CMOS BIOTECHNOLOGY

Edited by
Hakho Lee
Donhee Ham
Robert M. Westervelt

Springer

CMOS Biotechnology

Series on Integrated Circuits and Systems

Series Editor: Anantha Chandrakasan
Massachusetts Institute of Technology
Cambridge, Massachusetts

CMOS Biotechnology
Hakho Lee, Donhee Ham and Robert M. Westervelt
ISBN 978-0-387-36836-8

SAT-Based Scalable Formal Verification Solutions
Malay Ganai and Aarti Gupta
ISBN 978-0-387-69166-4, 2007

Ultra-Low Voltage Nano-Scale Memories
Kiyoo Itoh, Masashi Horiguchi and Hitoshi Tanaka
ISBN 978-0-387-33398-4, 2007

Routing Congestion in VLSI Circuits: Estimation and Optimization
Prashant Saxena, Rupesh S. Shelar, Sachin Sapatnekar
ISBN 978-0-387-30037-5, 2007

Ultra-Low Power Wireless Technologies for Sensor Networks
Brian Otis and Jan Rabaey
ISBN 978-0-387-30930-9, 2007

Sub-Threshold Design for Ultra Low-Power Systems
Alice Wang, Benton H. Calhoun and Anantha Chandrakasan
ISBN 978-0-387-33515-5, 2006

High Performance Energy Efficient Microprocessor Design
Vojin Oklibdzija and Ram Krishnamurthy (Eds.)
ISBN 978-0-387-28594-8, 2006

Abstraction Refinement for Large Scale Model Checking
Chao Wang, Gary D. Hachtel, and Fabio Somenzi
ISBN 978-0-387-28594-2, 2006

A Practical Introduction to PSL
Cindy Eisner and Dana Fisman
ISBN 978-0-387-35313-5, 2006

Thermal and Power Management of Integrated Systems
Arman Vassighi and Manoj Sachdev
ISBN 978-0-387-25762-4, 2006

Leakage in Nanometer CMOS Technologies
Siva G. Narendra and Anantha Chandrakasan
ISBN 978-0-387-25737-2, 2005

Statistical Analysis and Optimization for VLSI: Timing and Power
Ashish Srivastava, Dennis Sylvester, and David Blaauw
ISBN 978-0-387-26049-9, 2005

Hakho Lee
Donhee Ham
Robert M. Westervelt (Editors)

CMOS Biotechnology

Springer

Editors:
Hakho Lee
Center for Molecular Imaging Research
Massachusetts General Hospital,
Harvard Medical School
Charlestown, MA 02129
USA

Donhee Ham
School of Engineering and Applied Sciences
Harvard University
Cambridge, MA 02138
USA

Robert M. Westervelt
School of Engineering and Applied Sciences,
Department of Physics
Harvard University
Cambridge, MA 02138
USA

Series Editor:
Anantha Chandrakasan
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology
Cambridge, MA 02139
USA

Library of Congress Control Number: 2007921679

ISBN 0-387-36836-1 e-ISBN 0-387-68913-3
ISBN 978-0-387-36836-8 e-ISBN 978-0-387-68913-5

Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

springer.com

— Great things are not done by impulse,
but by a series of small things brought together.

Vincent van Gogh

PREFACE

Understanding biology at a cellular and molecular level is an important challenge for the coming century. Biological systems can be incredibly complicated and sophisticated, with many processes occurring simultaneously. Powerful tools will be needed to manipulate and test interacting cells and biomolecules, and to observe their behavior.

Microfluidic chips are well suited to these tasks – they provide a biocompatible environment for cells in a fluid with the proper surfaces, held at the right temperature. Using microchannels, one can sort or assemble cells according to their characteristics, and perform chemical tests. Advances in microfluidics are very promising for biology and medicine.

The complexity of biological systems and their parallel nature is well matched to integrated circuits. The CMOS industry produces programmable microprocessors containing over a hundred million transistors that operate at GHz speeds, as well as high-resolution displays and imaging chips. One can adapt the power of CMOS chips to biotechnology by joining the integrated circuit with a microfluidic system to form a hybrid chip. In this way, one can control the position of cells or biomolecules in fluid using spatially patterned electromagnetic fields, and sensitively sense their response for observations and tests.

The aim of the book is to explore this powerful new approach for biotechnology where the sophistication of CMOS integrated circuits is joined with the biocompatibility of microfluidic systems. Broad research activities of high current interest are covered, with each chapter contributed by experts in the field. We hope that the volume will provide a timely overview of the exciting developments in this nascent field, serving as a springboard for readers to join in.

This book is the culmination of the concerted effort from many people. First and foremost, we thank all the participating authors for their invaluable contributions. Our deep gratitude goes to Professor Chandrakasan at MIT for his helpful suggestions and W. Andress at Harvard for his kindhearted advice. We also gratefully acknowledge the support by Springer, especially from C. Harris and K. Stanne. Last but not least, we give our sincere thanks to our families for their patience and encouragement during the preparation of the book.

Cambridge, Massachusetts
January 2007

Hakho Lee
Donhee Ham
Robert M. Westervelt

CONTENTS

Part II. CMOS Actuators

Part IV. CMOS Optical Sensors

1 INTRODUCTION

Donhee Ham[1]*, Hakho Lee[2,3] and Robert M. Westervelt[1,2]

[1] School of Engineering and Applied Sciences, Harvard University
[2] Department of Physics, Harvard University
[3] Center for Molecular Imaging Research, Massachusetts General Hospital, Harvard Medical School
*donhee@deas.harvard.edu

The second half of the 20th century witnessed the metamorphosis of silicon, an element common in the Earth's crust, into silicon integrated circuits (ICs), complex superstructures that can contain hundreds of millions of complementary-metal-oxide-semiconductor (CMOS) transistors in a tiny footprint of only a few square centimeters. An arsenal of planar microfabrication technologies made possible the rock-to-IC transformation of silicon at surprisingly low costs. Now consider the phenomenal ability of silicon ICs. Hundreds of millions of CMOS transistors interconnected via a labyrinthine maze of metallic wires all work together to process data at gigahertz frequencies.

With their amazing capabilities, inexpensive production, and tiny physical size, ICs have come to have a major affect on our daily lives. As computer microprocessors, they significantly assist our intellectual endeavors. Silicon ICs profoundly enrich our ability to communicate, enabling communication technologies with high speed and data capacity over long distances. In addition, they provide entertainment: music and movies from handheld multimedia devices are a 21st century triumph of silicon technology. People on treadmills with tiny iPods can choose from thousands of musical tunes, ranging from Mariah Carey's songs to Ludwig van Beethoven's symphonies. This scene would be hard to imagine, were it not for silicon ICs.

While the dominance of silicon ICs in data processing, communication, and multimedia will undoubtedly continue into the foreseeable future, there are growing efforts to utilize the power of silicon technology for new types

of applications. An important area is biology and medicine. Bioanalytical instruments are being miniaturized to make labs on a chip to perform a variety of experiments: to probe DNA, to monitor electrochemical activity, to examine neural functioning, and to actuate biological cells, for example. Microfluidic systems are being developed to provide a biocompatible environment on chips. By exploiting the power of silicon technology, one can combine CMOS ICs with microfluidic systems to make hybrid chips that perform standardized, and repeatable biological experiments more quickly, with a smaller sample volume, at lower costs than conventional approaches. Research activities in this new field, which we call **CMOS Biotechnology**, are expected to enjoy substantial growth. This trend is reflected by an increasing number of publications in major conferences and journals for IC design.

This book, consisting of ten technical topics contributed by experts in the field, will share some of the exciting developments in CMOS Biotechnology with readers from different disciplines. A large amount of high quality research is being done in this rapidly developing field, making it difficult to select only ten topics. Our selection presents examples of outstanding work to form view of CMOS Biotechnology.

We structured this book by sub-grouping the ten select topics into four parts, based on shared themes.

Part-I Microfluidics for Electrical Engineers (Chapters 2-4) presents an introduction to microfluidics for electrical engineers. Microfluidic systems serve as a biocompatible way to interface biological samples suspended inside them with CMOS chips below. Chapter 2 offers a tutorial on the theoretical foundations of microfluidics. Chapter 3 describes biological applications of microfluidic systems, with special attention to bioanalytical separation operations. Chapter 4 discusses the basic concept and fabrication of a CMOS/Microfluidic hybrid chip consisting of a CMOS IC with a microfluidic system fabricated on top.

Part-II CMOS Actuators (Chapters 5-7) offers examples that show how the CMOS/Microfluidic hybrid chip can be used to manipulate (control the motions of) biological samples ranging from cells to DNA. In Chapters 5 and 6, magnetic and electric manipulations of biological cells are discussed, along with examples of biomedical applications that such manipulation can enable. Chapter 7 describes electrical manipulations of biological objects of much smaller nanoscale size, including DNA for DNA hybridizations.

Part-III CMOS Electrical Sensors (Chapters 8-9) is a sensor counterpart to **Part-II**. Chapter 8 describes a microelectrode array integrated in a

CMOS chip, which can be used to record neural and cardiac cell activities whose signatures are carried by electrical signals. In Chapter 9, a brain-implantable neural recording system based on a CMOS chip is presented. The approach is similar to Chapter 8, but it is more focused on the circuits tailored for implanted sensor applications.

Part-IV CMOS Optical Sensors (Chapters 10-11) presents optical bio-sensing systems built on solid-state imager chips in combination with microfluidic systems. Chapter 10 demonstrates a high-resolution cell imaging experiment made possible by a charge-coupled device (CCD) connected with a microfluidic system. Chapter 11 discusses an example of how a CMOS imager can be utilized to study biological objects at molecular size scales and how it can potentially be exploited for applications like DNA sequencing.

2 INTRODUCTION TO FLUID DYNAMICS FOR MICROFLUIDIC FLOWS

Howard A. Stone

School of Engineering and Applied Sciences, Harvard University
has@deas.harvard.edu

2.1 INTRODUCTION

This book is evidence of the many fields and researchers who are interested in devices for manipulating liquids on small (micro and nano) length scales. This particular chapter has been written from the perspective that the reader will come from an electrical engineering or physics training and so will have only had limited, if any, previous exposure to fluids and their motion. As such, the discussion is necessarily focused on the fundamental concepts most relevant to understanding flows of liquids and gases in small devices. A discussion of current research ideas and trends is given in recent review articles [1, 2] and books [3, 4].

We begin by describing in Section 2.2 a few elementary physical ideas helpful for describing fluid motions in channel-like configurations. In particular, we introduce the fluid viscosity, describe qualitatively the incompressible flow approximation, sketch the most common velocity distributions in channel flows, and introduce the Reynolds number, which is a dimensionless parameter useful for characterizing different possible fluid motions. In Section 2.3 we mention briefly the well-known elementary laws of Ohm and Kirchhoff for electrical circuits and give their fluid analogs, which serve to introduce the notions of flow rate, pressure drop, and viscous resistance that are useful in the most basic characterizations of fluid motions. These introductory connections should hopefully assist the reader in developing physical intuition for simple fluid flows as well as thinking about

network-like ideas useful for design considerations in microfluidics. The partial differential equations for describing fluid flows are given and typical estimates are summarized in Section 2.4; these equations are frequently too difficult to solve analytically but there are now available commercial packages in computational mechanics that can numerically solve these equations for many practical configurations. A few elementary solutions relevant to microfluidics are given in Section 2.5.

The use of pipes and channels to convey fluids in an organized manner is essentially as old as the living world, since living systems have veins and arteries to transport water, air, gas, etc. The use of channels for the transport and mixing of gases and liquids is part of the industrial and civil infrastructure of our society and so, not surprisingly, many aspects of the dynamics of flow in channels are well understood. At the larger scales (e.g. length scales and typical speeds) of many common flows, the inertia of the motion is most relevant to the dynamics, and in this case turbulence is the rule: such flows are irregular, stochastic, dominated by fluctuations, and often require statistical ideas, correlations or large-scale numerical computation to quantify (if that is even possible).

The recent explosion of interest in fluid flows, and indeed their active manipulation and control, in micro- and nano-environments has turned attention to dynamics where viscous effects, which can be thought of, as a first approximation, as frictional influences interior to the fluid, are most significant: such flows are regular, reproducible, and generally laminar, which makes detailed control possible at small length scales. In many cases relatively simple quantitative estimates of important flow parameters are possible. Several examples are provided in this chapter.

We are all familiar with the concept of force. In mechanics it is generally important to speak in terms of the *stress* or force/area. The term *fluid* refers to either a liquid or a gas, or more generally any material that flows (the specialist might say "deforms continuously") in response to tangential stresses. It is best to think first about the flow of a single phase fluid in a channel. The most common way to create such a flow is to apply a pressure difference across the two ends of the channel: the resulting flow speed, or flow rate (volume per time), typically varies linearly with the applied pressure difference, at least at low enough flow speeds or Reynolds numbers, which is a dimensionless parameter introduced below. This flow may be used to transport some chemical species or suspended particles (e.g. cells). It is important to recognize that the velocity varies across the channel, with the highest speed at or near the center of the channel and the lowest speed (zero) at the boundaries. Because different points in the fluid move at dif-

ferent speeds there is a natural *dispersion* of suspended matter: a tracer put in the flow, for example, moves much faster in the middle of a channel than near the walls and so spreading of a tracer takes place at a rate controlled by the flow.

The potential opportunities to use microfluidic "plumbing" to create the lab-on-a-chip concept requires the integration of channels, valves, etc. in a systematic way that allows control. Two examples from the lab of Steve Quake, which make clear the distinct fluid bearing components that have been successfully integrated into microfluidic devices used for mixing and

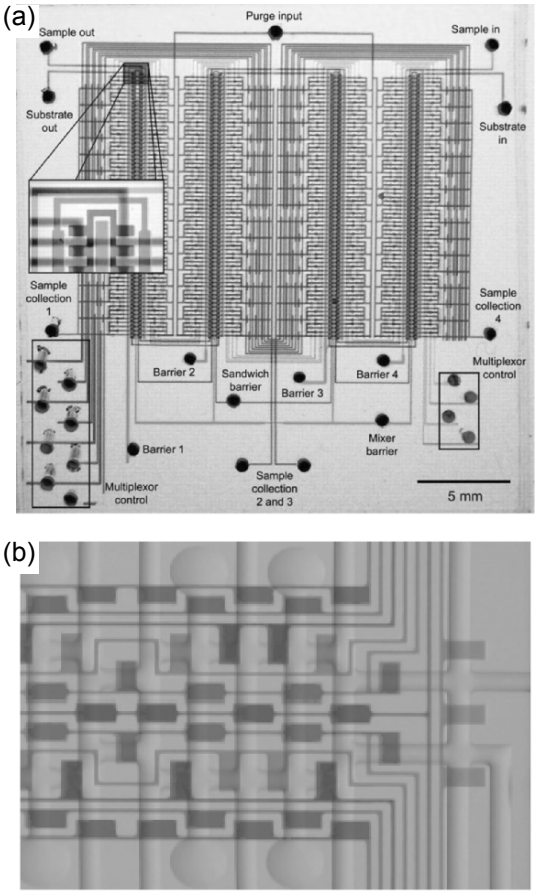


Figure 2.1 Large-scale integration of channels and valves for microfluidic systems. (a) Hundreds of channels and distinct chambers that have been integrated with more than two thousand valves [5]. Food dyes have been used to visualize the different channels and distinct chambers. (b) Example of large-scale integrated microfluidic system for measuring protein interactions [1]. Circular chambers are 250 μm in diameter. Figures courtesy of S. Quake.

reactions, are shown in Fig. 2.1; the different gray levels are dyes labeling distinct aqueous streams.

In other cases, two phases flow in a channel. In one variant of this kind of situation the two fluids are miscible (one fluid can dissolve in the other); see Fig. 2.2. Because the flows are generally laminar, in the absence of significant density variations, the two fluids can flow side-by-side down the channel which means that placement of the streams can be controlled and manipulated. For example, for these two-phase flows, chemical reactions between the two phases can be controlled and in cases such as shown in Fig. 2.2, the interface between the two fluids is the reaction zone [7]. Mixing common to turbulent flows, which is much faster than simple molecular diffusion, then does not take place. Instead, when mixing in a laminar flow in small devices is desired, some strategies need to be implemented [1, 2]. Recent research has made much progress in this area of microfluidic mixing [8].

In other two-phase flows the two fluids are immiscible: interfacial tension acts at the interface to minimize interface deformations. Here it is com-

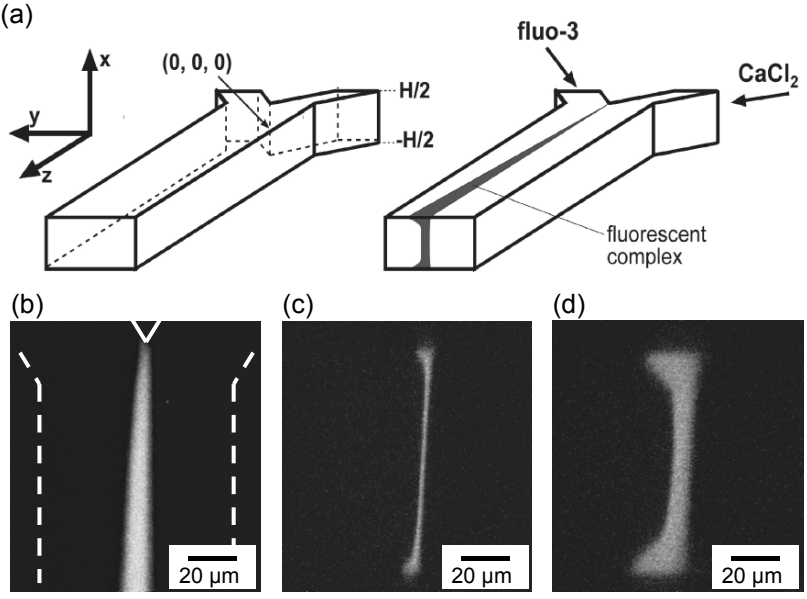


Figure 2.2 Flow of two miscible fluids along a rectangular microchannel [6]. (a) Schematic showing the region of interdiffusion (black). (b) Top view of an experiment using fluorescence to visualize the part of the flow where diffusion mixes the liquids. (c, d) Images taken with a confocal microscope at two different locations downstream, which show the diffusion between the two streams. The liquid near the wall has slower speeds so diffuses further near the wall than in the middle of the channel. These effects can be quantified [6].

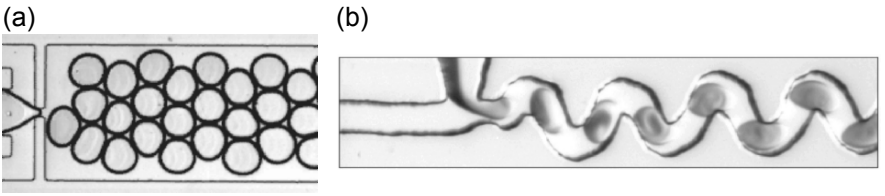


Figure 2.3 Formation of bubbles and drops in microdevices. (a) Formation of disk-shaped gas bubbles in a continuous liquid phase (e.g. [9]); figure courtesy of P. Garstecki. (b) Mixing and reaction inside of droplets, which serve as isolated chemical containers. The mixing is enhanced by the waviness of the channel [10]; figure courtesy of R. Ismagilov.

mon to disperse one phase as droplets in a continuous phase (see Fig. 2.3). The droplets may be used as small chemical reactors, can simply be plugs to separate distinct regions of a fluid column in order to minimize chemical dispersion in the continuous phase, or can be made into solid particles of a variety of size, shapes and compositions [11, 12]. The number of applications of these two-phase flows seems quite large and they are finding many uses in chemistry and biology as well as basic material science.

2.2 CONCEPTS IMPORTANT TO THE DESCRIPTION OF FLUID MOTIONS

2.2.1 Basic Properties in the Physics of Fluids

To introduce concepts needed to describe motion we note that the term fluid generally refers to either liquids or gases. One property needed to characterize a material is the *density* ρ, which measures the mass/volume. A second property important for understanding the flow response of a material is the *viscosity*. Fluids, like solids, can support normal forces without undergoing motion - the equilibrium fluid *pressure* measures the normal force under equilibrium conditions. Recall that the SI unit of pressure is the Pascal, which is defined as 1 N/m. For example, in a container of fluid at rest in a gravitational field, the pressure increases with vertical distance downward simply because of the mass of the column of material vertically above any position. In small devices, because of the small lengths scales involved, such vertical variations of *hydrostatic* pressure are generally not significant relative to other flow-associated stresses.

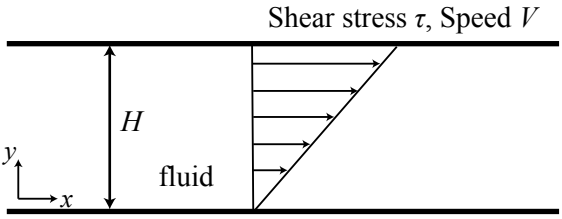


Figure 2.4 Shear flow: relative motion between two planes. A model experiment of this type is how the viscosity of a fluid is defined and measured.

2.2.2 Viscosity and the Velocity Gradient

In contrast, an ordinary (simple) fluid, such as air, water or oil, is set into motion whenever any kind of tangential force, or stress, is applied. The *viscosity* of a fluid measures the resistance to flow or the resistance to the rate of deformation. The simplest way to introduce the physical meaning of the viscosity is to consider the relative motion of two planar surfaces a distance H apart, as sketched in Fig. 2.4. A force/area, or tangential stress τ is needed to translate one surface at speed V relative to the other. In this case, the viscosity μ is the proportionality coefficient between τ and the shear rate V/H: $\tau = \mu(V/H)$. Notice that the dimensions of viscosity are mass/length/time. Water at room temperature has a viscosity $\mu_{\text{water}} \approx 10^{-3}$ Pa·sec. Typical cooking oils have viscosities 10-100 times that of water.

Once this idea is appreciated we note that in the situation considered in Fig. 2.4 it is more precise to introduce x and y coordinates, respectively, parallel and perpendicular to the boundary, and denote the shear stress along the upper boundary (whose normal is in the y-direction) as τ_{yx}. It is a fact of mechanics (unfortunate for the uninitiated) that the detailed description of the state of stress in a material requires two subscripts, one to indicate the direction of the normal to a surface and the second to indicate the direction of a force. In the simplest cases to think about for describing the viscosity (see Fig. 2.4), the velocity is directed parallel to the boundaries, $\mathbf{u} = u_x(y)\mathbf{e}_\mathbf{x}$, where $\mathbf{u}$ denotes the velocity field (it is a vector) and $\mathbf{e}_\mathbf{x}$ is a unit vector in the x-direction. The velocity is expected to vary linearly between the two surfaces, so $u_x = Vy/H$, and consequently, consistent with the definition in the preceding paragraph, the definition of viscosity can be written

$$\text{shear stress} = \frac{\text{force}}{\text{area}} = \tau_{yx} = \mu \frac{du_x}{dy}. \tag{2.1}$$

The *velocity gradient*, du_x/dy has dimensions 1/time and is frequently referred to as the *shear rate*. Equation (2.1) is a *linear* relation between the shear stress τ_{yx} and the shear rate, du_x/dy; materials that satisfy such linear relations are referred to as *Newtonian fluids*. For practical purposes, all gases and most common small molecule liquids, such as water and other aqueous solutions containing dissolved ions, are Newtonian. On the other hand, small amounts of dissolved macromolecules introduce additional molecular scale stresses – most significantly, these microstructural elements deform in response to the flow, and the description of the motion of the fluid, when viewed on length scales much larger than the macromolecules, generally requires nonlinear relations between the stress and the strain rate. These responses are termed *nonNewtonian*; dilute polymer solutions, biological solutions such as blood, or aqueous solutions containing proteins are in this category.

2.2.3 Compressible Fluids and Incompressible Flows

All materials are compressible to some degree; increasing the pressure usually decreases the volume. Nevertheless, when considering the motion of liquids and gases there are many cases where the density remains close to a constant value, in which case we refer to the "flow as incompressible." Such an incompressible flow is a significant simplification since we then take the density as constant for all calculations and estimates. It then follows that within the incompressibility assumption, the flows of liquids and gases are treated the same. It also follows that the value of the background or reference pressure plays no role in the dynamics other than setting the conditions where the fluid properties of density and viscosity are evaluated. There is only a need when describing fluid flows to distinguish liquids from gases when compressibility of the fluid is important.

A few additional comments may be helpful. It is almost always true under ordinary flow conditions where the pressure changes are modest (e.g. fractions of an atmosphere) that liquids can be treated as *incompressible*, i.e. the density can be taken as a constant. Even though gases are compressible in the sense that according to the ideal gas law their mass density ρ varies linearly with pressure, under common experimental conditions the pressure changes in the gas are sufficiently small that density changes in the gas are usually also small: again, we have the approximation of an incompressible flow. The one case of gas flows where more care is needed is when microchannels are sufficiently long that a gas flow is accompanied by a significant pressure change (say a 20% change in pressure); then the density will also

change by approximately this amount. Unless otherwise stated below we will assume that the motions of the fluids occur under conditions where the incompressible flow approximation is valid.

2.2.4 The Reynolds Number

For the simplest qualitative description of a fluid motion we need to recall Newton's second law: the product of mass and acceleration, or more generally the time rate of change of its linear momentum, equals the sum of the forces acting on the body. When we apply this law to fluids it is convenient to consider labeling some set of material points (imagine placing a small amount of dye in the fluid) and following their motion through the system. In addition, we can think about the hydrodynamic pressure as acting to either accelerate the fluid elements (i.e. overcome the inertia) or to overcome friction (viscosity) to maintain the motion. It then follows that the mechanical response to pressure forces that cause flow depends on the relative magnitude of the inertial response to the viscous response; this ratio is known as the *Reynolds number*.

To quantify this idea, consider some body (say a sphere) with radius ℓ translating with speed U through a liquid with viscosity μ and density ρ. The typical acceleration of fluid moving around the object is $U/\Delta t$, where $\Delta t \approx \ell/U$ is the typical time over which changes occur in the fluid when the body moves. Then, we compare (the symbol O simply indicates the order of magnitude)

$$\frac{\text{mass} \cdot \text{acceleration}}{\text{viscous forces}} = \frac{O(\rho \ell^3 U/(\ell/U))}{O(\mu U/\ell^2)} = \frac{\rho U \ell}{\mu} = \mathcal{R} = \text{Reynolds number}. \tag{2.2}$$

The reader can verify that the Reynolds number $\mathcal{R}$ is dimensionless, i.e. it has no dimensions.[1]

The Reynolds number is a *dimensionless* parameter that is useful for characterizing flow situations; it is not simply a property of the fluid but rather combines fluid properties (ρ and μ), geometric properties (a length scale ℓ)

[1] The Reynolds number is named after Osborne Reynolds (1842-1912), a professor at Manchester University, who introduced this ratio of variables in 1883 when characterizing the different observed motions for pressure-driven flow in a pipe. The parameter was apparently named the Reynolds number some years later by the German physicist Arnold Sommerfeld. For some historical remarks the reader can refer to [13].

and the typical flow speed. Roughly speaking, high Reynolds number flows tend towards turbulence, they have large domains that tend to be in uniform motion (possibly in a statistical sense) and narrow boundary layers where viscous effects are especially important and viscous stresses are large. Low Reynolds number flows are laminar. In most small devices, because the length scale is measured in tens of microns or less, and the speeds are typically tens of cm/sec, then even for water ($\mu = 10^{-3}$ Pa·sec), the Reynolds number tends to be on the order of unity or smaller. For example, with these numbers $\mathcal{R} = [10^3\ (\text{kg/m}^3) \times 0.1\ (\text{m/sec}) \times 10^{-5}\ (\text{m})]/[10^{-3}\ (\text{Pa·sec})] = 1$.

As a final remark we note that as the Reynolds number is increased from small to large there is a typical progression in "complexity" of a flow. For example, for flow past objects the laminar flow first develops a downstream wake of increasing length, before there is a transition to turbulence, i.e. stochastically fluctuating velocities in at least some region of the flow. For channel flows the laminar flow becomes suddenly turbulent if the Reynolds is large enough. This transition Reynolds number is known to depend on the magnitude of fluctuations in the inlet flow, but for common lab conditions is on the order of 2000 (based on the mean velocity and the diameter) for flow in a circular pipe.

2.2.5 Pressure-driven and Shear-driven Flows in Pipes or Channels

We have already mentioned that a detailed examination of the velocity distribution shows that it varies across a channel. As sketched in Fig. 2.5, the two typical cases are a shear flow, caused by relative movement of two surfaces, and a flow driven by a pressure gradient. In a shear flow the velocity varies linearly across the channel. In a pressure-driven flow in a cylindrical

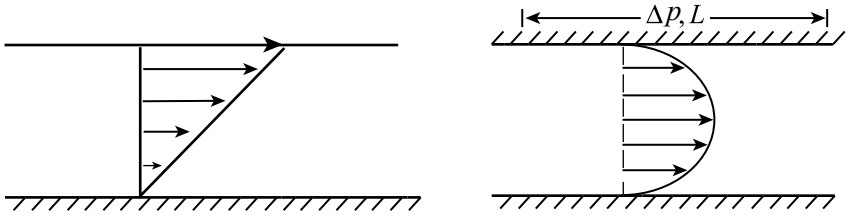


Figure 2.5 (Left) Shear flow driven by a moving boundary. The velocity varies linearly with position perpendicular to the flow direction. (Right) Flow driven by an applied pressure gradient. The velocity varies parabolically with position perpendicular to the flow direction. This situation is commonly called Poiseuille flow.

tube or between two parallel plates the velocity varies parabolically across the channel, with a maximum velocity in the center of the channel (see Subsection 2.5.1). For a rectangular cross section, the velocity distribution is somewhat more complicated but the basic velocity distribution is parabolic across the shortest dimension (see Subsection 2.5.2). Elementary textbooks give the details of the most common laminar flow configurations.

2.3 ELECTRICAL NETWORKS AND THEIR FLUID ANALOGS

2.3.1 Ohm's and Kirchhoff's Laws

The first elementary rule of circuit design is Ohm's law, which relates the change in electrical potential ΔV to the current I in the familiar form $\Delta V = IR$. The resistance R is dependent on the system, the materials, etc. The second rule involves charge conservation, which for steady states gives Kirchhoff's law for a node of a circuit where N currents I_m meet: $\Sigma_{m=1}^{N} I_m = 0$.

The above description by way of algebraic relations for characterizing elementary circuits has analogs when transport of fluid in microchannels is considered (See Table 2.1 for summary). For now we consider single-phase flow. Rather than charge, potential drop, and resistance, for example, as basic

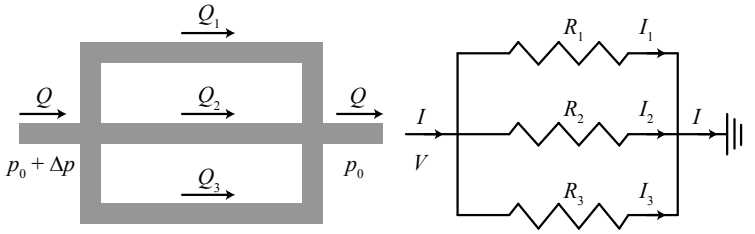


Figure 2.6 Pressure-driven flow in a network of parallel channels. The pressure at the inlet is Δp higher than the pressure at the exit. (Left) Schematic of the flows rates Q_i in each of three channels in parallel. The individual flow rates are related to the total flow rate by $Q = Q_1 + Q_2 + Q_3$ and the pressure drop Δp and Q are related by $\Delta p = QR_H$, where R_H is the equivalent hydrodynamic resistance. In this case, since the same pressure drop acts across each channel, then for each element $\Delta p = Q_i R_{Hi}$. (Right) The equivalent representation as an electrical circuit with resistances in parallel.

parameters, in micro-fluidics we keep track of the rate of mass (or more commonly volume) of fluid transported, a measure of the force needed to move the fluid (e.g. a change of pressure), and the viscosity of the fluid, which is a measure of the resistance of the fluid to motion. In many circumstances the motion of fluids is sufficiently simple that linear relation may be applied betweeen the forcing (e.g. pressure) and the output, e.g. the flow rate of liquid. Hence, given a pressure difference Δp across a fluid-filled channel, there is a flow rate Q (volume/time) with $\Delta p = QR_H$, where the hydrodynamic resistance R_H is a function of the geometry of the channel and the viscosity of the fluid in the channel. At sufficiently high Reynolds number, turbulence is expected and then the resistance R_H is also a function of the average fluid velocity or the Reynolds number.

Table 2.1 Comparing the basic relations for electrical systems.

Electrical relation	Fluid mechanical relation
$\Delta V = IR$	$\Delta p = QR_H$
$\sum_{k=1}^{N} I_k = 0$	$\sum_{k=1}^{N} Q_k = 0$
Power: $I\Delta V = I^2 R$	Power: $Q\Delta p = Q^2 R_H$
$\mathbf{j} = \sigma \mathbf{E}$ (Ohm's law)	$\mathbf{u}_{\text{avg}} = -\frac{k}{\mu}\nabla p$ (Darcy's law)
$R = \frac{\sigma^{-1} L}{\pi a^2}$	$R_H = \frac{8\mu L}{\pi a^4}$

For electrical systems, ΔV is the potential difference, I is the current and R is the resistance. With single-phase laminar flow, Δp is the pressure difference, Q is the volumetric flow rate, and R_H is the hydrodynamic resistance. The power is used in the usual way as the energy/time that is dissipated, so equals I^2R for an electrical circuit and Q^2R_H for a fluid circuit. In the next to last row, we first give Ohm's law in the form relating the current density $\mathbf{j}$ to the electrical field $\mathbf{E}$ (here σ is the conductivity of the material) while the corresponding flow description relating the average velocity $\mathbf{u}_{\text{avg}}$ to the pressure gradient is known as Darcy's law; it is an approximation valid for low Reynolds number motions. For this description we also introduce the permeability k which, for example, depends on the detailed cross-sectional shape of a channel. In the last row of the table we consider a circular element of radius a, length L, and resistivity σ^{-1}, and a circular pipe of radius a and length L, which is filled with fluid of viscosity μ. The electrical resistance R has a similar though not identical functional form to the hydrodynamic resistance R_H.

2.3.2 Channels in Parallel or in Series

Consider N elements with individual resistances R_m, $m = 1,\ldots, N$. In the two simplest situations the resistances are either in series or in parallel, as is familiar from elementary courses:

- When the elements are placed in series, the effective resistance R_{eff} of the combination is $R_{\text{eff}} = \sum_{m=1}^{N} R_m$.
- In contrast, when the elements are placed in parallel, the effective resistance is $R_{\text{eff}} = 1 \Big/ \sum_{m=1}^{N} R_m^{-1}$.

The same results apply to fluid circuits except that the hydrodynamic resistances R_H are used instead of the electrical resistances R.

2.3.3 Resistances in terms of Resistivities, Viscosities and Geometry

The electrical and fluid analogs extend to more microscopic pictures. The "local" version of Ohm's law relates the current density $\mathbf{j}$ (a vector) to the local electric field $\mathbf{j} = \sigma\mathbf{E}$, where σ is the conductivity and $\mathbf{E}$ is the local electric field. Thus, in a homogeneous circular wire of radius a and length L, the electrical resistance is $R = \sigma^{-1}L/(\pi a^2)$, where we have written this expression using the resistivity σ^{-1}.

As will be shown below, the corresponding hydrodynamic resistance, $R_H = \Delta p/Q$, which relates the pressure drop to the volumetric flow rate, has a similar form, $R_H = c\mu L/a^4$, where $c = 8/\pi$. The only significant difference in the fluid case is the dependence on the fourth power of the radius rather than the second power of the radius that occurs in the electrical case, which results from the distribution of velocities across the channel (see Fig. 2.5). The variation with the fourth power of the radius obviously will have significant influence when designs of significantly different radius are considered. As stressed in most elementary fluid mechanics treatments, the dependence of R_H on the fourth power of the radius means that a 10% change in radius produces approximately a 40% change in flow for a given pressure drop (see also Subsection 2.5.1).

Rectangular channels are common in microfluidic configurations simply because these are formed naturally by most existing fabrication methods. For a rectangular channel of height h and width w, with $h < w$, then to a good approximation $R_H = c_r\mu L/(h^3 w)$, where $c_r = 12$. In this case, the variation of

the hydrodynamic resistance with the third power of the height is to be kept in mind when design is considered.

To be more specific, consider a water-filled rectangular microfluidic channel with height and width 20 μm. Using the above results, if the channel is 5 mm long and the pressure drop is 5 psi (≈ 1/3 atm), the hydrodynamic resistance is $R_H \approx 3/8 \cdot 10^{15}$ kg/m^4·sec. The corresponding flow rate is about 10^{-10} m^3/sec ≈ 0.1 μl/sec, with a typical velocity 25 cm/sec. The corresponding Reynolds number is $\mathcal{R} = \rho u \ell / \mu = 5$ which is much smaller than the Reynolds number (2000) for transition from laminar to turbulent flow. If the cross-sectional dimensions are reduced by a factor of 2, the hydrodynamic resistance increases by 2^4, the flow rate decreases by 2^4, the average velocity decreases by 2^2, and the Reynolds number decreases by 2^3.

2.4 BASIC FLUID DYNAMICS VIA THE GOVERNING DIFFERENTIAL EQUATIONS

2.4.1 Goals

In this subsection we provide a brief survey of the most important quantitative ideas from fluid dynamics, starting with the basic differential equations, that are needed for describing flows in small devices. Of course, there are many undergraduate and graduate texts available, and even specialized texts for microfluidics. We have selected mathematical material with the view towards a stepwise, yet terse, description of fundamental ideas that recur frequently when characterizing microflows. In particular, we use the familiar continuum description of materials and speak of the velocity, pressure, and stress fields. Our goal is to give the reader familiarity with the concepts and equations with which fluid motions are described, solved for using computational packages, and estimated using order-of-magnitude arguments.

The mathematical description is familiar both at the level of conservation laws for charge and current, and at the scale of transport processes such as thermal conduction or molecular diffusion. Here the basic equations needed for studying the fluid motion are given with emphasis on incompressible flows where the fluid density can be treated as a constant, which is applicable to the overwhelming majority of flows on small scales (even most cases when gases are considered). As is common in engineering and physics, dimensionless parameters play an important role in characterizing problems.

Some elementary flows representative of motions in microdevices are given in Section 2.5.

2.4.2 Continuum Descriptions

Although our subject will be micro- and nanofluidics, our basic starting point for estimates and calculations will be the familiar continuum equations from classical physics. Why should these equations apply at the "small scales" of microdevices? The answer is simply that the equations represent the average of what all the molecules in the fluid are doing, and, if we consider a liquid in a cubic volume with side of length just five times that of a molecule, then there are already more than 100 molecules: averages computed with statistics based on a hundred objects are usually pretty good.

In order to take this idea one step further and to appreciate why the continuum equations are quite reasonable, even at such small scales, imagine computing some average extensive material property f based on N molecules in a small measuring volume. We then expect $f \propto N$. Because of thermal fluctuations there will be variations δN of the number of molecules in the measuring volume with $\delta N \propto N^{1/2}$ (this is a standard result in statistical physics). Hence, the corresponding variation of the average extensive property is $\delta f \propto \delta N$. Next compare the fluctuations with the average value. We see that

$$\frac{\delta f}{f} = \frac{\delta N}{N} = N^{-1/2}. \tag{2.3}$$

Thus, for the fluctuations to be small compared to the mean value, i.e. $\delta f/f \ll 1$, we need the number of molecules to exceed $N > (f/\delta f)^2$. In particular, fluctuations are less than 10% of a measured property, when $\delta f/f \approx 0.1$ or by Eq. (2.3) when $N > 100$. This number of molecules can be found in a cube with only 4-5 molecules on an edge, which is a very small length scale indeed, and justifies the characterization given in the first paragraph of this subsection. For example, a water molecule has a typical dimension of only a few angstroms so five water molecules corresponds to a length smaller than 2 nm. We see that it should not be surprising that continuum calculations often provide reasonable estimates even down towards the molecular scale.

As a final way to introduce the continuum description, one can imagine measuring a fluid property (Fig. 2.7). We distinguish molecular and intermolecular dimensions, ℓ_{mol}, a length scale over which a measuring devices makes a local measurement of the system, ℓ_{avg}, and distances over which

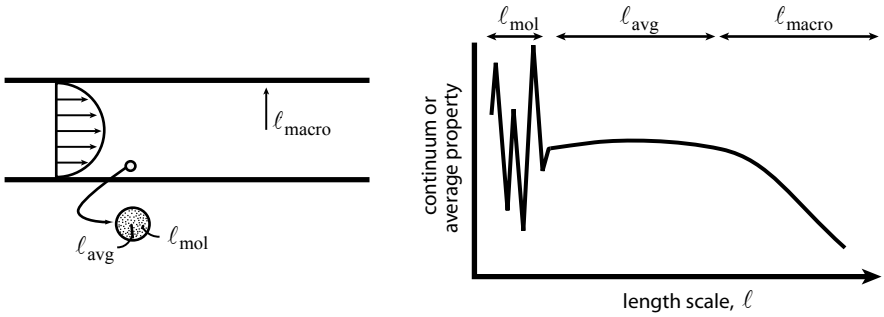


Figure 2.7 Different length scales in a flow when considering the continuum description.

variations are measured, ℓ_{macro}. The continuum variables measured on the scale ℓ_{avg} are such that $\ell_{mol} < \ell_{avg} < \ell_{macro}$.

In such a continuum description all field variables, such as velocity and pressure, are function of position $\mathbf{x} = (x, y, z)$ and time t, i.e. $\mathbf{u}(\mathbf{x}, t)$ and $p(\mathbf{x}, t)$. These variables are related by partial differential equations obtained from mass, momentum and energy considerations in the fluid.

2.4.3 The Continuity and Navier-Stokes Equations

2.4.3.1 Continuity - Local Mass Conservation

We know from courses in basic physics and chemistry that all materials are compressible to some degree (gases much more so than liquids or solids). Thus, for the most general case of motion of a material the velocity ($\mathbf{u}$) and the density (ρ) variations are linked, and this is expressed by the *continuity equation*:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \tag{2.4}$$

where the gradient operator, $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$.

Here we will primarily be concerned with fluid motions that are essentially incompressible, as introduced qualitatively in Subsection 2.2.3. Effectively, this means that the pressure variations that accompany the flow create insignificant density changes ($\delta\rho$), i.e. $\delta\rho/\rho \ll 1$. Hence, we treat the density of the fluid as constant. Justification for this approximation in terms of features of the flow are given in Subsection 2.4.5. From Eq. (2.4) we then see that the velocity field for incompressible motions satisfies

$$\nabla \cdot \mathbf{u} = 0. \tag{2.5}$$

In Cartesian coordinates $\mathbf{u} = (u_x, u_y, u_z)$, the continuity equation is written $\partial u_x/\partial x + \partial u_y/\partial y + \partial u_z/\partial z = 0$, which may be considered a constraint on the allowed form of the velocity variations.

2.4.3.2 The Navier-Stokes Equation - a Linear Momentum Balance

Here we only consider the form of the linear momentum statement for incompressible fluid motions of a Newtonian fluid; a Newtonian fluid is where the stress is linearly related to the rate-of-strain in a generalization of the simplified discussion given in Subsection 2.2.2. In this case, the velocity vector **u** and pressure p are related by the *Navier-Stokes equation* (here we are assuming that the viscosity is constant as well)

$$\rho\left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u}\right) = -\nabla p + \mu\nabla^2\mathbf{u} + \mathbf{f}_b, \tag{2.6}$$

where $\mathbf{f}_b$ represents body forces (per unit volume) that act on the fluid. The physical interpretation of this equation is that the left-hand side of Eq. (2.6) refers to the product of mass and acceleration (per unit volume) when following a fluid element (think of a dyed piece of fluid) moving along with the flow. The right-hand side of Eq. (2.6) corresponds to all of the forces acting on the fluid element. For a good physical discussion of the basic equations and a wide range of fluid flow phenomena, the reader is referred to [14].

The most common body force in most fluid systems is the gravitational force, though for most microfluidic (and smaller) flows this body force is generally negligible. Electric fields can also apply forces and we refer the reader to review articles [1, 2]. Note that in writing Eq. (2.6) we have also assumed that the fluid viscosity μ is constant. In some microfluidic flows, temperature variations occur and these can cause significant variations in μ, which can modify the velocity profiles.

Equations (2.5) and (2.6) represent four equations for the four unknowns of the velocity vector and the pressure. These equations are difficult to solve analytically for all but the simplest geometries, some of which fortunately arise in microfluidics (see Section 2.5). For example, when there is incompressible flow along a straight channel with uniform cross-sectional shape, then away from an entrance or an exit the nonlinear term in Eq. (2.6) $\mathbf{u}\cdot\nabla\mathbf{u} = \mathbf{0}$; the elimination of the nonlinear term is a significant simplification and many analytical results are available. Even then, it is worth noting that

steady solutions for simple geometries are only realistic if the flow speeds are sufficiently small (as measured by the Reynolds number), since otherwise the steady solutions are unstable to small perturbations and evolve to turbulent states. It is also now feasible to solve numerically these equations (several commercial packages exist), even for complicated three-dimensional geometries, at least in those cases where the Reynolds number is small enough that there is a stable steady flow solution. Cases there the geometry is complicated and the flow is continually evolving in time are generally still subjects of research and code development.

In order to solve the Navier-Stokes equations, it is necessary to impose boundary conditions. Most commonly the velocity is zero on all stationary boundaries, which is referred to as the no-slip boundary condition, and the pressure, velocity distribution or flow rate are prescribed on the two ends of the device.

2.4.4 The Reynolds Number

A useful characterization of any flow is obtained by comparing the ratio of the typical inertia of the fluid motion – the left-hand side of Eq. (2.6) – to the viscous terms – the second term on the right-hand side. This ratio is referred to as the Reynolds number $\mathcal{R}$, which was introduced in Subsection 2.2.4. For flows with a typical speed u, in a geometry with typical channel height h (assumed to be smaller than the width), simple dimensional estimates applied to Eq. (2.6) yield

$$\mathcal{R} = \frac{O(\rho \mathbf{u} \cdot \nabla \mathbf{u})}{O(\mu \nabla^2 \mathbf{u})} = \frac{\rho u^2/h}{\mu u/h^2} = \frac{\rho u h}{\mu}. \tag{2.7}$$

In many microfluidic applications the Reynolds number is not too large. For example, consider the flow in a device with $h = 100$ μm, $u = 1$ cm/sec, density comparable to water and a viscosity $\mu = 10\mu_{\text{water}}$. Then, $\mathcal{R} \approx 0.1$ and we should expect inertial influences not to be significant.

The above estimates in fact are rather conservative. In common microfluidic geometries there is a channel length $\ell \gg h$ over which the fluid flows, often changing the velocity at least in part because the channel height or width changes. In these steady flow cases a better estimate of the effective Reynolds number $\mathcal{R}_{\text{eff}}$ of the flow, based on the ratio of the inertial to viscous terms is

$$\mathcal{R}_{\text{eff}} = \frac{O(\rho \mathbf{u}\cdot\nabla\mathbf{u})}{O(\mu\nabla^2\mathbf{u})} = \frac{\rho u^2/\ell}{\mu u/h^2} = \frac{\rho u h}{\mu}\frac{h}{\ell} = \mathcal{R}\frac{h}{\ell} \tag{2.8}$$

With $h/\ell = 1/50$, such an effective Reynolds number is small even if the usual Reynolds number $\mathcal{R} = 10$. This kind of estimate is useful in so-called "lubrication" configurations, which are flows that are nearly unidirectional and parallel to the bounding surfaces.

In those cases where the Reynolds number is small – formally we consider $\mathcal{R} \ll 1$, but in practice often $\mathcal{R} < 1$ is sufficient – we neglect inertial terms all together. If the forcing of the flow is steady then we simplify Eq. (2.6) to the Stokes equations:

$$0 = -\nabla p + \mu\nabla^2\mathbf{u} + \mathbf{f}_b \quad \text{and} \quad \nabla\cdot\mathbf{u} = 0. \tag{2.9}$$

In this case the only length scales that matter are geometric. Equation (2.9) is the starting point for many detailed calculations of micro- and nanoflows. As a final remark for this viscously dominated flow case, the typical order of magnitude of pressures and stresses are $O(\mu u/\ell)$ where ℓ is the smallest dimension in the flow.

2.4.5 Brief Justification for the Incompressibility Assumption

In any given flow situation we have to ask under what conditions is $\delta\rho/\rho \ll 1$? The thermodynamic state of a fluid relates the density, pressure and temperature. For isothermal conditions, $\rho(p)$ and then $\delta\rho \approx \delta p(d\rho/dp) = dp/c^2$, where c is the (isothermal or adiabatic) sound speed, which is about 10^3 m/sec for small molecule liquids.

In high-speed flows commonly discussed in first-year physics or undergraduate fluid mechanics, the pressure-velocity relation is algebraic and given simply by the Bernoulli relation ($p + \frac{1}{2}\rho\mathbf{u}^2 = $ constant, neglecting gravity and viscous effects), so that pressure changes vary quadratically with the typical flow speed. Hence, the assumption of incompressibility requires

$$\frac{\delta\rho}{\rho} \approx \frac{\delta p}{\rho}\frac{d\rho}{dp} = \frac{u^2}{c^2} = \mathcal{M}^2 \ll 1, \tag{2.10}$$

where the dimensionless ratio $\mathcal{M} = u/c$ is the *Mach number*. Even speeds of liquids about 10 m/sec, which are much faster than those found in typical small devices, have $\mathcal{M} \ll 1$, so density variations can be neglected.

Much more common in small devices are lower Reynolds number flows. In such flows the pressure changes occur owing to viscous effects in the liquid. In this case, in a device with typical dimension ℓ (the small geometric dimension that impacts the flow), $\delta p \approx \mu u/\ell$. Hence, the assumption of incompressibility requires

$$\frac{\delta\rho}{\rho} \approx \frac{\delta p}{\rho}\frac{d\rho}{dp} = \frac{\mu u}{\rho \ell c^2} = \frac{\mathcal{M}^2}{\mathcal{R}} \ll 1, \tag{2.11}$$

where $\mathcal{R} = \rho u \ell/\mu$ is the Reynolds number. Since most microfluidic flows occur with conditions such that $u < 1$ m/sec then $\mathcal{M}^2 \approx 10^{-6}$ which is much smaller than the usual Reynolds numbers of the flows and again density variations can be neglected.

In other words, in almost all cases of interest in micro- and nanofluidics, the fluid flows can be treated as incompressible. The only caveat worth mentioning, as discussed briefly above in Section 2.2, is the case of gas flows in long channels where the pressure change is sufficiently large that the gas density, which varies in proportion to the pressure, needs to be taken into account.

2.5 MODEL FLOWS

There are some simple geometries where the Navier-Stokes equations can be solved exactly. Since some of the solutions for these laminar flows are useful for estimating flow speeds, shear rates, or the effect of a change in geometry, we give a few results, largely without derivation, in the hope that readers will find them helpful.

2.5.1 Pressure-driven Flow in a Circular Tube

Consider steady pressure-driven flow in a circular tube of radius a. Using cylindrical coordinates (r, z), the velocity distribution $u(r)$ is *parabolic* (recall Fig. 2.5) with the form

$$u(r) = \frac{1}{4\mu}\left|\frac{dp}{dz}\right|\left(a^2 - r^2\right), \tag{2.12}$$

where dp/dz is the axial pressure gradient, which is a constant for this flow. The corresponding flow rate Q (volume/time) is then

$$Q = 2\pi\int_0^a u(r) r \, \mathrm{d}r = \frac{\pi a^4}{8\mu}\left|\frac{dp}{dz}\right|. \tag{2.13}$$

The dependence of the flow rate on the *fourth* power of the radius has a significant impact on flow in small systems. For the same pressure gradient, a factor of two reduction in radius produce a 16-fold reduction in flow rate. To appreciate further how this small effect has such a big influence, consider flow in small blood vessels: a 10% decrease in the radius (i.e. due to eating too much fatty foods!) produces more than a 40% decrease in the flow rate of blood.

The moral here is that small changes in geometry can have a big impact on flow in micro- and nanoscale devices. In fact, poor comparison of theory and experiment in some studies with small geometries led some authors to conclude that the Navier-Stokes equation did not apply. However, more careful studies have concluded that, in large part because of the significant effect of small changes in scale, very careful experiments are needed and then theory based on the Navier-Stokes equation and experiment are in excellent agreement [15].

It is also convenient to work in terms of an average velocity, $U = \frac{Q}{\pi a^2} = \frac{a^2}{8\mu}\left|\frac{dp}{dz}\right|$. Recall the definition of the Reynolds number: $\mathcal{R} = \frac{\rho U a}{\mu}$. For a fixed pressure gradient, $\mathcal{R} \propto a^3$ and so a factor of two change in radius produces a factor of eight change in Reynolds number.

Since the average velocity is proportional to the pressure gradient it is convenient to write the flow field in vector form as

$$\frac{\mu \mathbf{U}}{k} = -\nabla p \tag{2.14}$$

where k is known as the permeability (k has dimensions length^2). We thus see that the permeability of a channel of radius a is $k = a^2/8$. Equation (2.14) is referred to as Darcy's law (see Table 2.1); the basic steps at the origin of this linear relation between velocity and pressure gradient can be traced to the lack of an effect of inertia in these laminar flows. As a general point, the

order-of-magnitude of the permeability of a uniform system is typically the square of the smallest dimension.

In Section 2.3 we discussed the hydrodynamic resistance of an element of a channel. With the above results we see that the hydrodynamic resistance, which is defined as $R_H = Q^{-1}\Delta p$, where $-dp/dz = \Delta p/L$, leads to $R_H = 8\mu L/(\pi a^4)$ (see Table 2.1).

The shear rate (often denoted $\dot{\gamma}$) at the wall is important to estimate. Using Eq. (2.12) we see that

$$\dot{\gamma}_{\text{wall}} = \left|\frac{du}{dr}\right|_{r=a} = \frac{4U}{a}. \tag{2.15}$$

The typical dimensional idea that the shear rate is the maximum centerline velocity divided by the distance from the centerline to the wall should typically be within a factor of two of the actual value for any geometry simply because cross-sectional velocities for most geometries tend to be parabolic.

We make two further remarks that involve physical and dimensional ideas.

- The basic results in this subsection could have been anticipated using only the characteristic scale for velocity established when making the problem statement non-dimensional. For example, consider the dimensions representative of Eq. (2.9). The magnitude of the velocity in laminar pipe flows is $u_c = a^2\Delta p/(\mu L)$ and so the typical volume flow rate is $Q \approx u_c a^2 \approx a^4\Delta p/(\mu L)$. Note that we have arrived at physical conclusions (e.g. $Q \propto a^4\Delta p$) regarding the order of magnitude of quantities without actually having solved the differential equation.
- The fluid motion arises as a balance between the pressure drop driving motion and the frictional (viscous) resistance from the bounding walls. This balance applies independent of the cross-sectional shape of the channel.

2.5.2 Pressure-driven Flow in a Rectangular Channel

Most microfluidic fabrication methods lead to channels with rectangular or nearly rectangular cross sections. Here we summarize the main flow features using the same concepts as described in the previous subsection. First, we consider the simplest case, which is a channel much wider than the height, $h \ll w$. For pressure-driven flow where a pressure drop Δp acts over a length L, the velocity profile across the channel is parabolic:

$$u(y) = \frac{\Delta p}{2\mu L}\left(\left(\frac{h}{2}\right)^2 - y^2\right). \tag{2.16}$$

The corresponding flow rate in a channel of width w is then approximately

$$Q = w\int_0^h u(y)\,\mathrm{d}y = \frac{wh^3\Delta p}{12\mu L}, \tag{2.17}$$

with an average velocity

$$U = \frac{h^2\Delta p}{12\mu L}. \tag{2.18}$$

The corresponding permeability is $k = h^2/12$ and the hydrodynamic resistance is $R_H = 12\mu L/(wh^3)$.

For the general case, the axial velocity in the cross-sectional (x, y) plane is known analytically in terms of a Fourier series:

$$\begin{aligned} u(x,y) &= \frac{\Delta p}{2\mu L}\left\{\left[\left(\frac{h}{2}\right)^2 - y^2\right] - \sum_{n=0}^{\infty} a_n \cos\left(\tfrac{\lambda_n y}{h/2}\right)\cosh\left(\tfrac{\lambda_n x}{h/2}\right)\right\} \\ \lambda_n &= \frac{(2n+1)\pi}{2}. \end{aligned} \tag{2.19}$$

The coefficients a_n follow from the no-slip boundary conditions at x = ±w/2, which leads to $a_n = h^2(-1)^n/[(\lambda_n)^3\cdot\cosh(\lambda_n w/h)]$. The corresponding flow rate is given by

$$\begin{aligned} Q &= 4\int_0^{w/2}\int_0^{h/2} u(x,y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \frac{wh^3\Delta p}{12\mu L}\left[1 - 6\left(\frac{h}{w}\right)\sum_{n=0}^{\infty}\lambda_n^{-5}\tanh\left(\frac{\lambda_n w}{h}\right)\right]. \end{aligned} \tag{2.20}$$

This equation can be evaluated for $0 \le h/w \le 1$ and in fact the result is nearly linear in the aspect ratio h/w. Consequently, a reasonable approximation is

$$\frac{12\mu LQ}{wh^3\Delta p} = 1 - \frac{6(2^5)}{\pi^5}\frac{h}{w} \quad \text{(less than 10\% error for } h/w \le 0.7\text{)}. \tag{2.21}$$

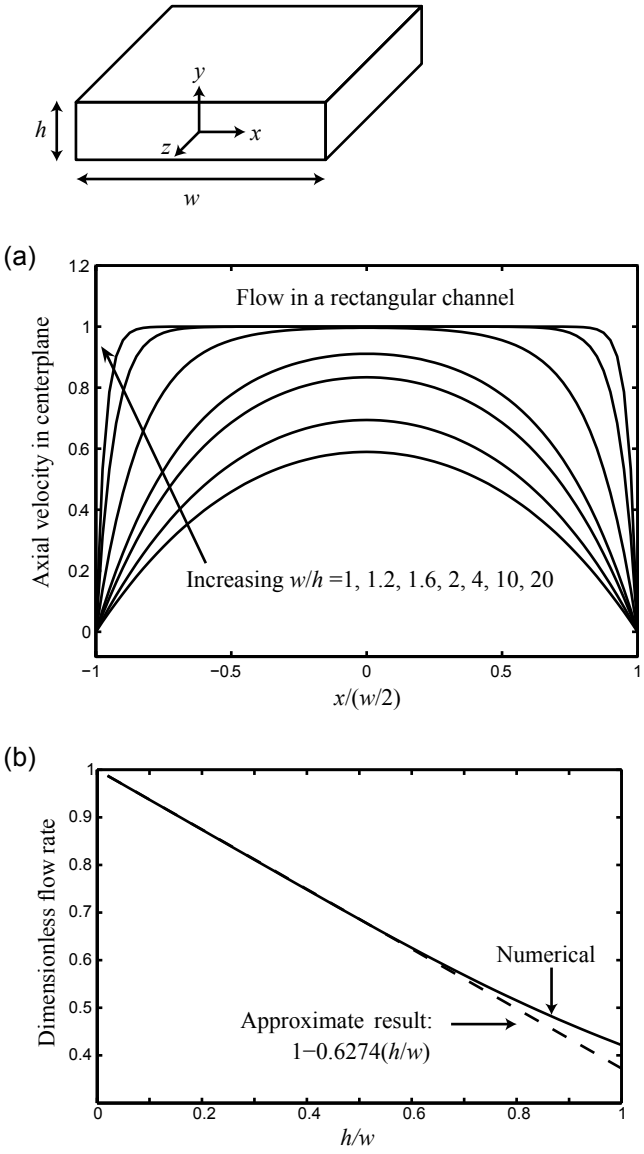


Figure 2.8 (a) Axial velocity distribution evaluated along the center-plane ($y = 0$), when viewed from above, in a channel with rectangular cross-section. The velocity has been scaled proportional to the average velocity. Note that as $w/h > 2$ the velocity distribution in the center of the channel becomes increasingly flat. (b) Dimensionless flow rate $(12\mu LQ)/(wh^3\Delta p)$ as a function of the channel aspect ratio, w/h.

To get a physical feel for the kinds of speeds and flow rates to expect, consider a water-filled rectangular channel of height $h = 20$ μm, width $w = 50$ μm, and length 1 cm. The viscosity of water is $\mu_{\text{water}} \approx 10^{-3}$ kg/m/sec. If a pressure drop of 1 atm (10^5 Pa) is applied then we find $U \approx 33$ cm/sec, $\mathcal{R} \approx 7$, and $Q = 33\times10^{-4}\,\text{cm}^3/\text{sec} = 0.3$ μl/sec.

There is an important feature of the velocity distribution to appreciate when considering rectangular cross sections with $h \ll w$. In the *shortest* dimensions, which is the height, the velocity distribution is parabolic. However, what is the distribution of the average velocity when viewed from above? In this case, across the width of the cross section the velocity distribution is largely unchanged, except for regions $O(h)$ near the walls where the fluid velocity has to change to satisfy the no-slip conditions. The evolution of the velocity distribution as the aspect ratio of a rectangular channel is changed is shown in Fig. 2.8: when the cross section is square the velocity distribution is very nearly parabolic and as the width increases then the velocity distribution across the width gets progressively flatter.

2.6 CONCLUSIONS AND OUTLOOK

The purpose of this chapter was to introduce some of the basic ideas and concepts of fluid dynamics as they frequently appear in applications involving micro- and nanofluidics. The basic equations, including the Reynolds number, were introduced and some elementary but useful order-of-magnitude estimates were given.

The fluid environment is finding many uses in biology, chemistry, and material science. The discussion here focused on pressure-driven flows but it is frequently the case that electrically driven motions are used as the device scale shrinks. In addition to simply transporting material, it is necessary to mix fluids, minimize dispersion, implement combinatorial approaches, sense and detect chemicals, etc. All of these topics are areas of active research and new applications appear regularly. Hopefully, the discussion given in this chapter will serve as a useful introduction for those who wish to go deeper into the physics of fluids in small-scale devices.

ACKNOWLEDGMENTS

I am indebted to all of my collaborators who have contributed to my education in fluid mechanics and microfluidics in particular. Piotr Garstecki,

Rustem Ismagilov and Steve Quake are thanked for allowing use of their beautiful images. I especially thank Armand Ajdari for many helpful conversations about basic ideas useful for thinking about microfluidics and its applications.

REFERENCES

[1] T. Squires and S. Quake, "Microfluidics: Fluid physics at the nanoliter scale," Reviews of Modern Physics, vol. 77, pp. 977-1026, 2005.
[2] H.A. Stone, A.D. Stroock and A. Ajdari, "Engineering flows in small devices: Microfluidics toward a lab-on-a-chip," Annual Reviews of Fluid Mechanics, vol. 36, pp. 381-411, 2004.
[3] P. Tabeling, *Introduction to Microfluidics* (Oxford University Press, 2005).
[4] G. Karniadakis, A. Beshok and N. Aluru, *Microflows and Nanoflows: Fundamentals and Simulations* (Springer, 2005).
[5] T. Thorsen, S.J. Maerkl and S.R. Quake, "Microfluidic large-scale integration," Science, vol. 298, pp. 580-584, 2002.
[6] R.F. Ismagilov, A.D. Stroock, P.J.A. Kenis, G. Whitesides and H.A. Stone, "Experimental and theoretical scaling laws for transverse diffusive broadening in two-phase laminar flows in microchannels," Appl Phys Lett, vol. 76, pp. 2376-2378, 2000.
[7] P.J.A. Kenis, R.F. Ismagilov and G.M. Whitesides, "Microfabrication inside capillaries using multiphase laminar flow patterning," Science, vol. 284, pp. 83-85, 1999.
[8] S. Wiggins and J.M. Ottino, "Foundations of chaotic mixing," Phil Trans Roy Soc Lond Ser A, vol. 362, pp. 937-970, 2004; see also other paper in this special issue.
[9] P. Garstecki, I. Gitlin, W. DiLuzio, G.M. Whitesides, E. Kumacheva and H.A. Stone, "Formation of monodisperse bubbles in a microfluidic flow-focusing device," Appl Phys Lett, vol. 85, pp. 2649-2651, 2004.
[10] H. Song, J.D. Tice and R. Ismagilov, "A microfluidic system for controlling reaction networks in time," Angew Chem Int Ed, vol. 42, pp. 768-772, 2003.
[11] M. Joanicot and A. Ajdari, "Droplet control for microfluidics," Science, vol. 309, pp. 887-888, 2005.
[12] K. Jensen and A. Lee, "The science and application of droplets in microfluidic devices," Lab on a Chip, vol. 4, pp. 31N-32N; see also the other articles in this special issue.
[13] N. Rott, "Note on the history of the Reynolds number," Annual Review of Fluid Mechanics, vol. 22, pp. 1-11, 1990.
[14] D.J. Tritton, *Physical Fluid Dynamics*. (Oxford University Press, 1988).

[15] K.V.R. Sharp, R.J. Adrian, J.G. Santiago and J.I. Molho, “Liquid flow in microchannels,” in MEMS Handbook, M. Gad-El-Hak editor, (Boca Raton, FL, CRC Press) 2001.

AUTHOR BIOGRAPHY

Howard A. Stone is Vicky Joseph Professor of Engineering and Applied Mathematics in the Division of Engineering and Applied Sciences at Harvard University. He received his Bachelor of Science degree in 1982 from the University of California at Davis and his PhD in 1988 from Caltech, both in Chemical Engineering. After spending a year as a postdoctoral fellow in the Department of Applied Mathematics and Theoretical Physics at Cambridge University he joined the Harvard faculty in 1989. His research is mostly in fluid mechanics and his current interests include transport processes and applications in microfluidics, multiphase flows, and wetting phenomena.

3 MICRO- AND NANOFLUIDICS FOR BIOLOGICAL SEPARATIONS

Joshua D. Cross and Harold G. Craighead*

School of Applied and Engineering Physics, Cornell University
*hgc1@cornell.edu

3.1 INTRODUCTION

Current research on analytical techniques for biological applications is being conducted using micro- and nanofluidic devices fabricated with CMOS processes. Materials such as silicon, silicon nitride, and silicon dioxide are used as device substrates because they are compatible with the lithographic and etching processes required to manufacture nanometer-scale structures. Micro- and nanoscale structures have been fabricated in order to probe and confine molecules on length scales that are comparable to the size of the molecules [1]. At these size scales, advanced separation techniques are possible as are single molecule studies. Micro- and nanofluidic devices have enabled new methods of DNA separation, such as the rapid separation of genomic length DNA. The ability to manipulate, elongate, and detect individual molecules has opened the door for single molecule restriction mapping, directly observing protein binding, and perhaps even single molecule sequencing [2]. Labs-on-chips, of which microfluidics is the key enabling component, hold the promise of facilitating faster bioanalytical techniques using less reagents with more sensitivity and less variability [3].

This chapter summarizes one subsection of what has become a large field of multidisciplinary research known as microfluidics. Because this chapter is embedded in a book called *CMOS Biotechnology*, and because one of the author's expertise is in biological separations in glass- and silicon-based micro- and nanofluidic devices, what follows will be oriented in this direction.

Common fabrication techniques, likely familiar to the CMOS engineering community, will be discussed. DNA and proteins will be briefly discussed as these biological molecules are the primary analytes of most of the microfluidic devices discussed. A brief history of the move from "conventional" microfluidics to chip-based microfluidics will be presented. After the fabrication, the biology, and the history, the remainder of the chapter will describe new micro- and nanoscale systems that have been developed to interrogate and analyze biological samples.

3.2 FABRICATION OF FLUIDIC STRUCTURE

As CMOS fabrication techniques, academic cleanrooms, and advanced lithographic tools have become more available, engineers have increasingly used these techniques for applications other than the manufacture of "classical" CMOS devices. Using CMOS fabrication techniques to manufacture devices for biological applications is now common in the research community. Conventional techniques for separating and purifying biological molecules almost always involve capillary electrophoresis, which is electrically driving molecules through fused silica capillaries [4-6]. Using CMOS fabrication techniques, these capillaries can now be etched directly into silicon or glass wafers.

With creative geometries, wafers can hold hundreds of capillaries, each meters in length. In addition to the miniaturization of existing technologies, CMOS fabrication methods allow features to be made with dimensions of comparable size as biomolecules. Currently, the demands of making features on the order of nanometers or tens of nanometers requires "hard" substrates familiar to the world of integrated circuits such as silicon and silicon dioxide. The lithographic methods and the pattern transfer processes (as described below) will no doubt seem familiar to the general CMOS community. The ability to make channels and structures on the scale of biomolecules has enabled entirely new biological applications as well as elucidated many biological and physical phenomena.

Using standard CMOS fabrication technologies, many types of microfluidic devices can be made (see Figs. 3.1-3). As previously mentioned, capillaries can be easily placed on chips. Channel widths of tens of microns are defined using standard photolithographic techniques. A resist is spun onto a silicon or glass wafer. The resist is patterned with photolithography or electron beam lithography and developed. Once developed, reactive ion etching or chemical etching can be used to transfer the lithographic pattern

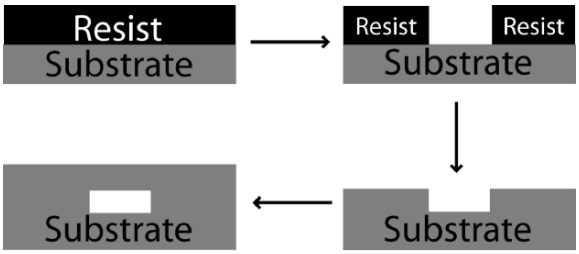


Figure 3.1 Schematic diagrams showing the fabrication process for a simple microfluidic channel. Resist is spin-coated on a SiO_2 substrate. The resist is then patterned with photolithography or electron beam lithography. Wet or reactive ion etching is used to transfer the pattern into the SiO_2 and the resist is removed. Finally, a SiO_2 cover is bonded to the SiO_2 substrate to encapsulate the channels.

into the substrate. The walls and floor of the chip-based microfluidic capillary are thereby defined. The ceiling of the channel is established when a cover wafer is bonded to the etched wafer. For the simplest implementation of capillaries-on-a-chip, this is all the fabrication required.

Sacrificial layer techniques are sometimes employed to produce very thin channels or when a thin capping layer is required (to reduce background fluorescence, for instance; see Fig. 3.2). A sacrificial film is deposited, spun, or grown on the surface of the substrate (examples of sacrificial layers include chrome, photoresist, or polysilicon). Lithography is used to pattern the sacrificial film. A capping layer is grown on the sacrificial film, which then has access holes patterned into it. Once the access holes are opened by wet or dry etching, the sacrificial layer is removed with an appropriate chemical etchant. Another film is then deposited to close the access holes and the microfluidic channels are complete. Capping layers of tens to hundreds of nanometers are possible. The microfluidic channels are defined by the sacrificial layer and can be as shallow as nanometers (see Fig. 3.10, for example).

More sophisticated fabrication techniques can be used to manufacture progressively more exotic microfluidic devices. An example of high-aspect ratio nano-pillars used to fractionate long DNA molecules is shown in Fig. 3.3 (from [8]). Multi-layer lithography, with aligned features, allows one to manufacture microfluidic channels that include holes, grooves in the channel floor, thin slits through otherwise micron deep channels, and similarly shaped gaps or restrictions. As many of these features can be made with sizes comparable to the size of things like DNA and proteins, many biophysical applications are possible in devices with these somewhat sophisticated geometries.

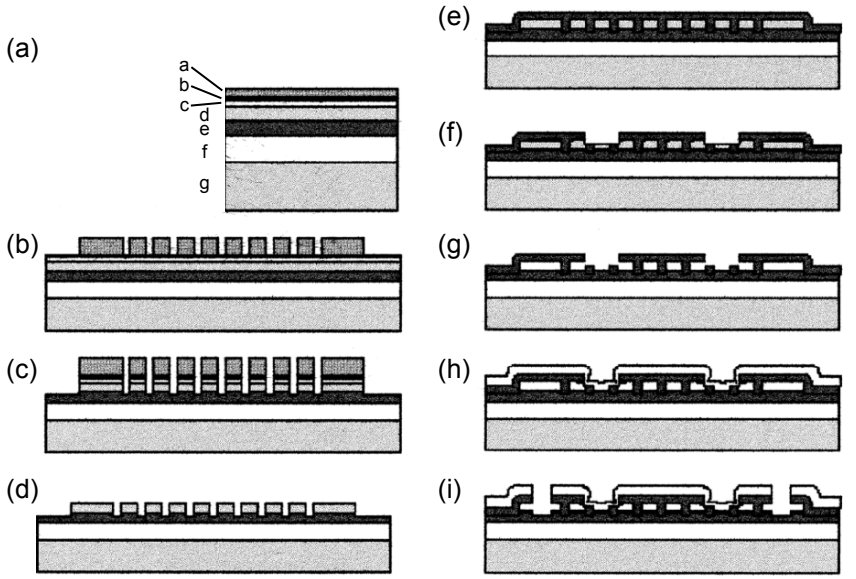


Figure 3.2 An overview of a sacrificial layer fabrication process (reprinted with permission from [7], Copyright 2000 Wiley-VCH Verlag GmbH & Co KG). For an example of a device made using this technique, see Fig. 3.10. (a) The film stack is composed of: a-PMMA resist, b-aluminum conducting layer, c-silicon dioxide hard mask, d-polysilicon sacrificial layer, e-silicon nitride floor layer, f-silicon dioxide insulating layer, and g-silicon substrate. (b) Pattern the resist. (c) Etch pattern into sacrificial layer. (d) Remove resist and pattern transfer layers. (e) Apply ceiling silicon nitride layer. (f) Add access holes to facilitate sacrificial layer removal. (g) Remove sacrificial layer. (h) Seal access holes with silicon dioxide. (i) Add reservoirs via new access holes.

As will be discussed below, the ability to manufacture microfluidic channels with features of comparable size as biomolecules permits new research on fundamental biophysical phenomena and applications. Using confining microfluidic geometries and single molecule imaging techniques, individual molecule properties such as conformation, length, affinity for substrates, and susceptibility to localized chemical treatments can be studied more directly and with better resolution than previously possible. Artificial gel-like structures can be created to separate biomolecules by length, diffusion constant, affinity for patternable substrates, or other physical properties. Additionally, due to the high surface-to-volume ratio achievable in microfluidic channels, miniaturized capillary systems are able to perform better than their macroscopic versions as higher electric fields can be used to drive molecules because heat dissipation is greatly improved.

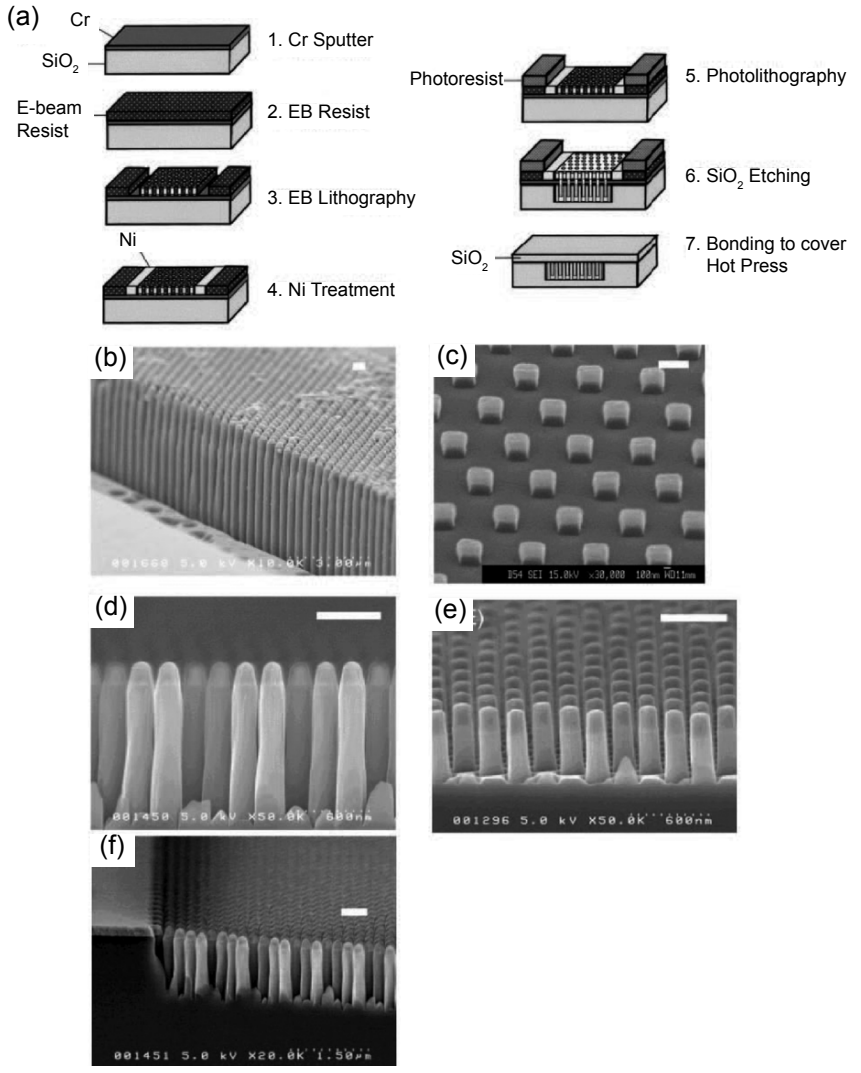


Figure 3.3 A schematic diagram of the fabrication process used to produce the high aspect ratio nanopillars shown in the lower scanning electron micrographs (reprinted with permission from [8], Copyright 2004 American Chemical Society). The devices are fabricated on a glass substrate. Chrome is used as a conducting layer under the electron beam resist. Nickel is electroplated to act as a mask for the silicon dioxide dry etch. The nanopillars are dry etched using a neutral planar discharge technique. The channels are sealed with a quartz cover plate. The micrographs show different aspect ratios of pillars (height to width, scale bar in each is 500 nm): b-20, c-1, d-6, e-3. The last micrograph (f), shows defects in the nanopillars.

3.3 BIOLOGICAL APPLICATIONS

As the majority of the microfluidic systems described in this chapter are used to separate mixtures of biological molecules, we present a brief discussion of the importance of these molecules. Proteins, RNA, DNA, and cells are the key biological molecules at work in all living organisms [9]. Proteins are responsible for gene expression and regulation, for shuttling other molecules throughout cells, and for catalyzing biochemical reactions within cells (just to name a few key roles they play) [6, 10]. Analytical techniques used to separate proteins are useful for diagnosing diseases or for investigating when certain genes are turned on or off throughout the cell cycle [6, 11]. Similarly, techniques for separating mixtures of DNA molecules are useful for studying genetic variation, which is important for determining how individual genes function, elucidating the foundations of genetic diseases, or anticipating reactions to medicines [5]. Separating DNA by size is also useful for characterizing DNA libraries and is necessary to sequence DNA molecules or do forensic DNA fingerprinting [5, 12]. While we do not mean to ignore RNA and cells, most of the techniques described below focus on DNA and protein samples, so only these molecules will be described in any detail here.

DNA forms the genetic code of all living things [5]. Variations in DNA length or sequence (described below) are important indications of genetic differences or mutations, are indications of gene function (as observed via "knock out" experiments), and can be used in DNA fingerprinting applications [12]. Chemically, DNA is a polymer that is made up of four monomer constituents that can be arranged in any order along the entire length of the chain. The monomers are typically referred to as bases and are called individually: A (adenine), C (cytosine), G (guanine), and T (thymine). The bases are of comparable molecular weight, and each is attached to the next by an identical piece of backbone. Therefore, the polymer mass is essentially proportional to the number of bases. The particular series of bases that form a DNA polymer is referred to as a sequence. DNA sequencing refers to the process of taking segments of DNA with unknown base composition and learning what that composition is [4, 5]. Sequencing DNA is useful for applications ranging from detecting genetic diseases to tracing evolutionary pathways.

DNA is found in both single and double stranded configurations. The single strand is a single polymer chain, whereas the double stranded version is composed of two single strands wrapped about each other in a stable, helical configuration [13]. DNA only forms stable double stranded configu-

rations when each strand is the complement of the other. Complimentary base pairings are: A with T and G with C. In bioanalytical applications, both single and double stranded DNA molecules are of interest, and DNA polymer lengths commonly used in experiments range from tens of bases to millions of base pairs.

DNA is a charged polymer. Over most of the pH range, the backbone of a DNA polymer carries one free negative charge per base. Because both the charge and mass of the DNA polymer scale linearly with the number of bases, DNA molecules have a constant charge-to-mass ratio. This constant charge-to-mass ratio is important because it makes separating DNA molecules by length challenging.

As a polymer, DNA is well-described using the terminology of polymer physics. When in an electrolyte solution of sufficient strength, a DNA molecule's backbone charge is sufficiently screened by counterions in the solution so that the molecule can relax into a blob configuration. A polymer blob is often described by a radius of gyration, R_g, which can be calculated from Flory's theory that takes into account the excluded volume of the polymer [14]

$$R_g \propto N^{3/5}, \tag{3.1}$$

where N is the total number of monomers. Another useful means of describing a polymer is by its persistence length, P. The persistence length of a polymer gives an indication of the stiffness of polymer (a large P means a stiffer polymer). A typical double stranded DNA molecule has a width of 2 nm and the monomer length is 0.34 nm [5, 15]. Single stranded DNA has a persistence length of ~3 nm, whereas double stranded DNA has a persistence length of ~50 nm. The persistence length is dependent upon the salt concentration [16].

As an electro-negative molecule, DNA in solution can be moved with electric fields. This technique is called electrophoresis. When an electric field is established in a fluid containing DNA molecules, the molecules move toward the positive electrode. To account for variability in the electrophoretic velocity caused by variations in environmental parameters such as salt type and salt concentration in the buffer, molecular mobility is typically cited in lieu of molecular velocity. The mobility is related to the velocity by

$$v = \mu E, \tag{3.2}$$

where v is the velocity of the molecules (cm/sec), E is the applied electric field (typically measured in V/cm) and μ is the mobility ($cm^2/V/sec$).

Unfortunately, in a solution free of obstructions, DNA molecules of different lengths almost always move with the same velocity [17-19]. This is unfortunate because if they had an inherent size-dependent velocity, then separating DNA molecules by length would be much easier than it is. Most DNA molecules migrate with the same velocity because both the charge and the mass of the molecules scale directly with the length of molecules. Thus, while the electromotive element (the charge) increases as the molecules get longer, so too does the drag (which is directly related to the mass). This unfortunate scaling relationship is the primary reason why so much effort has been put into developing sieving matrices like slab gels or microfluidic obstacle courses [18-20]. These restrictive physical environments impart DNA molecules with size-dependent velocities generally through friction, but not always, as will be highlighted later.

Proteins are also polymers, but are made of combinations of 20 different monomers known as amino acids [6]. The amino acids are also not all negatively charged. Each amino acid of the protein polymer can have one or multiple ionizable hydrogen atoms. Therefore, across a range of pH values, proteins can have dramatically different electrical properties. Additionally, amino acids are much more disparate in size than are the four DNA bases. Thus, proteins do not have a constant charge-to-mass ratio. Some analytical techniques take advantage of the variability of proteins' charge-to-mass ratio to directly separate proteins by free solution electrophoresis [6]. Another analytical technique that can be used to separate proteins is isoelectric focusing [6]. With this technique, a pH gradient is established in a column through which proteins are electrically driven. Ionization occurs at various points along the pH gradient for different proteins, and at particular pH each protein becomes electrically neutral and stops moving. Direct electrophoresis of proteins and isoelectric focusing are useful, but are not always helpful for samples of interest and are not always capable of adequately resolving mixed samples. In these cases, proteins can be immersed in a buffer containing a constant charge-to-mass ratio molecule that uniformly binds along the protein polymer chain (such as sodium dodecylsulfate or urea). These so-called denaturants impart all proteins with essentially the same charge-to-mass ratio. Then, sieving matrices similar to those used to separate DNA molecules by size can be used to separate the protein mixture by size.

Proteins and DNA are ubiquitous as specimens for biological assays because they are at the heart of many biological processes and diseases. DNA sequences encode genes that are turned into proteins, so mutations in DNA can cause more-, less-, or non-functional proteins to be created [9, 10]. Any of these deviations from normalcy has the potential to cause drastic,

observable biological consequences (such as disease) [11]. Thus, analytical techniques such as separating DNA molecules by size are used to detect disease causing mutations. Proteins are truly pervasive in biological processes. Proteins are the worker-molecules that perform diverse biological tasks such as packing and unpacking DNA from its storage place in cell nuclei, transcribing DNA into RNA which then gets translated into new proteins, transporting molecules across cell membranes, and participating in metabolic processes that convert stored fuel into usable energy [6, 9]. Protein separations are used to detect the presence of proteins at various points in the cell cycle or to detect mutated forms of proteins. Protein assays are critical for understanding what exactly proteins do and what happens to biological systems when proteins do not do what they are supposed to do. For these reasons – and many more which were omitted for brevity – there is tremendous interest in developing tools for separating proteins and DNA.

The ability to use CMOS fabrication techniques to manufacture microfluidic channels in silicon and glass chips has enabled miniaturized and improved versions of well-established biological and chemical techniques. Additionally, these microfluidic chips afford new analytical techniques. Separations of biomolecules by a range of physical properties are possible. Microfluidic filtration systems can be used to remove unwanted constituents of a complex mixture as well as concentrate and localize critical constituents that are often present at only a fraction of the concentration required for quantitating. Most or all of the research effort in microfluidics is in some way directed at achieving a so-called micro-total-analytical-system (μTAS) or lab-on-a-chip.

The allure of a μTAS is that there are potentially significant advantages over traditional systems. Very low sample volumes can be used, thus reducing the amount of reagents needed to perform experiments. Additionally, single molecule fluorescence imaging techniques permit extremely low concentrations of analytes to be detected in microfluidic channels. As microfluidic cell handling techniques improve and on-chip PCR and filtration are better refined, we will see more and improved experiments in which an individual cell is loaded into a chip reservoir and its DNA or proteins sorted and quantified. A μTAS has the potential to be massively parallel, dramatically improving analysis throughput. Improved heat dissipation, leading to better resolution, is another μTAS advantage. And, increasingly, key analytical components such as optics and detectors are being fabricated in such a way as to be packaged directly with the microfluidic channels. All of these advantages are driving the effort to develop μTAS technology that is widely deployed in the medical and research communities.

3.4 MICROFLUIDIC EXPERIMENTS

Chip-based microfluidic devices have emerged over the last decade as viable – and perhaps superior – platforms for separating biological molecules. Microfluidics, as exemplified by the promise of μTAS technology, have the potential to perform "classic" biological separations using less reagents, in less time, and with better resolution than current methods can deliver. A diagram of a typical microfluidic channel is shown in Fig. 3.4. Figure 3.4(a) shows a top-down view of a standard, crossed-channel design used for separating biomolecules based upon size, charge, or affinity to a material in the separation channel. Analytes would typically be loaded into reservoir 1 and driven towards reservoir 2, filling the intersection (Fig. 3.5). Well-defined plugs of the sample are injected into the separation channel, which is the channel extending from the intersection to reservoir 4, by switching the

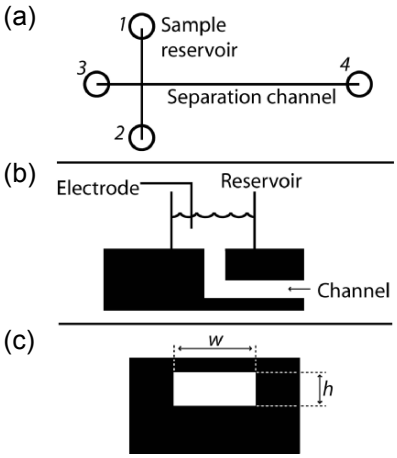


Figure 3.4 (a) A schematic representation of a typical microfluidic device. One such device would constitute a microchip and tens to hundreds could populate a four or six inch wafer if carefully designed. Biological samples are loaded into one of the reservoirs (reservoir 1 in the figure). The channel from reservoir 1 to reservoir 2 is the "loading" channel. The channel from reservoir 3 to 4 is the "separation" channel. Pressure- or electrokinetic-driven flow is used to move biological analytes through the channels. (b) A side-view of the interface between the macro- and micro-world. The reservoir is often a pipette tip glued to the chip. The access hole is a hole etched or drilled through the substrate in which the channels were fabricated. (c) A typical square cross-section microchannel (that would result from reactive ion etching). The width, *w*, is defined lithographically and is limited only by lithographic resolution. The channel height, *h*, is established through the etching process. This height is limited by practical etch limitations, and is sometimes also limited by the method used to bond a cover to encapsulate the channels.

direction of the flow to point from reservoir 3 to reservoir 4. In almost all circumstances, reservoirs 1 and 2 must be "back-biased" to prevent leakage of sample from these side-channels into the separation channel.

A generic schematic of a laboratory-based experimental setup is shown in Fig. 3.6. The chip containing the microfluidic channels is placed upon a microscope (typically an inverted microscope to allow for facile introduction of wires and tubing to the reservoirs on the chip). The microscope is equipped with a broad spectrum light source such as a mercury arc lamp for fluorescence illumination. Lasers can also be used to illuminate fluorescently-labeled biomolecules. Biomolecules can readily be labeled with fluorophores that have well-defined excitation and emission wavelengths. Bandpass or longpass filters are used to select wavelengths of light near where fluorophores are excited and emit fluorescent light. A CCD or photomultiplier is generally used to collect the photons emitted from the fluorescent molecules. The detector is positioned at some point along the separation channel and measures intensity or absorbance of the analytes as they are illuminated by the light source. Bio-separations are typically quantified with intensity versus time data from which other important values such as analyte mobility or abundance are calculated.

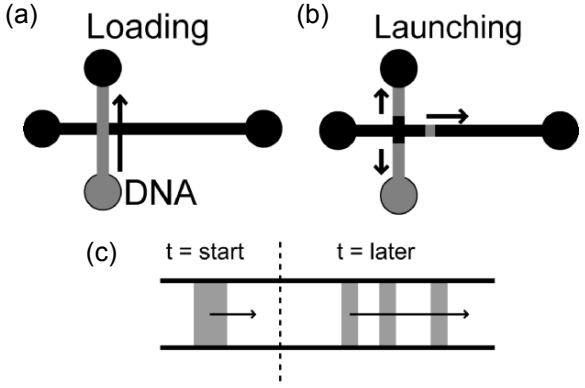


Figure 3.5 Schematics showing a typical microfluidic device and a DNA separation experiment. (a) DNA is driven electrophoretically from one reservoir to another in the loading phase. (b) The electric field is switched to direct a plug of DNA in the intersection to move down along the separation channel. In DNA separation experiments, the separation channel is typically filled with a sieving matrix. DNA near the intersection but not launched in the plug is typically forced to move away from the intersection with an electric field. This is done to prevent sample leakage into the separation channel. (c) An example of the desired effect when a plug of molecules (t = start) is launched into a separation channel and separates into individual constituents (t = later).

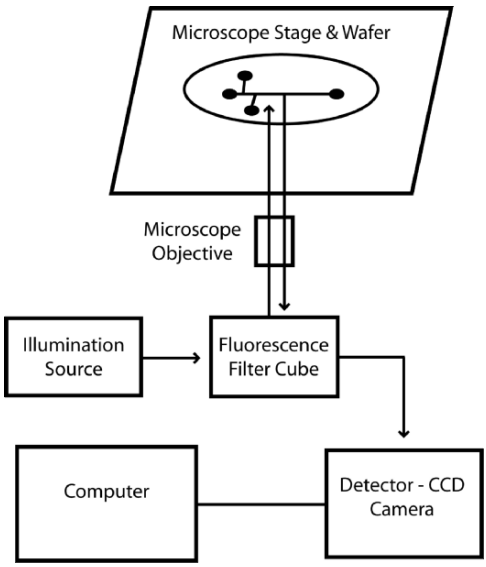


Figure 3.6 A schematic of a typical microfluidic experimental setup. A microchip or full wafer containing microfluidic devices is mounted on a microscope stage. A broad-spectrum light source or laser is typically used to illuminate fluorescently labeled molecules in the microfluidic channels. A photodetector such as a CCD camera or photomultiplier tube is used to quantify the emitted intensity from the interrogated molecules.

A schematic side-view of a reservoir and the access hole into the microfluidic channel is shown in Fig. 3.4(b). Though pressure-driven flow is sometimes used in microfluidic experiments, electric fields are typically used to move either the liquid (electroosmotic or electrokinetic flow) or the analytes in the channel (electrophoresis). Electrodes from a power supply at placed in the reservoirs to establish an electric field in the microfluidic channels (sometimes these electrodes are integrated onto the chip through the fabrication process as in [21]).

A cross-section of a microfluidic channel, as seen looking along the direction of the channel, is shown schematically in Fig. 3.4(c). The channel width, w, is typically defined and set lithographically, so it has dimensions between tens of nanometers (electron beam or electrospinning techniques) and tens of microns (photolithography). The channel height, h, is usually established through etching or a sacrificial layer process. Micron-scale depths are typically obtained with chemical etches such as hydrofluoric acid for glass substrates. Chemical etches are usually isotropic, so channels have rounded bottoms and sidewalls. If reactive ion etching is used to etch the

channels, then a square channel bottom is achieved, and the height is usually a couple of microns down to tens of nanometers.

The last step in most microfluidic fabrication processes is the bonding of a cover wafer to the processed wafer containing the channels (this is typically not necessary in sacrificial layer processes). Once the channels are sealed, they must be filled, and this is not always a trivial process (though a new, super-critical wetting technique was just reported [22] that is capable of filling even otherwise difficult to fill channels). When all of the channel dimensions are at least 1 μm, the filling process is usually straightforward. Whatever buffer is to be used for experiments can be directly introduced into the channels. The channels will fill by capillary force with few air bubbles. Once the channels are filled, any trapped air bubbles dissolve over time or can be driven out with electroosmotic flow (described later). Channels with a dimension that is less than 1 μm are more difficult to fill. Significant numbers of air bubbles form, especially in long channels or near intersections and corners. Additionally, even relatively large channels that have features such as pits, grooves, lateral pockets, or slits can be expected to require at least some massaging to remove trapped air bubbles. Low surface tension solvents, such as ethanol, can be used to help mitigate the formation of air bubbles in microfluidic channels.

Once a continuous fluid path is established between two reservoirs, an electric field can be used to help drive fluid through the channel. This can be done even in the presence of large bubbles and is often necessary to dissolve or drive out bubbles. Though capillary force is a convenient means of filling microfluidic channels, pressure can also be used to hasten the process. Sometimes pressure is necessary due to complicated channel geometries or because air bubbles are difficult to eliminate passively. Both positive and negative pressure can be used if appropriate reservoirs are securely fixed to the microfluidic channel access holes.

The interface between the macro-world and the micro-world is important, especially when thinking about a μTAS. In the research environment, microfluidic channels are most frequently made in chips or wafers with at least a 1 cm^2 footprint. The only access point between the big world and the small world is through the reservoir. Reservoirs are typically made from cut pipette tips glued around an access hole in the wafer. Reservoirs made to withstand high pressures and made with standard thread gauges for pressure-compatible fittings can also be purchased from commercial vendors. In commercial applications, reservoirs are sometimes nothing more than the access hole itself. Mechanized sample loading and electrode placement enables a direct connection to the microscale. Furthermore, pressurized or

sealed sample environments in commercial microfluidic devices prevent annoying effects such as evaporation, so significantly smaller volumes of liquid are needed.

Biological samples always require particular liquid environments in order to assume physically relevant forms. Proteins, for example, adopt complex structures that are meaningful and must be preserved for some types of experiments or eliminated for others. Buffer pH and buffer additives (such as sodium dodecylsulfate) can be used to control the structural properties of biomolecules. Of course, pH and additives often impact the electrical properties of the buffer, which subsequently affect the electrical properties of the microchannel surfaces, which finally affect the electrokinetics of the flow in the channels. For example, high ionic strength salt in a buffer significantly reduces the surface potential in glass microchannels causing decreased electroosmotic flow (bulk flow of liquid from the positive electrode to the negative electrode) [23-25].

There are primarily two types of flow used in microchannel experiments: pressure-driven flow and electrokinetic flow [26]. Each has associated advantages and disadvantages. The technique employed is often dictated by experimental conditions such as the buffer necessary for the biomolecules.

Pressure-driven flow can be used in almost all cases. Reservoirs must be securely affixed to the wafer in order to provide leak-proof sealing. Pressure-driven flow moves the bulk fluid through the channel carrying analytes along with it. Pressure-driven flow is relatively simple in that everything in the microchannel moves in the same direction. That is, positively and negatively charged analytes move with the bulk fluid. This is a straightforward means of driving both fluid and analytes through microchannels.

One of the biggest drawbacks of using pressure to drive flow is that the liquid velocity profile is parabolic. This arises from the no-slip boundary conditions on the liquid at the channel walls [26]. The parabolic velocity profile automatically disperses analytes plugs across the channel width. Thus, experiments that rely upon discrete plugs of analytes being separated with high resolution do not lend themselves to pressure-driven flow. Corners, side channels, and obstacles are also problematic for pressure-driven flow setups. Flow instabilities can arise in channels that are not straight. These can lead to vortices, stagnation regions, and generally unstable flow. Pressure drops in the fluidic system must be well understood, because it is possible that the flow rates required for an experiment are not achievable with a given pressure setup (it might not be possible to achieve high enough pressures in the microfluidic channels for instance). Additionally, because many traditional biological experiments depend upon the electrical force to

move charged molecules in response to an applied electric field, pressure-driven flow does not lend itself to performing certain "classical" biological experiments.

Electrokinetic flow involves moving the analytes, the bulk fluid, or both, by applying an electric field through the buffer. Electroosmosis is the term used to describe bulk fluid flow in a channel [27]. This phenomenon occurs when there is a surface charge on channel walls. SiO_2 channel walls, for instance, have a negative surface charge over most of the pH range. The surface charge leads to a wall potential (the so-called zeta potential) that decays into the channel where counterions in the buffer solution effectively neutralize it. Near the walls, however, the surface potential is significant and draws counterions to it. These counterions are not tightly bound to the surface, however, and when an electric field is applied in the channel, they move towards the positive electrode. These counterions drag fluid along with them and the net effect is that the entire buffer solution moves toward the positive electrode. Unlike the pressure-driven velocity profile, however, the electroosmotic velocity profile is nearly uniform over most of the channel (assuming channels are more than about 1 μm in their smallest dimension [23]). Thus, electroosmosis is quite effective at moving samples (via bulk fluid transport) with minimal dispersion.

Just as the counterions in the solution move towards the positive electrode, charged analytes also move. Negatively charged analytes move with the fluid toward the positive electrode, while positively charged analytes move against the fluid toward the negative electrode. Neutral molecules simply have the velocity of the bulk fluid. If the electroosmotic flow is strong enough, then all molecules will migrate in the same direction. The phenomenon of charged molecules moving in an electric field is called electrophoresis. The total mobility of electrokinetically driven molecules is

$$\mu_{total} = \mu_{EOF} + \mu_{EP}, \tag{3.3}$$

where μ_{EOF} is the electroosmotic (bulk) velocity, and μ_{EP} is the electrophoretic velocity.

Electroosmotic flow can be reduced or eliminated by minimizing the wall potential. Minimizing the wall potential can be accomplished with high concentrations of salt [25] or with surface coatings [8]. If electroosmotic flow can be minimized or eliminated, then purely electrophoresis is observed. There are some advantages to eliminating electroosmotic flow. If negatively charged analytes are used, eliminating electroosmotic flow speeds up the overall analysis time (because the electroosmotic flow slows down the analytes). Electroosmotic flow is also somewhat unstable, especially in devices

that are have one dimension less than 1 μm or in devices that have many intersections, corners, or otherwise complicated geometries. Because there is bulk fluid flow, there is the potential for stagnation points or vortices in the flow. These fluid effects are decoupled from electrophoresis, so it is possible to create very interesting (but usually not useful) analyte flow patterns when instabilities arise. Stagnation points and vortices are avoided when electrophoresis is used because the fluid does not move and charged molecules follow electrical field lines through the microchannels. These electrical field lines typically fill microchannels quite well [28].

In summary, the simplest microfluidic devices involve etching trenches in substrates such as glass or silicon. The fluidic channel is completed by sealing the etched substrate with a cover. The chip or wafer containing the microfluidic channel is mounted on a microscope with an illumination source and a detector. Biological samples are introduced to the microfluidic device via reservoirs that provide a gateway between the micro- and macroworlds. Pressure or electrokinetic drive is used to move fluids and analytes through the channels where they separated. Analysis is accomplished by observing and quantifying what occurs within the channels or by removing processed samples from another reservoir and performing an off-chip analysis. These steps comprise the essential features of microfluidic experiments and are almost all necessary in any experiment.

3.5 MICROCHANNEL CAPILLARY ELECTROPHORESIS

In this section we describe initial efforts to miniaturize conventional capillary electrophoresis experiments. Microfluidic devices were fabricated that replicated the essential qualities of capillary systems. These early microfluidic chips incorporated electrokinetic sample handling protocols to facilitate easy sample manipulation within the microfluidic environment. Additional improvements over conventional capillary systems were realized through decreased detection limits and decreased analysis time.

One of the first steps towards miniaturizing conventional biological assays and developing a μTAS was the work of Harrison *et al.* [29]. Over a decade ago, these researchers micromachined a glass substrate with 30 μm wide channels for use in separating two charged, fluorescent molecules. The experiment was essentially the microfluidic analog of capillary electrophoresis. Their results indicated that glass substrates and microfabrication techniques could effectively be used to miniaturize capillary electrophoresis equipment. Furthermore, their results showed that microfluidics

offered excellent fluid and sample handling characteristics. Rather than using pressure in conjunction with separate sample vials, they were able to introduce samples from on-chip reservoirs with electroosmotic pumping. Electrokinetic sample handling is significantly simpler than pressure-based methods and is very compatible with microfluidic systems.

They report their separation efficiency in terms of the theoretical number of plates (a common chromatographic measure of separation efficiency), and obtain a maximum number of theoretical plates for calcein of 35,000. This is comparable to conventional capillary electrophoresis methods. Finally, they estimate that their microfluidic separation system performs nearly ideally in the sense that the only measurable dispersion is due to the finite size of the detector and the size of the injected sample plug. Joule heating and analyte-wall interactions play no role in band broadening.

Another effort to miniaturize conventional capillary electrophoresis experiments was conducted by [30]. One of their first reports describes a 2.5 cm^2 glass chip upon which a serpentine channel was etched. This unique geometry enables much longer microfluidic channels to be made on a much smaller footprint. In addition to using a new geometry to minimize the overall chip-area, the group describes techniques for minimizing band dispersion caused by inefficient injection methods. In order to minimize the width of the injected analyte plug, it is necessary to apply appropriate voltages to all reservoirs, not just the sample loading and sample loading waste reservoirs. The separation reservoirs must be biased with voltage (as opposed to floated relative to ground) to prevent leakage during the loading phase. The group showed that these serpentine devices could be used to separate two fluorescent dyes as depicted in Fig. 3.7.

Shortly after their work in 1992, Harrison *et al*. reported results showing the separation of six amino acids (the monomer units of proteins) in similar microfluidic channels [31]. Results of the amino acid separation experiment are shown in Fig. 3.8. The reported separation efficiency was 40,000 to 75,000 theoretical plates, as compared to 400,000 theoretical plates reported in conventional capillary electrophoresis experiments [32]. While the microfluidic system's separation efficiency is an order of magnitude lower than the capillary system, the microfluidic-based experiment takes about 15 sec while the conventional experiment takes about 15 min.

In 1995, Jacobsen *et al*. used fused quartz microchips to electrophoretically separate metal ions bound to 8-hydroxyquinoline-5-sulfonic acid [33]. The researchers used zinc, cadmium, and aluminum, all of which possessed a net negative charge in solution. Because the negative charge on the metals would cause the metal electrophoretic mobility to be in the opposite direc-

tion as the bulk fluid electroosmotic mobility, the researchers coated the channels with linear acrylamide. This coating, which is applied by flowing acrylamide through the channels and covalently linking it to the negatively charged fused quartz surface prior to electrophoresis experiments, has the effect of suppressing the zeta potential which in turn essentially eliminates electroosmotic flow of the bulk fluid. When only negatively charged analytes are used in electrophoresis experiments, it is generally recommended to eliminate the bulk electroosmotic flow to increase analyte mobility and decrease analysis time. The researchers were able to separate the three metal complexes in about 15 sec. Owing to the very low background fluorescence of the fused quartz microchip, quantities of metals as low as 46, 57, and 30 ppb for Zn, Cd, and Al respectively. Similar experiments done with con-

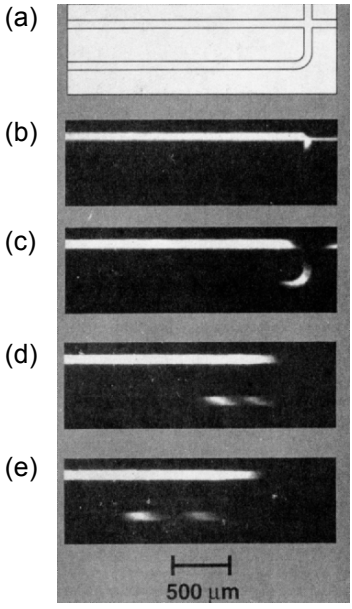


Figure 3.7 A five panel collage of optical micrographs of the serpentine device described in [30], reprinted with permission from [30], Copyright 1994 American Chemical Society. (a) bright field image of the intersection of the loading and separation channels. The beginning of the serpentine separation channel extends to the bottom and bottom-left of the image. (b)-(e) 1 sec interval fluorescent images of plug injection and separation. (b) Loading fluorescent dye through the intersection. (c) The plug just after injection into the serpentine separation channel. (d) The plug has separated into its two constituents, rhodamine (faster) and sulforhodamine (slower). (e) The samples are fully resolved. Note that the slight back-bias along the loading channel has drawn the sample in the loading channels away from the intersection, preventing sample leakage into the separation channel.

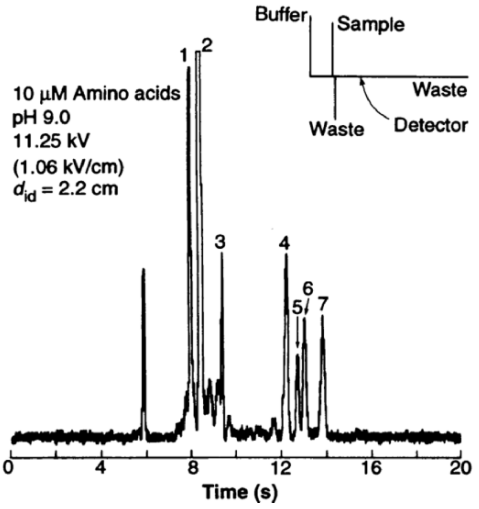


Figure 3.8 Results showing the separation of six amino acids in a microfluidic device, reprinted with permission from [31], Copyright 1993 AAAS. Peaks 1 and 3-7 are the amino acids. Peak 2 is a reactive by-product involving the fluorescent dye used in the experiments. The unlabeled peak at ~6 sec is not mentioned in the paper. The schematic in the upper right portion of the image shows the channel configuration. Translating the inset names into the terminology of Fig. 3.3(a), 'sample' is '1', 'buffer' is '3', (lower) 'waste' is '2', and (right) 'waste' is '4'. Electrokinetic pumping is used to move analytes throughout the microfluidic network. The separation efficiency of the microchip capillary electrophoresis device is about 50,000 theoretical plates which is comparable to conventional techniques.

ventional capillary electrophoresis systems result in detection limits from 46 to 613 ppb in times of at least two or three minutes [34].

These initial microchip-based capillary electrophoresis experiments all involved separations of molecules with different charge-to-mass ratios (or quantification of charged molecules in the case of [33]). No sieving matrix was present in the microchannels. When electrophoresis if performed in capillaries or channels with no matrix, the process is typically referred to as free solution electrophoresis. When the charge-to-mass ratio of molecules is different, then free solution electrophoresis is capable of separating the molecules given enough resolution or channel length. For proteins in their native conformation (or amino acids) or for biomolecules that can be selectively bound to "carrier" molecules of different charge-to-mass ratios, free solution electrophoresis is typically sufficient for separation. For DNA or denatured proteins (which have a constant weight and charge molecule bound along the entirety of the polymer), free solution electrophoresis is not capable of effecting length-based separation. A sieving matrix must be

introduced to break the so-called charge-to-mass symmetry. Slab gels of agarose or polyacrylamide are typically used to separate DNA and denatured proteins. Agarose, polyacrylamide, and similar substances can be introduced into microfluidic channels creating microfluidic analogues of gel electrophoresis techniques. These gel-filled channels are discussed in the next section.

3.6 FILLED MICROFLUIDIC CHANNELS

With the initial success in miniaturizing capillaries to the microchip format, researchers moved to improve the performance of microchip devices. For capillary electrophoresis or microchannel-based electrophoresis experiments improved performance generally means increased sample resolution. As capillaries had been filled with sieving matrices to improve their resolving power, so too were microfluidic channels filled with sieving matrices. Just as filling capillaries with viscous polymer solutions is difficult, so too is filling microfluidic channels. A number of techniques for filling channels have been developed and here we will describe the basic methods as well as some of the successes of filled microfluidic channels in rapidly separating biomolecules.

There are basically two types of polymer solutions that can be used to fill microfluidic channels: polymers that are cross-linked to form a gel-like matrix, and linear polymers that do not cross-link. For polymers that cross-link to form gels, the viscous nature of the gels makes introducing gels into microfluidic channels difficult. Due to the laminar nature of microfluidic flows, unless geometries are specifically chosen to force mixing to occur, mixing is typically difficult to achieve inside the microchannels. Therefore, chemical cross-linking agents can only be used if channels are filled immediately after all chemicals are mixed and a rapid, reliable channel-filling procedure is used. Then, because gelation of the matrix begins outside of the microfluidic channel, pressure-filling the channels with the mixture is usually necessary. Filling the channels with the polymer and then trying to achieve mixing of the cross-linking agent via diffusion is too slow to be practical. To overcome the problem of filling channels with a viscous polymer solution that is in the process of forming a sieving matrix, researchers often use photoinitiation to induce the gel-forming chemical reaction. Photoinitiation works for most polymers, such as polyacrylamide, and is relatively straightforward to implement. In gel-filled channels, analyte separation is achieved via transient entanglement of the analyte upon the gel

fibers, through frictional interactions between the analyte and the fibers, or through the process of an analyte hopping between restrictive regions and voids (a process discussed in Subsection 3.7.3).

Linear polymer solutions are generally easier to introduce into microfluidic channels than gels. If not too viscous, these polymer solutions can either be pressure-driven into microfluidic channels or carried into and through channels using electrokinetic pumping. Linear polymers that are frequently used are linear acrylamide or hydroxyethylcelluose. Numerous others exist as well (see, for example, the review [35]). Typical polymer concentrations in solution range from 0.0% to 2%. For linear polymer solutions, separation typically occurs as a result of transient entanglement of analyte molecules with the polymer fibers. Because the polymer fibers are not linked to each other or to the channel wall, they can be dragged by the analyte molecules through the microchannel, slowing down the migration of the analytes.

The reason one fills a microfluidic channel with a polymer solution (whether it is cross-linked or not) is to increase the resolution of biomolecular separations. When DNA and proteins are to be separated, for instance, a sieving matrix is almost always required to facilitate size-dependent mobility differences. Because DNA has a nearly uniform charge-to-mass ratio, frictional interactions with polymer solutions or gels must be used to separate mixed samples of molecules by size. Proteins to be separated by mass are often denatured with chemicals like sodium dodecylsulfate, which effectively renders all proteins globular and of similar charge-to-mass ratio, so a sieving matrix is required to separate these molecules as well.

In one of the first reported realizations of a microchip-based gel electrophoresis system, Effenhauser *et al.* separated single stranded DNA oligonucleotides 10-25 bases in length [36]. Though slightly modified in layout, their devices had essentially the same injection and separation geometry that has been previously discussed. Electrophoresis was used to move samples throughout the microfluidic channels. The researchers were able to separate DNA oligomers differing in length by a single base in a 10% polyacrylamide gel, a similar concentration to what would typically be used in a conventional slab gel. The ability to resolve length differences for these short DNA fragments is potentially useful for analytical DNA techniques such as single nucleotide polymorphism detection or for screening short DNA fragments used in hybridization experiments. The separation takes approximately 40 sec in the microfluidic format, as opposed to tens of minutes in a conventional slab polyacrylamide gel. Additionally, the microfluidic format offers highly automated and reproducible electrokinetic sample injection. The researchers showed a series of repeated, back-to-back injections across

which the migration times varied by less than 1% and peak heights varied by less than 2% (relative standard deviations).

Shortly after the work of [36], Mathies' group used a sieving matrix of hydroxyethylcelluose (HEC) to separate short DNA fragments in glass microfluidic channels [37]. Separations of ϕX174/HaeIII digested DNA fragments were performed in microfluidic channels with acrylamide bound to the channel surfaces (to reduce surface interactions) and 0.75% (w/v) HEC in the separation channel buffer. The HEC matrix was pressure-driven into the separation channel. Researchers were able to separate the DNA digestion fragments in less than two minutes with resolution comparable to that of conventional capillary gel electrophoresis experiments. The microfluidic-based separation time is approximately ten times faster than possible with conventional techniques. The researchers report comparable run-to-run variability as with conventional techniques and with other microfluidic capillary gel electrophoresis experiments, and note that at least 75 consecutive separation experiments could be performed in the microfluidic channels without replacing the sieving matrix.

These works, and others, established that gel-filled capillary electrophoresis techniques could be miniaturized successfully. Many commercially available microchip separation products incorporate (usually proprietary or non-disclosed) sieving matrices in their fluidic channels. A wide range of polymers exist that can be more or less easily incorporated into microfluidic systems. One of the ultimate measures of system performance is whether or not long DNA strands (500-1000 bp) can be separated with single base pair resolution. This is the resolution achievable with current capillary electrophoresis techniques employed to sequence DNA. Just as we examined the move to miniaturize free solution and gel-filled capillary electrophoresis techniques, we briefly discuss the effort to miniaturize sequencing techniques.

One of the primary motivations for developing microchip-based capillary gel electrophoresis technology was to speed up and reduce the cost of sequencing DNA. During the 1990s, the Human Genome Project was begun with the goal of sequencing the entire human genome [4]. The project was actually completed ahead of the original schedule because of new technology (capillary electrophoresis and computational techniques) that was developed and implemented within the time frame of the project. It was recognized by astute researchers that the potential benefits of microfluidic technology (such as decreased sample volumes and sample handling, automation, and decreased analysis time) for highly demanding sequencing applications were worth investing time and effort into developing.

One of the first demonstrations of a microchip device capable of achieving single base resolution for sequencing was published in 1995 by Mathies' group [38]. In their 1995 work, they used polyacrylamide-filled microfluidic channels to fully resolve (at the 0.5 resolution level that is standard for declaring adjacent peaks resolved) single stranded DNA fragments differing by one base out to fragment lengths of about 150 bases. A biochemical reaction is used to generate DNA fragments differing by one base where the DNA fragments are terminated by a fluorescent dye molecule. This fluorescent dye is correlated with the last base in the fragment. For example, fragments ending in the base A are red, fragments ending in the base T are blue,

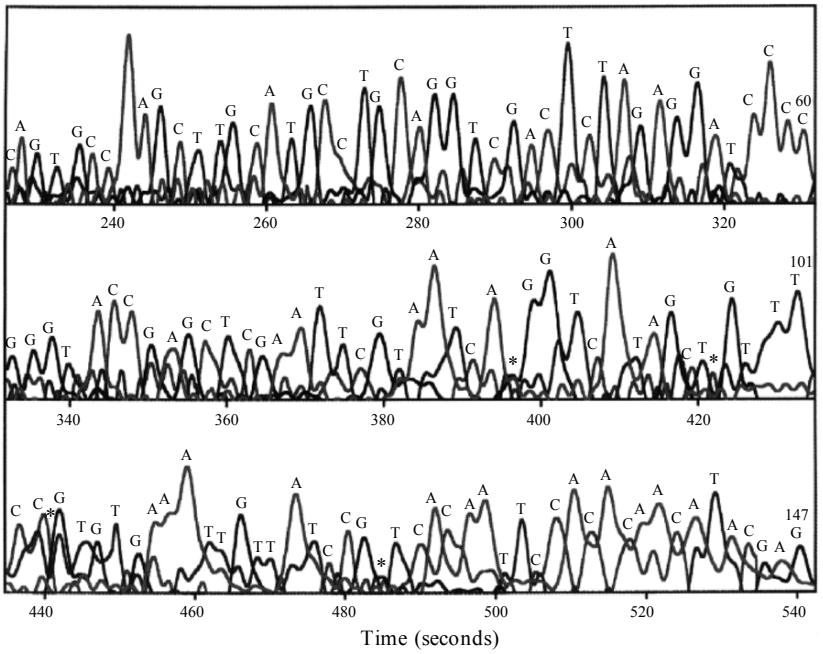


Figure 3.9 A black and white version of four color sequence data obtained using a microchip-based gel electrophoresis experiment, reprinted with permission from [38], Copyright 1995 American Chemical Society. Bases A, G, C, and T would normally be color-coded in the image to correspond to the fluorescent dyes attached to the DNA fragments that end with each base (bases and fluorescent dyes are matched during the sample generation process). The experiment was performed using a polyacrylamide gel in a microfluidic channel. Good resolution of peaks was obtained up to about 150 bases before signal-to-noise in two of the color channels caused peak resolution to degrade below the 0.5 level required to "call" each base. 97% accuracy was obtained for the 150 bases sequenced (asterisks denote the 3 bases that were incorrectly read, at times ~395 sec, ~440 sec, and ~485 sec).

and so on. Sequencing is then a two-dimensional process. One separates DNA fragments by length and correlates the length information with the color information obtained in the detection process. A sample electropherogram from a sequencing experiment is shown in Fig. 3.9. The gel-filled microfluidic channels were capable of a read accuracy of 97% over the fully resolved region of 150 bases. While not quite comparable to gel-filled capillary electrophoresis techniques available at time and used in the Human Genome Project (>99% read accuracy), the gel-filled microchannel results demonstrated that even the technically challenging task of sequencing DNA is possible in microfluidic channels.

Work towards developing microchip-based sequencing systems continues. Additionally, integration of detection and sample preparation steps are being incorporated into these microfluidic systems. The ability to manufacture micro- and nanoscale structures, however, has enabled researchers to explore entirely new separation mechanisms. It is to these new, fundamentally nanoscale systems that we now turn.

3.7 FABRICATED MICRO- AND NANOSTRUCTURES

The miniaturization of conventional bioanalytical techniques proved that microfluidics was a platform capable of competing with existing technology. Currently, work continues along the lines of miniaturizing conventional techniques and developing commercial products [39, 40]. Some of the most exciting bioanalytical work, however, has taken place in micro- and nanofluidic devices containing novel structures. Because of the ability to manufacture structures on the same size scale as the biomolecules, physical interaction mechanisms can be more carefully studied and new interaction mechanisms can be leveraged for separation applications. Below we describe a number of devices that are essentially nanoscale obstacle courses. These devices have found great utility in DNA studies and applications.

3.7.1 Artificial Sieving Matrices

In one of the first reports of a microfabricated sieving matrix, Volkmuth and Austin describe a two-dimensional array of posts through which a variety of sizes of DNA molecules are driven [41]. Posts are patterned with optical lithography and have diameters of 1 μm. The entire device (walls, floor and ceiling, and posts) is made from SiO_2. The post centers are spaced by 2 μm

and they are 150 nm tall. In their work, the authors deliberately shear 100 kb DNA (~30 μm contour length) to produce a DNA sample of mixed lengths.

When compared with an agarose gel typically used for separating long DNA molecules by size, the microfabricated post array corresponds to a very low concentration (~0.05%) gel that would not be stable. The radius of gyration of the longest DNA molecule is about 1.5 μm which is comparable to the distance between the edges of adjacent posts. DNA molecules are observed to move through the post array and periodically collide with posts in such a way as to elongate around the posts in a rope-over-pulley fashion. This configuration results in both arms of the elongated DNA molecule being extended along the direction of the flow. Eventually, the longer arm of DNA will lead the disentanglement of the molecule from the post. Once disengaged from a post, the DNA molecule will relax into its blob configuration as it migrates through the array.

When longer molecules entangle upon posts they must spend more time extracting themselves from the posts. Thus, their net velocity through the array is smaller than that of shorter molecules. With fluorescent microscopy and a CCD, it is possible to observe individual fluorescently labeled DNA molecules in the post array. Individual molecules can be tracked and their velocities in the array determined. In their work, Volkmuth and Austin show that a nearly two-fold velocity difference is observed between the longest (~28 μm or ~100 kb) DNA molecules and the shortest (~3 μm or ~10 kb) DNA molecules.

This early work established the principle of using a microfabricated, artificial gel to separate DNA in essentially the same manner as conventional gels. A few years later, one of the same scientists, along with others, established that pulsed field electrophoresis could be achieved in similar arrays of posts [42]. One advantageous feature of using microchip-based separation techniques is that it is straightforward to observe individual molecules migrating through the microfluidic channels. Thus, it is possible to carefully study physical phenomena such as molecular entanglement with obstructions, molecular response to the electric field, and what happens to molecules elongated in an array of pillars when the direction of the electric field is switch from being essentially parallel to the direction of molecular elongation to being perpendicular to the direction of elongation. This last point is particularly helpful in optimizing pulsed field electrophoresis experiments as there is a balance that must be established between switching frequency, duration, and electric field angles. With microfluidic systems and fluorescence microscopy, one is able to observe individual molecules as these parameters are adjusted and qualitatively get a feel for what is hap-

pening. Alternatively, the effects can be quantified by performing multiple separations using a variety of parameters and determining what combination of parameters produces the best resolution in the shortest time for the widest range of molecular sizes.

Sacrificial layer fabrication has also been used to manufacture microfluidic sieving structures (see Figs. 3.2 and 10). In one of the first demonstrations of the process, Turner *et al.* used a sacrificial layer, patterned with electron beam lithography, to produce 100 nm diameter posts separated by 100 nm gaps extending 400 nm between the floor and ceiling [43]. Figure 3.10 shows two cross-sectional scanning electron micrographs of the structures. The microfluidic elements of the device are made from silicon nitride and polysilicon is used as the sacrificial layer. These devices were used to measure the velocities of linear and circular DNA as it migrated through the pillars under a DC electric field. The results indicate that circular, M13mp8 phage DNA (7 kb) travels more slowly than does linearized lambda phage DNA (48 kb) in the artificial gel. This is in contrast to what one would expect based upon gel electrophoresis of the same fragments.

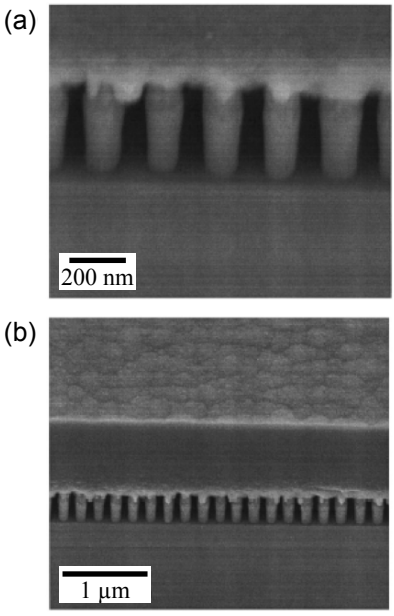


Figure 3.10 Scanning electron micrographs of a cross-section of an artificial sieving matrix fabricated using a sacrificial layer process, reprinted with permission from [43], Copyright 1998 American Institute of Physics. The image is taken at a 45º angle and shows the silicon nitride structures after removal of the sacrificial layer. The pillar aspect ratio is approximately 5:1.

Typically circular DNA moves faster than linear DNA, and shorter DNA moves faster than longer DNA. The authors attribute the altered behavior to supercoiling of the circular DNA that might cause the DNA to entangle more readily upon pillars than linear DNA does. Additionally, if the supercoiling is insufficient to relieve stress in the DNA polymer, then the polymer may actually be stiffer than the linear DNA. This stiffness would probably cause DNA molecules to have a more difficult time moving through the array.

Artificial sieving matrices like those described above are necessary for μTASs as putting gel-based sieving matrices in microfluidic channels is not always possible or desirable. Being able to combine the manufacturing of the microfluidic system with the creation of the sieving matrix greatly simplifies the entire production process. Furthermore, conditions not realizable in gel environments are possible with microfabricated structures. As individual molecules can readily be observed in these environments, polymer properties such as supercoiling and stiffness can be directly measured. Both of these latter properties are relevant for understanding how DNA is packed in cell nuclei and how genes are regulated.

3.7.2 Entropic Recoil

Turner *et al.* used essentially the same devices as described above to harness the configurational entropy of DNA molecules for separation [44]. A DNA polymer, being composed of flexible monomer units (the bases), exists in a relaxed blob configuration when in a spatial environment with dimensions larger than the radius of gyration of the molecule. When forced into an environment with dimensions smaller than its radius of gyration, the DNA molecule is squished into a more confined state. The molecule loses entropy. If the molecule straddles two regions, one in which the molecule is not squished and another in which the molecule is squished, there is an entropy associated with each region. Assuming that the molecule is free to move, it will maximize its entropy by extricating itself from the region of low entropy.

This effect, dubbed entropic recoil, was investigated and used to separate DNA molecules of different lengths [45]. Pillar-filled regions were created using electron beam lithography and the sacrificial layer fabrication technique described previously. Pillars were roughly 35 nm in diameter with a center-to-center spacing of 160 nm resulting in gaps of about 130 nm through which DNA molecules could move. The researchers used long DNA molecules (T2, T4, or lambda DNA) which have radii of gyration

of at least 300 nm. Thus, DNA molecules were forced to elongate to enter the pillar-filled region.

In entropic recoil experiments, DNA molecules are brought to the interface between a pillar-filled region and a pillar-free region (which has dimensions large enough that DNA molecules are in their relaxed, blob configuration). DNA is forced to partially extend into the pillar-filled region by applying an electric field. When the driving field is turned off, molecules that straddle the interface extricate themselves from the low entropy-pillar region. An example of this process is shown in Fig. 3.11.

To separate molecules by length using this technique, a multi-step electric field program is used. First, a collection of molecules of different lengths is brought to the interface region using an electric field small enough not to drive the molecules into the pillar region. Then, a higher electric field is briefly pulsed to drive a portion of the molecules into the pillar region. The high field pulse duration is chosen to be long enough to push the shortest molecules into the pillar region. The short molecules that find themselves entirely within the pillar region do not see the high entropy region and remain within the pillar matrix. Longer molecules that straddle the boundary experience the pull of the high entropy region and recoil into the high entropy region where they can relax. Subsequent gathering steps and high field pulses drive the short molecules already in the pillars through the pillar region and separate them from the long molecules. This process can be repeated with suitable high field pulses so as to separate molecules of different lengths from each other. Cabodi *et al*. separated T2 DNA (167 kb) and T7 DNA (40 kb) in an entropic recoil experiment [45]. Their experimental demonstration of the technique took hours, though this time could likely be significantly reduced by choosing electric field pulse durations more suited to the time required to drive DNA molecules into the pillar matrix. The authors speculate that based upon their results, a multi-stage entropic recoil separation device could be developed capable of separating DNA molecules in under an hour with 1% length resolution.

In the aforementioned entropic recoil experiments, polymers were observed to herniate in the pillars. This reduced the resolution of the entropic recoil separation technique. Recent work within the Craighead research group has focused on nanometer scale channels that preserve the advantageous entropy-reducing, confining effect but prevent molecule motion perpendicular to the driving electrical field. Using nanochannels as opposed to pillar-arrays prevents looping and herniation. These newer devices were fabricated with electron beam lithography directly in fused silica substrates, circumventing the need to use sacrificial layer fabrication. Mannion *et al*.

showed that similar recoil and frictional effects are obtained in these nanochannel structures [46]. Additionally, because the nanochannel design prevented herniation, a much more detailed study of the motion of the DNA polymers in the confining regions was permitted. The scientists discovered that molecules enter nanochannels stretched beyond their relaxed contour length. Furthermore, molecules often enter nanometer-scale spaces significantly folded. This means that in addition to a de-stretching process (caused by the initial over-stretching), and a recoil process (caused when the molecule straddles the boundary), there is also an unfolding process in which the molecule behaves much like a string sliding off of a table. An example

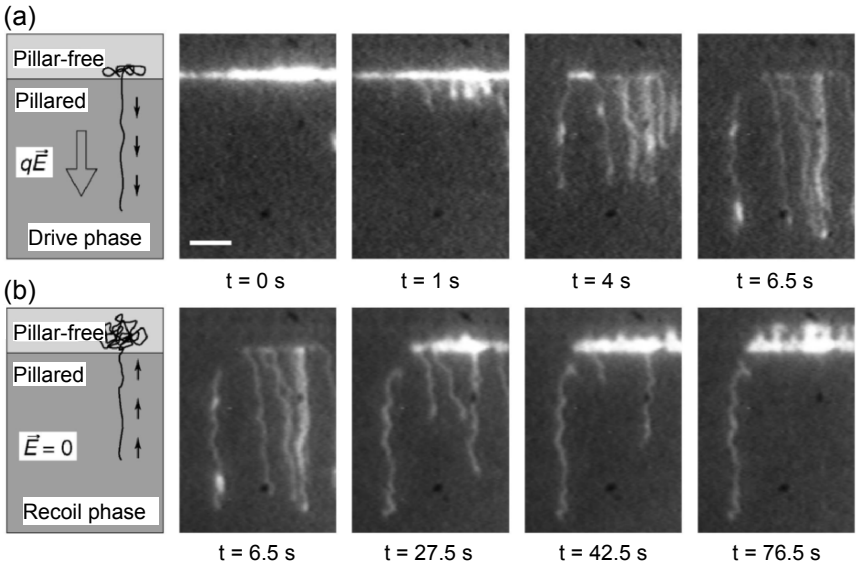


Figure 3.11 Schematics of the drive and recoil phases shown with fluorescence micrographs of T2 DNA molecules in the drive and recoil phases, reprinted with permission from [44], Copyright 2002 American Physical Society. (a) In the drive phase, a high electric field is used to enable the molecules to overcome the energy barrier posed by the confining pillars. The boundary between the pillar-free and pillared regions shown in the schematic is at the same place as in the micrographs to the right of the schematic. During the drive phase, molecules are driven into the pillared region. The initial insertion is probabilistic in nature, so not all molecules enter at the same time. Thus, for a given drive duration, some molecules may end up straddling the boundary. (b) In the recoil phase the driving field is turned off. Those molecules left straddling the boundary are essentially driven out of the pillared region by the high entropy of the pillar-free region. After 76.5 sec, all molecules straddling the boundary have recoiled and one molecule that was fully inserted remains in the pillared region.

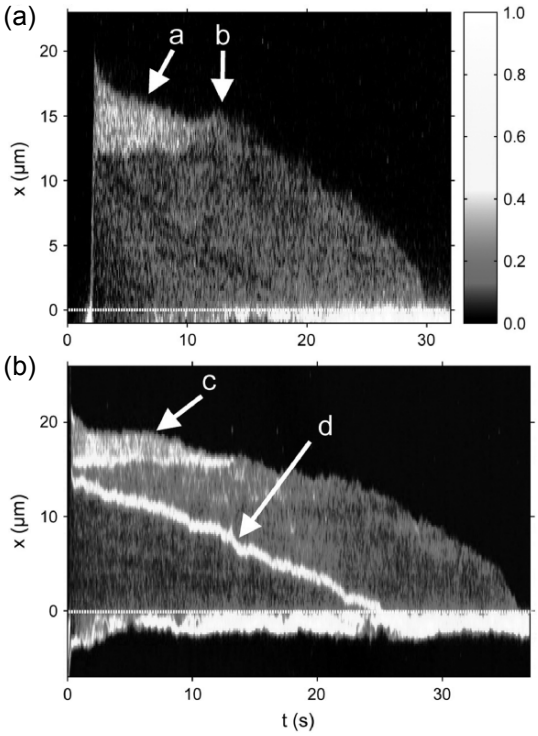


Figure 3.12 A time-series of intensity images along nanochannels in which T4 DNA molecules are inserted, reprinted with permission from [46], Copyright 2006 Biophysical Society. The vertical dimension shows the intensity along the length of the molecule. Bright parts of the graph correspond to more dyed DNA than do dim parts of the graph. The horizontal dimension shows the time evolution of the molecule in the nanochannel. The dashed white line at x = 0 marks the boundary between the regions of high and low entropy. (a) A DNA molecule is driven into the nanochannel region at about t = 2 sec. It is initially inserted about 20 μm into the channel and straddles the high and low entropy interface (the dashed line at x = 0). Once the molecule is inserted the field is immediately turned off and the molecule is left to recoil. Note the high intensity region near the far end of the molecule (13-20 μm). This portion of the molecule is folded-over on itself, an occurrence that is not uncommon when long molecules are forced into nanochannels. The molecule simultaneously undergoes relaxation, unfolding, and recoil. The relaxation process is very fast and most evident from t = 2 sec until about t = 5 sec. The unfolding process occurs until 'b', at which point the remaining folded portion flops open. From 'b' onwards, the motion is entirely recoil as the molecule extricates itself from the nanochannel. (b) Another molecule inserted into a nanochannel undergoing relaxation, unfolding, and recoil. Additionally, the molecule contains a knot that never unties. The knot is apparent as the short, bright segment that recoils with the leading end of the molecule (highlighted by the arrow 'd').

of an unfolding process coupled with the entropic recoil process is shown in Fig. 3.12. Similar channels were used by Reccius *et al.* to investigate the physics of DNA molecules in nanochannels as they are compressed against an impenetrable constriction and then allowed to expand back to their full contour length [47].

Entropic recoil-based separation of DNA is an excellent example of a technique made possible through the use of micro- and nanofabrication techniques. No macroscale analogue technique exists. The physics of the polymers – their stiffness and entropy – can be studied directly with these devices. Additionally, long DNA molecules can be readily separated. A multi-stage entropic recoil separation device has the potential to eliminate diffusion-based band broadening, which is a ubiquitous resolution limiting factor in conventional separation techniques.

3.7.3 Entropic Trapping

Entropy can be used in other ways to effect a molecular separation based upon length. Extensive theoretical work was done in the 80s and 90s that described the effects of confining environments on polymer diffusion, entanglement, and mobility [48-52]. Experiments in gels confirmed that when polyelectrolytes such as DNA are forced to move through entropically restrictive environments, the polymer mobility is not accurately described by either the Ogston sieving model or the reptation model [53-56]. Instead, a regime of entropic trapping exists where molecules are effectively trapped in pores larger than their relaxed blob size and only periodically hop through restrictive areas into new pores.

Well-characterized, "ideal" periodic structures were used in modeling the motion of polymers through restrictive environments, yet the experimental work involved random gel-matrices similar to those conventionally used to separate DNA molecules by length. It was not until the work of Han *et al*. that a lithographically defined array of entropic traps was used to show that DNA molecules are trapped and can be separated by length in a real "ideal" structure [25]. In their first work, Han *et al*. showed that 40 kb DNA molecules and 160 kb DNA molecules exhibit significantly different trapping times when forced to hop across barriers and into large "pores". Subsequent research showed that the mobility difference caused by the difference in delay times at each trap was sufficiently large to separate these molecules by length in a device 1.5 cm long containing thousands of traps. In a variety of experiments with so-called entropic trap arrays, Han *et al*. were able to separate a wide range of DNA lengths [57, 58]. T2 and T7 phage

DNA (160 kb and 40 kb respectively) were separated in 15 min in an entropic trap array. A comparable experiment using pulsed field gel electrophoresis would have required between 12 and 24 hours. Additional experiments showed that DNA from 5 kb to 40 kb could be separated in about 30 min with resolution comparable to that obtained using gel electrophoresis.

As originally described, entropic trapping relies upon an energy barrier and interfacial contact between a molecule and that energy barrier. DNA is a flexible polymer that has a characteristic radius of gyration set by the total number of bases in the polymer chain. When a relaxed blob of DNA is forced against a constriction in a microfluidic channel by an electric field, the blob makes interfacial contact with the restriction. This contact area is proportional to the radius of the blob: larger molecules "contact" the gap more than do small molecules. The molecule is subjected to two forces: an electric driving force and an entropic force discouraging it from entering the restriction. This free energy landscape is depicted in Fig. 3.13. The entropic barrier energy is sharp and significant. However, once a portion of the molecule overcomes the barrier, the large electric field in the restricted region pulls the entire molecule into the restriction and toward the next open region. Because the DNA molecule is "free draining" to both the electric field and whatever fluid flow is present, the process of entering the restriction is local in nature. Portions of the molecule randomly diffuse near the entrance to the restriction and when one happens to find itself on the downhill side of the energy landscape, it is pulled in and the rest of the molecule follows. To summarize: larger molecules have more interfacial area with the restriction and make more attempts to enter the high field shallow region. Consequently, larger molecules move between restrictions more readily than do smaller molecules.

Quantitatively, the following model was proposed to describe the entropic trapping process [59]. The energy landscape is composed of an electric and entropic component

$$F \sim xT - x^2E, \tag{3.4}$$

where x is the distance the molecule has penetrated the shallow region, T is the temperature, and E is the electric field. Equation (3.4) represents the energy barrier that must be overcome to pass through the restriction. The probability of escape is then given by

$$P = a\exp(-\Delta F/k_BT), \tag{3.5}$$

where k_B is the Boltzmann constant, and the prefactor a depends upon the number of escapes attempted (that is, the interfacial area). Equation (3.5) is

the crux of the size-dependence in entropic trapping: the prefactor contains all of the size-dependence while the exponential term represents the thermodynamics of the energy barrier and does not depend on any aspect of the molecule. The time molecules remained trapped is given by

$$\tau = \tau_0 \exp(-\Delta F/k_B T) = \tau_0 \exp(b/E_s k_B T), \tag{3.6}$$

where τ_0 is the size-dependent prefactor, b is a geometric constant, and E_s is the electric field at the edge of the shallow region. From the trapping time, the mobility can be written

$$\mu = \mu_0/(1 + \tau/t), \tag{3.7}$$

where t is the time required for a molecule to move through a distance equal to the length of the restriction if the restriction were not there. The length

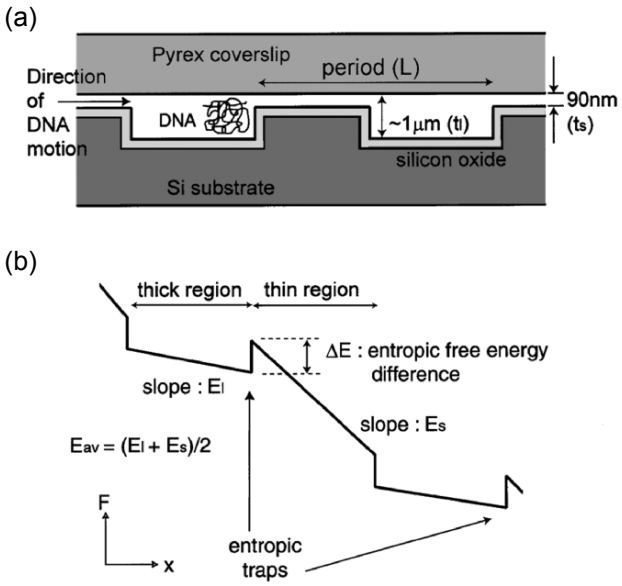


Figure 3.13 (a) A cross-sectional schematic of an entropic trap device, reprinted with permission from [59], Copyright 1999 American Physical Society. DNA is "trapped" in deep regions of high entropy and forced against shallow regions. The shallow regions are smaller than the radius of gyration of the molecule, and entry into those regions would require sacrificing configurational entropy. (b) The energy landscape of consecutive entropic traps. Molecules require considerable energy to overcome the barriers. The energy is attained when random, "beachhead" events extend a portion of the molecule a critical distance into the shallow region. Once a critical insertion length is obtained, the electrical force is sufficient to overcome the entropic force and the molecule jumps the barrier.

dependence in the mobility comes from τ and the length dependence is such that longer molecules migrate faster than shorter ones. Later work by Han *et al.* confirmed the applicability of the model over a wide range of DNA lengths and discussed optimizing the device for maximizing resolution and separating chromosomal-length DNA [57, 58].

After the original entropic trap array separation experiments were published, a new wave of theoretical and experimental papers surfaced [60-62]. In the first of such theoretical comparisons, Tessier *et al.* confirmed that the model proposed by Han was essentially correct in that Monte Carlo simulations based upon Han's model produced results qualitatively similar to the experimental results [60]. The shapes of the experimental and theoretical mobility versus electric field graphs, for instance, are identical. Furthermore, the essential physical mechanism of escape – herniation of a portion of the molecule into the restrictive region leading to the entire molecule moving through the gap – is confirmed by the theoretical results. Chen and Escobedo examined wider ranges of molecule sizes and electric field strengths than were examined experimentally [61]. Even though their results indicate that the physical model proposed by Han *et al.* might not fully describe the system at low electric fields or for very short DNA molecules, their simulations match Han's results for the conditions used in experiments. Additional simulations by Streek *et al.* showed that there might be a second, independent physical mechanism contributing to the time molecules spend migrating between restrictions [62]. In their simulations, they show that while in transit across the deep regions, molecules can diffuse out of the high field region of the channel. That is, molecules can diffuse into the corners of the entropic traps where the electric field is weak. In the corners, molecules are effectively left to diffuse back into a region of higher electric field. This phenomenon operates on a much slower time scale than the mechanism proposed by Han *et al.* and was probably not observed because of the sizes of molecules and electric field strengths used in their experiments.

Entropic traps separate DNA molecules by length and have been used to separate quite long DNA molecules. This artificial sieving matrix is straightforward to fabricate and has great potential as a separation matrix within a μTAS environment. The entropic trap array could potentially be used to handle the large DNA molecules that would be released from a cell lysed within a chip. Judicious use of entropic trap-like restrictions could also be used to separate or filter cellular debris and proteins from the DNA. Recent work has shown that by using very thin shallow regions in entropic trap array-like devices, one can separate very short DNA molecules and

proteins [63, 64]. Thus, the general model of patterning very narrow restrictions followed by relatively large gaps can be used for biological applications across a wide range of biomolecular sizes. This broad-utility is exactly what is needed for μTAS systems that start with whole cells and end with useful analytical results.

3.7.4 Asymmetric Potentials

In one of the famous Feynman lectures, Richard Feynman discusses a pawl and ratchet device that under certain conditions turns in the expected direction, but in other, quite realizable, conditions, the ratchet device can be "made" to turn in the opposite direction [65]. It has been proposed theoretically that asymmetric potentials, with shapes that look like the teeth of a ratchet, could be used to separate biomolecules [66-71]. A number of physical properties of the biomolecules could in principle be used to separate them with asymmetric potentials, though the molecular diffusion constant is the focus of the cited studies. Essentially, the proposed devices act as diffusion rectifiers, allowing net motion in only one direction. Another similarity of these proposals is that the asymmetric potential is established via arrays of asymmetric structures.

Figure 3.14 shows examples from each of the proposals. Molecules are driven through an array of obstacles (electrophoresis is typically used). The obstacles are asymmetric to the flow. That is, as one moves from row to row, moving to the left looks different than moving to the right. The asymmetry can be accomplished by "tilting" the structures relative to the direction the molecules are driven (through the fabrication process), or by bringing the molecule flow into the array slightly skewed. As molecules move from row to row they diffuse laterally (they also diffuse forwards and backwards but this motion is not important for the separation). If they are moving fast enough, then it is highly unlikely that they will move to the left by an entire column [for the structures shown in Fig. 3.14(b) and (c)]. It is not so unlikely that they will move sufficiently far to the right to jump to the next column to the right. The likelihood of jumping to the next column depends upon the ratio of the diffusion constant to the product of the velocity and the structure width. Note that if the molecules move very slowly, then they will have a significant chance of moving to the left column (and forward and backward diffusion will start to matter). Also, if the molecules move very quickly, then they will never have a chance to diffuse to the next right column. The sweet spot of the molecular velocity for the separation is entirely dependent upon the diffusion constants of the molecules and the dimensions of the

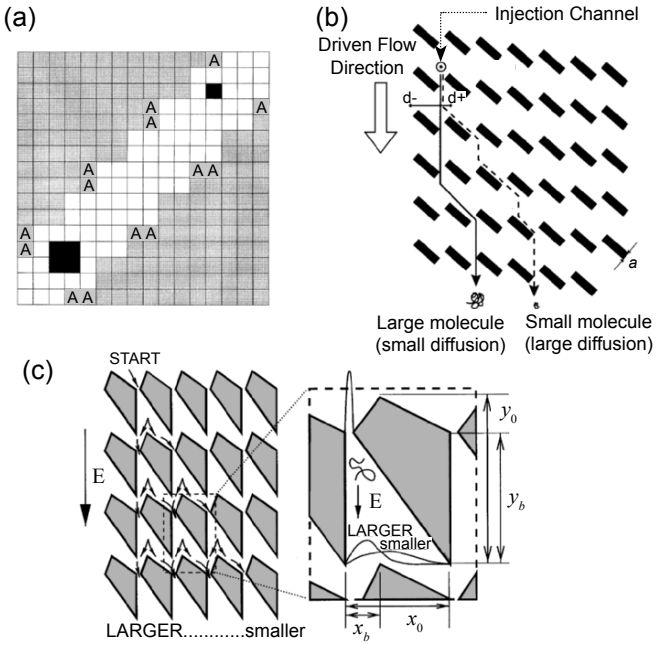


Figure 3.14 (a) The channel modeled by Slater *et al.*, reprinted with permission from [68], Copyright 1997 American Physical Society. The shaded squares are the boundaries of the channel and when the squares labeled with 'A' are removed, the channel is asymmetric. In the symmetric channel, Slater *et al.* showed that for 2×2 and 1×1 hard, diffusing objects can be made to move along the diagonal even when the net driving field is zero. Furthermore, when the channel is made asymmetric, there are critical AC field pulses (field strength times time) such that the 2×2 and 1×1 objects move in different directions. Again, this result is for a time-averaged mean field of zero. (b) The asymmetric diffusion array of Cabodi *et al.*, reprinted with permission from [28], Copyright 2002 Wiley-VCH Verlag GmbH & Co KG. Once molecules pass the downstream edge of an obstacle, they must diffuse a distance d− in order to skip a column to the left, d+ to skip a column to the right, or continue to move down along the same column. Due to the asymmetric shape of the obstacles, it is nearly impossible that molecules move to the left d−. Larger molecules have a significantly smaller chance of moving to the right a distance d+ than do smaller molecules with larger diffusion constants. Thus, smaller molecules tend to skip columns to the right, whereas larger molecules tend to move straight down the array. (c) A different array of asymmetric obstacles proposed by Ertas, reprinted with permission from [70], Copyright 1998 American Physical Society. In the right panel, the probabilities of spatial location are shown for larger and smaller diffusing species. The probability that larger molecules move beyond the peak of the asymmetric spatial "potential" is much smaller than the probability that smaller molecules do. This geometry is similar to that actually employed by Boxer to separate charged molecules in a lipid membrane.

device (both the structure dimensions and the spacing between structures). The diffusion array is an excellent example of a separation device that is only possible with carefully designed micro- and nanoscale structures.

In one of the first experimental demonstrations of this phenomenon, Chou *et al*. showed that DNA molecules with different diffusion constants (lengths) had significantly different probabilities of moving from column to column across an array of asymmetric obstacles [72]. 15 kb DNA exhibited more lateral displacement than did 33.5 kb DNA when they were driven electrophoretically through the array. These researchers encountered problems when loading mixed DNA samples, so they were unable show that their device could be used for separating DNA. Though a true separation of the species was not performed in their experiments the authors note that the resolution of the displacements is 6% of the molecular length for the lengths studied, suggesting that about fourteen equally size-spaced DNA molecules could be separated using this technique.

Shortly after Chou *et al*. published their results, Cabodi *et al*. solved the technical problem of sample loading that prevented the earlier researchers from separating multiple DNA sizes from each other [28]. Their work describes separation of T2 DNA from T7 DNA in essentially the same type of asymmetric array as [72] used. Out-of-plane sample injection permitted a thin stream of molecules to be drawn from the sample reservoir and into the separation array portion of the device. 3.8 cm from the injection point (about 2200 rows), the two DNA molecules were observed to be separated by about 440 μm. Based upon the theoretical model proposed by Austin and Duke [69], the molecules should have been nearly twice as far apart (750 μm). The authors note that the discrepancy is likely due to the observed deformation of the molecules as they interacted with the obstacles. This would cause the molecules to deform into non-spherical shapes or elongate, either of which would certainly alter the diffusion constant of the molecule. Additionally, the original model supposed conducting obstacles which is not the case in the experiment. The insulating obstacles cause the electric field lines to deform leading to electric field gradients in each unit cell of the array. These gradients could contribute to molecular deformation such as elongation [73].

van Oudenaarden and Boxer used an array of asymmetric obstacles to separate charged molecules in lipid membranes [74]. Lipids are polar molecules that are of biological interest because they constitute cell membranes. Many proteins reside in and interact with lipid bilayers. Scientists often use lipid bilayers to study molecular diffusion. In their lipid membrane experiments, van Oudenaarden and Boxer observed that diffusion was rectified

by an array of asymmetric obstacles and net lateral motion was observed even though the driving force (an electric field) was oriented directly along a column. In addition to supporting the theoretical aspects of the previously discussed DNA separation experiments, the lipid experiments prove that the phenomenon of rectified diffusion is generalizable and useful across a range of biological systems.

It should be pointed out that all of the previously described molecular separation experiments (those described prior to this subsection) are essentially discrete in nature. That is, biomolecules are loaded into a device and a band of molecules is defined. The band is then separated based upon some interaction with the device or environment within the device. While bands may be loaded and separated in rapid succession, each experiment has a definite beginning and a definite end. The primary reason for this "discreteness" is that all molecules must traverse the same path as they are separated and detected. Note that this is true of conventional molecular separation experiments as well (slab gel electrophoresis, for example). All of the diffusion array devices described in this subsection can be operated in a continuous mode, however, setting them apart from most other separation schemes. Molecules are introduced to the obstacle array from a reservoir and are separated laterally in space as they move through the array. There is no need to define plugs of molecules as the separated molecules move along entirely different physical paths. In addition to the continuous nature afforded by this scheme, another distinct advantage of laterally separating the molecules is that sample recovery is potentially much easier than in the discrete separation process. If one wanted to do analysis, purification, and sample recovery, then an asymmetric obstacle course would be very useful.

3.8 CONCLUSIONS

We have described one version of the history of microfluidics as it progressed from microchip-based capillary electrophoresis experiments to novel structures capable of elucidating new physics and permitting previously unrealizable methods of separating biomolecules. The story has focused on the myriad of applications made possible through creative uses of standard CMOS fabrication techniques. None of the advances in microfluidics would have been possible were it not for the preceding fabrication efforts of the solid state physics and electronics communities.

Our focus has been highly separation-centric, and we have omitted discussing exciting work that is taking place with applications as diverse as

microfluidic-based cell and tissue culture [75-77] and micro-chemical reactors [78-82]. There are numerous excellent reviews available covering a wide range of topics within the discipline of biological applications of micro- and nanofluidic devices. Many of the devices discussed herein are further discussed in [1, 7]. A variety of microchip-based capillary electrophoresis-like devices and many applications of these devices are discussed in [83]. In addition to separating biomolecules by size through some kind of charge variability or frictional interaction, microchip capillary electrophoresis affords other possibilities for analyzing biological samples. Many of these are discussed in [84]. From the standpoint of examining the new directions in which standard CMOS-type fabrication tools are being employed, our focus on separation techniques tells essentially the same story as if we had considered any of these other applications.

One of the primary future applications of micro- and nanofluidics is to develop μTASs. This effort will almost certainly involve an even stronger bond between the CMOS fabrication, optics, electronics, fluidics, and systems integration communities. Packaging and power requirements may prove difficult, though even these issues are currently being addressed [85]. Current engineering efforts include integrating microfluidic chips with micro-optical lenses, lasers, and detectors, all of which can be miniaturized in some way ([21, 86-89], and see Chapter 4 for examples. Once substantially all of these components are put on-chip, these μTASs could be widely deployed for applications such as rapid DNA sequencing and inexpensive medical diagnostics.

ACKNOWLEDGMENT

These authors would like to thank all of the authors whose work is referenced herein, especially those who have graciously allowed us to reprint their figures.

REFERENCES

[1] J. O. Tegenfeldt, C. Prinz, H. Cao, R. L. Huang, R. H. Austin, S. Y. Chou, E. C. Cox, and J. C. Sturm, "Micro- and nanofluidics for DNA analysis," Analytical and Bioanalytical Chemistry, vol. 378, pp. 1678-1692, 2004.

[2] P. S. Dittrich and A. Manz, "Single-molecule fluorescence detection in microfluidic channels - the Holy Grail in mu TAS?," Analytical and Bioanalytical Chemistry, vol. 382, pp. 1771-1782, 2005.

[3] A. Manz, N. Graber, and H. M. Widmer, "Miniaturized Total Chemical-Analysis Systems - a Novel Concept for Chemical Sensing," Sensors and Actuators B-Chemical, vol. 1, pp. 244-248, 1990.
[4] http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml.
[5] D. L. Nelson and M. M. Cox, "Lehninger Principles of Biochemistry." New York: Worth, 2000, pp. 325-359.
[6] D. L. Nelson and M. M. Cox, "Lehninger Principles of Biochemistry." New York: Worth, 2000, pp. 115-153.
[7] C. F. Chou, R. H. Austin, O. Bakajin, J. O. Tegenfeldt, J. A. Castelino, S. S. Chan, E. C. Cox, H. Craighead, N. Darnton, T. Duke, J. Y. Han, and S. Turner, "Sorting biomolecules with microdevices," Electrophoresis, vol. 21, pp. 81-90, 2000.
[8] N. Kaji, Y. Tezuka, Y. Takamura, M. Ueda, T. Nishimoto, H. Nakanishi, Y. Horiike, and Y. Baba, "Separation of long DNA molecules by quartz nanopillar chips under a direct current electric field," Anal Chem, vol. 76, pp. 15-22, 2004.
[9] D. L. Nelson and M. M. Cox, "Lehninger Principles of Biochemistry." New York: Worth, 2000, pp. 3-18.
[10] D. L. Nelson and M. M. Cox, "Lehninger Principles of Biochemistry." New York: Worth, 2000, pp. 1020-1068.
[11] D. L. Nelson and M. M. Cox, "Lehninger Principles of Biochemistry." New York: Worth, 2000, p. 196.
[12] D. L. Nelson and M. M. Cox, "Lehninger Principles of Biochemistry." New York: Worth, 2000, pp. 1132-1133.
[13] J. D. Watson and F. H. C. Crick, "Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid," Nature, vol. 171, pp. 737-738, 1953.
[14] M. Doi, "Introduction to Polymer Physics." Oxford: Oxford University Press, 2001, pp. 1-19.
[15] C. R. Calladine and H. R. Drew, "Understanding DNA: The Molecule and How it Works," 2nd ed. San Diego: Academic Press, Inc., 1997, pp. 1-15.
[16] C. G. Baumann, S. B. Smith, V. A. Bloomfield, and C. Bustamante, "Ionic effects on the elasticity of single DNA molecules," Proc Natl Acad Sci U S A, vol. 94, pp. 6185-90, 1997.
[17] N. C. Stellwagen, C. Gelfi, and P. G. Righetti, "The free solution mobility of DNA," Biopolymers, vol. 42, pp. 687-703, 1997.
[18] C. Heller, G. W. Slater, P. Mayer, N. Dovichi, D. Pinto, J. L. Viovy, and G. Drouin, "Free-solution electrophoresis of DNA," Journal of Chromatography A, vol. 806, pp. 113-121, 1998.
[19] J. Noolandi, "A New Concept for Sequencing DNA by Capillary Electrophoresis," Electrophoresis, vol. 13, pp. 394-395, 1992.
[20] G. W. Slater, C. Desrulsseaux, S. J. Hubert, J. F. Mercier, J. Labrie, J. Boileau, F. Tessier, and M. P. Pepin, "Theory of DNA electrophoresis: A look at some current challenges," Electrophoresis, vol. 21, pp. 3873-3887, 2000.

[21] J. R. Webster, M. A. Burns, D. T. Burke, and C. H. Mastrangelo, "Monolithic capillary electrophoresis device with integrated fluorescence detector," Analytical Chemistry, vol. 73, pp. 1622-1626, 2001.
[22] R. Riehn and R. H. Austin, "Wetting micro- and nanofluidic devices using supercritical water," Analytical Chemistry, vol. 78, pp. 5933-5934, 2006.
[23] S. Pennathur and J. G. Santiago, "Electrokinetic transport in nanochannels. 2. Experiments," Anal Chem, vol. 77, pp. 6782-9, 2005.
[24] S. Pennathur and J. G. Santiago, "Electrokinetic transport in nanochannels. 1. Theory," Anal Chem, vol. 77, pp. 6772-81, 2005.
[25] J. Han and H. G. Craighead, "Entropic trapping and sieving of long DNA molecules in a nanofluidic channel," Journal of Vacuum Science & Technology a-Vacuum Surfaces and Films, vol. 17, pp. 2142-2147, 1999.
[26] H. A. Stone, A. D. Stroock, and A. Ajdari, "Engineering flows in small devices: Microfluidics toward a lab-on-a-chip," Annual Review of Fluid Mechanics, vol. 36, pp. 381-411, 2004.
[27] R. F. Probstein, "Physicochemical Hydrodynamics: An Introduction," pp. 185-198, 1989.
[28] M. Cabodi, Y. F. Chen, S. W. Turner, H. G. Craighead, and R. H. Austin, "Continuous separation of biomolecules by the laterally asymmetric diffusion array with out-of-plane sample injection," Electrophoresis, vol. 23, pp. 3496-503, 2002.
[29] D. J. Harrison, A. Manz, Z. H. Fan, H. Ludi, and H. M. Widmer, "Capillary Electrophoresis and Sample Injection Systems Integrated on a Planar Glass Chip," Analytical Chemistry, vol. 64, pp. 1926-1932, 1992.
[30] S. C. Jacobson, R. Hergenroder, L. B. Koutny, R. J. Warmack, and J. M. Ramsey, "Effects of Injection Schemes and Column Geometry on the Performance of Microchip Electrophoresis Devices," Analytical Chemistry, vol. 66, pp. 1107-1113, 1994.
[31] D. J. Harrison, K. Fluri, K. Seiler, Z. H. Fan, C. S. Effenhauser, and A. Manz, "Micromachining a Miniaturized Capillary Electrophoresis-Based Chemical-Analysis System on a Chip," Science, vol. 261, pp. 895-897, 1993.
[32] Y. F. Cheng and N. J. Dovichi, "Subattomole amino acid analysis by capillary zone electrophoresis and laser-induced fluorescence," Science, vol. 242, pp. 562-4, 1988.
[33] S. C. Jacobson, A. W. Moore, and J. M. Ramsey, "Fused Quartz Substrates for Microchip Electrophoresis," Analytical Chemistry, vol. 67, pp. 2059-2063, 1995.
[34] D. F. Swaile and M. J. Sepaniak, "Determination of Metal-Ions by Capillary Zone Electrophoresis with on-Column Chelation Using 8-Hydroxyquinoline-5-Sulfonic Acid," Analytical Chemistry, vol. 63, pp. 179-184, 1991.
[35] M. N. Albargheuthi and A. E. Barron, "Polymeric matrices for DNA sequencing by capillary electrophoresis," Electrophoresis, vol. 21, pp. 4096-4111, 2000.

[36] C. S. Effenhauser, A. Paulus, A. Manz, and H. M. Widmer, "High-Speed Separation of Antisense Oligonucleotides on a Micromachined Capillary Electrophoresis Device," Analytical Chemistry, vol. 66, pp. 2949-2953, 1994.
[37] A. T. Woolley and R. A. Mathies, "Ultra-high-speed DNA fragment separations using microfabricated capillary array electrophoresis chips," Proc Natl Acad Sci U S A, vol. 91, pp. 11348-52, 1994.
[38] A. T. Woolley and R. A. Mathies, "Ultra-High-Speed DNA-Sequencing Using Capillary Electrophoresis Chips," Analytical Chemistry, vol. 67, pp. 3676-3680, 1995.
[39] C. Lifesciences, "The Application Notebook."
[40] A. Technologies, "Agilent 2100 bioanalyzer."
[41] W. D. Volkmuth and R. H. Austin, "DNA electrophoresis in microlithographic arrays," Nature, vol. 358, pp. 600-2, 1992.
[42] T. A. Duke, R. H. Austin, E. C. Cox, and S. S. Chan, "Pulsed-field electrophoresis in microlithographic arrays," Electrophoresis, vol. 17, pp. 1075-9, 1996.
[43] S. W. Turner, A. M. Perez, A. Lopez, and H. G. Craighead, "Monolithic nanofluid sieving structures for DNA manipulation," Journal of Vacuum Science & Technology B, vol. 16, pp. 3835-3840, 1998.
[44] S. W. Turner, M. Cabodi, and H. G. Craighead, "Confinement-induced entropic recoil of single DNA molecules in a nanofluidic structure," Phys Rev Lett, vol. 88, pp. 128103, 2002.
[45] M. Cabodi, S. W. Turner, and H. G. Craighead, "Entropic recoil separation of long DNA molecules," Anal Chem, vol. 74, pp. 5169-74, 2002.
[46] J. T. Mannion, C. H. Reccius, J. D. Cross, and H. G. Craighead, "Conformational analysis of single DNA molecules undergoing entropically induced motion in nanochannels," Biophys J, vol. 90, pp. 4538-45, 2006.
[47] C. H. Reccius, J. T. Mannion, J. D. Cross, and H. G. Craighead, "Compression and free expansion of single DNA molecules in nanochannels," Phys Rev Lett, vol. 95, pp. 268101, 2005.
[48] A. Baumgartner and M. Muthukumar, "A Trapped Polymer-Chain in Random Porous-Media," Journal of Chemical Physics, vol. 87, pp. 3082-3088, 1987.
[49] M. Muthukumar and A. Baumgartner, "Effects of Entropic Barriers on Polymer Dynamics," Macromolecules, vol. 22, pp. 1937-1941, 1989.
[50] M. Muthukumar and A. Baumgartner, "Diffusion of a Polymer-Chain in Random-Media," Macromolecules, vol. 22, pp. 1941-1946, 1989.
[51] G. I. Nixon and G. W. Slater, "Entropic trapping and electrophoretic drift of a polyelectrolyte down a channel with a periodically oscillating width," Physical Review E, vol. 53, pp. 4969-4980, 1996.
[52] G. W. Slater and S. Y. Wu, "Reptation, Entropic Trapping, Percolation, and Rouse Dynamics of Polymers in Random-Environments," Physical Review Letters, vol. 75, pp. 164-167, 1995.

[53] L. Liu, P. S. Li, and A. S. Asher, "Crystalline colloidal array of water voids in hydrogels: Direct evidence for entropic trapping of flexible polymers," Journal of the American Chemical Society, vol. 121, pp. 4040-4046, 1999.
[54] L. Liu, P. S. Li, and S. A. Asher, "Entropic trapping of macromolecules by mesoscopic periodic voids in a polymer hydrogel," Nature, vol. 397, pp. 141-144, 1999.
[55] J. Rousseau, G. Drouin, and G. W. Slater, "Entropic trapping of DNA during gel electrophoresis: Effect of field intensity and gel concentration," Physical Review Letters, vol. 79, pp. 1945-1948, 1997.
[56] D. L. Smisek and D. A. Hoagland, "Electrophoresis of Flexible Macromolecules - Evidence for a New Mode of Transport in Gels," Science, vol. 248, pp. 1221-1223, 1990.
[57] J. Han and H. G. Craighead, "Separation of long DNA molecules in a microfabricated entropic trap array," Science, vol. 288, pp. 1026-1029, 2000.
[58] J. Y. Han and H. G. Craighead, "Characterization and optimization of an entropic trap for DNA separation," Analytical Chemistry, vol. 74, pp. 394-401, 2002.
[59] J. Han, S. W. Turner, and H. G. Craighead, "Entropic trapping and escape of long DNA molecules at submicron size constriction," Physical Review Letters, vol. 83, pp. 1688-1691, 1999.
[60] F. Tessier, J. Labrie, and G. W. Slater, "Electrophoretic separation of long polyelectrolytes in submolecular-size constrictions: A Monte Carlo study," Macromolecules, vol. 35, pp. 4791-4800, 2002.
[61] Z. Chen and F. A. Escobedo, "Simulation of chain-length partitioning in a microfabricated channel via entropic trapping," Molecular Simulation, vol. 29, pp. 417-425, 2003.
[62] M. Streek, F. Schmid, T. T. Duong, and A. Ros, "Mechanisms of DNA separation in entropic trap arrays: a Brownian dynamics simulation," Journal of Biotechnology, vol. 112, pp. 79-89, 2004.
[63] J. P. Fu, J. Yoo, and J. Y. Han, "Molecular sieving in periodic free-energy landscapes created by patterned nanofilter arrays," Physical Review Letters, vol. 97, 2006.
[64] J. P. Fu, P. Mao, and J. Y. Han, "Nanofilter array chip for fast gel-free biomolecule separation," Applied Physics Letters, vol. 87, 2005.
[65] R. P. Feynman, R. B. Leighton, and M. Sands, "The Feynman Lectures on Physics," in The Feynman Lectures on Physics, vol. 1. Reading: Addison-Wesley, 1963, pp. 46-1-9.
[66] J. Rousselet, L. Salome, A. Ajdari, and J. Prost, "Directional Motion of Brownian Particles Induced by a Periodic Asymmetric Potential," Nature, vol. 370, pp. 446-448, 1994.
[67] D. R. Chialvo and M. M. Millonas, "Asymmetric Unbiased Fluctuations Are Sufficient for the Operation of a Correlation Ratchet," Physics Letters A, vol. 209, pp. 26-30, 1995.

[68] G. W. Slater, H. L. Guo, and G. I. Nixon, "Bidirectional transport of polyelectrolytes using self-modulating entropic ratchets," Physical Review Letters, vol. 78, pp. 1170-1173, 1997.
[69] T. A. J. Duke and R. H. Austin, "Microfabricated sieve for the continuous sorting of macromolecules," Physical Review Letters, vol. 80, pp. 1552-1555, 1998.
[70] D. Ertas, "Lateral separation of macromolecules and polyelectrolytes in microlithographic arrays," Physical Review Letters, vol. 80, pp. 1548-1551, 1998.
[71] S. Leibler, "Brownian-Motion - Moving Forward Noisily," Nature, vol. 370, pp. 412-413, 1994.
[72] C. F. Chou, O. Bakajin, S. W. P. Turner, T. A. J. Duke, S. S. Chan, E. C. Cox, H. G. Craighead, and R. H. Austin, "Sorting by diffusion: An asymmetric obstacle course for continuous molecular separation," Proceedings of the National Academy of Sciences of the United States of America, vol. 96, pp. 13762-13765, 1999.
[73] D. E. Smith, H. P. Babcock, and S. Chu, "Single-polymer dynamics in steady shear flow," Science, vol. 283, pp. 1724-1727, 1999.
[74] A. van Oudenaarden and S. G. Boxer, "Brownian ratchets: Molecular separations in lipid bilayers supported on patterned arrays," Science, vol. 285, pp. 1046-1048, 1999.
[75] B. J. Kane, M. J. Zinner, M. L. Yarmush, and M. Toner, "Liver-specific functional studies in a microfluidic array of primary mammalian hepatocytes," Analytical Chemistry, vol. 78, pp. 4291-4298, 2006.
[76] J. H. Wittig, A. F. Ryan, and P. M. Asbeck, "A reusable microfluldic plate with alternate-choice architecture for assessing growth preference in tissue culture," Journal of Neuroscience Methods, vol. 144, pp. 79-89, 2005.
[77] J. El-Ali, P. K. Sorger, and K. F. Jensen, "Cells on chips," Nature, vol. 442, pp. 403-411, 2006.
[78] J. R. Burns and C. Ramshaw, "The intensification of rapid reactions in multiphase systems using slug flow in capillaries," Lab on a Chip, vol. 1, pp. 10-15, 2001.
[79] S. L. Anna, N. Bontoux, and H. A. Stone, "Formation of dispersions using "flow focusing" in microchannels," Applied Physics Letters, vol. 82, pp. 364-366, 2003.
[80] I. Shestopalov, J. D. Tice, and R. F. Ismagilov, "Multi-step synthesis of nanoparticles performed on millisecond time scale in a microfluidic droplet-based system," Lab on a Chip, vol. 4, pp. 316-321, 2004.
[81] E. M. Chan, A. P. Alivisatos, and R. A. Mathies, "High-temperature microfluidic synthesis of CdSe nanocrystals in nanoliter droplets," Journal of the American Chemical Society, vol. 127, pp. 13854-13861, 2005.
[82] A. J. deMello, "Control and detection of chemical reactions in microfluidic systems," Nature, vol. 442, pp. 394-402, 2006.

[83] G. J. M. Bruin, "Recent developments in electrokinetically driven analysis on microfabricated devices," Electrophoresis, vol. 21, pp. 3931-3951, 2000.
[84] V. Dolnik, S. R. Liu, and S. Jovanovich, "Capillary electrophoresis on microchip," Electrophoresis, vol. 21, pp. 41-54, 2000.
[85] C. T. Culbertson, Y. Tugnawat, A. R. Meyer, G. T. Roman, J. M. Ramsey, and S. R. Gonda, "Microchip separations in reduced-gravity and hypergravity environments," Analytical Chemistry, vol. 77, pp. 7933-7940, 2005.
[86] J. C. Roulet, R. Volkel, H. P. Herzig, E. Verpoorte, N. F. de Rooij, and R. Dandliker, "Microlens systems for fluorescence detection in chemical microsystems," Optical Engineering, vol. 40, pp. 814-821, 2001.
[87] J. C. Roulet, R. Volkel, H. P. Herzig, E. Verpoorte, N. F. de Rooij, and R. Dandliker, "Performance of an integrated microoptical system for fluorescence detection in microfluidic systems," Analytical Chemistry, vol. 74, pp. 3400-3407, 2002.
[88] J. Hubner, K. B. Mogensen, A. M. Jorgensen, P. Friis, P. Telleman, and J. P. Kutter, "Integrated optical measurement system for fluorescence spectroscopy in microfluidic channels," Review of Scientific Instruments, vol. 72, pp. 229-233, 2001.
[89] J. Seo and L. P. Lee, "Disposable integrated microfluidics with self-aligned planar microlenses," Sensors and Actuators B-Chemical, vol. 99, pp. 615-622, 2004.

AUTHOR BIOGRAPHY

Joshua D. Cross received his Ph.D. degree in Applied Physics from Cornell University in 2007. His thesis focused on micro- and nanofluidic structures for separating DNA molecules by length and sequence. He has worked on projects ranging from single molecule detection at entropic traps to studying the physics of long DNA molecules that are confined in nanometer-scale channels or polymer fibers. He is currently working on commercializing MEMS and NEMS technologies developed in the Craighead Research Group.

Harold G. Craighead is a professor of applied physics and director of the Nanobiotechnology Center at Cornell University. He received his Ph.D. in physics from Cornell University and spent the early part of his professional career in industrial research on device physics at Bell Labs and Bellcore. His current research activity involves the use of mechanical, fluidic, and optical nanostructures for molecular manipulation, chemical analysis and for studies of biological systems with high spatial resolution.

4 CMOS/MICROFLUIDIC HYBRID SYSTEMS

Hakho Lee[1,3]*, Donhee Ham[2] and Robert M. Westervelt[1,2]

[1] Department of Physics, Harvard University
[2] School of Engineering and Applied Sciences, Harvard University
[3] Center for Molecular Imaging Research, Massachusetts General Hospital, Harvard Medical School
*redhawk@post.harvard.edu

4.1 INTRODUCTION

Current trend in most device engineering can be characterized by two keywords: miniaturization and integration. Undoubtedly, the biggest spur to the trend has been the rapid advance of integrated circuit (IC) technology. Since its first demonstration in 1958 with just one transistor [1], IC technology has undergone a revolutionary development [2] – in only about half a century, we can put together several millions of transistors on stamp-size area [Fig. 4.1(a)]. Small, fast and cheap electronic devices enabled by IC technology are now indispensable in every aspect of modern life.

In 1980s, the trend of miniaturization and integration was ushered in fluidic system engineering. By applying IC fabrication technologies to make channels and reservoirs not for electrons but for fluids, researchers began to shrink down the size of fluidic devices for chemical and biomedical analysis (See Chapter 3, for example). The development of small and integrated fluidic systems has been accelerated in 1990s [Fig. 4.1(b)] and eventually established itself as a major research field, microfluidics [3, 4]. Although still in its infancy, microfluidics holds a promise to become a revolutionary technology in chemical, biomedical analysis with its potential for rapid and reliable measurement while consuming a tiny amount of samples [5].

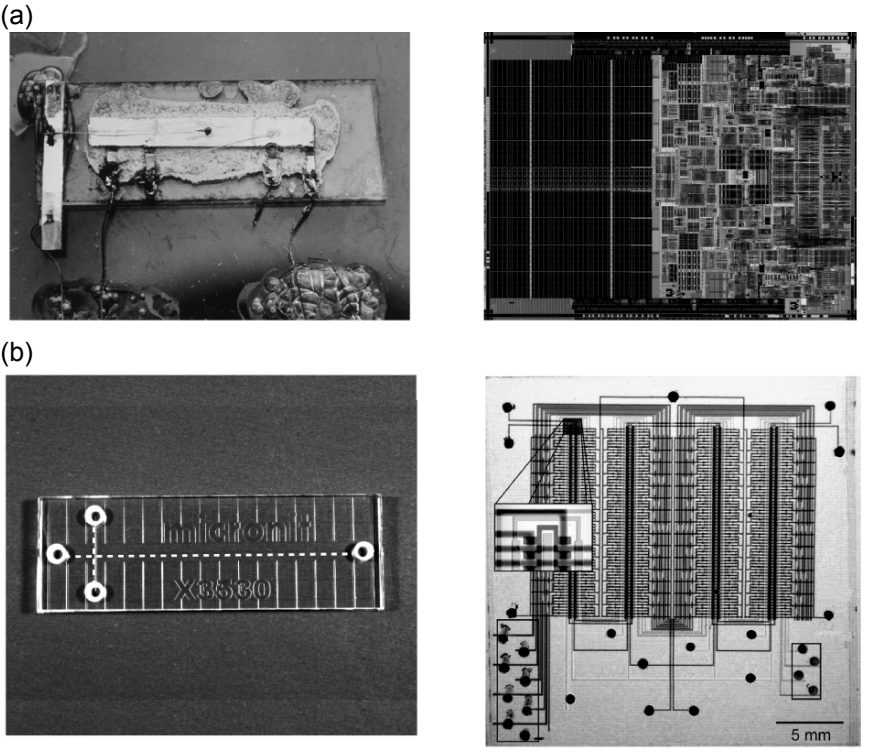


Figure 4.1 (a) Development of integrated circuit (IC) chips. (Left) Photo of the first integrated circuit in 1958. One transistor and passive components are integrated on a germanium substrate. The area of the substrate is 18 mm^2. Courtesy of Texas Instruments. (Right) Die micrograph of the Intel® Core™2 Duo processor in 2006. More than 290 million transistors are integrated in 143 mm^2 area. Courtesy of Intel Corporation. (b) Representative examples of microfluidic systems. (Left) Simple microfluidic system for capillary electrophoresis fabricated on a glass substrate (45 mm × 15 mm). Microfluidic channels are indicated by the dotted line. Such type of device was the dominant player in the early development of microfluidics [6]. Courtesy of Micronit Microfluidics. (Right) Integrated, multiplexed microfluidic system fabricated by soft lithography. Over 2000 valves are implemented in the device and are individually addressed by multiplexers. Reprinted with permission from [7].

The main objective of this chapter is to introduce a new device concept that combines the advantages of the aforementioned two distinctive systems, IC and microfluidics, through hybrid integration. Specifically, we propose a *CMOS/microfluidic hybrid system* where CMOS IC chips perform active functions inside microfluidic frameworks [8, 9]. Although there have been efforts to make hybrid systems from both electrical engineering and microfluidics communities, the implemented devices often assumed a "partial" integration level: the device may have a full-fledged CMOS IC but with a simple fluidic chamber (usually from the electrical engineering community) or it may have a sophisticated microfluidic network with just simple electrodes fabricated in-house (usually from the microfluidic community). What we aim in this chapter is to promote the integration of CMOS ICs and microfluidic systems on an equal footing, to fully harness the advantages of both systems.

In the next section, we describe the concept and the advantages of the hybrid approach. Then we devote the rest of the chapter to practical issues in the development of a hybrid system. In Section 4.3, we describe the representative fabrication technologies to implement hybrid systems; especially, we focus on how microfluidic systems can be fabricated and integrated with CMOS chips. Section 4.4 briefly discusses the device packaging which is a crucial but often overlooked process in the device development.

4.2 CMOS/MICROFLUIDIC HYBRID SYSTEM – CONCEPT AND ADVANTAGES

In a CMOS/microfluidic hybrid system, the CMOS chip is not just a back-end device for signal processing. Rather, the CMOS chip takes an active and essential role inside a microfluidic network. Figure 4.2 illustrates the concept of the hybrid system – on top of a substrate where CMOS chips are fabricated or placed, microfluidic networks are fabricated. The CMOS chip, embedded in the fluidic networks, exposes its *surface* to the fluidic flow. Functional elements such as sensors or actuators are implemented on the CMOS chip surface and they directly interact with fluids or objects suspended in the fluids. The microfluidic networks are a crucial component in the hybrid systems. Microfluidic networks are used 1) to introduce fluids or samples on to the CMOS chips, 2) to keep biocompatible environments, and 3) to provide sophisticated flow control for sample mixing, separation and reaction. The CMOS chips can be used for various purposes as described in the following subsection.

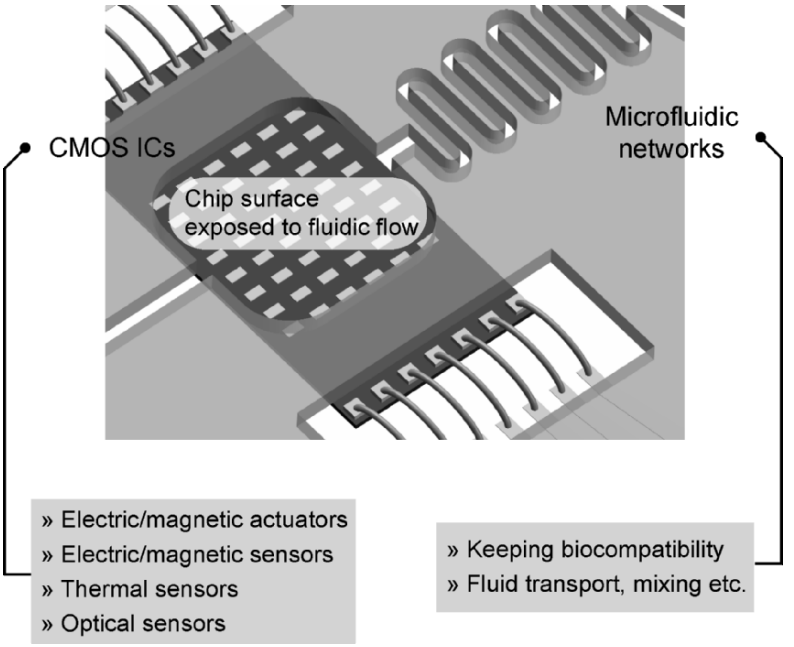


Figure 4.2 Concept of the CMOS/Microfluidic hybrid system. The CMOS chip is integrated into the microfluidic networks. The chip surface where actuators and sensors are located is exposed to the fluidic flow. Microfluidic networks keep biocompatible environments and provide a fine control on fluids.

4.2.1 Application of CMOS ICs in a Hybrid System

The function of the CMOS chip in a hybrid can fall into three major categories: actuators, sensors, and stimulators. In the actuation applications, CMOS chips generate mechanical forces on the fluids or objects via electromagnetic interactions. As a sensor, CMOS chips are used to measure electric, optical, or thermal properties of the target samples. The stimulation function is closely coupled with the sensing counterparts. The on-chip stimulators can affect the target samples by generating electric/magnetic field or heat, and the responses of the samples can be measured by the on-chip sensors. Note that with modern IC technology, monolithic CMOS chips that perform multiple or all of these functions can be easily implemented.

4.2.1.1 Electric and Magnetic Actuators

As an actuator, CMOS chips exert forces on small objects such as DNAs, biological cells and microspheres that are suspended in fluids. Both components of electromagnetic fields, electric and magnetic, can be used to create

forces and control the motion of objects in fluids. For the electric actuation, microelectrodes are implemented on a CMOS chip surface to generate electric fields that interact directly with target samples and the surrounding media. For example, in DNA microarray DC electric fields interact with the static charges of DNAs (See Chapter 7); cell dielectrophoresis exploits the difference of polarizabilities between the objects and media under AC electric fields to trap and move the objects (See Chapter 6). For magnetic actuation, micro-electromagnets are implemented on the chip surface to generate local magnetic fields. Magnetic beads or biological cells attached to the magnetic beads then can be manipulated on the chip surface (See Chapter 5). Because the field sources, microelectrodes and micro-electromagnets, can be small (~a few micrometers) and be in an array form, microscopic and parallel manipulation of target samples can be performed with high manipulation accuracy and throughput.

4.2.1.2 Electrical Sensors and Stimulators

Combined with integrated signal conditioning circuits (filters and amplifiers), microelectrodes on a CMOS chip surface can from high-sensitive electric sensors. Biological calls, for example, can be capacitively-coupled to microelectrodes and their bioelectric activities can be measured non-invasively. By implementing a large array of microelectrodes with multiplexing circuits, the propagation of the electrical signals in biological tissues can be monitored, providing valuable information on intercellular communication. Note that by reversing its function, a microelectrode can provide electrical stimulations to cells to initiate cellular responses. CMOS-based microelectrodes have been used to monitor the electrical activity of cultured neurons, heart cells and muscle cells (See Chapter 8), to record the brain activity as an implanted sensor (See Chapter 9), and to detect DNAs through hybridizations [10].

4.2.1.3 Thermal Sensors and Stimulators

For accurate temperature measurement, precision temperature sensors can be easily integrated in CMOS IC chips along with other actuating and/or sensing components. In a hybrid system, monitoring the temperature of the IC, especially the chip surface, is important for biocompatibility proposes as the IC generates heat during its operation. To enhance biocompatibility of the system, the output of the temperature sensors can be coupled with external coolers or heaters for temperature regulation (see Chapter 5 for an example).

CMOS ICs can also serve as a thermal actuator. Implementing conductors on an IC surface and flowing electrical currents, the temperature of the chip surface can be locally increased. This capability can be used, for example, to move fluids by creating temperature gradient [11, 12]. Local heating can also switch a heat sensitive polymer, poly-N-isopropylacrylamide (PNIPAM), to adsorb or lease proteins, which can be applied to spatially and temporally pattern proteins on the chip surface [13].

4.2.1.4 Optical Sensors

Optical imaging is the most sensitive and wide-spread detection method in biological and medical applications. At the same time, it is the weakest area in the aspect of device miniaturization; optical detection often involves bulky equipment including light sources and microscopes. Combined with microfluidic systems, CCDs (charge-coupled device) and CMOS imagers can image biological samples without relying on intermediary, bulky optical components (see Chapters 10 and 11). In these applications, samples are positioned right on top of the imaging chip surface, which is achieved by microfluidic systems. The microfluidic systems can further be used to tailor optical paths and manipulate lights [14]. Using solid-state light sources such as light emitting diodes (LEDs), the whole system (light source, optical detector, and microfluidic channel) can be packaged together to a make a portable, point-of-care imaging device.

4.2.2 Advantages of the CMOS/Microfluidic Hybrid Approach

The hybrid systems combine the advantages of both microelectronics and microfluidics. The merits of the microfluidic systems can be found in Chapter 2 and in many reviews (for example, see [4, 5, 15, 16]). The use of CMOS ICs brings the following advantages and promises.

- *Increased spatial and temporal resolution*: First and foremost, hybrid systems provide a way to perform experiments with high spatiotemporal resolution. Apparently the sensors and the actuators in a CMOS chip can have small footprints to increase the spatial resolution of experiments. The accuracy of experiments can be enhanced by electrically connecting the sensors and the actuators using the metal layers away from the chip surface, exposing only the active elements (sensors and actuators) to target biological samples. The CMOS chips can also be designed to operate in GHz ranges [17] for high speed measurement. By incorporating standard on-chip ADCs and memories, the data

sampling rate can be increased, enhancing the temporal resolution of the measurement.

- *Parallel and automated measurement*: With actuators or sensors laid out in an array format, hybrid systems allow users to perform measurements in parallel fashion for high throughput operation. Multiplexer circuitry can be integrated together to handle large arrays of actuators and sensors with a small number of external inputs. The measurement sequences can be easily automated through the programmed control of CMOS ICs, which minimizes the need for labor-intensive, potentially error-prone manual procedures.
- *Robustness and user-friendly interface*: One of the significant merits of the hybrid system is its durability and reproducibility. The CMOS ICs in a hybrid system are manufactured in a commercial foundry with sophisticated process control. The chip-to-chip variation is maintained minimal, which ensures the consistency in the chip performance. In addition, empowered by integrated control circuits, a hybrid ICs can present simple interfaces for external inputs; users can enjoy the same level of the user-friendly interfaces which are now standard in consumer electronics.
- *Cheap, portable and disposable*: The hybrid systems can be a cost-effective, portable unit. The ICs in a hybrid system can be easily mass-produced using the fabrication lines in conventional semiconductor foundries; as described in the following section, microfluidics systems can also be fabricated at low cost using polymers. The resulting device can be self-contained with most of control electronics already embedded, yet it can be cheap from high volume production. The hybrid system, therefore, can be a cheap and disposable device, which is especially attractive to clinical and medical applications.
- *New opportunity for IC industry*: The hybrid system presents a new, novel use of conventional ICs. Semiconductor industries which are looking for future markets for ICs can easily join in this field. In turn, the participation of the industry is highly desirable to further develop the hybrid system as a major technology and to promote its commercialization.

4.3 FABRICATION OF MICROFLUIDIC NETWORKS FOR HYBRID SYSTEMS

To complete a hybrid system, microfluidic systems has to be implemented with CMOS IC chips exposed to the fluidic path. Even though various fabrication methods have been developed to fabricate microfluidic systems [18], integrating CMOS chips with microfluidic systems is still a non-trivial task with the following challenges.

1. Because microfluidic systems are usually post-fabricated over CMOS chips, fabrication methods should be safe and compatible to CMOS ICs, which imposes restrictions on fabrication conditions. For example, fabrication methods requiring high temperature (>400ºC) or high electric fields, such as silicon fusion bonding or anodic bonding, cannot be used as they can cause permanent damages to the CMOS chips.
2. The footprint of microfluidic systems is usually large (~ cm^2) compared to the size of CMOS chips (~ mm^2). To facilitate the fabrication of microfluidic systems, it is desirable to increase the lateral dimension of the CMOS chips to match the size of microfluidic systems. Embedding empty spaces on a CMOS chip itself is a simple but clearly a non-practical solution. A cost-effective alternative is to attach a CMOS chip to a separate, bigger substrate. In this method, the substrate serves as a platform for the subsequent fabrication processes. Planarization steps are required, however, as the CMOS chip extrudes above the substrates.
3. Besides fluidic connections, CMOS/microfluidic hybrid systems need a number of electrical connections for operation, which can complicate the microfluidic system design. For example, when the size of the CMOS chips is smaller than that of the microfluidic systems, open areas or through holes have to be fabricated inside the microfluidic systems to allow electrical connections.

Polymers are a good candidate material for the microfluidic structures in CMOS/microfluidic hybrid systems, possibly offering a solution to the aforementioned challenges. First, polymers can be processed at relatively low temperate (<300ºC) that are safe to CMOS chips. Yet cured polymers show high chemical resistance and physical integrity. Second, due to their fluidic nature, polymers can be easily transformed into complex shapes. Especially, polymers provide an easy way to planarize bumpy surfaces, as they can reflow over the surfaces. Third, polymer-based microfluidic

systems can be of low cost and easily mass-produced; polymers are cheap and their fabrication techniques are well-established and industrialized [19].

This section reviews representative fabrication methods for polymer-based microfluidic structures. Four methods, direct patterning, casting, lamination, and hot embossing, are reviewed with the examples of how each method can be applied to complete CMOS/microfluidic hybrid systems.

4.3.1 Direct Patterning of Thick Resins

By spin-coating and patterning photo-definable resins, microfluidic systems can be fabricated directly on top of CMOS chips. This method has the following advantages.

1. Because the fabrication processes are compatible to those of CMOS chips, this method can be easily adapted in semiconductor foundries for the wafer-level fabrication of microfluidic structures.
2. Strong and watertight adhesion between the CMOS chips and the fluidic systems can be achieved, as resins are directly coated and cured over CMOS chips.
3. By coating and patterning multiple layers of resins, sophisticated fluidic structures with a low profile can be created.

The main drawback to the direct patterning is that this method produces open fluidic channels; a separate sealing step is necessary to make covered microfluidic systems. Novel fabrication techniques to implement closed microfluidic channels have been demonstrated, but the processes can be rather involved [20, 21].

Various resins, including epoxy, polyimide, and bisbenzocyclobutene, are commercially available and have been applied to implement microfluidic systems [22-24]. Especially, photo-definable epoxy (SU-8; MicroChem, MA) has been extensively used as a structural material for microfluidic systems. The SU-8 resin can be spin-coated into a film with a wide range of thicknesses, covering from <1 μm to >1000 μm. Using standard photolithograph, the resin can be patterned into high aspect ratio (up to 20) structures. Cured SU-8 shows high physical strengths, chemical resistances, and optical transparencies.

As an example of the direct patterning method, Fig. 4.3 illustrates how microfluidic channels can be fabricated using SU-8 resins [25]. [Step 1] First, a CMOS chip is attached to a bigger Si/SiO_2 substrate. The substrate can contain electrical leads to facilitate electrical connection to the CMOS chip. [Step 2] Subsequently, a thick layer of SU-8 is spin-coated and

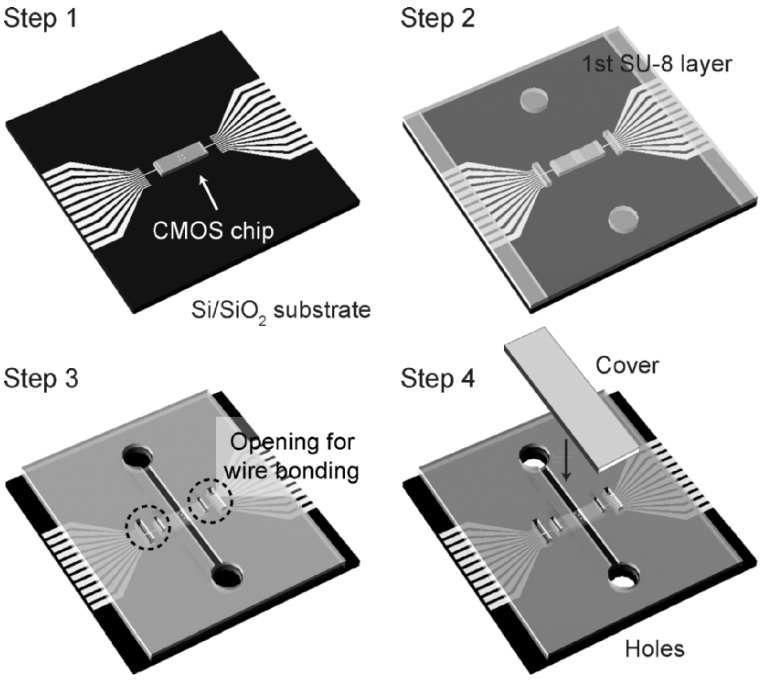


Figure 4.3 Example of direct patterning of thick resists. [Step 1] A CMOS chip is attached to a bigger substrate with electrical leads. The substrate serves as a platform for the subsequent processes. [Step 2] A thick SU-8 layer is spin-coated to planarize the surface. Fluidic ports and openings for electrical connections are patterned, but the resist is not developed. [Step 3] A second SU-8 layer is spin-coated and patterned for microfluidic channels. Both SU-8 layers are then developed. [Step 4] Holes are drilled on the substrate and the open microfluidic channel is sealed by gluing a cover. Reprinted with permission from [25].

exposed to UV light to pattern fluidic ports and openings for electrical connections, but the resist is not developed. The thickness of the SU-8 layer is controlled to be comparable to that of the CMOS chip to provide a leveled surface. [Step 3] A second layer of SU-8 is spin-coated and patterned for microfluidic channels, and then both SU-8 layers are developed. Because of the thick layer coated in Step 2, the resulting microfluidic channel can have relatively uniform depth. The whole structure is then hard-cured on a hotplate to further strengthen the SU-8. After the hard-cure, fluidic inlets are formed by drilling holes on the Si/SiO_2 substrate. [Step 4] The microfluidic channel is then sealed by gluing a thin glass slip.

4.3.2 Casting of Polymers

Casting is one of the easiest methods to shape microfluidic structures from polymers. In this method, a pre-polymer mixed with a curing agent is poured into a mold that has the negative profile of the desired fluidic structure. The polymer is then cured into a solid form and released from the mold. The microfluidic structure can be bonded or glued to a CMOS chip to implement a hybrid system. The casting method has the following advantages.

1. A mold can be structured on various materials with many different methods, offering flexibility in fabrication. For example, a mold can be cut on metal or PMMA (polymethylmethacrylate) blocks using CNC (computer-numerical-control) machines, or it can be fabricated on silicon or glass substrates using the conventional optical lithography technique. The mold, once fabricated, can be reused many times to replicate the microfluidic systems, reducing the labor and the time in the subsequent fabrication steps.
2. In contrast to the direct patterning of resins (Subsection 4.3.1), where microfluidic systems are structured on top of CMOS chips, the casting method allows the separate fabrication of the microfluidic systems. CMOS chips are needed in the final assembly step where the chips are attached to microfluidic systems. Therefore, one can develop and test many different fluidic structures without sacrificing CMOS chips, which is especially beneficial for device prototyping.

A choice of polymers, for example, polyurethane, silicone rubbers, polyimide and epoxies are available in a liquid form at room temperature, and can be used for casting. Among these materials, a silicone-base elastomer, polydimethylsiloxane or PDMS, is now widely used for microfluidic systems, even spawning a new field of research, soft lithography. The PDMS elastomer has many desirable properties as a structural material for microfluidic systems: 1) PDMS is chemically and thermally stable and does not swell with humidity, 2) cured PDMS layers can be vertically stacked via irreversibly bonding, making it possible to create sophisticated, 3-dimensional fluidic networks, and 3) PDMS is optically transparent, biocompatible and even gas-permeable; PDMS microfluidic systems are well-fitted to culture and monitor biological cells.

Figure 4.4 shows the general procedure to fabricate microfluidic systems by casting PDMS. [Step 1] First, a mold for a microfluidic system is patterned. Conventional optical lithography as illustrated in the figure or other methods can be employed to construct the mold. [Step 2] PDMS mixed with a curing agent is poured onto the mold. Before pouring PDMS, the mixture

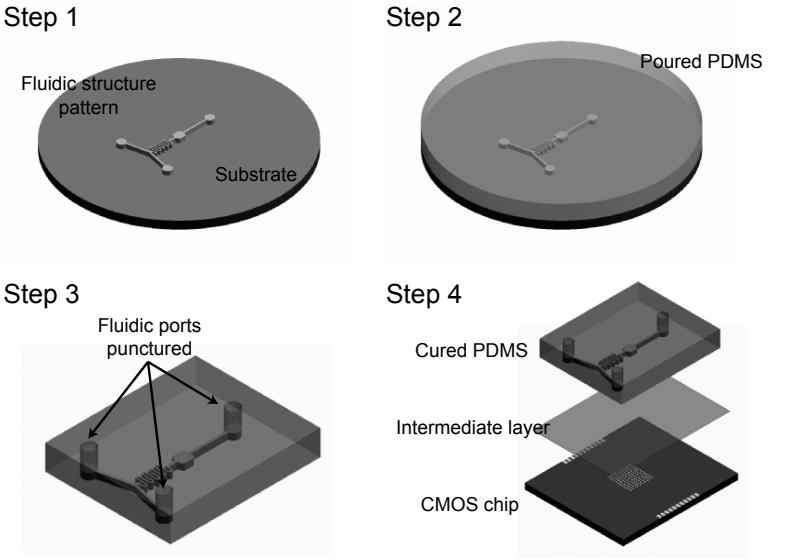


Figure 4.4 Casting method for implementing microfluidic systems. [Step 1] A mold is formed on a substrate. The mold can be fabricated on various substrates using different methods. [Step 2] PDMS (polydimethylsiloxane) with curing agent is poured over the mold and thermally cured. [Step 3] Cured PDMS is peeled off and cut into a desired shape. Fluidic ports can be punched out. [Step 4] The PDMS microfluidic structure is bonded to a CMOS chip. To promote the adhesion, a thin layer of PDMS pre-polymer can be applied to the cured PDMS structure or a SiO_2 layer can be deposited on the CMOS chip surface (See the text for detail).

can be degassed under vacuum to remove air bubbles from mixing. The poured PDMS is then thermally cured on a hotplate or in an oven. At 70ºC, typical cure time is ~1 hr. [Step 3] The PDMS structure is then peeled off from the mold and cut into a desired shape.

The final sealing of the PDMS structure to the CMOS chip deserves special attention. Due to their low interfacial free energy, cured PDMS layers rarely bond irreversibly to most of materials [26]. This property is helpful when detaching a cured PDMS from a mold, but becomes problematic when one tries to permanently attach a PDMS microfluidic system to a CMOS chip. One solution to promote the adhesion between the PDMS layer and the CMOS chip is to placing an intermediate layer on the chip surface. [Step 4] For example, a thin layer of PDMS pre-polymer can be applied to the surface of the cured PDMS microfluidic structure. The PDMS structure is then placed on top of the CMOS chip, followed by an additional thermal cure. Alternatively, a thin layer of SiO_2 can be deposited and patterned on top of the CMOS chip. Both surfaces of the CMOS chip and the PDMS structure

are then treated with O_2 plasma and made contact to initiate irreversible bonding [27].

4.3.3 Lamination of Dry Film Resists

Although relatively new to microfluidic system fabrication, dry film resists are standard material in the printed circuit board (PCB) industry, supporting easy and fast patterning of circuit layouts. Dry film resists, which consist of photo-definable resins sandwiched between protective polymer layers, can be easily coated on to various substrates via thermal lamination, and can be patterned by standard optical lithography. This technology can be easily adopted to fabricate microfluidic systems directly on top of CMOS chips. Main advantages of the dry film resist technology are following.

1. Conformal layers with uniform thickness can be achieved across the whole laminated area. Lamination does not suffer from the bead-edge formation as in the spin-coating of liquid resins. Furthermore, due to their excellent adhesion properties, dry film resists can be laminated on to almost any material.
2. Dry films resists can be used for various purposes in microfluidic systems. Patterned dry films can be vertically stacked via repeated lamination to form multilevel microfluidic systems. Thin dry film resists can function as an intermediate adhesion layer to bond two different substrates. Most of all, dry films can be laminated to open microfluidic structures without using additional adhesives, drastically simplifying sealing steps.

Main drawbacks of the lamination method are 1) the available thickness or the depth of the fluidic structure is limited to the multiples of the dry film thickness used, and 2) high aspect ratio structures are hard to realize; typical aspect ratio was found to be $1 \sim 3$ with dry films [28]. However, note that these shortcomings can be easily overcome by combining the lamination method with the direct patterning (Subsection 4.3.1). For example, by directly patterning SU-8, open fluidic structures with various depths and high aspect ratios (up to 20) can be patterned. Dry film resists then can be laminated to make closed fluidic systems.

Figure 4.5(a) illustrates the fabrication of microfluidic channels using dry film resists. [Step 1] An open microfluidic structure is formed either by spin-coating and patterning a photoresist or by laminating and patterning a dry film resist. If the CMOS chip has a small footprint, the chip is attached to a bigger substrate and leveled by spin-coating a thick layer of resists (See

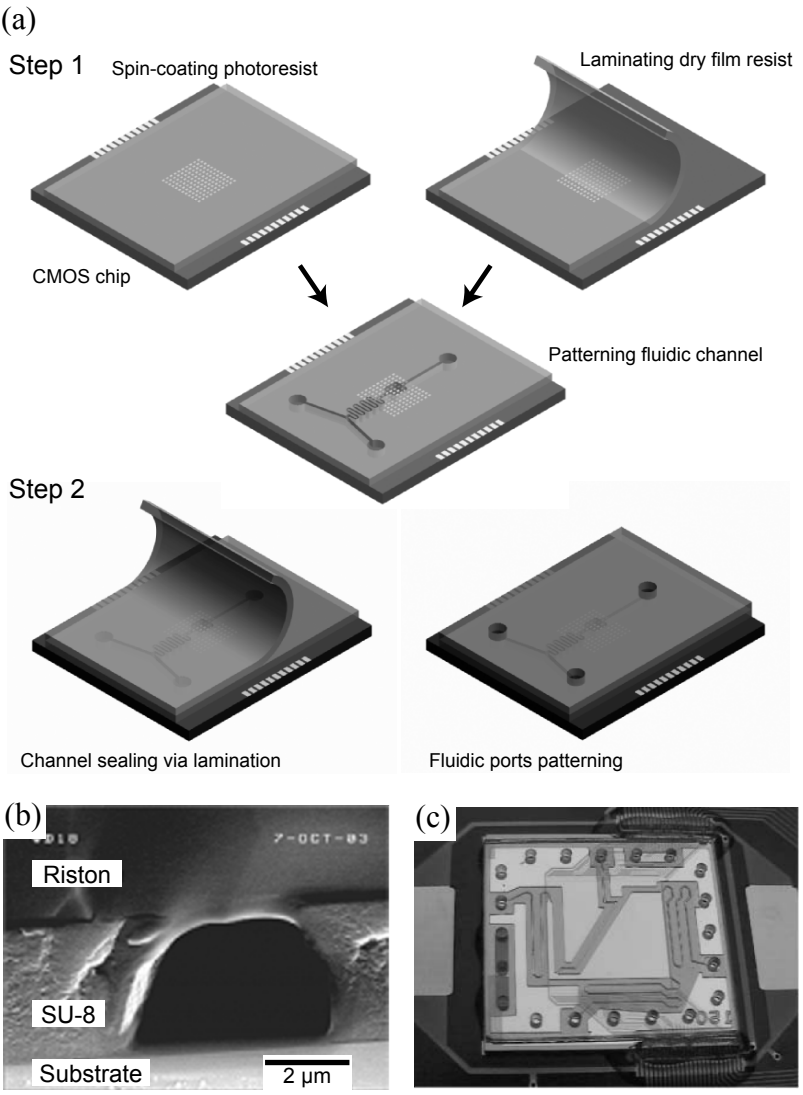


Figure 4.5 Lamination of dry film resists to fabricate microfluidic systems. (a) Fabrication procedure. [Step 1] An open microfluidic channel is patterned on a spin-coated photoresist or on a laminated dry film resist. [Step 2] A dry film resist is laminated to seal the channel and the fluidic ports are patterned lithographically. (b) and (c) Examples of microfluidic channels using dry film resists. In (b), an open channel is formed by patterning SU-8, and sealed by laminating a dry film resist (Riston). Reprinted with permission from [29]. In (c), an open channel is structured by laminating and patterning double layers of a dry film resist (Ordyl SY550). The open channel is then sealed with a glass cover using the underlying dry film as an adhesive. Reprinted with permission from [28].

Step 2 in Fig. 4.3). [Step 2] The open fluidic structure is sealed by laminating and patterning a dry film resist. To make multilayer structures, Step 2 can be repeated using additional dry film resists.

Figures 4.5(b) and (c) shows the micrographs of microfluidic systems fabricated using dry films. In Fig. 4.5(b), an open microfluidic channel is formed first by directly patterning SU-8, and then sealed with a Riston (DuPont, Research Triangle Park, NC) layer by lamination [29]. In Fig. 4.5(c), an open microfluidic structure is formed on a silicon device by laminating and patterning double layers of a dry film resist (Ordyl SY550, Elga Europe, Italy) [28]. A glass lid is then sealed on top, using the underlying dry film resist as an adhesive.

4.3.4 Hot Embossing

Hot embossing is one of the widely used methods to pattern features in polymers [19]. In this process, a rigid polymer sheet is heated up and pressed against a bossing master that has the negative pattern for the desired structure. Like the casting method described in Subsection 4.3.2, hot embossing is a replication technology; once a bossing master is prepared, microfluidic structures can be stamped out repeatedly. Hot embossing has the same merits as in the casting method: flexibility in the preparation of bossing masters and possibility to develop microfluidic systems separately from CMOS ICs. In addition, the hot embossing method offers the following additional advantages.

1. The fabrication cycle is considerably short (<10 min). In the casting process, typical fabrication time is relatively long (~1 hr) mainly due to the slow cure processes of polymers.
2. High aspect ratio (>10) structures can be achieved compared to the casting method where typical aspect ratio of structure is ~1 [30]. During the casting process, polymers flow and solidify over the whole molds. Cured polymers with high aspect ratio features are often hard to separate from the molds due to frictions. In the hot embossing process, however, only a small amount of polymers are melt and transported, which reduces the friction between a bossing master and a stamped polymer sheet [31].

The hot embossing processes start with the preparation of a boss master. To withstand the high pressing force (~10 kN/cm^2) during the embossing procedure, the masters are usually formed with metal. Large features can be directly cut on metal substrates by CNC (computer-numerical-control)

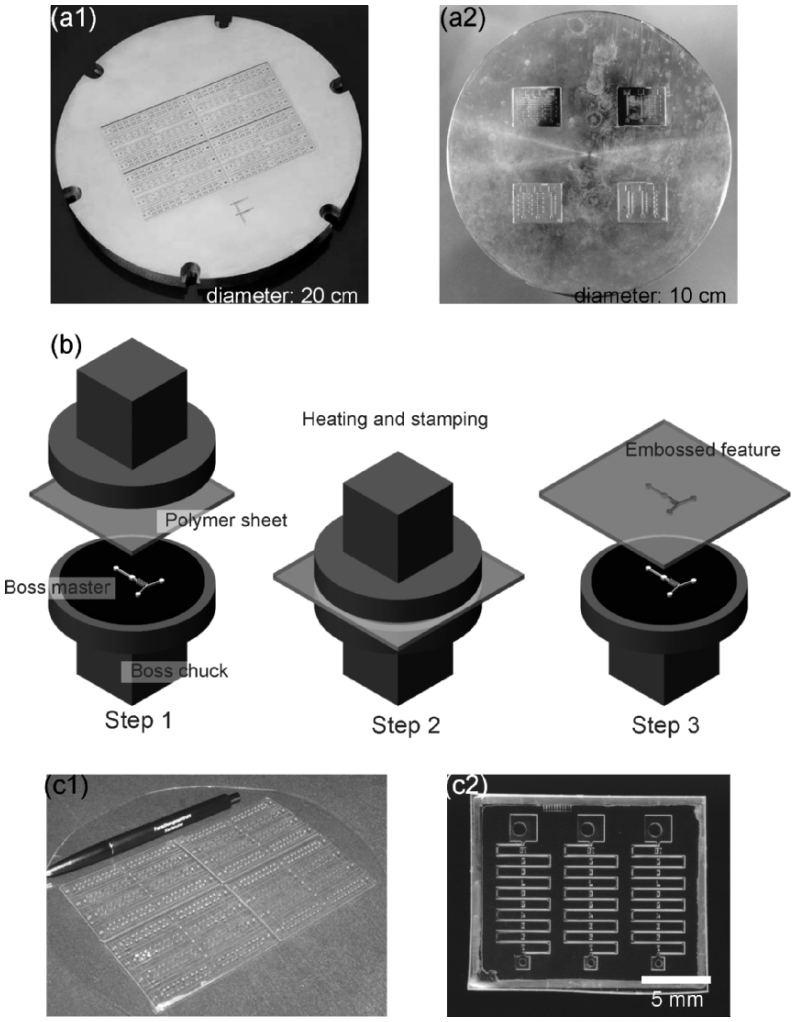


Figure 4.6 Hot embossing method for implementing microfluidic systems. (a) Bossing masters can be fabricated on various substrates. (a1) A bossing master is micromachined on a brass block (diameter 20 cm). (a2) A boss master (diameter 10 cm) is prepared by electroplating nickel on four negative patterns of fluidic structures. (b) Hot embossing processes. [Step 1] The boss master and a polymer substrate are mounted in the bossing machine. [Step 2] While the boss master and the substrate are heated above the glass transition temperature of the polymer, the master is pressed on to the polymer. [Step 3] The whole system is cooled down and the boss master is withdrawn from the polymer sheet. (c) Examples of final microfluidic structures. (c1) show the patterns transferred on PMMA substrates using the boss master in (a1); (c2) shows one of the fluidic chips patterned on PMMA using the boss master in (a2). Reproduced with permissions from [32] and [33].

machining; micrometer scale features can be patterned via optical lithography and metal-electroplating. Figures 4.6(a1) and (a2) show the examples of boss masters. In Fig. 4.6(a1), the master is structured by micromachining a brass substrate [32]. To make the master in Fig. 4.6(a2), initially the negatives of fluidic patterns are etched in silicon wafer and the wafer is diced into chips (17 mm × 14 mm); four chips are then glued to a substrate and the whole assembly is nickel electroplated [33].

Actual embossing processes are performed with a dedicated machine as schematically shown in Fig. 4.6(b). [Step 1] The boss master and a polymer sheet are mounted in the machine, and the chamber containing the setup is evacuated. Vacuum is required to prevent the formation of air bubbles during the subsequent stamping process. Polymethylmethacrylate (PMMA), polycarbonate (PC), and cycloolefin copolymers (COCs) are frequently used material as a substrate [34]. [Step 2] Both the polymer sheet and the boss master are heated above the glass transition temperature (T_g) of the polymer, and mechanical pressure is applied to stamp the boss master to the substrate. [Step 3] With the master still pressed, the polymer and the master are cooled below T_g, and then the master is mechanically withdrawn from the polymer sheet. The final structure can be cut and glued on top of a CMOS chip to form a hybrid system. Figure 4.6(c) shows the photos of embossed structures on PMMA substrates using the masters in Fig. 4.6(a). In Fig. 4.6(c1), the whole PMMA substrate is patterned with the mold in Fig. 4.6(a1). Figure 4.6(c2) shows one of the fluidic chips patterned using the mold in Fig. 4.6(a2) and capped with flat a PMMA sheet.

4.4 PACKAGING OF CMOS/MICROFLUIDIC HYBRID SYSTEMS

Packaging is an important and often the most time-consuming process in the fabrication of hybrid systems. Through packaging, hybrid systems acquire interfaces to external environments for operation; packaging provides electrical, fluidic, and thermal connections to and from hybrid systems. Hybrid systems operate under a unique condition where solid state devices (CMOS chips) coexist with fluids and biological species, which calls for a special attention in package design. For example, electrical connections and fluidic connections should be well isolated, and temperature of the packaged system should be regulated for biocompatibility. To save time, labor and cost, it is highly recommended to plan the packaging at the same time a hybrid system is designed.

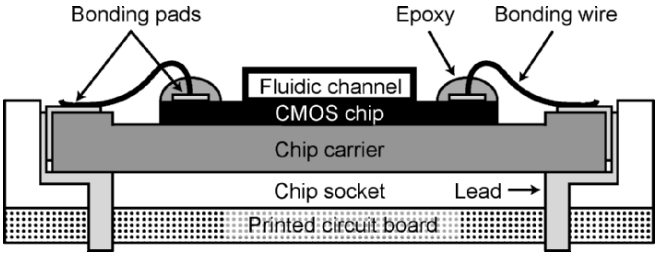


Figure 4.7 Electrical packaging of a hybrid system using a chip carrier and a chip socket. The CMOS chip is electrically connected to the chip carrier via wire bonding. The bonding sites can be protected by applying epoxy resins.

4.4.1 Electrical Connection

In IC packaging, CMOS chips or dies are electrically connected to packages via two distinctive and well-established techniques, wire bonding and flip-chip bonding [35]. In hybrid system packaging, wire bonding is a dominant choice for electrical connections; the CMOS chip surface where the bonding pads are located has to be exposed to fluidic systems.

Figure 4.7 illustrates a typical electronic interconnect scheme for a hybrid system. To facilitate wire bonding, commercial chip carriers with gold-plated bonding sites are used as a housing. After making wire connections, the bonding sites on CMOS chips can be covered with epoxy for protection. Chip carriers then can be plugged into matching chip sockets to make electrical connections with external electronics.

4.4.2 Fluidic Connection

Compared to the electrical connection techniques, the method for fluidic connection to microfluidic chips is still in its infancy; in many cases, fluidic connectors are manually installed by inserting and sealing capillary tubes to fluidic ports. For the high volume production of hybrid systems, standardized and automatic fluidic connection schemes should be developed as happened in electrical connection [36, 37]. One promising direction is to incorporate fluidic ports as well as electrical connection sites into chip carriers and chip sockets.

Figure 4.8 shows two examples of fluidic connection method used for hybrid systems. In Fig. 4.8(a), a microfluidic channel is formed on top of an IC chip by directly patterning a thick resist (polyimide) and by sealing the channel with a glass lid. Custom-made fluidic ports are glued to the

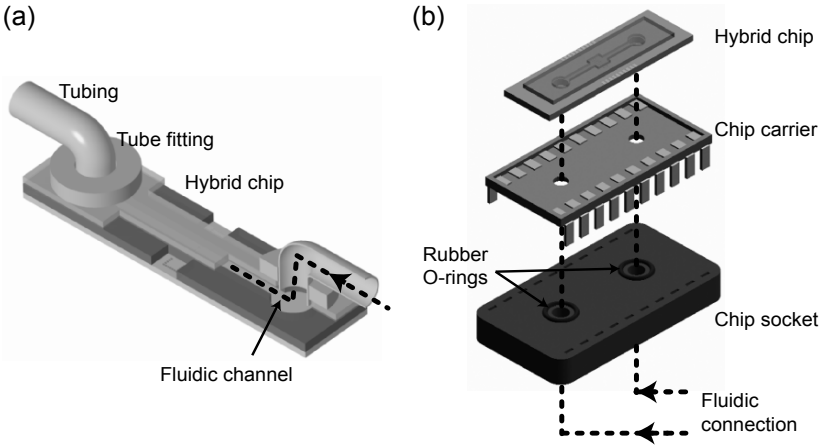


Figure 4.8 Examples of the fluidic packaging for hybrid systems. (a) Fluidic connections are made by inserting tubes to the tube fittings. Each device needs the tube fittings installed, which can be time-consuming and non-economical. Adapted from [8] (b) Modular connection scheme using a chip carrier and a chip socket. Note that fluidic connections as well as the electrical connections are automatically made by inserting the chip carrier to the chip socket. Adapted from [38].

device and fluidic connections are made by inserting elbow tube fittings to the ports. Although simple and straightforward, this connection method is time-consuming and non-economical, because fluidic connection components have to be implemented in each device.

Figure 4.8(b) illustrates an improved, modular fluidic connection method that uses a chip carrier and a chip sockets [38]. A hybrid system is prepared by patterning a microfluidic channel with SU-8 on top of a microelectronic chip. Fluidic ports are drilled on the chip itself and the microfluidic channel is sealed by gluing a glass lid. The assembled device is then mounted on a ceramic chip carrier that has pre-drilled matching holes for the fluidic ports. Fluidic connections are complete by inserting the chip carrier to a chip socket which has external fluidic ports installed. Fluidic seal between the chip carrier and the socket is achieved by the rubber O-rings. This modular connection scheme has several advantages: 1) Once installed, the chip socket and the fluidic connections can be reused, 2) the packaged hybrid system has a low profile, because the fluidic connectors are implemented under the chip carrier. This makes it easy to mount the system under microscopes for observation, and 3) electrical connections are automatically achieved via chip sockets and chip carriers.

4.4.3 Temperature Regulation

Temperature regulation or thermal management is an important issue in electronic packaging as integrated circuits become smaller with increasing operation speed [39]. To prevent thermal failure, ICs have to be cooled down, and packaging should provide efficient thermal paths between the ICs and external environments. In hybrid systems, temperature regulation can be more challenging; 1) to maximize biocompatibility, the temperature of hybrid systems has to be maintained constant with small fluctuations, and 2) with the presence of fluid on the CMOS chip surfaces, relatively large temperature gradient can exist between CMOS chips and the rest of hybrid systems.

At the packaging level, temperature regulation are facilitated 1) by gluing hybrid systems on ceramic or metallic chip carriers with thermally conductive adhesives and 2) by attaching external, active cooling components such as fans or thermal electric coolers to the packaging. For more accurate temperature regulation, temperature sensors should be incorporated in a CMOS chip to read out its local temperature. The output from the sensors then can be used to control the external cooling systems by forming a feedback loop (See Chapter 5 for example).

4.5 CONCLUSIONS AND OUTLOOK

In this chapter, we have introduced the CMOS/microfluidic hybrid system as a new approach to implementing highly functional, sophisticated devices for biological and biomedical applications. The hybrid system exploits the advantages of both integrated circuits and microfluidic networks; the microfluidics provides a fine control on fluid flow and the ICs function as versatile actuators and/or sensitive sensors *inside* the microfluidic network.

The hybrid system presents an unprecedented opportunity in biotechnology. It enables users to perform tests, assays and diagnoses with micrometer-scale spatial and microsecond temporal resolution. With its high throughout and automatic operations, the hybrid system offers the possibility of replacing time-consuming and labor-intensive laboratory procedures with mechanized batch-processing. The mass-production of the hybrid system is already feasible in semiconductor industry – cheap and portable hybrid systems can find their applications in biological labs, clinics and even in patients own hands.

We close the chapter by identifying three important tasks to further mature the hybrid technology. First, more experiments are required to assay the long-term compatibility of CMOS chips with biological samples. For example, we need to examine whether the passivation layers on chip surfaces can effectively protect the CMOS chips from salts and organic compounds in biological samples. Second, multidisciplinary collaborations between electrical, microfluidic, *and* biological/medical communities should be accelerated. In particular, the involvement of the biological/medical community, the end user of a hybrid system, is highly recommended to develop devices that are targeted at practical applications. Third, we should encourage the participation of industries, especially the semiconductor industry in this effort. Besides their financial and technological advantages, industries will bring into a device development a set of stringent quality measures that are often overlooked in academia: robustness, reproducibility and marketability. With these measures, the hybrid system can go beyond a demonstration in academic research and establish itself as a major enabling technology for biological/clinical applications.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support by NSF through NSEC (Nanoscale Science and Engineering Center) at Harvard.

REFERENCES

[1] J. S. Kilby, "Invention of the integrated circuit," IEEE Transactions on Electron Devices, vol. ED-23, pp. 648-654, 1976.

[2] G. E. Moore, "Cramming more components onto integrated circuits," Electronics, vol. 38, pp. 114-117, 1965.

[3] A. Manz, N. Graber, and H. M. Widmer, "Miniaturized total chemical analysis systems: a novel concept for chemical sensing," vol. 1, pp. 244-248, 1990.

[4] G. M. Whitesides, "The origins and the future of microfluidics," Nature, vol. 442, pp. 368-373, Jul 27 2006.

[5] J. El-Ali, P. K. Sorger, and K. F. Jensen, "Cells on chips," Nature, vol. 442, pp. 403-411, Jul 27 2006.

[6] D. J. Harrison, K. Fluri, K. Seiler, F. Zhonghui, C. S. Effenhauser, and A. Manz, "Micromachining a miniaturized capillary electrophoresis-based chemical analysis system on a chip," Science, vol. 261, pp. 895-897, 1993.

[7] T. Thorsen, S. J. Maerkl, and S. R. Quake, "Microfluidic large-scale integration," Science, vol. 298, pp. 580-584, Oct 18 2002.

[8] H. Lee, Y. Liu, E. Alsberg, D. E. Ingber, R. M. Westervelt, and D. Ham, "An IC/microfluidic hybrid microsystem for 2D magnetic manipulation of individual biological cells," 2005 IEEE International Solid-State Circuits Conference Dig. Tech. Papers, vol. 1, pp. 80-81, 2005.

[9] H. Lee, Y. Liu, R. M. Westervelt, and D. Ham, "IC/microfluidic hybrid system for magnetic manipulation of biological cells," IEEE Journal of Solid-State Circuits, vol. 41, pp. 1471-1480, 2006.

[10] M. Schienle, C. Paulus, A. Frey, F. Hofmann, B. Holzapβ, P. Schindler-Bauer, and R. Thewes, "A fully electronic DNA sensor with 128 positions and inpixel A/D conversion," IEEE Journal of Solid-State Circuits, vol. 39, pp. 2438-2445, 2004.

[11] R. H. Farahi, A. Passian, T. L. Ferrell, and T. Thundat, "Microfluidic manipulation via Marangoni forces," Applied Physics Letters, vol. 85, pp. 4237-4239, 2004.

[12] T. K. Jun and K. Chang-Jin, "Valveless pumping using traversing vapor bubbles in microchannels," Journal of Applied Physics, vol. 83, pp. 5658-5664, 1998.

[13] D. L. Huber, R. P. Manginell, M. A. Samara, B. I. Kim, and B. C. Bunker, "Programmed adsorption and release of proteins in a microfluidic device," Science, vol. 301, pp. 352-354, Jul 18 2003.

[14] D. Psaltis, S. R. Quake, and C. Yang, "Developing optofluidic technology through the fusion of microfluidics and optics," Nature, vol. 442, pp. 381-386, Jul 27 2006.

[15] A. J. DeMello, "Control and detection of chemical reactions in microfluidic systems," Nature, vol. 442, pp. 394-402, Jul 27 2006.

[16] H. A. Stone, A. D. Stroock, and A. Ajdari, "Engineering flows in small devices: microfluidics toward a lab-on-a-chip," Annual review of fluid mechanics, vol. 36, pp. 381-411, 2004.

[17] T. H. Lee, The design of CMOS radio-frequency integrated circuits. Cambridge, [England] New York: Cambridge University Press, 1998.

[18] N.-T. Nguyen and S. T. Wereley, Fundamentals and applications of microfluidics. Boston, MA: Artech House, 2002.

[19] H. Becker and C. Gartner, "Polymer microfabrication methods for microfluidic analytical applications," Electrophoresis, vol. 21, pp. 12-26, 2000.

[20] F. E. H. Tay, J. A. van Kan, F. Watt, and W. O. Choong, "A novel micromachining method for the fabrication of thick-film SU-8 embedded microchannels," Journal of Micromechanics and Microengineering, vol. 11, pp. 27-32, 2001.

[21] B. E. J. Alderman, C. M. Mann, D. P. Steenson, and J. M. Chamberlain, "Microfabrication of channels using an embedded mask in negative resist," Journal of Micromechanics and Microengineering, vol. 11, pp. 703-705, 2001.

[22] K. Lee, J. He, R. Clement, S. Massia, and B. Kim, "Biocompatible benzocyclobutene (BCB)-based neural implants with micro-fluidic channel," Biosensors & Bioelectronics, vol. 20, pp. 404-407, 2004.
[23] S. Metz, R. Holzer, and P. Renaud, "Polyimide-based microfluidic devices," Lab Chip, vol. 1, pp. 29-34, Sep 2001.
[24] M. Agirregabiria, F. J. Blanco, J. Berganzo, M. T. Arroyo, A. Fullaondo, K. Mayora, and J. M. Ruano-Lopez, "Fabrication of SU-8 multilayer microstructures based on successive CMOS compatible adhesive bonding and releasing steps," Lab on a Chip, vol. 5, pp. 545-552, 2005.
[25] H. Lee, Y. Liu, D. Ham, and R. M. Westervelt, "Integrated Cell Manipulation System - CMOS/Microfluidic Hybrid," Lab on a Chip, DOI:10.1039/B700373K.
[26] Y. N. Xia and G. M. Whitesides, "Soft lithography," Annual Review of Materials Science, vol. 28, pp. 153-184, 1998.
[27] W. J. Chang, D. Akin, M. Sedlak, M. R. Ladisch, and R. Bashir, "Poly(dimethylsiloxane) (PDMS) and silicon hybrid biochip for bacterial culture," Biomedical Microdevices, vol. 5, pp. 281-290, 2003.
[28] P. Vulto, N. Glade, L. Altomare, J. Bablet, L. Del Tin, G. Medoro, I. Chartier, A. N. Manaresi, M. Tartagni, and R. Guerrieri, "Microfluidic channel fabrication in dry film resist for production and prototyping of hybrid chips," Lab on a Chip, vol. 5, pp. 158-162, 2005.
[29] M. O. Heuschkel, L. Guerin, B. Buisson, D. Bertrand, and P. Renaud, "Buried microchannels in photopolymer for delivering of solutions to neurons in a network," Sensors and Actuators B, vol. B48, pp. 356-361, 1998.
[30] H. Becker and U. Heim, "Hot embossing as a method for the fabrication of polymer high aspect ratio structures," Sensors and Actuators A: Physical, vol. 83, pp. 130-135, 2000.
[31] M. Heckele and W. K. Schomburg, "Review on micro molding of thermoplastic polymers," Journal of Micromechanics and Microengineering, vol. 14, pp. R1-14, 2004.
[32] M. Worgull, J. F. Hetu, K. K. Kabanemi, and M. Heckele, "Modeling and optimization of the hot embossing process for micro- and nanocomponent fabrication," Microsystem Technologies, vol. 12, pp. 947-952, 2006.
[33] L. J. Kricka, P. Fortina, N. J. Panaro, P. Wilding, G. Alonso-Amigo, and H. Becker, "Fabrication of plastic microchips by hot embossing," Lab on a Chip, vol. 2, pp. 1-4, 2002.
[34] W. Ehrfeld, V. Hessel, H. Lowe, C. Schulz, and L. Weber, "Materials of LIGA technology," Microsystem Technologies, vol. 5, pp. 105-112, 1999.
[35] N. Maluf and K. Williams, Introduction to microelectromechanical systems engineering, 2nd ed. Boston: Artech House, 2004.
[36] C. K. Fredrickson and Z. H. Fan, "Macro-to-micro interfaces for microfluidic devices," Lab Chip, vol. 4, pp. 526-33, Dec 2004.
[37] B. L. Gray, D. Jaeggi, N. J. Mourlas, B. P. van Drieenhuizen, K. R. Williams, N. I. Maluf, and G. T. A. Kovacs, "Novel interconnection technologies for in-

tegrated microfluidic systems," Sensors and Actuators A (Physical), vol. 77, pp. 57-65, 1999.
[38] T. P. Hunt, H. Lee, R. M. Westervelt, "Addressable micropost array for the dielectrophoretic manipulation of particles in fluid," Applied Physics Letters, vol. 85, pp. 6421-6423, 2004.
[39] J. H. Lau and S.-W. R. Lee, Chip scale package (CSP): design, materials, processes, and applications. New York: McGraw-Hill, 1999.

AUTHOR BIOGRAPHY

Hakho Lee received the B.S. degree in physics from Seoul National University, Seoul, Korea, in 1998, and the Ph.D. degree in physics from Harvard University, Cambridge, MA, in 2005. For the Ph.D., he developed a new type of micro-total-analysis-systems (μTAS) for biological experiments based on integrated circuits and microfluidics. In 2005, he joined the Center for Molecular Imaging Research (CMIR) at the Massachusetts General Hospital, Harvard Medical School as a Research Associate. His current research interests include microelectronic actuators for biological cell manipulation, and magnetic/optical sensors for medical imaging.

Donhee Ham is an Associate Professor of electrical engineering, School of Engineering and Applied Sciences, Harvard University. He received the B.S. degree in physics in 1996 from Seoul National University, Korea, graduating with Presidential Honor atop the Natural Science College, and the PhD degree in electrical engineering from California Institute of Technology (Caltech) in 2002, winning the Charles Wilts Doctoral Thesis Prize for outstanding PhD research in electrical engineering. He was the recipient of the Caltech Li Ming Scholarship and IBM Graduate Research Fellowship. He was also the recipient of the 2003 IBM Faculty Partnership Award. He shared Harvard's Hoopes prize (best senior thesis award) in 2003 with Mr. William Andress. His work experiences also include Laser Interferometer Gravitational Wave Observatory (LIGO), Pasadena, CA, 1997/98, IBM T. J. Watson Research Center, NY, 2000, IEEE conference technical program committees including International Solid-State Circuits Conference, and industry/government technical advisory positions on subjects including high-speed electronics and future nanoelectronics technologies in the post-50nm era. His current research focus is on (1) RF & microwave ICs, (2) nanoscale quantum-effect devices for GHz & THz circuits, and (3) soliton & nonlinear wave electronics. His research also examines (4) biological laboratories on an IC. Donhee Ham's research details can be found at http://www.deas.harvard.edu/~donhee.

Robert M. Westervelt received the Ph.D. degree in physics from the University of California, Berkeley, in 1977. Following a postdoctoral appointment at Berkeley, he moved to Harvard University, Cambridge, MA, where he is currently a Mallinckrodt Professor of Applied Physics and Physics. His group investigates the

quantum behavior of electrons inside nanoscale semiconductor structures and develops tools for the manipulation of biological systems. Current research topics include: imaging electron flow through nanoscale devices at low temperatures using scanning probe microscopy, making artificial molecules composed of few-electron quantum dots for quantum information processing, and the development of hybrid integrated circuit/microfluidic chips to trap, move, assemble, and sort biological cells and small particles in fluids. He is Director of the NSF-funded Nanoscale Science and Engineering Center at Harvard University, which includes participants at the Massachusetts Institute of Technology, the University of California, Santa Barbara and the Museum of Science, Boston. Previously Westervelt was Director of the Materials Research Science and Engineering Center, and Co-Director of the Joint Services Electronics Program at Harvard.

5 CMOS-BASED MAGNETIC CELL MANIPULATION SYSTEM

Yong Liu[1], Hakho Lee[2,3], Robert M. Westervelt[1,2] and Donhee Ham[1]*

[1] School of Engineering and Applied Sciences, Harvard University
[2] Department of Physics, Harvard University
[3] Center for Molecular Imaging Research, Massachusetts General Hospital, Harvard Medical School
*donhee@deas.harvard.edu

5.1 INTRODUCTION

CMOS integrated circuits (ICs) are one of the most significant technologies in our period. With the advance of the CMOS IC technology, over hundred millions of transistors are integrated in a single chip to process gigabyte data. CMOS IC chips are ubiquitous in modern life; they can be found in computing systems, communication systems, and entertainment hardware. With all of these successful applications, lately there has been a growing interest in exploiting the benefits of the CMOS ICs for biological and biomedical applications. A variety of on-chip *analytical* instruments have been developed for different biological applications, such as studying neural activities and tissue dynamics, monitoring ion channels and electrochemical activities, and probing DNAs [1-5]. These systems can be viewed as *lab-on-IC* systems, which utilize the power of ICs to facilitate faster, programmable and standardized biological experiments with a small volume of biological sample (See chapters 8 and 9 for examples). The research activities on lab-on-IC are expected to enjoy substantial growth in the foreseeable future.

As a new type of lab-on-IC, this chapter introduces a CMOS IC manipulation system that uses IC chips as an actuator to control the motion of biological cells. Specifically, we describe a cell manipulation system based

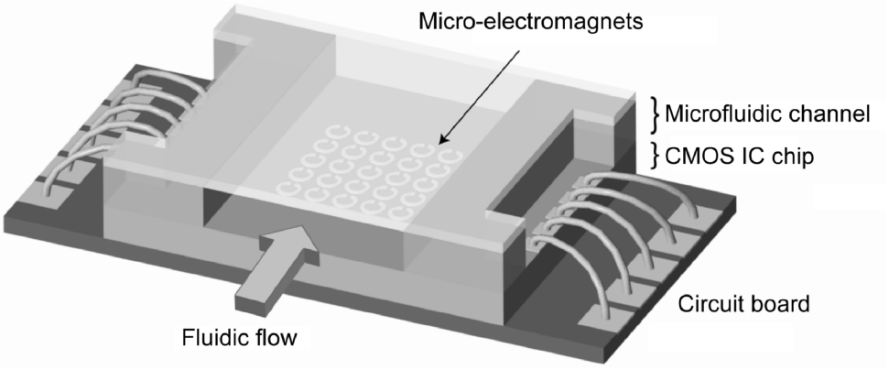


Figure 5.1 Schematic of the CMOS-based magnetic manipulation system. Reprinted with permission from [24]. Copyright © [2005] IEEE.

on the interactions between magnetic beads and micro-electromagnets; cell manipulation based on electric fields is described in [6] and Chapter 6. The CMOS cell manipulator adopts the hybrid system approach described in Chapter 4; it consists of a custom-designed CMOS chip and microfluidic channels fabricated on top as schematically shown in Fig. 5.1. The CMOS chip, with micro-electromagnets [7, 8] implemented on its surface, directly controls the motion of individual biological cells by generating local magnetic fields. Biological cells are labeled with magnetic beads to be attracted to magnetic fields. The microfluidic channels provide a pathway to introduce cells to the chip, and maintain biocompatible environments. To further enhance the biocompatibility, on-chip temperature sensors can be integrated to monitor temperature during experiments.

The combination of CMOS ICs and microfluidics in the CMOS cell manipulator brings unique advantages over conventional cell manipulation systems:

- *High Resolution*: magnetic beads have been widely used in molecular and cell biology for biological and clinical applications, such as cell and molecule manipulation and separation *in vitro*, and drug delivery *in vivo* [9]. In the conventional approaches, a large number of bead-bound molecules or cells are statistically handled. For example, in the magnetic separation of cells, a large number of bead-bound cells are statistically pulled all at once using magnetic fields with low spatial resolution. Such approach limits the application of magnetic bead in some cases where a high resolution is required. In contrast, the proposed CMOS-based cell manipulation system utilizes microscopic magnetic field patterns generated by the micro-electromagnets in the

CMOS IC chip, which permits manipulation of many individual cells with a high spatial resolution, moving each cell along a different path.

- *Programmability*: by programming the CMOS IC, the spatial magnetic field patterns for cell manipulation can be easily reconfigured in our hybrid system, which brings high flexibility inside microfluidic systems. In the conventional microfluidic applications, biological samples are moved within a fixed channel network of predetermined valve controls and conduits, and therefore, different cell manipulation functions would require different microfluidic structures. In contrast, our hybrid system can perform many different types of cell manipulations not by necessarily modifying the microfluidic system structure, but by reconfiguring the spatial pattern of the magnetic fields. In this sense the hybrid system is a *programmable* microfluidic system.

This chapter describes the operation principles, design, and demonstration of the CMOS-based cell manipulation system. General principle of cell manipulation using magnetic fields is reviewed first in the following section. Then we present the detailed description of the complete system: the CMOS chip, microfluidic channels, system packaging and the auxiliary experiment setup. The operation of the system is demonstrated by manipulating magnetic beads and biological cells.

5.2 PRINCIPLE OF MAGNETIC MANIPULATION OF CELLS

Both components of electromagnetic fields, electric and magnetic, can be used to control the motion of biological cells in fluids. In electric manipulation, electric fields interact directly with the electrical properties of biological cells as well as surrounding media to generate force on the cells. For example, in cell electrophoresis DC electric fields interact with the static charges of cells and media [10]; cell dielectrophoresis exploits the difference of polarizabilities between cells and media under AC electric fields [11]. In magnetic manipulation, however, neither biological cells nor media directly interact with magnetic fields: because of their intrinsically low magnetic susceptibilities, both biological cells and media are transparent to magnetic fields [12]. To manipulate cells using magnetic fields, extrinsic magnetic moments should be imparted to cells, which is achieved by labeling cells with magnetically responsive microspheres, magnetic beads.

In this section, we review the magnetic bead technology. Magnetic properties of magnetic beads are described, followed by the discussion on the motion of magnetic beads under external magnetic fields generated by

micro-electromagnets. Then, we describe how magnetic beads can be tagged to biological cells to give magnetic moments to the cells.

5.2.1 Magnetic Beads

Magnetic beads are polymer microspheres embedded with a large number of magnetite (Fe_3O_4) nanoparticles. In the absence of external magnetic fields, a magnetic bead has negligible magnetic moments; the magnetic moments of magnetite nanoparticles in the bead will be randomly oriented due to the thermal agitations. When external magnetic fields are applied, the bead will assume magnetic moments as the magnetic moments of the magnetite will line up in one direction. Subsequent interactions between the external magnetic fields and the acquired magnetic moments in the bead will exert force on the bead.

Magnetic beads are widely used in biomedical and clinical applications [9]. The surface of the beads can be functionalized, for example, with monoclonal antibodies to make highly specific bindings to target biological systems. By applying external magnetic fields, the target samples suspended in a fluid can be pulled out from the remaining contents, allowing fast and selective sorting of the target systems. The process is non-invasive due to the biocompatibility of magnetic fields.

Figure 5.2 shows the micrographs of commercially available magnetic beads. Magnetic beads are usually synthesized in 3-step processes [13].

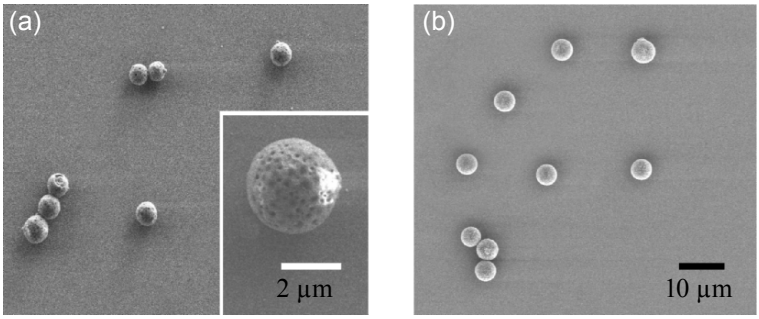


Figure 5.2 Micrographs of magnetic beads. Magnetic beads are polymer microspheres containing a large number of magnetite nanoparticles. By functionalizing the surface of the beads, biological systems, including cells, proteins, and macromolecules, can be linked to the beads to assume magnetic moments. Various magnetic beads with different sizes and surface chemistry are commercially available. (a) Magnetic beads with diameter 2.8 μm (M-280, Dynal Biotech, Oslo, Norway). (b) Magnetic beads with diameter 8.5 μm (UMC4F-6548, Bangs Laboratories Inc., IN).

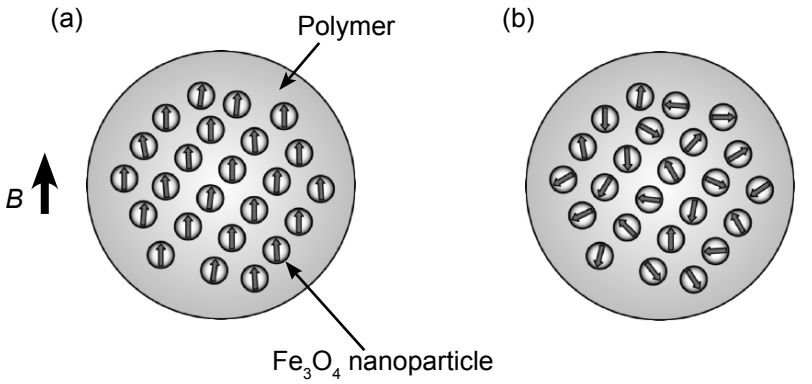


Figure 5.3 Schematics of magnetic properties of magnetic beads. The bead itself shows a paramagnetic behavior. (a) In the presence of external magnetic fields ***B***, the magnetic moments of Fe_3O_4 particles (small arrows) line up in parallel with ***B***. The bead will have appreciable magnetic moments. (b) When external fields are removed, the bead will have negligible magnetic moments because the moments of Fe_3O_4 particles are randomly oriented from thermal agitations.

First, porous polymer microspheres with a uniform size are prepared. With the advance of synthetic methods, the pores can be evenly distributed inside the spheres. In the second step, the microspheres are treated with a solution of iron salts. Upon the initiation of chemical reaction, the iron solution that fills up the pores in microspheres will precipitate, depositing single-domain magnetite nanoparticles inside the pores. Because the pores are evenly distributed in a microsphere, the concentration of magnetite particles will be uniform across the microsphere. Finally, the microspheres are covered with a polymer to fill the pores.

Even though magnetic beads contain magnetic particles with permanent magnetic moments, the beads themselves show paramagnetic characteristics, which is termed superparamagnetism. Figure 5.3 illustrates the magnetic behavior of a magnetic bead. When external magnetic fields are applied [Fig. 5.3(a)], the magnetic moments of nanoparticles will line up, giving appreciable net magnetic moments to the bead. At a given temperature T and under an external magnetic field B, the average magnetic moment m of a magnetic bead can be expressed using the Langevin function [14]

$$m = nV\mu_p \left[\coth\left(\frac{\mu_p B}{k_B T}\right) - \left(\frac{\mu_p B}{k_B T}\right)^{-1} \right], \tag{5.1}$$

where n and μ_p are the number density and the magnetic moment of nanoparticles respectively, V is the volume of a bead, and k_B is the Boltzmann

Table 5.1 Comparison of magnetic beads. The values for M-280 and M-450 are from the specification by the manufacturer (Dynal Biotech, Oslo, Norway). The magnetic susceptibility of UMC4F (Bangs Laboratories Inc., IN) is measured using a SQUID (superconducting quantum interference device) magnetometer. Magnetic susceptibilities of most biological cells are negligible compared to those of magnetic beads. Magnetic fields are biocompatible mainly due to the small magnetic susceptibility of cells.

	M-280	M-450	UMC4F	Biological cells
Diameter (μm)	2.8	4.5	8.5	1 ~ 100
Magnetic susceptibility χ	0.17	0.24	0.18	$\sim 10^{-5}$

constant. Typically, a magnetic bead is characterized by a volume magnetic susceptibility χ

$$\chi = \frac{\mu_0}{V}\frac{m}{B} \cong \mu_0 \frac{n \cdot \mu_p^2}{3k_B T}, \tag{5.2}$$

where μ_0 is the magnetic permeability of vacuum. Table 5.1 lists magnetic susceptibilities of commercial magnetic beads at room temperature. Note that because of the large differences of magnetic susceptibilities between magnetic beads and biological cells, high contrast can be achieved between magnetically tagged and untagged cell populations, which enables the selective manipulation of target cells.

When the external magnetic fields are removed [Fig. 5.3(b)], the magnetic moments of nanoparticles will spontaneously fluctuate from thermal agitation and reach thermal equilibrium. The net magnetic moment of a bead vanishes as

$$m(t) = m_0 \exp(-t/\tau), \tag{5.3}$$

where m_0 is the magnetic moment before removing the external magnetic field, t is the time lapse after the field removal, and τ is the relaxation time constant for the process. The relaxation time is given by the Néel-Brown model [15]

$$\tau = \tau_0 \exp(KV_p / k_B T), \tag{5.4}$$

where τ_0 is a time factor of the order of 10^{-9} sec, V_p is the volume of magnetic nanoparticle, and K is the anisotropy energy constant of the particle.

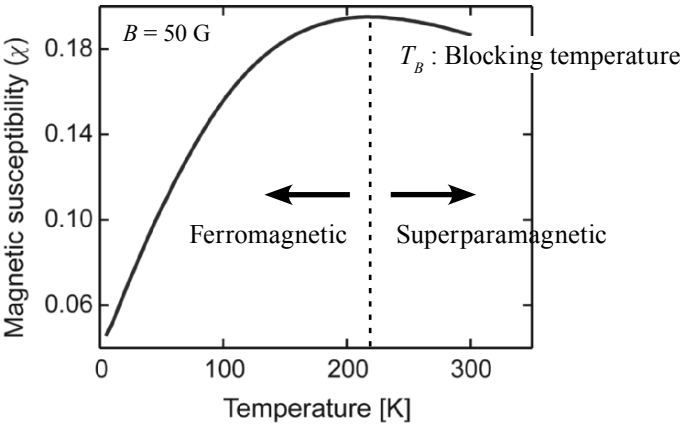


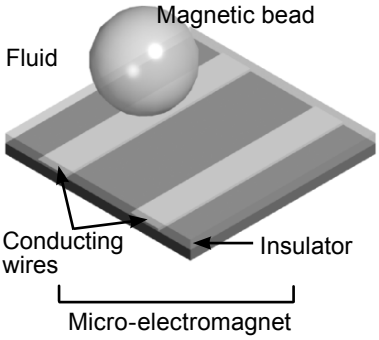
Figure 5.4 The beads shown in Fig. 5.2(b) are characterized using a SQUID magnetometer. Magnetic susceptibility χ of the bead is measured at different temperature. The bead becomes superparamagnetic when T is larger than the blocking temperature T_B. In $T > T_B$ region, $\chi \propto 1/T$ as expected in paramagnetic material.

Commercial magnetic beads are usually designed to have $KV_p < 25k_BT$, which gives $\tau < 10^2$ sec [9]. Therefore, biological cells collected by external magnetic fields can be easily dispersed again once the fields are removed.

Figure 5.4 shows the magnetic susceptibility χ of a bead in Fig. 5.2(b) as a function of temperature. The data are taken using a SQUID (superconducting quantum interference device) magnetometer (MPMS2, Quantum Design, CA). In this measurement, magnetic beads are cooled down first at the absence of external magnetic fields (ZFC: zero-field cooling) and then the magnetic susceptibility χ is measured by applying an external magnetic field $B = 5$ mT. Note that χ has the maximum value of 0.19 at $T_B = 215$ K. The temperature T_B, called a blocking temperature, indicates the onset of superparamagnetic state of a magnetic bead. When $T < T_B$, the magnetic bead is still in ferromagnetic state: χ is a monotonically increasing function of T, because thermal energy helps magnetic domains rotate to align with external magnetic fields. When $T > T_B$, thermal energy is large enough to cause random fluctuations of magnetic domains. This leads to the decrease of χ as T increases, which is a signature of paramagnetic behavior.

5.2.2 Motion of Magnetic Beads

When external magnetic fields are applied, magnetic beads are magnetized and pulled towards local field maxima. In this subsection, we examine how local magnetic fields influence the motion of a magnetic bead in a fluid.



Magnetic bead
- Radius (r_B) 4.25 μm
- Susceptibility (χ) 0.19

Micro-electromagnet
- Wire width 3 μm
- Wire thickness 0.3 μm
- Wire pitch (d) 10 μm
- Insulator thickness 1 μm

Environment
- Temperature 20°C
- Fluid viscosity (η) 1.0×10^{-3} N·sec/m^2

Figure 5.5 Model system to examine the motion of a magnetic bead. A micro-electromagnet consisting of two wires and an insulating layer generates magnetic fields. The magnetic bead shown in Fig. 5.2(b) is used as a manipulation target.

As a representative example, we consider the manipulation of a magnetic bead using a model system illustrated in Fig. 5.5. The magnetic bead has the radius r_B and the magnetic susceptibility χ and is suspended in a fluid of viscosity η. Micro-electromagnets [16] that consist of two conducing wires covered with an insulating layer are used as a magnetic field source. Microfluidic channel is placed on top of the micro-electromagnets. For the sake of simplicity, we divide the motion of a magnetic bead into two modes, a trapping and a transport. Note that magnetic beads are manipulated using the both modes in practice.

5.2.2.1 Trapping of a Magnetic Bead

Figure 5.6(a) shows the schematic of the trapping operation. With an electric current flow I, the micro-electromagnet generates a local magnetic field $\boldsymbol{B}$ which magnetizes and holds the magnetic bead. The trapping potential energy U is

$$U = -\frac{1}{2}\boldsymbol{m} \cdot \boldsymbol{B}, \tag{5.5}$$

where $\boldsymbol{m}$ is the magnetic moment of the bead. The factor 1/2 comes from the fact that $\boldsymbol{m}$ is induced by $\boldsymbol{B}$. Due to the paramagnetic nature of the bead, $\boldsymbol{m} \propto \boldsymbol{B}$. From the definition of χ in Eq. (5.2).

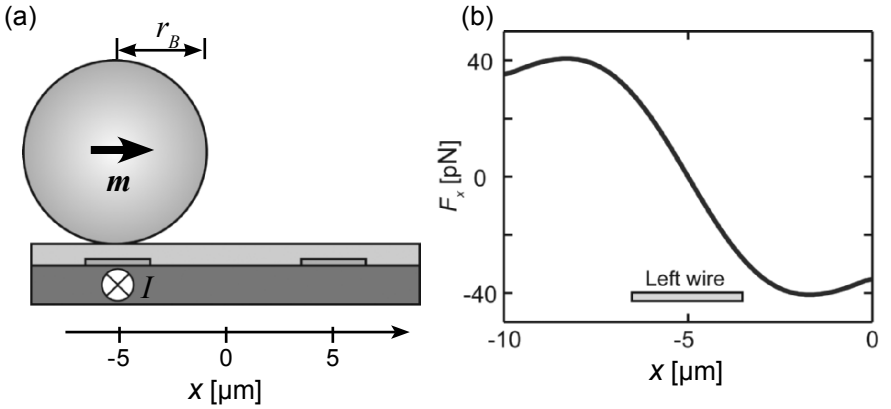


Figure 5.6 Trapping operation using the model system. (a) A magnetic bead is trapped by flowing a current I to the left wire. (b) Estimated lateral trapping force F_x. The bead has $r_B = 4.25$ μm and $\chi = 0.19$, and $I = 100$ mA is flowing through the wire. The maximum force on the bead across the wire is ~40 pN.

$$\boldsymbol{m} = \frac{\chi V}{\mu_0}\boldsymbol{B}, \tag{5.6}$$

where V is the volume of the bead. This leads to

$$U = -\frac{1}{2}\frac{\chi V}{\mu_0}B^2. \tag{5.7}$$

From U, the trapping force $\boldsymbol{F}_m$ by the magnetic field is

$$\boldsymbol{F}_m = -\nabla U = \frac{1}{2}\frac{\chi V}{\mu_0}\nabla B^2. \tag{5.8}$$

By flowing a large current in the micro-electromagnet, the trapping force can be easily increased because $F_m \propto B^2 \propto I^2$. However, to achieve a large current flow, the micro-electromagnet should be cooled down to prevent thermal breakdown. With an active cooling, the maximum current density in a metal wire can be up to 400 mA/μm^2 [16]. Figure 5.6(b) shows the estimated lateral trapping force by the model system. The magnetic bead shown in Fig. 5.2(b) with $r = 4.25$ μm and $\chi = 0.19$ is used for the calculation. With $I = 100$ mA, the maximum trapping force on the bead is 40 pN. Note that typical trapping force on magnetic beads using micro-electromagnets is on the order of 100 pN [8].

When there is a fluidic flow, the trapped magnetic bead will experience viscous drag

$$\boldsymbol{F}_d = -6\alpha\pi\eta r_B \boldsymbol{v}_0, \tag{5.9}$$

where $\boldsymbol{v}_0$ is the velocity of the bead relative to the fluidic flow and α is the correction factor to Stokes law from Faxen's formula [17]. When a bead is trapped on the surface of the device, $\alpha = 3.1$. For stable trapping, the lateral trapping force should be greater than that of viscous drag, which sets the upper limit in a fluidic flow. In the model system shown in Fig. 5.5(a),

$$v_0 < \frac{\chi}{36\alpha\eta}\frac{r_B^{\,2}}{\mu_0}\left|\frac{\partial B^2}{\partial x}\right|. \tag{5.10}$$

Using the force profile shown in Fig. 5.6(b), $v_0 < 160$ μm/sec for the magnetic bead with $r_B = 4.25$ μm.

5.2.2.2 Transport of a Magnetic Bead

By changing magnetic field patterns, a trapped magnetic bead can be moved over the surface of the device. Figure 5.7 illustrates the transport of a magnetic bead between two wires with the wire pitch *d*. Initially, a magnetic bead is trapped at the left wire by flowing a current in the wire. Subsequently, the current in the left wire is turned off and the current in the right wire is turned on to move the bead. To describe the motion of bead, following assumptions are made.

1. During the transport, the magnetic bead is assumed to have a fixed magnetic moment inside. While being trapped by the left wire, the bead acquires an induced magnetic moment $\boldsymbol{m}_0$ given by Eq. (5.6). When the right wire is activated, the magnetic moment of the bead will change as a response to the altered external magnetic field $\boldsymbol{B}$. The response time of the moment change is $\tau_r = \tau \exp(-\mu_p B/k_B T)$, where τ is the relaxation time given by Eq. (5.4) and μ_p is the moment of a magnetic nanoparticle in a bead. In a typical experimental condition where a magnetic bead is moved by a micro-electromagnet, $\tau_r \sim 10$ sec with $\mu_p B/k_B T \sim 1$ and $\tau \sim 10^2$ sec, whereas the transport time of a bead itself is usually $<10^{-1}$ sec. Therefore, the magnitude of the magnetic moment of the bead can be approximated as a constant m_0 during the transport.

2. Because the bead is manipulated in a fluidic environment where the viscous effect by the fluid is dominant, the inertial motion of the bead can be neglected. Therefore, the total force and the total torque on the bead vanish, giving

$$\boldsymbol{F}_m + \boldsymbol{F}_d + \boldsymbol{F}_n = 0, \tag{5.11}$$

$$\boldsymbol{N}_m + \boldsymbol{N}_d = 0, \tag{5.12}$$

where $\boldsymbol{F}_n$ is the normal force by the surface of the device, and $\boldsymbol{N}_m$ and $\boldsymbol{N}_d$ are torque on the bead from magnetic fields and viscous drag, respectively.

Based on the assumptions given above, the equation for translational motion of the bead in Fig. 5.7 is

$$\begin{aligned} &(\boldsymbol{F}_m + \boldsymbol{F}_d)\cdot \mathbf{e}_x = (\boldsymbol{m}_0 \cdot \nabla \boldsymbol{B})\cdot \mathbf{e}_x - 6\alpha\pi\eta r_B \dot{\boldsymbol{x}}\cdot \mathbf{e}_x = 0 \\ &\Rightarrow \dot{x} = \frac{m_0}{6\alpha\pi\eta r_B}\left(\cos\theta \cdot \frac{\partial B_x}{\partial x} + \sin\theta \cdot \frac{\partial B_y}{\partial x}\right) \end{aligned} \tag{5.13}$$

where $\mathbf{e}_x$ is the unit vector in the x-direction. For rotational motion, torques are given as

$$\boldsymbol{N}_m = \boldsymbol{m}_0 \times \boldsymbol{B}, \tag{5.14}$$

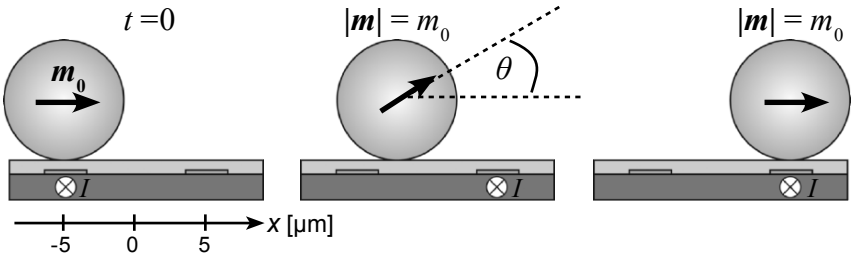


Figure 5.7 Transport operation using the model system. A magnetic bead is initially trapped by the left wire with the current *I*. The bead is then moved by turning off *I* in the left and turning on *I* in the right wire. The bead undergoes translational as well as rotational motion. To simplify the motion of the bead, the magnetic moment of the bead is assumed to have a constant magnitude m_0 acquired during the trapping, and inertial force is ignored.

$$\boldsymbol{N}_d = -8\pi\eta r_B^{\,3}\left(\dot{\theta} + \frac{15}{256 r_B}\dot{x}\right)\mathbf{e}_\theta \tag{5.15}$$

where $\mathbf{e}_\theta$ is the unit vector. Note that the torque by the viscous drag is coupled with a translational motion [17]. The equation for rotational motion is then

$$\dot{\theta} = \frac{m_0}{8\pi\eta r_B^{\,3}}\left(\cos\theta \cdot B_y - \sin\theta \cdot B_y\right) + \frac{15}{256 r_B}\dot{x}. \tag{5.16}$$

It is worthwhile to compare the time scale between the translational and the rotational motion of a magnetic bead. From Eq. (5.13), the time τ_T taken by the bead to move from the left to the right wire is on the order of

$$\tau_T \sim O\left(\frac{d}{\dot{x}}\right) = O\left(\frac{d\eta r_B}{m_0}\cdot\frac{d}{B}\right) = O\left(\frac{d^2\eta r_B}{m_0 B}\right), \tag{5.17}$$

where d is the distance between two wires. From Eq. (5.16), the order of time τ_R for rotational motion is estimated as

$$\tau_R \sim O\left(\frac{\pi}{\dot{\theta}}\right) = O\left(\frac{\eta r_B^{\,3}}{m_0 B}\right). \tag{5.18}$$

From Eqs. (5.17) and (5.18),

$$\tau_R / \tau_T \sim O(r_B^{\,2} / d^2). \tag{5.19}$$

Because $r_B < d$ for typical experimental conditions, $\tau_R << \tau_T$, which means that the bead quickly rotates itself during the transport, aligning $\boldsymbol{m}_0$ to the external $\boldsymbol{B}$. Therefore, the rotational motion of the bead can be ignored for the first approximation during the transport operation.

Using Eqs. (5.13) and (5.16), the trajectory of a magnetic bead transported by the model system is calculated and shown in Fig. 5.8. The magnetic bead has $r_B = 4.25$ μm and $\chi = 0.19$. The bead is initially trapped by the left wire in Fig. 5.7 by flowing $I = 100$ mA. At $t = 0$, the left wire is turned off and the right wire is turned on with $I = 100$ mA. Figure 5.8(a) shows the position of the bead center as a function of time. The bead is moved to the right wire in <80 msec, which corroborates the assumption that the magnitude of $\boldsymbol{m}$ of the bead can be approximated as a constant. Figure 5.8(b) plots the

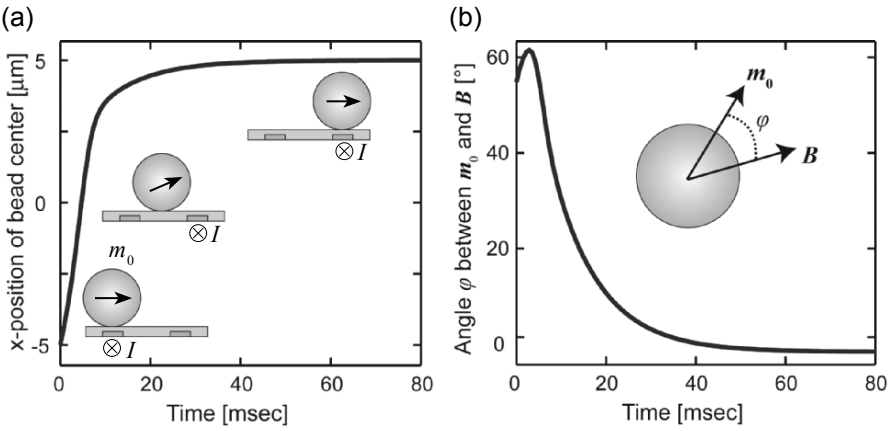


Figure 5.8 Simulated motion of a magnetic bead. The bead has $r_B = 4.25$ μm and $\chi = 0.19$. A bead is trapped by the left wire with $I = 100$ mA. Induced magnetic moment of the bead is $m_0 = 1.4 \times 10^{-13}$ Am^2, which is assumed to be constant during the transport. The bead is moved by flowing $I = 100$ mA in the right wire. (a) Lateral position of the bead center as a function of time. The bead reaches the right wire in ~80 msec. (b) The bead also undergoes rotational motion to align $\boldsymbol{m}_0$ to external magnetic fields $\boldsymbol{B}$. Note that the discrepancy in the directions of $\boldsymbol{m}_0$ and $\boldsymbol{B}$ decreases quickly (<30 msec). As the first-order approximation, rotational motion can be ignored when describing the motion of beads in a typical operation of micro-electromagnets.

instantaneous angles between $\boldsymbol{m}_0$ and $\boldsymbol{B}$ while the bead is moving. Note that the direction of the magnetic moment closely follows that of external fields when $t > 30$ msec.

5.2.3 Tagging Biological Cells with Magnetic Beads

To tag biological systems, including eukaryotic and prokaryotic cells, proteins, and macromolecules, with magnetic beads, the surface of the beads is coated with molecules that can recognize the target sample. Molecular-specific bindings between the coated beads and the target biological systems enable the highly selective manipulation of target samples using magnetic fields. A variety of magnetic beads are commercially available that already have functional coatings for widely-used types of cells and proteins. Customized bead-coating can also be performed using generic magnetic beads to meet experimental purposes.

In this subsection, we discuss a general approach to functionalizing magnetic beads with biological ligand. Three representative coating methods: adsorption, covalent coupling and affinity binding are described in the

following. Note that detailed, well-documented coating protocols are usually available from magnetic bead manufacturers.

5.2.3.1 Adsorption

Adsorption provides the simplest way to coat magnetic beads with proteins. Proteins readily adsorbs onto hydrophobic surfaces, especially that of polymer [18]. Because magnetic beads are made of polymer, usually polystyrene, simply mixing the beads with a solution of protein can lead to surface coating. The protein coated beads then can be linked to target biological systems.

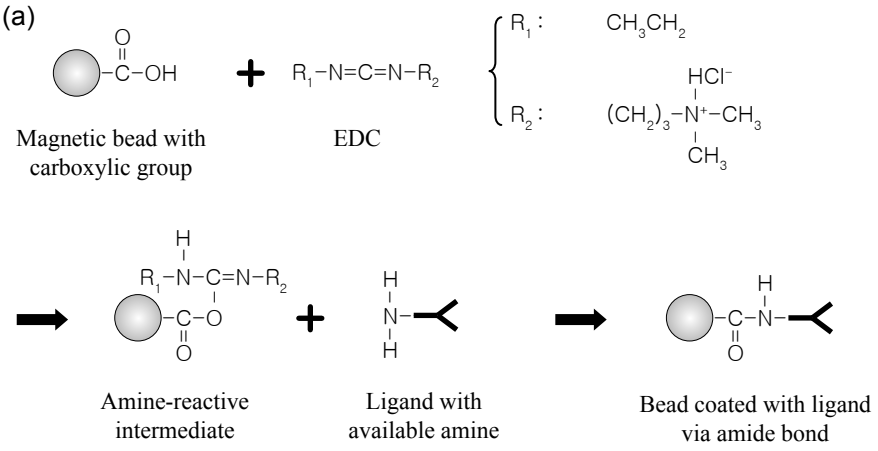
For effective coating, the concentrations of the protein and the magnetic bead are optimized based on the binding capacity of the bead for the protein used. For example, the surface binding capacity of polystyrene for bovine serum albumin (BSA) and bovine γ-globulin (BIgG) are 3 mg/m^2 and 2.5 mg/m^2, respectively [19]. To coat a monolayer of BSA or BIgG on magnetic beads that have the diameter 1 μm and are suspended with the concentration 10^9 bead/ml, required concentrations of BSA and BIgG are 9.4 μg/ml and 7.8 μg/ml, respectively. In practice, 3 to 10 times excess proteins are added to the protein solution.

5.2.3.2 Covalent Coupling

Magnetic bead with surface functional groups, including carboxylic (–COOH) or amino (–NH_2) groups, can be coated with proteins via covalent bonding. Covalent coating offers several advantages over the adsorption method. First, the coating can be highly selective, because only the proteins that can react with the surface groups on the beads are bound. Second, molecules are permanently attached to the surface with better thermal stability. Especially, small molecules should be linked to the bead surface using covalent coupling to prevent desorption over time. Third, the covalent coupling is a preferable method to coat small magnetic beads. Magnetic beads with –COOH or –NH_2 surface groups are hydrophilic and can easily be dispersed in water-based solution regardless the size of the beads. In contrast, hydrophobic magnetic beads used in the adsorption method can flocculate as the size of the beads becomes <1 μm.

As an example of covalent coupling, Fig. 5.9(a) shows the procedure of attaching proteins to magnetic beads with –COOH surface group. Carbodiimide is used as a coupling agent to make a link between –COOH on the bead surface and –NH_2 in proteins [20]. First, magnetic beads are activated by adding a water-soluble carbodiimide, 1-ethyl-3-(3-dimethyla

minopropyl)carbodiimide hydrochloride (EDC), which leads to the formation of an amine-reactive intermediate, O-acylisourea, on the bead surface. Then ligand with $–NH_2$ group is introduced. From the reaction between O-acylisourea and $–NH_2$, amide bond (CONH) is formed to bind the ligand to the bead surface.



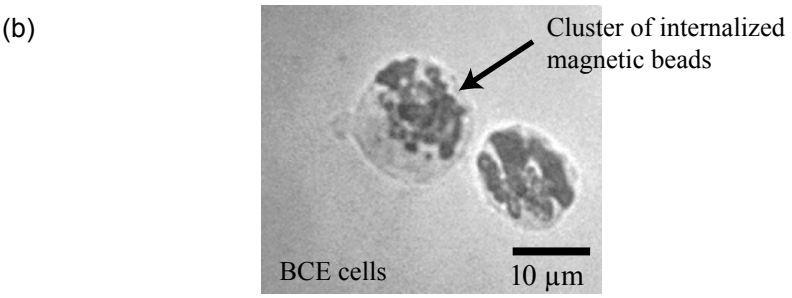


Figure 5.9 Covalent coating of magnetic beads. Magnetic beads with functional groups can be covalently coupled to desired ligands. (a) Schematic of bead coating using carbodiimide chemistry. Magnetic beads with carboxylic group (–COOH) are activated using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and then are mixed with the target ligand that has amine group ($–NH_2$). The ligand is linked to the bead surface via amide (–CONH–) bond. (b) Example of magnetically tagged cells using covalently coupled magnetic beads. A sequence of peptides is coupled to magnetic beads of diameter 250 nm via carbodiimide chemistry. Bovine capillary endothelial (BCE) cells are cultured with the coated beads, which leads to internalization of the beads through endocytosis. These BCE cells are used to demonstrate cell manipulation in Section 5.6.

Carbodiimide coupling can be used to bind peptides to magnetic beads. The peptide-coated beads then can be attached to or even engulfed by biological cells. Figure 5.9(b) shows the micrograph of bovine capillary endothelial (BCE) cells tagged with magnetic beads. The beads are coated with RGD peptides via carbodiimide chemistry. Co-culturing BCE cells with the coated beads leads to the internalization of the beads by BCE cells through endocytosis. The magnetically tagged BCE cells are individually manipulated as described in Section 5.6.

5.2.3.3 Affinity Binding

Magnetic beads that are already conjugated with various types of generic binding proteins are commercially available. As illustrated in Fig. 5.10, these pre-conjugated beads can be coated with desired proteins via affinity binding between proteins. For example, magnetic beads with streptavidin pre-coating can be easily covered with a second ligand with biotin [21]. Because most of biological ligands can be linked to biotin (biotinylation) [22], streptavidin-coated magnetic beads can be functionalized with virtually any kind of ligand. Furthermore, the streptavidin and biotin bonds are one of the strongest non-covalent bindings [23], providing stable attachment of ligand.

One of most important advantages of using affinity binding comes from the commercial availability of the pre-coated magnetic beads that are well characterized by manufacturers. Second, compared to covalent coupling of proteins, the affinity binding offers a simpler way to coat beads, which typically involves a one-step procedure. Yet, the binding strength is greater than that from adsorption method, nearly achieving the strength of covalent bonding.

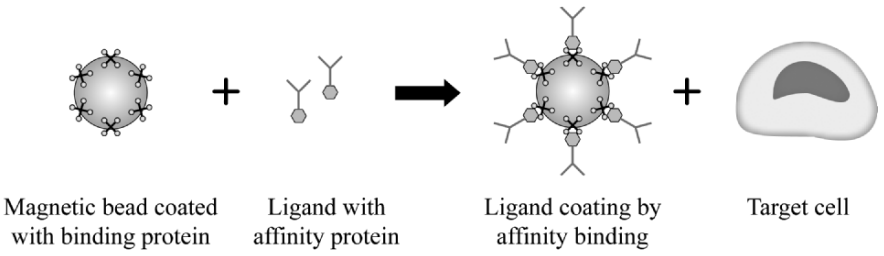


Figure 5.10 Coating of magnetic beads via affinity binding. Magnetic beads that are pre-coated with generic binding proteins, such as streptavidin or Protein A, are commercially available. By preparing ligands with proper proteins, the beads can be coated from the affinity binding of proteins.

5.3 DESIGN OF THE CMOS IC CHIP

The CMOS chip for cell manipulation is designed to generate multiple, localized magnetic field peaks on its surface by using an array of surface microcoils. Generating multiple magnetic peaks is desirable to control the positions of many individual cells in parallel, which can be used, for example, to study intercellular communications and to assemble artificial cellular structures.

Figure 5.11 shows the functional diagram and micrographs of the implemented CMOS chip for cell manipulation. As shown in Fig. 5.11(a), the chip contains both analog and digital circuits. The analog circuits include 8×8 microcoil arrays that generate magnetic field patterns for cell manipulation, a variable 8-step current source for the arrays, and on-chip temperature sensors to measure local temperature of the chip. As the digital parts, implemented are row/column decoders to choose a specific microcoil in an array and FET (field-effect transistor) switch networks to control the direction and the magnitude of the electrical current in the selected microcoil. The digital circuits facilitate the operation of the chip by reducing the number of external inputs for chip control. In total, ten synchronous external inputs are required to operate an 8×8 microcoil array: six signals to address individual microcoils, three to set the magnitude of the electrical current in a microcoil, and one to set the direction of the current.

Figure 5.11(b) shows the die micrograph of the CMOS chip. The lateral size of the chip is 2 mm × 5 mm, and the supply voltage is V_{dd} = 3.3 V. The chip contains five microcoil arrays with slightly different structure; only one array is enabled during experiments. Figure 5.11(c) shows the close-up of one of the microcoil arrays. The array consists of 64 identical microcoils arranged in 8 rows and 8 columns. The outer diameter of the microcoil is 20 μm and the center-to-center distance between two adjacent coils is 21 μm. To accurately monitor the temperature during the cell manipulation, on-chip temperature sensors are placed in the vicinity of the microcoil array.

In the following subsections, we describe the key building blocks of the CMOS chip: the microcoil array, control circuitry, and temperature sensors.

5.3.1 Microcoil Array

In order to manipulate the magnetic-bead-bound biological samples in the hybrid system, micro-electromagnets are implemented to produce magnetic

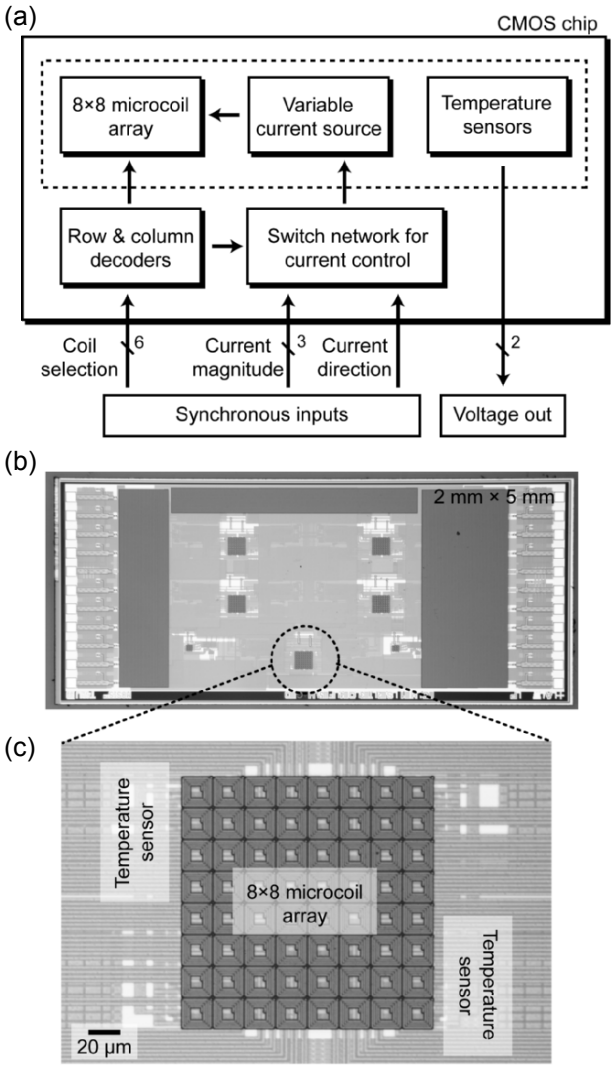


Figure 5.11 (a) The CMOS chip contains both analog and digital circuits. Analog circuits (inside the dotted box) include an 8×8 microcoil array, a current source to the array, and temperature sensors. Digital logic circuits (decoders and switch networks) control the array. The temperature sensors output voltages proportional to local temperature near the array. (b) Die micrograph of the CMOS chip. The chip has the footprint of 2×5 mm^2, and contains five microcoil arrays. (c) Close-up photo of an 8×8 microcoil array. The outer diameter of a microcoil and the center-to-center pitch of two adjacent microcoils are 20 μm and 21 μm, respectively. Temperature sensors, which are on the substrate level, are placed near the array. Reprinted from [27] by permission of The Royal Society of Chemistry.

field peaks on the surface of the chip. A variety of micro-electromagnets have been developed for cell manipulation, including a ring trap, a matrix, and a microcoil [8, 26]. We have chosen the microcoil as a magnetic field source in the current system, because the microcoil generates a highly localized peak in the magnetic field magnitude on the surface of the chip. By laying out microcoils into a 2-dimensional array, many magnetic field peaks can be produced simultaneously to independently manipulate individual biological cells.

The dimensions of the microcoil should 1) obey the design rule set by the foundries and 2) be optimized to produce a single magnetic field peak on the chip surface. Following guidelines are applied in the design of the microcoil.

1. The minimum width (W) of the metal line in the microcoil is determined $W = I/D$, where D is the maximum line current density (A/m) for the metal given by the chip foundry and I is the target current flowing in the metal line.
2. The outer diameter of the microcoil is similar to the size of biological cells to be manipulated, typically, tens of micrometers. This condition ensures the trapping of a single biological cell per each microcoil.
3. Once the outer diameter is set, the number of turns in the coil is swept to make sure that a single magnetic field peak is generated on the surface of the device.
4. Interconnections between the microcoil and the current sources are made using the metal layer farthest away from the chip surface to minimize stray magnetic fields.

Figure 5.12 shows the characteristics of the microcoil implemented on the chip. As schematically shown in Fig. 5.12(a), the microcoil is composed of two planar coils that are patterned in the top and the second metal layer and vertically connected by a via. The metal layers are separated by 1 μm thick SiO_2, and the top surface of the device is covered with a 2.5 μm thick passivation layer. Each planar coil has four turns with a 20-μm outer diameter. Figure 5.12(b) shows the calculated magnetic field profiles across the microcoil and on the chip surface; a single magnetic field peak is produced at the coil center on the surface of the top insulation layer. Note that most of the contribution to the field magnitude comes from the planar coil in the top metal layer; the planar coil in the second metal layer is primarily used to shape a single peak at the center of the microcoil on the chip surface. To form an array, 64 microcoils are arranged in 8 rows and 8 columns (an 8 × 8 microcoil array) as shown in Fig. 5.11(c).

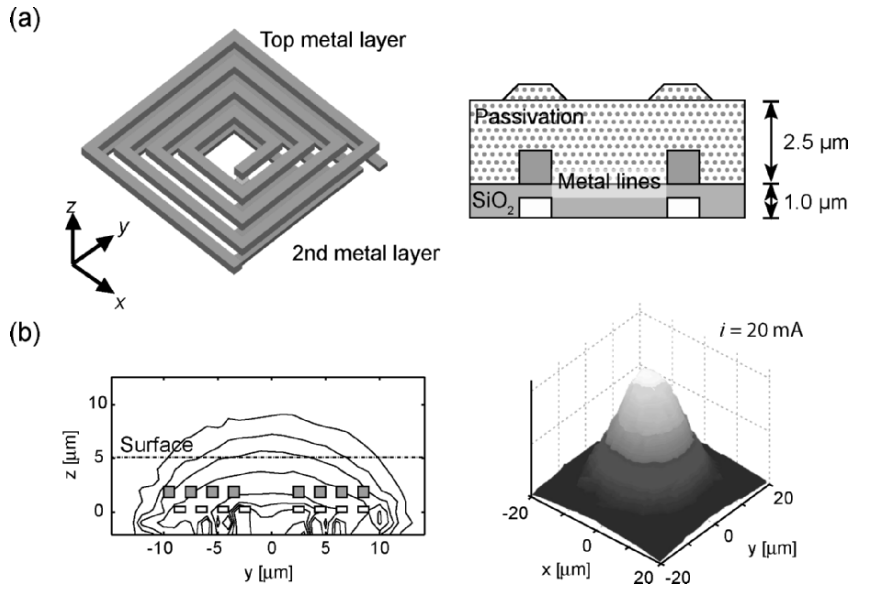


Figure 5.12 Design and magnetic field profiles of a microcoil. The microcoil is designed to generate a single peak in the magnetic field magnitude at its center. (a) The structure of a microcoil. Two planar coils, with each coil in a different metal layer, are connected by vias to form a microcoil. (b) Calculated magnetic field profiles across the microcoil and on the surface of the CMOS chip. A single peak is formed on the chip surface at the center of the microcoil.

5.3.2 Control Circuitry

The CMOS IC chip incorporates digital logic circuit to control a large microcoil array with low power consumption and to provide high spatial resolution in cell manipulation. With the full exploitation of CMOS electronics, an efficient, simultaneous manipulation of multiple bead-bound cells can be performed.

5.3.2.1 Low Power Operation of the Microcoil Array

To operate an $N \times N$ microcoil array, the simplest protocol is to connect one separate on-chip current source to each microcoil, and to continuously flow a current through a microcoil for transport or trapping of a bead-bound cell. However, when a large number of cells are handled by a large microcoil array, this operation protocol causes a severe power dissipation problem. Large power dissipation might cause thermal failure of the chip and damage to cells through heat shock. To solve this problem, a low-power operation

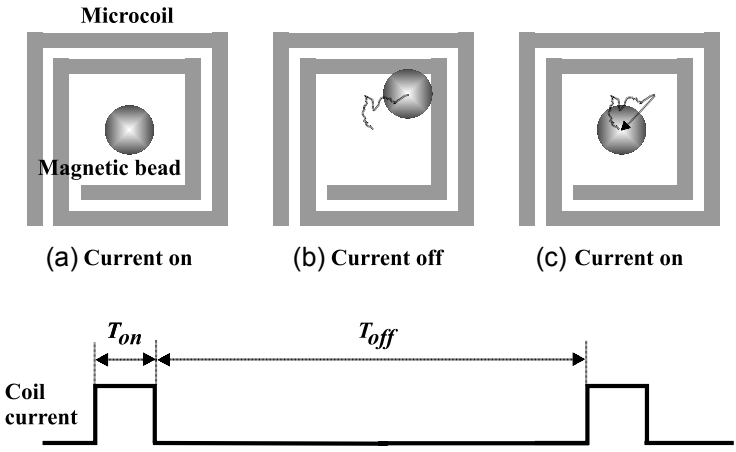


Figure 5.13 Basic working principle of the low power operation. (a) With a current in a microcoil, a magnetic bead is trapped. (b) When the current is off, the bead diffuses out of the microcoil center because of Brownian motion. (c) If the current is on before the bead escapes the microcoil, the bead can be trapped again. Repeating these steps, the bead can remain trapped by the same microcoil without flowing a continuous DC current. Reprinted with permission from [25]. Copyright © [2006] IEEE.

protocol of a large microcoil array is adopted in the CMOS IC, exploiting the fact that CMOS electronics is much faster than the motion of the cells in fluid.

The basic concept of the low-power protocol is illustrated in Fig. 5.13 using one microcoil. To trap a magnetic bead (or a bead-bound cell) without flowing a continuous current through the microcoil, the microcoil is activated for a time duration of T_{on} and is then deactivated for the following time duration of T_{off}. Such a current flow in the period of $T_p = T_{on} + T_{off}$ is repeated. During the on-time T_{on}, the magnetic bead remains trapped at the center of the microcoil [Fig. 5.13(a)]. While during the following off-time T_{off}, due to Brownian motion, the trapped bead walks away from the center of the microcoil [Fig. 5.13(b)]. However, before the bead completely escapes the microcoil, the microcoil is activated again, hence pulling the bead back to the center of the microcoil [Fig. 5.13(c)]. Therefore, with this protocol, the magnetic bead can remain virtually trapped with an effective power dissipation $T_{on}/(T_{on} + T_{off})$ times lower than that of supplying a continuous current to the microcoil.

The low-power operation protocol for a single microcoil can be extended to the case when an $N \times N$ microcoil array is used to handle a larger number of bead-bound cells. This can be done by connecting all the microcoils

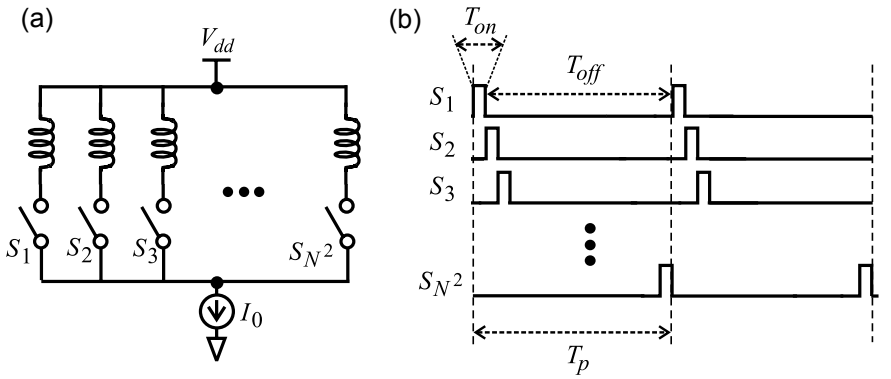


Figure 5.14 Application of the low power operation to $N \times N$ microcoil array. (a) Each microcoil in the array is connected to a single current source through a switch (S_i). (b) Current timing diagram for sequential activation of N^2 microcoils. Note that only one microcoil is on at any given moment. Reprinted with permission from [25]. Copyright © [2006] IEEE.

to one single current source controlled by switches [Fig. 5.14(a)] and by sequentially turning on and off a selected group of microcoils needed for a specific manipulation by properly clocking the switches. Figure 5.14(b) shows an example of this sequential current sharing operation when N^2 bead-bound cells are needed to be trapped by all of the microcoils, one cell for each microcoil. In a given period $T_p = N^2 \cdot T_{on}$, each microcoil is activated only once for a duration of T_{on}. At any given time, there is only one activated microcoil. As long as the off-time of each microcoil $T_{off} = (N^2 - 1) \cdot T_{on}$, is shorter than the characteristic coil escape time due to Brownian motion, the microcoil array can maintain the trap of the N^2 cells, sharing the common current source at different times, hence allowing low power operation. Because CMOS electronics is orders of magnitude faster than the cell motion in the fluid, this sequential sharing of the common current can be executed with high speed to work with a large number of microcoils.

One can analyze the limit of the current sharing due to Brownian motion as follows. During the off-time T_{off}, a magnetic bead trapped in a microcoil will diffuse from the center of the microcoil with the mean displacement

$$s = (D \cdot T_{off})^{1/2} = [D \cdot (N^2 - 1) \cdot T_{on}]^{1/2}, \tag{5.20}$$

where D is the diffusion constant of the bead in the fluid. For a magnetic bead of radius r_B in a fluid with viscosity η,

$$D = k_B T/(6\pi\eta r_B), \tag{5.21}$$

where k_B is the Boltzmann constant and T is the absolute temperature. To trap the diffused bead back to the coil center during the on-time, the dis-

placement s should be smaller than the radius r of the microcoil. This condition sets the upper limit on the on-time

$$T_{on} < r^2/ [D \cdot (N^2 - 1)], \tag{5.22}$$

and the period

$$T_p < r^2N^2/ [D \cdot (N^2 - 1)]. \tag{5.23}$$

For example, if the radius of the microcoils is 5 μm and the diameter of magnetic beads is 1 μm, for a 10 × 10 microcoil array, $T_{on} < 0.6$ sec and $T_p < 60$ sec when the beads are suspended in water at 25°C. These conditions are easily satisfied as electronic circuits in the chip operate at much higher speed.

The previous example clearly demonstrates the benefit of the sequential current sharing protocol to simultaneously "trap" multiple bead-bound cells. To simultaneous "transport" multiple bead-bound cells, while a low-power operation is still achieved with the sequential current sharing scheme, it comes at the cost of an increased transport time. This is because bead-bound cells are pulled by the microcoil not continuously but for a short duration (T_{on}) in each period (T_p). This power–time trade-off is analogous to *principle of work* exploited widely in engineering systems. In our case, reduction of power at the expense of the increased transport time is desired to avoid chip failure and to maintain biocompatible environment.

5.3.2.2 Enhancement of the Spatial Manipulation Resolution

Besides creating a single magnetic peak at each microcoil, one can generate many different magnetic field patterns by changing current distributions in a microcoil array. One application of such technique is to increase the spatial resolution of the magnetic field peaks.

Figure 5.15 illustrates how a peak in the magnetic field magnitude can be produced at the locations other than the centers of microcoils, increasing the spatial resolution of cell trapping. Both the magnitude and the direction of currents need to be adjusted as shown in Fig. 5.15(b). Note that during the peak movements, the magnitude of the peak remains nearly constant, but the direction of the magnetic field is changing. For example, in the Step 1 (Step 5), the main field component is in the $+z$ ($-z$) direction at the peak position, whereas in Step 3, the field is in the $+x$ direction at the peak position.

In the actual chip operation, the resolution enhancement scheme is combined with the low-power protocol explained in Subsection 5.3.2.1; the two microcoils in Fig. 5.15 have to share the same current source. Therefore, to mimic the magnetic field patterns shown in Fig. 5.15, currents with appro-

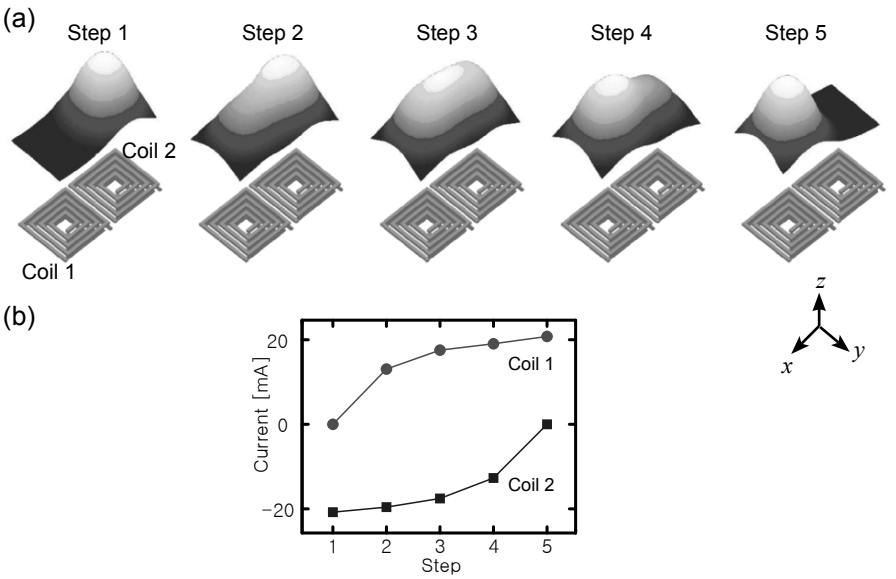


Figure 5.15 Dynamic magnetic field generation using a microcoil array. Versatile magnetic field patterns can be created by dynamically changing current distributions in a microcoil array. (a) A single magnetic peak is created and continuously moved between two adjacent microcoils. (b) Current distribution to generate magnetic field patterns in (a).

priate magnitude and polarity are pulsed to each coil alternatively. In this dynamic operation procedure, the static magnetic field magnitude profile shown in Fig. 5.15 is valid as an *average* field magnitude that a magnetic bead experiences.

5.3.2.3 Overall Architecture of the Control Circuit

The overall architecture of the control circuit that executes both the current sharing protocol and the manipulation resolution enhancement scheme is shown in Fig. 5.16. In this architecture, each microcoil is connected to three switches composed of CMOS transistors. The two switches, S_{D1} and S_{D2}, control the current direction of the microcoil. The switches labeled S_C connect the microcoils to a *common* eight-step current source. The current sharing protocol is achieved by sequentially activating and deactivating a selected group of S_C's using the column and row decoders.

To implement the resolution enhancement scheme, both current magnitude and direction need to be controlled. The current magnitude in the microcoil is set by the common eight-step current source in conjunction with the on-chip thermometer encoder. The thermometer encoder translates the

3-bit binary input signal, Control[2:0], into an output code, c[7:0], that enables individual current sources. Because each branch in the current source draws 2.5 mA, the microcoil current can reach up to 20 mA in the increments of 2.5 mA. The current direction is set by the two switches, S_{D1} and S_{D2}, which are gated by a current direction control signal. Since only one microcoil is activated at a given time from the current sharing protocol, one single direction control signal line is shared by all the microcoils. When the

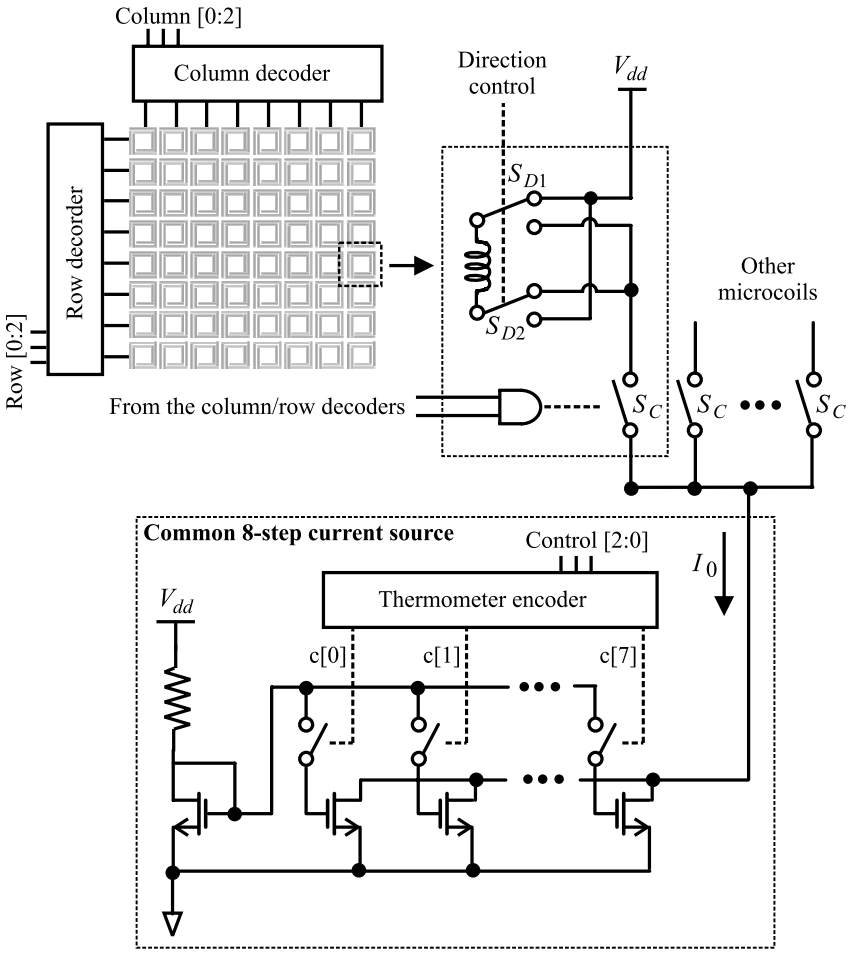


Figure 5.16 Overall architecture of the control circuit. Each microcoil is connected to the common current source through the switch S_C. The direction of the current is controlled by the switches S_{D1} and S_{D2}. The magnitude of the current is changed by individually activating each branches in the 8-fold current mirror. Reprinted with permission from [25]. Copyright © [2006] IEEE.

common eight-step current source is switched from one microcoil to another according to the current sharing protocol, the direction control signal changes its polarity if the current direction has to be changed. If the current magnitude is to be modified with the switching, the thermometer encoder will send a new control code to the eight-step current source to set a proper magnitude of DC current.

5.3.3 Temperature Sensor

To prevent the thermal failure of the CMOS IC chip and to minimize thermal shock to cells, it is essential to regulate the temperature of the hybrid system throughout the cell manipulation. To facilitate the temperature control, on-chip temperature sensors have been implemented across the CMOS IC chip. The output information from the temperature sensors is sent to an external thermoelectric cooler (TEC) controller. The TEC controller then cools down or heats up the microscope stage to regulate the temperature of the chip within a biocompatible range (4°C to 37°C).

Figure 5.17(a) shows the schematic of the temperature sensor that uses PNP bipolar transistors available in a standard CMOS process [13]. With a constant current I_1, the emitter voltage V_{E1} in the transistor Q_1 is proportional to the absolute temperature T

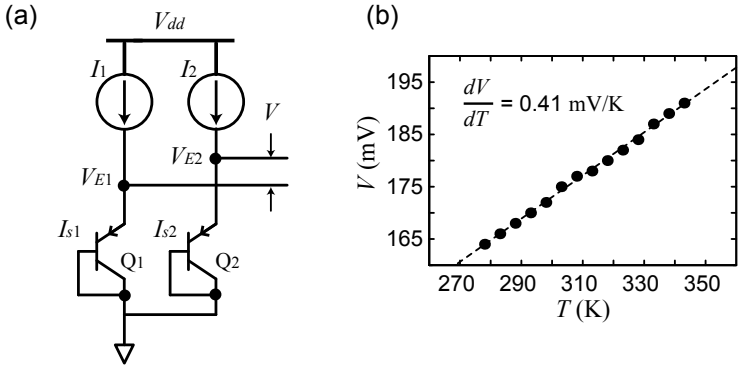


Figure 5.17 Circuit schematic and measured characteristics of the on-chip temperature sensor. (a) The sensor uses two PNP transistors in a diode connection. The emitter voltages (V_{E1} and V_{E2}) of the transistors (Q_1 and Q_2) are directly proportional to absolute temperature. By measuring the difference V between V_{E1} and V_{E2}, errors can be minimized. (b) Measured output V from the temperature sensor as a function of temperature T shows a good linear relationship with dV/dT = 1 mV/K. Reprinted from [27] by permission of The Royal Society of Chemistry.

$$V_{E1} = \frac{k_B T}{q} \log \frac{I_{s1}}{I_1}, \tag{5.24}$$

where I_{S1} is the current constant of Q_1 and q is the electron charge. Placing another transistor Q_2 and by measuring the voltage difference,

$$V = V_{E2} - V_{E1} = \frac{k_B T}{q} \log \frac{I_{s2} \cdot I_1}{I_{s1} \cdot I_2}, \tag{5.25}$$

errors caused by variation of current sources can be reduced, because the term in the *log* operation is determined by the dimensional ratio of Q_1 and Q_2. The output voltage V is more predictable, less affected by the process variations. In the actual chip design, $I_2 = 0.25I_1$ and $I_{S2} = 25I_{S1}$, which gives dV/dT = 0.4 mV/K. Figure 5.17(b) shows the measured characteristics of the temperature sensor. The measured data (black dots) agree well with the calculation result (dotted line), which expects linear dependency between V and T.

5.4 COMPLETE CELL MANIPULATION SYSTEM

In this section, we describe two very important processes to complete the CMOS cell manipulator: fabrication of microfluidic channels and packaging. The microfluidic channels are used to transport biological cells to actuator region, the microcoil array, as well as to maintain biocompatible environments. The packaging provides electrical connections to the CMOS chip and facilitates thermal cooling of the device to keep the temperature of the system in the biocompatible range.

5.4.1 Fabrication of Microfluidic Channels

Microfluidic systems in the current hybrid system are fabricated by spin-coating and patterning polymer resins on top of the CMOS IC chip (See Chapter 4 for general information). This method ensures good adhesion between the CMOS chip and the fluidic structure. Figure 5.18 shows the fabrication process of the microfluidic channel.

1. The CMOS IC die (2 mm × 3 mm) is glued onto an in-house fabricated silicon substrate that contains lithographically-patterned electrical

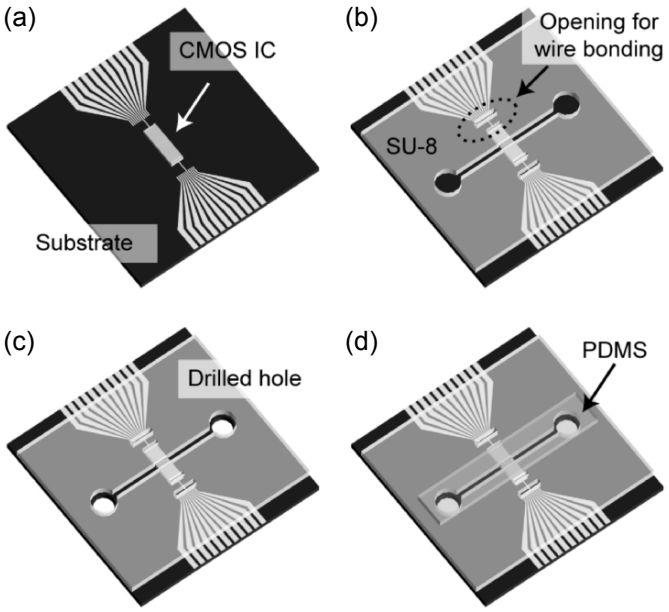


Figure 5.18 Fabrication procedure of microfluidic channels. (a) The CMOS chip is glued to a Si/SiO_2 substrate that contains lithographically patterned electrical leads. (b) SU-8 is spin-coated and patterned to form microfluidic channel sidewalls and to open up spaces for wire bonding. (c) Holes are drilled on the substrate using a diamond drill bit. (d) A PDMS layer, separately prepared by cast-curing the polymer, is placed on top of the microfluidic channel sidewalls. Reprinted with permission from [25]. Copyright © [2006] IEEE.

leads [Fig. 5.18(a)]. The substrate facilitates electrical interconnections and easy handling. It also serves as a platform on which a microfluidic system is fabricated.

2. A negative photoresist (SU-8, Microchem, MA) layer with thickness ~350 μm is spin-coated on the substrate. The SU-8 layer is then lithographically patterned to form the sidewalls of a microfluidic channel and to open up pad areas for wire bonding [Fig. 5.18(b)].
3. After patterning the sidewall structure, two holes are drilled on the substrate to form fluidic ports [Fig. 5.18(c)].
4. A PDMS (polydimethylsiloxane) layer of ~500 μm thickness is prepared by cast-coating and cure [13]. The cured PDMS layer is cut into a desired shape and placed on top of the sidewall structure to seal the microfluidic channel [Fig. 5.18(d)]. The PDMS layer forms a conformal sealing.

5.4.2 Packaging

Once the microfluidic channel is fabricated on top of the CMOS IC chip, the hybrid system is packaged for operation. The micrograph of the complete cell manipulation system is shown in Fig. 5.19(a). The size of the in-house fabricated silicon substrate is 2.5 cm × 2.5 cm. Gold wires are bonded between the CMOS chip and the substrate through the opening in the SU-8 layer. The hybrid system is electrically connected to a printed circuit board (PCB). Figure 5.19(b) shows the detailed schematic of a packaged CMOS cell manipulation system. The fluidic connections are to be made from the bottom of the hybrid system, which frees the top of the system from fluidic tubing components. This connection scheme allows the hybrid system to have a low profile, making it easier to mount the hybrid system below a microscope for observation. External electrical connections are made through the D-subminiature connector on the PCB. The hybrid system is mounted to a copper block for temperature cooling, which will be discussed in detail in Subsection 5.5.2.

5.5 EXPERIMENT SETUP

Besides a microscope to monitor the manipulation processes, the experiment setup to operate the cell manipulation system needs three major parts: a thermal cooling stage for temperature regulation, an electronic system to

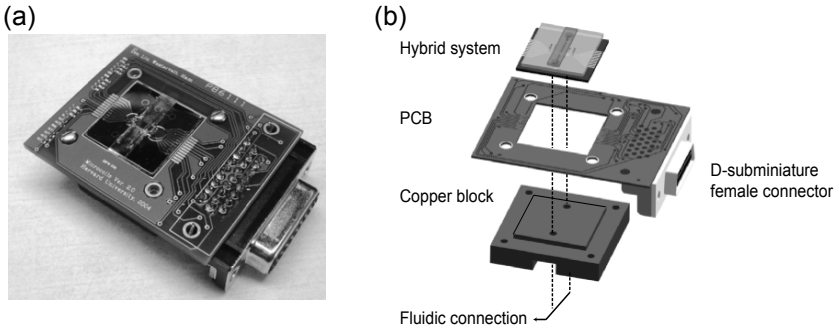


Figure 5.19 Complete cell manipulation system (a) The assembled cell manipulation system is packaged using a copper block and a PCB. (b) Schematic of the packaging. Fluidic connections are made through the backside of the hybrid system, freeing the front side of the system from fluidic tubing components. The copper block is used for cooling. Reprinted from [27] by permission of The Royal Society of Chemistry.

control the CMOS chip, and software. The overall experiment platform is built on a probe station where a bright field microscope with a charge-coupled device (CCD) camera is installed. A cooling stage is assembled on the probe station chuck which can translate and rotate laterally. The packaged cell manipulation system is then mounted to the cooling stage.

5.5.1 Temperature Control System

To regulate the temperature of the hybrid system, a cooling stage with a thermal electric cooler (TEC) is assembled as schematically shown in Fig. 5.20(a). A copper block with a temperature sensor (AD 590, Analog Device Inc., MA) is placed on the cooling side of a TEC (UltraTEC, Melcor, NJ). To cool down the TEC during its operation, another copper block is attached on the back side of the TEC, and cooled by circulating chilled water. The temperature information either from the on-chip sensors (Subsection 5.3.3) or the external sensor (AD 590) is fed to a TEC controller, which adjusts the TEC current to maintain constant temperature.

Figure 5.20(b) shows the photo of a microscope stage along with cooling system. The temperature of the microscope stage can be kept constant with the error ±0.1°C by the TEC controller (TEC 2000, ThorLabs Inc., NJ). A refrigeration cooler (CFT-25, NESLAB, NJ) is used to run chilled water through the copper block on the back side of the TEC.

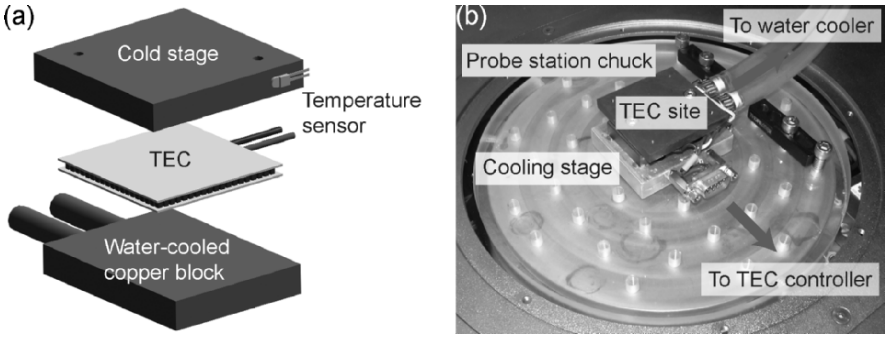


Figure 5.20 Assembly of the cooling stage. (a) A thermoelectric cooler (TEC) is sandwiched between a cold stage (copper) and a water-cooled copper block. (b) The complete cooling stage is connected to a TEC controller and a refrigeration water cooler. The cooling stage is mounted on a chuck that can translate and rotate laterally.

5.5.2 Control Electronics

The control circuit board contains a microcontroller and peripheral circuits. A microcontroller is a self-sufficient computer-on-a-chip containing a central processing unit (CPU), memory, timers, and input/output (I/O) ports. A microcontroller can be programmed to perform many different functions. A programmed microcontroller can run stand-alone, making it possible to package the whole cell-manipulation system into a hand-held device.

Figure 5.21(a) shows the block diagram of the microcontroller-based control circuit board. In this prototype, an 8-bit RISC microcontroller (Atmega32, Atmel Corporation, CA) is used. It has 32 programmable I/O ports, a 10-bit analog-to-digital converter (ADC), and a serial communication port. The microcontroller communicates with a host computer (PC) via serial communication (RS-232) to receive all necessary information to run a microcoil array: the magnitude and the direction of current in every microcoil as well as the on-time T_{on} are stored in the microcontroller. Then the

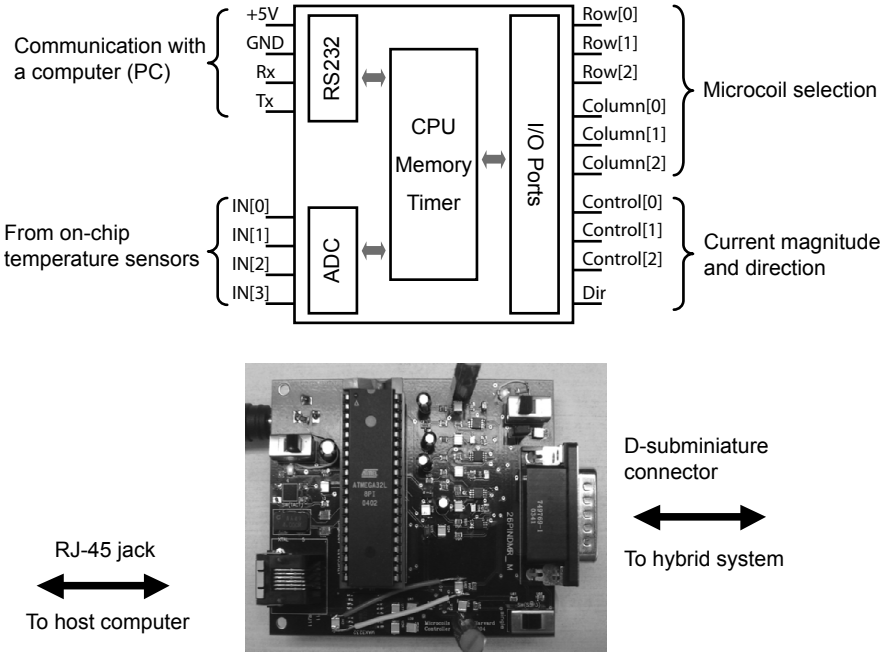


Figure 5.21 Block diagram and the photo of a microcontroller circuit. (a) The microcontroller (ATmega32) generates all external signals required to operate the CMOS chip in the hybrid system, reads out the on-chip temperature sensors, and communicates with a computer. (b) A PCB board is custom-designed for the microcontroller. The board itself functions as a stand-alone computer.

microcontroller sequentially selects each microcoil by generating proper Row[2:0] & Column[2:0] signals, and sets the current in the microcoil with Control[2:0] & Dir signals. The microcontroller turns on only one microcoil for T_{on}, realizing the time-sharing of current. Upon the request from the host computer, the microcontroller measures the analog outputs of the on-chip temperature sensors through IN[3:0] channels of the ADC. The microcontroller then sends back converted digital outputs to the host computer via RS-232 communication.

The photo of the microcontroller circuit board is shown in Fig. 5.21(b). A PCB board is custom-designed to accommodate the microcontroller and other peripheral parts. The board communicates with the host PC through the registered jack-type 45 (RJ-45) connector, which requires only 4 wires. The board directly mates with the PCB board in the packaged hybrid system using a D-subminiature connector pair.

5.5.3 Control Software

The computer software plays an important role in the operation of the cell manipulation system. Besides providing easy access to the microcoil array, software can streamline the whole experiment procedures: users can manipulate cells in real time just pressing coil buttons on the computer screen. By combining the power of the CMOS chip and the user-friendly interface of software, the manipulation process can be highly simplified and automated.

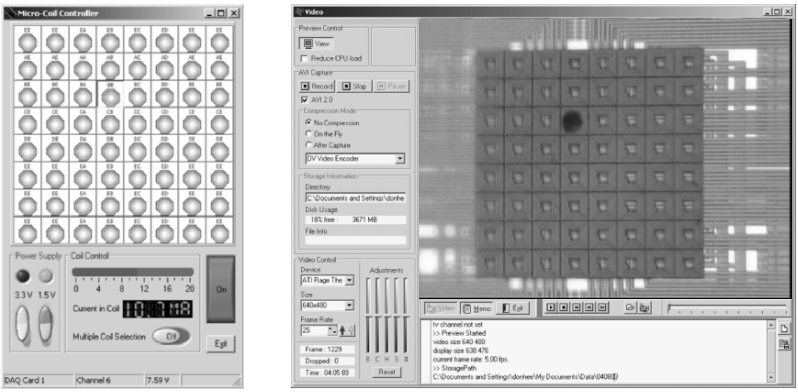


Figure 5.22 Screen captures of the programs for the microcoil array control (left) and real-time monitoring (right). The user-friendly interface facilitates the operation of the hybrid system.

Figure 5.22 shows the screen captures of the control program in a host PC. The program has an intuitive interface to control the microcoil array; the layout of buttons with a coil icon reflects the actual layout of the microcoils in the IC chip, and the images from the CCD camera in the micro scope are displayed in real time. Users can activate microcoils by just clicking corresponding buttons in the program at the same time monitoring manipulation processes.

5.6 DEMONSTRATION OF MAGNETIC CELL MANIPULATION SYSTEM

We have used two different samples, magnetic beads and bovine capillary endothelial (BCE) cells that are tagged with magnetic beads, as manipulation targets to test and demonstrate the capabilities of the CMOS cell manipulation system. A single magnetic bead is manipulated first to check the operation of the CMOS chip and to optimize the control program in the microcontroller. Multiple magnetic beads are then controlled individually using the sequential current sharing method to verify its working principle. We have also manipulated BCE cells to demonstrate cell assembly on micrometer length scales.

Before performing experiments, the surface of the CMOS chip in the hybrid system is treated bovine serum albumin (BSA) to suppress unwanted binding of the target samples. The microfluidic channel in the hybrid system is filled with BSA for an hour and then washed with phosphate buffered saline (PBS) solution. After surface treatment, target samples are introduced into the microfluidic channel, and the fluidic flow is stopped. The temperature of the hybrid system is maintained at ~10°C using the temperature control system.

5.6.1 Manipulation of Magnetic Beads

The working of the CMOS chip as well as the control electronics is tested by moving a magnetic bead across the microcoil array as shown in the sequence of images of Fig. 5.23. The magnetic bead has the diameter 8.5 μm and the magnetic susceptibility $\chi = 0.19$. By applying a continuous DC current of 20 mA to a microcoil, a magnetic field peak with $B = 30$ G is produced, trapping a single magnetic bead as shown in Fig. 5.23(a). Estimated trapping force on the bead is ~100 pN. Note that because only one magnetic bead is manipulated, the time-sharing method is not used in this demonstra-

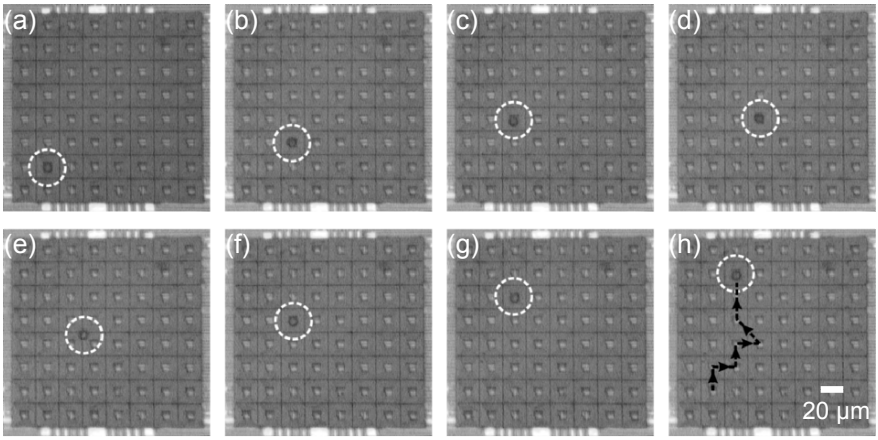


Figure 5.23 Manipulation of a single magnetic bead using the CMOS magnetic cell manipulation system. Because only one bead is manipulated, a continuous DC current is applied to a microcoil that traps the bead. The trapping magnetic field is ~30 G at the center of the microcoil, and the estimated trapping force is ~100 pN. Reprinted with permission from [25]. Copyright © [2006] IEEE.

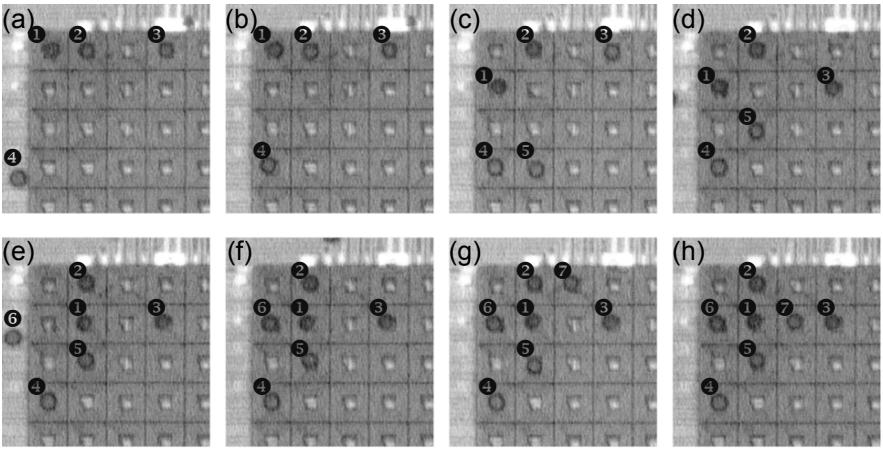


Figure 5.24 Manipulation of multiple magnetic beads with the current sharing method. Selected microcoils are activated, one by one, for $T_{on} = 10$ msec by the microcontroller circuit. (a) & (b) Four magnetic beads are trapped by producing four magnetic field peaks. (c) & (d) Additional magnetic peak is created and moved to trap another bead. (e) to (h) Total seven magnetic beads are trapped with one bead per microcoil and manipulated individually. Reprinted with permission from [25]. Copyright © [2006] IEEE.

tion. After initial trapping, the bead is moved to an adjacent microcoil by switching off the initial microcoil and switching on the adjacent microcoil. Repeating the same process, the bead is transported across the chip over a distance of 200 μm with the average speed 11 μm/s.

The sequential current sharing for low power consumption is demonstrated by manipulating multiple magnetic beads as shown in Fig. 5.24. Select microcoils in the 8×8 array are sequentially activated for T_{on} = 10 msec. During the on-time T_{on}, an electrical current of 20 mA or 0 mA is applied to a microcoil either to trap or release a magnetic bead. Initially four magnetic beads, one bead at one microcoil, are trapped [Figs. 5.24(a) and (b)]. Four microcoils are activated to generate four magnetic peaks with B = 30 G. While rearranging the initial four beads by changing the positions of the four peaks, another magnetic bead is trapped and moved [Figs. 5.24(c) and (d)] by adding and moving an additional magnetic peak. The same scheme, increasing the number of selected microcoils one by one and independently controlling the positions of the magnetic peaks, is repeated to control the motion of seven magnetic beads as shown in Figs. 5.24(e) to (h). The operation shown above is the clear verification of the current-sharing method as an effective way to operate large microcoil arrays with minimal current consumption and small number of external control signals.

5.6.2 Manipulation of Biological Cells

Experiments with biological cells are performed to demonstrate the several important capabilities of the hybrid system: 1) the control of individual cells with microscopic resolution, 2) the dynamic reconfiguration of magnetic fields patterns for versatile cell manipulation, and 3) the stable trapping and transport of multiple cells while sharing a single current source in time domain. As a representative manipulation target, bovine capillary endothelial (BCE) cells are used. To impart magnetic moment to the cells, peptide-coated magnetic beads are co-cultured with BCE cells, which leads to the uptake of the beads by the cells through endocytosis (See Subsection 5.2.3.2).

Figure 5.25 shows how a trapped cell can be positioned at locations other than the centers of microcoils, increasing the spatial resolution of cell trapping. By controlling the current distribution in two adjacent microcoils, a single magnetic peak can be moved in the steps less than the center-to-center pitch of the microcoils [Fig. 5.25(a)]. During the peak movements, the magnitude of the peak remains nearly constant, but the direction of the magnetic field is changing. Figure 5.25(b) shows the step-wise transport of a BCE trapped cell over two neighboring microcoils. To mimic the mag-

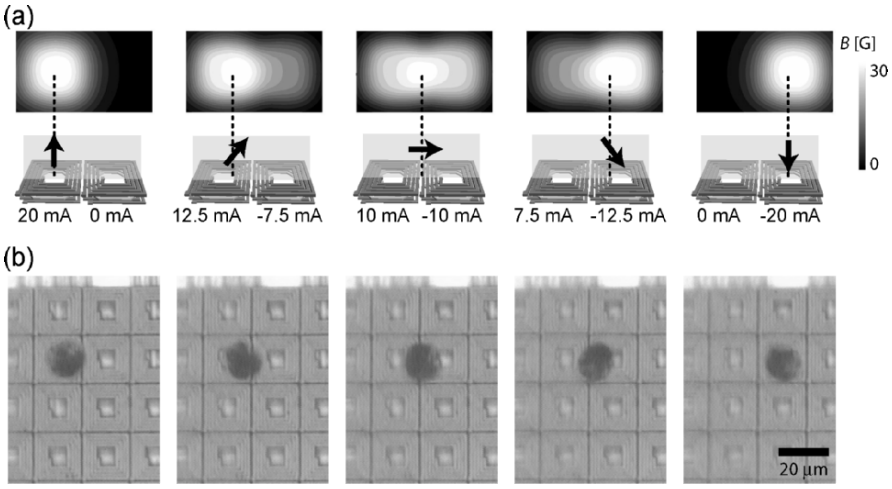


Figure 5.25 By controlling currents in the microcoils, a single peak in the magnetic field magnitude is created and moved in steps over two neighboring microcoils. (a) Calculated magnetic field patterns with current distribution in microcoils. The magnitude of the field peak remains nearly the same, but the direction of the field rotates. The arrows indicate the magnetic field direction in the plane normal to the chip surface. (b) A single BCE cell is trapped and moved using the field pattern in (a). To generate the fields, current pulses with appropriate values and directions are alternatively applied to two microcoils. The cell is observed rolling on the chip surface during the transport due the change of the field direction. Reprinted from [27] by permission of The Royal Society of Chemistry.

netic field patterns shown in Fig. 5.25(a), currents of appropriate magnitude and polarity are pulsed to each coil alternatively with the period of 20 msec and the duty cycle of 50%. The trapped cell is observed to roll on the chip surface during the transport, as the direction of the magnetic field is changing at each peak position.

One of the primary applications of the CMOS-based cell manipulation system is to assemble biological tissues on microscopic length scales. By bringing cells, one by one, with accurate spatial control, the hybrid system can build an artificial tissue in a standardized and repeatable manner. The assembled tissue can be used for many purposes, such as studying the communications between different types of cells and testing the efficiency of drugs [9].

A preliminary demonstration of the microscope tissue assembly is demonstrated in Fig. 5.26. Initially, three BCE cells are trapped by sequentially pulsing currents to three microcoils. The period, the duty cycle, and the magnitude of the current pulse are 30 msec, 33%, and 20 mA, respectively. With a single cell held still, the rest of the cells are moved independently

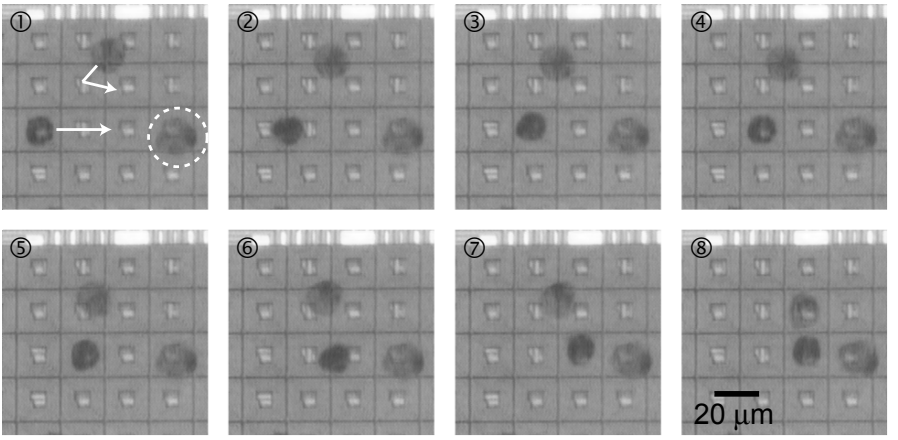


Figure 5.26 Manipulation of multiple cells. Three BCE cells are trapped and independently controlled to sit next to each other; the cell in the dotted circle is held still, and the rest of the cells are moved. The on-chip current source is sequentially shared in time domain among three microcoils; the current of 20 mA is on for 10 msec (T_{on}) and off for 20 msec (T_{off}) in each microcoil. Because the diffusion of trapped cells is negligible during the off-time of currents, multiple magnetic traps can be effectively created simultaneously. Reprinted from [27] by permission of The Royal Society of Chemistry.

by activating appropriate microcoils with the same pulse profile, and finally all cells are arranged next to each other. No cell is lost during this operation, confirming the time-sharing method as an effective way to generate multiple magnetic traps.

5.7 CONCLUSIONS AND OUTLOOK

In this chapter, a CMOS based cell manipulation system is introduced as a new type of lab-on-ICs. It combines the programmability and speed of the CMOS IC chip and the biocompatibility of the microfluidic system, providing efficient, accurate, and versatile manipulation of many individual magnetic beads and bead-bound cells. The miniaturized, self-contained hybrid system can be packaged for portable applications. Thanks to the low fabrication cost of the CMOS chips, the hybrid system can also serve as a single-used, disposable device.

There still remain challenges to further mature the proposed hybrid technology. One immediate challenge is to verify the long-term biocompatibility of the hybrid system by assembling and culturing biological cells over the CMOS chip. For this purpose, it would be necessary 1) to coat the chip

surface with extracellular matrix (ECM) proteins such that cells can adhere to and spread over the chip surface, and 2) to improve the temperature regulation to keep the system temperature at 37ºC. Another challenging yet very important task is to develop an all-electrical on-chip cell sensor (for each microcoil) that can map the location of all bead-bound cells without relying on optics. Since the magnetic bead changes the effective inductance of the microcoil due to the magnetic flux from the bead, a radio-frequency (RF) signal may be used to sense the magnetic bead. Such RF sensors will render bulky optics unnecessary, realizing a true lab-on-IC.

The CMOS-based cell manipulation system is promising technology for many new biological experiments. By bringing advanced microelectronics and microfluidics together, the hybrid system offers unique opportunities to perform cell-by-cell assays in a highly repeatable way. The work presented herein marks the onset of this new technology. With further advances in device deign and more biological experiments, the hybrid system can be an indispensable and powerful instrument for biological and biomedical research.

ACKNOWLEDGMENT

The authors thank L. DeVito and S. Feindt of Analog Devices Inc. and B. Park of Samsung Electronics for their continued support and chip fabrications. The authors also thank E. Nam of Electronics and Telecommunication Research Institute (ETRI), Korea, for support through foundry service and gift. E. Alsberg and D. Ingber of Harvard Medical School and Children's Hospital (Boston) provided the bovine capillary endothelial cells. AnSoft, Sonnet, and Agilent donated their electromagnetic field solvers.

REFERENCES

[1] R. A. Kaul, N. I. Syed, and P. Fromherz, "Neuron-semiconductor chip with chemical synapse between identified neurons," *Physical Review Letters,* vol. 92, p. 038102, 2004.

[2] P. Fromherz, "Joining ionics and electronics: semiconductor chips with ion channels, nerve cells, and brain tissue," *IEEE International Solid-State Circuits Conference Digest of Technical Papers,* vol. 1, pp. 76-77, 2005.

[3] A. Hassibi and T. H. Lee, "A programmable electrochemical biosensor array in 0.18 μm standard CMOS," *IEEE International Solid-State Circuits Conference Digest of Technical Papers,* vol. 1, pp. 564-565, 2005.

[4] B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert, M. Steinhauser, G. Eckstein, D. Schmitt-Landsiedel, and R. Thewes, "A 128 × 128 CMOS biosensor array for extracellular recording of neural activity," *IEEE Journal of Solid-State Circuits,* vol. 38, pp. 2306-2317, 2003.

[5] P. Cailat, M. Belleville, F. Clerc, C. Massit, J. H. Wuorinen, and J. H. Wuorinen, "Active CMOS biochips: an electro-addressed DNA probe," *IEEE International Solid-State Circuits Conference Digest of Technical Papers,* vol. 1, pp. 272-273, 1998.

[6] N. Manaresi, A. Romani, G. Medoro, L. Altomare, A. Leonardi, M. Tartagni, and R. Guerrieri, "A CMOS chip for individual cell manipulation and detection," *IEEE Journal of Solid-State Circuits,* vol. 38, pp. 2297-2305, 2003.

[7] C. Lee, "Control and manipulation of magnetic nanoparticles and cold atoms using micro-electromagnets," Ph.D thesis, Harvard University, 2002.

[8] H. Lee, A. M. Purdon, and R. M. Westervelt, "Manipulation of biological cells using a microelectromagnet matrix," *Applied Physics Letters,* vol. 85, pp. 1063-1065, 2004.

[9] U. Häfeli, *Scientific and clinical applications of magnetic carriers*. New York: Plenum Press, 1997.

[10] J. Bauer, *Cell electrophoresis*. Boca Raton: CRC Press, 1994.

[11] H. A. Pohl, *Dielectrophoresis: the behavior of neutral matter in nonuniform electric fields*. Cambridge; New York: Cambridge University Press, 1978.

[12] A. Radbruch, B. Mechtold, A. Thiel, S. Miltenyi, and E. Pfluger, "High-gradient magnetic cell sorting," *Methods Cell Biol,* vol. 42 Pt B, pp. 387-403, 1994.

[13] J. Ugelstad, P. Stenstad, L. Kilaas, W. S. Prestvik, R. Herje, A. Berge, and E. Hornes, "Monodisperse magnetic polymer particles. New biochemical and biomedical applications," *Blood Purif,* vol. 11, pp. 349-369, 1993.

[14] G. T. Rado and H. Suhl, *Magnetism* vol. 3. New York: Academic Press, 1963.

[15] W. F. Brown, "Thermal Fluctuations of a Single-Domain Particle," *Physical Review,* vol. 130, pp. 1677-1686, 1963.

[16] C. S. Lee, H. Lee, and R. M. Westervelt, "Microelectromagnets for the control of magnetic nanoparticles," *Applied Physics Letters,* vol. 79, pp. 3308-3310, 2001.

[17] J. Happel and H. Brenner, *Low Reynolds number hydrodynamics, with special applications to particulate media*. Englewood Cliffs, N.J.: Prentice-Hall, 1965.

[18] M. D. Bale, S. J. Danielson, J. L. Daiss, K. E. Goppert, and R. C. Sutton, "Influence of Copolymer Composition on Protein Adsorption and Structural Rearrangements at the Polymer Surface," *Journal of Colloid and Interface Science,* vol. 132, pp. 176-187, 1989.

[19] L. A. Cantarero, J. E. Butler, and J. W. Osborne, "The adsorptive characteristics of proteins for polystyrene and their significance in solid-phase imunoassays," *Anal Biochem,* vol. 105, pp. 375-382, 1980.
[20] C. Deen, E. Claassen, K. Gerritse, N. D. Zegers, and W. J. Boersma, "A novel carbodiimide coupling method for synthetic peptides. Enhanced anti-peptide antibody responses," *J Immunol Methods,* vol. 129, pp. 119-125, 1990.
[21] H. Heitzmann and F. M. Richards, "Use of the avidin-biotin complex for specific staining of biological membranes in electron microscopy," *Proc Natl Acad Sci U S A,* vol. 71, pp. 3537-3541, 1974.
[22] R. P. Haugland and W. W. You, "Coupling of antibodies with biotin," *Methods Mol Biol,* vol. 80, pp. 173-183, 1998.
[23] T. Sano and C. R. Cantor, "Intersubunit contacts made by tryptophan 120 with biotin are essential for both strong biotin binding and biotin-induced tighter subunit association of streptavidin," *Proc Natl Acad Sci U S A,* vol. 92, pp. 3180-3184, 1995.
[24] H. Lee, Y. Liu, E. Alsberg, D. E. Ingber, and R. M. Westervelt, "An IC/Microfluidic hybrid microsystem for 2D magnetic manipulation of individual biological cells," *IEEE International Solid-State Circuits Conference Digest of Technical Papers,* pp. 80-81, 2005.
[25] H. Lee, Y. Liu, R. M. Westervelt, and D. Ham, "IC/microfluidic hybrid system for magnetic manipulation of biological cells," *IEEE Journal of Solid-State Circuits,* vol. 41, pp. 1471-1480, 2006.
[26] H. Lee, A. M. Purdon, V. Chu, and R. M. Westervelt, "Controlled assembly of magnetic nanoparticles from magnetotactic bacteria using microelectromagnets arrays," *Nano Letters,* vol. 4, pp. 995-998, 2004.
[27] H. Lee, Y. Liu, D. Ham, and R. M. Westervelt, "Integrated cell manipulation system - CMOS/Microfluidic hybrid," Lab on a Chip, vol. 7, pp. 331-337, 2007.

AUTHOR BIOGRAPHY

Yong Liu received the B.S.E.E. and M.S.E.E. degrees from Tsinghua University, Beijing, China, in 2000 and 2003, respectively. He is currently working toward the Ph.D. degree in electrical engineering at Harvard University, Cambridge, MA. In 2001, he was with Tsinghua Tongfang Microelectronics Company, Beijing, working on the second-generation Chinese RF ID card. During the summers of 2005 and 2006, he was with the Mixed-signal Communications IC Design Group at the IBM T. J. Watson Research Center, Yorktown Heights, NY, where he designed a high-performance autonomic PLLs and on-chip jitter monitor for high-speed data links. His current research interests include biological laboratories on an integrated circuit and RF and mixed-signal integrated circuits for communications. Mr. Liu received the Seagate Scholarship for Extraordinary Freshman and the Motorola Scholarship at Tsinghua University in 1996 and 2001, respectively. He was the

recipient of the 2nd Prize in the National Graduate EDA Competition, China, in 2002, and the winner of the Analog Devices Outstanding Student Designer Award in 2004.

Hakho Lee received the B.S. degree in physics from Seoul National University, Seoul, Korea, in 1998, and the Ph.D. degree in physics from Harvard University, Cambridge, MA, in 2005. For the Ph.D., he developed a new type of micro-total-analysis-systems (μTAS) for biological experiments based on integrated circuits and microfluidics. In 2005, he joined the Center for Molecular Imaging Research (CMIR) at the Massachusetts General Hospital, Harvard Medical School as a Research Associate. His current research interests include microelectronic actuators for biological cell manipulation, and magnetic/optical sensors for medical imaging.

Robert M. Westervelt received the Ph.D. degree in physics from the University of California, Berkeley, in 1977. Following a postdoctoral appointment at Berkeley, he moved to Harvard University, Cambridge, MA, where he is currently a Mallinckrodt Professor of Applied Physics and Physics. His group investigates the quantum behavior of electrons inside nanoscale semiconductor structures and develops tools for the manipulation of biological systems. Current research topics include: imaging electron flow through nanoscale devices at low temperatures using scanning probe microscopy, making artificial molecules composed of few-electron quantum dots for quantum information processing, and the development of hybrid integrated circuit/microfluidic chips to trap, move, assemble, and sort biological cells and small particles in fluids. He is Director of the NSF-funded Nanoscale Science and Engineering Center at Harvard University, which includes participants at the Massachusetts Institute of Technology, the University of California, Santa Barbara and the Museum of Science, Boston. Previously Westervelt was Director of the Materials Research Science and Engineering Center, and Co-Director of the Joint Services Electronics Program at Harvard.

Donhee Ham is an Associate Professor of electrical engineering, School of Engineering and Applied Sciences, Harvard University. He received the B.S. degree in physics in 1996 from Seoul National University, Korea, graduating with Presidential Honor atop the Natural Science College, and the PhD degree in electrical engineering from California Institute of Technology (Caltech) in 2002, winning the Charles Wilts Doctoral Thesis Prize for outstanding PhD research in electrical engineering. He was the recipient of the Caltech Li Ming Scholarship and IBM Graduate Research Fellowship. He was also the recipient of the 2003 IBM Faculty Partnership Award. He shared Harvard's Hoopes prize (best senior thesis award) in 2003 with Mr. William Andress. His work experiences also include Laser Interferometer Gravitational Wave Observatory (LIGO), Pasadena, CA, 1997/98, IBM T. J. Watson Research Center, NY, 2000, IEEE conference technical program committees including International Solid-State Circuits Conference, and industry/government technical advisory positions on subjects including high-speed electronics and future nanoelectronics technologies in the post-50nm era. His current

research focus is on (1) RF & microwave ICs, (2) nanoscale quantum-effect devices for GHz & THz circuits, and (3) soliton & nonlinear wave electronics. His research also examines (4) biological laboratories on an IC. Donhee Ham's research details can be found at http://www.deas.harvard.edu/~donhee.

6 APPLICATIONS OF DIELECTROPHORESIS-BASED LAB-ON-A-CHIP DEVICES IN PHARMACEUTICAL SCIENCES AND BIOMEDICINE

Claudio Nastruzzi[1], Azzurra Tosi[1], Monica Borgatti[2], Roberto Guerrieri[3]*, Gianni Medoro[4] and Roberto Gambari[2]

[1] Department of Chemistry and Technology of Drugs, University of Perugia, Italy
[2] ER-GenTech, Biotechnology Center, Department of Biochemistry and Molecular Biology, University of Ferrara, Italy
[3] Center of Excellence on Electronic Systems (ARCES), University of Bologna, Italy
[4] Silicon Biosystems, S.p.A., Italy
*roberto.guerrieri@unibo.it

6.1 GENERAL INTRODUCTION

Severe and chronic diseases such as coronary artery disease, cancer and respiratory disorders afflict millions of people each year and are becoming more prevalent as the population ages. Despite continuing advances in medical and pharmaceutical technologies, current treatment options for many of these diseases are inadequate. Drug treatments for these diseases, where they exist, often offer only limited therapeutic benefits and cause adverse side effects.

In the last ten years, a large worldwide market has emerged based on technologies that improve the delivery of established drugs in novel, cost-

Abbreviations: DEP, dielectrophoresis; PAGE, polyacrylamide gel electrophoresis; EF, electric field; CMOS, complementary metal oxide semiconductor.

effective ways that in some cases provide significant benefits such as improved safety and efficacy, increased patient compliance, greater ease of use and expanded indications.

The development of advanced analytical and bioseparation methodologies based on micro- or macro-arrays and biosensors is one of the strategic objectives of the so-called postgenomics, but it also has an impact on strategic application fields, such as predictive oncology, diagnostics in the biomedical field, and drug research [1-3]. A comprehensive lists of recently published articles describing works employing commercial biosensors and emerging applications in areas of drug discovery, clinical support, food and environment monitoring, and cell membrane biology were reported by Rich and Myszka [4].

Many applications in the biological, pharmaceutical and medical field are characterized by complex experimental protocols which need both microorganism detection and manipulation. So-called "Lab-on-a-chip" (LOAC) can integrate functions such as sensing, processing, and actuation on the same platform. As discussed in this chapter, this approach has already allowed the execution of experiments that are at least impractical or, in several cases, impossible, if standard laboratory tools were used. However, the most important contribution of this technology is that it will enable highly complex laboratory testing to move from the central laboratory into non-laboratory settings [5-7].

An example of this trend to the integration of different functionalities is given by a new LOAC, recently described [8]. Thanks to the use of the integrated circuits technology, it can generate dielectrophoretic fields [9, 10] that immobilize and allow one to control single biological objects, such as cells, liposomes, or microspheres immersed in a liquid overhanging and in contact with the same chip. Since many biotechnological experiments can be recast as manipulation of microparticles, such as cell separation and detection, PCR techniques, DNA quantification and many others [11], the potential of this technology looms significant.

The need to integrate different functionalities is apparent when commercial LOAC are considered. Almost all of them [12, 13] available today are designed to replace gel electrophoresis or to study biomolecular interactions and complexes. This is an extremely valuable and flexible technique but results are very operator dependent and quantization is poor since sample preparation is left to the skill of the operator. LOACs that automate the entire analytical process including sample injection, microparticle separation and detection, thereby improve the current situation in terms of speed, reproducibility and quantitative results, essential for an application to real world

problems. In the following paragraphs, examples of applications that will greatly benefit from the availability of integrated platforms are reviewed.

6.1.1 Gene Expression Studies

A growing area in the study of diseases is the use of DNA microarrays, which work by providing a matrix of thousands of probes to which a target molecule is exposed. Expressed genes hybridise to the appropriate targets on the array [11]. The target, cDNA or cRNA, is derived from mRNA, which in turn is isolated from total RNA. Both total RNA and mRNA are very susceptible to degradation by RNase contamination. Since array experiments are very expensive, it is important that the quality of total and messenger RNA is checked prior to hybridisation. With traditional gel electrophoresis, this is a slow process and requires a relatively large amount of the precious RNA samples [11]. This problem is overcome by using LOAC able to integrate all steps of sample preparation, including cell selection, lysis and extraction of the DNA sample. This technology is fast, provides both qualitative and quantitative data and requires very small sample amounts [12].

6.1.2 Protein Studies

Proteins are much more difficult to handle than nucleic acids; they easily adsorb, degrade or aggregate, and present substantially greater analytical challenges. The current standard technique for the separation of proteins is SDS-PAGE which has the disadvantages of being very slow, messy, using toxic chemicals and being relatively expensive. The analysis of proteins has been implemented in LOAC and is fast, easy to use and less costly than sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE) [11-13]. In a typical SDS-PAGE electrophoresis large amounts of proteins are electrophoresed, blotted, and bound to specific monoclonal antibodies. The length of time and the cost of SDS-PAGE are much higher in respect to LOAC. In addition LOAC is far more quantitative than SDS-PAGE [12-13].

With respect to methods for identifying biomolecular interactions and complexes, two predominant methods are currently under investigation: protein microarrays and microfluidics LOACs. These technologies allow scientists to study the specificity of binding between two biomolecules, the concentration of a biomolecule, the rate of association/dissociation and the affinity for binding [11-13].

6.1.3 Quality Assurance and Quality Control (QA/QC) in Pharmaceutical Sciences

New technologies are typically first applied to solving novel or intractable problems in research and, especially in the highly regulated pharmaceutical industry. In the case of LOACs, the interest for QA/QC testing of protein-based drugs has been instantaneous. The reason for this is that there are an increasing number of protein-based drugs which require quality control methods. The existing technique, SDS-PAGE, is slow, difficult to automate, uses toxic chemicals and, most importantly, has very poor day-to-day and inter-lab reproducibility and is therefore inherently very difficult to validate. Moreover, four types of factors. (i.e., technical, instrumental, computational and interpretative) affect the outcome of microarray studies, and a major concern about microarray studies has been the lack of reproducibility and accuracy [14]. LOAC systems could resolve these problems and are a highly attractive alternative [11-14].

6.2 DIELECTROPHORESIS-BASED APPROACHES

The need to integrate different functionalities has prompted researchers to look for a technology that provides flexibility and is suitable for the steps needed to provide a complete sample preparation.

From this standpoint, several DEP-based devices have been proposed as very appealing candidates for the manipulation of objects of interest in biotechnology, biomedicine and for pharmaceutical analysis [15-17]. Dielectrophoresis is the physical phenomenon whereby polarizable particles, suspended in a liquid medium and subjected to a spatially non-uniform electric field, experience movement [20-22]. Thanks to the technological miniaturization trend, it is possible to develop new micro-fabricated electrode structures [18, 19] able to manipulate sub-micrometer particles by means of a non-uniform AC electric field [20-22]. The beauty of these approaches is that they do not hamper the viability of living organisms and allow their recovery after having performed complex operations. To achieve these goals, precise and at the same time flexible manipulation approaches are required. Existing CMOS technology already allows us to integrate manipulation of living cells under software control, offering a powerful technology for the development of new diagnostic and therapeutic protocols.

In particular, it is possible to show that under suitable conditions [20] and applying an AC electric field, the time-averaged DEP force exerted by the electric field on a small particle is given by Eq. (6.1) [20-22].

$$\left\langle \vec{F}_{DEP} \right\rangle = \pi \varepsilon_m r^3 \mathrm{Re}\left[f_{CM} \right] \nabla \left| \vec{E}_0 \right|^2 , \tag{6.1}$$

where ε_m is the particle permittivity, r is the particle radius, f_{CM} is the Clausius-Mossotti factor and E_0 is the applied AC electric field. The Clausius-Mossotti factor is given by Eq. (6.2) [20, 23].

$$f_{CM} = \frac{\varepsilon_p^* - \varepsilon_m^*}{\varepsilon_p^* + 2\varepsilon_m^*}, \tag{6.2}$$

where ε_p^* and ε_m^* are the complex permittivities of the particle and the medium, respectively, that are ruled by Eq. (6.3).

$$\varepsilon^* = \varepsilon - i\frac{\sigma}{2\pi f}, \tag{6.3}$$

where ε is the permittivity, σ is the conductivity and f is the frequency. Considering the effect of viscous friction on dragged particles, the dielectrophoretic mobility can be defined as in Eq. (6.4) [24].

$$\mu_{DEP} = \left(\varepsilon_m \mathrm{Re}\left[f_{CM} \right] r^2 \right) / \left(6\eta \right), \tag{6.4}$$

where η is the medium viscosity. When particles move at constant speed, the time required to cover the small distance Δd is given by Eq. (6.5),

$$\tau_{DEP} = \Delta d / \left(\mu_{DEP} \nabla E_0^2 \right). \tag{6.5}$$

This equation allows us to estimate the impact that improved CMOS technology has on the implementation of LOACs. The most obvious impact of the technology scaling provided by the existing CMOS roadmaps is that it will be possible to reduce the size of the structures used to generate the electric fields in the LOAC. A reasonable assumption is that future devices will scale keeping the electric field amplitude constant. This is needed to avoid excessive exposure of living microorganisms to electric fields. Hence, assuming that the following scaling rules apply,

$$\begin{cases} E_0 = constant \\ \Delta d' = \Delta d / K \\ s' = s / K \end{cases}, \tag{6.6}$$

where s is a generic device dimension and K is the scaling factor. τ_{DEP} scales downs as in Eq. (6.7),

$$\tau'_{DEP} = K^{-2}\tau_{DEP}. \tag{6.7}$$

A typical example of application is a 0.35 μm CMOS device [19], where the value of s is 20 μm and the amplitude of the sinusoidal signal V_0, applied between the electrodes, is 3.3 V. Since the radius of a eukaryotic cell is about 7-10 μm, τ_{DEP} is about 1 sec (Figure 6.1 shows an example of individual lymphoblastoid cell line (LCL) trapped by means of dielectrophoresis), while for a virus it is about 2 hours; single viruses are therefore not manageable with this CMOS technology. Scaling the technology down to 90 nm, this time scales down to about 7 minutes, making it theoretically possible to control individual viral particles.

However, the combined effects of heat generation and the double layer of charge, that occur when an electric field is applied to an electrolytic solution, induce alteration in E_0. The consequence is a modification of the DEP forces. Other parameters that should be carefully taken in consideration are the Brownian motions and fluid flows, including electrothermal flow (ETF) and AC electroosmosis.

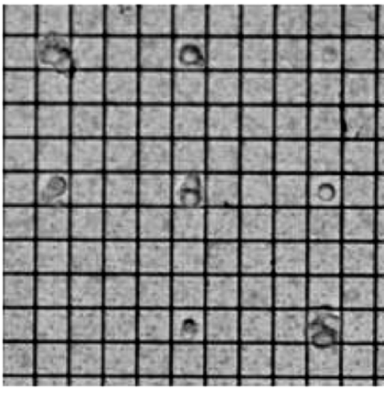


Figure 6.1 Microphotograph showing the trapping of lymphoblastoid cell line (LCL) in dielectrophoresis generated cages (DEP-cages) in a Lab-on-a-chip device.

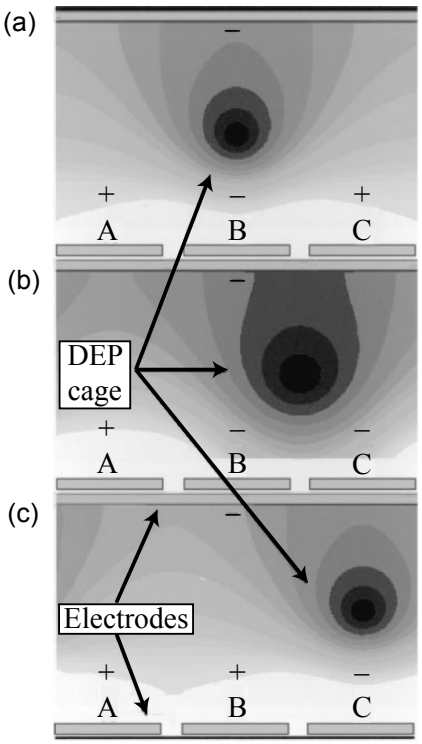


Figure 6.2 Computer generated analysis of dielectrophoresis generated cages (DEP-cages) in a Lab-on-a-chip platform. The dark areas represent local minima of electric fields associated with the presence of DEP-cages. In (a), (b) and (c) are depicted the changes in cages position as a consequence of variations of the potentials applied to the single electrodes.

Figure 6.2 shows a scheme depicting an example of generation of DEP fields in a LOAC. By applying suitable potentials to the electrodes, it is possible to generate time-dependent electric fields in the liquid. These fields can then generate DEP fields acting on the particles in the fluid. A DEP force is then generated thanks to the differences in the dielectric permittivities of the different materials. An important point of this approach is that the overall system can be designed to force the DEP fields to create closed cages that can trap particles in a stable way. By looking at Fig. 6.2, it is possible to see a local minimum of electric fields associated with the presence of the DEP-cage. Since these electric potentials can be applied under software control, it is possible to change in time the location of these closed DEP cages by changing the potentials applied to the electrodes. If proper timings are chosen, the moving cages displace also the particles contained by them.

6.3 DIELECTROPHORESIS BASED LAB-ON-A-CHIP PLATFORMS

The general set-up of typical dielectrophoresis-based LOACs is shown schematically in Fig. 6.3. The assembly is generally constituted by a direct optical microscope [Fig. 6.3(a)], eventually connected with camera and computer, and a platforms suitable for the management of the LOAC. In Figs. 6.3(b) and (c) are shown two different examples of LOACs, respectively described in [19, 25-27]; Fig. 6.3(b) shows a LOAC whose electrodes are structured as a one dimensional array, while Fig. 6.3(c) represents a device with a two dimensional array of electrodes that are controlled individually by CMOS transistors. The two different choices are appropriate for different types of experiments and have different manufacturing complexity.

In general, several DEP-based LOACs were recently described and found to be suitable for biotechnological applications in the field of isolation of single cells from a population as well as manipulation of single biological objects, including cells and microspheres. The following review, far from exhaustive, demonstrates that in this field the number of architectural choices proposed by different groups is still very large. Hence we expect that in the next few years we will experience a consolidation with a few platforms emerging as the most suitable candidates in their application domain, in a way similar to what we have today in the field of DNA array where optical devices are the most widespread platform.

6.3.1 Lab-on-a-chip with Spiral Electrodes

A LOAC with spiral electrodes has been reported by Wang *et al.* [28] as schematically represented in Fig. 6.4(a). By using concentric spiral electrodes, it was demonstrated that MDA-MB231 breast cancer cells can be concentrated at the center of the chip. This approach is very attractive for concentration and isolation of highly purified cell populations from samples constituted by low numbers of cells. In addition, for some biomedical applications, concentration of cells from dilute suspension to a target location is necessary. Using LOAC with spiral electrodes, it is possible to concentrate bacteria from urine and discriminate between Gram-positive and Gram-negative bacteria.

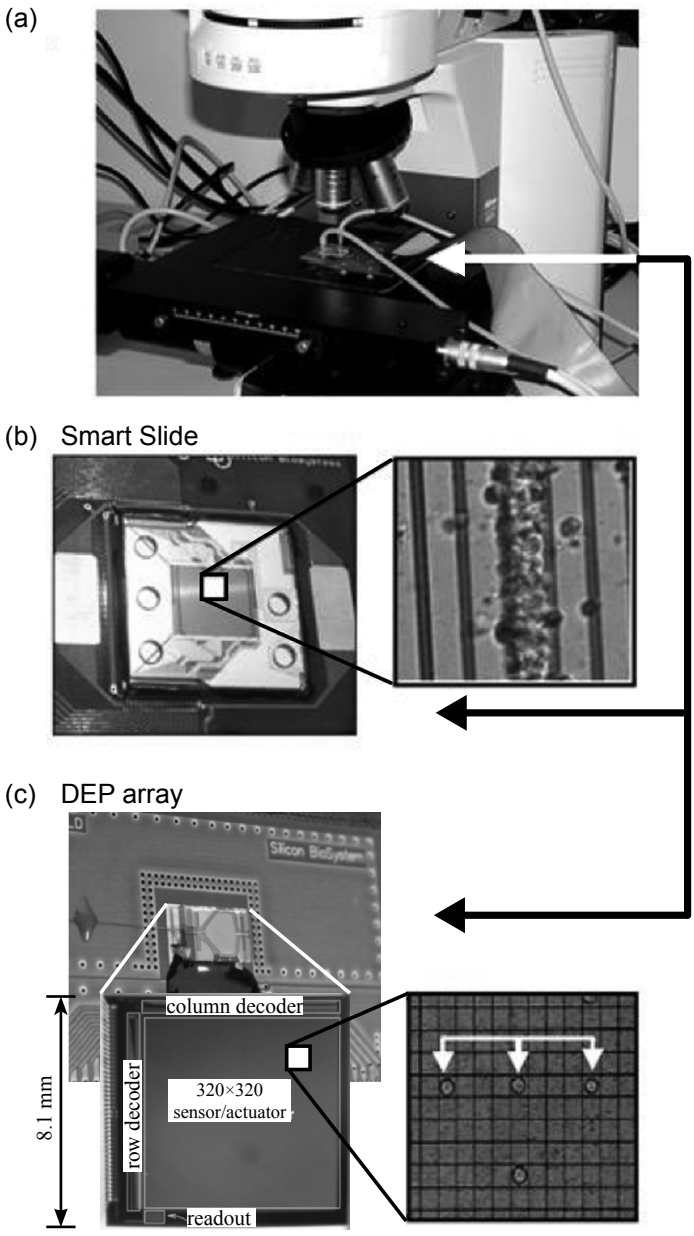


Figure 6.3 (a) General set-up of a LOAC assembly. The LOACs are located under an optical microscope and connected to a mother-board. (b) DEP-based LOACs with parallel 1D electrode. The inset shows the separation of cell populations. (c) DEP-based LOACs with 2D electrode array. The inset shows the manipulation of individual cells.

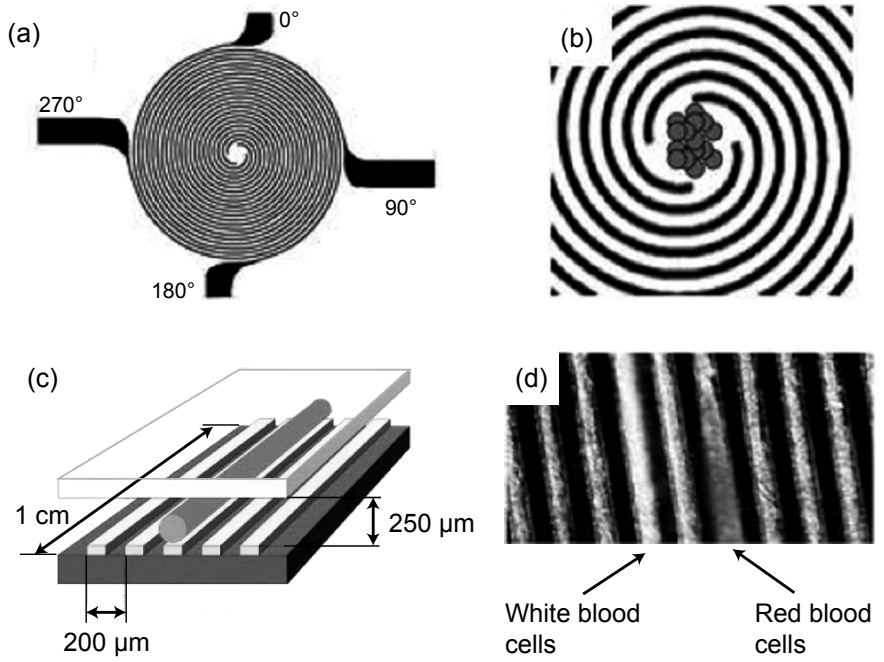


Figure 6.4 Schematic representation of different DEP-based devices. (a) and (b) Spiral electrodes chip for DEP-based cell concentration lab-on-a-chip with parallel electrodes. (c) and (d) Separation of red blood cells from human erythroleukemic cells (K562 cells).

6.3.2 Lab-on-a-chip with Parallel Electrodes

An example of LOAC based on parallel electrodes [29] is depicted in Fig. 6.4(c). This type of device is composed of a microchamber, delimited on the top by a conductive and transparent lid (which is itself an electrode and is electrically connected to the device by means of a conductive glue) and on the bottom by a support. A spacer determines the chamber height, which can vary from 20 to about 100 µm, while a silicon elastomer gasket delimits and seals the microchamber on the sides. A mother-board is used to generate and distribute to each electrode in the device the proper voltage phases needed to create and move the DEP cages and to perform the sensing operations, while a software tool allows to control the actuation and sensing operations. By changing the programming of the electrodes, each DEP cage can be independently moved from electrode to electrode along the whole microchamber, dragging with it the trapped elements. This device allows the separation and/or concentration of cells and/or microparticles in a fully-electronic system without the need for fluid flow control, which is usually more complex and expensive in an integrated system perspective.

A first device with 39 parallel electrodes which can generate from 0 to 19 cylinder-shaped DEP cages [29] has been applied to separate white blood cells from red blood cells [Fig. 6.4(d)]. The same system was recently employed to demonstrate that these cells are suitable for genomic studies. This is one of the several studies available in literature that show that DEP manipulation can be performed without damaging the vitality of the cells.

6.3.3 Lab-on-a-chip with Two-dimensional Electrode Array

Cells belonging to the human monocytic U937, glioma HTB, neuroblastoma SH-SY5Y, T-lymphoid Ind-2 cell lines have been moved throughout Lab-on-a-chips carrying 5×5 electrode arrays [30]. In this case, the cells are moved as clusters of 5-30 cells and cell separation and isolation performed. Interestingly, the authors of this paper demonstrated the efficient separation of cells by combining DEP with fluid flow. In particular, efficient separation of HTB and SH-SY5Y cells was obtained. Gene expression profiling studies demonstrated minimal stress response of DEP-separated cells as indicated measuring activity of stress-responsive genes (such as c-fos) [30]. A second example is a Lab-on-a-chip (the DEP-array) constituted of high-density arrayed electrodes and shown in Fig. 6.3(c). In this case, levitation and movement of thousands of single objects is obtained without the need of fluid flow. Moreover, this platform allows easy and reproducible handling of single cells.

6.4 APPLICATIONS OF LAB-ON-A-CHIP TO PHARMACEUTICAL SCIENCES

6.4.1 Microparticles for Lab-on-a-chip Applications

As previously discussed, DEP-based LOACs provide a flexible platform to displace microparticles, possibly loaded by suitable compounds, and use them as vectors to deliver precise amounts of these compounds to the relevant targets.

Even though the current literature contains a few examples of microparticle applications on LOACs, a comprehensive analysis of the relationship between size, chemical composition, polarity, viscosity and funtionalization of microparticles is not yet available. This missing information will

however be a major issue if microparticles have to be used as reliable carriers of useful compounds and we expect substantial advances in this area in the coming years.

6.4.1.1 Recent Advances

Our groups have recently published a paper describing the application of two dielecrophoretic-based microdevices to levitate and move microspheres and cells in order to obtain a forced interaction between microspheres and target cells [32]. A recent contribution reports that the Brownian motion of fluorescent microspheres was analyzed to validate a total internal reflection-based biochip [33]. Finally, the development of a novel, spectroscopic technique for high-sensitivity, label-free DNA quantification is reported in a work demonstrating that an optical resonance (whispering gallery mode) excited in a micron-sized silica sphere can be used to detect and measure nucleic acids. This all-photonic whispering gallery mode biosensor can be integrated on a semiconductor chip that makes it an easy to manufacture analytic component for a portable, robust lab-on-a-chip device [34]. These studies indicate the usefulness of microparticles for specific biotechnological applications of LOAC.

6.4.1.2 Commercial Microparticles

In this paragraph we describe different types of microparticles that are suitable for LOAC applications. Microparticles can be generally divided in two sub categories namely those available commercially and those custom-made in a laboratory. This second type of microparticles is usually tailored to very specific applications and therefore has a narrower applicability.

Many different companies are presently offering a large choice of different types of microparticles for a number of application protocols including the possibility to be loaded on lab-on-a-chip platforms. Many types of special microparticles are commercially available, for various applications. The microparticles are made from different materials, including: polystyrene, silica, melamine, glass, magnetite, carboxymethylcellulose, biodegradable polymers (such as polylactide, poly(d,l-lactide co-glycolide), triglycerides (tristearin), partial glycerides (monostearate), fatty acids (stearic acid), steroids (cholesterol), waxes (cetyl palmitate) and are available in different sizes usually ranging from 100 nm up to 1 mm. Table 6.1 summarizes some of the microparticles commercially available for lab-on-a-chip applications together with their main characteristics.

The possibility of purchasing microparticles exhibiting different features is very important, considering the wide spectrum of applications we could imagine. For instance, the dimension of microparticles is a crucial parameter for controlling transfection. Cationic microparticles are required for efficient delivery of DNA based gene expression modifiers. The chemical composition of the microparticles is expected play a role in DEP-based motion of them.

As far as stability, most microparticles are supplied in deionized water and have a shelf life of several years when stored at 4°C. Biodegradable microparticles, made on basis of e.g. starch, have a shorter shelf life and should be used faster. These microparticles are of great interest for the biological activity of delivered molecules.

6.4.1.3 Experimental Microparticles

In this subsection are summarized data referring to different type of microparticles (lipospheres, cellulose microparticles), specially tailored in our group for specific lab-on-a-chip applications.

6.4.1.3.1 Lipospheres

Lipospheres (LS) have been proposed as a new type of fat-based encapsulation system for drug delivery of bioactive compounds (especially lipophilic compounds) [35, 36]. LS consist of solid microparticles with a mean diameter usually comprised between 0.2 and 500 μm, composed of a solid hydrophobic fat matrix, where the bioactive compound(s) is dissolved or dispersed. LS have some advantages over other delivery systems, such as (a) good physical stability, (b) low cost of ingredients, (c) ease of preparation and scale-up and (d) high entrapment yields for hydrophobic drugs. Due to their large particle size range, LS can be administered by different routes, such as oral, subcutaneous, intramuscular, topic or used for cells encapsulation being thus proposed for the treatment of a number of diseases. To obtain a biocompatible formulation suitable for human administration, triglycerides and monoglycerides have been chosen as LS materials, in reason of their high biocompatibility, high physico-chemical stability and drug delivery release.

Lipospheres can be prepared both by melt-dispersion technique and solvent evaporation procedures (see scheme in Fig. 6.5). In particular for lipospheres produced with the melt-dispersion technique it was possible by adjusting the various experimental parameters: stirring speed, lipid composition and type of stabilizer, to obtain lipospheres perfectly fitting the special

Table 6.1 Examples of commercially available microparticles for lab-on-a-chip applications.

Material	Characteristics
Silica particles	Density of 2.0 g/cm^3, simple to dispense and to separate
Magnetic silica particles	Paramagnetic particles, with 80% magnetite content, allow simple magnet separation.
Silanol-ROB	Particles functionalized in different ways including carboxyl: binding of bioactive substances by the carbodiimide method; amino: binding of bioactive substances by the carbodiimide method; streptavidin: biotin labelled molecules; biotin: avidin/streptavidin labelled molecules.
Glass beads	Made from borosilicate. Non porous, chemically inert and appear as white powder.
Gold particles	Particles have a very narrow size distribution and are available from 2 nm to 250 nm.
Alumina particles	Excellent spherical shape, high thermal conductivity, heat and mechanical resistance.
Polymethacrylate particles	Spherical shape and a narrow size distribution, they are available in diameters from 1 to 100 μm. They are stable at temperatures of up to 200°C
Polystyrene (latex beads)	Extremely monodisperse size distribution (usually available in diameters from 1 to 100 μm).
Coated polystyrene magnetic particles	Prepared by coating a layer of iron oxide and polystyrene onto the core of the particles. They are relatively uniform in size, spherical in shape and paramagnetic in nature,
Microparticles based on melamine resin	New class of polymer particles with excellent physical and chemical properties, such as high temperature resistance up to 300°C, superior mechanical stability (pressure stable)

Table 6.1 (continued) Examples of commercially available microparticles for lab-on-a-chip applications.

Material	Applications
Silica particles	Useful for coupling of DNA, oligonucleotides, oligopeptides, proteins, lectines and antibodies.
Magnetic silica particles	The large surface area of 50 m^2/g provides a high capacity to bind cells, bacteria, DNA/RNA, proteins or antibodies.
Silanol-ROB	Functional microparticles are especially designed for use in DNA/RNA separation with robotic systems.
Glass beads	Usually designed for applications where inorganic beads with high density are required.
Gold particles	Used in production of diagnostic tests as well as conjugation studies of proteins and antibodies.
Alumina particles	Alumina spheres are available for various applications due to the size uniformity
Polymethacrylate particles	Indicated for applications where good biocompatibility and reduced non-specific protein binding are required
Polystyrene (latex beads)	Used in a wide range of immunodiagnostic assays, as size-standards for calibration of equipment, in cell biology applications etc. The physical adsorption of the polystyrene particles is used to bind ligands to the surface of the particles.
Coated polystyrene magnetic particles	These magnetic particles can be used for cell separation, affinity purification, DNA probe assays, magnetic particle EIA etc.
Microparticles based on melamine resin	MF-particles possess many functional groups on the surface (methylol groups, amino groups), which can be used for a covalent attachment of biological ligands (antigens, antibodies, enzymes).

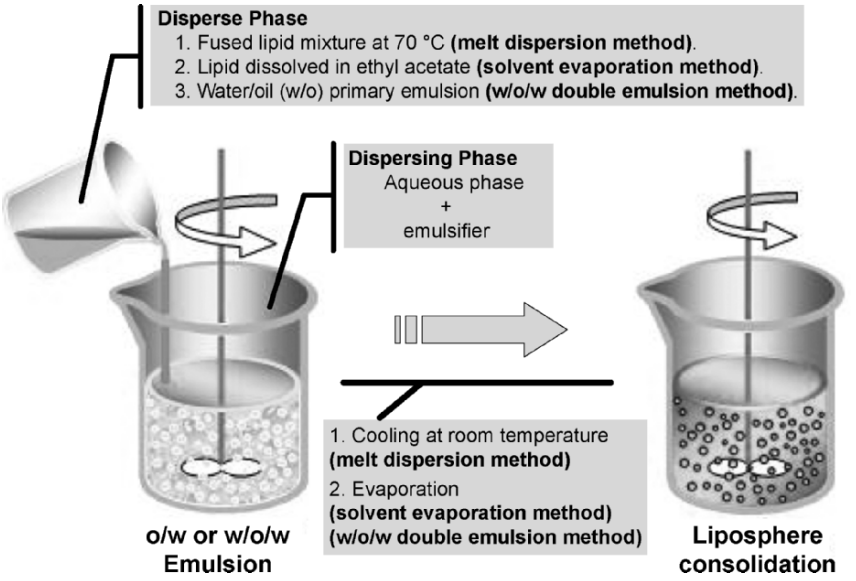


Figure 6.5 Schematic representation of methods of LS production: melt dispersion and solvent evaporation method.

geometric and change requirements of lab-on-a-chip as demonstrated in Figs. 6.6 and 6.7. The results shown in Fig. 6.7 demonstrate that these lipospheres can be easily moved throughout the Lab-on-a-chip device, thus rendering possible experiments aimed at the programmed interactions of these objects to target cells.

6.4.1.3.2 Cellulosic Microparticles

Cellulose esters were selected for microparticle preparation since they represent a large family of cellulose derivatives that have a long history of use in pharmaceutical and industrial applications. Cellulose acetate (CA) has found extensive use in solid pharmaceutical dosage forms, where it is typically used for controlled drug delivery. Typical technologies that employ cellulose esters include semipermeable membranes for osmotic pump drug delivery applications, sustained release from cellulose ester-based matrix formulations, and microparticles formed from cellulose esters and drugs. For microparticle preparation, CA was chosen as a highly lipophilic polymer, in reason of its very low swellability and solubility in water.

CA microspheres were prepared by an oil-in-water emulsion (o/w) method (see scheme in Fig. 6.8). Microparticle morphology, size and size distribution

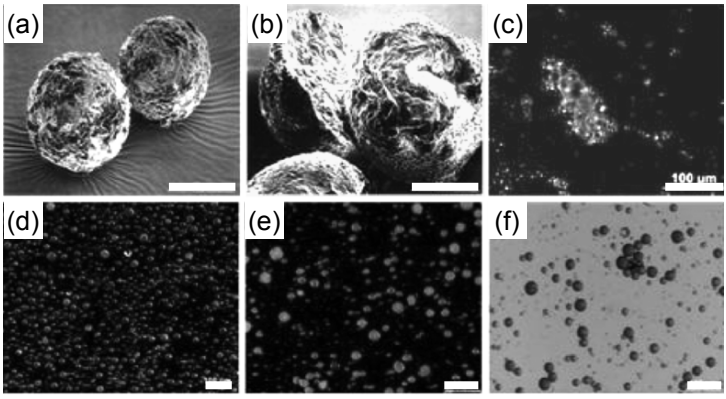


Figure 6.6 Electron scanning and optical stereo photomicrographs of lipospheres. (a) neutral LS, tripalmitin:glycerol monostearate 80:20 (w/w). (b) Section of neutral LS ,tripalmitin:glycerol monostearate 80:20 (w/w). (c) Neutral LS, tripalmitin:glycerol monostearate 80:20 (w/w) and encapsulating the red dye 1,1-[2-methyl-4-(2-methylphenylazo)phenylazo]-2-naphthol (Scarlet Red). (d) Neutral LS, tripalmitin:glycerol monostearate 70:30 (w/w) and encapsulating the red dye 1-[2,5-Dimethyl-4-(2,5-dimethylphenylazo)phenylazo]-2-naphthol (Oil Red O). (e) Cationic LS, tripalmitin:glycerol monostearate 80:20 (w/w) containing as cationic component the dihexadecyl-dimethyl-ammonium bromide and encapsulating the red dye commercially named "Oil Red O". (f) Anionic LS, tripalmitin:stearic acid 80:20 (w/w) and encapsulating the red dye commercially named "Oil Red O". Samples in (a), (b), and (d)-(f) were prepared by melt dispersion technique with a high-shear homogenizer; the sample in (c) was prepared by melt dispersion technique with a ultrasonic generator. In (a) and (b), the bars correspond to 10 µm; in (c)-(f), bars correspond to 100 µm.

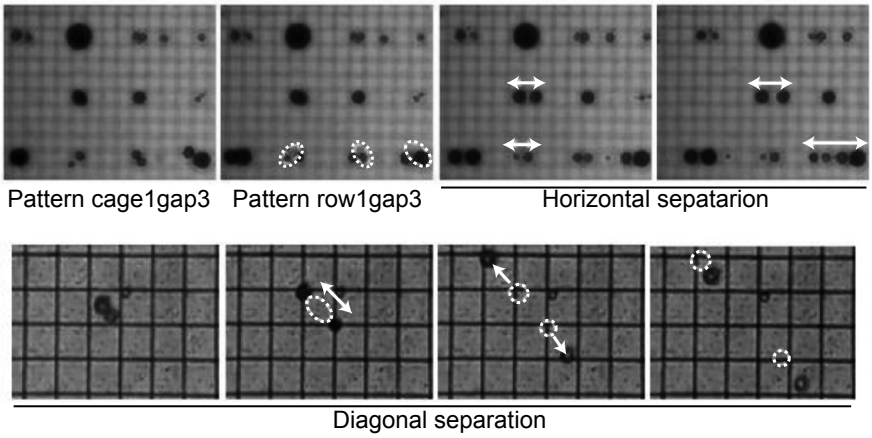


Figure 6.7 Programmed sequential moving of lipospheres on the DEParray™ Chip. Schematic representation of a DEParray™ Chip is shown in Fig. 6.3(c).

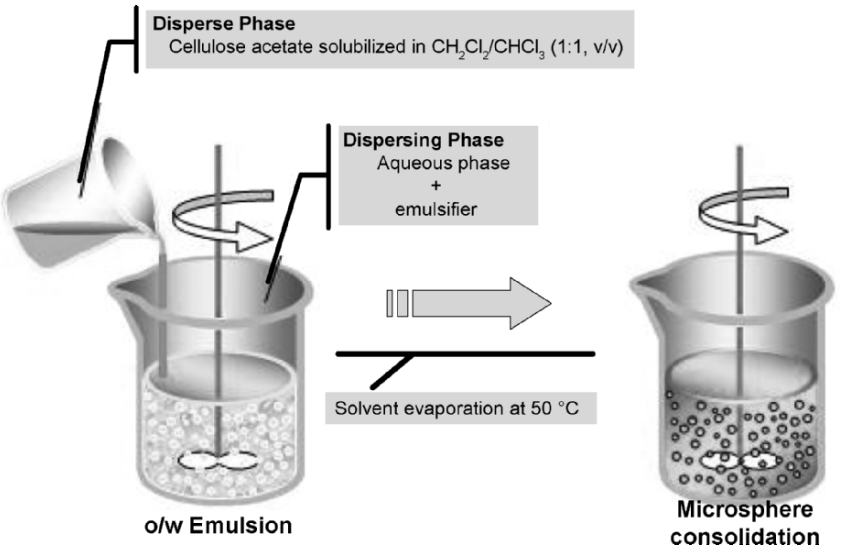


Figure 6.8 Schematic representation of solvent evaporation method for producing cellulose acetate microparticles.

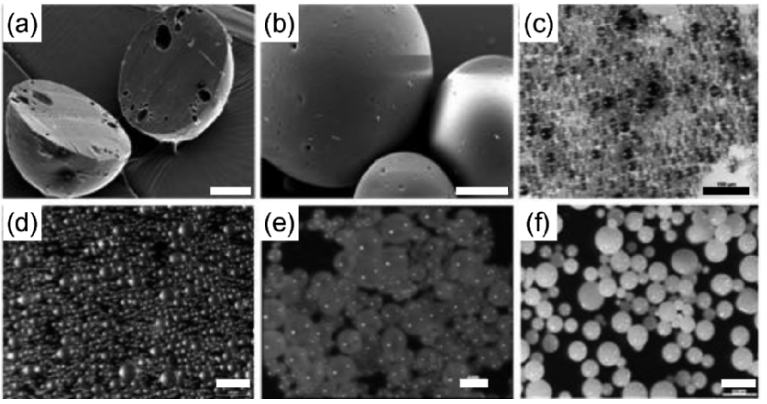


Figure 6.9 Electron scanning and optical stereo photomicrographs of neutral (a-c, f) and cationic (d, e) cellulose acetate microspheres. (c) Microspheres encapsulating the black dye commercially named "Sudan Black B". (d) Cationic microparticles containing as cationic component dimethyl-dioctadecy-lammonium chloride. (e) Cationic microparticles containing as cationic component dimethyl-dioctadecy-lammonium chloride and encapsulating the red dye 1-[2,5-Dimethyl-4-(2,5-dimet hylphenylazo)phenylazo]-2-naphthol (Oil Red O). Scale bars correspond to 20 μm in (a) & (b), 100 μm in (c)-(e), and 200 μm in (f).

were investigated. These parameters are particularly important since the small volume of the location site (the LOAC microchamber) can only accept a limited amount of microparticles with a very narrow size-distribution.

Different experimental variables such as type of oil and water phases, stirring speed, polymer and plasticizer type, surfactant type and concentration and presence of dyes and drugs were analyzed. The effects of some of these variables on morphological characteristics of microspheres are described elsewhere [37].

Light and scanning electron microscopy were employed to analyze the external and internal structure of the microparticles since it is well known that the microcapsule's morphology can influence both the applicability to lab-on-a-chip. In Fig. 6.9 are reported the stereo and electron micrographs of the external and internal structure of CA microparticles. As it is clearly evident, the preparation procedure did not lead to aggregation/agglomeration of the microparticles that, on the contrary, showed regular spherical geometry. Moreover, the internal morphology of the microspheres demonstrated that the particles were compact and homogeneous.

Since our aim was to obtain microspheres to be loaded in a LOAC prototype, we modulated the stirring speed, and succeeded in obtaining microspheres within suitable dimensional range (see Fig. 6.10).

With the aim to produce colored microparticles to improve their detection when loaded in the LOAC, under optical microscopes, we investigated the possibility to load dyes in the CA microparticles. For all the tested dyes, we were able to produce and isolated microparticles with optimal morphological characteristics and a homogeneous distribution of the dye. These microparticles are of great interest for the use on projects aimed at the characterization of the membrane phenotype of target cells. Combined with suitable imaging software, multiple characterization of target cells can be performed in parallel using microspheres loaded with different dyes and functionalized with different antibodies.

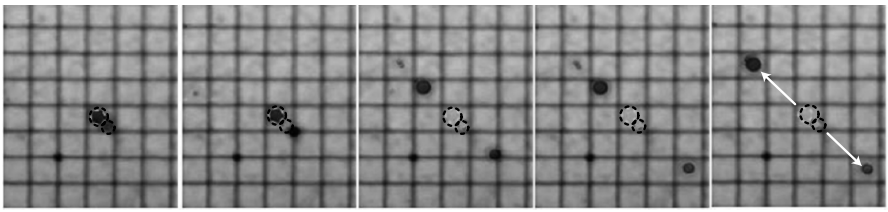


Figure 6.10 Programmed sequential diagonal separation of cationic red microspheres on the DEParray™ Chip.

6.4.2 Microparticles-cell Interactions on Lab-on-a-chip

Software-guided interactions between microspheres and target cells can be efficiently achieved with the use of DEP-based LOACs. For instance, a LOAC carrying parallel electrodes can be employed in order to force the interactions between microspheres and K562 cells [32], after moving them at the central electrode within the corresponding DEP cage.

A second example is shown in Fig. 6.11 and was performed with the LOAC described in [25-27] in order to determine whether this system would be suitable for directing single microspheres to a single identified target cell. In the experiment shown in Fig. 6.11, three cationic microspheres (M1, M2 and M3) and two K562 cells are shown, entrapped in six independent spherical DEP-cages. Only one of the two cells, named T-K562, was designed to be the cellular "target" of the three microspheres. We first moved the microsphere M1, generating the cell-microsphere complex shown in Fig. 6.11(b), then the microsphere M2, generating the T-K562/M1M2 complex shown in Fig. 6.11(c). Finally, we moved the M3 microsphere for a further targeting of the T-K562 cell, obtaining the T-K562/M1M2M3 complex shown in Fig. 6.11(d). The detail of this experiment is described in Borgatti *et al.* [32].

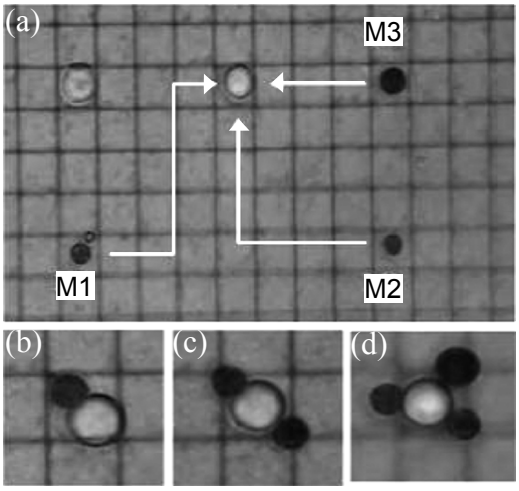


Figure 6.11 Programmed sequential interactions between three cationic microspheres (M1, M2 and M3) and one target K562 cell. The microsphere M1 was moved first in order to obtain the T-K562/M1 complex shown in (b). Then the M2 and M3 microspheres were moved to obtain the T-K562/M1M2 and T-K562/M1M2M3 complexes shown in (c) and (d), respectively. Data are adopted from [32].

As far as possible applications to pharmaceutical sciences, these results are of interest as an enabling technology to understand the controlled delivery of a precise amount of pharmaceutical products to target cells. This understanding can improve screening of new compounds with therapeutic effects since it allows the selection of cells resistant to specific treatments. In diagnostics, this approach allows the analysis of heterogeneous cell populations with microspheres carrying a same antibody or, conversely, the multivariate characterization of selected homogenous cell populations with different microspheres functionalized with a variety of antibodies.

6.5 LAB-ON-A-CHIP FOR BIOMEDICINE AND CELLULAR BIOTECHNOLOGY

6.5.1 Applications of Lab-on-a-chip for Cell Isolation

The possibility to isolate cell populations or single cells by the aid of DEP-based LOAC platforms has been reported in several papers [28, 38-41]. In this respects, isolating rare cells from biological fluids, including whole blood or bone marrow, is an interesting biological problem. Rare cells isolation would provide novel possibilities to cell-based diagnosis, cell-based therapy, and phenotypic characterization of cell population sub-sets. For instance, characterization of a few metastatic cells from cancer patients for further molecular analysis is desirable for prognosis/diagnosis. In this respect, isolation of the so called "cancer stem cells" might be of great relevance.

An example of the application of DEP-based protocols for cell isolation has been recently described by Yu *et al*. [40], who reported the use of an efficient method for trapping neurons and constructing ordered neuronal networks on Cell Positioning Bioelectronic Chip (CPBC) by using arrayed negative dielectrophoretic (DEP) forces [40]. In this case, a special CPBC with well defined positioning of the electrodes was designed and fabricated on silicon substrate. When a high frequency AC signal was applied, the CPBC is able to provide a well-defined non-uniform electric field, and thus generate negative DEP forces. Interestingly, when a neuron suspension was added onto the energized CPBC, the neurons were immediately trapped and quickly formed the predetermined pattern. According with the protocol developed by Yu *et al*., neurons may adhere and then be cultured directly on

the CPBC, and show good neuron viability and neurite development [40], suggesting that this approach could be used to characterize functional activities of neuronal networks. Hence the development of novel technologies aimed to pattern neurons into regular networks is of great scientific interest in the field of neurological research.

6.5.2 Separation of Cell Populations Exhibiting Different DEP Properties

Several examples have been reported on this specific issue [16, 26, 42, 43]. For instance, Huang *et al*. reported a dielectrophoretic field-flow-fractionation (DEP-FFF) method and used it to purge human breast cancer MDA-435 cells from hematopoietic $CD34^+$ stem cells [42]. In the case reported by these authors, an array of interdigitated microelectrodes lining the bottom surface of a thin chamber was used to generate dielectrophoretic forces that levitated the cell mixture in a fluid flow profile. $CD34^+$ stem cells were levitated higher, were carried faster by the fluid flow, and exited the separation chamber earlier than the cancer cells. The same group was able to demonstrate high separation performance in separating T- (or B-) lymphocytes from monocytes, T- (or B-) lymphocytes from granulocytes, and monocytes from granulocytes [42].

Another interesting application of DEP-based Lab-on-a-chip is the possibility to separate viable from non-viable yeast using dielectrophoresis. Markx *et al*. in [43] employed known mixtures of viable and heat-treated cells of *Saccharomyces cerevisiae* (60% non-viable cells present); these two cell populations were separated and selectively isolated using positive and negative dielectrophoretic forces generated by microelectrodes in a small chamber. Good correlations with the initial known relative compositions were obtained by direct microscopic counting of cells at the electrodes after initial dielectrophoretic separation, from methylene blue staining and by optical absorption measurements of the effluent after selectively flushing out the viable and non-viable cells from the chamber. The results obtained demonstrated that DEP-separated non-viable fraction contained 3% viable cells and the viable fraction 8% dead cells. Importantly, cell viability was not affected by the separation procedure.

A further example is represented by the possibility of using DEP-based LOAC for the isolation of infected cells from uninfected counterparts. This was shown by recent papers demonstrating that platforms carrying spiral electrodes can be used for isolating malaria-infected cells from blood [44-45]. In this case, during development of *Plasmodium falciparum*

(the malaria pathogen) there is an increase in the ionic permeability of the plasma membrane of infected erythrocytes. When challenged by suspension in a low conductivity medium, infected cells lose internal ions while uninfected cells retain them. The resultant dielectric differences between infected and uninfected cells were exploited by DEP manipulation employing spiral electrodes, with parasitized cells focused at the centre of a spiral electrode array. Notably, the dielectrophoretic methods require only a few micro liters of blood, and should be applicable to the production of small, low-cost automated devices for assessing parasite concentrations with potential applicability to drug sensitivity studies and the diagnosis of malaria. In agreement with these results, DEP-based protocols offer great potential in cell discrimination and isolation.

Furthermore, we have demonstrated the application of a printed circuit board (PCB)-based chip, generating DEP-based cylinder-shaped cages for separation and recovery of white blood cells from erythrocytes [46]. This possibility is of interest to develop low-cost LOACs for diagnostic purposes. Accordingly, we demonstrate that white blood cells recovered from this Lab-on-a-chip device are suitable for PCR-based molecular diagnosis procedures employing DNA sequencing or biospecific interaction analysis (BIA) using surface plasmon resonance (SPR) and biosensor technology [46].

6.5.3 DEP-based, Marker-Specific Sorting of Rare Cells

Most of the techniques currently available for high-speed cell sorting, especially in the field of isolation of rare cells, are limited by the inherent coupling among three competing parameters of performance: throughput, purity, and rare cell recovery. To efficiently isolate marker-specifically selected rare cells from complex mixtures, one possibility is represented by a recently described electro kinetic sorting methodology exploiting DEP in microfluidic channels. In this approach, the dielectrophoretic amplitude response of rare target cells is modulated by labeling cells with particles that differ in polarization response. Cell mixtures were analyzed in a continuous-flow DEP-activated cell sorter (DACS), wherein the electric fields were engineered to achieve efficient separation between dielectrophoretically labeled and unlabeled cells. This strategy was applied for affinity-based enrichment of rare bacteria expressing a specific surface marker from an excess of non target bacteria not expressing the marker used for selection. Rare target cells were enriched more than 200-fold in a single round of sorting at a single-channel throughput of 10,000 cells per second. DACS offers

the potential for automated, surface marker-specific cell sorting in a disposable format that is capable of simultaneously achieving high throughput, purity, and rare cell recovery [47].

A second approach is the DEP-based routing of identified rare cells to selected recovery fields of the LOAC. In this case, cells are identified in virtue of fluorescence labeling using monoclonal antibodies. A proof of principle of this strategy has been published by Fuchs *et al*. [48]. Sorting and recovering specific live cells from samples containing less than a few thousand cells were approached in this study employing a new technology described in [52] which allows individual and parallel single cell manipulation of up to 10,000 cells. One of the most important features of this approach is related to the fact that the cells maintain viability and proliferation capabilities. Manipulation was carried out using dynamic dielectrophoretic traps controlled by an electronic interface. These authors also demonstrate the capabilities of the chip by sorting and recovering individual live fluorescent cells from an unlabeled population [48].

6.6 FUTURE PERSPECTIVES: INTEGRATED SENSORS FOR CELL BIOLOGY

The development of effective integrated biosensors plays an essential role in LOAC technology for the automation of protocols, cost saving, and speed-up of processes. Feedback from such microsystems is essential for analyses and for carrying out complex tasks in an automated way. Moreover, though at the price of additional non-standard steps, CMOS technology has demonstrated that it is suitable to exploit different sensing principles for different biological applications, ranging from detection of molecules (e.g. DNA strands) to up to the cell level. Electrochemical sensors [49] have proven to be suitable for DNA detection in an integrated array structure, as well as sensing principles based on electric properties of analytes, such as conductivity [50].

However, one of the most challenging topics in the field of bioelectronics is the development of integrated microsystems capable of interactions with living cells and organisms. In this field, recent CMOS designs have brought innovative contributions.

Eversmann *et al*. [51] described a biosensor array chip allows high resolution imaging of extracellular signals from neural cells or tissue cultured on the chip surface. The chip provides 128×128 capacitive sensors in an area of 1 mm^2. Detection circuitry is based on a sensor-MOSFET mismatch-

compensated current mode technique. A potentiostat is used to adjust the potential of the electrolyte to the value of a programmable off-chip reference voltage and electrophysiological references are required for chip operation. Signal levels above 100 μV are detected while monitoring neural activity.

The ability to detect objects at a cell level is also required for LOAC applications in rare cell population analyses and individual cell biology. In fact, the CMOS-compatible DEP technique may be paired with specific biosensors in order to achieve automated analysis protocols. Integrated detection in CMOS LOACs has been successfully performed either with optical sensors [19] or impedance measurements [52]. Bioparticles induce variations in the impinging light or in the surrounding dielectric properties which can be monitored by these devices. A massively-parallel integration on the same substrate of 320 × 320, 20 μm × 20 μm sensors/actuators allows the manipulation under software control of more than 10,000 particles and the detection of their positions along the chip surface.

Circuit noise reduction is essential when sensing bioparticles by means of embedded CMOS circuitry. In the above devices, small signal variations have to be detected. Output voltages of sensing elements are brought to a reference value during the reset phase, then signal changes due to presence or absence of particles (i.e. in impinging light or in the electric field) will occur during the sensing phase upon proper circuit stimulation. Differential sensing of output voltage variation after the reset of the cell, usually referred to as correlated double sensing (CDS) technique, allows compensation of noised induced by KT/C effect and charge injection of reset transistor, as well as for the variations of parameters of cell readout transistors within the array, called fixed pattern noise. Additionally, CDS techniques further improve the signal to noise ratio of the readout process by greatly reducing low bandwidth noise sources such as the $1/f$ noise [53]. Furthermore, a programmable offset subtraction in the readout stage allows for higher gain while keeping signals in proper ranges, thus boosting signal dynamics.

The above circuits allow the detection of different types of bioparticles, ranging from polystyrene micro-beads to living cells like yeasts and human erithroleukaemia cells. Some results are shown in Fig. 6.12. Circuit noise measured at the output of the device shows that a significant limit to sensitivity is due to external noise (i.e. from external components and ion currents in suspended solutions), which depends on the implementation of the system. Features like on-chip or in-pixel A/D conversion may help improve sensing quality.

Another important issue is the size of sensing micro sites. Strong miniaturization is technologically possible, but benefits may be relevant only

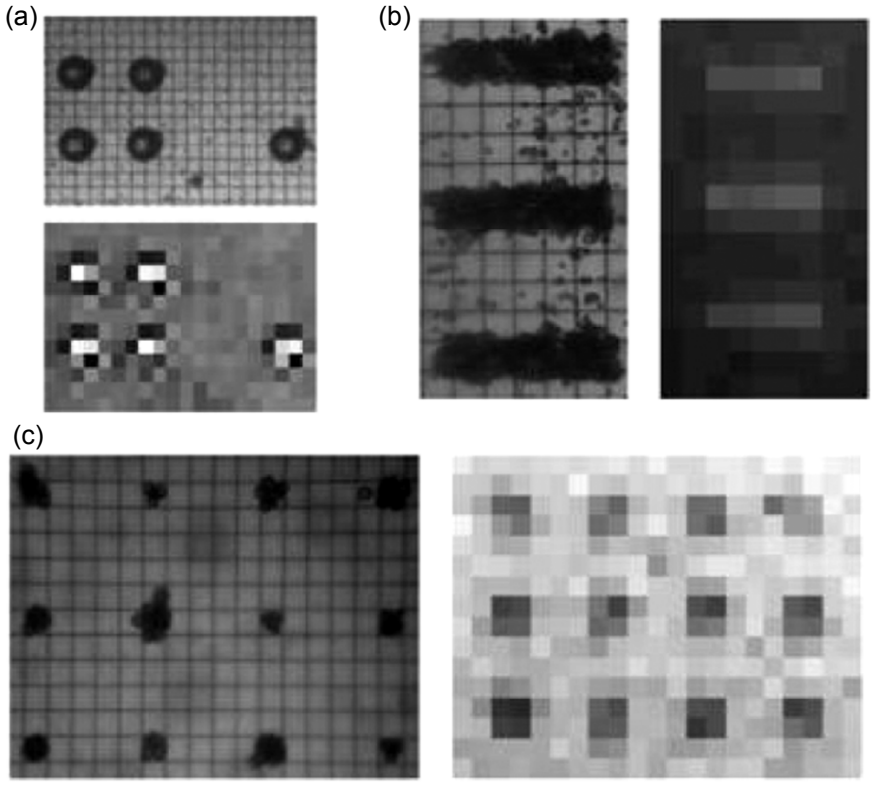


Figure 6.12 Comparison of images from optical microscope and sensors. (a) 50 μm polystyrene microbeads are detected optically (bottom). In (b) and (c) capacitive sensors images of yeasts (b) and 10 μm beads (c) are shown respectively.

when sensitivity scales favorably. Detectable amounts of some chemicals and micro organisms may require relatively large volumes of samples due to their low concentrations; however, concentration techniques based on DEP may improve the detection of sparse low-concentration particles. Moreover, although the volume of liquids scales down with size, the surface to volume ratio increases. Thus, evaporation becomes a relevant issue to be handled when dealing with sample volumes of a few μL. In addition, the scaling trend allows for the design of highly parallel sensor arrays, which means larger amount of available data from our systems. CMOS technology is mandatory to allow proper connectivity in such microarrays. Scaling of geometries also offers a closer coupling and stronger interactions between sensors and samples, which may be essential in capacitive and optical sensing schemes. A more quantitative and exhaustive analysis of these and other related topics is presented by Madou and Cubicciotti [54].

The current limits of detection of overlying living or not-living particles (e.g. cells or microbeads) and their relationships to scaling trends relate signal to noise ratio with parameters linked to the minimum feature size scaling trend of CMOS technology. In fact, the reduction in size of the CMOS devices creates a very complex interplay between the different noise sources induced by the physics of the device itself and the architectural possibilities made available by such scaling down. In fact since cells remain constant in size, we are going to see useful computing power in areas compatible with that located under a cell. Hence, it is clear that the possibility to provide signal conditioning exactly where the signal is generated offers unparalleled approaches to improve the crucial issue of the signal to noise ratio that still limits existing microsystems.

6.7 CONCLUSIONS

The data here reported sustain the concept that DEP-based Lab-on-a-chip devices are of interest for future applications in the field of applied cellular biology, diagnosis, therapy, pharmaceutical sciences. An expected achievement in applied cellular biology is the isolation of rare cell from heterogeneous cell population (examples: isolation of rare stem cells, including cancer stem cells; isolation of trophoblasts and nucleated red blood cells from maternal circulation in non invasive prenatal diagnosis). In the field of diagnosis, the possibility of easy manipulation of microspheres will allow the immunophenotypization of target cells. As far as therapy and pharmaceutical sciences, the improvement is expected on the quantitative loading of target cells with drugs to be evaluated for alteration of gene expression (including the possibility to select cells resistant to the treatment).

To achieve these results, a substantial contribution will come from the engineering side of this field. After years of creative research, major questions have found positive answers as discussed in this chapter. The need to consolidate these results in products that make available to the health care practitioner these new approaches requires a strong effort to define standards that allow interoperability of LOACs in a laboratory setting and the definition of platforms that can obtain the approval of regulatory bodies. The impact of these requirements is likely to change in a significant way the existing landscape of the LOACs.

ACKNOWLEDGMENT

This work is supported by FIRB-2001 (Italian Fund for basic research from the Italian Ministry of University) to Roberto Guerrieri, Roberto Gambari, and Claudio Nastruzzi, by the IST 6FP project Cochise to Roberto Guerrieri and Roberto Gambari and Associazione Veneta per la Lotta alla Talassemia, Rovigo to Roberto Gambari.

REFERENCES

[1] G. Gastrock, K. Lemke, J. Metze, "Sampling and monitoring in bioprocessing using microtechniques," J Biotechnol, vol. 82, pp. 123-135, 2001.

[2] H.M. Haake, A. Schutz, G. Gauglitz, "Label-free detection of biomolecular interaction by optical sensors," Fresenius J Anal Chem, vol. 366, pp. 576-585, 2000.

[3] R.C. McGlennen, "Miniaturization technologies for molecular diagnostics," Clin Chem, vol. 47, pp. 393-402, 2001.

[4] R.L. Rich, D.G. Myszka, "Survey of the year 2000 commercial optical biosensor literature", J Mol Recogn, vol. 14, pp. 273-294, 2001.

[5] P. Fortina, S. Surrey, L.J. Kricka, "Molecular diagnostics: hurdles for clinical implementation,", Trends Mol Med, vol. 8, pp. 264-266, 2002.

[6] K.K. Jain, "Cambridge Healthtech Institute's Third Annual Conference on Lab-on-a-Chip and Microarrays," Zurich, Switzerland, Pharmacogenomics, vol. 2, pp. 73-77, 2001.

[7] L.J. Kricka, "Microchips, microarrays, biochips and nanochips: Personal laboratories for the 21st century," Clin Chim Acta, vol. 307, pp. 219-223, 2001.

[8] N. Manaresi, A. Romani, G. Medoro, L. Altomare, A. Leonardi, M. Tartagni, R. Guerrieri, "A CMOS Chip for Individual Cell Manipulation and Detection," 2003 IEEE ISSCC Dig Tech Papers, pp. 192-193, 2003.

[9] H. Morgan, M.P. Hughes, N.G. Green, "Separation of submicron bioparticles by dielectrophoresis," Biophys J, vol. 77, pp. 516-525, 1999.

[10] J. Voldman, R.A. Braff, M. Toner, M.L. Gray, M.A. Schmidt, "Holding forces of single-particle dielectrophoretic traps," Biophys J, vol. 80, pp. 531-541, 2001.

[11] L.J. Kricka, "Microchips, microarrays, biochips and nanochips: personal laboratories for the 21st century", Clin Chim Acta, vol. 307, pp. 219-223, 2001.

[12] P. Hawtin, I. Hardern, R. Wittig, J. Mollenhauer, A. Poustka, R. Salowsky, T. Wulff, C. Rizzo, B. Wilson, "Utility of lab-on-a-chip technology for high-throughput nucleic acid and protein analysis," Electrophoresis, vol. 26, pp. 3674-3681, 2005

[13] H. Goetz, M. Kuschel, T. Wulff, C. Sauber, C. Miller, S. Fisher, C. Woodward, "Comparison of selected analytical techniques for protein sizing, quantitation and molecular weight determination," J Biochem Biophys Methods, vol. 60, pp. 281-293, 2004.

[14] L. Shi, W. Tong, F. Goodsaid, F.W. Frueh, H. Fang, T. Han, J.C. Fuscoe, D.A: Casciano, "QA/QC: challenges and pitfalls facing the microarray community and regulatory agencies," Expert Rev Mol Diagn, vol. 4, pp. 761-777, 2004.

[15] T. Muller, A. Pfennig, P. Klein, G. Gradl, M. Jager, T. Schnelle, "The potential of dielectrophoresis for single-cell experiments," IEEE Eng Med Biol, vol. 22, pp. 51-61, 2003.

[16] J. Yang, Y. Huang, X.B. Wang, F.F. Becker, P.R. Gascoyne, "Cell separation on microfabricated electrodes using dielectrophoretic/gravitational field-flow fractionation," Anal Chem, vol. 71, pp. 911-918, 1999.

[17] P.R. Gascoyne, J. Vykoukal, "Particle separation by dielectrophoresis," Electrophoresis, vol. 23, pp. 1973-1983, 2002.

[18] T. Schnelle, T. Müller, G. Gradl, G. Shirley, G. Fuhr, "Paired Microelectrode System: Dielectrophoretic Particle Sorting and Force Calibration," J Electrost, vol. 47, pp. 121-132, 1999.

[19] N. Manaresi, A. Romani, G. Medoro, L. Altomare, A. Leonardi, M. Tartagni, R. Guerrieri, "A CMOS Chip for Individual Cell Manipulation and Detection," IEEE J Solid-State Circuit, vol. 38, pp. 2297-2305, 2003.

[20] H.A. Pohl, "The Motion and the Precipitation of Suspensoid in Divergent Electric Fields," J Appl Phys D: Appl Phys, vol. 22, pp. 869-871, 1951.

[21] G. Fuhr, H. Glasser, T. Müller, T. Schnelle, "Cell Manipulation and Cultivation Under A.C. Electric Field Influence in Highly Conductive Culture Media," Biochim. and Biophys. Acta, vol. 1201, pp. 353-360, 1994.

[22] R. Pethig, Y. Huang, X. Wang, J.P.H. Burt, "Positive and Negative Dielectrophoretic Collection of Colloidal Particles Using Interdigitated Castellated Microelectrodes," J Phys D: Appl Phys, vol. 25, pp. 881-888, 1992.

[23] H.A. Pohl, Dielectrophoresis, Cambridge University Press, 1978.

[24] H. Morgan, N.G. Green, AC Electrokinetics: colloids and nanoparticles, Research Studies Press, 2003.

[25] G. Medoro, N. Manaresi, M. Tartagni, R. Guerrieri, " CMOS-only sensor and manipulator for microorganisms," Proceedings IEDM, pp. 415-418, 2000.

[26] G. Medoro, N. Manaresi, A. Leonardi, L. Altomare, M. Tartagni, R. Guerrieri, "A Lab-on-a-chip for cell detection and manipulation," Proceedings IEEE Sensors Conference, 2002.

[27] G. Medoro, N. Manaresi, M. Tartagni, L. Altomare, A. Leonardi, R. Guerrieri, "A Lab-on-a-chip for cell separation based on the moving-cages approach," Eurosensors, 2002.

[28] X.B. Wang, Y. Huang, X. Wang, F.F. Becker, P.R. Gascoyne , "Dielectrophoretic manipulation of cells with spiral electrodes," Biophys J, vol. 72, pp. 1887-1899, 1997.

[29] L. Altomare, M. Borgatti, G. Medoro, N. Manaresi, M. Tartagni, R. Guerrieri, R. Gambari, "Levitation and Movement of Human Tumor Cells using a Printed Circuit Board Device Based on Software-controlled Dielectrophoresis," Biotechnol Bioeng, vol. 82, pp. 474-479, 2003.

[30] Y. Huang, S. Joo, M. Duhon, M. Heller, B. Wallace, X. Xu, "Dielectrophoretic cell separation and gene expression profiling on microelectronic chip arrays," Anal Chem, vol. 74, pp. 3362-3371, 2002.

[31] J.O. Koskinen, N.J. Meltola, E. Soini, A.E. Soini, "A lab-on-a-chip compatible bioaffinity assay method for human alpha-fetoprotein," Lab Chip, vol. 5, pp. 1408-1411, 2005.

[32] M. Borgatti, L. Altomare, M. Abonnec, E. Fabbri, N. Manaresi, G. Medoro, A. Romani, M. Tartagni, C. Nastruzzi, S. Di Croce, A. Tosi, I. Mancini, R. Guerrieri, R. Gambari, "Dielectrophoresis-based 'Lab-on-a-chip' devices for programmable binding of microspheres to target cells," Int J Onc, vol. 27, pp. 1559-1566, 2005.

[33] N. Chronis, L.P. Lee, "Total internal reflection-based biochip utilizing a polymer-filled cavity with a micromirror sidewall," Lab Chip, vol. 4, pp. 125-130, 2004.

[34] F. Vollmer, S. Arnold, D. Braun, I. Teraoka, A. Libchaber, "Multiplexed DNA quantification by spectroscopic shift of two microsphere cavities," Biophys J, vol. 85, pp. 1974-1979, 2003.

[35] R. Cortesi, E. Esposito, G. Luca, C. Nastruzzi, "Production of lipospheres as carriers for bioactive compounds," Biomaterials, vol. 23, pp. 2283-2294, 2002.

[36] Claudio Nastruzzi (Editor), Lipospheres in drug targets and delivery, CRC Press, Boca Raton, 2005.

[37] S. Di Croce, A. Tosi, L. Bilancetti, S. Mazzitelli, N. Bozzuto, M. Borgatti, E. Fabbri, I. Mancini, R. Gambari, C. Nastruzzi, "Cellulose acetate microparticles for Lab-on-a-chip applications," New trends in bioencapsulation-Minerva Biotecnologica, vol. 17, pp. 259-269, 2005.

[38] S. Fiedler, S.G. Shirley, T. Schnelle, G. Fuhr, "Dielectrophoretic sorting of particles and cells in a microsystem," Anal Chem, vol. 70, pp. 1909-1915, 1998.

[39] R. Gambari, M. Borgatti, L. Altomare, N. Manaresi, G. Medoro, A. Romani, M. Tartagni, R. Guerrieri, "Applications to cancer research of lab-on-a-chip devices based on dielectrophoresis (DEP)," Technol Cancer Res Treat, vol. 2, pp. 31-40, 2003.

[40] Z. Yu, G. Xiang, L. Pan, L. Huang, Z. Yu, W. Xing, J. Cheng, "Negative dielectrophoretic force assisted construction of ordered neuronal networks on cell positioning bioelectronic chips," Biomed Microdevices, vol. 6, pp. 311-324, 2004.

[41] C. Xu, Y. Wang, M. Cao, Z. Lu, "Dielectrophoresis of human red cells in microchips," Electrophoresis, vol. 20, pp. 1829-1831, 1999.

[42] Y. Huang, J. Yang, X.B. Wang, F.F. Becker, P.R. Gascoyne, "The removal of human breast cancer cells from hematopoietic CD34+ stem cells by dielectrophoretic field-flow-fractionation," J Hematother Stem Cell Res, vol. 8, pp. 481-490, 1999.

[43] G.H. Markx, M.S. Talary, R. Pethig, "Separation of viable and non-viable yeast using dielectrophoresis," J Biotechnol, vol. 32, pp. 29-37, 1994.

[44] P. Gascoyne, C. Mahidol, M. Ruchirawat, J. Satayavivad, P. Watcharasit, F.F. Becker, "Microsample preparation by dielectrophoresis: isolation of malaria," Lab Chip, vol. 2, pp. 70-75, 2002.

[45] P. Gascoyne, J. Satayavivad, M. Ruchirawat, "Microfluidic approaches to malaria detection," Acta Trop, vol. 89, pp. 357-369, 2004.

[46] M. Borgatti, L. Altomare, M. Baruffa, E. Fabbri, G. Breveglieri, G. Feriotto, N. Manaresi, G. Medoro, A. Romani, M. Tartagni, R. Gambari, R. Guerrieri, "Separation of white blood cells from erythrocytes on a dielectrophoresis (DEP) based 'Lab-on-a-chip' device," Int J Mol Med, vol. 15, pp. 913-920, 2005.

[47] X. Hu, P.H. Bessette, J. Qian, C.D. Meinhart, P.S. Daugherty, H.T. Soh, "Marker-specific sorting of rare cells using dielectrophoresis," Proc Natl Acad Sci USA, vol. 102, pp. 15757-15761, 2005.

[48] A.B. Fuchs, A. Romani, D. Freida, G. Medoro, M. Abonnenc, L. Altomare, I. Chartier, D. Guergour, C. Villiers, P.N. Marche, M. Tartagni, R. Guerrieri, F. Chatelain, N. Manaresi, "Electronic sorting and recovery of single live cells from microlitre sized samples," Lab Chip, vol. 6, pp. 121-126, 2006.

[49] R. Thewes, F. Hofmann, A. Frey, B. Holzapfl, M. Schienle, C. Paulus *et al*., "Sensor Arrays for Fully-Electronic DNA Detection on CMOS," 2002 IEEE ISSCC Dig Tech Papers, pp. 350-351, 2002,

[50] M. Xue, J. Li, Z. Lu, C. Feng, Z. Zhang, P.K. Ko *et al*., "A High-Density Conduction Based Micro DNA Identification Array Fabricated with a CMOS Compatible Process," IEEE Trans Electron Devices, vol. 50, pp. 2165-2170, 2003.

[51] B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert, M. Steinhauser, G. Eckstein, D. Schmitt-Landsiedel, R. Thewes, "A 128×128 CMOS Biosensor Array for Extracellular Recording of Neural Activity," IEEE J Solid-State Circuit, vol. 38, pp. 2306-2317, 2003.

[52] A. Romani, N. Manaresi, L. Marzocchi, G. Medoro, A. Leonardi, L. Altomare, M. Tartagni, R. Guerrieri, "Capacitive Sensor Array for Localization of Bioparticles in CMOS Lab-on-a-chip," 2004 IEEE ISSCC Dig Tech Papers, pp. 224-225, 2004.

[53] Aldo Romani, Roberto Guerrieri, Marco Tartagni, Nicolò Manaresi, Gianni Medoro, "Beyond the Microscope: Embedded Detectors for Cell Biology Applications;" Proc of ISCAS, 2005.

[54] M.J. Madou, R. Cubicciotti, "Scaling Issues in Chemical and Biological Sensors," Proc IEEE, vol. 91, pp. 830-838, 2003.

AUTHOR BIOGRAPHY

Claudio Nastruzzi obtained his Ph.D. in Pharmaceutical Science at University of Ferrara in 1988. From 1988-1990, he worked as a post-doctoral fellow at the Institute for Polymers of the Eidgenissische Technische Hochschule, (Swiss Federal Institute of Technology) (ETH), Zürich in the group of Prof. P.L. Luisi. From 1990-1991, he was a post-doctoral fellow at the Department of Pharmaceutical Sciences of the University of Ferrara, having as main topic the production and characterization of liposomes especially designed for retinoids delivery. Since 1998 he is an associated Professor at the Dipartimento di Chimica e Tecnologia del Farmaco of the University of Perugia, Italy. The main research topics of Prof. Nastruzzi are: (a) the production and characterization of innovative dermatologic and cosmetic formulations, for dermal and trans-dermal delivery (phospholipid based microemulsion gels, cubic phases, liposomes and niosomes; (b) the production and characterization of microspheres and solid lipid microparticles; and (c) the study of liposome based formulations for gene delivery. More recently he investigated new research areas including the application of microparticles to lab-on-chip platforms and the production of polymeric scaffolds for tissue engineering and cell transplantation protocols. He published more than 100 papers in international journals and presented more than 80 contributions to national and international congress.

Azzurra Tosi is born in Ferrara, Italy on April 13th 1977. In July 2003 she spent a short period of time at the University of Verona to learn the main method applied to the proteomic analysis. In 2003 she received the Master Degree in Chemistry at the University of Ferrara. The main subject of her thesis is the decodification of two-dimensional polyacrylamide gel electrophoresis complex mapped by autocovariance. From 2004 to 2006, she worked at the Laboratory of Biomaterials and Bioencapsulation at the University of Perugia on the project "Production and characterization of nano- and microparticular systems for biopharmaceutical applications on a Lab-on-a-chip, based on microelectronic technology". She is currently a PhD fellow at the Department of Chemistry and Pharmaceutical Technology at the University of Siena and the main topic of her research is the development of lipid microparticles for topical applications.

Monica Borgatti received the degree in Chemical and Pharmaceutical Technologies from Ferrara University in 1999 and the Ph.D. degree in Biotechnology at the Department of Biochemistry and Molecular Biology, Ferrara University in 2004. She is a Professor of Structural Biochemistry since 2005 and member of Directive Council of Laboratory on Pharmacological and Pharmacogenomic Therapy of Thalassemia, at Ferrara University, since 2000. In 2003, Monica Borgatti received the CIB Young Researcher prize with the best poster, presenting the study: "A printed circuit board device based on software-controlled dielectrophoresis: levitation, movement and separation of human cells" (6th Biotechnology National Congress, Padova, Italy). In 2000, Dr. Borgatti was a scientific advisory of the Young Researchers Project "Use of PNAs (Peptide Nucleic Acids) for the identifi-

cation of GMO using advanced technologies". Her research interests are on various aspects of molecular biology, genetic medicine, gene therapy, pharmacogenomics. More than 35 papers, published by Monica Borgatti, are available in Medline. She is a co-inventor of two patents.

Roberto Guerrieri received the Dr.Eng. and Ph.D. degree from the University of Bologna, Italy, in 1980 and 1986 respectively. After working at the Department of Electrical Engineering and Computer Sciences of the University of California at Berkeley as Visiting Researcher and at the MIT as visiting scientist, he joined the University of Bologna where he is a Full Professor and teaches design of integrated systems. His research interests are in various aspects of integrated circuit modeling and design, including digital systems and biometric sensors, and applications of microelectronics to biotechnology. His work on VLSI design has been cited by widely read magazines, such as the Nikkei and Electronic Design and documented in more than 90 scientific papers. In 1992 he won the best paper award of the IEEE Transactions on Semiconductor Manufacturing for his research carried out on issues related to the modeling of various IC manufacturing steps. In 2004 he received an ISSCC best paper award for his work in the area of silicon-based lab-on-a-chip. His group invented the concept of silicon-based sensing of fingerprints currently produced by UPEK, and is cofounder of two companies in the field of biomedical devices. One of these, Silicon Biosystems, has developed a platform for the manipulation of rare cells. A second one has recently started a research activity in the field of immuno-oncology and is developing a technology to understand the response of the immune system to the development of cancer.

Gianni Medoro received his Ph.D. in Electrical Engineering and Computer Sciences from the University of Bologna, Italy in 2004. From 1999 to 2005 he held several positions as research fellow and post-docctoral fellow at the Advanced Research Center on Electronic Systems (ARCES) University of Bologna, Italy. In 1999 Gianni Medoro was a co-founder of Silicon Biosystems S.p.A. and has served as its Chief Scientific Officer (CSO) ever since. He is the inventor of the core technology patent on Moving DEP cages, and co-inventor in more than 10 other company patents. In Silicon Biosystems he has developed a platform for the manipulation of rare cells and is currently working to provide a non-invasive solution to the problem of prenatal diagnosis. This company has already closed a first round with angel investors. His research interests are in various aspects of Lab-on-a-chip modeling and design, including microfluidics and dielectrophoretic cell manipulation. He is co-author of more than 40 scientific papers in the field of lab-on-a-chip for cell analysis among which "A CMOS Chip for Individual Cell Manipulation and Detection" received the IEEE 2003 International Solid Circuit Conference Jan Van Vessem Award for outstanding European papers.

Roberto Gambari received the Dr. Biological Sciences degree from the University of Rome in 1975. After working at the Cancer Center of Columbia University as PhD fellow, he joined the University of Rome. He is as a Full Professor of Biochemistry at University of Ferrara since 2001, and currently is a chairman of the

PhD course in Biotechnology and Director of the Biotechnology Center of Ferrara University, and a Director of the ThalLab (Laboratory for the Pharmacological and Pharmacogenomic Therapy of Thalassemia). His research interest are on various aspect of molecular biology, genetic medicine, gene therapy, pharmacogenomics with the aim of developing novel therapeutic and diagnostic approaches of thalassemia, neoplastic diseases, cystic fibrosis. For the first time, his group has demonstrated a transcription factor decoy activity of PNA-DNA chimeras. His group identified several inducers of fetal hemoglobin (HbF) for possible use in the therapy of thalassemia. Recently, his group has conducted experiments on non invasive prenatal diagnosis of genetic diseases using biosensors and Lab-on-a-chip technologies. More than 220 papers, published by Roberto Gambari, are present in Medline. He is a co-author of more than 8 patents, including one on the applications of Lab-on-a-chip platforms to detect highly lytic cytotoxic T-lymphocytes.

7 CMOS ELECTRONIC MICROARRAYS IN DIAGNOSTICS AND NANOTECHNOLOGY

Dalibor Hodko[1], Paul Swanson[1], Dietrich Dehlinger[2], Benjamin Sullivan[3] and Michael J. Heller[2,3]*

[1] Nanogen, San Diego, CA 92121
[2] Department of Electrical and Computer Engineering, University of California, San Diego
[3] Department of Bioengineering, University of California, San Diego
*mheller@bioeng.ucsd.edu

7.1 INTRODUCTION

In the past decade, the development of microarray technology has greatly expanded our analytical capabilities for DNA and protein analysis and diagnostics [1]. A variety of new techniques now allows the simultaneous analysis of large numbers of DNA target sequences in small sample volumes to be carried out with both high selectivity and sensitivity. Some examples include Affymetrix's GeneChip™ [2-4], Nanosphere's gold nanoparticle technology [5], and Nanogen's electronically active Nanochip® [6] technologies. For DNA hybridization, assays have been developed for gene expression analysis [7], short tandem repeats (STRs) for forensics [8], single nucleotide polymorphisms (SNPs) for cancer and genetic disease analysis [9], and a variety of other formats for various research and diagnostic applications.

Other non-microarray based DNA detection systems have also been developed and are now used for applications such as the detection of bio-warfare agents and other pathogens. Some of these include Idaho Technology's Ruggedized Advanced Pathogen Identification Device (R.A.P.I.D System) and Rapid Cycler systems, Autonomous Pathogen Detection System (APDS) developed at the Laurence Livermore National Laboratories [10-12], and

Cepheid's Smart Cycler and GeneExpert systems [13-15] that are capable of integrating on-chip lysis of microorganisms, amplification of their characteristic DNA using the polymerase chain reaction (PCR) and subsequent fluorescence detection of DNA. Although most of these systems offer solutions to detection of a limited number of gene target sequences, the number of optical channels available somewhat limits their application when a larger number of agents or genes need to be detected. One of the major advantages of microarray type platforms is that they allow the multiplexed detection of a relatively larger number of pathogens or analytes and multiple gene characterization. More recently, a portable electronic microarray system has been developed (based on the Nanochip® microarray) which accommodates a CMOS 400 site microarray and uses fluorescence based detection to identify target DNA sequences [16].

Electronic microarray technology is fundamentally different from more conventional or passive DNA microarrays because it uses electric fields to electrophoretically transport the target DNA and/or DNA probe molecules to specific sites on the microarray. This transport over the microarray surface is electronically controlled by activating the appropriate geometry of microelectrodes (Fig. 7.1). The electronic addressing of DNA as well as other biomolecules (proteins, antibodies, etc.) can accelerate molecular binding interactions (hybridization, biotin/streptavidin ligation) more than 1,000 times faster than the passive binding. For example, DNA hybridization on passive microarrays may take up to several hours or more when low concentrations of target DNA molecules are being analyzed. On an electronic microarray the hybridization can be completed in less than one minute. Other advantages of electronic microarrays are that they can be reused, analyze multiple samples, and analyze more than one patient sample on the same array. This is possible because the hybridization of an analyte (patient sample) which occurs in seconds, will only take place at the specific test sites on the array which have been electronically activated. Other DNA molecules present in solution do not passively hybridize because of the short exposure time and special buffer solution which does not promote passive hybridization. The newest version of electronic microarray is a CMOS controlled array device with 400 platinum electrodes (Nanochip®), produced by Nanogen (San Diego, CA). The electrodes are 50 μm in diameter and each is independently controlled and monitored by CMOS (complementary metal oxide semiconductor) circuitry built into the underlying silicon structure of the chip. Another important feature of electronic microarrays is their permeation layer. This is a thin, hydrogel layer containing co-polymerized streptavidin which covers the surface of the microarray including the elec-

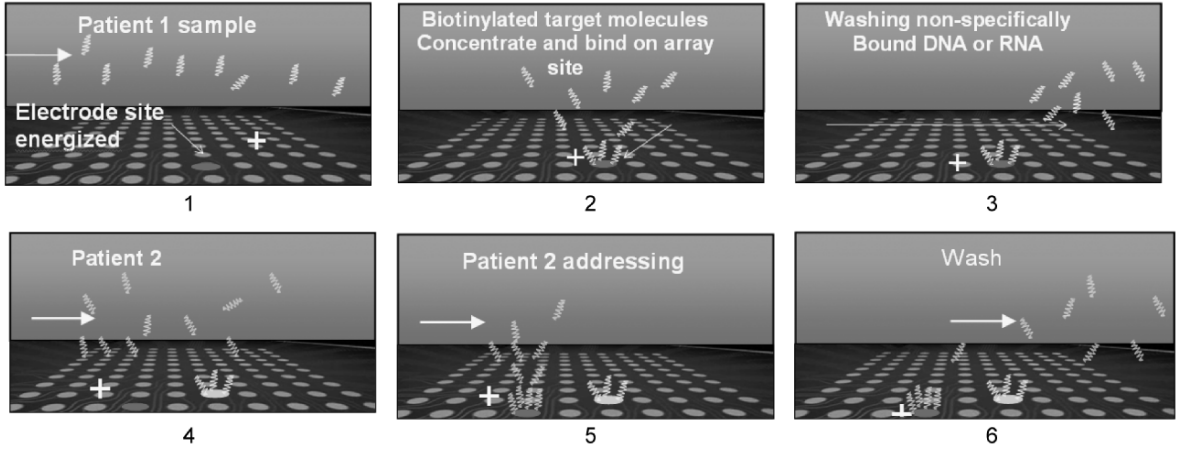


Figure 7.1 Schematic presentation of multiplexed addressing of nucleic acid samples from multiple analyte or patient sample and for multiple genes on the electronic microarray. The electric field applied (1) attracts negatively charged nucleic acid samples (e.g., DNA from patient 1) to a positively charged pad, or electrode. The electrode array is covered with a hydrogel permeation layer containing streptavidin which serves to bind the biotinylated sample or oligonucleotide probes. Addressing even very small concentrations of DNA is enabled by electronic concentration on the array site that occurs within 1 – 2 minutes. Once the sample is bound to the pad, e.g,. from patient 1, non-specifically bound nucleic acids are hydraulically washed from the chip (3). A second sample (patient 2) is introduced and addressed to a different pad or a series of pads (4 and 5). Excess or non-specifically bound DNAs are washed away (6). Oligonucleotide capture probes or reporter probes can be hybridized in the same way. This unique method allows a highly precise spatial control of addressing probes or amplicons onto a desired location on the array. Because passive hybridization usually occurs within hours (as in other competitive passive microarray technologies, e.g., Gene Chip), no binding occurs on non-electronically addressed pads. Fluorescent oligonucleotide reporters or discriminators could be incorporated in the assay design which allows for the determination of large number of genes on the array. The entire hybridization and fluorescent detection on the array is performed within minutes.

trodes. The permeation layer provides binding sites for biotin labeled DNA target and probe molecules, and also protects the DNA molecules from the adverse effects of the electrolysis products. To date, two commercial instruments have been produced by Nanogen for DNA hybridization research and diagnostic applications; the first was the Molecular Biology Workstation and more recently, the Nanochip® 400 System [Figs. 7.2(a) and (b)]. The earlier system utilized the 100 test site electronic microarray incorporated into a specialized cartridge which provided the fluidics, electrical contacts and an optical window [Figs. 7.3(a) and (b)]. The newer system utilizes the 400 test site CMOS controlled electronic microarray which is incorporated into an even smaller cartridge which provides fluidics, electrical contacts and an optical window [Figs. 7.4(a) and (b)]. Both of these systems allow the use of electronic, thermal and chemical techniques to control hybridization stringency. The instrument systems provide a relatively flexible platform that allows several types of assays to be formatted for multiplexed DNA hybridization analyses. For example, the determination of multiple genes in one sample, multiple samples with one gene, or multiple samples with multiple genes. The ability to electronically control individual test sites on the microarray permits biochemically unrelated molecules to be used simultaneously on the same array. This is not possible on a conventional DNA array where each test site cannot be controlled separately. Both the 100 test site and 400 test site electronic microarray systems use fluorescence based detection for the monitoring and analysis of the DNA hybridization reactions.

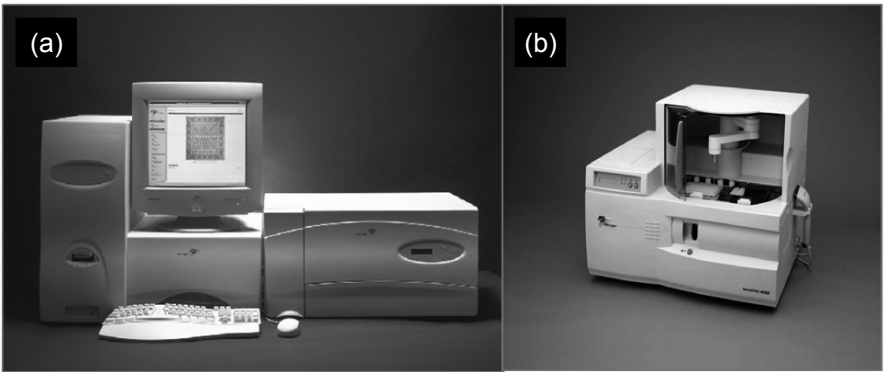


Figure 7.2 (a) The original Molecular Biology Workstation. (b) The newer Nanochip® 400 System. These are both open systems that uniquely combine microfluidics-based solution handling using a robotized arm with highly controlled electrophoretic addressing of biomolecules on the array. The user has fluidics inputs through standard microtiter plates (96 or 384 well plates) and/or specific reagent packages and reservoirs for buffer solutions mounted within the instrument.

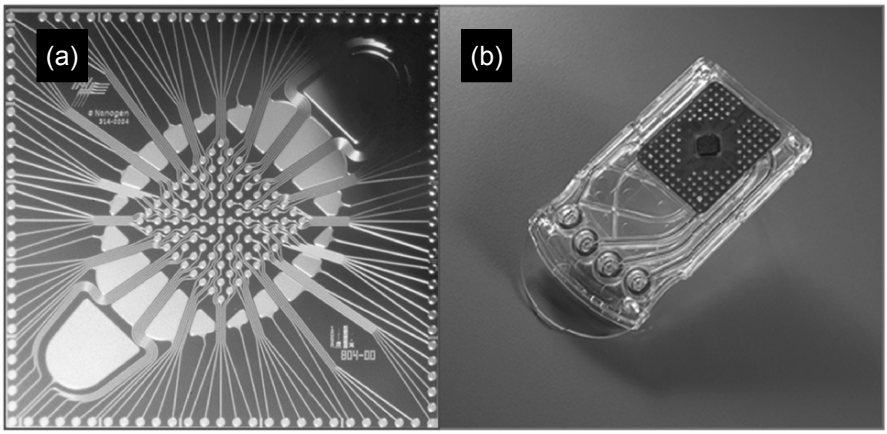


Figure 7.3 (a) The directly wired 100 site microarray device with an inner set of 80 μm diameter platinum microelectrodes and an outside set of auxiliary microelectrodes. (b) The 100 test site electronic microarray NanoChipTM cartridge which provides the electronic, optical, and fluidic interfacing.

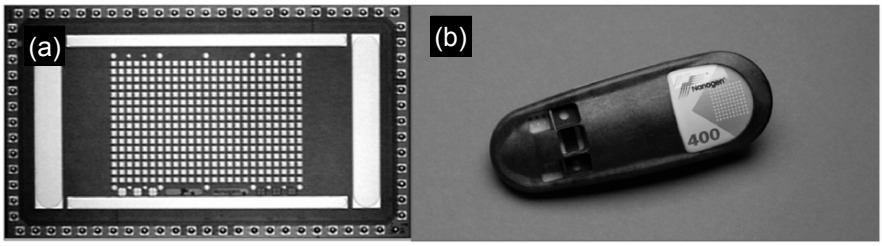


Figure 7.4 (a) The 400 test site CMOS electronic microarray device with 400 array microelectrodes (54 μm diameter) and four perimeter counter electrodes; the device is approximately 5 mm by 7 mm in size. (b) The NanoChip 400 plastic cartridge housing which provides the fluidic delivery system and outside electrical connections.

Generally, either the DNA target sequence or the oligonucleotide reporter probe is labeled with a suitable fluorophore group. While electronic microarrays were originally developed for DNA related bioresearch and clinical diagnostic applications, these active devices also demonstrated the ability to rapidly transport and bind a variety of other charged molecules and entities. Some of these included proteins (antibodies, enzymes), nanostructures, cells and even micron-scale structures. When DNA hybridization is carried out on these electronic microarrays, the device is actually using the electric fields to "direct" the self-assembly of the DNA hybrid at the test site. In principle, these active microarray devices can serve as "motherboards or hostboards" that can assist in the self-assembly of DNA molecules, as well as other moieties such as nanostructures or even microscale components. Active microarray electric field directed assembly represents a type of "Pick & Place" process that has the potential to be used for the heterogeneous integration and nanofabrication of molecular and nanoscale components into higher order materials, structures and devices.

Thus, electronic microarray technology which provides rapid transport and hybridization of DNA sequences to a specific location on the chip, offers a number of advantages over passive hybridization [17-25]. A completely automated system with laser based fluorescence detection allows DNA hybridization research and diagnostic assays to be carried out in a number of useful formats. New applications of this technology may even provide processes for the nanofabrication and nano-manufacturing of higher order structures from molecular and nanoscale components [26-32].

7.2 ELECTRONIC MICROARRAYS

7.2.1 Direct Wired Microarrays

The development of electronic microarrays started with early stage prototype devices with 25 directly wired microelectrodes, and progressed to the 100 microelectrode array device. The 100 site device was the first to be commercialized by Nanogen. The 100 site microarray device has an inner set of 80 μm diameter underlying platinum microelectrodes, and an outside set of auxiliary microelectrodes [Fig. 7.3(a)]. The outer group of electrodes can be biased negative, which allows DNA in the bulk sample solution to be concentrated at the specific internal test-sites (biased positive). Each microelectrode has an individual wire interconnect through which current and

voltage are applied and regulated. The 100 site microarray is about 7 mm^2 in size, with an active test site array area of about 2 mm^2. The microarrays are fabricated from silicon wafers, with insulating layers of silicon dioxide, platinum microelectrodes and gold connecting wires. Silicon dioxide/silicon nitride is used to insulate the conducting wires, but does not cover the surface of the platinum microelectrodes. The whole microarray surface is covered with several microns of a hydrogel (agarose or polyacrylamide) which serves as permeation layer. The permeation layer is impregnated with a coupling agent (streptavidin), which allows attachment of biotinylated DNA probes or other entities. The permeation layer allows current (counter ions) to flow to and from the microelectrode surface, and also serves to protect the DNA molecules from the adverse effects of the electrolysis products (H^+, OH^-, O_2, H_2, and free radicals), which are produced on the electrode surfaces when activated. (In order to produce the electrophoretic transport of DNA molecules in the bulk solution, voltages greater than 1.2 V must be used which cause the electrolysis of water to occur on the surface of the platinum microelectrodes). The 100 test site electronic microarray device has been incorporated into a cartridge package (NanoChip™ cartridge) which provides for the electronic, optical, and fluidic interfacing [Fig. 7.3(b)]. The chip itself is mounted (flip chip bonded) onto a ceramic plate and pinned out for the electrical connections. The chip/ceramic plate component is mounted into a plastic cartridge that provides several fluidic input and output ports for addition and removal of DNA samples and reagents. The area over the active test site portion of the array is an enclosed sample chamber covered with a quartz glass window. This window allows for fluorescent detection to be carried out on the hybridization reactions that occur at the test sites on the array surface. The complete instrument system (NanoChip™ Molecular Biology Workstation) for the 100 site microarray provides a chip loader component, fluorescent detection/reader component, computer control interface and data display screen component [Fig. 7.2(a)]. The probe loading component allows DNA probes or target sequences (DNA, RNA, PCR amplicons) to be selectively addressed to the array test sites, and provides a "make your own chip" capability. The automated probe loader system allows four 100 site NanoChip™ cartridges to be loaded with DNA probes or DNA samples from ninety-six or three hundred eighty four well microtiter plates. The DNA probes or target sequences are usually biotinylated, which allows them to become bound to streptavidin within the permeation layer of the specified test-site. In the electronic addressing procedure the probe loader component delivers the desired biotinylated probe to specified test sites which are biased positive. The electric (electrophoretic) field causes

the negatively charged DNA molecules to concentrate onto the positively activated test sites, with subsequent binding via the biotin/streptavidin reaction.

7.2.2 CMOS Microarrays

The use of individual wires to externally control the voltages and currents to a large number of microelectrodes becomes less viable for higher density arrays with more closely spaced and smaller microelectrodes. Thus, for higher density microarrays (>100 test sites) it is advantageous to integrate the microelectrode control circuitry directly into the microchip silicon structure itself. To accomplish this, prototype 400 site microarrays were designed and developed with integrated CMOS circuitry [33]. These first 400 site CMOS integrated microarrays used a current mirror circuit at each electrode in the array to produce an externally defined current at each enabled electrode (Figure 7.5(a) shows an original prototype 400 site CMOS device; and Figure 7.5(b) shows an enlarged image of the 54 μm microelectrodes and some of the underlying CMOS circuitry). The microelectrodes were enabled by switches at each row and column that controlled which current mirrors were connected to the +5 V rail and the ground rail. The switches on the rows controlled access to the +5 V and the switches on the columns controlled access to the ground. This enabling circuitry allowed any group of electrodes to be turned on, but did not allow any random pattern of electrodes to be turned on. Limitations of this design included the lack of ability to measure the current and voltage at each electrode, to determine if a current mirror had reached a rail voltage and to source the desired current. Nevertheless, in addition to the 400 site CMOS microarray two

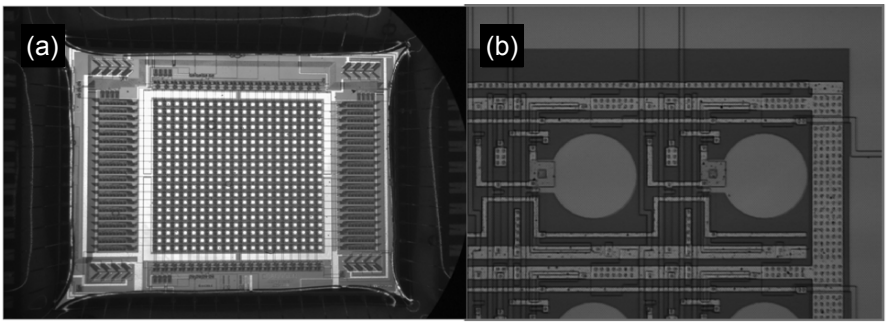


Figure 7.5 (a) A prototype 400 site CMOS electronic microarray which used current mirror circuits at each electrode. (b) An enlarged view of the 54 micron microelectrodes and the underlying CMOS circuitry.

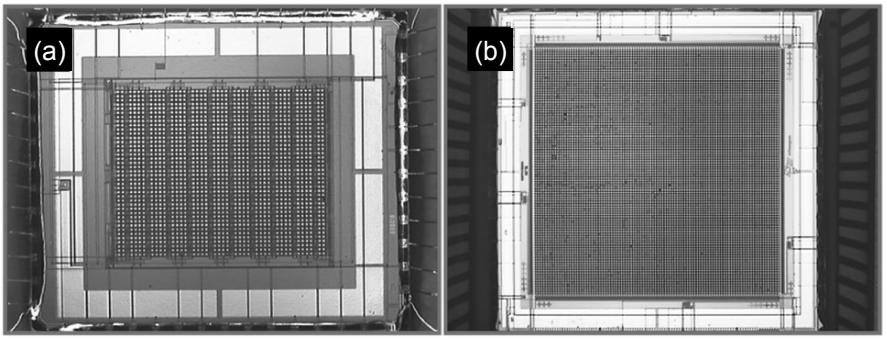


Figure 7.6 (a) A prototype 1600 site CMOS electronic microarray. (b) A prototype 10,000 site CMOS electronic microarray.

other prototype devices, a 1600 site microarray and a 10,000 site microarray were also fabricated [Figs. 7.6(a) and (b)]. These early CMOS microarrays were used for internal projects, collaborations, and government grants.

The next generation CMOS electronic microarray devices were designed with feedback loops to provide independent control of the biasing at each microelectrode. These devices could measure the current and voltage at each enabled electrode, allowing the controlling software to adjust the bias at each location to the desired value. Each microelectrode could be biased with a sample and hold circuit, which is periodically updated by a digital to analog converter and a dual port memory. CMOS switches controlled by AND gates tied to row and column indicators cycle through each electrode measuring the voltage and current at each location with a single analog to digital converter. An EE PROM and a p-n junction based thermal sensor were also added to the chip. Fabrication of the platinum microelectrodes above the CMOS circuitry initially caused some problems, but these were ultimately solved. Standard CMOS metallization does not generally make electrodes resistant to electrochemistry. For example, positively biased aluminum or titanium will irreversibly oxidize in an aqueous buffer, creating a steadily increasing charge transfer resistance at the surface of the electrode. A strong positive current on a gold electrode will cause the gold to ionize and dissolve into the solution. Platinum is most ideal for electrochemistry, however few CMOS foundries are willing to use this material in their standard CMOS process. Thus, it was necessary to pre-process the CMOS wafers at a separate facility, and then fabricate the upper level chip structure and platinum electrodes in a completely separate process. Figure 7.7 shows a cross section of the CMOS wafer as received from the foundry. The contact pads are exposed for testing, however the electrode array contacts remains buried under the final CMOS passivation layers of silicon dioxide and sili-

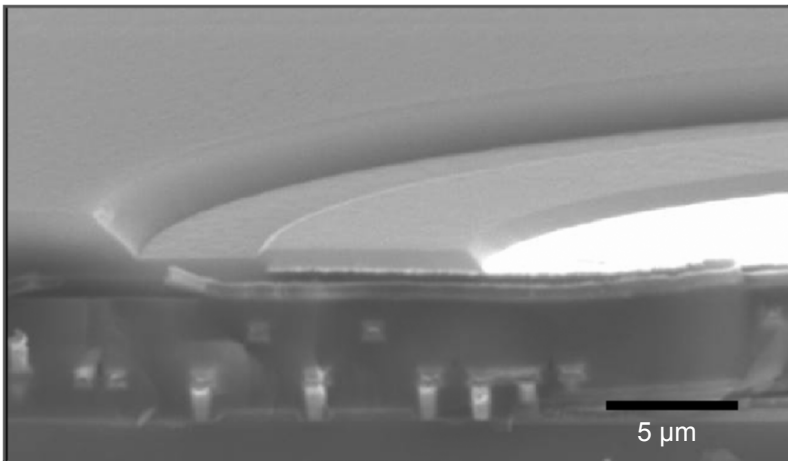


Figure 7.7 A cross-section through a 400 site CMOS microarray showing the platinum microelectrode and the underlying CMOS structure.

con nitride. After patterning round holes in photoresist over the electrode contacts, the silicon nitride is dry etched in a Reactive Ion Etcher (RIE) and the silicon oxide is wet etched in a Buffered Oxide Etch (BOE) (Fig. 7.8). The photoresist is stripped off and new photoresist is deposited and patterned with smaller holes over the electrodes. This photoresist is used as a stencil to pattern a sputtered Titanium/Tungsten and platinum. The edge of this photoresist employs an overhang to insure that the metallization being lifted-off is not connected to the metallization to remain. After the deposition and lift-off, a layer of silicon dioxide is deposited over the entire surface of the wafer using Plasma Enhanced Chemical Vapor Deposition (PECVD). A final photoresist is deposited and patterned with holes that will define the electrode surface available for electrochemistry. The pre-deposited PECVD SiO_2 is then wet etched with BOE down the platinum electrodes through the holes in the photo resist. The present CMOS microarray device consists of a 16 by 25 array of 50 µm diameter microelectrodes spaced 150 µm center to center. Figure 7.4(a) shows the 400 test site CMOS electronic microarray device which is approximately 5 mm by 7 mm in size.

In addition to "on chip" voltage/current sourcing and biasing, integrated CMOS circuitry also provides voltage/current measuring circuits, digital communication, memory and temperature sensing. The ability to monitor the voltage and current at each electrode provides an effective diagnostic tool for detecting buffer contamination. If a transport buffer is contaminated with salt, the current and voltage at the electrode will deviate from an established norm. Having a calibrated thermal sensor on chip provides

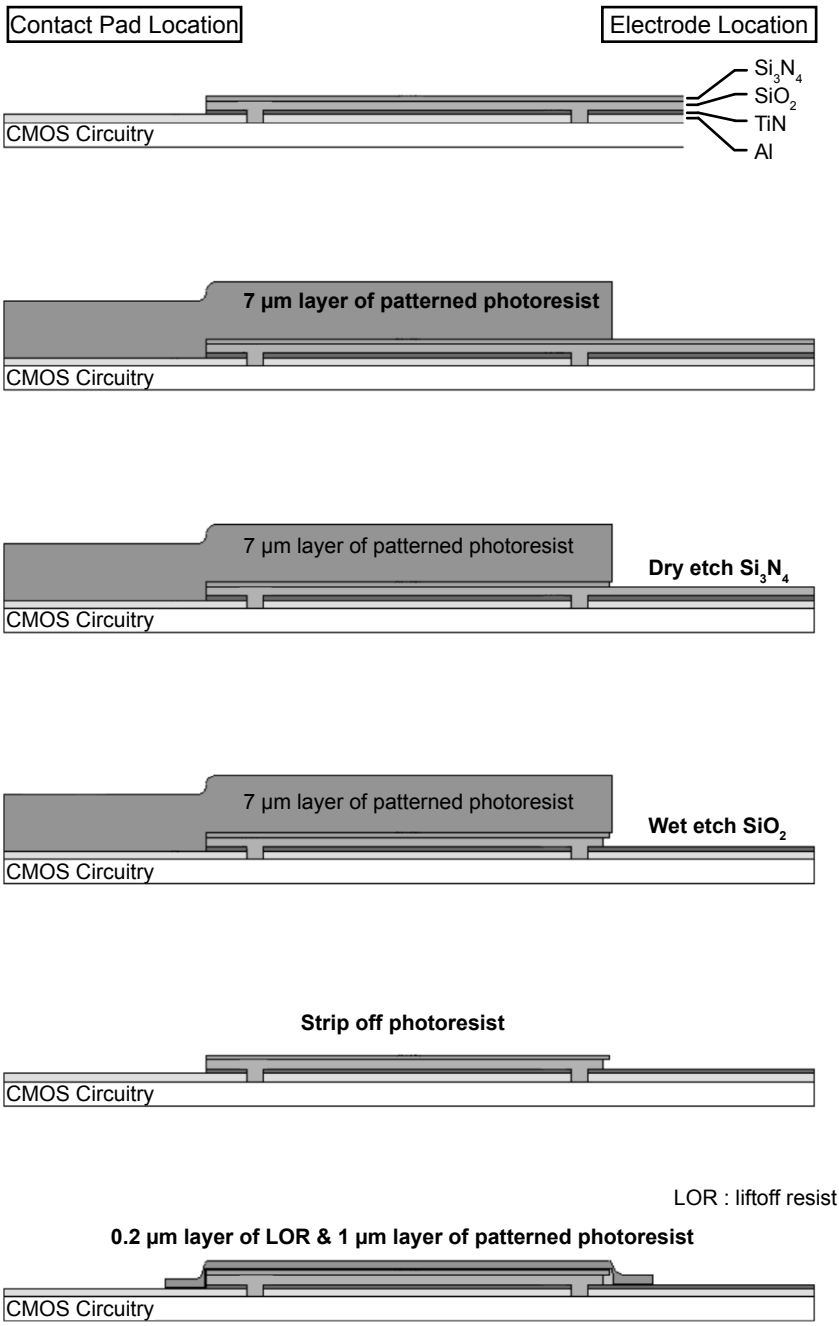


Figure 7.8 Patterning steps used in the fabrication of the platinum microelectrode structure over the CMOS circuitry.

accurate robust temperature control, minimizing the effects of variation of thermal transfer coefficients. The on-chip memory allows history of the multiple-use array to travel with the cartridge (chip and fluidic assembly). The 400 site CMOS chip has only 12 external electrical connections, +5 V and ground for the digital circuits, +5 V and ground for the analog circuits, digital signal in, digital signal out, clock signal, reset, and two terminals for an external current sampling resistor. For structural support 76 flip-chip solder bonds are used between the chip and a ceramic substrate which doubles as both a fluidic chamber and as an electrical contact interface. The CMOS process requires 16 masking steps followed by an additional 3 masking steps for the platinum electrodes. After the finished wafers have been inspected for defects, the wafers are diced into individual chips then flip-chip bonded onto the ceramic substrates. The Flip-Chip on Substrates (FCOS) are thermally calibrated and then finally assembled into the NanoChip 400 plastic cartridge housing which provides the fluidic delivery system and outside electrical connections [Fig. 7.4(b)]. The newer Nanochip® 400 System integrates both the loader and the reader into a single instrument [Fig. 7.2(a)]. The Nanochip® 400 System and 400 CMOS electronic microarray provide a tremendous amount of flexibility and control, each of the 400 test sites on the microarray can be easily configured and modified for a range of electronic DNA hybridization assays formats.

7.3 ELECTRONIC TRANSPORT AND HYBRIDIZATION OF DNA

The electronic control of the DNA and other molecular transport over the electronic microarray offers unique conditions of accelerating the concentration of these molecules as well as confining the hybridization exclusively at the addressed test site on the microarray. With the use of special low conductance buffers, such as histidine, the actual DNA hybridization will now only occur at those test sites on the microarray which have been biased positive. This phenomena is due indirectly to the electrochemically generated lower pH conditions which is produced at the positively charged microelectrodes. In order to carry out electronic (electrophoretic) addressing of DNA target and probe molecules to the microarray test sites, the bias potential at the platinum microelectrodes must be sufficiently high enough to cause electrolysis of water occur, i.e., water oxidation at the positive electrode and water reduction at the negative electrode [Eqs. (7.1) and (7.2)].

Oxidation $H_2O \rightarrow 2H^+ + \frac{1}{2}O_2 + 2e^-$ (7.1)

Reduction $2e^- + 2H_2O \rightarrow 2OH^- + H_2$ (7.2)

The products of water oxidation at the positive electrode are hydrogen ion (H^+) and molecular oxygen (O_2). When histidine is used as a zwitterionic buffer to carry out DNA addressing and hybridization, it not only allows rapid transport of DNA molecules, but also promotes hybridization in the local micro-environment around the positively biased electrode. This occurs because upon buffering the acid conditions at the microelectrode, the zwitterionic histidine becomes protonated and positively charged. This positively charged histidine is now able to neutralize some of the negatively charged phosphate groups on the backbone of the DNA, thus promoting and stabilizing DNA hybridization. The electric field can also be used to aid in hybridization specificity by reversing the polarity at the electrode (negative activation). In this case, any non-specifically bound (negatively charged) DNA molecules are driven away from the vicinity of the negatively charged electrode. This ability to bias an electrode (or series of electrodes) positively to promote the hybridization and then negatively to reduce non-specific binding provides a unique tool to improve DNA hybridization specificity. This "electronic stringency" enables single base differentiation and determination of single nucleotide polymorphisms (SNPs) to be carried out more effectively. Electronic microarrays provide an open platform for flexibility in the design of DNA hybridization assays for research or clinical diagnostic applications. For example, assays can be designed in either a DNA target (PCR amplicon) down format to screen one or more patients for one or more single nucleotide polymorphisms (SNPs), or in a DNA capture probe down format which allows one to screen many patients for one or more SNPs. In addition, sandwich-type assays are easy to perform where oligonucleotide discriminators and/or fluorescent probe labeled oligonucleotides are hybridized, electronically or passively, to captured amplicons or probes (Fig. 7.9). These assay design tools enable the researcher or clinician to increase the discrimination at a single base resolution as well as minimize the non-specific binding. Multiple probes or discriminators can be attached at a single site which facilitates the detection of multiple DNA targets on a single test site. In multiple gene detection, blocker oligonucleotides have been used that block SNPs which are not reported at a particular location. This allows a common universal reporter probe and specific discriminators to be used for reporting one or more SNPs on another site. The Nanochip® 400 System provides two lasers and a detection system which is able to recognize fluorescent signal from two different reporter dyes [Fig. 7.4(a)]. All

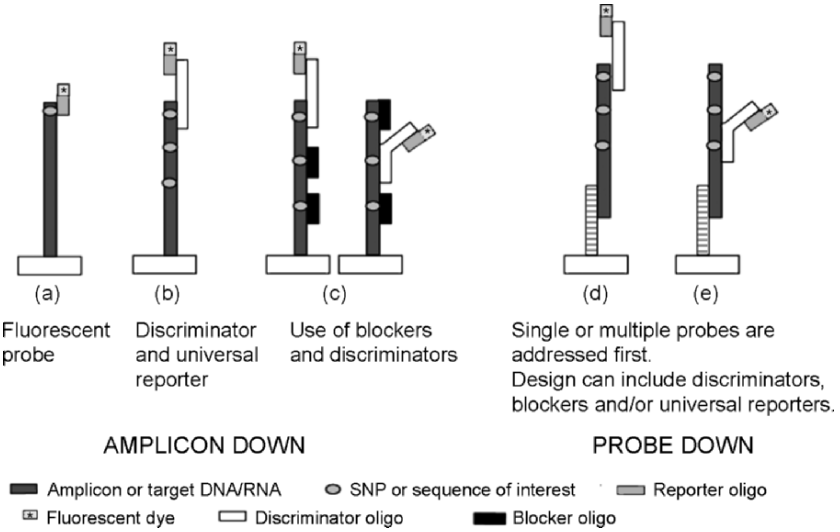


Figure 7.9 Demonstration of how the electronic microarray platform allows versatility in DNA hybridization assay design. Two basic designs include PCR amplicons (target DNA) down and oligonucleotide probe down. Amplicons are biotinylated in the amplification reactions and electronically addressed onto the array electrode sites where they remain bound within the hydrogel permeation layer containing streptavidin.

the steps are extremely fast so the users can design and generate their own assays and incorporate the preparation of the array with oligos specific to multiple targets and samples as a part of the assay. This is different from the approach used in high-throughput arrays where the oligo probes are pre-addressed. However, target specific oligos can also be pre-addressed on the electronic microarray.

7.4 NANOFABRICATION USING CMOS MICROARRAYS

One of the grand challenges in nanotechnology is the development of new nanofabrication methods that will lead to viable cost effective nano-manufacturing processes. In addition to top-down processes such as photolithography, so called bottom-up processes are being developed for carrying out hierarchical self-assembly of nanostructures into higher-order materials structures and devices [34-39]. To this end, considerable efforts have been carried out on Layer-by-Layer (LBL) self-assembly nanofabrication processes as a way to make three dimensional layered nanostructures which

have macroscopic x-y dimensions [40-42]. A variety of techniques have been designed and used to carry out passive LBL self-assembly processes. Generally, most processes involve dipping an activated substrate material (electrostatic charged, chemically reactive, etc) into a solution which contains a macromolecule, polymer, or nanoparticle with an opposite charge or reactivity to the substrate material. The reaction is given some time to complete, the substrate is washed extensively and then the next layering step is carried out [43-64]. When an x-y patterned structure is required, a pre-patterned substrate material must be used [65, 66]. Generally, LBL processes are relatively slow and non-parallel unless the substrate material is first patterned using a lithographic or other imprinting process. Such limitations of LBL and other self-assembly processes provide considerable incentive to discover other paradigms for nanofabrication and heterogeneous integration processes [67, 68]. To this end, electric filed assisted assembly devices and processes exhibited promise as such a technology [69, 26-32].

As was discussed in the earlier section, sophisticated CMOS electronic microarray devices produced by top-down photolithography processes have been developed and commercialized for DNA hybridization diagnostic applications. These devices are able to rapidly transport and accelerate the hybridization (binding) of the DNA molecules to complementary DNA sequences immobilized at specific locations on the microarray surface. Thus, these 'top-down" devices are able to control, assist and accelerate the self-assembly "bottom-up" process of DNA hybridization occurring on the surface of the microarray. In addition to DNA hybridization reactions, these devices are also able to rapidly accelerate protein/ligand binding reactions such as streptavidin/biotin and antibody/antigen by vastly increasing the local concentration at the test-site. In addition to DNA and immunodiagnostic applications, these devices have also demonstrated the ability to control the transport and positioning of almost any type of charged nanostructure or micron sized structure on the microarray surface. This includes most types of polymeric or metallic nanoparticles, quantum dots, cells, and even micron sized semiconductor components, such as 20 μm light emitting diodes (LEDs) [70-77]. The new 400 test-site CMOS microelectronic array device and the CMOS microarray controller system now provide a relatively inexpensive and compact control unit that can be run with a lap top computer. In short, these reconfigurable CMOS microarray systems can serve as a host nanofabrication platform for a highly parallel electric field "Pick & Place" or assisted assembly process that allows almost any charged molecule (DNA, protein, etc.) or nanostructure (quantum dots, nanoparticles,

etc.) to be hierarchically organized into two and three dimensional structures on the array surface.

7.4.1 Electric Field Directed Nanoparticle Assembly Process

In order to use the 400 site CMOS array device as a nano-manufacturing platform, standard procedures were developed to determine the optimal parameters for parallel 3D nanoparticle layering. The CMOS array device is initially prepared by first washing it several times with ultra-purified water to remove a protective carbohydrate layer from the permeation layer. After washing, the permeation layer surface is reacted with a 20 µl of a 2 µM biotin dextran solution for 30 minutes. Figure 7.10 shows the biotin dextran layer deposited on the permeation layer. The array is then washed with a 100 mM L-Histidine solution. To test various addressing (deposition) conditions, the microarray device is programmed to be activated in columns with currents varying from 0.025 µA to 0.4 µA in 0.025 µA increments. Since the nanoparticles used in most experiments have a net negative charge, the electrodes at the desired addressing sites on the array are biased positive, and the larger counter electrodes on the perimeter of the device are biased negative. Typically a group of alternating columns on the array is activated in parallel at the different current levels and addressing times, while the intervening columns of electrodes are not activated and thus served as negative controls for the nanoparticle addressing and binding process. Array

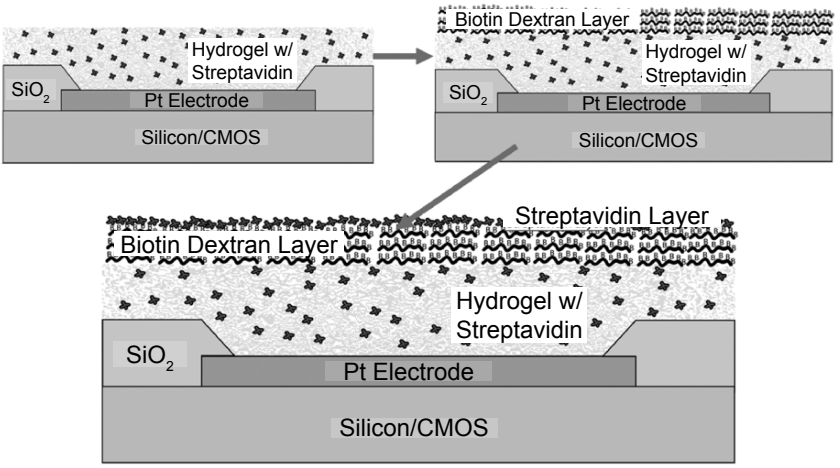


Figure 7.10 Representation of the biotin dextran layer deposited on the polyacrylamide/streptavidin permeation layer with the underlying platinum electrode.

addressing is usually carried out using 10 µl of 100 mM L-Histidine buffer, containing from 1-10 nM of the derivatized nanoparticles. The two types of derivatized nanoparticles used in the nanoparticle layering experiments being described were 40 nm red fluorescent polystyrene nanoparticles derivatized with streptavidin (Ex 580 nm, Em 605 nm) and 40 nm yellow-green fluorescent polystyrene nanoparticles derivatized with biotin (Ex 505 nm, Em 515 nm). The biotin-streptavidin ligand binding reaction allows the two different types of fluorescent nanoparticles to be bound to each other, but nanoparticles of the same type do not bind to each other.

The nanoparticle addressing, binding and layering experiments are carried out as follows: (1) First Nanoparticle Layer - About 20 µl of a 10 nM solution of 40 nm streptavidin nanoparticles (red fluorescence) in 100 mM L-Histidine is placed on the microarray and the selected columns of electrodes are activated at the different current levels (0.025 µA - 0.4 µA), with addressing times of 5, 15 and 30 seconds. (2) Wash - The array is then immediately washed (manually) three times with 100 mM L-Histidine, which usually takes less than one minute. Epifluorescence (Ex 580 nm, Em 605 nm) monitoring of the array is carried out during the process. (3) Second Nanoparticle Layer - About 20 µl of a 10 nM solution of 40 nm biotin nanoparticles (green fluorescence) in 100 mM L-Histidine is now placed on the array and selected electrodes were activated (currents form 0.025-0.4 µA, with addressing times of 5, 15 and 30 seconds). Epifluorescence (Ex 505 nm, Em 515 nm) monitoring of the array is carried out during the process. (4) Wash - The array is now washed three times with 100 mM L-Histidine. (5) Successive Nanoparticle Layers - Steps 1-4 are now repeated to achieve desired number of nanoparticle layers, and epifluorescence (red and green) monitoring of the array is carried during the whole process. (6) Final Wash - The array is finally washed several times with deionized water to remove L-Histidine. Figure 7.11 shows a diagram of a multilayered structure composed of alternating layers of biotin and streptavidin nanoparticles. Using the process described above, about forty to fifty alternating addressing and depositions of the 40 nm streptavidin nanoparticles (red fluorescence) and 40 nm biotin nanoparticles (green fluorescence) can be carried out in less than one hour. To directly view the resulting layers, the array device with the layered nanostructures can be sputtered with a 10 nm layer of gold to create a thin conductive layer. The array can then be imaged using a scanning electron microscopy (SEM). The 40 nm nanoparticles are easily resolvable by SEM and the nature of the deposited material can be observed.

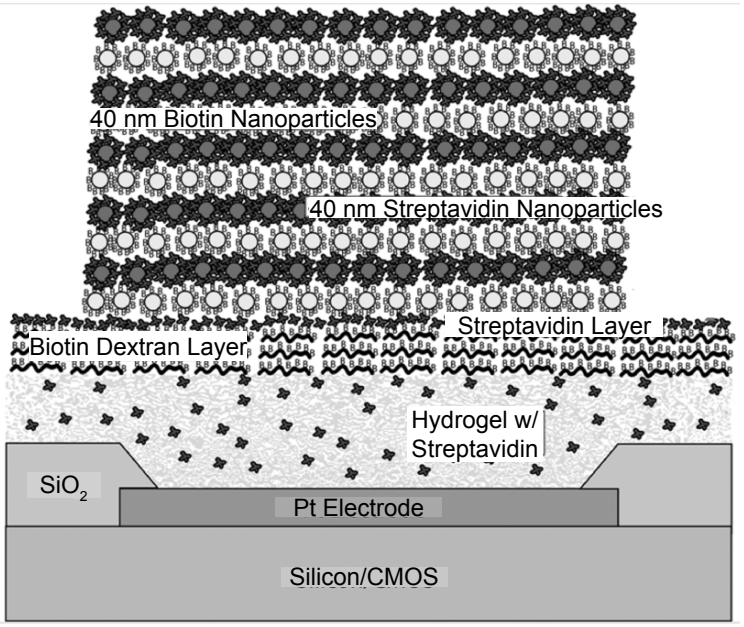


Figure 7.11 Diagram of a multilayered structure composed of alternating layers of biotin and streptavidin nanoparticles.

In order to determine optimal conditions for derivatized nanoparticle layering, experiments are carried out at addressing times of 5 seconds, 15 seconds and for 30 seconds. For each of the addressing time experiments, ten columns (16 sites) are activated with DC current levels that ranged from 0.025 μA to 0.4 μA, in increments of 0.025 μA. The activation of all 160 sites (at different current levels) can be carried out in parallel. For each addressing time experiment (5 seconds, 15 seconds and 30 seconds) the addressing process is carried out forty times with alternating 40 nm red fluorescent streptavidin nanoparticles and green fluorescent biotin nanoparticles. In these experiments, the alternate columns are not activated. Figure 7.12 shows fluorescent imaging results for 39 addressings at 5 seconds [Fig. 7.12(a)], at 15 seconds [Fig. 7.12(b)] and at 30 seconds [Fig. 7.12(c)]. The inserts in each of the three photos are enlargements showing four of the 54 μm sites, which were in the rows that received 0.375 μA levels of DC current. The images show no detectable fluorescence is observed on any of the alternate sites which were not activated. By the relative fluorescent intensity of the activated sites, the best conditions for nanoparticle layering appear to be at the 5 second and 15 second addressing times in the 0.30 to 0.40 μA current level range. At the lower current levels (<0.30 μA) the overall fluorescent intensity for the layers begins to decrease. At the longer 30 second address-

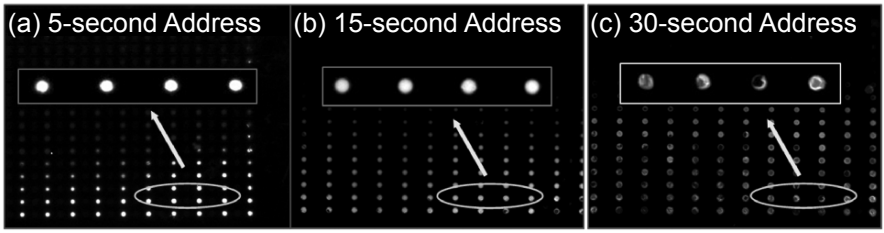


Figure 7.12 Fluorescent analysis results for highly parallel electric field 3D layering of 39 alternate layers of 40 nm red fluorescent streptavidin nanoparticles and green fluorescent biotin nanoparticles onto the 50 μm diameter test sites of a CMOS 400 site microarray device. Addressing times were 5 seconds (a), 15 seconds (b), and 30 seconds (c) at DC current levels that ranged from 0.025 μA to 0.40 μA. Inserts are enlargements of a small section of the microarray each showing the results for 0.375 μA on four test sites. Total layering time is one minute or less per layer, including the washing steps.

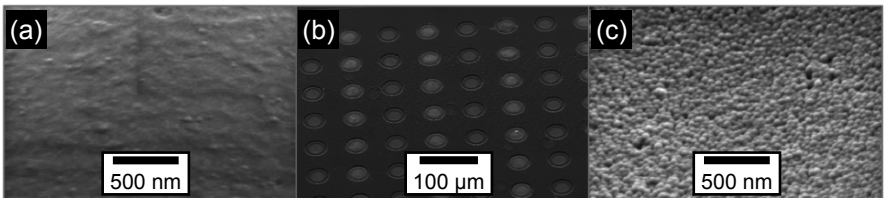


Figure 7.13 SEM images of the microarray after 40 layers of 40 nm biotin and streptavidin nanoparticles where assembled. (a) The surface of a 50 μm site which was not addressed with either type of nanoparticle. (b) A section of the microarray where every other column was addressed with nanoparticles (the brighter sites). (c) The surface of a 50 μm site which was addressed with nanoparticles, the upper surface layer of 40 nm nanoparticles is evident.

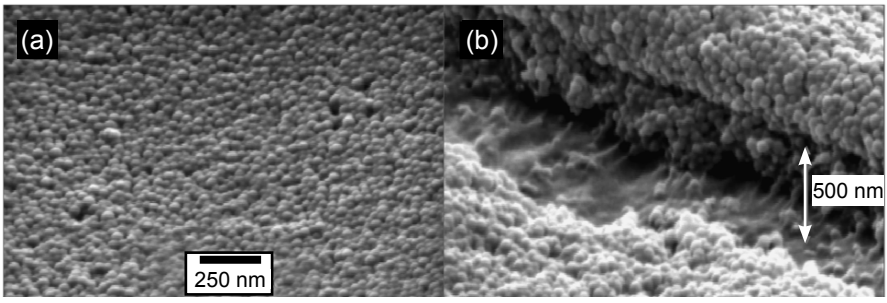


Figure 7.14 (a) The top layer of a 40 layer structure composed of alternating layers of 40 nm biotin and streptavidin nanoparticles. (b) A cross section through a 40 layer structure composed of alternating layers of 40 nm biotin and streptavidin nanoparticles.

Table 7.1 Heterogeneous Nanoconstruction Materials.

- Streptavidin
- Polyacrylamide Streptavidin
- Dextran-Biotin Polymer (10,000 MW)
- Silica Particles (~10 nm – 20 nm)
- 40 nm Nanoparticles (Biotin, Green Fluorescence)
- 40 nm Nanoparticles (Streptavidin, Red Fluorescence)
- 200 nm Nanoparticles (Streptavidin, Red Fluorescence)
- 15 nm Quantum Dots (Streptavidin, Red Emission 610 nm)
- 12 nm Quantum Dots (Biotin, Green Emission 506 nm)
- 50 nm Gold Nanoparticles (Streptavidin)
- 5'-Biotin-GAA-CAG-CTT-TGA-GGT-GCG-TG-3' (Initial Template)
- 40 nm Streptavidin Nanoparticle-5'-Biotin-GAA-CAG-CTT-TGA-GGT-GCG-TG-3'
- 40 nm Streptavidin Nanoparticle-5'-Biotin-CAC-GCA-CCT-CAA-AGC-TGT-TC-3"
- 5'-Biotin-GAA-CAG-CTT-TGA-GGT-GCG-TGT-TTG-TGC-CTG-TCC-TGG-GAG-AGA-CCG-GCG-CAC-3' (Initial Template)
- 40 nm Streptavidin Nanoparticle-5'-Biotin-GAA-CAG-CTT-TGA-GGT-GCG-TGT-TTG-TGC-CTG-TCC-TGG-GAG-AGA-CCG-GCG-CAC-3'
- 40 nm Streptavidin Nanoparticle-5'-Biotin-GTG-CGC-CGG-TCT-CTC-CCA-GGA-CAG-G-CAA-ACA-CGC-ACC-TCA-AAG-CTG-TTC-3'

ing time, the nanoparticle layers become visibly damaged. The SEM images of a microarray after assembly of 40 layers of 40 nm biotin and streptavidin nanoparticles is shown in Fig. 7.13. The first SEM image shows the surface of a 50 μm site which was not addressed (negative control) with either type of derivatized nanoparticles [Fig. 7.13(a)]. The middle SEM image shows a section of the microarray where every other column was addressed with nanoparticles, these are the brighter 50 μm sites [Fig. 7.13(b)]. The right hand SEM image shows the surface of a 50 μm site which was addressed with nanoparticles. The upper surface layer of 40 nm nanoparticles is clearly evident in this image [Fig. 7.13(c)].

Under real time epifluorescent microscope observation some of the fractured layers could actually be observed to flap when the sites were activated. Scanning electronic microscopy was used to examine the forty layer nanoparticle structures in more detail. Figure 7.14 shows SEM images of both a top layer [Fig. 7.14(a)] and a cross section of an activated site on which forty addressings of nanoparticles was carried out [Fig. 7.14(b)]. A number of nanoparticle layers can be seen from the top nanoparticle layer down to what appears to be the lower surface of the permeation layer. To

date, our work has demonstrated that a wide variety of different molecules, nanoparticles and other entities can be assembled into higher order structures using CMOS microarrays (Table 7.1).

7.5 DISCUSSION AND CONCLUSIONS

CMOS microarrays have now demonstrated the ability to not only carry out DNA hybridization diagnostics, but may also be used for nanofabrication applications. Initial work has shown that electric field assisted self-assembly of forty multiple layer nanoparticle structures can be carried out in a rapid and highly parallel format using a CMOS electronic microarray device. In this process, efficient nanoparticle addressing/deposition was achieved in 15 seconds or less. With a washing step of about 45 seconds, the total time for creating a nanoparticle layer was about one minute. Thus, the forty layer nanoparticle structures could be completed in less than one hour. The optimal electronic addressing window for creating high quality 3D layered structures appears to be at current levels in the 0.25 μA to 0.35 μA range. Because the nanoparticles used in these experiments were of the same size (40 nm), it is difficult to determine from the SEM images the homogeneity of individual biotin and streptavidin nanoparticle layers within the multilayered structures. Nevertheless, when taken together with fluorescent monitoring of depositions and the experiments showing no layering of like nanoparticles, the SEM images are consistent with a forty layer alternating nanoparticle structure. Work is in progress with different sized nanoparticles to better determine true quality of individual layers within the multilayer structures.

The results of these experiments have demonstrated a number of advantages relevant to 3D self-assembly nanofabrication using CMOS microarrays. First, it was shown that high electric field strengths and electrophoretic transport from the bulk solution could be maintained even through many layers of nanoparticles, allowing the new upper layers of nanoparticles to still be rapidly addressed to the activated sites. Second, results showed that the integrity of the biotin-streptavidin ligand binding was maintained even after forty layers of nanoparticles were addressed. This is extremely important for any 3D self-assembly process that is based on biomolecular binding using highly specific protein or DNA ligands. This ability to use high electric field strength with minimal side effects is most certainly due to the permeation layer which separates the electrode from the nanostructures, as well as to the relatively short addressing times (5-15 seconds) needed

for nanoparticle addressing and deposition. Third, because the high electric field strengths produce enormous concentration effects at the activated sites, only minimal concentrations of nanoparticles were needed for addressing and deposition. In fact, the concentration of nanoparticles used was so low that no significant non-specific binding of nanoparticles occurred during the whole process. Fourth, CMOS electronic array devices appear to allow nanoparticle layering to be carried out much faster (one to two orders of magnitude) than comparable passive Layer-By-Layer (LBL) self-assembly nanofabrication processes. Finally, with regard to x-y patterning during the fabrication process, we believe that CMOS array devices may also have advantages over lithographic and ITO patterning processes. Lithographic patterning of the substrate and subsequent layers would be a relatively time consuming process, while electronic microarray patterning could be carried almost instantaneously. In the case of ITO (optical) patterning, the patterning step could be rapid, but ITO substrates are not stable at the high DC electric fields needed for rapid addressing and deposition. Overall, the use of CMOS electronic microarray devices for assisted self-assembly represents a unique example of combining "top-down" and "bottom-up" technologies into a potentially useful nanofabrication process. Such a process may be useful for the hierarchal assembly of integrated nano/micro/macrostructures for a variety of electronic, photonic, materials and other applications.

REFERENCES

[1] Nature Genetics Supplement, vol. 21, pp. 1-60, 1999.

[2] Chee M, Yang R, Hubbell E, Berno A, Huang X, Stern D, Winkler J, Lockhart D, Morris M, Fodor S, "Accessing Genetic Information with High-Density DNA Arrays," Science, vol. 274, pp. 610-614, 1996.

[3] Pease A, Solas D, Sullivan E, Cronin M, Holmes C, Fodor S, "Light-Generated Oligonucleotide Arrays for Rapid DNA Sequence Analysis," PNAS, vol. 99, pp. 5022-5026, 1994.

[4] Lipshutz RJ, Morris D, Chee M, Hubbell E, Kozal MJ, Shah N, Shen N, Yang R, Fodor SP, "Using oligonucleotide probe arrays to access genetic diversity," Biotechniques, vol. 19, pp. 442-447, 1995.

[5] Taton A, Mirkin C, Letsinger R, "Scanometric DNA Array Detection with Nanoparticle Probes," Science, vol. 289, pp. 1757-1760, 2000.

[6] Sosnowski R, Tu E, Butler W, O'connel J, Heller M, "Rapid Determination of Single Base Mismatch Mutations in DNA Hybrids by Direct Electric Field Control," PNAS, vol. 94, pp. 1119-1123, 1997.

[7] Harrington CA, Rosenow C, "Monitoring gene expression using DNA microarrays," J Curr Opin Microbiol, vol. 3, pp. 285-291, 2000.

[8] Radtkey R, Feng L, Muralhidar M, Duhon M, Cantor D, DiPierro D, Fallon S, Tu E, McElfresh K, Nerenberg M, Sosnowski R, "Rapid, High Fidelity Analysis of Simple Sequence Repeats on an Electronically Active DNA Microchip," Nucleic Acids Research, vol. 28, pp. 1-6, 2000.

[9] Gilles P, Wu D, Foster C, Dillon P, Chancock S, "Single Nucleotide Polymorphic Discrimination by and Electronic Dot Blot Assay on Semiconductor Microchips," Nature Biotechnology, vol. 17, pp. 365-370, 1999.

[10] http://www.idahotech.com/products/index.html

[11] Belgrader P, Smith JK, Weedn VW, Northrup MA.,"Rapid PCR for identity testing using a battery-powered miniature thermal cycler," J Forens Sci, vol. 43, p. 315, 1998.

[12] Ibrahim MS, Lofts RS, Jahrling PB, Henchal EA, Weedn VW, Northrup MA, Belgrader P, "Real-Time Microchip PCR for Detecting Single-Base Differences in Viral and Human DNA," Anal Chem, vol. 70, p. 2013, 1998.

[13] K. Petersen, "DNA-chip technologies Part 2: State-of-the-art and competing technologies" IVD Technology, Nov./Dec., p. 35, 1998.

[14] Clinical Laboratory News, February, 1998.

[15] Christel L. A., K. Petersen, W. McMillan and M.A. Northrup, "Rapid, automated nucleic acid probe assays using silicon microstructures for nucleic acid concentration," J Biomech Eng - Trans ASME, vol. 121, pp. 22-27, 1999.

[16] Hodko D, "DNA Microarrays in Portable Diagnostic Systems", at "Miniaturization in Biomedicine" Conference, University of California, Irvine, October 30, 2003.

[17] Heller ML, "An active microelectronics device for multiplex DNA analysis," IEEE Engineering In Medicine and Biology, March/April, p. 100, 1996.

[18] Heller MJ, Tu E, US Patent No. 5,605,662 "Active Programmable Electronic Devices for Molecular Biological Analysis and Diagnostics", February 25, 1997.

[19] Ackley DE, Swanson PD, Graham SO, Mather EL, LeClair TL, Butler WF, US Patent No. 6,726,880 "Advanced Active Electronic Devices for Molecular Biological Analysis and Diagnostics and Methods of Manufacturing Same", April 27, 2004.

[20] Kricka, LJ, "Revolution on a square centimeter," Nature Biotechnology, vol. 16, pp. 513-514, 1998.

[21] Cheng J, Sheldon EL, Wu L, Uribe A. Gerrue LO, Carrino, J, Heller, MJ, O'Connell J, "Electric field controlled preparation and hybridization analysis of DNA/RNA from E. coli on microfabricated bioelectronic chips", Nature Biotechnology, vol. 16, pp. 541-546, 1998.

[22] Kassegne K, Reese H, Hodko D, Yang JM, Sarkar K, Smolko D, Swanson P, Raymond DE, Heller MJ, Madou MJ, "Numerical modeling of transport and accumulation of DNA on electronically active biochips," Sensors and Actuators B, vol. 94, p. 81, 2003.

[23] Huang Y, Mather EL, Bell JL, Madou MJ, "MEMS-based sample preparation for molecular diagnostics," Anal Bioanal Chem, vol. 372, p. 49, 2002.
[24] Westin L, Xu X, Miller C, Wang L, Edman CF, Nerenberg M, "Anchored multiplex amplification on a microelectronic chip array," Nature Biotechnology, vol. 18, p. 199, 2000.
[25] Yang JM, Bell J, Huang Y, Tirado M, Thomas T, Forster AH, Haigis RW, Swanson PD, Wallace RB, Martinsons B, Krihak M, "An integrated, stacked microlaboratory for biological agent detection with DNA and immunoassays," Biosen Bioelect, vol. 17, p. 605, 2002.
[26] Heller MJ, Cable JM, Esener SC, US Patent No. 6,652,808 "Methods for the Electronic Assembly and Fabrication of Devices", November 25, 2003.
[27] Edman CF, Heller MJ, Gurtner C, Formosa R, US Patent No. 6,706,473 "Systems and Devices for Photoelectronic Transport and Hybridization of Oligonucleotides", March 16, 2004.
[28] Dehlinger D, Sullivan B, Esener S, Swanson P, Hodko D, and Heller MJ, "Reconfigurable CMOS Electronic Microarray System for the Assisted Self-Assembly of Higher-Order Nanostructures", in Nanomanufacturing Handbook, Busnaina A Ed. Chapter 5, CRC Press, 2006.
[29] Sullivan B, Dehlinger D, Zlatanovic, Esener S, Heller MJ, "Electrophoretically Actuated Nanoscale Optoentropic Transduction Mechanisms," NSTI Nanotech 2006, vol. 2, pp. 209-212, 2006.
[30] Dehlinger D, Sullivan B, Esener S, Swanson P, Hodko D, and Heller MJ, "Next Generation Microelectronic Array Devices", in Handbook of Nanotechnology, Bhushan, B Ed. Part B, Chapter 14, Springer, 2006.
[31] Heller MJ, Ozkan CS and Ozkan M, "Electric Field Devices for Assisted Assembly of DNA Nanocomponents and Other Nanofabrication Applications", in BioMEMS and Biomedical Nanotechnology VI, Ferrari M Ed., vol. 2, Chapter 6, Springer, 2006.
[32] Heller MJ, Dehlinger DA Sullivan BD, "Parallel Assisted Assembly of Multilayer DNA and Protein Nanoparticle Structures Using a CMOS Electronic Array", DNA-Based Nanoscale Integration Symposium, W. Fritzsche (ed), American Institute of Physics, NY, pp. 73-81, 2006.
[33] Swanson P, Gelbart R, Atlas E, Yang L, Grogan T, Butler WF, Ackley DE, Sheldon E, "A fully multiplexed CMOS biochip for DNA Analysis," Sensors and Actuators B, vol. 64, pp. 22-30, 2000.
[34] Svrcek V, Slaoui A, Muller JC, "Silicon nanocrystals as light converter for solar cells," Thin Solid Films, vol. 451, pp. 384-388, 2004.
[35] Mwaura JK, Pinto MR, Witker D, Ananthakrishnan N, Schanze KS, Reynolds JR, "Photovoltaic cells based on sequentially adsorbed multilayers of conjugated poly(p-phenylene ethynylene)s and a water-soluble fullerene derivative," Langmuir, vol. 22, pp. 10119-10126, 2005.
[36] Kapur VK, Bansal A, Le P, Asensio O, "Non-vacuum processing of CuIn1-xGaxSe2 solar cells on rigid and flexible substrates using nanoparticle precursor inks," Thin Solid Films, vol. 431, pp. 53-57, 2003.

[37] Kim DW, Choi HS, Lee C, Blumstein A, Kang Y, "Investigation on methanol permeability of Nafion modified by self-assembled clay-nanocomposite multilayers," Electrochimica Acta, vol. 50, pp. 659-662, 2004.
[38] Hacker V, Wallnofer E, Baumgartner W, Schaffer T, Besenhard JO, Schrottner H, Schmied M, "Carbon nanofiber-based active layers for fuel cell cathodes - preparation and characterization," Electrochemistry Communications, vol. 7, pp. 377-382, 2005.
[39] Catledge SA, Fries MD, Vohra YK, Lacefield WR, Lemons JE, Woodard S, Venugopalan R, "Nanostructured Ceramics for Biomedical Implants," Nanosci Nanotechno, vol. 2, pp. 293-312, 2002.
[40] Akerman ME, Chan WC, Laakkonen P, Bhatia SN, Ruoslahti E, "Nanocrystal Targeting *In Vivo*," PNAS, vol. 99, pp. 12617-12621, 2002.
[41] Gao J, Gao T, Sailor MJ, "A porous silicon vapor sensor based on laser interferometry," Appl Phys Lett, vol. 77, pp. 901-903, 2000.
[42] Dancil K-PS, Greiner DP, Sailor MJ., "A Porous Silicon Optical Biosensor: Detection of Reversible Binding of IgG to a Protein A-Modified Surface," J Am Chem Soc, vol. 121, pp. 7925-7930, 1999.
[43] Chan S, Fauchet PM, Li Y, Rothberg LJ, Miller BL, "Porous Silicon Microcavities for Biosensing Applications," Phys Stat Sol A, vol. 182. pp. 541-546, 2000.
[44] Crommelin DJ, Strom G, Jiskoot W, Stenkes R, Mastrobattista E, Hennick WE, "Nanotechnological approaches for the delivery of macromolecules," Control Release, vol. 87, pp. 81-88, 2003.
[45] Prime KL and Whitesides GM, "Self-Assembled Organic Monolayers: Model Systems for Studying Adsorption of Proteins at Surfaces," Science, vol. 252, pp. 1164-1167, 1991.
[46] Kim B, Tripp SL, and Wei A, "Self-organization of large gold nanoparticle arrays," J Am Chem Soc, vol. 123, pp. 7955-7956, 2001.
[47] Bowden N, Terfort A, Carbeck J, and Whitesides GM, "Self-assembly of mesoscale objects into ordered two-dimensional arrays," Science, vol. 276, pp. 233-235, 1997.
[48] Fink J, Kiely CJ, Bethell D, and Schiffrin DJ, "Self-organization of nanosized gold particles," Chem Mater, vol. 10, pp. 922-926, 1998.
[49] Lee S-W, Mao C, Flynn CE, and Belcher AM, "Ordering of quantum dots using genetically engineered viruses," Science, vol. 296, pp. 892-895, 2002.
[50] Mirkin CA, Letsinger RL, Mucic RC, Storhoff JJ, "A DNA-based method for rationally assembling nanoparticles into macroscopic materials," Nature, vol. 382, pp. 607-609, 1996.
[51] Decher, D, Schlenoff, JB, Multilayer Thin Films-Sequential Assembly of Nanocomposite Materials, Wiley-VCH Verlag, Weinheim, 2003.
[52] Goddard, Brenner, Lyashevski, Lafrate, Handbook of Nanoscience, Engineering and Technology, CRC Press, Boca Raton, 2003.

[53] Wanunu M, Popovitz-Biro R, Cohen H, Vaskevich A, Rubinstein I, "Coordination-based gold nanoparticle layers," J Am Chem Soc, vol. 127, pp. 9207-9215, 2005.
[54] Artyukhin AB, Bakajin O, Stroeve P, Noy A, "Layer-by-layer electrostatic self-assembly of polyelectrolyte nanoshells on individual carbon nanotube templates," Langmuir, vol. 20, pp. 1442-1448, 2004.
[55] Yuan JJ, Zhou SX, You B, Wu LM, "Organic pigment particles coated with colloidal nano-silica particles via layer-by-layer assembly," Chem Mater, vol. 17, pp. 3587-3594, 2005.
[56] Ma N, Zhang HY, Song B, Wang ZQ, Zhang X, "Polymer micelles as building blocks for layer-by-layer assembly: An approach for incorporation and controlled release of water-insoluble dyes," Chem Mater, vol. 17, pp. 5065-5069, 2005.
[57] Zapotoczny S, Golonka M, Nowakowska M, "Novel photoactive polymeric multilayer films formed via electrostatic self-assembly," Macromolecular Rapid Communications, vol. 26, pp. 1049-1054, 2005.
[58] Hammond PT, "Form and function in multilayer assembly: New applications at the nanoscale," Advanced Materials, vol. 16, pp. 1271-1293, 2004.
[59] Jacobs HO, Campbell SA, Steward MG, "Approaching nanoxerography: The use of electrostatic forces to position nanoparticles with 100 nm scale resolution," Advanced Materials, vol. 14, pp. 1553-1557, 2002.
[60] Barry CR, Gu J, Jacobs HO, "Charging process and coulomb-force-directed printing of nanoparticles with sub-100-nm lateral resolution," Nano Lett, vol. 5, pp. 2078-2084, 2005.
[61] Mardilovich P, Kornilovitch P, "Electrochemical fabrication of nanodimensional multilayer films," Nano Lett, vol. 5, pp. 1899-1904, 2005.
[62] Allred DB, Sarikaya M, Baneyx F, Schwartz DT, "Electrochemical nanofabrication using crystalline protein masks," Nano Lett, vol. 5, pp. 609-613, 2005.
[63] Tsai DH, Kim SH, Corrigan TD, Phaneuf RJ, Zachariah MR, "Electrostatic-directed deposition of nanoparticles on a field generating substrate," Nanotechnology, vol. 16, pp. 1856-1862, 2005.
[64] Lumsdon SO, Kaler EW, Velev OD, "Two-dimensional crystallization of microspheres by a coplanar AC electric field," Langmuir, vol. 20, pp. 2108-2116, 2004.
[65] Hua F, Shi J, Lvov Y, Cui T, "Patterning of layer-by-layer self-assembled multiple types of nanoparticle thin films by lithographic technique," Nano Lett, vol. 2, pp. 1219-1222, 2002.
[66] Ko HH, Jiang CY, Tsukruk VV, "Encapsulating nanoparticle arrays into layer-by-layer multilayers by capillary transfer lithography," Chemistry of Materials, vol. 17, pp. 5489-5497, 2005.
[67] Small Wonders, Endless Frontiers: Review of the National Nanotechnology Initiative, National Research Council, 2002.

[68] The National Nanotechnology Initiative - Strategic Plan, National Science and Technology Council, December 2004.
[69] Heller MJ, Tu E, Martinsons R, Anderson RR, Gurtner C, Forster A, Sosnowski R, "Active microelectronic array systems for DNA hybridization, genotyping, pharmacogenomics and Nanofabrication Applications," in Integrated Microfabricated Devices, Chapter 10, 2002.
[70] Huang Y, Sunghae J, Duhon M, Heller MJ, Wallace B, Xu X, "Dielectrophoretic separation and gene expression profiling on microelectronic chip arrays," Anal Chem, vol. 74, pp. 3362-3371, 2002.
[71] Esener SC, Hartmann D, Heller MJ, and Cable JM, "DNA Assisted Micro-Assembly: A Heterogeneous Integration Technology For Optoelectronics," Proc. SPIE Critical Reviews of Optical Science and Technology, Heterogeneous Integration, Ed. A. Hussain, CR70, Chapter 7, 1998.
[72] Gurtner C, Edman CF, Formosa RE, Heller MJ, "Photoelectrophoretic Transport and Hybridization of DNA on Unpatterned Silicon Substrates," J Am Chem Soc, vol. 122, pp. 8589-8594, 2000.
[73] Huang Y, Ewalt KL, Tirado M, Haigis R, Forster A, Ackley D, Heller MJ, O'Connell JP, Krihak M, "Electric manipulation of bioparticles and macromolecules on microfabricated electrodes," Anal Chem, vol. 73, pp. 1549-1559, 2001.
[74] Edman CF, Gurtner C, Formosa RE, Coleman JJ, Heller MJ, "Electric-Field-Directed Pick-and-Place Assembly," HDI, vol. 10, pp. 30-35, 2000.
[75] Edman CF, Swint RB, Gurthner C, Formosa RE, Roh SD, Lee KE, Swanson PD, Ackley DE, Colman JJ, Heller MJ, "Electric Field Directed Assembly of an InGaAs LED onto Silicon Circuitry," IEEE Photonics Tech. Letters, vol. 12, pp. 1198-1200, 2000.
[76] Daniel M. Hartmann, David Schwartz, Gene Tu, Mike Heller, Sadik C. Esener, "Selective DNA attachment of particles to substrates," Journal of Materials Research, vol. 17, pp. 473-478, 2002.
[77] US # 6,569,382 "Methods and Apparatus for the Electronic Homogeneous Assembly and Fabrication of Devices", Issued May 27, 2003.

AUTHOR BIOGRAPHY

Dalibor Hodko received his B.Sc., M.Sc., Ph.D., and postdoctoral education in chemical engineering and chemistry from the University of Zagreb, Croatia and Texas A&M University. He has more than 20 years of industrial experience in the development of innovative technologies (17 issued patents) and products in the areas of bioanalytical and micro- and nano-fluidics devices as well as large systems with applications in physical and environmental chemistry. He has intensive experience in managing Government supported projects and was a Principal Investigator and manager on more than 25 federally funded projects including NIH, NASA, NSF,

DOD, EPA and DOE. Dr. Hodko is presently a Director of Advanced Technology at Nanogen, Inc.

Paul Swanson is a Principal Engineer at Nanogen, Inc. He received his Ph.D. degree in Electrical Engineering from the University of Illinois. He has designed generations of microelectrode arrays, with and without integrated active circuitry, for the electrophoretic manipulation of charged nanoscaled materials. He is an inventor on patents ranging from devices for molecular biological analysis to laser logic devices.

Dietrich Dehlinger recieved a B.A. in Physics from Reed College in 2001. He is currently pursuing a graduate degree in Electrical and Computer Engineering at the University of California, San Diego.

Benjamin Sullivan is a biomedical engineer with a research focus on nanoscale optoentropic transduction mechanisms, electrophoretic transport, and tear film biophysics. He is currently a graduate student at the University of California, San Diego.

Michael J. Heller began his faculty position at University of California, San Diego in July 2001. He has a joint appointment between the department's of Bioengineering and Electrical and Computer Engineering (ECE). His experience (academic and industrial) includes many areas of biotechnology and biomedical instrumentation, with particular expertise in DNA synthesis, DNA microarray diagnostics and optoelectronic based biosensor technologies. Dr. Heller has been the co-founder of three high-tech companies: Nanogen, Nanotronics and Integrated DNA Technologies. Dr. Heller's most recent work involved the development of an integrated microelectronic array based system for genotyping, genetic and infectious disease diagnostics, protein analysis, cell separations and for nanofabrication applications. Dr. Heller has a respectable publication record, and has been an invited speaker to a large number of scientific conferences and meetings related to DNA microarrays, biosensors, lab-on-a-chip devices, bio-MEMS and nanotechnology. He has over 35 issued US patents related to microelectronic chips, microarrays and integrated devices for DNA hybridization, miniaturized sample to answer diagnostic devices, biosensors, genomics, proteomics, nanotechnology and nanofabrication, nano-based DNA optical storage and for fluorescent energy transfer in DNA nanostructures. Dr. Heller has been a panel member for the NAS (NAE) Review of National Nanotechnology Initiative 2001-2002; the NAS(NAE) – Engineer for the 2020 - 2001/2002; the White House (OSTP) National Nanotechnology Initiative 1999/2000; and has also been involved in a number of NSF Nanotechnology Workshops.

8 INTEGRATED MICROELECTRODE ARRAYS

Flavio Heer* and Andreas Hierlemann

Physical Electronics Laboratory, ETH Zurich
*fheer@phys.ethz.ch

8.1 INTRODUCTION

A high degree of connectivity and the coordinated electrical activity of neural cells or networks are believed to be the reason that the brain is capable of highly sophisticated information processing. Likewise, the effectiveness of an animal heart largely depends on such coordinated cell activity. To advance our understanding of these complex biological systems, high spatiotemporal-resolution techniques to monitor the cell electrical activity and an ideally seamless interaction between cells and recording devices are desired. Electrogenic cells grown on substrate-integrated microelectrode arrays (MEAs) constitute an established system to study, e.g. neuronal information processing strategies [1-5] or to characterize the cellular response upon dosing biologically active agents [6-9]. Studying electrogenic cells *in vitro* provides different experimental conditions in comparison to investigating cells *in vivo* or in intact organisms: Most animal and plant tissues comprise a variety of different cell types, whereas a single specific cell type of interest can be grown in culture. Moreover, the experimental parameters, such as temperature or concentration of a chemical compound, can be controlled much more rigorously in cell cultures than in an organism. On the other hand it is not possible to directly draw from *in vitro* experiments upon the *in vivo* behavior of cells. The use of acute or organotypic slices, e.g. brain slices, which is also possible with planar multitransducer or multielectrode chips [10, 11] may help to bridge the gap between the *in vitro* and *in vivo* experiments.

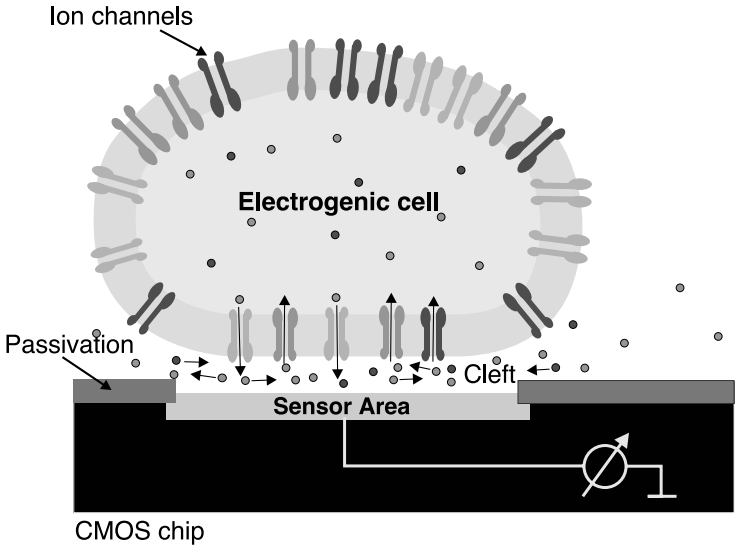


Figure 8.1 Schematic of a cell attached to a sensor surface. A cell featuring ion channels sits on a planar microtransducer (open-gate transistor or electrode). Moving ions in the electrode vicinity generate an electric field or voltage that is recorded by the respective microtransducer.

Methods to directly measure the electrical activity of cultured electrogenic cells (e.g. neurons, heart cells, retina cells, or muscle cells) include two fundamentally different techniques: (a) transmembrane measurements by inserting one of the electrodes into the cell, the so-called "patch clamp" technique [12, 13], and (b) extracellular recordings, e.g. by means of external microtransducers [14-25]. Additionally, there are indirect methods like optical measurements using voltage-sensitive or fluorescent dyes [26-28]. The patch clamp technique yields very accurate information on the electrophysiological properties of entire cells, or, alternatively, on currents flowing through single ion channels. However, it is an invasive method and is limited in the cell viability time (usually hours) and in the overall number of cells that can be simultaneously recorded from. For extracellular recordings, the cells are cultured directly on top of a transducer or an array of transducers (Fig. 8.1). When electrical activity or a so-called "action potential" occurs in a cell, ions flow across the cell membrane within milliseconds. These moving ions generate an electric field, which either directly influences the open-gate region of a field-effect transistor [29, 30] or can be recorded by a metallic microelectrode [8, 31-34]. Extracellular recordings are non-invasive (no puncturing of the cell membrane), which entails a potentially

long measurement time. Multi-site measurements are possible by arranging many transducers in an array. For stimulation, voltage transients can be applied by means of electrodes, which then evoke a depolarization of the cell membrane and solicit subsequent electrical cell activity.

8.1.1 Why using IC or CMOS Technology

The two main advantages in using integrated-circuit (IC) or complementary metal oxide semiconductor (CMOS) technology include: (a) *connectivity*; larger numbers of transducers or electrodes can be addressed by on-chip multiplexing architectures, and (b) *signal quality*; the signal is conditioned right at the electrode by means of dedicated circuitry units (filters, amplifiers). The use of CMOS technology hence offers to realize on a small system chip a large number of electrodes that provide good signal quality owing to the respective associated circuitry units.

On-chip microelectronics as provided by the use of IC or CMOS technology translate into system capability: signal conditioning can be performed on-chip, ensuring that weak neural signals are faithfully recorded; a CMOS system also enables a bidirectional communication via the electrodes (stimulation and recording); smart switching schemes allow for stimulating the cell ensemble via an arbitrarily selectable set of electrodes, all while recording from other electrodes uninterruptedly during stimulation; on-chip analog-to-digital conversion means that such chip produces a robust signal that may be easily manipulated and transferred without compromising its information content. Moreover, the use of on-chip electronics allows for the monolithic integration of the complete system on a single chip, which leads to small system dimensions and low power consumption, which is a key requirement for, e.g. implantable devices.

The use of multiplexers enables the integration of a large number of electrodes or transducers so that measurements at high spatiotemporal resolution become feasible. Traditional MEAs without multiplexers usually offer 64 electrodes with each electrode needing a connection to external circuitry, which adds parasitic capacitance and attenuates the small signals, whereas CMOS-based MEAs comprise up to 16,384 electrodes and the needed addressing circuitry on the same chip [35]. The high transducer density, which cannot be realized without on chip multiplexers and addressing, simply since it is impossible to establish thousands of physical or electrical connections from a millimeter-size chip, is an important asset to, e.g. study the behavior of neural networks or other electrophysiologically relevant processes at cellular resolution.

A disadvantage of CMOS chips is that silicon is not transparent to visible light in contrast to standard cell culture substrates used in biology. Additionally, the chip or its components can corrode upon operation and long-term exposure to liquids (salt water). Therefore, a good packaging solution is needed to, on the one hand, protect the chip against metabolism products and chemicals of the cell culture, and, on the other hand, to prevent the cells from being poisoned or disturbed by toxic materials released by the chip, such as the CMOS metal aluminum that dissolves in saline solution.

This book chapter gives an overview on the design, system integration and application of CMOS-based microarrays as bio-electronic devices for recording from and for stimulation of electrogenic cells. Several state-of-the-art CMOS-based systems for extracellular recording will be presented in this chapter, and a CMOS microelectrode array featuring 128 bidirectional interfaces/electrodes as well as complete on-chip signal conditioning, including A/D and D/A conversion will be treated in great detail.

8.2 FUNDAMENTALS OF RECORDING OF ELECTRICAL CELL ACTIVITY

8.2.1 Electrogenic Cells

Many types of cells in the body have the ability to undergo a transient electrical depolarization and repolarization that is either triggered by external mechanisms (e.g. motor-nerve stimulation of skeletal muscle, or cell-to-cell depolarization in the heart) or by intracellular, spontaneous mechanisms (e.g. cardiac pacemaker cells). Cells that exhibit the ability to generate electrical signals are called electrogenic cells, the most prominent of which include brain cells or neurons and heart cells or cardiomyocytes. The biology of these electrogenic cells is rather complex because many interdependent electrical, electro-physiological and biochemical processes are involved.

A brief description of the biology of electrogenic cells will be given in this subsection. The mechanisms underlying the functioning of voltage-gated ion channels, the generation of action potentials and the conduction of action potentials through a cellular network will be described for neuronal cells. Some of these mechanisms are similar for other electrogenic cells like the cardiomyocytes. For more details on electrogenic cells, the reader is directed to various books on this topic, see refs. [36-38].

8.2.1.1 The Neuron

A huge number of neuronal cells form the nervous system that regulates all aspects of body functions. The human brain contains about 10^{12} neurons (nerve cells) with each neuron forming thousands of connections to other neurons so that a large network of electrical connections with massively parallel information processing characteristics results. The nervous system also contains glial cells that occupy the spaces between neurons and modulate their function. They surround the soma and axons of the neurons and are metabolically coupled to the neurons. The output of a nervous system is the result of its inputs and its circuit properties, that is, of the wiring or interconnections (synapses) between the single neurons, and of the strength of these interconnections. The synaptic connections between neurons can be reorganized, which is known as synaptic plasticity, and is believed to be the mechanism of learning in our brain. Basically, three different types of neurons can be distinguished by their physiology and function in the body:

- interneurons, neurons in contact with other neurons (most of the interneurons are located in the brain)
- motor neurons, neurons, which control the actions of the organism mainly via contact with muscle cells
- sensor neurons, neurons, which receive stimuli from the external environment, such as the retina.

The neuron exhibits four distinct regions with different functions: the *cell body*, the *dendrites*, the *axon* and the *axon terminals* (Fig. 8.2). The cell body with a diameter of 10 to 100 µm contains the nucleus and is the production site of most neuronal proteins. Almost every neuron has a single *axon*, whose diameter varies from a micrometer in the human brain to a millimeter in the giant squid. Axons are specialized for the conduction of electrical pulses, termed action potentials, away from the cell body towards the axon terminus.

Neurons communicate with one another through specialized contact zones, which are called *synapses*. Synapses can be either electrical or chemical. Synapses have two important functions in the transmission of impulses from one cell to the other. The first is signal amplification, which is common at nerve-muscle synapses. A single motor neuron can cause a contraction of multiple muscle cells because the release of relatively few signaling molecules is required at the synapse to stimulate contractions. The second advantage is signal computation, which is common at synapses involving interneurons. A single neuron can be affected simultaneously by signals received at multiple excitatory, inhibitory and also electrical synapses. The

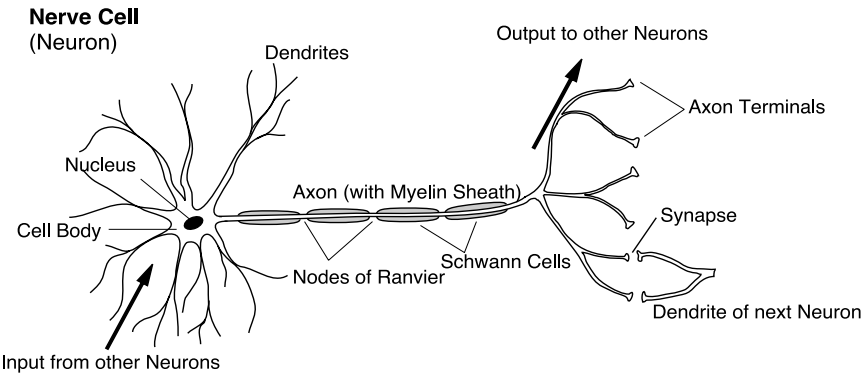


Figure 8.2 Schematic view of a neuron.

neuron averages these signals continuously and determines whether or not to generate an action potential. The connection between two neurons is therefore weighted by the property of the synapse, which connects the two neurons. This is the basis of the ability of neural networks to learn and to perform computation.

Most neurons have multiple *dendrites*, which extend outward from the cell body and are specialized to receive chemical or electrical signals from the axon terminals of other neurons via the *synapses*. Dendrites convert these signals into small electric impulses and transmit them towards the cell body. Particularly in the central nervous system, neurons have extremely long dendrites exhibiting complex branching.

8.2.1.2 Gated Ion Channels

Ion channels are embedded transmembrane proteins and allow for the formation of a concentration gradient of ions between the extracellular and the intracellular space (the intracellular space is also termed cytosol). This concentration gradient leads to an electrical potential across the cell membrane (typically between –30 mV and –70 mV). Ion channels are termed *gated* if they can be opened or closed. There are three types of gated ion channels: *ligand-gated*, *mechanically gated* and *voltage-gated* channels.

The best studied model for the mechanisms in excitable membranes is the squid giant axon as described by Hodgkin and Huxley [39]. In order to address both, voltage and time dependence of ion conductances in the squid axon, Hodgkin and Huxley developed the gate model, which proposes that ion currents are flowing through trans-membrane channel proteins that form aqueous pores, through which ions can diffuse down their concentration gradients. These pores have gates that are controlled by voltage-sensi-

tive gating charges or gating particles (Fig. 8.1). This model will be briefly introduced here, for more information the interested reader is directed to, e.g. ref. [37] or, for cardiac cells, to ref. [40].

The equivalent circuit of the gate model consists of parallel conductors describing the ionic currents and the membrane capacitance. The resulting total membrane current is then described by the following equation:

$$I_M = C_M \frac{dV_M}{dt} + \sum_i I_i = C_M \frac{dV_M}{dt} + \sum_i g_i(V_M, t) \cdot (V_M - E_i), \qquad (8.1)$$

where I_M is the total membrane current, C_M is the membrane capacitance, V_M the membrane potential, I_i the current from the ions i (e.g. Na, K, Cl), E_i the equilibrium potential of ion i, and the g_i are the specific voltage- and time-dependent conductances for ion i.

As an example Fig. 8.3(a) shows the equivalent circuit model of the squid giant axon as described by Hodgkin and Huxley. The main ionic currents are the potassium current, I_K, the sodium current, I_{Na}, and some small currents summarized in a leakage current, I_L. The conductances, g_i, are voltage- and time-dependent and can be represented as complex, fourth-order differential equations. The resulting electrical signals, the so-called action potentials and the respective ionic conductances are shown in Fig. 8.3(b).

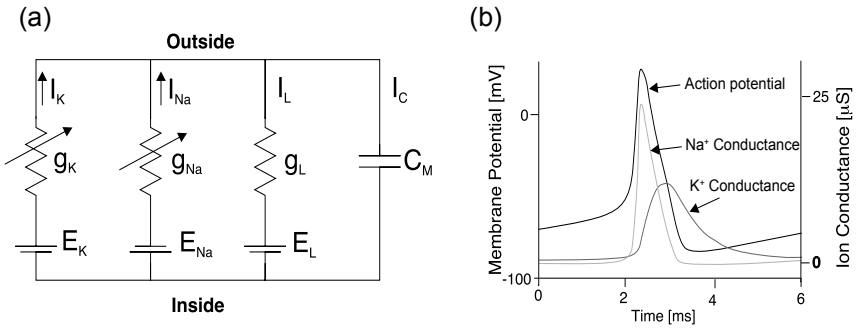


Figure 8.3 (a) Hodgkin-Huxley model of the squid axon. The sodium and potassium ionic conductances, g_K and g_{Na}, are voltage- and time-dependent, and g_L is assumed to be constant. The total membrane current is described by Eq. (8.1). (b) Action potential of the neuron and the transmembrane ionic conductances of K^+ and Na^+, which are the dominant ionic currents involved in the action potential generation.

8.2.2 Recording and Stimulation Techniques and Tools

The action potential of an electrogenic cell can either be measured using fluorescence methods, e.g. voltage sensitive dyes [26-28], or directly by using electrical recording techniques. Standard methods to electrically measure the activity of electrogenic cells are based on two different techniques: (a) trans-membrane measurements of the action potential using the patch-clamp technique with one electrode inserted into the cell and (b) noninvasive extracellular recordings by means of planar microelectrodes.

8.2.2.1 Intracellular Recording

The electrical activity of single cells can be recorded by means of glass pipettes that are inserted through the membrane into the cell to establish an intracellular electrical contact. The patch-clamp technique [12] yields very accurate and rich information about electrophysiological phenomena in the cells and played a key role in understanding the ion channel activities and the mechanisms underlying a cellular action potential. The recording from entire cells, or alternatively of currents flowing through single ion-channels, can be performed. It is an invasive method since a glass pipette electrode is used that penetrates the cell membrane thereby gaining access to the intracellular medium. While this method is rich in information it is limited by the number of cells that can be simultaneously monitored and by short-term cell viability (of usually hours).

8.2.2.2 Extracellular Recording

The electrical activity of cells can also be recorded without disrupting the cell membrane using extracellular recording techniques. The recording of electrical oscillations in the brain began in 1875, when Richard Caton discovered that currents could be recorded from deep inside the brain [41]. First extracellular recordings using an array of micro electrodes date back to the 1970s [42, 43].

For extracellular recordings, the cells are cultured directly on top of a transducing element, which is, in most cases, either a metallic electrode or an open-gate transistor (Fig. 8.4). When electrical activity or a so-called *action potential* in a cell occurs, ions flow across the cell membrane within milliseconds. When a cell is close to a transducer, the moving ions generate an electric field or voltage, which either directly influences the open gate region of the field-effect transistor or can be recorded by the metallic microelectrode (Fig. 8.4). Extracellular recordings are non-invasive

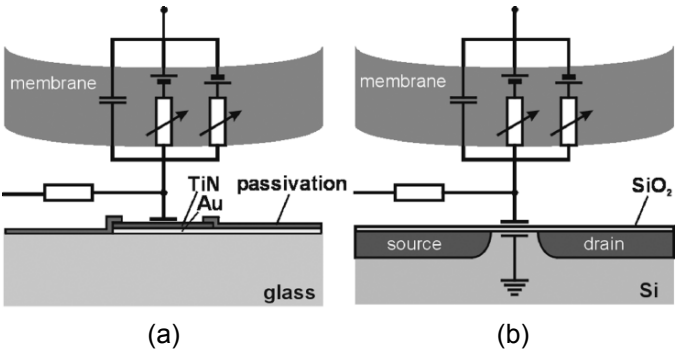


Figure 8.4 Devices for extracellular recording. (a) Metal electrode. (b) Open-gate field-effect transistor. Reprinted from [44] with permission.

(no puncturing of the cell membrane), which entails a potentially long measurement time, and microtransducer arrays offer multi-site measurement capabilities.

The magnitude of the extracellular signal depends on the ionic transmembrane currents and on the cleft resistance (Fig. 8.4). The transmembrane currents depend on the type and the number of contributing ion channels, which, in turn, depend on the size of the cell and on the spatial distribution of the ion channels. The cleft resistance is a function of the contact area, of the width of the cleft, and of the electrolyte resistance. The transmembrane peak-to-peak value of an action potential as measured by the patch-clamp method is about 100 mV. The action potential-induced voltage in the cleft between the cell and the transducer is on the order of a few microvolts to a few millivolts, and represents the signal that can be measured on the electrodes.

Several laboratories have established the capability to perform extracellular recordings from neuronal or cardiac cells cultured on MEAs [7, 14, 21, 32, 45-49]. Such MEAs have been used to study the effects of pharmacological agents on neuronal net activity [6], to detect neurotoxicants [50], or for investigating neural network dynamics [33, 47]. Commercially available systems are provided by several companies[1]. All these commercial systems are based on passive electrode arrays on a glass or silicon substrate with external signal conditioning and recording components.

[1] Multi Channel Systems GmbH, Germany, www.multichannelsystems.com; Panasonic, Japan, www.med64.com; Plexon Inc., USA, www.plexonic.com; Bionas, Rostock, Germany, www.bionas.de.

8.2.2.3 Field-Effect Transistors: FETs

It is possible to record electrical signals of cultured cells attached to the transducer surface with an open-gate field-effect transistor. Firing neurons are located on the exposed gate oxide of the transistor as it is illustrated in Fig. 8.1. A change of the local extracellular potential in the cleft between the cell and the transducer directly modulates the source-drain current (Fig. 8.5). The cell-chip junction may be considered as the microscopic input region of an electronic amplifier that does not have an electrolyte-metal contact. In common metal-oxide silicon (MOS) field-effect transistors, the low-frequency noise is caused by electrons tunneling between silicon and traps in the gate oxide. This noise is reduced when the electron channel of the transistor is buried in the silicon substrate, a few nanometers away from the interface.

The specific capacitance of an open-gate transistor is small in comparison to the capacitance of a metal electrode. As a consequence, the capacitive current through the transducer can be neglected. This approximation implies that the relaxation of the junction is fast in comparison to the dynamics of the transducer response. It is valid for recordings in the millisecond range [65].

8.2.2.4 Metal Electrodes

Metal electrodes are widely used to measure the electrical activity of cultured cells in biological and medical research [Fig. 8.5(c)]. The metal-electrode-electrolyte impedance can be very low, which renders metal electrodes suitable for extracellular stimulation and recording. However, metal electrodes are not suitable for the measurement of dc-potentials in contrast to the FETs [68]. The impedance of the microelectrode is an important parameter for extracellular recording, since it determines the noise of the electrode and the signal attenuation. Therefore, methods to increase the overall electrode surface area are applied. Commonly used materials include, e.g. platinum-black [69-71], titanium nitride [31], ITO [6], or iridium oxide [72].

There are two types of processes that can occur at metal electrode/electrolyte interfaces, a *capacitive process* resulting from the redistribution of charged and polar particles in the electrode vicinity with no charge-transfer between the solution and the electrode, and a component resulting from the electron exchange between the electrode and a redox species in the solution termed *faradaic* process (Fig. 8.6). In the non-faradaic case the electrode is typically highly polarized and behaves almost as an ideal capacitor. For any electrode-electrolyte interface there is a range of potentials, in which

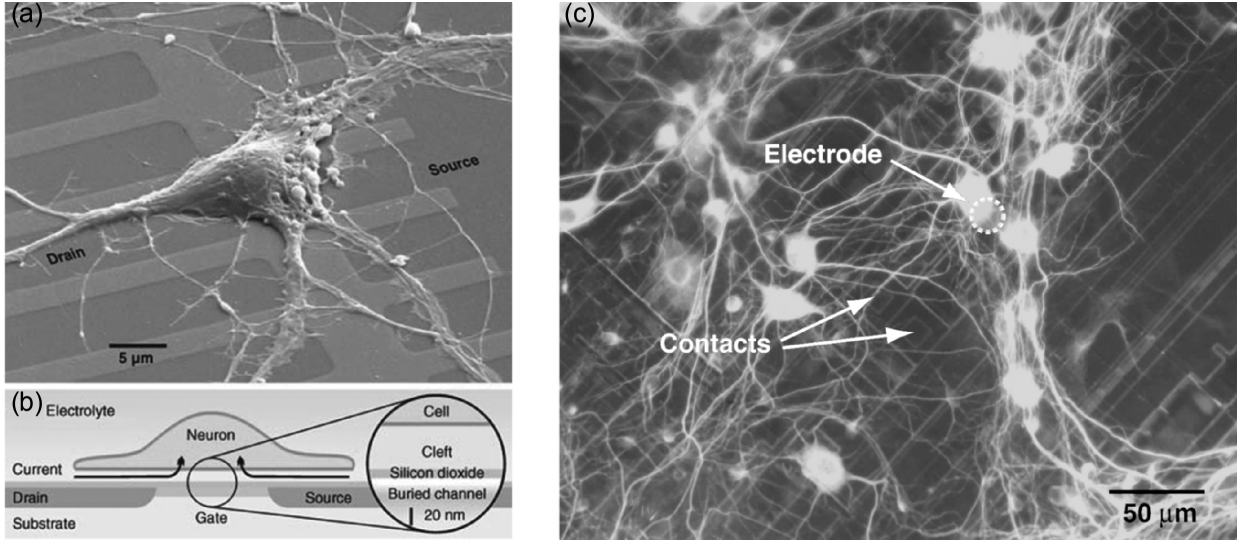


Figure 8.5 (a) Electron micrograph of a hippocampal rat neuron on a silicon chip featuring a linear array of p-type buried-channel transistors after eight days in culture. The open voltage-sensitive gates are located between the source and drain leads. The surface of the chip is chemically and structurally homogeneous, and consists of silica with a surface features smaller than 20 nm. (b) Schematic cross section of a neuron on a buried-channel field-effect transistor with a close-up (drawn to scale) of the contact area. During an action potential, current flows through the adhering cell membrane and along the resistance of the cleft between chip and cell. The resulting voltage in the cleft modulates the source–drain current (Reprinted from [29] with permission). (c) Fluorescence image of a neural network as grown on the CMOS chip; the neurons originated from the rat hippocampal tissue of newborn Sprague-Dawley rats. Neural networks were cultured for 17 days in supplemented serum-free Neurobasal medium. MAP2 (microtubuli-associated protein-2) immunostaining of the neurons was visualized using an FITC-conjugated secondary antibody. The structure of the CMOS chip surface is visible in the background; arrows show the position of the electrode and the contacts to the circuitry [57].

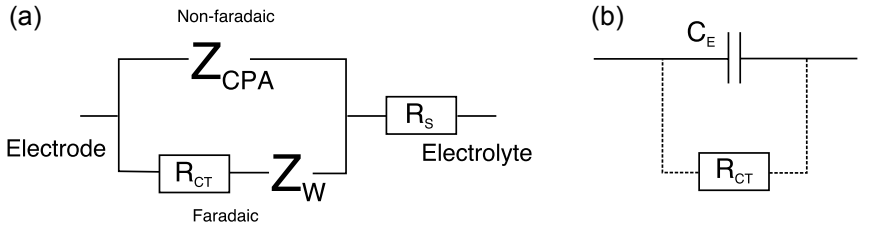


Figure 8.6 (a) Commonly used equivalent circuit model of the electrode-electrolyte interface. (b) Simplified model, which is a suitable approximation for extracellular recordings.

charged particles such as electrons are able to pass across the interface. Charge transfer typically involves oxidation or reduction reactions.

Platinum electrodes in physiological saline are generally modeled using both, faradaic and non-faradaic processes. Figure 8.6 shows the commonly used equivalent-circuit model of the electrode-electrolyte interface. The model consists of a constant-phase-angle impedance, Z_{CPA}, a charge-transfer resistance, R_{CT}, the Warburg impedance, Z_W, and a serial resistance, R_S. These elements will be explained in more detail below.

The immersion of a metal electrode in an electrolyte results in the formation of space charge layers near the electrode surface, the so-called inner and outer Helmholtz planes [73]. These layers of water molecules or opposite-charge ions form a capacitor with a capacitance that can be approximated according to $C_H = \varepsilon A/d$, where ε is the permittivity of the medium, A is the electrode area and d the distance between the electrode surface and the plane through the charged or polarized molecules. Assuming a distance of 0.5 nm, we get a capacitance of 11 μF/m² (note: this is based on the geometric area of a rough surface not on the absolute area of the electrode). A more accurate model can be derived, if the distribution of the charges within a certain distance from the electrode is taken into account, as it is done in the Gouy-Chapman model [73].

Using a constant-phase-angle impedance, Z_{CPA}, instead of a capacitance results in a better agreement of the model with the measurements [74-76]. The empirical non-faradaic constant phase-angle-impedance is given by: $Z_{CPA} = K(i\omega)^{-\beta}$, where K is a measure of the magnitude, and β is a measure of the deviation from purely capacitive behavior ($\beta = 1$). Typically, β is between 0.7 and 0.9 [67, 71]. R_{CT} is the faradaic charge transfer resistance. For platinum, e.g. the charge-transfer resistance is in the range of several hundred MΩ [67]. The Warburg impedance, denoted Z_W, takes into account that ions, produced on the surface of the electrode, need to diffuse away. This impedance is negligible for large charge-transfer resistances, which

produce only small faradaic currents. R_S is the electrical resistance between the electrode and counter electrode. This resistance is on the order of kΩ and can be neglected.

A simplified model version is given in Fig. 8.8(b) consisting of a capacitor, C_E, in parallel with a resistor, R_{CT}. The time constant of this structure is given by $R_{CT} \cdot C_E$ and is outside the signal band of typical cellular signals so that the electrode-electrolyte interface can be approximated by a capacitor. In contrast to the FETs, the dynamics of the cell-electrode junction have to be taken into account since the capacitance, C_E, for metal electrodes is significantly larger than for FETs.

8.2.2.5 Multisite Recording

Metal electrodes and open-gate field-effect transistors can be arranged in 2D-arrays for *in vitro* applications. Conventional and commercially available so-called "microelectrode arrays" (MEAs) are arrangements of usually 60 or 64 circular electrodes with diameters ranging between 10 µm and 30 µm covering an area of ~1 mm^2. While the same metal electrodes can be used for recording and for stimulation, an open field-effect transistor used for recording has to be accompanied by a capacitive-type stimulation spot that then triggers cell or neural activity [51].

8.2.2.6 Extracellular Stimulation

Stimulation capability is a crucial feature, especially in the field of neuroscience to, e.g. expose the neuronal culture to certain stimulus patterns, or to repeatedly apply identical patterns to achieve a "training" of the neuronal colonies, or to do closed-loop experiments with such neuronal cultures. To advance our understanding of network dynamics, we need a large number of bidirectional electrical interfaces to the neuronal network, i.e. stimulation and recording sites. An ideal system for stimulation includes the following features:

- Possibility to stimulate any subset of electrodes
- Flexibility in the stimulation waveform
- Current and voltage stimulation capability
- Capability to do a fast switching from stimulation to recording even on the same electrode

The so-called RACS (real-time all-channel stimulator) by Wagenaar and Potter [52] based on discrete off-chip components meets these requirements.

A monolithic CMOS implementation, which meets all criteria except for the ability to do current stimulation will be presented in Subsection 8.3.5.

Stimulation of neuronal cells can be performed by means of current or voltage pulses. Intracellular stimulation during patch clamp measurements is usually done in the current mode, since the stimulation of neurons in their natural environment also occurs through current inputs to their dendritic tree. For extracellular stimulation both, voltage stimulation [25, 51, 53] and current stimulation [25, 54-57] methods are widely used.

For current stimulation, a voltage-limiting element is required in order to prevent electrolysis, which may change the local pH, produce undesired electrochemistry, or damage the electrode. It has been found that in most cases negative current pulses are very effective to excite neurons to fire action potentials [52, 54]. Also voltage-controlled stimuli can be applied. Biphasic voltage-controlled pulses with the first part of the pulse having a positive voltage sign constitute most effective stimuli. This is easy to understand since the sharp downward voltage transient between the positive and the negative phase corresponds to a strong negative current pulse.

The pulse amplitude and its duration determine the stimulus efficacy [52, 55]. The number of cells that is directly stimulated through a defined, voltage-controlled pulse grows linearly with the amplitude of the respective pulse [52]. The pulse must be strong and long enough to allow the cell membrane and all parasitic capacitances in the system to charge. Normally, the pulse amplitude is in the range of hundred to several hundred millivolts. Another important parameter is the electrode impedance (Subsection 8.2.2.4), since it determines the current for a given voltage transient and the voltage on the electrode for a given current transient, respectively.

Electrical stimulation pulses are on the order of 1 V, whereas the cell signals are typically on the order of 100 μV. The large disparity between the voltage ranges for stimulation and recording leads to artifacts in the recorded signals, which can easily interfere with or completely cover action potentials, even on electrodes that are rather distant from the stimulation site. This stimulation artifact can be removed by using software approaches [58, 59], which are computationally expensive, and often require real-time processing. Moreover, hardware approaches based on active suppression circuits have been successfully developed (see Subsection 8.4.2) [32, 60].

8.3 INTEGRATED CMOS-BASED SYSTEMS

Silicon MEAs based on open-gate field-effect transistors (OGFETs) as electrodes have been developed in the group of P. Fromherz, MPI Munich, [14, 29, 77] and also by Offenhäusser *et al*. [78, 79]. A CMOS-based approach with palladium (Pd) electrodes but yet without on-chip electronics has been reported on by Baumann *et al*. [80]. None of these devices exhibit any on-chip circuitry or electronic components other than the transducers themselves (electrodes, FETs). The use of on-chip microelectronics is imperative for the realization of larger multitransducer or multielectrode arrays with stimulation and recording capabilities. An multielectrode array with integrated multiplexers and in-pixel amplifiers, which can simultaneously record from 4096 electrodes featuring a 20 μm pitch was presented by Berdondini *et al*. [70]. A multitransducer design based on FETs featuring 16384 transducer sites at 7.8 μm pitch was presented by Eversmann *et al*. [11, 35]. Both high-density arrays exhibit high noise levels due to the limited area for the in-pixel amplifiers, and stimulation capability is not included. Electrical stimulation is an important feature to trigger activity in networks of electrogenic cells and to, e.g. study the connectivity in neuronal networks. A CMOS chip with on-chip stimulation capability realized as an integrated stimulation electrode located outside but close to the recording-electrode array has been presented by the group of Kovacs at Stanford [15]. For *in vivo* applications, stimulation and recording probes, each featuring 64 electrodes and 8 channels, have been incorporated in a 3D-electrode array in the groups of Wise and Najafi [16]. A monolithic CMOS MEA featuring 128 bidirectional electrodes, complete on-chip signal conditioning and analog-to-digital conversion has been presented by Heer *et al*. [60]. All these systems will be presented in more detail in the subsequent subsections.

8.3.1 High-Density-Recording Devices

Considering the dimensions of neurons, which range from below 10 μm diameter for vertebrates up to 100 μm for invertebrates, it would be very interesting to conduct biological or electrophysiological experiments at cellular or subcellular resolution so that the availability of high-density arrays is highly desirable. In order to significantly increase the number and density of transducer or electrodes, the integration of active circuitry is mandatory to (i) overcome the connectivity problem by using multiplexers and dedicated readout circuitry, as well as to (ii) ensure signal fidelity and integrity by signal preconditioning (filtering and amplification) on chip.

8.3.1.1 A 128 × 128 CMOS Biosensor Array for Extracellular Recording of Neural Activity

A CMOS-based array featuring 16,384 sensors at a pitch of 7.8 μm has been developed by the Max Planck Institute for Biochemistry, Martinsried, Germany in collaboration with Infineon Technologies Corporate Research, Munich, Germany [35]. The system setup and chip architecture are depicted in Fig. 8.7(c). The monolithically integrated sensor array includes 128 × 128 sensors or pixels. The column decoder periodically selects one of the 128 columns of the pixel array. Moreover, control signals are provided by this decoder, which determine whether the pixels within the selected column are operated in the readout mode or in a calibration mode [Fig. 8.7(a)]. Calibration of all pixels needs to be performed before the pixels are operated in the readout mode, and this calibration has to be periodically repeated after a certain number of readout frames (typically 100). This calibration procedure is needed as a consequence of the mismatch of the transistor characteristics and of leakage occurring at the gate of the transistors.

The full frame readout rate is 2 kfps. Each of the 128 rows is connected to a separate readout amplifier [Fig. 8.7(c)]. The outputs of these amplifiers are connected to the chip output drivers or to dummy loads via sixteen 8-to-1 multiplexers. The 16 output drivers provide an output current to I/V converters arranged on an off-chip printed circuit board. Finally, the buffered output voltages of the I/V converters are converted by 16 analog-to-digital converters with an effective resolution higher than 8 bit, which are part of a Gage PC-based measurement system (www.gage-applied.com).

To integrate analog signal processing units in the pixel is not feasible if high spatial resolution is required. In order to have a high-density array, the pixel circuit shown in Fig. 8.7(a) consists of only three transistors: One transistor is required for sensing the extracellular voltage, another one for pixel selection. Due to the mismatch of the electrical parameters of the relatively small sensor transistors, a calibration transistor shortens the gate and drain of the sensor transistors in the calibration mode of the pixel. The calibration technique is described in detail in [35]. The pixel finally requires 7.8 μm × 7.8 μm area, the electrically sensitive area features a diameter of 4.5 μm so that an array of 128 × 128 pixels can be integrated within an area of 1 mm × 1 mm as shown in Fig. 8.7(b). In order to read out the 16,384 sensors of the array with a reasonable number of lines, the column scanning is extended by a row multiplexing at a ratio of 1:8.

Figure 8.7(d) illustrates the different sensor-site densities of conventional passive metal electrode arrays (25/mm^2) and the CMOS-based sensor array (1.6×10^4/mm^2). The high-density design offers two major benefits: (i) The

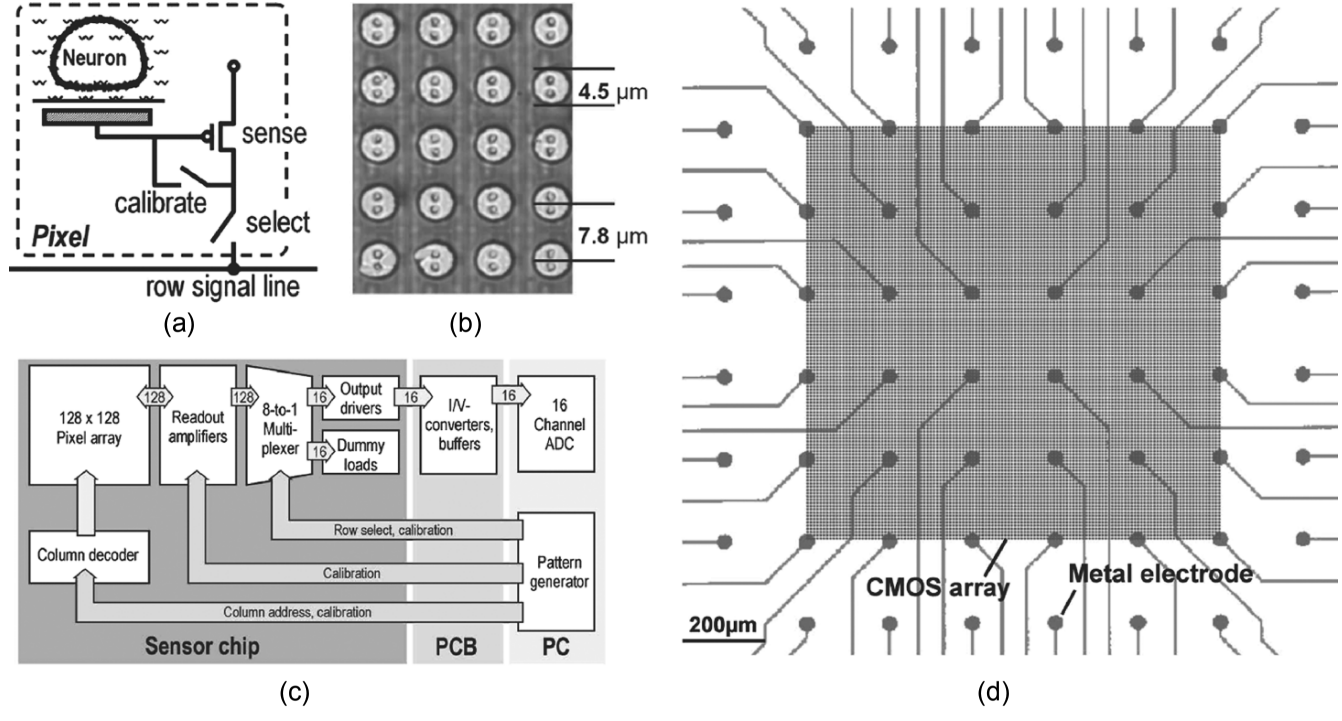


Figure 8.7 High-density sensor array with 16384 sensor pixels on 1 × 1 mm². (a) Pixel circuit. (b) Blow-up of the sensor array showing pixel diameter and pitch. (c) Chip architecture. (d) Comparison of the sensor density of the high-density array with that of a conventional 64-metal electrode array (adapted from [44, 81] with permission).

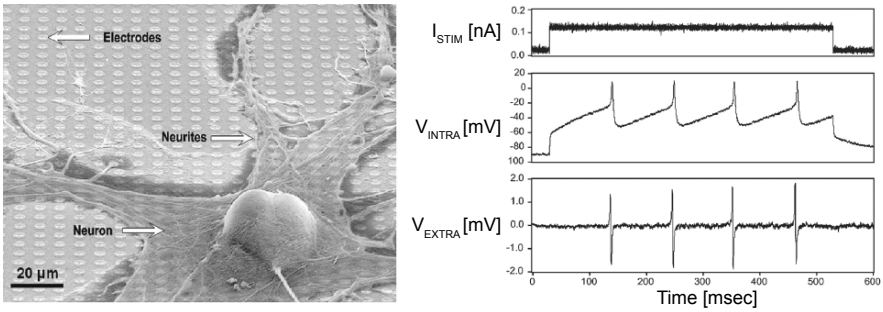


Figure 8.8 Left: Snail neuron on the sensor chip in culture. Right: Measured data from a snail neuron. The neuron is stimulated and monitored for reference purposes with a micropipette. Stimulation current, I_{STIM} , intracellular potential, V_{INTRA}, and extracellular potential, V_{EXTRA}, as recorded with the sensor array (Reprinted from [35] with permission).

yield of successful cell-sensor contacts is hardly influenced by biological dynamics. Cells can be placed randomly or may move during a dynamic network formation without escaping from the sensitive area. (ii) Dense arrays with small pixels enable neurobiological imaging that ranges from the mapping of ion channel densities of individual neurons to the electrical activity mapping of large brain slices.

The extracellular recording capabilities have been tested with neurons from a pond snail, *lymnaea stagnalis*. Figure 8.8 shows the result of a neuron contacted by a an invasive microelectrode and stimulated by the injection of a constant current of 0.1 nA for 500 msec. The intracellular voltage is recorded with the same microelectrode as used for stimulation. Four actions potentials with an amplitude of 60 mV have been observed. The change of the intracellular potential at the beginning and the end of the current injection is a measurement artifact due to the voltage drop, which occurs owing to the microelectrode resistance.

An area of 32 × 40 pixels of the sensor array has been selected for readout with a sample rate of 8 kHz. Data of the extracellular potential recorded from one pixel underneath the neuron are shown in Fig. 8.8. Due to the high-pass transfer function of the cleft, the extracellular signal represents the derivate of the intracellular signal. The same chip has also been used to record field potentials from a cultured brain slice [11]. The noise level has been found to be about 250 μV_{RMS} [11].

8.3.1.2 64 × 64 High-Density Electrode Array for Imaging in vitro Electrophysiological Activity

A second high-density array has been developed at the Institute of Microtechnology at the University of Neuchatel [70]. The aim of this work is to devise high-density metallic microelectrode arrays for high spatio-temporal resolution imaging of the electrophysiological activity of electrogenic cell cultures. Metallic microelectrodes offer the potential advantage of low-noise recordings.

The CMOS design is based on a solid-state active pixel sensor (APS) concept that was originally developed for image sensors. The chip consists of an array of 64 × 64 pixel elements on an overall active area of 2.5 mm × 2.5 mm. Each pixel occupies an area of 40 μm × 40 μm and comprises a square-shape gold microelectrode of 20 μm × 20 μm as well as a pre-amplifier.

The in-pixel pre-amplifier has to fit into the small pixel dimensions so that only circuits with a low number of transistors have been considered. A five-transistor operational transconductance amplifier [5-OTA in Fig. 8.9(b1)] optimized for low noise and small area has been used for the preamplifier. Three additional transistors (T1, T2 and T3) have been integrated in addition to the 5-OTAs in each pixel as can be seen in Fig. 8.9(b2). The presence of these additional transistors allows for operating the pre-amplifier in open- or closed-loop mode. In the open-loop mode, the signal is directly amplified by the open-loop gain, while in the closed-loop mode, the gain is set to one, and the pre-amplifier acts as an impedance transformer.

Electrophysiological measurements have been performed by culturing neonatal rat cardiomyocytes on the chip. These cells show a spontaneous electrical activity already after two days *in vitro* with high-amplitudes (extracellular potentials up to 1–2 mV_{PP}). Signals from several recording sites have been obtained by manually addressing the microelectrodes. The data acquisition was carried out using an oscilloscope. Figure 8.9(c) shows the recorded spontaneous activity from different pixels using different time frames and different sampling frequencies. The recorded spontaneous activity of cardiomyocytes showed signal amplitudes between 130 μV_{PP} and 300 μV_{PP} with a total signal duration (positive and negative phase) of 2–3 msec. It has to be noted that the signal shapes are affected by the low sampling rate of the oscilloscope. The overall noise of the system amounts to 80 μV_{PP}. The chips have been reported to be reusable after cleaning the devices with isopropanol and after rinsing in DI water.

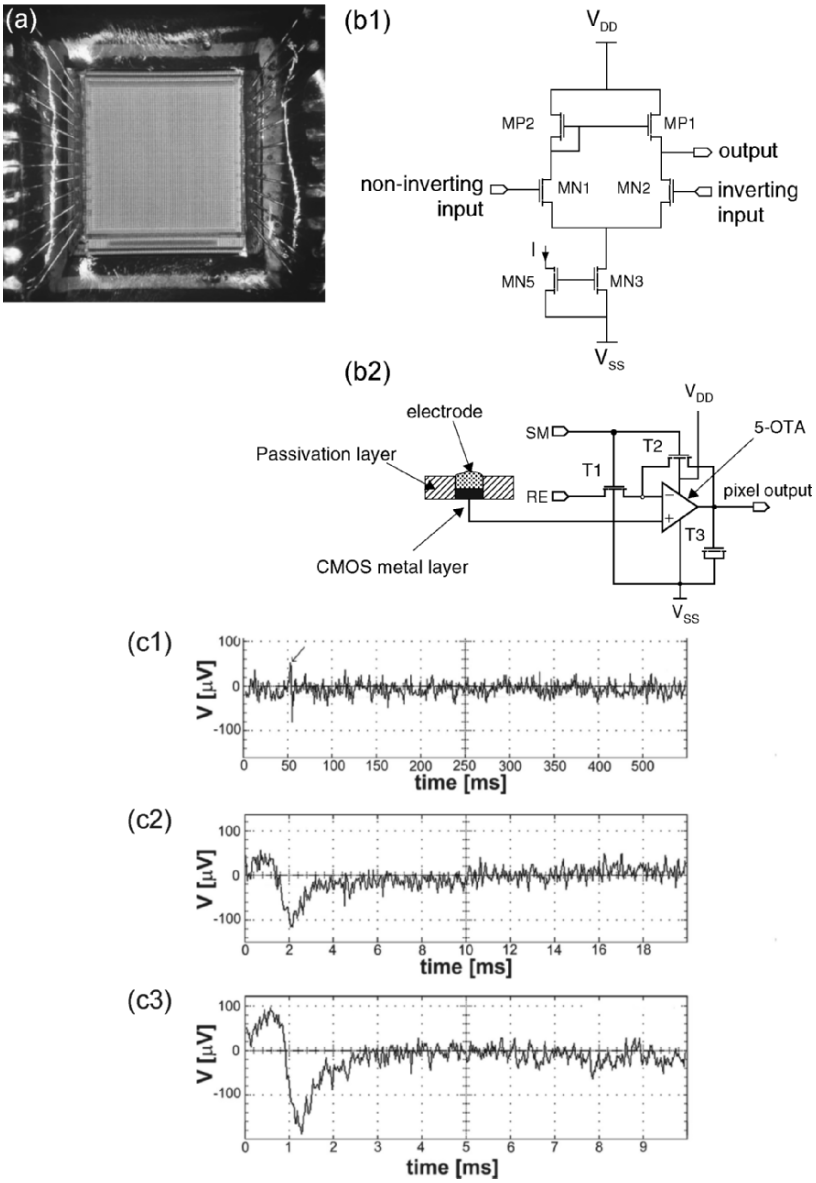


Figure 8.9 (a) Micrograph of a packaged high-density chip. (b1) Schematic of the in-pixel preamplifier and (b2) a detailed schematic of the OTA used in the preamplifier. (c) Recordings of spontaneous electrical activity of rat cardiomyocytes cultured on the chip for 2 days from different pixels at different time scales: (c1) Sampling rate of 1 kS/s, signal amplitude of 136 μV_{pp}. (c2) Sampling rate of 25 kS/s, signal amplitude of 168 μV_{pp}. (c3) Sampling rate of 50 kS/s, signal amplitude of 284 μV_{pp} (Reprinted from [70] with permission).

8.3.2 Multiparameter Sensor Chip

A very interesting approach includes the integration of several different sensor types on a single chip for monitoring not only the electrical cell activity, but also the temperature, the cell metabolism and the cell adhesion. Such a multiparameter sensor chip has been developed by the University of Rostock in cooperation with the semiconductor company Micronas (www.micronas.com) [80, 82]. The aim is to develop a sensor system, with which metabolic parameters as well as electrical signals can be measured with one sensor chip. The applications are in the field of basic research and drug screening.

The cellular environment of living cells in *in vitro* investigations differs considerably from their native environment *in vivo*. The so called Cell Monitoring System (CMS®) has been developed as a first approach to online monitoring of cellular reactions under well controlled experimental conditions. It allows for the parallel and non-invasive measurement of different parameters of cellular systems by means of microsensors.

These CMS sensor systems include:

- Cell-potential field-effect transistors (CPFET, sensitive gate areas of $6\times1\ \mu m^2$) and palladium electrodes (10-µm diameter) to measure the electrical cell activity, as shown in Fig. 8.10(b) [80]. The palladium coating was performed in a self-adjusting back end process. [80].
- A temperature sensor to monitor the temperature of the cell culture.
- Ion-sensitive field effect transistors (ISFETs) to monitor the pH in the cellular microenvironment, as shown in Fig. 8.10(c) [83]. The ISFETs allows for monitoring local acidification and respiration in *in vitro* cell networks.
- Interdigitated electrodes to measure the cell adhesion by means of impedance measurements [84]. The quality of the contact between the electrically active cells and the transducers is of pivotal importance for applications in basic and biomedical research. Impedimetric measurements using interdigitated electrode structures have been found, according to the authors, to provide information on the cell number, the cell adhesion and the cellular morphology since an AC current between the electrodes is influenced by the presence and structural properties of living cells growing on these electrode structures.

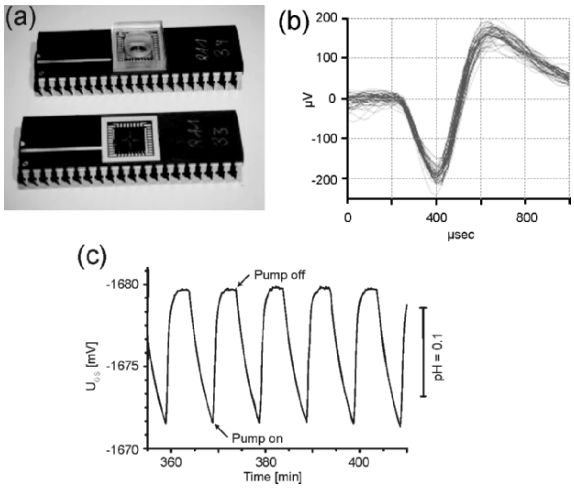


Figure 8.10 (a) CMOS chip mounted on a 40-pin IC ceramic socket. (b) Selection of extracellular recordings from one of the chip electrodes featuring quite good S/N ratios; the figure shows ~40 superimposed action potentials. (c) Extracellular acidification measurements in a neuronal network on a silicon chip as performed with ISFETs in a flow-through system. The acidification was measured during the time when the pump was shut off. When the pump was on, the medium was completely exchanged with fresh medium (Reprinted from [80] with permission).

8.3.3 Portable Cell-based Biosensor

A field-portable and a hand-held cell monitoring system have been developed at the Department of Electrical Engineering at Stanford University [15]. To realize field-portable detectors for chemical and biological agents, considerable changes must be made relative to the equipment used in typical laboratory-based situations. First, the entire system must be robust, relatively compact, and able to self-power for a reasonable operational period. Second, the cells must be packaged in a way as to maintain sterility and proper environmental conditions for the culture, and yet allow the introduction of unknown agents. Typically, the cells, microelectrode arrays, and potentially some fluidic interfaces are combined into a user-replaceable cartridge. The cells must also be maintained at constant pH, osmolarity, and temperature, typically via chemical buffering, humidification, and closed-loop thermal control [7, 24, 85]. The high sensitivities of cells to changes in these parameters can lead to false "signals" if they are not tightly controlled. In addition, sample preparation, which is typically done manually in the laboratory, must be automated so that it can be carried out under realistic environmental conditions.

Beginning with the first such work presented in [7], there has been steady progress toward realizing such practical, field-portable systems. Gilchrist *et al*. demonstrated a field-portable incubator and a cell-based biosensor data recording system in field tests and were able to acquire and to pharmaco-

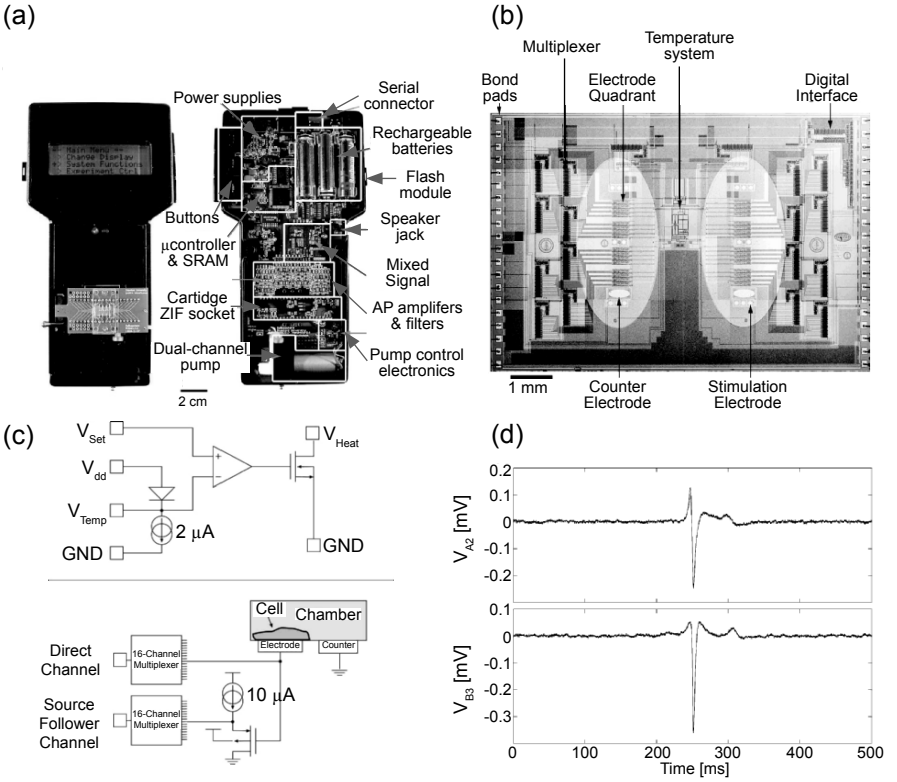


Figure 8.11 (a) Handheld cell-based biosensor system. Front view showing the cartridge, pump, display, and control buttons; back internal view showing the electronic subsystems on the right. (b) Microphotograph of the 6×9 mm^2 die. The lightened elliptical shading was added to highlight the cell chamber surface areas. Each chamber has four arrays (quadrants) of 16 microelectrodes each. The serial digital interface controls the eighteen 16-channel multiplexers. (c) Circuit block diagrams of the CMOS cell-cartridge. [Up] Temperature control system showing components and off-chip connections. [Down] Block diagram of one microelectrode sensor channel showing the unbuffered and source-follower-buffered connections. (d) Extracellular action potential waveforms recorded with the handheld system from HL-1 cells (cardiomyocyte cell line) growing for 8 days in chamber A (V_{A2}) and chamber B (V_{B3}) of a cell-cartridge. These signals were averaged 16 times and are of typical shape and amplitude (300–400 μV_{PP}). Reprinted from [24] with permission.

logically modulate action potential signals in a desert setting [85]. The field portable system is composed of two modules. A first module is a cell-transport case, which, through closed-loop thermal and gas control maintains the viability of the cells on the individual electrode arrays. The second module is a data acquisition system, which consists of the necessary electronics to amplify and filter the small action potential signals, to digitize the signals from multiple electrodes, and to stream the resulting data into a computer hard disk for analysis. Both modules must be fully self-contained, with on-board power, and be ruggedly packaged to survive the rigors of a transport to realistic settings.

A hand-held recording system (Fig. 8.11) with integrated cell cartridges incorporating closed-loop thermal control and front-end amplification and multiplexing electronics was also developed [24]. For cell types that require temperature regulation, shrinking the controlled volume has definite advantages in terms of power consumption and, hence, weight and volume. A die microphotograph of the CMOS chip used in this portable device with a floor plan overlay is shown in Fig. 8.11(b). Each chamber has one large pseudo-reference electrode, two stimulation electrodes, four arrays of 16 microelectrodes, multiple temperature sensors, and nine 16-channel multiplexers. Digital logic, which uses a three wire-serial interface, controls the multiplexers. The PDMS gasket seal contacts the top surface of the die except for the two elliptical chamber areas and the bond pads.

The temperature is regulated by an on-chip temperature controller. The goal of the temperature regulation system is to maintain the substrate beneath the cells at a defined temperature even when exposed to environmental transients. An NMOS transistor is used as the heating element [Fig. 8.11(c)].

The primary biological sensors of the cell-cartridge are multiple arrays of 16 gold-coated planar microelectrodes of 10 μm diameter. Both chambers have four arrays of 16 electrodes, resulting in 128 total sensor channels on a single cartridge. Each microelectrode has a connection to two separate multiplexers. One connection is directly to a multiplexer (for platinization, impedance measurements, and unbuffered action potential measurements), while the other connection goes to a low-noise source follower, the output of which is routed through the second multiplexer (for buffered action potential measurements). This configuration allows for simultaneous recordings from four buffered and four unbuffered microelectrodes per chamber. A large pseudo-reference/counter electrode in each chamber is used to hold the solution at a nearly constant potential during the measurements. The analog portion of the handheld device includes low-noise amplifiers, the

auxiliary channel amplifiers and A/D converters (temperature, system power, reference voltage, battery voltage), the stimulation system, and a pump control system. The low-noise amplifiers connect to the 8 channels of buffered source-follower cartridge outputs. The amplifier system was designed to provide a gain of 505 across a 16 Hz–3 kHz bandwidth with 1 μV_{RMS} of input-referred noise. Signals from HL-1 cells (cardiomyocyte cell line) were recorded at 8 days *in vitro* [Fig. 8.11(d)]. The typical amplitude of the recorded signals was in the 400 μV_{PP} range with the largest observed signal amplitudes being 1.3 mV_{PP}.

8.3.4 Wireless Implantable Microsystem

Needle-type probes for recording and stimulation featuring 64 electrodes and 8 channels have been developed and also incorporated in a 3D-electrode array [Fig. 8.12(a)] for *in vivo* applications at the Engineering Research Center for Wireless Integrated MicroSystems, at the University of Michigan [16, 17].

Silicon micromachined electrode arrays permit the long-term monitoring of neural activity *in vivo* as well as the insertion of electronic signals into neural ensembles or networks at the cellular level. Such electrode arrays are facilitating significant advances in the understanding of the nervous system, and merged with on-chip circuitry, signal processing, microfluidics, and wireless interfaces, they are forming the basis for a family of neural prostheses for the possible treatment of disorders such as blindness, deafness, paralysis, severe epilepsy, and Parkinson's disease.

Wireless operation of implantable systems is key to their successful deployment in clinical applications. Wires, typically used for power and data transfer between the implant and the outside world, are a primary source of infection, failure, manufacturing cost, and discomfort to the patient. Wireless transmission of power and data circumvents all of these problems. Power and data signals can be transmitted using electromagnetic radio frequency (RF), infrared, or acoustic energy; however, wireless telemetry based on RF transmission between two closely coupled coils is most commonly used. A typical wireless interface must satisfy several basic requirements. First, sufficient power has to be transmitted to the implant to enable operation of its circuitry and, in the case of stimulation, deliver the respective charge to the tissue. Second, the telemetry technique used must feature a sufficient range. This requirement depends on the application, but a range of a few centimeters is adequate for most prosthetic applications. Third, the wireless link should provide a high data transfer rate (bandwidth). This

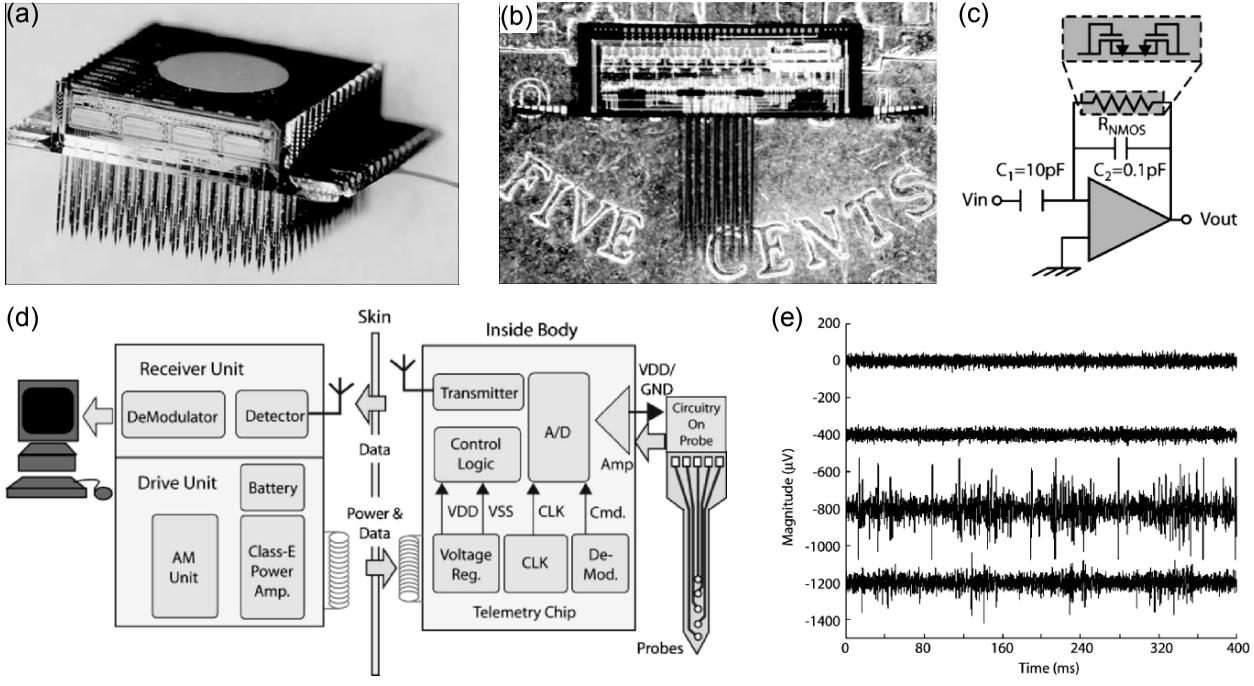


Figure 8.12 (a) 3-D 1024-site, 128-channel neuroelectronic interface. (b) A 64-site 8-channel recording probe, shown on a U.S. nickel. The probe dissipates 0.8 mW while providing a gain of 40 dB over a bandwidth of 10 Hz to 10 kHz with an input noise of less than 10 μV. (c) A schematic of the recording amplifier is shown at the right. (d) Block diagram of an inductive RF telemetry link for an implantable microsystem. (e) Recordings from a guinea pig cochlear nucleus using a multisite, multishank, multiplexed, buffered probe. The shanks providing the top two signal traces were not in the neural tissue and show the noise of the recording system without any background neural noise (Reprinted from [16] with permission).

requirement is also application-dependent, although in most emerging recording and stimulating systems, bandwidths in excess of 1–2 Mb/s are needed. Fourth, the telemetry approach chosen should be immune to most environmental conditions and should be able to pass through tissue. Finally, the wireless link should be adaptable so it can satisfy the needs of different applications.

For many neuroscience applications, the desired structure of a wireless implantable microsystem would be similar to that shown in Fig. 8.12(a). Here, a 3-D array of electrodes interfaces with the tissue and resides in a platform that rests on the cortical surface. The platform supports electronics for signal processing (e.g. spike identification and event recognition) and for wireless communication with the outside world. To avoid significant heating of the tissue, the *in vivo* power dissipation should be limited to 20–30 mW, depending on the size of the array.

Figure 8.12(b) shows a typical 64-site recording probe. The probe dissipates 0.8 mW and has a circuit area of 4.3 mm^2 in 3 μm CMOS technology. The amplifier must provide a stable gain and limit the high-frequency response to around 12 kHz to prevent aliasing in the multiplexer. It must also AC-couple the signal to minimize offsets and suppress the unstable DC potential of the site. It must provide these features while not increasing the overall system noise, occupying very little area, and dissipating less than 10 mW to avoid appreciable heating of the tissue. Temperature rises of more than 2°C would damage the surrounding neurons. The capacitively coupled recording amplifier shown in Fig. 8.12(c) provides these features. A 10 -pF input capacitor and a 100 fF feedback capacitor set the midband gain at 40 dB. A diode-connected subthreshold nMOS transistor in the feedback loop sets the lower band-pass corner at less than 10 Hz, while a Miller capacitor in the operational amplifier sets the high-frequency limit. The equivalent-input noise is less than 10 μV [86], and the amplifier [Fig. 8.12(c)] dissipates less than 100 μW from 1.5 V supplies.

Figure 8.12(d) shows the diagram of a wireless telemetry link. Power and program data are transmitted via an inductive link to an implanted antenna (receiver). The implanted circuitry generates a DC power supply from the RF carrier, demodulates program data from the carrier, and generates a clock signal with which to operate the electronics. The data from the probe are digitized and transmitted back to the outside world over a second data link, often at a higher frequency. The external circuitry receives the transmitted data from the implant and reconstructs the transmitted signals.

Figure 8.12(e) shows neural activity as recorded with a multiplexed buffered probe in a guinea pig cochlear nucleus, driven by white noise bursts.

Channels 3 and 4 show neural activity with a high SNR; no neural activity is recorded on the first two channels because the corresponding sites were not in the tissue due to the convexity of the cortical surface. These channels (which were nonetheless in the fluid) show the noise of the system without any background neural activity. For a bandwidth from 300 Hz to 3 kHz, the overall noise level of the system is less than 9 μV.

8.3.5 Fully Integrated Bidirectional 128-Electrode System

A monolithic microsystem in CMOS technology that provides bidirectional communication (stimulation and recording) to cultured electrogenic cells is presented in this subsection. The microchip can be directly used as a substrate for cell culturing, it features circuitry units per electrode for stimulation and immediate cell signal treatment, and it provides on-chip signal transformation as well as a digital interface so that a very fast, almost real-time interaction (2 msec loop time from event recognition to, e.g. a defined stimulation) is possible at remarkable signal quality. The corresponding spontaneous and stimulated electrical activity recordings with neuronal and cardiac cell cultures will be shown. The system can be used to, e.g. study the development of neural networks, to reveal the effects of neuronal plasticity and to study cellular or network activity in response to pharmacological treatments. After a short system overview, the design, implementation, fabrication and use of the complete system will be presented.

8.3.5.1 System Overview

Figure 8.13 shows a micrograph of the monolithic CMOS microsystem. The 6.5 mm by 6.5 mm chip comprises 128 stimulation- and recording-capable electrodes in an 8×16 array and an integrated reference electrode. The system is structured in a modular design (Fig. 8.13 and Fig. 8.16). Each pixel of the micro-electrode array incorporates the signal-transducing electrode, a fully differential band-pass filter for immediate signal conditioning, a mode storage unit and a buffer for stimulation. The pitch of the pixel units is 250 μm. Electrode pitch, size and shape are very flexible, and the electrode material can be selected from a large variety of materials, since the electrodes are realized during the post-CMOS processing, as described in Subsection 8.3.5.2. A digital control unit is also integrated on the chip and controls the multiplexing, the electrode selection for stimulation, the reset of single electrodes, and it contains the successive-approximation registers of the A/D converters and the interface to the outside world. Implementing

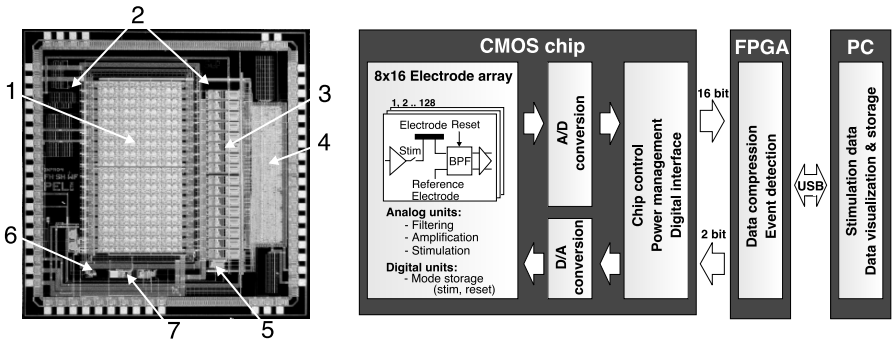


Figure 8.13 Left: Micrograph of the CMOS system chip (6.5×6.5 mm^2). The different components include: (1) 128-electrode array (8 × 16) including circuitry units per electrode; (2) space for on-chip platinum reference electrode; (3) sixteen analog-to-digital converters; (4) digital control and interface; (5) digital-to-analog converter for stimulation; (6) temperature sensor; (7) biasing circuitry. Right: Schematic of the setup for the CMOS microsystem: CMOS chip with 128 microelectrodes, repeated analog units, analog-to-digital converters, a digital-to-analog converter, chip-control and power-management unit, and digital interface. A FPGA board for data compression and event detection is interfaced on the one hand to the chip, and, on the other hand, via USB to a computer for further data treatment, visualization and stimulus generation [57].

filters and buffers at each electrode offers important advantages: (i) The signal is amplified and filtered in close proximity to the electrodes, which makes the design less sensitive to noise and interference picked up along connection lines; (ii) a buffer per electrode renders the stimulation signal independent of the number of activated electrodes; (iii) the high-pass filter removes offset and drift of the biochemical signals so that the signal can be amplified before it is multiplexed; (vi) the low-pass filter limits the noise bandwidth and works as an anti-aliasing filter for the multiplexing and for the subsequent A/D-conversion.

All in-pixel circuitry components have been optimized for low noise and small area. A total equivalent input noise of the pixel circuitry of 11.7 μV_{RMS} (0.1 Hz to 100 kHz) has been measured. The electrodes are continuously read out at a sampling rate of 20 kHz per electrode. A gain of 1000 or 3000 can be selected. The overall power consumption of the chip is 120 mW at 5 V supply. Electrogenic cells are very sensitive to temperature so that temperature changes may change the cell activity level and may even lead to cell death. An on-chip temperature sensor monitors the chip operating temperature. Operating the chip with liquid on the surface leads to a temperature rise of less than 1°C with respect to ambient temperature upon operation, so that additional cooling of the system is not required. Circuits operating

at low frequency (down to 1 Hz) might be sensitive to leakage currents, the effect of which has been reduced by the fully differential design of the in-pixel readout circuitry. Furthermore, electromagnetic coupling is also generally reduced in a fully differential architecture. The electrode units also provide stimulation capabilities. Any arbitrary stimulation pattern (with a maximum sampling rate of 60 kHz and 8 bit resolution) can be applied to any subset of electrodes. The readout circuitry at each electrode can individually be reset to its operating point in order to suppress artifacts evoked by the stimulation pulses from the stimulated electrode itself or from neighboring electrodes.

8.3.5.2 Fabrication and Packaging

The 6.5×6.5 mm^2 chip has been fabricated using an industrial 0.6 µm CMOS process with three metal layers, two polysilicon layers and a high-resistance polysilicon layer at XFAB, Germany (www.xfab.com). Special post-CMOS processing and packaging steps are necessary to make the packaged chip ready for operation under physiological conditions in liquid phase.

Due to a limited selection of materials available in a commercial CMOS process, the electrode material as received from the foundry is aluminum, a known neurotoxicant. A two-mask post-processing procedure was applied to cover the Al electrodes with biocompatible platinum. During this processing the electrodes were shifted (Fig. 8.14), and the aluminum was sealed with the electrode metal and a passivation stack. The metal electrode that is exposed to the liquid consists of 50 nm titanium/tungsten, an adhesion promoter, and 270 nm of Pt. These metals were sputter-deposited and structured in a lift-off process.

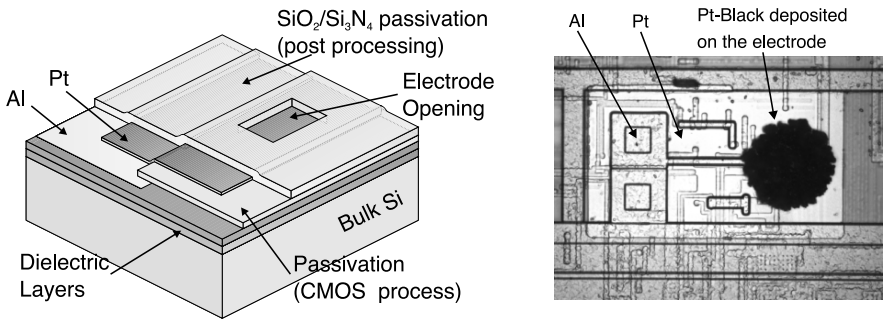


Figure 8.14 Post processing of the microelectrode. A schematic representation of the relevant layers is given on the left, and a micrograph of the shifted electrode is shown on the right side (Reprinted from [71]).

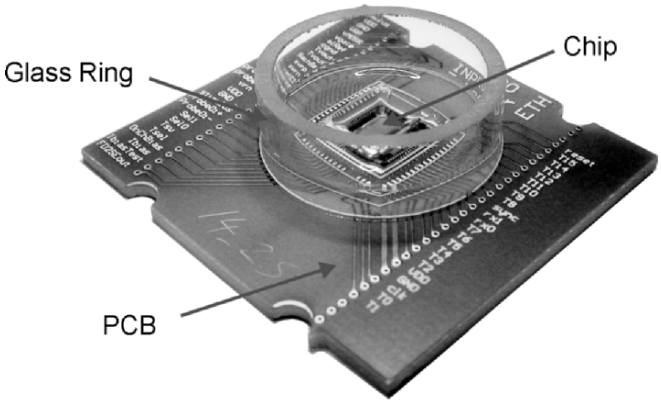


Figure 8.15 Packaged bioelectronic CMOS chip ready for measurements [60].

In order to prevent the formation of pinholes in the chip protection layer an alternating Si_3N_4/SiO_2 passivation stack consisting of 1 µm of Si_3N_4, and two times 100 nm SiO_2 and 200 nm Si_3N_4 (total layer thickness of 1.6 µm) has been deposited using plasma-enhanced chemical vapor deposition (PECVD). A mixed-frequency PECVD process was used to match the stress with that of the underlying Si_3N_4 deposited during the CMOS process. A simpler solution featuring a single 1 µm Si_3N_4 layer has been attempted, but proved to be ineffective in sealing the Al [71]. A reactive-ion-etching (RIE) step was then used to open the nitride stack and to define the size and shape of the electrodes. This increases design flexibility since the electrode dimensions are not defined by the CMOS process. Neurons have been cultured for many months on the chips with the 1.6 µm passivation stack, and the chips have been re-used with no degradation effects observed.

The processed chip has then been mounted and wire-bonded on a custom-designed printed-circuit board (PCB) (Fig. 8.15). A water-resistant medical epoxy (EPOTEK 302-3M, www.epotek.com) was used to encapsulate the bond wires and pads. A glass ring forms a bath that contains a suitable amount of cell medium. After packaging, the chips were sterilized with ethanol and exposed to UV light under a sterile bench.

To reduce the electrode impedance Pt-black can be electrochemically deposited using 1.0 nA/m^2 current density in a solution containing 7 mM hexachloroplatinic acid, 0.3 mM lead acetate and hydrochloric acid to adjust the solution pH to 1. A platinum wire has been used as the counter electrode connected to an external current source, and the on-chip stimulation circuitry was used to set the electrodes to a fixed potential.

8.3.5.3 Recording Electronics

The circuitry was designed to meet the specifications of extracellular capacitive recordings of neural or cardiac cell activity. The signal frequency range of an action potential is from approximately 1 Hz to 5 kHz, and expected signal amplitudes on the electrode are between 100 μV and 2 mV, both depending on the cell type. The minimum Nyquist rate is therefore 10 kHz and a sampling frequency of 20 kHz per electrode has been chosen in order to achieve a good resolution of the action or field potentials.

Another issue is the resolution of the A/D converter. The expected maximum signal amplitude is 500 μV for single action potentials (single cell signals) and several millivolts for field potentials (signals from an ensemble of cells). The noise of the electrodes is in the 5-10 μV range [21, 87, 88]. To achieve a good signal-to-noise ratio, an 8-bit resolution for the A/D converters seems to be sufficient.

Offset and drift of the electrode potential need to be eliminated in order to be able to amplify the small signals. To this end, a high-pass filter has been used as the first read-out stage. The signals then pass a low-pass filter with a corner frequency of about 5 kHz to prevent high-frequency aliasing and to limit the noise bandwidth. Finally, a buffer with a larger bandwidth has been realized to further amplify the signal and to allow for multiplexing.

8.3.5.3.1 Pixel Readout Circuitry

The first readout stage is a high-pass filter (see Fig. 8.16) to eliminate drift and offset of the metal electrode. The gain of the filter is 20 dB defined by the capacitance ratio C1/C2. The corner frequency is given by the capacitor C2 and the Resistor R. The value of C2 is a trade-off between gain accuracy and corner frequency. Given the spatial constraints imposed by the 250 μm pitch, a 200 fF capacitance was chosen for C2. To realize a corner frequency of 1 Hz a large resistance, R, is required. Such large resistance has been realized by using a MOS resistor (see, e.g. ref. [89]) with a gate length of 100 μm and a gate width of 1 μm. The absolute value of the resistor can be externally controlled by changing the gate voltage. The corner frequency can be tuned from approximately 1 Hz to about 1 kHz. A second design that makes use of a MOS diode to realize such large resistances has been implemented as a test structure (compare refs. [88, 90]). For the MOS-diode-based approach no reference voltage is required, but the filter stage is not tunable as it is with the MOS resistor. The corner frequency of the diode-based design was measured to be approximately 1 Hz.

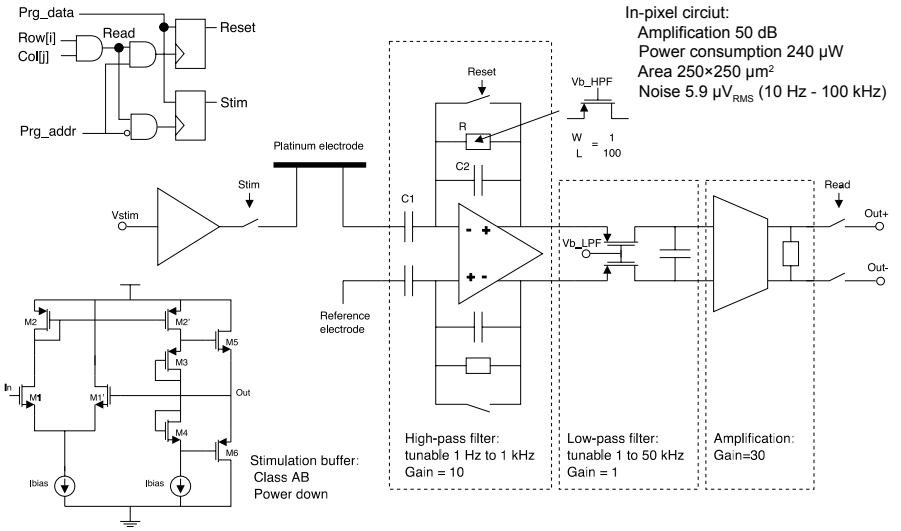


Figure 8.16 Circuitry unit, which is repeated at each electrode. Each unit comprises a stimulation buffer, a high-pass filter, a low-pass filter and a final amplification stage. The low-frequency corners of the filters are realized by using MOS resistors. The cut-off frequency of the filters can then be adjusted by changing the gate voltages (Vb_HPF and Vb_LPF) of the MOS resistors (adapted from [57]).

In the subsequent filter stage, a passive MOSFET-C low-pass filter (LPF) limits the noise bandwidth and prevents aliasing. The area (90×45 µm^2) of the passive low-pass filter was minimized by using MOS-resistors that are tunable by modification of the gate voltage and by using MOS-capacitors. The corner frequency of the filter can be tuned from about 1 kHz to 30 kHz. Following the LPF, an additional gain stage amplifies the signals and allows for fast multiplexing. This buffer is a fully differential version of the amplifier presented in [91] and has a gain of 30 dB.

As shown in Fig. 8.16, each electrode has a small digital circuit to control the pixel operation (Read, Reset, Stimulate). It is possible to set one of the flip-flops every 50 µsec using the signals Row[i], Col[j], Prg_data and Prg_addr, which are produced by the digital core.

Electrical stimulation pulses will affect the neural recordings on the stimulating electrode and on neighboring electrodes by creating large artifacts, which may superimpose on action potentials occurring during the decay time of this artifact. There are several recent papers on eliminating the stimulation artifact [58, 59]. However the methods are computationally expensive, and often require real-time data processing. In the design presented here, an artifact cancellation is implemented directly in the pixel circuitry. The high-pass filter saturates, when large signals are applied. The

recovery time is given by the time constant of the feedback, R and C2, and is, therefore, rather long. However, when the Reset switch is closed, the OTA is in the buffer mode and brings the filter back into the operation range within less than 100 μsec.

Each row comprising eight electrodes is connected to one A/D converter. A single-ended A/D converter is used so that the fully differential signals need to be translated into single-ended signals. For each row, a switched-capacitor amplifier is used to convert the fully differential signal to a single-ended signal. It features an operational amplifier offset voltage cancellation. Clock feed-through and charge injection are also reduced by additional switching circuitry and by the differential design [92]. The signal is then digitalized by a successive-approximation A/D converter with 8-bit resolution. All sixteen successive-approximation registers of the sixteen converters are implemented in the digital part. The intention is to completely separate all clocked digital circuitry from the analog parts.

8.3.5.4 Stimulation Electronics

For current stimulation, a voltage-limiting element would be required in order to prevent electrolysis, which may change the local pH or damage the electrode. Voltage stimulation was implemented in this design, because the applied voltage at the electrode during stimulation is then well known, and no electrolysis can occur as long as the voltage is small enough. Further considerations concerning the stimulation electronics included that the number of activated stimulation electrodes and, therefore, the load for the stimulation circuitry can vary. This makes it necessary to implement a buffer at each electrode.

The electrode impedance is important for the circuit implementation of the stimulation buffer. Studies on the electrode impedance have been reported in [67, 71, 93], and it is evident that the impedance of the electrode varies with its size and material and as a consequence of cell growth [94]. The capacitance of a 30-μm diameter electrode has been measured to vary from 200 pF for blank platinum (Pt) to 20 nF for Pt-black. The charge-transfer resistance of the Pt-electrode is on the order of several 100 MΩ . Sampling rates for the stimulation as reported in the literature are as high as 10 kHz when a square wave tetanus signal has been applied to the electrode [20]. In order to be able to stimulate with a variety of waveforms, a stimulation sampling rate of 60 kHz has been chosen, which entails the use of a buffer with a slew rate of 3V/16 μsec = 0.2 V/μsec at a load of 20 nF. Since the overall power consumption should be kept low and the needed chip real estate small, we decided to implement a buffer with a class-AB output stage

(Fig. 8.16), which is powered down if the electrode is not selected for stimulation.

The stimulation buffer (Fig. 8.16) comprises a differential input stage (M1, M1') with an active load (M2) and is connected in a unity gain configuration (OUT connected to gate of M1'). The class-AB stage is formed by the transistors M3, M4, M5 and M6, where M3 and M4, both of which are diode connected, bias the output transistors M5 and M6 [71]. The power consumption is 45 µW for very small inputs at a current, I_{bias}, of 2 µA, and a supply voltage of 5 V. However, when slewing, the amplifier can deliver up to several milliamps to the electrode. The whole buffer occupies an area of

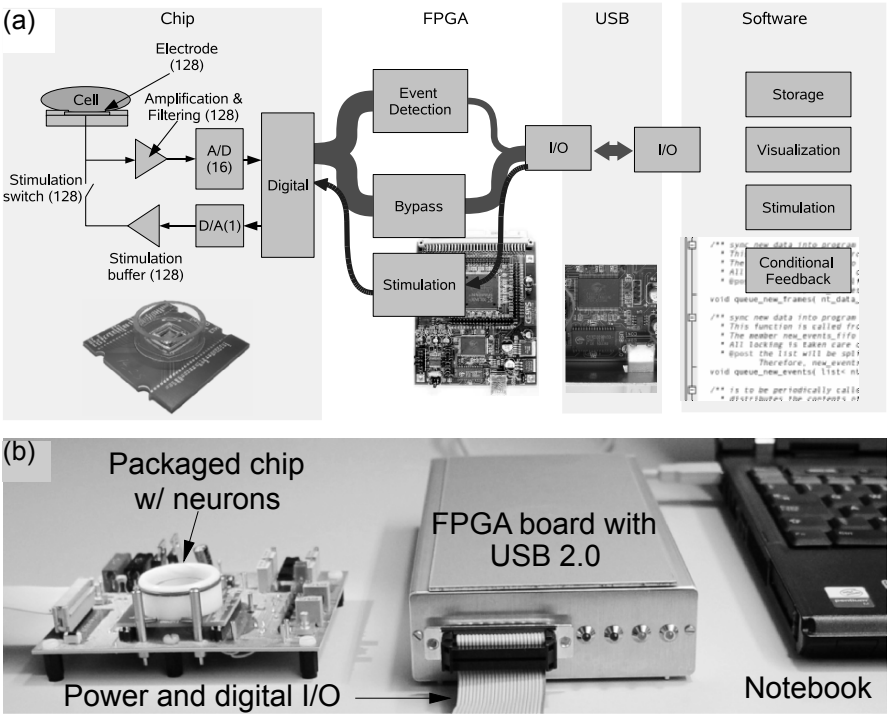


Figure 8.17 (a) System overview and data flow. Left: Recorded data propagate from the left to the right (chip to PC), while stimulation data propagate from the right to the left (PC to chip). The outermost left block labeled "Chip", includes many components in multiple realizations; the numbers in brackets indicate the number of realizations per chip, e.g. 128 electrodes or 16 A/D converters. For the sake of clarity, the analog signal processing chain is highly simplified. A more detailed chip schematic is given in Fig. 8.16. (b) Photo of the entire system setup. The PCB provides a stabilized power supply and reference voltages for the CMOS chip [95].

42 μm × 32 μm. Transmission gates are used as switches (Stim) to connect the buffer to the electrode.

8.3.5.5 Digital Control, FPGA and Software

The digital circuitry unit on chip controls the multiplexing, the electrode selection for stimulation, and the reset of single electrodes. Moreover, it controls the converters and it interfaces with the outside world (Fig. 8.13 and Fig. 8.17). To manage the data output rate of 3.2 MB/s and the input rate of 0.4 MB/s, an FPGA running at 48 MHz in conjunction with an USB 2.0 interface chip and 1 MB SRAM has been utilized. I/O buffering and digital signal processing like low-pass filtering and event detection are implemented on the FPGA to reduce the data volume transmitted to the PC.

Event detection is mandatory since data volumes are beyond the bandwidth of a human observer. In [96] a review of common approaches is given. The approach as presented in this paper is realized as a two-stage processing:

1. During *event detection*, events are isolated from the continuous flow of the data. In this context an event is a segment of recorded data that potentially contains a single spike, multiple (possibly overlapping) spikes, or a spike train or barrage.
2. During *spike discrimination*, events are further processed to isolate spikes, to perform a unit separation and to resolve overlaps.

The benefits of performing event detection on the FPGA as opposed to a PC are twofold. First, the PC is relieved from approximately 32 million arithmetic operations per second. Second, the load on the I/O system is decimated, which is a prerequisite for a future realization of larger arrays. The freed PC resources can be utilized for more complex tasks like on-line spike sorting, which are hard to perform in hardware. Additionally, the latency time of the whole system is reduced.

The architecture of the chip allows for selecting an arbitrary set of electrodes for stimulation with the same stimulus waveform. Reconfiguration of the stimulation subset and reset of the filters proceed at the speed of the readout of the array, since the addressing scheme for the readout is also used for the programming of the electrodes: whenever an electrode is sampled, either its Reset state or its Stim state may be reprogrammed. Consequently, the array may be reconfigured from the stimulation of a certain set of electrodes to the stimulation of any other set of electrodes within 50 μsec. In case that simultaneous stimulation of electrodes with different stimulus waveforms is desired, two consecutive simulations with a small delay of

50 μsec have to be performed, which, in most cases, can be considered to be simultaneous with respect to the time scale of the cellular responses or processes in, e.g. a neuronal culture.

The software is a crucial component of the system. It organizes the USB data handling, the spike sorting, the data visualization, the stimulus generation and the automated feedback generation. Low latency time and high data throughput are key requirements. It is implemented in C++ using the Qt library (Trolltech, Inc., Oslo, Norway) for the graphical user interface and runs on Linux. A program named *Neurotalker* runs in 3 main threads: the I/O-handling, the event processing, and the user interface [95]. The I/O thread interfaces the USB subsystem to the main loops of the program. Since USB is a pure Master-Slave bus, and since it relies on block transfers, a further protocol layer is required to have the capability to transfer data packages of arbitrary size at any defined time. A combined software and VHDL package, named ce-usb-kit that provides this additional protocol layer, has been developed and encapsulated into an independent library containing both VHDL and C code. It is available from the *sourceforge* project (http://sourceforge.net/projects/ce-usb-kit). The *ce-usb-kit* provides a VHDL code that includes an external-RAM I/O buffer, a receiver finite-state machine that decodes data packages sent by the PC, a sender finite-state machine that encodes data, performs padding, and calculates an 8-bit cyclic redundancy checksum [95].

8.4 MEASUREMENT RESULTS

8.4.1 Recordings from Neural and Cardiac Cell Cultures

In a first experiment primary neonatal rat cardiomyocytes were cultured on the chips, since these cells become electrically active very quickly and provide large signal amplitudes. Recordings from cardiomyocytes can be obtained after three days in culture, whereas dissociated neurons show larger-amplitude spikes in most cases after about two to three weeks. For both cell types, the recording parameters included a bandpass filter range between 10 Hz and 5 kHz at a sampling frequency of 20 kHz to faithfully record the fast transient signals. The cardiomyocytes form a confluent layer on the chip and show spontaneous activity, a regular beating driven by pacemaker heart cells in the culture. Usually, the signals of several electrodes show the same beating rhythm so that so-called “field potentials” from patches of

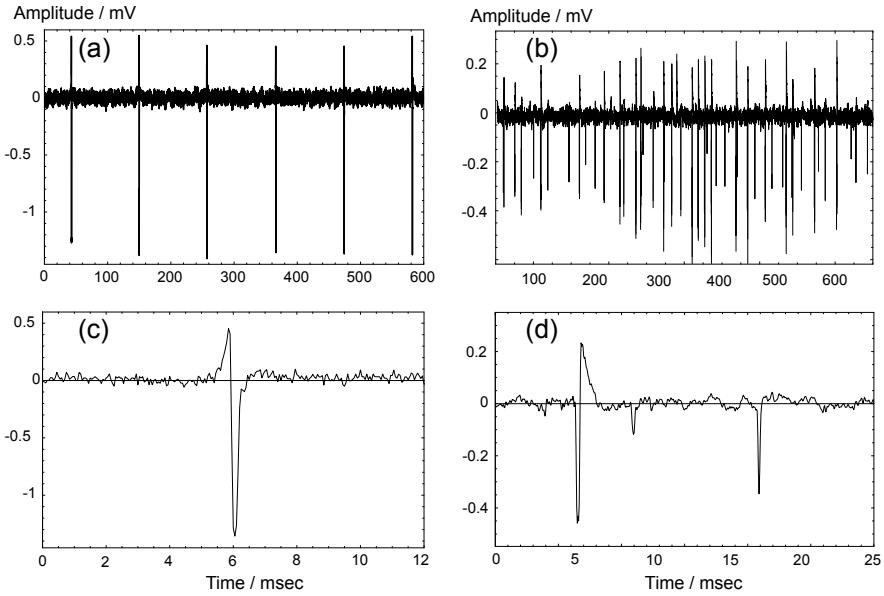


Figure 8.18 (a) Field potential recording of a confluent layer of regularly beating cardiomyocytes from neonatal rat (embryonic day 5) after phenylephrine dosage (100 µM) after 5 days *in vitro*. (b) Close-up of one individual spike. (c) Activity recordings from dissociated chicken cortical tissue (embryonic day 10) after 27 days *in vitro*. (d) Trace at higher temporal resolution showing several spikes [57].

heart cells are recorded. Recordings of the spontaneously beating cells after 5 days *in vitro* are shown in Fig. 8.18. In this example, the cells have been activated by dosing phenylephrine and beat at a rate of about 10 Hz. The signals were recorded on 20-µm-diameter electrodes. The cardiomyocyte signals are rather large featuring amplitudes between 1 and 1.5 mV.

In a next experiment, neural networks originating from dissociated cortical tissue of fertilized chicken eggs (*gallus domesticus*) at embryonic day 9 [97] were grown on the CMOS chips. These neurons survived more than 9 months on the CMOS chip, which suffices to possibly perform some kind of learning or information processing. Exemplary signals from spontaneously firing cells after 27 days *in vitro* are shown in Fig. 8.18. The noise level in these recordings was 17 μV_{RMS}. The signals recorded from 20-µm-diameter electrodes showed rather large signal amplitudes between 500 µV and 700 µV.

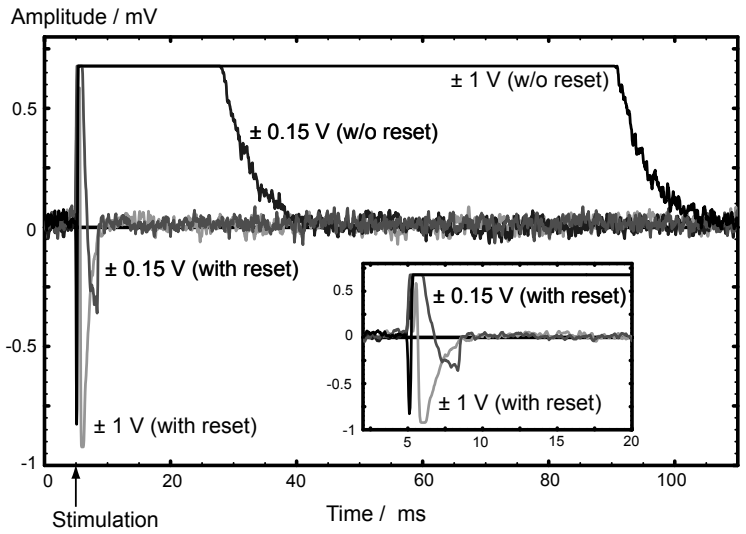


Figure 8.19 Efficiency of the on-chip reset function for stimulation. Bipolar pulses (220 µsec overall duration) of different amplitudes (±0.15 V and ±1 V) in saline solution have been applied, and the reset has been either operational or not. Without reset, it takes 20 to 100 msec for the recording circuitry to return to the measurement range. The inset shows a close-up of the initial 20 msec at higher temporal resolution.

8.4.2 Stimulation Artifact Suppression

Figure 8.19 shows the efficiency of the on-chip reset function for stimulation artifact suppression. Bipolar pulses (220 µsec overall duration) of different amplitudes (±0.15 V and ±1 V) in saline solution have been applied with the reset being either operational or not. The recordings from the stimulation electrode are shown. For the traces with reset, the readout filter was reset 50 µsec before the stimulation pulse was applied and kept at least until 50 µsec after the stimulation pulse was finished. Without reset, it takes 20 to 100 msec (depending on the stimulation amplitude) for the recording circuitry to return to the measurement range. With reset, it takes less than 5 msec for the recording circuitry to return to the measurement range. It is worth mentioning that a stimulation sequence always ends with a stimulation of a value close to the equilibrium potential of the electrode so that the electrode quickly returns to its equilibrium potential after stimulation.

8.4.3 Stimulation of Neural and Cardiac Cell Cultures

Stimulation experiments have been carried out using again cardiomyocytes first (Fig. 8.20). Half of the array (64 electrodes) was used for this purpose. The spontaneous regular beating of the cells every 1.2 seconds is shown in Fig. 8.20(a), which is an electrode event plot versus time, i.e., each electrical signal on an electrode is represented as a dot. The stimulation included 20 bipolar pulses of ±1 V at 1.5 kHz as can be seen in Fig. 8.20(b), which shows the stimulation signal on an electrode at 750 µm distance from the stimulated electrodes (the reset function was not activated at this electrode

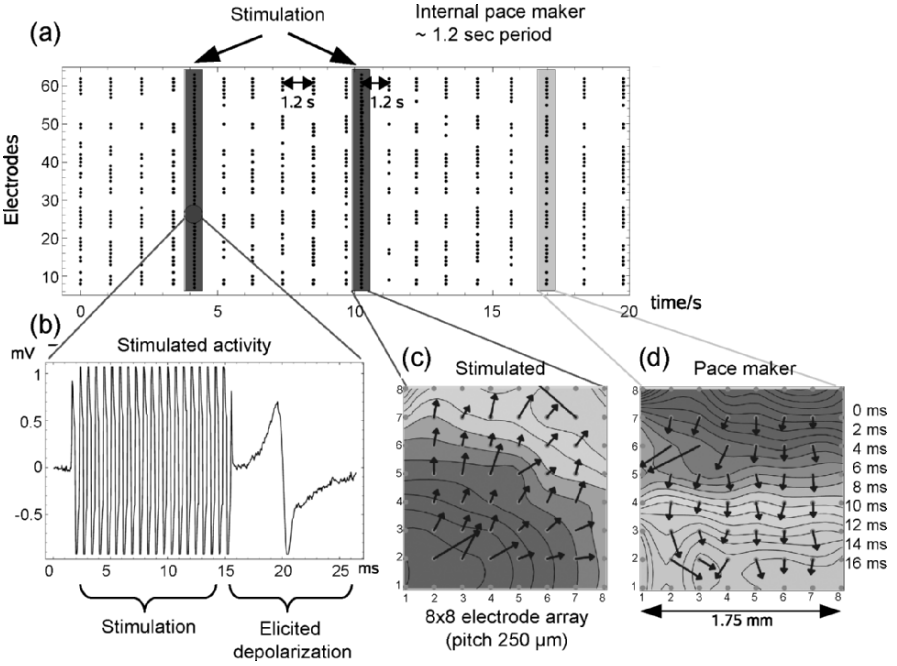


Figure 8.20 Stimulation and recording in a neonatal rat cardiomyocyte culture (embryonic day 5) after 3 days *in vitro*. (a) Electrode event plot versus time (each electrode signal represented as a dot) showing spontaneous regular beating of the cells every 1.2 seconds and the stimulations. (b) Stimulation (20 bipolar pulses of ±1 V at 1.5 kHz) and elicited cell activity on an electrode at 750 µm distance from the stimulated electrodes (reset was not activated to show the stimulation). (c) Isochronous map of a stimulated electrical wave propagating from the lower left corner of the electrode array (stimulation on electrodes 1-8); the times are encoded in different gray values. (d) Isochronous map of a spontaneous electrical wave propagating from a pacemaker region at the upper right corner. The propagation of the stimulated wave (12-14 msec across the array) is somewhat faster than that of the spontaneous wave (16-18 msec) [57].

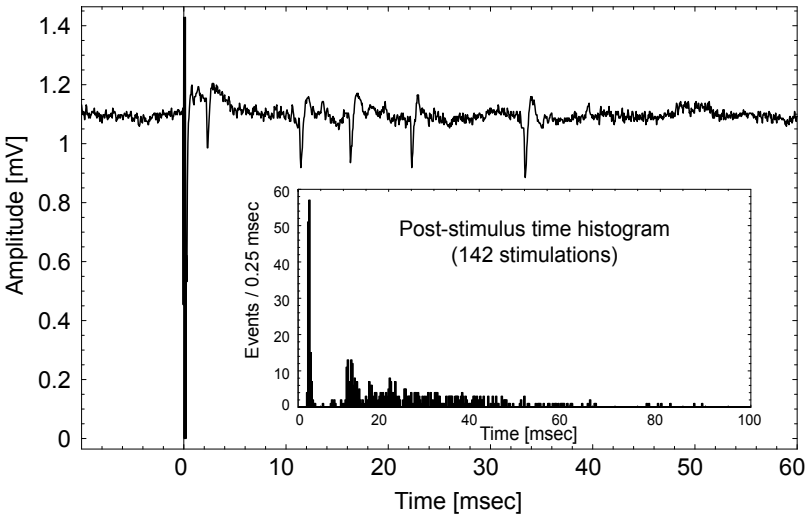


Figure 8.21 Successful excitation of spikes in cortical neurons from rat brain at 250 μm distance from the stimulation site. A single bipolar stimulation pulse of ±800 mV and 50 μsec duration was used to stimulate the cells in this example. The inset shows the post-stimulus time histogram of this channel based on 142 stimulation pulses [57].

to show the stimulation), and the evoked electrical activity after the stimulation. The cell culture reacts upon stimulation and adopts the stimulation signal as a new starting point for its spontaneous activity, as can be seen even more clearly for the second stimulation. Figures 8.20(c) and (d) show the propagation of spontaneous and triggered electrical waves over the electrode area. The pacemaker cells that produce spontaneous activity are in the upper right corner, and the wave propagates from there within about 16 msec over the electrode array [Fig. 8.20(d)]. For stimulation, electrodes in the lower left corner have been selected, which caused a signal wave in the opposite direction.

Finally, stimulation experiments using cortical neurons from the rat brain have been carried out. Figure 8.21 shows an example of a successful excitation of spikes recorded at 250 μm distance from the stimulation site. A single bipolar stimulation pulse of ±800 mV and 50 μsec duration was used to stimulate the cells in this example. The inset shows the post-stimulus time histogram of this channel that includes the results of 142 subsequent stimulation pulses. The first event or neuronal signal generally occurs between two and three milliseconds after the stimulation pulse with a very high probability of 96%.

8.5 CONCLUSIONS AND OUTLOOK

The use of CMOS technology offers decisive advantages for recording electrical signals from cells *in vitro* or *in vivo* with regard to signal quality and the number and density of recording sites or electrodes that can be achieved. Integrated multiplexers and on-chip electronics enable high-quality recording or stimulation through a large number of transducers and allow for attaining cellular or even subcellular resolution. The presented single-chip systems fabricated in industrial CMOS technology combined with post-CMOS processing are very compact.

An important research and development thrust at the moment is to further increase the number of stimulation and recording electrodes. In pursuing this goal it has to be taken into account that a large number of electrodes also generates a huge volume of recording data so that data sorting, standardized or automated data interpretation, and data compression will become pivotal issues. Methods like online spike detection using, e.g. an FPGA, will be very important to be able to handle the data volume of long-term recordings from electrode arrays with several hundreds or thousands of electrodes. The high-density electrode arrays still suffer from rather high noise levels on the order of hundreds of μV_{RMS} [11]. A new approach to achieve high-spatiotemporal-resolution and low-noise recordings is based on the idea to have a high-density array of electrodes, where only a selected subset of the electrodes is connected to recording and stimulation channels using a smart switching matrix [98]. Besides serving all applications of commercial-type multielectrode arrays this approach is particularly useful to study acute or organotypic brain or organ slices. Another important trend includes the integration of other sensor types into the system, such as chemical sensors (pH and oxygen content) to observe cell metabolism characteristics, or impedance sensors to study the cell attachment. Efforts in this direction have been described in Subsection 8.3.2.

CMOS microsystem platforms for bidirectional communication between cultured neurons and electronic systems will have, most probably, a major scientific and technological impact, particularly in neuroscience research and artificial intelligence. They will enable fundamental neurophysiological research at the cellular and circuit levels for investigating dynamic alterations in neuronal connectivity patterns and provide a new approach to investigate the behavior and information processing capabilities (including learning) of large populations of neurons *in vitro* or in slices. They may even open routes to the design of dedicated hybrid “intelligent” systems

combining evolutionary optimized highly parallel information processing capabilities of, e.g. brain cells with the more serial ones of microelectronics.

The CMOS-based systems will also find numerous applications in medical diagnostics and pharmacology. A large market can be expected for MEA-based biosensors targeted at the detection of chemical and biological toxins or the screening of pharmacologically active compounds since, e.g. *in vitro* networks show remarkable sensitivities to the dosing of chemicals at minute concentrations. Therefore, CMOS-based systems can be used for assessing dose-related effects of all kinds of chemicals or pharmacologically active substances [6-8, 99]. This holds particularly true for long-term measurements and multiple dosings or dosing sequences, which cannot be studied using the patch clamp method owing to the inherent short cell viability time (usually hours) as a consequence of the invasive nature of the method. Moreover, CMOS-based microsystems may be used as point-of-care devices to, e.g. test the effects of medical treatment on patient cells (accessible through biopsy) before administration.

Finally, the systems will have a large impact in the field of neural prosthetics: Microelectronics-based devices may become a viable approach to recover skeletal muscle function [100, 101], to restore visual perception [102] or functions of the auditory system [103, 104].

APPENDIX

Preparation of Cardiomyocytes: The NRCs (neonatal rat cardiomyocytes) were dispersed from the ventricles of 1-3-day-old Sprague-Dawley rats by digestion with collagenase II (Worthington Biochemical Corp., U.S.A.) and pancreatin (Sigma-Aldrich, U.S.A.). To obtain cultures comprising more than 95% cardiomyocytes, the cell suspensions were separated on a discontinuous Percoll gradient [105]. Cells were seeded in plating medium consisting of 68% DMEM (Amimed, Switzerland), 17% Medium M199 (Amimed), 10% horse serum (Life Technologies, USA), 5% fetal calf serum (Life Technologies), 4 mM glutamine (Amimed), and 1% penicillin-streptomycin (Amimed). After 24 h the plating medium was exchanged for maintenance medium, consisting of 78% DMEM (Amimed), 20% Medium M1999 (Amimed), 1% horse serum (Life Technologies), 1% penicillin-streptomycin (Amimed), and 4 mM glutamine, (Amimed).

Preparation of Neurons: Hippocampal tissue was extracted from newborn Sprague-Dawley rats or embryonic chicken (*Gallus domesticus*) at embryonic day 10-12 (according to [97]) and dissociated by trypsinization (0.25% Trypsin/EDTA) and gentle trituration. Cells were plated on laminin (10 g/ml) and Poly-L-Lysine (10 g/ml) coated CMOS chips at a density of 15000 cells/mm^2 and held in

Neurobasal medium (Gibco) containing B27 supplement (2%, Gibco), fatty acid supplement (0.1%, Sigma), lipid mixture (0.1%, Sigma), Alanyl-Glutamine (2 mM, Gibco) and sodium pyruvate (1 mM, Sigma). Prior to coating, the chips were prepared as follows: To optimize cellular adhesion, the chips were treated with oxygen plasma for 1 min at 40 W. The chips were then sterilized by immersing them in 70% ethanol for 1 min. After the ethanol was completely evaporated, the electrode area was rinsed several times with ultrapure water.

ACKNOWLEDGMENT

The authors are greatly indebted to all people involved in the biochip development: Wendy Franks, Urs Frey, Frauke Greve, Sadik Hafizovic, Kay-Uwe Kirstein and Branka Roscic from the Physical Electronics Laboratory at ETH Zurich; and Axel Blau, Simone Riedel, Tanja Ugniwenko, and Christiane Ziegler from the University of Kaiserslautern. Moreover, the authors are grateful to Prof. Henry Baltes, Physical Electronics Laboratory, ETH Zurich for sharing laboratory resources and for his ongoing stimulating interest in their work. The authors also want to thank Evelyne Perriard at ETH Zurich for help with the cardiomyocyte preparations. Funding has been generously provided by the Information Societies Technology (IST) Future and Emerging Technologies program of the European Union, and the Swiss Bundesamt für Bildung und Wissenschaft (BBW) under contract number IST-2000-26463.

REFERENCES

[1] Y. Jimbo, T. Tateno, and H. P. C. Robinson, "Simultaneous induction of pathway-specific potentiation and depression in networks of cortical neurons," Biophysical Journal, vol. 76, pp. 670-678, 1999.

[2] S. Marom and G. Shahaf, "Development, learning and memory in large random networks of cortical neurons: lessons beyond anatomy," Quarterly Reviews of Biophysics, vol. 35, pp. 63-87, 2002.

[3] G. Shahaf and S. Marom, "Learning in networks of cortical neurons," Journal of Neuroscience, vol. 21, pp. 8782-8788, 2001.

[4] M. E. Ruaro, P. Bonifazi, V. Torre, "Towards the neurocomputer: image processing and pattern recognition with neuronal cultures," IEEE Transactions on Biomedical Engineering, 2005.

[5] D. A. Wagenaar, Z. Nadasdy, and S. M. Potter, "Persistent dynamic attractors in activity patterns of cultured neuronal networks," Phys Rev E Stat Nonlin Soft Matter Phys, vol. 73, p. 051907, 2006.

[6] G. Gross, B. K. Rhoades, H. M. E. Azzazy, and W. Ming Chi, "The use of neuronal networks on multielectrode arrays as biosensors," Biosensors & Bioelectronics, vol. 10, pp. 553-567, 1995.

[7] J. J. Pancrazio, P. P. Bey, Jr., D. S. Cuttino, J. K. Kusel, D. A. Borkholder, K. M. Shaffer, G. T. A. Kovacs, and D. A. Stenger, "Portable cell-based biosensor system for toxin detection," Sensors and Actuators B (Chemical), vol. B53, pp. 179-185, 1998.

[8] A. Stett, U. Egert, E. Guenther, F. Hofmann, T. Meyer, W. Nisch, and H. Haemmerle, "Biological application of microelectrode arrays in drug discovery and basic research," Anal Bioanal Chem, vol. 377, pp. 486-495, 2003.

[9] T. Meyer, C. Leisgen, B. Gonser, and E. Gunther, "QT-screen: high-throughput cardiac safety pharmacology by extracellular electrophysiology on primary cardiac myocytes," Assay Drug Dev Technol, vol. 2, pp. 507-514, 2004.

[10] U. Egert, D. Heck, and A. Aertsen, "Two-dimensional monitoring of spiking networks in acute brain slices," Exp Brain Res, vol. 142, pp. 268-274, 2002.

[11] M. Hutzler, A. Lambacher, B. Eversmann, M. Jenkner, R. Thewes, and P. Fromherz, "High-resolution multi-transistor array recording of electrical field potentials in cultured brain slices," J Neurophysiol, 2006.

[12] E. Neher and B. Sakmann, "Single-channel currents recorded from membrane of denervated frog muscle fibres," Nature, vol. 260, pp. 799-802, 1976.

[13] K. Cole, "Dynamic electrical characteristics of the squid axon membrane," Arch. Sci. physiol., vol. 3, pp. 253-258, 1949.

[14] P. Fromherz, "Electrical interfacing of nerve cells and semiconductor chips," Chemphyschem, vol. 3, pp. 276-284, 2002.

[15] G. T. A. Kovacs, "Electronic sensors with living cellular components," Proceedings of IEEE, vol. 91, pp. 915-929, 2003.

[16] K. D. Wise, D. J. Anderson, J. F. Hetke, D. R. Kipke, and K. Najafi, "Wireless implantable microsystems: high-density electronic interfaces to the nervous system," Proceedings of the IEEE, vol. 92, pp. 76-97, 2004.

[17] R. H. Olsson, 3rd, D. L. Buhl, A. M. Sirota, G. Buzsaki, and K. D. Wise, "Band-tunable and multiplexed integrated circuits for simultaneous recording and stimulation with microelectrode arrays," IEEE Trans Biomed Eng, vol. 52, pp. 1303-11, 2005.

[18] A. G. Kleber and Y. Rudy, "Basic mechanisms of cardiac impulse propagation and associated arrhythmias," Physiol. Rev., vol. 84, pp. 431-488, 2003.

[19] W. L. C. Rutten, "Selective electrical interfaces with the nervous system," Annual Review of Biomedical Engineering, vol. 4, pp. 407-452, 2002.

[20] Y. Jimbo and H. P. Robinson, "Propagation of spontaneous synchronized activity in cortical slice cultures recorded by planar electrode arrays," Bioelectrochemistry, vol. 51, pp. 107-15, 2000.

[21] J. van Pelt, P. S. Wolters, M. A. Corner, W. L. Rutten, and G. J. Ramakers, "Long-term characterization of firing dynamics of spontaneous bursts in

cultured neural networks," IEEE Trans Biomed Eng, vol. 51, pp. 2051-2062, 2004.

[22] S. I. Morefield, E. W. Keefer, K. D. Chapman, and G. W. Gross, "Drug evaluations using neuronal networks cultured on microelectrode arrays," Biosens Bioelectron, vol. 15, pp. 383-396, 2000.

[23] J. J. Pancrazio, S. A. Gray, Y. S. Shubin, N. Kulagina, D. S. Cuttino, K. M. Shaffer, K. Eisemann, A. Curran, B. Zim, G. W. Gross, and T. J. O'Shaughnessy, "A portable microelectrode array recording system incorporating cultured neuronal networks for neurotoxin detection," Biosensors & Bioelectronics, vol. 18, pp. 1339-1347, 2003.

[24] B. D. DeBusschere and G. T. A. Kovacs, "Portable cell-based biosensor system using integrated CMOS cell-cartridges," Biosensors-&-Bioelectronics, vol. 16, pp. 543-556, 2001.

[25] M. Taketani, Advances in network electrophysiology using multi-electrode arrays. New York: Springer, 2006.

[26] B. J. Baker, E. K. Kosmidis, D. Vucinic, C. X. Falk, L. B. Cohen, M. Djurisic, and D. Zecevic, "Imaging brain activity with voltage- and calcium-sensitive dyes," Cell Mol Neurobiol, vol. 25, pp. 245-282, 2005.

[27] Z. A. Peterlin, J. Kozloski, B. Q. Mao, A. Tsiola, and R. Yuste, "Optical probing of neuronal circuits with calcium indicators," Proc Natl Acad Sci U S A, vol. 97, pp. 3619-3624, 2000.

[28] A. L. Obaid, L. M. Loew, J. P. Wuskell, and B. M. Salzberg, "Novel naphthylstyryl-pyridium potentiometric dyes offer advantages for neural network analysis," J Neurosci Methods, vol. 134, pp. 179-190, 2004.

[29] M. Voelker and P. Fromherz, "Signal transmission from individual mammalian nerve cell to field-effect transistor," SMALL, vol. 1, pp. 206-210, 2005.

[30] M. Jenkner, B. Muller, and P. Fromherz, "Interfacing a silicon chip to pairs of snail neurons connected by electrical synapses," Biological Cybernetics, vol. 84, pp. 239-249, 2001.

[31] M. O. Heuschkel, M. Fejtl, M. Raggenbass, D. Bertrand, and P. Renaud, "A three-dimensional multi-electrode array for multi-site stimulation and recording in acute brain slices," Journal of Neuroscience Methods, vol. 114, pp. 135-148, 2002.

[32] Y. Jimbo, N. Kasai, K. Torimitsu, T. Tateno, and H. P. Robinson, "A system for MEA-based multisite stimulation," IEEE Trans Biomed Eng, vol. 50, pp. 241-248, 2003.

[33] M. P. Maher, J. Pine, J. Wright, and T. Yu Chong, "The neurochip: A new multielectrode device for stimulating and recording from cultured neurons," Journal of Neuroscience Methods, vol. 87, pp. 45-56, 1999.

[34] S. Martinoia, P. Massobrio, M. Bove, and G. Massobrio, "Cultured neurons coupled to microelectrode arrays: circuit models, simulations and experimental data," IEEE Trans Biomed Eng, vol. 51, pp. 859-864, 2004.

[35] B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert,

M. Steinhauser, G. Eckstein, D. Schmitt-Landsiedel, and R. Thewes, "A 128 x 128 CMOS biosensor array for extracellular recording of neural activity," IEEE Journal of Solid-State Circuits, vol. 38, pp. 2306-2317, 2003.

[36] H. F. Lodish, Molecular cell biology, Third ed. New York: Scientific American Books, 1995.

[37] D. Johnston, S. Wu, Foundations of Cellular Neurophysiology. London: The MIT Press, 1995.

[38] E. R. Kandel, J. H. Schwartz, T. M. Jessell, Principles of Neural Science, Fourth ed. London: McGraw-Hill, 2000.

[39] A. L. Hodgkin and A. F. Huxley, "A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve," Journal of Physiology, vol. 117, pp. 500-544, 1952.

[40] A. G. Kleber and Y. Rudy, "Basic mechanisms of cardiac impulse propagation and associated arrhythmias," Physiol Rev, vol. 84, pp. 431-488, 2004.

[41] R. Caton, "The electric currents of the brain," Br. Med. J., vol. 2, p. 278, 1875.

[42] K. D. Wise and J. B. Angeli, "A Microprobe with Integrated Amplifiers for Neurophysiology " ISSCC, Proc. of pp. 100-101, 1971.

[43] C. A. Thomas, P. A. Springer, G. E. Loeb, Y. Berwald-Netter, L. M. Okun, "A miniature microelectrode array to monitor the bioelectric activity of cultured cells," Experimental Cell Research, vol. 74, pp. 61-66, 1972.

[44] M. Jenkner, M. Tartagni, A. Hierlemann, and R. Thewes, "Cell-Based CMOS Sensor and Actuator Arrays," Solid State Circuits, vol. 39, pp. 2431-2437, 2004.

[45] G. W. Gross, W. Y. Wen, and J. W. Lin, "Transparent Indium Tin Oxide Electrode Patterns for Extracellular, Multisite Recording in Neuronal Cultures," Journal of Neuroscience Methods, vol. 15, pp. 243-252, 1985.

[46] V. Bucher, M. Graf, M. Steizle, and W. Nisch, "Low-impedance thin-film polycrystalline silicon microelectrodes for extracellular stimulation and recording," Biosensors-&-Bioelectronics, vol. 14, pp. 639-649, 1999.

[47] Y. Jimbo, A. Kawana, P. Parodi, and V. Torre, "The dynamics of a neuronal culture of dissociated cortical neurons of neonatal rats," Biological Cybernetics, vol. 83, pp. 1-20, 2000.

[48] D. A. Wagenaar, R. Madhavan, J. Pine, and S. M. Potter, "Controlling bursting in cortical cultures with closed-loop multi-electrode stimulation," J Neurosci, vol. 25, pp. 680-688, 2005.

[49] M. Chiappalone, A. Vato, M. B. Tedesco, M. Marcoli, F. Davide, and S. Martinoia, "Networks of neurons coupled to microelectrode arrays: a neuronal sensory system for pharmacological applications," Biosensors & Bioelectronics, vol. 18, pp. 627-634, 2003.

[50] E. W. Keefer, A. Gramowski, D. A. Stenger, J. J. Pancrazio, and G. W. Gross, "Characterization of acute neurotoxic effects of trimethylolpropane phosphate via neuronal network biosensors," Biosensors & Bioelectronics, vol. 16, pp. 513-525, 2001.

[51] P. Bonifazi and P. Fromherz, "Silicon chip for electronic communication between nerve cells by non-invasive interfacing and analog-digital processing," Advanced Materials, vol. 14, pp. 1190-1193, 2002.
[52] D. A. Wagenaar and S. M. Potter, "A versatile all-channel stimulator for electrode arrays, with real-time control," J Neural Eng, vol. 1, pp. 39-45, 2004.
[53] G. W. Gross, B. K. Rhoades, D. L. Reust, and F. U. Schwalm, "Stimulation of monolayer networks in culture through thin-film indium-tin oxide recording electrodes," Journal-of-Neuroscience-Methods, vol. 50, pp. 131-143, 1993.
[54] J. R. Buitenweg, W. L. C. Rutten, and E. Marani, "Extracellular stimulation window explained by a geometry-based model of the Neuron-electrode contact," IEEE-Transactions-on-Biomedical-Engineering, vol. 49, pp. 1591-1599, 2002.
[55] C. C. McIntyre and W. M. Grill, "Extracellular stimulation of central neurons: influence of stimulus waveform and frequency on neuronal output," J Neurophysiol, vol. 88, pp. 1592-1604, 2002.
[56] Y. Jimbo and A. Kawana, "Electrical-Stimulation and Recording from Cultured Neurons Using a Planar Electrode Array," Bioelectrochemistry and Bioenergetics, vol. 29, pp. 193-204, 1992.
[57] F. Heer, S. Hafizovic, T. Ugniwenko, W. Franks, A. Blau, C. Ziegler, J. C. Perriard, and A. Hierlemann, "Single-chip microelectronic system to interface with living cells," Biosens Bioelectron, in press, 2006.
[58] J. W. Gnadt, S. D. Echols, A. Yildirim, H. Zhang, and K. Paul, "Spectral cancellation of microstimulation artifact for simultaneous neural recording in situ," IEEE Trans Biomed Eng, vol. 50, pp. 1129-1135, 2003.
[59] D. A. Wagenaar and S. M. Potter, "Real-time multi-channel stimulus artifact suppression by local curve fitting," J Neurosci Methods, vol. 120, pp. 113-120, 2002.
[60] F. Heer, S. Hafizovic, W. Franks, A. Blau, C. Ziegler, and A. Hierlemann, "CMOS Microelectrode Array for Bidirectional Interaction with Neuronal Networks," IEEE Journal of Solid-State Circuits, vol. 41, 2006.
[61] P. Fromherz, "Extracellular recording with transistors and the distribution of ionic conductances in a cell membrane," European Biophysics Journal with Biophysics Letters, vol. 28, pp. 254-258, 1999.
[62] D. Braun and M. P. J. Fromherz, "Fluorescence Interferometry of Neuronal Cell Adhesion on Microstructured Silicon," Phys. Rev. Lett., vol. 81, pp. 5241-5244, 1998.
[63] J. R. Buitenweg, W. L. Rutten, and E. Marani, "Geometry-based finite-element modeling of the electrical contact between a cultured neuron and a microelectrode," IEEE Trans Biomed Eng, vol. 50, pp. 501-509, 2003.
[64] R. Weis and P. Fromherz, "Frequency dependent signal transfer in neuron transistors," Physical Review E (Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics), vol. 55, pp. 877-889, 1997.
[65] S. Vassanelli and P. Fromherz, "Transistor records of rat hippocampal neurons," European Journal of Neuroscience, vol. 10, p. 2838, 1998.

[66] R. Schatzthauer and P. Fromherz, "Neuron-silicon junction with voltage-gated ionic currents," Eur J Neurosci, vol. 10, pp. 1956-1962, 1998.
[67] W. Franks, Schenker, I., Schmutz, P., Hierlemann, A., "Impedance Characterization and Modeling of Electrodes for Biomedical Applications," Transactions on Biomedical Engineering, 2005.
[68] R. C. Gesteland, B. Howland, J. Y. Lettvin, and W. H. Pitts, "Comments on Microelectrodes," Proceedings of the Institute of Radio Engineers, vol. 47, pp. 1856-1862, 1959.
[69] J. J. Pancrazio, J. P. Whelan, D. A. Borkholder, W. Ma, and D. A. Stenger, "Development and Application of Cell-Based Biosensors," Annals of Biomedical Engineering, vol. 27, pp. 687-711, 1999.
[70] L. Berdondini, P. D. van der Wal, O. Guenat, N. F. de Rooij, M. Koudelka-Hep, P. Seitz, R. Kaufmann, P. Metzler, N. Blanc, and S. Rohr, "High-density electrode array for imaging *in vitro* electrophysiological activity," Biosens Bioelectron, vol. 21, pp. 167-174, 2005.
[71] F. Heer, W. Franks, A. Blau, S. Taschini, C. Ziegler, A. Hierlemann, and H. Baltes, "CMOS microelectrode array for the monitoring of electrogenic cells," Biosensors & Bioelectronics, vol. 20, pp. 358-366, 2004.
[72] J. D. Weiland, D. J. Anderson, and M. S. Humayun, "*In vitro* electrical properties for iridium oxide versus titanium nitride stimulating electrodes," IEEE Trans Biomed Eng, vol. 49, pp. 1574-1579, 2002.
[73] A. J. Bard and L. R. Faulkner, Electrochemical methods fundamentals and applications, 2nd ed. New York: Wiley, 2001.
[74] R. W. De Boer and A. Van Oosterom, "Electrical properties of platinum electrodes: impedance measurements and time-domain analysis," Medical & Biological Engineering & Computing, vol. 16, pp. 1-10, 1978.
[75] B. Onaral and H. P. Schwan, "Linear and nonlinear properties of platinum electrode polarisation. III. Equivalence of frequency- and time-domain behaviour," Medical & Biological Engineering & Computing, vol. 23, pp. 28-32, 1985.
[76] E. Barsoukov, Impedance spectroscopy theory, experiment, and applications, 2nd ed. Hoboken, N.J.: Wiley-Interscience, 2005.
[77] M. Merz and P. Fromherz, "Silicon chip interfaced with a geometrically defined net of snail neurons," Advanced Functional Materials, vol. 15, pp. 739-744, 2005.
[78] C. Sprossler, M. Denyer, S. Britland, W. Knoll, and A. Offenhausser, "Electrical recordings from rat cardiac muscle cells using field-effect transistors," Physical Review E, vol. 60, pp. 2171-2176, 1999.
[79] A. Offenhausser and W. Knoll, "Cell-transistor hybrid systems and their potential applications," Trends in Biotechnology, vol. 19, pp. 62-66, 2001.
[80] W. Baumann, E. Schreiber, G. Krause, and S. Stüwe, "Multiparametric neurosensor microchip," presented at Proceedings Eurosensors XVI, Prag, 2002.

[81] U. Egert, B. Schlosshauer, S. Fennrich, W. Nisch, M. Fejtl, T. Knott, T. Muller, and H. Hammerle, "A novel organotypic long-term culture of the rat hippocampus on substrate-integrated multielectrode arrays," Brain Research Protocols, vol. 2, pp. 229-242, 1998.

[82] R. Ehret, W. Baumann, M. Brischwein, M. Lehmann, T. Henning, I. Freund, S. Drechsler, U. Friedrich, M. L. Hubert, E. Motrescu, A. Kob, H. Palzer, H. Grothe, and B. Wolf, "Multiparametric microsensor chips for screening applications," Fresenius J Anal Chem, vol. 369, pp. 30-35, 2001.

[83] M. Lehmann, W. Baumann, M. Brischwein, R. Ehret, M. Kraus, A. Schwinde, M. Bitzenhofer, I. Freund, and B. Wolf, "Non-invasive measurement of cell membrane associated proton gradients by ion-sensitive field effect transistor arrays for microphysiological and bioelectronical applications," Biosensors & Bioelectronics, vol. 15, pp. 117-124, 2000.

[84] R. Ehret, W. Baumann, M. Brischwein, A. Schwinde, and B. Wolf, "On-line control of cellular adhesion with impedance measurements using interdigitated electrode structures," Med Biol Eng Comput, vol. 36, pp. 365-370, 1998.

[85] K. H. Gilchrist, V. N. Barker, L. E. Fletcher, B. D. DeBusschere, P. Ghanouni, L. Giovangrandi, and G. T. A. Kovacs, "General purpose, field-portable cell-based biosensor platform," Biosensors & Bioelectronics, vol. 16, pp. 557-564, 2001.

[86] R. H. Olsson, 3rd, M. N. Gulari, and K. D. Wise, "Silicon neural recording arrays with on-chip electronics for *in vivo* data acquisition," Proc. IEEE-EMBS Int. Conf. Microtechnology Medicine and Biology, pp. 237-240, 2002.

[87] T. Akin, K. Najafi, and R. M. Bradley, "A Wireless Implantable Multichannel Digital Neural Recording System for a Micromachined Sieve Electrode," Solid State Circuits, vol. 33, pp. 109-118, 1998.

[88] R. R. Harrison and C. Charles, "A Low-Power Low-Noise {CMOS} Amplifier for Neural Recording Applications," Solid State Circuits, vol. 38, pp. 958-965, 2003.

[89] Y. Papananos, T. Georgantas, and Y. Tsividis, "Design considerations and implementation of very low frequencycontinuous-time CMOS monolithic filters," IEE Proceedings on Circuits, Devices and Systems, vol. 144, pp. 68-74, 1997.

[90] T. Delbruck and C. A. Mead, "Adaptive photoreceptor with wide dynamic range," Proceedings of ISCAS, vol. 4, pp. 339-342, 1994.

[91] J. J. F. Rijns, "CMOS Low-Distortion High-Frequency Variable-Gain Amplifier " Solid State Circuits, vol. 31, pp. 1029-1034, 1996.

[92] K. Martin, L. Ozcolak, and Y. S. Lee, "Differential Switched-Capacitor Amplifier " Solid State Circuits, vol. 22, pp. 104-106, 1987.

[93] J. B. Bates and Y. T. Chu, "Electrode-electrolyte interface impedance: experiments and model," Annals of Biomedical Engineering, vol. 20, pp. 349-362, 1992.

[94] X. Huang, D. Nguyen, D. W. Greve, and M. M. Domach, "Simulation of Microelectrode Impedance Changes Due to Cell Growth "IEEE Sensors, vol. 4, pp. 576-583, 2004.
[95] S. Hafizovic, F. Heer, T. Ugniwenko, A. Blau, C. Ziegler, and A. Hierlemann, "System Integration of a CMOS-based Microelectrode Array for Interaction with Neuronal Cultures," J Neurosci Methods, vol. submitted, 2006.
[96] M. S. Lewicki, "A review of methods for spike sorting: the detection and classification of neural action potentials," Network-Computation in Neural Systems, vol. 9, pp. R53-R78, 1998.
[97] V. Hamburger and H. L. Hamilton, "A series of normal stages in the development of the chick embryo," J. Morph., vol. 88, pp. 49-92, 1951.
[98] U. Frey, F. Heer, R. Pedron, F. Greve, S. Hafizovic, K. U. Kirstein, and A. Hierlemann, "11'000 Electrode-, 126 Channel-CMOS Microelectrode Array for Electrogenic Cells," Proc IEEE MEMS, in press, 2007.
[99] FDA, "ICH S7B Guideline: The Nonclinical Evaluation of the Potential for Delayed Ventricular Repolarization (QT Interval Prolongation) by Human Pharmaceuticals.," RockvilleMD 20857-0001, USA: U. S. Food and Drug Administration, 2004.
[100] T. Stieglitz and M. Gross, "Flexible BIOMEMS with electrode arrangements on front and back side as key component in neural prostheses and biohybrid systems," Sensors and Actuators B (Chemical), vol. 83, pp. 8-14, 2002.
[101] M. A. Nicolelis, "Brain-machine interfaces to restore motor function and probe neural circuits," Nat Rev Neurosci, vol. 4, pp. 417-422, 2003.
[102] E. Zrenner, "Will retinal implants restore vision?," Science, vol. 295, pp. 1022-1025, 2002.
[103] G. E. Loeb, "Cochlear prosthetics," Annu Rev Neurosci, vol. 13, pp. 357-371, 1990.
[104] J. P. Rauschecker and R. V. Shannon, "Sending sound to the brain," Science, vol. 295, pp. 1025-1029, 2002.
[105] A. Sen, P. Dunnmon, S. A. Henderson, R. D. Gerard, and K. R. Chien, "Terminally differentiated neonatal rat myocardial cells proliferate and maintain specific differentiated functions following expression of SV40 large T antigen," J Biol Chem, vol. 263, pp. 19132-19136, 1988.

AUTHOR BIOGRAPHY

Flavio Heer received the diploma in physics in 2001 and the Ph.D. degree in 2005 from ETH Zurich, Zurich, Switzerland. He is currently working as a Postdoc at the Physical Electronics Laboratory at ETH Zurich. The focus of his research activities is on system design and interface circuitry for cell-based microsensors and biochemical sensors.

Andreas Hierlemann received his diploma in chemistry in 1992 and the Ph.D. degree in physical chemistry in 1996 from the University of Tübingen, Germany. After working as a Postdoc at Texas A&M University, College Station, TX (1997), and Sandia National Laboratories, Albuquerque, NM (1998), he moved to ETH Zurich in 1999 and is currently associate professor at the Physical Electronics laboratory at ETH Zurich in Switzerland. The focus of his research activities is on CMOS-based chemical and biomicrosensors and on interfacing CMOS electronics with electrogenic cells.

9 CMOS ICS FOR BRAIN IMPLANTABLE NEURAL RECORDING MICROSYSTEMS

William R. Patterson III[1], Yoon-kyu Song[1], Christopher W. Bull[1], Farah L. Laiwalla[1], Arto Nurmikko[1,2]* and John P. Donoghue[3]

[1] Division of Engineering, Brown University
[2] Department of Physics, Brown University
[3] Department of Neuroscience, Brown University
*Arto_Nurmikko@brown.edu

9.1 INTRODUCTION

While paying homage to the technological revolutions empowered by modern microelectronics and the continued march towards ever increasing sophistication of nanoscale high-density digital circuits, it may be useful to pause and draw a reference to a computer of very different variety, namely the human brain. A product of millions of years of evolution, this non-silicon, biochemical processor features about 10^{15} interconnects amongst its 3D network of 10^{11} neurons ('neuristors') while consuming about 10 watts of power. Approximately 10^{16} synaptic events per second occur within the sublime architecture of our 'cortical computer' whose detailed design and operational principles define the main quest of modern neuroscience.

Science fiction has frequently visited the question of direct communication with the brain, including the prospects of interfacing the brain with external machines and extrasensory communication such as telepathy. But there is a nonetheless a serious and compelling medical rationale to pursue direct retrieval of command and control signals from the brain. There are millions of individuals who suffer from serious neurological illnesses, ranging, e.g. from Parkinson's disease to complete paralysis. Spinal cord injuries and degenerative diseases such as ALS in particular involve very large

populations of people whose quality of life is much impaired, even if the brain is cognitively functional. Enabling such individuals to translate their command thoughts from the brain to direct activation of external assistive devices offers a potential intersection with modern technologies of significant societal and health care value. A Neuromotor Prosthesis (NMP) is a neurotechnology, which may restore lost movements in paralyzed humans; the crucial role of specialized CMOS IC development for an implantable cortical NMP is the subject of this chapter.

Invasive, permanently implanted cortical microelectrodes yield signals from regions of the brain that are directly related to intended movements. For example, chronically implanted, multielectrode arrays have allowed long term (>1 year) exploration of a significant amount of motor cortex terrain in monkeys. In conjunction with newly developed decoding techniques using probabilistic analysis, good correlation has been achieved for the arm movement of a monkey between the signals recorded directly from the brain ("thought-for-action") and the real physical action by the animal [1]. A recent culmination of the research in this area is the first human clinical pretrials, where intention-driven neuronal ensemble activity has been converted into a control signal that has enabled a tetraplegic patient (and others) to perform useful tasks [2]. Figure 9.1 shows the human tetraplegic patient in the clinical test setting, with the implanted multielectrode array equipped with a head mounted module which mainly houses preamplifiers and multiplexing circuitry, subsequently cabled to multiple electronic system modules. As

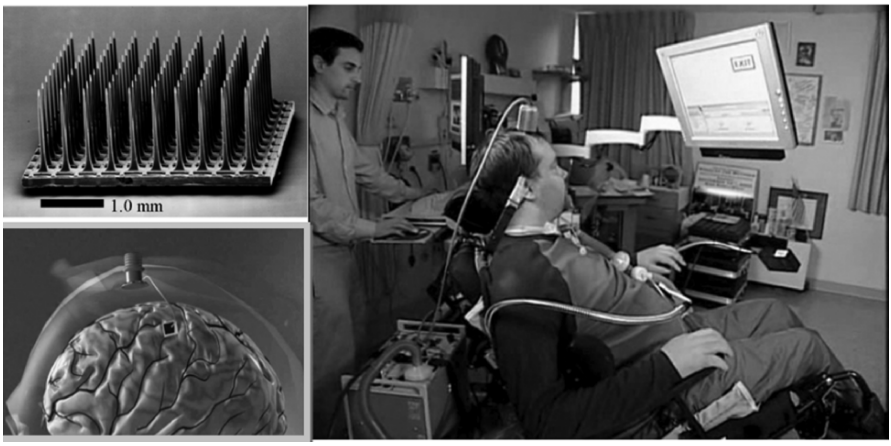


Figure 9.1 (upper left): a silicon-based cortical microelectrode array; (lower left): schematic of insert of the array and percutaneous connection to a skull mounted pedestal connector; (right) a patient in a human clinical trial (courtesy Cyberkinetics Inc., Foxboro, MA).

summarized in Ref. [2], this clinical trial, conducted over nearly one year has enabled the patient, whose spinal cord is severed at the neck, to control a cursor on a computer screen for communication activities such as reading email, typing messages, drawing elementary free form shapes, and operating an open-close prosthetic hand. Other assistive devices under such direct brain control, which are being tested and which are under further development include connecting the cortical recording system to a wheelchair and a robotic arm. It is the culmination of much research and development of brain recording devices which has not only led to the these striking results in the first human trials but is now motivating attempts to microminiaturize and enhance the performance of the neuroprobe systems by turning to advanced microelectronics.

The integration of a significant portion or all of an NMP system onto an implantable platform is highly desirable for future portable and wearable prosthetics, including on-board auxiliary telemetry to guide neural signals within the body or to transmit these to external remote processors and other assistive technologies. Implantable cortical NMP microsystems present a multi-faceted technical challenge, which includes the development of ultra-low power, integrable CMOS ICs to the neuroprobe recording platform, choices of (broadband) on-board telemetry, as well as the means to deliver power for the active components. Finally, encapsulation and biocompatibility of such an electrically active multi-element implant are challenges of fundamental importance for future human applications, for surgical and other clinical considerations.

There are several ongoing efforts to develop implantable cortical microsystems [3-5], including work presently underway in the authors' laboratories. The internal architecture of one version of our particular microsystem is shown schematically as a block diagram in Fig. 9.2. Within this design, we have fully integrated a CMOS chip composed of low-noise analog preamplifiers and multiplexers onto the so-called "Utah/Cyberkinetics" (CKI) silicon-based microelectrode cortical probe array [6]. The full microsystem design is based on specific device layout geometry and architecture, aimed firstly at non-human primates, and is composed of analog and digital microelectronic components as well as infrared digital optical telemetry. A version of the fully fabricated microsystem in shown in Fig. 9.3, configured for fiber optic extraction of digital neural data stream (supplied by a low current semiconductor laser) and for power delivery likewise by a second fiber link to a high efficiency photovoltaic microcrystal converter ("optical power supply", OPS). The microsystem is fabricated on a flexible liquid crystal polymer (LCP) substrate and has a soft and flexible encapsulation and its

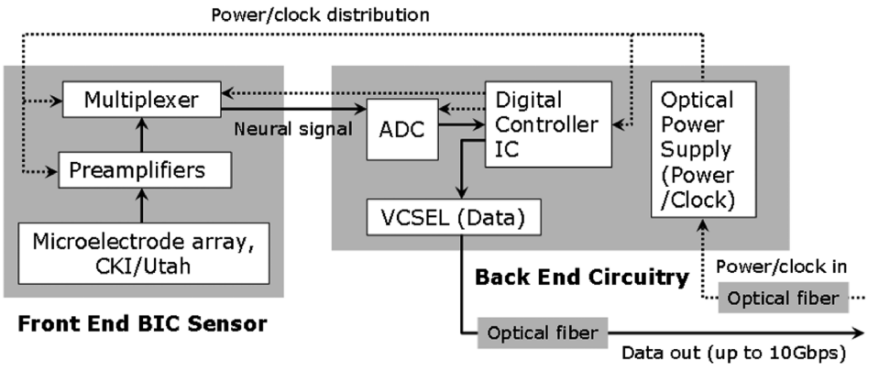


Figure 9.2 Block diagram for the implantable microsystem, detailing the key micro- and optoelectronic on-chip circuit elements.

design principles emphasize scalability for adapting to ever larger amounts of neural information transmitted from the cortex to assistive technologies, aimed ultimately for human neuroprosthetic applications. We have evaluated such an on-chip system both in bench-top experiments, by recording from thalamacortical rat brain slices [7], and from the somatosensory cortex of a rat [8], with first trials in monkeys near implementation.

From a purely engineering viewpoint, the concept of using the cortical microelectrode array as a platform onto which *all* the microelectronic and telemetric components are directly integrated as a single monolithic cortical implant module might at first appear to represent an attractive approach [5]. However, by weighing a number of neurophysiological, biomedical, telemetric, and surgical considerations, we have chosen to pursue the more spatially distributed microsystem architecture that is shown as a fabricated prototype in Fig. 9.3. The layout in the figure reflects our "dual-panel" microsystem design concept, composed of two intimately electrically interconnected islands that share a common a flexible liquid crystal polymer (LCP) substrate. The dual-panel system is composed of a "front end" which is directly implanted into the cortex. This front panel is flexibly connected to the back panel which threads through a sealable "burr" hole in the skull and resides between the skull and the skin. The front-end panel houses the cortical microelectrode array (in the present prototype version a 4×4 unit), directly flip-chip bonded onto an ultra-low power analog CMOS chip described in detail in this chapter. The custom designed and fabricated CMOS IC including low power preamplifier and multiplexing circuitry is the central focus of this discussion and other details and performance results can be found in [6] and [7]. A custom hybrid flip-chip bonding technique was developed at Brown to produce the functional, encapsulated microminiatur-

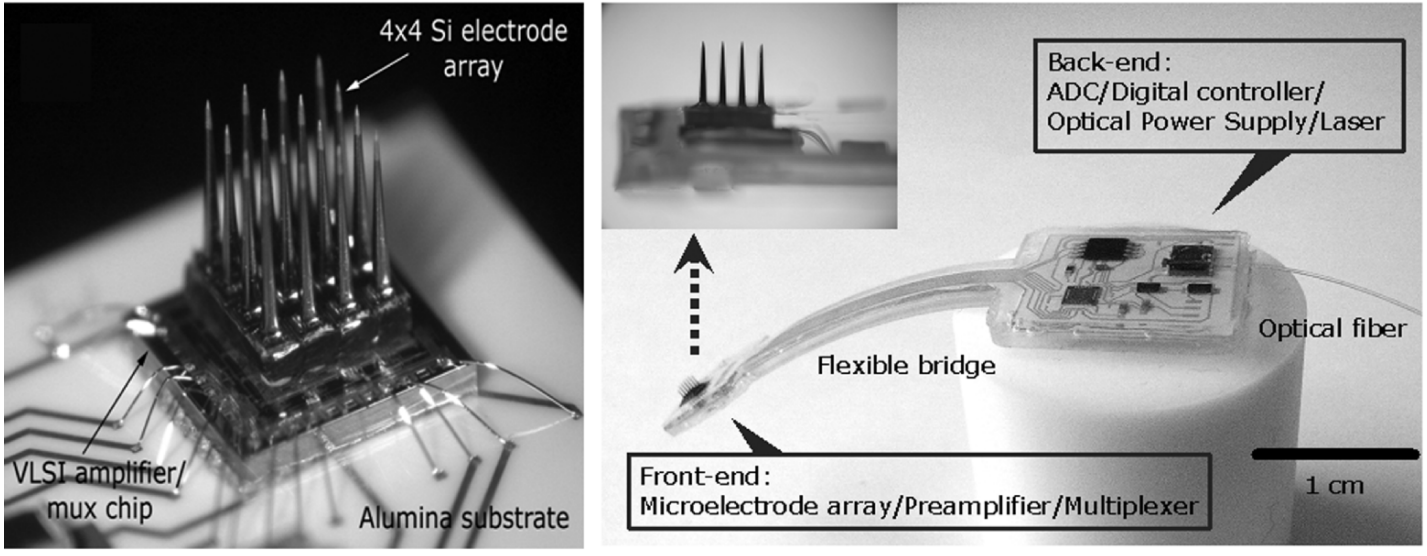


Figure 9.3 (Left): Photograph of a 4×4 element microelectrode array flip-chip bonded to a VLSI CMOS chip as an integrated cortical neuroprobe; (right): the full implantable microsystem fabricated on a flexible LCP substrate and encapsulated in silicone. The "back-end" panel houses A/D and command/control CMOS chips, together with a photovoltaic converter for incoming power/clock, and a VCSEL for outgoing IR digitized data stream, both using optical fiber.

ized front-end neuroprobe device whose separate recording characteristics have been reported in detail [7].

The dual-panel microsystem distributes its micro- and optoelectronic components in specific functional groupings. The multiplexed analog neural signals are routed from the front panel to the peripheral circuits of the second panel ("back end") of the implant. The back end circuitry integrates a low-power Analog-to-Digital converter (ADC), a digital control-and-command chip, an optional microcrystal photovoltaic energy converter, and an infrared (IR) data-out transmission link employing a very low threshold current IR vertical cavity surface emitting semiconductor laser (VCSEL). The concept of an optoelectronic data link enables considerable system flexibility for future neural prosthesis applications, including very large bandwidth capability. Extraction of the digitized IR neural signals from the neural interface unit can be performed either via a free space beam from the laser emerging directly through the skin, which is sufficiently transparent in the near infrared, or via a fiber optic strand threaded subcutaneously to other information linking sites in the body (e.g. a thoracic unit). Likewise, our platform enables different options for power delivery to the active components of the implant. In this implementation, the >50% efficient photovoltaic GaAs/AlGaAs three tandem cell photovoltaic energy converter is employed, but the system can be adapted to an inductively coupled RF link which we have also demonstrated in the laboratory.

To summarize the rationale for the architecture and design of our implantable, compact dual platform neural interface, we note that it is based on the convergence of engineering considerations, input from neurosurgeons and neurologists, all leading to a number of practical electronic and biomedical device choices, compromises, and solutions, with the ultimate goal of human implants providing the project motivation. Important factors include practical surgical considerations, minimizing the heat transferred by the unit directly to the cortex (temperature rise below 1°C based on thermal modeling scaled up to a full 10×10 array, i.e. a 100 channel system), mechanical stability against stress on the tether from movement of brain, the system's scalability to ever higher performance in terms of its signal bandwidth, its convenient modularity, adaptability to different wireless telemetries and powering (infrared, either free space or via optical fiber, RF), adaptability to multiple encapsulation strategies, and so on.

In the following, we focus specifically on the microelectronic design and implementation of the IC components that form the electronic backbone of the microsystem, and describe our approaches and solutions to a number of

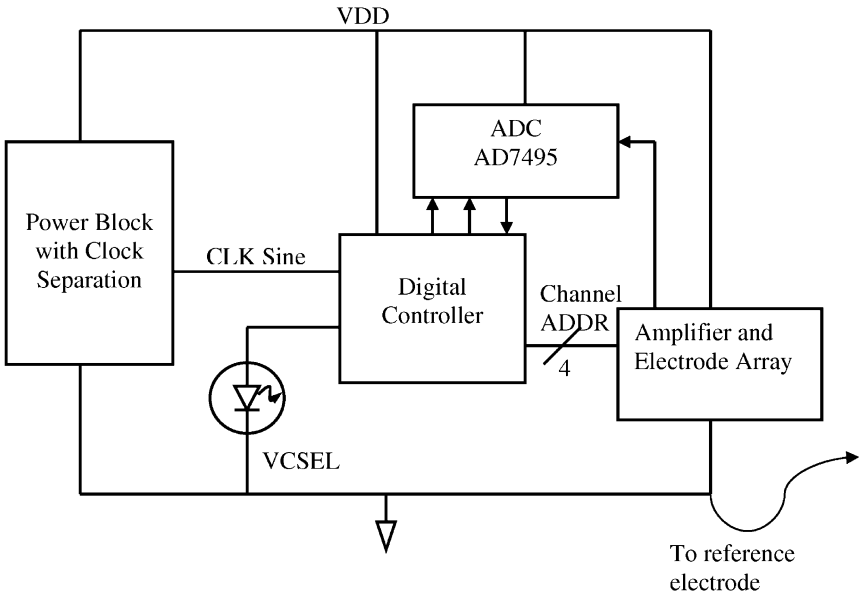


Figure 9.4 Block diagram of the electrical system of a 16-channel implantable sensor system incorporating the Utah/Cyberkinetics extracellular electrode array showing the use of a reconfigurable power module and the simple control signal interconnections

technically demanding problems many of which are generic challenges to any brain/head implantable biomedical device.

9.2 ELECTRICAL MICROSYSTEM OVERVIEW

A translation of the microsystem of Fig. 9.2 into an equivalent overall view of the system electrical architecture is shown in Fig. 9.4. The CKI microelectrode array is attached directly to the amplifier integrated circuit and is not shown explicitly on the block diagram. That electrode-amplifier assembly requires only seven connections to the backend circuitry for analog to digital conversion (ADC) and data handling, two lines for power and ground, four for channel/electrode selection, and one for the return of a multiplexed analog signal. The only other component on the front-end panel is a bypass capacitor for the power supply.

The analog-to-digital converter on the back-end panel is a type AD7594 manufactured by Analog Devices and packaged in a standard micro-SO8 package. It receives a clock signal and a start-of-conversion signal from a digital controller integrated circuit that also supplies the channel address

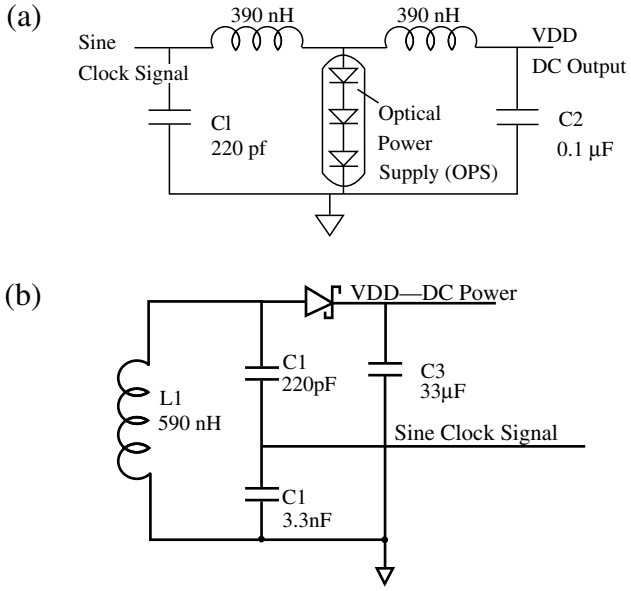


Figure 9.5 Power module configurations: (a) using a photovoltaic source power over an optical fiber. The three series connected diodes are an Optical Power Supply (OPS) built with stacked layers of GaAs/GaAlAs. (b) Power module configuration for RF inductive coupling to a half-wave rectifier.

data to the amplifier circuitry. The controller is a custom integrated circuit built in half-micron technology through MOSIS and is discussed in more detail below. The controller has two other functions. First, it multiplexes the ADC data with a periodic synchronization code word and converts the multiplexed data into a drive current to a vertical cavity surface emitting laser diode (VCSEL) to provide the optical output. Finally, the system clock is sent from outside the body by modulating the power source. The controller contains a comparator that derives its digital clock from a small sinusoidal signal separated from the DC power in the power module.

The power module can be configured two ways: the system can be powered by optical energy at 850 nm over a fiber connection or it can be powered through RF inductive coupling at 13.56 MHz. When power is delivered by fiber coupling, the VCSEL is also coupled to a second fiber for data return. When RF power is used, the VCSEL exploits the transparency of the skin at this wavelength to return data by free-space coupling transcutaneously.

Figure 9.5 shows the circuits for the two power module configurations. The fiber-coupled system of Fig. 9.5(a) uses an optical power supply (OPS) device having three stacked photovoltaic p-i-n junctions built with GaAs and

GaAlAs layers by molecular beam epitaxy. The thickness of each intrinsic layer is optimized to absorb 1/3 of the incident optical power at 850 nm. The device achieves greater than 40% electrical conversion efficiency with an open circuit voltage of 3.2 V and an optimum operating point of 2.8 V. (The electrical components are designed to function within specifications for $2.7\ \text{V} \leq V_{DD} \leq 4.5\ \text{V}$ although power is obviously minimized by operation at the lower end of that range.) More details of the photovoltaic structure are discussed in [9]. The 850 nm laser diode source that supplies energy to the OPS is modulated at 15.36 MHz to give an ADC sampling and data return of 40 ksps for all channels. One LC filter section separates the DC part of the output of the OPS, using a large bypass capacitor to assure a low noise output. That capacitor is essentially a ground connection to the RF signal. The two inductors and the small capacitor at the sinusoidal output of the module form a parallel resonant circuit that is driven by the RF current from the OPS. The sinusoidal output is about 0.2 Vpp.

When the system is powered by inductive coupling, the circuit is configured as shown in Fig. 9.5(b). The inductive coupling coil, L1, is actually a spiral trace on the underside of the liquid crystal polymer (LCP) substrate. It has a total inductance of 0.59 μH. Capacitors C1 and C2 in series form a parallel resonant circuit with L1 at the excitation frequency of 13.56 MHz. The capacitors also form a voltage divider to supply an appropriate sinusoidal signal to the clock comparator input to the digital controller. With this clock frequency the system sampling rate is 35.3 ksps. Finally, a simple schottky diode half-wave rectifier supplies the DC power.

9.3 PREAMPLIFIER AND MULTIPLEXOR INTEGRATED CIRCUIT

Figure 9.6 shows the architecture of the amplifier integrated circuit in block diagram form. An array of sixteen preamplifiers is arranged in four rows of four amplifiers, and each row has a separate circuit at one end for row selection and biasing. Amplifier selection is done by connecting all amplifiers in one row to a set of column lines and by using an analog multiplexer to select one column to be connected to a unity-gain output buffer amplifier. A common bias circuit generates cascode bias voltages and multiple copies of a reference current with which to set the amplifier bias points.

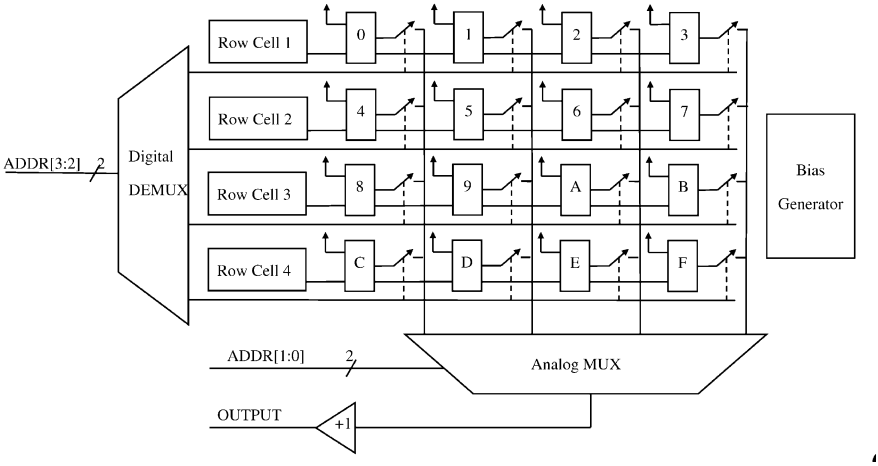


Figure 9.6 Block diagram of the amplifier integrated circuit showing the preamplifier-multiplexer-buffer signal paths.

9.3.1 Preamplifiers

We have built two versions of this circuit, which differ primarily in the choice of preamplifier feedback circuit. While our clinical results have primarily used the earlier device, we report the more recent circuit here because its noise performance is substantially improved. The basic preamplifier block is a folded cascode operational amplifier as shown in simplified form in Fig. 9.7. The primary simplification is that the figure does not show the segmentation of the current source devices into multiple identical parallel devices to improve current ratio matching. The selection of the folded cascode circuit was driven by its ease of frequency compensation, by its very high DC gain and by its ability to operate with low bias on its input gates. This last consideration is less important in the present circuit because the input-transistor gates are deliberately biased at several tenths of a volt.

Of the five bias voltages entering the left side of Fig. 9.7, three (Vbias2, Vbias3 and Vbias5) are derived in a central bias generator and are distributed throughout the chip. Vbias2 and Vbias3 are used throughout the system for the gates of N-channel and P-channel cascode transistors respectively and all cascoding transistors have approximately the same current per unit width. The central bias circuit also produces a 1.25 mA master reference current and copies of that current go to each row bias cell where a pair of current mirrors unique to that row derive Vbias1 and Vbias4 to set the values of current sources in those preamplifiers.

An unusual feature of the circuit is a common-drain output buffer stage consisting of M11 and its bias current source M12. The purpose of the stage

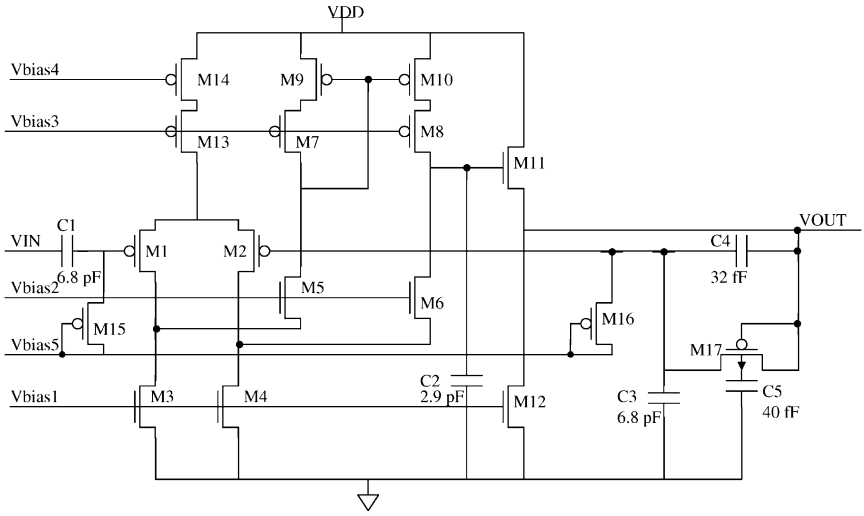


Figure 9.7 Simplified schematic of a preamplifier, including on-chip RC input coupling and a MOSFET feedback network.

is to lower the output impedance and to isolate the compensation capacitor from the output. The row-select circuit contains a switch connecting the output of the preamplifier to a column line shared by multiple preamplifiers and the column multiplexer. Without isolation, there would be charge sharing between the compensation capacitor and the column capacitance on switch actuation. In a 10×10 array, the column capacitance is comparable to the compensation capacitance, which would result in large errors. We have tried to design the system so it is scalable to use with a 10×10 set of electrodes. Even in the present 4×4 case, the error can be significant. The switching rate is too fast for the amplifier's closed loop response to correct for this effect within the settling time available for the first column conversion cycle.

The input coupling network (C1 and M15) and the feedback network (C3, C4, M16 and M17) are similar to those employed by Harrison [10]. We have chosen to use a single P-MOSFET in the branch from output to inverting input and in the input-coupling network instead of two. This sacrifices some low frequency response but provides a highly stable and predictable DC operating point. The three MOSFETs are completely cut off and therefore are essentially being used as unbiased diodes. M17 is similar to the configuration used originally by Ji, Wise and Najafi [11]. The bias voltage, Vbias5, applied to one end of M15 and M16, is about 0.65 V and is derived by passing a current from the bias generation block through a large (51 kΩ) on-chip resistor. The drain junctions of M15 and M16 are at zero bias and

their zero-bias resistances serve to set the gate node voltages for the input transistors to the 0.65 V level. From the observed step function response of the amplifier, the dynamic resistance of these minimum-sized junctions appears to be $\geq 1 \times 10^{12}$ Ω.

The DC operating point of the amplifier output is determined by feedback through M17 to the inverting input of the operational amplifier. The amplifier input transistors, M1 and M2, are sufficiently large and the DC gain is low enough that there is no significant offset from the input transistors. The circuit comes to equilibrium when the output forces the body of M17 to the 0.65 V bias on the input coupling resistance. This leaves the source-body junction on the amplifier output end of M17 slightly forward biased to compensate for leakage current between the body of M17 and the substrate. We are unable to measure that difference voltage directly, but we estimate it to be about 0.1 V.

Capacitors C3 and C4 form a voltage divider for AC signals and their ratio determines the in-band AC gain of the amplifier as $G_{AC} = 1 + C3/C4$. To make this gain reproducible from run to run, these capacitors are designed using two polysilicon layers, for which the area ratio will be the capacitance ratio within some small percentage error. However, C4 is bridged by M17, which has junction and gate capacitances that are significant compared to C4. Particularly the capacitance of the forward biased junction is poorly determined. To reduce the effects of the drain and source junctions, the capacitance from body to ground is augmented by C5 sufficiently to reduce feed through from the diode junction capacitances by an order of magnitude. Our design target for the mid-band gain was 45.6 dB and the measured result was 44.3 dB. We believe the discrepancy is a failure to account for all the parasitic capacitance bridging C4 in the actual layout.

The input coupling capacitor, C1, is also built using two layers of polysilicon. To keep the total amplifier area within the available space below an electrode, these capacitors were built underneath the input bonding pads. Such a structure might be a reliability concern if conventional ultrasonic or thermocompression bonding were used to attach the electrode array. However, we use an epoxy bonding technique that does not appreciably stress either the pads or the electrode array. The input pads do not have explicit electrostatic protection circuitry but with careful handling we have not seen any malfunctions due to electrical or mechanical stress.

The high frequency response of the preamplifiers is essentially a single pole with a gain-bandwidth product set by the current gain of the input stage and the compensation capacitor C2 as $f_{GBW} = g_{M1}/(2\pi \cdot C2)$. With this assumption, the high-frequency 3 dB point will be $f_{-3dB} = g_{M1}/(2\pi \cdot C3 \cdot G_{AC})$.

The central bias generation circuit uses a constant transconductance circuit to set the drain current of the input transistors, implying that the bandwidth will be set primarily by the time constant of a resistor-capacitor pair. The design goal was a bandwidth of 7.5 kHz based on the bandwidth of a successful commercial amplifier system [12]. The measured cutoff of the circuit is 7.3 kHz. Figure 9.8 shows the frequency response as measured with a Hewlett Packard gain-phase meter model HP3575A.

There are three sources of noise that limit the detectable signal, thermal noise from the device channel resistances, flicker noise primarily in the N-channel devices, and thermal noise in the input coupling and feedback resistances as filtered by the input and feedback capacitances. MOSFET thermal noise referred to gate-source voltage is approximately independent of frequency so its contribution to the spot noise variance of the output follows the frequency response of the amplifier. By using relatively large transistors, it is possible to move the contribution of flicker noise to the system to frequencies below 1 kHz. The noise effects of the input and feedback resistances primarily affect very low frequencies at or just above the low frequency cutoff. This contribution to low frequency noise sets a limit to the utility of the device for detecting small signals from such low frequency activity as local field potentials.

The primary contributors to thermal noise are transistors M1, M2, M3, M4, M9 and M10. Contributions from the cascode transistors M5 – M8 are reduced by source degeneration. Because their noise currents are uncorrelated, the mean square drain current fluctuations of each of these transistors

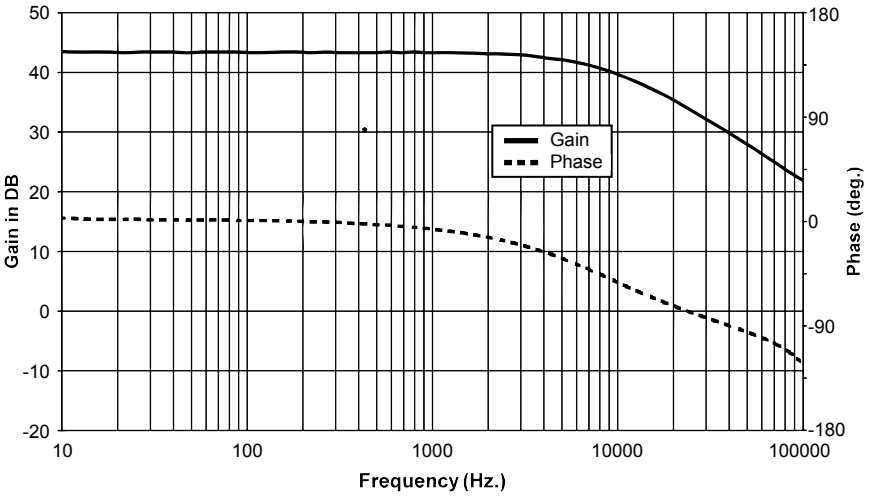


Figure 9.8 Bode plot of the gain and phase of a typical channel.

add directly at the compensation capacitor node, the common node for the drains of M8 and M6 at the gate of M11. This is obvious for the currents of M3, M4, M9 and M10 but slightly less obvious for M1 and M2. For example, if the current source, M13, is near ideal, then source degeneration of M1 by the impedance of the source of M2 reduces the drain current fluctuation of M1 by a factor of 1/2 but an equal amount of correlated noise current enters the source of M2. The two currents add together again at the compensation current node.

For a single long-channel transistor, the mean square drain noise current is given by [13]

$$\overline{i_d^2} = 4k_B T \frac{2g_m}{3} \Delta f \tag{9.1}$$

where $k_B T$ is the product of Boltzmann's constant with absolute temperature and g_m is the device transconductance. Summing the noise currents of the most important transistors and dividing by the square of the transconductance of the amplifier, gives an input referred mean square noise voltage density as

$$\frac{\overline{v_n^2}}{\Delta f} = \frac{16k_B T}{3g_{m1}} \left(1 + \frac{g_{m3}}{g_{m1}} + \frac{g_{m9}}{g_{m1}} \right). \tag{9.2}$$

This input referred noise is amplified with the same frequency response as the input signal. Since that response is essentially single pole, the total input referred noise from device thermal noise is

$$\begin{aligned} \sqrt{\overline{v_n^2}} &= \sqrt{\frac{16k_B T}{3g_{m1}} \left(1 + \frac{g_{m3}}{g_{m1}} + \frac{g_{m9}}{g_{m1}} \right) \cdot \int_0^\infty \frac{df}{\left(1 + (f / f_c)^2 \right)}} \\ &= \sqrt{\frac{16k_B T}{3g_{m1}} \left(1 + \frac{g_{m3}}{g_{m1}} + \frac{g_{m9}}{g_{m1}} \right) \cdot \frac{f_c \pi}{2}} \end{aligned} \tag{9.3}$$

where f_c is the closed-loop 3 dB bandwidth. In our amplifier, M1 and M2 have a W/L ratio of 11.2 μm/168 μm and are biased with $I_D = 4.5$ μA. Their typical transconductance is $4.1 \times 10^{-5}\ \Omega^{-1}$. Transistors M3 and M9 are biased with higher $V_{GS} - V_{TH}$ to lower their transconductances. The total input-referred noise voltage from thermal noise is estimated by Eq. (9.3) to be 3.9 μV_{RMS}.

Flicker or 1/*f* noise in this circuit is primarily due to M3 and M4 because N-channel MOSFETs exhibit this effect to a much greater degree than P-channel devices. For both device types, the mean square drain noise current is inversely proportional to the device gate area [14]. For a given drain current and transconductance, this means the noise is inversely proportional to the square of the device length, L_{eff}, that is,

$$\overline{i_d^2} \propto \frac{\Delta f}{f L_{eff}^2}. \tag{9.4}$$

For this reason, the device lengths were chosen to be as large as practical given that the preamplifier had to fit exactly beneath one electrode of the Utah array and area had to be allocated for both the electrode attachment pad and for substantial on-chip capacitors. The N-channel devices, M3 and M4 were the longest ones in the circuit with effective lengths of 28 μm. SPICE simulation suggested that the 1/*f* corner would be at 500 Hz for typical device parameters. As with thermal noise, flicker noise appears as an equivalent source at the input of the operational amplifier and its spectrum is modified by the amplifier's frequency response.

We have attempted to maintain sufficient low frequency response that the same device can be used to monitor local field potentials in the cortex, signals that require bandwidths down to a few Hz. Two factors limit the utility of the circuit for this purpose. The first is the low frequency cutoff due primarily to the dynamic resistance of M17, the feedback MOSFET, in parallel with feedback capacitor C4. Both a sinusoidal measurement manually with a low frequency generator and measurement of the response to a 500-μV step function give a low-frequency -3 dB point of 2.4 Hz. The second factor is that there is a marked increase in input-referred noise at low frequency due to both flicker noise in the operational amplifier and thermal noise in the transistors M15 – M17 used as large value resistors in the input coupling and feedback networks. The dynamic resistance of these devices is a source of Johnson noise at frequencies near the low frequency cutoff of the circuit.

Figure 9.9 shows an equivalent circuit for low frequency noise from the resistances of the input coupling and feedback transistors. The current sources in the figure are uncorrelated noise currents that are Norton transforms of the usual Johnson thermal noise given by

$$\overline{i_n^2} = \frac{4k_B T \Delta f}{R} \tag{9.5}$$

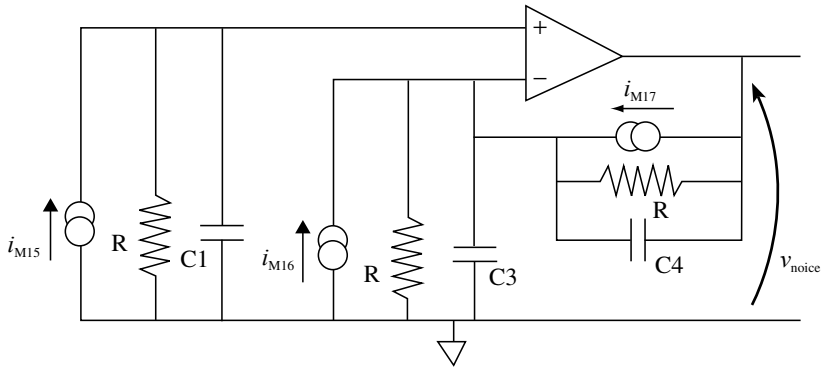


Figure 9.9 Noise equivalent circuit for the input coupling and feedback networks.

For purposes of estimating the effect of this source of noise, we have approximated all three MOSFET resistances as equal, assumed that the electrode impedance is negligible compared to R, and ignored the effect of C5. (The reported impedance at 1 kHz of typical Utah electrodes is in the range of 50 to 500 kΩ and the dynamic resistance of the input MOSFET is estimated as 1×10^{12} Ω.) It is well known [9] that the integrated noise of an RC combination like that from the non-inverting terminal of the operational amplifier to ground is $(k_B T/C)^{½}$. With an on-chip capacitor of 6.8 pF, this total noise voltage is 25 μV_{RMS}. This is much larger than the thermal noise from the devices in the amplifier circuit. However, much of the input-coupling noise is below the low-frequency cutoff of the circuit, requiring more careful analysis to show its effect on the spot noise density and on the ability to detect signals near but above this cutoff.

Each resistor in Fig. 9.9 represents the dynamic resistance of a MOSFET in either the input coupling or feedback network. The current sources are Norton transforms of the usual Johnson noise voltage of the resistance, having uncorrelated mean square noise currents given by Eq. (9.5). We have assumed that the electrode impedance is small compared to the MOSFET resistance, that the operational amplifier is ideal, and that MOSFET resistances are equal. (The electrode is omitted in this figure but is a voltage source in series with C1 between it and ground.) The reported values of electrode impedance at 1 kHz are in the range of 50 to 500 kΩ [16]. As discussed above, dynamic resistance of a zero-biased MOSFET junction is estimated to be 1×10^{12} Ω.

The noise voltage across C1 appears directly at the input of the preamplifier, and its noise voltage density is given by

$$\frac{\overline{v_1^2}}{\Delta f} = \frac{4k_B TR}{1+(\omega\tau_1)^2} \tag{9.6}$$

where $\tau_1 \equiv R{\cdot}C1$. The output-referred noise density due to the other two current sources is

$$\frac{\overline{v_2^2}}{\Delta f} = \frac{8k_B TR}{1+(\omega\tau_2)^2}$$

where τ_2 is defined as $R{\cdot}C2$. In practical measurements, one calculates an input-referred noise by measuring the output noise spectrum and dividing that by the mid-band gain of the system. The frequency dependant, low-frequency gain is

$$|G|^2 = \left|\frac{2+s(\tau_1+\tau_2)}{1+s\tau_2}\right|^2 \approx \frac{4+(\omega\tau_1)^2}{1+(\omega\tau_2)^2} \tag{9.7}$$

and the midband gain is $|G_{MB}|^2 \approx (\tau_1/\tau_2)^2$. Defining a frequency $f_1 = 1/(2\pi\tau_1)$ and combining equations yields an input-referred total mean square noise voltage density of

$$\frac{\overline{v_{in}^2}}{\Delta f} = \left[\frac{4k_B TR}{1+(f/f_1)^2}\cdot\frac{4+(\omega\tau_1)^2}{1+(\omega\tau_2)^2} + \frac{8k_B TR}{4+(f/f_1)^2}\right]\cdot\left[\frac{\tau_2}{\tau_1}\right]^2. \tag{9.8}$$

The frequency f_1 is the frequency at which the amplifier gain is 3 dB above its DC value and is the low-frequency, -3 dB cutoff frequency divided by the midband gain or 0.015 Hz.

Figure 9.10 shows the relative theoretical contributions to the low frequency noise spectrum from the three simple source models discussed here and Fig. 9.11 compares their total to the measured noise spectrum. The noise spectrum above 5 kHz was measured with a Hewlett Packard model HP3585B spectrum analyzer and below 10 kHz by power spectral density (PSD) transforms of ADC output data. With the input connected to ground, ten measurements of 3.2 sec each of 40 ksps digital output data (128 kB) were converted to PSDs and averaged together. The result was divided by the midband gain to derive an input-referred noise.

Below 10 Hz, noise from the dynamic resistances of the input coupling and feedback network MOSFETs dominates the spectrum. Flicker noise is

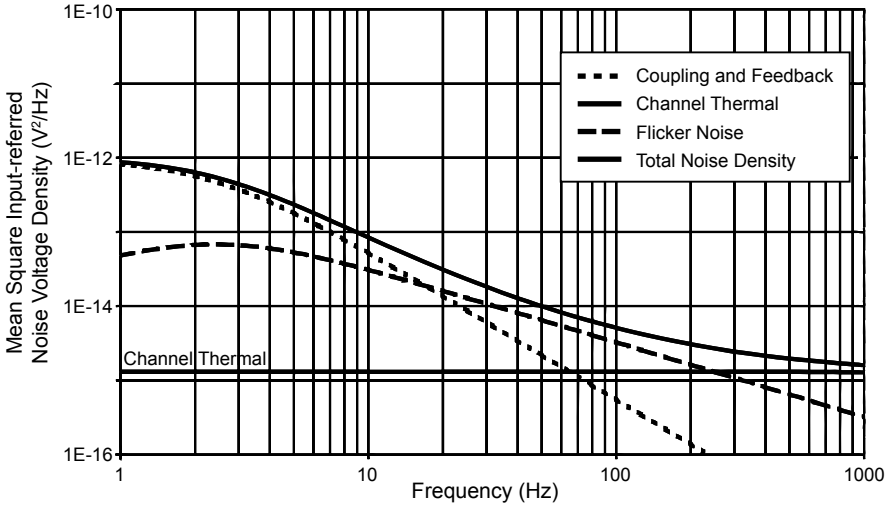


Figure 9.10 Simple theoretical model of the relative contributions of channel thermal noise, flicker noise, and input coupling noise to the overall mean square noise voltage spectral density.

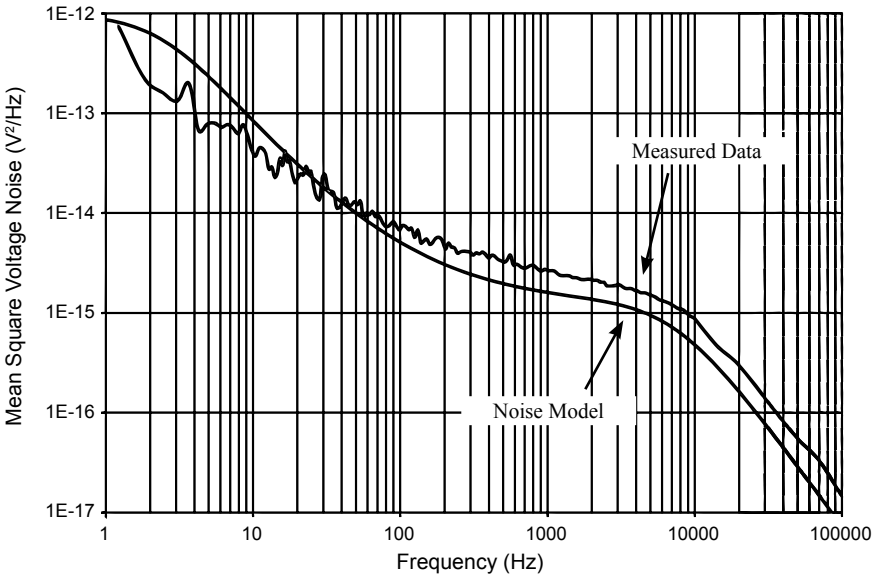


Figure 9.11 Comparison of the measured and theoretical amplifier spot noise spectral density distribution.

Table 9.1 RMS input-referred noise.

Frequency range	DC to 20 kHz (No filter)	50 Hz to 20 kHz	300 Hz to 20 kHz
Noise	5.20 μV	4.77 μV	4.50 μV

the principal contributor from 10 to 250 Hz. Finally, the present model underestimates the thermal noise by about 1.1 dB. Partly this is due to contributions from other transistors in the signal path, but we do not fully understand the discrepancy.

In addition to measuring the noise spectrum, we have computed the total mean square noise at the output divided by the midband gain with three slightly different measurement conditions. The first calculation computed the RMS value of the same data used to draw Fig. 9.11. The other two calculations were made from similar data sets taken with a single pole, high-pass filter between the amplifier output and the ADC. Cutoffs of 50 Hz and 300 Hz were used to see how the low-frequency part of the spectrum affected the detectable signal level. As there was no anti-aliasing filter beyond the amplifier's own response, this data represents the amplifier noise that will be seen in actual biological applications. It includes some noise above 20 kHz aliased into the baseband from above the Nyquist sampling limit. The results are shown in Table 9.1.

9.3.2 Column Multiplexing

The circuit in Fig. 9.12 shows how the output of a preamplifier connects to a column line. A conventional bi-directional switch implemented in M19 and M20 makes the actual connection. It is activated by a ROW signal and its complement. These are 3 V digital signals in a system with noise levels under 5 μV_{RMS}. Since their traces actually cross all the preamplifier cells, the layout includes careful shielding with grounded metal to eliminate any appreciable capacitive coupling to either the input pads or to the electrodes sitting above the preamplifiers. The source follower, M18, serves two purposes: first, it continues the strategy of protecting the amplifier output from the effects of charge sharing with the parasitic column capacitance as discussed above. Second, it increases the current available to drive the column line, lowering the output impedance of the preamplifier cell. In operation, the follower bias current is 8 μA, and the quiescent potential of the output of the circuit is 1.68 V. To save power, the current source for the follower is turned off when the row is not selected. When the system is scaled for a

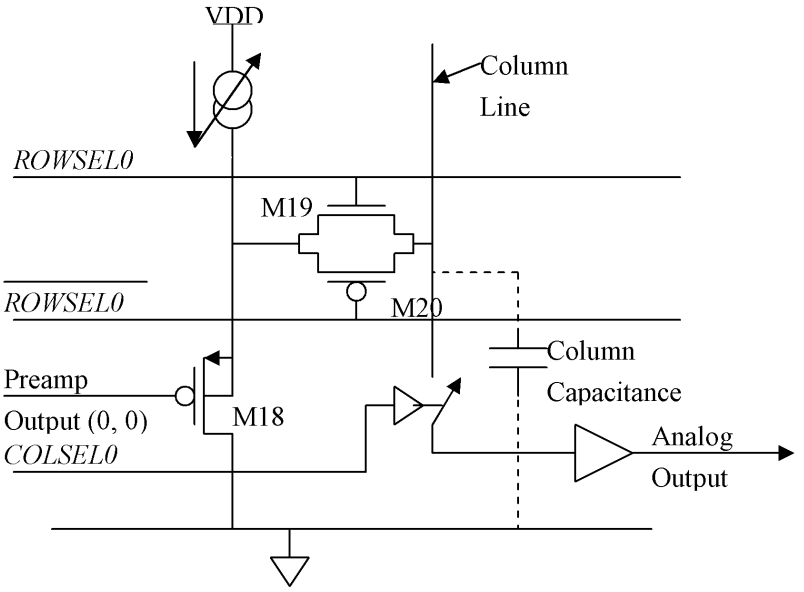


Figure 9.12 Column multiplexer schematic, showing a final source follower driver and the column capacitance it is necessary to drive.

10 × 10 electrode array, the current will be turned on shortly before the row to reduce the settling time of the output.

9.3.3 Output Buffer Amplifier

The output of the column multiplexer is amplified in a 1× buffer amplifier implemented as an operational amplifier with unity feedback with the circuit shown in Fig. 9.13. The principal requirement on this circuit beyond its settling time is a result of the properties of the ADC. The Analog Devices AD7495 is a switched capacitor, successive approximation ADC and, as is common in such devices, it has no input buffer amplifier. As a result, its input appears as a 20 pF capacitance during signal acquisition and 4 pF during conversion. The 16 pF sampling capacitor is always fully discharged as it switches to the input connection. In addition to the ADC capacitance, the buffer amplifier must also drive the capacitance of its own output pad and a significant substrate capacitance. (The liquid crystal polymer substrate is 75 μm thick and the analog output trace runs directly over a ground trace for 3 cm to reduce RF induced noise.) The buffer amplifier must settle quickly from this step excitation and exhibit no instability. To mimic the anticipated

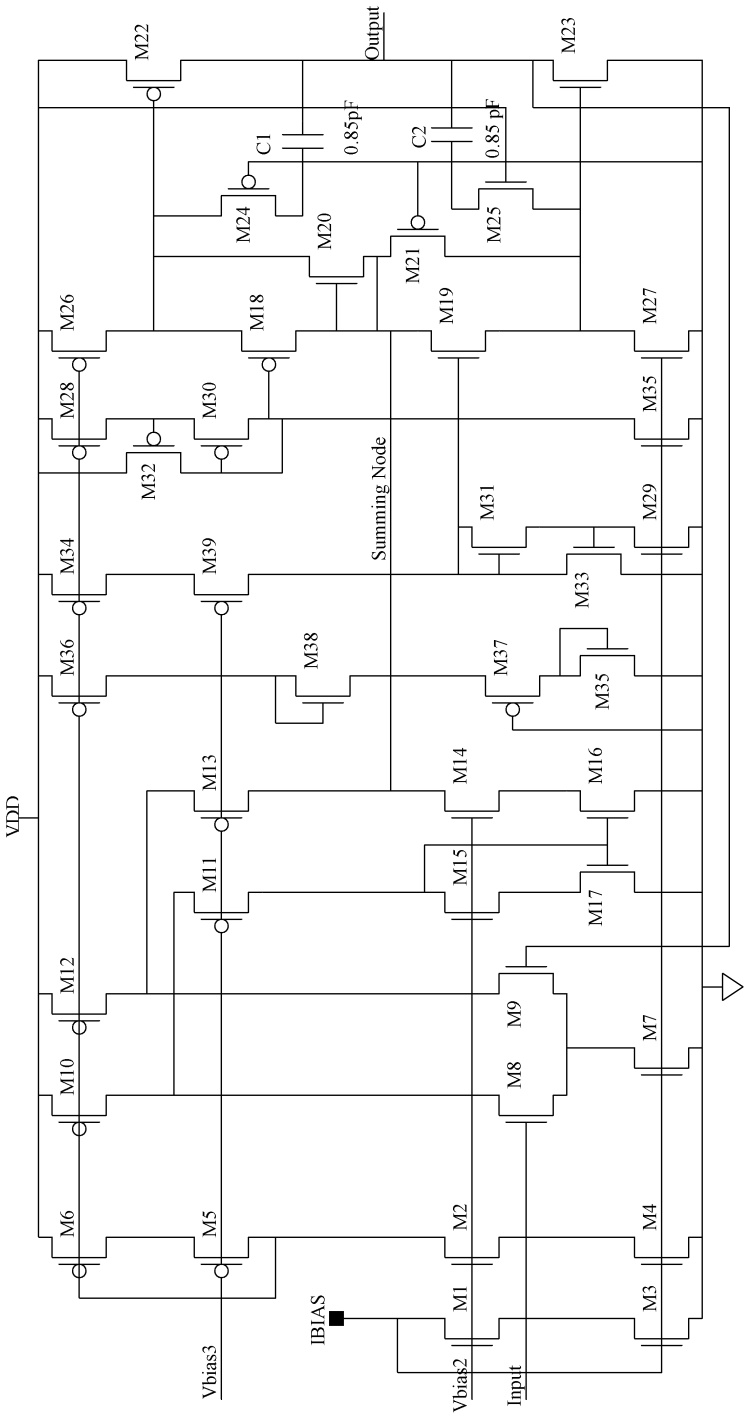


Figure 9.13 Simplified schematic diagram of the output buffer amplifier. The primary simplification is omission of the paralleled devices used to assure accurate scaling between current sources.

situation of a 10 × 10 electrode array system, in which we will use two ADCs doing alternate conversion and acquisition processes, we have set the present system to use 24 clock cycles to do a 12-bit conversion. Thus the amplifier sees 6 pF and 22 pF switching loads for 12 cycles or 0.77 μsec each.

The buffer amplifier has three sections, the first of which is a current mirror implemented in M1 – M6. One copy of the 1.25 μA reference current flows through M1 and M3 establishing control signals for both N and P current sources. These are cascoded sources and the bias voltages for the cascoding transistors are the same signals, Vbias2 and Vbias3, as are used for this purpose throughout the system. The circuit diagram does not show the division of current sources into multiple copies of a single-size transistor to produce different scaled values of current as required. This segmentation is also used on the output transistors M21 and M22 to set their quiescent currents based on matching to transistors in the current sources.

The second section is a straightforward folded cascode amplifier implemented by M7 – M17. The 1.68 V quiescent voltage at the output of the channel multiplexer dictates the choice of an N-channel differential input stage. The input transistors M8 and M9 have a quiescent current of 2.5 μA each. The output current of this stage appears at the junction of the drains of M11 and M14 and is applied to a summing node at the input of the output stage. (M20 and M21 are common gate amplifiers, the sources of which are connected together. That gives this node a low impedance whenever either transistor has an appreciably more current than the other. The drain currents of the input differential stage sum at this node and pass into the source of M20 or M21 depending on the sign of the net current. The impedance of this node is a more subtle question at quiescence because of the local feedback mechanism of M19 and M18 which are responsible for the class AB action of the output stage.)

The output stage is a class AB design using a circuit devised by Duisters and Dijkmans [17] and the reader is referred to their paper for further details. Capacitors C1 and C2 are compensation capacitors for positive or negative going output signals respectively and transistors M24 and M25 are effectively resistors in series with the compensation capacitors. This resistance performs two functions, moving the zero in the transfer function from the compensation capacitance to the left half plane and preventing the turn-on of an output transistor during slew when it is supposed to be turned off.

SPICE simulation has shown this circuit to be stable with the switched load and to be robust against process variation. It has a slew rate of 5.3 V/μsec with the 22 pF load, which requires a peak output current of 125 μA. Settling time for the overall amplifier is about 750 nsec, which is adequate

for our present system. The unity-gain frequency is 3.9 MHz and the phase margin is 57° in the typical case. Power consumption is 210 μW with a 3 V power supply.

9.3.4 Biasing and the Bias Generator

As mentioned above, there is a central bias generator circuit, the schematic for which is shown in Fig. 9.14. It produces the two cascode bias voltages, Vbias2 and Vbias3, as well as a master reference current, Ibias. A bootstrap bias circuit consisting of an unbalanced current mirror/current source (M1 – M4) attached to an N-channel current mirror (M5 – M8) generates the 1.25 μA reference current, Ibias [18]. The choice of putting the unbalanced mirror in the P-channel section is dictated by the N-well process that only allows a P-device to have a floating body. The segmentation of current sources is not shown explicitly but rather indirectly by a notation next to each of the P-channel source indicating the relative width of that source, for example, M1 is 4× the width of the output source M21. In the unbalanced mirror, M2 is four times wider than M1 and a voltage drop across the Rset resistor makes up the difference in $V_{GS1} - V_{GS2}$. At low currents, the current in M2 exceeds that in M1 because M2 is larger. At high currents, the gain of M2 is limited by degeneration from the resistor, so the current in M1 is larger. There are only two currents at which this configuration can come to equilibrium: zero amperes and a stable bias current determined by when

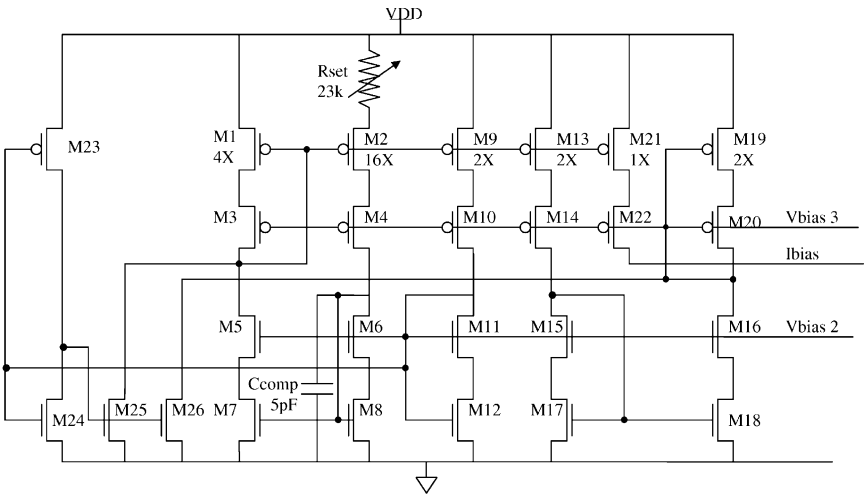


Figure 9.14 Global bias generator for the N and P-channel cascode gate voltages and for a 1.25 μA master bias reference current.

$g_{m1}=1/R_{set}$. (This derivation assumes that these are square-law devices.) The large set of 16 parallel transistors making up M2 are set in a separate well so that the well can be tied to the source of M2, eliminating body effect in that device.

The resistor is a long, narrow rectangle of N-well material, the apparent resistivity of which is not well controlled. We have found that the sheet resistivity of an N-well structure depends strongly on the width of the well. For the AMI ABN 1.5 micron process, the resistivity only approaches the value specified in the manufacturing process characterization when the width is greater than 25 microns. (The well depth is about 5 microns and the width dependence of the resistance is probably due to loss of effective dopant due to lateral diffusion at the edge of the well.) We have used resistors that are 20 microns wide but have found that compensating for the width effect increases the uncertainty in the final resistance. Because the drain current is inversely proportional to the square of the resistance, that is, $I_{D1} \propto 1/(R_{set})^2$, we have made the resistor value adjustable. The resistor has three taps on increments of 6% of the total resistance. The taps are at the VDD end of the resistor and are connected to bonding pads next to the VDD pad. An external wire-bond jumper connects the appropriate tap to VDD at assembly so that the drain current and hence all the circuit currents can be set within about 3% of their nominal value.

This bootstrap current source is a feedback system, which can be unstable. The potential instability is aggravated by the derivation of the cascode voltages from the current source itself. The compensation capacitor Ccomp provides a dominant pole in the loop that is sufficient to suppress any oscillation.

The bias voltages, Vbias2 and Vbias3 are derived by passing a scaled copy of the reference current through what is effectively a diode connected MOSFET. For example, M9 and M10 provide a constant current of 2.5 μA that generates the N-cascode bias, Vbias2, when it passes through M11 and M12. M14 through M19 serve the same purpose for Vbias3, the P-cascode bias. Segmenting the diode connected MOSFET into two parts serves the usual purpose of simplifying the analysis of simulation data during design. Because zero current is also a stable point of the bootstrap source, the system must have a circuit to start the current flow. This is the function of M23 – M26. M25 provides a startup current for the bootstrap mirror circuit and M26 does the same for the cascode voltage supply. A low-threshold inverter (M23 – M24) turns off the startup current when the N-cascode voltage, Vbias2, reaches a minimum level.

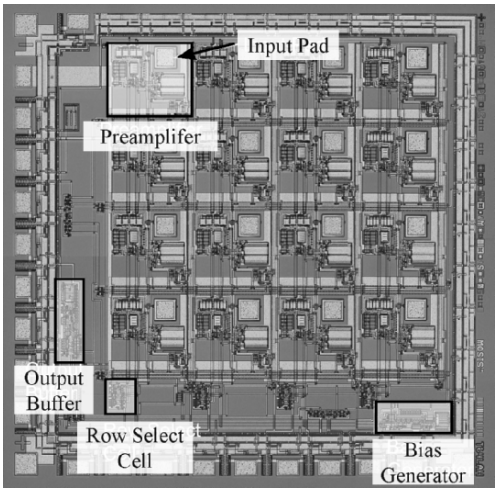


Figure 9.15 Photomicrograph of the amplifier integrated circuit, showing the size and location of the preamplifiers, bias generator and output buffer.

9.3.5 Amplifier Performance

Figure 9.15 shows a photomicrograph of the final amplifier integrated circuit and Table 9.2 gives the principal specifications and the measured results. The circuit was fabricated in the AMI ABN 1.5-micron process through MOSIS on a die size 2200 μm × 2200 μm. There are wire-bonding pads only along two sides of the die, some of which are used for diagnostic connections. Triple ground connections are used reduce any possible coupling between the single-ended preamplifier inputs and the ground currents of the output buffer or digital demultiplexer inputs.

Table 9.2 Amplifier characterization.

Parameter	Simulation	Measured
Noise (RTI 300Hz to 20 kHz)	3.9 μV_{RMS}	4.5 μV_{RMS}
High frequency cutoff (-3 dB)	7.5 kHz	7.3 kHz
Low frequency cutoff (-3 dB)	N.A.	2.5 Hz
Power per amplifier	50 μW	N.A.
Power	1.3 mW	1.4 mW

9.4 DIGITAL CONTROLLER INTEGRATED CIRCUIT

Figure 9.16 is a block diagram of the digital controller that selects electrodes and uses the multiplexed digital data to drive a VCSEL for data return. This device was built using the AMI 0.5-micron process through MOSIS on a 28-pin 2200 μm × 2200 μm die. The first function of this controller is to convert a small (≈200 mV) sinusoidal signal from the power supply modulation into a full digital clock. That is the purpose of the clock comparator, the circuit for which is shown in Fig. 9.17. To avoid injection of digital noise at the comparator input from coupling through the VDD and GND bonding wires, there are separate power and ground supply connections for the comparator.

The input of the clock comparator is capacitively coupled to a simple differential amplifier (M1 – M4) and a minimum-size, diode-connected P-MOSFET (M8) prevents overdriving M5. The second stage, M5 and its current source M14, is a simple inverter with minimal input capacitance. It is followed by a conventional inverter (M6 – M7) to supply the output clock signal. The bias voltage for the differential amplifier is derived by filtering the DC component of the output signal with the network of M9, M10 and C2. When the digital output is high, M10 acts as a constant current source to charge C2. Similarly when it is low, M9 discharges C2. By choosing the size ratio of these two MOSFETs properly, one can force the duty cycle of

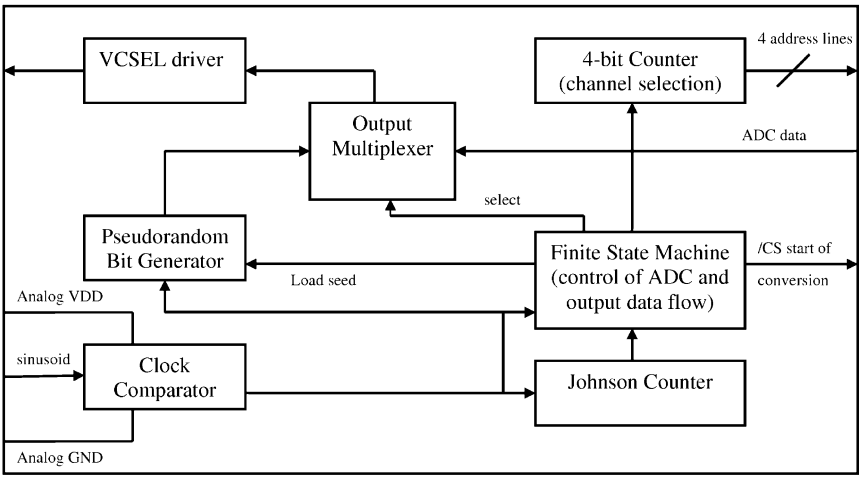


Figure 9.16 Block diagram of the digital controller integrated circuit that separates the clock signal from modulated power and uses that to select channels, to clock the ADC, and to merge synchronization and neural data onto a VCSEL-based data link.

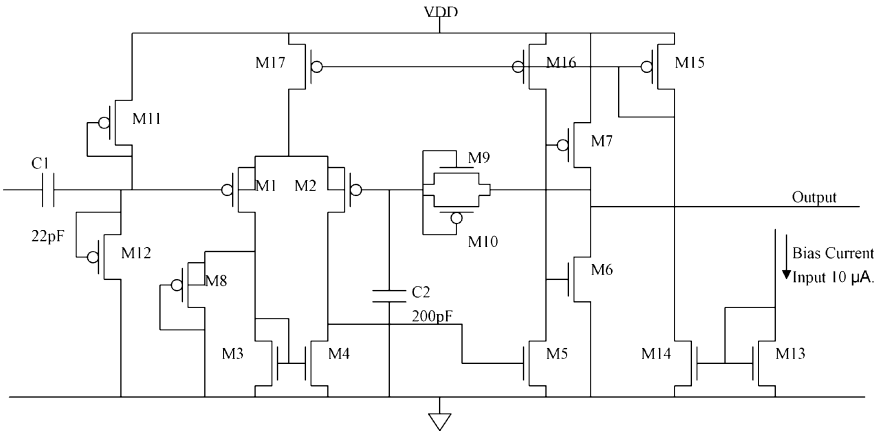


Figure 9.17 Clock comparator schematic. This circuit generates a full CMOS clock from a 0.2 Vpp sinusoidal signal derived from the power modulation. The relative sizes of M9 and M10 determine the output duty cycle.

the clock to be between 40 and 60% over all process variations. C2 is implemented as a large-area accumulation N-MOSFET built in an N-well. That gate-source capacitance is augmented by a second layer of ground polysilicon on top. The comparator frequency response sets the limiting clock rate and has demonstrated full functionality from 10 to 30 MHz.

A modified Johnson counter divides the clock down by 24:1 to the ADC conversion rate. Because all flip-flops in the system are either clocked continuously or held in reset, we used simple dynamic flip-flops throughout. The advantage of a Johnson counter in this application is the ease of generating glitch-free time-shifted outputs of any duty cycle. A very small finite state machine uses the ADC clock signal to increment the channel selection counter and to control the insertion of a synchronization word into the output data in place of the data from the 16^{th} channel. The synch word itself is generated by a pseudorandom bit generator (PRBG). The PRBG is a presetable 5-bit shift register with appropriate XOR feedback [13.]. The preset seed assures that the first 24 bits of the 31-bit output pattern are very improbable in the analog data, so the receiver detects this sequential bit pattern to determine the beginning of a frame of data. The output multiplexer also retimes the ADC data as it merges that data with the synch word. The VCSEL requires a 2.4 mA current for full output and a 2 mA current in the off condition to minimize the turn-on time. The VCSEL driver is a simple PMOS current source in parallel with a PMOS switch-resistor combination to supply the on-current.

9.5 CONCLUSIONS

In this chapter we have focused on the microelectronic design and implementation of components which form a brain implantable neural recording microsystem presently under development in our laboratories. The microsystem has been tested to date both in physical experiments as well as in acute recording from rats [8], with the next goal being trials in monkeys. A performance demonstration result of the full 16 channel microsystem is shown in Fig. 9.18, in a benchtop evaluation test, where the microelectrode array has been excited by a series of 'pseudospikes' (or artificial neurons) by applying a bipolar transient voltage through a wire to a saline solution (ACSF, which mimics closely the electrolyte background in the brain) into which the microsystem unit was immersed [8] [labeled "electrical" in Fig. 9.18(a)]. After detection and reconstruction of the received IR optical data stream transmitted by the full microsystem into analog form (labeled "all optical system"), a typical series of pseudospikes at high fidelity and good signal-to-noise ration were obtained as shown. (The upper trace is acquired from electrically accessing the front end of the system only). The signals acquired from the all-optical system in the lower trace shows almost the same preamplifier gain of 42 dB as the front end microelectronic system, with comparable input noise characteristics. In Fig. 9.18(b), a detail of the of the digital optical data stream trace is shown on microsecond time-scale in the upper trace, in comparison with the digital sampling rate as indicated in the bottom clock trace. The pulse coded modulation (PCM) data at the photoreceiver demonstrates a robust fiber optic data link in the present microsystem. The lower trace originates from the sinusoidal waveform applied to the external diode laser power source, current modulated to generate an infrared optical input carrying 15.36 MHz clock signal.

In terms of the "electrical microsystem", our next efforts in the microelectronic area are primarily aimed at device deployment. Problems of encapsulation longevity and surgical compatibility are being explored in conjunction with animal models and physical testing. Circuit development has three parallel aims. First, the system must be scaled to significantly large number of channel (100) operation for full exploitation of the neural signal base, while the materials and assembly problems are being solved. The design of the basic amplifier and controller circuits should be relatively straightforward, although there are challenges with the speed of settling in the multiplexer. Second, we wish to scale the system to network with multiple cortical array recording sites using a common power and ADC panel to service them. The problems in this area include achieving ade-

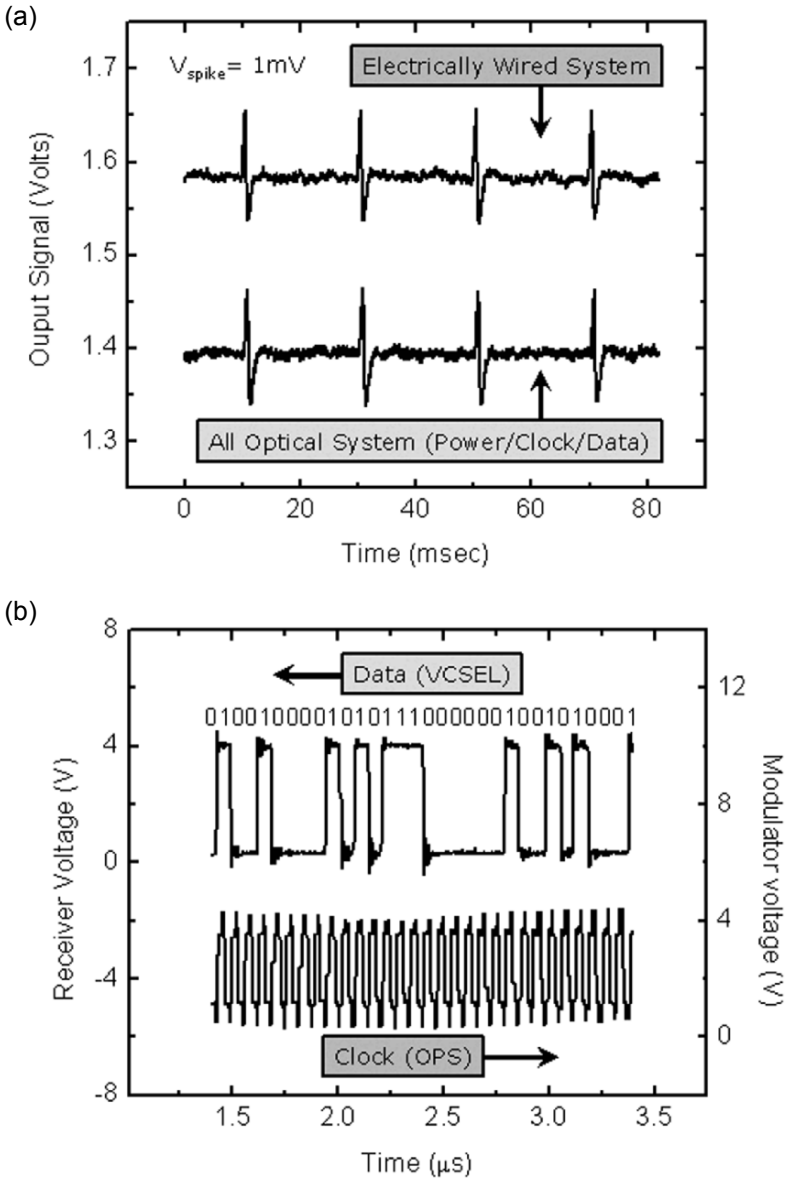


Figure 9.18 (a) Comparison of electrically recorded train of "pseudospikes", generated in a saline solution, with those retrieved from the digital optical data stream transmitted from the full microsystem by the IR optical fiber. (b) Upper trace shows a piece of the digitized outgoing optical data train while lower trace shows the incoming clock signal retrieved by photovoltaic conversion of modulated IR laser light entering the microsystem via the fiber.

quate ADC rates and improved flexibility of the tethering portions of the substrate. Finally, we have already begun the design of stimulus circuits both for cortical and peripheral neurostimulation. Our long-term view is that system-on-chip CMOS circuitry can be implanted almost anywhere in the body with the use of soft, high-quality packaging. The use of fiber optic communication, supplemented with RF power delivery, may someday allow substantial "rewiring" of damaged neurosystems in humans. In the process, what is learned of neuroencoding will contribute to our understanding of the central problem of 21st century neuroscience, namely the question of how intentions and thoughts are actually translated into action.

ACKNOWLEDGMENT

We wish to acknowledge M. Serruya, J. Simeral, M. Fellows, K. Davitt, J. Beals, N.-J. Hwang, J. Zhang, and H. Xu for their many contributions to the brain implantable microsystem project. We would like to thank several former students for their contributions to the design process for these circuits, especially Andrew DeAngelis, Christopher Lay, Michael A. Viamari, and Paula Petrica. The research was supported by grants from National Institutes of Health, the National Science Foundation, and DARPA.

REFERENCES

[1] M. D. Serruya, N. G. Hatsopoulos, L. Paninski, M. R. Fellows, and J. P. Donoghue, "Instant neural control of a movement signal," Nature, vol. 16, pp. 141–142 (2002); D. M. Taylor, S. I. Tillery, and A. B. Schwartz, "Information conveyed through brain-control: Cursor versus robot," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11 (2), pp. 195-199, 2003.

[2] L. R. Hochberg, M. D. Serruya, G. M. Friehs, J. A. Mukand, M. Saleh, A. H. Caplan, A. Branner, D. Chen, R. D. Penn, and J. P. Donoghue, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia," *Nature*, vol. 442, pp. 164-171 (2006); the website includes video clips

[3] M. Mojarradi, D. Binkley, B. Blalock, R. Andersen, N. Uslhoefer, T. Johnson, L. Del Castillo, "A miniaturized neuroprosthesis suitable for implantation into the brain", *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, pp. 38-42, 2003.

[4] I. Obeid, J. Morizio, K. Moxon, M. A. Nicolelis, P. D. Wolf, "Two Multichannel Integrated Circuits for Neural Recording and Signal Processing," *IEEE Trans. Biomed. Eng.*, vol. 50, pp. 255–258, 2003.

[5] R. Harrison, P. Watkins, R. Kier, R. Lovejoy, D. Black, R. Normann, F. Solzbacher, "A Low-Power Integrated Circuit for a Wireless 100-Electrode

Neural Recording System", *International Solid State Circuits Conference, February 2006*, Session 30, Silicon for Biology.

[6] W. R. Patterson, Y.-K. Song, C. W. Bull, A. P. Deangelis, C. Lay, J. L. McKay, A. V. Nurmikko, J. D. Donoghue, and B. W. Connors, "A Microelectrode/ Microelectronic Hybrid Device for Brain Implantable Neuroprosthesis Applications", *IEEE Trans. Biomed. Eng.*, vol. 51, pp. 1845-1853, 2004.

[7] Y.-K. Song, W. R. Patterson, C. W. Bull, J. Beals, N.-J. Hwang, A. P. Deangelis, C. Lay, J. L. McKay, A. V. Nurmikko, M. R. Fellows, J. D. Simeral, J. P. Donoghue, and B. W. Connors, "Development of a Chipscale Integrated Microelectrode/Microelectronic Device for Brain Implantable Neuroengineering Applications", *IEEE Trans. Neural Rehabil. Eng.*, vol. 13, p. 220, 2005.

[8] Y.-K. Song, W. R. Patterson, C. W. Bull, A. V. Nurmikko, M. D. Serruya, M. R. Fellows, J. D. Simeral, and J. P. Donoghue, "Development of Brain Implantable Microsystems With Infrared Optical Telemetry For Neuromotor Prosthesis", submitted to *IEEE Trans. Biomed. Eng.,* 2006.

[9] M. D. Serruya, Y.-K. Song, W. R. Patterson, C. W. Bull, A. V. Nurmikko, M. R. Fellows, J. D. Simeral, J. P. Donoghue, "Development of a brain implantable microsystem with infrared optical telemetry for advanced neuromotor prosthesis," Program No. 148.15. 2006 Neuroscience Meeting Planner. Atlanta, GA: Society for Neuroscience, 2006. Online.

[10] R. R. Harrison, "A low-power, low-noise CMOS amplifier for neural recording applications," IEEE J. Sold-State Circuits, vol. 38. pp. 958-965, 2003.

[11] Ji, J.; Najafi, K.; Wise, K.D., "A low-noise demultiplexing system for active multichannel microelectrode arrays," *Biomedical Engineering, IEEE Transactions on*, vol. 38, pp. 75-81, 1991.

[12] http://www.cyberkineticsinc.com/images/researchproducts/cerebus2.jpg

[13] P. R. Gray, P. J. Hurst, S. H. Lewis, R. G. Meyer, *Analysis and Design of Analog Integrated Circuits, 4th ed.* New York: John Wiley & Sons, 2001, p. 759.

[14] D. A. Johns, K. Martin, *Analog Integrated Circuit Design*, New York: John Wiley & Sons, 1997, p. 200.

[15] D. A. Johns, K. Martin, *Analog Integrated Circuit Design*, New York: John Wiley & Sons, 1997, p. 203.

[16] K. E. Jones, P. K. Campbell, and R. A. Normann, "A glass/silicon composite intrcortical electrode array," *Ann. Biomed. Eng.*, vol. 20, pp. 423-437, 1992.

[17] T. S. F. Duisters and E. C. Dijkmans, "A –90dB THD Rail-to-Rail Input Opamp Using a New Local Charge Pump in CMOS," IEEE J. Solid-State Circuits, vol. 33, pp. 947-955, 1998.

[18] D. A. Johns, K. Martin, *Analog Integrated Circuit Design*, New York: John Wiley & Sons, 1997, p. 248.

[19] W. T. Vetterling, *et al*., *Numerical Recipes in C 2nd ed.*, Cambridge [England]; New York: Cambridge University Press, 1992, pp. 130-139.

AUTHOR BIOGRAPHY

William R. Patterson was born in Bronxville, NY, in 1941. He received the Sc.B. degree in physics and the Sc.M. in electrical engineering from Brown University, Providence, RI, in 1963 and 1966. His current research interests include low-power analog circuit design for biomedical applications, circuits and architectures for microphone array technology, and instrumentation for geological spectroscopy. Since 1977, he has been with the Electrical Sciences Group in the Division of Engineering, Brown University, where he is currently a Senior Lecturer and Senior Research Engineer. In 1977, Mr. Patterson received the NASA Public Service Medal for contributions to the Viking program.

Yoon-Kyu Song received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 1992 and 1994 and the Ph.D. degree from Brown University, Providence, RI, in 1999. He was a Research Scientist at Agilent Technologies until rejoining Brown University as an Assistant Professor (Research) in 2003. His research interests include basic and applied semiconductor optoelectronics and neuroengineering.

Christopher W. Bull received the Sc.B. degree in mechanical engineering and the M.S. and Ph.D. degrees in materials engineering from Brown University, Providence, RI, in 1979, 1985, and 2006, respectively. He is a Senior Research Engineer and Lecturer at Brown University. His research interests include the neural-electronic interface, deformation of single crystals, and the strain rate dependence of material behavior.

Farah L. Laiwalla is born in Pakistan. She received the B.S degree in Electrical and Computer Engineering from Lafayette College, Easton, PA in 2003. From 2003 until 2006, she worked as a research associate at the Yale Medical School, working on the design of ultra-low noise amplifiers for electrophysiological recordings. She is currently pursuing an MD/PhD at Brown University. Her research interests include implantable electronics and the challenges of integrating such technologies into clinical practice, and she is currently involved in the design of a microsystem for cortical stimulation.

Arto V. Nurmikko was born in Finland. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of California at Berkeley. He is the L. Herbert Ballou University Professor of Engineering and Physics at Brown University, Providence, RI. His current research involves basic and device science of wide bandgap semiconductors, studies of ultrafast processes in ferromagnetically ordered systems, and high-resolution imaging of neural circuits, and brain-machine interfaces. He is a Fellow of the IEEE, of the American Physical Society and the Optical Society of America.

John D. Donoghue received the Ph.D. degree in neuroscience from Brown University, Providence, RI, in 1979. He is Henry Merritt Wriston Professor in and Chairman of the Department of Neuroscience at Brown University. He is a

cofounder and Executive Director of Brown University's interdisciplinary Brain Science program, which includes physical, biological, and medical scientists, engineers, and mathematicians. His personal research program is aimed at understanding neural computations used to turn thought into movements and to produce brain machine interfaces that restore lost neurological functions. He is co-Founder and chief scientific officer of Cyberkinetics, Inc., a biotechnology startup that is developing neural prosthetics devices. Prof. Donoghue has received a Basil O'Connor Fellowship (March of Dimes) and a Javits Award (NIH).

10 OPTOFLUIDIC MICROSCOPE – FITTING A MICROSCOPE ONTO A SENSOR CHIP

Changhuei Yang[1,3]*, Xin Heng[1], Xiquan Cui[1] and Demetri Psaltis[1,2]

[1] Department of Electrical Engineering, California Institute of Technology
[2] School of Engineering, Ecole Polytechnique Federale de Lausanne
[3] Department of Bioengineering, California Institute of Technology
*chyang@caltech.edu

10.1 INTRODUCTION

The invention of optical microscopy in the 16th - 17th century enabled the direct viewing of cells and other microscopic objects for the first time. Since then, microscopes have become ubiquitous and vital laboratory tools. While different microscope variants have been developed which provide additional functionalities, the basic microscope design has undergone very few fundamental changes over the years.

In general, a microscope must fulfill three functions. Historically, a microscope should be able to replicate an image of the object onto a person's retina. With the development of CMOS (complementary metal oxide semiconductor) and CCD (charge coupled device) detector grids, they can be and have been substituted in place of the retina. Nevertheless, the vast majority of microscopes still operate by relaying and replicating an object's image onto a sensor grid of some sort. Second, a microscope should be able to magnify the image so that microscopic objects are adequately resolved. 20/20 vision roughly corresponds to a pixel size of about 5 μm on the retina (this rough comparison is simply meant to give a sense of scale). To resolve sub-micron feature, magnification is invariably required. Third, a microscope should be able to select a specific plane in the object for imaging – optical sectioning. These functions invariably entail the use of expensive

and sophisticated optical components. Worse, optical relaying and magnification requires the use of substantial propagating space – as such, microscopes tend to be bulky.

In the past couple of decades, the advances and great strides in lab-on-chip platforms for chemical and bioscience applications fueled the need for a small, on-chip implementation of a high resolution optical microscope. The reason is simple – researchers have developed a comprehensive suite of miniaturized tools and systems, ranging from fluorescence-activated cell sorters [1], microfluidic mass spectrometry analysis systems [2-4], immunoassays systems [5], proteolytic digestion systems [6] to DNA separation, analysis and sequencing devices [7-9], yet, imaging requirements in lab-on-a-chip systems are still fulfilled by using large and expensive conventional microscopes. This obviates the size and cost advantages of lab-on-a-chip systems.

In addition to being an important component in a lab-on-a-chip suite of tools, a cheap, compact and high resolution on-chip optical microscope can dramatically change the way bioscience and clinical work are conducted. For example, a compact and cheap microscope can be used in white blood counting cytometry devices as image based cell type discriminators. In comparison to the cell type discrimination approach which relies on the sample's scattering properties, this should lead to more accurate cell typing. Further, the cheapness and compactness of such units will be a boon for clinicians as they can use such systems as disposable point-of-care units. Healthcare in developing countries can also benefit from a cheap and compact microscope that health workers can fit into their pockets with ease. For example, they can use such a microscope in place of an expensive and bulky conventional microscope for malaria diagnosis (microscope examination of blood sample remains the definitive diagnostic approach for determining malaria infection). A cheap and compact on-chip microscope can change the way a bio-scientist design his/her experiments. If a microscope can be made cheaply with existing semiconductor and microfluidic fabrication technologies, they can envision building tens or even hundreds of microscopes on a single chip. Such a device can be used to parallelize the imaging of a large number of samples and dramatically improve throughput. Taking a long view, an on-chip microscope can even form the imaging component of an implantable device that may be implanted into a person to provide continual monitoring of objects in the blood stream. Such a device may be useful for screening circulating tumor cells [10] and other abnormal objects to provide early warnings of developing diseases.

The Optofluidic Microscope (OFM) that we have recently implemented at Caltech is, to our knowledge, the first high resolution compact microscope that is implemented on a chip. We believe that this microscopy method can address a significant range of imaging needs in lab-on-a-chip systems described above and can change the way bio-scientists and clinicians use microscopes. The OFM technique is remarkable in several aspects. First, it does not require the use of any bulk optical elements, including lenses. This simplifies the fabrication process and enables compact implementation in a lab-on-a-chip system. Second, the method has the potential of reaching resolution that is better than that of a conventional microscope. Third, the method is simple and easily allows for the incorporation of additional functionalities, such as fluorescence detection capability and phase detection.

This chapter is divided into 7 sections. In Section 10.2, we will introduce the OFM's operating principle. In Section 10.3, we will summarize the experimental implementation and evaluation of the first OFM prototypes. In Section 10.4, we will discuss the relationship of image resolution, aperture size and sample height for the OFM. In Section 10.5, we will discuss the tradeoff between resolution and sensitivity. In Section 10.6, we will highlight some of the interesting OFM variations that provide additional image functionalities. Section 10.7 concludes this chapter by summarizing the key results.

10.2 OPERATING PRINCIPLE

The OFM imaging method can be best understood by taking a closer look at the first two aforementioned functions of a conventional microscope: image replication and magnification. We can do away with the need for image replication if we are willing to rethink the way we use a CCD or CMOS sensor grid. Unlike the human retina, we can place the object of interest directly onto such a sensor grid. By illuminating the object uniformly, an image of the object can be directly recorded by the sensor grid. The absence of optical elements in this arrangement implies that there is no aberration or collection numerical aperture to worry about. This imaging strategy mirrors the phenomenon of floaters in our eyes and has been implemented as an imaging method recently [11, 12].

This imaging method is non-magnifying and its resolution is fundamentally limited by the pixel size of the sensor grid. As commercial CCD or CMOS sensors have pixels that range upwards of 5 μm, such an imaging scheme cannot be expected to perform better than a conventional micro-

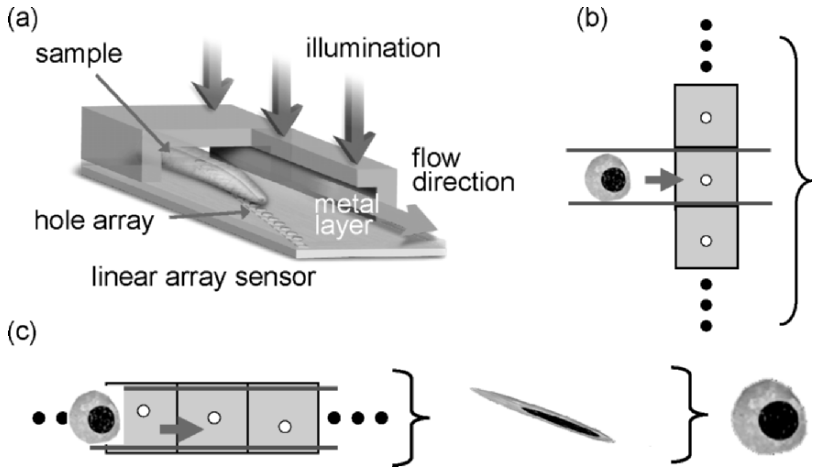


Figure 10.1 (a) OFM imaging scheme. (b) A transmission image of the object can be obtained by simply stacking up the transmission time traces collected as the object passes over the holes. In this figure, the size of the sample is approximately the size of a sensor pixel, as such, it passes over only a single hole. To achieve high resolution imaging, we need to space the holes closely across the channel. This particular arrangement give poor image resolution as we cannot space the holes closer than a sensor pixel width without mapping multiple holes onto a sensor pixel. (c) The OFM skews the line of holes along the channel. This way, we can space the holes as closely across the channel as we want, while at the same time ensure that there is a unique 1-to-1 mapping of holes to sensor pixels. As the object will pass over each hole at different times, we need to correct for the time delays in the transmission traces prior to image construction. This is easily done by unskewing the traces based on the flow velocity of the sample.

scope which has a resolution range of ~1 μm to 0.2 μm. While we can reasonably expect that sensors with ever smaller pixels will become available in the future, there are significant technological barriers to be breached. For example, with ever smaller pixels, the amount of light leakage from light incident on one pixel to adjacent pixels can become very significant.

The nascent field of optofluidics – the fusion of microfluidics and optical technologies [13], offers us a way to improve upon this imaging approach. The OFM enables the imaging of fluid immersible objects with microscope level resolution [14] by exploiting flow in its imaging strategy.

An OFM device may be fabricated as follows. First, a layer of metal is coated onto a linear sensor array to block out light. A line of apertures is then punched into the metal layer. Finally, a microfluidic channel is added on top of the entire chip.

In operation, the device is uniformly illuminated from the top and the target object is flowed across the aperture array [see Fig. 10.1(a)]. The time varying light transmission through the aperture constitutes a transmission image line trace across the object. By stacking the lines traces from all the apertures together, we will be able to construct a transmission image of the object. In this situation, the resolution is fundamentally limited by the aperture size and not the size of the underlying pixel. The exact arrangement of the aperture array with respect to the underlying pixel grid is crucial. To achieve high resolution imaging, we will need to space the apertures closely across the channel. An arrangement where we simply arrange the aperture array across the channel (the orientation of the aperture array is normal to the channel length) [see Fig. 10.1(b)] is sub-optimal as we will not be able to space the apertures closer than a sensor pixel width without mapping multiple apertures onto a sensor pixel. The apertures will under-sample the object. If instead we skew the line of apertures along the channel as shown in Fig. 10.1(c), we can then space the apertures as closely across the channel as we want, while at the same time ensure that there is a unique 1-to-1 mapping of apertures to sensor pixels. Every part of the object will be sampled in this arrangement.

As the object will pass over each aperture at different times, we need to correct for the time delays in the transmission traces prior to image construction. This can be done as long as we know the flow speed of the object. The OFM imaging strategy does require two conditions to function properly. First, the target should not change orientation or change shape during the imaging process. Second, the flow velocity should be uniform during the imaging process.

In the next section, we will discuss the implementation of the OFM in detail.

10.3 IMPLEMENTATION

10.3.1 Experimental Setup

Our initial implementation of the OFM is fairly straightforward. The prototype device consists of an opaque metal film with an etched array of submicron apertures and a PDMS microfluidic chip that is bonded onto the metal film. In our experiments, the fabrication of the aperture array and the microfluidic chip was carried out in two separate steps. To fabricate the aperture

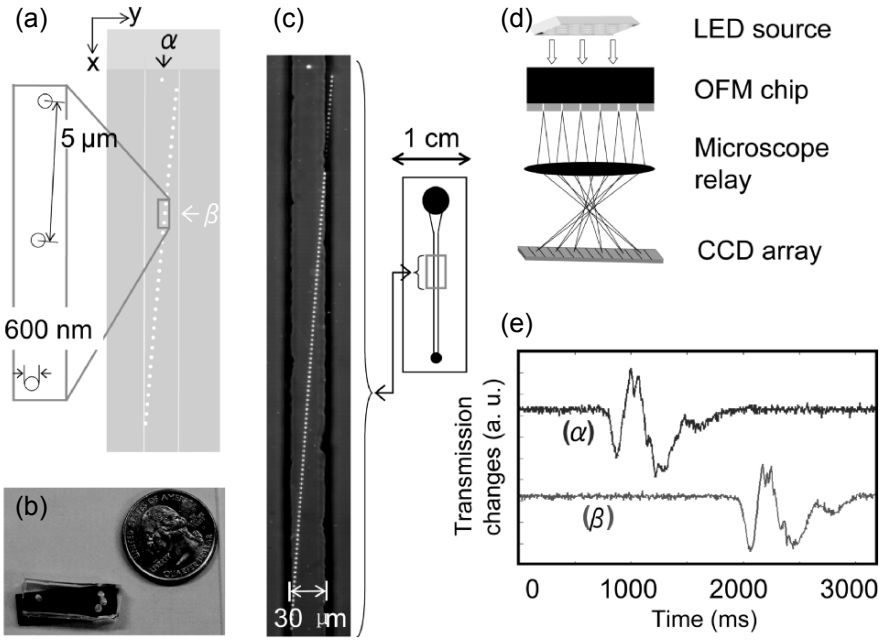


Figure 10.2 (a) Aperture array orientation and the relevant dimensions in the prototype. (b) Prototype in relation to a quarter. (c) Microscope image of the prototype. (d) Imaging arrangement for the prototype. (e) Line scans collected through hole α and β.

array, 90 nm thick aluminum was first evaporated onto a quartz wafer. Then the pattern of the aperture array (diameter D = 600 nm; spacing = 5 μm) was defined on PMMA resist by e-beam lithography (JEOL 9300) and subsequently transferred into the aluminum layer by reactive ion etching. The microfluidic structure was fabricated on Microchem SU-8 resist using photolithography (Karl Suss MA6) and then transferred onto PDMS elastomer. The microfluidic chip was attached onto the metallic layer and their accurate alignment was accomplished with a mask aligner. Note that the aperture array needs to be oriented at a small angle relative to the micro-channel [see Figs. 10.2(a) and (c)]. In our prototype, the microfluidic channel was 30 μm wide and 15 μm tall. A uniform light source (Lamina BL3000 white LED) at normal incidence to the device provided the illumination (~0.1 W/cm^2). The device can be fabricated directly onto a CCD or CMOS array to create a truly compact microscope system [see Fig. 10.1(a)]. In this case, the aperture spacing should be equal to array's pitch size so that the transmission through each aperture maps uniquely onto a single sensor pixel. Fabrication of the OFM directly onto a sensor array is straightforward [11, 12] and can be efficiently performed commercially.

For the purpose of demonstrating feasibility of the OFM method, our prototype was not fabricated onto a sensor array directly. Instead, the transmitted light through the aperture array were relayed onto a CCD array (Princeton Instruments, Spec10-100, pitch: 20 μm × 20 μm) via a microscope (Olympus IX-51) [see Fig. 10.2(d)]. Note that the microscope served as a relay system and played no direct role in imaging. This arrangement allowed us to more easily evaluate the performance of the OFM prototype.

Imaging with an OFM device is accomplished by flowing the target across the aperture array. During the transit, the transmission signal through each aperture will vary and the acquired time trace will constitute a line scan across the object. In our prototype, the aperture spacing in y-direction (perpendicular to the sample flow) is chosen to be 300 nm (half the aperture size) so that the line scans span the entire target without undersampling. Therefore, the resolution limit along y-direction is given by the aperture size. The resolution along x-direction (along the flow direction) is given by the larger of the following two parameters: the aperture size or the product of the sample velocity and the pixel acquisition time.

As the target passes over each aperture at a different time, the compilation of the line scans will form a distorted transmission image. If the flow velocity V of the target is known, the line scans can be un-skewed appropriately and an undistorted image can be created. In our implementation, this parameter is determined through the use of an isolated aperture α at one end of the array [see Figs. 10.2(a) and (c)] and a corresponding aperture β in the aperture array that has the same y coordinate as α. As the two apertures share the same y coordinate, they should scan the same line across the target. A flowing target passes over aperture α prior to aperture β; V can be calculated by dividing the distance between the two apertures by the time difference between when the target first passes over each aperture. This aperture pair serves an additional function – tumble monitor for the target. If the target rotates or tumbles during its passage over the aperture array, the two line scans will appear dramatically different. This enables us to effectively screen out rotating targets. Figure 10.2(e) shows a pair of line scans acquired through the aperture pair during the passage of an unrotated target.

10.3.2 Imaging *C. Elegans*

To demonstrate the OFM's imaging capability, we imaged *Caenorhabditis elegans* (*C. elegans*) at the first larval stage. The sample was prepared by first euthanizing the *C. elegans* in a 70ºC heat bath for 3 minutes, and then

mixing them with 0.1% bovine serum albumin (BSA) solution. The BSA solution reduces adhesion of *C. elegans* with the channel walls. Pressure driven flow was used to drive the targets through the channel. The nematode concentration in the microfluidic reservoir was approximately 5 *C. elegans* per nL. The average transport speed was 300 μm/s. Given that the CCD pixel acquisition time is 2.5 msec, this velocity gives an average pixel size of about 750 nm along the x-direction. The maximum achieved image acquisition rate was 40 worms per minute. On average, approximately 45% of the acquired images were rejected due to sample rotation (detected through the aperture pair α, β) and nematode aggregation. Figures 10.3(a-c) show several OFM images of the nematodes, and Fig. 10.3(d) shows the image of a *C. elegans* acquired with an inverted microscope (Olympus IX-51, 40× non-immersion objective) under similar illumination condition. The nematode pharynges are discernible in the OFM images, providing a clear indication of OFM's ability to image targets with resolution comparable to a conventional microscope.

The OFM can be treated as a set of parallel near-field scanning optical microscopes (NSOM) [15], operating in the collection mode. The resolu-

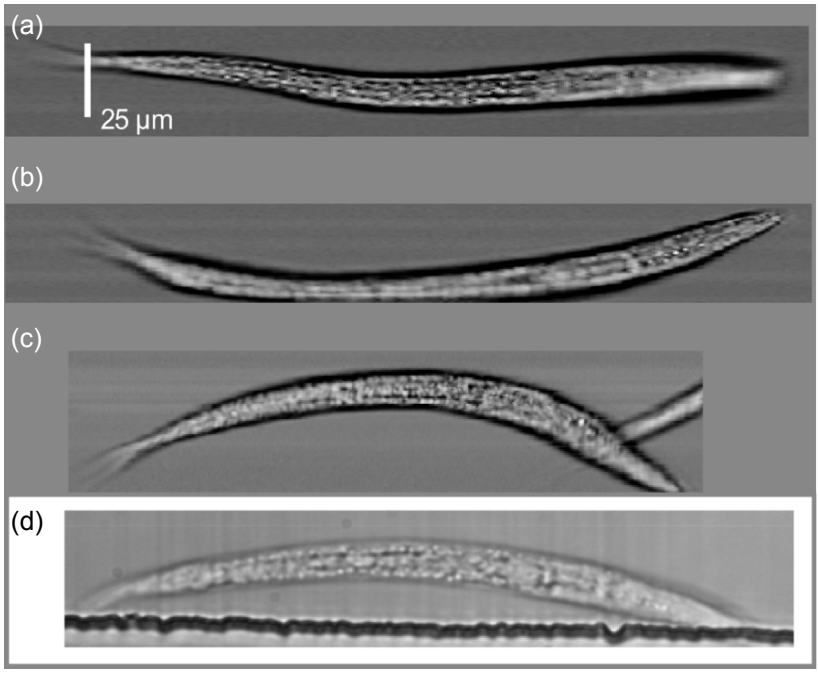


Figure 10.3 (a)-(c) OFM images of *C. elegans*. (d) Conventional microscope image of a *C. elegans*. The rough edge of the microfluidic channel is visible here.

tion limit of the OFM can be established by measuring the collection mode point spread function (CPSF) of the aperture. The CPSF is defined as the variation in the transmission through the aperture as a point source is laterally scanned across it. In other words, the vertical axis of a CPSF plot is the transmission and the horizontal axis is the lateral displacement of the point source from the aperture's center [e.g., Fig. 10.4(c)]. The resolution of the OFM for a target at any given plane above the aperture array is given by the width of the CPSF for that plane. As the CPSF broadens with increasing height from the aperture array, the resolving power of the OFM, likewise, deteriorates with increasing height. The exact relationship between resolution and height is discussed in detail in the next section. The resolution limit of a given OFM is equal to the width of the CPSF at zero height from the aperture array. Practically, this resolution is only attained for parts of the targets that are in the near-field zone of the aperture array. Figure 10.4(a) illustrates this relationship between resolution

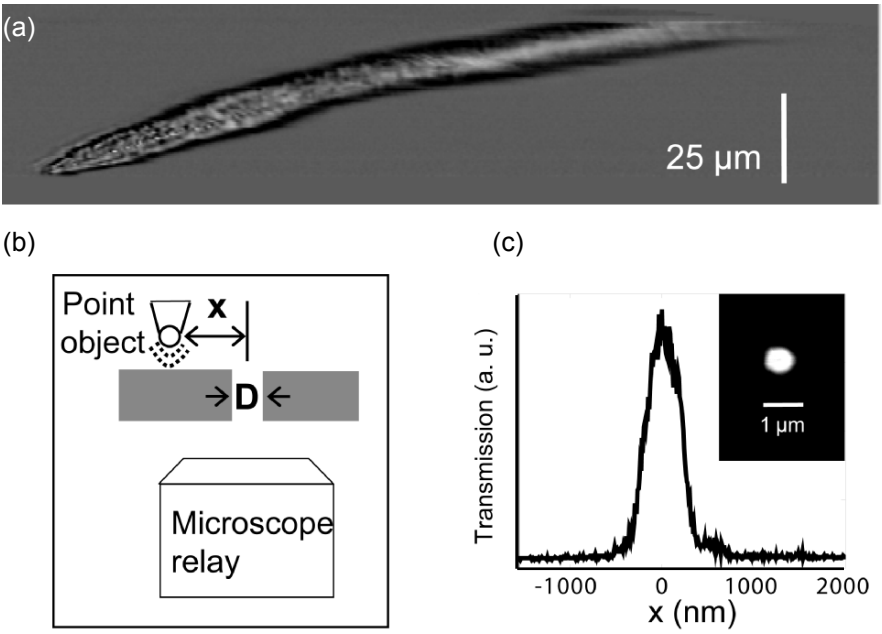


Figure 10.4 (a) OFM image of a *C. elegans*. The segment of the worm on the left hand side was close to the aperture array and was therefore imaged with good resolution. The rest of the worm was distant from the hole array and was therefore imaged with poor resolution. (b) We determine the resolution limit of our prototype by walking an NSOM tip over the hole and monitoring the transmission. (c) The result of the experiment.

and sample separation from the aperture array. The segment of the worm that was close to the aperture array was imaged with good resolution while the segment that was distant was poorly imaged.

We experimentally measured the near-field CPSF of the apertures used in our prototype with a transmission mode NSOM (Alpha-SNOM, WITec GmbH). The experimental setup is illustrated in Fig. 10.4(b) where the NSOM tip was taken as a pseudo point object. The illumination was provided by a red laser diode ($\lambda = 650$ nm, 30 mW). Given that the NSOM tip is 80 nm wide and the aperture is 600 nm in diameter, the approximation of the NSOM tip as a point source is reasonable. Figure 10.4(c) shows the measured CPSF profile of the OFM prototype. Based on measurements of 8 different apertures, the resolution limit of our OFM prototype is calculated to be 490 ± 40 nm (Sparrow's criterion [16]). However, to reach this resolution during the imaging process, the target will have to be in contact with the aperture array plane and the pixel acquisition time will need to be sufficiently short.

10.4 RESOLUTION

10.4.1 Putting Resolution in Context

Sub-wavelength apertures are known to possess interesting characteristics. The unique transmission and diffraction properties of a subwavelength circular aperture [17], a single nanoaperture with designed shapes [18, 19], a single nanoaperture surrounded by corrugated surface [20] or a tightly-spaced nanoaperture array [21, 22] have all been studied and reported in the recent past. While the transmission properties of subwavelength apertures are yet to be fully understood from the theoretical standpoint, subwavelength apertures have been used to demonstrate some remarkable improvements in areas such as bio-molecule analysis [23, 24], sensing [25], recording & storage [26], laser technology [27], as well as photolithography [28].

A number of optical imaging and sensing devices based on a single subwavelength aperture or a subwavelength aperture array [29, 30] have recently been developed. These devices have the distinct advantage of enabling compact and portable analysis systems. In addition, optical imaging devices based on subwavelength apertures are particularly important in biological applications as they have the potential to deliver resolution that is beyond

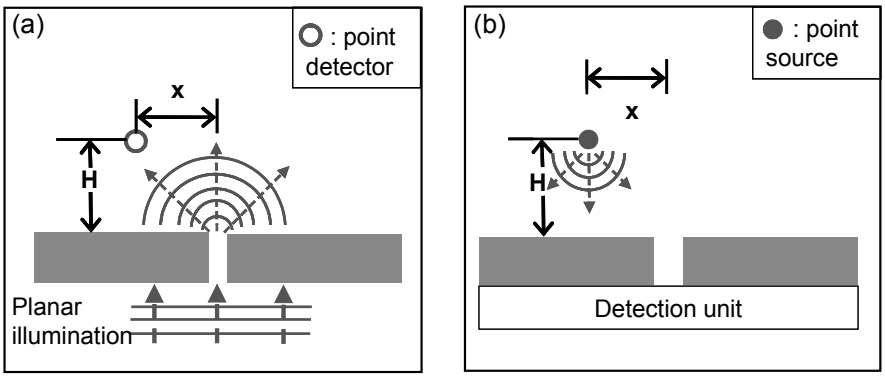


Figure 10.5 Two schemes of subwavelength aperture based imaging devices (ABIDs): (a) type-I, (b) type-II. H: vertical separation of point from aperture. x: the lateral displacement of point from the aperture.

the diffraction limit; the resolution of such devices is fundamentally limited by the aperture's size.

In general, there are two categories of subwavelength aperture based imaging devices (ABID): in one, the apertures are the illumination sources [type-I ABID, Fig. 10.5(a)]; in the other, the apertures are the collection units [type-II ABID, Fig. 10.5(b)]. The device developed by Tegenfeldt, *et al*. [29] is a type I device. In that device, a nanoslit trio was used to map the length of extended DNA molecules within a nanofluidic chip. The nanoslits functioned as near field illumination sources, and a resolution of 200 nm was reported when the device was used to image fluorescent beads. The OFM is a type-II ABID device.

The optical geometry for type-I ABIDs is straightforward. A planar illumination onto the aperture results in a specific transmission and diffraction pattern of the near and far field components on the exit side of the aperture [see Fig. 10.5(a)]. The associated diffraction pattern is directly related to the resolution of type-I ABIDs. To our knowledge, almost all of the publications to date that have investigated light transmission through apertures [17-22], employed this geometry or other closely related geometries.

Unfortunately, this geometry is not suitable for characterizing type-II ABIDs. The performance of a type-II ABID should be studied with a point source that is located at varying height (H) and lateral displacement (x) from the aperture [see Fig. 10.5(b)]. While the optical geometries of type-I and type-II ABIDs appear to be complements of each other by optical reversibility, in reality, they can behave quite differently. The two would indeed be optical complements under the condition that the detection unit in Fig. 10.5(b) is sensitive only to the plane wave component of the transmit-

ted light that exits the aperture in the normal incidence direction [an exact complement to the incident plane wave in Fig. 10.5(a)]. However, a realistic detection unit is sensitive to the total power flux and is indiscriminate of the light field wavefront; this breaks the required symmetry.

The primary objective of this section is to theoretically model and experimentally measure the transmission characteristics of a point source at varying height and lateral displacement from a subwavelength aperture [Fig. 10.5(b)]. While our focus is to apply our findings to design better OFMs, this geometry may also be of relevance to research areas, such as near field Raman scattering [31] and near field radiation force [32]. In Subsection 10.4.2, we will report on our experimental measurements of the collection mode point spread functions (CPSFs) (formally defined in Subsection 10.4.2) for different aperture sizes. In Subsection 10.4.3, we will describe the 3D electromagnetic simulation method that we used to model this optical geometry. In Subsection 10.4.4, we will discuss the validation of our simulation method with our experimental data. In the Results and Discussions (Subsection 10.4.5), we will use our experimental and simulation results to address the following issues:

1. The resolutions of type-II ABIDs and the far field trend. We will also compare type-II ABIDs and type-I ABIDs.
2. The dependency of the CPSF of the aperture on the numerical aperture (NA) of the collection optics.
3. The relationship between aperture size and the CPSF when the point source is close to the aperture, as well as the theoretical resolution limit of a type-II ABID imposed by the aperture size.

10.4.2 Experimental Method

The geometry shown in Fig. 10.5(b) can be used for the characterization of the transmission through an aperture from a point object. For clarity, we define the collection mode point spread function (CPSF) as the function of the transmission versus the lateral displacement of an isotropic point source that is scanned over the aperture at a fixed height. This CPSF is analogous to the point spread function (PSF) generally defined for a near field scanning optical microscope (NSOM) [33, 34].

The resolving power of a type-II ABID can be related to the CPSF as follows. Consider two non-interfering point sources of equal strength at the same height above the aperture. The resolution of the type-II ABID is equal to the minimum lateral separation between these two point objects such

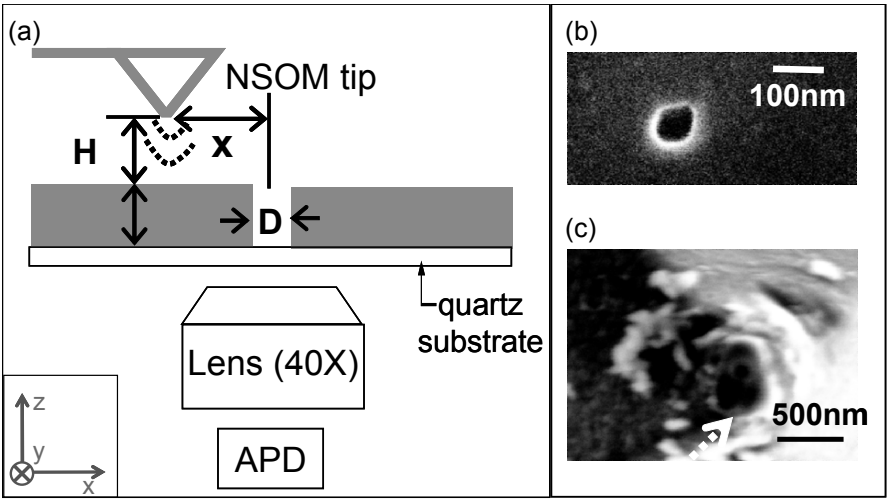


Figure 10.6 (a) Illustration of the experimental scheme. The NSOM tip scans over the subwavelength aperture (diameter: D) milled in a thin aluminum film (thickness: t) at a constant height (H). Substrate is an ultra-clean quartz wafer. x: the lateral distance between the NSOM tip and the aperture center. (b) SEM image of a FIB milled aperture; D = 100 nm. (c) SEM image of a NSOM tip with a diameter of ~100 nm.

that the peaks of their CPSFs are sufficiently separated to be distinguishable. One criterion – the Sparrow's limit [16], defines the resolution as the separation between the two point sources to be such that the composite point spread function has a zero null second derivative in the center. This criterion does not assume a particular PSF profile and forms a good basis for an unbiased comparison of resolution between systems with different PSF profiles (for example, a conventional microscope's PSF is described by a Bessel function [35, 36], while a type-II ABID's CPSF is not). We will use this criterion in our comparison of type-II ABIDs with conventional microscopes in Subsection 10.5.3. To keep our measurements and analysis general, we choose to quantify the CPSFs in the following subsections by their full width at half maximum (FWHM) – a quantity that is more widely used for characterizing PSFs as they are easily measured and conceptually simple to understand.

In principle, the experiment should be performed with an ideal point source. In the experiment, an NSOM tip (WITec GmbH, $\lambda_0 = 650$ nm) with a 100 nm characteristic diameter [Fig. 10.6(c)] was used as the light source. We approximated the NSOM tip as a point source. The impact of the finite tip size and its non-isotropic radiation pattern will be discussed in Subsection 10.4.4. We scanned the NSOM tip at various heights (H) and

measured the transmission through the aperture. We characterized the CPSF based on the collected data.

The subwavelength apertures were milled in an aluminum film coated on quartz wafer by using high resolution focused ion beam milling (FEI nova 200 dual-beam FIB). Aluminum was chosen in our experiment, as it is a highly optically absorptive metal. The refractive index of aluminum is given by $n = 1.24 + 6.6 \times i$ at $\lambda_0 = 650$ nm [37, 38], which implies an amplitude-associated skin depth of 15.7 nm. The thickness of the aluminum layer was 60 nm or ~4 times the skin depth. The aperture diameter ranged from 100 nm to 1000 nm; the aperture size was measured with a scanning electron microscope (SEM) (Hitachi S4100). The SEM image of an aperture with a diameter of 100 nm is shown in Fig. 10.6(b). For each diameter value, multiple apertures were perforated on the metal film with sufficient inter-aperture spacing (more than 20 μm), so that the surface plasmon induced anomalous transmission of the sub-wavelength apertures did not occur.

Our NSOM system was a commercial near field microscope (Alpha-SNOM, WITec GmbH.). In our experiments, as illustrated in Fig. 10.6(a), the NSOM tip scanned over the aperture (diameter: D), at a constant height (H) above the metal surface. The illumination through the NSOM tip was provided by a red laser diode ($\lambda_0 = 650$ nm, 30 mW). The total transmission through the aperture was collected by a 40× microscope objective and detected with an avalanche photodiode (AQ series, Perkin Elmer Optoelectronics).

As this specific NSOM tool only operates in the contact mode, we used the following method to vary the scan height of the probe tip from the aperture. Thin films of silicon dioxide (refractive index = 1.5) of specific thickness were coated on top of the aperture by plasmon enhanced chemical vapor deposition (STS PECVD system) successively. The NSOM tip was scanned across the aperture to collect the CPSF profile between each coating. The accumulative thickness of silicon dioxide was accurately measured by a film thickness measurement tool (SCI FilmTek 2000) after each coating step. This process was repeated until a height of 1.15 μm was reached. The apertures were planarized with diluted SU-8 resin; this prevented the NSOM tip from dipping into the apertures during scanning. We note that refractive index of the silicon dioxide layers is accounted for in the simulations described in Subsections 10.4.3 and 10.4.4.

Figure 10.7 shows an example of our collected NSOM data; the size of aperture in this example is 300 nm. From Fig. 10.7(a2 → c2) and (a3 → c3), it can be clearly seen that the measured CPSF of the aperture widens as H becomes larger.

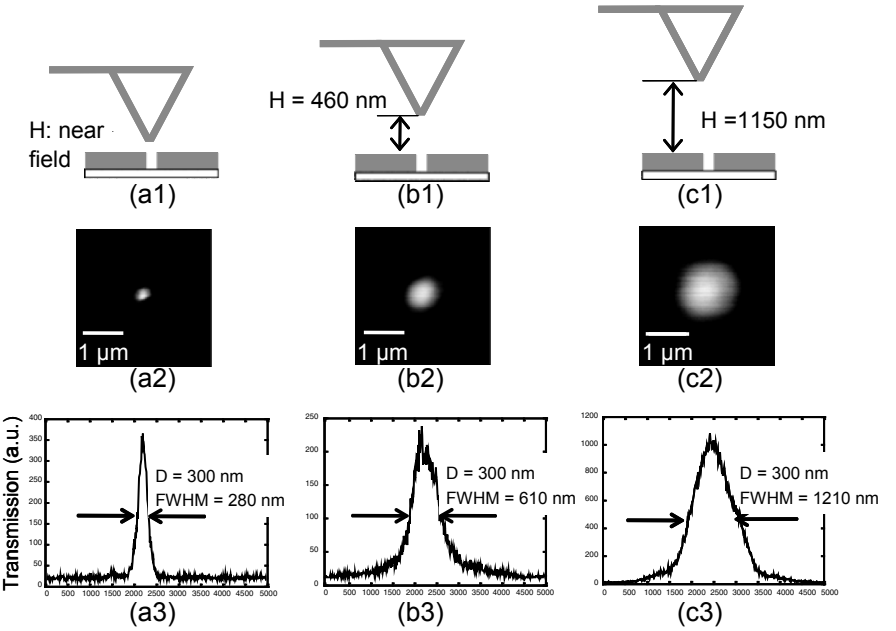


Figure 10.7 NSOM measurements of a subwavelength aperture (D = 300 nm). (a1) Illustration of the experimental scheme with NSOM tip in near field; (a2) the NSOM image under condition (a1); (a3) the CPSF curve extracted from the NSOM image (a2). (b1) The experimental scheme with H = 460 nm; (b2) the NSOM image under condition (b2); (b3) the CPSF curve extracted from (b2). (c1) The experimental scheme with H = 1150 nm; (c2) the NSOM image under condition (c1); (c3) the CPSF curve extracted from (c2).

The described experiment departs from an ideal point-source experiment in three ways. First, as mentioned above, the NSOM tip was approximated as a point source. The finite NSOM tip size can be expected to cause experimental inaccuracy when used to quantify the CPSFs of small apertures. Second, the radiation dipole components of the tip are largely confined to the x-y plane [39], which is a deviation from an ideal point source where the radiation pattern is isotropic. Finally, the transmission through the aperture is collected with a finite NA objective in the experiment. Ideally, we will like to detect the entire transmitted component. These deviations from an ideal experiment are not the result of flawed experimental design, but are instead due to the practical limitations of existing NSOM technology.

10.4.3 Simulation Method

Besides the NSOM walking experiment described in Subsection 10.4.2, we also applied an electromagnetic (EM) simulation scheme to study the same point source scanning geometry. In this study, the commercial software, COMSOL Multiphysics 3.2 [40], was used. The time averaged power flow (S), i.e. Poynting vector, was chosen to quantify the total transmission of the aperture.

Hertzian dipole [41] (i.e. a short line source) of length 10 nm (much shorter than the optical wavelength) was adopted in our simulations. In each simulation run, the dipole was placed at a given height (H) and lateral displacement (x) from the aperture (Fig. 10.8). For each source location, the orientation of the dipole was alternated through three orthogonal orientations – along x-, y- and z-axis in three simulation runs. The combination of the resulting Poynting vector fields for the three runs enabled us to create an isotropic radiation field emanating from that source location – an effective point source. We calculate the total transmission through the aperture as the summation of the transmissions from all three dipole orientations. The parameters of the simulations were chosen to match with those of the experiments. Figures 10.9(a) and (c) and Figs. 10.9(b) and (d) show examples of the simulation results with the effective point source at different lateral

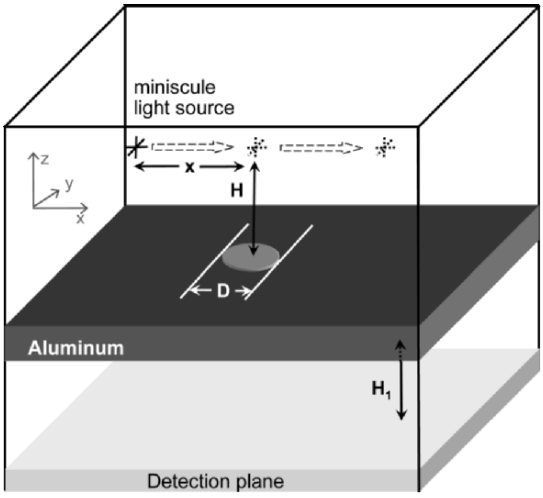


Figure 10.8 Schematic of the simulation geometry. The arrows indicate the movement of the light source in the simulation space. The detection plane is 0.2 μm underneath the subwavelength aperture. $H_1 = 0.2$ μm (distance between metal's bottom surface and the detector plane). D, H and x are defined the same way as Fig. 10.6.

displacement (H = 50 nm, D = 300 nm). As are apparent in the figures, the transmission through the aperture can be significantly altered by varying the lateral position of the light source.

A CPSF can thus be generated by moving the light source across the aperture in the simulation and plotting the transmission as a function of the source's lateral displacement. As illustrated in Fig. 10.8, the total transmission power is recorded by a virtual detector located 0.2 μm below the metal surface. Figures 10.9(e) to (g) clearly illustrate the widening trend of the CPSFs with increasing H.

The choice of numerical boundaries deserves some discussions. In this software, the outer boundaries that enclose the entire simulation space incur spurious numerical reflections, thus adding errors to simulation results. The low-order scattering boundary conditions provided by the software were not sufficient to suppress the reflection errors. Among all the possible means of solving this problem, perfectly matched layers (PMLs) have been accepted as one of the most effective. As such, we incorporated this scheme into our

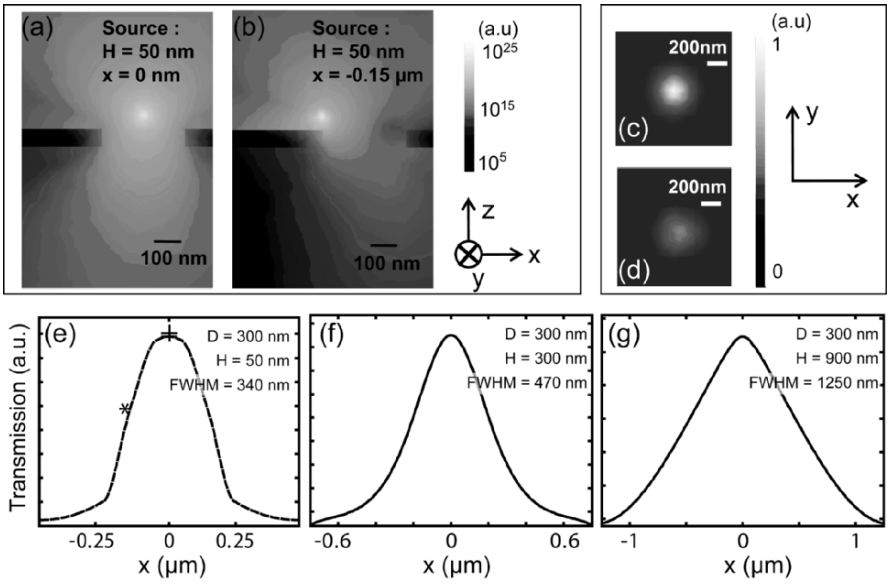


Figure 10.9 Example of the simulation results with all three polarizations considered. (a) Cross sectional plot of power flow (|S|), time averaged; H = 50 nm, D = 300 nm, x = 0. (c) Plot of |S| on the detector plane. (b) and (d) are similar to (a) and (c) except that x = –0.15 μm. (e)-(g) Simulation generated CPSF plots for source at different heights from the aperture. The cross and the asterisk indicate the transmissions corresponding to the point source geometries in (a) and (b), respectively.

simulation. However, we note that the numerical reflection on the interface between the constant-value PMLs and the real medium (i.e. SiO_2 in our case), though greatly reduced, still exists because of the finite mesh discretization at this interface [42, 43]. Therefore, a method of variant-PMLs was adopted to truncate the computation domains, with the coordinate stretching factor (a_z) defined as [43]:

$$a_z = 1 - j\left[(z - z_0)/L\right]^m \cdot \delta \tag{10.1}$$

where z_0 is the z-coordinate of the interface between PML and the glass medium and L is the thickness of PML; m and δ define the absorption property of the PMLs. In x and y directions, a_x and a_y can be similarly defined. In our study, we chose $L = 0.4$ μm, $m = 2$, $\delta = 1$, which are consistent with the PML optimization designs of other computational physicists [42-45]. After defining PMLs, we chose an iterative solver, GMRES (generalized minimum residual method) with a geometric multigrid preconditioner to solve the 3D computation problem. The mesh step size was no larger than 100 nm, which resulted in a rapid convergence during the process of iterative computation.

10.4.4 Comparison between Simulation and Experimental Results

As mentioned in Subsection 10.4.2, the experimental limitations are expected to cause systematic errors in the experimentally measured CPSFs. On the other hand, the simulation model requires cross-verifications. In this subsection, we first compare the experimental and simulation results by adjusting the simulation parameters to match with the experimental conditions. Specifically, we adapt the light source in the simulations to more closely resemble the actual NSOM tip. A comparison of the experimental data and the simulation result serves to validate the simulation model. Finally, we run the simulations based on ideal experimental conditions and compute the CPSFs for a range of aperture sizes and varying point source heights.

We experimentally measured the CPSFs for four aperture sizes: 100 nm, 150 nm, 300 nm, 450 nm and 550 nm by using the experimental procedure described in Subsection 10.4.2. The dielectric medium refractive index was 1.5. In our experiments, H ranged from a near field value (≤5 nm) to 1.15 μm. The FWHMs of the measured CPSFs are plotted in Fig. 10.10(a) (log-log scale).

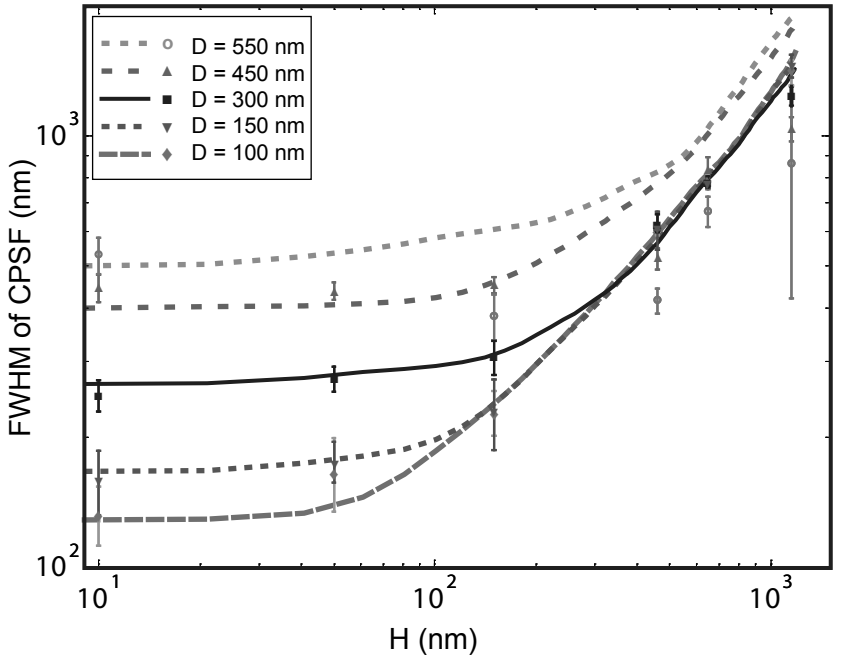


Figure 10.10 (log-log scale) The CPSF's FWHM versus the gap height (H) plots for a range of aperture sizes. The lines represent simulation results and the circles represent experimental data. The simulation model was adapted to match with the NSOM radiation characteristics.

The data is fitted with the theoretical CPSFs profiles created by using the EM simulation scheme described in Subsection 10.4.3 with the following modifications. First, the point source is simplified into a combination of x- and y- dipole components to better match with a NSOM's radiation pattern [39]. Second, we convolved the tip's aperture function (Gaussian with FWHM of 100 nm) with the simulated CPSFs to generate broadened CPSFs to account for the finite size of the NSOM tip. The simulation results are plotted in Fig. 10.10.

The goodness of fit between simulation result and experimental data (see Fig. 10.10) is excellent for most of the data points. However, the experiment and simulation do deviate significantly for the last three points of the data set for the 450 nm diameter aperture and most of the data set for the 550 nm diameter aperture. The difference is especially large for the data set for 550 nm diameter aperture. The measured width of the CPSF for that aperture size actually dipped as a function of H – a counter intuitive behavior.

We believe this discrepancy between simulation and experiment is attributable to the finite collection NA associated with the collection optics in the

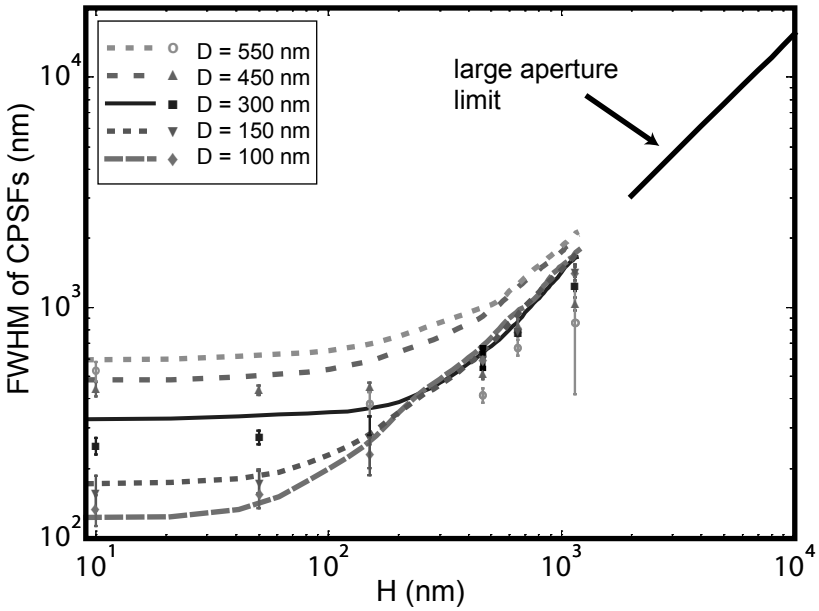


Figure 10.11 (log-log scale) The CPSF's FWHM versus the gap height (H) plots for a range of aperture sizes. The lines represent simulation results and the circles represent experimental data. The simulation models the light source as an effective isotropic point source. Black line: The far field trend of large apertures (FWHM ~1.53 H).

experiment. In comparison, the simulation uses a geometry that provides nearly perfect transmitted light collection. The transmission through a large aperture tends to propagate dominantly along the input light field direction; in comparison, a small aperture diffracts the transmission broadly. Due to the transmission's directionality, a large fraction of the transmitted light will miss the collection optics in the experiment when the incident angle of the incoming light is large, as in the case where the lateral displacement of the light source is large and the light source is sufficiently distant from the aperture. The drop-off in the collected signal for large lateral displacements leads to an effectively narrower CPSF and lower FWHM value. The observed lower experimental measured FWHM value is consistent with this explanation (see 450 nm and 500 nm apertures in Fig. 10.10). We studied the interplay between measured CPSF and the NA of the collection optics more closely in a different set of experiments and an account of our findings is reported in Subsection 10.4.5.

With the simulation model validated, we proceeded to compute the CPSFs for apertures of varying sizes based on our model description in

Subsection 10.4.3. The FWHMs of the CPSFs are plotted in Fig. 10.11. We can see that the widths of the CPSFs are at their minimum when H is small; the respective width matches the apertures' size. This corresponds well with our expectation that the resolution of type-II ABIDs is determined by the aperture size and that the best resolution is achieved when the aperture is close to the target - a well known property of near field optical imaging.

10.4.5 Results and Discussions

10.4.5.1 Resolution of Type-II ABIDs and the Far Field Resolution Trends

The plots in Fig. 10.11 can be used to characterize the resolution in type-II ABID devices. There are a few issues of note. First, as the CPSF are narrowest when H is small, we can conclude that a type-II ABID should provide highest possible resolution when the target sample is at close proximity to the aperture. For clarity in our nomenclature, we shall refer to this quantity as the resolution limit. Second, we can see in Fig. 10.11 that as the height (H) of the point source increases, the CPSFs begin to widen. This corresponds to a degradation of the resolution when the target is at a larger distance from the aperture. This height-dependent increase in the CPSFs' widths can be quantified by the depth of field (DOF), a quantity that corresponds to the axial working range of an imaging system. We define the DOF to be the height at which the width of the CPSF is twice that of the CPSF when the height is zero. It is worth noting that the aperture size D appears to be a good estimate to the DOF. This implies that, as a rough rule of thumb, to achieve good resolution with type-II ABIDs, the plane of interest in the target object should be maintained at a height (H) above the aperture plane that is no larger than the value of the aperture size (D).

The behavior of the CPSFs' width at large H values deserves some discussion. For large H, the light field incident on the aperture simplifies to a plane wave propagating along the direction from the light source to the aperture. This far field simplification allows us to analytically derive the CPSF behaviors for very large apertures. In the large aperture limit, the width of the CPSF will vary linearly with the point source height. The following paragraphs derive the exact relationship.

The transmission areas of very large apertures are the same as their physical areas. Therefore, the total transmission (T) through a large aperture is given by the surface integral of the incident Poynting vector field over the

area of the aperture. In the far field limit, the treatment of the incident light field as a plane wave simplifies the integration to a product of the magnitude of the Poynting vector normal to the aperture ($S_\perp$), and the area of aperture (A):

$$T = S_\perp \cdot A = S(x) \cdot A\cos\theta = S(x) \cdot A\sqrt{\frac{H^2}{x^2+H^2}} \propto \frac{A^2}{x^2+H^2} \cdot \sqrt{\frac{H^2}{x^2+H^2}}, \tag{10.2}$$

where $S_\perp$ is the average energy density above the open area of the aperture; x (H) is the lateral (vertical) distance between the point source and the aperture; θ is the incident angle with $\theta = 0$ representing normal incidence. Note that x has to been considered because it determines the amount of light energy that impinges on the aperture. The profile of CPSF of the aperture is given by the expression in Eq. (10.2), and as such the FWHM of the resulting CPSF is equal to:

$$\text{FWHM} = 2\cdot(2^{2/3} - 1)^{1/2}\cdot H \approx 1.53H. \tag{10.3}$$

It is clear from Fig. 10.11 that all of the CPSFs' FWHM trends taper into a proportionality relationship with the source height (H). More specifically, the relationship is linear in nature. The large aperture limit trend given by Eq. (10.3) is plotted in Fig. 10.11 for comparison. While we expect the simulation results for the larger apertures to show convergence to this trend, it turns out that all of the simulation curves show fair convergence to this trend. This convergence suggests that aperture size has weak impact on the CPSF in the far field limit. At present, we do not have a good explanation for this behavior.

Finally, we will like to point out that the PSF behavior in the far field limit for very large apertures in type-I ABID configuration [see Fig. 10.5(a)] is very different. In this case, an incident planar illumination of wavelength (λ) on the input side of the aperture of size D will result in a well characterized diffraction pattern on the output side. In the far field regime, Fraunhofer diffraction calculation is valid. Based on this calculation, the diffraction pattern is described by Airy's disk [36]. Therefore, FWHM of the diffraction profile (i.e. PSF) is:

$$\text{FWHM} = 1.03\times(\lambda/D)\cdot H. \tag{10.4}$$

Unlike the CPSF for large aperture type-II ABIDs, the width of the PSF for large aperture type-I ABIDs is dependent on the illumination's wavelength, aperture size and the height of the plane of interest. This difference can be attributed to the fact that the two optical geometries are not optical complements for the reason pointed out in Subsection 10.4.1.

10.4.5.2 Impact of the Collection Optics' Numerical Aperture (NA) on the Observed CPSFs

As mentioned earlier, the collection optics' NA can impact on the observed CPSFs. Our experimental geometry enabled us to investigate the influence of the collection optics' NA in more details. This phenomenon can be further appreciated by taking a closer look at the following two scenarios:

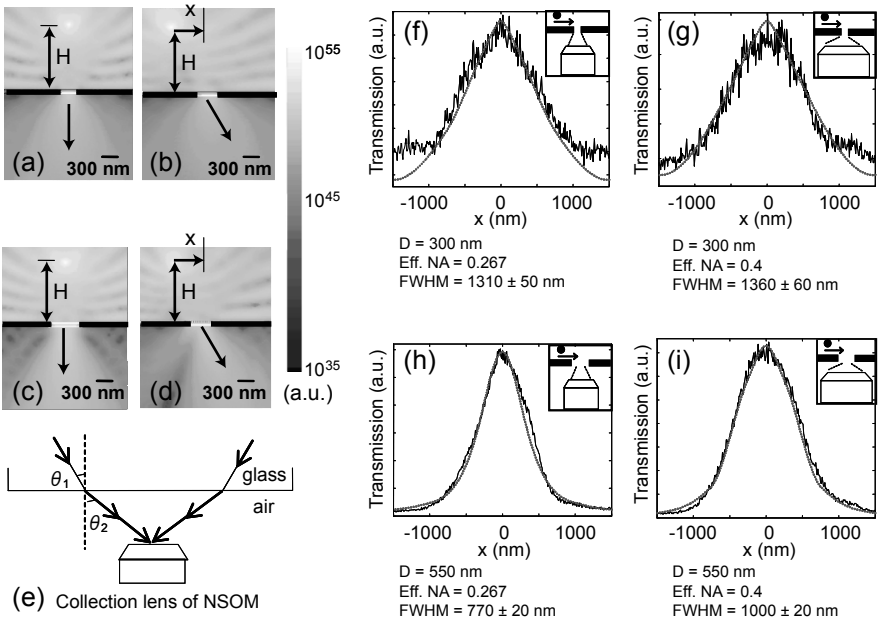


Figure 10.12 (a)-(d) 2D cross sectional plots of power flow, |S| for two different aperture sizes and two different source lateral displacements. The plots clearly show that the directionality of the transmission is strongly influenced by the aperture size. For (a) and (b), aperture size is 300 nm. For (c) and (d), aperture size is 550 nm. The source lateral displacement is 0 nm in (a) and (c), and –500 nm in (b) and (d). (e) Light collection geometry in the experiment showing that the effective numerical aperture is reduced by the presence of the quartz wafer. (f)-(h) Experimental data and corresponding simulation results of the CPSFs for two different aperture sizes. The height was kept at 1150 nm.

10.4.5.2.1 Small Apertures

First, consider a sufficiently small aperture such that it functions as a zero mode waveguide [17, 23]. In this case, the transmitted field pattern is simply the diffraction pattern associated with the aperture. The input light field pattern incident on the aperture has no effect on the transmission pattern. Barring non-linear effects, the input light field strength can only collectively increase the transmission, but will not be able to alter the output diffraction pattern. The independency of the output diffraction pattern on the input light field pattern extends approximately to apertures that are small enough to support only a small number of propagating modes. Figures 10.12(a) and (b) show the results of our simulations of the transmission through a 300 nm wide aperture for a point source at height H = 1.15 μm and two different lateral displacements (x = 0 nm and x = 500 nm). It can be seen from our simulation that the strength of the transmission changes but the output diffraction pattern remains almost unchanged. This implies that the use of a high or low NA objective to collect the transmitted light should yield almost the same shape of a CPSF profile for the aperture. The use of a high NA objective merely enables the collection of more light from the output diffraction transmission.

We performed an experiment with a 300 nm aperture to verify this prediction. The point source was scanned at a height of 1.15 μm above the aperture. Two objectives (NA = 0.6 and NA = 0.4) were used. Due to the intervening free space between the glass substrate and the collection optics, the effective collection NA is lower than the cited objective value (see Fig. 10.12(e) for an illustration of this consideration). The effective NA of these two objectives is 0.4 and 0.267, respectively.

The measured CPSFs are plotted in Figs. 10.12(f) and (g). As can be seen from the figures, our simulation results agree well with our experimental findings. The measured signal associated with the NA = 0.4 setup [Fig. 10.12(f)] was lower than that for the NA = 0.6 setup [Fig. 10.12(g)]; this is consistent with the fact that the lower NA system will collect less light from the diffracted transmission. More importantly, the measured FWHMs of the two CPSFs match up well with each other (4% difference). This illustrates the relative insensitivity of the measured CPSFs on the collection optic's NA for small apertures.

10.4.5.2.2 Large Apertures

In contrast to small apertures, large apertures have a large number of propagating modes. In this case, the output diffraction pattern from the aperture

depends strongly on the input light field profile and direction. We repeated the simulations corresponding to the results shown in Figs. 10.12(a) and (b) and substituted a 550 nm wide aperture in place of the 300 nm aperture. The results are shown in Figs. 10.12(c) and (d). As can be seen in Fig. 10.12(d), a laterally displaced point source (x = 500 nm) gives rise to a severely skewed output diffraction pattern.

In this situation, the choice of the collection optics' NA has a significant impact on the observed CPSF. Explicitly, a small NA objective will fail to collect most of the transmitted light when the skew angle of the central diffraction lobe is larger than the collection angle of the objective. As such, a smaller collection NA system will give a narrower CPSF in comparison to a higher NA system.

This prediction is corroborated by an experiment with an aperture of size 550 nm. The experimental setup and parameters are similar to those for the above described experiment with the 300 nm aperture. The measured CPSFs and the simulation profiles are plotted in Figs. 10.12(h) and (i). The experiment and simulation are again in good agreement. The measured FWHM of the CPSF for the experiment with the smaller collection NA is 770 nm and it is 1000 nm for the large collection NA experiment. The difference in FWHM is 33%.

The finding suggests that it is possible to design type-II ABIDs to have a significantly reduced CPSF for large H values (i.e. far field), and consequently a longer DOF by restricting NA of the collection optics. This method is valid for relatively large apertures, but will have little impact on type-II ABIDs that use small apertures. However, the reduction of collection optics' NA will lead to decreased light collection efficiency and therefore reduced device's sensitivity.

10.4.5.3 Comparison of the Resolution between Type-II ABIDs and Conventional Microscopes

The CPSF of the aperture is a direct correspondence to the PSF of a conventional microscope, as both of them can be used to characterize the ability of the specific imaging system to resolve objects. In this subsection, we will compare the resolution of type-II ABIDs to that of conventional microscopes. As mentioned in Subsection 10.4.4, the resolution limit of type-II ABIDs is reached only when the target is in close proximity to the aperture. The CPSF widens with an increasing target height (H). This behavior can be found in a conventional microscope as well – the associated PSF rapidly widens when a target moves away from the focal plane of a conventional microscope.

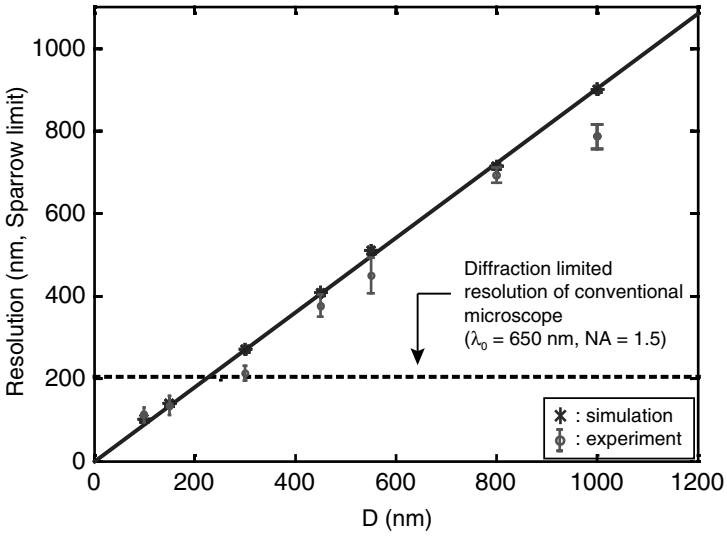


Figure 10.13 Comparison of type II ABIDs with conventional microscopes. Asterisk: Near field resolution of type II ABIDs vs. D (simulation). Simulation assumed an effective isotropic point source. Solid line: linear fit of the resolution limit of type II ABIDs. Error bars: Near field resolution of type II ABIDs based on our experiment. Horizontal dashed line: diffraction limited resolution of an ideal conventional microscope.

Figure 10.13 shows the measured resolution limit for type-II ABIDs based on our simulation results for a range of aperture sizes. We chose the Sparrow's criterion for this determination for the reason stated in Subsection 10.4.2. We have also included the estimates based on our experimental data in the figure. However, due to the previously mentioned limitations in NSOM experiments, the estimates based on our experimental data tend to over estimate the resolution. The simulation results are more accurate and unbiased.

Within the range of our measurements, the result shows an approximately linear trend. A linear fit of the simulation data was added to the plot to indicate the trend, but we note that this linearity may not extend beyond the plotted range and is purely an empirical observation.

Within the range plotted in Fig. 10.13, the resolution limit of type-II ABIDs is consistently smaller than the physical size of the aperture. In fact, the resolution limit is approximately 90% of the aperture's diameter based on our measurements. This empirical result is useful as a rule of thumb for estimating a type-II ABIDs' resolution when dealing with aperture dia-

meter in the range of 100 nm to 1000 nm. We do not expect this trend to hold for aperture of all sizes. In the limiting case where the aperture is large (e.g. 100 μm), geometric optics is valid to analyze the system. The optical cross section of a large aperture is the same as its physical area in the geometric optics regime. The CPSF at zero height should then be close to a top hat function with a width that is equal to the aperture's width. In that situation, the resolution limit should be equal to the aperture's diameter.

The dashed black line was also added to the plot to show the resolution limit (based on Sparrow's criterion) for an ideal diffraction limited conventional microscope with perfect NA. The medium's refractive index of 1.5 and a free-space wavelength of 650 nm were used in the calculation to match with the medium and illumination parameters in our sub-wavelength aperture study. Such a system has a resolution limit of about 200 nm. Based on the observed crossing point of this line with the simulation data fit, we can see that an aperture of size 225 nm can achieve the same resolution limit. Apertures that are smaller should therefore exceed the resolution limit of an ideal diffraction limited conventional microscope operating under the same wavelength and refractive index.

10.4.5.4 Summary

In summary, the light collection characteristics of subwavelength apertures with respect to point sources at varying heights and lateral displacements from the nanoaperture were systematically studied for the first time. We performed a set of NSOM-probe scanning experiments with apertures that ranges in size from 100 nm to 550 nm. We also ran a set of EM simulations based on the same geometry. The experimental results were used to verify our simulation results. We used our findings to gain a better understanding of type-II subwavelength aperture based imaging devices (ABIDs) – devices in which sub-wavelength apertures serve as light collection elements. We defined a parameter – collection mode point spread function (CPSF), which is similar to the point spread function for conventional microscopes, and used it to characterize the ABIDs.

Three significant findings were made in the study. 1) We found that, in the far field limit, the CPSF's width increases linearly with source height (H). Our simulation results indicate that, for far field, the CPSF's width is only weakly dependent on the size of the aperture. In the far field and large aperture limit, the CPSF's width is independent of the aperture size. This implies that the CPSF behavior of type-II ABIDs is quite different from the PSF behavior of type-I ABIDs at those limits. 2) In experiment, the observed CPSF is dependent on the NA of the collection optics that gathers the transmission

from the exit side of the aperture. Our study shows that the depth of field associated with apertures can be increased by restricting the collection optics' NA. 3) We observed an empirical linear relationship between the resolution limit of type-II ABIDs and the aperture size for size ranging from 100 nm to 1000 nm. Our study shows that type-II ABIDs based on small nano-apertures can deliver ultrahigh resolution beyond the diffraction limit – a restriction on the resolution of an ideal conventional microscope.

10.5 RESOLUTION AND SENSITIVITY

As we can see in Section 10.4, a high resolution OFM requires small aperture sizes. However, the amount of light transmission decreases with aperture size (D). This decrease is particularly dramatic for apertures that are much smaller than the optical wavelength, where the transmission is proportional to D^6. This claim was originally stated by Bethe in [46]. The tradeoff of resolution and sensitivity deserves some discussion.

We will briefly address this issue *in the limit where the metal layer in which the apertures are etched is perfectly conductive*. We may examine the problem in 2 different regimes:

- $D >> \lambda$ (large aperture limit) In this regime, the effective transmission area (A) is simply equal to the physical cross-section of the aperture.
- $D << \lambda$ (small aperture limit) In this regime and in the limit that the aperture is in a perfect conductor sheet that is *infinitesimally thin*, Bethe [46] showed that the effective transmission area is proportional to the sixth power of the aperture diameter. de Abajo in a recent work [47] observed that the transmission is further attenuated exponentially as a function of aperture depth, in the case where the perfect conductor sheet has a finite depth. The exact form of the exponential drop off was simulated and reported in Ref. [47], though, the normalization factor was omitted. Combining these information, we postulate that the effective transmission cross-section can be expressed as:

$$A = \frac{16\pi^3}{27} \cdot \frac{D^6}{\lambda^4} \cdot \exp\left(-4\pi d \sqrt{\frac{0.586^2}{D^2} - \frac{1}{\lambda^2}} \right). \tag{10.5}$$

This empirical expression agrees well with the simulation data that de Abajo reported. The above expressions are useful for estimating the achievable system sensitivity for a given system resolution. The total

transmission photon count for a pixel dwell time of τ is given by:

$$N_{transmission} = \frac{\varepsilon I A \tau}{hc/\lambda}. \tag{10.6}$$

A convenient definition for sensitivity of OFM devices can be based on the following simple model. The nature of this derivation is quite similar to the theoretical signal-to-noise computation that is used to calculate the sensitivity of a given optical coherence tomography (OCT) system. Knowing the total number of signal photons that is detectable and the uncertainty attributable to the presence of noise photons, the sensitivity can be ascribed to any given OCT system. A similar calculation can be performed here. Let p equals to the proportion of light occluded from an aperture by the presence of a target object over it. The threshold for this occlusion to be observable is given by equating the number change in the total detected photon counts to the noise term. The noise term is a function of the total photon count and the receiver noise ($n_r\tau$); simply, the noise contributed uncertainty term is equal to the square root of the total photon count that contributes to the DC signal. Mathematically,

$$\begin{aligned} p_{threshold} N_{transmission} &= \sqrt{N_{transmission} + (n_r \tau)} \\ Sensitivity &= \frac{1}{p_{threshold}} = \frac{N_{transmission}}{\sqrt{N_{transmission} + (n_r \tau)}}. \end{aligned} \tag{10.7}$$

This remarkably simple expression is a powerful design tool. Knowing the approximate extinction and scattering coefficient of the target object, we can compute the p characteristic for the object. For an OFM device to function well in imaging the target object, its sensitivity, $1/p_{threshold}$ must be higher than the $1/p$ value of the object.

We may draw a few conclusions from this analysis. First, a resolution of 200 nm for an object with a $1/p$ value of 55 dB is readily achievable under realistic experimental conditions (light intensity at optical tweezing limit: 10 MW/cm^2). Second, the use of high intensity illumination light field has diminishing returns in terms of achieving higher resolution; an 8 order magnitude illumination intensity increase from natural light illumination (sunlight) (~5 W/cm^2) to high laser intensity illumination (~10 MW/cm^2) is only capable of improving the resolution limit by less than an order of magnitude.

10.6 OFM VARIATIONS

There are two OFM variants that are readily implementable – fluorescence OFM and differential interference contrast (DIC) OFM. In this section, we will examine the implementation issues and explore possible applications.

10.6.1 Fluorescence OFM

A fluorescence capable OFM can be implemented by inserting filter that reflects the excitation light and transmits only the emission spectrum between the metal layer and the underlying sensor. The fabrication of a simple quarter-wave stack filter directly onto a sensor has been demonstrated previously [12].

The signal strength is a potential point of concern. Unlike in a conventional microscope, each location on the target object can only be sampled for a relatively short time interval. This time interval is equal to the time it takes any given point on the object to move across an aperture in the array. This implies that if an image acquired with the OFM with N resolvable points along the axis of the channel's length, the effective fluorescence acquisition time per pixel is N times less than if the image is acquired with a conventional microscope with the same image integration time.

On the other hand, the light collection solid angle in an OFM is effectively 2π steradian for fluorophores that are close to the aperture. This is true even for large aperture array OFM with low resolution. The large collection solid angle implies better light collection efficiency. In comparison, low resolution in a conventional microscope implies a low collection numerical aperture.

Fluorescence signal collection per pixel, resolution and throughput are codependent factors in a fluorescence capable OFM. A low resolution system allows more fluorescence signal collection through the large apertures, and decreasing sample flow speed implies that each fluorophore will take longer to flow over across an aperture and thereby increase the collected fluorescence signal.

To provide a point of reference, consider a fluorescence-capable OFM system with 1 μm diameter apertures and an object flow speed such that each fluorophore in the object takes 10 msec to pass over an aperture in the array. If we consider imaging a sample under the following conditions: sample contains 1 μM concentration of fluorescent species, the fluorescence lifetime is 1 nsec, the sample is 10 μm thick, and the incident excitation light field is sufficient high to cause saturation excitation, the total number

of fluorophores over an aperture is equal to 4×10^4 and the total number of photons collected over a 10 msec time interval is equal to 4×10^{11}. The thermal noise and other noise sources of the selected detector must be sufficiently low compared to this quantity to ensure good signal detection.

10.6.2 Differential Interference Contrast OFM

The DIC microscope is a very popular imaging tool that is used in the clinical and bioscience setting for imaging unstained samples. In this method, the contrast is provided by refractive index variations in the sample. In principle, a conventional microscope can be adapted into a DIC microscope by the inclusion of several optical elements. However, the placements are critical and a properly designed system can be very expensive. The prohibitive cost prevents wide spread usage of the DIC microscope. A low cost and robust DIC microscope is highly desirable in this context.

The imaging strategy of the DIC microscope is as follows. Consider a conventional microscope where the transmission through the sample is imaged onto an image plane (which may be a 2D sensor or the retina). In a DIC microscope, additional optical elements are added into the optical path to split the transmission into two. This creates two images on the image plane. The images are arranged to overlap; the images are laterally displaced with respect to each other by a small distance (typically equal to the resolution of the microscope). The interference of the two images translates phase variations in the original transmission into intensity variations.

In essence, this approach performs a differentiation of the transmission phase image of the object and maps the result onto an intensity map. This explains the method's name and explains the ability of DIC to highlight object edges clearly – the edges are the locations where phase varies rapidly. This also explains the opposite intensity contrast of left and right object edges, as a differentiation across a rising phase change and a falling phase change leads to positive and negative quantities.

A DIC variant of the OFM can be implemented by replacing each aperture with a pair of apertures. The arrangement of the underlying sensor will need to be slightly changed as well – a spacer will be inserted between the metal layer and the underlying sensor. In this case, the aperture pair functions like a Young's double slit. The transmission through the apertures will interfere and create an interference pattern on the underlying sensor. The interference pattern is the product of the interference fringe pattern from two point sources separated by a distance equal to the separation between the two apertures, and the diffraction pattern associated with an aperture.

The interference pattern should have a large zero order fringe and weaker higher order fringes. If the relative phase of the light incident on the apertures are not equal, this will lead to a shift of the zero order fringe off-center. By measuring the shift, we can then determine the phase skew. The total transmission through the apertures can be determined by simply integrating over the entire transmission – interference redistributes the light over the entire diffraction but does not alter total transmission.

A phase image of the target object can be collected by scanning the object with this arrangement using the OFM approach. The phase image should correspond to a DIC image where the image displacement equals to the separation of the apertures for each aperture pairs.

In Figs. 10.14(a) and (b), we illustrate the feasibility of this method with an aperture pair (hole size = 500 nm, aperture separation = 1 μm). The metal layer is spaced 150 μm from the sensor with a layer of SU-8 photoresist. In Fig. 10.14(c), one can see the shift in the collected interference pattern as the relative phase of the light incident on the apertures is varied. The minimum number of sensor pixels required is two – the determination of the shift can be made by measuring the change in the center of mass of the interference pattern.

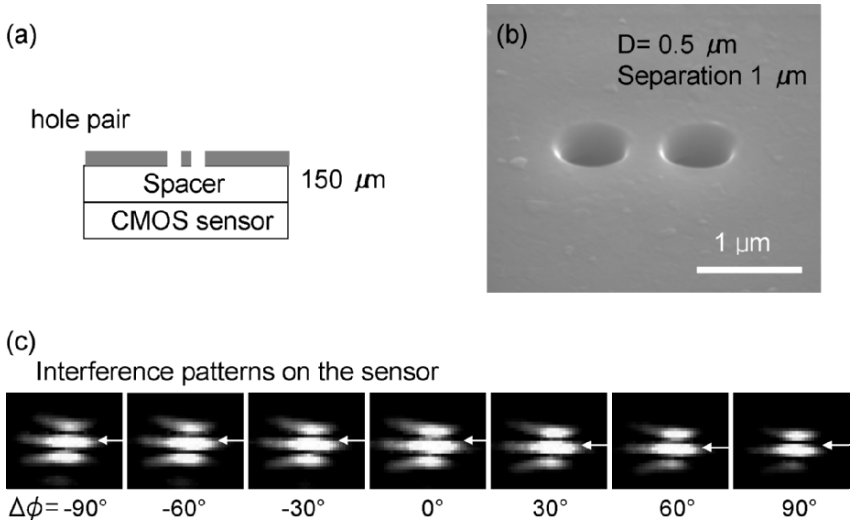


Figure 10.14 (a) Geometry of a Difference Interference Contrast OFM aperture pair. (b) SEM image of the aperture pair. (c) Changing the relative phase of light incident on the apertures, we can see a shift in the main lobe of the interference pattern on the CMOS sensor.

A DIC OFM has the ability to make quantitative phase and amplitude measurements; a conventional DIC microscope is unable to disentangle the two. Complete phase and amplitude knowledge of the object's transmission at the plane of the aperture array is very useful beyond providing DIC images. In principle, it will be possible to process the transmission wavefront to place an effective focus plane at a different depth from the aperture array. This is similar to the optical transform that occurs in a conventional optical microscope when a plane of focus in the object is selected for imaging. This is a method to achieve optical sectioning in the OFM.

10.7 CONCLUSIONS

In this chapter, we presented optofluidic microscopy, a new type of imaging scheme that can be readily integrated into microfluidic platform. Its working principle and its imaging properties (e.g. resolution, angular collection) have been explained in details. OFM's application in biology has been demonstrated in nematode imaging.

This microscopy method does not perform image replication and magnification – two functions that are associated with conventional microscopy, and yet is capable of delivering high resolution. More importantly, it achieves this without using bulk optical elements. OFM devices can be built onto chips with existing fabrication techniques and can be potentially mass manufactured at low cost (~ $10's). While the method does not provide optical sectioning, we are presently working on a variant that is potentially capable of such. This variant will be capable of high resolution phase and amplitude measurements of the optical wavefront with a similar on-chip system; knowing the wavefront, we can then perform virtual focusing into the sample of interest.

In addition to its demonstrated application in nematode imaging, the OFM can also be applied to image individual cells, e.g. blood cells and cells in suspensions. Fluorescence and interference imaging capability can also be added into the OFM platform as described in the Section 10.6.

Finally, the simplicity and compactness of the OFM systems lead to the possibility of fabricating multiple OFMs onto a single microfluidic chip. The microscopes can be used in parallel to increase imaging throughput or linked in serial to perform sequential imaging of the same targets.

ACKNOWLEDGMENT

The authors acknowledge funding from DARPA Center for Optofluidic Integration.

REFERENCES

[1] A. Y. Fu, C. Spence, A. Scherer, F. H. Arnold, and S. R. Quake, "A microfabricated fluorescence-activated cell sorter," Nature Biotechnology, vol. 17, pp. 1109-1111, 1999.

[2] Y. C. Tai, J. Xie, Q. He, J. Liu, and T. Lee, "Integrated micro/nano fluidics for mass-spectrometry protein analysis," International Journal of Nonlinear Sciences and Numerical Simulation, vol. 3, pp. 739-741, 2002.

[3] J. Xie, Y. N. Miao, J. Shih, Q. He, J. Liu, Y. C. Tai, and T. D. Lee, "An electrochemical pumping system for on-chip gradient generation," Analytical Chemistry, vol. 76, pp. 3756-3763, 2004.

[4] L. Licklider, X. Q. Wang, A. Desai, Y. C. Tai, and T. D. Lee, "A micromachined chip-based electrospray source for mass spectrometry," Analytical Chemistry, vol. 72, pp. 367-375, 2000.

[5] M. Tokeshi, Y. Kikutani, A. Hibara, K. Sato, H. Hisamoto, and T. Kitamori, "Chemical processing on microchips for analysis, synthesis, and bioassay," Electrophoresis, vol. 24, pp. 3583-3594, 2003.

[6] L. J. Jin, J. Ferrance, and J. P. Landers, "Miniaturized electrophoresis: An evolving role in laboratory medicine," Biotechniques, vol. 31, p. 1332, 2001.

[7] P. S. Doyle, J. Bibette, A. Bancaud, and J. L. Viovy, "Self-assembled magnetic matrices for DNA separation chips," Science, vol. 295, pp. 2237-2237, 2002.

[8] D. Trau, T. M. H. Lee, A. I. K. Lao, R. Lenigk, I. M. Hsing, N. Y. Ip, M. C. Carles, and N. J. Sucher, "Genotyping on a complementary metal oxide semiconductor silicon polymerase chain reaction chip with integrated DNA microarray," Analytical Chemistry, vol. 74, pp. 3168-3173, 2002.

[9] S. R. Liu, "A microfabricated hybrid device for DNA sequencing," Electrophoresis, vol. 24, pp. 3755-3761, 2003.

[10] R. D. Loberg, Y. Fridman, B. A. Pienta, E. T. Keller, L. K. McCauley, R. S. Taichman, and K. J. Pienta, "Detection and isolation of circulating tumor cells in urologic cancers: A review," Neoplasia, vol. 6, pp. 302-309, 2004.

[11] D. Lange, C. W. Storment, C. A. Conley, and G. T. A. Kovacs, "A microfluidic shadow imaging system for the study of the nematode Caenorhabditis elegans in space," Sensors and Actuators B-Chemical, vol. 107, pp. 904-914, 2005.

[12] M. L. Adams, M. Enzelberger, S. Quake, and A. Scherer, "Microfluidic integration on detector arrays for absorption and fluorescence micro-spectrometers," Sensors and Actuators a-Physical, vol. 104, pp. 25-31, 2003.
[13] D. Psaltis, R. S. Quake, and C. Yang, "Developing optofluidic technology through the fusion of microfluidics and optics," Nature, vol. 442, p. 381, 2006.
[14] X. Heng, D. Erickson, L. R. Baugh, Z. Yaqoob, P. W. Sternberg, D. Psaltis and C. Yang, "Optofluidic microscopy- a method for implementing a high resolution optical microscope on a chip," Lab on a Chip, vol. 6, pp. 1274-1276, 2006.
[15] B. Hecht, B. Sick, U. P. Wild, V. Deckert, R. Zenobi, O. J. F. Martin, and D. W. Pohl, "Scanning near-field optical microscopy with aperture probes: Fundamentals and applications," Journal Of Chemical Physics, vol. 112, pp. 7761-7774, 2000.
[16] T. R. Corle, and G. S. Kino, Confocal scanning optical microscopy and related imaging systems: San Diego: Academic Press, 1996.
[17] E. Popov, M. Neviere, P. Boyer, and N. Bonod, "Light transmission through a subwavelength hole," Optics Communications, vol. 255, pp. 338-348, 2005.
[18] E. X. Jin, and X. F. Xu, "Obtaining super resolution light spot using surface plasmon assisted sharp ridge nanoaperture," Applied Physics Letters, vol. 86, pp. 111106-1-3, 2005.
[19] X. L. Shi, L. Hesselink, and R. L. Thornton, "Ultrahigh light transmission through a C-shaped nanoaperture," Optics Letters, vol. 28, pp. 1320-1322, 2003.
[20] H. J. Lezec, A. Degiron, E. Devaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal, and T. W. Ebbesen, "Beaming light from a subwavelength aperture," Science, vol. 297, pp. 820-822, 2002.
[21] H. J. Lezec, and T. Thio, "Diffracted evanescent wave model for enhanced and suppressed optical transmission through subwavelength hole arrays," Optics Express, vol. 12, pp. 3629-3651, 2004.
[22] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, "Extraordinary optical transmission through sub-wavelength hole arrays," Nature, vol. 391, pp. 667-669, 1998.
[23] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, and W. W. Webb, "Zero-mode waveguides for single-molecule analysis at high concentrations," Science, vol. 299, pp. 682-686, 2003.
[24] J. Wenger, P. F. Lenne, E. Popov, H. Rigneault, J. Dintinger, and T. W. Ebbesen, "Single molecule fluorescence in rectangular nano-apertures," Optics Express, vol. 13, pp. 7035-7044, 2005.
[25] A. G. Brolo, R. Gordon, B. Leathem, and K. L. Kavanagh, "Surface plasmon sensor based on the enhanced light transmission through arrays of nanoholes in gold films," Langmuir, vol. 20, pp. 4813-4815, 2004.
[26] A. Partovi, D. Peale, M. Wuttig, C. A. Murray, G. Zydzik, L. Hopkins, K. Baldwin, W. S. Hobson, J. Wynn, J. Lopata, L. Dhar, R. Chichester, and J. H.

J. Yeh, "High-power laser light source for near-field optics and its application to high-density optical data storage," Applied Physics Letters, vol. 75, pp. 1515-1517, 1999.

[27] F. Chen, A. Itagi, J. A. Bain, D. D. Stancil, T. E. Schlesinger, L. Stebounova, G. C. Walker, and B. B. Akhremitchev, "Imaging of optical field confinement in ridge waveguides fabricated on very-small-aperture laser," Applied Physics Letters, vol. 83, pp. 3245-3247, 2003.

[28] A. Sundaramurthy, P. J. Schuck, N. R. Conley, D. P. Fromm, G. S. Kino, and W. E. Moerner, "Toward nanometer-scale optical photolithography: Utilizing the near-field of bowtie optical nanoantennas," Nano Letters, vol. 6, pp. 355-360, 2006.

[29] J. O. Tegenfeldt, O. Bakajin, C. F. Chou, S. S. Chan, R. Austin, W. Fann, L. Liou, E. Chan, T. Duke, and E. C. Cox, "Near-field scanner for moving molecules," Physical review letters, vol. 86, pp. 1378-1381, 2001.

[30] X. Heng, D. Erickson, L. R. Baugh, Z. Yaqoob, P. W. Sternberg, D. Psaltis and C. Yang, "Optofluidic microscopy- a method for implementing a high resolution optical microscope on a chip," Lab on a Chip, vol. 6, pp. 1274-1276, 2006.

[31] D. P. Tsai, A. Othonos, M. Moskovits, and D. Uttamchandani, "Raman-Spectroscopy Using a Fiber Optic Probe with Subwavelength Aperture," Applied Physics Letters, vol. 64, pp. 1768-1770, 1994.

[32] K. Okamoto, and S. Kawata, "Radiation force exerted on subwavelength particles near a nanoaperture," Physical Review Letters, vol. 83, pp. 4534-4537, 1999.

[33] E. Betzig, J. K. Trautman, T. D. Harris, J. S. Weiner, and R. L. Kostelak, "Breaking the Diffraction Barrier - Optical Microscopy on a Nanometric Scale," Science, vol. 251, pp. 1468-1470, 1991.

[34] D. Courjon, Near-field microscopy and near-field optics: London: Imperial College Press, 2003.

[35] S. Inoue, and K. R. Spring, Video microscopy: the fundamentals, 2nd ed: New York: Plenum Press, 1997).

[36] J. W. Goodman, Introduction to Fourier optics, 3rd ed: New York: McGraw-Hill, 2004).

[37] L. G. Schulz, and F. R. Tangherlini, "Optical Constants of Silver, Gold, Copper, and Aluminum.2. the Index of Refraction-N," Journal of the Optical Society of America, vol. 44, pp. 362-368, 1954.

[38] L. G. Schulz, "The Optical Constants of Silver, Gold, Copper, and Aluminum.1. the Absorption Coefficient-K," Journal of the Optical Society of America, vol. 44, pp. 357-362, 1954.

[39] D. E. Grupp, H. J. Lezec, T. Thio, and T. W. Ebbesen, "Beyond the Bethe limit: Tunable enhanced light transmission through a single sub-wavelength aperture," Advanced Materials, vol. 11, pp. 860-862, 1999.

[40] "COMSOL Multiphysics 3.2 (2006)," in COMSOL Inc. (http://www.comsol.com/).

[41] N. N. Rao, Elements of engineering electromagnetics, 6th ed: Upper Saddle River, N.J.: Pearson Prentice Hall, 2004.
[42] J. P. Berenger, "Three-dimensional perfectly matched layer for the absorption of electromagnetic waves," Journal of Computational Physics, vol. 127, pp. 363-379, 1996.
[43] J. Jin, The finite element method in electromagnetics, 2nd ed: New York: Wiley, 2002.
[44] F. Collino, and P. Monk, "The perfectly matched layer in curvilinear coordinates," SIAM Journal on Scientific Computing, vol. 19, pp. 2061-2090, 1998.
[45] S. D. Gedney, "An anisotropic perfectly matched layer-absorbing medium for the truncation of FDTD lattices," IEEE Transactions on Antennas and Propagation, vol. 44, pp. 1630-1639, 1996.
[46] H. A. Bethe, "Theory of diffraction by small holes," Physical Review, vol. 66, p. 163, 1944.
[47] F. de Abajo, "Light transmission through a single cylindrical hole in a metallic film," Optics Express, vol. 10, pp. 1475-1484, 2002.

AUTHOR BIOGRAPHY

Changhuei Yang received his Ph.D. degree in Electrical Engineering from the Massachusetts Institute of Technology in 2002. Following postdoctoral appointments at ESPCI, Paris and Duke University, he joined the faculty at the California Institute of Technology where he is an assistant professor in both the department of Electrical Engineering and the department of Bioengineering. His group's research area is biophotonics—the imaging and extraction of information from biological targets through the use of light. His research efforts can be categorized into two major groups – Optofluidics for biosensing and interferometry based imaging. Current research topics includes: optofluidic biosensors, quadrature phase imaging, fast microscopy techniques, phase conjugation based imaging, forward imaging needle endoscopy. He is currently the associate director of the DARPA funded Center for Optofluidic Integration at Caltech. http://www.biophot.caltech.edu/index.html

Xin Heng received his B.S degree in physics from Nanjing University, China in 2002 and the M.S. degree in Applied Physics from California Institute of Technology (Caltech) in 2003. He is currently pursuing his Ph.D. degree from Electrical Engineering in Caltech. By fusing nanophotonics with biofluidics, he is focused on developing new imaging and sensing devices that have superior properties to the conventional tools.

Xiquan Cui received his B.S. and M.S. degrees in Optical Engineering from Zhejiang University, China in 2003 and his M.S. degree in Physics from Portland State University in 2005. Now he is pursuing his Ph.D. degree from Electrical

Engineering in Caltech. His research is primarily focused on studying how light behaves and interacts with nano systems, and developing novel micro/nano opto-fluidic devices.

Demetri Psaltis obtained his PhD degree in Electrical Engineering from Carnegie Mellon University (1977). He is the Thomas G. Myers professor of Electrical Engineering at the California Institute of Technology, a full professor in Electrical Engineering and Applied Optics, and the Dean of the School of Engineering at Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. Professor Psaltis is a world-renowned expert in the area of optical information processing, holography, optical networks, imaging, optical memories and optical devices, and is known for his pioneering work in holographic memory and optical neural networks. His fundamental work has had a significant impact in the scientific community, with more than a thousand citations. He served as director of the NSF Research Center on Neuromorphic Systems Engineering from 1996-1999, and is currently leading the multi-institutional DARPA Center for Optofluidic Integration. In addition to his important contributions in fundamental and applied research, Professor Psaltis has obtained 39 patents and co-founded two start-up companies.

11 CMOS SENSORS FOR OPTICAL MOLECULAR IMAGING†

Abbas El Gamal[1]*, Helmy Eltoukhy[2] and Khaled Salama[3]

[1] Department of Electrical Engineering, Stanford University
[2] Stanford Genome Technology Center, Stanford University
[3] Department of Electrical Engineering, Rensselaer Polytechnic Institute
*abbas@ee.stanford.edu

11.1 INTRODUCTION

Optical molecular imaging applications involve the measurement and characterization of biological processes and species on the cellular and molecular level through the detection of bioreporter photoemission at visible (400-700 nm) or near-infrared (700-1000 nm) wavelengths. Such applications include systems biology, genetics, proteomics, as well as pharmacological research and development, which are key to basic life science research. For example, *in vitro* optical molecular imaging techniques have and continue to be used for DNA sequencing, gene expression studies, pathogen identification, and protein detection. Additionally, *in vivo* optical molecular imaging has been used in the study of gene expression and cell migration during mammalian and other vertebrate development, in the evaluation of animal models of human cancer, in drug discovery, and in revealing patterns of infectious disease and host response to infection. These techniques are expected to have a revolutionary impact on medicine, environmental testing, agriculture, and biodefense, especially through the development of

† Portions reprinted, with permission, from "CMOS Image Sensors," IEEE Circuits and Devices Magazine, Vol. 21. Iss. 3, © [2005] IEEE and "A 0.18 µm CMOS Bioluminescence Detection Lab-on-Chip," IEEE Journal of Solid-State Circuits, Vol. 41, No. 3, © [2006] IEEE.

hand-held, highly automated testing platforms for point-of-care and on-site environmental, forensics, and biohazard studies [1]-[3].

Current optical molecular imaging platforms, however, are expensive, bulky and/or require large amounts of reagents, making them ill-suited for such applications. This is in part due to the use of photo-multiplier tubes and ultra-cooled charge-coupled devices (CCDs), which are needed to detect very low photoemission rates. Recent developments in CMOS image sensor technology offer the promise of realizing hand-held molecular imaging systems. In addition to achieving sensitivity close to that of CCDs, CMOS image sensors consume lower power, operate at higher speeds, and offer much higher level of integration. They can also be customized to suit the characteristics of a specific molecular imaging application [4]-[8].

This chapter provides an introduction to CMOS image sensors for optical molecular biology applications. Since basic understanding of the photoemission processes used in optical molecular imaging is essential to the design of any detection system, we first briefly describe the most commonly used processes, namely fluorescence and chemiluminescence. In Section 11.3, we provide a primer on solid-state image sensors. We briefly describe their architectures and operation, explain the main differences between CCD- and CMOS-based sensors, discuss their non-idealities and performance metrics, and the multiple sampling methods used to improve their sensitivity. In Section 11.4, we focus on the question of when CMOS image sensors can or should be used in molecular imaging. We provide a methodology for answering this question and give several examples of CMOS biosensors designed specifically for molecular biology applications. Finally, in Section 11.5, we detail a recently developed CMOS-based lab-on-chip for *de novo* DNA sequencing consisting of a microfluidics substrate for reagent delivery and DNA sample confinement directly coupled via a fiber-optic faceplate to a CMOS detection system-on-chip. Following the methodology developed in Section 11.4, we first discuss the application requirements for the lab-on-chip. We determined the need for detecting emission rates below 10^{-6} lux over 30 sec of integration time at room temperature. We show how this is achieved by a combination of assay pitch matched pinned photodiodes, low noise circuit design, and on-chip averaging for noise reduction. We also report experimental results.

11.2 LUMINESCENCE

Optical molecular imaging involves the detection of photoemission through luminescence processes, such as phosphorescence, fluorescence, and bioluminescence. Such photoemission is the result of chemical, biochemical, or crystallographic changes, the motions of subatomic particles, or radiation-induced excitation of an atomic system. The intensity of photoemission in a luminescence assay is typically well correlated with the relative concentration of the analyte, thereby enabling identification as well as quantification of biological species that can be coupled to a luminescence-based marker or chemistry [9].

11.2.1 Fluorescence

Generation of luminescence through excitation of a molecule by ultraviolet or visible light photons is a phenomenon termed photoluminescence and is formally divided into two categories, fluorescence and phosphorescence, depending upon the electronic configuration of the excited state and the emission pathway. Fluorescence is the property of some atoms and molecules to absorb light at a particular wavelength and to subsequently emit light at a longer wavelength after a brief interval, termed the fluorescence lifetime.

More specifically, such photoemission is the result of fast intramolecular vibrational relaxations that emit photons at a longer wavelength than

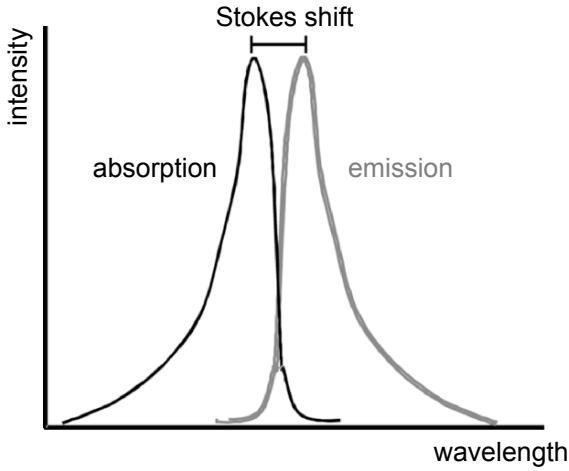


Figure 11.1 Absorption and emission spectra for a typical fluorescence dye. The separation in peak wavelengths is known as the Stokes shift.

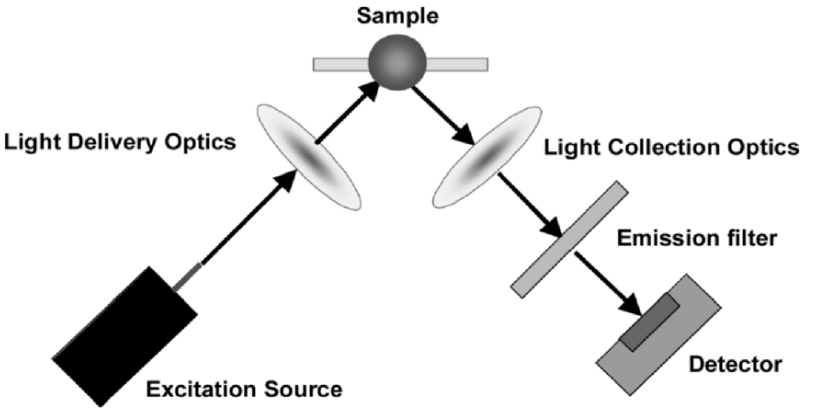


Figure 11.2 Typical fluorometric setup with excitation source, focusing lens, fluorescent sample, emission filter, and photodetector.

that of the incident photons. The difference between the two wavelengths is referred to as the Stokes shift (see Fig. 11.1). This property is the basis of the simple separation of emitted fluorescence from excitation light, which renders fluorescence such a powerful tool. The emitted photon is detected within a typical delay (lifetime) after absorption of the excitation photon that depends on the species studied and its local environment. Organic dyes have typical lifetimes from several tens of picoseconds to several nanoseconds. Longer lifetimes are obtained for fluorescent semiconductor nanocrystals (NCs) (tens of nsec), organometallic compounds (hundreds of nsec), and lanthanide complexes (up to msec). The probability distribution of emission times is usually monoexponential, characterized by a lifetime $\tau = (k_r + k_{nr})^{-1}$, where k_r and k_{nr} are the radiative and non-radiative decay rates, respectively. The intensity of photoemission is usually a direct function of the input power of the excitation source, i.e. $P_{emission} = \alpha P_{excitation}$ [10].

Fluorescence is most known for its use in Sanger sequencing and DNA microarrays. In these applications, short DNA sequences of interest, known as oligonucleotides, are detected through the presence of a bound fluorescence tag or marker (typically Cy3 or Cy5). These tags are detected through a fluorometric setup, whereby an excitation source, e.g. a laser, initiates the fluorescence excitement and a detection system with appropriate optical interference filters is used to record the resultant photoemission (see Fig. 11.2). The most challenging aspect in fluorometry is the need to filter the unwanted background radiation from the excitation source and sample medium, which can be very close to the emission in wavelength. Advanced

systems sometimes employing evanescent wave optical setups can achieve a detection sensitivity of one fluorescent tag, i.e. one fluorophore.

11.2.2 Bio-/Chemi-Luminescence

Emission of light as a result of a chemical reaction at room temperatures is known as chemiluminescence. When the reaction takes place inside a living organism it is known as bioluminescence. The time-dependent light generation from a typical chemiluminescence process is a function of the underlying chemical reaction kinetics. The rate of a reaction, in general, is the speed at which reactants are converted into products. If enzyme species E (the catalyst) converts the substrate molecule S into product P, the stoichiometric formula is given by

$$S + E \xleftrightarrow{k_f, k_r} E + P, \tag{11.1}$$

where k_f and k_r are the association and disassociation rate constants. The reaction rate is defined by

$$rate = \frac{d[P]}{dt} = -\frac{d[S]}{dt} = k_f[S][E] - k_r[E][P], \tag{11.2}$$

where $[E]$, $[S]$ and $[P]$ are the concentrations of the enzyme, substrate and product in the medium, respectively. Now if we suppose that the above process is a luminescence enzymatic reaction with quantum yield α, then the photon generation rate I in volume V of the reaction medium (A is Avogadro's number) would be

$$I = (\alpha VA)\frac{d[P]}{dt} = (\alpha VA)\cdot[E]\cdot\left(k_f[S] - k_r[P]\right). \tag{11.3}$$

The total number of photons generated by this luminescence process $N_{ph}(T)$, in the time interval T, would be

$$N_{ph}(T) = (\alpha VA)\cdot[E]\cdot\int_T (k_f[S] - k_r[P])\cdot dt. \tag{11.4}$$

In luminescence assays the experiment is typically set up in such a way that the luminescence probe, e.g. a light generating enzyme, either reports the quantity of a substrate molecule, e.g. adenosine tri-phosphate (ATP), or the molecule to which it binds, e.g. luciferase-based labels in immunoas-

Table 11.1 Summary.

	Fluorescence	Chemiluminescence
Wavelength	Visible, near-IR	Visible (425, 560 nm)
Emission Time	psec ~ msec	sec ~ hours
Emission Rate per reporter	Up to 10^5 ph/sec	0.01 – 1 ph/sec
Sensitivity	Single molecule	1000 molecules
Background	High (need filters)	Relatively low

says. The photon generation rate from the luminescence reaction, which is a function of the target concentration, is then measured and correlated to the target concentration [11].

The firefly reaction has the highest known quantum yield or efficiency of 88%. In this reaction, ATP reacts with luciferin with the aid of the enzyme luciferase to yield an intermediate complex as well as photoemission as a by-product. Its high efficiency makes it one of the most widely used chemiluminescent chemistries.

Numerous bioluminescent and chemiluminescent reactions are studied using luminometry and are commonly used in biotechnology research, environmental testing, industrial applications, and clinical research. Among its many applications are the measurements of gene expression using reporter gene assays, the determination of intracellular ATP, and DNA sequencing [12]. Chemiluminometry is also more amenable to lab-on-chip integration than the more widely used fluorometric methods. Signal photoemission is triggered simply upon the addition of the final chemical required to precipitate the luminescence reaction, with very little associated background emission and without the need for interference filters. This is the approach we have taken for our lab-on-chip system described in Section 11.5.

11.3 SOLID-STATE IMAGE SENSORS

Various detection systems have been developed for luminometry. By far the most sensitive detection devices are photomultiplier tubes (PMTs), which via a photocathode and a series of amplifying dynodes can generate up to one million electrons for every incident photon. These devices are very sensitive since their noise can be removed easily using a level-discriminator. However, they are costly and require high operating voltages (1000 VDC), precluding their use in a hand-held system. Furthermore, the overall photon detection efficiency of PMT-based systems is limited to 1-4% by the optics and the low quantum efficiency (10-20%) of PMTs. Finally, multiplexed

imaging is not readily feasible using PMT's, as they are relatively large and thus unsuitable for dense arrays. It is this aspect of PMT's that has most limited their use and applicability [9]. For the remainder of the section, we will focus on solid-state imaging, namely through CCD and CMOS-based architectures.

We begin with a general discussion of the main differences between CCD and CMOS image sensors. In the rest of the section we discuss the signal path through a CMOS sensor from the incident photon flux up to the output voltage. We divide the discussion into two subsections, the first dealing with the photodetection operation and the second dealing with CMOS readout architectures.

An image sensor consists of a 1D or 2D array of pixels, each containing a photodetector that converts incident light into photocurrent and some of the readout circuits needed to convert the photocurrent into electric charge or voltage and to read it off the array. The fraction of area occupied by the photodetector in a pixel is known as *fill factor*. The rest of the readout circuits are located at the periphery of the array and are multiplexed by the pixels. Array sizes can be as large as tens of megapixels for high-end applications, while individual pixel sizes can be as small as 1.6 μm × 1.6 μm [13]. Color filters and microlens array are typically deposited on top of the pixel array to perform color imaging and increase the amount of light incident on each photodetector.

Many of the differences between CCD and CMOS image sensors arise from differences in their readout architectures. In a CCD [see Fig. 11.3(a)], charge is shifted out of the array via vertical and horizontal CCDs, then

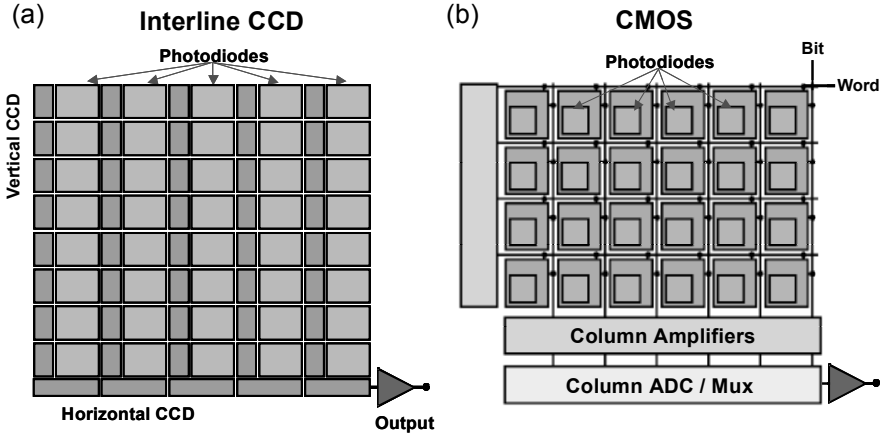


Figure 11.3 (a) Readout architectures of interline transfer CCD and (b) CMOS image sensors.

converted into voltage via a simple follower amplifier and serially read out. In a CMOS image sensor, charge or voltage signals are read out one row at a time in a manner similar to a random access memory using row and column select circuits [see Fig. 11.3(b)]. The main advantage of the CCD readout architecture is that it requires minimal pixel overhead, making it possible to design image sensors with very small pixels. Another important advantage is that charge transfer is passive and therefore does not introduce temporal noise or pixel to pixel variations due to device mismatches, known as *fixed pattern noise* (FPN). Charge transfer readout, however, is serial resulting in limited readout speed. It is also high power due to the need for high rate, high voltage clocks to achieve near perfect charge transfer efficiency. By comparison, the random access readout of CMOS image sensors provides the potential for high-speed readout and window-of-interest operations at low power consumption. There are several recent examples of CMOS image sensors operating at hundreds of frames per second with mega pixel or more resolution [14]-[16]. The high-speed readout also makes CMOS image sensors ideally suited for implementing very high-resolution imagers with multi-mega pixel resolutions, especially for video or high frame-rate applications. The readout path in a CMOS image sensor, however, contains several active devices that introduce both temporal noise and FPN.

Other differences between CCDs and CMOS image sensors arise from differences in their fabrication technologies. CCDs are fabricated in specialized technologies solely optimized for imaging and charge transfer. The disadvantage of using such specialized technologies, however, is the inability to customize the sensor front-end for a specific application as well as the inability to integrate analog and digital circuits on the same chip with the sensor. CMOS image sensors, on the other hand, are fabricated in standard technologies and thus can be readily customized for a particular application. Such customization combined with the integration possibilities offered by CMOS can reduce imaging system power and size and enables the implementation of new sensor functionalities.

11.3.1 Photodetection

An image sensor converts incident photon flux to output voltage (see Fig. 11.4). The conversion from photon flux to photocurrent is performed by the photodetector. As the range of photocurrents produced under typical illumination conditions is too low (in the range of femto- to pico-amperes) to be read directly, it is integrated and read out as charge or voltage at the end

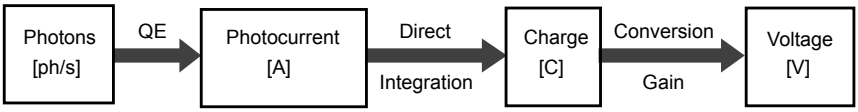


Figure 11.4 Block diagram of signal path for an image sensor.

of the integration time. This operation is referred to as *direct integration*. In the following we discuss the photodetection operation in more detail.

11.3.1.1 Photodetectors

The most popular types of photodetectors used in image sensors are the reverse-biased PN junction photodiode and the P+NP pinned diode (see Fig. 11.5). The structure of the pinned diode provides improved photoresponsivity at shorter wavelengths relative to the standard PN junction [17]. Moreover, the pinned diode exhibits lower thermal noise due to the passivation of defect and surface states at the Si/SiO_2 interface, as well as a customizable photodiode capacitance via the charge transfer operation through transistor TX. The main imaging parameters of a photodetector are quantum efficiency (QE) and dark current.

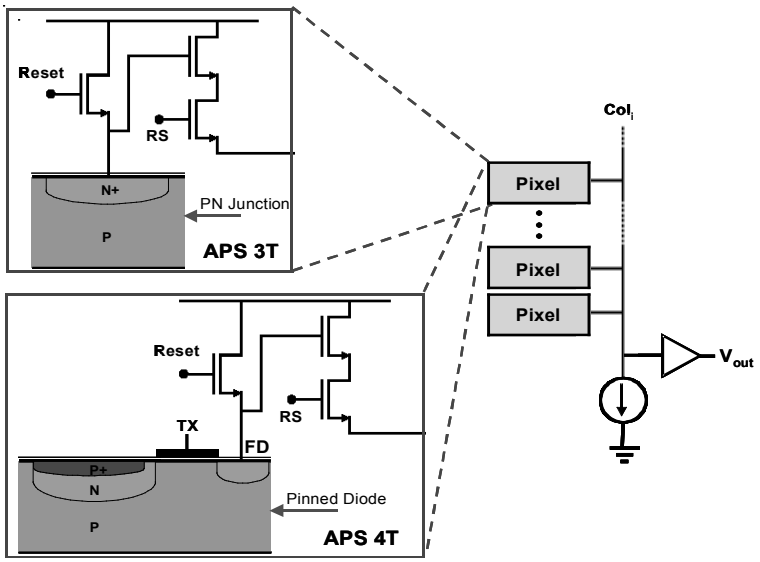


Figure 11.5 Schematic of a 3T and 4T active pixel sensor (APS).

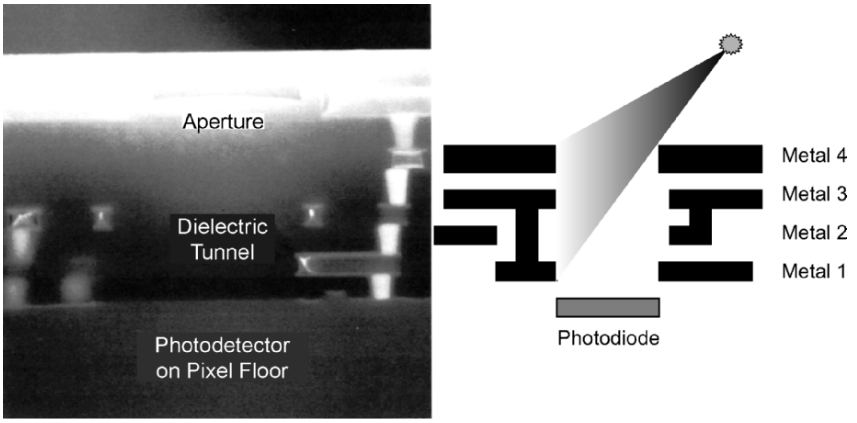


Figure 11.6 Illustration of optical "tunnel" above photodetector and pixel vignetting phenomenon.

11.3.1.1.1 Quantum Efficiency

QE is the fraction of incident photon flux that contributes to the photocurrent in a photodetector as a function of wavelength. It is the product of optical efficiency, which is the photon-to-photon efficiency from the pixel surface to the photodetector's surface, and the *internal QE,* which is the fraction of photons incident on the photodetector surface that contributes to the photocurrent. Optical efficiency depends on the geometric arrangement of the photodetector with respect to other elements of the pixel structure, i.e. shape and size of the aperture, length of the dielectric "tunnel", and position, shape, and size of the photodetector. Figure 11.6 is an SEM photograph of a cross section through a CMOS image sensor pixel illustrating the tunnel through which light must travel before reaching the photodetector. Internal QE is a function mainly of photodetector geometry and doping concentrations and is always less than one for the above silicon photodetectors. The QE for a typical CMOS image sensor fabricated in a modified 0.18 μm process is shown in Fig. 11.7.

CMOS technology scaling has detrimental effects on pixel optical efficiency. The use of silicon dioxide/nitride materials reduces light transmission to the photodetector. Moreover, as CMOS technology scales, the distance from the surface of the chip to the photodiode increases relative to the photodiode lateral dimension (see Fig. 11.6). This is due to the reduction in pixel size and the fact that the thickness of the interconnect layers scales slower than the planar dimensions. As a result, light must travel through an increasingly deeper and/or narrower "tunnel" before reaching the photo-

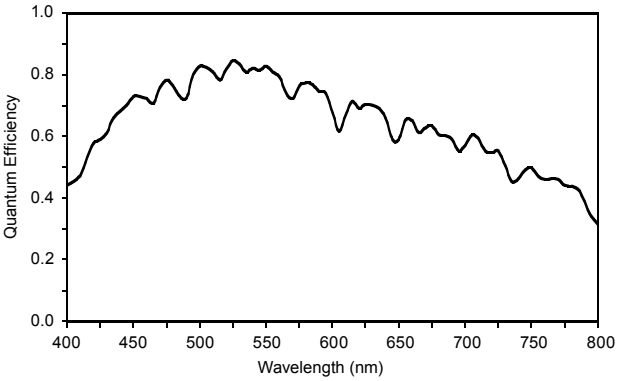


Figure 11.7 Quantum efficiency curve for a typical 0.18 μm CMOS image sensor.

diode surface. This is especially problematic for light incident at an oblique angle. In this case, the tunnel walls cast a shadow on the photodiode area, a non-ideality known as *pixel vignetting*, resulting both in a severe reduction in QE and in optical color crosstalk between adjacent pixels [18]. Several process modifications are being made in order to increase optical efficiency:

- Oxide materials with better light transmission properties are being used.
- Thinning of metal and oxide layers is used to decrease the aspect ratio of the tunnel above each photodetector, thereby reducing pixel vignetting. For example, in [19] a CMOS-based 1P2M process with 30% thinner metal and dielectric layers is developed and used to increase pixel sensitivity.
- Placement of air gaps around each pixel is another technique used in order to create a rudimentary optical waveguide whereby incident light at the surface is guided to the correct pixel below via total internal reflection. The air gaps also serve to significantly reduce optical spatial crosstalk, which can be particularly problematic as pixel sizes decrease [20].

11.3.1.1.2 Dark Current

The second important imaging characteristic of a photodetector is its leakage or dark current, which is the photodetector current when no illumination is present. It is generated by several sources, including:

- carrier thermal generation and diffusion in the neutral bulk,
- thermal generation in the depletion region,

- thermal generation due to surface states at the silicon-silicon dioxide interface, and
- thermal generation due to interface traps (caused by defects) at the diode perimeter.

Dark current is detrimental to imaging performance under low illumination as it introduces shot noise that cannot be corrected for as well as non-uniformity due to its large variation over the sensor array. Much attention is paid to minimizing dark current in CCDs, which can be as low as 1-2 pA/cm^2, through the use of gettered, high-resistivity wafers to minimize traps from metallic contamination as well as buried channels and multi-phase pinned operation to minimize surface generated dark current [21]. Dark current in standard submicron CMOS processes is orders of magnitude higher than in a CCD and several process modifications are used to reduce it down to the 30-200 pA/cm^2 range [22]. Fortunately, somewhat higher dark current can be tolerated in CMOS image sensors, since in a CCD dark current affects both photodetection and charge transfer.

CMOS technology scaling results in poor photoresponsivity and high leakage in several ways. The use of shallow junctions and high doping result in low photoresponsivity, and the use of shallow trench isolation (STI), thin gate oxide, and silicide result in unacceptably high dark current. Furthermore, in-pixel transistor leakage becomes a significant source of dark current. Indeed in a standard process, dark current due to the reset transistor off-current and the follower transistor gate leakage current in an APS pixel can be orders of magnitude higher than the diode leakage itself. To address these problems, several modifications to CMOS process have been made. These include:

- To improve photoresponsivity, non-silicided deep junction diodes with optimized doping profiles are added to a standard process.
- To reduce dark current non-silicided, double-diffused source/drain implantation as well as pinned diode structures are included. Hydrogen annealing is also used to reduce leakage by passivating defects [22].
- To reduce transistor leakage, both the reset and follower transistors in an APS use thick gate oxide (70 Å). The reset transistor threshold is increased to reduce its off-current, while the follower transistor threshold is decreased to improve voltage swing.

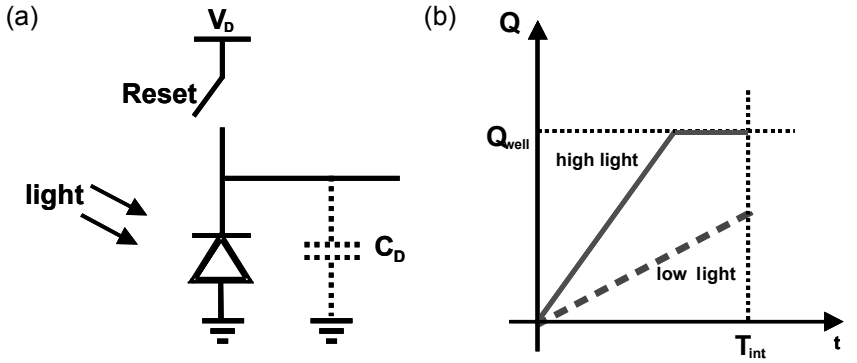


Figure 11.8 (a) Schematic of pixel operating in direct integration. (b) Charge versus time for two photocurrent values.

11.3.1.2 Direct Integration

Direct integration is used to convert photocurrent to charge or voltage as illustrated in Fig. 11.8. The photodiode is first reset to a voltage V_D. The reset switch is then opened and the photocurrent i_{ph} as well as the dark current, i_{dc}, are integrated over the diode capacitance C_D. At the end of integration, the charge accumulated over the capacitance is either directly read out as in CCDs or PPS and then converted to voltage or directly converted to voltage and then read out as in APS. In both cases, the charge-to-voltage conversion is linear and the sensor *conversion gain* is measured in microvolts per electron. The charge versus time for two photocurrent values is illustrated in Fig. 11.8(b). In the low light case the charge at the end of integration is proportional to the light intensity, while in the high light case the diode saturates and the output charge is equal to the well capacity Q_{well}, which is defined as the maximum amount of charge (in electrons) that can be held by the integration capacitance.

11.3.2 CMOS Architectures

There are different flavors of CMOS image sensor readout architectures. We describe the passive pixel sensor, which is the earliest CMOS image sensor architecture (see Fig. 11.9), the 3 and 4 transistor per pixel active pixel sensor, which are the most popular architectures at present (see Fig. 11.5), and the digital pixel sensor (see Fig. 11.11).

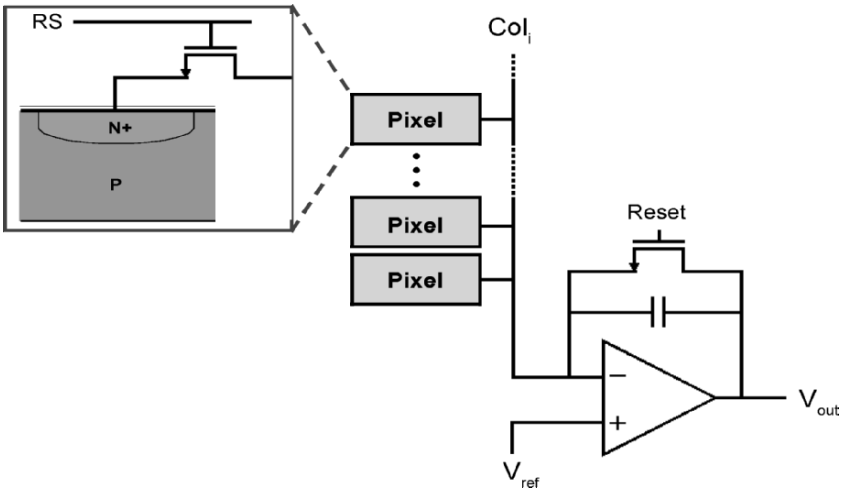


Figure 11.9 Schematic of a passive pixel sensor (PPS).

11.3.2.1 Passive Pixel Sensor (PPS)

The PPS pixel includes a photodiode and a row select transistor (see Fig. 11.9). The readout is performed one row at a time in a staggered "rolling shutter" fashion. At the end of integration, charge is read out via the column charge-to-voltage amplifiers. The amplifiers and the photodiodes in the row are then reset before the next row readout commences. The main advantage of PPS is its small pixel size. The column readout, however, is slow and vulnerable to noise and disturbances. The APS and DPS architectures solve these problems, but at the cost of adding more transistors to each pixel.

11.3.2.2 3T-Active Pixel Sensor (APS)

The 3T APS pixel includes a reset transistor, a source follower transistor to isolate the sense node from the large column bus capacitance and a row select transistor. The current source component of the follower amplifier is shared by a column of pixels. Readout is performed one row at a time. Each row of pixels is reset after it is read out to the column capacitors via the row access transistors and column amplifiers. Note that unlike CCD and PPS, APS readout is non-destructive. Although the main purpose of the extra transistors in the APS pixel is to provide signal buffering to improve sensor readout speed and SNR, they have been used to perform other useful functions. By appropriately setting the gate voltage of the reset transistor in an APS pixel, *blooming*, which is the overflow of charge from a saturated pixel into its neighboring pixels, can be mitigated [23]. The reset transistor

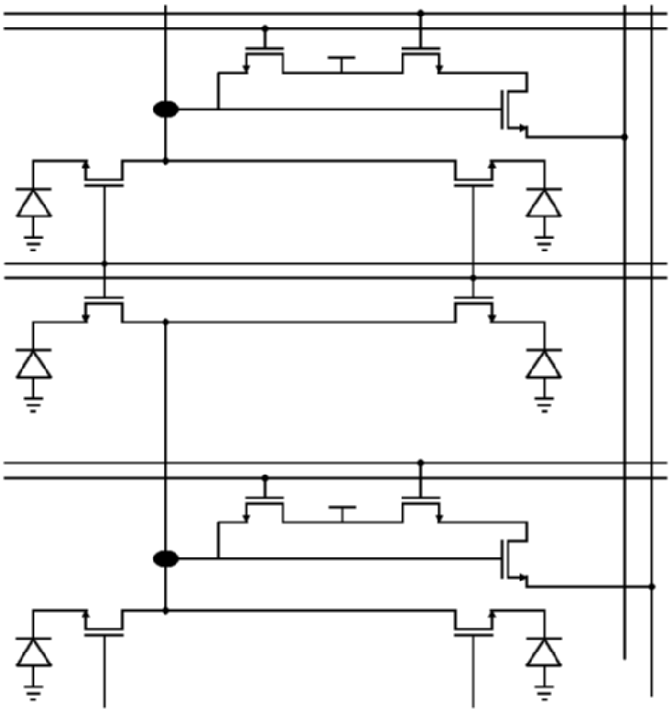


Figure 11.10 Schematic of a shared transistor 4T APS pixel.

can also be used to enhance dynamic range via well capacity adjusting as described in [24].

11.3.2.3 4T-Active Pixel Sensor (APS)

The 4T APS architecture employs a pinned diode, which adds a transfer gate and a floating diffusion (FD) node to the basic 3T APS pixel architecture. At the end of integration, the accumulated charge on the photodiode is transferred to the FD node. The transferred charge is then read out as voltage in the same manner as in the 3T architecture.

In addition to the lower dark current and better QE of a pinned diode, the 4T APS has several other advantages over 3T. The use of a transfer gate and the FD node in a 4T pixel decouples the read and reset operations from the integration period enabling true correlated double sampling (CDS) as will be discussed in detail in Subsection 11.3.4. Moreover, in a 3T pixel conversion gain is set primarily by the photodiode capacitance, while in a 4T pixel, the capacitance of the FD node can be selected independently of the photodiode size allowing conversion gain to be optimized for the sensor

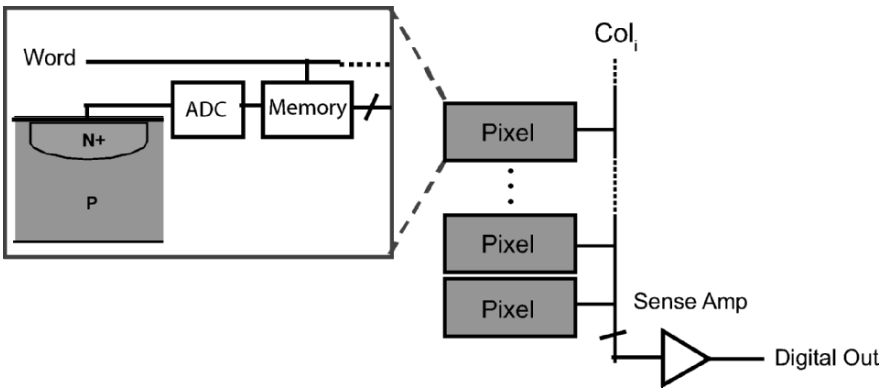


Figure 11.11 Schematic of a digital pixel sensor (DPS).

application. Another advantage of 4T over 3T is pixel size. Even though 4T pixel has more devices than a 3T pixel, some of these devices can be shared between neighboring pixels resulting in very small pixel sizes. One example is the 1.75 transistor per pixel APS described in [25]. In this architecture, the buffer of the 4T APS pixel is shared among each 4 neighboring pixels using the transfer gates as a multiplexer (Fig. 11.10).

11.3.2.4 Digital Pixel Sensor (DPS)

The third and most recently developed CMOS image sensor architecture is DPS where A/D conversion is performed at each pixel or group of neighboring pixels and digital data is read out from the pixel array in a manner similar to a random access digital memory. Figure 11.11 depicts a block diagram of a DPS pixel consisting of a photodetector, an ADC, and digital memory for temporary storage of data before digital readout via the bit-lines. DPS offers several advantages over analog image sensors, such as PPS and APS, including better scaling with CMOS technology due to reduced analog circuit performance demands and the elimination of read-related column fixed-pattern noise (FPN) and column readout noise. More significantly, employing an ADC and memory at each pixel to enable massively parallel A/D conversion and high-speed digital readout, provides unlimited potential for high-speed "snap-shot" digital imaging [26, 27]. The main drawback of DPS is that it requires the use of more transistors per pixel than conventional image sensors, resulting in larger pixel sizes or lower fill factors This problem is becoming less severe as CMOS technology scales down to 0.18 μm and below.

11.3.3 Non-idealities and Performance Measures

Image sensors suffer from several fundamental and technology related non-idealities that limit their performance. In this section we discuss these non-idealities and their effect on image sensor SNR and dynamic range.

11.3.3.1 Temporal and Fixed Pattern Noise

Temporal noise is the most fundamental non-ideality in an image sensor as it sets the ultimate limit on signal fidelity. This type of noise is independent across pixels and varies from frame to frame. Sources of temporal noise include dark and photocurrent generated shot noise, pixel reset circuit noise, readout circuit thermal and flicker noise, and quantization noise. There are more sources of readout noise in CMOS image sensors than CCDs introduced by the pixel and column active circuits.

In addition to temporal noise, image sensors also suffer from FPN, which is the pixel-to-pixel output variation under uniform illumination due to device and interconnect mismatches across the image sensor array. These variations cause two types of FPN: offset FPN, which is independent of pixel signal, and gain FPN or photo response nonuniformity (PRNU), which increases with signal level. Offset FPN is fixed from frame to frame but varies from one sensor array to another. Again, there are more sources of FPN in CMOS image sensors than CCDs introduced by the active readout circuits. The most serious additional source of FPN is the column FPN introduced by the column amplifiers. Such FPN can cause visually objectionable streaks in the image. Additionally, offset FPN due to pixel-to-pixel dark current variation is commonly referred to as dark signal non-uniformity (DSNU). Since dark current in CMOS image sensors can be much higher than in CCDs, DSNU can greatly degrade CMOS image sensor performance under low illumination levels. This is most pronounced at high temperatures, as dark current and thus DSNU exponentially increases with temperature, roughly doubling every 7 °C. DSNU can be corrected for using digital calibration. However, strong dependence on temperature makes accurate calibration difficult.

Temporal noise and FPN determine the range of illumination that can be detected by the image sensor, known as its dynamic range (DR), and the quality of the signals it produces within the detection range measured by the signal-to-noise ratio (SNR). We define sensor SNR and DR and discuss their dependence on various sensor parameters.

11.3.3.2 Signal-to-Noise Ratio (SNR)

The noise power at the end of integration can be expressed as the sum of six independent components:

- integrated shot noise Q_{shot}, which has zero mean and average power $(i_{ph}+i_{dc})t_{int}/q = (Q_s+Q_{dc})e^2$,
- reset (or kTC) noise Q_{reset},
- thermal noise of the readout circuit noise Q_{read} with zero mean and average power σ_{read}^2,
- the flicker (or $1/f$) component of readout noise $Q_{1/f}$,
- offset FPN Q_{fpn} due to photodetector and device mismatches,
- gain FPN, also known as PRNU, Q_{PRNU} and is a function of t_{int}.

Assuming the mean of the background signal is subtracted off by capturing several images before the desired photoemission process is triggered, which is feasible in many molecular imaging scenarios, the output charge including these six noise components as well as the signal charge Q_s is thus given by

$$S = Q_s + Q_{shot} + Q_{reset} + Q_{read} + Q_{fpn} + Q_{1/f} + Q_{PRNU} \text{ electrons.} \qquad (11.5)$$

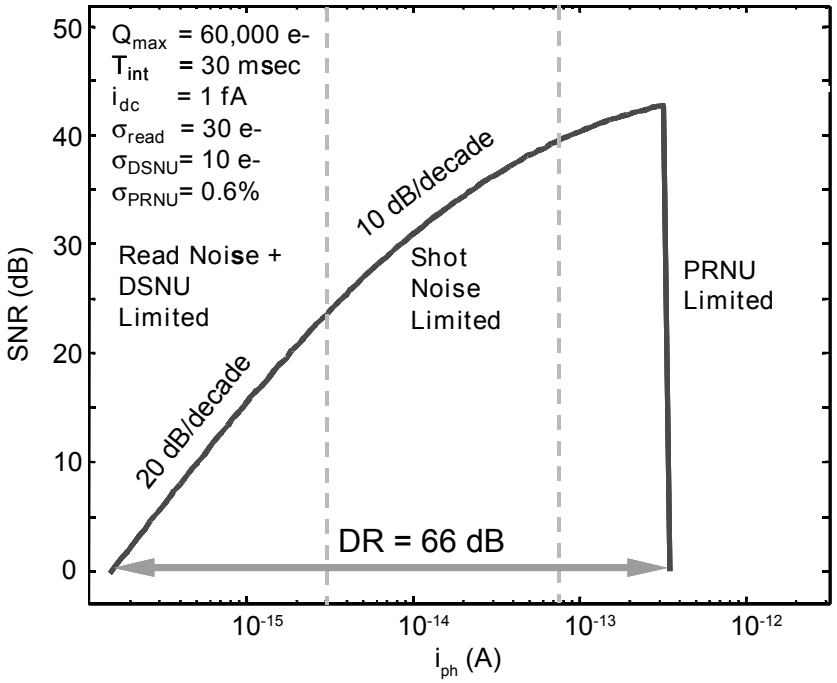


Figure 11.12 SNR versus photocurrent (i_{ph}) for a typical image sensor.

With this noise model, we can quantify pixel signal fidelity using the Signal-to-Noise Ratio (SNR), which is the ratio of the signal power to the total average noise power, as

$$\mathrm{SNR} = 10\log_{10}\frac{Q_s^2}{Q_{shot}+Q_{reset}+Q_{read}+Q_{1/f}+Q_{fpn}+Q_{PRNU}}. \tag{11.6}$$

SNR for a set of typical sensor parameters is plotted in Fig. 11.12. Note that it increases with photocurrent ($i_{ph}=qQ_s/t_{int}$), first at 20 dB per decade when readout noise and shot noise due to dark current dominate, then at 10 dB per decade when shot noise dominates, and then flattens out when PRNU is dominant, achieving a maximum roughly equal to the well capacity Q_{well} before saturation. SNR also increases with integration time. Thus it is always preferred to have as long an integration time as possible.

11.3.3.3 Dynamic Range

Sensor dynamic range quantifies its ability to image scenes with wide spatial variations in illumination. It is defined as the ratio of a pixel's largest non-saturating photocurrent i_{max} to its smallest detectable photocurrent i_{min}. The largest saturating photocurrent is determined by the well capacity and integration time as $i_{max} = q(Q_{well} - Q_{dc})/t_{int}$. The smallest detectable signal is set by the RMS of the noise under dark conditions as

$$i_{min} = q\left(\sqrt{Q_{dc}^2+Q_{reset}^2+Q_{read}^2+Q_{1/f}^2+Q_{fpn}^2}\right)/t_{int}. \tag{11.7}$$

This limit is particularly important in applications requiring low light detection. It can be improved by reducing dark current and read noise. It can be improved also by increasing integration time. However, motion blur among other factors impose constraints on integration time.

The sensor dynamic range can be expressed as

$$DR = 20\log_{10}\frac{Q_{well}-Q_{dc}}{\sqrt{Q_{dc}^2+Q_{reset}^2+Q_{read}^2+Q_{1/f}^2+Q_{fpn}^2}}. \tag{11.8}$$

Note that even though i_{min} decreases with integration time, dynamic range decreases as integration time increases due to the adverse effects of dark current and background. On the other hand, increasing well capacity, decreasing read noise and decreasing FPN, increases sensor dynamic range.

Example 1

Suppose we want to calculate the minimum detectable signal for a CMOS image sensor with 10 μm pixel pitch, 50% fill factor, 50% QE, $i_{dc} = 0.2$ fA, $Q_{well} = 50{,}000e^-$, conversion gain = 20 μV/e$^-$, $t_{int} = 1$ sec, $Q_{reset} = 30e^-$, $Q_{read} = 10e^-$, $Q_{1/f} = 3e^-$, and $Q_{fpn} = 100e^-$. Using the dark current and integration time, we calculate $Q_{dc} = 1250e^-$. Thus, given $q = 1.6\times10^{-19}$ C/e-, $i_{min} = 0.018$ fA corresponding to 5×10^8 photons/sec·cm^2. Also, we can use the above relation to determine that the dynamic range of the sensor is approximately 52 dB.

Sensor dynamic range is generally not wide enough to image the range of possible photoemissions encountered in some molecular biology applications. This is especially the case for CMOS image sensors, since their read noise and DSNU are typically larger than CCDs. For reference, standard CMOS image sensors have a dynamic range of 40-60 dB, CCDs around 60-70 dB, while the human eye exceeds 90 dB by some measures. In contrast, some molecular biology applications often require greater than 100 dB of dynamic range. To solve this problem, several dynamic range extension techniques such as *well-capacity adjusting* [24], *multiple capture* [26], *time-to-saturation* [28], and *self-reset* [29] have been proposed. These techniques extend dynamic range at the high illumination end by increasing i_{max}. In multiple capture and time-to-saturation this is achieved by adapting each pixel's integration time to its photocurrent value, while in self-reset the effective well capacity is increased by "recycling" the well. To perform these functions most of these schemes require per-pixel processing. A comparative analysis of these schemes based primarily on SNR is presented in [30]-[32].

11.3.3.4 Spatial Resolution

An image sensor is a spatial (as well as temporal) sampling device. As a result, its spatial resolution is governed by the Nyquist sampling Theorem. Spatial frequencies (in lp/mm) that are above the Nyquist rate cause aliasing and cannot be recovered. Below the Nyquist rate, low pass filtering due to the optics, spatial integration of photocurrent, and crosstalk between pixels, cause the pixel response to fall off with spatial frequency. Spatial resolution below the Nyquist rate is measured by the modulation transfer function (MTF), which is the contrast in the output image as a function of frequency.

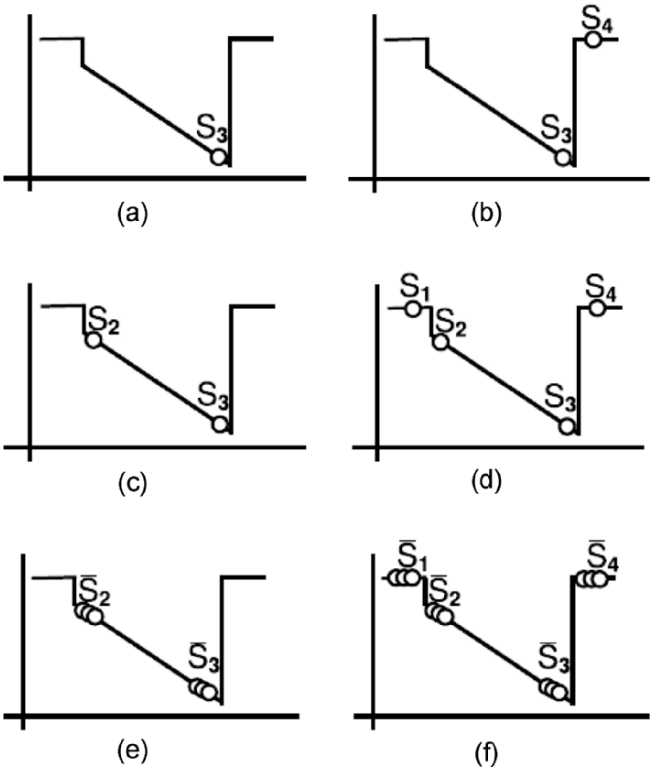


Figure 11.13 Sampling schemes: (a) Video, (b) Delta-Reset, (c) Correlated Double Sampling, (d) Correlated Triple Sampling, (e) Multiple Correlated Sampling and (f) Correlated Multiple Sampling.

11.3.4 Sampling Techniques for Noise Reduction

Low light detection sensitivity can be greatly enhanced using various multiple sampling techniques. We briefly describe these techniques and analyze their noise suppression effectiveness using the above noise model. As these sampling techniques are aimed at improving low light detection, we ignore PRNU in the following analysis and assume that accurate background subtraction has been performed.

11.3.4.1 Delta-Reset

This is the scheme used in photodiode Active Pixel Sensors (APS). Two samples are taken, one at the end of integration and the other after or during the following reset as shown in Fig. 11.13(b). The signal is then estimated as the difference between these two samples. To explain the effect of this

operation we index the noise components for each sample by their sample number to obtain

$$\begin{aligned} S_3 - S_4 &= Q_s + Q_{shot} + (Q_3 - Q_4)_{reset} + (Q_3 - Q_4)_{read} + (Q_3 - Q_4)_{fpn} \\ &\approx Q_s + Q_{shot} + (Q_3 - Q_4)_{reset} + (Q_3 - Q_4)_{read}. \end{aligned} \tag{11.9}$$

Note that this operation cancels offset FPN. It also reduces $1/f$ noise [33], since the two samples are very close in time. However, it doubles reset and read noise power, since they are uncorrelated.

11.3.4.2 Correlated Double Sampling (CDS)

This is the scheme used in most CCDs and 4T pinned photodiode architectures [33]-[34]. Two samples are taken, one right after reset of the floating diffusion node and the second after the signal charge is transferred [see Fig. 11.13(c)]. The signal is estimated as

$$\begin{aligned} S_3 - S_2 &= Q_s + Q_{shot} + (Q_3 - Q_2)_{reset} + (Q_3 - Q_2)_{read} + (Q_3 - Q_2)_{fpn} \\ &\approx Q_s + Q_{shot} + (Q_3 - Q_2)_{read}. \end{aligned} \tag{11.10}$$

This operation cancels reset noise, offset FPN and flicker noise but at the expense of doubling read noise power. Since the time between the two reads is short, flicker noise is negligible as its components in the two samples are correlated. For image sensors not employing charge transfer and capable of non-destructive reads (our case), two samples are taken, one right after reset and the second at the end of integration. The signal is estimated as

$$\begin{aligned} S_3 - S_2 &= Q_s + Q_{shot} + (Q_3 - Q_2)_{reset} + (Q_3 - Q_2)_{read} + \\ &\quad (Q_3 - Q_2)_{fpn} + (Q_3 - Q_2)_{1/f} \\ &\approx Q_s + Q_{shot} + (Q_3 - Q_2)_{read} + (Q_3 - Q_2)_{1/f}. \end{aligned} \tag{11.11}$$

This operation cancels reset noise and offset FPN but at the expense of doubling the power of the other noise components. When integration is short, flicker noise is negligible. In our case, however, integration time can be very long, leading to large values of $1/f$ noise as the two samples are quite uncorrelated.

11.3.4.3 Correlated Triple Sampling (CTS)

This scheme was developed for astronomical imaging using CMOS focal plane arrays characterized by long integration times and thus subject to the adversities of flicker noise [35]. It combines the benefits of both delta-reset sampling and CDS by sampling the pixel values four times: during reset, just after reset, at the end of integration and during the following reset [see Fig. 11.13(d)]. The signal estimate is given by

$$\begin{aligned}(S_3 - S_2) - (S_4 - S_1) \quad &= \quad Q_s + Q_{shot} + (Q_1 - Q_4)_{reset} + \\ &\quad (Q_3 + Q_1 - Q_4 - Q_2)_{read}.\end{aligned} \tag{11.12}$$

Note that $(S_3 - S_2)$ corresponds to a CDS operation. The subtraction of $(S_4 - S_1)$ reduces the flicker noise of the readout chain, which is important when circuit observation time is long, as in our application. Read noise power is quadrupled as a result of this operation. Reset noise, however, may not be fully doubled as a result of the low pass filtering by the readout chain.

11.3.4.4 Multiple Correlated Sampling (MCS)

This scheme is a modified version of correlated double sampling [36] and is intended to reduce read noise. The signal is sampled n times at the beginning of integration and at the end of integration and each set of samples is averaged to reduce read noise [see Fig. 11.13(e)]. The signal estimate is given by

$$\overline{S}_3 - \overline{S}_2 \approx Q_s + Q_{shot} + n^{-1/2}(Q_3 - Q_2)_{read} + (Q_3 - Q_2)_{1/f}. \tag{11.13}$$

Note that this scheme does not reduce flicker noise (for non-FD architectures), and as a result is not well suited for applications with long integration times.

11.3.4.5 Correlated Multiple Sampling (CMS)

This is the scheme we use in the operation of our chip. It combines the reset, FPN and flicker noise suppression advantage of CTS with the low read noise of MCS. The pixel value is sampled n times at each of the four locations and each set of samples is averaged [see Fig. 11.13(f)]. The signal is estimated by (when flicker noise is dominant)

$$(\overline{S}_3 - \overline{S}_2) - (\overline{S}_4 - \overline{S}_1) \approx Q_s + Q_{shot} + n^{-1/2}(Q_1 - Q_4)_{reset} + n^{-1/2}(Q_3 + Q_1 - Q_4 - Q_2)_{read}. \quad (11.14)$$

In general, when flicker noise is on the order of the read noise (uncorrelated noise), we can use

$$(\overline{S}_3 - \overline{S}_2) - \kappa(\overline{S}_4 - \overline{S}_1), \quad (11.15)$$

where

$$\kappa = \sigma^2_{1/f} / (\sigma^2_{1/f} + n^{-1}\sigma^2_{read}) \quad (11.16)$$

to ensure we are not unduly penalized by the additional read noise.

Determining which sampling scheme is appropriate for a particular application depends on the capture rate, dynamic range and the sensitivity requirements.

Example 2

Assuming the same image sensor properties in Example 1, we notice that reset noise and FPN are the two major sources of noise. Thus it would be prudent to employ a correlated double sampling scheme. Indeed, performing CDS improves the minimum detectable photosignal by three-fold to 1.7×10^8 photons/sec·cm^2. However, detecting such low signal levels in practice (below 1 mV) may require a high precision ADC with greater than 12 bits of resolution.

In the remaining sections, we outline a methodology for the design of CMOS image sensors for specific biological applications as well as discuss several application specific CMOS imager sensor designs.

11.4 CMOS IMAGE SENSORS FOR MOLECULAR BIOLOGY

The greatest promise of CMOS image sensor technology arises from the ability to customize the sensor front-end and integrate processing with sensing to suit the specific needs of different applications. For example in molecular biology applications, the imaging array can be matched to the assay pitch, high-speed imaging architectures can be used for fluorometry, noise bandwidth can be tailored and signal processing can be performed

on-chip to improve sensitivity or reduce data bandwidth requirements. Even though recent process modifications have made CMOS a viable alternative for low-noise, high-sensitivity applications, it is best suited to applications requiring customization, large-scale integration, low power consumption, very high-speed, and/or extraordinarily large array sizes, and low cost.

To evaluate the feasibility of using CMOS for a particular biological application, it is first necessary to quantify the application requirements. These include:

- *Photon Flux*: This refers both to the range of emission rate and its dynamics over time, that is, the number of photons emitted per second over time. This is important both for determining the required sensitivity and the dynamic range. Some quantitative assays have emission rates as low as tens of photons per second and exhibit a dynamic range of greater than 10^6
- *Wavelength*: This refers to the characteristic emission wavelengths. Visible and near IR wavelengths are the most commonly used in molecular biology applications.
- *Spot size*: This refers to the size of the assay spot or well, typically on the order of 5-150 microns for microarrays.
- *Resolution*: This refers to the size of the spot array in 1 or 2D. In microscopy applications high resolution are needed for imaging small features such as subcellular structure or processes.
- *Capture Rate*: This refers to the required number of image captures per second. Some applications require very short capture rates in the nanosecond range, such as for fluorescence lifetime measurements, while many chemiluminescent chemistries are characterized by long emission times on the order of many tens of minutes with a relatively steady emission rate.
- *Background*: The background signal of the chemistry may be many orders of magnitude larger than the signal level itself. This imposes additional requirements on the image sensor design, in terms of read noise, dynamic range, and ADC resolution.

Once the above requirements are specified for a particular application, one can begin to explore the question of if CMOS can be used. The first step is to evaluate the suitability of commercially available CMOS image sensors. This is quite straightforward and involves determining if basic sensor parameters, such as pixel size, array size, capture rate, ADC resolution, SNR/low light sensitivity, and dynamic range can satisfy the application requirements. If no commercially available sensor can satisfy the require-

ments, the next step is to evaluate the feasibility of a customized solution. Because of the design flexibility provided by CMOS, many system architecture and circuit design options and trade-offs may need to be explored before a conclusion can be reached regarding the feasibility of CMOS.

- What pixel architecture should be used? This depends on well capacity, conversion gain, linearity, need for non-destructive reads, and readout speed.
- Does the array need to be contiguous, or can it be divided into small subarrays? Being able to divide the array into smaller subarrays provides the ability to integrate additional functionality within the array.
- What readout architecture should be used? This depends on array size, readout speed, linearity, resolution, dynamic range, as well as system data acquisition rate.
- What multiple sampling scheme should be used? As explained in Section 11.3, this depends on integration time, noise, and available data acquisition rate.
- How much processing should be done on-chip? How much should be done in analog versus digital? This involves careful analysis of power consumption and noise considerations.

To illustrate this process we first briefly describe some recent examples of customized CMOS image sensors for biological applications. Then in the next section, we detail the work of our group on the design of a highly sensitive CMOS-based detection platform for *de novo* DNA sequencing.

11.4.1 CMOS for Fluorometry

In one method of fluorometry, known as time-gated fluorescence detection, the need for expensive optical filters is potentially eliminated. This is due to the fact that photodetection is performed only immediately after the excitation source is briefly pulsed, thus capturing photoemission only during the period of exponential decay dictated by the fluorophore's lifetime. This mode of detection can potentially decrease system cost, size, and complexity paving the way for a low-cost, handheld fluorometer. However, CCDs are not well-suited to such applications, since ultra-high speed imaging with sub-nanosecond precision is required. CMOS image sensors, on the other hand, built in deep sub-micron semiconductor processes are well equipped to handle such speed requirements. To address these application requirements, the authors in [38] built a CMOS-based biosensor comprising an 8×4 pixel array, 4 sigma-delta ADCs, an SRAM block and a laser driver.

The biosensor performs time-gated fluorometry with a timing precision of 230 psec and sensitivity of 1.15×10^8 photons/cm^2, thus obviating the need for interference filters for a wide variety of fluorophores.

11.4.2 CMOS for Bio-/Chemi-Luminescence

Another potentially important application is the remote monitoring of adverse substances and pathogens in the environment. One technique is through the use of bioreporter cells, which are single-celled organisms engineered to produce bioluminescence in response to specific environmental chemicals or substances. Such cells can be attached to the surface of an image sensor to produce a sort of environmental monitor. Photoemission rates from bioreporter cells are typically very low, so high-sensitivity detection is required. Furthermore, remote monitoring at many node locations is typically desired as well, thus requiring low-cost sensors with low-power consumption, a high level of integration, and output data amenable to wireless transmission.

Once again, such application requirements are ideally suited to the use of a customized CMOS image sensor. To target this application, the authors in [6], [7] built a CMOS-based luminometer comprising a single large area (1.47 mm^2) photodiode coupled to a light-to-frequency ADC for low-level luminescence sensing implemented in a 0.5 μm CMOS process. The light-to-frequency ADC both extended the dynamic range of the sensor as well as allowed for easy integration with an RF transmitter. The sensor was able to detect photocurrents in the femtoamp range, thus detecting relatively low bioluminescence photoemission rates.

11.5 LAB-ON-CHIP FOR *DE NOVO* DNA SEQUENCING

There is a growing need to perform biological testing such as DNA sequencing, immunoassays, and gene expression analyses at the point-of-care and for on site environmental, forensics, and biohazard studies [2],[3]. Clinical platforms for performing such testing today are bulky, high power and require large amounts of reagents, making them ill-suited for such applications. Moreover, the complexity of biological systems makes testing expensive, labor intensive, and time consuming, often requiring the experiments to be repeated several times to achieve the required level of reliability. Consequently, there is a need to develop hand-held, miniaturized, and highly automated testing platforms for molecular biology applications

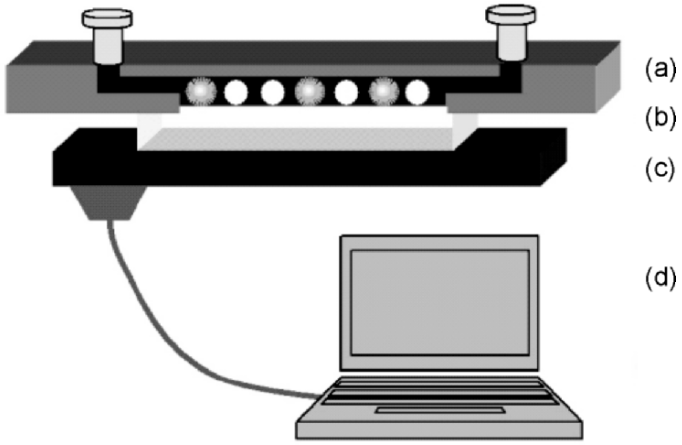


Figure 11.14 Lab-on-chip based de novo sequencing system consisting of a (a) microfluidics substrate for reagent delivery and confinement of the DNA samples directly coupled via a (b) fiber-optic faceplate to a (c) CMOS detection system-on-chip and (d) software for external signal processing.

to enable robust, low-cost point-of-care analysis. To address this need, there have been recent efforts to develop lab-on-chip solutions in which sampling, preparation, analysis and reporting of a wide range of chemical compounds are integrated and automated [39]-[42]. In addition to reducing system size, cost, and power consumption, such lab-on-chips can increase system throughput, thus reducing testing time and labor.

Because of the size and cost constraints of a lab-on-chip, it is also necessary to reduce the amounts of reagents used. Such reduction, however, decreases the levels of the biochemical signals generated, and special efforts must be devoted to enhancing detection sensitivity [8],[43, 44]. It should be noted that most chemiluminescence detection systems used in the biological field today perform detection using cooled CCD-based camera systems. Cooling, which is used to reduce detector dark current to enhance its sensitivity, is costly and consumes too much power, making miniaturization of these systems difficult. The use of intermediary optics also makes miniaturization difficult. More specifically, lower-end CCD-based systems employ fairly simple optics, incurring significant optical loss in the imaging system and hence large amounts of reagents are needed for adequate detection. In contrast, high-end imaging systems avoid this loss by using multi-piece optical systems or custom-fitted fiber optic components, which makes them quite costly and bulky.

CMOS image sensors offer the promise of achieving high sensitivity in a portable lab-on-chip format, especially when coupled to a MEMS-based

microfluidic platform for reagent delivery and analyte confinement. We developed such a CMOS lab-on-chip for *de novo* DNA sequencing in cooperation with the Stanford Genome Technology Center and the Stanford Center for Integrated Systems [45]. The overall system consists of a microfluidics substrate for reagent delivery and confinement of the DNA samples directly coupled via a fiber-optic faceplate to a CMOS chemiluminescence detection system-on-chip (SoC). The data from the SoC is sent to a PC to perform the processing needed for base-calling (see Fig. 11.14). Direct coupling is used to improve optical efficiency and reduce cost by eliminating the optics [46]-[47]. The challenge with this approach is the stringent design constraints arising from the need to operate the detection system near room temperature (to ensure reasonable reactivity of the applied chemistry). Such a detection system needs to achieve high sensitivity despite the presence of relatively high dark currents.

11.5.1 Lab-on-Chip Application Requirements

The range of possible applications for our lab-on-chip are not limited to DNA sequencing, but include on-site pathogen detection, gene expression using reporter gene assays, and determination of intracellular ATP for food contamination testing to name a few. These applications require moderate sized assay arrays with approximately 100 sites, each site of a relatively large area of around 150 μm × 150 μm. This site size is compatible with existing microarray spotting equipment, microfabricated microfluidic systems for reagent delivery and recycling as well as manual spotting. The emission times in these applications are long, ranging from 1 sec to over 30 sec. The photoemission rates can be extremely low, with rates down to 4×10^5 photons/cm^{2}sec (or 10^{-6} lux) over a period of around 30 sec. Finally, analyte concentrations can easily exceed 5 orders of magnitude, i.e. 100 dB of dynamic range.

Since we directly couple the assay to the luminescence detection system, it must achieve these stringent sensitivity and dynamic range requirements at room temperature, i.e. with relatively high dark current. This is the case since most luminescence chemistries degrade at low temperatures. It also means the system must be low power as well to ensure the reagents are not overheated.

Unfortunately, commercially available off-the-shelf CCD and CMOS imagers cannot meet the aforementioned sensitivity and dynamic range requirements in part due to their small pixel sizes and high dark current at

room temperature [48]. In the next section we detail the detection SoC that we have designed to meet the above goals.

11.5.2 Luminescence Detection System-on-Chip

A block diagram of the detection SoC, which is fabricated in a 0.18 μm CMOS image sensor process (CIS) [22], is given in Fig. 11.15(a). It comprises three main sections, an 8 × 16 pseudo-differential pixel array, a 128 channel 13-bit 2-step ADC, and a column-level DSP single-instruction, multiple-data (SIMD) array. It contains 492,000 transistors on a 5 × 5 mm^2 die. Each pseudo-differential pixel is connected to its own ADC and all ADCs operate simultaneously with no multiplexing or switching between pixels. An on-chip ramp generator outputs a differential signal that is routed globally to all ADCs. Each column DSP comprises 8 20-bit loadable shift registers, a 20-bit adder/subtractor, data registers for intermediate storage as well as for the 4-bit DAC values, and a 64 × 20 SRAM register file. Each DSP can perform read/write/add/subtract/accumulate/shift operations on the ADC data and operates at 40 MHz. Accordingly, each pixel can be thought of as a self-contained unit, with its own ADC and 9 20-bit storage registers and access to a shared column-level processor. The on-chip memory is required for offset/background correction and multiple signal sampling without the need for off-chip data transfers during the acquisition period, thereby keeping on-chip disturbances to a minimum. Multiple sampling is used for dynamic range enhancement and for noise reduction to enhance low-light signal detection. To facilitate off-chip readout and provide for testability, the 16 column-level 20-bit DSP shift registers are strung together into 8 320-bit

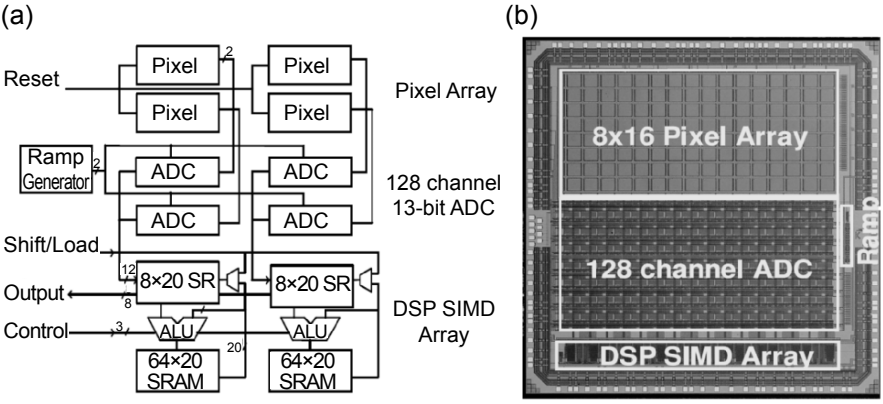


Figure 11.15 (a) Detection SoC block diagram. (b) Detection SoC photomicrograph.

Table 11.2 Overall chip characteristics.

Process	0.18 μm 4-metal CMOS
Die size	5 mm × 5 mm
Array size	8 × 16 pixels
Number of transistors	492K
Readout Architecture	8-bit (40 MHz)
Readout rate	320 Mbit/s
Pixel size	240 μm × 210 μm
Photodetector type	P+/N/Psub
ADC Per-pixel	2-step 13-bit ADC
DSP type	SIMD
Static power	26 mW

shift registers. The periphery contains various biasing circuitry, clock generators, and a globally routed counter for ADC operation.

A photomicrograph of the chip is shown in Fig. 11.15(b) and the main chip characteristics are listed in Table 11.2. Special care was taken in chip floor planning, layout of the critical cells, and chip operation to avoid noise injection from the digital portion to the analog and photodetector array sections. The pixel array and ADC sections each use separate supply voltage networks, while the SIMD array section has a separate pad ring. 1.5 nF of decoupling capacitors are integrated on chip. Switching lines are kept away from sensitive nodes. The ADC and the clock signal are powered down during integration. In the digital portion, adjacent column-level SRAM register files are consolidated to reduce overhead and conserve power. For speed and power considerations the SRAM blocks operate on a different clock phase than the adders and shift registers. Furthermore, the clock inputs to each SRAM block are phase-shifted from other blocks to minimize the peak transient current during memory access across the chip. To minimize digital switching and resulting injected substrate currents, a pulse generation block is used so that the global counter value is latched in each pixel's respective shift register only once during analog-to-digital conversion. The digital section of the array is completely testable, with an input port for scanning in test or calibration data that can be written into any of the registers including the SRAM.

A test chip was designed and fabricated in order to ensure that each of the individual components would meet the targeted noise, linearity and functionality requirements. Furthermore, the test chip allowed us to characterize the pinned photodiode specifications in this fairly new 0.18 μm CIS process

and fine-tune the pixel accordingly. In the following sections, we discuss details of the pixel and ADC design.

11.5.2.1 Pseudo-differential Pixel

To achieve low read noise and some immunity to common-mode disturbances, a pseudo-differential pixel and fully differential ADC architecture is used [see Fig. 11.16(a)]. The 240 μm × 210 μm pixel comprises a P+/N/Psub photodiode, an NMOS reset gate and two PMOS source follower amplifiers, one connected to the photodiode and the other connected to a global reference signal VREF. The P+ implant used in the photodiode structure helps passivate surface-state traps to prevent the formation of interface states on the photodiode surface, leading to a very low dark current of 2.6 mV/s. Several annealing steps are performed to ensure net leakage current is not adversely affected due to defect states created by the P+ implant. The P+/N/Psub doping profile results in very low photodiode junction capacitance, which is desirable for achieving high conversion gain. This low capacitance results from the fully depleted nature of the intermediate N-region even at very low bias voltages. The electric field profile is such that electrons are continually swept from the photodiode junction towards the N+ diffusion of the reset transistor, i.e. no transfer gate is used. To achieve acceptable well capacity at low swing (as well as acceptable linearity) as a result of the low diode capacitance, we added Metal 1 capacitance to the photodiode node. Although sensitivity is somewhat decreased by the resultant reduction in conversion gain, the additional well capacity is essential for chemistries with high background luminescence in order to achieve acceptable dynamic range.

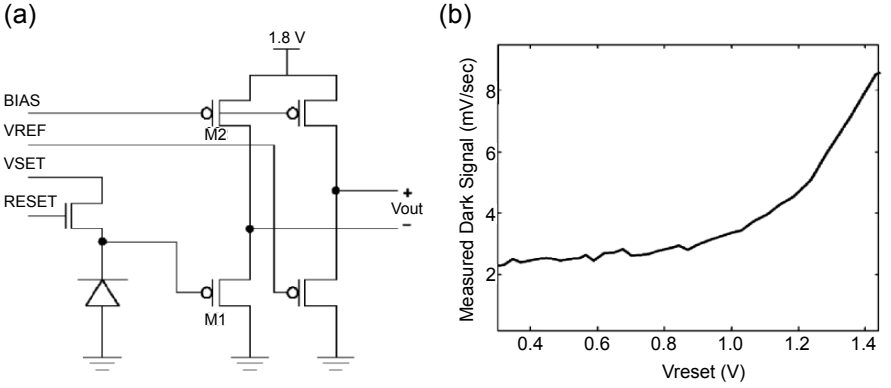


Figure 11.16 (a) Pseudo-differential Pixel. (b) Dark signal versus reset voltage Vreset.

Table 11.3 Pixel characterization summary.

Photodetector area Voltage	150 µm × 150 µm
Range	0.1-0.9 V
Dark Signal	2.6 mV/s
QE @ 560 nm	0.4
Sensitivity	160 V/lux.sec
Capacitance	150 fF
Follower non-linearity	0.10%
Input-referred SNR	58 dB
Dynamic Range	61 dB

A PMOS follower instead of the NMOS follower more commonly used in Active-Pixel Sensors is employed to further reduce dark current and to achieve high linearity. Figure 11.16(b) plots dark current as a function of reset voltage. As can be seen, dark current increases rapidly as the reset voltage is increased above 1 V. The use of a PMOS follower makes it possible to operate with a reset voltage of around 1 V or below. High follower linearity is achieved by tying the body of the PMOS transistors to their sources to cancel the body effect.

Special considerations were taken for the pixel layout. Diode and PMOS N-well corners are angled to minimize injected charge carriers from high radius of curvature implants. Each diode is surrounded by 35 microns of Metal 4 shielding to minimize crosstalk between adjacent pixels. Additionally, the entire pixel array is surrounded by an N-implant guard ring to isolate the array from substrate noise disturbances. The optical and electrical characteristics of the pixel are summarized in Table 11.3.

11.5.2.2 Two-step ADC Design and Characterization

One of the key decisions made in the design of this image sensor was the choice of the ADC architecture. As discussed earlier, the chip must be capable of detecting less than 10^{-6} lux over 30 sec integration time at room temperature. The challenge here is the large background due to the integration of dark current as well as non-negligible background light due to the chemistry for over 30 sec time. As an example, we use the hypothetical, worst-case photodetector size, QE, conversion gain, background current density, and voltage swing values listed in Table 11.4. For this particular case, we can quantify the design challenge as detecting a signal charge of 10 ke^- amidst a background charge of 1.3 Me^- with shot and readout circuit rms noise of around 1.5 ke^-. To achieve such low detection with an acceptable

Table 11.4 Worst case photodetector model.

Pixel size	240 μm × 210 μm
Photodetector area	150 μm × 150 μm
Voltage Swing	1 V
Dark Signal	1 nA/cm^2
QE @ 560 nm	0.4
Capacitance	400 fF

SNR we need to choose our ADC step size to be less than 100e$^-$. Since well capacity is 2.9 Me$^-$ in this example, this implies that the ADC should have around 15 bits of resolution. As such resolution is difficult to attain in the standard 0.18 μm CMOS process used (due to both mismatch and noise), we decided to digitize the signal multiple times using a lower resolution ADC and then average the results to achieve an acceptable SNR. Figure 11.17 shows a Monte-Carlo simulation result plotting the corresponding SNR as a function of ADC resolution for different numbers of averaged samples. Guided by such analysis, we decided on 13-bit ADC resolution with an on-chip averaging capability of at least 128 samples.

To perform the 13-bit A/D conversions and averaging for the entire pixel array at a fast enough speed as dictated by the dynamic range requirements requires a total conversion rate of at least 6 Mpixels/s to ensure the signal

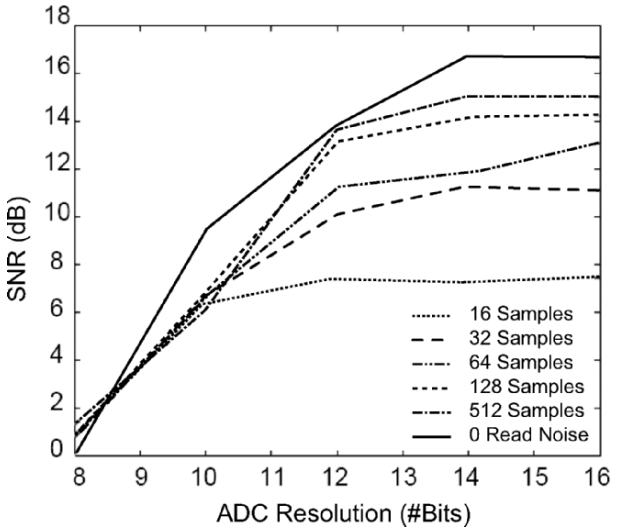


Figure 11.17 SNR versus ADC resolution for several values of sample averaging (16, 32, 64, 128, 512 and 0 read noise).

Table 11.5 Nyquist ADC architectures.

	Flash	Pipeline	Single Slope	Algorithmic
Conversion rate	Very High	High	Moderate	Low
Area	Large	Large	Small	Small
Power	High	High	Low	Low
Resolution	8	10-14	8-11	12-16

estimate is not corrupted by the non-zero aperture time. This can be performed using a single ADC for the entire array, an array of per-column ADCs, or an array of per-pixel ADC. We decided to use a per-pixel architecture for several reasons: (i) to avoid the use of sample-and-hold capacitors and the analog multiplexing needed in the other options, (ii) to provide scalability with array size, (iii) to reduce the circuit design requirements, and (iv) to provide low power consumption [49]. We decided to use a Nyquist rate ADC architecture instead of an oversampled ADC, since a very high oversampling rate would be required to achieve such high resolution at low signal levels. Such high clock rates could have an adverse effect on the photodiode array due to switching induced substrate noise. Successive approximation ADCs were also ruled out due to the fact that mismatch errors can cause adjacent output codes to not have a uniform (or near uniform) distribution, that is, the codes are not equally likely under typical circuit mismatch conditions [50]. As such, averaging is sub-optimal for this type of ADC. Table 11.5 provides a comparison of existing Nyquist rate ADC architectures based on their area, speed and typical resolution. From the table, it is clear that no single architecture combines the small area needed for per-pixel implementation at the required speed and resolution.

With the above considerations in mind, a 13-bit pixel-level 2 step single-slope ADC was chosen, which meets the resolution, speed and averaging requirements. The conceptual operation of the per-pixel 2-step ADC used in the chip is illustrated in Fig. 11.18(a). A coarse 4-bit conversion is first performed. The output of the 4-bit DAC is then subtracted from the pixel output, amplified by a factor of 16 and digitized to 9-bits. Single slope A/D conversion is performed in each step using a differential ramp signal and a digital ramp sequence that are generated on-chip and globally distributed to the comparators and column-level shift registers, respectively. When a comparator output flips, the corresponding digital code is latched into the shift register. The resulting 13-bit digital values are subsequently stored in the DSP SRAM. In reality, the 9-bit conversion contains more than 512 codes to provide enough overlap between the 16 (4-bits) DAC regions to ensure

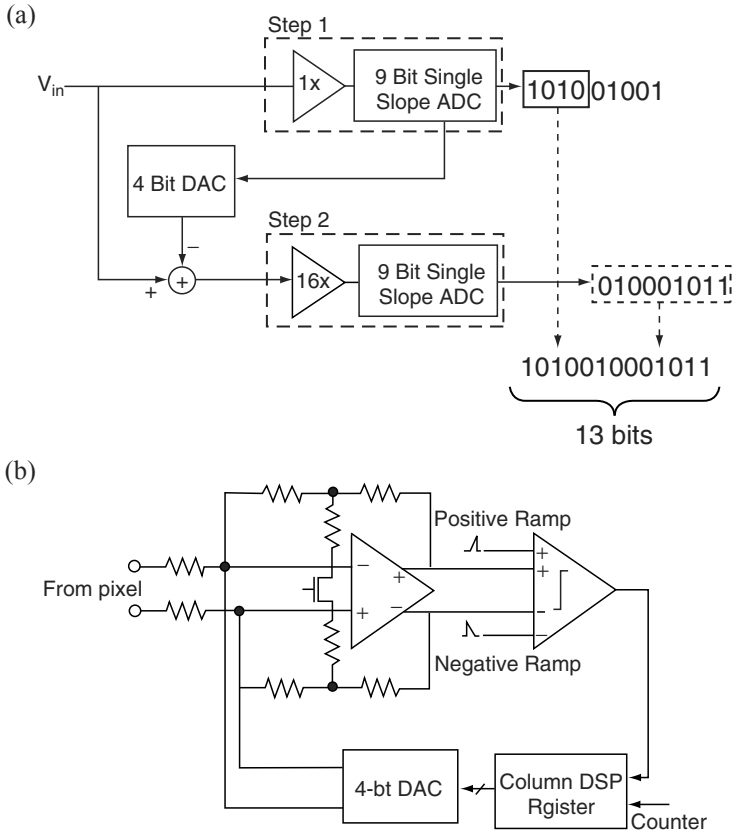


Figure 11.18 (a) 2-step 13-bit ADC principle of operation (b) 2-step 13-bit circuit block diagram.

successful conversion despite circuit mismatches across the array and to provide enough signal drift margin for subsequent averaging operations.

A fully differential ADC architecture is employed to provide high power supply and common-mode noise rejection and double the dynamic range of the system. Each pixel's pseudo-differential output is connected to a separate ADC channel consisting of a variable gain amplifier (VGA), a 1-bit differential comparator and a 4-bit DAC as shown Fig. 11.18(b). The VGA, comparator, DAC and DSP RAM circuits can be electrically tested by applying analog signals to the sense node through the "Reset Voltage" signal, performing A/D conversion using the on-chip differential ramp signal generator and counter and then reading out the digitized values. In this way, except for the photodetectors, the chip can be electrically tested and characterized without the need for light or optics. We now discuss each of the various components of the 2-step ADC.

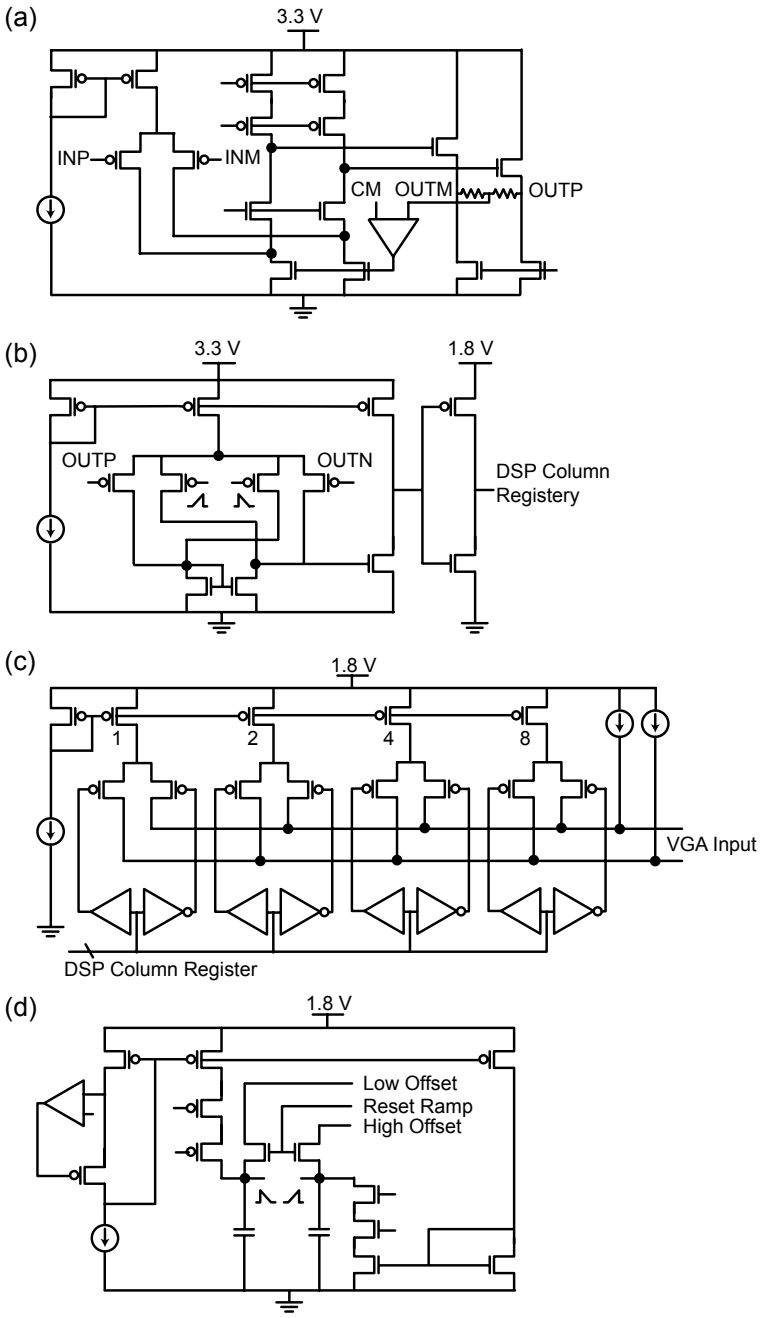


Figure 11.19 Schematics of the various ADC components. (a) Variable-gain amplifier. (b) Differential-difference amplifier. (c) Current-steering DAC. (d) Ramp generator.

Table 11.6 Variable Gain Amplifier Characterization Summary.

Architecture	Resistive T Network
Gain	2×, 30× (differential)
Resistor type	Non-silicided P+ poly
Resistor size	6×500 kΩ, 2×30 kΩ
Linearity	11 bits
Open loop DC gain	109.5 dB
BW	3.5 MHz

The VGA employs a fully differential amplifier with resistive feedback T-network [see Fig. 11.19(a)]. The differential gain of the VGA can be switched between 2 and 30 (and to any value in between) via switch transistor M. Resistive feedback is used instead of a switched capacitor design to eliminate the adverse effects of clocking and switching and to save area. The resistors are implemented using non-silicided P^+ poly. A PMOS folded cascode topology with a low impedance output stage and amplifier-based common-mode feedback is used. The amplifier DC gain and bandwidth are designed for near resistor limited linearity and low noise. The VGA was characterized independently using the test chip and the results are summarized in Table 11.6. Its linear output voltage swing range is 4 volts differential with a measured linearity of over 11 bits. In the design of the VGA, we paid special attention to $1/f$ noise due to the long integration time, which can be as large as 3 min. The total noise injected is about 26 μV (0.26 LSB) which is within our design tolerance.

The 1-bit comparator shown in Fig. 11.18(b) is based on a differential difference amplifier (DDA) comprising two parallel PMOS differential pairs followed by an inverting gain stage [51]. The DDA is characterized by the following equation, $V_o = A[(V_1 - V_2) - (V_3 - V_4)]$.

For high gain open loop operation the DDA voltage output flips when the differential ramp signal exceeds the differential VGA output. This signal is level shifted from 3.3 V to 1.8 V and is used to latch a counter value that is synchronized with the ramp signal in the column DSP register. The DDA has a DC gain of 87.6 dB and a switching time (from 0 to 1.8 V) of 50 nsec.

The 4-bit DAC uses a differential current steering architecture as shown in Fig. 11.19(c). Since the DAC is reset after each ADC and not stepped, we are able to save area by using a binary weighted instead of a thermometer coded architecture. The DAC uses a 1.8 V power supply and occupies 40% of the ADC area in order to achieve the accuracy needed, since this is the

Table 11.7 ADC Characterization Summary.

Architecture	128 channel 2-step
Resolution	13-bit
Area	7 mm^2
Conversion Rate	7.1 Msamples/sec
INL	< 2.5 LSB
DNL	< 50 μV (0.5 LSB)
FOM = Power/(2^{b+1}BW)	1.2 pJ

most critical component in terms of ADC linearity. It also features two trimming current sources in order to compensate for any offset variations.

The on-chip differential ramp generator circuit is shown in Fig. 11.19(d). The circuit integrates an externally adjustable constant bias current over 2 on-chip 5 pF MiM capacitors. The charging and discharging of the two capacitors are synchronized and triggered by a single reset signal. The timing error between the ramps is negligible for the conversion times used (i.e. 25 μsec). Furthermore, integral linearity of the ramp differential ramp signal is greater than 10 bits. The circuit has a differential output that is buffered and routed globally to all ADCs.

The effective performance of the 128 channel 13-bit 2-step ADC is summarized in Table 11.7.

11.5.3 Low Light Detection

To achieve high detection sensitivity, efforts must be devoted to minimizing both the light loss in the optical path and the read noise. We achieved both goals by a combination of the following techniques:

- The loss in the optical path is minimized by direct coupling of the luminescent assay array to the photodetector array as well as by array pitch matching. Pitch matching ensures minimal loss of sensitivity due to less than unity pixel fill-factors as well as only one added read noise component per assay signal read.
- The use of a very low dark current photodiode reduces background noise and thus increases signal detectability.
- Background subtraction is performed by acquiring several dark frames and averaging them prior to acquiring the signal frame. This removes the mean value of dark current without introducing too much additional shot noise.

- Low noise circuits are used with virtually no multiplexing or switching during integration.
- Fully differential circuits are used to enhance noise performance and provide immunity to common-mode disturbances.
- Correlated multiple sampling with multiple capture is used to suppress offsets, reset noise, and flicker noise as well as increase dynamic range as explained in detail in the next subsection.
- Repeated digitizations and averaging are performed by the ADC and SIMD DSP to significantly reduce the uncorrelated readout noise components.

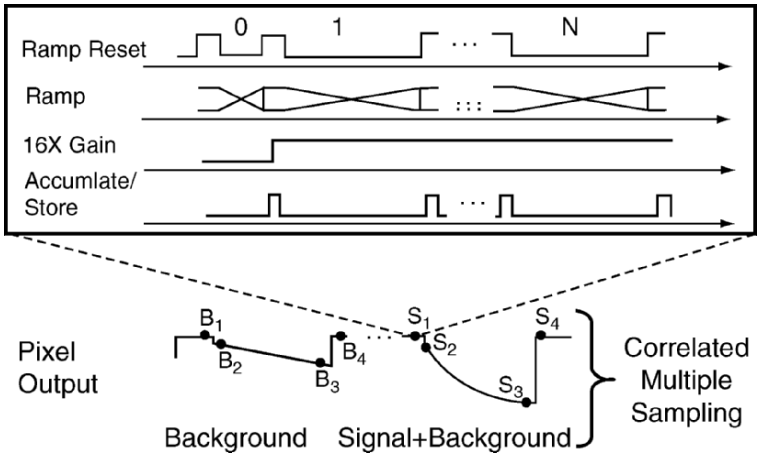


Figure 11.20 Detection SoC timing diagram.

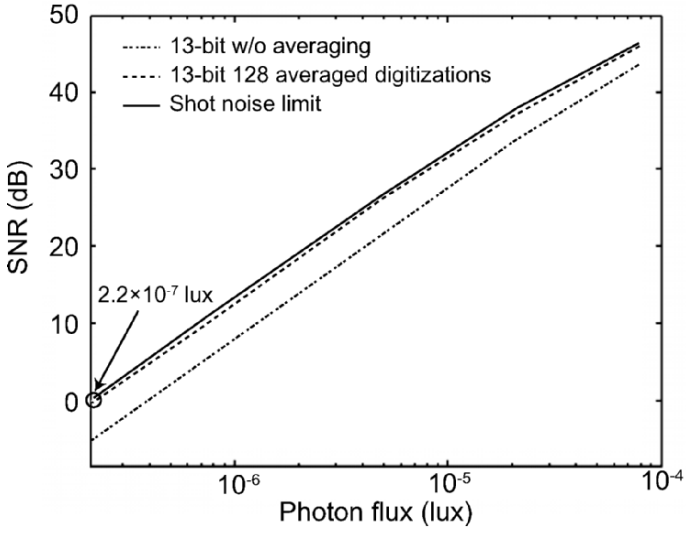


Figure 11.21 Simulated luminescence detection SoC SNR.

Table 11.8 Sampling Schemes Experimental Results.

Technique	Offset	Reset	Flicker	Read	Loss
Delta-Reset			✓	2×	–15.44 dB
Correlated Double Sampling	✓	✓		2×	–6.94 dB
Correlated Triple Sampling	✓		✓	4×	–9.42 dB
Multiple Correlated Sampling	✓	✓		2×/N	–2.05 dB
Correlated Multiple Sampling	✓	✓	✓	4×/N	– 1.12 dB

The chip operation is illustrated in Fig. 11.20. Several dark frames having the same integration time as will be used for signal detection are first captured to accurately estimate the background signal. Light emission is then triggered by the addition of a chemical reagent and the signal frame is captured. The chip is capable of implementing different types of sampling schemes to suppress offset, reset noise, and flicker noise [see Fig. 11.21] and to increase dynamic range (DR). In the following, we compare the sampling techniques highlighted in Section 11.3 and demonstrate the benefits of Correlated Multiple Sampling (CMS).

In order to evaluate the performance of the CMS technique we performed the following experiment. We reset the photodiode and allow the detector to integrate the dark current for 80 sec. We sample the signal at 4 different times during reset, just after reset and at the end of integration. At each time, a sample consists of 16 reads at 25 μsec intervals which are averaged to reduce read noise. We compare different schemes by calculating how far they are from the shot noise (of both the signal and background) limited case and refer to that as the loss term shown in Table 11.8. Note that using CMS, we are only 1.12 dB away from the shot-noise limited case.

The added advantage that CMS combined with our imaging architecture provides over MCS employed by charge transfer imaging devices, is that dynamic range can be significantly improved through multiple capture without loss of low light sensitivity. Currently in practice, for chemistries with over 100 dB of DR, CCD-based schemes require at least two sets of experiments to be performed in order to match the dynamic range required since non-destructive reads are not feasible. This is problematic, as the chemical and temporal variations from one experiment to another can degrade the accuracy of the test results. In contrast, employing CMS with multiple capture allows us to extend the DR almost arbitrarily without compromising signal fidelity for the weakest signals. For example, although the intrinsic DR of

our sensor is 61 dB for a 30 sec integration time, by interspersing additional captures within the exposure period we can increase DR by t_{int}/t_{min}, where t_{int} is the total integration time and t_{min} is the time of the first signal capture. For our sensor, this increase (for t_{int} = 30 sec) can be as large 6×10^5 or 266 dB. For interested readers, the trade-off between the number and positions of captures versus SNR and DR is explored in more detail in [30]. We should note that the multiple sample requirement (for read noise averaging) also does not limit the high light performance, since we have the luxury of choosing to read only once for intermediate samples. This can be done since intermediate samples are taken to extend dynamic range at the high end and these samples accordingly have relatively high SNR.

In Fig. 11.21, we plot the SNR versus the incident photon flux in lux for 30 sec integration (without multiple capture) for (i) single 13-bit A/D conversion, (ii) averaged 128 13-bit A/D conversions, and (iii) optimal shot noise limited performance. Note that the SoC performance exceeds the 10^{-6} lux detection requirement and is close to optimal and that averaging provides around 5.3 dB improvement over single 13-bit A/D conversion.

11.5.4 Applications

One of the several bioluminescence assays that we characterized using our lab-on-chip is Pyrosequencing [52, 53], which is a recently developed DNA sequencing method based on the detection of released inorganic pyrophosphate during DNA synthesis shown in Fig. 11.22, where APS is adenosine phosphosulfate, dNTP is deoxynucleotide triphosphate, and PPi is pyrophosphate. Using a linked enzymatic reaction, visible light (λ = 562 nm) is generated that is proportional to the number of incorporated nucleotides.

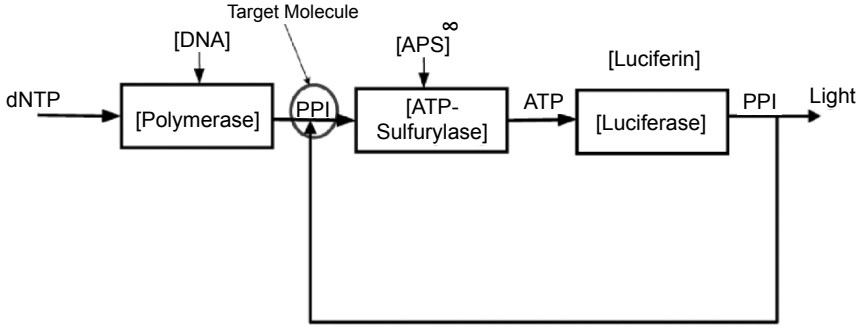


Figure 11.22 Block diagram of Pyrosequencing in which PPi release from DNA polymerization is measured via an enzymatic cascade consisting of luciferase and resulting in photoemission.

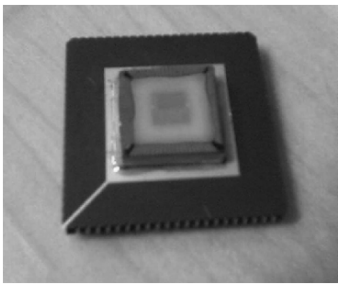


Figure 11.23 Luminescence detection lab-on-chip: a fiber optic faceplate is coupled to the detection SoC.

Consequently, cycling through the addition of the four nucleotides results in serial readout of the target DNA strand.

The enzymatic cascade starts with a DNA polymerization reaction in which inorganic pyrophosphate (PPi) is released as a result of nucleotide incorporation by polymerase. The released PPi is subsequently converted to ATP by ATP sulfurylase. The synthesized ATP provides the energy for luciferase to generate photons. Unincorporated deoxy-nucleotides and ATP are degraded by the enzyme apyrase to chemically reset the enzymatic system after the incorporation test. The enzymatic reaction is given by

$$\begin{aligned} &\mathrm{ATP} + \mathrm{luciferin} + \mathrm{O_2} \\ &\quad \xrightarrow{luciferase} \mathrm{AMP} + \mathrm{oxyluciferin} + \mathrm{CO_2} + \mathrm{PPi} + h\upsilon, \end{aligned} \tag{11.17}$$

where AMP is adenosine monophosphate. When product inhibition is negligible, the rate of ATP consumption is given by [48]

$$\frac{d[\mathrm{ATP}]}{dt} = -k_L[E][\mathrm{luciferin}_{\max}][\mathrm{ATP}_0] \cdot \exp\left\{-k_f[E][\mathrm{luciferin}_{\max}]t\right\}. \tag{11.18}$$

The quantum efficiency of luciferase is about 0.88, and the turnover rate of the enzyme, depending on the ratio of luciferase to ATP, can vary from 0.1 to 1 sec^{-1} (glow compared to flash). In the glow process, the light intensity is approximated by

$$N_{ph}(t) = (\alpha VA) \cdot k_t \cdot [\mathrm{ATP}_0] \cdot \exp[-k_t t]. \tag{11.19}$$

We tested the light generation portion of the Pyrosequencing reaction by adding 200 fmole of ATP to a 10 μL reaction well containing an enzyme-substrate mixture of luciferase, luciferin, APS and ATP sulfurylase to stimulate light generation. The area of the well was just slightly larger

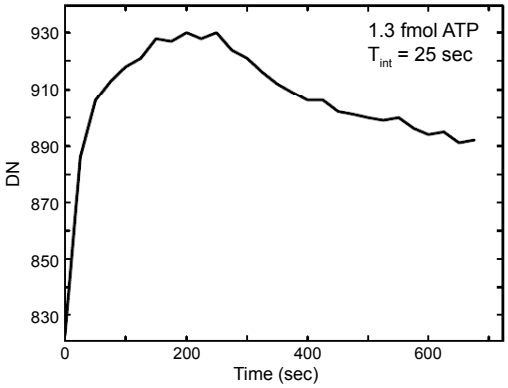


Figure 11.24 Measured light signal using detection SoC for a Pyrosequencing reaction.

than that of the photodetector portion of the detection SoC. The well itself was affixed to the surface of a fiber optic faceplate that is directly coupled to the SoC using an index-matching oil and peripheral epoxy as a sealant (see Fig. 11.23). Figure 11.24 plots the resultant ADC value against time for a single pixel. We were able to detect reliably 1.3 fmole of ATP which is two orders of magnitude lower that the current state of the art Pyrosequencing machines [53]. Furthermore, the signal to noise ratio was high enough, such that no background subtraction or read noise reduction was needed. This implies that detection of even lower amounts of ATP, well into the attomole range, is readily feasible with this system. Indeed, this is on par with commercially available PMT-based luminometers (e.g. the Beckman Coulter LD400 with a 50 amol ATP detection sensitivity [54]).

ACKNOWLEDGMENT

The authors thank M. Ronaghi, S. Kavusi, K. Ghosh, and A. Ercan for their valuable insights and helpful comments.

REFERENCES

[1] T. Massoud and S. Gambhir, “Molecular imaging in living subjects: seeing fundamental biological processes in a new light,” Genes & Dev. vol. 17, pp. 545-590, 2003.

[2] D.M. Wilson, S. Hoyt, J. Janata, L. Abando, and K. Booksh, "Chemical Sensors for Portable, Handheld Field Instruments," IEEE Sensors Journal, vol. 1, No. 4, pp. 256-276, 2001.

[3] R. Davies, D. Bartholomeusz and J. Andrade, "Personal Sensors for the Diagnosis and Management of Metabolic Disorders,"IEEE Engineering in Medicine and Biology Magazine, vol. 22, No. 1, pp. 32-42, 2003.

[4] U. Lu , B. Hu , Y. Shih , C. Wu and Y. Yang, "The design of a novel complementary metal oxide semiconductor detection system for biochemical luminescence," Biosensors and Bioelectronics, vol. 19, No. 10, pp. 1185-1191, 2004.

[5] S. Huang, Y. Shih, C. Wu, C. Yuan, Y. Yang, Y. Li and T. Wu, "Detection of serum uric acid using the optical polymeric enzyme biochip system," Biosensors and Bioelectronics, vol. 19, No. 12, pp. 1627-1633, 2004.

[6] M.L. Simpson, G.S. Sayler, B.M. Applegate, S. Ripp, D.E. Nivens, M.J. Paulus and G.E. Jellison, "Bioluminescent bioreporter integrated circuits form novel whole-cell biosensors," Trends in Biotechnology, vol. 16, pp. 332-333, 1998.

[7] E.K. Bolton, G.S. Sayler, D.E. Nivens, J.M. Rochelle, S. Ripp, M.L. Simpson, "Integrated CMOS photodetectors and signal processing for very low-level chemical sensing with the bioluminescent bioreporter integrated circuit," Sensors and Actuators B: Chemical, vol. 85, pp. 179-185, June 2002.

[8] J.P. Golden and F.S. Ligler, "A comparison of imaging methods for use in an array biosensor," Biosensors and Bioelectronics, vol. 17, No. 9, pp. 719-725, 2002.

[9] N. Van Dyke, C. Van Dyke, K. Woodfork, (Eds.), 2002. Luminescence Biotechnology: Instruments and Applications. CEC Press, Boca Raton, New York.

[10] X. Michalet, "The Power and Prospects of Fluorescence Microscopies and Spectroscopies," Annual Review of Biophysics Biomolecular Structures, vol. 32, pp. 161-182, 2003.

[11] L. Brovko, O. Gandel'man, T. Polenova, N. Ugarova, "Kinetics of Bioluminescence in the Firefly Luciferin-Luciferase System. Biochemistry," vol. 59, pp. 195-201, 1994.

[12] L.J. Kricka, "Clinical and biological applications of luciferases and luciferins," Analytical Biochemistry, vol. 175, pp. 14-22, 1988.

[13] M. Oda, M, T. Kaida, S. Izawa, T. Ogo, K. Itsumi, Y. Okada, K. Sasada, "A 1/4.5in 3.1M pixel FT-CCD with 1.56 μm pixel size for mobile applications," ISSCC Digest of Technical Papers, pp. 346-347, 2005.

[14] A. Krymski, D. Van Blerkom, A. Andersson, N Block, B. Mansoorian and E.R. Fossum, "A High Speed, 500 Frames/s, 1024 ×1024 CMOS Active Pixel Sensor", Proceedings of the 1999 Symposium on VLSI Circuits, pp. 137-138, June 1999.

[15] N. Stevanovic, M. Hillegrand, B.J. Hostica, and A. Teuner, "A CMOS Image Sensor for High Speed Imaging," ISSCC Digest of Technical Papers, vol. 43, pp. 104-105, February 2000.

[16] S. Kleinfelder, S.H. Lim, X.Q. Liu and A. El Gamal, "A 10,000 Frames/s 0.18 µm CMOS Digital Pixel Sensor with Pixel-Level Memory," ISSCC Digest of Technical Papers, pp. 88-89, San Francisco, CA, February 2001.

[17] B. C. Burkey, *et al.*, "The Pinned Photodiode for an Interline Transfer CCD Imager," Proceedings of IEDM, pp. 28-31, Dec. 1984.

[18] X. Liu, P. Catrysse, and A. El Gamal, "QE Reduction due to Pixel Vignetting in CMOS Image Sensors," In Proceedings of the SPIE Electronic Imaging 2000 conference, vol. 3965, San Jose, CA, January 2000.

[19] H. Takahashi, M. Kinoshita, K. Morita, T. Shirai, T. Sato, T. Kimura, H. Yuzurihara, S. Inoue, "A 3.9 µm Pixel Pitch VGA Format 10b Digital Image Sensor with 1.5-Transistor/Pixel," ISSCC Digest of Technical Papers, vol. 47, pp. 108-109, February 2004.

[20] T.H. Hsu, Y.K. Fang, D.N. Yaung, S.G. Wuu, H.C. Chien, C.S. Wang, J.S. Lin, C.H. Tseng, S.F. Chen, C.S. Lin, C.Y. Lin, "Dramatic Reduction of Optical Crosstalk in Deep-Submicrometer CMOS Imager With Air Gap Guard Ring," Electron Device Letters, vol. 25, pp. 375-377, June 2004.

[21] N.S. Saks, "A Technique for Suppressing Dark Current Generated by Interface States in Buried Channel CCD Imagers," IEEE Electron Device Letters, EDL-1, No. 7, pp. 131-133, July 1980.

[22] H.C. Chien *et al*., "Active Pixel Image Sensor Scale Down in 0.18¹m CMOS Technology," IEDM Technical Digest, pp. 813-816, 2002.

[23] S. Ohba, M. Nakai, H. Ando, S. Hanamura, S. Shimda, K. Satoh, K. Takahashi, M. Kubo, T. Fujita, "MOS Area Sensor: Part II – Low-Noise MOS Area Sensor with Antiblooming Photodiodes," JSSC, vol. 15, pp. 747-752, Aug. 1980.

[24] S. J. Decker, R. D. McGrath, K. Brehmer, and C. G. Sodini, "A 256×256 CMOS imaging array with wide dynamic range pixels and column-parallel-digital output," IEEE Journal of Solid State Circuits, vol. 33, pp. 2081-2091, December 1998.

[25] M. Mori, M. Katsuno, S. Kasuga, T. Murata, T. Yamaguchi, "A 1/4in 2M Pixel CMOS Image Sensor with 1.75 Transistor/Pixel," ISSCC Digest of Technical Papers, vol. 47, pp. 110-111, February 2004.

[26] D. Yang, A. El Gamal, B. Fowler, and H. Tian, "A 640 × 512 CMOS Image Sensor with Ultrawide Dynamic Range Floating-Point Pixel-Level ADC," Journal of Solid State Circuits, vol. 34, No. 12, pp. 1821-1834, December 1999.

[27] W. Bidermann, A. El Gamal, S. Ewedemi, J. Reyneri, H. Tian, D. Wile, and D. Yang, "A 0.18 µm high dynamic range NTSC/PAL imaging system-on-chip with embedded DRAM frame buffer," ISSCC Digest of Technical Papers, pp. 212-213, February 2003.

[28] D. Stoppa, A. Simoni, L. Gonzo, M. Gottardi, and G.-F. Dalla Betta, "A 138 dB dynamic range CMOS image sensor with new pixel architecture," ISSCC Digest of Technical Papers, vol. 45, pp. 40-41, February 2002.
[29] L. McIlrath, "A low-power low-noise ultrawide-dynamic-range CMOS imager with pixel-parallel A/D conversion," IEEE Journal of Solid-State Circuits, vol. 36, pp. 846-853, May 2001.
[30] David Yang, and Abbas El Gamal, "Comparative Analysis of SNR for Image Sensors with Enhanced Dynamic Range," Proceedings of SPIE, vol. 3649, pp. 197-211, San Jose, CA, January 25-26, 1999.
[31] S. Kavusi and A. El Gamal, "A Quantitative Study of High Dynamic Range Image Sensor Architectures," to appear in Proceedings of SPIE, January 2004.
[32] S. Kavusi and A. El Gamal, "Quantitative study of high-dynamic range SigmaDelta-based focal plane array architectures," to appear in Proceedings of SPIE Defense and Security Symposium, April 2004.
[33] M.H. White, D.R. Lampe, F.C. Blaha and I.A. Mack, "Charactarization of surface channel CCD image arrays at low light levels," IEEE Journal of Solid State Circuits, vol. 9, pp. 1-13, 1974.
[34] R.W. Brodersen and S.P. Emmons, "The measurment of noise in buried channel charge coupled devices," Proceedings of International Conference Applications of CCD's, pp. 331-345, 1975.
[35] A.M. Fowler, I. Gatley, "Demonstration of an algorithm for read-noise reduction in infrared arrays," Astrophysics Journal, vol. 353, pp. L33-L34, 1990.
[36] A. Fowler and I. Gatley, "Noise reduction strategy for hybrid IR focal plane arrays," Proceedings of SPIE, vol. 1541, pp. 127-133, 1991.
[37] K. Yoon, C. Kim, B. Lee, and D. Lee, "Single-chip CMOS Image Sensor for Mobile Applications," ISSCC Digest of Technical Papers, vol. 45, pp. 36-37, February 2002.
[38] G. Patoukanis, K. Shepard, R. Levicky, "Active CMOS Biochip for Time-Resolved Fluorescence Detection," VLSI Symposium Digest of Technical Papers, pp. 68-71, 2005.
[39] B. Eversmann, M. Jenkner, F. Hofmann, C. Paulus, R. Brederlow, B. Holzapfl, P. Fromherz, M. Merz, M. Brenner, M. Schreiter, R. Gabl, K. Plehnert, M. Steinhauser, G. Eckstein, D. Schmitt-Landsiedel, R. Thewes, "A 128×128 CMOS biosensor array for extracellular recording of neural activity," IEEE Journal of Solid-State Circuits, vol. 38, No. 12, pp. 2306-2317, 2003.
[40] N. Manaresi, A. Romani, G. Medoro, L. Altomare, A. Leonardi, M. Tartagni, R. Guerrieri, "A CMOS chip for individual cell manipulation and detection," IEEE Journal of Solid-State Circuits, vol. 38, No. 12, pp. 2297-2305, 2003.
[41] A. Romani, N. Manaresi, L. Marzocchi, G. Medoro, A. Leonardi, L. Altomare, M. Tartagni, R. Guerrieri, "Capacitive sensor array for localization of bioparticles in CMOS lab-on-a-chip," IEEE ISSCC Digest of Technical Papers, vol. 47, pp. 224-225, 2004.

[42] M. Schienle, C. Paulus, A. Frey, F. Hofmann, B. Holzapfl, P. Schindler-Bauer, R. Thewes, "A Fully Electronic DNA Sensor With 128 Positions and In-Pixel A/D Conversion," IEEE Journal of Solid-State Circuits, vol. 39, No. 12, pp. 2438-2445, 2004.

[43] R. Yotter, L. Lee and D.M. Wilson, "Sensor Technologies for Monitoring Metabolic Activity in Single Cells–Part I: Optical Methods," IEEE Sensors Journal, vol. 4, No. 4, pp. 395-411, 2004.

[44] R. Yotter, L. Lee and D.M. Wilson, "Sensor Technologies for Monitoring Metabolic Activity in Single Cells–Part II: Non optical Methods and Applications," IEEE Sensors Journal, vol. 4, No. 4, pp. 412-429, 2004.

[45] H. Eltoukhy, K. Salama and A. El Gamal, "A 0.18 μm CMOS Luminescence Detection Lab-on-Chip," IEEE Journal of Solid-State Circuits, vol. 41, pp. 551-562, 2006.

[46] M. Eggers, *et al*., "A microchip for quantitative detection of molecules utilizing luminescent and radioisotope reporter groups," Biotechniques, vol. 17, pp. 516-25, 1994.

[47] J.B. Lamture and *et al*., "Direct detection of nucleic acid hybridization on the surface of a charge coupled device," Nucleic Acids Research, vol. 22, pp. 2121-2125, 1994.

[48] K. Salama, H. Eltoukhy, A. Hassibi and Abbas El Gamal, "Modeling and simulation of integrated bioluminescence detection platforms," Biosensors and Bioelectronics Special issue: Micro and Nano Bioengineering, vol. 19, pp. 1377-1386, 2004.

[49] B. Fowler, A. El Gamal, and D. Yang, "Techniques for pixel-level analog-to-digital conversion," Proceedings of SPIE, Infrared Readout Electronics IV, vol. 3360, pp. 2-12, 1998.

[50] T.M. Souders, "Code probability distributions of A/D converters with random input noise," IEEE Transactions on Instrumentation and Measurement, vol. 47, pp. 1042-1045 Oct. 1998.

[51] E. Sackinger and W. Guggenbuhl, "A versatile building block: the CMOS differential difference amplifier," IEEE Journal of Solid-State Circuits, vol. 22, pp. 287-294, 1987.

[52] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlen and P. Nyren, "Real-Time DNA Sequencing using Detection of Pyrophosphate Release," Analytical Biochemistry, vol. 242, pp. 84-89, 1996.

[53] M. Ronaghi, "Pyrosequencing Sheds Light on DNA Sequencing," Genome Research, vol. 11, pp. 3-11, 2001.

[54] Beckman Coulter LD400 Specifications, "http://www.beckman.com/products/instrument/automatedsolutions/integsystems/ld400 inst dcr.asp," 2005.

AUTHOR BIOGRAPHY

Abbas El Gamal Abbas El Gamal (S'71-M'73-SM'83-F'00) received his B.Sc. degree in Electrical Engineering from Cairo University in 1972, the M.S. in Statistics and the PhD in Electrical Engineering from Stanford in 1977 and 1978, respectively. From 1978 to 1980 he was an Assistant Professor of Electrical Engineering at USC. He has been on the Stanford faculty since 1981, where he is currently Professor of Electrical Engineering and the Director of the Information Systems Laboratory. He was on leave from Stanford from 1984 to 1988 first as Director of LSI Logic Research Lab, then as cofounder and Chief Scientist of Actel Corporation. In 1990 he co-founded Silicon Architects, which was later acquired by Synopsys. His research has spanned several areas, including information theory, digital imaging, and integrated circuit design and design automation. He has authored or coauthored over 180 papers and 25 patents in these areas.

Helmy Eltoukhy received the B.S. (with distinction), M.S. and Ph.D. degrees, all in electrical engineering, from Stanford University, Stanford, CA, in 2000, 2001 and 2006, respectively, where he currently is a postdoctoral fellow at the Stanford Genome Technology Center. His current research interests include low-noise CMOS image sensor design, hand-held biosensors for environmental and medical diagnostics, and biological applications of statistical signal processing. He is currently working on a joint project with the Stanford Genome Technology Center to develop low-cost, ultra-high throughput DNA sequencing platforms based on CMOS image sensors. His work is currently supported by the National Institute of Health.

Khaled Salama received the B.Sc. degree with honors from the Electronics and Communications Department, Cairo University, Cairo, Egypt, in 1997, and the M.Sc. and Ph.D. degrees from the Electrical Engineering Department, Stanford University, Stanford, CA, in 2000 and 2005, respectively. He is currently an Assistant Professor in the Electrical, Computers and Systems and the Bioengineering Departments at Rensselaer Polytechnic Institute, Troy, NY. He has authored or coauthored 25 papers and holds two patents in the areas of CMOS image sensors, integrated biosensors, VLSI architectures for bioimaging and analog filters.

INDEX