The fields of biological and medical physics and biomedical engineering are broad, multidisciplinary and dynamic. They lie at the crossroads of frontier research in physics, biology, chemistry, and medicine. The Biological and Medical Physics, Biomedical Engineering Series is intended to be comprehensive, covering a broad range of topics important to the study of the physical, chemical and biological sciences. Its goal is to provide scientists and engineers with textbooks, monographs, and reference works to address the growing need for information.

Books in the series emphasize established and emergent areas of science including molecular, membrane, and mathematical biophysics; photosynthetic energy harvesting and conversion; information processing; physical principles of genetics; sensory communications; automata networks, neural networks, and cellular automata. Equally important will be coverage of applied aspects of biological and medical physics and biomedical engineering such as molecular electronic components and devices, biosensors, medicine, imaging, physical principles of renewable energy production, advanced prostheses, and environmental control and engineering.

Editor-in-Chief:
Elias Greenbaum, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

Editorial Board:
Masuo Aizawa, Department of Bioengineering, Tokyo Institute of Technology, Yokohama, Japan
Olaf S. Andersen, Department of Physiology, Biophysics & Molecular Medicine, Cornell University, New York, USA
Robert H. Austin, Department of Physics, Princeton University, Princeton, New Jersey, USA
James Barber, Department of Biochemistry, Imperial College of Science, Technology and Medicine, London, England
Howard C. Berg, Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts, USA
Victor Bloomfield, Department of Biochemistry, University of Minnesota, St. Paul, Minnesota, USA
Robert Callender, Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York, USA
Britton Chance, Department of Biochemistry/Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania, USA
Steven Chu, Lawrence Berkeley National Laboratory, Berkeley, California, USA
Louis J. DeFelice, Department of Pharmacology, Vanderbilt University, Nashville, Tennessee, USA
Johann Deisenhofer, Howard Hughes Medical Institute, The University of Texas, Dallas, Texas, USA
George Feher, Department of Physics, University of California, San Diego, La Jolla, California, USA
Hans Frauenfelder, Los Alamos National Laboratory, Los Alamos, New Mexico, USA
Ivar Giaever, Rensselaer Polytechnic Institute, Troy, New York, USA
Sol M. Gruner, Cornell University, Ithaca, New York, USA
Judith Herzfeld, Department of Chemistry, Brandeis University, Waltham, Massachusetts, USA
Mark S. Humayun, Doheny Eye Institute, Los Angeles, California, USA
Pierre Joliot, Institute de Biologie Physico-Chimique, Fondation Edmond de Rothschild, Paris, France
Lajos Keszthelyi, Institute of Biophysics, Hungarian Academy of Sciences, Szeged, Hungary
Robert S. Knox, Department of Physics and Astronomy, University of Rochester, Rochester, New York, USA
Aaron Lewis, Department of Applied Physics, Hebrew University, Jerusalem, Israel
Stuart M. Lindsay, Department of Physics and Astronomy, Arizona State University, Tempe, Arizona, USA
David Mauzerall, Rockefeller University, New York, New York, USA
Eugenie V. Mielczarek, Department of Physics and Astronomy, George Mason University, Fairfax, Virginia, USA
Markolf Niemz, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
V. Adrian Parsegian, Physical Science Laboratory, National Institutes of Health, Bethesda, Maryland, USA
Linda S. Powers, University of Arizona, Tucson, Arizona, USA
Earl W. Prohofsky, Department of Physics, Purdue University, West Lafayette, Indiana, USA
Andrew Rubin, Department of Biophysics, Moscow State University, Moscow, Russia
Michael Seibert, National Renewable Energy Laboratory, Golden, Colorado, USA
David Thomas, Department of Biochemistry, University of Minnesota Medical School, Minneapolis, Minnesota, USA

Steven Chu, Lawrence Berkeley National Laboratory, Berkeley, California, USA
Over the last several decades, our ability to understand and diagnose disease has dramatically progressed due to advances in quantitative measurements of molecular biomarkers. Techniques such as genetic sequencing, mass spectrometry, flow cytometry, and nuclear magnetic resonance have enabled sophisticated and highly sensitive measurements of biological material that promise improvements in early diagnosis, accurate monitoring of existing diseases, and understanding of the underlying causes and mechanisms of disease. However, the widespread translation of these advanced techniques toward improving patient care has been severely limited by cost and the requisite infrastructure that these technologies demand.

Medical diagnostics at the point-of-care (POC), e.g., performed by a clinician or by a patient without the need for a clinical lab, can greatly improve access to medical diagnosis. Simple POC diagnostics have been available for many years, such as glucose monitors for diabetics and home pregnancy tests, and have provided tremendous benefits for patients. Motivated by the success of these simple tests, there has been a strong push to develop more complex chips that automate and miniaturize advanced diagnostics currently performed in laboratories. These proposed “lab-on-a-chip” devices combine advances in electronics, photonics, microelectromechanical systems (MEMS), and microfluidics to build small, low-cost chips that can perform extremely complex tasks. Much research has gone towards the development of these lab-on-a-chip systems, but only recently has this technology begun to reach the level of maturity for practical use.

It is appropriate that many of the researchers working on developing these biomedical chips have backgrounds in semiconductor physics and engineering. Over the last 50 years, the semiconductor industry, through feats of integration and miniaturization, has turned computers from million-dollar, room-sized machines into the pocket-sized smart phones that are now so globally ubiquitous. Following the example of the semiconductor industry, researchers today imagine handheld devices the size of a cell phone, that with a drop of blood, spit, urine, or sputum
will run a battery of medical diagnostics at a cost and speed that cannot be matched by the manual labs of today.

In addition to the logistical and economical improvements that come from miniaturizing diagnostics, such chips can outperform the larger and more expensive traditional diagnostics tools. Sensors and actuators perform best when they are sized appropriately for the objects that they are measuring. In the case of biology, great improvements in performance come from having micro- and nanometer-sized tools to control and measure cells and molecules respectively.

The invention of new techniques to inexpensively and rapidly diagnose disease at the POC promise enormous impacts on many pressing current health issues. One example of the urgent need for this technology is the emergence of drug resistant strains of tuberculosis (TB). These multi-drug resistant strains (MDR-TB) are mutated strains of TB that do not respond to the standard treatments for the disease. For the effective treatment of this potential pandemic, clinicians need to be able to accurately target the more expensive and more aggressive second-line-of-defense drugs to patients with MDR-TB. However, the best method to rapidly identify MDR-TB in patients is with polymerase chain reaction (PCR), which, despite recent efforts, is still largely unavailable in resource-poor regions where the MDR-TB epidemic is currently localized.

This book consists of chapters by leading researchers working on biomedical chips to bring advanced diagnostics to the point of care. This work is separated into two broad categories: (1) the development of miniaturized chips to control and prepare biological samples and reagents and (2) the development of sensors that can be miniaturized onto chips for the accurate detection of biomarkers.

Beginning with work in fluid handling, Sam Sia describes the work of his group at Columbia University to engineer extremely low-cost microfluidics for POC diagnostics using injection molded chips preloaded with reagents. Sia’s work has led to a start-up company Claros Diagnostics Inc. and has recently been used in a clinical trial to diagnose STDs in rural areas of Rwanda. Next, Robert M. Westervelt describes the work by his group to demonstrate a programmable platform to control single cells and pL drops of fluid that utilizes the complexity, small feature size, and low cost of integrated circuits (ICs) combined with microfluidics.

A variety of techniques to measure molecular biomarkers on biomedical chips are described. Peter Kiesel at the Palo Alto Research Center and Aydogan Ozcan at UCLA describe two very different approaches toward miniaturizing optical measurements for the high-throughput bright-field and fluorescence screening of cells. Yang-Kyu-Choi at Korean Advanced Institute of Technology, Toshiya Sakata at University of Tokyo, and Shan Wang at Stanford discuss recent work by their groups to utilize the complexity and small feature sizes of microchip technology to build extremely sensitive microfabricated sensors for molecular biomarkers. Donhee Ham at Harvard University and Hakho Lee at Massachusetts General Hospital describe the work by their groups to miniaturize nuclear magnetic resonance to detect sparse biomarkers in turbid biological samples.
Overall, the work by these researchers paints a very hopeful picture. Through the continuing advancement of miniaturized diagnostic technology, future clinicians and patients will have access to more accurate, timelier, and far less expensive diagnosis of disease.

Philadelphia, PA, USA  
Cambridge, MA, USA

David Issadore  
Robert M. Westervelt
## Contents

### Part I Fluid Handling

1 **Low-Cost Microdevices for Point-of-Care Testing** ............................... 3  
Curtis D. Chin, Sau Yin Chin, Tassaneewan Laksanasopin,  
and Samuel K. Sia  
1.1 Introduction .................................................................................. 3  
1.2 Comparisons Between Point-of-Care Tests in Developed  
    Versus Developing Worlds .............................................................. 4  
1.3 Classes of Analytes ....................................................................... 6  
   1.3.1 Proteins .................................................................................. 6  
   1.3.2 Cells .................................................................................... 6  
   1.3.3 Nucleic Acids ........................................................................ 7  
   1.3.4 Small Molecules .................................................................... 7  
1.4 Applications .................................................................................. 8  
   1.4.1 Immunoassays ........................................................................ 8  
   1.4.2 Cell-Based Assays ................................................................. 11  
   1.4.3 Nucleic Acid Amplification Testing ....................................... 12  
   1.4.4 Clinical Chemistry Assays ..................................................... 16  
   1.4.5 Hematology ........................................................................... 17  
1.5 Outlook ......................................................................................... 17  
References .......................................................................................... 18

2 **Programmable Hybrid Integrated Circuit/Microfluidic Chips** ............ 23  
Caspar Floryan, David Issadore, and Robert M. Westervelt  
2.1 Introduction ................................................................................... 23  
2.2 Principles of Dielectrophoresis ..................................................... 24  
2.3 Integrated Circuit/Microfluidic Chips .......................................... 28  
2.4 Functions Performed by Hybrid Integrated  
    Circuit/Microfluidic Chips ............................................................ 30  
   2.4.1 Positioning/Programmable Channels ................................... 30  
   2.4.2 Deform .................................................................................. 33
Part II Detection

3 Flow Cytometry on a Chip ............................................................. 47
   Peter Kiesel, Joerg Martini, Michael I. Recht,
   Marshall W. Bern, Noble M. Johnson, and Malte Huck
   3.1 Flow Cytometry ............................................................. 47
       3.1.1 Introduction ..................................................... 47
       3.1.2 Miniaturized Flow Cytometers and Microfluidic-
              Based Approaches ............................................. 49
   3.2 On-the-Flow Analyte Characterization Based on Spatial
       Modulation Technique .................................................. 51
       3.2.1 Spatially Modulated Fluorescence Emission:
              The Enabling Technique ........................................ 51
       3.2.2 Data Evaluation and Correlation Analysis ................. 52
       3.2.3 CD4 Count in Whole Blood .................................... 56
       3.2.4 Handheld Prototype ............................................. 58
   3.3 Pathogen Detection in Water ............................................. 59
       3.3.1 Background and Currently Used Techniques ............... 59
       3.3.2 Pathogen Detection in Water with Spatially
              Modulated Emission ............................................ 60
   3.4 Multiplexed Flow Assay ................................................ 61
   3.5 Multicolor Detection with Single Large-Area Detector ........... 63
   3.6 Summary and Outlook .................................................. 64
   References ..................................................................... 66

4 Lensfree Computational Microscopy Tools for On-Chip
   Imaging of Biochips .......................................................... 71
   Serhan O. Isikman, Waheb Bishara, Onur Mudanyali,
   Ting-Wei Su, Derek Tseng, and Aydogan Ozcan
   4.1 Introduction ............................................................. 71
   4.2 Principles of Partially Coherent Lensfree On-Chip
       Holographic Microscopy ............................................... 72
   4.3 Lensfree Holographic Microscopy of Biochips Using
       a Single Partially Coherent Source (LED) ......................... 74
   4.4 Lensfree Holographic Microscopy on a Cell Phone ............... 77
5 Dielectric Detection Using Biochemical Assays
Yang-Kyu Choi, Chang-Hoon Kim, Jae-Hyuk Ahn, Jee-Yeon Kim, and Sungho Kim
5.1 Introduction ............................................................. 97
5.2 Electrical Detection Based on Dielectric and Charge Properties ... 98
  5.2.1 Electrochemical Impedance Spectroscopy ..................... 98
  5.2.2 FET-Based Biosensor ........................................... 100
5.3 Dielectric-Modulated Field-Effect Transistor (DMFET) .............. 103
  5.3.1 Basic Structure and Theory .................................... 103
  5.3.2 Proof of Concept and DNA Detection with DMFET ...... 106
5.4 Structural Modification of a DMFET ................................ 109
  5.4.1 Underlap FET ................................................... 109
  5.4.2 Double-Gate FET ............................................... 109
5.5 Sensing Metrics in a DMFET .......................................... 113
  5.5.1 Interface Trap Charge Method ................................. 113
  5.5.2 Substrate Current Method ...................................... 117
5.6 Environmental Effect ................................................... 118
References ..................................................................... 120

6 Semiconductor-Based Biosensing Chip for Point-of-Care Diagnostics
Toshiya Sakata
6.1 Background .................................................................. 125
6.2 Concept of Semiconductor-Based Biosensing Devices ............. 127
  6.2.1 Ion-Sensitive Field-Effect Transistor (IS-FET) ............... 127
  6.2.2 Device Structure ................................................ 128
6.3 Signal Transduction Interface .......................................... 133
  6.3.1 Physically Structured Interface ................................. 133
  6.3.2 Chemically Synthesized Interface ............................. 133
  6.3.3 Biologically Induced Interface .................................. 134
6.4 Molecular Recognition Based on Intrinsic Molecular Charges ....... 135
  6.4.1 Genetic Analysis ................................................ 135
  6.4.2 Detection of Low-Molecular-Weight Antigen ............... 143
6.5 Noninvasive and Real-Time Monitoring of Cell Functions ......... 146
  6.5.1 Metabolism ...................................................... 146
6.6 In Vitro Cell Sensing with Semiconductor-Based Biosensing Technology .................................................. 149
References ..................................................................... 151
7 Magneto-Nanosensor Diagnostic Chips .................................................. 153
Richard S. Gaster, Drew A. Hall, and Shan X. Wang

7.1 Introduction .............................................................................. 154
7.2 Magnetic Biosensing Modalities .............................................. 154
7.3 Magnetic Nanoparticle Tags ..................................................... 159
  7.3.1 Superparamagnetism ....................................................... 159
  7.3.2 Application of Magnetic Nanoparticles for Magnetic Biosensing .................................................. 160
7.4 Protein Detection Assay Using Magnetic Nanoparticles ............... 162
  7.4.1 Standard Protein Detection Assay ....................................... 162
  7.4.2 Wash-Free Protein Detection Assay for POC ....................... 164
  7.4.3 Microfluidic Integration of Magnetic Biosensors for POC .......................................................... 165
7.5 Miniaturization of Desktop Biostation ....................................... 167
7.6 POC Detection Results via GMR Biosensor Arrays ..................... 171
7.7 Conclusions ........................................................................... 173
References ..................................................................................... 174

8 Handheld NMR Systems and Their Applications for Biomolecular Sensing ................................................................................... 177
Nan Sun and Donhee Ham

8.1 Introduction .............................................................................. 177
8.2 NMR RF Transceiver IC Design .............................................. 181
  8.2.1 Overall Architecture and Operation .................................. 181
  8.2.2 Characteristics of the Coils ............................................. 183
  8.2.3 Transmitter with Proton Filter ........................................ 183
  8.2.4 Heterodyne Receiver with Passive Amplification ............... 186
8.3 Transceiver Measurements ....................................................... 187
  8.3.1 Receiver Measurement ................................................... 188
  8.3.2 Transmitter Measurements ............................................. 189
8.4 NMR Experiments and NMR-Based Biomolecular Sensing .......... 190
  8.4.1 Proton NMR Experiments .......................................... 190
  8.4.2 NMR-Based Biomolecular Sensing ................................ 191
8.5 Comparison to Other Miniaturization Works .............................. 195
8.6 Conclusion ............................................................................. 195
References ..................................................................................... 196

9 Diagnostic Magnetic Resonance Technology ...................................... 197
Changwook Min, Huilin Shao, David Issadore, Monty Liong,
Ralph Weissleder, and Hakho Lee

9.1 Introduction .............................................................................. 198
9.2 Principle of DMR Detection .................................................... 199
9.3 New Magnetic Nanoparticles ................................................. 200
  9.3.1 Ferrite-Based MNPs .................................................. 201
  9.3.2 Fe-Core MNPs .......................................................... 202
9.4 Optimizing MNPs for DMR Applications ........................................... 204
  9.4.1 Biocompatible Coating on Hydrophobic MNPs ...................... 204
  9.4.2 New Labeling Method for High MNP Loading ..................... 206
9.5 Miniaturized NMR System .......................................................... 206
  9.5.1 System Concept .................................................................. 208
  9.5.2 Optimal NMR Probe Design .............................................. 209
  9.5.3 Clinical System .................................................................. 211
9.6 Biological Applications .............................................................. 214
  9.6.1 Cancer Detection and Profiling .......................................... 214
  9.6.2 Clinical Trial .................................................................... 215
9.7 Conclusions and Outlook ............................................................ 219
References ...................................................................................... 219

Index .................................................................................................. 223
Contributors

Yang-Kyu Choi  Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea, ykchoi@ee.kaist.ac.kr

Donhee Ham  School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA, donhee@seas.harvard.edu

Peter Kiesel  Palo Alto Research Center, Palo Alto, CA, USA, Peter.Kiesel@parc.com

Hakho Lee  Center for Systems Biology, Massachusetts General Hospital/Harvard Medical School, Boston, MA, USA, hlee@mgh.harvard.edu

Aydogan Ozcan  Electrical Engineering Department, Bioengineering Department, University of California, Los Angeles, USA, ozcan@ee.ucla.edu

Toshiya Sakata  Department of Materials Engineering, The University of Tokyo, Tokyo, Japan, sakata@biofet.t.u-tokyo.ac.jp

Samuel K. Sia  Department of Biomedical Engineering, Columbia University, New York, NY, USA, ss2735@columbia.edu

Shan X. Wang  Department of Bioengineering, Stanford University, Stanford, CA, USA, sxwang@stanford.edu

Robert M. Westervelt  Department of Physics, School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA, westervelt@seas.harvard.edu
Part I
Fluid Handling
Chapter 1
Low-Cost Microdevices for Point-of-Care Testing

Curtis D. Chin, Sau Yin Chin, Tassaneewan Laksanasopin, and Samuel K. Sia

Abstract Microdevices enable clinical diagnostics to be miniaturized for use at the point-of-care (POC). Microdevices can be composed of microfilters, microchannels, microarrays, micropumps, microvalves, and microelectronics, and these mechanical and electrical components can be integrated onto chips to analyze and control biological objects at the microscale. The miniaturization of diagnostic tests offers many advantages over centralized laboratory testing, such as small reagent volumes, rapid analysis, small size, low power consumption, parallel analysis, and functional integration of multiple devices. Here, we review work on the development of microdevices to diagnose disease at POC settings.

1.1 Introduction

An exciting application of microdevices is miniaturized clinical diagnostic methods for point-of-care testing. Microdevices are defined as systems which employ analytical components and techniques at microscopic scales (ranging from 1 to 500 μm), including microfilters, microchannels, microarrays, micropumps, microvalves, and bioelectronic chips [1, 2]. Advantages over tests in centralized laboratories include small reagent volumes, rapid analysis, small size, low power consumption, parallel analysis, and functional integration of multiple devices [3–5]. Because of these advantages, microdevices are capable of being low cost and portable — two qualities critical for implementation of point-of-care (POC) diagnostics.

POC tests are important in settings where timing is critical (e.g., emergency triage), where laboratory facilities are nonexistent (e.g., military or extraterrestrial...
functions), and where resources are low (e.g., in developing countries). They help deliver appropriate and prompt treatments and improve clinical outcomes [5–7].

The development of microdevice-based diagnostics originated from driving forces in molecular analysis, biodefence, molecular biology, and microelectronics, the last of which provided tools for miniaturization and microfabrication [1,8]. However, miniaturized silicon-based devices were not suited to biological applications with physiological fluids because of difficulty of optical detection, high material cost, low biocompatibility, and lack of available rapid prototyping. The extension of microfabrication techniques to first glass then polymers such as elastomers and hard plastics enabled the construction of a set of microfluidic mechanical devices grouped under the broad category of microelectromechanical systems (MEMS) or lab-on-a-chip (LOC) systems [9]. These technologies, in turn, have led to developments in sensors for medical, environmental, and life science applications [1].

1.2 Comparisons Between Point-of-Care Tests in Developed Versus Developing Worlds

Integration, portability, low power consumption, automation, and ruggedness are several important qualities common to point-of-care tests in both developed and developing countries. (Here we define the developed world as the group of high-income countries and the developing world as the groups of low- and middle-income countries, based on classifications by the World Bank). These general design constraints suggest some LOC components and procedures to be more appropriate than others in point-of-care settings (Fig. 1.1). In general, the differences in appropriate LOC procedures are more pronounced between tests performed at centralized settings versus point-of-care settings than between point-of-care tests performed at high-income versus low-income settings [10]. For instance, the presence of reliable ground electricity at centralized testing facilities allows for active mixing on-chip and bulky signal detection systems which draw significant power. At point-of-care, however, the lack of ground electricity requires passive or low-power approaches for mixing on-chip and portable signal detection systems which can be powered by battery (or read by eye) (Fig. 1.1). By virtue of the common design criteria in point-of-care tests, engineers interested in developing microdevices for point-of-care testing in resource-limited settings can leverage the large body of existing research on LOC designs for point-of-care testing for physicians and home use [11, 12], devices for military applications [13] and first responders, and extraterrestrial sensors [14–16].

There are, however, at least two important distinctions between LOC point-of-care tests intended for developed countries versus developing countries. The first is cost. Due to vastly smaller budgets via public financing, the budget available to spend on medical diagnostics devices in developing countries is limited compared to that of developed countries. Thus, the cost of the microfluidic device (which includes both the material and the manufacturing process) must be kept extremely low in POC testing in developing countries; the fixed instrument must be portable
Fig. 1.1 Appropriateness of technologies for important LOC procedures according to setting (centralized versus point-of-care) and resource-level (high income versus low income) ([5] – reproduced by permission of the Royal Society of Chemistry)

Fig. 1.2 Distribution of disease burden (DALYs) in high-income, middle-income and low-income countries [17]

and cheap, and the disposable must be extremely cheap. All components of the device (including the instrument and disposable) must be robust and rugged under a variety of environmental conditions.

Clinical need is another feature distinguishing POC tests in the developed versus the developing worlds (Fig. 1.2). The burden of disease in the developed world skews heavily toward noncommunicable diseases: cardiovascular disease (stroke, heart disease), cancer, respiratory disease (COPD and asthma), and neuropsychiatric conditions are some of the more significant causes [17]. By contrast, in the developing world, while noncommunicable diseases are prevalent, a significant burden of disease is attributed to infectious diseases, many of which can be treated and prevented. Major infectious diseases according to the burden of disease include HIV/AIDS, tuberculosis, malaria, diarrheal diseases, and lower respiratory infections. Maternal and perinatal conditions and nutritional deficiencies are also significant causes of disease and disability in the developing world.

Point-of-care tests based on LOC technologies have the potential for rapid and portable diagnosis of a range of conditions. The rest of this chapter will describe
the range of analytes detected by low-cost microdevices at point-of-care and their specific applications.

1.3 Classes of Analytes

Microdevices have been used in academic, government, nonprofit, and commercial settings to detect a range of analytes using many types of biological fluids. These studies have demonstrated assays with better performance than benchtop tests with regards to speed, portability, sensitivity, and multiplexing. The classes of analytes detected include proteins, nucleic acids, cells, and small molecules [5]. Each analyte class will be described in the section on applications.

1.3.1 Proteins

Proteins in clinical specimens in whole blood, serum/plasma, saliva, urine, and other sample matrices have been used for clinical diagnostics and monitoring disease states [18]. Microdevice-based technologies at point-of-care for detecting proteins have included both immunoassays and enzymatic assays. Currently available clinical tests for microdevices at point-of-care include viral infections (anti-HIV antibodies, antibodies against influenza A/B virus, rotavirus antigens), bacterial infections (antibodies against Streptococcus A and B, Chlamydia trachomatis, Treponema pallidum), parasitic infections (histidine-rich protein 2 for P. falciparum, trichomonas antigens), and noncommunicable diseases (PSA for prostate cancer, C-reactive protein for inflammation, HbA1c for plasma glucose concentration) [5, 19, 20]. For example, a saliva-based nano-biochip immunoassay has been constructed to detect a panel of C-reactive protein, myoglobin, and myeloperoxidase in acute myocardial infarction patients [21].

1.3.2 Cells

Cell-based POC testing is often needed for disease diagnosis and hematological analysis. Full blood cell counts provide information in diagnosing and monitoring conditions such as anemia and HIV/AIDS. For instance, CD4+ lymphocyte counting is used to monitor the progression of HIV/AIDS, and microdevices have been developed to substitute the conventional method (bulky and expensive instruments such as flow cytometry) for POC testing. Traditional methods also often require trained personnel and ground electricity. On-chip cell sorting techniques from heterogeneous cell suspensions (such as whole blood) can be classified into two main categories: size- and density-based techniques and affinity-based
techniques (chemical, electrical, or magnetic) [22]. Several novel size-based separation methods force a stream of cells through a series of channels by hydrodynamic flow switching, electrokinetic flow switching, dielectrophoresis, electrowetting-assisted flow switching, or valves [5] which provide predictable and reproducible cell movement [22]. These methods do not employ the use of any proteins or biologies to achieve cell separation, hence, these size-based devices would have longer shelf lives. Affinity-based methods include the specific capture of cells using antibodies such as with magnetic beads coated with antibodies, which is a common laboratory protocol and can be miniaturized easily. Another affinity-based method makes use of dielectrophoresis. This method makes use of the differential response of cells to electric fields due to their density, physiological, and metabolic states. Other approaches involve lysing the undesired cells, leaving behind the desired cells in the reaction chamber [23]. This technique can reduce the processing time and therefore avoid cell damage from prolonged contact with lysis agents and centrifugation [22]. These approaches are easily implemented on microfluidic devices and usually followed by downstream analysis such as fluorescence microscopy or proteomics/genomics analysis following the lysing of the captured cells [24].

1.3.3 Nucleic Acids

Clinical diagnoses can be made based on the analysis of DNA or RNA sequences. Nucleic acid detection and analysis can identify the type of infection or pathogen and diseases. It can be used in prenatal diagnosis of inherited disorders, clinical disease diagnosis (genetic disease, infection, disease staging, drug resistance mutation, and pathogen presence/abundance), and forensic investigations. Nucleic acid testing (NAT) offers detection that is highly sensitive (due to amplification) and specific (due to specific base pairing of complementary nucleotides). NAT also reduces the window period between infection and detectability of disease. Currently available clinical tests mostly target infectious diseases, for instance, HIV (diagnosis and viral load monitoring), H1N1 influenza, tuberculosis, and group B streptococcal disease. These available systems are mainly used in hospitals and centralized laboratories with complex operation steps and high-cost instruments [25]. In order to achieve NAT for POC diagnosis, fully integrated system is preferable; for instance, to avoid contamination issues, reduce worker steps, and deliver rapid results.

1.3.4 Small Molecules

Analyzing electrolytes from body fluids is used to monitor health parameters in disease prevention. The ranges of electrolytes (Na$^+$, K$^+$, Cl$^-$, Ca$^{2+}$), general chemistries (pH, urea, glucose), blood gases (pCO$_2$, pO$_2$), and hematology
(hematocrit) are analyzed. Currently, methods are based on electrochemical detection such as potentiometry, amperometry, and conductance [5] (see applications for more details). Another microfluidic approach to ion analysis is optical sensing [26]. For example, a microfluidic device can be used for pH analysis based on the color change of pH-sensitive dyes immobilized in patterned hydrogels [27].

1.4 Applications

1.4.1 Immunoassays

One of the most common protein assays that are adapted for point-of-care microfluidic-based diagnostics is the immunoassay. This assay makes use of the binding interactions between antigens and antibodies to detect protein markers from either pathogen or host immune responses. Analytes for immunoassays span the entire range of pathogen types, from viruses (e.g., anti-HIV antibodies and p24 antigen for HIV) to bacteria (e.g., anti-treponemal antibodies for syphilis and early secretory antigenic target 6 for tuberculosis) and parasites (e.g., histidine-rich protein 2 for malaria). Immunoassays are also commonly used for detection of noncommunicable diseases, such as prostate cancer via measurement of levels of prostate specific antigen, as in a system by OPKO Diagnostics (formerly Claros Diagnostics) or heart disease via measurement of levels of B-type natriuretic peptide, as in the Triage BNP Test by Biosite [28].

Heterogeneous immunoassays are a popular immunoassay format where capture of analytes occurs at protein-modified surfaces. The reference standard of many heterogeneous immunoassays is the enzyme-linked immunosorbent assay (ELISA), which uses multiwell plates and can be automated for high-throughput processing at well-equipped central laboratories. Most ELISAs use colorimetric- or chemiluminescence-based detection and have detection limits typically in the pM range due to enzyme-mediated signal amplification and serial washing. Unfortunately, the traditional ELISA format, which requires expensive and bulky instrumentation (for liquid handling and signal detection) and trained workers, is ill-suited for point-of-care testing.

Microfluidics is an attractive technology for point-of-care immunoassays. Heterogeneous immunoassays, because they involve capture of analytes at surfaces, are well-suited to exploit the large surface-to-volume ratios encountered in microfluidics. Faster analysis times can be achieved because of the replenishment of analytes and detection reagents in the boundary layer above the surface in standard wellplate formats [29]. Below, we highlight recent advances in microfluidic immunoassays that are low cost, portable, and can be used at point-of-care settings.
1.4.1.1 Signal Detection

Our group has demonstrated signal amplification using reduction of silver ions on gold nanoparticles [30] in cassettes made of injection-molded plastic (from Claros Diagnostics) for the simultaneous detection of HIV and syphilis antibodies (Fig. 1.3). Gold-catalyzed silver reduction is an attractive method of signal amplification and detection because the signals can be developed under continuous flow and can be read using low-cost optics. In a preclinical evaluation at several clinical sites in Rwanda, the technology showed sensitivity and specificity rivaling those of reference benchtop assays on hundreds of patient samples, with a time-to-result of no more than 20 min [31]. We also demonstrated excellent performance of HIV diagnosis using only $\frac{1}{157}$ L of unprocessed whole blood. The potential social impact of this test is high when used in remote settings (e.g., clinics providing antenatal care), since treatments for HIV and syphilis are affordable and effective for minimizing disease transmission from mother to child.

Surface plasmon resonance is another suitable detection method, given appropriate signal amplification schemes. Sensata Technologies has previously developed a low-cost, disposable device which can detect heterogeneous antigen/antibody binding with changes in refractive index measured at the gold surface.

1.4.1.2 Fluid Control

Valves are an essential component of fluid control on microdevices. Pneumatic deflection of elastomers, such as PDMS, in multilayer systems has been demonstrated to be robust, leak-free, and capable of facilitating highly multiplexed systems
and have been utilized in development of bead-based microfluidic ELISA systems. However, these micromechanical valves require high pressure often supplied by gas tanks. Active microvalves which are based on elastomeric deflections and are better suited at point-of-care settings include hydraulic (liquid) deflection based on magnetic actuation with solenoids and torque-actuated deflection using small machine screws.

Passive approaches, which leverage differences in fluid behavior from varying microchannel geometries in capillary systems, are attractive because they do not require external power and moving off-chip parts. Such control mechanisms include delay valves, which merge smaller channels into larger channels to allow for smooth collection of incoming fluid streams at different flowrates; stop valves, which are the assembly of multiple stop valves preventing further fluid flow into a common outlet until the arrival of all inlet streams. Some of these microfluidic control elements have been integrated on a chip for detection of C-reactive protein. In addition, check valves have been implemented on a microfluidic device for multistep ELISA, detecting botulinum neurotoxin.

We have developed valveless delivery of reagents in microfluidic systems which have been demonstrated for detecting anti-HIV antibodies and anti-treponemal antibodies. The plug-based reagent delivery is a robust and low-cost approach for delivering multiple reagents without the need for on-chip valves.

1.4.1.3 Fluid Actuation and Delivery

Movement of fluids by capillary forces is reliable and does not require external power or moving parts. Miniaturized immunoassays based on capillary forces have been used for detecting cardiac markers and luteinizing hormone. Paper-based microfluidic systems also leverage capillary flow and have been used to detect anti-HIV antibodies. Despite the need for external power, electrophoretic immunoassays in capillaries can be utilized if the power requirement is low (allowing for battery operation) and external instrumentation integrated in a single, easy-to-use device; such has been demonstrated in a promising proof-of-concept device for integrated, rapid, point-of-care testing of biotoxins ricin, Shiga toxin I, and Staphylococcal enterotoxin B, as well as in a rapid bioassay for endogenous matrix metalloproteinase-8 in saliva. Pneumatic-based actuation of fluids can be suited for point-of-care settings, for example, in the manual operation of on-card bellows in the commercially available ABO blood typing chip by Micronics. A hand pump can also be used for pneumatic fluid actuation in microfluidic immunoassays. Injection-molded centrifugal-based platforms rely on spin frequency to drive fluid movement, and movements are gated by capillary or hydrophobic valves; these can be suited for point-of-care testing in resource-limited settings, where in one example a centrifugal bead-based immunoassay was developed for the detection of antigen and antibody to hepatitis B virus.
1.4.2 Cell-Based Assays

POC testing for cell-based diagnostics has been increasingly important in hematology laboratory practices. It involves the numeration of specific cell types, typically white blood cells, from whole blood for the diagnosis, monitoring, and staging of diseases such as HIV/AIDS. A common use of microfluidics for POC cellular analysis is to replace traditional, resource-intensive technologies such as those used for flow cytometry and cellular biosensors. Such microfluidic methods may involve cell (or microparticle) separation based on size [45] and are based on methods such as filtration [46], laminar flow [47], dielectrophoretic force [48], optical gradient force [49], magnetic force [50], and acoustic force [51]. Other methods are based on antibody-specific capture of cells [52].

POC testing has been increasingly important in hematology practices, especially in obtaining white cell counts. The HemoCue WBC system (HemoCue AB, Sweden) measures total white blood cells in just 10 μL of sample [53] and consists of a photomicroscope, a microcuvette holder, and an LCD display (Fig. 1.4a). It is able to operate on batteries and has low power consumption, making it suitable for POCT. The Chempaq analyzer is a self-contained system that uses differential impedance to detect different types of cells to return full blood counts within 3 min [54].

Due to the pressing need for a point-of-care CD4 T-lymphocyte enumeration method for the management of HIV/AIDS epidemic, especially in underserved regions, much work has been done to develop a microfluidic approach. A lysate impedance measurement method has been developed where the specific cells of interest are captured and lysed, and the resulting change in impedance is measured...
using electrodes integrated on-chip [55]. This is the underlying technology behind the Daktari CD4 system, a POC microdevice developed by Daktari Diagnostics that is portable, handheld, and does not require complex sample processing [24] (Fig. 1.4b). Another method, being commercialized by LabNow, uses semiconductor quantum dots to label captured CD4 cells, and an automated software is then used to analyze the fluorescent images [56]. An interesting use of microfluidics to create POC microdevices for cell-based assays is the integration of commercially available fiber optics with the microdevices so as to miniaturize the whole system by reducing any required optical and mechanical hardware [57]. Finally, the Pima CD4 test is being tested in the developing world [58].

We have utilized ultrasensitive chemiluminescence for detection of CD4+ cells on a microfluidic platform [59]. Compared to other optical-based LOC approaches for CD4 counting, this method used an instrument that requires no external light source and no image processing to produce a digitally displayed result only seconds after running the test.

1.4.3 Nucleic Acid Amplification Testing

In order to perform nucleic acid testing on a microfluidic chip, the main functionalities that have to be integrated on-chip include sample preparation, nucleic acid amplification, and the detection of the amplified product. While these steps can be easily integrated together and performed in a general laboratory, miniaturization of such disparate processes onto a single microfluidic chip is still a topic of intense research [60–62]. A fully integrated system could avoid contamination, reduce worker steps, and deliver rapid results. However, most of the devices that have been reported are single-function and require separate modules for detection and analysis due to the challenges involved in combining and miniaturizing these functionalities into a cost-effective, simple, and robust platform which these processes require much effort and time [25, 60].

1.4.3.1 Sample Preparation

After biological samples are collected, cell isolation and lysis followed by nucleic acid extraction, purification, and preconcentration may be performed [63]. This “sample processing” or “sample preparation” step has been less developed than other assay steps because of its intrinsic complexity. Moreover, contamination and inhibitors for subsequent amplification steps, and nucleic acid degradation are also critical and influence diagnostic testing as these factors impede quantitative assessment of the analyte in question, leading to misinterpretation of results [62, 64]. Therefore, sample preparation tends to be performed off-chip (using laboratory equipment such as centrifuge), while amplification and detection can be accomplished in microfluidic systems [65]. Nonetheless, efforts have been made
to integrate cell capture, cell lysis, and nucleic acid purification into a single microfluidic device [66]. Integration of sample pretreatment with analysis could lead to improvements in sensitivity (as less sample is lost in between steps) and convenience [5]. One of the first microfluidics-based DNA purification procedures has been demonstrated using silicon microdevices, taking advantage of high-aspect ratio features to increase the capture surface area and hence the loading capacity [67]. Other work has focused on single-cell mRNA extraction and analysis via cDNA synthesis on a microfluidic chip [66]. The Cepheid’s GeneXpert has employed a macrofluidic approach integrated with a miniaturized sonicator [68]; this system has shown success in detecting drug-resistant tuberculosis in the developing world [69]. In another approach, Claremont BioSolutions has developed a miniaturized bead blender for sample preparation.

1.4.3.2 Signal Amplification

Since the amount of nucleic acid acquired from either the preparation step or from the raw sample is usually low for immediate identification and quantification [60], a method of amplification is needed to obtain a sufficiently nucleic acid detection signal. The most common technique is the polymerase chain reaction (PCR), for which miniaturization promotes the ability to reduce the reagent consumption, reduce the cycle time, and automate the process [5]. PCR requires thermal cycling for the reaction. Miniaturization of PCR provides many advantages, such as decreased cost of fabrication and operation, decreased reaction time for DNA amplification, reduced cross talk of the PCR reaction, and ability to perform large numbers of parallel amplification analyses on a single PCR microfluidic chip. Also, microfluidics allows for increased portability and integration of the PCR device. One of the first silicon-based stationary PCR chip was described several years after the introduction of PCR itself [70]. Since then, many research groups began to develop microdevice-based PCR devices. Most of these devices are based on silicon and glass, but more recently, polymer materials such as PDMS [71], PMMA, polycarbonate, SU-8, polyimide, poly(cyclic olefin), and epoxy are being used. HandyLab (HandyLab, Inc., is now part of Becton Dickinson) has also developed a disposable microfluidic chip that implements heat and pressure gradients to move microliter-sized plugs via valves and gates through different temperature zones within the chip [49, 72]. Flow-through designs also decrease the possibility of cross-contamination between samples as well as allow for the incorporation of many other functions, which is appealing as it leads toward the development of a micro total analysis system. As PCR is a temperature-controlled, enzyme catalyzed biochemical reaction system, the method in which the different temperature zones are generated and maintained is crucial to the design of PCR microfluidics. Various heating methods have been employed, and they can be broadly categorized into contact and noncontact heating methods. Contact heating has been implemented in PCR microfluidics using integrated thin-film platinum resistors as both the heating and sensing elements on these chips [73] and noncontact methods include hot-air
cycling [74]. Both methods have low power consumption and are amenable for use at the point-of-care.

By contrast, isothermal techniques have been developed to perform nucleic acid amplification without thermal cycling. This feature removes the need of using different temperatures by making use of enzymes to perform reaction at a single temperature which translates into less complex and cheaper instrumentation that is amenable for point-of-care use [25, 69]. Current commercially available platforms [25], for example, helicase-dependent amplification (HDA; BioHelix), transcription-mediated amplification (TMA; Gen-Probe), nucleic acid sequence-based amplification (NASBA; BioMerieux), strand displacement amplification (SDA; Becton Dickinson), loop-mediated amplification (LAMP; Eiken), and recombinase polymerase amplification (RPA; TwistDx) are implemented in laboratory-based NAT and microdevices for point-of-care.

1.4.3.3 Product Detection

Amplified product can be examined at the end of reaction (end-point detection) or during the reaction (real-time detection). There are many techniques to detect nucleic acids, which one of the primary methods is quantifying nucleic acids by ultraviolet light (260 nm) absorption. Nonetheless, fluorescence-based techniques remain the most commonly employed due to its high level of sensitivity and low background noise [64]. Fluorescent dyes can either bind to nonspecific locations (general interactions) or specific locations of molecule, depending on the application. Cepheid Inc., for example, has developed a real-time fluorescent PCR detection (fluorescently labeled probes), requiring instruments that may be used in some but not all point-of-care settings. Cepheid’s GeneXpert test platform (Fig. 1.5a) has been tested for clinical trials in four developing countries and showed promising results for detecting tuberculosis [69]. However, this system requires uninterrupted and stable electrical power supply and annual validation of the system, and generates considerable more waste than microscopy technique [75] which may lead to problems of waste management in resource-limited settings.

At the point-of-care, electrochemical methods may also be suitable due to their compatibility with low-cost and portable analyzers [60]. For example, Nanosphere is building a scanner-based detector to detect DNA via nanoparticle probes [60]. Gold nanoparticles functionalized with oligonucleotides are used as probes for DNA sequences complementary to the sequences of those oligonucleotides (Fig. 1.5b). Visualization of the gold nanoparticle, and hence DNA content, is performed using a signal amplification method in which silver is reduced at the surface of the gold, and a scanner is used to measure amount of light scattered.

Lateral flow devices for end-point detection have been used for simple and low-cost method. A DNA strip using specific probes functionalized on membrane to capture target sequences from amplification reaction can be used for identification of target strains. GenoType (Hain Lifescience GmbH) and INNO-LiPA (Innogenetics NV) have developed assays to detect infectious diseases. BioHelix has developed a strip cassette for the isothermal amplification assay (Fig. 1.6).
Fig. 1.5 (a) Cepheid’s GeneXpert test platform (right) integrates sample processing and PCR in a disposable plastic cartridge (left) containing reagents for cell lysis, nucleic acid extraction, amplification and amplicon detection. This system has been used to detect drug-resistant tuberculosis cases (Taken from www.cepheid.com with permission from Cepheid, Inc.) (b) Nanosphere Verigene system. Picture of a Nanosphere cartridge with loaded reagent wells (left). Picture of Verigene readers with touch-screen control panels (right) (Taken from www.nanosphere.us with permission from Nanosphere, Inc.)

Other commercial effort toward a point-of-care nucleic acid device includes Spartan RX CYP2C19, by Spartan Bioscience, which is the first point-of-care genetic testing system to receive the CE marking from the European regulatory health authorities [53]. T2 Biosystems is based on magnetic resonance signal detection from nanoparticles coupled with target analytes which does not require sample purification; therefore, rapid turnaround time can be achieved.
1.4.4 Clinical Chemistry Assays

Clinical chemistry encompasses the measurement of a wide range of blood parameters, including gases, electrolytes, hemoglobin, pH, enzymes, metabolites, lipids, hormones, vitamins and trace factors, inflammatory markers and cytokines, coagulation proteins, therapeutic drugs, and drugs of abuse [12,76]. Frequent sampling and fast diagnostic monitoring (in situations such as intensive care units and operating rooms) drive the need for clinical chemistry tests at the point-of-care [12]. In addition, POC testing for clinical chemistry markers can be more accurate by avoiding analyte changes during sample transport to laboratories. These analyte changes can be caused by the delayed release of analytes (e.g., release of K$^+$ from red blood cells during refrigerated storage), by continued metabolism (e.g., decrease in glucose and pH and increase in lactate from active red blood cells under hypoxic conditions), and by protein/peptide degradation in whole blood [76].
Electrochemical detection methods are popular for quantitative analysis of blood parameters [77]. The iSTAT device from Abbott Laboratories (Abbott Park, IL, USA), which uses microfabricated thin-film electrodes to measure levels of electrolytes, general chemistries, blood gases, and hematocrit, has an electrochemical detection system which includes amperometry, voltammetry, and conductance, depending on the analyte [5]. The iSTAT device uses disposable cartridges for performing sample preconcentration and separation, and a portable instrument for electrochemical detection [12]. Conventional glucose meters are also based on electrochemical assays [77, 78]. However, devices that feature electrochemical detection have several disadvantages with regard to POC testing. Electrodes are required to be integrated onto the disposable, increasing per-test costs; detecting hundreds of simultaneous reactions at once is difficult to achieve with electrodes; and incompatibilities exist between capabilities of existing microelectronics fabrication facilities and specifications of sensors, with regard to dimensions, materials, passivation layer, and others [5, 77]. With the increasing availability of cheap, high-quality optoelectronic components such as CCDs and laser diodes, optical detection may be more suited for POC testing of blood chemistry parameters. For situations where semiquantitative analysis is sufficient, paper-based microfluidic devices may be attractive [78, 79].

1.4.5 Hematology

The most widely tested hematology parameters in point-of-care situations are hemoglobin and hematocrit (for anemia, red blood cell transfusion therapy, and acute hemodilution during surgery), and coagulation/clotting time (for cardiac surgical and catheterization procedures) [80]. Sphere Medical (Cambridge, United Kingdom) has developed a microanalyzer technology with silicon chips for signal processing and sensing based on three transducer technologies: potentiometry, amperometry, and conductimetry. Different membranes and receptor materials are used to cover the sensors to detect specific analytes. The technology can measure hematocrit and small molecule analytes.

Activated clotting time, activated partial thromboplastin time, and prothrombin time are monitored for high-dose heparin therapy (a blood clot inhibitor) and blood-thinning medications (such as warfarin). Companies with POC instruments for monitoring coagulation include Medtronic (Minneapolis, MN, USA), Abbott Diagnostics (Abbott Park, IL, USA), Roche Diagnostics (Indianapolis, IN, USA), and Hemosense (owned by Alere) (Waltham, MA, USA).

1.5 Outlook

Over the last decade, microdevices for point-of-care testing have generated great progress in both academic and industrial laboratories. Several challenges remain before these technologies can be translated to be effective POC tests, including
the need to miniaturize all peripheral instrumentation. An important recent trend is the increased attention in developing POC tests for the developing world [4, 5]. Although targeted toward use in emerging and low-income countries, the design constraints are remarkably similar across the world (Fig. 1.1). Trends toward decentralized healthcare and personalized medicine ensure the importance in the future of microdevices for point-of-care testing.

References

1. P.N. Floriano, Microchip-based assay systems: methods and applications: (Humana Press, Totowa, 2007)
Abstract The miniaturization of laboratory functions onto microfluidic chips is leading a paradigm shift in biotechnology, analogous to the transformation of electronics by the integrated circuit (IC) 50 years ago. The microfabricated pipes, pumps, valves, and mixers of microfluidics enable small volumes of reagents, samples, and individual living cells to be controlled on low-cost, portable chips. However, a microfluidic chip that can be programmed to perform the wide range of chemical and biological tasks required for medical and scientific analysis, akin to a microprocessor in electronics, remains a challenge. Here, we review work done by our group to develop hybrid IC/microfluidic chips that can simultaneously control thousands of living cells and picoliter volumes of fluid, enabling a wide variety of chemical and biological tasks.

2.1 Introduction

Advances in microfluidic technologies are revolutionizing the way medicine and biology are approached. Ever smaller samples are being manipulated and analyzed more quickly and with greater accuracy than ever. As these performance metrics
improve, new microfluidic applications in enzymatic assays, genomics, proteomics,
and clinical pathology are becoming feasible [1–5].

Combining microfluidics with the power of electronics is a recent and growing
trend empowering many new biological and medical applications. Microelectronics
offer new and exciting ways to position and analyze cells and fluids [6–9]. They
enable tremendous miniaturization, greater accuracy, discrete sample handling, and
the ability to perform hundreds or even thousands of functions in parallel.

Integrated circuits (ICs) have the ability to empower even further miniaturiza-
tion and complexity of biomanipulation and analysis devices. Integrated circuits
combine the power of programmability with CMOS technology to fit billions
of transistors on a single chip. Hybrid integrated circuits/microfluidic chips are
capable of performing intricate manipulations and analysis on single cells and small
chemical volumes [5, 10–17]. These devices pave the way for a new generation
of miniaturized biomedical experiments which can be performed on single drops of
human physiological fluids to diagnose illness and disease in a point-of-care setting.

Currently, the analysis of human physiological fluids is an often unpleasant and
slow process. A large and painful syringe is first used to collect blood samples,
after which the sample must be sent to a laboratory for analysis, delaying the
results by days [18]. The vision driving integrated circuits/microfluidic technology
is to take hundreds of existing tests and perform them quickly and simultaneously
using a single nanoliter droplet of body fluid. The tests will be completed at the
patient’s bedside and, in minutes, empowering doctors with real-time knowledge of
the patient’s condition. The tests can also be performed in resource-limited settings
such as developing countries, remote locations, and war zones.

Figure 2.1 shows an illustration of a hybrid integrated circuit/microfluidic chip
for point-of-care diagnostics. Cells and reagents enter the device through inlets.
Single cells and droplets are pinched off into the microfluidic chamber. The
integrated circuit then positions the cells and droplets using electric fields. Sensors
built into the integrated circuit analyze various properties of the cells and droplets
and monitor the outcomes of chemical reactions.

Numerous new technologies and methods are needed to make integrated circuits
for point-of-care diagnostics a reality. Many have already been developed and many
more are in the pipeline. This chapter presents an overview of the current state of
the field and a vision of future work needed to make ICs a reality in point-of-care
diagnostics. Table 2.1 below summarizes functions which can be performed and on
IC/microfluidic chips.

2.2 Principles of Dielectrophoresis

Most functions listed in the summary table above make use of electrical phenomena
to manipulate and analyze droplets and cells. Dielectrophoresis is the bedrock
phenomenon used to transport droplets. Electroporation and dielectric heating are
Fig. 2.1 An illustration of a hybrid integrated circuit/microfluidic chip for point-of-care diagnostics. The device contains an integrated circuit with a microfluidic chamber built on top. Cells and reagents enter through the inlets. Single cells and droplets are pinched off into the microfluidic chamber where they are positioned and analyzed used to release cell contents, merge droplets, and heat the chip. All make use of alternating electric fields in different frequency bandwidths.

Dielectrophoresis (DEP) is a phenomenon where dielectric objects are attracted to electric field maxima. The object’s polarizability relative to the surrounding medium determines the strength of DEP. It comes in two flavors, positive DEP and negative DEP, as seen in Fig. 2.2. With positive DEP, a dielectric particle is more polarizable than the surrounding medium and is attracted to the electric field maximum. This is shown with the upper particle of Fig. 2.2. In negative DEP, seen at the bottom of Fig. 2.2, the particle is less polarizable than the surrounding medium and is pushed away from the electric field maximum. This is analogous to a helium-filled balloon. Even though gravity is pulling it down, it floats upward because it is less dense than the surrounding air.

The force generated by dielectrophoresis is described by the following equations [19]:

\[
\text{force} = \frac{\epsilon \Delta \varepsilon E^2}{2} \frac{\partial^2 \Phi}{\partial x^2}
\]

where \(\epsilon\) is the dielectric constant, \(\Delta \varepsilon\) is the difference in polarizability between the particle and the medium, \(E\) is the electric field, and \(\Phi\) is the electric potential.
Table 2.1 Functions performed on integrated circuits for medical diagnostics

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
<th>Diagram</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transport / Programmable Channels</td>
<td>Droplets of physiological fluids, reagents, and cells are moved using dielectrophoresis and magnetophoresis</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Deform</td>
<td>Electric and fields deform droplets by pulling them from different directions.</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Porate</td>
<td>Controllable mixing between the interior and exterior of cells and droplets.</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Merge</td>
<td>Cells and droplets are merged with reagents to initiate chemical reactions.</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Temperature Control</td>
<td>Droplets and cells are heated by applying a microwave frequency electric field (1-3GHz).</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Temperature Sensor</td>
<td>Temperature is monitored to initiate or terminate temperature-dependent reactions.</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Capacitance Sensor</td>
<td>Capacitance measurements monitor electrical properties of cells and droplets.</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Droplet Maker</td>
<td>Droplets of external fluid are dispensed onto the chip using a droplet maker.</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Mix</td>
<td>Mixing to initiate chemical reactions.</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Imaging</td>
<td>Fluorescence and color detection monitors reaction outcomes.</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
</tbody>
</table>

\[
F_{DEP} = 2\pi \varepsilon_m r^3 CM(\omega) r^2 \sqrt{E_{RMS}} \tag{2.1}
\]

\[
CM(\omega) = \text{Re} \left( \frac{c_p - c_m}{c_p + 2c_m} \right) \tag{2.2}
\]

where \( F_{DEP} \) is the dielectrophoretic force, \( \varepsilon_m \) is the permittivity of the surrounding medium, \( r \) is the particle’s radius, \( CM \) is the Clausius–Mossotti factor, \( E \) is the
Fig. 2.2 An illustration of positive and negative dielectrophoresis (DEP). The top object is undergoing positive DEP, being pulled toward the electric field maxima. Its dielectric constant is greater than that of the surrounding medium. The bottom object is being pushed away from the maximum by negative DEP. Its dielectric constant is less than that of the surrounding medium.

Fig. 2.3 A plot of three frequency regimes for vesicles. In the 1-Hz–1-kHz range, electroporation and electrofusion dominate. In the 1-kHz–10-MHz range, dielectrophoresis (DEP) dominates. Above 1 GHz, microwave heating dominates [10].

electric field, and \( \varepsilon_p \) is the particle’s permittivity. The strength and sign up the DEP force depends heavily on the frequency of the electric field, which is embedded in the complex permittivities in the Clausius–Mossotti factor. DEP for cells floating in water is optimal at a frequency of approximately 1 MHz. This is illustrated in Fig. 2.3 with the curve labeled “DEP.”

As shown in Fig. 2.3, different functions can be performed on cells and vesicles at various electric field frequencies. At low frequencies, cells can be porated and fused. The 1-Hz–1-kHz range destabilizes vesicles by creating a potential across their membranes. A large enough potential causes dielectric breakdown of the
membrane. At higher frequencies in the 1-kHz–10-MHz range, DEP dominates over electroporation and electrofusion. The electric field is switching faster than the vesicle membrane’s time constant, preventing voltage from building up across it. At even higher frequencies in the 1-GHz range, microwave heating becomes dominant.

2.3 Integrated Circuit/Microfluidic Chips

Integrated circuit/microfluidic chips contain displays of electrical pixels which are used to trap and position thousands of cells and droplets. Activating a pixel produces an electric field which dielectrophoretically attracts and traps cells. By simultaneously activating thousands of pixels, large numbers of cells can be precisely trapped and positioned.

Integrated circuit/microfluidic chips have undergone an evolution in the past decade, progressively increasing in complexity and incorporating more functions. Table 2.2 details the evolution of these chips – beginning with built-in-house DEP chips containing only 25 pixels and leading to present CMOS devices capable of producing both electric and magnetic fields and containing up to 32,000 pixels.

The second generation of hybrid IC/microfluidic chips shown in Table 2.3 uses CMOS integrated circuit technology. The chip is shown in Fig. 2.4 and consists of a display of 128 × 256 electrical pixels [14]. Each pixel is individually addressable and contains a single SRAM memory element. The inset in Fig. 2.3 shows two active pixels radiating electric fields and polarizing and trapping a nearby dielectric particle. Multiple pixels can be activated simultaneously. The pixels are 11 × 11 μm² in area, about the size of a living cell, and the entire chip is 2.3 × 3.3 mm². The pixels are covered in a 2-μm thick layer of polyamide insulator to prevent short circuiting.

Active pixels are charged by an AC square wave running between 0 and 5 V. The frequency is set between 1 Hz and 1.8 MHz. Inactive pixels are also connected to an AC square wave, but are run out of phase with the active pixels. This creates electric field maxima at the interface between active and inactive pixels – this is where cells are trapped by DEP.

The third-generation chip, shown in Fig. 2.5, incorporates larger voltages and the ability to generate magnetic fields [6]. Larger voltages generate stronger DEP forces which position cells faster and more reliably. Higher voltages also allow cells to be electroporated which enables cell fusion and the transport of foreign particles across the cell membrane. Magnetic fields are generated to enable functionalized magnetic particles – often used in the biomedical community – to be used with this chip.

This chip is a CMOS integrated circuit with an array of 60 × 61 electric pixels, as seen in Fig. 2.5b. Each pixel is individually addressable and can be activated to 0 or 50 V. Similar to the second-generation chips, pixels are activated by sending an AC electrical signal. Inactive pixels are run out of phase, creating electric field maxima at the interface between active and inactive pixels. This is where cells are trapped by DEP. This chip can also generate magnetic fields using a 60 × 60 array of
Table 2.2 Progression of integrated circuit/microfluidic chips

<table>
<thead>
<tr>
<th>Chip Chip</th>
<th>Field</th>
<th>Number of pixels</th>
<th>Technology</th>
<th>Picture</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEP Chip</td>
<td>Electric</td>
<td>5 x 5</td>
<td>Built in-house</td>
<td>![Picture]</td>
</tr>
<tr>
<td>1st generation [15]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd generation [14]</td>
<td>Electric</td>
<td>128 x 256</td>
<td>CMOS</td>
<td>![Picture]</td>
</tr>
<tr>
<td>3rd generation [11]</td>
<td>Electric and Magnetic</td>
<td>60 x 61 electric 60 x 60 magnetic</td>
<td>CMOS</td>
<td>![Picture]</td>
</tr>
</tbody>
</table>

Table 2.3 Characteristics of the second-generation DEP chip

<table>
<thead>
<tr>
<th>Process</th>
<th>MOSIS TSMC 0.35 μm gate length 2PM4 process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pixel size</td>
<td>11 x 11 μm</td>
</tr>
<tr>
<td>Pixels</td>
<td>128 x 256</td>
</tr>
<tr>
<td>Chip size</td>
<td>2.3 x 3.3 mm²</td>
</tr>
<tr>
<td>Addressing</td>
<td>8-bit word line decoder, 128-bit, two-phase clocked shift register for bit lines</td>
</tr>
<tr>
<td>Transistor count</td>
<td>&gt;360,000</td>
</tr>
<tr>
<td>Pixel voltage</td>
<td>V = 3-5 V, DC = 1.8 MHz</td>
</tr>
<tr>
<td>Operating current</td>
<td>30–100 mA</td>
</tr>
</tbody>
</table>

Fig. 2.4 The second generation of hybrid IC/microfluidic chip incorporates a CMOS integrated circuit. This chip contains a display of 128 x 256 pixels, each individually addressable and 11 x 11 μm² in area. The entire chip is 2.3 x 3.3 mm² large. The inset shows two active pixels producing electric fields and attracting a nearby dielectric particle.
Fig. 2.5 (a) The third-generation chip consists of an array of $60 \times 61$ pixels and $60 \times 60$ wires. (b) The pixels are charged to 0 or 50 V to generate electric fields. (c) The wires carry current to generate magnetic fields. Magnetic field maxima occur where two orthogonal wires overlap.

wires, shown in Fig. 2.5b. The wires are activated by running a 120-mA DC current through them. Magnetic field maxima are generated where two perpendicular wires overlap, creating a trap for magnetic particles [13].

2.4 Functions Performed by Hybrid Integrated Circuit/Microfluidic Chips

Many functions have been implemented on integrated circuits for medical diagnostics. Droplets and cells can be transported, deformed, porated, merged, and heated [10–15]. Virtual microfluidic channels, defined by electric field boundaries, have also been implemented. Temperature and conductance sensors have also been built into IC’s [14, 17]. These functionalities are described in further detail in the remainder of this chapter. Several important operations have also yet to be implemented on an integrated circuit platform for diagnostics. These include droplet making, mixing, and color and fluorescence sensing. These will round off the remainder of this chapter.

2.4.1 Positioning/Programmable Channels

Positioning cells and droplets, shown in Fig. 2.6, is a critical function underpinning the usefulness of integrated circuit technology for medical diagnostics. Reagents must be transported from storage areas onto the chip and mixed and reacted with
blood, urine, cells, and other human samples. The resulting reactants are positioned over sensing regions of the chip for analysis. The positioning process must be capable of multiplexing hundreds of droplets simultaneously and without collisions to allow multiple diagnostic tests to be performed in parallel.

Figure 2.6 shows cells and droplets being trapped and positioned using a hybrid integrated circuit/microfluidic chip. A cell is trapped over an active pixel and positioned by shifting the pixel’s voltage onto neighboring pixels [14]. Cells and droplets follow the electric field maximum as it moves from pixel to pixel. Figure 2.6
Virtual microfluidic channels are created using trains of moving pixels. Every tenth pixel in a line is active, with each one transporting several cells. Linear channels can also be combined with junctions to create the branching configurations seen on the right.

Fig. 2.7 Virtual microfluidic channels are created using trains of moving pixels. Every tenth pixel in a line is active, with each one transporting several cells. Linear channels can also be combined with junctions to create the branching configurations seen on the right.

shows (a) two similar cells being moved by the second-generation integrated circuit and (b) two different strains of cells moving over the integrated circuit. An active pixel potentiated to 5 V creates approximately 5 pN of force on a cell, transporting it at up to 300 μm/s. The cells are exposed to electric fields of up to 50 kV/m at 1 MHz, well within the range where they remain healthy. Thousands of cells were also positioned simultaneously to form a complex and well-defined structure, as seen in Fig. 2.6c. The phrase “lab on a chip” is spelled using yeast cells.

Virtual microfluidic channels were created using the hybrid integrated circuit/microfluidic chip. Figure 2.7 shows two such arrangements where three channels are run in parallel and where one channel is branched into four. Cells are carried by trains of pixels where active pixels are separated by several inactive pixels. In Fig. 2.7, every tenth pixel was active. The active pixels each move in the same direction and each one carries several cells. Virtual microfluidic channels are created similarly to videos on the chip. A video file contains a number of frames where a train of pixels moves one position with each subsequent frame. Each video encodes a different configuration of channels, allowing a single chip to perform many tasks simply by playing a different video.

Thousands of cells can also be moved simultaneously on the chip. Figure 2.8 shows frames from a video where thousands of yeast cells were positioned to move like a dancer. The dancer is approximately 1.5 mm tall. Active pixels are located inside the dancer, drawing cells to the edge of the dancer where the electric field is strongest. This video was created by sending a GIF video file to the chip. A new frame is displayed on the chip several times every second, moving the cells as each new frame is displayed. Figure 2.8 shows frames 5 s apart with one additional frame inserted at 0:17 s to show the dancer’s full range of motion.
A video is played on the chip where thousands of yeast cells are moved in the shape of a dancer. The video is sent to the chip as a GIF file.

2.4.2 Deform

The chip can be used to deform vesicles into different shapes, shown in Fig. 2.9 [10]. This has important implications for single-vesicle and single-cell rheology where microscopic control is essential for studying these complex systems. It represents an alternative to microcontact patterning [20] and other existing methods for single-cell studies. Figure 2.9 shows vesicles being stretched, compressed, and being deformed into various shapes with flat edges. The first column in the figure shows the active pixels shaded. The second column shows the results of electric field simulations mapping the field amplitude where brighter red colors represent stronger fields. The third column shows images of vesicles being deformed. The vesicles are approximately 50 μm in diameter and can be stretched, compressed, and deformed into squares, diamond, and hexagons.
The hybrid integrated circuit/microfluidic chip is used to deform a vesicle containing magnetic particles, shown in Fig. 2.10 [10, 11]. The chip combines both an array of electric pixels and an array of magnetic wires. The vesicle is first trapped dielectrophoretically by activating the pixels below it, while a magnetic field traps the iron oxide particle. The magnetic field maximum is moved away from the vesicle, dragging the magnetic particle and elongating the vesicle. Once
the magnetic field is turned off, the vesicle contracts back to its original shape. This chip can be useful in single-cell rheological studies for point-of-care diagnostics.

### 2.4.3 Porate

Poration is used to release vesicle contents, as seen in Fig. 2.11. Electroporation involves using electric fields to destabilize the vesicle membrane and create small, short-lived holes [21]. The holes are large enough to allow macromolecules to enter and leave the vesicle. AC electric fields in the 1-Hz–10-kHz range induce transmembrane voltages large enough to cause a dielectric breakdown of the membrane, thus destabilizing it [21, 22]. In Fig. 2.11, a vesicle was placed in a fluid containing fluorescein, a fluorescent molecule [10, 12]. At time $t = 0$ s, the vesicle was electroporated, allowing fluorescein to enter. In a second experiment, a vesicle was prefilled with fluorescein and submerged in water. It was similarly electroporated, diffusing away its fluorescein and losing its fluorescence.
Fig. 2.11  Vesicles were electroporated using a hybrid integrated circuit/microfluidic device. At \( t = 0 \text{s} \) in the upper left corner, a vesicle surrounded by fluid containing fluorescein, a fluorescent molecule, was electroporated. The molecule entered the vesicle, making it fluoresce at \( t = 8 \text{s} \). At time \( t = 0 \text{s} \) in the lower right corner, a fluorescent vesicle was electroporated. The fluorescein diffused into the surrounding fluid.

Fig. 2.12  A cell is electroporated using the hybrid integrated circuit/microfluidic chip. The cell membrane is normally impermeable to trypan blue, a dye commonly used to stain cells, but after electroporation, it enters and darkens the cell.

Cells can also be electroporated to insert contents into the cell and to release its contents, as seen in Fig. 2.12. This is a critical function enabling biological vectors and functionalized particles to be introduced into the cell and to analyze the cell contents [22–24]. Cell electroporation is demonstrated by placing a yeast cell on the hybrid integrated circuit/microfluidic chip in a solution containing trypan blue, a dye commonly used for staining biological samples [10, 12]. The cell membrane is impermeable to trypan blue. The cell was trapped above a
pixel using DEP, after which the electric field was switched to a lower frequency, triggering electroporation. After 40 s, the trypan blue had entered and stained the cell, confirming that electroporation had indeed occurred.

### 2.4.4 Merge

Figures 2.13 and 2.14 demonstrate how a simple fluid droplet can be split into two, with one individually manipulated, and then rejoined into one. Merging droplets is an essential function for performing chemical reactions on the chip. A sample and multiple reagents are each contained in individual droplets. The reaction is instigated by merging the droplets together, bringing the different reagents into contact with the sample. Droplet merging is performed by moving the active pixels under each droplet together, as seen in Fig. 2.13 [10, 12]. When the two neighboring droplets come into contact, they merge into a single, larger drop. After completion of a reaction, the analytes are transported to an area of the chip with sensing capabilities.

Cell fusion is a critical technique enabling cellular reprogramming, cloning, and hybridoma formation [25–31]. A method for fusing cells using a hybrid integrated circuit/microfluidic chip was developed and is shown in Fig. 2.15. The chip traps and moves cells using dielectrophoresis, positioning the cells into pairs in preparation for cellular fusion. Fusion was performed using electroporation or by introducing PEG (polyethylene glycol) into the system. Our technique has the advantage of precise control over every cell, allowing perfect pairing of cells every time and thus significantly increasing the yield. Fusion was performed in three different configurations: An individual pair of cells was isolated and fused, several pairs of cells were isolated and fused, and thousands of cell pairs were fused simultaneously.
Cells were fused both stochastically and deterministically—stochastic pairing results in random cells being fused together, while deterministic pairing ensures that each pair contains one cell from both cell lines. Viable fused cells with a mean yield of 89% were observed. Figure 2.15a shows multiple pairs of cells being aligned in rows and fused, and Fig. 2.15b shows cells being arranged in a circle for fusion.
2.4.5 Temperature Sensor

Temperature sensors are an important component for integrated circuits to become viable in medical diagnostics. Cells must often be viable for extended periods of time, requiring on-chip incubation and temperature monitoring. Analytes and physiological fluids must frequently be maintained at specific temperatures for proper chemical activation. Biochemical reactions are also often initiated and terminated by temperature changes—a well-known example being polymerase chain reaction (PCR). To this end, recent IC/microfluidic chips have had multiple temperature sensors integrated into their architecture, allowing the temperature to be monitored independently over different regions of the chip [12]. They consist of micron-scale thermistors built under the IC surface. They work by monitoring temperature-dependent resistance changes and calibrating them to a temperature scale. A feedback loop with an external cooling apparatus then maintains the chip at the desired temperature.

2.4.6 Microwave Dielectric Heating

Microwave dielectric heating has been developed for warming small, micron-sized droplets in oil [32]. Electric fields in the 1–3-GHz bandwidth— in the microwave spectrum—are transmitted across droplets, causing rapid and well controlled heating of the droplets without heating the surrounding environment directly. The advantage of heating such small-length scales is the high surface-to-volume ratio, allowing rapid thermal cycling in the 15-ms range. The technology is scalable onto an integrated circuit platform, with each individual pixel acting as a microwave heater. This enables the highly localized heating and incubation of droplets and cells.

2.4.7 Capacitance Sensor

Sensing the capacitative coupling between a pixel and an object on top can be used to inform the device where droplets and cells are located. This is seen in Fig. 2.16. This is critical for giving the chip feedback capabilities and is a method for error checking that a cell or droplet indeed moved to the position it was instructed. Capacitative sensing can eliminate the need for the hybrid integrated circuit/microfluidic chip to be placed under an optical microscope. In Fig. 2.16, capacitative sensors were embedded under every pixel, giving high-resolution images of cells and beads resting on the chip’s surface [17]. The sensors discerned differences in electrical properties in the medium directly above the chip, and mapped them as grayscale images. Fig. 2.16 shows the image obtained from the chip’s embedded sensors next to an image of the same region taken with an optical microscope.
Fig. 2.16 Built-in sensors in the chip are capable of creating images by mapping electrical properties of the medium in contact with the chip’s surface. On the left-hand side is an image of beads created using this method. On the right-hand side is the same image taken with an optical microscope [17]

2.4.8 Integrating Hybrid IC/Microfluidic Chips with Other Technologies

Hybrid integrated circuit/microfluidic chips can be integrated with other technologies to increase their ability to solve problems in point-of-care diagnostics. Improved methods for intaking fluids – blood, urine, and chemical reagents – from the outside environment are critical for the success of hybrid IC/microfluidic chips. Methods to mix very small volumes of fluid are also necessary to allow reagents to properly and fully mix with the analytes. This is a difficult problem at small-size scales, and inspiration can be borrowed from electrowetting and magneto-mechanical mixing techniques.

There is a strong need for more automated and sophisticated methods of pulling fluids directly out of the environment and dispensing them as droplets on the hybrid IC/microfluidic chip. This functionality promises to make the chip more self-contained and independent, eliminating the need for technicians and external syringe pumps to deliver fluids. Electrical methods of pulling droplets from fluid reservoirs have previously been demonstrated. Electrowetting is one commonly cited method [33, 46], using electrically induced differences in the wettability of adjacent surfaces to move and split droplets [39].

Mixing is an important function in microfluidic devices, significantly accelerating the rate and accuracy of biochemical reactions. Small fluid volumes are
especially difficult to mix as they have a tendency to remain in the laminar flow regime. In this regime, flow-vector lines do not intersect, and mixing remains slow. Several methods have been demonstrated to overcome this challenge for droplets in the millimeter-size range. Such droplets were rapidly actuated between neighboring pixels using electrowetting, inducing internal flow fields [37]. Dielectrophoresis-induced flows can also be used. Another approach uses magneto-mechanical mixing [47]. Superparamagnetic beads were inserted into unilamellar vesicles, and an external magnetic field was applied, aligning the beads into chains. As the magnetic field rotated, the vesicle contents were mixed by the spinning chains. The vesicles were 10–20 μm in diameter – the same size-scale as droplets and cells used in hybrid IC/microfluidic chips.

### 2.5 Conclusions

This chapter describes the development of a versatile platform for point-of-care diagnostics using integrated circuit (IC) technology. This work is an important step toward developing automated, portable, and inexpensive devices to perform complex chemical and biological tasks. Such a device would revolutionize the way that biological and chemical information is collected for medical diagnostics, allowing doctors to make near real-time prognoses.

The hybrid IC/microfluidic chips developed thus far control living cells and small volumes of fluid. Table 2.1 summarized the basic lab-on-a-chip functions that the hybrid chips can perform. The chips can be programmed to transport, deform, porate, and merge droplets and cells; they can control temperature, sense temperature, sense capacitance, and sense color and fluorescence, and they can make and mix droplets. These basic functions can be strung together to perform complex chemical and biological tasks. The fast electronics and complex circuitry of ICs enable thousands of living cells and droplets to be simultaneously controlled, allowing many well-controlled biological and chemical operations to be performed in parallel. With ICs becoming more powerful each year and microfluidics beginning to enter the commercial arena, IC/microfluidic chips are poised to play an important role in clinical diagnostics.

### References


Part II
Detection
Chapter 3
Flow Cytometry on a Chip

Peter Kiesel, Joerg Martini, Michael I. Recht, Marshall W. Bern, Noble M. Johnson, and Malte Huck

Abstract  Flow cytometers are indispensable tools in medical research and clinical diagnostics for medical treatment, such as in diagnosing cancer, AIDS, and infectious diseases. The cost, complexity, and size of existing flow cytometers preclude their use in point-of-care (POC) diagnostics, doctor’s offices, small clinics, on-site water monitoring, agriculture/veterinary diagnostics, and rapidly deployable bio-threat detection. Here, we present a fundamentally new design for a flow cytometer for the POC that delivers high effective sensitivity without complex optics or bulky, expensive light sources. The enabling technology is spatially modulated emission, which utilizes the relative motion between fluorescing bioparticles and a selectively patterned environment to produce time-modulated signals that can be analyzed with real-time correlation techniques.

3.1 Flow Cytometry

3.1.1 Introduction

Flow cytometers are indispensable tools in medical research and clinical diagnostics for medical treatment, such as in diagnosing cancer, AIDS, and infectious diseases [1]. Chemical and physical information are obtained from functionalized microbeads or bioparticles, for example, cells, viruses, or subcellular complexes, as they are transported in a fluid stream [2]. Conventional flow cytometers can analyze microparticles at rates of \( \sim 50,000 \) per second [3] and allow for extremely sensitive measurement of particle-associated probes (<100 fluorophores per particle) [2].

P. Kiesel (✉) · J. Martini · M.I. Recht · M.W. Bern · N.M. Johnson · M. Huck
PARC (Palo Alto Research Center Incorporated), 3333 Coyote Hill Rd., Palo Alto, CA 94304, USA
e-mail: Peter.Kiesel@parc.com

D. Issadore and R.M. Westervelt (eds.), Point-of-Care Diagnostics on a Chip, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-642-29268-2_3, © Springer-Verlag Berlin Heidelberg 2013
Higher analysis rates are limited by detector sensitivity, data acquisition electronics, and cell coincidences. The stochastic arrival of particles in the detection volume limits their concentration to avoid an intolerable number of coincidences.

The cost, complexity, and size of existing flow cytometers preclude their use in point-of-care (POC) diagnostics, doctor’s offices, small clinics, on-site water monitoring, agriculture/veterinary diagnostics, and rapidly deployable biothreat detection. The conventional design of flow cytometers is not readily extendable to applications where high performance, robustness, compactness, low cost, and ease of use are required in a single instrument. To date, all fluorescence-based flow cytometers employ the same basic optical configuration, namely, intense illumination of the bioparticle as it speeds through a highly localized light spot, generally generated by a laser [2, 4], an elaborate arrangement of precision optics, and sensitive detectors to record fluorescence and scattered light.

The excitation region covers usually the lateral width of the flow channel and expands some tens of micrometers along the flow direction. A number of commercially available flow cytometers use multiple excitation sources, each focused on a well-defined location or region separate from the others.

The detection region(s) in flow cytometers are commonly defined by – not necessarily diffraction limited – confocal light collection optics with high numerical aperture lenses. Light emitted from each source’s region is typically analyzed with a series of dichroic beam splitters, filters, and photomultiplier tubes (PMTs) in order to detect and distinguish differently stained particles including those that simultaneously carry multiple dyes.

The exciting light spot, the detection area, and the particle stream need to reliably overlap in any flow cytometer. Therefore, the size, position, and flow speed of the particle stream need to be accurately controlled, which is typically realized by hydrodynamic focusing. A common implementation of flow focusing is the use of sheath flow, where buffer liquid surrounds the analyte and thereby effectively dilutes the sample, lines up the particles, prevents channel clogging, and maintains clean channel walls. However, the sheath-flow flux can be thousands of times higher than the analyte flux. Therefore, the necessity for large amounts of sheath liquid and waste makes the use of sheath flow impractical for POC testing.

In sheath-flow systems, particles travel at a speed of up to several meters per second resulting in transit times of microseconds. This requires the use of expensive, high-power, low-noise lasers and high-speed data systems, which increases the cost and power requirements of a flow cytometer. In addition, since the detection region is small and the objects traverse it rapidly, such flow cytometers have serious signal-to-noise ratio (SNR) limitations for weakly fluorescing cells. These limitations become more acute if multiple targets must be characterized and distinguished for counting or sorting.

A major cost associated with the use of flow cytometers applied for clinical diagnostics applications is the cost of reagents (e.g., antibodies and conjugated dyes). There are two ways to reduce the amount of consumables: first, one can reduce the required amount of analyte (e.g., by employing microfluidic techniques), and second, one can reduce the amount of consumable per analyte volume which requires improved signal-to-noise discrimination.
In the following, we will discuss various approaches to resolve the drawbacks of conventional flow cytometers regarding their use for POC diagnostics.

### 3.1.2 Miniaturized Flow Cytometers and Microfluidic-Based Approaches

In recent years, a number of scaled-down “high-end” flow cytometer instruments and microfluidic-based devices for POC diagnostics have been developed with the promise of portability and reduced cost. The main driver for this development is the urgent need to perform CD4 T-lymphocyte counts in resource-limited settings. This is required for screening, initiation of treatment, and monitoring of HIV-infected patients [5].

Flow-based analysis with instruments such as the FACSCount™ (Becton Dickinson), EPICS XL/MCL™ (Beckman Coulter), Guava EasyCD4™ (Millipore/Merck), PointCare NOW™ (PointCare Technologies), or CyFlow CD4™ (Partec GmbH) is the established method for CD4 counting. However, these are still quite expensive and sophisticated instruments, requiring a lab environment and skilled operators. In addition to the established instruments, there are several under development or have recently entered the market that are especially designed for POC testing (e.g., from Axxin Ltd., Alere, Daktari Diagnostics, mBio Diagnostics, Partec and Zyomyx). These instruments are discussed in greater detail in a recent review article by Boyle et al. [6]. They are fast, quite robust, use a disposable cartridge, and need only a small amount of analyte (e.g., finger prick of whole blood). From this list, the recently introduced CyFlow miniPOC from Partec is the only instrument that represents a miniaturized flow cytometer, tailored to the needs of CD4 counting in the field. The others use different detection schemes to provide the cell count.

The CD4 test reader from Axxin measures the concentration of a cell-associated CD4 protein in whole blood rather than actually counting the CD4 cells. The PIMA™ instrument from Alere is based on cell capturing, microfluidic sample processing, and digital dual-color fluorescence image analysis. The device under development by Daktari Diagnostics uses specific cell capturing in a microfluidic cartridge functionalized with CD4 antibody. Differences in binding affinity and effective shear forces are used to differentiate between lymphocytes and monocytes. The CD4 count is determined by measuring the impedance change induced by the released ions after lysing the captured CD4 cells. The SnapCount™ instrument developed by mBio Diagnostics is based on immunostaining and cell capturing in combination with a two-color fluorescence imaging system. The instrument-free CD4 counting device developed by Zyomyx uses a sedimentation technique based on functionalized magnetic beads. The CD4 count can be determined by eye from a scale reading on the tubing.

Besides the commercial developments discussed above, there are many very promising concepts and technologies for POC diagnostics discussed in the literature.
In the following, we will focus on techniques suitable for on-the-flow analyte characterization. Concepts relying on specific cell capturing, for example, in functionalized microfluidic devices with automated electrical or imaging detection will not be addressed.

Over the past decade, many concepts have been developed to simplify and miniaturize on-the-flow analyte characterization by utilizing microfluidic channels that integrate fluidic handling (e.g., pumping, valves), on-chip sample preparation (e.g., mixing), particle manipulation (e.g., flow focusing, on-chip sorting), and miniaturized optics [7, 8].

Several recent review articles give an excellent overview of microfluidic flow cytometer technologies [9–11]. We will concentrate on their fluidic handling and how this is relevant for POC flow cytometers.

Conventional flow cytometers use sophisticated flow cells which allow for hydrodynamic flow focusing, in which large amounts of sheath flow are used to confine the analyte particles into a narrow stream. Flow focusing guarantees a constant speed for all particles and prevents sticking to the flow cell wall. A major goal driving the development of microfluidic flow cytometers is to reduce overall size of the instrument and the required amount of analyte and sheath fluid. Microfabrication techniques can be used to realize low-cost and miniaturized flow chips; however, in order to enable analyte focusing with little to no sheath fluid, new concepts have to be implemented. Due to the parabolic flow profile in microfluidic channels, it is essential to either confine the particle path in order to ensure uniform particle velocity or to implement detection schemes which can handle the large velocity distribution. Many concepts have been suggested to achieve analyte focusing to align the particles in microfluidic channel. They include inclusion of mechanical structures in flow channels [12, 13] and the use ultrasound effects [14], to confine and align the cell in microfluidic channel. A very interesting approach uses the inertia of the fluid acting on particles in shaped microchannels to enable precise cell positioning in the stream [15–19]. This technique allows for sheathless positioning and can be used to concentrate particles to a focused position. Moreover, it has been suggested that this approach also evenly spaces cells and particles along the direction of flow and potentially minimizes coincident detection [20]. Unfortunately, the inertial focusing depends on particle properties (e.g., size, shape), dimensions of the channel, and process parameters (e.g., particle speed). For instance, smaller particles need a longer distance to reach their stable positions in microfluidic channels. Consequently, this concept does not represent a general solution for sheathless analyte focusing in microfluidic-based flow cytometer. Even though designed and optimized for a specific application, this approach might be a very good solution. Using curved microfluidic channels combined with two-dimensional sheath flow provides a solution for particle focusing which is less elegant but also less critical [21]. The inclusion of chevron-shaped mechanical structures at the channel wall is another very interesting concept for hydrodynamic focusing in microfluidic channels [22]. The chevrons cause the sheath fluid provided from one or two sides to embed the analyte stream from all sides.
Microflow cytometers are being developed and tested in many labs for various applications, including blood analysis [23], CD4 counting [13], multiplexed bead assays [24], and ocean water monitoring [25, 26]. A group at NRL has demonstrated simultaneous diagnosis of 12 different infectious diseases in serum [24] using bead-based assays and is currently integrating on-chip sample processing for POC use as a 30-min test. This group is also testing their microflow cytometer for real-time measurements of marine phytoplankton populations based on size and intrinsic fluorescence from chlorophyll and other light-harvesting pigments [25, 26]. This group claims that unlike most other microflow cytometers, their system can analyze cells ranging in size from 1 to over 100 μm. Researchers at Los Alamos National Laboratory have tested and evaluated very inexpensive lasers [27] and data systems [28]. Their goal is to reduce the cost of conventional flow cytometry and make them more widely available for point-of-care applications. A group at Purdue University is developing a LED-based microfluidic cytometer with integrated sample preparation. They have demonstrated the feasibility of using magnetic nanoparticles for on-chip sorting of human white blood cells from red blood cells and subsequent analysis of white blood cell subsets using antibody-labeled quantum dots [23]. This work as well as many other activities on this field is driven by the vision to develop an integrated handheld, portable, battery-powered unit for rapid blood analysis from a single drop of whole blood that is completely processed on-chip.

To date no flow cytometer meets all technical requirements for POC detection; in particular, the cost target remains extremely challenging. However, there are many interesting concepts under development. Especially the ones based on microfluidic flow cells look very promising. Based on the recent progress in this field, one can be cautiously optimistic that low-cost microcytometers for POC diagnostics will be available soon.

### 3.2 On-the-Flow Analyte Characterization Based on Spatial Modulation Technique

#### 3.2.1 Spatially Modulated Fluorescence Emission: The Enabling Technique

PARC has demonstrated a new optical detection technique that delivers high signal-to-noise discrimination without precision optics to enable an optofluidic detector that can combine high performance, robustness, compactness, low cost, and ease of use. Detection sensitivity and analyte throughput can meet or even exceed the specifications of currently available (commercial) high-performance flow cytometers with the addition of point-of-need compatibility.

The enabling technique is termed “spatially modulated emission” and generates a time-dependent signal as a continuously fluorescing (bio-)particle traverses a
predefined pattern for optical transmission. Correlating the detected signal with the known pattern achieves high discrimination of the particle signal from background noise. In conventional flow cytometry, the size of the excitation area is restricted to approximately the size of the particle. With the spatial modulation technique, a large excitation area (ca. 0.1 × 1 mm) is used to increase the total flux of fluorescence light that originates from a particle. Despite the large excitation area, the mask pattern enables high spatial resolution which permits independent detection and characterization of near-coincident particles, with a separation (in the flow direction) that can approach the dimension of individual particles. In addition, the concept is intrinsically tolerant to background fluorescence originating from constituents in solution, the materials of the fluidic structures, or contaminants on surfaces.

To apply the spatially modulated fluorescence emission technique to particles moving through a fluid channel, a spatially patterned mask modulates the intensity of the fluorescent light incident on the photo detector over a large excitation area as illustrated in Fig. 3.1. The time dependence of the signal is defined by the spatial structure of the stripes of the mask and the speed of the particle. The recorded signal is analyzed by correlation techniques, and the intensity and time when the particle traverses the detection zone are accurately calculated. With state-of-the-art real-time correlation techniques, characterization for particle speeds up to a few meters per second is possible. The correlation analysis not only allows for very sensitive and reliable particle detection, but it also reveals the speed of each particle (see Fig. 3.1) which ultimately enables simple fluidic handling and true volumetric determination of the analyte. The basic concept, its lab implementation, and first proof-of-concept demonstration are described in [29]. The correlation analysis is described below in Sect. 3.2.2, and applications of the spatial modulation technique for detection bioparticles are presented in Sects. 3.2.3, 3.3, and 3.4.

### 3.2.2 Data Evaluation and Correlation Analysis

The data analysis relies on the concept of a matched filter. A matched filter correlates a known model signal, or template, with the stream of sensor data in order to detect the presence of the model signal within the stream. The mathematical operation computes the dot product of the template, $T = (T_1, T_2, \ldots, T_n)$, with the time series, $t_1, t_2, \ldots$, at each possible position $j$, that is, we compute $D_j = \sum_i T_i \cdot t_{j+i}$, for each choice of $j$, and then declare a detection whenever $D_j$ exceeds some threshold chosen to balance the two types of errors, false positives and false negatives. The template may be a theoretical pattern or, as shown in Fig. 3.2, derived from previously detected signals. As a speedup, we can compute $D_j$ for many choices of $j$ at once using the fast Fourier transform (FFT) to compute a convolution, rather than separately computing $D_j$ for each choice of $j$. The matched filter technique, which dates back to World War II applications in radar and sonar, is a mathematically optimal detector for well-separated signals with additive white Gaussian noise. The technique can detect signals at remarkably low signal-to-noise
Fig. 3.1 Schematics illustrating the concept of spatially modulated emission. (Left) A patterned mask modulates the fluorescent light emitted from a particle flowing through a μ-fluidic channel directed to a large-area detector. Correlation analysis compares the measured signal and expected particle signatures. (Right) Data evaluation results in an intensity histogram and a speed profile of the detected particles that allows extraction of particle count and analyte volume, respectively.

Fig. 3.2 Time-resolved signals with spatial modulation: (a) high SNR particles serve as templates for the expected mask pattern. An improved model signal can be generated by aligning and averaging multiple observed signals. (b) Overlapping signals from calibration beads with similar SNR. (c) Overlapping signals from calibration beads with different SNR. The second particle is approximately a factor of four weaker in intensity and 30% faster than the first one.

ratios (SNR), often picking out a signal that cannot be detected by eye from a plot of the time series. Complicating the situation, however, are several factors: nonwhite noise, varying speeds of particles, overlapping signals, multiply tagged particles, and multiplexed masks.

3.2.2.1 Nonwhite Noise

If the noise in the sensor and electronics really were additive white Gaussian noise, spatial modulation would offer no sensitivity advantage, because a mask with no pattern on it would give a one-pulse signal just as detectable as a periodic signal from a striped mask. White noise, however, is a mathematical abstraction, and the
Fig. 3.3 Analysis of spatial modulation signal: (a) the time series above shows 5 ms of fluorescence intensity data. A particle is passing by the mask and its fluorescence signal is detected. While the absolute fluorescence intensity is not higher than the background, a clear pattern can be discerned. Therefore a typical threshold algorithm would fail to detect this particle. (b) Smoothened FFT of Figure 3.3a: The periodic signal from a particle passing a striped mask appears between 17 kHz and 40 kHz, depending upon the speed of the particle. Thus even a signal with low SNR still results in a strong peak in the frequency domain. Additionally an envelope peak for the detection duration of the particle occurs at a lower frequency. In Figure 3.3a, a particle is detected for approximately 1 ms, resulting in an envelope peak at 1 kHz in the FFT spectrum.

noise from the sensor and electronics (Fig. 3.3) actually has power level decreasing quite markedly with frequency. In this case, spatial modulation offers a great advantage: by varying the width of the stripes in the mask, we can place the signal in a higher frequency region and thereby improve the effective SNR by 10× or more. The advantage is greatest for periodic signals of known frequency and long duration.

3.2.2.2 Varying Speeds of Particles

Our simple, low-cost, compact flow cytometer design moves complexity from hardware to software by allowing the speed of particles to vary within the channel and compensating for this variability computationally. The software computes the
correlation of the time series with many different templates, each a resampled version of a master template with a different sample rate. The software declares a detection if any dot product exceeds a settable threshold. There is one subtlety: faster particles give smaller dot products, so for accurate measurement of fluorescence intensity, templates must be scaled to correct for this (not-quite-linear) effect. With coded masks, and hence aperiodic signals, there may also be an issue with computational speed. Simultaneous correlation with a great many templates may be slow, even with FFTs, so we generally use templates at fairly coarse steps, for example, 1 m per second, 1.02, 1.02², 1.02³, and so forth. Dependent upon the channel geometry and particle sizes, speed can vary by 3× from slowest to fastest, so that up to ~60 templates may be necessary.

With a regularly striped mask and periodic signals, there is no need to compute the correlation with many different templates. In this case, the signal is sufficiently concentrated in the frequency domain that a simple algorithm suffices (Fig. 3.2), one that declares a detection if any single power spectrum value within the expected frequency range exceeds background values by a preset threshold. We have found, however, that this simple algorithm is best used only to detect the presence and approximate speed of a particle, with accurate intensity computed using time-domain correlation as before.

3.2.2.3 Overlapping Signals

With a large-area sensor, there is a significant chance of sensing two or more particles at once. Overlapping signals (Fig. 3.2) can be hard to detect and separate, especially in the case of periodic signals and unequal intensities. The data analysis software takes a two-pass approach: it detects isolated particles first, removes them, and then detects what is left of overlapping particles. This approach appears successful for particles that overlap by less than ~80%, at which point even the human eye cannot reliably separate the signals. Except in the case of particles that physically stick to each other, we have found that overlapping signals occur at the rate one would predict by a Poisson process, that is, if there is at least one particle in the detector 20% of the time, then there are two particles about 0.2 × 0.2 = 4% of the time. Aperiodic masks are more reliable than periodic masks for the detection of overlapping particles because two particles of the same speed but opposite phase can sum to a long blur rather than a periodic signal as they pass behind a periodic mask. Not all aperiodic masks are equally good at separating overlapping particles, and mask designs that give low autocorrelation side lobes (e.g., masks based on Barker codes) outperform other masks.

3.2.2.4 Multiply Tagged Particles

As described below, we have run particles tagged with two dyes, an identifier and a reporter, through a detector equipped with a periodic mask and two read-out
channels. The identifier channel has high SNR, so we can use the identifier channel to produce the template for correlation with the reporter channel. That is, the software detects particles by looking for a strong peak in the power spectrum of the identifier channel and then correlates the detected time-domain signal from the identifier channel, normalized for total intensity, with the reporter channel to measure reporter intensity. There is no need to compute the full convolution because the two channels are exactly synchronized.

### 3.2.2.5 Multiplexed Mask

Also as described below, we have also measured two differently tagged types of particles with a single read-out channel, by using a multiplexed mask modulated at one spatial frequency for one color and at another spatial frequency for another color. The two frequencies are at a known ratio of r. For this case, the algorithm detects particles by looking for a strong peak in the power spectrum and then classifies the particle by checking the power levels of frequencies r and 1/r times the detected peak. This algorithm gave error rate about 1%, with the few errors resulting from overlapping signals.

### 3.2.3 CD4 Count in Whole Blood

The spatial modulation technique has been extensively evaluated with measurements of absolute CD4+ and percentage CD4 counts in human blood, which are required for screening, initiation of treatment, and monitoring of HIV-infected patients. And the technique has been benchmarked against a commercial instrument (BD FACSCount) with a direct one-to-one comparison of measurements on the same labeled blood samples, with excellent agreement for both absolute CD4 and CD4% as discussed in [30].

The evaluation was performed with a prototype bench-top instrument capable of measuring absolute CD4, CD8, and percentage CD4 in whole blood. A sample of tagged blood could be analyzed in less than 5 min. Sample preparation required only simple dilution, mixing, and incubation steps, with no lysing or washing step. The mixing was performed by repeatedly pipetting the sample-analyte mixture into a vial.

Measurements of CD4 were conducted on samples of whole blood. The samples were prepared with the standard BD CD4 reagent (PE-CD4, PE/CY5-CD3, and known number of fluorescent microbeads) and a recently introduced FACSCount CD4% reagent kit (BD#339010) that consists of a single tube containing a mixture of the following: three monoclonal antibodies CD4/CD14/CD15 (conjugated with PE/PE-Cy5/PE-Cy5, respectively), a nucleic acid dye, and a known number of fluorescent microbeads. The antibody to CD14 recognizes a human monocyte/macrophage antigen, whereas the antibody to CD15 recognizes
a human myelomonocytic antigen that is present on the majority of granulocytes.
No lysing of the red blood cells or washing steps to separate the tagged blood
cells from unbound dye were used. We have tested a variety of blood samples with
different dilutions (1:10 to 1:1 blood to buffer ratio), incubation times (10–40 min),
and temperatures (RT, 37°C).

A histogram of detected CD4 cells in whole blood (no lyse, no wash, dye: PE) is
shown in Fig. 3.4a as a function of fluorescent intensity. Sample preparation was as
follows: 25 μl whole blood, 2 μl CD4-PE, and 123 μl PBS, dilution of 1:5. This
illustrates a CD4 count in whole blood (no lyse, no wash, dye: PE). The plots
exhibit two peaks that are attributed to CD4 lymphocytes (right peak) and CD4
monocytes (left peak). This histogram is representative for many measurements on
this donor blood (i.e., for repeated measurements on the same sample and samples
with modified sample preparation). The average absolute CD4 count was ~1,800
CD4 cells per μl blood with a variation of about ±6%. The recorded CD4 count
is at the upper end but within the expected range for human blood. The relative
count rate of lymphocytes and monocytes and, more importantly, the peak distance
(intensity ratio) are in good agreement with data reported in the literature [31].

Measurements of %CD4 were conducted on samples of whole blood that
were stained with BD CD4% reagent (BD#339010). For this measurement, two
fluorescence signals were simultaneously recorded from the same detection area.
They were recorded from opposite sides of the fluidic chip. A volume of 30 μl
of analyte containing 3 μl of whole blood (specified protocol for the CD4%
reagent) was analyzed within 5 min. The sample was prepared by following the
recommended BD sample preparation protocol to obtain samples with a dilution of
1:10.
A representative density plot is shown in Fig. 3.4b with patterns for the major constituents of white blood cells (WBC). For comparison and benchmarking, measurements were performed on the same samples with a FACSCount (BD Biosciences). The patterns from the two instruments were in good agreement as discussed in [30].

Measurements of absolute CD4 were also successfully performed on samples with much higher blood concentrations of up to 1:2. The ability to analyze a large amount of whole blood per unit time enables either reduced analysis time or improved data statistics.

### 3.2.4 Handheld Prototype

The first generation of a compact, single-parameter optofluidic detector based on the spatial modulation technique is shown in Fig. 3.5. Its size is $\sim 5 \times 3 \times 2$ in., and we anticipate that the dimensions will shrink further in the final, engineered layout that will include laser diodes for excitation. For detection, we use a small high numerical aperture aspheric lens (NA 0.6, 6.3-mm diameter) and a basic pin photodiode rather than a photomultiplier tube (PMT). The prototype was assembled with off-the-shelf components ($\sim$350, excluding pumps and housing), and the optical unit can be battery powered (2 x 9 V, 4 x 1.5 V AA-type are sufficient for $>10$ h continuous operation).
The prototype uses sheath flow from two sides of the 25 μm-high and 120 μm-wide fluidic channel. This channel is part of a custom fluidic chip that has the spatial shadow mask directly attached to it and that also serves as a light guide for the excitation laser. Measurements of the sensitivity and dynamic range of the prototype were conducted with 2 μm Rainbow calibration beads (Spherotech) and yielded a detection limit ~200 MEPE (molecules of equivalent phycoerythrin) \[32\], which meets the needs for a wide range of bioparticle detection applications. By replacing the pin photodiode by a pixelated avalanche photodiode, the sensitivity increases to 50 MEPE which meets or even exceeds the specifications of current commercial high-performance flow cytometers.

3.3 Pathogen Detection in Water

3.3.1 Background and Currently Used Techniques

Water-quality monitoring is an essential priority for global health. The United States Environmental Protection Agency (EPA) and Centers for Disease Control and Prevention (CDC) estimate that there are 4–12 million cases of acute gastrointestinal illness annually attributable to public drinking water systems in the USA \[33\]. With microorganisms a primary cause for the occurrence of infectious diseases, the concentrations of harmful microbes should be routinely monitored to maintain microbiological quality control of drinking water.

Because of the difficulty and cost of directly measuring all microbial pathogens in water samples, organisms like coliform bacteria, *Giardia*, and *Cryptosporidium* that indicate the presence of sewage and fecal contamination have been targeted for measurement \[34\]. There is a strong need for an inexpensive, rugged, and fast detection instrument to monitor both beaches and drinking water at the point of need.

Bacterial quantification is currently performed by labs primarily using plate culture assay techniques that have supported microbiology for more than 100 years. The gold standard to determine bacterial coliform count in water starts with the membrane filter technique, then incubation growth in a plate culture followed by counting of the colony-forming units. Unfortunately, culture assay techniques for quantification are costly, labor intensive, and time-consuming to conduct with measurement times greater than 24 h due to incubation needs. Culture-independent techniques have used fluorescent microscopes, but the method is labor intensive and differences in the numbers of bacteria observed can arise due to staining technique, physicochemical characteristics of the samples, and investigator bias \[35\]. A method using TaqMan PCR has been developed to quantify indicator bacteria rapidly \[34\], but the instrument is bulky and expensive so the sample is still sent to the lab for analysis with the samples usually kept on ice during the time before analysis.
Flow cytometry as an effective and well-established method for counting cells on a large scale is also used to detect microorganisms in water [36]. Flow cytometers allow sensitive and reliable quantification of individual cells; however, as noted earlier, this technique requires expensive equipment and is skill- and labor intensive. Microfluidic devices have the potential to increase ease of use by integrating sample pretreatment and separation strategies. Recently, flow cytometer research with microfluidic devices has shown detection and quantification of bacteria [37,38], by using fluorescently labeled anti-\textit{E. coli} antibodies to selectively detect \textit{E. coli}. Flow cytometry has also been used to detect quantitatively \textit{Giardia} cysts and \textit{Cryptosporidium} oocysts [39–41]. These studies evaluated the staining efficiencies for commercial antibodies and suggest that flow cytometry is a precise method for the detection of \textit{Giardia} and \textit{Cryptosporidium} in water. Antibodies for immunofluorescence staining are available for most of the targeted waterborne pathogens.

The monitoring of drinking water is currently mostly performed by filtering and culturing techniques and represents a substantial part of workload of microbiology laboratories [42]. Waterborne bacterial pathogens and indicators are often physiologically altered/stressed and sometimes cannot be cultured efficiently with standard techniques [43]. This can lead to a considerable underestimation of the concentration of these bacteria in water and therefore of their risks to human health. In contrast, flow cytometry can detect bacteria in all stages. With appropriate staining (e.g., propidium iodide (PI)), flow cytometers can be used to distinguish between viable and nonviable cells [44].

### 3.3.2 Pathogen Detection in Water with Spatially Modulated Emission

Our prototype instrument can also be used to reliably identify and count specifically tagged pathogens (e.g., \textit{E. coli}, \textit{Giardia}, and \textit{Cryptosporidium}) in water. For example, Fig. 3.6a shows the intensity histogram and the speed profile for \textit{Giardia lamblia} stained with an anti-\textit{Giardia} monoclonal antibody conjugated with Cy3. Incubation studies were performed to determine the required staining time and amount of reagent. As shown in Fig. 3.6b for \textit{G. lamblia}, a reagent-to-analyte volumetric ratio of 1:100 and an incubation time of less than 2 min are sufficient for reliable detection. Note that even for this data point the pathogen signal (~ 10^5 MEPE) is more than two orders of magnitude separated from the noise. A one-step staining method consisted of addition of the fluorescently tagged antibody to the \textit{G. lamblia} sample, mixing, and incubation steps. No additional steps, such as washing to remove unbound antibody, were necessary prior to measurement. The results from the incubation study indicate that this particular application is compatible with rapid on-chip sample preparation.
3.4 Multiplexed Flow Assay

For advanced diagnostics, multiple biomarkers within a complex biological sample need to be simultaneously detected and quantified (proteins, cells, DNA fragments, (bio)molecules, etc.). For quantification of cellular markers, for example, CD4 cells, flow cytometry is an obvious choice. The size of cells is well compatible with the flow cytometer technology. Thousands of staining assays provide the sensitivity and specificity for intra- and cell surface markers that can be detected, enumerated, and quantified in flow cytometers. The detection of smaller individual bioparticles, for example, viruses, is less common [45] because fewer specific stains are available, fluorescent brightness and scatter signals are generally lower, and the small size of objects leads to inaccurate detection when multiple particles are present in the detection area. Especially the latter failure mechanism prevents specific direct quantification of biomolecules (e.g., proteins, DNA fragments) in solution.

In order to detect and quantify incorporated dyes of (bio-)molecules simultaneously, various detection schemes have been developed – enzyme-linked immunosorbent assay (ELISA) [46,47], DNA microchip [48,49], multiplexed SPR [50,51], etc.

The common characteristic of these techniques is the fact that different detection reagents are spaced in close proximity and the detection scheme takes position-resolved measurements. Different positions therefore identify different analytes. For any of these techniques, the detection positions are located on a surface, making the lateral diffusion lengths and times of analytes the relevant ones for the measurement.

For flow cytometer applications, multiplexed (fluorescent) particle-based assay have been developed and commercialized. These assays consist of different types of beads which can be distinguished by size (e.g., Assay Designs), emission
intensity of incorporated dyes (e.g., BD™ Cytometric Bead Array), or color of incorporated dyes (e.g., Luminex). In the assays that use particles filled with fluorescent dyes, different dye concentrations are used to label different classes of beads, while size-encoded classes of beads are measured with scatter signals. The dye intensity/color and thereby the class of particle are determined in one or more fluorescence excitation/emission channels, while the quantification of the target molecules is performed in an additional fluorescence channel. Beads of one class are functionalized with a specific primary antibody. Such a bead “collects” antigens (e.g., proteins) from the solution and binds them. After incubation and washing, a fluorescently labeled secondary antibody is added. This antibody binds specifically to all target proteins. The amount of secondary antibody bound to the bead is a measure for the concentration of the target molecule in the sample; it is detected by the fluorescence intensity of the secondary antibody’s label. This means that the identifying fluorescence of the particle (class) provides a trigger and the specificity of the system, while the fluorescence of the secondary antibody quantifies the target analyte concentration – it provides the sensitivity. The concept of using two antibodies is often referred to as “sandwich assay,” and it is also commonly used in ELISA. Therefore, bead-based analyte identification and quantification is frequently called “ELISA on the flow.”

Such multiplexed assays have been developed and commercialized, for example, for human cytokine profiling [52, 53]. In order to provide additional flexibility in microfluidic sample preparation, magnetic beads that carry fluorescent identifiers have also been developed and are commercially available [54].

A prominent example for a color-coded multiplexed bead assay is the Luminex xMap technology [55] which provides up to 100 classes by using concentration combinations of two dyes. These dyes can be excited at 635 nm, and their emission maxima are around 660 and 710 nm. PE-fluorescence – excitable at 532 nm with an emission maximum at 575 nm – is used for the quantification of the secondary antibody. Therefore, the three fluorescent labels sufficiently differ in their excitation/emission spectra to ensure a correct identification of the beads without cross talk from the identifying dyes into the quantifying channel.

We have demonstrated that our microfluidic detection platform is capable of analyzing multiplexed flow assays. We have set up an instrument with two lasers (532 and 635 nm) and three fluorescence channels. Fluorescence light was collected and filtered with a 535-nm-long pass filter. Subsequently, the detected light was sent through a 648-nm dichroic mirror. The reflected portion of the light was sent through a 585/40-nm band-pass filter onto the quantifier channel detector. Light that was transmitted through the first dichroic mirror was directed through a 635-nm-long pass filter to a second dichroic mirror (685 nm). This mirror splits the classifying dye emission to two additional detectors, one filtered with a 660/32 nm and the other filtered with a 716/40-nm band-pass filter.

Luminex xMap beads were identified, and thyroid-stimulating hormone concentrations were detected down to 0.1 µIU/ml in this setup. Figure 3.7 shows results for Luminex calibration beads for testing the identifier (CON1) and reporter (CON2)
3 Flow Cytometry on a Chip

Fig. 3.7 Results for Luminex xMap bead. The results for the CON1 calibration beads (identifier channel) illustrate the ability to distinguish between different bead types, and the results for CON2 show the ability to detect low analyte concentrations (reporter channel).

channel. This measurement illustrates that we can clearly distinguish the different bead types and that we have a sufficient sensitivity in the reporter channel to measure low analyte concentrations.

3.5 Multicolor Detection with Single Large-Area Detector

The particular strength of our technique is to enable sensitive, rugged, and low-cost instruments for PoC applications. The critical alignment within the system is substituted by the alignment of the shadow mask relative to the channel thereby providing a rugged system. To prove multicolor detection can be achieved without costly dielectric filter sets and multiple detectors, we demonstrated an elegant and effective solution for two-color detection using a single larger-area detector. Replacing the patterned shadow mask with a patterned color mask (see Fig. 3.8a) allows recording multiple fluorescence channels with a single large-area detector. Different fluorescence emission characteristics of different fluorophores are transmitted at different positions of the mask. Therefore, different fluorophores create a different fluorescence intensity-time profile when they traverse the detection area. Information of the particle’s emission spectrum is encoded in the time-dependent signal.

In our measurements, we used a 532-nm excitation laser, a pin photodiode, and a red/green-patterned color mask to distinguish between PE and PECy5 Spherotech beads. PE-fluorescence emission is transmitted by the green parts of the transmission mask shown in Fig. 3.8a, while PECy5-fluorescence emission is transmitted by the complementary red parts of the mask. As shown in Fig. 3.8b, the fluorescence signal from PE and PECy5 Spherotech beads which have the same excitation spectra but different emission spectra shows a (almost) complementary
Fig. 3.8 Multicolor detection: (a) schematics of the transmission characteristics of a patterned color mask. Light from “green” particles are only visible through the green area of the mask. (b) PE and PECy5 Spherotech Beads excited at 532 nm and detected with a red/green patterned color mask. The transmission maxima of the PE bead coincide with transmission minima of the PECy5 Bead. With the color mask spectral information are converted into time variation of the detected intensity as indicated in Fig 3.8b by the overlaid mask pattern.

fluorescence intensity-time pattern. Therefore, they can be identified based on their emission characteristics in a new way, that is, one that does not need multiple detectors, multiple filters, or dichroic mirrors.

We have also tested this characterization method by using the superposition of two periodic red and green masks with corresponding 33 and 23 transmission periods. PE and PECy5 beads showed clearly distinguishable time patterns in this setup. In particular, the fast Fourier transformation of the signal showed different ratios for the contribution of the fast emission frequency (resulting from the red parts of the mask) and the slow emission frequency (resulting from the green parts of the mask) for different beads. Based on this information, we characterized beads automatically while simultaneously determining their fluorescence intensity and their speed.

3.6 Summary and Outlook

The majority of biological and biomedical tests are performed at major, centralized clinical laboratories because the availability of compact, robust, and inexpensive instruments for POC testing is very limited. Yet there are compelling factors driving
the development of POC testing that include reduced costs, timely test results, lower mortality rates, and reduced morbidity.

Today’s flow cytometers are sophisticated analytical instruments that are extensively used in research and clinical laboratories. Their complex measurement principles make it challenging to package such a system into a mechanically rugged, compact, low-power, and inexpensive instrument. However, for the clearly defined use of CD4 and CD4% testing, such an instrument is now commercially available (Partec, CyFlow miniPOC), and we are certain that similar devices for other applications will follow. The applicability of a flow cytometer in the field will be evaluated, and this device and others will be continuously improved based on field experiences.

In particular, the cost of devices and the cost per test will be a driver for further development for any POC testing device. Smaller sizes, more functionality, better accuracy, etc., will be naturally included in the next generation of these devices. Blood glucose meters for diabetics are probably the best example for this development outline. Today, however, no POC flow cytometer has a maturity that is comparable to any given blood glucose meter.

The technical challenges for a POC flow cytometer can be grouped into two sets of problems – fluidic handling and optical detection.

Most advanced POC devices will require sample preparation steps. Sample preparation represents a major source of errors, in particular if performed by minimally trained operators. Skilled operators on the other hand are a significant cost factor in testing, but they still do not guarantee error-free operation. Therefore, automatic sample acquisition and preparation of minimal amounts is a benefit for any biomedical POC (or conventional) testing device. Numerous microfluidic and fluidic chip-based sample handling approaches have been presented and implemented: on-chip cell lysing, microfluidic separation of white blood cells, and automated specific staining of cells only to name a few.

The basis for on-chip sample preparation is in most cases reliable fluidic handling. This includes actuating, valving, mixing, and metering nano- to microliters of various liquids. To avoid large external pumps, the need for pressure lines, and complex pressure interfacing with a fluidic chip, which would contradict the idea of a portable POC devices, on-chip fluid handling has been developed. An impressive toolbox of fluid handling techniques has been developed, many different complete fluidic solutions have been demonstrated, and the field is still rapidly progressing.

At PARC, we have mainly focused on demonstrating and benchmarking a new optical detection technique that meets the requirements of POC devices. “Spatially modulated fluorescence emission” delivers high signal-to-noise discrimination without precision optics to enable robustness, compactness, and low cost in POC flow cytometers. The design is based on a large-area optical encoding-decoding mechanism rather than a redesign of a classical flow cytometer.

Our detection technique generates a time-dependent signal as a continuously fluorescing bioparticle traverses a predefined pattern for optical transmission. Correlating the detected signal with the known pattern achieves high discrimination of the particle signal from background noise.
Our method is intended to overfill a large excitation area to increase the total flux of fluorescence light that originates from a particle and to be relatively alignment insensitive. The large excitation area could be illuminated by an LED rather than a laser, a feature that cannot be achieved in classical flow cytometer setups. The mask pattern enables a high spatial resolution – comparable to classical flow cytometers – and it is aligned relative to the flow channel during the production of the fluidic chip.

The detection technique has been extensively evaluated with measurements of absolute CD4+ and percentage CD4 counts in human blood. More recent experiments demonstrate that the platform can address a large variety of diagnostic needs including multiplexed bead-based assays and identification and enumeration of pathogens (e.g., *Giardia*, *Cryptosporidium*, and *E. Coli*) in fluids.

We foresee that the future of most POC testing devices lies in the integration of sample collection and preparation into a disposable cartridge. We have described an optical detection technique that is compatible to this approach because it provides the necessary alignment tolerance, sensitivity, and price reduction.

Acknowledgment The work on the spatial modulation technique and its application in flow cytometry was partially supported by grants from the National Institute of Health (5R21EB011662-02) and the US Army Research Office (W911NF-10-1-0479).

References


Chapter 4

Lensfree Computational Microscopy Tools for On-Chip Imaging of Biochips

Serhan O. Isikman, Waheb Bishara, Onur Mudanyali, Ting-Wei Su, Derek Tseng, and Aydogan Ozcan

Abstract The use of optical imaging for medical diagnostics at the point of care (POC) has great potential, but is limited by cost and the need for highly trained personnel. To this end, the cost, complexity, and size of optical microscopy devices can be reduced through the use of computation. These techniques can perform particularly well at specific tasks such as cytometry, water quality management, and disease diagnostics. This chapter focuses on lensfree on-chip imaging techniques that are based on partially coherent digital in-line holography and are especially promising for imaging of biochips toward field-use and telemedicine applications. This emerging imaging platform discards most optical components that are found in traditional microscopes such as lenses and compensates for the lack of physical components in the digital domain. Widely available image sensors and abundant computational power are used to digitally process the acquired raw data to recover traditional microscope-like images with submicron resolution over large sample volumes within biochips.

4.1 Introduction

Optical imaging has been a cornerstone of science and medicine for centuries. Despite the recent developments that optical imaging has witnessed in terms of spatial resolution, speed, sensitivity, throughput, and other performance metrics...
[1–8], the cost and complexity of microscopic imaging tools have relatively increased, partially limiting their use to well-equipped laboratories.

A different imaging trend has emerged in recent years, where the focus is on reducing the cost and complexity of optical microscopy devices through the use of computation [9–25]. These techniques can perform particularly well at specific tasks such as cytometry, water quality management, and disease diagnostics. If these devices can be made sufficiently simple, cost-effective, and robust, that could allow lowering the cost of, for example, medical tests and point-of-care diagnostics while increasing their availability even to resource-limited settings.

This chapter focuses on this timely opportunity by discussing in greater detail a family of lensfree on-chip imaging techniques that are based on partially coherent digital in-line holography and are especially promising for imaging of biochips toward field use and telemedicine applications [9–22]. This emerging imaging platform discards most optical components that are found in traditional microscopes such as lenses and compensates for the lack of physical components in the digital domain. Widely available image sensors and abundant computational power are used to digitally process the acquired raw data to recover traditional microscope-like images with submicron resolution over large sample volumes within biochips. Though will not be covered in this chapter, other than lensfree on-chip holography, other approaches to reducing the size and cost of microscopic imaging devices are also actively pursued. Among those, one can cite optofluidic microscopy (OFM) [23,24], which is based on microfluidic channels coupled with submicron fabricated apertures designed to enhance contact imaging by increasing the spatial sampling frequency, and the CellScope, which miniaturizes the traditional design of an optical microscope to fit directly onto a cell phone [25].

4.2 Principles of Partially Coherent Lensfree On-Chip Holographic Microscopy

The operating principles of all the lens-free holographic imaging devices discussed in this chapter are based on partially coherent digital in-line holography that is operated under unit magnification [26, 27].

In an in-line holography setup, a single beam or wavefront illuminates the object. If the wavefront is spatially coherent at the sensor plane, then the light that is scattered and transmitted through the object can interfere with the portion of the light that is not scattered, forming an interference pattern (i.e., a hologram [28]) at the digital sensor plane that exhibits interference minima and maxima as shown in Fig. 4.1b:

\[
I(x, y) = |R(x, y, z_0) + s(x, y, z_0)|^2 = |R(x, y, z_0)|^2 + |s(x, y, z_0)|^2 + R^*(x, y, z_0)s(x, y, z_0) + R(x, y, z_0)s^*(x, y, z_0)
\]
Fig. 4.1 Schematic illustration of our partially coherent lensfree on-chip holography platform is shown (left). The objects are placed directly on a digital sensor array with typically <5 mm distance to its active area. A partially coherent light source, such as an LED, is placed ~4–10 cm away from the sensor to illuminate the objects recording their digital in-line holograms with unit fringe magnification over a large field of view. A typical full FOV (~24 mm²) holographic image of a heterogeneous sample recorded using a 5MP sensor with 2.2 μm pixel size is shown (right). The insets show zoomed lensfree holograms for different micro-objects.

where \( R(x, y, z) \) is the reference wave, that is, the part of the wavefront that is not scattered, \( s(x, y, z) \) is the scattered field, and \( z_0 \) is the position of the digital detector array.

Typically in digital in-line holography [26,27], the object is placed much closer to the illumination source than the detector, and a laser source (after being filtered by, e.g., a submicron aperture) is used for illumination to provide sufficient coherence and power. As a result of this, the imaging field of view (FOV) is typically much smaller than the detector size. In the lensfree on-chip imaging configuration discussed in this chapter (see Fig. 4.1a), the object to be imaged is placed much closer to the detector array rather than to the illumination source. This permits using a simple light source such as a light-emitting diode (LED) while maintaining a sufficiently large spatial coherence diameter for hologram formation at the detector array. Also, since the object is placed several centimeters away from the illumination source, a large aperture of, for example, ~0.1 mm diameter can be used, allowing easier and more efficient light coupling compared to submicron apertures used in the traditional in-line holography schemes. More importantly, from the perspective of imaging applications that require high throughput, the presented configuration with unit magnification has the entire active area of the sensor as the imaging FOV, which can be up to several squared centimeters. A major trade-off that is made in return for such advantages and simplicity is that the spatial resolution in our scheme is now strongly affected by the pixel size at the detector array. This limitation, however, can be addressed by pixel super-resolution techniques which will be discussed in greater detail in Sects. 4.5 and 4.6.

Once a lensfree hologram is recorded, it can be digitally propagated to the object plane to undo the effect of diffraction that occurred between the object and sensor planes. In our geometry, due to its long distance from the sensor, the
incident beam can be approximated as a plane wave, \( R = e^{ik s \cdot r} \), where \( k \) is the wave number of illumination, \( s \) is the direction of incidence, and \( r \) is the position vector. Therefore, the back propagation of the hologram needs to be done along the direction illumination, \( s \). For this purpose, digital beam propagation is achieved by using the angular spectrum approach [29], which involves a Fourier domain implementation of convolution. That is, the Fourier transform of the field to be propagated is multiplied by the transfer function of free space propagation, and an inverse Fourier transformation yields the complex field in the plane of interest. This process, referred to as digital holographic reconstruction, results in a complex valued microscopic image of the object. However, unlike other holography schemes such as phase-shifting [30] and off-axis holography [31], digital in-line holography suffers from an artifact which compromises the quality of the microscope image, referred to as the twin image, and can be seen in the interference terms of Eq. 4.1. Since the sensor is only sensitive to the intensity of the optical field, the phase of the field is lost. As a consequence of this, in addition to the desired field \( s(x, y, z_0) \), the term \( s^*(x, y, z_0) \) is also generated. This last term, also known as twin image, translates to a defocused image of the object spatially overlapping with the microscopic image of the object. There are several techniques in the literature that allow the retrieval of the phase of the field through the knowledge/measurement of its intensity and certain physical constraints on the field [32, 33]. For this end, one commonly used phase-retrieval technique in lensfree on-chip holography is an iterative one, which propagates the optical field back and forth between the sensor and the object planes. At each iteration, the amplitude of the field that is measured at the sensor plane is enforced, but the phase is allowed to be updated. In addition to this, at the object plane, the spatial support of the object is also enforced at each iteration. Note that this support can still be determined despite the presence of the twin-image artifact. This algorithm has been successful in retrieving the lost optical phase at the detector plane, eliminating the twin image for practical considerations [9–15].

4.3 Lensfree Holographic Microscopy of Biochips Using a Single Partially Coherent Source (LED)

The initial versions of lensfree microscopy prototypes that we developed comprise a single partially coherent light source (LED) butt-coupled to a large pinhole (with \(~0.1\) mm diameter) and an optoelectronic sensor array (e.g., CMOS or CCD chip) to record digital in-line holograms of objects within biochips without using any lenses, as illustrated in Fig. 4.1. The architectural simplicity of these platforms enabled us to build lightweight (\(~46\) g), compact (\(4.2 \times 4.2 \times 5.8\) cm), mechanically robust, and cost-effective telemedicine microscopes as shown in Fig. 4.2 [10]. Using a 5-megapixel CMOS sensor with 2.2-\(\mu\)m pixel size, these field-portable microscopes can achieve a sub-pixel lateral spatial resolution of \(~1.5\) \(\mu\)m over an FOV of \(~24\) mm\(^2\), which can be particularly useful for high-throughput diagnostic imaging applications such as blood analysis and rare-cell detection in low-resource settings.
**Fig. 4.2** A schematic drawing (left) and a photograph (right) of a lensfree telemedicine microscope that employs a single LED butt-coupled to a large pinhole (diameter ~0.1 mm) and a CMOS sensor. The cylindrical structure is a simple hollow tube within which the spatially filtered LED light propagates before impinging on the samples, which are loaded onto the sensor chip using the sample tray. Light tubes of different lengths can be interchangeably utilized to adjust the degree of spatial coherence at the sensor plane.

**Fig. 4.3** Shows the measured holograms (using the lensfree microscope of Fig. 4.2) and the corresponding reconstructed lensfree images of different types of micro-objects including microbeads, RBCs, WBCs, platelets, and *G. lamblia*. For visual comparison, conventional microscope images of the same objects obtained with a 40x objective lens (NA: 0.65) are also presented.

To demonstrate the potential of this platform for global health applications, we imaged various types of micro-objects such as red blood cells (RBC), white blood cells (WBC), platelets, microparticles, and *Giardia lamblia* cysts (*G. lamblia*: a waterborne parasite), using the imaging platform shown in Fig. 4.2. As presented in Fig. 4.3, the lensfree images obtained with our telemedicine microscope compare successfully against the conventional bright-field microscope images of the same samples and provide subcellular structural details, which can be particularly useful for cytometry applications [10, 14]. It should be noted that these images are cropped.
Despite their simplicity, lensfree telemedicine microscopes can perform advanced tasks for point-of-care diagnostics such as automated blood analysis. Toward this end, we recently demonstrated cytometry on a chip using this lensfree imaging modality [12]. Owing to the large FOV of our platform, large numbers of cells can be accurately counted in less than 1 s to measure the density of different cell types in whole blood. To achieve that, either the cell holograms can be enumerated by automated pattern matching using a library of cell holograms, or the reconstructed cell images can be counted. The advantage of the latter is that larger numbers of cells can be accurately counted in the image domain as cell images do not overlap at high densities as much as their diffraction holograms do (see Fig. 4.4). As a result, RBCs can be counted with <5 % error up to densities reaching 0.4 million cells/µL. In addition to counting, volume histograms of RBCs can also be calculated as validated against a commercial Coulter counter (see the inset in the plot shown in Fig. 4.4). The phase imaging capability of lensfree holographic microscope plays a key role for this task, as the reconstructed phase information is used to estimate the volume of each individual cell. The same platform can also be utilized to determine the hemoglobin concentration within whole blood by measuring the photon transmission through a cuvette filled with whole blood (see Fig. 4.5a). Additionally, WBCs can also be counted with less than 10 % error.
Fig. 4.5 (a) Whole blood samples are placed on a sensor chip with a cuvette to measure their photon transmission, which enables measurement of the hemoglobin density of the blood samples with <3 % error, as compared against a commercial hematology analyzer. (b) The plot shows that WBCs can be automatically counted with <10 % error compared to Coulter counter results. The large FOV is a key enabler to achieve this, as it allows counting statistically significant numbers of WBCs with a single holographic image despite their 1000× lower density in whole blood compared to RBCs (Fig. 4.5b). Furthermore, the subcellular spatial resolution in our lensfree images can also permit three-part differential imaging of white blood cells [12].

This simple lensfree holographic microscopy platform can further be enhanced by differential interference contrast (DIC) imaging techniques [10]. To achieve that, cost-effective thin nonlinear crystals can be placed on the sensor to shear the optical field transmitted by the sample into two orthogonal polarizations, which can interfere through the use of a plastic polarizer, allowing us to record DIC-enhanced holograms that can be processed as explained in Sect. 4.2 to reconstruct lensfree DIC images [9, 10]. Consequently, without adding significant cost or complexity to the system, our lensfree imagers can also perform DIC microscopy, which can especially be useful to image weakly scattering transparent objects with increased natural contrast.

4.4 Lensfree Holographic Microscopy on a Cell Phone

The ubiquity of cell phones throughout the world is making them increasingly attractive for use in point-of-care diagnostics. As of 2011, more than 80 % of the world population lives in regions that are covered by GSM networks, and around 90 % of the entire world population is expected to own a cell phone by 2015 [34]. This rapid growth in wireless communications has enabled cell phones to be equipped with advanced functionalities at significantly lower manufacturing costs. Being integrated with imaging, sensing, and communication interfaces that operate almost everywhere in the world, cell phones can actually transform healthcare by facilitating, for example, diagnosis through telemedicine in low-resource settings. In this respect, much research has been done to develop ubiquitous imaging and diagnosis
platforms running on cell phones [25, 35–38]. For instance, a mobile phone-mounted light microscope has been demonstrated that is capable of bright-field and fluorescence imaging to identify *Plasmodium falciparum*-infected and sickle red blood cells as well as *Mycobacterium tuberculosis*-infected sputum samples [25].

Along the same lines, we have also developed a compact and cost-effective microscope integrated on a cell phone [11] that does not utilize lenses and other bulky optical components. This telemedicine microscope, shown in Fig. 4.6, is based on the partially coherent lensfree digital in-line holography technique introduced in the previous section and inherits its advantages such as large field of view (e.g., \( \sim 24 \text{ mm}^2 \)), architectural simplicity, and mechanical robustness. This mobile platform utilizes a lightweight add-on unit that attaches to the cell phone to convert it to a telemedicine microscope. The add-on unit simply consists of a battery-operated LED (center wavelength at 587 nm) that is butt-coupled to a large pinhole (\( \sim 100 \mu \text{m} \) diameter), a hollow tube for light propagation and a sample-loading tray to mount the objects on top of the built-in digital sensor of the cell-phone camera unit (5 MP, \( \sim 24 \text{ mm}^2 \) active area), whose lens is physically removed. The objects placed on the sensor with \(< 2 \text{ mm} \) distance to its active area are then illuminated by the installed LED to record digital in-line holograms of the objects using the color (i.e., RGB) sensor chip of the cell phone. Nevertheless, the sensors that are employed in cell-phone cameras comprise color filters tiled in a Bayer pattern, which are optimized for color photography. This renders the cell-phone sensors nonideal for holographic microscopy, where a quasi-monochromatic light source (e.g., an LED) is employed for illumination, as these color filters lead to nonuniform pixel response partially distorting the holographic information. To minimize this distortion due to color filters, we utilized an additional digital correction step in our holographic reconstruction algorithm, summarized in Fig. 4.7, whose aim is to
calculate a grayscale de-Bayered holographic image with high signal-to-noise ratio (SNR) using the uncompressed raw Bayer pattern image captured with the cell-phone camera [11]. We specifically capture raw images as opposed to compressed color images as the color images are demosaiced, which may result in loss of high-frequency information in the holograms. In order to utilize most of the pixels of the color sensor with high SNR, we employed an LED with a center wavelength of $\lambda = 587\,\text{nm}$, where both the green and the red pixels of the sensor are highly responsive. Since a Bayer pattern is composed of two green pixels, one blue and one red pixel, only the blue pixels constituting 25% of the total pixels are inefficiently illuminated by the LED. As a result, the information in these pixels have very low SNR, and our de-Bayering algorithm aims to recover the values of these blue pixels with high SNR based on the information in the neighboring green and red pixels, using an edge-aware interpolation algorithm [11]. In this computation step, the red and green channels of a recorded raw holographic image are initially equalized using a background image, serving as a calibration measurement that only needs to be done once, which is recorded with identical illumination conditions in the absence of objects. After channel equalization and the initial interpolation to estimate the values of the missing blue pixels, we further refine the recovered pixel values through an iterative recovery process that uses the object size as a priori information, which is indeed digitally obtained with an automatically generated object-support mask. In this iterative refinement process, holograms are propagated back and forth between the sensor and the object planes using the object-support constraint as described in Sect. 4.2, and both the amplitude and phase values of the recovered holograms are updated for the blue pixels, while only the phase is updated.
Fig. 4.8 Holograms recorded with the cell-phone microscope (Fig. 4.6), corresponding reconstructed images and conventional bright-field microscope images are provided for different microparticles (3 and 7 μm diameter), RBCs, WBCs (monocyte and granulocytes), and G. lamblia cysts.

<table>
<thead>
<tr>
<th>Measured Hologram</th>
<th>Cell phone Microscope Image</th>
<th>Brightfield Microscope (10x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7μm bead</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
<tr>
<td>3μm bead</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
<tr>
<td>RBC</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
<tr>
<td>Platelets</td>
<td>[Image]</td>
<td>[Image]</td>
</tr>
</tbody>
</table>

for green and red pixels to be faithful to the measured amplitude values in these high SNR pixels of the raw image. Once convergence is achieved, a complex optical field at the hologram plane is obtained, which can be digitally back-propagated to the object plane to reconstruct a lensfree microscopic image of the samples.

We investigated the imaging performance of our cell-phone microscope by conducting experiments with different samples including spherical microparticles, RBCs, WBCs, platelets, and waterborne parasites (G. lamblia cysts). As demonstrated in Fig. 4.8, the lensfree images obtained with our cell-phone microscope correlate well with images obtained using a conventional benchtop optical microscope (10× objective lens with a numerical aperture (NA) of 0.25). The spatial resolution of our cell-phone imager is sufficiently high to reveal subcellular details of cells, which is especially noticeable in the granulocyte images where multiple nuclei within the WBC can be discerned. The high contrast of our lensfree images (which is due to reconstruction of optical phase information of the specimen) makes our approach particularly useful to image weakly scattering phase objects as revealed by the cell-phone image of the G. lamblia cyst in Fig. 4.8.

These experimental results obtained with our holographic cell-phone microscope suggest that this platform can be utilized to perform blood analysis similar to what has been presented in Sect. 4.3. Moreover, an additional advantage of using a cell-phone-integrated telemedicine microscope is the ability to wirelessly transmit the acquired holographic data to a remote station, for example, to a computer located in a hospital or clinic, for rapid digital processing. As demonstrated in [11], the raw
lensfree holograms captured with our cell-phone microscope can be significantly compressed for faster wireless transmission, such that an image corresponding to an FOV of \( \sim 5 \text{mm}^2 \) (i.e., \( \sim 1 \text{ megapixels} \)) can be transmitted using 375 kBytes of data at \( \sim 3–4 \text{ bits/pixel} \) in a common picture format such as portable network graphics (PNG) format.

4.5 Lensfree On-Chip Super-Resolution Microscopy of Biochips

Microscopic biochip imaging devices described in the previous sections have a lateral resolution of about \( 1.5 \mu m \), which proved sufficient for various applications such as counting of red and white blood cells or waterborne parasites. Note that this resolution level is still sub-pixel since a pixel size of \( \sim 2.2 \mu m \) is used to acquire lensfree holograms in our unit magnification geometry. Nonetheless, a smaller physical pixel size (e.g., \(<2 \mu m\)) will allow better sampling of the lensfree holograms and in turn a finer spatial resolution, since a 2.2-\( \mu m \) pixel size may still lead to undersampling and aliasing of the raw holograms, limiting the resolution of our microscopy platform.

One approach to overcome this spatial sampling limitation is to cover the physical pixel with pinholes, which are much smaller than the pixel size. When these pinholes are arranged in a certain configuration and the object is controllably moved across the pinholes, finer sampling of the object can be obtained, compared to the sampling without the pinholes [23, 24]. This scanning-based technique, however, requires careful fabrication and alignment of the pinholes, and very good control over the movement of the object, which might be rather limiting for imaging of heterogeneous samples in, for example, field settings. Refer to Sect. 4.7 for further discussion on optofluidic imaging of biochips using lensless holographic approaches.

Using computational on-chip imaging, it is also possible to digitally reach a smaller effective pixel size with simple mechanical modifications to the devices described in Sects. 4.3 and 4.4. Once more, the cheap and widely available computational power allows transferring some of the burden to the digital domain. To effectively achieve a smaller pixel size in our on-chip imaging setup, the same in-line holography configuration is used but now with multiple LEDs, each butt-coupled to a multimode optical fiber, to sequentially illuminate the objects of interest (see Fig. 4.9). The free ends of these fibers are arranged along a line and act as the illumination pinholes. A simple and inexpensive microcontroller turns on the LEDs one at a time, which is equivalent to moving the illumination source, causing a physical shift of the object hologram at the sensor plane. A schematic of this configuration is shown in Fig. 4.9. Due to the ratio between the source-object \((z_1)\) and object-sensor \((z_2)\) distances, the shift of the hologram on the sensor is much smaller than the distances between the free ends of the fibers.

These spatial shifts between different holograms allow extracting high-resolution content of the objects through the use of digital pixel super-resolution (PSR)
algorithms [39, 40]. First, the shifts between the different holograms need to be calculated, which can be done without the prior knowledge of any of the physical parameters of our imaging systems, adding to the robustness of the approach. Integer pixel shifts between holograms are redundant and add no useful information regarding high-frequency content. Therefore, as a first step, the integer part of the shifts between holograms is calculated through a simple correlation, and the images are approximately aligned to each other. The more useful sub-pixel shifts between the holograms can be computed, for example, using an iterative gradient method [40]. If the hologram $I_2$ is a slightly shifted version of the hologram $I_1$, that is, $I_2(x, y) = I_1(x + a, y + b)$, then $I_2$ can be approximated using $I_1$ and its derivatives:

$$
\tilde{I}_2(x, y) \approx I_1(x, y) + \left( \frac{\partial I_1}{\partial x} \right) \times a + \left( \frac{\partial I_1}{\partial y} \right) \times b
$$

(4.2)

The parameters $a$ and $b$ can be estimated by minimizing the squared distance between the measured $I_2$ and its approximation $\tilde{I}_2$. If desired, $I_1$ can be shifted by the calculated parameters, and then, the same linear approximation can be repeated for a better estimate of the shifts.
A single raw lensfree hologram is compared against a super-resolved hologram, which is the result of processing multiple raw and shifted holograms [19]. The sub-pixel shifts between the different frames are shown on the right. The super-resolved hologram reveals more fringes and resolves the aliasing apparent in the raw hologram. The additional fringes in the super-resolution hologram translate to a higher spatial resolution image (after appropriate twin-image elimination processing), as illustrated in the lower reconstructed images.

After all the shifts of the lensfree raw holograms (each captured using an individual LED within the source-array) are computed, the PSR algorithm can be invoked. The goal of PSR is to find a single high-resolution hologram, which recovers all the shifted low-resolution holograms once downsampled with the appropriate shifts. A simple way to reach this desired hologram is to minimize the following cost function:

$$C(Y) = \frac{1}{2} \sum_{k=1}^{p} \sum_{i=1}^{M} (x_{k,i} - \tilde{x}_{k,i})^2 + \frac{\alpha}{2} \left( Y_{f1}^T \bullet Y_{f1} \right),$$  \hspace{1cm} (4.3)

where $x_k$ are the measured holograms, $\tilde{x}_k$ are the corresponding images which are obtained from downsampling the high-resolution image $Y$, and the index $i$ runs over all pixels of a given hologram. The last term in Eq. 4.3 penalizes very high-frequency components which could be artifacts of the optimization process, and the strength of this penalty can be adjusted using the parameter $\alpha$. The cost function of Eq. 4.3 is a quadratic function of the pixels of the high-resolution images and can therefore be straightforwardly minimized using, for example, the conjugate gradient descent method.

An experimental demonstration of the resolution enhancement due to the multiframe PSR is shown in Fig. 4.10. A single raw low-resolution hologram and its
Fig. 4.11 Images of red blood cells infected with malaria parasites in a thin blood smear, obtained using the field-portable lensfree super-resolution microscope (see Fig. 4.9). The parasites within the cells are visible in the amplitude and phase images. A bright-field microscope image is also provided for comparison, in which the parasites have a different color due to staining.

The object that was imaged is a “UCLA” pattern etched in glass using focused ion beam (FIB) milling; therefore, it is nearly a phase-only object. The width of the letters and spacing between them in this pattern are on the order of $\frac{1}{c}$, and they are visibly resolved in the super-resolution microscope image.

The PSR holographic on-chip microscope is designed to be lightweight, robust, and cost-effective, with global health-related applications in mind, such as disease diagnostics or water quality monitoring. As an initial demonstration of the capabilities of this microscope in tackling such issues, this field-portable device was used to image red blood cells infected with malaria parasites ($P. falciparum$) in a standard thin blood smear. Figure 4.11 shows the PSR microscope images of healthy...
and infected red blood cells, both in amplitude and phase and the corresponding comparison images captured using a 40× bright-field objective lens (NA = 0.625). This figure illustrates that the combination of lensfree amplitude and phase images allows distinguishing the infected cells from the healthy ones. The FOV of this super-resolution lensless microscope is approximately 24 mm² (i.e., >20 fold larger than the FOV of a typical 10× objective lens) and would contain thousands of cells. The digital nature of the microscope coupled with the statistically significant number of cells imaged in a single FOV form a very promising starting point for possible automated detection and enumeration of malaria parasites in blood smears.

4.6 Field-Portable Lensfree Tomographic Microscope

Even though holography inherently provides the ability to reconstruct an optical field at different depths, which can in principle enable three-dimensional (3D) imaging, it cannot achieve truly tomographic imaging of samples due to its generally low axial resolution compared to the lateral one [41–43]. Similarly, in lensfree on-chip holography, the axial resolution can be more than an order of magnitude worse than the lateral resolution despite the use of pixel super-resolution techniques. To illustrate this, we digitally reconstructed a lower-resolution (LR) and a pixel super-resolved (SR) hologram of a microparticle (with a diameter of 2 μm) at different depths along the optic axis (along z). Although faithful images are obtained in the x-y plane in both cases, the FWHM of line profiles in the y-z and x-z cross sections in Fig. 4.12(a2–a3, b2–b3) are measured to be ~90 and ~45 μm with a single LR hologram and a SR hologram, respectively, giving rise to drastically elongated images along the optic axis. Such a long depth of focus (DOF) does not permit obtaining high-quality 3D images of a sample with the lensfree microscopes, where only a vertical illumination hologram (either LR or SR) is used.

On the other hand, numerous successful 3D imaging modalities have been developed to obtain volumetric structural information, in the micron scale and beyond. These systems, including but not limited to optical coherence tomography, confocal microscopy, diffraction tomography, digital holography, and light-sheet microscopy, can provide high-quality 3D images of samples such as red blood cells, cancer cells, tissue, and the like [41–58]. Along these lines, lensfree optical tomography was recently introduced that can achieve a 3D spatial resolution of <1 × <1 × <3 μm (in x, y, and z, respectively) over a large imaging volume of 15 mm³ within a biochip [22]. In this technique, SR holograms of the samples are synthesized for different directions of illumination (spanning, e.g., −50° : 50° over two-orthogonal directions, employing a dual-axis illumination scheme [22]), which are then reconstructed as described in Sect. 4.2 to obtain lensfree projection images of objects from multiple viewing angles. These projection images can be digitally back-projected using well-established algorithms used in, for example, X-ray and electron tomography [59] to compute tomograms (sectional images) of the samples with significantly improved axial resolution, for example, <3 μm.
Fig. 4.12  (a1–a3) show the reconstructed images of a 2 \( \mu m \) diameter microsphere in x-y, y-z, and x-z planes, respectively, using a single LR hologram. (b1–b3) show the reconstructed images of a 2 \( \mu m \) diameter microsphere in x-y, y-z, and x-z planes, respectively, using a SR hologram. (c1–c3) show the computed tomograms in the x-y, y-z, and x-z planes for the same microparticle, obtained by using the field-portable lensless tomographic microscope shown in Fig. 4.13.

Since we correct for the diffraction between the object and the sensor (i.e., hologram plane) by digital holographic reconstruction algorithms as discussed in earlier sections, the use of a back-projection algorithm, as opposed to a diffraction tomography approach, only ignores the diffraction within the object. This approximation can be justified by the modest NA (0.3–0.4) and the relatively long DOF of our lensfree projection images. We can denote the sample’s 3D transmission function as \( s(x, y, z) / \text{DOF} \), where \( (x, y, z) \) defines a coordinate system whose z-axis is aligned with the illumination angle (\( \theta \)) at a particular projection. Ignoring multiple scattering within the sample and by assuming that it weakly scatters the incident light [60], after phase recovery (or twin-image elimination) steps, each amplitude projection image yields the 2D line integral of our 3D object function, that is, \( \int s(x, y, z) / \text{DOF} \times dz \). That is, a projection image along a given angle can be approximated to represent a rectilinear summation of the amplitudes of the transmission coefficients of the 3D object over a length scale of one depth of focus (DOF) around \( z_0 \), where \( z_0 \) is the depth for which the tomogram is to be
reconstructed. Therefore, a back-projection approach can be used to calculate the 3D object function [22, 49, 55, 59].

Based on our initial promising results with lensfree optical tomography [22], we demonstrated a cost-effective and field-portable implementation of an optical tomographic microscope that employs a single axis of illumination spanning an angular range of $\pm 50^\circ$ [20]. This compact, cost-effective, and high-throughput lensfree tomographic microscope weighs only $\sim 110$ g (see Fig. 4.13) and achieves $\sim 1\mu m$ lateral resolution and $< 7\mu m$ axial resolution over a large imaging volume of $\sim 20$ mm$^3$ [20]. To implement this portable tomographic microscope, we used 24 LEDs ($< 0.3USD$ per piece) that are individually butt-coupled to an array of multi-mode fiber-optic waveguides (with a core diameter of $\sim 0.1$ mm) tiled along an arc as illustrated in Fig. 4.13. These LEDs provide partially coherent illumination at different angles along a single axis. To automate the illumination process, an inexpensive microcontroller and a custom-built LabView interface is employed that sequentially turns on these LEDs and captures projection holograms at different angles. To slightly increase the temporal coherence of illumination (i.e., to create a sufficiently rich hologram at especially large illumination angles, where temporal coherence requirements increase proportional to the path length increase between the object and the sensor planes), we used interference-based color filters centered at $\sim 640$ nm with $\sim 10$ nm bandwidth ($< 50USD$ total cost), which are mounted on a piecewise arc to match the arc-shaped geometry of the fiber-optic array (see Fig. 4.13).

To increase the 3D spatial resolution, in addition to multiangle illumination, we also utilized pixel SR techniques for each projection image. To achieve this, we chose to devote a single optical fiber to each angle and then physically displace the fiber tips by small amounts ($< 500\mu m$) to record sub-pixel shifted holograms at each viewing direction. In this scheme, the optical fibers are connected to a
Fig. 4.14 (a1–a3) Projection images obtained by reconstructing their corresponding SR holograms are shown for three different illumination angles. These projection images are registered with respect to the bead at the center of the FOV to demonstrate how the projection images change as a function of angle for the same volume of the sample.

common moveable bridge (arc-shaped plastic piece shown in Fig. 4.13). Low-cost magnets are attached on both ends of this bridge for electromagnetic actuation with low power consumption. By applying different voltages across the coils (which are mounted facing the magnets inside the housing), the movable bridge can be translated, leading to simultaneous shifts of all the fibers along both the x and y directions. Once fibers are shifted, a new set of projection images is obtained for all the angles. The exact amount of displacement for these fiber-ends does not need to be repeatable or accurately controlled as the hologram shifts are digitally estimated with no prior knowledge required as explained in Sect. 4.5. By recording 10–15 projection holograms, one SR hologram for each illumination angle is generated (see Fig. 4.14), which are used to digitally compute tomograms [20, 22]. A set of 24 images can be acquired in ~6 s at 4 frames/s, which can be significantly sped up using a sensor with higher frame rate (e.g., >15–20 fps).

To demonstrate tomographic imaging using the device shown in see Fig. 4.13, we imaged microbeads of 5 μm diameter (refractive index ~1.68) distributed within a chamber (~50 μm thick) filled with optical gel (refractive index ~1.52). As shown by the slice images in Fig. 4.15(b1–b5), randomly distributed beads in the chamber are successfully imaged in their corresponding depths as validated by conventional microscope (40×, 0.65-NA) images. The inset in Fig. 4.15, enclosed with the dashed rectangle, further demonstrates the depth-sectioning performance, where two axially overlapping microbeads are discerned. In a separate experiment shown in Fig. 4.12c3, the FWHM along the axial line-profile for a 2 μm bead tomogram was measured as 7.8 μm [20]. These results suggest an axial resolution of <7 μm, that is, >13× improvement over what is achieved using a single LR hologram. Computing the tomograms for the region-of-interest (with ~50 μm thickness) shown in Fig. 4.15 takes <1 min using a graphical processing unit (GPU, NVidia Geforce GTX 480).

To investigate the performance of our field-portable lensfree tomographic microscope for potential biomedical imaging applications, we conducted experiments with a Hymenolepis nana (H. nana) egg, which is an infectious parasitic flatworm having an approximately spherical shape. As demonstrated in Fig. 4.16, computed
Fig. 4.15 (a1–a5) Bright-field microscope images (40×, 0.65-NA) of different depths of a chamber filled with randomly distributed micro-beads (5 μm diameter). (b1–b5) Lensfree tomograms for the corresponding layers are shown to demonstrate depth-sectioning capability. The solid arrows in each image point to the beads that are in focus at a given depth. (c) Zoomed tomographic image through the center of an arbitrary 5 μm bead. (d) The axial line profile (along z) and its spatial derivative for the same bead as in (c). The inset enclosed with the dashed rectangle (see b4–b5) shows sectioning of two axially overlapping micro-beads both by lensfree on-chip tomography and conventional microscopy (40×, 0.65-NA).

Fig. 4.16 (a1–a3) Computed tomograms for different depths of an H. nana egg are shown. (b1–b3) 40× microscope images of the same object provided for comparison purposes. Tomograms through the object provide distinct details at different layers, demonstrating successful optical depth sectioning. We also verified that the thickness estimated from the tomograms of the egg matches its actual physical thickness of ~40 μm [20].

Even though the results presented in this section are shown for small volumes of interest, the imaging performance is maintained over a large FOV of ~20 mm² and a depth of field of ~1 mm, enabling our tomographic microscope to probe a large biochip volume of ~20 mm³ with a decent 3D spatial resolution [20]. The main reason that our axial resolution is limited to ~7 μm is the fact that we
captured projection holograms along a *single* axis with a limited angular range of $\pm 50^\circ$, since the sensor offered poor response at higher illumination angles. Therefore, a significantly better axial resolution could be achieved using a dual-axis tomography scheme as already demonstrated in [22], as well as using different sensors offering better angular response (i.e., larger pixel NA), that could permit recording of holograms at larger angles, for example, $70\text{–}80^\circ$.

### 4.7 Holographic Lensfree Optofluidic Microscopy and Tomography

In the imaging devices described above that utilize multiple-shifted frames and PSR algorithms to create a higher resolution image, the spatial shifts between different frames were achieved by utilizing multiple-shifted light sources or equivalently by mechanically shifting a single source of illumination.

A different method of obtaining these spatial shifts required for PSR is to shift the object being imaged, rather than the light source. Considering the wide interest in microfluidic devices in recent years for biological and chemical analysis of biochips [61] and even optical devices [62], utilizing the flow of an object in a microfluidic channel appears to be a promising approach for generating the shifts required for PSR. This would allow incorporating optical imaging into existing microfluidic devices with our simple on-chip in-line holography scheme and extending the possible functionalities of microfluidic devices in general.

This optofluidic approach has been implemented to create a lensless holographic optofluidic microscope on a chip [21, 63]. The system configuration is similar to the on-chip in-line holography configuration described earlier in this chapter. A microfluidic channel where the objects are to flow is placed atop a CMOS sensor, with the illumination filtered through a large aperture placed a few centimeters away from the channel, as shown in Fig. 4.17. The objects flow through the microchannel, either due to a pressure gradient or electrokinetically driven, and the digital sensor continuously captures holograms, which are shifted due to the movement/flow of the objects. Since the shifts between consecutive lensfree holograms are calculated solely from the captured images, the fluidic flow does not need to be particularly uniform in speed or direction. The requirements on the object while it flows are that it is rigid and does not rotate out of plane.

These multiple-shifted holographic frames during the flow of the objects are processed in a similar fashion to the processing done in the portable super-resolution microscope discussed earlier. The shifts between different frames are estimated, and then, the PSR algorithm finds the best high-resolution hologram compatible with all the shifted images. To obtain microscopic images of the flowing objects, performing PSR for only a single viewing angle (e.g., vertical illumination) can be sufficient, while the illumination angle can also be varied to obtain multiple SR projection images if tomographic microscopy is of interest. In Fig. 4.17, the imaging result for
Fig. 4.17  A schematic of holographic opto-fluidic microscopy (HOM), which can provide super-resolution (SR) holograms of flowing objects within the micro-channel at any illumination angle (left). For this end, multiple lensless holograms are captured at a given illumination angle as the object flows through the microfluidic channel. To the right, the images of a *Caenorhabditis elegans* worm with vertical illumination are shown, where the multi-frame PSR result shows noticeably enhanced resolution when compared to the single hologram microscope image. In this case, 15 shifted holograms were used as input. Using different viewing angles within an angular range of $\pm 50^\circ$, the same platform can also serve as an opto-fluidic tomographic microscope as illustrated in Fig. 4.18.

A *C. elegans* worm is shown. If a single hologram is used to recover a microscopic image of the worm (ignoring the flow), the resolution of the reconstructed image is constrained by the physical pixel size of the CMOS sensor, as we discussed earlier. However, when the flow is utilized to capture multiple frames for PSR processing, the resulting reconstructed microscopic image is sharper and reveals more details and better contrast due to the smaller effective pixel size after PSR processing (see Fig. 4.17). In this case, 15 shifted consecutive holograms were used for PSR, which took 3 s of acquisition time at 5 frames per second, with the object flowing at a speed of $\sim 1 \mu m$ per second. A higher frame rate sensor would allow a shorter total acquisition time and faster flow of the samples within the biochip.

Owing to the almost alignment-free, robust, and simple nature of our illumination scheme, the same holographic optofluidic microscope (HOM) can be conveniently converted to an optofluidic tomography platform as shown in Fig. 4.17 [21]. By illuminating the objects from different angles as they flow through the microchannel, it is possible to record multiple projection holograms for each angle and then synthesize SR holograms of the objects, as described above for a *C. elegans* worm that is electrokinetically driven through a microchannel. Using the procedure described in Sect. 4.6, these SR holograms can then be used to compute tomograms of the flowing objects. As shown in Fig. 4.18, optofluidic tomography can achieve
Fig. 4.18 Tomographic optofluidic imaging of *C. elegans* is illustrated. Different depth slices through the worm from \( z = -6 \) to \( +6 \mu m \) reveal distinct details. A conventional microscope (10×) image is also provided for the same worm.

optical sectioning of micro-objects to reveal distinct details at different depths of the objects. To prevent distortion in the final tomograms, the objects should ideally flow rigidly through the chamber during the entire acquisition since uncontrolled movements such as out-of-plane rotations may not always be digitally corrected.

### 4.8 Discussion and Conclusions

Lensfree on-chip holography is a rapidly emerging computational microscopy modality that offers high-throughput imaging of biochips in compact, cost-effective, and field-deployable architectures, as well as offering straightforward integration with microfluidic platforms. Since objects are placed directly on the sensor and away from the illumination source, holograms of micro-objects, for example, cells, are recorded with unit magnification; as a result of which, the entire active area of the sensor serves as the imaging FOV. This FOV is typically \( 24 \, \text{mm}^2 \) and could further increase with next generation sensor chips that employ a larger active area and smaller pixels. As we discussed earlier, the immediate drawback of our unit magnification imaging scheme is that it limits the achievable spatial resolution to \( \sim 1.5 \, \mu m \) (for a typical pixel size of, e.g., \( 2.2 \, \mu m \)). To mitigate this limitation, pixel super-resolution is shown to offer a unique solution by significantly increasing the lateral resolution to \(< 1 \, \mu m \) without compromising the large FOV of our computational microscopes. As a further improvement to on-chip microscopy, lensfree optical tomography has also been developed to significantly improve axial resolution to \(< 3 \, \mu m \) and achieve high-throughput 3D imaging of micro-objects within biochips. As cost-effective LEDs are utilized without the need for sensitive alignment, achieving multiangle illumination with multiple LEDs does not add a
significant cost or complexity to our field-portable tomographic microscopes, which is an important advantage compared to laser-based holographic imaging approaches.

Another important advantage of holographic recording with unit fringe magnification is the elimination of highly coherent illumination sources (e.g., lasers), which significantly reduces various noise terms such as speckle and multiple interference noise. That is, the partially coherent illumination acquires sufficient spatial coherence as it propagates 40–100 mm after the pinhole toward the sensor chip, which permits holographic recording of the interference between the background light (reference wave) and the scattered light (object wave). Moreover, owing to the quasi-planar reference wave and the short sample-to-sensor distance, the optical path difference between the reference and object waves is typically within the temporal coherence length range of the illumination that has \( \sim 1–15 \) nm spectral bandwidth. This reduced coherence requirement at the source end permits the use of a large pinhole (e.g., 0.05–0.1 mm diameter), thereby removing the need for any coupling optics (e.g., an objective lens mounted on a mechanical stage) between the light source and the pinhole. As shown in [10], in partially coherent digital in-line holography, the demagnified image of the pinhole at the detector plane is effectively convolved with the recorded optical field. Nevertheless, this does not pose any limitation in our imaging geometry due to the large \( z_1/z_2 \) ratio, giving rise to a demagnification factor of \( \sim 100–200 \), which effectively scales the pinhole size down to \( <1 \mu m \) at the detector plane.

An important limitation of lensfree on-chip holography, which is common to all in-line transmission holography schemes, is that it requires the samples to have relatively low optical density. Since a separate reference wave is not generated in an in-line imaging geometry, dense samples lead to excessive distortion of the reference wave, and the detected intensity at the sensor chip becomes dominated by the non-holographic self-interference terms (i.e., the second term in Eq. 4.1 of Sect. 4.2). As a result, our lensfree telemedicine microscopes, which work in transmission geometry, cannot image dense samples such as histopathology slides. This task, however, can be achieved by alternative reflection-based lensfree holography approaches, which are not discussed in this chapter.

In addition to sharing the common requirements of in-line holographic imaging, especially regarding the relatively lower spatial density of the samples, lensfree tomographic microscopes additionally require the objects to satisfy the so-called projection approximation [22, 49, 59]. This approximation necessitates that the digitally reconstructed images represent a line integral of a property of interest of the object (scattering, absorption, phase, etc.). This requires that each spatially connected object should be a relatively weak scatterer such that it is not thicker than the depth of field of the reconstructed images.

In conclusion, lensfree on-chip holography is a promising computational microscopy platform offering high-throughput imaging of biochips within a compact, cost-effective, lightweight, and mechanically robust architecture. These devices can provide a new toolset for point-of-care diagnostics and telemedicine applications by facilitating important tasks such as microscopic analysis, cytometry, and detection of infectious diseases.
References


Chapter 5
Dielectric Detection Using Biochemical Assays

Yang-Kyu Choi, Chang-Hoon Kim, Jae-Hyuk Ahn, Jee-Yeon Kim, and Sungho Kim

Abstract  Point-of-care (POC) diagnostics typically make use of labeling techniques that employ fluorescent, chemiluminescent, redox, or radioactive probes. Although such methods provide high sensitivity, they are complicated because their labeling steps require a significant amount of time and labor in their execution and in the analysis of their results. Thus, the portability, which is meant to be the primary advantage of POC systems, is sacrificed. The use of electronic devices for POC systems circumvents this problem, enabling label-free detection, miniaturization, and low costs. Label-free detection is made possible by direct electrical measurement of the sample molecules, which works by monitoring changes in their intrinsic electrical properties. Miniaturization and the integration of sensors and readout circuitry have been enabled by industrialized microfabrication technology. By integrating the sensors and circuitry onto a monolithic substrate, the fabrication cost can be remarkably reduced.

5.1 Introduction

Point-of-care test (POCT) systems have shown great promise as diagnostic techniques that provide fast, convenient results at or near the site of patient care [1]. These methods, which are widely used, often involve labeling techniques that employ fluorescent, chemiluminescent, redox, or radioactive probes [2]. Although such methods provide high sensitivity, they are complicated because their labeling steps require a significant amount of time and labor in their execution and in the analysis of their results. A transducer that converts a molecular signal to an electrical signal is an indispensable part of the aforementioned approaches, but
transducers are typically not well miniaturized. Thus, the portability, which is meant to be the primary advantage of POCT systems, is sacrificed. The use of electronic devices for POCT systems circumvents this problem, enabling label-free detection, miniaturization, and low costs [3]. Label-free detection is made possible by direct electrical measurement of the sample molecules, which works by monitoring changes in their intrinsic electrical properties. Miniaturization and the integration of sensors and readout circuitry have been enabled by industrialized microfabrication technology. If the sensors and circuitry are monolithically integrated on the same substrate, then the fabrication cost can be remarkably reduced.

Label-free electrical detection is usually based on the electrical properties of biomolecules. The binding of charged molecules leads to changes in the surface potential, which can be measured by changes in conductivity or capacitance. Although this type of charge-based detection achieves high sensitivity, the sensor signal can be adversely affected by environmental conditions such as pH and ionic strength. Moreover, weakly charged or neutral biomolecules can be difficult to detect with charge-based methods. However, dielectric detection, which is a type of detection based on the dielectric properties of biomolecules, is less sensitive to environmental variations, which allows it to be used to detect weakly charged or neutral biomolecules.

The aim of this chapter is to describe recent advances in the dielectric detection of biomolecules for POCT systems. Several electrical detection techniques will be reviewed. A nanogap-embedded device that is well suited to detecting dielectric changes will be described, and experimental results obtained with this device will be discussed. The discussions will also address the structural modifications of dielectric sensors, different options for sensing metrics, and the effects of environmental conditions on this technology.

5.2 Electrical Detection Based on Dielectric and Charge Properties

5.2.1 Electrochemical Impedance Spectroscopy

Electrochemical impedance spectroscopy (EIS) is suitable for the electrical detection of biomolecular interactions on the transducer surface [4, 5]. In EIS, a voltage perturbation with a small amplitude applied to an electrochemical cell generates a current response. The current response depends on the impedance of biomolecules, which is related to the resistive and capacitive properties of the biomolecules. The impedance is defined as the ratio of the applied voltage and the current response. The impedance between the electrode and the electrolyte solution can be simply modeled using the Randles equivalent circuit, as shown in Fig. 5.1a [4], where $R_s$ denotes the resistance of the electrolyte solution. The charge can be stored in the electrical double layer at the interface, resulting in the double layer capacitance $C_{dl}$. 


Dielectric Detection Using Biochemical Assays

The charge transfer resistance $R_{ct}$ is related to the current flow caused by the redox reaction at the interface.

Using the resistance $R_s$ in series with the parallel combination of the resistance $R_{ct}$ and the capacitance $C_{dl}$, as shown in Fig. 5.1a, the total impedance can be given by

$$ Z = R_s + \left( \frac{1}{R_{ct}} + j \omega C_{dl} \right)^{-1} = R_s + \frac{R_{ct}}{1 + (\omega R_{ct} C_{dl})^2} - j \frac{\omega R_{ct} C_{dl}}{1 + (\omega R_{ct} C_{dl})^2}. \quad (5.1) $$

The real part and the imaginary part of the impedance are expressed as

$$ Z_{Re} = R_s + \frac{R_{ct}}{1 + (\omega R_{ct} C_{dl})^2}, \quad Z_{Im} = -j \frac{\omega R_{ct} C_{dl}}{1 + (\omega R_{ct} C_{dl})^2}. \quad (5.2) $$

Impedance data can be represented in Nyquist form as shown in Fig. 5.1b [4]. Each data point corresponds to a different frequency value. The impedance is limited to $R_s$ at high frequencies and $R_s + R_{ct}$ at low frequencies. As shown in Fig. 5.1b, the limited impedance at low frequency increases as $R_{ct}$ increases. The maximum of the semicircle or the maximum of $Z_{Im}$ occurs when the frequency is equal to $(R_{ct} C_{dl})^{-1}$.

As shown in Fig. 5.2a, receptors, recognition elements, are immobilized on the surface of the transducer electrode. When analytes, target biomolecules, bind to the receptors, the charge transfer between the redox mediator and the electrode is interrupted, resulting in an increased value of $R_{ct}$. Concomitantly, $C_{dl}$ decreases because the biomolecular thickness increases. By monitoring the impedance change, the binding of the analyte to the receptor at the interface of the electrode can be detected. However, in the presence of the insulator at the interface of the electrode or the absence of the redox mediator, no charge transfer occurs due to the blocking
of the redox reaction, as shown in Fig. 5.2b. In this case, the change in $C_{dl}$ is the only parameter with which to detect analyte binding.

The EIS method has been used for immunoassays [6–8] and nucleic acid detection [8–11]. It is possible to detect target biomolecules with this label-free electrical method, but one of its disadvantages is low sensitivity. Several amplification techniques had been proposed to improve detection sensitivity, including techniques that use enzymes [9, 12], liposomes [13, 14], conducting polymers [15], or nanomaterials [16, 17].

To achieve high sensitivity, a field-effect transistor (FET)-based biosensor has been suggested [18, 19]. Miniaturization and compatibility with complementary metal-oxide semiconductor (CMOS) technology are additional advantages of this type of biosensor. FET-based biosensors are reviewed in the following section.

### 5.2.2 FET-Based Biosensor

FET-based sensors are attractive because they allow changes in the solution pH or the binding of analytes on the surface to be directly monitored based on changes in the electrical properties of the target molecules.

An ion-sensitive field-effect transistor (ISFET), in which the electrical properties change according to the pH or ionic strength of the solution, was first reported in the 1970s [20]. The structure of an ISFET is shown in Fig. 5.3. An ISFET is essentially a FET in which a gate electrode is replaced by a reference electrode and electrolyte solution. It is noteworthy that the gate dielectric is exposed to the electrolyte solution. The reference electrode supplies a stable potential in the solution and produces a channel under the gate dielectric layer.

The ISFET is not a biosensor but rather a chemical sensor. The gate dielectric of the ISFET, which is a chemically sensitive material, changes the surface potential of the channel depending on the $H^+$ ion concentration; hence, the source-to-drain
current changes \[21\]. Figure 5.4 shows a typical change in current–voltage characteristics due to the pH changes. Surface hydroxyl groups on the ISFET dielectric layer react with the solution in different ways depending on pH; for a basic solution, this layer shows larger negative charge characteristics compared to an acidic solution. Therefore, the threshold voltage is larger in the case of a higher pH value.

Because the ISFET is very sensitive to any electrical interaction on the surface, it may also be used to detect DNA hybridizations or immunological reactions on the surface. With single-stranded DNAs or antibody/antigen molecules immobilized on the surface, an equivalent threshold voltage change is expected when a specific binding reaction occurs. However, the unreliable operational behaviors of ISFETs have limited their performance as sensors [22]. Unexpected responses such as drift and hysteresis have been observed during sensor operation, resulting in misleading signal changes during measurement. ISFETs are also sensitive to external light and temperature fluctuations [23] as the sensor is made of a semiconductor with a channel area that is exposed to solution. Additionally, the requirement of a reference electrode makes it difficult to build the sensor on a silicon chip using integrated circuit technology.

The limited performance of ISFETs can be overcome by improving the architecture from a planar structure to a nanoscale three-dimensional structure, that is, semiconductor nanowires. Because nanowire has a higher surface-to-volume ratio than a planar device, it displays a higher sensitivity [24]. The molecules bound on the surface affect the channel potential of both devices, but in the case of nanowires, electrical carriers in the bulk of the nanowire are affected as well as in the surface of the nanowire.

![Fig. 5.3 A schematic representation of an ISFET. The reference electrode and electrolyte solution play a role as a gate electrode in an ISFET.](image-url)
Fig. 5.4 Drain current ($I_D$) versus gate voltage ($V_G$) characteristics and their changes with pH. The aqueous solution with the higher pH shows the higher threshold voltage.

Silicon nanowire biosensors have been used to detect ions [24], small molecules [25], proteins [26], DNA [27], and viruses [28] by taking advantage of the changes in surface charge, depending on the binding of target molecules. Surface charges serve as a gate and induce a potential change in the nanowire channel, thus leading to a conductance change in the nanowire. For example, as shown in Fig. 5.5, when prostate specific antigen (PSA), a cancer marker protein, binds to a nanowire coated with the PSA antibody, the conductance is increased because PSA carries a negative charge at pH 7.4.

Fig. 5.5 Conductance versus time, recorded for a PSA antibody-modified p-type silicon nanowire for (1) 9 pg/ml PSA, (2) 0.9 pg/ml PSA, (3) 0.9 pg/ml PSA and 10 μg/ml bovine serum albumin (BSA), (4) 10 μg/ml BSA, and (5) 9 pg/ml PSA (Copyright 2005 Nature Publishing Group)
In the early days of silicon nanowire biosensor research, “bottom-up” nanowires were widely used. However, in terms of integration, alternative “top-down” fabrication methods are becoming more attractive. The use of “top-down” method has also been demonstrated in a sensor application [29]. The sensor response is shown in Fig. 5.6.

Up to this point, we have summarized electrical detection-based biosensors and dielectric-based biosensors. To utilize both types of sensor technology, a dielectric-modulated field-effect transistor (DMFET) has been suggested. The DMFET has many advantages, such as label-free detection, easy integration of readout systems, compatibility with low-cost CMOS technology, and high applicability for detecting various types of biomolecules, including those that are electrically neutral. The details of DMFET will be explained below.

### 5.3 Dielectric-Modulated Field-Effect Transistor (DMFET)

#### 5.3.1 Basic Structure and Theory

The DMFET structure can be obtained via the modification of a conventional FET (Fig. 5.7a) [30–32]. The gate is suspended above the gate oxide, and a nanogap is formed by a carving process between the gate and the gate oxide, as shown in Fig. 5.7b. Biomolecules can be introduced and bound within the nanogap using nanofluidics [33]. To create biosensors, a DMFET can be functionalized with receptors (Fig. 5.7c) that capture specific analytes (Fig. 5.7d) in the sample solution. The electrical characteristics of DMFETs are subsequently affected by the properties of the biomolecules introduced into the nanogap; in particular, the charge density and the dielectric constant of the biomolecules alter the electrical properties of the DMFET.
The nanogap is typically fabricated using advanced lithographic techniques such as electron beam lithography and dry etching [34, 35], which make the process expensive and complex. The novelty of the DMFET is its simple nanogap fabrication method, which provides more design flexibility than conventional methods. The nanoscale gap is defined not by lithography but by a thin film deposition and a wet etching process, which are conventional CMOS fabrication methods. The nanogap size corresponds to the deposited thin film, which can be controlled at atomic resolution using an atomic layer deposition (ALD) technique. Thus, the nanogap size is not constrained by a lithographic resolution limit.

When analyzing the effects of biomolecules in DMFETs, an analogy between a conventional FET and the DMFET may be noted. In both cases, the gate voltage ($V_G$) modulates the drain current, which flows horizontally from the drain (D) to the source (S), as illustrated in Fig. 5.7a. The current starts to flow when the gate voltage exceeds the threshold voltage ($V_T$). In the operation of conventional FETs, trapped charges in the gate oxide induce a $V_T$ shift. This $V_T$ shift, caused by trapped charges, can be exploited in memory applications [36]. In a similar manner, a $V_T$ shift occurs in DMFETs due to the intrinsic charges of the biomolecules bound in the nanogap. The electrical characteristics of FET are mostly governed by the gate field, which is applied across the gate dielectric. For example, the vertical field from the gate is strengthened as the dielectric constant ($k$) of the gate dielectric increases. Hence, the drain current can be further increased by the use of a “high-$k$” gate dielectric material [37]. Similarly, when biomolecules are introduced onto the nanogap of DMFETs, the dielectric constant is increased ($k > 1$) from unity. Thus, it is evident that the dielectric properties of the biomolecules affect the electrical characteristics of DMFETs, especially the threshold voltage and the corresponding drain current.
Fig. 5.8 Simplified model of the total gate capacitance ($C_g$) of a DMFET, corresponding to receptor immobilization and analyte binding. Gate oxide, airgap, analyte, and receptor can be modeled as the capacitances $C_{ox}$, $C_{air}$, $C_{anlt}$, and $C_{rcpt}$, respectively.

To determine the dielectric effect of biomolecules, we initially assume that the analyte is weakly charged or neutral. The $V_T$ value of the DMFET can then be adapted from a modification of the $V_T$ value of a conventional FET and is defined by the following equation:

$$V_T = V_{FB} \pm 2\psi B \pm \frac{Q_{DEP}}{C_g}. \tag{5.3}$$

Here, $V_{FB}$ is the flat band voltage, $2\psi B$ is the surface potential, $C_g$ is the total gate capacitance, and $Q_{depl}$ is the depletion-layer charge. Equation 5.3 is considered to have a positive sign for n-channel FETs and a negative sign for p-channel FETs. As shown in Fig. 5.8, the gate oxide, airgap, analyte, and receptor can be modeled as the capacitances $C_{ox}$, $C_{air}$, $C_{anlt}$, and $C_{rcpt}$, respectively. They are connected in series, resulting in the total gate capacitance ($C_g$) given by the following equation:

$$\frac{1}{C_g} = \frac{1}{C_{rcpt}} + \frac{1}{C_{anlt}} + \frac{1}{C_{air}} + \frac{1}{C_{ox}} = \frac{t_{rcpt}}{k_{rcpt}\varepsilon_0} + \frac{t_{anlt}}{k_{anlt}\varepsilon_0} + \frac{t_{air}}{\varepsilon_0} + \frac{t_{ox}}{k_{ox}\varepsilon_0}. \tag{5.4}$$

In this equation, $\varepsilon_0$ is the permittivity of air, $k_{rcpt}$ is the dielectric constant of the receptors, $k_{anlt}$ is the dielectric constant of the analytes, $k_{ox}$ is the dielectric constant of gate oxide, $t_{rcpt}$ is the thickness of the receptors, $t_{anlt}$ is the thickness of the analyte, $t_{air}$ is the thickness of air, and $t_{ox}$ is the thickness of the gate oxide.

Filling this nanogap ($k = 1$) with analytes ($k_{anlt} > 1$) bound to receptors increases the total gate capacitance and results in a signal change, that is, $\Delta V_T$. According to Eqs. 5.3 and 5.4, $\Delta V_T$ can be given as

$$\Delta V_T = V_{T,anlt} - V_{T,anlt} = \pm |Q_{dep}| \left( \frac{1}{C_{g,anlt}} - \frac{1}{C_{g,anlt}} \right)$$

$$= \pm |Q_{dep}| \left( \frac{t_{anlt}}{k_{anlt}\varepsilon_0} - \frac{t_{anlt}}{\varepsilon_0} \right) = K \left( \frac{1}{k_{anlt}} - 1 \right) \tag{5.5}$$

where $K = \pm |Q_{dep}|t_{anlt}/\varepsilon_0$ denotes the response coefficient. Again, the equation has a positive sign for n-channel FETs and a negative sign for p-channel FETs.
Table 5.1 The direction of $V_T$ according to the properties of biomolecules and the types of FETs

<table>
<thead>
<tr>
<th>Dielectric effect</th>
<th>n-Channel FETs</th>
<th>p-Channel FETs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\Delta V_T &lt; 0$</td>
<td>$\Delta V_T &gt; 0$</td>
</tr>
<tr>
<td>Charge effect</td>
<td>Negative</td>
<td>$\Delta V_T &gt; 0$</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>$\Delta V_T &lt; 0$</td>
</tr>
</tbody>
</table>

Thus, it is possible to detect the specific bindings of biomolecules by monitoring $\Delta V_T$. As shown by Eq. 5.5, the amount of signal change increases as the dielectric constant of the analyte ($k_{anlt}$) increases. Moreover, the signal change can be enhanced via the response coefficient, which is determined by $t_{anlt}$ and $Q_{dep}$ that, in turn, depend on the substrate doping concentration. Although we previously assumed that the analyte is weakly charged or neutral, strong charges exist on some of the most analyzed biomolecules: proteins and nucleic acids. When the analyte is negatively/positively charged, it leads to a positive/negative $V_T$ shift ($\Delta V_T > 0/\Delta V_T < 0$) in both n-channel FETs and p-channel FETs.

It is important to note that the direction of the $V_T$ shift depends on the dielectric and charge effects of the biomolecules, as well as the types of FETs, as shown in Table 5.1. Thus, the properties of the biomolecules and the device type should be considered to maximize the signal change.

It is well known that the charge effect is inversely proportional to the distance from the sensor surface (here, the channel). As the charged analytes move far from the silicon channel, the charge effect tends to be weaker, resulting in a smaller $V_T$ shift. Thus, one should consider a binding site where receptors are immobilized and subsequent analytes are bound. When the binding site is close to the silicon channel, the charge effect is the dominant factor, exceeding the dielectric effect. However, when the binding site is far from the silicon channel, the charge effect is weaker, and the dielectric effect is relatively more influential in the detection of the analytes. Additional details and experimental data are described in the following section.

5.3.2 Proof of Concept and DNA Detection with DMFET

The first result involving a DMFET was reported in 2007 [30]. In that work, researchers concentrated on the proof of concept of a DMFET with weakly charged biomolecules: specifically biotin and a streptavidin biomolecules, which are the most widely used biomolecules in verifications of the operation of a biosensor. The fabricated DMFET had a thick gate oxide (10 nm) and a gold gate with nanogaps at the edges of the gate dielectric, as depicted in Fig. 5.9. The thick gate dielectric reduced the $V_T$ shift caused by trapped charges or intrinsic charges from biomolecules; hence, only a $V_T$ shift due to a change in the dielectric constant was, in fact, observed (Fig. 5.10).
After molecule immobilization, the gate oxide layer was reduced because of the dielectric constant of the dielectric layer was increased from 1 to a higher number ($k > 1$ for biomolecules). As shown in Fig. 5.10, the $V_T$ shift after the binding of streptavidin and biotin was 0.73 V. The subsequent research on DMFETs concentrated on the charge effect related to the $V_T$ shift [38]. To maximize the $V_T$ shift caused by trapped charges or intrinsic charges, a newly designed DMFET with a very thin gate oxide (4 nm) was fabricated. Two types of devices, an n-channel DMFET and a p-channel DMFET, were fabricated at the same time to verify the charge polarity effect according to
the type of FET. The target biomolecules were also replaced with DNA, which is a negatively charged biomolecule. PNA, which has no electrical charge, was used for control experiments to verify the dielectric constant effect.

Figure 5.11 shows the transfer $I_D - V_G$ characteristics and statistical variation of $V_T$ according to the biomolecule binding steps for the n-channel and the p-channel DMFETs. In the graphs, the charge effect and the dielectric constant effect counteract each other in the n-channel DMFET, whereas they are acting in the same direction to shift $V_T$ in the p-channel DMFET. With these data, it was verified that the dielectric constant increment and negative charges in DNA were competing against each other in the n-channel DMFET, as shown in Table 5.1. The $V_T$ value of the p-channel DMFET shifted toward the positive side either after target DNA (analyte) or target PNA (analyte) hybridization, whereas the $V_T$ value of the n-channel DMFET shifted to the positive side only when the target DNA was hybridized. In addition, it shifted to the negative side when the target PNA was hybridized to probe DNA (receptor). A notable result was that the differences in the $V_T$ shift between the PNA hybridization and DNA hybridization to the probe DNA had the same value, indicating that the differences were entirely caused by the negative charges in the DNA. From this result, it was confirmed that an increment in the sensing margin is possible via the proper selection of the FET type, that is, n- or p-channel, according to the charge polarity of the analyte.
5.4 Structural Modification of a DMFET

5.4.1 Underlap FET

The DMFET has the advantages of high compatibility with the conventional CMOS process and adaptability to readout circuits for on-chip integration; however, the bio-reaction probability is lower than that of other biosensors because the sensing part in a DMFET is retracted and covered by the gate material. To overcome this shortcoming while maintaining the advantages of the DMFET, an underlap FET was proposed (Fig. 5.12) [39]. With a novel structure derived from a conventional FET, the underlap FET has all of the merits of the DMFET, as well as additional advantages originating from the underlap structure [39]. The underlap structure shows higher bio-reaction probability and structural stability compared to the carved nanogap architecture due to its opened sensing area. In addition, small changes in the current can be accurately measured because the channel potential in the underlap region is highly sensitive to external charges [39].

Figure 5.12 shows a schematic of the underlap FET in which an offset (underlap region) is introduced between the gate and the drain. The underlap region serves as the sensing area; target molecules on the underlap region will affect the channel potential of the underlap region, which results in a drain current change. For more details, see Fig. 5.13, which shows an expanded view of an underlap FET to explain its operational principle. After the immobilization of charged biomolecules on the underlap region, the number of inverted electrons in the underlap region is modulated, thus causing a conductance change to occur. As a result, the drain current at a particular gate voltage will change.

The operation of an underlap FET was demonstrated using an avian-influenza (AI) antigen/antibody. As shown in Fig. 5.14, after the binding of the AI antibody (anti-AI), the drain current was significantly decreased. The abrupt drop in drain current was attributed to the negative charges of the anti-AI: negatively charged molecules increase the channel potential, resulting in a decrease in the drain current. With its simple fabrication process, CMOS process compatibility, and enhanced sensitivity, the underlap FET is a promising candidate for use in chip-based biosensors.

5.4.2 Double-Gate FET

As mentioned previously, there have been two main approaches used in preparing nanowire biosensors: bottom-up and top-down [40]. In the bottom-up process, integration issues and incompatibilities with the conventional CMOS process cannot be avoided, even though the size of the nanowire can be reduced beyond the limit of lithographic resolution. In contrast, the top-down approach is restricted by the lithographic resolution limit, but it enables more precise control of the position
of the nanowire and enables the fabrication of perfectly ordered nanowire arrays [41]. Thus, it is timely to consider the structural and/or operational modifications of nanowire biosensors to overcome the aforementioned challenges while also utilizing the well-established CMOS technology.

A double-gate nanowire biosensor was proposed to avoid the aggressive scaling of silicon nanowires in biosensor applications [42]. Compared to conventional nanowire FET biosensors, which are operated using a single bottom gate, the remarkable difference in the double-gate FET is that independent double gates (G1 and G2) are positioned vertically beside the silicon nanowire and facing each other, as shown in Fig. 5.15.
There are two ways to drive the FET in the double-gate structure: the single-gate (SG) mode and the tied double-gate (DG) mode [43]. In the SG mode, G1 is used as a drive gate, and G2 is used as a supplementary gate to pin the channel potential at a fixed voltage. In contrast, in the DG mode, G1 and G2 are electrically connected, which implies that the same voltage is always applied to G1 and G2; that is, it has a symmetrical bias ($V_{G1} = V_{G2}$).

The data measured in the SG mode (hollow circles) in Fig. 5.16a show that the drain current by G1 can be modulated according to the bias condition of G2. The increment of the G2 voltage ($V_{G2}$) from a negative to positive value tends to lower the $V_T$ value and degrade the subthreshold slope (SS), which is defined as $d(V_{G1})/d(\log I_D)$; that is, it becomes less steep. However, the characteristics of the DG mode (filled squares) show a steeper SS than that of the SG mode due to the greater control over current in the double – gate FET [43].

As shown in Fig. 5.16b, when $V_{G2}$ is larger than $V_{TDG}$ ($V_T$ in the DG mode), $V_T$ changes significantly in response to small changes in $V_{G2}$. However, changes in $V_T$ are less sensitive to $V_{G2}$ when $V_{G2}$ is lower than $V_{TDG}$. Thus, the sensitivity in terms of $V_T$ is less affected by the condition of $V_{G2}$. 

---

**Fig. 5.15** Schematics of (a) a conventional nanowire FET and (b) a double gate nanowire FET

**Fig. 5.16** (a) $I_D - V_{G1}$ for various $V_{G2}$ conditions and (b) impact of the change in $V_{G2}$ on the change in the threshold voltage (Copyright 2010 American Chemical Society)
As shown in Fig. 5.17, charged analytes (e.g., antibodies) bound to receptors immobilized on the nanowire attract/repel the inversion layer (channel) depending on their charge polarity. In this way, the $V_T$ value is changed. As shown in Fig. 5.17a, because the channel is formed close to the G1 side under the condition of $V_{G2} < V_{T,DG}$, G1 can control the channel efficiently, resulting in a small $\Delta V_T$ value. Under the condition of $V_{G2} > V_{T,DG}$ as shown in Fig. 5.17b, however, the channel is relatively far from the G1 side; hence, G1 loses its ability to control the channel conductivity, leading to a large $\Delta V_T$ value [44].

To test sensing ability, bio-experiments using a specific analyte-receptor binding system for the detection of the anti-AI were performed, as shown in Fig. 5.18a. An AI antigen (Ala) fused with silica-binding protein (SBP) was immobilized on the surface of the nanowire via the SBP domain which serves as an anchor. The specific binding between SBP-Ala and anti-AI was then accomplished by introducing anti-AI onto an SBP-Ala immobilized device.
The inherent negative charges of anti-AI increase the $V_T$ and lead to a positive $V_T$ shift, as shown in Fig. 5.18b. When a $V_{G2}$ value of -2.0 V ($< V_{T,DG}$) is applied, there is no remarkable change in $V_T$. In contrast, $V_T$ changes significantly when a $V_{G2}$ value of 0.5 V ($> V_{T,DG}$) is used. It is noteworthy that this bias condition shows enhanced sensitivity compared to the condition of a floating $V_{G2}$, which is the most similar to a conventional single-gate nanowire FET.

Even for the same nanowire dimensions, the double-gate nanowire FET with a supportive gate (G2) can deliver enhanced sensitivity compared to the single-gate nanowire FET. Thus, the difficulties in scaling down nanowire biosensors can be overcome by implementing a double gate using matured CMOS technology.

### 5.5 Sensing Metrics in a DMFET

Currently, the detected response from FET-based biosensors, that is, the sensing parameter, is always a $V_T$ shift. However, this restricted sensing scheme leads to several problems: (1) The scope of possible analyses is limited because only the detection of biomolecules themselves is possible. An attempt to calculate the charge-trapping properties of DNA using the $V_T$ shift was made [45]; however, this did not provide any physical or electrical meaning about the DNA regarding the amount of the trapped charge. (2) In addition, as only a small amount of electrostatic force under a low concentration of target molecules participates in the biosensing procedure, the net response is generally less than an order of magnitude; thus, this type of sensing scheme is associated with a low signal-to-noise ratio [46]. To improve the sensitivity relative to noise, the dimensions of the sensor device must be scaled down, which complicates the fabrication of the device [47].

Actually, a FET has many useful and sensitive device parameters aside from $V_T$. Every device parameter can be utilized as a sensing parameter for detecting biomolecules. In this section, new sensing parameters for use in a DMFET are discussed.

#### 5.5.1 Interface Trap Charge Method

One device parameter that can be utilized as a sensing parameter is the number of interface states at the surface of the channel, or the channel/gate dielectric interface (normally Si/SiO$_2$), which is also referred to as the interface trap density ($D_{it}$). The interface trap is a very sensitive parameter that can affect the device characteristics severely. Therefore, many techniques to investigate and extract $D_{it}$ electrically have been developed; among them, charge-pumping [48] and $1/f$ noise measurements [49] have been studied since the 1960s. A recent and related development is the advent of a biosensing technique that detects the interface state modulation resulting from biomolecular interactions.
In this section, two newly developed sensing techniques for FET-based biosensors, charge-pumping [50–53], and 1/f noise measurement [54] are discussed. Their advantages and weaknesses, operation principles, and technical issues are also described.

5.5.1.1 Charge-Pumping Technique

The operating principle of the charge-pumping technique is as follows: a DMFET is fabricated in which the gate dielectric is partially etched to form a nanogap, as shown in Fig.5.19a. Interface traps are located at the interface between the channel and the gate dielectric, and the measured charge-pumping current \( I_{CP} \) is proportional to the interface trap density [55]. Therefore, because the nanogap exposes the silicon channel to biomolecules directly, additional traps can be provided by biomolecules immobilized inside the nanogap; consequently, variation of the trap density can occur with a measurable quantity \( I_{CP} \) that is highly sensitivity.

The great advantage of the charge-pumping technique is its sensitivity; when the frequency and the level of the applied pulse were optimized by the prediction from the derived analytical model [52], the micro-sized FET showed high sensitivity that was comparable to a nanowire biosensor without a dimension scaled to the nanoscale. This makes the fabrication of a highly sensitive biosensor at a low cost feasible. Figure 5.20 shows the measured values of \( I_{CP} \) as a function of the charge-pumping frequency \( f_{CP} \). It was verified experimentally that the sensitivity can be improved if a lower pulse frequency is used during the charge-pumping measurement [52]. The sensing margin can be improved if \( f_{CP} \) is lowered; consequently, the sensitivity can fall below the picomolar concentration regime without scaling the physical size of the sensor.

Another advantage is that the charge-pumping technique is able to analyze various properties of biomolecules electrically. Hence, not only does it enable the detection of biomolecules, but it also extracts their fundamental electrical properties. For example, the identification of the biomolecular charge polarity was demonstrated using a charge-pumping technique [53]. When negatively or positively charged biomolecules are immobilized in the nanogaps, the \( V_T \) of a FET is not uniform along the channel but instead varies locally, as shown in Fig. 5.21a. Accordingly, if the maximum peak level of the pulse \( V_h \) is increased gradually, a lateral \( V_T \) profile can be expected. Consequently, the biomolecular charge polarity can be identified. The experimental results are provided in Fig. 5.21b and c, showing that the biomolecular charge polarity was successfully determined by the shift of the direction of the \( dI_{CP}/dV_h - V_h \) curves. Therefore, the charge-pumping technique is useful in that it enables the analysis of various electrical properties of biomolecules and can be utilized as an investigational tool to extract their fundamental properties and their biosensing characteristics.
Fig. 5.19 (a) A schematic diagram of a DMFET and the experimental setup for the charge-pumping measurements [52]. (b) Experimental results comparing the measured $I_{cp}$ values before and after the binding of biotin-streptavidin in the nanogap. When the concentration of a biotin/phosphate-buffered saline (PBS) solution is fixed at 10 mM, the modulation of $I_{cp}$ depends on the streptavidin/PBS solution concentration, which ranged from 20 pM to 2 nM in this case [51] (Copyright 2010 American Institute of Physics)

Fig. 5.20 Measured $I_{cp}$ values as a function of the frequency. The sensing margin is increased in the low-frequency range. The sensitivity can fall below the picomolar concentration range [52] (Copyright 2010 American Institute of Physics)
Fig. 5.21 (a) A schematic diagram showing the operating principle of the negatively charged biomolecules and the expected $dI/dV_h$ versus $V_h$ [53]. (b) The $dI_{cp}/dV_h$ versus $V_h$ characteristics. The shift direction in the $dI_{cp}/dV_h - V_h$ plot indicates the charge polarity of the biomolecules [53]. (c) The measured $dI_{cp}/dV_h$ data with two different pH solutions. When streptavidin is at pH 3.5, the expected charge polarity of streptavidin is positive. Therefore, the peak of $dI_{cp}/dV_h$ is shifted to the left side from initial ($V_0$) value due to the locally increased $V_T$ near the nanogap region (Copyright 2010 American Institute of Physics)

5.5.1.2 1/f Noise Measurement

The $1/f$ (flicker) noise in FET systems has been studied for more than four decades. It is known that the $1/f$ noise comes from the random trapping and detrapping processes of charges in the interface traps located at the Si/SiO$_2$ interface (channel/gate dielectric). The charge fluctuation results in fluctuation of the surface potential, which in turn modulates the channel carrier density and conductance. This type of noise is one of the limiting factors for biosensing. To distinguish the signal from noise clearly, the noise spectra in the frequency domain may allow contributions from different noise sources to be analyzed directly.

As shown in Fig. 5.22a, the concentration-dependent conductance change indicates that it is difficult to distinguish the signal from noise when the PSA concentration is at or below 0.15 pM. In contrast, when the device is in pure buffer, a clear $1/f$ spectrum (noise) can be observed in the frequency domain (Fig. 5.22b). When solutions of 0.15 pM PSA were delivered, the power spectra showed a curved-shape signal that was clearly different from that measured in buffer.

Therefore, the main benefit of $1/f$ noise technique is that sub-picomolar detection has been routinely achieved based on the fact that the characteristic frequencies associated with protein binding are well separated from other noise sources. An increase in the detection sensitivity of more than tenfold has been achieved with the frequency domain compared to time domain measurements from the same device.
The techniques described above, charge-pumping and \(1/f\) noise measurement, were recently developed. Therefore, only a few preliminary studies of the techniques have been reported thus far. Both techniques demonstrated higher sensitivity compared to previous FET-based biosensors; however, the stability, reliability, and reproducibility of the biosensor operation have yet to be confirmed. In particular, the measurement procedure of both techniques is complicated, which could be problematic for their use in POCT systems. For the charge-pumping technique, a continuous pulse stream must be applied to the gate electrode, which requires an additional peripheral circuit to generate pulse. To measure the \(1/f\) noise, additional equipment (generally an amplifier and a spectrum analyzer) is required for high-quality measurements, and these parts are not compatible with the miniaturization desired in POCT systems.

### 5.5.2 Substrate Current Method

Another parameter significantly affected by the status of the nanogap site in a DMFET is the substrate current \(I_{\text{sub}}\). The substrate current is generated when a high drain bias is applied while the channel is inverted. The high drain bias makes a high lateral electric field near the drain junction edge. Electrons accelerated by the
lateral electric field from the drain voltage trigger impact ionization, thus generating electron–hole pairs. The resulting electron–hole pairs are driven by the electric force produced by the applied bias on the substrate, and therefore, they create the substrate current, which is comprised of holes in an n-channel FET. Equations 5.6 and 5.7 express the substrate current and saturated drain voltage ($V_{D_{sat}}$) which predominantly affect the substrate current:

$$I_{sub} \propto (V_D - V_{D_{sat}})I_D \exp \left(-\frac{B}{V_D - V_{D_{sat}}}\right)$$

(5.6)

$$V_{D_{sat}} = V_G - V_T$$

(5.7)

Here, $B$ is the impact ionization coefficient [56]. The substrate current increases monotonically as $V_G$ increases at a low or intermediate level of $V_G$ because the first linear term on the right side in Eq. 5.6 is dominant over the substrate current. However, when $V_G$ is very high, the substrate current is reduced by the second exponent term in Eq. 5.6. Thus, $I_{sub}$ appears bell-shaped, with a crucial voltage ($V_G@I_{sub,max}$) that shows the maximum $I_{sub}$ value upon the first increment of $I_{sub}$ led by an increase in the $V_G$ value, with the next decrement of $I_{sub}$ driven by the exponent decrease in Eq. 5.6. The $V_G@I_{sub,max}$ value is significantly affected by the maximum electric field in which impact ionization occurs. Thus, the changed electric field near the gate edges, that is, near the drain junctions due to air and biomolecules in the nanogap, results in different $V_G@I_{sub,max}$ values. The results as they pertain to the substrate current have yet to be reported, but $I_{sub}$ could be one of the sensing parameters used in a DMFET.

## 5.6 Environmental Effect

Most previously reported biosensors were characterized under aqueous conditions [29, 47]. However, a few biomolecular detection experiments were also performed in ambient air environments [18, 30] (depending on the exposed condition, these are referred to as a “watery environment” or a “dry environment,” respectively). Although measurements in a watery environment are common in biosensor characterizations because aqueous conditions maintain the functionality of biomolecules, measurements in a dry environment facilitate various device structures without consideration of the isolation between the aqueous solution and the device. Hence, characterizations of a biosensor in a bionic solution and in air ambient have been performed [57]. With the same device structure as an underlap FET, the researchers validated the biosensor functionality in a dry environment by comparing the result in watery environment.

To maintain biosensor operation in a bionic solution, an additional passivation layer that prevents leakage current through the bionic solution was implemented in the underlap FET, as shown in Fig. 5.23. The device was submerged in a bionic
solution or kept in ambient air for characterization in watery or dry environment, respectively.

As mentioned in Sect. 5.4.1, the channel potential of the underlap region is highly sensitive to the charge on it. An external charge on an open area will affect the drain current. Thus, a change in the drain current can be observed after target molecule binding in both wet and dry environments. The drain current reduction in the dry environment was shown in Sect. 5.4.1, and the result under the watery environment is demonstrated in Fig. 5.24. The drain current reduction in Fig. 5.24 can also explained by the electrical effect of the negatively charged biomolecules (anti-AI) in the underlap region.

To ensure the specific binding of anti-AI and AIa on the underlap FET, false-positive tests were performed in both watery and dry environments. First, a non-AIa-functionalized device was immersed in an anti-AI solution for 1 h; after which it was rinsed several times. As shown in Fig. 5.25, there was no change in the drain current for the nonfunctionalized device, even after anti-AI was added. In another control experiment, an AIa-functionalized device was exposed to an anti-rabbit-IgG antibody solution which showed nonspecific binding with AIa for 1 h; after which
it was rinsed several times. This experiment also resulted in a negligible change in the drain current as expected. The results clearly confirmed that the considerable decrease in the drain current for both environments was due to the specific binding of anti-AI and AIa. Thus, it was verified that the biosensor characteristics measured in dry environments are valid to the same extent as they are in watery environments.

Acknowledgements This work was supported by the National Research and Development Program under grant NRDP, 2012-0001131 for the development of biomedical function monitoring biosensors and by the Center for Integrated Smart Sensor through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology under Grant CISS-2011-0031845.

References


Chapter 6
Semiconductor-Based Biosensing Chip for Point-of-Care Diagnostics

Toshiya Sakata

Abstract Nano-micro and semiconductor technologies can be used to develop novel biosensing devices for molecular recognition and cellular functional analysis. These devices offer key advantages over conventional techniques, such as high sensitivity, high accuracy, and high throughput. By use of field-effect transistors (FETs) based on semiconductor technology, we investigate electrical characteristics of cell membrane with molecular charges on the gate surface as sensing area. Moreover, we analyze electrical phenomena of cell membranes and variability of drug responses for each organ (liver, pancreas, and so on) using FETs. For high-performance screening and sensing devices in regenerative medicine and pharmaceutical lead discovery, we utilize field-effect-based drug screening devices crossing disciplines between biology and electronics.

6.1 Background

In 1956, John Bardeen, Walter H. Brattain, and William B. Shockley were jointly awarded the Nobel Prize in Physics “for their researches on semiconductors and their discovery of the transistor effect.” The semiconductor technology has been developed in the field of electronic device industry according to Moore’s Law [1]. The size of gate has reached to nm scale (~ a few 10 nm) in the present day, and reducing of gate size has been reported to be difficult in case of silicon-based semiconductor technology. The downsizing of gate will be realized by the use of carbon nanotube- or graphene-based nano-materials in the future.
Biological phenomena are found as molecular behaviors such as DNA, antigen-antibody reaction, or ligand protein-receptor interaction at cell membrane at the nanoscale. On the other hand, microelectromechanical systems (MEMS) technology is useful to produce various micro- or nanoscale devices by top-down approach [2]. The developed devices are composed of regular and functional structure, on which biomolecules are arrayed and detected in a multiplex system. Functional biomolecules are controlled on the multiarrayed devices and in the microfluidic system, which are detected with an optical or electrochemical method.

In 1970, Bergverd et al. showed electrical detection method of pH variation based on change of positive charges of hydrogen ions using field-effect transistor (FET) on the basis of semiconductor principle [3]. This is called ion-sensitive field-effect transistor (IS-FET), as shown in Fig. 6.1. Semiconductor material is separated with solution across gate insulator, of which the thickness is not more than a few hundred nm. The gate insulator is usually composed of oxide such as SiO$_2$, Ta$_2$O$_5$, and Al$_2$O$_3$ or nitride such as Si$_3$N$_4$ [4]. The hydroxyl groups are formed at the surface between solution and gate insulators and are so sensitive to hydrogen ions. These positive charges at the gate surface interact electrostatically with electrons at the channel in silicon crystal. The field effect caused by charge density changes at the gate induces the change of drain-source current and threshold voltage ($V_T$). This electrical response to hydrogen ions is applicable to Nernstian response, 59.1 mV/pH.

Recently, FET biosensor is being studied and developed to apply for clinical diagnosis, drug discovery, tissue engineering, and so on. Since FET biosensor can detect molecular recognition events accompanied by charge density changes without labeled materials and be easily arrayed by use of the conventional semiconductor processes in order to measure multisamples, the platform based on FET chips is suitable for a simple and cost-effective system for chip-based diagnosis. The downsizing of system is significant for personalized medicine at home. Moreover, the
electrical signals based on FET devices result in direct and quantitative analyses of bio-samples. One positive or negative charge of ion or ionic molecule interacts electrostatically with one electron charge in semiconductor device. Therefore, ion behaviors based on biological phenomena can be directly detected using semiconductor devices. Most of biological phenomena in vivo are closely related to charged mediums, for example, such as DNA molecules with negative charges based on phosphate groups, ions (potassium, sodium, and so on) through ion channel at cell membrane keeping homeostasis.

6.2 Concept of Semiconductor-Based Biosensing Devices

6.2.1 Ion-Sensitive Field-Effect Transistor (IS-FET)

As a reason of a bad tooth, three elements of “Streptococcus mutans,” “quality of tooth,” and “saccharinity” affect on a bad tooth in the course of time. S. mutans induces acidification by dissolving saccharide in food and drink. As a result, acidification of dental plaque is in progress on a tooth. That is to say, enamel of tooth begins to dissolve less than pH 5.5, resulting a bad tooth. Therefore, it is important to control meal considering pH variation in a mouth in order to prevent a bad tooth. Thus, pH measurement is needed even for health care in daily life and can be accomplished by engineering such as semiconductor technology.

The principle of IS-FET is based on potentiometric detection of charge density changes induced at a gate insulator/solution interface accompanied by pH variation. Hydrogen ions with positive charges at the gate insulator electrostatically interact with electrons in silicon crystal across the thin gate insulator, resulting in the $V_T$ change.

Typical drain voltage ($V_{DS}$)–drain current ($I_D$) characteristic of the FET is shown in Fig. 6.2. It is found that the FET can be operated correctly. Since the fabricated FET is depletion type as can be seen in Fig. 6.2, the reference electrode is usually connected to the ground for the measurement of the interface potential between gate insulator and solution using the circuit shown in Fig. 6.3. The pH-response characteristics of the FET with a Si$_3$N$_4$ gate are shown in Fig. 6.4. The time course of the interface potential was measured during calibration and is shown in Fig. 6.4a. The arrows indicate the timing to change the buffer solutions. The interface potential changed rapidly after changing the buffer solution and became stable within 1 min. The calibration curve for the Si$_3$N$_4$ gate FET is shown in Fig. 6.4b. The relationship between pH and the output voltage is linear in the range from pH 1.68 to 9.18 with a correlation coefficient of 0.9999. The slope of the calibration curve was 57.52 mV/pH, which is close to the theoretical slope at 25°. On the basis of these results, the operation of the Si$_3$N$_4$ gate FET was considered to be stable, and no leakage through the gate insulator and no defect of the encapsulation could be observed.
6.2.2 Device Structure

Insulated gate field-effect transistor (IG-FET) is composed of electrolyte-insulator-semiconductor. The structure is utilized as pH sensor of IS-FET. The electrical characteristics of IG-FET are the same as the one of IS-FET shown in Sect. 2.1.

On the other hand, extended gate field-effect transistor (EG-FET) is shown in Fig. 6.5. The gate electrode is separated and extended from metal oxide semiconductor (MOS) FET. Using the EG-FET, materials and structures of gate electrode can be varied and gate sensing membranes can be easily arrayed by sputtering, and MOS-FETs are not replaced and can be reused. Various kinds of gate materials can be designed as follows.
6.2.2.1 Gold (Au) Gate Electrode

It is simple to immobilize biomolecules on Au electrode, because spacer molecules with -SH (thiol groups) are easily tethered. DNA or antibody molecules are immobilized by binding with spacer molecules on Au. However, Au surface is polarized in solutions and a constant potential based on ion charges is not determined. Therefore, it is effective to tether self-assembled monolayer (SAM) with both of functional groups (−COO−, −NH3+, and so on) with charges and −SH at each end on Au surface for electrical stabilization of Au surface potential.

Transparent electrode is available for microscopic observation of cultured cells as if performed using the conventional culture dish. The surface of ITO is composed
of oxide so that it is covered by hydroxyl groups in solutions. The potential of EG-FET with ITO-gate shows near Nernstian response to pH variation [5]. However, it is very important to control making process of ITO surface.

Using the EG-FET, the structure of gate itself can be varied corresponding to applications. In the field of tissue engineering, the safety and functionality of artificial organs are required to transplant to living body actually. This is why 3D culture of cells is required for making artificial organs. In order to perform 3D culture on electrode, nano- or microfiber scaffold is able to be developed as gate electrode of EG-FET. Scaffold architecture affects cell binding and spreading. Cells binding to scaffolds with microscale architectures flatten and spread as if cultured on flat surfaces. Scaffolds with nanoscale architectures have larger surface areas to adsorb proteins, presenting many binding sites to cell membrane receptors. The adsorbed proteins may also change conformation, exposing additional cryptic binding sites [6].

The fiber scaffold gate electrode (FSG) can be prepared by the electrospray deposition (ESD) method [7]. The gold micro-FSG is made by sputtering after silica fiber is prepared using the ESD method, as shown in Fig. 6.6. The sputtering of gold is performed for both side of silica fiber membrane. The thickness of silica fiber membrane is about a few hundred μm, and the several sputtered fiber membranes are piled up for the FSG. The FSG is connected to the gate of MOS-FET, and the measurement system is composed as shown in Fig. 6.7. Using the system with the FSG, the constant potential of interface between fiber surface and solution can be measured in a phosphate buffer solution of pH 7.41 or a culture medium (Dulbecco’s Modified Eagle’s Medium; DMEM). Moreover, HeLa cells, which are a cell type in an immortal cell line derived from cervical cancer cells taken from Henrietta Lacks, a patient who eventually died of her cancer, are cultured on the FSG as shown in Fig. 6.8. Calcein acetoxymethyl ester (AM) is used as fluorescent dye for live cell staining. HeLa cells are found even inside the fiber.
The measurement of one of cell functions, programmed cell death (apoptosis), has been tried using the FSG-based FET. When apoptosis is induced by ligand proteins such as TNF-related apoptosis-inducing ligand (TRAIL), volume reduction of cells is observed accompanied by DNA fragmentation, chromatin concentration, and so on in a cell resulting in fragmentation of cells, as shown in Fig. 6.9. In particular, the previous work has shown the possibility of \( \text{H}_2\text{O}, \text{K}^+, \text{and Cl}^- \) release through cell membrane before shrinkage of cells in the early stage of apoptosis [8]. Therefore, their ionic behaviors are detected by use of the principle of field effect of semiconductor devices. When TRAIL is introduced to HeLa cells on the FSG, the interfacial potential between the FSG and the culture medium increased gradually after the addition of TRAIL, as shown in Fig. 6.10. The positive shift of potential shows the increase of positive charges of potassium ions based on apoptosis or the detachment of cells due to apoptosis (the decrease of negative charges based on sialic acid at cell membrane) at the interface between the FSG and the solution.
Maybe both changes would be included in the electrical signal. Moreover, the FSG-based FET without cells showed no potential shift after the addition of TRAIL, although the effect of background noises such as temperature change, change in ion concentration, and so on was observed in introducing it into the sample. Thus, it is important to compare the signal of target sensor with the one of control sensor to consider various background noises. Actually, the potential shift for the FSG-based FET seemed to be larger than the signal for apoptosis measured by the IG-FET. This may be because the surface area of FSG was larger than that of IG-FET, resulting in the increase of number of adhered cells due to fibrous electrode. However, the FSG-based FET needs to be improved in order to detect potassium ion specifically and selectively by modifying chemically the fiber surface.

A biosensing technique to measure molecular charge distribution at cell membrane was developed using a principle of semiconductor. In this research, a cantilever of atomic force microscopy (AFM) was utilized in order to contact with cell membrane and to move electrode at nanoscale, and a “gatelever” with platinum tip was developed to detect charges at cell membrane using AFM. The gatelever as electrode is extended from MOS-FET as one of the EG-FETs. Actually, negative charges based on sialic acids and so on at cell membrane were detected using AFM with gatelever-based FET. As a result, membrane protein with some charges at cell membrane can be detected at nanoscale using the proposed system.
6.3 Signal Transduction Interface

6.3.1 Physically Structured Interface

Micro- and nanostructure on gate electrode of semiconductor devices contribute to increase of surface area for immobilization of probe molecules and amplification of signals. Nanoparticles or nanopillar can be designed as nanostructure, of which gold, silica, carbon nanotube, and so on are utilized. In particular, detection limit of FET devices is closely related to Debye length at the interface between gate surface and solution. When nanopillar structure is made within Debye length, biomolecules with longer or larger size such as DNA or protein can be detected within Debye length. In case of DNA molecules, particularly, more extended DNA molecules based on one base extension reaction can be detected parallel to gate surface even for longer DNA molecules (Fig. 6.11), resulting in label-free DNA sequencing based on intrinsic molecular charges.

6.3.2 Chemically Synthesized Interface

Functional interfaces synthesized chemically are proposed as monolayer or polymer membrane. A functional membrane as monolayer sensitive to $K^+$ at the gate insulator is prepared using 18-crown-6 ether derivative, as shown in Fig. 6.12. First, carboxyethylsilanetriol was used as silane coupling agent. The sensor was immersed in carboxyethylsilanetriol sodium salt 25 wt% in water for 24 h, then rinsed with Milli-Q, and dried in a vacuum at 120°C for 1 h. Second, 4-aminobenzo-18-crown-6 (AB18C6) was used as ionophore of $K^+$, and 1-ethyl-3-(3-dimethyaminopropyl) carbodiimide (EDC) was used as condensation agent. The sensor was immersed in AB18C6 (1 mM) and EDC (10 mM) in $N,N$-dimethylformamide (DMF) for 24 h.
and rinsed. Finally, the chemical structure shown in Fig. 6.12 was synthesized. This monolayer does not contain plasticizer, and ionophore is chemically incorporated to the gate insulator, so there is no expected cytotoxicity.

In order to investigate the electrical characteristics of FET, a semiconductor parameter analyzer was used. Measurement was performed in 2[4-(2-hydroxyethyl)-1-piperazinyl] ethanesulfonic acid (HEPES) buffer solution, and pH of the sample solution was adjusted to 7.4 with Ca(OH)$_2$ ([Ca(OH)$_2$] = 1.7 mM). And K$^+$ concentration was adjusted to 10$^{-4}$, 10$^{-3}$, 10$^{-2}$, and 10$^{-1}$ M with KCl. The prepared FET was immersed in the sample solution, and then the $V_T$ change was measured. The effect of K$^+$ concentration on the electrical signal of prepared FET was investigated. The electrical response of prepared FET to K$^+$ concentration was 4.2 mV/decade. The slope was relatively small compared to that of the Nernstian response. This could be because the density of ionophore immobilized onto the FET gate surface is low. Further investigation toward improved sensitivity and selectivity to K$^+$ is currently underway. The surface potential shift to negative direction indicates the increase of K$^+$ with positive charge on the crown ether-modified gate insulator. Thus, the variation of K$^+$ concentration was successfully detected using the prepared FET. The prepared device showed the electrical signals of potassium ion release through cell membrane caused by programmed cell death “apoptosis.” Also, measurement of DNA molecular recognition or antigen-antibody reaction requires chemical immobilization of probe molecules for target ones using spacer molecules of monolayer or polymer membrane.

6.3.3 Biologically Induced Interface

Figure 6.13 shows a schematic illustration of biologically induced interface. A cell A is composed as an interface functional cell and induce a signal based on a target cell B to a semiconductor device as if an interface material plays its role. A cultured cell
A causes a function of cell B due to cell-cell interaction on the gate sensing surface. In this chapter, vascular endothelial cell and tumor cell correspond to cell A and B, respectively, for example. The noninvasive real-time measurement of the invasion of cancer cells to the vascular endothelial cell layer and basement membranes was performed using vascular endothelial cell-based field-effect transistor (VEC-FET), which is based on potentiometric detection of molecular recognition on the gate insulator. The shift of the $V_T$ caused by the charge density change on the gate insulator can be monitored during the invasion process of tumor cells. The negative shift of the $V_T$ has been successfully detected around 4 h after addition of invasive tumor cells (HeLa cells). This result indicates that positive change of charge density was induced on the surface of the gate insulator as a result of degradation of a negative-charged basement membrane on the gate insulator by secretion of enzyme from tumor cells. The platform based on the VEC-FET is suitable for a real-time and simple invasion assay system.

### 6.4 Molecular Recognition Based on Intrinsic Molecular Charges

#### 6.4.1 Genetic Analysis

##### 6.4.1.1 Detection of DNA Molecular Recognition Events

The principle of genetic FET is based on the detection of charge density change on the gate surface which is induced by the specific binding of DNA molecules [9–17]. Oligonucleotide probes are immobilized on the surface of the gate insulator. The genetic FET is immersed in a measurement solution together with an Ag/AgCl reference electrode with saturated KCl solution. The potential of a measurement solution is controlled and fixed by the gate voltage ($V_G$) through the reference electrode. When complementary DNA molecules are contained in a sample solution, hybridization occurs at the surface of the gate area. Since DNA molecules are negatively charged in an aqueous solution, a hybridization event can be detected.
by measuring a shift of the $V_T$. When a n-channel FET is used, the $V_T$ shifts in the positive direction in response to DNA molecules. When intercalators are introduced into the double-stranded DNA after hybridization, the $V_T$ of the genetic FET shifts in the negative direction, because the intercalators are positively charged. Since the intercalators react specifically with double-stranded DNA, an undesirable background noise caused by nonspecific adsorption of single-stranded target DNA can be eliminated, and a more precise and reliable detection of a hybridization event can be realized.

A specific binding of charged biomolecules at the gate surface can be detected as a shift of the $V_T$, which can be determined in the gate voltage $V_G$ and the drain current $I_D(V_G - I_D)$ characteristics of the genetic FET (Fig. 6.14). The $V_G - I_D$ characteristics of the genetic FET shifted along the gate voltage $V_G$ axis in the positive direction after immobilization of oligonucleotide probes. Immobilized oligonucleotide probes were a normal type for R353Q locus of factor VII gene (probe: 5$'$-amino group-CCACTACCGGGCAGT-3$'$ (17mer), target: 5$'$-ACGTGCCCCGGTAGTGG-3$'$ (17mer), melting temperature: 60°C) [18]. In order to evaluate the $V_T$ shift in more detail, the local area shown in Fig. 6.14b (surrounded area) was magnified. The $V_T$ shifts after hybridization and specific binding of DNA binder are also shown in Fig. 6.14. When oligonucleotide probes were immobilized on the gate surface, the $V_T$ shifted along the $V_G$ axis by the amount of 32 mV. The positive shift is due to negative charges induced at the gate surface after immobilization process including cleaning, silanization, glutaraldehyde treatment, immobilization of oligonucleotide probes, and blocking with glycine. Immobilization of oligonucleotide probes and glycine blocking is considered to contribute to the $V_T$ shift to a large extent. When the complementary target DNA was introduced to the gate surface and hybridized with oligonucleotide probes, the $V_T$ shifted in the positive direction by the amount of 12 mV. This is due to increase of negative charges of the target DNA by hybridization. After hybridization, a DNA binder, Hoechst 33258, was introduced to the gate surface. The $V_T$ shifted in the negative direction by the amount of 14 mV. The negative shift
of the $V_T$ indicates increase of positive charges at the gate surface and is due to specific binding of Hoechst 33258 to the double-stranded DNA. This is in contrast to the positive change of the $V_T$ due to negatively charged DNA molecules. Thus, the charge density change at the gate surface after each molecular recognition event can be successfully detected using genetic FET.

DNA recognition events such as primer extension reaction can be also directly detected as electrical signal by use of the genetic FET. The 11-base ligonucleotide probes on the genetic FET were hybridized with the 21-base target DNA at first. In order to extract small changes of the output voltages of the genetic FET after extension reaction, the output voltages before the introduction of DNA polymerase for both genetic and reference FET were initialized and adjusted to zero as shown in Fig. 6.15 by adding or subtracting offset voltages, although the absolute output voltages are not zero and different between active and reference FET. After washing, the genetic FETs were immersed in a reaction mixture and thermostable DNA polymerase was introduced into the gate surface. The $V_T$ of the FET changed during primer extension reaction as shown in Fig. 6.15. Differential measurement was performed using a pair of FETs; one is the genetic FET with immobilized oligonucleotides probes, and the other is the reference FET without oligonucleotide probes.
The common changes of the $V_T$ due to ambient salt and pH changes and temperature difference in the buffer solution can be canceled out using the differential measurement. The difference of the $\Delta V_T$ between the genetic FET and the reference FET during extension reaction, $\Delta V_{T\text{diff}}$, increased drastically up to about 10 mV. This positive change in the $\Delta V_{T\text{diff}}$ is rightly due to negative charges of polynucleotide extended by primer extension reaction. The $\Delta V_T$ of the reference FET is considered to be mainly due to the temperature change. From Fig. 6.15, we could demonstrate that the primer extension event on the gate surface was transduced directly into the electrical signal by the use of the genetic FET.

The effect of base length of the target DNA on the $\Delta V_{T\text{diff}}$ was investigated (Fig. 6.16 and Table 6.1) [18]. The linear relationship between the base length and the $\Delta V_{T\text{diff}}$ was obtained up to 41 bases. The $\Delta V_{T\text{diff}}$ after extension reaction increased to 24 mV, when target DNA with 41 bases was used. This is because the number of charges on the gate surface increased after primer extension with increasing template base length. However, the $\Delta V_{T\text{diff}}$ did not follow the linear relationship, when the target DNA samples with 51 bases and 61 bases were used. The reason for this nonlinearity is considered to be related to the width of the electrical double layer at the interface between the gate insulator and an aqueous solution. The width of the electrical double layer, the Debye length, which is expressed in Eq. 6.1, is about 10 nm in the diluted salt solution (approximately 1 mM) and about 1 nm in the physiological solution (approximately 100 mM).

$$\delta = \left(\frac{\varepsilon_0 k T}{2 \varepsilon^2 q^2 I}\right)^{1/2}$$

(6.1)
Table 6.1 Base length of probe and target DNA

<table>
<thead>
<tr>
<th>Locus</th>
<th>Function</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor VII</td>
<td>-122–normal</td>
<td>5’-amino group-CGTCCTCTGAA-3’ (11mer)</td>
</tr>
<tr>
<td>gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>probe</td>
<td></td>
<td>5’-AGCTGGGGTGTTCAGAGGACG-3’ (21mer)</td>
</tr>
<tr>
<td>target</td>
<td></td>
<td>5’-TGCAAGCTTCTCAGCTGGGGGTGTTCAGAGGACG-3’ (31mer)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-GGCGGCCAGGTGCAGCTCTCAGCTGGGGGTGTTCAGAGGACG-3’ (41mer)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-CATGGCCACTGGCGGCCAGGTGCAGCTCTCAGCTGGGGGTGTTCAGAGGACG-3’ (51mer)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5’-GCAGGGGATGCGATGGCCACTGGCGGCCAGGTGCAGCTCTCAGCTGGGGGTGTTCAGAGGACG-3’ (61mer)</td>
</tr>
</tbody>
</table>

where \( \delta \) is the Debye length, \( \varepsilon \) is the permittivity of the electrolyte solution, \( \varepsilon_0 \) is vacuum permittivity, \( k \) is the Boltzman constant, \( T \) is the absolute temperature in Kelvin, \( z \) is the valency of the ions in the electrolyte, \( q \) is elementary charge, and \( I \) is the ionic strength of the electrolyte.

The charge density change induced within the Debye length can be detected with the genetic FETs, while the charge density change induced outside the Debye length is shielded by counter ions and cannot be detected with the genetic FETs. In the present study, a 25 mM phosphate buffer solution was used for the VT measurement. The Debye length at the gate insulator surface is therefore considered to be a few nanometers. The length of the target DNA with 41 bases is 13.94 nm, when it is straight. But oligonucleotide probes and the target DNA are flexible in the aqueous solution and oligonucleotide probes are not always perpendicular to the surface of the gate insulator. It is therefore reasonable to ascribe saturation of the linear relationship between the VT shift and the base length over 51 bases to the Debye length. Since the Debye length is dependent on the ionic strength of the aqueous solution as shown in Eq. 6.1, it is important to optimize the buffer concentration used for the measurement of the VT shift.

### 6.4.1.2 Immobilization Density of Oligonucleotide Probes

The molecular recognition events such as hybridization and specific binding of DNA binder could be directly transduced into electrical signal using the genetic FETs. The change in the surface charge density could be detected as a shift of the VT of the genetic FETs. The VT shift after hybridization shown in Fig. 6.14, \( \Delta V_T \), can be expressed in Eq. 6.2, where \( Q_{ds-DNA} \) is the charge per unit area of the double-stranded DNA after hybridization, \( Q_{ss-DNA} \) is the charge per unit area of the single-stranded oligonucleotide probes, \( \Delta Q_{DNA} \) is the charge difference per unit area after hybridization, and \( C_i \) is the gate capacitance per unit area.

\[
\Delta V_T = \left( Q_{ds-DNA} - Q_{ss-DNA} \right) / C_i = \Delta Q_{DNA} / C_i \tag{6.2}
\]
Since $\Delta V_T = 12 \text{ mV}$ and $C_i = 4.3 \times 10^{-4} \text{F/m}^2$ for the genetic FET, the amount of charges increased after hybridization is calculated to be $5.1 \times 10^{-6} \text{C/m}^2$. The base lengths of the oligonucleotide probe and the target DNA used in the experiment are both 17 bases, which correspond to 5.78 nm in length. Negative charges derived from phosphate groups are distributed along the double-stranded DNA from the gate surface to the bulk of the sample solution. We assume that these negative charges along the DNA molecules contributed to the $V_T$ shift equally and that all the oligonucleotide probes were hybridized with the target DNA. Under these assumptions, the number of oligonucleotide probes on the channel region can be calculated to be $2.3 \times 10^4$, which corresponds to $1.9 \times 10^8 \text{cm}^2$. The surface density of oligonucleotide probes immobilized on glass, silicon dioxide, and gold has been reported to be in the order of $10^9 - 10^{13} \text{cm}^2$, which was determined by different methods [19–24]. Since the density of the oligonucleotide probes is strongly dependent on the method and materials used for a substrate and immobilization, the number of oligonucleotide probes immobilized on silicon nitride could be increased by optimizing the immobilization method. It is noted that hybridization with $2.3 \times 10^4$ target DNA molecules resulted in the $V_T$ shift of 12 mV. Therefore, detection of DNA molecules by the use of the genetic FET can be very sensitive, if hybridization is carried out sufficiently.

### 6.4.1.3 Label-Free DNA Sequencing Based on Intrinsic Molecular Charges

Although a number of methods for single nucleotide polymorphism (SNP) analysis have been developed as described in the previous works, DNA sequencing techniques are still required to be improved in terms of cost, simplicity, and high throughput in order to analyze not only SNPs but also genomic variations such as insertion/deletion and short tandem repeat. A new method for DNA sequencing is introduced in this section, which is based on detection of intrinsic charges of DNA molecules using the field effect.

Oligonucleotide probes are immobilized on the Si$_3$N$_4$ gate surface. The complementary target DNA is hybridized with the oligonucleotide probes on the gate surface. The hybridization events are followed by the introduction of DNA polymerase and one of each deoxynucleotide (dTTP, dATP, dGTP, or dCTP). DNA polymerase extends the immobilized oligonucleotide probes in a template-dependent manner (Fig. 6.17). As a result of extension reaction, negative charges increase at the gate surface of the FETs because of intrinsic negative charges of incorporated molecules. This charge density change can be detected as a shift of the $V_T$ of the FETs. Thus, iterative addition of each deoxynucleotide and measurement of $V_T$ allow a direct, simple, and nonlabeled DNA sequencing.

The base sequences of factor VII gene including two SNP sites and that of hereditary hemochromatosis (Table 6.2) [18] were used to demonstrate the principle of DNA sequencing based on the FETs. We have paid special attention to the
buffer concentration to be used for measuring charge density change at the gate surface. The potential change induced by adsorption of proteins at the gate surface was reported to be dependent on the electrolyte concentration [25]. It is therefore important to optimize the Debye length at the gate insulator/solution interface. In the present study, a 0.025 M phosphate buffer solution was used for measuring charge density change at the gate surface, while the conventional reaction mixture was used for single-base extension reaction.

The 11-base oligonucleotide probes were immobilized on the gate surface and hybridized with the 21-base target DNA for the base sequence of -122 (Table 6.2) [18]. We evaluated the FETs in combination with single-base extension for DNA sequencing. We prepared four kinds of buffer solution containing both DNA polymerase and one of dCTP, dATP, dGTP, or dTTP, respectively. The FETs hybridized with target DNA were immersed into the conventional reaction solutions for single-base extension reaction and the $V_T$ shift was measured in a 0.025 M phosphate buffer solution after washing the FETs. The cycle of single-base extension and measurement of the $V_T$ was repeated iteratively to determine the base sequence of the target DNA. When the base sequence of R353Q region of the factor VII gene was used as a target DNA, the $V_T$ shifted in the positive direction only after single-base

---

**Fig. 6.17** Principle of label-free DNA sequencing based on intrinsic molecular charges

**Table 6.2** Base length of probe and target DNA

<table>
<thead>
<tr>
<th>Locus</th>
<th>Function</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>R353Q</td>
<td>probe</td>
<td>5'-amino group-CCACTACCG-3' (9mer)</td>
</tr>
<tr>
<td>R353Q</td>
<td>target</td>
<td>5'-ACGTGCCCCGTTAGTG-3' (17mer)</td>
</tr>
<tr>
<td>-122</td>
<td>probe</td>
<td>5'-amino group-CGTCTCTGCAG-3' (11mer)</td>
</tr>
<tr>
<td>-122</td>
<td>target</td>
<td>5'-AGCTGGGGTCTAGGACG-3' (21mer)</td>
</tr>
<tr>
<td>C282Y</td>
<td>probe</td>
<td>5'-amino group-AGATATACGTG-3' (11mer)</td>
</tr>
<tr>
<td>C282Y</td>
<td>target</td>
<td>5'-CTCCACCTGGACGTATAC-3' (21mer)</td>
</tr>
</tbody>
</table>
extension with the specific deoxynucleotides which were complementary to the base sequence of the target DNA (Fig. 6.18a). The $V_T$ change for three-base extension, GGG, was 6.9 mV, which was bigger than that for one-base extension, but was not three times as big as that expected from the number of intrinsic charges. Although the linear relationship between the base length synthesized by the extension reaction and the $V_T$ shift was obtained in the range from 0 to 30 bases (Fig. 6.16), it is important to detect single-base extension quantitatively, in order to reduce base call error especially for continuous sequence of the same base. The density and orientation of the immobilized oligonucleotide probes have to be controlled during a series of extension reactions at 72°C. Further improvement of precision of the base call is also expected by automation of extension reaction and VT measurements.

C282Y region of hereditary hemochromatosis gene was used as another example of DNA sequencing using the FET and single-base extension (Fig. 6.18b). The positive $V_T$ shifts could be detected in accordance with the base sequence of the target DNA. In this case, the average $V_T$ shifts for two-base incorporation was 5.8 mV with the standard deviation of 0.4 mV, while the average $V_T$ shift for single-base extension was 3.2 mV. The $V_T$ shifts for two-base extension was approximately twice as big as that for single-base extension. Thus, the results of iterative extension reaction and detection of the $V_T$ indicated the ability of a direct, simple, and potentially precise DNA sequencing analysis using the FETs. The number of bases which can be analyzed by the proposed method is about ten bases at present. The $V_T$ shift for single-base extension became gradually smaller as the number of bases increased more than ten bases. One of the reasons for this limitation would be the Debye length at the gate insulator/solution interface. Any charge density change induced outside the Debye length cannot be detected with the FETs. Lateral extension reaction in which DNA probes are extended in parallel with the gate surface would be effective for DNA sequencing with long bases. Another reason for the limitation would be peeling off the immobilized oligonucleotide probes from the surface of the gate insulator as the temperature stress of the extension reaction at 72°C is applied repeatedly. The stronger immobilization method for oligonucleotide probes on the Si$_3$N$_4$ surface has to be adopted to analyze longer base sequence.
6.4.2 Detection of Low-Molecular-Weight Antigen

About a decade ago, a novel immunoassay principle called open sandwich immunoassay (OS-IA) was proposed for the detection of mainly low-molecular-weight antigens in a noncompetitive mode [26–31]. This method is based on the antigen-dependent interchain interaction of separated $V_L$ and $V_H$ chains from a single antibody variable region. To perform OS-IA, for example, one of the separated chains is immobilized on the plate. When the sample containing the antigen and another chain that is labeled with an enzyme is added on, both chains reassociate according to the antigen concentration in the sample, resulting in the positive colorimetric or chemiluminescent signal on the plate after washing and substrate addition. In this distinguished method, the detection sensitivity depends on several parameters. One is the affinity of the antibody used, and another is the antigen dependency of the $V_H/V_L$ association. However, equally important is the sensitivity of the detection method, and typical detection limits obtained in the described OS enzyme-linked assays (OS-ELISA) were in the few nanomolar range [27–31].

Since the separated variable regions are smaller than (approximately $2 \times 2 \times 3.5 \text{ nm}$) the original antibody to be captured within the Debye length at the gate surface, the detection principle of bio-FET will overcome the detection limit of OS-ELISA. Compared with the OS-ELISA, the OS-FET has the advantages of label-free and quantitative detection. Because of the label-free measurement, we can omit the experimental process of labeling. Moreover, we can estimate quantitatively the number of target molecules with charges and directly transduce the charge density changes based on the capture of target on the gate into the electrical signals using the OS-FET. In the present study, we propose a new detection method, open sandwich-based immunofield-effect transistor (OS-FET) for a label-free and highly sensitive detection of low-molecular-weight antigen, which is based on the detection of intrinsic molecular charges of one of the separated chains using the field effect. Here, we report on the direct transduction of open sandwich immunoassay at the gate surface into an electrical signal using the FET.

In our setup, the $V_H$ chain is chemically immobilized on the $\text{Si}_3\text{N}_4$ gate surface. Then, the small antigen is sandwiched with the free $V_L$ chain tethered with a negatively charged protein (MBP), and the $V_H$ chain is immobilized on the gate surface (Fig. 6.19). In this study, bisphenol A (BPA) is utilized as model antigen with a molecular mass of 228 and negligible charge at neutral pH [32]. The compound is a widely used small monomer in the manufacture of polycarbonate plastics and epoxy resins and has been reported to have estrogenic effects. While many methods were devised to detect BPA, the detection limit attained was more than the nanomolar range [33–38]. However, recent research suggested that the exposure of BPA at the picomolar range changed some cell functions [38]. Therefore, the development of a more sensitive method to detect BPA is urgently needed. As shown in Fig. 6.19, the free $V_L$ chain used is tethered with MBP with 13 negative charges...
per molecule (pI=5.2), which is one of routinely used fusion tags for recombinant protein production in *Escherichia coli*, partly because it acts as an intramolecular chaperone and increases the expression yields of the fused proteins. Due to the large negative charge of MBP, the reassociation of the two chains is expected to induce the charge density changes at the gate surface. This charge density change can be detected as a shift of the $V_T$ of the FET. Thus, the measurement of $V_T$ allows a direct, simple, and nonlabeled OS-IA.

The FET chip is immersed in a measurement solution together with an Ag/AgCl reference electrode with saturated KCl solution. The potential of a measurement solution is controlled and fixed by the gate voltage ($V_G$) through the reference electrode. We have paid special attention to the buffer concentration used for measuring charge density change at the gate surface. The potential change induced by adsorption of proteins at the gate surface was reported to be dependent on the electrolyte concentration. It is, therefore, important to optimize the Debye length at the gate insulator/solution interface. In the present study, a 0.025 M phosphate buffer solution was used for measuring charge density change at the gate surface, while the conventional reaction mixture was used during OS-IA reaction.

The shift of the gate voltage ($V_G$)-drain current ($I_D$) characteristic, the $V_T$ shift, was measured after reacting BPA with MBP-linked $V_L$ (MBP-VL) and $V_H$ chains (Fig. 6.20). The initial concentration of BPA added to the OS-FET chip was 1 µM. The resultant $V_T$ shifted in the positive direction by the amount of 13.1 mV due to the negative charge of MBP (Fig. 6.20a). On the other hand, the effect of nonspecific adsorption of MBP-VL on the electrical signal was investigated by use of the control FET (Fig. 6.20b). When the control FET chip with the immobilized $V_H$ chains was reacted with the MBP-VL solution without BPA, the $V_T$ shift was 4.5 mV to the positive direction. The electrical signal of the control FET implies not only the antigen-independent basal binding of MBP-VL but also the presence of unexpected BPA, which might be included in the conventional disposal tube used in the experiments. Thus, for the highly sensitive detection of small uncharged antigens using the OS-FET, it is important to perform the differential measurement by the use of the control FET and the OS-FET.
The above results demonstrate that the reassociation of MBP $V_L$ and $V_H$ chains with a small antigen can be directly transduced into electrical signal using the FETs and that the OS-IA can be realized essentially on the basis of intrinsic molecular charges without any labels. From the point of view of practical application, it is required to detect BPA at the possible lowest adverse effect concentration. According to a report of the National Toxicology Program in 2008 (http://ntp.niehs.nih.gov/), the lowest, but controversial, dose levels that can induce a variety of adverse effects are $0.0024 - 0.01 \text{ g/kg/day}$. Assuming that the amounts of BPA are taken from beverages, the concentration is estimated to be $10 - 60 \text{ ng/ml (43–260 nM)}$, which is completely within the working range of OS-FET.

It is possible to integrate multiple FETs and signal processing circuits in a single chip using advanced semiconductor technology. Simultaneous analyses of various small antigens in the environment or in our body can be realized on the basis of the FETs. Since the output of the FET is an electrical signal, standardization of the results obtained is easier than those with the chemiluminescence-based analyses. Therefore, the platform based on the FETs is suitable for a label-free, highly sensitive, and quantitative detection system for small antigen analysis in environmental, food, and clinical research.
6.5 Noninvasive and Real-Time Monitoring of Cell Functions

6.5.1 Metabolism

Glucose-stimulated electrical activity in pancreatic β-cells was provided as evidence of cell electrical activity [39–49], although muscle and nerve cells were believed to be specialized cells with electrical excitability. After the investigation, numerous works on the electrophysiology of β-cells were carried out and it has been elucidated [50–52]. The patch-clamp technique makes it possible to demonstrate the channels that contribute to the generation of electrical activity. The adenosine triphosphate (ATP) regulated \( K^+ \) channels (\( K_{ATP} \) channels) provide the crucial link between metabolic and electrophysiological effects of glucose. \( K^+ \) flows out of the cell through spontaneously active \( K_{ATP} \) channels at low glucose and ATP concentrations. The \( K^+ \) gradients existing over the β-cell membrane results in a negative membrane potential of approximately -70 mV. Increasing glucose concentrations subsequently elevates intracellular ATP concentrations and closes \( K_{ATP} \) channels. Moreover, the generation of electrical activity is a key step in glucose-induced insulin release, and it is widely thought to represent the primary mechanism by which \( Ca^{2+} \) is imported into the β-cell cytosol, causing an increase in intracellular \( Ca^{2+} \) concentration and triggering several processes that ultimately result in insulin exocytosis [53].

Basically, the semiconductor-based biosensing device that was used in this study is an ion-sensitive field-effect transistor (IS-FET), which has been commonly used as a pH sensor [3]. The principle of IS-FET is based on the direct detection of hydrogen ion concentration based on the site-binding model [54] at the gate insulator, as also described above. On the other hand, the glucose-induced insulin secretion process includes the activation of ATP generation, which is closely related to respiration activity in the mitochondrion. The induction of respiration activity would increase hydrogen ion concentration at the cell/gate interface because of \( CO_2 \) release in solutions [55]. As a result of activation of respiration, the increase of hydrogen ion concentration at the cell/gate interface will be directly detected as pH variation using IS-FET. In this section, we introduce the real-time, label-free, and noninvasive monitoring of electrical activity of rat pancreatic β-cells using a cell-based FET, focusing on the correlation between respiration activity accompanied by insulin secretion process due to glucose induction and electrical activity of pancreatic β-cells.

Rat pancreatic β-cells (RIN-5F) were used for the electrical detection of glucose response using the FETs in the present study. The β-cells were introduced to the sensing areas of the FET chip in a culture medium RPMI1640 medium, pH 7.4 (Invitrogen) including 2 mg/ml glucose supplemented with 10% fetal bovine serum (FBS; JRH Biosciences) and 1% penicillin/streptomycin (Invitrogen) at 37°C in an atmosphere of 5.0% \( CO_2 \) for 3 days after being maintained on a culture dish with the controlled RPMI1640 medium at 37°C in an atmosphere of 5.0% \( CO_2 \) for 1 week. In the electrical measurements, 400 \( \mu l \) of controlled medium without glucose was placed in the FET chamber, and 10 mg/ml glucose was added to the culture medium.
Figure 6.21 shows the concept of the cell-based FET. The diagram for the measurement of electrical signals is shown in Fig. 6.21a. The pancreatic β-cells were cultured on the gate surface of the FET, which is composed of the Ta₂O₅/Si₃N₄/SiO₂ layers. The Ta₂O₅ gate surface was not chemically modified with functional molecules. The potential of a measurement solution is controlled using a Ag/AgCl reference electrode with saturated KCl solution. Two types of FETs were prepared in the present study; one is the cell-based FET with pancreatic β-cells cultured on the gate surface, and the other is the reference FET without cells (Fig. 6.21b). Using these FETs, differential measurements were performed in order to eliminate the common background noises such as temperature changes and changes in ion concentration.

Figure 6.22 shows the shift of the surface potential of gate sensing area after introduction of glucose at a concentration of 10 mg/ml using the pancreatic β-cell-based FET. The β-cells aggregated on the Ta₂O₅ gate surface in the culture medium, because it was difficult for β-cells to adhere to the nonmodified gate but easy for them to aggregate like spheroid. As shown in Fig. 6.22, the surface potential of the β-cell-based FET shifted gradually in the positive direction at about 8.5 mV for about 3 h. The average of surface potential changes due to the introduction of glucose was 10.3 mV for about 3 h using five-cell-based FETs. On the other hand, almost no surface potential changes of the reference FET were found, although a smaller shift was detected owing to the background noises such as temperature changes when glucose was injected to the gate. The electrical signal of the β-cell-based FET indicates the increase in positive charges or the decrease in negative charges based on ions or charged molecules through β-cells at the β-cell/gate interface.
following the introduction of glucose into the β-cell-based FET. About five β-cells were cultured on the gate sensing area, of which the gate width was 10 μm and the gate length was 340 μm.

The surface potential changes were based on the charge density changes of the glucose response of about five β-cells. The uptake of glucose into β-cells causes the increase in ATP/ADP concentration, the depolarization of K\textsubscript{ATP} channel, the influx of Ca\textsuperscript{2+} through ion channel, and the exocytosis of insulin and C-peptide together with Zn\textsuperscript{2+} to form microcrystals within secretory granules [56]. In this study, insulin and C-peptide secretion has been detected by use of enzyme-linked immunosorbent assay (ELISA) method based on the reaction with anti-insulin monoclonal antibody, of which the concentration was about 32 ng/ml at 2 h after glucose addition. The resulting molecular and ion charges at the β-cell/gate interface induced the surface potential changes of the cell-based FET that prove the electrical activity of pancreatic β-cells. Furthermore, the positive shift of surface potential can be assumed to indicate the increase in the concentration of hydrogen ions at the β-cell/gate interface, because the nonmodified FET device with the Ta\textsubscript{2}O\textsubscript{5} gate insulator is basically more sensitive to hydrogen ion concentration than those of other ions. This is why we assume that the respiration of β-cells would have been activated during the insulin secretion process following glucose addition. After the increase of electrical signal, the surface potential of β-cell-based FET decreased gradually because of diffusion of hydrogen ions from the cell/gate interface to the bulk solution accompanied by degradation of respiration activity. The insulin secretion process would include the respiration activity of β-cells so that the electrical monitoring by use of the β-cell-based FET might demonstrate the reduction and exhaustion of glucose-induced insulin secretion.
The pH variation at the β-cell/gate interface determined on the basis of respiration activity can be estimated from the electrical signals of the cell-based FET. Basically, the FET chip utilized in this measurement showed the change of gate voltage 51.4 mV/pH for pH variation. The average change in surface potentials after the addition of glucose was about 10.3 mV with a standard error of ±2.2 mV for five-cell-based FETs. Therefore, the respiration activity triggered by glucose caused the change of about pH 0.2 at the interface between the pancreatic β-cells and the gate surface. Strictly, the shift of pH from 7.4 to 7.2 was detected at the β-cell/gate interface by the cell-based FETs. The amount of eliminated CO₂ was calculated to be about $1.2 \times 10^{-8}$ M on the basis of principle that changes in hydrogen ion concentration corresponds to pH variation, according to the equilibrium of CO₂ in solutions. In the calculation, O₂ consumption, ATP synthesis, and so on in the citric acid cycle inside the mitochondrion could be estimated from the amount of eliminated CO₂, which was determined from the electrical signals of the cell-based FET. However, the effect of buffer solution on pH and the diffusion of CO₂ at the β-cell/gate interface has to be considered in order to estimate accurately the consumption and the generation based on the respiration activity. Also, the electrical signals of the β-cell-based FET seem to include the effect of ion or molecular charges in the insulin secretion process other than pH variation based on respiration activity. The electrical signals of the cell-based FET related to biochemical information, however, are a significant factor for the convenient and noninvasive evaluation of cell functions. In a previous work [45], the noninvasive monitoring of transporter function was shown using oocyte-based FETs; the size of an oocyte was larger than that of a somatic cell, and the amplification of electrical signals based on transporter function was expected in the case of using oocyte-based FETs. However, the electrical detection of β-cell activity obtained in this study has demonstrated the possibility of detection of somatic cell function using β-cell-based FETs.

6.6 In Vitro Cell Sensing with Semiconductor-Based Biosensing Technology

Lastly, I would like to summarize in vitro cell sensing with semiconductor, as follows. In our laboratory, we focus on a direct detection of ions or ionic molecules through ion channels at cell membrane, because most of cell functions are closely related to transferring of charged conductors from cell to cell. We have clarified that a principle of semiconductor devices based on field effect realizes it in a direct, label-free, real-time, and noninvasive manner for cell functional analysis.

The principle of semiconductor-based biosensing devices is based on the potentiometric detection of charge density changes induced at a gate insulator/solution interface accompanied by specific biomolecular recognition events. Ionic charges
of ions or biomolecules at the gate insulator electrostatically interact with electrons in silicon crystal across the thin gate insulator resulting in the threshold voltage change. Particularly, we are interested in ion transportations through membrane proteins such as ion channels and ion pumps at cell membrane and trying to detect ionic behaviors based on biological phenomena using a cell-coupled gate semiconductor. The semiconductor-based biosensing devices have good advantages of label-free, real-time, and noninvasive method, and we can make an arrayed device for a multitarget analysis by use of the conventional semiconductor processes.

In the point of detection of cell functions, we propose the device structure with three components such as target, signal transduction interface, and detection device. Since we utilize the cell-coupled gate semiconductor, we are trying to design the signal transduction interface in order to detect ion charges specifically and selectively based on each cell function.

For example, in order to detect drug effect on cancer cells, we need to detect ion charges based on programmed cell death “apoptosis” using the cell-coupled gate semiconductor and develop the signal transduction interface to trap them. The previous work showed the possibility of potassium ions, chloride ions, and water release in the early stage of apoptosis. Therefore, we have focused on potassium ion release based on apoptosis and succeeded in the real-time, direct, and noninvasive monitoring of their flow. In particular, this result was accomplished by use of crown ether monolayer to trap selectively potassium ion as signal transduction interface of the cell-coupled gate semiconductor. Moreover, we have found the possibility of multitarget detection for high-throughput screening of drug effect using the cell-coupled gate semiconductor with some transfected cancer cells in our study.

Using the cell-coupled gate semiconductor, furthermore, we have found the possibilities of a real-time and noninvasive monitoring of various cell functions, as follows:

- Interaction between substrate and transporter at cell membrane for drug effect detection
- Embryo activity based on in vitro fertilization (IVF) for assisted reproductive technology (ART)
- Glucose response of pancreatic β-cells for insulin secretion
- Differentiation of stem cells such as murine or human iPS cells
- Autophagy for accommodation to starvation
- Expression of sialic acid at cell membrane of cancer cells
- Cell motility at biomaterials
- Other cell function

Acknowledgements  The authors wish to thank Professor K. Kataoka of the University of Tokyo in Japan and Professor Y. Miyahara of the National Institute for Materials Science (NIMS) in Japan for their help and useful discussion. These works were partly supported by Center for NanoBio Integration (CNBI), Sentan-Keisoku, and CREST projects, Japan Science and Technology Agency (JST).
References

Chapter 7
Magneto-Nanosensor Diagnostic Chips

Richard S. Gaster, Drew A. Hall, and Shan X. Wang

Abstract  We have developed an automated assay for disease biomarker detection that can be run on a handheld sensing platform. By coupling magnetic nanotechnology with an array of magnetically responsive nanosensors, we demonstrate a rapid, multiplex immunoassay that eliminates the need for trained technicians to run molecular diagnostic tests. A major limitation for other detection modalities is signal distortion that occurs due to background heterogeneity in ionic strength, pH, temperature, and autofluorescence. Here, we present a magnetic nanosensor technology that is insensitive to background yet still capable of rapid, multiplex protein detection with resolution down to attomolar concentrations and extensive linear dynamic range. The insensitivity of our detector to various media enables our technology to be directly applied to a variety of settings such as molecular biology and clinical diagnostics.

R.S. Gaster
Department of Bioengineering, Medical Scientist Training Program, School of Medicine, Stanford, CA 94305, USA

D.A. Hall
Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
e-mail: drewhall@UCSD.EDU

S.X. Wang
Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA
Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA
e-mail: sxwang@stanford.edu

D. Issadore and R.M. Westervelt (eds.), Point-of-Care Diagnostics on a Chip, Biological and Medical Physics, Biomedical Engineering, DOI 10.1007/978-3-642-29268-2_7, © Springer-Verlag Berlin Heidelberg 2013
7.1 Introduction

Clinical diagnostics have tremendous potential to revolutionize the practice of medicine. The worldwide research community has made great strides toward developing faster, more sensitive, and more cost-effective diagnostic technologies. Despite extensive progress, however, worldwide access to many revolutionary technologies remains limited to large centralized laboratories in the developed world. Some of the most promising technological advances have combined engineering, material science, chemistry, and physics in the development of devices on the nanoscale, 1–100 nm. The size scale of these devices has been demonstrated to have unique physical and/or chemical properties that can be exploited for biological applications. In addition, the exceptionally small scale enables the detection platform to be highly portable, require relatively low power, and compile numerous sensors into one high-density array for multiplex detection. Accordingly, nanotechnology has been leveraged broadly in both diagnostic medicine and therapeutics.

The diagnosis of a disease requires the identification of a disease marker. The earlier that marker is detected, the earlier the disease can be treated. Typically these disease markers are quantified because ascertaining the concentration aids in determining the prognosis, ideal treatment, or the progression of the disease. Proteins and oligonucleotides, for example, serve as the most common biomolecular markers, termed “biomarkers.” Abnormalities in the structure, function, or amount of a biomarker present are what predispose a patient to disease or may be indicative of a particular disease. Because nanosensors are of the same scale as these naturally occurring biomarkers, with nanoscale detection, one can readily interface a highly sensitive sensor with the biological molecule of interest. The use of magnetic nanotechnology for detecting the disease biomarkers holds great promise because magnetically responsive nanosensors often allow for improvement in the lower limit of detection and have the additional advantage of portability of the detection apparatus and ease of use for point-of-care (POC) application. The magnetoresistive biosensors described in this chapter involve three distinct components: the magnetically responsive biosensor itself, the magnetic nanoparticle (MNP) tags, and the protein detection assay. Each aspect will be discussed in detail in this chapter.

7.2 Magnetic Biosensing Modalities

Magnetic biosensing offers several significant advantages over conventional optical techniques and other sensing modalities. The samples (blood, urine, serum, etc.) naturally lack any detectable magnetic content, providing a sensing platform with a very low background. Additionally, MNP tags are not subject to problems that have plagued fluorescent labels such as label bleaching and autofluorescence. Furthermore, the sensors can be arrayed and multiplexed to perform quantitative
protein and/or nucleic acid detection in a single assay without relying on optical scanning. Lastly, the sensors are compatible with standard silicon integrated circuit (IC) technology which allows them to be manufactured with a built-in electronic readout at low cost, in mass quantities, and to be deployed in a one-time use disposable format.

To our knowledge, the first use of MNPs as labels in immunoassays was reported in the literature in 1997 by a group of German researchers [1]. The measurement was achieved by use of a superconducting quantum interference device (SQUID) to detect binding events of magnetically labeled antibodies. While successful, the operating conditions required liquid helium cooling and magnetic shielding, limiting the practicality of the SQUID-based biosensors. In 1998, Baselt first demonstrated detection of MNPs using giant magnetoresistive (GMR) sensors with GMR multilayers [2]. GMR sensors have the advantage of room-temperature operation and simpler instrumentation, making them more attractive, particularly for portable applications.

All magnetoresistive sensors share a common principle of operation where the magnetization of a free magnetic layer (or layers) responds to a change in the local magnetic field and causes a change in the resistance of the sensor. A material that exhibits magnetoresistance transduces a change in an external magnetic field into a change in resistance. This effect was first discovered by Lord Kelvin in 1856 when he found that the resistance of an iron bar increased when the current flowing through the bar was in the same direction as the magnetic field. Furthermore, the resistance decreased when the magnetic field was perpendicular to the current. In actuality, most conductors exhibit magnetoresistance, albeit on an incredibly small scale not useful for transduction. This effect, known as anisotropic magnetoresistance, is commonly used in many sensors today, generally with more efficient materials such as Permalloy (Ni_{0.2}Fe_{0.8}). These sensors have typical magnetoresistance (MR) ratios on the order of 2% at room temperature [3] where the MR ratio is defined as

\[
MR = \frac{R_{\text{max}} - R_{\text{min}}}{R_{\text{min}}} = \frac{\Delta R}{R_{\text{min}}} \quad (7.1)
\]

Many years later came the discovery of GMR which is a quantum mechanical effect wherein a change in magnetic flux is transduced into a change in electrical resistance through spin-dependent scattering. GMR was first observed in a Fe/Cr/Fe thin film stack in 1988 independently by Albert Fert and Peter Grünberg, both of whom went on to win the 2007 Nobel Prize in Physics for their discovery. The most basic device exhibiting this behavior is the multilayer GMR stack where two or more ferromagnetic layers are separated by a thin non-ferromagnetic spacer. The thickness of this non-ferromagnetic spacer is typically only a few nanometers and is critical to the operation of the device. At certain thicknesses, the Ruderman-Kittel-Kasuya-Yosida (RKKY) coupling between the ferromagnetic layers becomes antiferromagnetic, thus causing the magnetization of the adjacent layers to align in an antiparallel state. An external magnetic field rotates the magnetization of the
upper layer causing it to align with the external field, minimizing the total energy of the system.

Qualitatively, the operation of the device can be understood by examining the two extreme cases. In the antiparallel state, as an electron in a spin-up state (designated with an arrow pointing to the right) passes through the film stack, it will scatter when it travels through each ferromagnetic layer (Fig. 7.1a). As it travels through the first ferromagnetic layer, the scattering is relatively small and leads to a low resistance since the spin of the electron is in the same direction as the majority spin of this layer. As the electron continues into the second ferromagnetic layer of the opposite magnetization, it will again scatter. This scattering event, however, is relatively large and leads to a higher resistance because the spin of the electron is in the same direction as the minority spin of this layer. The electron in a spin-down state (designated with an arrow pointing to the left), traveling through a GMR sensor in the antiparallel state, will have a similar resistance to the spin-up electron except in a reversed sequence, where the first layer it travels through is of high resistance and the second layer it travels through is of low resistance. In contrast, in the parallel state, the electron in the spin-up state passes through the first layer and the second layer with relatively few scattering events and thus has a low resistance in its entire path because the spin of the electron is always in the same direction as the majority spin of the layers. The electron in the spin-down state passes through both layers with relatively high resistance because the spin of the electron is always in the same direction as the minority spin of the layers. The overall resistance in each state can be understood using a circuit model where the resistance of the path taken depends on the spin polarization of the electron. For the antiparallel state, each path has a high resistance in series with a low resistance. In the other extreme where a large external magnetic field has caused the two layers to be in the parallel state, the spin-up electron will pass through the structure with minimal scattering. The spin-down electron will undergo significantly more scattering in both layers and thus have a higher resistance as seen in the equivalent circuit model (Fig. 7.1b). The parallel state has two paths: one with two low resistances in series and one with two high resistances in series. If the distribution of conducting electron spins is equal in spin-up and spin-down states, the circuit in the parallel state has an overall lower resistance than that in the antiparallel state.
One drawback of the basic multilayer GMR film stack is the high magnetic field needed to fully switch the magnetization from the parallel to the antiparallel state (up to 2,000 Oe), which tends to cause the sensors to have poor low-field sensitivity [4], limiting their utility as biosensors. GMR spin-valves (GMR SV) overcome this drawback by introducing a synthetic antiferromagnet to the film stack. This antiferromagnet formed by the CoFe/Ru/CoFe structure pins the magnetization with exchange coupling. The other ferromagnetic layer, called the free layer, rotates freely with the applied magnetic field. A pinning layer, typically PtMn or IrMn, defines the magnetization of the synthetic antiferromagnetic and is also used to direct the microstructural texture of the subsequent thin films. As can be seen in Fig. 7.2, GMR SV sensors are elaborately engineered film stacks, typically only a few tens of nanometers thick, passivated with an ultrathin oxide [5].

The transfer curve of a GMR SV sensor is shown in Fig. 7.3a. The sensor adopted in this work has a minimum resistance of \( 2.190 \Omega \) in the parallel state and a maximum resistance of \( 2.465 \Omega \) in the antiparallel state, corresponding to 12% MR ratio. The sensitivity of the sensor was calculated by differentiating the transfer curve (Fig. 7.3b). The sensor is most sensitive when no field is applied, tapering off as the field strength is increased. The saturation field is only 200 Oe, rendering GMR SV excellent low-field sensors.

The quest for higher MR ratios has led researchers to more exotic materials and even more elaborate film stacks. However, possibly the simplest way to increase the MR ratio is by changing the mode of operation. The original devices passed current vertically through the device as depicted in Fig. 7.1, referred to as the current-perpendicular-to-plane (CPP) mode of operation. The CPP mode of operation has a limited MR ratio because the electrons only pass through the ferromagnetic layers once and have limited opportunities to undergo spin-dependent scattering. In contrast is the current-in-plane (CIP) mode where the current flows parallel to the Cu layer and, given that the length of the sensor is significantly longer than the thickness, has many more opportunities to scatter. Figure 7.2b is an image of a device operated in CIP mode where each segment of the sensor is very long (90 μm). Typically GMR SV sensors often exhibit a 10–15% MR ratio at room temperature.
The lower-field sensitivity of GMR SV sensors made them very attractive for many sensor applications, such as the read head in hard disk drives, current sensing, earth field sensing, and biosensing. GMR SV sensors replaced the inductive read heads in hard disk drives in the late 1990s and were key in enabling the rapid increase in areal density and larger hard drives [6]. These GMR SV sensors have since been replaced by magnetic tunnel junction (MTJ) sensors, which exhibit even higher MR ratios. MTJs rely on an entirely different quantum mechanical effect known as tunneling magnetoresistance (TMR). Structurally the devices look very similar to a GMR SV, but the conductive Cu layer is replaced with an insulator. The orientation between the pinned layer and the free layer changes the probability that electrons can tunnel through the oxide, thus modulating the conductivity of the device. The MR ratio of these devices is often 100 % or more with the current record at 604 % at room temperature and 1,144 % at 5 K [7]. Presently, the main issue limiting the adoption of MTJs as biosensors is that with the large area of the devices, a single pin-hole defect renders the device unusable.

Despite the differences in the origin of the magnetoresistance, all magnetoresistive biosensors can be made to operate in a similar fashion. A magnetic immunoassay tethers MNP tags to the surface of the sensor. The underlying magnetically responsive biosensor detects the stray field from the MNP tags through a change in resistance. Since the stray field of the MNP tags falls off rapidly as the distance between the sensor and the tags increases, the magnetoresistive sensors can be referred to as proximity-based sensors.
7.3 Magnetic Nanoparticle Tags

Magnetic particles are commonly used in a variety of different applications such as cell sorting, magnetic resonance imaging, data storage, environmental remediation, and other applications. Here, we focus on using MNPs as tags for proteomic analysis in biomedicine. In this section, the MNP design requirements for applications to magnetically responsive nanosensors are discussed. The ideal design parameters, however, are often in conflict, thus requiring optimization choices to be made. For example, the highest signal per particle would originate from relatively large magnetic particles (on the order of 1 \( \mu \)m or larger). Larger particles, however, are not necessarily optimal because they tend to settle as they lack colloidal stability and they have significantly slower diffusion times. Therefore, in selecting the ideal MNP tag, competing factors must be considered and trade-offs must be made.

7.3.1 Superparamagnetism

An important design requirement for this technology is that the magnetic tags must not aggregate, chain, or precipitate during the course of any given experiment. Thus, the particles used in magnetically responsive and proximity-based detection systems ideally should be superparamagnetic, where the volume of the ferromagnetic core is so small that thermal energy alone is large enough to cause the magnetic moment of the cores to fluctuate rapidly. The average magnetic moment over time of any given superparamagnetic core is therefore zero, resulting in zero remnant moment. However, when an external magnetic field is applied, the nanoparticles magnetize with a much greater magnetic susceptibility than paramagnetic materials.

More specifically, a superparamagnetic material is a magnetic material of such small size that at temperatures below the blocking temperature, it behaves like a paramagnetic material. As the size of these superparamagnetic particles increase, they lose their superparamagnetic nature and become ferromagnetic. This limit is known as the “superparamagnetic radius.” This superparamagnetic radius can be calculated by the following equation:

\[
P = v_0 e^{-\frac{\Delta f V}{k_B T}}
\]

where \( P \) is the probability per unit time that the magnetization will change direction, \( v_0 \) is the attempt frequency (\( \sim 10^9 \) s\(^{-1} \)), \( \Delta f V \) is the free-energy barrier that the particle must overcome in order for the moment of the particle to switch directions, \( k_B \) is Boltzmann’s constant, and \( T \) is the temperature. Therefore, if iron oxide nanoparticles, for example, are to be utilized in our assay, it is favorable to use them at sizes smaller than the critical size (so they remain in the superparamagnetic regime). The challenge in using such small MNPs, however, is that as the size of the
Fig. 7.4 Schematic representation of a magnetic nanoparticle (MNP) labeled with antibody drawn to scale. The magnetic tag is comprised of a dozen iron oxide cores embedded in a dextran polymer and then functionalized with antibody or receptor [8].

nanoparticle decreases, the magnetic moment decreases as well, causing the signal per particle to decrease.

The solution to this problem is to cluster many MNPs into a dextran polymer (Fig. 7.4). In this way, the MNPs are physically isolated, allowing them to remain superparamagnetic, while the overall magnetic content per magnetic tag is relatively large due to the multiplicity of cores in each cluster.

### 7.3.2 Application of Magnetic Nanoparticles for Magnetic Biosensing

Recent work has adapted magnetically responsive sensors for the detection of biological species in solution by implementing a traditional sandwich assay directly on these magnetically responsive nanosensors and utilizing superparamagnetic nanoparticles as tags. If a magnetic particle similar to the one described above is introduced to label the biomolecule of interest, magnetically responsive sensors are capable of highly sensitive protein and oligonucleotide detection [2, 9–12].

Among the more commonly used MNP tags in GMR biosensors are those comprised of clusters of monodisperse Fe$_2$O$_3$, superparamagnetic particles each with a 10 nm diameter embedded in a dextran polymer and functionalized with streptavidin (Fig. 7.4), as determined by TEM analysis [13]. The entire nanoparticle averages 46 ± 13 nm in diameter (measured by number-weighted dynamic light scattering). Based on the Stokes-Einstein relation, these particles have a translational diffusion coefficient of approximately $8.56 \times 10^{-12}$ m$^2$s$^{-1}$. The MNPs have a reported zeta potential of $-11$ mV [14]. These particles are superparamagnetic and colloidally stable, so they do not aggregate or precipitate during the reaction. Therefore, a major advantage of using these tags is that the magnetically responsive sensors detect the exact same signal before and after washing (Fig. 7.5). This means that it is equally valid to read the sensor signal prior to a final wash that removes unreacted
MNPs. Importantly, the magnetically responsive sensors operate as proximity-based detectors of the dipole fields from the magnetic tags. Therefore, unbound MNP tags contribute negligible signal in the absence of binding, rendering this unique nanosensor-MNP system ideal for real-time kinetic analysis [8].

It is apparent from plotting the magnetic moment per particle versus the applied external magnetic field that there is very minimal coercivity in these particles confirming their superparamagnetic nature (Fig. 7.6). In our experience, these MNPs have performed the best in terms of the kinetics of binding to detection antibody, minimal nonspecific binding, and high reproducibility.

In summary, choosing the magnetic label is a vital aspect to any magnetic biosensor where important design trade-offs must be made. On the one hand, larger, micron-sized magnetic particles are desirable as they will generate a high signal per particle. Larger particles, however, are kinetically unfavorable since detection requires diffusion of the magnetic tags to the surface-immobilized detection antibody. Furthermore, larger particles are undesirable because nonspecific binding events of micron-sized particles will have a much greater effect on the overall signal than will a nonspecific binding event of a nanoparticle. Accordingly, the optimal balance is to use MNPs that are comprised of a cluster of small superparamagnetic particles imbedded in a dextran polymer. This configuration will increase the magnetic content of each particle while remaining superparamagnetic.
7.4 Protein Detection Assay Using Magnetic Nanoparticles

7.4.1 Standard Protein Detection Assay

Biomarkers can be detected either by their accumulation at a stationary binding site over a sensor surface or by the binding of biomarkers to tags in a solution inducing aggregation of the magnetic tags. For the GMR nanosensors discussed in this chapter, the former will be described. For NMR-based biosensors discussed in Chaps. 9 and 10, the latter method of protein detection will be described. One of the most effective and specific methods of detecting proteins on a sensor surface, like a GMR nanosensor, is by means of a “sandwich assay.” Typically known for its use in the enzyme-linked immunosorbent (ELISA), the sandwich assay involves the formation of a three-layered structure where two antibodies (or aptamers, diabodies, Fab fragments, etc.) form a sandwich around a protein (also called the “analyte”) of interest (Fig. 7.7a). One of the antibodies, generally referred to as the “capture antibody,” is directly immobilized on the sensor surface. In order to make the sandwich assay highly specific, a monoclonal capture antibody is traditionally used. A solution of monoclonal antibodies means that every antibody in the solution has the exact same Fab region and therefore will bind to only one epitope on one protein. The capture antibody makes up the foundation of the sandwich assay and serves to selectively immobilize a specific protein of interest directly over the sensor surface.

The second antibody, known as the detection antibody, is delivered in solution and binds to a second epitope on the captured protein of interest. The detection antibody is typically polyclonal and pre-modified with a reactive chemistry, enabling facile attachment of the detection antibody to the tag of interest. A polyclonal antibody solution is one in which all the antibodies react with the same protein; however, they may bind to different epitopes on that protein with varying affinities. Therefore, the Fab region is not uniform across all the antibodies in a polyclonal solution but recognizes different regions of the same protein. Typically the detection
antibody is modified with biotin, and the tag is modified with streptavidin since the biotin-streptavidin interaction is one of the strongest non-covalent receptor-ligand interactions in biochemistry (association constant, $K_a \approx 10^{14} \text{–} 10^{15} \text{M}^{-1}$). In the ELISA, the tag of interest is typically colorimetric or fluorescent. However, when using magnetically responsive biosensors, the tag of interest is magnetic. Therefore, the more protein that is present in the system, the more detection antibodies bind and the more magnetic tags bind. As the number of magnetic tags increases over the sensor, the MR in the underlying magnetically responsive sensor changes proportionally, producing larger signals. In this way, quantitative protein detection is possible with this assay. Similar to a protein sandwich assay, it is also possible to form a DNA or RNA sandwich structure by utilizing a capture oligonucleotide sequence and biotinylated detection oligonucleotide sequence for highly sensitive nucleic acid detection (Fig. 7.7b).

A “reverse-phase” assay can be used to detect proteins of interest in many patients’ blood samples simultaneously by reorganizing the traditional sandwich assay (Fig. 7.7c). In the reverse-phase assay, instead of functionalizing a capture antibody onto the sensor surface, patient samples containing proteins of interest such as cell lysates are immobilized directly onto the GMR sensor array. Then, a solution containing detection antibodies complementary to the protein of interest is introduced and will bind to the immobilized protein of interest over the GMR sensor. Since the detection antibody is biotinylated, it can then bind to magnetic nanotags coated with streptavidin in the same way as the detection antibody and magnetic nanotag interaction in the traditional sandwich assay. In addition, if one separates
the protein of interest from the sample prior to sensor immobilization using protein purification techniques, it is possible to increase capture protein density, thereby maximizing signal.

### 7.4.2 Wash-Free Protein Detection Assay for POC

With magnetic biosensors, the traditional sandwich assay described above can be redesigned to leverage the proximity-based detection capabilities and unique magnetic properties of a magnetically responsive biosensor system. The GMR nanosensors, for example, can be built with an ultrathin passivation layer. As discussed above, these GMR sensors are proximity-based sensors, and therefore, only magnetic nanotags within ~150 nm of the surface are detected [5]. Because the magnetic nanotags typically employed are on the order of 50 nm in diameter and colloidally stable, they do not settle or precipitate on the sensor surface, contributing negligible signal in the absence of the target protein or detection antibody. Only in the presence of both the biomolecule of interest and detection antibody will the magnetic nanotags congregate over the appropriate sensor in close enough proximity and in high enough density for the GMR sensor to experience a measurable magnetoresistance change. This is a significant advantage over the vast majority of protein detection platforms in which the excess/unreacted tags must be washed away prior to detection, preventing their ability to utilize a wash-free detection method. As a result, while performing a traditional sandwich assay requires washing steps to remove excess antibodies or nanotags, with magnetic nanotechnology, the signal remains unchanged with a final wash step. Accordingly, the protein content can be determined in the assay without implementing washing (Fig. 7.5). Thus, because this assay obviates the need for washing steps, it offers a faster, simpler testing process that untrained users can easily perform in point-of-care settings.

By taking advantage of the "autoassembly" nature of this assay, only minimal human intervention is required to run a test, removing the dependence on the end user to have prior laboratory training. Moreover, the assay can be run in an open-well format, removing the need for complex microfluidic plumbing or external pneumatic pressure controllers. The wash-free assay entails only three steps: (1) The operator places the biological sample into the reaction well which is equipped with an array of GMR sensors pre-functionalized with a panel of antibodies against predetermined proteins of interest. As the sample incubates in the well, the proteins of interest are captured by the immobilized antibodies directly over individually addressable GMR sensors. (2) The user adds a solution of magnetic nanotags labeled with streptavidin. At this point, no detectable reaction takes place because no biotin is present in the reaction well (Fig. 7.8a). (3) Finally, detection antibodies labeled with biotin are introduced. These detection antibodies subsequently link the magnetic nanotags to the captured analyte, thus inducing a measurable signal in the underlying GMR
Fig. 7.8 Schematic representation of the autoassembly immunoassay where each square represents a unique GMR nanosensor in the array. (a) After immobilizing unique capture antibodies over individually addressable sensors, and incubating with the protein of interest, the magnetic nanotags are added in solution above the sensor. Since there is no chemistry to link the magnetic nanotags to the captured antigen, no signal is detected by the underlying sensor. (b) Once the detection antibody in solution is added, the detection antibody which is labeled with biotin is capable of linking the streptavidin-labeled magnetic nanotag to the captured analyte. In the presence of captured analyte, the magnetic nanotags will congregate over the corresponding GMR sensors in high enough concentration to be detected. Insert: optical microscopy of a section of the array of nanosensors. Each square in the array is one sensor and each circle is a nanoliter droplet of capture antibody uniquely functionalized over the sensor surface [15].

Among the advantages of magnetic nanosensors is that they can be fabricated into high-density arrays with minimal increase in cost or size of the overall chip. This allows for highly multiplex protein detection in a single reaction well. While there are significant advantages to open-well protein detection systems, there are several limitations as well. An open-well format is limited to running only one sample per chip. For high-throughput analysis with fewer than 5 biomarkers of interest
Fig. 7.9  Real-time monitoring of sensors during autoassembly immunoassay. Addition of the sample and magnetic nanotags contribute negligible signal (upon addition of the magnetic nanotags, there is a very small signal rise due to detection of magnetic nanotags in solution above the sensor). However, once the detection antibody is introduced, the magnetic nanotags are clearly measurable on sensors functionalized with the appropriate capture antibody and antigen. The negative control sensors, coated with anti-insulin antibody, remain flat, indicating negligible nonspecific binding. The y-axis units are the change in magnetoresistance normalized to the initial magnetoresistance presented in parts per million (ppm).

per sample, it is wasteful and underutilized to have on the order of 100 sensors per nanosensor array investigate only five biomarkers. It would be more efficient if the high-density sensor array could be subdivided where several patient samples could be run simultaneously on a single chip using parallel microfluidic channels. In addition, with cross-reactive antibodies, reagents can be separated into their own reaction chambers when implementing microfluidic integration in order to minimize this phenomenon. Further, the use of microfluidics can be optimal for handling biological samples when only very small sample volumes are available. Fortunately, microfluidic chip integration is highly compatible with magnetic nanosensor arrays (Fig. 7.10). In this very basic microfluidic chip, each microfluidic channel contains eight sensors for up to 8-plex protein detection on any given sample in any given channel. This will yield eightfold more tests per hour and amortize the chip cost over multiple samples.

The microfluidic chips are fabricated using standard soft lithography techniques. Polydimethylsiloxane (PDMS) is cast onto an SU8-based mold. The PDMS is then cured and peeled from the mold. The thickness of the mold is used to form the fluidic channels. External connections are then punched into the inlets and outlets of the PDMS blocks. The final microfluidic chip design is comprised of 200 – μm-wide channels that are each 20 μm high and the channel pitch is 400 μm. For more details on microfluidic biosensors, please refer to Chap. 2.
7.5 Miniaturization of Desktop Biostation

In order to facilitate effective deployment in the field by nontechnical users, it is important that the wash-free assay be integrated into an ultraportable and battery-powered detection module [16]. This capability should obviate the need for a constant supply of electricity or a designated laboratory. Since the form factor of GMR nanosensors is very small, dictated only by lithography used in their fabrication, by miniaturizing the electronic components, it is possible to replace a laboratory full of equipment with a handheld and battery-powered device (Fig. 7.11). No lasers or expensive charge-coupled device (CCD) cameras are required for the platform, which uniquely positions GMR-based biosensors for ultraportable, POC applications.

The detection platform has been designed to have two components: a reusable handheld detection module the size of a handheld calculator (Fig. 7.12a) and a disposable detection stick (Fig. 7.12b). The handheld detection module consists of two boards. First, the data acquisition board (DAQ) has both analog and digital subcircuits. The analog circuits are comprised of the excitation signal generation, the field signal generation, and the front end. The front end for the sensor is a classical Wheatstone bridge with a high-gain instrument amplifier. Also contained on the DAQ board is the microprocessor which does the digital signal processing (DSP) and handles all of the user interactions. The other board in the detection module is the coil board which contains a power amplifier and a planar electromagnet used to generate the magnetic field to modulate the sensors. The disposable stick, which is the second component of the overall detection platform, contains no electronics, just the GMR nanosensor array with 8 individually addressable sensors.
Fig. 7.11  Miniaturization of the original biostation into a handheld, battery-powered device. Demonstration of miniaturization from (a) the initial laboratory test station which occupied an entire room into (b) handheld point-of-care diagnostic device. All components of the laboratory setup have been miniaturized and incorporated into an ultraportable platform [17]

Fig. 7.12  (a) Photograph of the disposable stick and reaction well in which the assay is run. Insert: Inside the reaction well is an array of GMR sensors capable of simultaneously monitoring multiple different proteins in a 20–50 μL sample. (b) Image of the handheld device with case and test stick

for multiplex detection, mounted on a printed circuit board (PCB). The disposable stick also has an open well surrounding the GMR nanosensor array where the wash-free protein detection assay is run. The disposable stick can be pre-functionalized with capture antibodies to detect biomarkers for cardiovascular disease, cancer, influenza, HIV, and a variety of other chronic and infectious diseases.

One of the largest and most difficult elements to miniaturize in the research grade biostation was the Helmholtz electromagnet. This large component alone weighs over 100 kg and when coupled with the associated power amplifier, consumes over a hundred watts of power drawn from a wall outlet. In the typical research setting (Fig. 7.11a), the size and cost of a magnetic test station are not critical factors because the main goal is often to maximize the sensitivity, linear dynamic range, and throughput.
In contrast, reducing the form factor and power consumption to create a handheld, ultraportable device is essential for POC application, but posed several engineering challenges. To accomplish this miniaturization, a planar electromagnet was designed using 1.27-mm (50 mil) traces on a four-layer PCB (Fig. 7.13a). The orientation of the current flowing through the coil alternates between clockwise and counterclockwise to avoid the need for any crossover traces that would reduce the number of available routing layers. The magnetic field is generated out of the plane (perpendicular to the PCB) and reoriented by soft magnetic flux guides manufactured out of cold rolled steel. The flux guides concentrate the field over a smaller region, acting as a form of passive amplification, and are used as heat sinks for the electromagnet. In addition to the flux guides above the coil, there are flux guides below the coil to close the flux loop and increase the efficiency. Due to the off-axis nature and the use of magnetic flux guides, analytical models are not tractable for design. Instead, finite element modeling (FEM) is needed to determine the required number of turns (11 turns per layer) and current, which includes both field strength and frequency. The miniature electromagnet is driven by a custom-designed class-A power amplifier. Figure 7.13b illustrates the relationship between the current through the electromagnet and the measured field across the GMR SV. Power consumption was minimized by cycling the power amplifier and electromagnet when they are not being used.
Currently, proteomics in clinical (not POC) settings utilize fluorescent detection based on the ELISA, which report detection limits on the order of 1 pM with 2 orders of linear dynamic range. The wash-free assay presented here has a similar dynamic range but achieves over an order of magnitude higher sensitivity in a fraction of the time. The higher sensitivity is primarily attributable to using MNP tags rather than fluorescent labeling. With the magnetic nanotechnology described in this chapter, detection down to 50 fM in a 25 µL sample has been demonstrated [17]. While the sensitivity needs for POC settings is generally less stringent, having higher sensitivity allows for a shorter assay time, leading to faster diagnostic times. The higher sensitivity of this technology may also facilitate the earlier diagnosis of disease.

The MNP tags require an external magnetic field to induce a magnetic moment, and the sensors require the magnetic field to modulate the sensor response to a higher frequency. The optimal magnetic field for this particular combination of sensor and MNP has been shown previously to be 25 Oe [18]. Because the optimum is fairly shallow, this allows the field to be reduced without a significant loss in sensitivity. At a magnetic field of 15 Oe (60% of the optimum), the signal per MNP decreases by only 20%. With this small reduction in sensitivity, the power consumption can be significantly reduced.

GMR spin-valves typically exhibit high flicker noise (also known as 1/F noise because it is inversely proportional to frequency). To increase the signal-to-noise ratio (SNR) and improve the detection capability of the device, the signal from the MNP tags is modulated away from the low-frequency noise to a higher frequency [19]. To recover this signal, the microprocessor digitizes the response from the GMR nanosensors and performs the filtering and demodulation. Figure 7.13c, d illustrates this process with the incoming modulated signal and the clean output signal after a 113th order digital filter has been applied. A minimalistic version of the computationally intensive signal processing algorithms used in our desktop station was implemented due to the limited computational power of the microprocessor [20]. With the integration of a power source, signal processing, and display functionality into the handheld detection module, no additional components are required to run and measure an assay, allowing it to truly be a POC testing device.

A fundamental element of the handheld device is a microchip microprocessor (dsPIC30F6012a) which runs at 80 MHz (20 MIPS). The microprocessor has an integrated 12-bit analog to digital converter used to digitize the signals from the sensors. Furthermore, the microprocessor communicates to the direct digital synthesizer chips via an integrated SPI bus. However, the primary reason for choosing a high-end microprocessor is for the heavy DSP algorithms that it performs. To extract the single tone from the spectrum with the double modulation scheme, the 113 tap digital FIR bandpass filter is applied to the incoming samples. The tap count was chosen after all of the code had been written such that it filled the remaining memory of the microprocessor to minimize the noise bandwidth of the extracted tone. The root mean square value of the filter output is proportional to the magnetoresistance of the sensor and is saved to an internal buffer. The sensors are scanned in a round robin fashion, rotating from sensor 1 through sensor 8.
sensor, the signal is acquired for 200 ms and the subsequent analysis takes 800 ms. While the signal processing could be overlapped with the data acquisition or applied in real-time, the tasks are undertaken sequentially to simplify the timing between the acquisition and the processing steps. This signal acquisition and processing is repeated throughout the duration of the test.

In POC settings, it is not practical to perform sample preparation prior to running the diagnostic test. Accordingly, the platform must have reproducible detection despite differences in the sample fluid (buccal swab, serum, urine, cell lysates, etc.), pH, and temperature. Fortunately, GMR spin-valve sensors have been reported to be insensitive to different sample matrices, rendering the platform highly generalizable to a diversity of biologically relevant samples and removing the need for any complex sample preparation [21]. This subtle requirement is often overlooked or ignored when discussing POC diagnostics, but in fact is critical to the utility of such a diagnostic device in real-world settings.

7.6 POC Detection Results via GMR Biosensor Arrays

The user interface of the detection module has been designed to provide both a rapid readout and a user-friendly, easy-to-comprehend display. The microprocessor monitors the real-time binding events and predicts the saturation signal based on the initial binding trajectory. Monitoring the binding trajectory in real-time significantly reduces the assay time and produces a more reliable final readout than taking a single-point measurement at an arbitrary time prior to signal saturation. Figure 7.14a shows the binding curves of various concentrations of human immunodeficiency virus (HIV) p24 protein ranging from 100 ng/mL down to 32 pg/mL. We used these binding trajectories to train the microprocessor for future experiments. The assay runs for 15 min to allow sufficient time for differentiable signals to emerge while still providing rapid results for POC utility.

Each disposable stick, which is inserted into the detection module, is equipped with eight sensors allowing for up to 8-plex protein detection simultaneously in a single assay and permitting entire panels of markers to be monitored in real-time. The signals detected by each sensor on the stick can be displayed to the user via colored light-emitting diodes (LEDs) on the detection module. The microprocessor is preprogrammed with tables that contain calibration curves for each target protein as well as for the corresponding concentration thresholds (undetectable, low, medium, and high) which are predetermined by physicians according to clinically relevant therapy regimens (Fig. 7.14b). As the assay runs, the colored LEDs dynamically change and thereby present the results in real-time to the end user. The display of the device can alternatively be equipped with a quantitative digital readout, but the LED color reporting system shown here suffices to indicate relative levels of protein content for untrained users. After a 15 min incubation period, the predicted signal at saturation is compared with threshold values, and the microprocessor selects the appropriate indicative color for each LED. For example, when a 10 ng/mL of p24 capsid protein was tested on the handheld device, a signal of 39 ppm was measured
Fig. 7.14 Detection characteristics and readout of the handheld magnetic biosensor. (a) Real-time binding curves of diluted HIV p24 protein at concentrations ranging from 100 ng/mL to 32 pg/mL. The sensors, functionalized with bovine serum albumin (BSA) as a negative control, gave minimal signal indicating negligible nonspecific binding of the wash-free assay. (b) Calibration curves for each marker of interest were generated after 15 min of incubation time. The final curve is subsequently divided into four predetermined concentration ranges and will be presented via color coded LEDs to the end user. If the signal is undetectable, the indicator light will not be lit. If the signal is low, medium, or high, then the light indicator will display green, orange, or red light, respectively. (c) To demonstrate the specificity and readout of the device, we functionalized each of the eight sensors with a different capture antibody. For example, sensor S3 was functionalized with anti-p24 antibody. When 10 ng/mL of p24 antigen was spiked into the reaction well, only sensor S3 lit up in the medium concentration regime [17].

and the LED for sensor 3 (on the device labeled S3) turned orange, indicating a moderate level of protein content (Fig. 7.14b and c). As appropriate, all the other sensors, functionalized with noncomplementary antibodies, registered no signal. Similar experiments have been demonstrated with detection of hepatitis C virus (HCV) capsid protein [22], presented in Fig. 7.15. The combination of a rapid, wash-free assay and user-friendly display system can help facilitate the rapid adoption of this platform in both urban centers as well as remote field settings.

Another important consideration for POC applications is that the platform should be cost-effective. The total cost of each disposable stick (including the antibodies, magnetic tags, sensors, and assembly) in high volume is less than $3.50 (Table 7.1), which means that this diagnostic tool is sufficiently cost-effective to be used in both developing and developed nations. Furthermore, we believe the cost could be substantially reduced to less than $1 with slight changes to the sensor array and by preparing the MNPs in-house.
Fig. 7.15 (a) Real-time binding curves of diluted HCV capsid protein at concentrations ranging from 1,000 ng/mL to 100 pg/mL. The sensors, functionalized with bovine serum albumin (BSA) as a negative control, gave minimal signal indicating negligible non-specific binding of the wash-free assay. (b) Calibration curves for each marker of interest were generated after 15 min of incubation time on a log-log scale [17]

Table 7.1 The itemized cost of the one-time use disposable stick, including the circuit board, sensors, capture and detection antibodies for eight sensors, magnetic nanoparticles, surface chemistry, and assembly, is shown below. The volume of production is calculated for one million units per year. It is further assumed that the antibodies are robotically spotted to reduce the required volume and assembly cost

<table>
<thead>
<tr>
<th>Item</th>
<th>Unit cost ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circuit board and connector</td>
<td>0.34</td>
</tr>
<tr>
<td>GMR sensor die</td>
<td>1.00</td>
</tr>
<tr>
<td>HCV capture antibody (abcam 2583)</td>
<td>0.03</td>
</tr>
<tr>
<td>HCV detection antibody (abcam 58713)</td>
<td>0.03</td>
</tr>
<tr>
<td>Magnetic nanoparticles (Miltenyi Biotec 130-048-101)</td>
<td>1.88</td>
</tr>
<tr>
<td>Surface chemistry reagents</td>
<td>0.01</td>
</tr>
<tr>
<td>Assembly</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3.49</strong></td>
</tr>
</tbody>
</table>

7.7 Conclusions

The handheld device described in this chapter has the potential to provide a significant contribution to the future of POC medical diagnostics. As the population continues to expand and societal mandates for universal healthcare grow, innovations in diagnostic testing will be required to provide timely, easily accessible, and inexpensive results. To meet this need, it is necessary to develop cost-effective, portable, and easy-to-use devices which allow individuals to conduct their own molecular diagnostic tests without the need for a centralized laboratory, laboratory technicians, clinic or emergency room visits, or in some instances visits to a physician’s office.
While significant strides have been made toward developing a true POC testing device using magnetic nanotechnology [23–25], systems designed prior to this work have relied on an external power source and an external PDA (either a pocket PC or a laptop) for signal processing, data logging, and display. In the work presented in this chapter, a fully integrated and cost-effective unit incorporating a built-in microprocessor and miniature electromagnet has been designed to perform all of these tasks. The handheld biomarker detection platform utilizing magnetically responsive biosensors and MNP tags has tremendous potential for POC diagnostics and personalized medicine. Throughout the development of this technology, a conscious effort was made to create a platform that is both cost-effective and power efficient for portable applications. The handheld detection module consumes an average of 3.7 W from a rechargeable battery and weighs only 0.34 kg (0.75 lbs). Moreover, the sensitivity of this handheld device and its multiplex capability are noteworthy. The protein detection limit of 50 fM and 8-plex protein detection achieved by this device are on par with or exceed many of the current desktop protein detection platforms.

With the presented magnetic nanotechnology, patients can receive accurate molecular-based diagnosis on their own in a matter of minutes. Furthermore, due to the versatility of the sensing platform, the potential applications are vast, particularly in the realm of infectious diseases. By providing disposable sticks pre-functionalized with different capture antibodies, this technology can be deployed for detection of a range of infectious diseases that pose large-scale public health risks, such as HIV, HCV, tuberculosis, *Salmonella typhi*, and toxigenic *E. coli*, as well as swine (H1N1) flu and avian (H5N1) flu. In addition, screening for enteric infections is of particular interest, as the device will enable public health officials to inspect and immediately detect contamination on-site, to aid in safeguarding food and water sources for populations worldwide. This technology has the potential to reshape the practice of medicine by providing societies in both the developed and developing world with a new medical infrastructure: one that gives individuals the tools to literally take healthcare into their own hands.

**Acknowledgements** This work was supported, in part, by the United States National Cancer Institute (grants 1U54CA119367, 1U54CA143907, 1U54CA151459, 1U01CA152737, 5R33CA138330, and N44CM-2009–00011), the United States National Science Foundation (grant ECCS-0801385–000), the United States Defense Advanced Research Projects Agency/Navy (grant N00014–02–1–0807), a Gates Foundation Grand Challenge Exploration Award, and the National Semiconductor Corporation. RSG acknowledges financial support from the Stanford Medical School Medical Scientist Training Program and the National Science Foundation Graduate Research Fellowship Program.

**References**


Chapter 8  
Handheld NMR Systems and Their Applications for Biomolecular Sensing

Nan Sun and Donhee Ham

Abstract We have developed a miniature nuclear magnetic resonance (NMR) system. By combining the physics of NMR with CMOS radio-frequency ICs, we developed a 0.1-kg palm NMR system that is 1,200 times smaller, 1,200 times lighter, and yet 150 times more spin-mass sensitive than a state-of-the-art 120-kg commercial benchtop system. The small NMR system can be used for disease detection and medical diagnostics. It was demonstrated capable of detecting human cancer cells and cancer marker proteins.

8.1 Introduction

Nuclear magnetic resonance (NMR) is the energy exchange between a radio-frequency (RF)-varying magnetic field and an atomic nucleus such as a hydrogen proton, which acts like a tiny bar magnet due to its spin. Since the detailed resonance behavior is influenced by the environment of the nucleic magnets, NMR can be used to examine properties of a material, and thus, it has a wide array of applications in technology and science, such as biomolecular sensing, medical imaging, and oil detection, just to name a few.

The benefits of NMR would be broadly available, if NMR instruments can be made small, thus, at low cost. For example, a miniature NMR biosensor may enable disease screening in a doctor’s office or a patient’s home at an affordable...
cost. Nonetheless, NMR systems remain bulky, heavy, and expensive, with their use limited in hospitals, testing facilities, and laboratories. A case in point is the state-of-the-art commercial benchtop NMR system of [1], which weighs approximately 120 kg.

The large size is due to the following reason. An NMR system consists of a magnet to produce a static magnetic field, a sample coil, and an RF transceiver to generate an RF magnetic field and to monitor the resonance. Since a larger-sized magnet tends to yield a stronger NMR signal even for the same static magnetic field strength and hence relaxes the sensitivity requirement on the transceiver design, large magnets are used, leading to the bulky size, where the magnet is by far the largest component.

To miniaturize an NMR system, we took an approach opposite to the convention: we chose to use small magnets, and to detect the NMR signal substantially weakened by the small magnets, we developed highly sensitive RF transceivers. Furthermore, we integrated the RF transceiver onto silicon integrated circuit (IC) chips. As small magnets are used, the transceiver integration makes sense: in a conventional system where a large and expensive magnet dominates the system size and cost, integration of the RF transceiver would hardly reduce either the system size or the cost.

Our efforts first led to the construction of a portable NMR system shown in Fig. 8.1 [2]. Occupying only 2.5 L and weighing only 2 kg, this system is 60 times lighter, 40 times smaller, yet 60 times more spin-mass sensitive than the state-of-the-art benchtop NMR system of [1]. It uses a small magnet, the size of a hamburger.
(this magnet and the magnet of the commercial system both produce the static field of approximately 0.5 T). To receive the NMR signal weakened by the small magnet, we developed a partially integrated, high-performance RF transceiver and a separate high-quality planar coil.

After the design of the 2-kg portable system, we developed two new miniature NMR systems, which we call the palm NMR system and the 1-chip NMR system [3]. They represent yet another orders-of-magnitude size reduction and lab-on-a-chip capability. The palm NMR system, shown in Fig. 8.2, is the smallest complete NMR system to our best knowledge. It is 1,200 times lighter, 1,200 times smaller, yet 150 times more spin-mass sensitive than the 120-kg commercial system of [1]. As compared to the 2-kg portable NMR system, the palm system is 20 times lighter, 30 times smaller, and yet 2.5 times more spin-mass sensitive. To attain this further substantial size and cost reduction, we use a tiny magnet only the size of a ping-pong ball. This considerably lowers the NMR signal, which we overcome by designing a new, high-performance RF transceiver. As the signal is already lowered by the ping-pong-ball-sized magnet, the palm system uses a high-quality solenoidal coil, not to further weaken the signal.

The 1-chip NMR system is shown in Fig. 8.3. In this system, even the NMR coil is integrated as a planar spiral in the silicon IC chip along with the transceiver developed for the palm system. The transceiver’s performance allows the use of the lossy on-chip coil that lowers the signal-to-noise ratio. Not to further weaken the signal-to-noise ratio, the 1-chip system operates with the same hamburger-sized magnet of the 2-kg portable system. Due to this magnet, the weight reduction from the portable system is by 25%, but the point of the 1-chip system is lab-on-a-chip.
operation. For example, a biological sample can be placed directly on the coil of the chip for on-chip screening of disease markers. The chip can be disposable for one-time diagnostic testing. The direct interface may also enable oil detection [4] and quantum computing [5] on a silicon chip. The 1-chip system has the same spin-mass sensitivity as the 2-kg portable system while 60 times more spin-mass sensitive than the commercial system of [1].

The key to these two new developments is the new RF transceiver, an advance from the transceiver in the 2-kg portable system. First, the new transceiver achieves the sensitivity to cope with the signal-to-noise ratio lowered by the ping-pong-ball-sized magnet (palm system) or the lossy on-chip coil (1-chip system). Second, the new transceiver attains the highest level of integration among existing NMR transceivers. The transceiver in the 2-kg portable system did not integrate a power amplifier (PA), as meeting the large power tuning requirement of NMR was nontrivial with an integrated PA. We integrate a PA in the new transceiver by devising a power tuning scheme that exploits the natural high-$Q$ ($\sim 10^4$) filtering ability of atomic nuclei.

We used our three miniature NMR systems to detect various kinds of biomolecules relevant to disease screening. For example, we detected avidin, folic acid, and human chorionic gonadotropin (hCG), which can be used as a cancer marker for male patients for cancers such as choriocarcinoma, germ cell tumor, and islet cell tumor. At the cell level, we detected human bladder cancer cells.
This chapter explains the construction and experiment of our miniature NMR systems. Its organization is as follows. Sections 8.2 and 8.3 present the design and measurement of the silicon RF transceivers. Section 8.4 reports NMR experiments and NMR-based biomolecular sensing. Section 8.5 compares our miniature NMR systems to other NMR miniaturization efforts. We recommend readers unfamiliar with NMR to read our publication of [2] for quick introduction to NMR basics relevant to our work.

8.2 NMR RF Transceiver IC Design

8.2.1 Overall Architecture and Operation

We focus our discussion on the transceivers in the palm and 1-chip systems (Figs. 8.2 and 8.3), for they are more advanced than the transceiver in the 2-kg portable system (Fig. 8.1). Figure 8.4a and b shows the architectures of the NMR RF transceivers in the palm and 1-chip systems, respectively. The palm system (Fig. 8.4a) uses an off-chip solenoidal coil; the 1-chip system (Fig. 8.4b) employs an on-chip planar spiral coil. The dashed lines in the figures indicate silicon integration boundaries for the two systems. The transceiver architecture is essentially the same between the two systems, but the transceiver-coil matching networks are different for a reason explained in Sect. 8.2.4; thus, the two separate figures were prepared to avoid confusion. The electrical characteristics of the coils will be described in Sect. 8.2.4. All NMR experiments in our work, including biomolecular sensing, are done with protons in hydrogen atoms in aqueous samples. In the palm system, a sample is placed inside the solenoidal coil and is subjected to a static magnetic field \( B_0 \) of 0.56 T produced by the ping-pong-ball-sized magnet. In the 1-chip system, a sample placed on the planar coil is subjected to a static magnetic field \( B_0 \) of 0.49 T produced by the hamburger-sized magnet. The NMR frequency for protons subjected to is given by \( \nu_0 = \frac{2}{\mu_0} \frac{1}{4\pi} B_0 \) MHz: 23.9 MHz for the palm system and 20.9 MHz for the 1-chip system.

In the excitation phase of NMR, switches S1 and S2 are closed, and the transmitter (upper half of Fig. 8.4a or b) sends in an RF current to the coil to produce an RF magnetic field in the sample. If the RF magnetic field's frequency is tuned into the NMR frequency, \( \omega_0 \), it resonantly excites the protons, increasing their energy. During this excitation phase, the receiver amplifier stages (in the lower half of Fig. 8.4a or b), except the front-end stage, are isolated from the large excitation signal by short-circuiting their inputs and open-circuiting the RF signal path, using switches S3 through S11 controlled by the ENA command signal. The front-end stage remains connected to the large excitation signal, in order not to place switches in front of it, as lossy switches at the front end would compromise the receiver noise figure.

After protons acquire sufficient energy, the RF transmission is ceased by turning off switches S1 and S2. Nearly simultaneously, the receiver path (lower half of
Fig. 8.4a or b) is activated by operating switches S3 through S11 in the configuration that is opposite to their configuration during the excitation phase. In this reception phase of NMR, the excited protons electromagnetically interact with the coil, inducing an AC voltage signal with the NMR frequency, $\omega_0$, across the coil. This NMR signal, whose peak-to-peak voltage is on the order of 100 nV and bandwidth is about 1 kHz, is amplified and frequency down-converted by the heterodyne receiver. The intermediate frequency (IF) for the receiver is set at 3 kHz, which is high...
enough to mitigate the $1/f$ noise and low enough to facilitate the rejection of out-of-band noise by placing an off-chip band-pass filter at the outputs of the mixers. Two mixers are used to perform the frequency down-conversion with quadrature oscillator outputs. The outputs of the mixers, after the band-pass filtering, are digitized by an off-chip analog-to-digital converter and subsequently undergo an image rejection signal processing, to avoid the extra 3-dB noise figure degradation brought by the frequency down-conversion.

NMR transceivers usually employ two separate clocks, one with the NMR frequency for the proton excitation and the other as the local oscillator (LO) with the frequency different than the NMR frequency by the target IF to produce the correct IF. In contrast, in our transceiver, both the transmitter excitation signal and the receiver local oscillator signal share the identical frequency, both derived from the same clock (Fig. 8.4). In this scheme the clock frequency is set at a value 3 kHz larger than the NMR frequency $\omega_0$ so as to produce the target IF of 3 kHz. Therefore, the excitation signal is 3 kHz off from the NMR frequency. Nonetheless, it can still excite protons, for it has a nonzero bandwidth due to its finite duration and the bandwidth can be made large enough to cover $\omega_0$. The advantage of this single-clock scheme is simplicity: we only need to tune one frequency in the NMR experiment, instead of tuning two clock frequencies while maintaining their difference at 3 kHz.

### 8.2.2 Characteristics of the Coils

The off-chip solenoidal coil of the palm system (Figs. 8.2 and 8.4a) has 14 turns around a capillary tube (inner diameter, 0.75 mm; outer diameter, 1 mm). The sample volume inside the coil is 2 µL. The coil has an inductance of 100 nH, a resistance of 0.5 Ω, and a $Q$ of 28, all measured at the NMR frequency 23.9 MHz of the palm system.

The on-chip planar spiral coil of the 1-chip system (Figs. 8.3 and 8.4b) has 25 turns and occupies an area of 2.5 by 2.5 mm. We use a package that exposes the coil part while encapsulating the rest of the chip (Fig. 8.3). The open part of the package above the coil can hold a 5 µL of sample. To reduce the coil resistance, five metal layers are connected in parallel. SONNET electromagnetic field solver is used in the coil design. The coil has an inductance of 430 nH, a resistance of 31 Ω, and a $Q$ of 1.9, all measured at the NMR frequency 20.9 MHz of the 1-chip system. The low $Q$ is due mainly to the coil’s dc resistance, while the substrate and skin effect are less pronounced at the NMR frequency.

### 8.2.3 Transmitter with Proton Filter

The power amplifier (PA) of the NMR transmitter in general needs to have a large output power tuning capability in order to control the amount of energy
that goes into the protons during the NMR excitation phase. The transmitter in the 2-kg portable system did not integrate a PA, since meeting the power tuning requirement was not trivial with an integrated PA. The new transmitter in the palm and 1-chip systems, which is shown in detail in Fig. 8.5, integrates the entire front-end transmitter chain, including a PA. We manage to tune the PA’s output power by exploiting the proton’s natural high-$Q$ ($\sim 10^4$) filtering ability.

To start with, the PA is realized as a differential chain of cascaded four inverter stages (Fig. 8.5, bottom right). The inverters are consecutively quadrupled in size to sequentially amplify power and ensure drivability at the output. This class-D arrangement is simple to design and does not consume static power, but it produces a square wave output with fixed voltage amplitude of the power supply $V_{DD}$, thus, calling for a technique to tune its output power.

To this end, we tune the duty cycle of the transmitted signal. A given transmitted square wave (frequency: $\omega_0$; amplitude: $V_{DD}$) with a specific duty cycle (Fig. 8.5, top) assumes a particular power distribution of the fundamental tone at $\omega_0$ and higher-order harmonics. The power distribution over the harmonics varies with the duty cycle. Here we only need to look at the variation of the power at $\omega_0$ with the duty cycle, for higher-order harmonics lie outside the “proton filter” band: protons are a high-$Q$ ($\sim 10^4$) band-pass filter centered at $\omega_0$, in the sense that they are not excited by signals that lie outside the frequency band. As the duty cycle is altered from 0 % to 50 %, the $\omega_0$-component changes its voltage from 0 to $\frac{\sqrt{2}}{\pi} V_{DD}$ (Fig. 8.5, top right). This effectively corresponds to the output power tuning.
The duty cycle is tuned by the duty cycle controller that consists of cascaded voltage-controlled delay lines (VCDL) and AND gates (Fig. 8.5, bottom left). Both of its quadrature square-wave inputs, \( I_{in} \) and \( Q_{in} \) (frequency: \( \omega_0 \); amplitude: \( V_{DD} \)), have a 50% duty cycle. The AND operation on \( I_{in} \) and its delayed version yields \( I_{out} \), whose duty cycle varies with the amount of the total delay. The same principle applies to \( Q_{in} \) and \( Q_{out} \). As the total delay changes from 0 to \( \pi/\omega_0 \), the duty cycle shifts from 50% to 0%.

Each voltage-controlled delay line in the duty cycle controller consists of three voltage-controlled delay cells, and each voltage-controlled delay cell consists of three voltage-controlled inverters in parallel (Fig. 8.6). Inv-1 is a standard current-starved inverter. Its delay is not linear with control voltage \( V_C \): with \( V_C \) below a certain threshold, the delay tends toward infinity; with large \( V_C \), the delay hardly tunes. Inv-2, a complementary current-starved inverter with a size smaller than Inv-1, prevents the steep delay increase for small \( V_C \). Inv-3, a current-starved inverter with \( V_C \) fed after a source follower, sustains a delay reduction with increasing \( V_C \). These combine together to yield a more linear tuning characteristics.

The digital pulse sequence generator (Fig. 8.5, upper left) controls the MUX and switches S1 and S2 to produce an adequate NMR excitation pulse sequence such as the CPMG pulse sequence, an essential task in practical NMR works [6].
pulse sequence generator also sets the timing scheme for the receiver by controlling switches S3 through S11 (see also Fig. 8.4) in the way explained in Sect. 8.2.1.

### 8.2.4 Heterodyne Receiver with Passive Amplification

Figure 8.7 shows the detailed structure of the heterodyne receiver, which consists of a low-noise amplifier (LNA), a variable-gain amplifier (VGA), two mixers, and switches S3 through S11, whose usage was explained in Sect. 8.2.1. To handle the NMR signal-to-noise ratio substantially lowered by the ping-pong-ball-sized magnet in the palm system or the lossy on-chip coil in the 1-chip system, the noise figure (NF) of the receiver should be minimized. To this end, both minimization of the LNA’s input-referred noise and optimum LNA-coil noise matching are necessary.

To minimize the LNA’s input-referred noise, we take the following measures in our new LNA design: (1) resistive loads are used in place of active loads. This obviates the need for a common-mode feedback circuit, thus reducing the noise sources. To compensate for the low gain due to the passive loads, we use a two-stage amplifier; (2) PMOS transistors are used as input devices to minimize $1/f$ noise and substrate coupling from digital circuits; and (3) the cascode configuration attenuates coupling between the local oscillator and the LNA. For the optimum LNA-coil noise
matching, we place a capacitor $C_m$ in parallel with the coil (Fig. 8.7, left), where $C_m$ resonates with the coil inductance at the NMR frequency $\omega_0$. This network forms a band-pass filter, whose frequency response peaks at $\omega_0$. The filter may be viewed as a preamplifier with “passive” voltage gain of $Q$, where $Q$ is the coil quality. As $C_m$ has negligible loss compared to the coil, the passive amplification hardly adds any noise and maintains the original signal-to-noise ratio from the coil. In other words, the passive amplifier has an NF close to 0 dB but with the gain of $Q$, which leads to a low receiver NF according to the Friis equation [7]. While this passive amplification scheme is not applicable for wideband signals due to the frequency-dependent transfer function, it suits well the NMR signal, which in general has a very narrow bandwidth (about 1 kHz in our case). Nonetheless, nonoptimal coil-LNA impedance matching at 50Ω, instead of the optimum noise matching based on the passive amplification, has been a conventional choice, as the former is convenient in the conventional NMR electronics that have largely been realized at the discrete level. This shows an advantage of the integrated NMR electronics.

The LNA-coil resonance matching for minimum noise figure corresponds to an impedance mismatch between the LNA and the coil. In contrast, the PA and the coil need to be impedance matched for maximum power delivery. In order to simultaneously achieve both the optimum LNA-coil noise matching and PA-coil power matching, the palm system adopts an advanced matching network (Fig. 8.4a, right). In the 1-chip system, on the other hand, we provide only the LNA-coil resonance matching using $C_m$ (Fig. 8.4b, right) without using the advanced network; thus, the PA and the coil are not impedance matched. This is because the components of the advanced network are too large to be integrated, and using discrete components defeats the purpose of constructing a 1-chip system. Nonetheless, the 1-chip system manages to deliver a reasonable amount of power to the coil for proton excitation.

The VGA (Fig. 8.7) is to handle both the palm and 1-chip systems, whose NMR signal strengths are different. It is a differential common-source amplifier with tunable loads. By tuning the load impedance through $V_{ctrl}$, we can change the gain of the VGA from 0.8 to 22. The mixer is a double-balanced Gilbert mixer with an active load. It provides a voltage gain of 26 dB.

### 8.3 Transceiver Measurements

We implemented two variations of the NMR RF transceiver with essentially the same architecture (Fig. 8.4), one for the palm system and the other for the 1-chip system, in 0.18-μm complementary-metal-oxide-semiconductor (CMOS) technology. The transceiver IC for the palm system occupies an area of 1.4 by 1.4 mm (Fig. 8.2) and is packaged in a 32-lead QFP package. The transceiver IC with the on-chip planar coil for the 1-chip system occupies an area of 4.5 by 2.5 mm (Fig. 8.3) and is so packaged in a 56-lead TSSOP package that the coil part of the
chip, on which an aqueous sample is placed, is left exposed while the rest of the chip is encapsulated.

### 8.3.1 Receiver Measurement

To measure the receiver input-referred noise, we feed a 100-dBm, 21-MHz RF signal to the receiver’s LNA input and use a 21.003-MHz square wave as a local oscillator signal. From the signal and noise power spectrum measured at the receiver’s mixer output (Fig. 8.8) using an Agilent E4448 spectrum analyzer, the receiver gain is inferred, and then, by dividing the measured output noise with the gain, the receiver input-referred noise of $1.26 \text{nV}/\sqrt{Hz}$ is extracted. In this process, the image effect due to the frequency down-conversion is factored.
out, for the actual operation indeed performs image rejection via back-end digital signal processing (Sect. 8.2.1). Using the measured input-referred noise and the coil impedance (Sect. 8.2.2), we infer the receiver NF. In the palm system, a passive gain of 28 offered by the resonance matching (Sect. 8.2.4) lowers the NF from 22.5 to 0.9 dB. In the 1-chip system, a passive gain of 2.1 lowers the NF from 6.1 to 2.2 dB. The combination of the small input-referred noise and the resonance matching (optimum noise matching) leads to sufficiently low NF, making the ping-pong-ball-sized magnet and the lossy on-chip coil viable system options.

8.3.2 Transmitter Measurements

The measured output impedance of the differential PA (Sect. 8.2.3) is 27. With the power supply $V_{DD}$ of 3.3 V, the maximum deliverable power at the fundamental tone is 82 mW. The measured delay versus control voltage, $V_C$, of the entire voltage-controlled delay line (VCDL) in the duty cycle controller (Fig. 8.5, bottom left) is shown in Fig. 8.9. The delay is altered from 29 to 2 ns, as $V_C$ is varied from 0 to 3.3 V.

The measured duty cycle as a function of $V_C$ for a 21-MHz excitation signal is shown in Fig. 8.10. As $V_C$ is changed from 0 to 3.3 V, the duty cycle increases from 0 % to 45 %. This translates to the tuning of output power at the fundamental tone from 0 to 80 mW (98 % of the total deliverable power of 82 mW).
8.4 NMR Experiments and NMR-Based Biomolecular Sensing

8.4.1 Proton NMR Experiments

NMR is performed on protons of hydrogen atoms in a 2-μL water sample using the palm system. Figure 8.11 shows a measured, down-converted NMR signal. The repeated ringings, which are the result of the proton excitations using a CPMG pulse sequence [6], constitute the NMR signal. It decays with characteristic time called $T_2$, one of the key parameters in NMR experiments [6], which we use in our NMR-based biomolecular sensing, as seen shortly. $T_2 = 100$ ms is extracted from the exponentially decaying envelope of the NMR signal, shown as a dotted line. The repeated spikes between the ringings are due to the coupling of the large excitation signals, but they do not compromise the observation of the NMR signal (ringings), as they occur at different time instances. The spin-mass (the minimum mass of water that produces a detectable NMR signal; a smaller minimum mass corresponds to a higher spin-mass sensitivity) sensitivity is 2.5 times higher than that of the 2-kg portable system and 150 times higher than that of the state-of-the-art commercial system [1].

Figure 8.12a shows a measured, down-converted NMR signal obtained in a proton NMR experiment (5 – μL water sample) done with the 1-chip system, from which we obtain $T_2 = 722$ ms. The spikes coupled from the large excitation signals are more pronounced, but once again, they do not hamper the observation.
of the NMR signal due to their occurrence at different time instances. Note the difference between the values measured using the 1-chip and palm systems. While $T_2 = 722$ ms obtained with the 1-chip system well approximates the true value, $T_2 = 100$ ms obtained with the palm system is a substantial underestimation of the true value, which is due to the pronounced static magnetic field inhomogeneity of the ping-pong-ball-sized magnet used in the palm system [6]. Nonetheless, this is not a fundamental problem, as repetition of the CPMG pulses at a faster rate, which the current implementation has no provision for but is easy to incorporate, can readily yield the correct value [6]. Moreover, in our $T_2$-based biomolecular sensing experiments that will be presented shortly, we focus on the relative measure of values.

Figure 8.12b shows a measured, down-converted NMR signal obtained in another water proton NMR experiment using the 1-chip system, this time, after 0.05 mM magnetic nanoparticles [Fe] (30 nm) are added in the water sample. The measured $T_2$ is decreased to 93 ms. This reduction of the $T_2$ value in the presence of magnetic nanoparticles, which perturb the NMR behavior, is expected from the NMR theory [8].

## 8.4.2 NMR-Based Biomolecular Sensing

Figure 8.13 shows the detection of avidin protein using the palm system. Magnetic particles (38 nm) coated with biotins are put into a 2-$\mu$L water inside the solenoidal...
Fig. 8.12 Measured proton NMR signal with the 1-chip system. (a) Water. (b) Water with magnetic nanoparticles (0.05 mM) (From Ref. [3])

coil. In the absence of avidin (Fig. 8.13, top), the particles stay uniformly dispersed, yielding $T_2$ of 48 ms. In the presence of avidin (Fig. 8.13, bottom), the biotinylated magnetic particles bind to avidin to self-assemble into clusters [9]. The effectively larger magnetic particles reduce $T_2$ to 40 ms [9]. The reduction in $T_2$ corresponds to the detection of avidin. The palm system detects down to 1 avidin molecule in 600 million water molecules.

We use the 1-chip system to detect human chorionic gonadotropin (hCG), the hormonal protein found in blood or urine (Fig. 8.14). Its primary use for female
patients is as a pregnancy indicator, but for male patients it serves as a marker for certain cancers, such as choriocarcinoma, germ cell tumors, and islet cell tumors. Magnetic particles (38 nm) coated with mouse monoclonal antibody to hCG are put into a 5 – μL water sample placed on the on-chip planar coil. $T_2 = 169$ ms in the absence of hCG, and $T_2 = 141$ ms in its presence, which corresponds to the detection of hCG. The 1-chip system detects down to 1 hCG molecule in 12 billion water molecules.

We also detect human bladder cancer cells using the 1-chip system (Fig. 8.15). Magnetic particles (40 nm) coated with monoclonal antibody to bladder cancer cell surface markers are put into a 5 – μL water sample placed on the on-chip coil. In the absence of the cancer cells, the magnetic particles are monodispersed (Fig. 8.15, top left); in the presence of the cancer cells, magnetic particles bind to the cell surface (Fig. 8.15, bottom left). A following centrifugation step [10] separates the cells and unbounded magnetic particles in different layers, and then the unbounded magnetic particles are washed out. As a result, for the sample without cancer cells, all magnetic particles are removed in the washout process, leading to a longer $T_2$ of 506 ms (Fig. 8.15, top); for the sample with cancer cells, the magnetic particles that bind to the cell surface remain, leading to a shorter $T_2$ of 440 ms (Fig. 8.15, bottom).
Fig. 8.14  hCG detection with the 1-chip system (From Ref. [3])

Fig. 8.15  Human bladder cancer cell detection with the 1-chip system (From Ref. [3])
The reduction in $T_2$ indicates the existence of cancer cells. The concentration detection threshold is 17.5 cells per $\mu$L.

### 8.5 Comparison to Other Miniaturization Works

Before concluding, we compare our miniature NMR systems to other NMR system miniaturization efforts [11–17]. As shown in Table 8.1, other works use small magnets [11, 12] or integrated transceivers [13–17], but not both as we did. In addition, the integrated transceivers used in [13–17] have integration levels considerably lower than ours. Overall, in terms of the entire system dimension and integration level, our miniature NMR systems represent a meaningful advance from previous miniaturization efforts.

### 8.6 Conclusion

By combining the physics of NMR with silicon RF integrated circuits, we developed three miniature NMR systems for biomolecular sensing ultimately aimed at disease screening. We would like to view the value of our work from a few different angles. First, NMR has a broad array of applications in addition to biomolecular sensing, and from this general NMR point of view, our work on the small, low-cost NMR is a meaningful development that can help make the benefits of NMR closer to our lives. Second, from the circuit point of view, our work showcases how silicon RF integrated circuits can be used not only for wireless applications but also in 1 day for human health care and disease screening in direct interface with biological systems. Third, from the biotechnology point of view, our work suggests a way to perform general-purpose disease screening in the low-cost, handheld platform.
References

Chapter 9
Diagnostic Magnetic Resonance Technology

Changwook Min, Huilin Shao, David Issadore, Monty Liong, Ralph Weissleder, and Hakho Lee

Abstract For the sensitive and quantitative measurement of protein biomarkers, pathogens, and cells in clinical samples, magnetic nanoparticles (MNPs) offer unique advantages over traditional detection methods. Specifically, due to the inherently negligible magnetic background of biological material, MNPs can be used to obtain highly sensitive measurements in minimally processed samples. Our detection platform, termed diagnostic magnetic resonance (DMR), exploits MNPs to modulate the nuclear magnetic spin-spin relaxation time of water. Here, we review work done by our group to develop more effective MNP biosensors, advanced conjugational strategies to target the MNPs to molecular targets, and highly sensitive miniaturized NMR systems. We demonstrate this platform as a robust and easy-to-use system for the detection of a wide range of targets in clinical settings including whole cells, proteins, DNA/mRNA, metabolites, drugs, viruses, and bacteria.
9.1 Introduction

Robust, sensitive, and easy-to-use biosensors for the detection and quantification of rare biomarkers will have significant applications in both basic research and clinical practice. If made available, these platforms could aid in understanding of fundamental biology, in accurately detecting diseases at their early stage, and in evaluating and monitoring the efficacy of therapy [1–3]. To realize such sensors, the underlying detection technology should ideally (1) enable high sensitivity and accuracy, with minimal false positives and negatives; (2) support short assay time with minimal sample processing; and (3) allow for multiplexed detection in a single parent sample [4]. Different types of sensing platforms, fulfilling some of these requirements, have been developed based on optical [5, 6], electronic [7, 8], or mass-based [9] detection. These systems, however, often require lengthy sample purification, large sample volumes, or long assay time, which can potentially limit their clinical utility and adaption.

Biosensors based on magnetic detection have recently emerged as a promising diagnostic platform. Due to the intrinsically negligible magnetic susceptibilities of biological entities, magnetic detection experiences little interference from native biological samples; even optically turbid samples will often appear transparent to magnetic fields. Biomarkers of interests, when magnetically labeled, however, can attain a high contrast against the biological background. Recent progresses in the synthesis of magnetic nanoparticles (MNPs) have further advanced the magnetic detection technology. With their size scale similar to that of biological molecules, MNPs can efficiently and abundantly bind to biological targets, amplifying analytical signals [10–13]. Various detection technologies have been developed based on this magnetic-tagging concept. These include techniques that use magnetometers, such as superconducting quantum interference device (SQUID) [14–16], magnetoresistive sensors [17–20], and Hall sensors [21], all of which directly measure the magnetic fields arising from the magnetically labeled targets.

We have recently developed a new magnetic sensing platform, diagnostic magnetic resonance (DMR) [22]. Contrary to directly measuring the magnetic moments of the labeled targets, the DMR uses nuclear magnetic resonance (NMR) as the detection mechanism. When placed in NMR magnetic fields, MNPs create local magnetic fields and change the relaxation rate of surrounding water molecules [23]. The detection offers an intrinsic signal amplification mechanism, as more than millions of water molecules can be affected by a single MNP. Moreover, since the signal is generated from the entire sample volume, the assay procedure is significantly simpler than the direct magnetic detection in which MNP-labeled targets have to be closely positioned to the sensing elements.

By optimizing MNPs and miniaturizing NMR detectors, the DMR detection sensitivities for various target types have been considerably improved over the last few years. These developments enable rapid and multiplexed detection on a wide range of targets in microliter sample volumes, including nucleic acids [24], proteins [22], drugs, bacteria [25], and tumor cells [26–28]. With the recent integration
of bioorthogonal targeting strategies [27, 29] as well as accurate and real-time control of device temperature [23], the DMR platform has become more robust and sensitive, allowing operation in clinical settings [26]. This chapter reviews the latest development of the DMR technology, focusing on its three major components: magnetic nanoagents, miniature NMR systems, and optimized assay protocols. Specific biomedical and clinical DMR applications will also be discussed.

### 9.2 Principle of DMR Detection

The DMR detection of MNP-labeled cells is realized by exploiting the “$T_2$-shortening” effect of MNPs in NMR measurements [30]. When placed in static, polarizing magnetic fields for NMR detection, MNPs produce local dipole fields with strong spatial dependence, which efficiently destroy the coherence in the spin-spin relaxation of water protons. MNP-labeled objects consequently cause faster decay of NMR signal, or shorter transverse relaxation time $T_2$, than nontargeted ones (Fig. 9.1).

The capability of MNPs to induce $T_2$ changes is defined as transverse relaxivity ($r_2$) [31]. With MNPs in solution, the relaxation rate ($R_2 = 1/T_2$) can be expressed as [28]

$$R_2 = R_W + r_2 \cdot \frac{N_p}{V}, \quad (9.1)$$

---

Fig. 9.1 Principle of DMR detection. Biological objects (e.g., cells) tagged with MNPs can accelerate the transverse relaxation of water protons. Compared to the nontagged samples, the NMR signal will decay faster in time domain, therefore providing a sensing mechanism.
where $R_w$ is the relaxation rate of the background (usually water), $V$ is the NMR detection volume, and $N_P$ is the total number of MNPs in $V$. If each biological cell has $n$ MNPs and the total number of cells is $N_C (= N_P/n)$, the net change of $R_2 (\Delta R_2 = R_2 - R_w)$ is given as

$$\Delta R_2 = r_2 \cdot \frac{N_P}{V} = \frac{n \cdot r_2}{V} N_C = r_2^{\text{cell}} \cdot \frac{N_C}{V}, \quad (9.2)$$

where $r_2^{\text{cell}} (= n \cdot r_2)$ is defined as the cellular relaxivity (transverse relaxivity per given cell concentration). Note that $r_2^{\text{cell}}$ is indicative of the abundance of relevant surface biomarkers. NMR thus can be used effectively for molecular profiling of target cells [27, 28]. Equation 9.2 provides valuable insights into how to increase the sensitivity and specificity of NMR-based sensors:

- **MNPs with high $r_2$ relaxivity**. Pronounced $R_2$ changes will occur when cells are labeled with MNPs of high $r_2$ relaxivity [23]. Because $r_2$ is proportional to the magnetic moment ($\mu_p$) of particles [32, 33], making magnetically stronger MNPs will benefit the measurements.

- **Maximal MNP labeling on cell**. The $R_2$ changes are also directly proportional to the number of MNPs loaded onto cells. In addition to increasing the $r_2$ potency of individual MNPs, it is equivalently important to establish a labeling protocol to maximize and/or amplify MNP loading on cells.

- **Miniaturized NMR probes**. Higher sensitivity can be achieved on the device level by decreasing the NMR detection volume ($V$). This approach can effectively increase the analyte concentration ($N_C / V$), leading to large $\Delta R_2$. Furthermore, smaller NMR probes assume higher SNR (signal-to-noise) ratio due to the increased sample filling factor. It can be shown that the sensitivity of NMR coils (with a typical dimension of $d$) scales as $d^{-1/2}$ [34].

These recognitions motivated us to explore three major activities in DMR development: synthesis of new MNPs, optimization of such MNPs for cellular labeling, and miniaturization of NMR systems. The following sections will describe these accomplishments.

### 9.3 New Magnetic Nanoparticles

We have developed many different types of MNPs (Table 9.1) and engineered the particle size and composition to enhance the transverse relaxivity. According to the outer-sphere model of transverse relaxation, the $r_2$ value of an MNP is proportional to $\tau_d \cdot M^2$, where $\tau_d$ is the residence time of water molecules around the particle and $M$ is the particle magnetization [35]. The efforts to enhance $r_2$ were thus focused on synthesizing larger MNPs using magnetically stronger material. We herein introduce two types of such particles: metal-doped ferrite and iron (Fe)-based MNPs.
Table 9.1  MNPs developed for DMR applications

<table>
<thead>
<tr>
<th>MNP core material</th>
<th>Core size (nm)</th>
<th>Per nanoparticle</th>
<th>Magnetic moments ($\times 10^{-15}$ emu)</th>
<th>$r_2$ relaxivity ($\times 10^{-15}$ s$^{-1}$ L$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ferrite</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MION</td>
<td>3</td>
<td>0.003</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>CLIO</td>
<td>7</td>
<td>0.03</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>PION</td>
<td>11</td>
<td>0.1</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Fe$_3$O$_4$</td>
<td>16</td>
<td>0.7</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Doped ferrite</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CoFe$_2$O$_4$</td>
<td>16</td>
<td>0.7</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>MnFe$_2$O$_4$</td>
<td>16</td>
<td>0.8</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Fe-core particles</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fe@FeO</td>
<td>16</td>
<td>1.1</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Fe@Fe$_3$O$_4$</td>
<td>16</td>
<td>1.5</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Fe@MnFe$_2$O$_4$</td>
<td>16</td>
<td><strong>1.6</strong></td>
<td><strong>68</strong></td>
<td></td>
</tr>
</tbody>
</table>

### 9.3.1 Ferrite-Based MNPs

With their excellent stability and biocompatibility, cross-linked iron oxide (CLIO) nanoparticles have been widely used for DMR applications [22,36]. CLIO nanoparticles contain a superparamagnetic iron oxide core (3–5-nm monocrystalline iron oxide) composed of ferrimagnetic magnetite (Fe$_3$O$_4$) and/or maghemite ($\gamma$Fe$_2$O$_3$). The metallic core is encased with biocompatible dextran, which is cross-linked and functionalized with primary amine. Amine-terminated CLIO nanoparticles have an average hydrodynamic diameter of 38 nm, and about 60 amine groups are available for bioconjugation per nanoparticle. The $r_2$ of CLIOs is $\sim 50$ s$^{-1}$ mM$^{-1}$ [Fe] [24,37] as measured at 40°C and at the external field of $B_0 = 0.5$ T.

Two main strategies have been employed to further improve the magnetization of ferrite nanoparticles and thereby the $r_2$ relaxivity: magnetic doping and nanoparticle sizing. Magnetic doping with ferromagnetic elements such as manganese (Mn), cobalt (Co), or nickel (Ni) has been known to modulate the overall magnetization of MNPs [38,39]. Among these doped ferrite MNPs, MnFe$_2$O$_4$ nanoparticles have the highest magnetization, as Mn$^{2+}$ ions have the highest spin quantum number (5/2). Moreover, larger nanoparticles are also known to have increasing magnetization [40]. Spin canting, a feature that decreases the overall magnetic moment of small nanoparticle due to tilted surface spins, can be reduced in bigger nanoparticles to increase the overall magnetization. Concurrently, larger particle size further enhances the particle $r_2$ by increasing $r_d$.

We employed both magnetic doping and sizing strategies to produce MnFe$_2$O$_4$ nanoparticles with superior $r_2$ relaxivity [28]. These particles were synthesized in the organic phase by reacting iron (III) acetylacetonate [Fe(acac)$_3$], manganese (II) acetylacetonate [Mn(acac)$_2$], and 1,2-hexadecanediol at elevated temperature (300°C). Through a seed-mediated growth approach, the particle size was stepwise increased from 10 nm to 12, 16, or 22 nm. MnFe$_2$O$_4$ nanoparticles with diameter $\leq 16$ nm were found to be highly monodisperse and superparamagnetic at 300 K.
Fig. 9.2 Mn-doped ferrite for DMR detection. (a) Transmission electron microscope (TEM) images of MnFe$_2$O$_4$ MNPs prepared in our laboratory. The particles have a narrow size distribution and consisted of a single crystal (insets). (b) The particles in (a) showed superparamagnetic behavior at room temperature (Adopted from [28]. Copyright 2009 National Academy of Sciences, USA)

(Fig. 9.2). Because of their larger magnetic core, these MnFe$_2$O$_4$ nanoparticles assumed high relaxivities with $r_2$ values approaching 420 s$^{-1}$·mM$^{-1}$[metal] (equal to $6 \times 10^{-14} L \cdot s^{-1}$ per particle), more than eight times greater than CLIO nanoparticles in metal basis (50 s$^{-1}$·mM$^{-1}$[metal] or $7 \times 10^{-16} L \cdot s^{-1}$ per particle) [28].

### 9.3.2 Fe-Core MNPs

Ferromagnetic metals, instead of their oxides, have been suggested as an ideal constituent of MNPs for their superior magnetization. However, while Fe-core MNPs can achieve high $r_2$ relaxivities [41], these monometallic MNPs are extremely reactive and require protective layers to prevent rapid oxidation.

Recently, a unique 16-nm Fe-core/ferrite-shell MNP (Fe@ferrite) has been developed for DMR applications (Fig. 9.3) [42]. The particle consists of an elemental iron core (not iron oxide) and a protecting oxide shell. A novel synthetic route was established, as summarized below, that allows for the preparation of large yet monodisperse Fe MNPs and the growth of protective ferrite shells around existing Fe MNPs (Fig. 9.3a).

- **Larger Fe cores.** Fe MNPs were formed by thermally decomposing metal complexes [Fe(CO)$_3$] in the presence of surfactant (oleylamine) under air-free condition. By increasing the reaction temperature during Fe MNP synthesis, the particle size could be proportionally increased [42]. The phenomenon can be attributed to the higher reactivity of Fe ions at elevated temperatures during particle formation [43]. Applying this approach, Fe MNPs with diameters up to 18 nm could be prepared while maintaining the relative size variations <5%.
Fig. 9.3 Fe@ferrite MNPs. (a) New synthetic route. The size of the Fe core is increased by elevating the reaction temperature. The core is then overcoated with a ferrite shell. To further enhance the particle magnetization, the shell is metal doped. (b) TEM image of Fe@ferrite MNPs. Fe MNPs (dotted circles) were preserved during the coating process. (c) Fe@ferrite MNPs showed stable magnetic properties over time with small changes (<10%) in saturation magnetization ($M_s$). (d) The field-dependent magnetization ($M$) of Fe@MFe$_2$O$_4$ MNPs at 300 K showed an unusual feature: negligible remanent moments but the presence of hysteresis (Adopted from [42]. Copyright 2011 John Wiley and Sons, Inc.)

- Ferrite-shell overcoat. Native Fe MNPs undergo rapid oxidation, which necessitates the presence of protective shells. To prepare such shells, we carried out the reaction for ferrite synthesis in the presence of as-prepared Fe MNPs, based on the hypothesis that Fe MNPs could serve as nucleation sites for ferrite formation. Indeed, the resulting particles (Fe@ferrite, Fig. 9.3b) had Fe cores whose sizes were similar to that of the initial Fe MNPs (dashed circles in Fig. 9.3b). Importantly, Fe@ferrite MNPs maintained their shape and magnetic properties over time (Fig. 9.3c), verifying the sturdy protection against oxidation.
by the shell. Note that we further enhanced the overall magnetic moment and $r_2$ relaxivity by doping the ferrite with Mn$^{2+}$ during the shell formation (Fe@MnFe$_2$O$_4$ MNPs).

The Fe@ferrite MNPs assumed higher saturation magnetization ($796 \text{kA m}^{-1}$) and $r_2$ ($7 \times 10^{-14} \text{ L s}^{-1}$ per particle, $430 \text{s}^{-1} \text{mM}^{-1}$[metal]) than similarly sized ferrite MNPs, primarily due to the large Fe cores. It is noteworthy that the Fe core is in a thermally stable ferromagnetic state with nonzero coercivity. The ferrite shell, which is superparamagnetic, however, effectively reduces the overall coercivity of particles by leading the magnetization processes at small external magnetic fields [42]. The resultant Fe@ferrite MNPs thus displayed a unique magnetic feature, namely, the presence of hysteresis with negligible coercivity (Fig. 9.3d). This property is crucial in preventing interparticle aggregations from magnetic interactions. When applied for DMR assays, these Fe@ferrite MNPs achieved superior performance, capable of detecting picomolar avidin and single cancer cells in whole blood samples.

### 9.4 Optimizing MNPs for DMR Applications

In addition to the above-mentioned strategies to improve nanoparticle relaxivities through inorganic chemistry, postsynthesis modifications such as better particle surface chemistry and new labeling approaches have also been developed for DMR applications. These novel postsynthesis modifications not only improve the detection sensitivities but also simplify the targeting assays, making the DMR platform easily applicable to detect a wide range of biological entities and translatable for effective clinical utility.

#### 9.4.1 Biocompatible Coating on Hydrophobic MNPs

Most MNPs, synthesized via the thermal decomposition method, are suspended in nonpolar solvents and coated with hydrophobic surfactant. For biological applications, these particles should be transferred into aqueous phase. We have traditionally used a small bifunctional molecule DMSA (meso-2,3-dimercaptosuccinic acid) to replace hydrophobic capping layers (e.g., oleic acid or oleylamine) on MNP surfaces [44,45]. The resulting particles, however, displayed short-term stability (<3 month), gradually precipitating out in physiological buffers [44].

Overcoming the issue, we have established a new, polymer-based surface coating that can render MNPs hydrophilic with superb stability under varying pH and ionic strength (Fig. 9.4a) [46]. As a coating substrate, we selected polyvinyl alcohol (PVA) since the material is synthetic, inexpensive, hydrophilic, and biodegradable [47–49]. The polymer was further modified into carboxymethyl polyvinyl alcohol (CMPVA); we hypothesized that multidentate carboxylic (-COOH) groups would
Fig. 9.4 Highly stable polymer coating on MNPs. (a) Hydrophobic layers on MNPs were first replaced with TMAH (tetramethylammonium hydroxide) and then further coated with carboxymethyl polyvinyl alcohol (CMPVA). TEM confirms that the particles are well dispersed in water after the CMPVA coating. (b) The presence of carboxylic acids on polyvinyl alcohol (PVA) is crucial to the stability of the coating; PVA-coated MNPs aggregated in aqueous solutions. (c) Cancer cells (SkBr3) were labeled with fluorescent CMPVA-MNPs by targeting HER2/neu surface receptors. CMPVA coating displayed remarkably low nonspecific binding to the cells (Reproduced from [46]. Copyright 2010 John Wiley and Sons, Inc.)

allow CMPVA to strongly bind to the metal oxide surface. Indeed, when hydrophobic MnFe$_2$O$_4$ MNPs were coated with CMPVA, the particles showed excellent, long-term solubility (>12 months) in aqueous buffers, whereas PVA-coated particles precipitated spontaneously (Fig. 9.4b). CMPVA coating also provided free amine (–NH$_2$) groups for bioconjugation.

When tested for cellular labeling, the CMPVA coating exhibited extremely low nonspecific binding. Figure 9.4c shows an example of cellular labeling with CMPVA-MNPs. Cancer cells (SkBr3) were first labeled with biotinylated anti-HER2/neu monoclonal antibodies, followed by an incubation with neutravidin-modified, fluorescent CMPVA-MNPs. Control samples were prepared in the same way but without antibody injection. Strong fluorescence signal could be visualized on the surface of targeted cells, whereas the signal from control samples was
negligible. This obvious difference between the control and positively labeled samples is highly important for DMR assay in reducing false positives in biosensing.

### 9.4.2 New Labeling Method for High MNP Loading

Besides improving the physical properties of MNPs, equally important for the DMR assay is to develop an efficient method for MNP labeling on target cells. A conventional way is to use MNPs preconjugated with target-specific affinity ligands [22], which often requires extensive optimization of the affinity ligands and the conjugation method for each new target.

We have developed a new targeting strategy, BOND (bioorthogonal nanoparticle detection), that is modular and broadly applicable and can amplify MNP binding to biological objects [27]. The BOND is based on $[4 + 2]$ Diels-Alder cycloaddition, especially between tetrazine (Tz) and trans-cyclooctene (TCO; Fig. 9.5a) [50]. The reaction is fast and irreversible (covalent) and can be performed at room temperature without using any catalyst (copper). We have adapted the chemistry for MNP labeling of cells (Fig. 9.5b), wherein cells are pretargeted with TCO-modified antibodies and subsequently incubated with Tz-loaded MNPs (Tz-MNPs). Multiple TCO tags (usually $\sim 20$) could be incorporated onto an antibody, and such modified antibodies were found to maintain their affinity. Consequently, the antibodies functioned as a scaffold to promote multiple attachments of Tz-MNPs. Indeed, flow cytometry measurements showed that the BOND yielded $\sim 15$-fold improvement in MNP loading on cells, compared to labeling with antibody-MNP direct conjugates (Fig. 9.5c). The trend was further confirmed in NMR-based cell detection (Fig. 9.5d); the BOND method yielded more pronounced $T_2$ changes and improved the cellular detection limit.

BOND has been successfully adapted for DMR molecular profiling of experimental cellular samples [27] and clinical fine-needle aspirate samples [26]. Recently, the BOND technology was further generalized by developing newer two-step detection schemes based on complementary oligonucleotide approaches [29], alternative cycloaddition chemistries [51], and cyclodextrin/adamantine supramolecular interactions [52].

### 9.5 Miniaturized NMR System

The development of miniaturized nuclear magnetic resonance (μNMR) systems [22] represents a key milestone in sensitive detection in DMR. Device miniaturization brings several distinctive advantages for sensitive detection and clinical translation. First, it provides a promising way to improve the detection sensitivity. Smaller NMR systems reduce the detection volume, which in turn effectively increases the concentration of MNP-targeted cells for large $R_2$ changes (Eq. 9.3).
Fig. 9.5 Bioorthogonal nanoparticle detection (BOND). (a) The method is based on the Diels-Alder cycloaddition between trans-cyclooctene (TCO) and tetrazine (Tz). (b) Cells are prelabeled with TCO antibodies and targeted with Tz-MNPs. The antibody provides sites for multiple MNP couplings. (c, d) Compared to the direct targeting with MNP-antibody conjugates, BOND method enabled higher MNP loading on target cells as confirmed by fluorescent (c) and μNMR (d) measurements (Reproduced from [27]. Copyright 2010 Nature Publishing Group)

Second, miniaturized NMR probes (coils) produce much stronger radio-frequency (RF) magnetic fields per unit current, leading to higher signal-to-noise per unit sample volume [53]. Third, with smaller RF coils, the requirement for spatial homogeneity of static magnetic fields becomes less stringent, making it possible to use small, portable magnets [22]. The entire measurement system can be realized as a portable device for point-of-care operations.
Fig. 9.6 Schematic of microNMR system. (a) The system consists of an array of microcoils for NMR measurements, microfluidic networks for sample handling and mixing, embedded NMR electronics, and a permanent magnet for polarizing magnetic field generation. The whole setup can be packaged as a handheld device for portable operation. (b) Micrograph of a microcoil array. The microcoil (inset) generates RF magnetic fields to excite samples and receives the resulting NMR signal. (c) Schematic of the NMR electronics. The circuit is designed to perform $T_1$ and $T_2$ measurements via inversion recovery and CPMG (Carr-Purcell-Meiboom-Gill) pulse sequences, respectively. (d) Example of a microfluidic network for effective mixing between magnetic nanoparticles and the samples (Reproduced from [22]. Copyright 2008 Nature Publishing Group)

9.5.1 System Concept

NMR-based cell detection has been previously performed on benchtop relaxometers (e.g., minispec, Bruker) [36,54–56]. The systems are equipped with permanent, low-field (<1 T) magnets for field generation, which simplifies operation and housing of the equipments; however, the main drawbacks of the benchtop system include the use of relatively large sample volumes (~100 μL) and the lack of capability for parallel measurements.

Overcoming these limits, the first μNMR prototype was designed and tested for the feasibility of miniaturization. Figure 9.6 shows the main features of the μNMR
system: planar microcoils, microfluidic networks, onboard NMR spectrometer, and a portable magnet. The microcoils are used for NMR detection and are arranged in an array format for parallel measurements. The microfluidic networks facilitate the handling and distribution of small volumes of samples. A small, portable magnet (NdFeB, $B_0 = 0.5$ T) was employed to generate NMR field. The system measured the $T_1$ relaxation time using inversion recovery pulse sequences; for $T_2$ measurements, Carr-Purcell-Meiboom-Gill (CPMG) spin-echo pulse sequences were used to compensate for the inhomogeneity of the polarizing magnetic field. To generate versatile pulse sequences while using minimal electronic parts, we devised a new circuit schematic for NMR electronics that has served as a blueprint for subsequent NMR systems.

### 9.5.2 Optimal NMR Probe Design

Reducing sample volume requirements can lead to the effective increase of cell concentrations. However, it can also lead to degradation of the signal-to-noise ratio (SNR), as the absolute level of the NMR signal is proportional to the sample volume. System miniaturization thus should be accompanied by measures to maintain or enhance SNR to truly improve the detection sensitivity. In the second generation of μNMR system, we focused on improving SNR by engineering the NMR probes.

The SNR of a NMR probe can be expressed as

$$SNR = \kappa M_0 \sqrt{\frac{\mu_0 Q \omega_0 V_c}{4k_B T \Delta f}}, \quad (9.3)$$

where $\kappa$ is the fraction of the coil volume ($V_c$) occupied by the samples (filling factor), $M_0$ is the nuclear magnetization of the sample, $\mu_0$ is the vacuum permeability, $\omega_0$ is the Larmor frequency, $Q$ is the quality factor of an NMR coil, and $\Delta f$ is the bandwidth of a receiver electronics. For a given NMR setup (i.e., the same magnets and electronics), SNR could be improved by increasing $\kappa$ and $Q$, which are properties of the NMR probes. Indeed, we have demonstrated a new probe design that achieves both maximal $\kappa (\approx 1)$ and high $Q (\sim 25)$ [28]. In this design, the probe consisted of a solenoidal microcoil embedded in a microfluidic structure (Fig. 9.7a). Solenoidal coils were chosen for their higher SNR than planar or birdcage coils [53]. To increase $\kappa$, we adopted the cast-molding technique in device fabrication. First, the coils were wound around polyethylene tubes and subsequently immersed into a polymer (PDMS). After PDMS cure, the tubes were withdrawn to open up the fluidic channels. With this design, the entire bore of the coil was available for samples; in one example, the new probe displayed $>350\%$ larger SNR than a similar coil wrapped around a tube (Fig. 9.7b). Compared to the lithographically patterned planar coils in our previous systems [22], the improvement in SNR was much more significant ($>20$-fold enhancements); the solenoidal coil excited larger volumes of
Fig. 9.7 New NMR coils for higher sensitivity. (a) A solenoidal coil is embedded along with a microfluidic channel. The entire bore of the coil is available to samples, maximizing the filling factor ($\approx 1$). (b) The new coil, due to its high filling factor, offered $>350\%$ enhancement in SNR compared to a similar coil wrapped around a tubing. (c) The $R_2$ changes can be amplified by concentrating MNP-labeled samples inside the microcoil. Based on the coil design in (a), we incorporated a filter that captures target objects. The photo shows a prototype device containing 100-nm pore filter (Adopted from [28] and [25]. Copyright 2009 National Academy of Sciences, USA and John Wiley and Sons, Inc.)

samples by producing more homogeneous radio-frequency magnetic fields and had smaller electrical resistance (e.g., 0.3 vs. $3.2\Omega$ at 20 MHz).

We have further modified the embedded coil to improve the overall SNR and to streamline the assay procedure. In the new probe design, a cell-capturing membrane filter was incorporated at the outlet side of the coil (Fig. 9.7c). The filter serves two essential functions. First, it size selectively captures cells and concentrates them inside the NMR detection coil, which leads to more pronounced $R_2$ changes (Eq. 9.2). The filter also provides a way to detect a small number of targets from large sample volumes. Second, the filter enables on-chip separation of cells from unbound MNPs, therefore obviating the need for separate off-chip purification steps (e.g., centrifugation). Together with the highly magnetic Fe@ferrite MNPs, this new probe has been applied to diagnose tuberculosis [25]. Using the attenuated bacillus Calmette-Guérin (BCG) as a surrogate for Mycobacterium tuberculosis, we detected as few as 20 colony-forming units (CFUs) in sputum (1 mL). Importantly, the entire detection procedure was performed in a single-chip format, minimizing sample loss and making the assay simple and fast (in less than 30 min).
9.5.3 Clinical System

Our third-generation μNMR system is designed for practical applications in a clinical environment (Fig. 9.8) [57]. Its small size, easy accessibility, and high robustness provide end users with fast and stable measurements of biological samples. The new μNMR system is composed of three core parts: a newly designed probe for clinical samples, NMR electronics, and a user-friendly software.

The system design is based on that of previous generations of μNMR and consists of a small portable magnet ($B_0 = 0.5$ T) and a solenoidal coil for higher SNR (Fig. 9.8 left). A custom-made PMMA (polymethyl methacrylate) housing cages the magnet, the microcoil, and RF matching circuit. With its major improvement focused on clinical translatability, this new system uses disposable thin-walled polyimide tubes to load biological samples for measurement, thereby eliminating potential contamination of the NMR probe. The disposable tubes are filled with samples (~5 μL) and are inserted into the coil bore for NMR detection. Modular coils made in a variety of sizes can be plugged into the system to optimally accommodate available sample volumes (1–100 μL).

A small form-factor (20 × 12 × 5 cm) NMR electronics is implemented using off-the-shelf integrated circuit (IC) chips to achieve cost-effective (<$200) and highly programmable NMR platform (Fig. 9.8 middle). The NMR electronics generates the NMR pulse sequences, acquires the NMR signal, and communicates with external terminals (computer, mobile devices). It has three main parts: a microcontroller unit (MCU), an RF transmitter, and a signal receiver. The MCU (TMS320F28235, Texas Instruments) controls overall RF transceiver operations as well as data communication with external terminals. As an RF transmitter, a direct digital synthesis chip (AD9954, Analog Devices) is employed to generate two RF signals with 90° phase difference that are modulated by voltage-controlled switches (ADG1419BRMZ, Analog Devices). The RF heterodyne system is implemented to process the NMR signal. First, the signal is amplified by low-noise amplifier (AD604, Analog Devices) and down-converted to baseband (1–10 kHz) by a mixer (ADE-6, Mini-Circuits). Baseband signals subsequently pass a low-pass filter and are digitized by an analog-to-digital converter (AD7625, Analog Devices). Mobile devices (e.g., iPhone™, iPad™), which are connected to the MCU via a standard communication channel (e.g., USB, Bluetooth), receive digital data and show graphical outputs.

The NMR software incorporates graphical user interface (GUI), data acquisition/process, data logging/sharing, and a temperature compensation engine, which deliver a user-friendly interface and contribute to the robustness of the new μNMR system (Fig. 9.8 right). The software is programmed with Objective-C using Cocoa and Cocoa Touch frameworks and operates on iOS™ and OS X™. In order to implement a graphical representation of $R_2$ relaxation curve, open-source plotting framework (Core Plot) is cross-linked with Cocoa and Cocoa Touch frameworks. Time domain NMR data are acquired at the negative edge of spin echoes and processed real time. Negative edge-triggered data acquisition reduces the size of
The entire system was redesigned for robust operation in clinics. The NMR probe accepts samples sealed in disposable tubes to prevent probe contamination (left). NMR coils are embedded into the polymer support to open up the entire coil bore, and a thin-walled (thickness ~65 μm) tube is used to minimize loss in filling factor. Block diagram of the NMR electronics for sample excitation and NMR signal reception (middle). The electronics is highly programmable and can be controlled via external terminals. Functional view of NMR software that processes data and shows graphical output of NMR measurements (right).

The data to be transferred and consequently decreases the data acquisition time via USB/Bluetooth connections by more than 10-fold. Transferred NMR data are then processed to obtain $T_2$ relaxation time. The overall signal pattern is displayed, and the data is stored in the terminal device. In addition, wireless communication module is embedded in the software for data logging/sharing over an encrypted wireless network, promoting usability in a remote clinical site.

To ensure stable and reliable $T_2$ relaxation time measurements, a temperature compensation engine has been implemented to run independently from the main thread. This temperature compensation engine transforms NMR data from time domain to frequency domain using fast Fourier transformation (FFT) and detects any changes in the NMR frequency due to environmental temperature fluctuation (described below).
9.5.3.1 Temperature Compensation

The NMR frequency $f_0$ is the parameter that requires most frequent adjustments and affects the measured NMR signals most significantly. The frequency $f_0$ changes as the magnetic field ($B_0$) from the permanent magnet drifts with temperature [58]. For example, with a $1^\circ$C increase in temperature, $B_0$ field from a NdFeB magnet will drop $\sim 0.1$% from its initial value, and $f_0$ will proportionally decrease by 0.1%; when the initial $f_0$ is 20 MHz ($B = 0.47$ T), the frequency change is then $\sim 20$ kHz. Such changes can place down-converted NMR signals near or beyond the low-frequency cutoff in the amplification stage, distorting the measured signal. Commercial benchtop NMR systems address the problem by housing the entire magnet block inside a heated container. This solution, however, significantly undermines the portability of the system due to the use of bulky and power-consuming parts. In the new $\mu$NMR system, we employed a dynamic control approach. Namely, programmable hardware in the NMR electronics and temperature compensation engine in the NMR software are designed to track and compensate for temperature dependency of the system. These implementations ensure optimal measurement settings for reliable and robust performance.

Figure 9.9a and b show the algorithm for temperature compensation. The feedback loop tracks the Larmor frequency $f_0$ and reconfigures the frequency $f$ of NMR excitation. Coarse-tuning mode starts with initial NMR excitation frequency $f_i$ and increases $f$ by $\Delta f$. When the spectral power ($P$) of NMR spin echo reaches a predefined threshold $P_{th}$, a fine-tuning mode takes over to measure the frequency offset $f_d (= |f - f_0|)$. The fine tuning iterates until $f_d$ reaches a target value. The target $f_d$ value is carefully selected to keep down-converted NMR signal within the passband of the low-pass filter. Once the new NMR excitation frequency $f$ has been established, CPMG pulse sequence is used to measure the $T_2$ relaxation time of the sample.

Figure 9.9c demonstrates the effectiveness of the developed temperature compensation method. When $f_0$ was allowed to drift but the RF frequency ($f$) for sample excitation was fixed, $T_2$ values varied up to 200% relative to its starting value with typical fluctuation of room temperature ($\Delta T \sim 2^\circ$C). When the temperature compensation engine was activated, however, $T_2$ variations were significantly reduced to $<1$%. We further tested the system in environmental settings with a wide range of temperature differences ($4$–$50^\circ$C). To determine the measurement accuracy, the linear dependence of $T_2$ on temperature was utilized. Figure 9.9d shows the $T_2$ values of an MNP solution monitored at different temperature. For a given environment setting, the $\mu$NMR system was operated with the temperature tracking activated to compensate for minute temperature variations ($\sim 1^\circ$C). The results show a linear relationship ($R^2 > 98\%$) as theoretically predicted, demonstrating the capacity for reliable $T_2$ measurements in various settings.
Fig. 9.9 Automated temperature tracking in the new portable μNMR system. (a) A flow diagram of the temperature compensation engine. The tracking routine searches for the Larmor frequency $f_0$ at a given environmental setting, using the successive tracking routines (coarse and fine). Once $f_0$ is determined, a full NMR measurement is performed. (b) Temperature tracking algorithm. For a given RF excitation with the frequency $f$, a spin echo signal is measured and transformed into the frequency domain via fast Fourier transform (FFT). The coarse search mode checks whether the amplitude ($P$) of a peak in the spectral power is larger than the predefined value ($P_{TH}$). In the next stage, the fine search mode iteratively tunes $f$ until the frequency offset $f_d$ reaches the optimal offset value ($\approx 3$ kHz). (c) The effectiveness of temperature compensation was evaluated using an MNP solution ($T_2 \approx 100$ ms). When the tracking routine was turned off (square), the $T_2$ variation was $>200\%$ of its initial value. With the tracking routine turned on (circle), the fluctuation was significantly reduced ($<1\%$). (d) The robustness of μNMR measurements across a broad range of temperature ($4–50^\circ$C) was demonstrated. The dotted line indicates a theoretical prediction (Reproduced from [57]. Copyright 2011 RSC Publishing)

9.6 Biological Applications

9.6.1 Cancer Detection and Profiling

Sensitive detection and rapid characterization of tumor cells in minimally processed biological samples will have significant impact on both biomedical research and clinical practice. Using the first-generation DMR device (μNMR-1) [22], DMR
molecular profiling of cancer markers (Her2/neu, EGFR, and EpCAM) on human cells was demonstrated with CLIO nanoparticles directly conjugated to monoclonal antibodies. With the development of the second-generation $\mu$NMR-2 and the highly magnetic MnFe$_2$O$_4$ nanoparticles, subsequent cellular detection sensitivity was remarkably improved to approximately single-cell level, far surpassing the sensitivity of other conventional clinical methods (Fig. 9.10a). Furthermore, a new assay protocol was established that reports the expression level of a specific biomarker and the target cell density.

To independently measure cell numbers by $\mu$NMR, we exploited a phenomenon of low-grade phagocytosis of nontargeted MNPs by tumor cells [59]. When mammalian cells were incubated (15 min at 37°C) with unmodified MNPs (MNP-$\Phi$), linear and cell-number-dependent $R_2$ changes ($\Delta R_2^\phi$) were observed (Fig. 9.10b). Interestingly, these changes were similar across a wide variety of cell types. The results were fitted to Eq. 9.2: $\Delta R_2^\phi = r_2^\phi \cdot n_C$, where $r_2^\phi$ is the cellular relaxivity for MNP-$\Phi$ and $n_C$ is the cell concentration ($N_C / V$). The cellular relaxivities ($r_2^\phi$) were statistically identical ($p > 0.99$) among different tumor cell lines, suggesting that the method may provide a universal measure for estimating $n_C$.

Using Eq. 9.2 and the cell density information (above), the expression level ($\xi$) of a select marker was defined as

$$\xi^{Ab} = \frac{\Delta R_2^{Ab}}{\Delta R_2^\phi} = \frac{r_2^{Ab} \cdot n_C}{r_2^\phi \cdot n_C} = \frac{r_2^{Ab}}{r_2^\phi},$$

(9.4)

where $\Delta R_2^{Ab}$ and $r_2^{Ab}$ are $R_2$ changes and the cellular relaxivity, respectively, with a marker-specific MNP. In this normalized form, $\xi$ now reports the cellular expression level of a targeted marker, providing a way to molecularly profile target cells regardless of cell numbers in a sample. The method was extensively verified by comparing $\mu$NMR measurements to other standard methods (flow cytometry, Western blotting) [22, 27, 28]. In one set of experiments, we measured the expression level of HER2 in breast cancer cell lines (Fig. 9.10c). The measured $\xi^{HER2}$ ($= \Delta R_2^{HER2} / \Delta R_2^\phi$) from $\mu$NMR (requiring $\sim 10^3$ cells) showed good agreement ($R^2 > 98\%$) with both flow cytometry (requiring $\sim 10^5$ cells) and Western blotting (requiring $\sim 10^7$ cells), validating the analytical capability of DMR. Note that DMR detection was much faster ($\sim 15$ min) and performed using $> 10^2$ times fewer cells.

### 9.6.2 Clinical Trial

Through the integration of the complementary DMR and BOND technologies, this chip-based NMR detection platform has been applied in clinical trials of cancer cell profiling [26]. A total of 50 patients with suspected abdominal malignancies were enrolled. Each patient underwent fine-needle aspiration (FNA) using a 22-G needle, followed by routine core biopsies (17-G needle) for conventional analysis. The FNA samples were aliquoted and profiled for 11 predefined cellular markers:
Fig. 9.10 Cellular detection via DMR. (a) The detection limit by DMR is nearly at single cell level (~2 cells), which is superior to the current clinical methods (cytology and histology). (b) The uptake of unmodified MNPs by cells was exploited to estimate the cell population in the samples. The measured $\Delta R_2$ was linearly proportional to the cell concentrations. Importantly, the linear trends were statistically identical in different cell types. (c) The analytical accuracy of the $\mu$NMR was benchmarked against flow cytometry and Western blotting by measuring HER2/neu expression on breast cancer cells (Adopted from [28]. Copyright 2009 National Academy of Sciences, USA)
Fig. 9.11  Expression levels of different biomarkers arranged by patient number. Note the high degree of heterogeneity of marker expression per person. Patients 5, 12, 17, 18, 21, and 42 had benign lesions. x-axis, patient number; y-axis, expression level ($\xi$) of a marker (Reproduced from [26]. Copyright 2011 American Association for the Advancement of Science AAAS)

9 cancer-related markers (EpCAM, MUC-1, HER2, EGFR, B7H3, CK18, Ki-67, p53, Vimentin), CD45 for leukocyte counts, and control (MNP-$\Phi$) to measure the total cell density. A priori selection of cancer markers was based on current practice (e.g., EpCAM, CK18) [60, 61] or on reports of clinically relevant overexpression [62–64]. On average, 3,850 cells were obtained per patient via FNA. Of these, approximately one-third were CD45-positive leukocytes (mean: 1,273 cells), and the remaining were nonleukocytic, primarily tumor cells as determined by flow cytometry. The aliquots (containing $\sim$350 cells) were labeled using the BOND method and measured by the clinical $\mu$NMR system. The $R_2$ data with CD45 was used to account for the contribution by leukocytes; for each cancer marker, its cellular expression level ($\xi$) was obtained based on Eq. 9.3.

Figure 9.11 shows the profiling results of 9 cancer markers for 50 patients. When plotted for each patient, the expression level of the markers showed considerable
Table 9.2  Sensitivity, specificity, and accuracy with DMR

<table>
<thead>
<tr>
<th>Number</th>
<th>Marker</th>
<th>$\xi$ value</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single</td>
<td>MUC1</td>
<td>0.25</td>
<td>66</td>
<td>83</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>EGFR</td>
<td>0.20</td>
<td>64</td>
<td>83</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>B7H3</td>
<td>0.11</td>
<td>68</td>
<td>67</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>HER2</td>
<td>0.24</td>
<td>64</td>
<td>100</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Ki-67</td>
<td>0.10</td>
<td>68</td>
<td>67</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>EpCAM</td>
<td>0.21</td>
<td>59</td>
<td>67</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Vimentin</td>
<td>0.08</td>
<td>59</td>
<td>67</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>CK18</td>
<td>0.06</td>
<td>73</td>
<td>50</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>p53</td>
<td>0.38</td>
<td>41</td>
<td>83</td>
<td>46</td>
</tr>
<tr>
<td>Dual</td>
<td>EpCam + CK18</td>
<td>0.27</td>
<td>84</td>
<td>50</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>(unweighted)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triple</td>
<td>MUC1 + HER2 + EGFR</td>
<td>1.23</td>
<td>95</td>
<td>67</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>(weighted)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quad</td>
<td>MUC1 + HER2 + EGFR</td>
<td>1.60</td>
<td>100</td>
<td>67</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>+ EpCAM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(weighted)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Cutoff $\xi$ value for identification of malignancy. The value was determined from the point on a receiver operating characteristic curve that has the minimal distance between the 0% false negative and the 100% true positive.

heterogeneity across samples, which reconfirmed the importance of multiple-marker screening in cancer detection. Indeed, when analyzed for the diagnostic sensitivity, specificity, and accuracy (Table 9.2), a single marker showed accuracy of <70%. The highest accuracy for cancer diagnosis with our cohort was obtained with a quadruple marker combination (MUC1 + EGFR + HER2 + EpCAM, 96% accuracy), followed closely by weighted triple markers (MUC1 + EGFR + HER2, 92% accuracy). Interestingly, the EpCAM and CK18 combination, which is routinely used in detecting circulating tumor cells in peripheral blood, achieved an overall diagnostic accuracy of 80%, wherein high sensitivity (84%) was offset by low specificity (50%).

The DMR results were also compared to those by standard-of-care methods. Conventional cytology on FNA specimen was performed in 49 of 50 cases and was diagnostic in 36 cases with 11 misdiagnoses (accuracy 74%). Conventional histology on all 50 biopsy samples correctly diagnosed 37 cases out of 45 diagnostic samples (accuracy 84%), with the remaining results nondiagnostic (5 cases). DMR consistently outperformed the other methods, with 2 misdiagnoses in all 50 samples (accuracy up to 96%). Also note that DMR permitted fast detection (in less than 60 min for each patient), whereas the mean clinical turnaround time (from sample submission to final report) was 3 days for cytology (1–8 days) and 4 days for surgical pathology (1–11 days).
9.7 Conclusions and Outlook

As a novel biosensing platform, DMR offers many synergistic advantages over traditional approaches, such as high detection sensitivity, multiplexed capability, rapid measurement, and small sample volume requirement with minimal sample processing. Indeed, DMR thrives through the complement of several cutting-edge technologies, namely, magnetic nanomaterials, bioconjugation chemistry, and microfabrication. With new developments such as chip-based μNMR devices, optimized magnetic nanoparticles, and advanced labeling techniques, the DMR technology has proven itself as a robust and sensitive approach for quantitative and molecular analyses for biomedical research. Moreover, our recent clinical trial has confirmed that the DMR is capable of detecting and molecularly profiling cells with minimal false negatives. Its remarkable performance and potential impact on clinical disease management would no doubt accelerate the advance of personalized treatment by providing valuable information on molecular signature of individual patients.

We further envision broader application of the DMR in global healthcare. The DMR technology does not require extensive sample purification and can be packaged as a portable device. The system thus is well suited for rapid and point-of-care (POC) testing, especially in resource-limited primary clinics where majority of diagnoses are made based on physical symptoms only. Indeed, we plan to evaluate the developed system for TB (tuberculosis) detection in fields. Successful completion of this research will be a cornerstone for realizing POC technology for TB detection, which will bring significant societal benefits worldwide.

Acknowledgements We thank T. J. Yoon, J. H. Chung, J. B. Haun, and N. Sergeyev for assistance with experiments and R. M. Westervelt (Harvard) and C. Castro (MGH) for many helpful discussions. This work was supported in part by the following NIH grants: U54CA151884, R01EB010011, R01EB004626, P01CA069246, P50CA86355, U01CA141556, U24CA092782, and R21CA14122. H. Shao acknowledges financial support from the B.S.-Ph.D. National Science Scholarship awarded by the Agency for Science, Technology and Research, Singapore.

References


45. N. Fauconnier et al., Thiolation of maghemite nanoparticles by dimercaptosuccinic acid. J. Colloid Interface Sci. 194, 427–433 (1997)


Index

AC, 28
Activated clotting time, 17
Affinity-based techniques, 6–7
AFM. See Atomic force microscopy (AFM)
ALD. See Atomic layer deposition (ALD)
Anisotropic magnetoresistance, 155
Antibody, 193
Apoptosis, 131, 134
Atomic force microscopy (AFM), 132
Atomic layer deposition (ALD), 104

Back propagation, 79
Bandpass filter, 170
Base length, 138
3 µm Bead, 75
7 µm Bead, 75
10 µm Bead, 75
Biochemical reactions, 39
Biocompatibility, 201
Biocompatible, 204
Biomarkers, 154
Biomolecular sensing, 177
Bioorthogonal nanoparticle detection (BOND), 206
Bioorthogonal targeting, 199
Bioparticles, 61
Biosensor, 177
BOND. See Bioorthogonal nanoparticle detection (BOND)
Burden of disease, 5

Cancer cells, 193
Cancer markers, 215
Capacitance, 26
Capillary systems, 10

Capture antibody, 162
CD4+, 49, 56
CD45, 217
Cell phone, 80
Cells, 6
  fusion, 37
  sorting, 6
Centrifugal-based platforms, 10
Cepheid’s GeneXpert, 13
Charge-pumping, 114
Chemiluminescence, 12
1-Chip NMR system, 179
Chips, 29
Clausius–Mossotti, 26
Clinical trial, 215
CLIO. See Cross-linked iron oxide (CLIO)
Clock, 183
CMOS. See Complementary metal-oxide semiconductor (CMOS)
Coil, 178
Complementary metal-oxide semiconductor (CMOS), 28
Controller, 185
Conventional microscope, 75
Correlation, 52
Correlation analysis, 52
Cross-linked iron oxide (CLIO), 201
Culture assay, 59
Cytometry, 72, 75, 76, 93

Daktari CD4 system, 12
Dancer, 32
Debye length, 133, 139, 142
Deform, 33
Dental plaque, 127
DEP. See Dielectrophoresis (DEP)

D. Issadore and R.M. Westervelt (eds.), Point-of-Care Diagnostics on a Chip, Biological and Medical Physics, Biomedical Engineering,
DOI 10.1007/978-3-642-29268-2, © Springer-Verlag Berlin Heidelberg 2013
DEP/magnetic chip, 35
Detection antibody, 162
Developed world, 4
Developing world, 4
Diagnostic magnetic resonance (DMR), 198
Dielectrophoresis (DEP), 25, 27
Diels-Alder cycloaddition, 206
Disease screening, 180
Disposable, 13
DMR. See Diagnostic magnetic resonance (DMR)
DNA, 135
DNA-microchips, 61
Double gate FET, 109
Drug-resistant tuberculosis, 13
Duty cycle, 184

EGFR, 215
Electrochemical, 14
Electrochemical detection, 8
Electrofusion, 28
Electrolytes, 7
Electromagnet, 167
Electroporated, 36
Electroporation, 28
Electrowetting, 40
Enzyme-linked immunosorbent assay (ELISA), 8, 61
EpCAM, 215
Equalization of red and green pixels, 79
Equalized output, 79
Estimation of blue pixels, 79
Excitation, 181

$1/f$, 114, 116
Fast fourier transform (FFT), 52
Fe-core MNPs, 202
Ferrite-based MNPs, 201
FET. See Field-effect transistor (FET)
FFT. See Fast Fourier transform (FFT)
Field-effect transistor (FET), 126
Fine-needle aspiration, 215
Flicker, 116
Flicker noise, 170
Flow cytometry, 47, 60
Fluorescein, 35
Fluorescence, 14
Fourier domain, 74
Frames, 32

Giant magnetoresistive (GMR), 155
Giardia lamblia, 75

Global healthcare, 219
GMR. See Giant magnetoresistive (GMR)
Gold (Au), 129
Ground electricity, 4

Handheld, 58
HandyLab, 13
Hela cells, 130
Helmholtz electromagnet, 168
Hematology, 7, 17
HemoCue WBC system, 11
Hepatitis C virus (HCV), 172
Her2/neu, 215
Heterodyne, 186
High-income, point-of-care, 5
HIV/AIDS. See Human immunodeficiency virus (HIV/AIDS)
Holographic reconstruction, 79
Holography, 74, 78, 81, 85, 90, 92, 93
Human immunodeficiency virus (HIV/AIDS), 5, 56, 171
Hybrid integrated circuits/microfluidic chips, 24
Hybridization, 139
Hybridoma, 37
Hydrogen, 181

IC. See Integrated circuits (ICs)
Immunooassay, 8
Impedance, 11, 98, 99
In vitro, 149
Infectious diseases, 5
Input-referred noise, 188
Integrated circuits (ICs), 24, 211
Integration, 180
Interface trap charge, 113
Ion sensitive field effect transistor (IS-FET), 126, 127
Isothermal, 14
iSTAT, 17
Iterative refinement of blue pixels + phase recovery, 79

Label-free, 98, 100, 103, 140
LabNow, 12
Lateral flow, 14
Lensfree, 72, 73, 75, 78, 79, 82–84, 87, 89, 93
Lensfree images, 75
Leukocytes, 217
Lithography, 167
Local oscillator (LO), 183
Index

Low cost, 3, 5
Low-cost, portable detector, 5
Low-income, point-of-care, 5
Lysing, 49, 56

Magnet, 178
Magnetic, 30
biosensing, 154
flux guides, 169
particles, 191
sensing, 198
susceptibilities, 198
Magnetic nanoparticles (MNP), 154, 198
Magnetic tunnel junction (MTJ), 158
Magnetoresistance, 155
Magnetoresistance (MR) ratios, 155
Malaria, 5, 84, 85
Mask, 79
Matching networks, 181, 187
Measured amplitude, 75
Measured hologram (Bayer pattern image), 79
MEMS. See Microelectromechanical systems (MEMS)
Merging, 37
Metabolism, 146
Microcontroller, 211
Microelectromechanical systems (MEMS), 126
Microfluidic, 23
Microwave dielectric heating, 39
Miniature NMR systems, 179
Miniaturization, 195
Miniaturized nuclear magnetic resonance (μNMR), 206
Mitochondrion, 149
Mixing, 40
MNP. See Magnetic nanoparticles (MNP)
Monoclonal antibodies, 162
Moore’s Law, 125
MTJ. See Magnetic tunnel junction (MTJ)
Multiplexed, 56
Multiplexed SPR, 61

NA. See Numerical aperture (NA)
Nanogap, 98, 114, 117
Nanoparticles, 15
Nanowire, 101, 103, 110
Neutrophil, 75
NMR. See Nuclear magnetic resonance (NMR)
Noise figure, 187
Noise matching, 187
Non-communicable diseases, 5

Non-white noise, 53
Nuclear magnetic resonance (NMR), 177, 198
Nucleic acids, 6
extraction, 12
testing, 7
Numerical aperture (NA), 80, 86
Nyquist, 99

Object support, 79
OFM. See Opto-fluidic microscopy (OFM)
Oligonucleotide, 139
On-chip, 5
Optical detection, 17
Opto-fluidic, 90, 91
Opto-fluidic imaging, 81
Opto-fluidic microscopy (OFM), 72
Output power, 189

Palm NMR system, 179
Paper-based microfluidic, 10
Parasites, 75, 80
Passive, 5
Passive amplification, 187
Pathogen, 59
PCB. See Printed circuit board (PCB)
PCR. See Polymerase chain reaction (PCR)
Permittivities, 27
pH measurement, 127
Physiological fluids, 24
Pima CD4 test, 12
Platelets, 75
Plug-based reagent delivery, 10
Pneumatic, capillary, 5
Point-of-care (POC), 3, 97
diagnostics, 24
testing, 219
Polarizable, 25
Polyclonal antibody, 162
Polymerase chain reaction (PCR), 59
Portable, 3
detector, 5
NMR system, 178
Positioning, 30
Power tuning, 184
Predicted hologram, 79
Printed circuit board (PCB), 168
Proteins, 6, 192
Proton filter, 183
Protons, 181
Proximity based detection, 164
Proximity based sensors, 158
Pulse sequence, 190
Pulse sequence generator, 185

Quantum dots, 12

Radiofrequency (RF), 207
Rare biomarkers, 198
RBC. See Red blood cells (RBC)
Receiver, 181
Reconstructed lensfree image, 79
Recovered hologram, 79
Red blood cells (RBC), 75
Relaxivity, 200
Resonance matching, 189
Reverse phase, 163
Rheological, 35
Rough handling, 5

Sample preparation, 12
Sandwich assay, 162
Sandwich immunoassay, 143
Self-contained, 5
Sensitivity, 190
Shadow mask, 59, 63
Sheath-flow, 48
Signal amplification, 9
Signal-to-noise ratio, 180
Silicon microdevices, 13
Silver reduction, 9
Single nucleotide polymorphism (SNP), 140
Size- and density-based techniques, 6
Small molecules, 6
SNP. See Single nucleotide polymorphism (SNP)
Spatially modulated fluorescence emission, 51
Spin-mass, 190

SQUID. See Superconducting quantum interference device (SQUID)
Substrate current, 117
Superconducting quantum interference device (SQUID), 155
Superparamagnetic, 201, 204
Synthetic antiferromagnet, 157

$T_2$, 190, 199
TB. See Tuberculosis (TB)
Temperature sensors, 39
Tetrazine, 206
Thermistors, 39
Thiol groups, 129
TMR. See Tunneling magnetoresistance (TMR)
Tomograms, 89
Transceiver, 178
Trans-cyclooctene, 206
Transmitter, 181
Transparent electrode, 129
Trapped, 31
Trypan, 36
Tuberculosis (TB), 5, 219
Tunneling magnetoresistance (TMR), 158

Valves, 9
Very low cost, 5
Vesicles, 34
Video, 33
Wash-free assay, 164
Water, 190
Wires, 30
Yeast cells, 31