A nanoelectrode lined nanochannel for single-molecule DNA sequencing

by

Peter Schmidt

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science
in
Electrical and Computer Engineering

August, 2004

Dept. of Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Advisor: Dr. Tamal Mukherjee

Abstract

This thesis presents a novel idea for a device for single-molecule DNA sequencing. The device consists of a nanometer scale water channel embedded in glass, with an array of evenly spaced nanoelectrodes running at a 90 degree angle with the channel and terminating on both sides of it. These nanoelectrodes permit an electric characterization of the DNA, differentiating between the bases on the basis of their permanent dipoles. The contribution of the DNA's charged backbone is easily eliminated by clever choice of applied signal and signal processing.

Whereas traditional DNA sequencing methods rely on costly chemical reactions and unscalable machinery, the device only uses DNA and water and allows system-on-a-chip identification. Our theoretical analysis, supported by finite element data, shows that it surpasses the capabilities of all other non-traditional DNA sequencing technologies: We show that a feasibly manufacturable device should be able to classify with approximately 2~19 base accuracy a DNA molecule of arbitrarily long length comprised of two concatenated strands of a single base type.

We also discuss the manufacturing process, the tolerance in the various physical design parameters and their effects on performance. As current molecular dynamics technology is insufficient to accurately capture all the intricacies of the system, we present sufficient information for a full molecular dynamics simulation to be implemented in future work.

Acknowledgement

I would like to thank Dr. Tamal Mukherjee for much assistance with the presentation of the paper and invaluable advice on the physical manufacturing of the device, as well as convincing me that dielectric spectroscopy was worth a look. I would also like to thank Professor James Hoburg for advice on electrostatics as well as for reading my thesis, and fellow graduate student Ryan Magargle for help with FEMLAB. Additionally, I would like to thank Mary Moore, Lynn Philbin and Elaine Lawrence for invaluable administrative assistance, as well as my family for their support. Finally, I would like to thank Kristen for her assistance and support during the writing of this thesis, and for yelling at me to get back to work.

Table of Contents

1 Introduction

1.1 DNA

Deoxyribonucleic acid, commonly called DNA, is a polymer consisting of nucleotides; nucleotide is an umbrella term for the three components of the monomers in DNA. An individual nucleotide consists of a phosphate group, a sugar (deoxyribose) and a nitrogen base. Single-stranded DNA, or ssDNA, is a poly-

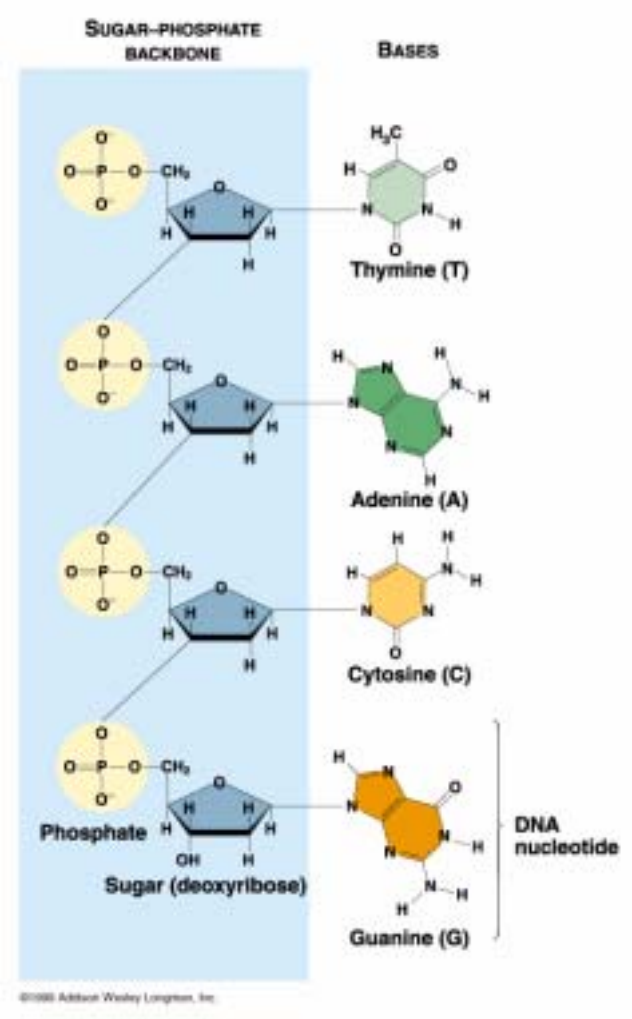

Figure 1-1. Single-stranded DNA (ssDNA) [1].

mer consisting of nucleotide monomers joined by bonds between adjacent sugar and phosphate groups as shown in Figure 1-1.

For a geneticist, the elements of interest in DNA are the nitrogen bases, of which there are four different types: thymine, adenine, cytosine, and guanine, typically labeled 'T,' 'A,' 'C,' and 'G' respectively. All of the genetic information in a strand of DNA is encoded in these bases; the nucleotide's remaining components are only relevant structurally.

It is possible for two complete single-stranded molecules to hydrogen bond at every base-pair to form a double-stranded molecule (called dsDNA); this process is termed hybridization. For this to occur two requirements need to be met:

1. The two ssDNA molecules need to be of opposite "orientation" (described below)
2. The two ssDNA molecules need to have the appropriate "base-pairs" at every position (described below)

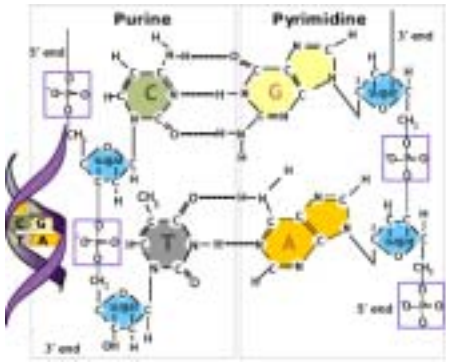

Figure 1-2. Double-stranded DNA (dsDNA); dsDNA consists of two hydrogen-bonded ssDNA molecules [2].

In ssDNA one end is necessarily terminated with a phosphate group (called the 5' end), and the other a sugar (called the 3' end). This determines a strand's orientation: either 5'-3' or 3'-5'.

There are two separate categories of bases (see Figure 1-2): purines, cytosine and thymine, and pyrimidines, guanine and adenine. As shown, cytosine can form three hydrogen bonds with guanine and thymine can form two hydrogen bonds with adenine; as a result, "cytosine and guanine" and "thymine and adenine" are called base-pairs (abbreviated "bp").

Complete hybridization requires both condition #1 and #2 described above to be met, but incomplete hybridization is still possible if there exists only a small percentage of inappropriate base pairs. If two

ssDNA molecules are capable of complete hybridization they are called “complimentary” (note that a given ssDNA molecule has exactly one “complimentary”).

1.2 History of DNA Sequencing

Frederick Sanger is accredited with the first successful technique for rapid DNA sequencing in 1978[1]. It is as follows (note DNA can trivially be copied nucleotide-by-nucleotide):

1. Make a multitude of copies of a strand of DNA (for reasons not discussed here, these are made of “dummy” nucleotides called dNTP (deoxyribonucleic triphosphate), i.e. dATP, dTTP, dGTP and dCTP).
2. Divide these copies into four pools of approximately equal size.
3. In addition to the dNTP, add a different ddNTP (dideoxyribonucleic triphosphate) to every pool (one gets ddATP, one gets ddTTP, one gets ddGTP and one gets ddCTP). These are similar to the dNTPs, except that
   a) Copying (done nucleotide by nucleotide) stops immediately upon adding a ddNTP
   b) ddNTPs are easily detectable in X-ray machines
4. Separate the copies in each pool by size (easily accomplished by gel electrophoresis)
5. Apply traditional X-ray technology to view the results. Assuming that sufficient copies were made, the sequence can be trivially read off. For 5’-GATTACA-3’
   a) The “ddATP” pool should have a 2-long, a 5-long and a 7-long strand
   b) The “ddTTP” pool should have a 3-long and 4-long strand
   c) The “ddGTP” pool should have a 1-long strand
   d) The “ddCTP” pool should have a 6-long strand

1. For more information on this and the other topics in this section, please refer to references [3] and [4].

Sanger's technique is extremely accurate but exceedingly slow: to do a million bases typically takes more than a year whereas DNA is usually many millions to billions of bases long.

In 1986 Leroy Hood responded to this by developing a technology that allows replacement of the ddNTPs by 4 different electronically detectable fluorescent dyes. This so called system of "color coding" eliminates the need for four separate pools as well as arduous hand analysis of X-ray data, making it over four times as fast as Sanger's technique. Hood's developments made electrophoresis the slowest step.

In 1991 Craig Venter developed "shotgunning," a process consisting of mechanically chopping up a strand of DNA, taking the pieces and running them in parallel, and then statistically re-assembling the original. Although initially dismissed as inaccurate, Venter's technique has proven as accurate as its predecessors, with a vast improvement in speed, as the speed of electrophoresis is limited by the longest piece of DNA present.

Finally in 1999 machines capable of electrophoresis became available, enabling fully automated DNA processing. The technology was called "capillary array electrophoresis," referring to the large arrays of tiny capillaries included to accommodate shotgunning. As of this writing this is the present state-of-the-art technology. Interestingly, even 25 years later this technology is rooted in Sanger's work. To date nobody has managed to sequence a strand of DNA without making copies and doing size-based separation.

1.3 Nanopores

Sequencing DNA is inherently difficult because of the extremely small size of DNA bases (12-15 atoms). The DNA bases and a scale are illustrated in Figure 1-3. To determine what a given base is, it is necessary to interact with it individually, requiring a detector of similar size. Traditional technology avoids this problem by repeated copying, chemically generating a multitude of labeled polymers for every base present; determining the presence or absence of a big group of polymers is trivial compared to determining what a single molecule is. Recent developments in nanotechnology with biological nanopores, silicon nan-

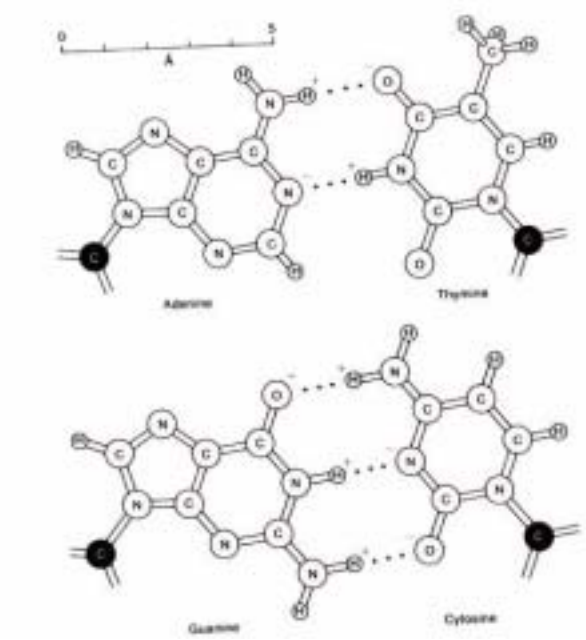

Figure 1-3. DNA bases to scale [5].

opores, and nanochannels have attempted to address the issue of single-molecule sequencing (i.e. without copies); cartoons describing these developments, as well as our concept of an nanoelectrode-lined nanochannel (NLN) are shown in Figure 1-4.

1.3.1 α-hemolysin nanopores

In 1996, John J. Kasianowicz (Harvard) [6] discovered that measurements of the general length of individual DNA strands were possible by passing DNA through α-hemolysin, a transmembrane protein that self-assembles in a lipid bilayer (see cross-section in Figure 1-5). α-hemolysin forms a channel with a limiting inner diameter of approximately 1.5nm [7], a value extremely close to the diameter of ssDNA or RNA (but too small for dsDNA) [8].

Kasianowicz did as follows:

1. Built a chamber and divided it in half with a lipid bilayer membrane

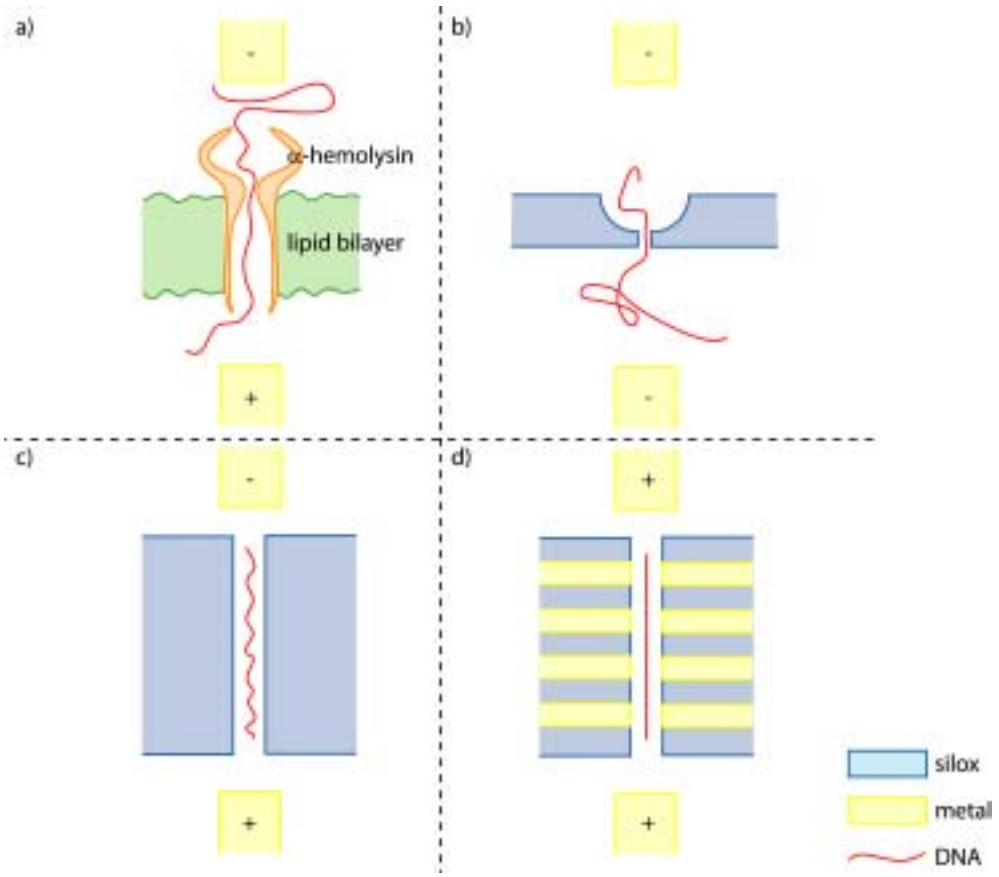

Figure 1-4. Cartoons of experimental setup for a) α-hemolysin nanopore, b) silicon nanopore, c) nanochannel, and d) our proposed design, the nanoelectrode-lined nanochannel. For a more complete illustration of how these are connected to an experimental apparatus, see [10]. Of interest is the entropy difference between parts of the DNA on different sides of the nanopores (a, b). The nanochannel (c) puts all parts of the DNA at approximately equal entropy. The nanoelectrode-lined nanochannel (d) is able to further reduce the entropy by stretching the DNA; the others are incapable of this since their sensing methods is intimately coupled to their control (positioning) method.

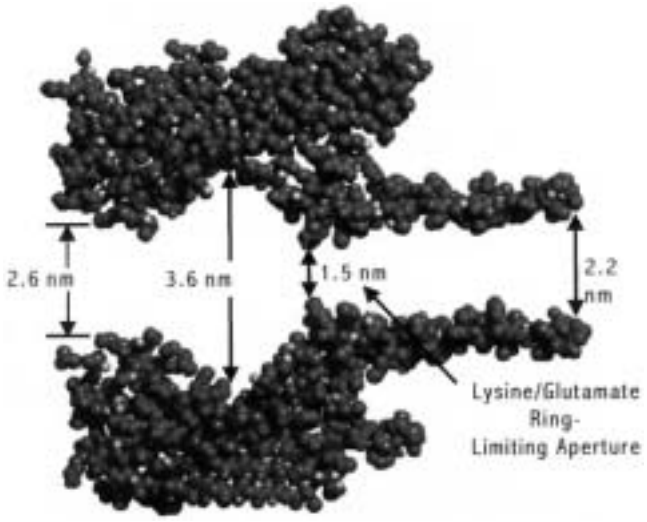

Figure 1-5. α-hemolysin cross-section [8]: the side on the left is *cis*-; the side on the right is *trans*-.

2. Allowed α-hemolysin to self-assemble in the membrane, but then stopped the process as soon as one finished, thus obtaining two reservoirs connected by a single nanopore (called cis- and trans-, a reference to the orientation of the nanopore; see Figure 1-5)
3. Added KCl to the cis- and trans- reservoirs
4. Added ssDNA to the cis- reservoir
5. Put a negative electrode on the cis- side and a positive electrode in the trans- side
6. Applied a voltage and measured the current by means of a patch clamp amplifier.

In the following discussion, please refer to Figure 1-6.

The electric field naturally causes the K^+Cl^- ions to pass from one side to the other (1-6A, top). The phosphate group in DNA has an excess electron, and as a result it as well moves in an applied electric field (this is the basis of electrophoresis). Thus in this setup it is eventually forced to pass through the nanopore

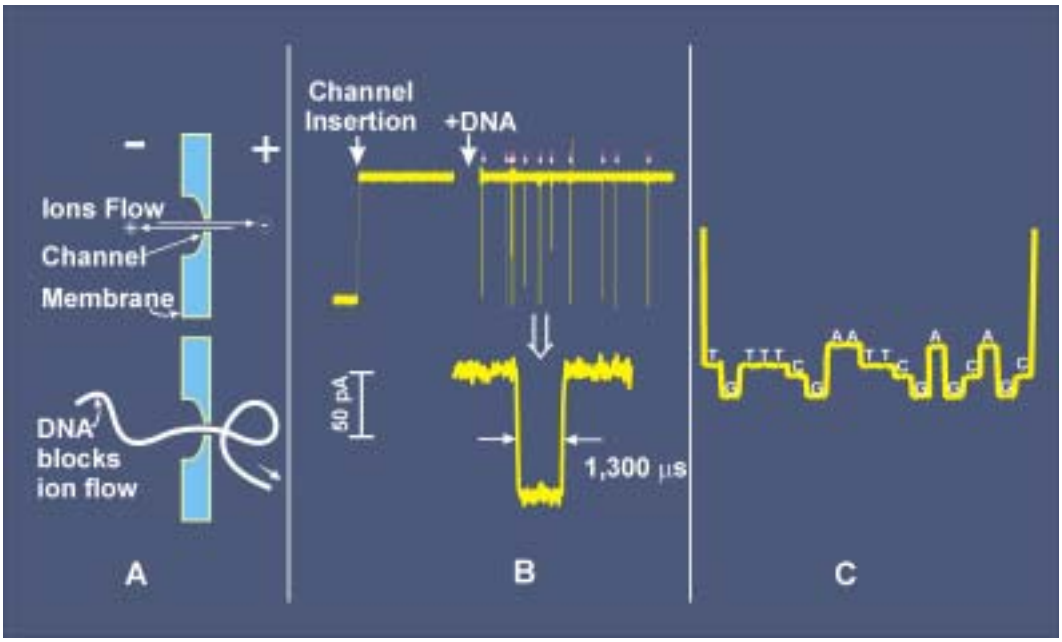

Figure 1-6. The concept behind the nanopore sequencing system (not actual data) [9].
a) (top) A differential voltage applied between electrodes on either side of the nanopore (not pictured) force K^+Cl^- ions to pass through the nanopore, resulting in a measurable current.
a) (bottom) The electric field between the electrodes also forces DNA molecules to pass through (translocate).
b) When an ssDNA molecule translocates, the K^+Cl^- current is impeded because it occupies some of the cross-sectional area the ions were passing through; this is shown in this current versus time graph. The bottom graph is a zoomed-in view of the top one.
c) The amount of current falloff depends on the blockage present in the nanopore; if single base precision could be achieved, the DNA could be sequenced by reading off the different of current blockage levels.

(commonly called translocation; see 1-6a, bottom). While present in the nanopore the DNA occupies a part of the cross-sectional area available for the K^+Cl^- to pass through, and thus current between the electrodes on the cis- and trans- sides of the nanopore is temporarily impeded (1-6B).

Kasianowicz discovered that the length of one of these "events" depended on the length of the ssDNA in the nanopore; however, there was a wide distribution of transit times, meaning the velocity of the ssDNA was inconsistent. [8]).

In 1999, Mark Akeson (Harvard), using a setup similar to Kasianowicz's, managed to achieve a time varying current loss as a strand of $A_{30}C_{70}Gp$ RNA travelled through an α-hemolysin nanopore, with one level corresponding to the A_{30} section and one level corresponding to the C_{70} section - if this could be accomplished with single base precision and an arbitrary strand of DNA, it would allow sequencing (1-6C) [10].

Much to his dismay, however, this is witnessed in neither ssDNA nor different RNA. The amazing ability of $A_{30}C_{70}Gp$ was attributed to the fact that poly C forms an unusually small diameter helix, beneath the limiting diameter of the nanopore. It thus traverses the pore as a helix, as opposed to the extended form poly A traverses the pore in. This is shown in Figure 1-7. Essentially, the issue is the size difference between poly A and poly C is bigger than the size difference between A and C (in RNA as well as DNA), meaning the nanopore would have to be smaller; however, this is not the only problem. The event durations were

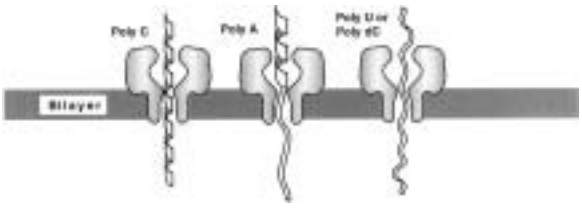

Figure 1-7. Poly C (RNA) versus Poly A (RNA), Poly U (RNA) and Poly dC (ssDNA) [10].

inconsistent too: the ratio of the blockage time for the A section to the blockage time for the C section was inconsistent from event to event, meaning it was not apparent if it was $A_{30}C_{70}$ or $A_{70}C_{30}$. This emphasized the relevance of Kasianowicz's result that the velocity of the DNA in the nanopore had a high variability.

In 2000, Amit Meller (Harvard), while confirming this [11], determined that composition dependent characteristics of individual DNA molecules could still be seen by examining the level and duration of the whole event. The disappointing implication of this is that a detector resolution of 4^n would be needed for a strand of n nucleotides, far beneath the noise floor of the system. Meller also had a large distribution of transportation times, consistent with Kasianowicz and Akeson.

One additional problem inherent to the nanopore system is that same electric field that moves the DNA moves the ions; as a result, if the field is lowered in an attempt to slow or stop it (even assuming there were no issues with entropy - see Section 1.4), the ions are similarly slowed or stopped, eliminating the current needed to do measurements.

In 2001 Wenonah Vercoutere (UCSC) finally managed to avoid this problem (to some degree) [12] with the clever use of DNA hairpins. A DNA hairpin consists of folded ssDNA: the ssDNA is self-complimentary, permitting its two halves to be hydrogen-bonded to each other. For stability the middle 4 bases are usually non-self-complimentary and as a result it has a "ball" at one end. This "ball" has sufficient diameter that it cannot fit into the α-hemolysin nanopore (see Figure 1-8).

This allowed researchers to "dangle" the terminal two hydrogen-bonded bases in the nanopore, thereby permitting them to do measurements for an arbitrary length of time. By using machine learning's Hidden Markov Models and Support Vector Machines [13], they were able to identify the termini with better than 99% accuracy. This, however, was more of a verification of the problem than a possible solution as just as many hairpins would be needed with this technology as copies are needed with traditional methodologies.

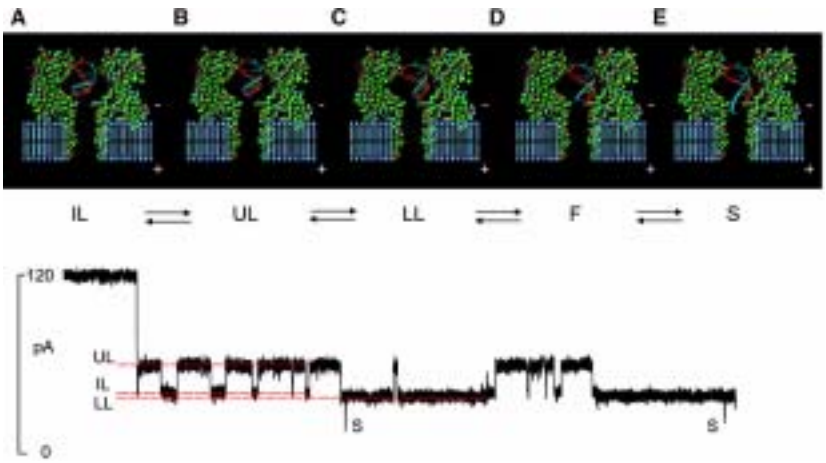

Figure 1-8. Time-dependent currents for different hairpin conformations [14]. The top shows the different conformations of the DNA in the nanopore and the bottom has a current versus-time plot with the corresponding conformation labeled.

The amazingly good resolution of this system has been attributed to its ability to detect the dynamic behavior of a base-pair in the nanopore (Figure 1-8). Theoretical analyses appear to be in good agreement [13] [14], giving additional credibility to the claim that the primary barrier to nanopore-based technologies is the uncontrollable velocity of the DNA. This is also supported by simulation: G. Timp (UIUC) concluded "control of the molecular transit time through the pore is a key aspect of the [nanopore] sensor" [15].

In 2003, Mark Bates (Harvard) applied active control technologies to further characterize DNA behavior in an α-hemolysin nanopore [16]. They had two key discoveries for the future of nanopore technology: first, the feasibility of feedback control (as a result of the comparably slow movement of the DNA versus the KCl ions), and second, that there was a possibility of the DNA getting temporarily "stuck" in the nanopore (attributed to interactions between the ssDNA bases and the walls of the nanopore).

1.3.2 Solid-state nanopores

This problem, in combination with the fixed size of the nanopore and the general instability of α-hemolysin, provided enough motivation for solid-state silicon nanopores. Three different groups did so independently in 2003: Jaili Li (Harvard) managed to make pores as small as 3nm [17], whereas A. J. Storm (Brown) achieved pores of 2nm [18]. J. B. Heng (UIUC) surpassed both with pores of 1.4±0.3nm [19] (as well as was first).

There are two main criteria in judging the quality of a nanopore (as applied to DNA sequencing): consistency in event duration (translocation times) and consistency in event magnitude (current falloff). Consistency in event duration is necessary to tell *how many* bases correspond to a certain event, whereas consistency in event magnitude, is needed to tell *what* their type is. Both Li and Heng did Meller-style [11] measurements on their nanopores [17] [19]. Both recorded improvements in consistency of event duration of silicon nanopores over their biological counterparts discussed in the proceeding section. Li's measurements showed that this consistency rapidly degraded with decreases in translocation speed (as a result of decreases in voltage or increases in number of bases in a strand). As for consistency in event magnitude, bigger nanopores have larger cross-sectional area and therefore naturally, a larger error in event magnitude, as the ions can more easily go around the DNA in the nanopore. Li verified this by recording a narrower distribution of translocation times for identical experiments on 10 nanometer pores, thus showing a tradeoff between consistency in event magnitude and consistency in event duration.

3 nanometer pores have insufficient consistency in both event duration and event magnitude to allow sequencing, and thus it appears that even if nanopores were ever made small enough to allow differentiable current levels, they would have far too inconsistent transportation times to be of any use, meaning Kasianowicz's system is fundamentally incapable of sequencing. Nonetheless, this is still an area of active research, with some groups modifying this system in an attempt to overcome these limits on the pareto-optimal nanopore technology.

1.4 Nanochannels

Entropy differences play a large role in the dynamics of DNA passage through a nanopore. Due to DNA's affinity for coiling, the side of the nanopore with more nucleotides on it is entropically preferable. This is readily apparent in recent work by Bates (Harvard) [16] wherein DNA was shown to leave a nanopore independently when an applied electric field was shut off, with a speed dependent on the number of

nucleotides present on one side or the other. Entropy is a statistically variable factor in translocation times, and thus it follows that it plays a large part in their inconsistencies.

In an attempt to reduce the entropic contribution to inconsistency, some scientists have proposed nano*channels*. A nanochannel is just an axially extruded nanopore, such that its depth is much greater than its cross-sectional area.

Once a strand is completely inside the nanochannel, all of its nucleotides are at approximately the same free energy and thus its movement is no longer entropically motivated. This effect is such that the difference between a strand of DNA partially versus completely inside a nanochannel (or nanopore) is so great that nanochannels have been successfully used for size-based separation of DNA [22] as an alternative to electrophoresis. The nanochannel's ability to "linearize" DNA inside the nanochannel has also been documented [23] [24].

Unfortunately, the nanopore sensing technique (an electrode at either end of the nanopore/channel) is rather useless in nanochannels. It relies on detecting different current blockages; these are primarily determined by the biggest nucleotide present. As a nanochannel has a large number of a strand's nucleotides in it all at once, a more localized sensing method is needed, such as additional electrodes *in* the nanochannel.

1.5 Thesis Outline

The rest of this thesis is dedicated to the analysis and design of a solid-state *nanoelectrode-lined* nanochannel (hereafter referred to as an NLN) with the advantages of both nanopores *and* nanochannels. Chapter 2 presents an overview of the device and a summary of its capabilities. Chapter 3 discusses a complete electrostatic characterization of the NLN, which is sufficient to determine its ability to detect DNA molecules (given their conformation), as well as the electric field it experiences. Chapter 4 concludes with several areas for future work, and applies the results of Chapter 3 to obtain some order-of-magnitude estimates of the NLN's performance.

2 The nanoelectrode-lined nanochannel

As aforementioned, we propose the development of a solid-state *nanoelectrode-lined* nanochannel. The nanochannel fixes the consistency problems of the nanopore but has fatally delocalized measurements. The nanoelectrode-lined nanochannel (NLN) has the nanochannel's consistency, localized measurements, and additional advantages of repeatability and control. The progression from nanopore to nanochannel to NLN was shown in Figure 1-4 on page 6. The NLN is detailed in Figure 2-1. It is targeted to be generated in 6 steps):

1. Deposit a layer of SiO_2 (silox1)
2. Deposit a layer of metal

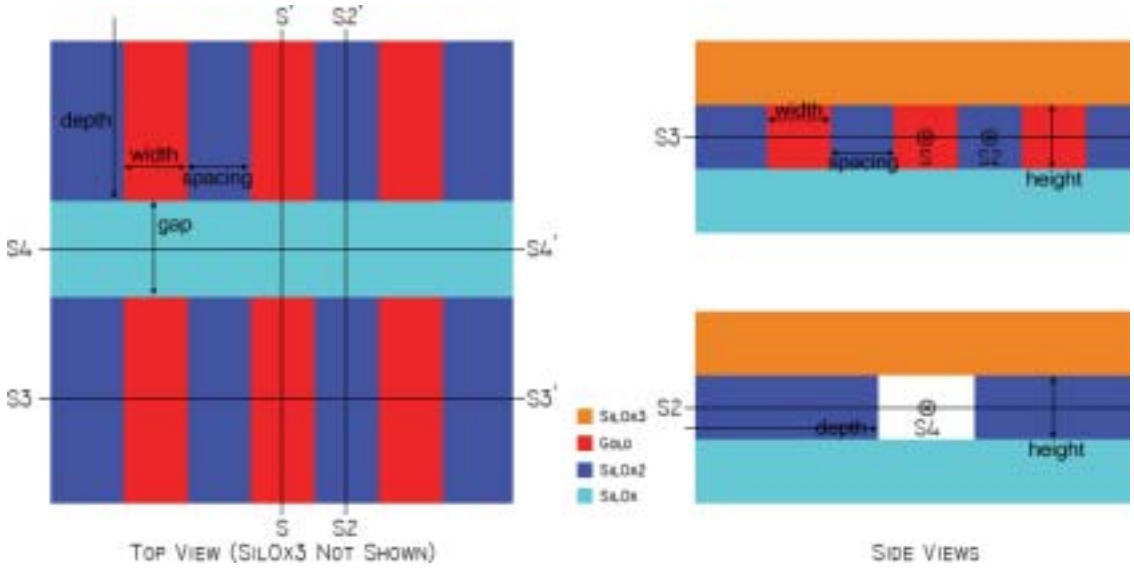

Figure 2-1. General layout of the NLN. Several dimensions of interest are labeled: depth, width, gap, spacing, and height; these are discussed more in Chapter 3.

3. Pattern metal to create an array of evenly spaced striplines; these should be as small as possible with good repeatability (illustrated top-to-bottom in Figure 2-1)
4. Conformally deposit SiO_2 (silox2) between striplines
5. Cut orthogonally to the striplines and as small as possible (illustrated left-to-right in Figure 2-1)
6. Apply glass cover slip (silox3) and anodically bond

Additional details of the fabrication of the NLN are beyond the scope of this thesis, which focuses on its simulation.

2.1 Advantages

The aforementioned advantages of the NLN are discussed in detail below.

2.1.1 Repeatability

The small separation (*spacing* in Fig. 2-1) of the electrodes allows nanopore-style current blockage characterization of the intervening bases. As well as overcoming the problem inherent in nanochannels, this allows for repeatability of measurements, which is critically important to the success of any current blockage characterization of DNA given the width of the distributions seen thus far [6] [8] [10] [11] [15] [17] as well as past successes with characterizations involving good repeatability [12] [13] [14]. In addition, the fact that the measurements should represent time-shifted data allows for cross-correlation to further validate them and help with control mechanisms (see 2.1.2 below).

2.1.2 Control

The biggest advantage of the nanoelectrode-lined nanochannel (NLN) is control. As aforementioned, all nanopore (as well as nanochannel) designs suffer from the fundamental limitation that the same electric field that moves the DNA moves the ions (see page 9). The electrode array in the NLN decouples

the control-measurement relationship; some electrodes can be programmed to position the DNA while others do only measurements. The following methods of control are all possible:

1. DC "stretching" via polarity differences
2. DC electrophoresis via feedback control
3. AC electrophoresis (see below)
4. AC dielectrophoresis

AC electrophoresis is accomplished by accounting for the different mobilities of K^+/Cl^- ions and DNA in electrophoresis. Any molecule's mobility in electrophoresis is entirely dependent on its charge to mass ratio. K^+ and Cl^-'s charge to mass ratios are bigger than DNA's (and therefore their mobilities are as well) and thus there exists a limiting electrode sizing and separation beneath which it will be possible to apply an electric field parallel to the channel that will only move DNA negligibly but move the K^+ and Cl^- ions two electrodes per half-cycle (the minimum needed for a complete circuit and thus the minimum needed for current measurements), thereby allowing measurements without DNA movement. This also keeps a large axial force on the DNA at all times, eliminating drift.

R. Austin's group (Princeton) [20] has successfully been able to manufacture nanochannels as small as 10x50nm; however, the lack of control technologies included resulted in prohibitive complications due to polymeric folding of the DNA (dsDNA has a persistence length of 50nm whereas ssDNA's is only 3 [21]). Despite contriving a novel injection mechanism, the molecules still continuously contorted in the channel; this was even an issue with channels later developed as small as 10nm in diameter [21].

2.1.3 Entropy

The NLN also has all the translocation-time consistency advantages inherent to the nanochannel, meaning that if its cross-sectional area can be made small enough to achieve sufficiently consistent event magnitudes, consistent translocation times will not be sacrificed as in a nanopore (see Section 1.4 on page 11).

2.1.4 Manufacturability

Unfortunately, current technologies for making short solid-state nanopores do not scale to long nanochannels (or NLNs), and so we are limited to using electron beam lithography. Assuming reproducibility for a nanochannel with cross-sectional area of 10nmx10nm, the ratio of the cross-sectional area of ssDNA to this is less than 0.04, meaning it would still have (extremely) inconsistent event magnitudes; therefore, current-blockage measurements in this system are relegated to the future of lithography technology. Nonetheless, dielectric spectroscopy style measurements of the DNA still allow all of the repeatability, control and entropy advantages of the NLN. The electrode-less nature of nanopores and nanochannels prohibits these type of measurements and thus there is no prior nanoscale work to use as a basis. There have, however, been many macroscopic dielectric characterizations done on bulk DNA.

2.2 Prior dielectric analysis of DNA

Apart from anomalous data [25] [26] later dismissed as a result of a faulty apparatus [27] [28], it is generally accepted that the 1-20GHz dielectric spectra of DNA is featureless [27] [28] [29] [30] [31] [32]; this is well supported by theory [33] [34]. Resonances were, however, discovered in the high gigahertz and terahertz ranges [36]; these were also determined to be sequence dependent [37] [38] [39]. As opposed to those of [25] and [26], the presence of resonant absorption at these frequencies is well supported by theory [37] [40] [41] [42] [43] [44] [45]. In addition, these resonances depend on the salt type as well as its concentration in a salt-DNA mixture [36] [37] [46] [47]. Impressively, THz spectroscopy has been able to successfully detect single base-pair mismatches in dsDNA [48].

Since our analysis method uses electrodes, we are incapable of optical analyses and are limited to those frequencies we are able to realistically apply to them (anything less than approximately 50GHz). Whereas all past research in the GHz region has been done on *coiled* dsDNA, the NLN analyzes *linear* ssDNA. The NLN also has the key ability that it can assure (as a result of its control capabilities) that the ssDNA's axis is always at a right angle to the applied field, leaving the bases as free as possible to rotate to

align with the applied field. We expect this will allow detectability of differences between different bases' *permanent dipoles* (described below).

A molecule's dielectric characteristics are a result of its dipole moment. A molecule's dipole moment consists of two components, both dependent on its atomic constituents: a permanent part determined by their relative electronegativities as well as an induced part determined by their polarizabilities (both scalars) [49]. Elementary physics tells us that a dipole in an oscillating electric field will attempt to align with it; however, at a sufficiently high frequency (dependent on several factors) it will be incapable of keeping up. This is known as dipole relaxation; relaxation of permanent dipoles typically occurs in the gigahertz range whereas relaxation of induced dipoles typically occurs in the terahertz range. This is a result of a movement of the electrons in induced dipoles, as opposed to permanent dipoles in which this is a result of a movement of the molecule itself (and therefore is dependent on a variety of factors: anything that promotes or impedes movement).

Due to their differing compositions, the bases in DNA all have different permanent dipole moments. Values were initially determined in 1962 by Howard Devoe and Ignacio Tinoco [50]. Since then numerous researchers have verified and modified the exact values, but it is still accepted that they are different (for a good comparison, see M. Preuss' recent paper [51]). The bases have different polarizabilities as well.

Terahertz spectroscopy relies on detecting different induced dipoles as a result of different polarizabilities; this is substantially easier than detecting different permanent dipoles for two reasons: first, the induced dipoles are induced; thus, even in a long, coiled piece of DNA in which the permanent dipoles would probabalistically cancel, the induced dipoles will still all be aligned. Second, the permanent dipoles are subject to anything that physically impedes movement; for example, in a piece of dsDNA the rigidity of the double-helix makes it nearly impossible for two bases to rotate independently to align with the field.

However, as much as differentiating the bases by their induced dipoles is preferable, as aforementioned we are limited to applied signals of frequencies less than 50GHz, meaning we must differentiate the

bases by their permanent dipoles. We assert that this is nonetheless sufficient for a device capable of single-molecule sequencing with better resolution than any existing technologies.

3 Electrostatic Characterization

To fully characterize the performance of the NLN we need to know both the effect of the DNA on the electrodes and the effect of the electrodes on the DNA: the DNA induces charge on the electrodes as a function of its position and orientation. However, the DNA's position and orientation changes as a result of this charge on the electrodes (in addition to charge appearing as a result of an applied voltage). By knowing these two things (DNA's effect on the electrodes and the electrodes' effect on DNA) as a function of the physical parameters of the NLN (electrode width, height, etc.), we will be able to evaluate the NLN's ability to analyze DNA.

3.1 Model

We use a continuum model for the DNA's environment: the electrodes, the silicon dioxide and the water. We assume no salt in the system, as salt is detrimental to the NLN's performance (this is discussed in detail in Section 4.4.2 on page 71). We model the DNA (electrically) as a collection of infinitely small partial charges. We analyze the system with FEMLAB, a finite element analysis program, operating in electrostatics (es) mode.

It has become common practice to model a large biomolecule as a collection of partial charges and its environment with continuum electrostatics ([53], [54], [55], [56], [57], many others). It should be noted, however, that the electrostatic approximation becomes less and less valid as frequency of operation increases; for it to be valid, the frequency of all applied signals should be sufficiently small that the corresponding wavelength is much bigger than any of the physical dimensions of the device. The use of finite elements to solve these problems is also well established [58]. We selected FEMLAB for its versatility and its interoperability with MATLAB. A boundary element solver would be equally well suited to this problem,

and could be used to further verify these results, although their agreement with analytic approximations and the “ideal” case (see Section 3.3) should be sufficient to show their accuracy.

3.1.1 Ionic interference

This model assumes *completely* deionized water; *incompletely* deionized water presents a variety of issues:

1. Ions in a fluid dielectric create a double-layer at the electrode-water interface and additionally isolate and neutralize any charges in the channel
2. Ions in a fluid dielectric cause a real conductivity.
3. Ions will induce charge on the electrodes.
4. Salt affects the polarizability of DNA [71].
5. Salt affects the persistence length of DNA; high concentrations are likely to promote hairpin formation and entanglement [20]. At extremely low (25mM) levels, the Kuhn length is as small as 3nm, meaning that it acts as a freely joined chain with segment length 3nm. At levels as high as 2M it is as low as 1.5nm [72]. At extremely high levels it reaches 0.796nm [70].

DNA can be easily purified by attaching it to the interior of a (commercially obtainable) device called an elution column and washing it over and over again until all impurities are removed. Something called an elution buffer is then applied, freeing the DNA. The elution buffer needed depends on a variety of factors beyond the scope of this thesis, but there exist commercial systems that can be eluted with either 10mM Tris-Cl (pH 8.5), or normal water with a similar pH. A pH>8 is needed for our system to work at all as a pH<8 causes excess H^+ to react with the PO_4^- on the DNA’s backbone, netting HPO_4, which is electrically neutral and thus unaffected by electrophoresis (severely limiting the NLN’s control capabilities).

It would seem that water is preferable as it contains less salt than 10mM Tris-Cl, which has

$$10\times10^{-3}\frac{\text{mol}}{\text{L}}\times 6.022\times10^{23}\frac{\text{molecules}}{\text{mol}}\times\frac{1}{1000}\frac{\text{L}}{\text{mL}}\times\frac{1}{1}\frac{\text{mL}}{\text{cm}^3}\times\frac{(1\times10^2)^3\,\text{cm}^3}{(1\times10^9)^3\,\text{nm}^3}\approx 0.006\frac{\text{molecules}}{\text{nm}^3} \quad (3.1)$$

meaning approximately .38 +/- ion pairs per base in ssDNA, assuming an ssDNA contour length per base of .63nm (see Section 4.5.2.1 on page 72) and a cross sectional nanochannel of 100nm^2.

Exceedingly pure water is commercially obtainable in which a majority of ions are a result of dissociation of the water molecules [59]: if KOH or NaOH is added to pH 7 water until its pH is 8.5 then $[H_3O^+]\sim10^{-8.5}$M, $[OH^-]\sim10^{-5.5}$M, and $[K^+]$ or $[Na^+]\sim10^{-5.5}$-10^{-7}M. The total ion concentration is then

$$(10^{-8.5}+2\times10^{-5.5}-10^{-7})\frac{\text{mol}}{\text{L}}\times 6.022\times10^{23}\frac{\text{molecules}}{\text{mol}}\times\frac{1}{1000}\frac{\text{L}}{\text{mL}}\times\frac{1}{1}\frac{\text{mL}}{\text{cm}^3}\times\frac{(1\times10^2)^3\,\text{cm}^3}{(1\times10^9)^3\,\text{nm}^3}\approx 3.75\times10^{-6}\frac{\text{molecules}}{\text{nm}^3} \quad (3.2)$$

meaning approximately .00024 +/- ion pairs per base in ssDNA, assuming an ssDNA contour length per base of .63nm and a cross sectional nanochannel of 100nm^2.

This is clearly preferable, but this pH is difficult to maintain: without a buffer water is highly susceptible to pH fluctuations. Extremely pure water will even react with atmospheric CO_2 to form carbonic acid, lowering the water's pH as low as 4.5 within a few hours [59]; thus, if NaOH/KOH in water is to be used as an elution buffer, extreme care must be taken to prevent contamination.

Note that normally, in a continuum treatment, neutralization of charges in a conductive fluid will occur with a characteristic time constant τ given by

$$\tau = \sigma/\varepsilon \quad (3.3)$$

However, needless to say continuum measurements of conductivity do not apply at the ion concentrations discussed above: there are approximately 30 ions with Tris-Cl in a 5000nm^3 nanogap (and even less with NaOH/KOH in water) as opposed to ~170,000 water molecules in the same volume. Nonetheless, if the Tris-Cl is utilized there still may be a deleterious charge neutralization effect, as well as a parasitic ion-induced charge on the electrodes. The charge neutralization is preventable by applying a sufficiently high frequency sinusoid to keep the DNA in motion, while the parasitic ion-induced charge is preventable by a clever signal processing technique used to separate monopole induced charge and dipole induced charge

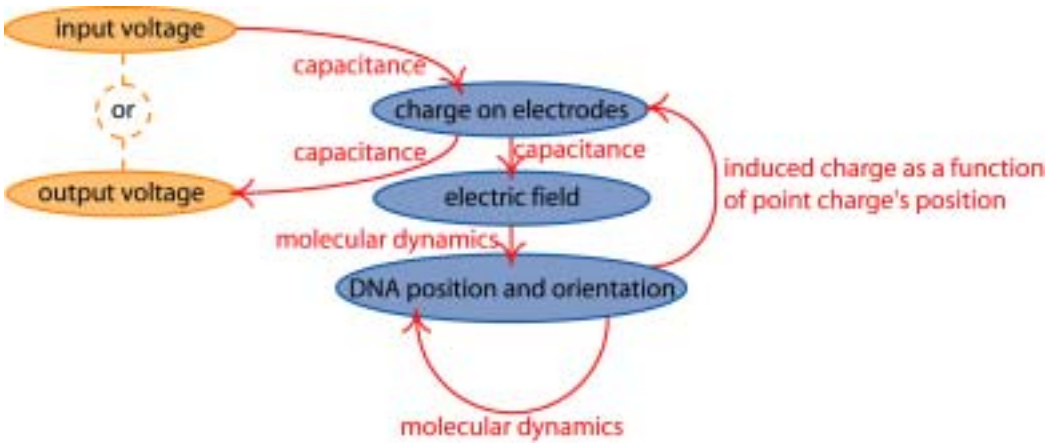

Figure 3-1. Schematic representation of the system. The bubbles represent physical quantities and the arrows represent their effects on one another. The text next to the arrows states the data needed to model the effect the arrow represents. Which of the two orange bubbles is used depends on how the system is used (see text for details); either a) a charge is applied to the electrodes (via a current source) and the voltage measured (via a volt meter), or b) a voltage is applied to the electrodes (via a voltage source) and the charge measured (via a current meter)

(see Section 4.3). In both cases salt effects on polarizability and persistence length are expected to be negligible (these ion levels are far below those usually used with DNA), and ionic effects on the real resistance seen between the electrodes are irrelevant to this analysis. Thus, ions are not included in the model.

3.2 Characterization overview

We split the characterization into several parts: First, we need to know the capacitances between the various electrodes in the NLN; since capacitance is a relationship between charge and voltage, it tells us the charge appearing as a result of an applied voltage, *or* the voltage appearing as a result of DNA induced charge. (which depends on how the system is connected - the various electrodes can either be connected to current sources (and the DNA-induced charge detected as a voltage variation), or they can be connected to voltage sources (and the DNA-induced charge detected as a current variation)[1]. In addition, the capacitance can tell us the amount of the electric energy concentrated in the fringing field (this must be small or else it will be difficult to control the DNA); this is accomplished by comparing it to an analytic approximation that

1. A voltage-source connected electrode is probably experimentally preferable: any ions in the aqueous channel will lead to a non-infinite real resistance; this will cause a charge leakage across the channel. The charge difference across an electrode gap will be maintained with a voltage source as it will replace any leaked charge in maintaining the voltage difference. A current source will not, and as a result both the charge on the electrodes and the voltage will drift.

neglects the fringing field. When the two are in agreement, we can use another analytic approximation that tells us the resultant electric field the DNA experiences. Finally, the capacitance can tell us the high-frequency coupling (cross-talk) between different pairs of electrodes.

Second, we need to know the DNA induced charge on the electrodes. We accomplish this by determining the charge induced on a single electrode by a single point charge as a function of the point charge's position; since we modeled the DNA as a set of point charges, we obtain the DNA induced charge by superposition. Once we know the charge induced on any one electrode; the others are obtainable by symmetry.

This is all the information that is needed to do a molecular dynamics simulation of the DNA in the NLN; the molecular dynamics simulation closes the loop and tells us the position and orientation of the DNA as a result of its current position and orientation and the applied signal. All of this is schematically illustrated in Figure 3-1.

3.3 Ideal case

In the remaining parts of this chapter we determine the range of geometrical parameters (width, height, spacing, gap and depth; see Figure 2-1 on page 13) over which the capacitances between the different electrodes and the DNA induced charges on the electrodes are in good agreement with the "ideal case," in which

- all of the capacitance in the system is across the NLN's nanogaps (a nanogap is the region of the channel between a pair of electrodes; it has a volume of width x height x gap as shown in Figure 2-1)
- this capacitance is equal to the no-fringing field formula $C = \varepsilon wh/g$ where ε is the dielectric constant of water (approximately $=80\varepsilon_0$), w is width, h is height, and g is gap as shown in Figure 2-1
- the dielectric disparity between water and silicon dioxide ($\varepsilon=3.925\varepsilon_0$) can be considered sufficiently large that
    - the charge induced on an electrode's face by a charge in the channel is identical to that induced on a water-silicon dioxide interface in the same position[1]

1. This is equivalent to saying that the charge induced on an electrode's face is simply whatever portion of the charge induced a water-silicon dioxide interface that is subsumed by said face.

- the charge induced on an electrode's face by a charge in the channel is not affected by movement of the charge in the channel in the direction that is perpendicular to both the channel axis and the electrode's face (this is along its height, as labeled in Figure 2-1)

• the *total* charge induced on an electrode's face by a charge in the channel is equal to $1/2 \pm y_0/g$ where y_0 is the position of the charge in the direction normal to the electrodes' faces (0 in the middle) and g is the gap as shown in Figure 2-1.

As shall be shown in this chapter, these assumptions are met if gap<<width, width ~=height, and spacing ~=width. With this set of conditions, the NLN is optimized for orienting, and detecting the orientation of, DNA's permanent dipoles (additional justification for this statement is in Section 3.5.6 on page 62).

3.4 Capacitance

3.4.1 Analytic model

The closest thing available to an analytic model of the capacitance of the electrodes is that of a parallel-plate capacitor. Utilizing conformal mapping, H. Palmer published an analytic solution for parallel-plates of negligible thickness in 1937 [60]. In 2000, H. Yang published an analytic solution for parallel-plates of non-negligible thickness, also by utilizing conformal mapping [61]. Yang's solution, however, was founded in an inaccurate equation for the negligible thickness capacitance by R. Elliott [62]. This issue was addressed by V. Leus and D. Elata in 2004 [63] by swapping the term from Elliott's approximation with that from Palmer's. This is the most accurate formula to date, being in agreement within 0.3% of finite element data for typical stripline-stripline capacitances.

Unfortunately, a stripline is a poor approximation to make when analyzing the capacitance across the electrode gap. Nonetheless, it's a good approximation to make when analyzing the capacitance across the electrode spacing (for definitions of gap and spacing, see Figure 3-16 on page 43) In the latter scenario,

$$C_{\text{spacing}} = \varepsilon \frac{h}{s}\left[1 + \frac{s}{\pi h} + \frac{s}{\pi h} ln\left(\frac{2\pi h}{s}\right) + \frac{s}{\pi h} ln\left(1 + \frac{2w}{s} + 2\sqrt{\frac{w}{s} + \frac{w^2}{s^2}}\right)\right] \tag{3.4}$$

is the Leus-Elata formula, where *C*=capacitance/depth, *h*=height, *w*=width and *s*=spacing. We might apply the Leus-Elata formula to the electrode gap (swapping *s* with *g* and *w* with *d* (depth)) but this is useless as it approaches infinity for large depth. We could also apply H. Palmer's formula,

$$C_{\text{gap}} = \varepsilon\frac{h}{g}\left[1 + \frac{g}{\pi h} + \frac{g}{\pi h} ln\left(\frac{2\pi h}{g}\right)\right] \tag{3.5}$$

where *C*=capacitance/width, *h*=height and *g*=gap; In reality it is probably in-between this and

$$C_{\text{gap, no fringing field}} = \varepsilon\frac{h}{g} \tag{3.6}$$

where C=capacitance/width, *h*=height and *g*=gap.

3.4.2 Finite element result

Equation 3.6 is actually the most accurate formula for the capacitance across the electrode gap. This is attributed to several factors limiting the fringing field:

1. The electric field lines terminating on the sides and back of the electrode are attenuated by the poor dielectric constant of SiO_2 (ε_r=3.925) as opposed to water (ε_r=80).
2. The electric field lines terminating on the back of the electrode assumed in Equation 3.5 actually terminate on adjacent electrodes.

The second effect is sufficiently dominant that if (3.5) is expanded, the 1st term's (the part of the capacitance as a result of the field lines terminating on the face, identical to Equation 3.6) ε using ε_r=80 but the 2nd and 3rd terms' (the part of the capacitance as a result of the field lines terminating on the back) ε's using ε_r=3.925, (3.5) is still more accurate.

We assemble a complete characterization of the NLN's capacitance from three separate 2D analyses. The charge densities at the electrode's corners do not add a significant amount to their total charges, and as a result 2D is preferable to 3D as a finer mesh is possible given the same computational power (since the mesh elements have one less dimension, speed and memory requirements are significantly lower). The

three analyses are a z=0 analysis (Figure 3-2 on page 27), a $y>>g$ analysis (Figure 3-7 on page 32), and an x=0 analysis (Figure 3-12 on page 36).

3.4.2.1 Definitions

We choose an arbitrary reference electrode and assign a coordinate system such that the electrode is defined by -width/2<x<width/2, -height/2<z<height/2, y>gap/2. We separate the capacitance between the reference electrode and the other electrodes into separate capacitances between it and each of the others' surfaces. We define symbols for each of these as follows (these are shown in the aforementioned figures):

- C_{ij} is the capacitance to electrode i, surface j, excluding depth dependent contributions.
- C_i is the capacitance to electrode i (sum over all faces in the current analysis)
- i=2,4,6,8,... with decreasing x for electrodes on the side of the reference
- i=1,3,5,7,... with decreasing x for electrodes on the side opposite the reference
- j=1 (y=g/2 plane),2 (x=+d/2 plane),3 (x=-d/2 plane),4 (z=+h/2 plane),5 (z=-h/2 plane)
- C_{ij}' and C_i' are similar to C_{ij} and C_i, but are the depth dependent contributions (the total capacitance to electrode i, surface j is C_{ij}+C_{ij}')
- C_{i1}' is meaningless as there is no depth dependent contribution to the j=1 surface (y=g/2 plane).
- C_{ij}^a is an analytic approximation to C_{ij}
- Note that all C_{ij}, C_{ij}' are not necessarily defined in every analysis

All analyses were done in FEMLAB 3.0 electrostatics mode. Capacitances were determined using C=Q/V. All data was solved for via a normal mesh with three refinements (about the limit of the computing power available on our computer, an Intel Pentium 4 2.0GHz with 512MB ram). All other settings were left on their defaults. The base values for all parameters were:

- gap (g) ~5nm
- height (h) ~20nm
- width (w) ~50nm
- spacing (s) ~50nm
- depth (d) ~300nm (not shown in Figure 3-16)
- padding to edge of analysis where $\hat{n} \bullet \vec{D} = 0$ ~200nm

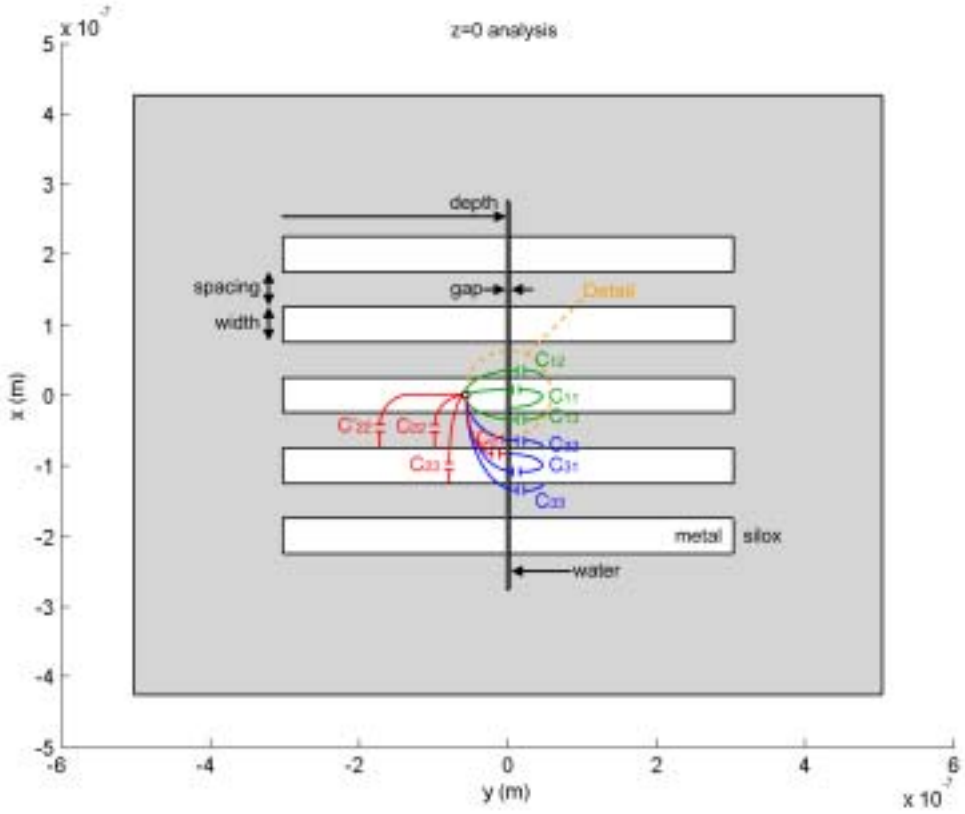

Figure 3-2. "z=0 2D analysis. Also see Figure 3-3 on page 28 (detail).

50nm is considered possible with present technology whereas 5nm is considered the edge of feasibility; the state-of-the-art devices for nanofabrication advertise a maximum resolution of 5nm whereas a factor of 10 is usually needed for reproducibility [64]. As aforementioned, nanochannels (not electrode-lined) have been manufactured as small as 10nm in diameter [21], with which these values are consistent.

3.4.2.2 *z*=0 2D analysis

This analysis has the following characteristics:

- Assumes a large height
- All data per unit height
- Accuracy increases as height approaches infinity

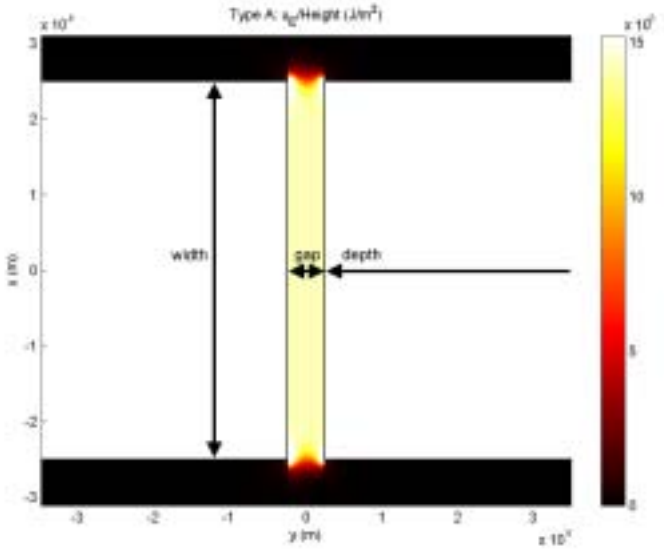

Figure 3-3. *z*=0 electric energy density. (*w*=50nm, *g*=5nm). The vast majority of the electric energy density appears in the nanogap.

- No height dependent effects
- No z-field

It is anticipated that the electrode's height will be the smallest component of its size (nominally ~20nm with current technology). Nonetheless, the assumptions of the *z*=0 analysis are still acceptably good due to the vastly different dielectricities of H_2O and SiO_2 limiting edge effects; thus, the *z*=0 analysis is a good predictor of cross-channel (across gap) and "channel-end" (between the surfaces of electrodes near the channel) capacitances. The electric energy density of the detail in Figure 3-2 is shown in Figure 3-3.

This analysis sweeps *g* from 5~50nm and *w* from *g*~50nm. It is somewhat useless to do *w*<*g* as it takes several electron beam cuts to obtain a given *w* but it takes only one to obtain a given *g*. Thus, for an obtainable *w*, *g*=*w* should easily be possible. *s* was not swept as there is no real reason for decreases in *s* (as *s* increases are mainly useful for eliminating C_2 and C_3, and at 50nm, C_2 and C_3 are already extremely small).

Depth is not too relevant in determining C_{i2} and C_{i3} (for even i) as the capacitive charge appearing on their corresponding edges falls off very rapidly with y: for even d>100nm, these data are extremely invariant (<<1%).

Figure 3-4 shows all of the different capacitances as a function of width and gap sizing.

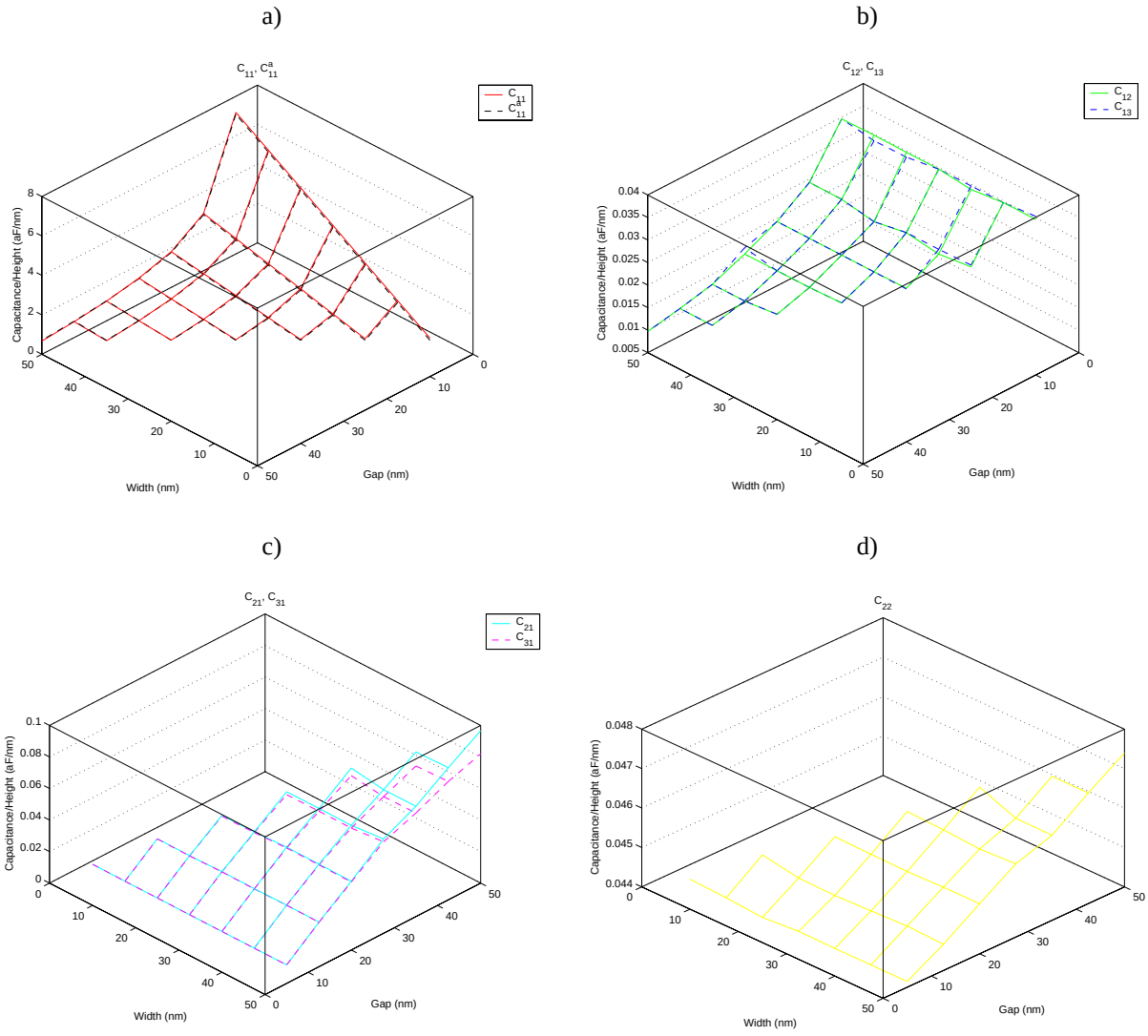

Figure 3-4. a) C_{11} and $C_{11}{}^{a}$ (equation (3.6))
b) C_{12} and C_{13}. C_{12} and C_{13} go as $1/g$, independent of w. By symmetry, C_{12}=C_{13}; as such this is a good example of the accuracy of finite elements.
c) C_{21} and C_{31}. C_{21} and C_{31} are approximately linear in g in this analysis region, with a transition to a different linear relationship at g~27.5nm. For g<27.5nm, note their similarity, attributed to the channeling effect of the dielectricity of water.
d)C_{22}. This C_{22} is not the total capacitance as shown in Figure 3-2 (see below). C_{22} is linear in g and independent of w

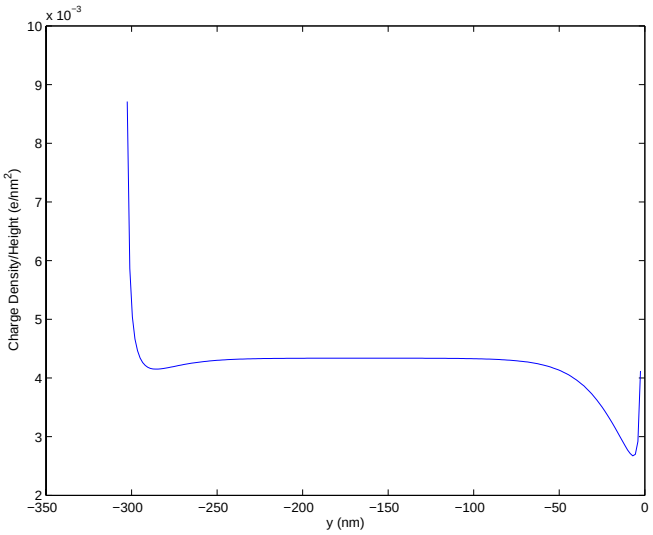

Figure 3-5. Charge appearing on boundary corresponding to C_{22} as a function of y-position (w=5nm, g=5nm)

A z=0 analysis is only valid for determining the channel-end contribution to C_{22} (this is the C_{22} shown in Figure 3-4). This channel-end contribution was determined as follows (see Figure 3-5):

1. Find the w, g pair with the longest transient decay (w=5nm, g=5nm)
2. Find the y-value where the corresponding transient falls off to 1% of its final value (y=-23.05nm)
3. Determine how far in this is (~72nm)
4. Integrate on all curves this far in to determine total charge (and therefore capacitance)

(the total charge appearing on this boundary looks similar to Figure 3-5 for all w, g, except that the height of the transient is dependent on g).

Figure 3-6 shows a summary relating all capacitances to C_{11}.

As suggested in 3.4.1 (as well as seen in Figure 3-4a), the analytic approximation in (3.6) is excellent over all w, g considered. The maximum error is <+6%/-14% over all values and <±5% over all values wherein w>=20nm and g<=42.5nm. (underapproximates for small w and overapproximates for large g)

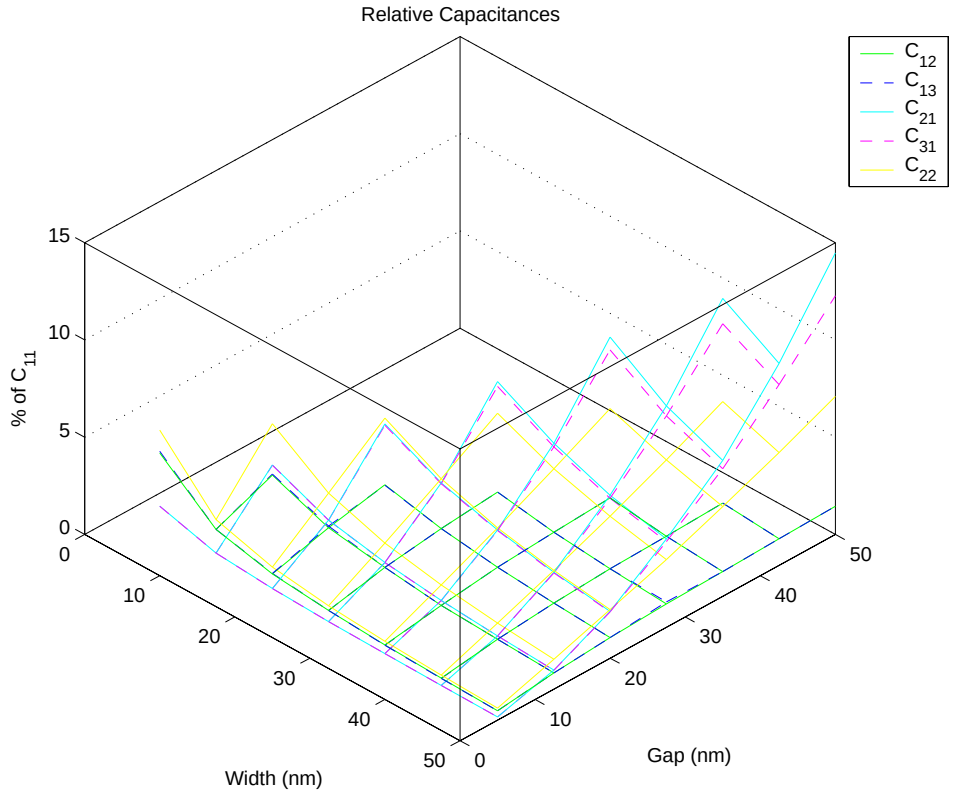

Figure 3-6. All capacitances as a percentage of C_{11}.

Note C_{32}, C_{23} and C_{33} were not included in this analysis as C_{32} was always less than 1% of C_{11} over all w, g considered while C_{23} and C_{33} were always less than.02% of C_{11}.

To address the issue of s dependence (although it is somewhat irrelevant as aforementioned), for s=200, the worst-case C_{12} and C_{13} (g=5nm) rose by ~45% while the corresponding C_{22}, C_{21} and C_{31} fell by ~74%[1]; This suggests C_{21} and C_{31} s-dependence very similar to C_{22}, meaning the results of the $y>>g/2$ analysis (see below) are probably applicable to approximating their fall-off as well (something probably not necessary in nanochannel development, however).

3.4.2.3 z=0 summary

The z=0 analysis has shown that

1. The biggest percent increases, however, were at g=50, where C_{12} and C_{13} saw increases of ~151% (still negligible). C_{11} witnessed ~17.8% increases (no appreciable increases (~ 0.26%) in C_{11} occurred at g=5). C_{21}, C_{31} fell 67% and 71%, strengthening the claim at the end of this paragraph.

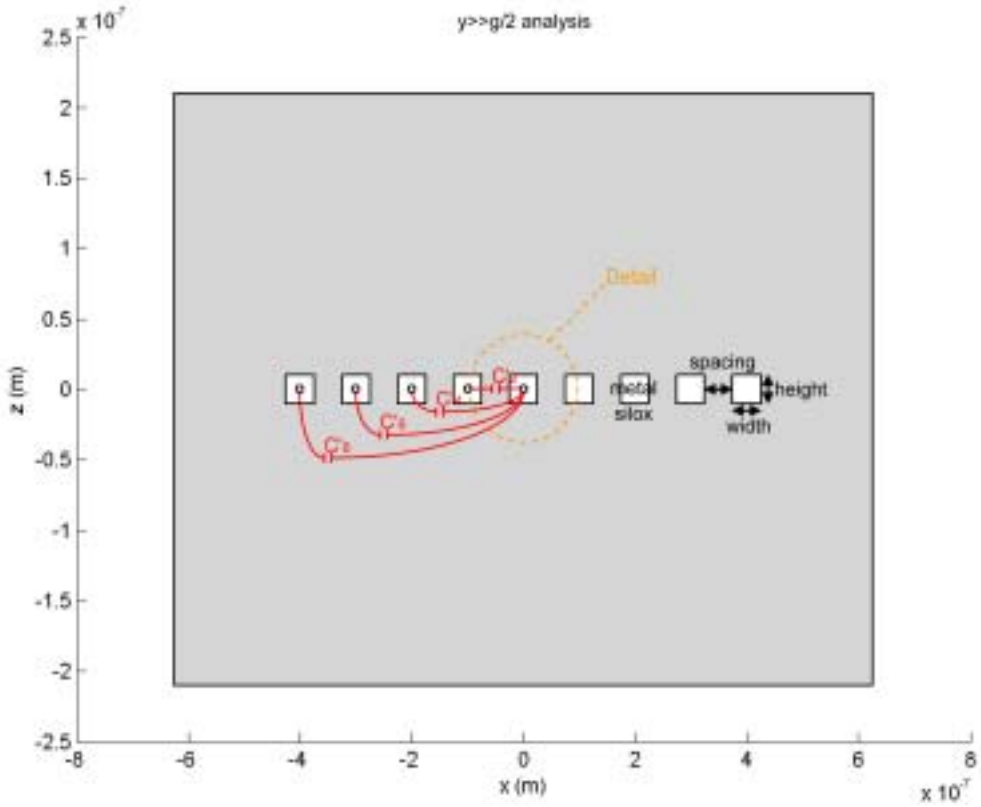

Figure 3-7. *y*>>*g*/2 analysis. Also see Figure 3-3 (detail).

- The finite element gap capacitance is extremely close to the analytic no-fringing field approximation, meaning that the fringing field is small. Thus, we can expect an electric field perpendicular to the electrode's end, with magnitude $|E|=\sigma/2\varepsilon$.
- The capacitance across the gap dominates all other channel-end capacitances; this dominance decreases with smaller widths and bigger gaps.

3.4.2.4 *y*>>*g*/2 2D analysis

This analysis has the following characteristics:

- Assumes a large depth
- All data per unit depth
- Accuracy increases as depth approaches infinity
- No depth dependent effects
- No *y*-field

The z=0 analysis does not account for capacitances due to electric fields in the z-direction, requiring the $y>>g/2$ and x=0 analyses. The cross-channel capacitances (C_{ij}, for all odd i) are generally independent of these effects as C_{i1} (the main component) uses the high dielectricity of the channel, whereas C_{i5} and C_{i6}, those occurring as a result of electric fields in the z-direction, do not[1] (see Fig. 3-12); as such, the z=0 analysis is good at determining cross-channel and channel-end capacitances, while bad at determining capacitances between parallel electrodes away from the channel, like those seen in Figure 3-7.

This analysis sweeps w from 0~50nm and s from 50~200nm. h was not considered as height is not a good parameter to adjust to affect these types of capacitances. While lessening h decreases C_2, it increases C_4, C_6, C_8, etc. In addition, it decreases C_{11}, C_{12}, and C_{13} while maintaining C_{14} and C_{15}, meaning less capacitance total but a poorer ratio of capacitance across the channel to capacitance not across the channel as well as a less ideal C_{11}; these effects both contribute to worsened control over DNA in the nanochannel that does not occur as a result of an equivalent decrease in C_2 obtained by increasing s. For comparison to Figure 3-3, the electric energy density of the detail in Figure 3-3 is shown in Figure 3-8.

As opposed to analyzing the capacitance component-wise, the low complexity of the $y>>g/2$ analysis is well tailored to curve-fitting. Figure 3-9 shows C_2 with its fit, C_2^a(3.7) where

$$C_2^a = \frac{7.376\times10^{-3}w + 4.545\times10^{-1}}{s^{0.85}} + 1.120\times10^{-4}w + 9.707\times10^{-3} \quad (3.7)$$

for C_2^a in aF/nm and s and w in nanometers. (3.7) is accurate to ±2% as long as s<200nm. For extremely large s C_2^a is extremely close to 1/s. (with error smaller than FEMLAB's).

The dependency of C_i for even i>2 on w, s is much less than that of C_2. Their dependency is so small, in fact, that it is difficult to determine it due to the inaccuracies in the finite elements. However, the relationship between C_2 and these future C_i's is not: The inaccuracy in determining this at w=50nm, s=50nm

1. *Ci2* and *Ci3* gain some benefit from the high dielectricity of the channel but are, for s<200nm, usually still smaller as a result of neighboring electrodes terminating electric field lines exiting bottom and top faces (see below).

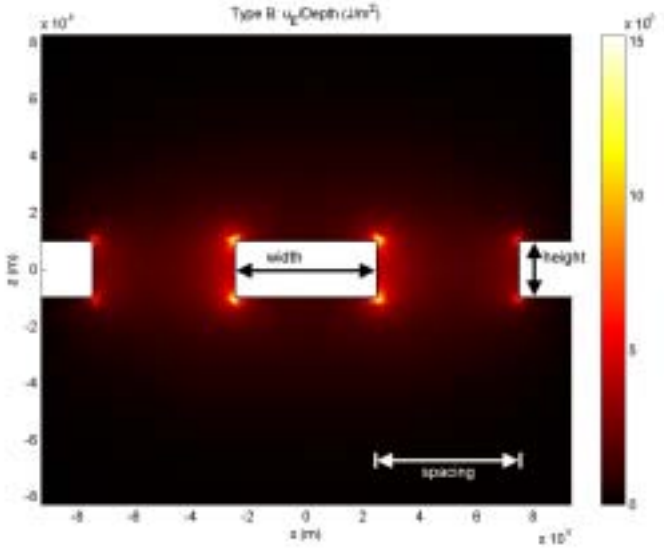

Figure 3-8. $y>>g/2$ electric energy density (w=50nm, s=50nm). The electric energy density is lower and less concentrated than that in Figure 3-3.

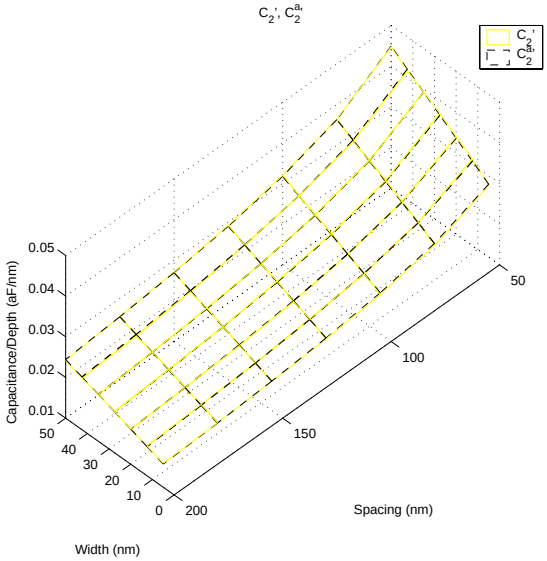

Figure 3-9. C_2 and C_2^a

and assuming independency on w, s for $i>2$ underestimates C_4 by <15% of C_2 and C_6, C_8, by <9% of C_2, the worst case errors occurring at w=50nm, s=200nm. The equation is

$$C_n^a = \frac{4.341\times10^{-2}}{n^3} + 1.402\times10^{-3} \tag{3.8}$$

for $C_n{}^a$ in aF/nm. A plot of the fit is shown in Figure 3-10.The dominance of the difference between different *n* over the variation in a given *n* as well as the dominance of the variation in C_2 over C_i for future *i* can be seen in Figure 3-11, validating the approximation.

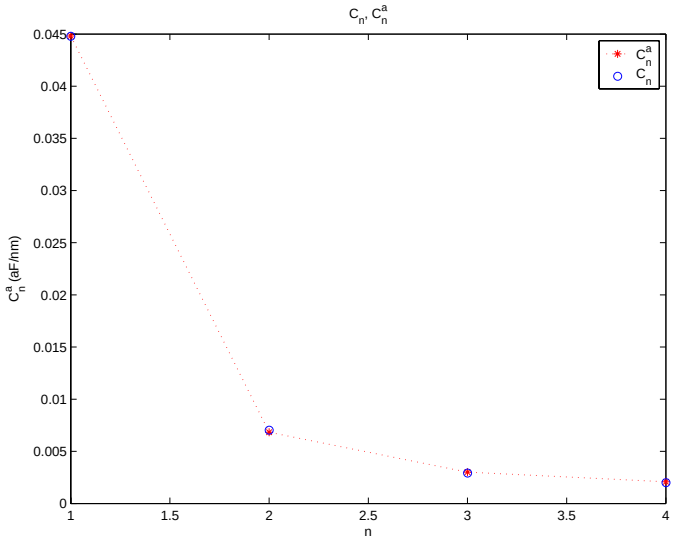

Figure 3-10. $C_n{}^a$, C_n

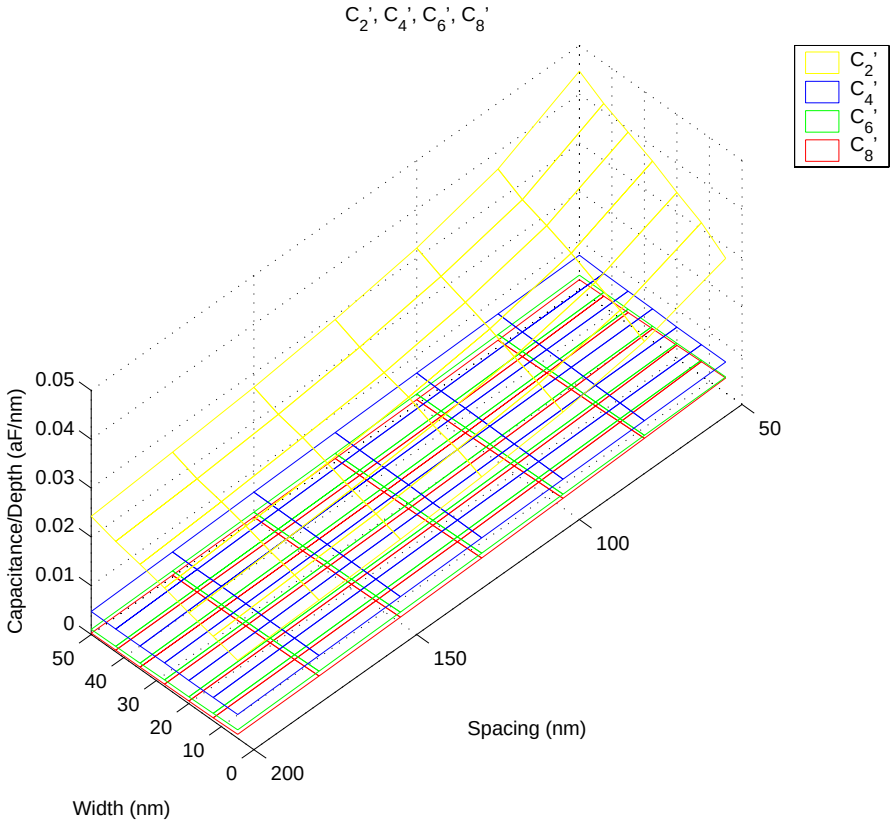

Figure 3-11. Comparison of C_n, *n*=1 to 4 (the range of the *y*>>*g*/2 analysis)

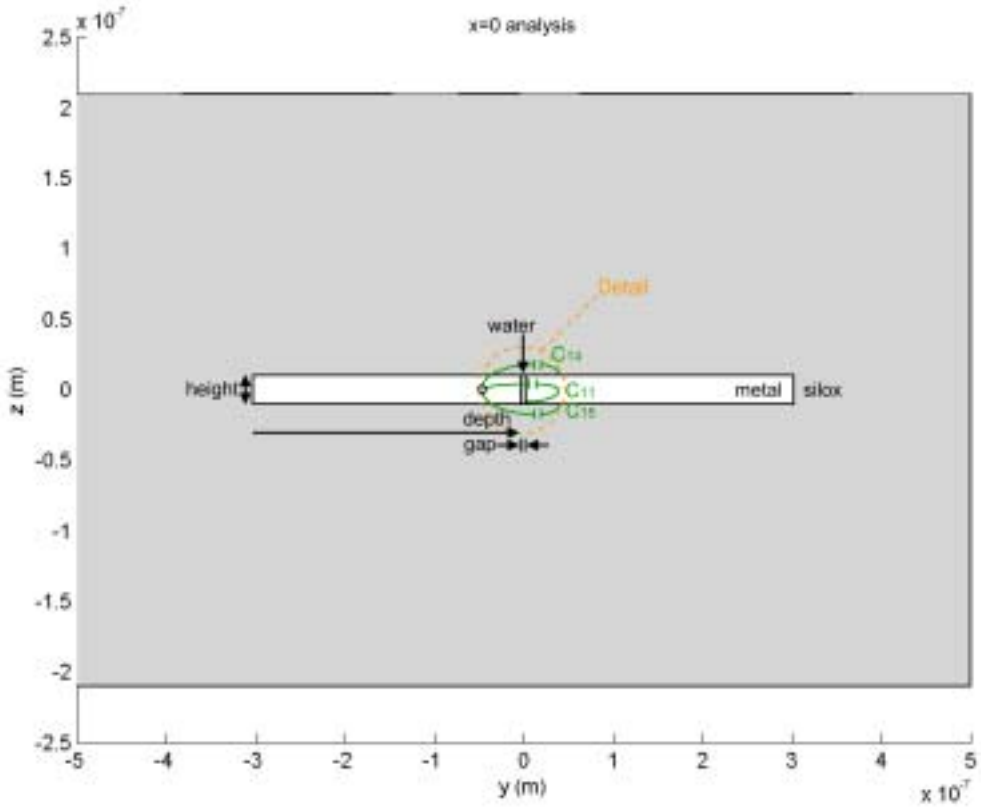

Figure 3-12. *x*=0 analysis. Also see Figure 3-13 (detail).

In practice, *s* can probably safely be incremented until a satisfactory capacitance is achieved with few tradeoffs. Only for extremely large *s* should there be a noticeable effect on performance, as *s* only affects controllability and repeatability.

3.4.2.5 *y*>>*g*/2 summary

The *y*>>*g*/2 analysis is intended as a guide to set the electrode spacing *s* on the basis of the depth needed for interconnect and the width and gap achievable. spacing=width seems satisfactory for the design parameter ranges considered in this thesis.

3.4.2.6 *x*=0 2D analysis

This analysis has the following characteristics:

- Assumes a large width
- All data per unit width

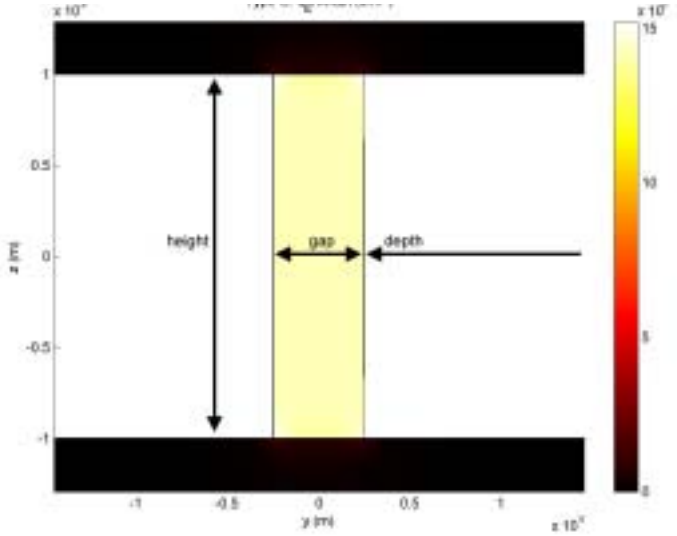

Figure 3-13. *x*=0 electric energy density (*h*=20nm, *g*=5nm). The electric energy density does not leak above and below the nanogap as it does to its sides (as seen in Figure 3-3).

- Accuracy increases as width approaches infinity
- No width dependent effects
- No *x*-field

For comparison to Figure 3-3, the electric energy density of the detail in Figure 3-12 is shown in Figure 3-13.The *x*=0 analysis has two key uses. First of all, it corroborates the *z*=0's C_{11}, as shown in Table 3-1. As these data are similar to the error in the analytic approximation to C_{11}, this is a decent corroboration.

Secondly, it is useful in determining C_{14} and C_{15}. Similar to C_{12} and C_{13}, these only differ in that they don't have either the intervening dielectric or interference from adjacent electrodes. The latter effect is dominant over the gap and spacing values considered in this analysis[1]. As such, C_{14} and C_{15} (per unit width) double as upper limits on C_{12} and C_{13} (per unit height) as spacing gets extremely large (see Table 3-2).

	g=5	*g=50*
s=50	+1.2%	-8.5%
s=200	+1.4%[a]	+7.7%[b]

Table 3-1. Difference in C_{11} (*w*=50/20**h*) between *x*=0 and *z*=0 analyses; data positive for *x*=0>*z*=0.

a. C_{11} at *s*=200 0.26% bigger than C_{11} at *s=50* in the *z*=0 analysis

b. C_{11} at *s*=200 17.8% bigger than C_{11} at *s=50* in the *z*=0 analysis

1. Attributed to the mean path distance from the reference electrode to electrode *i*=1 being much bigger than the gap, which determines the amount of the field benefitting from the aqueous dielectric.

	g=5	*g=50*
s=50	-35.8%	-66.1%
s=200	-6.72%[a]	-14.7%[b]

Table 3-2. Comparison of C_{12}, C_{13} with C_{14}, C_{15} (per unit width/height); data positive for C_{12}, C_{13}>C_{14}, C_{15}

a. C_{12}, C_{13} at *s*=200 45.4% bigger than C_{12}, C_{13} at *s=50* in the *z*=0 analysis

b. C_{12}, C_{13} at *s*=200 151% bigger than C_{12}, C_{13} at *s=50* in the *z*=0 analysis

3.4.2.7 *x*=0 summary

This analysis verifies that the capacitance across the nanogap is dominated by C_{11}. Whereas the *z*=0 analysis showed that C_{12} and C_{13} were negligible, the *x*=0 analysis shows that C_{14} and C_{15} are negligible. In addition, it shows that, since C_{12} and C_{13} are on the order of C_{14} and C_{15}, and C_{i4} and C_{i5} for all odd *i* are obviously less than C_{14} and C_{15}, we know that we do not have to consider them as well. It also validates the *z*=0's approximation that C_{11} is very close to Equation 3.6.

3.4.2.8 Notes on 2D analyses

The NLN should be developed attempting to maximize C_{11} and minimize all other capacitances. The 2D analyses should be used as a guide to the relationship between design parameters and capacitances, as well as to the ranges over which the analytic approximations (which the rest of this work assumes) are accurate.

3.5 DNA induced charge

This section discusses the charge appearing on a pair of electrodes separated by the nanogap as a result of monopoles and dipoles in the nanogap. We desire the solution for a pair of electrodes (as opposed to a single electrode) for two reasons: a) it is across a pair that the differential charge is applied to get the dipoles to align and b) the dipoles, by having two poles, tend to induce differential charge.

As aforementioned, this analysis is accomplished by determining the charge induced on a single electrode by a single point charge as a function of the point charge's position. We extend this to an electrode

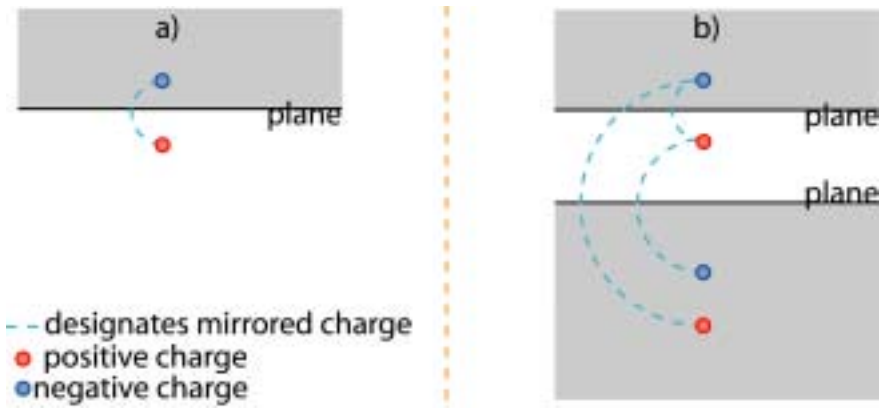

Figure 3-14. Cartoon of method of images. A positive charge is in the vicinity of a) an infinite grounded conducting plane, b) a pair of infinite parallel grounded conducting planes. In a) the charge is mirrored by a negative charge. In b) the charge is mirrored by both planes, and all of the mirrored charges are mirrored as well (1 shown). Note that the solution for a non-grounded plane can be obtained by superposition with a uniform oppositely charged surface charge density (the same surface charge density that would appear for a capacitor consisting of infinite parallel conducting planes)

pair by symmetry, and extend it to a dipole by superposition. The solution for different pairs also be generated by symmetry.

In the end the DNA can be modeled as a chain of dipoles (as we will do in the following chapter) for simple, "back of the envelope" calculations or as a set of monopoles for molecular dynamics simulation.

3.5.1 Analytic solution

An analytic solution is possible only for a pair of infinite parallel planes. This is accomplished by applying the method of images (see cartoon in Figure 3-14). First, take a grounded infinite plane at $y=g/2$. The electric field generated by some point charge q at $y=0$ is cancelled everywhere in the plane by a point charge $-q$ at $y=g$ [65]. Thus, the plane can be viewed as "mirroring" the point charge. Next, assume a point charge q at $y=0$ with infinite planes at distances $y=g/2$ and $y=-g/2$. The result is a "hall of mirrors" effect: a plane mirrors the original charge as well as all of the mirror images produced by the other plane. Here the total voltage V is[1]

1. In these equations the physical representation of 'p' is the positions of the point charges.

$$V(x,y,z) = \lim_{n\to\infty}\left\{\sum_{i=-n}^{n}\alpha(x,y,z)\Big|_{p=2ig+y_0} - \sum_{i=-n}^{n-1}\alpha(x,y,z)\Big|_{p=2(i+1)g-y_0}\right\} \tag{3.9}$$

$$\alpha(x,y,z) = \frac{kq}{\sqrt{x^2+z^2+(y-p)^2}}$$

The y-electric field E_y is

$$E_y(x,y,z) = -\frac{\partial V}{\partial y}$$

$$= \lim_{n\to\infty}\left\{\sum_{i=-n}^{n}\beta(x,y,z)\Big|_{p=2ig+y_0} - \sum_{i=-n}^{n-1}\beta(x,y,z)\Big|_{p=2(i+1)g-y_0}\right\} \tag{3.10}$$

$$\beta = \frac{kqp}{[x^2+z^2+(y-p)^2]^{3/2}}$$

The y-electric field, however, may seem nonphysical as it was derived as a result of a charge distribution in the interior of a conductor; however, an identical (physically possible) field can be generated by a charge density σ on the conductors' surfaces, where

$$\sigma_{\pm}(x,y) = \varepsilon E_y\Big|_{y=\pm g/2}$$

$$= \lim_{n\to\infty}\varepsilon\left\{\sum_{i=-n}^{n}\beta\left(x,\pm\frac{g}{2},z\right)\Big|_{p=2ig+y_0} - \sum_{i=-n}^{n-1}\beta\left(x,\pm\frac{g}{2},z\right)\Big|_{p=2(i+1)g-y_0}\right\} \tag{3.11}$$

The total surface charge Q appearing on a plate - per q - is therefore

$$Q_{\pm} = \lim_{w\to\infty}\lim_{h\to\infty}\int_{-w/2}^{w/2}\int_{-h/2}^{h/2}\sigma_{\pm}(x,z)\,dz\,dx/q$$

$$= \lim_{w\to\infty}\lim_{h\to\infty}\lim_{n\to\infty}\left\{\sum_{i=-n}^{n}\gamma_{\pm}(w,h)\Big|_{p=2ig+y_0} - \sum_{i=-n}^{n-1}\gamma_{\pm}(w,h)\Big|_{p=2(i+1)g-y_0}\right\} \quad (3.12)$$

$$\gamma_{\pm}(w,h) = \frac{p}{\pi\left(\pm\frac{g}{2}-p\right)}\,atan\left[\frac{\frac{wh}{4}}{\left(\pm\frac{g}{2}-p\right)\sqrt{\left(\pm\frac{g}{2}-p\right)^2+\frac{w^2}{4}+\frac{h^2}{4}}}\right]$$

Q_+=Q_-=1/2 (independent of y_0). A Q independent of y_0 suggests that a dipole induces no net charge on a plate in a parallel-plate capacitor, putting Equation 3.12 in seeming conflict with the witnessed dielectric effect. But this is only a seeming conflict: Note that an infinite approximation is legitimate whenever $\frac{w}{g} \gg 1 \wedge \frac{h}{g} \gg 1$, but

$$\lim_{g\to 0} Q \neq \lim_{w,\,h\to\infty} Q \quad (3.13)$$

Equation 3.12 approaches

$$Q_{\pm} = \frac{1}{2} \pm \frac{y_0}{g} \quad (3.14)$$

as n approaches infinity for finite w, h, but approaches 1/2 as w, h approach infinity for finite n; the effect of linear increases in n for finite h=w=r_0 bigger than the zero of σ always seems to be less than logarithmic, and therefore as w, h, n all approach infinity Equation 3.12 approaches 1/2, hiding the second term in (3.14) (i.e. we get the limit as g approaches 0 instead of the limit as w, h go to infinity). Figure 3-15 shows a) Equation 3.12 and b) the difference between (3.12) and (3.14) as a function of z_0 and w=h=r_0.

3.5.2 Finite element solution

The FEMLAB model used is shown in Figure 3-16, labeled with the various design parameters (previously seen in Figure 2-1 on page 13).

a) b)

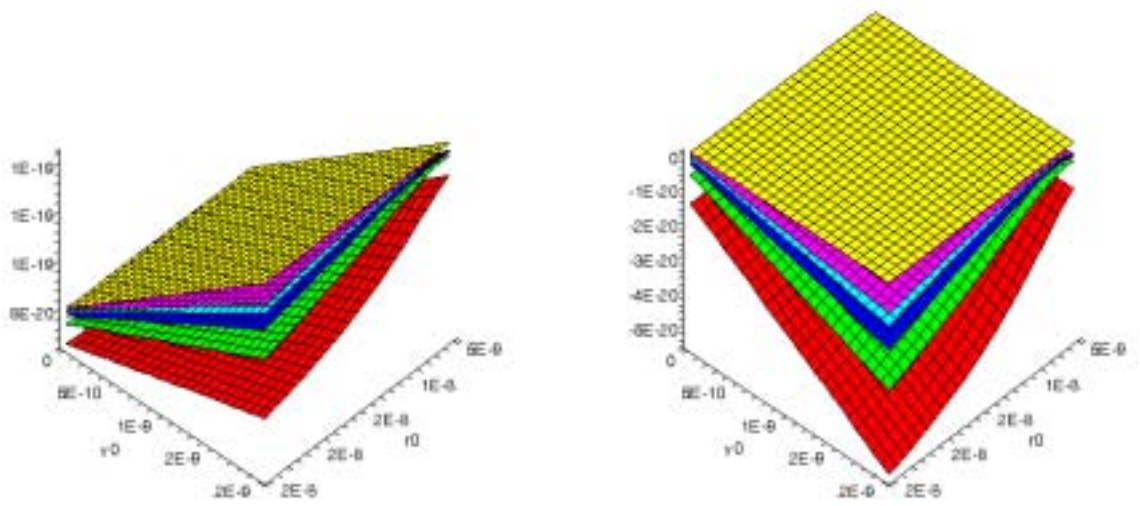

Figure 3-15. a) Equation 3.12 b) Equation 3.14-Equation 3.12. The axes are *z*-position of the point charge (z_0) and the width, height of the electrode (assuming they are equal), $w=h=r_0$. The number of mirror images n=1 (red), 2 (green), 3 (blue), 4 (cyan), 5 (magenta); yellow illustrates the limit as n approaches infinity. The *z*-axis as labeled assumes g=5nm, q=e, and shows total charge induced on the parallel plate at $+g/2$. Note how increases in n make Equation 3.12 closer and closer to Equation 3.14 over a wider and wider range. Outside this range it approaches 1/2 (see text).

. Two complete electrodes are modeled along with a small sphere of radius .25nm that has a total charge e uniformly distributed on its surface. This is an accurate representation of a point charge except that its electric field is not allowed to deviate from normality to its surface as far away as .25nm from its center. This is not much of a limitation, however, as .25nm is near enough to atomic sizes that any time the sphere is in a bizarre enough situation that this is not a good assumption, a point charge is probably a bad approx-imation to an electron cloud anyway.

We redefine our coordinate axes such that the origin is in the middle of the space between two electrode pairs; this is preferable for this type of analysis. g is swept from 5~50nm, and x_0, y_0 and z_0 swept to within 0.25nm of all of the nanochannel's surfaces.

The nominal parameter values (similar to the 2D analyses) are shown in Figure 3-16 on page 43:

- gap=5nm (g)
- height=20nm (h)
- width=50nm (w)
- spacing=50nm (s)
- depth=50nm (d)

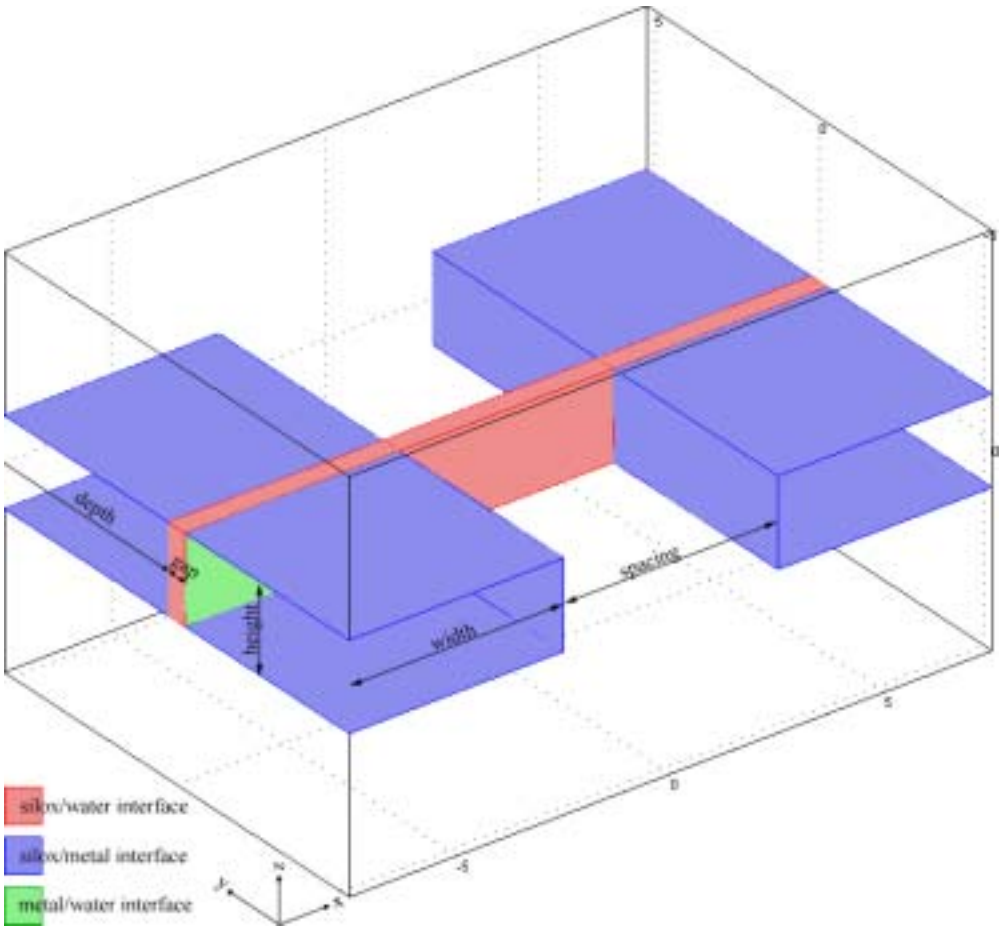

Figure 3-16. FEMLAB model of the nanochannel. Two complete electrodes are included. The electrode's depth is assumed infinite as it must be substantially larger than the other dimensions for interconnect purposes. As depicted, gap=5nm, height=20nm, width=50nm, spacing=50nm.

- the sphere is nominally at x_0=-50nm, y_0=0 and z_0=0.

The depth *d* is fixed and was selected such that the worst-case induced charge on any depth-dependent surface was beneath the noise floor of the system at the edge of the simulation environment. The system is extremely resilient to fluctuations in the height *h*. Even those by a factor of +100%/-50% from the nominal case cause less than a ~1% change in induced charge. The spacing *s* only has effects at the system's extreme edge cases, and these are hardly noteworthy. *w* data can be easily extrapolated and as thus is not swept either. Even though z_0 **was** swept, none of the data for these sweeps appears as a result of all z_0 dependencies being beneath the noise floor of the system[1].

It may seem odd, but noise nonetheless occurs in the data as a result of FEMLAB's random meshing algorithm: two attempts with the same input produce different output. To account for this, all sweeps were repeated; this was more computationally efficient than refining or turning up the fineness of the meshing algorithm, as neither refined meshes nor finer meshes resulted in a statistically significant improvement in the data (i.e. the difference between a run and a run with one refinement was less than the difference between two of the same runs). Additional refinements were not attempted as limitations on computational power rendered them impractical.

3.5.3 Measurement of *Q*

See Equation 3.9 on page 40 for the definition of Q.

We determine the charge induced on a single electrode by a single point charge (Q, in units of charge induced per unit point charge) as a function of the point charge's position (given by x_0, y_0, and z_0). We separate Q into components, one component per surface, just as we did with capacitance. We retain two subscripts for subscript consistency, however, we only show the data for one electrode (Q_1), defined by $-s/2-w<x<-s/2$, $y>g/2$, and $-h/2<z<h/2$ (an arbitrary choice). Q_{1j} is thus the charge induced on surface j, where

- j=1 ($y=g/2$ plane),2 ($x=-s/2$ plane),3 ($x=-s/2-w$ plane),4 ($z=+h/2$ plane),5 ($z=-h/2$ plane)
- Q's for the other electrodes can be trivially obtained by symmetry (a coordinate transformation).

Rather than averaging the data from different runs, it is shown as a set of semi-transparent surfaces, so as to illustrate its precision. Four runs were done for the base case, g=5nm (as it had a significantly higher noise floor than the other analyses), and two runs were done for g=12.5, 27.5, 42.5, and 50nm.

In the following figures, the x- and y-axes are consistent with the analytic approximation and Figure 3-16.

1. Although z_0 dependencies on electrode's sides were characterizable in a few cases, the simulation-to-simulation fluctuation from random mesh effects on its face was always bigger, meaning these were statistically insignificant.

The data for the g=5 run is shown in Figure 3-17.

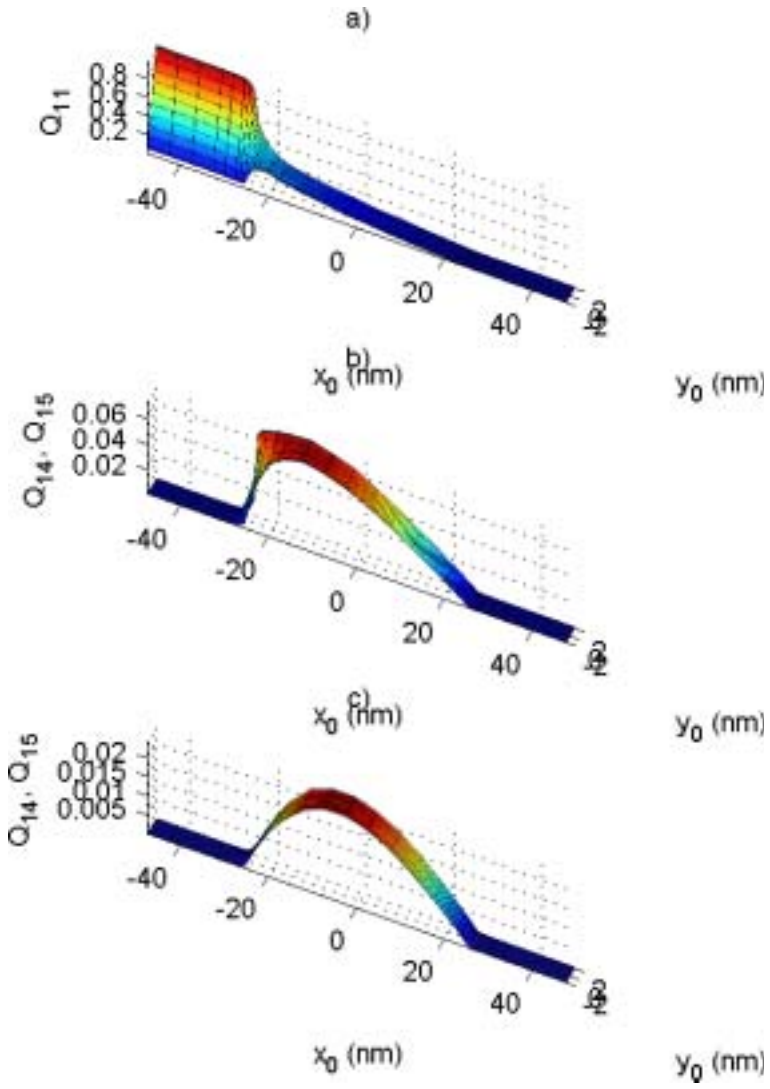

Figure 3-17. Q_1 for g=5

The data for the g=12.5 run is shown in Figure 3-18.

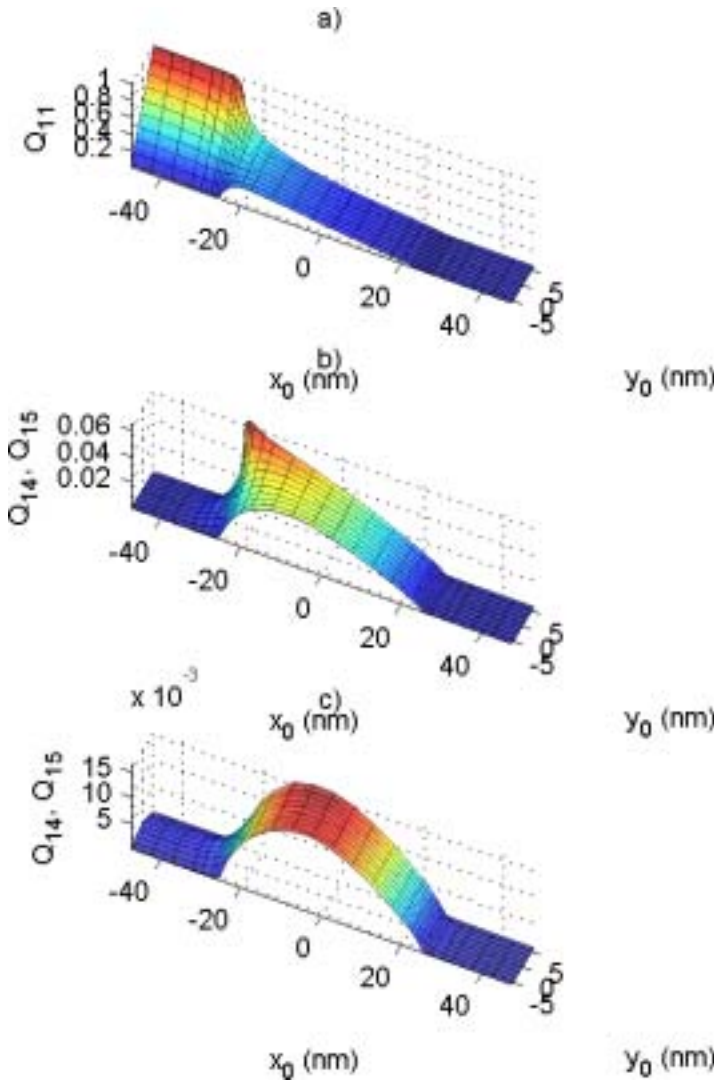

Figure 3-18. Q_1 for g=12.5

The data for the g=27.5 run is shown in Figure 3-19.

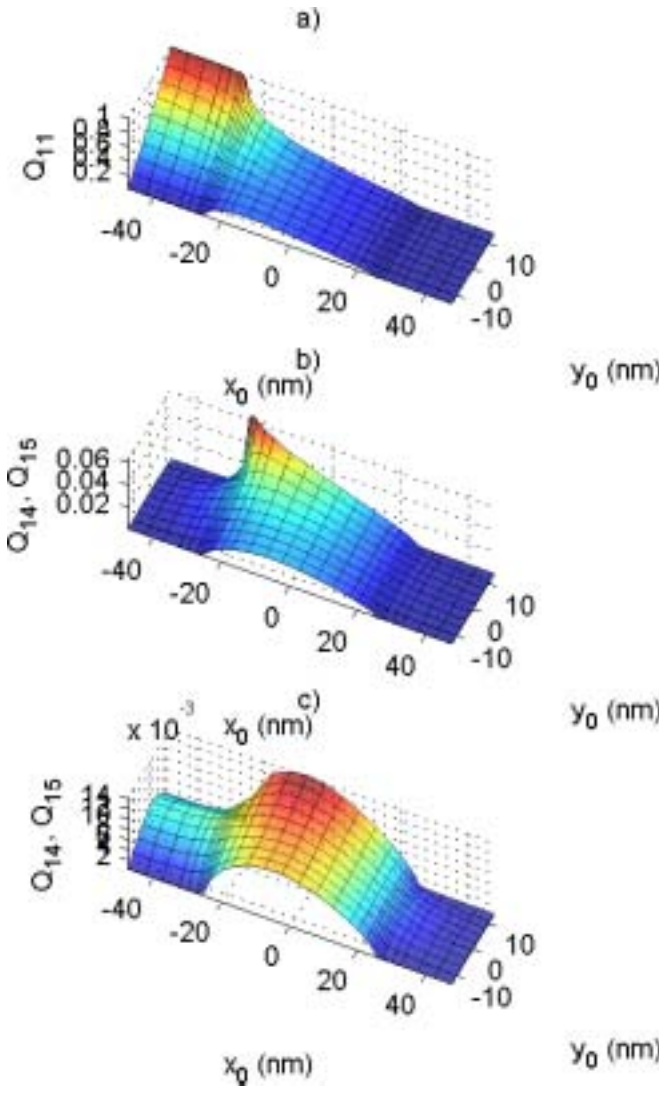

Figure 3-19. Q_1 for g=27.5

The data for the g=42.5 run is shown in Figure 3-20.

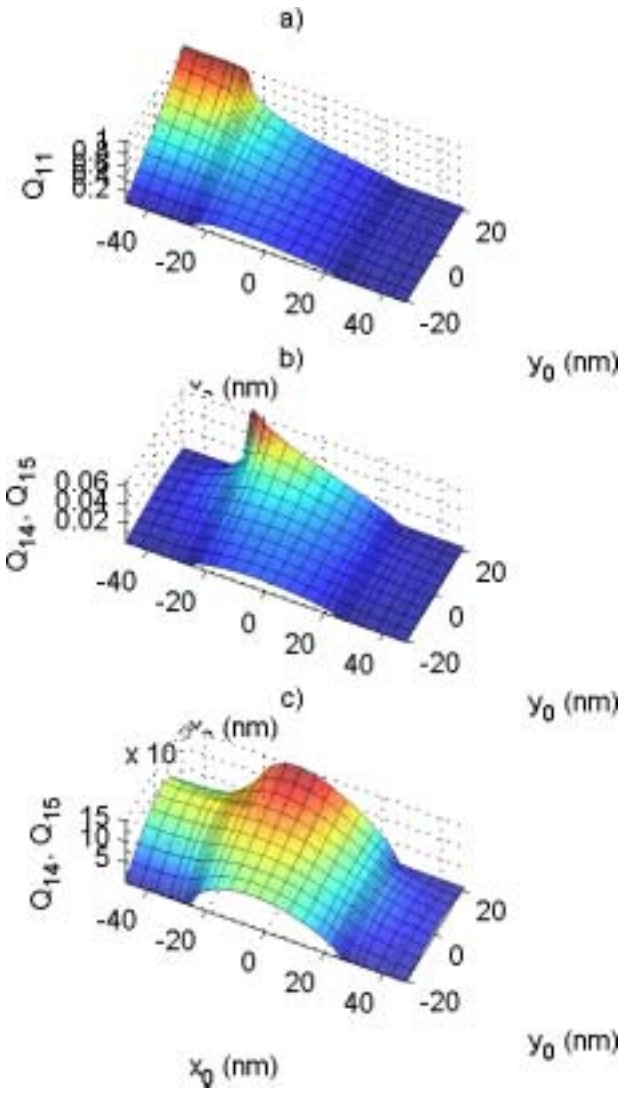

Figure 3-20. Q_1 for g=42.5

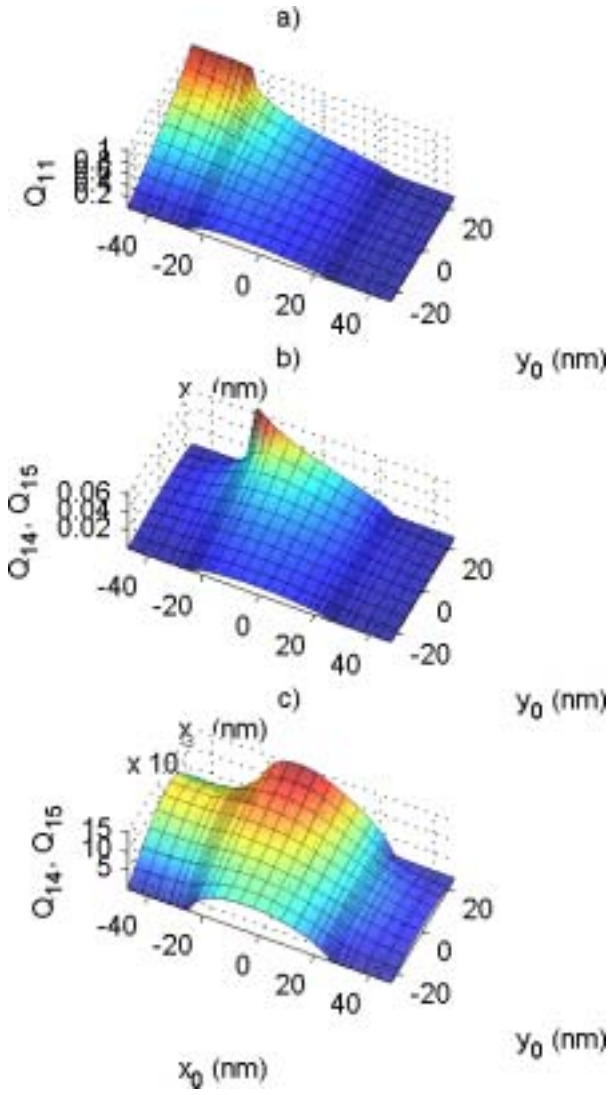

Figure 3-21. Q_1 for g=50

The data for the g=50 run is shown in Figure 3-21.

3.5.4 Analysis of *Q*

There are several relevant points of interest in these plots: first of all, the left hand-side area in the nanogap is linear for small g, going between 0 and 1, just as in the analytic approximation. As g increases, nanogap performance becomes less and less linear. For optimum performance, a very linear behavior in the nanogap is highly desirable (the reason why this is desirable is discussed in Section 3.5.5). The reason why they occur is discussed here: a plot of the position-dependent charge distribution (σ in Equation 3.11 on

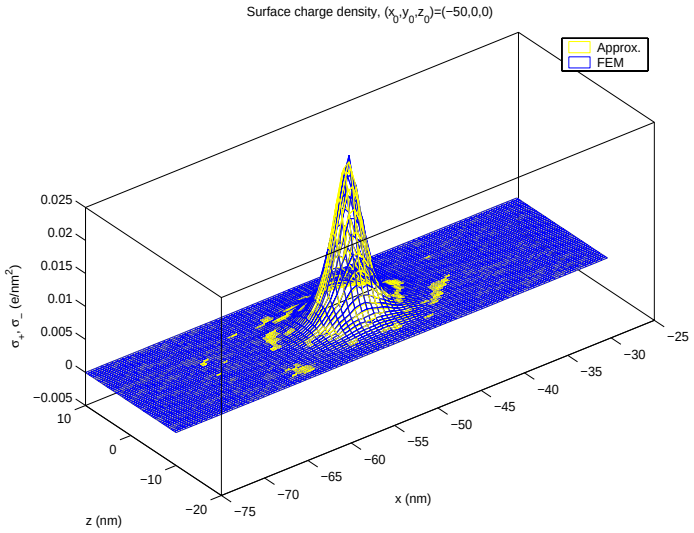

Figure 3-22. The charge density for $\{x_0, y_0, z_0\}$={-50, 0, 0}, g=5nm

page 40) is shown in Figure 3-22 for x_0=-50, y_0=0, z_0=0 (sphere centered in the nanogap), g=5, along with the analytic solution (Equation 3.11 on page 40)

For $\{x_0, y_0, z_0\}$={-23, 0, 0}, (sphere at the electrode's x-edge), it looks like Figure 3-23: σ is still approximately ideal; however, here, a part of it is now no longer on the electrode's surface and thus Q_{11} decreases by the integral of the part of σ that is "off the edge."

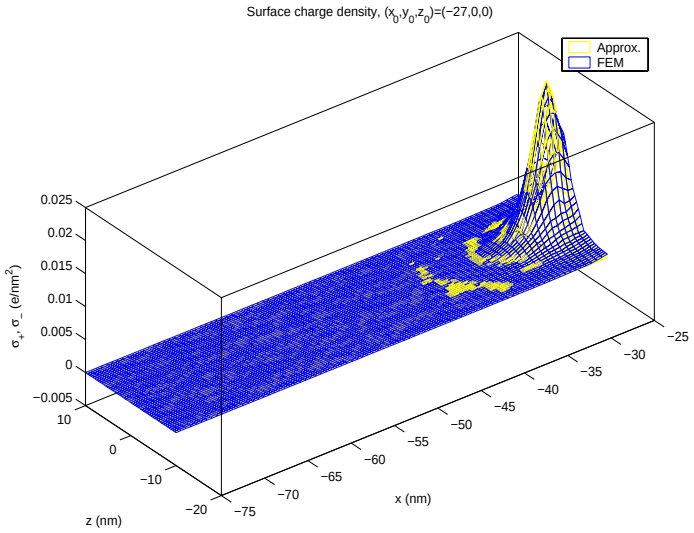

Figure 3-23. The charge density for $\{x_0, y_0, z_0\}$={-23, 0, 0}, g=5nm

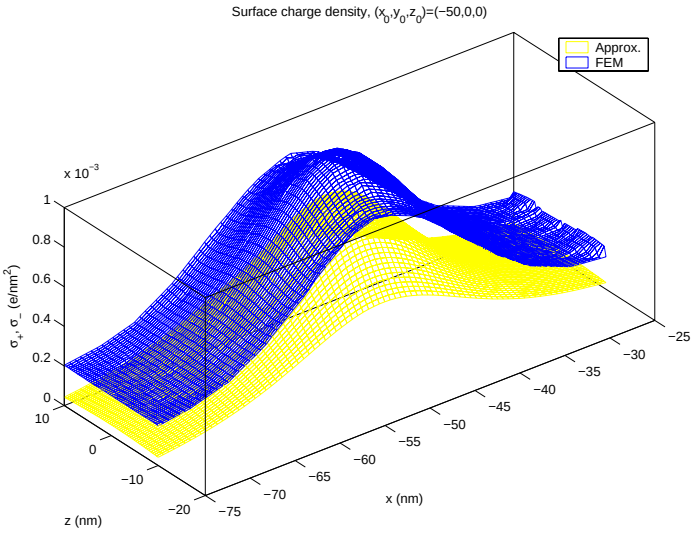

Figure 3-24. The charge density for $\{x_0, y_0, z_0\}$={-50, 0, 0}, g=27.5nm. The behavior at x=~25nm occurs as a result of artifacts in the finite element analysis; these are accounted for in determining the total charge, which uses interpolation.

If the sphere is repositioned at $\{x_0, y_0, z_0\}$={-50, 0, 0} while the gap is now widened to g=50, it looks like Figure 3-24. At g=27.5nm a good deal of σ is, even for a charge in the middle of the channel, not in the nanogap. σ is in addition no longer in good agreement with the analytic solution, but this is a result of a side effect (see below). The more relevant "spreading" effect occurs in general if a point charge is moved away from an electrode (on a path normal to its surface); this is shown in Fig. 3-25 which shows the charge induced on an electrode for a point charge at two different y-positions (y_0's).

a) b)

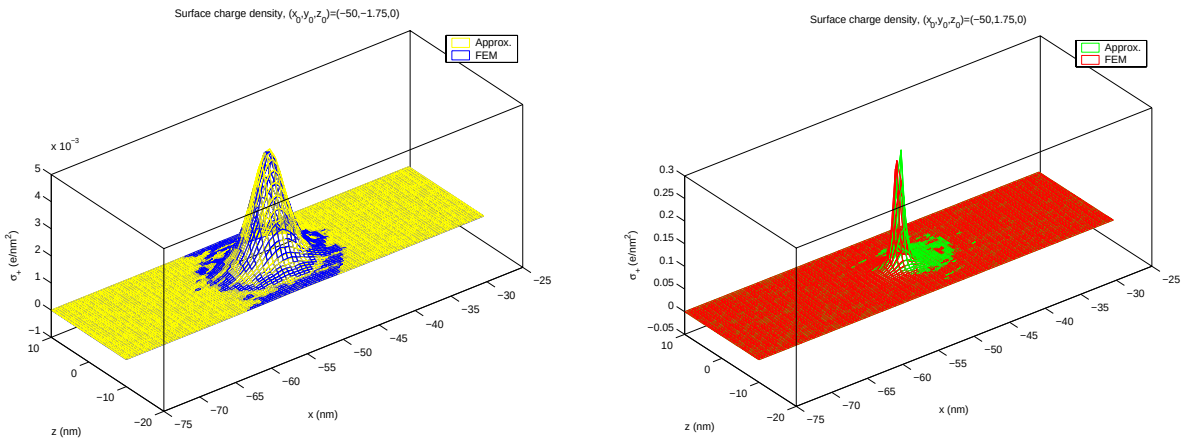

Figure 3-25. a) The charge density for $\{x_0, y_0, z_0\}$={-50, -1.75, 0}, g=5nm b) The charge density for $\{x_0, y_0, z_0\}$={-50, +1.75, 0}, g=5nm (both of these plots show σ_+ (i.e. the σ corresponding to C_{11}); note that by symmetry σ_- in a) is the same as σ_+ in b) and vice-versa).

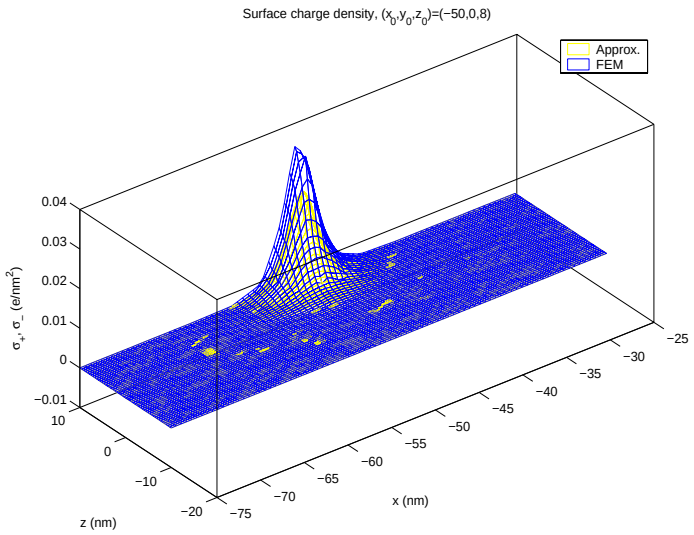

Figure 3-26. The charge density for $\{x_0, y_0, z_0\}=\{-23, 0, 7\}$, g=5nm

The charge in Figure 3-25 b) is more concentrated, as well as bigger. Note that the non-idealities of Figure 3-24 are not present here, despite the spreading effect occurring. We attribute these non-idealities to the same effect that prevents *z*-dependence: The part of σ that would fall outside the electrode on one of its *z*-edges is channeled is greatly inhibited by the silicon-water high dielectric disparity as shown in Figure 3-26; this is similar to Figure 3-22 on page 50, except that the sphere/point charge has been moved to the *z*-edge. σ increases to compensate for a good deal of whatever would otherwise have been "lost" over the *z*-edge.

If we change x_0, y_0, z_0 in Equation 3.12 on page 41 to account for different positions of the electrode relative to the charge ("electrode relative to the charge" since (3.12) is in charge centered coordinates) and replace the height *h* by a larger **effective** height *h'* (as a result of the dielectric interface), we obtain an approximate *Q*. Equation 3.12 on page 41 becomes

$$
\begin{aligned}
Q_{\pm} &= \int_{w'}^{w''}\int_{-h'}^{h'} \sigma_{\pm}\,(x,z)\,dz\,dx/q \\
&= \lim_{n\to\infty}\left\{ \sum_{i=-n}^{n} \gamma_{\pm}\,(w', w'', h, z_0)\Big|_{p=2ig+y_0} - \sum_{i=-n}^{n-1} \gamma_{\pm}\,(w', w'', h, z_0)\Big|_{p=2(i+1)g-y_0} \right\} \qquad (3.15)\\
\gamma_{\pm} &= \frac{p}{2\pi\left(\pm\frac{g}{2}-p\right)}\left\{ atan\left[\frac{w''h'}{\left(\pm\frac{g}{2}-p\right)\sqrt{\left(\pm\frac{g}{2}-p\right)^2+w''^2+h'^2}}\right] - atan\left[\frac{w'h'}{\left(\pm\frac{g}{2}-p\right)\sqrt{\left(\pm\frac{g}{2}-p\right)^2+w'^2+h'^2}}\right]\right\}
\end{aligned}
$$

where $w'' = -x_0 - s/2$, $w' = w'' - w$ and the effective height $h'/2$ is a parameter of fit.

This does a somewhat decent job at predicting linearity in the nanogap but grossly overoptimizes the falloff for x_0 outside it. Figures 3-27 and 3-28 show the approximation for 5 and 50 nanometers respectively:

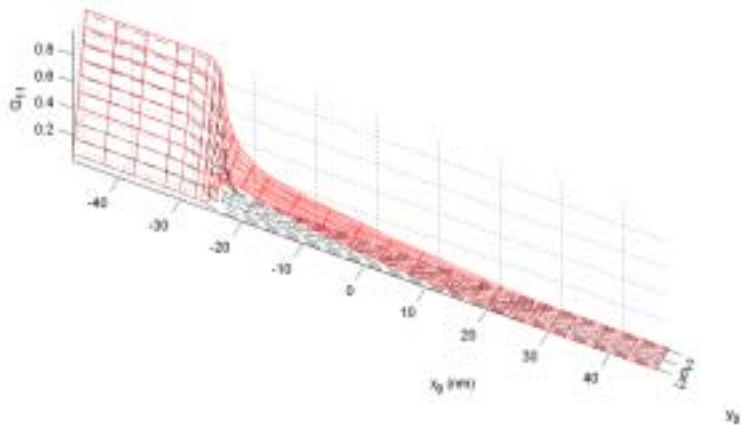

Figure 3-27. Comparison of analytic *Q* with large effective height (*h'*=500nm) (black) and FEM (red) Q for *g*=5. (n=30)

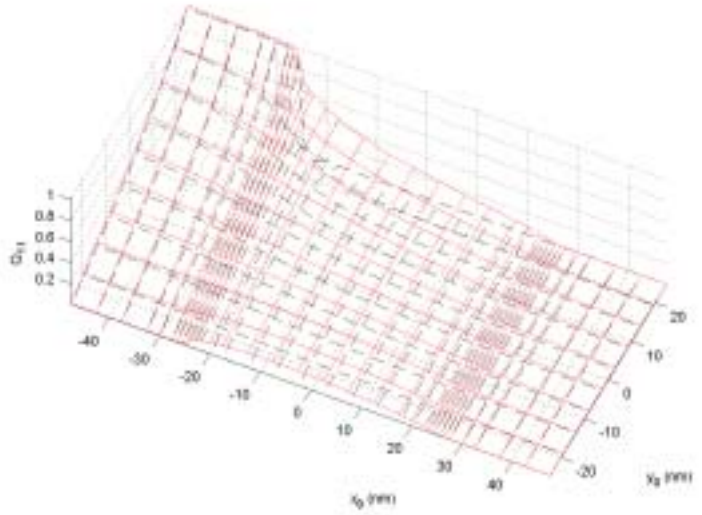

Figure 3-28. Comparison of analytic *Q* with large effective height (*h'*=500nm) (black) and FEM (red) Q for *g*=50. (n=30)

In summary, as y_0 gets further from the electrode's surface, σ becomes more and more spread out. As x_0 gets further from the middle of the electrode, more and more of this spread out σ "falls off the edge" of the electrode's surface. As more and more σ "falls off the edge," *Q* becomes more and more non-linear. The implications this has for *w* and *g* are cartooned in Figure 3-29.

3.5.5 Decomposition of *Q*

While the plots of Figures 3-17~3-21 can be invoked to determine the induced charge as a result of DNA dipole orientation, there are better representations for optimizing them from a human perspective. For *Q* data with a small 2nd derivative, one can do a linearization over the dipole separation: for a monopole let

$$\begin{aligned} M_{cm} &= q \cdot \mu_{cm} \\ \mu_{cm} &= \frac{Q(x,y) + Q(x,-y)}{2} \end{aligned} \tag{3.16}$$

$$\begin{aligned} M_d &= q \cdot \mu_d \\ \mu_d &= \frac{Q(x,y) - Q(x,-y)}{2} \end{aligned} \tag{3.17}$$

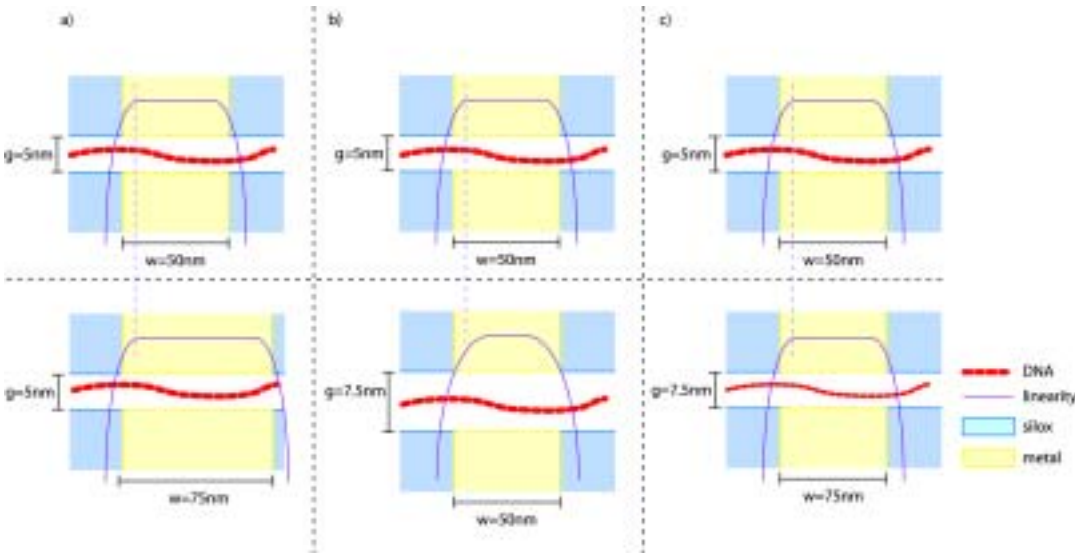

Figure 3-29. Cartoon of linearity of *Q* overlaid over nanogaps with a variety of *w*'s and *g*'s. The top three are all *w*=50, *g*=5nm, and the bottom three are a) 50% increase in *w* b) a 50% increase in *g* and c) a 50% increase in *w* and *g*. a) shows how increases in *w* increase the size of the linear region while maintaining the non-linear (transition) region. b) shows how increases in *g* increase the size of the non-linear region while decreasing the linear region. c) shows how proportional increases in *w* and *g* maintain the relative sizes of the non-linear and linear regions. Note, however, that in c) there is an equal number of DNA bases in the linear region in a) and the non-linear region in b). This scaling cannot be applied recklessly, however, as eventually increases in *h* will need to occur: for a large enough *w* and *g* the dielectric disparity's ability to contain the leakage will be compromised. In addition, one needs to allow for increases in *s* to maintain dominance of C_{11} over the other capacitances, and limit *Q* cross-coupling (i.e. charge being induced on two neighboring electrode pairs), although a good deal of the *Q* cross-coupling suggested by Figures 3-17~3-21 can be filtered (in the decomposition of *Q* in the next section, note that the components of *Q* outside the nanogap are separated out). This is why *s*, as aforementioned, can probably be neglected in optimizing *Q*.

therefore in a pair of electrodes one has total charge $q=M_{cm}+M_d$ whereas the other has total charge $q=M_{cm}-M_d$. For a dipole with $+q$ at x_1, y_1 and $-q$ at x_2, y_2, let

$$\Delta_{cm} = q\frac{[Q(x_1,y_1)-Q(x_2,y_2)]+[Q(x_1,-y_1)-Q(x_2,-y_2)]}{2} \tag{3.18}$$

$$\Delta_{d} = q\frac{[Q(x_1,y_1)-Q(x_2,y_2)]-[Q(x_1,-y_1)-Q(x_2,-y_2)]}{2} \tag{3.19}$$

If we assume *Q* linear in-between the two separated charges of a dipole and representing the dipole as a dipole moment $\vec{p}$ at a position *x, y*,

$$\begin{aligned} q[Q(x_1,y_1)-Q(x_2,y_2)] &= [\nabla Q \bullet \vec{p}](x,y) \\ &= \left|\vec{p}\right|[Q_x cos\angle\vec{p} + Q_y sin\angle\vec{p}](x,y) \end{aligned} \tag{3.20}$$

$$
\begin{aligned}
q[Q(x_1, -y_1) - Q(x_2, -y_2)] &= [\nabla Q \bullet \vec{p}^{*}](x, -y) \\
&= \left|\vec{p}\right|[Q_x \cos\angle\vec{p} - Q_y \sin\angle\vec{p}](x, -y)
\end{aligned}
\tag{3.21}
$$

therefore

$$
\begin{aligned}
\Delta_{cm}(x, y) &= \left|\vec{p}\right|(\delta^x_{cm} \cos\angle\vec{p} + \delta^y_{cm} \sin\angle\vec{p}) \\
&= \vec{p} \bullet \langle\delta^x_{cm}|\delta^y_{cm}\rangle = \vec{p} \bullet \vec{\delta}_{cm} \\
\delta^x_{cm}(x, y) &= [Q_x(x, y) + Q_x(x, -y)]/2 \\
\delta^y_{cm}(x, y) &= [Q_y(x, y) - Q_y(x, -y)]/2
\end{aligned}
\tag{3.22}
$$

$$
\begin{aligned}
\Delta_d(x, y) &= \left|\vec{p}\right|(\delta^x_d \cos\angle\vec{p} + \delta^y_d \sin\angle\vec{p}) \\
&= \vec{p} \bullet \langle\delta^x_d|\delta^y_d\rangle = \vec{p} \bullet \vec{\delta}_d \\
\delta^x_d(x, y) &= [Q_x(x, y) - Q_x(x, -y)]/2 \\
\delta^y_d(x, y) &= [Q_y(x, y) + Q_y(x, -y)]/2
\end{aligned}
\tag{3.23}
$$

$\vec{\delta}_{cm}$ and $\vec{\delta}_d$ for dipoles are analogous to μ_{cm} and μ_d for monopoles. Δ_{cm} and Δ_d are dot products of a vector ($\vec{p}$) with a vector $\vec{\delta}_{cm}$ or $\vec{\delta}_d$ as dipoles have position and orientation; M_{cm} and M_d are products of a scalar (q) with a scalar μ_{cm} or μ_d as monopoles have position only. The $\vec{\delta}$ and μ functions are much more interesting than the Q functions because they directly tell us what extractable signals will occur on an electrode as a result of a monopole or dipole in the nanogap, the dipole result being much more relevant for DNA.

Note that the $\vec{\delta}$ functions are dependent on ∇Q which is heavily dependent on gap size ($\nabla Q \approx 1/g$ in the middle of the nanogap) The claim that s is easily filtered, is because signals occurring as a result of monopoles and dipoles **not in** the nanogap tend to be common mode, but the signals occurring as a result of monopoles and dipoles **in** the nanogap tend to be differential. Far away enough from the nanogap that Q_y is negligible (a small distance), these appear as a part of $\delta_{cm}{}^x$, implying two things: First of all, since the applied field is differential, they will statistically appear on both sides equally with opposite signs,

meaning that they probably sum to zero. Secondly, the fact that they occur out of the nanogap assures they will appear on the adjoining electrodes as well (thus a multi-electrode common-mode filter could be applied). While one could apply opposite common mode signals to two neighboring electrode pairs in an attempt to align them in the *x*-direction (and use this as some sort of additional sensing method), this is pointless as there would be strictly less signal than the nanogap, and in addition it would add axial movement of the DNA by electrophoresis.

One other thing of interest is that the slope of μ_{cm} for -s/2<*x*<s/2 (i.e not in a nanogap) is extremely linear while μ_d for -s/2<*x*<s/2 is extremely small. This suggests a small enough *s* that there is substantial mirroring of charge between neighboring pairs of electrodes[1]. While noteworthy, this has no real implications for NLN design for the reasons discussed in the preceding paragraph.

The following analysis of $\vec{\delta}$ and μ for the Q's in 3.5.3 verifies the above claims as well as providing some good characterizations of nanochannel quality. *In the following plots,* $Q = \sum_i Q_{1j}$*:*

1. Note that in the nanogap, which has significant mirroring about the *y*-axis, μ_d is highly linear in *y*, μ_{cm} small and flat in *x*. For *x* outside the nanogap, μ_{cm} is highly linear in *x*, small and flat in *y*, suggesting mirroring about the *x*-axis.

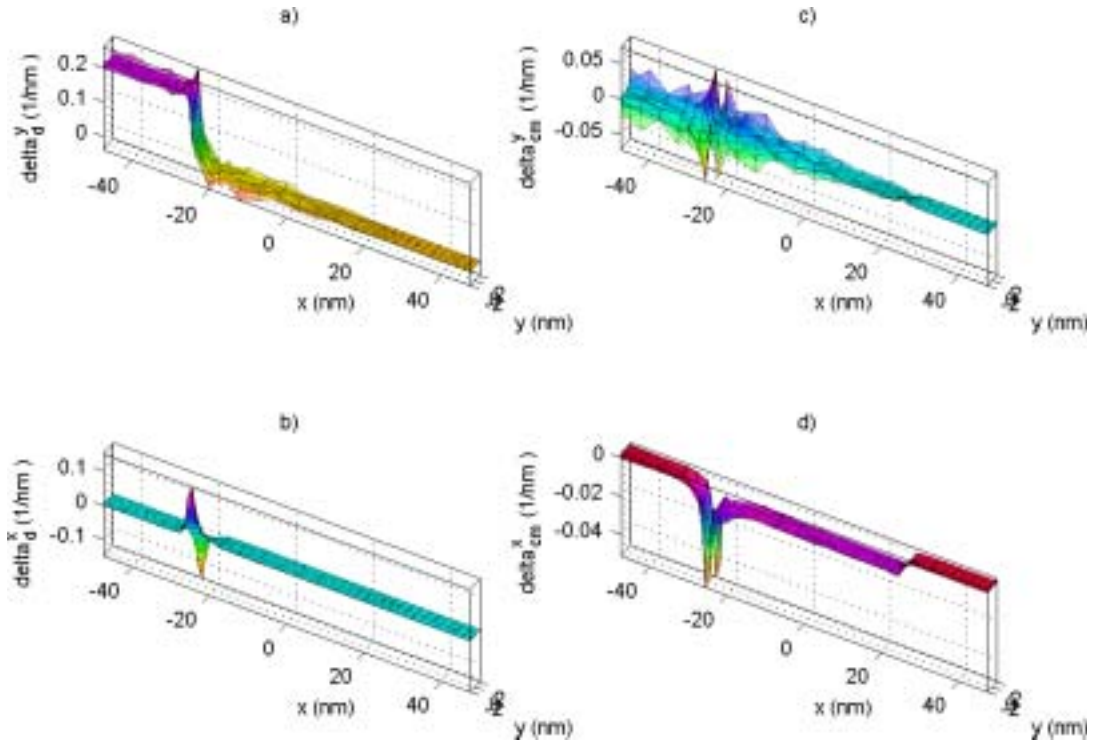

Figure 3-30. $\hat{\delta}$ for g=5nm.

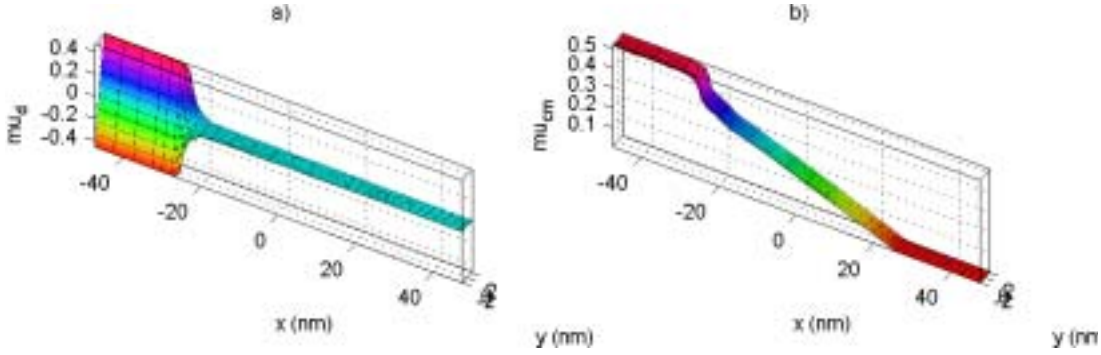

Figure 3-31. μ for g=5nm.

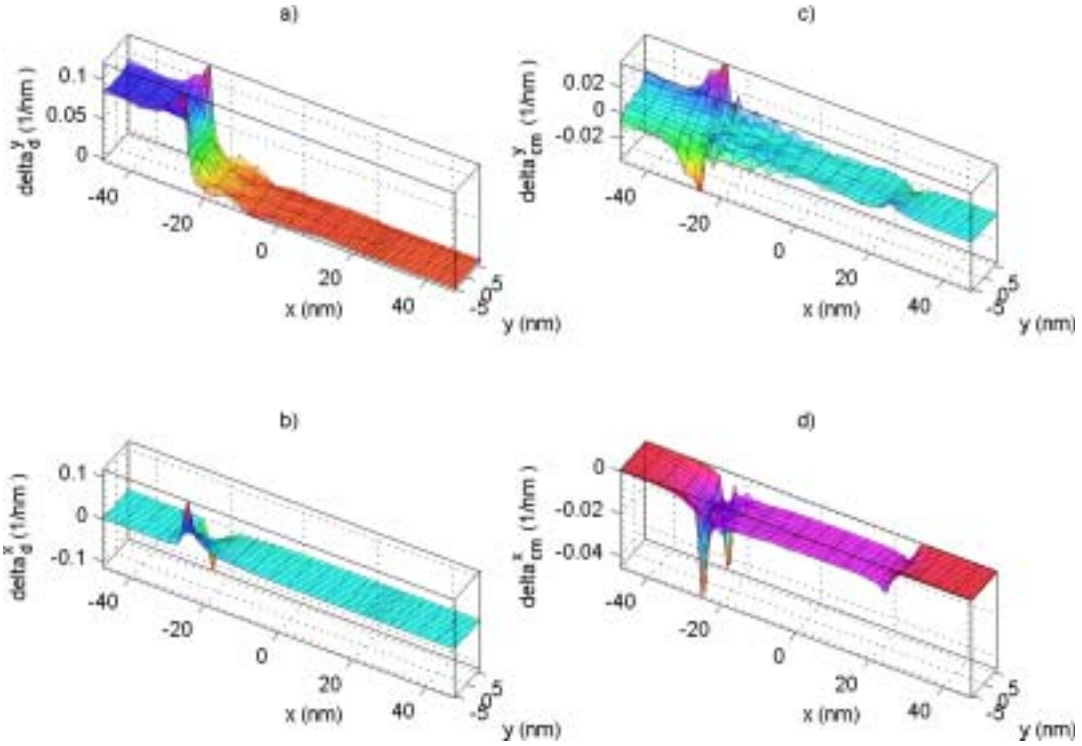

Figure 3-32. $\vec{\delta}$ for g=12.5nm.

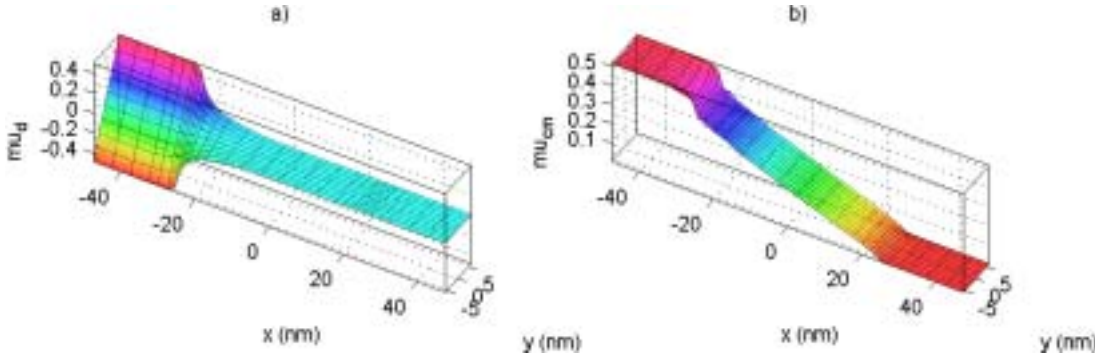

Figure 3-33. μ for g=12.5nm.

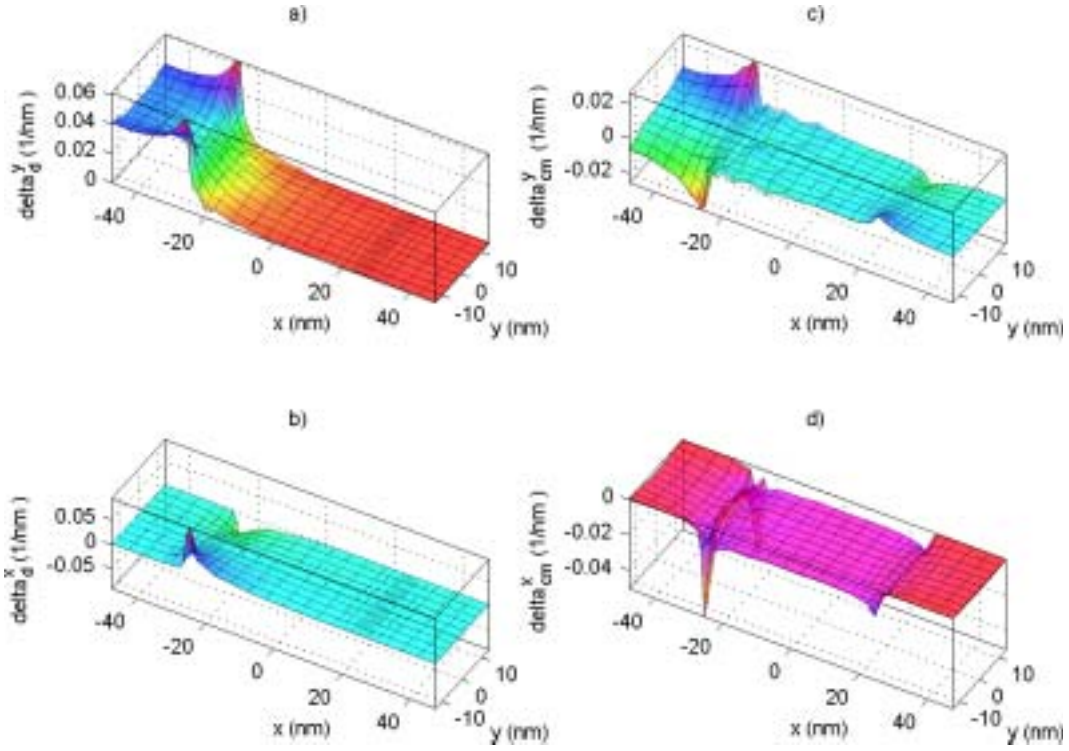

Figure 3-34. $\vec{\delta}$ for g=27.5nm.

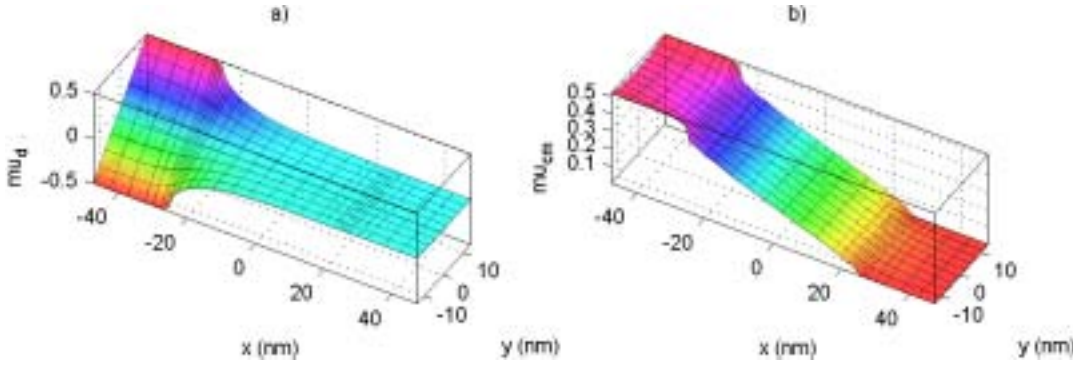

Figure 3-35. μ for g=27.5nm.

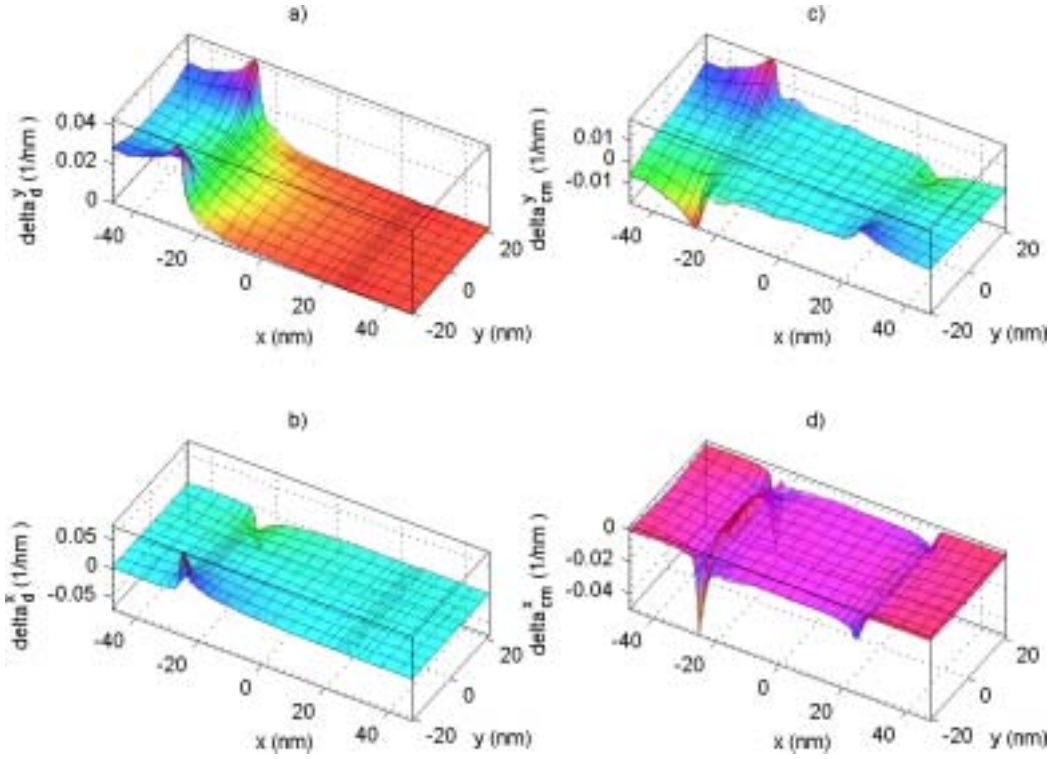

Figure 3-36. $\vec{\delta}$ for g=42.5nm.

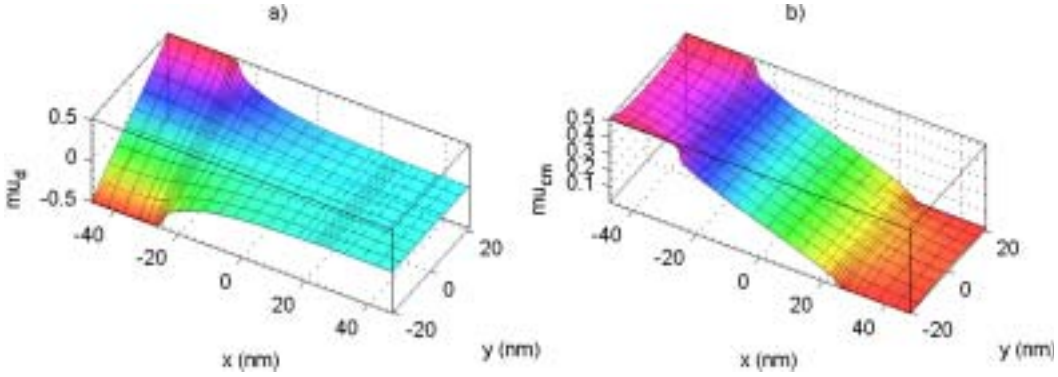

Figure 3-37. μ for g=42.5nm.

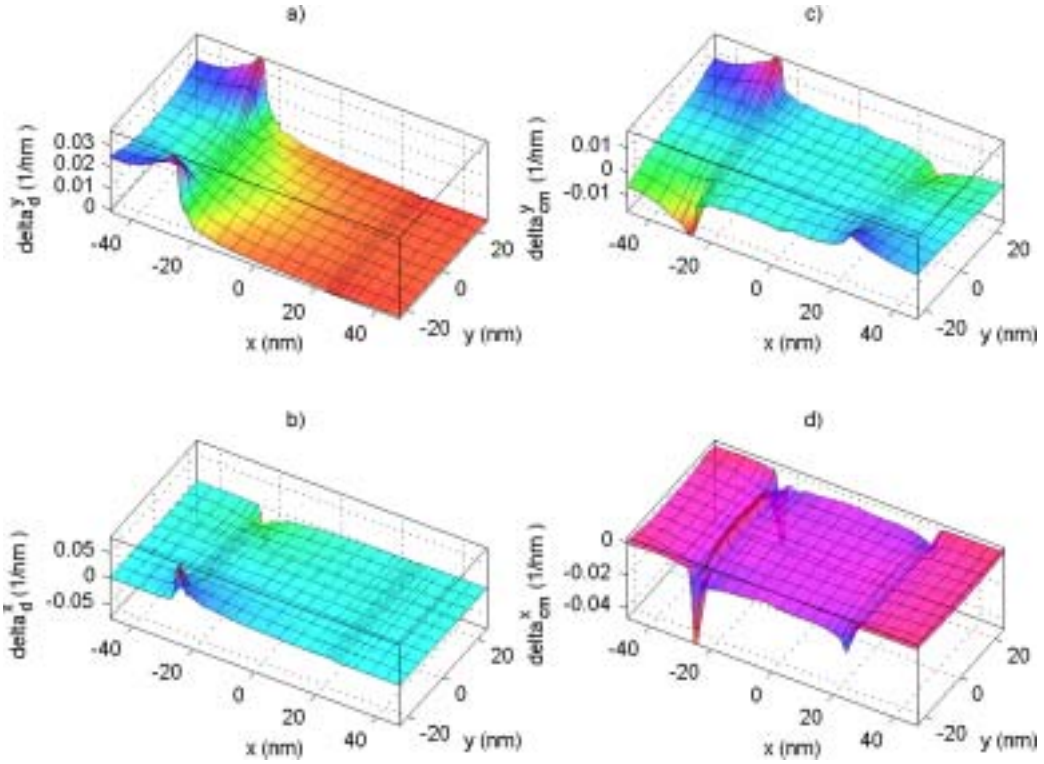

Figure 3-38. $\vec{\delta}$ for g=50nm.

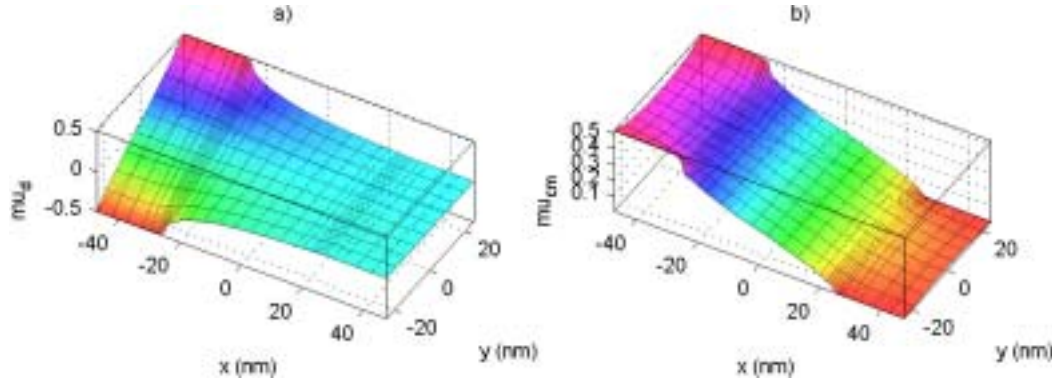

Figure 3-39. μ for g=50nm.

3.5.6 Optimum design

For the detection of DNA bases, this thesis assumes that the following design is optimum:

1. Maximally flat $\delta_d^y = 1/g^1$ in the nanogap

1. Theoretical maximum

2. Rapid falloff of δ_d^y outside the nanogap
3. Minimal δ_d^x
4. Minimal δ_{cm}^y
5. δ_{cm}^x spike at channel's edge, independent of y-position.
6. Minimal μ_{cm} and μ_d (realistically for any channel having #1 true μ_{cm}=.5 and μ_d=y_0/g).

Note that these are consistent with the aforementioned "ideal case" (Section 3.3).

This is not to say that there is no better method; there may exist a bizarre mode of DNA movement that, together with a nonlinear filter allows for ideal discrimination among the bases. However, determination of what that mode is and how to discriminate bases with it would necessitate advanced simulation technologies beyond what is currently available (discussed in detail in Section 4.2 on page 68), or equally unavailable experimental data, as well application of machine learning, all of which are beyond the scope of this thesis. Thus, this thesis characterizes nanochannels on the basis of their ability to apply a minimal-fringing electric field, which should cause a DNA dipole to align with it, and then detect the magnitude of that dipole. Given this, the justifications for each of the above requirements are therefore:

1. i) Signal with minimum dependency on position (in the nanogap)
   ii) Signal with maximum dependency on orientation
2. i) Signal with minimum dependency on position (in the nanogap, none outside it)
   ii) Minimal signal on neighboring electrodes
3. Signal with maximum dependency on orientation
4. This is only desirable as a non-zero δ_{cm}^y necessitates a non-constant δ_d^y
5. This is consistent with a square-wave type cutoff in δ_d^y; in addition, this would allow filtering of dipoles at the edge of the nanogap, where δ_d^y is dependent on position, unfortunately it is relatively uncorrelated with the falloff except for small g.

6. Minimum interference from PO_4^-

The data in this section meets these characteristics with the following dependency on *g*:

1. Magnitude if δ_d^y is approximately 1/*g*, y-flatness in the middle of the nanogap becomes worse and worse with increasing *g* due to non-linearity in Q (as discussed beginning on page 49) as well as increasing *x* due to a larger transition region (see #2).

2. The falloff (transition region) becomes less and less well defined with increasing *g*, compounding the poor linearity problem of #1 (i.e. d(*y*-non-linearity)/d*x* degrades with *g* as well as *y*-non-linearity in the middle of the nanogap) as well as making δ_d^y non-zero outside the nanogap. Any non-linearities at the nanogap's edge also generate "peaks" - these "peaks" are bigger with smaller *g* but wider with bigger *g*.

3. A high δ_d^x is a necessary evil to achieve #2. If δ_d^x was even in *y* (this is physically impossible as long as *Q* is linear in the nanogap by definition of δ) this would not be an issue as dipoles at the edge of the nanogap would have rotational independence and thus would just constitute a DC-offset in the differential signal; however, it is odd and as such will generate interference. Also see #4 below.

4. δ_{cm}^y is correlated with the "peaks" at the edge of the nanogap as mentioned in #2; however, it is odd in y. It is possible that it could be used to filter out those "peaks," but will be difficult because δ_{cm}^x will interfere. (orientation should be difficult to control at the edge of the nanogap due to the fringing field).

5. δ_{cm}^x also "peaks" at the nanogap's edge; however, these are uncorrelated to the aforementioned peaks yet unfortunately bigger than those of δ_{cm}^y for bigger *g* (where they are more problematic). It also falls off to a low level for the duration of *s*. Unfortu-

nately, as *g* increases the transition region's loss of definition is correlated with lesser $\delta_{cm}{}^{x}$, meaning that using it to filter out the transition region dipoles gets less and less feasible with increasing *g* (where it is needed more and more).

6. μ_{cm} is always linear in the nanogap from -.5 to .5 and has an odd y-falloff for x outside the nanogap that approaches 0 faster and faster with increasing *g.* In the nanogap μ_d has flatness correlated to the non-linearity in Q and has an even y-falloff for x outside the nanogap that is approximately linear to the next nanogap (and thus dependent on *s*) as a result of the charge appearing on the sides, top and bottom of the electrode. μ essentially decomposes *Q* into μ_d, a linear-in-y, non-linear in x component and μ_{cm}, a linear-in-x, non-linear-in-y components.

3.5.7 Summary

Table 3-3 on page 66 summarizes the effect of the geometrical parameters discussed in this chapter on the NLN's analysis capabilities.

As can be seen in Table 3-3, DNA induced charge is dependent on gap size whereas DNA induced voltage is not. DNA induced voltage is DNA induced charge divided by capacitance, and since both go as 1/*g*, their ratio is independent of it. However, note that since total charge is less with a bigger gap, noise is likely to be more problematic. In addition, a lower capacitance means more charge is needed to get an equivalent electric field in the nanogap, meaning a higher *applied* voltage and a higher ratio of *applied* voltage to *induced* voltage (equivalently, an equally high ratio of applied *charge* to induced *charge*)[1].

This chapter discussed the usage of finite element software to determine the effects of the geometrical design parameters on the electrostatic characteristics of the NLN, presented the results, and compared them to some analytic formulae. The next chapter discusses applying them to predict the NLN's ability to distinguish DNA bases.

1. Since detectors usually have a resolution in parts per whatever, this means worse system performance.

	dependency on g (gap)	*dependency on h* (height)	*dependency on w* (width)	*dependency on s* (spacing)	*dependency on d* (depth)
nanogap electric field (prop. to nanogap cap.)	$1/g$	h	w		
C_{1j}, j=4,5[a]	$1/g$		w		$1/d^n$*
C_{1j}, j=2,3[a]	$1/g$	h		c-$1/s^n$	$1/d^n$*
C_2[b]	g*	h	w	$1/s$	d
C_i, even i not =2[b]	g*	increases with h for small h, then decreases*	increases with w for small w, then decreases*	$1/s^n$	d
C_i, odd i not =1[b]	g	h		$1/s^n$	
monopole induced charge (μ)	$1/g$	y-linearity for x_0 in nanogap increases w/ h*	y-linearity for x_0 in nanogap increases w/ w	x-linearity for x_0 outside nanogap decreases w/ s*	
dipole induced charge (δ)	1/g	position dependency decreases with d*	position dependency decreases with w[c]		
induced voltage ($q\mu/C$, $q\delta/C$)[d]		$1/h$	$1/w$	c-$1/s^n$*	$1/d$
# of bases in nanogap			w		
retroactive electrostatic force[e]	$1/g$	accuracy of (3.12) increases w/ h*	accuracy of (3.12) increases w/ w		

Table 3-3. Summary of effects of NLN's design parameter. n is a generic undetermined exponent, possibly 1; c is a generic undetermined constant. Descriptive statements are given for relationships that were not quantitatively analyzed.
*effect unnoticable at typical values

a. this capacitance causes more charge as a result of an applied voltage yet does not contribute to the electric field in the nanogap.

b. this capacitance causes signals to appear on neighboring electrode pairs.

c. the system scales for equivalent increases in w, g, as discussed on page 54, meaning the ratio of bases in the nanogap to those outside the nanogap is unaffected, but the total *number* in the transition region (near the edge of the nanogap) still increases, thus, average position dependency is constant but total position dependent dipole induced charge increases.

d. only applicable when electrodes are driven by a current source.

e. electrostatic force the DNA as a result of DNA induced charge; this is unsupported in current molecular dynamics technology (see next chapter) and could possibly make the electrodes have a non-linear transfer function.

4 Minimum capabilities

4.1 Why "Minimum capabilities?"

The data presented in Chapter 3 is sufficient to determine the voltage induced on a pair of electrodes by an arbitrary collection of bonded and non-bonded atoms present in the nanogap between them. This relationship is half of what is needed to fully characterize the NLN's capabilities: It is dependent on their positions and charges, and therefore for a molecule like DNA, dependent on its conformation. Thus while this gives us voltage as a function of conformation, we still need the converse, or conformation as a function of voltage (both as a result of an applied voltage and as a result of the DNA itself). Ideally this would be accomplished (from a theoretical perspective; experimentation is also possible) with molecular dynamics simulations. All of the necessary data is in Chapter 3; however, accurate simulation of the NLN has several requirements not met by any currently available molecular dynamics package, and so we are limited to analyzing fixed conformations.

In this section we shall analyze only dipoles completely aligned with the NLN's electrodes. This is sufficient to defend our claim that the NLN is more discerning than existing solid-state (nanopore and nanochannel) technologies; however, the NLN's capabilities should exceed those suggested by this approximation. Complete simulation is necessary to determine a variety of anticipated non-linearities that will likely aid in detection and classification (see Section 4.3 below).

First, however, we address what is still to be done, discussing what enhancements would be necessary to existing molecular dynamics packages, a qualitative argument for base-dependent non-linearities, and a simple example of a non-linearity allowing discrimination between the bases and the backbone, as well as some operating suggestions to be considered if full molecular dynamics or experimentation is done.

4.2 Limitations of molecular dynamics

Although there are force fields[1] that have been optimized for DNA, CHARMM27 [65] [66] being the best for our purposes [67], there are no available packages that can account for

1. forces occurring as a result a charge induced on the electrodes
2. a spacially-varying external electric field
3. a temporally-varying external electric field
4. polarization

CHARMm, NAMD, and AMBER are the most popular suites for molecular dynamics. The first two support CHARMM27; AMBER supports AMBER, a force field slightly less optimized but still acceptably so [67]. CHARMm allows sinusoidally varying fields and AMBER has limited polarization abilities. None supports all of the necessary features, however (1-4 above). The relevancy of these features is this:

1. While the voltage induced per base is several microvolts, which may seem trivial next to several volts of applied voltage, recall that this is a result of a very localized (see Figures 3-22~3-26 beginning on page 52) charge density, especially so for small gap sizes and dipoles close to the electrodes' surfaces.
2. Polarization is less critical as dipole moments as a result of polarization are approximately 1% of the permanent dipole moment of DNA (assuming 50nmx20nm/5nm nanogap with 0.5V applied, a value for the minimum (see below, [50], [71]) electric field needed to overcome entropy of 10^7 V/m), but could be comparable to permanent dipole effects with voltages as low as 10 V (but realize polarization is dependent on several factors such as salt levels). Nonetheless, the fact that the bases have different

1. a "force field" in this context refers to the conformation-dependent force used to determine atomic motion.

polarizabilities suggests that it could possibly be used as an aid to discrimination (especially for guanine and cytosine which have similar permanent dipole moments). Polarization relaxation, however, is not anticipated in the gigahertz range of operation.

3. A spacially varying electric field is needed to model control mechanisms and edge effects

4. A temporally varying electric field is needed to model any frequency dependent effects at all.

For future work, NAMD is the best candidate for adding these features to as it can be freely modified. The external electric field is applied in ComputeEField.C. Adding support for arbitrary spacially varying fields would be extremely easy as the atoms' positions are available in the FullAtom class (see NamdTypes.h), and ComputeEField is passed an array of all FullAtoms in its "patch"; a "patch" is a NAMD term for a local collection of atoms (since NAMD is designed to work on clusters of machines, a machine only has a local collection of atoms to which it has access. Adding temporally varying fields would require much more extensive modifications as the current time is not available to ComputeEField. (also see ComputeExt.C). Adding support for forces occurring as a result of induced charge could be implemented by adding a term from Equation 3.10 on page 40 (the "hall-of-mirrors" approximation) to ComputeNonbondedUtil.C, where electrostatic interactions are considered. However, this is not as easy as the others due to forces crossing patch edges, several speed considerations (atom-atom forces are usually not included at every time-step and/or between every pair), etc.

4.3 Projected non-linearities

The motivation for a full molecular dynamics treatment, as aforementioned, is that it could discover non-linearities in the transfer function of the nanogap. These occur for three reasons. First, the bases rotate as a result of an applied electric field, and the applied electric field changes due to induced charge changing as a result of the bases' rotations. Second, non-linearities ought to occur as a result of mechanical coupling

via atomic bonds. These so-called mechanical non-linearities would likely depend on the types of adjacent bases: bases can rotate freely about base-sugar bonds and sugars can rotate freely about sugar-phosphate bonds (more-or-less)[1]. Thus, two bases brought in close proximity in their attempts to align with the applied field will experience a mutual electrostatic interaction. This interaction will depend on their atomic constituents, which are different from base to base, and thus should generate sequence dependent non-linearities. Thirdly, there is a non-linearity inherent in permanent dipole alignment itself: An AC applied signal causes a rotation of the bases as well as a vibration of the backbone due to electrophoresis of the PO_4^-. For a sufficiently strong, sufficiently low frequency sine wave, the dipoles will align extremely quickly, causing a square wave-like output; the monopoles, however, will continue to move with the applied field, causing a sine wave-like output. This type of non-linear behavior should be helpful in separating out the PO_4^-'s signal and determining the strand's position in the nanogap. This is cartooned in Figure 4-1 on page 71

4.4 Operating considerations

4.4.1 Stretching

Stretching the DNA by a small amount will probably benefit overall performance for two reasons: First of all, it increases the base-to-base spacing, meaning less bases in the nanogap. Secondly, it eliminates the DNA's ability to kink to relieve mechanical stress occurring as its bases attempt to align with the field. By kinking it limits mechanical non-linearities, as well as pulls more bases into the nanogap. Note that excessive stretching will not be beneficial as it will limit the rotation of the sugars about the sugar-phosphate bonds (inhibiting alignment), while excessively increasing the base-to-base spacing will lower the electrostatic (primary) contribution to mechanical non-linearity.

1. This rotation actually causes a small amount of kinking due to rigidity of the backbone's bond angles. This is small, however: note that the contour length difference between un-stretched (lots of kinking) and stretched DNA is worst case ~.7nm (see 4.5.2.1); also note the small ratio of Kuhn length to contour length (see 4.4.2).

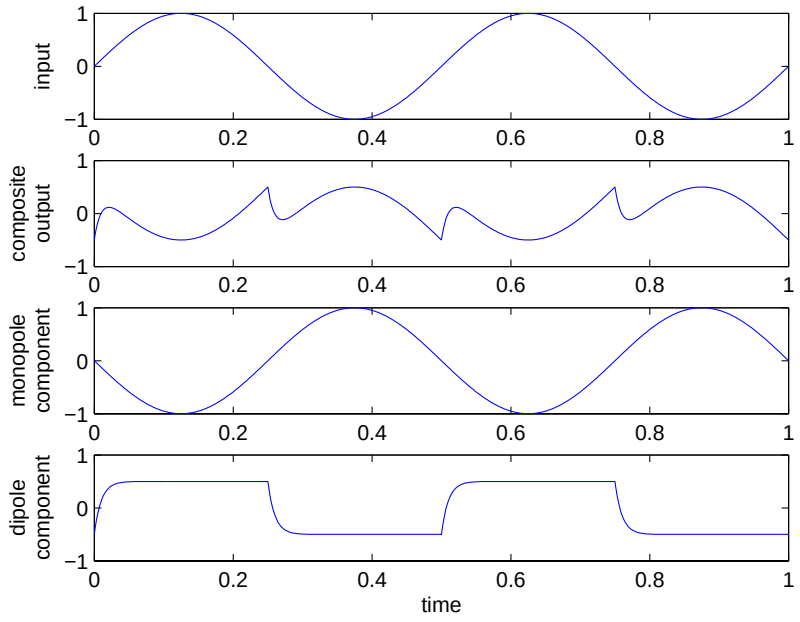

Figure 4-1. Cartoon of separation of monopole and dipole components of the output signal (no drag); if there is drag, the system still works as the monopole component is still linear in the input whereas the dipole component is still not: the dipole component can be **always** be reconstructed by notch filtering the output at the input frequency (varying the filter strength until the result takes the correct form; i.e. asymptotic to a constant). Note that if there are mechanical non-linearities, the dipole component would probably look more "underdamped."

4.4.2 Solvent

For a detailed discussion of the solvent, see Section 3.1.1 on page 20.

4.5 Fixed conformation analysis

We can still analyze the NLN with DNA of a fixed conformation. We choose a conformation where all the bases' dipoles are completely normal to the electrodes' faces. This is a "DC" solution in the sense that it is the conformation the DNA will tend to if left in a normal applied field for a long time. This serves as a lower bound on the resolution actually obtainable since it ignores separation in the frequency domain. Note that resonances, per se, will never be obtainable as the most charge a base's dipole can induce is $\left|\vec{p}\right|\vec{\delta}_d \bullet \hat{y}$, which is independent of frequency.

4.5.1 Assumptions

We assume that the DNA is sufficiently stretched that no kinking occurs (this could even be accomplished by electrophoretic drag from an axially applied field), that there is a negligible concentration of salt present, and that the applied field is sufficiently strong and sufficiently low frequency that all of the bases' dipoles are at a 90 degree angle with the electrodes' surfaces as well as δ separable from μ (see Section 4.3).

4.5.2 Relevant data

4.5.2.1 Contour length per base

Several different values for the contour length per base of ssDNA have been published: .59nm by S. Smith [68], .56nm by S. Cocco [69], .539nm by S. Koch [70] and .63nm by M. Murphy [72]. With a high enough force the contour length per base has been witnessed as large as .7nm [68], but herein we assume enough force only to limit kinking, and therefore use the most recent unstretched data, .63nm (a conservative approximation).

4.5.2.2 Dipole moments

As aforementioned, the bases' dipole moments have been determined by a variety of methods, experimental and ab-initio. The experimental values were determined indirectly, and as such thought to be inaccurate [50]. This thesis uses M. Preuss' values determined by density functional theory using ultrasoft psuedopotentials [51]; these are the most recent available and in good agreement with quantum-chemical

results. CHARMM27 uses similar ab-initio data for its representation of dipole moments (see [65] supplement). M. Preuss' data is shown visually in Figure 4-2.

The first thing to notice about this data is that guanine and cytosine's dipoles are very close in magnitude and so some other effect will need to play a part for them to be discernible. Polarization is promising as adenine and guanine have polarizability α=~14 A^3 whereas cytosine and thymine have α=~11 A^3 [50]. (note these vary with a variety of conditions, the most prominent of which is probably going to be salt concentration in our setup; more recent experimental data suggests even bigger differences, albeit only visible in the terahertz regime [48]). Mechanical non-linearities may provide an alternate discrimination method.

4.5.2.3 Applied field

The applied electric field needed to overcome thermal effects is on the order of 10^7 V/m [71]. Because the electric field in the nanogap is very close to ideal, this translates into a surface charge density

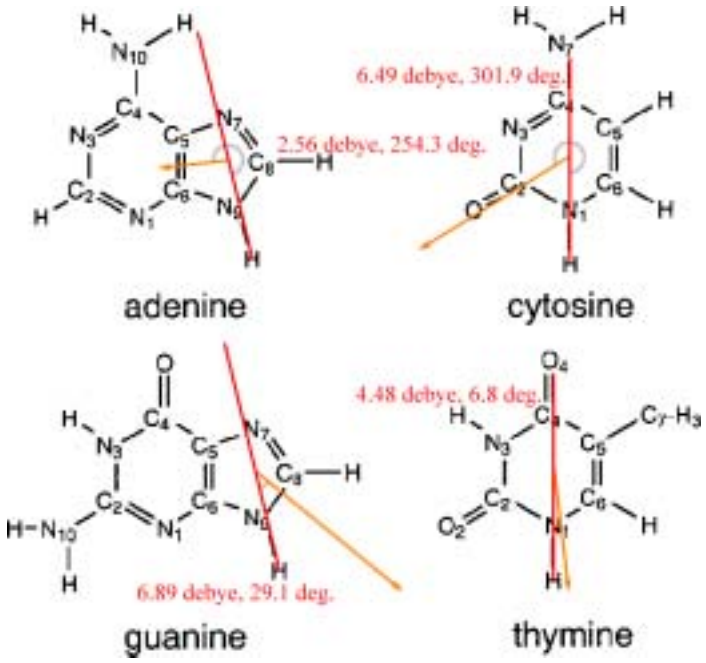

Figure 4-2. Data from ref. 51 (dipole moments to scale). In addition to the angles shown:
cytosine bends out of the page by +1.9 degrees
guanine bends out of the page by +1.3 degrees
thymine bends out of the page by +0.3 degrees
(adenine does not bend out of the page at all)
Figure modified from [51].

of $\sigma=\varepsilon E=$ ~.042e/nm^2= ~0.05V for a w=50nm, h=20nm, g=5nm electrode pair; a larger voltage may be needed for alignment with the electric field to overcome dipole-dipole interactions between the bases.

4.5.3 Ideal induced voltage

The following table lists the voltage induced as a result of a given dipole in ideal alignment (90^0) with the faces of an ideal (completely flat δ_d, no parasitic capacitance) nanogap with wxh=50x20nm^2 ($V=\left|\vec{p}\right|\vec{\delta}_d \bullet \hat{y}/C_{11}$ where $\vec{\delta}_d \bullet \hat{y} = 1/g$ and $C_{11} = \varepsilon_{water} wh/g$; the end result is independent of g (also see page 65).

Although voltage is independent of g, a large w/g is still needed to maintain a flat $\vec{\delta}_d$ and a high capacitance. Not maintaining a flat[1] $\vec{\delta}_d$ will invalidate $\vec{\delta}_d \bullet \hat{y} = 1/g$, whereas failing to maintain a high capacitance raises noise as a result of there being less total electrons on the electrodes for a given voltage (as previously mentioned on page 65).

4.5.4 The effect of non-ideal $\vec{\delta}_d$

The data in Table 4-1 suggests that, in a strand of two different types of bases of known type, the quantity of one of the bases versus the other can be determined with single-base precision (assuming a sufficient system resolution, for example 32-12=20μV for guanine versus adenine). However, as stated, this assumes a large ratio of w to g and bases near the middle of the nanogap. In reality, if w and g are comparable non-idealities in $\vec{\delta}_d$ severely limit the resolution of the channel.

base	***differential voltage (μV), h=20nm, g<<w***		
	w=50nm	*w=27.5nm*	*w=5nm*
adenine	12	22	120
cytosine	31	56	310
guanine	32	59	320
thymine	21	38	210

Table 4-1. Voltage induced as a result of a given dipole in ideal alignment with the faces of an ideal nanogap.

1. The error in Table 4-1 becomes uncorrelated with position (i.e. very near to the noise floor of FEMLAB) for dipoles in the middle 75% (in the x-direction) of the nanogap (meaning $\vec{\delta}_d$ is approximately flat in this area) for w=50nm, g=5nm, and therefore for **all** w/g=10, assuming $\vec{\delta}_d$ scales with a constant ratio of w to g ($\vec{\delta}_d$ is linear in Q, and therefore should scale in the same way; see page 54).

In the following analysis we will make the assumption that all the bases of a given type are grouped together ($A_{30}C_{70}$, not $A_{15}C_{70}A_{15}$), as well as that there are two types of bases in a given strand (assumptions similar to Akeson's experiment [10]). Table 4-2 shows the possible base compositions corresponding to a given δ_d^y differential voltage across an electrode pair for a w=50, h=20, g=5 nanogap under these conditions (note that a w=50 nanogap has 80 bases present at a time).

Table 4-2 can be interpreted as follows: If one were to read a δ_d^y induced differential voltage of 1.8mV across an electrode pair, it could occur as a result of any of the following:

1. an adenine-guanine strand with 33~43 guanines (the rest adenines)
2. an adenine-cytosine strand with 36~47 cytosines
3. an adenine-thymine strand with 76~79 thymines
4. a thymine-guanine strand with 1~15 guanines
5. a thymine-cytosine strand with 1~17 cytosines
6. a strand of all thymine

- this voltage is too low to correspond to a strand with only guanine and/or cytosine

voltage (mV)	# guanine in adenine	# cytosine in adenine	# thymine in adenine	#guanine in thymine	#cytosine in thymine	#guanine in cytosine
.89	0	0	0	X	X	X
1.1	3~12	4~13	8~24	X	X	X
1.3	12~22	14~24	30~47	X	X	X
1.5	23~33	25~36	54~71	X	X	X
1.8	33~43	36~47	76~80	0~15	0~17	X
2.0	43~53	48~59	X	13~34	15~40	X
2.2	53~64	59~70	X	31~52	38~62	X
2.4	64~75	70~80	X	49~71	60~80	0~36
2.6	74~80	X	X	68~80	X	1~80
2.8	80	X	X	80	X	80

Table 4-2. Possible base compositions corresponding to a given δ_d^y induced differential voltage across an electrode pair. X=not possible to obtain this voltage with a strand of these constituents. y-movement of up to y=g-0.25nm is considered.

The explanation for the lowered resolution is simple: Take a strand of 64 thymines and 16 cytosines, for example. If the voltage is initially 1.8mV and then increases to 2.0mV, this could have occurred as a result of *y*- or *x*- movement of the strand:

- *y*- movement would have increased the differential voltage caused by the cytosine and thymine already present as a result of *y*-dependency in δ_d^y
- *x*- movement would have increased the differential voltage caused by existing cytosine, decreased that cause by thymine (due to *x*-dependency in δ_d^y) **and** brought additional cytosine into the nanogap

The ambiguity in the relative numbers of the two nucleotides present occurs because it is impossible (without time or frequency domain aids) to separate out which of these factors (or a combination of the two) is responsible for a change in voltage.

We define a resolution r_{mn} equal to the biggest range over all voltages for a strand consisting of bases of types *m* and *n*. This is a good characterization of the NLN in that experimentally, given a strand known to consist of bases *m* and *n*, if one experimentally measures a voltage *V*, the relative number of bases present is then the average number corresponding to *V*, $\pm\lceil (r_{mn} - 1)/2 \rceil$ (as a non-integer number of bases is impossible). If a technique was developed to filter out the voltage as a result of data near the x- and y-edges of the nanogap (even one as simple as comparison with δ_{cm}^y, which is correlated with positions at the edges of the nanogap), the resolutions would dramatically improve. Decreasing *w*/*g* (or decreasing the width for a given gap size), on the other hand, causes a dramatic degradation. Table 4-3 shows the result of selected variations on the resolution. The data is presented with varying *g* since varying *w* affects the number of bases in the nanogap (*w*=50nm, *h*=20nm in the data below)

	r	r_{ga}		r_{ca}		r_{ta}		r_{gt}		r_{ct}		r_{cg}	
g	*filters*	*all x*	*75% x*[a]	*all x*	*75% x*	*all x*	*75% x*	*all x*	*75% x*	*all x*	*75% x*	*all x*	*75% x*
5 nm	*all y.*	12	2[b]	13	2[b]	19	2[b]	23	3[b]	26	4[b]	80[c]	21[b]
	80% y[d]	9	2[b]	9	2[b]	14	2[b]	17	3[b]	19	4[b]	80[c]	19[b]
12.5 nm	*all y*	31	12	31	13	45	18	58	23	62	26	80[c]	60[c]
	80% y	21	8	21	9	30	12	39	16	43	18	80[c]	60[c]
27.5 nm	*all y*	53	28	53	30	80[c]	42	80[c]	60[c]	80[c]	60[c]	80[c]	60[c]
	80% y	37	20	37	22	48	30	67	37	80[c]	45	80[c]	60[c]
42.5 nm	*all y*	62	37	61	42	80[c]	60[c]	80[c]	60[c]	80[c]	60[c]	80[c]	60[c]
	80% y	47	28	47	31	66	39	80[c]	60[c]	80[c]	60[c]	80[c]	60[c]
50 nm	*all y*	59	42	66	47	80[c]	60[c]	80[c]	60[c]	80[c]	60[c]	80[c]	60[c]
	80% y	46	32	52	36	75	60[c]	80[c]	60[c]	80[c]	60[c]	80[c]	60[c]

Table 4-3. Effect of x- and y- position-based filtering and gap size on resolution. This does not account for increases in noise that will probably accompany smaller w/g (see paragraph beneath Table 4-1 on page 74).

a. resolution if only data in the middle 75% (in the x-direction) of the nanogap is considered; note that these are out of 60 bases, not 80

b. FEMLAB error on the order of one base (see paragraph beneath Table 4-1 on page 74)

c. these two bases are indistinguishable in a strand consisting of the two of them

d. resolution if only data in the middle 80% (in the y-direction) of the nanogap is considered

Note that the resolution would be slightly worse if the bases are randomly distributed in a strand if δ_d^y has a large x-dependency (i.e. w not $>>g$) as the bases of one type could be more heavily concentrated in areas of high δ_d^y.

4.6 Comparison to existing technologies and conclusion

Many technologies have been developed for sequencing DNA over the years, however, nobody has yet managed to sequence a single strand of DNA without making numerous copies, a process requiring large equipment and expensive chemical reactions. α-hemolysin nanopores provided hope for such an accomplishment, but suffered from biological limitations. Silicon nanopores resolved this problem but still suffered from conformational inconsistencies that severely limited resolution. Nanochannels provided a way to maintain conformational consistency but suffered from an inability to do localized measurements, limiting resolution in a different way. Nanoelectrode lined nanochannels provide a “best of both worlds” compromise between nanopores and nanochannels.

As can be seen in Table 4-3, an NLN with cross-sectional area of 20nmx5nm=100nm^2, an area that has been physically realized in a nanochannel [21], electrodes an order of magnitude bigger than the channel width, and extremely basic signal processing techniques one could achieve a resolution orders of magnitude better than the best that has been previously accomplished [10] (with the exception of DNA hairpin-based sequencing, but that technique is inherently more difficult than traditional sequencing). This resolution could be further improved by applying time-domain and frequency-domain signal processing as well as with smaller channels. In addition, the design naturally lends itself to fully integrated system-on-a-chip designs, which should have exceptional noise characteristics compared to existing technologies (which suffer from the ancient technology of patch-clamp amplifiers and erratic ion flow), while being much smaller.

4.7 Future work

The two most important things that need to be done are both theoretical in nature: a) an analysis of interconnect so as to determine if it is possible to even get a sufficient signal to or from the nanogap and b) a molecular dynamics simulation of the DNA in the nanogap as discussed at the beginning of this Chapter. The effect of Tris-Cl and any other ions in the elution buffer should be included.

After b) above is accomplished, signal processing specialists can determine what signals to apply to the input as well as how to filter the output. Machine learning would probably be of considerable utility here, given its past successes in classifying DNA [13], [14].

Also of immediate interest would be a nanochannel with only one electrode (two additional electrodes in the reservoirs on either side of the nanochannel could be used for stretching), a gap of approximately 5nm, a height of approximately 20nm (probably ±50% would be okay), and a reasonable depth (there are no adjacent electrodes, so height is somewhat irrelevant). If this “NLN” could discriminate between an ssDNA molecule of all adenines and an ssDNA molecule of all guanines, it would serve as proof-of-principle.

Finally, it should be noted that we have provided data sufficient to predict the NLN's ability to characterize other (non-DNA) biological molecules: the NLN provides a wealth of opportunities for electrically sensing arbitrary charged molecules.

Bibliography

[1] http://departments.oxy.edu/biology/Franck/Bio130S_2002/Images/Ch16/fig16_3.JPG

[2] http://www.chm.bris.ac.uk/webprojects2002/bdavies/structure.gif

[3] http://www.uweb.ucsb.edu/~trevorc/histseq.html

[4] http://www.genomenewsnetwork.org/timeline/timeline_home.shtml

[5] http://www.chm.bris.ac.uk/webprojects2002/bdavies/dna1.jpg

[6] Kasianowicz JJ, Brandin E, Branton D, Deamer DW. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13770-3

[7] Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE. (1996) *Science* **274**, 1859-66.

[8] Deamer DW, Branton D. (2002) *Acc Chem Res.* **35**, 817-25.

[9] http://www.mcb.harvard.edu/branton/Blockade&SeqApplic.jpg

[10] Akeson M, Branton D, Kasianowicz JJ, Brandin E, Deamer DW. (1999) *Biophys. J.* **77**, 3227-33.

[11] Meller A, Nivon L, Brandin E, Golovchenko J, Branton D. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 1079-84.

[12] Vercoutere W, Winters-Hilt S, Olsen H, Deamer D, Haussler D, Akeson M. (2001) *Nature Biotech.* **19**, 248-52.

[13] Winters-Hilt S, Vercoutere W, DeGuzman VS, Deamer D, Akeson M, Haussler D. (2003) *Biophys J.* **84**, 967-76.

[14] Vercoutere WA, Winters-Hilt S, DeGuzman VS, Deamer D, Ridino SE, Rodgers JT, Olsen HE, Marziali A, Akeson M. (2003) *Nucleic Acids Res.* **31**, 1311-8.

[15] Timp G, Schulten K, Leburton JP, Bezryadin A. “Nanometer-scale Gene Chip,” *NSF Nanoscale Sci. and Engr. Grantees Conference*, Dec. 16-18, 2003.

[16] Bates M, Burns M, Meller A. (2003) *Biophys. J.* **84**, 2366-72.

[17] Li J, Gershow M, Stein D, Brandin E, Golovchenko JA. (2003) *Nature Materials* **2**, 611-615.

[18] Storm AJ, Chen JH, Ling XS, Zandbergen HW, Dekker C. (2003) *Nature Materials* **2**, 537-540.

[19] Heng JB, Dimitrov V, Grinkova YV, Ho C, Kim T, Muller D, Sligar S, Sorsch T, Twesten R, Timp R, Timp G. *Electron Devices Meeting, IEDM '03 Technical Digest*, Dec. 2-3, 2003, 8-10.

[20] Cao H, Tegenfeldt JO, Austin RH, Chou SY (2002). *Appl. Phys. Lett.* **82**, 3058-60.

[21] Austin RH (2003). *Nature Materials* **2**, 567-8.

[22] Cabodi M, Turner SWP, Craighead HG. (2002) *Anal. Chem*. **74**, 5169-74.

[23] Guo LJ, Cheng X, Chou C (2004) *Nano Lett.* **4**, 69-73.

[24] Tegenfeldt J, Prinz C, Cao H, Chou S, Reisner W, Riehn R, Wang YM, Cox EC, Sturm JC, Silberzan P, Austin RH (2004) *PNAS* **101**, 10979-83.

[25] Edwards GS, Davis CC, Saffer JD, Swicord ML. (1984) *Phys. Rev. Lett.* **53**, 1284-7.

[26] Edwards GS, Davis CC, Saffer JD, Swicord ML. (1985) *Biophys. J.* **47**, 799-807.

[27] Foster KR, Epstein BR, Gealt MA. (1987) *Biophys. J.* **52**, 412-5.

[28] Bigio IJ, Gosnell TR, Mukherjee P, Saffer JD. (1993) *Biopolymers* **33**, 147-50.

[29] Swicord M, Davis C. (1982) *Biopolymers* **21**, 2453-60.

[30] Gabriel C, Grant EH, Tata R, Brown PR, Gestblom B, Noreland E. (1987) *Nature* **328**, 145-6.

[31] Gabriel C, Grant EH, Tata R, Brown PR, Gestblom B, Noreland E. (1989) *Biophys J.* **55**, 29-34.

[32] Garner HR, Ohkawa T, Tuason O, Lee RL. (1990) *Phys. Rev. A* **42**, 7264-70.

[33] Van Zandt LL, Davis ME. (1986) *JBSD* **3**, 1045-53.

[34] Van Zandt LL. (1987) *JBSD* **4**, 569-82.

[35] Van Zandt LL, Saxena VK. (1989) *Phys. Rev. A* **39**, 2672-4.

[36] Wittlin A, Genzel L, Kremer F, Haseler S, Poglitsch A, Rupprecht A. (1986) *Phys. Rev. A* **34**, 493-500.

[37] Powell JW, Edwards GS, Genzel L, Kremer F, Wittlin A, Kubasek W, Peticolas W. (1987) *Phys. Rev. A* **35**, 3929-39.

[38] Woolard DL, Globus TR, Gelmont BL, Bykhovskaia M, Samuels AC, Cookmeyer D, Hesler JL, Crowe TW, Jensen JO, Jensen JL, Loerop WR. (2002) *Phys. Rev. E* **65**, 051903-1-11.

[39] Fischer BM, Walther M, Jepsen PU. (2002) *Phys. Med. Biol.* **47**, 3807-14.

[40] Young L, Prabhu VV, Prohofsky EW. (1989) *Phys. Rev. A* **39**, 3173-80.

[41] Van Zandt LL, Saxena VK. (1989) *Phys. Rev. A* **39** 2672-4.

[42] Saxena VK, Van Zandt LL. (1990) *Phys. Rev. A* **42** 4993-7.

[43] Saxena VK, Dorfman BH, Van Zandt LL. (1991) *Phys. Rev. A* **43**, 4510-6.

[44] Saxena VK, Van Zandt LL. (1994) *JBSD* **11**, 1149-59.

[45] Van Zandt LL, Saxena VK. (1994) *Biophys. J.* **67**, 2448-53.

[46] Saxena VK, Van Zandt LL. (1992) *Phys. Rev. A* **45**, 7610-20.

[47] Van Zandt LL, Saxena VK. (1992) *JBSD* **10**, 227-37.

[48] Nagel M, Bolivar PH, Brucherseifer M, Kurz H, Bosserhoff A, Buttner, R (2002) *Appl. Phys. Lett.* **80**, 154-56.

[49] Hasted JB, Aqueous Dielectrics. Chapman and Hall: London, 1973.

[50] DeVoe H, Tinoco I Jr. (1962) *J. Mol. Biol.* **4**, 500-17.

[51] Preuss M, Schmidt WG, Seino K, Furthmuller J, Bechstedt F. (2004) *J. Comput. Chem.* **25**, 112-122.

[52] Smythe W, Static and Dynamic Electricity (3rd Edition). McGraw-Hill, 1968.

[53] Gilson MK, Rashin A, Fine R, Honig B (1985). *J. Mol. Biol.* **184**, 503-16.

[54] Hendsch ZS, Tidor B (1994). *Protein Sci*.**3**, 211-26.

[55] Honig B, Nicholls A (1995). *Science* **268**, 1144-49

[56] Vorobjev YN, Almagro, JC, and Hermans, J (1998). *Proteins* **32**, 399-413

[57] Norel H, Sheinerman F, Petrey D, Honig B (2001). *Protein Sci.***10**, 2147-61.

[58] Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001). *PNAS USA* **98**, 10037-41.

[59] http://www.lenntech.com/deionised-demineralised-water.htm

[60] Palmer HB (1937). *Trans. AIEE.* **56**, 363-6.

[61] Yang H (2000). "Microgyroscope and microdynamics," *Ph. D. Dissertation* [see M. H. Bao, Micro Mechanical Transducers. Elsevier Science, 2000.]

[62] Elliott RS, Electromagnetics. McGraw-Hill, 1966.

[63] Leus V, Elata D (2004). Tech. Report ETR-2004-2, Israel Institute of Technology.

[64] Private communication with Tamal Mukherjee, Research Scientist at Carnegie Mellon.

[65] Foloppe N, Mackerell A (1999). *J. Comput. Chem.* **21**, 86-104.

[66] Mackerell A, Banavali N (2000). *J. Comput. Chem.* **21**, 105-120.

[67] Reddy S, Leclerc F, Karplus M (2003). *Biophys. J.* **84**, 1421-49.

[68] Smith SB, Cui Y, Bustamente C (1996). *Science* **271**, 795.799.

[69] Cocco, S, Monasson R, Marko JF (2001). *PNAS* **98**, 8608-13.

[70] Koch SJ, Shundrovsky A, Jantzen BC, Wang MD (2002). *Biophys. J.* **83**, 1098-105.

[71] Dewarrat, F. (2002). "Electric Characterization of DNA," *Ph. D. Thesis*

[72] Murphy MC, Rasnik I, Cheng W, Lohman T, Ha T (2004). *Biophys. J.* **86**, 2530-7.