2010-08-09

Thin Film Microfluidic and Nanofluidic Devices

Mark Noble Hamblin

Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Electrical and Computer Engineering Commons

BYU ScholarsArchive Citation

Hamblin, Mark Noble, "Thin Film Microfluidic and Nanofluidic Devices" (2010). All Theses and Dissertations. 2281.

https://scholarsarchive.byu.edu/etd/2281

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Thin Film Microfluidic and Nanofluidic Devices

Mark N. Hamblin

A dissertation submitted to the faculty of Brigham Young University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Aaron R. Hawkins, Chair
Stephen M. Schultz
Gregory P. Nordin
Milton L. Lee
Adam T. Woolley

Department of Electrical and Computer Engineering
Brigham Young University
December 2010

Copyright © 2010 Mark N. Hamblin
All Rights Reserved
ABSTRACT

Thin Film Microfluidic and Nanofluidic Devices

Mark N. Hamblin
Department of Electrical and Computer Engineering
Doctor of Philosophy

Lab-on-a-chip devices, also known as micro total analysis systems (μTAS), are implementations of chemical analysis systems on microchips. These systems can be fabricated using standard thin film processing techniques. Microfluidic and nanofluidic channels are fabricated in this work through sacrificial etching. Microchannels are fabricated utilizing cores made from AZ3330 and SU8 photoresist. Multi-channel electroosmotic (EO) pumps are evaluated and the accompanying channel zeta potentials are calculated. Capillary flow is studied as an effective filling mechanism for nanochannels. Experimental departure from the Washburn model is considered, where capillary flow rates lie within 10% to 70% of theoretical values. Nanochannels are fabricated utilizing cores made from aluminum, germanium, and chromium. Nanochannels are made with 5 μm thick top layers of oxide to prevent dynamic channel deformation. Nanochannel separation schemes are considered, including Ogston sieving, entropic trapping, reptation, electrostatic sieving, and immutable trapping. Immutable trapping is studied through dual-segment nanochannels that capture analytes that are too large to pass from one channel into a second, smaller channel. Polymer nanoparticles, Herpes simplex virus type 1 capsids, and hepatitis B virus capsids are trapped and detected. The signal-to-noise ratio of the fluorescently-detected signal is shown to be greater than 3 for all analyte concentrations considered.

Keywords: thin film, chemical analysis, microfluidics, nanofluidics, electroosmotic pump, nanosieve, sacrificial etching, capillary action, lab-on-a-chip, μTAS
ACKNOWLEDGMENTS

Above all that everyone has done for me, I want to thank my amazing wife, Andrea. She has always encouraged me to pursue my dreams, and continues to walk every challenging and exciting step of life with me. I would like to thank my advisor, Dr. Aaron Hawkins, for providing me the opportunity to research and discover what I am capable of. His continual encouragement and insight has been invaluable, and I can never thank him enough. I also need to thank Dr. Milton Lee, Dr. Adam Woolley, Dr. Daniel Maynes, Dr. Stephen Schultz, Dr. Gregory Nordin, and Dr. Dennis Tolley for their continued guidance and expertise.

I would like to thank my parents for their support. They have always believed that I could accomplish anything, and they have provided me with the foundation of confidence that I needed to pursue my dreams. I would also like to thank my mother- and father-in-law, who have been second only to my parents in their belief that I can accomplish anything I set my mind to. This work would not have been possible without the work and help of such excellent colleagues like John Edwards, Bridget Peeni, Megan Larsen, Jie Xuan, Dallin Murray, Lynnell Wright, Mallory Phillips, Megan Bracken, and John Stout. I would be remiss if I did not mention the great help that other fellow researchers have provided me over the course of the last several years, including Tao Shang, Evan Lunt, John Hulbert, Brian Phillips, Matt Holmes, Michael Johnson and, Josh Beutler. I would also like to thank Dr. John Gardner and Michael Standing for all of their help with fluorescence and scanning electron imaging.
Table of Contents

List of Tables xi
List of Figures xiii

1 Introduction 1
   1.1 Lab-on-a-Chip/µTAS ........................................ 1

2 Analytical Separations 5
   2.1 Capillary Liquid Chromatography .......................... 5
   2.2 Capillary Electrophoresis .................................. 6
   2.3 Microchip Separations ...................................... 9
       2.3.1 Microfluidic Separations .............................. 9
       2.3.2 Nanofluidic Separations ............................. 11
   2.4 Fluorescence Detection .................................... 12
   2.5 Size Separation ........................................... 14
       2.5.1 Size Exclusion Chromatography ..................... 14
       2.5.2 Capillary Gel Electrophoresis ....................... 16
       2.5.3 Steric Field-Flow Fractionation ..................... 16

3 Microfluidic/Nanofluidic Channel Fabrication 19
   3.1 Bonding .................................................. 20
   3.2 Shadow Deposition ....................................... 21
3.3 Imprint Molding ......................................................... 23
3.4 Sacrificial Etching ....................................................... 24
  3.4.1 Fabrication ......................................................... 25
  3.4.2 Channel Characterization ........................................ 30

4 Pumping Mechanisms ................................................. 35
  4.1 Pressurized Injection .................................................. 35
    4.1.1 Derivation ....................................................... 36
  4.2 Electroosmosis ....................................................... 40
    4.2.1 Physics of Electroosmosis .................................... 41
    4.2.2 Derivation ....................................................... 42
    4.2.3 Electroosmotic Pump Theory ................................ 47
    4.2.4 Experimental Results ......................................... 52
  4.3 Capillarity .......................................................... 64
    4.3.1 Derivation ....................................................... 65
    4.3.2 Experimental Results ......................................... 67

5 Nanochannel-Based Separations ...................................... 79
  5.1 Ogston Sieving ....................................................... 81
  5.2 Entropic Trapping .................................................... 83
  5.3 Reptation .......................................................... 84
  5.4 Electrostatic Sieving ................................................. 85
  5.5 Immutable Trapping ................................................. 86
    5.5.1 Dual-Segment Device Fabrication ............................ 87
    5.5.2 Dual-Height Trapping Operation ............................... 89
    5.5.3 Molecular Trapping ............................................ 91
## List of Tables

3.1 Sacrificial Materials ................................................. 30

4.1 Relating Electroosmotic Velocity ................................. 46

4.2 Nanochannel Bowing ................................................ 77
# List of Figures

1.1  Lab-on-a-chip device. .................................................. 2

2.1  Liquid chromatography. .............................................. 7

2.2  Capillary electrophoresis .......................................... 10

2.3  Microchip capillary liquid chromatography. ..................... 11

2.4  Microchip capillary electrophoresis. ............................ 12

2.5  Fluorescence detection ............................................. 13

2.6  Size exclusion chromatography. .................................. 15

2.7  Field flow fractionation. .......................................... 18

3.1  Fabrication flow chart illustrating microchannel fabrication using bonding. . 22

3.2  Microfluidic or nanofluidic channel creation using shadow deposition. .... 23

3.3  Microfluidic or nanofluidic channel creation using imprint molding. ....... 25

3.4  Single-layer sacrificial etching. .................................. 27

3.5  Double-layer sacrificial etching. .................................. 29

3.6  Etching of sacrificial aluminum cores. .......................... 30

3.7  Profiles of sacrificial cores. ...................................... 31

3.8  Scanning electron micrographs of rounded-core sacrificially-etched channels. 33

3.9  Scanning electron micrograph of the cross-section of a 145-nm tall nanofluidic channel. .............................................. 34

4.1  Poiseuille flow profile. ............................................ 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.24</td>
<td>Transient capillary filling of nanochannels.</td>
<td>73</td>
</tr>
<tr>
<td>4.25</td>
<td>Capillary flow speeds, $x^2/t$, versus height for various nanofluidic channels.</td>
<td>74</td>
</tr>
<tr>
<td>4.26</td>
<td>Calculated dynamic contact angle versus channel height for various channels.</td>
<td>75</td>
</tr>
<tr>
<td>4.27</td>
<td>Theoretical critical channel width versus channel height.</td>
<td>77</td>
</tr>
<tr>
<td>5.1</td>
<td>Nanofluidic separation mechanisms.</td>
<td>81</td>
</tr>
<tr>
<td>5.2</td>
<td>Two-segment immutable trapping fabrication steps.</td>
<td>89</td>
</tr>
<tr>
<td>5.3</td>
<td>Photos of completed two-segment immutable trapping devices.</td>
<td>90</td>
</tr>
<tr>
<td>5.4</td>
<td>Dual-segment immutable trapping system setup.</td>
<td>91</td>
</tr>
<tr>
<td>5.5</td>
<td>Dual-segment immutable trapping.</td>
<td>92</td>
</tr>
<tr>
<td>5.6</td>
<td>Enveloped virion.</td>
<td>93</td>
</tr>
<tr>
<td>5.7</td>
<td>Polymer beads and viral capsids used for testing two-segment immutable trapping.</td>
<td>94</td>
</tr>
<tr>
<td>5.8</td>
<td>Time-lapse sequence of the fluorescent signal of trapped HSV-1 capsids and HBV capsids.</td>
<td>95</td>
</tr>
<tr>
<td>5.9</td>
<td>Fluorescent signal intensity normalized to the maximum signal versus time for trapped analytes.</td>
<td>97</td>
</tr>
<tr>
<td>5.10</td>
<td>Signal-to-noise ratio (SNR) of trapped 120-nm and 30-nm beads.</td>
<td>98</td>
</tr>
<tr>
<td>5.11</td>
<td>Lipoprotein classes.</td>
<td>99</td>
</tr>
<tr>
<td>5.12</td>
<td>Trapped lipoproteins in a dual-segment immutable trap.</td>
<td>99</td>
</tr>
<tr>
<td>5.13</td>
<td>Immutable trapping methods.</td>
<td>101</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

We live in an amazing time where high technology reigns. It influences and affects every facet of our day-to-day lives. We drive automobiles, fly in airplanes, connect instantaneously around the globe through the Internet, and carry cell phones. One powerful and extremely important application of modern technology is our ability to perform clinical and chemical analyses to better our quality of life. We have extensive laboratories to perform studies which provide a wealth of useful information in many analyses. But as our technology improves, we also look for a way to improve the analytical laboratory as we know it today. This is possible through a chemical analysis revolution synonymous with the microelectronics revolution. As these technologies expand, we will see the cost and time of clinical and chemical analyses drop substantially, ushering in a new era of affordable health care, efficient environmental monitoring, and rapid chemical threat assessment.

1.1 Lab-on-a-Chip/µTAS

One of the main functions of laboratory analyses is to perform a variety of tests to determine the types and quantities of molecules and particles present in a solution. Many specialized laboratories exist to perform specific tests, however, all standard laboratories suffer from the same drawbacks. One of these drawbacks is the high cost associated with specialized analysis. Analysis times may also take several days or weeks. Such laboratories also require very costly specialized equipment with a very highly-skilled workforce to operate the equipment. Furthermore, laboratories are stationary, and samples often have to be transported across country. Any improvements that could be leveraged against these drawbacks would be extremely beneficial.
A new breed of analysis systems is coming to life, falling under the heading of *lab-on-a-chip* or *micro total analysis systems* (μTAS). As the name *lab-on-a-chip* implies, the underlying concept is to take all of the functionality of an analytical laboratory and move it onto a single microchip, in much the same way that classical electronics were moved to microchips. The name *micro total analysis systems* implies the ability of the chip to perform a wide variety of analysis techniques. As analytical systems are moved to microsystems, they will result in lower cost, faster analysis times, simple usage, point-of-care capability, and minute sample volumes. A sample lab-on-a-chip device is demonstrated in Figure 1.1.

![Figure 1.1: Lab-on-a-chip device. Adapted from http://commons.wikimedia.org.](image)

The point of this work is to expound upon chemical analysis systems that incorporate thin film microprocessing techniques. Chapter 2 is an overview of analytical separation techniques that can be applied to lab-on-a-chip systems. Chapter 3 provides a review of fabrication methods for producing microchannels and nanochannels. This chapter also delves into greater detail concerning sacrificial etching, the fabrication technique employed for all devices fabricated in this work. Chapter 4 explores liquid pumping techniques employed in microfluidic and nanofluidic systems. It contains background information on several flow-
generating forces, as well as presents original research related to these techniques. Chapter 5 focuses on separation techniques that are unique to nanochannels, and treats immutable trapping in great detail. Immutable trapping is a novel separation technique developed and explored in this work. Chapter 6 provides a brief conclusion of the results described in this work.
Chapter 2

Analytical Separations

Analytical chemistry is the field that deals with separating, identifying, and quantifying a substance. Analytical chemistry is commonly used in areas such as clinical analysis, genome sequencing, and environmental monitoring. Analytical separation science is one of the major areas of analytical chemistry, and is usually essential before identification and quantitative analysis. The purpose of this chapter is to introduce some of the general features of separation science, especially microseparations and microfabricated and nanofabricated devices. The first and second sections treat capillary liquid chromatography and capillary electrophoresis. The third section treats microchip separations for microfluidic and nanofluidic devices. The fourth section of this chapter addresses fluorescence detection, which is the detection method used for experiments in this work. The final section deals with size separation, since the main nanofabrication technique described in this dissertation involves size separation.

2.1 Capillary Liquid Chromatography

For capillary liquid chromatography, separation is based on differential solubility of an analyte between the stationary phase and the mobile phase. In other words, an analyte will have a preference due to its ability to solvate in one phase over the other. Depending on which way this preference goes, the analyte will either move into the mobile phase or into the stationary phase. Analytes that migrate into the stationary phase will be retained and elute later, while analytes that stay in the mobile phase will elute earlier. Thus, detection order goes from analytes with low differential solubility to analytes with high differential solubility. In liquid chromatography, the stationary phase and the mobile phase are liquids.
In Figure 2.1, the stationary phase is represented by the yellow film along the channel walls. As the sample is introduced in (a), analytes with a high differential solubility will partition from the mobile phase to the stationary phase and be retained there. As the mobile phase flows through in (b), any analytes which have greater solubility in the mobile phase will leave the stationary phase and elute through the channel. In (c) and (d), the differential solubility of the retained analytes decreases as the mobile phase is changed over time, and this causes all of the retained analytes to eventually partition back to the mobile phase and elute. An example chromatogram is shown on the bottom of (c) and (d), corresponding to the blue analytes as the first peak and the orange analytes as the last peak.

2.2 Capillary Electrophoresis

The word electrophoresis can be broken into two components: *electro*, meaning of or having to do with electric fields, and *-phoresis*, meaning movement in a medium. Thus, electrophoresis is the movement of charged analytes in a medium due to electric fields. The speed with which an analyte will migrate in the presence of an electric field is called its electrophoretic mobility, and depends directly on the charge and size of the analyte, and on the viscosity of the solution.

To understand how electrophoresis works, first consider a charged particle suspended in solution. As an electric field is applied, it will exert a force, \( F_e \), on the analyte. This force is equal to

\[
F_e = QE,
\]

where \( Q \) is the charge of the particle and \( E \) is the strength of the applied electric field.

While this particle is moved by the electric field, a resistive, frictional force opposes this movement. George Gabriel Stokes showed that this force, for a spherical object, can be shown as

\[
F_s = 6\pi \eta ru_e,
\]

where \( F_s \) is the resistive force of the liquid, \( r \) is the Stokes radius of the sphere, \( \eta \) is the dynamic viscosity of the fluid, and \( u_e \) is the velocity of the analyte. Now if Equations (2.1)
Figure 2.1: Liquid chromatography. (a) A sample is introduced, and high differential solubility causes solutes to partition into the stationary phase. (b) As the mobile phase flows, any solutes with high solubility with respect to the mobile phase partition out of the stationary phase. (c) The differential solubility increases, resulting in more analyte partitioning. (d) Eventually the solubility of the mobile phase exceeds the solubility of the stationary phase for all analytes, and all molecules elute to be detected.
and (2.2) are set equal to each other, we obtain

\[ u_{ep} = \frac{QE}{6\pi\eta r}. \]  

(2.3)

This is the electrophoretic velocity of an analyte. We can define an analyte in terms of its electrophoretic mobility, which is defined as

\[ \mu_{ep} = \frac{Q}{6\pi\eta r}. \]  

(2.4)

This leads to redefining Equation (2.3) as

\[ u_{ep} = \mu_{ep}E. \]  

(2.5)

One more thing to note about Equation (2.3) is defining the Stokes radius, \( r \). The Stokes radius is often stated to represent the hydrated radius of a molecule, although more accurately it represents a sphere diffusing at the same rate as the respective molecule. The Stokes radius can be defined as

\[ r = \frac{k_BT}{6\pi\eta D}, \]  

(2.6)

where \( k_B \) is Boltzmann’s constant, \( T \) is temperature, and \( D \) is the diffusion coefficient of the molecule. In other words, if the diffusion constant of the molecule is known, it is possible to model that molecule using the Stokes radius.

From Equation (2.5), it is clear that analytes with a high electrophoretic mobility will elute faster than those with a low electrophoretic mobility and, thus, separation is achieved. Another interesting effect to note is that the same mechanism that causes these analytes to separate also causes them to move down the length of the channel toward the detector. However, in most cases, migration limited only to electrophoresis can often times be much slower than is desired for most analytical separation systems, such that another mechanism, such as electroosmosis, is usually coupled with electrophoresis (electroosmosis will be discussed in detail in Chapter 4). Another point to note about electrophoresis is
that it can only separate charged molecules. Neutral analytes will be unaffected by this separation method.

Figure 2.2 demonstrates how electrophoresis operates. In (a), a sample of analytes with charges of the same sign is introduced into the column. In (b), an electric field is established by placing an anode and cathode at opposite ends of the column. For most cases, the electrodes will come into direct contact with the solution, and great care must be taken to ensure bubbles do not enter the capillary. In (c), the separation of these analytes becomes more pronounced, until by (d) each of the analytes is separated into discrete bands. Each of these detected bands is represented by the numbered peaks in the electropherogram at the bottom of Figure 2.2(d).

### 2.3 Microchip Separations

Various analytical systems have been miniaturized to the point where they are housed on a microchip. Such microchip separation devices rely on microfabrication techniques and materials. As such, any microchip separation device must be designed to accommodate this approach. Microchip devices can incorporate microfluidic channels and/or nanofluidic channels.

#### 2.3.1 Microfluidic Separations

Two separation methods mentioned previously that can be implemented with microfluidic channels include capillary liquid chromatography and capillary electrophoresis. Figure 2.3 illustrates an implementation of a microfluidic capillary liquid chromatography system. In (a), a microfluidic device is constructed that utilizes a liquid stationary phase. One common implementation in microfluidic liquid chromatography systems is to pattern a network of micropillars within the channels that acts as a support for the stationary phase. A buffer is then introduced into the device. In (b), a sample plug is injected into the system and is pumped toward the waste reservoir. In (c), as the sample interacts with the stationary phase, separation occurs until separated bands elute past the detector.

Capillary electrophoresis is another implementation in microfluidic devices. A common structure that utilizes capillary electrophoresis is shown in Figure 2.4. In (a), the device
Figure 2.2: Capillary electrophoresis. (a) A sample of ions is introduced into the column. (b) An electric field is applied along the length of the column, and ionic species begin to migrate according to their respective charges and mobilities (assume that, in this illustration, larger ions correspond to greater charge and mobility). (c) As the ions move according to their electrophoretic mobility, they begin to separate until (d) all of the analytes have been isolated in their respective bands.
Figure 2.3: Microchip capillary liquid chromatography. (a) A microfluidic network is initially filled with a buffer. (b) A sample plug is introduced and is pumped through the channel. (c) After some time, individual bands begin to separate within the column and are recorded by the detector.

is initially filled with a buffer. In (b), a sample is introduced into the left reservoir, and a voltage is applied between the left and right reservoirs to electrokinetically transport the sample across this segment. Because of the offset T-shaped junction in this device, a small sample plug will be present in the long channel segment that runs from top to bottom. In (c), a voltage is applied across the top and bottom reservoirs to move the sample plug toward the bottom waste reservoir. As the sample elutes, electrophoresis will cause charged analytes to separate into bands that will be monitored by the detector.

2.3.2 Nanofluidic Separations

Microchip separation devices can also be created with nanofluidic channels. However, there are few separation mechanisms that operate effectively on this scale. For example, capillary liquid chromatography is limited in scope with nanofluidic channels because most capillary liquid chromatography systems utilize pressurized pumping. Because of the small cross-sectional areas of nanofluidic channels, extremely high pumping pressures would be
Figure 2.4: Microchip capillary electrophoresis. (a) A typical T-channel initially filled with a buffer. (b) A sample is injected in the left reservoir, followed by a positive voltage dropped across the left and right reservoirs, causing the sample to flow from left to right. (c) A positive voltage is dropped across the top and bottom reservoir, causing a plug of sample to migrate toward the bottom reservoir. (d) As the sample migrates, species will separate electrophoretically to form bands that will be measured by the detector.

required. Capillary electrophoresis can be used with nanofluidic channels, but typically suffers from reduced electrokinetic transport. To date, the most common separation method used with nanofluidic devices is analogous to capillary gel electrophoresis (which will be discussed in Section 2.5).

2.4 Fluorescence Detection

Fluorescence detection is a method that is commonly used in miniaturized separation systems. Figure 2.5 illustrates how fluorescence detection operates. In most applications, a wideband light source will be used – typically an ultraviolet lamp. In this case, the emitted light will be narrowband filtered before it reaches the sample. This is done to ensure that the precise fluorescence excitation wavelength is used to target the desired fluorophores in the
sample. As this light passes through the transparent capillary window, photons will interact with the analytes. If the analytes contain fluorophores with excitation wavelengths matching those of the incident light, then they will fluoresce. Fluorescence is the phenomenon where a molecule will absorb electromagnetic radiation at one wavelength and emit at another wavelength. This emitted wavelength will then provide a fingerprint for the presence of the analyte. In the figure, the fluorescently emitted light will pass to the photodetector. To isolate the emitted wavelength, every other wavelength is filtered out. This way, only the signal generated by fluorescence will reach the detector, and indicate the presence of separated analyte bands.

**Figure 2.5:** Fluorescence detection. A narrowband light source at the wavelength of fluorescence excitation transmits through a transparent section of the column. Lasers and LEDs exhibit narrow bandwidths, but another method is to filter a powerful light source like a UV lamp. Fluorophores excite and emit light at a different wavelength than the excitation wavelength, such that a filter that blocks the excitation wavelength will allow all fluorescent light to pass to the photodetector. Each fluorescent band will appear as a peak in a chromatogram.

Fluorescence detection is a very powerful method, however the excitation and emission wavelengths for the sample must be known ahead of time. Furthermore, analytes that do not fluoresce cannot be detected using this method. It is possible to detect non-fluorescing molecules by using indirect fluorescence. In this scheme, the mobile phase consists of fluo-
rescent molecules while the sample does not fluoresce. Dips in the intensity of the detected signal indicate a band of analytes, since the lack of emission results in a diminished signal. The downside to this method, however, is that it lacks the sensitivity that direct fluorescence detection exhibits.

2.5 Size Separation

One method of separating analytes is by size. This approach allows separation that does not rely on partitioning or analyte charge. As such, it can be considered a fairly universal method. Size-separation methods are particularly useful for nanofluidic channels, where most other separation methods are difficult or impossible to implement.

2.5.1 Size Exclusion Chromatography

Size-exclusion chromatography can be considered a subclass of liquid chromatography, but with the distinction that the separation mechanism does not rely on chemical interaction. The stationary phase consists of porous materials into which analytes can diffuse. When an analyte passes into a pore, its diffusion path length will determine how and when it exits the pore and continues to elute through the column again. Large analytes will diffuse into pores much less frequently than smaller ones, and thus elute sooner. The smallest analytes will be continuously diffusing in and out of pores along the length of the column, resulting in long retention times. Thus, the elution order goes from largest analytes to smallest analytes.

Figure 2.6 illustrates the process of size-exclusion chromatography. In (a), the sample is introduced, and analytes begin to diffuse in and out of the pores of the particles that compose the stationary phase. In (b) - (c), the mobile phase continues to flow through the column as the analytes diffuse in and out of the pores. Example chromatograms are shown in the bottom of (c) and (d) to illustrate the order of elution, where the first peak pertains to the big, blue analytes in the figure and the last peak corresponds to the small, orange analytes.
Figure 2.6: Size exclusion chromatography. (a) A sample is introduced in a column filled with a porous stationary phase. The largest analytes are too large for the pores and elute unretained. (b) The red particles can diffuse into the pores, but are sufficiently large that they are retained only occasionally. (c) The green particles diffuse in and out of the pores more frequently, and (d) the orange particles are the smallest in the sample and, thus, are retained the longest, being the last to elute and be detected.
2.5.2 Capillary Gel Electrophoresis

Capillary gel electrophoresis is a size-selective separation method that combines electrophoresis with a suspended gel matrix within a channel. A sample solution is mobilized electrophoretically, and analytes will move through pores within the gel. Small analytes will pass quickly through the pores and elute first, while large analytes will experience increased steric interactions with the gel matrix and elute later.

2.5.3 Steric Field-Flow Fractionation

Field-flow fractionation is a technique that relies on an external field applied perpendicular to the length of the column to cause analytes to migrate to one side. A mobile phase then flows through the column to move the analytes through the system. If the flow is generated by pressure, then the flow will be laminar in most microfluidic and nanofluidic systems. This means that the flow velocity profile will look like a convex parabola, meaning that the flow in the center of the channel is highest, while the flow at the channel walls approaches zero. Therefore, larger analytes will stick further out into the channel and are, thus, subject to higher velocity flow fields. As such, large analytes will elute faster, while small analytes that are stuck in the slow-flow regions near the channel walls will elute last.

Figure 2.7 illustrates the process of field-flow fractionation. First, a sample is introduced into the column, as shown in (a). Once the sample is in place, the analytes need to be moved to one side of the column. This is achieved by applying an external field that is perpendicular to the length of the channel, as illustrated in (b). The external field can be electric, magnetic, thermal, or even a pressurized flow that moves through a semi-permeable membrane. Once all of the analytes have migrated to one channel side, a mobile phase flows through the channel to move the analytes to the detector. The flow of the mobile phase also acts as the separation mechanism. If the flow is pressurized, then the flow velocity is Poiseuille in nature. This means that the velocity profiles of the mobile phase assume that of a parabola across the width of the column. This flow profile is shown in Figure 2.7(c). The greatest velocity vector occurs in the middle of the channel, and falls off on each side until the wall of the column is reached. At the column wall, the velocity profile is zero. Larger analytes stick-out further into the channel and are thus subject to higher velocity lines, and
this causes them to elute quickly. Smaller analytes are affected by slower velocity lines and thus elute slower, as illustrated in Figure 2.7(d). Thus, the elution order is from largest analytes (blue analytes in the figure) to smallest analytes (orange analytes in the figure). It is important to note that the external field is continuously applied during this separation process such that analytes do not cross into higher velocity lines due to Brownian motion and elute faster, causing separated bands to smear. The bottom of (c) and (d) shows sample fractograms, illustrating detected peaks of analytes as they elute past the detector.

Analytical systems can be implemented into miniaturized, microchip systems. Such an approach would minimize the size and portability of the system, and if a microfabrication system were tooled to fabricate such devices, the cost would also be greatly reduced. Common separation mechanisms used with microfluidic devices include capillary liquid chromatography and capillary electrophoresis, while nanofluidic channels typically utilize capillary gel electrophoresis. One of the simplest and most powerful detection methods that can be easily incorporated with microchip separation devices is fluorescence detection. Size-separation techniques offer an attractive approach to separate analytes on microchip devices because of their universal application, especially for nanofluidic devices which typically cannot utilize other separation methods.
Figure 2.7: Field flow fractionation. (a) A sample is introduced into a column, after which (b) an external field is applied perpendicular to the length of the channel, causing the analytes to move to one side of the channel. (c) A pressurized mobile phase is through the channel, resulting in a parabolic velocity profile. Larger analytes are exposed to higher-velocity vectors and elute quickly, while smaller analytes are subject to slow velocity vectors. (d) The elution order is blue, red, green, and orange.
Chapter 3

Microfluidic/Nanofluidic Channel Fabrication

At the heart of thin film separation systems lies fabrication methods. Several fabrication methods have been developed over the years to create microfluidic and nanofluidic channels. Each method offers unique advantages as well as disadvantages. The intended application of the channel will be the most significant factor in determining which method to employ. Other relevant factors include cost and fabrication time.

The channel fabrication methods discussed in this chapter utilize thin film technology, which has been extensively developed over the last several decades in the microelectronics industry. In fact, most microfluidic and nanofluidic devices are created using the same technology that makes complex circuits, like microprocessors, possible. Microchannels and nanochannels fall into a category of devices called microelectromechanical systems (MEMS). As the name implies, MEMS utilize electrical and mechanical characteristics. MEMS include devices such as membranes for pressure sensors, micro-cantilevers, oscillators, and of course, microchannels and nanochannels [1]. The mechanical properties of channels must be taken into account to ensure channel stability and reliable behavior.

Thin film fabrication methods seem to be a counterintuitive approach for fabricating microchannels, given that thin film methods are planar in nature. The question then arises: how is it possible to create 3D structures with a 2D technology? Thanks to the ingenuity of many researchers over the years, thin film processes have been applied to the development of several types of microfluidic and nanofluidic channels. Some of these methods include bonding, shadow deposition, imprint molding, and sacrificial etching. An extensive review of nanofluidic fabrication technologies is presented by Mijatovic et al. [2]. The purpose of this chapter is to review these fabrication methods, and more extensively explore the
process of sacrificial etching, the process by which all microfluidic and nanofluidic channels were generated in this work.

3.1 Bonding

The most popular method for fabricating microchannels is bonding technology. Bonding has a fairly long history, with the birth of the modern-day wafer-bonding approach generally attributed to Lasky et al. [3] in 1985. The first MEMS sensors utilizing fusion bonding were implemented by Petersen et al. [4] in 1988. For a more thorough background of wafer bonding, refer to a review by Gösele and Tong [5].

Wafer bonding is simplistic in its approach: one wafer is placed in immediate contact with another, and because of the nature of the highly-polished wafer surfaces, they bond together through van der Waals forces. When this process is performed at room temperature, however, the surface bond can be broken fairly easily. As such, several bonding methods have been developed, including anodic, fusion, direct, chemical, and eutectic bonding. Anodic bonding is achieved by applying a high voltage across the joined wafers. In this approach, one wafer contains ions on its surface that will chemically react the surface of the other wafer under the influence of a high electric potential. Anodic bonding has been used to fabricate nanofluidic channels [6, 7]. Fusion bonding operates by heating bonded wafers to high temperatures so they anneal. Fusion bonding has been used to fabricate nanofluidic channels [8, 9, 10, 11, 12]. Direct bonding relies solely on van der Waals forces to maintain the bond, and this method has been used to create nanochannels [13]. Chemical bonding relies on a chemical reaction between the wafer surfaces to form a strong bond, and this method has been used to fabricate microfluidic and nanofluidic channels [14, 15]. Eutectic bonding joins two wafers under high temperatures, where the lower melting point of one wafer with respect to the other results in alloying at the bond union.

Two variant bonding techniques to create microchannels are illustrated in Figure 3.1. Figure 3.1(a) illustrates bonding a glass wafer to a silicon wafer. For the process shown in (a), a silicon wafer (1) is coated in a layer of deposited silicon dioxide (2). Next, a masking layer of photoresist is spun over the top and is patterned. The exposed oxide is etched away, typically using a dry-etching process since a high degree of anisotropy is generally desired.
Following the etching step, the masking photoresist is removed, and a thin layer of oxide is deposited to ensure wettability along the bottom of the trench. In the final step, a glass wafer is bonded to the top of the processed wafer, which results in a hollow channel. Figure 3.1(b) illustrates channel generation by pattern transfer via bonding. Two silicon wafers are used, as shown in 1a and 1b. On the first wafer, a masking layer of photoresist is applied and developed. The exposed silicon is then etched, typically via a wet-etch process, resulting in a trench etched into the silicon. The other wafer is coated in a thick layer of oxide. The first wafer is then cleared of the masking photoresist, and a layer of oxide is deposited over the etched trenches. The two wafer-faces are then brought into bonding contact, and the first wafer is entirely etched away, leaving behind a hollow channel made of silicon oxide.

Both bonding methods illustrated in Figure 3.1 exhibit some advantages. The process shown in Figure 3.1(a) allows for cross-sectional geometries with very orthogonal edges. However, the top layer of oxide may potentially be excessively thick for some applications, since this layer will be as thick as the glass wafer bonded to the silicon wafer. The process shown in Figure 3.1(b) results in channels with a controllable, thin film cover. However, because of the nature of wet-etching, the cross-sectional profile of the channel will not be rectangular, but trapezoidal in shape.

3.2 Shadow Deposition

Shadow deposition is a technique that was explored by Johnson et al. as a means of generating grooves and rounded profiles by depositing films at an angle off-normal to a series of relief patterns on a substrate. Over the years, this method has been extended to enclose patterned grooves to result in hollow microfluidic and nanofluidic channels. Cao et al. demonstrated channel fabrication with dimensions as small as 10 nm by utilizing shadow-deposited silicon dioxide. Han et al. fabricated 27-nm tall channels by shadow depositing oxide over a pedestal of silicon nitride resting atop a narrow base. Other microfluidic and nanofluidic channels have been created using phospho silicate glass (PSG), where non-conformal deposition sealed a hollow channel within a trench structure.
Figure 3.1: Fabrication flow chart illustrating microchannel fabrication using bonding. Two main bonding methods are illustrated under (a) and (b). The first method, (a), involves bonding a glass substrate to a trench etched in a silicon dioxide thin film. The second method, (b), involves etching a trench in silicon, coating the trench in glass, and bonding this to a second silicon wafer with a surface layer of glass. The first wafer is completely etched away, leaving behind a microchannel on top of the second wafer.

Figure 3.2 illustrates the general concept of channel fabrication using shadow deposition. In (a), relief structures are etched into a substrate or thin film. The substrate is then
rotated at an angle $\theta$ to the normal path of an evaporant stream. In (b), as the evaporant deposits on the relief structures, portions of the trench will become increasingly shadowed as the film increases in thickness. Depending on the setup, there may be areas within the trench that will not be covered by the evaporated material at all. By (c), sufficient film has been deposited to enclose the trenches, resulting in the generation of hollow microfluidic or nanofluidic channels.

![Figure 3.2: Microfluidic or nanofluidic channel creation using shadow deposition. (a) A patterned thin film is rotated an angle $\theta$ normal to the path of an evaporant. (b) The trenches patterned in the initial film partially obscure evaporant from penetrating all the way to the bottom. Over time, a layer of evaporated thin film begins to form along the top of the trenches and the sides exposed to the evaporant path. (c) Eventually, the film closes together and seals off a hollow section that will be used as a microchannel or nanochannel.](image)

The primary benefit of this method is that it produces channels quickly. Once a relief grating is etched into a film, channel fabrication time is only as long as the time required to deposit a thin film layer. Another advantage of this method is that it relies solely on microfabrication technology, such that additional wafers need not be introduced, as is the case with bonding. On the downside, however, channel geometries will be highly very asymmetrical, and absolute channel dimensions can be difficult to control.

### 3.3 Imprint Molding

Imprint molding came about as a means of repeating fabrication of high-resolution devices without the need of high-resolution lithography equipment (i.e. extreme ultraviolet
lithography [21], X-ray lithography, or electron beam lithography [22, 23]). Imprint molding begins with generating a mold with very fine features requiring high-resolution lithographic processes. Once the mold has been created, films can be deposited over or impressed into the mold, and the film can then be transferred to a substrate. Stewart et al. demonstrated the ability to create structures down to 20 nm in dimension using imprint lithography [24]. Imprint molding has been used to fabricate nanofluidic channels ranging in size from approximately 100 nm to 250 nm [25, 26].

Figure 3.3 illustrates the process of imprint molding to create microchannels and nanochannels. In (a), a mold with fine features has been generated previously, utilizing a high-resolution lithography technique. In this particular example, a polymer film is coated over a substrate, and is brought into contact with the mold in (b). At this point, the mold pattern can be imprinted into the film through mechanisms such as pressure or heat. In (c), the substrate is removed from the film, and in (d), the imprinted polymer is peeled away from the mold and placed over a flat substrate surface. In (e), the imprinted film is brought into contact with the substrate. Adhesion of this film can be promoted through mechanisms such as pressure or elevated temperatures. The indented portions of the film form hollow channels.

Perhaps the greatest advantage of imprint molding is the ability to fabricate non-planar nanofluidic channels without using high-resolution lithography. Another advantage is the high throughput available since laborious lithography steps are not required. Some disadvantages include the ability to ensure high conformality between the film and the mold features. Similarly, the method of imprinted film transfer to another substrate must be addressed. Another disadvantage is that aligning the imprinted film to pre-existing features on a substrate can be difficult.

### 3.4 Sacrificial Etching

Sacrificial etching is the fabrication method pursued in this work to generate all microfluidic and nanofluidic channels. Sacrificial etching relies on a thin-film core that is encased within a second film. Once the encapsulating second film is in place, the core is
removed by utilizing a highly-selective etchant. Once the core is etched away, only the outer film remains, which serves as a hollow channel.

### 3.4.1 Fabrication

Sacrificial etching is a method that can utilize several combinations of encapsulating materials and sacrificial cores. Some of these materials include silicon, polymers, and metals. Using amorphous silicon, Stern *et al.* [27] fabricated a 20-nm tall channel. Other researchers have fabricated nanochannels based on polysilicon cores that range in height from 5 nm
to 270 nm [28, 29, 30, 31, 32]. Polymer cores have also been used to produce a variety of microfluidic channels [33, 34, 35, 36, 37, 38, 39] as well as nanofluidic channels as shallow as 100 nm in height [25]. Thin film metals comprise another set of sacrificial materials that can be deposited to heights of a few nanometers, and many metals benefit from rapid etching compared to other sacrificial materials [40, 41, 42, 43, 44].

The general process of sacrificial etching is illustrated in Figure 3.4. In (a), a layer of oxide is first deposited over a substrate. This layer serves as the bottom wall of the channel and must be highly wettable to maximize flow. A sacrificial layer which will form the cores of the channels is deposited in (b). In (c), if the sacrificial material is a photoresist, it can be directly patterned and developed to form the channel core. Otherwise, masking photoresist must first be patterned on top of the sacrificial core, followed by etching away the exposed material. After etching, the masking photoresist is removed. A thick layer of oxide is then deposited over the top, as shown in (d). This top layer will serve as the ceiling and sidewalls of the channel, and it must also be highly wettable to maximize flow rates. In (e), the sacrificial core is removed with an appropriate etchant. It is necessary to select an etchant that is highly selective so that it will remove the core while not damaging the structural oxide coating. Once the core is completely removed, a hollow oxide channel remains.

Sacrificial etching experiences several advantages. One advantage is that all device fabrication can be performed using standard photolithographic processes; there is no need to transfer patterns from one substrate to another or bond one wafer to another. Another benefit is the high channel resolution that is possible with thin-film depositions. The major drawback of sacrificial etching, however, is that removing a sacrificial core can take a long time. As a sacrificial core is etched away, the product of the reaction will eventually have to diffuse back out through the opening of the channel so that more reactants can be introduced to further the etching process. In other words, sacrificial etching times are diffusion limited. As the core etches further and further back, the diffusion path length for reacted products is increased. The etching time required for a sacrificial core can be expressed in the terms of Equation (3.1) [27],

$$l \approx \sqrt{2k_n D_{c_{o}}t},$$ (3.1)
where $l$ is the total etched length of the core, $k_n$ is a geometric constant, $D$ is the diffusion constant of the limiting agent, $c_o$ is the molar concentration of the limiting agent, and $t$ is the total elapsed time. Therefore, the etching speed can be maximized if the diffusion constant and reagent concentration are maximized. The main thing to note from Equation (3.1) is that $t \propto l^2$, meaning that a long channel can result in very long etching times. Another point to note from this equation is that $k_n$ is proportional to the channel height, such that shallower channels will etch slower than taller ones.

As mentioned previously, polymer cores are frequently employed when fabricating sacrificially-etched microfluidic channels. However, many etchants that target polymers exhibit low diffusion constants. As such, polymer cores may require prohibitively-long etch times to form channels [33]. One method that speeds the core-etching process is coupling sacrificial cores. If one core can be etched more quickly than the other, this leaves the second core in place with a large exposed surface area. Because of the increase in exposed surface area, the second core can be removed quickly. One approach that utilizes this concept cou-
pled a thin sacrificial aluminum core with a thick photoresist core [45, 46]. The benefit of such a process is that it is possible to fabricate microchannels much more quickly than by utilizing a single polymer core. Aluminum cores etch quickly [41] but cannot be easily deposited to heights in excess of a couple hundred nanometers. Photoresist is easily spun to heights in the micron range, but these cores etch slowly. By combining these cores, the aluminum can be quickly etched away, leaving a photoresist core with a large exposed surface area that can be quickly removed.

Figure 3.5 illustrates the concept of double-core sacrificial etching. In (a), a substrate is coated with a layer of oxide, followed by the deposition of the first sacrificial layer in (b). A second sacrificial layer is then deposited in (c), where the sum of the thicknesses of both sacrificial layers will determine the overall height of the channel. In (d), if the second sacrificial layer is a photoresist, it can be directly patterned and developed. If not, a masking photoresist needs to be applied over the top of the film, and the exposed film will be etched away, followed by the removal of the masking photoresist. In (e), the second sacrificial film acts as a mask for the first sacrificial film, and the exposed areas of the first sacrificial film will be etched away. A thick layer of structural oxide is then deposited in (f). In (g), the first sacrificial layer is removed. This exposes the underside of the second sacrificial layer, allowing the following etchant to come into contact with high surface area of the core. By (h), each of the cores has been completely removed, and only a hollow oxide channel remains.

When selecting a core/etchant scheme for generating microfluidic or nanofluidic channels, it is necessary to select cores and etchants such that the chemical reaction does not damage the channels, and the etching process can be completed in a realistic time frame. In this work, AZ3330 photoresist, SU8 photoresist, aluminum, germanium, and chromium have all been utilized as sacrificial cores. Table 3.1 summarizes these sacrificial cores along with their accompanying etchants, typical etch lengths, characteristic channel dimensions, and typical etch times. In Table 3.1, Nanostrip is a commercially-available stable form of piranha. The piranha listed in the table corresponds to a mixture of sulfuric acid and hydrogen peroxide (1:1), and aqua regia is a mixture of hydrochloric acid and nitric acid (2:1). The chromium etchant is a commercially-available etchant from Micro-Chrome Technologies (CEP 200). It is apparent from Table 3.1 that the fastest etching material tested was alu-
Figure 3.5: Double-layer sacrificial etching. (a) Oxide is deposited over a substrate, followed by (b) the deposition of the first sacrificial layer. (c) A second sacrificial layer, usually a photoresist, is then deposited. (d) The photoresist layer is then patterned and developed, after which (e) the exposed areas of the first sacrificial layer are etched away. (f) A thick layer of oxide is deposited over the top of the sacrificial cores, following which (g) the first sacrificial core is removed by an appropriate etchant. (h) The second etchant will then quickly remove the remaining sacrificial material due to the highly-exposed surface area of the second core, resulting in a hollow channel with oxide on all sides.

Minum and, hence, it was utilized for one of two layers in most microfluidic channels and as the sacrificial layer in most nanofluidic channels. The progression of etching an aluminum sacrificial core is shown for a nanofluidic channel in Figure 3.6. The top half of the image contains aluminum cores that are still intact and appear as a bright gray or white color. The bottom half of the image shows channels that have been completely cleared of their sacrificial cores. The jagged end of the most recently-etched portion of the aluminum cores is visible.
Table 3.1: Materials tested as sacrificial cores for microfluidic and nanofluidic channels.

<table>
<thead>
<tr>
<th>Material</th>
<th>Etchant</th>
<th>Etch Length</th>
<th>Dimensions</th>
<th>Etch Durations</th>
</tr>
</thead>
<tbody>
<tr>
<td>AZ3330</td>
<td>Nanostrip</td>
<td>~1 cm</td>
<td>~3 µm</td>
<td>~1 day*</td>
</tr>
<tr>
<td>SU8</td>
<td>piranha</td>
<td>~1 cm</td>
<td>~5 µm</td>
<td>~3 days</td>
</tr>
<tr>
<td>Al</td>
<td>aqua regia</td>
<td>~1 mm</td>
<td>20 nm – 200 nm</td>
<td>~4 hours</td>
</tr>
<tr>
<td>Ge</td>
<td>hydrogen peroxide</td>
<td>~1 mm</td>
<td>20 nm – 100 nm</td>
<td>~1 week</td>
</tr>
<tr>
<td>Cr</td>
<td>chromium etchant</td>
<td>~1 mm</td>
<td>20 nm – 100 nm</td>
<td>~4 weeks</td>
</tr>
</tbody>
</table>

*core was coupled with a thin aluminum core to speed etching time.

Figure 3.6: Etching of sacrificial aluminum cores. The top half of each channel still contains sacrificial aluminum cores. The bottom half of each channel is etched clear. It is possible to see the jagged edge of the aluminum cores where the etchant is stripping the material away.

3.4.2 Channel Characterization

Once channel fabrication has been performed, it is essential to characterize those channels. It is important to know if the channels are intact, if they exhibit the appropriate dimensions, and if any anomalies occurred during fabrication that can be corrected in future fabrication processes.

The majority of experiments performed in this work depend heavily on a knowledge of exact channel dimensions. Such knowledge allows characterizing aspects such as flow rates, zeta potentials, contact angles, and analyte trapping ranges. Perhaps the most straightforward and accurate way of determining channel dimensions is to profile the sacrificial core
before the top layer of oxide is deposited. This is necessary because once the oxide layer is deposited, information about the shape and absolute height of the channel is lost due to non-conformal film coverage. Topographical profiling of the sacrificial core was performed using a Tencor AlphaStep 200 surface profilometer. Figure 3.7 shows two core profiles measured with a profilometer. Figure 3.7(a) is a profile of an SU8 core with a height around 5 \( \mu \text{m} \), and Figure 3.7(b) is a profile of an aluminum core with a height around 95 nm. From these profile images it is possible to determine measurements such as surface roughness and average height, as well as provide information about the shape of the core.

![Figure 3.7](image)

**Figure 3.7:** Profiles of sacrificial cores. (a) Profile of a core fabricated from SU8. (b) Profile of a core fabricated from aluminum.

Another powerful method that allows for channel characterization once all fabrication steps are completed is through scanning electron microscopy (SEM). A scanning electron micrograph will allow high-resolution imaging of a channel at just about any angle or position imaginable. Figure 3.8 contains a collection of scanning electron micrographs captured from a microfluidic device that was fabricated using a coupled aluminum/AZ3330 photoresist sacrificial core. Figure 3.8(a) provides an overhead view of a group of parallel microchannels. These channels are coated in PECVD silicon dioxide, and small defects and inhomogeneities can be seen across the channels. Even so, the channels appear otherwise to be intact. Figure 3.8(b) is an image showing the openings at the end of a group of microfluidic channels. These openings are made by wet-etching through oxide with a buffered oxide etchant composed of diluted hydrofluoric acid. Through these openings, etchants such as aqua regia and
Nanostrip are able to access and remove the sacrificial cores. Figure 3.8(c) is a scanning electron micrograph highlighting a microchannel cross-section. Imaging channel cross-sections is possible if the device is first cleaved perpendicular to the length of the channels, and the device is then rotated so that the plane of the cross-section faces the microscope pole piece. A cross-sectional image like the one in Figure 3.8(c) provides useful information about the dimensions and shape of the fully-fabricated channel. It also provides information about the integrity of the channel. For example, extensive seams begin at the bottom corners of the cross-section and propagate up toward the upper corners of the viewing window. Identifying the presence of seams can lead to adapted fabrication steps to correct the problem and increase device yield, since seams can result in channel structural failure. Corrective measures, such as thick oxide depositions, aid in increasing structural stability. Other possible approaches include annealing a completed device, although this method was not attempted in this work.

SEM analysis can also be applied to nanofluidic channels. The nanofluidic channels fabricated here are planar in nature, meaning that they exhibit a very high width-to-height aspect ratio. As such, imaging can be difficult because a scanning electron microscope scans with a one-to-one ratio for the width and height component. Due to this complication, initial attempts to cleave a channel and view the cross-section proved fruitless. However, an alternative approach proved successful. This method involved ion-beam milling a cross-section out of a nanochannel followed by imaging with a scanning electron microscope. Figure 3.9 is a scanning electron micrograph of a 145-nm tall nanofluidic channel imaged using this method. This image provides information such as channel height variability along the width of the channel. By imaging nanochannel cross-sections, it was discovered that thin channel tops led to channel bowing due to high compressive film stresses. This problem was corrected by increasing the deposition thickness, followed by imaging the channels to ensure a constant channel height, as exhibited by the channel shown in Figure 3.9.

Microfluidic and nanofluidic channels can be fabricated using techniques such as bonding, shadow deposition, imprint molding, and sacrificial etching. Sacrificial etching is the fabrication method pursued in this work. Sacrificial etching benefits from the ability to utilize thin film fabrication methods, but etch times can be prohibitively long if the device
Figure 3.8: Scanning electron micrographs of rounded-core sacrificially-etched channels. (a) Top view of oxide rounded-core channels. (b) Channel openings created by reservoir etching. (c) Cross-section of a microfluidic channel created from a reflowed photoresist core combined with a thin aluminum core.

is not carefully designed. Sacrificially-etched microfluidic and nanofluidic channels can be characterized using techniques such as profilometry and SEM. General fabrication information has been provided in this chapter, but more specific fabrication processes will be given in subsequent chapters as they apply to specific devices.
**Figure 3.9:** Scanning electron micrograph of the cross-section of a 145-nm tall nanofluidic channel. The sacrificial core was aluminum. With a 5 µm-thick coating of oxide, the channel height is constant across the channel width.
Chapter 4

Pumping Mechanisms

At the heart of analytical separations lie filling methods. If the separated analytes cannot be moved to a point where they are detected, the separation is useless. Just as there are nearly innumerable separation methods, similarly there are numerous methods to pump fluids through microfluidic systems. The scope of this work is not to exhaustively treat the various methods that exist but to present and analyze some of the most common: pressurized (or Poiseuille) flow, electroosmotic (EO) flow, and capillary flow, also known as capillarity. Extensive experiments have been performed to characterize EO flow in microchannels and capillary flow in nanochannels.

4.1 Pressurized Injection

Perhaps the most common fluid injection method in microfluidic systems is pressurized injection. Pressurized injection is a popular method because stable pressure generation is well founded and easily integrable with silica capillaries. For typical pressure injections, one end of a channel is connected to a high-pressure pump while the other end is open to atmospheric pressure. This results in a net force directed toward the atmospheric end of the channel, which drives fluids through the system.

The flow profile resulting from pressurized injection is parabolic in shape. This flow profile is shown in Figure 4.1. It is assumed that a constant pressure is applied across the fluid, but the meniscus takes on a parabolic, convex shape. This is due to the friction generated between the fluid with the channel walls. Because of these interactions the average velocity of the fluid immediately at the channel/fluid interface approaches zero, while the maximum average velocity occurs at the center of the channel.
Figure 4.1: Pressurized flow, also known as Poiseuille flow, is generated by applying a pressure drop across a fluidic channel. The resulting flow velocity profile is convex, where the maximum flow rate occurs at the center of the channel and the velocity falls off to zero at the walls.

While pressurized injections can be simple to implement and result in high flow rates, the parabolic flow profile shown in Figure 4.1 causes band dispersion, resulting in lower separation efficiency. Because of this drawback, many approaches have been designed to break-up this parabolic profile when using pressurized injections. For example, packing a column with microparticles is one method of disrupting the Poiseuille flow profile, causing it to approximate more closely to a plug shape, thus increasing the resolution of separated bands.

The purpose of this section is to provide an introduction to Poiseuille flow and demonstrate how commonly-used equations reflecting Poiseuille flow are derived. Iverson and Garimella provide a very thorough review of micropumps, including pressurized pumps [47], and Amirouche et al. review mechanical pumps for biomedical applications [48].

4.1.1 Derivation

To understand how Poiseuille flow operates, it is important to understand where it comes from. Poiseuille flow can be derived from the Navier-Stokes equations, which are analogous to the continuum fluid mechanics version of Newton’s second law ($F = ma$). The Navier-Stokes equation for incompressible flow can be expressed as

$$
\rho \left( \frac{\partial u_p}{\partial t} + u_p \cdot \nabla u_p \right) = -\nabla p + \eta \nabla^2 u_p + f, \quad (4.1)
$$
where $\rho$ is the fluid density, $\mathbf{u}_p$ is the fluid velocity vector for Poiseuille flow, $t$ is time, $p$ is the applied pressure, $\eta$ is the dynamic fluid viscosity, and $\mathbf{f}$ represents any externally applied forces. For Poiseuille flow, no external field is applied, such that $\mathbf{f} = 0$.

To simplify the left-hand side of Equation (4.1), we need to consider the role of inertial and viscous forces in the microfluidic scale. One way to do this is by considering the Reynolds number,

$$
\text{Re} \equiv \frac{\rho U_o L_o}{\eta},
$$

(4.2)

where $U_o$ is a velocity scale and $L_o$ is the characteristic length scale. The Reynolds number represents the ratio of inertial forces to viscous forces acting on a moving fluid. For a typical system, where water is the fluid of interest, $\rho$ is around $1.0 \text{ g cm}^{-3}$ and $\eta$ is around $1 \times 10^{-2} \text{ g cm}^{-1} \text{ s}^{-1}$. Assuming a large microfluidic channel with a radius of 100 $\mu$m and a high flow rate of 1 cm s$^{-1}$, these values lead to a Reynolds number of $\text{Re} = 1$, where the laminar flow regime extends up to $\text{Re} = 10$. Flow is not considered to be fully turbulent until around $\text{Re} = 2000$, such that for the majority of microfluidic systems, viscous forces greatly outweigh inertial forces. When advective inertial forces are insignificant compared to viscous forces (i.e., at low Re), the left-hand side of Equation (4.1) vanishes, leaving a simplified form of the Navier-Stokes equation known as the Stokes equation,

$$
\eta \nabla^2 \mathbf{u}_p = \nabla p.
$$

(4.3)

Many microfluidic and nanofluidic systems can be thought of as planar, meaning that only two dimensions are significant. Adopting the coordinate system in Figure 4.1, pressure is applied only along the $z$-direction, and flow velocities only change along the $y$-direction. This simplifies Equation (4.3) to

$$
\eta \frac{\partial^2 \mathbf{u}_p}{\partial y^2} = \frac{\partial p}{\partial z}.
$$

(4.4)

Ultimately, it is desired to solve for $\mathbf{u}_p$ to gain an understanding of flow velocities in channels using Poiseuille flow. This can be accomplished by integrating both sides of
Equation (4.4) with respect to $y$, such that

$$\eta \frac{\partial u_p}{\partial y} = \frac{\partial p}{\partial z} y + C_1. \quad (4.5)$$

A Poiseuille flow profile is parabolic in nature, where the maximum velocity occurs at the center of the channel. As such, the boundary condition for the maximum velocity at the center of the channel can be expressed as

$$\frac{\partial u_p}{\partial y} \bigg|_{y=0} = 0, \quad (4.6)$$

where it is assumed that $y = 0$ is the center line of the channel. Applying this boundary condition to Equation (4.5) leads to $C_1 = 0$. Now it is possible to integrate Equation (4.5), resulting in

$$u_p = \frac{y^2 \partial p}{2\eta \partial z} + C_2. \quad (4.7)$$

For a Poiseuille flow profile, the velocity at the channel walls is equal to zero. This is referred to as the no-slip boundary condition, and although this condition is never entirely true in real systems, the assumption is usually considered sufficient. The no-slip boundary condition is

$$u_p \left( \frac{h}{2} \right) = u_p \left( -\frac{h}{2} \right) = 0, \quad (4.8)$$

leading to $C_2 = -h^2/8\eta \cdot \partial p/\partial z$, and Equation (4.7) becomes

$$u_p = \frac{1}{8\eta \partial z} \left( 4y^2 - h^2 \right). \quad (4.9)$$

Equation (4.9) expresses the Poiseuille flow velocity, $u_p$, at any position, $y$.

Oftentimes it is desirable to know the maximum fluid velocity in a channel. Since the maximum velocity occurs at $y = 0$, it becomes

$$u_{p\text{-max}} = \frac{-h^2 \partial p}{8\eta \partial z} \quad (4.10)$$

for planar channels driven by Poiseuille flow.
While maximum flow rate may be interesting, the average flow rate is usually a more meaningful measure. The average velocity of Poiseuille flow is found by integrating Equation (4.9) from \(-h/2\) to \(h/2\) and dividing by the total length, \(h\), resulting in

\[
u_{p-avg} = \frac{1}{8h\eta} \frac{\partial p}{\partial z} \int_{-h/2}^{h/2} (4y^2 - h^2) \, dy = -\frac{h^2}{12\eta} \frac{\partial p}{\partial z}. \tag{4.11}
\]

For most systems, the pressure drop across the length of the channel is constant, such that \(\frac{\partial p}{\partial z}\) can be replaced with \(\Delta p/L\), where \(L\) is the length of the channel, leading to

\[
u_{p-avg} = -\frac{h^2}{12\eta L}. \tag{4.12}
\]

This is referred to as the Hagen-Poiseuille equation, and is used to model flow rates due to a pressure drop across a planar channel. Equation 4.12 can be easily extended to describe the average volumetric flow rate,

\[
Q_{p-avg} = -\frac{h^2\Delta p A}{12\eta L}, \tag{4.13}
\]

where \(A\) is the cross-sectional area of the fluidic channel.

If the same derivation is followed as outlined above using cylindrical coordinates, it is possible to solve the Hagen-Poiseuille equation for cylindrical channels instead of planar ones. The result is

\[
u_{o-avg} = -\frac{d^2\Delta p}{32\eta L}, \tag{4.14}
\]

\[
Q_{o-avg} = -\frac{\pi d^4\Delta p}{128\eta L}, \tag{4.15}
\]

where \(d\) is the diameter of the channel. Notice that the average flow rate in a cylindrical channel is lower than that of a planar channel with a height equal to the diameter of the cylindrical channel. This is not surprising since a cylindrical channel incorporates wall effects on all sides, while a planar channel considers only a top and bottom wall.

Pressurized injections are important to understand due to their popularity in separation systems. Maximum flow rates occur toward the center of the channel while flow velocity
drops off to zero at the channel walls. Pressurized flow also provides the basis for solving for capillary flow.

4.2 Electroosmosis

Following pressurized injections, electroosmotic (EO) flow could be considered the second-most popular method for pumping liquids. For most charged surfaces, when an electric field is applied across the channel length, a bulk fluid motion, called electroosmosis, will ensue. The direction of the flow depends on the polarization of the electrodes as well as the charge on the channel wall surface. Figure 4.2 illustrates EO flow occurring in the positive \( z \) direction. One of the distinct features to note about EO flow is that the flow profile is very flat compared to most other methods, which greatly increases the efficiency of separation experiments. This led to the popularity and increasing use of electroosmosis as a pumping method for many microfluidic applications.

![Electroosmosis](image)

**Figure 4.2:** Electroosmosis is generated by applying an electric field across an electrolyte. The resulting flow velocity profile is very flat, except very near the channel walls where the velocity drops off to zero.

The purpose of this section is to provide an introduction to EO flow and its uses in microfluidic systems. EO flow velocities will be derived, and these derivations will aid in explaining advanced EO pump systems. Experimental results of flow rates and calculated surface charges will be discussed.
4.2.1 Physics of Electroosmosis

The first observation of electroosmosis is attributed to F. F. Reuss. He reported the movement of water across a clay membrane in the presence of an electric field and named the phenomenon electroosmosis, since it appeared that the water migrated across a semipermeable membrane. In fact, the motion was due to ions in the solution being attracted to the negatively-charged walls of naturally-occurring microchannels within the clay, which then migrated under the influence of the electric field from one end of the clay to the other. The term “electroosmosis” is then somewhat of a misnomer, but the name is too deeply-entrenched at this point and will continue to be the accepted nomenclature.

Electroosmosis operates due to charge imbalances that exist on most surfaces. Homogeneous materials will generally be charge-neutral throughout their bulk, because bulk atoms are surrounded on all sides by neighbors with whom they can easily share electrons to fill vacancies in their outermost valence shell. However, at the surface the material forms an interface with something other than the bulk material (whether that be another solid, air, or some liquid), such that there are no neighboring atoms of similar composition, resulting in exposed charges or weak bonds that are formed with other available molecules. These loosely bound molecules can then be removed when they come into contact with species with greater affinity.

EO flow occurs as an electric field is applied across a channel with charged layers present at its walls. To form charged channel layers, glass is an excellent candidate. Glass, in all its forms, is composed primarily of SiO$_2$. In microfabrication, glass can exist in the form of a highly porous spin-on glass, in more dense LPCVD or PECVD oxide, or as very dense oxide grown in a high-temperature thermal oven. In all cases, the surface of glass is predominately occupied by silanol sites, SiOH, as shown in Figure 4.3(a). Although silanol is neutral, the hydrogen is only loosely bound to the oxygen. By bringing the surface into contact with a solution with a pH of 3 or higher, the silanol sites ionize, leaving behind negatively-charged oxygen atoms, shown in Figure 4.3(b). These exposed negative charges attract free cations available in the solution, which will ultimately form two distinct layers: the Stern layer and the diffuse layer, as shown in Figure 4.3(c). The Stern layer consists of ions that are in the closest proximity to the channel wall, and are bound so tightly by
their electrostatic interaction that they become immobile. The diffuse layer forms a second, less-tightly bound layer adjacent to the Stern layer. The diffuse layer consists of ions that act to fully shield the remaining negative charge present at the channel wall, such that the overall electric potential seen in the bulk fluid is zero. The diffuse layer is named such because, unlike in the Stern layer, ions here are mobile, particularly under the influence of a tangential electric field. The combination of the Stern layer and diffuse layer comprises what is referred to as the electric double layer. Figure 4.3(d) illustrates what occurs when an electric potential is applied across the length of the channel. While the ions that comprise the Stern layer remain fixed, the concentrated ions in the diffuse layer now drift toward the cathode. Because the ions in the diffuse layer are solvated, they drag the bulk fluid along with them toward the cathode, such that the resulting electroosmotic mobility will be identical for all cations and anions in the bulk solution. In other words, all molecules will flow at the same rate, which results in a flat flow velocity profile (except at the extreme edges where the velocity of the ions in the Stern layer is zero). A flat flow profile is desirable because it will minimize band broadening in analytical separations.

There are several surfaces that will support electroosmosis. The previous explanation of the workings of electroosmosis was based upon using a glass wall, which results in a negatively-charged surface. Electroosmosis will also work for positively-charged surfaces, where anions would form the electric double layer instead of cations.

### 4.2.2 Derivation

Now that a qualitative description of electroosmosis has been given, it is of interest to form a quantitative form that describes EO flow. As with Poiseuille flow, EO flow can be derived from the Navier-Stokes equations [49, 50, 51, 52],

\[ \rho \left( \frac{\partial \mathbf{u}_{eo}}{\partial t} + \mathbf{u}_{eo} \cdot \nabla \mathbf{u}_{eo} \right) = -\nabla p + \eta \nabla^2 \mathbf{u}_{eo} + \mathbf{f}, \]  

(4.16)

where \( \rho \) is the fluid density, \( \mathbf{u}_{eo} \) is the liquid velocity vector for EO flow, \( p \) is the applied pressure, \( \eta \) is the dynamic fluid viscosity, and \( \mathbf{f} \) is the body forces per unit volume acting on the fluid. For a system driven purely by electroosmosis there will be no pressure gradient,
Figure 4.3: How electroosmosis works. (a) A glass wall consists of exposed silanol sites, SiOH. (b) As a solution with a pH of 3 or greater is introduced, the silanol sites ionize to leave behind negatively-charged oxygen atoms. (c) Cations are attracted to the exposed negative charges, and form a tightly-bound Stern layer and a mobile diffuse layer. The combination of these layers comprise the electric double layer. (d) As an electric field is applied across the channel length, solvated ions in the diffuse layer migrate to the cathode and drag the bulk fluid with them. The resulting flow profile is flat. It should be noted that the thicknesses of the Stern and diffuse layers in relation to the bulk fluid layer in these illustrations is greatly exaggerated for most microfluidic systems.

such that $\nabla p = 0$. Since an electric field is applied across the length of the channel, the external body force is equal to $f = \rho_e E$, where $\rho_e$ is the charge density and $E$ is the applied electric field. Furthermore, for most microfluidic and nanofluidic systems, the viscous forces greatly outweigh the inertial forces, which means the left-hand side of Equation (4.16) vanishes, leaving

$$\eta \nabla^2 u_{eo} = \rho_e E.$$  \hspace{1cm} (4.17)
We can further simplify Equation (4.17) by assuming that the channel of interest is planar, and that the dimensions of the channel are much greater than the thickness of the electric double layer. Under these assumptions, the electric field is applied along the \( z \)-direction, as shown in the coordinate system in Figure 4.2. This allows the one-dimensional model of Equation (4.17),

\[
\eta \frac{\partial^2 u_{eo}}{\partial y^2} = - \rho_e E_z.
\]  

(4.18)

The charge density, \( \rho_e \), needs to be described in terms of the electrostatic potential, \( \phi \), resulting from the formation of the electric double layer. Figure 4.4 illustrates the relationship between \( \phi \) and \( y \). Since the channel wall attraction is strongest at its surface, the greatest concentration of ions will be located near the surface within the Stern layer. The electric potential decreases throughout the diffuse layer, until the field generated by the ions and the surface charges cancel each other in the bulk fluid. The slipping boundary is the point where the fluid begins to move, and is usually approximated to occur at the interface between the Stern and diffuse layer. The potential at the slip boundary is called the zeta potential, \( \zeta \), and is usually approximated to equal the potential at the channel surface \( (\phi(0) = \phi_s \approx \zeta) \). The thickness of the electric double layer is called the Debye layer thickness, \( \lambda_D \). Also shown in Figure 4.4 is a dotted line showing the approximate flow rate. From this discussion it is possible to specify two sets of important boundary conditions,

\[
\begin{align*}
\frac{\partial u}{\partial y} \bigg|_{\text{bulk}} &= 0, & \frac{\partial \phi}{\partial y} \bigg|_{\text{bulk}} &= 0, \\
\phi(0) &= \zeta. 
\end{align*}
\]  

(4.19), (4.20)

Understanding the electric potential set up by the electric double layer allows for solving for EO flow rates. The electric potential, \( \phi \), is related to the charge density, \( \rho_e \), by Poisson’s equation, which states

\[
\nabla^2 \phi = \frac{\partial^2 \phi}{\partial y^2} = - \frac{\rho_e}{\epsilon},
\]  

(4.21)
where $\epsilon$ is the electrical permittivity of the solution. Plugging Equation (4.21) into Equation (4.18) results in

$$
\eta \frac{\partial^2 u_{eo}}{\partial y^2} = \epsilon E_z \frac{\partial^2 \phi}{\partial y^2}.
$$

(4.22)

Integrating Equation (4.22) with respect to $y$ results in $\eta \partial u / \partial y = \epsilon E_z \partial \phi / \partial y + C_1$. Applying the boundary conditions outlined in Equation (4.19) leads to $C_1 = 0$, so that

$$
\eta \frac{\partial u_{eo}}{\partial y} = \epsilon E_z \frac{\partial \phi}{\partial y}.
$$

(4.23)

Integrating Equation (4.23) with respect to $y$ leads to $\eta u = \epsilon E_z \phi + C_2$, and after applying the boundary conditions specified in Equation (4.20), $C_2 = -\epsilon E_z \zeta$, so that

$$
u_{eo} = \frac{\epsilon E_z}{\eta} \left( \phi - \zeta \right).
$$

(4.24)
If we now assume that $y \gg \lambda_D$, which comprises the majority of the flow area, then $\phi = 0$ and Equation (4.24) becomes

$$u_{eo} = -\frac{\epsilon E_z \zeta}{\eta} = -\mu_{eo} E_z,$$  \hspace{1cm} (4.25)

where $\mu_{eo}$ is equal to $\epsilon \zeta/\eta$ and is called the electroosmotic mobility. Equation (4.25) is the Helmholtz-Smoluchowski equation, and is used to model fluid flow due to electroosmosis when the channel dimensions are much greater than the Debye screening length.

For most practical systems, the volumetric flow rate, $Q_{eo}$, is the measure of interest, and can be expressed as

$$Q_{eo} = -\mu_{eo} E_z A,$$  \hspace{1cm} (4.26)

where $A$ is the cross-sectional area of the fluidic channel. $Q_{eo}$ is useful in determining the capacity of an electroosmotic pumping system.

Electroosmosis is possible due to the electric potential generated by the electric double layer formed at the channel walls in the presence of a solution. The flow velocity profile is very flat, which aids in maintaining high efficiencies in chemical separations. EO flow rates are directly proportional to the applied electric fields, and other factors that affect EO flow are listed in Table 4.1 as a reference.

**Table 4.1:** Relationship between electroosmotic velocity, $u_{eo}$, electroosmotic volumetric flow, $Q_{eo}$, and several dynamic factors.

<table>
<thead>
<tr>
<th>Relationship Variable</th>
<th>Common Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_{eo} \propto E$</td>
<td>volts per meter (V cm$^{-1}$)</td>
</tr>
<tr>
<td>$u_{eo} \propto \zeta$</td>
<td>volts (mV)</td>
</tr>
<tr>
<td>$u_{eo} \propto \text{pH}^*$</td>
<td>unitless</td>
</tr>
<tr>
<td>$u_{eo} \propto \frac{1}{\sqrt{M}}$</td>
<td>moles per liter (moles L$^{-1}$)</td>
</tr>
<tr>
<td>$u_{eo} \propto \lambda_D$</td>
<td>length (nm)</td>
</tr>
<tr>
<td>$D \not\propto u_{eo}$, but $Q_{eo} \propto D^2$</td>
<td>characteristic dimension length (µm)</td>
</tr>
</tbody>
</table>

$u_{eo}$ and pH are proportional up until a pH of about 6.0, after which $u_{eo}$ is approximately constant.
4.2.3 Electroosmotic Pump Theory

Electroosmosis is a powerful pumping technique which benefits from flat flow profiles. However, EO flow does suffer from some drawbacks, the greatest of which, perhaps, is the requirement of very high voltages. As a comparison, pressurized micropumps with diameters around 200 µm can easily generate flow rates of 10 µL min\(^{-1}\). To generate a similar flow rate using electroosmosis, voltages in excess of 10 kV would be required (assuming lengths around 10 cm). For small, portable systems, such high voltages are unfeasible.

One approach that utilizes the effects of electroosmosis while minimizing the large voltages required is the EO pump. In an EO pump, an electric potential is applied across one portion of the channel while there are no other body forces affecting the remaining channel segments. Such a pump is illustrated in Figure 4.5. \(L_1\) defines the total length over which an electric field is in effect, while the length \(L_2\) is not under the influence of any field. An EO flow profile, \(u_{eo}\), is initially generated in the first segment, and once it reaches the interface with the second segment, EO flow ceases. However, fluid is continuously being driven electroosmotically in the first segment, such that fluid is forced into the second segment. At this point, viscous forces resist the advancing fluid front, which results in a component of pressurized back-flow, \(u_b\), in the first segment. The difference between EO flow and back-flow results in Poiseuille flow, \(u\), through the system.

**Figure 4.5:** Single-channel EO pump. An electric field is applied across only the first section of length \(L_1\). The resulting EO flow, \(u_{eo}\), generates a pressurized flow, \(u\), in the second section of length \(L_2\). A back-pressure, \(u_b\), is also generated in the first section which acts to lower the overall translation of flow from the first segment into the second segment.
EO pumps are ultimately limited by the amount of back-pressure they generate. In order to reduce the back-flow, the first channel segment can be made much narrower than the second segment. This causes the hydraulic resistance of the first segment to be much greater than the second, such that most of the generated EO flow is translated into pressurized flow in the main channel. To understand how hydraulic resistance, \( R_h \), affects back-flow, consider an equation analogous to Ohm’s law,

\[
\Delta p = R_h Q, \quad (4.27)
\]

where \( \Delta p \) is the pressure drop across the channel and \( Q \) is the volumetric flow rate. Since back-pressure is Poiseuille in nature, Equation (4.27) can be plugged into Equation (4.13) to produce

\[
R_h = \frac{12 \eta L}{h^2 A}. \quad (4.28)
\]

Analogous to a conventional current resistor, cross-sectional area is inversely proportional to hydraulic resistance. In other words, as the channel becomes narrower, the hydraulic resistance increases dramatically. To illustrate this point, consider a cylindrical channel feeding into a second, larger channel of the same length, where the ratio of the cross-sectional areas between the two is 1:10. By applying Equation (4.28), the ratio of hydraulic resistance between the smaller and larger channel would be 10,000, which produces a very small back-pressure.

A straight-forward method of representing EO pump flow is to compare the ratio of the generated flow to the EO flow. The generated flow is the difference of the EO flow and the back-flow, \( u = u_{eo} - u_b \). By rearranging this equation and plugging in Equations (4.12) and (4.25), the ratio of \( u_{eo} \) to \( u \) is [53]

\[
\frac{u}{u_{eo}} = \frac{\frac{h_2}{h_1}^{\frac{h_2}{h_1}}}{\frac{h_2}{h_1}^{\frac{L_2}{L_1}} + \frac{L_2}{L_1}} = \frac{a^2}{a^2 + b}, \quad (4.29)
\]

where \( h_1 \) and \( h_2 \) are the first and second channel heights, respectively, and \( L_1 \) and \( L_2 \) are the first and second channel lengths, respectively, \( a = h_2/h_1 \), and \( b = L_2/L_1 \). This ratio is illustrated graphically in Figure 4.6, where the different curves represent various values of
The translation of EO flow into pressurized flow is maximized as the ratio \( L_2/L_1 \) is minimized and the ratio \( h_2/h_1 \) is maximized. In other words, an EO pump is most efficient when the electroosmotic channel segment is long and narrow.

While it is possible to generate highly translated pressurized flow using a single-channel electroosmotic pump, the volumetric flow rates will generally be too small for those required in realistic systems. To illustrate this point, consider channel segments of equal length where \( h_2/h_1 = 10 \). This leads to \( u/u_{eo} = 100/101 \approx 1 \), such that the majority of EO flow translates into pressurized flow. Now assume the channels are cylindrical and \( h_1 \) is equal to the diameter of the channel, \( d_1 \), where \( d_1 = 10 \, \mu \text{m} \) and \( E_z = 1000 \, \text{V cm}^{-1} \). The resulting volumetric flow rate is approximately 5.3 nL min\(^{-1}\). For microfluidic devices of this scale, the total channel volume can easily approach 1 \( \mu \text{L} \), such that it would take hours to flush an entire channel volume. In many cases, such time scales are unacceptably long.

One approach to increase the pumping capacity of an EO pump is to combine several small channels in parallel to form the electroosmotically-induced segment, which then feeds into the single, large channel. This is illustrated in Figure 4.7, where there are a total of \( n \) small channels combined in parallel. As an electric field is applied across this parallel-channel

![Figure 4.6: Ratio of generated pressurized flow versus electroosmotic flow, \( u/u_{eo} \), versus the ratio of channel segment lengths, \( L_2/L_1 \). The ratio approaches 1 as \( L_2/L_1 \) becomes small and \( h_2/h_1 \) becomes large.](image-url)
segment, each channel produces a flow rate directly proportional to the electric field, and each flow additively contributes to the main pressurized flow in the large channel. Each of the channels in the parallel set experience a pressurized back-flow, where the overall solution volume lost to back-flow is multiplied by \( n \). However, with appropriate planning of length and scale ratios between the two channel segments, back-flow can be greatly minimized while increasing the volumetric flow rate.

\[ \frac{u}{u_{eo - multi}} = \frac{h_2^2}{h_1^2} + n \frac{L_2}{L_1} = \frac{a^2}{a^2 + nb^2}. \] (4.30)

This has the net effect of increasing the role of the ratio \( L_2/L_1 \). This is shown in Figure 4.8, which plots the ratio \( u/u_{eo} \) versus the ratio \( L_2/L_1 \). Separate curves are shown for various numbers of parallel channels, \( n \). The efficiency of flow translation increases with \( L_2/L_1 \), as shown before, but efficiency now also increases with the number of parallel channels provided that \( h_2/nh_1 \) remains constant for all values of \( n \).

It would now be useful to extend the previous example of a single-channel EO pump to a multi-channel EO pump. Assume that \( h_2/h_1 = 10 \), and \( n = 10 \) parallel channels. This

\[Figure 4.7: \text{Multi-channel EO pump. An electric field is applied across only the first section of length } L_1. \text{ The resulting sum of EO flows, } u_{eo}, \text{ generates a pressurized flow, } u, \text{ in the second section of Length } L_2. \text{ A back-pressure, } u_b, \text{ is also generated in the first section, which acts to lower the overall translation of flow from the first segment into the second segment.} \]
leads to a translation ratio of \( u/u_{eo} = 100/110 = 0.91 \) and, assuming \( d_1 = 10 \) µm and \( E_z = 1000 \) V cm\(^{-1} \), the volumetric flow rate becomes 48 nL min\(^{-1} \), which is nearly \( n \) times the flow rate of the single-channel case. Although the efficiency dropped by about 9%, the volumetric flow rate increased by more than 800%.

To conclude the discussion of EO pump theory, it is useful to compare the flow rates of a multi-channel electroosmotic pump against a single-channel EO pump. For the purposes of this comparison, assume that the electroosmotic segment and the main channel segment possess the same cumulative cross-sectional area, such that \( nA_1 = A_2 \). Taking the quotient of Equations (4.30) and (4.29) results in

\[
\frac{u_{multi}}{u_{single}} = \frac{n \left( 1 + \frac{L_2}{L_1} \right)}{n + \frac{L_2}{L_1}} = \frac{n (1 + b)}{n + b}.
\]

(4.31)

This relationship is shown graphically in Figure 4.9. From the figure, it is apparent that the generated flow rate of a multi-channel pump compared to a single-channel pump increases with \( n \) and with an increasing ratio of \( L_2/L_1 \). This plot suggests that pumping efficiency is greater for larger values of \( L_2/L_1 \), but it must be remembered that this is a comparison against a single-channel pump and does not suggest that the highest flow rates are achieved by increasing \( L_2/L_1 \). In fact, by referring to Figure 4.8, it is evident that to maximize flow
rates, $L_2/L_1$ should be as small as possible, while larger values of $n$ allow larger values of $L_2/L_1$ without much loss of translational efficiency.

![Graph showing the ratio of multi-channel EO flow rate to single-channel EO flow rate versus total number of parallel channels in the multi-channel pump. For this plot, it is assumed that the cumulative cross-sectional area of all segments is equal, such that $nA_1 = A_2$.]

**Figure 4.9:** Ratio of multi-channel EO flow rate to single-channel EO flow rate versus total number of parallel channels in the multi-channel pump. For this plot, it is assumed that the cumulative cross-sectional area of all segments is equal, such that $nA_1 = A_2$.

### 4.2.4 Experimental Results

Microfluidic EO pumps have been studied in detail over the course of the last several years. Wang *et al.* [54] provides a thorough review of EO pumps, and Laser *et al.* [55] provides an in-depth review of several microfluidic pumping systems, including EO pumps. Single-channel EO pumps have been demonstrated with flow rates ranging from 6 pL min$^{-1}$ to 15 mL min$^{-1}$ and pressures generated as high as 33 kPa [32, 56, 57]. Other single-channel EO pumps have been integrated with chromatography systems to perform separations, resulting in theoretical plates as high as $N = 2000$ [58] and plate heights as low as $H = 500 \, \mu m$ [59]. Multi-channel pumps have been created that generate flow rates ranging from 20 nL min$^{-1}$ to 20 µL min$^{-1}$ and pressures ranging from 25 kPa to 551 kPa [53, 60, 61, 62, 63, 64, 65].
Another common implementation of the EO pump utilizes packed columns or monoliths. In the case of a packed column pump, a capillary or microfluidic channel is pressure-filled with a collection of micro-particles. EO flow takes place within the through-pores created by this packing technique. Flow rates ranging from 3.6 µL min\(^{-1}\) to 0.8 mL min\(^{-1}\) have been reported, and pressures between 14 kPa and 140 kPa have also been reported \[66, 67, 68\]. Another implementation similar to the use of packing particles is filling the channel with monolith. Tripp \etal\ [69] demonstrated the use of monolith discs to generate flow rates as high as 0.41 mL min\(^{-1}\) and pressures as high as 380 kPa.

A novel approach coupling the utility of electroosmotic pumping and micro-valves was demonstrated by Brask \etal\ [70], where an alternating field was used to actuate a set of valves in a setup very similar to a full-wave rectifying circuit. The result was steady flow at rates as high as 10 µL min\(^{-1}\) with pressures as high as 100 kPa.

EO pumps are fabricated using a variety of techniques, the most common of which is bonding. Benefits of bonding include high resolution when using well-established dry- and wet-etching techniques. Another advantage is that bonded channel fabrication can be a relatively fast process. Some disadvantages include the need for special equipment and conditions to bond the cover plate to the main substrate. Common complications that result from bonding include channel sag, which results in altered channel dimensions or complete channel collapse. Furthermore, bonding involves a step that cannot be easily integrated into an autonomous microfabrication process. The microfluidic channels fabricated in this work utilize sacrificial etching \[71, 72\]. A major benefit of sacrificial etching is that it is easily integrated into standard microfabrication processes. Another advantage is that the core height can be controlled to within a nanometer or so, depending on the deposition method chosen for the sacrificial material. The main disadvantage of sacrificial etching is the long times necessary to etch the core, which vary as the square of the etch length. However, the overall etch rate can be increased by including through-holes in the encapsulating layer which can be later filled through a deposition process to ensure total channel enclosure \[29\].

The work presented here employs sacrificially-etched microfluidic channels. The specific fabrication process for creating electroosmotic pumps is outlined in Figure 4.10. First, a layer of oxide is deposited to a thickness of approximately 200 nm over a silicon or glass
substrate using plasma-enhanced chemical vapor deposition (PECVD) (a). This first layer serves as a wettable surface for the bottom of the channel. A layer of approximately 200 nm of aluminum is then deposited using a thermal or e-beam evaporator (b). This layer of aluminum serves as a quick-etching section of the sacrificial core, which, when removed, greatly increases the exposed area of the second part of the sacrificial core. The increased surface area results in much shorter etching times for the second sacrificial core material.

Following the aluminum deposition, AZ3330 photoresist is spun over the top of the aluminum layer and is patterned to form the second sacrificial core (c-d). The photoresist core is then temperature-ramped from 90 °C to 250 °C (e). Under these conditions, the photoresist softens and refloows to minimize its surface area. The result is a shift from a rectangular to a rounded core with a peak height of approximately 3 µm. In the next step, the exposed aluminum is removed with a commercial aluminum etchant (Transene) at 50 °C (f). A second, thick layer of oxide is then deposited using PECVD, resulting in a top layer that is approximately 3 µm thick (g). This top layer serves as the channel structure and provides a wettable surface for liquid flow. Finally, the ends of the channels are exposed, and etchants remove the sacrificial cores (h). The first etchant is aqua regia, a mixture of hydrochloric acid and nitric acid heated to 90 °C. This chemical targets the thin aluminum and etches it very quickly. With all of the aluminum removed, the photoresist core is exposed all along its bottom area, which allows for rapid etching. The photoresist core is removed with a stable form of piranha called Nanostrip (Cyantek, Fremont, CA) heated to 50 °C. Once the photoresist core is etched away, a hollow, oxide microfluidic channel remains.

Single- and multi-channel EO pumps were fabricated using dual-core sacrificial etching. Figure 4.11(a) shows a completed device containing a single-channel electroosmotic pump. The area illustrated by the dashed rectangle indicates the typical testing area when measuring flow rates. In the case of a multi-channel EO pump, several identical microfluidic channels were set side-by-side to produce the pumping region. Figure 4.11(b) shows a multi-channel EO pump device nearly identical to that shown in Figure 4.11(a), where in this case the single-channel pump has been replaced with a multi-channel pump. The bright pump area in this image contains 75 parallel channels that are 6-µm wide and 4 mm long. The pump array feeds a single channel that is 50 µm wide.
Figure 4.10: Fabrication process to create EO pumps with dual sacrificial cores. (a) A thin layer of oxide is deposited on the top of a substrate, following which (b) a thin layer of aluminum is then deposited. (c) Photoresist is spun-on over the aluminum layer and (d) patterned to define the microfluidic channels. (e) The photoresist lines are reflowed by ramping a hot plate from 90 °C to 250 °C, resulting in rounded cores. (f) The exposed aluminum is etched away, and (g) a thick layer of oxide is deposited over the top of the cores. (h) Reservoirs are etched through the oxide to access the cores, and the sacrificial cores are etched with aqua regia to remove the aluminum and nanostrip to remove the photoresist. The end product is a hollow rounded-core microfluidic channel.

To further investigate microfluidic channels fabricated from sacrificial etching, scanning electron micrographs (SEMs) can be used to highlight channel details. Figure 4.12(a) is an overhead view of the small parallel pump arms interfacing with the larger, main channel. Figure 4.12(b) is a cross-sectional image of one of the pump channels. Notice the rounded channel shape generated from reflowing the sacrificial photoresist core. Figure 4.12(c) is a cross-sectional image of one of the large main channels. It is apparent that the main channel is wider than the pump channels, and its shape also takes on a rounded form due to reflowing the photoresist core. The heights of both channels are comparable, with the maximum height measuring approximately 3.5 µm at the center of the channel.

A method must be developed to measure the flow rate generated by the EO pumps. One common method is to include microparticles in the solution to act as tracers. The
difference in position of the tracer over time is used to determine the velocity of the fluid flow, and the volumetric flow rate can then be determined if the channel dimensions are known. The present work implements a simple measurement technique illustrated in Figure 4.13. Reservoirs are laser-cut from a sheet of PMMA and are attached over the microchannel inlets using epoxy. A pair of Teflon tubes with an inner diameter of 0.22 inches is attached to each reservoir. This allows solution injection through one tube and evacuation through the other. This ensures that the solution effectively reaches the reservoirs etched into the microfluidic device. In the case when only one injection tube is attached to the reservoir, air becomes trapped near the surface of the chip, effectively cutting off liquid flow to the system. With reservoirs attached and the solution injected, electrodes are inserted into one tube for each reservoir. In order to minimize electrolysis and bubble formation in the presence of high electric fields, palladium wires are used for the electrodes. Palladium possess the ability to reabsorb hydrogen gases that dissociate in a liquid under the influence of a large electric potential. A second method that aids in bubble minimization involves mixing equal concentrations (10 mM) of hydroquinone and p-benzoquinone. Hydroquinone was shown to be an effective bubble-suppressor in microfluidic devices [73], and aided in reducing bubble formation.
Figure 4.12: SEMs highlighting completed EO pumps. (a) Multichannel network of small pump channels feed into a single, larger channel. (b) The pump arms are approximately 6 µm wide, while (c) the large, main channel is approximately 50 µm wide. The height of each channel is approximately 3 µm.

formation in the work presented here. The fluid level in each of the four tubes is then recorded using a marker, and the experiment is ready to begin. Since the microfluidic channels are composed of oxide, a negative zeta potential exists between the solution and the channel walls. This results in EO fluid flow migrating toward the cathode. As such, the anode is in contact with the reservoir connected to the electroosmotic pump, while the cathode is connected to the end of the larger, single channel. When a voltage is applied, fluid begins its migration through the channels. When sufficient time has elapsed a noticeable change in fluid levels is apparent in the Teflon tubes. The difference in fluid levels can then be measured using a standard ruler, and the quotient of the displacement and the elapsed time
of the experiment produces a flow velocity. Since the inner diameter of the tubing is well characterized, it is a simple calculation to determine the volumetric flow rate.

![Diagram of experimental setup](image)

**Figure 4.13:** Experimental setup to measure flow rates generated by an EO pump. Buffer is injected through the tubing at both reservoirs. Palladium electrodes are inserted at each end and biased to result in electroosmotic flow from the multichannel pump segment to the single channel. As solution flows from one end to the other, displacement is measured using a ruler. Dividing this displacement by the elapsed time results in the total flow rate of the system.

Several microfluidic networks and electroosmotic pumps were tested. In each case, the microchannels were constructed from PECVD oxide. Substrates of silicon and Borofloat were both used, although there was no significant difference in results between channels built on either substrate. The solutions used include citrate buffer, phosphate buffer, and carbonate buffer, ranging in pH from 2.6 to 3.9, 4.8 to 8.3, and 9.2, respectively. Each buffer was mixed at a concentration of 10 mM, including 10 mM of hydroquinone and p-
benzoquinone added to suppress bubble formation. Voltage was applied using a step-up transformer, which allowed values as high as 3 kV. The equivalent electrical circuit for this test setup is illustrated in Figure 4.14. In this figure, $V_s$ is the DC source voltage, which is increased 100 times using a step-up transformer. The resulting amplified voltage is then applied across the microfluidic network, with ammeter $A$ and voltmeter $V$ monitoring the current through the network and the voltage across it.

![Figure 4.14: Circuit schematic representation of the test setup for the EO pump systems.](image)

The source voltage, $V_s$, is stepped-up by a transformer 100 times. This voltage is applied across the microfluidic network, which includes the EO pump and the main microfluidic channel. The ammeter $A$ monitors current levels through the microfluidic network, and the voltmeter $V$ measures the voltage drop across the channel.

Determining the voltage drop across the EO pump segment is an essential step to characterize EO pump flow rates. In order to accomplish this, consider the microfluidic network, represented by a resistor, in Figure 4.14. This network consists of the serial impedance presented by the fluid in the Teflon tubes, the fluid in the EO pump channels, and the fluid in the main microfluidic channel. Given the cross-sectional areas of each of these channel segments, it is possible to calculate a ratio of local resistance, $R$, to total resistance, $R_T$. These values vary from experiment to experiment depending upon which flow path is chosen for analysis in the microfluidic system, but typical values fall around $0.01R_T$, $0.08R_T$, and $0.91R_T$ for the Teflon tubes, the EO pump, and the main channel section, respectively. It is apparent that the majority of the voltage across the microfluidic system drops across the main channel segment, and not the EO pump. In order to ensure that all of the applied voltage is limited to the area of the EO pump, integrated electrodes would need to be in-
cluded at both ends of the pump. Even though the current configuration lacks this direct electrode integration, the EO pump greatly increases the volumetric flow rate compared to the rate achievable implementing only the main channel segment by itself. To illustrate this point, consider the following: compare channel $a$ of length $L$ with cross-sectional area $A_2$ to channel $b$, which consists of a multi-channel EO pump of length $L/2$ and cross-sectional area $A_1$ for each of $n$ channels in the pump, which connect to a larger single channel of length $L/2$ and cross-sectional area $A_2$ (refer to Figure 4.7). If a voltage, $V$, is applied across each of the aforementioned systems, from Equation (4.26) the following equations are produced,

$$Q_a = \frac{\mu_{eo} V A_2}{L},$$

$$Q_b = \frac{2\mu_{eo} \alpha V n A_1}{L},$$

where $\alpha$ represents the ratio of the total voltage, $V$, that is dropped across the multi-channel EO pump. In order to illustrate the benefit of an EO pump in this setup, consider the ratio of $Q_b$ to $Q_a$,

$$\frac{Q_b}{Q_a} = 2\alpha n \frac{A_1}{A_2}.$$

For the case of the microfluidic channels fabricated in this work, $Q_b/Q_a = 1.8$, which means that the flow generated using a multichannel EO pump increases the flow rate possible in a single channel of cross-sectional area $A_2$ by nearly two times, even when the voltage is dropped across the entire microfluidic system.

It should be noted that the main purpose of a multi-channel EO pump is to generate a pressurized flow through the main channel network. If, as in this case, an electric field exists across the entire microfluidic network, electrophoretic forces will separate analytes while the pump generates flow. If the separation mechanism is not intended to be electrophoretic, it is necessary to ensure that the electric field is confined solely to the multi-channel EO pump area.

Volumetric flow rates were measured using multi-channel and single-channel EO pumps. Figure 4.15 shows flow data for silicon dioxide multi-channel and single-channel EO pumps using a phosphate buffer at a pH of 8.3. The $x$-axis represents the electric field
dropped across the segment of interest. For the multi-channel EO pump, the data are plotted for the electric field across the pump section, while for the single-channel EO pump, the data are plotted for the electric field across the entire channel. The data represented in the plot were generated from two independently-tested multi-channel EO pump systems and two independently-tested single-channel EO pump systems. The right-hand side of the y-axis shows the calculated pressure generated by the multi-channel EO pumps.

![Graph](image)

**Figure 4.15:** Volumetric flow rate of silicon oxide EO pumps versus applied electric field. The applied electric field corresponds to the field across the multi-channel pumps (○) and single-channel pumps (□). The solution used for these experiments was phosphate buffer at a concentration of 10 mM and a pH of 8.3.

Volumetric flow rates were also measured for pump systems fabricated using silicon nitride instead of oxide. To create nitride microchannels, the fabrication steps are identical to those shown in Figure 4.10, with the difference that the deposited oxide layers in steps (a) and (g) are replaced with PECVD nitride. Figure 4.16 shows flow data for silicon nitride multi-channel and single-channel EO pumps using a phosphate buffer at a pH of 8.0. The experimental flow rates are slightly lower than those shown in Figure 4.15, which suggests
that nitride walls produce a slightly lower zeta potential. Again, the right-hand side of the $y$-axis shows the calculated pressure generated by the multi-channel EO pumps. Figure 4.17 shows more experimental results for nitride EO pump systems, but in this case, a carbonate buffer at a pH of 9.2 is used. With the increase in pH compared to the experiment represented by Figure 4.16 comes a slight increase in flow rate values. Again, the right-handed $y$-axis shows calculated pressures generated by the multi-channel EO pumps.

![Figure 4.16: Volumetric flow rate of silicon nitride EO pumps versus applied electric field. The applied electric field corresponds to the field across the multi-channel pumps (○) and single-channel pumps (□). The solution used for these experiments was phosphate buffer at a concentration of 10 mM and a pH of 8.0.](image)

The zeta potential for oxide and nitride surfaces, along with several buffers at different pH values, can be calculated using experimental flow rates. Figure 4.18 shows the calculated zeta potential for various values of pH. As is expected, the magnitude of the zeta potential increases with increasing pH, which results in higher flow rates. This curve remains fairly linear up until a pH of about six. For increasing pH values beyond this point, the overall calculated zeta potential remains fairly constant. The plotted curve suggests that if the
Figure 4.17: Volumetric flow rate of silicon nitride EO pumps versus applied electric field. The applied electric field corresponds to the field across the multi-channel pumps (○) and single-channel pumps (□). The solution used for these experiments was carbonate buffer at a concentration of 10 mM and a pH of 9.2.

pH were lowered to values under two, the zeta potential would eventually become positive, resulting in electroosmotic flow moving in the opposite direction (i.e., from cathode to anode). The figure also illustrates the difference in calculated zeta potential for oxide channels (hollow shapes) and nitride channels (solid shapes). It should be noted that the zeta potentials calculated here exhibit significantly higher magnitudes for nitride films than those measured by Bousse and Mostarshed [74]. This discrepancy could be explained by the presence of a higher concentration of silanol sites on the surface of the PECVD nitride than exhibited in the LPCVD nitride studied by Bousse and Mostarshed. Three different buffers are represented in Figure 4.18, including citrate buffer, phosphate buffer, and carbonate buffer, all mixed at a concentration of 10 mM.

Electroosmosis is a powerful force for pumping fluids through microfluidic channels. EO flow produces a flat flow velocity profile, which is particularly desirable for separation systems. EO flow can be used to generate very efficient Poiseuille flow in a microfluidic network. EO pumps can be created by interfacing several, small parallel channels to a
Figure 4.18: Plots of calculated zeta potential versus pH. The magnitude of zeta potential is seen to increase with increasing pH. The solid shapes represent nitride channels, while the hollow shapes represent oxide channels. The solutions used were citrate buffer (○), phosphate buffer (□), and carbonate buffer (♦).

4.3 Capillarity

Capillary action, or capillarity, is a very attractive filling mechanism because it requires no external forces. This eliminates the need for external power sources which would normally be required to drive the fluid. The downside to this method is that it is short-lived – once a channel has filled for the first time, capillary action ceases and further fluid flow must be controlled by another mechanism.

Capillary action occurs due to the interaction of surface energies between the solution and the channel walls. The surface energy of the solid-gas interface must exceed the sum of the surface energies of the liquid-gas and solid-liquid interfaces to pull the liquid through the channel. This pulling action continues only as long as a liquid-gas interface exists within the channel, such that once the channel has filled, capillary action ceases. Due to the pulling
action of the walls, the flow velocity profile assumes a concave shape, as shown in Figure 4.19. The channel height is represented by $h$, and the resulting pressure that is generated across the meniscus is represented by $\Delta P$. Due to the shape of the meniscus, capillary flow results in a negative pressure.

Figure 4.19: Capillary flow is generated due to the interaction of surface energies at the channel walls, which in effect “pulls” the fluid through the channel. The resulting flow velocity profile is concave, where the maximum flow rate occurs along at the channel walls and the minimum flow velocity occurs at the center of the channel.

4.3.1 Derivation

Modelling capillary flow can be done by utilizing the Navier-Stokes equations, such as has been shown previously for Poiseuille and EO flow. Although the meniscus shape resulting from capillary flow is inverted from that of Poiseuille flow, capillary flow is still a pressure-generating mechanism. As such, we can begin with Equation (4.12), which represents the average flow velocity of a pressurized fluid in a planar channel. However, in the case of capillary filling, flow velocity, $u_{c-avg}$ will change along the length of the channel, $L$, such that $u_{c-avg} = f(z)$. Replacing $L$ with $z$ in Equation (4.12) leads to

$$ u_{c-avg} = -\frac{h^2 \Delta p}{12 \eta z} \quad (4.32) $$

For capillary filling, the pressure due to the surface tension between a liquid-gas interface can be expressed by the Young-Laplace equation as

$$ \Delta p = -\frac{2\gamma}{R} \quad (4.33) $$
where $\gamma$ is the surface tension between the liquid-gas interface, and $R$ is the principal radius of curvature of the meniscus. The negative sign in Equation (4.33) arises because the radius of curvature for a meniscus formed by capillary pressure will be concave. The radius of curvature can be expressed in terms of the channel height, $h$, by considering the liquid-solid contact angle, $\theta$. To understand this relationship, refer to Figure 4.20, which shows the meniscus of a fluid filling a channel due to capillary action. Through simple geometry it is apparent that $R = h/(2\cos \theta)$. Plugging this value back into Equation (4.33) results in

$$\Delta p = -\frac{4\gamma \cos \theta}{h}.$$  (4.34)

![Figure 4.20: Relating radius of curvature, $R$, to channel height, $h$, and fluid contact angle, $\theta$. The relationship shown here is $R = h/(2\cos \theta)$.

The governing equation that describes capillary flow in channels is called the Washburn equation. To derive this equation, it is first necessary to describe the average flow velocity due to capillary filling, which is done by combining Equations (4.34) and (4.32),

$$u_{c-avg} = \frac{\partial z}{\partial t} = \frac{h\gamma \cos \theta}{3\eta z}.$$  (4.35)

Integrating Equation (4.35) with respect to $t$ results in $z = h\gamma t \cos \theta/3\eta z + C_1$. By setting $t = 0, z|_{t=0} = 0$, such that $C_1 = 0$, which leads to the Washburn equation for planar channels [75],

$$z^2 = \frac{h\gamma t \cos \theta}{3\eta},$$  (4.36)
which describes the advancement of the liquid meniscus position over time due to capillary forces. From Equation (4.36), it is apparent that the velocity of the fluid increases with increasing channel height and surface tension, as well as increasing with decreasing contact angle and fluid viscosity.

### 4.3.2 Experimental Results

Capillary filling has been exploited over several years for microfluidic devices. In more recent years, nanofluidics has emerged as a promising technology, and capillarity has been employed as one of the primary driving forces for such systems [76, 77]. Nanofluidic channels have been successfully created using various methods such as bonding [8, 10], shadow deposition [17, 20], imprint lithography [25, 26], and sacrificial etching [29, 31, 36, 37, 38], with heights ranging from 10 nm to 250 nm. Many of these channels have been utilized to characterize capillary filling on the nanofluidic scale.

In this treatment, sacrificial etching was the method employed to fabricate planar nanochannels [78]. These fabrication steps are illustrated in Figure 4.21. A substrate is initially coated in a thin layer of PECVD silicon dioxide, usually to a thickness of about 200 nm (a). This layer is important in studying capillary action, because oxide is a highly wettable film, and this layer will form the bottom wall of the nanochannels. A sacrificial layer is deposited to the thickness of the desired channel height (b). In this treatment, aluminum, germanium, and chromium have been used as sacrificial layers. These layers were deposited using evaporation with a thermal or electron-beam evaporator. These materials were chosen due to high deposition resolution as well as the availability of highly-selective etchants to remove the sacrificial cores. In the next fabrication step, a masking photoresist layer is patterned and developed, and the exposed sacrificial material is removed (c). The masking photoresist is then stripped away, leaving behind patterned sacrificial cores. The cores are then enveloped in a 5 μm-thick layer of silicon dioxide (d). This top layer will serve as the structural layer of the nanochannels, and since it is composed of oxide, it will also maximize the available capillary forces. Masking photoresist is patterned and developed to expose reservoirs over the channel ends. These reservoirs are etched into the oxide using a buffered oxide etchant (BOE), following which the masking photoresist is removed (e). The
sacrificial cores are etched away using an appropriate etchant. For aluminum, the etchant used was aqua regia (hydrochloric acid and nitric acid at a ratio of 2:1) heated to 90 °C. For germanium, the etchant was hydrogen peroxide heated to 50 °C. For chromium, the etchant was a commercial chromium etchant heated to 50 °C. Once the cores are removed, hollow planar nanochannels remain.

![Fabrication process to create planar nanofluidic channels.](image)

**Figure 4.21:** Fabrication process to create planar nanofluidic channels. (a) A thin layer of oxide is deposited on a substrate. (b) A sacrificial material is deposited at a thickness that translates to the height of the nanochannel. (c) Cores are patterned from the sacrificial layer. (d) A thick layer of oxide is deposited over the top of the cores. (e) The ends of the cores are exposed by etching reservoirs into the oxide. (f) The exposed cores are then removed using an appropriate etchant, resulting in hollow nanofluidic channels.

A micrograph of completed planar nanochannels with heights of 88.5 nm are shown in Figure 4.22. These channels were fabricated using aluminum as the sacrificial core. This micrograph shows empty channels along with reservoirs visible at the bottom of the image. The reservoir is the point where the channel ends are open to air, and this is the point where solutions can be introduced. The bottom of the reservoir is mostly silicon coated with a thin layer of naturally occurring oxide. Each of the channels in Figure 4.22 is approximately 15 µm wide, 1.2 mm long, and there are 200 parallel channels per device. These parallel
channels can be filled and measured simultaneously to statistically characterize capillary filling in nanochannels.

Figure 4.22: Microscope image of 88.5-nm tall channels made from aluminum cores. There are a total of 200 parallel channels per chip. A reservoir is visible spanning the width of the figure at the bottom of the image.

To determine that the sacrificially-etched nanochannels are intact and have resulted in the expected height, it is useful to capture scanning electron micrographs of the channel cross-sections. Figure 4.23 shows an SEM of the cross-section of a nanochannel 85 nm in height. Since sacrificially etched nanochannels are planar, the width of the devices is much greater than the height. As such, the channel shown in Figure 4.23 is a thin, long slit appearing toward the middle of the image. From this figure it is apparent that the channel is intact and at no point along its width does the channel collapse in on itself. Furthermore, the height of the channel is fairly constant all along its width. These are important factors to experimentally determine when recording accurate capillary flow measurements.

Capillary filling for planar channels was explained previously and is modeled by the Washburn equation, Equation (4.36). This model has been widely accepted to predict capillary flow behavior in microchannels, and in that role it has proven to be very effective. However, strange phenomena result in increasingly larger deviations from the model as the
channel dimensions shrink to tens of nanometers and smaller. In almost all cases reported by researchers, there is some range of channel heights where capillary filling speeds decrease significantly from what Washburn’s model would predict, and several explanations have been offered to account for this behavior. Hibara et al. noticed a slowing in predicted capillary flow for nanochannels with heights of 330 nm [15]. They attributed this behavior to an apparent increase in fluid viscosity. A couple of years later, Tas et al. took this idea of apparent viscosity increase one step further to suggest the phenomenon at play was the electroviscous effect [79]. Electroviscosity is attributed to the build-up of ions at the meniscus front, as ions within the mobile layer of the electric double layer are pulled along by capillary forces. This charge imbalance results in a streaming potential which generates EO flow that opposes the direction of the capillary filling. The net effect is that the liquid fills more slowly than it otherwise would by capillary forces alone. This phenomenon is pronounced in nanochannels, because for many solutions at this scale, channels will possess dimensions on the same order as the Debye length. The idea of the electroviscous effect slowing capillary filling of nanochannels was further perpetuated by later studies [11, 80, 81, 82]. However, in spite of the feasibility of the electroviscous effect, most researchers agree that the magnitude of the effect cannot completely account for the large departure from continuum theory. Either there are other effects at play or there are several effects, including the electroviscous effect, that need to be considered.
Another aspect that results in departure from the Washburn equation is channel top layers that are thin enough to dynamically deform under high capillary pressures. Tas et al. characterized the high negative pressures that are generated in nanochannels during capillary filling [83], and later demonstrated the extent to which nanochannels would bow and eventually collapse for various top layer thicknesses [9, 84, 85]. Dynamically-deforming nanochannels change the nature of capillary filling, since the dimensions of the channel change during filling. To ensure accurate measurements for a given channel height, the top layer of the channel must be sufficiently thick to prevent deformation.

Bubble formation inside nanochannels has been another reason attributed to slower filling speeds. A theoretical model of bubble formation in nanochannels was pursued by Nagayama et al., and their results suggested that bubble nucleation in nanochannels departs from continuum models and depends strongly on channel surfaces [86]. They also found that very round bubbles formed on hydrophilic surfaces, and low-density liquids spontaneously formed larger bubbles than high-density liquids. In other words, spontaneous bubble formation during capillary filling of nanochannels is plausible. An experimental capillary filling experiment demonstrating bubble formation was demonstrated by Thamdrup et al., where they showed that bubbles pinned inside nanochannels increased the hydraulic resistance and, thus, resulted in slower filling speeds than continuum theory predicts [87].

A fourth effect that has been considered to contribute to deviations in the Washburn model for nanochannels is the dynamic contact angle. Macroscopic contact angles of water on oxide surfaces rarely produce a $\theta$ greater than $20^\circ$. However, little is known about the contact angle that occurs during filling, which is referred to as the dynamic contact angle. The dynamic contact angle forms due to instantaneous force balancing that occurs at the meniscus during capillary filling, which results in contact angles that are greater than macroscopically-measured contact angles. Han et al. reported water capillary filling speeds slower than expected for channels 50 nm in height [18], and they attributed this to a high dynamic contact angle of $68^\circ$. Zhu and Petrovic-Duran showed that there is a significant departure in dynamic contact angle for various microchannels with various geometries [88]. However, there appears to be little or no correlation between dynamic contact angles and
channel sizes—it is only apparent that the dynamic contact angle for small channels is much larger than the macroscopic contact angle.

After all of the examples of experiments demonstrating slowing of capillary filling in nanochannels, it is worthwhile to note that there has also been at least one report that suggests the opposite behavior. Supple and Quirke demonstrated water filling in carbon nanotubes with diameters around 5 nm [89]. They showed that capillary filling speeds exceeded what the Washburn model would predict for round nanotubes. Furthermore, the meniscus advanced linearly with time, opposed to the square of the meniscus position as suggested by the Washburn equation. They attributed this behavior to single-molecule effects at this scale. With all of the reports on capillary filling in nanochannels, the general consensus is as follows: no one seems to know exactly what is happening.

Presented here are capillary filling experiments performed using nanochannels fabricated via sacrificial etching. Although other researchers have fabricated nanochannels using sacrificial-etching methods, this is the first work that addresses capillary filling behavior in sacrificially-etched nanochannels. Capillary filling tests were performed by introducing a minute drop of DI water with a syringe into one of the reservoirs on a nanofluidic device. A microscope-mounted CCD camera then captured video of the meniscus advancement over time. The recorded data were analyzed frame-by-frame using computer software to track and record the meniscus advancement. Figure 4.24 shows capillary filling of a nanochannel device with a height of 88.5 nm. The main thing to note in Figure 4.24(a) is the difference in contrast between filled channel segments and empty channel segments. In the picture, the top half of the channels are filled with water, and appear dark in color. Below that line, the channels are empty, and appear bright in color. Figure 4.24(b) shows the meniscus position squared, $x^2$, versus elapsed time for channels with heights of 88.5 nm and 19.5 nm. Each of these curves represents the average of some 50 independently measured channels. As is illustrated in this plot, the curves are approximately linear, as predicted by the Washburn equation ($x^2 \propto t$). Similarly, taller channels fill more rapidly than shorter channels ($x^2/t \propto h$).

Even though the shape of the curves in Figure 4.24(b) is nearly linear as predicted by the Washburn equation, it is necessary to evaluate the actual data points compared
**Figure 4.24:** Transient capillary filling of nanochannels. (a) Water filling channels that are 88.5-nm tall. The top half of the image shows filled channels (dark in color), while the bottom half of the image shows empty channels (light in color). (b) Meniscus position squared versus elapsed time for 88.5-nm tall channels and 19.5-nm tall channels.

to the Washburn equation to see how closely they match. Figure 4.25 demonstrates the filling behavior for channels of various heights fabricated from aluminum, chromium, and germanium sacrificial cores. The factor $x^2/t$ can be thought of as a measure of the capillary filling speed. The dashed line represents the Washburn equation for a perfectly wettable
channel, i.e., $\theta = 0$. It becomes quickly apparent from the plot that the experimental data points all lie well below the theoretical curve. Given the history of nanochannel capillary filling experiments outlined earlier, this behavior is not unexpected. Another thing to note from this figure is that there appears to be little or no difference in filling behavior between one sacrificial core and another.

**Figure 4.25:** Capillary flow speeds, $x^2/t$, versus height for various nanofluidic channels. Three different sacrificial cores are represented in this figure, including aluminum (Al), chromium (Cr), and germanium (Ge). The dotted line represents the theoretical curve for perfectly wettable channels. As channel dimensions decrease, the experimental data fall further and further below the ideal line.

To gain a greater understanding of the deviation of the experimental data points from the theoretical model, Figure 4.26 plots the same points in Figure 4.25, but in terms of the dynamic contact angle. A new data point is also added, which represents a microfluidic channel fabricated using an SU8 sacrificial core. This plot assumes that the contact angle is dynamic and changes with channel height, although the data could have also been plotted against apparent viscosity. From this plot, there is a decreasing trend in dynamic contact angle as the channel height increases. This matches most trends witnessed by previous researchers using capillary action to fill nanochannels. However, this trend appears to be
more exaggerated than in other reports, with the exception of a close match with results published by Han et al. for water-filling experiments [18]. Another interesting point to note in Figure 4.26 is that the microchannel, with a height exceeding 5 µm, exhibits a large dynamic contact angle of 47°. From the data points, it appears that the dynamic contact angle levels off with increasing channel height, such that the experimental results of capillary filling speeds of sacrificial channels will always fall somewhere around 70% of the theoretical value or lower. Regardless of the effect responsible for the departure from continuum capillary theory, it is important to characterize nanochannel filling speeds to understand how a fluidic system will be affected.

![Figure 4.26: Calculated dynamic contact angle versus channel height for various channels. Channels are made from aluminum (Al), chromium (Cr), and germanium (Ge) sacrificial cores. It is apparent that the dynamic contact angle becomes increasingly larger with decreasing channel heights. The curve representing dynamic contact angle levels off around 45°, even with channels as tall as 5 µm.](image)

Several initial capillary filling experiments exhibited odd behavior, such as an elongated meniscus where filling occurred faster along the channel ends than at the channel center. Other channels filled much slower than the results shown in Figure 4.25. One of the effects mentioned previously that alters capillary flow behavior is an elastic channel top
layer, and this elastic layer was shown to affect sacrificially-etched nanochannels. Tas et al. outlined an equation that determines the critical width, $w_c$, of a channel, which is the point where the top of the channel flexes sufficiently to come into permanent contact with the bottom of the channel [85]. This equation can be represented as

$$w_c = \frac{8.16h_o^2 f(u)D}{\gamma \cos \theta}, \quad (4.37)$$

where $h_o$ is the center channel height, $f(u)$ is the plate-stiffening number, which for water is around 3.2, $D$ is the flexural rigidity of the channel, which for oxide is typically around $6.1 \times 10^{-11}$ N m, $\gamma$ is the dynamic surface tension between water and air, and $\theta$ is the contact angle. Figure 4.27 is a plot of theoretical curves for the critical channel width versus channel height generated from Equation (4.37). The curves represent channels with top oxide layers with thicknesses of 1 µm, 3 µm, 5 µm, and 7 µm. As shown in the figure, taller channels can be wider than shallower channels without experiencing collapse. It is important to design a channel such that the value of the top layer thickness lies well above the curve of collapse for a channel with a given width and height. If the point lies just above the curve of critical collapse, the channel will not be destroyed when it fills, but the top layer will bend significantly. Therefore, the top layer must be adequately thick so that it remains rigid during capillary filling. Table 4.2 lists some experimentally-determined capillary filling speeds, $x^2/t$, for various thicknesses of top oxide layers. From this table, it is shown that thinner oxide top layers resulted in slower channel filling presumably due to dynamic channel deformation. At top layer thicknesses of 5 µm and 7 µm, the ratio of experimental values is approximately equal. As such, all nanofluidic channels presented previously were fabricated with oxide top layers 5 µm thick to prevent dynamic channel deformation.

Filling mechanisms are a very important aspect of all microfluidic and nanofluidic channels. Without a method to pump fluids through a channel, any separation technique is rendered useless. Three main filling mechanisms of microfluidic and nanofluidic channels include Poiseuille flow, EO flow, and capillary flow. By understanding filling mechanisms, they can be employed in various ways to achieve the desired research goal.
**Figure 4.27:** Theoretical critical channel width versus channel height. Curves are shown for different thicknesses of top oxide layers, including 1 µm, 3 µm, 5 µm, and 7 µm. The y-axis illustrates the point at which the top channel layer is sufficiently flexible that it will come into permanent contact with the bottom of the channel.

**Table 4.2:** Filling speeds of nanochannels, \( x^2/t \), for varying thicknesses of oxide top layers. The height for each channel in this table is 56.5 nm, and the sacrificial core was germanium.

<table>
<thead>
<tr>
<th>Oxide Thickness</th>
<th>( x^2/t )</th>
<th>Exp. to Theo. Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 µm</td>
<td>0.516 mm(^2) s(^{-1})</td>
<td>38.1%</td>
</tr>
<tr>
<td>5 µm</td>
<td>0.845 mm(^2) s(^{-1})</td>
<td>62.3%</td>
</tr>
<tr>
<td>7 µm</td>
<td>0.848 mm(^2) s(^{-1})</td>
<td>62.5%</td>
</tr>
</tbody>
</table>
Chapter 5

Nanochannel-Based Separations

Just as the trend in microelectronics leads to increasingly smaller devices, so microfluidic devices evolve toward the nanofluidic regime. Nanofluidic channels are defined as devices whose characteristic dimension lies in the range from 1 nm to 100 nm. Channels with characteristic dimensions that lie between 100 nm and 1 µm are typically referred to as sub-micron channels, although in many instances the term nanofluidics is used for this range as well. Nanofluidics differ from microfluidics in their nano-scale dimensions and, as such, offer some unique benefits as well as suffer from some unique disadvantages compared to microfluidics [90]. Some of these advantages include the use of even smaller sample volumes than those used in microfluidics. Another advantage is the ability of nanochannels to employ channel dimensions and features in size-based separations [91]. Some challenges include developing fabrication methods that are reproducible, robust, and fast. Furthermore, as channel dimensions decrease, several well-founded pumping techniques become ineffective. For example, electroosmosis becomes negligible as electric double layers begin to overlap and reduce the net charge that causes electroosmotic flow [92, 93]. Pressurized pumping is difficult to implement because nanofluidic channels generate such high hydraulic resistances, such that impossibly high pressures must be used in order to generate appreciable flow [92]. Therefore, there is a need to research and create systems that take advantage of the benefits of nanofluidics while taking the necessary steps to avoid the drawbacks.

Nanochannel cross-sectional profiles can be divided into two main categories, i.e., non-planar and planar. Non-planar methods produce nanochannels with cross-sectional widths and heights that are approximately on the same scale, while planar nanochannels typically have one cross-sectional dimension that is many times larger than the other dimension. For non-planar profiles, lithographic techniques define lateral dimensions that are nanometers
Lithographic dimensions in the nanometer range are possible using techniques such as extreme ultraviolet (EUV) lithography, X-ray lithography, and electron beam lithography (EBL). However, each of these systems suffers from drawbacks, including high tooling costs and slow operation [22, 23]. In contrast, planar nanochannels rely on standard thin film processes which are inexpensive, fast, and can define the channel height to within several nanometers.

Most nanofluidic separations rely on electrophoresis, porous membranes, networks of micro-pillars, or variable-height nanofluidic channels [94]. Electrophoretic separations have been demonstrated in capillaries ranging in diameter from 770 nm to 270 nm [30, 95]. However, when the thickness of the electric double layer is on the order of the characteristic dimensions of a nanofluidic channel, the flow profile becomes highly nonuniform and the nature of the separation deviates from classical electrophoresis. [96, 97] Another separation medium, membranes, has been fabricated with pores ranging in diameter from 260 nm to 5 nm, [98, 99, 100] but the membrane can easily break when pressurized. Arrays of micro-pillars [28, 101] and variable-height nanofluidic channels [7, 102] have been shown to successfully sieve and separate DNA. However, these separation methods are limited to entropic trapping and reptation, which limits the scope of the separation to analytes which can easily deform.

The purpose of this chapter is to familiarize the reader with a few of the common nanofluidic-specific separation techniques available. Specifically, methods that directly rely upon nanochannel dimensions to achieve separation will be discussed. Most nanofluidic separation systems are composed of either (1) arrays of micro- or nano-pillars or (2) networks of nanochannels. By far the most common analyte separated by nanofluidic systems is DNA [103]. Other nano-scale biological molecules that are candidates for nanofluidic separations include proteins and viruses. Figure 5.1 illustrates the most common nanofluidic separation mechanisms. They include Ogston sieving, entropic trapping, reptation, electrostatic sieving, and immutable trapping. Each of these methods operates according to different principles, and specific methods can be used to target different analytes. Several reviews have been authored to elucidate the differences and advantages of each of these approaches [94, 104, 105, 106, 107]. The following sections briefly treat each of the methods shown in Figure 5.1. The final section explores immutable trapping experiments performed at BYU.
Figure 5.1: Nanofluidic separation mechanisms. (a) Ogston sieving is dominated by steric interactions between the analytes and the channel walls. In this regime, \( R_g < h \). (b) Entropic trapping relies on analyte herniation overcoming an entropic energy barrier. In this regime, \( R_g \sim h \). (c) Reptation involves unraveling a polymer chain, such as DNA, and separation occurs depending on the total strand length. In this regime, \( R_g > h \). (d) Electrostatic sieving relies on relatively thick electric double layers that act to repel ions of equal sign and pass those of opposite sign. In this regime, \( \lambda_D \sim h \). (e) Immutable trapping relies on permanently capturing a molecule at a point where the diameter of the analyte exceeds the dimensions of the nanofluidic channel. In this regime, \( h_n > D_n > h_{n+1} \).

5.1 Ogston Sieving

Ogston sieving is the oldest and best understood of the nanofluidic separation mechanisms. It is named after Alexander Ogston (1911–1996) for the work that he performed studying steric interactions of spherical particles in suspended fibers [108]. Ogston sieving relies on steric interactions between analytes and the walls of the nanochannels to perform
sample separation. This method is illustrated in Figure 5.1(a). The top half of the figure
demonstrates Ogston sieving in a pillared array, whereas the bottom half illustrates Ogston
sieving in a variable-height nanochannel. In a pillared array, the radius of gyration, $R_g$,
of the analyte will be less than the distance between the pillars. Most molecules analyzed
in this regime are considered fairly rigid, and do not deform significantly when interacting
with other bodies. Retention is possible as molecules pass near a pillar and experience steric
repulsion. The effect is similar for an analyte that migrates from one channel into a sec-
ond, smaller channel. Due to steric repulsion, analytes will be momentarily retained at the
entrance to the second channel. For Ogston sieving, analytes elute from smallest to largest.

Ogston sieving has been demonstrated in several experiments. One such experiment
utilized a colloidal bed suspension developed by Zeng et al. \cite{109}. In their experiment,
silica particles were suspended inside a microfluidic channel to create void lengths ranging
between 24 nm and 135 nm. This system separated DNA strands between 50 base pairs (bp)
and 50 kbp, as well as separated proteins with masses ranging from 20 kDa to 200 kDa. In
addition, the authors demonstrated the transition from Ogston sieving to entropic trapping
with high-bp DNA strands.

Another advance in understanding Ogston sieving came from Li et al., when they
generated a theoretical model to fit Ogston sieving data \cite{110}. This model was created from
a periodic array of dual-height nanochannels, which is a system used extensively by Fu et
al., who demonstrated DNA and protein separation using channels ranging in heights from
40 nm to 180 nm \cite{111}. In these experiments, they separated DNA strands between 50
bp and 766 bp, as well as proteins with masses between 11.4 kDa and 179 kDa. Later,
they demonstrated the transition from Ogston sieving to entropic trapping for DNA strands
between 500 bp and 8 kbp \cite{102}, as well as created a continuous-flow system for separating
DNA strands \cite{112}.

Nanopillars have also been used to affect Ogston sieving. Baba et al. created arrays
of nanopillars spaced 250 nm apart \cite{91}. This system separated DNA strands between 2 kbp
and 10 kbp in length with the smallest strands eluting first. They also fabricated a system
with micro-scale ridges through which only the smallest strands could pass, while flow was
directed through a microchannel segment tangential to the length of the channel between the
ridges [91]. The result was that larger strands eluted first because they could not pass into the channel between the ridges, while smaller strands pass through the constricted channels and were subsequently slowed through steric interactions.

5.2 Entropic Trapping

Entropic trapping is a separation mechanism that relies on the movement of an analyte through an opening that is narrower than the incident analyte. Therefore, this mechanism can only be used with analytes that can deform or herniate. For example, because of the coiled structure of DNA, it proves to be an excellent candidate for entropic trapping. Figure 5.1(b) illustrates the principle of entropic trapping. In the top half of the figure, entropic trapping takes place within a pillared array, whereas the bottom half of the figure illustrates entropic trapping in a variable-height nanochannel. Entropic trapping operates on the principle of an entropic energy barrier. While an analyte resides in the large channel segment, it possesses a high level of entropy. As the analyte proceeds toward the entrance to the small channel segment, it must deform in order to advance. For this to happen, the analyte must overcome an entropic free energy barrier. When this occurs, the analyte can deform to enter the small channel, and will rapidly elute until it reaches another large-channel region, where it will again be retained until it gains sufficient energy to overcome the next entropic barrier. In this separation scheme, the largest analytes will elute first, and the smallest last. This elution order initially seems counter-intuitive, and is attributed to the high surface area of a large analyte that is available at the entropic barrier, which increases the probability of overcoming the barrier per unit time.

As was mentioned in the previous section, it is common to transition between Ogston sieving and entropic trapping within the same system, provided that a sufficiently large range of analyte sizes is used [102, 109, 112]. Han et al. developed several separation systems that utilized entropic trapping [7, 113]. They fabricated an array of entropic traps ranging in heights anywhere from 3 μm for the tall segments to 75 nm for the short segments. They demonstrated the ability to separate DNA strands with lengths between 37.9 kbp and 164 kbp. They also developed a system that would pre-concentrate the DNA using low electric
fields, which hindered the ability of analytes to overcome the initial entropic barrier. After concentration was complete, high electric fields were applied to drive the separation [114].

Entropic studies have been performed to determine the response of DNA as it uncoils and is subsequently allowed to relax back to its lowest energy state. Turner et al. [115] tested DNA recoil by stretching the polymer through a region of nanopillars. A similar experiment was performed by Mannion et al. [116], except nanofluidic channels were used instead of nanopillars to unravel the DNA.

5.3 Reptation

Reptation is an extreme condition of entropic trapping. As a polymer overcomes the entropic barrier and migrates into a much smaller secondary channel, the strand will completely uncoil. The ensuing motion of the stretched polymer chain is similar to a snake slithering across the ground, hence the usage of the term reptation. Figure 5.1(c) illustrates the principle of polymeric reptation. The top half of the figure shows reptation occurring through a series of closely-spaced nanopillars, while the bottom half of the image illustrates reptation occurring in a sufficiently shallow nanochannel. When a system utilizes nanopillars, it is highly probable that long analyte strands will entangle around pillars as they elute. For this reason, it is often necessary to apply alternating voltages across orthogonal vectors to enable long analyte strands to migrate through the system. Separation order corresponds to the length of the strand, where short strands will elute first and long ones will elute last.

The most common structures for effecting reptation are arrays of micropillars and nanopillars or nanopillars. Turner et al. [28] demonstrated the ability to separate DNA strands between 7.2 kbp and 43 kbp with 100-nm wide nanopillars. Huang et al. [117] separated strands of DNA between 61 kbp and 201 kbp with micropillars 2-µm wide. They also applied alternating electric fields at angles of 120° with respect to each other to disentangle and move long DNA strands. Kaji et al. [101] developed a system consisting of periodic arrays of empty channel segments adjoining segments containing nanopillars, and they used this system to separate DNA strands between 48.5 kbp and 165.6 kbp in length. Gu et al. [13] fabricated a system where channel segments containing nanopillars immediately fed into
arrays of nanochannels. The nanopillars were used primarily to begin uncoiling the DNA strand, while the nanochannels were employed to complete the reptation process.

5.4 Electrostatic Sieving

Electrostatic sieving is a separation mechanism that is dependent upon the electric double layer that forms at the boundary between an electrolyte and a channel wall. When a strong electrolyte is used, the electric double layer will be very thin, usually on the order of 5 Å. However, as the ion concentration is decreased, the electric double layer expands. In most microfluidic systems, a wide electric double layer affects the system little, since the extent of the layer will typically be on the order of tens or hundreds of nanometers. However, in nanochannels, it is easy to generate a Debye length on the order of the channel height.

Figure 5.1(d) illustrates the concept of electrostatic sieving. The top half of the figure illustrates negatively-charged nanopillars which have formed wide electric double layers on their surfaces, while the bottom half of the figure illustrates a variable-height nanochannel containing a relatively wide electric double layer. As charged analytes come into close proximity of the nanopillars, they will be either repulsed or attracted, depending on their charge. In the example shown in Figure 5.1(d), the greater the negative charge or the greater the electrophoretic mobility, the more quickly analytes will be pulled through the system of pillars. Positively charged analytes, on the other hand, will be repulsed by the electric double layer, increasing their travel time through the pillars. The concept is similar for the nanochannel case, where negatively-charged analytes will easily enter a nanochannel with nearly-overlapping electric double layers, while positively-charged analytes will be repulsed. Through electrostatic sieving, analytes can be separated according to their electrophoretic mobility without the necessity of an applied electric field, although electrokinetic pumping is commonly employed in nanofluidic channels.

Stein et al. [118] created channels with heights ranging from 1015 nm to 70 nm, and showed that the conductance of the channel approached a constant value for a KCl solution at a concentration of 1 mM or less. Schoch et al. [119] demonstrated nearly the same results with channels 50 nm in height. Electrostatic sieving was demonstrated by Fu et al. [112] using a periodic entropic trapping array. They utilized a Tris-borate buffer at
a concentration of 1.3 mM, which generated a Debye length of 8.4 nm. They successfully separated the proteins lectin and streptavidin using this method.

5.5 Immutable Trapping

Immutable trapping is an emerging separation method in the field of nanofluidics. Unlike most other nanofluidic separation methods, analytes are permanently retained, such that once an analyte is trapped, it will not traverse the remaining length of the channel. Because analytes are immutably trapped, detection must take place at the locations of entrapment across the system. Immutable trapping is very simplistic in its approach: if an analyte is too large to pass into a channel, it will become physically stuck, or trapped, at the entrance. Figure 5.1(e) illustrates the principles behind immutable trapping. The top half of the figure demonstrates trapping using a nanopillar array while the bottom half demonstrates trapping using an increasingly shrinking nanochannel. In the nanopillar case, the periodic spacing between pillars in each row will continue to decrease along the length of the channel in order to capture increasingly smaller analytes. The change in spacing period is shown to the right of the figure for each row, where \( \lambda \) represents the spacing period of the last row. The bottom of Figure 5.1(e) illustrates immutable trapping in a nanochannel that is stepwise decreasing in height along the length of the channel. As analytes enter a constricting region, they will continue along until they encounter a physical, impassable barrier. If the heights of each channel segment are well-characterized, it is possible to determine a size distribution of analytes at each trapping point. It should be noted that while, from Figure 5.1(e), it appears that a trapped analyte will hinder any further elution of particles behind it, the width of the channel is several orders of magnitude greater than its height, such that there are numerous voids through which following analytes may pass.

Perhaps the first experiments that exploited immutable trapping were carried out by Wang et al. [120, 12]. In their experiments, they fabricated a microchannel that shrunk to a nanochannel 40 nm in height toward the middle of the device. Analytes were adsorbed to gold particles with an average diameter of 60 nm, and the solution was then pumped through the channel. The gold particles became trapped at the interface between the microchannel and the nanochannel, and the analytes adsorbed to the surface of the gold particles were detected
using surface-enhanced Raman spectroscopy (SERS). Using this method, they were able to detect adenine, ethanol, BSA, and insulin. This approach utilized a single-step change, such that no separation occurred. However, detection was possible due to the unique Raman shifts exhibited by different adsorbed species. Some of the benefits of this method include using the change in channel dimensions to trap and concentrate analytes. Another advantage is that SERS is a universal detection method. However, the ability to utilize varying channel dimensions for separation is not addressed in this experiment.

The remainder of this section is devoted to the design, fabrication, and experimental testing of a dual-height immutable trap [121]. The purpose of this device is to prove the feasibility of capture based on analyte size and the restriction size of the channel. This system is shown to trap polymer beads 30-nm and 120-nm in diameter, as well as trap hepatitis B virus (HBV) capsids and herpes simplex type 1 virus (HSV-1) capsids. This section concludes with discussion about considerations for future immutable trapping approaches.

5.5.1 Dual-Segment Device Fabrication

Fabrication of dual-segment immutable traps is based on sacrificial etching techniques. One of the greatest advantages of sacrificially etching nanochannels is that the process can be carried out quickly. Nanochannels can be made from aluminum cores, and aluminum can be sacrificially etched much more quickly than most polymer cores used to fabricate microchannels. Another advantage offered by sacrificial etching is high device yield.

Figure 5.2 is a flowchart illustrating the steps involved in fabricating a dual-step trapping device. In (a), a thin layer of silicon dioxide, approximately 200 µm in thickness, is deposited over the top of a bare silicon wafer. This layer forms the bottom wall of the device and creates a wettable layer to aid in capillary flow. Next, a layer of aluminum is evaporated to a thickness that equals the height of the shorter of the two channel segments. In (b), AZ nLOF 2020 photoresist is spun-on the wafer, patterned, and developed. The unmasked areas on the wafer surface correspond to where the second, thicker layer of aluminum will be deposited to form the taller of the two channel segments. The second aluminum layer is deposited in (c), and the summation of the thicknesses of the first aluminum layer and the second layer determines the height of the tall channel segments. In (d), the masking
photoresist is lifted-off using N-methylpyrrolidone (NMP) heated to 50 °C. In (e), AZ3330 photoresist is spun-on the wafer and patterned, where the masked regions define the lateral dimensions of the channels. The exposed aluminum is etched away in (f) using aluminum etchant Type A (Transene) heated to 50 °C. The remaining photoresist is dissolved using treatments of acetone and isopropyl alcohol. In (g), a 5-µm thick layer of silicon dioxide is deposited over the entire face of the wafer. To access the aluminum cores for etching, AZ3330 photoresist is spun on the wafer and patterned to mask everything except rectangular regions immediately over the core ends in (h). The exposed oxide is etched with a buffered oxide etchant (BOE), which creates a fluid reservoir as well as exposes the tips of the aluminum cores. In (i), the aluminum cores are etched away with a mixture of aqua regia (hydrochloric acid and nitric acid) heated to 80 °C. Aqua regia is a very aggressive aluminum etchant. It is also non-viscous, which contributes to the diffusion path length of spent etchant, effectively increasing the overall etching speed compared to other more-viscous etchants. The result is that channel cores 1.2-mm in length can be completely etched-out in a matter of a few hours.

A completed dual-height immutable trapping device is shown in Figure 5.3. Figure 5.3(a) is a photograph of a single 1-cm² die containing dual-height traps set next to a U.S. quarter for size comparison. A photograph taken through a light microscope showing this same device is demonstrated in Figure 5.3(b). The entire trapping device spans a width of 8 mm, and each channel is 1.2-mm long. There are a total of 200 identical trapping channels set side-by-side in parallel on each die, and each channel is approximately 15 µm wide. Figure 5.3(c) provides a photograph of the same device taken at a higher magnification. The tall channels on this particular device are 200-nm in height, and can be clearly distinguished by their orange appearance. The short channels are 45-nm tall and are discernible by their teal color. Tall and short channel segments are highlighted in this image, as are channel openings to one of the two reservoirs. The most important component of this device is the channel interface between the tall and short segments, which is also highlighted in the figure. As analytes move from the tall channels toward the short channels, they will become trapped if their critical dimension exceeds that of the short channels.
Figure 5.2: Two-segment immutable trapping fabrication steps. (a) A thin layer of oxide and a sacrificial layer of aluminum are deposited. (b) Photoresist is spun on and patterned to define the area for the tall channels. (c) A second layer of sacrificial aluminum is deposited over the top of the masking photoresist and the previous aluminum layer. (d) The masking photoresist is lifted off. (e) Photoresist is spun-on and patterned to define the channel dimensions. (f) The exposed aluminum is etched away, and the photoresist is stripped off. (g) A thick layer of oxide is deposited over the sacrificial cores. (h) A reservoir is lithography masked and etched over the ends of the sacrificial cores. (i) The exposed aluminum cores are etched away, leaving hollow nanofluidic channels.

5.5.2 Dual-Height Trapping Operation

Figure 5.4 outlines the steps involved in trapping, detecting, recording, and analyzing analytes trapped at the dual-height interface. The device is set up on a microscope stage that interfaces with a highly-sensitive CCD camera (CoolSNAP HQ). As shown in (a), the sample is injected through the liquid reservoirs. As the solution fills the channels, analytes that are too large to enter the second channel segment become trapped at the interface. As analytes trap, they are monitored and detected through fluorescence detection, which is possible due to fluorophores attached to the surface of the trapped analytes. In (b), incoming light is
initially filtered to an appropriate band of wavelengths (blue light in this illustration). The filtered light is incident on the device on the microscope platform, and causes the trapped analytes to emit fluorescent signals at wavelengths different from the filtered wavelengths (green light in this illustration). The fluorescent signal passes through a dichroic mirror and is collected by the CCD camera. In (c), the signal captured by the camera is recorded by a computer. In this illustration, the bands of white correspond to regions where analytes are trapped. Once the data have been recorded, they can be interpreted and analyzed, as shown in (d). In this particular example, a curve demonstrating fluorescent signal intensity versus elapsed time is shown. From fluorescence intensity information, it is possible to characterize the rate of analyte trapping, the signal-to-noise ratio of the system, and the concentrations of trapped analytes.
Figure 5.4: Dual-segment immutable trapping system setup. (a) The sample is injected into the dual-height immutable trapping chip. As the analytes become trapped at the interface, (b) their fluorescent signal is isolated by a filter and detected by a cooled CCD camera. (c) The images of fluorescent intensity over time are recorded by a computer and (d) processed for analysis.

To illustrate the mechanism of dual-step immutable trapping, refer to Figure 5.5, where a top view is shown in (a) and a side view is shown in (b). The solution is injected in the reservoir on the left-hand side, and fills the channels moving from left to right. The height of the first channel segment is $h_1$, and the height of the second channel segment is $h_2$. Any analytes with average diameters that lie between $h_1$ and $h_2$ become trapped at the interface between the channel segments. The trapped analytes are then interrogated through their fluorescently-excited signal. One important point to notice in this setup is that the channels initially fill via capillary action. However, capillary action ceases once the channels have completely filled. Further flow occurs due to evaporation of the solution off of the waste reservoir end. Evaporative flow is slower than capillary flow, and should be accounted for in experiments with time constraints.

5.5.3 Molecular Trapping

In order to test the feasibility of an immutable trapping system, it is essential to identify analytes whose natural shapes and sizes would lend to such a separation mecha-
Figure 5.5: Dual-segment immutable trapping. In (a) and (b), the solution is injected from the left-hand side of the device and flows via capillary action to the right. Analytes with diameters larger than the second channel height, $h_2$, will become trapped at the interface between the two channel segments. Their fluorescently-excited signals confirm their presence and allow for quantification.

nism. One requirement is a fairly rigid, non-deformable structure. For example, DNA is very elastic, and as such is a poor candidate for immutable trapping. Another requirement is that the analyte will not significantly interact for prolonged periods of time with the channel walls. Such interactions could include Coulombic forces, significant steric interactions, adsorption, or some form of partitioning. If any of these effects are anticipated, additional accommodations must be made. For example, the addition of surfactants or varying the ionic strength or pH level of the buffer can aid in minimizing some of these effects. In summary, for immutable trapping to operate well, it is important to choose analytes with well-defined dimensions that will not be hindered in their journey to the channel interface.

Several analytes have been selected and trapped using the dual-step immutable trap. Initially, trapping tests were performed with commercially-available polymer beads. Polymer beads were an attractive choice due to their well-defined dimensions, spherical shape, fluorescent labeling, and lack of surface proteins or other biological components that could hinder their progress through the channels. As trapping and detection of polymer beads proved successful, virus capsids were employed as well. Virus capsids are an attractive choice be-
cause they represent a common biological sample requiring detection in clinical analyses. Capsids are also ideal analyze candidates due to their near-spherical shapes and fairly rigid structures. Figure 5.6 illustrates virion structures that employ viral envelopes and utilize DNA for their genome. Intact, one of these viruses posses a viral envelope that encompasses the entire capsid and contains surface protein binding sites. It will also contain strands of viral DNA within the capsid structure. Analyzing such a complete virion would be difficult, since the viral envelope can swell or shrink depending on the surrounding conditions, and binding proteins can easily adsorb to channel walls. Furthermore, there is an inherent risk in handling virions containing live viral DNA. To analyze these viruses, the envelope was first stripped and the DNA was neutralized, leaving behind hollow capsids for trapping test.

**Figure 5.6:** Enveloped virion. Herpes simplex virus type 1 (HSV-1) and hepatitis B virus (HBV) incorporate viral envelopes, as illustrated here. The viral genome (DNA in this case) is located within the protein capsid. The capsid is protected by a viral envelope which contains protein binding sites on its surface that allow it to enter a host cell. Once the envelope is removed and the DNA is neutralized, it is possible to handle and trap the rigid capsid. HSV-1 capsids are approximately 125 nm in diameter and HBV capsids are approximately 30 nm in diameter.

To characterize the dimensions of the analytes used in this work, scanning electron micrographs and transmission electron micrographs were captured of the polymer beads and virus capsids, as shown in Figure 5.7. A scanning electron micrograph of polymer beads with an average diameter of 120 nm is shown in (a), while polymer beads with an average diameter of 30 nm are shown in (b). A transmission electron micrograph of HSV-1 capsids is
shown in (c), and a transmission electron micrograph of HBV capsids is shown in (d). The average diameter of HSV-1 capsids is 125 nm, and the average diameter of HBV capsids is 30 nm.

Figure 5.7: Polymer beads and viral capsids used for testing two-segment immutable trapping. (a) Polymer beads with an average diameter of 120 nm, (b) polymer beads with an average diameter of 30 nm, (c) HSV-1 capsids with an average diameter of 125 nm, and (d) HBV capsids with an average diameter of 30 nm.

The initial trapping tests were performed using the 120-nm and 30-nm polymer beads. These were chosen first due to the simplicity of their shapes and structures. These beads also provided insight to how well the virus capsids would be trapped, because HSV-1 capsids possess nearly the same average diameters as those of the beads (a 125 nm diameter for HSV-1 capsids and a 35 nm diameter for HBV capsids) [122]. Once it was demonstrated that trapping and detecting polymer beads was feasible, fluorescently-labeled virus capsids were tested next. Figure 5.8 shows time-lapsed photographs of trapped fluorescently-detected HSV-1 capsids (a) and HBV capsids (b). In each case, it is apparent that the fluorescent signal intensity increases over time, suggesting that molecules are continually trapping at the channel interfaces. Furthermore, it is interesting to note the pattern of trapped capsids. HSV-1 capsids tend to form a uniform line along the interface, whereas HBV capsids appear to aggregate more strongly at the corners of the interface. This is most likely due to the difference in interaction between the virus capsids and the channel walls. The channels in
Figure 5.8(a) change in height from 190 nm to 65 nm, and the channels in Figure 5.8(b) change in height from 145 nm to 25 nm.

![Figure 5.8](image)

**Figure 5.8**: Time-lapse sequence of trapped (a) HSV-1 capsids and (b) HBV capsids. Intensity of the trapped capsids increases over time. In (a) the channels change height from 190 nm to 65 nm, and in (b) the channels change height from 145 nm to 25 nm.

An effective way to characterize immutable trapping behavior is to study the change in fluorescent signal intensity over time. Figure 5.9 shows the fluorescent signal of trapped analytes over time for 120-nm polymer beads at a concentration of $4.55 \times 10^{10}$ particles mL$^{-1}$, 30-nm polymer beads at a concentration of $3.54 \times 10^{12}$ particles mL$^{-1}$, HSV-1 capsids at a concentration of $5.4 \times 10^{12}$ particles mL$^{-1}$, and HBV capsids at a concentration of $9.44 \times 10^{14}$ particles mL$^{-1}$. From this figure, it is apparent that there are at least two different intensity slopes for the beads and the virus capsids. This change in slope appears to occur around 12 s for the polymer beads and around 6 s for the virus capsids. These variations can be interpreted in a couple of ways. One possibility is that the first, steeper slope for each set
represents the time when molecules are becoming trapped during the fast capillary filling stage. The second, shallower slope could then be attributed to the time when analytes are becoming trapped when driven by evaporative flow, which is a slower filling mechanism than capillary action. However, it would be expected that the filling times would reflect those predicted by the Washburn equation (Equation (4.36)) coupled with the dynamic contact angles reported in Figure 4.26. Calculations based on this information indicate capillary filling times of the second, short segment would be limited to only a couple of seconds, even for the shortest channels used in these experiments. As such, there is likely another mechanism at play. The initial slope of each curve may be due to the time delay between channel filling and the first detected signal, which may be attributed to titration of the solution by acidic silanol groups on the silicon dioxide walls [123]. As the titrated solution flowed through the channel due to evaporation at the channel outlet, the solution surrounding the trapped analytes became less acidic and the fluorescent signal intensified. This effect continued until the solution at the trapping interface reached a pH at which fluorescence was no longer suppressed (i.e. at $t \approx 12$ s for the polymer beads and at $t \approx 6$ for the virus capsids). The difference in times at which the change in slope occurred for the beads and the virus capsids can be attributed to the difference in fluorophores used to label the virus capsids and the polymer beads. The increase in signal intensity thereafter was caused by additional analytes being trapped at the interface due to evaporative flow. Furthermore, in Figure 5.9 the curves representing the virus capsids were more flattened than the curves representing the polymer beads. This result was likely due to some retention mechanism caused by interaction between the virus capsids and the channel walls.

Another important aspect to quantify in an immutable trapping system is the signal-to-noise ratio. This ratio provides the ability to determine the minimum detectable analyte concentration. Figure 5.10 shows the signal-to-noise ratio of 120-nm polymer beads (a) and 30-nm polymer beads (b) at various concentrations, as indicated by the bottom x-axis. The data for determining the signal-to-noise ratio were recorded at approximately 24 seconds following initial signal detection in each case. The vertical error bars correspond to the standard deviation of the average computed taken over several channels. The horizontal error bars correspond to the top x-axis, particles loaded per channel. This top x-axis provides
Figure 5.9: Normalized fluorescent signal intensity versus time for trapped analytes. The 120-nm beads and HSV-1 capsids were trapped in channels changing in height from 190 nm to 65 nm, and the 30-nm beads and the HBV capsids were trapped in channels changing in height from 145 nm to 25 nm. The plotted data are the normalized, mean intensity values taken for 17 to 32 data sets per point.

The estimated number of analytes trapped at the interface. This calculation assumes that the total number of particles loaded in one channel is subsequently trapped at the interface. The plots in (a) and (b) demonstrate linear behavior ($R^2 = 0.994$ for the 120-nm beads, and $R^2 = 0.992$ for the 30-nm beads), suggesting that signal strength increases linearly with analyte concentration. Figure 5.10(c) shows a 3D plot of the intensity of a fluorescent signal of trapped analytes and the immediate surrounding area. The particles in this instance were 120-nm beads at a concentration of $4.55 \times 10^{10}$ particles mL$^{-1}$. This 3D image illustrates the intensity of the detected signal compared to the background.

5.5.4 Trapping Considerations

There are many candidate analytes that could be separated and detected using immutable trapping. Lipoproteins are naturally occurring complexes found in blood composed of lipids and proteins. Lipoproteins can be broken down into several fractions, including very low density lipoproteins (VLDLs), intermediate density lipoproteins (IDLs), low density lipoproteins (LDLs), and high density lipoproteins (HDLs). These fractions are shown
Figure 5.10: SNR of (a) trapped 120-nm beads and (b) trapped 30-nm beads. The relationship between the SNR and the particle concentration is nearly linear in (a) at $R^2 = 0.994$ and (b) at $R^2 = 0.992$. The top x-axis in both plots represents the calculated number of particles initially loaded into each channel, providing a first-order approximation of the number of particles trapped at the interface. (c) Typical contour plot of the fluorescent signal of trapped 120-nm beads at a concentration of $4.55 \times 10^{10}$ particles mL$^{-1}$, illustrating the intensity of the signal against the low background noise.

in Figure 5.11, along with the range of diameters for each fraction. Lipoprotein fraction size ranges were taken from Myant [124]. Some lipoprotein fractions, such as LDLs and HDLs, have been extensively studied as indicators of cardiovascular disease. In most instances, LDLs are thought of as “bad” lipoproteins and HDLs are thought of as “good” lipoproteins [125]. However, such an assignment is incomplete and does not fully explain cardiovascular risk. For example, large subfractions of HDLs exhibit a positive correlation with cardiovascular disease, while small subfractions of LDLs exhibit a negative correlation with cardiovascular disease. An immutable trapping system could be designed to separate and identify concentrations of lipoprotein subfractions for cardiovascular disease threat analysis.

To demonstrate the ability of an immutable trap to separate lipoproteins, a sample of fluorescently-labeled lipoproteins was introduced into a dual-segment immutable trap with channel heights changing from 125 nm to 25 nm. The resulting trapped lipoprotein fluorescent signals are shown in Figure 5.12. The interfaces that exhibit trapped lipoproteins are highlighted in the figure, as well as those fractions small enough to completely elute
Figure 5.11: Lipoprotein classes. The pink regions represent the phospholipid layer, and the purple regions represent proteins. Cholesterol is located at the center of each lipoprotein.

the length of the channel. A dual-segment device will not provide much information on lipoprotein subclasses, such that a more expanded, sophisticated system would need to be applied.

Figure 5.12: Trapped lipoproteins in a dual-segment immutable trap. The channel decreases in height from 125 nm to 25 nm at the interface. This system would trap most VLDLs and IDLs, and some LDLs. All HDLs should elute to the end of the channel. The majority of lipoproteins in this sample are VLDLs, which can be as small as 25 nm in diameter.

The concept of the immutable trap relies on its ability to separate a variety of analytes suspended in a solution. Figure 5.13 illustrates three methods of separation through
immutable trapping. In Figure 5.13(a), separation occurs along the length of a single, long channel. In this case, the height of the channel decreases in discrete steps from $h_1$ down to $h_N$ (where $N = 5$ in the figure). As analytes flow through the channel from left to right, they will be immutably trapped as they encounter a segment shorter than the diameter of the analyte. The bottom portion of Figure 5.13(a) illustrates a top view of the same process to demonstrate that these devices are planar and will not become clogged at the interfaces unless the analyte concentration is extremely high. Once analytes are trapped at discrete locations, their resulting signals can be recorded and plotted versus position. Due to the discrete nature of the channel segments, each peak will exhibit high spatial selectivity. The area of the curve can then be calculated to determine quantities such as molecule concentration. The concentration at each point will then correspond to a unique size range. For example, the first point in the second plot of Figure 5.13(a) will correspond to a range of $d > h_2$, assuming $d$ is the diameter of the trapped analytes. For the next point, the analyte size range is $h_2 > d > h_3$, and so forth. Using this approach, it is possible to determine the concentration of discrete size-ranges.

In Figure 5.13(b), a single long channel is employed as well, but in this case there are no discrete height changes. Instead, the channel linearly decreases in height along its length. Currently, there are evaporation and exposure methods under investigation that allow the creation of a graded thin film, which are necessary to create such channels. As analytes flow from left to right, they become trapped as they reach the specific point where the height of the channel equals the diameter of the analyte. Again, a top view is shown to illustrate the concept that channel clogging is minimized when a planar channel is filled with a solution containing a sufficiently low analyte concentration. The trapped analyte signals are shown in the plot in Figure 5.13(b). It becomes apparent that one of the drawbacks of this method is the possibility of poor resolution between peaks of analytes with similar dimensions. The main advantage, however, is that if the analytes are resolved, the exact size of a specific analyte can be determined depending on the spatial location of the detected signal.

Figure 5.13(c) offers an approach that is more conducive to sacrificial etching techniques. Since sacrificial etching times are proportional to the square of the length of the channel, the methods shown in (a) and (b) may require prohibitively long etch times. The
Figure 5.13: Immutable trapping methods. (a) A channel decreases in height in a step-wise fashion along its length, as shown in the side view. Analytes become trapped as they encounter a segment shorter than their diameter. The signal intensity of the trapped analytes is recorded, and the area of the peak intensities can be used to determine the number or concentration of analytes within specific size ranges. (b) A channel decreases linearly in height along its length. Analytes become trapped as they reach a point of critical restriction, as shown in the side view. The signal intensity of the trapped analytes is recorded, and if the resolution between peaks is sufficient, analyte concentrations within a specific size range can be determined. (c) Parallel channels are utilized to reduce the total length required in a single-channel system. The channel heights, \( h_{2n} \), decrease from left to right. Most analytes will pass through the tallest segment, \( h_{21} \), while all analytes will be trapped at the interface of the segment with height \( h_{25} \). The difference in intensity peak area between parallel channels will provide information about the concentration and number of channels of a specific size range at each trapping interface.
scheme in Figure 5.13(c) employs several short parallel channels. For each channel in this illustration, \( h_1 > h_{2n} \). In this case, \( h_1 \) is the same height for each channel, but the last section will decrease in height from \( h_{21} \) down to \( h_{2N} \) (\( h_{25} \) in the figure). As equal concentrations of analytes enter each channel, the majority of the analytes will pass through the interface between segments \( h_1 \) and \( h_{21} \), while most of the analytes will be trapped at the interface between segments \( h_1 \) and \( h_{25} \). The detected signals at the interfaces between segments \( h_1 \) and \( h_{2n} \) are shown in first plot of Figure 5.13(c). Since increasing numbers of analytes are trapped with each subsequent channel, the fluorescent signal at the interface will increase with each channel. The area of the detected peaks can then be measured. By taking the difference between peak areas, a distribution is created that allows quantification of the particles in a specific size range. For example, by taking the difference between the first and second peak areas in Figure 5.13(c), the analyte size distribution is \( h_{21} > d > h_{22} \). By taking the difference between the second and third peak areas, the analyte size distribution is \( h_{22} > d > h_{23} \), and so forth. The greater the number of parallel channels, the greater the resolution of the size distributions.

Separation mechanisms that rely on nanochannels are becoming increasingly popular. It is important to understand the different separation mechanisms at work, and choose a system appropriate to separate the analytes of interest. Ogston sieving is a mechanism that relies on steric interactions between the analytes and the channel walls. Entropic trapping relies on the finite probability that an analyte will deform at some point in time to squeeze into a channel smaller than the radius of gyration of the analyte. Reptation is the extreme case of entropic trapping where the analyte completely unravels to pass through a very narrow channel. Electrostatic sieving relies on the Coulombic interaction between analytes and the charges along a channel wall to either pass or retain an analyte. Immutable trapping is a mechanism which permanently traps an analyte that is too large to pass into a small opening. Immutable trapping was studied in this work by fabricating a dual-segment trapping device. Nano-scale polymer beads and virus capsids were trapped and detected, and the resulting signals demonstrated the ability to quantify trapped samples as well as determine the SNR of the system.
Chapter 6

Conclusion

Analytical separations allow for clinical diagnostics, environmental monitoring, and chemical threat detection. When analytical separation systems are incorporated onto a single microchip, they gain several unique advantages over traditional methods. An understanding of analytical techniques will allow those in engineering to apply analytical capabilities to their discipline.

Microfluidic and nanofluidic channels have been fabricated using sacrificial etching. This fabrication method allows the usage of standard thin film processing techniques and equipment. This method also allows for straight-forward integration with other microelectronic features. It was shown that when presented with various possible sacrificial cores for fabricating nanochannels, aluminum proved to perform best in terms of rapid etching. Polymer cores were also shown to etch more quickly when coupled with aluminum cores.

A thorough understanding of pumping mechanisms in microchannels and nanochannels is essential when designing any thin film chemical analysis device. The most common pumping methods include pressurized flow, electroosmotic flow, and capillary action. Electroosmotic flow was characterized in round-core microfluidic channels, while capillary flow was characterized in planar nanofluidic channels. Both were shown to be effective pumping mechanisms.

There are several unique separation methods available within the realm nanochannels. Some of these methods include Ogston sieving, entropic trapping, reptation, and electrostatic sieving. An alternative method, immutable trapping, was developed and explored as a powerful tool for separating nano-particles based on analyte dimensions.
Bibliography


105


107


108


111


Appendix A

Appendix – Publications

The following comprises the list of publications of which I have been a part.

Archival Journal Publications


Conference Papers and Presentations


Appendix B

Appendix – Process Recipes

B.1 Introduction

This appendix lists several processing recipes used for the devices fabricated in this work. Due to the nature of the fabrication process, these recipes will be very specific to the BYU Integrated Microfabrication Laboratory (IML), but due considerations can be made to adapt these devices to other facilities. Specific processing information for particular machines can be found at http://cleanroom.byu.edu.

B.2 Electroosmotic Pump

1. Bottom oxide deposition
   (a) Deposit \( \sim 200 \) nm PECVD oxide

2. Aluminum core evaporation
   (a) Blow-off surface of wafer with nitrogen gun
   (b) Evaporate \( \sim 200 \) nm aluminum using the thermal or e-beam evaporator

3. Photoresist core masking and reflow
   (a) Evaporate-bake wafer at 150 °C for 30 min
   (b) Spin-on HMDS for 10 s at 2000 RPM
   (c) Spin-on AZ 3330 photoresist (positive) for 1 minute at 2000 RPM
   (d) Soft-bake for 60 s at 90 °C
   (e) Align and expose mask set uFLUID C-1 CORE (clear-field) for \( \sim 14 \) s
   (f) Develop in AZ300 MIF developer for \( \sim 30 \) s, or until fully developed
   (g) Rinse in DI water and dry with nitrogen gun
   (h) Hard-bake ramp from 90 °C to 250 °C

4. Core creation
   (a) Descum wafer in PEII for \( \sim 1 \) minute
   (b) Etch exposed aluminum in aluminum etchant heated to 50 °C until removed
   (c) Thoroughly rinse wafer in DI water
(d) Dry wafer with nitrogen gun

5. Top oxide deposition
   (a) Descum wafer in PEII for $\sim$1 minute
   (b) Deposit 5 $\mu$m PECVD oxide

6. Core expose
   (a) Evaporate-bake wafer at 150 °C for 30 min
   (b) Spin-on HMDS for 10 s at 2000 RPM
   (c) Spin-on AZ 3330 photoresist (positive) for 1 minute at 2000 RPM
   (d) Soft-bake for 60 s at 90 °C
   (e) Align and expose mask set \textbf{uFLUID C-1 CORE EXPOSE} (dark-field) for $\sim$14 s
   (f) Develop in AZ300 MIF developer for $\sim$30 s, or until fully developed
   (g) Rinse in DI water and dry with nitrogen gun
   (h) Hard-bake for 2 min at 110 °C
   (i) Descum wafer in PEII for $\sim$1 minute
   (j) Etch exposed oxide in BOE for $\sim$8 min or until reservoir is fully etched down to silicon
   (k) Rinse wafer in DI water
   (l) Dry with nitrogen gun
   (m) Strip masking photoresist using acetone and IPA on the Solitec spinner

7. Sacrificial core etch
   (a) Create mixture of aqua regia
      i. Add nitric acid (HNO$_3$) to hydrochloric acid (HCl) at a ratio of 1:2
   (b) Submerge wafer in dish filled with aqua regia
   (c) Heat aqua regia to 130 °C (plate temperature)
   (d) Allow cores to etch for several hours or overnight
   (e) When aluminum cores are completely removed, rinse wafer in DI water
   (f) Submerge wafer in dish filled with DI water
   (g) Allow to soak for several hours or overnight
   (h) Dry wafer with nitrogen gun
   (i) Submerge wafer in dish filled with Nanostrip
   (j) Heat Nanostrip to 90 °C (plate temperature)
   (k) Allow to soak for several hours or until photoresist core is completely removed
   (l) Rinse wafer in DI water
(m) Submerge wafer in dish filled with DI water
(n) Allow to soak for several hours or overnight
(o) Dry wafer with nitrogen gun
(p) Dry channels
   i. Place wafer on cooled hot plate
   ii. Set temperature to 50 °C
   iii. Increase temperature in 10 °C increments every ~5 min until temperature reaches 110 °C

B.3 Dual-Height Immutable Trap

1. Bottom oxide deposition
   (a) Deposit ~200 nm PECVD oxide

2. Short-segment aluminum evaporation
   (a) Blow-off surface of wafer with nitrogen gun
   (b) Evaporate aluminum to desired thickness of the short channel segment using the thermal or e-beam evaporator

3. Tall-segment area masking
   (a) Evaporate-bake wafer at 150 °C for 30 min
   (b) Spin-on HMDS for 10 s at 2750 RPM
   (c) Spin-on nLOF 2020 photoresist (negative) for 1 minute at 2750 RPM
   (d) Soft-bake for 60 s at 110 °C
   (e) Expose mask set NANOSIEVE A-2 METAL 2 (clear-field) for ~10 s
   (f) Hard-bake for 60 s at 110 °C
   (g) Develop in AZ300 MIF developer for ~1 minute, or until fully developed
   (h) Rinse in DI water and dry with nitrogen gun

4. Tall-segment aluminum evaporation
   (a) Descum wafer in PEII for ~1 minute
   (b) Blow-off surface of wafer with nitrogen gun
   (c) Evaporate aluminum to desired thickness of the short channel segment using the thermal or e-beam evaporator
      i. Total thickness of the tall segment is the summation of the short-segment evaporation and this evaporation
      ii. AZ300 MIF developer etches aluminum, so by this step the exposed short-segment aluminum may be ~10 nm shorter than it was originally ($d_T = d_1 + d_2 - 10 \text{ nm}$)
5. Aluminum lift-off
   (a) Submerge wafer in dish filled with NMP
   (b) Cover dish with watch glass
   (c) Heat to 90 °C (plate temperature)
   (d) Label sample
   (e) Allow to soak for ~1 hour or until all of the photoresist/aluminum combination lifts off
   (f) If lift-off is taking awhile, place entire dish with sample in sonicator for ~10 min
   (g) If residue still remains, gently scrub wafer surface with foam swab
   (h) While wet with NMP, rinse wafer with IPA or methanol in fume hood
   (i) Dry wafer with nitrogen gun

6. Core masking
   (a) Evaporate-bake wafer at 150 °C for 30 min
   (b) Spin-on HMDS for 10 s at 2000 RPM
   (c) Spin-on AZ 3330 photoresist (positive) for 1 minute at 2000 RPM
   (d) Soft-bake for 60 s at 90 °C
   (e) Align and expose mask set **NANOSIEVE A-3 CORE** (clear-field) for ~14 s
   (f) Develop in AZ300 MIF developer for ~1 minute, or until fully developed
   (g) Rinse in DI water and dry with nitrogen gun
   (h) Hard-bake for 2 min at 110 °C

7. Core creation
   (a) Descum wafer in PEII for ~1 minute
   (b) Etch exposed aluminum in aluminum etchant heated to 50 °C until removed
   (c) Thoroughly rinse wafer in DI water
   (d) Dry wafer with nitrogen gun
   (e) Strip masking photoresist using acetone and IPA on the Solitec spinner

8. Top oxide deposition
   (a) Descum wafer in PEII for ~1 minute
   (b) Deposit 5 µm PECVD oxide

9. Core expose
   (a) Evaporate-bake wafer at 150 °C for 30 min
   (b) Spin-on HMDS for 10 s at 2000 RPM
(c) Spin-on AZ 3330 photoresist (positive) for 1 minute at 2000 RPM
(d) Soft-bake for 60 s at 90 °C
(e) Align and expose mask set NANOSIEVE A-4 CORE EXPOSE (dark-field) for ~14 s
(f) Develop in AZ300 MIF developer for ~1 minute, or until fully developed
(g) Rinse in DI water and dry with nitrogen gun
(h) Hard-bake for 2 min at 110 °C
(i) Descum wafer in PEII for ~1 minute
(j) Etch exposed oxide in BOE for ~8 min or until reservoir is fully etched down to silicon
(k) Rinse wafer in DI water
(l) Dry with nitrogen gun
(m) Strip masking photoresist using acetone and IPA on the Solitec spinner

10. Sacrificial core etch

(a) Create mixture of aqua regia
   i. Add nitric acid (HNO₃) to hydrochloric acid (HCl) at a ratio of 1:2
(b) Submerge wafer in dish filled with aqua regia
(c) Heat aqua regia to 130 °C (plate temperature)
(d) Allow cores to etch for several hours or overnight
(e) When aluminum cores are completely removed, rinse wafer in DI water
(f) Submerge wafer in dish filled with DI water
(g) Allow to soak for several hours or overnight
(h) Dry wafer with nitrogen gun
(i) Submerge wafer in dish filled with Nanostrip
(j) Heat Nanostrip to 90 °C (plate temperature)
(k) Allow to soak for several hours or overnight
(l) Rinse wafer in DI water
(m) Submerge wafer in dish filled with DI water
(n) Allow to soak for several hours or overnight
(o) Dry wafer with nitrogen gun
(p) Dry channels
   i. Place wafer on cooled hot plate
   ii. Set temperature to 50 °C
   iii. Increase temperature in 10 °C increments every ~5 min until temperature reaches 110 °C
B.4 Dual-Height Immutable Trap With Sweeper Channel

1. Bottom oxide deposition
   (a) Deposit ∼200 nm PECVD oxide

2. Short-segment aluminum evaporation
   (a) Blow-off surface of wafer with nitrogen gun
   (b) Evaporate aluminum to desired thickness of the short channel segment using the thermal or e-beam evaporator

3. Tall-segment area masking
   (a) Evaporate-bake wafer at 150 °C for 30 min
   (b) Spin-on HMDS for 10 s at 2750 RPM
   (c) Spin-on nLOF 2020 photoresist (negative) for 1 minute at 2750 RPM
   (d) Soft-bake for 60 s at 110 °C
   (e) Expose mask set NANOSIEVE A-2 METAL 2 (clear-field) for ∼10 s
   (f) Hard-bake for 60 s at 110 °C
   (g) Develop in AZ300 MIF developer for ∼1 minute, or until fully developed
   (h) Rinse in DI water and dry with nitrogen gun

4. Tall-segment aluminum evaporation
   (a) Descum wafer in PEII for ∼1 minute
   (b) Blow-off surface of wafer with nitrogen gun
   (c) Evaporate aluminum to desired thickness of the short channel segment using the thermal or e-beam evaporator
      i. Total thickness of the tall segment is the summation of the short-segment evaporation and this evaporation
      ii. AZ300 MIF developer etches aluminum, so by this step the exposed short-segment aluminum may be ∼20 nm shorter than it was originally ($d_T = d_1 + d_2 - 20$ nm)

5. Aluminum lift-off
   (a) Submerge wafer in dish filled with NMP
   (b) Cover dish with watch glass
   (c) Heat to 90 °C (plate temperature)
   (d) Label sample
   (e) Allow to soak for ∼1 hour or until all of the photoresist/aluminum combination lifts off

122
(f) If lift-off is taking awhile, place entire dish with sample in sonicator for \( \sim 10 \) min
(g) If residue still remains, gently scrub wafer surface with foam swab
(h) While wet with NMP, rinse wafer with IPA or methanol in fume hood
(i) Dry wafer with nitrogen gun

6. Core masking
   (a) Evaporate-bake wafer at 150 °C for 30 min
   (b) Spin-on HMDS for 10 s at 2000 RPM
   (c) Spin-on AZ 3330 photoresist (positive) for 1 minute at 2000 RPM
   (d) Soft-bake for 60 s at 90 °C
   (e) Align and expose mask set \textbf{NANOSIEVE A-3 CORE} (clear-field) for \( \sim 14 \) s
   (f) Develop in AZ300 MIF developer for \( \sim 1 \) minute, or until fully developed
   (g) Rinse in DI water and dry with nitrogen gun
   (h) Hard-bake for 2 min at 110 °C

7. Core creation
   (a) Descum wafer in PEII for \( \sim 1 \) minute
   (b) Etch exposed aluminum in aluminum etchant heated to 50 °C until removed
   (c) Thoroughly rinse wafer in DI water
   (d) Dry wafer with nitrogen gun
   (e) Strip masking photoresist using acetone and IPA on the Solitec spinner

8. Sweeper channel creation
   (a) Evaporate-bake wafer at 150 °C for 30 min
   (b) Spin-on HMDS for 10 s at 2000 RPM
   (c) Spin-on AZ 3330 photoresist (positive) for 1 minute at 2000 RPM
   (d) Soft-bake for 60 s at 90 °C
   (e) Align and expose mask set \textbf{NANOSIEVE A-5 INLET} (clear-field) for \( \sim 14 \) s
      i. This alignment step is very precise, and you may need to shift the mask up a few microns to ensure that the sweeper channel touches the aluminum cores
   (f) Develop in AZ300 MIF developer for \( \sim 1 \) minute, or until fully developed
   (g) Rinse in DI water and dry with nitrogen gun
   (h) Hard-bake ramp from 90 °C to 250 °C

9. Top oxide deposition
   (a) Descum wafer in PEII for \( \sim 1 \) minute
(b) Deposit 5 μm PECVD oxide

10. Core expose

(a) Evaporate-bake wafer at 150 °C for 30 min
(b) Spin-on HMDS for 10 s at 2000 RPM
(c) Spin-on AZ 3330 photoresist (positive) for 1 minute at 2000 RPM
(d) Soft-bake for 60 s at 90 °C
(e) Align and expose mask set **NANOSIEVE A-6 CORE EXPOSE 2** (dark-field) for ~14 s
(f) Develop in AZ300 MIF developer for ~1 minute, or until fully developed
(g) Rinse in DI water and dry with nitrogen gun
(h) Hard-bake for 2 min at 110 °C
(i) Descum wafer in PEII for ~1 minute
(j) Etch exposed oxide in BOE for ~8 min or until reservoir is fully etched down to silicon
(k) Rinse wafer in DI water
(l) Dry with nitrogen gun
(m) Strip masking photoresist using acetone and IPA on the Solitec spinner

11. Photoresist sacrificial core etch

(a) Create mixture of piranha
   i. Add hydrogen peroxide (H₂O₂) to sulfuric acid (H₂SO₄) at a ratio of 1:1
(b) Submerge wafer in dish filled with piranha
(c) Heat piranha to 130 °C (plate temperature)
(d) Allow photoresist core to etch for a couple days until completely removed
(e) Replace the piranha solution every 24 hours
(f) Rinse wafer in DI water
(g) Submerge wafer in dish filled with DI water
(h) Allow to soak for several hours or overnight
(i) Dry wafer with nitrogen gun

12. Aluminum sacrificial core etch

(a) Create mixture of aqua regia
   i. Add nitric acid (HNO₃) to hydrochloric acid (HCl) at a ratio of 1:2
(b) Submerge wafer in dish filled with aqua regia
(c) Heat aqua regia to 130 °C (plate temperature)
(d) Allow aluminum cores to etch for several hours or overnight
(e) When aluminum cores are completely removed, rinse wafer in DI water
(f) Submerge wafer in dish filled with DI water
(g) Allow to soak for several hours or overnight
(h) Dry wafer with nitrogen gun
(i) Submerge wafer in dish filled with Nanostrip
(j) Heat Nanostrip to 90 °C (plate temperature)
(k) Allow to soak for several hours or overnight
(l) Rinse wafer in DI water
(m) Submerge wafer in dish filled with DI water
(n) Allow to soak for several hours or overnight
(o) Dry wafer with nitrogen gun
(p) Dry channels
   i. Place wafer on cooled hot plate
   ii. Set temperature to 50 °C
   iii. Increase temperature in 10 °C increments every ~5 min until temperature reaches 110 °C