Semiconductor design methodologies for epigenetic monitoring

by

Dora Ma

September 2017

Supervised by Regius Professor Christofer Toumazou

A thesis submitted for the degree of Doctor of Philosophy of Imperial College London

Department of Electrical and Electronic Engineering
Imperial College of Science, Technology and Medicine
Declaration of Originality

I hereby declare that this thesis and the work described is my own and that everything else is appropriately referenced and attributed.
Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives license. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the license terms of this work.
Abstract

Epigenetics governs environmental influences on gene expressions, and plays a crucial role in chronic disease progressions. The binary nature of methylation and microRNAs (miRNA), defined by their “on” and “off” states or “up” and “down” regulations respectively, makes them accessible as potential biomarkers and therapeutics. Integration of such epigenetic modifications as part of future healthcare depends on the development of appropriate technology for monitoring. Therefore the thesis envisions an ISFET (ion-sensitive field effect transistor) based epigenetic microchip, to enable portability in methylation detection and miRNA quantification.

Four circuits were designed and implemented to carry out or improve the accuracy of suitable analysis methods that are compatible with ISFETs. For methylation-specific PCR (MSP), a novel configuration of ISFETs as a current mirror was proposed for differential sensing of opposite strand primers. With the circuit operating in weak inversion, drift and temperature sensitivities were reduced while achieving sufficient pH resolution to detect nominal changes expected in MSP reactions. For single-base resolution, a novel ISFET switched current integrator fabricated in unmodified CMOS was proposed to overcome one of the primary limitations in ion-semiconductor sequencing, namely base calling for repeated nucleotides known as homopolymers. The application of integration can potentially increase both accuracy and window for detection of base incorporations. Additionally, the feasibility of a new compressed work flow unique to ISFET-based sequencing for methylation is investigated. For evaluating miRNA expressions in relative quantification analysis, the first implementation of an on-chip back-end using the Derivative method to compute threshold cycles is presented. The process also led to the development of a new algorithm that compares three data points to provide real-time assessment of positive amplifications based on exponential characteristics, enabling detection times to be decreased whilst providing similar estimates of threshold cycles to the established Derivative method.

Overall work express potential of implementing epigenetics detection using ISFETs, with CMOS microchips presented as a fast and affordable solution for future monitoring.
Dedicated to my parents, my sister
Acknowledgments

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Toumazou. He has been an inspiration ever since I attended his lectures during my undergraduate studies, and has continued to do so with his ability to innovate and cultivate research in the real world. Not only has he given me countless opportunities academically, whether it be my first taste in research as a summer UROP student, or the chance to pursue a PhD, he has always actively involved me in industries and many other experiences. Without him I would not be where I am today, and I am forever grateful.

I want to give special thanks to Dr. Pantelis Georgiou for his valuable discussions and advise, often beyond his obligations. The kindness and continuous support he offers to everyone and to myself is exceptional and aspiring. Together with Dr. Timothy Constantdinou, they have been prominent role models and taught me how to be a better researcher.

I would like to thank Ms. Gifty Kugbleenu, Ms. Izabela Wojcicka-Grzesiak and Ms. Wiesia Hsissen, who have been so caring and supportive throughout the whole process. Their help have been incremental to ensuring all aspects of my studies went smoothly. As a group, we are extremely appreciative of their hard work on a day-to-day basis.

I have been so fortunate to work amongst not only talented but many wonderful colleagues and friends. In particular, I want to show my deep appreciations to Dr. Melina Kalofonou and Dr. Mohammadreza Solhbat who have been my mentors academically as well as personal cheer leaders. Their kindness are infectious and have motivated me in the most difficult times. Part of the thesis could not have been completed without Dr Jesus Rodriguez-Manzano, who kindly provided real experimental data of fluorescent amplification reactions. Special mentions must also go to my dear friend Nicolas Moser, who is always willing to extend help and put others’ needs before himself. My gratitude goes towards all my friends from the Centre of bio-inspired technology, Nicoletta, Pete, Pau, Nick Miscourides, Ermis, Dorian, Song, Yan, Lieuve, Sara, Bernard... the list goes beyond that could be mentioned here.
I would like to thank all the staff who have contributed towards my learning during my degree and thesis. Ms Maria Khaleeq, who not only bonded microchips for us, but kindly passed on her knowledge before she retired. I am grateful for the opportunity to be a teaching assistant for Ms. Esther Perea, and Dr. Kristel Fobelets for her support and feedback at the early stage assessment. In addition, I am thankful to the Department of Electrical and Electronic Engineering for sponsoring my PhD studies through scholarship, and Imperial College for the nurture and resources over the years.

Above all, none of this would be possible without my loving family. My parents who have provided me with an environment and confidence to freely pursue an academic career. And my sister, who taught me the importance of critical thinking. From her actions and presence, she has been, and will always be an inspiration to me.
# Contents

1 Introduction .................................................. 26  
1.1 Motivation .................................................. 26  
1.2 Aim & Vision ................................................. 31  
1.3 Overview ..................................................... 32  
1.3.1 The role of Epigenetics in chronic diseases ............ 32  
1.3.2 Ion-semiconductors for DNA analysis .................. 32  
1.3.3 Analogue current mode design approach for methylation 32  
1.3.4 On-chip threshold cycle evaluation and Early amplification 33  
1.3.5 Conclusion and Future work ............................. 33  

2 The role of Epigenetics in chronic disease ............... 37  
2.1 Epigenetics .................................................. 38  
2.1.1 DNA methylation ....................................... 39  
2.1.2 miRNA Regulation ...................................... 41  
2.2 Chronic Kidney Disease ................................... 42  
2.2.1 Glomerular Filtration Rate (GFR) ...................... 43  
2.2.2 Fibrosis ................................................... 44  
2.2.3 Epigenetics in CKD ..................................... 46  
2.3 Methylation Analysis ....................................... 51  
2.3.1 Preparation methods ................................... 51  
2.3.2 Post bisulfite analyses ................................ 53  
2.3.3 Bisulphite free methods ............................... 59  
2.4 Quantifying miRNAs ....................................... 60  
2.4.1 Post Reverse transcription ............................. 62  
2.4.2 Absolute, Relative and Digital quantification ........ 65  
2.5 Summary ..................................................... 72
3 Ion-semiconductors for DNA analysis

3.1 Label free electronic DNA detection .................................................. 92
  3.1.1 Impedance and Dielectric Spectroscopy ........................................ 93
  3.1.2 DNA charge measurements ........................................................... 94
  3.1.3 RedOx-based .............................................................................. 96
  3.1.4 Pyrophosphate-based ..................................................................... 99
  3.1.5 pH-based ..................................................................................... 100

3.2 ISFET sensor ....................................................................................... 102
  3.2.1 Operation ...................................................................................... 103
  3.2.2 pH sensitivity ............................................................................... 104
  3.2.3 ISFET modeling ........................................................................... 107

3.3 ISFET design in unmodified CMOS .................................................... 108
  3.3.1 Sensor area ................................................................................. 109
  3.3.2 DC offset ..................................................................................... 110
  3.3.3 Measurement Times ..................................................................... 111
  3.3.4 Temperature and Noise ................................................................. 113

3.4 ISFET-based epigenetic detection ....................................................... 114
  3.4.1 Sequencing by ISFETs ................................................................. 114
  3.4.2 Application-specific IC ................................................................. 116
  3.4.3 Real-time amplification ................................................................. 117

3.5 Summary ............................................................................................. 120

4 Analogue current mode design approach for methylation .................. 131

4.1 MSP-based Circuit Design ................................................................. 133
  4.1.1 Differential ISFET monitoring ....................................................... 133
  4.1.2 ISFET weak inversion current mirror ........................................... 136
  4.1.3 Conclusion .................................................................................... 148

4.2 Signal processing for Bisulfite-Sequencing ....................................... 149
  4.2.1 Current mode integrators ............................................................. 150
  4.2.2 ISFET based Switched Current integrator ..................................... 153
  4.2.3 Fabrication and Testing ............................................................... 157
  4.2.4 Frequency Response ................................................................. 158
  4.2.5 Base calling performance ............................................................ 160
  4.2.6 Conclusion ................................................................................. 172

4.3 Compressed targeted DNA methylation sequencing .......................... 174
  4.3.1 Compressed incorporations ......................................................... 174
  4.3.2 Simulations ............................................................................... 175
4.3.3 Conclusion ........................................ 182
4.4 Summary ............................................. 184

5 On-chip threshold cycle evaluation and Early amplification 189
5.1 System for determining the threshold cycle, $C_T$ .......................... 191
  5.1.1 Fit point and Derivative method comparisons ................ 191
  5.1.2 VHDL Implementation .................................. 193
  5.1.3 Fabricated system ...................................... 195
  5.1.4 Measured Results ....................................... 198
  5.1.5 ISFET as front-end ..................................... 213
  5.1.6 Conclusion ............................................. 218
5.2 Early amplification detection .......................................... 220
  5.2.1 3-point exponential evaluation system ...................... 220
  5.2.2 Simulations ............................................. 221
  5.2.3 Measured Results ....................................... 222
  5.2.4 Conclusion ............................................. 228
5.3 Summary ................................................. 229

6 Conclusion and Future Work 232
6.1 Summary and Contributions ........................................ 233
  6.1.1 The role of epigenetics in chronic disease .................. 233
  6.1.2 Ion semiconductors in DNA detection ....................... 233
  6.1.3 Analogue current mode design approach for methylation .... 234
  6.1.4 On-chip threshold cycle evaluation and Early amplification 236
6.2 Future work ................................................ 236
  6.2.1 Development of differential micro-fluidics .................. 237
  6.2.2 Computational readout for quantitative MSP current mirror 237
  6.2.3 ISFET shapes ........................................... 237
  6.2.4 Multiple base incorporation measurements ................ 238
  6.2.5 Validation of systems with DNA amplification ............... 238
  6.2.6 Characterising relative $C_T$s and sampling for pH-LAMP .... 238

Appendix A Publications 240
  A.1 Conference papers ........................................ 240
  A.2 Journal papers .......................................... 240

Appendix B Principles of DNA hybridisation 241
# List of Figures

1.1 Global deaths projection for 2004 to 2030 by WHO: chronic diseases such as cancers and heart conditions on the rise compared to infectious diseases

1.2 Mechanisms of chronic kidney disease based on [12] [13]

1.3 Historical trend in computer memory and storage

1.4 Technological vision for an epigenetics microchip

2.1 Interactions of epigenetic mechanisms based on [12]

2.2 DNA methylation is the addition of a methyl group to the 5th carbon of cytosines in CpG sites (top). Methylated CpG islands typically affects gene expression by switching transcription off (bottom)

2.3 Biogenesis of miRNA: First pri-miRNAs are transcribed inside the nucleus. After cleavage, subsequent pre-miRNAs are transported out of the nucleus and hairpin structure is cleaved by Dicer. Combination with argo2 forms active RISC mature miRNA which then partially or fully bind to messenger RNA, resulting in blocked translation or cleavage respectively [22]

2.4 Different signaling pathways of renal fibrosis: TGF-β is triggered by epithelial cell injury, causing downstream signaling of Smad2 and Smad3 to bind with Smad 4 to form a Smad complex. It then travels into the nucleus and regulates transcriptions of miRNAs or target genes that promote fibrosis. Specifically Smad7 is a negative regulator of Smad complex and NF-kappaB which promotes Inflammation. Hypertension and high glucose can also affect Smad pathway without TGF-β [44]

2.5 Different types of analysis: Methylation content (left); Methylation level (middle); Methylation pattern (right). Each horizontal line represents double stranded DNA clones derived from different cells, red circles represent methylated cytosines, green are unmethylated [88]

2.6 Preparation methods for preserving methylation: restriction enzymes (left); affinity enrichment (middle); bisulfite conversion (right)
2.7 Second generation sequencing techniques: First common adaptors are used to select fragments of DNA, then amplified with Bridge or Emulsion PCR. Sequencing by synthesis detects nucleotide incorporations by identifying different coloured fluorescent tags; Pyrosequencing captures light released during extension; Ion semiconductor detects hydrogen ions and Sequencing by ligation uses a library of probes labeled with different tags according to the first and second base [95].

2.8 Third generation sequencing methods

2.9 Reverse transcription: stem loop primer contains an overhang structure that is complementary to the target miRNA, after annealing, the hairpin loop unfolds when denatured, allowing primers to bind for amplification (left) [126]; Reverse transcription by poly-A tail starts with attachment of miRNA, a universal primer is used to bind to the poly-A tail, subsequent amplification can be carried out with miRNA-specific primers (right) [127].

2.10 Fluorescent labeling of DNA by: SYBR green II dyes are trapped into double stranded DNA during extension and releases fluorescence accordingly (left); Taqman probes containing reporter and quencher dyes anneal to the target sequence, during DNA extension the probe is cleaved by DNA polymerase, releasing the reporter fluorophore (right).

2.11 A typical qPCR output signal with threshold cycle (C_T) (left); Derivative method (right) [134].

2.12 Methods to determine C_T: Fit point method (left); Derivative method (right).

2.13 Standard curve of threshold cycles from different known dilutions of DNA copies plotted on a logarithmic scale.

2.14 Loop-mediated Isothermal Amplification (LAMP): a. 6 primers on each DNA strand; b. Inner primers (FIP, BIP) bind to multiple site and outer primers (F2, B2) binds to one site and displaces the strand, self hybridising loops form a dumbbell structure from the singled stranded DNA; c. Dumbbell structure provides multiple sites for amplification, generating various regions of the original sequence [144].

3.1 Equivalent circuit of two-electrode system: INT_A and INT_B represent surface of the electrodes treated with DNA probes and in contact with electrolyte, C_{INT_A} and C_{INT_B} are the interface capacitances, R_S and R_P are the solution and interface impedance respectively. Inset illustrates the total external capacitance C_{TOT} between the electrodes. [6]
3.2 Proposed floating gate device in [16], including a control capacitor to bias the gate and active area for sensing ................. 95
3.3 RedOx reactions occurring at two electrodes [21] .................. 96
3.4 Interdigitated electrode in circular configuration (left), RedOx-cycling process (right) [23] .......................... 97
3.5 (a) Cyclic voltammetry input, (b) Output voltammogram [25] ...... 98
3.6 FET-based DNA detection by capturing pyrophosphate (PPi) using chelators [29] .................................................. 100
3.7 ISFET platform reported in [37], with micro-fluidics chambers positioned on top of ISFETs for DNA reactions and detection .................. 101
3.8 Cross sectional diagram of the original ISFET based on [40] .......... 102
3.9 Illustration of the site binding model, Helmholtz and Gouy-Chapman planes with a graph of the charge potential distribution ............ 105
3.10 ISFET behavioural macromodel [50] ................................. 107
3.11 Cross sectional diagram of an ISFET fabricated in unmodified CMOS with the corresponding behavioural model, \( C_{\text{pass}} \) represents passivation layer capacitance ........................................ 108
3.12 Illustrations of (a) DC offset by trapped charge and (b) attenuation by passivation capacitance on ISFET’s characteristics ................. 110
3.13 Behavioural macromodel of programmable gate ISFET reported in [59] 111
3.14 Drift due to different sensing membrane material [61] ............. 112
3.15 Comparison of flicker noise between MOSFETs and ISFETs from [57] 113
3.16 Ion Torrent sequencing platform [1]: (a) Illustration of a well containing single bead with DNA templates, sitting on top of an ISFET device, (b) Incorporation signal of one nucleotide: red trace is modeling of diffusion; blue trace is measured signal ........................................ 114
3.17 Performance of Ion Torrent platforms: (a) Read length improvements proportional to number of wells [73], (b) Accuracy of Ion Torrent PGM decreases with length of homopolymer (results from different template kits) [74] ........................................ 115
3.18 System block diagram of ASIC reported in [78] ...................... 116
3.19 Amplification curves from reactions carried out on ISFET platform reported in [84] ........................................ 118
3.20 Photograph of microchip mounted on SD card next to electronic reader, penny for scale [84] ........................................ 119
4.1 Differential ISFET-REFET readout configuration [3] ................. 134
4.2 ISFET based Gilbert cell [8] .................................................. 135
4.3 ISFET current mirror .......................................................... 137
4.4 ISFET current mirror with multiple output stages for quantitative MSP 138
4.5 Simulated output current to different variables: (a) pH as exponential gain, (b) bias current as linear gain, (c) trapped charge as exponential gain .................................................. 139
4.6 Simulated pH response at different passivation capacitances ...... 140
4.7 Simulated pH response at $C_{pass}$=1fF with, (a) matched trapped charges, (b) unmatched trapped charges .............................................. 141
4.8 Experimental set up with 2 dipsticks and salt bridge ................. 142
4.9 Comparing simulated and experimental pH response at different bias currents: (a) 100nA, (b) 50nA, (c) 10nA ..................... 143
4.10 $I_D - V_{GS}$ sweep of ISFET devices ...................................... 144
4.11 (a) Output current as tunable gain, (b) pH sensitivity at different bias current ................................................................. 144
4.12 Graph showing drift from a single-ended device and current mirror . . 145
4.13 Noise analysis derived from drift of single-ended device and ISFET current mirror ................................................................. 146
4.14 Graph showing effects of temperature in a single device and current mirror 147
4.15 (a) Incorporation signal model developed by Sohbati [25], based on [16], (b) Integral of base incorporation ........................................... 149
4.16 Typical OTA circuit (left): First order OTA-C low pass filter [18] (right) 151
4.17 ISFET-based log domain filter circuit by [19] .............................. 152
4.18 Switched current memory cell [22] .......................................... 152
4.19 ISFET based switched current integrator ................................. 153
4.20 Signals of 1 incorporation and no incorporation from [24] ............. 154
4.21 Diagram of non-overlapping clock signal and circuit .................. 155
4.22 Modeling noise in SI memory cell ........................................... 156
4.23 Micro-photographs of two fabricated chips with annotated dimensions in $\mu$m: red boxes represent ISFET sensors; black outlines contain integrator circuits ................................................................. 157
4.24 Photo of test board with flow cell ........................................... 158
4.25 Simulated PAC response for difference scaling factors ............... 159
4.26 Simulated PAC response for difference clock frequencies .......... 159
4.27 Comparison of simulated and experimental AC response at different clock frequencies ................................................................. 160
4.28 Ideal input signal modeling 0 to 9-nucleotide incorporation events .... 161
4.29 Simulated output transient response for different clock frequencies .... 161
4.30 Modeling 0 to 9 repeated base calls with peak-to-peak noise amplitude
   of 0.01pH ................................. 162
4.31 Simulated output response for 0 to 9 repeated base calls, with 0.01pH
   noise at different frequency spectrums ............................ 163
4.32 .......................... 164
4.33 Input current to emulate 0-9 repeated base calls .......................... 165
4.34 Electrical output response to 0-9 repeated base calls ...................... 165
4.35 $I_D - V_{GS}$ characteristics of ISFET after UV exposures ................. 166
4.36 Diagram illustrating set up using ISFET as input to integrator circuit .. 167
4.37 Variations in source voltage to different reference electrode voltage with
   positive and negative trapped charge .............................. 168
4.38 Effects of input resistance $R_{in}$: 5MΩ (top); 1.2MΩ (middle); 560kΩ (bot-
   tom), red trace represents input square wave, blue trace is integrator’s
   output current ................................. 169
4.39 Measured output response at various input ranges ............................ 170
4.40 Simulated output current noise spectrum ................................. 171
4.41 Measured output response of integrator without signal smoothing ...... 172
4.42 Measured output voltage noise spectrum ................................. 172
4.43 Step 1: dATPs, dTTPs, and dCTPs are added to the sample, DNA
   extension occurs until there is a cytosine in the sequence. Hydrogen ions
   are released in each nucleotide incorporation .......................... 174
4.44 Step 2: dGTPs are added to the sample, an incorporation occurs if there
   is a cytosine, releasing a hydrogen ion ................................. 175
4.45 Representation of results from the new proposed method, with X as any
   nucleotide except cytosine ................................. 175
4.46 Homopolymer distributions of sequences at 200bp ......................... 177
4.47 Homopolymer distributions of sequences at 50bp ......................... 178
4.48 Simulated input and output response of homopolymers for 0 to 90 bases 179
4.49 Effects of incorporation errors due to polymerase on percentage of correct
   methylated cytosines ................................. 181
5.1 Block diagram of the proposed on-chip system for computing $C_T$ using
   the Derivative method ................................. 191
5.2 Graphs showing Matlab generated qPCR curves (top), Fit point method
   (middle) and comparisons of dilution curves (bottom) ....................... 192
5.3 Block diagram of fabricated system with SPI

5.4 Micro-photograph of the fabricated system

5.5 Test board version 2

5.6 Matlab GUI for testing

5.7 Graph and table generated by GUI when testing: blue trace is input curve, red trace shows second derivative computed by the chip

5.8 Graphs showing generated ideal qPCR curves of various efficiencies (top), corresponding second derivative (middle) and dilution curves (bottom)

5.9 Graphs showing qPCR reaction up to 99 cycles (left), and corresponding second derivative (right) with coherent peak occurring in the exponential phase

5.10 Graphs showing fluorescent qPCR data (P1) with primers FAM, JOE, and ROX of three different template concentrations, each with 32 wells

5.11 Graphs showing fluorescent qPCR data (P2) for six different primer combinations, with 2 wells for every 10-fold dilution

5.12 Comparison of average $C_T$ for first set of PCR data (P1)

5.13 Comparison of average $C_T$ for second set of data P2

5.14 Graphs of fluorescent data from pH-LAMP data L1 for four different primer, with 4 wells for every 10-fold dilution

5.15 Graphs of fluorescent data from pH-LAMP data L2 for four different primer, with 4 wells for every 10-fold dilution

5.16 Comparisons of average $C_T$ from pH-LAMP data L1

5.17 Comparisons of average $C_T$ from pH-LAMP data L2

5.18 Block diagram of ISFET read out circuit with system

5.19 Graph of ISFET source voltage for $V_{ref} = 700 \mu V$ to 3.3V

5.20 Graph of measured $C_T$ with ISFET front-end and input through electrode

5.21 Block diagram of ISFET read out circuit with calibration and system

5.22 Experimental set up for pH infusion

5.23 Graph of source voltage of ISFET from change of pH and corresponding second derivative

5.24 Diagram illustrating logic flow of proposed system

5.25 Graph showing theoretical exponential phase of qPCR curve

5.26 Graph of scaled fluorescence data of primer ecolinon-NDM from $L2$

5.27 Graph of ranges of $C_{exp}$ for each primer with positive amplification

5.28 Dilution curves based on $C_{exp}$

6.1 ISFETs with same total top metal areas, and different surface areas
B.1 Structures of complementary DNA strands: thymine with adenine; cytosine with guanine [1] .................................................. 241

B.2 Primers annealing: downstream (top); Upstream (bottom) [2] ........... 242

C.1 Illustration of wire bonding steps: Step 1, position wedge above pad and complete first bond; Step 2, clamp of the wire opens and move wedge in straight line to second bond pad; Step 3: complete second bond; Step 4, clamp closes to cut wire ............................................. 244

C.2 Controls for wire bonding parameters: top row for first bond; bottom row for second bond .................................................. 245

C.3 View of die through wire bonder’s microscope: double row pads can be bonded using a larger loop parameter than the inner interconnects . . . 245

C.4 Close up photograph of a dipstick with glob top encapsulation ......... 246

D.1 CAD drawing of flow cell layers: red represent lines to be laser cut, inlet and outlet holes should be placed aligned to the end corners of the chamber to minimise air pockets ................................. 247

E.1 CAD drawing of modified flow cell layers: an addition opening was added to include salt bridge in the chamber ................................... 248

E.2 Flexible salt bridge connecting two flow cells on dipsticks ............... 249
List of Tables

2.1 Classification of stages in chronic kidney disease [33] ...................... 43
2.2 Table summarising DNA methylation and miRNA regulation associated
with CKD ......................................................................................... 50
2.3 Table of bisulfite conversion based methods for detecting methylation . 58
2.4 Table summarising different methods for quantifying miRNA [150] .... 71

3.1 pH sensitivity for different sensing layer material [41] ......................... 103
3.2 Comparison of existing portable real-time PCR machines .................. 119

4.1 Power consumption at different bias current and pH ......................... 145
4.2 minimum detectable pH based on integrated noise ............................. 147
4.3 Achieved Specifications for ISFET weak inversion current mirror ....... 148
4.4 Fabrication Specifications .......................................................... 157
4.5 Table of theoretical change in pH and measured resolution for each base
call .................................................................................................. 171
4.6 Achieved specifications for ISFET switched current integrator .......... 173
4.7 Composition of Rasal1 sequences .................................................. 176
4.8 Table illustrating different patterns of homopolymers at 10 base resolu-
tion for promoter and CpG island sequences of 50bp ......................... 180
4.9 Table showing percentage accuracy of various polymerase in different
read lengths .................................................................................... 181

5.1 Table comparing $C_T$ obtained from Fit point and Derivative method 193
5.2 pseudo code for “dx” block ......................................................... 194
5.3 pseudo code for “max” block ....................................................... 194
5.4 pseudo code for “counter” block .................................................. 195
5.5 Theoretical slopes of dilution curves for different efficiencies .......... 198
5.6 Average $C_T$ for different calibration methods ............................... 203
5.7 $C_T$ difference between high and low concentration for $P1$ ............... 205
5.8 Average $\Delta C_T$ in pH-LAMP for 10 fold dilutions .......................... 212
5.9 Average pH change in pH-LAMP experiments ................................. 216
5.10 Simulation results of 3-point system to qPCR curves at various efficiencies 222
5.11 Measured results summarising accuracy of the fabricated chip .......... 223
5.12 Results on accuracy of the proposed system with adjusted parameters . 225
5.13 $\Delta C_T$ comparisons for qPCR reactions ................................. 226
5.14 $\Delta C_T$ comparisons for pH-LAMP reactions ............................... 226

C.1 Descriptions of wire bonding parameters ................................. 245
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3’UTR</td>
<td>Three prime Untranslated Region</td>
</tr>
<tr>
<td>AC</td>
<td>Alternating Current</td>
</tr>
<tr>
<td>ACR</td>
<td>Albumin to Creatinine Ratio</td>
</tr>
<tr>
<td>ADC</td>
<td>Analogue to Digital Converter</td>
</tr>
<tr>
<td>AER</td>
<td>Albumin Excretion Rate</td>
</tr>
<tr>
<td>AGE</td>
<td>Advanced Glycation End-product</td>
</tr>
<tr>
<td>AKI</td>
<td>Acute Kidney Injury</td>
</tr>
<tr>
<td>APS</td>
<td>Adenosine 5’ Phosphosulfate</td>
</tr>
<tr>
<td>ASIC</td>
<td>Application-Specific Integrated Circuit</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td>BNC</td>
<td>Bayonet Neill-Concelman</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer-Aided Design</td>
</tr>
<tr>
<td>CDS</td>
<td>Correlated Double Sampling</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin Immunoprecipitation</td>
</tr>
<tr>
<td>CKD</td>
<td>Chronic Kidney Disease</td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal-Oxide-Semiconductor</td>
</tr>
<tr>
<td>COBRA</td>
<td>Combined Bisulfite Restriction Analysis</td>
</tr>
<tr>
<td>CpG</td>
<td>Cytosine-phosphate-Guanine</td>
</tr>
<tr>
<td>CS</td>
<td>Chip Select</td>
</tr>
<tr>
<td>CuuuT</td>
<td>Threshold Cycle</td>
</tr>
<tr>
<td>CV</td>
<td>Cyclic Voltammetry</td>
</tr>
<tr>
<td>DAC</td>
<td>Digital to Analogue Converter</td>
</tr>
<tr>
<td>DC</td>
<td>Direct Current</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA Methyltransferase</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleic Nucleoside triphosphate</td>
</tr>
<tr>
<td>dsDNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>ECM</td>
<td>Extracellular Matrix</td>
</tr>
<tr>
<td>EIS</td>
<td>Electrochemical Impedance Spectroscopy</td>
</tr>
<tr>
<td>EPROM</td>
<td>Erasable Programmable Read-Only Memory</td>
</tr>
<tr>
<td>ESRD</td>
<td>End-Stage-Renal-Disease</td>
</tr>
<tr>
<td>FET</td>
<td>Field-Effect-Transistor</td>
</tr>
<tr>
<td>FFC</td>
<td>Flexible Flat Cable</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>FIR</td>
<td>Finite Impulse Response</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>GFR</td>
<td>Glomerular Filtration Rate</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphic User Interface</td>
</tr>
<tr>
<td>HF</td>
<td>High-Fidelity</td>
</tr>
<tr>
<td>HRM</td>
<td>High Resolution Melting analysis</td>
</tr>
<tr>
<td>IC</td>
<td>Integrated Circuit</td>
</tr>
<tr>
<td>IHP</td>
<td>Inner Helmholtz Plane</td>
</tr>
<tr>
<td>IL-1</td>
<td>Interleukin-1</td>
</tr>
<tr>
<td>IOTA</td>
<td>ISFET Operational Transconductance Amplifier</td>
</tr>
<tr>
<td>ISE</td>
<td>Ion-Sensitive Electrode</td>
</tr>
<tr>
<td>ISFET</td>
<td>Ion-sensitive Field Effect Transistor</td>
</tr>
<tr>
<td>ISH</td>
<td>In Situ Hybridisation</td>
</tr>
<tr>
<td>KP</td>
<td>Klebsiella Pneumoniae</td>
</tr>
<tr>
<td>KPC</td>
<td>Klebsiella Pneumoniae Carbapenemase</td>
</tr>
<tr>
<td>LAMP</td>
<td>Loop-mediated isothermal amplification</td>
</tr>
<tr>
<td>LED</td>
<td>Light Emitting Diode</td>
</tr>
<tr>
<td>LNA</td>
<td>Locked Nucleic Acid</td>
</tr>
<tr>
<td>MALDI-TOF MS</td>
<td>Matrix-Assisted Laser Desorption/Ionization-Time Of Flight Mass Spectrometry</td>
</tr>
<tr>
<td>MBD</td>
<td>Methyl-CpG-Binding Domain</td>
</tr>
<tr>
<td>MeDIP</td>
<td>Methylated DNA Immunoprecipitation</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>MS-SNuPE</td>
<td>Methylation-Sensitive Single Nucleotide Primer Extension</td>
</tr>
<tr>
<td>MSP</td>
<td>Methylation-Specific PCR</td>
</tr>
<tr>
<td>NCD</td>
<td>Non-Communicable Disease</td>
</tr>
<tr>
<td>NDM</td>
<td>New Delhi Metallo-beta-lactamase-1</td>
</tr>
<tr>
<td>NF-KAPPA</td>
<td>Nuclear Factor Kappa-light-chain-enhancer of activated B cells</td>
</tr>
<tr>
<td>NGS</td>
<td>Next-Generation Sequencing</td>
</tr>
<tr>
<td>NHS</td>
<td>National Health Service</td>
</tr>
<tr>
<td>OHP</td>
<td>Outer Helmholtz Plane</td>
</tr>
<tr>
<td>OTA</td>
<td>Operational Transconductance Amplifier</td>
</tr>
<tr>
<td>p-AP</td>
<td>p-aminophenol</td>
</tr>
<tr>
<td>PAC</td>
<td>Periodic-AC</td>
</tr>
<tr>
<td>PC</td>
<td>Personal Computer</td>
</tr>
<tr>
<td>PCB</td>
<td>Printed Circuit Board</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PG-ISFET</td>
<td>Programmable-Gate ISFET</td>
</tr>
<tr>
<td>PGM</td>
<td>Personal Genome Machine</td>
</tr>
<tr>
<td>PoC</td>
<td>Point-of-Care</td>
</tr>
<tr>
<td>PVC</td>
<td>Polyvinyl Chloride</td>
</tr>
<tr>
<td>PWM</td>
<td>Pulse Width Modulator</td>
</tr>
<tr>
<td>RAS</td>
<td>Renin-Angiotensin System</td>
</tr>
<tr>
<td>RC</td>
<td>Resistor-Capacitor</td>
</tr>
<tr>
<td>RCA</td>
<td>Rolling Circle PCR</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>REFET</td>
<td>Reference-ISFET</td>
</tr>
<tr>
<td>RIE</td>
<td>Reactive Ion Etching</td>
</tr>
<tr>
<td>RISC</td>
<td>RNA Induced Silencing Complex</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>RRT</td>
<td>Renal Replacement Therapy</td>
</tr>
<tr>
<td>RSD</td>
<td>Relative Standard Deviation</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SI</td>
<td>Switched Current</td>
</tr>
<tr>
<td>SiNW</td>
<td>Silicon Nanowire</td>
</tr>
<tr>
<td>SMRT</td>
<td>Single Molecule Real Time sequencing</td>
</tr>
<tr>
<td>SNP</td>
<td>Single-Nucleotide Polymorphism</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SoC</td>
<td>System-on-Chip</td>
</tr>
<tr>
<td>SPI</td>
<td>Serial Peripheral Interface bus</td>
</tr>
<tr>
<td>TET</td>
<td>Ten-Eleven Translocation methylcytosine dioxygenase</td>
</tr>
<tr>
<td>TGF- beta</td>
<td>Transforming Growth Factor - beta</td>
</tr>
<tr>
<td>TNF alpha</td>
<td>Tumour Necrosis Factor alpha</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>VHDL</td>
<td>VHSIC Hardware Description Language</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>WI</td>
<td>Weak Inversion</td>
</tr>
<tr>
<td>ZMW</td>
<td>Zero Mode Waveguide</td>
</tr>
</tbody>
</table>
### Mathematical Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>deviation factor of ISFET to Nerstian pH sensitivity</td>
</tr>
<tr>
<td>( \beta_{int} )</td>
<td>intrinsic buffer capacity</td>
</tr>
<tr>
<td>( \chi_{sol} )</td>
<td>solution dipole potential</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>grouping of pH independent chemical potentials</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>channel length modulation</td>
</tr>
<tr>
<td>( \mu )</td>
<td>carrier mobility</td>
</tr>
<tr>
<td>( \phi_0 )</td>
<td>surface dipole potential</td>
</tr>
<tr>
<td>( \phi_b )</td>
<td>bulk solution potential</td>
</tr>
<tr>
<td>( \phi_s )</td>
<td>oxide surface potential</td>
</tr>
<tr>
<td>( \psi_{m}, \psi_{Si} )</td>
<td>metal, silicon work function</td>
</tr>
<tr>
<td>( \psi_f )</td>
<td>Fermi potential</td>
</tr>
<tr>
<td>( \varepsilon, \varepsilon_0 )</td>
<td>permittivity, permittivity of free space</td>
</tr>
<tr>
<td>( \zeta )</td>
<td>ratio of oxide capacitance to passivation capacitance</td>
</tr>
<tr>
<td>( A_{chem}, A_{elec} )</td>
<td>chemical sensing area, polysilicon area</td>
</tr>
<tr>
<td>( b )</td>
<td>y-intercept</td>
</tr>
<tr>
<td>( C_{CF} )</td>
<td>control capacitance (in charge-modulated FET)</td>
</tr>
<tr>
<td>( C_{dl} )</td>
<td>double layer capacitance</td>
</tr>
<tr>
<td>( C_{DNA} )</td>
<td>equivalent capacitance of DNA layer</td>
</tr>
<tr>
<td>( C_{exp} )</td>
<td>threshold cycle computed using 3 point algorithm</td>
</tr>
<tr>
<td>( C_{FB} )</td>
<td>floating gate to silicon bulk capacitance</td>
</tr>
<tr>
<td>( C_{Gouy} )</td>
<td>Gouy-Chapman capacitance</td>
</tr>
<tr>
<td>( C_{Helm} )</td>
<td>Helmholtz capacitance</td>
</tr>
<tr>
<td>( C_{INT} )</td>
<td>interface capacitance (between electrode and solution)</td>
</tr>
<tr>
<td>( C_{ox} )</td>
<td>gate oxide capacitance</td>
</tr>
<tr>
<td>( C_{pass} )</td>
<td>passivation capacitance</td>
</tr>
<tr>
<td>( C_{TOT} )</td>
<td>total external capacitance (between electrodes)</td>
</tr>
<tr>
<td>( C_T )</td>
<td>threshold cycle</td>
</tr>
<tr>
<td>( C' )</td>
<td>capacitance per area</td>
</tr>
<tr>
<td>( E_{reference}, E_{target} )</td>
<td>efficiency of reference, target reaction</td>
</tr>
<tr>
<td>( g_m )</td>
<td>transconductance</td>
</tr>
<tr>
<td>( H(jw), H(s) )</td>
<td>transfer function</td>
</tr>
<tr>
<td>( i_{in} )</td>
<td>small signal input current</td>
</tr>
<tr>
<td>( I_{out}, i_{out} )</td>
<td>large signal, small signal output current</td>
</tr>
<tr>
<td>( I_{ref} )</td>
<td>reference bias current</td>
</tr>
<tr>
<td>( I_0 )</td>
<td>MOSFET characteristic current</td>
</tr>
</tbody>
</table>
\( I_B \) bias current
\( I_D \) drain current
\( i_f \) feedback current
\( I_{meth}, I_{unmeth} \) current from methylated, unmethylated primer reactions
\( i_n \) noise current
\( i_{pH} \) small signal current due to pH
\( k \) Boltzmann constant
\( K_a, K_d \) association, dissociation constant
\( L \) gate length
\( M \) number of methylated templates
\( m \) slope
\( n \) weak inversion slope factor
\( N_s \) surface binding sites
\( pH_b, pH_s \) bulk solution pH, sensing surface pH
\( q \) electron charge
\( Q_{depl} \) depletion charge
\( Q_{dl} \) Double layer charge
\( Q_{DNA} \) charge from double stranded DNA
\( Q_{ox} \) gate oxide charge
\( Q_{ss} \) silicon/gate oxide interface charge
\( Q_s \) sensing surface charge
\( R \) ratio (of threshold cycles)
\( R_{DNA} \) equivalent impedance of DNA layer
\( R_{in} \) input impedance
\( R_P \) interface impedance (between electrode and solution)
\( R_S \) solution impedance
\( S_{ISFET}, S_N \) pH sensitivity of ISFET, Nerstian response
\( S_{pass} \) scaling factor due to passivation layer
\( T \) temperature
\( t \) thickness
\( U \) number of unmethylated templates
\( U_t \) thermal voltage
\( V_{chem} \) ISFET chemical dependent voltages
\( V_{DS} \) drain-source voltage
\( V_{G\nu} \) voltage at polysilicon gate
\( V_{G'} \) gate potential at ISFET’s sensing membrane
\( V_{GS} \) gate-source voltage
\( V_{in} \) input voltage
\( V_{ref} \) reference electrode voltage
\( V_{tc} \) trapped charge voltage
\( V_{th} \) threshold voltage
\( V_S \) source voltage
\( W \) gate width
Chapter 1

Introduction

1.1 Motivation

The future of healthcare is personalised medicine. Over the years, progress in genetic research has been tremendous. From when the Human genome project [1] was first completed in 2001, spanning over 15 years with a total cost of $2.7 billion dollars, sequencing a whole genome now takes only one day for less than $1,000 dollars\(^1\). Genetic testing is already becoming increasingly a norm with many companies offering commercial at-home testing kits, from genotyping across all chromosomes by 23andme\(^2\) to individual SNP\(^3\) detection by GeneU\(^4\). The public’s enthusiasm in genetic testing is not only restricted to their own genome. uBiome\(^5\) is a sequencing-based screening test for your microbiome, giving insights on your diet and gut health. Coupled with wearables and smart watches which are dominating the fitness and technology industries, clearly there is a huge interest in collecting health data specific to individuals.

Suitably, the field of epigenetics research is growing more rapidly than ever. While the study of genetics focuses on our inherent genetic make up, epigenetics examines how environmental stimulants can affect gene expressions without modifying the order of DNA sequences. For example, DNA methylation\(^6\) can block transcription, or microRNAs\(^7\) can degrade messenger RNAs responsible for translating genes to proteins. These mechanisms are reversible and regulated in most mammals. However, long term exposure to external stimulants can cause abnormal epigenetic modifications, which may or may not lead to a disease state.

\(^1\)http://www.nature.com/news/is-the-1-000-genome-for-real-1.14530
\(^2\)https://www.23andme.com/en-gb/
\(^3\)Single base variations in a DNA sequence that occurs at a specific position
\(^4\)https://www.geneu.com/
\(^5\)https://ubiome.com/
\(^6\)addition of methyl group to cytosine
\(^7\)short non-coding RNAs
Since its first discovery in the 1950s [2], epigenetics has been used to explain the effects of environmental influences, providing evidence for the debate “nature versus nurture”. Notably, the famous “Dutch famine study” [3] reveals that epigenetic modifications can be heritable. Prenatal exposure to the lack of food constituted to epigenetic changes that were found in adults up to 6 decades later. And whilst we know smoking is linked to higher risks of lung cancer from very early on [4], the exact mechanisms in which this happens were only recently described in terms of epigenetics in the past 10 years or so [5] [6]. In an era where personal data gathering is becoming increasingly a trend, there is no doubt epigenetics would play a vital role in providing health monitoring. Imagine a future where people giving up smoking can oversee their progress through reversed epigenetic modifications, or check whether effects of your diet choices will be passed down to your children. It would completely revolutionise how people make lifestyle choices.

Although we are still a long way away from providing everyday monitoring on your epigenome, recent breakthroughs in linking various diseases with epigenetic mechanisms proves its integration to healthcare is impending [8]. For most of us living in the modern world, the rising trend in chronic disease-related mortalities shown in Fig. 1.1 demonstrates our biggest threat is no longer dominated by fatal infectious dis-

---

8World Health Organisation
Introduction

eases, of which many can be prevented or treated with vaccines and antibiotics. The increasing number of incidence and prevalence of chronic diseases in both developed and developing countries suggests that modern lifestyle is the main contributor to a person’s decline in health [9]. In 2012, cardiovascular diseases and cancer alone were found to be responsible for 25.7 million deaths across the world [10]. Generally these non-communicable diseases (NCD) are not contagious, and are a result of sustained environmental factors. In particular, 10% of cardiovascular related mortality can be attributed to excess sodium intake [10].

While many non-communicable diseases can be prevented by lifestyle changes, it is clearly not an effective solution. Despite efforts in educating healthy diets and exercise, percentage of obesity in the population is continuously rising [11]. Furthermore, between the years of 2008 and 2012, number of deaths caused by NCDs have increased even though global mortality has decreased [9] [10]. It would appear that adequate treatments and diagnosis must also be included in plans to reduce cases of chronic disease. Early detection could raise awareness for the patient to make lifestyle changes to slow progression, and implement treatments as soon as possible. However, most NCDs are diagnosed at later stages due to tendencies of having none or non-specific symptoms initially. In many cases, the disease is not curable and treatments are provided only as a form of management.

![Figure 1.2: Mechanisms of chronic kidney disease based on [12] [13]](image)

By contrast, epigenetics have the potential to shed new light on the development of biomarkers and therapeutics. Using chronic kidney disease as an example: current methods rely on detection of metabolic imbalances such as filtration rates or uremic toxins as an indicator of reduced kidney functions. Often, the changes are so small in the early stages of the disease that they are rarely detected. In addition, these biomarkers are highly variable within and between individuals, depending on their genetics or even diet on the day of analysis. Typically, only when symptoms are severe
or during routine check up of high risks individuals would the disease be diagnosed. Even then, the primary treatment for chronic kidney disease is medications for hypertension. Regulating blood pressure does not recover kidney functions but decelerates progression in an attempt to prevent renal failure. Literature has shown the effects of high glucose and salt on renal scarring through epigenetic mechanisms [14] [15]. There is also evidence that prolonged acute kidney injury in turn triggers aberrant methylation [16]. Identifying the associated alterations would provide a new form of biomarker that is specific to the disease, and in effect bypasses the current methods of diagnosis relying on highly variable imbalances as illustrated in Fig. 1.2. The added therapeutic potentials of reversing these modifications further incentivise the development of such epigenetic biomarkers.

As medical research in epigenetics advances, technology to realise its potential should be developed in parallel. As it stands, semiconductors remain the pinnacle of modern society’s progress. From the development of mobile phones to computers, its success can be largely credited to its scalability and low costs. Motivation to continuously downsize technology led to efforts to miniaturise transistors, fulfilling and
even exceeding Moore’s Law\textsuperscript{9}. The use of silicon, a readily available material in abundance, combined with the now fully mature CMOS fabrication process has resulted in a consistent decrease in manufacturing costs. From 1960 to 2015, Fig. 1.3 shows prices for computer memory has dropped logarithmically by 12 folds\textsuperscript{10}. Nowadays, 1GB of memory will cost $4, in comparison to over $6 million back in 1980\textsuperscript{11}. Furthermore, designing with transistors which are fundamentally a controllable switch has led to a seemingly limitless and diverse portfolio of circuits and functions. Unsurprisingly, efforts to integrate electronics into medicine is on-going. Particularly, research in Lab-on-chips focuses on developing disposable diagnostic tests that can provide faster and cheaper detection than traditional laboratory analysis.

The area of semiconductor-based DNA detection is fairly new. In comparison to fluorescent-based sequencing [17], the first fully electronic DNA detection platform was reported in 2002 [18], twenty-five years later. Nonetheless this did not halt its progress into commercial use in 2010\textsuperscript{12}, with pH-based analysis using ISFETs\textsuperscript{13} emerging as the practical solution amongst other label-free electronic methods. The technique involves direct detection of hydrogen ions released during DNA reactions, and therefore does not require additional tagging yet maintaining a degree of specificity. At the same time, the technology is built on unmodified CMOS devices that need none or very little post processing. This potential for scalability has resulted in high throughput sequencers that can output up to 10Gb in 2 to 4 hours [19]. Low costs in CMOS manufacturing has also allowed for disposable cartridges that can sequence bacteria associated with sepsis, providing end-to-end detection with a tube of whole blood in 3 hours\textsuperscript{14}.

Moving forward, to integrate personalised medicine into healthcare requires cheap and portable solutions that can monitor the impact of environmental influences. To accelerate epigenetics research, we require high throughput sequencing for investigating the whole genome; and to implement subsequent findings, we need fast detection that can be used outside of the laboratory in routine check-ups. Clearly, ISFETs can provide solutions for both. In addition, designing with ISFETs benefits from an already sophisticated archive of robust and state of the art circuit techniques. With both areas of epigenetics and semiconductor-based DNA detection still at the forefront of research, the possibilities for innovation are enormous.

\textsuperscript{9}Number of transistors on an integrated circuit predicted to double every two years
\textsuperscript{10}https://hblok.net/blog/storage/
\textsuperscript{11}http://www.statisticbrain.com/average-historic-price-of-ram/
\textsuperscript{12}by Ion Torrent, later acquired by Thermal Fisher Scientific https://www.thermofisher.com/uk/en/home/brands/ion-torrent.html
\textsuperscript{13}Ion-sensitive Field Effect transistors
\textsuperscript{14}LiDia by DNAe http://www.dnae.com/lidia.html


1.2 Aim & Vision

For this thesis, we aim to truly integrate the areas of semiconductor technology with epigenetic monitoring. By fully understanding what biological information and criteria are needed, as well as limitations in the technology, simple CMOS circuitries can be designed or re-purposed with ISFETs as an application-specific solution.

Our vision is an epigenetic microchip for chronic disease management in future healthcare. In particular, we are interested in detecting DNA methylation and miRNAs as they act as binary modifications such that the “on” or “off” states of methylated cytosines, and increase or decrease in miRNA levels govern gene expressions. ISFETs allow for the development of sensor circuits and easy integration with CMOS back-ends for further processing. For example, both methylation and miRNA analyses can be achieved through amplification-based methods, namely Methylation-specific PCR and quantitative PCR. On-chip processing can then be used to evaluate threshold cycles (C_T) for quantifying miRNAs, with the possibility of shortening detection time since computations are carried out simultaneously. Similarly, base calling accuracies in sequencing when investigating methylation at single base resolution can be improved through integrated front-end or back-end processing. Together, the various systems can be contemplated as a combined epigenetics microchip depicted in Fig. 1.4. Implementation of such an ISFET-based platform has the potential to carry out cheap, portable and simultaneous detection of multiple epigenetic biomarkers without the need of a computer.

Figure 1.4: Technological vision for an epigenetics microchip
1.3 Overview

The overall format of the thesis follows an introduction of necessary background and reviews of available methods and technology, followed by proposed circuits for each epigenetic mechanism with technical findings. The structures of each chapter is outlined below.

1.3.1 The role of Epigenetics in chronic diseases

This chapter introduces epigenetics and describes the modifications: namely methylation and miRNA regulation. A close examination on chronic kidney disease is presented, along with a review of specific epigenetic modifications associated with its pathology. This is followed by an overview outlining the pros and cons of the different analysis available for detecting methylation and miRNAs.

1.3.2 Ion-semiconductors for DNA analysis

A literature review of label-free electronic DNA detection is presented, followed by close inspection of ISFET’s operation, sensitivity to pH and non-ideal effects when fabricating in unmodified CMOS process. The chapter ends with discussions of current ISFET-based technologies for detecting epigenetic mechanisms.

1.3.3 Analogue current mode design approach for methylation

This chapter explores how analogue circuits can be designed to suit the appropriate methylation analyses, namely Methylation-specific PCR and bisulfite sequencing for Point-of-care and therapeutics research applications respectively. Using simulations and experimental results to support findings, two circuits are designed and presented. For MSP, the focus is on differential sensing for opposite primers, therefore an ISFET current mirror is proposed. A brief review of past differential ISFET configurations shows that weak inversion operation can reduce drift and temperature sensitivity. For bisulfite sequencing, the focus is on improving base call accuracies particularly for homopolymers through signal processing. An ISFET switched current integrator is proposed based on the merits of discrete time current mode operations and potential increase in pH-resolution through averaging by summation. Additionally, a novel protocol specific for methylation analysis in ion-semiconductor sequencing is investigated. The feasibility of the proposed method is investigated by Matlab simulations.
1.3.4 On-chip threshold cycle evaluation and Early amplification

On the basis of an existing ISFET platform capable of real-time amplifications, this chapter develops two on-chip back-ends using VHDL to carry out miRNA quantification analysis and reduce total detection time. The first system is implemented based on the Derivative method for evaluating threshold cycles, the second system employs a simple algorithm to detect exponential phase in an amplification reaction. The fabricated systems are tested with experimental fluorescent-based PCR and pH-LAMP data sets, with a demonstration of an end-to-end operation.

1.3.5 Conclusion and Future work

Findings from each chapter are summarised. Concluding remarks of merits and limitations of each circuit design along with future work are discussed.
Bibliography


Chapter 2

The role of Epigenetics in chronic disease

According to the World Health Organisation (WHO), non-communicable diseases are responsible for 70% of global mortality [1]. Such conditions include cancer, cardiovascular disorders, diabetes, chronic respiratory and kidney disease. The nature of non-communicable diseases is not transmittable between person to person, and develops gradually particularly in aging populations. By 2050, the number of people aged 60 or over will outnumber adolescents and young adults between the ages of 10-24. With global life expectancy steadily increasing, together with the obesity epidemic, the total annual deaths from NCDs are expected to rise to 55 million by 2030 [2]. The current and anticipated economic burden on health services and global GDP is enormous.

The likelihood of developing chronic diseases can often be attributed to environmental influences with varying degrees of heritability [3]. Epigenetic changes have long been suspected to contribute to the onset of non-communicable disorders [4]. Aberrant DNA methylation patterns are found in tumour suppressor genes in early cancer development [5, 6]. Epigenetic alterations acting as mediators for translating environmental stimulants to disease risks is supported by a body of literature [7, 8]. Understanding its mechanisms can provide a means to track the impact of environmental factors on a cellular and molecular level. As there are no cures for most chronic diseases, the reversible nature of epigenetics provides a huge prospect for developing new therapeutics [9]. Even if the pathway in which epigenetics affects the disease is found to be non-causal, it can still act as a powerful predictor for NCDs prior to symptoms.

Taking chronic kidney disease as a case study, this chapter examines the mechanisms in which epigenetics can influence disease development, namely DNA methylation and microRNA regulation. Followed by an overview of detection techniques for each modification.
2.1 Epigenetics

Epigenetics can be defined as changes in gene expressions without altering the order of genetic sequences. Derived from the Greek word “epi”, it literally translates to “in addition to”, and can be thought of as a layer above genetics. Since every cell in our bodies possesses the same genetic code, it explains how the same genotype can influence different phenotypes. The understanding of cell differentiation is key in developmental biology, and is famously depicted by Waddington’s epigenetics landscape in 1957 [10]. The imagery likens the many paths a marble can roll down a hill to the various differentiations a cell can develop to. By all means, the developmental process comprises many dynamic feedbacks and involvements from many different regulatory genes.

Whilst naturally occurring epigenetic modifications are essential, when triggered incorrectly can cause health and behavioural problems. Known environmental drivers such as heavy metals, pesticides, air pollutants have been linked to chronic diseases related to cardiovascular, respiratory dysfunction, cancerous mutations and many more [7, 8]. As well, early studies in mice discovered the effects of maternal behaviour in altering the epigenome [11]. The study suggested tendencies to anxious behaviour in adult rats were a result of reduced physical contact such as licking by their mothers as a pup. By silencing alleles through imprinting, these epigenetic modifications that occur in life can be passed down from parents to offspring, demonstrating epigenetics’ role in both nature and nurture.

There are 3 main mechanisms in which epigenetics can affect gene expressions: histone modifications, DNA methylation and RNA interference [12]. Both histone modifications and DNA methylation affect expressions through transcriptional silencing. Histones enable the entire genome to be stored in every cell nucleus by acting like a spool to thread for DNA to wrap around. The accessibility of the DNA sequence therefore determines whether transcription can take place. Modifications imply the histone tails are acetylated, phosphorylated, ubiquitylated or methylated [13]. Transcriptionally active regions typically contains high levels of histone acetylations which corresponds to a sector of loosely coiled DNA sequence. Such histone modifications may encourage DNA methylation and vice versa., hence reinforcing the intended transcriptional activation or repression. Meanwhile small-interfering RNAs can influence both DNA methylation and histone modifications via regulation of enzymes that carry out the alterations. The 3 mechanisms exist and interact dynamically to affect gene expressions (see Fig. 2.1).

Development of epigenetic-based biomarkers have been primarily focused on DNA methylation and small-interfering RNAs such as microRNA [5, 14, 15]. The binary
Figure 2.1: Interactions of epigenetic mechanisms based on [12]

states of methylation in regions can be readily correlated to specific gene silencing. In
comparison, the role of histone modifications in disease development are not as exten-
sively studied due to the complexity of the histone code. MicroRNA expressions are
primarily altered by different disease conditions, and high levels of extracellular mi-
croRNAs can be found in bodily fluid, showing its potential for non-invasive diagnosis.
The following sections examine DNA methylation and microRNA regulation in more
details.

2.1.1 DNA methylation

DNA methylation refers to the addition of a methyl group (CH$_3$) to the 5th carbon
of the base cytosine [16]. The conversion to 5-methylcytosine is carried out by DNA
methyltransferase (DNMT) enzymes. DNMT3a and 3b are responsible for de novo$^1$
methylation typically during early development, while DNMT1 maintains methylation
patterns in cell replications [17]. With the exception of embryonic stem cells, DNA
methylation occurs almost exclusively at CpG sites where a guanine follows a cytosine.
CpG sites occur globally throughout the genomic sequence, though more frequently

$^1$new, in this case previously unmethylated
as clusters named CpG islands\(^2\) near promoter regions. Majority of CpG sites in the genome are methylated, while CpG islands typically remain unmethylated \([19]\). The additional methyl group in DNA can down-regulate gene expressions by recruiting more histone repressors or preventing access to the sequence, effectively blocking transcription.

Often named as the “5th base”, DNA methylation plays an important role in gene silencing associated in diseases \([4]\). It is widely accepted that aberrant methylation is linked to carcinogenesis, and the pattern may be tumour specific \([5, 6]\). In addition to cancer, irregular methylation is found to play a role in neurological disorders and autoimmunes diseases \([8]\). Hypermethylation in promoter regions have been found to be associated with gene silencing. Interestingly, genomic methylation was found to decrease in general as mammals age \([20]\) suggesting a link in which epigenetics may play a part in the development of chronic diseases in older population.

\(^2\)Regions of DNA greater than 500bp, with percentage of total guanine and cytosine contents equal or greater than 55% and observed/expected CpG ratio of 0.65 \([18]\)
The process of demethylation used in epigenetics reprogramming can also provide a means for potential therapeutics to be developed. Passive demethylation is carried out by the maintenance methyltransferase DNMT1 through cells divisions; active demethylation is the removal of methyl group by enzymes. The use of TET enzymes to convert methylated cytosine to naked cytosines via hydroxymethyl conversion in recent publications show promising results in reversing aberrant methylation patterns [21].

2.1.2 miRNA Regulation

MicroRNAs (miRNA) are small, single stranded non-coding RNAs responsible for regulating a range of biological cell functions such as differentiation, proliferation and apoptosis\(^3\) [22]. They vary from 21-25 nucleotides in length, and contain uracils in

---

\(^3\) programmed cell death
place of thymines. Initially thought to be useless in mammalian species, miRNAs are now known to play a crucial role in post transcriptional silencing of expressions [23].

The regulation of miRNAs is considered an epigenetic mechanism. Two main enzymes named Drosha and Dicer are responsible for the biogenesis of miRNA [24]. Primary miRNAs (pri-miRNA) are first transcribed from the corresponding gene in cell nucleus. The initial double stranded miRNA is around 100-120 nucleotides long, and cleaved by Drosha to about 70 nucleotides named precursor miRNA (pre-miRNA). The subsequent pre-miRNA is transported out of the nucleus into the cytoplasm with exportin5, which can act as a rate limiting factor for the biogenesis process. Dicer then cleaves off the hairpin loop structure leaving the remaining miRNA duplex which interacts with argo2⁴ and associated proteins to form RISC (RNA induced silencing complex) [25]. The resultant single stranded mature miRNA is now capable of binding to the 3'-UTR⁵ region of messenger RNAs to cleave the target or block translation [26]. A fully complementary match to the target mRNA will result in cleavage and subsequent degradation; partial match requires a region of approximately 6-7 nucleotides in length, and inhibits binding by preventing access to the messenger RNA. An average miRNA has 100 target sites [27]. The ability to bind to sites even with mismatches allow miRNA to regulate many different gene expressions.

2.2 Chronic Kidney Disease

Chronic Kidney Disease (CKD) is one of the non-communicable diseases rapidly on the rise. From 1990 to 2010, [28] reported CKD to have increased from the 27th to 18th leading cause of death globally. Defined as the long term decline in kidney functions, risk factors related to modern day living such as obesity, smoking, inactivity and alcohol consumption contribute to the increase in CKD patients. Despite advances in healthcare, incidence and prevalence for CKD remain high, affecting 6-13% of the adult population [29]. As well as an increase of mortality, CKD and its complications affect a wide scope of organs. Diabetes and hypertension act as the leading cause and mutual risk multipliers for renal failure resulting in kidney replacement therapies. Between 2009 and 2010, the NHS spent £1.45 billion on CKD and related conditions, and more than half of the budget was spent on dialysis and transplants [30]. Early detection and accurate monitoring have been shown to increase patients’ survival rate and reduce risks for cardiovascular and other comorbid diseases [31, 32].

⁴Argonaute 2
⁵three prime untranslated region - section following coding sequence in messenger RNA, typically not translated
The role of Epigenetics in chronic disease currently no cure for CKD. The most effective treatment to date is Renin-angiotensin system (RAS) inhibitors, which is used to regulate the renal haemodynamics to alleviate stress on the blood vessels. Such treatments can only manage and slow progression, thus emphasising the importance of early detection in CKD and need for new therapeutics.

2.2.1 Glomerular Filtration Rate (GFR)

Table 2.1: Classification of stages in chronic kidney disease [33]

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
<th>GFR</th>
<th>Action</th>
<th>Symptom</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kidney damage with normal or increased GFR</td>
<td>&gt;= 90</td>
<td>Diagnosis and treatment; treat comorbid conditions, slow progression; reduce risk for Cardiovascular disease</td>
<td>None, may be slightly elevated creatinine and urea</td>
</tr>
<tr>
<td>2</td>
<td>Kidney damage with mild decrease of GFR 60-89</td>
<td></td>
<td>Estimate progression</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Moderate decrease in GFR 30-59</td>
<td></td>
<td>Evaluate &amp; treat complications</td>
<td>fatigue, loss of appetite, itching</td>
</tr>
<tr>
<td>4</td>
<td>Severe decrease in GFR 15-29</td>
<td></td>
<td>Prepare for kidney replacement therapy</td>
<td>In addition: shortness of breath, vomiting, change in skin colour, ammonia breath</td>
</tr>
<tr>
<td>5</td>
<td>Kidney failure</td>
<td>&lt;15 (or dialysis)</td>
<td>Kidney replacement (if uremia present)</td>
<td></td>
</tr>
</tbody>
</table>

The clinical definition of chronic kidney disease is determined by the level of kidney damage measured by glomerular filtration rate (GFR). Glomerulus refers to the cluster of capillaries at the entrance of nephrons, which are small basic filtering units that make up the kidneys. The filtration rate can evaluate kidney function by measuring the amount of waste products filtered from blood per minute (mL/min). CKD is characterised based on kidney damage or a GFR of less than 60mL/min per 1.73m² for 3 months or more [33]. The condition is further classified into 5 stages based on GFR shown in Table. 2.1. The direct impact each stage has on the kidneys is unclear, although later stages have a higher risk of mortality, and clinical significance starts in stage 3. The early stages of CKD are often asymptomatic, even when presented at stage 3, they appear as non-specific symptoms. Stage 5 is classified as end-stage-renal-disease (ESRD), when almost all kidney functions have stopped and the patients would require renal replacement therapy (RRT) in the form of dialysis or transplant. Once on RRT, mortality rate in CKD patients is significantly increased with cardiovascular conditions as a primary contributor [34]. In addition, the quality of life in dialysis patients deteriorates greatly, with some stating they would trade 10 years on dialysis with only 6 years of normal kidney function [35].
So far the number of biomarkers used to diagnose CKD are limited to estimated GFR (eGFR) and urinary proteins [36]. Measured GFR requires continuous delivery and multiple measurements over time of filtration markers like Inulin through Intravenous fusion [37]. Therefore estimations based on serum creatinine and prediction formula taking into account age, ethnicity and sex are often used instead [38, 39]. However, in the early stages of CKD (stage 1-2), GFR may remain unchanged or even increases due to healthy nephrons overcompensating for initial glomerulus injuries through hyperfiltration and hypertrophy. Differences in creatinine levels due to mild kidney damage is comparable in magnitude to other variables such as diet or rate of creatinine break down. Small decreases in eGFR may also be due to tubular secretion of creatinine which can vary up to 10% for a healthy subject. In addition, differing choice of assays between laboratories and approximations made in the formulas contribute to inaccuracies of eGFR [33]. Alternatively, the presence of proteins in urine like levels of albumin can reflect the extent of kidney damage. Though urinary albumin is highly variable due to metabolism, albumin excretion rate (AER), or albumin to creatinine ratio (ACR) is considered to be a better indicator for severity, and often used in routine testing for high risk patients with diabetes or high blood pressure [40].

2.2.2 Fibrosis

Understanding the pathogenesis of renal fibrosis, the process leading to ESRD, is key to identifying new potential biomarkers for diagnosis and therapeutics of the disease. A common feature of CKD regardless of etiology is the accumulation of extracellular matrix (ECM) resulting in a fibrotic kidney [41]. Renal fibrosis commonly occurs in the glomerulus or tubules of a nephron [42]. As a response to epithelial injury, transforming growth factor TGF-β1 binds to its corresponding receptors and triggers a downstream Smad signaling [43]. Phosphorylated Smad2 and Smad3 bind with Smad4 to form a Smad complex, which then translocates to the nucleus and regulates the transcription of target genes and miRNAs [44]. The role of Smad3 is pathogenic in promoting collagen production and preventing ECM degradation, while Smad2 can mediate fibrosis by competitively blocking Smad3 phosphorylation or translocation [45]. One of the regulated target gene expressions include Smad7, which can inhibit Smad complex formation as well as inflammation triggered by NF-κB. Pro-inflammatory cell signaling proteins such as TNFα and IL-1 are responsible for the activation of NF-κB as part of the body’s immune system [46]. In addition, advanced glycation end-product (AGE) and angiotensin II induced by diabetes and hypertension respectively can activate the Smad pathway independently without TGF-β1 [47]. The promotion
and suppression of fibrosis through regulating gene and miRNA expressions suggest the likely involvements of epigenetic alterations.

Figure 2.4: Different signaling pathways of renal fibrosis: TGF-β is triggered by epithelial cell injury, causing downstream signaling of Smad2 and Smad3 to bind with Smad 4 to form a Smad complex. It then travels into the nucleus and regulates transcriptions of miRNAs or target genes that promote fibrosis. Specifically Smad7 is a negative regulator of Smad complex and NF-κappaB which promotes Inflammation. Hypertension and high glucose can also affect Smad pathway without TGF-β [44]
2.2.3 Epigenetics in CKD

Recently, epigenetic mechanisms have gained attention as potential biomarkers for chronic kidney disease [12, 48, 49]. For example, miRNAs have been shown to be essential in maintaining homeostasis in the kidneys. The inactivation of Dicer, a crucial enzyme in the biogenesis of miRNAs, can lead to excessive proteins in the urine, renal failure and eventually death in mice [50]. Global hypermethylation was also found in CKD patients, and correlated with inflammation and oxidative stress [51]. Numerous studies conducted using DNA and miRNA extracted from whole blood, saliva and urine demonstrates their potential as non-invasive biomarkers [52–54]. Using only saliva, multiple differentially methylated CpG sites were found on 187 genes comparing end stage CKD patients and diabetic patients with no nephropathy [53]. Particularly, circulating miRNAs are extremely stable in body fluid such as plasma and urine. A high abundance of miRNAs were found in plasma despite the presence of RNA-degrading enzymes (RNase) [55]. Additionally, expression levels of miR-21 in urine sample were unaffected after storing in room temperature for up to 24 hours [54].

Pathogenesis

Identifying specific epigenetic modifications can help further understanding in the pathogenesis of chronic kidney damage. The key differentiation between normal kidney repair recovering from acute kidney injury (AKI), to scarring of the tissues lies in the persistent production of TGF-β1 [56, 57]. Under prolonged exposure to the growth factor hormone, Betchel et al. [58] showed that Rasal1 promoter becomes hypermethylated in vitro. Subsequent decrease in its expressions caused downstream fibroblasts activation from signaling proteins. In a recent genome wide study, Cux1 was identified as one of six genes to have significantly differential methylation levels between CKD patients and healthy controls [52]. Previously it had been associated with cell proliferation [59] and fibrosis in the glomerular sections of the kidneys [60]. [61] reported that a decrease in Cux1 expression levels were found in fibriotic tissues in mice. Interestingly they found TGF-β was only able to activate collagen deposition in the absence of Cux1, suggesting anti-fibrotic properties of Cux1. Evidently, methylation plays an important role in regulating the delicate balance of gene expressions to maintain a healthy kidney.

Severity and Progression

Several studies have linked the severity of CKD with methylation and miRNA regulations. A study carried out using DNA extracted from both whole blood and renal

---

6increase in methylation level at individual or multiple CpG sites
tissues from CKD patients showed an inverse correlation between eGFR and promoter methylation in the anti-aging gene Klotho [62, 63]. Previously, [64] showed promoter methylation in Klotho reduces its expressions, implying the gene is less expressed in CKD kidneys. [65] reported expression levels of 5 circulating miRNAs in plasma to be positively proportional with eGFR, with significant differences between stage 3-4 and ESRD patients. miR-29c was also found to be positively correlated with eGFR, and low abundance was associated with severe tubulointerstitial fibrosis [66]. In the study, miR-29c, miR200a and c were able to distinguish mild and moderate fibrosis while proteinuria did not show any correlation. Recent literature examined 100-200 non-coding RNAs in various stages of CKD, and found 16 miRNAs to have significant differences in expression levels between healthy and CKD patients [67]. Particularly, miR-181a was down-regulated by 200 fold in both early and late stage CKD patients, demonstrating a potential as an early detection biomarker.

**Diabetic nephropathy**

Methylation may also be used to identify susceptibility of renal related complications such as nephropathy in diabetic patients. [68] determined 19 possible CpG sites that correlated with the time to onset of kidney disease in type I diabetic patients. Specific genes such as Elmo was previously linked to susceptibility of developing diabetic nephropathy [69, 70]. Elmo was also recently reported to be hypomethylated in a genome wide study using blood derived DNA from CKD patients [52]. Similarly, promoter methylation in p66sHC was found to be reduced in ESRD patients [71], in agreement with an increase of expressions found in diabetic nephropathy patients [72]

**Dietary influences**

The effects of dietary stimulants on chronic kidneys can be closely studied in epigenetic modifications. For example, the miR-29 family exhibit diverse characteristics under different glucose and salt conditions, despite having the same seed region and many overlaps in target genes [73–77]. Members of miR-29 are generally considered to be antifibrotic and pro-apoptosis. Under high glucose exposure, miR-29a was found to be down regulated via TGF-β1 signaling in vitro, with expressions falling by 40-50% in 7 days [74]. The decrease in abundance contributed to increased expressions of collagen production genes, suggesting miR-29a act as a repressor for collagen deposition and may be used as a marker for diabetic nephrothy. Similarly, miR-29c was reported to influence cell-apoptosis and fibroblasts synthesis in diabetic mice under hyperglycemia conditions.
[77]. Using rats that were sensitive to salt-induced hypertension, [75] reported a higher abundance of miR-29b in the control compared to salt-sensitive rats under a high salt diet. Increase in several extracellular matrix genes implies miR-29b may act as an inhibitor of fibrosis.

Environmental stimulants

Other environmental factors such as metal exposure have also been shown to alter the epigenome. A long term study on the effects of Cadmium contaminant exposure was conducted in China over an 8 year period [78]. Zhang et al. found that there was hypermethylation in Rasal1 and Klotho genes when comparing subjects from highly exposed to unpolluted areas. They also correlated the degree of methylation with eGFR and found them to be statistically significant even after adjusting for other epigenetic related factors such as age, sex, smoking and drinking. Elevated methylation levels in CKD for both genes are also supported by multiple studies that were mentioned above [58, 62]. Methyltranferase is suspected to play a part in altering these methylations. Injection of undialyzable uremic toxins into cultured kidney cells showed an increase of DNMT1, 3a and 3b expressions, Klotho hypermethylation and subsequent fibrosis [79]. [58] also reported an increase in DNMT1 when investigating Rasal1.

Potential therapeutics

The therapeutic potential of CKD epigenetic modifications has been widely demonstrated in numerous literature. Promoter hypermethylation of Rasal1, which has been reported as a key mediator in sustaining fibroblasts activation in CKD, was normalised using BMP7 (bone morphogenic protein 7) treatment [80]. The study reported reduction in expression levels of demethylating enzymes TET3 by growth factor hormones (TGF-β1), followed by successful restoration with BMP7 in fibrotic mice. In a separate study, a low dose of antihypertensive drug (hydralazine) was used to reverse aberrant methylation in Rasal1 [81]. The resulting demethylation was able to attenuate fibrosis independently of the therapeutic effects of lowering blood pressure. The application of treatment prevented acute renal injury from developing into chronic kidney damage in mice. Similarly, [82] reported the up-regulation of Klotho expressions after Rhein treatment in mice with induced renal fibrosis. Klotho is known to increase resistance to oxidative stress [83], and expression levels were found to be reduced in CKD patients [62]. The corrected aberrant methylation as well as DNMT1 and 3a expressions were able to reverse repression and alleviate injury.

miRNA regulation can be potentially corrected with appropriate plasmids using
ultrasound delivery systems. miR-21 was reported to be up-regulated as a result of TGF-β1 signaling [84]. The subsequent increase in collagen and fibrotic proteins by miR-21 was then minimized when miR-21 knockdown plasmids were delivered. Qin et al. [85] were also able to normalise expression levels of miR-29b to block fibrosis in mice, demonstrating the possibility of miRNAs as therapeutic targets for chronic kidney disease.

**Summary**

In summary, we have shown the potential of using DNA methylation and miRNA regulation in CKD as non-invasive biomarkers for early detection, severity, progression, etiology and environmental stimulants (summarised in Table. 2.2). Other areas such as monitoring effectiveness of existing treatments may also be explored. Differing expressions of miR-21 and miR-200b were found between kidney transplant patients with and without atrophy, demonstrating miRNA may also be used as an indicator of allograft success [86]. In the literature review, some discrepancies in changes of expressions levels were found between studies of the same miRNA. This may be due to experiments carried out in different samples or analysis methods. Particularly, regulation of circulating miRNA may be opposite to tissue-based results as diseased cells may release an increased number of exosomes containing miRNAs into blood or urine. Data normalisation (discussed in more detail in Section.2.4) also plays a part in the consistency of results between different laboratories. Provided that the relevant alterations are studied extensively in the desired medium for detection, epigenetic modifications have huge potential as new therapeutic targets and diagnostic biomarkers for CKD.
Table 2.2: Table summarising DNA methylation and miRNA regulation associated with CKD

<table>
<thead>
<tr>
<th>Gene/ miRNA</th>
<th>Modification</th>
<th>Subject</th>
<th>Sample Type</th>
<th>Description of study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cux1 mC↑</td>
<td>Human</td>
<td>Blood</td>
<td>Genome wide study [52]</td>
<td>Associated with scarring in the glomeruli [60], cell proliferation [59], inhibits of collagen [61]</td>
</tr>
<tr>
<td></td>
<td>Mice</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elmo1 mC↓</td>
<td>Human</td>
<td>Blood</td>
<td>Genomic wide study [52]</td>
<td>Elevated expressions associated with enhanced extracellular matrix deposition in chronic kidney injury [69], [70]</td>
</tr>
<tr>
<td></td>
<td>Mice</td>
<td>Tissue</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Klotho mC↑</td>
<td>Blood, Tissue</td>
<td>Blood, Urine</td>
<td>eGFR decline correlated with increase in promoter methylation [62]</td>
<td>Cadmium exposure correlated with increase levels of methylation [78]</td>
</tr>
<tr>
<td></td>
<td>Tissue</td>
<td>Tissue</td>
<td>Associated with anti-aging [63], resistance to oxidative stress [83]</td>
<td></td>
</tr>
<tr>
<td>p66Shc mC↓</td>
<td>Blood</td>
<td>Blood, Tissue</td>
<td>Promoter methylation reduced in ESRD patients [71]</td>
<td>Decreased expressions associated with tubular kidney injuries in diabetic nephropathy [72]</td>
</tr>
<tr>
<td>Rasal1 mC↑</td>
<td>Human, mice</td>
<td>Tissue, urine</td>
<td>Progression from acute to chronic kidney injury, sustained fibroblasts activation [58]</td>
<td>Cadmium exposure correlated with increase levels of methylation [78]</td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>Blood</td>
<td>Increase collagen deposition, fibronectin production [84]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mice</td>
<td>Tissue</td>
<td>Positive correlation with eGFR in kidney transplant patients [86]</td>
<td></td>
</tr>
<tr>
<td>miR-21↑</td>
<td>Mice</td>
<td>Urine</td>
<td>Negative correlation with eGFR in stage-3-5 CKD patients [65]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>Blood</td>
<td>Down-regulated with high glucose exposure, Suppresses collagen production [74]</td>
<td></td>
</tr>
<tr>
<td>miR-29↓</td>
<td>Mice</td>
<td>Tissue</td>
<td>Targets collagen matrix, down-regulated by Smad3 in TGF-β signalling in induced fibrosis [85]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Human</td>
<td>Tissue</td>
<td>Master regulator of multiple collagen-related genes, down-regulated in high salt diet [75]</td>
<td></td>
</tr>
<tr>
<td>miR200a,b,c↓</td>
<td>Human</td>
<td>Urine</td>
<td>Correlated with severity and degree of tubular-based fibrosis. Can distinguish mild and moderate fibrosis in CKD patients [66]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urine</td>
<td>Significant inverse correlation with eGFR in transplant patients [86]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.3 Methylation Analysis

There are a number of methods for detecting methylation. The choice of technique depends on a variety of factors, such as the aim of the study, amount and quality of starting DNA, specificity and sensitivity required, costs etc [87]. More importantly, there are different types of analyses depending on the required information: methylation content refers to all methyl-cytosines in every position, such as global methylation; methylation level can be defined quantitatively as percentage methylation based on the same CpG site assessed across different clones; methylation pattern is a series of CpG in a region of DNA. Methylation can be fully, partially (50%), or unmethylated (0%) in each cell since there exists two copies of chromosomes. A genomic study would contain DNA molecules derived from a variety of cells [88]. Percentage methylation can be calculated based on the following eq. 2.1, where M and U represent the number of methylated and unmethylated cytosines respectively:

\[
\% \text{ methylation} = \frac{M}{M+U}
\]  

(2.1)

Figure 2.5: Different types of analysis: Methylation content (left); Methylation level (middle); Methylation pattern (right). Each horizontal line represents double stranded DNA clones derived from different cells, red circles represent methylated cytosines, green are unmethylated [88]

2.3.1 Preparation methods

Initial preparation that is sensitive to methyl-cytosines is necessary prior to amplification or hybridisation (see Appendix B) in all methylation analyses. Without the
The role of Epigenetics in chronic disease

In the presence of DNMT, 5-methyl cytosines are not maintained when DNA replicates. It has been speculated that with the correct conditions and DNMT, methylation status could be preserved during amplification [89]. In general, most techniques rely on methylation dependent treatments to distinguish the methylated DNA using Restriction enzymes, Affinity enrichment or Bisulfite conversion, summarised in Fig. 2.6 below.

![Diagram of methylation preparation methods](image)

**Figure 2.6: Preparation methods for preserving methylation: restriction enzymes (left); affinity enrichment (middle); bisulfite conversion (right)**

Restriction enzymes are widely used to make recombinant DNA in gene cloning as well as DNA fingerprinting [90]. Restriction enzymes will digest the DNA at specific sites with the same recognition sequence. For methylation analysis, enzymes such as SmaI and HpaII are used to cleave only unmethylated sequences, leaving the methylated cytosines intact. For example, the enzyme HpaII shown in Fig. 2.6 (left) cleaves any unmethylated CpG sites that exist in the sequence “CCGG”. The remaining fragments are size selected and then hybridised or sequenced to determine the methylations status. This approach has limited genome coverage, restricted to areas close to the enzymes’ recognition sequences. Global methylation can be determined by the digestion ratio of restriction enzymes. For example, HpaII and MspI have the same recognition sequence, but MspI does not differentiate methylation and will cleave all cytosines. Low ratios of HpaII to MspI in an analysis therefore indicates a high methylation content.

Affinity enrichment is a method to separate methyled cytosines based on immuno-precipitation [91]. Similar to chromatin Immunoprecipitation (ChIP) used to detect histone modifications, crosslinking proteins first bind to specific areas of chromatin. The DNA is sheared based on protein positioning using endonuclease or sonication. Antibodies will then attach to the targeted protein and separate the regions. In the
The role of Epigenetics in chronic disease

case of MeDIP (methylated DNA immunoprecipitation), antibodies specific to 5-methyl cytosines or methyl binding proteins (MBD) are used. The enriched DNA therefore contain methylated sequences and can be detected using array hybridisation or sequencing. Regions with more CG contents are more likely to be enriched than CG poor areas, even if both are fully methylated [92].

Originally developed in the 1970s [93], bisulfite conversion is considered the gold standard for preparing methylated DNA. In bisulfite conversion, sodium bisulfite is treated to the sample DNA to preserve 5-methyl cytosines [94]. The chemical reaction converts all unmethylated cytosines to uracils, while 5-methyl cytosine remain unchanged. The converted uracils are later transformed to thymines in the amplification step. The resultant sequences will be either T-rich if the sequence is high in unmethylated C contents, or G/T-rich if the sequence is highly methylated.

Unlike bisulfite conversion, restriction enzymes and affinity enrichment based methods cannot provide single base resolution on CpG sites. However, they tend to be more cost effective and some library loss can occur during bisulfite conversion from fragmentation of the sample as well as reduced sequencing complexity from the converted bases. Nonetheless, the ability to identify individual CpG sites and compatibility with next generation sequencing (NGS) makes bisulfite conversion a commonly used approach.

2.3.2 Post bisulfite analyses

Bisulfite conversion can be combined with many different techniques to extract information on the methylated states [94]. The following sections examine some of the commonly used methods.

Bisulfite-Sequencing (BiSeq)

There is keen interest in developing sequencing based methylation analysis due to its digital compatibility, which allows for integration of data from different runs and studies. Sequencing can be directly from bisulfite PCR products or clones. Although clonal sequencing can provide single molecule information, it is very labour intensive and time consuming. Sanger sequencing is classically used to carry out analysis as the first sequencing technique to be developed [96]. However, quantification of methylation level using Sanger sequencing is immensely difficult and typically requires cloning.

---

8 Also recently known as Second generation sequencing

9 DNA fragments terminated at different lengths are extended using single type nucleotides called dNTPs. Electrophoresis is used to separate the strands by size. Since the dNTPs were labeled with different coloured fluorescence, an optical detector reads the terminating base and the resultant sequence can be assembled according to size.
particularly if the sites are partially methylated [97]. Therefore it is often used instead to validate results generated by other sequencing methods for when there is severe amplification bias\textsuperscript{10} [99].

Second generation sequencing [100] have emerged as prominent techniques for di-

\textbf{Figure 2.7:} Second generation sequencing techniques: First common adaptors are used to select fragments of DNA, then amplified with Bridge or Emulsion PCR. Sequencing by synthesis detects nucleotide incorporations by identifying different coloured fluorescent tags; Pyrosequencing captures light released during extension; Ion semiconductor detects hydrogen ions and Sequencing by ligation uses a library of probes labeled with different tags according to the first and second base [95]

\textsuperscript{10}PCR is found to favour amplification of unmethylated templates when using methylation independent primers [98]
The role of Epigenetics in chronic disease

rect quantitative sequencing due to their reproducibility and ease of use. Efficiency of massively parallel and high throughput systems have allowed for great reduction in sequencing time and costs. Prior to sequencing, a library of templates has to be prepared. Firstly, common adapters are used to align the fragmented DNA. Depending on the sequencing method, Bridge or Emulsion PCR are used to carry out amplification. In Bridge PCR, primers (See Appendix B for further details on primers) from both ends are attached to a surface, resulting in clusters of amplicons generated from DNA extensions chained to the surface. For Emulsion PCR, only one primer is bound to a bead, and subsequent amplification occurs on any strands that have been captured [95].

There are four main types of second generation sequencing [101]: pyrosequencing [97], sequencing by synthesis, sequencing by ligation and ion semiconductor sequencing [102]. Pyrosequencing relies on the release of pyrophosphate (PPi) to detect DNA extension. When a nucleotide is incorporated into the sequence, PPi is released and converted through a series of chemical reactions into visible light, which is then detected by a camera. One type of dNTP is added at one time, therefore the intensity of light corresponds to the number of bases incorporated. In sequencing by synthesis, all four nucleotides labeled with different colour fluorescence are added but only a single dNTP can be incorporated and detected at one time. Sequencing by ligation uses a set of probes consisting of two known bases followed by six universal/degenerate bases. The probes are labeled accordingly with four different fluorescence and the specific di-bases. The enzyme ligase binds the probe with matched bases to the sequence and the template is imaged. The last three bases and fluorophore at the end of the probe is then cleaved and the process is repeated for 7-10 cycles. After the series of probe extensions, a shift in one base is applied to the primer used for anchoring and the whole process begins again. Generally a template’s sequence can be determined with five offset primers. Finally ion semiconductor sequencing uses direction detection of hydrogen ions released during DNA extension for base calling (discussed in more details in Chapter 3).

**Mass spectrometric based**

Mass spectrometric based methods can provide quantitative information at multiple CpG sites, subject to areas of fragmentation [103, 104]. In vitro transcription is carried on bisulfite PCR products, and the resultant RNA are cleaved at the uracils. The RNA fragments of different size and mass are then quantitatively read by MALDI-TOF MS\(^\text{11}\) to evaluate methylation level. Epityper is reported to provide the highest sample throughput when compared with other NGS sequencers (MiSeq and Infinium),

\(^\text{11}\)Matrix-assisted Laser Desorption/Ionization - Time of Flight Mass Spectrometry
with slightly less accuracy [105].

**High-resolution melting analysis**

MS-HRM, methylation sensitive high-resolution melting can be used to quantify methylation template relatively [106, 107]. The melting temperature is defined as a specific temperature at which the double helix structure is denatured into two single strands. Depending on the DNA length and contents, each would have a characteristic melting profile. During PCR, fluorescent molecules are intercalated and emits light when bound to double stranded DNA. A melting analysis is performed at the end of a PCR reaction and the temperature is raised continuously. At melting temperature, the DNA strands separate and the dye is released, resulting in a large drop in fluorescence. In MS-HRM, the melting temperature of the bisulfite converted templates is compared with a dilution series of known methylated and unmethylated DNA concentrations. A sequence with high G:C content (methylated) will melt at a higher temperature than unmethylated amplicons. Percentage methylation on the overall CpG contents can be deduced relatively.

**COBRA and MS-SNuPE**

Other non-selective amplification based techniques include COBRA (combined bisulfite restriction analysis) [108] and MS-SNuPE (methylation sensitive single nucleotide primer extension) [109]. COBRA uses restriction enzymes to cleave bisulfite converted DNA into fragments containing CpG sites. Electrophoresis can then be applied to quantify methylation levels for each fragment size. Ms-SNuPE requires extensive and complex primer designs for each CpG site. The oligonucleotides\(^{12}\) are made to terminate at cytosines, after which a radiolabel is added. The templates terminating at different CpG sites can be read by a phosphorimager. Both methods are able to perform relative quantification on methylation, but not all CpG sites are necessarily available from enzyme cleavage or primers.

**Methylation-specific PCR**

Methylation specific PCR can provide fast qualitative analysis on CpG sites based on selective amplification [110]. Two sets of PCR primers are designed to anneal to the methylated and unmethylated sequence of the same CpG locus. The subsequent amplification indicates whether the sequence is methylated or unmethylated. The analysis is site specific and restricted to single or at most two CpG sites included in the primers.

\(^{12}\)specifically designed DNA sequences
Real time PCR can be combined with MSP to provide relative quantification on the specific CpG site [111]. In particular, Methylight can provide quantitative MSP analysis by introducing additional probes to ensure coverage of CpG sites not included in the primers and all possible methylation combinations [112].

Certainly, each method has its advantages and disadvantages. Table 2.3 summarises the aforementioned techniques compatible with bisulfite conversion. Evidently, sequencing is the only one that can provide quantitative analysis at single base resolution after bisulfite treatment. Other than considerations to the type of analysis, choice of detection method will often depend on readily available equipment and expertise. For example, PCR machines are typically included in molecular biology laboratories, and melting curve analysis are carried out routinely to check specificity of amplicons, therefore MSP and MS-HRM may be easier to implement. For clinical applications, time and costs are a priority, as well as the handling of dangerous reagents. Any methods based on electrophoresis which requires processing over days and incorporation of radionucleotides are not well suitable to clinical practice.
Table 2.3: Table of bisulfite conversion based methods for detecting methylation

<table>
<thead>
<tr>
<th>Technique</th>
<th>Type of analysis</th>
<th>Principles</th>
<th>Advantages and Disadvantages</th>
<th>Starting Quantity</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisulfite-Sequencing [97]</td>
<td>Absolute</td>
<td>Any second generation sequencing technique</td>
<td>✓ Single base resolution with exact location of CpG site, high specificity, high throughput</td>
<td>ng-µg</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>× High cost, time consuming, computationally intensive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COBRA [108]</td>
<td>Relative</td>
<td>Restriction enzymes and electrophoresis of fragments</td>
<td>✓ Specific regions selected by restriction enzymes</td>
<td>ng-µg</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>× False positive from incomplete digestion, cannot detect multiple CpG sites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass Spectrometry-based [103]</td>
<td>Absolute</td>
<td>In vitro transcription to generate RNA, subsequently cleaved fragments quantified by mass spectrometer</td>
<td>✓ High throughput</td>
<td>ng</td>
<td>High</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>× Expensive, require specialised instrumentation and knowledge</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS-HRM [106]</td>
<td>Relative</td>
<td>Methylation-independent amplification followed by melting curve analysis</td>
<td>✓ Simple overall estimation of methylation</td>
<td>ng</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>× Low sensitivity or becomes qualitative if sample has varying methylation levels (heterogeneous)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS-SNuPE [109]</td>
<td>Relative</td>
<td>Primer extension at every CpG site, after which a single radiomucleotide is introduced and detected</td>
<td>✓ Fast, small amount of DNA required</td>
<td>&lt;ng</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>× Handling of radioisotopes, Only single CpG site analysed, labour intensive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSP [110]</td>
<td>Qualitative</td>
<td>Selective amplification with methylated and unmethylated primers</td>
<td>✓ High sensitivity, high specificity (for target CpG site), fast, low amount of DNA required</td>
<td>&lt;ng</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>× Not quantitative, false positives from incomplete conversion, primer design</td>
<td></td>
<td></td>
</tr>
<tr>
<td>qMSP (Methy-light) [111]</td>
<td>Relative</td>
<td>MSP by real time PCR (with methylation-specific probe)</td>
<td>✓ High throughput, fast, sensitive, reduced false positives, small amount of DNA required</td>
<td>ng</td>
<td>Low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>× Primers optimisation, cost increases (compared to MSP)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations in order of appearance: COBRA, Combined Bisulfite Recognition Analysis; MS-HRM, Methylation-Specific High Resolution Melting analysis; MS-SNuPE, Methylation-Sensitive Single Nucleotide Primer Extension; MSP, Methylation-Specific PCR; qMSP, quantitative Methylation-Specific PCR.
2.3.3 Bisulphite free methods

Direction detection of methylation without chemical preparations is possible with Third generation sequencing techniques based on single molecules [100]. Sequencing from a single DNA template avoids bias and artifacts that arise from amplification steps. In addition, methylated cytosines can be identified based on signal variations without prior methylation sensitive treatments. Oxford Nanopore Technology and Single Molecule Real Time sequencing (SMRT)\(^{13}\) are the two most prominent commercialised technologies in the area.

![Single Molecule Real Time sequencing](image1)
![Nanopore sequencing](image2)

Figure 2.8: Third generation sequencing methods

SMRT has two main features that enables sequencing on single molecules: phospholinked nucleotides and Zero mode waveguides (ZMW) [114]. Zero mode waveguides are cylindrical chambers that allow for very small volume detection. Light illuminating from at the bottom of the ZMW cannot penetrate through due to its wavelength. Acting like a microscope, the attenuated light is contained to only nm regions of 100\(\mu m\) depth well, creating an extremely small focal volume. Nucleotides are introduced to a single strand of DNA immobilised at bottom of the ZMW, then DNA polymerase cleaves away the fluorophore after each incorporation, releasing fluorescence for detection. Thousands of ZMWs are placed in parallel with DNA replications occurring simultaneously in real time. Methylated cytosines are detected based on polymerase kinetics. Reactions occurring at a speed that differs from the expected incorporation rate implies the presence of methylation.

Nanopore technology sequences single molecule DNA based on disruptions of ion-currents [116]. The nanopore is placed in a synthetic polymer membrane with a very high resistance and potential is applied. Restricted by motor protein, single stranded

\(^{13}\)Helicos and Pacific Bioscience
DNA passes through the nanopore one base at a time, and the type of base can be distinguished from the ion-current signal. Similarly, 5-methyl cytosines have distinct characteristics in the ion-current, allowing for direct detection.

Third generation sequencing is emerging as a promising field, particularly for direct detection of methylation. In addition, Oxford Nanopore technology can achieve very long reads up to 1000bp [117]. Run times for each technology are also very fast, in the ranges of 100Mbp per hour. In comparison, NGS technologies can reach up to 30Mbp per hour, with the exception of ion semiconductor sequencing which can achieve similar speeds to single molecule based methods. However, third generation sequencing are still at a stage of development and generated much higher error rates when compared with other platforms [118]. Nonetheless, its ability to detect methyl-cytosines directly along with other intermediates of methylation such as hydroxymethyl-cytosines would prove to be valuable in epigenetics research.

2.4 Quantifying miRNAs

As demonstrated in Section 2.2.3, measuring miRNA levels can indicate disease states via gene expressions. While miRNA are found to be exceptionally well-preserved in body fluid specimens such as plasma, serum and urine, its potential as a biomarker is subject to challenges associated with its profiling. Occupying only 0.01% of the total RNA mass [23], it is essential to selectively detect or isolate miRNAs from the sample. Due to the short length of miRNAs, conventional primers for synthesising cDNA\textsuperscript{14} and PCR cannot be used. And unlike messenger RNAs, they do not contain a poly-A tail\textsuperscript{15} that can allow for universal primer binding onto a common sequence. Any profiling techniques must also be able to distinguish mature miRNAs from the longer sequences (pri-miRNA and pre-miRNA) in which it is cleaved from during biogenesis. In addition, heterogeneity vary between different miRNAs, with some miRNAs from the same family (such as let-7), differing only by one base [119].

Normalization

Normalization of data is essential for accurate measurements of miRNA expressions [54]. Quantification errors can arise from sample collection, preparation and varying chemical or enzymatic reaction efficiencies. The aim of normalisation is to modify the data to ensure any variations in expression levels are not due to bias but biological differences. There are three main approaches for normalisation [120, 121]. The use of endogenous

\textsuperscript{14}complementary DNA, synthesised from RNA counterparts
\textsuperscript{15}long chain of adenines
miRNA is one of the most common approach for normalisation. A housekeeping miRNA that is abundantly expressed and not affected by the studied condition is selected as a control. Choosing an appropriate and stably expressed housekeeping miRNA is non-trivial, often two or more are used for proper normalisation. Similarly, exogenous, or synthetic miRNAs that are not normally expected in the sample can be added as a reference. Such external spike in miRNAs are well-suited to normalising circulating samples such as serum and plasma by monitoring differences in extraction yields [122]. For example, the addition of a plant-based miRNA would be a good exogenous control when studying samples from human. Alternatively, a global mean expression value of all the miRNAs can be used to correct differences of specific miRNA in samples, since the pattern of expression can be expected to be the same. However this requires large scale profiling of all the miRNAs.

**Northern blot**

Northern blot is one of the most established methods for detecting miRNA [123]. Named after the first blotting technique, Southern blot for DNA analysis, Northern blot is widely used to quantify RNA [124]. After miRNA are isolated from the sample, it is fractionated by electrophoresis on a high percentage\(^{16}\) gel that allows for separation of small molecules. The “blotting” refers to the transfer of miRNA from gel to membrane. The molecules are then fixed to the membrane via UV crosslinking, or baking at around 80\(^{\circ}\)C. Radiolabeled or non-isotopically labeled probes are hybridised to the miRNA and the blot is exposed onto film using autoradiography. Northern blots can be used to detect mature as well as primary and precursor miRNAs since electrophoresis separate molecules by size. Although the technique does not require synthesis of cDNA, the whole process requires at least 2 days. With the exception of locked nucleic acid\(^{17}\) (LNA) probes [125], conventional probes have low specificity that need at least 3 unmatched bases to distinguish between sequences [23].

**In situ hybridisation (ISH)**

In situ hybridisation (ISH) is used to analyse miRNA on a cellular level [120]. ISH is performed directly in the tissues of interest to assess the location and levels of expressions. The process involves a predigestion step, where the cells are cleaved by protease\(^{18}\) to provide access to the miRNAs. Then LNA or standard probes anneal to the target

---

\(^{16}\)percentage of agarose gel is inversely proportional to the size of pores  
\(^{17}\)specially modified nucleic acid with high affinity and thermal stability  
\(^{18}\)protein digesting enzymes
miRNAs and results are imaged by autoradiography or immunohistochemistry\textsuperscript{19}. In situ hybridisation is crucial in the understanding of biological processes. It is a well-established technique, but detection in miRNA poses some challenges. ISH can frequently fail to set up due to errors arising from RNase contaminations and optimising hybridisation conditions in terms of temperature, buffer concentrations etc. Quantification of miRNA is restricted to measuring relative abundance based on intensity.

2.4.1 Post Reverse transcription

Unlike Northern blotting and ISH, most methods require reverse transcription (RT) to generate cDNA of the target miRNA for amplification or hybridisation based analyses. Specially designed primers are required to reverse transcribe miRNA due to its short length, namely stem loop \textsuperscript{126} or universal poly-A tail primers \textsuperscript{127}. Stem loop primers are used to reverse transcribe specific miRNAs via its short single strand region that is complementary to the 3' end\textsuperscript{20} of the target. cDNA is extended with dNTPs using reverse transcriptase. For amplification, forward primers bind to the 5' end\textsuperscript{20} of the cDNA sequence, and the loop structure unfolds after denaturing, allowing a reverse

\begin{figure}
\centering
\includegraphics[width=\textwidth]{reverse_transcription_diagram.png}
\caption{Reverse transcription: stem loop primer contains an overhang structure that is complementary to the target miRNA, after annealing, the hairpin loop unfolds when denatured, allowing primers to bind for amplification (left) \textsuperscript{126}; Reverse transcription by poly-A tail starts with attachment of miRNA, a universal primer is used to bind to the poly-A tail, subsequent amplification can be carried out with miRNA-specific primers (right) \textsuperscript{127}}
\end{figure}

\textsuperscript{19} detection of labeled antibodies
\textsuperscript{20} refers to hydroxyl group’s positioning in relation to carbon of a sugar ring, as an indicator of direction
primer to bind. Although difficult to design, stem loop primers provide high specificity by reducing tenancies to anneal to primary and precursor miRNA [128]. When analysing many different miRNAs, poly-A tails can be attached to the 3' end of miRNAs to provide a common sequence. Reverse transcription can then be carried out using a universal primer. Evidently, poly-A tails have the advantage of transcribing multiple miRNAs, including any normalisation controls, though at a higher cost than stem loop primers [23]. Since reverse transcription by poly-A tails do not discriminate different miRNAs, the required targets should be selected later during amplification step with specific primers.

RNA sequencing

Reverse transcription can be used to prepare a library of cDNA for RNA sequencing. Any massive parallel sequencing technique mentioned in Section 2.3.2 can be applied to profile miRNAs. Notably, profiling by sequencing has advantages of detecting both known and novel miRNAs with high accuracies, at the expense of computational power. Requiring a substantial amount of miRNA, quantification by sequencing is only relative, where the number of reads are compared with the total transcripts. In addition, normalisation techniques are not well-established in RNA-sequencing due to the reduction of available reads for the target miRNA in the presence of highly expressed controls [119].

Microarray

Based on hybridisation of probes, microarrays can detect many different miRNAs in parallel [129, 130]. Probes made from reverse transcribed cDNA are attached to a microarray and fluorescently labeled miRNA\(^{21}\) are introduced. Molecules complementary to the DNA-based probes will bind onto the microarray, the assay is then washed and scanned with an imager. Expression levels are correlated to the intensity of fluorescence emissions, with a limited linear range. Nanostring nCounter is a variation of hybridisation based technology that can directly detect the number of molecules digitally [119]. Using a library of florescent probes, every molecule captured can then be counted as each unique miRNA sequence has a corresponding “barcode”. With the exception of Nanostring nCounter, microarrays are more suited to comparing different expression levels between miRNAs due to its inability to carry out absolute quantification (see Section 2.4 for more details).

\(^{21}\)Typically by enzymatic ligation of fluorophore-conjugated nucleotides to the 3' end [119]
Real time quantitative PCR (qPCR)

Reverse transcription combined with real time quantitative PCR (qPCR) is considered the gold standard for quantifying RNAs [131]. Traditional endpoint PCR visualise the finished amplification products on agarose gel by electrophoresis, to detect the presence or absence of amplicons with semiquantitative properties [90]. In comparison, real time PCR monitors the products generated at the end of every thermal cycle. Amplification by thermal cycling consists of 3 main steps: nucleic acids are denatured into two single strands at a temperature of 90-96°C; the temperature is then lowered to 50-65°C for annealing PCR forward and reverse primers to 5’ and 3’ end of the target gene respectively; finally the sequences are extended with dNTPs and DNA polymerase to form 2 new DNA copies at 75-80°C. Reverse transcription can be performed in one step [132] or prior to amplification separately. In one step RT-qPCR, only one mastermix is required but RT primers must be specific and conditions should be optimised for both reverse transcription and amplification.

SYBR green II

Taqman probe

Figure 2.10: Fluorescent labeling of DNA by: SYBR green II dyes are trapped into double stranded DNA during extension and releases fluorescence accordingly (left); Taqman probes containing reporter and quencher dyes anneal to the target sequence, during DNA extension the probe is cleaved by DNA polymerase, releasing the reporter fluorophore (right)

Generally there are two main methods of fluorescent labeling in qPCR, namely Taqman and SYBR green [133]. SYBR green II is an intercalating dye that emits light when bound to double stranded DNA. Therefore detection of SYBR green occurs at the end of elongation. Since SYBR green dyes can bind non-specifically, melting curves analysis are often carried out at the end of PCR reaction to confirm amplification of the target sequence. Subsequently, melting analysis are based on the large drop of fluorescence
when SYBR dyes are released when double stranded DNA denatures at its melting temperature. Taqman uses special probes with reporter and quencher fluorophores to anneal to amplicons. Due to their proximity, the quencher inhibits fluorescent emissions from the reporter dye by absorption. During extension, the exonuclease activity of Taq DNA polymerase displaces and cleaves the probe, releasing the reporter’s fluorescence. Taqman probes are designed to be specific to the target sequence, enabling multiplexing of different DNA amplifications in the same run. SYBR green tends to be more widely used due to its low cost and availability.

2.4.2 Absolute, Relative and Digital quantification

Quantification in real time PCR is based on the evaluation of threshold cycles [124]. In a PCR run, the reaction goes through three phases: exponential, linear and plateau as shown in Fig. 2.11. In the exponential phase, there is doubling of amplicons at every cycle as each double stranded DNA is denatured and elongated to two new copies. The reaction will continue until no primers and reagents are available, slowing to a linear phase and eventually plateau. The growth rate in the linear phase is highly variable, and a direct correlation to cycle number can only be defined in the exponential phase. A threshold cycle ($C_T$), or crossing point ($C_P$) is used to evaluate expressions by defining a cycle number in the exponential period above background noise, thus $C_T$ is inversely proportional to the number of copies in the sample. A low $C_T$ implies a high concentration of starting DNA while a high $C_T$ implies a low number of copies as the reaction takes longer to reach exponential growth.

![Figure 2.11: A typical qPCR output signal with threshold cycle ($C_T$)](image)

---

$^{22}$Cleavage of nucleotide at the end of a chain by hydrolysis
There are several methods to determining the threshold cycle [134]. One of the most well known techniques is the Fit point method, which chooses the crossing point according to the log-linear phase of the reaction. Considering that an exponential curve plotted in a logarithmic scale is linear, $C_T$ can be determined as intersection points between a parallel line drawn to the x-axis and qPCR signal as shown in Fig. 2.12 (left). The Derivative method, as the name suggests, calculates the second order derivative of the real time PCR signal and sets the $C_T$ as the cycle in which maximum occurs. Since the second derivative describes the rate at which the reaction is changing, choosing the maximum point in the curve ensures the threshold cycle lies within the exponential phase. Many PCR machines nowadays adapt a technique that measures the baseline (typically an average of cycles 3 to 15) and sets the threshold fluorescence to be a number of standard deviations above the baseline [135]. It is also not uncommon to allow users to manually set their own background cycles or thresholds. As such, $C_T$s determined by Fit point or baseline method are highly subjective and variable depending on users.

![Figure 2.12: Methods to determine $C_T$: Fit point method (left); Derivative method (right) [134]](image_url)

Depending on the application, two types of quantification can be used [136]. Absolute quantification estimates the concentration of nucleic acid in a sample interpolated from a standard curve. A template of known quantity diluted at various concentrations, typically at 10-fold dilutions, is used to generate the standard curve as shown in Fig. 2.13. The threshold cycles of the dilution series is plotted against the logarithm of the dilution factors with a line of best fit. The copy number in the unknown sample can be evaluated by plotting its $C_T$ value on the same graph. Since the equation for a regression line is $y = mx + b$, where $m$ is the slope and $b$ is the y-intercept, the copy
number (y) can be derived by rearranging for x and considering the logarithmic scale using the following equation:

\[ \text{quantity} = 10^{\frac{C_T-b}{m}} \]  

(2.2)

Quality of the standard curve affects the precision and accuracy of quantification. Linearity outside of the standard curve is undefined, therefore the dynamic range ideally should be as broad as possible, starting from 1 molecule per sample up to 9 orders of magnitude. The number of copies is defined with respect to the sample, such as copies per ml blood or copies per cell and the volume specified act as a normaliser. Absolute quantification is useful for doctors when determining viral particles in a patient’s sample. Generation of a standard curve is a labourious procedure, and require stable templates that do not degrade over time. Errors can arise if conditions between the standard curve and sample amplifications differ.

![Figure 2.13: Standard curve of threshold cycles from different known dilutions of DNA copies plotted on a logarithmic scale](image)

The efficiency of a qPCR reaction can also be derived from the slope (m) of a standard curve [137]. The formula based on 10-fold dilutions is shown in eq. 2.3. A fully efficient reaction corresponds to a slope of -3.3, which translates to the doubling of DNA copies. Amplification rate can be affected by many factors, causing variations in steepness, background levels or plateau value in the qPCR curve. Certain inhibitors or enhancers that prevent or accelerate amplifications can originate from different sources like blood, tissue or hair etc. They interfere with the reaction by direct binding or impacting polymerase activities. Careful handling of samples can minimise bias, though some are inherent to the source and requires purification.
\[ Efficiency = 10^{-\frac{1}{m}} - 1 \] (2.3)

Relative quantification is commonly used to investigate the relative differences, also known as fold difference in expressions between healthy and disease samples [138]. A ratio (R) representing the difference in \( C_T \)s between test and control samples are normalised to a reference gene. Assuming the efficiency of the reactions are 100%, Livak [139] defined the ratio as:

\[
\Delta C_T^{(test)} = C_T^{(target,test)} - C_T^{(reference,test)} \quad (2.4)
\]

\[
\Delta C_T^{(control)} = C_T^{(target,control)} - C_T^{(reference,control)} \quad (2.5)
\]

\[
\Delta \Delta C_T = \Delta C_T^{(test)} - \Delta C_T^{(control)} \quad (2.6)
\]

\[
R = 2^{-\Delta \Delta C_T} \quad (2.7)
\]

Also known as the “delta delta” method. A variation of the formula can be used to evaluate normalised expressions within the same sample:

\[
R = C_T^{(target)} - C_T^{(reference)} \quad (2.8)
\]

Reactions with significantly different or suboptimal efficiencies can falsely estimate copy number. Pfaffl [140] proposes an efficiency correction model that takes into account the amplification rate of each reaction:

\[
Ratio = \frac{(E_{target})^{\Delta C_T^{target}(control-test)}}{(E_{reference})^{\Delta C_T^{reference}(control-test)}} \quad (2.9)
\]

Clearly, when efficiency is 100%, eq. 2.9 can be written as Livak’s equation.

**Digital PCR (dPCR)**

Digital PCR (dPCR) can provide absolute quantification without the need of a standard curve [141]. Initially reported in 1990, its application was not widespread until the new developments of micro-fluidics in recent years [142]. The method is based on separating the sample into many small volume partitions for individual endpoint PCR reactions. The presence and absence of templates in each partition is then used to calculate the total copies in the sample. In commercially available technologies like droplet
digital PCR (ddPCR) [143], nano-litre droplets are generated with a water-oil emulsion cartridge. Errors due to deviations from one template per partition is corrected using statistical analysis based on Poisson distribution. Digital PCR has high sensitivity and tolerance to inhibitors as the effects are reduced by the use of tiny partitions. The increase in volume of sample allows for detection of rare target templates, on the other hand, careful design of dilutions is required to avoid many partitions with multiple copies if an abundance of templates is expected.

**quantitative LAMP (Loop-mediated Isothermal Amplification)**

![Diagram of LAMP](image)

Figure 2.14: Loop-mediated Isothermal Amplification (LAMP): **a.** 6 primers on each DNA strand; **b.** Inner primers (FIP, BIP) bind to multiple site and outer primers (F2, B2) binds to one site and displaces the strand, self hybridising loops form a dumbbell structure from the singled stranded DNA; **c.** Dumbell structure provides multiple sites for amplification, generating various regions of the original sequence [144]

An alternative amplification method, Loop-mediated Isothermal Amplification (LAMP) has gained increasing attention in diagnostic applications in the recent years [145]. Since its first development by Notomi et al. in 2000 [146], there has been substantial literature reporting its use in detecting virus, pathogens in infectious diseases or food etc. Amplification in LAMP is carried out at constant temperature (typically 60-65 °C),
eliminating the need for a thermal cycling machine. Six primers are designed to anneal to six regions of the target sequence, promoting high specificity amplifications. For example regions found on a single strand, F1, F2, F3, B1c, B2c, B3c, correspond to forward and backward ends of the template with ‘c’ to indicate complementary sequences. Inner primers FIP, BIP bind to multiple sites and outer primers F3 and B3 bind only to the corresponding sites. The process can be described to have two main phases, a starting structure producing step and cyclic amplification phase. After denaturation, forward inner primers bind to F2c site in the 3’ end of the template, and polymerase extends the sequence with dNTPs. Primer F3 then binds upstream and displaces the newly formed strand. The displaced single stranded DNA forms a self-hybridising loop while the backward inner primer repeats a similar bind and displacement reaction at its 3’ end. The result is a LAMP dumbbell structure that is used in cyclic amplification. The dumbbell structure acts as a seed sequence with many initiation sites for inner primers to extend from both ends, generating double stranded DNA and therefore more binding sites. Additional loop primers can also be used to further accelerate the reaction [147]. The result is a rapid multiplication of templates that contain different number of sequence repeats. The amplicons can be detected by incorporating fluorescence labeling such as SYBR green dyes or visual turbidity due to LAMP byproducts [148].

Like qPCR, LAMP can also be monitored in real time to provide relative or absolute quantitative analysis [149]. “Time to threshold” is often used in place of cycles since the reaction is carried out at constant temperature. Since there are no formulas estimating copy number with time due to difficulties in modeling the reaction kinetics, both relative and absolute quantification utilise a reference gene and standard curve. Nonetheless, LAMP with reverse transcription provides a cost effective solution that enables easy implementation of miRNA profiling outside of laboratory.
### Table 2.4: Table summarising different methods for quantifying miRNA [150]

<table>
<thead>
<tr>
<th>Method</th>
<th>Principle</th>
<th>Advantages and Disadvantages</th>
<th>Starting quantity</th>
<th>Time</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>In situ hybridisation [120]</td>
<td>Detection of labeled antibodies/probes inside cells/tissues</td>
<td>✓Can determine distribution of miRNA, does not require reverse transcription&lt;br&gt;×Low sensitivity &amp; specificity, can only detect high abundance</td>
<td>N/A</td>
<td>1 day</td>
<td>High</td>
</tr>
<tr>
<td>Microarray [129] [130]</td>
<td>Hybridisation of miRNA to reverse transcription attached to NAs simultaneously&lt;br&gt;Array</td>
<td>✓High throughput, can screen many different miRNAs simultaneously&lt;br&gt;×Restricted linear range of quantification, low specificity</td>
<td>ng-µg</td>
<td>1 - 2 days</td>
<td>Low</td>
</tr>
<tr>
<td>Northern blot [123]</td>
<td>Electrophoresis of miRNA then transferred to membrane</td>
<td>✓Can detect both mature and precursor miRNA easily, low cost&lt;br&gt;×Labour intensive, low throughput and sensitivity, time consuming</td>
<td>&gt; µg</td>
<td>&gt; 2 days</td>
<td>Low</td>
</tr>
<tr>
<td>RNA-sequencing [119]</td>
<td>Reverse transcription followed by any sequencing method</td>
<td>✓Can profile novel and known miRNA, high throughput and accuracy&lt;br&gt;×High costs, no established normalisation technique, computation complexity</td>
<td>ng-µg</td>
<td>2-7 days</td>
<td>High</td>
</tr>
<tr>
<td>RT-quantitative PCR [131]</td>
<td>Reverse transcription followed by real time PCR (detection at every thermal cycle)&lt;br&gt;Reference cycle</td>
<td>✓Can carry out absolute quantification, normalisation&lt;br&gt;×Low throughput for genome wide profiling</td>
<td>&lt;ng or ng-µg</td>
<td>6 Hours - 1 day</td>
<td>Low</td>
</tr>
</tbody>
</table>
2.5 Summary

This chapter introduces epigenetics and its vital role in chronic disease developments, namely chronic kidney disease. Gene expressions can be affected by epigenetic mechanisms such as methylation, the addition of a methyl group to cytosines and number of miRNAs through transcription and translation interference respectively. A close examination on CKD demonstrated the need for new biomarkers and treatments. The disease’s development is closely linked to regulations of gene expressions, as illustrated in the pathology of renal fibrosis. A review of current literature based on methylation and miRNA’s influences on the disease’s etiology and progression was presented, including their subsequent therapeutics potential.

Some commonly used techniques for methylation and miRNA analysis were summarised. Sample preparation to conserve methylation contents is essential prior to any analyses, with the exception of third generation sequencing techniques. While three main preparation methods are available, namely restriction enzymes, affinity enrichment and bisulfite conversion, the use of sodium bisulfite is typically preferred as it maintains information at every CpG site. The treatment converts any unmethylated cytosines to uracils and can be coupled with a variety of techniques for analysis. There are merits to each method, though ones that can be applied with readily available equipment and expertise are more favourable. Particularly, Methylation-specific PCR can provide fast detection with high sensitivity and specificity based on amplification results of methylation-dependent primers. Quantitative analysis at single base resolution must be carried out by sequencing bisulfite converted templates. Similarly, majority of miRNA detection requires reverse transcription to obtain cDNA of adequate length for primers to anneal. Real time quantitative PCR is a well-established technique for assessing expression levels, capable of both absolute and relative quantification. Evaluation of threshold cycle is essential for quantification analysis, from which various equations can be used to compute concentration based on standard dilutions or relative to other genes.


abstract{\%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/8383116{\%5Cnhttp://www.nature.com/doifinder/10.1038/ncomms7716{\%5Cnhttp://www.nature.com/doifinder/10.1038/ncb3110{\%5Cn


Chapter 3

Ion-semiconductors for DNA analysis

From the previous chapter, it is evident that many existing DNA detection techniques can be adapted to analyse epigenetic modifications. Additionally, the trend to move towards solid state detection is demonstrated by the developments of ion-semiconductor sequencing [1] and solid state nanopores [2]. Such techniques are revolutionary in an area traditionally dominated by optics. Not to mention the advantages of costs and scaling when manufacturing in CMOS technology.

In this chapter we focus solely on DNA detection using semiconductors. In particular, looking at label-free techniques to prevent additional sample preparations on top of the necessary treatments for detecting methylation and miRNA. With ISFETs as our primary candidate for electronic detection, the fundamentals of its operation and effects of fabricating in unmodified process are examined. The methods in which ISFETs can be translated to detect epigenetic modifications are presented through a review of existing literature.
3.1 Label free electronic DNA detection

DNA hybridisation due to base pairing is a robust and specific reaction well suited to bio-sensing. In general, biosensors consist of a recognition layer which is sensitive to the target molecules, and a transducer for converting the signal to the same domain as the readout method. Over the years, a range of biosensors based on different methods of analysis have been developed, including optical, thermal, mass and many more [3]. Whilst each technique has its advantages, the need to convert final readout signals to electrical often requires the design and fabrication of specialised instruments. Electrochemical sensors on the other hand can detect DNA hybridisations directly based on incorporations or RedOx\(^1\) reactions from labeled molecules. Also, fabrication in an already mature technology such as CMOS dramatically lowers costs. With reduction in cost, analysis time and size, genetic and epigenetic testing can potentially be included in routine diagnostic check ups where high accuracy and dynamic range are secondary requirements.

Most biosensors require labeling of the target molecule in order to read out the results of the reaction. The incorporation of enzymes or fluorescent tags is a cumbersome procedure that is expensive and can affect binding properties and yield. There are many incentives for label free DNA detection, primarily minimising sample handling, which would greatly benefit Point-of-Care applications. Label free reactions are also more likely to allow for real time detection of the DNA binding for improving measurement accuracies [4].

In the following sections, we examine a variety of semiconductor-based sensors for label-free DNA detection. The different analytical principles can be summarised as voltammetry, potentiometry and conductometry [5]. Typically arranged as a three-electrode system, voltammetric systems observes current as a function of variable potential. The reference electrode is used to provide a well-defined voltage, while the counter electrode supplies current which is measured at the working (or indicator) electrode. Amperometry is a type of voltammetry where the potential is kept constant for detecting specific molecules. Conductometry, as the name suggests measures conductance of an analyte. An alternating potential is applied to a pair of inert electrodes in which the resistance can be deduced from the resultant current. In potentiometry, the voltage difference is measured between two electrodes, typically an ion-sensitive electrode (ISE) and a reference electrode, with no current flowing through the analyte. Of the three techniques, potentiometry is the most widely used in electrochemical sensors.

\(^1\)Reduction (gain of electron) and Oxidation (loss of electron)
3.1.1 Impedance and Dielectric Spectroscopy

Electrochemical Impedance Spectroscopy (EIS) is a powerful tool for characterising electrochemical systems using two or three electrodes [6]. In general, a small time varying signal such as a sinusoidal voltage is applied to the electrolyte, and the subsequent current reflecting the state of the cell is measured. Therefore the impedance can be obtained by considering the magnitude and phase of the applied voltage and measured current. The results can be represented in the frequency domain or with real and imaginary components of impedance. The use of an AC signal has the advantage of generating more information about the system than DC. Equivalent circuit models with passive components like resistors and capacitors are often used to represent the electrochemical interactions.

For detecting DNA hybridisations, oligonucleotides containing complementary sequence of the target are immobilised to one of the electrodes in an EIS system. In [7], the gates of a FET device array were treated with DNA probes that have different number of nucleotide mismatches to the target. As a result, corner frequency of the circuit decreases with complementary probes. Based on this, analysis of SNP\textsuperscript{2} mutation was done by differential transfer function of the sensor before and after hybridisation. A sinusoidal voltage of varying frequencies is applied through the reference electrode to the solution and the theoretical transfer function \( H(j\omega) \) was described as:

\[
H(j\omega) = \frac{1 + j\omega R_{DNA}C_{DNA}}{1 + j\omega R_{DNA}(C_{DNA} + C_{ox})}
\]

Where \( C_{ox} \) is the device’s gate oxide capacitance, and the layer of DNA is modeled by a resistor \( R_{DNA} \) and capacitor \( C_{DNA} \). Clearly, the time constants of the pole and zero gives information on the impedance and capacitance of the DNA layer. According to the paper, such electrochemical system has low pass filtering characteristics. The values extrapolated from the measured frequency response curves resulted in an increase of layer resistance after DNA hybridisation of matched probes. The higher observed time constant can be elucidated as an increase in density of molecules on the surface. A drop in relative capacitance was also observed in [8], who reported a 20% reduction in diffusion capacitance.

Charge based measurements for detecting DNA hybridisations have also been reported in several publications [6, 9–11]. They adopt a two-electrode configuration, typically one is gold and the other is coated in DNA probes. The equivalent circuit model

\[\text{Single Nucleotide Polymorphism: single base mutations found across the population occurring at the same position}\]
Figure 3.1: Equivalent circuit of two-electrode system: $INT_A$ and $INT_B$ represent surface of the electrodes treated with DNA probes and in contact with electrolyte, $C_{INT_A}$ and $C_{INT_B}$ are the interface capacitances, $R_S$ and $R_P$ are the solution and interface impedance respectively. Inset illustrates the total external capacitance $C_{TOT}$ between the electrodes [6].

modeled in [6] is shown in Fig. 3.1 The solution containing target DNA molecules is characterised as a conductor ($R_S$), the interface resistance and capacitance between the electrode and solution are modeled as $R_P$ and $C_{INT}$ respectively. Typically a square wave pulse is used to repeatedly charge and discharge the surface-interface capacitance, and its value can be obtained from the average current [12] or time constant [10] in relation to the frequency and applied voltage. Guiducci et al. [6] reported an increase in dielectric permittivity from 1.9 and 2.5 and relative capacitance from hybridisation of probes. [6, 9, 10] all reported a reduction of 25-50\% in capacitance between matched and mismatched sequences. Again, the change in thickness of equivalent circuit model capacitance due to increase of DNA molecules from binding have been theorised as the mechanism behind the drop in value.

One of the drawbacks of EIS is the need for post processing. Electrodes are typically coated in gold, or any inert bio-compatible metal can be used. Even when the system can be fabricated in standard CMOS technology [8], the electrode requires deposition in which a passivation layer opening must be formed. Then an initial deposition of Nickel must be carried out before gold since gold cannot be plated onto the aluminum top metal directly. As well, depending on the architecture of the system, only end point measurements of before and after the reaction can be taken [7].

### 3.1.2 DNA charge measurements

The intrinsic charge of DNA can be used as a method of detection. The theory is described in detail in [13]. Based on floating gate devices, the paper considers charge distribution of the diffusion layer with surface of the FET. In order to maintain constant
current, and therefore electric field in the device, any rearrangements of charge in the electrolyte or semiconductor to liquid junction will be compensated by changes in the electrode voltage of the same magnitude. Indeed, the paper observed opposite shifts in applied voltage due to accumulation of oppositely charged molecules, resulting in positive increase when biopolymer poly(L-lysine) was deposited and negative shift in voltage when DNA was immobilised in p channel devices.

Based on the principle of negatively charged DNA, several variations of sensors were reported over the years. [14] adopted a FET device with thermally grown $SiO_2$ and $Si_3N_4$ as gate membrane for immobilising oligonucleotides, which incidentally is the same as the passivation layers used in unmodified CMOS fabrication. The accumulation of negatively charged DNA from immobilisation caused a shift in threshold voltage which was measured as an increase in gate voltage at constant current. Further hybridisation of probes resulted in a smaller shift of $11mV$ for $2.6 \times 10^3$ target molecules. Subsequent application of SNP detection was demonstrated by introducing single base extension in [15], with an average of $3.2mV$ change for elongation from a single type of complementary dNTP.

![Figure 3.2: Proposed floating gate device in [16], including a control capacitor to bias the gate and active area for sensing](image)

Similarly, Barbaro [17] reported an increase in threshold voltage from DNA charge using a “charge-modulated FET”. A comprehensive analysis of the device was presented in [16], which describes it to be a floating gate FET with an active area for charge induction and control capacitor for providing reference electrode voltage. The inclusion of control capacitor ($C_{CF}$) eliminates the need for external electrode to establish voltage difference between the electrolyte and substrate. Depending on the doping of the device, accumulation of charge can either switch on or off the device by inducing opposing charge in the channel. Since the reported n-channel FET device is turned on via $C_{CF}$, hybridisation is detected by differences in threshold voltages rather than absolute values. The change in threshold voltage ($V_{th}$) due to hybridisation from immobilised
probes can be estimated as:

$$\Delta V_{th} \simeq \frac{Q_{DNA}}{C_{CF} + C_{FB}}$$

(3.2)

Where $Q_{DNA}$ is the charge from double stranded DNA, $C_{CF}$ is the control capacitance and $C_{FB}$ is the capacitance from floating gate to silicon bulk.

Immobilising oligonucleotide probes is a complicated procedure, as highlighted in [18]. It involves cleaning, functionalisation/modification of the surface and then anchoring of probes [17]. [18] described the detection of DNA hybridisation without immobilisation of probes using a FET with exposed aluminum top metal made from the same mask as bond pads openings. A platinum layer acting as the reference electrode was deposited on top of the passivation in post processing. The circuit architecture adopts correlated double sampling for offset and low frequency noise reduction. Relative voltages were reported to increase with various concentrations of DNA, demonstrating potential for real time PCR applications [19].

Detecting hybridisation by DNA charge is a simple technique that have potentials for fabricating in unmodified CMOS. Particularly, the design of an integrated electrode presented in [17] is an attractive feature. Sensitivity in general poses an issue in analyses as any small variations in the double layer will contribute to the output signal. Sensors’ selectivity is extremely low since charge is present in all molecules. This was demonstrated in [13] where the effects of sodium hydroxide surface treatment and other polymers was measured. Nonetheless, detection by DNA charge is a fundamental technique that has encouraged the development of other sensors such as silicon nanowire devices (SiNW-FET) [20].

### 3.1.3 RedOx-based

Figure 3.3: RedOx reactions occurring at two electrodes [21]
The principle of RedOx reactions is widely used in electrochemical sensors. Redox, short for oxidation and reduction, is based on the loss and gain of electrons respectively. In general, DNA analysis by RedOx can be described as follows [22]: oligonucleotide probes are immobilised onto electrodes, then the target DNA strands tagged with enzymatic label (i.e. alkaline phosphatase) are introduced during the measurement phase and hybridisations occur with complementary sequences. The sensor is then washed, and a chemical substrate (i.e. para-aminophenylphosphate) is applied. In the presence of a suitable substrate, the enzyme label then cleaves and generates a RedOx species (i.e. p-AP). An oxidation and reduction potential are applied simultaneously to a generator and collector electrode respectively, where the RedOx species will repeatedly lose and gain electrons, resulting in a RedOx current. Thus the current is proportional to the number of hybridised DNA strands.

Figure 3.4: Interdigitated electrode in circular configuration (left), RedOx-cycling process (right) [23]

One of the first RedOx based DNA sensors was proposed by Hofmann et al. [22]. [21–23] all reported a CMOS sensor architecture made up of interdigitated gold electrodes, enclosed in a circular compartment to avoid contaminations (Fig. 3.4). For each sensor pixel, a potentiostat was included to maintain constant voltage in the electrolyte via a reference and counter electrode. Each pixel sets the RedOx cycling voltages, typically around ±200mV and the current is measured. Artifacts in the analyte causes offsets in current, therefore the rate of change with respect to time is monitored. A broad dynamic range up to 5 decades was required, prompting subthreshold operation [22].
or analog to digital conversion within pixel [23]. Regardless of the readout method, all papers reported an increase in current with DNA hybridisations.

Cyclic Voltammetry (CV) is a specific measurement technique based on RedOx reactions. Periodically ramping the potential between working and reference electrodes up and down (see Fig. 3.5 (left)) results in a working electrode current that corresponds to the subsequent oxidation and reduction reactions. When plotted against the applied voltage or time, it generates a cyclic voltammogram (see Fig. 3.5 (right)) which gives information on the magnitude of RedOx peaks, in addition to analyte concentrations and kinetics of the reaction. [24] applied this technique in an array of gold electrodes and found the area of the RedOx current waveform increased with DNA hybridisation. In particular, sequences with a single nucleotide mismatch resulted in smaller peak currents in comparison to fully complementary DNA, which can be applied in SNP detection. Levine et al. also investigated the relationship between electrode sizing and scan rates, and found a linear correlation with area in both cases. This implies that an increase in electrode area and frequency of ramping voltage could in theory improve sensor performance.

![Figure 3.5: (a) Cyclic voltammetry input, (b) Output voltammogram [25]](image)

Label-free DNA detection is possible using cyclic voltammetry based on changes in current from surface charge variations. The aforementioned literature so far have been based on directly measuring the faradaic current due to targets labeled with RedOx tags. [26] presented a platinum electrode sensor with an electropolymer deposited for oxidation and reduction. The layer is then immobilised with oligonucleotide probes. The paper reported the area of the cyclic voltammetry curve to decrease with hybridisations, due to accumulation of dsDNA on the surface which blocks electron transfer. The percentage decrease in area corresponded to concentrations of the sample, with
10% for 10nM and 38% for 100nM. The reduction in CV curve area is supported by Jafari et al. [25], who used a RedOx complex (potassium ferricyanide) instead of a polymer layer. The authors demonstrated an increase in CV curve area is proportional to concentrations of potassium ferricyanide, and then decrease in current peaks when synthetic prostate cancer DNA were hybridised. The paper explains the reduction as a result of increase in negatively charged DNA repelling the RedOx complex of the same charge. Concentration as low as 10aM was detected by the grown nanostructured gold electrodes, which is known to have better sensitivity than flat electrodes due to a larger active area [27].

Apart from the obvious limitations of post processing electrodes and immobilising target probes, RedOx-based methods are restricted only to hybridisation and cannot be used for sequencing. The technique also demands for systems with difficult to achieve specifications such as large dynamic range or high frequency for sensitivity.

### 3.1.4 Pyrophosphate-based

As mentioned in Section 2.3.2, Pyrosequencing is a label-free sequencing technique based on the direct detection of pyrophosphate (PPi) released during base extension. The series of chemical conversions from PPi to light are summarised in eq. 3.3 below:

\[
\begin{align*}
(DNA)_n + dNTP & \xrightarrow{\text{polymerase}} (DNA)_{n+1} + \text{pyrophosphate (PPi)} \\
\text{pyrophosphate (PPi)} + APS & \xrightarrow{\text{ATP sulphydrylase}} \text{ATP} \\
\text{ATP} + \text{luciferin} & \xrightarrow{\text{luciferase}} \text{oxyluciferin} + \text{light}
\end{align*}
\]

Pyrophosphate is first converted to ATP in the presence of APS catalysed by ATP-sulphydrylase enzymes. The subsequent production of ATP with luciferin is driven to oxyluciferin by luciferase, generating light as a byproduct. CMOS circuits can be used to detect the resultant visible light optically. Different from fluorometric sequencing techniques such as sequencing by synthesis, it does not require an excitable source for illuminating fluorescent tags. An example of such system is presented in [28]. A photodiode array is coupled directly with a fibre-optic faceplate used for immobilising luminescent probes. Strict specifications such as very low dark currents, dynamic range and low power (to prevent overheating of reagents) were required to achieve the necessary sensitivity for detection. The presented lab-on-chip includes differential pixel, correlated multiple sampling, no switching when integrating delayed images, and repeated averaging with digital signal processing. The system was able to detect low levels of light simulated by the addition of ATP and the relevant enzymes.
Alternatively, [29] presented a method for detecting pyrophosphate directly using chelators\(^3\) immobilised on sensors (Fig. 3.6). The chelators captures PPI as target molecules, which are negatively charged. As a result, threshold voltage was reported to become more positive with base extensions due to an increase of holes accumulation in the p-channel FET. Rolling circle PCR (RCA) was used to rise the number of parallel binding sites at each DNA strand, effectively amplifying the signal. Evidently, pyrophosphate techniques can be used for both hybridisation of probes and base extensions.

### 3.1.5 pH-based

The principle of pH-based detection is similar to Pyrosequencing, in which hydrogen ions are detected instead of light. When a dNTP is incorporated into a single stranded DNA, catalysed by polymerase, the pyrophosphate released is also hydrolysed, generating a hydrogen ion as a by-product (see eq. 3.4). The addition of hydrogen ions in the solution alters its pH, and can be detected by ion-sensitive FET devices with a suitable sensing gate membrane. The characteristics of such device changes with pH concentrations so read out of its current or voltage can indicate DNA hybridisation.

\[
(DNA)_n + dNTP \xrightarrow{\text{polymerase}} (DNA)_{n+1} + \text{pyrophosphate}
\]

\[
\text{pyrophosphate}(P_2O_7^{4-}) + H_2O \rightarrow \text{orthophosphate}(PO_4^{3-}) + H^+
\]  

(3.4)

The rise of pH-based DNA detection with FETs began in 1990s and took off in the 2000s. The first reported DNA polymerase reaction detected by ISFET was in 1992

\(^3\)Molecule capable of binding to certain ions or metal ions
by [30], though the paper claimed an increase in pH during hybridisation due to uptake of hydrogen ions. Various attempts were also documented in other disciplines such as Journals of Immunological methods [31, 32], the papers used commercially available ISFET sensors to record results from end-point and real time PCR reactions, with primary focus on optimising biological procedures. Souteyrand et al. [33] reported a shift in measured gate voltage in response to hybridisation, suggesting correlations to changes in surface charge on the gate membrane. Eventually the principle of pH-based DNA analysis was described in detail with regards to theory and detection method by Toumazou and Purushothaman in [34, 35]. The decrease in pH due to nucleotide incorporations was successfully measured for the purpose of SNP detection [36]. Based on this method of analysis, [37] reported an integrated system on chip (SoC) with 40 ISFETs, temperature sensors, heaters, biasing and signal processing circuits for DNA Point-of-Care applications. A micro-fluidic assembly of chambers and channels was designed to sit on the top of the chip to incorporate the chemical samples, and the whole system can be considered as a disposable cartridge for single-time use. Adopting one ISFET as a reference device, the reported SoC is capable of detecting 19 SNPs simultaneously. When loaded with the target probes, the detection of a matched incorporation is determined by the readout voltage crossing a predefined threshold value. The paper reported an average time to threshold of 20 to 40s.

Figure 3.7: ISFET platform reported in [37], with micro-fluidics chambers positioned on top of ISFETs for DNA reactions and detection
3.2 ISFET sensor

Of all the described techniques on label-free electronic DNA detection, pH-based method using ISFETs can provide CMOS compatibility without the need for immobilisation of probes. The option to manufacture in unmodified CMOS without post-processing is appealing, particularly for large scale productions of disposable diagnostic platforms. The direct detection of hydrogen ions can provide an extent of specificity to DNA hybridisation, unlike DNA negative charges measurements.

Despite its earlier discovery by Piet Bergveld in the 1970s [38, 39], majority of ISFETs’ applications have been centered around research purposes until recently [40]. ISFET was originally realised by removing the gate metal of a conventional MOSFET, and exposing the oxide layer (SiO₂) to an electrolyte. A gate electrode, typically silver/silver chloride (Ag/AgCl) is used to apply a constant bias to the solution. The insulating membrane traps or releases protons depending on ion activities in the solution. The accumulation of charge at the insulator surface induces an opposing potential in the ISFET’s channel. Therefore changes in ionic concentration is observed as a threshold voltage variation in the device.

The material of gate insulator defines an ISFET’s characteristics to ions. At the beginning, silicon dioxide was used as the sensing membrane as it was easily obtainable from heating silicon. Hydration of the oxide layer naturally led to pH sensitivity, however it also caused instability in operation due to cracking of the surface. This prompted a double layer structure in which a protective insulator under a sensing membrane is used to ensure isolation of the channel from solution. Subsequent characterisation of various gate materials formed the basis of how we evaluate ISFETs today: sensitivity, namely changes in gate voltage due to pH; selectivity is the device’s sensitivity to other ions; hysteresis describes variations in response when following the same changes of pH.
in opposite paths; drift is the stability of output values over time. Initial results by Matsuo [41] showed tantalum pentoxide (\(Ta_2O_5\)) and aluminium oxide (\(Al_2O_3\)) had superior performance in the aforementioned parameters over silicon nitride (\(Si_3N_4\)), particularly in pH sensitivity (Table. 3.1). On the other hand, \(Si_3N_4\) is inherent in the passivation layer of commercial CMOS fabrication, and can therefore be achieved without additional deposition after manufacturing.

<table>
<thead>
<tr>
<th>pH sensitivity (mV/pH)</th>
<th>SiO\textsubscript{2}</th>
<th>Si\textsubscript{3}N\textsubscript{4}</th>
<th>Al\textsubscript{2}O\textsubscript{3}</th>
<th>Ta\textsubscript{2}O\textsubscript{5}</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 - 35 &gt; pH\textsubscript{7}</td>
<td>46 - 56</td>
<td>48.5 - 57 [42]</td>
<td>56 - 58.5 [43]</td>
<td></td>
</tr>
<tr>
<td>37 - 48 &lt; pH\textsubscript{7}</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3.1: pH sensitivity for different sensing layer material [41]

### 3.2.1 Operation

The operation of an ISFET is similar to its MOSFET equivalent [44]. Its threshold voltage dependence on pH can be derived from the MOSFET’s threshold voltage equation. The threshold voltage of a MOSFET (\(V_{th(MOSFET)}\)) is defined to be the potential required for the inverted carrier concentration to equal to the carrier concentration in the bulk. This can be described mathematically by three potentials: \(\phi_m - \phi_{Si}\) is the difference between the metal and silicon work function; \(Q_{ox}\), \(Q_{ss}\), \(Q_{depl}\) is the charge in the oxide, silicon/gate oxide interface and depletion respectively; \(\phi_f\) is the Fermi potential of the semiconductor.

\[
V_{th(MOSFET)} = \frac{\phi_m - \phi_{Si}}{q} - \frac{Q_{ox} + Q_{ss} + Q_{depl}}{C_{ox}} + 2\phi_f \tag{3.5}
\]

Since there is no metal gate on the ISFET, its threshold voltage (\(V_{th(ISFET)}\)) can be partly defined as the threshold voltage of its underlying MOSFET with exception of the metal work function:

\[
V_{th(ISFET)} = V_{ref} - \psi_0 + \chi_{sol} - \frac{\phi_m}{q} + V_{th(MOSFET)} \tag{3.6}
\]

In which the effects of the floating gate and electrolyte can be considered with sums of the reference electrode potential (\(V_{ref}\)), surface dipole potential (\(\psi_0\)) which is pH dependent and dipole potential (\(\chi_{sol}\)) from the solution. These additional terms can be grouped as a chemically dependent voltage called \(V_{chem}\) [45]:

\[
V_{chem} = V_{ref} - \psi_0 + \chi_{sol} - \frac{\phi_m}{q} \tag{3.7}
\]
In particular, $\psi_0$ is the voltage change due to activities on the insulator, and can be defined as the difference in potential between the oxide surface ($\psi_s$) and bulk solution ($\psi_b$) \[46\]:

$$\psi_0 = \psi_s - \psi_b$$ \hfill (3.8)

### 3.2.2 pH sensitivity

The pH sensitivity of an ISFET can be described using a combination of site-binding theory \[47\] and the Gouy-Chapman-Stern model \[48\]. Examining closely at the surface of the sensing membrane, in this case $SiO_2$, there exists dangling bonds of oxides terminated by hydrogen (OH). When placed in a solution, some of the hydroxide bonds will protonate or deprotonate with hydrogen ions in the liquid. These reactions can be described by the following equations:

$$[SiOH] + [H^+]_s \rightleftharpoons [SiOH_2^+] \quad \text{with} \quad K_a = \frac{[SiOH_2^+]}{[SiOH][H^+]_s} \hfill (3.9)$$

$$[SiOH] \rightleftharpoons [SiO^-] + [H^+]_s \quad \text{with} \quad K_d = \frac{[SiO^-][H^+]_s}{[SiOH]} \hfill (3.10)$$

where $[SiO]$ is the concentration of proton donors, $[SiOH_2^+]$ present the acceptors and $[H^+]_s$ is the concentration of hydrogen ions at the insulator surface. $K_a$ and $K_d$ are the association and dissociation constants respectively, defining the ratio of reactive species’ concentrations at equilibrium. Clearly, concentration of hydrogen ions determines the number of charged binding sites and therefore overall charge of the surface membrane.

The surface charge per unit area on the sensing membrane $Q_s$, can then be expressed as the sum of positive and negatively charged binding sites, taking into account electron charge $q$:

$$Q_s = q([SiOH_2^+] - [SiO^-]) \hfill (3.11)$$

Similarly, the number of total sites $N_s$ is defined by all the binding sites including the neutral OH bonds:

$$N_s = [SiOH_2^+] + [SiOH] + [SiO^-] \hfill (3.12)$$

Substituting the equilibrium constants $K_a$ and $K_d$ into eq. 3.12 and eq. 3.11, the surface charge can therefore be expressed as a function of hydrogen ions on the surface of the insulator:

$$Q_s = qN_s \left( \frac{[H^+]_s/K_a - K_d/[H^+]_s}{1 + K_d/[H^+]_s + [H^+]_s/K_a} \right) \hfill (3.13)$$
Certainly, variations in charge inside the ISFET channel can be deduced by substituting values for $[H^+]_s$, and defining the surface charge with respect to surface pH ($pH_s$) is known as the intrinsic buffer capacity ($\beta_{int}$):

$$\frac{\delta [Q_s]}{\delta pH_s} = -q\beta_{int}$$ (3.14)

Figure 3.9: Illustration of the site binding model, Helmholtz and Gouy-Chapman planes with a graph of the charge potential distribution

However, the concentration of hydrogen ions on the surface of the membrane is not the same as in the bulk solution. This phenomenon can be characterised using the double layer capacitor proposed by the Gouy-Chapman-Stern model. The Helmholtz layer, also known as the Stern plane describes the behaviour of ions closest to the sensing membrane. The inner Helmholtz plane (IHP) consists of ions that are attracted to the binding sites on the insulator; the outer Helmholtz plane (OHP) refers to a line
of charge that are surrounded by water molecules, preventing them from approaching the surface. These specifically absorbed charges and hydrated ions cause more or less a linear drop in potential from the surface membrane. The Gouy-Chapman model refers to the diffusion of ions away from the outer layer and into the bulk electrolyte. The effect is an exponential drop in potential as the charges follow a Boltzmann distribution before becoming evenly spaced in the solution. The different charge regions can be modeled as two series capacitors, $C_{Helm}$ and $C_{Gouy}$, forming a double layer capacitor $C_{dl}$:

$$\frac{1}{C_{dl}} = \frac{1}{C_{Helm}} + \frac{1}{C_{Gouy}}$$ (3.15)

Since the law of conservation of charge dictates that net charge must sum to zero, the sum of charges from the surface and double layer ($Q_{dl}$) must also equate to zero:

$$Q_{dl} = -Q_s$$ (3.16)

Subsequently, the double layer capacitance can be written with respect to the change in potential across the solution ($\psi_0$):

$$\frac{\delta Q_{dl}}{\delta \psi_0} = -\frac{\delta Q_s}{\delta \psi_0} = -C_{dl}$$ (3.17)

Based on eq. 3.14 and 3.17, variation in insulator-electrolyte potential with respect to surface pH can be expressed as a function of the intrinsic buffer capacity and double layer capacitance:

$$\frac{\delta \psi_0}{\delta \psi_s} = \frac{\delta \psi_s}{\delta Q_s} \frac{\delta Q_s}{\delta \psi_0} = -\frac{q\beta_{int}}{C_{dl}}$$ (3.18)

According to Gouy-Chapman model, the relationship between concentration of hydrogen ions in the bulk solution ($[H^+]_b$) and insulator surface can be described using the Poisson-Boltzmann equation as a function of the insulator-electrolyte potential ($\psi_0$) from eq. 3.8, with $k$ as the Boltzmann constant:

$$[H^+]_s = [H^+]_b \exp \left( -\frac{q\psi_0}{kT} \right)$$ (3.19)

Taking the logarithm of the above eq. 3.19 and substituting into eq. 3.18, we can obtain the change in potential with respect to solution pH, which is the ISFET’s pH sensitivity ($S_{ISFET}$):

$$S_{ISFET} = \frac{\delta \psi_0}{\delta \psi_s} = -2.303kT \frac{q}{q^2\beta_{int}} \left( \frac{2.303kTC_{dl}}{q^2\beta_{int}} + 1 \right)^{-1}$$ (3.20)

The limit in pH sensitivity can be deduced theoretically based on the required reference electrode voltage in order to maintain the same current for different pH solutions. Incorporating the electrode potential into eq. 3.19, and given that pH is the
logarithm of hydrogen ion concentration \((pH = \log[H^+])\), the surface hydrogen ion concentration can be rewritten as the following:

\[
[H^+]_s = \exp \left( -2.303pH - \frac{q}{kT}(\psi_0 - V_{ref}) \right)
\]  \(3.21\)

Since the current is proportional to surface charge concentration, which is a function of hydrogen ions on the insulator surface, equating the above eq. 3.21 for two different pH, we can obtain the change of reference electrode voltage with respect to pH variation. Substituting values for Boltzmann constant \(k\) at room temperature 298K, the potential to maintain current for a unit pH change is 59mV, which is in agreement with the sensitivity of an ideal Nernstian response \(S_N\).

\[
\frac{\delta V_{ref}}{\delta pH} = \frac{-2.303kT}{q} = -59\text{mV/pH}
\]  \(3.22\)

Therefore, deviation of an ISFET’s pH sensitivity to the ideal Nernstian value can be defined by the factor \(\alpha\) [49]:

\[
\alpha = \frac{S_{ISFET}}{S_N} = (\frac{2.303kTC_{dl}}{q^2\beta_{int}} + 1)^{-1}
\]  \(3.23\)

As a result, \(V_{chem}\) can be rewritten in a form that directly relates to pH by substituting \(\psi_0\) into the earlier eq. 3.7, with \(\gamma\) as the grouping of all pH independent chemical potentials and \(U_t\) is the thermal voltage \((\frac{kT}{q})\):

\[
V_{chem} = \gamma + 2.303\alpha U_t pH
\]  \(3.24\)

### 3.2.3 ISFET modeling

The effects of the double layer capacitance and the ISFET’s sensitivity to pH can be summarised in a behavioural macromodel proposed by [50] shown in Fig. 3.10. The Gouy-Chapman \((C_{Gouy})\) and Helmholtz \((C_{Helm})\) capacitances are represented as...
a variable and linear capacitor respectively. \( V_{G'} \) denotes the potential seen after the double layer, so the voltage drop across the electrolyte solution can be given in terms of \( V_{chem} \):

\[
V_{G'} = V_{ref} - V_{chem}
\]  

(3.25)

We can substitute eq. 3.25 to the MOSFET’s characteristics equations to obtain drain current for ISFET devices in the prospective modes of operation. When the gate-source voltage \( (V_{GS}) \) exceeds threshold voltage, and \( V_{DS} \geq V_{GS} - V_{th} \), the device is in saturation:

Strong inversion: \( I_{D(SAT)} = \frac{\mu C_{ox} W}{2L} (V_{ref} - V_{S} - V_{chem} - V_{th(MOS)})^2 (1 + \lambda V_{DS}) \)  

(3.26)

Where \( \lambda \) is the factor for channel length modulation. When the device operates below threshold voltage \( (V_{GS} < V_{th}) \), the drain current is dominated by diffusion of carriers which has exponential characteristics:

Weak inversion: \( I_{D(WI)} = I_0 \exp \left( \frac{V_{ref} - V_{S} - V_{chem}}{nU_T} \right) \left( 1 - \exp \frac{-V_{DS}}{U_T} \right) \)  

(3.27)

3.3 ISFET design in unmodified CMOS

![Diagram of ISFET](image)

**Figure 3.11:** Cross sectional diagram of an ISFET fabricated in unmodified CMOS with the corresponding behavioural model, \( C_{pass} \) represents passivation layer capacitance
Fabrication of ISFETs in unmodified CMOS process allows for easy integration of the device into circuits on wafer. The extended gate approach was first proposed by Bausells et al. in 1999 [51]. Metal stacks are used to connect polysilicon gate from the buried MOSFET to top metal, therefore coupling Si$_3$N$_4$ from the top passivation layer for pH sensitivity (Fig. 3.11). Adapting such technique has benefits of low cost mass production, miniaturisation and scalability. However, performance also suffers in terms of drift, reduced sensitivity and offsets in threshold voltage when compared to devices whose pH sensitivity is achieved through post processing of gate insulator [52]. It is possible to compensate for these effects to an extent through various design techniques, which are discussed in the following sections.

3.3.1 Sensor area

When manufacturing ISFETs in unmodified CMOS technology, the passivation layer is coupled to the polysilicon gate as the sensing membrane. By doing so, it effectively adds a series capacitance $C_{pass}$ between the electrolyte and underlying MOSFET. Considering the materials of the passivation layer to be silicon nitride and silicon dioxide, $C_{pass}$ value can be estimated as a combination of the two dielectrics [53]:

$$C_{pass} = \varepsilon_0 A_{chem} \frac{\varepsilon_{Si_3N_4} \varepsilon_{SiO_2}}{\varepsilon_{Si_3N_4} t_{SiO_2} + \varepsilon_{SiO_2} t_{Si_3N_4}}$$

(3.28)

Where $\varepsilon_0$ is the permittivity of free space, $A_{chem}$ represents the chemical sensing area, $\varepsilon$ and $t$ are the permittivity and thickness of respective materials. According to the specifications in AMS 0.35µm process, the passivation capacitance is around 22.65µF/m$^2$ [54]. In comparison, the double layer capacitance is orders of magnitude bigger at around 0.16F/m$^2$ [49]. The equation for coupling of the reference electrode voltage should now take into account the passivation capacitance:

$$V_G^* = (V_{ref} - V_{chem}) \frac{C_{pass}}{C_{pass} + (C_{ox}C_d/(C_{ox} + C_d))}$$

(3.29)

Where $V_G^*$ represents the attenuated voltage sensed at the polysilicon gate, due to the capacitive divider formed by $C_{pass}$ and the MOSFET’s oxide and depletion capacitances, $C_{ox}$ and $C_d$ respectively. The effect is similar to a decrease in the device’s transconductance, as shown in Fig. 3.12b. For simplicity, the effect of capacitive attenuation can be defined as the scaling factor $S_{pass}$ based on $\zeta$:

$$S_{pass} = \frac{1}{1 + \zeta}; \zeta = \frac{C'_{ox}}{C'_{pass}} \frac{A_{elec}}{A_{pass}}$$

(3.30)
Where $C'$ represents the value of capacitance per area, $A_{elec}$ is the electrical device area and the equivalent capacitance of the gate oxide and depletion are estimated as $C_{ox}$. Based on eq. 3.30, [55] proposed that passivation attenuation can be minimised by considering the geometry of ISFET devices. Particularly, $\zeta$ can be reduced by maximising the ratio of chemical area to electrical area. With $C''_{ox}$ estimated to be $4821 \mu F/m^2$ in 0.35 $\mu m$ process, a scaling factor of 0.9 would require a sensing area to be almost 2000 times bigger than the buried MOSFET. At half attenuation, $A_{chem}$ needs to be 200 times larger than $A_{elec}$.

\[
V_{G''} = V_{ref} - V_{chem} - V_{tc}
\]

Figure 3.12: Illustrations of (a) DC offset by trapped charge and (b) attenuation by passivation capacitance on ISFET’s characteristics

3.3.2 DC offset

Trapped charge is a phenomenon that arises as a result from the CMOS fabrication process. The accumulation of charge originate from the absence of dissipation path to ground during manufacturing, and can exist in the polysilicon gate or between the passivation layers in the ISFET [56]. The effect can be modeled as a DC offset $V_{tc}$ in the device characteristics, similar to a threshold voltage shift illustrated in Fig. 3.12a. Of 30 fabricated ISFET sensors, [57] found the threshold voltage varied from -14V to +8V, with a standard deviation of 5.95V. The diverse distribution of trapped charge should be accounted for when designing biasing points and voltage ranges to turn transistors on.

Compensation for trapped charge may be achieved through design and post fabrication treatment. Attempts to remove trapped charge via UV radiation was demonstrated in [56, 58]. On a similar basis to programmable read-only memory devices (EPROM),
electrons can be excited with sufficient energy to overcome the barrier of the passivation or floating gate. Reduction in threshold voltage offset was reported in both papers after UV exposure using EPROM eraser. Alternatively, [59] proposed a programmable gate ISFET (PG-ISFET) to overcome effects of trapped charge, shown in Fig. 3.13. The structure of the reported PG-ISFET includes a control gate capacitor ($C_{CG}$) to provide biasing directly to the floating gate, thus canceling trapped charge. While PG-ISFETs allow the DC biasing point to be set externally, the additional capacitance also further reduces voltage at the floating gate and occupy additional area [60].

![Figure 3.13: Behavioural macromodel of programmable gate ISFET reported in [59]](image)

### 3.3.3 Measurement Times

Practical applications of ISFETs have been largely limited by the instability of long term pH measurements. A slow and gradual change in threshold voltage occurs in the device when it is exposed to an electrolyte over a long period of time. This “drift” phenomenon has been generally considered as a consequence of ionic fluctuations on the pH-sensing layer [61]. Hydration of the insulator surface modifies the oxide layer in contact with the solution, progressively changing the combined capacitance of the dielectric’s surface and bulk. As a result, the paper reported a temporal drift in gate voltage of several millivolts per hour, in agreement with later publications [62]. Other explanations of drift include ion migration within the membrane due to electric field [41], or injection of electrons from the solution to insulator under large reference electrode voltage [63]. Certainly, there are many possible contributions to drift making it very difficult to model and predict. It had long been accepted that drift is monotonic and follows a single direction, but as demonstrated in [54] this is not guaranteed to be the case. Having said that, the rate of drift can be expected to be fastest at the beginning before slowing down, presumably to counteract the initial non-equilibrium conditions.
in the electrochemical system.

Since the behaviour of drift is highly variable, any compensation schemes must consider the practical aspects of the intended application. For example, the range of pH to be detected can impact the extent of drift as a larger temporal change in voltage was observed at higher pH in [64]. Sensing membranes of different materials will also have varying resistance to surface hydration (Fig. 3.14). However, the real decisive factor when compensating for drift is whether it is necessary in the first place. The effect of drift is only critical if the application demands for high accuracy in absolute values over a long time. Notably if the detection period required is several seconds or minutes, such as in DNA hybridisations, the impact of drift would be negligible. When considering applications for long term monitoring, circuit techniques such as differential measurements can be applied to cancel drift. The downside of such approach is that drift pattern between two devices is not guaranteed to be similar, with values varying from 1.5 to 8.5 mV/h measured from ISFETs on the same die [57]. Alternatively, [65] proposed resetting drift by switching the reference electrode voltage using a PWM. This effectively cycles the vertical electric field between the insulator and electrolyte, allowing any accumulated ions to diffuse away from the surface. Experimental results showed the drain current of the ISFET is able to return to the starting value at each reference voltage cycling [66]. Power consumption would be the main drawback when using oscillators for switching.

![Graphs showing drift due to different sensing membrane materials](image)

(a) $\text{Al}_2\text{O}_3$  
(b) $\text{Si}_3\text{N}_4$

Figure 3.14: Drift due to different sensing membrane material [61]
3.3.4 Temperature and Noise

Other design considerations when using ISFETs for chemical sensing is sensitivity to noise and temperature. The modeling of these non-idealities are not dissimilar to its MOSFET counterpart. ISFETs suffer the same noise sources such as thermal and flicker noise (also known as 1/f noise) [67]. In addition to electrical noise, chemical noise that arise from ion activities is significant. As demonstrated in [57], the increase in 1/f noise for ISFETs can be up to two orders of magnitude bigger than MOSFETs of the same electrical area (Fig. 3.15).

Figure 3.15: Comparison of flicker noise between MOSFETs and ISFETs from [57]

Temperature sensitivity in ISFETs is more complex than its equivalent MOS transistor, as dependency arise not only from the body effect in threshold voltage, the reference electrode, electrolyte and sensing membrane are all susceptible to thermal changes [68]. The reported experiments carried out on ISFETs with silicon nitride sensing layer showed increased pH sensitivity at higher temperatures, (58 to 70mV/pH), in agreement with simulation results published in [69]. A straight forward solution to output variations from temperature is to ensure readings are taken in the same conditions. ISFETs operating in weak inversion have also been proposed to reduce temperature sensitivity by thermal voltage cancellation [70]. Stable monitoring independent of temperature fluctuations have been shown to be achievable by intelligent systems incorporating compensation algorithms [71] or calibrations using predefined lookup tables [72]. Though such compensation schemes to accommodate ISFETs for long term measurements come at the heavy expense of computation power and additional software and hardware implementations.
3.4 ISFET-based epigenetic detection

Fabricating in unmodified CMOS process has without a doubt led to some commercial success. While non-ideal effects from the passivation layer and drift have prevented ISFETs to be widely adapted in biomedical applications, ISFETs are well-suited to DNA detection as it does not require high precision and long term measurements. Based on the principle of pH-based detection, there are several options for implementing methylation and miRNA analysis: sequencing is essential for discovering new modifications, particularly as a means to provide single base resolution in methylation; application specific ICs can be designed to carry out a particular type of analysis; real-time PCR is for quantitative measurements, namely miRNA quantification. The following sections examine some existing ISFET platforms that can be applied to analyse epigenetic modifications.

3.4.1 Sequencing by ISFETs

In 2011, Rothberg et al. [1] from Ion Torrent Life Technologies published the first whole genome sequencing results using ISFET devices. The massively parallel chip contains 1.2 million sensors fabricated in modern CMOS process. Tantalum oxide was added on top as the pH sensing layer to give a higher sensitivity (58mV/pH) than the one provided by the passivation layer manufactured in unmodified CMOS process.

---

4Technology was licensed by DNAe [http://www.dnae.com/index.html]
Genomic DNA templates are fragmented then amplified by emulsion PCR, with the resultant beads distributed into wells by placing the chip into a spinning centrifuge. The depth of each well was designed to contain only one bead for detection. As one type of dNTPs are introduced to the chip sequentially with washes in between, any complementary nucleotides would release hydrogen ions as a result of base extension. The signal generated by the change of pH resembles an instantaneous peak shown in Fig. 3.16b, since any hydrogen ions released in the incorporation event diffuse away causing the signal to return to baseline. Base calling is carried out off-chip using software that incorporates physical modeling of diffusion rates and other chemical interactions. The reported system can sequence 25 million bases in a 2-hour run, with average read lengths of 100 to 200bp. Results obtained from sequencing three bacterial and human genomes achieved coverage of at least 96.8% and per-base accuracy of 98%, which is comparable and in some cases higher than optical based methods.

Detecting methylation with semiconductor-based sequencing machines is becoming an increasingly popular choice [75, 76], particularly for genome wide studies. In [77], the authors applied an affinity enrichment based protocol, MeDIP (Methylated DNA immunoprecipitation) on the Ion Torrent sequencing platform to detect total methyl-
cytosine contents and differential methylation. As well as validating previously known differentially methylated loci, the platform identified additional sites that were later confirmed on a fluorescent array (Illumina 450k array). Such semiconductor-based methylation analysis is an attractive option due to its fast run time and low start up and operating costs. The potential for scalability also means it can achieve ultra high throughput sequencing, with read length improvements increasing linearly at the speed of, if not faster than Moore’s law (Fig. 3.17a). Since the first commercial Ion Torrent sequencing machine, the number of on-chip ISFETs have increased from 1.2M to 154M, and more recently to 660M in the Proton II chip [73]. However, the relative downsizing of wells to accommodate for higher sensor density comes with the undesirable effect of faster diffusion rate which results in a shorter detection window. The diffusion of hydrogen ions between neighbouring wells also contributes to the deterioration of accuracy when trying to identify multiple incorporation events. When a sequence contains a succession of the same base, known as a homopolymer sequence, the proportional increase in hydrogen ions released from the multiple base extension corresponds to a larger signal. As the length of the homopolymer increases, the ability to accurately estimate the number of bases from the amplitude of the signal rapidly decreases as seen in Fig.3.17b. This is one of the biggest limitations of ion-semiconductor sequencing, and its effect becomes more prominent after bisulfite conversion as the modification reduces sequence complexity, therefore increasing the probability of homopolymerisms.

3.4.2 Application-specific IC

Figure 3.18: System block diagram of ASIC reported in [78]

One of the major advantages of fabricating in CMOS process is the potential of designing application-specific ICs. [79] proposed a circuit based on translinear principle\footnote{The product of weak inversion currents are equal in clockwise and anticlockwise junctions, provided that there are an even number of junctions.} for methylation analysis. With the main focus on detecting aberrant methylation...
levels for cancer diagnosis, the author looked at applying methylation-specific PCR and computing the percentage methylation via circuitry. This was achieved by using two ISFET sensors for methylated and unmethylated DNA templates, and using a translinear loop to express the output current ($I_{out}$) as a ratio of the resultant ISFETs’ drain currents, $I_{meth}$ and $I_{unmeth}$ from MSP. Indeed, the author used a commercial pH sensor to test products from in-tube MSP reactions and confirmed a higher drop in pH with matched templates. $I_{out}$ is then compared with an adjustable reference current that represents a threshold value before the input of an inverter. The final output of the circuit is a voltage that swings to supply or ground, depending on if the output current is over and below the threshold respectively. The proposed circuit was fabricated and its operation was confirmed by measured results in [78].

Other pH-based methylation analysis ASIC also operate on the basis of methylation-specific PCR. [80] proposed a system that monitors progression by comparing the initial and current status of methylation. The paper recommends subtracting the present and previous methylated and unmethylated values as a newly defined ratio for percentage methylation. Computations are carried out by a digital back-end following three ISFET readout circuits, with one acting as the reference. Drift is accounted for by taking the differential currents between sensor and REFET for calculations. DC offset is sampled when the ISFETs are biased in weak inversion and canceled with autozeroing (CDS). Unlike the first ASIC described in [78], emphasis of this system is its compensation schemes rather than MSP analysis. The same author from [78, 79] also presented a proof of concept for on-chip methylation detection [81] using the SNP analysis platform from [37]. The chambers were loaded with methylated and unmethylated primers to carry out isothermal MSP. An average of 30mV and 20mV change were observed on the sensors from two genes, demonstrating the potential of a common platform for genotyping and methylation.

3.4.3 Real-time amplification

As described in Section 2.4, quantification of miRNA can be carried out by real-time PCR. Several attempts of quantitative analysis using ISFETs were reported in [18, 83]. The majority of publications used externally amplified products of different concentrations or various cycles of PCR, with none demonstrating integrated real-time on-chip amplification.

The first real time monitoring of DNA amplification on an integrated ISFET platform was reported by Toumazou et al. [84]. Fabricated in unmodified CMOS process, pH-LAMP is LAMP amplification with optimised conditions for pH change (i.e. low buffering) [82]
the system consists of ISFET sensors, embedded heaters with temperature control circuitry for thermal cycling. ADC is included to perform on-chip digitisation to improve SNR, as well as allow communication via SPI which minimises off-chip connections. Micro-fluidics chambers of 2µl reaction volumes are mounted on top of the chip, covering at least one or two ISFETs for redundancy, as well as to relax alignment specifications. Three chambers are used to carry amplification, with the middle as the reference containing no templates. A PCR reaction of 40 cycles completed in 35 minutes, with results comparable to fluorescence reactions running in parallel. The serial dilution obtained from both on-chip PCR and pH-LAMP amplification showed good linearity, detecting as low as 10 copies consistently. The paper also found threshold cycle variations between sensors to be low when compared within the same chip, showing reliability for quantitative pH-based amplification measurements.

In addition to scalability, one of the biggest appeals of the reported platform is its small size. The microchip itself is 4.5mm x 5.5mm, the size of a fingernail, and it is operated by a battery powered microelectronic reader slightly larger than an SD card shown in Fig. 3.20. This level of miniaturisation is likely to be only achievable by pH-based ISFET detection. Table. 3.2 compares some of the portable real time PCR machines available commercially and in literature. With the exception of the aforementioned system, all portable machines are optical based, typically using LED and photodiodes as a excitation source and detection respectively. [85] reported a system utilizing the image camera on androids and tablets for easy integration to a portable thermal cycler. Evidently, Toumazon's integrated platform is significantly smaller than any of the reported systems. As well, majority of the machines are limited by number of samples which can fit inside the instrument since they all use tube-based reactions. Consequently their reaction volumes are larger, and are more prone to contaminations.
and higher costs. Other specifications such as amplification time and sensitivity are also comparable, if not superior to optical-based detection. Without a doubt, Toumazou et al. have shown transistor-based monitoring as a promising choice for future portable real-time amplification systems.

Table 3.2: Comparison of existing portable real-time PCR machines

<table>
<thead>
<tr>
<th>System</th>
<th>Method</th>
<th>No. of samples</th>
<th>Min. copy</th>
<th>Size (mm)</th>
<th>Volume</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mini8⁷</td>
<td>Optical</td>
<td>8</td>
<td>1</td>
<td>205x190x98</td>
<td>15-150µl</td>
<td>-</td>
</tr>
<tr>
<td>Hunter⁸</td>
<td>Optical</td>
<td>13 (1 for control)</td>
<td>-</td>
<td>172x412x412</td>
<td>20µl</td>
<td>1-2hr</td>
</tr>
<tr>
<td>Mic qPCR⁹</td>
<td>Optical</td>
<td>48</td>
<td>1-3</td>
<td>150x150x13</td>
<td>5-30µl</td>
<td>25mins</td>
</tr>
<tr>
<td>Toumazou [84]</td>
<td>pH-based</td>
<td>up to 40</td>
<td>10</td>
<td>4.8x5.5</td>
<td>2µl</td>
<td>35mins</td>
</tr>
<tr>
<td>Bialek [85]</td>
<td>Optical</td>
<td>3</td>
<td>-</td>
<td>10inch wide</td>
<td>25-30ul</td>
<td>&gt;3hr</td>
</tr>
</tbody>
</table>

Figure 3.20: Photograph of microchip mounted on SD card next to electronic reader, penny for scale [84]

---

⁷https://www.mobitec.com/cms/products/bio/10_lab_suppl/mini8_real_time_pcr_cycler.html
⁸https://instantlabs.com/technology/
⁹https://biomolecularsystems.com/mic-qpcr/
3.5 Summary

This chapter reviewed existing semiconductor-based technology for DNA detection. The incentive for solid state analysis stems from advantages of CMOS design and manufacturing process. Electronic detection allows for platforms with portable architecture, scalability and low costs. In particular, label-free techniques can minimise sample preparation and subsequent reagent costs. An overview of label-free electronic DNA detection systems based on various principles was presented: impedance spectroscopy; DNA negative charge; RedOx; pyrophosphate and pH. A common drawback was found to be post processing of electrodes. Sensitivity is very low in DNA negative charge techniques, since charges that did not originate from DNA reaction could also influence the output. Of the few platforms that could carry out label-free detection based on RedOx principle, they all require immobilisation of probes and are restricted to hybridisation only applications. Techniques that rely on immobilizing DNA molecules not only suffer from the added specialised preparation step, the density and even distribution of templates would impact results. Finally, pH-based analysis detects hydrogen ions released during DNA extension, and can be achieved using ISFETs fabricated in unmodified CMOS process with micro-fluidics assembly to contain DNA templates.

Following on from pH-based detection, the principles of ISFET’s operation and pH sensitivity were studied. The device is loosely based on a MOSFET without its metal gate and exposing its gate oxide to an electrolyte, thereby charges in the channel is related to ion concentration. More specifically, pH sensitivity arise from surface binding sites that can become positive or negative from dissociation reactions, and the distribution of ions through the solution can be described using the double layer capacitance model. As a result, voltage drop due to pH of the solution can be defined as $V_{\text{chem}}$ and incorporated into the equivalent MOSFET’s equations to characterise behaviour in ISFETs. Notably, drop in pH expected in DNA base extension will result in increase of drain current since $V_{\text{chem}}$ is decreased. Design considerations should be taken into account when fabricating in unmodified CMOS process due to non-ideal effects from the passivation layer, namely DC offset, capacitive attenuation, drift, noise and temperature sensitivity. Compensation can be achieved through various techniques such as area ratios of sensing membrane with polysilicon gate to minimise passivation capacitance, or removal of trapped charge by UV radiation.

Lastly, existing ISFET platforms capable of detecting epigenetics modifications were examined. Reported systems include Ion Torrent sequencing, ASIC for Methylation-specific PCR and platform for real-time amplification. Ability for high throughput and scalable architecture was reflected in the various generations of Ion Torrent se-
quencer chips, though limitations arise from diffusion of hydrogen ions between neigh-
bouning wells, and thus increasing error rate when calling long homopolymers. The
reported ASIC showed possibility to apply processing circuits with ISFET sensors to
achieve output that corresponds to methylation results. Ultimately, the portability
of semiconductor-based detection was clearly demonstrated when comparing the in-
tegrated ISFET platform for real-time amplification with qPCR machines currently
available on the market.
Bibliography


Chapter 4

Analogue current mode design approach for methylation

As shown in chapter 2, aberrant methylation plays a crucial role in the study of chronic disease developments. With the appropriate detection method implemented in a suitable technology, there is potential for methylation to be integrated as a stable biomarker in the foreseeable future. With ISFET-based DNA detection, the reduced costs of manufacturing in an already mature CMOS fabrication process, along with a well-established database of circuit techniques, can help make methylation detection microchips into a reality. There are two main areas that we envision methylation to play a part in future healthcare: as a routine check up for prevention and diagnoses of chronic diseases; and for developing targeted therapeutics which would require detailed analysis and monitoring of methylation status. Considering these applications and the current available detection techniques that are compatible with ISFETs, we want to design circuits that are optimised for each of the two scenarios.

For developing Point-of-Care systems, the main priorities are cost and time. Ease of use and size are also important criteria to satisfy in a portable application setting. Chemically, Methylation-specific PCR is the simplest and fastest methylation detection method for one or two designated CpG sites. Based on the amplification results of methylated and unmethylated primers, the essence of the technique naturally leads to differential measurements. With a brief overview of existing ISFETs differential configurations, we decided to look at the simplest way of achieving MSP detection with the least amount of sensors. Together with low power considerations and compensation for non-ideal effects, we propose a weak inversion ISFET-based current mirror intended for MSP detection with prospect of quantitative analysis.

As for comprehensive methylation analysis, bisulfite sequencing is without a doubt the most effective method in providing information at single CpG resolution. De-
scribed in section 3.4.1, the Ion Torrent platforms have made huge progress in the field of semiconductor-based sequencing, by reducing costs of sample and steady improvements in per-base accuracy with number of sensors per chip. Still, the biggest challenge in the technology lies in the upper limit of detecting same-base occurring sequences. Incidences of homopolymerisms in a sample is intensified by reduced sequence complexity after bisulfite conversion. Fundamentally, the problem of distinguishing long lengths of homopolymers arise from diffusion of hydrogen ions, which distorts the proportionate increase in signal and limits the window of detection. We propose incorporating existing signal processing circuits, such as a switched current integrator with ISFET sensors in hope to resolve this issue by providing a stable averaging of the output. In addition, we feel that there could be improvements in the current sequencing protocol to be better adapted for methylation purposes. After all, sequencing every base seems excessive when only methyl-cytosines are of interest. Together with the idea of a DNA integrator that improves accuracies in homopolymers base calling, we suggest a new compressed method of loading bases on ion-semiconductor sequencing systems that can provide same information on the location and number of methylated CpG sites with less steps.

The chapter is split into three main sections: the first section presents the results of an ISFET current mirror for differential sensing based on MSP; the second section considers integration as a solution to improving homopolymerism base calling in ion-semiconductor sequencing, and subsequent ISFET integrator design and measurements; the final section proposes a novel methylation specific protocol that may be applied to ISFET sequencing to improve efficiency.
4.1 MSP-based Circuit Design

This section proposes an ISFET current mirror in weak inversion (WI) for detection based on Methylation-Specific PCR (see Section 2.3.2). In MSP, methyl-cytosine is detected by positive amplification results using designated primers that explicitly defines the methylated target site. Another set of primers are used in conjunction to anneal to the same CpG without methylation. This can be described as a form of differential sensing, and the section begins by examining some existing differential ISFET topologies.

4.1.1 Differential ISFET monitoring

Differential ISFET sensing originally emerged as a solution to developing a reference electrode that was compatible with solid state fabrication. The ISFET as a pH sensor requires a well-defined liquid junction potential that does not vary with analyte composition. This is typically provided by an Ag/AgCl arrangement, consisting of a silver wire contained in an inert solution such as KCl with a porous frit. The structure of such electrode is bulky, therefore defeats the purpose of small detection volumes that can be achieved with CMOS-based sensors. As a result, a reference ISFET, termed as a REFET was proposed [1]. REFET would ideally be insensitive to all ions and thus act as an alternative to the conventional glass electrode, with the sampled solution grounded by a noble metal (i.e. platinum).

The initial approach for achieving ion-insensitivity in REFETs was focused on developing a membrane with low site density. After all, a high number of binding sites means more are available for protonation and deprotonation, particularly at high pH concentrations. Dissociation of protons is directly linked to changes in the surface charge and therefore intrinsic buffer capacity (see eq. 3.14). According to Matsuo et al. [1], an insulator site density of less than $1 \times 10^{12}/cm^2$ could provide negligible pH sensitivity by limiting the change in surface potential to below 1mV. However, this is difficult to achieve in practice, as even a hydrophobic material such as Parylene has an estimated site density of $3 \times 10^{12}/cm^2$, and any further reduction would require chemical modifications.

In addition to the number of binding sites, total charge density in the semiconductor also influences surface potential on the gate insulator. For the device to be unresponsive to ion concentration, the effect of the double layer capacitance must be compensated. This could be achieved by controlling the potential of the device via substrate contact, so that the net charge between the surface and diffusion layer is balanced (see eq.
As well as trying to compensate for the surface potential on the REFET device alone, electrical characteristics between an ISFET and REFET must be matched. As demonstrated in Hammond’s paper [2], deposition of PVC membrane to form a REFET caused further mismatch in threshold voltages between the sensors.

Figure 4.1: Differential ISFET-REFET readout configuration [3]

With limited success in pursuing the “perfect REFET” with minimal ion sensitivity [4, 5], attention shifted to using ISFETs with differing pH responses in various circuit configurations. Wang et al. [3] reported a differential pH-sensing circuit consisting of two integrated ISFET op amps connected as voltage followers, and the outputs were differentially amplified off-chip. One of the op amps contains an ISFET with high pH sensitivity (∼58mV/pH) provided by $Ta_2O_5$ as the gate insulator; the other has a reduced pH sensitivity (∼18mV/pH) from oxidisation of $Si_3N_4$ at a very high temperature. Experimental results using on-chip gold electrode for biasing the solution achieved an overall pH sensitivity of 40-43mV/pH, and common mode rejection of noise that occur in both ISFETs.

Alternatively, [6] and [7] showed that differential pH sensing can also be attained without post processing of membranes using ISFETs made in an unmodified CMOS process. Shepherd et al. [7] fabricated a REFET by excluding contacts that coupled the polysilicon gate to the top metal sensing area. Each ISFET and REFET sensor was connected to a voltage clamping circuit and combined to form a differential pH readout with fixed biases. In [6], Chodavarapu et al. proposed a current mode equivalent of the configuration reported in [3]. Deliberate variation in response to pH between the two IOTA (ISFET Operational Transconductance Amplifiers) was designed through sizing of the transistor loads, achieving similar performance in terms of overall pH sensitivity.
So far, majority of differential ISFET and REFET configurations have been primarily concerned with maintaining good pH sensitivity in readout whilst allowing for downsizing and integration of solid state electrodes. In conventional MOSFET topologies, differential inputs are used to provide many benefits such as noise reduction, possibility of feedback, and namely common mode signals rejection. Since many non-idealities in ISFETs can be common in devices (see Section 3.3), differential ISFET measurements can greatly improve its potential in bio-sensing applications. As Sohbati [9] suggested, the transistor characteristics of an ISFET should be exploited as part of the conditioning circuit, rather than viewed as a pH-sensor alone with readout done by fixing constant bias. In the paper, the author connects an ISFET and REFET like a differential pair as part of a pixel in a genotyping system. In such application, the REFET is identical to the ISFET in terms of device structure, and is used as a reference by coating of non-matching primers. This idea of an inert reaction-based REFET was carried forward by Kalofonou and Toumazou [8], in which the input currents of a Gilbert cell were provided as a result of the reaction and buffer chambers positioned on top of two ISFETs. Variations of pH only occur in the reaction chamber containing biological samples and reagents; the REFET chamber contains only buffer solution therefore no pH changes can occur. Incorporating ISFET sensors as part of a
Gilbert cell provide advantages such as tunable current gain and low power operation that are intrinsic to the circuit’s design. In addition, the differential structure reduced drift by 100 fold compared to an ISFET as a single ended device.

To sum up, differential ISFET sensing can be used as a means to miniaturise reference electrode or for compensating non-ideal effects of the sensor. Initial attempts at developing self-contained pH sensors with solid state reference electrode lead to the proposal of a reference ISFET that did not respond to changes in surface potential. Creating such a device that was completely insensitive to ion activities proved to be challenging in terms of developing site-free insulator material, compensation of surface charge from the semiconductor, and matching electrical properties with the paired pH sensitive ISFET. Consequently, later publications focused on using fabrication and processing methods, or circuit techniques to achieve differential readout with ISFETs of sufficiently distinct pH sensitivities. More recently, differential ISFET measurements were applied in conjunction with biomedical analysis, such that the REFET serves as a control biologically and in the circuit. In such configurations where the differential biosensing forms part of the signal processing circuit, drift can be greatly reduced along with common mode noise. On the other hand, any mismatch in non-idealities such as trapped charge are magnified under high gain [9]. Nevertheless, differential ISFET measurements have enormous potential in practical applications when designed as part of a conditioning circuit by carrying out pH sensing and signal processing simultaneously.

4.1.2 ISFET weak inversion current mirror

In this section, we set out with the aim to design a circuit for Methylation-Specific PCR with the least number of transistors. Clearly, a minimum of two devices are required for differential sensing. In CMOS, a current mirror is used to copy currents and often used as a basic circuit building block. Though it is not traditionally intended for differential measurements when constructed with MOSFETs, any variations in gate-source voltages between the two devices would appear in the output current. Therefore, by replacing the transistors with ISFETs, we can achieve differential sensing by independently changing the pH of each device. As we learnt in the review of differential ISFETs, biasing the sensors in weak inversion has the potential for drift and thermal voltage cancellation. This is particularly suited to amplification based detection that requires thermal cycling over time. Furthermore, multiple output stages can be easily cascaded to potentially provide different sample detection or integrated as part of other conditioning circuits. Based on this, we propose an ISFET-based current mirror operating in weak inversion for MSP analysis, in which configuring ISFETs in a current mirror structure is novel.
**Circuit description**

Fig. 4.3 shows the circuit and sensor layout for the proposed ISFET current mirror. We chose to use n-channel ISFETs as we expect a drop in pH from positive amplifications to give an increase in current, as opposed to a decrease of current in p-channel ISFETs. For the device dimension, we know from section 3.3.1 that $\zeta$ should be small to minimise passivation attenuation, with $A_{elec}$ significantly smaller than $A_{chem}$. The output current of the circuit can be derived using the equation for chemical gate voltages (eq.4.1) and ISFET in weak inversion (eq. 4.2):

\[ V_{ref} - V_{G''} = V_{chem} + V_{tc} \]  
\[ I_{D(W)} = I_0 \exp \left( \frac{V_{G''} - V_s}{nU_T} \right) \]  
\[ I_{out} = I_{ref} \exp \left( \frac{2.3\alpha(pH1 - pH2)}{n} \right) \exp \left( \frac{V_{tc1} - V_{tc2}}{nU_T} \right) \]
Eq. 4.3 shows the output current becomes an exponential function of the difference in pH between the input ISFET (pH1) and output ISFET (pH2) with thermal voltage cancellation. If the pH and trapped charge on both ISFETs are the same, it operates as a conventional current mirror; if there is a difference in pH, any DNA reactions occurring at ISFET2 is positively amplified. Bias current $I_{\text{ref}}$ acts as a linear tunable gain, while trapped charge from the two devices appear as a differential gain if unmatched. Any common drift is expected to cancel as the ISFET sensors are in a differential configuration.

![Image of ISFET current mirror with multiple output stages for quantitative MSP](image)

Quantitative MSP could be achieved by cascading multiple output stages in the ISFET current mirror. Suppose the reference ISFET is treated with unmethylated primers, and the remaining output ISFETs are coated with methylated primers and different clones of the same target DNA. The ratio of mirrored branches showing positive and no amplification would represent the percentage of methylation for that site. For example, if 8 out of 10 output branches have an increase in current compared to $I_{\text{ref}}$, the percentage methylation is 80%.

**Circuit simulations**

The circuit was simulated in 0.35μm technology in Cadence Spectre, using the ISFET veriloga model reported in [10]. The ISFETs were implemented with $\alpha = 0.95$ and $\gamma = -196$ mV which gives $V_{\text{chem}} = 56$ mV/pH. First we want to confirm the circuit’s operation agrees with the derived equation for output current. Fig. 4.5a illustrates $I_{\text{out}}$ does indeed vary exponentially with differences between pH1 and pH2. Adjusting for drain voltage variations so that $V_{DS1} = V_{DS2}$, output current is shown to be directly
proportional with bias current in Fig. 4.5b. Threshold voltage shift caused by trapped charge in the device is simulated using DC voltage sources at the gate of the ISFETs. The positive exponential gain caused by difference in trapped charge shown in Fig.4.5c implies the circuit should be connected such that the output current is amplified via the device with a smaller trapped charge.

Figure 4.5: Simulated output current to different variables: (a) pH as exponential gain, (b) bias current as linear gain, (c) trapped charge as exponential gain
Passivation attenuation is a prominent non-ideal effect of ISFETs manufactured in unmodified CMOS. As mentioned in Section 3.3.1, we can model the effects of attenuation as a passivation capacitance $C_{pass}$. Based on its theoretical formula (eq. 3.28) and specifications of 0.35$\mu$m process, the passivation capacitor can be estimated to be 226fF for a chemical area of 100x100$\mu$m$^2$. However, the extensive experimental data published in [11] using ISFETs of various sizing combinations suggests $C_{pass}$ is in the range of 1fF or less when considering an electrical area of 1x1$\mu$m$^2$. Consequently, we simulated a range of values from 100aF to 1fF on the ISFET current mirror with results shown Fig. 4.6. We found the passivation capacitance to be inversely proportional to the gradient of the exponential curve relating pH to output current. A smaller capacitance contributes to a bigger voltage drop at the gate, resulting in a decrease in current and almost linearising the response. Maximising $C_{pass}$ will increase the exponential gain on pH.

Next we simulated the circuit’s pH response to a combination of trapped charge and passivation capacitance. Fig. 4.7a shows the output current for $C_{pass} = 1fF$, with same trapped charge in both devices. The absolute value of matched $V_{tc}$ was found to attenuate the exponential response through drain voltage variation. Since the gate voltage bias is determined by what is required to support the bias current on top of trapped charge, as $V_{tc}$ increases, it also affects the difference in drain voltages between the two devices. Clearly, the output current becomes negligible as the value of trapped charge becomes larger than the required gate voltage. If the trapped charges are unmatched (Fig. 4.7b), together with the passivation capacitance, a positive difference in $V_{tc1}$ and $V_{tc2}$ seems to linearise the response as the positive gain dominates, causing $I_{out}$ to increase by an order of magnitude relative to $I_{ref}$. Conversely, a negative
Analogue current mode design approach for methylation

Figure 4.7: Simulated pH response at $C_{\text{pass}}=1\text{fF}$ with, (a) matched trapped charges, (b) unmatched trapped charges

**Experimental Setup**

The ISFET sensors were fabricated in AMS 0.35$\mu$m technology and connected externally as the proposed current mirror. Dimensions of the devices were designed to minimise $\zeta$, with $(W/L)_{\text{elec}} = 1\mu m/1\mu m$, and $(W/L)_{\text{chem}} = 100\mu m/100\mu m$. The silicon die was bonded and encapsulated with glob top epoxy on PCB dipsticks (See Appendix C). As a proof of concept, the current mirror was tested using two microchips in separate beakers connected via two Ag/AgCl reference electrodes. A semiconductor parameter analyser, Keithley 4200 Semiconductor Characterization System was used to characterise and provide bias current for the ISFETs via a generic test board.

Since each beaker containing a dipstick and reference electrode effectively forms a voltaic half cell, a salt bridge was required to maintain charge neutrality. Acting as an ionic conductor, it allows the transfer of positive and negative ions to achieve a zero net charge. For our experiments, we used a U-shaped tube sealed with porous frits containing an inert electrolyte such as potassium chloride (KCl). The positive potassium ($K^+$) and negative chloride ($Cl^-$) ions move into the solutions to balance the charges. A high concentration of KCL was chosen to minimise junction potential and effects of different diffusion rates. Alternatively, when miniaturising the set up, micro-fluidics with separate chambers can be partitioned with a porous material to act as a salt bridge channel (See Appendix E).
Experimental Results

The pH response of the current mirror was obtained by varying pH on the output ISFET with the input ISFET kept at pH=7. Fig. 4.9 compares the measured and simulated results for different bias currents. From the device characteristics sweep of reference electrode voltage in Fig. 4.10, a difference of $\sim +0.5V$ trapped charge can be observed between the input and output ISFETs, and passivation capacitance appears to be similar. As noted from simulations, ISFET with the smaller trapped charge (ISFET2) should be configured as the output device in a mirror configuration to provide positive amplification of the pH signal.

The combination of unmatched trapped charge with passivation effects causes $I_{out}$ to appear more linear than exponential. Comparing results from different bias currents, the shape and magnitude of $I_{out}$ shown in Fig. 4.9a and Fig.4.9b suggest the passivation capacitance on the fabricated chip is between 1fF to 500aF, with an average gain of 117nA/pH and 52.1nA/pH respectively. In Fig. 4.9c, the output current is very small with a gain of 1.7nA/pH and well below simulated values, which implies the effects of
attenuation is relatively bigger at lower bias currents.

Figure 4.9: Comparing simulated and experimental pH response at different bias currents: (a) 100nA, (b) 50nA, (c) 10nA
Plotting output against bias current, it is clear that $I_{ref}$ indeed operates as a linear tunable gain for the current mirror. Fig. 4.11a shows the gain is bigger than 1, likely due to trapped charge gain and drain voltage variation. Despite attenuation at low bias currents, pH sensitivity of the circuit remains consistent. The values shown in Fig. 4.11b is similar, with an average sensitivity of $27.4mV/pH \pm 4mV$. The reduction from ideal Nerstian response is expected in unmodified CMOS fabrication, however the value is lower than generally reported in the literature [12–14].

Power consumption

Power consumption of the circuit is dominated by the magnitude of output current $I_{out}$ as a result of gain. Subsequently, overall power consumption is proportional to pH, bias current, readout voltage as well as passivation attenuation and differences in DC
offsets between the devices. In this experiment, we biased the output at supply voltage of 3.3V to ensure $V_{DS} > 4U_T$, so the circuit is always in weak inversion saturation. Table 4.1 shows the power consumption for various bias current with input ISFET at pH7 and different output pH. In general, the power consumed is within a few $\mu W$, with a maximum of 3.34$\mu W$ at 100nA bias current.

<table>
<thead>
<tr>
<th>Ibias/nA</th>
<th>Power consumption/$\mu W$ @ pH4</th>
<th>@ pH7</th>
<th>@ pH10</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3.15</td>
<td>1.93</td>
<td>0.83</td>
</tr>
<tr>
<td>50</td>
<td>3.23</td>
<td>2.02</td>
<td>0.91</td>
</tr>
<tr>
<td>100</td>
<td>3.34</td>
<td>2.12</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Table 4.1: Power consumption at different bias current and pH

**Drift noise**

Drift between a single ISFET device and current mirror configuration was investigated. The duration of PCR and LAMP amplifications are typically around 30-40 minutes, so we ran the experiments in pH7 solution for 60 minutes. Fig. 4.12 shows the normalised drift in the two configurations: an average of 9nA/min observed on the single-ended ISFETs is significantly larger than the same devices connected as a current mirror (0.22nA/min). To investigate the effects of the salt bridge, we connected ISFETs from the same microchip as a current mirror. This way, we can bias the dipstick using the

![Figure 4.12: Graph showing drift from a single-ended device and current mirror](attachment:image.png)
same electrode and pH solution beaker without the need of a salt bridge. The initial
rate of drift in the ISFET current mirror with salt bridge is bigger, but quickly settles to
a value close to the starting current, with the rate of change similar to the configuration
using the same electrode.

The noise analyses of the mirror and single-ended device are derived by applying
FFT on the drift data. Difference in smoothness of the noise signals shown in Fig.
4.13 is a result of varied sampling frequencies in readings. All drift data was sampled
every half a minute, with the exception of the ISFET current mirror with salt bridge
measured at every second. As expected, noise is dominated at low frequencies by 1/f
noise, and the noise floor of single-ended ISFETs are higher than in a current mirror
configuration. The use of salt bridge adds noise by a small amount, likely as a result
of the transfer of ions for charge neutrality. Based on the integrated noise values,
minimum detectable pH can be deduced for each configuration (Table. 4.2). The pH
resolution of a single device over an hour period is well over 1pH due to drift, while
the proposed ISFET current mirror has 0.3pH resolution. As demonstrated in [15], up
to 1pH change can be expected in a pH-PCR experiment. Therefore the resolution of
an ISFET current mirror is more than adequate to detect the amplification results of
from Methylation-specific PCR.

![Noise analysis derived from drift of single-ended device and ISFET current mirror](image)

Figure 4.13: Noise analysis derived from drift of single-ended device and ISFET current
mirror
Table 4.2: minimum detectable pH based on integrated noise

<table>
<thead>
<tr>
<th>Resolution /pH</th>
<th>ISFET1</th>
<th>ISFET2</th>
<th>ISFET mirror (same electrode)</th>
<th>ISFET mirror (salt bridge)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution /pH</td>
<td>2.65</td>
<td>1.85</td>
<td>0.04</td>
<td>0.26</td>
</tr>
</tbody>
</table>

**Temperature**

Similarly, we tested the temperature drift between a single device and current mirror to validate thermal voltage cancellation shown in eq. 4.3. After leaving the microchips in pH7 solution for 60 minutes to settle temporal drift, we transferred the dipsticks from room temperature to beakers of solutions with the same pH heated to $80^\circ C$ to create a step change. The choice of temperature was based on an average of thermal cycling ($50^\circ C - 90^\circ C$) and LAMP reactions ($60^\circ C$). Fig. 4.14 shows the rate of drift increases with change of temperature for both sensor and current mirror. In comparison, the average value of device temperature drift (115nA/min) is more significant than that of the mirror configuration (4.4nA/min). However the increase in drift due to temperature when compared with itself is ten-fold in a single ISFET, and 15-fold in the current mirror. Since the term defining exponential trapped charge contains thermal voltage, so the circuit has maintains some temperature dependencies. Nonetheless, the temperature drift in the mirror is substantially small relative to the magnitude of output current.
4.1.3 Conclusion

In this section, we proposed an ISFET-based current mirror for detecting methylation with MSP. Methylation specific PCR uses amplification results with methylated and unmethylated primers to identify methylation at a targeted CpG site, naturally leading to differential detection. Using only the minimum number of devices required for differential detection, the circuit operates in subthreshold to provide three main features: exponential amplification relative to the difference in pH between two sensors; bias current as a tunable gain; drift and temperature cancellation. In addition, simple cascading of output stages could be used to provide quantitative MSP analysis.

Based on a combination of simulations and experimental results, the theoretical operation of the circuit and effects of the passivation layer was validated. Passivation capacitance contributed to attenuation of the exponential pH response; positive differences in trapped charge between the input and output ISFET can be used as an additional gain. The circuit was designed and fabricated to minimise $C_{pass}$ via ratio of electrical and chemical sensing area. Comparing the experimental results to simulations, we conclude a positive differential trapped charge can linearise the exponential response, with bias current as a linear gain. While pH sensitivity remain consistent for different input currents, it would seem that a large WI bias current is preferred to minimise attenuation in the circuit. Results comparing temporal and temperature drift between a single ISFET and as current mirror demonstrates the circuit’s ability for thermal voltage and drift cancellation, which is ideal for analyses requiring up to 60 minutes and thermal cycling. The minimum detectable pH derived from noise analyses was also in line with the value expected in amplification reactions.

Preliminary results of the ISFET current mirror suggest its suitability as a basic circuit building block for MSP detection. Through simulations, we can anticipate behaviour of combined non-ideal effects arising from the passivation layer, and provide estimation of the circuit’s response.

Table 4.3: Achieved Specifications for ISFET weak inversion current mirror

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>AMS 0.35 µm CMOS</td>
</tr>
<tr>
<td>Supply voltage</td>
<td>3.3V</td>
</tr>
<tr>
<td>Power Consumption</td>
<td>2.12 µW @ $I_{bias} = 100nA$ and pH7</td>
</tr>
<tr>
<td>pH gain</td>
<td>117nA/pH at $I_{bias} = 100nA$</td>
</tr>
<tr>
<td>pH sensitivity</td>
<td>27.4mA/pH</td>
</tr>
<tr>
<td>Drift</td>
<td>0.27nA/min</td>
</tr>
<tr>
<td>Temperature drift</td>
<td>4.4nA/min @ 80°C</td>
</tr>
<tr>
<td>Resolution</td>
<td>0.26pH @ $I_{bias} = 100nA$</td>
</tr>
</tbody>
</table>
4.2 Signal processing for Bisulfite-Sequencing

This section explores how to improve base calling accuracies in ion-semiconductor sequencing, notably to increase the length of detectable homopolymers. When combined with sodium bisulfite treatment, ISFET-based sequencing is an efficient technique for analysing methylation at individual CpG resolution. A major drawback of this approach is the reduced sequence complexity after chemical preparation, since any unmethylated cytosines are converted to thymines. The likelihood of the same base appearing consecutively, also known as homopolymers, therefore increases as there are less frequent variations of nucleotides in a sequence. The detection of homopolymerisms is characteristically challenging in ion-semiconductor sequencing in the first place due to diffusion of hydrogen ions between neighbouring wells. In addition to limiting the window of detection, its effect becomes more severe when trying to determine a high number of identical bases as the increments of pH decreases following the logarithmic relationship of hydrogen ions and pH.

![Figure 4.15](image)

Figure 4.15: (a) Incorporation signal model developed by Sohbati [25], based on [16], (b) Integral of base incorporation

With considerations to the above, we want to apply processing to better define and lengthen the output signal corresponding to the number of base incorporations. An integral represents the total amount of the measured quantity as it calculates area under the curve. Any fast varying amplitudes such as noise would have little effect on the final value, similar to a low pass filter. This form of averaging by summation could be used to improve accuracy in base calling generally. As well, real time integration is commonly used to convert waveforms, famously a square to triangular wave etc. Based on the shape of an incorporation signal published in the Ion Torrent sequencing
platform [16], it can be easily deduced that its integral would resemble the inverse of a
half cosine wave. To illustrate how integration can be applied to improve base-calling
accuracies, we used the model developed by Sohbati that simulates diffusion of ions
during base incorporations. Fig. 4.15 shows the Matlab reconstructed base call signal
with and without noise. Clearly, the integral is much less susceptible to noise and
effectively extends the duration of detectable amplitude. As a result, a more precise
distinction can be made on the two base calls.

4.2.1 Current mode integrators

Rationale behind designing an ISFET based integrator in current mode is primarily
due to its performance advantages [17]. In current mode, signals are represented by
branch currents rather than voltages at each node. Compared to voltage mode designs,
dynamic range is not limited to power supply but based on the magnitude of current
sourced andsinked. With that, low power operation is easily attainable by biasing
in weak inversion and limiting total currents. Since each capacitive nodes are already
charged, current mode circuits are fast and can achieve high bandwidths. Moreover,
analogue signal processing favours current mode architectures as computations can be
readily synthesised using the translinear principle.

Within the current mode domain, there are many available integrator topologies
such as Gm-C, current conveyor based, switched current and log domain filters [18].
Considering our application, possibility to assemble as part of a pixel is desirable for
sequencing. This implies the design of the circuit should aim to minimise the number of
transistors and silicon area. Also, with the time of a DNA reaction lasting for at least
more than 1s on average, implementation of an integrator as first order filter would
require very low and likely to be sub-Hertz cut off frequencies. Therefore, topologies
like a current conveyor integrator is not suitable as it comprises a current mode op amp
as well as active components that would demand a large area to achieve the required
frequency specification.

Gm-C, or OTA based integrators are effectively RC filters with transconductance
implemented by MOSFETs to be more CMOS compatible. The transconductance stage
of the filter is often realised using an OTA (operational transconductance amplifier). A
typical OTA structure can be seen in Fig. 4.16, consisting of a differential pair with a
current mirror. Following Kirchoff’s current law, the output current is a result of any
small signal difference between the positive and negative voltage inputs $V_{in+}$ and $V_{in-}$
respectively. An equation for output voltage $V_{out}$ can therefore be derived with $gm$
as the overall transconductance of the OTA circuit, and $C$ as the output capacitor:
An option to adapt this configuration as an ISFET-based integrator would be to use the pH sensors as the differential pair, similar to in [9]. Though as mentioned in the said publication, the intrinsically high gain of differential configuration amplifies any mismatch in DC offset between the two devices. As well, differential pairs are notorious for their very small linear range of $gm$. Any techniques to extend the range would require either more transistors or resistors, with limited improvements in results. For a low corner frequency, the circuit would require a small overall transconductance with a large output capacitance. Although it is possible to reduce the value of $Gm$ via transistor sizing and bias current, sub-Hertz filters would only be achievable with a substantial output capacitor.

$$V_{out} = \frac{1}{sC}(V_{in+} - V_{in-})gm$$ \hspace{1cm} (4.4)

The theory of log-domain filtering [20] is a product of the translinear principle with application of companding theory. The log-domain integrator operates in weak inversion and its circuit can be considered as three stages: the input current is logarithmically compressed to the voltage at the left current mirror node; linear integration is carried by the capacitor; the resultant output current is an exponential expansion of the capacitor voltage. This process of companding integration has the benefits of increased SNR since the signal is amplified whilst the noise floor is suppressed. In [19], Georgiou et al. demonstrated the potential of adapting ISFETs into a log-domain structure. In the first instance, the input transistor was replaced by an ISFET as shown in Fig. 4.17. Simulated results showed the gain of the frequency response increased with ion concentration. In the second configuration, bias current $I_D$ was supplied by an ISFET. This resulted in a tunable filter whose gain and cut off frequency were inversely and
directly proportional to ion concentration respectively.

Switched current (SI) is a discrete time analogue signal processing technique first proposed by Hughes et al. in 1989 [21]. It can be viewed as the current mode equivalent of switched capacitor circuits, a popular topology for filters due to their insensitivity to absolute component values\(^1\). Utilising the inherent MOSFET gate-source capacitance to store charge, switched current circuits can implement digital filters using a combination of switches and transistors to form delays. The basic building block of all SI circuits is a memory cell, based on the design of a simple current mirror shown in Fig. 4.18.

\[^1\text{Frequency response in switched capacitor networks depend on capacitor ratios}\]
Switch opens in period \((n - 1)\), current through transistor M1 is a summation of the input \(i_{in}\) and bias current \(I_B\).

\[
I_{D1}(n - 1) = I_B + i_{in}(n - 1) \tag{4.5}
\]

Switch closes in the following period \(n\), drain current of the input transistor is mirrored to the output scaled by sizing factor \(A\). As the bias currents cancel when \(A=1\), the output current \(i_{out}\) is therefore a delayed version of the input.

\[
i_{out}(n) = i_{in}(n - 1) \tag{4.6}
\]

From the SI memory circuit, it can be seen that further delays can be formed by a cascading multiple cells [22]. In essence, any discrete time function such as integration and differentiation can be implemented with delays, summation and feedback. This ease of design and compatibility with the digital domain makes switched current a versatile signal processing technique. And as we demonstrated in the last section, it is possible to create an ISFET-based current mirror. Thus for continuity, we propose the novel application of integration to improve base calling accuracy, by altering an existing switched current integrator to include ISFET sensors as part of its circuit. Results of the new architecture are presented in the following sections.

### 4.2.2 ISFET based Switched Current integrator

The proposed circuit is based on the design of a switched current integrator presented in [23]. Composing of n and p channel current memory cells with weighted output

![ISFET based switched current integrator](image.png)

Figure 4.19: ISFET based switched current integrator
stages, the circuit carries out integration by summation of delayed input and output currents from the previous phase. Typically, the small signal input current is provided externally at the input summing node. We propose implementing the initial n-channel memory cell with an ISFET and REFET pair to provide an input current related to the difference in pH. A diagram of the ISFET integrator is shown in Fig. 4.19.

![Figure 4.20: Signals of 1 incorporation and no incorporation from [24]](image)

To distinguish a base incorporation event in ion-semiconductor sequencing, background correction is required by comparing wells with and without DNA extension. The raw ISFET sequencing signals shown in Fig. 4.20 illustrates a well with no incorporation displaying a steadily decreasing pH, with the amplitude from incorporation being significantly smaller than the signal due to drift and diffusion of hydrogen ions. By configuring the input memory cell with ISFET and REFET, the differential pH measurement can be considered as background correction. The REFET can be achieved by not coating the device’s gate with any reagents, alternatively as [25] suggested, the well can be made small enough to not contain any DNA micro-beads. As incorporation occurs, the drop in pH from the release of hydrogen ions will induce small changes in current, denoted as $i_{pH}$.

**Circuit Operation**

The transfer function of the circuit can be derived based on the branch currents in each clock period. A diagram of the two-phase switch signals are shown in Fig. 4.21, along with a non-overlapping clock generation circuit made up of a NOR-gate flip flop and delays. Phase 1 and phase 2 make up one clock period and are control the switches $\phi_1$ and $\phi_2$. During phase 2 in $(n - 1)^{th}$ period, $\phi_2$ is closed, current in ISFET M1 is
Analogue current mode design approach for methylation

Figure 4.21: Diagram of non-overlapping clock signal and circuit

copied to REFET M2 via the current mirror, where $I_1$ is the sum of the bias current $I_B$, feedback current $i_f(n-1)$ and small signal pH current $i_{pH}$:

$$I_1(n-1) = I_B + i_f(n-1) - i_{pH}(n-1)$$  \hspace{1cm} (4.7)

Feedback current can be deduced by assuming $\phi 1$ was closed just before the start of phase 2, so $i_f(n-1)$ is a scaled version of the output current during phase 1:

$$i_f(n-1) = -\frac{B}{A}i_{out}(n-1)$$  \hspace{1cm} (4.8)

In the following period $n$, $\phi 1$ is closed and $\phi 2$ is open, so $I_2$ is held at the value from the previous period, and mirrored to the output branch by factor A.

$$AI_1 = I_B - i_{out}(n)$$  \hspace{1cm} (4.9)

Substituting eq. 4.7 and 4.8 to eq. 4.9, we can write the output current as:

$$i_{out}(n) = Bi_{out}(n-1) + Ai_{pH}(n-1)$$  \hspace{1cm} (4.10)

And therefore the transfer function in the z-domain as,

$$\frac{i_o}{i_{pH}}(z) = \frac{Az^{-1}}{1 - Bz^{-1}}$$  \hspace{1cm} (4.11)

Similar to the original switched current circuit [23], the transfer function corresponds to a non-inverting integrator with the forward Euler rule $s = \frac{1-z^{-1}}{Tz^{-1}}$. The circuit is a lossless integrator for scaling factor $B=1$.

**Compensating for non-ideal effects**

To minimise several non-ideal effects of the circuit simultaneously, we added a capacitor at the node after each switch. Switching is the main source of error in sampled-data circuits. When the switch goes off, voltage from the clock transition is coupled through
the capacitive divider formed between the switch and output transistor. Charges escaping from on-phase of the switch are injected into the gate capacitance of the connected transistor, and thermal capacitive noise, $kT/C$ generated in the switches are stored similarly. The addition of a capacitor reduces the magnitude of error voltage produced. The value of the capacitor is designed such that the time it requires to charge the node at the transistor’s gate is smaller than the clock period to avoid any settling errors, since the transconductance should be made relatively small to minimise errors in the drain current from gate voltage variations. However this is not a difficult specification to fulfill as the length of incorporation reactions do not demand for high frequency operations. The use of transmission gates can also reduce charge injection to an extent through charge cancellation depending on the value of the voltage.

Other non-idealities can be summarised by mismatch, drain voltage variation and transistor noise. Cascode current mirror configuration with appropriate sizing was adopted to reduce effects of channel length modulation and percentage error arising from fluctuations in fabrication. The nature of switched current memory cells provide measures against 1/f noise through correlated double sampling [26]. When a switch is open between a current mirror, the noise current ($i_{n1}(t)$) at the input transistor is sampled. When the switch closes, the noise currents from the output and input transistors are subtracted:

$$i_{out}(t + \frac{T}{2}) = i_{n2}(t + \frac{T}{2}) - i_{n1}(t) \quad (4.12)$$

So any low frequency noise is canceled through correlated double sampling. The cancellation is more effective in a single transistor memory cell configuration where the diode connection at the gate is switched. Nonetheless, if the transistors are well-matched in the mirror memory cell, the switching will provide a degree of noise protection.

![Figure 4.22: Modeling noise in SI memory cell](image-url)
4.2.3 Fabrication and Testing

The ISFET integrator was fabricated in AMS 0.35\(\mu m\) technology using the specifications shown in Table 4.4. Two generations of the circuit were manufactured on chips named “Optimus” (Fig. 4.23a) and “Batman” (Fig. 4.23b). Optimus contains only the basic design of the ISFET integrator without cascode. In this version of the chip, ISFET sensors could not be accessed separately for characterisation, which made debugging very difficult. Trying to generate well-defined non-overlapping clock phases externally was also problematic. In Batman (Fig. 4.23b), we included multiplexers to interchange ISFETs and MOSFETs into the circuit, allowing for both electrical and chemical testing. Particularly, we considered a design that would allow the use of two separate dipsticks for switching and biasing the input memory cell in different pH solutions, like the ISFET current mirror in Section 4.1.2. The sensors were positioned away from the bond ponds which was necessary for encapsulation. Additionally, they were also placed sufficiently apart to allow for possibility of miniaturising the chemical set up by adopting a two-chamber manifold such as the one described in [27].

<table>
<thead>
<tr>
<th>Table 4.4: Fabrication Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
</tr>
<tr>
<td>Supply Voltage</td>
</tr>
<tr>
<td>Transistor size /(\mu m)</td>
</tr>
<tr>
<td>ISFET size /(\mu m)</td>
</tr>
<tr>
<td>sensing area/ (\mu m)</td>
</tr>
</tbody>
</table>

Figure 4.23: Micro-photographs of two fabricated chips with annotated dimensions in \(\mu m\): red boxes represent ISFET sensors; black outlines contain integrator circuits
Common centroid layout was applied to all current mirrors to minimise mismatch. A test board was manufactured for convenience in carrying out different modes of testing. It includes various types of connectors such as FFC or BNC to convert signals from different sources. Electrical testing can be carried out using the socket for dry packaging of the die; whilst chemical testing is done either by dipstick via ribbon cable, or a different PCB that slots into the test board with a flow cell on top. A flow cell (See Appendix D) allows for easy change or infusion of different pH in small volumes compared to beakers: a syringe is used to deliver the pH buffers into the chamber with an outlet for waste. A chlorinated silver electrode is used to bias the solution in a flow cell.

![Flow cell setup](image)

4.2.4 Frequency Response

The frequency response of the integrator was first simulated in Cadence 0.35µm technology. The electrical version of the circuit was used with a current source at the input summing node to provide a small signal current and generate results for periodic AC response (PAC).

Fig. 4.25 shows the current gain of the circuit for different scaling factors using

---

2This is done by biasing two silver wires as anodes and cathodes in potassium chloride solution over a period of time at ~ 0.9V. The cathode wire will become coated in chloride ions appearing whiter, and used as a reference electrode.
a sampling frequency of 5Hz. As expected, for scaling factor B bigger than 1, the integrator is lossy and has a frequency response below 0dB. For unity scaling factor B and A bigger or equal to 1, the shape of the curves are similar to an ideal integrator’s transfer function $H(s) = 1/s$. Gain increases with scaling factor A, resulting in unity gain frequencies of 1Hz, 1.8Hz and 2.7Hz for A equal to 1, 2 and 3. A lossless and high gain integrator is preferred, with area as a trade off. Unity scaling factors were used in the fabricated circuit.

Similar to switched capacitors networks, the cut off frequency of the circuit can be adjusted using the clock. Fig. 4.26 shows the PAC response for various sampling frequencies. The clock periods 2s, 0.2s and 0.02s correspond to unity gain frequencies.
of 0.5Hz, 1Hz and 9Hz. According to Nyquist theorem, any sampling periods smaller than 1s would avoid aliasing assuming an incorporation time of around 2-3s.

![Graph showing AC response comparison](image)

Figure 4.27: Comparison of simulated and experimental AC response at different clock frequencies

The MOS-equivalent of the fabricated integrator circuit was used to generate its frequency response. Keithley 6221 DC and AC current source was used to provide different types of small signal input current, and the output current was measured using the voltage bias as a current monitor unit on Keithley 4200 semiconductor characterization system. The measured frequency response was deduced from transient waveforms of the output current to different input square wave frequencies. Fig. 4.27 shows a comparison of the AC response with the previous simulations for 3 different sampling periods. Measurements at certain frequencies were not possible due to the Keithley’s sampling capacity and Nyquist restriction between input and clock periods. The graph shows a positive correlation between clock frequency and gain as expected. The shape of the measured response is similar to the simulated results with a higher gain of about 10dB. The increase in measured gain could be due to harmonics that exist in a square wave input.

4.2.5 Base calling performance

Above all, we are most interested in the transient response of the proposed ISFET-REFET integrator to investigate its performance in base calling. The drop in ISFET gate voltage due to pH was implemented using veriloga component from [10], with parameters $\alpha = 0.95$ and $\gamma = -196$ mV to give $V_{chem} = 56.5$ mV/pH. We modeled the
input signal in Cadence with the positive cycle of a sine wave and decreasing amplitudes of pH relative to the number of repeated bases. pH values were computed based on \( \text{pH} = -\log[H^+] \), with 0.02pH as the first incorporation change \([16]\). The DC value for no base calls was set to be pH8, similar to the buffers used in pH-sensitive amplifications \([15]\). Fig. 4.28 shows the pH signals for 0 to 9 incorporations.

### Clock frequency

![Figure 4.28: Ideal input signal modeling 0 to 9-nucleotide incorporation events](image)

Fig. 4.29 shows the integrator’s response to the simulated pH signal for 2 different clock frequencies, 5Hz and 20Hz. In general, the output resembles the curves initially simulated in Matlab (Fig. 4.15b), with the final integral value remaining constant when the reaction has ended. We computed the gain per pH from each base call simulation and obtained an average and standard deviation (\( \sigma \)). For sampling frequency of 5Hz,
the gain is $2.4 \mu A/pH$ with $\sigma \pm 50nA$ and relative standard deviation RSD=4.2%; for 20Hz, gain is $6.4 \mu A/pH$, with $\sigma \pm 0.7 \mu A$ and RSD=22%. As expected the gain is higher at a faster clock frequency, however the standard deviation relative to mean also significantly increases, most likely due to an increase of switching errors.

**Simulated output response**

Next, we explored the effects of noise on the accuracy of base calling. Fig. 4.30 shows the input pH signals with random gaussian noise added on top. Clearly, noise reduces the number of bases that can be determined as the signal peaks become overlapped with different incorporation events. We simulated this in Cadence and found the integrator is able to retain similar gain and standard deviation for noise as big 0.01pH, which is half the magnitude of the first base call as shown in Fig. 4.31. Though the nature of low pass filtering means it is more prone to low frequency noise, where it appears as artifacts in the output current.

![Figure 4.30: Modeling 0 to 9 repeated base calls with peak-to-peak noise amplitude of 0.01pH](image)

The simulated proposed integrator has shown its ability to remove limitations on the window for detection and robustness against noise. Identification of bases is no longer restricted to signal peaks as the circuit provides constant output levels that corresponds to the number of incorporations. In practice, the difference between each integrated base call may in fact be bigger as ions take longer to diffuse away when there are more base extensions. So far, simulations have shown the circuit is capable of accurately identifying up to 9 bases even in the presence of noise, whereas Ion Torrent’s base calling accuracy has been largely limited to 6-base length [24].
Figure 4.31: Simulated output response for 0 to 9 repeated base calls, with 0.01pH noise at different frequency spectrums.
Electrical testing

Before moving onto testing base call response chemically on-chip, we want to ensure a linear correlation exists between the input and output amplitudes electrically. Based on results from transient simulations and measured frequency response, we biased the circuit at an appropriate clock frequency to input. The input current period was set to be 2s, in line with the length of a typical nucleotide incorporation in ion semiconductor sequencing [16]. Therefore the clock period was chosen to be 0.2s, to maximise gain and minimise standard deviations. Fig. 4.32a shows that an increase in small signal input current is directly proportional to output, so it would be reasonable to assume the same applies for pH changes when ISFETs are substituted into the circuit.

![Graph of measured output current to different amplitudes of input currents](image1)

![Graph showing detection limits based on measured standard deviations](image2)

Figure 4.32

We can speculate the theoretical limitation on the number of base calls depending on variations in measurements. Repeated readings of $i_{out}$ gave a relative standard deviation of 2.4% at $i_{in}=100\text{nA}$. Since the output is linearly proportional to the input, we can calculate the RSD on simulated integrals of the pH incorporation signals and examine the overlaps. From Fig. 4.32b, it would appear that the standard deviations at bases 8 and 9 are slightly overlapping, and very close between 7 and 8. Evidently this would result in some inaccuracies in base calling at those lengths.

To test the base call response of the circuit electrically, nucleotide incorporations were emulated by modulating $i_{in}$ as shown in Fig. 4.33. This was achieved by using Matlab to simulate the signal as input to the integrator through a micro-controller’s DAC and resistor. The measured output of the integrator is shown in Fig. 4.34. Unlike the simulated response, output currents shown here are not at constant levels after the reaction because the output of the DAC is not equal to the voltage at the input.
summing node, so there exists a DC input current when the signal ends. In chemical testing, we expect the integrals to remain constant as the input is provided by pH changes in the ISFET-REFET pair. With this in mind, the integrated peaks show clear distinctions between each base call. The measured standard deviations from 10 repetitions shows a general increase as $i_{out}$ increases as expected. In fact, the measured values appear to be slightly better than theorised in Fig. 4.32b considering no overlaps occur at 8 and 9 base calls.

Figure 4.33: Input current to emulate 0-9 repeated base calls

Figure 4.34: Electrical output response to 0-9 repeated base calls
Trapped charge

Prior to testing the integrator circuit with the proposed ISFET-REFET inputs, the sensors were characterised. We found voltage offsets due to trapped charge were significantly large across 10 chips, with some exceeding 20V. This could be due to sizing of the devices, since more charges can be trapped in a bigger area of polysilicon. In contrast, the 1x1\(\mu\)m ISFETs used in Section 4.1.2 have trapped charge typically in the region of 1 to 3V, with a maximum of 7V. Based on [28, 29], we tried to remove some of the trapped charge using UV radiation from an EPROM eraser. Fig. 4.35 shows the \(I_D - V_{GS}\) characteristics of an ISFET after being exposed to different durations of UV.

![Figure 4.35: \(I_D - V_{GS}\) characteristics of ISFET after UV exposures](image)

Fig. 4.35 shows trapped charge is indeed reduced after each UV exposure, though it presented 2 main problems. Reliability was an issue such that sometimes even with repeated exposure, trapped charge did not decrease. Secondly, there appears to be a limit at how much trapped charge can be removed. This is in agreement with findings from [28] where no further reduction in offset could be achieved past an equilibrium value. The lack of openings on the top metal to expose the polysilicon to UV radiation also limited charge removal to the passivation layer only. To further decrease the threshold voltage, we would need a substrate contact to bias the bulk positively during exposure.

Following on, we tried several methods to compensate for the effects of trapped charge. When designing the chip, we made the source terminals of the ISFETs available for external biasing to adjust the gate-source voltages. However the magnitude of
trapped charge meant a large negative source voltage that exceeds -0.7V would be required to provide sufficient $V_{GS}$. Attempts to bias the ISFETs at saturation using external non-inverting gain amplifiers between the circuit and reference electrode did not succeed. As we discovered at re-simulation, the trapped charge in the ISFETs must be less than or equal to zero, and matched exactly in order for integration to take place. This was tricky in practice as only one of the sensors’ terminals was available. In future, we should consider calibration of the ISFET-REFET by disabling the feedback loop and closing all the switches simultaneously, where trapped charge is matched when the output current is zero.

**ISFET as external input sensor**

As an alternative, we used a smaller external ISFET to provide the input to the integrator for proof of concept. Fig. 4.36 shows the set up configuration. Using reference electrode voltage to simulate pH during incorporation, the changes were read out via the source of the ISFET with drain voltage and current held constant. An input resistor $R_{in}$ was used to convert the voltage to a small signal input current.

![Figure 4.36: Diagram illustrating set up using ISFET as input to integrator circuit](image)

Proper biasing of the input ISFET is essential to optimise the integrator’s performance. Depending on the value of the trapped charge, the range of source voltage can vary according to the current that the device is biased at. Trapped charge can vary from device to device, or within the same ISFET after repeated use. We can generalise differences in biasing into 2 scenarios: positive or negative DC offsets. Fig. 4.37 shows how source voltage varies with reference electrode voltage depending on different source
currents $I_S$. For $V_{tc} > 0$, the range of $V_S$ increases as $I_S$ decreases since source voltage is 0 for $V_{ref}$ below the required value; for $V_{tc} < 0$, range of $V_S$ increases with $I_S$ as it saturates to supply voltage for $V_{ref}$ bigger than the required value. Choice of source current also affects the DC biasing point for $i_{in}$. The range of operation should be set such that $V_S$ is biased around half of the power supply. The effect of passivation capacitance can be measured as the gradient for a given range of electrode voltage. The attenuation for this device is 0.7, with a geometry of $A_{chem} = \frac{(100\times100)\mu m}{(1\times1)\mu m}$. In comparison, device with $A_{chem} = \frac{(100\times100)\mu m}{(40\times40)\mu m}$ was measured to be 0.08, an order of magnitude smaller.

![Figure 4.37: Variations in source voltage to different reference electrode voltage with positive and negative trapped charge](image)

**Input resistance**

The magnitude of the input current also depends on the value of input resistor. A small resistance will give a bigger $i_{in}$, but at the risk of generating a pole at an undesired frequency. Fig. 4.38 shows the integrator’s output to an input square wave at different resistor values. According to the frequency response measured in Fig. 4.27, an input period of 8s should correspond to a current gain of around 24dB. A $5\Omega$ resistor gives a gain of 20dB, $R_{in}=1.2\Omega$ and $560k\Omega$ give a gain of 17.6dB and 13.3dB respectively. Moreover, at 1.2$\Omega$ and $560k\Omega$, output current resembles more like a low pass filtered square wave than a triangular wave. Evidently, the attenuated gain and distorted waveform suggests the minimum value of input resistor should be $5\Omega$ to ensure the circuit’s time constant is sufficiently large compared to the period of the input signal.
Figure 4.38: Effects of input resistance $R_{in}$: 5MΩ (top); 1.2MΩ (middle); 560kΩ (bottom), red trace represents input square wave, blue trace is integrator’s output current

**Chemical testing**

The final measurements of the circuit tests the integrator’s limit for detection. Since Keithley semiconductor characterisation system is prone to noise and limited sampling rate, we biased the circuit with a battery and the output current was read directly through a 16bit ADC to Matlab. Varying the input ISFET’s reference electrode voltage, we emulated 0 to 9 nucleotide incorporations in amplitudes relative to different maximum voltage limits. The raw signals from the readout circuit contain noise from ADC conversion, so we applied a moving average filter in Matlab to smooth out the output. Results from 12 repeated measurements at each voltage range are shown in Fig. 4.39, with mean and standard deviation plotted at a time when the maximum integral occurs on average.

Clearly, the integrator’s base calling accuracies are correlated with input amplitudes. The output corresponding to electrode voltage ranging from 0 to power supply is well defined for all base calls up to 9. The response for input up to 1.65V remain distinct, with occasional errors and overlaps occurring at incorporations of 7 bases or more. In these voltage ranges, we see maximum $i_{out}$ occurring later with more bases since the
tail of the reaction is longer. Standard deviation also increases with base repetitions as expected. The performance starts to deteriorate significantly as the input drops to 1V and below, with the range of output current falls to 2.5nA or less. At 1V input, the integrator can still identify the first 4 bases clearly (Fig. 4.39b), and 2 bases at 0.8V (Fig. 4.39d). The output is practically indistinguishable and dominated by noise as $V_{ref}$ range falls to 0.5V in Fig. 4.39f. Evidently, the fabricated circuit cannot integrate voltages equivalent to practical pH values in sequencing.

![Graphs showing output response at various input ranges](image)

Figure 4.39: Measured output response at various input ranges
Noise

Noise is a major factor on the circuit’s ability to correctly identify nucleotide incorporations. Simulated output electrical noise from Cadence shows the spectrum is primarily dominated by flicker noise, with a noise floor of 1.35pA. We estimated the minimum detectable pH based on the integrated input referred noise ($0.38nA_{RMS}$), and dividing by the transconductance of the input transistor and pH sensitivity. We used the average measured value of pH sensitivity from the fabricated chips ($\sim$30mV), and obtained a pH resolution of $4\times10^{-4}$pH. This corresponds to a SNR of 34dB when applied to the first nucleotide incorporation, reported to be $\Delta pH=0.02$ in [16]. Considering the theoretical value of pH change from 8 to 9 homopolymer lengths is 0.003pH (see Table. 4.5), the simulated resolution is more than capable of detecting at least 9, if not more base incorporations.

![Output noise spectrum](image)

Figure 4.40: Simulated output current noise spectrum

<table>
<thead>
<tr>
<th>Base call</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>theoretical $\Delta pH$</td>
<td>0.020</td>
<td>0.032</td>
<td>0.040</td>
<td>0.046</td>
<td>0.052</td>
<td>0.056</td>
<td>0.060</td>
<td>0.063</td>
<td>0.066</td>
</tr>
<tr>
<td>Measured pH resolution</td>
<td>0.043</td>
<td>0.044</td>
<td>0.042</td>
<td>0.046</td>
<td>0.047</td>
<td>0.046</td>
<td>0.048</td>
<td>0.050</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Table 4.5: Table of theoretical change in pH and measured resolution for each base call

Naturally, the circuit’s noise performance is worse in practice with more sources of contribution. Using the raw signal of the base call response, such as the one shown in Fig. 4.41, we obtained the transient noise for all measurements by subtracting the data with the average signal. Subsequently we were able to compare the simulated noise with experimental power spectrum by applying FFT. Clearly, the measured noise is significantly larger than simulations, by 3 to 4 orders of magnitude. Also, flicker
Analogue current mode design approach for methylation

noise no longer dominates in comparison. Calculating the measured integrated noise suggests the integrator has an average pH resolution of 0.047, with small proportional correlation to number of bases possibly due to higher RSD at larger amplitudes (Table. 4.5). This resolution is notably worse than the simulated value, and cannot detect even the first base call. The decrease in measured resolution explains the reduction in detection limits when factoring in chemical as well as additional readout circuit noise.

Figure 4.41: Measured output response of integrator without signal smoothing

Figure 4.42: Measured output voltage noise spectrum

4.2.6 Conclusion

In this section, we proposed an ISFET switched current integrator for detecting homopolymerisms in bisulfite sequencing with ion semiconductors. Based on theory and
Analogue current mode design approach for methylation simulations, integrating the differential pH signal from incorporations can remove the window for detection by maintaining constant output values depending on the duration and magnitude of input. As well as improving the number of base calls, averaging through integration can provide robustness against noise.

Trapped charge in the fabricated system proved to be challenging for testing. Large device dimension were chosen to minimise mismatch, but consequently contributed to a huge accumulation of trapped charge which far exceeded the range that could be calibrated by adjusting ISFET’s source voltages. Attempts to remove trapped charge with UV radiation and other methods were not sufficient to achieve matched DC offsets below or equal to zero in the input ISFET-REFET memory cell. As an alternative, chemical testing of the circuit was carried out with an external ISFET for proof of concept.

With pH of the incorporation signal imitated as reference electrode voltage, the electrical integrator circuit showed potential for improving the number of base calls with limitations to input ranges. Despite simulations showing the circuit’s ability to integrate signals as small as 0.02pH, both in transient and noise analysis, the noise contributions from readout circuits and chemical gate were significant in preventing similar performance in practice. Further investigations into the measured pH resolution reveals limits of detection on the fabricated circuit to be below the value required for identifying the first base call. This limitation on the range of detectable input would exist also if the trapped charge was removed in the proposed integrator circuit with ISFET-REFET input memory cell. So far, focus had been on establishing the integrator circuit’s functionality. However as the experiments demonstrated, appropriate sensor calibration and design of read out circuits that can accommodate the range of detection were not trivial. Mechanisms for converting and reading current mode inputs and outputs should be considered as part of the system at initial stage of design. Nonetheless, the proposed integrator circuit was able to accurately carry out integration on various inputs and base incorporation signals, subject to their magnitudes.

Table 4.6: Achieved specifications for ISFET switched current integrator

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technology</td>
<td>AMS 0.35um CMOS</td>
</tr>
<tr>
<td>Supply Voltage</td>
<td>3.3V</td>
</tr>
<tr>
<td>Power consumption</td>
<td>$\sim 16.5\mu W @ I_{\text{bias}} = 1\mu A$</td>
</tr>
<tr>
<td>Clock frequency</td>
<td>5Hz</td>
</tr>
<tr>
<td>0dB input frequency</td>
<td>0.5Hz</td>
</tr>
<tr>
<td>pH resolution</td>
<td>0.47pH</td>
</tr>
</tbody>
</table>
4.3 Compressed targeted DNA methylation sequencing

Whilst many methods exist for methylation analysis, the majority of techniques are adapted from conventional sequencing and genotyping, with very few that have been designed specifically for methylation. In particular, bisulfite sequencing is the only method that can provide information at single CpG resolution. Considering methylation is only concerned with the location and number of methylated cytosines, sequencing every base seems to be overly labour intensive.

We propose designing a new protocol that is specifically optimised for methylation and ion semiconductor sequencing. The technique should aim to provide a comprehensive picture of where the methylation occurs, as well as density and quantitative information on targeted DNA strands, while reducing complexity of the sequencing process. In this section we describe the steps and investigate the feasibility of the proposal via simulations.

4.3.1 Compressed incorporations

The newly proposed protocol relies on the basis that pH change is proportional to the number of nucleotide incorporations for every type of base. As only cytosines from a bisulfite converted sequence are useful for indicating methylation, we can treat the other bases as generically the same. This implies we can incorporate the bases complementary to guanine, thymine and adenine into the target DNA at the same time, and therefore compressing the number of flows and washes that are normally required. The steps after bisulfite treatment and PCR are described below:

**Step 1.** Add dATPs\(^3\), dTTPs\(^4\) and dCTPs\(^5\) into the sample. This will trigger DNA extensions along the sequence until there is a cytosine.

\[
\text{ATATTAACGATCGGAATCGTATTGACGTTAGATAGT} \quad \text{H}^+ \times 10
\]

**Step 2.** When the reaction in step 1 has terminated, a wash follows to rid of any

---
\(^3\) adenine nucleotides  
\(^4\) thymine nucleotides  
\(^5\) cytosine nucleotides
unincorporated dNTPs. Then add only dGTPs\textsuperscript{6} to the target DNA which will bind only with methyl-cytosines left over from the bisulfite treatment. Then repeat step 1 and 2.

$$\text{5'}\text{ATATTAACGATCGGAATCGTATTGACGTAGATAGT} \text{3'}$$

As a result, there is effectively “double confirmation” in the procedures: termination of the reaction in step 1 implies the existence of a cytosine; and a positive result (decrease in pH) from the incorporation of dGTP in step 2 also confirms a cytosine in the sequence. Figure 4.45 is a representation of the final results, with X as a generic base such as G, T or A.

$$\text{5'}\text{XXXXXXX}C\text{XXXXX}C\text{XXXXXX}C\text{XXXXXXX}C\text{XXXXXXXXXX} \text{3'}$$

Evidently the main difficulty of this new approach lies in correctly identifying what appears to be long stretches of homopolymer sequences. As we proposed in the last section, an ISFET integrator has the potential to improve the length and accuracy of homopolymers detected in ion semiconductor sequencing. Regardless, there exists a limit to the resolution which we have to consider through simulations.

### 4.3.2 Simulations

To examine whether this new approach can be realised in practice, we look at the range, length and resolution of homopolymers that would be required. Taking the promoter region of the gene Rasal1 (See Section 2.2.3) as a test sequence, we simulated various combinations in which the sequence can be methylated. Then we measured the number of bases between methylated cytosines to get an idea of what should be expected. Randomising the locations of methyl-cytosines, we ran the simulations 1000 times for various methylation patterns.

\textsuperscript{6}guanine nucleotides
We used two different types of sequences for our simulations: Promoter and CpG islands. CpG islands have a high concentration of CG dinucleotides, and are typically found unmethylated inside Promoter regions. Promoter regions are linked to gene silencing often via hypermethylation. According to Ion Torrent, the read lengths of their sequencers range from 50-200bps [24]. Table 4.7 shows a breakdown of contents from each test sequence. The percentages of each base are fairly evenly distributed in the promoters, conversely CpG islands have a high concentration of Cs and Gs. The higher number of CpG sites in comparison to promoters are expected. In fact, the number of CpG found in Rasal1 promoter ranges from 1 to 6 every 200bp, whereas it ranges from 13 to 19 inside CpG islands.

First, looking at the 200bp promoter and CpG island sequences, the frequencies at which a particular length of homopolymer occurs in different combinations of methylcytosines are displayed in Fig. 4.46a and 4.46b. Full methylation represents the scenario when all CpG sites are methylated, half is when half of the sites are methylated, and Two-C is defined as the minimum methylation in these simulations where any two CpG sites are methylated. The distribution of homopolymers are concentrated in clusters when simulated in the promoter region. Since there are only 6 possible sites in which methylation can occur in the sequence, there are only a small number of possible combinations and lengths that can exist between the methyl-cytosines. In comparison, the distribution is very sparse in CpG island where there are 18 CpG sites in total. Fig.
4.46b shows the deviation in the number of bases is large particularly at lower density of methylation, where there are more combinations of pattern available.

It is reported that the Ion Torrent sequencing platform [16] can accurately detect homopolymers up to 6 bases long [24]. As the simulations show, the length of homopolymers go way beyond the known detection limit. Even with an integrator, there
can only be up to a few bases in improvements. Technically, we only require resolution that would allow us to correctly deduce the locations of methyl-cytosines. If the lengths of expected homopolymers in the target sequence are all distinctly different, i.e. 10 and 20 bases, such that the distribution is similar to in Fig. 4.46b, we do not require single base resolution after all.

Figure 4.47: Homopolymer distributions of sequences at 50bp
So far the simulations have shown a low number of CpG sites within a sequence can provide distinctly different lengths of homopolymers, while a high concentration in a long sequence would produce a widespread of bases. In theory, shorter sequences will provide a cluster-like distribution. Fig. 4.47 contains simulation results for promoter and CpG island sequences with 50bp. Both sequences show majority of expected homopolymers are distinctly different. Though for higher concentrations of CpGs, the number of bases between some homopolymers in the CpG island are inevitably closer together. Having said that, methylation analyses are typically only performed on known targeted sequences. So we can possibly combine prior knowledge of the sequence with results for the final analysis.

Figure 4.48: Simulated input and output response of homopolymers for 0 to 90 bases

Using the ISFET integrator proposed in the earlier Section 4.2.2, we examine limitations on the lengths and resolution of detectable homopolymers. Incorporation signals and corresponding integrator’s output for 0 to 90 bases with a 10 base resolution are shown in Fig. 4.48. There are considerable separations between output current levels up to 40 bases, after which they become very close together. Based on this, we can assume the signals for incorporations up to 40 bases at 10 base resolution will be clearly distinguishable.

Table. 4.8 illustrates an example of how we can use prior knowledge on the sequence to reduce the necessary resolution of homopolymers. Considering that the total number of methyl-cytosines can be determined by the frequencies of terminations and incorporations of step 1 and step 2 respectively, the methylation pattern can be deduced based on the combination of homopolymer lengths. As an example, we examine the case for Two-C methylation, since low densities of converted cytosines generate more possible combinations of patterns. For the promoter region with 50bp, three possible
Table 4.8: Table illustrating different patterns of homopolymers at 10 base resolution for promoter and CpG island sequences of 50bp

<table>
<thead>
<tr>
<th>colour code</th>
<th>=&lt;10</th>
<th>11-20</th>
<th>21-30</th>
<th>31-40</th>
<th>&gt;40</th>
</tr>
</thead>
</table>

### Promoter 50bp

<table>
<thead>
<tr>
<th>Pattern</th>
<th>H1 1st mC position</th>
<th>H2 2nd mC position</th>
<th>H3 3rd mC position</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>30</td>
<td>C1</td>
<td>2</td>
</tr>
<tr>
<td>U2</td>
<td>30</td>
<td>C1</td>
<td>11</td>
</tr>
<tr>
<td>U3</td>
<td>30</td>
<td>C2</td>
<td>9</td>
</tr>
</tbody>
</table>

### CpG island 50bp

<table>
<thead>
<tr>
<th>Pattern</th>
<th>H1 1st mC position</th>
<th>H2 2nd mC position</th>
<th>H3 3rd mC position</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>39</td>
<td>C4</td>
<td>3</td>
</tr>
<tr>
<td>P1</td>
<td>27</td>
<td>C3</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>C3</td>
<td>12</td>
</tr>
<tr>
<td>P2</td>
<td>13</td>
<td>C2</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>C2</td>
<td>26</td>
</tr>
<tr>
<td>U2</td>
<td>13</td>
<td>C2</td>
<td>14</td>
</tr>
<tr>
<td>U3</td>
<td>1</td>
<td>C1</td>
<td>41</td>
</tr>
<tr>
<td>U4</td>
<td>1</td>
<td>C1</td>
<td>39</td>
</tr>
<tr>
<td>U5</td>
<td>1</td>
<td>C1</td>
<td>27</td>
</tr>
<tr>
<td>U6</td>
<td>1</td>
<td>C1</td>
<td>13</td>
</tr>
</tbody>
</table>

Combinations exist for 2 out of 3 CpG sites (C1, C2 and C3) to be methylated. Assuming the signal for every 10 base are clearly distinguishable, we colour coded the expected homopolymers according to their 10-base resolution brackets. U1, U2, and U3 represent three unique colour patterns, and therefore each methylation pattern can be detected with an exclusive set of homopolymers with reduced resolution. Similarly for CpG island of 50bp, there are ten possible combinations for 2 out of 5 CpG sites to be methylated. The majority of combinations have unique patterns, with the except of P1 and P2. In each case, the third homopolymer H3 should be able to provide distinctions within the same pattern as 6-base homopolymers fall within the conventional detectable limit.

Based on these examples, it would appear to be feasible to use prior knowledge of the sequence with the new protocol to deduce methylation levels and locations. Notably, the target DNA region should be short (50bp) and fragmented or selectively amplified.
Analogue current mode design approach for methylation accordingly. The stretch of sequence should be strategically selected with appropriate number and locations of CpG sites to ensure every combination of methylation pattern can be detected using limited resolution of homopolymers.

A factor that will affect the accuracy of the results is the fidelity of the polymerase. This is defined as the error rate of reading each base in the sequence. For example, if a cytosine is wrongly incorporated by dATP, dCTP or dGTP during step 1, the lengths of the resultant set of homopolymers will be different. Table 4.9 shows a list of available polymerase. A commonly used DNA polymerase such as Taq has a very low fidelity, where the percentage error in reading sequences is high. Proof-reading polymerase

Table 4.9: Table showing percentage accuracy of various polymerase in different read lengths

<table>
<thead>
<tr>
<th>Polymerase</th>
<th>1 kb template</th>
<th>3 kb template</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taq DNA polymerase</td>
<td>68.40%</td>
<td>205.20%</td>
</tr>
<tr>
<td>Pyrococcus furiosus DNA polymerase</td>
<td>8.40%</td>
<td>25.20%</td>
</tr>
<tr>
<td>Phusion High-Fidelity DNA Polymerases (HF Buffer)</td>
<td>1.32%</td>
<td>3.96%</td>
</tr>
</tbody>
</table>

![Promoter 50bp](image)

![CpG Island 50bp](image)

Figure 4.49: Effects of incorporation errors due to polymerase on percentage of correct methylated cytosines
are often used to correct errors by excising the wrongly incorporated nucleotide and allowing for the correct dNTP to anneal. However this is not applicable in our approach as addition signal from the excision and corrected incorporation will create errors in the homopolymers. Adapting a high fidelity polymerase such as Phusion or Pyrococcus can in turn improve the accuracy in sequencing, lowering the error percentage to 1.3 or 8.4 percent respectively.

Based on the values from Table 4.9, we randomly inserted wrong bases into the promoter and CpG island sequences to simulate effects of Taq and Phusion polymerase. The simulations ran for 1000 times for each methylation pattern. The percentage of correct mC reflects the probability that a wrong incorporation occurs at a methyl-cytosine. The results show when using Taq polymerase, the average number of correctly incorporated cytosines fall below 50%, with a large deviation. By contrast, Phusion polymerase produce almost no error to the number of methyl-cytosines. This emphasises the importance of high fidelity polymerase when considering this new protocol.

4.3.3 Conclusion

In this section, we proposed to modify the current procedures for electronic sequencing to be better suited for methylation analysis. With the aim to reduce labour intensity, whilst still able to provide information at individual CpG sites, the new method compresses the extensions of bases other than cytosines. Since the incorporation signal for different types of base is the same in ion semiconductor sequencing, the results from adding all the dNTPs complementary to guanine, thymine and adenine simultaneously will appear as a homopolymer sequence. By identifying the number of bases in the hypothetical homopolymer, along with incorporation of dGTP confirms the location and existence of methyl-cytosine after bisulfite conversion. The biggest challenge for this technique would be limitation on length and accuracy in detecting these long stretches of homopolymers.

Exploring the feasibility of this new technique, we examined required lengths and resolution of detectable homopolymers through simulations. Taking the promoter region and a CpG island from Rasal1 gene as test sequences, findings from simulations showed the length of target sequence and concentration of CpG sites are related to the number of bases expected in homopolymers. A high concentration of CpG sites give a widespread distribution in homopolymer lengths as more combinations of methylated cytosines are available. More importantly, it is desirable to have shorter length sequences where only certain values of distinctly different lengths of homopolymers can reduce the required detectable resolution. Simulations by the previously proposed
ISFET based integrator suggest homopolymers up to 40 bases with 10 base resolution can be considered distinctly different. We illustrated the possibility of combining prior knowledge of the sequence to deduce methylation levels with unique patterns of distinctly different homopolymers for promoter and CpG island of 50bp. Since the method relies on the magnitude of signal released during incorporations, a high fidelity polymerase is essential for the process to prevent errors.

With careful considerations of the target sequence region and length, simulations showed the practicality of applying the new proposed protocol for optimising sequencing process for methylation analysis. To further validate its feasibility, experiments exploring the detectable limit of mixed nucleotide incorporations should be investigated.
4.4 Summary

Using CMOS design methodologies, this chapter proposed two circuits along with a novel protocol tailored for ISFET-based DNA-methylation detection and sequencing. Motivated by possible scenarios that methylation analysis may be integrated into future health care, namely Point-of-Care in routine check ups or therapeutics for chronic diseases, we focused on developing circuits for the best suited detection methods, Methylation-specific PCR and bisulfite sequencing.

Designing a differential sensing circuit became the focus for methylation-specific PCR, since the technique requires application of two oppositely designed primers. A brief review of various differential ISFET configurations was presented, within which the concept of a REFET was explained. As a result, an ISFET weak inversion current mirror was proposed to take advantage of biasing in subthreshold for drift and temperature cancellation using the minimum number of transistors. Through simulations and experiments, we confirmed the output current increased with positive difference in pH between the input and output devices. Particularly, bias current could be used as a tunable gain, while non-ideal effects like passivation capacitance and DC offsets affected amplitudes and shape of the pH response. Reduction in rate of temporal and temperature drift was evident when comparing experimental results with single ended ISFET devices, from which the minimum detectable pH was also deduced. Sufficient pH resolution and insensitivity to time and temperature over the range used in DNA amplifications suggest the proposed ISFET current mirror is a good candidate circuit for MSP detection, with possibility of quantitative analysis using cascaded output stages.

For bisulfite sequencing, we developed a signal processing circuit that performs integration to improve base calling accuracies. Particularly, repetitions of base are difficult to determine due to hydrogen diffusion in Ion semiconductor sequencing. We hypothesised that applying integration will increase SNR of the incorporation signal as well as lengthens the window for detection. Considering advantages of current mode operation and continuity from the ISFET current mirror, an integrator was proposed based on an existing switched current design with addition of an ISFET-REFET pair incorporated as part of the circuit. This new inherent differential sensing acts as background correction which is essential for identifying nucleotide incorporation events. Despite simulations of the circuit showing promising results, in practice we encountered two main issues. Firstly, large DC offsets on the ISFET sensors prevented the circuit to be biased correctly, even when reduced by UV radiation, the matching of trapped charge was essential for the circuit to carry out integration. Consequently, we used an external ISFET as an alternative to provide input for the electrical version of the integrator with
MOSFET substitutes. Secondly, chemical and additional readout circuits noise limited the base calling performance to large input amplitudes. Further analyses into the raw signal noise revealed significant deterioration of pH resolution from theoretical value. Even so, the proposed integrator can carry out integration functionally up to an input range of 0 to 1.65V in the current setup.

Lastly, to reduce the labour intensity of sequencing, we proposed a modified protocol specific for methylation and ion semiconductors. Since we are only interested in information on cytosines, specifically CpG sites that can become methylated, the protocol suggests introducing dGTP separately and incorporating the remaining three types of bases at the same time. This halves the number of sequencing steps required from 4 to 2, with signals from the simultaneous incorporation step detected as generic repeated bases. The incorporation of guanine and termination of base extensions from remaining non-complementary nucleotides will both confirm the presence of a methylated cytosine. Using a sample sequence, a previously identified gene associated with methylation in chronic kidney disease, Matlab simulations were carried out to determine the feasibility of the proposed protocol. With considerations to practical read length and achievable base call resolutions, we demonstrated the potential of applying the technique to short sequences with a limited number of CpG sites. This effectively restricts sequencing results to a few possible scenarios, in which carefully chosen regions of sequence can produce unique combinations of distinctly different lengths of base repetitions. With a reduction in the required base calling resolution, the proposed technique may be applied in practice if used with a high fidelity polymerase.
Bibliography


Chapter 5

On-chip threshold cycle evaluation and Early amplification

Part of this thesis investigates the potential of using miRNA regulation as a means for developing new chronic disease biomarkers and therapeutics. Specifically, miRNAs are inherently well-suited for non-invasive Point-of-Care applications due to its stability in circulating samples. As described in Section 2.4, quantification of miRNA or nucleic acids in general can be achieved through real time PCR. Due to the nature of fluorescent tube-based amplifications, efforts to downsize qPCR machines are limited to physical dimensions in the number of samples that can fit into the instrument. In contrast, the ISFET-based platform reported by [1] can carry out real time PCR on up to 40 samples with a microchip and reader the size of a credit card.

The next steps to developing a full-integrated microchip for miRNA quantification would be to include on-chip evaluation of threshold cycles. When considering disease progressions or diagnoses, relative quantification is more appropriate for comparing the difference between miRNAs that are epigenetically modified or unaffected by the condition. Regardless, both absolute and relative quantification requires $C_T$ definition. By integrating this evaluation onto the platform can greatly reduce the amount of data having to be transmitted off-chip, or even eliminate the need altogether by including computation of relative ratios. Measurements at each cycle in real-time PCR instinctively leads to sampled circuits. As well, each thermal cycle takes 5-10 minutes to complete, during which the time can be spent doing processing. While there are two common techniques, namely Fit point and Derivative method available for assessing threshold cycles, differentiation can be easily implemented in discrete time domain processing. Therefore in this chapter, we present the design and results of a digital
back-end system that computes $C_T$ based on the maximum of the second derivative qPCR signal.

One of the possible applications of on-chip back-end processing can be extended to detect amplification early. When using qPCR as a form of analysis rather than generating clones of the target sample, completing all the thermal cycles is not necessary if the required information can be extracted prior to the end. Since qPCR reactions are defined as the doubling of DNA copies at every cycle, we can potentially verify amplification before the plateau phase by identifying exponential characteristics in the curve. Based on this, we propose a logic system that estimates the rate of exponential increase using three data points to determine amplification in real time.
5.1 System for determining the threshold cycle, $C_T$

Design of the proposed system for evaluating the threshold cycle using the Derivative method is shown in Fig. 5.1. The technique carries out double differentiation and chooses its maximum as $C_T$ since it represents the fastest rate of change, therefore ensuring the point lies within the exponential phase. Notably, the advantage of this method is that $C_T$ is explicitly defined without a user interference, and only three consecutive data points are needed for computing the second order derivative. Implementation in hardware over PC also means computations can occur in real time, increasing the processing speed and reduce data required for transmission. Discrete time differentiation is commonly applied in FIR digital filters, based on the approximation equation:

$$\dot{x}[n] = x[n] - x[n-1]$$ (5.1)

where $\dot{x}[n]$ is the derivative of the function in the current period $n$, and $x[n-1]$ is the value of the previous sample. Using this principle, we can implement the Derivative method by cascading four simple processing blocks (Fig. 5.1). Suppose we sample at the end of each qPCR cycle, the second derivative of the signal can be calculated by passing through twice, a block that computes eq. 5.1. Subsequent block “max” then continuously checks and updates the maximum value. Meanwhile a counter is used to keep track of the number of cycles, which is also acting as the clock signal, and outputs the cycle number when a maximum occurs.

![Figure 5.1: Block diagram of the proposed on-chip system for computing $C_T$ using the Derivative method](image)

5.1.1 Fit point and Derivative method comparisons

A proof of concept of the proposed system was first simulated in Matlab. An ideal qPCR curve was constructed using numbers that doubled after each cycle to represent the exponential phase with 100% efficiency, then followed by a linear and plateau
phase. Using this method we generated qPCR curves that simulated different starting quantities, from 1 to 10-fold dilutions. Fig. 5.2 shows the comparison between threshold cycles obtained from both the Fit point and Derivative methods. In particular, two log-linear threshold values (threshold1 and threshold2) were selected to derive $C_T$ values in the Fit point method.

![Matlab generated qPCR curves](image)

![Fit point method](image)

![theoretical threshold cycles for various methods](image)

Figure 5.2: Graphs showing Matlab generated qPCR curves (top), Fit point method (middle) and comparisons of dilution curves (bottom)

The theoretical standard curve for a 100% efficient qPCR reaction has a slope that corresponds to -3.3 cycle change in a 10-fold dilution (See Section 2.4.2 ). Due to the nature of qPCR cycles, $C_T$ values are limited to integers and cannot display
small variations in starting quantities that corresponds to threshold cycles less than 1. Interpolating from the lines of best fit, we get dilution curve gradients of -2.9 for threshold1, -3.2 for threshold2 from the Fit point method and -2.8 cycles from the Derivative method. Evidently, the Derivative method produced a slope that has the largest deviation from theoretical value. On the other hand, discrepancies in \( C_T \)s from the Fit point method depending on the choice of thresholds in the log-linear curve are significant, especially when examining specific data points. For instance, threshold2 has the same \( C_T \) for 5 to 9 folds increase in starting quantity whereas threshold1 is able to give 1 cycle difference within that change. That being said, threshold1’s estimate for 10-fold \( C_T \) is further from the theoretical value. Clearly, accuracy of the Fit point method is very susceptible to the user’s selection. By contrast, the second derivative technique generates only one value of \( C_T \) for each dilution, and appears fairly evenly distributed across starting quantities. In addition, the 10-fold \( C_T \) calculated using the true data points at 1 and 10-fold dilutions is -3 cycles, which is the nearest possible integer to -3.3. All things considered, the Derivative method seems more suited for relative quantification in Point-of-Care applications where we want to minimise variability in results with the least amount of processing.

Table 5.1: Table comparing \( C_T \) obtained from Fit point and Derivative method

<table>
<thead>
<tr>
<th>10-fold ( C_T )</th>
<th>Theoretical</th>
<th>Fit point1</th>
<th>Fit point2</th>
<th>Derivative method</th>
</tr>
</thead>
<tbody>
<tr>
<td>line of best fit</td>
<td>-3.3</td>
<td>-2.9</td>
<td>-3.2</td>
<td>-2.8</td>
</tr>
<tr>
<td>data points</td>
<td>-3</td>
<td>-4</td>
<td>-4</td>
<td>-3</td>
</tr>
</tbody>
</table>

### 5.1.2 VHDL Implementation

The system was designed and simulated in VHDL, a programming language used to describe logic circuits and is commonly used in CMOS integrated circuit design. The tool allows for translation of written operations into transistors through synthesis. For this initial design, inputs and outputs were chosen to be 8 bits since qPCR cycles rarely exceeds 40. We designed the system to have asynchronous resets and assumes the data is clocked in at the end of each qPCR cycle. Below shows the pseudo code of each block from Fig. 5.1.

The derivative \((dx)\) block calculates the first order derivative by subtracting the current new input value \( x[n] \) with the previous stored input \( x[n-1] \). As we are storing the current and previous inputs, as well as cascading 2 blocks for double differentiation, there will be a 4 cycle delay between input and dx2.
Table 5.2: pseudo code for “dx” block

<table>
<thead>
<tr>
<th>Instructions for “dx”</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF reset is 1 THEN</td>
<td>reset all registers and output to zero</td>
</tr>
<tr>
<td>ELSE if reset is 0 THEN on the rising edge of clock</td>
<td>moves value from register 1 to register 0</td>
</tr>
<tr>
<td>move new input value to register 1</td>
<td>compares new and previous inputs</td>
</tr>
<tr>
<td>IF register 1 is larger than register 0 THEN</td>
<td>computes dx</td>
</tr>
<tr>
<td>output is the difference of reg1 and reg0</td>
<td>set dx to 0 if the signal is not increasing</td>
</tr>
<tr>
<td>ELSE</td>
<td></td>
</tr>
<tr>
<td>output is zero</td>
<td></td>
</tr>
<tr>
<td>END</td>
<td></td>
</tr>
<tr>
<td>END</td>
<td></td>
</tr>
</tbody>
</table>

The “max” block is used to evaluate the peak of the second derivative by constantly assessing whether the input (dx2) is increasing. If the newest dx2 value is larger than the previously stored maximum, the block outputs a “store” signal accordingly. Again, we are storing and comparing the inputs in two registers, so there are two clock delays.

Table 5.3: pseudo code for “max” block

<table>
<thead>
<tr>
<th>Instructions for “max”</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF reset is 1 THEN</td>
<td>reset all registers and output to zero</td>
</tr>
<tr>
<td>ELSE if reset is 0 true THEN on the rising edge of clock</td>
<td>sets “store” to 0</td>
</tr>
<tr>
<td>move value from register 1 to register 0</td>
<td>“store” is used to signal Counter block to output cycle number</td>
</tr>
<tr>
<td>move new input value to register 1</td>
<td>compares new input to previous and maximum value so far</td>
</tr>
<tr>
<td>IF register 1 is larger than register 0 and max THEN</td>
<td>store is true if the new input is the new maximum</td>
</tr>
<tr>
<td>new input becomes max</td>
<td></td>
</tr>
<tr>
<td>“store” becomes 1</td>
<td></td>
</tr>
<tr>
<td>END</td>
<td></td>
</tr>
<tr>
<td>END</td>
<td></td>
</tr>
</tbody>
</table>

The “counter” block accumulates the clock which effectively counts the number of qPCR cycles. It computes the current cycle number subtracted by 7 to account for the previous and current computational delays. \( C_T \) is only outputted if two conditions are true: when there is a new dx2 maximum and if the current second derivative is 0. The condition dx2=0 is to ensure only the latest maximum, therefore peak of second derivative is outputted as \( C_T \).
Table 5.4: pseudo code for “counter” block

<table>
<thead>
<tr>
<th>Instructions for “counter”</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>IF reset is 1 THEN</td>
<td>reset all registers and output to zero</td>
</tr>
<tr>
<td>ELSE if reset is 0 THEN</td>
<td>accumulates counter according to clock</td>
</tr>
<tr>
<td>add 1 to “counter”</td>
<td></td>
</tr>
<tr>
<td>IF “store” is 1 AND “dx2” is 0 THEN</td>
<td>dx2=0 implies the end of exponential phase</td>
</tr>
<tr>
<td>IF “counter” is bigger than 7</td>
<td>minus 7 to account for register delays</td>
</tr>
<tr>
<td>output counter-7 as cycle number</td>
<td></td>
</tr>
<tr>
<td>ELSE IF “counter” is less than or equal to 7</td>
<td></td>
</tr>
<tr>
<td>output 0 as cycle number</td>
<td></td>
</tr>
<tr>
<td>END</td>
<td></td>
</tr>
<tr>
<td>END</td>
<td></td>
</tr>
<tr>
<td>END</td>
<td></td>
</tr>
</tbody>
</table>

The VHDL simulations in Modelsim of the whole system gave us exactly the same threshold cycles as the ones generated in Matlab. We then synthesised the blocks individually and imported them into Cadence for further transistor level simulations and fabrication.

5.1.3 Fabricated system

The system was fabricated in 0.35μm AMS technology. For ease of reading and writing digital data, a Serial Peripheral Interface bus (SPI) was built onto the chip using the same circuit from [2]. A micro-controller (KL26Z) acting as the ‘master’ was used to communicate with the SPI (‘slave’). It automatically sets up chip select (CS) to be active low and serial data sent from Matlab has to change prior to rising edge of the SPI clock (3.7MHz). The output from each processing block is connected to a multiplexer for debugging where 16 bits of data can be read by the SPI at one time. For testing, we planned to use simulated and experimental data provided by Matlab. An
ISFET sensor with ADC were included as an option for chemical input to the system. The two different input sources were multiplexed and selected via SPI.

Fig. 5.4 shows the fabricated system on chip with SPI, ISFET and ADC. Using VHDL we were able to specify the shape and dimensions in which the system should be synthesised, in order to fit the layout of the remaining circuits.

![Micro-photograph of the fabricated system](image)

Figure 5.4: Micro-photograph of the fabricated system

An updated version of the PCB test board from Section 4.2.3 was made. In this generation, plug in connectors for the micro-controller and routed test points were included for ease of debugging. Sockets for ICs and discrete components were also put in for off-chip processing (discussed further in Section 5.1.5).

![Test board version 2](image)

Figure 5.5: Test board version 2
Matlab GUI

We created a Matlab Graphical User Interface (GUI) to select the desired inputs and decode the output into meaningful formats for experiments (Fig. 5.6). The GUI lets us write individual values to the chip and read the selected output accordingly. We also included a function that can write the whole data set for an ideal qPCR curve specifying its efficiency.

![Matlab GUI for testing](image)

Figure 5.6: Matlab GUI for testing

The GUI then plots and displays the output from the chip, Fig. 5.7 shows an example of the graph and table of results for a 100% efficient ideal curve. Note that the peak of the second derivative curve plotted, highlighted with a red box in the third data column, does not appear in the exponential phase of the input qPCR curve due
to computational delay. However this delay has been taken into account in the VHDL design before the final cycle number is outputted, which is shown numerically in the green box in the second column.

![Figure 5.7: Graph and table generated by GUI when testing: blue trace is input curve, red trace shows second derivative computed by the chip](image)

### 5.1.4 Measured Results

Various data sets were applied to the fabricated system to test its performance.

**Ideal qPCR curves**

First we tested the fabricated chip using ideal qPCR curves generated in Section 5.1.1 at different efficiencies. The effect of efficiency is similar to having differences in DNA concentration: the less efficient a reaction is, the slower the amplification and therefore $C_T$ increases. Using eq. 2.3 from Section 2.4.2, we can calculate the slopes of the theoretical dilution curves for various efficiencies, as shown in Table 5.5. Accordingly at 10% efficiency, it would take 8 times longer to reach the same amplification as a fully efficient qPCR reaction.

<table>
<thead>
<tr>
<th>Efficiency/%</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
</table>

The inputs and results from the fabricated system have been plotted in Fig. 5.8. Note that the graph of the second derivative shows only positive values as the system was designed for unsigned binary. From the bottom plot, we can see the threshold
cycles computed on-chip are practically identical to the theoretical curve. And the relationship between efficiency and $C_T$s are inversely exponential, where the curve is fairly flat at higher efficiencies. This suggests that as long as we have a highly efficient qPCR reaction, we can apply relative quantification without efficiency correction and maintain good accuracies. In practice, the efficiency of the overall qPCR reaction is mainly dependent on primer design, other factors such as the master mix or the machine can also play a part. Typically we can expect most amplifications to be at least 80%, if not 90% to 100% efficient [3].

Figure 5.8: Graphs showing generated ideal qPCR curves of various efficiencies (top), corresponding second derivative (middle) and dilution curves (bottom)
Fluorescent-based qPCR data

Having tested the system under ideal conditions, the chip was then evaluated with real laboratory data. Results from two different fluorescent qPCR runs were obtained. Initial investigations found the length of the qPCR reactions played a role in the accuracy of the Derivative method. For one of the amplification runs, 99 cycles were carried out. In comparison, the number of qPCR cycles are normally limited to 30-40 only. This is because reagents get used up, dNTPs will start to degrade, but specifically for real time PCR, the length of run time coincides with an increase of by-products (i.e. primer dimers, where primers hybridise to each other). Generation of by-products can be rapid and causes peaks in the second derivative. This is illustrated in Fig. 5.9, where there is a coherent second derivative peak during the exponential phase prior to 40 cycles, after which spontaneous amplification of different unspecific products produce erratic maxima. From this, it can be demonstrated that examining the fluorescent signal alone is not sufficient to determine the exponential phase, whereas the Derivative method can provide a clear indicator of exponential amplification, provided that the reaction is limited to 40 cycles.

The first set of data, P1, contains a large number of reactions for three different primers (FAM, JOE and ROX). Each primer set had 96 wells containing three different template concentrations: high, low and no template/control (ntc). Fig. 5.10 shows there are amplifications occurring in the high and low concentrations using FAM and JOE primers, and none in primer ROX as it acts as a control. No amplification was observed in wells without DNA copies for all three primers.
Figure 5.10: Graphs showing fluorescent qPCR data (P1) with primers FAM, JOE and ROX of three different template concentrations, each with 32 wells.

The second set of data (P2) contains less reactions but more primer combinations and concentrations. qPCR was carried out on four different bacterial samples: Ecoli, Ecoli-NDM non-producer; Klebsiella pneumoniae (KP) and KP non-producer. Two specific primers NDM and KPC (Klebsiella pneumoniae Carbapenemase) were used to test Ecoli and KP respectively for mutations that produce its corresponding antibiotic resistant enzymes. Additionally, primer 16S was used to confirm if the samples are in fact a bacteria, producing six different primer combinations in total. Each combination had six different starting quantities with two wells for each concentration, and qPCR ran for 40 cycles. With 10 fold difference between each concentration, this data set is useful for testing the accuracy of relative quantifications computed using $C_T$S produced...
by the fabricated system. From the raw input data plotted in Fig. 5.11, there are clearly staggered exponential phases for each starting quantity. In addition, there appears to be slow amplifications for primer combinations ecoli-NDM and KP-KPC, in which some lower concentrations did not amplify at all.

![Graphs showing fluorescent qPCR data (P2) for six different primer combinations, with 2 wells for every 10-fold dilution](image)

Figure 5.11: Graphs showing fluorescent qPCR data (P2) for six different primer combinations, with 2 wells for every 10-fold dilution
Background Calibration

One of the aspects we had not considered prior to using real experimental data was the large variation in starting signal values. This is particularly apparent in $P_1$ where deviations from zero in the initial fluorescence can be clearly seen in Fig. 5.10. Consequently, this is problematic in our system as we initialise all registers and outputs to zero, therefore if the background is not close to zero it may appear as a maximum in the second derivative.

In light of this, we must deduct this initial offset in order to obtain results from the fabricated chip. Seeing as the starting value differs from sample to sample, we need to calibrate to zero at the start of each qPCR reaction. With considerations to future work, there are two ways to subtract this background offset: prior to the system using analogue circuits or built into the processing blocks digitally. In qPCR, amplifications typically occur after 15-20 cycles, so the first 15 cycles are usually considered to be all background noise. To optimise the calibration, we decided to test the chip using a variety of background values as offsets: cycle 1, cycle 3 and an average of cycles 3 to 15. In order to simulate the effect of prior or post system calibration, we subtracted the background value before or after scaled rounding of the data to fit the input 8 bit range respectively.

<table>
<thead>
<tr>
<th>Samples</th>
<th>theoretical $\Delta C_T$</th>
<th>Prior</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>cycle1</td>
<td>cycle3</td>
</tr>
<tr>
<td>FAM</td>
<td>2.58 (8 fold)</td>
<td>5.47</td>
<td>2.50</td>
</tr>
<tr>
<td>JOE</td>
<td></td>
<td>2.75</td>
<td>2.72</td>
</tr>
<tr>
<td>ecoli-16S</td>
<td>2.75</td>
<td>3.17</td>
<td>2.89</td>
</tr>
<tr>
<td>ecoli-non-16S</td>
<td>2.72</td>
<td>3.22</td>
<td>3.33</td>
</tr>
<tr>
<td>KPC-16S</td>
<td>4.22</td>
<td>4.22</td>
<td>4.33</td>
</tr>
<tr>
<td>KPCnon-16S</td>
<td>3.33</td>
<td>3.33</td>
<td>3.33</td>
</tr>
</tbody>
</table>

Table 5.6: Average $C_T$ for different calibration methods

Results from the fabrication chip using various calibration methods are displayed in Table 5.6. We removed the control primer ROX, along with ecoli-NDM and KPC-KPC from the comparison table as they either did not amplify or had slow amplification rates that were not representative of normal qPCR reactions. The high and low concentrations in $P_1$ approximates to an 8-fold dilution which corresponds to a theoretical $\Delta C_T$ of 2.58 cycles. Similarly, the 10 fold dilutions in $P_2$ are defined as 3.32 cycle change. Results that deviate more than 10% from the theoretical $\Delta C_T$ are marked red. Clearly, calibration prior to system generated higher number of $\Delta C_T$ close to the theoretical value. Common practice considers any signal before cycle 3 as noise, and the table
shows that an averaging of cycle 3 to 15 did not produce significantly more accurate results than using cycle 3 as the calibration offset. In favour of less computations and therefore simpler circuitries in practice, we chose cycle 3 as the value to calibrate prior to system.

Comparisons of measured results

Having optimised the calibration method, we assessed the system performance by comparing the fabricated chip with theoretical derivative computed in Matlab and industry standard methods. As mentioned in Section 2.4.2, many qPCR machines use the baseline method where threshold fluorescence is set to be 10 standard deviations above the mean of cycles 3 to 15 [4]. We computed this value and corresponding $C_T$ for each individual well using the raw qPCR data, and the theoretical $C_T$ using the Derivative Method. The qPCR data going into the chip was first calibrated using cycle 3 from each well, and then scaled to 8 bits with the maximum value of the whole qPCR run set to be 255. Following sections discuss results presented in Figures. 5.12 and 5.13 for data sets $P1$ and $P2$ respectively.

qPCR data $P1$

![Graphs showing the comparison of average $C_T$ for first set of PCR data ($P1$)](image)

Figure 5.12: Comparison of average $C_T$ for first set of PCR data ($P1$)
At first glance, results from the first qPCR data set across all three methods seem comparable. Using the baseline, particularly for primer JOE produced bigger standard deviations in $C_T$s than the Derivative method theoretically and on-chip. Neither the baseline or the fabricated system produced $C_T$s for the control primer ROX, as the curve never crosses the default threshold and the input signal after calibration remained close to zero respectively. The theoretical $C_T$ using raw data still produced some results as there always exists a maximum in the second order derivative. The large standard deviations demonstrate that these threshold cycles are undoubtedly based on noise rather than amplification.

<table>
<thead>
<tr>
<th>Samples</th>
<th>8-fold $\Delta C_T$</th>
<th>Baseline</th>
<th>Chip</th>
<th>Theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAM</td>
<td>2.58</td>
<td>3.25</td>
<td>2.50</td>
<td>2.38</td>
</tr>
<tr>
<td>JOE</td>
<td>2.31</td>
<td></td>
<td>2.72</td>
<td>2.78</td>
</tr>
</tbody>
</table>

Table 5.7: $C_T$ difference between high and low concentration for $P1$

For relative quantification, differences in threshold cycles between starting quantities are more relevant than absolute values. Table 5.7 displays the average difference in $C_T$s for high and low concentrations from qPCR data set $P1$. The results computed by the chip and Matlab are shown to be similar, implying there is little rounding errors in the fabricated back-end. Again, as we highlighted the values that deviate more than 10% to the 8-fold $\Delta C_T$, it is clear that the threshold set by baseline vary further from the standard curve. Moreover, values of $C_T$ using the Derivative method are similar between the two primers, demonstrating consistency between different qPCR runs which is essential for relative quantification.

**PCR data P2**

Results from qPCR data set $P2$ comparing $C_T$s for six primer combinations are plotted in Fig. 5.13. Starting quantities vary in 10 folds from 1000 copies/µl to 0.01 copies/µl, corresponding to a relative $C_T$ of -3.3 cycles between concentrations. On the whole, with the exception of ecoliNDM and KPCKPC, all methods produced almost identical threshold cycles. The baseline method tends to give much bigger $\Delta C_T$ than 3.3 in higher concentrations. Evidently, the relationship between $C_T$ and starting quantities becomes less linear as the concentrations become smaller. $\Delta C_T$ will typically fall below 3.3 for concentrations less than 1 copy/µl. As for ecoliNDM and KPCKPC, the original qPCR signal from Fig. 5.10 shows only the 2 to 3 largest concentrations amplified towards the end of the reaction. This explains why the chip produced lower or no threshold
cycles for these reactions as the computational delay meant the output would only be computed after 40 cycles. This is supported by theoretical $C_T$s calculated in Matlab to be similar to the baseline method. The problem could be potentially solved by manually clocking in the last qPCR data repeatedly at the end of a reaction to ensure the final cycle number outputted is in fact the threshold cycle.

Figure 5.13: Comparison of average $C_T$ for second set of data P2
Cₜ output condition

It should be noted that threshold cycles used in Fig. 5.12 and Fig. 5.13 were derived from the “store” signal generated from processing block “max”. When we were designing the system, test simulations were based on ideal qPCR curves, therefore assuming a strictly exponential-linear-plateau shape. We set the cycle number to be outputted from “counter” when the second derivative is zero, implying the reaction has entered linear or plateau phase. The intention was to ensure only the peak of the second derivative was outputted from the system, and avoid intermediate maximums to be considered as Cₜs. In this way, results can also be obtained as soon as a value is generated before the end of the reaction. However as we saw earlier in Fig. 5.9, the rate of change does not necessarily fall immediately after the exponential period. In fact, qPCR curves are likely to increase continuously with non-specific amplifications occurring. As the condition dx²=0 is not always satisfied, cycle numbers may not be outputted from the final “counter” block at all. In practice, we found the cycle number was generated 77% of the time in data set P1 and 99% in P2. As a simple solution, we can eliminate this condition and take the final value at the end of the qPCR reaction to be the threshold cycle.

pH-LAMP data

With the intent of developing the system for Point-of-Care applications, amplification method such as pH-LAMP would be more appropriate as it does not require thermal cycling. The proposed system for evaluating Cₜ can be applied in the same way to LAMP signals. However, there is no established techniques for relative quantification in LAMP since there are no cycles or definitive relationship between the reaction time and DNA copies in the sample. Therefore in this section, we investigate how well the fabricated chip processes pH-LAMP data as well as how their Cₜs change with concentrations.

Two sets of pH-LAMP data were obtained, L1 and L2. The fluorescent signals have been plotted in Fig. 5.14 and Fig. 5.15. The first pH-LAMP run was based on the bacteria Klebsiella pneumoniae (KP), testing whether it is in fact the Klebsiella bacteria and if it has the mutation that produces KPC which is antibiotic resistant. Likewise in the second run, Ecoli is tested for Ecoli bacteria confirmation and mutation of NDM (New Delhi Metallo). So for both data sets there are four different primer combinations, each combination has six concentrations of 10 fold dilutions and four wells for each starting quantity. We expect 3 out of 4 primer combinations to have positive amplifications for both data sets.
Background Calibration adjustments

In LAMP amplifications, the readings are taken at regular intervals (i.e. 1 minute), and are often still named cycles for consistency despite the lack of thermal cycling. Although there are no theoretical values as to how much $\Delta C_T$ should be according to dilutions, we can expect the threshold cycles to increase as template concentrations decrease. Considering the primers are designed to be more specific than qPCR, amplifications tend to happen much faster in LAMP reactions, where some can occur as early as cycle 3. With this in mind we move our calibration cycle to the start (cycle 1) of the reaction to prevent subtracting any amplification signals.
Comparisons of measured results

Similar to the qPCR data sets, we assessed results from the fabricated chip by comparing them to the machine’s threshold cycles and Derivative method calculated in Matlab. Without a set of theoretical $\Delta C_T$ for different concentrations, linearity of the resultant dilution curve slopes can act as an indicator for threshold cycle accuracies and validity.

pH-LAMP data L1

Results from the first pH-LAMP data set L1 are shown in Fig. 5.16. At high starting quantities, the fabricated chip gave almost identical results as the machine and Matlab for the first three primer combinations where there are amplifications. In the first two
plots, it would appear that the system was able to give threshold cycles at concentrations below the machine’s capabilities. However, looking at the fluorescent signals of those starting quantities shows no amplifications occurred. Seeing as the theoretical threshold cycle differs from the chip, it is likely that $C_T$s computed by the proposed system simply happened to coincide with the linearity of the dilution curves. We saw briefly from the earlier qPCR data sets that the Derivative method has a tendency to produce threshold cycles despite no amplifications. Similarly in this case, large standard deviations shown in the last plot suggest the $C_T$s stem from background signal and not amplification. In the future, we should consider combining the system with one that confirms amplification has occurred (i.e. comparing start and finish value) so we avoid using the $C_T$s that are a result of noise.

Figure 5.16: Comparisons of average $C_T$ from pH-LAMP data L1
The second set of pH-LAMP sees similar results to the first, with measured $C_T$s from the chip following closely to the machine’s generated values. With the exception of 1copy/$\mu l$ concentration, where the chip and theoretical derivative have lower $C_T$s for ecoli-e and ecolinon-e. This is because for both primers only 1 out of 4 wells was amplified, so the machine only took into account that particular well to be its average $C_T$. For the fabricated system and Matlab result, we computed an average of the threshold cycles across all 4 wells therefore the final mean was skewed by ones that did not amplify. Again this indicates the need to combine amplification confirmation with the Derivative method.

Figure 5.17: Comparisons of average $C_T$ from pH-LAMP data $L2$
CT correlations in pH-LAMP

Finally we examine if there are any correlations between $\Delta C_T$ and copy number in pH-LAMP reactions from the two data sets. Like the observations made from the qPCR runs, linearity in the dilution curves degrades as the amount of starting quantity decreases. In the case of pH-LAMP, it seems there is a consistent relationship down to 100 copies/µl across all primers, in comparison to 1 copy/µl for qPCR. Table 5.8 contains the average 10-fold $\Delta C_T$s for each primer combination, calculated from $10^5$ copies/µl to $10^2$ copies/µl. Similar values were observed within each pH-LAMP run, but a bigger difference in $\Delta C_T$ between the two can be seen with an overall average of 1 and 2.2 for L1 and L2 respectively. It is difficult to clarify from these data whether the numerical variation is due to different target DNA or the reactions themselves. Although in L2, there seems to also be significant differences between the primer combinations, which suggests primers are affecting the efficiency of the reactions. More experiments should be carried out to further develop the relationship between concentrations and $\Delta C_T$, though we speculate that each primer may have its own dilution curve slope value. If this is the case, we can characterise the primers prior to use for relative quantification in PoC. In addition, we should explore the optimum time period to define each cycle in pH-LAMP. A faster sampling rate could potentially provide a more accurate estimation of $\Delta C_T$s.

<table>
<thead>
<tr>
<th>Samples</th>
<th>Baseline</th>
<th>Chip</th>
<th>Theoretical</th>
</tr>
</thead>
<tbody>
<tr>
<td>KP-KP</td>
<td>0.86</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>KP-KPC</td>
<td>1.2</td>
<td>1.25</td>
<td>1.25</td>
</tr>
<tr>
<td>Kpnon-KP</td>
<td>0.78</td>
<td>0.92</td>
<td>0.92</td>
</tr>
<tr>
<td>Ecoli</td>
<td>2.15</td>
<td>2.08</td>
<td>2.08</td>
</tr>
<tr>
<td>Ecolinon-e</td>
<td>1.82</td>
<td>1.75</td>
<td>1.75</td>
</tr>
<tr>
<td>EcoliNDM</td>
<td>2.6</td>
<td>2.75</td>
<td>2.75</td>
</tr>
</tbody>
</table>

Table 5.8: Average $\Delta C_T$ in pH-LAMP for 10 fold dilutions
5.1.5 ISFET as front-end

Finally, an end-to-end proof of concept of the proposed system was tested. The purpose of the $C_T$ evaluation is to provide back-end processing to ISFET-based real time PCR or LAMP amplifications. So far, we have been testing the fabricated chip with experimental data inputted digitally through Matlab. Here, an ISFET was used to provide pH signals that imitates the amplification curve. The dimensions of the device used is $(W/L)_{elec} = 1 \times 1 \mu m$, with a sensing area of $(W/L)_{chem} = 100 \times 100 \mu m$ in order to minimise trapped charge and passivation attenuation (see Sections 3.3.2 and 3.3.1).

Readout circuit

![Block diagram of ISFET read out circuit with system](image)

Figure 5.18: Block diagram of ISFET read out circuit with system

Firstly, a readout circuit of the ISFET sensor with digital conversion was tested. The structure presented in Fig. 5.18 is similar to one used in the ISFET integrator circuit from the previous Chapter 4, Section 4.2.5. Changes in pH was measured via the ISFET’s source voltage ($V_s$), by setting the drain and reference electrode at a constant voltage and biasing the device at a set current. For the initial tests, we emulated changes in pH by varying the electrode voltage ($V_{ref}$) and using deionized water on the ISFET. Matlab input signals were converted to $V_{ref}$ using micro-controller’s DAC. Fig. 5.19 shows the output source voltage read by an on-chip ADC when applying the full 8-bit range at the gate of the ISFET. This corresponded to a $V_{ref}$ range of 700$\mu$V to 3.3V and $V_s$ of 0.25V to 2V, equating to a passivation attenuation of 0.53. More importantly, the graph shows a linear proportional relationship between the range of gate and source voltage.
Ideal qPCR curve

We used the generated ideal qPCR curves of various efficiencies as input to the system and compared the results from Section 5.1.1 in Fig. 5.20. The first thing to note is there are some outliers in cycle number from the curve. Due to conversion noise from the DAC and ADC, the condition \( dx2=0 \) was often not satisfied and therefore the final \( C_T \) was not outputted. However, \( C_T \)s derived from the signal “store” are very similar to the results using external digital input, demonstrating very little difference between using an ISFET front-end.
pH variations

Next we tested the system by varying pH on the ISFET. According to one of the figures in [5], infusing different pH solutions in a flow cell can achieve a transient response with a shape similar to a qPCR curve. In the previous measurements, we were able to utilise the full range of the DAC to vary the reference electrode voltage, which resulted in a source voltage swing of at least half of the power supply. By contrast, a pH change approximates to 30mV in gate voltage based on our average measured pH sensitivity\(^1\) for ISFETs fabricated in unmodified CMOS. Furthermore, characterisation from the readout circuit revealed attenuation of the device to be about half. With such a small input signal, any voltage offsets from ground at the start are likely to dominate as the maximum in the second derivative. To prevent this, the calibration method mentioned in Section 5.1.4 was implemented off-chip on the test board with a sample and hold circuit followed by a differential amplifier that subtracts the initial offset from the current signal (Fig. 5.21). Ratio of resistors was chosen to give a gain of 37, which roughly translates to a change of 0.4V/pH after attenuation.

\(^{1}\text{For these experiments, we had an even smaller sensitivity due to the addition of a polyimide layer from this particular wafer run. The resultant sensitivity of the chip was }\sim20\text{mV after applying RIE (Reactive Ion etching) in an attempt to remove the unwanted polyimide}}\)
the initial signal for subtraction. The experimental set up is shown in Fig. 5.22.

Table 5.9: Average pH change in pH-LAMP experiments

<table>
<thead>
<tr>
<th>pH-LAMP</th>
<th>Amplification</th>
<th>No amplification</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>1.75</td>
<td>0.53</td>
</tr>
<tr>
<td>L2</td>
<td>1.68</td>
<td>0.63</td>
</tr>
</tbody>
</table>

![Experimental set up for pH infusion](image)

Figure 5.22: Experimental set up for pH infusion

The output of the ADC and second derivative computed from the chip are plotted in Fig. 5.23. The blue trace shows the measured data sampled at 0.1s intervals in Matlab. The shape of the input data into the system looks similar to the shape of a
qPCR curve, however the peak of the derivative occurs in the plateau part of the signal. It seems the sampling rate contributed to a lot of noise artifacts, and the maximum of the second derivative is therefore caused by noise rather than the decrease in pH. If we increased the sampling period to 1s as shown by the red trace, the second derivative peak appears in a location similar to the exponential phase of a qPCR curve. However, considering the signal follows the shape of a qPCR curve only aesthetically and does not consist of a real exponential-linear-plateau format, it is difficult to assess where the second derivative maximum should lie. Having ran the experiment numerous times, we found the value is largely dependent on the sampling rate, with the optimal frequency range to be around 0.8Hz to 3Hz. Although this issue of sampling frequency is specific for this set up and not for qPCR where there are default periods set by thermal cycling, it gives a small insight into the idea of optimising measurement intervals for LAMP reactions.

Figure 5.23: Graph of source voltage of ISFET from change of pH and corresponding second derivative
5.1.6 Conclusion

In this section, the focus was on developing processing circuits to carry out on-chip quantification. Real time PCR is commonly used to analyse expression levels like miRNA regulation through evaluations of threshold cycle. With Point-of-care applications and ease of implementation in mind, we proposed a system that computes $C_T$ using the Derivative method. As the technique is based on choosing the maximum of the second order derivative to be the threshold cycle, a processing block was designed according to each mathematical operation in discrete time. The fabricated system was simulated and synthesised using VHDL.

Initial measured results using generated ideal qPCR curves showed practically identical results with theoretical values. Several problems were encountered when using real fluorescent based data, primarily differences in background fluorescence. This required calibration to remove the offset in order to prevent the increase from zero to background from becoming the maximum of the second derivative curve. Six offset calibration options were explored and compared, with background subtraction prior to input performing slightly better than within the system. This is not surprising as some accuracies are lost during scaling and rounding from digital conversions. A non-essential condition ($dx^2=0$) based on ideal data was included in one of the processing blocks and prevented $C_T$ from being outputted on occasions, so should be removed in future generations of the system. A scenario that had not been considered previously was late amplifications. $C_T$ was not calculated if the difference between the final cycle number and threshold cycle is more than the computational delay. A design feature such as adding additional clock cycles at the end of a reaction should be considered in the future.

The main findings from using qPCR and pH-LAMP data sets showed the fabricated system gave results comparable to industry standard methods. At times, it was even able to give a more linear relationship that was closer to the standard dilution curve. Typically, the chip gave a much smaller standard deviation than ones generated from the machine or baseline method, with the exceptions of no amplifications where the resultant $C_T$s were due to noise. This suggests the need to validate amplifications so that only threshold cycles from positive reactions are used. Constant temperature conditions in pH-LAMP is appealing particularly for Point-of-Care uses, on the other hand the lack of thermal cycling also means there are no standard dilution curves. From our results based on two sets of pH-LAMP data, it would seem that the rate of amplification can be correlated to dilutions, but the value is specific to each run. It is not clear if the variation stems from differences in sample and primers, or efficiency
of separate experiments. LAMP amplifications would require more characterisation before it may be applied to relative quantification.

Lastly, a proof of concept of an end-to-end system was tested using an ISFET front-end. Calibration was necessary to avoid false peaks in the second derivative that originate from DC offsets. A sample and hold circuit combined with a differential amplifier was implemented accordingly to subtract the initial background. Based on the data from pH-LAMP experiments, we used a flow cell and infusion of different pH buffers to imitate the shape of a qPCR curve. In this set up, sampling frequency played a role in the location of the second derivative peak and it was inconclusive where the maximum should occur as it was not an authentic qPCR reaction. Nonetheless, the chip was able to carry out real time evaluations of \( C_T \) from a chemical pH input with the appropriate scaling and calibration.

In conclusion, the proposed system based on the Derivative method was able to carry out real time evaluation of threshold cycles similar to the standard dilution curve. Although there were some design considerations that were overlooked, such as the need for calibration and slow or no amplifications, overall findings demonstrated the method is sound and corrections in hardware can be made in future systems. Brief investigations tried to establish correlation between threshold cycle and DNA starting quantities in pH-LAMP, though it was inconclusive due to sample size. Nonetheless, an end-to-end system with an ISFET demonstrated the proposed system’s potential to integrate with sensors and applied to ISFET-based amplifications.
5.2 Early amplification detection

As a continuation to develop on-chip real time processing for DNA amplification, this section explores the possibility of reducing qPCR reaction time by identifying the exponential phase as early as possible. When using amplification as a means for detection where the end-products are not used for further analyses, it is not essential for the reaction to complete if we can identify amplification has occurred. Currently there are no existing protocols that are tailored to shorten the time of amplification for such purpose.

In the previous system, although the second derivative is continuously assessed in real time, the threshold cycle is only determined at the end of the amplification reaction to ensure a maximum has occurred. In addition, a value for $C_T$ is always computed when using the Derivative method even when there are no amplifications, since there exist second derivatives even for noise. Hence there is motivation to develop a new algorithm that can provide confirmation of positive amplification prior to the end of the reaction. Such algorithm can then be used in conjunction with the Derivative method, or as replacement of the previous system if the new proposed algorithm can also provide sufficiently accurate estimates for threshold cycles.

5.2.1 3-point exponential evaluation system

The proposed system is built on the basis of exponential verification. Using 3 points to check if the number of templates are doubling, we can confirm there is positive amplification as soon as it happens. In theory, the amount of DNA in the sample should double every cycle during the exponential phase:

\[ f(n) = 2\delta(n) \quad (5.2) \]

where $n$ is the cycle number and $\delta(n)$ is the difference between the current cycle and previous cycle. Based on this, a processing block was designed to check if the above eq. 5.2 holds true for three consecutive data points. We chose to use three points to minimise the amount of memory and register delays, whilst still able to establish if the curve is monotonically increasing.

Fig. 5.24 illustrates the logic flow of the VHDL code used to synthesis the system. Assuming a clock is provided at the end of each thermal cycling, three data points are continuously evaluated by cascading them along registers $x1$ and $x0$ and storing the newest input in $x2$. For the first condition, we compare if the signal is monotonically increasing. Theoretically the signal should be doubling during exponential amplification.
On-chip threshold cycle evaluation and Early amplification

Figure 5.24: Diagram illustrating logic flow of proposed system

Figure 5.25: Graph showing theoretical exponential phase of qPCR curve

...however we have to account for reactions that are less than 100% efficient. Instead we check if the input has increased by at least half of its previous value consecutively. \( \delta_1 \) and \( \delta_2 \) are defined to be the differences between the first two and last two inputs respectively, as illustrated in Fig. 5.25. These are compared with the values stored in \( x_1 \) and \( x_0 \) divided by 2. As a result the output condition of the system holds true if the inputs are monotonically increasing by a factor of 0.5 or more. We chose a factor of 0.5 as the operation to divide by two is a simple right shift in binary registers. This is likely to be an underestimate for typical efficiency but can be optimised according to experimental results.

5.2.2 Simulations

Using Matlab, we tested the theory of the proposed 3 point system with the same ideal qPCR curves of various efficiencies generated in Section 6.6.1. To assess the system, we...
define the first cycle at which it outputs true to be $C_{\text{exp}}$, and therefore $C_{\text{exp}}=0$ implies no amplification. Table 5.10 shows the system correctly identifies positive reaction for 50% efficiency or above. This is expected as the rate of increase was set to be 0.5 in the proposed system, which effectively corresponds to a qPCR reaction with 50% efficiency.

<table>
<thead>
<tr>
<th>Efficiency</th>
<th>10%</th>
<th>20%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_{\text{exp}}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 5.10: Simulation results of 3-point system to qPCR curves at various efficiencies

In these ideal simulations, the exponential phase starts at cycle 1. Since simulations in Matlab have not taken into account the input and output clocking delays, we should expect $C_{\text{exp}}$ to be 3 from using three data points. Evidently, the computational delay exceeds 3 cycles as the initial inputs round down to 0 at low values from the binary divide operation. It would seem that $C_{\text{exp}}$ increases at lower efficiencies since amplification is slower.

5.2.3 Measured Results

The system was synthesised using VHDL to Cadence and fabricated in 0.35µm AMS technology. Fluorescent-based qPCR and pH-LAMP data sets from the previous Section 5.1.4 were used as input to test the fabricated system. Performance is based on accurately detecting amplification as well as potential savings on time. In addition, we also investigate the possibility of using $C_{\text{exp}}$ for relative quantification.

Accuracy

Table 5.11 displays results using the fabricated chip to identify amplifications in various primer combinations. As well as comparing percentage accuracy, we define the scenario where the system outputs true when there is no reaction as “false positive” \(^2\). This is in line with the previous Section 5.1 where we mentioned the need to develop a method to distinguish when there is no amplification. In the table, false positives are presented as a fraction, with the denominator denoting the total number of reactions that did not amplify.

Generally the system performed with high percentage of accuracy, with the exceptions of $P2$ and $L2$ data sets. First, considering reactions that have less than 100% efficiency.

\(^2\)False positives are typically defined in qPCR reactions when amplification occurs despite the absence of target templates
On-chip threshold cycle evaluation and Early amplification 223

Table 5.11: Measured results summarising accuracy of the fabricated chip

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th>False positives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>qPCR (P1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAM</td>
<td>100%</td>
<td>0/32</td>
</tr>
<tr>
<td>JOE</td>
<td>95%</td>
<td>0/32</td>
</tr>
<tr>
<td>ROX</td>
<td>100%</td>
<td>0/96</td>
</tr>
<tr>
<td><strong>qPCR (P2)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ecoliNDM</td>
<td>50%</td>
<td>0/7</td>
</tr>
<tr>
<td>ecoli16S</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>ecolinon-16S</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>KPC-KPC</td>
<td>71%</td>
<td>0/10</td>
</tr>
<tr>
<td>KPC-16S</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>KPCnon-16s</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td><strong>pH-LAMP (L1)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KP-KP</td>
<td>100%</td>
<td>0/8</td>
</tr>
<tr>
<td>KP-KPC</td>
<td>100%</td>
<td>0/8</td>
</tr>
<tr>
<td>Kpno-KP</td>
<td>100%</td>
<td>-</td>
</tr>
<tr>
<td>Kpno-KPC</td>
<td>100%</td>
<td>0/24</td>
</tr>
<tr>
<td><strong>pH-LAMP (L2)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ecoli-e</td>
<td>96%</td>
<td>1/2</td>
</tr>
<tr>
<td>ecoli-NDM</td>
<td>83%</td>
<td>3/4</td>
</tr>
<tr>
<td>ecolinon-e</td>
<td>100%</td>
<td>0/3</td>
</tr>
<tr>
<td>ecolinon-NDM</td>
<td>33%</td>
<td>16/24</td>
</tr>
</tbody>
</table>

accuracies but no false positives (P1 and P2), we discovered the error arose from slow amplifications. Specifically when the reaction increases at a rate less than half of its copy number every cycle. Visual verification from Fig. 5.11 shows primers ecoliNDM and KPC-KPC do indeed have low gradients in their amplification curves. Using Matlab to simulate alternative rates at which the data increases by, we can correctly identify all the reactions in P2 by using a factor as low as 0.25. For primer JOE, a factor of 0.4 would yield 100% accuracy. Notably, adopting a lower factor will give higher percentage accuracy in some cases, but will also likely to contribute to the number of false positives in other reactions.

Next we examine results containing false positives in the pH-LAMP data set L2. We suspect false positives may stem from the factor 0.5 being too low. However through simulations, false positives were found to exist even when factor was set to 2, implying an 100% efficient reaction. Examining closely at the scaled input data to the system,
specifically ecolinon-NDM, there is a gradual increase in the signal as illustrated in Fig. 5.26. Evidently, drift is appearing as positive amplification to the system. Particularly, the error tends to emerge at the beginning of the reactions where the system is comparing input data with registers that have been reset to zero. This effect is accentuated by the fact that the condition, $f(n) - f(n - 1) > f(n - 1)/2$ is more easily satisfied by noise when the values are low.

![Graph of scaled fluorescence data of primer ecolinon-NDM from L2](image)

**Figure 5.26: Graph of scaled fluorescence data of primer ecolinon-NDM from L2**

**False positive corrections**

We considered various ways to rectify the errors in accuracy, such as comparing more than three data points, or accumulating a number of true outputs for confirmation of an exponential phase etc. Most modifications would add delay to the final output which is undesirable since we want to identify amplification as early as possible. Instead, we decided to adjust and optimise some of the parameters in the current structure. Using the initial Matlab simulation, we reset all registers to 255 instead of 0, and added an additional condition to prevent errors at small values. For instance, when the input is 0 or 1, we store it as 2 because exponential comparisons at low numbers give higher rates of error.

Table 5.12 shows the results from the new parameters. Most of the false positives were able to be corrected except one, whilst retaining the same percentage accuracy in the other primers. As a trade off, we could increase the default amplification rate from 0.5 which would eliminate the remaining false positive but cause a drop in percentage accuracy in the other reactions. On the whole, overall accuracy of the system is 96% across all primers, and can correctly identify no amplification 99.6% of the time.
Table 5.12: Results on accuracy of the proposed system with adjusted parameters

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th>False postives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>qPCR (P1)</strong></td>
<td>96 wells</td>
<td></td>
</tr>
<tr>
<td>FAM</td>
<td>100%</td>
<td>0/32</td>
</tr>
<tr>
<td>JOE</td>
<td>93%</td>
<td>0/32</td>
</tr>
<tr>
<td>ROX</td>
<td>100%</td>
<td>0/96</td>
</tr>
</tbody>
</table>

| **qPCR (P2)**        | 14 wells|               |
| ecoliNDM             | 50%     | 0/7           |
| ecoli16S             | 100%    | N/A           |
| ecolinon-16S         | 100%    | N/A           |
| KPC-KPC              | 71%     | 0/10          |
| KPC-16S              | 100%    | N/A           |
| KPCnon-16s           | 100%    | N/A           |

| **pH-LAMP (L1)**     | 24 wells|               |
| KP-KP                | 100%    | 0/8           |
| KP-KPC               | 100%    | 0/8           |
| Kpno-KP              | 100%    | N/A           |
| Kpno-KPC             | 100%    | 0/24          |

| **pH-LAMP (L2)**     | 24 wells|               |
| ecoli-e              | 100%    | 0/2           |
| ecoli-NDM            | 92%     | 0/4           |
| ecolinon-e           | 100%    | 0/3           |
| ecolinon-NDM         | 96%     | 1/24          |

**Time savings**

To investigate the prospect of shortening detection time in amplifications, the range of \( C_{exp} \) for each primer combination are plotted in Fig. 5.27. Fluorescent based qPCR data are plotted on the left hand side of the red dotted line, pH-LAMP reactions are on the right. The line bars represent range of values in \( C_{exp} \) depending on the DNA concentration. The distance between the data and the top of the y-axis (40 cycles) represents the number of cycles that are presumably “redundant” in a detection setting. Clearly, the potential number of cycles saved in qPCR is significantly smaller than pH-LAMP as amplification tends to happen later in the reaction. Considering each thermal cycling or period of sampling averages on 1 minute per cycle, we may see savings of 5 to 15 minutes in qPCR. In comparison to pH-LAMP, we can expect to cut the length of the reaction to 20 minutes majority of the time.
C_{exp} for relative quantification

As an additional function, we want to investigate the possibility of using this method to perform relative quantification. Since \( C_{exp} \) is computed based on an exponential phase occurring, it can be considered as a threshold cycle. Fig. 5.28a and Fig. 5.28b are dilution curves based on \( C_{exp} \) for qPCR and pH-LAMP reactions respectively. Looking at the qPCR data, there is a general trend of \( C_{exp} \) increasing with less starting quantity except for primer JOE. Similar to the Derivative method, linearity is maintained up to a sample concentration of 1 copy/\( \mu l \), after which the 10 fold difference between \( C_{exp} \) becomes less than 3. Likewise, \( C_{exp} \) is inversely proportional to copy number in pH-LAMP reactions up to 100 copies/\( \mu l \). However, there appears to be slightly more variants particularly at lower concentrations.

<table>
<thead>
<tr>
<th>On-chip system</th>
<th>FAM</th>
<th>JOE</th>
<th>Ecoli16S</th>
<th>Ecolinon-16S</th>
<th>KPC-16S</th>
<th>KPCnon-16S</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-point exp.</td>
<td>2.5</td>
<td>-0.8</td>
<td>2.9</td>
<td>3.5</td>
<td>4.3</td>
<td>3.3</td>
</tr>
<tr>
<td>Derivative method</td>
<td>2.5</td>
<td>2.7</td>
<td>3.2</td>
<td>3.3</td>
<td>4.1</td>
<td>3.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>On-chip system</th>
<th>KP-KP</th>
<th>KP-KPC</th>
<th>KPno-KP</th>
<th>Ecoli-e</th>
<th>Ecoli-NDM</th>
<th>Ecolinon-e</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-point exp.</td>
<td>0.9</td>
<td>1.4</td>
<td>0.9</td>
<td>2.2</td>
<td>1.8</td>
<td>2.4</td>
</tr>
<tr>
<td>Derivative method</td>
<td>1.0</td>
<td>1.3</td>
<td>0.9</td>
<td>2.1</td>
<td>1.8</td>
<td>2.8</td>
</tr>
</tbody>
</table>

On closer inspection, we compared average \( \Delta C_T \) obtained from both fabricated system using starting quantities of 1 to \( 10^3 \) and 100 to \( 10^5 \) copies/\( \mu l \) for qPCR and pH-
LAMP reactions respectively. It would appear from Tables 5.13 and 5.14 that using three points gives almost the same threshold cycles as the Derivative method. Further investigations are required to understand the outlier in primer JOE. Nonetheless, values obtained from the proposed 3-point exponential evaluation system demonstrates the potential of carrying out relative quantification with early detection of amplification simultaneously.

Figure 5.28: Dilution curves based on $C_{exp}$
5.2.4 Conclusion

A system to indicate positive amplification before the end of a reaction was proposed. Based on DNA copies doubling every thermal cycle, the system determines exponential phase of a qPCR curve using 3 consecutive points. By comparing the differences between current data and previous inputs, the system outputs true when the signal is monotonically increasing by a certain multiple.

Similar to Section 5.1, the system was fabricated using VHDL and tests were carried out based on fluorescent-qPCR and pH-LAMP data sets. Initial results showed the fabricated system can correctly identify amplification with decent accuracies, with exceptions of false positives in the second set of pH-LAMP data. Errors resulting in false positives tend to appear mainly at the beginning of a reaction, when conditions are more easily fulfilled and registers were reset to zero. With adjusted parameters, the system was able to achieve 96% overall accuracy and 99.6% when there are no amplifications.

The potential time that could be saved using the proposed 3-point algorithm was investigated by defining $C_{\text{exp}}$ as the cycle at which the system classifies as exponential phase. The number of redundant cycles in qPCR is significantly less than that in pH-LAMP reactions since amplifications tend to occur later due to lower specificity in primers. In addition, we investigated the possibility of using $C_{\text{exp}}$ as threshold cycle for relative quantification. After initial examinations, it would seem that the fabricated system was able to provide relative estimates in starting quantity very similar to the Derivative method, though generally with higher standard deviations.

Overall, the proposed system can distinguish between positive amplifications through a 3-point exponential evaluation method. Early detection varies with concentration due to the nature of qPCR and LAMP amplifications, with most time savings at high starting quantities and Isothermal conditions. Furthermore, threshold cycles computed from the fabricated system can be correlated to template concentrations for relative quantification. With further verifications and refinements to parameters, the proposed method shows promising results for early amplification detection as well as quantification.
5.3 Summary

This chapter was centered around integrating back-end circuits for real-time processing in ISFET-based amplifications. We focused on two aspects: developing on-chip analysis for miRNA expressions, and shortening detection time in amplification reactions. Expression levels are commonly evaluated by real-time PCR, in which the threshold cycle can be used to quantify templates. In such detection-based amplifications where the final amplicons are not required for further analyses, reaction time can be reduced based on obtained results. The rationale behind integrating hardware to carry out on-chip processing is to minimise data transmission, or eliminate the need for PC all together, resulting in a truly portable platform. As well, processing in real-time can produce results prior to the end and decrease overall time required for analyses. Tools such as VHDL aids implementation of digital back-ends, by allowing synthesis of written operations into logic circuits that can be easily included in CMOS microchip’s design and layout.

Two systems were presented: one for evaluating threshold cycle based on the Derivative method; the other identifies exponential phase using 3 points for early amplification detection. The Derivative method was considered as it does not require user input and can be easily translated to discrete time operations. The first proposed system contains modules that collectively computes maximum of the second order derivative and outputs the corresponding cycle as $C_T$. The second system contains a simple algorithm developed to estimate the rate of amplification by comparing 3 points, thereby determining if there is positive amplification. Using only 3 points was intended to minimise register memory and therefore area, as well as computational delays.

The fabricated systems were tested with a customised Matlab GUI to transmit signals and plot received data from the micro-controller used to communicate with the SPI on chip. Four sets of data from fluorescent-based qPCR and pH-LAMP were used for testing. The first set of qPCR data contains a large number of reactions, with a total of 288 wells for three primers and compares two different concentrations with control. The second set of qPCR data is smaller in comparison but contains more variations in dilutions, ranging from 0.001 to 1000 copies/µl. Similarly, the pH-LAMP reactions contain six different concentrations from 1 to $10^5$ copies/µl. The need for background calibration was put into light when using experimental data. Variations in starting value caused errors as the systems were initialised to zero. An appropriate offset from the background phase was chosen to be subtracted prior to input.

Performance of the Derivative method system focused on computing accurate and consistent $\Delta C_T$s. For our application, we are interested in relative quantification which
can be carried out based on threshold cycles alone, without having to produce a separate standard curve with known dilutions. Results from the fabricated system were generally close to the theoretical values, and produced smaller standard deviations than industry’s baseline method. Limitations arise at low concentrations and slow amplifications, with linear correlations of \( C_T \) to starting quantity in qPCR restricted to above 1 copy/\( \mu l \) and 100 copies/\( \mu l \) for pH-LAMP. Considering there are no standard values for \( \Delta C_T \) in LAMP, attempts were made to relate threshold cycle with concentration. Although there was clearly a \( C_T \) value associated with each pH-LAMP data set, it was ambiguous whether the difference between the two runs was a result of efficiency variations or specific to primer combinations. Implementation of end-to-end operation using an ISFET sensor shows the system can carry out real-time threshold cycle evaluation when combined with off-chip calibration circuits and proper scaling of input.

For the 3-point exponential evaluation system, high accuracy was an essential criteria for identifying amplifications early. In general, the fabricated chip was able to confirm amplification with high accuracies, with exceptions of some false positives. The false positive could be corrected by adjusting some conditions and parameters in the proposed algorithm, with overall accuracy of 96%. Potential time saved by the system varies with concentrations as exponential phase occurs later for low starting quantities. Particularly, durations of LAMP-based reactions could be reduced by half due to faster amplifications from high primer specificity. Although the system was originally intended for early detection, the cycle at which amplification is confirmed can also be considered as the threshold cycle. Investigations into the possibility of relative quantification found \( C_T \)s produced by the proposed 3-point algorithm were nearly identical to the Derivative method. The findings illustrate potential to provide quantification prior to the end of an amplification reaction.
Bibliography


Chapter 6

Conclusion and Future Work

At a time when chronic diseases are the key contributors to mortality, epigenetics bring about new prospects for effective biomarkers and potential therapeutics. Development of technology to accommodate fast detection at low costs is essential for integrating epigenetic monitoring into future healthcare.

This thesis focuses on applying CMOS design techniques to develop circuits for various types of methylation analyses and miRNA quantification. For Methylation-specific PCR, an ISFET current mirror was proposed to provide differential sensing of opposing reactions. Combined with weak inversion operation, the configuration achieved the necessary pH resolution for detecting amplification, with reduced sensitivity to drift and temperature. More detailed analysis of methylation requires sequencing with bisulfite sample preparation. Integration was introduced as a solution for improving base calling accuracies by extending window for detection and averaging by summation. A switched current implementation was fabricated with integration carried out functionally subject to input range. Corresponding protocol was proposed to compress workflow for methylation-based sequencing. According to simulations, the technique can be applied to sequences of certain lengths and density of methylated sites. Building on the basis of an existing ISFET-based platform for real-time amplifications, on-chip back-end processing circuits were proposed for miRNA quantification. Two digital systems were implemented with VHDL and tests were carried out on experimental data sets from fluorescent-qPCR and pH-LAMP reactions. A circuit for evaluating threshold cycles based on Derivative method showed consistency in computing relative differences, and an end-to-end concept was demonstrated using an ISFET sensor. In addition, detection time in LAMP-based experiments can potentially be halved with real-time assessment of exponential rate from three data points.
6.1 Summary and Contributions

This thesis set out to apply ISFET-based detection for methylation and miRNA analyses. Contributions and concluding remarks from each chapter are summarised below.

6.1.1 The role of epigenetics in chronic disease

To formulate a well-designed circuit, its application and the information required must be fully understood. Chapter 2 provides a basic understanding of epigenetics, its role in chronic diseases and the different types of detection methods available. Focusing specifically on chronic kidney disease, it was found that current diagnostic techniques have limited sensitivity for early detection as well as ineffective treatments, therefore would benefit from new epigenetic-based biomarkers and therapeutics. A comprehensive review of frequently reported epigenetic modifications linked to CKD was compiled. Specific gene methylation and miRNA regulations were linked to the disease’s origin, severity, effects of environmental influences, and may be reversible through targeted therapies. Thus forming a picture as to how epigenetics can be applied to chronic diseases, though it was noted that studies should be conducted in the sample intended for detection to prevent discrepancies.

Followed by a review of current available analyses, it was established that methylation detection can be quantitative, qualitative or global while miRNA regulation is defined by quantification of its expression levels. Prior to analysis, both methylation and miRNA require sample preparation in order to preserve methyl-groups or obtain cDNA with adequate length for primer-based techniques respectively. While there are some methods that do not need chemical modifications, they either lack sensitivity or are at an infancy stage of development. Limitations in methylation methods are very much application dependent, with Methylation-specific PCR standing out as the fastest technique for detecting targeted methylation sites with high specificity, and sequencing as the only quantitative analysis for single base resolution. Similarly, real-time PCR emerged as the preferred choice of technique due to easy integration into existing work flow and ability to carry out both absolute and relative quantification. Ultimately, the best detection method is one that can provide essential information with the least cost and time for the intended application.

6.1.2 Ion semiconductors in DNA detection

Following on from methylation and miRNA analyses, chapter 3 looks into semiconductor based detection with focus on ISFETs. The incentive for label-free electronic
DNA analyses stems from omitting procedures for attaching chemical tags and bypassing the need for signal conversion when using electrochemical sensors. Highlights of past and current semiconductor technology summarised several principles for label-free detection: impedance spectroscopy, negative charge, RedOx reactions and release of pyrophosphate. Majority of the reported literature required post processing of electrodes, primer immobilisations or had limitations in sensitivity. pH-based method is founded on detecting hydrogen ions released during DNA hybridisation, and can be achieved with ISFETs without post-processing.

A closer examination of ISFET’s operations illustrates its pH sensitivity arise from a combination of proton transfers at the insulator surface and distribution of charge in the electrolyte. The effects can be modeled as a chemical voltage drop, $V_{chem}$ from the reference electrode to gate, and incorporated into MOSFET’s characteristics in circuit analyses. Non-ideal phenomenons, trapped charge and attenuation are a result of passivation layer coupling when ISFETs are fabricated in unmodified CMOS process. A number of techniques including UV radiation and area design can be applied to minimise DC offsets and passivation capacitance. Other aspects of ISFET measurements such as temporal drift and temperature sensitivity can be considered on an application basis.

ISFET-based technology available for epigenetic detection to date were presented. Sequencing by Ion torrent demonstrates capacity for fast high-throughput analyses, with limitations in base call numbers due to hydrogen ions diffusion. Microchips designed to compute percentage methylation in cancer showed potential in developing application-specific circuits. Lastly, an ISFET-based platform was reported to carry out real time application and emerges as the smallest portable system amongst other optical-based machines available in the market. While there exists draw backs in ISFET sensing, ability to provide sensitive DNA detection without post processing makes it stand out from other fluorescent-based and semiconductor platforms.

6.1.3 Analogue current mode design approach for methylation

Chapter 4 is centered around analogue circuit design based on different methods of methylation detection. Methylation-specific PCR and bisulfite sequencing were concluded as most suited to future monitoring in routine check-ups and obtaining methylation results at single CpG resolution when developing therapeutics. As a result, a novel configuration of an ISFET current mirror was presented for MSP, and an ISFET switched current integrator for signal processing in sequencing. The rationale behind each circuit design was focused on application: a weak inversion current mirror achieves differential sensing required for MSP with the minimal number of devices and drift can-
cellation, relevant for Point-of-Care; integration in switched current domain increases window for detection and base calling for homopolymers at low frequencies appropriate for nucleotide incorporations without large passive components.

Based on experimental results, the effects of trapped charge and passivation capacitance were particularly significant. In the case of the ISFET current mirror, DC offset was incorporated into the circuit’s operation as an alternative to compensation. Connecting the devices strategically according to trapped charge added favourably to the output current gain. While attenuation altered the shape of the exponential response, the necessary pH resolution for MSP was achievable with appropriate bias current. As well, differential configuration reduced drift over time and temperature in ranges practical for amplification. However, experiments were carried out with pH buffers only and miniaturisation of the chemical set up would require a porous partition to act as a salt bridge between micro-fluidics chambers. Whether there will be DNA contamination through the barrier is unknown.

In contrast to the current mirror, trapped charge was detrimental in the ISFET integrator’s circuit operation. The significant increase in DC offsets was likely to be a result of larger devices. Prior to chemical testing, the electrical response of the fabricated integrator was promising therefore prompted the concept to be tested with an external ISFET as input sensor instead. However, the demand for resolution in sequencing proved to be much greater than one provided by the circuit due to previously unaccounted noise from the chemical gate and readout circuits. The results demonstrated the need to consider system as a whole, particularly when stringent specifications are required, to include biasing and conversion from analogue and current mode into the design flow. Although not considered in this thesis, processing inside or outside of a pixel would further impact constraints to structure and area.

A new protocol was proposed to increase efficiency for bisulfite sequencing. From the literature review, it became apparent that only information on cytosines are essential for methylation analysis, yet every base is sequenced. Compressing the incorporation of other bases in one step have the potential of retaining single base resolution at methyl-cytosines, while the remaining bases are read as long stretches of generic homopolymers. Matlab results concluded feasibility of the technique largely depends on dynamic range and resolution required for homopolymers, which is determined by the length of sequence and density of CpG sites. Operating on the assumption that homopolymers differing in 10 bases will produce sufficient variations in response, locations of the methyl-cytosines can be deduced from occurrences by fitting possible patterns of methylation. Clearly, the assumption of a 10-base resolution needs to be
verified beyond simulation. As well, the investigation was limited to data from one gene sequence. While the presented results demonstrate potential, further analysis should include practical limits of detection and percentage of accuracies derived from a large cohort of genes.

6.1.4 On-chip threshold cycle evaluation and Early amplification

Contributions from chapter 5 are towards miRNA quantification by implementing on-chip back-ends for evaluating threshold cycles and detecting amplifications early. Although there is an existing ISFET platform for conducting real time amplifications, the addition of integrated processing blocks to compute $C_T$s can make a truly portable device for carrying out relative quantification without PC. Real-time processing also allows for possible reduction in detection time when the qPCR end-products are not required. Subsequently two systems were proposed and fabricated: $C_T$ evaluation based on Derivative method and early amplification using three data points.

Primary findings showed both systems written in VHDL performed exactly as described in the synthesised programming language. Complications mainly arose from overlooked issues such as background calibration or conditioning statements that were not satisfied in previously unaccounted scenarios. Although not always obtainable, the systems’ design process should have included practical input data from the start since many complications were a result of deviations from an ideal qPCR curve. By large, each fabricated system produced respectable outcomes for its application, namely consistent threshold cycles close to theoretical values for relative quantification and high percentage accuracy for false positives when reducing detection time. These results were based on four input data sets from fluorescent-qPCR and pH-LAMP reactions, therefore are limited to the variations of signals provided. In addition, measured results of an end-to-end system with ISFET sensor highlighted the importance of input range. Achieved specifications of the front-end will dictate required number of bits and therefore how data should be scaled for testing. Evidently, design flow of digital back-end circuits with VHDL should be focused on ensuring robustness in algorithms and optimising parameters through testing large number of data sets, since correct implementation of functionality is almost guaranteed.

6.2 Future work

This thesis provided some initial circuit ideas for various epigenetic analyses. Some recommendations to further the designs into practical applications are discussed.
6.2.1 Development of differential micro-fluidics

Practicality in MSP application of the ISFET current mirror from Section 4.1.2 is dependent on successful miniaturisation of micro-fluidics. In order to carry out further experiments with DNA samples, a setup that accommodates for small volumes while incorporating a salt bridge is essential. As mentioned in the chapter, a porous partition is sufficient as a salt bridge between two chambers. So the focus when developing a differential micro-fluidics should be on choosing appropriate materials and size of pores that prevent DNA contaminations between 2 samples whilst allowing for transfer of ions.

6.2.2 Computational readout for quantitative MSP current mirror

In Section 4.1.2, it was mentioned that quantitative analysis can be implemented by cascading output stages. The number of output branches with increased current indicate positive amplifications. A readout circuit that sums all the positively amplified currents and divide by a current scaled by multiples of the total output branches would give the percentage methylation. There is potential for both analogue and digital implementations.

6.2.3 ISFET shapes

Figure 6.1: ISFETs with same total top metal areas, and different surface areas

The most prominent non-ideal effects of ISFETs manufactured in unmodified CMOS process are trapped charge and passivation capacitance. As we found in our experiments, particularly in the ISFET integrator, DC offsets and gate voltage attenuation
can debilitate a circuit’s operation through improper biasing and decrease in signal amplitudes. Since the value of passivation capacitance is linked to its chemical sensing area and a portion of trapped charge is located in the dielectric, ISFETs of different shapes but same top metal area should result in varying device characteristics due to the additional surface area provided by irregular perimeters. In our last wafer run, we included various ISFETs with assorted shapes in sensing areas such as multiple squares, criss cross, strips etc as shown in Fig. 6.1 to investigate if there is a relationship between sensing perimeter with attenuation and offset.

6.2.4 Multiple base incorporation measurements

Several practical assumptions were made when simulating the proposed sequencing protocol in Section 4.3. Theoretically, incorporating several types of base simultaneously should generate a signal proportional to the length of the extended sequence, however this has not been confirmed in practice. So prior to further simulations, the limits and resolution of incorporating different nucleotides should be defined. For example, experiments should be carried out using a high resolution commercial pH probe\textsuperscript{1} to verify that the value of pH change from incorporating two of the same base is equal to two types of base adjacent to each other. Then ISFETs should be used to confirm if they also generate similar transient responses.

6.2.5 Validation of systems with DNA amplification

The two proposed systems from chapter 5 should be validated with DNA amplification. As it stands, the set up of the ribbon cable and test board presented in Section 5.1.5 should allow for the dipstick to be placed in a PCR machine. The system’s clock needs to be be synchronised with the thermal cycling, either manually or automated from PC. In theory, if the rate of ISFET sensor drift is slower than the exponential increase in pH from amplification, it should not interfere with computation of \( C_T \). Otherwise it must be compensated prior to input of the system. As an alternative, future generations of the system can include drift subtraction into its algorithm. Ultimately, the systems should be tested as a fully integrated platform with real time amplification carried out by ISFETs.

6.2.6 Characterising relative \( C_T \)s and sampling for pH-LAMP

Chapter 5 briefly discussed that theoretical \( \Delta C_T \)s did not exist for LAMP-based reactions due to lack of thermal cycling. From our results, it would appear that a specific

\textsuperscript{1}http://www.sentron.nl/ph-meters/ph-meter-line/
ΔCT relationship with template concentration may exist for a primer combination. In addition, sampling at a higher rate than one minute could generate higher resolutions of relative threshold cycles, or more noise from oversampling. An optimum frequency should be investigated, from which a ΔCT to starting quantity may be derived.
Appendix A

Publications

A.1 Conference papers


A.2 Journal papers


240
Appendix B

Principles of DNA hybridisation

DNA hybridisation is a mechanism that forms the basis of genetic testing. It can determine the degree of similarity between sequences on the principle of complementary bases. Complementary bases are held together by hydrogen bonds, where thymine is complementary to adenine, and cytosine bonds with guanine. When double stranded DNA are heated to 90°C, hydrogen bonds between the bases break and the strands separate. When the temperature is lowered (50 to 60°C), any complementary bases or sequences will align and anneal to the singled stranded DNA. The pairing of bases therefore reflects the order of sequence.

Figure B.1: Structures of complementary DNA strands: thymine with adenine; cytosine with guanine [1]
Fragments of sequences or oligonucleotides are often used to carry out DNA detection. Typically labeled with fluorescence or radioactive tags, hybridisation of such probes indicates complementary sequences. Alternatively, oligonucleotides can be used to initiate DNA extension by allowing dNTPs to be added to the sequence. Oligonucleotides used for copying DNA, typically known as primers, anneal in complementary pairs to top and bottom strands. The forward primers refers to the oligonucleotide that binds upstream to the bottom strand; and reverse primer anneals downstream to the top strand. Since the hydrogen bonds binding cytosines with guanine are more stable than thymines to adenines, with three instead of two, the GC contents will dictate the optimum temperature at which primers will anneal to template. A low temperature would promote non-specific hybridisations, while a high temperature would decrease the percentage of primers binding. Generally, the forward and reverse primers should have similar melting temperatures to optimise the reaction. Chain extension is then catalysed by DNA polymerase from the 3 end, referring to the positioning of carbon on the sugar backbone. DNA polymerase are enzymes integral in genetic testing as well as in vivo transcription for inserting individual nucleotides into pre-existing sequences according to the template DNA.

![Diagram showing primers annealing: downstream (top); Upstream (bottom)](image)

Depending on the application and information required, different types of DNA analyses should be implemented. Information on the whole genome or long stretches of sequences are typically obtained through sequencing. The detection of fluorescence or other byproducts from individual dNTP extension are often used to determine the order of sequence, with exceptions such as sequencing by ligation which uses probes. Identification of a specific gene or variants in the genome can be achieved through
genotyping or SNP detection. In this case, only a region of DNA is of interest, which can be selected by primers in targeted sequencing. Alternatively, probes are a rapid way of identifying short stretches of sequence, particularly for single base mutations in SNPs. Similarly, methods such as Southern blot adopts hybridisation probes to evaluate DNA expression levels by intensity of radioactivity or fluorescence.
Appendix C

ISFET Packaging

To carry out chemical testing on ISFETs, surface of the silicon die needs to be exposed. This required customised packaging of the fabricated die to PCB with liquid protection for wire bonds. The processes of wire bonding and encapsulation are described below.

C.1 Wirebonding

Firstly, the die is glued onto the ground plane of the PCB using conductive silver-filled epoxy (EPO-TEK H20E), and cured in the oven for 15 minutes at 120°C. Then wire bonding was carried out using the wire bonder model K&S 4123, based on wedge bonding. In wedge bonding, welds are made by a combination of ultrasonic energy and pressure applied to the wire. A 45 to 60 degree feed is used to ensure the wire is directly under the surface of the wedge. Aluminum or gold wires are commonly used in wedge bonding. Additional thermal energy is required when bonding with gold wire (∼120°C), and is supplied through the jig used to clamp the PCB. Control parameters described in Table. C.1 and Fig. C.2 are adjusted to ensure a clean secure welds that do not short circuit to neighbouring bond pads.

Figure C.1: Illustration of wire bonding steps\(^1\): Step 1, position wedge above pad and complete first bond; Step 2, clamp of the wire opens and move wedge in straight line to second bond pad; Step 3: complete second bond; Step 4, clamp closes to cut wire

\(^{1}\text{http://www.palomartechnologies.com/blog/bid/206846/ball-bonding-vs-wedge-bonding}\)
Figure C.2: Controls for wire bonding parameters: top row for first bond; bottom row for second bond

Table C.1: Descriptions of wire bonding parameters

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
<th>Typical values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search</td>
<td>Height of bonding wedge above bonding surface</td>
<td>0.5 - 1*</td>
</tr>
<tr>
<td>Force</td>
<td>Downward force applied by wedge to wire against bonding surface</td>
<td>2 - 3</td>
</tr>
<tr>
<td>Time</td>
<td>Duration of ultrasonic energy applied</td>
<td>1 - 3</td>
</tr>
<tr>
<td>Power</td>
<td>Magnitude of ultrasonic energy applied</td>
<td>2 - 3</td>
</tr>
<tr>
<td>Loop</td>
<td>Height of bonding wedge after the first bond (curvature)</td>
<td>1 - 2</td>
</tr>
</tbody>
</table>

*When bonding from substrate to die (reverse bonding), Search on the second bond is usually 0.3 bigger than the first bond to account for thickness of silicon (and vice versa for forward bonding)

Figure C.3: View of die through wire bonder’s microscope: double row pads can be bonded using a larger loop parameter than the inner interconnects
C.2 Encapsulation

An electrically insulating glob top epoxy (EPO-TEK T7139) was used to encapsulate bonds and other areas of exposed aluminum on the chip. The epoxy requires curing at 125°C for 1 hour. The consistency of the glob top after mixing the two-part components is fluidic\(^2\) and may harden outside of the applied area after curing. As a result it requires two to three layers of gradual applications to ensure the window for the ISFET sensor is not covered up.

![First layer of glob top](image1)

Any PCB vias or exposed tracks should also be covered.

![3rd layer](image2)

![2nd layer](image3)

Figure C.4: Close up photograph of a dipstick with glob top encapsulation

\(^2\)Consistency is also affected by exact measurements of the two-part ratio (10:1). From practice, we found too much of part B contributed to a more runny epoxy. After mixing, the glob top should be left in air for 30 minutes before being applied to the chip.
Appendix D

CAD drawings for flow cell

The flow cell was based on the design used in [3]. The version used in this thesis was made by stacking two layers of clear acrylic plastic cut by laser cutter (Fig. D.1) and a soft foam material on the bottom to form a water tight seal when clamped to the PCB. Plastic tubing was used as the inlet and outlet, and also acts as a watertight opening for the reference electrode which requires regular renewing after repeated use. A clear adhesive epoxy was also used to seal any gaps which were not watertight.

Figure D.1: CAD drawing of flow cell layers: red represent lines to be laser cut, inlet and outlet holes should be placed aligned to the end corners of the chamber to minimise air pockets
Appendix E

Miniaturisation of salt bridge

An attempt was made to downsize the differential set up used in experiments for the ISFET current mirror in Chapter 4, Section 4.1.2. To incorporate a salt bridge into the existing flow cells\(^1\) (used in Sections 4.2 and 5.1), we made a flexible\(^2\) salt bridge with smaller diameters to connect two flow cell chambers. The ends of the salt bridge was filled with cotton wool to allow ion diffusions, and a syringe dispenser tip\(^3\) was used to deliver KCl solution into the tube. The top opening for injecting KCl must be closed to prevent leakage through both ends. In addition, there should be no air bubbles in the tubing so the opening must be air tight immediately after the salt bridge is filled.

From our experiments, we found this design was not compatible with a flow cell set up. Since injecting solution through the inlet would also force some of the electrolyte to flow through into the other flow cell’s chamber via the salt bridge. This was confirmed

---

\(^1\)A small modification was applied to include an opening in the flow cell chamber for the salt bridge

\(^2\)This allows flexibility to connect two chips clamped on separate boards

\(^3\)Since the cotton wool ends would dry up between usage and requires changing after every use, the dispenser tip was attached to the design for easy refill of KCl
by using colouring dye in the pH buffers. A possible solution would be to eliminate the inlets and outlets and pipette solution into the chamber from an opening.

Figure E.2: Flexible salt bridge connecting two flow cells on dipsticks
Bibliography

