencing etching rates in specified etching processes and the related phenomena and effects. These influence factors and the related effects interact and determine etching rate together. The effective terminal test is very important for the precise control of the isolation trench depth in the HgCdTe micro-mesa array etching process.

b. Etching profile

The etching profile is defined as the sidewall profile of etching zone [10, 21, 24]. There are two kinds of basic etching profiles: isotropic and anisotropic. The

Table 5.1 Influence factors of etching rates in specified etching processes and related phenomena and effects [11, 22]

Influence factor	Related phenomena	Related effects
Total area of etching zone	The larger the total area of etching zone, the more etchant it will consume, which makes the etching rate slower; in contrast, the etching rate will become faster	Loading effect: large area etching could consume more etchant
Pattern size of etching zone	If the pattern size of etching zone is quite small, the etchant cannot get into etching zone easily and the etching product cannot leave etching zone rapidly, which makes the etching rate slower; in contrast, the etching rate tends to be normal	Aspect ratio dependence etching (ARDE), also called microloading effect
Surface status and trench depth of etching zone	With increasing etching time, more and more polymer product in etching zone will hinder the subsequent etching; meanwhile, with the increase of etching, the trench depth of etching zone increases constantly, which makes the etching rate slower	Time nonlinearity of etching rate: the etching rate becomes slower and slower in the course of time
Cd component x of $Hg_{1-x}Cd_xTe$ material	Closely related to the detailed etching process. Usually, the higher the component x is, the slower the etching rate becomes	Vertical nonlinearity: the new generation of HgCdTe detectors employ the heterojunction of the vertical component

isotropic etching is defined as etching in all directions (transversal and vertical) that is carried out at the same rate. It results in the undercutting of etching materials beneath the mask, which could bring the undesirable linewidth loss. In substance, the wet chemical etching is isotropic. The anisotropic etching can be classified into moderately anisotropic (small sidewall angle) and highly anisotropic (vertical sidewall). The anisotropic etchings are mostly achieved by plasma etching, and they will not bring the undesirable linewidth loss mainly.

In order to satisfy the demand of high duty cycle of a micro-mesa pixel and complicated functional structure fabrication on the micro-mesa, the isolating trenches should have high aspect ratio (AR), i.e., the etching process is required to have good anisotropy. It demands that the etching is carried out only in the direction vertical to HgCdTe surface and the transversal etching is suppressed. It is easier to acquire high AR etching profile by dry etching through smaller mask patterns [21].

However, for a high AR patterning opening, the etchant is difficult to enter, while the reaction product is hard to come out. It will lead to obvious ARDE. In order to solve the problem, it is expected that the plasma could be pushed into the high AR opening directionally. If the ions in plasma are directional (vertical to HgCdTe surface), the bottom of patterning opening should suffer bombardment rather than the sidewall of the etching patterns. Hence, the technological parameters need to be optimized constantly to force etchant into the high AR opening in the condition for small undercutting.

c. Selectivity ratio

The selectivity ratio means the comparative difference of etching rate between two materials under the same etching condition [10, 21, 24]. It is defined as the etching rate ratio between etching material (HgCdTe) and another material (usually, the mask). The high selectivity ratio means the etching is only carried out on the material layer A which needs to be removed. For a high selectivity ratio etching process A, the etching can stop at the proper depth. Both the electrical connection layer C and photoresist or mask B are not etched. Commonly, the smaller the size of patterns to be transferred, the thinner the thickness of photoresist is required to be. Therefore, to ensure key dimension and profile control, the high selectivity ratio is indispensable in advanced semiconductor etching process. Particularly, for micro-mesa trench isolation, the selectivity ratio between HgCdTe etching and mask material etching is required to be higher.

The selectivity ratio S_R between etching material A and mask material B can be calculated by

$$S_R = \frac{E_A}{E_B},$$

where E_A is the etching rate of etching material, and E_B is the etching rate of mask material. If the selectivity ratio of an etching process, i.e., the S_R, is small (e.g., 1:1), it means the etching material and mask material are etched at the same rate; if the

selectivity ratio S_R of an etching process is large (e.g., 100:1), it indicates that the etching rate of etching material is hundred times higher than the material not to be etched.

d. Uniformity

The etching uniformity is a parameter judging the etching capability on one HgCdTe chip, or one batch, or between batches. And the nonuniformity of the etching rate can be calculated by the following equation [6, 10, 21, 25]:

$$\text{Nonuniformity } (\%) = \frac{\text{Rate maxium} - \text{Rate minimum}}{2 \cdot \text{Etching rate mean}} \times 100\,\%$$

The uniformity of micro-mesa isolation etching for HgCdTe photoelectric pixels is one of the critical ingredients ensuring the uniformity of chip performance. However, the uniformity of the etching process is difficult to be acquired because of the different densities of patterns on HgCdTe. For example, it is an extreme challenge to achieve good uniformity on the patterns of high density, large interval, and high AR.

The reason why the uniformity is difficult to control is that the etching rate is related to etching profile, pattern dimension, and pattern density. As shown in Table 5.1, the etching rate could be lowered in small patterning openings, and the etching may even stop on small high AR patterns. For instance, the etching rate of micro-mesa isolating trenches with a high AR is lower than that for low AR trenches, which is called ARDE or microloading effect [21, 23]. Because of the small dimension of the etching zone, the etchant cannot enter the etching zone easily and the products also cannot leave the etching zone rapidly, thereby making the etching rate lower until it actually stops. In order to improve uniformity, the ARDE on a HgCdTe surface during etching process must be reduced to a minimum. At the same time, when designing the etching patterns of HgCdTe detector chip, the dimensions and densities of etching patterns should be distributed uniformly.

e. Residue

The etching residue is the unwanted material remaining on the HgCdTe surface after etching [6, 10, 11, 21]. It usually covers the chamber inner wall or the bottom of etching patterns. There are many reasons for the generation of residue, e.g., the surface contamination before etching, improper proportion of chemical etchant (e.g., etching rate is too high), the contamination in the chamber, and/or the impurities distributed unevenly in the films. Etching residues have different names, including long fine strip, shield, coronal, and grid. The long fine strips are fine etching material residues not removed completely and have electrical activity, which can cause undesirable short circuits between patterns. However, the etching residues can be removed during the removal of photoresist or by wet chemical etching currently.

f. Polymer

During the etching process, some polymers could form by the combination of etching products and carbon in photoresist or chemical etchants. It can cause the internal components in the chamber of the equipment to be covered with polymers [6, 10, 11], thus the chamber should be cleaned periodically to remove the polymers and even the components that cannot be cleaned must be replaced. If deposited on the chip, the polymer could hinder the further etching, and then the time nonlinearity effect of etching rate could become obvious and etching stagnation might appear. It is undesirable for the technological researchers.

Sometimes, the formation of a polymer is intentional, which is to form a corrosion resistant film on the sidewall of etching patterns thereby preventing transversal etching. Because the polymer can prevent the sidewall from being etched, the directionality of etching is enhanced. Therefore, highly anisotropic patterns can form so that the good control of critical dimension of patterns could be achieved [10, 21]. However, it is very difficult to realize in etching HgCdTe.

Accordingly, the etching process of micro-mesa isolation in HgCdTe detector array demands the optimization of type, proportion, and physical etching effect of etching gas, in order to realize the reasonable control of the balance between deposition and desorption of polymer, so that the influence of polymer deposition on etching process can be suppressed.

g. Plasma induced damage

In the process of etching, plasma-induced damages are formed easily on the surface of HgCdTe etching zone due to the plasma containing ions, electrons, and excited molecules [6, 7, 11]. The plasma-induced damage of HgCdTe material is quite different from that of the other semiconductor materials, such as silicon or GaAs. It contains not only the lattice damage caused by physical collision, but also the electrical damage caused by physical collision and chemical etchant movement [6, 7, 11]. This is directly related to the characteristics of HgCdTe material and the etchant used in dry etching.

The formation of electrical damage is due to the distinctiveness of HgCdTe material. That is, the Hg vacancy defects are p-type intrinsic dopants, while the Hg interstitial defects are n-type intrinsic dopants. Therefore, the physical collision between plasma and HgCdTe material can cause some Hg atoms to diffuse into the material and then become Hg interstitial, thereby resulting in the transformation of HgCdTe material from p-type to n-type after etching [7]. Etching of HgCdTe material must be achieved by plasma based on methane/hydrogen gas in which there are a large number of H atoms and H ions. Most of H atoms and ions could react with the atoms on the surface of HgCdTe material and the formed volatile compound can be removed from the surface of the etching zone and the chamber. However, after reaching the HgCdTe surface, a small amount of hydrogen may enter the material even into the deeper portions without reacting with the surface atoms of HgCdTe material. It is well known that the electrical activity of hydrogen is good, so the hydrogen could combine with the Hg interstitials, Hg vacancies, and

other acceptor dopants. Accordingly, the electrical property of the material may change, which leads to an increase in the thickness of the electrical damage layer [6, 7, 11, 26].

For the dry etching process of HgCdTe micro-mesa pixels, the etching damage is undesirable, especially the change of electrical property. Hence, the suppression and recovery of electrical damage induced by dry etching attracts much attention in micro-mesa isolation technologies.

h. Surface roughness [21]

The etching surface roughness is usually directly related to the energy of the etching plasma and the concentration and proportion of etchant. If the etching surface is too rough, the contact between the dielectric passive film and the surface of micro-mesa isolating trenches could be influenced so that the dark current, 1/*f* noise, and detectivity of micro-mesa array chip are also affected [6, 7, 11]. Therefore, during the micro-mesa isolation etching, the plasma energy and etchant concentration should be reduced to that just required to achieve a smooth surface.

(2) Main methods of etching HgCdTe

HgCdTe is a II-VI group compound semiconductor material, and its detection of IR energy in different bands can be realized by adjusting the stoichiometric ratio of Hg in the material [5, 19, 27]. Due to the weak Hg–Te bonds, the Hg in HgCdTe is unstable and sensitive to external particle bombardment and temperature. Usually, in the p-type zones in the device, when particles with a certain energy bombard the material surface, some Hg bonds in the surface layer may be broken and the Hg atoms displace forming a surface inversion [6, 7, 10, 11]. In addition, the Hg bonds in the material can be broken at high temperatures so that some Hg atoms could escape from material surface forming a surface inversion. In virtue of the distinctiveness of HgCdTe material, the dry etching process must satisfy the demand of ultralow damage to device. So low-damage etching is one of the key technologies for the fabrication of HgCdTe detector chip.

At present, the IR detectors with novel structures, such as high density FPAs, APDs, two-color and multicolor devices, are isolated by the high duty cycle trenches with little damage on surface and sidewalls [6, 7, 28]. Accordingly, the etching process is the precondition for achieving pixel isolation. If the material is damaged and forms a surface inversion during the etching process, an inverse parasitic diode will be produced in the surface layer of the pixel [6, 7, 11], resulting in the degradation of device performance. Therefore, in order to ensure the device performance of micro-mesa array, the etching damage induced by isolation etching technology should be minimized [7, 29, 30]. Meanwhile, good etching morphology should be acquired under the precondition of ultralow damage and low processing temperature. The low damage dry etching technique of micro-mesa devices consists of low damage dry etching, heat treatment for recovery of etching induced damage, surface cleaning, etc.

The conventional etching process for HgCdTe material includes wet etching and dry etching. Due to the disadvantages of poor anisotropy, severe undercutting, poor repeatability of etching depth and bad uniformity, the wet etching process is not suitable for the formation of mesa devices with fine linewidth. In this section, the etching techniques adopted in the fabrication of HgCdTe devices are summarized, and the information on the HgCdTe etching of low energy and high plasma density will be emphasized [2, 6, 7, 10, 11].

a. Development of HgCdTe devices and etching technique

The fabrication process of HgCdTe devices has been always full of challenges owing to the frangibility of HgCdTe material [6, 7, 11]. Compared with the wide band gap semiconductors (Si, GaAs, GaN), the HgCdTe material has the problems of low damage threshold and high defect density. Each step of the fabrication process (including etching) for HgCdTe devices should be carefully monitored. In the fabrication process of devices, the etching that copies the photoresist patterns to the chip surface is a critical step.

Recently, HgCdTe FPA techniques have developed rapidly. Although the reduction in the critical dimension of HgCdTe devices may not follow Moore's law, not similar to Si memory, the increase in the number of the pixels in HgCdTe scanning or staring FPAs follows this law similar to Si devices [5, 6, 19, 31]. Dry etching has high anisotropy, but the etching of HgCdTe material is a complicated process. In order to understand the mechanism of dry etching, the information in the fields of chemistry, physics, and electronics is necessary. The ingredients influencing etching performance include etching rate, selectivity, uniformity, repeatability, microloading effect, surface roughness, device damage, critical dimension, and profile control. Therefore, a great number of technological parameters are related to the etching process, such as activation energy, adhesion (dynamics in gas phase etching), etching rate, etchant component, surface inversion, and lattice damage are necessary to accumulate through many experiments.

The etching process is an important part in the fabrication of HgCdTe IR FPAs. Figure 5.2 shows various wet etching and dry etching techniques adopted in the fabrication of HgCdTe devices up to now [7]. The critical dimension of the first generation of linear detector arrays is 50–60 μm, so that the conventional wet chemical etching can satisfy the demands of etching process. Although wet chemical etching techniques are applied since the growth of materials, so far the most widely applied wet etchants are only bromine alcoholic solution (Br_2/methanol) and bromine acid solution (Br_2/HBr) [7, 32]. Due to the disadvantages of isotropy and poor uniformity, wet etching is difficult to satisfy the demands of the manufacturing industry on controllability of etching parameters. The second generation of photoelectric diode 2-D arrays, high-performance FPAs, and superlattice structures are realized by the micro-mesa isolation in a pixel array. Also, the development of the dry etching technique has been promoted by the demands of high duty cycle, smooth etching surface, and high uniformity for isolating trenches. Subsequently, a variety of plasma etching techniques appear, such as Methyl glow

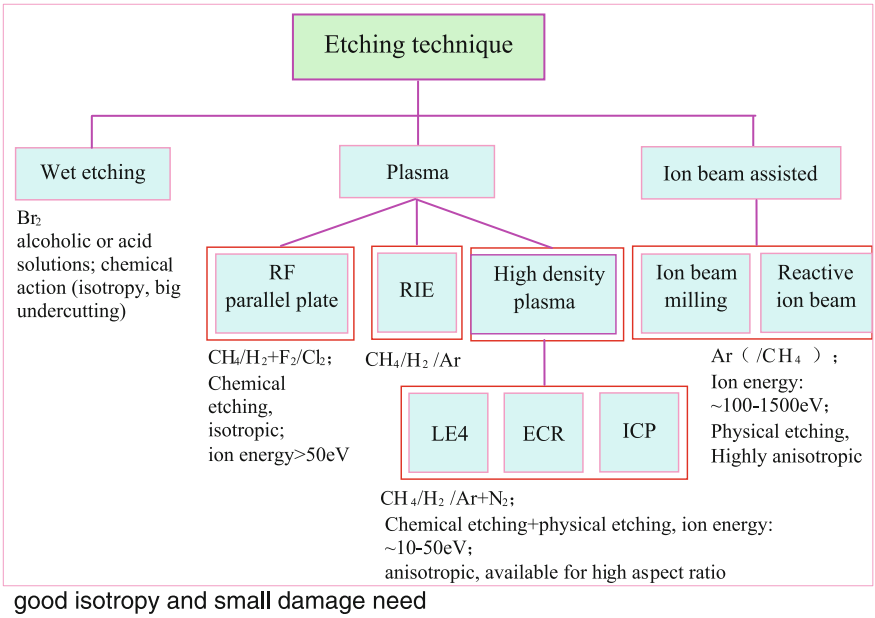


Fig. 5.2 Various wet and dry etching techniques adopted in the fabrication of HgCdTe devices up to now

discharge etching based on Cl_2, CH_4, H_2, and F atoms, reactive ion etching (RIE) based on CH_4 and H_2, the magnetic enhanced electron cyclotron resonance (ECR) and electron cyclotron resonance RIE based on H_2/Ar. Currently, the developing third-generation IR detectors, such as two-color detectors, APDs, and hyperspectral arrays, are required to possess the isolating trenches with high AR (a pixel isolation of 20 μm needs trenches 15 μm deep). Meanwhile, the etching of APD loopholes and damage detectors demands very small surface damage. In the nanostructure devices, the high density patterns with good isotropy and small damage need to be achieved by the enhanced HDP reactor.

b. Wet chemical etching methods

The information on the etchants adopted in the wet chemical etching of HgCdTe is rarely reported in literature. The conventional etchants of HgCdTe are Br_2-based chemical solutions [2, 7]. These solutions, such as Br_2/methanol, Br_2/butanol, Br_2/ethanediol, are usually used for the etching of CdTe and HgCdTe. The process of wet chemical etching is that the semiconductor elements are first oxidized, and then the products are dissolved by the proper solutions. The Br_2/HBr/ethanediol etchant is used to etch HgCdTe mesa and can give smooth (the roughness is 2–3 nm) and nondamage surface and sidewalls repeatedly in the vertical and transversal directions. Wet etching is superior to dry etching with better chip architecture and no electrical damage. In a large number of papers on wet etching,

the surface analysis results of wet chemical etching acquired by the methods such as photoluminescence (PL) spectra, XPS, Rutherford backscattering, and Auger electron spectroscopy (AES) have been reported [7]. In this paper [7], wet chemical etchants are summarized, and surface features after etching and etching rates of HgCdTe are discussed. Etching $Hg_{1-x}Cd_xTe$ materials with given component x requires suitable oxidizing (neutral/acidic) solutions for etching under the strict controls of technological conditions. The main processing variables are etching thickness, operating temperature and etching time. Sometimes, the stirring of the etching solution is an ingredient of the diffusion controlling reaction. However, the wet etching is usually a rate-controlled reaction. Due to the inhomogeneity and isotropy of etching, a large etching deviation is easy to occur. Thus, the wet etching is usually only applied in the surface treatment of HgCdTe.

c. Dry etching and induced damage

(a) Dry etching methods

During the late 1980s and early 1990s, the dry etching technique of II-VI group compound semiconductors began to develop. It was realized gradually that the etching of fragile materials by ions with lower energy ($\sim$100 eV or lower) is indispensable for reducing ion damage. At present, the methods adopted in the dry etching of HgCdTe include IBE, ECR, RIE, and ICP [6, 7, 11, 33]. In fact, the development of various etching techniques of II-VI group compound semiconductors is all obtained directly or indirectly from the dry etching techniques of IV and III-V group compound semiconductors. Most of the etching methods of IV and III-V group semiconductors are based on the halogen plasma (F_2, Cl_2, Br_2, and I_2). The etchant is required to remove three components of HgCdTe material simultaneously, but the chemical volatility of the products is inconsistent. Initially, halogen (F_2, Cl_2) glow discharge was used to etch II-VI group semiconductor compounds. However, the vapor pressure of cadmium halide is low, so that the chemical etching of HgCdTe cannot be completed using Cl_2 and F_2. Moreover, these etchants and their reactive products have causticity and toxicity. Subsequently, chemical etchants such as CH_4/H_2 and $CH_4/H_2/Ar$ attracted attention in semiconductor plasma etching. Although the design of a plasma reactor is different from RF glow discharge, RIE, magnet enhanced RIE, low-energy electron enhanced etching (LE4) and ECR systems, in fact the etchant hydrocarbon/H_2 has been applied widely [7, 23, 29, 34–36]. All of these plasmas are called low pressure (10^{-4}–10 Torr) or cold plasma, and their charged ion and free radical densities are lower than that of neutral plasma. Because the energies of ions and free radicals are very low, these etchants can be applied in the low-damage etching of II-VI group compounds. The ion beam dry etching methods include reactive ion beam etching (RIBE), chemistry-assisted ion beam etching (CAIBE), and ion beam-assisted etching (IBAE). The ion energies of these etching methods are high (100–1000 eV), but they can be applied in etching fine patterns of HgCdTe due to good anisotropy.

High AR etching is required in the fabrication of multicolor detectors and APDs, i.e., the depth of trenches should be 10–15 μm and the width should be only

3–5 μm. This requires good anisotropy of the etching process to achieve a high filling factor. Therefore, for the above-mentioned purpose, the plasma etching techniques such as ECR have been applied extensively.

The problems to be settled in the etching process include [7]:

- Low roughness of sidewall and surface;
- No changes in the chemical proportion and photoelectric characteristics of etching surface;
- Avoid the excessive deposition of polymers and redeposition of products;
- Avoid the reverse layer in the etching surface layer.

The technological parameters such as gas phase components of plasma, chamber pressure, input power of ECR or RF source, self-bias, incident ion energy, and chip temperature can affect the above-mentioned etching standard of HgCdTe. Recently, Stoltz et al. [37] acquired the experimental results of the influence of photoresist thickness, sidewall angle, ion angular distribution (IAD), and ion energy distribution (IED) on the AR of etching trenches.

(b) Dry etching induced damage

In the research of dry plasma etching techniques of HgCdTe, some changes in material characteristics should be noticed [7]:

- Change in conduction type
- Change in chemical proportion and introduction of defect
- Surface roughness
- Deposition of polymers

Due to weak Hg–Te bonds and poor volatility of CdTe, the damage threshold of HgCdTe material is very low. In order to acquire anisotropic cross section and maintain the surface chemical proportion, dry etching requires the assistance of physical action. Also, to overcome the disadvantages of the low volatility of Cd and help the surface desorption of assisted etching products, a certain extent ion-induced reaction is necessary. However, during the dry etching process, the bombardment of ions can change the electrical and optical characteristics of HgCdTe materials [7, 11, 18, 20, 25, 28, 33].

During the ECR or ICP etching of HgCdTe, the low-energy Ar ion can produce a sputtering effect under high DC bias, thereby enhancing the escape of Hg. Under the operating bias, the damage depth caused by sputtering is less than 10 nm. By the analysis of reflection high-energy electron diffraction (RHEED) and low-energy electron diffraction (LEED), it was discovered that the etching surface of ECR for a crystal surface is different from the noncrystal surface of III-V group semiconductors. However, the twin and small-angle crystal surface can be observed, and all the other defects can be recombined to form a crystal surface. Although the chemical proportion and the change in transportation properties for the etching surface layer can be analyzed comfortably, the features of the sidewall may be different and difficult to characterize. The issues such as sidewall damage, change in

n-type doping, and introduction of minority carrier recombination center can reduce the related device performance. When the trenches are very narrow or the trenches are very deep and the etching lag effect appears, the *I–V* characteristics decrease with breakdown voltage being reduced and the inverse current increased.

Dry etching can reduce pixel size and greatly improve the duty cycle of pixels, satisfying the demands of preparing high density detector arrays [7]. Hence, for the design parameters (etching depth, etching width, mesa profile, and etching time) of specified pixel size, the dry etching process must be optimized to achieve high-performance photoelectric diode detector arrays.

2. ICP etching process

(1) Low energy with HDP etching equipment

The ECR-enhanced RIE equipment is supplemented with a microwave source generated plasma on the basic conventional RIE. The ECR plasma arises from the interaction of the microwave source and a magnetic field. Since entering the magnetic field from the chamber wall of microwave source, the electrons act a cyclotron motion under the role of the Lorentz force and collide with gas molecules, thereby generating plasma. When the motion frequency is equal to the microwave frequency, the probability of collision increases due to the electron/cyclotron resonance, so that the density of the plasma is increased. The energy of the plasma acting on the etching samples is controlled by another RF source emanating from the bottom electrode. By virtue of the advantages of high anisotropy, small etching damage, high etching rate, etc., the ECR-enhanced RIE system has been widely applied in the isolation etching of micro-mesas in HgCdTe high-density FPAs [7, 23].

ICP plasma is another HDP technique following the ECR source, using a RF source connected with a coil instead of the microwave source and magnetic field of the ECR [23]. The forming principle of plasma is that an RF electrical field is induced by a RF magnetic field generated from the RF coil in the chamber, which can accelerate electrons and then produce high-density ICP plasma. Compared with ECR source, the ICP source has the advantages of no requirement for a high magnetic field setting, simple structure, operating stability, and better uniformity of plasma. It makes the ICP-enhanced RIE technique more popular in the plasma etching of low-energy and high-density material, and is especially applied extensively in the etching of HgCdTe micro-mesa arrays for a new generation of photoelectric devices.

Figure 5.3 shows the structure diagram of a ICP-enhanced RIE system from Oxford Inc. As shown in Fig. 5.3, the system has two RF sources: one source is connected with an inductance coil and used for generating ICP plasma and controlling the density of plasma, while another one is connected with the bottom electrode and used for controlling the energy of plasma during etching. For the inductive coupling, the electrostatic shielding layer between the cryogenic inductance coil and the chamber is used to avoid capacitive coupling. This is a patented technique of ICP source presented by Oxford Inc.

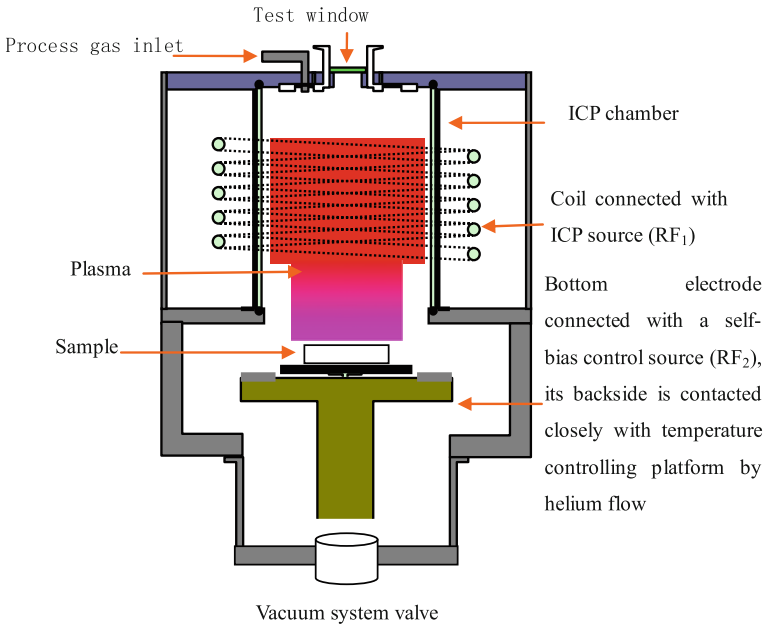


Fig. 5.3 Structure diagram of ICP enhanced RIE system of Oxford Inc. (provided by Oxford)

A conventional RIE system has only one RF source. In order to achieve high density of plasma and high etching rate, the energy of plasma must be increased. The ICP enhanced RIE system has two RF sources, separating the plasma generating source from the source controlling the etching energy of plasma. This configuration can realize plasma etching of low-energy and high-density materials. Accordingly, the ICP-enhanced RIE technique can not only ensure high etching rate through increasing the density of plasma, but also suppress the etching induced structure and electrical damage by controlling the energy of plasma. This is important for the isolation of micro-mesas in high density HgCdTe pixel arrays achieved by dry etching technique.

(2) Technological process, parameters, and testing methods of ICP etching

a. Technological process of ICP etching

In fact, the RIE process of HgCdTe based on the mixed gas of CH_4/H_2 is an inverse process of metal oxide vapor phase epitaxy (MOVPE) used for preparing the HgCdTe material [2]. This process is very complicated and involves some dynamic equilibrium issues, such as physical etching effect versus chemical etching effect, deposition versus desorption of reactive products, deposition versus desorption of polymers, etc.

For the etching of II or VI group elements in compounds using CH_4/H_2-based plasma, the key process is that the Methyl groups and hydrogen atoms derived from

the decomposition of CH_4 and H_2 molecules colliding with high-speed moving electrons diffuse to the bottom electrodes and react with Te, Cd, and Hg atoms on the surface of HgCdTe. Then the volatile products, such as $Te(CH_3)_2$, $Cd(CH_3)_2$, $Hg(CH_3)_2$, and TeH_2, are generated. Moreover, the content of $–CH_3$ group and H atom in plasma is related directly with the etching rate of HgCdTe.

When the $–CH_3$ groups are derived from the decomposition of CH_4 molecules, $–CH_2$ and –CH groups are also generated, which are the precursors for producing hydrocarbon polymers. In order to suppress the formation of these precursors, the CH_4 gas enters the chamber through the bottom gas inlet, while the other process gases enter the chamber through the top inlet. This configuration reduces the distance between the generated $–CH_3$ groups and the sample surface and lowers the probability of producing $–CH_2$ and –CH groups from the collision of plasma and $–CH_3$ groups, thereby suppressing the deposition of polymer. In the development of equipment, the two-inlet method is applied.

In addition, the introduction of the nitrogen gas as an etching process gas can also suppress the generation of polymers [38]. This is because the *N* atoms generated from N_2 gas in plasma can react with $–CH_2$ and –CH groups and suppress the generation of polymers by reducing the concentrations of polymer precursors. Meanwhile, the *N* atoms can also react with H atoms in plasma to produce NH_3 thus reducing the concentration of H atoms in plasma, so that the surface roughness after the RIE process is reduced and the problem of Cd-rich surface is solved.

Accordingly, the mixed gas of $CH_4/H_2/N_2/Ar$ is used for the RIE process gas in the etching of HgCdTe by the ICP-enhanced RIE system [2, 7]. CH_4 and H_2 are chemical etchants and play a chemical etching role, N_2 is used to reduce the polymer precursors, e.g., $–CH_2$, –CH, C_2H_4, and C_2H_2 and optimize the etching surface, Ar is a diluent which can make plasma stable and physical etching dominant to optimize the equilibrium between deposition and desorption of polymers and etching products and adjust the relative dominance of physical etching and chemical etching.

b. ICP etching technological parameters

Similar to the ECR etching technique, ICP etching is also influenced by the components such as type and proportion of process gas, chamber operating pressure, input power of ICP source, RF power of bottom electrode (self-bias), chip temperature, selectivity ratio of mask, and the pattern profiles of mask. These components are quite significant for the micro-mesa etching of HgCdTe and their influences on etching and corresponding reasons are introduced as follows:

- Components and proportion of process gas: related directly with the type and volatility of etching products, the roughness and chemical proportion of etching surface, dynamic equilibrium between physical etching and chemical etching, and dynamic equilibrium between deposition and desorption of polymers;
- Input power of ICP source: critical component determining decomposition level and ionization of process gas, which is related directly with plasma density and can influence type, energy, and angle distribution of particles in plasma;

- RF power of bottom electrode: critical component determining physical etching and etching energy of ions in plasma, and is closely related to the etching induced electrical damage in HgCdTe material. May influence the dynamic equilibrium of physical etching versus chemical etching, deposition versus desorption of products, and deposition versus desorption of polymers;
- Operating pressure: directly related to density and etching energy of plasma. Usually, the high operating pressure could increase plasma density, but reduce the mean free path of ions, resulting in the reduction in particle energy of plasma; meanwhile, a high operating pressure could also suppress the volatilization of etching products, so that the etching products cannot be expelled and redeposited;
- Etching selectivity ratio of the mask: related directly to the etching deviation. If the selectivity ratio of etching material and mask is low, the mask layer should be very thick. It results in the ARDE effect and rough sidewall;
- Pattern profile of mask: it is related to the incident angle of ions, and its influence changes constantly during etching;
- Substrate temperature: related to the chemical reaction rate of etchant and HgCdTe, the volatilization rate of etching products, and the deposition rate of polymers. Meanwhile, the substrate temperature is also related with levels of etching induced damage, which is because the H atoms entering the material surface without reaction with HgCdTe and the Hg interstitials generated from etching can diffuse deeply into HgCdTe, thereby inducing a very thick damage layer.

In fact, the above technological parameters are interactive and belong to a complicated technological system. During etching process, DC bias is a very important etching state parameter [6, 11, 28]. It can not only reflect the etching energy of plasma related with technological parameters such as ICP input power, RF power, and operating pressure, but also can be used to estimate etching state such as polymer deposition, etching product deposition, and etching rate according to its change during etching.

c. Test methods of ICP etching

The process of ICP etching for HgCdTe is complicated, and there are many factors influencing the process. Therefore, in order to acquire dry etching process for micro-mesa isolation of HgCdTe FPAs, its technological conditions must be optimized through a large number of experiments, in which the surface chemical proportion, surface morphology, trench profile, and electrical damage need to be tested. The specific experimental methods are:

- Surface chemical proportional
  - Microzone XPS: used to determine type, chemical proportion, and percentage of elements existing on the surface of etching zones;
  - AES: the functional relationship between component and depth of samples can be determined by auger electrons with characteristic energies emitted from the sample surface. It is sensitive to C and O elements.

- Surface morphology
  - Scanning electron microscopy (SEM): check surface morphology of etching sample and observe whether the etching residuals or polymers are left;
  - Atom force microscopy (AFM): check surface roughness of etching zone.
- Optical characteristics
  - Room temperature PL: check energy levels of dopant defects introduced by etching;
  - X-ray diffraction (XRD): check the dry etching influences on the physical properties of crystals, e.g., lattice constant, double crystal diffraction peak half width.
- Electrical characteristics
  - *I–V* characteristics test: estimate the surface layer of etching zone transforms (including inverse) by checking the *I–V* characteristics of etching zone and non-etching zone;
  - LBIC: to acquire the transverse distribution of etching damages or etching damage induced inverse zones by checking the LBIC signal from the surface of etching samples;
  - Differential Hall: to acquire the longitudinal distribution of etching damage induced inverse zone by checking the dependence of the Hall concentrations on etching samples in depth;
  - Microzone surface potential method (SPM): check the surface potentials of etching induced electrical damage zone and nondamage zone.
- Etching rate
  - Profilometer: only check the etching patterns with low ARs;
  - Confocal microscopy: check the etching patterns with high ARs.

Since most of these test methods are the common techniques, there is no need to illustrate them here. Based on material chip technologies, the laser beam induced current (LBIC) checking of etching induced electrical damages and the *I–V* characteristics measuring methods will be presented in detail.

(a) LBIC measurement

LBIC is a nondamage, high resolution, facile-measuring method which can acquire the spatial distribution of the electrical field in the entire sample [6, 11, 26, 30]. When the laser beam is focused on the sample surface, photon-generated carriers are created, and then are separated by the local electrical field or the field in the other locations through diffusion. The local LBIC signal is collected by two remote electrodes. Thus, by linear or planar laser beam scanning, the relationship between laser-induced current and spatial location of the sample is obtained, and the spatial distribution of the electrical field on the surface of sample is determined. The detailed operating principle of LBIC method is mentioned in Sect. 2.3.

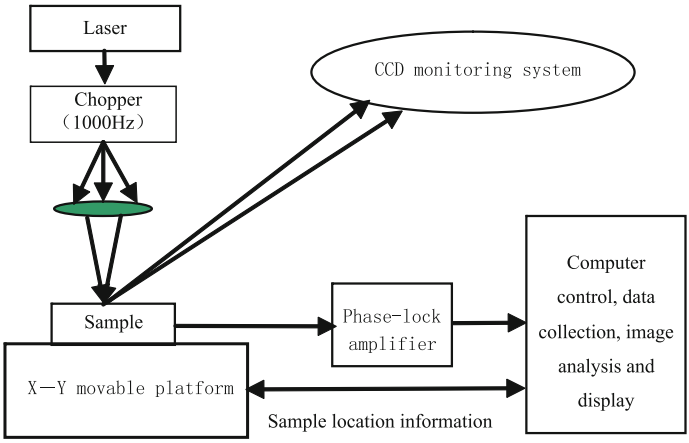


Fig. 5.4 Structure diagram of LBIC measuring system

Figure 5.4 shows the structure diagram of experimental LBIC test system. The system is composed of a laser, chopper, *x*-*y* movable platform, CCD monitoring system in the test location, a phase lock amplifier and signal collector of LBIC signals, and computer system for processing and displaying. In the system, the laser is a 632.8 nm He–Ne laser, and the diameter of the laser spot is approximately 0.5 μm The EG&G Model 650 chopper system is adopted, with the chopping frequency set to 1000 Hz. The stepping precision of the *x*-*y* movable platform is 0.1 μm. In the measurement, the phase-locked measuring equipment (SR830 DSP lock-in amplifier) detection sensitivity is on the order of femtoampere. A low noise current amplifier may be needed to amplify the signal in the primary step.

The electrical damages induced by dry etching can produce a strong electrical field distribution at the interface between the damage zone and the nondamage zone in the surface layer, while there is no electrical field distribution in the other zone. Therefore, the LBIC method can be applied to measure the electrical field distribution induced by etching, and the spatial location of interface between damage zone and nondamage zone can be obtained indirectly.

(b) Measuring method of *I*–*V* and *R*–*V* characteristics of electrical damage

The damages induced by dry etching are classified as lattice damage and electrical damage. It is necessary to avoid electrical damage for the micro-mesa isolation of high-density FPAs. In order to assess the dry etching process without electrical damage, the characterization of electrical damage induced by etching is indispensable. Thus, the visualized and reliable measuring method is the measurement of *I*–*V* and *R*–*V* characteristics of etching induced electrical damage layer and nondamage layer [6, 11].

Figure 5.5 shows the diagram of a measuring method used for checking the thickness of etching induced electrical damage layer. The material chip technique is

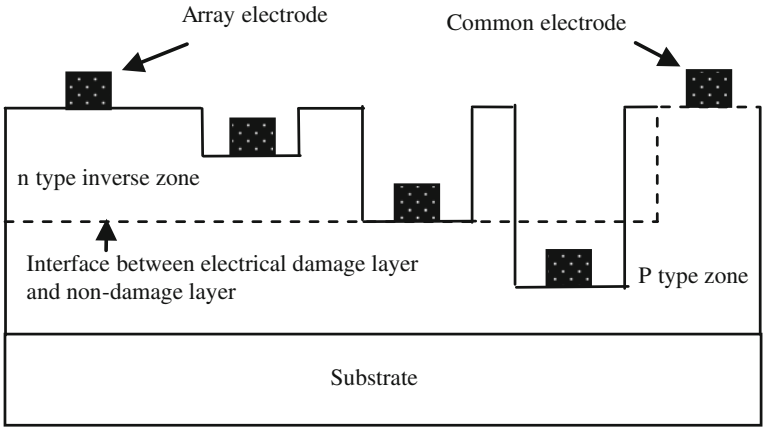


Fig. 5.5 Diagram of measuring method of *I–V* and *R–V* characteristics for dry etching induced damage

well adopted in this method. As shown in Fig. 5.5, the photoelectric diode arrays with different thickness of damage layers and no damages are prepared on the same etching sample, and the thickness of electrical damage layer is estimated by measuring the *I–V* and *R–V* characteristic curves. The thickness of different damage layers in the array is controlled by wet chemical etching with no electrical damage. When the electrical damage layer is removed completely by wet chemical etching, that is, the electrode of the photoelectric diode is located at the interface between electrical damage layer and nondamage layer, the *I–V* characteristic curves of this location should be linear. Accordingly, the thickness of the electrical damage layer is the etching depth of the damage layer, which can be obtained by profilometer or confocal microscopy.

Because the photoelectric diodes are acquired through dry etching of the same batch of material and the measured *I–V* characteristic curves are obtained from the etching damage layer and non-etching zone, the method is visualized and creditable. This method has a good potential for measuring the *p-n* junction depth of ion implantation, the longitudinal distribution of in situ *p-n* junction, etc. [6, 11, 28].

(3) Research on etching condition of HgCdTe

a. Influence of etching process parameters on DC bias

The plasma-induced damage is most concerned in many evaluation parameters of etching techniques. For the ICP-enhanced RIE technique, the etching energy of plasma is estimated by the state parameter of DC bias. Therefore, the influences of various technological parameters on DC bias are investigated in this section.

Figure 5.6 indicates the influences of various technological parameters on DC bias. As shown in Fig. 5.6, the DC bias changes obviously with the variation of RF power of bottom electrode and increases rapidly at a high rate. However, the DC bias changes slowly with RF power, when the RF power of bottom electrode is

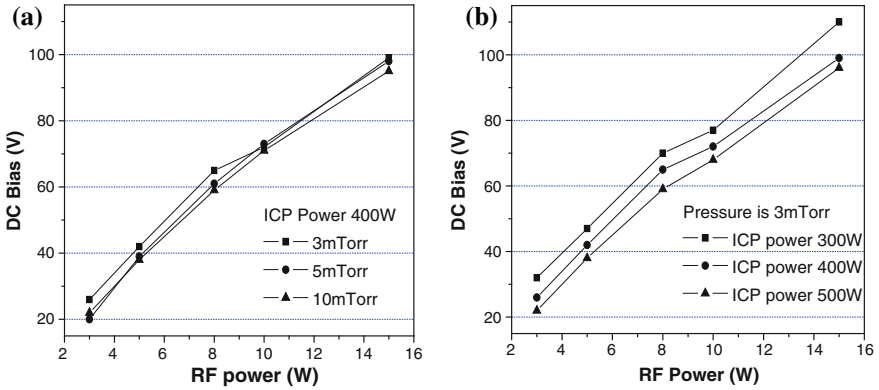


Fig. 5.6 Influences of various process parameters on DC bias. **a** Influences of chamber operating pressure and RF power of bottom electrode on DC bias; **b** Influences of ICP input power and RF power of bottom electrode on DC bias

8–10 W. Figure 5.6a shows the influence of different chamber operating pressures and RF powers of the bottom electrode on the DC bias at the same ICP input power. As shown in Fig. 5.6a, there are three curves intersecting under some identified RF power of the bottom electrode, which indicates that the influence of operating pressure on DC bias is complicated. However, the general trend is that the higher the operating pressure, the lower the DC bias. This is because the mean free path of excited plasma decreases with the increase of operating pressure and the collision of ions in plasma could lead to the loss of plasma energy.

In Fig. 5.6a, it indicates that the DC bias reduces with an increase in operating pressure, however, the operating pressure cannot be too high in plasma etching. When the operating pressure is too high, both the density of plasma and the chemical etching rate become very high. This causes the etching products to not volatilize promptly and then be deposited on the etching surface under high operating pressure. In fact, it not only makes etching residuals increase, but also prevents the next etching step.

Figure 5.6b illustrates the influences of different ICP input powers and RF powers of the bottom electrode on DC bias under the same chamber operating pressure. As shown in Fig. 5.6b, the influence of different ICP input powers on DC bias is very obvious, that is, the higher the ICP input power, the lower the DC bias. This is due to the fact that if the ICP input power becomes higher, the density of ICP plasma will increase, and then the same RF power of bottom electrode can be obtained under smaller electrical field intensity of bottom electrode, thereby reducing the DC bias.

Also, Fig. 5.6b shows that the high ICP input power leads to the reduction of DC bias. However, the ICP input power cannot be set too high, on the one hand, because excessive ICP input power could lead to the $–CH_3$ group being broken into $–CH_2$ and –CH groups forming polymer precursors, which aggravates the

deposition of polymer and affects etching rate. On the other hand, it is because the excessive ICP input power could lead the $-CH_3$ groups to be ionized to CH_3^+, thereby decreasing the etching rate.

b. Etching surface and corresponding etching mechanism

The mixture of methyl compounds and hydrides (CH_4 and H_2 commonly), which can react with HgCdTe material and form volatile products, is usually used for process gas etching HgCdTe. Additionally, some Ar is added for diluting process gases and stabilizing the plasma.

In the reaction, CH_4 and H_2 are cracked into methyl groups ($-CH_3$), H atoms and H^+ ions, which react with Te, Cd, and Hg and generate different volatile products, respectively [7, 38]:

$$\begin{aligned}
&Hg + 2\,CH_3 \rightarrow Hg(CH_3)_2\\
&Te + 2\,H \rightarrow TeH_2\\
&Cd + 2CH_3 \rightarrow Cd(CH_3)_2\\
&Te + 2CH_3 \rightarrow Te(CH_3)_2
\end{aligned}$$

Under different proportions of process gases, the etching morphology and etching rate are quite different. The experimental result indicates that the increase in Ar proportion could reduce etching rate significantly. Ar ions usually play a role of physical bombardment, so the etching morphology becomes poor when the Ar proportion increases. The chemical reaction rate of H^+/H and HgCdTe is relatively low, while the reaction of methyl groups ($-CH_3$) produced by the pyrolysis of methane and HgCdTe is drastic. The etching rate and morphology can be changed by adjusting the proportions of CH_4 and H_2: with the proportion of methane increasing, the reaction rate could increase and the bottom morphology of mesa patterns can be improved. However, when the proportion of methane is very high, a large number of reaction products can remain on the sidewalls of the mesa patterns. The experimental result indicates that flat bottom morphology can be obtained under the high proportion of methane, but obvious products will accumulate on the sidewall. When the proportion of methane is low, the sidewall is clean and the bottom morphology is rough [2].

By microzone XPS and SEM, the influences of various ICP process parameters on the morphology of etching surface can be investigated. It is helpful for the investigations on the reaction mechanism of ICP etching HgCdTe and its micro-cosmic process under different chamber pressures and gas proportions.

Here, a QUANTUM 2000 Scanning ESCA Microprobe XPS system produced by American PHI Inc. is adopted. The base pressure of the analysis chamber is better than 5×10^{-8} Pa, and the excitation source is a monochromatic Al anode (the energy is 1458.6 eV). The beam spot size and the power of the X-ray are 10 μm and 2.5 W, respectively. The sputtering condition is that the energy of Ar^+ is 1 kV, the sputtering time is 1 min, and the area is 2 mm × 2 mm. The surface morphologies of samples are observed by SEM (TS5136, Tescan Inc.). The experimental sample is

the p-type HgCdTe film grown by MBE. The mask materials of etching patterns include photoresist and SiO_2, and the etching gas is $CH_4/H_2/N_2/Ar$. The details of etching samples, processes, and conditions are provided in Table 5.2.

Because the samples for XPS analysis are maintained in air for a while, the contaminations of C and O on the sample surface are serious. In order to acquire precise information, the data after the sputtering of Ar^+ at 1 kV for 1 min are adopted. For Si, surface material of $\sim$2 nm thick is sputtered under the same condition. The sputtering rate on the HgCdTe sample can be estimated by the calibration of sputtering rate with a Si sample under the same condition. Therefore, it can be considered that the data after sputtering still represents the surface conditions of the samples. In microzone XPS analysis, the X-ray beam spot is 10 μm and the beam is focused on the surface etched by ICP. Hence, the obtained data represent the information of HgCdTe surface after etching.

c. Influences of process parameters on etching surface

(a) Chamber pressure [6, 10, 11]

For the samples in Fig. 5.7, photoresist is used for the etching mask, and the etching gases are CH_4:H_2:N_2:Ar = 1:0:5:20. When the chamber pressure is higher (10 mTorr), a large number of etching products are redeposited on the etching surface after desorption, resulting in surface blackening (as shown in Fig. 5.7a). When the other conditions are identical and the chamber pressure is lower (3 mTorr), the surface of samples is bright and the etching result is good (as shown in Fig. 5.7b). It can be seen from SEM images that when the chamber pressure is high, there are residuals on the etching zone, while when the chamber pressure is low, there are no residuals left.

The samples with black surface are analyzed by microzone XPS. According to the bonding energy of C element, the deposit on the surface is deduced to be $C_6H_5(CH_3)$. Because there is no H_2 content in the etching gases of this sample, it is deduced that CH_4 may react with photoresist generating $C_6H_5(CH_3)$. In the XPS spectra of bright surface samples (Fig. 5.7b), there are no obvious peaks for C1s and O1s, and the percentage compositions of C and O are 2.6 % and less than 0.1 %, respectively; nevertheless, the percentage compositions of Hg, Cd, and Te are 33.7, 16.9, and 46.7 %, respectively. It indicates that the etching surface is HgCdTe rather than polymer deposits under low chamber pressure.

Although only chamber pressure is changed in all process parameters under the above etching conditions, the etching results are quite different. When the chamber pressure is low, the products can be expelled without the redeposition of etching products on the surface. Moreover, the etching particles possess a longer mean free path under low operating pressure, so that the dominant physical etching is helpful for the desorption of etching products, resulting in the bright and clean etching surface. However, when the chamber pressure is high, the etching products cannot be expelled promptly and are redeposited on the sample surface, preventing the subsequent etching process. When the surface of etching zone is covered

Table 5.2 Etching samples and process parameter setup [10]

Serial number	Process gas proportions				Pressure (mTorr)	RF power (W)	ICP power (W)	Etching time (min)	DC Bias (V)	Cooling gas flow (sccm)	Etching rate (nm/min)	Mask
	CH_4	H_2	N_2	Ar								
run25	0	2	5	20	3	8	300	60	43	2.3	26.7	PR
run26	1	0	5	20	10	8	300	45	45	2.3	16.3	PR
run28	1	0	5	20	3	8	300	60	65	2.3	22.0	PR
run39	0	1	5			8	300			2.2		PR
run51	1	0	5	20	3	8	500	40		1.8	43.3	SiO_2
run52	1	0	5	20	3	8	400	240		1.8	31.0	SiO_2
run53	2	0	5	20	3	8	400	240		1.8	38.6	SiO_2

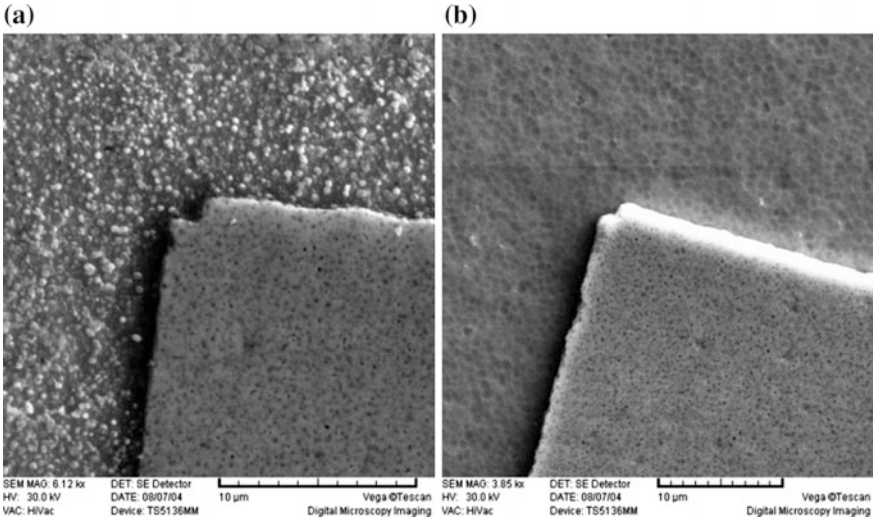


Fig. 5.7 SEM images of sample surfaces under different chamber pressures. **a** Chamber pressure is 10 mTorr; **b** Chamber pressure is 3 mTorr

completely by the redeposited products, the $-CH_3$ groups in the plasma cannot react with HgCdTe surface layer, but react with photoresist mask, so that the polymers are generated and deposited. Thus, the influence of chamber pressure on the surface morphology of etching samples is very obvious.

(b) Different etching gases

i. Influence of the reaction between H_2 and photoresist

Figure 5.8 shows the ICP etching results when photoresist is used for a mask and the etching gas includes H_2 rather than CH_4. The proportions of etching gases are CH_4:H_2:N_2:Ar = 0:2:5:20. As shown in SEM images of Fig. 5.8a, there are many nets in the etching zone. In Fig. 5.8b, some peaks of C1s are discovered by microzone XPS analysis, and the C element exists in the form of organism by analyzing the bonding energy of C. Due to no CH_4 presence in the process gas, and only photoresist and vacuum grease (used for adhering sample on the sample platform) in the chamber, it can be deduced that the C element should be generated by the reaction between H_2 and photoresist or vacuum grease. In the experiment of run51, it is found that if H_2 reacts with vacuum grease, the O content in XPS spectra should be high; however, the O content is very low in this experiment. Hence, it can be deduced that the polymers on the surface may be generated from the reaction between H_2 plasma and photoresist [6, 10, 11].

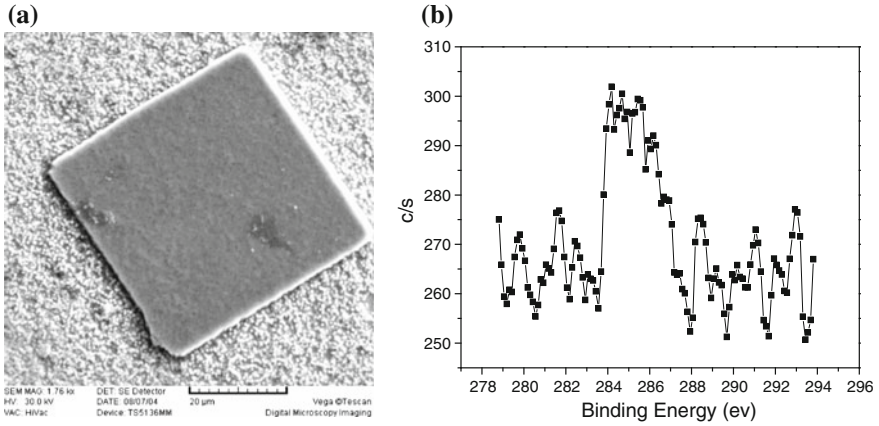


Fig. 5.8 Experimental results of H-based etching. **a** SEM images of etching surface; **b** XPS spectra of C1s

ii. Influence of CH_4 introduction

Figure 5.9 provides the XPS spectra of O1s in all etching samples. As shown in Fig. 5.9, the O content of the etching surface in run51 samples is very high, and the bonding energy is 532.6–533.1 eV. It is probably due to the generation of methyl silicone (532.6 eV) and phenyl silicone (532.7 eV). Since the device contains C element, the bonding energy of C should be measured in order to analyze further the products in run51. Figure 5.10 shows the C1s spectra of the run51 sample. The C bonding energy is 284.85 eV, consistent with the C bonding energy in C–H bonds of polyester film. Therefore, it is likely that the vacuum grease for adhering samples is exposed to the plasma and reacts with CH_4 resulting in the deposition of polymers.

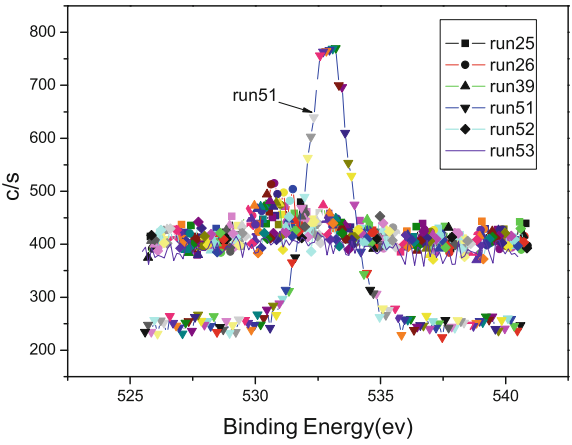


Fig. 5.9 XPS spectra of O1s in all etching samples

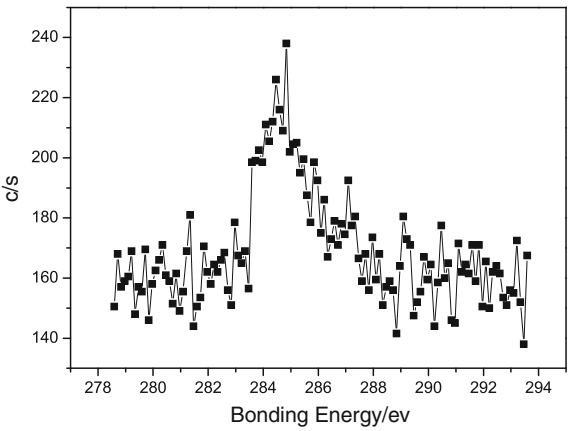


Fig. 5.10 XPS spectra of C1s in run51 sample

(c) RF power of bottom electrode

The RF power of bottom electrode determines the self-bias of etching process, that is, the energy of ions and active groups arriving at the sample surface. It is a significant parameter for estimating the etching process damage to the material. The roles of self-bias include: (1) providing kinetic energy for physical bombardment in the reaction process; (2) providing energy for active groups and chemical reaction.

When the self-bias is high, the physical bombardment becomes strong and the sidewall is abrupt, but the bottom morphology is poor and the damage is obvious. When the self-bias is low, the chemical effect is dominant and the bottom morphology is better, but the abruptness of sidewall is not good, probably forming undercutting, however, the damages are reduced.

(d) ICP power

The ICP power of the inductance coupling coil around the reaction chamber determines the density of plasma in chamber. The density of the plasma becomes higher and the etching rate increases under a high ICP power, while the etching rate becomes lower under a low ICP power.

(e) Substrate temperature

In this process, there are many components affecting substrate temperature, including bottom electrode power, ICP power, He pressure of back surface cooling of the bottom electrode, etc. The substrate temperature is an important parameter in ICP etching process, and its influences on etching rate and morphology are remarkable. The reaction products are not easy to volatilize at low sample temperature and redeposited into the contact hole. A higher substrate temperature can promote the rapid volatilization of products and avoid the redeposition of products. Owing to the limited tolerance of HgCdTe to temperature, the process temperature should be controlled below 80 °C. Through the comparison of etching results at different temperatures, it is found that the morphology of etching patterns becomes

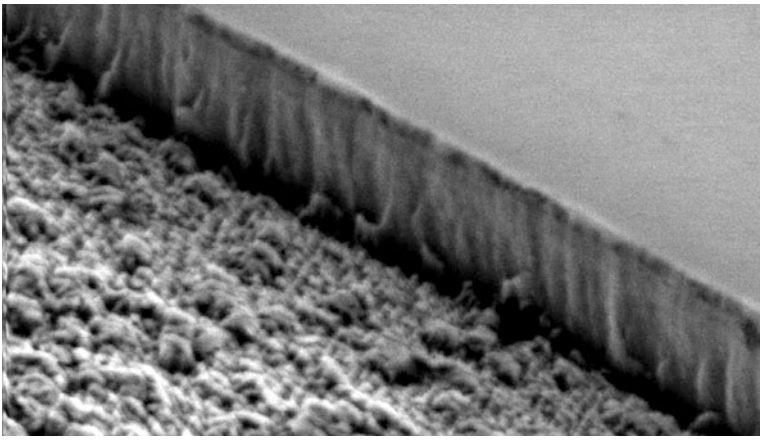


Fig. 5.11 SEM image of etching sample at low substrate temperature

better at 45–55 °C. Figure 5.11 provides the etching results at low substrate temperature. It can be seen from this figure that there are a large number of residuals on the sidewall and bottom of patterns because of the lower volatility of products.

d. Surface residuals and their origin

In the exploration for an optimized ICP etching process, a series of experiments are carried out through varying some parameters, such as etching gas, mask material, chamber pressure, etc. According to the morphology of residuals, they could be classified into: ball, net, grass, and pimple. The analysis indicates that they are generated from the reaction of H_2 and photoresist, high roughness of etching surface and reaction residuals [6, 10, 11].

(a) Ball

As shown in Fig. 5.7a, there are a large number of balls accumulating on the surface of etching zone, so that the etching sample appears black. When the chamber pressure is high, the products cannot be expelled promptly and redeposit on the sample surface, so that the subsequent etching process is hindered. When the redeposited products cover completely the surface of etching zone, a large number of $-CH_3$ groups react with photoresist rather than HgCdTe, and the generated polymers redeposit. Therefore, when the chamber pressure is too high, the balls are coated by polymers on the etching surface, which appears black.

(b) Net and grass

Figure 5.8a illustrates that there are many grainy residuals on the etching zone surface of sample run25. In Fig. 5.8b, by analysis of microzone XPS, some peaks of C1s and no peaks of the O element are observed, and the C element exists in the form of organic matter according to bonding energy. Because there is no CH_4

content in process gas, the grainy residuals shown in Fig. 5.8a are probably related to the reaction of H_2 and photoresist.

In order to determine the relationship between grainy residuals and the reaction of H_2 and photoresist, the samples with SiO_2 mask and no mask are adopted to repeat the same experiment as sample run25. Figure 5.12 provides the SEM images of etching samples with SiO_2 mask and no mask. It can be seen that their morphologies after etching are identical and quite different from that with photoresist mask, though both of them have white grasses (shown in Fig. 5.8a). Accordingly, the residuals on the surface of sample run25 are indeed related to the reaction products of H_2 and photoresist.

In the reaction of H_2 plasma and photoresist, the plasma etchant containing H etches the Te element in HgCdTe first. Due to good volatility of Hg, the rough Cd-rich layer appears on the etching surface. With the increase in the thickness of the Cd-rich layer in the entire etching zone, the reaction between the H plasma and the Cd-rich surface can be weakened, while the probability of the reaction between the H plasma and photoresist forming polymers could increase. Moreover, the polymers deposit on the surface of etching zone constantly, so that the etching surface is covered by black nets.

However, the etching surfaces with SiO_2 mask and no mask are quite different from that of photoresist mask. As shown in Fig. 5.12, there are sparse white grasses on the surface of etching sample, and the surface appears bright. This is due to the appearance of Cd-rich layer on the local regions of etching zone caused by the plasma etchant containing H. These tiny Cd-rich layers prevent the subsequent etching, equivalent to micro mask.

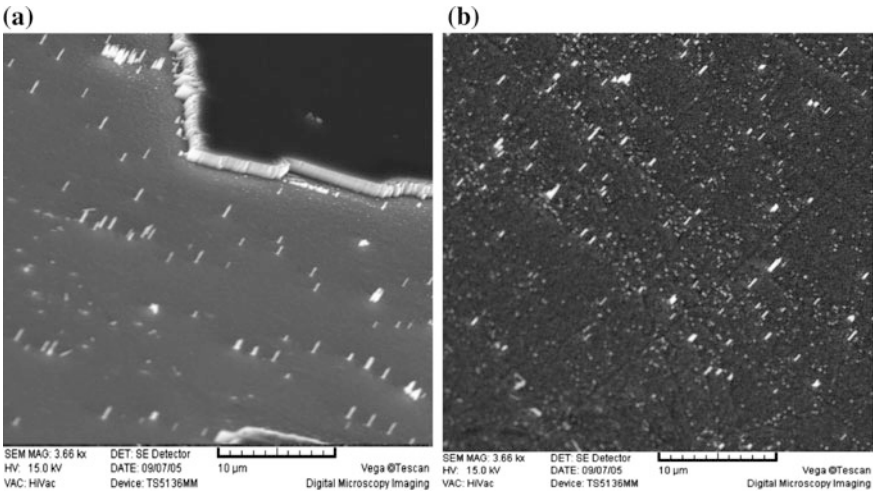


Fig. 5.12 SEM images of etching sample surfaces with different masks. **a** SiO_2, b no mask

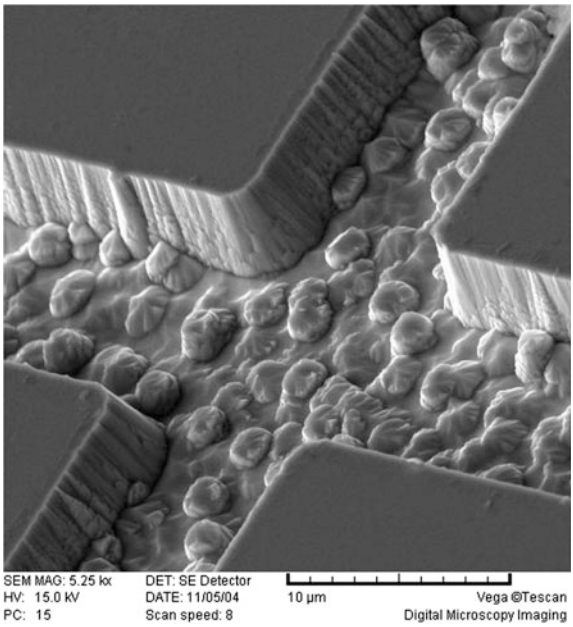


Fig. 5.13 Surface morphology of etching sample with the increase in CH_4 proportion

(c) Pimple

Figure 5.13 provides the morphology of etching samples utilizing a SiO_2 mask and double flow of CH_4. From the quantitative analysis by XPS, it is obtained that the percentage compositions of C, O, Hg, Cd, and Te atoms are 13.1, 1.7, 22.8, 7.5, and 44.9 %, respectively. This indicates that the surface residuals contain some C element and little O element. The sample without a mask was used to repeat the experiment of sample run53, and the acquired morphology of the etching surface is identical with that of run53. Although the etching surface of sample run53 appears good by eye measurement, there are a large number of similar pimples existing on the etching surface in the SEM image shown in Fig. 5.13. In this case, the etching rate is high, $\sim$386 Å/min.

Figure 5.14 provides the surface morphology of etching samples with a reduced CH_4 proportion, ICP input power of 600 W, chamber pressure of 5 mTorr, and bottom electrode RF power of 8 W. In comparison with the sample run53, only the CH_4 proportion and the ICP input power are different. Although the flow of CH_4 gas is lower than sample run53, the increase in ICP input power causes the plasma density to increase, thereby improving the etching rate. As illustrated in Fig. 5.14, the surface morphology of the etching sample is better, and there are a few pimples similar to those in Fig. 5.13.

The size of residuals in Figs. 5.13 and 5.14 is similar, and these residuals should be the residuals that are to be desorbed and expelled out of chamber. The etching rate in run53 is high, so that the residuals that can prevent subsequent etching are not generated on the surface of etching zone. It is shown that the surface residuals

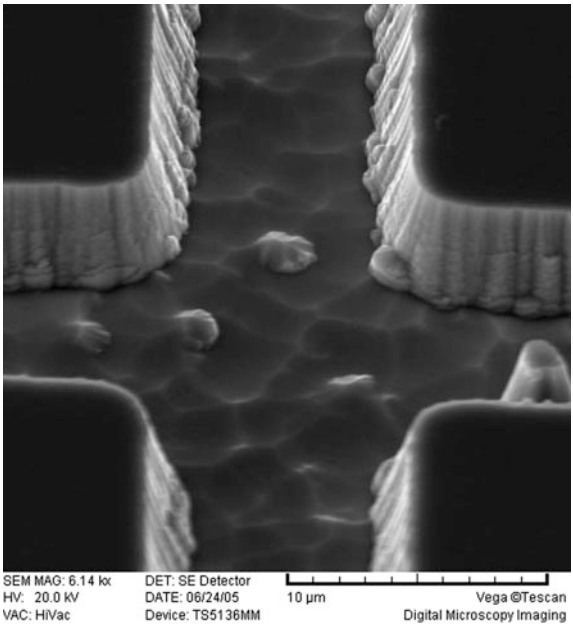


Fig. 5.14 Surface morphology of etching sample with reduced CH_4 proportion and the ICP input power of 600 W

are the etching products generated by the reaction of Hg, Cd, Te, and CH_4 by quantitative analysis of XPS. The pimple residuals in Fig. 5.14 are obviously fewer than that in Fig. 5.13. On the one hand, this is because that the CH_4 proportion is lower so that the etchant concentration becomes low; on the other hand, it is due to the fact that the ICP concentration is higher so that the concentration of Ar^+ in plasma increases, which is helpful for the rapid desorption of etching products.

e. Etching surface and its etching mechanism

Five types of surface morphologies were discovered from the experiments on the influences of various process parameters on etching surface and the analyses on surface residuals and their origin. These morphologies represent five different microcosmic reaction processes [6, 10, 11]. A summary of these processes follows:

(a) Bright etching surface

In this case, the ICP etching process possesses a normal etching rate, neither the redeposition of etching products nor serious polymer deposition appear on the etching surface. During the etching process, the plasma formed by the process gases of $CH_4/H_2/Ar$ reacts with HgCdTe, and then the etching products can be rapidly desorbed and expelled by vacuum system. The deposition and desorption of polymers and etching products are in a dynamic equilibrium. Meanwhile, the physical etching effect under this process condition becomes obvious, which is helpful to accelerate the desorption of etching products and polymers and acquire a bright etching surface.

(b) Surface with pimples

In this case, the ICP etching process has a relatively normal etching rate. Although the etching surface seems bright, there are some etching products, such as Hg, $Hg(CH_3)_2$, $Cd(CH_3)_2$, $Te(CH_3)_2$, etc., not desorbed completely and adhere to the etching surface, as shown in Fig. 5.13. That is, the generation rate of etching products is higher than that of desorption, so that some pimple products are accumulated on the surface of the etching zone. They are remarkably different from the deposition of polymers.

(c) Surface with grass

In this case, the ICP etching process has a relatively normal etching rate, and the etching surface is bright but with obvious grasses. These grasses are caused by a micro mask. The micro mask introduced in the fabrication process of mask patterns can be avoided by improving the fabrication process of the mask, however, those micro masks produced in the etching process are difficult to avoid. For the etching sample in Fig. 5.12, the plasma etchant containing H leads to the Cd-rich layer in local regions of the etching surface, and these tiny Cd-rich zones prevent subsequent etching as micro masks.

(d) Surface with balls

In this case, the etching rate of the ICP process is low, and the etching surface appears black. As illustrated in Fig. 5.7a, when the chamber pressure is high, the etching products are redeposited on the surface of samples after desorption and prevent subsequent etching. On the other hand, a large number of $–CH_3$ groups react with photoresist mask to cause the deposition of polymers rather than HgCdTe. Therefore, the balls are covered by polymers on the etching surface, and the sample surface appears black.

(e) Surface nets

In this case, the etching rate of ICP process is obviously lower, and the etching surface appears black. As shown in Fig. 5.8a, due to H plasma, the rough Cd-rich layer could appear on the etching surface. When the thickness of Cd-rich layer in entire etching zone increases, the H plasma could react with photoresist constantly to generate polymers, which are deposited on the rough etching surface. Hence, the black net-shaped surface with polymer coating appears.

(f) Etching process for good surface

From experiments on different process gas proportions, chamber pressures, and ICP input powers, a stable etching technological condition for clean and smooth etching profiles is acquired. The process gas proportion is CH_4:H_2:N_2:Ar = 1:0:5:20, the ICP input power is 400 W, the RF power of bottom electrode is 8 W, the chamber pressure is 3 mTorr, and the temperature of sample platform is 20 °C. The parameter setting is usually adopted in the etching process mentioned

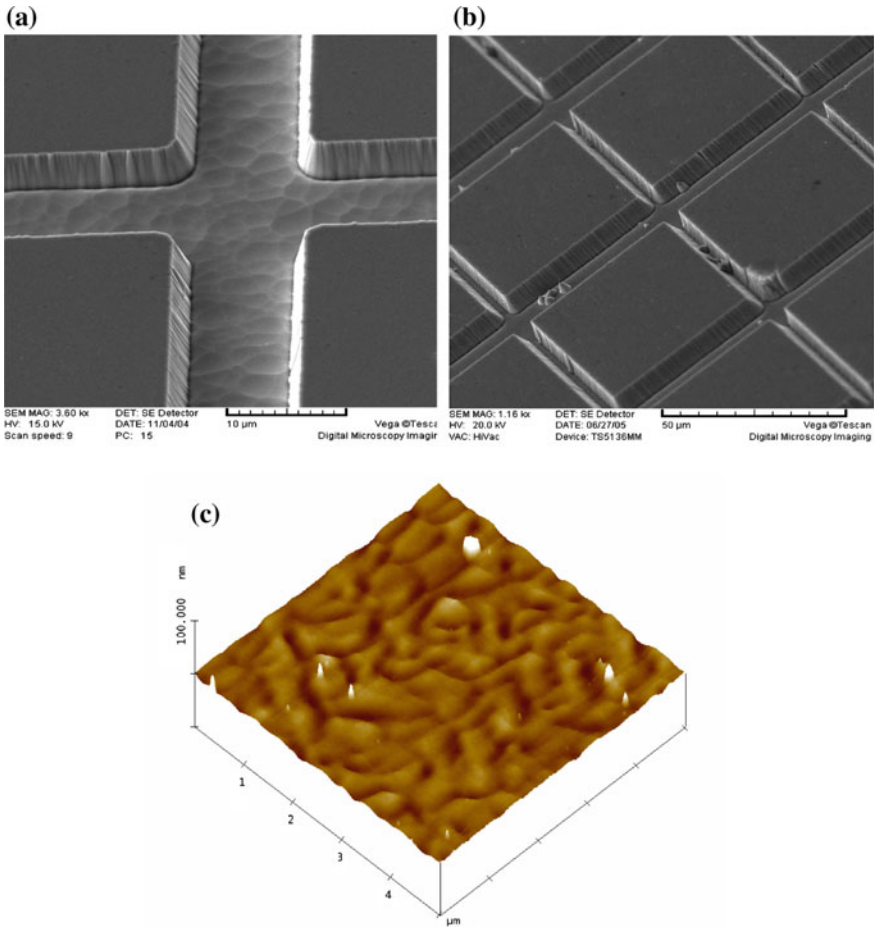


Fig. 5.15 SEM images and roughness of etching surface of different etching patterning openings under stable technological condition. **a** Etching opening of 10 μm; **b** Etching opening of 4 μm; **c** The surface roughness of less than 1.5 nm

later. Figure 5.15 shows the SEM images and surface roughness of etching surface in the above-mentioned etching condition. As shown in Fig. 5.15a, when the etching opening is 10 μm, a clean and smooth etching profile could be obtained. However, as illustrated in Fig. 5.15b, when the etching opening is 4 μm, grasses appear on the surface of the etching zone. They are likely introduced during the fabrication of the etching mask with very small linewidth. Furthermore, it is found that the trenches approach "V" shape, preventing subsequent etching, that is, they are influenced by the ARDE effect [6, 7, 10, 11]. Accordingly, the linewidth minimum of etching process is around 4 μm.

(4) Etching rate of ICP-enhanced RIE

The etching rate of ICP-enhanced RIE is influenced by the etching process conditions, as well as total area of etching zone, etching opening size, surface morphology of etching zone and Cd component [37, 39, 40]. It is also related to the loading effect, the ARDE effect, and nonlinearity of RIE and the selectivity of the etchant to compound semiconductor materials. In order to improve the uniformity, repeatability, and controllability of dry etching HgCdTe, an investigation on etching rate must be carried out. The relationship between dry etching rate and etching time, etching area opening size, as well as uniformity of etching rate, is now introduced.

a. Nonlinearity of etching rate with time [6, 7, 10, 11]

Here, the dependence of etching rate in an empty chamber (called etching rate later) on etching time is introduced. The optimized process condition is used to acquire a clean and smooth profile. The specific gas proportion is CH_4:H_2:N_2:Ar = 1:0:5:20, the ICP power is 300 W, RF power is 8 W, chamber pressure is 3 mTorr, and temperature of sample platform is 20 °C.

Figure 5.16 illustrates the dependences of etching rate and depth as a function of processing time. As shown in Fig. 5.16, the etching rate decreases linearly at the rate of 0.34 Å/min with time, so that the dependence of etching depth on time is not linear. The nonlinearity of the etching process is related to the pattern dimension and other technical parameters. On one hand, along with etching, the AR of the pattern opening increases constantly, so the etching rate could decrease because the etchant is more difficult to enter and the products are more difficult to be expelled. On the other hand, in the etching process, there could be more and more polymers and residuals deposited on the surface of the etching zone, thereby influencing the subsequent etching.

b. Loading effect [6, 11, 21, 22]

The loading effect refers to the influence of etching area on etching rate. The larger the etching area, the lower the etching rate becomes in the same process

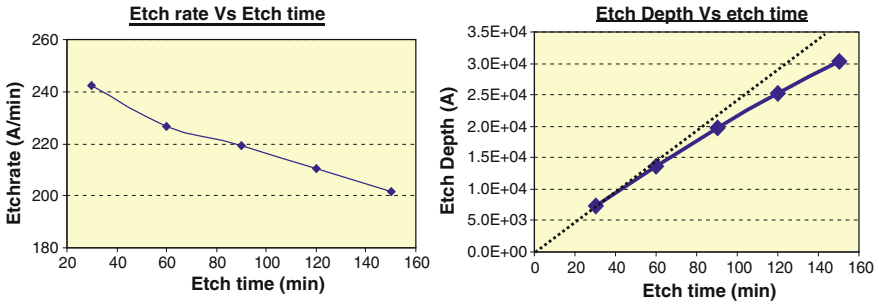


Fig. 5.16 Dependences of etching rate and depth on time

condition. This can significantly affect the controllability of the plasma. It is because of the difference of etching areas between different batches of samples that a change in etching rate, thereby resulting in the disagreement of practical etching depth and designed etching depth can occur. Since the etching rate of the samples of different dimensions, batches and amounts could become uncertain in the same etching condition due to loading effect, the dependence of etching rate on etching area can be given as follows [6, 7, 10, 11]:

$$\text{EtchRate} = \text{EtchRate}_0 - m \cdot \text{EA},$$

where EtchRate is the etching rate of practical sample, EtchRate_0 is etching rate in an empty chamber (it can be measured by samples with tiny etching area), EA is the practical area of etching zone, and m is a constant related to specified etching condition. Therefore, as long as EtchRate_0 and m are obtained by experiments, the etching rate can be calculated according to the area A of etching zone.

In order to acquire the m value in the equation (e.g., the dependence of etching rate on etching area) and the role of loading effect on etching, the layout that the etching area can control precisely is designed for a series of etching experiments. Finally, the dependence of etching rate on etching area is provided in Fig. 5.17.

As illustrated in Fig. 5.17, the etching rate decreases linearly with increasing etching area. The plasma density in the etching zone decreases because the increase in etching area demands more etchant. By linear fitting on experimental data, the formula of etching rate on etching area can be given by

$$\text{EtchRate} = 372.92 - 0.69 \cdot \text{EA}$$

where EcthRate unit is Å/min, EA is etching area, mm^2. The coefficient of determination of this formula is 0.98013. This formula represents the dependence of etching rate on the etching area in experimental etching conditions and only reflects

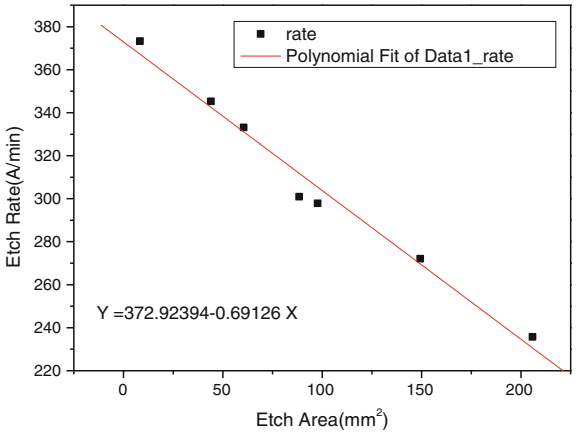


Fig. 5.17 Dependence of etching rate on etching area

an experimental result of loading effect in this condition. For the other etching conditions, the values of EtchRate_0 and m need to be determined by a series of similar experiments.

c. ARDE [6, 11, 21, 22]

The ARDE is contrary to the loading effect, that is, the smaller the linewidth of etching zone, the lower the etching rate becomes. Because the linewidth of the etching zone is smaller, the products are difficult to be desorbed and the plasma etchant is difficult to enter. With the same linewidth, the etching delay effect becomes more obvious with increasing etching depth. That is, the etching delay effect depends on the AR of etching trenches, which is called the microloading effect [6, 7, 10, 11, 28].

The ARDE also significantly influences the controllability of the plasma etching process. It is due to the effect that the etching delay could make the etching depths of trenches with different linewidths on the same chip different. Furthermore, when the AR of the etching trenches reaches a certain value, the etching will stop. Therefore, in the design of etching patterns, the linewidth of etching openings that is smaller and bigger than the critical value should be avoided for improving the uniformity of etching HgCdTe.

In order to understand the influence of the linewidth of etching trench opening on etching rate, a special mask is fabricated. The patterns on mask are mesas of 50 μm × 50 μm, and the spacing of the mesas are 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, and 16 μm, respectively. In the etching process, the SiO_2 layer with high selectivity ratio is applied for mask.

Figure 5.18 provides SEM images of the etching patterns fabricated by the above-mentioned mask. The etching process required 240 min. It can be seen that

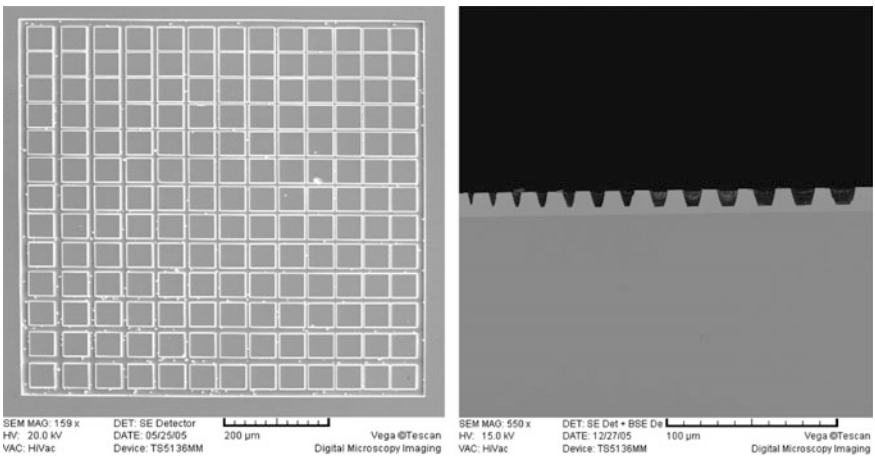


Fig. 5.18 SEM morphologies of top view and section of etching samples with different opening widths

when the opening width of the etching patterns is less than 4 μm, the profiles of etching trenches are "V" shape. This proves the influence on the limit of linewidth minimum of the above-mentioned etching condition.

In the experiment, the etching times for the samples are 240, 300, and 360 min, respectively. Due to the small linewidth of the etching trenches, the profiler cannot be used for measurement. Thus, the confocal laser scanning microscope (CLSM) is utilized to measure 3D surface profile and AR of etching trenches.

Figure 5.19 shows the CLSM images of surface and section of etching samples with different opening widths. Figure 5.20 shows the profile of an etching sample acquired by CLSM measurement. As shown in this figure, the etching depth decreases with decreasing linewidth. Moreover, when the opening width is less than 4 μm, the profile of etching trenches begins to appear as a "V" shape, which is consistent with SEM morphology images in the etching sample section. It indicates that the CLSM system can be applied for the nondamage measurement of the etching profile for HgCdTe samples and the precise acquisition of significant parameters of etching trenches, such as opening width, bottom width, and etching depth.

Figure 5.21 shows the dependencies of etching depth on opening width of etching patterns as measured by the CLSM system. In the figure, the three curves

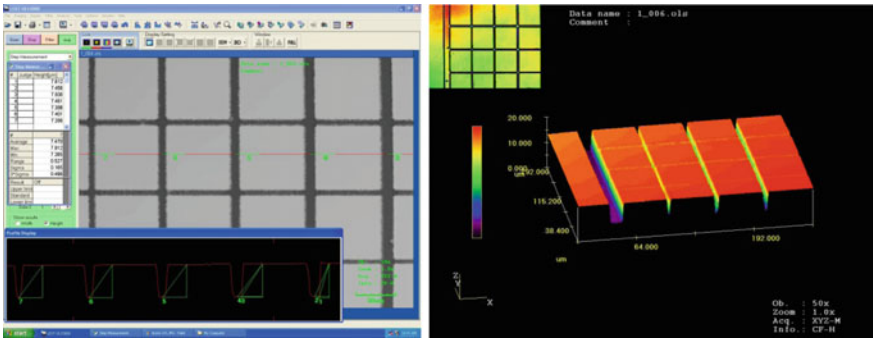


Fig. 5.19 CLSM images of surface and section of etching sample with different opening widths

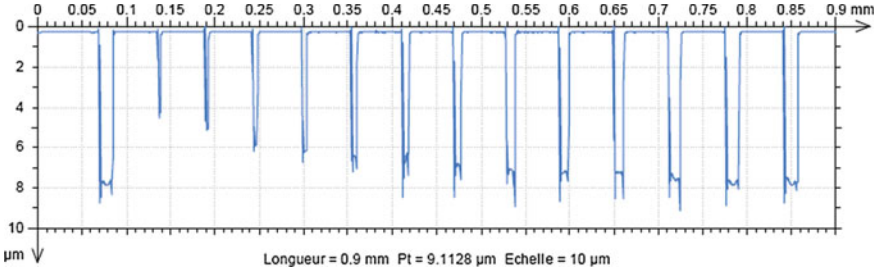


Fig. 5.20 Section profile of etching sample acquired by CLSM measurement

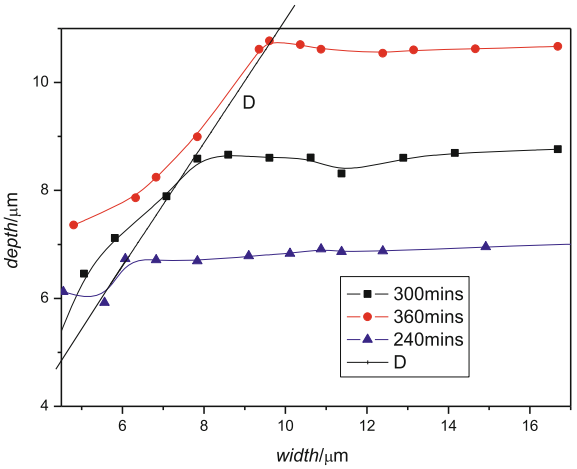


Fig. 5.21 Dependences of etching depth on opening width of etching patterns

correspond to the etching periods of 240, 300, and 360 min. When the opening width is large, the etching depth is constant basically with the change in opening width. However, when the width is less than a certain value, the etching depth reduces with the decreasing opening width. That is, when the opening width is less than the value of inflection point, the etching delay effect begins to act.

Nevertheless, the values of the inflection points for different etching periods are not fixed and increase with increasing etching period. As shown in Fig. 5.21, the etching depth of different etching periods and the values of inflection points on the opening width curves are almost on the same line D. Therefore, the opening width of etching patterns is not the only factor causing etching delay.

In order to investigate the critical condition of etching delay effect appearance, the curves of etching depth on opening width are transformed to the curves of the AR of etching trenches on etching rate. Figure 5.22 provides the relation curves of AR of etching trenches on etching rate as measured by experiments. As shown in this figure, when the AR of etching trenches is less than the critical value of 1.1, the etching rates for different opening widths are basically constant, while when the AR is greater than 1.1, the etching rate decreases promptly with the AR increasing.

The differences of experimental conditions corresponding to each data point, such as chamber cleanliness, also make the etching rate different. Hence, to reflect the role of this critical condition, the etching rates are normalized [6, 7, 10, 11]. Figure 5.23 provides the normalized relation curves of AR of etching trenches on etching rate. As shown in this figure, the inflection points of the curves at different etching periods are quite consistent. So it can be considered that the critical value of etching delay is determined by the AR of etching trenches. The reason for this is that when the AR of etching trenches is increased to a certain value, the plasma etchant cannot reach the etching zones and the reaction products cannot be desorbed promptly.

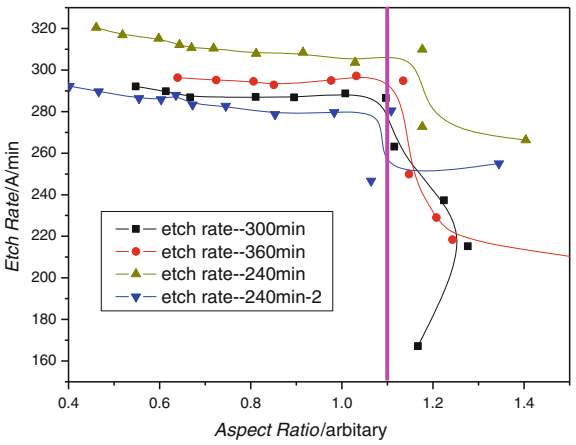


Fig. 5.22 Relation curves of aspect ratio of etching trenches on etching rate

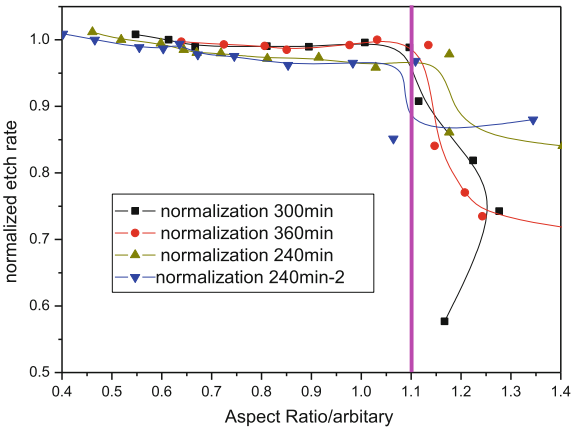


Fig. 5.23 Normalized relation curves of aspect ratio of etching trenches on etching rate

However, for different etching process conditions, the critical values of ARs are different. When the physical etching effect of ICP-enhanced RIE etching is strengthened, the critical dimension AR0 will increase. On the contrary, when the chemical etching effect is strengthened, AR0 could reduce. At present, the critical AR measured by experimental (AR0 = 1.1) is less than the critical dimension of dry etching for silicon. It indicates that in the etching process of HgCdTe, in order to reduce the electrical damage of materials, the physical etching effect of plasma should be relatively weakened, while the chemical etching effect should be enhanced.

d. Uniformity of etching rate

Since the ingredients influencing the etching rate of dry etching have been analyzed, the uniformity of large area or different batches is introduced here.

Table 5.3 provides the experimental results of uniformity within the range of 2 in. in diameter [6, 7]. In the experiment, five samples are placed on north, south, west, east, and center of the sample platform within the range of 2 in., respectively. By measuring the mean etching rate on each point of sample, the etching rate on each point can be obtained. By the uniformity formula

$$\text{nonuniformity} = \frac{\text{Max} - \text{Min}}{2 \cdot \text{Average}} \times 100\,\%,$$

The nonuniformity within the range of 2 in. can be calculated to be 9.98 %. If the minimum on the north point is ignored, the uniformity is 5.79 %. The low etching rate on the north point is likely due to the designs of chamber and inlet of etching gas. By increasing operating pressure and RF power of bottom properly, the etching uniformity within large area can be improved further.

Table 5.4 shows experimental results of etching uniformities in different batches. In this table, the experiments of run2, run3, and run4 satisfy the preconditions of the same sample dimension, the same location on sample platform, the same process parameter setting and equal etching time. According to the mean etching rates, the etching nonuniformity of different batches is obtained to be 2.53 % using the above-mentioned calculation method of uniformity. The etchings of different batches have good uniformity, which indicates that the etching process condition has good repeatability and stability.

(5) Selectivity of mask and influence of its profile on etching

The trench profiles of HgCdTe achieved by dry etching could be influenced by the selectivity of the mask and etching sample and the profiles of patterning mask. The reason is that the magnitude of etching selectivity can affect the etching deviation. On the other hand, because the etching plasma is in angular distribution, the profile of the mask pattern has a significant influence on the shape of etching trenches. Therefore, the influence of mask selectivity on etching deviation is introduced, and the influence of the profile of pattering mask on trench profile is analyzed in depth next.

a. Influence of mask selectivity on etching [6, 7, 10, 11]

Figure 5.24 provides the SEM images of etching pattern morphology with different etching selectivity. Figure 5.24a is obtained from the photoresist mask with a selectivity of less than 0.75:1, where the etching sidewall is not vertical and appears to be a zigzag ladder. This is because low etching selectivity leads to the contraction of the ambient pattering mask during etching process. In this case, the highly vertical sidewall can be not only acquired, but also can result in the loss of linewidth and large etching deviation. Figure 5.24b is obtained from the SiO_2 films grown by magnetron sputtering, of which the selectivity is more than 20:1. It can be seen in Fig. 5.24b that the sidewall has good vertical extent. Although the etching conditions of Fig. 5.24a, b are the same, their sidewalls have a great difference. Accordingly, a high etching selectivity is a significant factor to achieve good

Table 5.3 Experimental results of uniformity within range of 2 in. in diameter

run1	5 points							Ignoring north	
	North	West	Center	East	South	Total mean	1.24	Total mean	1.27
Measured value 1	1.13	1.32	1.37	1.24	1.22	Max	1.36	Max	1.36
Measured value 2	1.13	1.21	1.34	1.14	1.19	Min	1.11	Min	1.21
Measured value 3	1.08	1.28	1.37	1.28	1.23	Max–min	0.25	Max–min	0.15
Mean for each point	1.11	1.27	1.36	1.22	1.21	Nonuniformity	9.98 %	Nonuniformity	5.79 %

Table 5.4 Experimental results of etching uniformities in different batches

Sample	run2	run3	run4	Mean	2.57
Measured value 1	2.67	2.58	2.53	Max	2.64
Measured value 2	2.65	2.56	2.57	Min	2.51
Measured value 3	2.6	2.51	2.43	Max–min	0.13
Mean	2.64	2.55	2.51	Uniformity	2.53 %

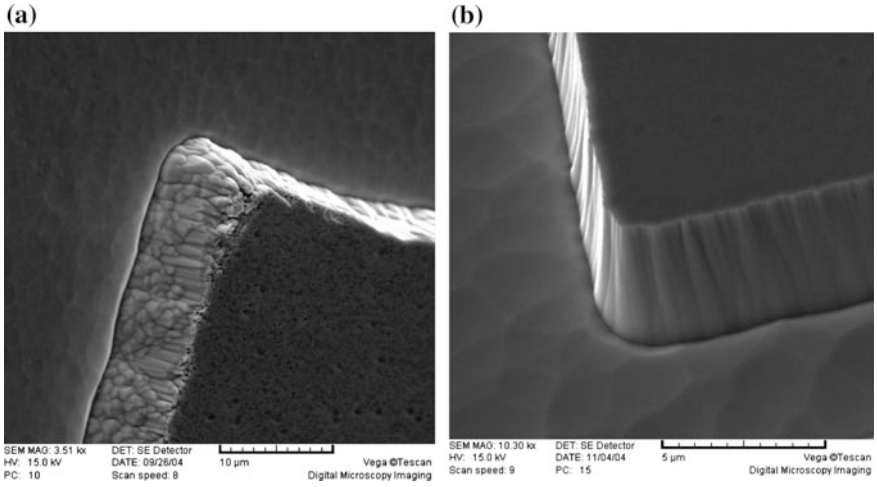


Fig. 5.24 SEM images of etching pattern morphologies with different selectivity. **a** Selectivity <0.75:1, **b** selectivity >20:1

etching profiles and small etching deviations. When the selectivity is high, the etching deviation will decrease; on the other hand, the etching deviation will become large with low selectivity. In fact, the etching deviation represents the extent of contraction of patterning mask during etching process, which can be calculated by $E_b = W_e - W_r = 2 \cdot L$, where W_r is the width of trench bottom of patterning mask, We is the width of the top of trench (correlated to the fill factor of pixels in the detector), and L is the distance of the contraction of trench sidewall.

In that way, the key to fabricating a mask pattern of high selectivity is to control the growth process of mask material. Due to the poor temperature stability of HgCdTe material, the growth process of the mask must satisfy the demand of low temperature growth. Thus, the magnetron sputtering method where the growth temperature can be controlled in the lower range is selected to prepare a SiO_2 mask. Moreover, the technical damage of HgCdTe material induced by the mask growth by magnetron sputtering can be reduced through decreasing the RF power of magnetron sputtering system.

Electrical damages of HgCdTe material induced in the mask growth under the RF powers of 200 and 20 W are provided next. The technical damages are

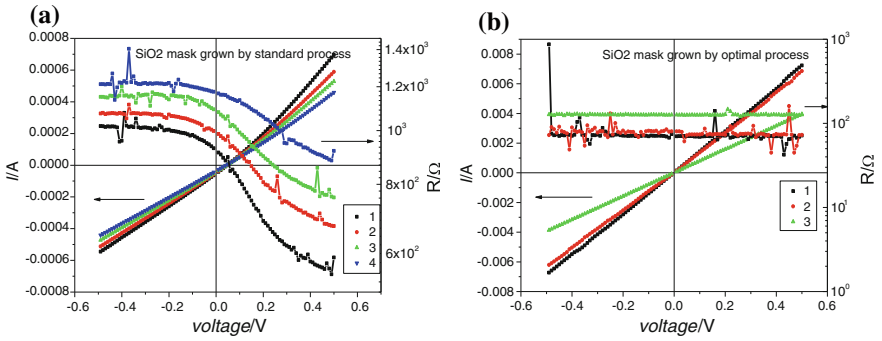


Fig. 5.25 *I–V* and *R–V* characteristic curves of mask zone and nonmask zone. **a** RF power is 200 W; **b** RF power is 20 W

estimated by measuring the *I–V* and *R–V* characteristics of the HgCdTe surface in both the mask zone and the nonmask zone. When the *R–V* characteristic curve is a line, it indicates that no technical damage is introduced during mask growth. On the other hand, when the *R–V* characteristic curve appears to have diode characteristics, it indicates that the technical damages are introduced. Figure 5.25 provides *I–V* and *R–V* characteristic curves of mask zone and nonmask zone. As shown in Fig. 5.25, the electrical damages are introduced for the mask grown at the RF power of 200 W, while there is rarely electrical damage at the RF power of 20 W.

b. Influence of mask pattern profile on etching trench profile

The geometrical profile of an etching trench depends on the etching selectivity, thickness and mask sidewall angle, and the bottom electrode power [6, 7, 10, 11, 28, 39]. Figure 5.26 shows the schematic of the influence of the mask pattern geometrical profile on the etching trench profile. As shown in this figure, the ICP plasma has a specific IAD and IED. The IAD is typically a Gaussian distribution with the maximum at the center ($\theta = 0$). Here, θc is defined to be twice the half peak width of the IAD. Meanwhile, the AR of the trenches is defined as $A_r = D/W_e$, where W_e is the top width of etching trench, and D is the vertical depth of trench. For a fixed etching time, D is actually a function of W_e due to the etching lag effect which is actually limited by the amount of etching ions that could pass through the trench and reach the desired depth. As shown in Fig. 5.26, the AR value can be calculated by $(2 \cdot tg\phi)^{-1}$, where ϕ is the steepness angle of sidewalls in etching trenches. When $\theta_c < \phi$, the ions of the plasma can reach the bottom of etching trench without etching lag effect in most directions. However, when $\theta_c > \phi$, only the ions in a very few directions (in vertical directions) can reach the bottom of the trenches to take part in the etching reaction, resulting in the etching lag effect.

Notably, the IAD of plasma is affected by the DC bias. If all the other parameters remain constant and the DC bias increases, then the range of IAD and θ_c will decrease. Therefore, with an increase in the bottom electrode power, the AR of the etching trenches is correspondingly increased. However, due to the increase in the

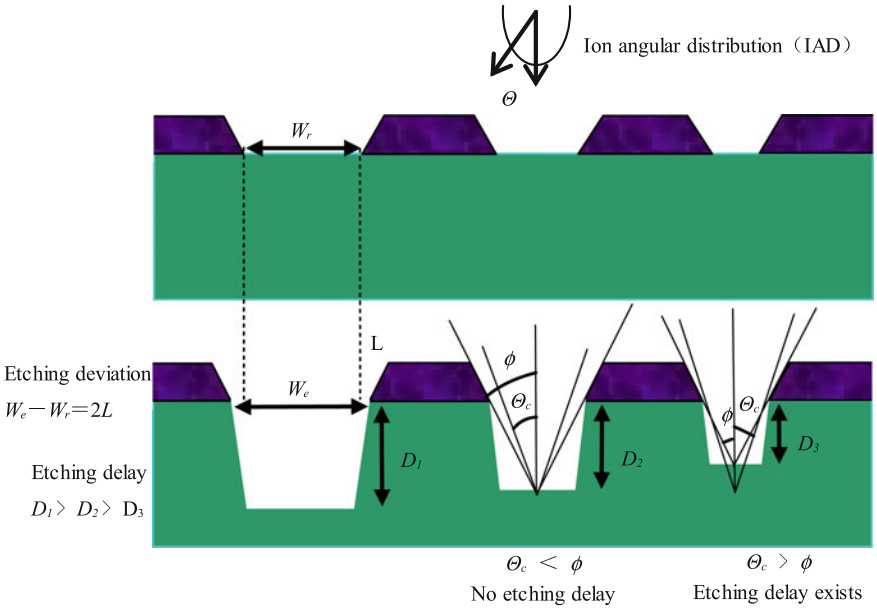


Fig. 5.26 Schematic of influence of mask pattern geometrical profile on etching trench profile

RF power of bottom electrode, the electrical damages induced by dry etching must also increase. Accordingly, for optimizing the etching process, the contradiction between the decrease in the electrical damages induced by dry etching and avoiding etching lag effect should be balanced.

Moreover, in order to improve the duty cycle of pixel mesas, the etching lag effect necessarily exists in the etching of mesh trenches for micro-mesa isolation. The reason for is that the section of trenches for micro-mesa isolation is desired to be a "V" shape satisfying the design requirement of trench depth in order to acquire the highest duty cycle when designing the etching process. Furthermore, during dry etching, the high AR mask pattern of is usually needed to reduce the influence of the long duration bombardment of the etching ions deviating from the vertical direction to the sidewall of the etching trenches [7, 10, 39]. Figure 5.27 provides the schematic of the thick mask pattern with high AR used to suppress the sidewall damage of etching trenches. As shown in this figure, when the mask is thin, the etching ions deviating from the vertical direction could bombard the sidewall of the etching trenches for a long time and induce the electrical damage, thereby influencing the performance of the micro-mesa array device. On the other hand, when the mask is thick, only the etching ions in the vertical and nearly vertical directions can pass through trenches and reach the bottom of trenches to complete the etching reaction, while those etching ions with the incident angles deviating from the vertical direction obviously only bombard the sidewalls of the masks or take part in the reaction, rather than bombard the sidewalls of etching trenches. This

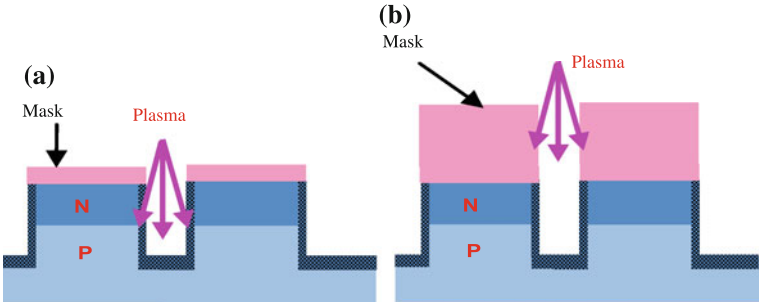


Fig. 5.27 Schematic of the mask pattern with a high aspect ratio used to suppress the sidewall damage of etching trenches. **a** Thin mask; **b** Thick mask

can not only reduce the sidewall damage in dry etching, but also can improve the anisotropy of etching trenches.

3. Etching damage of HgCdTe and its estimation

(1) Mechanism analysis of etching damage

Usually, it is considered that the plasma induced damage of HgCdTe is closely correlated to the physical effect in dry etching [41]. Due to the weak Hg–Te bonds in HgCdTe material, the bonds can be broken to generate free Hg. Because the surface temperature of the etching sample is high, a portion of free Hg can diffuse into the material in the form of Hg interstitial atoms and cause the induced damage. According to the literature [41], for MWIR HgCdTe material, an Hg interstitial atom could be generated after etching 48,000 HgCdTe molecules. On the one hand, these Hg interstitial atoms could cause lattice damages and form donor defects similar with the ion implantation of HgCdTe; on the other hand, they can react with metal vacancies and generate equivalent n-type dopants. For Hg vacancy doped p-type HgCdTe materials, the Hg interstitial atoms can diffuse into HgCdTe and neutralize the vacancies in HgCdTe, and equivalently the Hg interstitial atoms annihilate the Hg vacancies in HgCdTe.

$$\mathrm{Hg_i^+} + \mathrm{V_{Hg}^-} \rightarrow \mathrm{Hg_{Hg}^0}$$

Due to the annihilation of Hg vacancies, the results are equivalent to n-type dopants electrically. However, for As-doped or Au-doped p-type HgCdTe, similar electrical-induced damages also appear during dry etching. Generally, it is considered that the Hg interstitial atoms generated from the etching surface diffuse into the material and drive the donor dopants to form the displacement in the lattice, which is called the "kick out" inversion mechanism [41].

It has been proved in many experiments that the diffusion of Hg interstitial atoms induced by physical effect is the main factor in generating induced inversion damage for RIE or Ion milling dry etching of HgCdTe. However, the physical

damages in the ICP etching of HgCdTe are small, so the influence of the selection of chemical etchant in dry etching process on etching damage becomes very significant.

The etching mechanisms of CH_4 and H_2 etching HgCdTe start from the decomposition and ionization, and the reaction equations are given by [41–43]

$$\begin{aligned}
&CH_4 + 4.4\,eV \rightarrow CH_3 + H\\
&CH_3 + 3.8\,eV \rightarrow CH_2 + H\\
&CH_2 + 4.0\,eV \rightarrow CH + H\\
&H_2 + 4.5\,eV \rightarrow 2H\\
&H + 13.6\,eV \rightarrow H^+ + e^-\\
&H_2 + 15.4\,eV \rightarrow H_2^+ + e^-
\end{aligned}$$

The decomposition and ionization of H_2 only generate H, while CH_4 mainly generates $–CH_3$ and H. As known that the etching damage of H_2-based process gas is remarkably more than that of CH_4-based process gas, the H ions in plasma may play an important role on the etching damage. The H element in plasma, especially H ions (protons), has strong penetrability and activity, due to tiny volume and high speed. The H element may diffuse into HgCdTe and react with the other atoms to lead to the change in the electrical properties of the material, thereby forming thick induced damage layers [6, 33, 34, 44, 45].

(2) Test methods of etching induced damage

a. Measurement of *I–V* and *R–V* characteristics [6, 11]

It is well known that ion beam milling, traditional RIE- and HDP-type ECR-enhanced RIE can result in a change in the electrical properties of HgCdTe surface layers, and even induce the transformation of p-type to n-type, which is called inversion phenomenon. The electrical measurement of inversions induced by etching is one of the investigations on dry etching. Measuring the *I–V* characteristics of etching zone and non-etching zone can be used to estimate the inversion of material after etching. An example is used to estimate if the low-energy ICP-enhanced RIE leads to the transformation of the etching surface layer of HgCdTe material from p-type to n-type as the etching energy of plasma is lower than 100 eV or even lower than 50 eV.

A series of 50 × 50 μm^2 holes arranged regularly are fabricated on the sample, and then the sample is etched. The etching conditions are: the process gas proportion is CH_4:H_2:N_2:Ar = 1:0:5:20, the ICP input power is 400 W, the RF power of bottom electrode is 8 W, the chamber pressure is 3 mTorr, and the sample platform temperature of bottom electrode is 20 °C. During etching, the etching state parameter DC bias is 70 V. After etching, for measuring the electrical properties, ohmic contact electrodes are fabricated on the etching zone and non-etching zone, respectively. If the etching really leads to the inversion, the structure is quite similar to ion-implanted n^+-on-p planar photovoltaic detectors.

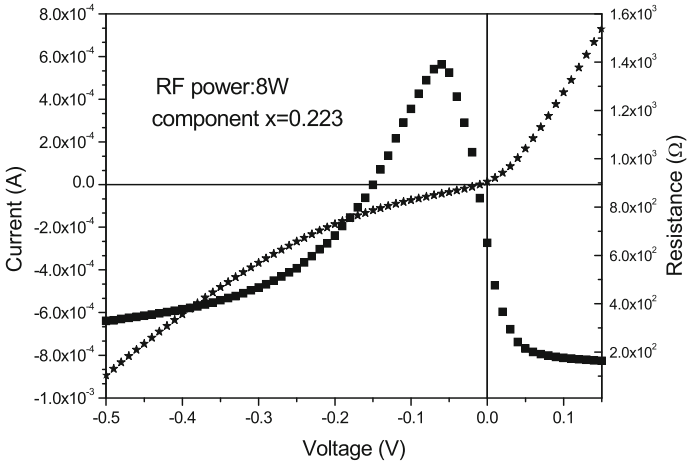


Fig. 5.28 *I–V* and *R–V* characteristic curves of etching zone and non-etching zone at liquid nitrogen temperature

Figure 5.28 provides the *I–V* and *R–V* characteristic curves of the etching zone and non-etching zone of p-type HgCdTe material ($x = 0.223$) at liquid nitrogen temperature after plasma treatment. As shown in this figure, the *I–V* and *R–V* curves have diode characteristics, which indicate that there is a *p-n* junction structure existing between etching zone and non-etching zone. The result indicates that even if the DC bias is controlled to be 70 V, the etching plasma still induces some damage to HgCdTe, so that the surface layers in the etching zone of p-type material are transformed to n-type. Therefore, the etching condition cannot be applied directly in the forming process of HgCdTe micro-mesa array.

b. Measurement of transverse distribution of electrical damage by LBIC

The measurement results of the above-mentioned *I–V* and *R–V* characteristics indicate that the low-energy ICP-enhanced RIE still leads to severe electrical damage induced by dry etching, so that the inversion layer occurs in the surface layers in the etching zone of p-type material. Accordingly, the distribution of the inversion layers induced by plasma in this etching condition is investigated further based on the LBIC measurement.

Figure 5.29 provides the SEM morphology of the LBIC test sample and the measurement results of its electrical damage after dry etching. The etching condition of the sample is the same with that of *I–V* and *R–V* characteristics test samples. Figure 5.29a shows the SEM morphology of the LBIC test sample and the locations of measuring electrodes after dry etching. As shown in Fig. 5.29a, after ICP-enhanced RIE etching, two remote LBIC measuring electrodes A and B are extracted from the non-etching surface on the edge of the sample of p-type HgCdTe ($x = 0.482$). A series of square etching mesas arranged periodically (the spacing is 42 μm and the sidelength is 50 μm) is fabricated between the electrodes A and B.

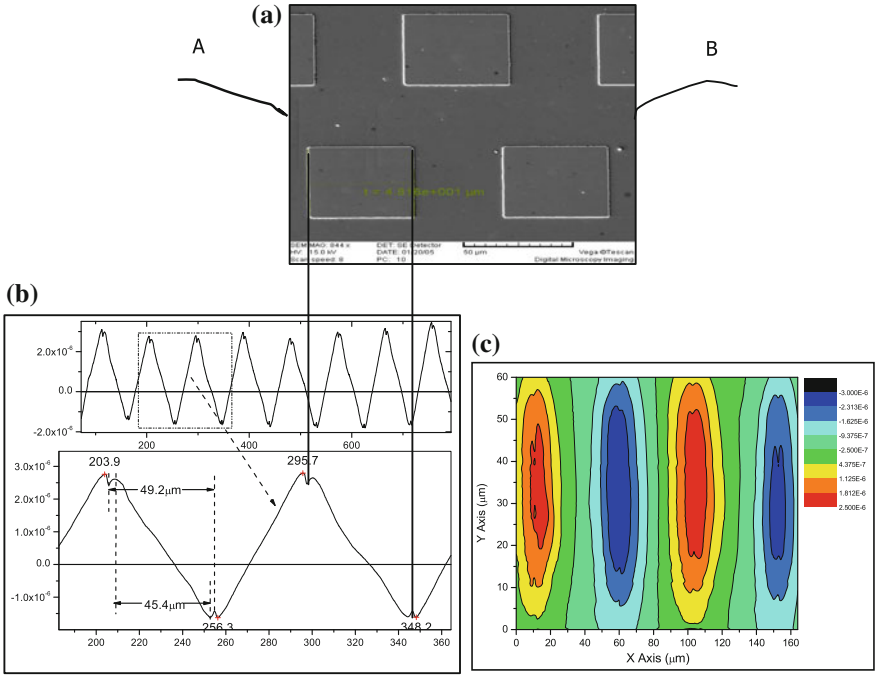


Fig. 5.29 SEM morphology of the LBIC test sample and its measurement results after dry etching. **a** SEM morphology of the LBIC test sample and measuring electrodes; **b** LBIC line scanning results; **c** LBIC area scanning results

Figure 5.29b, c provides LBIC line scanning and area scanning results of dry etching samples [6, 11, 41]. As shown in this figure, the peak at 295.7 μm and the trough at 348.2 μm correspond to the two sidewalls of an etching mesa. It can be observed that there are two small peaks at the main peak, while there are two troughs at the main trough. In the range from 295.7 to 348.2 μm, the small peaks and small trough inboard correspond to the interfaces of *p-n* junctions (comprised of the n-type inversion zone of etching induced damage and p-type zone of non-etching at the two sidewalls of a mesa), and the spacing between them is approximately 45.4 μm; while the pit at the peak and the bump at the trough correspond to the two sidewalls of a mesa, and the spacing between them is about 49.2 μm. The spacing between two sidewalls of a mesa is less than 50 μm of layout design, which is due to over development during the developing of the photolithographic patterns. According to the mesa sidelength of 49.2 μm and the spacing between *p-n* junction interfaces of 45.4 μm, it can be determined that the etching plasma induces the transversal damage inversion layers of $\sim$1.9 μm on each sidewall of a p-type HgCdTe mesa.

Two small peaks and two small troughs shown in Fig. 5.29b have been not reported yet. Of the two small peaks, one peak corresponds to the location of the etching induced *p-n* junction; another peak occurs in the etching zone with a

distance of ~1.5 μm to the mesa sidewall, and its cause is not well known. The probable explanation is that narrow and shallow accessory trenches are caused by the etching process on the both sides of the trench bottom for micro-mesa isolation, resulting in the appearance of the maximum LBIC signal at these locations. However, it can be seen in SEM morphology that the above-mentioned small trenches are not discovered. Hence, the mechanism of the double peaks and double troughs is not clear, and it should be analyzed thoroughly according to the physical mechanism of LBIC signal generation and the geometrical structure features of the etching samples.

c. Measurement of longitudinal distribution of electrical damage layer by differential Hall methods

The Hall mobility and Hall concentration at room temperature and 77 K can be obtained by the differential Hall measurement based on van der Pauw theory. The change in the Hall concentration (77 K) of the etching sample as a function of depth is measured [6, 11, 45]. The samples of differential Hall measurement are obtained by preparing ohmic contact electrodes on the four corners of p-type HgCdTe material etched by plasma. The four ohmic contact electrodes are connected to the etching damage inversion layer, as well as the non-etching zone inside the sample. The fractional removal of layers in differential Hall measurement is achieved by wet chemical etching with a specified concentration of Br_2/HBr solution. The removed thickness of each layer is obtained by measuring Fourier infrared transmission spectra and fitting calculation and determined in comparison to the estimated value of the product of empirical etching rate and etching time. Figure 5.30 provides the results of differential Hall measurement of etching samples with different bottom electrode RF powers and different Cd components. Figure 5.30a shows the measurement results of p-type HgCdTe materials (Cd component is 0.319 and doping concentration is 9.45×10^{15}) with bottom electrode RF powers of 8 and 20 W.

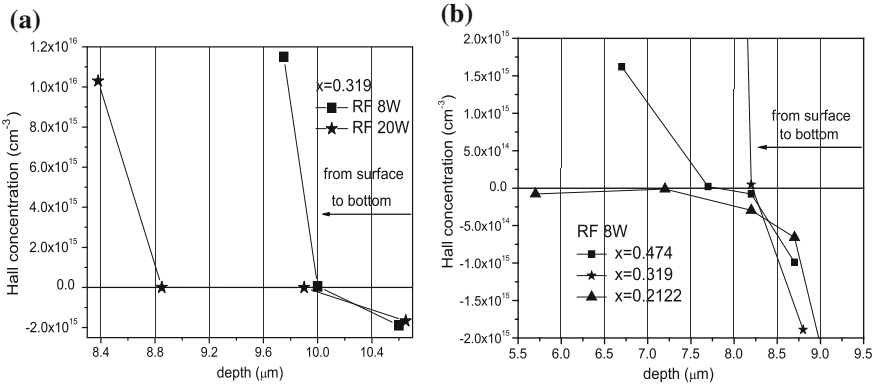


Fig. 5.30 Results of differential Hall measurement of etching samples. **a** Different bottom electrode RF powers; **b** Different Cd components

The other parameter settings of the etching conditions in the experiment are the same as those for measuring sample *I–V* and *R–V* characteristics. As shown in Fig. 5.30a, when the RF power of the bottom electrode is 8 W, the longitudinal thickness of the etching induced inversion layer is approximately 1.0 μm. When the RF power of bottom electrode is 20 W, the longitudinal thickness of etching induced inversion layer is about 2.3 μm. This indicates that the higher the RF power of bottom electrode, the thicker the inversion layer induced by plasma etching becomes.

Figure 5.30b provides the results of differential Hall measurement of etching samples (x = 0.474, 0.319 and 0.2122). The three samples were etched in the same batch, and the parameter settings of the process conditions are the same as those of samples for *I–V* and *R–V* characteristics measurement. As shown in Fig. 5.30b, the damage inversion layer of p-type HgCdTe material (x = 0.2122) becomes very thick and is not removed completely after the wet etching for 5.0 μm. The thickness of the inversion layer of short wave material (x = 0.474, doping concentration is 3.7×10^{17}) is approximately 2.0 μm; while the thickness of the inversion layer of medium wave material (x = 0.319, doping concentration is 9.47×10^{15}) is approximately 1.0 μm. The etching inversion layer of p-type short wave HgCdTe is thicker than that of medium wave, which is contradictory to the fact that the HgCdTe material with high Cd component is usually considered to have low damage. However, the cause is not well known.

d. Measurement of longitudinal distribution of damage by material chip technique [6, 11, 45]

For the measurement method of dry etching HgCdTe introduced in the above section, the mechanism and features of measuring the thickness of etching induced electrical damage layer by material chip techniques have been presented in detail. The photodiode array between the damage zones of different thicknesses and the nonelectrical damage zones is fabricated on the same etching sample. By measuring the *I–V* and *R–V* characteristics of this array, the thickness of electrical damage layer can be estimated precisely according to the change in these curves. With the decrease in the thickness of damage layer, the *I–V* and *R–V* characteristics of photodiodes become more and more obscure. When the electrodes of the photodiodes are located at the interface between the electrical damage layer and non-damage layer, the *I–V* and *R–V* characteristic curves of this location should be a line. Accordingly, the thickness of electrical damage layer is the etching depth of damage layer, which can be measured by the profiler or confocal scanning laser microscope. Since the photodiodes are obtained from the same batch and the same dry etching process and the *I–V* characteristic curves of etching damage layer and non-etching zone are measured, the method is intuitive and the acquired data are reliable. Therefore, the method is commonly adopted in the experiment on the longitudinal distribution of electrical damages undertaken later.

Figure 5.31 provides measurement results of the etching induced electrical damage by the material chip technique. The sample in this figure is long wave HgCdTe material (x = 0.215), and the dry etching condition is the above-mentioned

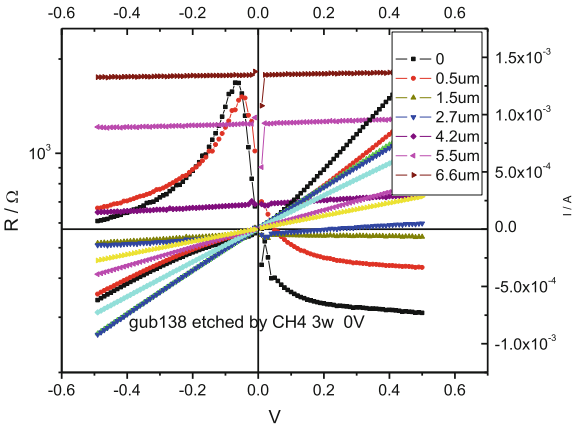


Fig. 5.31 Measurement results of the etching induced electrical damage by material chip technique

stable etching process offering good etching surface. As shown in Fig. 5.31, when the electrical damage layers induced by dry etching are not removed or removed for 0.5 μm, the etching zone and the non-etching zone still represent remarkable diode characteristics. When the electrical damage layers are removed for 1.5 μm, the etching zone and the non-etching zone do not possess diode characteristics. This indicates that the longitudinal thickness of electrical damage layer induced by dry etching is less than 1.5 μm for long wave materials.

e. Measurement of longitudinal distribution of damage by microzone SPM

Although the measurement of longitudinal distribution of electrical damages by material chip techniques has high creditability, the preparation of measuring samples based on this method has many processing steps, involving cleaning, dry etching, passivation, contact electrode, and indium bump preparation. These processes almost include all the fabrication technologies of monochromatic FPA chips, so that there are many factors influencing the experimental results and the experiment period is also prolonged. Additionally, the microzone SPM applied in the semiconductor (Si, GaAs) field can be adopted to measure the longitudinal distribution of electrical damages induced by dry etching. The specific method is that the etched HgCdTe is cleaved and then coated with a thin passivation dielectric layer on the cleavage surface to prevent the oxidization of the cleavage surface, which could affect the measurement accuracy of surface potential.

Figure 5.32 shows the measurement results of etching induced electrical damages using the microzone SPM [11], where the sample #13 is not etched by plasma, while the sample #12 suffers the dry etching of plasma. The Cd component and etching condition here are identical as shown in Fig. 5.31. As shown in Fig. 5.32, no obvious inflection points are discovered mainly on the surface potential curve of the sample #13 longitudinally, except for the small zone in the surface layer; while the notable inflection point occurs at a distance of 1.5 μm on the surface potential curve of sample #12. Accordingly, it can be estimated that the dry etching induces

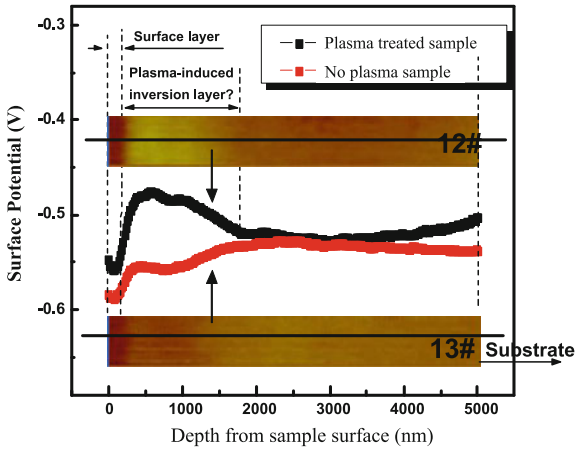


Fig. 5.32 Measurement results of etching induced electrical damages using microzone surface potential method

these electrical damages near the surface, which is consistent with the measurement results of material chip technique.

(3) Research on the reduction of etching induced damage

By these comparison experiments and precise measurement of induced damage, the generating mechanism of etching damage is preliminarily achieved. Here, starting from the main parameters of etching process, the mechanism of damage is analyzed in order to achieve an ICP etching condition with low damage.

a. Relationship between self-bias modulating RF power and damage layer thickness [11]

Regarding the aspect of physical etching damage, the high-energy ions can break the weak Hg–Te bonds in HgCdTe material and produce free Hg interstitial atoms, which diffuse into the material, thereby causing the *p*-to-*n* induced damage. For chemical etching damage, the energetic H^+ ions have a strong penetration and result in the deeper induced damage layer. Hence, the higher the ion energy in dry etching, the worse the etching induced damage becomes. In the research on the reduction of damage, first, the relationship between the etching induced damage and etching energy should be explored for the ICP etching technique with high density and low energy.

Figure 5.33 provides the thickness of damage inversion layer obtained from the measurement of *I*–*V* and *R*–*V* characteristics of p-type HgCdTe material (x = 0.2325) etched by ICP etching at different RF powers. The role of self-bias modulating RF source is to modulate the DC bias of the ion-directional acceleration voltage in the chamber, and its power depends on the ion energy in plasma. In the experiment, only the RF power is varied, while the other process parameters such as process gas proportion, ICP power, and chamber pressure remain constant. Although the variation range of RF power is 0.1–0.6 W/cm^2 (still in the scope of low-energy etching), the change in etching damage becomes remarkable (the thickness of

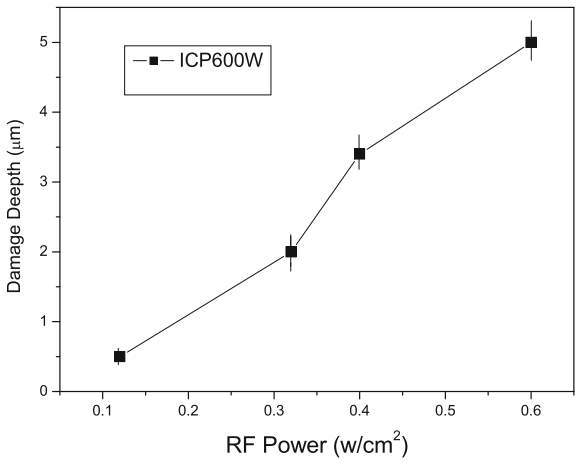


Fig. 5.33 Relationship between self-bias modulating RF power and induced damage layer thickness

induced damage layer increases from 0.5 to 5 μm, for one order of magnitude). The reason is considered that on the one hand, the physical damage increases; on the other hand, with RF power increased, the concentration of H^+ ions generated from the decomposition of chemical etchant increases continuously as well as the kinetic energy of H^+ ions for the low-energy etching process. The latter causes the infusion of more H^+ ions with higher energies into the material, and the strong penetration and deep penetrating depth of H^+ ions enhances the obvious etching induced electrical damage and increases the damage depth.

b. Relationship between ICP power and damage layer thickness [11]

In order to reduce the damage to HgCdTe material induced by ICP etching, the ion energy of plasma must be lowered. Furthermore, one of the features of ICP etching technique is to make up the influence of low energy on etching rate through improving the density of plasma. Therefore, increasing plasma density and reducing ion energy is an effective way to reduce etching damage without the loss of etching rate.

In the same conditions, increasing ICP power can generate higher density of the plasma. So, if the RF power remains constant, the energy of individual ions in plasma will decrease. A significant process parameter characterizing the energy of individual ion is DC bias. Figure 5.34 provides the relationship between DC bias and ICP power with constant RF power.

The DC bias determines directly the energy of ions in the chamber longitudinally (e.g., in the etching direction). In Fig. 5.34, the DC bias decreases with the increasing ICP power, except for a certain increase at the beginning. That is, increasing the ICP power can reduce the DC bias under high ICP power. In other words, decreasing the kinetic energy of ions in the etching direction can reduce the plasma induced damage effectively in theory.

Figure 5.35 shows the relationship between the ICP powers measured in the experiments and the thicknesses of induced damage inversion layers. With the ICP

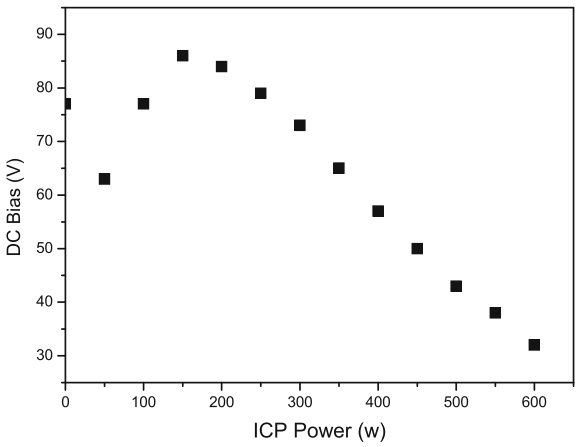


Fig. 5.34 Relationship between DC bias and ICP source power

power increasing, the thickness of damage inversion layer decrease gradually. It indicates that an increase in ICP power can reduce the etching damage effectively. Meanwhile, according to the results of etching profile optimization, improving the plasma density can increase the etching rate as an effective way of achieving good dry etching process for HgCdTe.

c. Influence of chamber pressure on etching damage [11]

The variation of chamber pressure and gas flow also can change the plasma density in the etching system. In theory, improving the chamber pressure can increase the plasma density in the chamber and the ion density correspondingly. So the kinetic energy of etching ions will decrease under the constant RF power.

However, the measurement results of *I–V* and *R–V* characteristics indicate that the change in the thickness of inversion layer with chamber pressure increasing is

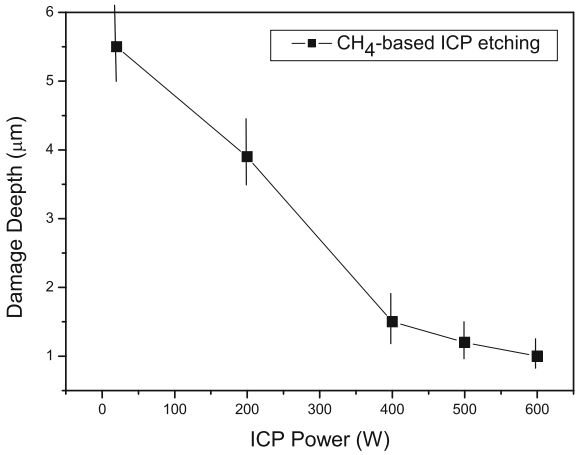


Fig. 5.35 Relationship between ICP power and induced damage inversion layer

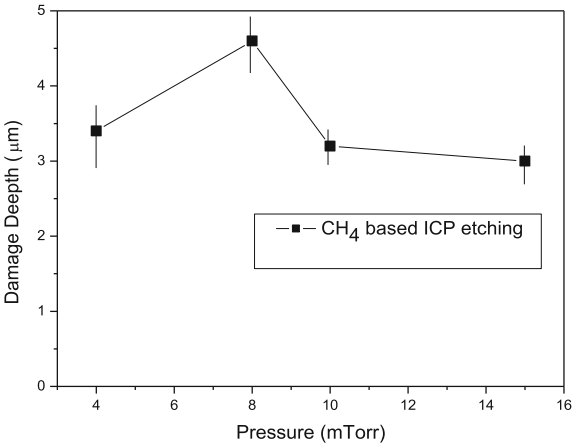


Fig. 5.36 Relationship between chamber pressure and induced damage inversion layer

not obvious, and even the phenomenon of damage increasing occurs, as shown in Fig. 5.36. By analysis of experimental results, the reason is that the increase in chamber pressure can reduce the energy of etching ions but also lead to the reduction in the mean free path of ions in the chamber while H^+ ions in the plasma have very low probability to collide with other atoms and reach the sample surface more easily due to small volume and high speed. That is, the increase in the proportion of H^+ ions taking part in etching reaction aggravates the etching damage. The Ar/H_2 plasma ratio in ECR etchant was analyzed using Langmuir probe by Stoltz et al. [43]. It was discovered in the measurement that the proportion of $Ar^+/(Ar^+ + H^+)$ could decrease with the chamber pressure increasing. The experimental results verify the above-mentioned analysis. Although improving the chamber pressure can increase the plasma density and lower the ion energy, the increase in the proportion of H ions does not decrease the induced damage with the plasma density increasing.

d. Relationship between concentrations of H^+ and H and the thickness of damage layer

In the above sections, the optimization of process parameters has been introduced to reduce the induced damage. The H ions in the etchant gas of H_2/Ar could have a significant influence on etching damage. The origin of H ions is the ionization of H_2 gas in ICP chamber. The probable decomposition of H_2 gas occurring in plasma could be [11, 42, 46]:

$$H_2 + 4.5\,eV \rightarrow 2H$$
$$H + 13.6\,eV \rightarrow H^+ + e^-$$
$$H_2 + 15.4\,eV \rightarrow H_2^+ + e^-$$

The H ion is actually a proton, with high speed and strong penetration. However, it can be seen that the products of H_2 molecules in the plasma are also H atoms and H_2 ions, except for H^+ ions, and all of them can be used as chemical etchants for HgCdTe. The H atoms not only have larger volume than H^+ ions, but also have no charge. Therefore, they cannot be accelerated directionally in the accelerating electrical field, and the plasma-induced damages generated longitudinally are also fewer.

(a) Suppression of H^+ ion etching

From the Bohm's equation in plasma physics, the proportional relationship between ionization potential and electron temperature can be given by Eq. (5.1) [7, 11, 41, 46, 47]:

$$V_f - V_p = \frac{kT_e}{2q}\left[\ln\left(2\pi\frac{m_e}{m_i}\right) - 1\right] \tag{5.1}$$

where m_e is the electron mass, m_i is the ion mass, and the electron temperature kT_e and the ionization potential $V_f - V_p$ are interactive. The relative atomic mass of the ion is set to be M, as the proton mass is 1836 times more than that of electron, it can be obtained [7, 11, 41, 46–48] that:

$$ER = \frac{q(V_f - V_p)}{kT_e} = \frac{1}{2}\left[\ln\left(2\pi\frac{1}{1836\cdot M}\right) - 1\right], \tag{5.2}$$

where ER is the energy rate of ions presented by Elwenspoek; for H^+ ions, $M = 1$, and its energy rate $(ER)_H = -3.34$; for Ar^+ ions, $M = 40$, and $(ER)_{Ar} = -5.18$.

From Eq. (5.2), it can be obtained that ER is closely correlated to the atomic mass. By assuming that the plasma of H_2/Ar has only H^+ ions and Ar^+ ions, the concentration distribution of ions can be achieved by analyzing ER. It is because [7, 11, 41, 46, 47]

$$C_{Ar}\cdot(ER)_{Ar} + (1 - C_{Ar})\cdot(ER)_H = (ER)_{Measured}, \quad 0 < C_{Ar} < 1, \tag{5.3}$$

where $C_{Ar} = N(Ar^+)/N(Ar^+ + H^+)$ represents the percentage of Ar^+ ions in the total amount of ions in plasma. Equation (5.3) is applicable to all the binary ion systems, especially the case of two ions with a great difference in mass. Substituting the ERs of the H^+ ions and Ar^+ ions into the equation, the percentage of Ar^+ ion concentration in the plasma of H_2/Ar for dry etching HgCdTe can be given by [46]

$$C_{Ar} = \frac{(ER)_{Measured} + 3.34}{-1.84} \tag{5.4}$$

The ER of ions in the plasma can be acquired by Langmuir probe measurement. The measurement results of plasma etching for HgCdTe reported by Storltz et al. indicate that even if the H_2 gas enters the chamber from the inlet at the bottom of

the chamber, the C_{Ar} value of H_2-based plasma is still proportional to the proportion of Ar in the process gas. If the H_2 molecules in the plasma are dissociated mainly into H atoms, the C_{Ar} value should be a constant approaching to 1 and invariable with the change in gas proportion. The experimental results indicate that even if the H_2 gas enter the chamber from the inlet at the bottom, the H_2 molecules in the plasma are still dissociated into H^+ ions. Similar results are also presented in other research work. Based on the modeling calculation presented by Neyts et al. [35], it was given that the main components of ions in the plasma of H_2/Ar are ArH^+ and H^+ at the plasma pressure of 100–250 mTorr. The results of mass spectrometric analysis presented by Lee et al. [48] indicate that the main components of ions in the plasma of H_2/Ar are H^+, H_2^+, Ar^+, and ArH^+ at the low plasma pressure of 1–10 mTorr. Even if the H_2 gas enters the chamber from the bottom of the chamber, the generation of H^+ ions is difficult to suppress. Also, the measurement results of etching damage indicate that the etching damages induced by H_2-based etchants are still significant with the H_2 gas entering the chamber from the bottom inlet.

(b) Introduction of hydrogen consumption etchant and related experiments

Because the etching damages caused by H_2-based etchants are not avoided, CH_4-based etching process attracts more attention. In fact, the CH_4-based etching process gases also generate a large number of H^+ ions in the plasma. The ionization process of the CH_4 gas in plasma is given as follows [7, 11, 41, 46, 47]:

$$\begin{aligned} CH_4 + 4.4\,eV &\rightarrow CH_3 + H \\ CH_3 + 3.8\,eV &\rightarrow CH_2 + H \\ CH_2 + 4.0\,eV &\rightarrow CH + H \end{aligned}$$

And the H atoms produced from the dissociation of CH_4 may be ionized in the plasma:

$$H + 13.6\,eV \rightarrow H^+ + e^-$$

Under the same etching conditions, the amount of H^+ ions in the CH_4-based process gas may approach that of the H_2-based process gas. However, the results of etching damage are greatly different, which indicates that the content of H^+ ions in the CH_4-based process gas is very low. In comparison with their gas proportion, there is N_2 gas existing in the CH_4-based process gas, except for CH_4 (as etchant) and Ar (as plasma stabilizer). Adding N_2 into the CH_4-based process gas is to suppress the generation of polymer precursors (C_2H_4, C_2H_2, CH_2, and CH) for reducing the deposition of polymers on the etching surface. In order to explore the role of N_2 on the etching damage, the N_2 gas is added to the H_2-based process gas. Accordingly, the following results are achieved.

The HgCdTe material of $x = 0.2325$ is used in the etching experiment in Fig. 5.37, and the ICP power is 600 W. The relationship of the measured thickness of etching induced damage layer and the RF power indicates that the damage

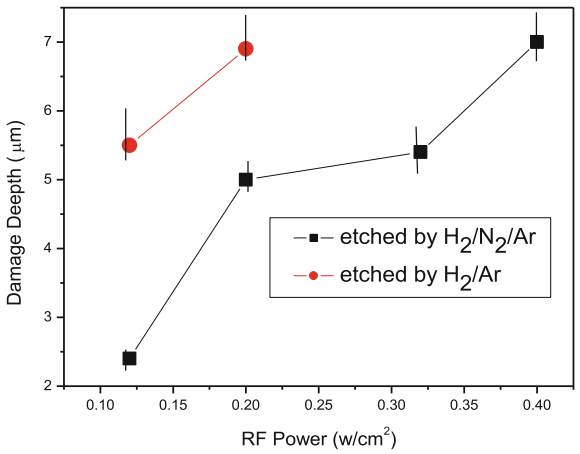


Fig. 5.37 Relationship between the thickness of plasma induced damage inversion layer and the RF power

induced by the H_2-based etchant containing N_2 decreases remarkably. However, this etchant could not satisfy the requirement of the device fabrication process. The influences of adding N_2 gas to the process gas include two points: (1) The N_2 gas could not affect the H_2 etching reaction, but it serves as the stabilizer and diluent of plasma like Ar^+, so that it can reduce the energy of etching ions in a matter; (2) The introduction of the N_2 gas can change the concentration of H^+ ions in the etching plasma. For the first point, since the flow of N_2 is lower than that of Ar, its role of diluting plasma cannot be obvious. Next, the influence of the N_2 gas on the H^+ concentration in etching plasma is analyzed.

The dissociation of N_2 in the plasma is [7, 11, 41, 46, 47]:

$$N_2 + e^- + 9.46\,eV \rightarrow N + N^-$$
$$N_2 + e^- + 6.17\,eV \rightarrow N_2^* + e^-$$

where N_2^* is the activated N_2 molecule. And the H_2 molecules in plasma can also be dissociated to H atoms and H^+ ions [7, 11, 41, 46, 47]:

$$H_2 + 4.5\,eV \rightarrow 2H$$
$$H + 13.6\,eV \rightarrow H^+ + e^-$$

Regardless of either H atoms or H^+ ions, both can react with N^- ions and N_2^* activated molecules in the plasma to generate NH_3 or $NH_4{}^+$, so that the content of H element (including H atoms and H^+ ions) in the plasma decreases and the etching damage can be reduced correspondingly. It can be seen from Fig. 5.38a that the rate of etching ZnS using $H_2/N_2/Ar$ is far lower than that of H_2/Ar and rarely changes with the RF power increasing. This indicates that the H content in the process plasma of $H_2/N_2/Ar$ is very low, and the N_2 gas plays a role of H consumption etchant in the etching process. However, the H element is the only chemical etchant

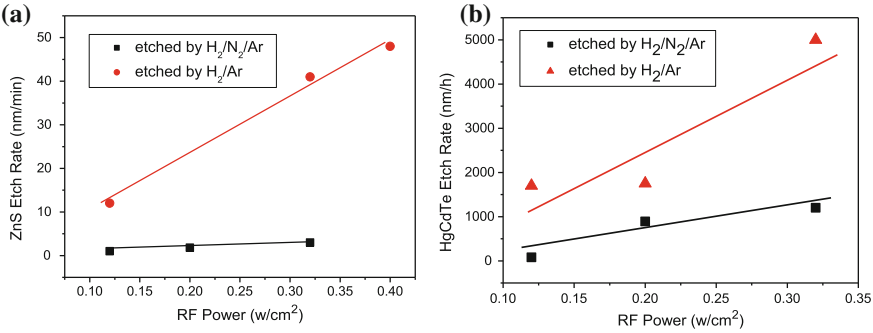


Fig. 5.38 Relationship between the etching rate of H2-based process and RF power

in the process gas of $H_2/N_2/Ar$, so the decrease in the H content (including H atoms and H^+ ions) could lead to the reduction of etching rate directly, which is verified by the experimental results. Figure 5.38b provides the comparison of the etching rates of the H_2-based process gas with and without N_2. The two rates have a great difference at the same RF power, and the rate difference could increase with the RF power increasing. This is because the H content in the H_2-based process gas containing N_2 gas is very low, and even with increasing RF power, the etching rate in this process remains very low in the low-energy etching condition due to the absence of chemical etchant.

For the H_2-based process gas containing N_2, the existing problem is that the rate of etching HgCdTe is excessively low. However, the CH_4-based process gas is different. The chemical etchants in the CH_4-based process gas include two free radicals: –H and $–CH_3$. The radical $–CH_3$ can serve as a chemical etchant in the case that the radical H is suppressed. Therefore, at the same RF power, the damage induced by the CH_4 etching process are significantly less than that of the H_2-based etching process without N_2, while its etching rate is higher than that of the H2-based etching process containing N_2. Accordingly, the CH4-based etching process, in which the N_2 gas proportion increases to acquire the desired reduction of etching damage, is mainly applied for the dry etching technique of HgCdTe.

e. Etching process for reducing damage

According to the above research results of induced damage reduction, the ICP etching isolation technique of reducing damage for HgCdTe micro-mesa arrays can be achieved by decreasing RF power, increasing plasma density and introducing the H consumption etchant. Figure 5.39 summarizes the thicknesses of plasma induced damage inversion layers caused by different etching process gases. It can be seen that the CH_4-based process gas with the enhanced N_2 gas proportion is a preferred etching process. On the one hand, it produces least etching damage and can be used for the preparation of HgCdTe mesa junction devices; on the other hand, its RF power is more suitable. The lower RF power could lead to a low etching rate and low duty cycle, so that the demands of preparing HgCdTe micro-mesa IRFPAs are

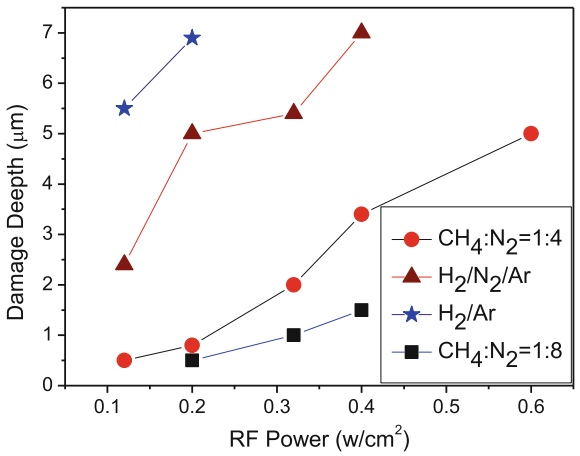


Fig. 5.39 Relationship between different process gas proportions and thicknesses of damage layers

difficult of satisfy. The measurement of *I*–*V* and *R*–*V* characteristics indicates that the thickness of electrical damage inversion layers in this process is less than 1 μm, so this process can be used for the dry etching of HgCdTe mesa junction devices.

(4) Electrical damage repair

Considering the features of plasma dry etching process and HgCdTe material, the damages are difficult to eliminate merely by improving etching process, and there are always induced damage layers remaining on the etching surface. Next, the investigations on the repair of weak damage layer on the etching surface using high temperature annealing treatment are introduced (Fig. 5.40).

The high temperature annealing treatment could repair the lattice damage induced by dry etching and reduce the defect density, so it can be used to repair the damages caused by plasma etching. It was reported [43] that baking can make the unbounded H atoms escape from the material surface. Thus, the heat treatment has

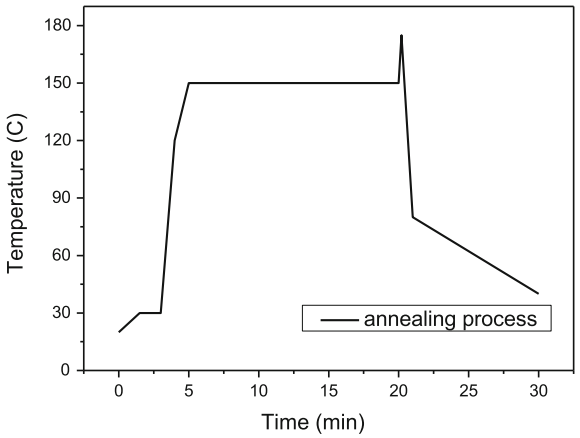


Fig. 5.40 Schematic diagram of high temperature annealing treatment

an important role on reducing the thickness of dry etching induced damage inversion layer.

Sun et al. studied the influence of heat treatment on the performance of HgCdTe IR detectors. The optimal heat treatment process was used to anneal HgCdTe samples etched by plasma dry etching. The samples were the HgCdTe material of $x = 0.239$. The etched samples were annealed using SRO-702 meltback furnace (ATV Inc.) in the N_2 atmosphere. The annealing condition is shown in Fig. 5.43.

The damage measurement results are provided in Fig. 5.41. The high temperature annealing has a certain effect on damage repair, especially in the case of severe damages. However, the effect is not remarkable for the H consumption $CH_4/N_2/Ar$ process, in which the plasma induced damages are less. It indicates that the existing high temperature annealing processes have a remarkable repair effect on etching damages caused by H etchant, but its effect on etching damages caused by Hg interstitials is not good.

Nevertheless, the measurement results of *I–V* and *R–V* characteristics indicate that the high temperature annealing process leads to a change in the material properties of HgCdTe, that is, the ohmic contact resistances increase. Figure 5.42 illustrates that the *R–V* curves of the samples that underwent wet chemical etching for a short time after high temperature annealing present a resistance property, while the unannealed samples still present a *p-n* junction property. This indicates that the etching damage inversion layer becomes thinner. After annealing, the dynamic resistance increases by an approximate order of magnitude. The reason may be that the performance of ohmic contact to HgCdTe material becomes poor after high temperature annealing. To apply the high temperature annealing process in the dry etching fabrication process of HgCdTe, the influence of high temperature on passivation layer is required to be studied. Therefore, although the high temperature annealing treatment has a certain effect on repairing plasma etching damages, further exploration and optimization are still required of the annealing process.

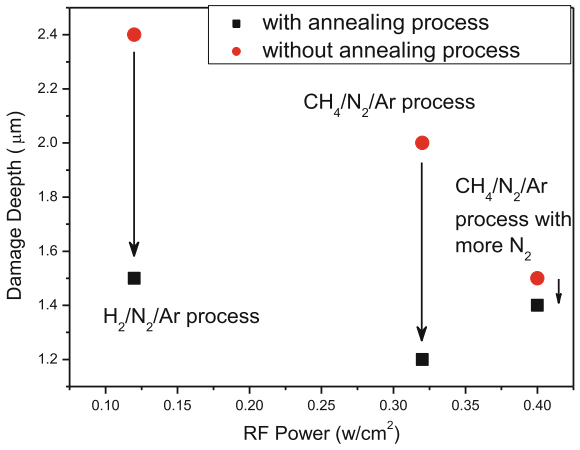


Fig. 5.41 Influence of high temperature annealing treatment on etching damage of HgCdTe

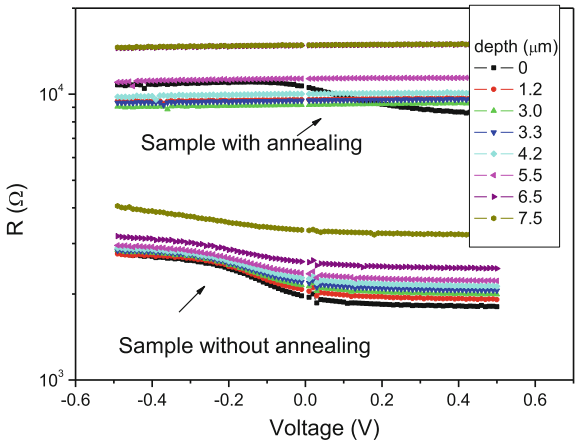


Fig. 5.42 Variation in the *R–V* curves before and after high temperature annealing treatment

4. Wet chemical process for micro-mesa array

The wet cleaning and etching are the processes of exposing semiconductor wafers to liquid chemicals to remove unwanted material (film or defect) from the surface. The wet chemical process for HgCdTe micro-mesa array includes many technologies, such as cleaning, etching, and removing of photoresist, especially various wet chemical processes after the formation of mesa. Due to the existence of trenches, these processes are quite different from the conventional planar wet chemical processes. By the experimental analysis of mesa devices and cleaning processes, it is found that since the sidewalls and surface junctures suffer from the wipe force due to the existence of mesa structure, the surface passivation layer is easy to lift off, thereby resulting in the large area abscission of surface passivation layer and poor process results.

Next, taking the cleaning and etching processes before passivation of micro-mesa device as examples, the differences between them are demonstrated. Before the passivation of a micro-mesa array device, the cleaning process is used to clean up the polymers residing on the surface of mesa device after dry etching, while the etching process is an effective way to properly eliminate the physical damage layer generated by mesa etching. In the specific process, it is discovered that the phenomenon of liftoff and abscission of surface passivation layer could occur as the mesa device is cleaned using the planar cleaning process (shown in Figs. 5.43 and 5.44).

In the specific process, the liftoff of surface passivation layer can be reduced by wiping softly, and the previous ZnS layer is removed after cleaning and the new ZnS layer is grown again to overcome the above problem. Bromine methanol and hydrochloric acid are used to etch the pattern films, respectively. Photoresist is used to protect the front surface, and then the bromine methanol is used to etch the HgCdTe and the ZnS layer is etched with hydrochloric acid. In this way, the damage layers on the sidewalls and bottom surfaces are etched by bromine

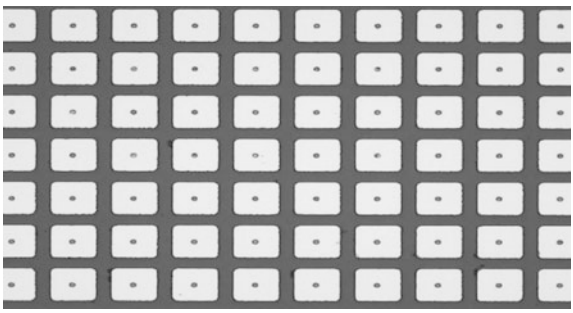


Fig. 5.43 Planar device surface

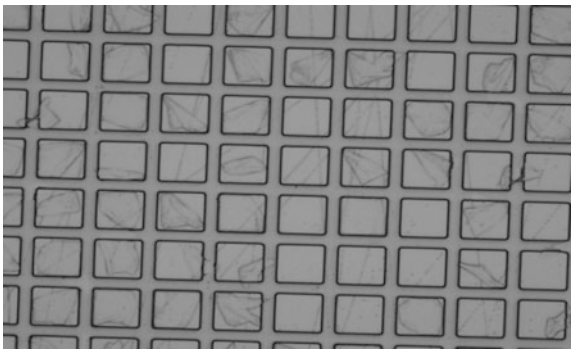


Fig. 5.44 Mesa device surface

methanol, and the CdTe/ZnS composite layers on the front surface are protected by photoresist. After removing the photoresist, the ZnS layer is etched by hydrochloric acid, and a good process effect can be achieved.

Currently, due to the small size and irregular shape of HgCdTe chips, no automatic wet chemical equipment can satisfy the process requirement of HgCdTe micro-mesa devices. Therefore, it is more feasible to select the wet chemical equipment with corresponding functions in association with a suitable manual operation.

5. Prototype chip of dry etching

A stable, dry etching process offering good etching surface and its etching rate, loading effect, ARDE, uniformity, and electrical damage has been introduced and now, based on this etching process, a long wave prototype chip is prepared and analyzed [11].

Figure 5.45 provides the electrical and spectral characteristic curves of the prototype chip etched by dry etching. The samples are based on p-type long wave HgCdTe material of $x = 0.215$, and the condition of dry etching is as mentioned above. After the formation of n^+-on-p structure through large area ion implantation, dry etching and wet chemical etching techniques for removing weak damage layers are used to isolate the micro-mesa array and open the common zone, respectively. After passivation, metallization, and indium bump preparation, the prototype chip is

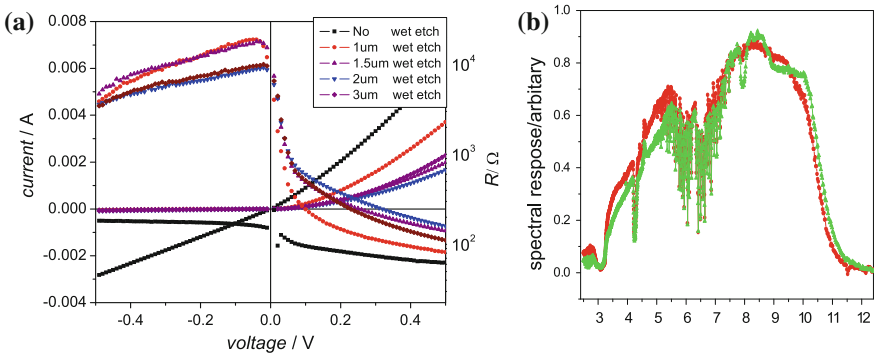


Fig. 5.45 Measurement results of prototype chip fabricated by dry etching. **a** *I–V* and *R–V* characteristic curves; **b** Spectral response curves

complete. According to the concept of the material chip technique, five test structures with wet chemical etching for 1, 1.5, 2, 3 μm and no etching are integrated on the same chip.

As shown in Fig. 5.45a, in the prototype chips, the diode property of the photodiode becomes very obscure without removal of damage layer, while the *I–V* and *R–V* characteristic curves of photodiodes with etching surface removal of 1 and 1.5 μm are normal. However, after removing the etching surface for 2 μm or more, the performances of the photodiodes are not improved but reduced. The reason is likely that when the etching depth for removing surface damages caused by dry etching using wet chemical etching is more than 2 μm, crescent-shaped trench sidewalls of micro-mesa will occur due to the isotropy of wet chemical etching, and the passivation film is difficult to coat the crescent-shaped sidewalls. Figure 5.45b shows the spectral response curves of the prototype chip, and it can be seen that the cutoff wavelength of this device is 10.7 μm.

The measurement results of the prototype chip indicate that the micro-mesa array isolation using dry etching cannot achieve desirable device performance. Nevertheless, after the surface is etched by dry etching and is removed to a depth of 1–1.5 μm, the device provides well in *I–V* and *R–V* characteristics. It indicates that the dry etching process indeed results in the electrical damage layer of 1–1.5 μm thick on etching surface. It is in agreement with the measurement results of the thickness of dry etching induced electrical damage layer.

5.2.2 *Micro-Mesa Photolithography*

1. Uniform coating technology of photolithography for deep micro-mesa

The chief problem of the photolithography technology for high-accuracy micro-mesa array is the uniform coating technology of photolithography on the top

of mesa, trench bottom, and sidewall [6]. Because the sidewall is very steep and the trench is very deep, the thickness uniformity of the photoresist on these locations is difficult to ensure. Usually, the photoresist is accumulated in the trenches, while there is rarely photoresist covering the top edge of the sidewall. This could make the exposure time and developing time difficult to control during the photolithography process. In order to ensure the integrity of patterns, the exposure time and developing time should be prolonged. However, this could cause the overdeveloping and loss of photoresist mask on the top and sidewall of micro-mesa, so that the areas of mesa surface regions protected by photoresist reduce and the thickness of photoresist on the sidewall becomes very thin or nonexistent. Finally, when the electrode isolation process is performed, the photoresist cannot effectively protect the passivation films on the top and sidewall of micro-mesa array.

The photoresist is difficult to coat on the surface and sidewall of the mesa structure using the conventional coating method. So, in the same exposure, some regions with thicker photoresist cannot be developed completely due to insufficient exposure, and, as a result, the desirable patterns could not be achieved. Moreover, after the subsequent processes such as etching, photoresist accumulated in the trenches may be difficult to remove (shown in Fig. 5.46). In view of the existing difficulties of photoresist coating technique for climbing metallization of micro-mesa array, the process should be optimized in the aspects of the fluidity performance of different photoresists, the shape feature of micro-mesa and photoresist coating rate.

The distribution of photoresist on the mesa top, the bottom, and sidewall of trenches becomes uniform after optimization of coating process shown in Fig. 5.47, especially, the proper thickness of photoresist is obtained on the mesa top and the top corner on the slope surface. It is not only helpful for the control of exposure time and developing time in different regions, but also ensures that the passivation films are destroyed when opening contact holes.

There are two technical approaches to adopt in order to solve the problems in the photolithography of deep micro-mesas using a conventional spin-coating method:

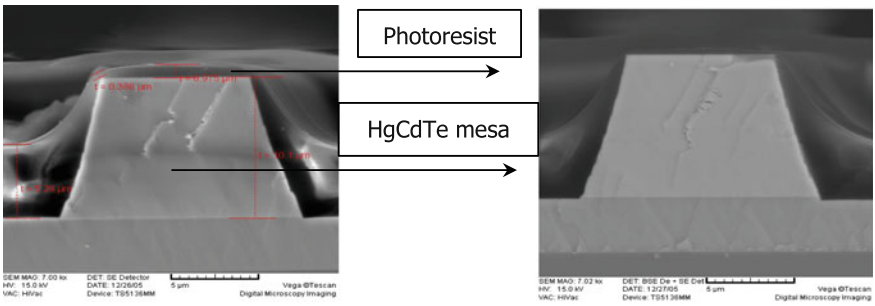


Fig. 5.46 Distribution of photoresist obtained from conventional spin-coating process

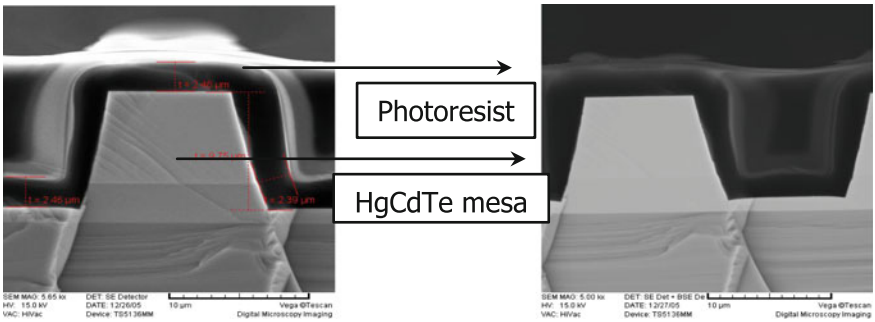


Fig. 5.47 Distribution of photoresist obtained by optimizing spin-coating process

- Surface planarization processing technique: a film is grown to reduce the surface roughness, and then the photoresist coating and the subsequent exposure and developing processes are carried out. This method is more complicated technologically.
- Photoresist spray coating technique: using the photoresist spray coating equipment, uniformly coated photoresist on the uneven surface can be achieved (as shown in Fig. 5.48), and the uniformity of exposure and developing processes is ensured. The method is relatively easy and feasible.

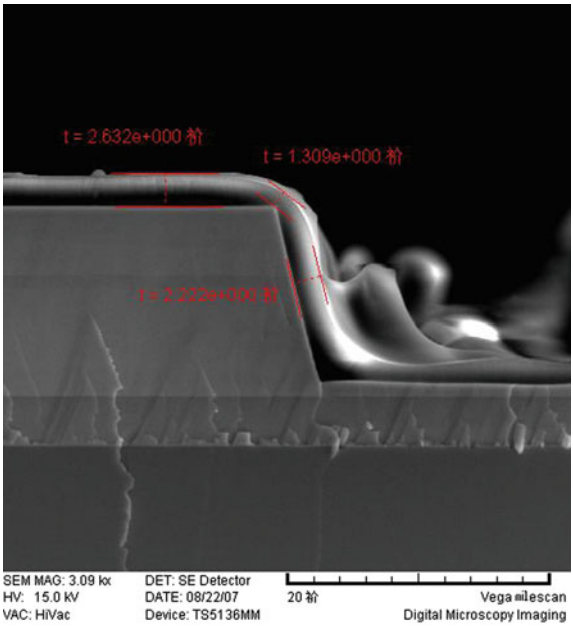


Fig. 5.48 Coating photoresist on the uneven surface

In the spray coating system, the photoresist is sprayed out of the jet with a certain pressure after which the photoresist does not possess high fluidity. This effectively reduces the local (edge) accumulation of photoresist using spin-coating method due to the chip morphology.

Based on the photoresist spray coating technique, research work is performed in the aspects of the fluidity performance of different photoresists, the shape feature of micro-mesa, photoresist coating rate, curing temperature, etc. In this way, the distribution of photoresist on the mesa top and the bottom and sidewall of trenches can be uniform, especially if the thickness of photoresist on the surrounding areas of micro-mesa top becomes very thick. It is not only helpful for the control of the exposure time and developing time in different regions, but also ensures that the passivation films are destroyed when opening contact holes.

2. Exposure technique for photolithography of deep micro-mesa

For the uneven surfaces on the mesa top and the bottom and sidewall of trenches, if the photolithography is required to be performed at the same time, the linewidth could change for the existing contact exposure method, and the contact at lower locations is unable to be good for exposure, thereby leading to distortion of patterns. In this case, the performance of the stepper should be improved. The advanced noncontact stepper could be used to achieve favorable patterns, which is helpful for the improvement of total uniformity of the mesa devices.

For HgCdTe micro-mesa devices, the more advanced direct write stepper can be selected except for the noncontact stepper. The comprehensive photolithography techniques involving the direct write continuous scanning and the step printing with different scaling factors are adopted in the nonmask stepper based on a digital micromirror device (DMD), and smooth pattern profiles can be developed. Also, the exposure efficiency can be improved. The light from the 405 nm (or 365 nm) LED source is projected onto the DMD array after the calibration of a group of optical beam splitter systems. The DMD array can provide the projecting patterns for exposure according to the exposure requirement, and the patterns are projected onto the wafers in a certain scaling factor through the optical projecting system to achieve the transfer of patterns. The conversion of scaling factors is achieved by selecting projecting lens with different magnifications. The projected patterns are obtained by the segmentation of wafer layout and data processing without the preparation of photo mask.

5.2.3 *High-Quality Sidewall Passivation Technique of Micro-Mesa Arrays*

After the formation of deep micro-mesas, the vertical passivation of micro-mesa FPA chips becomes a technical challenge, because it involves film growth in both vertical and horizontal directions, the distribution of bond stress and reliability.

The sidewall passivation is as important as the surface passivation, and both of them have a significant impact on the performance of detectors.

The performance of HgCdTe devices is closely correlated to the passivation process. A good passivation process of the surface and sidewalls can reduce surface damages, interface states, surface leakage current, surface recombination velocity, and $1/f$ noise on the surface of HgCdTe devices. The process can also increase the dynamic resistance and the reverse breakdown voltage of detectors, thereby improving the device performance. Accordingly, it is required to prepare a compact and high-resistance passivation layer in HgCdTe device technology.

Higher growth temperature and energy are required for the growth of compact and high-resistance passivation films, however, the HgCdTe material cannot endure long duration heating at temperatures higher than 80 °C, and the higher growth energy also causes new surface damages to HgCdTe devices. Thus, four aspects of investigations are required as follows:

1. Design of composite film system

For HgCdTe detectors, there are some passivation films reported early as follows:

Film system	Advantage	Disadvantage
ZnS	Energy gap = 3.9 eV (77 K), good insulation	Poor matching with HgCdTe
CdTe	Energy gap = 1.6 eV (77 K), good matching with HgCdTe	Poor insulation
CdZnTe	Perfect matching with HgCdTe	Poor insulation
CdTe/ZnS composite film	Perfect matching with HgCdTe	Good insulation

With the development of HgCdTe technology, the CdTe/ZnS composite passivation film is the most widely applied technique for HgCdTe. The lattice mismatch between CdTe and HgCdTe is small and fewer interface charges are obtained by a CdTe passivation film, which is helpful for the achievement of the surface flat-band condition. The CdTe film makes a graded layer form at the interface of HgCdTe and CdTe, while the ZnS film can serve as a good insulation layer.

2. Ultralow damage surface passivation technology

At present, the main methods applied in the passivation technology for HgCdTe devices include: thermal evaporation, sputtering, magnetron sputtering, MBE, etc.

The thermal evaporation technique is one in which the source material is heated to melt sufficiently and then evaporate. The passivation technology contains three stages:

1. Material evaporation (evaporation stage)
2. Transportation of particles to the substrate in vacuum (transportation stage)
3. Condensation on the substrate (condensation stage)

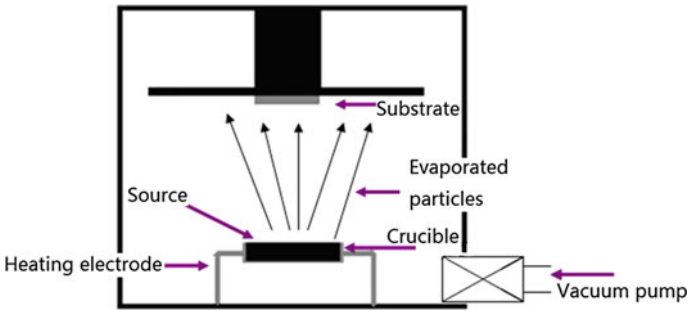


Fig. 5.49 Schematic diagram of vacuum thermal evaporation

The feature of this technology is that a high vacuum is required in the vacuum chamber, so that the evaporated particles have long mean free paths and reach the sample surface effectively, thereby improving evaporation efficiency. Hence, the vacuum pump of evaporation system is usually connected to the main vacuum reaction chamber through a pipeline with a large cross section (shown in Fig. 5.49).

For equilibrium pressure or saturated vapor pressure, the evaporation rate and the condensation rate of source material are equal. The saturated vapor pressure is given by

$$p_s = \mathrm{Ae}^{(-\mathrm{BT})},$$

where A is the integration constant, B is the parameter closely correlated to the evaporation process. Very small temperature changes can lead to a great change in evaporation rate. For evaporation, either the resistance heater or e-beam (in the case of e-beam evaporation) can be applied to the heating source material. For evaporation systems applied in the HgCdTe device passivation technology, control of substrate temperature must be carefully maintained.

The sputtering technology is a plasma process. During sputtering, inertial gas ions (usually Ar ion) are accelerated and bombard the target (cathode), and the emergent particles from target form a vapor column and then deposit on the substrate. The sputtering process includes four stages:

- The plasma is generated by the collision and ionization of the gas atoms (Ar) with high-energy electrons
- The plasma bombards the target
- The bombarded target atoms are transported onto the substrate
- The emergent target atom groups are deposited on the substrate

In order to achieve the maximum deposition rate of target atoms on the substrate, the distance between the substrate and the target is usually 5–10 cm (up to 35 cm in some sputtering systems). The atoms arriving at the substrate have higher energy (3–10 eV) than that in the evaporation technique (0.2 eV). The collision between target atoms and gas atoms causes the atoms to arrive at the substrate from different

directions. Therefore, the passivation process is largely isotropic, and the step coverage of sputtered films is better than that of evaporated films. Sputtering provides for a stronger particle collision to the substrate surface than evaporation, so that the lattice of substrate material is easier to destroy.

The ions are generated by the plasma discharge between two electrodes, that is, they are caused by the inelastic collision between free electrons and gas molecules (Ar mostly). The required ionization energy (15.7 eV for Ar) comes from the electron kinetic energy. The kinetic energy of Ar^+ ions is in the range of 10 and 1000 eV due to the acceleration by the electric field between the cathode and the anode. During collision with the substrate, the Ar^+ ions transfer a part of their kinetic energy to a finite (around 1000 atoms) lattice through a series of quasi elastic collisions. If the energy is higher than the surface bonding energy, then these atoms will escape from the lattice. The mean amount of bombarded atoms or the sputtering yield S is given by

$$S = \text{Amount of bombarded target atoms}/\text{Amount of colliding ions}$$

The sputtering yield is determined completely by the angle of Ar^+ ions bombarding the target. When the incident angle is 0°, the ions fly parallel over the target surface, so the probability of ejecting atoms out of the target is very low. When the collision angle is 90°, the momentum of ions is transferred onto the target, but cannot eject atoms from the target. Usually, the maximum sputtering yield is achieved at an incident angle of approximately 60°. Meanwhile, *S* is also determined by the mass relationship of ions (Ar) and target atoms and the bonding energy between atoms.

In the conventional sputtering systems, only a few electrons collide with Ar atoms, causing ionization. Most electrons accumulate on the anode so that the substrate is heated, which has a significant impact on HgCdTe devices. On the basis of common sputtering, the amount of electrons used for Ar^+ ionization is increased by the applied electromagnetic field in magnetron sputtering (shown in Fig. 5.50). The electrical field and the magnetic field cause the charged carriers to move in a circular motion in the plasma, where the deflection radius of electrons is more than that of ions. Accordingly, the electrons are collected near the target, so that a higher

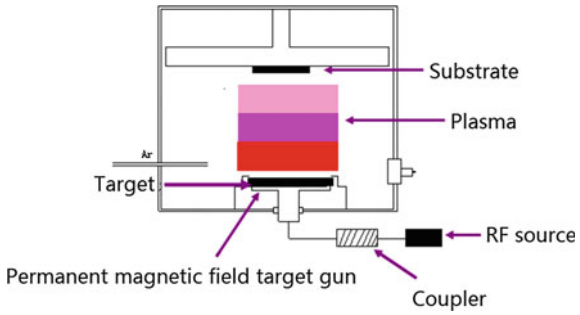


Fig. 5.50 Schematic diagram of magnetron sputtering system

ionization probability and sputtering rate are obtained. The ($E \times B$) trend makes the electrons flow onto the anode rather than the substrate, so the heating of the substrate is reduced. However, the mismatch of the stoichiometry in the grown compound films needs to be noticed.

The MBE process is a technique where molecular/atomic beams form the film on the substrate in a high vacuum environment (shown in Fig. 5.51). The main feature of MBE is that the motion of atoms or molecules with sufficiently long mean free paths causes them to have essentially ballistic characteristics, but without gas phase reaction. The adhesion coefficients of the elements with low vapor pressure tend to be consistent, while the elements with high vapor pressure need to be used excessively to make up for the low adhesion coefficient, so that the surface dynamics and diffusion energy satisfy the two-dimensional growth. However, the growth temperature of MBE is high, which has a significant impact on long wave HgCdTe material.

During the bombardment, the HgCdTe material is easily damaged and surface defects are produced, so that the device performance is reduced. On the other hand, the performance of the HgCdTe material is sensitive to temperature. So investigation into low damage surface passivation technology is critical. At present, the magnetron sputtering system is usually used for passivation. Magnetron sputtering has an advantage of low bombardment to the substrate, as a result of plasma constrained near the target. The process is helpful for reducing the damage to the substrate and lowering deposition temperature and is suitable for the passivation process of HgCdTe.

3. Difficulty analysis and technological method of passivation for HgCdTe micro-mesa devices

For the HgCdTe mesa devices, the surface passivation technique needs not only to solve the problem of surface leakage current in the planar monochromatic devices, but also to reduce the leakage current of mesa sidewalls through passivating the sidewalls of exposed HgCdTe mesas by growing high-quality passivation layers after deep micro-mesa etching.

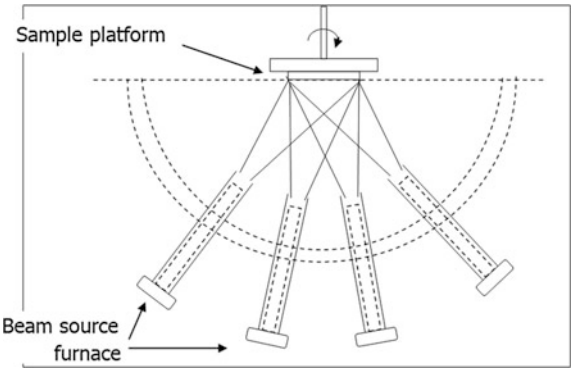


Fig. 5.51 Schematic diagram of molecular beam epitaxy system

(1) Low temperature and low damage composite layer passivation technique

For the exposed HgCdTe surface, the composite passivation film of CdTe/ZnS is required to be applied. The CdTe film is grown directly on the surface of HgCdTe, and it has a good lattice matching with HgCdTe and is helpful for reducing the surface state of HgCdTe while the ZnS film can provide favorable optical and high resistance properties. Considering the requirement that HgCdTe cannot endure high temperature and the collision of high-energy beam current, the following aspects should be studied:

- Low damage passivation method of CdTe film: the low temperature and low damage growth method of CdTe films is explored to reduce the damage caused by the CdTe passivation process to the HgCdTe surface and ensure the interfacial and high resistance properties of passivation layers;
- Temperature-change passivation method of ZnS film: for the problems that the low temperature growth of the ZnS film brings poor adhesion force and loose film structure and the HgCdTe devices cannot endure the temperature higher 80 °C for a long time, the temperature-change growth method of ZnS film at high initial temperature is studied, and a compact ZnS film with high resistance and good adhesion is achieved to improve the reliability of devices without the reduction in device performance;
- Component control technique of CdTe film: for the problem of the proportion mismatch of Cd and Te, the external high-purity Te technique and the secondary electron energy spectrum monitoring are applied to achieve Te-rich CdTe passivation film. This method can effectively reduce the Hg-rich defect state density on the HgCdTe surface caused by the heat treatment and greatly reduce the dark current, thereby improving the dynamic resistance of detectors.

(2) Optimal design of micro-mesa passivation technological flow

For the layered mesa devices, the micro-mesa passivation technology is divided into two steps: (1) the problem of surface leakage current can be solved based on the CdTe/ZnS composite film technique in the planar monochromatic device processing; (2) the problem of the uniform distribution of the passivation film on the sidewall of mesa devices is solved. Figures 5.52 and 5.53 provide the passivation of planar and mesa devices, respectively. The sidewall of mesa is covered by the CdTe/ZnS composite film with the thickness similar to the surface passivation film. If the film on the sidewall is too thin, the passivation effect becomes poor thereby influencing the device performance; if the film is too thick, it could result in a large change in the surface passivation film structure, so that the reliability of the device is affected.

On the basis of the existing passivation technique for HgCdTe planar monochromatic devices, the magnetron sputtering system or thermal evaporation system is selected, and the low temperature and low damage fractional passivation method is applied combined with wet chemical process. Also, the proper passivation film for sidewall is selected to achieve mesa devices with good surface and sidewall passivation.

Fig. 5.52 Passivation of planar devices

Fig. 5.53 Passivation of micro-mesa sidewall

(3) Optimization technique of surface and sidewall mesa coating

The sidewall passivation of deep micro-mesa is more difficult than that of planar devices for the micro-mesa passivation technique of HgCdTe devices. Regarding the growth of passivation layer on the sidewall of mesa, the nonuniformity of the thickness of passivation layers on surface and sidewall is easily caused thereby influencing the passivation quality due to the shadow effect. Specifically, a fracture at the juncture of sidewall and surface likely occurs, so that the exposed HgCdTe surface could generate remarkable surface leakage current.

So far, the conventional sidewall passivation methods for HgCdTe include magnetron sputtering passivation and thermal evaporation passivation. And the optimization methods of magnetron sputtering passivation process are given:

- Increase the sputtering pressure to enhance the scattering of sputtered particles colliding with Ar^+ ions, in order to improve the directivity of grown films;
- Reduce the distance between the target and the sample, in order to increase the incidence angle of sputtered particles entering the trenches;
- Confocal sputtering growth method: the magnetron sputtering method of focusing growth is performed on the sample using adjustable angle multiple targets. It has the advantage that the growth directivity is easy to control, as well as the disadvantages of large bombardment intensity to sample and high temperature rise.

The optimization of thermal evaporation methods is given:

- Planet carrier of sample platform: the sample holder has a planet orbiting structure, so that the sample can revolve around the bearing and on its own axis. It has the advantage of good uniformity of film, as well as the disadvantages of complicated loading and poor mechanical stability.
- Reducing the distance between the sample and the evaporation boat can increase the incident angle of evaporated particles entering the trenches, thereby improving the uniformity of the thickness of surface and sidewall passivation layers.

At present, the magnetron sputtering method and the thermal evaporation method could also be optimized further. In order to reduce or avoid the problem on the shadow of film growth, the planet carrier can be designed, and the thickness distribution of passivation layers on the surface and sidewall can be measured by SEM. Therefore, the thickness ratio of passivation layers on the surface and sidewall could be improved effectively.

(4) Evaluation of passivation effect on the surface and sidewalls in HgCdTe micro-mesa devices

The surface passivation on deep trench micro-mesas is more difficult than that in the planar device process. Due to the shadow effect, the phenomena of poor adhesion, too thin thickness, and loose structure occur easily in the growth of passivation layers on the sidewall of a mesa, thereby influencing the passivation effect significantly. Accordingly, it becomes quite important to evaluate the passivation effect of surface and sidewall precisely.

First, the thicknesses of passivation layers on surface and sidewall are observed applying SEM, shown in Fig. 5.54. The aim of the passivation process is to grow the CdTe/ZnS composite films and make the thicknesses of both films equivalent. However, in fact, the thicknesses of the composite films on sidewalls and at the juncture of surface and sidewall are relatively thinner. The distributions of the passivation layers on the top of the mesa and the bottom and sidewall of the trench are investigated using the SEM analysis of the cleavage profiles. Figure 5.55 provides the cross-section view of the distribution of passivation films in the HgCdTe micro-mesa detector array. It can be seen that the distribution of the passivation layers on the top of mesa and the bottom of trenches is uniform, and the thickness of the passivation layer on sidewall is more than 45 % of that on the mesa top. In other words, the passivation effect can be achieved effectively.

Next, in order to optimize the technical condition, the interfacial properties of passivation layers are investigated by the *C–V* measurement and the component is analyzed by photoelectron energy spectrum. The surface passivation device with a MIS structure is fabricated (shown in Fig. 5.56), and the *C–V* curves at high and

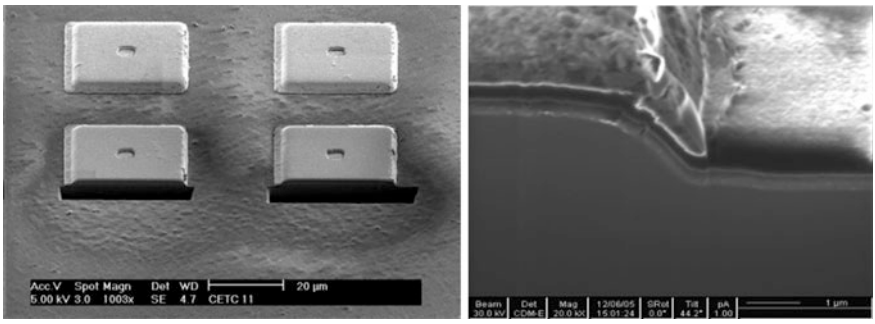


Fig. 5.54 SEM images of sidewall passivation

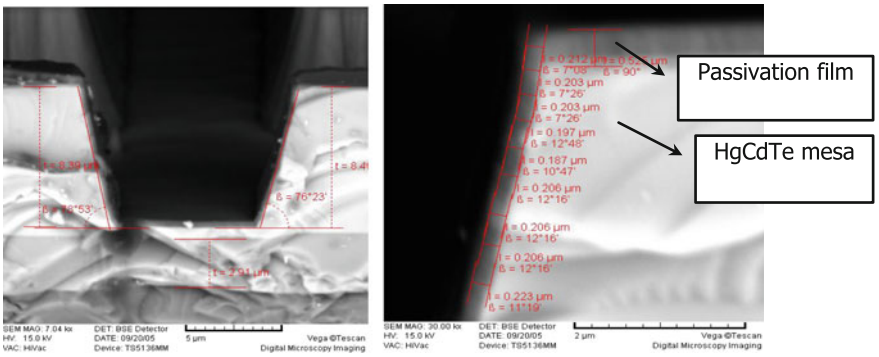


Fig. 5.55 Cross-section view of the distribution of passivation layers in a micro-mesa pixel array

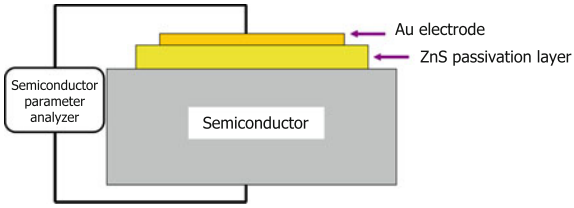


Fig. 5.56 *C–V* measurement structure of surface passivation

low frequencies are measured by semiconductor parameter analyzer. Finally, the fixed and movable charges at the interface of passivation layer can be analyzed further.

5.2.4 Metallization of Micro-Mesa Array

The electrode fabrication is an important step in the HgCdTe IRFPA process. The common port of micro-mesa detectors is required to be extracted from the bottom of trenches, which is more difficult than that of planar detectors in the small scale. Due to the specificity of HgCdTe material, the special requirement is brought to the electrode fabrication. That is, on the one hand, damage from technical processes to the material should be avoided; on the other hand, good thermodynamic contact with HgCdTe should be formed on the premise of ohmic contact formation. Next, the layer selection of deep micro-mesa signal and the climbing metallization method of lateral extraction of selected zone are introduced.

1. Ohmic contact of electrode

In the electrode fabrication, it is very significant to form an ohmic contact between metal and the semiconductor. If the metal and the semiconductor are connected simply, the work function difference between them could cause the

redistribution of carriers, so that a barrier is created at the interface forming the rectification characteristics. This additional unilateral conductivity leads the *p-n* junction to work abnormally. Therefore, the semiconductor surface layer should be treated to eliminate rectification effects. The ohmic contact should satisfy two conditions electrically: (1) the *I–V* characteristics at the contact interface present linear symmetry, (2) the contact resistance should be as low as possible.

In the MS contact theory of ideal semiconductors, it indicates [49] that the ohmic contact can be formed by selecting a metal with suitable work function (for n-type semiconductor, $W_m < W_s$; for p-type semiconductor, $W_m > W_s$). However, a large number of experimental measurements [50] demonstrate that the contact barrier is not determined by the work function of the metal due to massive surface states existing on the semiconductor surface. In the practical fabrication of chips, the ohmic contact is usually formed by tunneling effect.

Taking n-type semiconductor as an example, the rectification contact can be formed when the metal is contacted with the low-doping n-type semiconductor. When the doping concentration becomes higher, the rectification characteristics degrade and the *I–V* characteristic curves trend to be symmetric forming the ohmic contact. In the energy band theory, when the doping concentration is low, the barrier at the interface becomes wider and the electron transportation is based on the thermionic emission over the barrier. In the forward bias, the electrons can cross over the barrier into the metal; in the reverse bias, the electrons in the metal should cross over the barrier (the barrier height is Φ_B) into the semiconductor, so the *I–V* characteristics appear asymmetrical. When the doping concentration is high, the barrier becomes narrow and the probability of electron tunneling increases, so the electron transportation is based on the field emission effect caused by electron tunneling and the thermionic emission can be ignored. Hence, the *I–V* characteristics appear linearly symmetric (Fig. 5.57).

In general applications ($x \leq 0.5$), the energy gap of $Hg_{1-x}Cd_xTe$ is narrow. Many research papers indicate that the ohmic contact is formed easily between metal and n-type HgCdTe, while the p-type HgCdTe is difficult to form an ohmic contact. The theoretical calculation presented by Spicer [51] indicates that the Fermi energy levels at the metal/$Hg_{1-x}Cd_xTe$ interface are pinned in the forbidden band ($x > 0.4$) or in the conduction band ($x < 0.4$). So the intrinsic ohmic contact is difficult to form between metal and p-type HgCdTe. A very thin barrier is required to form at the interface of metal and HgCdTe, so that the carriers could tunnel. For the specific forming methods, each laboratory has its own optimal recipe. The reported metals that can form good ohmic contact with p-type HgCdTe include Sn, Cr, Au, etc. [52–54].

In many devices, there is an n^+ region of 1 μm thick in the surface layer of n-type implanted zone, and the doping concentration is 5×10^{17}–2×10^{18} cm^{-3}. The formation of the n^+ region is closely correlated with the surface damage of ion implantation and the distribution of implanted B element. The n^+ region of HgCdTe in high doping concentration could form a good ohmic contact with metal. The distribution of doping concentrations at the *p-n* junction is shown in Fig. 5.58. The distribution can be demonstrated by secondary ion mass spectroanalyzer (SIMS):

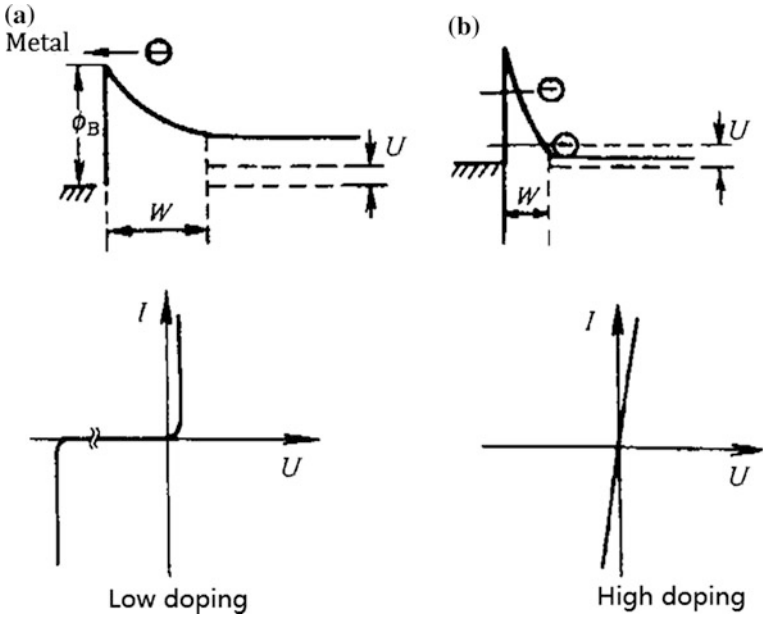


Fig. 5.57 Energy band diagram of ohmic contact formation

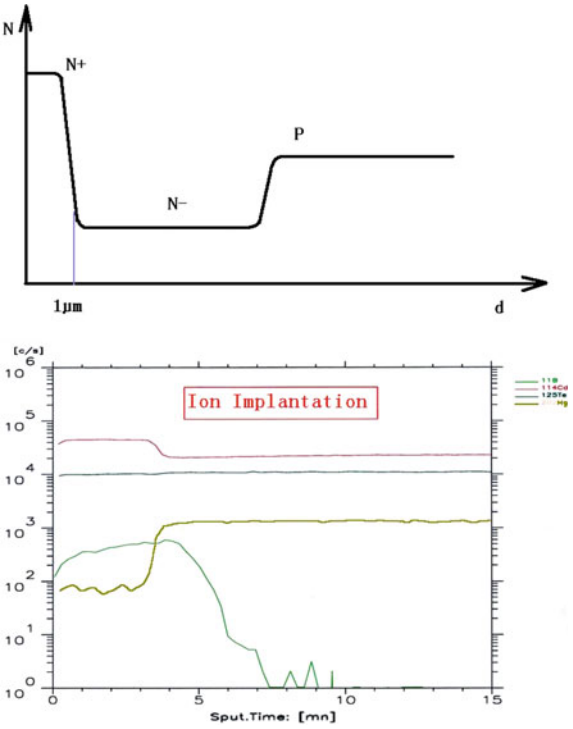


Fig. 5.58 Schematic diagram of doping concentration caused by ion implantation

According to the relationship between doping concentration and barrier width, in the high doping concentration of n^+ region, the barrier width becomes narrow, and the tunneling effect has dominated the electron transportation, that is, a good ohmic contact can be formed between HgCdTe and metal.

2. Multilayer electrode

During the fabrication of HgCdTe detectors, Hg loss or surface damage are easy to occur. Moreover, the detector is sensitive to temperature and doping. When the dopants of a deep energy level, especially Au, are introduced in the device process, the detector performance could be reduced.

The preparation of an electrode should satisfy the following requirements: (1) the process brings low damage to material; (2) the introduction of impurities should be avoided in the electrode material itself and the fabrication process; (3) deep interdiffusion should not occur between electrode material and HgCdTe; (4) the electrode has good adhesion and thermal matching to the substrate. In the device process, the fabrication of a metal contact requires good step coverage capability. Additionally, the adhesion of the electrode layer to the ZnS layer on the surface of devices is required to be good, and the obvious atomic interdiffusion influencing the performance of the passivation layer cannot occur between them. Above all, a good ohmic contact should be formed between metal and HgCdTe.

In the electrode structure, multilayer metal electrodes are adopted in HgCdTe devices. Functionally, the electrode system can be divided into four layers; ohmic contact, adhesion, barrier, and conduction. For the electrode structure in HgCdTe devices, the Au layer is used as a conduction layer, the Cr layer is used as ohmic contact, adhesion, and barrier layers, and the Pt layer is also used as a barrier layer to prevent the deep interdiffusion between Au and In in the interconnection process.

Au is a good conductor and its conductivity is $0.452 \times 10^6/\text{cm}\ \Omega$. Au has soft texture and good ductility. The Au layer with a certain thickness can play a role of relieving stress. Therefore, it serves as the conduction layer and the stress relief layer. The disadvantages of Au are that it is easy to interdiffuse with HgCdTe and then forms a deep energy level influencing the device performance, and the adhesion of Au layer to HgCdTe is poor and easy to break off. So, between Au and HgCdTe, a barrier layer is required to prevent the interdiffusion, and the adhesion layer is used to enhance the adhesion strength.

Taking Cr for an example, Cr is also a good conductor and its conductivity is $0.0774 \times 10^6/\text{cm}\ \Omega$. In the low current IR detectors, it could not produce the electromigration and can serve as an ohmic contact layer. Cr has an adhesion to nonmetal material, so it is usually used as a transition layer between the other metal and the nonmetal. The Cr atom has no strong diffusion into other materials. Usually, after the diffusion of several atomic layers (<100 Å) on the contact interface with the other material, the interatomic bonding reaches a stable value and the deep diffusion is stopped. Particularly, Cr is easy to react with other atoms such as oxygen (O) and sulfur (S) and the bonding energy is high, so a compact layer of several atoms thick could form thereby preventing the further diffusion between

materials. Thus, the Cr layer is difficult to diffuse deeply into ZnS and HgCdTe and can serve as an ideal barrier material. Hundreds of angstroms thick Cr film can play a good barrier role. The favorable barrier performance of Cr can be verified by SIMS. On the other hand, the chromium breaks away from the target surface in the form of an atom or ion during sputtering, so the Cr film has high compactness and good step coverage and is suitable for the filling process for high AR.

The disadvantage of the chromium is that it is easy to be oxidized so that its conductivity becomes poor. During sputtering, a high vacuum must be achieved near the substrate. The Cr film in the electrode can be covered by the Au film to prevent oxidization. In consideration, the chromium in the electrode can serve as the ohmic contact layer, adhesion layer, and barrier layer.

The flip chip interconnection technique is usually applied in HgCdTe detectors. As an interconnection material, indium can interfuse with Au to form an alloy. The shallow interfusion can make the interconnection more stable, while the deep interfusion could reduce the conductivity of Au. Therefore, the Pt layer is prepared on the Au electrode to prevent the interfusion between Au and In, and the thin Au film on the Pt film is used to form a shallow interfusion with In for the enhancement of firmness (Fig. 5.59).

3. Preparation and evaluation of electrode

In general, the deposition of metal electrodes in the HgCdTe device includes thermal evaporation, RF/DC sputtering, and ion beam sputtering.

The thermal evaporation process has the features of high evaporating rate and simple equipment configuration. Because of operating in the high vacuum, the films prepared by the thermal evaporation process have advantages of high compactness and small surface graininess. However, due to the long mean free path of particles in this process, the thermal evaporation has a poor coverage for the steps of high AR patterns. For micro-mesa devices, the poor step coverage can make the electrode thickness on the sidewall thinner and even create an incomplete electrode with the electrode climbing, thereby resulting in the failure or poor reliability of the electrode. Moreover, due to the low accuracy control of evaporating rate, the evaporating rate is not stable during thermal evaporation, which could make the control accuracy of the sample temperature poor. Accordingly, a high evaporating rate and long duration evaporation could cause a drastic increase in sample temperature. For HgCdTe devices sensitive to temperature, the excessively high sample

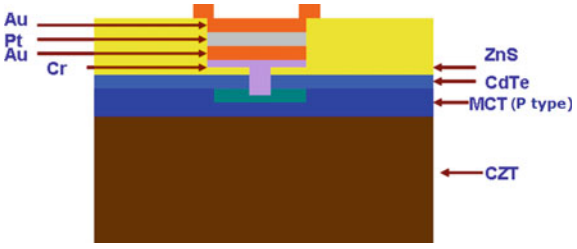


Fig. 5.59 Schematic diagram of the electrode in the HgCdTe device

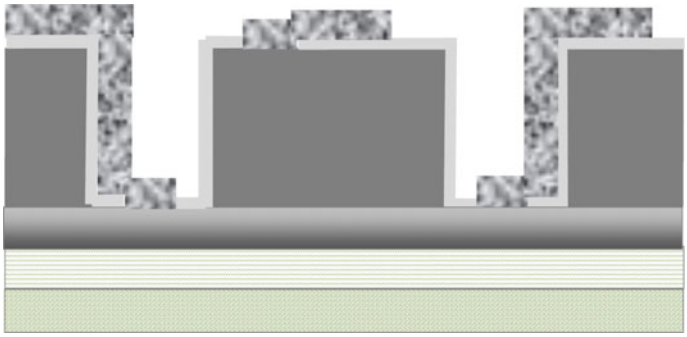


Fig. 5.60 Diagram of electrodes covering the micro-mesas

temperature could lead to failure, while an excessively low sample temperature could also cause a reduction in the adhesion of film and even the abscission of the electrode. Considering these various issues, the thermal evaporation process is not suitable for the electrode preparation of HgCdTe micro-mesa devices (Fig. 5.60).

The RF/DC sputtering process is also a conventional technology for preparing metal electrodes. In the sputtering process, the RF power is applied on the target to generate a HDP in the chamber. During the process, the target is subjected to the continuous bombardment of plasma and the target atoms are sputtered out of the material surface to deposit on the sample.

The major feature of the sputtering process is good film adhesion. Because the particles are subjected to the continuous bombardment of plasma during the deposition on the surface of substrate, the energy of deposited particles is very high and the combination of the particles and the substrate surface is better. Additionally, the sputtering rate is more stable and the technological temperature during deposition is low, so it could not cause thermal damage to the HgCdTe material.

In the ion beam sputtering process, a high-energy ion beam is used to bombard the target causing the target particles to break off from the target and deposit on the substrate surface. The critical part of the ion beam sputtering equipment is the ion beam generator—ion source. The ion source generates the HDP in the chamber by RF discharge, and then Ar^+ ions are extracted from the discharge chamber using the beam extractor to form the directional beam with a certain energy and speed bombarding the target, thereby producing the secondary particles. These particles are sputtered into the chamber directionally. If the sample is placed within the mean free path of the secondary particles, the particles could be deposited on the sample surface to form the film.

The technological parameters of the ion beam sputtering consist of beam pressure, beam current, and sample temperature. The beam pressure is the plasma potential in the chamber and determines the energy of ions bombarding the target. The beam current is the electrical quantity of Ar^+ ions emitted from the ion source per unit time. Moreover, the equipment provides the sample pallet with accurate temperature control.

The technological feature of ion beam sputtering is that the high vacuum can be achieved near the substrate (due to the Ar gas supplied during operating, the vacuum near the ion source is low, however, the distance between the substrate and the ion source is large, so the high vacuum of 10^{-4}–10^{-7} Pa can be also achieved near the substrate). The reduction in residual gases improves the film purity, compactness, uniformity, and step coverage and decreases defects. Advantages include: (1) suitability for the high AR and small aperture metallization processes such as the deposition of metal contacts; (2) the temperature rise and physical damages caused by the bombardment of plasma to the substrate placed in the plasma discharge environment for conventional sputtering are avoided; (3) the thermodynamic properties could be controlled by adjusting the film growth temperature, which determines the stress and adhesion of the film. Additionally, constant sample temperature could reduce the thermal damage caused by the heat accumulation during deposition of particles. Hence, ion beam sputtering can satisfy a series of stringent requirements such as low damage, high density, high purity, and high step coverage and has been extensively applied in the electrode preparation of HgCdTe devices.

The diffusions between metal and nonmetal and between metals are analyzed by SIMS. The distributions of the concentrations on time are provided in Fig. 5.61.

From the distribution curves of the concentration on sputtering time for MCT + CdTe + ZnS + electrode (Fig. 5.61a), the conclusions can be achieved as follows:

a. The concentration of Cr in Zn element and the concentration of Zn in Cr element are low, both of the concentration curves are steep at the junction. This indicates that the interdiffusion between ZnS and Cr is weak, and the thickness of the interdiffusion layer can be estimated to be on the order of dozens of angstroms.
b. From the interface of Au and Cr, the interdiffusion between them is also thin. This proves that the barrier effect of Cr to Au is remarkable. By the blocking of Cr, the concentration of Au in ZnS approaches to zero, which verifies the barrier properties of Cr layer.
c. At the interface of ZnS and Cr, there is a small drop occurring in the SIMS profile curve of the Cr atom concentration, as well as that of Zn. This indicates that the sputtering rate of the SIMS beam decreases so it can be estimated that the interatomic bonding energy here is high and the bonding between atoms is compact. It also verifies the properties that Cr atoms and S atoms are easy to bond firmly and there is a compact atom layer existing at the interface of ZnS and Cr. From the range of the curve drop, it can be seen that the small drop only occurs within a very small range, which indicates that the compact atom layer is very thin.

From the distribution curves of the concentration on sputtering time for MCT + electrode (Fig. 5.61b), it can be seen that: (a) the concentration of Au in HgCdTe is very low. By blocking Cr, the concentration distribution in HgCdTe

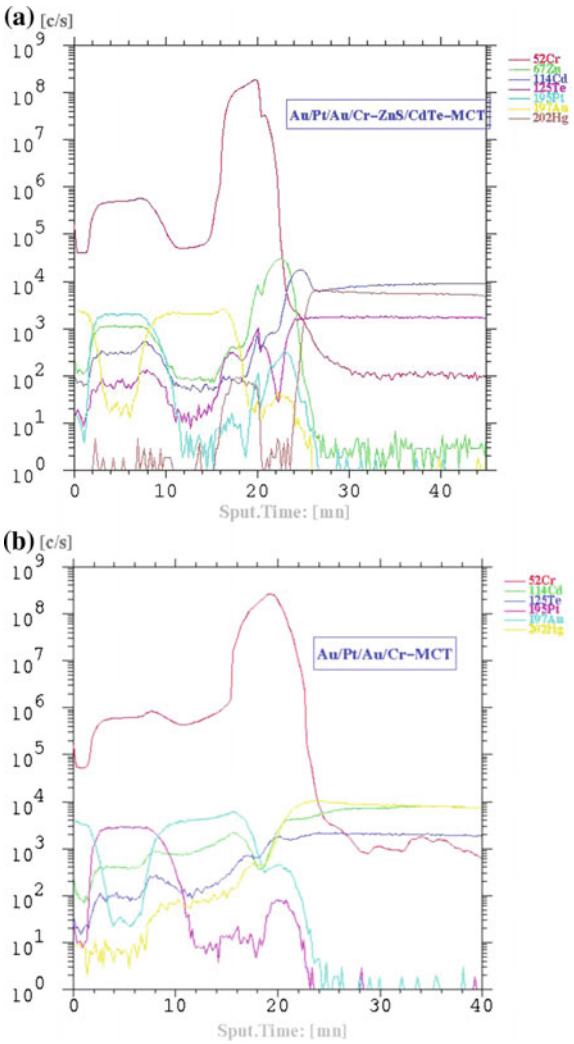


Fig. 5.61 SIMS profile results of the concentration versus the sputtering time. **a** The distribution curves of the concentration on sputtering time for MCT + CdTe + ZnS + electrode; **b** The distribution curves of the concentration on sputtering time for MCT + electrode

approaches to zero. It verifies the good blocking role of Cr layer. (b) the concentration of Cr at the interface of Cr and HgCdTe decreases to a very low value, which indicates that the diffusion of Cr to HgCdTe is very shallow. (c) the concentration of Hg in the Cr layer falls gradually, which indicates that Hg has a certain depth of diffusion in the Cr layer. This diffusion forms the low concentration of doping, which is helpful for the formation of an ohmic contact.

5.2.5 Indium Bump Preparation and Hybridized Interconnection of Micro-Mesa Array

The main role of flip chip interconnection is to connect the prepared detector chip and the corresponding Si-based readout circuit electrically [55–57]. The flip chip technique is a packaging technology pasting the chip to the base with the active surface facing the base (that is, different from the wire bonding technology, as chip with indium bumps is flipped over, and the active surface is placed downwards—so it is called flip chip) [58–60]. The flip chip is pasted onto the base using the alignment bonding equipment, and the solder bumps on the chip are positioned on the corresponding contact pads on the base. Usually, by heating air and applying a little pressure, the electrical and physical connections between the chip and the base are formed after the reflowing of solder bumps, which is a conventional packaging technology in the field of integrated circuits (IC) at present. In general, the epoxy resin filling technique is applied to improve the reliability of the flip chip [60].

1. Welding technique of flip chip

The IRFPA chip and the ROIC are welded together using the flip chip technique. However, the IRFPAs fabricated at ambient temperature operate at the temperature of 80–90 K, and the device suffers from hundreds and thousands times of thermal cycling, so the Sn/Pb alloy used usually in the IC industry cannot be selected for the bump material of FPAs. Because Indium (In) has good ductility at high, low, and ambient temperatures, and the large-scale array of tiny In bumps are prepared at ambient temperature, indium becomes the most ideal solder material for FPAs [61–63].

The flip chip technique of IRFPAs usually includes cold welding and reflowing self-alignment because the reflowing pulling welding technique is difficult to realize, its practical application has not been reported. These three welding techniques are introduced in detail as follows [64–67].

(1) Flip chip cold welding [65]

As a conventional welding technique applied in IRFPAs, the flip chip cold welding is the mainstream technology adopted by American IRFPA manufacturers. In this technique, the FPA chip with indium bump array and the ROIC with the same indium bump array are aligned using the alignment bonding system firstly and then are welded together by applying a certain amount of pressure. Although the method is simple and practical, there are some disadvantages of large extrusion force, low welding accuracy, uncontrollable weld shape and height, poor spacing uniformity, and some shortcomings in the electrical and mechanical connectivity.

The top area of the micro-mesa in a high density array is limited. Especially for the two-color and multicolor IRFPAs, so it is very difficult to prepare two indium bumps on the micro-mesa of limited area and ensure that two indium bumps on the same mesa will not be extruded to touch each other after the hybridized interconnection.

(2) Reflowing self-aligning welding [65, 66]

The detector chip and the ROIC are aligned first and welded together by applying a certain pressure, and then a reducing agent is introduced into the meltback furnace to cause the reflowing of the indium welds. Finally, the reflowing self-aligning welding of the detector chip and the ROIC is achieved under the action of the surface tension between the reflowing indium welds and the contact pads. The requirement of this method to the process is low, and it can be achieved in the most welding techniques. However, the accuracy control of the indium weld shape and height cannot be realized with this method.

A good electrical connection between the detector chip and the Si-based ROIC can be achieved under the proper interconnecting pressure and temperature, and the interconnected indium bumps can endure the temperature shock of the detector operating. Additionally, the reflowing can lead the melted indium and the metal electrode to form some alloys, thereby making the interconnection between the ROIC and the detector more firm.

A thrust damage experiment is performed on the IRFPA analog devices interconnected by reflowing self-aligning flip chip technique, and the contact between indium balls and welds is analyzed. Figure 5.65 provides the SEM images of the surface of a chip without indium bumps after splitting the chip and the ROIC. As shown in this figure, the indium traces are distributed on the contact pads of the surface without indium welds, and most of sections are neither on the contact pads nor on the under bump metallurgy (UBM) sites of indium welds, but they appear within the indium welds. It indicates that the indium balls have good wettability to the contact pads to ensure the electrical and mechanical connectivity of indium bump interconnection and the better reliability [6, 59] (Fig. 5.62).

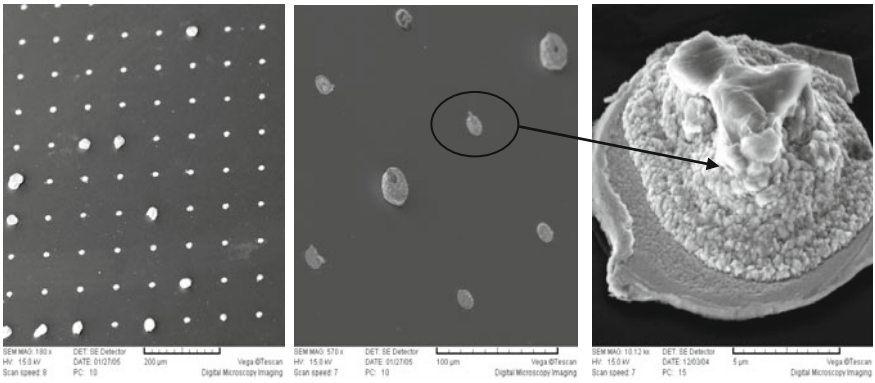


Fig. 5.62 SEM images of the surface of a chip without indium bumps after splitting the chip and the ROIC

(3) Reflowing pulling welding [68, 69]

The detector chip and the ROIC are aligned, and the shape and height of the melted indium balls can be controlled by regulating precisely the distance between the chip and the ROIC. Finally, the reflowing pulling welding is achieved after lowering the temperature. For the height h of indium bumps in the reflowing pulling welding, there are three state control parameters: the height of connecting h_0, the height of extruding h_1, and the expected height after welding h_2. Accordingly, if it is to be realized that $h_2 > h_1$, the pulling operation is necessary, so it is called reflowing pulling. In theory, it is feasible and helpful for improving the reliability of FPAs and the electrical and mechanical connectivity, but there are many difficulties such as the control of technological parameters, and the fabrication of samples.

2. Evaluation of flip chip welding

The flip chip welding for FPAs not only provides the electrical and mechanical connectivity between the pixel of the detector chip and the input terminal of the corresponding ROIC, but also relieves the stress caused by the mismatching of the coefficients of thermal expansion between the detector chip and the Si-based ROIC after the thermal cycling. Therefore, the flip chip welding technique can be estimated in the aspects of welding resistance, mechanical connectivity, weld height (interplanar spacing), and reliability [60, 63, 65, 68, 69].

(1) Welding resistance

The welding resistance is the series resistance between the chip and the ROIC, and it includes the contact resistance between the indium weld and the pixel, the self-resistance of indium weld and the contact resistance between the indium weld and the ROIC. The welding resistance is one of the critical parameters estimating the performance of welding.

A high contact resistance can influence directly the coupling between the pixels and the ROIC, and the influence becomes more notable in the LWIR and MWIR FPAs; moreover, the uniformity of the welding resistance can also influence the uniformity of the FPA performance.

However, due to the active properties of indium, a compact oxide dielectric layer could be formed on the surface of the prepared indium ball. Thus, in the flip chip reflowing welding, the key to achieve the low and uniform welding resistance is how to crack the oxide layer for the high wettability of the indium balls to the contact pads and how to prevent the further oxidization of the indium bumps.

(2) Mechanical connectivity

Mechanical connectivity is the premise of the electrical connection of the flip chip. Its magnitude is a mechanical parameter representing the welding firmness of the chip and the ROIC.

For the flip chip reflowing welding technique, the main issues influencing the mechanical strength of welding include the wettability of the indium weld to the contact pads, the contact area between the indium bump and the pad (related to the

alignment), and the wettability of the indium bump and its self-metal pedestal (UBM).

Accordingly, for the flip chip reflowing welding technique, by increasing the wettability of the indium bump to the pad and the UBM and improving the accuracy of welding alignment, the high mechanical strength can be achieved and the capability of enduring the thermal cycling impact and the mechanical vibration of FPAs can be improved.

(3) Weld height (interplanar spacing)

The weld height is actually the interplanar spacing between the detector chip and the ROIC, as another significant parameter of FPAs.

After undergoing hundreds to thousands of thermal cycles or rapid thermal cycling, the fatigue damage of the indium weld is caused by the mismatching of the coefficients of thermal expansion between the detector chip and the ROIC, until the mechanical and electrical connectivity between them cannot be guaranteed, thereby leading to the failure of IRFPAs. The number of times of the thermal cycling that the indium weld could endure is proportional to the square of the height of the indium bump after flip chip welding, and the relational expression is given [6, 62, 68]:

$$N_{\mathrm{f}} \approx \frac{1}{2} \cdot \left(\frac{\Delta\gamma}{0.65}\right)^{-2} = \frac{1}{2} \cdot (0.65h/L\Delta\alpha\Delta T)^2, \tag{5.8}$$

where $\Delta\gamma$ is the tangential displacement of the indium weld during the thermal cycling, N_{f} is the times of thermal cycling, h is the height of the indium bump, L is the distance between the indium weld of the chip and the center of the FPA, $\Delta\alpha$ is the mismatching of the coefficients of thermal expansion, and ΔT is the temperature range of thermal cycling. In addition to the variable parameters in the expression, L, $\Delta\alpha$ and ΔT are constant, so the times of the thermal cycling that the welded chip can endure can be only increased by improving the height h of the weld.

Therefore, the height of the weld is another significant technical parameter in the estimation of the flip chip reflowing welding technique.

(4) Reliability

The optimization of the mechanical strength and the weld height is also required to solve the problem of the reliability of the flip chip reflowing welding. Thus, the aim of the reliability optimization is to improve the welding technique in the aspects of the sample fabrication and the technical parameter controls of welding, so that the main issues resulting in the failures of the FPAs (the failures of the electrical and mechanical connectivity between the chip and the ROIC) are minimized.

3. Preparation of indium ball array

In the reflowing pulling welding or the reflowing self-aligning welding, the preparation of the indium ball array is a key first step. A fluxing agent is required in

the previous preparation technique of the indium ball array, but the conventional fluxing agent has a corrosion effect on the indium bumps and the other parts of the FPAs, thereby influencing significantly the storage and lifetime of the FPAs. Accordingly, the preparation techniques of the indium ball array without fluxing agent are introduced as follows:

(1) Meltback technique of indium ball without fluxing agent [6, 62, 68]

Figure 5.63 shows the SEM morphology of the indium bump array. The indium bump array is achieved by the processes such as the photolithography of the indium bump hole, the thermal evaporation growth of indium film and liftoff of photoresist. In the indium bump array prepared by thermal evaporation, every indium bump is formed by the accumulation of indium grains with the grain boundaries existing between grains.

The existence of these grain boundaries results in that the indium bump cannot be melted into a ball at the melting temperature or higher temperature. Therefore, in order to achieve the transformation from indium bump to indium ball, these grain boundaries between indium grains must be eliminated apart from satisfying the temperature requirement.

Hydrogen gas is a good reductant, but it presents reducing properties only at high temperatures, which is disadvantageous for the connection between the indium weld and its UBM pedestal and the performance of the ROIC.

However, as a weak reductant, formic acid (HCOOH) can eliminate the grain boundaries during the meltback treatment of the indium bump and make the indium

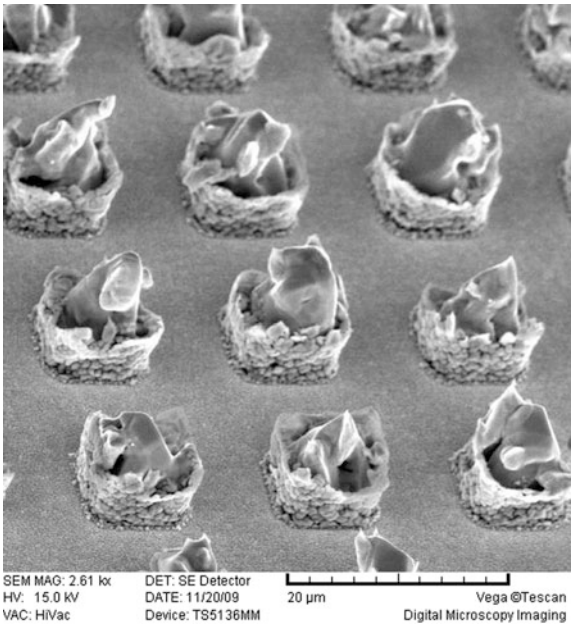


Fig. 5.63 SEM morphology of the indium bump array

grains join together at the melting temperature. This is because the formic acid can react with the oxide (In_2O_3) at the indium grain boundaries ($In_2O_3 + 6HCOOH = 2In(HCOO)_3 + 3H_2O$) and the products of $In(HCOO)_3$ and H_2O can be expelled with the gas flow containing formic acid.

When the temperature is rising in the meltback process, the oxides at the grain boundaries are eliminated continuously with the reaction between the formic acid and the oxide at the grain boundaries. It enhances the cohesion between indium grains and makes the indium bump melted into a ball at the melting point of indium due to the action of the surface tension.

Therefore, in the preparation of indium ball, formic acid is selected as the reductant of In_2O_3 for the melting of indium bump at lower temperature.

(2) Preparation of indium ball array and estimation of height h

In the preparation of an indium ball array, the indium bump array is placed into the meltback furnace with a certain gas and the accurate control of temperature and time, and formic acid is used as a reductant. Finally, the indium ball array can be achieved after high temperature treatment. A mechanical vacuum pump should be adopted in the equipment to avoid the corrosion effect of the formic acid gas flow on the vacuum system. During meltback, the nitrogen gas flow containing the formic acid is introduced into the chamber continuously; meanwhile the sample is heated rapidly to the melting point of indium and maintained for several minutes to make sure that all the indium bumps are melted into balls. Then, the large flow of nitrogen gas is introduced rapidly to expel the residual formic acid and the products in the chamber and rapidly lower the sample temperature.

Figure 5.64 provides the SEM morphologies of the indium ball array achieved by the meltback treatment. As shown in this figure, the height of indium ball is around 16.5 μm, and the diameter is about 19.5 μm.

The height h of the indium ball depends on the radius r_1 and the height h_1 of the indium bump before meltback treatment, as well as the radius r_2 of the UBM pedestal. Because the amount of indium is constant before and after the meltback treatment, that is, the following equation is satisfied [6, 59]:

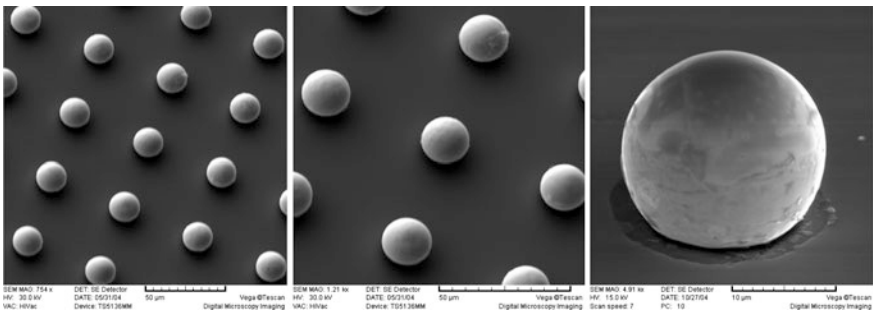


Fig. 5.64 SEM morphologies of the indium ball array achieved by the meltback treatment

$$V = \frac{\pi}{6} \cdot h \cdot \left(3 \cdot r_2^2 + h^2\right) = \pi \cdot r_1^2 \cdot h_1 \tag{5.9}$$

From Eq. (5.9), by increasing the radius r_1 and the height h_1 of the indium bump and reducing the radius r_2 of the UBM pedestal, the height of the indium ball can be raised up.

5.3 Two-Color Micro-Mesa Detector Chip

5.3.1 Selection of Two-Color Detector Chip Architecture

1. Structure type of two-color detector chip

As previously mentioned in this chapter, the FPA is an advanced imaging sensor with the integration of acquisition and processing of an IR signal. The integrated two-color IRFPA is comprised of the detector chip responding to two bands of IR radiation separately and the compatible Si-based ROIC. The detector architectures of two bands are integrated in the longitudinal direction.

According to the literature [70], there are six structure types existing in the back-illuminated integrated two-color IRFPAs (seen in Chap. 2). Their structural features are summarized in Table 5.5. From the numbers of extracted electrodes, they are divided into unipolar and bipolar. In the unipolar two-color IRFPAs, two back-to-back photodiodes are integrated longitudinally. The structure of this FPA is simple relatively. However, the operating band is selected and the photoelectrical properties on each band are optimized by applying bias during operation, and the operations on the two bands are asynchronous sequentially. For the bipolar two-color FPAs, either two photodiodes in the same direction or two photodiodes with opposite directions are integrated longitudinally. The architecture of the bipolar two-color IRFPA is obviously complicated. However, the photoelectrical properties on two bands can be optimized and equilibrated, respectively, and the temporal synchronization can be achieved in the bipolar two-color IRFPAs.

From the direction of two longitudinal photodiodes, the photodiode architecture can be divided into equidirectional diodes and back-to-back diodes. For two-color IRFPAs with equidirectional diodes, two photodiodes in the same direction are integrated longitudinally only based on the bipolar structure. Nevertheless, for two-color IRFPAs with back-to-back diodes, two photodiodes in opposite directions are integrated longitudinally in the back-to-back manner, using the bipolar or unipolar structure.

2. Features of two-color detector chips with different structures

(a) n-p-n type two-color HgCdTe IR detector with unipolar mode back-to-back diodes

This FPA is an n-P^+-N type sequential two-color detector based on MBE technology presented by the former HRL laboratory [29, 33], and the pixel is

Table 5.5 Structural features of back-illuminated integrated two-color IRFPAs

Type	Manufacturer	Wave band	Scale	Pixel size (μm)	Material	Pixel electrode	Polar direction	Operating mode
(a)	RVS	MW/LW	640 × 480 1280 × 720	20	n-P-N type HgCdTe film grown by MBE	Unipolar	Back-to-back	Sequential
(b)	RVS	MW/LW	256 × 256	40	n-P-N type HgCdTe film grown by MBE	Bipolar	Back-to-back	Simultaneous
(c)	BAE, SELEX	MW/LW	320 × 256 640 × 512	24	p-n-N-N-P type HgCdTe film grown by MOVPE	Bipolar	Back-to-back	Simultaneous
						Unipolar	Back-to-back	Sequential
(d)	LETI and Sofradir	MW/MW MW/LW	256 × 256 640 × 512	20	p-P-P-N type HgCdTe film grown by MBE	Bipolar	Back-to-back	Simultaneous
(e)	Rockwell	MW/LW	128 × 128	40	n-N-N type HgCdTe film grown by MBE	Bipolar	Equidirectional	Simultaneous
(f)	LETI and Sofradir	MW/LW	640 × 512	20	p-P-P type HgCdTe film grown by MBE	Bipolar	Equidirectional	Simultaneous

actually comprised of two back-to-back photodiodes. Because each pixel has only one indium bump, a bias is applied to select the operating photodiode. When the bias >0, the MW photodiode is in the reverse bias and the SW photodiode is in the front bias, the loop current is mainly determined by the MW photoelectric current and the MW photodiode works; when the bias <0, it is opposite. This structure presents high spatial reading characteristics and the duty cycle close to one hundred percent and makes the pixel smaller than 40 μm feasible.

(b) n-p-n type two-color HgCdTe IR detector with bipolar back-to-back diodes

This FPA is an n-P^+-N type simultaneous two-color detector based on MBE technology presented by the former HRL laboratory, and it can overcome the disadvantages of the sequential detector that two photodiodes cannot work in the optimal operating bias simultaneously and the MW spectral cross talk becomes remarkable. However, each pixel requests two contact electrodes to be welded with the indium bump and some reading structures to realize the current isolation and the simultaneous reading. Thus, the detector chip technique and the design of ROIC present a significant challenge.

(c) n-p-P-P-N type two-color HgCdTe IR detector with unipolar and bipolar back-to-back diodes

This FPA is an n-p-P-P-N type two-color heterojunction IR detector with unipolar and bipolar diodes produced by Leti/LIR in France. In this FPA, the MW photodiode is a planar implanted junction, while the SW photodiode is the in situ mesa junction. This structure has high quantum efficiency, small spectral cross talk, and high duty cycle.

(d) p-n-N-N-P type two-color HgCdTe IR detector with unipolar and bipolar back-to-back diodes

This FPA is a p-n-N-N-P type two-color simultaneous IR detector based on the MOCVD technique presented by Lockheed Martin. A component barrier of n-N-N structure is introduced between two photodiodes to block the interdiffusion of minority carrier holes on each band thereby reducing the MW spectral cross talk.

(e) Two-color HgCdTe IR detector with bipolar equidirectional diodes

This FPA is a Simultaneous Unipolar Multispectral Integrated Technology (SUMIT) type simultaneous two-color IR detector. For each pixel, the As^+ ion implantation is performed twice on the n-N-N type heterojunction material to achieve two longitudinally integrated p^+-on-n planar diodes. Because the diodes on two bands are equidirectional, their photoelectric signals are not required to be isolated. As long as each reading unit contains two individual single band reading

structures, the simultaneous operating mode can be achieved. Therefore, its ROIC structure is relatively simple, and the two-color FPA has better device performance.

(f) Two-color HgCdTe IR detector with two bipolar equidirectional monochromatic chips integrated longitudinally

The normal incidence structure (SW detector on top and MW detector below) is applied in this two-color FPA produced by DRS Corporation. The device includes two longitudinally integrated HgCdTe monochromatic FPA chips adhered onto the two-color Si ROIC. The SW diode is connected to the Si-based ROIC through the isolated paths in the MW diode. Both of the monochromatic detector chips are fabricated based on the p-type HgCdTe material grown on CdZnTe substrate using the LPE process. The devices are fabricated through ion implantation via hole etching and climbing metallization. They have a low generation–recombination current. Since the two band photodiodes are equidirectional, its ROIC is similar with that of SUMIT-type two-color detector presented by Rockwell and is less complex than the other back-illuminated ROIC. Accordingly, this device has good performance.

3. Selection of two-color detector preliminary chip architecture

Based on the theoretical calculation of the performance of two-color detectors with different structures in Sect. 2.2 and the feasibility of current technological realization, the back-to-back bipolar mode is selected as the two-color detector preliminary chip, and the configuration of the pixel and the common port is similar to an IR triode structure.

The advantages and disadvantages of this two-color FPA structure include simultaneous operating mode on SW and MW bands because there are two signal outputs within each pixel. Moreover, according to the theoretical calculation, this bipolar back-to-back two-color detector has higher quantum efficiency than other designs. However, there are certain technical challenges on the preparation of the SW and MW output electrodes and the indium bump on the same micro-mesa. On the other hand, in the triode structure of each pixel, the individual MW photoelectric signal is transported to the MW port, while the superposition of MW and SW signals is transported to the SW port. Thus, in the design of the ROIC, the dual-input reading design for two signals within the individual pixel and the accuracy separation technique of MW and SW mixed signals must be determined.

Although there are remarkable challenges in the aspects of two-color material growth, chip processing, and two-color ROIC for the bipolar back-to-back two-color IR FPA, this device has advantages of high quantum efficiencies on two bands and simultaneous operation. Therefore, the bipolar back-to-back two-color IR detector chip is selected as a starting point for the investigation on the fabrication of two-color detector chip.

5.3.2 *Fabrication of Two-Color HgCdTe Micro-Mesa Preliminary Detector*

1. Fabrication of two-color detector preliminary chip

Figure 5.65 provides the cross-section view and top view of the fabricated SW/MW integrated HgCdTe two-color detector chip. In the fabrication of the n-p-P-N type $Hg_{1-x}Cd_xTe$ two-color detector preliminary chip, the n-P-N type multilayer heterojunction material is grown by MBE, and the chip processing techniques such as selective B^+ ion implantation, mesa etching formation, micro-mesa photolithography, sidewall passivation, and climbing metallization are carried out to fabricate the device [6, 12, 13, 49, 50].

For the fabrication of $Hg_{1-x}Cd_xTe$ two-color detector chip, the p-P-N type multilayer heterojunction material is prepared on the GaAs substrate by MBE and in situ doping. As shown in Fig. 5.65, the P-on-N homojunction responding to SW radiation is grown first, and the Cd components of two layers are 0.4. The thickness of the N-type material for the common port is 2 μm and the doping concentration of in situ indium is 5×10^{17} cm^{-3},while the thickness of the p-type material for the

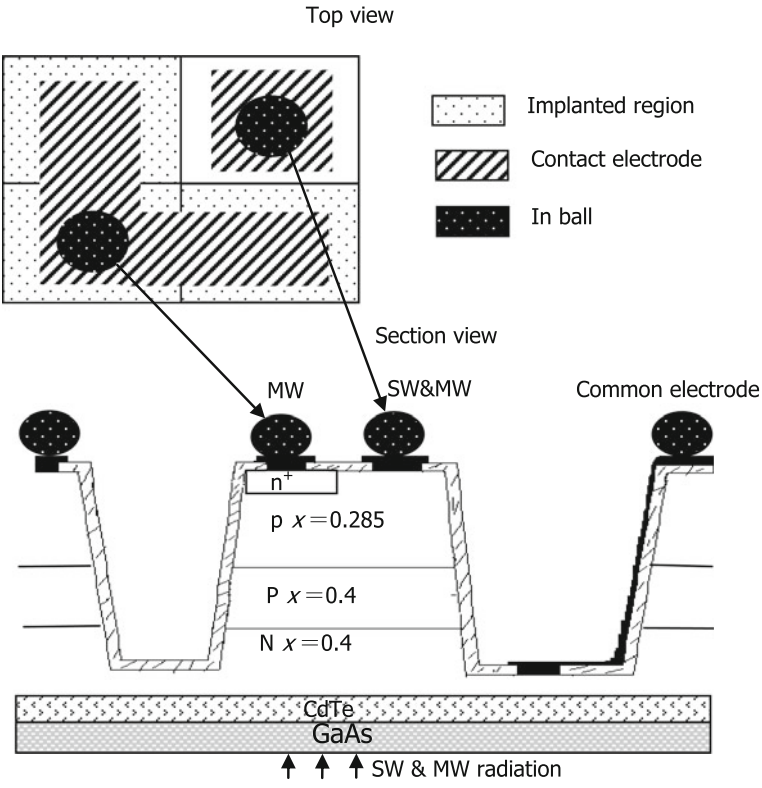


Fig. 5.65 Pixel structure of integrated SW/MW HgCdTe two-color detector chip

absorption layer for SW radiation is 6 μm; then, the 8 μm thick p-type layer responding to MW radiation of x = 0.285 is grown. In the prepared p-P-N type multilayer heterojunction material, the N-type layer is doped with in situ indium, while both of the p-type layers are doped with Hg vacancies obtained from the heat treatment after MBE growth.

As shown in Fig. 5.65, the photodiode array responding to MW radiation is fabricated through the surface treatment and selective B^+ ion implantation (in the region of 'L' shape with the 3/4 of the square area of a pixel) performed on the p-P-N type multilayer heterojunction material, while the photodiode responding to SW radiation is formed during the in situ doping of the growth process. The isolation in the two-color micro-mesa array is achieved by the ICP RIE and a wet etching with automatic rotating. The contact electrodes for the MW photoelectric current signal are prepared on the B^+ implanted n^+ regions within the square micro-mesa pixels, and the contact electrodes for SW photoelectric current signal (the common electrodes of MW and SW photodiodes) are prepared on the nonimplanted *p* region on the top of pixel. The common electrode of the entire device is extracted from the *n* layer nearest to the substrate through the climbing metallization. The micro-mesa pixel is the longitudinal back-to-back integration of the SW mesa photodiode and the MW planar photodiode, and it can respond to the IR radiation on SW and MW bands simultaneously. Specifically, the individual optimal operating bias can be selected independently to output photoelectric current signals.

Figure 5.66 provides the SEM morphology of a HgCdTe two-color IR detector chip. As shown in this figure, each pixel has two indium bumps, which output the photoelectric current signals of the SW and MW photodiodes, respectively, especially for bipolar IR FPAs. Because the area of the micro-mesa top is limited, the diameters of the indium bumps must be very small to ensure that two indium bump cannot be squeezed to touch each other after hybridized integration.

The individual two-color micro-mesa pixels are arranged periodically in a one-dimensional or two-dimensional array on the planar space to achieve the two-color HgCdTe detector chip. After the chip is welded with the corresponding ROIC by flip chip welding, the linear array or large format two-color FPAs can be achieved. Here, the two-color FPAs of 256 × 1 linear array and 128 × 128 area array are fabricated.

2. Test and estimation of HgCdTe two-color micro-mesa detector

Figure 5.67 shows the *I–V* and *R–V* characteristic curves of two longitudinal photodiodes of a SW/MW HgCdTe two-color detector chip. As shown in Fig. 5.67, the *I–V* curves of the SW and MW photodiodes in each pixel have good repeatability at $T = 78$ K, and the corresponding mean R_0A values are 3.852×10^5 and $3.015 \times 10^2\ \Omega\ \mathrm{cm}^{-2}$, respectively. The *I–V* curves present obvious photodiode characteristics, which indicates that the material and the chip architecture design are reasonable, and the chip processing technologies such as the growth of multilayer heterojunction material, the isolation of micro-mesa array, the sidewall passivation, the climbing metallization, and the preparation of dual indium bumps with a pixel are feasible.

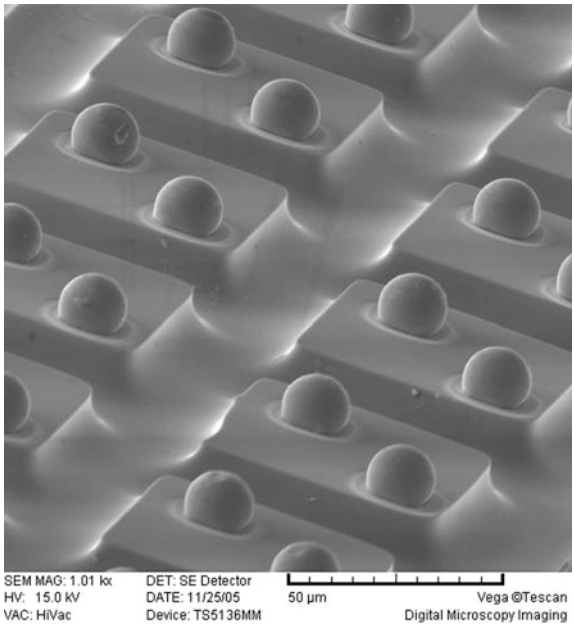


Fig. 5.66 SEM morphology of HgCdTe two-color IR detector chip of 256 × 1 linear array

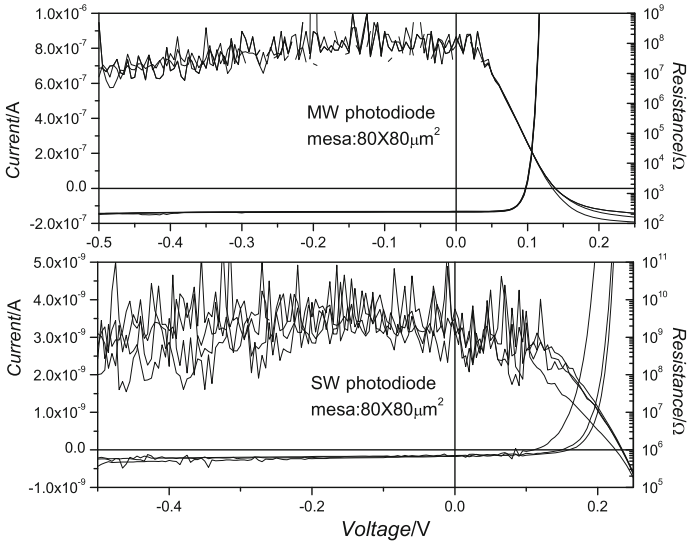


Fig. 5.67 *I–V* and *R–V* characteristics of SW and MW photodiodes in the two-color detector

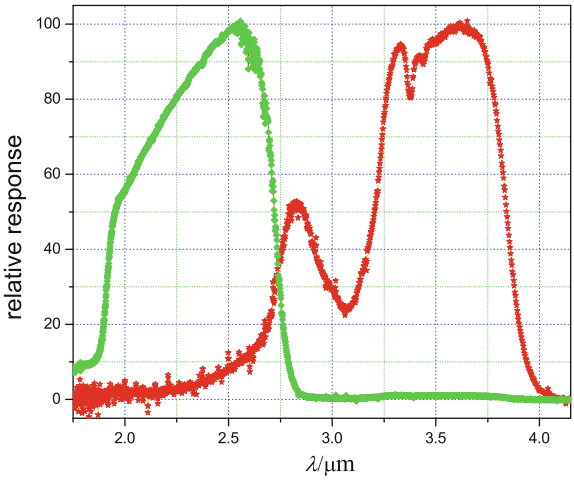


Fig. 5.68 Spectral response of the two-color FPA

The photoelectric current signals and dark current noise of the SW and MW photodiodes in a HgCdTe two-color detector chip are measured at liquid nitrogen temperature under 800 K blackbody radiation. The peak wavelength detectivity of the SW and MW photodiodes is calculated to be 9.3×10^{10} and 5.1×10^{10} cm $Hz^{1/2}/W$, and their standard deviations are 25.30 and 18.85 %, respectively. These properties approach the photodiode properties of HgCdTe monochromatic FPAs on the corresponding band.

Figure 5.68 provides the spectral response of the two-color FPA. From the figure, it is seen that the cutoff wavelengths of two band photodiodes are 2.70 and 3.8 μm, respectively. The cutoff wavelength of the SW photodiode is approximately equal to the cut-on wavelength of the MW photodiode (the influence of the atmospheric window absorption on the spectral response test of the MW photodiode, that is, the hollow of the MW spectral response at the wavelength of 3.0 μm is ignored). According to the shape of the spectral response curve, the MW spectral cross talk can be calculated to be 5 %, and the spectral cross talk of the SW photodiode can be basically ignored. The SWIR spectral response peak of the two-color FPA using the SWIR in situ heterojunction becomes wider indeed, and the n layer plays the window role of filtering the shorter wavelength part of the IR radiation [6, 12, 13].

5.3.3 *Simultaneous 128 × 128 Two-Color HgCdTe IRFPAs Detector*

High temporal and spatial coherent simultaneous long-wavelength/mid-wavelength (LW/MW) two-color FPA infrared detection is the cutting-edge technique for third-generation infrared remote sensing. HgCdTe LW/MW two-color infrared

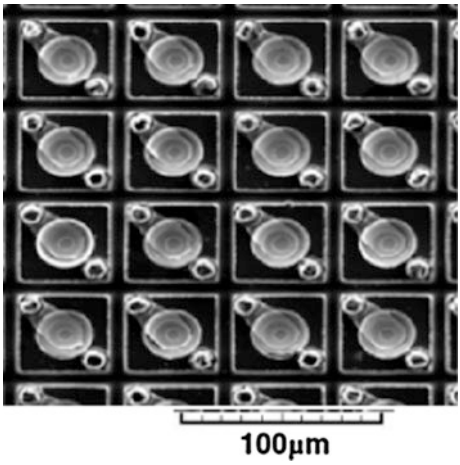


Fig. 5.69 SEM image of the 128 × 128 LW/MW HgCdTe two-color infrared FPA detector

detectors were designed and fabricated. The top long-wavelength and bottom mid-wavelength infrared planar photodiodes were processed by selective B^+-implantation after etching the long-wavelength epilayer into a curvature and exposing the mid-wavelength layers for the implantation of the n region of the MW photodiode by a micro-mesa array technique (shown in Fig. 5.69). A 128 × 128 MW/LW HgCdTe infrared FPA detector is fabricated photolithographically by simultaneous nonplanar B^+-implantation of the LW and MW photodiodes, passivation and metallization of the sidewalls, mesa isolation, and flip chip hybridization with a readout integrated circuit. The inner mechanisms for suppressing the cross talk and improving photoresponse have been carried out by combining experimental work with numerical simulations [14].

As shown in Fig. 5.70, it can be observed that the cutoff wavelength of the MW diode is close to the turn-on wavelength of the LW diode, in good agreement with the proposed structure and parameters. This effectively enables the MW diode to not only absorb and respond to the MW radiation but also acts as a window for any LW radiation.

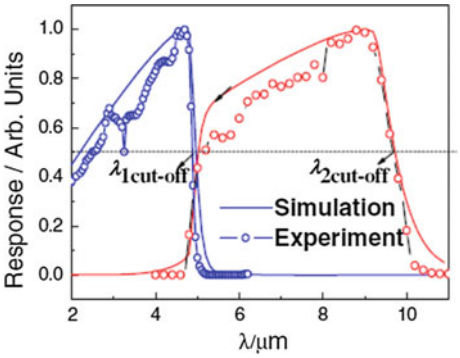


Fig. 5.70 Experimental (*dotted*) and simulated (*solid*) spectral photoresponse of the LW/MW HgCdTe two-color infrared detector with cutoff wavelengths of λ_1 cutoff 4.8 μm and λ_2 cutoff 9.7 μm for the LW and MW diodes, respectively. A voltage of 0.01 V is used in the simulation and experiment

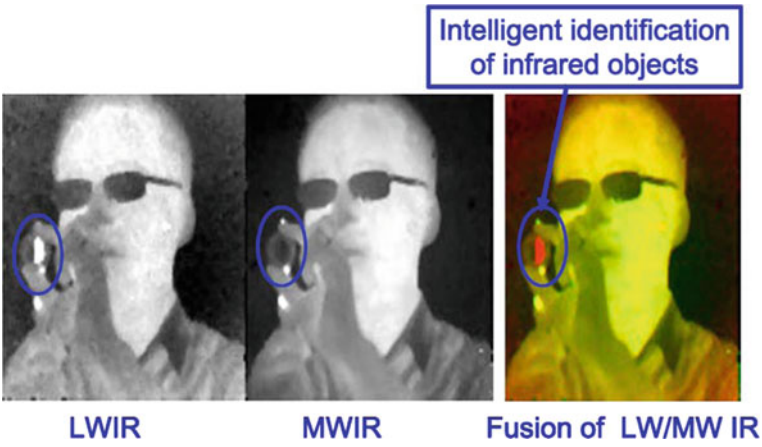


Fig. 5.71 Infrared image from the 128 is very promising two-color infrared FPA detector at 77 K. A MW optical filter is used to stop MW infrared radiation. The heated electric soldering iron behind the filter is invisible for the MWIR detector (*middle*) while visible for the LWIR detector (*left*)

The simulated spectral photoresponse is also compared to that of experimental results showing that the simulation and experiment are self-consistently in good agreement. The calculated peak detectivity is 3.2×10^{11} and 4.3×10^{10} cm $Hz^{1/2}$ for the MW and LW photodiodes, respectively. From Eqs. (5.1) and (5.2), LW-to-MW cross talk of 1.25 % and MW-to-LW cross talk of 0.7 % are obtained.

Figure 5.71 is the infrared image from the 128 × 128 LW/MW HgCdTe two-color infrared FPA detector. It is shown that the proposed two-color HgCdTe infrared FPA detector is very promising for the intelligent identification of infrared objects.

In summary, a 128 × 128 HgCdTe LW/MW two-color infrared FPA detector is fabricated with the cutoff wavelengths of 4.8 and 9.7 μm, and peak detectivity of 3.2×10^{11} and 4.3×10^{10} cm $Hz^{1/2}$, respectively. Ultralow LW-to-MW cross talk of 1.25 % and MW-to-LW cross talk of 0.7 % are obtained showing that the LW/MW two-color HgCdTe infrared FPA detector is very promising for third-generation intelligent infrared imaging.

5.4 Si-Based HgCdTe Processing Technology

The development of modern infrared system requires higher detectivity and recognition capability to rapidly detect, recognize, and monitor the dynamic change on the battlefield for the effective treatment of emergencies. In order to satisfy the military requirement, the third-generation IRFPA technologies acquire a great development and application. The Si-based HgCdTe FPA technique is one of these

technologies. In order to satisfy the high resolution and large area monitoring capability of IR systems, the high-performance large-scale and very-large-scale HgCdTe FPAs are indispensable. With the continuous expansion of array size, the problems on the size limit and the cost of the conventional substrate CdZnTe become increasingly prominent. The photovoltaic detector array prepared based on the HgCdTe material on Si substrate attracts great attention. The Si substrate has the advantages of large size, high mechanical strength, low cost, and good thermal matching with ROIC, which make the Si-based HgCdTe an effective way to break the limit of the CdZnTe substrate and realize the very-large-scale multicolor IR imaging detection [71, 72].

In this section, the new technologies of Si-based HgCdTe IR FPAs are introduced. Due to large lattice and thermal mismatch existing between Si and HgCdTe, the Si-based HgCdTe FPAs have different features from the conventional substrate devices in the aspects of thermal stress distribution, chip architecture, and chip preparation. In Sect. 5.4.1, the Si-based HgCdTe material and the stress distribution of the device are analyzed, and the chip architecture is optimized. In Sect. 5.4.2, the low damage processing technology of 3-in. Si-based HgCdTe chips is introduced, and the performance of the chip is analyzed.

5.4.1 *Stress Analysis and Structure Design of Si-Based HgCdTe*

1. Stress distribution analysis of Si-based HgCdTe FPAs

Selecting silicon as the epitaxial substrate of HgCdTe material has many technical difficulties. First, the nonpolar silicon surface could result in the polarity problems such as antiphase domain, twin, and surface crystals. In Sects. 3.2 and 3.3 , the surface polarity control technique and theoretical model of silicon surface have been shown. It is considered that As passivation is an effective way to acquire B surface polarity control and avoid the generation of an antiphase domain.

Second, the lattice mismatch between CdTe and Si is 19.3 % (Table 5.6 shows the comparison of lattice constants and coefficients of thermal expansion of some

Table 5.6 Comparison of lattice constants and coefficients of thermal expansion of some materials [76, 77]

Material	$Zn_{0.05}Cd_{0.95}Te$	GaAs	Si	CdTe	$Hg_{0.7}Cd_{0.3}Te$
Lattice constant	6.465	5.6537	5.4309	6.4827	6.4663
coefficient of thermal expansion (K^{-1})	5.0×10^{-6}	5.7×10^{-6}	2.5×10^{-6}	4.96×10^{-6}	4×10^{-6}

materials), and, therefore, could introduce high density perforation dislocations (they become carrier scattering centers and recombination centers in the material) into the active layer, thereby reducing the carrier mobility and carrier lifetime. With the progress of dislocation suppression technique and HgCdTe epitaxy technique on the Si-based CdTe composite substrate, the dislocation density in the Si-based HgCdTe is effectively suppressed. However, in comparison with the HgCdTe material grown on the CdZnTe substrate, the dislocation density is 1–2 orders of magnitude higher ($10^6/cm^2$). Therefore, the improvement of the device performance is affected to some extent.

Third, the difference of coefficients of thermal expansion between Si and CdTe/HgCdTe is very large, 2.5×10^{-6} and 4.96×10^{-6} K^{-1},respectively. When the material is cooled from the epitaxy temperature (near 200 °C) to the ambient temperature (~25 °C), and then from the ambient temperature to operating temperature (around −193 °C), the huge thermal mismatch between Si substrate and HgCdTe epitaxial film introduces a large internal stress in the material. Repeated temperature cycles could lead to increased material defects, thereby resulting in the degradation of device performance.

Thermal stress is one of the significant issues affecting the performance and uniformity of IRFPAs and also one of the critical difficulties in the modularization of FPAs. The stress distribution of Si-based HgCdTe FPAs is investigated based on finite element analysis method [73], and the chip architecture and FPA structure can be optimized according to the analysis results.

The structure of the FPA is given in Fig. 5.72. The stress distribution of the Si-based HgCdTe FPA is calculated by a one-dimensional model and finite element analysis. In the calculation, the indium bump is equivalent to the scale indium slices with the same height and Young's modulus. In the one-dimensional model [74], each layer material is supposed to be infinite, only the thickness is taken into consideration, and it is assumed that the coefficient of thermal expansion (CTE) and Young's modulus of each layer are constant with temperature, as shown in Table 5.7. In the finite element analysis, the change in CTE of HgCdTe and Si with temperature is considered as shown in Figs. 5.73 and 5.74; the CTE of the epoxy glue is 62.6 ppm/K ,and the vitrification temperature is considered to be constant. For Young's modulus of HgCdTe and Si, the anisotropy of the material and its change with temperature are taken into consideration. The Young's modulus of the wire base plate is 335 GPa (the Young's modulus of sapphire plate).

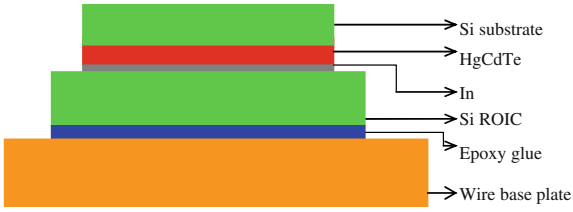


Fig. 5.72 Si-based HgCdTe FPA structure

Table 5.7 Coefficients of thermal expansion and Young's modulus in finite element analysis

Material	CTE (ppm/K)	Young's modulus (GPa)
MCT	5	50
Si	2	80
Sapphire plate	5.5	335
Epoxy glue	62.6	38

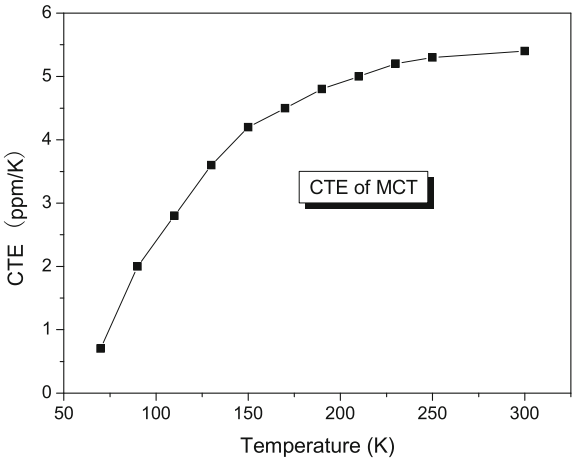


Fig. 5.73 Change in coefficient of thermal expansion of HgCdTe with temperature

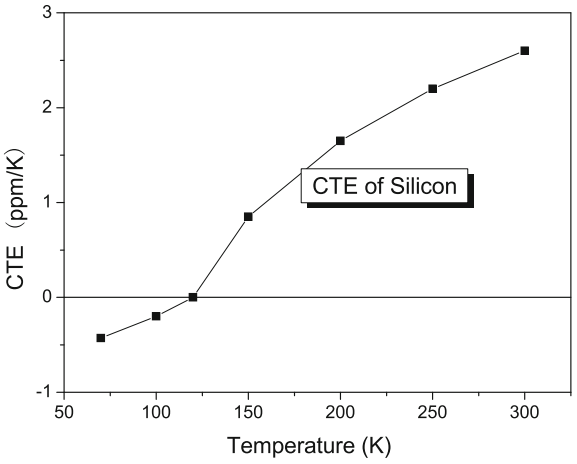


Fig. 5.74 Change in coefficient of thermal expansion of Si with temperature

First, considering the Si-based HgCdTe dual layer structure, the low temperature deformation and internal stress of Si-based HgCdTe film are calculated as shown in Figs. 5.75 and 5.76. The result indicates that when the thickness of Si substrate increases, the concave curvature of HgCdTe decreases, but the internal stress of HgCdTe film increases slightly.

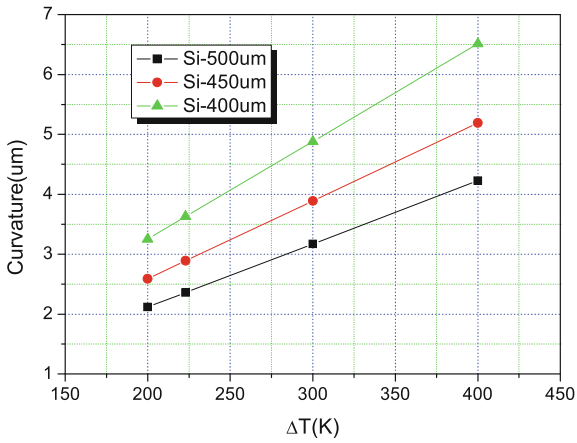


Fig. 5.75 Change in curvature of materials with temperature difference on the Si substrates of different thicknesses

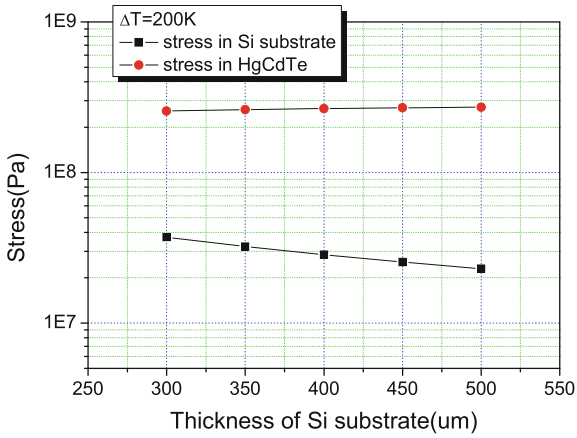


Fig. 5.76 Calculated stress conditions of Si and HgCdTe with Si substrates of different thicknesses when the temperature difference is 200 K (from the ambient temperature to the liquid nitrogen temperature)

For the Si-based HgCdTe FPA structure shown in Fig. 5.72, when the thickness of Si substrate is reduced gradually from 500 to 0 μm, the stress conditions of each layer in FPA is calculated, as shown in Fig. 5.77. Changing the CTE of wire base plate and the Young's modulus of assembly glue, the stress change of each layer is calculated, as shown in Figs. 5.78 and 5.79.

A finite element analysis is performed for verification of the calculation results, and the results are compared as shown in Fig. 5.80.

The analysis results indicate that 100 MPa thermal stress exists in the HgCdTe film. The Si substrate thickness, Young's modulus of assembly glue, and the CTE of wire base plate affect the stress on the HgCdTe film to varying degrees, and the influence of the CTE of the wire base plate on the stress on the HgCdTe film is the most remarkable.

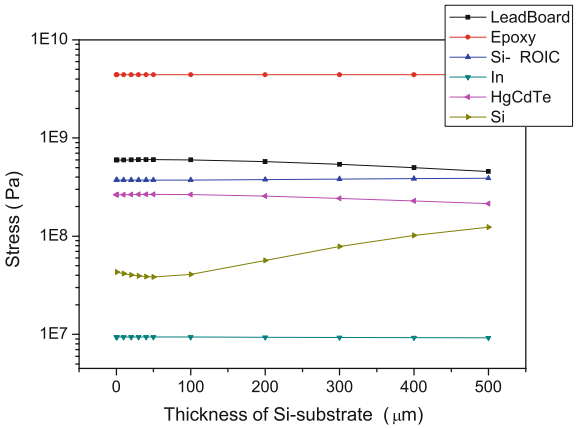


Fig. 5.77 Change in subjected thermal stress of each layer with the thickness of Si substrate

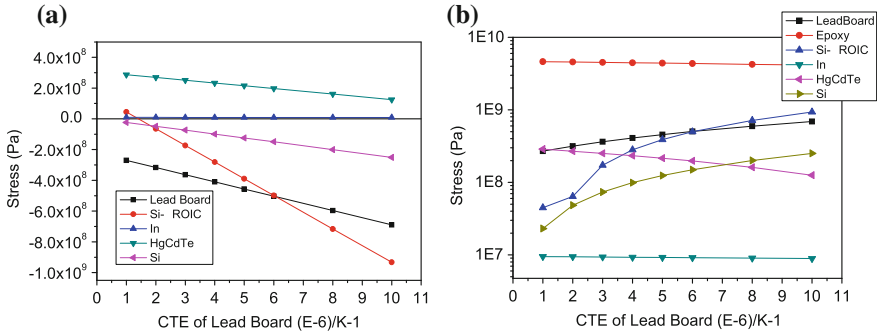


Fig. 5.78 Change in subjected stress of each layer with the CTE of wire base plate. **a** Linear coordinate, **b** logarithmic coordinate

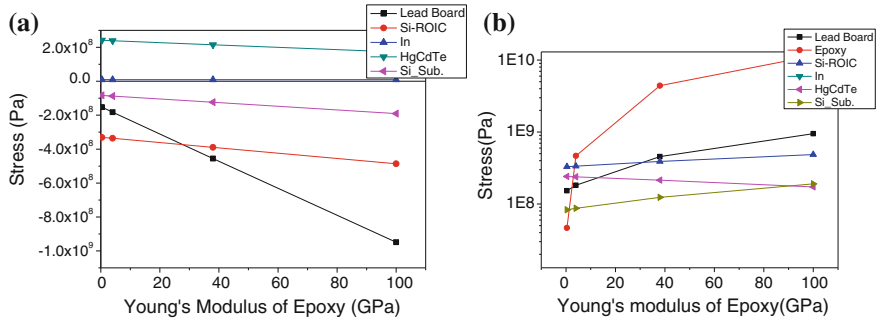


Fig. 5.79 Change in subjected stress of each layer with the Young's modulus of assembly glue. **a** Linear coordinate, **b** logarithmic coordinate

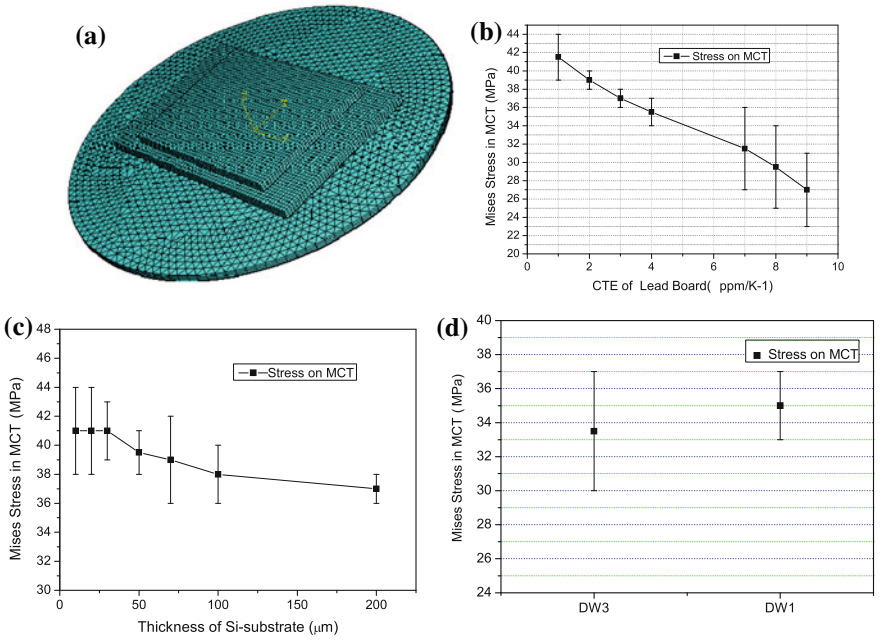


Fig. 5.80 Stress analysis of HgCdTe. **a** Mesh segmentation of finite element analysis for Si-based FPAs, **b** influence of coefficient of thermal expansion of wire base plate on Mises stress of Si-based HgCdTe, **c** influence of Si substrate thickness on Mises stress of Si-based HgCdTe, **d** influence of different assembly glues on Mises stress of Si-based HgCdTe

2. Chip stress state analysis and temperature stress experiment

It can be concluded from the calculation results that, although the selection of structure layer material in FPAs can affect the stress in the HgCdTe film to varying degrees, the location of thermal mismatch maximum is between the Si substrate and HgCdTe epitaxial film. Using finite element analysis, the internal stress distribution of the HgCdTe film on heterogeneous substrate is simulated, and the stress is found to concentrate on the middle position, as shown in Fig. 5.81. The Si-based HgCdTe chip is cooled to the liquid nitrogen temperature, and the low temperature deformation of the chip is measured by the profiler. During measurement, the chip temperature rebounds gradually. For the same position on the chip, the deformation curves of the liquid nitrogen temperature LT, $T1$, $T2$, and the ambient temperature RT (LT < $T1$ < $T2$ < RT) are measured successively, as shown in Fig. 5.82.

3. Effect of Thermal Stress on the Performance of Si-based HgCdTe Diodes and FPA

In the last section, the thermal stress existed in the Si-based HgCdTe materials and FPA was calculated and the distribution was analyzed. In this section, the performance of the HgCdTe/Si material, diodes, and FPA under different thermal stress condition were studied.

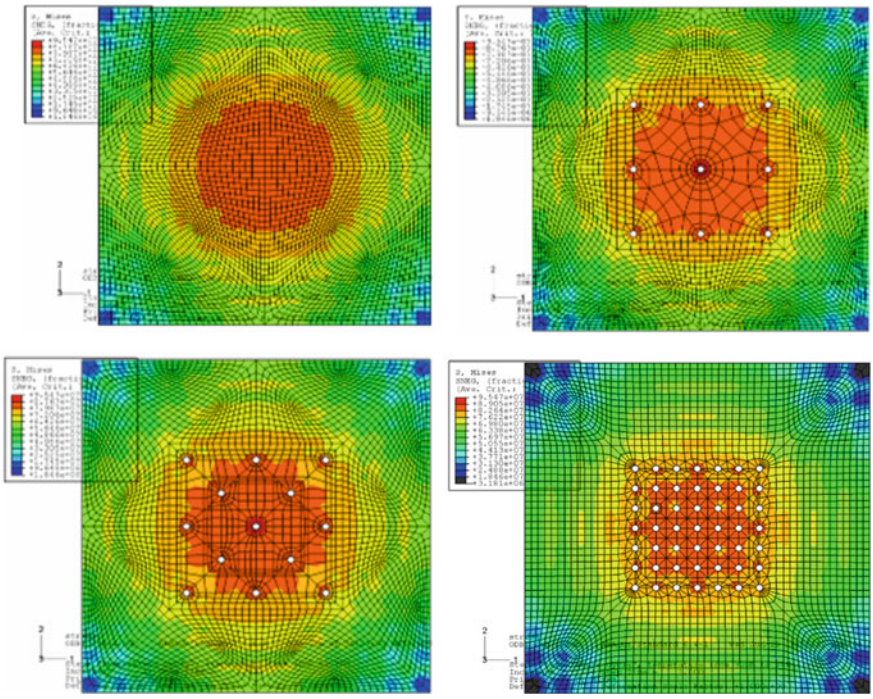


Fig. 5.81 Stress distribution in the Si-based chip

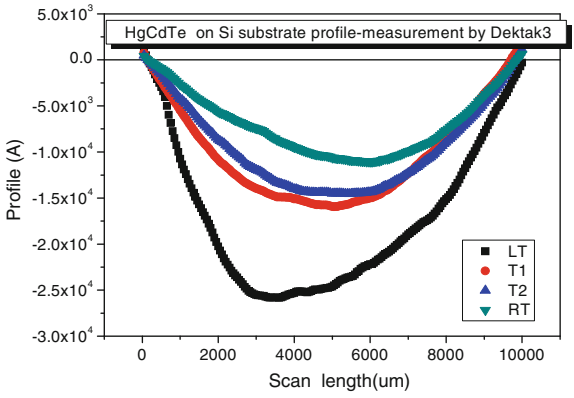


Fig. 5.82 Low temperature deformation of stress chip

In order to change the stress in the HgCdTe film, many materials are tried to tie to the devices whose thermal expansion coefficient is different with the reference material. The experimental measurements of HgCdTe/Si material, diode, and FPA performance are performed utilizing these different structures.

(1) Effect of thermal stress on material properties

For assessing the quality of the Si-based HgCdTe layers, a van de Pauw test structure is processed and the carrier mobility and concentration under different temperatures are estimated before and after a sapphire layer is attached to the back of the Si substrate.

Figure 5.83 shows the temperature dependence of carrier mobility and concentration for the HgCdTe/Si material before and after a sapphire layer applied to the back of the Si substrate. By comparing two different stress structures, it is shown that p-HgCdTe/Si material with sapphire layer tied to Si substrate has higher hole mobility and lower carrier concentration and the effect of stress on material performance is much more obvious at lower temperature.

(2) Effect of thermal stress on diodes performance

Diodes are fabricated by the Si-based HgCdTe material, diode's current–voltage (*I–V*) characteristics under IR illumination and nonillumination are evaluated, respectively, at 77 K. After a Kovar layer is tied to the back of the fan-out, the electrical characteristics are retested. In order to confirm the experiment results, the Kovar is removed, then reapplied and the diode performance is retested. Results of *I–V* under IR illumination and nonillumination are shown in Fig. 5.84. Table 5.8 provides photo current and dark current values at −100 mV versus the diodes with or without Kovar. As shown in Fig. 5.84a, b, the photo currents are dramatically increased while the dark current changed slightly. Experiment results indicate that the quantum efficiency of the diode increases after Kovar is applied to the back of the sapphire fan-out. Comparing the diode's performance after the removal and reapplication of the Kovar, it is shown that the effect of the stress on the performance can be duplicated and recoverable.

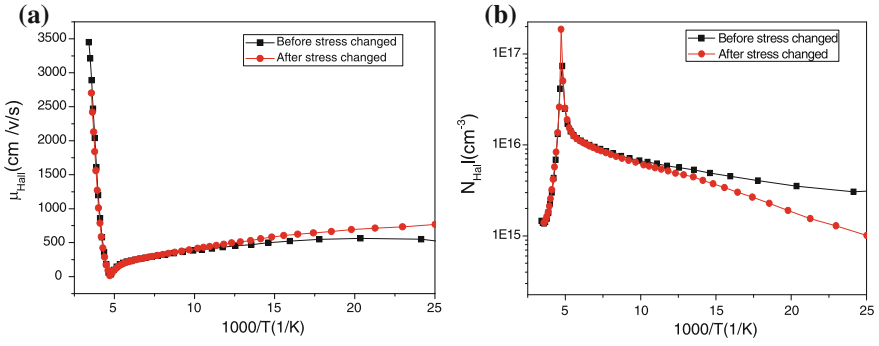


Fig. 5.83 Temperature dependence of carrier mobility (**a**) and concentration (**b**) for the HgCdTe/Si material before and after stress changed

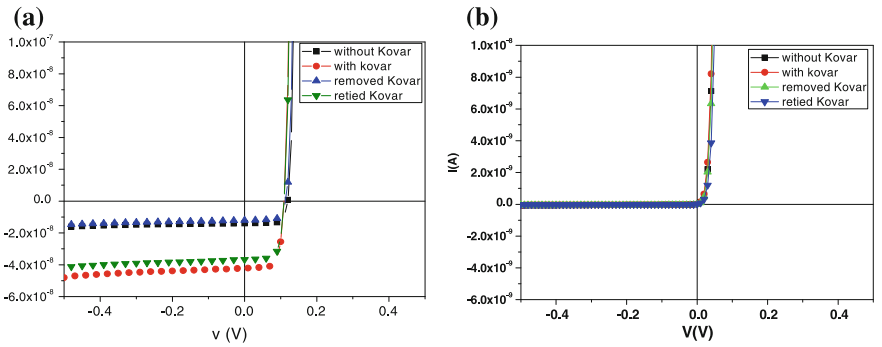


Fig. 5.84 *I–V* under IR illumination (**a**) and *I–V* in dark field (**b**) of HgCdTe/Si diodes with and without Kovar tied to the back of the sapphire fan-out

Table 5.8 Photo current and dark current values at −100 mV versus the diodes with and without Kovar

	Photo current at −100 mV (A)	Dark current at −100 mV (A)
Diodes without Kovar	1.3×10^{-8}	3.66×10^{-11}
Diodes with Kovar	4.1×10^{-8}	2.2×10^{-11}

(3) Effect of thermal stress on FPA performance

The cross-section view of the HgCdTe/Si FPA is shown as Fig. 5.72. After the Si substrate is thinned from 540 to 150 μm, and then to 98 μm, FPA performances are measured.

Table 5.9 shows the performance parameters of the 320 × 256 HgCdTe/Si FPA with different substrate thickness. All results are obtained at the same test condition (tint: 2000 μs, f/2.7). It is clear that the mean responsivity and detectivity increase while the root mean square (rms) values decrease. That is to say, the sensitivity and response uniformity of the FPA increase with substrate thinned to 150 and 98 μm, respectively.

By the experiment results, we can see that the performance of the Si-based HgCdTe material, diodes, and FPA regularly changes with the thermal stress and all results can be duplicated and recoverable.

Table 5.9 Performance parameters of the FPA with different substrate thickness

Thickness of substrate (μm)	540	150	98
Responsivity mean (V/W)	1.44E+07	1.54E+07	1.60E+07
Responsivity rms (V/W)	2.13E+06	1.84E+06	1.78E+06
Mean detectivity (cm/Hz1/2 W)	1.46E+09	1.67E+09	1.94E+09
Detectivity rms (cm/Hz1/2 W)	2.80E+08	2.77E+08	3.09E+08

(4) Theoretical residual films stress and discussions

The thermal stress along different direction in the Si-based films are calculated by the finite element analysis software ABAQUS. Figures 5.85, 5.86, and 5.87 show the simulation results of the thermal stress in the HgCdTe film for the material, diodes, FPA before and after the structure changed. The signs of the stress components are indicative of relative changes in the film. The positive values indicate the tensile stress, while negative values indicate the compressive stress. The unit of the stress value is MPa. The material parameters used in the calculation are shown in Table 5.7.

By comparison, the simulation results in Fig. 5.85 show that the stress in the HgCdTe is tensile, and the tensile stress decreases for the HgCdTe/Si material with sapphire tied to Si substrate. Figure 5.86 shows that the specified direction stress is compressive stress, and the stress is decreased until it changes to tensile stress for the diode with Kovar tied to the back of the sapphire fan-out. Figure 5.87 shows that, when Si substrate is thinned from 540 to 150 and 98 μm, the specified

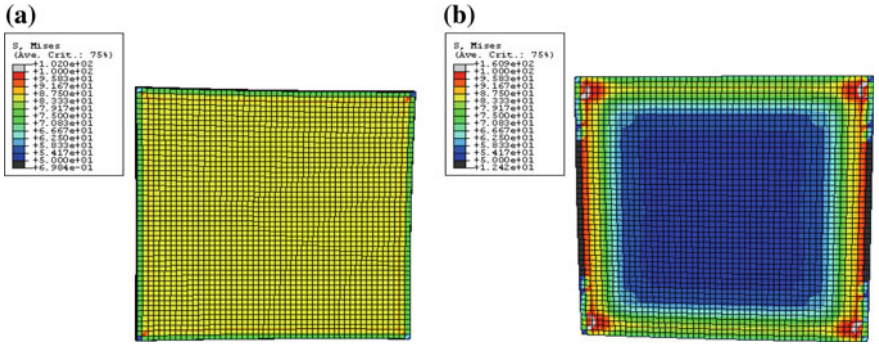


Fig. 5.85 Simulation results of average stress in HgCdTe film for material before (**a**) and after (**b**) stress changed

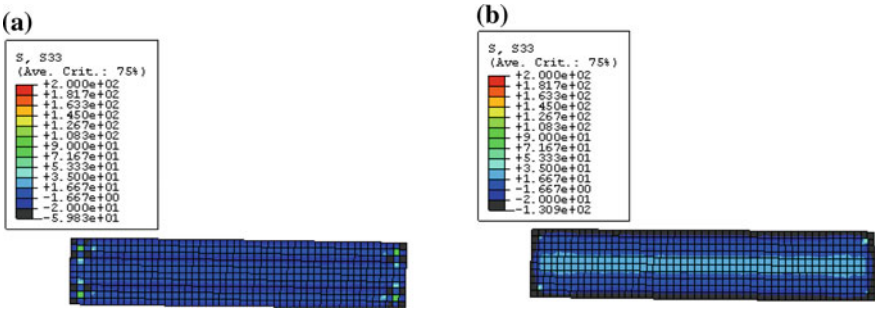


Fig. 5.86 Simulation stress of the HgCdTe film along the vertical direction to the HgCdTe film for the diodes without (**a**) and with (**b**) Kovar

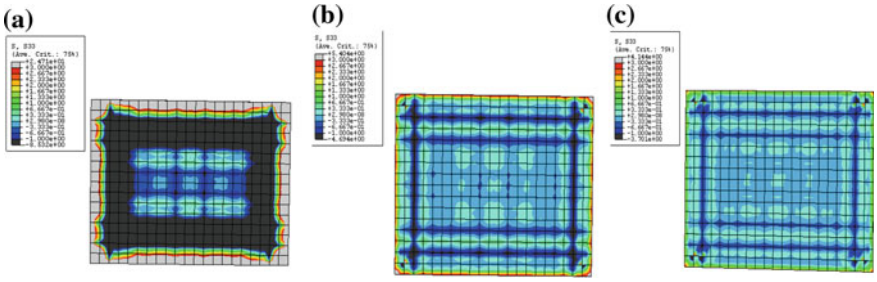


Fig. 5.87 Simulation stress of the HgCdTe along the vertical direction to the film for the FPA with different substrate thickness: **a** 540 μm; **b** 150 μm; **c** 98 μm

direction stress is compressive stress and the stress is decreasing until it changes to tensile stress which has the same change tendency as diodes. At the same time, the uniformity of the film stress is improved.

For the zinc blend compounds such as HgCdTe film, energy band under different thermal stress can analyze using the energy band theory, shown as Fig. 5.88.

In summary, the HgCdTe film with tensile stress exhibits higher electron mobility and lower hole mobility than that of the bulk HgCdTe crystal. For compressive uniaxial stress, the energy band of the HgCdTe changes inversely and results in hole mobility improvement and electron mobility decrease.

So based on energy band theory for HgCdTe under different uniaxial stress, the change of the performance with the thermal stress can be explained. Improvement of hole mobility in HgCdTe/Si material is found to be caused by a reduction of tensile in HgCdTe film. For HgCdTe/Si diodes, the mobility of electrons increases with increasing of tensile stress along the vertical direction to the HgCdTe film. Similar to diodes, improvement in FPA performance is found to be caused by the tensile stress increasing and improvement of the uniformity. In conclusion, the performance of the material, diodes, and FPA can be improved by optimization of thermal stress.

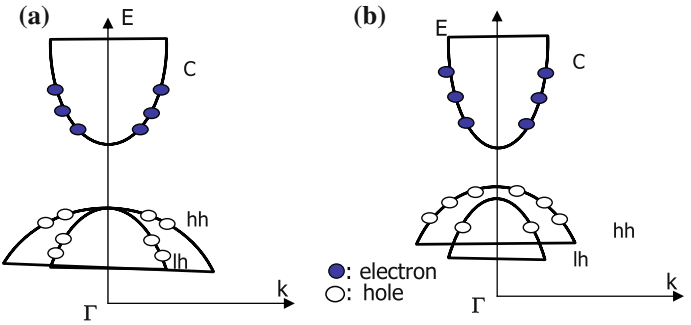


Fig. 5.88 Energy band of HgCdTe under different stress condition: **a** no stress; **b** uniaxial tensile stress

4. Structure design of Si-based HgCdTe chip

The stress analysis result indicates that the maximum thermal stress is located in the interface layer between Si substrate and HgCdTe epitaxial film, and the thermal stress of HgCdTe film is concentrated mainly in the center of chip. Thus, two different chip structures are designed, one is to introduce local stress releasing point onto the chip and the other is to apply the deep mesa etching to release or relieve the thermal stress.

In order to check the reliability of material in different chip structures, a cycle is defined by taking the Si-based HgCdTe chip with different chip structures and different thicknesses of epitaxial films from room temperature to liquid nitrogen temperature within 20 s, then back to room temperature within 10 min. The experiment result indicates that the capability of resistance to temperature cycles of the epitaxial film is enhanced after deep mesa etching, so the structure with the shallow and deep mesas is suitable for the Si-based HgCdTe chip.

5.4.2 *Low Damage Stress Chip Processing Technology of 3-Inch Si-Based HgCdTe Wafer*

1. Low damage processing technology

The Si-based HgCdTe FPA has an advantage of low cost. One reason is the low substrate cost; the other is that the large Si-based HgCdTe can take advantage of automated silicon processing equipment for increasing efficiency in device fabrication.

According to the chip structure mentioned in the above section, the Si-based HgCdTe FPA is divided into many periodic microzones by the crisscrossed deep trenches, and each microzone contains a number of micro-mesa pixels. The incomplete closed configuration of the deep trenches surrounding each microzone provides the electrical connecting narrow paths between the base region and the common electrode (seen in Fig. 5.89). During the thermal cycling, the shallow trench segmentation of micro-mesa pixel in the Si-based HgCdTe FPA can release

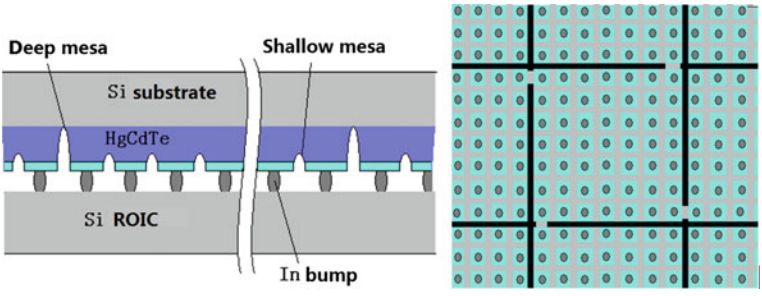


Fig. 5.89 Optimized micro-mesa FPA structure hybridizing deep and shallow mesas

partially the thermal mismatching stress of each microzone, and these connected microzones can also release or relieve the thermal mismatching stress in the large-scale IR FPA through free expansion. The above-mentioned chip architecture can be obtained by the mesa formation technique presented in Sect. 5.4.1.

The Si-based HgCdTe breaks through the size limit, so the device fabrication can draw on the experience of the Si microelectronic process for large wafers. However, due to the particular HgCdTe material characteristics such as fragility, poor thermal reliability, etc., and the existence of residual stress, the tools for the ion implantation, passivation, metallization, and tiny indium bump preparation of Si-based HgCdTe large area chips require a particular design. The large mismatching between the HgCdTe epitaxial layer and the Si substrate should be taken into consideration in the design of the process flow to avoid the increase in dislocations during the technological process. Additionally, the uniformity and reliability of wet chemical etching for the electrode hole should also be considered.

Through the successful development and confirmation of the chip processing technologies for the 3-in. Si-based HgCdTe on heterogeneous composite substrate wafers, the fabrication efficiency of IRFPAs can be improved and the development cost can be reduced. Figure 5.90 provides the picture of 3-in. Si-based HgCdTe detector chip, where nine 640 × 512 FPAs, eight 320 × 240 FPAs, and twelve 128 × 128 FPAs are achieved. It lays a solid foundation for the fabrication of large-scale, low-cost, and high-reliability IR FPAs.

2. Measurement and analysis of Si-based chip

Since huge lattice and thermal mismatching exist between the Si-based HgCdTe film and the substrate, the defect density and electrical properties of the material may be different from that on a conventional substrate. Therefore, it is necessary to investigate the performance of the detector prepared by the Si-based HgCdTe and

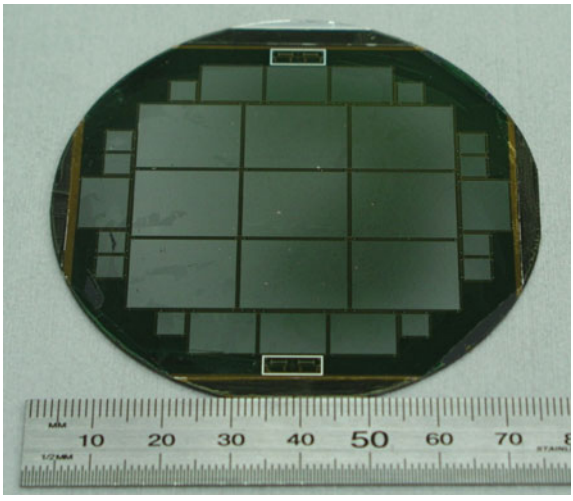


Fig. 5.90 Chip distribution on the 3-in. Si-based HgCdTe wafer

the performance changes after the thermal cycling between the ambient temperature and the liquid nitrogen temperature.

The n-on-p type photovoltaic detector is prepared based on the Si-based HgCdTe film. The *I–V* characteristics are measured [75], and the typical curves are shown in Fig. 5.91. The spectral response of the device is measured by the Fourier spectrometer as shown in Fig. 5.92. In comparison with the corresponding properties of HgCdTe epitaxial film on the conventional substrate, it indicates that the performance of the Si-based HgCdTe individual pixel is the same with that on the conventional substrate, but its uniformity and blind pixel rate is worse than the conventional substrate.

In order to investigate further the electrical properties of Si-based HgCdTe devices, the electrical measurement below 77 K is carried out on the device, and the measurement structure is shown in Fig. 5.93. The varying temperature *I–V* measurement is performed on the Si-based device, and the measurement results are shown in Figs. 5.94 and 5.95.

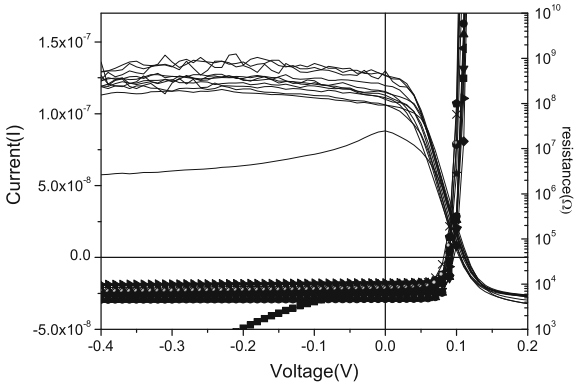


Fig. 5.91 *I–V* and *R–V* characteristic curves of the Si-based HgCdTe detector

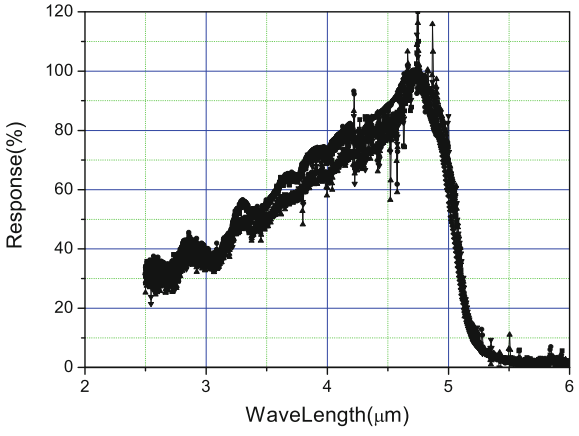


Fig. 5.92 Spectral response of the Si-based HgCdTe detector

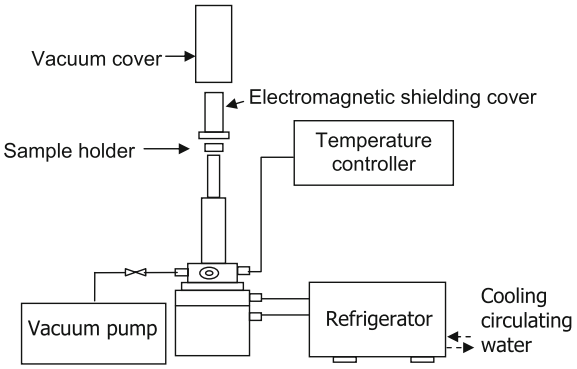


Fig. 5.93 Structure diagram of the varying temperature *I–V* characteristics measurement system

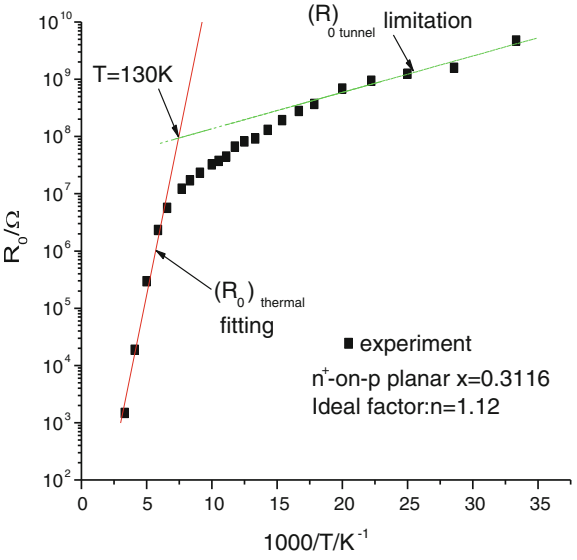


Fig. 5.94 R_0-1000/Tc experimental characteristic curve of Si-based HgCdTe detector

The electrical test data of the device are fitted by the *I–V* model of *p-n* junction, and the results are provided in Fig. 5.96 and Table 5.10.

The analysis results [75] indicate that in the positive bias, the dark current of the device depends mainly on the generation–recombination (G-R) current at low temperature and is determined partly by the G-R current and diffusion current with the increase of the temperature. In zero bias and small reverse bias, the dark current depends mainly on the trap assisted tunneling (TAT) current at low temperature and is determined partly by the TAT and the G-R current, especially the G-R current

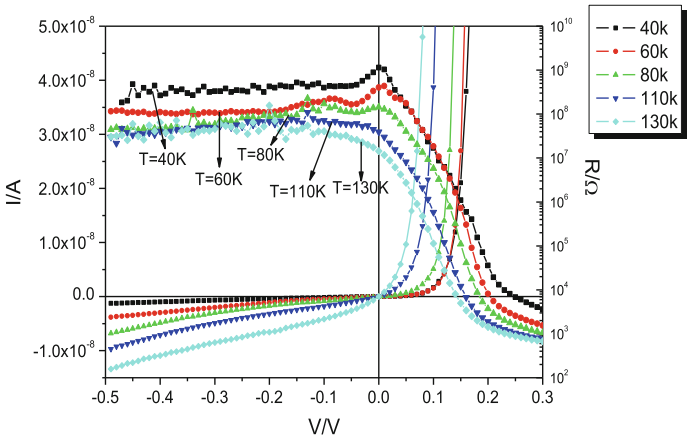


Fig. 5.95 The *I–V* and *R–V* characteristic curves at different temperature of Si-based HgCdTe detector

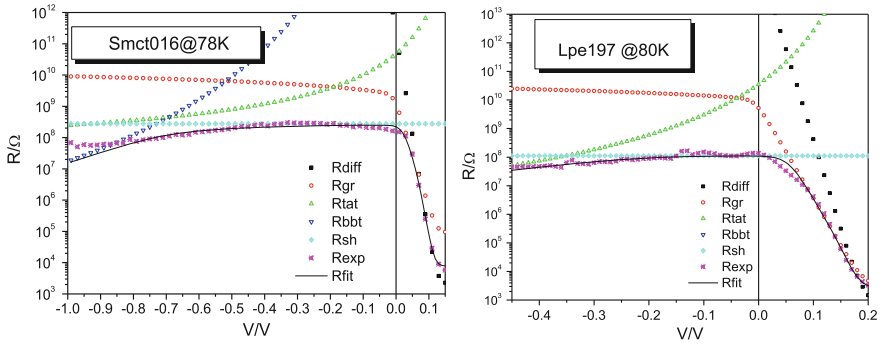


Fig. 5.96 Fitting analysis on the *R–V* experimental characteristic curves

Table 5.10 Fitting parameters of *I–V* characteristic curves

Sample	N_d/cm^{-3}	$\square_0$/ns	E_t/E_g	N_t/cm^{-3}	Rs/Ω	Rsh/Ω
L197-60K	2.3×10^{16}	0.012	0.68	9.6×10^{11}	4.0×10^{6}	1.9×10^{8}
L197-80K	2.3×10^{17}	0.51	0.57	2.8×10^{13}	2.0×10^{3}	1.1×10^{8}
L197-110K	8.1×10^{16}	0.05	0.54	6.4×10^{13}	1.2×10^{3}	5.8×10^{7}
S016-78K	2.2×10^{16}	0.12	0.56	1.3×10^{10}	5.6×10^{3}	2.8×10^{8}

plays a dominant role completely. In the medium reverse bias, the TAT current is dominant and the direct tunneling current rarely has contributions to the dark current. At the operating temperature of 80 K, the current mechanism near zero bias includes TAT current, G-R current, and surface leakage current. In order to improve

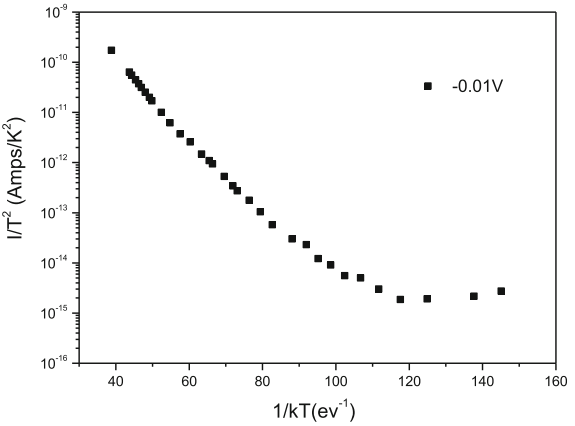


Fig. 5.97 The *I*-1/kT relationship of Si-based HgCdTe diodes under reverse bias 0.01 V

the device performance, the contributions of TAT current, G-R current, and surface leakage current to the dark current must be reduced.

The deep levels of the mid-wave infrared $Hg_xCd_{1-x}Te$ diodes (x = 0.31), which were fabricated on Si substrates have been studied using the varying temperature *I–V* measurement. The *I*-1/kT relationship is fitted when the reverse current is dominated by generation–recombination process, and the deep level 0.05135 eV (1/4 E_g) is calculated at the reverse bias 0.01 V. Figure 5.97 is the *I*-1/kT relationship of Si-based HgCdTe diodes under reverse bias 0.01 V. Figure 5.98 is the simulation results of the deep level.

Using the same method, the deep levels at different reverse biases are also investigated and shown in Fig. 5.99. The experimental observation that the deep

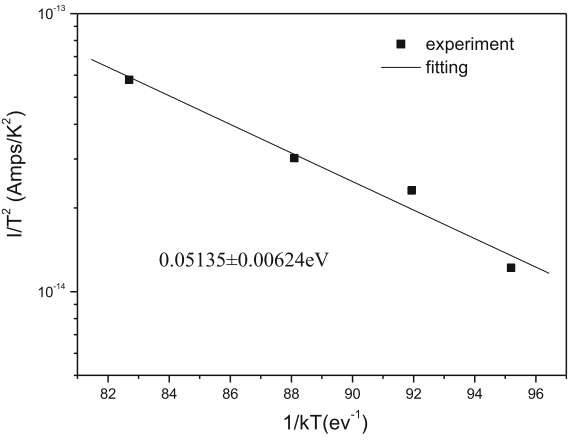


Fig. 5.98 The simulation results of the deep level at different temperature range 120–140 K

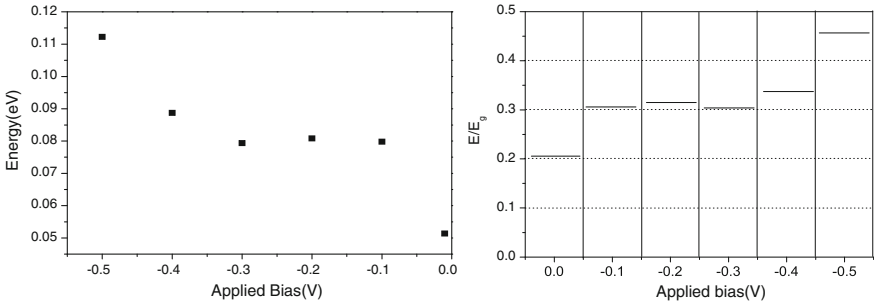


Fig. 5.99 The deep levels in Si-based HgCdTe diodes at different reverse biases

level at the 0 reverse bias was 0.2 E_g which is related to the effect caused by ion implanting at the deep level with a reverse bias of −0.1 to −0.3 V and was 0.3 E_g. The deep level at the reverse bias of 0.5 V was 0.45 E_g which may be caused by the defect energy related to the TAT current. So the source of these deep levels corresponds well with the reverse current mechanisms (Fig. 5.100).

The minority carrier diffusion length at different temperatures can be obtained by fitting the relationship between the product of zero-bias resistance and area (R_0A) and p/A. It shows that the minority carrier diffusion length increases with increasing of temperature below 200 K, while the minority carrier diffusion length decreases with increasing of temperature above 200 K. The minority carrier lifetime of Hg vacancy p-type HgCdTe on Si substrate was calculated by the minority diffusion length at different temperature. The magnitude and relationship with the temperature of HgCdTe/Si minority carrier lifetime is same with the HgCdTe/CdZnTe material by comparing the minority carrier lifetime in HgCdTe material grown on CdZnTe substrates.

3. Some results of Si-based HgCdTe staring FPAs

Recently, the Si-based HgCdTe epitaxy technology and the chip design and fabrication technologies have made progress. And the design technologies and technological processes mentioned above are applied to the development of MW Si-based HgCdTe FPA. Large format staring HgCdTe IRFPA based on Si is developed from 320 × 240 to 640 × 512 and 1024 × 1024, shown in Figs. 5.101 and 5.102. The 640 × 512 Si-based HgCdTe FPA IDDCA (Integrated Detector Dewar Cooler Assemblies) is shown in Fig. 5.103. The images taken by the 640 × 512 Si-based HgCdTe FPA IDDCA are provided in Fig. 5.104.

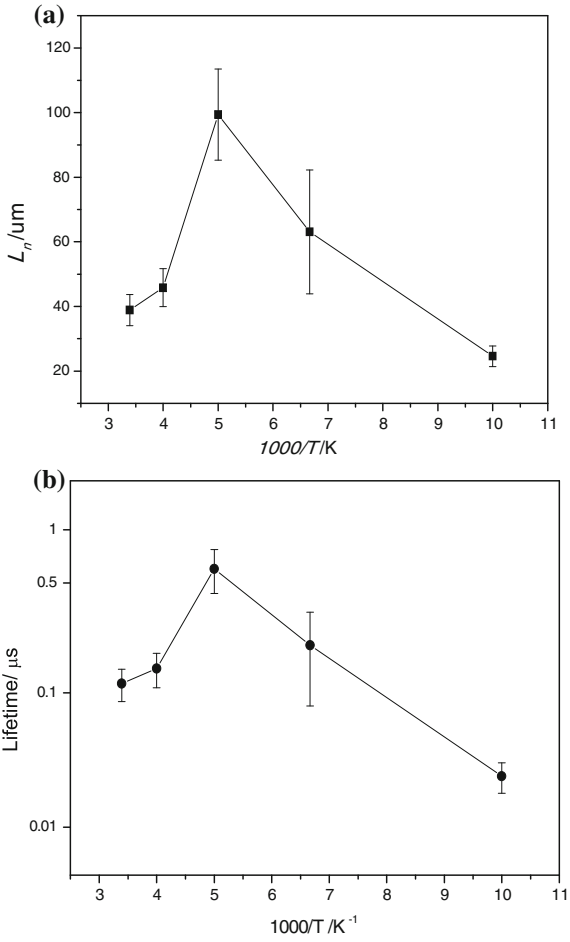


Fig. 5.100 The minority carrier diffusion length (**a**) and lifetime (**b**) of Si-based HgCdTe material versus temperature

Fig. 5.101 640 × 512 Si-based device

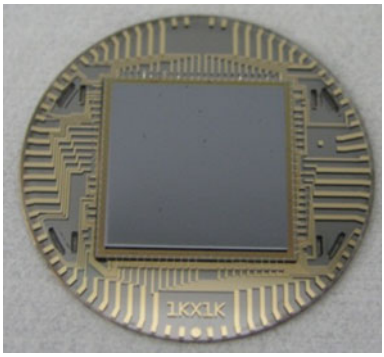


Fig. 5.102 1 K × 1 K Si-based device

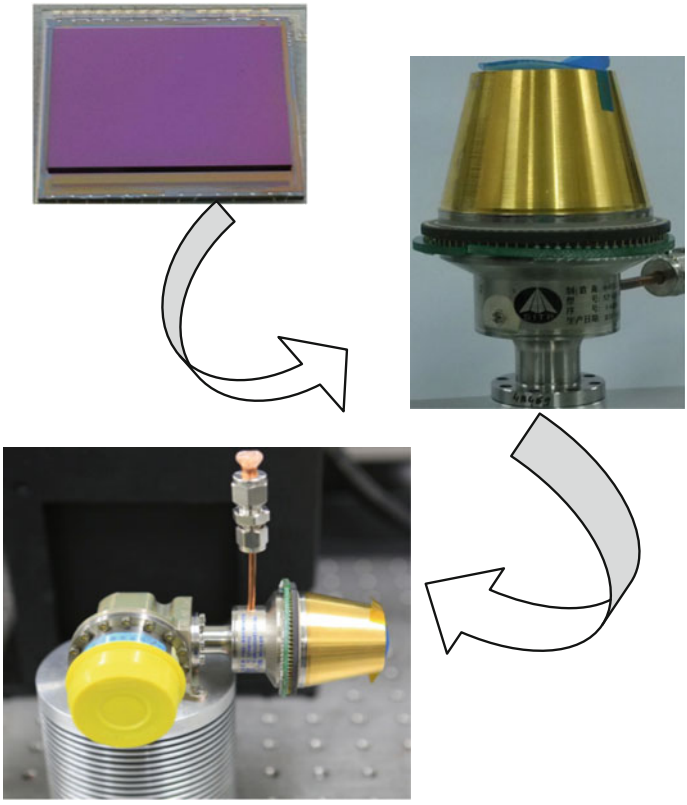


Fig. 5.103 640 × 512 Si-based HgCdTe IDDCA

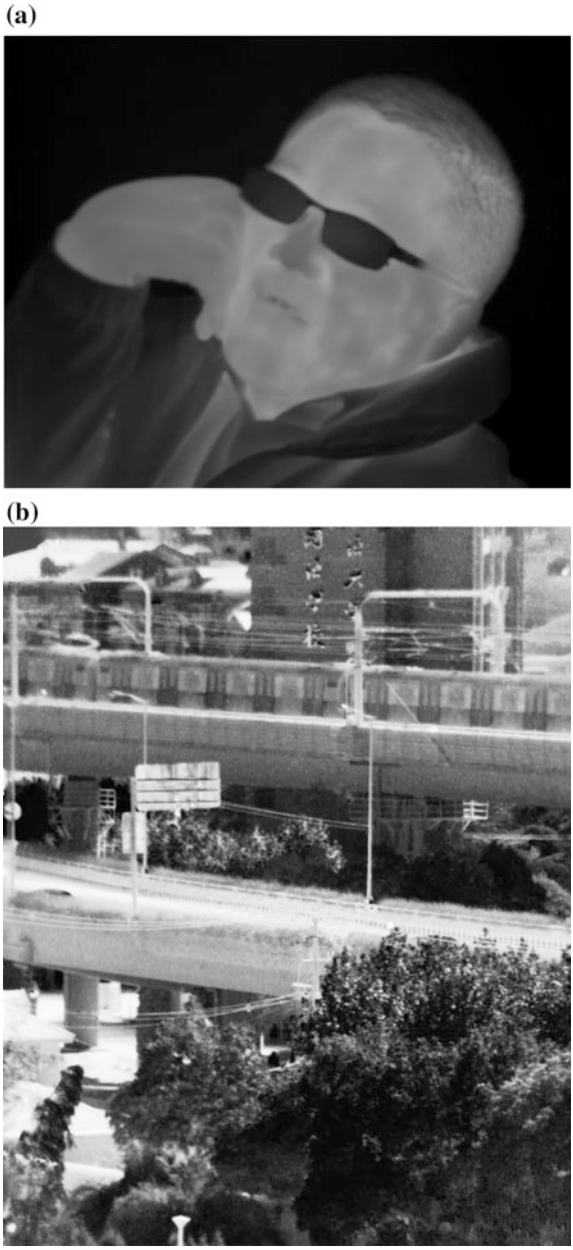


Fig. 5.104 Images taken by the 640 × 512 Si-based HgCdTe IDDCA

5.5 Summary

In this chapter, parts of the core technologies of the third-generation IR FPAs are introduced. The important aspect of advanced IR FPAs is that the very-large-scale FPA chips are applied to improve the spatial resolution and the multiband detection are achieved by multilayer heterojunction structures to improve the recognition capability of the target. The fabrication of the multilayer heterojunction structure detector chip involves the semiconductor technologies such as corrosion, etching, photolithography, and film growth, but different from that of the conventional chip. The difficulty of using the conventional technologies needs to be overcome, and the new processing technology for the deep micro-mesa HgCdTe chip should be further developed. In this chapter, the deep micro-mesa chip processing technologies are demonstrated systematically, mainly including ICP etching of mesa isolation, mesa passivation, metallization, and indium bump preparation and interconnection. Meanwhile, with the continuous expansion of the detector scale, Si-based HgCdTe is applied in the very-large-scale FPAs to solve the problems of the conventional substrate size limit and high cost. Here, in view of the material and chip features of Si-based HgCdTe and the problems on the fabrication of the stress chip, the technological process of the Si-based HgCdTe device is described.

For the development of the multilayer heterojunction small pixel devices, the processing technologies of deep micro-mesa chips are introduced, involving the etching formation of the high AR trenches for isolating micro-mesa array, the photolithography technique of micro-mesa array, the passivation of the concave surface in mesas, the fabrication of the climbing metal films, the preparation of indium bumps and flip chip welding. As these key technologies are introduced, the uniqueness of HgCdTe technologies is analyzed, and the differences from the conventional semiconductor process are discussed. In the isolation techniques of micro-mesa arrays, the feature that the Hg–Te bond is relatively weak in the HgCdTe material should be taken into consideration, and the physical damages and electrical damages must not be introduced in the specified technological process to ensure the normal photoelectric properties of the pixel after mesa isolation.

For the isolation technology of micro-mesa arrays, first, the main etching methods and the characterization parameters of HgCdTe are introduced. The technical requirements of isolating in the HgCdTe micro-mesa arrays are analyzed, and the conventional semiconductor etching techniques and the reported HgCdTe etching techniques are summarized. The limitation of the wet chemical etching technique is demonstrated, and then the ICP enhanced RIE etching methods suitable for HgCdTe are introduced, as well as the etching parameters for the characterization of the etching process, including etching rate, selectivity ratio, uniformity, surface morphology, etching damage, etc.

Second, the HDP-ICP etching process is introduced, including the high selectivity ratio mask technique of ICP process, analysis of the influencing factors of etching rate, selection of etching condition, and the technological process, parameter, and test method of ICP etching. Based on the current advanced ICP etching

technique, the research status of HgCdTe micro-mesa isolation technologies is introduced. The ICP enhanced RIE etching for HgCdTe is achieved based on the plasma of CH_4 and H_2, and the key process is that the plasma formed by CH_4/H_2 reacts with Te, Cd, or Hg of HgCdTe surface, and the volatile products such as $Te(CH_3)_2$, $Cd(CH_3)_2$, $Hg(CH_3)_2$, and TeH_2 are generated. The optimization of etching rate is described, as well as loading effect, microloading effect, and uniformity. The selectivity ratio of the mask and the influence of its profile on etching are discussed. When optimizing the etching process, the contradiction between reducing dry etching induced electrical damages and avoiding etching lag effect is required to be balanced properly. Meanwhile, in order to improve the duty cycle of micro-mesa, it is proposed that in the design of etching process, the section of the trenches for mesa isolation is expected to be "V" shaped with satisfying the design requirement of trench depth, and the sizes of etching patterns should not be more than and less than the critical dimension simultaneously.

Then, the etching damages of HgCdTe and their estimation methods are introduced. The mechanism and estimation methods of ICP etching induced electrical damages in HgCdTe micro-mesa isolation are analyzed, as well as the weakening condition of etching induced damages and the repair treatment of the electrical damages. By measuring the thickness of electrical damages, it is found that the ICP etching induced damages caused by H^+ and H are obviously more than that of the physical etching induced damages in the same condition. Therefore, the mechanism of the ICP etching induced electrical damages is investigated further. The influences of CH_4 the based process gas and the H_2 based process gas on the polymers and the etching damages produced by etching are compared. Based on the measurement methods such as *I–V*, LBIC, differential Hall, XRD, and microzone surface potential, the distribution and heat treatment repair technique of the electrical damages induced by dry etching p-type HgCdTe material are obtained. In the optimization of the etching process, it is presented that the high AR mask patterns are applied to reduce the long duration collision of the etching ions deviating from the vertical direction to the sidewall of trenches, thereby decreasing the etching induced electrical damages on the sidewalls.

Photolithography, sidewall passivation, climbing metallization, and hybridized interconnection after the formation of deep micro-mesa array are introduced in detail. These technologies are the core technologies for developing the third generation of FPAs. As an integrated application with these technologies, the development results of prototype two-color HgCdTe micro-mesa detector chips are provided.

In the aspect of Si-based HgCdTe FPAs, the HgCdTe epitaxial material grown on Si substrates and the features of the fabricated detector are introduced, and the stress structural distribution of Si-based FPAs is analyzed. In view of the feature of the stress structure, the chip architecture combining deep mesa and shallow mesa is adopted for the design of detector. The electrical measurement results for MW Si-based HgCdTe devices indicate that the R_0 values of the devices on different substrates are comparative in magnitude, but the homogeneity of *R–V* curves of Si-based HgCdTe devices is worse than that of the GaAs-based devices.

The analysis results indicate that this is due to the problem of the internal stress in Si-based HgCdTe devices at cryogenic temperatures. The performance parameters of the devices, such as response, noise, and detectivity, are compared, and the results show that the performance of the pixels in the Si-based HgCdTe detector is comparable to the detectors prepared on conventional substrates. The measurement and analysis of the varying temperature *I–V* characteristics are performed on the MW Si-based HgCdTe device, and the results show that the main dark current mechanism near zero bias at the operating temperature of 80 K includes TAT current, G-R current, and surface leakage current. To improve the device performance, the TAT current, the G-R current, and surface leakage current are reduced to lower the dark current. Accordingly, in order to lower the dark current of Si-based HgCdTe devices, the minority carrier lifetime and surface passivation are improved and the surface recombination velocity is reduced to lower the G-R current and surface leakage current. The material quality is improved and the trap concentration of *p-n* junction is reduced so as to reduce the contribution of TAT current.

The stress analysis result indicates that the location of thermal mismatch maximum is between the Si substrate and HgCdTe epitaxial film, and the internal stress of HgCdTe film is concentrated mainly on the middle position. Thus, different chip structures are designed, and two methods of introducing a local stress releasing point onto the chip and the deep mesa etching are applied to release or relieve the thermal stress.

In order to check the capability of resistance to the cleavage of material in different chip structures, the temperature cycle shock experiment (cooled rapidly to 80 K with liquid nitrogen within 20 s, heated rapidly up to 300 K within 1 min) is performed on the Si-based HgCdTe chip with different chip structures and different thicknesses of epitaxial films. The experimental result indicates that the capability of resistance to temperature shock of the epitaxial film is enhanced after deep mesa etching, so the structure combining the shallow and deep mesas is suitable for a Si-based HgCdTe chip.

References

1. Rogalski A. New material systems for third generation infrared detectors. Proc SPIE. 2009; V7388:P73880J-1.
2. Bratt PR, Johnson SM, Rhiger DR, et al. Historical perspectives on HgCdTe material and device development at Raytheon vision systems. Proc SPIE. 2009;7298.
3. Bensussan P, Tribolet P. 50 years of successful MCT research and production in France. Proc SPIE. 2009;7298.
4. D'Souza AI, Robinson EW, Stapelbroek MG. Visible to SWIR response of HgCdTe HDVIP detectors. Proc SPIE. 2009;7298.
5. Rogalski A, Antoszewski J, Faraone L. Third-generation infrared photodetector arrays. J Appl Phy. 2009;105(09).
6. Zhenhua Ye. Research on integrated HgCdTe two-color detector chip techology. Shanghai: Shanghai Institute of Technical Physics; 2005.

7. Srivastav V, Pal R, Vyas HP. Overview of etching technologies used for HgCdTe. Opto-Electron Rev. 2005;13(3):197–211.
8. Zhou W, Ye Z, Xing W, et al. The study on the profile of HgCdTe micro-mesa arrays isolated by dry-etch process. Laser Infrared. 2006;36(11):1029–31.
9. Ye Z, Hu X, Quan Z, et al. Study on etch pattern of dry technique for HgCdTe IRFPAS. J Infrared Millimeter Waves. 2006;25(5):325–8.
10. Guo J, Ye Z, Hu X, et al. XPS analysis of the HgCdTe surface etched by ICP. Laser Infrared. 2005;35(11):832–4.
11. Wenhong Zhou. Research on the ICP technique of HgCdTe micro-mesa dry isolation. Shanghai: Shanghai Institute of Technical Physics; 2009.
12. Ye Z, Wu J, Hu X, et al. Study of integrated MW1/MW2 two-color HgCdTe infrared detector arrays. J Infrared Millimeter Waves. 2004;23(3):193–196.
13. He L, Hu X, Ding R, et al. Recent progress of the 3rd generation infrared FPAs. Infrared Laser Eng. 2007;36(5):696–701.
14. Hu W, Ye Z, Liao L, et al. 128 × 128 long-wavelength/mid-wavelength two-color HgCdTe infrared focal plane array detector with ultralow spectral cross talk. Optics Lett. 2014;39(17).
15. Ye Zh H, Quan Zh J, Zhou W H, et al. The 2006 joint 31st international conference on infrared and millimeter waves and 14th international conference on terahertz electronics: an investigation on spectral-characteristic of HgCdTe two-color detector, Shanghai, 18–22 Sept 2006. China: Shanghai.
16. Baylet J, Gravrand Q, Laffosse E, et al. Study of the pixel-pitch reduction for HgCdTe infrared dual-band detectors. J Electron Mater. 2004;33(6):690–700.
17. Horn S, Norton P, Cincotta T, et al. Challenges for third-generation cooled imagers. Proc SPIE. 2003;5074:44–51.
18. Harris KA, Enders DW, Yanka RW, et al. Electron cyclotron resonance plasma etching of HgTe-CdTe superlattices grown by photo-assisted molecular beam epitaxy. J Electron Mater. 1995;24(9):1201–6.
19. Antoni Rogalski. Infrared detectors: status and trends. Prog Quantum Electron. 2003;27(2–3): 59–210.
20. Hu X, Ye Z, Ding R, et al. Low energy inductively coupled plasma etching of HgCdTe. Proc SPIE. 2005;5964.
21. Quirk M, Serda J. Semiconductor manufacturing technology. In: Han Z, Hai C, Xu Q, et al., editor. Beijing: Publishing House of Electronics Industry; 2004.
22. Ye Z, Guo J, Hu X, et al. A study on etch rate of dry technique for HgCdTe IRFPAs. Laser Infrared. 2005;35(11):829–31.
23. Chang CY, Sze SM. ULSI technology. New York: McGraw-Hill; 1996.
24. Zhou W, Ye Z, Hu X, et al. Study on mask technology of HgCdTe etched by inductively coupled plasma dry etching. Laser Infrared. 2007;37(348):928–30.
25. Smith EPG, Musca CA, Redfern DA, et al. H2-based dry plasma etching for mesa structuring of HgCdTe. J Electron Mater. 2000;29(6):853–8.
26. Haakenaaen R, Colin T, Steen H, et al. Electron beam induced current study of ion beam milling type conversion in molecular beam epitaxy vacancy-doped CdxHg1-xTe. J Electron Mater. 2000;29(6):849–52.
27. Ye Z, Wu J, He L, et al. A preliminary study on MBE-grown HgCdTe two-color FPAs. Proc SPIE. 2004;5640.
28. Smith EPG, Venzor GM, Goetz PM. Scablity of dry-etch processing for small unit-cell HgCdTe focal-plane arrays. J Electron Mater. 2003;32(7):821–6.
29. Antoszewski J, Musca CA, Dell JM, et al. Characterization of Hg0.7Cd0.3Te n-on p-type structure obtained by reactive ion etching induced p-to n conversion. J Electron Mater. 2000;29(6):837–40.
30. Ye Zhenhua Hu, Xiaoning Zhang Haiyan, et al. Study of dark current for mercury cadmium telluride long-wavelength photodiode detector with different structures. J Infrared Millimeter Waves. 2004;23(2):86–90.

31. Horn S, Norton P, Cincotta T, et al. Challenges for third-generation cooled imagers. Proc SPIE. 2003;5074:44–51.
32. Vossen JL, Kern W. Thin film processing. New York: Academic Press; 1978.
33. Yang KD, Lee YS, Lee HC. Annealing behavior of hydrogen-plasma-induced n-type HgCdTe. Appl Phys Lett. 2005;87(11).
34. White J, Pal R, Dell JM. p-to-n type conversion mechanisms for HgCdTe exposed to H2/CH4 plasmas. J Electron Mater. 2001;30(6):762–76.
35. Neyts E, Yan M, Bogaerts A, et al. Particle-in-cell/monte carlo simulations of a low-pressure capacitively coupled radio-frequency discharge: effect of adding H2 to an Ar discharge. J Appl Phys. 2003;93(9).
36. Dell JM, Antoszewski J, White JK, et al. RIE induced n-on-p junction HgCdTe photodiodes: effects of passivant technology on bake stability. Proc SPIE. 2001;4454:106–14.
37. Stoltz AJ, Benson JD, Baresi JB. Macro-loading effects of electron-cyclotron resonance etched II–VI materials. J Electron Mater. 2004;33(6):684–9.
38. Keller RC, Seelmann-Eggebert M, Richter HJ. Dry etching of HgCdTe using CH4/H2/Ar/N2 electron cyclotron resonance plasmas. J Electron Mater. 1996;25(8):1270–5.
39. Benson JD, Stoltz AJ, Boyd PR, et al. Lithography factors that determine the aspect ratio of electron cyclotron resonance plasma etched HgCdTe trenches. J Electron Mater. 2003;32 (7):686–91.
40. Mogab CJ. The loading effect in plasma etching. J Electrochem Soc. 1977;124(8):1262–8.
41. Laffosse E, Baylet J, Chamonal JP, et al. Inductively coupled plasma etching of HgCdTe using a CH4-based mixture. J Electron Mater. 2005;34(6):740–5.
42. Sonag KH, Yoon TH, Hahn SR, et al. Changes in surface characteristics of HgCdTe by dry etching. Proc SPIE. 1998;3436:77–83.
43. Stoltz AJ, Benson JD, Smith PJ, et al. Morphology of inductively coupled plasma processed HgCdTe surfaces. J Electron Mater. 2008;37(9):1225–30.
44. Bommena R, Velicua S, Boieriu P, et al. Uniformity studies of inductively coupled plasma etching in fabrication of HgCdTe detector arrays. Proc SPIE. 2007;6542.
45. Chandra D, Schaake HF, Aqariden F, et al. p to n conversion in SWIR mercury cadmium telluride with ion milling. J Electron Mater. 2006;35(6):1470–1473.
46. Stoltz AJ, Sperry MJ, Benson JD, et al. A langmuir probe investigation of electron cyclotron resonance argon-hydrogen plasmas. J Electron Mater. 2005;34(6):733–9.
47. Smith EPG, Musca CA, Redfern DA, et al. H2-based dry plasma etching for mesa structuring of HgCdTe. J Electron Mater. 2000;29(6):853–8.
48. Lee MH, Jang SH, Chung CW. Floating probe for electron temperature and ion density measurement applicable to processing plasmas. J Appl Phys. 2007;101(3).
49. Rhoderick EH. Metal-semiconductor contacts. Oxford: Clarendon Press; 1978.
50. Liu E, Zhu B. Semiconductor physics. 4th ed. Beijing: National Defence Industry Press; 2007. p. 211.
51. Spicer WE, Friedman DJ, Carey GP. The electrical properties of metallic contacts on $Hg_{1-x}Cd_xTe$. J Vac Sci Technol. 1988;6(4):2746–51.
52. Zimmermann H, Keller RC, Meisen P, et al. Interface formation between deposited Sn and $Hg_{0.8}Cd_{0.2}Te$. J Electron Mater. 1996;25(8):1293–9.
53. Leech PW. The specific contact resistance of ohmic contacts to $HgTe/Hg_{1-x}Cd_xTe$ heterostructures. J Appl Phys. 1990;68(2):907–9.
54. Beck WA, Davis GD, Goldberg AC. Resistance and 1/f noise of Au, Al, and Ge contacts to (Hg, Cd)Te. J Appl Phys. 1990;67(10):6340–9.
55. Merken P, John J, Zimmermann L, et al. Technology for very dense hybrid detector arrays using electroplated indium solder bumps. IEEE Trans Adv Packag. 2003;26(1):60–4.
56. Pierre C, Francois M, Jean-Luc M, et al. A megapixel HgCdTe MWIR focal plane array with a 15 μm pitch. Proc SPIE. 2004;5251:65–72.
57. Caccia M. The challenge of hybridization. Nucl Instrum Meth Phys Res A. 2001;465:195–9.

58. Burggraaf P. Chip scale and flip chip: attractive solutions. Solid State Technol. 1998;7: 239–46.
59. Chang C, Sze S. ULSI technology.New York: McGraw-Hill; 1996.
60. Babiarz A. Key process controls for underfilling flip chips. Solid State Technol. 1997;7:77–81.
61. Gemme C, Fiorello AM, Gagliardi G, et al. Study of indium bumps for the ATLAS pixel detector. Nucl Instrum Meth Phys Res A. 2001;465:200–3.
62. Jiang JT, Tsao S, O'Sullivan T, et al. Fabrication of indium bumps for hybrid infrared focal plane array applications. Infrared Phys Technol. 2004;45:143–151.
63. Cihangir S, Kwan S. Characterization of indium and solder bump bonding for pixel detectors. Nucl Instrum Meth Phys Res A. 2002;476:670–5.
64. Babiarz A. Key process controls for underfilling flip chips. Solid State Technol. 1997;7:77–81.
65. Tissot JL, Marion F. Collective flip-chip technology for HgCdTe IRFPAs. Proc SPIE. 1996;2894:115–22.
66. Destefanis G, Astier A, Baylet J, et al. Recent developments of high-complexity HgCdTe focal plane arrays at leti infrared laboratory. J Electron Mater. 2003;32(7):592–601.
67. Oppermann HH, Kallmayer C, Klein M, et al. Advanced flip chip technologies in RF, microwave and MEMS applications. Proc SPIE. 2000;4019:308–14.
68. Kwan S, Appel JA, Chiodini G, et al. The 7th workshop on electronics for LHC experiments: a study of thermal cycling and radiation effects on indium and solder bump bonding, Batavia, 10–14 Sept 2001. USA: Fermilab.
69. John J, Zinmermann L. High density hybrid interconnect technologies. Proc SPIE. 2004;5454:60–7.
70. Rogalski A, Antoszewski J, Faraone L. Third-generation infrared photodetector arrays. J Appl Phys. 2009;105(09):1103–46.
71. Ye Z, Yin W, Huang J, et al. 128 × 128 SW/MW two-color HgCdTe IRFPAs. J Infrared Millimeter Waves. 2010;29(6):415–8.
72. Brill G, Velicu S, Boieriu P, et al. MBE growth and device processing of MWIR HgCdTe on large area Si substrate. J Electron Mater. 2001;30(6):717–22.
73. Hu X, Zhang H, Li Y, et al. Thermal stress analysis of HgCdTe focal plane arrays on Si substrates. Laser Infrared. 2006;36(11):1020–1022.
74. Feng ZC, Liu HD. Generalized formula for curvature radius and layer stresses caused by thermal strain in semiconductor multilayer structures. J Appl Phys. 1983;54(1):83–5.
75. Yue T, Yin F, Hu X. Characterization analysis of dark current in HgCdTe/Si photodiodes. Laser Infrared. 2007;37:931–934.
76. He L, Fu XL, Zh Wei Q, et al. MBE HgCdTe on alternative substrates for FPA applications. J Electron Mater. 2008;37(9):1189–99.
77. Capper P. Properties of narrow gap cadmium based compounds. England: Short Run Press Ltd.; 1994. p. 399–407.

Chapter 6
Chip Technique of AlGaN Focal Plane Arrays

6.1 Introduction

After infrared and laser detection techniques were developed, ultraviolet detection techniques become another important photoelectric detection tool used in both military and civilian fields. As a complement of infrared technique, the ultraviolet detection technique is widely used in missile warning, precise guidance, secure communication, biochemical analysis, fire detection, biomedical analysis, ozone monitoring, oil contamination monitoring, solar illumination monitoring, and public security surveillance [1–3].

Photomultiplier tubes and some vacuum devices have been used as ultraviolet detectors with a high sensitivity for quite a long time. The UV-enhanced photodiode based on Si material was also used as one representative of UV solid state detectors. Along with the development of growth technique of wide band gap semiconductor material, intrinsic UV photoelectric detectors based on wide band gap semiconductor material became a new development direction. Among these detectors, the GaN detector is one with the best development potential. As a representative of third generation semiconductor materials, GaN-based thin film materials have many advantages, such as a direct band gap with a large changeable range, high electron drift velocity, small dielectric constant, and so on. Because wide band gap materials have excellent physical and chemical stability, the GaN-based detectors can work in harsh conditions, especially for AlGaN thin film materials. The band gap of AlGaN material can be adjusted by the composition of Al in a range from 3.4 to 6.2 eV. The corresponding cutoff wavelength is in a range from 200 to 365 nm. Therefore, AlGaN is an ideal material to fabricate UV detectors [4].

Because GaN detectors have high responsivity, high reliability, and the fabrication process is compatible with Si processes, GaN-based photoelectric detectors usually use back-illuminated p–i–n structure and can be integrated to form large-scale array detectors [5, 6]. The detection waveband of UV detectors can be

L. He et al., *Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN*,
DOI 10.1007/978-3-662-52718-4_6

divided into two wavelength bands corresponding to visible-blind (300–365 nm) and solar-blind (240–280 nm), respectively. In some papers, 250–280 nm or 260–280 nm are also defined as the solar-blind waveband. In the visible-blind waveband, the responsivity of GaN-based p–i–n UV detectors can exceed 0.20 A/W [7]. Because the solar-blind waveband has important applications in military use, AlGaN solar-blind detectors have become the research hotspot of UV detectors. In 2004, the UV detector with a responsivity of 0.136 A/W and a cutoff wavelength of 282 nm was achieved at zero bias by McClintock et al. [8]. The external quantum efficiency of the detector exceeds 60 %. A resonant-cavity-enhanced (RCE) GaN UV detector was proposed by Selim et al. [9].

The wide band gap of the material brings a series of problems in GaN UV detector processing, such as mesa etching, Ohmic contact, interface passivation, and so on. In this chapter, the response model and design of AlGaN p–i–n solar-blind UV detectors will be introduced in Sect. 6.2. In Sect. 6.3, the structure design and detector fabrication process of resonant-cavity-enhanced (RCE) GaN UV detectors with a precisely controlled cutoff wavelength are introduced. In Sect. 6.4, the fabrication processes of the detectors, such as micro-mosaic etching, passivation, and Ohm contact are introduced. As an important aspect of the detectors in aerospace application, adaptability, and reliability of the detectors under irradiation (including proton irradiation, electron irradiation, and gamma irradiation) are analyzed in Sect. 6.5. At the end of this chapter, some applications of GaN-based UV focal plane assembly are presented.

6.2 AlGaN P–I–N Solar-Blind UV Detectors-Model and Design

The current research of UV detectors focuses on the solar-blind spectrum. Photovoltaic devices, especially p–i–n structures, have several advantages, (i) fast response; (ii) both the cutoff and cut-on wavelength can be tuned through changing the composition and doping of the AlGaN layers, (iii) relative small capacitors, low noise, and excellent response uniformity, (iv) large fill factors and excellent reproducibility. Therefore, AlGaN p–i–n detectors have become the technology of choice for linear and area UV focal plane arrays. It is very important to study the theoretical model of optoelectronic response for material and device design.

A photodetector usually does not respond to light with a wavelength outside of its working wavelength bands as determined by the material structures. Within the working band, their characteristics, such as quantum efficiency, background noise, operation speeds, feasibility for integration for wide view field, and multi-spectrum closely depend on the material properties. The same characteristics apply to a UV detector [10, 11]. Therefore, the design of AlGaN UV detectors should start from the structural and electrical parameters of the materials and devices. Then calculate and simulate their optoelectronic properties, such as quantum efficiency, noise,

band width, responsivity, and sensitivity. In this chapter, we will follow the general theories of optoelectronic devices with special attention to the characteristics of wide band gap materials, to derive the response model of AlGaN p–i–n detectors.

6.2.1 Material Parameters of AlGaN Thin Films

The band gap E_g of $Al_xGa_{1-x}N$ varies with its composition x, which at room temperature (300 K) can be express as follows [12]:

$$E_g(x) = E_g(\mathrm{GaN})(1-x) + E_g(\mathrm{A1N})x - bx(1-x) \tag{6.1}$$

where E_g (GaN) = 3.4 eV is the band gap of GaN, E_g (AlN) = 6.2 eV the band gap of AlN and b = 1.0 eV, a bowing parameter.

The $Al_xGa_{1-x}N$ absorption coefficient and minority diffusion length depends on epitaxial techniques and processes and its reflectivity on the refractive index and the flatness of the surface layers. The reported absorption coefficients are not consistent in the literature to date and the reported minority diffusion length varies from 50 nm through 3.4 μm. Muth et al. [13] demonstrated the relationship of the absorption coefficient, refractive index on incident light energy of $Al_xGa_{1-x}N$. For intrinsic absorption, the coefficients are usually higher than 10^5 cm^{-1}.

6.2.2 Response Model and Design of AlGaN P–I–N Detector

1. Band diagram of back illuminating AlGaN p–i–n heterostructures

The analysis of back illuminating AlGaN p–i–n detectors usually adopts the Anderson model. In experiments, the i-AlGaN layers are weakly n-doped, and therefore, a back-illuminated AlGaN p–i–n detector actually has a p–n–N^+ structure (Here "N" means a layer with higher Al composition, i.e., the p-AlGaN and the i-AlGaN form a p–n homojunction, while the i-AlGaN and the N^+-AlGaN window layer form an n-N^+ heterojunction). Since it is difficult to obtain an analytical solution of the space charge contribution for an iso-type heterojunction and its barrier height is much smaller than that of a p–n junction, the n-N^+ barrier will be ignored for further analysis. Based on the above assumption, the structure of a back-illuminated detector is similar to a front-illuminated GaN p–i–n detector. In order to include the factor of Al composition and apply to a general case, we employ formulas for heterojunction structures. It is obvious that when the detectors have an identical Al composition in all layers, the above formulas will be the same as that of the homojunction devices.

Two sides in Fig. 6.1, the band diagram of two semiconductor materials with different band gaps is shown before they reach thermal equilibrium,

With an external bias, the band diagram of an abrupt p–n junction is shown in Fig. 6.2.

For a p–i AlGaN heterojunction (Here the band gap of i-region is wider than that of p-region, p corresponds to "1" in Fig. 6.3 and i to "2"). The width of the space charge region at the p-side X_p and i-side X_i can be expressed, respectively, as follows,

$$X_p = \left[\frac{2\varepsilon_p\varepsilon_i N_i}{qN_a(\varepsilon_p N_a + \varepsilon_i N_i)}(V_D - V_a)\right]^{1/2} \tag{6.2}$$

$$X_i = \left[\frac{2\varepsilon_p\varepsilon_i N_a}{qN_i(\varepsilon_p N_a + \varepsilon_i N_i)}(V_D - V_a)\right]^{1/2} \tag{6.3}$$

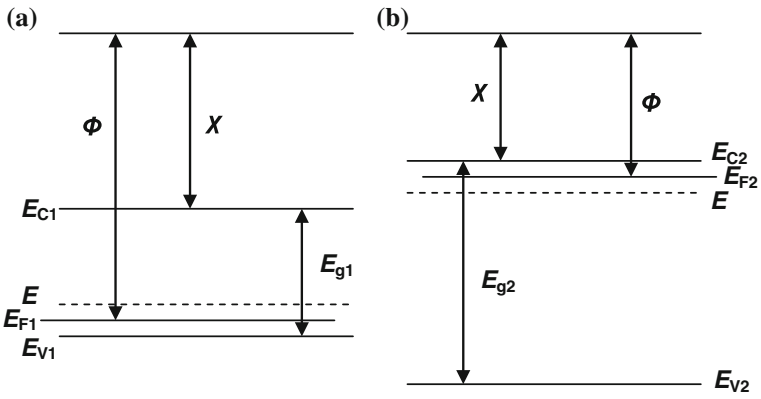


Fig. 6.1 Band diagram of two semiconductor materials with different band gaps, **a** semiconductor material with band gap E_{g1}, **b** semiconductor material with band gap E_{g2}

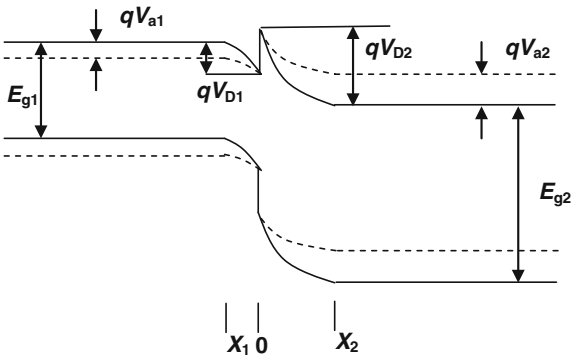


Fig. 6.2 Band diagram of an ideal abrupt p–n with external bias

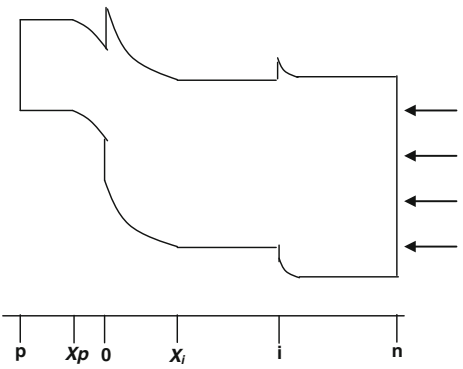


Fig. 6.3 Band diagram of an abrupt AlGaN p–i–n heterostructure

Where V_D is the built-in potential, V_a external bias, q electron charge, ε_p and ε_i, respectively, the dielectric constants of the p-AlGaN and i-AlGaN, N_a and N_i, respectively, the doping concentration of p and i-AlGaN.

$$V_D = \Phi_p - \Phi_i = \chi_p - \chi_i + E_{gp} - E_{Fp} - E_{Fi} \tag{6.4}$$

where Φ is the material working function, χ electron affinity potential, E_g band gap, and E_{Fp} (relative to valence band top) and E_{Fi} (relative to conduction band bottom) Fermi energy levels. The ratios of the built-in potentials and the external biases on the p-AlGaN and i-AlGaN can be expressed as

$$\frac{V_{D1}}{V_{D2}} = \frac{V_{a1}}{V_{a2}} = \frac{\varepsilon_i N_i}{\varepsilon_p N_a} \tag{6.5}$$

2. Internal quantum efficiency of AlGaN p–i–n heterojunction

The band diagram of an AlGaN abrupt p–i–n heterojunction is shown in Fig. 6.3, where *p*, *i,* and *n* represent, respectively, the thickness of p-, i-, and n-AlGaN layers. Based on the assumption and analysis in other works and ignoring the barrier of the i–N junction, with incident light of energy higher than the band gap of the N-AlGaN window layer, the window layer can contribute diffusion currents but not drift currents.

The internal quantum efficiency (QE) of a p–i–n structure consists of the following components.

Drifting QE in p-AlGaN

$$QE11(\lambda) = \exp(-\alpha_3 N) \cdot \exp(-\alpha_2 \mathrm{i}) \cdot (1 - \exp(-\alpha_1 X_p)) \tag{6.6}$$

Diffusion QE in p-AlGaN

$$QE12(\lambda) = \frac{\alpha_1 L_{np}}{1+\alpha_1 L_{np}} \exp(-\alpha_3 N) \cdot \exp(-\alpha_2 \mathrm{i}) \cdot \exp(-\alpha_1 X_p) \tag{6.7}$$

Drifting QE in i-AlGaN

$$QE21(\lambda) = \exp(-\alpha_3 N) \cdot \exp(-\alpha_2 \mathrm{i}) \cdot (\exp(\alpha_2 X_i) - 1) \tag{6.8}$$

Diffusion QE in i-AlGaN

$$QE22(\lambda) = \exp(-\alpha_3 N) \cdot \frac{\alpha_2 L_{pi}}{1-(\alpha_2 L_{pi})^2} \left[\begin{array}{l} \left(\alpha_2 L_{pi} + \sinh\left(\frac{i-X_i}{L_{pi}}\right) / \cosh\left(\frac{i-X_i}{L_{pi}}\right)\right) \exp(-\alpha_2 (i-X_i)) \\ -1/\cosh\left(\frac{i-X_i}{L_{pi}}\right) \alpha_2 L_{pi} \end{array} \right] \tag{6.9}$$

Diffusion QE in n-AlGaN

$$f1 = \sinh\left(\frac{i+N-X_i}{L_{pN}}\right) + \frac{SL_{pN}}{D_{pN}} \cosh\left(\frac{i+N-X_i}{L_{pN}}\right) \tag{6.10}$$

$$f2 = \cosh\left(\frac{i+N-X_i}{L_{pN}}\right) + \frac{SL_{pN}}{D_{pN}} \sinh\left(\frac{i+N-X_i}{L_{pN}}\right) \tag{6.11}$$

$$QE3(\lambda) = \frac{\alpha_3 L_{pN}}{1-(\alpha_3 L_{pN})^2} \left[\begin{array}{l} (\alpha_3 L_{pN} + f1/f2) \exp(-\alpha_3 (i+N-X_i)) \\ -\frac{1}{f2}\left(\frac{SL_{pN}}{D_{pN}} + \alpha_3 L_{pN}\right) \end{array} \right] \tag{6.12}$$

The overall quantum efficiency of the p–i–n device is

$$QE(\lambda) = QE11(\lambda) + QE12(\lambda) + QE21(\lambda) + QE22(\lambda) + QE3(\lambda) \tag{6.13}$$

Therefore, the responsivity of the device is,

$$R(\lambda) = (1-R) \cdot QE(\lambda) \cdot \frac{q\lambda}{hc} \tag{6.14}$$

λ is the wavelength of incident light and R the reflectivity at the sapphire/air surface. α_1, α_2, and α_3 are the absorption coefficients in, respectively, the p-, i- and n-AlGaN layers, and L_{np}, L_{pi}, and L_{pN} are the diffusion length in each layer. D_{pN} is the hole diffusion coefficient in the n-AlGaN and S the recombination rate at the interface of n-AlGaN/Sapphire substrate. h is Plank constant and c speed of light in vacuum.

Based on the basic principle of AlGaN p–i–n heterojunction detectors, the Al composition and the optical constants in the $Al_xGa_{1-x}N$ materials for the two operation windows, 300–365 nm visible-blind and 250–280 nm solar-blind can be determined. Then, based on the formulas of the key parameters, we can theoretically simulate the effects of composition, thickness, and doping concentration of each layer on the quantum efficiency, responsivity, and response time in an AlGaN p–i–n UV detector. In our simulation, defect-assisted carrier recombination is not considered. Table 6.1 lists the parameters in simulation.

Figures 6.4, 6.5 and 6.6 show the simulated dependence of the quantum efficiency and responsivity on the thickness of, respectively, n-, i-, and p-AlGaN layers of a back illuminating AlGaN p–i–n detector in the wavelength band of 250–280 nm. One can see that the optimal thicknesses of p-, i-, and n-AlGaN are

Table 6.1 Parameters used in simulation

Parameter	Value	Remark	Parameter	Value	Remark
S	10 cm/s	Reference [14]	N_C(GaN)	$4.3 \times 10^{14} \times T^{3/2}$ cm^{-3}	Reference [17]
T	300 K	Room temp.	N_V(GaN)	$8.9 \times 10^{15} \times T^{3/2}$ cm^{-3}	Reference [17]
c	2.998×10^8 m/s	Speed of light in vacuum	ε(GaN)	9.5 ε_0	Reference [18]
h	6.626×10^{-34} J·s	Plank's constant	ε(AlN)	8.5 ε_0	Reference [17]
k	1.381×10^{-23} J/K	Boltzmann's constant	m_n(GaN)	0.2 m_0	Reference [17]
n		Reference [15]	m_n(AlN)	0.4 m_0	Reference [17]
q	1.6×10^{-19} C	Magnitude of electronic charge	m_p(GaN)	1.5 m_0	Reference [17]
t_{pN}	10 ns	Reference [14]	m_p(AlN)	7.26 m_0	Reference [17]
α		Reference [16]	χ(GaN)	4.1 eV	Reference [17]
E_g(GaN)	3.4 eV	Reference [17]	χ(AlN)	0.6 eV	Reference [17]
E_g(AlN)	6.2 eV	Reference [17]	b	1.0 eV	Reference [12]

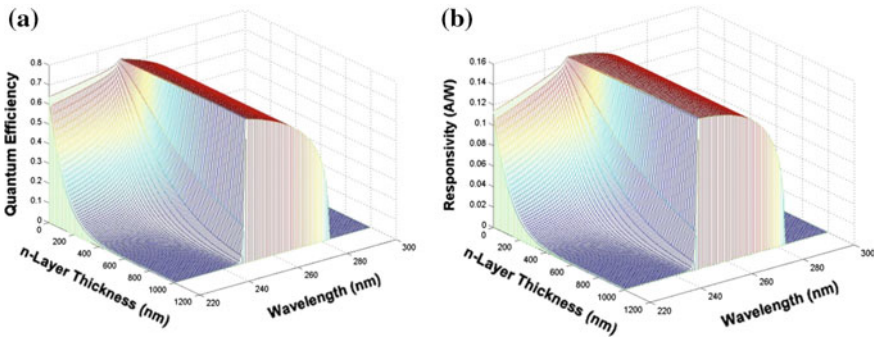


Fig. 6.4 Simulated dependence of **a** quantum efficiency and **b** responsivity on pn-AlGaN thickness

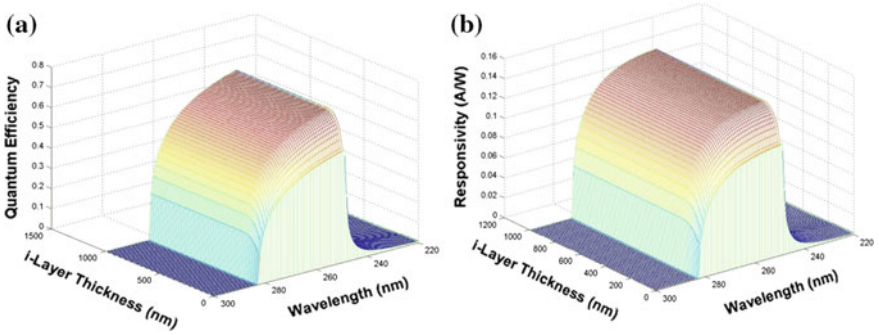


Fig. 6.5 Simulated dependence of **a** quantum efficiency and **b** responsivity on i-AlGaN thickness

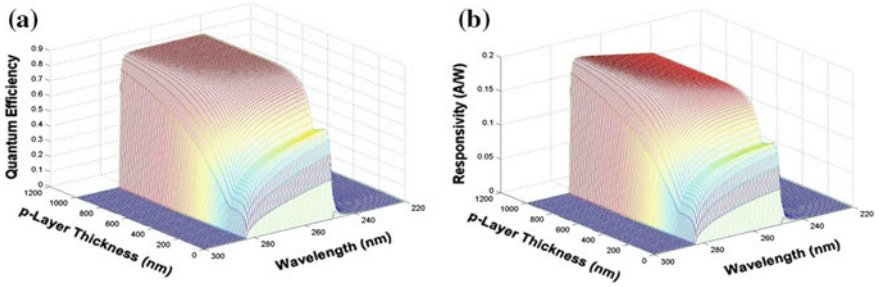


Fig. 6.6 Simulated dependence of **a** quantum efficiency and **b** responsivity on p-AlGaN thickness

about, respectively, 300, 200 and 500 nm. If the n-AlGaN is to be used as a filtering layer its thickness should be at least 500 nm.

The effect of the carrier concentration in the p- and i- region on the quantum efficiency and the response time can be obtained by analyzing the energy band. Figure 6.7 shows the energy band and electric field distribution, by the theoretical calculation, of a solar-blind (250–280 nm) back-illuminated heterojunction AlGaN p–i–n detector with carrier densities of 1×10^{17} cm^3 in the p and n region. Since the depletion layer extends toward the p-region, a built-in electric field does not exist in the area close to the substrate of the i-region, where photon-generated carriers transport through diffusion but not drifting. This mechanism does not favor quantum efficiency and response time in the back-illuminated devices. Therefore, it is important to increase the hole density in the p-region and decrease the electron density of the i-region.

The dependence of the response time on the p-layer concentration and specific contact resistance can be also obtained. The resistance-capacitor product (RC) is the main parameter to determine the response time. The p-layer carrier concentration in the AlGaN is usually at the range of $(1–3) \times 10^{17}$ cm^3, which results in high bulk resistance. There also exists large contact resistivity between the p-layer and the metal. These are the main factors to restrict the response speed. Figure 6.8 shows the

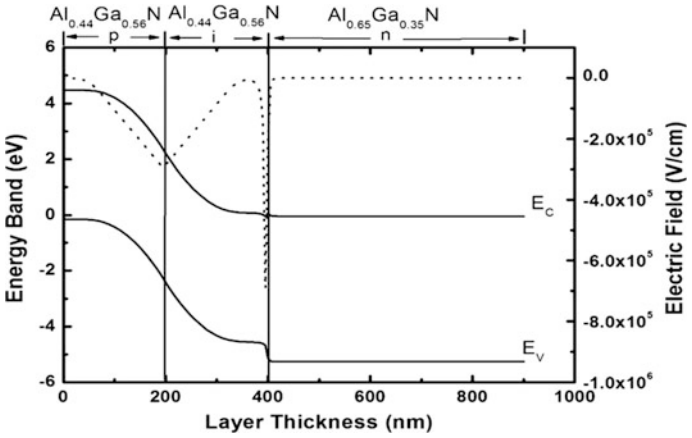


Fig. 6.7 Energy band and electric field distribution of a solar-blind (250–280 nm) back-illuminated heterojunction AlGaN p–i–n detector

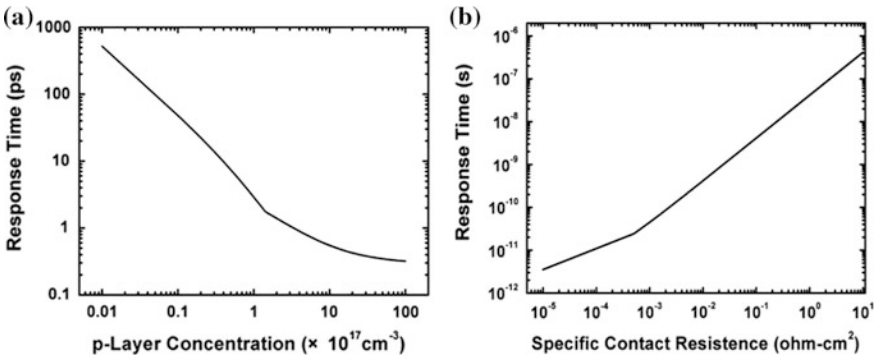


Fig. 6.8 Simulated relationship of response time with **a** the p-layer density and **b** contact resistivity of a back-illuminated p–i–n GaN detector within the wavelength from 300 to 365 nm

simulated relationship of the response time with, respectively, the p-layer doping density (Fig. 6.8a) and contact resistivity (Fig. 6.8b) for the back-illuminated p–i–n GaN detector within the wavelength range of 300 to 365 nm.

Because there are strong polarization electric fields at the GaN-based heterojunction interfaces, polarization will affect the detector's response spectrum [19, 20]. Figure 6.9a is the calculated energy band diagram and electric field distribution in a front-illuminated (wavelength range: 300–365 nm) GaN-based p–i–n heterojunction detector under 100 % polarization, and (b) photo response spectra under a different polarized situation. The results show that gradient composition interfaces can reduce the effect of polarization on the response spectrum. However,

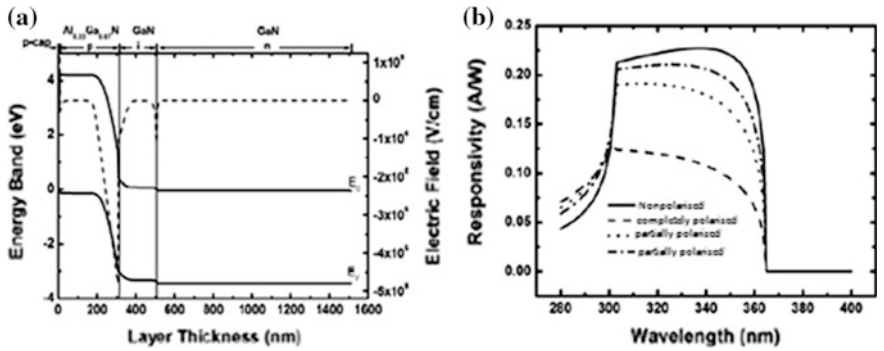


Fig. 6.9 **a** Simulated energy band of a front-illuminating 300–365 nm GaN p–i–n heterojunction detector under 100 % polarization and **b** photo-response spectra under different polarized situation

polarization has little effects on the responsivity of the back-illuminated heterojunction p–i–n detector and can be ignored.

3. Optimal designs of back- and front-illuminated p–i–n detector structures

p–i–n is the most used structure for depletion-type photodetectors. Its response wavelength and frequency is optimized under the condition that the intrinsic layer thickness is equivalent to the depletion layer width under the reverse bias. The intrinsic layer has the following functions,

1. Compared with the n and p regions, the intrinsic layer has high resistivity. Therefore, the reverse bias is imposed on the intrinsic layer. This increases the depletion layer width, extends the effective volume for photoelectronic conversion and improves the device sensitivity.
2. For the conventional diffusion-type p–n photodiodes, it is essential to use highly resistive materials to increase the breakdown voltage, which results in high series resistance and RC constant. On contrary, by introduction of the intrinsic layer, the breakdown voltage in a p–i–n detector is no longer restrained by the resistivity of the base material. This allows a significant reduction in the series resistance and time constant but still maintains a high breakdown voltage by using low resistivity base materials.
3. The photoelectronic conversion mainly takes place in the intrinsic layer and a region in the n- or p-layer within one diffusion length from the intrinsic layer. Since strong electric fields in the intrinsic layer can accelerate minority carrier transportation, properly increasing the intrinsic layer width hardly affects the transition time.
4. p–i–n photodiodes usually operate under zero bias or high reverse voltage (avalanche photodiode). The depletion layer is much wider than that of the conventional diffusion-type p–n photodiodes resulting in smaller capacitance and higher response speed.

5. As the thickness of the intrinsic layer increases, the drift current of photogenerated carriers in the depletion region increases as well. However, a too thick intrinsic layer will affect the transition time and the response frequency. The optimal depletion layer width (intrinsic layer thickness) is chosen in such way that the transition time equals half of the modulation frequency.

In design, optimization and fabrication of UV photodetectors, attention must be paid to the following items,

1. Reducing the junction depth so that the p–i–n junction is closer to illumination face. This can reduce the recombination loss of photogenerated carriers during the diffusion toward depletion region of the p–n junction.
2. Selecting proper dopants and surface concentration to avoid stress and dislocations induced from impurities diffusion. Removing the very short lifetime 'dead-layers' near surface caused by diffusion.
3. Improving the surface properties and reducing the surface recombination rate through passivation and other technologies.
4. Optimizing the interface region between the p–i–n structure and the substrate to release stress and decrease dislocation density as much as possible. Enhance the optoelectronic performances of the p–i–n structures by releasing stress and improving material quality.

During optimization, it is necessary, taking the specific issues of material growth and device process into account, to study the relationships of the detector performance with device structures, material defects, and processing conditions. This helps to develop a more practical design of the front-and back-illuminated p–i–n detector structures in both visible-blind and solar-blind spectra. Through feedback from experimental results, the theoretical model can be further revised. In solar-blind p–i–n detectors (including the RCE detector), the key factors that affect quantum efficiency are defects due to high Al composition in the $Al_xGa_{1-x}N$, accurate control of Al composition and layer thickness, p-type doping and its activation, p-and n-type Ohmic contact resistivity, and so on. Figure 6.10 shows the typical structures of a front-(a) and a back-illuminated (b) solar-blind AlGaN heterojunction p–i–n UV detector. Figure 6.11 presents the theoretical quantum efficiency (a) and response spectrum (b) of a back-illuminated detector. The highest theoretical response value 0.18 A/W can be observed.

6.3 AlGaN Resonant-Cavity-Enhanced UV Detectors

Because a conventional AlGaN UV detector usually does not possess in a single device all the necessary characteristics, such as high-quantum efficiency, suitable bandwidth, fast response, and insensitivity to light with wavelength outside of the working band, it may have difficulty meeting some practical application requirements. This partly limits the development of UV detection technology. Fortunately,

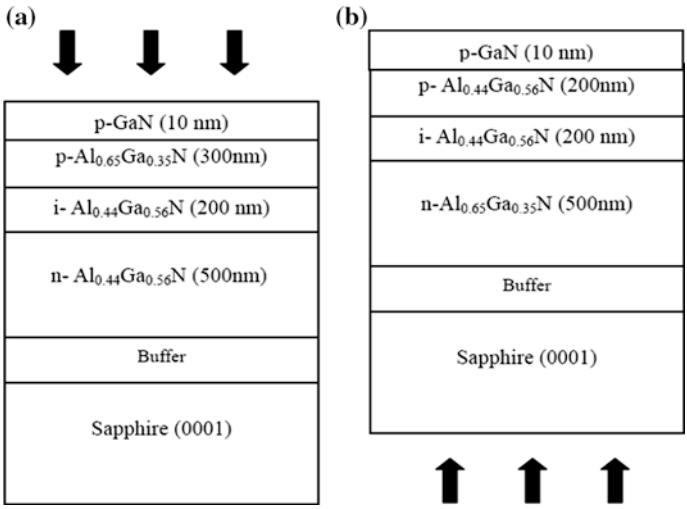


Fig. 6.10 Typical structures of solar-blind front- (**a**) and back-illuminated (**b**) AlGaN heterojunction p–i–n UV detector

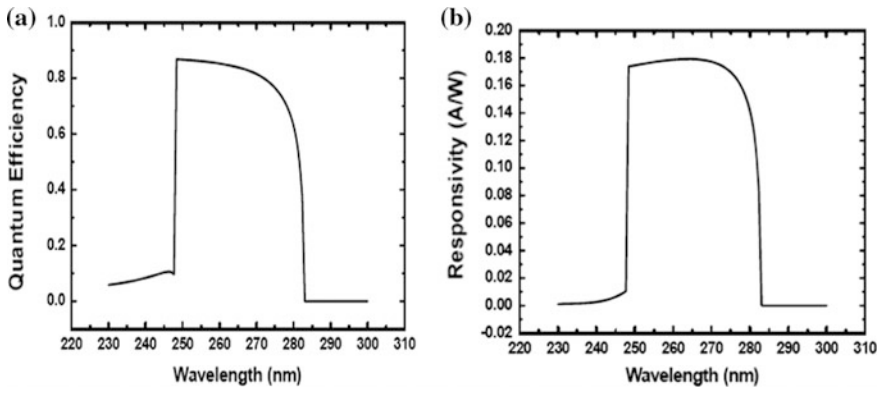


Fig. 6.11 Theoretical quantum efficiency (**a**) and response spectrum (**b**) of a solar-blind back-illuminated AlGaN heterojunction p–i–n detector

through hard work of many researchers, a new type of device, the resonant-cavity-enhanced (RCE) photodetector associated with its theoretical basis, was invented. The RCE photodetector was realized after the break through of the Fabry–Perot resonant-cavity epitaxy technology.

In this section, we first briefly introduce the basic structure of a RCE UV detector. Then, we describe the design and fabrication of GaN-based distributed Bragg reflectors (DBR) as well as single-and dual-color RCE UV detectors.

6.3.1 The Basic Structure of Resonant-Cavity-Enhanced UV Detectors

Kishino et al. [21] from Department of Electrical Science and Engineering of Sophia University, Japan started to study the AlGaN-based RCE metal-semiconductor-metal (MSM) detector in 2000. The device had a back-illuminated structure. The bottom reflector was made of six pairs of high reflectivity SiO_2/ZrO_2 DBR and the top mirror eight pairs of $Al_{0.14}Ga_{0.86}N$ (34.4 nm)/AlN(40.5 nm) DBR. The GaN absorption layer thickness was about 40 nm and the cavity length about 0.33 μm. The responsivity spectrum has FWHM of 15 nm with a significant enhancement at the wavelength 353 nm.

Li et al. [22] from Department of Electronics and Computers of Texas University started to study GaN-based RCE p–i–n detectors in 2001. Their device was designed to work at a wavelength of 330 nm. The DBR peak reflectivity was measured to be 57 %. However, due to p-type Ohmic contact issues, the RCE p–i–n device was not finally realized. Detailed properties were not reported as well.

Biyikli et al. [23] from Department of Electrical Science and Engineering of Bilkent University, Turkey reported a GaN-based RCE Schottky detector in 2003. The bottom mirror was made of 20 pairs of $AlN/Al_{0.2}Ga_{0.8}N$ DBR. The peak responsivity at the wavelength of 337 and 350 nm is 0.153 and 0.046 A/W, respectively. The pulse width is shorter than 77 ps at the wavelength of 357 nm. The bandwidth at 3 dB is 780 MHz.

The structure of an RCE detector, as shown in Fig. 6.12 mainly consists of two parts. One is the reflectors on the two sides of the cavity which will realize the reflection resonance. Another part is the cavity composed of an absorption layer and an isolation layer to tune the cavity length. The active layer in the F–P resonant cavity has the function to absorb and detect light. RCE can be MSM, Schottky, p–i–n or other detector types.

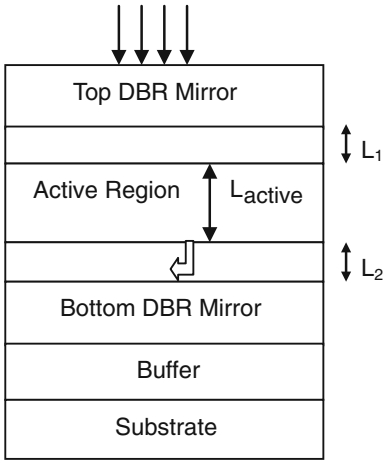


Fig. 6.12 The structure sketch of RCE detectors

In RCE detectors the active layers are inserted into the F–P cavity encompassed by the two DBRs. They can enhance the photoresponse of light with wavelength in resonance with the F–P cavity and suppress the response of light off-resonance. Therefore, they can achieve the function of wavelength-picking and spectrum-narrowing. Since light reflects back and forth many times in the cavity, high quantum efficiency can be attained even using a thin absorption layer, which reduces the transportation time of photogenerated carriers and improves the response speed and bandwidth-efficiency product. Additionally, the F–P resonant cavity can selectively respond to a predetermined narrow band due to its inherent property of frequency selection.

DBRs are the important components of a RCE detector. The materials to compose the DBRs could be GaN-based semiconductors, dielectric, or metal. If using GaN-based semiconductors the DBRs and the absorption layer can be grown successively within the same equipment. However, due to increasing difficulty in growing high Al composition GaN-based material, it is hard to fabricate high reflectivity reflectors. The dielectric materials systems have relatively mature manufacturing processes, and can be grown directly by plasma-enhanced chemical vapor deposition (PECVD). Two factors should be considered in choosing a UV DBR material, (1) a low absorption coefficient in the operation wavelength range from 250 to 365 nm, (2) easy preparation. The dielectrics that can meet the two requirements are SiO_2, HfO_2, Al_2O_3, MgF_2, CaF_2, ZrO_2, ITO, etc. Metal materials can select Al, Au, Ag, etc. A metal mirror can achieve a reflectivity of 90 % in UV band independent of wavelength. These metals are suitable for the top mirror in a Schottky-type RCE detector.

Design of GaN-based DBRs is in accord with the following processes, (1) determine the cutoff wavelength of the RCE UV detector (The target visible-blind wavelength is 330 nm, the target solar-blind wavelength 270 nm), (2) choose the material parameters and optical constants, such as absorption coefficient and refractive index, from the literature, or measure the parameters by experiment, (3) determine the Al composition in AlGaN. The structures can be calculated by transfer matrix method (TMM) and be grown by metal organic chemical vapor deposition (MOCVD). The structural and optical properties of the grown samples can be characterized by scanning electron microscopy (SEM), atomic force microscopy (AFM), transmission electron microscopy (TEM), high resolution X-ray diffraction (HRXRD), and optical reflection. By analyzing and comparing the differences between the measured reflection spectrum and the theoretical simulation, one can understand the factors that affect the DBR properties. The relationships of the DBR's center reflection wavelength, reflectivity at center wavelength and the reflection bandwidth with the refractive index, individual layer thickness, period number, interface roughness, stress, and defects can be extracted. Through such a process the DBR's theoretical model can be optimized.

Refractive index is a key parameter in design of the DBR layer thickness. The techniques of measuring refractive index are a prism coupler, transmission spectrum, reflection spectrum, and ellipsometry. Though the prism method is simple it can only measure the refractive index at a single wavelength and cannot obtain the

dispersion relation of the refractive index. The transmission spectrum can determine the refractive index by the interference fringes but can only apply to photon energies less than the material band gap ($E < E_g$). The reflection spectrum can not only determine the refractive index by the interference fringes, but also can determine the refractive index and the extinction coefficient by the measured spectral reflectivity and Kramers–Kronig relation. Ellipsometry obtains the refractive index through a fitting process. Xiaoli et al. [24], using the prism coupler method (the light source is a He–Ne laser with a wavelength of 632.8 nm), measured the ordinary refractive index in TE mode of an $Al_xGa_{1-x}N$ film grown on (0001) sapphire substrates by MOCVD. The measured refractive index of GaN, $Al_{0.26}Ga_{0.74}N$, and AlN are, respectively, 2.35, 2.27, and 2.04. They also calculated, through the transmission spectrum, the dispersions of refractive index of GaN in the energy range of $E < E_g$. Their results were consistent with the reported data from literature [25].

Through an optimization process between simulation and experiments [26–31] solved many issues in growth of multi-period $Al_xGa_{1-x}N$/AlN DBRs by MOCVD, such as stresses caused by lattice mismatch and thermal mismatch, selection of buffer layer material, precise control of Al composition, and layer thickness (<1 nm). They successfully grew visible-blind $Al_xGa_{1-x}N$/AlN DBRs with excellent repeatability and high reflectivity, and measured the refractive index of the DBR samples.

6.3.2 Design and Fabrication of RCE Ultraviolet Detectors

1. Design of RCE detectors

For a high-performance RCE detector, the absorption coefficient of the active layer in the operation wavelength range must be far bigger than that of the non-absorption region materials (the isolation layers and the reflecting mirror DBRs). The absorption coefficient of the absorption layer in the operation band should be appropriate ($1 \times 10^3\ cm^{-1} < \alpha < 5 \times 10^4\ cm^{-1}$). If the absorption coefficient is too small, the absorption layer thickness has to be thick, which will cause long carrier transportation time and degrade the device's speed performance. If the absorption coefficient is too large, the absorption layer has to be very thin, with which the standing wave effects on device performance will be significant. In design of multilayer mirrors, in order to avoid defects in the absorption region, one can choose two materials with good lattice-constant matching and larger refractive index differences to form the reflection mirror, and thus reduce the number of mirror layers, and thereby simplify the growth process and reduce device series resistance.

RCE detectors can be classified into single-color and dual-color RCE UV detectors according to the number of detecting wavelength bands in a single device.

Zheng et al. [32–34] have reported the designs of the two kinds of RCE UV detectors.

1. Design of single-color RCE UV detectors

(i) Design of the layer structures of RCE UV detectors

Single-color RCE UV detectors can be classified as Al-containing or Al-free detectors. At present, Al-containing RCE UV detectors are the most reported GaN-based RCE UV detectors [21–23]. The top reflection mirror of an Al-containing GaN-based RCE Schottky UV detector can be made by transparent electrodes or other mirror materials. The device structure is grown in the following sequence, an AlN buffer layer on a sapphire substrate, six pairs of AlN/$Al_{0.35}Ga_{0.65}N$ DBRs, an $Al_{0.35}Ga_{0.65}N$ isolation layer, and finally a GaN absorption layer, as shown in Fig. 6.13a. The Al-free RCE UV detector structure consists of an AlN buffer layer and a GaN absorption layer grown by MOCVD on two sides of the polished sapphire substrates. Then a SiO_2 isolation layer is grown using PECVD. Finally, the top and bottom reflection mirrors composed of periodic HfO_2/SiO_2 are grown, as shown in Fig. 6.13b.

(ii) Design of RCE UV detector

Single-color Al-containing GaN-based RCE UV detectors have different types, such as photoconductive, Schottky, and MSM structures. Figure 6.14 shows the profile and the top view of a single-color Al-free RCE photoconductive UV detector. Figure 6.15 shows the profile and top view of a single-color Al-free RCE Schottky UV detector structure. Figure 6.16 shows the profile view and top view of a single-color Al-free MSM RCE UV detector. The structures of single-color Al-containing GaN-based Schottky and p–i–n RCE UV detectors are shown in Fig. 6.17.

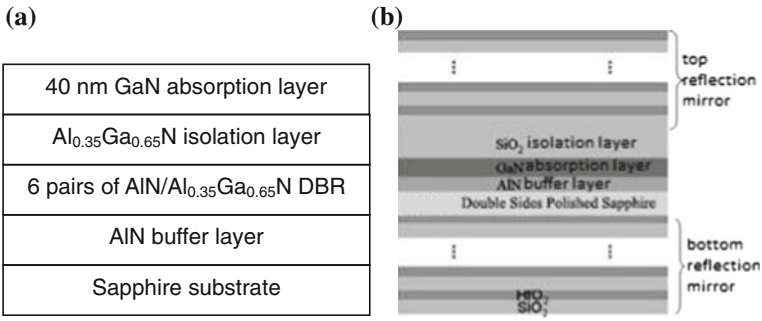


Fig. 6.13 Material structures of single-color **a** Al-containing **b** Al-free GaN-based RCE UV detectors

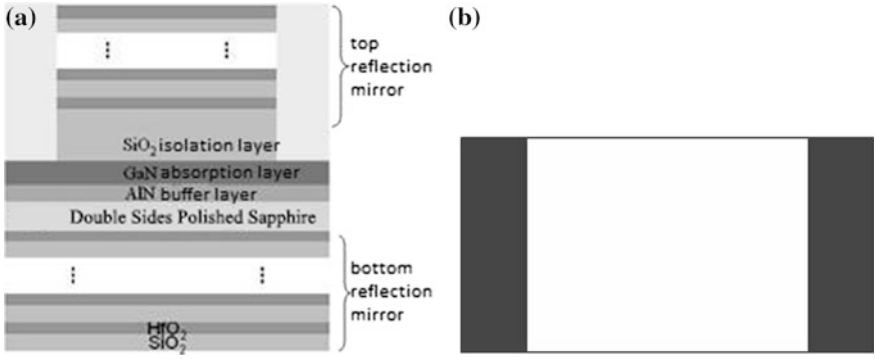


Fig. 6.14 **a** The profile and **b** top view of single-color Al-free RCE UV photoconductive detector

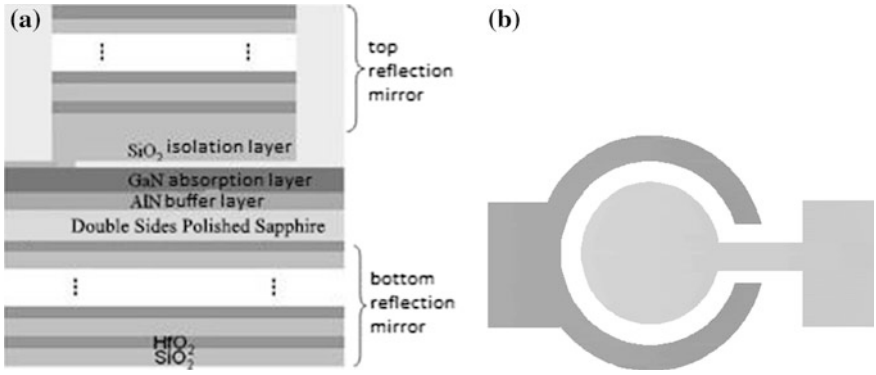


Fig. 6.15 **a** The profile and **b** top view of single-color Al-free Schottky RCE UV detector

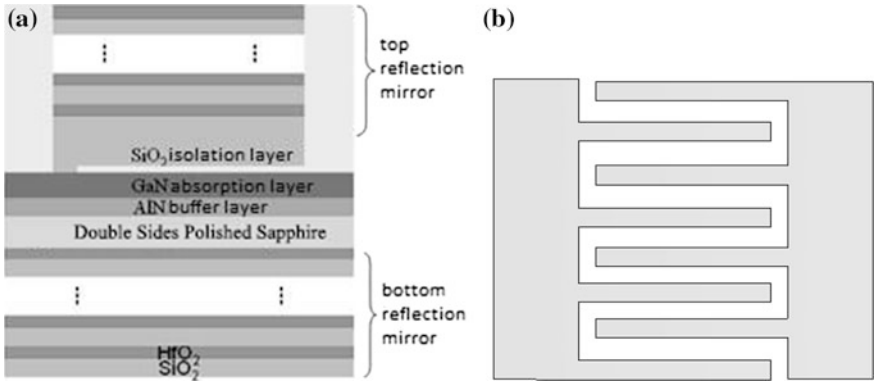


Fig. 6.16 **a** The profile and **b** top view of a single-color Al-free RCE UV MSM detector

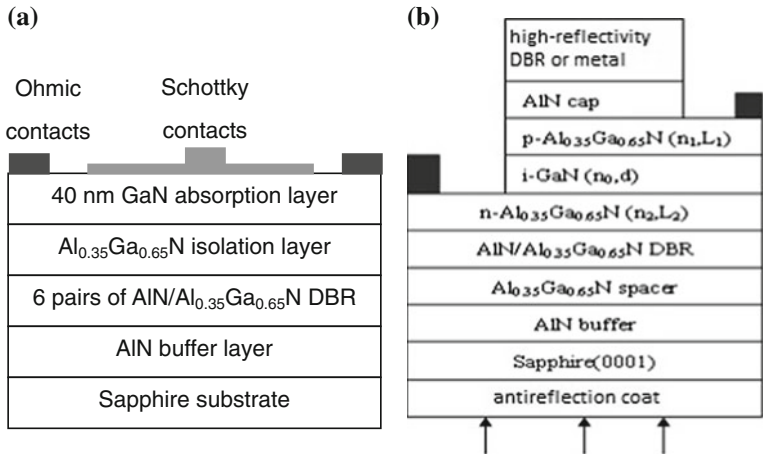


Fig. 6.17 **a** Schottky and **b** p–i–n single-color Al-containing GaN-based RCE UV detector

2. Design of dual-color RCE UV detectors

Although its design and fabrication are much difficult than that of a single-color RCE UV detector, a dual-color RCE UV detector can not only synchronously obtain in space two band information of a target but also suppress complex backgrounds, and thus can greatly enhance the efficiency of detection and the accuracy of target recognition. Therefore, it is very important in some special application fields.

The material structure of a dual-color RCE UV detector is shown in Fig. 6.18. The epitaxial growth is as follows. An AlN buffer layer and a GaN absorption layer are first grown on one side of a double-side polished sapphire substrate by MOCVD. An AlN buffer layer and $Al_{0.34}Ga_{0.66}N$ absorption layer are then grown on the other side of the substrate. Part of the area of GaN absorption layer is then etched away. After that, a SiO_2 isolation layer is grown by PECVD. Finally, the top and bottom reflection mirrors composed of repeated HfO_2/SiO_2 layers are grown.

Let us calculate the material parameters assuming the two detection bands are 320–365 nm and 250–280 nm. For the 320–365 nm band, GaN is used as the absorption material. According to the previous discussion, the reflectivity of the top and the bottom mirrors should be, respectively, R1 = 0.7 and R2 = 0.99, to achieve a quantum efficiency of 0.99. The absorption coefficient $\alpha_{GaN} = 1.25 \times 10^5\ cm^{-1}$ and the refractive index $n_{GaN} = 2.7$ at 340 nm are from the literature. From the equation $R_1 = R_2 e^{-2\alpha d}$, the theoretical thickness of the GaN absorption layer can be determined to be $\approx$ 14 nm. Similarly, for 250–280 nm band, $Al_{0.34}Ga_{0.66}N$ is used as the absorption material. The absorption coefficient of $Al_{0.34}Ga_{0.66}N$ at 265 nm is $1 \times 10^5\ cm^{-1}$. The reflectivity of the top and the bottom mirror should be, respectively, 0.7 and 0.99 to achieve a quantum efficiency of 0.99. From the equation $R_1 = R_2 e^{-2\alpha d}$, the theoretical thickness of the $Al_{0.34}Ga_{0.66}N$ absorption layer can be determined to be $\approx$ 17 nm.

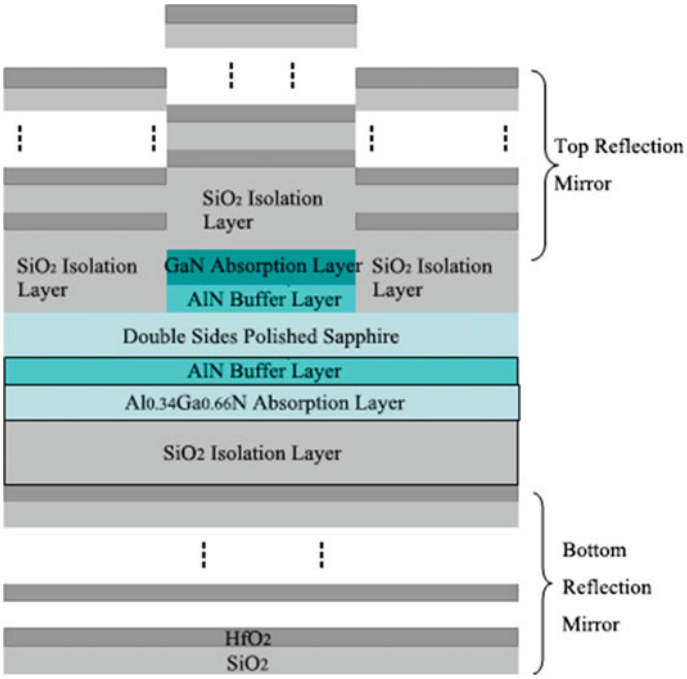


Fig. 6.18 Material structure of dual-color GaN-based RCE UV detector

Dual-color RCE UV detectors can also employ a photoconductive structure, a Schottky structure and a MSM structure. Figure 6.19 shows the profiles of (a) a photoconductive, (b) a MSM dual-color GaN-based RCE UV [32] detector.

A dual-color RCE ultraviolet detector operates as follows. Light penetrates into the structure from $Al_{0.34}Ga_{0.66}N$ layer side. Since the top and the bottom mirrors work in the wavelength band of 250–365 nm, when the incident wavelength is longer than 365 nm or shorter than 250 nm, the amounts of photogenerated carriers is negligible in the resonator. However, when the incident wavelength is in the range of 250–365 nm, it will be reflected back and forth through the absorption layers in the resonator by the top and bottom mirrors, until most of light is absorbed. Light in the wavelength range of 250–280 nm is absorbed by the $Al_{0.34}Ga_{0.66}N$ layer to produce strong photogenerated signals since the photosensitive area of the $Al_{0.34}Ga_{0.66}N$ layer is 8–10 times larger than that of the GaN absorption layer. It will also generate weak photo signals in the GaN layer because some light goes through the $Al_{034}Ga_{0.66}N$ and is absorbed in the GaN layer. But these signals are so weak they can be ignored. When the incident wavelength is in the range of 320–365 nm, almost all the light will pass through the $Al_{0.34}Ga_{0.66}N$ layer without absorption and be absorbed by the GaN layer. Therefore, the light does not generate photo signals in the $Al_{0.34}Ga_{0.66}N$ layer but produces strong photo signals in the GaN layer. When the incident light consists of wavelengths

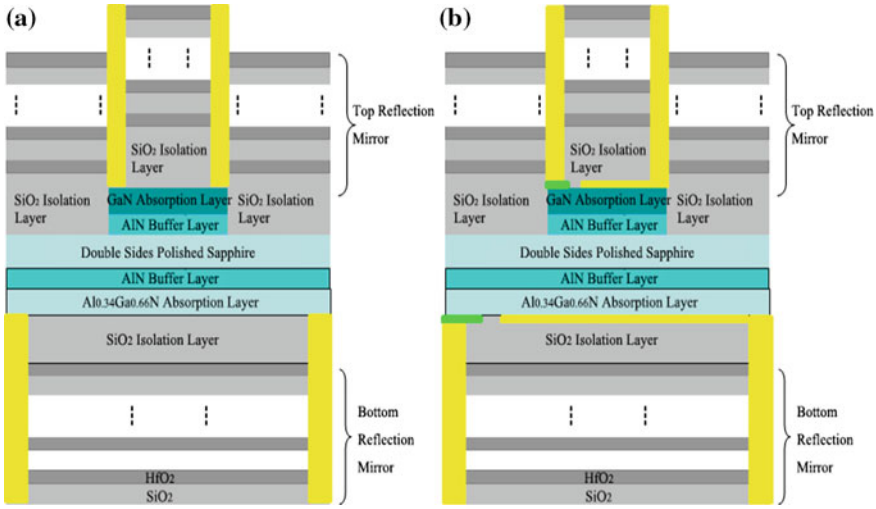


Fig. 6.19 Profiles of **a** a photoconductive, **b** a MSM dual-color GaN-based RCE UV detector

Table 6.2 Light with different wavelengths detected by dual-color RCE detector

Light wave band/nm	Signal at GaN layer	Signal at $Al_{0.34}Ga_{0.66}N$ layer
<250	No	No
250–280	Very weak	Very strong
250–280, 320–365	Very strong	Very strong
320–365	Very strong	No
>365	No	No

both in the range of 250–280 nm and of 320–365 nm, most of the light in the wavelength of 250–280 nm will be absorbed by the $Al_{0.34}Ga_{0.66}N$ layer and most of the light in wavelength of 320–365 nm will be absorbed by the GaN layer. Therefore, they will generate strong light signals on both sides. Table 6.2 summarizes the conditions of a dual-color RCE detector to detect and recognize light with different wavelength.

2. Experiments on single-color RCE UV detectors

In order to verify the designs, He et al. [32–35] reported experimental results of single-color RCE ultraviolet detectors. Starting from the basic principle and according to the targeting wavelength, they did a series of studies on RCE UV detectors, such as selection of the absorption and isolation layer materials, the effect of layer thickness on quantum efficiency and response time, the optimal position of the absorption layer in the RCE cavity by considering the standing wave effect. They also studied the effects of cavity length on quantum efficiency, spectral width,

and response line width as well as the number of the resonant modes. They grew a variety of RCE samples by MOCVD with MSM, p–i–n and Schottky structures for different targeting wavelength of 320 and 330 nm. They measured the samples' reflectance and response spectrum. By analyzing the deviation between measurements and the designs, they determined the key factors to achieve high-performance GaN UV detectors. In addition, they also initiated the material and device process studies of a novel 270 nm GaN RCE ultraviolet detector. The results were a theoretical design for the DBR and RCE detector structures with a central wavelength of 270 nm and fabricated solar-blind DBRs by MOCVD.

According to theoretical calculation, the quantum efficiency of a RCE detector shows periodical variation with thickness of the SiO_2 isolation layer. Hence, the thickness of SiO_2 should be strictly controlled in order to maintain a high-quantum efficiency of the RCE detector.

Figure 6.20a is the design diagram of a GaN-based p–i–n RCE detector with targeting wavelength of 320 nm (bottom mirror consists of 40.5 pairs of $Al_{0.3}Ga_{0.7}N$/AlN DBRs and top mirror is the air/GaN interface). Figure 6.20b is the measured response spectrum with the central peak at 315 nm, a slight deviation from the design wavelength. Figure 6.20c is the simulated reflectance and absorption spectrum. By comparing Fig. 6.20b and c, the response and absorption spectrum coincide well to each other.

Figure 6.21(a) shows the layer structure of a GaN-based p–i–n RCE detector with targeting wavelength of 330 nm, (b) a SEM picture of the device in fabrication, and (c) the response spectrum. The response spectrum shows the peak wavelength of 329 nm, the line width 5.5 nm and the responsivity 6.3 mA/W (under the 0 V bias). The response spectrum shows the peak wavelength of 329 nm, the line width 5.5 nm and the responsivity 6.3 mA/W (under the 0 V bias).

Figure 6.22(a) shows the layer structure of a GaN-based Schottky mesa type RCE detector employing an Al top mirror with a targeting wavelength of 330 nm, (b) a SEM picture of the device, and (c) the response spectrum. The spectrum shows the response central peak at 331 nm, responsivity of 24 mA/W (under 0 V bias). The measured peak response wavelength precisely coincides with the targeting wavelength of 330 nm. The spectrum shows the response central peak at 331 nm, responsivity of 24 mA/W (under 0 V bias).

Figure 6.23 shows the response spectrum of a MSM type GaN-based RCE detector with targeting wavelength of 330 nm. The detector has a short cavity. Ni/Au is employed as the metal electrodes. The spectrum shows the response peak at 329 nm, responsivity of 4.3 A/W (under 6 V bias). This type of all-nitrides MSM RCE detectors is difficult to be fabricated. The response peak is 329 nm, peak responsivity 4.3 A/W (under 6 V bias).

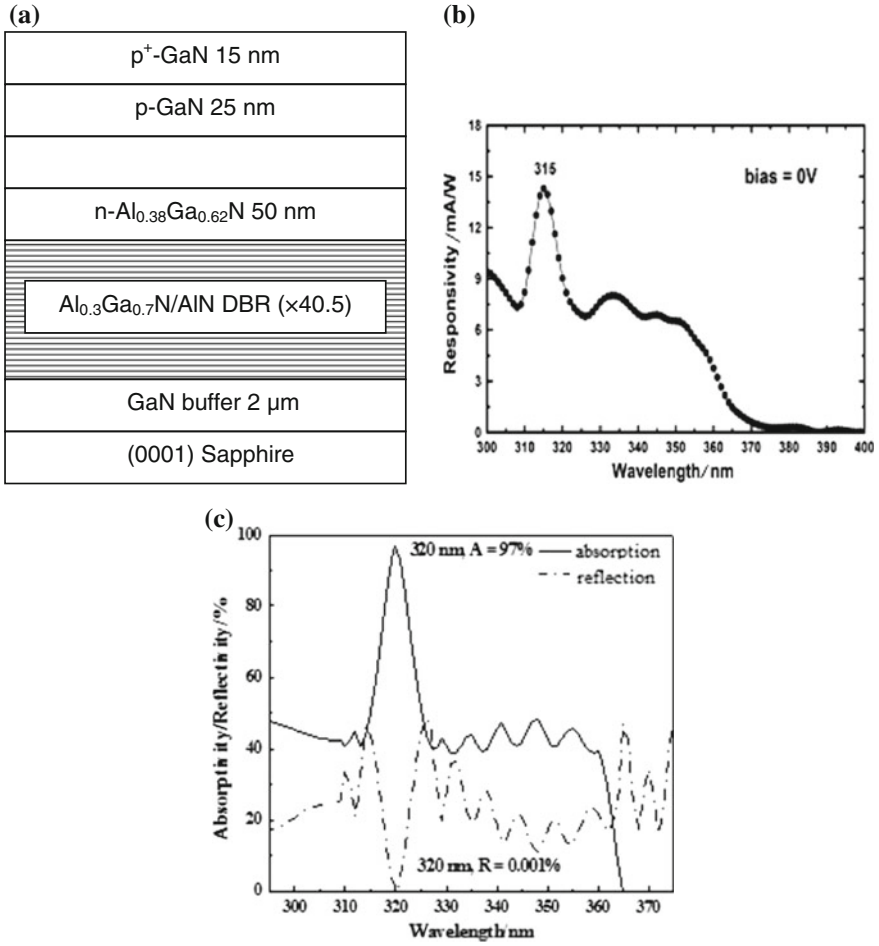


Fig. 6.20 RCE detector structure sketch diagram of target wavelength 320 nm GaN Base (**a**) measuring response spectrum (**b**) and light reflection spectrum and light absorption spectrum of the theoretical simulation of the detector structure (**c**)

6.4 AlGaN Detector Chip Fabrication

Devices of certain novel semiconductor develop and bloom, always initially originated from the maturity of the semiconductor material and doping (n-type and p-type) technology. It is also true for a wide band gap GaN device. In the 1990s, Shuji Nakamura successfully demonstrated high-quality GaN film growth technology by using a two-way gas configuration and by a low temperature buffer layer method [36]. Professor Amano demonstrated the p-type dopant active technology using the e-beam irradiation method, while Nakamura used rapid thermal annealing

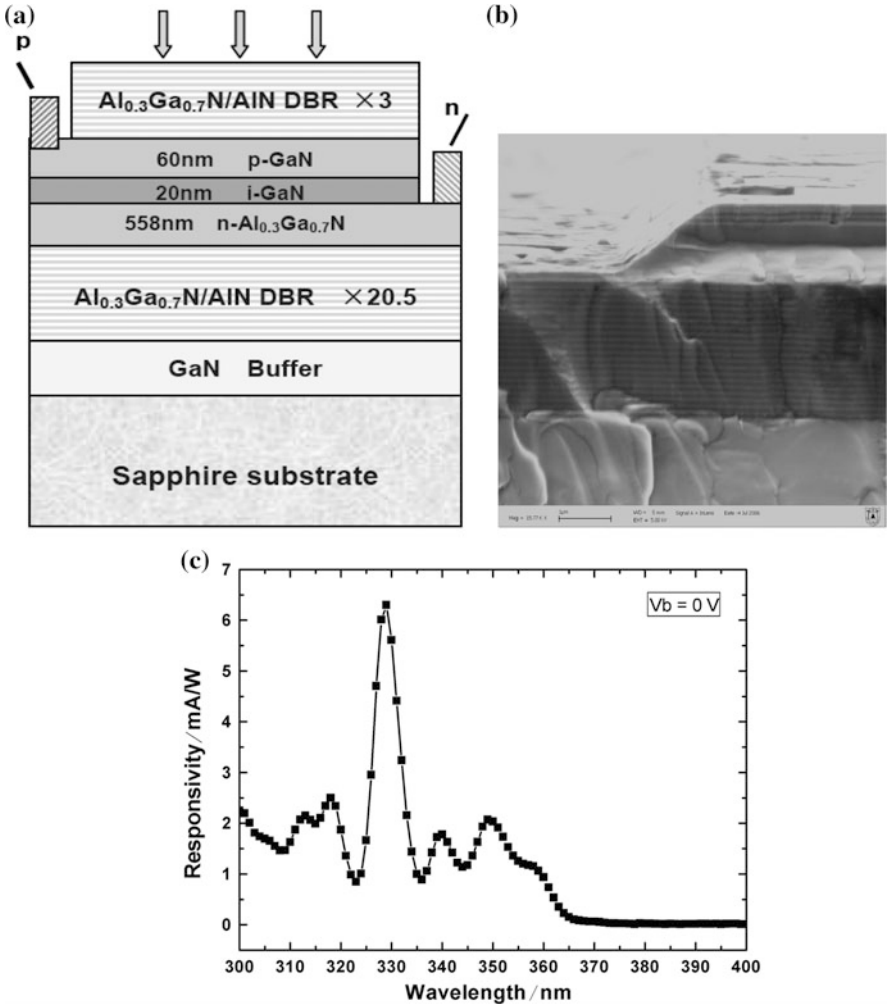


Fig. 6.21 **a** Layer structure of a GaN-based p-i-n RCE detector with targeting wavelength of 330 nm, **b** a SEM picture of the device in fabrication and **c** the response spectrum

in N_2 atmosphere [37, 38]. Thus, the solid foundation formed for III-nitride material and devices. Since then, the device fabrication and process study has become the main focus in the research of various devices.

In this section, AlGaN detector chip fabrication technology is discussed, including mesa technology, passivation and Ohmic contact. In mesa technology study, dry etching (mainly using inductively coupled plasma), and wet chemical treatment are discussed. The SiO_2 passivation and n-type, p-type AlGaN Ohmic contact are also discussed.

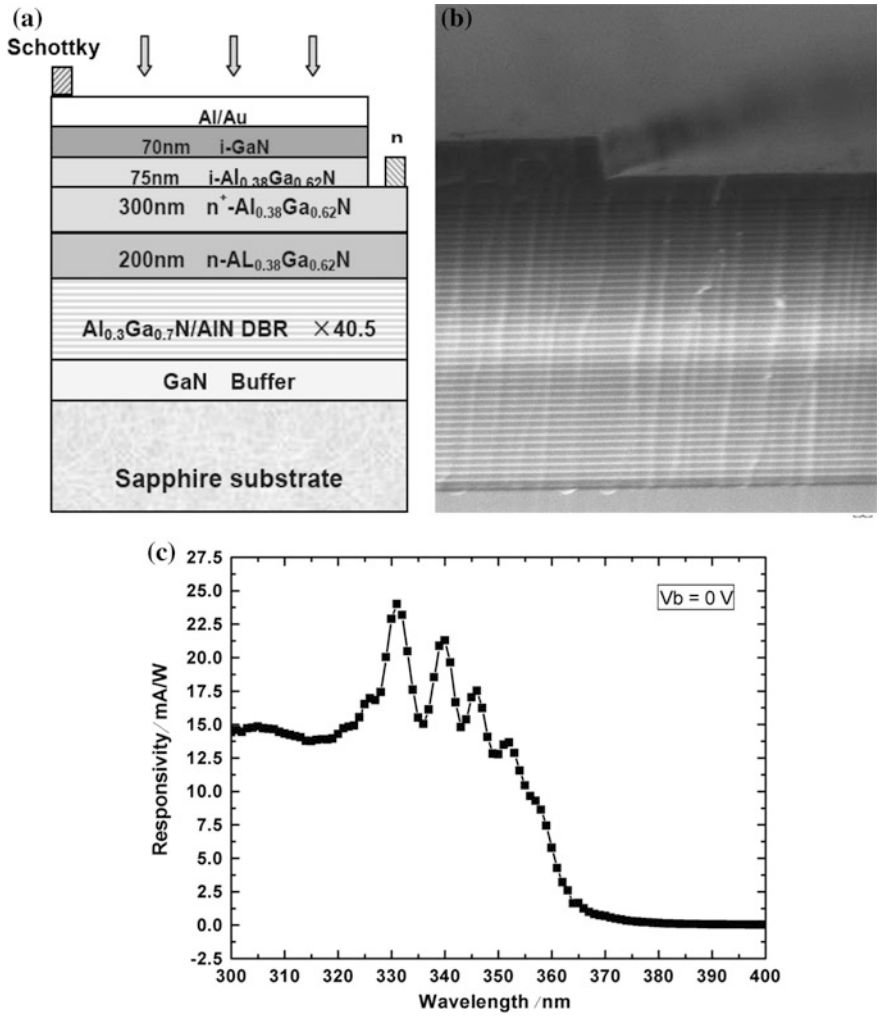


Fig. 6.22 **a** Layer structure of a GaN-based Schottky mesa type RCE detector employing an Al top mirror with a targeting wavelength of 330 nm, **b** a SEM picture of the device and **c** the response spectrum

6.4.1 Mesa Formation Technology

GaN-based material has the advantage of being stable to chemical corrosion, because of its high bond energy. Compared with other semiconductor material, III-nitride wide band gap materials have much stronger chemical bonds. The bond energies of GaN, InN, and AlN were 8.92, 7.72, and 11.52 eV/atom, respectively. While for GaAs it is 6.52 eV/atom. This fact gives III-nitride many special features,

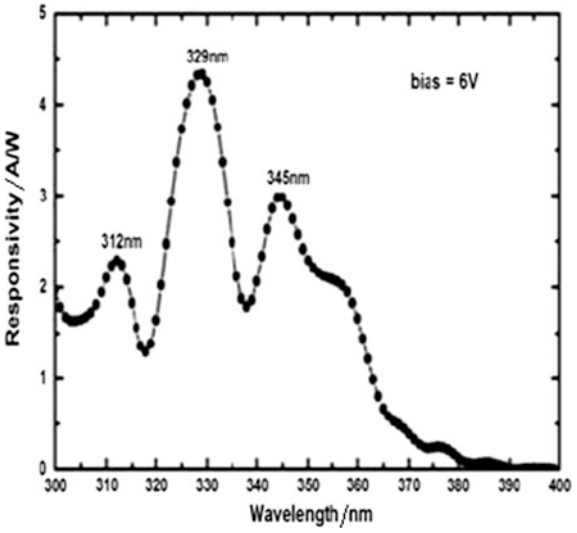


Fig. 6.23 Response spectrum of a MSM GaN-based RCE detector with targeting wavelength 330 nm

such as wide band gap, good thermal stability, and suitability for high-temperature work. It is used for optoelectronic devices from UV to yellow light and high-temperature, high-frequency, and high-power electronic devices [39]. But, on the other hand, it also leads to the requirement of much higher activation energy in etch reaction compared with other semiconductors. This made III-nitride dry etching and chemical etching difficult at room temperature, when all known acids etch GaN very slowly. Most alkali also etches GaN very slowly. Electrochemical etch or photo-enhanced chemical etch may give larger etch rates. But chemical wet etching process is low precision for pattern transfer. The surface and side walls are rough. Wet etching process is difficult to control. These issues made it difficult to use in practical fabrication. Etching processes with low damage and smooth side walls are desired in the mesa formation process of high-performance AlGaN devices, especially in large format high-performance GaN-based UVFPA devices.

1. Inductively coupled plasma etching

In recent years, inductively coupled plasma (ICP) etching, which is high density plasma etch technology, is under very intensive research. It is widely used in the III-nitride etch process. Compared to reaction ion etch (RIE), ICP has many advantages, especially for the ability of separate controlling of plasma density and energy. Plasma density can be achieved with 2–4 orders of magnitude higher density than that of RIE, which makes it easier to break the chemical bonds between III elements and N atoms by effective bombardment of the surface. Nitride material etch rate is enhanced by ICP. At the same time, ion energy is relatively low and adjustable (at about 20 eV). One can obtain anisotropic and low damage etched

surfaces with a high etch rate. ICP can maintain a stable glow discharge in low pressure. It guarantees the ion bombardment uniformity and collimation.

Shul et al. [40, 41] studied the GaN etch using ICP and $Cl_2/H_2/Ar$ gases. A high etch rate could be achieved to 6875 nm/min. Very good results were demonstrated with ICP and Cl_2/Ar gases in etching III-nitride. A smooth side wall and a roughness of less than 20 nm of the etched surface were obtained. It is believed that the process is suitable for GaN/InGaN/AlGaN heterojunction laser diode vessel fabrication [42].

ICP etching of III-nitride has advantages of high etch rate, anisotropy, good aspect ratio, low etching damage, large area of good uniformity, etch profile controllability and smooth etched surface, and so on. So it is widely used in III-nitride devices of large-scale production and manufacturing [43–45]. Studies for the gases recipe choice, different gas ratios, ICP power, and DC bias on ICP etching of AlGaN were carried out. By using scanning electron microscopy (SEM) and atomic force microscopy (AFM), etching morphology was characterized. By using X-ray photoelectron spectroscopy (XPS), the chemical changes were studied on the material surface before and after etching. The transmission line model (TLM) method was also used for material sheet resistance change characterization which may be associated with damage. While rapid thermal annealing was also studied to remove the etching damage.

In our ICP etching experiment, n-$Al_{0.33}Ga_{0.67}N$, n-$Al_{0.47}Ga_{0.53}N$, n-$Al_{0.45}Ga_{0.55}N$ samples are mainly used, grown by the MOCVD method with double-sided polishing of 2 in (0001) orientation of sapphire substrates. Samples were grown with a buffer layer of AlN and the epitaxial layer thickness is about 1 μm. For n-type material the dopant is Si, while the dopant is Mg for p-type. In order to measure the three properties in a single run, namely the changes of material sheet resistance, etching rate, and surface morphology, a square-shaped transmission line model (TLM) structure was used in the etching experimental, with a pad length of 250 μm and width of 120 μm and the distance between pads changes linearly from 10 μm by 2 μm step until to 28 μm.

Ti/Al/Ni/Au metal layers with thickness 200Å/300Å/100Å/100Å were deposited for Ohmic contact and thermal annealing of 40 s in N_2 atmosphere was accomplished. Then, a photoresist mask was made. The ICP etching machine is an ICP-98C high-density plasma unit. An AMI BOIS XP-2 stimulus profiler was used for the measurement of mask height and etching step depth (5 points average). A probe stage and Keithley 236 Source Meter were used for resistance measurement. The surface morphology was observed by using a Vega-Tescan TS5136MM scanning electron microscope, because the material resistivity is high and the substrate is not conductive, about 120 Å Au were deposited on the sample surface by electron beam evaporation before SEM observation. Etching the surface roughness was measured using DI Nanoscope IIIa atomic force microscope with scanning area of 10 μm × 10 μm.

Table 6.3 Boiling point of products in etching GaN based material using methyl or halide plasma

Etching products	Boiling point (°C)	Etching products	Boiling point (°C)
$AlCl_3$	183	$InCl_3$	600
AlF_3	NA	InF_3	>1200
AlI_3	360	InI_3	NA
$AlBr_3$	263	$InBr_3$	Sublimation
$(CH_3)_3Al$	126	$(CH_3)_3In$	134
$GaCl_3$	201	NCl_3	<71
GaF_3	1000	NF_3	−129
GaI_3	345 (sublimation)	NBr_3	NA
$GaBr_3$	279	NI_3	Explosive
$(CH_3)_3Ga$	55.7	NH_3	−33
$(CH_3)_3N$	−33	N_2	−196

(1) The choice of ICP etching gases

In the choice of the etching plasma, the reaction product boiling points are an important factor to consider. For Al, Ga, In elements, their methyl or halide compound is a better choice. Table 6.3 gives the boiling point [46, 47] of several reaction products and some related compounds. The etching process is a dynamic process of evaporation and deposition of etching products. As shown in the table, the best volatile etching product for GaN-based materials is a methyl compound, especially since the $(CH_3)_3$ in boiling point is much lower than indium halide. Therefore, in the etching of GaN-based materials with high. In composition, CH_4 is very helpful. But the chemical activity of neutral methyl groups is much lower than for the neutral halogen atom. It is easy to form polymers which are often deposited on the surface and the side walls thus affecting the etching quality. On the other hand, the side wall effect can be used for better control of anisotropy. The polymer deposition on the surface will make the etching rate requirements difficult to meet. Another problem in methyl-based plasma is the presence of H^+ ion, which has a large penetration depth in GaN-based material. Simulations of Ar^+, Cl^+ and H^+ ions injection into GaN were done using TRIM by Singh [46]. The results are shown in Table 6.4. In addition, H in the GaN-based materials often affects the impurity properties, especially the p-type material, by the formation of a Mg-H complex which dramatically decrease the conductivity of materials [48].

For GaN-based material etching, except for very high In content material, Cl-based plasma is a good choice. The etching gases commonly used are Cl_2, $SiCl_4$,

Table 6.4 TRIM simulation of ion implantation depth of different ions in GaN (ion energy 1.0 keV)

Ion species	Projected range (nm)	Straggling (nm)
Ar^+	1.8	1.0
Cl^+	1.9	1.0
H^+	14.0	7.3

and BCl_3. The physical etching effect can be enhanced by adding an inert gas. For high Al content AlGaN material etching, adding some amount of BCl_3 is necessary [46]. Due to the more active Al atom, it is easy to react with O atoms, which are absorbed by the vacuum chamber or gas pipeline, by forming Al_2O_3 which is very difficult to etch away from the surface of the material. This is the "micro mask" effect. This effect leads to a "needle" etched surface. BCl_3 can effectively remove O atoms in the atmosphere of plasma.

To sum up, Cl_2 and BCl_3 are better choices for chemical etching gases, while Ar and He are better physical etching gases generally. Since the rate of chemical etching is much faster than the physical etching. Usually chemical etching gases are more effective than physical gas. It consists of six processes: gas ionization, reactive group arrival at the material surface, reactive groups diffusion onto the material surface, reaction with material, reaction product desorption, reaction product is pumped out of the reaction chamber.

(2) DC bias Influence on AlGaN etching

For different DC bias voltages, the n-$Al_{0.45}Ga_{0.55}N$ and i-$Al_{0.45}Ga_{0.55}N$ etching rates are shown in Fig. 6.24. The etching gases were He, Ar, and $Cl_2$with a gas flow rate of 10 sccm, 6 sccm, and 15 sccm. The reaction chamber pressure of about 0.65 Pa. ICP main power was 350 W. While the DC bias voltage changes. The etching time was 4.5 min or 6 min (i-AlGaN). The larger the DC bias, the more energy of the ions which are incident to the sample surface and the physical etching effect is more obvious. The large energy of incident particles is also better for the etching product volatilization. Therefore, from the figure, we can see that the etching rate increased with ICP DC bias. For the n-type material, the etching rate at −60 V is 407 Å/min; and it increased to 1654 Å/min at −200 V. Under the same bias condition the n-type material etching rate is lower than that of i-type material. Figure 6.25 shows the etching selectivity of AlGaN and Ni mask at different DC bias voltages. With an increase of the incident ion energy, Ni mask anti-erosion

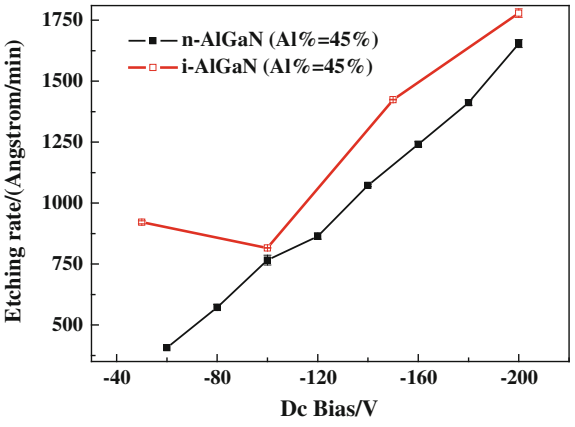


Fig. 6.24 Etching rate of $Al_{0.45}Ga_{0.55}N$ vs. dc bias

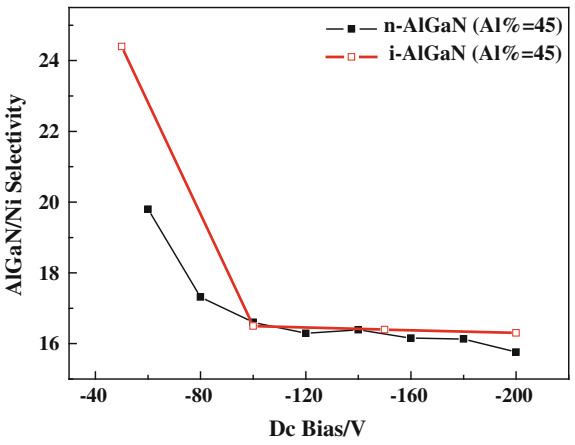


Fig. 6.25 Selectivity of $Al_{0.45}Ga_{0.55}N$ and Ni at different dc bias

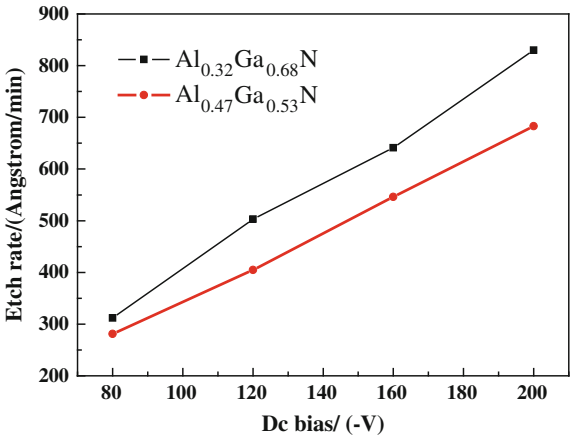


Fig. 6.26 The etching rate of n-$Al_{0.33}Ga_{0.67}N$ and n-$Al_{0.47}Ga_{0.53}N$ at different dc bias

ability gradually becomes poor. But the selectivity is still good and larger than 15 even at −200 V.

Figure 6.26 is the etching rate of n-$Al_{0.33}Ga_{0.67}N$ and n-$Al_{0.47}Ga_{0.53}N$ varied with dc bias voltage. From this figure, we can see that the etching rate of AlGaN increase monotonically with dc bias voltage. At −80 V, the etching rate of $Al_{0.33}Ga_{0.67}N$ is 312 Å/min, at same time, for $Al_{0.47}Ga_{0.53}N$ it is 281 Å/min. While at −200 V, the etching rate of $Al_{0.33}Ga_{0.67}N$ is 830 Å/min, for $Al_{0.47}Ga_{0.53}N$ it is 683 Å/min.

There are both physical and chemical processes in ICP dry etching. Compared with pure wet etching, ICP etching has an inevitable damage problem which tends to degrade the device performance [39]. Figure 6.27 is a relationship between TLM resistance and gap distance on the n-type material under different etching conditions. In Fig. 6.27a the gas flow rate is Cl_2/Ar/He (10/6/15 sccm), the ICP power is

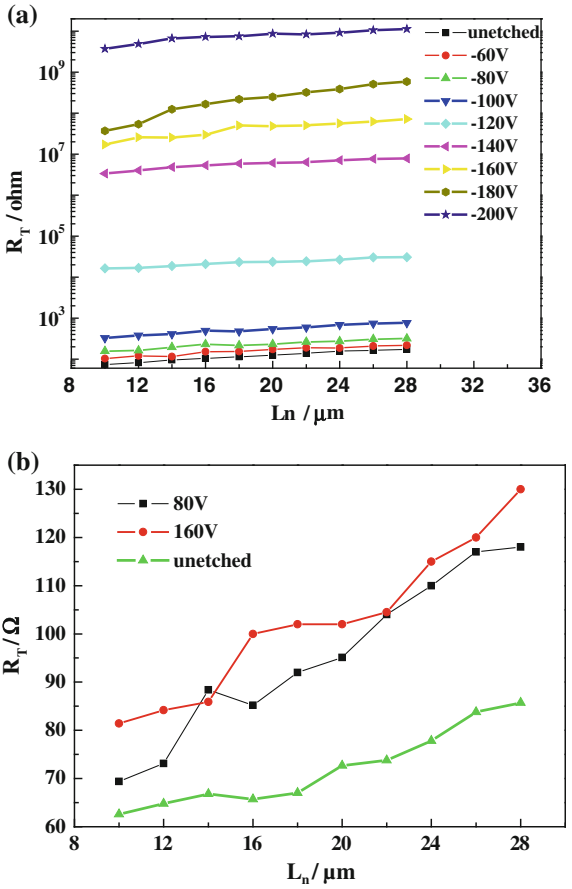


Fig. 6.27 Resistance between TLM pads with different gap distance under different etching condition **a** n-$Al_{0.45}Ga_{0.55}N$, **b** n-$Al_{0.33}Ga_{0.67}N$

350 W. Since the graph is in exponential scale, gap resistance, and square resistance is increased by orders of magnitude. The sheet resistance is about 10 times bigger after etching at −100 V. The sheet resistance increases about 100 times at −120 V. In Fig. 6.27b, the gas flow rate of Cl_2/Ar/BCl_3 is 8/2/20 sccm, ICP power is 350 W. This figure used a linear scale. Obviously, the sheet resistance of the latter is much smaller, so the etching damage is much less.

Using the data points in Fig. 6.27a, through linear fitting method, the sheet resistance value obtained. The change ratio of the sheet resistance (Re) before and after etching is shown in Fig. 6.28. In Fig. 6.28, the sheet resistance changes more and more obviously with the increase of dc bias. For the n-type material the change ratio is up to 10^7 at −200 V.

The sheet resistance increase after etching is likely due to etching introduced deep levels. These deep levels may capture free electrons, and not be ionized at room temperature, so as to compensate for donors which may account for the sheet resistance increase [49]. These deep levels will not only make the sheet resistance

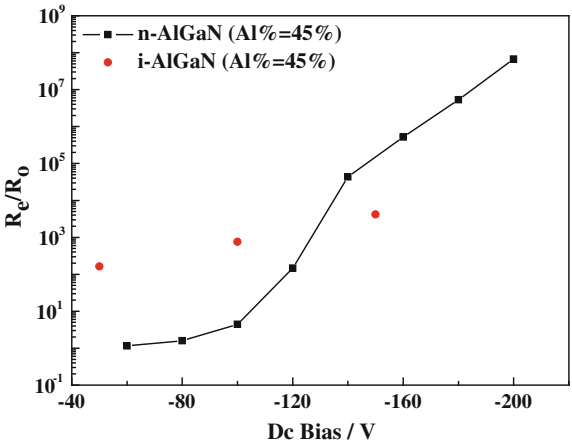


Fig. 6.28 The ratio of $Al_{0.45}Ga_{0.55}N$ sheet resistance after etching (R_e) over before etching (R_o) at different dc bias

larger, but also enhance carrier recombination, making the p–i–n junction leakage current increase, and degrade device performance [47]. The plasma components contribute differently to damage which is difficult to determine [50]. For the traditional III–V materials, such as GaAs, InP, reactive-ion etching of the material carrier loss has been studied more detailed [47, 50]. While for GaN-based semiconductor, it still needs further research.

Figure 6.29 is SEM photograph of n-type and i-type $Al_{0.45}Ga_{0.55}N$ mesa after etching. (a–c) is n-type $Al_{0.45}Ga_{0.55}N$ and (d–f) is i-type $Al_{0.45}Ga_{0.55}N$.

Figure 6.30 is AFM typical photograph of n-type $Al_{0.45}Ga_{0.55}N$ etched at different dc bias. Figure 6.31 is the AFM photograph of virgin n-$Al_{0.45}Ga_{0.55}N$ surface. Figure 6.32 depicts the RMS roughness of etched surface vs. dc bias voltage.

From the above test results it can be seen that, along with the increase of dc bias of etching, the etched surface becomes smoother. Larger ion energy leads to a more obvious bombardment which is conducive to the etching product volatility and a smooth etched surface can be obtained. In the SEM photos we can see that, for the etching of i-type material at the dc bias of −150 V the surface has been relatively smooth, as shown in Fig. 6.29e. But for the n-type materials it is at a bias of −200 V when the surface became smooth, see Fig. 6.29c. Considering the n-type and i-type material etching rate discussed above, Si-doped n-type material is difficult to etch and etched surface tend to be rough compared with i-type material which is unintentionally doped. However, with the increase of dc bias, a stripe structure gradually appears on mesa side wall; see Fig. 6.29c and f, which is due to erosion of the mask edge. This will lead to an increase of device reverse leakage current. At the same time, when the dc bias is high, the scattered ions from the etching surface bombard the side wall will become more numerous. This is also likely to cause damage to the device on the side wall which causes the reverse leakage current to increase [50]. However, if the dc bias is too small, the role of physical etching reduces, anisotropic etching will become poor, and prone to the undercutting phenomenon, which also reduces the performance of the device. In

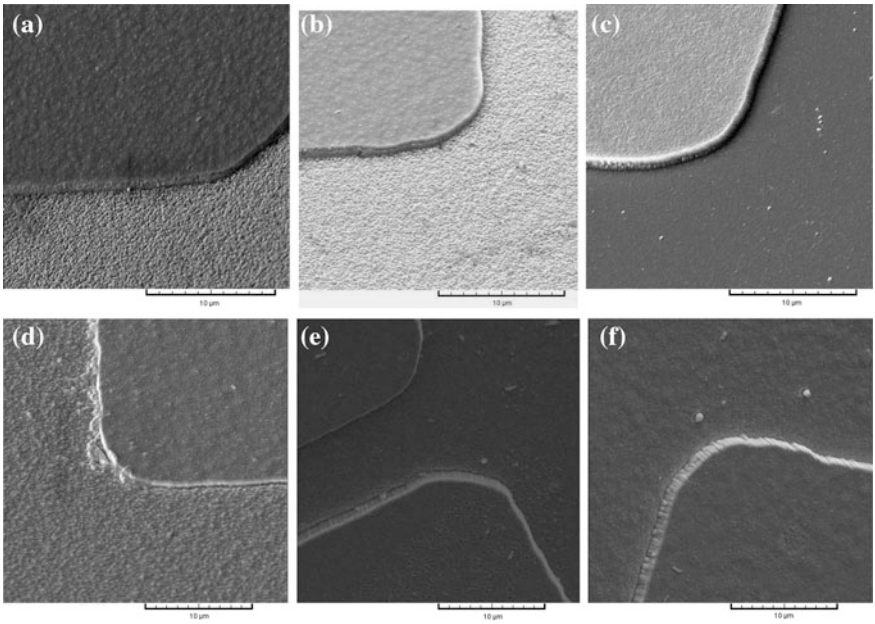


Fig. 6.29 SEM photo of mesa etched at different dc bias: **a** n-$Al_{0.45}Ga_{0.55}N$, −100 V; **b** n-$Al_{0.45}Ga_{0.55}N$, −160 V; **c** n-$Al_{0.45}Ga_{0.55}N$, −200 V; **d** i-$Al_{0.45}Ga_{0.55}N$, −100 V; **e** i-$Al_{0.45}Ga_{0.55}N$, −150 V; **e** i-$Al_{0.45}Ga_{0.55}N$, −200 V

this experiment, the undercutting phenomenon was not found to be apparent. This may be due to the relatively high bond energy and strong corrosion resistance of the GaN-based material.

Figure 6.33 shows the SEM photographs of n-$Al_{0.33}Ga_{0.67}N$ after etching at DC bias voltages of −120 and −160 V with a constant Cl_2/Ar/BCl_3 gas flow 8/2/20 sccm and 350 W ICP power. From Fig. 6.32 we can see that, for this material, at a DC bias of −80 V, the roughness is much bigger than that at −120 V, while, at −160 and −200 V, it remains small but shows a increasing trend with bias. From the SEM photograph we can see that some pits and hillocks exist. The pits appeared after etching are most likely to due to pores forming in the process of growth. The hillocks may be caused by the second phase inclusions and are difficult to remove in the etching process.

In conclusion, considering high-performance GaN-based UV detectors, the lowest level of damage is certainly preferred. But if dc bias was set too low, the etching rate will be too slow. Meanwhile, changing the composition of the etching gas may ruin the process. Therefore, the dc bias voltage value should be chosen carefully, balancing with other issues in certain processes. Generally, we recommend choosing a voltage of −100 V.

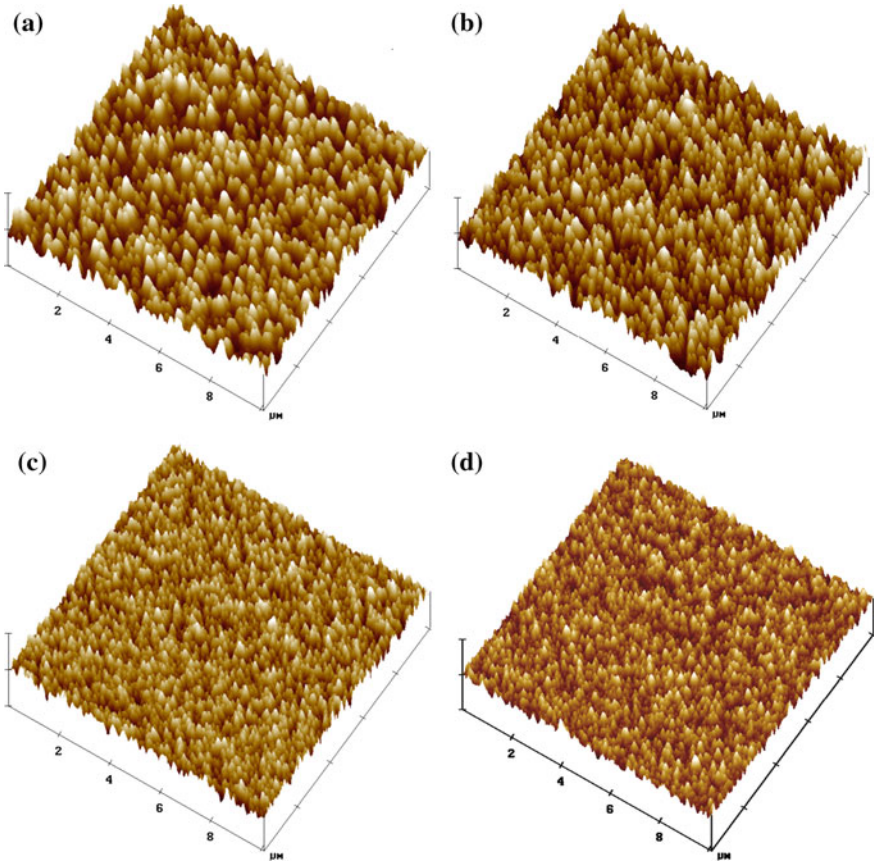


Fig. 6.30 The AFM photograph etched at different dc bias voltage: **a** −100 V, **b** −120 V, **c** −180 V, **d** −200 V

(3) The ratio effect of physical etching gases on AlGaN etching

Both He and Ar gases contribute to the physical etching effect in processes. The Ar/He ratio selection problem was studied by using etched surface roughness measurement. By using Ni Film as the etching mask, Chen [43] investigated the i-type $Al_{0.45}Ga_{0.55}N$ etching results with the gas flow rate of Ar/He at 9/0 sccm, 6/3 sccm, 3/6 sccm, while the Cl_2 gas kept at a flow rate of 20 sccm and the total gases flow rate remain unchanged at 29 sccm. The pressure of the reaction chamber was maintained at about 0.65 Pa. The main ICP power is 350 W. The dc bias voltage is −100 V.

Figure 6.34 depicted the etching rate versus He/Ar ratio. Figure 6.35 gives the surface roughness measurement results under different conditions. As the He ratio increases, the roughness of etched surface increases. These results show that maybe Ar is better than He in the AlGaN etching process.

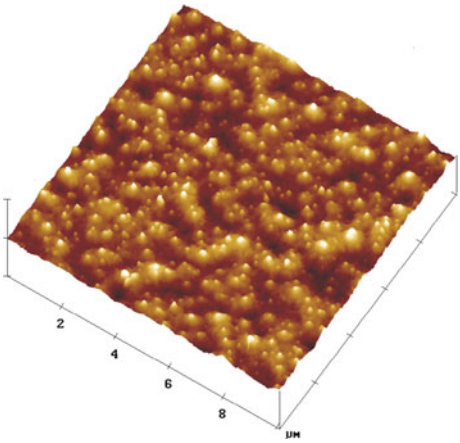


Fig. 6.31 The AFM photograph of virgin n-$Al_{0.45}Ga_{0.55}N$ surface

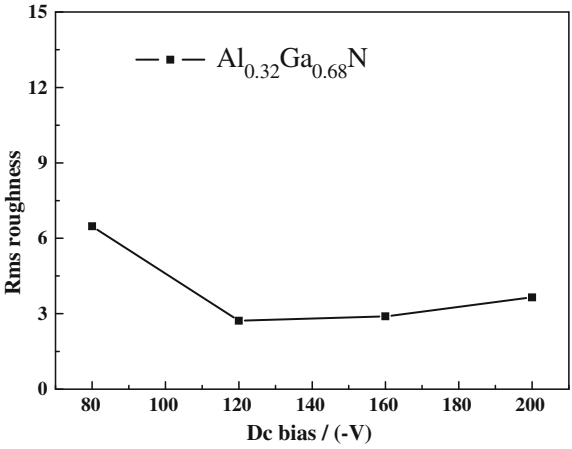


Fig. 6.32 The RMS roughness of etched surface vs. dc bias voltage (n-$Al_{0.33}Ga_{0.67}N$)

(4) Effect of physical /chemical etching gases ratio on AlGaN etching

Pure chemical wet etching will give almost zero damage to the lattice. It is believed that the same situation will hold in dry etching. Chemical etching gases, where the etching effect is mainly obtained by chemical reaction, will lead to less damage. Physical etching gases, which are accomplished by ion bombardment, tend to cause more damage. But, if etching is carried out with too much chemical gas, the results tend to be homogenous in three directions and cause rough surfaces. That is, undesirable in device fabrication. At the same time, the etching rate may be also undesirable. Experiments of chemical/physical etching gases ratio were carried out, focusing on the etching rate and surface roughness [43]. In these experiments, the Cl_2/Ar ratio changed while the total gas flow rate remains constant at 30 sccm. The chamber pressure was about 0.5–0.65 Pa (The pressure went a little higher with Cl_2

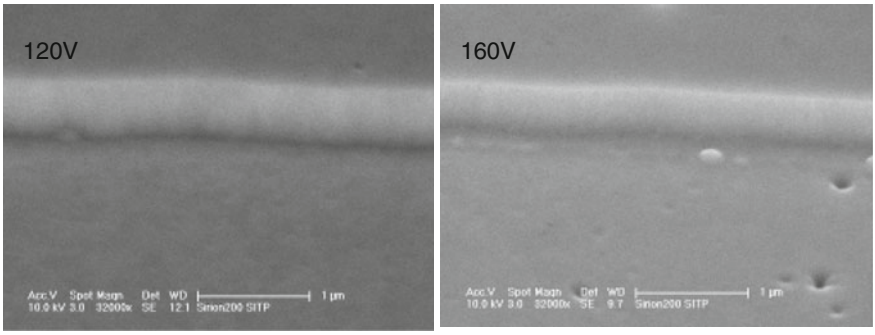


Fig. 6.33 SEM photograph of etched surface at different dc bias voltage: **a** −120 V **b** −160 V

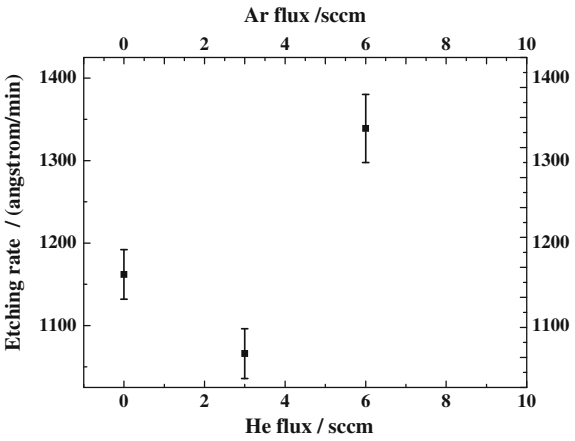


Fig. 6.34 The etching rate of $Al_{0.45}Ga_{0.55}N$ at different He/Ar ratio

content increase). ICP main power was 350 W. The dc bias was −100 V. Ni metal film was used as etching mask on p-type $Al_{0.45}Ga_{0.55}N$.

Figure 6.36 showed the etching rate vary with different Cl_2 content percentage. Figure 6.37 showed the surface roughness vary with different Cl_2 content percentage. Figure 6.38 are AFM photographs of different Cl_2/Ar gases ratio etched surfaces.

From Fig. 6.36 we can see that the etching rate goes high with Cl_2 content percentage increase; and it reach a maximum at Cl_2 content of 50 %. After that, the etching rate decreases with increasing Cl_2 content. From Figs. 6.37 and 6.38 we can see that roughness of the etched surface is the smallest also at a Cl_2 content of 50 %. These results may be explained as follows. When the Cl_2 content is small (<50 %), the chemical etching effect increases with Cl_2 content, and at the same time, the products of chemical reaction can be effectively removed from the surface. So, as etching rate increases, the surface becomes smoother. When the Cl_2 content is large (>50 %), Ar content is relatively small, chemical reaction is overwhelming

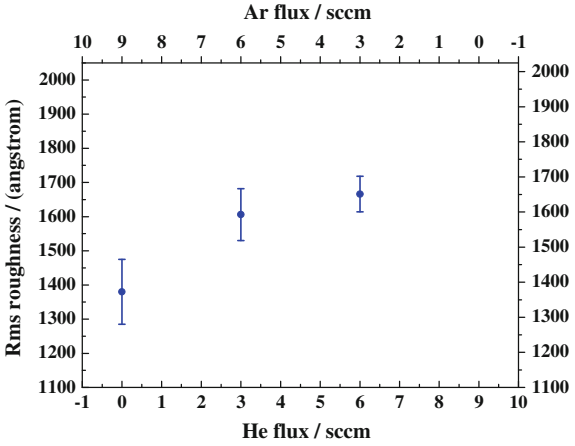


Fig. 6.35 The etched surface roughness at different He/Ar ratio

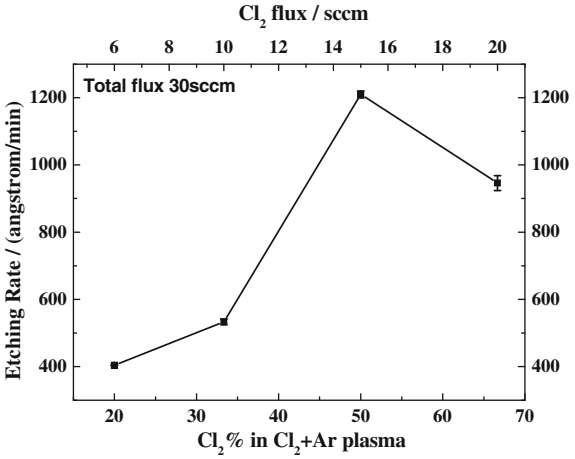


Fig. 6.36 The etching rate of $Al_{0.45}Ga_{0.55}N$ with different Cl_2 content percentage

which leads to a slow removal of reaction products. Reaction products deposited on the surface hinder the reaction. Therefore, the etching rate becomes slow and the etched surface becomes rough. These experiments show that there is an optimized content ratio at about $Cl_2/Ar = 1:1$. Generally, we choose a Cl_2/Ar ratio at 1:1.

(5) Effect of BCl_3 content on AlGaN etching

The Al atom in AlGaN is thought to be more active. It is intended to be oxidized by an O atom absorbed by ICP chamber surface or gas venting vessel surface. The product is Al_2O_3 which was a type of a micro mask, which led to a rough etched surface. Usually, BCl_3 is used to minimize this effect in the AlGaN etching process. Experiments were done to study the effect of BCl_3 content on AlGaN etching. The dc bias was −100 V, Cl_2/Ar ratio was 1:1, ICP main power was 350 W, total gases flow

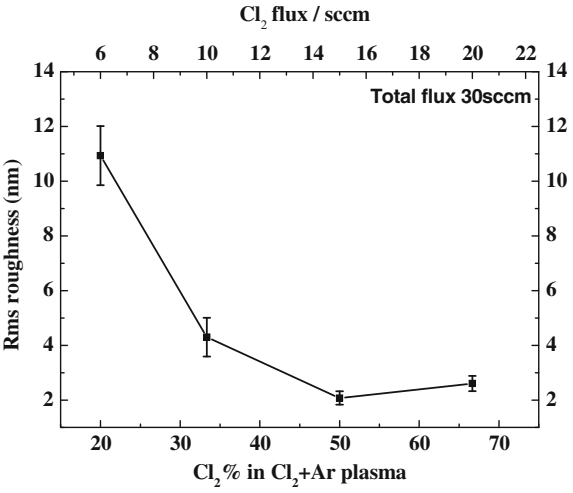


Fig. 6.37 The etched surface roughness with different Cl_2 content percentage

rate was 30 sccm, and the pressure of reaction chamber was 0.6–1.0 Pa (increased with BCl_3 content). Ni metal film mask was used and n-type $Al_{0.45}Ga_{0.55}N$ was used.

The chemical etching effect of BCl_3 is lower than that of Cl_2, because of its lower capability of forming Cl^- [46, 51]. As depicted in Fig. 6.39, according to the increasing of BCl_3 content, the etching rate decrease. From both Figs. 6.40 and 6.41, we can see that the roughness goes high and then decreases with BCl3 content increasing. At 80 % of BCl_3 content, the roughness reaches a minimum.

From Fig. 6.42 we can see that, the change ratio of sheet resistance, etched surface samples over virgin samples (R_{she}/R_{sho}), also gives a minimum at 80 %. This indicates that, etched at this condition, the surface damage may be smallest. Tripathy et al. [53] results of ICP Cl_2/BCl_3 plasma etching of GaN showed that, the intensity of the PL illuminated peak at the band edge decreases with Cl_2 content increase. This may be caused by Ga vacancy. The higher Cl concentration will lead to more Ga vacancies on the GaN surface. The same situation holds for AlGaN. $AlCl_3$, as a reaction product, has a lower boiling point than $GaCl_3$, which should be increase volatilization. But the bond energy of Al-N is higher than that of GaN. From Fig. 6.42, the obvious results that the sheet resistance change ratio decreases with BCl_3 content increase. This may be explained by the Cl- concentration effect on surface damage or surface carrier traps. The influence level on sheet resistance is low compared with the dc bias.

(6) Effect of $Cl_2/Ar/BCl_3$ content on AlGaN etching

In order to optimize the ICP etching recipe, systematic experiments of $Cl_2/Ar/BCl_3$ content effect on AlGaN etching were carried out [43]. AlGaN ICP etching characteristics depend on many factors such as plasma density, ion energy, and chamber pressure. Moreover, the etching rate was also affected by substrate temperature, because temperature will affect the desorption process of etching products.

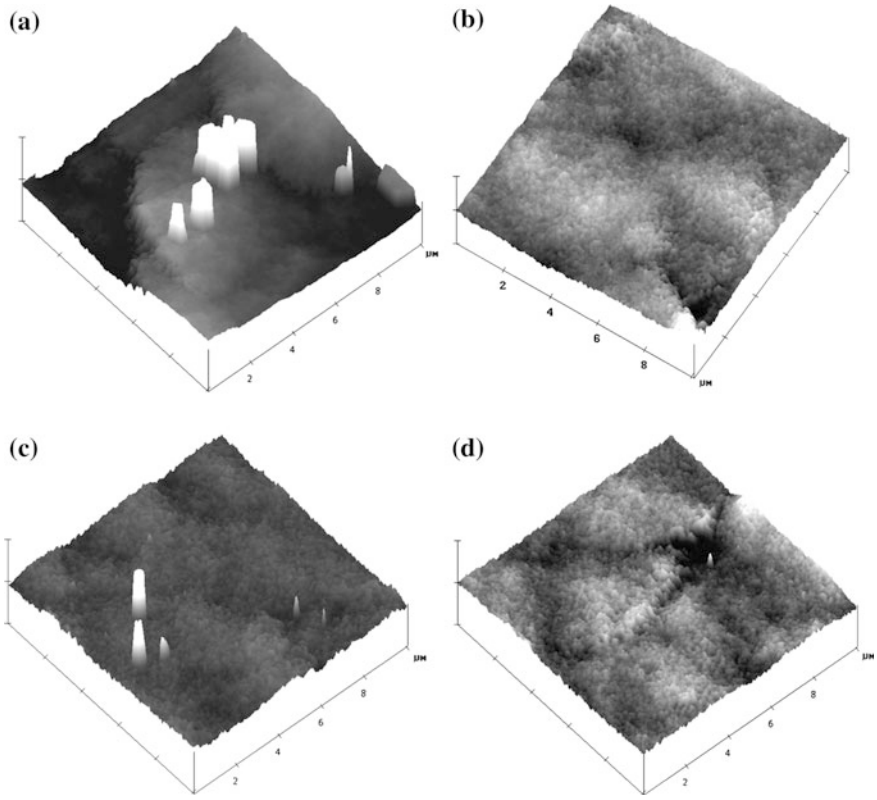


Fig. 6.38 AFM photograph of etched surface with different Cl_2/Ar ratio **a** Cl_2/Ar = 1:4; **b** Cl_2/Ar = 1:2; **c** Cl_2/Ar = 1:1; **d** Cl_2/Ar = 2:1

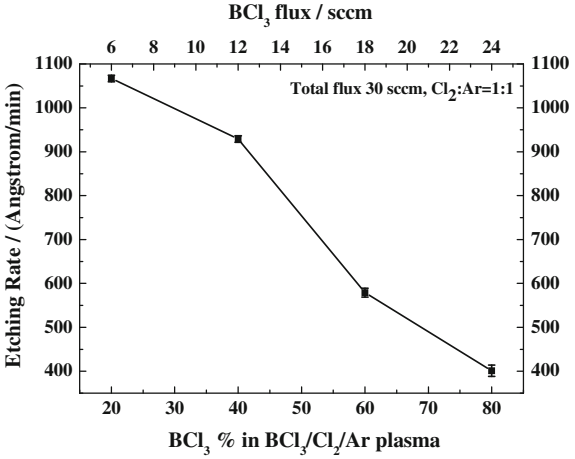


Fig. 6.39 Etching rate of $Al_{0.45}Ga_{0.55}N$ with different BCl_3 content

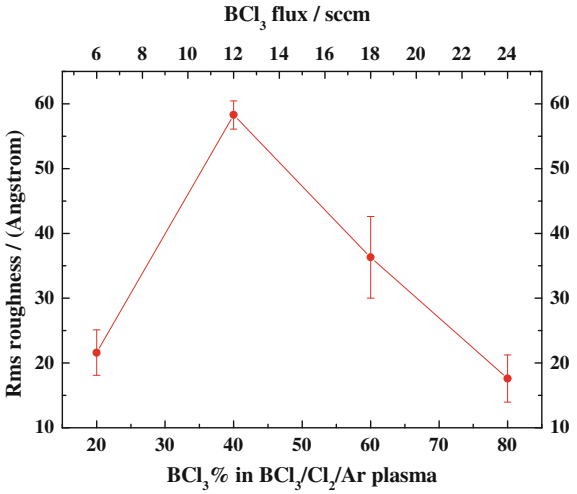


Fig. 6.40 Roughness of etched surface with different BCl_3 content

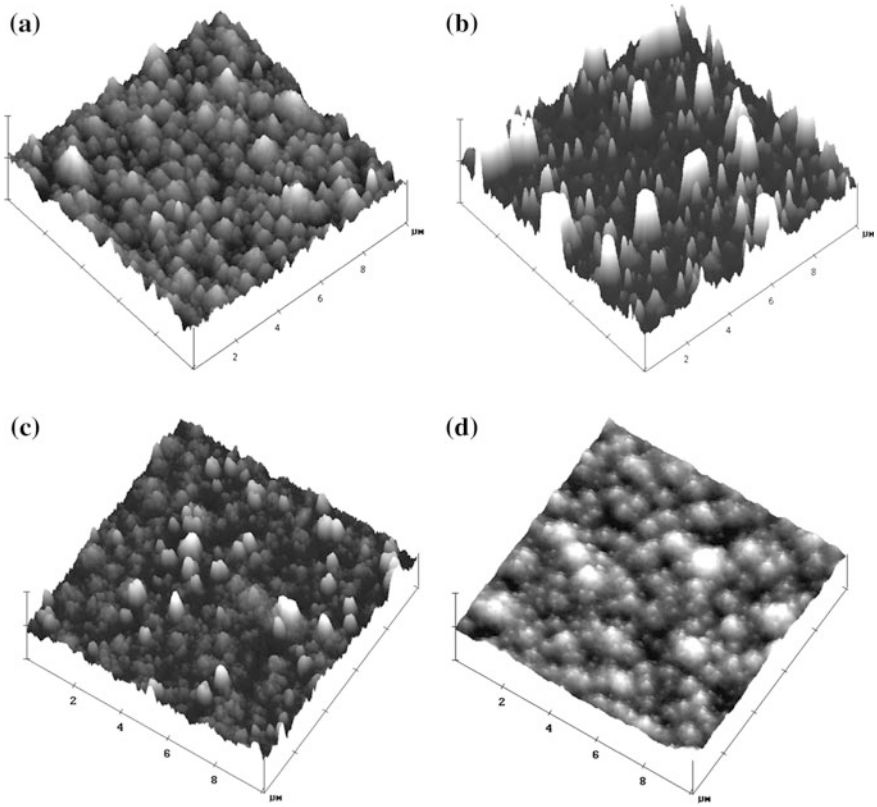


Fig. 6.41 AFM photograph of etched surface with different BCl_3 content: **a** BCl_3 % = 20 %; **b** BCl_3 % = 40 %; **c** BCl_3 % = 60 %; **d** BCl_3 % = 80 %

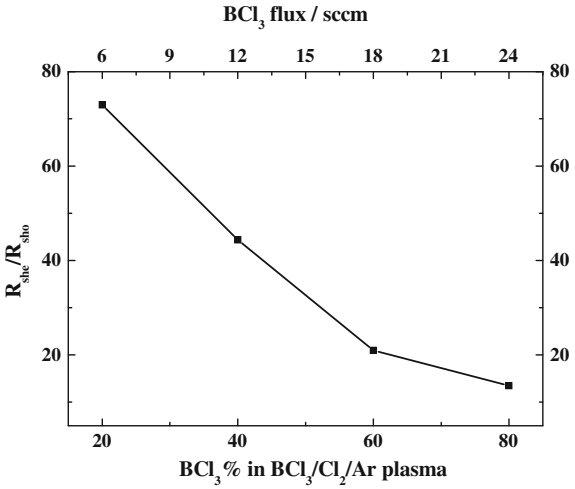


Fig. 6.42 The sheet resistance change ratio with different BCl_3 %

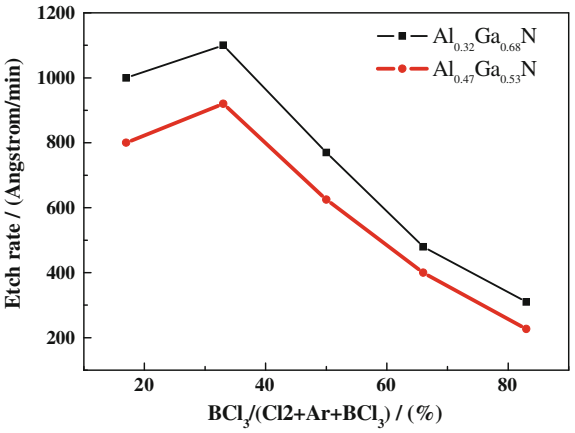


Fig. 6.43 Etching rate vs. gas percentage

Figure 6.43 shows AlGaN etching rate varies with the BCl_3 percentage of Cl_2/Ar/BCl_3. The materials are n-type $Al_{0.33}Ga_{0.67}N$ and $Al_{0.47}Ga_{0.53}N$. ICP main power is 350 W. The dc bias is −120 V. Reaction chamber pressure is 1 Pa. Total gases flow rate is 30 sccm. The gas flow rate of Cl_2/Ar/BCl_3 is 4/1/25 sccm, 8/2/20 sccm, 12/3/15 sccm, 16/4/10 sccm, and 20/5/5 sccm, respectively. From this Figure, we can see that etching rate depends on the gas recipe. As the BCl3 composition increases (that is Cl_2/Ar percentage goes low), the etching rate of AlGaN almost goes low linearly except for the first condition.

From Fig. 6.43 we can see that the etching rate increases by adding a certain amount of BCl_3. This is mainly because BCl_3 gas creates a protection effect. It prefers to react with an O atom and hinder the process of Al reacting with O. But further increasing the BCl_3 content will cause the etching rate to slow. The BCl_3 gas

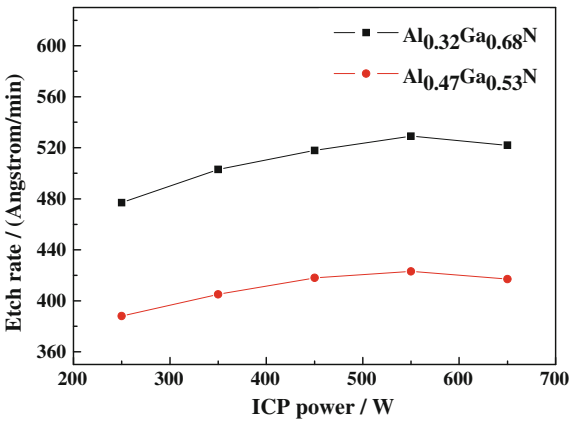


Fig. 6.44 Etching rate vary with ICP power

effect on etching rate is far below pure Cl_2. BCl3 will generate fewer Cl- active groups than pure Cl_2, neutral Cl is also far less than Cl_2. Neutral Cl is critical to the etching rate. Moreover, from Fig. 6.43 we can see that the etching rate of $Al_{0.53}Ga_{0.47}N$ is always higher than $Al_{0.33}Ga_{0.67}N$. This is caused by the Al and N combination energy which is higher than Ga and N [53].

Figure 6.44 is the etching rate. The figure shows the etching rate varying with ICP etching power. In etching, the $Cl_2/Ar/BCl_3$ gas flow rate is 8/2/20 sccm. From Fig. 6.44 we can see that, for those two samples ($Al_{0.33}Ga_{0.67}N$ and $Al_{0.47}Ga_{0.53}N$), the same trends result. The etching rate goes up slowly. After a maximum is reached, the etching rate goes down. For $Al_{0.33}Ga_{0.67}N$ maximum the etching rate of 529 Å/min is obtained at 550 W. While for $Al_{0.47}Ga_{0.53}N$ at 550 W, the maximum etching rate is 423 Å/min. ICP power has the dominant influence on plasma density, because determines the ionization rate of the etching gas. When the ICP power is relatively low, the gas ionization rate increases. The population of effective etching ion particles, known as neutral Cl atom, Ar^+ ion, Cl_2^+, and BCl_2^+, increases with ICP power. That leads to a growth of etching rate. When ICP power keeps increasing up to a certain level, the $Cl_2/Ar/BCl_3$ ionization will be saturated. At this condition, the thermal kinetic process dominates. Recombination between charged particles increases. Meantime, the neutral Cl particles collide more with the wall of reaction chamber, which also participated in the growth of recombination of neutral Cl particles. So, when ICP power goes even higher, the effective etching particles which can react with sample surface actually decrease. The etching rate will go down with ICP power. In fact, in this experiment, ICP power shows a small influence on the etching rate, which means that the dominant factor (the population of neutral Cl particle) changes little. Considering that high plasma density can be obtained at low ICP power, 350 W is thought to be an appropriate setting. Etching experiments at that condition may also avoid some bad side effects, such as overheating the substrate by large ICP power (i.e., 650 W).

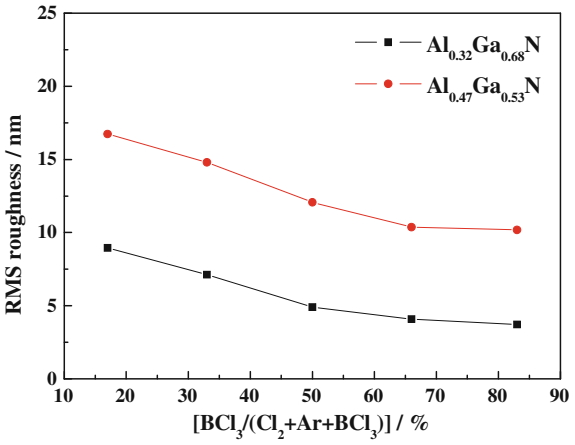


Fig. 6.45 Surface roughness vary with gases flow rate ratio

Table 6.5 XPS measured content of different elements on surface

Element	C (%)	O (%)	N (%)	Al (%)	Ga (%)	Cl	(Al + Ga)/N
As grown	41	15.1	20.4	11.6	12.9	4.40×10^5	1.20
Etched	36.4	18.8	18.4	11.2	15.2	7.61×10^5	1.43

Figure 6.45 shows the surface roughness of $Al_{0.33}Ga_{0.67}N$ and $Al_{0.47}Ga_{0.53}N$ vs. ratio of gas flow rate. Cl_2/Ar keep at 4:1, total gas flow rate is 30 sccm. BCl_3 percentage increase.

Figure 6.45 shows that the $Al_{0.33}Ga_{0.67}N$ sample has a better surface than $Al_{0.47}Ga_{0.53}N$. This may be due to different crystal quality. As the BCl_3 percentage increases, the surface roughness goes down. That means the BCl_3 contributed to surface smoothing. The results coincided with Tripathy's report [52].

XPS studies on the surface of samples before and after etching, respectively, acquired percentage measurements of C, O, N, Al, Ga, Cl, along with the deduced ratio of (Al + Ga) over N, and are shown in Table 6.5. Carbon percentage shows that the sample surface may suffer a severe contamination. After etching this percentage decrease a little. Oxygen percentage goes a little bit higher after etching, which show oxidation of surface was under control. For the Al, Ga, N components, the ratio of (Al + Ga) over N is 1.20 before etching. It becomes 1.43 after etching. It shows that etching process gives a surface with more N vacancies. That means, in ICP etching of AlGaN surface, N atoms are removed more effectively than Ga and Al atoms. This leads to a Ga and Al rich surface. Cl_2, as a main part of etching chemical gas, shows only a traceable concentration. That again coincided with Tripathy's report [52], except for the small N vacancy in our results. The smaller N vacancy may indicate that our etching recipe gives better results in term of surface damage. Since Cl_2 etching tends to give a N-rich surface [54], while Ar etching gives N-deficient surface. By combining Cl_2 and Ar in the etching recipe, a better etched surface with normal stoichiometry can be obtained.

2. Wet etching

At room temperature, a III-nitride compound shows a very slow etching rate in a wet etching solution. Pearton et al. [55] found a 2 nm/min etching rate of GaN in a 30–50 % NaOH solution, 3 nm/min of AlN in an HF solution, 1 nm/min of InN in an HCl/HNO_3 solution. At higher temperatures with highly chemical active etchants (KOH or NaOH), the etching rate becomes larger. At 60 °C in KOH or NaOH, InN the etching rate will be in range of 30–60 nm/min [56].

Chemical etching can eliminate the damage caused by the ion beam etching. Device performance may be improved. For UV detectors, AlGaN material chemical etching was studied at high temperatures with high concentrations of KOH solution. Devices were also made to compare the leakage current and spectral response [57–59].

In order to study the wet chemical etching effect on AlGaN, a n-type $Al_{0.63}Ga_{0.37}N$ material sample was made by MOCVD epitaxy on sapphire substrate with a AlN buffer layer. The AlGaN film thickness was 0.6 μm, carrier concentration is about 1.0×10^{17} cm^{-3}, the etching masks were made by e-beam evaporation and photolithograph with a structure of Ni/Au/Ni/Au (20/20/20/20 nm), followed with 550 °C RTA in air. This metal alloy mask on the surface provides a shield from KOH attack. After the etching process, a step structure will form which is used to measure the depth of etching. A 20 % wt KOH water solution was used. It was heated up to a temperature of 85, 95, and 102 °C, respectively. The heater is an oil bath controlled by a PID controller and a Pt resistor sensor. The temperature can be exactly set within ±1 °C. The experiment results show the wet chemical etching rate is 0.48, 0.64, and 0.96 μm/min, respectively. Figure 6.46 shows the relationship of etching depth with etching time at different temperature.

Figure 6.46 shows the relationship of etching depth with etching time at different temperatures. At the start period, the etching rate shows little difference for the three temperatures. But after 20 s, it diverged. The higher the temperature is, the faster

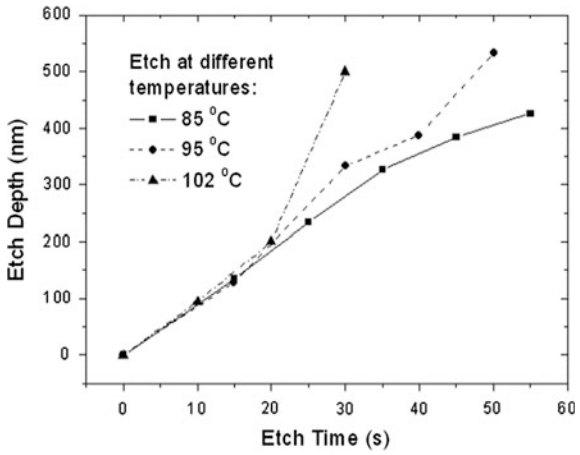


Fig. 6.46 Wet chemical etching curve of n-$Al_{0.63}Ga_{0.37}N$ at different temperature

the etching rate. It is learned that, in GaN chemical etching, OH- reacts with Ga. The equation is written as following:

$$2GaN + 3H_2O \rightarrow Ga_2O_3 + 2NH_3$$

Similar to AlGaN, Al oxidation will form. $Al_xGa_{2-x}O_3$ film will form on the sample surface according to the Al composition. This film will hinder the reaction from going further. But KOH will dissolve this film. The higher the temperature, the faster is the dissolving speed and, therefore, the etching rate increases. At the start period, the low change of etching rate between different temperatures may due to the natural oxidation film already formed on the sample surface which should be dissolved first.

We should not expect that only chemical wet etching was used in device fabrication. Some difficulties are encountered in wet etching, such as the uniformity issues, dependency on material quality, and Al composition. By combining the dry etching and wet chemical treatments, device fabrication will be improved. Dry etching, such as Ar ion milling and ICP processing, always tends to form some damage on etched surface and the side walls, which may introduce more leakage current in devices and degrade the device performance. Experiments were carried out to show that wet chemical etching of the Ar^+ ion milled surface can reduce the damage and leakage of device. Figure 6.47 shows the SEM photos of Ar^+ ion milling of a sample surface before and after wet chemical treatment. The virgin sample is p–i–n multilayer structure $Al_{0.63}Ga_{0.37}N$ material. A dry etching process was done to reach the n-layer. Those two samples are prepared with surface carbon spray for better electrical conductivity. The sample stage is tilted at a 60° angle to better view pits on the surface. By comparing the two photographs, it is obvious that the chemical treatment is effective in the removal of surface damage caused by dry etching.

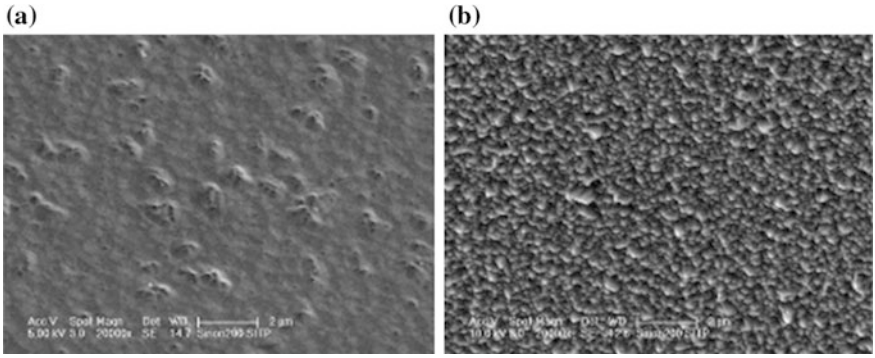


Fig. 6.47 SEM photograph: **a** surface after Ar^+ ion milling; **b** after Ar^+ ion milling, then 105 °C 20 % KOH solution treatment for 15 s

AES studies give the same conclusion as shown in Fig. 6.48. After wet chemical etching, oxidation of Al and Ga was removed. Al, Ga, and N surface composition percentage and the ratio are almost restored to the bulk characteristics.

In order to study the electrical property influence of the wet etching process, single element visual-blind p–i–n devices were made with a material structure shown in Fig. 6.49.

Some device samples were made using a wet chemical 20 %KOH treatment at 106 °C and in 15 s, while other samples were fabricated without that process. Current–voltage (I–V) measurement shows that, at −5 V bias voltage, the leakage current of the wet treated sample is about 5.09×10^{-9} A. while it is about 3.49×10^{-8} A for the untreated one. The results are shown in Fig. 6.50. Almost a decade degree improvement was obtained by the wet chemical treatment.

UV spectral response is another important feature of devices. The experiment set consisted of a 350 W xenon lamp and a computer controlled monochromator. Device response spectra under front and backside-illumination are shown in Fig. 6.51a and b, respectively. Comparison of wet chemical treated sample spectra with untreated ones was shown with different color and symbols. In both front

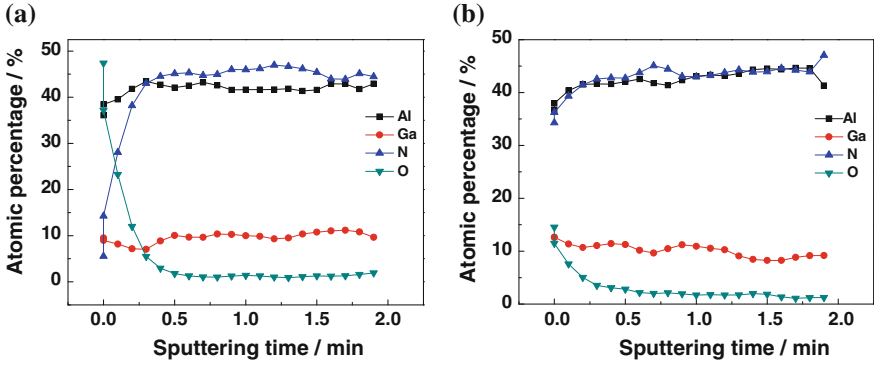


Fig. 6.48 AES measurement data: **a** before wet etching; **b** after 20 % KOH wet chemical etching

Fig. 6.49 Material layer structure

p-GaN 150 nm
i-GaN 180 nm
n-$Al_{0.33}Ga_{0.67}N$ 560 nm
AlN buffer layer
Sapphire

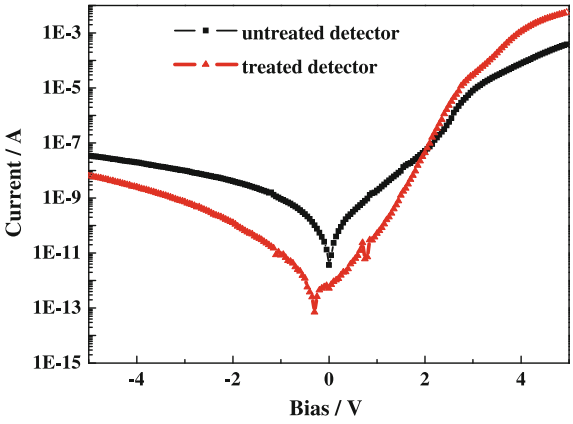


Fig. 6.50 I–V curves of wet chemical treated sample and untreated one (absolute value of reverse current)

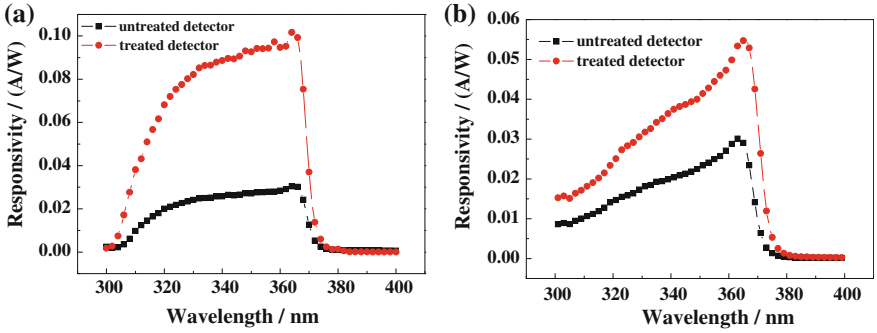


Fig. 6.51 Response spectra of wet chemical treated sample and untreated one: **a** back side illumination; **b** front side illumination

illumination and backside-illumination conditions, wet chemical treated samples show larger response over untreated ones.

For backside illumination, the treated sample has a peak current responsivity of 0.10 A/W. While the untreated one is 0.03 A/W for front illumination, it is 0.05 versus 0.02 A/W. We can see that wet chemical treatment can help to improve the device responsivity.

6.4.2 Passivation of the Chip

A passivation process is widely used in fabrication of UV detectors. Usually, UV detectors use a mesa structure. Surface leakage and side wall leakage current may be reduced by introducing a layer of passivation on the chip surface and side wall. It is not only better for device electrical properties but also for quantum efficiency and signal-to-noise ratio. SiO_2 or Si_3N_4 films were commonly used for GaN/AlGaN heterojunction device passivation.

SiO_2 film is used for leakage current reduction, for reliability improvement, and for radiation hardening [60, 61]. The passivation of HEMT was also reported [60, 62]. For UV detector passivation, especially the thermal annealing effects on SiO_2 were studied [63, 64].

We know that a SiO_2 passivation layer will change the device optical properties such as reflectance and transmittance of the chip. Thus, device external quantum efficiency will also change. PECVD SiO_2 film was deposited on a double-side polished sapphire substrate. It is thermally annealed in oxygen gas at different temperatures. Figure 6.52 shows the transmittance measurement results. Transmittance of the sample decreased at UV wavelength range. The shorter the wavelength, the more the transmittance value is decreased compared with the pure substrate. At the same time, thermal annealing also influences the SiO_2 passivation film transmittance in the UV spectral range. The 550 °C thermal annealing improved the transmittance by about 5 %. But the 750 °C annealing and the 650 °C annealing results show transmittance little improved. While for 900 °C annealing the SiO_2 transmittance decreased about 15–20 %. The shorter the wavelength is the greater decrease in SiO_2 transmittance.

Figure 6.53 shows sample reflectance measurement results. Unlike the transmittance, the value of reflectance changes only about 2–8 %. These results show that a SiO_2 passivation layer can reduce the reflectance to a certain degree. Therefore, considering the transmittance and reflectance results, a SiO_2 layer can be used for the 300–365 nm wavelength band without significant loss of light energy.

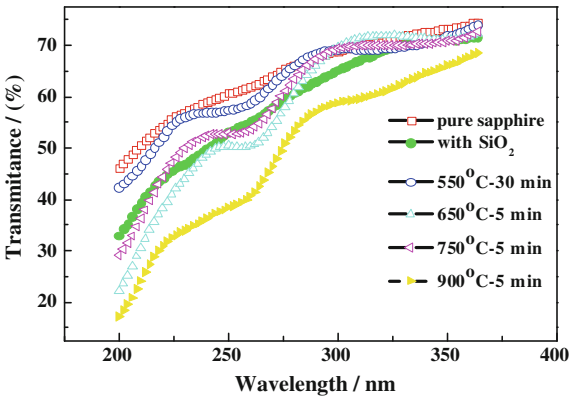


Fig. 6.52 Transmittance spectra of SiO_2 on sapphire and with different annealing conditions

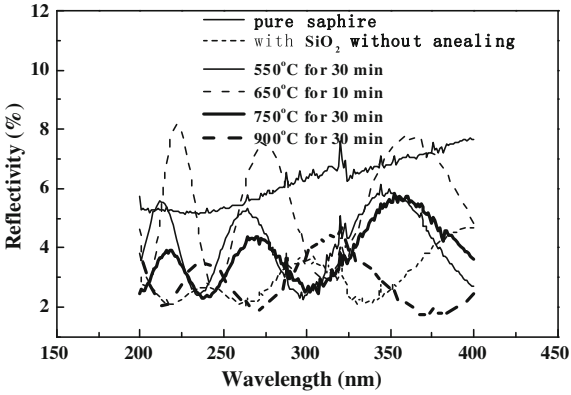


Fig. 6.53 Reflectance spectra of SiO_2 on sapphire and with different annealing conditions

While for 200–300 nm wavelength band usage one should take care of the transmittance lost. This may be caused by light scattering in the film.

6.4.3 Ohmic Contact

1. Ohmic contact on n-GaN

Compared with Si and GaAs, n-GaN has a high work function of 4.2 eV. According to the thermionic emission theory, metals with work function lower than or around 4.2 eV are preferred for fabricating Ohmic contacts on n-GaN. These possible metals are Ti, Al, Ta and W, etc.

It is generally believed that the conditions and mechanisms for forming good Ohmic contacts are as follows: first, during metal deposition the ratio of Ti and Al should meet a certain value, then in the subsequent annealing treatment, a portion of Ti remains after the formation of the Al_3Ti alloy, finally an Al_3Ti/Ti/GaN structure is formed. Additionally, itis found that Ti can react with N in GaN and generate TiN. TiN is a semimetal and conductive layer, which can reduce the specific contact resistance [67]. Furthermore, reactive-ion-etching (RIE) before metal deposition can produce a low contact resistance for Ohmic contacts on n-GaN. The possible mechanisms are: (a) removal of surface oxides. (b) Creation of a donor-like layer in the plasma-induced damage area. (c) Forming a donor-like N vacancy. A surface analysis technique should be used to further identify the c) mechanism.

Figure 6.54 shows the current–voltage (I–V) curves of Ti/Al contacts on i-GaN at 10 μm spacing under different annealing temperatures. Figure 6.55 shows the measured specific contact resistance as a function of the contact gaps. The specific contact resistances are calculated using a transmission line model. The I–V measurements clearly indicate that the rectifying contact forms at Ti/GaN interface before annealing, then the contact shows linear I–V characteristics after annealing at higher than 600 °C. The experiments show that with increasing annealing

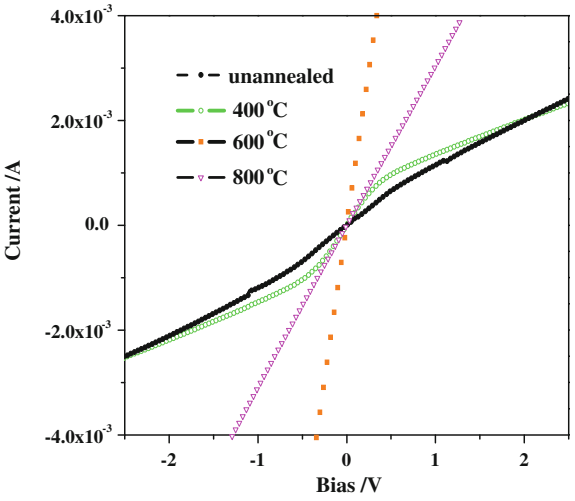


Fig. 6.54 Current–voltage curves of Ti/Al contacts under different annealing temperature

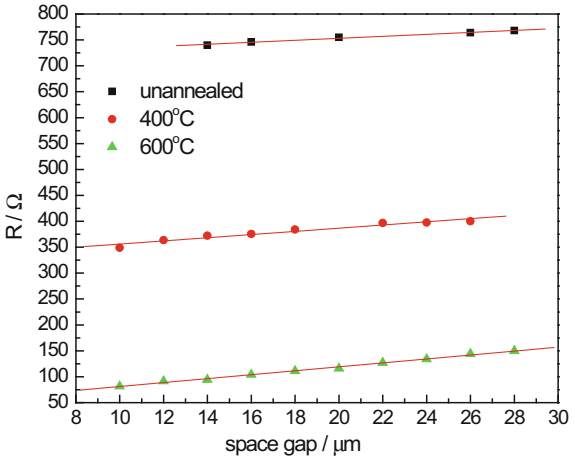


Fig. 6.55 The measuredspecific contact resistance as afunction of the contact gaps

temperatures, the specific contact resistances decrease from 1.66×10^{-1} Ω cm^2 to 3.03×10^{-4} Ω cm^2.

Moreover, to further investigate the effects of carrier concentration on the Ohmic contact, the same Ti/Al contacts are fabricated on i-GaN and n-GaN, then the contacts on i-GaN are annealed at 600 °C under N_2 ambient, and the contacts on n-GaN are not annealed. Figure 6.56 compares the space resistance of the two samples as a function of the space gaps. The two samples both show linear I–V curves, indicating that the Ti/Al/i-GaN and Ti/Al/n-GaN both form Ohmic contacts. For the annealed Ohmic contact at 600 °C on undoped i-GaN with a carrier concentration of 1.19×10^{17} cm^{-3}, the specific contact resistance can reach 3.03×10^{-4} Ω cm^2. While for the unannealed Ohmic contact on Si-doped n-GaN with a carrier concentration of 5.88×10^{18} cm^{-3}, the specific contact resistance is

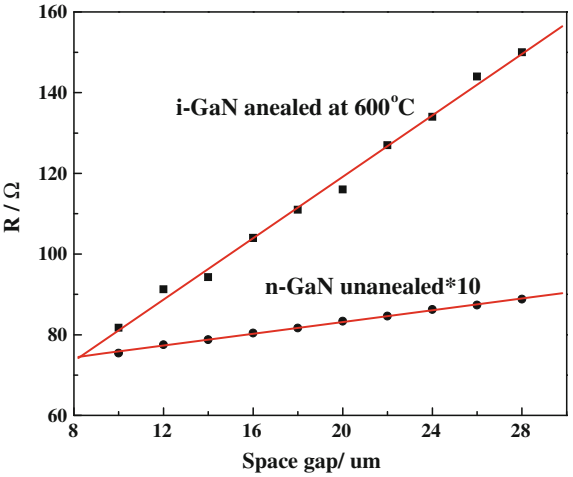


Fig. 6.56 The space resistance of the two contacts on i-GaN and n- GaN as a function of the space gaps in transmission line model

up to 4.03×10^{-4} Ω cm^2, after annealed at 600 °C, the specific contact resistance can reach 2.3×10^{-6} Ω cm^2.

As far as the material is concerned, the carrier concentration of the i-GaN is 1.19×10^{17} cm^{-3}, and the depletion layer width of the unannealed Ti/GaN interface is around 600 Å. The tunneling parameter is about 4.06 meV, $kT/E_{00} = 6.4$ illustrates that for undoped i-GaN before annealing, thermal electron emission is the main transportation mechanism and the interface shows a Schottky contact. The interface reaction induced N vacancy after annealing can generate a heavily doped region, which may facilitate electron tunneling. The carrier concentration of the n-GaN is 5.88×10^{18} cm^{-3} and the depletion layer width is around 85 Å. The tunneling parameter is about 32.7 meV and $kT/E_{00} = 0.8$ illustrates that before annealing the doping concentration in Si-doped n-GaN is high. The depletion layer becomes narrower, then the electrons can tunnel through the barrier, and thermal electron emission is the main transportation mechanism.

2. Ohmic contact on p-GaN

Compared with n-GaN, the successful preparation of an Ohmic contact on p-GaN is always a huge challenge and problem in developing GaN-based ultraviolet detectors. In fact, the high contact resistance to p-GaN is one of the major technical obstacles to realize long life and continuous working GaN-based devices.

The main reasons for the difficulty in fabricating Ohmic contact on p-GaN are as follows: first, compared with Si and GaAs, the work function of p-GaN is even higher than n-GaN (n-GaN:4.2 eV, p-GaN:7.5 eV [68]). The band gap of GaN is 3.4 eV, the electron affinity is 4.1 eV, and the largest work function of Pt is only 5.38 eV. It is difficult to find a metal with sufficiently high work function for fabricating low-resistance Ohmic contact. Second, it is difficult to grow heavily doped p-GaN, which is ascribed to the low activation efficiency of p-GaN. The doping concentration of Mg is 10^{20} cm^{-3}, and Mg can produce deep acceptor levels

and the ionization energy of Mg is rather high ($\sim$170 meV). According to Fermi–Dirac statistics theory, the ionization rate of Mg dopant should vary from 29 % (the concentration of Mg is 10^{17} cm^{-3}) to 1 % (the concentration of Mg is 10^{20} cm^{-3}), resulting in the hole concentration being difficult to reach 10^{18} cm^{-3}. The low hole concentration causes a high Schottky potential barrier at the metal/ p-GaN interface, limiting the tunnel current required for the formation of low-resistance Ohmic contact. Additionally, metallization processes, such as surface treatment, metal deposition, alloying processing, and conditions also affect the contact resistance to p-GaN. Not only is the adhesion of contact metal with p-GaN generally poor, but also the process is likely to cause the generation of N vacancies at GaN surface so that the surface of p-GaN might convert to n-type conductivity, thereby, seriously affecting the Ohmic contact on p-GaN.

Therefore, in order to obtain low contact resistance, reliable, high-quality Ohmic contact on p-GaN, we need to reduce the effective barrier height or realize heavy doping at the interface, which requires research on material structure and device process techniques. In terms of material, heavily doped p-GaN is preferred. The higher the dopant concentration, the thinner the barrier, the easier it is to achieve Ohmic contact. But this mechanism has its inherent defects, limited by the material growth process and device performance, the doping concentration at interface cannot reach too high, making it difficult to obtain low contact resistance. We generally use surface treatment and annealing to increase the doping concentration at interface, but these methods are limited. Another method is trying to reduce the barrier height or completely eliminate the barrier, such as epitaxial growth of a narrow band gap semiconductor layer (InGaN, etc.) on p-GaN by band gap engineering, or the use of a superlattice to lower surface barrier [69]. In terms of the device process, we usually use metal systems with high work function to produce electrode on p-GaN, such as Pt/Ni/Au, Pd/Au, Ni/Au, Co/Au, Cr/Au, Pt, Pd, etc. Moreover, annealing is necessary for fabricating Ohmic contacts on p-GaN, and the annealing temperature, time, ambient all have effects on the contact resistance [69, 70]. Also improving the hole concentration by using superlattice structures, removing the oxides at the metal/semiconductor interface by surface treatment, reducing the surface N vacancies, and increasing Ga vacancies are all effective methods used to reduce the contact resistance [71, 72].

The researchers have tried various metallization schemes, including Au, Ni, Ti, Pd, Pt, W, WSi, Ni/Au, Pt/Au, Cr/Au, Pd/Au, Au/Mg/Au, Ti/Pt/Au, Pd/Pt/Au, Ni/Cr/Au, Pt/Ni/Au, Ni/Au-Zn, Ni/Si, Ni/Mg/Ni/Si, etc. Generally, Ni/Au is a good choice for the preparation of low Ohmic contacts and is widely used in LEDs and photodetectors. It is found that the clean surface of GaN and the presence of oxygen during annealing facilitate the formation of Ohmic contact. In earlier reports, the specific contact resistances of Ni/Au on p-GaN are usually 10^{-4} Ω cm^2 of magnitude [71, 73, 74]. Suparna Pal and Takashi Sugino fabricated Ni/Au contacts on Mg-doped p-GaN ($N_A = 1 \times 10^{20}$ cm^{-3}), then obtained the lowest ρ_c of 10^{-2} Ω cm^2 after the sample was annealed at 700 °C for 30 s. They also pointed out that the contact resistance increased with increasing annealing temperature, which is due to the high-temperature-induced surface N vacancies. The donor-like N vacancies

give rise to the ρ_c. Compared with the alloyed Ni/Au contacts, the oxidized Ni/Au contacts exhibit lower resistance and more transparency, Ho et al. [75] reported that for the oxidized Ni (10 nm)/Au (5 nm) contacts on p-GaN ($N_A = 2 \times 10^{17}$ cm^{-3}), the contact resistance was lower than 1.0×10^{-4} Ω cm^2 and the transparency increased from 65 to 80 % at 450–550 nm. They pointed out that the *N* vacancies and the interstitial O of NiO at the interface are manifested as a p-type semiconductor, and the hole concentration in NiO could reach 1.3×10^{19} cm^{-3}, which was about one to two orders of magnitude more than the normal p-GaN and more easily to form an Ohmic contact. Jang et al. [76] used Pd (3 nm)/Ni (7 nm) to prepare low-resistance Ohmic contacts on p-GaN, after annealing at 500 °C in oxygen, the obtained minimum specific contact resistance was 5.7×10^{-5} Ω cm^2. The analysis showed that NiO formed at the surface prevents out diffusion of Pd atoms and formation of Ga_2Pd_5 and Ga_5Pd, and the Ga vacancies account for the reduction of contact resistance. Since Pt has the highest work function of 5.65 eV, a series of Pt-based metal stacks have been extensively studied as p-type contacts. Arai et al. investigated and compared the stability of electrical properties and microstructures of Pt, Pt/Au, Ni/Au, and Ta/Ti contacts.

Four kinds of contacts were fabricated on the Mg-doped p-GaN epitaxial layer, the hole concentration is 6–7 $\times 10^{17}$ cm^{-3}, and was annealed in O_2 and N_2 ambient. The results showed that Ta/Ti provided the lowest contact resistance, but the performance was unstable. The contact resistances of Pt (O_2) and Pt/Au (O_2) are 10^{-3} Ω cm^2 of magnitude, which cannot meet the requirements of blue LD ($<10^{-4}$ Ω cm^2), but they exhibit the most stable electrical properties under electrical injection. Zhou et al. investigated the Ti (15 nm)/ Pt (50 nm)/ Au (80 nm) contacts on p-GaN ($N_A = 3 \times 10^{17}$ cm^{-3}). After annealing at 800 °C for 2 min under N_2 ambient, the contact resistance is 4.2×10^{-5} Ω cm^2. Chen et al. utilized anodic oxidation to generate at the surface of p-GaN, and then removed GaO_x by wet etching, thereby creating more Ga vacancies and increasing the hole concentration. Next, the Ni/Au contacts were prepared, and the obtained contact was 6.1×10^{-6} Ω cm^2, which was currently the lowest contact resistance on p-GaN.

There are several models to explain the reduction of the contact resistance [75–78]. For example, some studies have shown that a Au/NiO/p-GaN structure is formed at the interface, the annealing process may break Mg–H bond, improving the activation efficiency of Mg dopant, or the interface reaction induced formation of Ga vacancies under the contacts, thereby improving the hole concentration and decreasing the Fermi level position.

We investigated the preparation process and theory of Ohmic contacts on p-GaN, and used SEM, AES depth analysis to study the formation mechanism of Ohmic contacts under different annealing conditions [66, 79–81]. Our p-GaN samples were grown on a sapphire substrate by MOCVD, the buffer layer was 25 nm, and the Mg-doped layer was about 2 μm thick, The hole concentration after activation as determined by room temperature Hall measurement was 1.12×10^{16} cm^{-3},the carrier mobility is 21.1 cm^2 V^{-1}s^{-1}. The wafer was cleaned by chloroform—diethyl ether—acetone—ethanol successively and blown dry by N_2. Then, the samples were soaked in aqua regia for 10 min, dipped in DI water for

10 times and blown dry by N_2. The transmission line model (TLM) was defined using standard photolithography, followed by deposition of Ni/Au (30 nm/100 nm) contacts by the ion beam sputtering system. The contact is 250 μm in length and 120 μm in width, the gaps varied from 10, 12, 14 to 28 μm. Finally the samples were annealed in a rapid thermal annealing furnace. The I–V measurements were all performed for the contacts with 10 μm gap. The test system is a computer controlled Keithley 236 measurement unit and the specimen sealed in the Dewar was placed in a metallic shielding box, so as to reduce electromagnetic interference. All tests were conducted at room temperature.

The interface changes of Ni/Au and GaN before and after annealing was observed by SEM, with the acceleration voltage of 25 kV. Auger electron spectroscopy (AES) depth profile analysis was performed using PHI550 ESCA/SAM multifunction photoelectron spectroscopy. An Ar ion beam was used as etching source, the beam diameter was 15 μm, the etching energy was 3 keV, and the etching rate was 35 s/cycles.

As for the formation mechanism of Ni/Au Ohmic contact on p-GaN, we can investigate the following aspects: (i) research on annealing conditions from the electrical perspectives. (ii) observation of the interfacial microstructures using SEM. (iii) study on interfacial components using AES depth profile analysis, and the effects of interfacial reaction between metal and GaN on the Ohmic contact. The transmission line model was defined on GaN epilayer using standard photolithography, followed by deposition of Ni/Au (30 nm/100 nm) contacts. The deposition rates of Ni and Au were 2.7 and 10 nm/min, respectively. Before metal deposition, the samples were soaked in aqua regia for 10 min, dipped in DI water for 10 times and blown dry by N_2, so as to remove surface oxides. Figure 6.57 shows the I–V curves of Ni/Au (30 nm/100 nm) contacts on p-GaN before and after annealing in N_2 and air at 550 °C for 5 min.

As shown in Fig. 6.57, the Ni/Au contact on p-GaN exhibits rectifying characteristics before annealing, the resistance can reach 10^{10} Ω at zero bias, the I–V curves change after annealing in N_2 and air, and the resistance rapidly decreases to

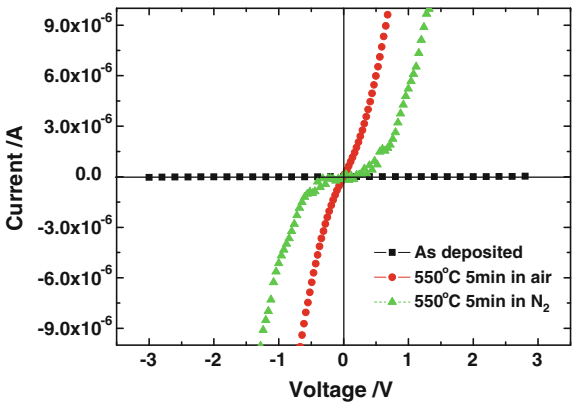


Fig. 6.57 The I–V curves of Ni/Au contacts on p-GaN before and after annealing in N_2 and air

10^5 Ω at zero bias. The sample annealed in N_2 does not show a linear I–V curve, while the sample annealed in air tends to exhibit a linear I–V curve. This phenomenon indicates two points: first, annealing is a key factor accounting for the formation of Ohmic contact, which may induce the change of doping concentration in GaN and interfacial reaction between metal and GaN. Second, the annealing ambient affects the formation of Ohmic contact. As seen from the experimental results, compared with the sample annealed in nitrogen, the I–V curves of samples annealed in air tend to be more linear, indicating the participation of oxygen during annealing is beneficial for the formation of Ni/ Au Ohmic contact on p-GaN. Then, we utilize SEM and AES depth profile to study the formation mechanism of Ni/ Au Ohmic contact on p-GaN.

As calculated from the AES results, we can obtain the depth distribution of atomic concentration for Au, Ni, O, Ga and N, as depicted in Fig. 6.58.

In Fig. 6.58a, the O doesn't exist at p-GaN surface before annealing, indicating that it is an effective method to remove surface oxides by aqua regia before metal deposition.

The content of Au in GaN is almost negligible, the observed Au element may be an error caused by the Ar ion beam, which is not perfectly perpendicular to the sample surface during the striking process whereas in the sample, the Au, Ni, GaN interfaces show a clear distribution, which is consistent with the SEM results. As seen from Fig. 6.58b, after annealing in air, the mutual diffusion of Au and Ni is very significant, resulting in diffusion of Au into the interfacial region completely.

In terms of the thermodynamic parameters, during the generation process of Au-Ni solid solution and NiO, the Gibbs free energy changes are listed in Table 6.6 [82]. As shown in the table, after annealing in air, it is easier to generate NiO than Au-Ni solid solution. In Fig. 6.58b, Au atoms diffuse to GaN interface, and Ga atoms diffuse to the metal layer. Considering the Ga atomic diameter (1.35 Å) is smaller than the diameter of Au atoms (1.44 Å), Ga atoms have a high solid solubility in Au, so the reaction products at the interface are mainly a mixture of Au-Ga solid solution and NiO.

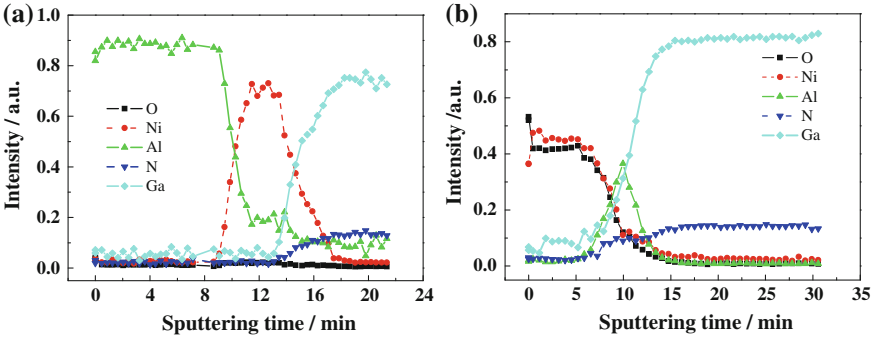


Fig. 6.58 The AES depth distribution of different atoms in sectional Ni/Au contact on p-GaN **a** before annealing, **b** after annealing in air for 5 min

Table 6.6 The changes of NiO and Au-Ni solid solution' Gibbs free energy

Ni alloys	$\Delta G_f a$ (kJ/mol)
Au:Ni(3:1)	−0.6
Au:Ni(1:1)	+4.3
Au:Ni(1:3)	+1.3
NiO	−83.7

Combining current–voltage measurements with AES experimental results, we studied the formation mechanisms of Ni/Au Ohmic contact on p-GaN. (a) First of all, it is worth noting that, after annealing, either in N_2 or air, judged from the I–V curves, the resistance at zero bias decreases significantly, indicating that the annealing process can effectively eliminate the H atoms in p-GaN, such that Mg-H complex at the metal/GaN interface decomposes. This is in accordance with earlier results that the Mg-H complex can be decomposed at 300 °C and higher temperatures which may cause the release of H atoms from the surface [83, 84]. The annealing-induced decomposition of the Mg-H complex can increase the interfacial acceptor concentration, so that the contact resistance is sharply decreased in favor of forming an Ohmic contact. However, the sample annealed in the air tends to exhibit a linear I–V curve, which requires further analysis of the interfacial reaction results. (b) According to the experimental results of AES, after annealing in the air, the Ohmic contact mechanism of Ni/Au on p-GaN can be explained from the conditions of surfaces and interfaces, as shown in Fig. 6.59. Ni atoms diffuse toward the surface, and combine with O atoms in the air to generate NiO at the surface. Jang et al. [85] have reported that NiO can suppress the diffusion of N atoms from the GaN decomposition at 500 °C. As is known, the formation of N vacancies deteriorates Ohmic contacts on p-GaN. Therefore, the generation of NiO can serve to reduce the contact resistance and facilitate the formation of Ohmic contacts.

(c) Judging from the experimental results, Au atoms diffuse to the GaN interface, so that the contact structure becomes NiO/Au/p-GaN, Ga atoms diffuse outwardly, Au-Ga solid solutions form at the interface, and large numbers of Ga vacancies are generated at the GaN surface. Then electrons can be captured by Ga vacancies in p-GaN, thereby enhancing the hole concentration under the contact. Thus, the depletion layer thickness and effective barrier height of the metal/semiconductor

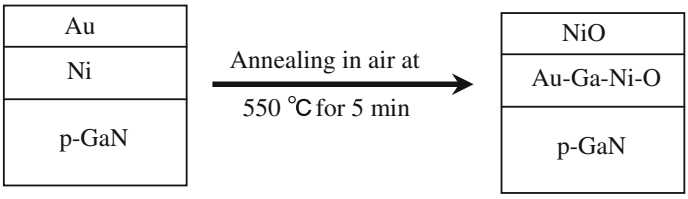


Fig. 6.59 The microstructure change of Ni/Au on p-GaN before and after annealing in the air

contact are simultaneously reduced, and it is beneficial to the hole tunneling and forming of an Ohmic contact.

The current transport mechanisms of metal/semiconductor contacts are mainly: diffusion current, thermionic emission, quantum mechanical tunneling through the barrier of the field emission, and generation and recombination of the space charge region. Wherein, the former two mechanisms depict that the carriers can cross the top of the Schottky barrier and transport from the semiconductor to the metal, and a rectifying contact will form. The latter two processes enable the Schottky junction to deviate from a rectifying characteristic. If the metallic work function is higher than the work function of the p-type semiconductor, there is no potential barrier and an Ohmic contact can form. But for p-GaN of the wurtzite structure used in the experiment, at room temperature the band gap is 3.44 eV, electron affinities is 4.1 eV, and its work function can reach as high as 7.5 eV, and there is no suitable metal with such high work function. The third mechanism contains two cases: in one case, the electrons with energy near E_F of the semiconductor can tunnel through the barrier, known as the field emission and in the other case, the electrons are excited to a higher energy, when the electrons encounter a thin and low barrier, the tunneling probability increases rapidly, called thermal field emission. For the semiconductor process, Ohmic contacts are mostly realized according to the principle of field emission. E_{00} is an important parameter in tunneling effect, known as tunneling probability parameter, and given in the following equation:

$$E_{00} = \frac{h}{4\pi}\left(\frac{N_A}{m^*\varepsilon_0\varepsilon_r}\right)^{1/2} = 1.85 \times 10^{-14}\left(\frac{N_A}{m^*\varepsilon_r}\right)^{1/2} \tag{6.15}$$

Wherein, m^* is the effective carrier mass, ε_r is the relative dielectric constant of the semiconductor, N_A is the acceptor concentration by m^{-3} units. The importance of E_{00} lies in the ratio of kT and E_{00}, which can determine the electron transport mechanism. If $E_{00} \gg kT$, the mechanism is expected to be the field emission, when $E_{00} \sim$ kT, it is the thermal field emission, when $E_{00} \ll$ kT, it is the thermal electron emission [85, 86]. For p-GaN, $m^* = 0.8\ m_0$, $\varepsilon_r = 9.5$, the relationship of tunneling probability E_{00} and doping concentration is shown in Figs. 6.60 and 6.61. With the carrier concentration increasing, the tunneling probability also increases. As seen from Fig. 6.61, when the doping concentration is less than 10^{17} cm^{-3}, $kT/E_{00} \gg 1$, thermionic emission plays a major role. Even if the doping concentration is 10^{19} cm^{-3}, due to the hole mass being large, $kT/E_{00} \sim 1$, the transport properties are due to the thermal field emission. So it is very difficult to realize an Ohmic contact on an ideal surface of the p-GaN.

In the experiment, Ni is used as contact metal for p-GaN. The work function of Ni is 5.2 eV [87], based on Schottky–Mott model, the barrier height of Ni/p-GaN can be written as:

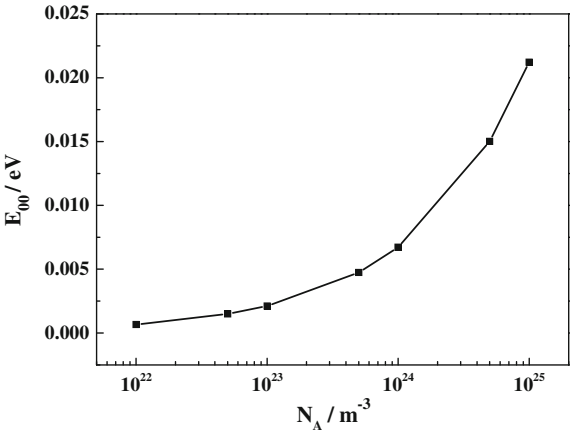


Fig. 6.60 The dependence of E_{00} on doping concentration

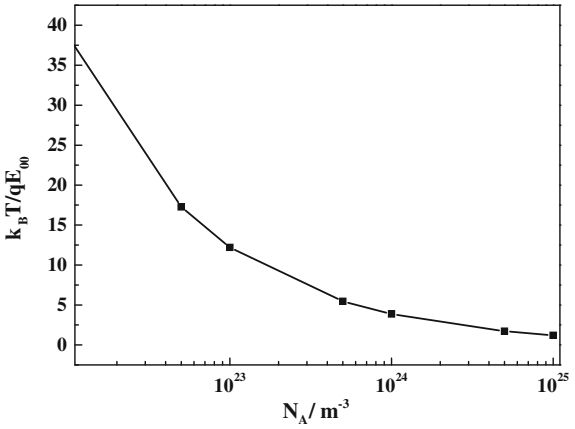


Fig. 6.61 The dependence of kT/E_{00} on doping concentration

$$\Phi_{bp} = E_g + \chi - \Phi_m = 2.34 \text{ eV} \tag{6.16}$$

In fact, for an ideal metal/semiconductor interface with Schottky barrier, thermionic emission is the main transport mechanism. According to the fitting of the I–V formulas of the reverse-biased Schottky diode in thermal emission mechanism, considering the parameters of p-GaN, $m^* = 0.8\ m_0$,[17] when $T = 300$ K, $S = 250*120*10^{-8}$ cm^2, we can obtain:

$$A^* = \frac{4\pi q m^* k^2}{h^3} = 96.1(\text{A/cm}^2 \cdot K^2) \tag{6.17}$$

Figure 6.62 demonstrates the fitting of the I–V curve of Ni/Au contact on p-GaN before annealing and the Ni/p-GaN barrier height may be extracted from the curve. The fitted Schottky barrier height is 0.86 eV with an ideal factor of 1.08. The

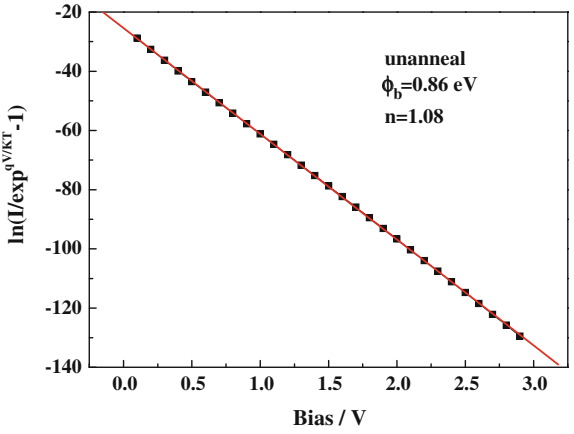


Fig. 6.62 The fitting I–V curve before annealing

interfacial Schottky barrier shows a departure from the Schottky–Mott model, indicating a significant impact beyond the work function of Ni/GaN interface. Since image force cannot reduce a barrier height reduction that high, considering that for the III–V compound semiconductors, surface state density is rather high. Based on the surface state theory by Bardeen, the barrier height of Ni/GaN is about 1/3Eg = 1.15 eV. In addition, the erosion of aqua regia on p-GaN surface will reduce the barrier height on the GaN surface.

After the annealing of Ni/Au contact on p-GaN, the interfacial interdiffusion and phase transitions cause the changes of chemical composition and structure. The interface is no longer a sudden mutation and the generation of interface state changes the band structure resulting in changes of electrical properties. Figure 6.63 demonstrates the fitting I–V curve of Ni/Au contact on p-GaN after annealing in air at 550 °C. As shown in the figure, the ideality factor at the contact interface rises. The ideal factor can be used to evaluate the deviation of current–voltage

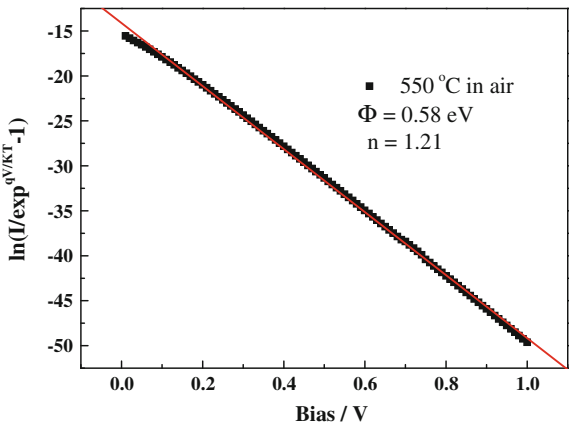


Fig. 6.63 The fitting I–V curve after annealing

characteristics from thermionic emission theory, such a high ideal factor reveals that, after annealing, field emission has become the main current transport mechanism. Meanwhile the extraction of the Schottky barrier height decreases more than that of the sample before annealing, which also confirms the Ga vacancies at GaN surface play a role of improving the interfacial hole carrier concentration, so that the effective barrier height decreases, resulting in the formation of an Ohmic contact.

In order to prepare a low-resistance Ohmic contact, many chemical treatment methods have been tried in the study of Ohmic contacts on p-GaN and may also affect the Ohmic contact on p-GaN. Herein, we investigate the effects of aqua regia and KOH solution on the Ohmic contact to p-GaN.

After the chemical treatments, the metal electrodes were fabricated using the standard photolithography processes. Then, the samples were annealed at 550 °C for 5 min and the resistances of different space gaps were measured. As shown in Fig. 6.64, the contact resistance and sheet resistance can be obtained by the *y*-axis intercept and slope of the linear curve, and the specific contact resistances were calculated. The sample treated with aqua regia provided a specific contact resistance of $1.63 \times 10^{-2}\ \Omega\ \text{cm}^2$, while the contact resistance of the KOH-treated sample was $1.78 \times 10^{-1}\ \Omega\ \text{cm}^2$.

The surface chemical compositions of the samples before and after chemical treatments are investigated using XPS. Figure 6.65 shows the XPS spectra of O 1s, C 1s, Ga 3d level for the untreated GaN sample. As seen from the figure, O 1s and C 1s photoelectrons were detected, suggesting that the O and C coexist at GaN surface. In addition, there is a satellite peak (marked by *) in the Ga 3d spectrum. Since the observed peak is similar to the Ga oxides peak at GaAs surface [88], the peak was identified as Ga_2O_3. The XPS results indicated the presence of Ga oxides and C atoms at the surface of untreated GaN sample.

To analyze the effects of chemical treatments on the GaN surface contamination layer, the chemical composition of the KOH-treated and aqua regia-treated samples was observed by XPS. Figure 6.66 shows XPS spectra of the O1s, C1s core level for the samples before and after chemical treatments. As seen from the Figure,

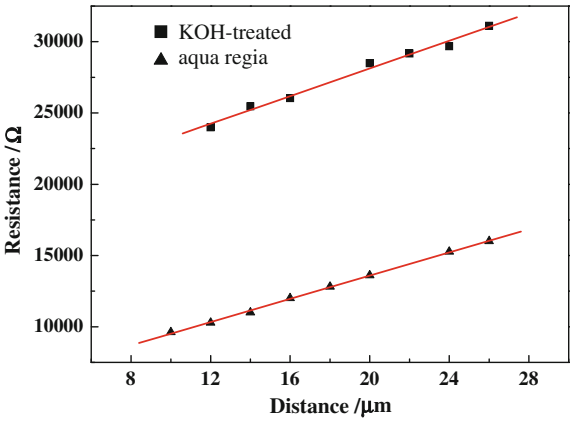


Fig. 6.64 The dependence of contact resistances in TLM on the space gaps for the samples treated with KOH and aqua regia

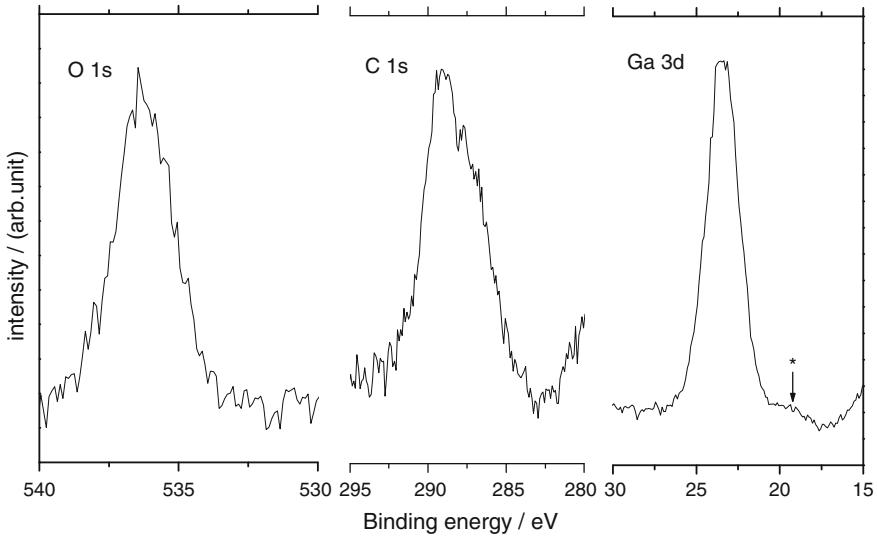


Fig. 6.65 XPS spectra of the O1s, C1s, Ga 3d core level for the untreated sample

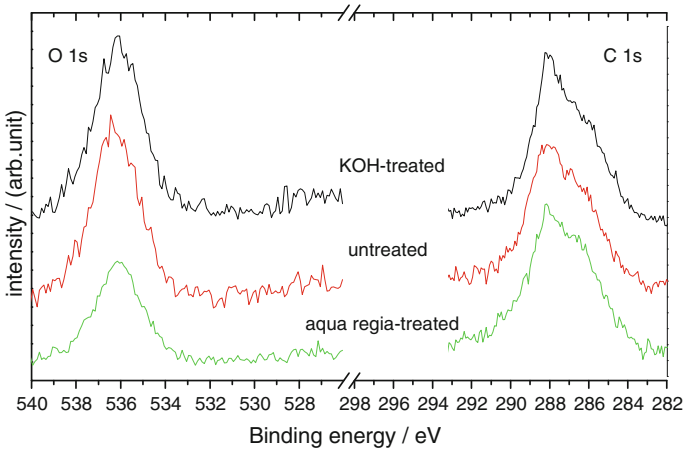


Fig. 6.66 XPS spectra of the O1s, C1s core level for the samples before and after chemical treatments

before and after the two treatments, the XPS spectrum of C1s did not change significantly. The O1s peak intensity of aqua regia-treated sample was significantly reduced. The possible explanation of the C1s spectra is that before being placed in the analysis chamber, the exposing of the sample of cleaned GaN to air led to the regrowth of the pollution layer because Koide et al. [89] have found that the Ar ion-etched GaN surface will soon be contaminated. On the other hand, the O1s peak

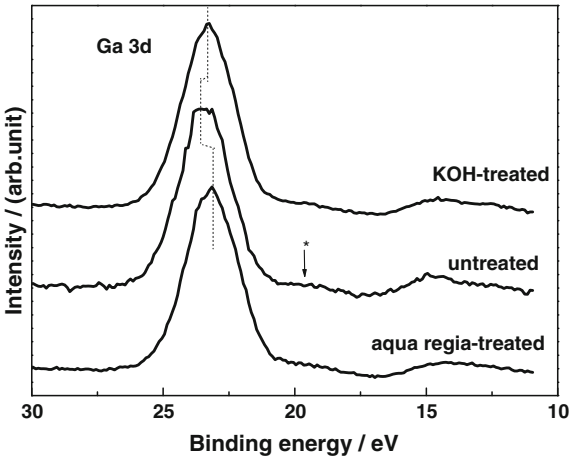


Fig. 6.67 XPS spectra of the Ga 3d core level for the samples before and after chemical treatments

intensity of the aqua regia-treated sample was significantly reduced, indicating that the aqua regia could effectively eliminate the surface oxides, while the elimination of oxide layer with KOH treatment is weaker. These study results concerning KOH solution have some differences from the report by Lee et al. [90]. It should be noted that the ratio and temperature of KOH solution have great impacts on the experimental results, but the solution ratio is not mentioned in the literature.

The surface Fermi level position can be obtained from the GaN peak level. This is currently the most accurate method of determining the Fermi level shift in bending band [91]. As shown in Fig. 6.67, after the aqua regia and KOH solution treatment, Ga 3d peaks move toward the low binding energy by 0.3 and 0.2 eV, respectively, indicating that the surface Fermi level moves toward the edge of valence band (E_V) by the same displacement. The movement of the surface Fermi level in this direction indicates the reduction of surface band bending, and therefore, the barrier height of the samples will decrease.

In the band gap of compound semiconductors, two kinds of energy levels can cause Fermi level pinning at the surface. One is the donor level in the vicinity of the conduction band, the other one is the acceptor level near the valence band. For p-GaN, the surface oxides GaO_x act as acceptor and the induced surface states lead to the Fermi level pinning [92]. For the samples treated with aqua regia, the oxides are removed, the GaN layer with a large amount of Ga vacancies is exposed, resulting in the shift of Fermi level toward the valence band by 0.3 eV, and the interfacial barrier height reduces by 0.3 eV. Koide et al. [89] have calculated that the presence of 2 nm thick oxide layer can improve the barrier height by 0.2–0.3 eV. For KOH-treated samples, the elimination of the oxides is not observed, but the Fermi level moves toward the valence band by 0.2 eV, which is more likely due to the formation of Ga-based hydroxide, [94] and the surface Ga vacancies reduce the barrier height. In summary, the decrease in barrier height reduces the Ohmic contact resistance.

3. Ohmic contact on n-AlGaN

In order to extend response wavelength to solar-blind, a high Al composition of $Al_xGa_{1-x}N$ epitaxial layer is needed and the Al component ranges from 0.10 to 0.65. The effective doping concentration of the epitaxial layer with high Al content is often greatly reduced due to the self-compensating Si-donor [94]. With the increase of the Al component, the metal/semiconductor contact barrier rises and the ratio of contact resistance varies greatly. This makes it difficult to prepare low resistivity Ohmic contacts and study the Ohmic contact formation mechanism. Because high-performance devices require high-quality Ohmic contacts, metal/semiconductor contact, especially in high Al composition of $Al_xGa_{1-x}N$, has become one of the major problems of $Al_xGa_{1-x}N$ device research in recent years [3, 95].

As described above, at the present, specific contact resistance in n-GaN material is typically on the order of 10^{-4}–10^{-8} Ω cm^2. With increasing Al content, specific contact resistance increases rapidly. Ruvimov et al. [96] has studied the microstructure of the AlGaN/Ti/Al Ohmic contact and found that the specific contact resistance was increased from 8.6×10^{-7} Ω cm^2 to 2.07×10^{-3} Ω cm^2, changed nearly three orders of magnitude, when the Al component increased from 9 to 22 %.

Next, we first study Ohmic contact characteristics in AlGaN material with an Al content 0.13 [66]. Research of n-GaN Ohmic contacts is a starting point. The AlGaN epitaxial films were grown in accordance with the same transmission line model of Ti/Al (30 nm/1000 nm) films. The Al content of AlGaN film was 13 %. Samples were annealed in nitrogen with an annealing temperature of 550 °C and an annealing time of 60 s. Figure 6.68 shows IV characteristics curves of two Ti/Al Ohmic contact pads in AlGaN samples before and after annealing. It can be seen that, before annealing, the curve is nonlinear and after annealing, a linear curve appears. This change means annealing helps the formation of Ohmic contacts. Figure 6.69 shows a typical result of testing the relationship between resistance and pitch by the transmission line model. From the figure, it can be calculated that after

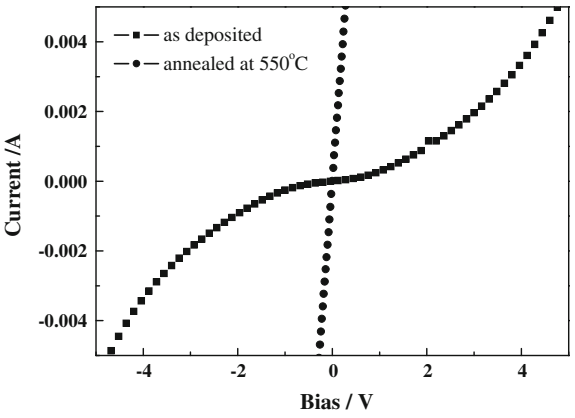


Fig. 6.68 IV curves of Al/Ti/AlGaN contact before and after annealing

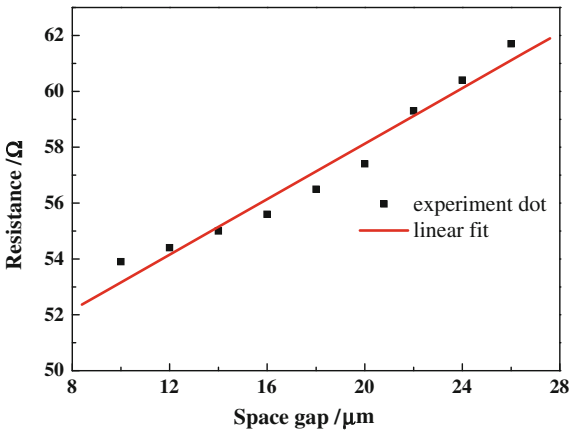


Fig. 6.69 Resistance vs the space gap by the transmission line model

annealing the specific contact resistance is 2.8×10^{-3} Ω cm^2 and the contact resistance R_c is 20.36 Ω.

Then, we study Ohmic contact characteristics in n-$Al_{0.45}Ga_{0.55}N$ [97, 98]. $Al_{0.45}Ga_{0.55}N$ samples were grown on 2 inch (0001) orientation sapphire substrates by MOCVD. Using a Si-doped n-$Al_{0.45}Ga_{0.55}N$ sample, the epitaxial layer has a thickness of about 1 μm with N_D $\sim 2.6 \times 10^{18}$ cm^{-3} and mobility μ_e 25 cm^2 V^{-1}s^{-1}. First, the wafer was cleaned successively with chloroform, diethyl ether, acetone, and alcohol, and then blown dry with nitrogen. A standard photolithography process was applied to define the transmission line model (TLM). A contact layer was sputtered with metal film Ti (30 nm)/Al (90 nm)/Ni (20 nm)/Au (20 nm) using an ion beam sputtering system vacuum of 10^{-4} Pa. After lift-up, a second photolithography was conducted and the sample was then ion etched to the substrate by an ion etching system to insulate the transmission line model area and the surrounding insulating material metal layer. The space gap was varied from 10, 12, up to 28 μm. The samples were rapidly annealed under a N_2 atmosphere in an annealing furnace. The annealing temperature was 400, 550 and 600 °C and the annealing time was 60 s and 30 s. The flow rate of Nitrogen was about 2.5 L/min. The surface morphology of samples in different annealing conditions was observed by SEM. IV testing samples were 10 μm space gaps. A Keithley 236 source measurement unit was computer controlled as testing system. The test sample was sealed and placed in a Dewar which was also put into a metallic shield box in order to reduce interference.

Figure 6.70 shows specific contact resistance at different annealing temperatures and times. Specific contact resistance was calculated from the measured resistance and the contact space gap using the transmission line model method. The results show that the specific contact resistance is affected by the annealing temperature and the annealing time. For the same annealing time (30 s), annealing at 600 °C minimum specific contact resistance; when the annealing time was 60 s, a minimum specific contact resistance of 2.75×10^{-4} Ω cm^2 was obtained at 550 °C.

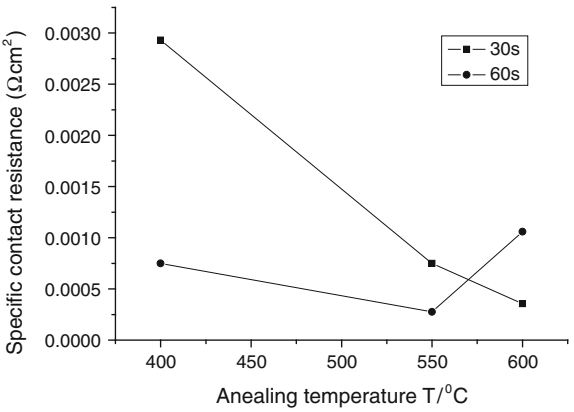


Fig. 6.70 Specific contact resistance at different annealing temperatures and annealing times

To confirm the change in the surface and the interface after annealing, AES was applied to characterize the surfaces and interfaces. Figure 6.71 is the AES spectra figure after the calculation processing. The depth distributions of Ti, Al, Au, Ni, O, and Ga elements are obtained before and after annealing. The presence of an oxide layer was observed on the GaN surface of both samples. An oxide layer on the AlGaN surface can lead to an increased barrier. Before annealing, the interface is clear between Ti/Al/Ni/Au and GaN; and Ti has a small amount of diffusion into GaN. After annealing, the oxide layer becomes thin in the interface, which shows that the introduction of annealing techniques helps eliminate the interfacial oxide, and result in the Schottky barrier reduction. Al diffuses to the surface through Ni and Au; while Al diffuses to the AlGaN surface through a thin layer of Ti and Ti also diffuses to Al. At the interface, Ti, Al, Ga, and N four elements coexist. The formation of Ti–Al alloy helps for conductive Ohmic contact [75, 99, 100].

Ohmic contact characteristics of n-$Al_{0.63}Ga_{0.37}N$ were studied. First, p-$Al_{0.45}Ga_{0.55}N$/i-$Al_{0.45}Ga_{0.55}N$/n-$Al_{0.63}Ga_{0.37}N$ multilayer material was etched to a n-$Al_{0.63}Ga_{0.37}N$ layer with Ar+ ion etching; Second, a Ti/Al/Ti/Au multilayer metal was grown with electron beam sputtering. The Ti/Al ratio was varied in several experiments. Then, multilayer metal electrodes were annealed in a nitrogen atmosphere at different annealing temperatures and times. Finally, I–V testing and Auger electron spectroscopy (AES) were applied to study the formation mechanism of an Ohmic contact.

Figure 6.72 show I–V characteristics before and after annealing. From the figure, it can be seen that before annealing a Schottky contact was formed between metal electrode and semiconductor material. In a nitrogen atmosphere, rapid thermal annealing was conducted under the condition of 700 °C/45 s, 750 °C/45 s and 800 °C/45 s. I–V characteristics were tested and optimum annealing conditions are for 750 °C/45 s. Annealing has a great influence on the metal/ semiconductor contact and the reaction of the alloy tends to reduce the Schottky barrier height.

Figure 6.73 show the element depth profile of an electrode under different annealing conditions. From the figure, it can be seen that the annealing favors

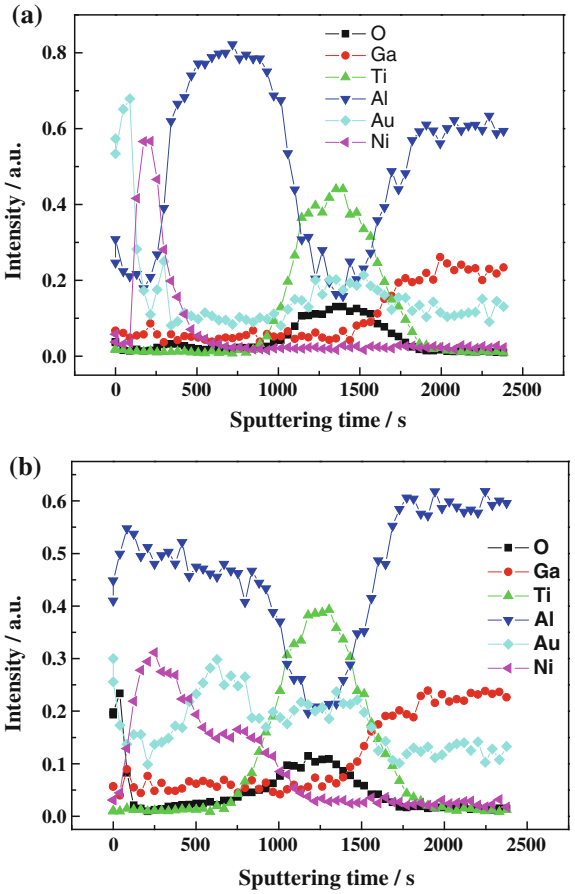


Fig. 6.71 Depth distribution before annealing and after annealing: **a** AES depth distribution before annealing; **b** depth profile of AES while annealed at 550 °C/60 s AES

interdiffusion between the metal layers, and to generate an alloy. The interface between metal layers is not obvious. Multicomponent alloys of Ti–Al-Ga-N-O were formed at the interface of the metal/semiconductor junction. Ti–Al-based alloys at the interface tend to generate Ti–Al-Ga-N multi-alloy and the presence of these reactants was helpful to the formation of an Ohmic contact. But in this experiment, the oxygen elements were present at the interface and had a large proportion. It can be caused by exposure to and being oxidized in the air after the sample was etched to n-AlGaN layer. For this reason, surface treatment was conducted by the use of boiling 5 % KOH solution after the material was etched. Depth profile (AES) show that the surface treatment can effectively remove the oxide and the etching damage layer in the surface, but wet etching increases the surface roughness.

According to the transmission line model, it can be obtained that the specific contact resistance is calculated as 4.9×10^{-2} Ω cm^2 after the multilayer metal of Ti/Al/Ti/Au (grown on the surface of n-$Al_{0.63}Ga_{0.37}N$) was annealed under N_2 at 750 °C/45 s. Thus, suitable annealing conditions are helpful to reduce the Schottky

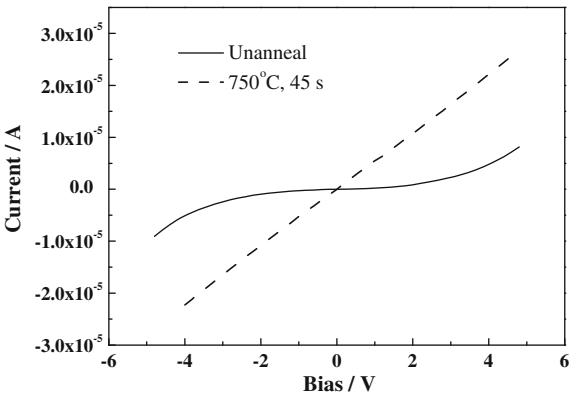


Fig. 6.72 I–V characteristic curves before and after annealing

barrier height and to the interdiffusion in metals. Further, specific contact resistance in high Al content n-AlGaN is indeed several orders of magnitude higher than in n-GaN material or in low Al composition n-AlGaN material. It is just a preliminary result in the study of Ohmic contact in high Al content material. On account of the influence of interface, surface treatment, and electrode structure can greatly improve the Ohmic contact.

4. Ohmic contact on p-AlGaN

Relatively speaking, it is more difficult to realize Ohmic contact on p-AlGaN and the relevant literature show little research for this issue. The following report demonstrates the Ohmic contact on p-$Al_{0.1}Ga_{0.9}N$ [98]. The sample was prepared on a single polished sapphire by MOCVD. First, a low temperature buffer layer was grown on the substrate, followed by the growth of a heavily doped n-GaN layer with a thickness of 2.5 μm and doping concentration of 5×10^{18} cm^{-3}. Then, a undoped GaN layer with a thickness of 0.4 μm and electron concentration of 4×10^{16} cm^{-3} was grown, followed by a p-$Al_{0.1}GaN_{0.9}N$ layer with hole concentration is 3×10^{17} cm^{-3} after activation. Finally, a 15-nm-thick GaN thin layer was deposited to avoid the surface oxidation.

Standard III–V device processes were used for the device fabrication. The sample was treated by the diluted HCl (37 %) solution, and Ni/Au/Ni/Au stacks were selected as the p-type contact. The sample was divided into four parts and annealed under different conditions with the same temperature at 550 °C, and the annealing time is 0, 1, 3, and 5 min, respectively. After the rapid thermal annealing in N_2, the Ti/Al/Ni/Au(30 nm/90 nm/20 nm/20 nm) stacks formed Ohmic contacts on n-GaN. The structure of the p–i–n AlGaN-based ultraviolet detector is show in Fig. 6.74.

The I–V characteristics of the UV detector are measured by a Keithley 236 system. The photoresponse system uses a 300 W xenon lamp as the light source. The modulated light emitted from the xenon lamp enters the monochromator then irradiates the detector. The current signal from the detector is converted into a

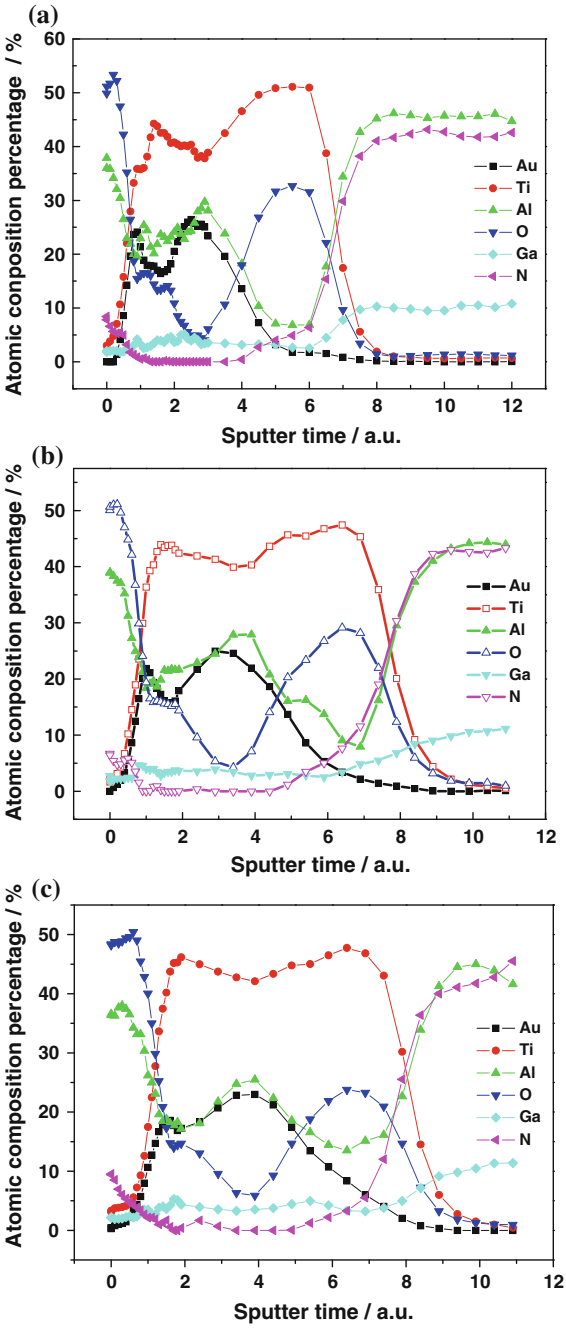


Fig. 6.73 The depth distribution by AES in conditions: **a** 700 °C/45 s; **b** 750 °C/45 s; **c** 800 °C/45 s

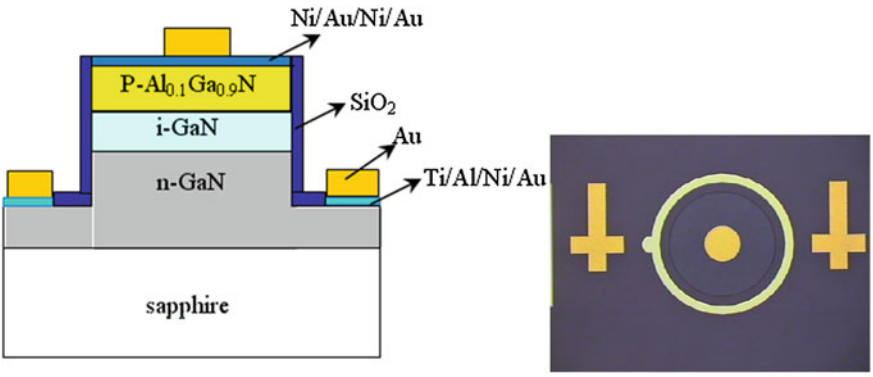


Fig. 6.74 The p–i–n GaN based ultraviolet detector: cross-sectional view (*left*) and top view (*right*)

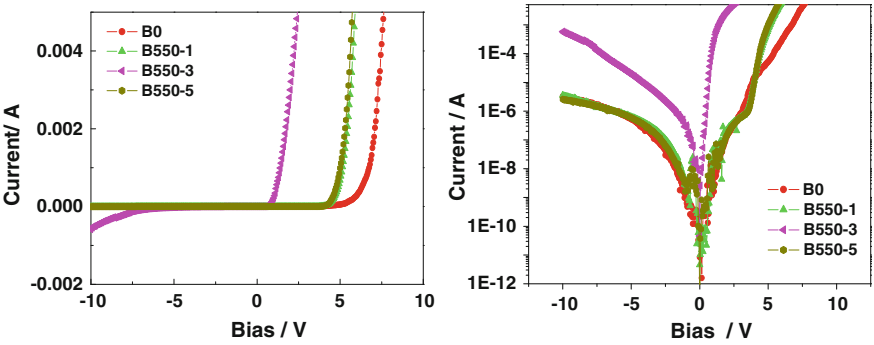


Fig. 6.75 The I–V curves of the p–i–n photodiodes (*left*: linear coordinate. *right* exponential coordinate)

voltage signal by a Stanford 830 phase-locked amplifier. Finally, the result from lock-in amplifier can be read on the computer, and we can obtain the photoresponse spectra after calibrating using a Si detector.

Figure 6.75 shows the I–V curves of the p–i–n GaN photodiodes annealed after different temperatures. It is obvious that the contacts have an effect on the p–i–n GaN detectors. At forward bias, the device annealed at 550 °C for 3 min has the lowest turn-on voltage of 0.69 V. The sample without annealing has the highest turn-on voltage of 5.96 V. The current is 10^{-3} A under −10 V reverse bias, and the current of the other samples is 5×10^{-6} A. The dynamic resistances at zero bias were 1.5×10^{9}, 5.3×10^{9}, 2.47×10^{8} and 2.7×10^{9} Ω.

For the p–i–n GaN detector, when the p-type region is considered as a Schottky junction, the device equivalent circuit diagram can be expressed by Fig. 6.76. D_d is a p–i–n junction, D_j is a Schottky junction, R_{sh} is a parallel resistance of p–i–n junction, and R_{ss} is the series resistance of the device.

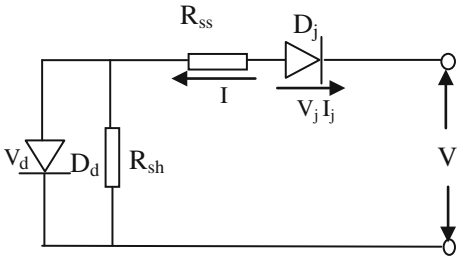


Fig. 6.76 The equivalent circuit diagram of the device

According to Fig. 6.76, the current–voltage equations can be expressed as follows:

$$\begin{aligned} IR_{ss} + V_d &= V + V_j \\ I &= I_{lk} + I_d \\ I &= -I_j \end{aligned} \tag{6.18}$$

I_d and V_d are the current and voltage of the p–i–n junction, respectively. I_j and V_j are the current and voltage of the Schottky junction. Under reverse voltage bias, the Schottky junction is equivalent to a part of the series resistance and the leakage current is the only possible electron transfer mechanism. The leakage current may be expressed as the following equation:

$$I_{lk} = \frac{V - R_{ss}I}{R_{sh}} \tag{6.19}$$

The value of leakage current depends on the series resistance and parallel resistance, which would explain the highest leakage after annealing at 550 °C for 3 min. It is ascribed to the low parallel resistance. When the voltage increases, because the Schottky junction plays an important role, the current is mainly caused by thermal excitation. Thermally stimulated current can be written as the following equation.

$$I = -I_j = -A^* S T^2 \exp\left(\frac{-q\varphi_b}{n_j kT}\right) \exp\left(\frac{qV_j}{n_j kT}\right) \left(1 - \exp\left(\frac{-qV_j}{kT}\right)\right) \tag{6.20}$$

In this equation, $V_j = -V + IR_{ss}$, n_j is the ideal factor. When the voltage is very high, the relationship between current and voltage can be expressed as follows:

$$\mathrm{q}V_j/\mathrm{kT} << 0, \quad \text{so } \exp(\mathrm{q}V_j/\mathrm{kT}) \; << 1 \tag{6.21}$$

$$I \approx A^* S T^2 \exp\left(\frac{-q\varphi_b}{n_j kT}\right) \exp\left(\frac{q(V - IR_{ss})}{kT}\left(1 - \frac{1}{n_j}\right)\right) \tag{6.22}$$

Table 6.7 The fitting parameters of I–V curves

Annealing temperature annealing time	n_j	φb/eV
No annealing	1.13	1.33
550-1	1.12	1.02
550-3	1.19	0.52
550-5	1.13	1.06

$$V = P1 * I + P2 * \ln I + P3 \tag{6.23}$$

$$P1 = R_{ss}, P2 = \left(\frac{q}{kT}\left(1 - \frac{1}{n_j}\right)\right)^{-1}, P3 = -\ln I_{j0} * \left(\frac{q}{kT}\left(1 - \frac{1}{n_j}\right)\right)^{-1} \tag{6.24}$$

The parameters of the metal/semiconductor interface, such as the ideal factor n_j and Schottky barrier height φ_b(SBH) can be obtained from the fitting relationship of the current–voltage equation. The fitting results are shown in Table 6.7.

As seen from the table, for the sample annealed at 550 °C for 3 min, the SBH and the ideality factor of metal/semiconductor contact are 0.52 eV and 1.19, respectively. Low SBH and high n_j show that the tunneling current mechanism at the metal/semiconductor interface plays a major role for the formation of an Ohmic contact. The samples annealed at 550 °C for 1 min and 5 min exhibit higher SBH. For the 5 min sample, the higher SBH may be ascribed to diffusion of metal Ni to p-type contact area. The photoresponse spectrum of the sample is shown in Fig. 6.77. It can be found that the metal/p-GaN contact has no obvious effect on the response spectra of the devices. The samples all have response in the range of 345–365 nm, and the weak response in the range of 250–340 nm is determined by the minority carriers' diffusion to the space charge region, which is caused by p-AlGaN's absorption of UV light.

Then Ohmic contact characteristics on p-AlGaN with Al content of 0.35 are also studied [43, 102]. The $Al_{0.35}Ga_{0.65}N$ sample is also prepared on the 2 in c-plane (0001) sapphire substrate by MOCVD. The p-doping concentration was 5×10^{19} cm^{-3}, the

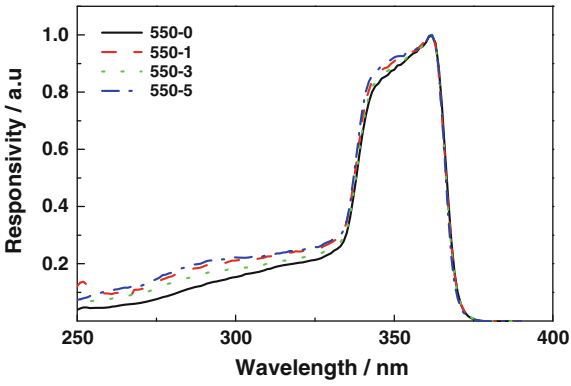


Fig. 6.77 Photoresponse spectra of the samples annealed at different temperatures

hole concentration is about 1×10^{16} cm^{-3} after activation. The wafer was washed by chloroform, acetone, and alcohol in turn, and then dried by nitrogen. The clean sample was soaked in alcoholic hydrochloric acid for 10 min, and then rinsed in deionized water for 10 times. The standard photolithography process was used to define the transmission line model (TLM). The Ni(100 Å)/Au(300 Å) films were sputtered by electron beam evaporation system at 10^{-4} Pa vacuum degree. The electrodes were 250 μm long, 120 μm wide, with space gaps varied from 10, 12, 14 to 28 μm. The sample was placed in the rapid thermal furnace for annealing in N_2 ambient, the annealing temperatures were 400, 500, 550, 600, and 700 °C, respectively. The annealing time was 4 min for all samples and the nitrogen flow rate was all about 2.5 L/min.

Figure 6.78 shows the I–V curves of the contacts under different annealing temperatures at 10 m space gaps. Figure 6.79 demonstrates the measured contact resistance as a function of the contact space gaps under different annealing temperatures. The specific contact resistances are calculated from the measured resistance and contact space gaps using transmission line model. Before annealing, the

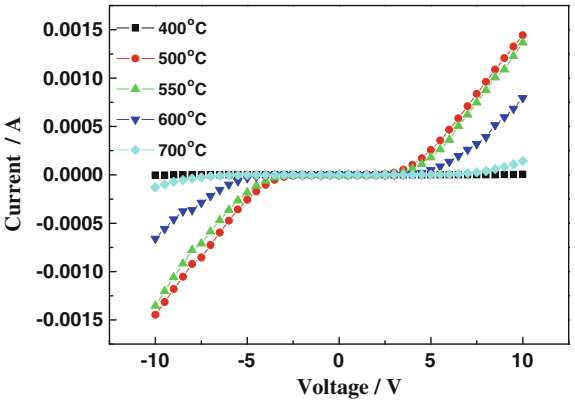


Fig. 6.78 The I–V curves of the samples under different annealing temperatures at 10 μm spacing

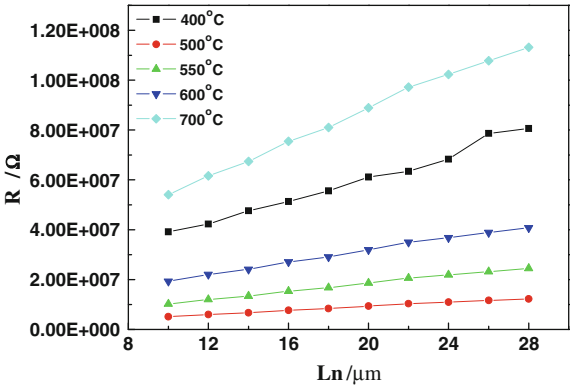


Fig. 6.79 The measured specific contact resistances as a function of the contact gaps in TLM

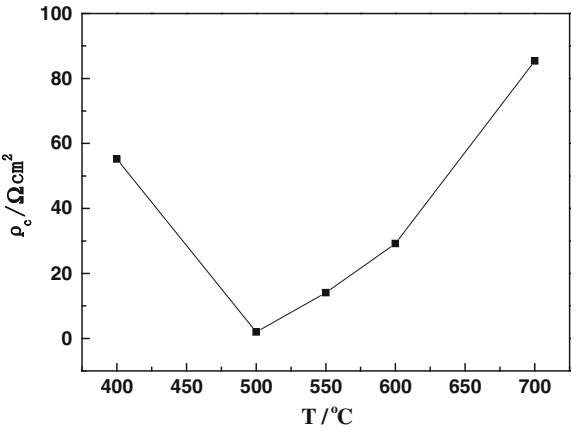


Fig. 6.80 The dependence of specific contact resistances on the annealing temperatures

contact resistance is too large to obtain an I–V curve. After annealing, the results clearly show that the interface of Ni/Au and AlGaN is still a rectifying contact, and the minimum contact resistance appeared at 500 °C.

With the least square fitting method, the dependence of the calculated specific contact resistance on the varying annealing temperatures is shown in Fig. 6.80. The minimum specific contact resistance is 1.98 Ω cm^2 when the annealing temperature is 500 °C. The metal stacks cannot form an alloy at very low temperatures and annealing at rather high temperatures will produce N vacancies at the surface of GaN, which may increase the contact resistance. At present, there are few reports of Ohmic contacts on p-AlGaN with a high Al component. Kim et al. [104] reported that the specific contact resistance on p-$Al_{0.15}Ga_{0.85}N$ can reach as low as 3.1×10^{-4} Ω cm^2.

The fundamental reason for the large contact resistance is mainly associated with the p-AlGaN material. The growth technology of p-type material is still not mature, and it is still difficult to keep doping and activation effective. High p-type contact resistance will seriously affect the device performance, increase the series resistance greatly and affect the response time of the device.

A present, the two main principles followed to prepare metal/semiconductor Ohmic contact are: on the one hand, for semiconductors approximately conforming to simple Mott theory, it is possible to obtain Ohmic contact if we can find a metal with work function lower than n-type semiconductor or larger than p-type semiconductor. On the other hand, the semiconductor layer close to the metal should be heavily doped (i.e., the carrier concentration is high). At this time, the depletion layer is thin enough to realize field emission (i.e., carriers can transit through the barrier by the tunneling effect). Therefore, very low resistance is obtained at the zero bias. Most of metal/semiconductor Ohmic contacts are fabricated according to this principle. However, for p-AlGaN, there are two major obstacles to realize Ohmic contact: First, a lack of metal or alloy system with work functions larger than that of the AlGaN. Second, although the activation energy of Mg (170 or

215 meV) is lower than the other acceptors, it is still too high. Only about 1 % Mg atoms can ionize at room temperature even if the Mg-doping concentration reaches 1×10^{20} cm^{-3}. In addition, Mg can also form complexes Mg-H (i.e., hydrogen passivation effect) with the H in the material, and partly compensate the residual donor impurities, which will further reduce the hole concentration of p-AlGaN. Therefore, the hole concentration of p-AlGaN prepared by methods used currently is difficult to reach 1×10^{18} cm^{-3}. In fact, the hole concentration is too low to achieve the level where a hole can tunnel through Schottky barrier. So it is unable to meet these two conditions mentioned above. It is still difficult to obtain perfect metal/p-AlGaN Ohmic contacts whether in theory or during the fabrication process.

5. Surface treatments and their impacts on Ohmic contact to GaN

In general, the device processing consists of photolithography, chemical treatment, and etching processes. Surface treatments must be carried out after each step so as to remove the oxides and pollutants introduced in the previous step. Surface treatments are usually divided into dry process and wet process, while the dry process makes use of surface plasma. Cl_2, Ar, N_2, etc. are commonly used to generate the plasma gas [107–111]. Wet process refers to the use of a variety of chemical reagents for surface treatment, HCl, KOH and NaOH, $(NH)_4S_{1-x}$ and aqua regia, etc. are commonly used [71, 72, 112]. Many groups have extensively studied the surface treatments in the fabrication process of GaN-based devices [113–115].

(1) Surface treatments and their impacts on GaN

For GaN, the chemical reagents used for the wet process are mainly acid solutions, alkali solutions, and molten alkali or salt, etc. And the general mechanism to oxidize the semiconductor surface, and then remove the oxides from the surface. Early research found that GaN can be corroded in NaOH and other alkali solution, however, the insoluble oxide ($Ga(OH)_3$) may terminate further corrosion because it will cover the surface [116, 117]. Therefore, it requires continuous injection if it needs further corrosion [117]. In the literature [118, 119], researchers used a variety of acid and alkali solutions to corrode AlN and GaN at 75 °C, but it had almost no effect on GaN grown by MBE. But Carosella et al. [120] found that AZ-400 k photoresist at 80 °C could corrode GaN grown by MBE, but could not corrode GaN grown by MOCVD or hydride vapor phase epitaxy (HVPE). These differences may be associated with the polarization or the different dislocation density of the material, as is well acknowledged that the N-pole GaN is more easily corroded than Ga-pole one [121].

Surface treatments have widespread impacts on the Ohmic contact, Schottky contact and the passivation of interface state [71, 72, 122, 123]. However, in the manufacturing process, surface treatments have not been unified. Surface analysis techniques are usually used to study the surface pollutants, oxides, metal particles and surface reconstruction, including auger electron spectroscopy (AES), X-ray photoelectron spectroscopy (XPS), low energy electron diffraction (LEED),

secondary ion mass spectrometry (SIMS). And atomic force microscope (AFM) is often used to observe surface morphology and monitor the surface cleanliness.

Since it is difficult to form good Ohmic contact on p-GaN, the surface treatments become very important. For this reason, many groups have extensively studied the impacts of surface chemical treatments on Ohmic contact and Schottky contact, such as HCl, KOH and NaOH, $(NH)_4S_{1-x}$ and aqua regia solution, etc. [71, 72, 112]. Smith et al. [113] have found that the cleanliness of GaN surface is closely related to the used chemical reagents. Lee et al. [73] reported that the HCl treatment could not effectively reduce the resistance metal contact on p-GaN, but it could decrease by two orders of magnitude when using KOH solution. Sun et al. [72] have investigated the mechanism for the decrease of resistance by two orders of magnitude with KOH solution. Kim et al. [112] used hydrochloric acid, aqua regia, $(NH)_4S_{1-x}$ to treat p-GaN, respectively and they found that it could decrease the contact resistance by three orders of magnitude, and $(NH)_4S_{1-x}$ processing could passivate the surface and reduce the surface oxidation. Huh et al. [124] treated p-GaN with the alcohol-based $(NH)_4S_{1-x}$ solution, finding that $(NH)_4S_{1-x}$ processing could remove the surface oxides and shift the Ga 2p peak toward the valence band, thus improving the properties of Ohmic contact. Lin et al. [125] have explored the mechanism of the reduced Ohmic contact impedance to the p-GaN treated with $(NH)_4S_{1-x}$ by photoluminescence spectroscopy (PL) and surface photo voltage spectrum, finding that a $(Mg_{Ga}\text{-}V_N)^{2+}$ complex on p-GaN surface was changed into a $(Mg_{Ga}\text{-}S_N)^0$ complex after $(NH)_4S_{1-x}$ treatment, which increased the surface hole concentration. Later on, Dialea et al. [126] analyzed he removal of surface contamination with various surface treatments, demonstrating that $(NH)_4S_{1-x}$ solution was the best cleaning agent because of the least amount of C,O pollution, minimum Ga/N ratio and roughness on the GaN surface. And the adhesion of electrodes can also be increased after surface treatment. Miller et al. [127] studied the influence of electrochemical treatment on the Schottky diodes, finding that the electrochemical treatment could decrease the reverse leakage current under 20 V bias by three orders of magnitude.

In recent years, many researchers have studied the impacts of different plasma treatments on the GaN surface electrical properties. Jang et al. [128] studied the effects of Cl_2 inductively coupled plasma (ICP) treatment on n-GaN Ohmic contact, finding that Cl_2 ICP processing induced N vacancies (V_N) could improve the Ohmic contact characteristics. Lee et al. [111] reported the N_2 plasma produced by PECVD could repair the etching damage on n-GaN and recover the electrical performance. Bae et al. [123] treated n-GaN with N_2/He plasma before the deposition of SiO_2 by PECVD, achieving an intensity of interface states lower than the conventional methods by one order of magnitude. Kim et al. [129] have found that the usage of N_2 plasma generated by PECVD under appropriate conditions to treat p-GaN can greatly improve the hole concentration, and increase the electrical conductivity. The main reason is the reduction of N vacancies (V_N), and N vacancies act as the donor for GaN, the reduction of N vacancies (V_N) means the alleviation of compensation effect, so the hole concentration near p-GaN surface increases.

In general, most of the literature reports that whether surface chemical treatments or plasma treatments, are all beneficial to the improvement of the Ohmic contact, the amelioration of the surface electrical properties and the reduction of interface states, but there are some reports considering that surface treatment have impact on the Ohmic contact as well [130]. In addition, in order to study the impacts of photolithography, surface chemical treatments and surface plasma treatments on Ohmic contact to GaN [121], the corrosion process can be used to avoid the influence of photoresist, and XPS technique can be used to study the effects of chemical treatments and plasma treatments on surface chemistry, surface Fermi level, and the Ohmic contact. The experimental results [121] demonstrate that the surface treatments are beneficial to the formation of low-resistance Ohmic contact, as well as the improvement of adhesion and adaptability.

(2) Effect of photolithography on Ohmic contact to p-GaN

The sample epitaxial structure was sapphire/buffer layer (0.49 μm)/i-GaN (0.93 μm)/n-GaN (0.71 μm)/i-GaN(1.2 μm)/p-GaN(0.5 μm). The hole carrier concentration after activation was 1.0×10^{17} cm^{-3} measured by Hall effect at room temperature. The wafer was cleaned by chloroform—diethyl ether—acetone —ethanol successively, then soaked in ethanol-based hydrochloric acid for 10 min, rinsed in deionized water and dried by nitrogen. When defining the transmission line model, the procedures for Sample A1 were photolithography with positive-photoresist-metal deposition-lift off, while the procedures for Sample A2 were metal deposition-photolithography with positive- photoresist-metal corrosion-lift off. The metal contacts of the two samples were fabricated under the same conditions: the vacuum level of the electron beam evaporation system was kept at 10^{-4} Pa, the evaporation rates of Ni/Au ($\sim$20 nm/20 nm) were 1–2 Å/s, 0.5–1 Å/s, respectively. The two samples were annealed simultaneously in the rapid thermal annealing furnace at 550 °C for 5 min with air ambient.

The I-V curves of the contacts with space gaps of 10–28 μm were measured with a computer controlled Keithley 236 test system at the room temperature, and the specimens were placed on a probe station. The samples were also characterized by X-ray photoelectron spectroscopy (XPS) in a Perkin–Elmer PHI-5400 spectrum system with a monochromatic Mg Ka X-rays source. The source energy was 1253.6 eV, the chamber pressure was 1×10^{-7} Torr. The binding energy of Ga3d, N1s, O1s, and C1s were measured simultaneously for each sample.

Figure 6.81 shows the I–V curves of the two samples with different fabrication procedures of the transmission line model. As shown in the figure, both two samples exhibited rectifying characteristics before annealing, while the rectifying feature of Sample A1 was stronger. After annealing in the air for 5 min, Sample A2 showed a good linear I–V curve, but Sample A1 still remained the rectifying characteristics. It indicated that part of photoresist might remain after development, and the residual photoresist had a great influence on the contact resistance. To examine the effects of the residual photoresist on the Ohmic contact, we used X-ray photoelectron spectroscopy for further analysis.

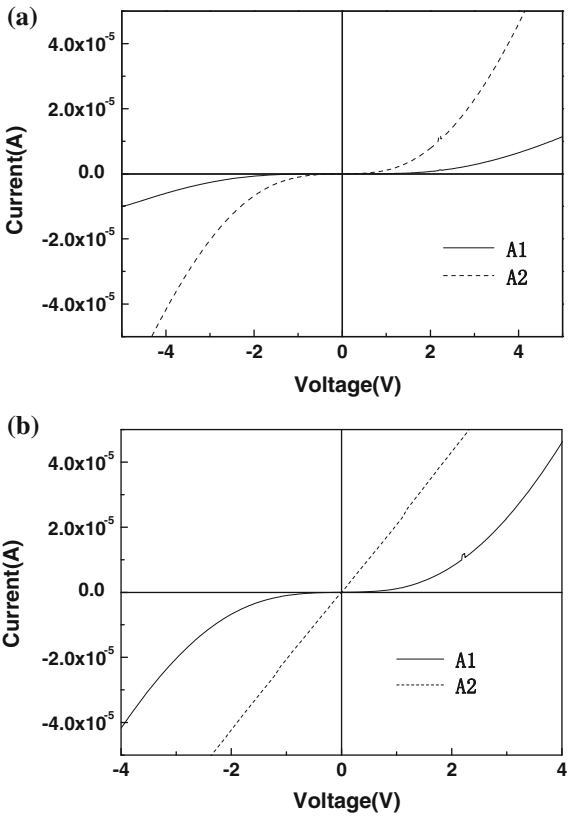


Fig. 6.81 The I–V curves of Sample A1 and A2 **a** before annealing **b** after annealing in air at 550 °C for 5 min

Figure 6.82 shows the XPS spectra of C1s and O1s for the developed wafer before metal evaporation and the wafer without photoresist, respectively. It is obvious that C1s and O1s can be found in these samples. The peak intensity and full width at half maximum of C1s on the wafer increase significantly after photoresist coating and development, indicating that the development cannot completely remove photoresist. But the binding energy of O1s keeps unchanged.

In conclusion, the fact that Sample A1 cannot form Ohmic contact is caused by the following reasons. The existence of residual photoresist after development, which is made of hydrocarbons, can be easily decomposed into carbon and hydrogen after annealing at 550 °C. In GaN, carbon atoms exist as the acceptor, and do not have adverse impacts on the holes, but hydrogen atoms can passivate Mg and form Mg-H complex, thereby, the complex can passivate the acceptor Mg. Then the decrease of surface hole concentration eventually results in the rectifying contact.

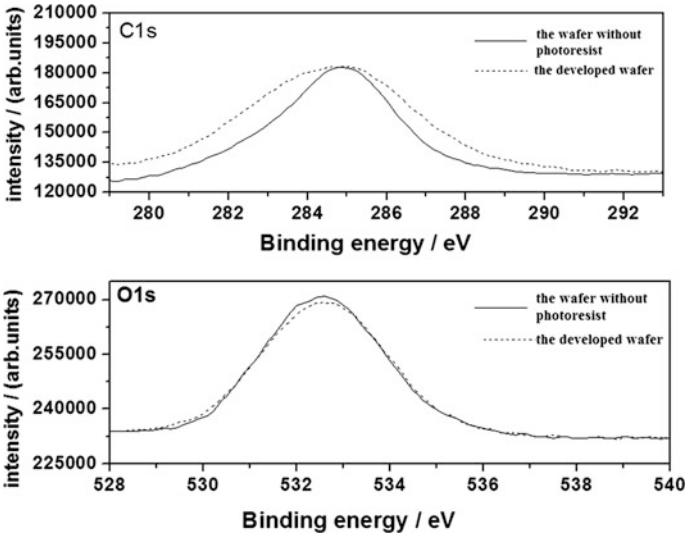


Fig. 6.82 The XPS spectra of C1s and O1s for the developed wafer before metal evaporation and the wafer without photoresist

(3) Effects cause by chemical treatments to P-GaN Ohmic contact

Chemical treatments were applied to the p-GaN surface with saturated H_2S solution. Then clean it with chloroform, diethyl ether, acetone, and alcohol, respectively. The sample B1, was washed in hydrochloric acid for 10 min, washed with deionized water for 1 min, and then dried with high-purity nitrogen. It is called the hydrochloric acid treated sample. In sample B2, take a cleaned sample and dip it into boiling aqua regia for 10 min, wash it with deionized water for 1 min, and then dry it with the high-purity nitrogen. It is called aqua regia-treated sample. In sample B3, take a cleaned sample and dip it into boiling $(NH)_4S_{1-x}$ solution for 10 min, and then dry it use the high-purity nitrogen. It is called $(NH)_4S_{1-x}$ treated sample. In sample B4, take a cleaned sample and dip it into boiling saturated H_2S solution for 10 min, and then dry it use the high-purity nitrogen. It is called H_2S treated sample. All the samples are listed at Table 6.8. The samples were evaporation grown with Ni/Au $\sim$ 20 nm/20 nm using the electron beam evaporate grown system at a vacuum of 10^{-4} Pa at the same time. The growth rate is 1–2 Å or 0.5–1 Å, respectively. Then use the standard photolithography technique to define the transmission line model (LTM). Get rid of the uncovered Ni/Au by photoresist by chemical method. The four samples were annealed in the quick annealing furnace at the same time. The annealing environment was: 5 min at 500 °C. The test conditions of I–V test, X-ray photoelectron energy spectral are same as in the last section. Photoelectron energy spectral is obtained from the stimulation by a 325 nm in He-Cd laser.

Figure 6.83 is the I–V characteristic curve of four different surface-treated samples at 10 um. The specific contact resistances calculated are listed in Table 6.9. It can be seen from Fig. 6.83 and Table 6.9 that, the sample treated by H_2S has the most obvious result.

The reason why oxide makes Ohmic contact worse is that it increases the height of Schottky barrier between metal and p-GaN. Figure 6.84a, b show the energy band between metal and p-GaN when there is a layer of oxide and there is no oxide respectively.

The existence of a certain thickness of oxide between the contact metal and p-GaN can increase the height of Schottky barrier [131]. The relation of oxide thickness and the height of Schottky barrier are shown below:

$$q\phi_B = q\phi_{B0} + \frac{2KT}{\hbar} 2m\chi^{1/2}\sigma \tag{6.25}$$

Table 6.8 Numbers of four surface-treated samples

B1	Chloroform-diethyl ether-acetone-alcohol	10 min in hydrochloric acid	Ni/Au	Annealed in air for 5 min
B2		10 min in boiling aqua regia		
B3		10 min in heated $(NH_4)_2$ S_x		
B4		10 min in heated saturated H_2S		

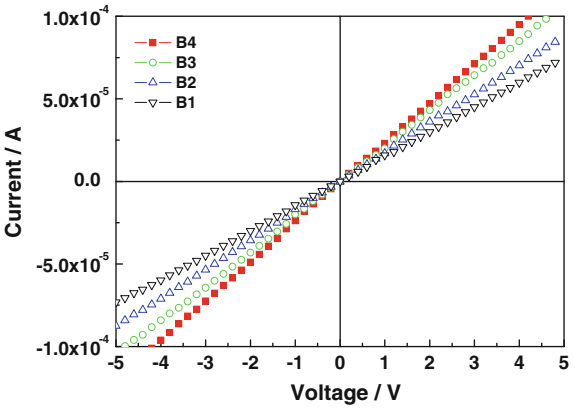


Fig. 6.83 I–V characteristic of four surface-treated samples after annealing

Table 6.9 Specific contact resistances of Ohmic contact of 4 surface treated samples

Sample number	Specific contact resistance/Ω cm^2
B1	8.4×10^{-2}
B2	3.13×10^{-2}
B3	2.13×10^{-3}
B4	2.46×10^{-4}

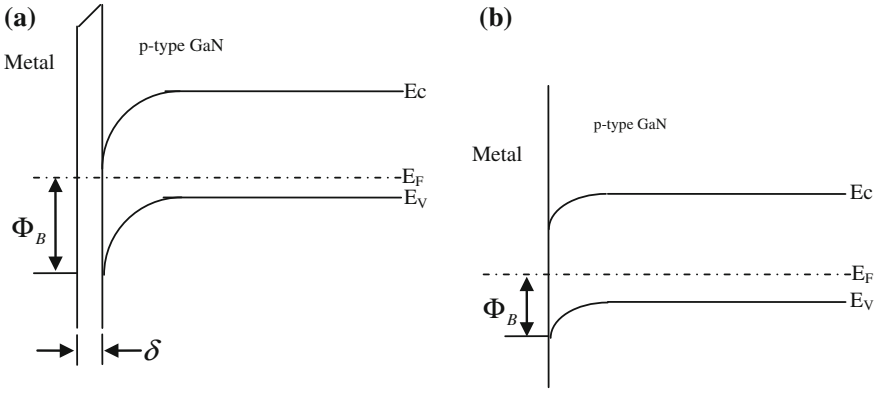


Fig. 6.84 Surface energy band between metal and p-GaN: **a** hydrochloric acid treated sample; **b** saturated H_2S solution treated sample

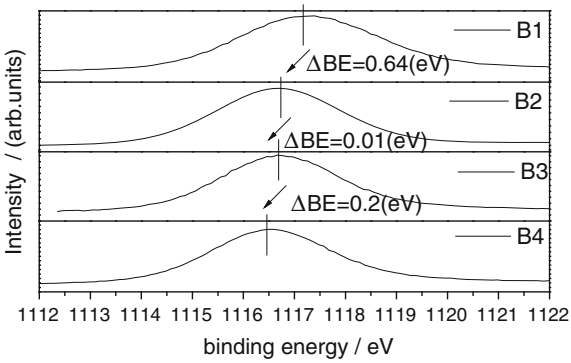


Fig. 6.85 The Ga2p XPS peak of different treated p-GaN surfaces

Here σ is the thickness of oxide, χ is the tunnel barrier the carriers transit from metal to p-GaN, m is the average effective mass of tunnel carriers. The second part at right side of the equation represents the increased barrier height caused by the existence of oxide. It can be seen from the equation that: the existence of GaO_x can widen the depletion layer under contact metal because it heightens the Schottky barrier. The calculation result of K. Hattori, etc. [132] shows that an oxide layer of 2 nm can increase the height of barrier by 0.2 to 0.3 eV.

Generally speaking, testing the movement of Ga2p energy level is a more accurate method to confirm the movement of Fermi level at GaN surface [71]. Then we can detect the surface Fermi level's shift by the movement of GaN-based central energy level peak. Figure 6.85 shows the Ga2p XPS peaks by different surface treatment. The figure shows the movements of Ga2p peak after four surface treatments clearly and the movements are caused by the movements of surface Fermi level. Taking the four Ga2p peaks after different surface treatment into comparison, we can see comparing the Ga2p peak of sample treated by saturated

H_2S solution with the sample treatment by hydrochloric acid, the Ga2p peak of sample treated by saturated H_2S moves to the low binding energy, in other words, the valence band, for 0.84 eV. That lowers the height of surface barrier, and makes the metal contact better.

As is known, the surface sulfur treatment has been extensive researched in GaAs devices. In the process of GaAs corrosion, it is essential to passivate with the sulfocompound. In passivation, sulfide reacts with GaAs and generates other sulfides, which can lessen the reoxidation at the surface. It is generally acknowledged that there are similar treatments in the vulcanization of GaN. The surface of GaN can generate Ga vacancies when it reacts with sulfide. The Ga vacancies act as accepters and make the surface Fermi level move to the edge of valence band. Yow-Jon Lina, etc. form Taiwan researched deeply into the $(NH_4)_2S_x$ treated GaN surfaces, and obtained the results that the $(Mg_{Ga}\text{-}V_N)^{2+}$ (Mg_{Ga}:Ga atoms are occupied by Mg atoms at the $(NH_4)_2S_x$ treated GaN surfaces (VN: N vacancy), then the complex changed to $(Mg_{Ga}\text{-}S_N)^0$ (S_N: N atoms are occupied by S atoms). This reduces the N vacancies and increases the hole density at the surface. But saturated H_2S solution is acid sulfide, has a more deoxidation capacity, and can generate more $(Mg_{Ga}\text{-}S_N)^0$, that can be proved by photoluminescence (PL).

Figure 6.86 shows the PL spectra of sample surfaces treated by aqua regia, $(NH_4)_2S_x$ and H_2S. As the intrinsic absorbing peak of heavy-doping p-GaN is difficult to test, we use the PL spectra near 2.8 eV to compare. It is generally acknowledged that the PL spectrum near 2.8 eV is generated from the transition from deep donor $Mg_{Ga}\text{-}V_N$ to shallow acceptor Mg_{Ga}, which is shown in Fig. 6.87.

We can infer from Figs. 6.86 and 6.87 that, the PL intensity of p-GaN H_2S treated is the smallest, the next is $(NH_4)_2S_x$ treated, and the PL intensity of aqua regia is the greatest. It is generally acknowledged that the intensity at 2.8 eV can represent V_N [125], which means the amount of V_N in the sample H_2S treated is the least. It is same with the movement of Ga2p XPS peak. In other words, using H_2S

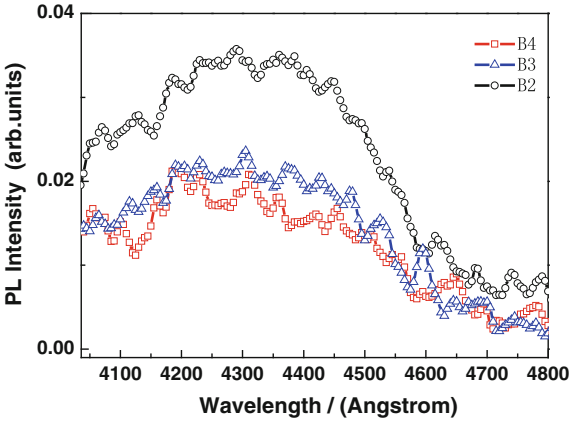


Fig. 6.86 The PL spectral at 2.8 eV after aqua regia, $(NH4)_2S_x$ and H_2S treated at room temperature

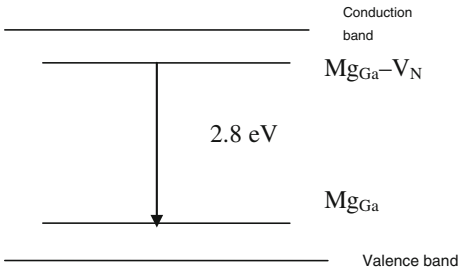


Fig. 6.87 The energy band of deep donor transition to shallow accepter of p-GaN at 2.8 eV

to treat the samples can lessen N vacancies, and accordingly reduce the bend of the surface energy band and lower the height of surface barrier.

(4) The effect of treatment to p-GaN Ohmic contact with nitrogen plasma

In this section, we research the effect of treatment of a p-GaN Ohmic contact with nitrogen plasma by the same p-GaN material. The wafers are washed by chloroform-diethyl ether-acetone-alcohol-based hydrochloric acid. Sample C1, take a clean wafer and wash it with hydrochloric acid for 10 min, wash it for 1 min with deionized water, then dry it with the high-purity nitrogen. It is called hydrochloric acid treated sample. Sample C2, treat it with nitrogen plasma for 10 min or more. It is called plasma treated sample. The plasma generation equipment is a PDC-32G-2 plasma cleaner, which is produced by Harrick in USA. Its power is 18 W. There is no gas flow meter, so we process it with the medium glow. After processing, the two samples are evaporation grown with Ni/Au $\sim$ 20 nm/20 nm using the electron beam system at a vacuum level of 10^{-4} Pa. The growth rate is 1–2 Å, 0.5–1 Å, respectively. Then, we use the standard photolithography technique to define the transmission line model. Then eliminate the Ni/Au uncovered by photoresist with a chemical method. The four samples are annealed in the quick annealing furnace at the same time. The annealing environment is: 5 min at 500 °C.

Figure 6.88 shows the I–V curve of the sample treated by hydrochloric acid before annealing at 10 um distance. It can be seen from the figure that the two treatments show the Schottky contact before annealing, but the sample treated by nitrogen plasma.

Figure 6.89 shows the I–V curve of hydrochloric acid and plasma treated sample after annealing. It can be seen that, after annealing, the two treatments show the Ohmic contact feature, but the sample treated by nitrogen plasma has a better performance. Their specific contact resistance: the sample treated by hydrochloric acid is 8.4×10^{-2} Ω cm^2, the sample treated by N_2 plasma is 3.1×10^{-4} Ω cm^2. That indicated that, treated by N_2 plasma causes better Ohmic contact.

Figure 6.90 shows the XPS energy spectral of Ga2p, N1s, O1s and C1s after two types of surface treatments. It can be seen for Fig. 6.90 that, Ga2p, N1s, C1s of sample treated by N_2 plasma have little change, but the intensity of O1s peak has obviously decreased. That means the oxygen atoms at GaN surface treated by N_2 plasma are fewer than that treated by hydrochloric acid. All the elements densities

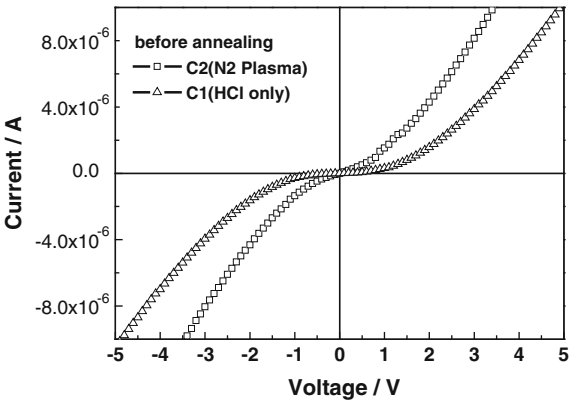


Fig. 6.88 The I–V curve of hydrochloric acid and plasma treated sample before annealing

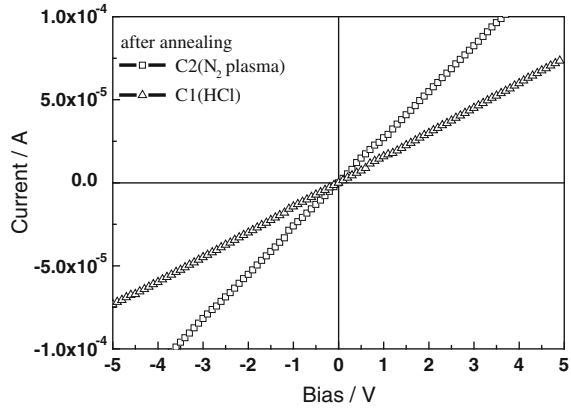


Fig. 6.89 The I–V curve of hydrochloric acid and plasma treated sample after annealing

are listed in Table 6.10. The table shows that, after N_2 plasma treatment, the Ga/N ratio reduced the half-wave width of Ga2p XPS peak the density of oxygen atoms is also lower. Figure 6.90 and Table 6.10 indicate that, N_2 plasma treatment can remove the insulating GaO_x at the surface of p-GaN [90], and the oxide can increase the surface barrier. This shows that, treatment by N_2 plasma is beneficial for the formation of Ohmic contact.

(5) The effects cause by Ar^+ etching to non-alloy n-GaN Ohmic contact

Ti is very important in forming a stable Ohmic contact of n-GaN. Ti layers on GaN surface have good adhesion, and the existence of a Ti layer on the interface also can provide for mechanical stabilization of the contact. This is extremely important to the reliability and repeatability in lift-up of photo resist after photolithography. It is generally believed that after annealing, Ti/GaN will generate TiN. There is no decomposition of the lattice structure as the N element will diffuse out from GaN lattices and form TiN and N vacancies. The N vacancies act as donors in GaN and make the surface area heavily doped. That provides a tunnel

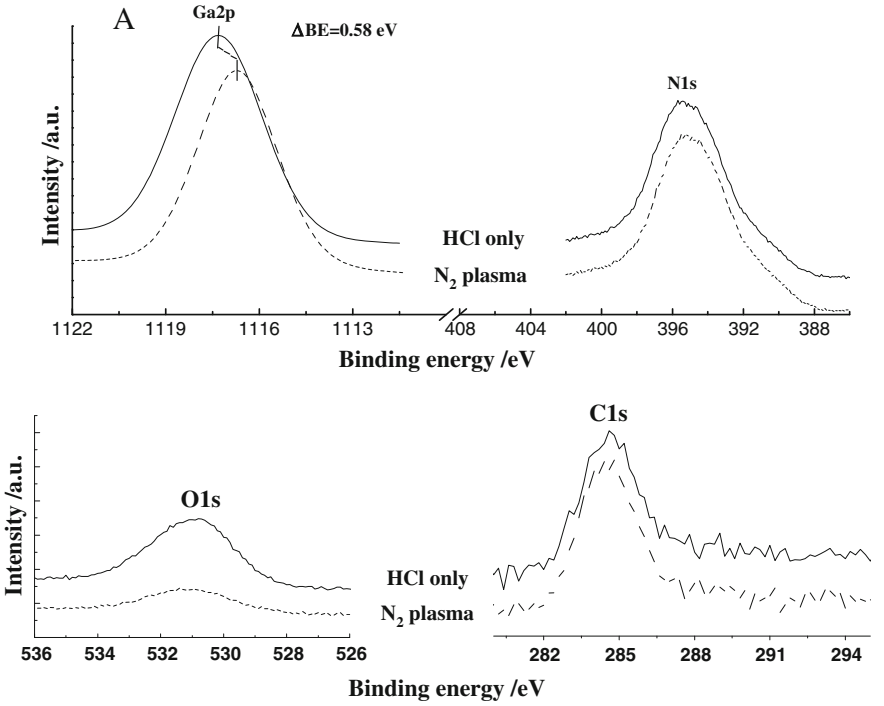


Fig. 6.90 The Ga2p, N1s, C1s and O1s XPS energy spectral after two kinds of surface treatments

Table 6.10 The density of Ga, N, C and O atoms and the half-wave width of Ga 2p peak of p-GaN after two kinds of surface treatments

Elements	HCl only (%)	N_2 Plasma (%)
Ga	17	19
N	53	61
C	12	14
O	18	6
Half-wave width(Ga 2p)	3.24 eV	3.0 eV

current mechanism and promotes the formation of the Ohmic contact. Thin TiN has been observed by the TEM microscope when the metal/GaN interface is annealed in N_2. The interface is abrupt, and only two single layers of TiN can make 10 nm GaN film's doping density to 10^{20} cm^{-3}. Furthermore, the melting point of Ti is 1660 °C and the melting point of Al is 660 °C. High temperature rapid annealing will form $TiAl_3$ and it can guarantee thermal stability, and is difficult to oxidize. But as the top layer, Al metal has the problem of being easy oxidized, so we always use a Ti/Al/Au alloy structure. The Au can protect but also be easy to defuse. To achieve isolation, it always mixed with Ni and Pt. The problems in n-GaN Ohmic contact have already been solved. But when we produce a detector, we should make the

p-GaN Ohmic contact first, next, to form good n-GaN Ohmic contact, the detector should be annealed in N_2 at high temperature, but this will make the performance of p-GaN Ohmic contact worse. This phenomenon is also observed in other experiments designed in this book. But in the process of making an n-GaN Ohmic contact, it is hard to use the reverse photo resist method as in making p-GaN Ohmic contact. Because when Ti/Al alloy is etched, it is inevitable to etch p-GaN Ohmic electrode. Generally, when making n-GaN Ohmic contacts, the high-temperature-annealing process may damage the superlattice and quantum well, that is detrimental to the design of electrical properties. But the non-alloy low contact resistance contact on n-GaN is hard to get in practice. The Ti/Al two-layer structure on surface-untreated GaN has Schottky properties when it is not annealed. One of the preconditions of forming an Ohmic contact is a clean GaN surface. The surface treatments are aimed at removing the oxide and hydroxide, or increasing the density of surface state. But the photoresist cannot bear a long duration chemical treatment, so we usually adopt the method that uses the Ar+ etching to get a fresh GaN surface and to reach the alloy contact.

The GaN film is grown using the MOCVD method on a sapphire ($\alpha-Al_2O_3$) of 2 inches in dimension and (0001) in lattice direction. The n-GaN is Si-doped and the thickness of epitaxial layer is 2.7 μm. The density and mobility of carriers are 1×10^{18} cm^{-3} and 198 cm^2 $V^{-1}s^{-1}$, respectively. The chips are cleaned with chloroform, diethyl ether, acetone, and alcohol, respectively. Then dipped into hydrochloric acid for 10 min, washed with deionized water for 1 min, and dried using high-purity nitrogen. After that, standard photolithography is used to define transport line model (TLM). One sample is bombarded with Ar^+ for 5 min. Then all the samples were evaporated and grown with Ti/Al/Ti/Au (~30 nm/90 nm/30 nm/300 nm) by electron beam evaporation system at a vacuum of 10^{-4} Pa. The growth rate is 1–2 Å, 1.5–3 Å, 0.5–1 Å, respectively.

Figure 6.91 shows the I–V characteristics of etched and unetched sample after annealing with 10 μm gap space. It can be seen from Fig. 6.91 that the sample shows Schottky contact features before etching by Ar^+. But one sample after

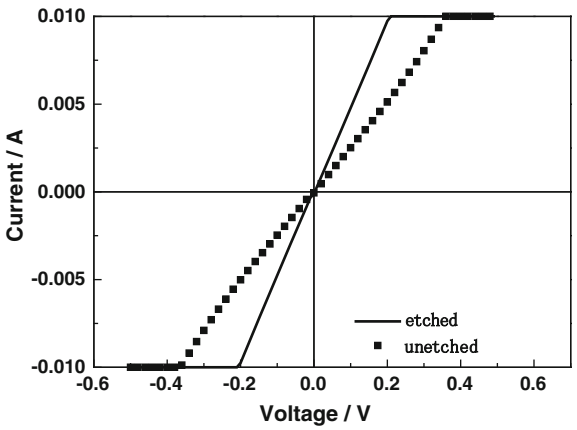


Fig. 6.91 The I–V characteristics of etched and unetched sample after annealing

etching shows Ohmic contact feathers. This indicated that the direct contact between Ti at the interface and n-GaN can reduce the contact barrier effectively and can improve features of an Ohmic contact. We have mentioned before that, in GaN, Ga vacancies act as acceptors and the N vacancies act as donors. The n-GaN after etching by Ar^+ can generate N vacancies, and this process increases the electron density at the surface. Ar^+ etching also can remove the oxide and photoresist, which can increase the height of the barrier at the surface. So the samples etched by Ar^+ shows Ohmic contact. Resistance of n-GaN film is very small, so it is not quite necessary to get a smaller specific contact resistance by annealing. And the annealing process also can create a rough electrode surface, lower adhesiveness, and so on. So the un-alloyed Ohmic contact has some advantages. Similarly, Ar^+ etching can produce pits at the surface of n-GaN, and the pits can bury Al into the n-GaN and improve the stability of contact. In conclusion, Ar^+ etching is a feasible and approving choice.

6.5 Irradiation Effects of AlGaN Ultraviolet Detectors

The irradiation effect of ultraviolet detectors has become a practical and urgent issue as increasingly the use of this type of devices can be found in more and more extensive applications. The improvement of irradiation hardening performance has also been a significant project for detectors operating with long term reliability in a serious and challenging environment with pervasive irradiation.

The irradiation environment includes nuclear irradiation, outer space, high-energy physics experiments, natural environment irradiation, and irradiation induced by processing technologies [137]. More and more human activities have appeared in deep space due to the rapid development of space technology. The high-energy particles which are active in space will be harmful even fatal to electrical and electronic systems and this issue must be seriously considered by scientists and engineers. Irradiation hardening electronics has become an independent subject in both defensive and commercial fields. Among many kinds of semiconductor materials, GaN has attracted more attention for its being more immune to irradiation due to its larger energy band gap than SiC, GaAs, and Si. Many studies have been carried out on GaN-based materials and devices to evaluate the effects caused by irradiation [138–152].

In the following sections, a systematic investigation will be described about the irradiation effects of GaN-based materials and detectors [153–157]. First, the displacement effect of irradiation will be discussed by Raman spectrum measurement of GaN-based material irradiated by different particles with different dosages. Second,the ionization effect was studied by irradiating GaN MIS structures passivated by different dielectric layers with different dosages. Third, the failure mechanism caused by various particle and dosage irradiation on the frontside illuminated GaN p–i–n UV detectors was investigated. And lastly, a preliminary

study of irradiation effects of AlGaN UV detectors was carried out and the results were presented.

Electron irradiation was carried out using a pulsed ILU-8 accelerator, and the energy of the electron was 0.8 MeV, and the dosages were 5×10^{15}, 5×10^{16}, 5×10^{17}, and 5×10^{18} n/cm^2, respectively. And all the irradiation experiments and measurements were carried out at room temperatures.

6.5.1 Proton Irradiation Effects

1. Effect on the Raman spectrum of GaN materials caused by proton irradiation

The energy of the proton is 2 MeV in the experiment and the dosages are 5×10^{14}, 2×10^{15}, and 1×10^{16} n/cm^2, respectively. The Raman spectra were measured after proton irradiation with different dosages, as shown in Fig. 6.92. The Raman scattering measurement was taken using a Jobin Yvon LabRam-INFINITY Raman micro spectrometer, and the wavelength of exiting light was 514.5 nm from an Ar^+ laser. The configuration of the experiment was back scattering with exciting light normally incident on the sample surface.

As shown in Fig. 6.92, the E2 mode of Raman spectra of GaN basically remained stable after proton irradiation, and meanwhile the A_1(LO) mode moved in the low frequency direction slowly as proton irradiation dosage increased. Figure 6.93 clearly demonstrated the different shift of these two modes caused by proton irradiation. As Kozawa et al. mentioned [158], the A_1(LO) mode of Raman spectrum shifted in the low wavenumber direction as the doping concentration decreased. Also Wang et al. found that red shift of the A_1(LO) mode might occur due to neutron irradiation [159]. The carrier mobility and concentration might be obtained by fitting A_1(LO) mode spectra of GaN material subjected to proton irradiation of different dosages, wherein the physical parameters were used in the way as mentioned before. The fitting results were demonstrated in Fig. 6.94.

According to the fitting results, it could be found that the carrier concentration decreased as the proton dosage increased, which were 6.85×10^{16}, 4.08×10^{16}, and 3.14×10^{16} cm^{-3}, respectively. However, the mobility of the carrier showed a capricious variation with dosage, wherein the mobility values were 199, 212.8, and 202.7 cm^2 V^{-1}s^{-1} which were also smaller than that of before irradiation.

2. The proton irradiation effect of PL (Photoluminescence) spectrum of GaN

PL spectra of the GaN sample exposed to proton irradiation were demonstrated in Fig. 6.85. Proton energy was 2 MeV and dosages were 5×10^{14}, 2×10^{15}, and 1×10^{16} n/cm^2, respectively. The irradiation and PL measurement were carried out at room temperatures. The PL spectrum measurement was taken using a Jobin Yvon hr 800 uv micro Raman PL spectrometer which was made in France. The exciting light was produced by a He-Cd (325 nm) laser, and the light power was

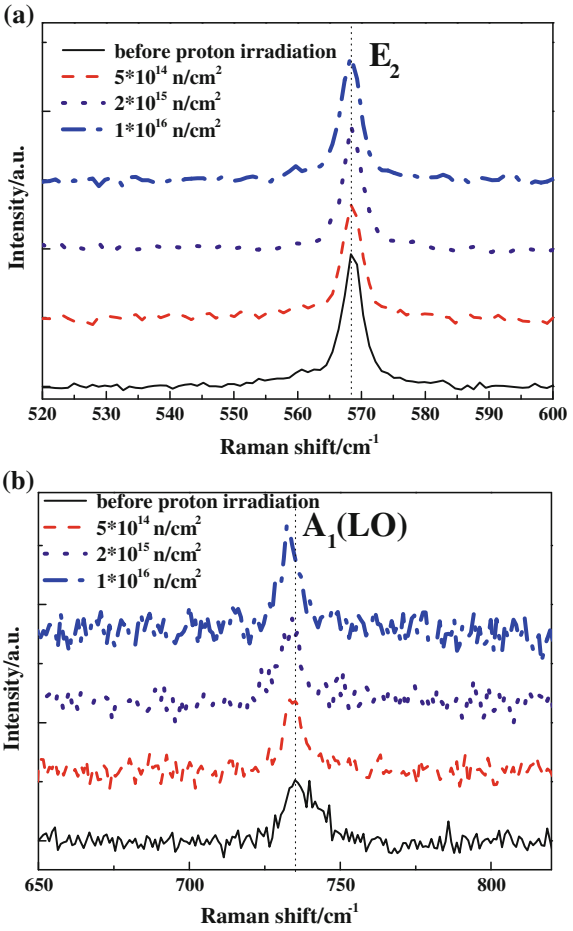


Fig. 6.92 Raman shift of GaN subjected to proton irradiation of different dosages: **a** E_2 mode **b** A_1(LO) mode

approximately 5 mW. Figure 6.95 was displayed in logarithm scale in order to demonstrate the variation of PL intensity clearly caused by proton irradiation. Every piece of the GaN sample was measured using the same condition and cleaved from the same large GaN material.

As shown in Fig. 6.95, the PL intensity due to irradiation at 3.4 eV decayed drastically as the proton dosage increased, and the yellow light intensity at 2.2 eV was also found reduced with an increase of proton irradiation dosage. It could be concluded from the experimental results that a minor shift happened to the main peak of PL spectrum in the short wavelength direction caused by proton irradiation, which was is also called "blue-shift." Meanwhile, new minor peaks appeared marked at both lower energy end of 3.36 eV and higher energy end of 3.45 and 3.54 eV.

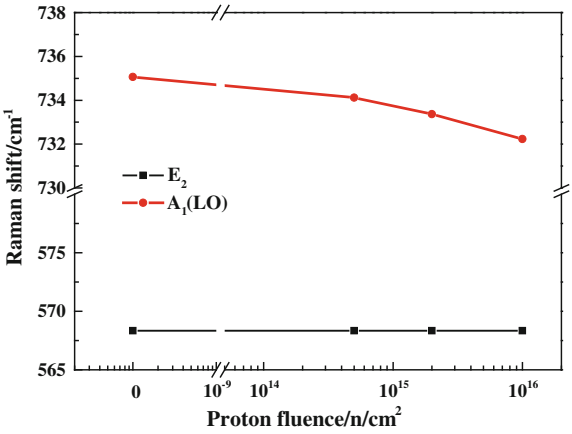


Fig. 6.93 Influence on the Raman peak of GaN subjected to proton irradiation of different dosages

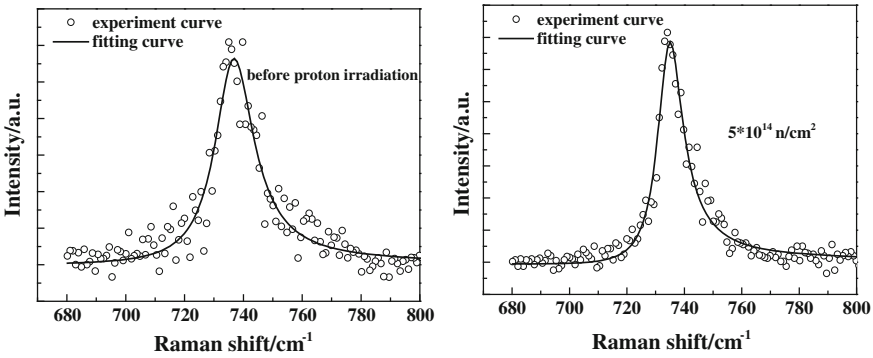


Fig. 6.94 Fitting results of A_1(LO) mode of Raman spectra for GaN subjected to proton irradiation of different dosages

6.5.2 Electron Irradiation Effects

1. The electron irradiation effects of PL(photoluminescence) spectrum of GaN material

Electron irradiation experiments of different dosages were performed for GaN material. It was found that the main peak intensity of PL spectra remained nearly stable and only a minute change occurred when the samples were irradiated with small dosages. However, it decreased significantly as the electron dosage increased, and this also happened with the yellow light PL intensity, as displayed in Fig. 6.96a. It could be seen that a blue-shift happened to the position of the main

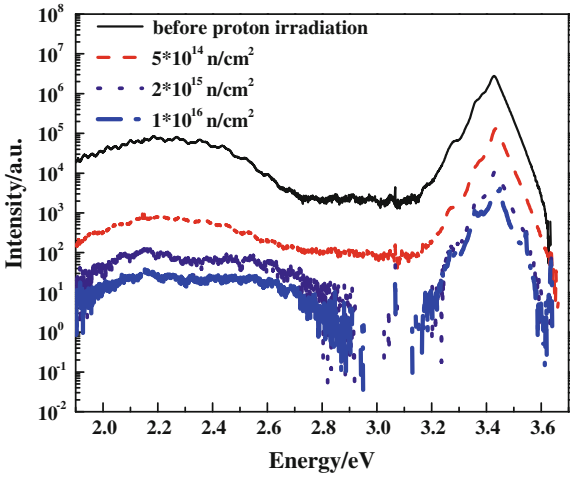


Fig. 6.95 PL spectra of GaN subjected to proton irradiation of different dosages

peak by normalizing the PL spectrum curves. And a weak new peak appeared at the position of 3.36 eV due to the biggest irradiation dosage, which was illustrated in Fig. 6.96b.

Several groups at home and abroad have taken a lot of studies on the irradiation effects of GaN materials by means of DLTS, PL, and Raman scattering spectrum [90, 160–162]. Researchers from Hong Kong University found that the A_1(LO) mode of the Raman spectrum moved in the low frequency direction due to neutron irradiation, and a blue-shift also happened to the main peak of the PL spectrum at low temperatures [159]. The PL intensities of the main peak and the yellow light peak decreased and increased, respectively caused by neutron irradiation, which was somewhat similar to the phenomena revealed by proton irradiation of GaN materials. Reduction of the carrier concentration could be concluded as a consequence of the blue-or red-shift in the Raman spectra, which was the well-known majority carrier removal effect caused by the displacement effect of irradiation.

The new peak appeared at the position of 3.36 eV due to irradiation in the long wavelength direction could be explained by the new generated traps caused by the irradiation displacement effect rather than the recombination of excitons because the bondage energy of the exciton of GaN material is only about 27 meV. New luminescence-active centers at 3.37 eV might be generated due to the high-energy electron irradiation which was also found by Buyanova et al. by means of irradiation of GaN materials using 2.5 MeV electrons [160]. But they did not give an explicit explanation for these defects. A new defect level which was 0.07 eV below the conduction band was found by Look et al. by means of variable-temperature Hall measurement of GaN materials irradiated by 0.7–1 MeV electrons, and they related this new energy level to the N vacancies caused by irradiation. According to Look's results [92], it could be inferred that the new PL peak at 3.36 eV was also the result of the N vacancies caused by irradiation. The PL spectra remained unchanged due to low electron dosages and this was due to the fact that GaN

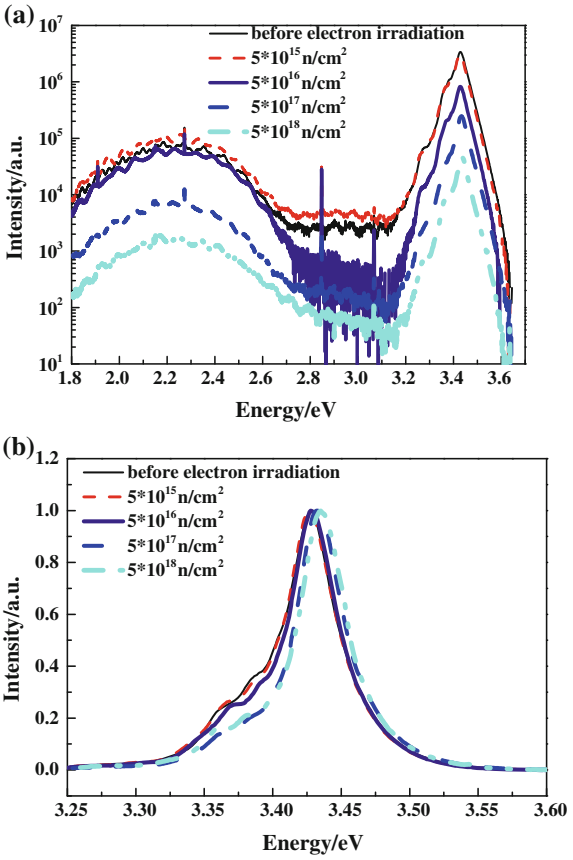


Fig. 6.96 Room temperature PL spectra of GaN subjected to electron irradiation of different dosages

materials share an excellent nature of irradiation tolerance. The new luminescent peaks at 3.45 eV and 3.54 of the PL spectra in the high-energy direction might be interpreted as luminescence of metastable states of defects which were related to the interstitial-vacancy pairs caused by irradiation.

It could be inferred that the crystal lattice quality of GaN was degraded due to irradiation from the reduction of the main PL peak intensity resulting from the variation in the recombination mechanism in GaN material due to irradiation, which also explained the attenuation of yellow light intensity [159]. Although an explicit interpretation of blue-shift of the main PL peak due to neutron irradiation was not given by Wang et al., this phenomenon could also be attributed to the changes of interaction among different atoms in GaN materials induced by the defects such as N vacancies caused by irradiation, which could be explained by the enlargement of energy band gap as a consequence of the weak changes of the crystal lattice caused by irradiation.

Many defect levels were induced by irradiation in the GaN materials, and this could lead to the result that the donor electrons or acceptance holes would be

excited onto these defect levels instead of conduction band or valence band. Accordingly, a decrease in density of the conduction band electrons or valence band holes occurred due to irradiation, which was in accordance with result achieved by fitting the Raman scattering spectra.

According to the results obtained by Buyanova [160], electron irradiation might widen the PL peak between 0.7 and 1.1 eV. Meanwhile, a new peak of 0.88 eV appeared in the PL spectrum of n-GaN material due to electron irradiation, and it could be interpreted as the transition between the base state and defect states caused by irradiation.

The inner and outer Van Allen Belts around the earth mainly consist of high-energy protons of 30–100 MeV and electrons of 0.4–1 MeV. The energy of the electron irradiation in the experiment was smaller than that of protons and more similar to the real space environment. It could be concluded from the above results that the proton irradiation would cause a more obvious physical change in GaN materials than electron irradiation of the same dosage.

2. The physical characterization of the interface between GaN material and different passivate layers and its irradiation effects

The performance of the semiconductor devices are closely correlated to the semiconductor surfaces, which could effectively affect influence the operational stability and irradiation tolerance of the devices. So it is very important and meaningful to study the semiconductor surfaces in order to improve the performance and reliability of the devices.

The ionization effects of irradiation have been studied through the fabrication of GaN MIS devices passivated by different dielectrics. SiO_2 film is usually used to passivate the surface of GaN detectors. The irradiation tolerance of SiN layer, however, is about 3–4 times of SiO_2 film, and more robust than SiO_2 [163]. So, GaN MIS devices passivated by SiN and SiO_2 have been fabricated to assess the different physical performance of the interfaces by means of electron irradiation at various dosages. C–V characteristics were measured before and after irradiation, and the physical performance changes were evaluated.

Electron irradiation was carried out using a pulsed ILU-8 accelerator. The energy of the electrons were 0.8 MeV, and the dosages 5×10^{13}, 5×10^{14}, and 5×10^{15} n/cm^2, respectively. All the irradiation and C–V measurements were carried out at room temperatures. C–V characteristic curves were measured for each irradiation level, as shown in Fig. 6.97.

The high frequency C–V characteristic curves for electron irradiation at each dosage are shown in Fig. 6.97, and it can be seen that a shift in the direction of negative voltage happened due to the electron irradiation for both SiN/GaN and SiO_2/GaN MIS devices, which meant that positive charges were induced in the passivating dielectrics by the irradiation.

By combining the high and low frequency C–V curves, the variation of interface states in both SiN/GaN and SiO_2/GaN devices caused by irradiation were achieved. The values for SiN/GaN MIS structure were 4.4×10^{11}, 1.11×10^{12}, and

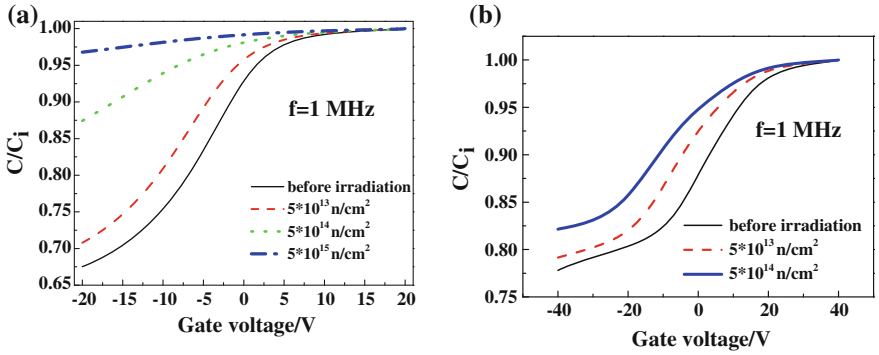


Fig. 6.97 C–V curves of SiN/GaN and SiO_2/GaN MIS devices subjected to electron irradiation of different dosages **a** SiN/GaN, **b** SiO_2/GaN

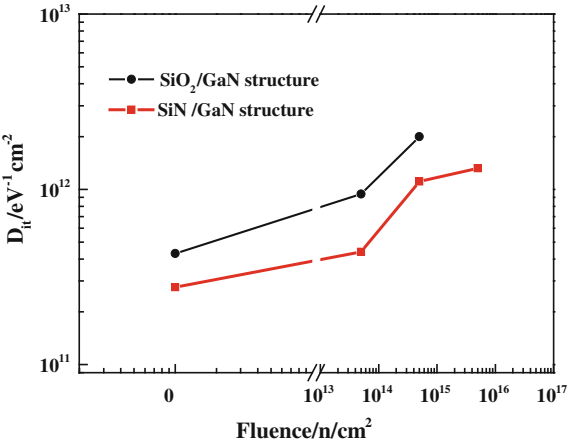


Fig. 6.98 Variation of density of interface state of GaN MIS devices with dosage of electron irradiation

1.32×10^{12} eV^{-1} cm^{-2}, respectively for the dosages of 5×10^{13}, 5×10^{14}, and 5×10^{15} n/cm^2 by use of high and low frequency C–V curves between 20 and −20 V. This led to the result that the interface states of SiN/GaN would increase proportional to the electron dosage (Fig. 6.98).

For the SiO_2/GaN MIS devices, the interface states were 9.4×10^{11} and 2.0×10^{12} /cm^{-2}, respectively for the dosages of 5×10^{13} and 5×10^{14} by use of high and low frequency C–V curves, which also led to the conclusion that the interface states of SiN/GaN would increase proportional to the electron dosage. However, the abnormal results appeared due to the third irradiation level of 5×10^{15}. The capacitance value under the condition of accumulation was smaller than the first two irradiation runs, and the capacitance values of high frequency were much smaller than those of low frequency under the condition of accumulation, as shown in Fig. 6.99.

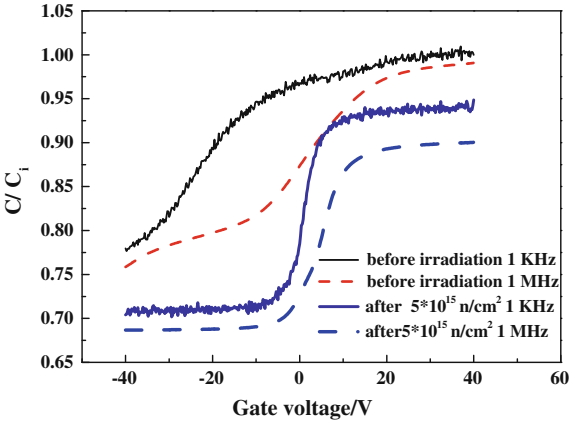


Fig. 6.99 C–V curves of SiO_2/GaN MIS devices before and after irradiation of 5×10^{15} n/cm^2

According to the results of Yang [164], for the MIS devices, if the capacitance at low frequencies in the condition of accumulation was different from that of high frequencies, it might be caused by the thin insulating oxide film along which a tunneling current passed when bias voltages were applied, and the current increased exponentially with the biasing voltage. Another reason to explain the noncoincidence of capacitance of high and low frequencies under the accumulation condition is ignoring the series resistance of the semiconductors [165]. The tunneling current might be traced to the deterioration of the insulation of the SiO_2 layer due to electron irradiation. Through the comparison of the interface states density changing rate due to the electron irradiation with dosages of 5×10^{13} and 5×10^{14} n/cm^2, which was 0.59 and 3, respectively for the SiN MIS device and 1.2 and 3.7, respectively for SiO_2 MIS device, it could be concluded that the SiN/GaN interface is more robust to the electron irradiation than SiO_2/GaN interface.

SiN dielectric film shares excellent tolerance of Na^+ pollution and irradiation. It has been found that the irradiation tolerance of SiN film is related to the depositing techniques. The SiN film usually consists of a certain amount of oxygen atoms, and it could be etched away with HF buffering etchant directly. Zaininger thought that the same mechanism acted when the SiN and SiO_2 were irradiated by electrons [166], and the positive charges could be induced in both dielectrics due to the generation of electron-hole pairs caused by the irradiation. Although most of these electron-hole pairs would have disappeared through recombination, a small number of electrons could leave the dielectrics through the diffusion effect, and leave a corresponding amount of positive charge thereafter. As a result, a shift in the negative voltage happened for the high frequency C–V curves.

Meanwhile, interface states also appeared due to the electron irradiation, and the density increased with the electron dosage. This would not only explain the variation of C–V curves of the MIS devices duet to electron irradiation, but also reveal the deterioration of the detectors through the influence on the passivating layer of the detectors.

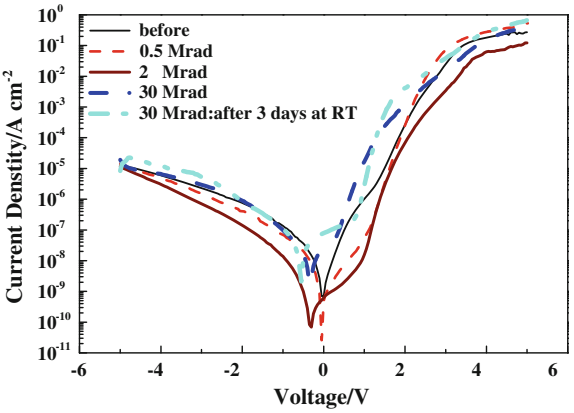


Fig. 6.100 I–V curves of AlGaN Schottky photodiode before and after γ irradiation

The micro-mechanism of the interface state induction due to electron irradiation is that the weak bonds present at the interface were broken by the irradiating electrons [167]. Many dangling bonds were left and new interface energy levels were generated thereafter, which was a complicated process. This micro-mechanism may also be used to explain the interface state generation at the interface between dielectrics and GaN. Up to now, however, few research reports are found focused on the chemical bond types and characteristics of GaN/dielectric interface, so more effort will be still needed to establish an accurate model to interpret the interface states induced by irradiation.

6.5.3 γ Irradiation Effects

AlGaN UV detectors were fabricated to assess their γ irradiation effects. The epi-layer structure was AlGaN/n^+-GaN/sapphire substrate with the carrier concentrations of the AlGaN layer and n^+-GaN were 10^{16} cm^{-3} and 10^{18} cm^{-3}, respectively. The aluminum composition of the AlGaN is 0.23. The fabrication of AlGaN detectors was that of typical mesa devices in which the mesa was dry etched to the n^+-GaN layer with a SiO_2 passivating layer deposited on the surface. The n contact layer was In/Au alloy with the contact hole etched by a photolithography method before the deposition of metal layers. The Schottky contact was achieved by Au deposition before which the contact hole was etched using photolithography as well.

The γ irradiation dosages were 0.5, 2, and 30 Mrad, and the I–V, C–f and responsive spectrum were measured before and after each dosage irradiation at room temperatures. The I–V and C–f curves were obtained using KEITHLEY 236 and HP4194. A respectively, and the response spectra were measured by a computer controlled monochromator with a 300 W xenon lamp as the light source. The current signal of the UV detectors was converted into a voltage signal by a Stanford

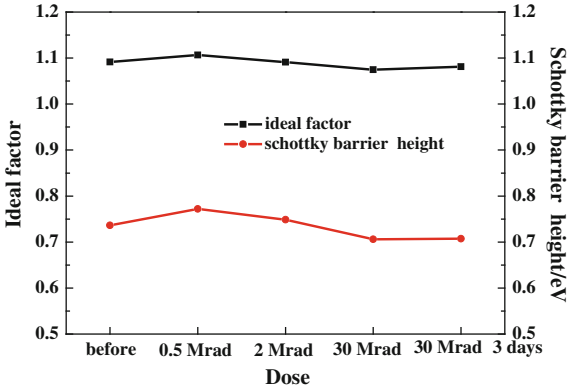


Fig. 6.101 Variation of ideality factor and height of Schottky barrier with dosage of γ irradiation

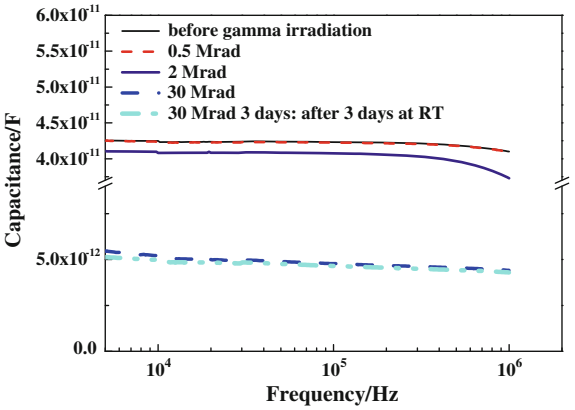


Fig. 6.102 C–f curves of AlGaN Schottky detectors before and after irradiation

830 lock-in amplifier. The I–V curves due to γ irradiation at different dosages are shown in Fig. 6.101, and it can be seen that γ irradiation caused a negligible effect on the reverse-biased dark current.

The ideality factors and Schottky barrier heights due to variable dosages were demonstrated in Fig. 6.101 and showed a decreasing trend with an increase in dosage.

The C–f curves with a zero bias voltage due to variable dosages of γ irradiation were shown as Fig. 6.102.

As illustrated in Fig. 6.102, the capacitance decreased as the frequency increased for the detectors which had not been irradiated by γ ray. The γ irradiation with a dosage of 0.5 Mrad had no impact and the C–f curves remained basically unchanged compared to the samples without irradiation. As the irradiation dosage

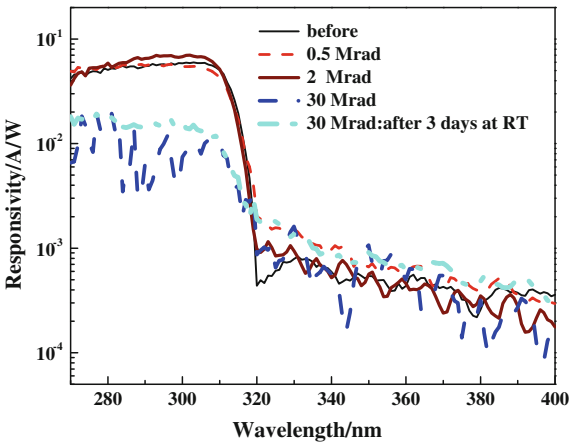


Fig. 6.103 Responsivity spectra of AlGaN Schottky detectors before and after γ irradiation

further increased to above 2 Mrad, the frequency feature began to deviate from its normal shape, wherein the capacitance at low frequencies began to decrease and the amplitude of the reduction was even an order of magnitude for a dosage of 30 Mrad. Continuous annealing at low temperature was conducted for the detectors due to an irradiation of 30 Mrad, and no obvious restoration appeared after 3 days. The fact that the capacitance of the detectors showed a variation when scanned from low frequency to high frequency was attributed to the existence of deep energy levels in the materials. The ionization rate varies with biased voltage if deep energy levels are present in the junction, which leads to the result that the response time of the systems will not make the capacitance vary with a very low frequency. So, the capacitance at a low frequency would be larger than that at a high frequency. This frequency effect on the capacitance is not obvious for the junctions with low density of traps. However, a high density of traps will cause a serious frequency effect on the capacitance. It could be concluded that the deviation of C–f curves was owing to the defect levels generated by γ irradiation.

The response spectra of the AlGaN detectors due to various dosages of γ irradiation were demonstrated in Fig. 6.103. It showed that a radiation dosage of no more than 2 Mrad played no evident effect on the responsivity of the detectors. However, an obvious reduction would be brought to light when the dosage was raised to 30 Mrad, which was illustrated in Table 6.11.

Table 6.11 Ideality factor, Schottky barrier height and peak responsivity of AlGaN Schottky diodes before and after γ irradiation

Dose/Mrad	Ideality factor	Barrier height/eV	Responsivity/A/W
0	1.09	0.74	0.058
0.5	1.11	0.77	0.055
2	1.09	0.75	0.068
30	1.07	0.71	0.008
30(3 days in room temperature)	1.08	0.71	0.015

Defect energy levels act as recombination centers in semiconductors. The excessive carriers generated by incident UV light would be eliminated by recombination through these recombination centers, which could lead to a decrease of the responsivity of the UV detectors. No restoration happened to the samples subjected to the largest radiation dosage after 3 days of room temperature annealing.

6.5.4 *Irradiation Hardening Study of the GaN-Based UV Detectors*

1. Irradiation effects of GaN-based p–i–n UV detectors passivated by different dielectrics

Surface passivation of the semiconductor devices could not only result in a decrease of the surface leakage current and an improvement of the performance and stability but also a hardening effect for the devices operating in a irradiation environment. For high-performance UV detectors, an excellent passivating dielectric layer is needed which could provide a low density of interface states and a high transmittance in the ultraviolet region.

The passivation effect of different dielectrics on GaN detectors could be evaluated by their performance and irradiation tolerance. SiN film could provide a more superior insulativity and irradiation hardening nature than SiO_2. In the following section, the influences in the performance and irradiation tolerance due to irradiation by electrons and protons with various dosages that the different dielectrics (i.e., SiO_2 and SiN) had on the GaN-based UV detectors is discussed and analyzed.

The p-i-n mesa type GaN-based UV detectors were studied. The structure of the epilayer material was as follows: an AlN buffer layer of 20 nm was deposited on the (0001) sapphire substrate at 550 °degrees centigradsse followed by an n-type GaN epilayer of 2.5 μm with a dopant density of 5×10^{18} cm^{-3} at 1100 °C, and an intrinsic GaN layer of 0.4 μm with an electron density of 4×10^{16} cm^{-3} was then epitaxially grown with a following p-type $Al_{0.1}Ga_{0.9}N$ layer of 0.15 μm the hole density which was 3×10^{17} cm^{-3}. The fabrication of the detectors was as follows: After a typical surface cleaning using organ solvents for the epilayer material, the p contact of Ni/Au of 20 nm/20 nm was deposited by electron evaporation with an annealing at 550° C for 1 min in order to form an Ohmic contact. Then the sample was delineated into several mesa pixels by an Ar ion dry etch to the n-type layer followed by deposition of SiN or SiO_2 as surface passivating layer respectively whose thickness was 200–300 nm. Thereafter, the n contact of In/Au alloy was deposited by sputtering method following after a selective wet etching of SiN or SiO_2 with buffered HF etchant.

2. Irradiation effects on GaN-based p–i–n ultraviolet detectors passivated by SiN and SiO_2

SiN and SiO_2 layers were used for surface passivation respectively, and a comparative study of electron irradiation was carried out for the resulting detectors. Electron irradiation was taken using a pulsed ILU-8 accelerator, and the energy of the electron is 0.8 MeV, and the dosages are 5×10^{14}, 5×10^{15}, and 5×10^{16} n/cm^2, respectively.

Eight UV detectors passivated by SiN layers together with the same number of those passivated by SiO_2 layer were irradiated with electrons at the same time, and the performance of these detectors were measured under the same condition. The I–V curves were measured successively after the detectors being exposed to the same irradiation dosage with a successive increase shown in Fig. 6.104. The dark current of the detectors without being exposed to irradiation passivated by SiN layer was 1×10^{-10} A when being biased by −5 V. No obvious change occurred to the dark current when the detectors were irradiated with relatively small dosages of 5×10^{14} and 5×10^{15} n/cm^2. However, an increase of one order of magnitude appeared due to a large dosage of 5×10^{16} n/cm^2. For the detectors passivated by SiO_2 layer, the dark current was 1×10^{-10} A when biased by −5 V without being exposed to irradiation. Slight increases occurred to the dark current when the detectors were irradiated with relatively small dosages of 5×10^{14} and 5×10^{15} n/cm^2, respectively; however, an increase of two orders of magnitude appeared due to a large dosage of 5×10^{16} n/cm^2.

The dark current of detectors passivated by different layers were illustrated in Fig. 6.105 due to electron irradiation of variable dosages when being biased by −5 V. It could be concluded that the threshold dosage due to which the dark current began to increase drastically was one order of magnitude bigger for the detectors passivated by SiN than those by SiO_2.

The responsivity spectra with electron irradiation dosages are shown as Fig. 6.106. It can be seen that the responsivity would decrease from 0.054 A/W to 0.050 A/W for the detectors passivated by SiN layer due to a dosage no more than 5×10^{16} n/cm^2, while from 0.072 A/W to 0.067 A/W for those passivated by SiO_2 layer due to the same dosage. The conclusion could therefore arrive that the electron irradiation had a negligible influence on the responsivity spectra of the GaN-based UV detectors.

It has been mentioned in the Sect. 6.5.2 that the intensity of main peak of PL spectra decreased drastically after the GaN materials were subjected to electron irradiation of the same dosage. The influence of 1 MeV electron irradiation on AlGaN materials was also studied by Hogsed [169], and four new energy levels marked as R1(0.15 ± 0.02 eV), R2(0.21 ± 0.02 eV), R3(0.26 ± 0.02 eV) and R4 (0.33 ± 0.02 eV) were found in AlGaN materials after being subjected to electron irradiation of 9×10^{16} n/cm^2 at 100 K. The energy levels of R1 and R2 were considered to be related to the nitrogen vacancies in AlGaN, and the R4 was unique to the materials subjected to electron irradiation. The detectors which were studied were exposed to electron irradiation consisted of a p-type layer of $Al_{0.1}Ga_{0.9}N$, and

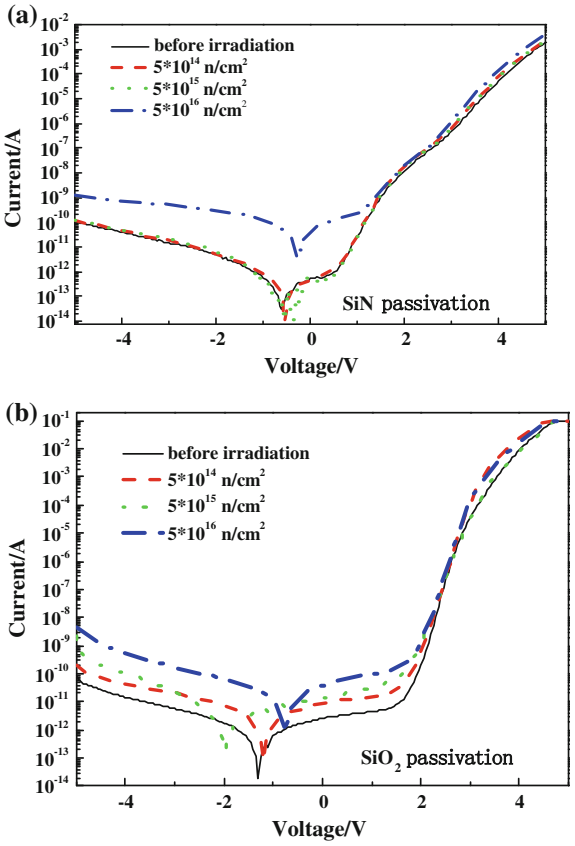


Fig. 6.104 I–V curves of GaN based p–i–n detectors subjected to electron irradiation of different dosages

the increase of negatively biased dark current due to irradiation was attributed to the deep defect energy levels induced by electron irradiation. As recombination centers, the deep levels would act as the source of excessive dark current when being negatively biased by releasing captured electrons and holes before recombination happened, and an obvious increase happened to the dark current as a result. Moreover, it has been manifested that new interface states could be induced between the passivation layers and GaN by electron irradiation studies of SiN/GaN and SiO_2/GaN MIS devices. These new interface states also acted as recombination centers and resulted in the increase of the dark current. The increase of dark current caused by electron irradiation was generally due to the recombination of both failure mechanisms mentioned above.

Although electron irradiation played a prominent role in the decrease of dark current, it exerted a negligible impact on the responsivity of the detectors. Meanwhile, considering the fact that the processes in the fabrication of detectors such as dry etching could influence the performance of the completed detectors by introducing damage or defects to the materials, it could be concluded that only

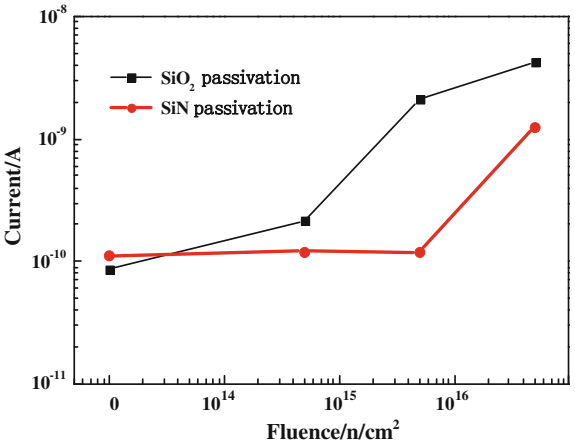


Fig. 6.105 Dark current at −5 V of GaN based p–i–n detectors passivated by different layers subjected to electron irradiation of different dosages

displacement effects of irradiation would lead to an obvious deterioration effect. So, the reason why electron irradiation exerted a negligible impact on the responsivity of the detectors is that the collection of the photo generated carriers was not affected by induced defect levels, and the ionization effect of irradiation played a more important role in spite of the fact that both displacement and ionization effects were present in the irradiation experiments.

3. Proton irradiation effects of GaN-based UV detectors

In order to make a comparison with electron irradiation of GaN-based p–i–n UV detectors, proton irradiation with variable dosages at room temperatures was carried out for the detectors made of the same material and fabrication procedures with zero bias exerted. The energy of the protons was 2 MeV, and the dosages were 5×10^{14} and 2×10^{15} n/cm^2, respectively.

Eighteen detectors were selected for the proton irradiation, and six failed to provide photo signals when being exposed to irradiation of three continual dosages, which meant that 33 % of the total detectors had failed to work. In comparison, the failure rate in the electron irradiation was only 13 %.

In order to make a comparison with electron irradiation of GaN-based p–i–n UV detectors, proton irradiation with dosages of 5×10^{14} and 2×10^{15} n/cm^2 was carried out, and I–V curves were measured before and after irradiation. It was found that an obvious increase happened to the negatively biased dark current when the detectors were irradiated with a dosage of 5×10^{14} n/cm^2, as shown in Fig. 6.107. Although the conclusion could not be arrived at that the irradiation effect of protons on negatively biased dark current was more serious than that of electron, for the dark current of the detectors used for proton irradiation was smaller than that for electron irradiation experiments from the beginning. Without being irradiated, it could still be illustrated that proton irradiation affected the positively biased dark current, wherein the turn-on voltage increased and the current after turn-on decreased with detectors being exposed to irradiation of 2×10^{15} n/cm^2.

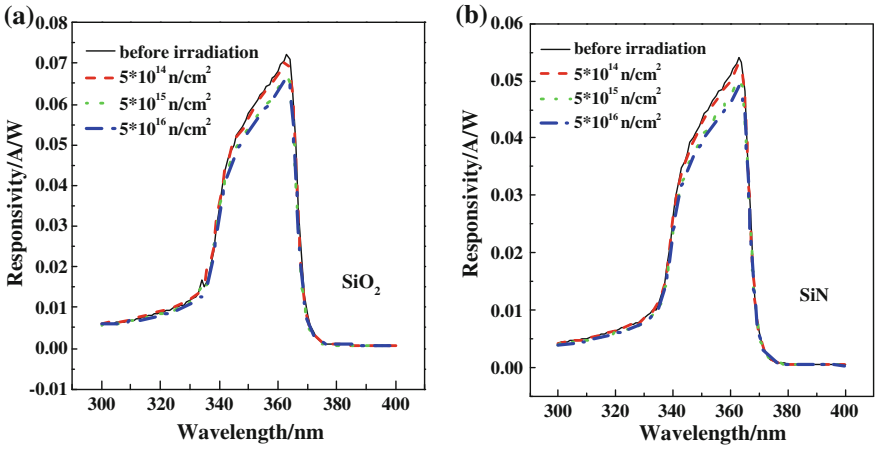


Fig. 6.106 Responsivity spectra of GaN based detectors subjected to electron irradiation of different dosages

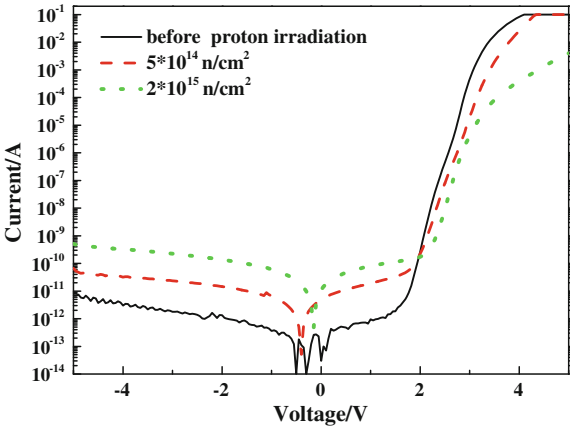


Fig. 6.107 I–V curves of GaN based detectors subjected to proton irradiation of different dosages

In comparison with electron irradiation, proton irradiation exerted a comparable effect on the negatively biased current but a more deteriorating effect on the positively biased current for the same dosage. This might be explained by the fact that the same dosage of protons affected the electrical nature due to decreased conductance and increased series resistance. This was coincident with the result provided in the last section that a decrease of both carrier concentration and mobility happened for the GaN materials being irradiated by protons obtained by fitting of Raman scattering spectra.

The responsivity spectra of the detectors before and after being irradiated are shown in Fig. 6.109, and it can be seen that a successive reduction happened to the responsivity with irradiation dosage being increased. The peak responsivity

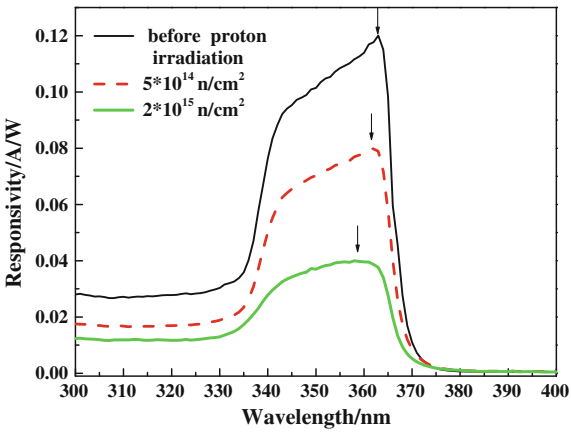


Fig. 6.108 Responsivity spectra of GaN based detectors subjected to proton irradiation of different dosages

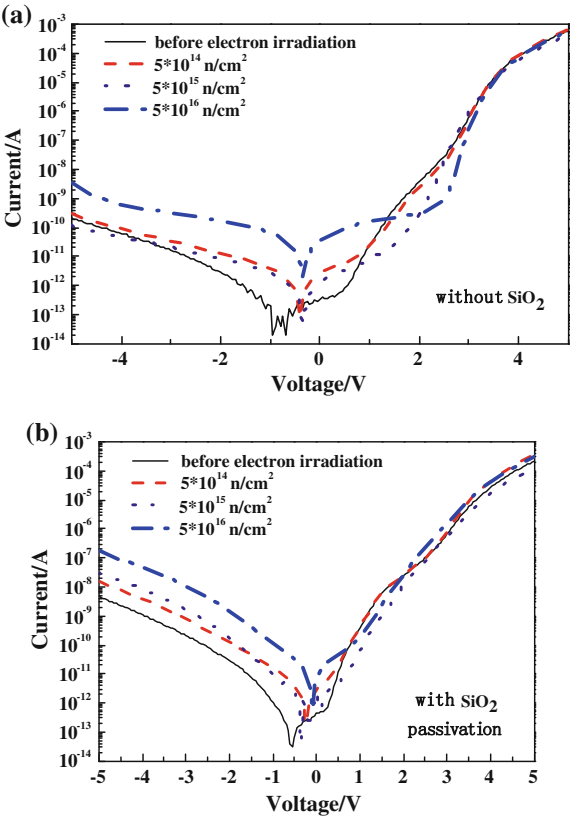


Fig. 6.109 I–V curves of AlGaN p–i–n detectors passivated by different layers subjected to electron irradiation of different dosages

decreased from 0.12 A/W to 0.08 A/W and 0.04 A/W for the detector after being irradiated with dosages of 5×10^{14} and 2×10^{15} n/cm^2, respectively. The responsivity also decreased obviously in the direction of short wavelengths. Meanwhile, the position of the responsivity peak moved slightly toward the shorter wavelength. The same condition was used for the responsivity measurement before and after irradiation in order to improve the reliability of the obtained results and avoid the deviation of results due to the variation of the measurement condition. The responsivity peak of the detectors used for proton irradiation was 363 nm before irradiation. The detectors were divided into two groups for the proton irradiation experiment of the same dosage. The peak position changed after the detectors were irradiated and moved in the direction of shorter wavelength as dosage increased.

As discussed in Sect. 6.5.1, it could be concluded that the reduction of the intensity and blue-shift of the main peak position of the PL spectrum occurred when unintentionally doped GaN samples were proton irradiated with dosages of 5×10^{14}, 2×10^{15}, and 1×10^{16} n/cm^2, respectively. This result might be interpreted from the blue-shift of responsivity peak of GaN-based p–i–n detectors after being irradiated by protons. Moreover, the defect levels induced by irradiation acted as recombination centers and led to a loss of photon-generated excessive carriers through recombination. As a consequence, responsivity would decrease drastically for the detectors being exposed to proton irradiation. It was concluded that both displacement and ionization effects resulted in the failure of the detectors subjected to proton irradiation.

The biggest threat to the spacecraft around the earth is the inner and outer Van Allen Belts over the equator which mainly consists of high-energy protons of 30–100 MeV and electrons of 0.4–1 MeV. The energy of the irradiating electrons and protons were 0.8 and 2 MeV, respectively in the experiments. It could be concluded from the above results that the proton irradiation would cause a more obvious electrical damage in the GaN materials than electron irradiation of the same dosage because the mass of a proton is 1836 times greater than that of an electron. According to Ionascut's results [170], the Rutherford scattering cross section of 2 MeV protons is 214 times larger than 2.5 MeV electrons, the detectors will be more affected by the irradiation of protons than of electrons.

4. Irradiation effects on AlGaN ultraviolet detectors

Up to now most of the studies about the irradiation effects are focused on GaN materials and detectors, and few are about the same effects on the AlGaN materials and detectors. This was due not only to the increased difficulty in acquiring high-quality AlGaN materials but also the limitation of analysis techniques for the materials with large energy band gap such as AlGaN [149, 169]. Recently, a preliminary study on the irradiation effects of AlGaN materials was carried out by a group from Russia led by Polyakov [145, 146]. These studies found that the Fermi level could be pinned to a level of 0.35 eV below the conduction band by fast neutron irradiation according to their results, in which only AlGaN Schottky and

HEMT devices were mentioned. So a gap still exists in the knowledge of irradiation effects of AlGaN materials and detectors, and it is the following section that would fill the gap in which the preliminary research was taken to explore these effects on AlGaN p–i–n UV detectors by means of I–V and responsivity spectrum measurements.

The p–i–n mesa type AlGaN UV detectors were studied. The structure of the epilayer material was as follows: an AlN buffer layer was deposited on the sapphire substrate followed by an n-type $Al_{0.35}Ga_{0.65}N$ epilayer of 0.55 μm, and an intrinsic $Al_{0.35}Ga_{0.65}N$ layer of 0.22 μm was then epitaxially grown with a following p-type GaN layer of 0.15 μm. The fabrication of the detectors was as follows: After the p contact was formed by electron evaporation, the sample was delineated into several mesa pixels by an Ar ion dry etching to the n-type layer followed by deposition of SiO_2 of 200 nm as surface passivating layer by PECVD followed by a selective wet etching of SiO_2 with a buffered HF etchant. Thereafter, the n contact of Ti/Al/Ti/Au alloy was deposited by electron evaporation followed by annealing for 30 s at a temperature of 750 °centigrade in order to form Ohmic contacts. In order to perform an evaluation of the role of the passivation layer in the irradiation effect, two comparative groups of detectors were fabricated of which one was passivated with SiO_2 and the other was not. Electron irradiation was performed using a pulsed ILU-8 accelerator, with the energy of the electron being 0.8 MeV, and the dosages are 5×10^{14}, 5×10^{15}, and 5×10^{16} n/cm^2, respectively.

The measurements and electron irradiation studies were carried out as soon as the detector samples were completed to avoid the possibility that the performance of the detectors might be affected due to their being laid aside for a long time which could misguide the irradiation effect studies. The I–V curves of the detectors before and after being irradiated are shown in Fig. 6.109.

It is illustrated by Fig. 6.109 that the dark current of the detectors being −5 V biased was 2×10^{-10} A and 5×10^{-9} A, respectively for those without a passivation layer and those passivated with SiO_2. The dark current of the former group was nearly one order of magnitude lower than the latter. This was due to the fact that the dark current of the AlGaN mesa type detectors with a fresh surface was fairly low. However, as the exposure time in the atmosphere environment increased, the dark current increased drastically for the reason of the surface being oxidized and polluted, and this provoked the necessity of a surface passivation layer. It can be seen that, for the detectors without passivation, negligible variation occurred to the dark current for the first low irradiation dosages. As the irradiation dosage increased to 5×10^{16} n/cm^2, the dark current was 3×10^{-9} A which was one order of magnitude larger than that without being irradiated. For the detectors with passivation, the dark current was 1.8×10^{-7} A which was also one order of magnitude larger as the irradiation dosage increased to 5×10^{16} n/cm^2.

It could concluded that the physical mechanism behind the increase of the dark current of the detectors being irradiated consisted of not only the influence of the

interface states between the passivation layer and the semiconductor surface which were induced by electron irradiation but also the defect energy levels in the AlGaN materials which acted as recombination centers. It can be concluded that these recombination centers were induced by irradiation results from the fact that a similar increase happened to the dark current for the detectors with surface being passivated not.

The responsivity spectra of the detectors were also measured before and after electron irradiation and the irradiation effects were assessed. The responsivity was small for the reason of frontside illumination and the peak wavelength was 303 nm. The measurement procedure was as follows: First, the power spectrum of the UV light source was obtained by the measurement of a standard Si detector whose responsivity spectrum had been provided by the manufacturer. Then the signals of the AlGaN detectors were tested, and lastly the responsivity spectra per watt were achieved by the calculation of the signals over the spectral power of the UV light source. For the detectors without passivation, a slight decrease in the shorter wavelength direction happened to the signal of the detectors when subjected to electron irradiation of the first two dosages, and a more obvious change happened for the third dosage, which is shown in Fig. 6.110a. It could be seen that a tail-shaped extension appeared in the long wavelength direction after irradiation wherein the curve was low and flat before irradiation due to the wavelength being larger than the cutoff wavelength of the device. The responsivity in the shorter wavelength also became larger. These phenomena were common to the detectors being passivated and uncommon to the unpassivated ones for which the only result was the decrease of signal as shown in Fig. 6.110b. All the spectra presented here were measured as one batch without any variation to the instrumental setup.

A lump in the region of 320–370 nm was also found in the spectra of the detectors without a passivation layer which only happened when the detectors were subjected to irradiation of a large dosage level. It was thought that this was related to the creation of new defect levels in the forbidden band by electron irradiation. These newly generated levels not only influenced the responsivity in the shorter wavelength but also contributed to the responsivity at the corresponding transition wavelength between the defect levels and conduction or valence band. Due to the fact that these phenomena were not found for the detectors without passivation layer, it could be concluded that the SiO_2 played an important role for the irradiation hardening of the detectors in spite of the creation of defect states at the interface between SiO_2 and AlGaN.

Irradiation effects were also studied for GaN detectors without a passivation layer in order to make a comparison with the AlGaN detectors. GaN detectors were manufactured without being passivated and irradiated together with the above mentioned AlGaN detectors. The responsivity spectra were also measured and no obvious change in the shape of the curves occurred which was similar to the result of AlGaN detectors with passivation layer. This result therefore indicated the necessity of the irradiation effect study of AlGaN materials and detectors.

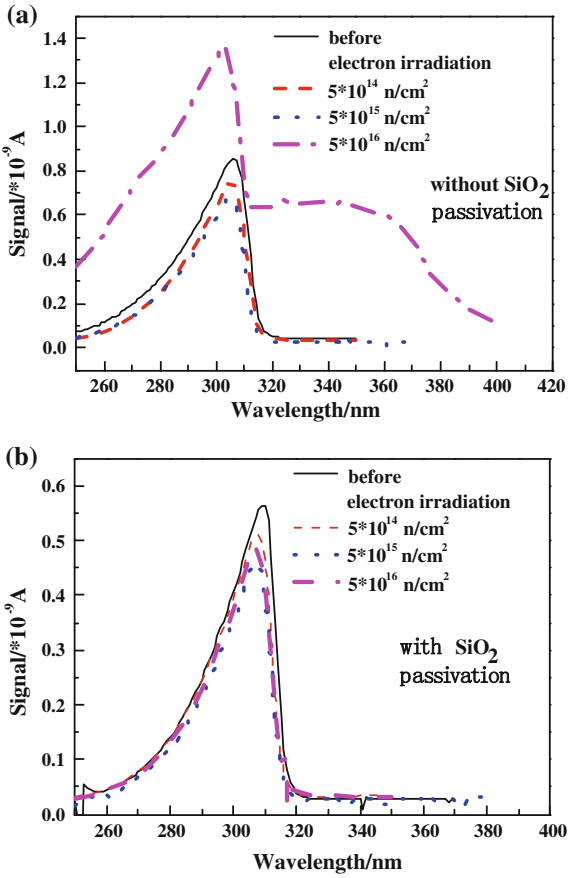


Fig. 6.110 Responsivity spectra of AlGaN p–i–n detectors with and without passivation layer subjected to electron irradiation of different dosages

6.6 Imaging and Application of UV Focal Plane Assembly

UV focal plane assemblies and imaging systems with the response wavelengths covering 240–280 nm, 300–365 nm or both visible- and solar-blind wavebands will be introduced in this section [171, 172]. UV imaging experiments were completed for some near objects and long distant scenes. Additionally, some aerial photographs are displayed.

6.6.1 Imaging of Quartz Tube Heated by Oxyhydrogen Flame

Figure 6.111 shows an experiment of UV imaging for a quartz tube heated by oxyhydrogen flame. Figure 6.111a is a photograph shot by a visible light camera.

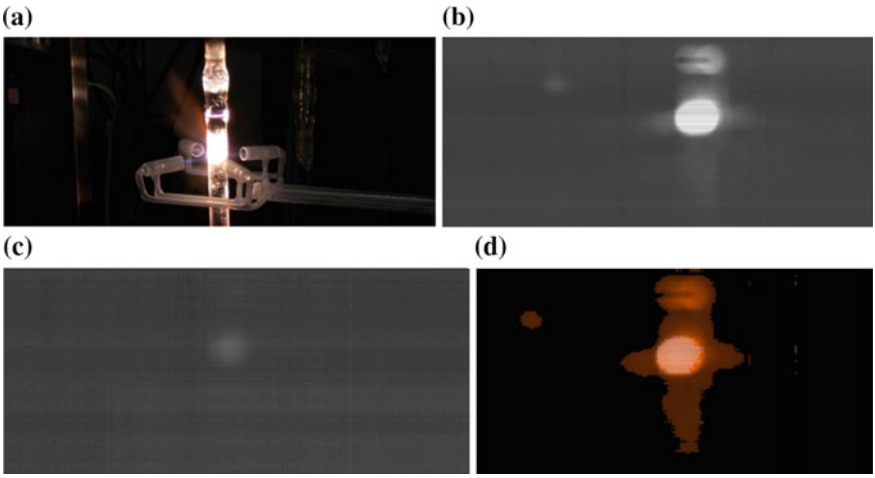


Fig. 6.111 UV imaging experiments for quartz tube heated by oxyhydrogen flame **a** photograph taken by visible light camera, **b** photograph of visible-blind waveband imaging, **c** photograph of solar-blind waveband imaging, **d** photograph obtained by using image fuse technique based on the photographs in (**b**) and (**c**)

Figures 6.111b and c are the photographs, respectively shot by the visible-blind channel and solar-blind channel of a 128 × 1 GaN based, two band focal plane array. The photograph in the solar-blind channel was obtained when the temperature of quartz tube is close to its melting point. At this time, the quartz tube will generate a strong UV radiant flux. Figure 6.111d is a fused image of both photographs shot by visible-blind channel and solar-blind channel.

6.6.2 Imaging in Visible-Blind Waveband

At the end of 2008, a push broom scanner with a 512 × 1 GaN-based visible-blind UV linear detector was made in the Shanghai Institute of Technical Physics, CAS. Figure 6.112 shows a picture of a light rail track and an elevated roadway obtained by using this scanner. The response of the detector in visible-blind waveband was

Fig. 6.112 Visible-blind waveband imaging of light rail track and elevated roadway in Shanghai

realized by selecting a suitable material composition. The optical system of the scanner has high transmission and the detector has high sensitivity. The weight and power of the scanner with a simple structure are less than 9 kg and 9 W respectively.

6.6.3 Imaging of Outside Scene

Figure 6.113 shows a building in clouds as background, which is a result of long distant scene imaging experiment. Shanghai Jinmao Building can be seen in the photograph. Jinmao Building can be seen in visible light camera, as indicated by the circle in Fig. 6.113a and has a distance of 6 km from the imaging system. The image of visible-blind waveband is shown in Fig. 6.113b, Jinmao Building can be also observed clearly. This means that the visible-blind UV detector can be well used for long distant imaging.

6.6.4 Aerial UV Photographs of Oil on Sea Surface

In September 2009, the UV imaging system made by Shanghai Institute of Technical Physics, CAS, was used in the aerial photography experiment. The target is an oil slick in sea area of Bohai, China. The experimental result is shown in Fig. 6.114. The feasibility of UV imaging to the oil target on sea surface was confirmed. A visible light camera with two channels (565 nm for channel 1 with

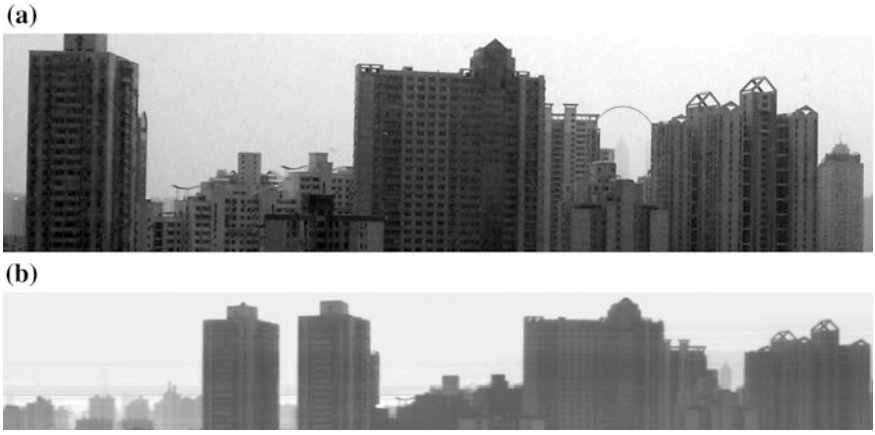


Fig. 6.113 Experimental results of long distant scene imaging **a** *gray* photograph of visible light imaging, **b** photograph of visible-blind waveband imaging

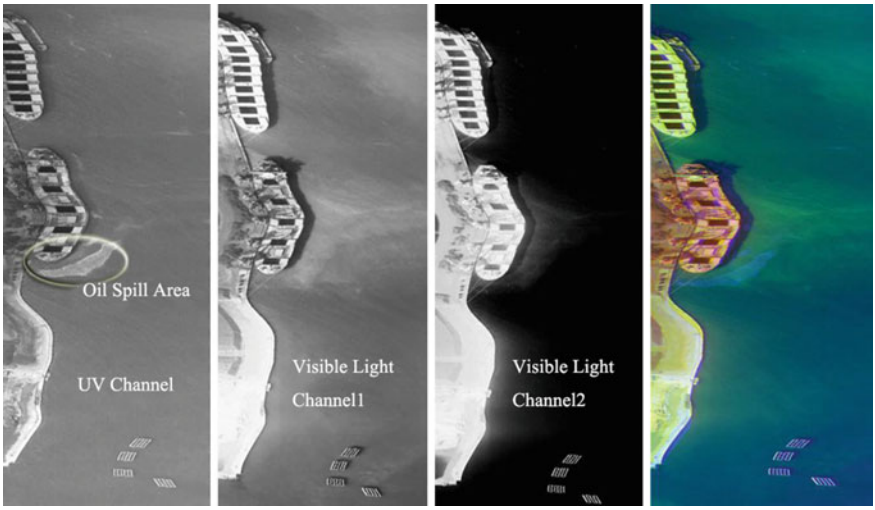


Fig. 6.114 Photograph of the oil target in sea area of Bohai obtained by the aerial photography experiment

bandwidth of 20 and 670 nm for channel 2 with bandwidth of 20 nm) is mounted in the imaging system besides of UV push broom scanner. The color photograph was obtained by image fusion technique of UV and visible light. The result shows than the oil slick can be detected more sensitively in visible-blind waveband than in visible light waveband.

6.7 Summary

The chip, fabricating process, and reliability of AlGaN detectors are introduced in this chapter. The principle and state of the art of AlGaN focal plane arrays, the basic structure of RCE UV detectors, the design and processing of the detectors, and irradiation effects of the detectors, are discussed systematically.

Compared with Si-based enhanced photoelectric diodes, solar-blind AlGaN UV detectors have a high rejection ratio to visible light for UV detection. The back-illuminated p-i-n diode is the main device structure for a UV detector, which has a high fill factor and can be easily connected with a Si-based readout circuit for linear or array fabrication. The photoelectric response characteristic of AlGaN UV detectors is analyzed systematically according to the design theory of photoelectric diodes and properties of UV film materials. An optimum device structure is obtained. In order to obtain high-quantum efficiency, suitable bandwidth, quick response time, and narrow-band response, a RCE UV detector can be used. The design and experiments for the basic structure of a RCE UV detector are introduced.

The overall design and experimental demonstration of RCE single-color and two-color UV detectors are completed.

Because there are strong chemical bonds between the atoms, AlGaN film materials have wide band gaps. Compared with traditional processing technology of the semiconductors (Si, Ge, and GaAs etc.), the processing of AlGaN UV devices faces new problems in Ohmic contacts, mosaic etching, and passivation. By studying the mosaic etching technique, it was found that the composite etching technique of ICP and wet etching can not only obtain good mosaic structure but also efficiently remove the physical damage formed in etching process. The main leakage current of the AlGaN UV devices at zero bias is the surface leakage induced by mosaic etching. The annealed SiO_2 passivation film deposited by PECVD is a better passivation technique to reduce surface leakage current. Ti/Al-based multilayer alloy film after annealing can form good Ohm contact with n-type AlGaN material. For the p-type AlGaN materials with large composition, it is still difficult to prepare good Ohmic electrodes at this time. Because p-type GaN material has a high work function and a low carrier concentration, the Ohmic contact can be improved by using a p-type GaN cap layer as a transition layer between the metal and p-type AlGaN.

The influence of ionizing radiation on detectors is very important for their applications in aerospace and military fields. By measuring the Raman spectroscopy of AlGaN materials irradiated by electrons and protons with different dosages, it was found that the Raman spectroscopy shows a displacement in peak frequency and the carrier concentration of the material will decrease as well. The experimental results for the photoluminescence spectrum show that the density of the spectrum will decrease and the frequency of main emission peak will move toward short wavelengths after the materials were irradiated by electrons and protons. The above phenomena may be caused by the defect energies induced by the irradiation. The experimental results also show that the effect of electron irradiation is a little weaker than that of proton irradiation. The variations of I–V curves and response spectrum of the detectors after the electron and proton irradiation with different dosages show that the leakage currents of GaN-based p–i–n device will increase. The responsivity of the detector does not change significantly after electron irradiation, while proton irradiation will make the responsivity of detector decline. The peak of the response spectrum will have a small displacement in the short-wave direction. This means that the influence of the proton irradiation on the detector is larger than that of electron irradiation. The influence of different passivation techniques on electron irradiation effects of AlGaN-based p–i–n UV detectors was also investigated. The results show that the leakage currents of the detectors will increase after the irradiation. The response spectrum of the detector without a passivation cap layer will change. The experimental results show that SiO_2 passivation layer can increase the ability of the detectors to withstand radiation effects. The properties of I–V curves, C–f curves, and response spectrum of the detectors after γ irradiation do not change obviously. But the Schottky barriers of metal-semiconductor contacts of the detectors will decrease after γ irradiation.

The applications of UV linear and array detectors with single-color response in visible-blind waveband and two-color responses in both visible-and solar-blind wavebands are introduced at the end of this chapter. The imaging results for near objects, long distance scenes, and the oil on the sea surface show that UV detection can play an important role in the applications of low altitude and high altitude imaging techniques.

References

1. Razeghi M, Rogalski A. Semiconductor ultraviolet detectors. J Appl Phys. 1996;79 (10):7433–73.
2. Lianghui C. III-Vsemiconductor multi-spectra FPA detector development. Infrared Laser Eng. 2008;37(1):1–8 (in Chinese).
3. Haimei G, Xianbgyang L, Yong K, et al. III-nitride UV detector and development. Laser Infrared. 2005;35(11):812–6 (in Chinese).
4. Walker D, Saxler A, Kung P, et al. Visible blind GaN p-i-n photodiodes. Appl Phys Lett. 1998;72(25):3303–5.
5. Monroy E, Calle F, Garrido JA, et al. Si-doped $Al_xGa_{1-x}N$ photoconductive detectors. Semicond Sci Technol. 1999;14(8):685–9.
6. Monroy E, Calle F, Pau JL, et al. Analysis and modeling of $Al_xGa_{1-x}N$ -based Schottky barrier photodiodes. J Appl Phys. 2000;88(4):2081–91.
7. Yang W, Nohova T, Krishnankutty S, et al. Back-illuminated GaN/AlGaN heterojunction photodiodes with high quantum efficiency and low noise. Appl Phys Lett. 1998;73(8):1086–8.
8. McClintock R, Yasan A, Mayes K, et al. High quantum efficiency AlGaN solar-blind p-i-n photodiodes. Appl Phys Lett. 2004;84(8):1248–50.
9. Selim M, Strite S. Resonant cavity enhanced photonic devices. J Appl Phys. 1995;78(2):607–39.
10. Charles LJ. Advances in astronomical UV image sensors and associated technologies. Proc SPIE. 1997;2999:244–258.
11. Adivarahan V. Design, fabrication and characterization of AlGaN solar-blind photodetectors. UMI Microfilm. 2001;3020921:4–5.
12. Koide Y, Itoh H, Khan MRH, et al. Energy band-gap parameter in an $Al_xGa_{1-x}N$ alloy. J Appl Phys. 1987;61(9):4540–3.
13. Muth JF, Brown JD, Johnson MAL, et al. Absorption coefficient and refractive index of GaN, AlN and AlGaN alloys. MRS Internet J Nitride Semicond Res. 1999;4S1, G5.2.
14. Bunker KL, González JC, Russell PE. EBIC measurements of minority carrier diffusion length in a GaN-based lighting emitting diode. http://spm.aif.ncsu.edu/aif.
15. Palik ED. Handbook of optical constants of solids. Orlando: Orlando Academic Press; 1985.
16. Yu G, Wang G, Ishikawa H, et al. Optical properties of wurtzite structure GaN on sapphire around fundamental absorption edge (0.78–4.77 eV) by spectroscopic ellipsometry and the optical transmission method. Appl Phys Lett. 1997;70(24):3209–11.
17. Levinshtein ME, Rumyantsev SL, Shur MS. Properties of advaced semiconductor materials. New York: Wiley; 2001.
18. Monemar B. III-V nitrides-important future electronic materials. J Mater Sci Mater Electron. 1999;10(4):227–54.
19. Yan WS, Zhang R, Xiu XQ, et al. Phenomenological model for the spontaneous polarization of GaN. Appl Phys Lett. 2007;90(18):182113(1–3).
20. Zhou J, Jiang R, Ji X et al. Interface polarization effect on properties of AlGaN/GaN heteo-junction PIN photodetector. J Semicond. 2007;28(6):947–950 (in Chinese).

21. Kishino K, Yonemaru M, Kikuchi A. AlGaN-based resonant cavity enhanced UV photodetectors. IEEE. 2002;1–2:835–6.
22. Li T, Carrano JC, Eiting CJ, et al. Design of a resonant-cavity-enhanced p-i-n GaN/$Al_xGa_{1-x}N$ photodetector. Fiber Integr Opt. 2001;20(2):125–31.
23. Biyikli N, Kartaloglu T, Aytur O, et al. High speed visible-blind resonant cavity enhanced AlGaN Schottky photodiodes. Mater Res Soc. 2003;8(8):1–7.
24. Xiaoli Ji, Ruolian J, Liang L, et al. Design and characteristics of AlGaN/GaN distributed bragg refect mirrors. Laser Infrared. 2005;35(11):888–90 (in Chinese).
25. Ambacher O, Arzberger M, Brunner D, et al. AlGaN based Bragg reflectors. MRS Internet J Nitride Semicond Res. 1997;2(22).
26. Liang L, Rong Z, Zili X, et al. AlGaN/GaN DBR grown by MOCVD. Laser Infrared. 2005;35(11):891–3 (in Chinese).
27. Liu B, Zhang R, Xie Z, et al. MOCVD method for AlGaN based multi-bands Bragg refelect mirrors. J Semicond. 2007;28:33–38 (in Chinese).
28. Liu B, Zhang R, Xie ZL, et al. The template effects on AlN/$Al_{0.3}Ga_{0.7}N$ distributed Bragg reflectors grown by MOCVD. J Cryst Growth. 2007;298:357–60.
29. Ji XL, Jiang RL, Zhou JJ, et al. Investigation into the energy band diagram and charge distribution in AlGaN/GaN double heterostructures by self-consistent Poisson-Schroginger calculation. Chin Phys Lett. 2005;22(2):454–6.
30. Ji XL, Jiang RL, Xie ZL, et al. High-reflectivity AlGaN/AlN distributed Bragg reflector in ultraviolet region. Chin Phys Lett. 2007;24(6):1735–7.
31. Ji XL, Jiang RL, Liu B, et al. Structural characterization of AlGaN/AlN Bragg reflector grown by metalorganic chemical vapor deposition. Phys Stat Sol (a). 2008;205(7):1572–4.
32. Zheng H. GaNBased RCE UV detector. Shanghai: Chinese Academy of Sciences; 2006.
33. Zheng H. GaN based RCE UV detector. Infrared. 2005;8:37–42 (in Chinese).
34. He Z, Kang Y, Tang YW, et al. Study on the spectral response of the Schottky photodetector of GaN. Chin Phys. 2006;15(6):1325–9.
35. Xiaoli J, Ruolian J, Jianjun Z, et al. AlGaN Base resonance enhanced pin UV detector. J Semicond. 2007;28(12):1957–60 (in Chinese).
36. Nakamura S. GaN growth using GaN buffer layer. Jpn J Appl Phys. 1991;30:L1705–7.
37. Amano H, Akasaki I, Kozawa T, et al. Electron beam effects on blue luminescence of Zinc-doped GaN. J Lumin. 1988;40–41:121–2.
38. Nakamura S, Mukai T, Senoh M, et al. Thermal annealing effects on P-Type Mg-doped GaN films. Jpn J Appl Phys. 1992;31:L139–42.
39. Pearton SJ, Zolper JC, Shul RJ, et al. GaN: processing, defects and devices. J Appl Phys. 1999;86(1):1–78.
40. Shul RJ, McClellan GB, Pearton SJ, et al. Comparison of dry etch techniques for GaN. Electron Lett. 1996;32(15):1408–9.
41. Shul RJ, McClellan GB, Casalnuovo SA, et al. Inductively coupled plasma etching of GaN. Appl Phys Lett. 1996;29(8):1119–21.
42. khan FA, Zhou L, Ping AT, et al. Inductively coupled plasma reactive ion etching of AlxGa1-xN for application laser facet formation. J Vac Sci Technol B. 1999;17(6):2750–2754.
43. Liang C. GaN/AlGaN UV detector fabrication and optoelectric properties. Shanghai: Chinese Academy of Sciences; 2008.
44. Yong K. Research on GaN based p-i-n Uvdector array and optoelectric properties. Shanghai: Chinese Academy of Sciences; 2006.
45. Liang C, Yong K, Longyuan Z, et al. Cl2/Ar/BCl3 ICP etching of AlGaN. Micro Nano Electron Technol. 2007;44(7):7–8 (in Chinese).
46. Singh R. High-density plasma etching of III-Nitrides: process development, device application and damage remediation, Boston University, 1997.
47. Pearton SJ, Shul RJ, Ren F. A review of dry etching of GaN and related materials. MRS Internet J Nitride Semicond Res. 2000;5:11.
48. Nakamura S, Fasol G. The blue laser diode. New York: Springer; 1997.

49. Pearton SJ, Lee JW, MacKenzie JD, et al. Dry etch damage in InN, InGaN, and InAlN. Appl Phys Lett. 1995;67(16):2329–31.
50. Seaward KL, Moll NJ. Semiconductor damage from inert and molecular gas plasmas. J Vac Sci Technol B. 1992;10(1):46–52.
51. Shul RJ, Zhang L, Baca AG, et al. Inductively coupled plasma-induced etch damage of GaN p-n junctions. J Vac Sci Technol A. 2000;18(4):1139–43.
52. Tripathy S, Ramam A, Chua SJ, et al. Characterization of inductively coupled plasma etched surface of GaN using Cl_2/BCl_3 chemistry. J Vac Sci Technol A. 2001;19(5):2522–32.
53. Crowell CR, Rideout VL. Normalized thermionic-field (T-F) emission in metal-semiconductor (Schottky) barriers. Solid-St.Electron. 1969;12(2):89–105.
54. Kim HS, Yeom GY, Lee JW, et al. A study of GaN etch mechanisms using inductively coupled Cl_2/Ar plasmas. Thin Solid Films. 1999;341(1–2):180–3.
55. Pearton SJ, Abernathy CR, Ren F, et al. Ar+-ion milling characteristics of III-V nitrides. J Appl Phys. 1994;76(2):1210–5.
56. Guo QX, Kato O, Yoshida A, et al. Chemical etching of indium nitride. J Electrocham Soc. 1992;139:2008–9.
57. Jie C, Jintong X, Ling W, et al. Low-damage wet chemical etching for GaN-based visible-blind p-i-n detector. Proc SPIE. 2008;6621:66211D.1–66211D.
58. Chen J, Xu J, Wang L, et al. Wet chemical etching for GaN-based material. Laser Infrared. 2007;37(9):961–3 (in Chinese).
59. Jie C. Research on wet chemical etching for GaN-based detector. Shanghai: Chinese Academy of Sciences; 2008.
60. Jogai B. Influence of surface state on the two-dimensional electron gas in AlGaN/GaN heterojunction field-effect transistors. J Appl Phys. 2003;93(3):1631–5.
61. Collins CJ, Chowdhury U, Wong MM, et al. Improved solar-blind detectivity using an AlGaN heterosturcture p-i-n photodiode. Appl Phys Lett. 2002;80(20):3754–6.
62. Kang BS, Kim S, Kim J, et al. Effects of external strain on the conductivity of AlGaN/GaN high-electron-mobility transistors. Appl Phys Lett. 2003;83(23):4845–7.
63. You D, Tang YW, Xu JT, et al. Effects of tensile stress induced by SiO_2 passivation layer on the properties of AlGaN/GaN hetero-structure photodiode. Appl Phys Lett. 2006;89:062107 (1)-(3).
64. Da Y. Research on GaN/AlGaN heterojunction UV detectors. Shanghai: Chinese Academy of Sciences; 2006.
65. Xue L. GaN based semiconductor MS contact and p-i-n UV detectors. Shanghai: Chinese Academy of Sciences; 2005.
66. Mohney SE, Wang Y, Cabassi MA, et al. Measuring the specific contact resistance of contacts to semiconductor nanowires. Solid State Electron. 2005;49(2):227–32.
67. Trexler JT, Pearton SJ, Holloway PH, et al. Comparison of Ni/Au, Pd/Au, Cr/Au metallizations for ohmic contacts to p-GaN[A]. Mater Res Soc Symp Proc. Boston, MA, USA, 1997:1091–1096.
68. Lin ME, Huang FY, MorkoG H. Nonalloyed ohmic contacts on GaN using InN/GaN short-period superlattices. Appl Phys Lett. 1994;64(19):2557.
69. Fischer S, Wetzel C, Haller FE, Meyer BK. On p-type doping in GaN-acceptor binding energies. Appl Phys Lett. 1995;67(9):1298–300.
70. Kim JK, Lee JL, Lee JW, et al. Effect of surface treatment by $(NH4)_2S_x$ solution on the reduction of ohmic contact resistivity of p-type GaN. J Vac Sci Technol B, 1999;17(2):497–499.
71. Sun J, Rickert KA, Redwing JM, et al. p-GaN surface treatments for metal contacts. Appl Phys Lett. 2000;76(4):415.
72. Lee JL, Webber M, Kim JK, et al. Ohmic contact formation mechanism of nonalloyed Pd contacts to p-type GaN observed by positron annihilation spectroscopy. Appl Phys Lett. 1999;74(16):2289.
73. Pal S, Sugino T. Fabrication and characterization of metal/GaN contacts. Appl Surf Sci. 2000;161(1–2):263–7.

74. Ho JK, Jong CS, Chiu CC, et al. Low-resistance ohmic contacts to p-type GaN achieved by the oxidation of Ni/Au films. J Appl Phys. 1999;86(8):4491.
75. Jang HW, Kim KH, Kim JK. Low-resistance and thermally stable ohmic contact on p-type GaN using Pd/Ni metallization. Appl Phys Lett. 2001;79(12):1822–4.
76. Chen LC, Ho JK, Jong CS, et al. Oxidized Ni/Pt and Ni/Au ohmic contacts to p-type GaN. Appl Phys Lett. 2000;76(25):3703.
77. Qiao D, Yu LS, Lau SS, et al. A study of the Au/Ni ohmic contact on p-GaN. J Appl Phys. 2000;88(7):4196.
78. Li X, Chen J, Xu JT, et al. Study on metal/p-GaN contacts on p-i-n GaN based UV detector. San Jose, California: SPIE Photonics West; 2007. p. 20–5.
79. Tang YW, Li X, Kang Y, et al. Influence of a new surface treatment method on ohmic contact resistivity of p-type GaN. Proc SPIE. 2004;5633:401–8.
80. Tang YW, You D, Xu JT, et al. Reduction of ohmic contact resistivity on p-GaN using N2 plasma surface treatment at room temperature. Semicond Sci Technol. 2006;21(12):1597–9.
81. Boer FR, Boom R, Mattens WCM, et al. Cohension in metals. North-Holland Amsterdam, 1989.
82. Neugebauer J, Van de Walle CG. Role of hydrogen in doping of GaN. Appl Phys Lett. 1996;68(13):1829–31.
83. Myers SM, Wright AF, Petersen GA, et al. Diffusion, release, and uptake of hydrogen in magnesium-doped gallium nitride: theory and experiment. J Appl Phys. 2001;89(6):3195.
84. Jang HW, Urbanek W, Yoo MC, et al. Low-resistant and high-transparent Ru/Ni ohmic contact on p-type GaN. Appl Phys Lett. 2002;80(16):2937.
85. Rideout VL, Crowell CR. Effects of image force and tunneling on current transport in metal-semiconductor (schottky-barrier) contacts. Solid-State Electron. 1970;13(7):993.
86. Kuznetsov NI, Dmitriev VA Kalinina EV, et al. Schottky barrier on n-GaN grown on SiC. J Electron Mater. 1996;25(5):831–834.
87. Moulder JF, Stickle WF, Sobol PE, et al. Hand-book of x-ray Photoeletron spectroscopy. MN: Perlin-Elmer Eden Prairie; 1991.
88. Koide Y, Ishikawa H, Kobayashi S, et al. Dependence of electrical properties on work functions of metals contacting to p-type GaN. Appl Surf Sci. 1997;117:373–9.
89. Lee JL, Kim JK, Lee JW, et al. Effect of surface treatment by KOH solution on ohmic contact formation of p-type GaN. Solid-state Electron. 1999;43(2):435–8.
90. Sun J, Seo DJ, O'Brien WL, et al. Chemical bonding and electronic properties of SeS_2-treated GaAs(100). J Appl Phys. 1999;85(2):969.
91. Look DC, Reynolds DC, Hemsky JW, et al. Defect donor and acceptor in GaN. Phys Rev B. 1997;79(12):2273.
92. Jain SC, Willander M, Narayan J, et al. III–nitrides: Growth, characterization, and properties. J Appl Phys. 2000;87(3):965–1006.
93. Yan Z, Haimei G, Yun B, et al. Space-borne UV detector and development of AlGaN detector. Laser Infrared. 2006;36(11):1009–12 (in Chinese).
94. Ruvimov S, Weber ZL, Washburn J. Microstructure of Ti/Al ohmic contacts for n-AlGaN. Appl Phys Lett. 1998;73(18):2582.
95. Chen J, Li X, Li X Y. Effect of rapid thermal annealing on Ti/Al/Ni/Au ohmic contact to n-$Al_{0.45}Ga_{0.55}N$. In: IEEE the 6th international workshop on junction technology; 2006. p. 273–275.
96. Chen Jun . MS contact of GaNbased semiconductor and p-i-n UVdetectors. Shanghai: Chinese Academu of Sciences; 2007.
97. Luther BP, Mohney SE, Jackson TN, et al. Investigation of the mechanism for Ohmic contact formation in Al and Ti/Al contacts to n-type GaN. Appl Phys Lett. 1997;70(1):57.
98. Kaminska E, Piotrowska A, Gaziewicz M, et al. Ohmic contact to n-GaN with TiN diffusion barrier. Mater Res Soc Symp Proc. 1997;449:1055–60.
99. Liang C, Jintong X, Yun B, et al. High Al content AlGaN and photoconductor devices. Laser Infrared. 2006;36(9):868–70 (in Chinese).

100. Kim HK, Seong TY, Adesida I, et al. Low-resitance Pt/Pd/Au ohmic contacts to p-type AlGaN. Appl Phys Lett. 2004;84(10):8.
101. Rhoderick EH. Metal-semiconductor contact. Beijing: Science Press; 1984.
102. Fan ZF, Mohammad SN, Kim W, et al. Very low resistance multilayer ohmic contact to n-GaN. Appl Phys Lett. 1996;68(12):1672–4.
103. Sheu JK, Su YK, Chi GC, et al. Investigation of the mechanism for Ti/Al ohmic contact on etched n-GaN surfaces. J Vac Sci Technol B. 2000;18(2):729–732.
104. Jang HW, Kim JK, Jeon CM, et al. Room Temperature Ohmic contact on n-type GaN using plasma treatment. MRS Internet J Nitride Semicond Res. 2001;6(8):1–7.
105. Rong B, Cheung R, Gao W, et al. Effects of reactive ion etching on the electrical characteristics of GaN. J Vac Sci Technol B. 2000;18(6):3467–70.
106. Lee JM, Chang KM, Kim SW, et al. Dry etch damage in n-type GaN and its recovery by treatment with an N2 plasma. J Appl Phys. 2000;87(11):7667–70.
107. Kim JK, Lee JL, Lee JW, et al. Low resistance Pd/Au ohmic contacts to p-type GaN using surface treatment. Appl Phys Lett. 1998;73(20):2953.
108. Smith LL, King SW, Nemanich RJ, et al. Cleaning of GaN surfaces. J Electron Mater. 1996;25(5):805–10.
109. King SW, Barnak JP, Bremster MD, et al. Cleaning of AlN and GaN surfaces. J Appl Phys. 1998;84(9):5248.
110. Lee KN, Donovan SM, Gila B, et al. Surface chemical treatment for the cleaning of AlN and GaN surfaces. J Electro Chem Soc. 2000;147(8):3087–90.
111. Chu TL. Gallium nitride films. J Electro Chem Soc. 1971;118(7):1200.
112. Pankove JI. Electrolytic etching of GaN. J Electrochem Soc. 1972;119(8):1118.
113. Cho H, Hays DC, Vartuli CB, Pearton SJ, Abernathy CR, MacKenzie JD, Ren F, Zolper JC. Wet chemical etching survey of III-nitrides. Power Semicond Mater Devices. 1998;483:265–70.
114. Vartuli CB, Pearton SJ, Abernathy CR, et al. Wet chemical etching survey of III-nitrides. Solid-State Electron. 1998;41(12):1947–50.
115. Carosella CA, Molnar B, Schiestel S, Sprague JA. Wet etching of ion-implanted GaN crystals by AZ-400 K photoresist. MRS Internet J Nitride Semicond Res. 2000;5(S1).
116. Yingwen Tang. Studies on surface and interface of GaN UV detectors and HgCdTe infrared detectors. Shanghai: Chinese Academy of Sciences; 2006.
117. Bardwell JA, Foulds I, Lamontagne B, et al. Fabrication of high performance GaN modulation doped field effect transistors. J Vac Sci Technol A. 2000;18(2):750–3.
118. Bae C, Rayner GB, Lucovsky G, et al. Device-quality GaN–dielectric interfaces by 300 8C remote plasma processing. Appl Surf Sci. 2003;216(1–4):119–23.
119. Huh C, Kim SW, Kim HM, et al. Effect of alcohol-based sulfur treatment on Pt Ohmic contacts to p-type GaN. Appl Phys Lett. 2001;78(13):1942.
120. Lin YJ, Chu YL, Huang YS, et al. Optical and electrical properties of heavily Mg-doped GaN upon (NH4)2Sx treatment. Appl Phys Lett. 2005;86(20):202107.
121. Dialea M, Aureta FD, van der Berg NG, et al. Analysis of GaN cleaning procedures. Appl Surf Sci. 2005;246(1–3):279–89.
122. Miller EJ, Schaadt DM, Yu ET, et al. Origin and microscopic mechanism for suppression of leakage currents in Schottky contacts to GaN grown by molecular-beam epitaxy. J Appl Phys. 2003;94(12):7611–5.
123. Jang HW, Jeon CM, Kim JK, et al. Room-temperature Ohmic contact on n-type GaN with surface treatment using Cl_2 inductively coupled plasma. Appl Phys Lett. 2001;78(14):2015.
124. Kim SW, Lee JM, Huh C. Reactivation of Mg acceptor in Mg-doped GaN by nitrogen plasma treatment. Appl Phys Lett. 2000;76(21):3079–81.
125. Ishikawa H, Kobayashi S, Koide Y, et al. Effects of surface treatments and metal work functions on electrical properties at p-GaN/metal interfaces. J Appl Phys. 1997;81(3):1315–22.
126. Sze SM, Coleman DJ, et al. Current transport in metal-semiconductor-metal structures. Solid-state Electron. 1971;14(12):1209–1218.
127. Hattori K, Izumi Y. The electrical characteristics of degenerate InP Schottky diodes with an interfacial layer. J Appl Phys. 1982;53(10):6906.

128. Claeys C, Simoen E. Radiation effects in advanced semiconductor materials and devices. Berlin: Springer-Verlag; 2008.
129. Fang ZQ, Hemsky JW, Look DC. Mack M P. Electron-irradiation-induced deep level in n-type GaN. Appl Phys Lett. 1998;72(4):448.
130. Polenta L, Fang ZQ, Look DC. On the main irradiation defect in GaN. Appl Phys Lett. 2000;76(15):2086.
131. Goodman SA, Auret FD, Koschnick FK, et al. Radiation induced defects in MOVPE grown n-GaN. Mater Sci Eng. 2000;71(1–3):100–3.
132. Legodi MJ, Hullavarad SS, Goodman SA, et al. Defect characterization by DLTS of AlGaN UV Schottky photodetectors. Phys B. 2001;308–310(12):1189.
133. Nedelcescu AI, Carlone C, Houdayer A, et al. Radiation hardness of gallium nitride. IEEE Trans Nucl Sci. 2002;49(6):2733.
134. Umana-Membreno GA, Dell JM, Hessler TP, et al. 60Co gamma-irradiation-induced defects in n-GaN. Appl Phys Lett. 2002;80(23):4354.
135. Hayes M, Auret FD, Wu L, et al. Electrical defects introduced during high-temperature irradiation of GaN and AlGaN. Physica B-Condens Matter. 2003;340:421–5.
136. Polyakov AY, Smirnov NB, Govorkov AV, et al. Neutron irradiation effects in undoped n-AlGaN. J Vac Sci Technol B. 2006;24(3):1094.
137. Polyakov AY, Smirnov NB, Govorkov AV, et al. Neutron irradiation effects on electrical properties and deep-level spectra in undoped n-AlGaN/GaN heterostructures. J Appl Phys. 2005;98(3):033529.
138. Polyakov AY, Smirnov NB, Govorkov AV, et al. Electrical and optical properties of hydrogen plasma treated n-AlGaN films grown by hydride vapor phase epitaxy. J Vac Sci Technol B. 2004;22(1):77.
139. Polyakov AY, Smirnov NB, Govorkov AV, et al. Proton implantation effects on electrical and optical properties of undoped AlGaN with high Al mole fraction. J Vac Sci Technol B. 2003;21(6):2500.
140. Hearne SM, Jaimeson DN, Yang C, et al. Electrical characteristics of proton irradiated AlGaN devices. Nucl Instrum Methods Phys Res B. 2002;190(1–4):873–7.
141. Hogsed MR, Yeo YK, Ahoujja M, et al. Radiation-induced electron traps in $Al_{0.14}Ga_{0.86}N$ by 1 MeV electron radiation. Appl Phys Lett. 2005;86(26):261906–1.
142. Li C, Subramanian S. Neutron irradation effects in GaN-based blue LEDs. Trans Nucl Sci. 2003;50(6):1998.
143. Hu XW, Karmarkar AP, Jun B, et al. Proton-irradiation effects on AlGaN/AlN/GaN high electron mobility transistors. Trans Nucl Sci. 2003;50(6):1791.
144. Yun B. Research on irradiation effect on GaN based UV detector. Shanghai: Chinese Academy of Sciences; 2007.
145. Yun B, Xiumei S, Yan Z, et al. electron irradiation effect of AlGaN/GaN P-I-N UV detector. Laser Infrared. 2007;37(9):867–9 (in Chinese).
146. Yun Bai, Xiumei Shao, Liang Chen, et al. proton irradiation effect on GaN based PIN UV detector. Laser Infrared. 2007;37(1):957–60 (in Chinese).
147. Yun B, Xiumei S, Liang C, et al. electron irradiation effect on GaN based Uvdetectors. Infrared Laser Eng. 2008;37(2):270–3 (in Chinese).
148. Yun B, Hui Q, Xiangyang L, et al. Low temperature annealling effect on γ-irradiated HgCdTe mid-wave photonconductor detectors. High Power Laser Beams. 2007;18(2):301 (in Chinese).
149. Kozawa T, Kachi T, Kano H, et al. Raman scattering from LO phonon-plasmon coupled modes in gallium nitride. J Appl Phys. 1994;75(2):1098.
150. Wang RX, Xu SJ, Fung S, et al. Micro-Raman and photoluminescence studies of neutron-irradiated gallium nitride epilayers. Appl Phys Lett. 2005;87(3):031906–1.
151. Buyanova IA, Wagner M, Chen WM, et al. Photoluminescence of GaN: effect of electron irradiation. Appl Phys Lett. 1998;73(20):2968.
152. Look DC, Farlow GC, Drevinsky PJ, et al. On the nitrogen vacancy in GaN. Appl Phys Lett. 2003;83(17):3525.

153. Khanna MS, Webb J, Tang H, et al. 2 MeV proton radiation damage studies of gallium nitride films through low temperature photoluminescence spectroscopy measurements. IEEE Trans Nucl Sci. 2000;47(6):2322.
154. Yongxiao D. Failure analysis of semiconductor. Beijing: Areo space press (in Chinese).
155. Yang JK, Hu C. MOS capacitance measurements for high-leakage thin dielectrics. IEEE. 1999;46(7):1500.
156. Foresi JS, Moustakas TD. Metal contacts to gallium nitride. Appl Phys Lett. 1993;62:2859.
157. Zaininger KH. Irradiation of MIS capacitors with high energy electrons. IEEE Trans Nucl Sci. 1966;13(6):237.
158. Jianzhong C. Irradiation effect of semiconductor materials. Beijing: Science Press; 1993 (in Chinese).
159. Hogsed MR, Yeo YK, Ahoujja M, Ryu MY, Petrosky JC, Hengehold RL. Radiation-induced electron traps in $Al_{0.14}Ga_{0.86}N$ by 1 MeV electron radiation. Appl Phys Lett. 2005;86:261906–1.
160. Ionascut A, Carlone C, Houdayer A. Radiation hardness of gallium nitride. IEEE Nucl Sci. 2002;49(6):2733.
161. Huang Y, Li XY, Gong HM. A demo-imaging system based on GaN UV detectors. Opt Syst Des SPIE. 2005;5964:286–93.
162. Yimin H. Research on dual band UV imaging system. Shanghai: Chinese Academy of sciences; 2007.

Chapter 7
Readout Integrated Circuit, Measurement, and Testing Technology for Advanced Focal Plane Array

7.1 Introduction

An infrared focal plane array (IRFPA) is an advanced imaging sensor with the ability to detect, acquire, and process infrared information. Generally, an IRFPA is composed of a photodiode array and specific readout integrated circuits. Being irradiated by infrared spectrum, each photodiode will generate the photocurrent corresponding to the energy of the radiation, and the photocurrent is processed and amplified to output by a readout integrated circuit (ROIC). The first-generation and second-generation infrared focal plane arrays have played an important role in the field of space, military, and civilian industries. In recent years, the third generation of IRFPA has been successfully developed with the characterization of high resolution, multicolor, and so on.

The ROIC is generally fabricated using standard CMOS integrated circuit technology. Primary functions of the ROIC are included as follows: to provide an adjustable bias voltage for the detector, to convert the photocurrent signal generated by the detector into the voltage signal, and to amplify and process signals along with time multiplexing of data from many detectors to just a minimum number of outputs. Basic elements of the ROIC include the input stage circuit signal preamplifiers, data multiplexers, and output video amplifiers. Since the basic architecture of the ROIC has been established with the development of infrared focal plane technology, more attention has been paid to enhance the signal process capability of IRFPA, and system on chip (SOC) has become an important development trend for IRFPA. First, an IRFPA is required to have various working modes to enhance the flexibility of application for the infrared imaging system. Second, the analog-to-digital conversion function is integrated on a readout circuit chip to eliminate signal noise and cross talk coupling problems caused by a long analog signal chain. By now, with the proposed concept of Focal Plane Processor (FPP), nonuniformity correction and other digital signal processing functions have been incorporated into IRFPA not only to meet the higher ability for real-time signal processing, but also to promote the

L. He et al., *Technology for Advanced Focal Plane Arrays of HgCdTe and AlGaN*,
DOI 10.1007/978-3-662-52718-4_7

miniaturization for infrared imaging systems. The implementation of a digital focal plane device with flexible operating modes and signal processing functions has been the trend of development for IRFPA, which demands the ROIC to be a SOC chip including functions with preamplification, digital signal processing, and output. With the development of Very-Large-Scale Integrated Circuits (VLSI) technology, the complicated SOC chip for infrared application has become technically possible by the shrinking of feature size and the increasing of integration level. Recently, many papers for infrared FPP are reported from Raytheon, LETI and Israel SCD companies. Raytheon has developed an Application-Specific Integrated Circuit (ASIC) with a 12-bit resolution analog-to-digital converter (ADC), and implemented the integration with 2048 × 2048 infrared focal plane device by the use of three-dimensional multilayer packaging technology [1]. In 2009, the Israel SCD Company developed a 640 × 512 FPP integrating a 12-bit resolution ADC and nonuniform correction function [2].

In this chapter, the development of ROIC is discussed. Dual-band ROIC design is then introduced, and the conventional and proposed topologies are presented. The implementation of a dual-band ROIC is illustrated. Second, the digital transmission SOC for IRFPA is explored in detail, including system architecture, topologies, and algorithms for ADC on focal plane, design, and implementation of ADC. Finally, since the focal plane testing technology is the basis of the evaluation for focal plane devices, testing techniques for ultraviolet and IRFPA are displayed separately.

7.2 Introduction and Development for Readout Integrated Circuit

Being an important part of IRFPA for processing photocurrent signal generated by the infrared detector array, the ROIC is generally provided with functions of photocurrent conversion, amplification, and buffered output. Applied in IRFPAs for detecting different wavelengths of infrared radiation, the ROIC may employ different topologies for the input stage circuit, including Self-Integrator (SI), Direct Injection (DI), Buffered Direct Injection (BDI), Source Follower Detector (SFD), Capacitance Trans-impedance Amplifier (CTIA), Gate Modulation Injection (GMI), and so on. Each topology may have respective advantages and disadvantages. Figure 7.1 is the schematic diagram of an ROIC with a 320 × 256 pixel array, which mainly consisted of the pixel array, column amplifier circuit, output amplified circuit, row address decoder, column address decoder, and timing control circuit. The output signal of the ROIC is in analog form. Since the analog signal is susceptible to various noise disturbances such as cross talk, clock jitter, and electromagnetic interference along the signal processing chain, a digital transmission SOC for the IRFPA could eliminate such noise sources related to analog transmission. Throughout current research in IRFPA, the development of the ROIC is on the way to digitalization.

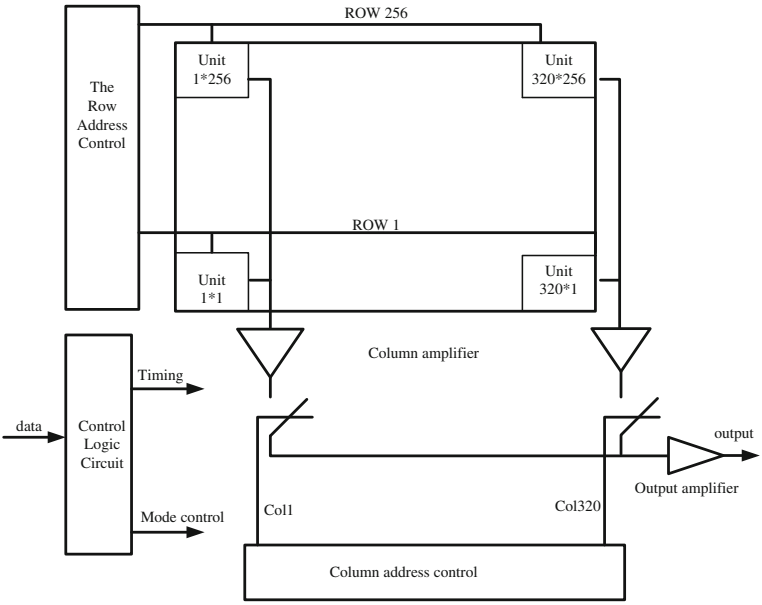


Fig. 7.1 The schematic diagram of an ROIC with a 320 × 256 pixel array

Through the development to now, the basic circuit architecture of the ROIC has been confirmed, and more attention is turning to enhance the potential capability for IRFPA and reduce the complexity of the system. Therefore, SOC is becoming an important development trend for ROIC.

The development trend of SOC for ROIC includes the following:

(1) Simplification of interface

With the improvement of microelectronics technology and circuit design, the ROIC is capable of providing bias voltages and timing control signals on chip, thereby greatly reducing the number of external pins for IRFPA. The external interface will become simpler. Digital and fiber-optic interfaces are becoming the development trend.

(2) Flexible IRFPA

In addition to the increasing array scale for IRFPA, the ROIC is provided with a windowing function, which allows a large array IRFPA (such as 640 × 480 or 1024 × 1024) to work as a low frame rate acquisition sensor with a wide visual field, high sensitivity, and moderate pixel rate, meanwhile allowing the IRFPA to read out a narrow visual field (such as 64 × 64, programmable) at a high frame rate when the pixel rate is still moderate.

(3) Standardization for universal application

In recent years, commercial companies for IRFPA have developed several standardized ROICs for the universal application. For example, the standardized 640 × 512 ROIC from the Indigo Company is designed for hybridization with various infrared detectors such as InSb, MCT, QWIP, InGaAs, etc., and is suitable for applications ranging from handheld infrared viewers to high-speed industrial infrared imaging systems [3]. The standardized ROIC is a sign of maturity for infrared focal plane technology and the result from the industrialization for ROIC.

(4) The enhancement of capability for signal processing

The new generation IRFPA demands higher resolution, smaller size, lighter weight, and lower power consumption. To solve these problems, new methods have to be introduced in the design of ROIC. Integrating more signal processing functionality and the ADC at system level into the focal plane has become the top priority of the development. By achieving analog-to-digital conversion on the focal plane, the low-noise and narrow-bandwidth analog signal does not have to be multiplexed into a wide-bandwidth analog channel, thus greatly increasing the response speed of the device at low temperature, eliminating the signal noise and cross talk coupling problem brought by the long analog signal chain, and directly providing the digital signal to post-stage circuits for processing.

With the integration of analog-to-digital conversion, the signal processing function of the ROIC will be greatly enhanced, including not only the preamplifier, the nonuniformity correction, analog-to-digital conversion, but also the memory, image intensification, image compression, feature extraction, and other image processing functionalities. Thereby the ROIC is enabling an intelligent focal plane assembly so as to greatly improve the system performance and portability of IRFPA.

With the development of microelectronic technology, the feature size of transistors is becoming smaller, and more complex circuits could be integrated into a limited area. The function of the ROIC will be much more powerful, and the development trend of SOC is becoming evident.

7.3 Dual-Band Readout Integrate Circuit

In recent years, in order to meet the demanding requirement for the IRFPA in terms of strategic and tactical application, not only is the size of pixel array increasing and the pitch of the pixels reducing for the single-band focal plane device, but also the integrated dual-band focal plane device has reached a high level of development. A dual-band IRFPA is capable of simultaneously acquiring the information from two bands of the infrared spectrum for the same input source so as to suppress a complex background. A dual-band IRFPA is provided with the function extending from just the detection of a target to the identification of the target, and eventually

with the function of quick discrimination of the target. Since dual-band IRFPAs play a significant role in military applications such as surveillance, early warning, and antimissile interception, it has been the research focus for IRFPA. Dual-band IRFPAs have a variety of combinations, in which the most common types are SWIR/MWIR, MWIR1/MWIR2, MWIR/LWIR, and LWIR1/LWIR2. The United States and France are in the leading group due to their earlier research and more advanced technology. Using the mode of Time-Division Multiplexed Integration (TDMI), the 1280 × 720 MWIR/LWIR IRFPA from Raytheon Vision Systems of the United States has been the largest dual-band IRFPA based on the material of HgCdTe up to now [4].

Compared to the single-band IRFPA, the dual-band IRFPA has the characteristic of high integration, high reliability, ultra-high speed and miniaturization, which greatly lowers the system's design, size, weight, and cost. It can simultaneously obtain the two bands of target information in the same radiation field, thus obtaining more meaningful target information and the benefit of eliminating false signals. An early dual-band infrared system is made up of two discrete single-band IRFPA detectors, which makes it difficult to ensure the spatial consistency and requires a complex optical system, independent cryogenic system, and complicated image processing technique. However, the integrated dual-band IRFPA not only ensures the spatial consistency of images for different infrared bands, but also greatly reduces the size of system, the complexity of system, and the requirement for refrigeration. In general, an infrared system composed of an integrated dual-band IRFPA can synchronously acquire information of target on two infrared bands in space and time, can make the suppression and identification for complicated background, can detect the target signal quickly and effectively, and can reduce the false alarm rate to improve the detection efficiency and accuracy in early warning, searching, and tracking system, thus significantly enhancing the performance of infrared imaging system and versatility in various weapon platforms.

The dual-band ROICs can convert the current signal of two bands received by the detector into the voltage signal through integration and output, thus realizing the independent information for signals of two bands. The capability of dual-band ROICs is directly relative to the performance of two-color IRFPA. In 1998, the American Lockheed Martin Company developed the simultaneous MW/LW dual-band MOVPE HgCdTe 64 × 64 IRFPA and the corresponding dual-band ROIC [5]. The detector pixel of IRFPA consists of two back-to-back MW/LW photodiodes. The input stage circuit of dual-band ROIC uses the DI architecture. In addition, the ROIC adopts Lockheed Martin's circuit algorithm of Compact Signal Averager (CSA) which has two integration capacitors, respectively, for MW and LW in each cell to achieve the separated integration of both bands, where the LW integration capacitor floats over the MW integration capacitor.

In 2002, a 128 × 128 MWIR1/MWIR2 dual-band IRFPA and ROIC were reported by the French LETI laboratory [6]. The detector structure is n^+ppn with a single indium bump and is operated in the sequence mode. The DI structure is applied for the input stage of the ROIC, where two integration capacitors are used for the photocurrent integration process of the two bands and the output signals can

be divided into two paths for synchronous readout. In 2005, Santa Barbara Research Center introduced the Time-Division Multiplexing (TDM) and Sub-Frame Average (SFA) technology [7].

The main function of the input stage circuit of ROIC is integrating the photocurrent generated by the detector and transforming the current signal to a voltage signal for amplification. The performance of the input stage circuit has the great influence on the whole performance of the ROIC and IRFPA. Since the circuit topology of the dual-band ROIC is dependent on the structure of detector, the research and design of the input stage circuit, which determines the performance of the dual-band IRFPA, is crucial. The two-color IR detectors and signal processing circuit structure are shown in Table 7.1. Table 7.2 displays the performance of multicolor IRFPAs.

7.3.1 Conventional Topologies of a Dual-Band ROIC

Different structures of the detector demand different input stage circuits to obtain the best performance. The circuit architecture shown in Fig. 7.2 is applied for a simultaneous dual-band IR detector, which is the vertical integration of two discrete photodiodes with same polar direction. Since this structure does not produce a mixed signal, the simultaneous operation mode can be easily achieved if each pixel contains two independent input stage circuits for a single spectral band. Simply speaking, this circuit is a mosaic of two single-band readout circuits. Many structures can be applied in the input stage circuit for each band of this type of detector, including SI, DI, BDI, GMI, and CTIA. The specific architecture can be adopted with the demanding of performance.

Figure 7.2 shows the DI structure of input stage circuits for two pixels. $D1\_1$ and $D1\_2$ are detectors for one band. $C1\_1$ and $C1\_2$ are integration capacitors of $D1\_1$ and $D1\_2$, respectively. Similarly, $D2\_1$ and $D2\_2$ are detectors for another band. $C2\_1$ and $C2\_2$ are integration capacitors of $D2\_1$ and $D2\_2$, respectively. $D1\_1$ and $D2\_1$ are in a pixel, and $D1\_2$ and $D2\_2$ are in another. Where, Φ_{start} is the initial pulse of circuit, V_{reset} is the reset voltage of the integration capacitor, and Φ_{reset} controls the reset of integration capacitor. V_{g1} is the gate bias voltage of input transistors $G1\_1$ and $G1\_2$ for band 1. Similarly, V_{g2} is the gate bias voltage of input transistors $G2\_1$ and $G2\_2$ for band2. V_{g1} and V_{g2} are, respectively, adjusted to ensure that transistors $G1\_1$, $G1\_2$, $G2\_1$, and $G2\_2$ can be in the threshold zone and simultaneously provide the low impendence interface to make the bias voltage constant for each detector. The N terminal of each detector (a PN junction) is respectively connected to the source terminal of integration control transistors $M1\_1$, $M1\_2$, $M2\_1$, and $M2\_2$ through the interconnection of the indium bump. The working principle of circuit is as follows: in the beginning, Φ_{start} is low and Φ_{reset} is low, therefore the switches $M1\_1$, $M1\_2$, $M2\_1$, $M2\_2$ are turned off and the switches $S1\_1$, $S1\_2$, $S2\_1$, $S2\_2$ are turned on to make the integration capacitors $C1\_1$, $C1\_2$, $C2\_1$, $C2\_2$ reset to the V_{reset}. Next, Φ_{start} is changed to high and Φ_{reset}

Table 7.1 The two-color IR detectors and signal processing circuit structures

Detector structure	Company	Scale	Polar	Band number	Operation mode	Indium bump number	Difficulty in signal readout
n-p-p-P-N heterojunction	Raytheon Vision Systems	256 × 256,640 × 480,1280 × 720	p-on-n/n-on-p	Dual-band MW/LW	Sequential	One	Relatively easy for the separation of two band signals
n-p-p-P-N heterojunction	French LETI LIR	320 × 256,256 × 256,128 × 128	p-on-n/n-on-p	Dual-band MW1/MW2	Sequential	One	Relatively easy for the separation of two band signals
n-p^+-n three-layer heterojunction	Santa Barbara Research Center and Hughes Laboratory	/	p-on-n/n-on-p	Dual-band MW/LW	Sequential	ONE	Relatively easy for the separation of two band signals
n-p^+-n three-layer heterojunction	Santa Barbara Research Center and Hughes Laboratory	/	p-on-n/n-on-p	Dual-band MW/LW	Simultaneous	Two	Relatively difficult for the separation of two band signals
pn-N-P homojunction	Lockheed Martin	64 × 64	p-on-n/n-on-p	Dual-band MW/LW	Simultaneous	Two	Relatively difficult for the separation of two band signals
n-n-n-N-P sumit	Rockwell Research Center	/	p-on-n/p-on-n	Dual-band MW1/MW2	Simultaneous	Two	Each photodiode pixel has two discrete single-band readout circuit; easy for signal separation
n^+/n/P on the Silicon base of ROIC	American DRS	/	p-on-n/n-on-p	Dual-band MW/LW	Simultaneous	Two	Each photodiode pixel has two discrete single-band readout circuit

Table 7.2 The performance of multicolor IRFPAs

Company	Band number	Material	Band combination	Array scale	Pixel dimension (μm^2)	Wavelength (μm)	NETD (mK)	Effective pixel (%)	Temperature (K)
RVS	Dual-band	HgCdTe	MW/LW	640 × 480	20 × 20	5.5	20.6	98.8	77
						10.5	21.8	96.6	
DRS	Dual-band	HgCdTe	MW/LW	320 × 240	50 × 50	5.2	9	97.1	77
						10.2	23	96.3	
LETI	Dual-band	HgCdTe	MW/MW	256 × 256	50 × 50	3.1	27.6	99.8	77
						5.2	9.6	99.8	
JPL	Dual-band	QWIP	LW/VLW	640 × 480	25 × 25	9.1	29	99.7	40
						15	74	98	
JPL	Four-band	QWIP	MW/LW1/LW2/VLW	640 × 512	25 × 25	10.2	20	>99	43

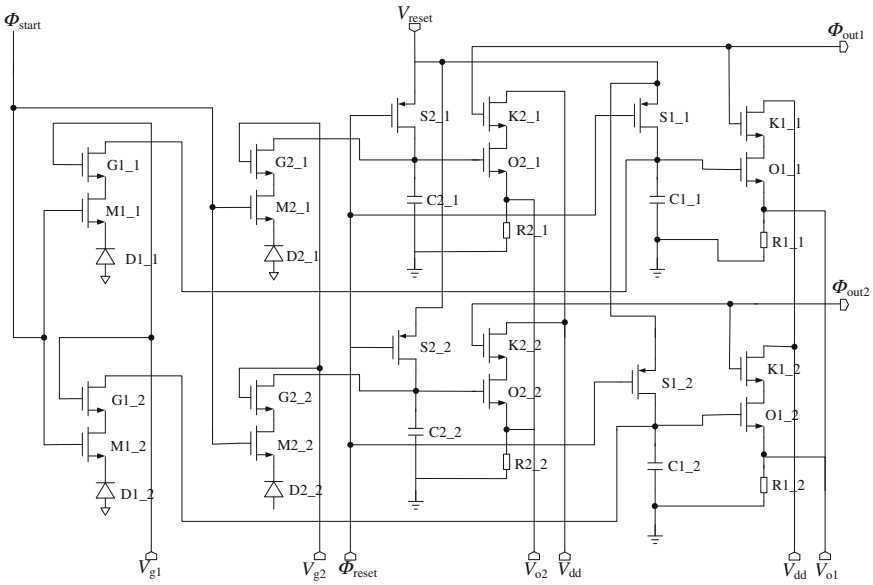


Fig. 7.2 The DI input stage circuit for a dual-band IR detector with simultaneous operation mode

is high, so the switches *S*1_1, *S*1_2, *S*2_1, *S*2_2 are turned off and the switches *M*1_1, *M*1_2, *M*2_1, *M*2_2 are turned on to allow the photocurrents generated by detectors *D*1_1, *D*1_2, *D*2_1, *D*2_2 to be integrated on the capacitors *C*1_1, *C*1_2, *C*2_1, *C*2_2. When the integration process is finished, all switches *M*1_1, *M*1_2, *M*2_1, *M*2_2 are turned off, and the current signal is transformed into the voltage signal which will be multiplexed to the output bus V_{o1} and V_{o2} through each source follower transistor *O*1_1, *O*1_2, *O*2_1, and *O*2_2 with the control of the pulse Φ_{out1} and Φ_{out2}. After the signals are readout, Φ_{reset} is set to low to begin the next period. Figure 7.3 shows the timing diagram of the ROIC, the rising edge of Φ_{reset} and the rising edge of Φ_{start} are at least at the same time or slightly overlapping. The first rising edge of Φ_{out} has some lag with the falling edge of Φ_{start}.

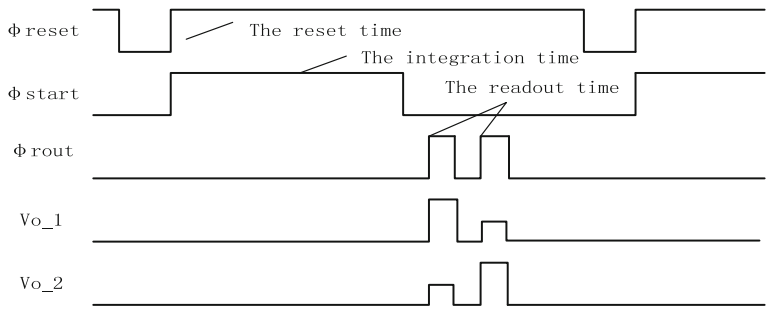


Fig. 7.3 The timing diagram of the simultaneous dual-band ROIC

In one working period, if the integration time is T, the integration current includes the photocurrent I_s and the dark current I_d, and η is injection efficiency, then the voltage of integration capacitor V_{int} can be obtained as Eq. 7.1 at the end of one working period.

$$V_{int} = V_{reset} - \frac{\eta(I_s + I_d)T}{C} \tag{7.1}$$

Equation 7.1 reveals the input stage circuit as an ideal integrator and the voltage of integration capacitor goes down almost as a constant slope with time.

The small signal mode of DI circuit is used to analyze which circuit parameters are concerned with the circuit's injection efficiency, as is shown in Fig. 7.4 [8]. Where I_s is the photocurrent, I_{nd} is the dark current, R_d is the internal resistance of the detector, C_d is the parasitic capacitor of the detector, g_m is the transconductance of the input transistor, and I_{nm} is the noise current of the input transistor. The injection efficiency η can be defined as

$$\eta = \frac{I_{int}}{I_{photocurrent}} \tag{7.2}$$

Under low frequency or dc input, the injection efficiency η can be expressed as

$$\eta = \frac{g_m R_d}{1 + g_m R_d} \tag{7.3}$$

Equation 7.3 shows that the injection efficiency depends on the resistance of the input transistor $1/g_m$ and the internal resistance of the detector R_d, so it can be improved by increasing g_m. The gate bias of the input transistor is lower than the threshold voltage since the input transistor is required to work in the weak inversion region. The transition voltage from weak inversion to strong inversion is

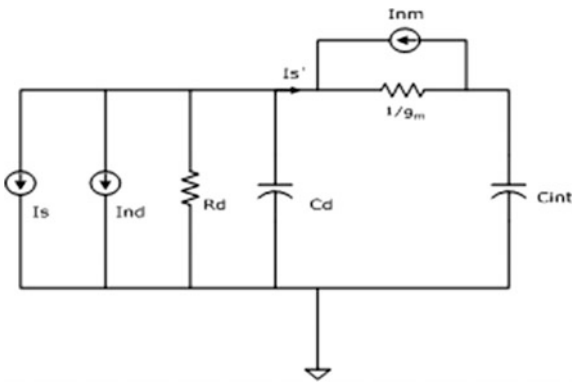


Fig. 7.4 The small signal mode of DI circuit

$$V_{GS} = V_T + \frac{2nkT}{q} \tag{7.4}$$

Here, n is a constant between 1 and 2, k is the Boltzmann constant, T is the absolute temperature, q is the electron charge, and $2kT/q$ is about 26 mV at room temperature. In weak inversion, the drain current versus the gate-source voltage is not the square law but the exponential relationship

$$I_{DS} = I_{D0} \exp\left[\frac{q}{nkT}(V_{GS} - V_T)\right]\left[1 - \exp\left(\frac{-q}{nkT}V_{DS}\right)\right] \tag{7.5}$$

Here, $I_{D0} = \frac{W}{L}\mu C_0\left(n\frac{kT}{q}\right)^2$, μ_n is the electron mobility, C_0 is the oxide capacitance. When $V_{DS} > 3\frac{kT}{q} = 3V_t$, the last term in (7.5) approaches unity and the equation can be simplified as

$$I_{DS} = I_{D0} \exp\left[\frac{(V_{GS} - V_T)}{nVT}\right] \tag{7.6}$$

So g_m can be obtained as

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}} = \frac{I_{DS}}{nVT} \tag{7.7}$$

Equation 7.7 shows that the transconductance g_m is independent of W/L of the input transistor and is proportional to the drain current.

The input stage circuit structure shown in Fig. 7.2 is simple. The bias voltage of the detector for each band can be set independently to help the detector be at the best working point, therefore the performance of device is relatively good.

Figure 7.5 shows the input stage circuit of ROIC which is developed for a sequential dual-band IRFPA and adopts the DI structure. Different from the circuit structure shown in Fig. 7.2, this structure has only one signal input terminal shared by detectors for two bands. The structure of detector consists of two photodiodes with vertical superposition, which respectively respond to the SWIR and MWIR bands. The switch transistor $M1$ is not only the integration time control transistor of photocurrent Isw for SWIR but also the bias control transistor of the detector. SWINT is the integration time control pulse for SWIR, whose signal amplitude should ensure that the $M1$ works in the sub-threshold region when $M1$ is turned on. $M1$ is the integration capacitor for the SWIR photocurrent signal. $M1$ is the reset switch of $M1$ and RST1 is the corresponding reset control pulse. Similarly, $M2$ is the integration time and bias control transistor of the photocurrent I_{mw} for MWIR. MWINT is the integration time control pulse providing the voltage for $M2$ to work in the sub-threshold region. $C2$, $S2$, and RST2 represent, respectively, the integration capacitor, the reset switch of integration capacitor, and the reset control pulse. V_{con} is the bias for the common terminal of the detector and is also a pulse

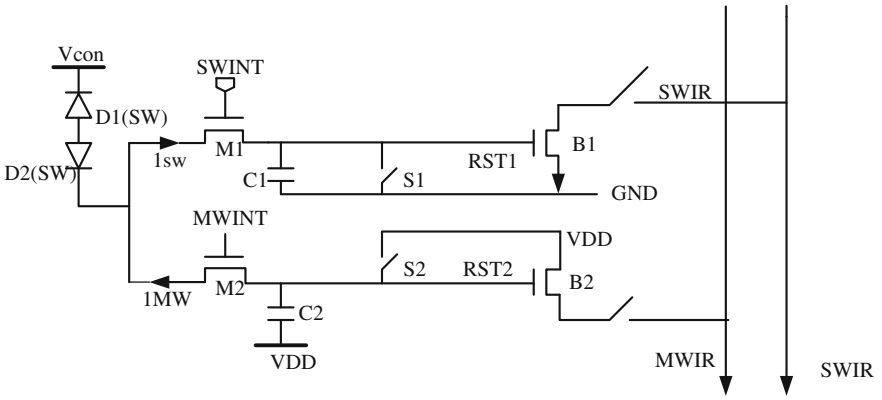


Fig. 7.5 The input stage circuit of ROIC for the sequential dual-band IRFPA

signal. The circuit working process is as follows: at the beginning, RST1 is high and *S*1 is turned on to make *C*1 reset to the zero potential, RST2 is low and *S*2 is open to make *C*2 reset to the VDD potential. Meanwhile, *M*1 and *M*2 are turned off with SWINT being high and MWINT being low. When the reset process is over, RST1 is changed to low and RST2 is changed to high. If I_{mw} is integrated at first, the detector *D*2 should be reverse biased and the detector *D*1 is forward biased by setting V_{con} low potential. To make *M*2 work in the sub-threshold region, there are

$$V_{\mathrm{GS}} = V_{\mathrm{MWINT}} - V_{\mathrm{CON}} = V_{\mathrm{MWINT}} - 0 \approx V_{\mathrm{TN}} \Rightarrow V_{\mathrm{MWINT}} \approx V_{\mathrm{TN}} \tag{7.8}$$

where V_{MWINT} is the amplitude of the pulse MWINT, V_{TN} is the threshold voltage (which is about 0.7 V at room temperature), so V_{MWINT} is close to 0.7 V according to the Eq. 7.8 and the accurate value can be determined by the simulation results. When the integration process of I_{MW} is finished, MWINT is changed to low, and the *M*2 is closed to start the integration of I_{SW} while V_{con} is high. Similarly, V_{SWINT} is the amplitude of the pulse SWINT to ensure that *M*1 work in the sub-threshold region, and there are

$$V_{\mathrm{CON}} - V_{\mathrm{SWINT}} = V_{\mathrm{DD}} - V_{\mathrm{SWINT}} \approx V_{\mathrm{TP}} \Rightarrow V_{\mathrm{SWINT}} \approx V_{\mathrm{DD}} - V_{\mathrm{TP}} \tag{7.9}$$

where V_{TP} is the threshold voltage for PMOS transistor. When the integration process of I_{SW} is finished, SWINT is changed to high and the *M*1 is closed. The integration voltage signals of SWIR and MWIR will be readout through the buffered output transistors *B*1 and *B*2. Fig. 7.6 describes the timing diagram of working pulse.

There are two methods to produce the two pulse signals MWINT and SWINT at a sub-threshold amplitude: one is to connect the terminal to the peripheral driving circuit so that the pulse can be directly set as needed. The second is to provide the pulse internally, which is shown in the dotted box of Fig. 7.7. The control signal

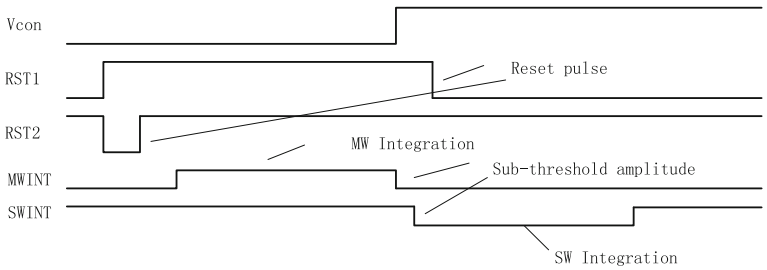


Fig. 7.6 The timing diagram of the sequential dual-band ROIC

MWINT is illustrated as follows: two switch transistors K2A and K2B are added and controlled by the pulse KP2. Two bias voltages V_{bimw} and GND are connected to the gate of the switch *M*2, respectively, through the transistor K2A and K2B. When *M*2 operates, KP2 is set to high to turn on the switch K2A and turn off the switch K2B, thus the gate voltage of *M*2 is set to the sub-threshold voltage of V_{bimw}. When the integration process is over, KP2 is changed to low so that K2B is opened and K2A is closed, then the gate voltage of *M*2 is zero and the switch *M*2 is off. Similarly, two switch transistors K1A and K1B controlled by the pulse KP1 are added and respectively connected to the bias V_{bisw} and V_{DD}. The working process is same as *M*2. The integration control pulses of all pixels share the circuit shown in the dotted box of Fig. 7.7. Compared to the pulse provided by the peripheral driving circuit, the circuit in Fig. 7.7 has two more pulse signals so that two more pins are added to the chip accordingly.

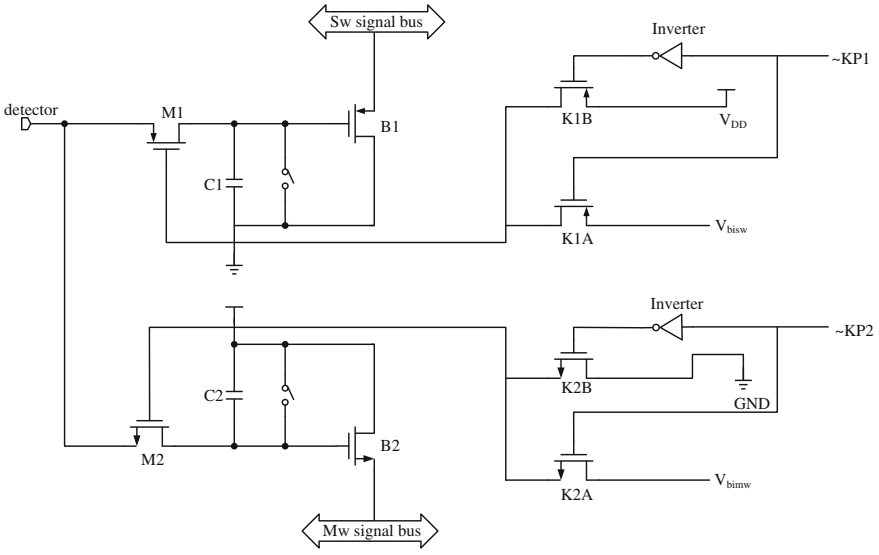


Fig. 7.7 The internal circuit to produce bias pulses for the sequential dual-band ROIC

The biggest difference between the circuit of Fig. 7.5 and the circuit of Fig. 7.2 is that two switches $M1$ and $M2$ in Fig. 7.5 not only control the integration time but also provide the bias for the detector. In this way, the circuit eliminates two transistors with the price of higher demanding on the control pulse, which is beneficial to simplify the input stage circuit and satisfy the requirements of miniaturization and high integration for IRFPA. Therefore, the circuit structure shown in Fig. 7.5 is the majority of the input stage circuit in the dual-band ROIC for IRFPA at present.

Adopting the DI structure, the input stage circuit shown in Fig. 7.8 is applied to a simultaneous mode ROIC for the back-to-back dual-band IR detector. Two transistors $M1$ and $M2$ are integration time and bias control transistors, respectively, for the long-wave photocurrent I_{lw} and the mixed photocurrent with long wave and middle wave ($I_{\mathrm{lw}} + I_{\mathrm{mw}}$). $C1$ is the integration capacitor for I_{lw} and $C2$ is the integration capacitor for the mixed photocurrent. The switch $M5$ connects two output terminals of the dual-band detector, and then the LW detector does not work for being zero-biased state while $M5$ is open. The reset switches and control pulses for capacitors $C1$ and $C2$ are not drawn in Fig. 7.8. The working process of the circuit is as follows: At the beginning, the capacitor $C1$ and $C2$ are reset to their respective reset voltage. Then the switches $M1$ and $M3$ are open, the photocurrent I_{lw} and ($I_{\mathrm{lw}} + I_{\mathrm{mw}}$) will be integrated on the capacitor $C1$ and $C2$, respectively. The switch $M5$ is always on when the LW detector does not work and no mixed photocurrent is generated in this time. When the LW detector starts to work, the switch $M5$ is turned off and the mixed photocurrent ($I_{\mathrm{lw}} + I_{\mathrm{mw}}$) is generated in the circuit. When the integration process is complete, the switches $M3$ and $M1$ are turned off, and the current signals have been transformed into the voltage signals, which will be readout through the control of the pulse READ. The timing diagram of the circuit of Fig. 7.8 is shown in Fig. 7.9.

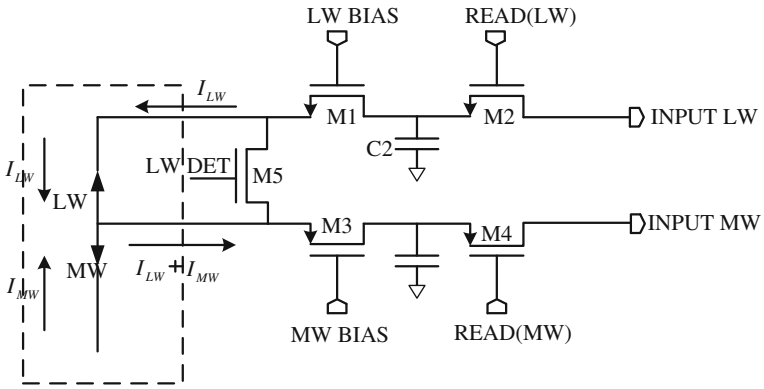


Fig. 7.8 The input stage circuit in a simultaneous mode ROIC for the back-to-back dual-band IR detector

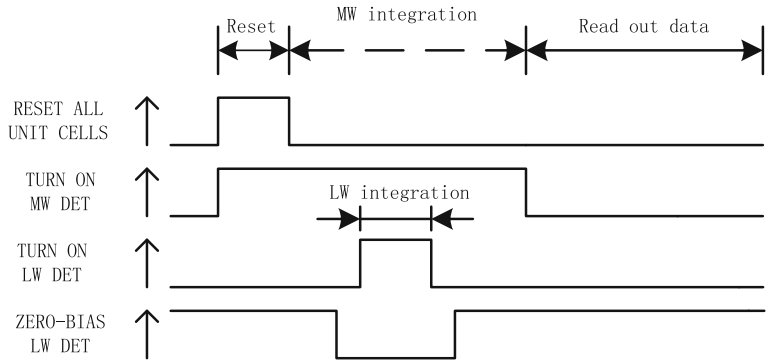


Fig. 7.9 The timing diagram of a simultaneous dual-band ROIC

The circuit structure of Fig. 7.8 just realized general functions of the input stage circuit by providing the bias for the detector and converting the current signal into the voltage signal. Since the mixed photocurrent signal exists at the terminal of a simultaneous back-to-back dual-band infrared detector due to its structure characteristic, the output of the input stage circuit is not the voltage signal for MW but the mixed voltage signal with LW and MW. Therefore, the signal separation circuit is required to be followed up the input stage circuit. The signal separation circuit structure of Fig. 7.10 is introduced for the circuit of Fig. 7.8. This circuit consists of three parts: the charge-sensing differential amplifier (CSDA), the clamp circuit, and the sample and hold (S/H) circuit. Where the CSDA can achieve a high precision only if the opening gain of the amplifier is very high (such as 500–1000) and both capacitances of C_{fb1} and C_{fb2} are completely equal. The working process of the CSDA is divided into two stages: In the first stage, the voltage signal of LW signal on the integration capacitor $C1$ of the input stage circuit is read out to the terminal Input LW of the CSDA, stored in the CSDA for the following signal subtraction process, then buffered through the clamp circuit and finally stored in the S/H circuit. In the second stage, the mixed signal with LW and MW is read out to the terminal Input MW of CSDA from the integration capacitor $C2$ of the input stage circuit, then subtracted from the signal LW stored in the CSDA to get only the MW signal, which is buffered through the clamp circuit and stored in the S/H stage.

The working procedure of the circuit is as follows: At the initial time, the pulse Φ_{RSTAMP} and Φ_{CLAMP} turn on the switches $M8$ and $M13$ to make the CSDA and the clamp circuit reset. The pulse Φ_{RSTUCL} and Φ_{RSTUCM} turn on the switch $M6$ and $M7$, so that the terminal Input LW and Input MW are reset to VRSTUCL and VRSTUCM, respectively. The switches $M8$, $M13$, and $M6$ are open, and the switch $M7$ is closed, then the CSDA is in the charging mode for the LW signal. The switch $M2$ of Fig. 7.8 is turned on, the voltage signal of LW is read out to the terminal Input LW, and stored on the capacitor C_{fb1} and C_{fbm}. Both voltages on the capacitor C_{fb1} and C_{fbm} are equal (since both capacitances are equal), which is very important for realizing the subsequent signal subtraction. The voltage signal of LW is buffered

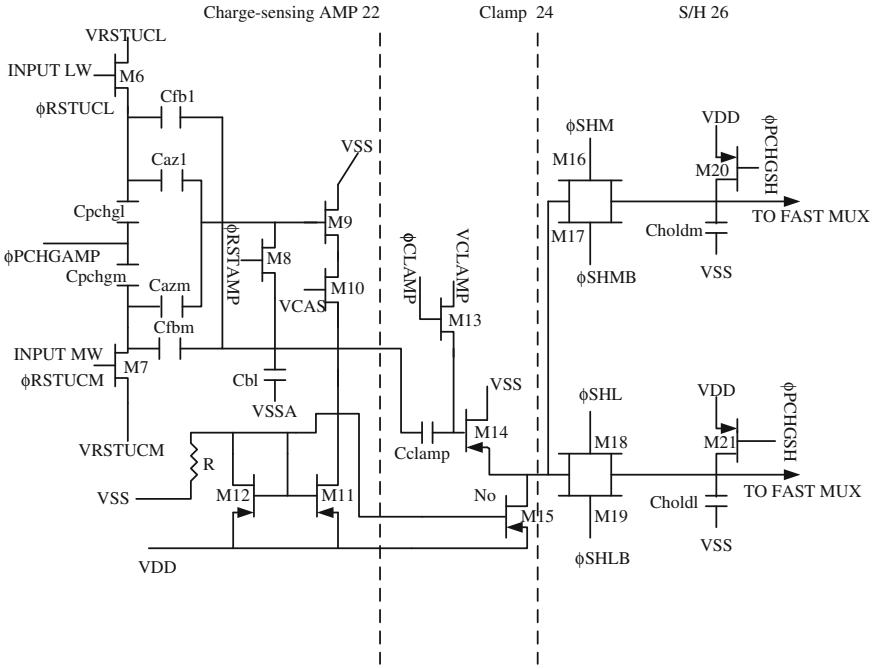


Fig. 7.10 The signal separation circuit for a simultaneous dual-band ROIC

out the node No, then sampled and hold on the capacitor C_{holdl} through the transmission gate by $M18$ and $M19$. The switch of $M7$ is turned off, and the $M6$ is turned on, then the CSDA is changed to the charging mode for the MW signal. The switch $M4$ of Fig. 7.8 is turned on, the mixed signal with LW and MW is read out to the terminal Input MW. Since the voltage signal of LW has been stored on the C_{fbm} previously, the signal of MW can be separated from the signal of (LW + MW) with subtraction. Then the MW signal is buffered to the S/H circuit through the node No, and stored on the capacitor C_{holdm} after the transmission gate by $M16$ and $M17$. The S/H circuit for the MW signal is optional and the MW signal can be directly read out from the node No.

The capacitor C_{azl} is an auto-zero capacitor. In the stage for reading out the signal of LW, the C_{azl} can ensure the voltage at the terminal Input LW to remain at the reset voltage of VRSTUCL, thus making the whole LW signal able to be charged into the C_{fb1}. The capacitor C_{azm} has a similar role. The capacitors C_{pchgl} and C_{pchgm} are optional, whose function is to supply the regulation on DC voltage for the output of CSDA. The cascode amplifier is composed of two transistors $M9$ and $M10$, and the gate of $M9$ is the input node N_{i} of the amplifier. The current source load of the amplifier is composed of two switches $M11$ and $M12$. C_{fb1} is the capacitor of bandwidth limitation as a filter. The source follower formed of the transistors $M14$ and $M15$ is an optional buffer stage for output. The main function of the clamp circuit is to reduce the reset noise of the preamplifier.

The circuit in Fig. 7.10 has a high precision on the separation of signal, while having a very complex structure. Although it could be shared by one row or one column, the circuit structure still brings up the problem of area and power consumption.

Figure 7.11 shows the input stage circuit of a sequential ROIC for the back-to-back dual-band IR detector [7]. Providing the bias for the detectors $D1$ (MW) and $D2$ (LW), the transistors $S1$ and $S2$ can be switched quickly between on and off through the pulses pDetRstM Wand pDetRst LW, thus changing the voltage at the node DN of the detector. DI1 and DI2 are the injection transistors of the detectors for MW and LW, respectively. C_{int} is the integration capacitor shared by the LW and the MW detectors. Controlled by the pulse of pResetMWUC, Rst1 is the transistor for the reset of the capacitor C_{int} before the integration of MW. The Rst2 has the same function as the Rst1 for the LW signal and is controlled by the pulse of pReset LWUC. The transistors SF1 and SF2 are used for transferring the voltage signal on C_{int} to C_{avgmw} and C_{avglw} near the end of integration for MW and LW, respectively. The capacitors C_{avgmw} and C_{avglw} are called the SFA integration capacitors, which are used to obtain the integration signal of MW and LW. The signals of MW and LW will be read out through different paths finally. The working process of the circuit is as follows: Assume the MW as the first signal, and the LW as the second signal to integrate. At the beginning, the switch $S1$ is turned on by the pulse of pDetReset MW, and the detectors are biased to bring the $D1$ into the working state and the $D2$ into the conductance state. The capacitors C_{avgmw} and C_{avglw} are reset to an initial value, and the switches SF1 and SF2 are open. Controlled by the pulse of pReset MWUC, the switch Rst1 is turned onto reset the capacitor C_{int} to vRstUCMW. The Rst1 is turned off when the reset is finished. Controlled by the pulse of vgate MW, DI1 is turned on to make the injection transistor work in the sub-threshold region, and then the current signals of MW start to integrate. SF1 is turned onto make the capacitors C_{avgmw} and C_{int} average the

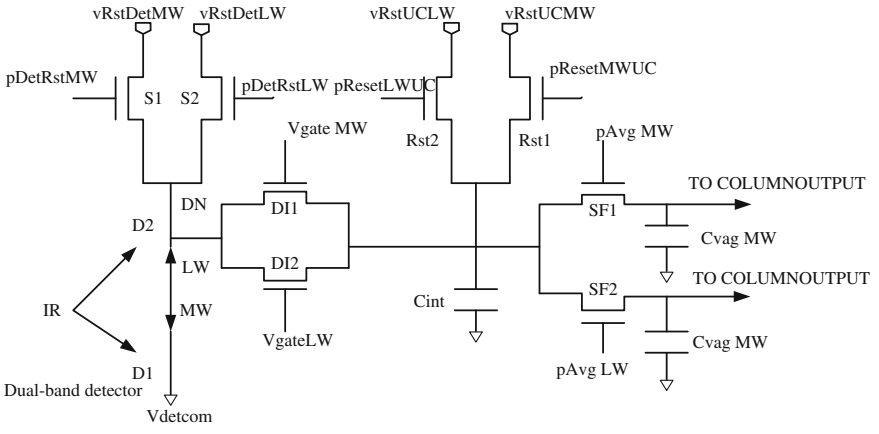


Fig. 7.11 The input stage circuit of a sequential ROIC for the back-to-back dual-band IR detector

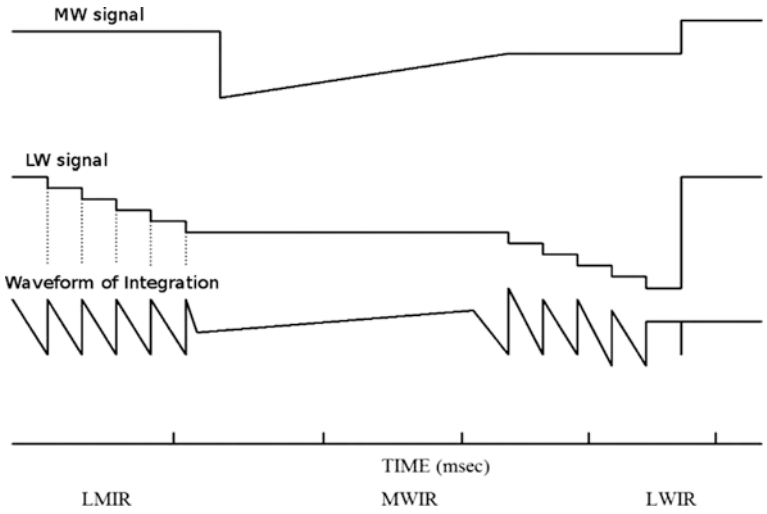


Fig. 7.12 The timing diagram of a dual-band ROIC with sequential mode

voltage signal near the end of integration. After the end of MW integration, the detector switches to the working state of LW. The working process is the same as for the MW. The timing diagram of the circuits is shows in Fig. 7.12.

In the timing diagram of Fig. 7.12, the LW current signal starts to integrate at first. The integration of the LW signal repeats several times, while the integration of the MW signal only once. Compared with the circuit of Fig. 7.5, the input stage circuit of Fig. 7.11 eliminates one integration capacitor with the price of a more complicated control logic.

The DI circuit has been widely used as an input stage circuit of a dual-band ROIC for IRFPA because that the structure of DI is simple and the area cost is small, which has a great advantage especially in the dual-band ROIC with high level of integration.

7.3.2 The Proposed Topology for a Simultaneous Dual-Band ROIC

The mixed photocurrent of two bands always exists at the output terminal of a simultaneous back-to-back dual-band infrared detector due to its structure characteristic. The functionality of signal separation has to be added to the ROIC. Since the difference of the amplitude between signals of two bands is very large, a high demand on precision and stability is required for the separation of signals. The reason why a specific signal separation circuit is provided to be subsequent to rather than incorporated in the input stage circuit of Fig. 7.8 is because the pixel area of a

large-scale dual-band IRFPA is small and the area used for the input stage circuit is limited. However, for a linear dual-band IRFPA, the area in horizontal direction used for the input stage circuit is limited, whereas the area in vertical direction is unlimited. Therefore, there is some feasibility that a signal separation circuit is incorporated with the conventional structure of input stage circuit for the design of a linear dual-band IRFPA.

Considering the current mirror gate modulation amplifier (CM) shown in Fig. 7.13, the structure consisting of $M1$ and $M2$ is called a current mirror, the capacitor C_{int} is the integration capacitor, the switch $M3$ controlled by the pulse Φ_{RST} is the reset switch of C_{int} and the control switch for integration, $M4$ is the buffer transistor of a source follower (the load is not shown in Fig. 7.13), and $M5$ is the sampling switch, which is controlled by the pulse Φ_{SAMPLE}. C_{h} is the capacitor for S/H. The basic principle of the current mirror gate modulation circuit is to employ the behavior characteristic of the transistors $M1$ and $M2$ operated in the saturation region. Neglecting channel length modulation, we can write for the drain current of $M1$ as follows

$$I_{\text{det}} = \frac{1}{2}\mu_{\text{p}} C_{\text{ox}} \left(\frac{W}{L}\right)_{M1} (V_{\text{GSM1}} - V_{\text{TP}})^2 \tag{7.10}$$

For the drain current of $M2$

$$I_{\text{int}} = \frac{1}{2}\mu_{\text{p}} C_{\text{ox}} \left(\frac{W}{L}\right)_{M2} (V_{\text{GSM2}} - V_{\text{TP}})^2 \tag{7.11}$$

In Fig. 7.13, both gates of $M1$ and $M2$ are connected together, and both sources of $M1$ and $M2$ connect VDDA, hence the voltage V_{GSM1} is equal to V_{GSM2}. We can derive

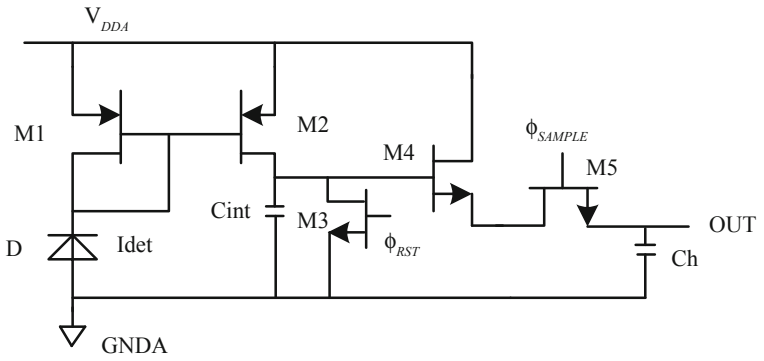


Fig. 7.13 Typical architecture of a current mirror gate modulation circuit

$$\frac{I_{\text{det}}}{I_{\text{int}}} = \frac{\left(\frac{W}{L}\right)_{M1}}{\left(\frac{W}{L}\right)_{M2}} \tag{7.12}$$

where the ratio of I_{det} and I_{int} is a constant, given by the ratio of device dimensions. If the size of $M1$ matched the size of $M2$ accurately, the current for integration is identical to the detector current.

Based on the principle that a current mirror can accurately reproduce the input current to the output, it can be designed to copy the photocurrent for one band of a dual-band detector (such as LW/MW) to two branches of the circuit by using a current mirror. If the output terminal of the LW detector connects to the input terminal of a CM, the photocurrent I_{LW} is reproduced to I_{LW1} and I_{LW2}. The current I_{LW1} is used for integration. The current I_{LW2} flows into the branch circuit for the mixed photocurrent ($I_{\text{LW}} + I_{\text{MW}}$), thus counteracting the current I_{LW} to obtain the current I_{MW}. Consider that the flow direction of the current I_{LW} and the current ($I_{\text{LW}} + I_{\text{MW}}$) are inverse ($I_{\text{LW}}$ flow into the detector, $I_{\text{LW}} + I_{\text{MW}}$ flow out the detector), the current direction of I_{LW1} and I_{LW2} are reversed through the current mirror and consistent with the current direction of ($I_{\text{LW}} + I_{\text{MW}}$), so if the MW detector is directly connected to the input stage circuit, the currents I_{LW2} and ($I_{\text{LW}} + I_{\text{MW}}$) are not offset but accumulated. Therefore, the output terminal of the MW detector is required to connect with a current mirror, and the flow direction of the output current ($I_{\text{LWD}} + I_{\text{MWD}}$) is also reversed, thus achieving the subtraction with the current I_{LW}. The implementation circuit is shown in Fig. 7.14. $M1$, $M2$, and $M3$ form a current mirror for the LW detector. The current mirror for the MW

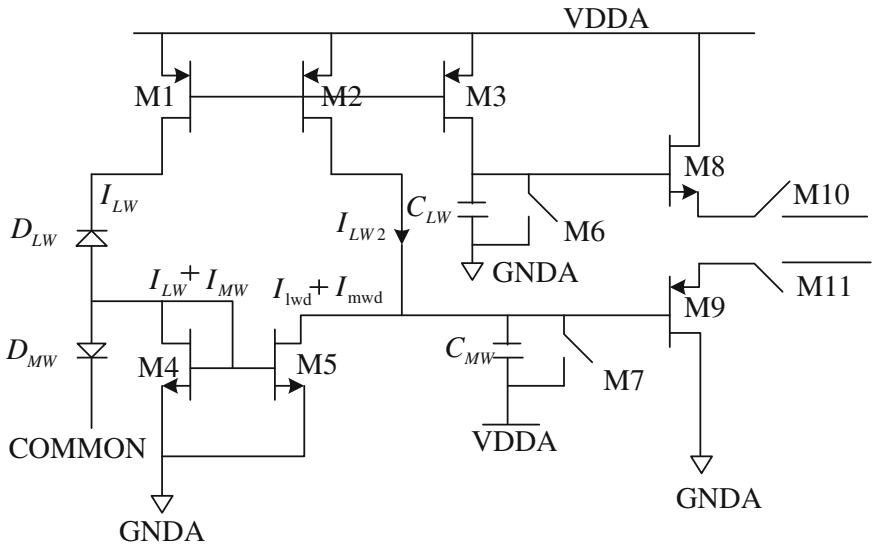


Fig. 7.14 The current mirror gate modulation circuit for a dual-band IR detector

detector is composed of $M4$ and $M5$. At node N, considering Kirchhoff's current law, we have

$$I + I_{\mathrm{LW2}} - (L_{\mathrm{LWD}} + I_{\mathrm{MWD}}) = 0 \Rightarrow I = (L_{\mathrm{LWD}} + I_{\mathrm{MWD}}) - I_{\mathrm{LW2}} = I_{\mathrm{MW}} \quad (7.13)$$

The integration current I_{MW} can be obtained from Eq. (7.13). In Fig. 7.14, $M6$ and $M7$ are the reset switch of the integration capacitor C_{LW} and C_{MW}, respectively. After the reset of capacitors is complete, $M6$ and $M7$ are turned off to start the integration of current signals. Near the end of integration, $M10$ and $M11$ are turned on to S/H the voltage signal of two integration capacitors (two sampling capacitors are now shown in Fig. 7.14).

The circuit shown in Fig. 7.14 not only possesses the function of a conventional input stage circuit but also the function for achieving the signal separation. However, this circuit cannot be applied directly because the precision for the signal separation is not high enough. As mentioned earlier, Eqs. (7.10), (7.11), and (7.12) are derived by ignoring second-order effects for a MOS transistor. Considering the effect of channel length modulation, we can write for the drain current of $M1$ and $M2$ in saturation region as follows:

$$I_{\mathrm{det}} = \frac{1}{2}\mu_{\mathrm{p}} C_{\mathrm{ox}} \left(\frac{W}{L}\right)_{M1} (V_{\mathrm{GS}} - V_{\mathrm{TP}})^2 (1 + \lambda V_{\mathrm{DSM1}}) \quad (7.14)$$

$$I_{\mathrm{int}} = \frac{1}{2}\mu_{\mathrm{p}} C_{\mathrm{ox}} \left(\frac{W}{L}\right)_{M2} (V_{\mathrm{GS}} - V_{\mathrm{TP}})^2 (1 + \lambda V_{\mathrm{DSM2}}) \quad (7.15)$$

Thus,

$$\frac{I_{\mathrm{int}}}{I_{\mathrm{det}}} = \frac{\left(\frac{W}{L}\right)_{M2} 1 + \lambda V_{\mathrm{DSM2}}}{\left(\frac{W}{L}\right)_{M1} 1 + \lambda V_{\mathrm{DSM1}}} \quad (7.16)$$

where λ is the channel length modulation coefficient. The drain-source voltage V_{DSM2} of $M2$ will change because the voltage of the capacitor C_{MW} will vary with integration. Therefore, the circuit of Fig. 7.14 cannot guarantee the precision on current reproduction. In order to suppress the effect of channel length modulation, a cascode current mirror shown in Fig. 7.15 can be used. To make $V_{\mathrm{DSM1}} = V_{\mathrm{DSM2}}$, that is, to make $V_X = V_Y$. In Fig. 7.15, $V_N = V_X - |V_{\mathrm{GS0}}| = V_Y - |V_{\mathrm{GS3}}|$, according to the size of $M3$, the size of $M0$ is properly selected to make $V_{\mathrm{GS0}} = V_{\mathrm{GS3}}$, then it can effectively suppress the effect of channel length modulation.

Figure 7.16 shows the structure of a cascode current mirror gate modulation circuit for a back-to-back dual-band IR detector [9]. The PMOS transistors M1–M6 form a current mirror CM1 to provide the bias for LW detector and the reproduction of LW photocurrent. The NMOS transistors M7–M10 form a current mirror CM2 to provide the bias for the MW detector and the reproduction of the mixed

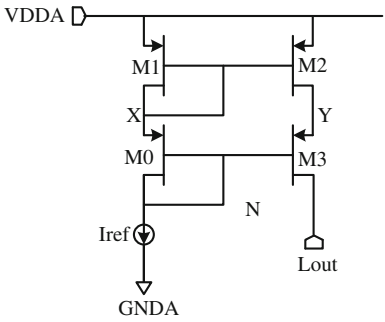


Fig. 7.15 Cascode current mirror

photocurrent. The separation of the current I_{MW} is realized at node A. $C1$ and $C2$ are integration capacitors for the LW and MW signals, respectively. $M11$ and $M12$ are the transistors for the reset and integration control of the $C1$ and $C2$ capacitors. $M13$ and $M14$ form a source follower, and so do $M15$ and $M16$. The working process of the circuit is as follows: At the beginning, the pulse LWINT turns on the $M11$, the pulse MWINT turns on the $M12$, the capacitor $C1$ and $C2$ are reset to the voltage of GND and VDD, respectively. When reset is complete, the $M11$ and $M12$ are turned off, and the LW and MW current start to integrate on the capacitor $C1$ and $C2$, respectively. Near the end of integration, the sampling transistors $M17$ and $M18$, controlled by the pulse SH1 and SH2, are turned onto S/H the voltage signals on C_{s1} and C_{s2}. The circuit structure of Fig. 7.16 provides a solution to the problem of the signal separation for a back-to-back dual-band IR detector. The simultaneous and independent integration of the photocurrent for each band can be realized within the input stage circuit. Therefore, a specific signal separation circuit is not necessary to be subsequent to the input stage circuit.

7.3.3 The Implementation of a Dual-Band Infrared ROIC and an Ultraviolet ROIC

1. **A simultaneous 256 × 2 dual-band ROIC**

(1) **The block diagram of the circuit**

Figure 7.17 is the overall block diagram of the circuit. The circuit consists of four parts: unit cell array (input stage circuit), correlated double sampling circuit, shift register, and output driver.

(2) **The input stage circuit**

The 256-pixel dual-band detector is a linear array with no dimensional limitation in vertical direction, so the area cost and structure for unit cell circuit can be relatively flexible. The circuit structure, a cascode current mirror gate modulation

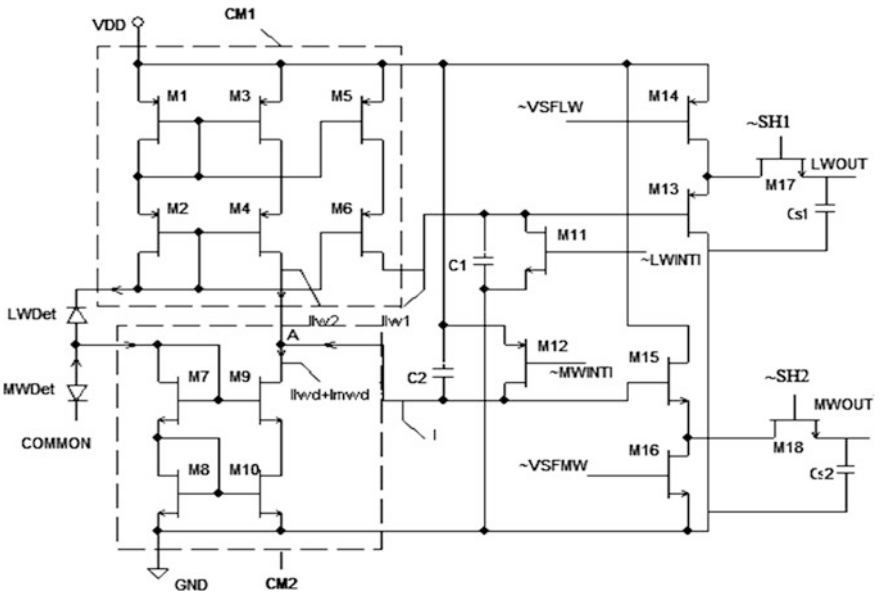


Fig. 7.16 The structure of a cascode current mirror gate modulation circuit for a back-to-back dual-band IR detector

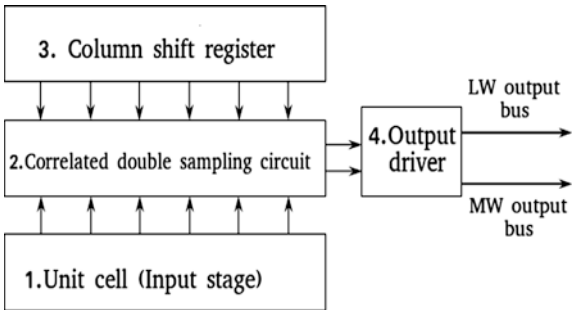


Fig. 7.17 The overall block diagram of a 256 × 2 dual-band ROIC

circuit with the function of signal separation shown in Fig. 7.16, is adopted as the input stage circuit. Section 7.3.2 details the working principle of the circuit, so this section mainly discusses the simulation of the circuit.

(a) Simulation for the precision of current mirrors

The current signal for the integration and the mixed current signal for the separation are all reproduced from the photocurrent of the detector through a current

Table 7.3 The value of the current signal for each copy of CM1 and CM2 (*unit* nA)

I_{LW}	CM1 (drain current of *M*6)	CM1 (drain current of *M*4)	CM2 (drain current of *M*9)
10	10.01	10.02	20.04
20	20.02	20.04	30.06
30	30.02	30.06	40.07
40	40.03	40.08	50.09
50	50.04	50.10	60.11
60	60.04	60.12	70.12
70	70.05	70.14	80.14
80	80.06	80.16	90.15
90	90.06	90.18	100.20
100	100.10	100.20	110.20
110	110.10	110.20	120.20
120	120.10	120.20	130.20
130	130.10	130.20	140.30
140	140.10	140.30	150.30
150	150.10	150.30	160.30

mirror. The current mirror requires high precision on the output so as to not introduce additional error on the signal. In the circuit of Fig. 7.16, the PMOS transistors *M*1–*M*6 forming a current mirror CM1 have the same W/L, the NMOS transistors *M*7–*M*10 forming a current mirror CM2 also have the same W/L. In the simulation, the input current I_{LW} is set to 10 nA at the beginning and increased to 150 nA by 10 nA for each step, the current I_{MW} is set to a constant 10 nA, and then the input range of the mixed current ($I_{LW} + I_{MW}$) is from 20 to 160 nA. Table 7.3 lists the value of the current signal for each copy of CM1 and CM2.

From Table 7.3, it can be seen that CM1 and CM2 have high precision on current reproduction. The maximum error appeared when $I_{LW} = 130$ nA. The current signal from CM1 is 130.2 nA, the mixed current signal from CM2 is 140.3 nA, then the current signal obtained by signal separation is $140.3 - 130.2 = 10.1$ nA, so the relative error for a 10 nA input signal is only 1 %.

(b) Simulation for the integration function of the input stage

According to the analysis in Sect. 7.3.2, for the circuit of Fig. 7.16, transistors *M*11 and *M*12 act as reset switches when both are turned on, while acting as integration controller when both are turned off. Figure 7.18 is the waveform diagram for the working process of MW and LW signal, which shows that the reset process and the integration process work alternatively. Figure 7.19 is the waveform diagram for integration voltages with different input current signals, which shows a good linear relationship between the integration voltage and the input current signal.

In the simulation, assuming the integration time of the LW signal is 120 μs, the integration time of the MW signal is 140 μs, the integration capacitors *C*1 and *C*2 all

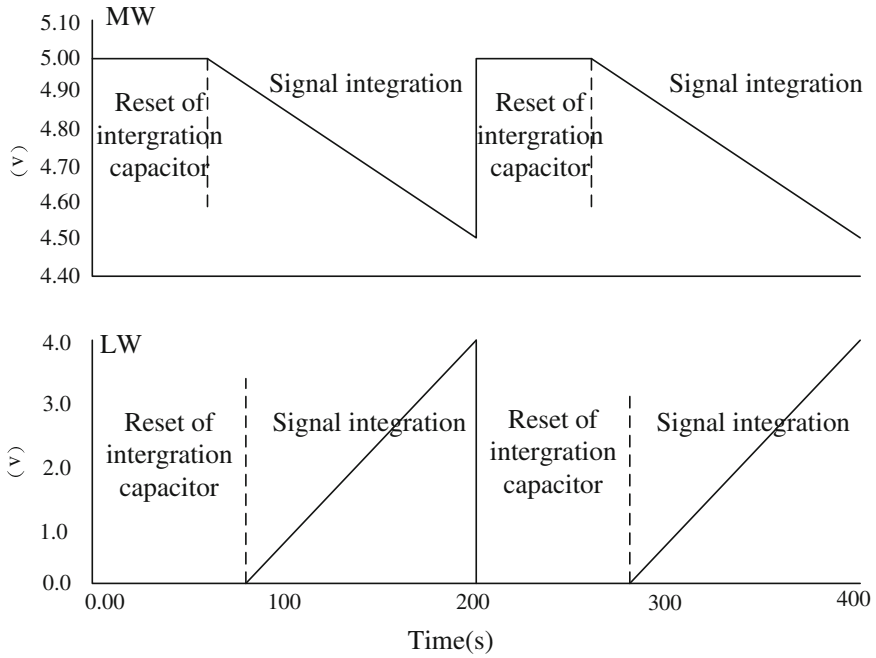


Fig. 7.18 The waveform diagram for the working process of MW and LW signal

are 3pf. Regardless of whether the voltage signal after the integration would drive the subsequent source follower in the working region, the ideal range for the LW current signal is from 10 to 110 nA, and for the MW current signal is from 10 to 100 nA. According to the simulation results in Fig. 7.19, the voltage signals at the time when the integration is complete are given by Eqs. (7.17) and (7.18), respectively

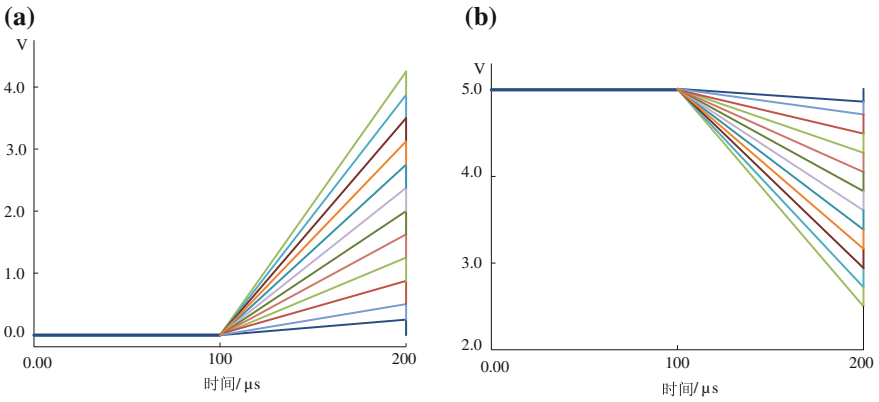


Fig. 7.19 The waveform diagram for integration voltages with different input current signals. **a** LW integration. **b** MW integration

$$V_{\mathrm{LW}} = 0 + \frac{I_{\mathrm{LW}} t_{\mathrm{LW}}}{C1} \tag{7.17}$$

$$V_{\mathrm{MW}} = V_{\mathrm{DD}} - \frac{I_{\mathrm{MW}} t_{\mathrm{MW}}}{C2} \tag{7.18}$$

Considering the voltage signal after the integration could drive the subsequent source follower in the working region, the ideal range for LW signal is unchanged, while the ideal range for MW signal is from 10 to 95 nA.

(3) **Correlated double sampling (CDS) circuit and simulation**

For the input stage circuit of Fig. 7.16, the reset process and the integration process work alternatively, the voltage signal cannot be held on the integration capacitor and is required to be sampled immediately at the end of integration, then the sampled signal is stored on the capacitors C_{s1} and C_{s2}. Replacing $M17$, $M18$, C_{s1}, and C_{s2} in Fig. 7.16 with the CDS circuit shown in Fig. 7.20 could not only realize the S/H but also be able to decrease or eliminate the effect of noise. The whole circuit structure is shown in Fig. 7.21.

Figure 7.22 is the waveform diagram for the working process of a CDS circuit. The current signal for LW band is 100 nA, and the integration time is 120 μs. The current signal for the MW band is 10 nA, and the integration time is 140 μs. The integration capacitors all are 3pf. In Fig. 7.22, the waveform of V_{LW} or V_{MW} is for the output voltage signal of the source follower consisting of $M13$ and $M14$ or $M15$ and $M16$, respectively. At the beginning of integration, the sampling switches $M17$ and $M18$ are turned off, the reset switches $S1$ and $S2$ are turned on to take the first sample and then turned off. Near the end of integration, the sampling switches $M17$ and $M18$ are turned on to take the second sample and then turned off.

The simulation results for the voltages of two times sampling for LW and MW, the voltage of the node LWOUT right to C_{h1} and the voltage of the node MWOUT right to C_{h2} are listed in Table 7.4. For MW, the voltage difference between two times sampling for the output of the source follower is −0.419 V, and the voltage variation on the capacitor C_{h2} is −0.395 V, so the error is about 5.7 %. For LW, the voltage difference between two times sampling for the output of the source follower

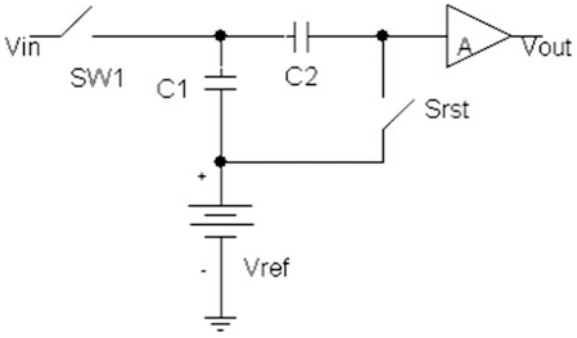


Fig. 7.20 Correlated double sampling circuit

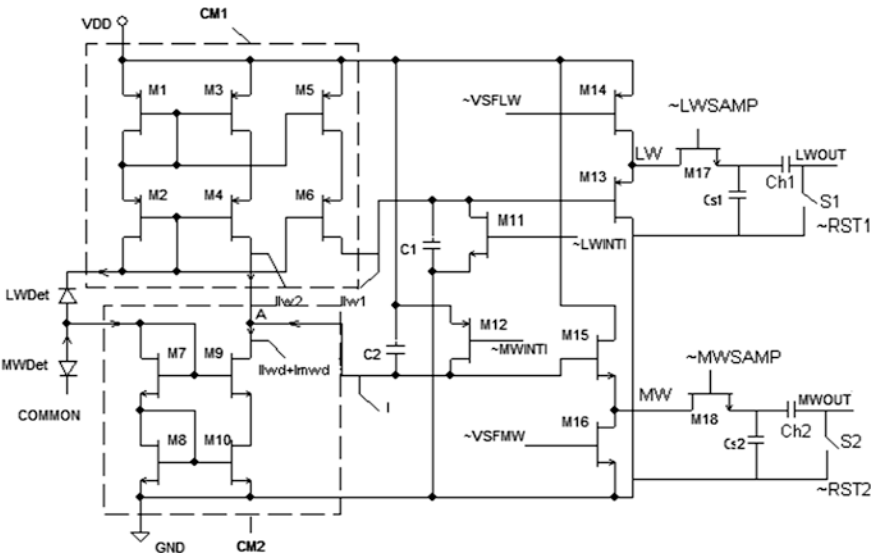


Fig. 7.21 The whole circuit structure of a simultaneous 256 × 2 dual-band ROIC

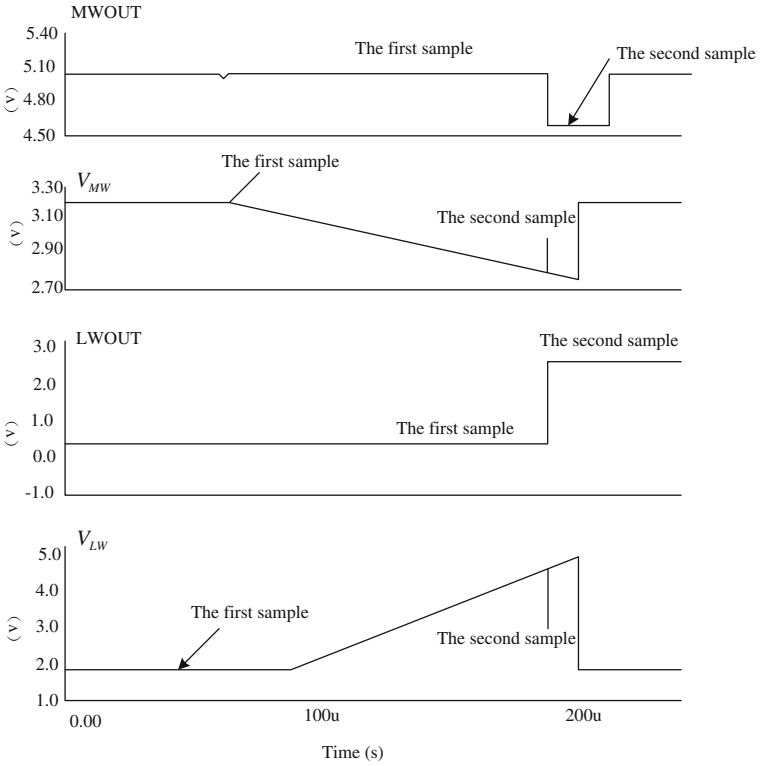


Fig. 7.22 The waveform diagram for the working process of a CDS circuit

Table 7.4 The output voltage of the source follower and the output voltage of the sample capacitor in a CDS circuit

Unit (V)	The output voltage of the source follower			The output voltage of the sample capacitor		
	The first sample	The second sample	$V2 - V1$	The first sample	The second sample	$V2' - V1'$
MW	3.206	2.787	−0.419	5	4.605	−0.395
LW	1.951	4.807	2.856	0	2.909	2.909

is 2.856 V, and the voltage variation on the capacitor C_{h1} is 2.909 V, so the error is about 1.8 %. The precision of the CDS circuit for each band is high enough to reduce the noise effectively. The error in the CDS circuit is dominated by stray capacitance of MOS switches, which introduces the redistribution of the charge in a capacitor and would affect the precision of sample.

(4) **Power analysis of the input stage circuit**

The circuit of Fig. 7.21 consists of a signal separation and integration sub-circuit and a CDS sub-circuit. In the signal separation and integration sub-circuit, the static current, set by the input photocurrent of a detector, exists both in the reset state and the integration state. Taking the maximum photocurrents for two bands, I_{LW} is 110 nA, I_{MW} is 95 nA, so (I_{LW} + I_{MW}) is 205 nA. For a source follower, the static current is determined by the drain current of the load transistor in saturation state, and is given by

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}(V_{GS} - V_{TH})^2 \tag{7.19}$$

Calculated from the above equation, the static current of the source follower for LW signal is 23 μA,the static current of the source follower for MW signal is 10.5 μA, and the maximum power of the two parts is given by

$$P = I \times V = \left\lfloor (110 \times 3 + 205 \times 2) \times 10^{-3} + 23 + 10.5 \right\rfloor \times 5 = 171.2\ \mu W \tag{7.20}$$

In the CDS sub-circuit, almost no static current exists. Compared to that of the signal separation and integration sub-circuit, the power consumption of this sub-circuit is small enough to be ignored.

(5) **The implementation of a 256 × 2 dual-band ROIC and IRFPA**

The 256 × 2 dual-band ROIC adopts the structure of the simultaneous integration circuit based on a current mirror. The input stage circuit has two terminals, and each terminal is connected to the detector for one band through the indium bump. The 256 input stage circuits are connected to subsequent CDS circuits, the signal of each channel is buffered out through the source follower. The 256 shift registers are used for addressing and to control 256 output switches of the CDS

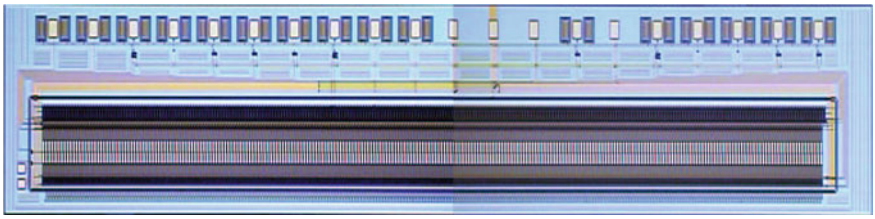


Fig. 7.23 The photograph of a simultaneous 256 × 2 dual-band ROIC chip

circuits. The ROIC is fabricated using the 0.6 μm double-poly double-metal (DPDM) CMOS process. Figure 7.23 is the photograph of a ROIC chip.

A 256 × 2 MW/SW dual-band IRFPA is assembled with a simultaneous 256 × 2 dual-band ROIC and a back-to-back dual-band detector array by indirect hybridization. The structure of the detector is n-P-P-P-n (as shown in Fig. 7.24). Based on the GaAs heteroepitaxy material, the dual-band detector is produced by using the technology of an in-place *p-n* junction combined with ion implantation. The cutoff wavelengths for two bands are 2.8 and 3.9 μm, respectively. The average detectivity are 2×10^{11} cm $Hz^{1/2}$/W and 1×10^{11} cm $Hz^{1/2}$/W, respectively. The dead pixel count is less than 8 %.

2. **A sequential 128 × 128 dual-band ROIC**

(1) **The circuit structure**

Figure 7.25 is the overall block diagram of a sequential 128 × 128 dual-band ROIC. The circuitry is composed of the input stage circuit, S/H circuit, output driver, and row/column shift register array.

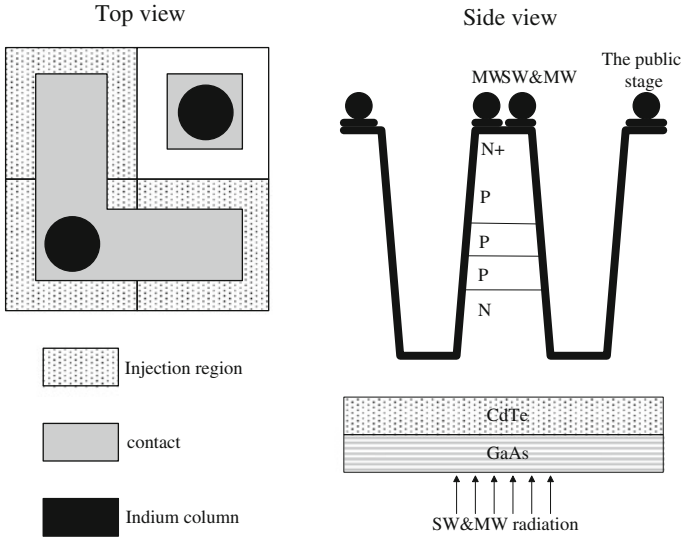


Fig. 7.24 Top view and cross section of a back-to-back SW/MW dual-band IR detector

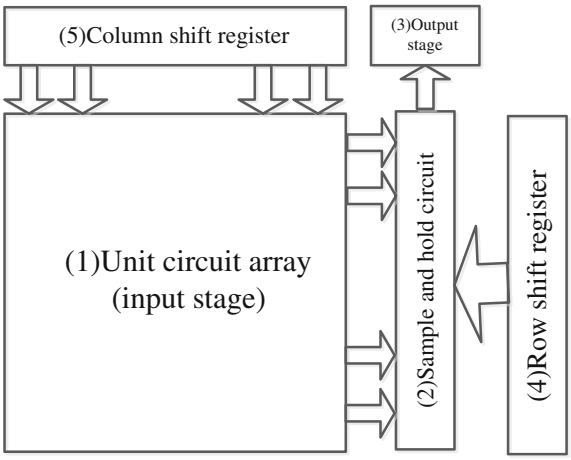


Fig. 7.25 The overall block diagram of a sequential 128 × 128 dual-band ROIC

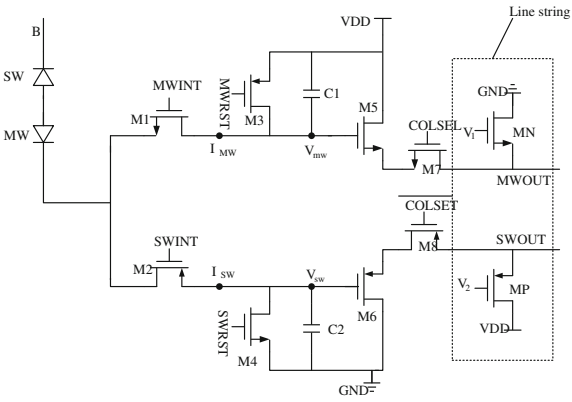


Fig. 7.26 The dual-DI structure for the input stage circuit

(2) **The input stage circuit**

The input stage circuit employs the dual-DI structure shown in Fig. 7.26, whose operating principle is described in Sect. 7.3.2. In the linear ROIC, each load transistor is only connected to a corresponding buffer transistor for the source follower of the input stage, whereas each load transistor can be multiplexed to a row of buffer transistors in the two-dimensional ROIC. Therefore, for a ROIC with 128 × 128 array, only 128 transistors are required to be loaded, so that not only is the number of MOS transistors decreased but also the power consumption is reduced.

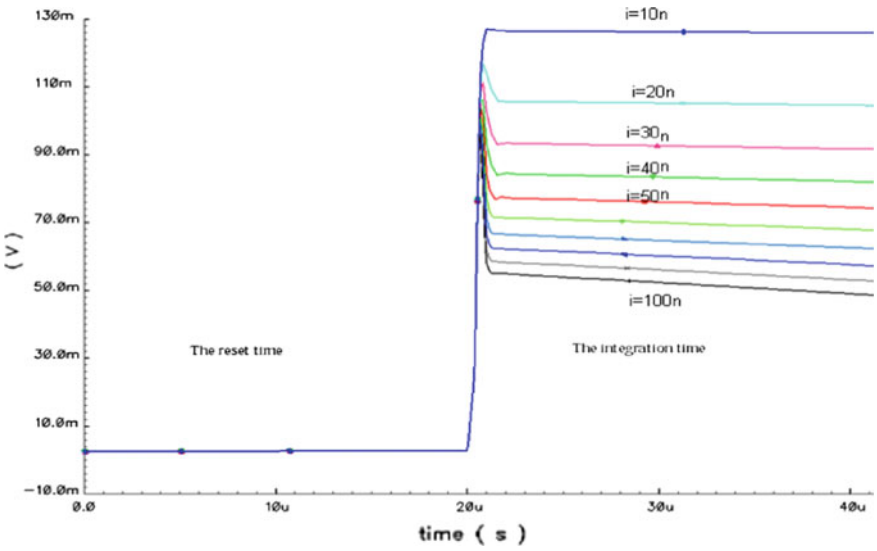


Fig. 7.27 The waveform diagram for the bias voltage of a detector with different photocurrents

Figure 7.27 is the waveform diagram for the bias voltage of a detector with different photocurrent. The range for the current is from 10 to 100 nA, the highest voltage is 126.4 mV, the lowest voltage is 54.2 mV, and the variation range of the bias is 0.0722 V. The simulation results verify that the bias of the detector provided in this circuit is much more stable than the bias provided in the gate modulation input stage circuit. So the noise introduced by the unstable bias can be reduced.

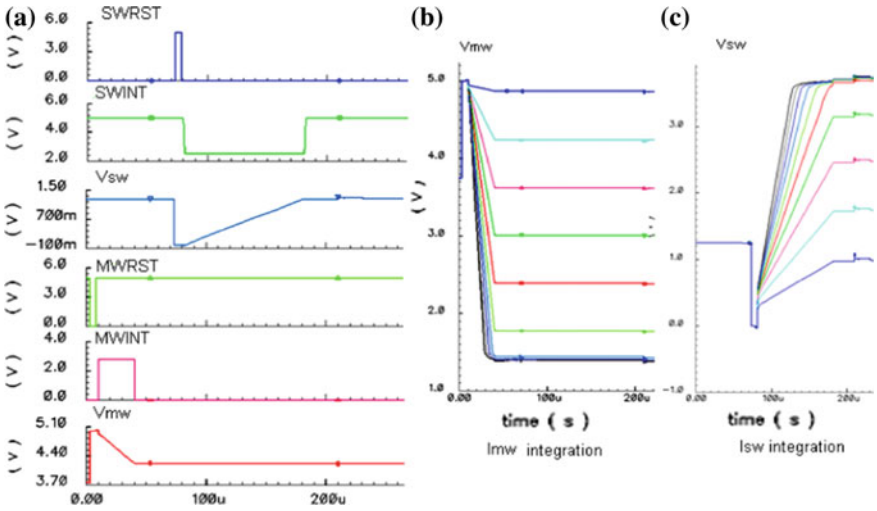


Fig. 7.28 The waveform diagram for the integration process of the input stage circuit. **a** Integration process for the sequential mode. **b** V_{MW}–I_{MW} curve. **c** V_{SW}–I_{SW} curve

Figure 7.28 is the waveform diagram for the integration process of the input stage circuit. Figure 7.28a is the waveform of the reset signal, the integration control signal and the voltage on the integration capacitor for two bands. The working process of a sequential dual-band ROIC is as follows: reset of the integration capacitor for MW → integration and hold for MW → reset of the integration capacitor for SW → integration and hold for SW → simultaneous readout of voltage signals for both bands, that is why the term "sequential" is used. In the simulation, the pulse width of the reset signal is 5 μs, the integration time for MW is 30 μs, the integration capacitance for MW is 0.5pf, the integration time for SW is 30 μs, and the integration capacitance for SW is 0.5pf. Figure 7.28b, c are, respectively, the waveform for the integration on the capacitor $C1$ and $C2$ with different input current. It is shown that the slope for integration changes linearly with the integration current.

(3) **Sample and hold circuit and output driver**

The S/H circuit and output driver are shown in Fig. 7.29. The voltage signals after integration are sampled and held on the capacitors C_{h1} and C_{h2} through respective column switches, and then readout through the row selection signal. $M1$ and $M2$ are the buffer transistors, and a large ratio of W/L is required to provide a drive current enough for the load.

Figure 7.30a is the waveform of the output voltage V_{out} versus the input voltage V_{in} for the output driver. It can be seen that when the input voltage is less than 0.7 V, there is almost no output. This is because when the input voltage is less than the threshold voltage V_{TN} of the MOSFET, the transistor stays in the cutoff region. After the input voltage exceeds 0.7 V, the output voltage goes up with the increase

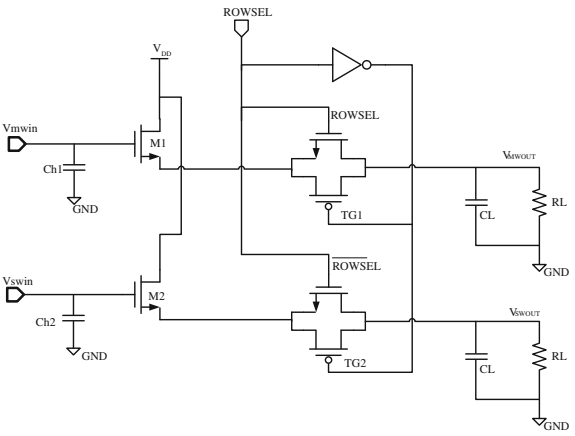


Fig. 7.29 Sample and hold circuit and output driver

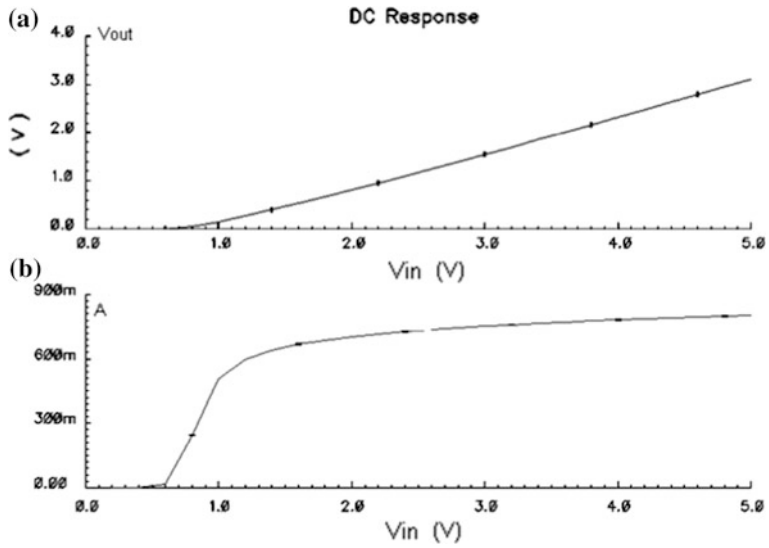


Fig. 7.30 The simulation for the source follower. **a** *V*out – *V*in. **b** *A* − *V*in

of the input voltage and the output and the input present an approximately linear relationship. To reduce the voltage drop between the output and the input, it is required to improve the gain of the source follower by increasing the ratio of W/L for the buffer transistor. Figure 7.30b shows how the gain A of the source follower changes with the input voltage *V*in. The gain of the source follower will reach more than 0.7 to achieve a good driving capability only when the input voltage is greater than 2.5 V, so the ratio of W/L for the buffer transistor has to be adjusted in order to have a wide input range. In addition, if the voltage difference between the output and the input is reduced, the maximum drain current will decrease, and the driving capability of the source follower will drop. Therefore, there is a trade-off between the output voltage and the driving capability.

(4) **Simulation of the whole circuit and Analysis of power consumption**

(a) Simulation of the whole circuit

The whole circuit schematic of a 128 × 128 dual-band ROIC is built up with the Cadence Composer tool. Digital pulses, input currents, and bias voltages are connected to the input terminal of the circuit. Capacitors and resistors as load are connected to the output terminal of the circuit. Some pads are set for testing. Thus a simulation platform is established to do the analysis on the function and performance of the whole circuit.

The working process of the circuit is divided into two stages: reset and integration stage and readout stage. In the stage of reset and integration, the pulse MWRST resets the integration capacitor for MW and is effective at a low level. When the reset is complete, the pulse MWINT turns on the switch to start the

integration for I_{MW}. After the integration for the MW signal, the pulse SWRST resets the integration capacitor for SW, and then the pulse SWINT turns on the switch to begin the integration for I_{SW}. When the integration for both bands is finished, the voltage signals are sampled and held on the capacitors $C1$ and $C2$, respectively, and are read out by the control of column selection signal. In readout stage, controlled by the first output pulse of column shift register array, the voltage signals on the first column are read out to the sampling capacitors before the output driver, and then are successively buffered to the output bus by the control of the output pulse of row shift register array. When the above process is complete, the voltage signals on the second column are read out to the sampling capacitors by using the second output pulse of column shift register array, and then are readout through the control of the output pulse of row shift register array. The frame period of the circuit is not ended until finishing the readout for the voltage signals on the 128th column.

Figure 7.31 is the waveform diagram of the output for 32 pixels of one column in the circuit. The simulation results show that the input current range for MW is from 0 to 90 nA, the output voltage is 1.553 V; the input current range for SW is from 2 to 10 nA, the output voltage is 1.695 V.

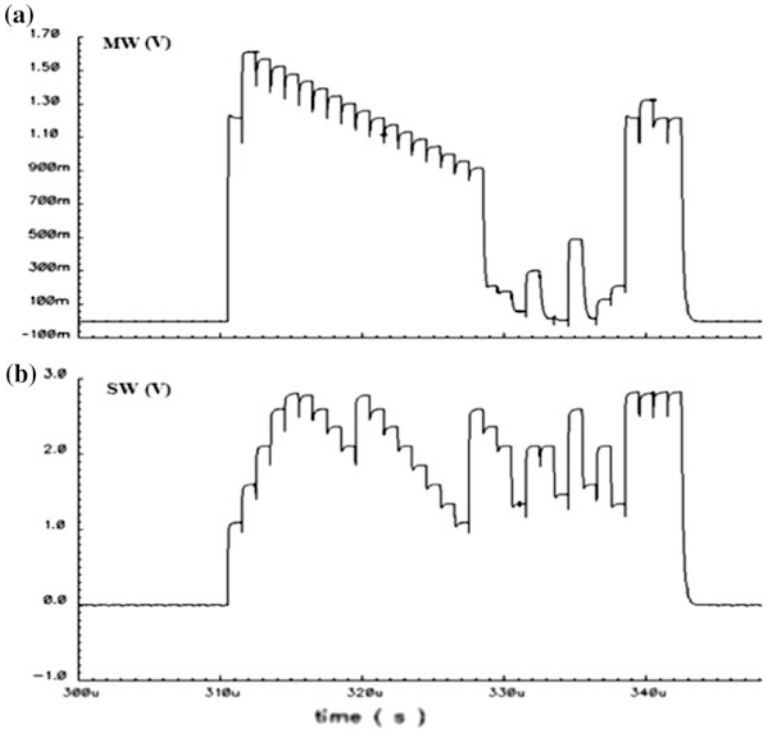


Fig. 7.31 The waveform diagram of the output for 32 pixels of one column

(b) Analysis of power consumption
The biggest integration current is 95 nA in the unit cell, and the integration time is a small part of the frame period, so its power consumption can be given by

$$P1 = 128 \times 128 \times 95\ \text{nA} \times 5 \times \frac{90\ \mu\text{s}}{9.31\ \text{ms}} = 15\ \mu\text{W} \tag{7.21}$$

In the unit cell, the load transistor of the source follower is shared by one row, the tail current for NMOS transistor is roughly 12 μA, the load current for PMOS transistor is about 5.5 μA, so the power consumption for the source follower is given by

$$P2 = 128 \times (12 + 5.5)\ \mu\text{A} + 5 = 11.2\ \text{mW} \tag{7.22}$$

Only one output driver is on during each readout time, the drain current of the driver is about 320 μA, and the power consumption is given by

$$P3 = 2 \times 5 \times 320\ \mu\text{A} = 3.2\ \text{mW} \tag{7.23}$$

According to the calculating formula for dynamic power consumption, the power consumption $P4$ of shift register array is about 10 μW. Therefore, the total power consumption of the circuit is $P = P1 + P2 + P3 + P4 = 14.4$ mW. Compared with the 256 pixels linear ROIC, the power consumption of this circuit has greatly reduced by sharing a load transistor of the source follower in one row.

(5) **A 128 × 128 dual-band IRFPA**

Figure 7.32 is the layout of a sequential 128 × 128 dual-band ROIC. The ROIC, with a 128 × 128 MW/SW dual-band IR detector array, is assembled

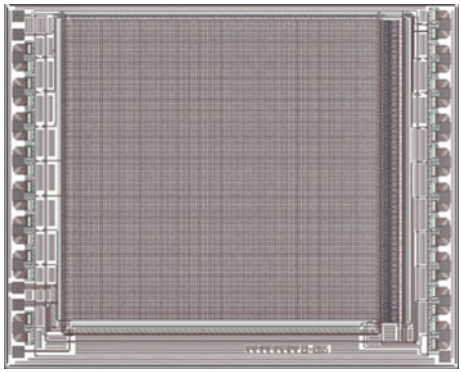


Fig. 7.32 The layout of a sequential 128 × 128 dual-band ROIC

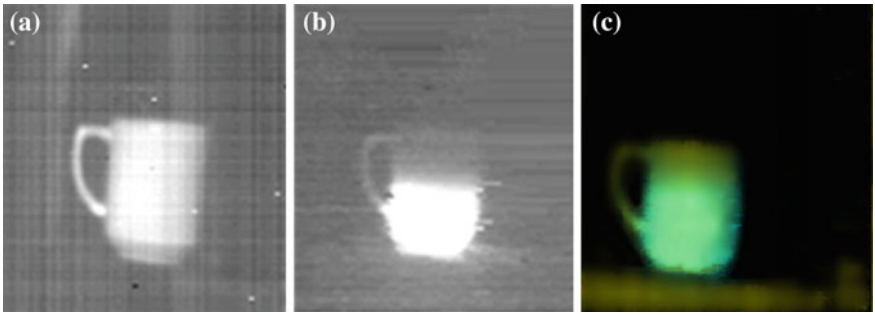


Fig. 7.33 Infrared images for MW and SW band of a 128 × 128 MW/SW dual-band IRFPA. **a** Infrared image for SW band. **b** Infrared image for MW band. **c** Image fusion with two infrared bands

to a 128 × 128 MW/SW dual-band IRFPA using direct hybridization. The average detectivity reaches 1.5×10^{11} cm $Hz^{1/2}/W$ (SW) and 2×10^{11} cm $Hz^{1/2}/W$ (MW). Figure 7.33 shows two infrared images respective for MW band and SW band of a 128 × 128 MW/SW dual-band IRFPA, and also the image with the fusion of MW signal and SW signal.

3. **The implementation of an ultraviolet ROIC**

Because of the characteristics of high impedance and small photocurrent for an ultraviolet photodiode, the CTIA structure is used as the input stage circuit of the ROIC for an ultraviolet IRFPA. In addition, the CTIA structure has the advantage of low noise. Figure 7.34 is the block diagram for a 128 × 128 ultraviolet ROIC. The pixel size is 50 μm × 50 μm, and the chip size is 8.5 mm × 8.5 mm. Figure 7.35 is the layout of a 128 × 128 ultraviolet ROIC and the photograph of

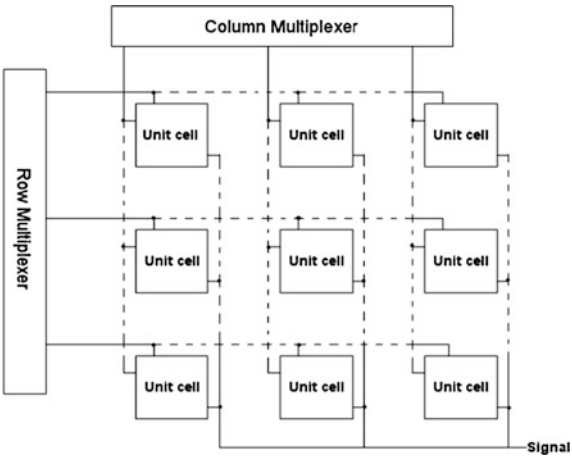


Fig. 7.34 The block diagram for a 128 × 128 ultraviolet ROIC

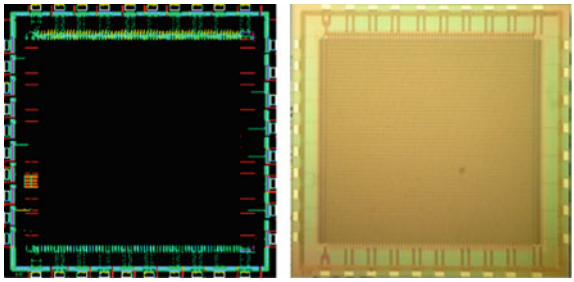


Fig. 7.35 The layout of a 128 × 128 ultraviolet ROIC and the photograph of chip

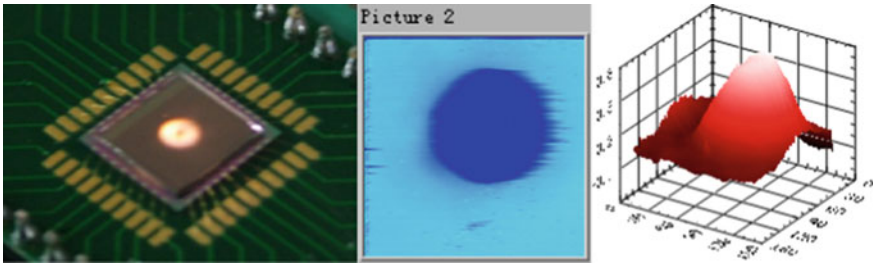


Fig. 7.36 The photograph for a visible–blind ultraviolet sensor assembly focused with circular light spot, an ultraviolet image, and a 3-D data graph

chip. The surface of the ROIC has been coated with a protective layer resistant to the light, thus avoiding the impact of visible light on the ROIC.

The 128 × 128 ultraviolet ROIC, with different ultraviolet detector arrays, is assembled to respectively form visible–blind and solar-blind type 128 × 128 ultraviolet IRFPA. The measured parameters for visible–blind 128 × 128 ultraviolet IRFPA are as follows: The maximum frame rate for readout is 150 Hz, the dead pixel count is 1.66 %, and the response uniformity is 7.15 %. Figure 7.36 is the photograph for a visible–blind ultraviolet sensor assembly focused with a

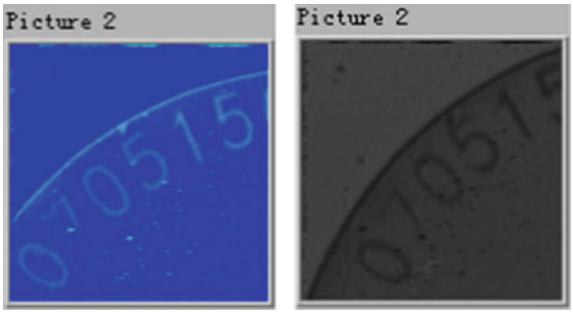


Fig. 7.37 The photograph for the laser marker of "070515" on a solar-blind ultraviolet sensor assembly

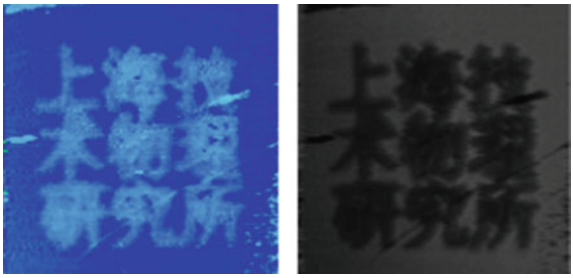


Fig. 7.38 The photograph for the imaging solar-blind ultraviolet sensor assembly

circular light spot, an ultraviolet image and a 3-D data graph. Figure 7.37 is the photograph for the laser marker of "070515" on a solar-blind ultraviolet sensor assembly. The measured parameters for the solar-blind 128 × 128 ultraviolet IRFPA are as follows: The maximum frame rate for readout is 150 Hz, the dead pixel count is 7.8 %, and the response uniformity is 3.3 %. Figure 7.38 is the photograph for the imaging solar-blind ultraviolet sensor assembly.

7.4 Digital Transmission System on Chip for IRFPA

The continuous expansion on the scale of IRFPA and the demand for improving the performance of infrared imaging bring a large amount of data processing requirements. The signal-to-noise ratio (SNR) and bandwidth of the IRFPA are influenced and restricted with such factors as environmental noise, various parasitic effects and the time for the charge and discharge of stray capacitance. The inherent limitation of analog circuits for the signal processing has become the constraint in the development of IRFPA to large-scale and high-frame rate. A digital transmission SOC for IRFPA could eliminate noise sources such as cross talk, clock jitter, and electromagnetic interference related to analog transmission. Taking advantage of the rapid development on CMOS integrated circuits, the digitalization of the ROIC not only enhances the anti-interference ability and improves the performance of focal plane imaging, but also can reduce the system cost and simplify the electronic interface. For a digital transmission SOC, the ADC is the core circuit. The development trend of on-chip ADC circuit is toward higher precision, lower power consumption, and higher speed. In recent years, on-chip ADC circuits have been realized in the practical application. The digitalization of the ROIC is indispensable for the incorporation of more signal processing functionality (such as automatic correction, data compression, etc.) in the future, and has been an important field for the development of IRFPA.

After the year of 1996, the Amain Microelectronics Company of the United States has developed the 64 × 64 and 256 × 256 LW MCT IRFPA, the pixel pitch is 60 and 30 μm, respectively, a chip-level ADC structure is adopted with the multiplexed oversampled analog-to-digital conversion (MOSAD) technology, and

the precision of the output is 12-bit [10]. In 2000, the Amain Company introduced a 640 × 480 MW MCT IRFPA using the MOSAD technology, the precision of the output is 12-bit, the pixel pitch is 27 μm, and the power consumption of the chip is less than 100 mW [11]. After that, a new 128 × 128 MW InSb IRFPA was developed with the MOSAD technology [12], the precision of the output is 14-bit, and the power consumption is less than 50 mW.

In 2003, Israel's Semiconductor Company reported a 640 × 512 MW InSb IRFPA using the column-level ADC structure with the integration of 1280 A/D converters [13]. The precision of the output is 15-bits, the frame rate is 100 Hz, the power consumption is less than 120 mW, the input charge-handling capacity is greater than 13 Me, and the algorithm of the ADC is dual-ramp single-slope (DRSS).

In 2006, the French Sofradir Company developed a 640 × 512 MW IRFPA with a digital interface [14]. The column-level ADC solution using the algorithm of single-ramp dual-slope is provided, the column ADC pitch is 30 μm, the frame rate is 120 Hz, the power consumption for each column ADC is 65 μW, the precision of the output is 15-bit, while the chip area was increased by 18 % greater than the ROIC with analog interface.

7.4.1 The Architecture of the Digital System for IRFPA

At present, most digital imaging systems are composed of discrete independent components. A typical discrete digital imaging system is shown in Fig. 7.39.

The clock driver module mainly provides control signals to IRFPA. The components of the IRFPA are designed to convert the infrared radiation signal to the voltage signal in analog form. Using the high-speed ADC circuit, the analog signal is converted to the digital signal that is stored in RAM. An ASIC module is mainly responsible for controlling the entire digital imaging system and achieving the necessary functionality of signal processing and data compression.

At present, the digital imaging system is limited to the application of microdevices because of its high power consumption and large area cost. The integration of

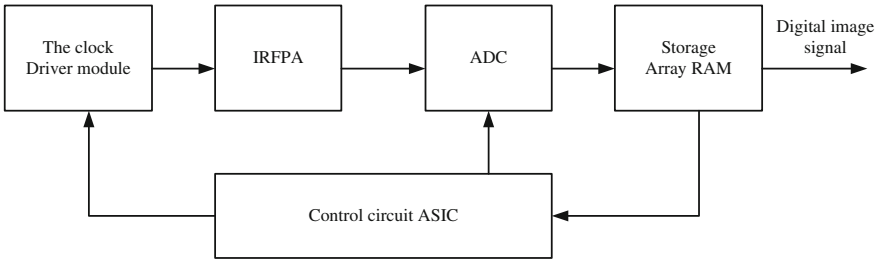


Fig. 7.39 The block diagram of a discrete digital imaging system

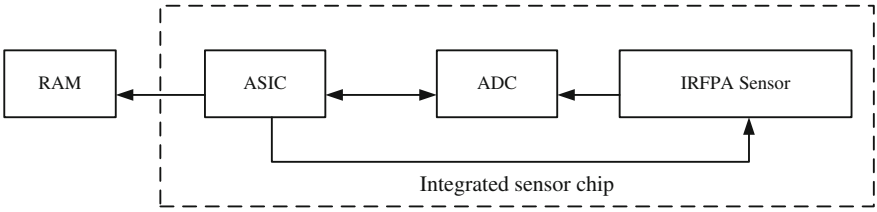


Fig. 7.40 The block diagram of the future digital imaging system

system requires a reduction in the system size and power consumption, and is limited because different circuit modules usually adopt different technologies. Therefore, clock driver, ASIC controller, and ADC modules shown in Fig. 7.39 shall be integrated into the ROIC of IRFPA. The future digital imaging system is shown in Fig. 7.40.

An integrated sensor chip consists of the IRFPA sensor, ADC circuit, ASIC controller, and signal processing circuit. According to the correspondence between the ADC circuit and the pixel array, the on-chip ADC implementation scheme mainly consists of three structures: chip-level ADC structure, column-level ADC structure, and pixel-level ADC structure.

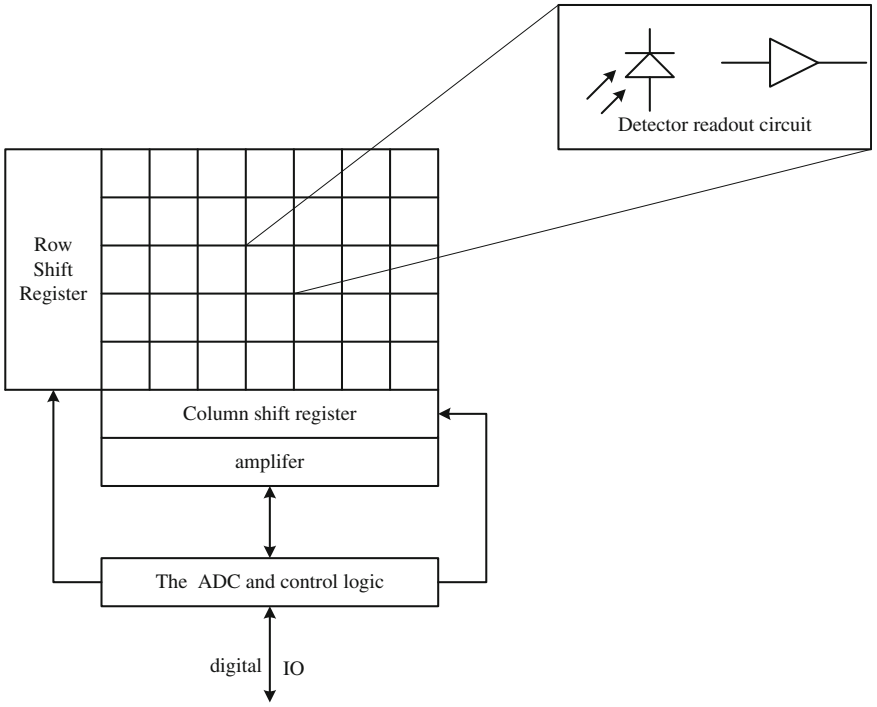


Fig. 7.41 The circuits of chip-level ADC structure

1. **Chip-level ADC structure**

As the simplest on-chip integration structure, the chip-level ADC structure would incorporate a single ADC circuit into a ROIC, that is, the output driver is followed by a high-speed ADC circuit, which converts the analog signal of each pixel into a digital output signal in turn. The circuit is shown in Fig. 7.41, and the working process is as follows: First, a unit cell circuit converts the photocurrent signal of a pixel into the analog voltage signal, which is multiplexed and amplified to the output bus through a row shift register and a column shift register; Then, each analog voltage signal is converted to the digital voltage signal by an ADC circuit and outputs in turn.

In the early days of the digital imaging system, the ADC module was a discrete circuit chip. The on-chip integration of the analog-to-digital conversion and other functionality has the potential benefit to reduce the size, power consumption, and cost of the system and enhance the reliability. Additionally, the on-chip integration of ADC could improve the antinoise performance of the circuit chip through the conversion of the analog signal to the digital signal at the beginning of the signal chain in the system.

Compared with traditional readout circuits (only for the amplification and readout of signal), the circuits of a chip-level ADC structure have an increase on the power consumption with the on-chip integration of the analog-to-digital conversion, and the power consumption is more with the higher operating frequency of ADC. The advantage of this structure is that the size of an ADC circuit is not constrained by the pixel pitch of the focal plane array (FPA), and the conventional technologies for CMOS circuit design and layout implementation are allowed. Therefore, the chip-level ADC structure is usually used in the application for low frame rate and low resolution (8-bits or lower). In 1991, the Peter Denyer research group from University of Edinburgh developed a 256 × 256 ROIC by using the chip-level ADC structure [15]. The on-chip ADC circuit adopts the algorithm of high-speed successive approximation, the precision of output is 8-bits, and the power consumption is 100 mW.

Compared with the conventional readout circuits, the size and power dissipation of the output bus driver in the circuit of Fig. 7.41 have become smaller because the driver is designed to drive the input capacitance of an ADC circuit but not the load capacitance induced by external signal chain and circuits board. The input sampling capacitor of an ADC circuit must be large enough to make the ADC circuit achieve a good SNR and linearity, otherwise it would be adversely affected by the KT/C noise and the constraint in component matching.

As mentioned above, the chip-level ADC structure still has some limitations in the application for IRFPA.

2. **The Column-level ADC structure**

For the circuits of chip-level ADC structure, when the scale of the FPA is large, the conversion speed of the ADC is demanded to be high and the power consumption would be very high also. In order to overcome these shortcomings, the

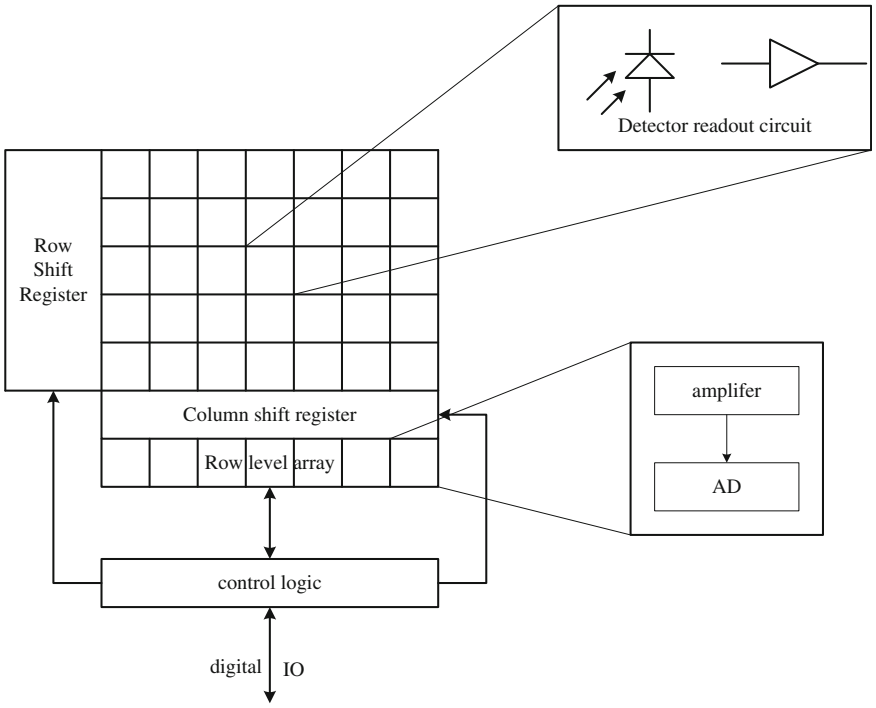


Fig. 7.42 The circuits of column-level ADC structure

column-level ADC structure could be used as shown in Fig. 7.42, where each column (row) or n columns (rows) ($n > 1$) share an ADC circuit, and multicolumn ADC circuits are parallel connected to the multiplexer for output.

The column-level ADC structure consists of a two-dimensional FPA and a linear ADC array, which has two correspondences: one is that each column of the FPA is followed with an ADC; the other is that multiple columns of the FPA share an ADC through a multiplexer. Thus, the conversion speed of the ADC in the column-level ADC structure could be much slower than that of the ADC in the chip-level ADC structure.

For the chip-level ADC structure, the transmission for serial data on the output bus has the highest rate in the system, which introduces a large amount of white noise and brings a high requirement for the bandwidth of the ADC. The column-level ADC structure can be to used to obtain the frame signal with a lower clock rate so as to reduce the clock noise,and the capacitive coupling noise is also reduced because it would increase with the increase of the clock rate.

This configuration is a benefit for the design of the ADC circuit because of the column-level ADC structure has a low requirement for the conversion speed. However, the layout implementation is more complex because the pitch of the two-dimensional FPA must match with that of the linear ADC array.

Dickinson et al. from AT&T and JPL of the United States developed a 176 × 144 ROIC by using the column-level ADC structure with the single-slope integration algorithm [16]. The frame rate is 30 Hz, the precision of the output is 8-bits, and the power consumption is 35 mW. A 640 × 512 InSb MWIR FPA introduced by the Israeli SCD Company and a 640 × 512 HgCdTe MWIR FPA developed by the French Sofradir Company in 2006 all adopt the column-level ADC structure in the design [13, 14].

The advantage of the column-level ADC structure is to eliminate the output driver, the multiplexer and the column buffer, which all have high power consumption. The analog signal path between the unit cell circuit and the input of the ADC circuit is also shortened. For the column-level ADC structure, the size of ADC is only limited by the pixel pitch in one direction, whereas the conversion speed of ADC is dependent on the scale of the FPA and the frame rate. The column-level ADC is a trade-off between the chip-level ADC structure and the pixel-level ADC structure, and is a circuit structure in practical application at present.

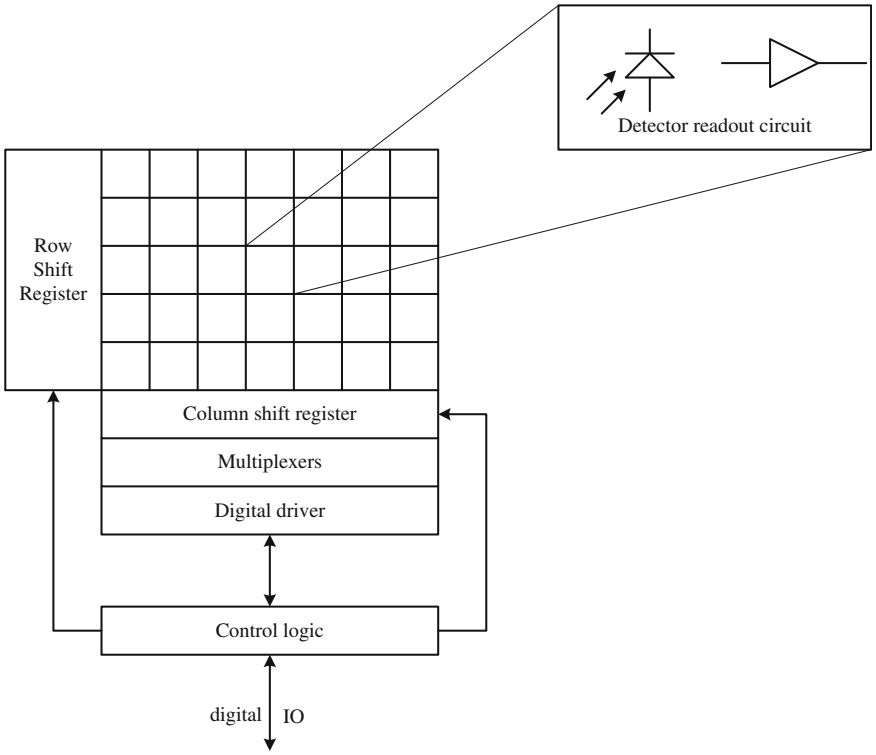


Fig. 7.43 The circuits of pixel-level ADC structure

3. **Pixel-level ADC structure**

The pixel-level ADC structure works like a fully parallel structure. As shown in Fig. 7.43, each pixel of the FPA contains an input stage circuit with A/D conversion or $n \times n$ pixels share an ADC circuit. The basic structure of the input stage circuit could be the DI, BDI or CTIA. The analog signal of each pixel is directly buffered to or multiplexed to the ADC, and then in turn is converted to the digital signal. For different scales of the FPA, the conversion speed of the ADC circuit is always constant.

The biggest advantage for the pixel-level ADC structure is that the requirement for the ADC conversion speed is the lowest, so an ADC with small power consumption can be used to help meet the low power demands for the integration on IRFPA. In addition, the analog pixel signal is directly connected to the ADC circuit, the analog signal path is the shortest, and so the noise is minimum. Adopting the pixel-level ADC structure, the digital signals of the FPA can be further processed to achieve more advanced functions, such as exposure control, dead pixel elimination, etc., and the image compression also can be implemented on-chip to reduce the requirement for external driver circuits. In 1999, the Amain Electronics Company of the United States developed an 640×480 MW MCT IRFPA by using the pixel-level ADC structure, where the MOSAD algorithm is adopted.

The pixel-level ADC structure has some disadvantages. Especially due to the limitation by the pixel area, the algorithm of the ADC circuit only has a few choices except for the pipeline ADC and the successive approximation ADC with a large area cost, so the trade-off is between the area cost, design precision, and power consumption. With the continuous increase in size for the FPA, the pixel pitch is smaller, the more difficult is the layout implementation for the pixel-level ADC structure, especially when the pixel pitch is very small and the demand for the ADC resolution is very high. At present, the progress for the pixel-level ADC structure is still in the stage of laboratory research. Taking advantage of the rapid development of CMOS process and the shrinking feature size, more design optimizations can be made for the pixel-level ADC structure. So considering the advantages compared with the chip-level ADC structure and the column-level ADC structure, the pixel-level ADC structure would be the future development trend for the digitalization of IRFPA.

7.4.2 Algorithms for the Implementation of ADC on the Focal Plane

1. **The principle of A/D conversion**

An ADC circuit is to convert an analog signal to the corresponding digital signal (binary code). The analog signal is continuous in time and amplitude, while the digital signal is discrete. First, the input analog signal is sampled at a series of time

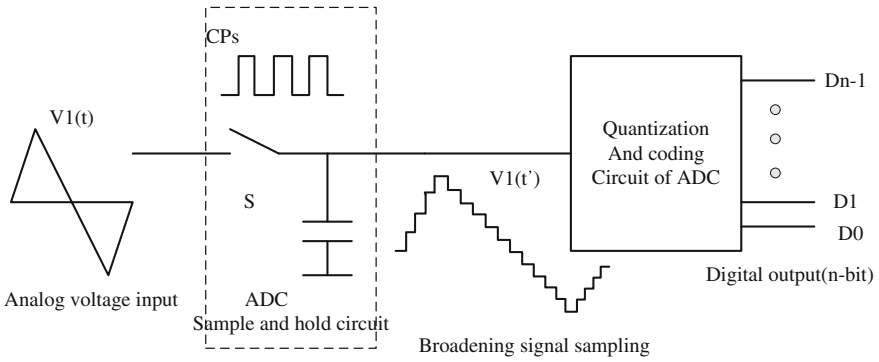


Fig. 7.44 The block diagram of the process of A/D conversion

points. Then the sampled value is converted to the corresponding digital value, which is discrete on the amplitude.

Figure 7.44 shows a block diagram of the process of A/D conversion, which consists of sample, hold, quantization, and coding. Controlled by a sampling pulse, the sampling circuit samples the input signal at a series of time points; the hold circuit maintains the sampled signal during the conversion; the quantization circuit converts the held signal to the binary value most approximate; and the coding circuit converts the binary value to the corresponding code with a sign bit.

2. **Algorithms for the implementation of an A/D conversion on focal plane**

With the column-level ADC structure or the pixel-level ADC structure, the ADC circuit demands a small area cost, simple implementation circuit, and slow conversion speed. So the main algorithms for the implementation of the ADC circuit on focal plane could consist of the voltage-frequency-modulation ADC, the sigma-delta ADC, the Multichannel Bit-Serial (MCBS) ADC, and the integral (single-slope/dual-slope) ADC, etc.

(1) **The voltage-frequency-modulation ADC**

The main function of a voltage-frequency-modulation ADC is to integrate the photocurrent by the integration capacitor and compare this value with a reference voltage. When the output of the comparator switches, a pulse signal is generated and the integration capacitor is reset. In the same time interval, the frequency of the pulse signal is proportional to the value of photocurrent, and the corresponding digital value is obtained through a counter. The block diagram of the ADC circuit is shown in Fig. 7.45, which mainly consists of the detector, the detector capacitance C_d, the comparator, a reset transistor, a feedback control circuit, and a counter.

The working process of the circuit is as follows: When light is irradiated onto the detector, the generated photocurrent causes the detector capacitor to discharge, resulting in a decrease of the potential V_d. The potential V_d and the reference level V_{ref} are compared in a comparator, the comparator is triggered by a clock signal

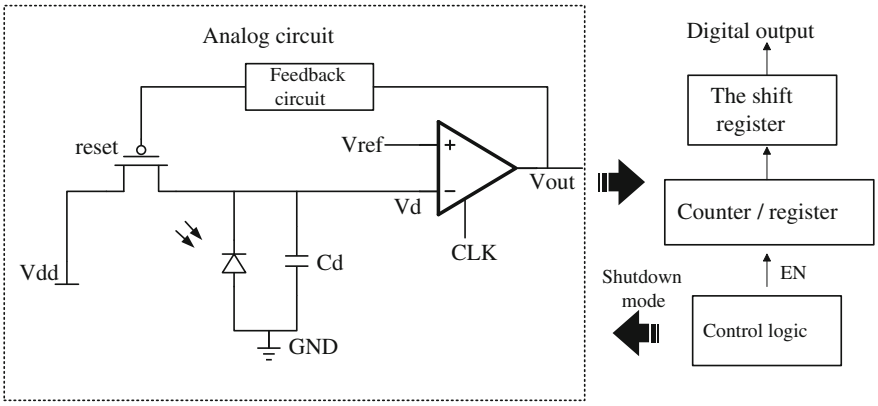


Fig. 7.45 The block diagram of the voltage-frequency-modulation ADC

CLK, and V_{ref} is produced by a digital-to-analog converter (DAC) on chip or a voltage divider. When the potential V_d is lower than V_{ref}, the comparator output turns to high, which generates a feedback to reset PMOS transistor, then the potential of the detector capacitor is reset to V_{dd}. In other words, when the potential V_d reaches V_{ref}, the output of the comparator will generate a pulse, the period of pulse is

$$T_d = \frac{C_d(V_{dd} - V_{ref})}{i_d}, \quad f = \frac{1}{T_d} \tag{7.24}$$

where i_d is the detector photocurrent and C_d is the internal capacitance, f is the frequency of pulse. Obviously, i_d is proportional to the frequency of pulse. Controlled by an external enable signal En, a counter only works in a cycle of T_{en}, and the number of pulses N is given by

$$N = \frac{T_{en}}{T_d} = \frac{2^n T_{clk}}{T_d} = \frac{2^n T_{clk}}{(V_{dd} - V_{ref})C_d} i_d \tag{7.25}$$

where n is the ADC resolution, T_{clk} is the period of the clock signal Clk.

Figure 7.46 are timing waveforms for the signals V_{dd}, V_{out}, Clk, and En. When the level of the signal En goes low, the counting is finished, then the ADC enters reading mode. The result of the counter is stored in the shift register and read out in serial. Meanwhile, the analog circuits of the ADC are in standby mode, thereby reducing the power consumption of the circuit.

The advantages of the voltage-frequency-modulation ADC are as follows:

(a) The circuit structure is simple and easily implemented.
(b) The dynamic range (DR) is large and suitable for different applications.

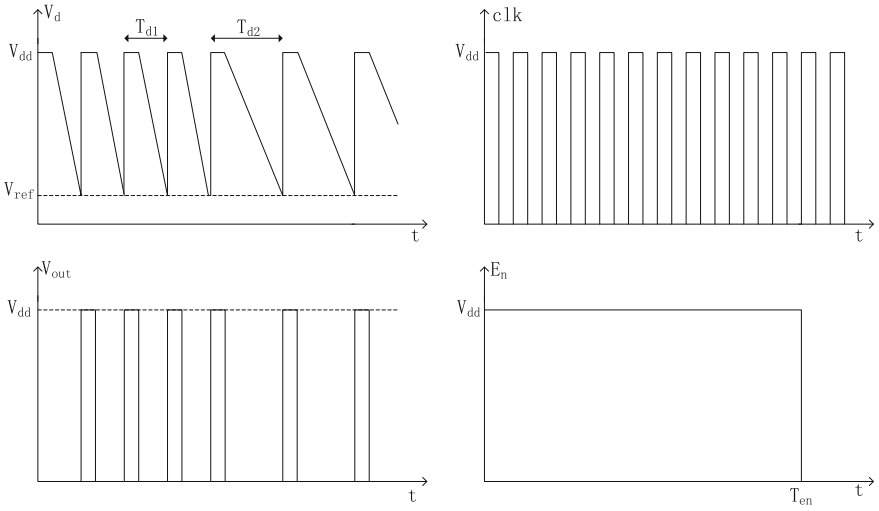


Fig. 7.46 The timing waveform for the signals V_{dd}, V_{out}, CLK, and En

The disadvantages are as follows:

(a) Each pixel requires a counter. The area cost of the counter is larger as the precision increases, and the conversion speed is not fast.
(b) The period of the signal En is determined by the maximum photocurrent and the integration capacitance C_d. It is difficult to accurately determine.

(2) **The sigma-delta (Σ-Δ) over sampling ADC**

Figure 7.47 shows the block diagram of a Σ-Δ oversampling ADC circuit, which mainly consists of an anti-aliasing filter, the Σ-Δ modulator and a low-pass decimation filter. The anti-aliasing filter is used to limit the input signal within a certain bandwidth. For the oversampling ADC, the bandwidth f_0 of the input signal is much smaller than the half of the sampling frequency f_s, so the transition zone ($f_s - 2f_0$) between the passband and the stop band of the anti-aliasing filter is wide, thus lowering the design requirement and making the usage of a low-order analog filter possible. The Σ-Δ modulator is used to convert the oversampled analog signal to a high-speed and low-precision digital signal, which is converted to the high-precision signal at Nyquist frequency through a low-pass decimation filter.

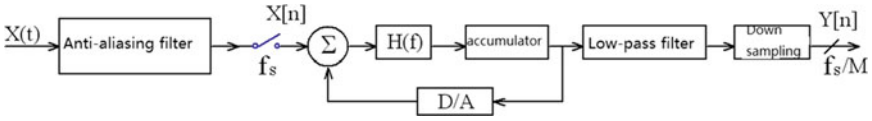


Fig. 7.47 The block diagram of a Σ-Δ oversampling ADC circuit

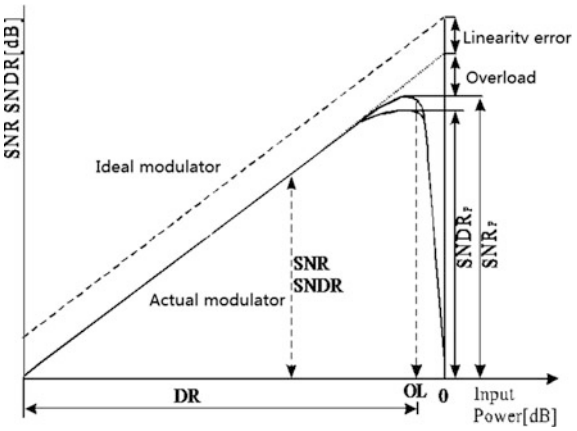


Fig. 7.48 Typical performance chart of a Σ-Δ oversampling ADC

The modulator could reduce the noise and nonlinearity introduced by the oversampling ADC so as to lower the requirement on the accuracy of analog devices. In addition, the S/H circuit could be eliminated for the switched-capacitor implementation of the oversampling ADC.

The significant parameters representing the performance of Σ-Δ oversampling ADC are as follows: DR, signal-to-noise ratio (SNR), signal-noise-distortion ratio (SNDR), effective bit (ENOB), and overload (OL). Fig. 7.48 is the typical performance chart of a Σ-Δ oversampling ADC, the horizontal axis is the normalized input, that is, V_{in}/V_{ref}, and the vertical axis is SNR or SNDR, both are in the units of dB. As can be seen from the diagram, when the input amplitude is small, the magnitudes of SNR and SNDR are equal. With an increase in the level of the input, distortion will reduce the performance of the modulator, so SNDR would be smaller than the SNR with large input. Figure 7.48 also shows that the performance of the nonideal modulator is lower than that of the ideal modulator because of the finite gain and the OL of the practical modulator. The SNR of a Σ-Δ oversampling ADC is improved mainly through the oversampling and the noise shaping techniques, thus obtaining the higher accuracy. In addition, the adoption of a multi-bit quantizer is also a basic method to enhance the SNR of a Σ-Δ oversampling ADC with broad bandwidth.

The Σ-Δ modulator makes the quantization noise spread in a broader band through over sampling. The majority of the quantization noise is moved to the higher band, so the quantization noise in baseband is much reduced. The low-pass digital filter is used to eliminate the quantization noise outside the base band. The reduction of the quantization noise in the baseband is equivalent to an increase the effective resolution of the ADC. After the quantization noise in high frequency is eliminated, the sampling rate is down to the Nyquist rate because the Nyquist rate would minimize the information necessary for subsequent transmission, storage,

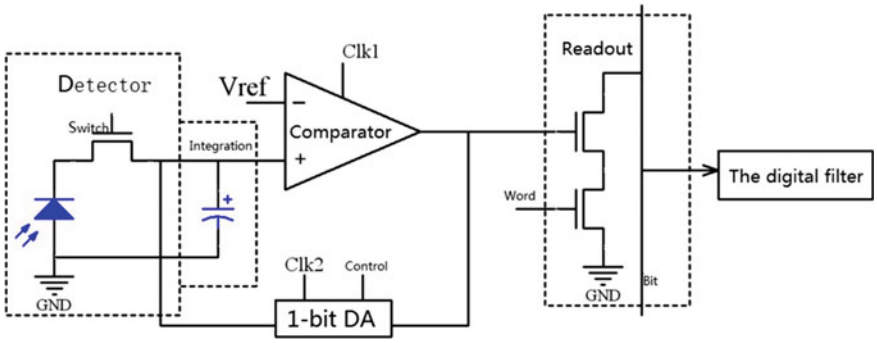


Fig. 7.49 The block diagram of the first-order Σ-Δ ADC circuit on the pixel level

and digital signal processing. The most economical and conventional digital filter for the decimation is a comb filter.

The Σ-Δ oversampling ADC is a high-precision ADC. The characteristic of the circuit is to gain the precision with the price of time through oversampling, to realize the simplification of analog circuits by increasing the complexity of digital circuits, thus achieving high precision and low power consumption. More than 95 % of the circuits are digital circuits. In addition, the Σ-Δ oversampling ADC circuit has a large charge-handling capacity and good quantization ability at the middle of the input range. The fixed pattern noise (FPN) of the circuit is quite small with respect to the infrared detector. Therefore, a Σ-Δ oversampling ADC is suitable for infrared detection applications.

The first-order Σ-Δ oversampling ADC has the simplest circuit structure. Its on-chip implementation easily meets the requirement for area cost and power consumption for IRFPA. The working principle of the first order is: Every time the analog input signal is sampled, a 1-bit digital code is produced to represent the amplitude of the input signal. Each digital code is sequentially read out in high speed to form a sequence of digital codes, which will be converted to a digital word corresponding to the Nyquist sampling rate through a low-pass filter and decimation.

Figure 7.49 shows a block diagram of a first-order Σ-Δ ADC circuit on the pixel level. The circuit consists of a clock-controlled comparator (controlled by Clk1), an integration capacitor, a 1-bit DAC, the readout structure and the digital filter exterior to the pixel array. The working process of the circuit is as follows: The photocurrent is integrated on the integration capacitor through a switch transistor, then the integration voltage is quantized to 1-bit digital value by the clock comparator, where the control pulse Clk1 is the oversampling frequency. The quantized value is converted to a voltage signal by a 1-bit D/A converter (controlled by Clk2), and the signal is fed back and added to the voltage on the integration capacitor, thus completing a conversion cycle. At the end of each cycle, a digital code will be

produced and buffered out through the BIT line, and finally read out through the digital filter on chip or external to the chip. Although each digital code generated cannot be predicted, the statistic of a digital bit-stream would reflect the value of the photocurrent.

In addition, a multistage error integrator can be used to improve the conversion speed. Although multiple integrators will create a stability problem, this could be solved with the use of a cascaded (MASH) structure.

The advantages of the first-order delta-sigma ADC are

(1) The charge-handling capacity increases due to the feedback of 1-bit DAC.
(2) Low noise. The integration reduces the noise of DAC; the feedback reduces the noise of comparator. Compared with the quantization noise, the above noise, the flicker (1/*f*) noise and the thermal noise are small enough to be ignored.
(3) The circuit structure is simple.
(4) High precision and high SNR. According to the oversampling theory, the precision of ADC increases one bit every time the sampling frequency increases four times.

The disadvantages of the first-order delta-sigma ADC are

(1) The requirement for a digital filter. If the digital filter is external to the chip, the demand for the I/O bandwidth of the chip is very high, especially for high-speed and large-scale array.
(2) Output data rate is too high.

(3) **The MCBS ADC**

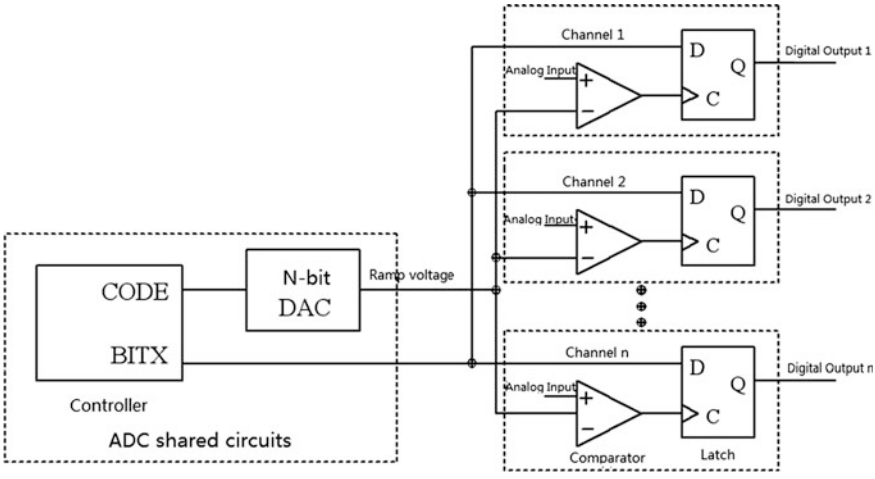


Fig. 7.50 The block diagram of the MCBS ADC circuit on the pixel level

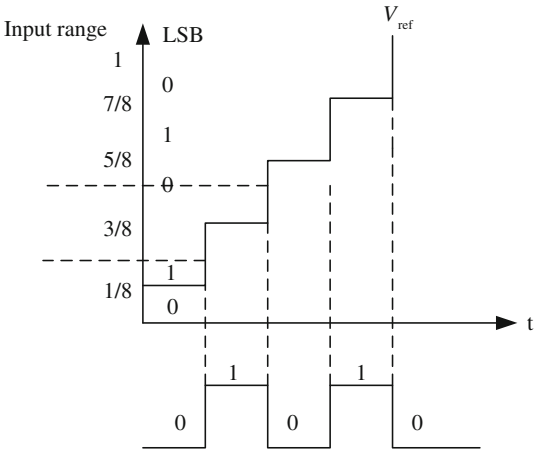


Fig. 7.51 The waveform for the operation of comparator/latch

In order to reduce the output data rate of the circuit, a MCBS ADC circuit working at the Nyquist frequency could be adopted on the pixel level. The block diagram of such an ADC circuit is shown in Fig. 7.50.

The pixel circuit of the MCBS ADC consists of a 1-bit comparator and a latch, and the remaining on-chip circuits include a finite state machine controller and a *N*-bit DAC. The control circuit outputs a BITX signal and the *N*-bit digital codes, which are converted to the corresponding analog ramp voltage signals though the DAC.

Similar to the traditional approximation comparison ADC, the input voltage signals and the ramp voltage signals are compared successively and quantized, the results are stored in the latches though the signal BITX. Different from the traditional approximation comparison ADC, the serial bit is read out independently, the digital output is not fed back to the *N*-bit DAC.

Figure 7.51 is the waveform for the operation of the comparator/latch. The input voltage signal S and the reference level V_{ref} (generated by the DAC circuit) are connected respectively to the positive input and negative input of the comparator. The output of the comparator is then connected to the clock input of the latch. The level of BIT is connected to the data input of the latch. V_{ref} starts from 0 and monotonously steps through the critical point of a range of 1. At the same time, BIT starts from 0 and switches as long as V_{ref} changes, as shown in Fig. 7.49. Once V_{ref} is equal or greater than *S*, the comparator would switch and latch the value of BIT, which is the desired LSB. Since V_{ref} is monotonous, the comparator only switches once so that the desired value can be latched. After V_{ref} across all critical points, the latched values are read out. Then V_{ref} and the BIT are reset for next comparison cycle. In Fig. 7.49, when the LSB is generated, the critical point of V_{ref} is (1/8, 3/8, 5/8, 7/8) of the input range. The MSB is generated by comparing S and 4/8 of the input range. The NMSB is generated by comparing *S* and (2/8, 6/8) of the input range. Similarly, the conversion principle of 3-bit ADC could be extended to an m-bit ADC, and $2^m - 1$ comparisons are required to complete the conversion in an *m*-bit ADC. Table 7.5 is the Gray

Table 7.5 Table for gray coding when me = 3

ADC input range	Code word
0–1/8	000
1/8–2/8	001
2/8–3/8	011
3/8–4/8	010
4/8–5/8	110
5/8–6/8	111
6/8–7/8	101
7/8–1	100

coding table when $m = 3$. The Gray coding is helpful not only to reduce the error code, but also to decrease the number of comparison.

The advantages of the MCBS ADC are

(a) The ramp voltage and the BITX signal are generated by the shared circuit. The level of Gain FPN and Offset FPN can be reduced.
(b) Having the complete testability. In the reset mode, the photodiode output can be compared with an arbitrary potential through the ramp voltage signal, and the potential can be digitized and used to calibrate the ADC.

The disadvantages of the MCBS ADC are

(a) The precision and conversion speed is lower than the oversampling Σ-Δ ADC. The maximum precision of ADC is determined by the gain of comparator, the conversion speed is determined by the gain and bandwidth of comparator.
(b) The area cost in the pixel is high.

(4) The Single-Slope/Dual-Slope Integral ADC

The principle of the integral ADC is to transform the analog voltage signal into a time interval, which is then converted to the corresponding digital word. A clock pulse with a stable frequency is applied for the conversion, and the result of conversion is the count number for the clock pulse by using a counter, which is proportional to the amplitude of the analog voltage.

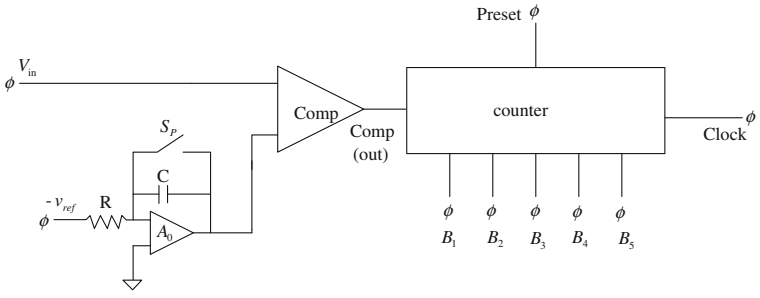


Fig. 7.52 The block diagram of a single-slope integral ADC circuit

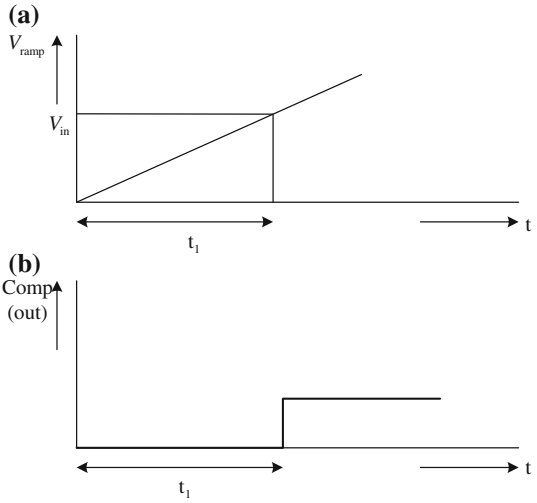


Fig. 7.53 The waveform for the conversion of a single-slope integral ADC

Figure 7.52 shows the block diagram of a single-slope integral ADC circuit, which is composed of an integrator, a comparator, and a counter. The integrator is used to generate a ramp voltage signal V_{ramp} with a reference input V_{ref}. V_{ramp} is connected to one input of the comparator, and the input voltage signal V_{in} is connected to the other input of the comparator. Figure 7.53 is the waveform for the conversion of a single-slope integral ADC. Before the conversion, the counter is cleared, and the integrator is reset. When V_{ramp} starts to ramp up, the output of the comparator initiates the counter and the conversion begins. When V_{ramp} reaches the level of V_{in}, the output of the comparator would switch and stop the counter, thus achieving the conversion from analog voltage to digital counts. The time of conversion is given by

$$t_1 = RC\frac{V_{\text{in}}}{V_{\text{ref}}} \tag{7.26}$$

The digital output of the counter is given by

$$N_{\text{digital}} = t_1 \times f_{\text{clock}} \tag{7.27}$$

The accuracy of the ADC circuit is dependent on the stability of the clock pulse, the RC time constant of the integrator, and the reference voltage V_{ref}. The offset of a comparator is measured with zero input voltage, and converted to a digital count, which can be used to preset the counter, thus achieving the self-compensation for the offset.

Compared with the single-slope integral ADC circuit, a dual-slope integral ADC circuit, as shown in Fig. 7.54, would effectively improve the accuracy and speed of analog-to-digital conversion. The circuit of Fig. 7.54 consists of the input switch, an integrator, a comparator, a counter, and a clock generator.

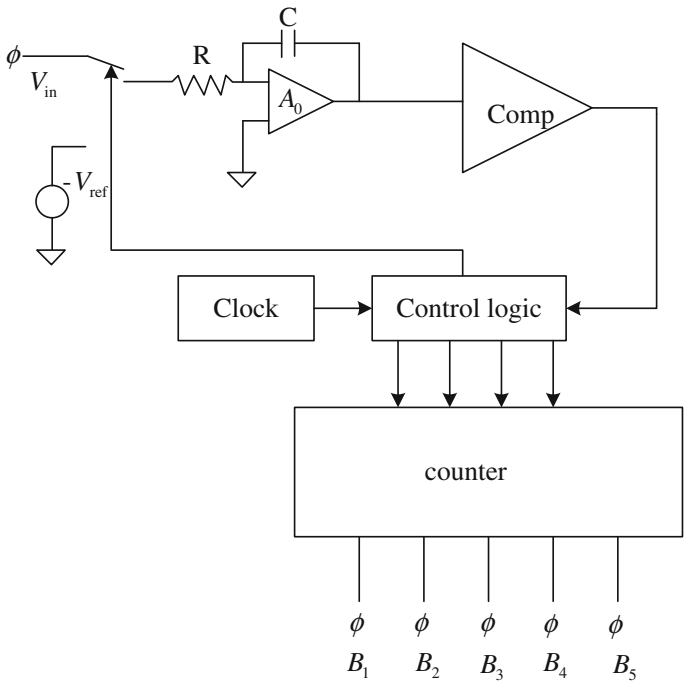


Fig. 7.54 The block diagram of a dual-slope integral ADC circuit

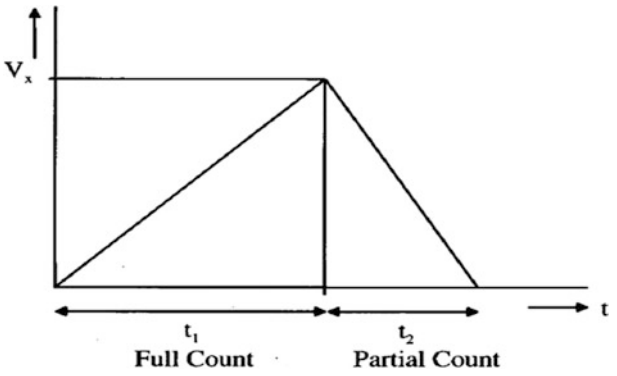


Fig. 7.55 The waveform for the conversion of a dual-slope integral ADC

Figure 7.55 is the waveform for the conversion of a dual-slope integral ADC. The working process is as follows: At first, the integrator is reset, and the input of the integrator is connected to the input voltage signal V_{in}. V_{in} integrates on the capacitor C during the time of t_1, which should correspond to the full-scale of the counter. After the integration, the input of the integrator is switched to the reference

voltage V_{ref}, whose polarity is opposite to V_{in}. The capacitor discharges during the time of t_2, and the counter is turned on to count the number of clock pulses. When the voltage of the capacitor reaches zero, the output of the comparator would switch and stop the counter, then the output of the counter is the digital value of V_{in}. The relationship between the input voltage and the reference voltage is given by

$$V_{\text{in}} = V_{\text{ref}} \times \frac{t_2}{t_1} \tag{7.28}$$

where t_2 is the time for the discharge of the capacitor C from the integration voltage to zero.

For the single-slope/dual-slope integral ADC circuit, the comparator is the core and used to compare the sampled input voltage with the reference ramp voltage. The performance of the comparator has a great influence on the accuracy of A/D conversion.

The gain of an ideal comparator is infinite, for a noninverting comparator, that is, when the noninverting input V_{P} is greater than the inverting input V_{N}, the output of comparator is high, when V_{P} is smaller than V_{N}, the output is low, while an inverting comparator is just the opposite. However, the gain A_{v} of a practical comparator is finite, so the voltage transmission characteristic for a practical noninverting comparator is shown in Fig. 7.56. The output of the comparator is given by

$$V_{\text{OUT}} = \begin{cases} V_{\text{OH}} & (V_{\text{P}} - V_{\text{N}}) > V_{\text{IH}} \\ A_V(V_{\text{P}} - V_{\text{N}}) & V_{\text{IL}} \le (V_{\text{P}} - V_{\text{N}}) \le V_{\text{IH}} \\ V_{\text{OL}} & (V_{\text{P}} - V_{\text{N}}) < V_{\text{IL}} \end{cases} \tag{7.29}$$

The performance of a practical comparator is mainly represented by the following parameters, which will directly affect the accuracy of A/D conversion

(a) Input offset voltage V_{OS}

In Fig. 7.56, when the $V_{\text{P}} - V_{\text{N}} = V_{\text{in}} = 0$, then theoretically $V_{\text{OUT}} = 0$. However, in the case when V_{OS} exists, then $V_{\text{OUT}} \neq 0$, and there is a constant

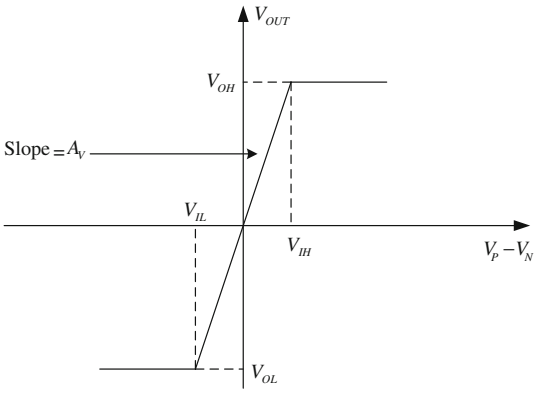


Fig. 7.56 The voltage transmission characteristic for a noninverting comparator

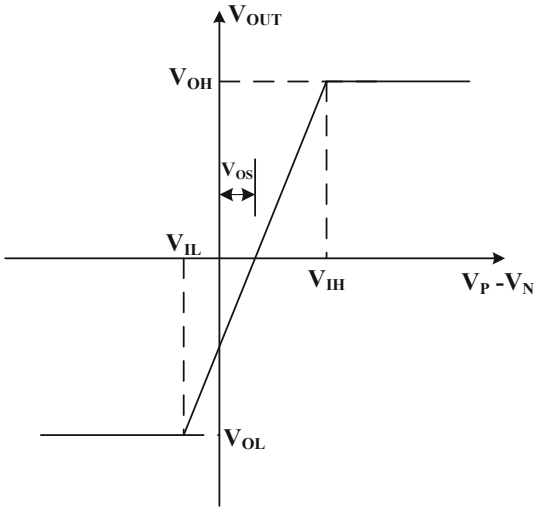


Fig. 7.57 The voltage transmission characteristic for a noninverting comparator with input offset voltage

"offset" for V_{OUT}. In fact, an input offset voltage V_{OS} is more useful, and defined as the input voltage when the output voltage is equal to zero, that is, $|V_{OS,in}| = |V_{OS,out}|/A_V$, as shown in Fig. 7.57.

The offset voltage is caused by two main reasons: one is from the mismatch in the current mirror, which is used as the load source of the comparator, due to the effect of channel length modulation. This mismatch can be eliminated through the optimization of the structure and parameters of the circuit. The other is from the mismatch on the device due to the CMOS process. The mismatch on the device can only be improved by using the advanced matching techniques on the layout implementation of the circuit. The typical value for V_{OS} due to the mismatch on device is about 5–15 mV [17].

The input offset voltage limits the minimum difference between the input voltage and the reference voltage for the comparator, thus reducing the accuracy of ADC.

(b) The resolution of a comparator

The resolution of a comparator is defined as

$$\Delta V = \frac{V_{OH} - V_{OL}}{A_V} = V_{IH} - V_{IL} \tag{7.30}$$

Assuming the input range of a comparator is V_{DR} and the accuracy of ADC is N-bit, then the conversion error of the comparator must be less than 1LSB (least significant bit), so the resolution of the comparator should reach at least $V_{DR}/2^N$, and the minimum gain A_V can also be derived from Eq. (7.30). Considering the effects of transmission time delay and input offset voltage, the practical gain of a comparator should be much higher.

(c) Propagation delay time T_P

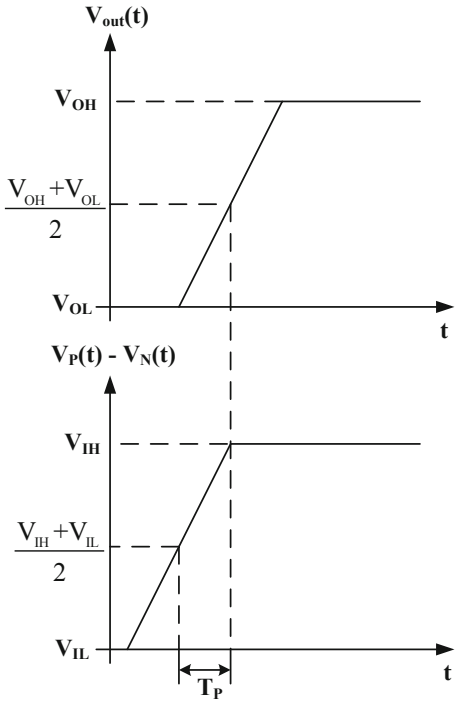


Fig. 7.58 The time response of a comparator

Table 7.6 Comparison of algorithms for the implementation of ADC on the focal plane

ADC algorithm	Quantization bits	Power	Area cost	Speed	Feasibility
Voltage-frequency-modulation	8–10	Low	Medium	Slow	Yes
Delta-sigma oversampling	18–20	Low	Medium	Medium	Yes
MCBS	10–14	Low	Medium	Medium	Yes
Successive approximation	10–20	Low	Large	Medium	Maybe
Single-slope/dual-slope integral ADC	8–14	Low	Small	Slow	Yes

Defined by the time delay between the input excitation and the output response, the parameter T_P represents how fast the output of a comparator could change when the input reaches the threshold value. As shown in Fig. 7.58, T_p is the time delay between the output voltage at 0.5 (V_{OH} + V_{OL}) and the input voltage at 0.5 (V_{IH} + V_{IL}).

The output transition time of a comparator would affect the width of the sequence of output pulses produced in subsequent circuit stages, so as T_P becomes smaller, the accuracy of A/D conversion becomes higher. For example, if the output transition time of a comparator for full-scale is 128 μs, to reach the precision of 8-bit, the error must be less than 1LSB, and then T_p must $\leq 128/2^8 \approx 0.5$ μs. Therefore, the output resistance and load capacitance of a comparator should be designed to be small enough to meet the requirement on T_P.

(5) **Summary of algorithms for the implementation of ADC on focal plane**

According to the above analysis on different algorithms for the implementation of ADC, their characteristics and the feasibility of the integration on the focal plane are summarized in Table 7.6.

7.4.3 Implementations for the ADC on Focal Plane

In practice, the preferred structure and algorithm for the implementation of ADC on focal plane is decided by the actual situation and comprehensive factors. For a cryogenic IRFPA, the best standard is how much the ADC circuit can reduce the size and power consumption of the whole chip.

The chip-level ADC structure is suitable for the applications for low frame rate and small resolution (less than 8-bit); The column-level ADC and pixel-level ADC structures are suitable for the applications for higher frame rate and bigger resolution. The pixel-level ADC structure combined with the algorithm of oversampling can achieve a high resolution of greater than 16-bits. With the continuous development of CMOS processing, the pixel-level ADC structure would take a very important position in the realization of a large-scale digital IRFPA with high resolution. The analog part of the pixel-level ADC structure is limited in the pixel, and the pixel output is the digital signal so that a high SNR can be obtained. For the pixel-level ADC structure, the conversion speed of the ADC could be very low, and the analog circuit would work in the sub-threshold region, thus reducing the power consumption of the circuit. Additionally, the oversampling technology can be used to reduce the noise, to improve the DR and increase the charge-handling capacity.

1. **A digital 128 × 2 dual-band ROIC based on the first-order Σ-Δ oversampling ADC**

Considering area cost and power consumption, the 128 × 2 digital ROIC adopts the circuit structure in which a first-order Σ-Δ oversampling ADC is shared by 8 pixels. There are two independent ADC circuits working in synchronization, one is for the A/D conversion of the voltage signal after the integration for source photocurrent, and the other is for the A/D conversion of the voltage signal after the integration of the sink photocurrent. So the digital 128 × 2 dual-band ROIC has 32 A/D converters in total.

Four pulse signals generated by the control circuit are output in parallel to the bus and each signal is corresponding to the control of 32 pixels. Then each pulse signal is branched out in four paths, and each signal path controls 8 pixels. All digital outputs of ADC circuits are read out in serial mode.

The block diagram of a digital 128 × 2 dual-band ROIC based on the first-order Σ-Δ oversampling ADC circuit is shown in Fig. 7.59. The circuit structure is composed of a dual-input stage circuit, the first-order Σ-Δ ADC circuit and the

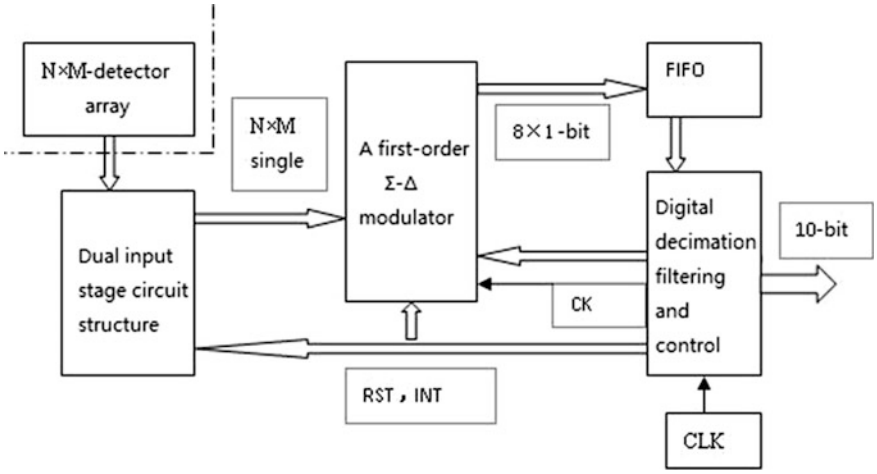


Fig. 7.59 The block diagram of a digital 128 × 2 dual-band ROIC based on the first-order Σ-Δ oversampling ADC circuit

Table 7.7 Performance comparison for different structures of input stage circuits

Injection structure	Dynamic range	NEI	Uniformity	Degree of linearity	Capacity	Maximum flux	Frequency response	Unit area
SI	Small	High	Bad	Bad	Low	Low	Medium	Small
SFD	Small	Medium	Medium	Medium	Medium	Low	Medium	Small
DI	Medium	Medium	Medium	Good	Low	Medium	Low	Small
CTIA	Big	Low	Good	Excellent	High	High	High	Big

control circuit, where the ADC circuit consists of the modulator and a digital decimation filter shared by the whole pixel array.

(1) Dual-input stage circuit

The structure of the input stage circuit of the ROIC for IRFPA has been established. The performance comparison for the most common families of input stage circuits is given in Table 7.7. Each structure has different advantages and disadvantages, thus being suitable for different types of detectors and applications.

By comparison, the DI circuit has advantages such as simple structure, small area cost, low power consumption, good linearity, and wide DR. The DI circuit has been widely used for MW and SW IRFPA.

Figure 7.60 shows a dual-band detector and the structure of a dual-DI input stage circuit, which consists of a sub-circuit for the integration of source photocurrent and a sub-circuit for the integration of sink photocurrent in parallel connection. The integration capacitor is implemented with a PIP structure.

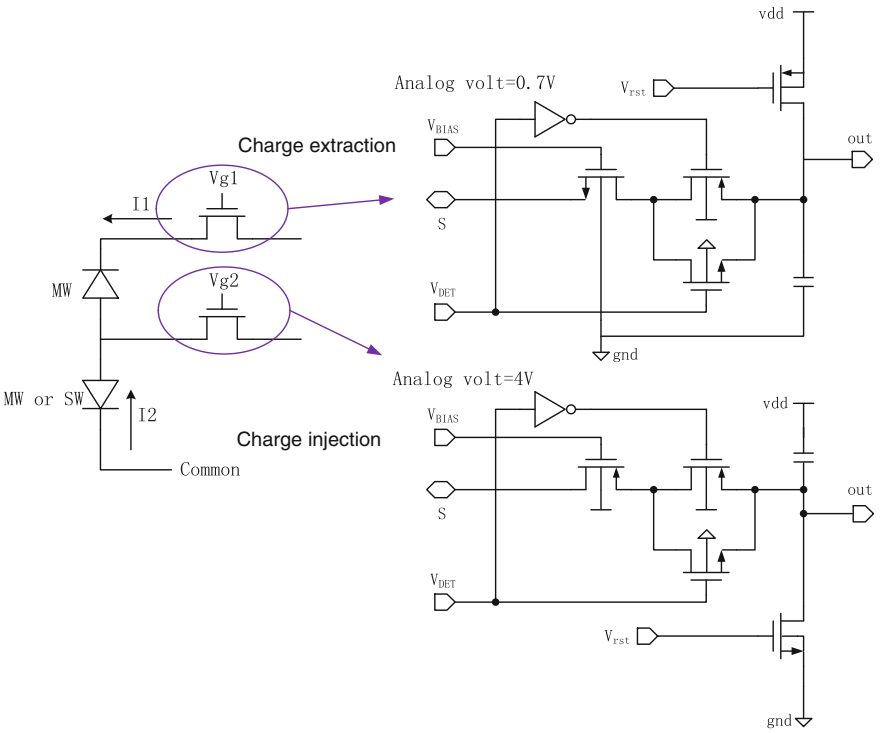


Fig. 7.60 The dual-band detector and the structure of a dual-DI input stage circuit

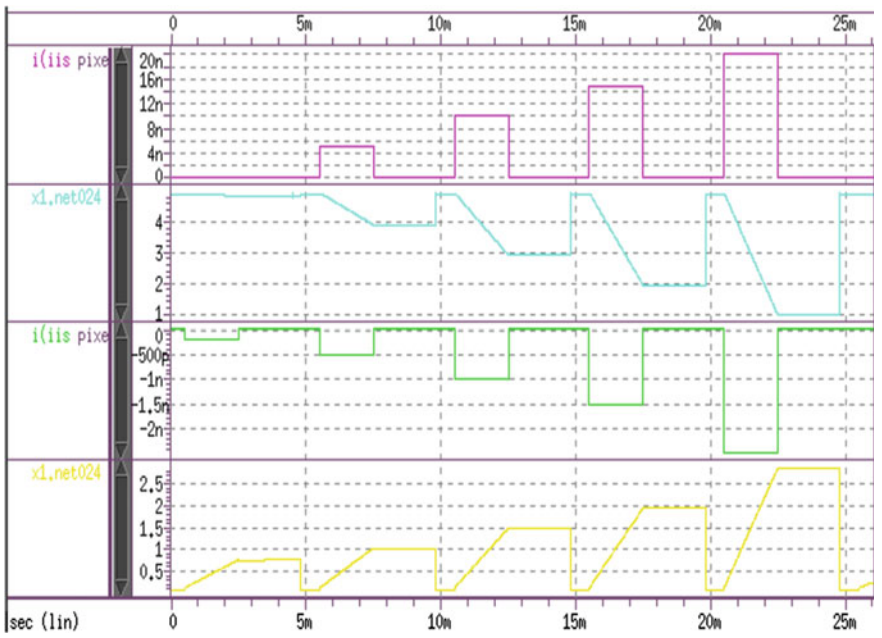


Fig. 7.61 Simulation waveforms for a dual-DI input stage circuit

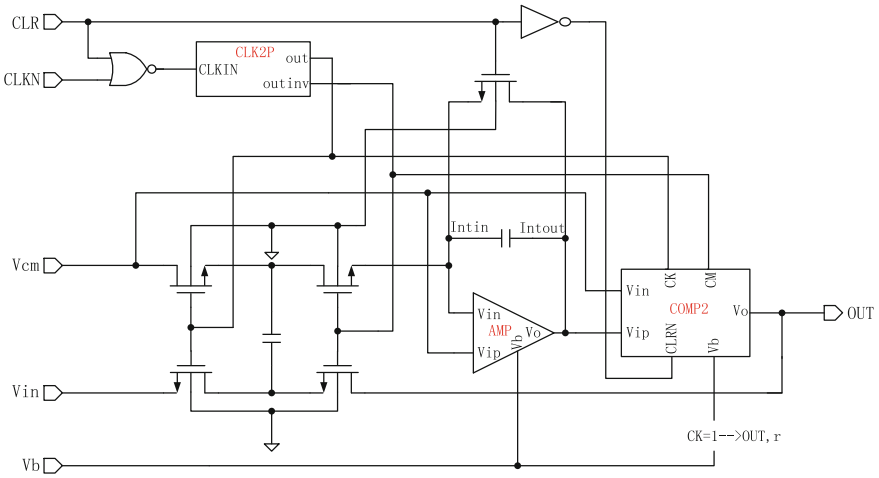


Fig. 7.62 The block diagram of the first-order sigma-delta modulator

The waveform for the simulation of the circuit is shown in Fig. 7.61, where the upper two graphs are for source photocurrent, the lower two graphs are for sink photocurrent. It can be seen that the voltage signal on the integration capacitor decreases or increases linearly with the photocurrent.

(2) The first-order Σ-Δ modulator

The first-order Σ-Δ modulator is adopted in view of the pixel area. Figure 7.62 shows a block diagram of the modulator, which is composed of a switched-capacitor integrator and a clocked comparator.

For a first-order Σ-Δ modulator, when the sampling frequency f_s is 2 MHz, then the oversampling ratio should reach 1024 times to achieve an accuracy of 10-bit for the ADC circuit. According to the design requirement, the gain of the operational amplifier in the modulator is required to be greater than the oversampling ratio, that is, more than 60 dB, and the smallest unity-gain bandwidth is half of the sampling frequency, that is, 1 MHz. Under the condition to meet the desired gain and bandwidth, the operational amplifier should be designed to realize low power consumption. A circuit structure of an op-amp is shown in Fig. 7.63. In the circuit, V_{in} and V_{ip} are the input, V_{out} is the output, V_{b1} and V_{b2} are two bias voltages. In simulation, V_{b1} and V_{b2} were set to 1.9 and 1.2 V, respectively. The simulation results for an op-amp are shown in Fig. 7.64, where the DC gain is 71.8 dB, the unity-gain bandwidth is 28 MHz, thus meeting the design requirements.

Through the calculation, when the output state of the comparator switches, the magnitude for the change is 1 V at least. Therefore, if the hysteresis voltage is less than 1 V, and the gain of the comparator is large enough, then the output of the comparator can achieve a full-scale transition with varied amplitude of the input

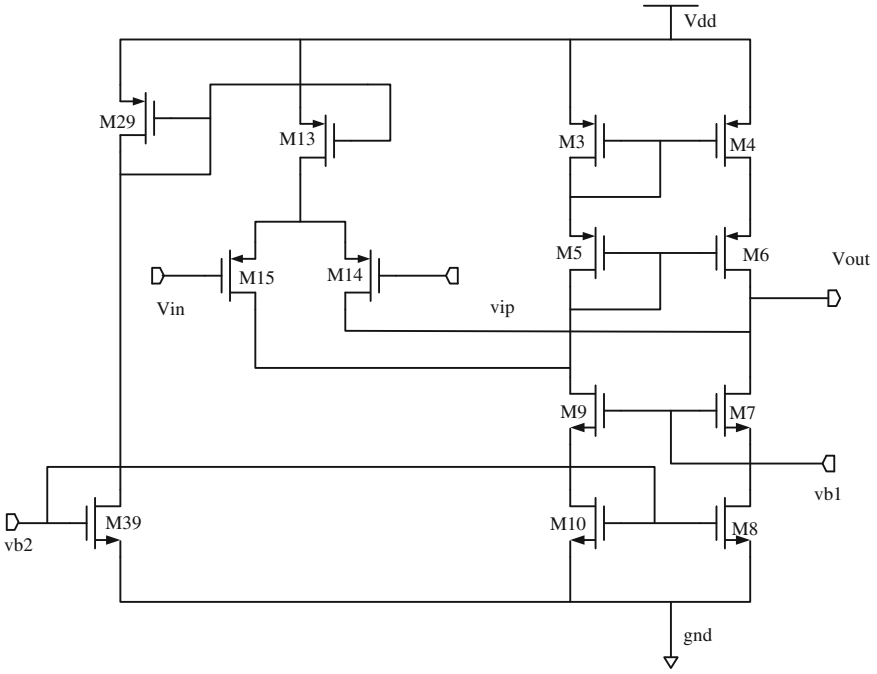


Fig. 7.63 The architecture of an op-amp in the modulator

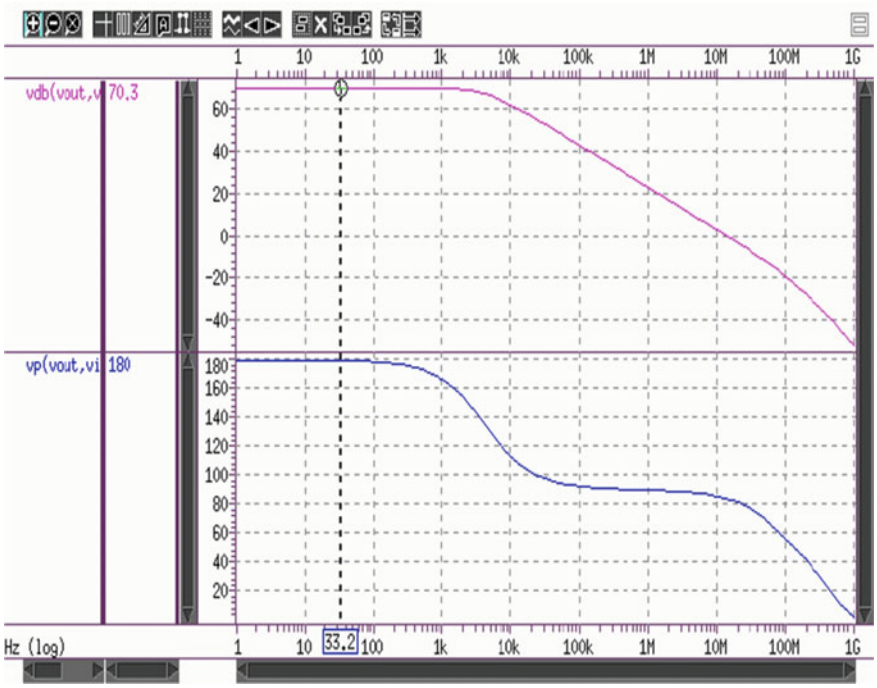


Fig. 7.64 Gain and phase versus frequency for an op-amp in the modulator

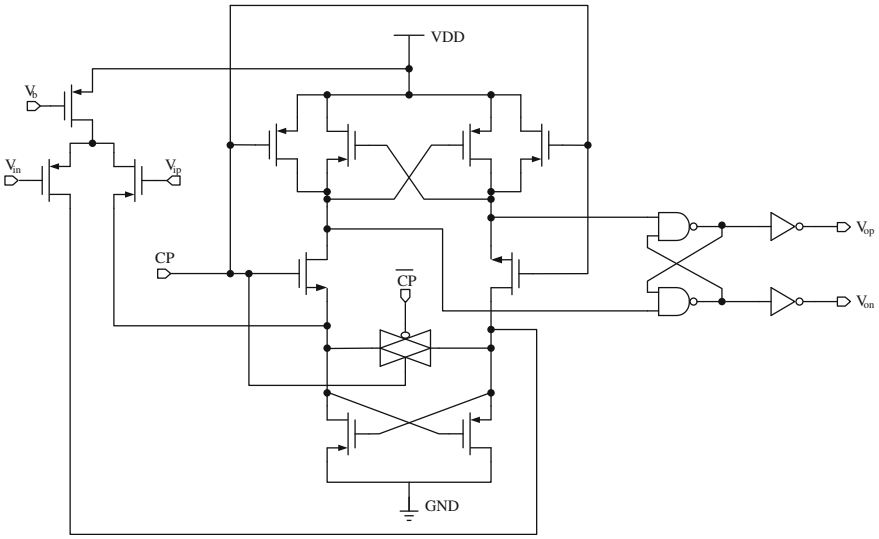


Fig. 7.65 The circuit structure of a clocked comparator

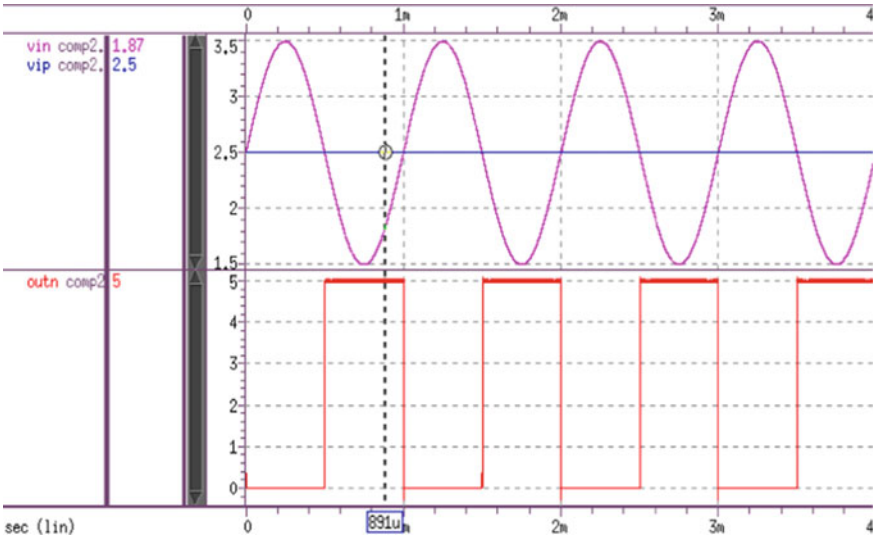


Fig. 7.66 Simulation waveforms of the comparator

signal, thus the comparator working properly. The circuit structure of a comparator is shown in Fig. 7.65, and the simulation waveforms are shown in Fig. 7.66.

The simulation waveform for the first-order Σ-Δ modulator is shown in Fig. 7.67. The simulation shows that the input analog voltage is modulated to get a 1-bit digital pulse sequence at high frequency.

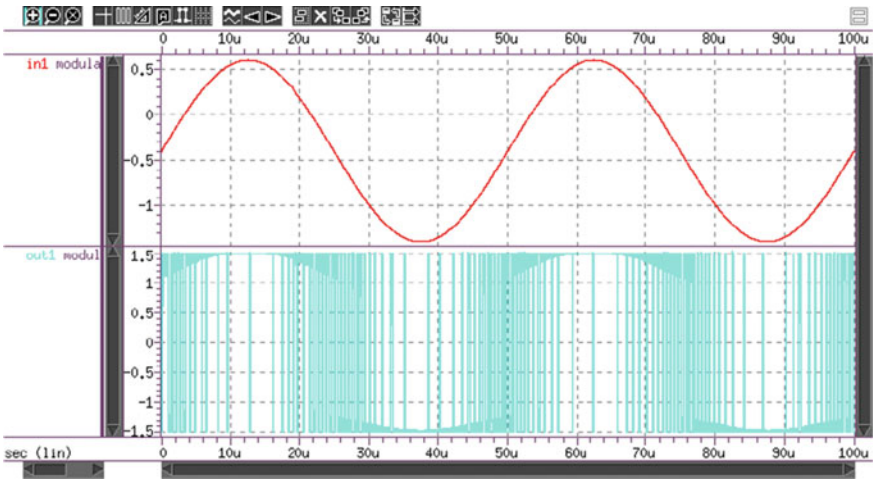


Fig. 7.67 The first-order Σ-Δ modulator simulation waveforms

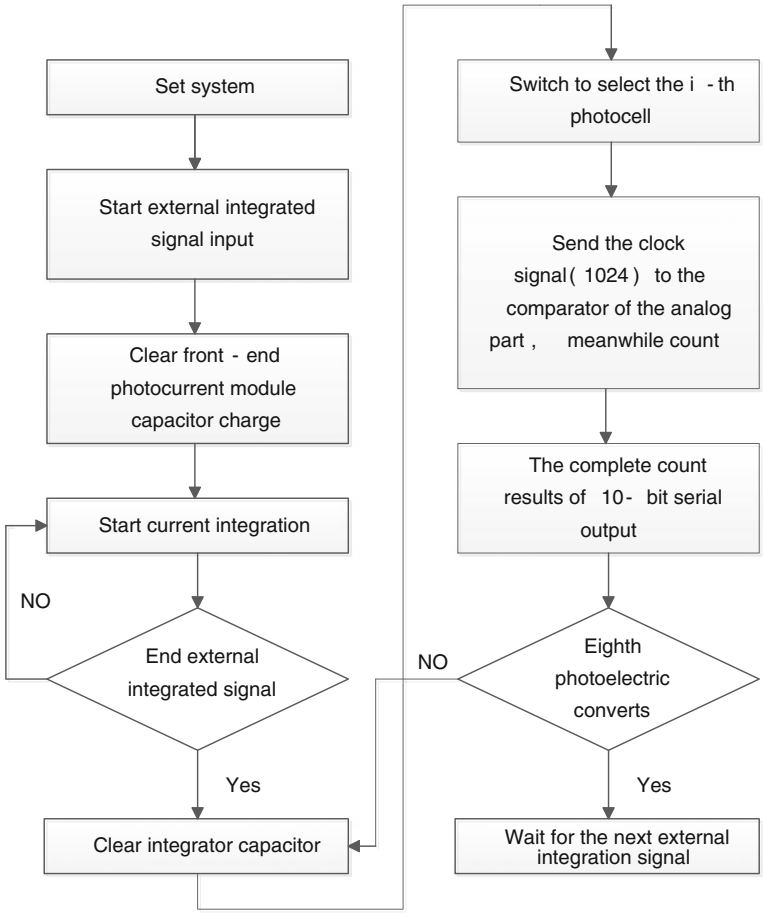


Fig. 7.68 The workflow of a decimation filter and control circuit

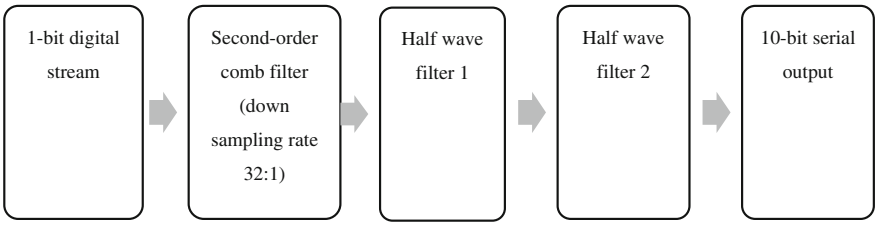


Fig. 7.69 The conventional structure of a digital comb filter

(3) The digital decimation filter and control circuit

The digital circuits of ROIC are designed to serve two modules: one is for the A/D conversion and control of the source photocurrent, and the other is for the A/D conversion and control of the sink photocurrent. Immediately after the conversion of one cell, the digital data is read out.

The workflow of a digital decimation filter and control circuit is shown in Fig. 7.68. The conventional digital decimation filter may use the structure shown in Fig. 7.69, and the increase in resolution, N_{inc}, through the oversampling is given by

$$N_{\text{inc}} = \frac{30 \log K - 5.17}{6.02} \tag{7.31}$$

From Eq. 7.31, a 128 times oversampling is required to get an accuracy of 10-bit. Because of the working characteristic of the input stage circuit and the modulator, a digital decimation filter needs to establish an initial state for each conversion, so the period of conversion actually costs about 1000 clock cycles instead of 128.

Since the voltage signal sampled from the integration capacitor of the DI input stage circuit is a dc signal, a 10-bit counter can be used to substitute the traditional structure of a comb filter, thus not only realizing the function of a low-pass decimation filter, but also greatly simplifying the structure of digital circuits.

Through the simulation, it is verified that the power consumption is lowered by one-third and the area cost is reduced two-thirds by using the counter structure. The simulation results for the counter structure and the traditional comb filter are shown in Fig. 7.70. The first graph of (a) and the first graph of (b) in Fig. 7.70 show that the filtering on high frequency for a comb filter is better than that for a counter structure, which is still higher than 100 dB and, therefore, meets the specification. The second graph of (a) and the second graph of (b) in Fig. 7.70 show that the noise suppress for the counter structure is two orders of magnitude higher than that for the comb filter.

(4) The simulation results for the whole circuit

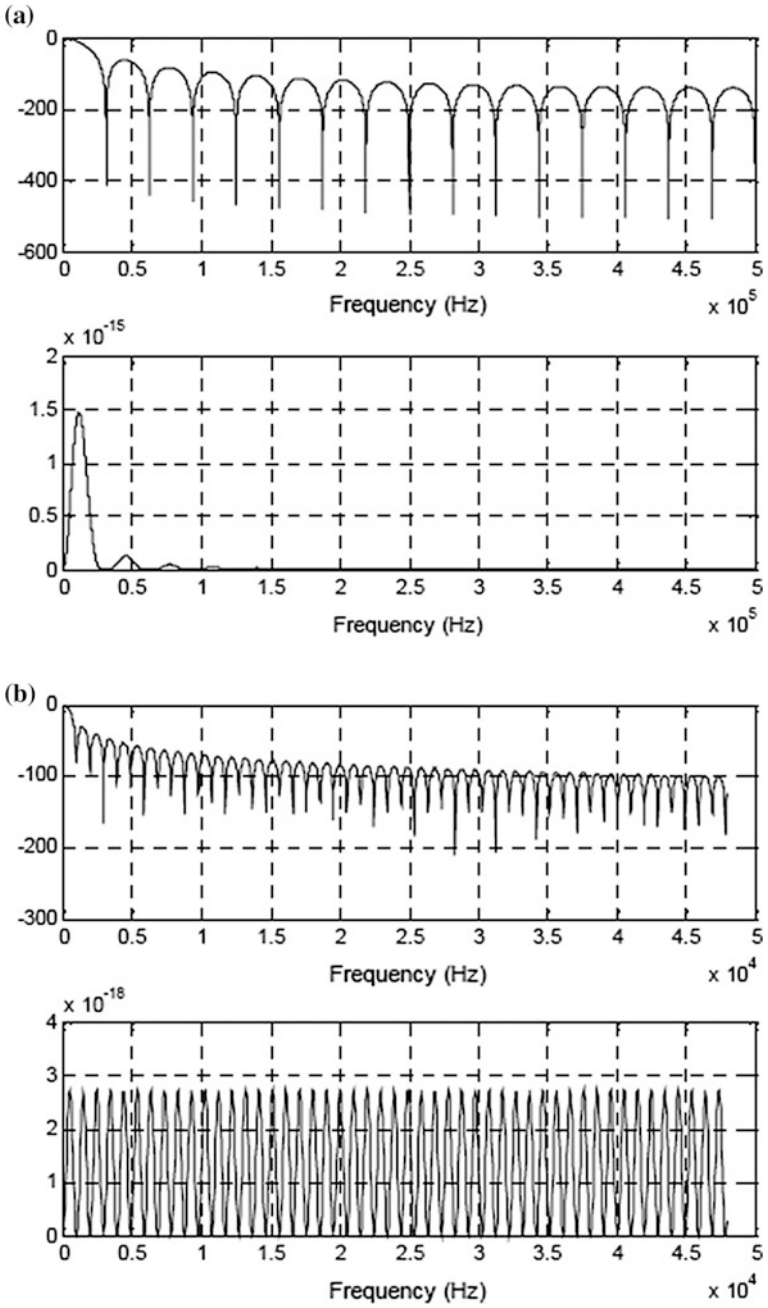


Fig. 7.70 The simulation results for the counter structure and the traditional comb filter. **a** Conventional second-order CIC filter. **b** The proposed structure

Table 7.8 The statistics of the simulation results for an 8 × 2 digital ROIC module

Input current (nA)			Pre-simulation results		Post-simulation results	
			Modulator input voltage (v)	Output 1: the number of zeros (take 50 clk cycles)	Modulator input voltage (v)	Output 1: the number of zeros (take 50 clk cycles)
Source current	*I*1	1	3.17	32:18	3.16	32:18
	*I*2	2	2.89	29:21	2.88	29:21
	*I*3	4	2.3	23:27	2.28	23:27
	*I*4	7	1.43	14:36	1.4	14:36
	*I*5	9	0.873	9:41	0.825	9:41
	*I*6	12	0.0897	1:49	0.0487	1:49
	*I*7	15	0	0:50	0	0:50
	*I*8	18	0	0:50	0	0:50
Sink current	*I*9	0.1	1.88	19:31	1.9	19:31
	*I*10	0.2	1.92	20:30	1.95	19:31
	*I*11	0.4	2.01	20:30	2.04	20:30
	*I*12	0.7	2.12	21:29	2.15	21:29
	*I*13	0.9	2.19	22:28	2.23	22:28
	*I*14	1.2	2.29	23:27	2.34	23:27
	*I*15	1.5	2.39	24:26	2.44	24:26
	*I*16	1.8	2.5	25:25	2.54	25:25

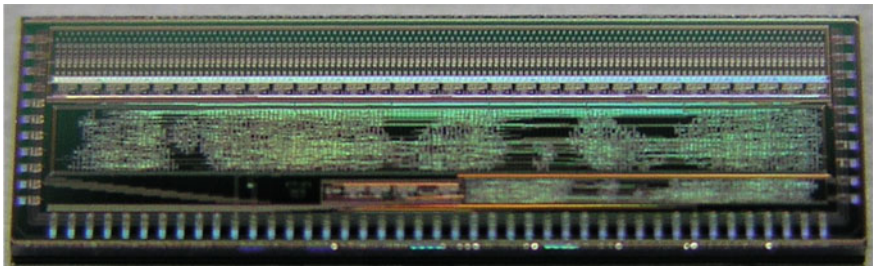


Fig. 7.71 The photograph of the chip of a digital 128 × 2 dual-band ROIC based on the first-order Σ-Δ oversampling ADC

For the 128 × 2 digital ROIC, the basic circuit module of the chip is a first-order Σ-Δ oversampling ADC circuit shared by 8-cell circuits. Therefore, the simulation results for an 8 × 2 array are used for statistical decision. From the statistics shown in Table 7.8, it can be seen that the function of the modulator is realized and the performance of the circuit meets the requirement on the accuracy.

The digital 128 × 2 dual-band ROIC based on the first-order Σ-Δ oversampling ADC is implemented and fabricated in the CSMC Foundry. The chip size is 12,906 μm × 3335 μm. Figure 7.71 is the photograph for the chip.

2. **A 8 × 8 digital ROIC based on the Pixel-level structure with the single-slope integral ADC**

For the design of a two-dimensional digital ROIC, the area cost for a unit cell readout circuit is strictly limited because the pixel size of detector array is small and fixed. Therefore, the area cost has the top priority among the design constraints, and a trade-off between the area cost and performance is made.

(1) The frame structure of the 8 × 8 digital ROIC

As described earlier, the single-slope integral ADC is composed of an integrator, a comparator, an output counter, and a parallel-to-serial converter. For a pixel-level ADC structure, there is a problem of limited pixel size. However, the output counter and parallel-to-serial converter could be moved out of the pixel array, and shared by eight unit cells of each column, thus not only ensuring the digital output from the pixel and a high SNR, but also reducing the area cost for the cell circuit.

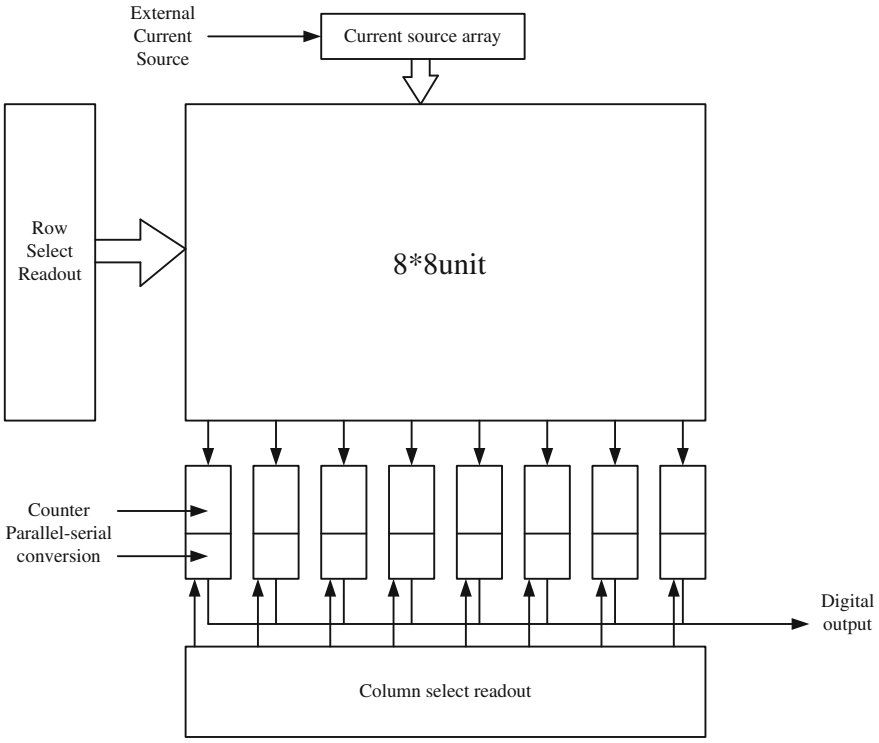


Fig. 7.72 The block diagram of a 8 × 8 digital ROIC based on the single-slope integral ADC

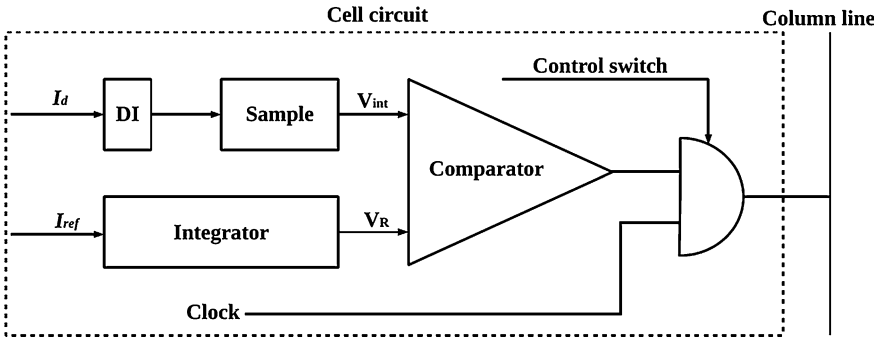


Fig. 7.73 The circuit structure of a unit cell

The block diagram of the whole circuit is shown in Fig. 7.72. The reference current source of each unit cell is provided through a current source array. The working process is as follows: a row of unit cell circuits is selected by a row selection signal and starts the integration, the output of each unit cell circuit is a sequence of pulses, which are buffered to the corresponding 8-bit counter and parallel-to-serial converter, thus generating a serial digital word; Controlled by the column selection signal, eight serial digital words are read out sequentially. Then all counters and parallel-to-serial converter are reset, and a new row selection signal starts to select the next row of unit cell circuits. And so on, a frame of signal data is not ended until each row of unit cells is read out.

(2) The circuit structure of a unit cell

Through a cell circuit, the photocurrent from the detector is integrated and converted to a sequence of digital pulses, whose number is proportional to the current intensity, that is, the A/D conversion is realized in the unit cell. Digital circuits out of the unit cell are used to count the number of digital pulses and output in serial. Figure 7.73 is the circuit structure of a unit cell [18]. The DI structure is adopted for the input stage circuit. Since the integrated voltage signal is a dc signal, the sample capacitor could be eliminated in the design, and the voltage signal on the integration capacitor could be directly converted to a digital signal.

One input of the comparator is connected with the input transistor and the integration capacitor. After a period of integration time, the voltage on the integration capacitor is V_{int}. The other input of the comparator is a ramp voltage V_{R} with good linearity, which is generated by a reference current source and an integrator. When the comparator output is high, the AND gate opens, and outputs the clock pulse. When $V_{\text{R}} < V_{\text{int}}$, the output of the comparator flips, and blocks the AND gate, thus the number of output clock pulses is proportional to the current intensity, and is waiting for the row selection signal to be read out to the column line.

(3) Design of the comparator

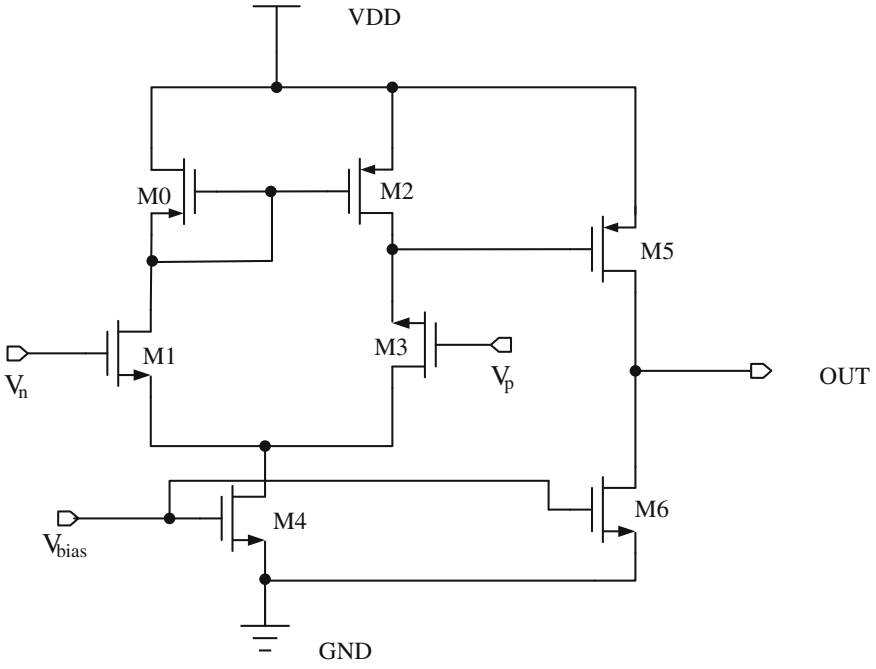


Fig. 7.74 Two-stage operational amplifier configuration diagram

Using a common differential comparator, the transition voltage can be precisely controlled; whereas the gain is very low, and the swing for the output voltage is too small to meet the requirement on resolution. In order to increase the gain, a two-stage op-amp is adopted as a comparator. Figure 7.74 is the circuit topology of a two-stage op-amp. Where the first stage is the input differential pair, reducing the offset and noise, providing high gain, and the second stage is a common-source stage, providing a large swing, thus reducing the output resistance.

In Fig. 7.74, we need to determine the size of the transistors $M1$–$M7$ and assume the current through them respectively to be $I1$–$I7$. First, the size of each transistor is set through hand calculation with approximate design equations, and then fine tuned through simulation. The steps are detailed as follows:

Known parameters for the circuit design are the DC gain (A_v), the input common-mode range [V_{in} (min) and V_{in}(max)], load capacitance (C_L), conversion rate (dV/dT), output voltage swing [V_{out} (min), and V_{out} (max)] and power (P_{diss}).

- Determine the equilibrium constraints (the relationship of current mirror)

The so-called circuit static equilibrium, meaning that when all devices are operated in the saturation region, all n-channel devices can absorb the same amount

of current come from their paired p-channel devices. In this case $M1$ and $M2$, $M3$ and $M4$ all are matched, and VGS3 = VGS4 = VGS6. If the inputs are in balance, then when the V_P and V_N are equal, the following equations hold

$$\frac{W_1}{L_1} = \frac{W_2}{L_2} \tag{7.32}$$

$$\frac{W_3}{L_3} = \frac{W_4}{L_4} \tag{7.33}$$

$$I_1 = I_2 = 0.5I_5 \tag{7.34}$$

$$\frac{W_6/L_6}{W_4/L_4} = 2 \cdot \frac{W_7/L_7}{W_5/L_5} \tag{7.35}$$

- Determine the output current $I7$ to meet the requirement of the slew rate

When no specific conversion rate is given, one can choose 10 times the propagation delay time as a starting point. In the circuit design, the propagation delay time T_P = 0.5 μs, power supply voltage V_{DD} = 5 V, and the conversion rate is $dV/dT \geq 25$ V/μs. Then the current $I7$ is given by

$$I_7 = C_L\left(\frac{dV}{dT}\right) \tag{7.36}$$

- Determine the minimum size of $M6$ and $M7$ to satisfy the output voltage swing. Using the following equation

$$V_{DS} = \left(\frac{2I_7}{\beta_{6,7}}\right)^{1/2} \tag{7.37}$$

- Knowing the current of the second stage and the minimum size of $M6$, the gain of the second stage is given by

$$A_{V2} = -\frac{g_{m6}}{g_{ds6} + g_{ds7}} \tag{7.38}$$

- According to specifications and the gain obtained above, the gain of the first stage is given by

$$A_{V1} = A_V / A_{V2} \tag{7.39}$$

- According to the relationship of current mirror and the minimum size of $M6$ and $M7$ determined earlier, the current of the first stage is determined. Then verify whether the requirement of power P_{diss} is satisfied.
- Knowing the current and gain of the first stage, the size of the transistor $M1$ is given by

$$A_{V1} = -\frac{g_{m1}}{g_{ds1} + g_{ds3}} \tag{7.40}$$

where

$$g_{m1} = \left[2K'\left(\frac{W}{L}\right)I_1\right]^{1/2} \tag{7.41}$$

$$g_{ds} \approx \lambda I_1 \tag{7.42}$$

- Determine the minimum size of $M5$ to meet the requirement of the common-mode voltage range

$$V_{G1}(\min) = V_{SS} + V_{DS5} + \left(\frac{I_5}{\beta_1}\right)^{1/2} + V_{T1}(\max) \tag{7.43}$$

where

$$V_{DS5} = \left(\frac{2I_5}{\beta_5}\right)^{1/2} \tag{7.44}$$

- Increase the size of $M5$ or $M7$ to satisfy the relationship of current mirror.
- Determine the minimum size of $M4$ to meet the requirement of the common-mode voltage range

$$V_{G1}(\max) = V_{DD} - \left(\frac{I_5}{\beta_3}\right)^{1/2} - |V_{T03}(\max)| + V_{T1} \tag{7.45}$$

- Increase the size of $M3$ or $M6$ to satisfy the relationship of current mirror.
- Using the Spectre simulation tool of Cadence for the circuit simulation, verify whether all specifications are satisfied and whether the propagation delay time meets the requirement.

In above Equations, K is the transconductance of the transistor ($\approx \mu C_{OX}$), β is KW/L, λ is the parameter of channel length modulation.

Meanwhile, design parameters are determined to satisfy design specifications. It is required that the output low voltage V_{min} be less than the logic low level and that

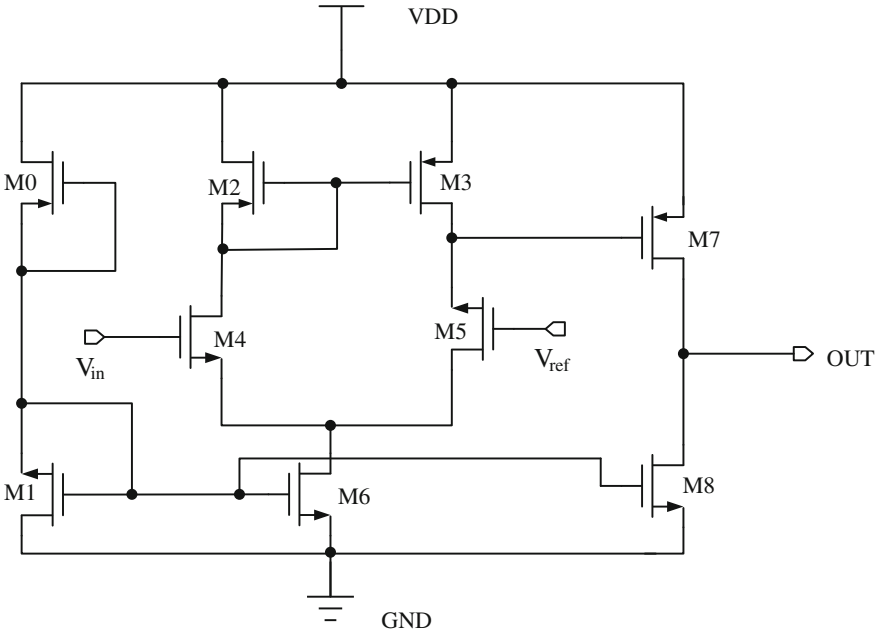


Fig. 7.75 The circuit diagram of the comparator with determined transistor size

the output high voltage V_{max} be more than the logic high level because the output voltage of the comparator is required to drive the subsequent digital circuits. Therefore, the swing range of V_{OUT} is about 0.3–4 V. It is required that the average power consumption for a unit cell must be less than 80 μW, and the power consumption mainly comes from the amplifier, while the amplifier operates no more than 128 μs within a frame period (about 2 ms), so the operating power P_{diss} should be less than $80 \times (2000/128) \approx 1000$ μW, that is, 1.25 mW. Other determined parameters include: C_L = 1pF, V_{DD} = 5 V, V_{SS} = 0 V, the swing of V_{OUT} is 0.5–4.5 V, the common-mode input range is 2.5–5 V, the conversion rate is 25 V/μs,

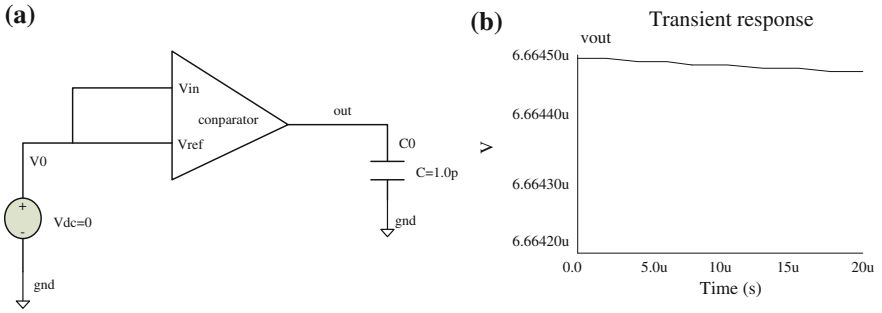


Fig. 7.76 The simulation for offset voltage. **a** Circuit schematic for the simulation of offset voltage. **b** Simulation result

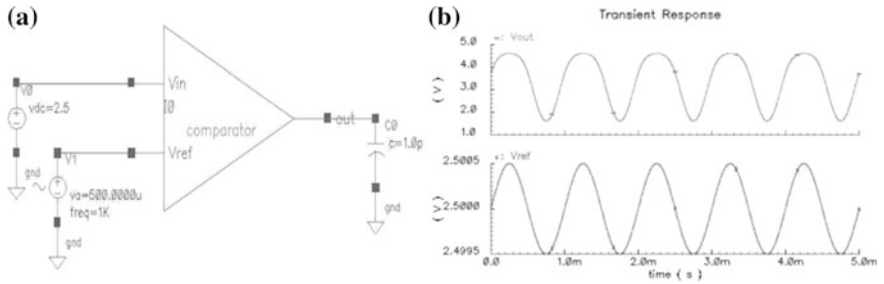


Fig. 7.77 Comparator gain simulation circuit schematics.

and the gain $A_V \geq 1000$. These parameters are substituted into the design procedure described above, and the width and length of each transistor can be determined finally. Figure 7.75 is the circuit diagram of the comparator with determined transistor size [19].

(4) Simulation for the circuit

After all transistors sizes are determined, the Software Spectre is used to simulate four important performance parameters mentioned earlier to determine whether they meet the design specifications.

The simulation for offset voltage is shown in Fig. 7.76, when two inputs of the comparator are zero, the output offset voltage is 6.66449 μV, divided by the gain, the derived input offset voltage V_{OS} is less than 6 nV, therefore the impact on the performance of a comparator can be ignored. Here we only simulate the offset voltage of the system, and the offset voltage caused by the fabrication process is determined through the test after chip tape out.

Figure 7.77 is the simulation for the gain of the comparator when the differential input is a sine wave signal whose frequency is 1 kHz and amplitude is 1 mV and $\Delta V_{in} = 1$ mV. Simulation results show that $\Delta V_{out} = 4.607\ \text{V} - 1.624\ \text{V} = 2.983\ \text{V}$, the gain $A_V = 2.983/0.001 \approx 2983$. So the DC gain of comparator is greater than or equal to 3000, which meets the design requirement and also has a great margin.

The simulation for propagation delay time of the comparator is shown in Fig. 7.78. Adding a DC voltage of 4.029 V at the input V_{int} of the comparator, and a linear ramp voltage at the input V_{ref}, and observing the output waveform and propagation delay time of the comparator, it can be seen that the comparator changed the output state when V_{ref} was reduced to near V_{int}. In order to measure the propagation delay time within the area of the threshold time, the waveform in Fig. 7.78 is zoomed in and is shown in Fig. 7.79. At the point B, V_{ref} is 4.02869 V, reaching the threshold value, $t_B = 206.381\ \mu\text{s} = t_{IL} \approx t_{IH}$, near the point A, $V_{out} = 2.5\ \text{V} = 0.5\ (V_{OH} + V_{OL})$, $t_A = 206.842\ \mu\text{s} = 0.5\ (t_{OL} + t_{OH})$, therefore the propagation delay time $T_P \approx t_A - t_B = 206.842 - 206.381 = 0.461\ \mu\text{s} < 0.5\ \mu\text{s}$, so it meets the design requirement.

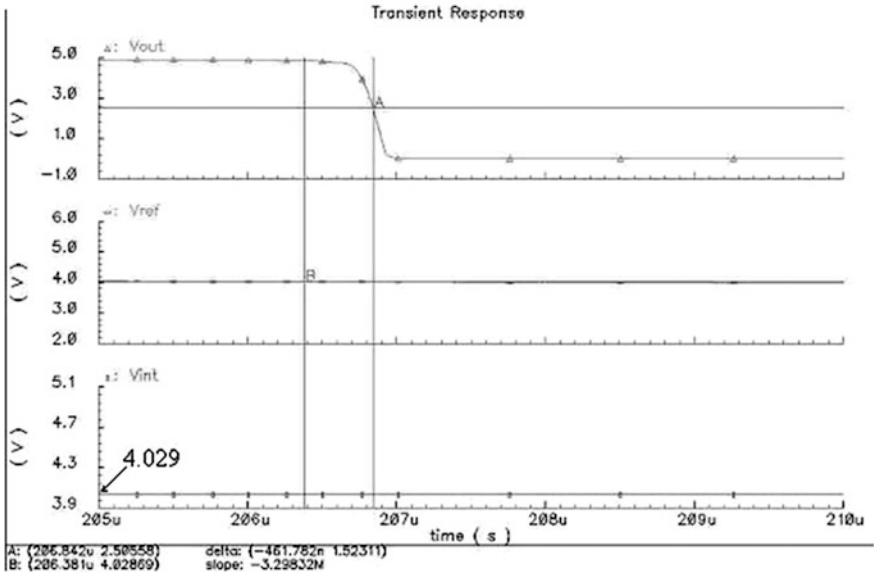


Fig. 7.78 The simulation for the propagation delay time of comparator

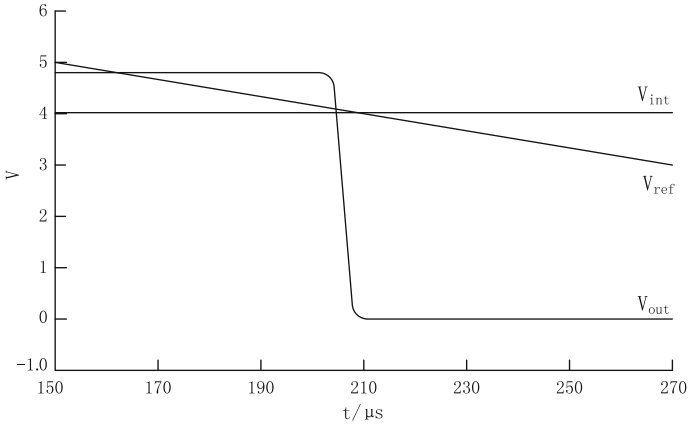


Fig. 7.79 The zoom-in waveform for Fig. 7.78

Figure 7.80 shows the operating waveform of the comparator when the input V_{int} is connected with a DC voltage and the input V_{ref} is connected with a linear ramp voltage. The comparator can operate normally when V_{int} is 2.5 and 4.981 V (close to 5 V), so the input common-mode voltage range is from 2.5 to 5 V.

Based on the above simulation results, the error of A/D conversion due to the performance of the comparator can be calculated. The offset error of the comparator is very small, so the error of the A/D conversion due to is negligible. The resolution

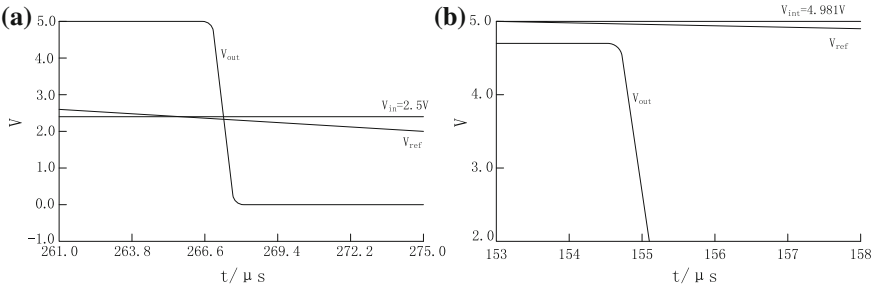


Fig. 7.80 The simulation for the input common-mode voltage range of comparator. **a** Operating waveform of comparator when the input is close to 2.5 V. **b** Operating waveforms of comparator when the input is close to 5 V

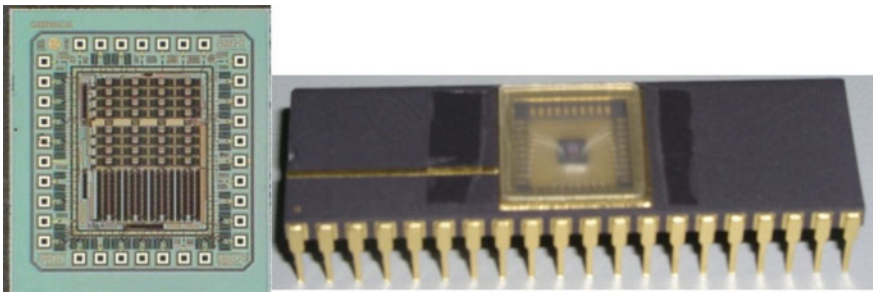


Fig. 7.81 Photographs for the 8 × 8 digital ROIC based on the pixel-level structure with the single-slope integral ADC and a chip with DIP package

Table 7.9 The measurement for the 8 × 8 digital ROIC based on the pixel-level structure with the single-slope integral ADC

1	Format	8 × 8
2	The pixel pitch	100
3	Resolution	8-bit
4	Power of unit ADC	37.5 μW
5	Offset error (E_o)	1LSB
6	Full-scale range (FSR)	252 (decimal)
7	Linearity error (E_T)	2.5 LSB
8	Gain error (E_G)	2.5 LSB
9	Conversion time of ADC (t_c)	144 μs
10	Power consumption of chip	$\leq$8.7 mW
11	Nonuniformity between pixels	≈2.54 %

of the comparator is higher than 5/3000 ≈ 0.00167 V = 1.67 mV, so then the error of A/D from it is less than 1.67/(2500/256) = 0.17 LSB. Propagation delay time is 0.46 μs, so the error of A/D from it is 0.46/(128/256) = 0.92 LSB. Therefore, the

maximum error of A/D conversion caused by the comparator is 0.17 LSB + 0.92 LSB = 1.09 LSB.

(5) The measurement for the circuit chip

The 8 × 8 digital ROIC based on the pixel-level structure with the single-slope integral ADC is implemented with the 0.6 μm Double-Well Double-metal (DPDM) CMOS process by the CSMC Foundry. The chip size is about 2.1 mm × 2.3 mm. Figure 7.81 is the photograph for the circuit chip and a chip with DIP package. The measurement for the whole chip and a unit cell circuit verifies that the input photocurrent can be read out and the A/D conversion can operate properly. The performance parameters of the measurement are shown in Table 7.9.

3. **The 8 × 8 digital ROIC based on the Pixel-level structure with first-order Σ-Δ oversampling ADC**

The first-order Σ-Δ oversampling ADC circuit has features such as high input charge-handling capacity, high resolution, and high SNR. The drawback is to

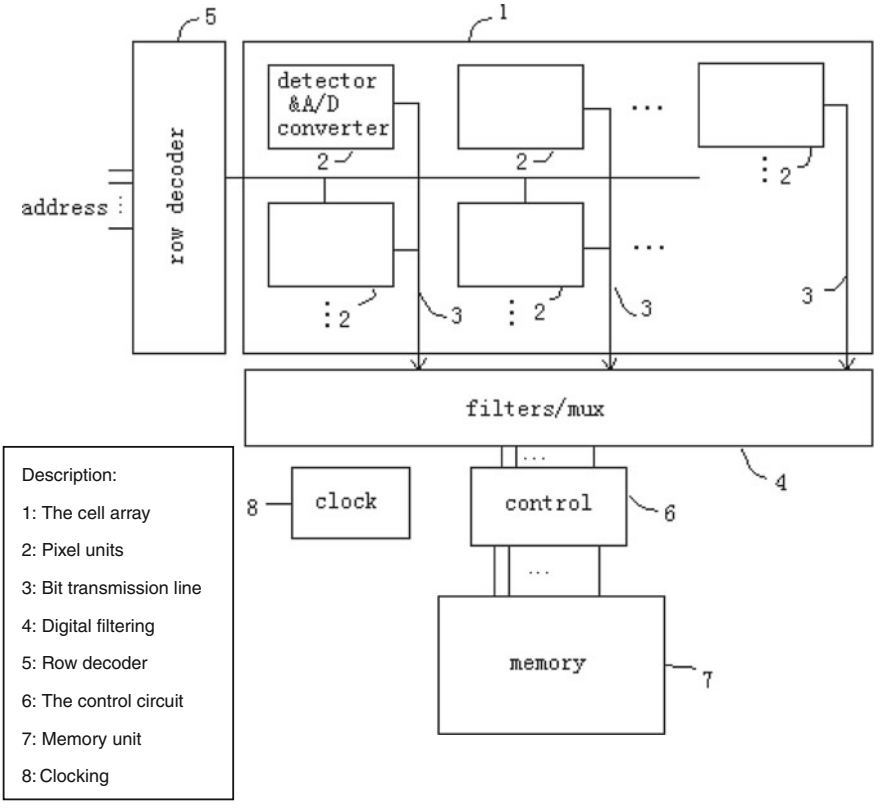


Fig. 7.82 The block diagram of an 8 × 8 digital ROIC based on the pixel-level structure with first-order Σ-Δ oversampling ADC

realize a digital filter outside the pixel array, which increased the circuit complexity, the input bandwidth, and the output bandwidth. With the continuous development of CMOS technology, the digital filter can be implemented not only off-chip, but also on-chip.

(1) The frame structure of the circuit

Figure 7.82 shows the overall structure of an 8 × 8 digital ROIC based on the pixel-level structure with a first-order Σ-Δ oversampling ADC. Each pixel is connected to an A/D converter, which outputs a digital bit-stream representing the analog input of each pixel. Each unit cell is independent. The digital filter is used for receiving the digital bit-stream and converting it to an *N*-bit digital word. The digital filter can also be implemented off-chip with hardware or software, and the digital bit-stream is transmitted to the receiving end of the filter for digital filtering through the signal line.

(2) The unit cell circuit structure of the 8 × 8 digital ROIC

The 8 × 8 digital ROIC is based on the pixel-level ADC structure. The algorithm of the first-order Σ-Δ oversampling ADC is adopted. The unit cell circuit structure is shown in Fig. 7.83.

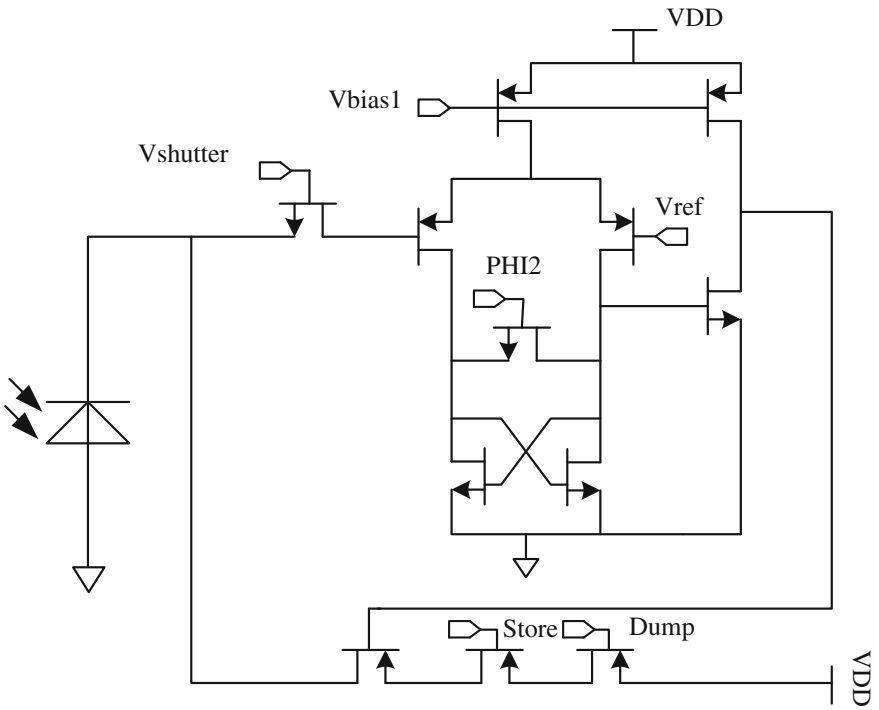


Fig. 7.83 The unit cell circuit structure based on the first-order Σ-Δ oversampling ADC

The photodiode is gated by a switch transistor, the photo charge integrates on the integration capacitor, and the integrated voltage signal is quantized. The comparator operates in the sub-threshold region, thus lowering the power consumption and noise, improving the gain, and reducing the leakage current of DAC. The quantized value is converted to the electric charge, which is fedback to the photodiode, by 1-bit DAC. The 1-bit DAC is realized with an analog shift register similar to the three-phase CCD. If the output of comparator is low, the sequence consisting of three control pulses will fill the integration capacitor with a fixed amount of charge.

$$Q_{\mathrm{d}} = \left(V_{\mathrm{dd}} - V_{\mathrm{store}} - V_{\mathrm{tp}}\right) \cdot C_{\mathrm{ox}} \cdot W \cdot L \tag{7.46}$$

In Eq. 7.46 V_{dd} is the supply voltage, V_{store} is the low voltage of the Store signal, C_{ox} is the gate oxide capacitance of the MOS transistor, and W and L are the width and length of the MOS transistor. The voltage amplitude of the Store signal can be used to control the amount of charge for transport, and one bit is generated and read out by the bit line in each clock cycle.

The ROIC works exactly in the same way as read-only memory, and adopts the structure composed of bit location decoding, column location decoding (multiplex for output), and the amplifier. Each pixel generates one bit for each access time, so

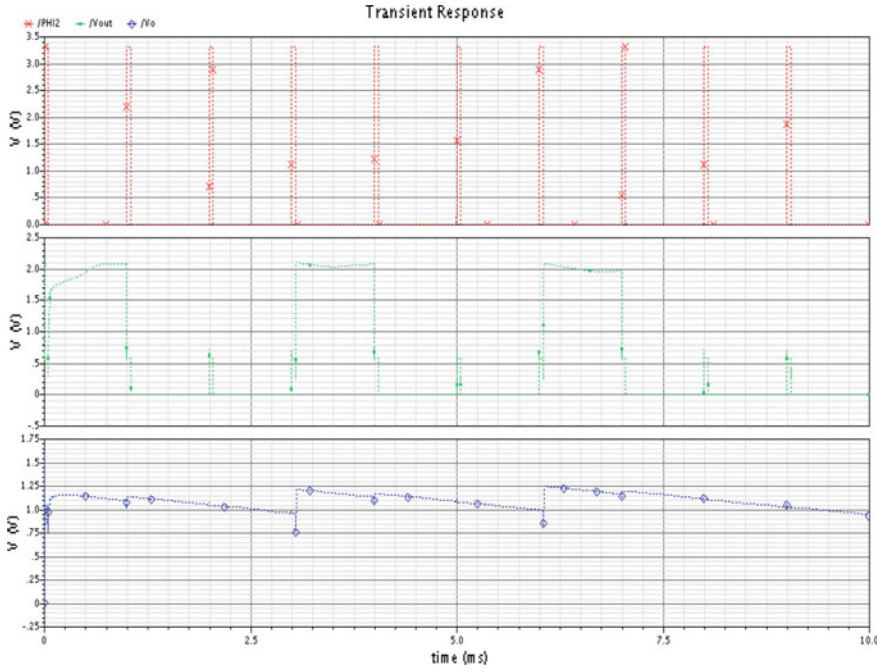


Fig. 7.84 The simulated output of a unit cell circuit when the photocurrent is 1/4 of the maximum

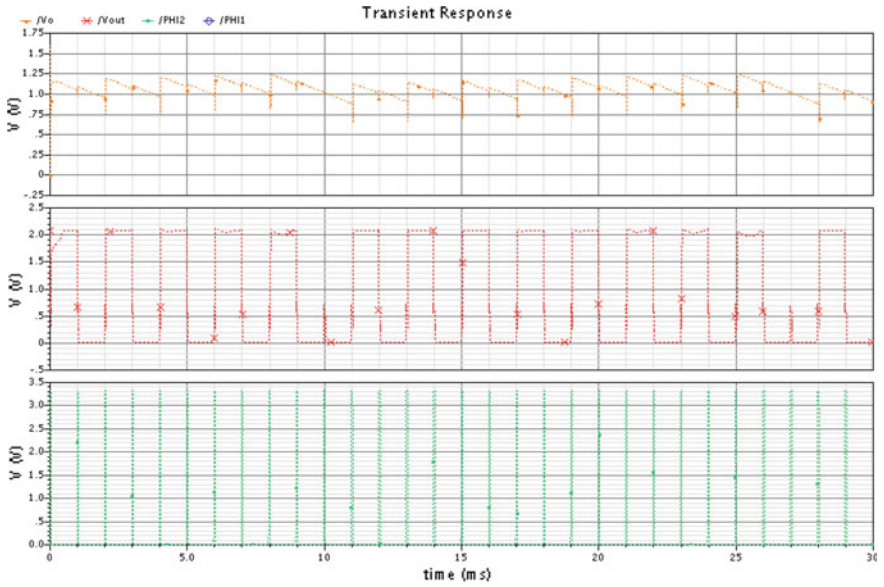


Fig. 7.85 The simulated output of a unit cell circuit when the photocurrent is half of the maximum

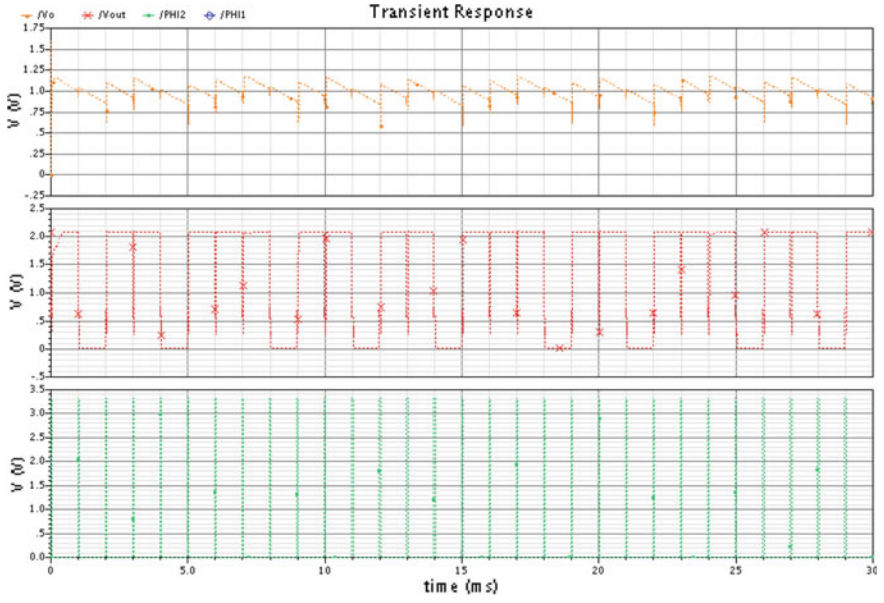


Fig. 7.86 The simulated output of a unit cell circuit when the photocurrent is 3/4 of the maximum

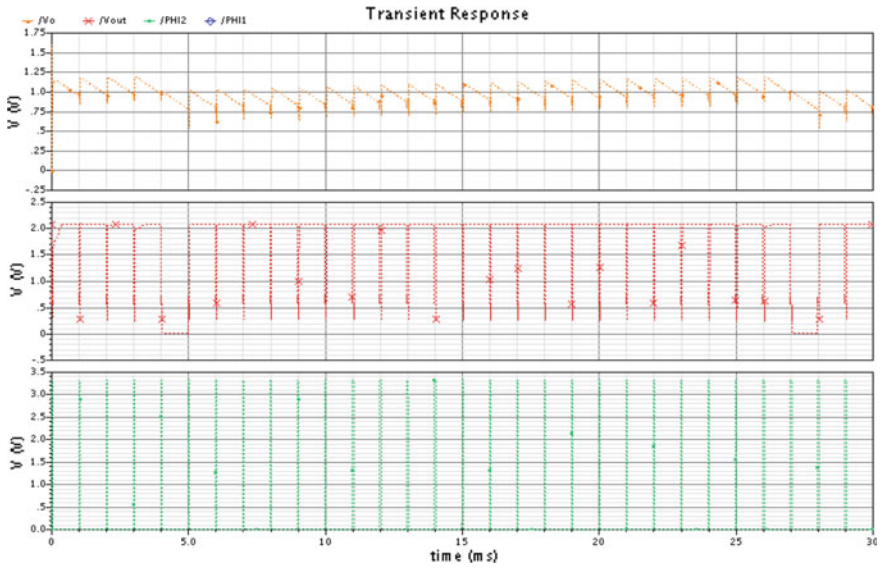


Fig. 7.87 The simulated output of a unit cell circuit when the photocurrent is the maximum

the whole focal plane generates a two-dimensional bit-plane each time. Each frame of the image is composed of L bit-planes. L is determined by the SNR

$$\mathrm{SNR} = 9\log_2(L) - 5.2\ \mathrm{dB} \tag{7.47}$$

(3) The simulation results

Although each bit of the digital output of the comparator does not represent the magnitude of the input analog signal, the statistical property of the digital bit-stream signal out from the comparator can reflect the magnitude of the input analog signal.

Figures 7.84, 7.85, 7.86 and 7.87 are the simulation results for a unit cell circuit when the photocurrent is 1/4 of the maximum, half of the maximum, 3/4 of the maximum, and the maximum, respectively. The upper graph is the sampling pulse, the middle graph is the output of a unit cell circuit, and the lower graph is the output of the integrator.

(4) The layout implementation

The layout of the 8 × 8 digital ROIC is implemented with the design rule of 0.5 μm DPDM CMOS process. The pixel size is 50 μm × 50 μm. The layout for a unit cell circuit is shown in Fig. 7.88. The layout for the chip of 8 × 8 digital ROIC is shown in Fig. 7.89. The circuit structure is suitable for an infrared detector with a high photocurrent. Each unit cell has an ADC circuit and the area of cell can be controlled within 50 μm × 50 μm. Because the comparator operates in the

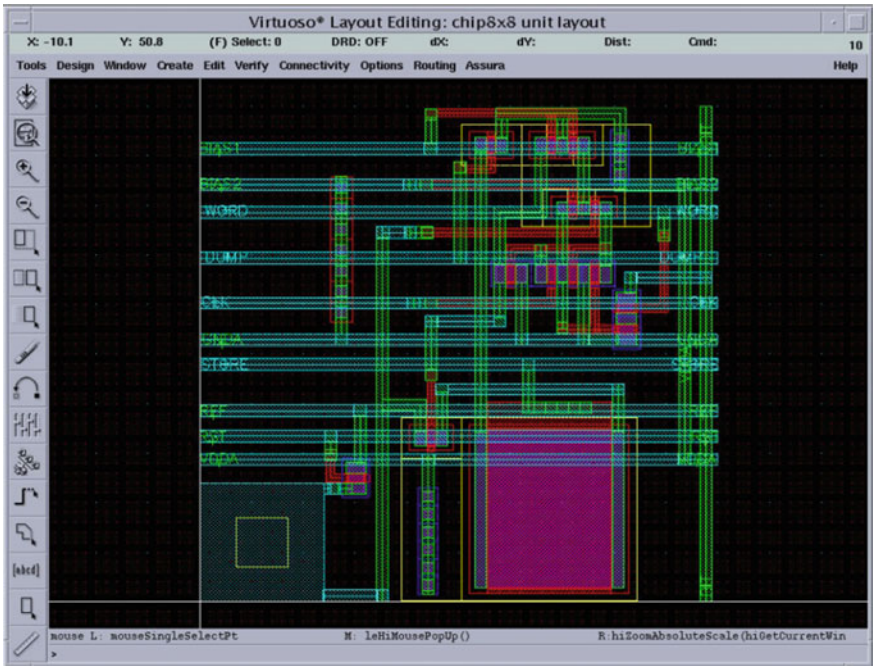


Fig. 7.88 The layout of a unit cell circuit

sub-threshold region, the simulation shows that the power consumption can be controlled within 10 μW.

4. **Several applied ADC circuits at the Pixel-level structure**

William Mandle of the American Amain Company and Boyd Fowler, Abbas EL Gamal of Stanford University and other researchers have been committed to the design of the pixel-level ADC circuit for IRFPA.

In 1994, Stanford University introduced a pixel-level ADC topology, which is a first-order Σ-Δ oversampling ADC circuit shared by four pixels [20]. The advantages of the circuit are small area cost and a large charge-handling capacity. The disadvantage is the high output data rate. To reduce the output data rate, a Nyquist sampling rate MCBS ADC was designed by Stanford University in 1999 [21]. The advantages of the circuit are small area cost and the low FPN for the gain and bias. The disadvantages are that the gain bandwidth of the comparator is small and that the response sensitivity of the photodiode is low. Through analyzing two algorithms for ADC, the circuit designers conclude that the Σ-Δ ADC is more suitable for infrared imaging, and MCBS ADC is more suitable for visible light imaging.

In 1996, using two-stage MOSAD algorithm, a 640 × 480 IRFPA with the pixel-level ADC was developed by William Mandle, and the structure of the input stage circuit is the BDI circuit [10]. Results from the paper showed that better than

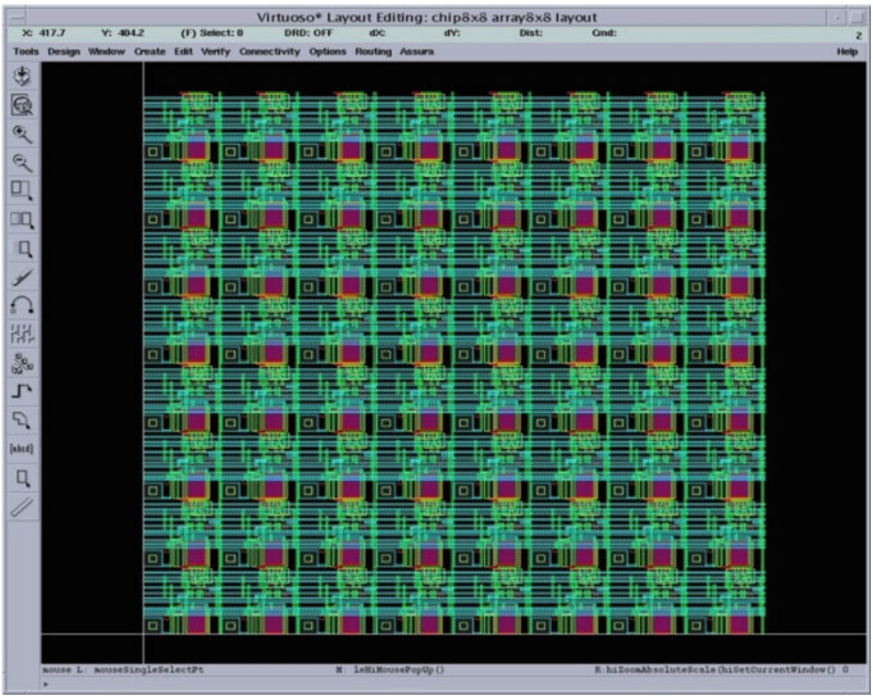


Fig. 7.89 The layout for the chip of 8 × 8 digital ROIC with the first-order Σ-Δ oversampling ADC

14 bits could be achieved and that well capacity could be increased to support high background needs. The Amain Electronics Company of the United States developed an 640 × 480 MW MCT IRFPA by adopting the pixel-level MOSAD ADC structure.

In 2007, the development on the ADC embedded in a 25 μm pixel is reported by the LETI of France [22]. With a 14-bit at 50/60 Hz video rate, the ADC structure has been applied to 25 μm pitch bolometer sensors with a dedicated 320 × 240 IRCMOS circuit, and the room temperature sensitivity and the DC offset are solved directly in the pixel. In 2010, LETI and SOFRADIA of France presented a new 15-bit pixel-level ADC which is implemented in a 25 μm pixel pitch [23]. The pixel-level ADC is based on charge packets counting technique and the ADC dimensioning has been optimized for LWIR imaging. This technique offers a large well capacity that paves the way for a breakthrough in thermal sensitivity. In electro-optical test, a peak NETD value of 2 mK has been reached with power consumption under 150 mW.

7.5 Measurement and Testing Technology for Focal Plane Array

With the development from "one pixels" to "focal plane array (FPA)" for infrared and ultraviolet detectors in the past twenty years, the measurement and testing technology has also improved. The early testing method for infrared detectors with one pixel or multiple pixels was as follows: a cavity black body was used as the infrared radiation source, the detectors were applied with bias current or bias voltage, the output of the detectors was amplified through a preamplifier and input to the acquisition card, and then the measurement results were output through the computer processing. With the introduction of an extended area blackbody to the measurement system for infrared focal plane, the new testing technology has been studied and developed. The signal of infrared detectors could be amplified through ROIC on the focal plane, so the performance of the output has a great improvement. Meanwhile, the measurement system is required to provide analog drivers and timing pulses for ROIC, so it is important to incorporate a low-noise driver circuit. With the increase in testing requirements, the measurement system has a continuous upgrade. Generally, a low-noise bias circuit, a pulse generator and a real-time imaging function would be integrated in a latest measurement system for FPA, which is versatile and suitable for various types of detectors for in-depth analysis. With the development of FPA, the measurement and testing technology is developed on various aspects such as the measurement on the photovoltaic effect of the detector, the testing of ROIC, the measurement and evaluation of the device after hybridization, the measurement on the photoelectric performance of FPA and reliability testing, etc.

Table 7.10 The characteristic parameters and related quantities to evaluate IRFPA

Parameters		Related quantities
Responsivity	Nonuniformity	Integration time
Detectivity	Operable pixel	Frame period
Noise	Dead pixel	Line period
Noise equivalent irradiation power	Over hot pixel	Charge capacity
Saturation irradiation power	Max pixel rate	Irradiation power
Dynamic range	Noise equivalent temperature difference	Irradiation energy
Relative spectral response	Cross talk	Field of view
Spectral response range	Modulation transfer function	Fill factor
Quantum efficiency	Fixed pattern noise	Pixel amount
Linearity		

Table 7.11 The evaluation parameters classified according to characteristics testing

Characteristic	Name of parameter
Photoelectric conversion characteristic	Responsivity, quantum efficiency, fill factor, detectivity
Noise characteristic	Noise voltage, noise equivalent irradiation power, noise equivalent temperature difference
Spectral characteristic	Relative spectral response, spectral response range
Dynamic characteristic	Dynamic range, saturation irradiation power, noise equivalent irradiation power, linearity
Spatial resolution feature	Modulation transfer function, cross talk
Temperature resolution characteristic	Noise equivalent temperature difference
Response uniformity characteristic	Nonuniformity, dead pixel, over hot pixel, FPN
Frequency characteristic	Maximum pixel rate, frame period, line period

7.5.1 Measurement of Parameters for Infrared FPA

1. **The characteristic parameters of infrared FPA**

According to physical properties of IRFPA, there are a total of 28 characteristic parameters and related quantities (including 19 characteristic parameters, 9 related quantities) to fully evaluate IRFPA, as shown in Tables 7.10 and 7.11, which comprehensively describe the characteristics of IRFPA from eight aspects [24].

2. **The classification of the testing for characteristic parameters of IRFPA**

There are many performance parameters which should be measured to evaluate the characteristics of IRFPA, including responsivity, noise, detectivity, noise equivalent temperature difference (NETD), linearity, modulation transfer function (MTF), minimum resolvable temperature difference (MRTD), etc. There are three features on the testing for IRFPA: First, the blackbody source is not modulated during the test (except spectrum testing), that is a constant irradiation, so the blackbody response signal is measured by eliminating the background; Second, the sample hold method must be used because the output signal is discrete. Third, the noise is usually measured as a root mean square (RMS) voltage by calculating signal statistical fluctuation. By now, the test methods of "responsivity", "noise equivalent temperature difference", and "dynamic range" are all based on these three features.

Basically, the testing for IRFPA can be divided into three categories: floodlight mode testing, focused mode testing, and system-level testing.

(1) **Floodlight mode testing**

The floodlight mode testing includes measurements on noise, responsivity, NETD, noise equivalent irradiation power (NEP), detectivity (D*), nonuniformity,

and linearity. During the measurement, a blackbody is used to floodlighting IRFPA at a series of temperature points, and then the output signal of each pixel under the corresponding temperature is acquired.

It is a key to measure the noise and responsivity for the evaluation of characteristics of a detector. The noise is the RMS deviation of the output voltage signal over time with a constant input signal. The noise testing is implemented by measuring the output voltage signal several times under a constant irradiation with a blackbody at or close to room temperature. The responsivity is the change of the output voltage signal under a unit irradiation. There are several different representations for the responsivity, such as blackbody responsivity, peak responsivity, etc. Responsivity and noise are the basis to calculate parameters of D*, NETD and NEP, etc. Therefore, the measurement accuracy of responsivity and noise is very important, and an error in measurement will lead to an error in other parameters.

Nonuniformity is the change for the output signal of each pixel under a constant and uniform irradiation, including the difference between the output signals under the fixed background (FPN) and the differences between the output signals with varied irradiation (response nonuniformity). Nonuniformity usually is expressed as a maximum deviation or standard deviation of the statistics on the output signals when the IRFPA is excited by infrared radiation from a blackbody.

Traditionally, the linearity testing is implemented by using an extended area blackbody or a cavity blackbody to floodlighting the IRFPA. The output signals of IRFPA would vary with the temperature change of blackbody, so the linearity is used to represent the relationship between the output signal and the irradiation power, and can be acquired by calculating the deviation from the fitted straight line. This testing method is straightforward, while only having an accuracy of approximately 2 %.

(2) **Focused mode testing**

In focused mode testing, the infrared radiation energy from a blackbody source is focused on the individual pixel of detector. Generally, focused mode testing is used to measure the cross talk and MTF which reflect the spatial characteristic of IRFPA.

Cross talk reflects the influence between the response signals of adjacent pixels. In principle, it is a simple test: a small light spot is focused on a pixel, and then the output signals of the pixel and adjacent pixels are measured. Since the pixel size of an advanced IRFPA is close to the optical diffraction limit, it is very difficult to get alight spot smaller than the pixel size, as well as to guarantee the accurate focus of the light spot in the center of pixel. Therefore, the cross talk testing actually is a very demanding technology.

Applying a standard optical imaging system on IRFPA, the MTF could be calculated by the Fourier transform from the curve of the output signal, which is obtained by scanning across the slit of the focal plane.

(3) **System-level testing**

System-level testing is generally the performance measurement of IRFPA with an optical lens, and mainly includes the MRTD, signal transfer function (SiTF), etc.

MRTD is generally measured by imaging a 4 bar target on the focal plane and then changing the temperature until the four bars can just be distinguished. MRTD depends on the size of the bar, as a larger bar is easy to distinguish. The curve of MRTD can be obtained by drawing the temperature difference according to the target space frequency change. MRTD is used to evaluate both sensitivity and spatial resolution of the system, so the measurement of MRTD is very important.

(4) **The analysis of error sources**

There are some standard components and test equipment involved in the test of characteristics of IRFPA, so the error can be resulted from many sources. The error may come from the inherent error of the measurement system, or be produced due to the no repeatability of equipment during the transformation between different measurements. There are three main types of error such as optical error, electrical error, and mechanical error.

Optical error is mainly produced by poor alignment. Many optical parts are often used to focus the infrared energy on the detector. Each optical part has at least three degrees of freedom. The position errors of these optical elements are cumulative, but it is difficult to determine the value. In the light string test, for example, the size of light spot is a key optical parameter. A standard optical system is required to generate a very small light spot on the detector, and is very sensitive to defocusing error. Scanning testing need to align the light spot for *x*-*y* translation with the orientation of detector array, otherwise there will be a stray light and vignette problems, which would cause the measurement error.

The measurement of parameters such as responsivity and noise can be referred to the measurement of response voltage. The main error is from the electrical noise of equipment, which is used to acquire the output signal. In the testing environment of IRFPA, the noise asynchronous with the operation of IRFPA comes from electronic components such as the computer, external power, digital signal processing equipment, and so on. Therefore, in the testing, we must eliminate possible pathways between the device under test and potential noise sources to ensure the accuracy of the measurement.

To realize the consistency of measurement results, the repeatability of testing equipment is a key. It is difficult to change the test configuration of the system consisting of many adjustable components. Therefore, a measurement system is generally composed of a set of component modules, which can be replaced rapidly. The system is quickly adjusted, and the configuration is accurately recorded, thus achieving the repeatability of test results. The method of integration eliminates many errors related to the assembling adjustment and the alignment for various test configurations.

3. **The measurement of parameters for IRFPA**

(1) **Measurement of parameters such as responsivity, detectivity, and NETD**

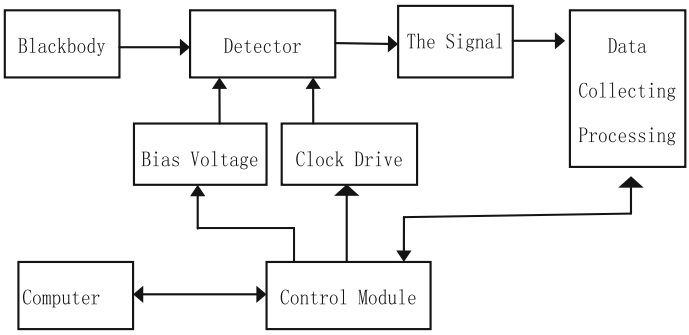


Fig. 7.90 The block diagram of a measurement system

The measurement of parameters such as responsivity, etc., can be referred to the measurement of response voltage signal under two different blackbody temperatures, that is the temperature $T0$ and the temperature $T1$. After the response voltage is measured, characteristic parameters such as responsivity can be calculated according to the definition. Figure 7.90 shows the block diagram of a measurement system, which is composed of the blackbody source, the detector array under test, the electronic circuits, computer, etc. The testing method is as follows: A blackbody with adjustable temperature is placed close to the detector array under test, the timing pulses and bias voltage is supplied. After system initialization, the blackbody

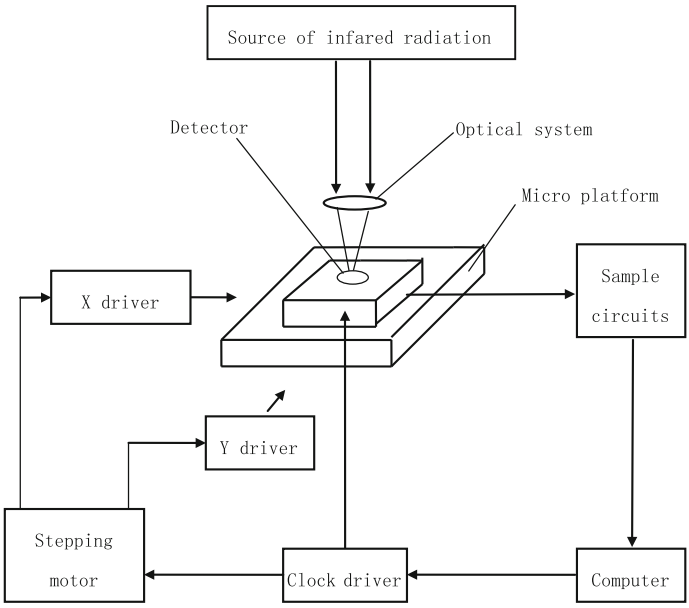


Fig. 7.91 The block diagram of a testing system for cross talk

temperature is changed to $T1$, and the output voltage signal $V1$ is recorded by the data acquisition system. Then the blackbody temperature is changed to $T2$, and the output voltage signal $V2$ is recorded again. The characteristic parameters of the detector array will be calculated by substituting $V1$ and $V2$ in corresponding formula. The Shanghai Institute of Technology and Physics, Chinese Academy of Sciences, has drafted GB/T17444, the technical norms for measurement and testing of characteristic parameters of infrared focal plane arrays [25].

(2) **Cross talk testing**

In the detector, the cross talk signal produced by the influence of the adjacent pixels is called cross talk [26]. Figure 7.91 displays the block diagram of a testing system for cross talk, which is composed of a small light spot system, the device under test, the micropositioning platform, electronic circuits, and other components.

The optical system installed between the blackbody and the detector is used to focus the blackbody radiation to one pixel of the detector. The response voltage of the tested pixel and adjacent pixels are measured at the same time, and the cross talk is calculated as the percentage of the voltage of adjacent pixels over that of the tested pixel.

(3) **Spectral response test**

The response of an infrared detector to different wavelengths of irradiation is called spectral response. Figure 7.92 shows the block diagram of spectral response test system, which is composed of the radiation source, monochromator, the detector under test, the signal processing circuit, and the computer for data acquisition.

Generally, a high temperature blackbody is used as the radiation source, which radiates to the detector through the monochromator. There are two kinds of detectors for the test, one of which is a standard pyroelectric detector, and the other is the detector under test.

For a standard pyroelectric detector

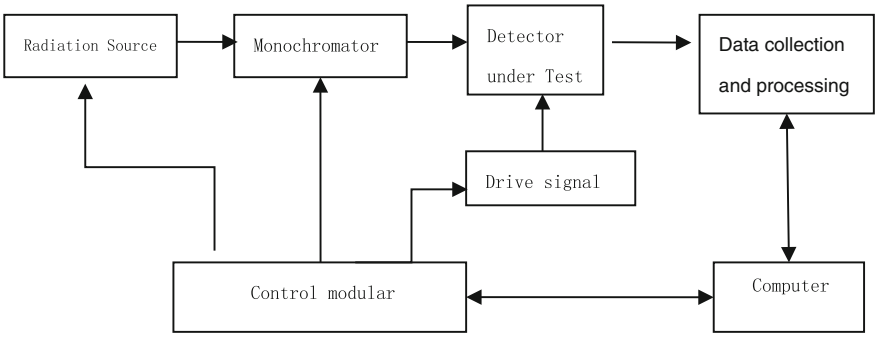


Fig. 7.92 The block diagram of spectral response test system

$$V_{\text{pyro}}(\lambda) = k \times R(\lambda) \times T(\lambda) \tag{7.48}$$

where $R(\lambda)$ is the blackbody radiation, $T(\lambda)$ is the transmittance of monochromator, and k is the electrical gain which is constant. The standard detector is a pyroelectric detector of full spectrum.

Similarly, for the detector under test

$$V_{\text{det}}(\lambda) = S(\lambda) \times R(\lambda) \times T(\lambda) \tag{7.49}$$

where $R(\lambda)$ is the blackbody radiation, $S(\lambda)$ is the spectral response of the detector, the spectral response curve $S(\lambda)$ is $V_{\text{det}}/V_{\text{pyro}}$. The final result should be normalized.

(4) **Frequency response and the response time**

The frequency response of a photovoltaic detector is mainly decided by three factors

(1) The diffusion time τ_{n} to junction (barrier) area for photon-generated carriers
(2) The drift timeτ_{d} of photon-generated carriers under the effect of electric field in the body
(3) Circuit constant τ_{c} determined by the junction capacitance C_{d} and load resistance R_{L}.

So the total response time of photovoltaic detector is as follows:

$$\tau = \tau_{\text{d}} + \tau_{\text{n}} + \tau_{\text{c}} \tag{7.50}$$

The above three response time parameters are related to photovoltaic detector materials, structure, working condition, and subsequent circuit.

4. **The testing system for parameters of IRFPA**

It is a complex task to test and evaluate the IRFPA with different materials, different readout integrated circuits, and different functions, so the testing system should be versatile, universal, and highly automatic. With the development of IRFPA, the corresponding testing system must be continuously updated. At present, the main components of an infrared testing system include a bias voltage card, clock card, preamplifier and acquisition card, and relevant control software. Its working principle is as follows: Through the software programming, the parameters of system hardware are configured, the power, bias voltage, and timing pulses are generated to drive the IRFPA, then the output data is acquired to calculate the characteristic parameters of the IRFPA (responsivity, the detectivity and nonuniformity, etc.). The main testing system is integrated with the function of real-time imaging, and the gray level of an IRFPA real-time image can be measured by histogram statistics. It will be more comprehensive and effective to evaluate the staring focal plane, which is mounted with a band-matching lens as is the same as a linear focal plane equipped with a swing lens.

Fig. 7.93 A practical testing system for comprehensive parameters of IRFPA

The rapid development of measurement and testing techniques is promoted by the technology progress of IRFPA. The demands for more convenient, real, and precise measurement have been the focus in the market. Currently, the major testing system suppliers include HGH of France, SBIR and PI of the United States, CI Company of Israel, etc.

Figure 7.93 shows a practical testing system for comprehensive parameters of IRFPA, which can measure the spectral response of short-wave, medium-wave, long-wave infrared focal planes, as well as responsivity, detectivity, noise, cross talk, NETD, and so on. It can locate the position of the noneffective pixel of the focal plane detectors and provide effective data to analyze the test result. Such a system can realize the real-time imaging of an array detector, and has the function of nonuniformity correction, which intuitively reflects the device performance and provides effective reference for detector applications.

7.5.2 Measurement of Parameters for Ultraviolet FPA

The main difference between the testing methods for ultraviolet FPA and infrared FPA is that the radiation source is different and the calibration for radiation power is varied. In the test of an ultraviolet FPA, the light source is one of the key components. The light source should be chosen to have a large radiance in the band of ultraviolet and provide high radiation, so the detector under irradiation would have a higher output response thus improving the SNR. Meanwhile, the light source should have good stability in order to improve the measurement repeatability and accuracy. Conventional ultraviolet light sources are deuterium lamp, xenon lamp, mercury lamp, etc. Deuterium lamps, which are commonly used as ultraviolet and vacuum ultraviolet transmission standard light sources, have the required stability,

repeatability, small size, and low power consumption. The lamp can emit strong ultraviolet radiation and produce continuous radiations ranging from the wavelength of 165–370 nm. However, a typical deuterium lamp only has an output power of 30 W. Through a spectral monochromator, the power of emit light is too weak to be detected by the detector. Xenon lamp has a continuous spectrum and high output power, generally more than 250 W. With a reasonable optical design and circuit design, the stable output optical power could be achieved for a xenon lamp, which can be applied to the testing for the ultraviolet FPA. Mercury lamp has a discontinuous line spectrum. Its output power generally can reach more than 250 W, while the energy is mainly set in dozens of spectral lines, of which the 365 nm spectral line output the biggest power. So a mercury lamp can provide a single spectral line that has enough light energy, and be applied to the measurement of responsivity and detectivity for the ultraviolet FPA at a particular wavelength.

The calibration of radiation power mainly has two types of standards: (1) The standard based on the light sources, which are commonly consisted of blackbody and synchrotron radiation. When the blackbody is used as the standard, the principle is based on the Planck's law, Stefan Boltzmann's law and Lambert's law. Its precision depends on the stability of electrical power supply, the precision of temperature control, the measurement precision for the exit aperture of blackbody, etc. The blackbody is normally used as infrared radiation standard. When the synchrotron radiation is used as the standard, its spectral radiation power can be accurately calculated by Schwinger formula, and the main factors affecting the accuracy are the magnetic field, the electron energy, electronic flow, and geometrical structure. A synchrotron radiation light source is often used for ultraviolet radiation standard. (2) The standard based on the detectors, which are mainly divided into the self-calibrating detector and the electric substitution radiometer (ESR). The self-calibration detector must be based on the proven technology. Based on the principle of photoelectric equivalence, the ESR is to convert the measurement of radiation power to the measurement of electric power, so it is an absolute calibration method. There are two types of ESR, one is a cryogenic radiometer and the other is a room temperature radiometer, which is based on the electrical

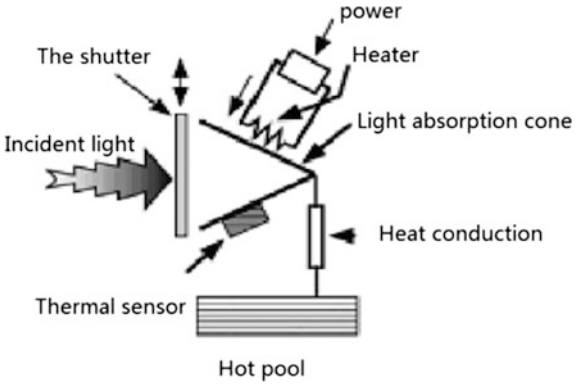


Fig. 7.94 The working principle diagram for a cryogenic radiometer

calibration pyroelectric detector. The working principle of a cryogenic radiometer is described in Fig. 7.94. Both light irradiation and electrical heating can make the temperature of the absorption cavity rise. When the increase of temperature is same, the electrical power is equivalent to the light power and its precision depends on the accurate measurement of the current and voltage and photoelectric equivalence effect. The working principle of the electrical calibration pyroelectric detector is similar to the cryogenic radiometer, while it is an absolute calibration method operating at room temperature, which is also based on photoelectric equivalence effect. A pyroelectric detector can respond to the ultraviolet radiation and the electrical heating, so it can convert the measurement of ultraviolet radiation power to the measurement of an equivalent electrical heating power.

In the ultraviolet waveband, during the transfer process from the standard light source based on synchrotron radiation to a ultraviolet light source such as deuterium lamp, the measurement of irradiation technology will further cause the error due to the limitation of stability and repeatability. The current development trend is to build a calibration method for ultraviolet radiation power by using the ESR. In recent years, a radiation standard was gradually established on the basis of an electric substitution cryogenic radiometer by the International standards organization and developed countries, which integrated advanced technology such as Electric Substitution, double cold shielding, high vacuum, superconducting, and so on. This evolution has enabled the radiation calibration technology to reach unprecedented accuracy. However, because the cryogenic radiometer needs some advanced technologies, only a few organizations, such as the NIST of America, NPL of British and PCB of Germany, have the ability to develop it.

The Shanghai Institute of Technical Physics has also carried out the research on the calibration method for ultraviolet radiation based on the ultraviolet electrical calibration pyroelectric detector [27]. The structure of an ultraviolet electrical calibration pyroelectric detector was designed and the practical calibration detector was developed. The test platform for the relative spectral response of an ultraviolet detector was established by exploring highly stable ultraviolet radiation sources and studying the factors of spectral measurements and multi-beam split processing. Although the uncertainty of the measurement for the responsivity of ultraviolet FPA can reach 5 %, the test method has been submitted for the establishment of a Chinese Military Standard.

7.6 Summary

As the representative of the third generation of IRFPA, dual-band/multiband IRFPA, digital IRFPA have characteristics such as high speed, high reliability, small size, low cost, and excellent performance for detection, and have become the hot topic for the technology development of IRFPA. The dual-band ROIC for signal processing on IRFPA and the digitalization of ROIC are introduced and illustrated with recent implementations in this chapter. In addition, the test techniques for the

infrared focal plane and the ultraviolet focal plane are the basis for the correct evaluation on FPA, so the testing techniques are also described.

This chapter is summarized as follows:

(1) A brief review of the development for dual-band detectors and readout circuits are presented. The circuit topology, operation mode, and features for the dual-band ROIC are discussed in detail. According to the characteristics of dual-band IR detectors with different structures, a simultaneous 256 × 2 dual-band ROIC and a sequential 128 × 128 dual-band ROIC are implemented. For the simultaneous dual-band ROIC, a current mirror gate modulation structure is adopted as the input stage circuit and the function of signal separation is also realized in the input stage. For the sequential dual-band ROIC, the DI structure is used as the input stage circuit and the load of output buffer is shared in column level. The infrared imaging is realized for IRFPA assembled with ROIC and dual-band detector array in hybridization. In addition, the implementation of a 128 × 128 ultraviolet ROIC is presented, and ultraviolet imaging is also displayed.

(2) The development for the digitalization of IRFPA is introduced. Three architectures including the chip level, the column level, and the pixel level for the implementation of A/D conversion on the focal plane are discussed. The main algorithms including the voltage-frequency-modulation ADC, the sigma-delta ADC, the MCBS ADC, and the single-slope/dual-slope integral ADC are presented and analyzed, respectively. The performance parameters, advantages, and disadvantages for each algorithm are listed. Two implementations based on the first order Σ-Δ oversampling ADC are illustrated respectively for a column-level 128 × 2 ROIC and a pixel-level 8 × 8 ROIC. An 8 × 8 digital ROIC based on the pixel-level structure with the single-slope integral ADC is implemented.

(3) The characteristic parameters of infrared focal planes are briefly introduced, and three basic types of infrared focal plane tests are described, including floodlight mode test, focus mode test, and system-level test. The sources of measurement deviation are analyzed. The test methods and equipment for infrared focal plane parameters including responsivity, detectivity, noise equivalent temperature spectral response, cross talk, and so on are presented. Meanwhile, basic test requirements and methods of UV focal plane parametric are introduced.

References

1. Bai Y, Bernd SG, Hosack JR, et al. Hybrid CMOS focal plane array with extended UV and NIR response for space applications. In: SPIE's 48th annual meeting, 3–8 Aug 2003, San Diego, CA.

2. Andresen BF, Fulop GF, Norton PR. The architecture and performance of SCD's 17 μm pitch VOx μ-bolometer detector. In: Proceedings of SPIE. vol. 7298. 2009. 72980R.
3. Andresen BF, Strojnik M. Standardized high-performance 640 × 512 readout integrated circuit for infrared applications. In: Proceedings of SPIE. vol. 3698. 1999. pp. 766–777.
4. Smith EPG, Bornfreund RE, Kasai I, et al. Status of two-color and large format HgCdTe FPA technology at Raytheon vision system. In: Proceedings of SPIE. vol. 6127. 2006. 61271F.
5. Reine MB,Hairston A,Dette PO,et al. Simultaneous MW/LW dual-band MOVPE HgCdTe 64x64 FPAs. In: Proceedings of SPIE. vol. 3379. 1998. pp. 200–212.
6. Baylet P, Zanatta J-P, Chance D, et al. Recent advances in the development of infrared multispectral 1282 FPAs. In: Proceedings of SPIE. vol. 4721. 2002. pp. 134–143.
7. Finch JA,Graham RW, et al. IRFPA ROIC with dual TDM reset integrators and sub-frame averaging functions per unit. US Patent No. 6885002. April 2005.
8. Johnson JF. Direct injection readout circuit model. In: Proceedings of SPIE. vol. 2226. 1994. pp. 120–129.
9. Liu F, Ding R, et al. 256 × 1 readout integrated circuit with simultaneous integration mode for two-color infrared detectors. In: Proceedings of SPIE. vol. 7158. 2009.
10. Mandl W, Kennedy J, Chu M. MOSAD IR focal plane per pixel A/D development. In: SPIE proceedings. vol. 2745. 1996. pp. 90–98.
11. Mandl W, Shen C, Martin P. Images and test results of MOSAD all digital 640 × 480 MWIR prototype camera. In: SPIE Proceedings, vol. 4131. 2000. pp. 355–363.
12. Mandl W. Four kiloframe, 14 bit, 128 × 128 digital imaging spectrometer. In: SPIE proceedings. vol. 4486. 2001. pp. 385–392.
13. Elkind S. Focal plane processor with a digital video output for InSb detectors. In: Proceedings of SPIE. vol. 4820, infrared technology and applications XXVIII. 2003. P751–P758.
14. Fillon P, Dugalleix S, Pistone F, Tribolet P. Digital output for high performance MCT staring arrays. In: Proceedings of the SPIE. vol. 6206. 2006. pp. 62060U.
15. Denyer PB, Renshaw D, Gouyu W, Ying LM. On-chip CMOS sensors for VLSI imaging systems. In VLSI 91, Edinburgh, UK, August 1991.
16. Dickinson A, Mendis S, Inglis D, Azadet K, Fossum ER. CMOS digital camera with parallel analog to digital conversion architecture. In Program of 1995 IEEE workshop on eeDs and advanced image sensors, Dana Point, CA, 20–22 April 1995.
17. Chen G, Shao Z, Chen J, et al. Design of CMOS integrated circuits. Xi'an: Xi'an Jiaotong University Press; 2000. P96–P101,P124–P126,P139–P156, .
18. Gan W,Ding R,Ni Y. Readout circuit with pixel level analog to digital conversion. Infrared components and their applications. In: SPIE, SPIE photonics Asia conference, China. vol. 5640. 2004.
19. Gan W, Ding R. Implementation of a readout circuit with pixel-level analog to digital conversion on IRFPA. In: Joint 31st international conference on infrared and millimeter waves and 14th international conference on terahertz electronics, Shanghai, China. 18–22 Sept 2006.
20. Fowler B, El Gamal A, Yang A. A CMOS area image sensor with pixel-level A/D conversion. ISSCC Digest of Technical Papers, San Francisco. 1994. pp. 226–227.
21. Yang D, Fowler B, El Gamal A. A nyquist-rate pixel-level ADC for CMOS image sensors. IEEE J Solid-State Circ. 1999;34(3):348–356.
22. Tchagaspanian M, Villard P,Dupont B. Design of ADC in 25um pixels pitch dedicated for IRFPA image processing at LETI.Infrared Technol Appl XXIII. 2007;6542:65421W1–65421W12.
23. Bisottoa S, de Borniola E, Mollard L, et al. A 25 μm pitch LWIR staring focal plane array with pixel-level 15-bit ADC ROIC achieving 2mK NETD. In: Proceedings of SPIE. vol. 7834. 2010. 78340J.
24. Liangchu Dong, Ruijun Ding, Zhengkai He. The study of the definition and measurement methods of characteristic parameters of IRFPA. Infrared Laser Eng. 1997;26(3):14–8.
25. Dong L, Ding R, Liang P. GB/T 17444 measuring methods for parameters of infrared focal plane arrays. Standards press of china; 1999 (Book number:155066.1–15431).

26. Dang Khoa V, Kauffman Christopher L, Derzko Zenon I. Infrared focal plane array crosstalk measurement. SPIE. 1992;1686:125–35.
27. Shao X. The development of a UV electrically calibrated detector and the study of calibration application. Dissertation, Jan 2009. Shanghai Institute of Technical Physics of The Chinese Academy of Sciences.