CHARGE PUMPS AND FLOATING GATE DEVICES FOR SWITCHING APPLICATIONS

by

Bongani Christopher Mabuza

Submitted in partial fulfilment of the requirements for the degree

Master of Engineering (Microelectronic Engineering)

in the

Department of Electrical, Electronic and Computer Engineering
Faculty of Engineering, Built Environment & Information Technology

University of Pretoria

July 2012

SUMMARY

CHARGE PUMPS AND FLOATING GATE DEVICES FOR SWITCHING APPLICATIONS

by

Bongani Christopher Mabuza

Supervisor: Prof. S. Sinha
Department: Department of Electrical, Electronic & Computer Engineering
University: University of Pretoria
Degree: Master of Engineering (Microelectronic Engineering)
Keywords: Field programmable gate arrays, CMOS technology, FG transistor, charge pumps, sense amplifiers, CHEI, FN tunnelling, millimetre wave integrated circuits, Dickson pumps, two-phase clock, velocity saturation, EEPROM, fuses, microelectromachanical systems and voltage doubler.

On-chip impedance tuning is used to overcome IC perturbations caused by packaging stress. Tuning is more important for matching networks of radio frequency (RF) systems. Possible package resonance and fabrication process variations may cause instability, which is a major problem in RF systems. Thus, precautions need to be taken in order to maintain the overall stability of components and the final system itself. Electrically erasable programmable read-only memory switches (EEPROMs) occupy less die area compared to e-fuses and microelectromechanical system (MEMS) switches, thus EEPROMs are proposed to be used as tuning switches in millimetre-wave (mm-wave) applications. It is anticipated that EEPROM switches will also enable multi-time programming because of the smaller area and the fact that more switches can be used for fine-tuning.

The problem addressed in this research is how suitable EEPROMs are for switching applications in the mm-wave region. The main focus of this dissertation is to characterise the suitability of EEPROM switches qualitatively for tuning with systems operating in the mm-wave spectrum.

130 nm SiGe BiCMOS IBM 8HP process technology was used for simulation and the fabricated prototypes. The Dickson charge pump (CP), two voltage doubler CPs and four floating gate (FG) devices were investigated. Literature and theoretical verification was done using computer aided design (CAD) Cadence software through circuit analysis and the layouts were also designed for integrated circuit (IC) prototype fabrication. The qualitative evaluation of the hypothesis was based on investigating reliability issues, switching characteristics, CP output drive capability and mm-wave characterisation.

The maximum measured drain current for FGs was 1.4 mA, 2.7 mA and 3 mA for devices 2, 3 and 4, respectively. The ratio between ON state switching current (after tunnelling) and OFF state switching current (after injection) was 1.5, 1.35 and 6 for devices 2, 3 and 4, respectively. The ratios correlated with the expected results in terms of FG transistor area: a high area results in a higher ratio. Despite the correlation, devices 2 and 3 may be unsuitable because the ratio is less than 2: a smaller ratio between the ON and OFF states could also result in higher losses. The Dickson CP achieved an output voltage of 2.96 V from an input of 1.2 V compared to 3.08 V as computed from the theoretical analysis and 4.5 V from the simulation results. The prototypes of the voltage doubler CP did not perform as expected: a maximum of 1 V was achieved compared to 4.1 – 5 V as in the simulation results.

The suitability of FG devices for switching applications depends on the ratio of the ON and OFF states (associated to insertion and isolation losses): the larger the FG transistor area, the higher the ratio. The reliability issues are dominated by the oxide thickness of the transistor, which contributes to charge leakages and charge trapping: smaller transistor length causes more uncertainties. Charge trapping in the oxide increases the probability of leakages and substrate conduction, thus introduces more losses. Based on the findings of this research work, the FG devices promise to be suitable for mm-wave switching applications and there is a need for further research investigation to characterise the devices in the mm-wave region fully.

OPSOMMING

LADINGSPOMPE EN WISSELHEK TOESTELLE VIR SKAKEL TOEPASSINGS

deur

Bongani Christopher Mabuza

Studieleier: Prof. S. Sinha

Departement: Departement van Elektriese, Elektroniese & Rekenaaringenieurswese

Universiteit: Universiteit van Pretoria

Graad: Magister in Ingenieurswese (Mikro-elektroniese Ingenieurswese)

Sleutelwoorde: Veldprogrammeerbare hekskikkings, CMOS-tegnologie, swewendehek-transistor, laaipompe, sinversterkers, CHEI, FN-tonnelling, millimetergolf- geïntegreerde stroombane, Dicksonpompe, tweefaseklok, snelheidversadiging, EEPROM, sekering, mikro-elektroniese meganiese stelsels en spanningverdubbelaar.

Impedansie-instelling op skyf word gebruik om steurings in geïntegreerde stroombane wat deur verpakkingstres veroorsaak word, te oorkom. Instelling is meer belangrik om netwerke van radiofrekwensiesisteme te paar. Moontlike verpakkingresonansie en variasies in die vervaardigingsproses kan onstabiliteit veroorsaak, wat ’n groot probleem is in radiofrekwensiesisteme. Voorsorg moet dus getref word om die oorhoofse stabiliteit van komponente en die finale sisteem self te handhaaf. Elektries uitveebare programmeerbare slegs-lees-geheueskakelaars (EEPROMs) neem minder matrysarea op as e-sekerings en die sekerings van mikro-elektromeganiese sisteme en word dus voorgestel vir gebruik as instellingskakelaars in millimetergolfaanwendings. Daar word verwag dat EEPROM-skakelaars ook multi-tydprogrammering sal moontlik maak as gevolg van die kleiner area en die feit dat meer skakelaars gebruik kan word vir fyn instellings.

Die probleem wat in hierdie navorsing aandag geniet, is die geskiktheid van EEPROMS vir skakelaanwendings in die millimetergolfstreek. The hooffokus van die verhandeling is om die geskiktheid van EEPROM-skakelaars kwalitatief te karakteriseer vir instelling met sisteme wat in die millimetergolfspektrum funksioneer.

130 nm SiGe BiCMOS IBM 8HP-prosestegnologie is gebruik vir simulasie en die vervaardigde prototipes. Die Dickson-laaipomp is gebruik vir simulasie en die vervaardigde prototipes. Die Dickson-laaipomp, twee spanningverdubbelinglaaipompe en vier swewendehektoestelle is ondersoek. Literatuur- en teoretiese verifikasie is gedoen met behulp van rekenaarondersteunde-ontwerp (CAD) Cadence-sagteware deur stroombaananalise en die uitleg is ook ontwerp vir die vervaardiging van geïntegreerdestroombaanprototipes. Die kwalitatiewe evaluasie van die hipotese is gebaseer op die ondersoek van betroubaarheidkwessies, skakelingeienskappe, laaipompuitsetdryfvermoë en millimetergolfkarakterisering.

Die maksimum gemete dreineerstroom vir swewende hekke was 1.4 mA, 2.7 mA en 3 mA vir onderskeidelik toestelle 2, 3 en 4. Die verhouding tussen die AAN-toestand van die skakelstroom (na tonnelling) en die AF-toestand van die skakelstroom (na inspuiting) was 1.5, 1.35 en 6 vir toestelle 2, 3 en 4, onderskeidelik. Die verhoudings het ooreengestem met die verwagte resultate rakende die swewendehek-transistorareas: ’n groot area het ’n hoër verhouding tot gevolg. Nieteenstaande die ooreenstemming, mag toestelle 2 en 3 moontlik nie geskik wees nie, omdat die verhouding kleiner as 2 is: ’n kleiner verhouding tussen die AAN- en AF-toestande mag ook hoër verliese tot gevolg hê. Die Dickson-laaipomp het ’n uitsetspanning van 2.96 V vanaf ’n inset van 1.2 V vergeleke met 3.08 V soos bereken volgens die teoretiese analise en 4.5 V volgens die simulasieresultate. Die prototipes van die spanningverdubbelinglaaipomp het nie gefunksioneer soos verwag is nie: ’n maksimum van 1 V is bereik vergeleke met 4.1 – 5 V soos in die simulasieresultate.

Die geskiktheid van swewendehektoestelle vir skakelingtoepassings hang af van die verhouding van die AAN- en AF-toestande (wat met invoer-en isolasieverlies geassosieer word): hoe groter die swewendehektransistorarea, hoe hoër die verhouding. Die betroubaarheidkwessies word oorheers deur die oksieddikte van die transistor, wat bydra tot ladinglekkasies en ladingvasvangs: korter transistorlengte veroorsaak meer onsekerheid. Ladingvasvangs in die oksied verhoog die moontlikheid van lekkasies en substraatgeleiding en veroorsaak dus groter verlies. Die bevindings van hierdie navorsing toon dat swewendehektoestelle waarskynlik geskik is vir millimetergolfaanwendings en verdere navorsing is nodig om die toestelle volledig in die millimetergolfstreek te karakteriseer.

ACKNOWLEDGEMENTS

The author would like thank some of the people who were involved during this research for an MEng degree (Microelectronic Engineering).

- Mrs Louise Euthimiou, Coordinator: International Relations (Department of International Relations), University of Pretoria, for extending bursary opportunities to all students at our university.
- Prof. Roelf Sandenbergh, Dean of the Faculty (Engineering, Built Environment and Information Technology), University of Pretoria for his guidance on my bursary applications to the faculty.
- My supervisor, Prof. Saurabh Sinha, for his support and availability during the entire process.
- Dr Daniel Foty, President: Gilgamesh Associates, LLC, for his introduction to Prof. Paul Hasler and also for making a recommendation to GT.
- The research group at GT for allowing me to use their devices and equipment. Special thanks to Farhan Adil for his assistance in enabling the measurement and Prof. Paul Hasler, who was my supervisor at GT.
- MOSIS Educational Program (MEP) for sponsoring the fabrication of the prototype IC.
- Erik-Jan Moes, microwave design engineer at SAAB Electronic Defence Systems (EDS), for his assistance in designing and manufacturing the PCB used for capturing measurements.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AMS	Analogue/mixed signal
BBHE	Band-to-band hot electron
BiCMOS	Bipolar CMOS
CAD	Computer-aided design
CHE	Channel hot electron
CHEI	CHE injection
CMOS	Complementary metal-oxide semiconductor
CP	Charge pump
DC	Direct current
DG	Drawing
DRC	Design rule check
EEPROM	Electrically erasable programmable read-only memory
FET	Field effect transistor
FG	Floating gate
FN	Fowler-Nordheim
GT	Georgia Tech
HBT	Heterojunction bipolar transistor
IBM	International Business Machines
IC	Integrated circuit
LVS	Layout versus schematic
MEMS	Microelectromechanical systems
MEP	MOSIS educational program
MOSFET	Metal-oxide semiconductor field effect transistor
mm	Millimetre
MOSIS	Metal Oxide Semiconductor Implementation Service
MPW	Multi-project wafer
NDA	Non-disclosure agreement
NM	Noise margin
PCB	Printed circuit board
PDK	Process design kit
RF	Radio frequency

RFID	Radio frequency identification
ROM	Read-only memory
SiGe	Silicon germanium
SPICE	Simulation Program with Integrated Circuit Emphasis
TEM	Transverse electromagnetic

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND TO THE RESEARCH

The demand for wireless transmission bandwidth, especially for high-volume data such as video streaming and local area networks, is ever increasing. In order to meet this requirement, devices with large memory storage capacity have been developed. However, since there is an urgent need to transfer data (for sharing, backups and even for other essential downloads) , the drawbacks of low data transfer rates limit the effectiveness of these devices. In complementary metal-oxide semiconductor (CMOS) design, higher bandwidths have been achieved through device scaling [1]. However, scaling has reached its limit and for over-scaled devices, a compromise is made between performance and the cost of the advanced fabrication processes [2]. Furthermore, the bandwidth gains that are achieved through device scaling are still low and limited [3].

New technologies have been proposed and are promising to yield better performance for high frequency applications. Some of these technologies are targeting the available 5 GHz to 7 GHz unlicensed bandwidth around the 60 GHz region [4]. Thus, there is great interest in millimetre-wave (mm-wave) applications. Bipolar CMOS (BiCMOS) is such an enabling technology, especially for analogue and radio frequency (RF) applications [5], while CMOS remains dominant for digital applications.

The frequency characteristics of transistors are dominated by parasitic parameters, capacitance and resistance [4], the *RC* time constant. The values of these parameters can differ from one fabrication process to another owing to operating temperatures, packaging stress and other controllable parameters such as supply voltage. Since designs are based on initial intrinsic values, these variations will also affect the system performance or specifications. In order to overcome this, post-assembly tuning is desired to compensate for such process variations and reduce packaging stress [3], [6]. Thus, reliable on-chip tuning circuits are required to vary the chip reactance accordingly. Testing against process variations can be expensive; by using programmable or reprogrammable tuning, testing cost can be greatly reduced. For example,

tuning can be used for the following applications: switch-mode regulators, oscillators, RF tags, radio frequency identification (RFID) transponders and transmission line matching networks [7].

The mm-wave applications have brought new challenges into the design of new systems [8]; the materials and operational behaviour have changed. Modelling of transistors is becoming more challenging and new models are necessary for analysis and circuit design [9]. Thus, performance evaluation needs to be done on current systems based on new developing technologies and a new design approach or methodology will need to be derived or followed based on these new challenges.

The aim of this research was to investigate and characterise the performance of electrically erasable programmable read-only memory (EEPROM) switches for mm-wave applications, and to test integration with other mm-wave modules and/or BiCMOS technology. The research questions were formulated based on this investigation and the research hypothesis was derived based on that.

1.2 RESEARCH PROBLEM AND HYPOTHESIS

The problem addressed in this research is:

- How suitable are EEPROMs for switching applications in the mm-wave region?

The EEPROMs depend on floating gate (FG) transistors' functionality. In addressing the research question, FG transistors have been evaluated against device scaling and for high-frequency operations. Charge pumps (CPs) are used to program the FG transistors and different CP structures have also been evaluated.

The following is the research hypothesis:

- If the current limits (or ratios) of the ON/OFF states are affected by device scaling, required for mm-wave applications, then EEPROMs' switching performance would not be suitable for mm-wave applications.

To expand on the primary research question, the following secondary research questions were addressed:

- For how long can a charge be retained in the FG device?
    - Charge retention depends on the oxide; as devices are scaled, the oxide thickness is also reduced, affecting charge retention. Leakages reduce the reliability of charge retention and also increase power dissipation.
- How is the output loading of CP affected by device scaling and thus reduced output currents?
    - The process technology limits voltages, current density, saturation velocity and other critical parameters; this affects device performance.
- How good are the switching characteristics (insertion and isolation losses) of FG devices?
    - Device scaling reduces the margin between the ON and OFF state, hence the insertion and isolation losses, which determine the switching performance, are affected.
- How will the FG devices behave at mm-wave frequencies; will high frequencies have a negative effect?
    - The process variations, interconnections and device parasitic affect performance and cannot be ignored for mm-wave applications. The voltage drop across interconnections limits operating voltages.

The questions addressed here are based on making decisions in terms of the programming method to be applied, the CP structure and a specific sense amplifier that will maximise the EEPROM's performance.

Read-only memory (ROM) programming methods can take two extremes: one with excellent density but very long process time or alternatively, one with better cycle time but poor density. There are different architectures (discussed in section 2.4.2) which trade off performance with density. Consequently, an architecture that provides acceptable performance for mm-wave applications has been investigated.

Because of scaling, the operating voltages and currents of transistors are reduced [10], [11], thus to measure low-level signals, sense amplifiers are used. Sense amplifiers have been investigated and evaluated for different structures [2], [11]-[12], i.e. differential or single-ended sense amplifiers. A conversion from a single-ended to a differential amplifier may be required for some designs. The choice depends on the memory device, voltage levels and overall memory architecture. The investigation relating to sense amplifiers determines the scaling and loading effects for integrating sense amplifiers with FG transistors.

The evaluation includes voltage references, power dissipation, and coupling and leakage currents. The evaluated parameters imposed limitations on the integrated circuit (IC) technologies suitable for mm-wave application. The threshold voltage, current, power and programming time parameters were also tested for high frequencies.

1.3 JUSTIFICATION OF THE RESEARCH

There are different methods by which tuning can be achieved and these are not limited to after-package trimming. RF circuits and matching networks are other examples where tuning is desired [7], [13]. As a result, the investigation and characterisation of EEPROM switches will play a vital role in terms of where and when tuning is required. The research questions addressed in this dissertation have enabled the evaluation of different structures that will yield better performance, reliability, low-cost and power-efficient (low-power) systems. Since EEPROMs occupy less area, more cells can be used and fine-tuning can be achieved. An advantage of having more cells is that fine-tuning can be easily achieved, thus the aim is to have as many cells as possible, while maintaining the desired performance. The number of cells, however, requires more power, implying a design trade-off.

EEPROM technology is a proven and applied technology, but its performance for mm-wave frequency has not been fully determined yet. At this stage, EEPROM switches seem like an ideal technology for reactance trimming against packaging stress and tuning for matching networks. This research has contributed to investigating, characterising and validating the suitability of EEPROM switches for mm-wave applications.

The size of antennas decrease as the frequency increases [3], thus at mm-wave the size of the antennas will be small enough to create possibility for antennas to be developed "on-package." On-chip impedance matching or tuning will be desired for such on-package antennas. The application of EEPROM switches at mm-wave can also be extended to tuneable filters and resonators [14]-[15]. It is evident that the application of EEPROM switches will contribute to current and future technologies, with mm-wave being one of the applications.

1.4 RESEARCH METHODOLOGY

A typical scientific method to be followed is shown in Figure 1.1 [16]-[17]. New technology developments are required to provide the same or better performance within the existing theories or observations from experiments. Thus, the hypothesis is formulated and validated based on the limitations or technology constraints. System description is based on design concepts and some of the modules are part of the design hypothesis, which needed to be verified.

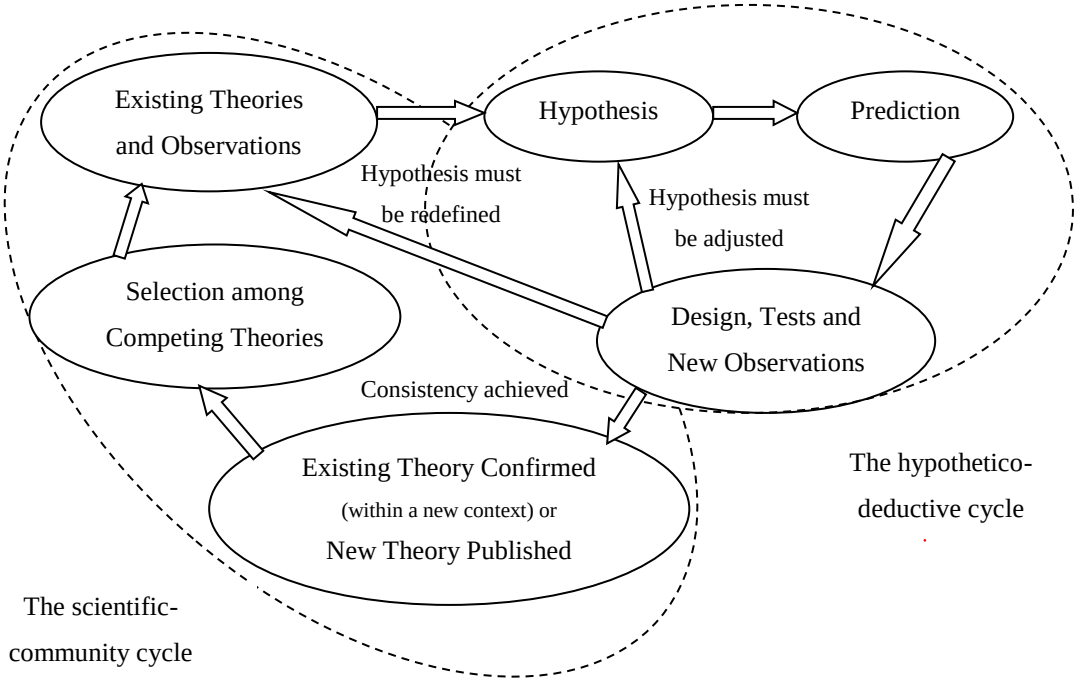


Figure 1.1. Research methodology illustrated [17], printed with permission.

The methodology shown in Figure 1.1 illustrates the scientific process followed during the research. The hypothesis needed to be either adjusted or redefined during the process until there was a correlation between the predicted outcome and the existing theories. There also had to be consistency during testing and when observations were done. Competing theories were compared and selected based on their performance for specific applications; the focus was on mm-wave applications.

1.5 OUTLINE OF THE DISSERTATION

Chapter 1 (*Introduction*)

This chapter briefly describes the field of study and the focus of this research. The motivation for relevance and the importance of this research are justified in this chapter. Delimitations of scope and key assumptions are also presented.

Chapter 2 (*Literature review*)

Previous relevant literature is reviewed and key research issues and challenges are identified. Different models and structures are evaluated to build a theoretical foundation for the research.

Chapter 3 (*Methodology*)

This chapter outlines the research methodology used to collect data. The experimental setup, qualification protocols and instruments used for prototype testing are described in detail. The process technology and models used are also presented. The roadmap developed in this chapter is the motivation for the chosen methodology and research decisions taken based on assumptions and approximations.

Chapter 4 (*Prototype designs and CAD software simulations*)

The prototype designs are developed from the mathematical and theoretical models presented in chapters 2 and 3. Simulation circuits and results are discussed. The prototype was fabricated in view of validating the research hypothesis.

Chapter 5 (*Results*)

The results and discussion of the practical measurements are presented here; reference is also made to the theoretical (simulation) results presented in chapter 4. The simulation and measurement results are evaluated for their relevancy to the research hypothesis. Results measured at Georgia Tech (GT) and those of the fabricated IC are discussed in this chapter.

Chapter 6 (*Conclusion*)

This chapter summarises the theories and hypothesis presented in this dissertation and provides an opinion regarding these based on the results obtained during mathematical, simulated, and measured results.

1.6 DELIMITATION OF SCOPE AND KEY ASSUMPTIONS

The delimitations applicable within the scope of work by fabrication process used, compact model approximations and key assumptions based on the literature review are given below:

- Scaling is directly proportional to high frequency, thus by determining the effects of scaling, deductions are made for high-frequency operations.
- The process technology places a limit on the design parameters (minimum transistor length, supply voltage, saturation velocity) and the operation ranges. Design parameters are specified by the foundry and restrict the designer.
- Compact models used for mm-wave frequencies are approximations, since there are no generalised models. Thus, the accuracy of measurements is limited by the best parameter fit of the compact models.
    - In other cases, parameters for design are extracted experimentally. Since no prior experiments have been performed, such design parameters are not known and are approximated, where possible, based on findings in literature.
- The initial charge of the FG is unknown and depends on the layout structure. For simulation purposes, an arbitrary value was chosen over a range of values.
- Current leakages through oxide and charge trapping during programming of FG devices cannot be determined directly. Thus, threshold voltage shifts and

changes in the drain currents are used to determine and measure the effects of leakages and charge trapping.

- Because of the number of fabricated devices (10 samples sponsored by Metal Oxide Semiconductor Implementation Service (MOSIS) Educational Program [MEP] were received), the number of samples was limited and not enough to perform statistical analysis. The measurements are also taken from low risk to high risk in terms of causing device failure or damaging the IC.

The research focus was on evaluating the suitability of using EEPROM for mm-wave applications. In addressing this, key sub-systems of EEPROM are evaluated in this research, i.e. CPs (to provide programming voltage) and FG devices (the cell/core of EEPROMs). Sense amplifiers have been evaluated according to the literature review only; no devices have been fabricated.

1.7 CONCLUSION

Silicon germanium (SiGe) BiCMOS is the preferred technology for mm-wave applications. The advantage of SiGe BiCMOS is its mixed signal isolation that allows analogue and digital modules to be easily integrated in a two-chip solution. (high-performance low-power circuits (SiGe) can be combined with analogue circuits (CMOS) in the same process, as separation of the SiGe BiCMOS (SiGe and CMOS) is possible through signal isolation), something that is less possible with CMOS. Highly scaled CMOS devices have difficulty with high-power applications, which is another gap addressed by BiCMOS. This is more evident at mm-wave frequencies, since more device scaling is required. SiGe BiCMOS is, therefore, a suitable technology for RF, analogue and power management applications.

Full application of e-fuses, microelectromechanical systems (MEMS) and EEPROMs for mm-wave is still an area of continuing research and more characterisation of these devices is still required for mm-wave applications. MEMS provide better performance for isolation and insertion losses, but occupy a larger area than EEPROMs. Production of EEPROMs can easily be repeated, unlike that of MEMS. Cost tends to go down if design or production can be repeated easily, thus this would mean lower cost for EEPROMs and higher ones for MEMS.

There is consequently a trade-off between cost, size, speed and reliability for these devices. Despite this, the main determining factor is the performance of these devices in the mm-wave spectrum.

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION

Although mm-wave technology exhibits a number of advantages such as small size, low mass, wide bandwidth or reduced interaction with other services, tolerance requirements due to shorter wavelengths, increased reflections at discontinuities and interconnections, and possible package resonances require precautions during design, fabrication and assembly [6]. Thus, an IC may work well at wafer level, but as a result of packaging and interconnection, it may change its behaviour and not work as expected. One method of adjusting IC perturbations is to have on-chip circuitry, modules or components that can be used to adjust the reactance of the IC. This is done by switching mechanisms that either connect or disconnect reactance within an IC. Tuning switches can be implemented by using fuses, MEMS or EEPROM.

2.2 TECHNOLOGY ISSUES

2.2.1 Interconnection

An important aspect of interconnections is that at high frequencies, parasitic components cannot be ignored. Thus, there may be a considerable voltage (IR) drop over the interconnecting wires, depending on the length of interconnections. An interconnection is electrically short if it is physically shorter than one-tenth of a wavelength, λ [18]. Interconnection effects include ringing, signal delay, distortion, reflections and crosstalk.

Influences of an interconnection are described through edge, proximity and skin effects [18]. Edge effects influence interconnection parameters in the low to medium frequency region and causes current to concentrate near sharp edges of the conductor, thus increasing resistance. The proximity effect causes current to concentrate in the sections of ground planes that are close to a signal conductor. The skin effect causes the current to concentrate in a thin layer at the conductor surface and has a greater influence at high frequencies. The skin depth, average depth of current penetration, is given by

$$\delta = \sqrt{\frac{2\rho}{\omega\mu\mu_0}}\ , \tag{2.1}$$

where, ρ is the volume resistivity, ω is the angular frequency and μ_0 is the magnetic permeability with a value of 1.257 µH/m. With aggressive CMOS scaling, distributed models (quasi-transverse electromagnetic [TEM] approximation) have become inaccurate in describing the interconnection performance. Full models that account for all possible field components and satisfying boundary conditions are required to give accurate estimation of the high-frequency effects [18]. A full analysis is required if the dielectric heights are comparable to the wavelength of propagation given by

$$\lambda_{\text{effective}} = \frac{v}{f}, \tag{2.2}$$

where v is the velocity of the wave and f is the frequency. Since operating voltages are low, interconnections constrain supply voltages. Another problem is that interconnections that are close to one another may act as a virtual short circuit. This effect is more evident in small-scale devices, such as those operating in mm-wave. Thus, low-k materials for inter-poly film are investigated to overcome this effect between signal conductors.

2.2.2 Packaging

One of the most important problems in RF systems is the overall stability of components and the final system itself. Thus, to reduce the effects of packaging, ceramic packaging is used in preference to plastic packaging. The penalty paid is that ceramic packaging is more expensive. Another after-package uncertainty is that intensive testing is required and such a testing process can be very expensive, which also contributes to material cost.

Plastic packaging is known to stress ICs mechanically, which can cause voltage threshold changes. Therefore, in-package trim is important in order to achieve high accuracy. Tuning (or trimming) is also required for chip identification, RF tuning and trimming impedance networks. The methods, characterisation and challenges of tuning for mm-wave applications have been discussed in previous work that forms part of this research [19].

Because of scaling and mm-wave frequencies, packaging cannot be designed independently from the overall system design. The circuit density is increasing, thus interconnections and chip heat dissipation are also increasing. At high frequencies, a signal propagating through a package is degraded owing to reflections and line resistance [18]. Hence, reliable package design, thermal compensation and controllable packaging processes are desirable.

2.2.3 Scaling

CMOS scaling has achieved its success through constant field scaling. Constant field scaling requires that as sizes are scaled down, the power supply must also be scaled down in proportion. As discussed in [3], the power supply is limited by the thermal voltage and thus the supply will saturate at about 1 V and its scaling will be less than the actual technology scaling below 180 nm. As a consequence, constant field scaling is violated and the field increases with technology scaling. This gives rise to the following issues [18]: transistor short-channel effects, device parameter variations, excessive sub-threshold leakage, junction leakage and gate oxide leakage. Other parameters that do not scale as desired are the intrinsic resistance and capacitance, which also limit delay improvements.

Increasing the electric field causes current leakages, thus high-*K* dielectrics are considered for gate dielectric materials [20]. The limit is that thermal emission becomes dominant for materials with very high *K* values and the bandgap reduces with increasing permittivity. Thus, a very high *K* value is not as effective for gate dielectric applications. A future trend may be to use a high *K* dielectric with metal gates.

Current leakages result from low and non-scalable threshold voltages. A high threshold voltage minimises leakage current and low threshold voltage enhances performance [18]. System performance can be optimised by using low threshold transistors (performance transistors) for critical paths and the high threshold transistor for the rest of the system. Through this optimisation, delay improvements of about 24 % can be achieved [18].

Scaling has also phased out the famous square-law metal-oxide semiconductor field-effect transistor (MOSFET) current equation. Since the field increases with technology scaling beyond 180 nm, velocity saturation is reached and thus, the square-law no longer holds. New current models must take velocity saturation into account.

2.2.4 CMOS

The scaling of devices is inversely proportional to frequency; as frequency increases the devices become smaller and smaller, thus there is high scaling of devices at mm-wave. An advantage of device scaling is that it also reduces operating voltages, thus low-voltage ratings can be achieved. These devices are then operating at low power. Despite this, it poses a new challenge to power devices and it is difficult to integrate CMOS with the power devices [5]. This is one of the setbacks of highly scaled CMOS devices and is even more important for RF circuits. For systems operating in the low GHz frequency (< 5 GHz), device scaling does not have a serious effect, as opposed to several GHz systems where parasitics of the devices cannot be ignored and have a greater influence on device performance [4]. While characterised - transistor modelling based on simulation program with integrated circuit emphasis (SPICE) is frequently available from IC foundries, and models such as the vertical bipolar inter-company and high-current model exist, usage of these models for selected circuits within the mm-wave spectrum is still required.

2.2.5 BiCMOS

Silicon ICs are well-suited to low-cost and high-volume microprocessors and memory applications, while RF circuits and mm-wave applications that operate at higher frequencies put more restrictive performance demands on the transistor building blocks [5]. This has led to the advancement of SiGe heterojunction bipolar transistor (HBT) technology and the progress of bringing this technology to reality has been exceptionally rapid. The SiGe HBT is a promising candidate for silicon-based monolithic integrated mm-wave circuits [21], [22]. SiGe has a lower bandgap and even in the case of high base doping, high emitter efficiency is achieved. The transit frequency f_T and the maximum oscillation frequency $f_{\max}$ are other important parameters of interest ($f_{\max} \geq f_T$). The maximum oscillation frequency for HBT is given by

$$f_{\max} = \sqrt{\frac{f_T}{8\pi C_{bc} r_b}}, \tag{2.3}$$

where, r_b is the base resistance, f_T is transition frequency and C_{bc} is the base collector capacitance. From (2.3) it can be seen that f_T and $f_{\max}$ are influenced by the base resistance and

capacitance and a trade-off occurs between the parasitic parameters and the maximum achievable frequency [5]. However, reducing building block power consumption remains an obstacle owing to the high V_{BE} of the SiGe HBT, which ultimately limits the supply voltage [7]. The f_T can also be traded against higher voltage breakdown V_B, for large signal RF drivers and power amplifiers [5].

BiCMOS technology takes the advantage of SiGe HBT and CMOS scaling across technology nodes and reduces the number of stacked transistors [1]. Thus, the performance improvement of SiGe BiCMOS follows from the performance roadmap of both bipolar and CMOS devices. In BiCMOS technology, the combination of self-aligned structures and oxide insulation allows extreme downscaling and also the realisation of extremely low parasitic capacitances [22]. The main advantage of SiGe BiCMOS technology is on-chip isolation, which is difficult to achieve in only CMOS-based technology.

On-chip isolation provides the ability to have mixed signal integration and this contributes to cost reduction. The integration of higher density passive elements, enabling scaling of analogue functions, can be integrated with new modules enabling a higher level of functional integration [5]. This is essential for mm-wave applications, since the device sizes are small and there is room to integrate RF modules with digital and analogue modules. The main advantages of SiGe BiCMOS technology is in RF, analogue and power management applications.

2.2.6 MOSFET and capacitance modelling

The MOSFET models are meant for long channel devices (> 0.18 μm) and short channel devices (≤ 0.18 μm). The short channel design model will be presented here [32].

2.2.6.1 Short channel

When V_{GS} (gate-to-source voltage) ≥ V_{th} (threshold voltage), the transistor operates either the saturation or linear region. If

$$V_{DS} \geq \frac{(V_{GS} - V_{th})E_C L}{V_{GS} - V_{th} + E_C L}, \tag{2.4}$$

then the transistor operates in the saturation region and drain current is given by

$$I_{DS} = W v_{SAT} C_{OX} \frac{(V_{GS} - V_{th})^2}{V_{GS} - V_{th} + E_C L}, \tag{2.5}$$

where V_{DS} is the drain-to-source voltage, E_C is the critical field, L is the length of the transistor, I_{DS} is the drain-to-source current, W is the width of the transistor, v_{SAT} is the saturation velocity, and C_{OX} is the oxide capacitance. For the purpose of this dissertation, process parameter values are not documented due to a non-disclosure agreement (NDA) with the foundry.

If

$$V_{DS} < \frac{(V_{GS} - V_{th}) E_C L}{V_{GS} - V_{th} + E_C L}, \tag{2.6}$$

the transistor operates in the linear region and drain current is given by

$$I_{DS} = \frac{W}{L} \frac{\mu_e C_{OX}}{1 + \frac{V_{DS}}{E_C L}} \left(V_{GS} - V_{th} - \frac{V_{DS}}{2} \right) V_{DS}, \tag{2.7}$$

where μ_e range from 574.19 – 660.21 cm^2/V.s (μ_p range from 81.16 – 87.11 cm^2/V.s) is the electron mobility. If $V_{GS} < V_{th}$, the transistor operates in the sub-threshold region.
The critical field, E_C, is determined by [30]

$$E_C = \frac{2 v_{SAT}}{\mu_V} \tag{2.8}$$

and

$$\mu_V = \frac{U0}{1 + (UA + UC + V_{BS}) \left(\frac{V_{GS} - V_{th}}{t_{OX}} \right) + V_B \left(\frac{V_{GS} - V_{th}}{t_{OX}} \right)^2}, \tag{2.9}$$

where V_{BS} is the bulk-to-source voltage, t_{OX} = 3.2 nm is the oxide thickness and V_B is the bulk voltage. *U*0, *UA* and *UC* are parameters specified by foundry process technology. The MOSFET model is shown in Figure 2.1. The gate capacitance and the diffusion capacitance are also shown.

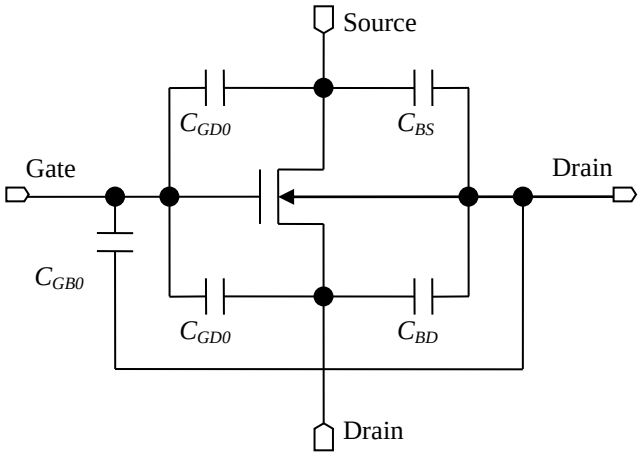


Figure 2.1. Transistor model with intrinsic capacitors.

The parasitic capacitance and all other capacitors are determined at each critical node to determine the effect on performance and also to compute the system delay and power. The computation of the capacitances in Figure 2.1 is shown in Table 2.1 [32].

Table 2.1. Parameters for intrinsic capacitors for all regions of operation [32]. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, New Jersey.

Parameter	Cut-off	Linear	Saturation
C_{GS}	C_{ol}	$C_{ol} + C_g/2$	$C_{ol} + 2C_g/3$
C_{GD}	C_{ol}	$C_{ol} + C_g/2$	C_{ol}
C_{GB}	$(C_g/2 + C_{jc}/2) < C_{gb} < C_g$	0	0

The junction capacitance in Table 2.1 can be determined by using a less accurate level 1 model or a higher level BSIM3 model. The drain and source junction capacitance for the level 1 model are given by

$$C_{jD} = \frac{C_J AD}{\left(1 - \frac{V_j}{PB}\right)^{M_J}} + \frac{C_{JSW} PD}{\left(1 - \frac{V_j}{PB}\right)^{M_{JSW}}}, \tag{2.10}$$

$$C_{jS} = \frac{C_J AS}{\left(1 - \frac{V_j}{PB}\right)^{M_J}} + \frac{C_{JSW} PS}{\left(1 - \frac{V_j}{PB}\right)^{M_{JSW}}}. \tag{2.11}$$

C_J is the zero-bias bulk junction bottom capacitance, C_{JSW} is the zero-bias bulk junction sidewall capacitance, V_j is the positive forward bias junction potential, PB is the bulk junction potential, M_J is the bulk junction grading coefficient and M_{JSW} is the bulk junction grading coefficient parameters specified by foundry process technology. AD is the junction drain area, AS is the junction source area, PD is the perimeter of the drain and PS is the perimeter of the source These parameters are layout-dependent. The BSIM3 model is slightly different from level 1 and the equivalent equations are

$$C_{jD} = \frac{C_J AD}{\left(1 - \frac{V_j}{PB}\right)^{M_J}} + \frac{C_{JSWG} PD}{\left(1 - \frac{V_j}{PB}\right)^{M_{JSWG}}}, \tag{2.12}$$

$$C_{jS} = \frac{C_J AS}{\left(1 - \frac{V_j}{PB}\right)^{M_J}} + \frac{C_{JSWG} PS}{\left(1 - \frac{V_j}{PB}\right)^{M_{JSWG}}}, \tag{2.13}$$

where M_{JSWG} is the gate-side sidewall grading coefficient and C_{JSWG} is the gate-side sidewall junction capacitance. The parasitic resistance, R_{DS}, is computed as follows,

$$R_{DS} = \frac{R_{DSW}}{W_{eff}}, \tag{2.14}$$

where R_{DSW} is the parasitic resistance per unit width.

2.3 TUNING METHODS

2.3.1 E-fuse

The types of fuses that are available include metal fuses, polysilicon fuses and lasercut fuses. All these fuses, before advanced polysilicon fuses, had the disadvantages of size, high blowing currents and a large separation distance required between fuses and all were limited to on-chip programming [23]. Another preferred requirement for tuning is the ability to tune on-chip and on-package. In addressing this, electrically programmable fuses (e-fuse) have been developed [23]-[25][27]. The electrical characterisation of the e-fuse is based on four things: the blowing current, the blowing voltage, the blowing time and the resistance before and after blowing.

Typical resistance values before and after blowing are in the range of <200 Ω and >1 kΩ. In 0.13 µm SiGe BiCMOS technology, values of 200 Ω and 5 kΩ have been reported; refer to Table 2.2 [25]. Programming currents can range between 10 mA and 40 mA and voltages between 3.3 V and 6 V. The fuses can be blown to achieve both open and short circuit and this process is irreversible. Field effect transistors (FETs) or personal identification number (PIN) diodes can be used, but the turn-on voltage, poor isolation and insertion loss make the use of these devices unfavourable.

Table 2.2. E-fuses across technologies [25] © 2007 IEEE.

Node [nm]	Programming voltage [V]	Pulse width [ms]	Pre-resistance [Ω]	Post-resistance [kΩ]
350	5.5	1	≤ 100	≥ 1
250	4.5	1	≤ 200	≥ 2
180	3.6	0.5	≤ 200	≥ 8
130	3.3	0.2	≤ 200	≥ 5

An ideal case for post-resistance is that its value should increase with technology scaling; based on Table 2.2 moving from 350 to 130 nm, this seemed to be the case, with the resistance increasing from 1 to 5 kΩ. The prediction based on this is that further scaling will result in even lower values, thus e-fuses may not be suitable for use in mm-wave applications.

2.3.2 MEMS

MEMS can also be used for tuning. There is considerable interest in the development of meso-scale MEMS devices fabricated using printed circuit board (PCB) processing techniques. PCB-MEMS technology is compatible with multi-chip module lamination technology [14]. Therefore, monolithic integration of embedded passives and surface mount electronics could be accomplished using well-established assembly and packaging techniques. Another advantage of MEMS is that RF-MEMS switches show better performance such as low loss, low power consumption and no measurable inter-modulation.

The types of switches that can be implemented with MEMS include capacitive shunt switches, ohmic series relays, capacitive series relays and capacitive series switches. The capacitive shunt switch and the ohmic series relay are the preferred types [28]. The series-mode passes a signal in actuated state, while the shunt-mode terminates transmission in the actuated state. The capacitive shunt is more compact and suitable for high frequencies.

A challenge or limitation for MEMS is to achieve fast switching, IC compatible devices and low actuation voltage. These disadvantages and challenges make MEMS a non-ideal alternative, when compared to semiconductor-based switches, but some of these challenges are addressed in the literature [14], [28] and improved performance is expected for future developments.

2.3.3 EEPROMS

Another alternative method for tuning is using EEPROM switches. The advantages of EEPROM are that it offers multi-time programming and is easy to trim once the die has been packaged. It occupies a small area and thus more cells can be included. This will also enable fine–tuning, which is more essential for RF and matching networks. One of the concerns about e-fuses is the programming current, but EEPROMs require low current for trimming. A disadvantage that is inherent in EEPROMs is their process cost. Previous EEPROMs were fabricated in multi-polysilicon processes and multi-oxidation for thin SiO_2 layers [7]. Thus, many masks were required, which resulted in longer process turnaround time, lower yield, higher cost and lower reliability.

Over the years single-poly EEPROMs have been investigated and developed for standard CMOS processes [29]-[31]. This technology advancement enabled on-chip EEPROM to be realised easily in digital and analogue circuits, and there are also merits in terms of lower cost, shorter turnaround time and higher reliability circuits. Another disadvantage of multi-poly EEPROMs is the high-voltage programming requirement, which is unfavourable for embedded applications. A novel single-poly EEPROM has been proposed [31] and the voltage levels for single poly have been lowered from 10 V to 6.5 V.

EEPROMs belong to a class of non-volatile memories, which are tabulated in Table 2.3 [32]. EPROMs were replaced by EEPROMs, because of their ineffective erasing method of ultra-violet light, high-write voltages due to hot electron injection and the low program/erase cycle. The EEPROMs use a reversible process for writing and erasing, Fowler-Nordheim (FN) tunnelling. A disadvantage is that an additional transistor is required to control the injection and removal of electrons in the FG.

Table 2.3. Comparison between non-volatile memories [32]. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, New Jersey.

Quantity	Cell No. of Trans.	Area (ratio wrt EPROM)	Erase-method	Write-method	Erase-voltage	Write-voltage	Program/erase cycle
Mask ROM	1 T	0.35 – 5	-	-	-	V_{DD}	0
EPROM	1 T	1	UV exposure	Hot Electrons	V_{PP}	V_{DD}	~100
EEPROM	2 T	3 – 5	FN Tunnelling	Hot Electrons	V_{PP} (int)	V_{DD}	104 – 105
Flash memory	1 T	1 – 2	FN Tunnelling	Hot Electrons	V_{PP}	V_{DD}	104 – 105
			FN Tunnelling	FN Tunnelling	V_{PP} (int)		

The introduction of flash memory was to combine the advantages of EPROMs and EEPROMs by using one transistor and an effective erasing method, as shown in Table 2.3. This increased the programming cycles. The flash memory still uses hot electron injection for writing a cell and unlike EEPROMs, it does not have the flexibility of erasing one cell at a time. To circumvent the problem of using an additional transistor for EEPROM, a dual FG has been proposed in the literature and is described in section 2.4.1.

2.4 EEPROM SWITCHES

The memory of EEPROM designs is evaluated based on different design metrics [32], namely:

- cost expressed by complexity and area,
- integrity and robustness, expressed by static (steady-state) behaviour,
- performance, determined by the dynamic (transient) response, and
- energy efficiency set by power consumption.

For static behaviour, switching thresholds and noise margins (NMs) need to be considered. The parameters of interest or design include threshold voltages (V_T, $V_{TP,N}$) and saturation voltages (V_{CESAT}, V_{DSSAT}). The dynamic behaviour depends on the charging and discharging rate of the capacitors, which also contributes to propagation delay. The minimum propagation delay is determined by intrinsic parameters and can be longer, depending on cascaded networks or external capacitances. The parameters of interest include intrinsic capacitances and resistances. These parameters also affect the maximum frequency of the transistors, which determines the switching rate.

Energy or power dissipation can be divided into three parts: dynamic power, static power and direct-path current dissipation. Dynamic power is dependent on the charging and discharging of the capacitors. Direct-path dissipation is due to the non-zero rise-and-fall time of the switching clock. Static or steady-state dissipation depends on the steady-state current and the supply voltage. The emphasis on low power dissipation is critical and the design should minimise the power dissipation. This is even more important for portable and wireless devices, since these devices have limited power supply.

There are different structures by which memory cells can be implemented and the cells are characterised by three operations; writing, erasing and reading operations. A channel hot electron (CHE) injection and FN tunnelling can be used for writing the cell. FN tunnelling is done between the gate and n^+ diffusion in the NMOS. Erasing the cell can be achieved by tunnelling between the gate and p^+ of the PMOS or between the n^+ and gate of the NMOS. The mode of operation is determined by the C_{PG} / C_{GN} ratio [14]. The general structure of the FG transistor [33] is shown in Figure 2.2.

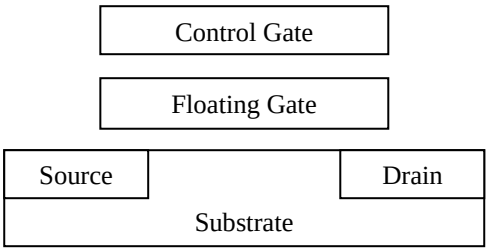


Figure 2.2. Cross-section of a floating gate transistor [33] © 1997 IEEE.

The capacitance at the drain and source in Figure 2.2 changes as the charge is accumulated or removed from the FG. Since the source and drain are doped differently, the drain and source capacitance would differ. Some EEPROM cells are programmed and erased via FN tunnelling and are composed of two transistors. In flash memory, the cell is programmed and erased electrically by a single transistor. A dual-control gate EEPROM cell can be used to eliminate the select transistor. Another advantage of a dual-control gate is that it reduces hot-hole injection degradation induced by band-to-band tunnelling [34].

2.4.1 Floating gate transistors – model description

The FG transistor operation is similar to a standard transistor, except that the threshold voltage is varied based on the charge stored in the FG. Thus, the linear and saturation drain currents are modified as given in the equations below:

$$I_D = \mu_n \cdot C_{ox} \frac{W}{L}\left(V_{FG} - V_{th} - \frac{V_D}{2}\right) \cdot V_D, \tag{2.15}$$

$$I_D = \frac{1}{2}\mu_n \cdot C_{ox} \frac{W}{L}(V_{FG} - V_{th})^2 . \tag{2.16}$$

Note that this is a standard current equation with gate voltage replaced by the FG voltage, V_{FG}. As mentioned earlier, owing to technology scaling, velocity saturation is caused by the increasing field. The velocity saturation affects the drain current by a velocity saturation factor of [18]

$$I_{D\_\text{modified}} = I_D \cdot \kappa(V_{DS}) \tag{2.17}$$

and

$$\kappa(V) = \frac{1}{1 + \left(V / \xi_C L \right)}, \tag{2.18}$$

where ξ_C is the critical field where velocity saturation begins.

For Figure 2.3, the FG voltage is found by using the charge neutrality equation and the FG charge equation.

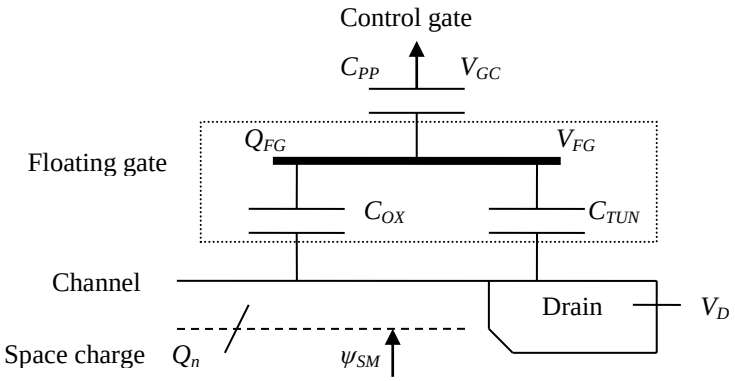


Figure 2.3. Electrical equivalent circuit of floating gate without the mechanism of electron tunnelling [35] © 2001, with permission from Elsevier.

The FG charge Q_{FG} of a single control gate in Figure 2.3 is given by [35]

$$\begin{aligned} Q_{FG} &= C_{PP}(V_{FG} - V_{GC}) + C_{OX}(V_{FG} - V_{FB} - \psi_{SM}), \\ &+ C_{TUN}(V_{FG} - V_D) \end{aligned} \tag{2.19}$$

where C_{PP} is the inter-poly capacitor, C_{OX} is the gate oxide capacitor, C_{TUN} is the tunnel oxide capacitor, V_{GC} is the control gate voltage, V_D is the drain voltage and ψ_{SM} is the average surface potential between drain and source. Note that V_{FG} depends on Q_{FG} and $Q_{FG,}$ which is a function of V_{FG}, where V_{FG} is computed iteratively. The charge above the channel and below FG is

$$Q_Z = \frac{C'_{ox}}{C_x}\left[Q_{FG} + C_{PP}V_{GC} + C_{TUN}V_D - (C_{PP} + C_{TUN})(V_{FB} + \psi_{SM})\right], \tag{2.20}$$

with

$$C_X = C_{PP} + C_{OX} + C_{TUN}\,. \tag{2.21}$$

The charge per unit area in substrate is

$$Q_{SC} = -\gamma \cdot C'_{OX} \sqrt{\psi_S - V_{th} + V_d}\,, \tag{2.22}$$

with

$$V_d = V_{th}\left[\exp\left(\frac{\psi_S - 2\psi_b - V_Y}{V_{th}}\right) + \exp\left(\frac{\psi_S}{V_{th}}\right)\right], \tag{2.23}$$

where γ is the substrate effect coefficient, ψ_S is the surface potential, ψ_b is the bulk Fermi level, V_{th} is the thermal voltage and V_Y is the potential along the channel. Q_{SC} can be approximated for weak drain potential in inversion mode by [11]

$$Q_{SC} = -(C_{HV} + C_{TUN})(V_{FG} - V_B), \tag{2.24}$$

where C_{HV} is the gate thick oxide capacitor. The amount of charge for the threshold voltage based on the FG charge is given by

$$\Delta V_{FG\_TH} = \frac{\Delta Q_{FG}}{C_T}\,, \tag{2.25}$$

where C_T is the sum of all capacitors that affect Q_{FG}. Note that the derivation of both single and dual structures is based on charge neutrality; the charge of FG and the charge above the channel and below FG must add to zero. The FG charge for a dual gate structure [11] is shown in Figure 2.4.

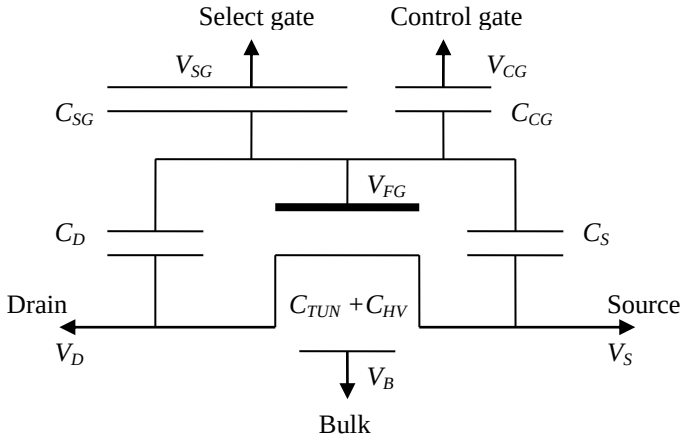


Figure 2.4. Electrical equivalent circuit of FG with dual control gate [11] © 2008, with permissions from Elsevier.

The charge in Figure 2.4 is given by

$$\begin{aligned} Q_{FG} &= C_{SG}(V_{FG}-V_{SG})+C_{CG}(V_{FG}-V_{CG})+C_D(V_{FG}-V_D) \\ &+C_S(V_{FG}-V_S)+(C_{HV}+C_{TUN})(V_{FG}-V_{FB}-\psi_S) \end{aligned}. \tag{2.26}$$

When comparing (2.23) and (2.30), which represent single- and dual-control gates, respectively, C_{SG} (select-to-gate capacitor) and C_{CG} (control-to-gate capacitor) are equivalent to $C_{PP,}$ the inter-poly capacitors, the sum of C_{HV}, and C_{TUN} is equal to C_{OX} and (2.30) also includes the tunnelling mechanism factors, C_D and C_S. The charge is evaluated iteratively by using the write and erase currents [33]. The tunnelling currents can be estimated as presented in [11].

FG devices have single- and two-poly structures. A two-poly FG synapse requires only a transistor and controlled sources and does not use the MOSFET's channel potential. The model can thus be implemented for any SPICE simulator. The synapse is a four-terminal device comprising two *p*-channel FETS; an FG *p*-channel FET for electron injection and readout, and a shorted *p*-channel FET for electron tunnelling. Only a single-poly device has been investigated in this dissertation.

A single-poly structure is used in [30] and it utilises an *n*-channel MOSFET and MOFET capacitor. The conceptual structure is shown in Figure 2.5.

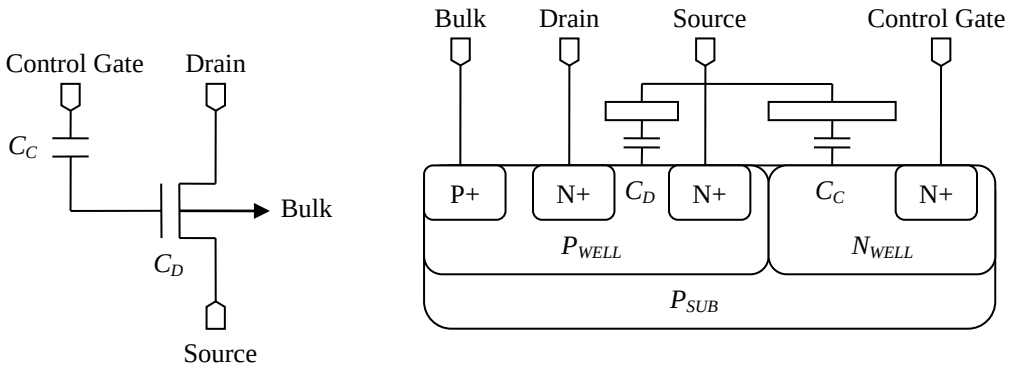


Figure 2.5. Structure of integrated transistor and control gate capacitor [30] © 2007, with permission from Elsevier.

The polysilicon in Figure 2.5 is shared between the NMOS gate and the MOSFET capacitor's top plate. Since the poly is not connected to an electrical lead, it forms an FG and the *n*-well serves as the control gate. Another structure that uses single poly is shown in Figure 2.6.

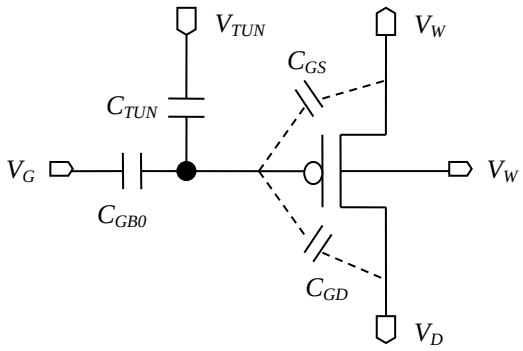


Figure 2.6. FG transistor model with control gate and tunnelling capacitors.

The FG in Figure 2.6 uses a control gate as in [30] for electron injection, but it also has an additional capacitor for tunnelling. The two capacitors and the transistor share the same poly, which is not connected to any lead, thus forming an FG. The tunnelling capacitor is smaller in size than the control gate capacitor.

The FG is programmed by applying a voltage to the control terminal, which is coupled to the gate of the transistor through the MOSFET capacitor. CHE created at the drain are then attracted by the gate potential and injected into it. The erasing is done by applying erasing

voltage at both the source and drain terminals, while the control gate is grounded. The stored electrons are then extracted from the gate by tunnelling.

The relationship between FG, CG and charge voltage is given by

$$\begin{aligned} V_{FG} &= \frac{C_C}{C_C + C_D} V_{CG} + \frac{Q_{FG}}{C_C + C_D} \\ &= \alpha_C \left(V_{CG} + \frac{Q_{FG}}{C_C} \right) \end{aligned}, \tag{2.27}$$

where α_C is the coupling ratio, which determines the performance of the cell. The coupling ratio has a strong influence on the FG voltage for values between ¼ and 4.

The synapse transistor circuit's micro-model is given in Figure 2.7 [48].

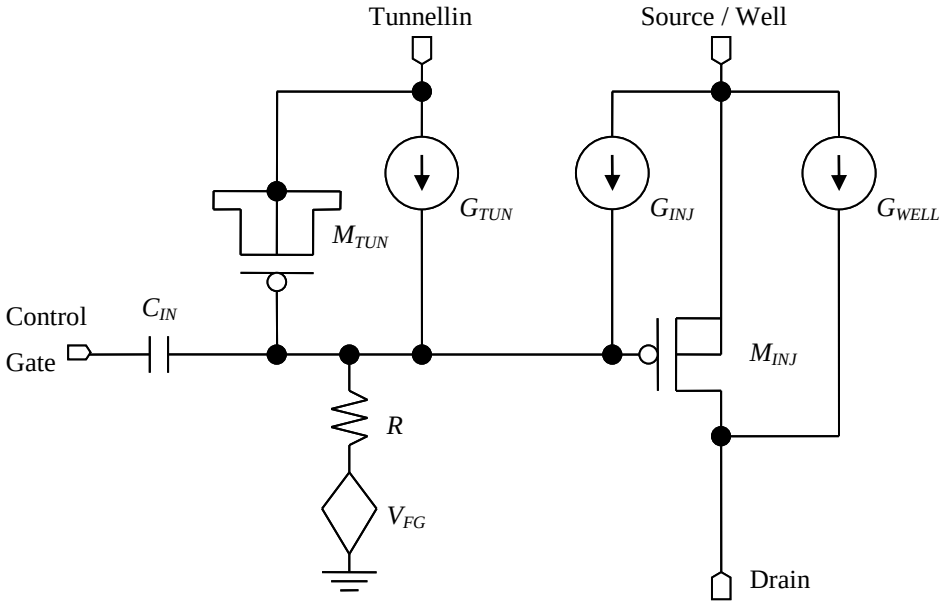


Figure 2.7. FG transistor micro-model [48] © 2002 IEEE.

The resistor and dependent voltage source in Figure 2.7 are included for DC convergence since the FG has no direct path to ground. The dependent source follows the FG voltage such that the voltage across the resistor is zero. Thus, the resistor value is chosen to be as large as possible and it does not affect the charge on the FG. The tunnelling current of the FG depends

on the oxide voltage and can be evaluated as presented in [45]. Another alternative model for simulation is shown in the Figure 2.8, where the resistor and the dependent voltage source are replaced by an amplifier [48].

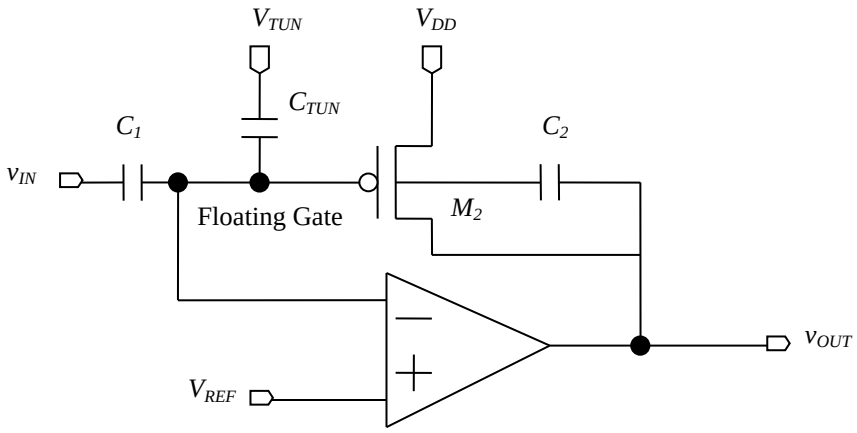


Figure 2.8. FG transistor with reference voltage used to establish initial FG voltage [48] © 2002 IEEE.

As can be seen from Figure 2.8, V_{REF} still follows the FG voltage and the input resistance is still as large as possible (ideally, infinite for this amplifier).

A voltage drop across the oxide effectively thins the barrier and exponentially increases the probability that a low-voltage electron will traverse it. When no voltage is applied to the oxide, the oxide thickness must be extremely small for any substantial amount of current to tunnel through.

The CHE injection depends on the source current and drain-to-channel potential; however, it is not explicitly available for simulation purposes. A first-order model for sub- and above-threshold, which does not depend on the drain-to-channel voltage, is presented in [45]. This model predicts injection current and is bias-dependent, since there is no bulk reference. The model is thus modified to include the bulk voltage to overcome both problems [46].

The injection model can be simulated as a dependent current source with a schematic as shown in Figure 2.9 [49].

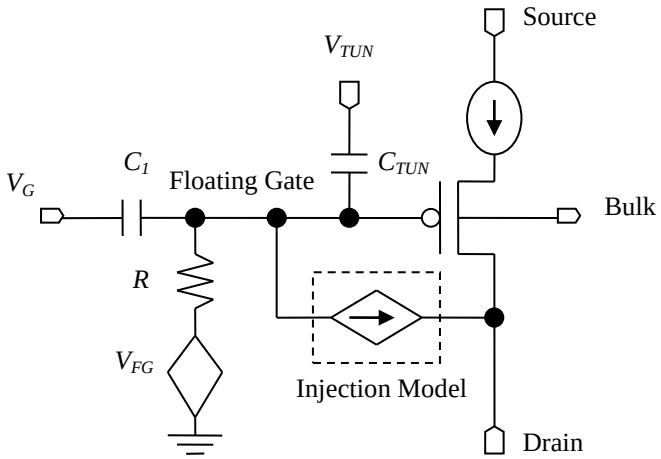


Figure 2.9. FG transistor with injection model included for simulations [49] © 2008 IEEE.

As in the previous case in Figure 2.8, the dependent voltage source and resistor in Figure 2.9 are included for DC convergence. The circuit is similar to the model presented in Figure 2.7, but it excludes the tunnelling and *WELL* modelling; a biasing or reference current is also applied at the source. The injection model is presented in [49].

2.4.2 Programming methods

2.4.2.1 Fowler-Nordheim tunnelling

The general equation for FN tunnelling current density is given by [36]

$$J_{tunnel} = \alpha \cdot E^2 \cdot e^{-\frac{\beta}{E}}, \tag{2.28}$$

where E is the electric field and α and β are FN parameters.

Oxide thickness is the primary factor that determines the accuracy of the parameter,

$$B = B_0\left(\frac{1}{1+t}\right), \tag{2.29}$$

where t is the uncertainty of oxide thickness [36]. The FN current is found by multiplying J with S, the tunnel area. The electric field is given by

$$E = \frac{V_{ox}}{t_{ox}}, \tag{2.30}$$

where V_{ox} is the voltage across the oxide and t_{ox}, the oxide thickness is

$$t_{ox} = \frac{\varepsilon_o \varepsilon_{ox} S}{C_{\max}}. \tag{2.31}$$

As stated earlier, the parasitic capacitors and resistors are not scaling desirably with the technology. Thus, S will scale more than $C_{\max}$, hence the overall oxide thickness will also scale down. The scaling of voltage is also not scaling as desired, owing to thermal voltage, thus the oxide voltage will scale less than the oxide thickness. It is based on this that the electric field keeps on increasing with technology and a constant field is not maintained. The effect is the induced current leakages.

2.4.2.2 Channel hot electron injection

The CHE injection (CHEI) current density is given by [37]

$$J_G = \int_{xt}^{\infty} J_p(x) \cdot M(x) \cdot P(x) \frac{dx}{W} \quad , \tag{2.32}$$

where $J_p(x)$ is the hole current position, $M(x)$ is the probability to generate an electron-hole pair, $P(x)$ is the probability of an electron at x to cross over the SiO_2 barrier and W is the width of the depletion region.

An electron injection mechanism can also exploit the band-to-band tunnelling current; this mechanism is known as the band-to-band hot electron injection (BBHE) [18]. Since most carriers generated by the band-to-band tunnelling process proceeds without forming an inversion layer, band-to-band tunnelling current can be estimated by the calculation of electric field distribution and the tunnelling probability. The injection current is then given by

$$J_{\text{injected}} = P \cdot P_{\text{prob\_tunnelling}} \cdot J(E), \tag{2.33}$$

where P is the scattering factor. Based on the depletion approximation and the assumption of uniform impurity distribution, the electric field E is in the depletion region.

A summarised comparison of CHEI and BBHE is presented in Table 2.4.

Table 2.4. Comparison of BBHE injection and CHEI as a programming method for stacked-gate devices. Republished, with permission of Taylor and Francis Group, LLC, from [18]; permission conveyed through Copyright Clearance Centre, Inc.

Quantity	BBHE	CHEI
Power consumption	Lower	Higher
Injection efficiency	Higher	Lower
Programming speed	Faster	Slower
Electron injection window	Wider	Narrower
Oxide field	Higher	Lower

From Table 2.4, it can be noted that BBHE offers better performance with power consumption, programming speed and injection efficiency. Because of a lower oxide field, CHEI has a thicker oxide thickness and less current leakage than BBHE. A further comparison between FN tunnelling and CHEI is given in Table 2.5.

Table 2.5. Comparison of FN Tunnelling and CHEI as a programming method for stacked-gate devices. Republished, with permission of Taylor and Francis Group, LLC, from [18]; permission conveyed through Copyright Clearance Centre, Inc.

Quantity	FN tunnelling	CHEI
Power consumption	Low	High
Oxide field	High	Low
Programming speed	Slower	Faster

The high oxide field of FN tunnelling, as shown in Table 2.5, causes it to have more current leakage. The CHEI has a low oxide field and thus suffers less velocity saturation, but because of its thicker oxide, it requires a higher programming voltage than FN tunnelling. Although it

is faster, the power consumption poses more limitations to low-voltage applications. Hence, FN tunnelling is the preferred method.

2.5 Charge pumps

For low power supply devices, e.g. new technologies and mm-wave systems, the programming voltage is higher than the voltage supplied. Thus, CP circuits are used to generate higher programming voltage [13]. The CP circuits must be designed to minimise the current drawn from the supply. This is important since the current drawn by the CP can be several times the current used by the cell. It is based on this that most power can be consumed by the CP circuit. The CP circuit is designed with *N*-stages, which can be controlled by either switches (Figure 2.10) or diodes (Figure 2.11) [38].

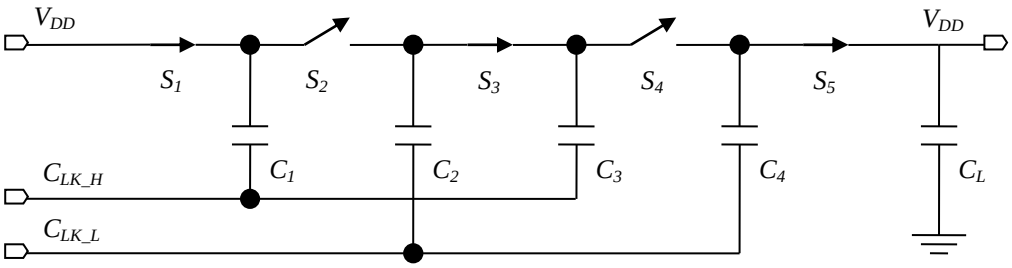


Figure 2.10. Charge pump with ideal switches [38] © 2002 IEEE.

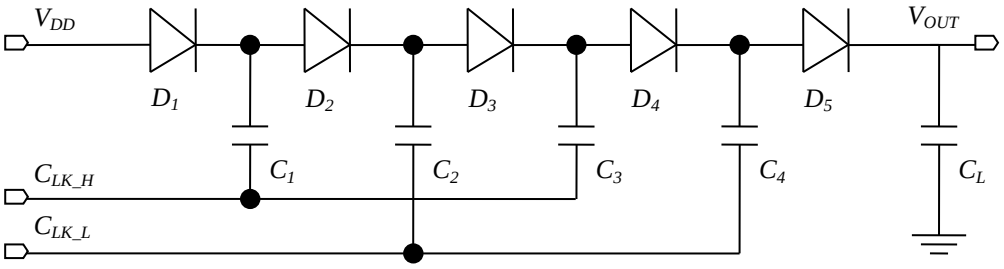


Figure 2.11. Charge pump with diodes [38] © 2002 IEEE.

The diodes in Figure 2.11 must have reduced turn-on voltage and the switches in Figure 2.10 must have good isolation and insertion loss. The structure of the circuit that has been used most often is the Dickson CP [40], as shown in Figure 2.12, but other structures, which can

offer other advantages, have been proposed in the literature [38]. The Dickson CP and voltage doubler have been investigated in this research. The hypothesis was formulated and an IC prototype was fabricated for testing and verification.

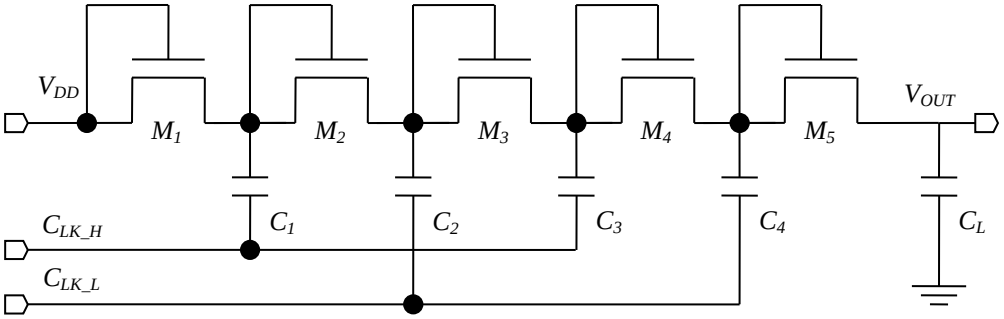


Figure 2.12. A four-stage Dickson charge pump [40] © 1998 IEEE.

The Dickson CP in Figure 2.12 has been expanded to offer other advantages, as proposed in the literature [38]. The Dickson CP and voltage doubler have been investigated in this research. The hypothesis was formulated and an IC prototype was fabricated for testing and verification. The multiple stages or pumping units of CPs are composed of a one-way switch and a capacitor. No two consecutive stages operate at the same time; the previous stage and the next stage serve as isolation switches. Thus, a two-phase clocking signal is required to proceed with the charge pumping operation. Device scaling or mm-wave applications also present challenges for clock distribution. Other alternatives for clock distribution are considered, including optical distribution [41].

The switches are designed to prevent the charge from flowing back to the input. The conceptual structure of CPs is shown in Figure 2.13.

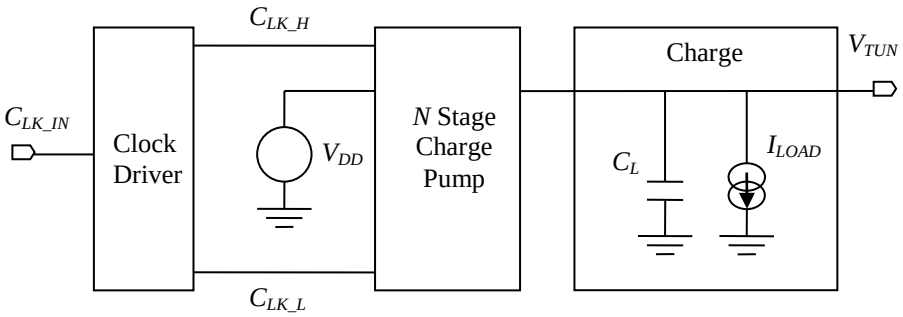


Figure 2.13. Conceptual structure of charge pump circuits [42] © 2007 IEEE.

The input clock in Figure 2.13 is generated from an oscillator and serves as the input to the clock driver, which further generates two non-overlapping clock signals. The CP stages are driven by the clock signals and the input is the supply voltage, which is limited by the technology node used (IBM 8HP). The number of stages is further limited by the increase in threshold voltage due to the body effect. Different structures have been proposed in the literature and the design considerations are given in the next subsections.

2.5.1 Two-phase non-overlapping clock signals

The problems of pass transistors for NMOS and PMOS are shown in Figure 2.14.

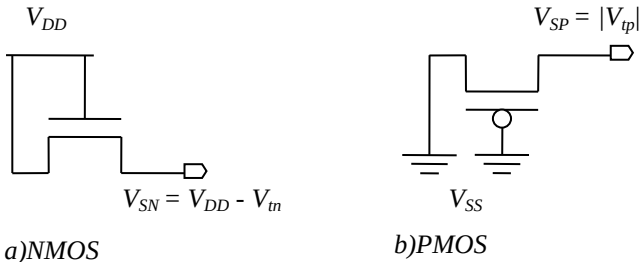


Figure 2.14. Voltage limitations for pass transistors.

In Figure 2.14 the source voltage, V_{SN}, for the NMOS transistor cannot be higher than V_{DD}-V_{tn} and thus its high voltage is limited by the threshold voltage. The lower voltage can be zero. A higher source voltage, V_{SP}, of V_{DD} can be achieved for the PMOS transistor, but its lower voltage is limited to V_{tp} by the threshold voltage. An inverter uses both a PMOS and NMOS transistor for maximum and minimum voltages, respectively. This ensures that the output voltage can reach maximum V_{DD} and a minimum of 0 V. An inverter is shown in Figure 2.15 [32].

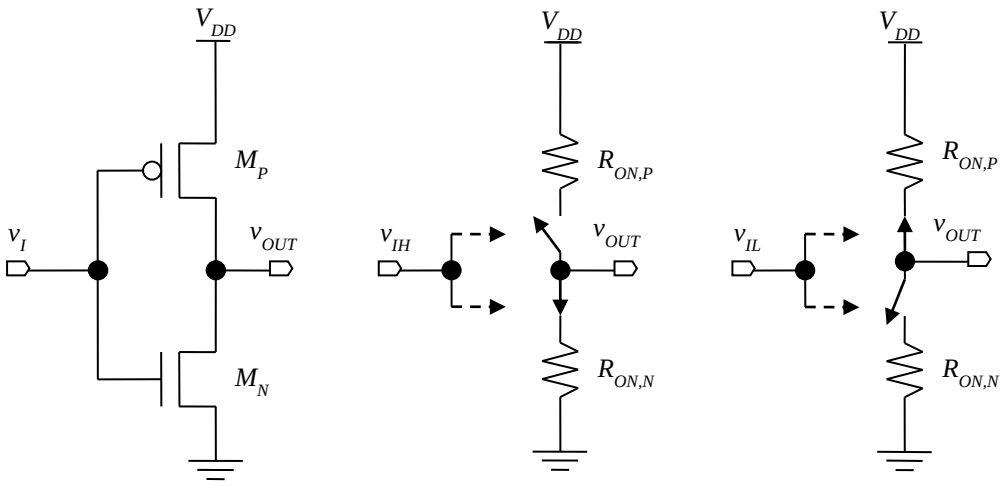


Figure 2.15. Inverter and its resistor modelling [32]. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, New Jersey.

As shown in Figure 2.15, when v_1 is equal to V_{DD} the PMOS is off and the NMOS is on and v_{OUT} is pulled down to V_{SS}. When v_1 is equal to zero, NMOS is off and PMOS is on and v_{OUT} is pulled up to VDD. The voltage transfer characteristics and operation regions are shown in Figure 2.16.

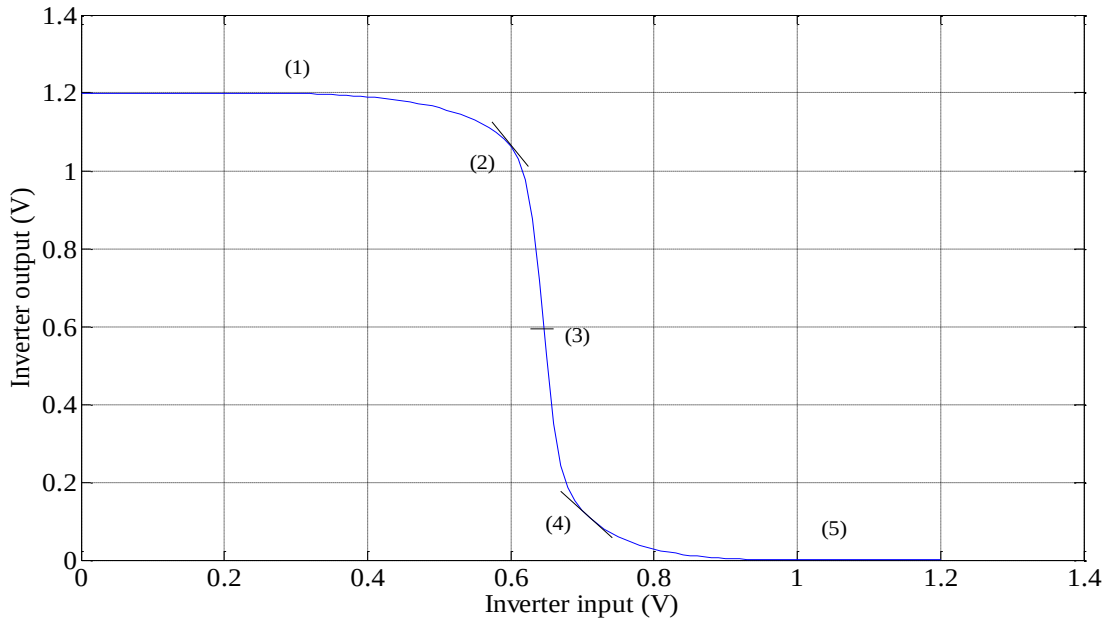


Figure 2.16. Inverter's transfer characteristic.

The regions of operation indicated by circled numbers in Figure 2.16 are:

(1) $v_0 = V_{0H}$, $v_1 < V_{tn}$

(2) $|V_{DS}| \geq |V_{GS}-V_{tp}|$

(3) M_N and M_P are saturated

(4) $V_{DS} \geq V_{GS}-V_{tn}$

(5) $v_0 = V_{0L}$, $v_1 > V_{DD} - |V_{tp}|$.

The propagation delay estimate of an inverter is given by

$$\tau_{PHL} = R_{ON} C \left\{ \ln\left(4\left[\frac{V_{OH} - V_{tn}}{V_{OH} + V_{OL}} \right] - 1 \right) + \frac{2V_{tn}}{V_{OH} - V_{tn}} \right\}, \tag{2.34}$$

where

$$R_{ON} = \frac{1}{(V_{OH} - V_{tn})k_n}. \tag{2.35}$$

For a symmetric inverter,

$$\tau_{PLH} = \tau_{PHL}. \tag{2.36}$$

Thus, the total delay is given by

$$\tau_p = \frac{\tau_{PLH} + \tau_{PHL}}{2} = \tau_{PHL}. \tag{2.37}$$

The rise and the fall times of an inverter can be expressed as,

$$t_f = \tau_{PHL}, \tag{2.38}$$

$$t_r = \tau_{PLH}. \tag{2.39}$$

2.5.1.1 CMOS NOR gate

For a CMOS NOR gate, transistors are sized such that delay time is the same as for the reference inverter. To accomplish this, the on-resistance on the PMOS branch of the NOR must be the same as that for an inverter. For a two-input NOR gate, the $(W/L)_p$ must be twice as large as that of an inverter.

The NOR gate has a body effect since the bottom PMOS bulk is not connected to its source, hence its threshold voltage changes as V_{SB} changes during switching. This effect disappears as V_0 approaches $V_{H;}$ however, the rise time is slowed down as $|V_{TP}|$ is a function of time.

2.5.1.2 Two-phase clock generator circuit

The non-overlapping clock generator consists of an inverter and two NOR gates. These gates have previously been discussed. A circuit for the clock generator is given in Figure 2.17.

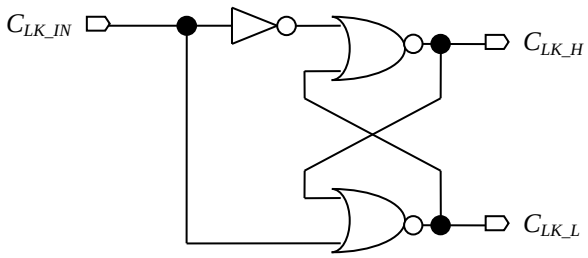


Figure 2.17. Two-phase clock generator circuit.

Delay, power and other parameters are calculated and estimated based on the equations derived for the inverter and the NOR gate. In Figure 2.17, C_{LK_H} and C_{LK_L} are the non-overlapping clock signals generated to drive the CPs.

2.5.2 Conventional Dickson charge pump

A four-stage Dickson CP is shown in Figure 2.18 [40].

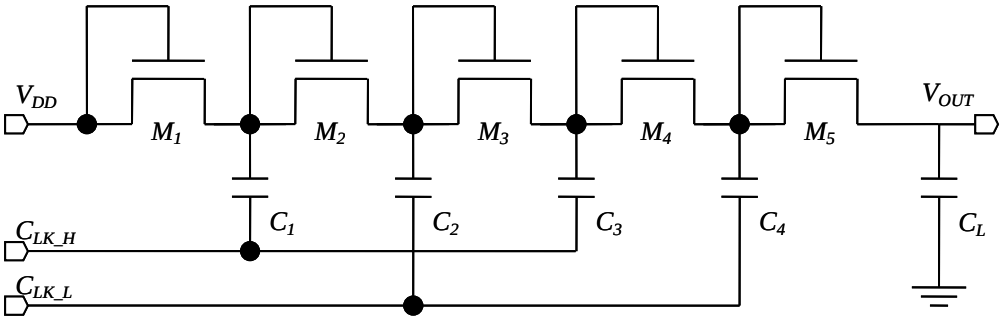


Figure 2.18. Dickson charge pump [40] © 1998 IEEE.

Diode-connected NMOS transistors in Figure 2.18 are used to transfer charge. The circuit consists of capacitor stages connected by MOSFET diode transistors and coupled in parallel with two-phase non-overlapping clock signals. The fluctuating voltage per stage is given by [40]

$$\Delta V = V_{CLK} \cdot \frac{C}{C + C_s} - \frac{I_0}{f(C + C_s)}, \tag{2.40}$$

where V_{CLK} is the amplitude of the voltage clock driver, C is the pumping capacitor, C_s is the parasitic capacitor at each node, I_0 is output current and f is clock frequency. For small values of C_s and I_0, fluctuating voltage is approximately equal to supply voltage. The N^{th} stage gain is given by

$$\begin{aligned} G_V &= V_N - V_{N-1} \\ &= \Delta V - V_{th}(N) \end{aligned}. \tag{2.41}$$

Since the bulk of NMOS is connected to ground, as the number of stages increase, the threshold voltage also increases in reaction to the body effect. The output voltage is given by

$$v_{OUT} = \sum_{N=1}^{N+1} \left(V_{DD} - V_{th}(N)\right). \tag{2.42}$$

The number of stages is limited by the increase in threshold voltage; maximum threshold voltage is limited to supply voltage. The performance of the Dickson CP is limited by the

threshold voltage and reverse charge sharing. The reverse charge problem comes from the fact that the NMOS switches cannot be completely turned off.

2.5.3 Static CTS charge pump

The static CTS CP, shown in Figure 2.19 [40], is used to overcome the threshold voltage problem. It uses an additional auxiliary switch chain, which is parallel to the main MOSFET diode switch chain.

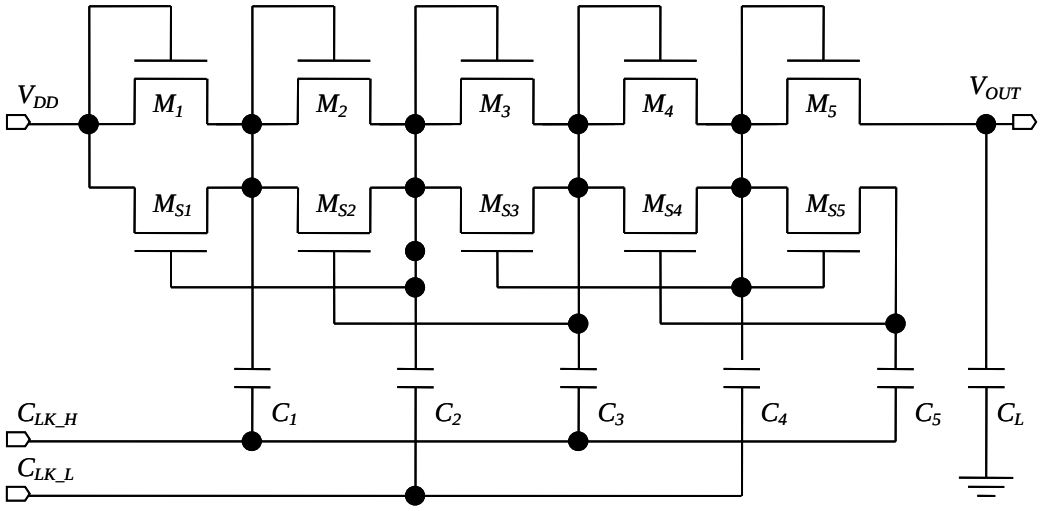


Figure 2.19. Static CTS charge pump [40] © 1998 IEEE.

The operation of CTS CP in Figure 2.19 is similar to the Dickson CP: the gate voltage of the auxiliary switch is derived from the next stage. Thus the gate voltage is high enough to keep the turn-on state of the switch, which results in reverse charge sharing.

A novel CTS control scheme [40] has also been considered. The NMOS CTS is backward-controlled by deriving high voltage from the following stage and the PMOS CTS is forward-controlled by deriving its low voltage from the previous stage. Some of the advantages are that NMOS CTS has no substrate current and PMOS CTS is able to eliminate the body effect.

2.5.4 NCP-2 charge pump

NCP-2, Figure 2.20 [40], uses dynamic CTS controlled by pass transistors; it avoids the threshold voltage problem and has no reverse charge problem.

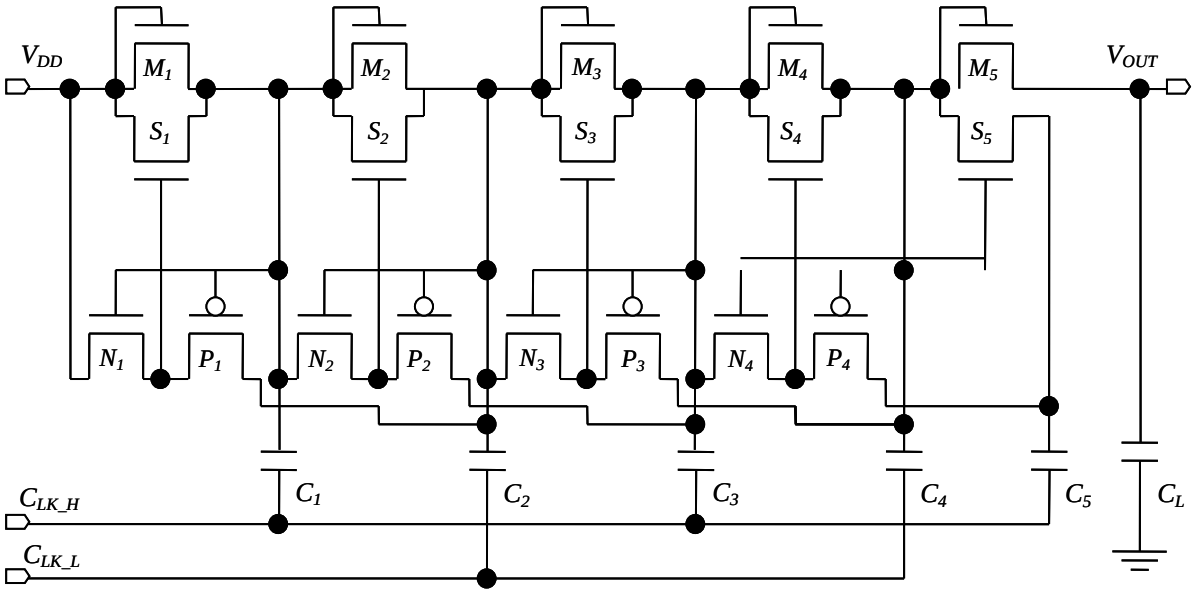


Figure 2.20. NCP-2 charge pump [40] © 1998 IEEE.

The output stage (in Figure 2.20) uses NMOS transistors and this leads to reduced output voltage due to body effect. An improved structure of NCP-2 has been proposed in [46] with an improved output stage, Figure 2.21.

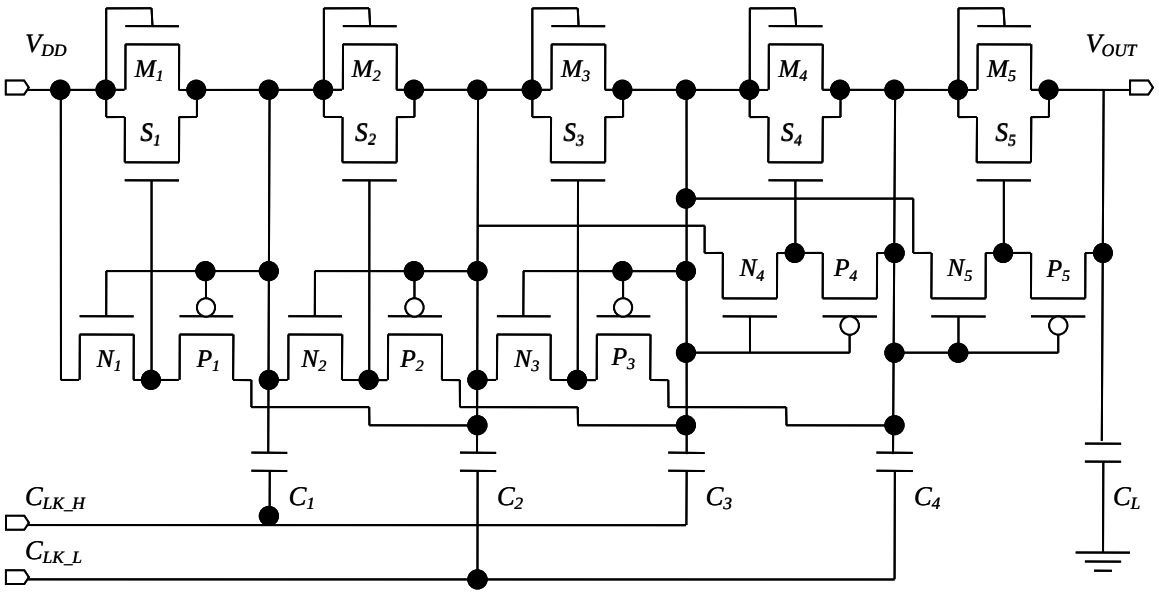


Figure 2.21. Improved NCP-2 charge pump [46] © 2009 IEEE.

The NMOS switch is controlled with backward control by using the following stage, while the PMOS transistor is controlled with forward control from a previous stage. The structure in

Figure 2.21 avoids the large forward substrate current to the *n*-well compared to all PMOS switches. It does not require clock-enhanced voltage amplitude like NCP-2.

2.5.5 Voltage doubler

The voltage doubler consists of a pair of storage capacitors, NMOS switches to charge the storage capacitors and PMOS transistors, which are used as transfer elements. Figure 2.22 [50] shows the schematic for the voltage doubler.

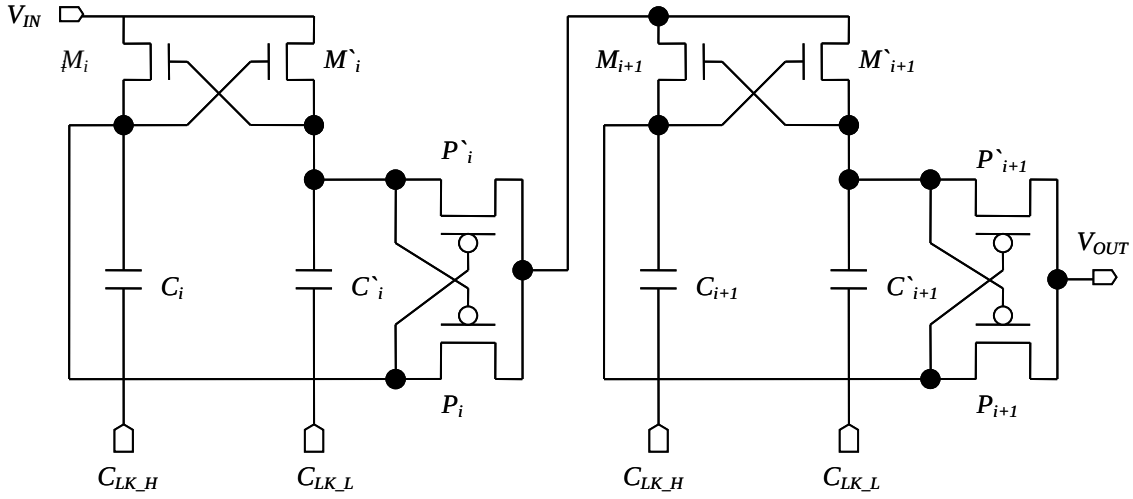


Figure 2.22. Voltage doubler charge pump [50] © 2007 IEEE.

In Figure 2.22, the voltage drop across any element is less than V_{DD}. To achieve charge transfer from the storage element $C_{,i}$ of the i^{th} stage to the storage element $C'_{,i+1}$, P_i and M'_{i+1} are activated. The structure has reduced on-resistance.

2.5.6 Power efficiency analysis

Efficiency is the measure of the ratio between the power delivered at the output and the power supplied to the CP. Power efficiency is given as

$$n = \frac{p_{OUT}}{p_{OUT} + p_{LOSS}}, \tag{2.43}$$

where p_{LOSS} is the total power loss and is calculated from the equation

$$p_{LOSS} = P_{RES} + p_{dyn} + P_{SC}, \tag{2.44}$$

where P_{RES} is due to a finite value of output resistance, R_{OUT}, of the pump. p_{dyn} is related to dynamic losses (which depend on frequency and capacitance), (2.54), and P_{SC} is the short circuit power consumption of the phase driver. The loss parameters are determined as follows

$$P_{RES} = R_{OUT} I_{OUT}^{\ 2}, \tag{2.45}$$

where I_{OUT} is the output current. The dynamic loss is

$$p_{dyn} = k_1 N f C_{par} V_{DD}, \tag{2.46}$$

where k_1 is a structure coefficient and is process-dependent (voltage swing across any parasitic capacitance is a fraction of V_{DD}) and C_{par} is the total parasitic capacitance [39]. The parameter C_{par} consists of the top plate parasitic capacitance and the capacitance of the MOSFET switches. For a Dickson CP, MOSFET switch capacitance is the sum of gate-bulk and gate-source capacitance of the pass transistor M_{i+1} and the source-bulk capacitor of the pass transistor M_i. For a voltage doubler, MOSFET switch capacitance is the sum of the gate capacitor of M'_i, the source-bulk capacitor of pre-charge M_i and gate and source-bulk capacitors of the PMOS pass transistors P'_i and P_i, respectively. The difference in the parasitic capacitance for both the voltage doubler and Dickson pump results in different voltage increments per stage. The short-circuit loss is given as

$$P_{SC} = k_2 N B_{stage}, \tag{2.47}$$

where k_2 is related to supply voltage, threshold voltage and the timing of the clock signal's rise or falling edge [39]. Since the inverters in the phase driver are symmetric, $B_P = B_N$ and the B_{stage} is given by

$$B_{stage} = \mu C_{ox} \left(\frac{W}{L} \right)_N. \tag{2.48}$$

The output power is determined by the output voltage and the maximum output current, i.e.

$$p_{OUT} = v_{OUT} i_{OUT(\max)} . \tag{2.49}$$

In general, the *N*-stage voltage elevator can be modelled with Thévenin's equivalent circuit; with an open circuit voltage source of

$$V_{Th\acute{e}venin} = (N+1)V_{DD} , \tag{2.50}$$

and a series output resistance of

$$R_{OUT} = \frac{N}{fC} . \tag{2.51}$$

Thus, ideal output voltage can be expressed as

$$v_{OUT} = (N+1)V_{DD} - R_{OUT} i_{OUT} . \tag{2.52}$$

The total parasitic capacitance can be expressed as

$$C_{par} = \alpha C , \tag{2.53}$$

where α depends on the kind of storage capacitor and on the topology of the CP. Because of parasitic capacitance, the output voltage and the effective output resistance are modified as follows

$$v_{OUT} = V_{DD} + \frac{1}{1+\alpha}\left[NV_{DD} - R_{OUT} i_{OUT}\right], \tag{2.54}$$

where α is the capacitive coupling ratio strongly influenced by the top plate of the capacitor. The effective resistance is

$$R_{OUT} = \frac{N}{f\left(C + C_{par}\right)} \tag{2.55}$$

and the maximum current is

$$\begin{aligned} i_{OUT(\max)} &= \frac{NV_{DD}}{R_{OUT}}, \\ &= fCV_{DD} \end{aligned} \tag{2.56}$$

where $i_{OUT(max)}$ is limited by the maximum charge transfer in one clock period and is not dependent on parasitic elements. The on-resistance of the cascaded voltage doubler connecting two capacitors of adjacent stages, C_i and C_{i+1}, results from the sum of the PMOS, P_i and NMOS, M'_{i+1} switches. The total resistance relates to on-resistance by

$$R = 2R_{on}, \tag{2.57}$$

with $R_{on,\,N} = R_{on,\,P}$. Thus,

$$C_{DK} = 2C_{DB}. \tag{2.58}$$

This is done in order to achieve equal time constants for both the voltage doubler and the Dickson CP.

2.5.6.1 Power consumption optimisation

The ideal output voltage can be expressed as

$$v_{OUT} = \left(N+1\right)\cdot V_{DD} - \frac{Ni_{OUT}}{fC}, \tag{2.59}$$

and the total current for power consumption taking into account the non-ideality is given by

$$i_{POWER} = \left(N+1\right)\cdot I_L + i_{ni}, \tag{2.60}$$

where i_{ni} is the non-ideality term. This is given by

$$\begin{aligned} i_{ni} &= NC_{par} fV_{DD} \\ &= \alpha NCfV_{DD} \end{aligned}, \tag{2.61}$$

where

$$\alpha = \alpha_B + \alpha_T . \tag{2.62}$$

From the ideal output voltage equation, the storage capacitor can be expressed as

$$C = \frac{Ni_{OUT}}{\left[(N+1)V_{DD} - v_{OUT}\right]f}, \tag{2.63}$$

thus,

$$i_{POWER} = \left\{(N+1) + \frac{\alpha N^2 C}{\left[(N+1)V_{DD} - v_{OUT}\right]}V_{DD}\right\} \cdot I_L . \tag{2.64}$$

The minimum power consumption is equivalent to the minimum current consumption, thus the current expression is differentiated and equated to zero. The resulting expression is then solved for N, the optimum number of stages for optimal power consumption,

$$\frac{di_{POWER}}{dN} = 0\Big|_{N_{opt}} . \tag{2.65}$$

This result gives the following number of stages for minimum power consumption,

$$N_{opt} = \left(1 + \sqrt{\frac{\alpha}{1+\alpha}}\right)(x-1), \tag{2.66}$$

where

$$x = \frac{V_{OUT}}{V_{DD}}. \tag{2.67}$$

The CP efficiency is given by

$$\begin{aligned} n_p &= \frac{i_{OUT} v_{OUT}}{i_{POWER} V_{DD}} \\ &= \frac{x}{N+1+\alpha N^2 / N+1-x} \end{aligned}. \tag{2.68}$$

From this it can be noted that maximum efficiency is achieved when an optimal number of stages are used.

2.6 SENSE AMPLIFIERS

The read operation is based on sense amplifiers, which can be implemented as either current- or voltage-mode sensors. Current-mode sensors have been a choice for most applications with conventional sense amplifiers [12] (Figure 2.23) having the disadvantage of large reference current and a small sense-line swing.

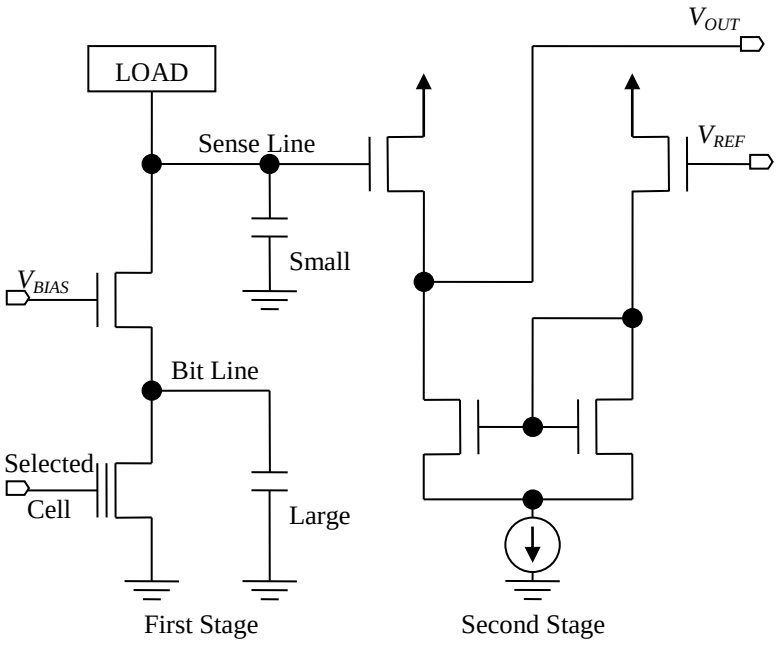


Figure 2.23. Conventional sense-amplifier circuit [12] © 2000 IEEE.

In minimising this problem associated with a conventional sense-amplifier (Figure 2.23), bit line direct sensing circuits have been proposed, as these circuits offer better sense-line swing for low supply voltage operations [12]. An advantage of current-mode is that the sensing of memory cells does not require a large voltage swing. A voltage mode is used to overcome degeneration of the FG transistor, which makes the reference current fluctuate with the average readout currents [43]. The sensing speed of the voltage mode is limited by high-speed large-size memories, where there is very large bit-line capacitance.

Fast pre-charging independent of the capacitive load, represented by a bit-line array, can be achieved by adapting cascading techniques using an inverter with a source follower output stage and unity gain feedback. Slew rate limits the speed to pre-charge a bit-line.

Since sense amplifiers determine the access time, fast sensing speed is required. Thus, a sense amplifier must be optimised for speeding up the sensing process. The second-stage sense amplifier can be pre-charged and prepared, while the first-stage sense amplifier is amplifying the signal [18]. As a result, sensing overhead can be reduced.

2.7 CHARACTERISATION

2.7.1 Reliability

The measure of reliability is determined based on retention and endurance tests. Endurance is a measure of the cell's performance based on the number of cycles for programming or erasing the memory cell. The stored information must be maintained over time; retention is a measure of this. A parameter that is also important for reliability concerns is the oxide thickness and its controllability to avoid defects. The oxide thickness is therefore chosen as a trade- off between the performance and reliability concerns. For example, a thin oxide layer poses problems, because it is close to the break-down between the substrate and source junction. In addressing this problem, an asymmetric structure is implemented between the source-substrate and the drain-substrate junction. The source is more doped than the drain, thus the diffusion for the source and the drain will be different [33]. The dependency of the tunnel current on the oxide-electric field also causes process control problems.

Endurance and retention are used to determine the effects and causes of charge gain/loss, which may change voltage thresholds. As the cell is programmed or erased, some charges are trapped or more charges can be drawn from the FG, thus this leads to threshold changes and also affects the durability of the cells.

2.7.2 Insertion and isolation losses

MEMS are known for their good insulation and isolation losses when compared to either EEPROMs or e-fuses. Even though this is the case, there are other reliability issues, which make MEMS unfavourable. The challenges of MEMS development are due to the multitude of failure mechanisms of such devices, i.e. cracking, creep, fatigue, deposition and growth on the contact surfaces, and static friction [44]. Other improvements that are still under investigation are fast switching and low actuating voltage.

Switches can be implemented as capacitive or resistive switches, but capacitive switches are mostly used in high-frequency applications and they avoid large resistive losses, which contribute to high power consumption. These are the same reasons diode switches are not preferred: high power consumption and large resistive losses. Diodes are non-linear, but have

the advantage of switching speed.

Since the switches are capacitive, the insertion and isolation loss can be characterised by using either current or capacitive ratios in the ON and OFF states. The current of an OFF state can be in the range of 10 nA and this may be a concern to some measuring instrument applications. This value may be at threshold or below minimum sensible current and thus it may pose challenges during the testing process. The ON state current is in the range of 100 µA. The operation of the memory cell is based on an FG. The cell is written/erased by either injecting or removing a charge from the FG. Thus, the current flow is controlled by the amount of charge present in the FG. The capacitance at the drain and source is also changed by the amount of charge. The capacitive ratio is thus different for the ON and OFF state and it can be used to determine the state of the switch. The isolation and insertion loss can therefore be measured using either current or capacitive ratios. An advantage of the FG is that it can be used at the extreme ends for an ON/OFF switch or it can be used for other characteristics, which operate between the extreme states by controlling the charge in the FG [15].

2.7.3 Performance

The memory cell can be implemented by using either an *n*- or *p*-channel. The *n*-channel has higher electron mobility and it suffers less drain disturbance than the *p*-channel, but it has the disadvantages of current degradation and closure occurring with increasing cycles. The *p*-channel achieves high speed, better reliability and lower power; it is fast in programming but slow in erasing [46]. Injection efficiency is one of the performance parameters which measures the ratio of the electrons collected at the FG to those generated in the channel. The *p*-channel offers better performance over the *n*-channel and it also has high injection efficiency.

Table 2.6 provides a comparison of MEMS switches against PIN and metal semiconductor FET (MOSFET) diode switches [46]. MEMS offer a number of advantages, such as small series resistance, low loss, good isolation and small size, but, they have a slow switching speed and may require a high control voltage. PIN diodes are also slow and require a high control current. MOSFET have high series resistance, higher insertion loss and their isolation can be lower by 20 dB. The advantages of MOSFET are low control current and fast switching speed. EEPROMs have the advantages of small size, fast switching and lower control

currents. Control voltage and series resistance are expected to be well comparable to those of MOSFET.

Table 2.6. RF MEMS vs. PIN and MOSFET Switch Comparison [46]. Reprinted with permission.

Quantity	MOSFET	PIN Diode	MEMS
Series resistance (Ω)	3 – 5	1	< 1
Loss at 1 GHz (dB)	0.5 – 1	0.5 – 1	0.1
Isolation at 1 GHz (dB)	20 – 40	40	> 40
IP3 (dBm)	40 – 60	30 – 45	> 66
Size (mm^2)	1 – 5	0.1	< 0.1
Switching speed	~ns	~µs	~µs
Control voltage (V)	8	3 – 5	3 – 30
Control current	< 10 µA	10 mA	10 µA

The parasitics of switches have an influence on the performance of the switch. Parasitic capacitance, resistance and transistor leakage currents [15] are the factors that affect the switches to function non-ideally. The parasitics are due to high technology scaling and high operating frequencies, thus parasitic effects for mm-wave application have more influence on the performance of the switches. For a short circuit, insertion loss is due to capacitive and resistive loss along the path, while for an open circuit, current leakages flowing through the switch contribute to isolation losses. Unlike low-frequency applications, for mm-wave applications the capacitive parasitics have a critical influence, which cannot be ignored. All factors that contribute to either insertion or isolation loss of EEPROM switches need to be determined and evaluated for mm-wave applications.

2.8 CONCLUSION

This chapter has highlighted some critical issues of technology scaling and influence of mm-wave frequencies for FG transistors. Device models, programming methods, peripheral circuits, packaging and other parameters were discussed.

Today, parasitic capacitances, parasitic resistances and threshold voltages are no longer scaling in proportion with technology scaling. This has the effect that the oxide field is increasing with technology and induces current leakages. Design approaches have to optimise the performance and minimise the effects introduced by scaling. Scaling is desired to achieve higher frequencies (mm-wave) with CMOS. Such operating frequencies are required to meet an ever-increasing demand for bandwidth.

Devices operating in mm-wave are highly sensitive to device or parameter variations. Thus, packaging is now playing a critical part in design and has to be done concurrently with the overall system design. Signal routing interconnections are more comparable with the operating system wavelength and thus add other influences such as voltage drop and delay, which limit system performance. The voltage drop is becoming more of an issue owing to low supply voltages that are decreasing with technology scaling.

The programming methods also need careful consideration. The electrons have to be injected to cross the oxide into or out of an FG. Oxide thickness is affected by technology scaling and requires that the programming methods be improved and chosen based on the technology trends and trade-off constraints. Peripheral circuits are also used to optimise the overall system performance. Hence, the peripheral circuit designs and performance need to be evaluated for mm-wave applications.

There are different methods by which tuning can be achieved and these are not limited to after-package trimming. RF circuits and matching networks are other examples where tuning is desired. The tuning circuits can be implemented by using either e-fuses, MEMS or EEPROMs. A full application of these devices for mm-wave is still under development and the main determining factor will be how these devices will perform in the mm-wave spectrum.

Different tuning methods have been qualitatively compared based on performance parameters. EEPROM switches are proposed for tuning, since they offer the advantages of speed, size and low power consumption. These are the most important parameters for mm-wave applications. Although insertion and isolation losses will be poor compared to MEMS, EEPROM switches can still achieve acceptable values for normal or desired operations.

CHAPTER 3: METHODOLOGY

3.1 INTRODUCTION

The research methodology followed in this research is described in this chapter. The method highlights the research hypothesis, mathematical modelling, computer-aided designs (CAD) and the IC fabrication process. Comparison from modelling to the final IC prototype has been used to validate the hypothesis and to reach conclusions for the research outcomes. Thus, mathematical modelling, CAD software tools and the measurement setup of the final prototype are described in this section.

3.2 RESEARCH METHODOLOGY

A typical research methodology is illustrated in the flow diagram as shown in Figure 3.1.

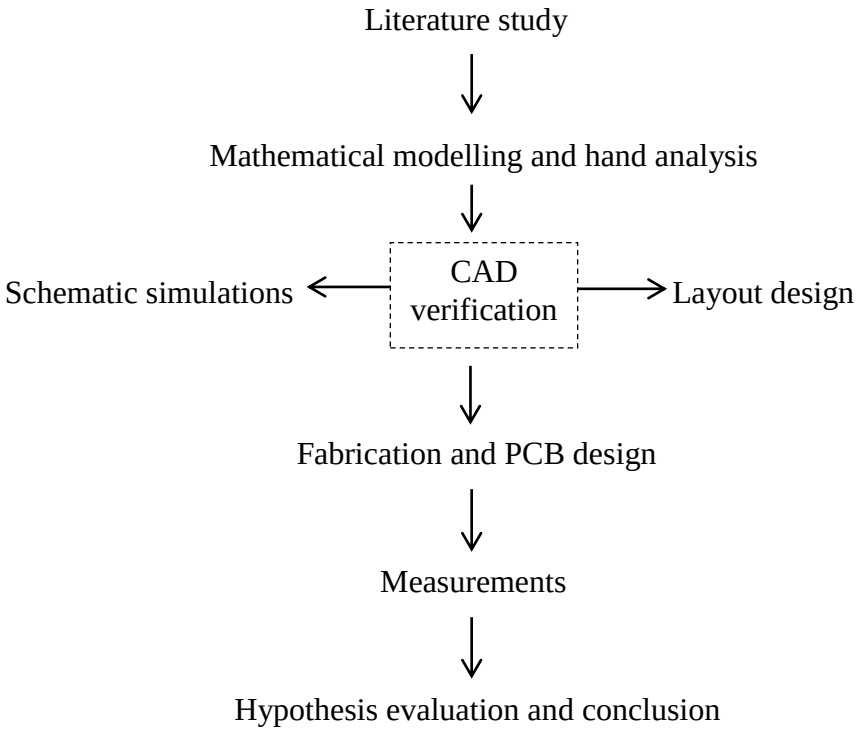


Figure 3.1. Research methodology.

The hypothesis testing was done iteratively using the research methodology in Figure 3.1. The process started with a literature study, as presented in Chapter 2. The fundamental theories, new and old concepts were studied in relation to the challenges regarding this research work. As part of an observation study, experiments to acquire more insight were conducted at GT for FG devices.

Mathematical modelling and hand analysis were sufficient for first-level testing; different models were compared and evaluated. Design verification was done in two steps using CAD tools. Firstly, schematic simulations were performed and compared to the hand or mathematical results. When definitive correlation was achieved, the second step was the layout design and simulations. The simulation software was used to compare the schematic and layout designs, by using the layout versus schematic (LVS) simulation feature. The literature review, simulations and mathematical model provided predictions for the expected results of the prototypes.

After satisfactory results had been achieved and the final layout design had been completed, the design was sent for fabrication in order to perform prototype testing and verification at a practical level. Practical measurements are provided in chapter 5. The designed hypothesis was then evaluated and validated based on the predictions, simulations and measured results; the IC prototyping was fabricated to enable characterisation of the hypothesis.

The use of FG transistors as switches for mm-wave applications forms the basis of the hypothesis for this dissertation. FG transistors are tested and characterised for both the ON and OFF state by determining the insertion and isolation losses, respectively. Other device performances (size, power efficiency, etc.) are also important, but insertion and isolation losses are the main characteristics for switches. Thus, the hypothesis relates to testing the suitability of FG devices for mm-wave applications.

3.3 MATHEMATICAL MODELLING

Transistor modelling is based on assumptions and approximations to simplify models for analysis; as CMOS scales down, some assumptions and approximations no longer hold. Thus, complex models are used with more parameters taken into account. For hand calculations, it is almost impossible to analyse circuits with such complex models. Hence, simpler models are still used for hand calculations to obtain the first order approximations and to have a rough idea of the operational regions. MATLAB software is used to analyse complex equations, taking into account most of the parameters to achieve more accurate results.

MATLAB has mathematical functions that can be used to develop algorithms easily and iterative calculations are easily performed with 'for' and 'while' statements. This makes the finalisation of the conceptual design and the testing of different models much easier. MATLAB is also used to analyse experimental results; for FG measurements performed at GT, it was used to acquire data for graphical presentation.

3.4 SIMULATION (CAD) SOFTWARE

The CAD software that was used for schematic and layout simulations is Cadence. The Cadence Virtuoso software suit consists of packages as described in Table 3.1.

Table 3.1. Cadence Tools

Package Name	Functionality
Virtuoso schematic editor	Graphical user interface for circuit schematic design
Analogue/mixed-signal (AMS) 2007.1	SPICE and Spectre based simulator
Virtuoso layout editor	Graphical user interface for drawing circuit layouts
Assura	Design rule check (DRC) and LVS check

Mathematical modelling is verified by constructing the equivalent circuits using the schematic editor. The Cadence analogue design environment was used to perform different simulation analyses for the designed schematic, using the packages in Table 3.1. The analyses that were performed using the software are DC (operating point and sweep), transient and frequency domain analysis.

Layout design and verification were done after satisfactory results had been achieved through mathematical and schematic simulations. LVS was used to ensure that the layout corresponded with the constructed schematic and the DRC was used to ensure that the layout complied with the foundry specifications to ensure successful chip fabrication.

The performance parameters of a final design are dependent on the layout. Thus parameter extraction of the layout is done and such parameters are used to perform LVS. The LVS confirms the accuracy and mapping of the physical layout based on the schematic net listing.

3.5 MANUFACTURING PROCESS AND IC PROTOTYPING

The process that was used for simulation is an International Business Machines (IBM) 8HP process; a 130 nm SiGe BiCMOS technology node was used. All simulation models used use parameters that were extracted from a 130 nm technology node; the technology files were supplied by the foundry. The foundry also provided DRC files for all technology nodes. All files and models were included in the 8HP process design kit (PDK) and some process parameter values are not disclosed due to the NDA; only the details available in the public domain are provided here. The maximum power supply was limited by a specific technology node; for the 8HP process the maximum supply was 1.2 V. A summary of the technology process and parameters is given below.

Technology features:

- V_{DD} = 1.2 V
- Maximum power supply voltage; 1.6 V for 2.2 nm FETs and 2.7 V for 5.2 nm FETs
- Operating temperature range of -55 to 125 ^{o}C
- Dual gate oxide with physical thicknesses 2.2 nm and 5.2 nm
- Minimum lithographic image 0.12 µm (gate only)
- 5, 6 or 7 levels of global layers (2-4 levels of 1x copper; M1-M4, 1 level of 2x copper; MQ and 2 levels of analogue metal; LY, AM)
- Common wiring level vias V1, V2 and VL.

Supported base feature devices:

- Thin oxide surface channel NFET and PFET (L = 0.12 µm)
- Bondpads (C4 and wirebond; 2 variants each)
- PCDCAP/MOSFET varactor (for 1.5 V operation)

- *p*+ polysilicon resistor (2 placement variants)
- NS resistor.

A design verification tool, Assura, was used for testing both DRC and LVS. The following design checks were completed before layout submission for fabrication:

- Design rule checking (drc.rul)
- Floating – gate and antenna (float.rul)
- Pattern density (global.rul)
- Local pattern density (local.rul)
- Line mode and orthogonal check (invalid shapes and off-grid placement; design geometry).

Layout versus schematic involved comparing the CDL netlists for schematic design and the layout extracted parameters.

The 8HP technology process used for prototyping had five metal levels. The 8HP BiCMOS transistors are suitable for mm-wave applications, as the process transistor has a unity gain frequency of 200 GHz (nominal). Thus, it provides support for mm-wave application modules. The 8HP is a 0.13 µm process technology, has a higher switching speed for CMOS applications and operates with a 1.2 V power supply. For FG devices, the functionality is highly dependent on the oxide thickness for programming and the capability to retain charge in the FG device. The oxide thickness for 8HP is 3.2 nm and this poses a challenge: charge leakages through the thin oxide and charge trapping in the oxide will affect the performance. Different vias are used to connect the metal levels. The PDK included the contacts already developed and the library instances were used in the layout design. The transistor parameters used for design are given in Table 3.2.

Table 3.2. *N*- and *p*-channel transistor parameters.

Size	Parameters	*W* / *L*	*n*-channel	*p*-channel	Units
Minimum	V_{th}	0.16 / 0.12	0.43	0.44	Volts
	I_{ds}		427	-173	µA/um
Short	V_{th}	20.0 / 0.12	0.47	-0.43	Volts
	V_{pt}		3.6	-3.6	Volts
Wide	I_{ds0}	20.0 / 0.12	139.5	-120.1	pA/µm
	V_{th}		0.13	-0.24	Volts

Large	V_{jbkd}	20.0 / 20.0	2.6	-2.6	Volts
	I_{jlk}		< 50.0	< 50.0	pA
	Gamma		0.26	0.26	$V^{1/2}$
-	$K' (U_o * C_{ox}/2)$	-	309.8	-47.0	$\mu A/V^2$
-	Low-field mobility	-	574.19	87.11	$cm^2/V.s$
-	t_{OX}	-	3.2	3.2	nm

The designs in chapter 4 are based on the process parameters in Table 3.2 and parasitic parameters as given by the foundry. The parameters are specified for different channel lengths for both the *n*- and *p*-type transistors.

3.5.1 Transistors

The process used set some design constraints, which had to be adhered to. Some of the parameters required for design are determined by the layout structure used; these include device area, perimeter and also other parasitic parameters dependent on the layout structure. This section presents some of the key parameters, definitions and the typical transistor layout structure is given in Appendix B.

As specified by the foundry, IBM, the maximum supply voltage is 1.6 V. However, additional layout and design precautions need to be followed when using the maximum supply. The nominal or typical supply voltage used is 1.2 V for a thin oxide, which is used in this dissertation; thick oxides are used for slower devices and thus not suitable for mm-wave application. The oxide thickness is 2.2 nm, but the effective oxide thickness is given by

$$T_{OX\_eff} = \frac{\kappa_{OX}\varepsilon_0}{C_{IN}}, \tag{3.1}$$

and it ranges between 3.15 and 3.2 nm for both NFET and PFET. The width or the *RX* layer and the length of the gate, as determined from the poly-layer in the layout, are different from those used in calculations. For calculation purposes, the effective width and length are given by

$$L_{eff} = L_{des} - \Delta L, \tag{3.2}$$

$$W_{eff} = W_{des} - \Delta W \,, \tag{3.3}$$

where ΔL and ΔW are the fabrication variation as specified by the foundry (not disclosed here due to a NDA). The threshold voltage is another important parameter for the transistor and this has been characterised by the foundry and also varied based on the layout parameters. The threshold voltage for NFET devices satisfy the following expression

$$\left|I_{DS}\right| = k_1 \frac{W_{eff}}{L_{eff}} nA \,, \tag{3.4}$$

and for PFET devices, it must satisfy

$$\left|I_{DS}\right| = k_2 \frac{W_{eff}}{L_{eff}} nA \,. \tag{3.5}$$

Here k_1 and k_2 are process parameters. Thus, the threshold voltage can also be determined experimentally by varying the gate-to-source voltage until the above equations are satisfied.

3.5.2 Capacitors

An ncap (MOSFET capacitors) was used for the CP stages. The ncap capacitors are dependent on three parameters that characterise the layout: effective width and length, and the number of fingers (parallel devices). Other parameters are based on material characteristics and are provided by the foundry, i.e. C_L, C_W and C_F. The nominal capacitance depends on the *n*-well substrate junction voltage and is given by

$$C_{NOM}(V) = C_A(V) L_{eff} W_{eff} f + 2C_L L_{eff} f + 2C_W W_{eff} f + C_F f \,. \tag{3.6}$$

The parasitic capacitance depends on the area and perimeter of the n-well layer as it characterises the value of the capacitor. As in the case of the nominal capacitance, the parasitic capacitances also depend on the *n*-well to substrate voltage; the area and perimeter components are given below; measured in fF/μm^2 and fF/μm respectively.

$$C_A(V) = \frac{C_{A0}}{\left(1 - \frac{V}{pb}\right)^{ma}}, \tag{3.7}$$

$$C_P(V) = \frac{C_{P0}}{\left(1 - \frac{V}{php}\right)^{mp}}. \tag{3.8}$$

The process parameter values are specified by foundry, IBM, and not disclosed due to the NDA.

The MOSEFT capacitor layout design rules and characterisation, as done by the foundry, set the minimum constraints for the width and length. The process parameters are used to determine the NFET capacitors and the parasitic capacitors depending on the transistor dimensions.

3.6 MEASUREMENT SETUP AND EQUIPMENT

This section presents the experimental setup that was used in GT (using GT devices) and the setup that was used for the fabricated prototypes. The equipment used is also provided, together with the experimental procedures. The relationship between the experiments conducted and the hypothesis is also presented here.

3.6.1 For GT

The simulation setup(s) was set to test the functionality of the CPs. Parameter sweep was also done by varying the clock frequency, stage capacitors, load capacitor and device sizes. This was done to determine the sensitivity of parameter variation on device performance and also to optimise the designed parameters. FG transistors were not simulated in Cadence because of the direct current (DC) convergence problem and a secondary problem was based on selecting an initial voltage, which depends on the final layout and has a more profound effect on the final device performance evaluation. An algorithm was developed to simulate the FG devices using MATLAB software. The parameters that were simulated included the coupling ratio for

the tunnelling capacitor and control gate capacitor. The lengths and widths of the FG transistors were varied and tested for different values.

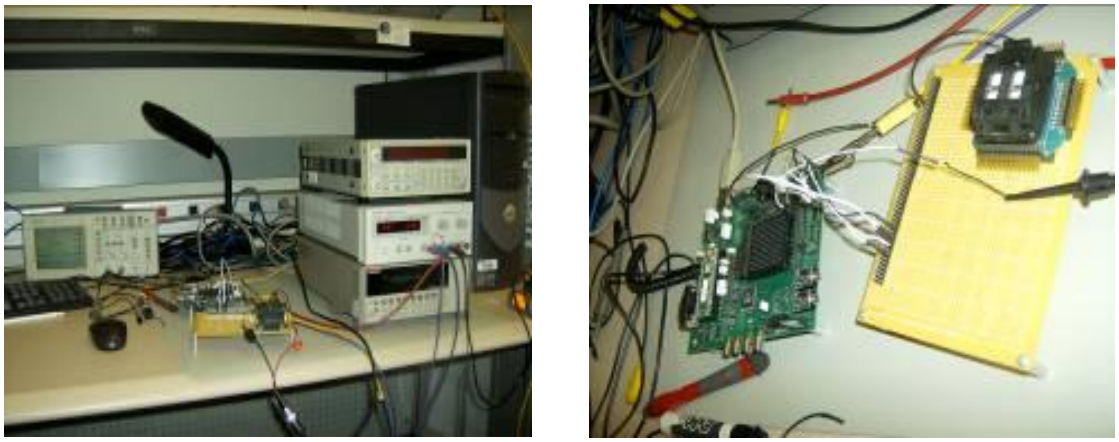
FG transistor prototypes developed at GT based on a 0.35 μm process were used for practical measurements and comparisons to the FG devices designed as a result of this research. The device dimensions for the prototypes that were used are provided in Table 3.3.

Table 3.3. Devices that were used for measurements

Parameter	Device 1	Device 2	Device 3	Device 4
Length (μm)	0.35	0.35	0.35	0.35
Width (μm)	2	2	10	10
C_{AREA} (μm^2)	1 × 1	10 × 10	1 × 1	10 × 10

The tested devices in Table 3.3 were developed by GT and these samples were made available for the purpose of the comparison referred to. The devices were used to validate the FG transistor hypothesis partially and this presented the first practical opportunity to do measurements on these devices. In Table 3.3, C_{AREA} is the area of the capacitor. Different capacitors were used and the effects of different coupling ratios were determined based on the experimental results.

The equipment and measurement setup for the prototypes that were used for measurements in GT are shown in Figure 3.2.



a) b)

Figure 3.2. a) shows some equipment used during measurements. The devices are interfaced using a PC card and the data acquisition accomplished using a MATLAB-based interface. b) shows the FG devices on a veroboard connected to a programmable device that is connected to the computer using the PC card.

The setup in Figure 3.2 was not modified; it was set up by the researchers at GT. An explanation of the experimental setup and testing procedure was given to ensure that testing was done successfully and within the device limits without damaging either the equipment or the prototype.

The setup was integrated with MATLAB for programming; MATLAB was also used for graphical display of the data. A PC card was used as an interfacing device between the external hardware and the programming software. The setup was done in such a way that the drain currents were recorded based on the equivalent gate, source and drain voltages.

The test measurements that were conducted at GT included gate and drain sweep, injection and tunnelling tests, a temperature sensitivity test and the lifetime of the devices. FG transistors were tested and measurements for gate and drain sweep, injection and tunnelling were obtained. Based on GT applications, some of the tests of interest to validate the research hypothesis were not done at GT. The GT experimental setup did not accommodate measurements for retention and endurance tests. Insertion and isolation losses were also not measured to the extreme values, constrained by the fact that some equipment was not integrated with the setup that was available.

As stated in Section 3.2, during the research process observational study and experimental measurements were conducted in GT, which enabled the researcher to develop insight on a practical level.

3.6.2 For fabricated IC

The design (chapter 4) with Cadence software and the IC fabrication were done to enable further measurements and also to enable validation for the mm-wave range. The fabricated chip was mounted on a custom-made PCB to characterise the FG transistors and CPs. The following devices were fabricated:

- Dickson CP
- Two voltage doublers; one with bulk connected to V_{DD} and another with V_{DD} connected to the next stage
- Four FG devices with different transistor ratios and the coupling capacitors used for programming.

A two-phase non-overlapping clock driver was integrated with the CPs and the voltage doubler. Thus, it is not shown as a separate sub-system. The qualification protocol for the experiments performed were generalised for frequency and device scaling dependency:

- Frequency is the independent variable used to characterise the normal and high-frequency applications for FG and CP devices.
- The programming time is dependent on the settling time for the output voltage of CPs. Thus, high frequency and settling time were evaluated for CPs.
- Although device sizes can be manipulated as independent variables, device scaling is dependent on the technology process used (IBM 8HP), which determines the minimum device sizes.
- Temperature tests and baking of the devices were measurement conditions used to estimate the devices' lifetime. The charge loss and current leakages are temperature-dependent; temperature affects the charge retention and device endurance.
- For dependent variables where direct measurements are not possible, drain/source current and extracted threshold voltages were used.

The charge in an FG must be maintained at a certain level for normal operation, thus charge loss was evaluated, as it relates to the effects of current leakages. An endurance test was performed to determine the degradation of the threshold voltage as a result of programming cycles. The experimental results were then used to characterise the ON (insertion) and OFF (isolation) states of the FG devices. The experimental setup for FG devices is shown in Figure 3.3..

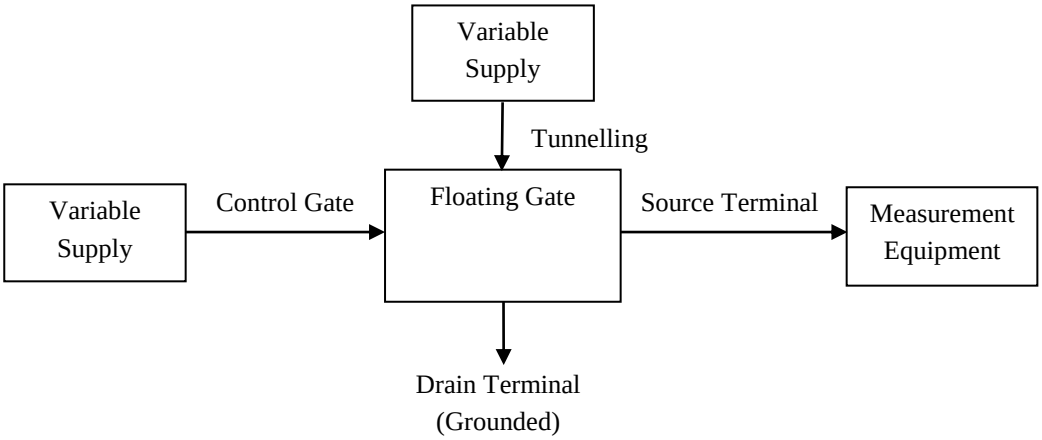


Figure 3.3. Measurement setup for FG devices.

FG devices of different sizes were used during the measurements. The devices were programmed under different conditions and evaluated for each test condition; the actual equipment used in Figure 3.3 is presented in Figure 3.7. The objective remained to validate the suitability of FG devices for mm-wave application. The tests, procedure and acceptance criteria for FG are discussed in Table 3.4.

Table 3.4.

The tests and experimental procedure for FG devices

Objective	Procedure	Acceptance Criteria
1) Tunnelling test: To determine programming efficiency and output current by applying different tunnelling voltages.	All terminals are grounded except the tunnelling voltage terminal. Different tunnelling voltages are applied and the source current is then measured. A gate sweep is then performed to determine the	Tunnelling voltage is usually supplied by CPs, thus a supply of $4V_{DD}$ (maximum supply from pump circuits) must be sufficient for tunnelling. The tunnelling time must be reduced as

	effects of tunnelling. Before tunnelling, injection is applied for a long time. To ensure that the device is erased, this is done for each tunnelling voltage.	tunnelling voltage increases.
2) Injection test: To determine the effects of erasing the devices by using different control gate and source voltages.	Different injection voltages are applied for a specific source voltage. Before injection, the devices must be tunnelled for a long time to ensure that they are programmed. A gate sweep is used to determine the tunnelling effects.	The combination of control gate and source voltage must provide sufficient injection. Programming time must be reduced by using higher voltages.
3) Retention test: To determine early lifetime failures for soft erase and programming.	The devices are baked at 150 °C for 72 hours. A gate sweep is then performed to determine the charge loss for each programmed state (1 or 0).	The devices must still function after being baked; this applies to both the programmed and non-programmed devices. (Not enough samples or PCBs for statistical analysis.)
4) Endurance test: To determine the minimum read/write cycles before device failure.	The devices are programmed or erased until failure, for a maximum of 10 000 cycles.	The devices must show negligible endurance failure for 1 000 read/write cycles. For 10 000 cycles, the threshold variation must be less than 40 %.
5) Temperature test: To measure the effects of temperature over normal operating conditions.	The devices are programmed or erased under different temperature conditions for the range between -25 °C and 55 °C.	For typical operating temperatures, -25 °C and 55 °C, the threshold variation must be less than 10 %.
6) Frequency test (MHz to GHz): for mm-wave validation.	Applying different frequency signals at control gate terminal from MHz to GHz. The source current is then measured at different frequencies.	The devices must be able to operate in the GHz range (> 10 GHz). The test was performed from low to high frequency and the maximum operating frequency had to be determined.

The experimental procedures presented in Table 3.4 provide detailed objectives for each test and the evaluation criteria used. The pumping efficiency and maximum voltage for CPs were measured. The loading effects and operating frequency determine the maximum performance

that can be achieved. The frequency was varied at discrete levels during measurements; continuous variation led to poor performance and this was verified at simulation level. The experimental setup for the CPs is shown in Figure 3.4.

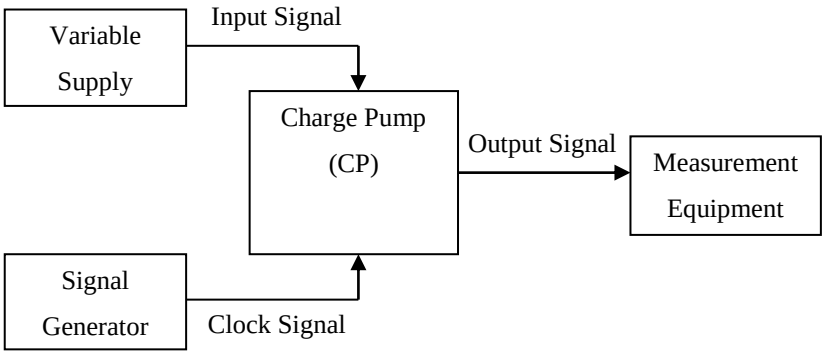


Figure 3.4. Measurement setup for charge pumps.

The functionality of the CP is based on charge transfer and also highly dependent on the stage capacitances and output load. Thus, charge leakages due to device scaling and non-negligible contribution of parasitic capacitances affect the performance of the CP. The Dickson CP and voltage doubler were measured and evaluated by varying different parameters (using the setup in Figure 3.4): clock frequency, input voltage and output load. Thus, the experiment was designed to validate the suitability of the CP for the IBM 8HP technology process. The tests, procedures and acceptance criteria for CPs are discussed in Table 3.5.

Table 3.5. The tests and experimental procedure for charge pumps devices

Objective	Procedure	Acceptance Criteria
1) Basic test: Determine output voltage and settling time.	Set input voltage to 1.2 V and apply a square wave clock signal with a 1.2 V. The output is connected to a measurement device and then results are determined.	The maximum output voltage should be $4V_{DD}$. An acceptable output voltage must be between $3V_{DD}$ and $4V_{DD}$.
2) Frequency test (MHz): Performing the first test at different frequencies.	If frequency is varied continuously, the charge transfer cycles are changed and there is insufficient time for stages to transfer charge. Thus, testing is done at discrete frequencies. The	The output voltage must show negligible dependence on frequency variation in the range of 10 – 50 MHz. Settling time must be less at higher

	range is 10 – 50 MHz with 10 MHz increments.	frequencies, because of faster charge transfer rates at high frequencies.
3) Loading effects test: To determine the variation of output voltage based on different loads.	The load is connected to the output of the pump circuit. Capacitive load is varied in the pF range and the resistive load is varied from kΩ to MΩ.	The loads must not severely affect the output voltage and source current. It is expected that resistance in range of 100 kΩ and few MΩ has a negligible influence on the output voltage.
4) Input voltage sweep test: Performing the first test with different input voltages.	The input voltage is varied from 0 – 2.4 V in steps of 0.4 V.	Because of threshold voltage, input voltages less than 0.4 V are expected to be insufficient to drive the pump circuits. Voltages higher than 1.5 V are expected to stress the oxide, which could result in higher current leakages and thus poorer performance for CPs.
5) Power measurements: Determine the output power for normal operation and the maximum achievable power.	A variable load resistance is connected to the output. The current and voltage are measured for different resistances and the output power is determined. The experiment is also performed to determine the maximum output power.	The maximum output power must be in the range of a few tens of µW.
6) High frequency test (GHz): For mm-wave validation.	Same procedure as second test, but the frequency range is in the GHz range.	The CPs must be able to operate in the GHz with fewer leakages and without suffering high losses due to parasitic effects. No specific criteria; this test will form the basis of characterising against the mm-wave frequencies (> 10 GHz).

The measurement procedures for CP in Table 3.5 provide the detailed objectives of the experimental tests and the evaluation criteria. The fabricated IC was placed on a PCB board, which enabled integration with an external power supply and the measurement equipment. MOSIS sponsored a piece (16 mm^2) of the wafer, shared with other universities/institutes

during a multi-project wafer (MPW) run, through the MOSIS Educational Programme (MEP). The particular institutional run was further shared by three postgraduate students. It is shown in Figure 3.5.

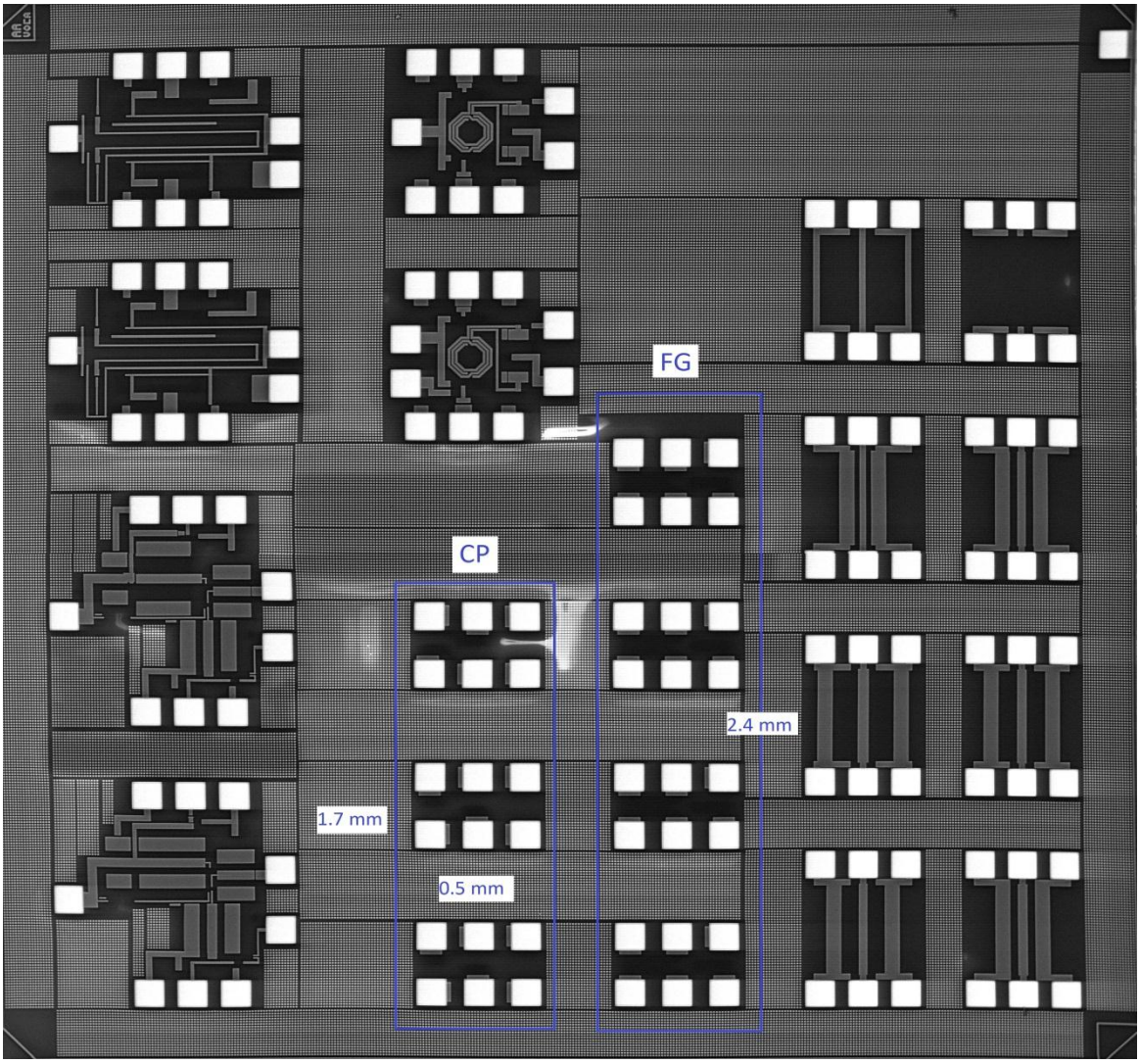


Figure 3.5. Fabricated chip of the MPW has an area of 4 × 4 mm^2. The CP and FG highlighted devices were fabricated for this project. The fabrication was done through MEP, which sponsors the fabrication run once a year.

The other circuits shown in Figure 3.5 were developed for different research projects in the MPW. The fabrication process was sponsored by MEP after submission and acceptance of the proposal to use the IC as a research prototype. There are three CP devices and four FG devices. The PCB design is shown in Figure 3.6.

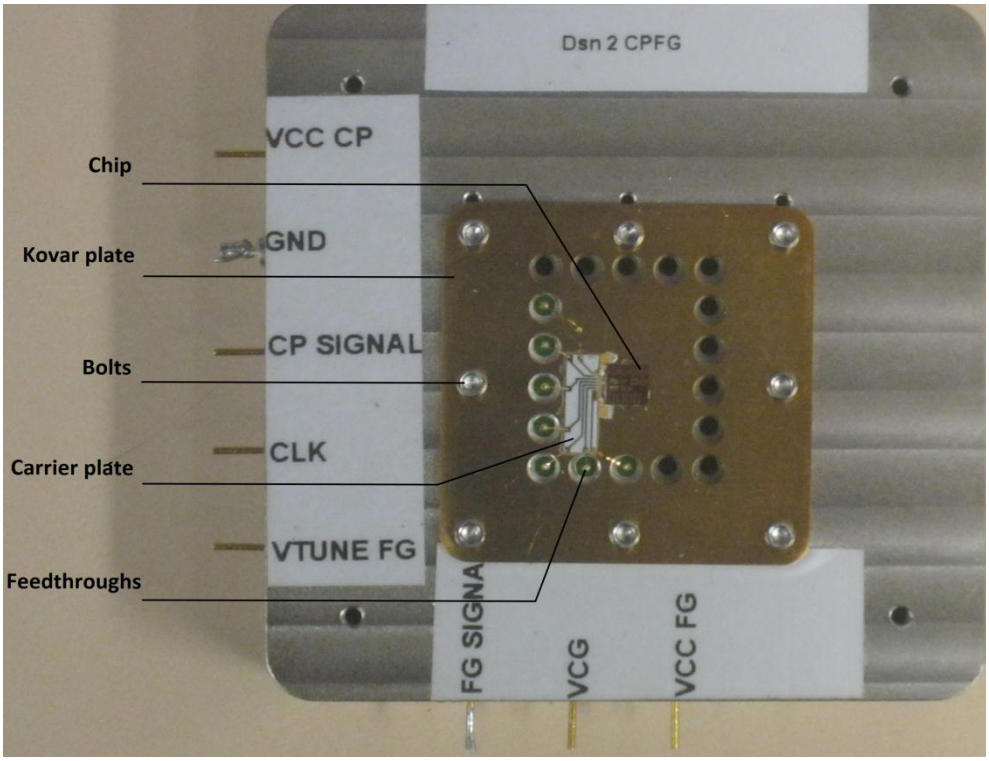
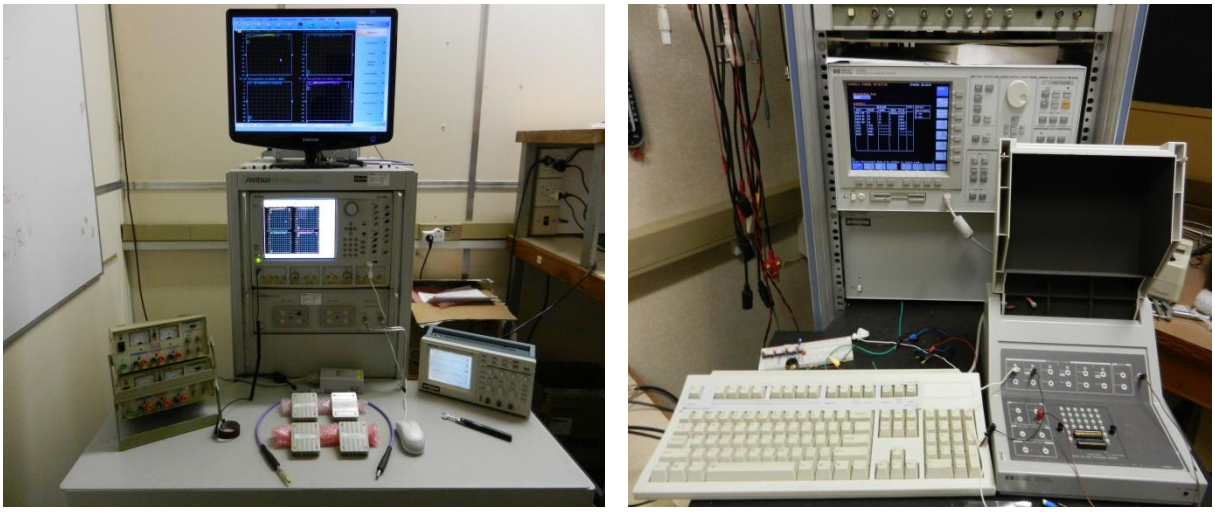


Figure 3.6. PCB designed for experimental measurements with physical dimensions: area of 7 × 7 cm^2 and a height of 2 cm.

The designed PCBs have the same structure as in Figure 3.6; in a single PCB there is one CP and one FG device. The FG terminal includes control gate voltage, tunnelling voltage, the output signal (drain/source) terminal and the supply voltage. The CP has the clock input, input signal and a supply voltage terminal; both the CP and FG devices share the same ground terminal. The diced wafer is placed on a gold-plated Kovar carrier plate by using conductive epoxy. The Kovar plate is used to provide a ground plane for the chip. An Aluminium substrate, which is used for wirebond pad connections, is soldered onto the Kovar plate. The Kovar carrier plate is screwed onto a larger nickel-plated aluminium casing, which has feedthrough connections to enable coupling with measurement equipment. The bondwires used are 18 µm in diameter.

The equipment used during measurements is shown below in Figure 3.7.



a) b)

Figure 3.7. a) shows two power supplies, oscilloscope, VNA and the four PCB prototype devices. b) shows the parameter analyser and its test fixture. The equipment not captured includes a signal generator and an oven for baking the devices.

The test equipment presented in Figure 3.7 was connected to a computer to capture the measured results. The two dual power supplies were used to provide up to four voltage levels during injection and tunnelling of FG devices. Before and after programming the FG devices, the parameter analyser (Agilent 4155B Semiconductor Parameter Analyser) and its test fixture (Agilent 16442A Test Fixture) were used to measure the gate and drain sweep of the FG devices. A vector network analyser (VNA) (Anritsu VectorStar Broadband VNA ME7828A) was used to measure the *S*-parameters at the output of the device; only one port was measured at the output owing to the limitations on the experimental setup design. Input measurements were prohibited by the programming capacitors used for tunnelling and injection. The signal generator (Agilent 33250A Function/Arbitrary Waveform Generator) was used to supply the clock signal to the CPs and an oscilloscope (Tektronix TD220) and multimeter were used to take measurements.

3.7 CONCLUSION

The methodology followed in this research has been described. The software and fabrication process used were also provided in this chapter. The schematic for the fabricated IC and the PCB design were presented with the experimental measurement setup used. The experimental

setup and procedures were discussed, as well as the devices tested during the measurements. At the time of this research work, on-chip probing equipment was unavailable locally, thus additional on-chip and PCB design wirebonding was required. The effects introduced by additional circuitry placed a limit on frequency measurements, as well as the effects introduced by connectors, cables and the substitute equipment.

CHAPTER 4: DESIGN AND SIMULATIONS

4.1 INTRODUCTION

This chapter discusses the design constraints and trade-offs. It then describes the design and analysis of various sub-systems implemented for fabrication. Layout designs and simulations results are also discussed. The fabricated IC was used for practical measurements, as described in Section 3.6.2. The approximations and assumptions were based on the reviewed literature and used during prototype development. The simulations were used to validate models and measure the accuracy of the designs. The effects of device scaling were evaluated for different sizes and by manipulating the controllable parameters and independent variables. Experimental results are provided in chapter 5 for comparison purposes and are also evaluated to characterise CPs and FG devices. Thus, the characterised devices were analysed to validate the hypothesis.

In this chapter, CPs, FG transistors and two-phase clock generator circuits are the main focus. The design is based on the IBM 8HP process technology and thus all process model parameters are provided and available from the PDK used for the schematic, layout design and simulation. To validate the theoretical analysis, a number of parameters were varied during the simulation analysis. Different simulation results are presented and discussed in this chapter.

4.2 CLOCK GENERATOR

The CPs use two-phase non-overlapping clock signals, reflected in Figure 2.17. An inverter and two NOR gates are used as instances for the clock generator. The NOR gates have been designed using an inverter as a reference, thus the primary primitive will be an inverter. The operating frequency for CPs ranges from 5 – 50 MHz; the period ranges from 20 – 200 ns. The pumps are highly dependent on the rise/fall time of the clock signal. The rise time was chosen to be 2 ns, smaller by an order of 10 to the minimum period. The clock generator’s rise time is required to be short in order to reduce the charge sharing between subsequent stages of CPs:

charge sharing limits the charge transfer and thus, the output drive capability. Hence the clock generator has a direct influence on the experiments designed for CP circuits. The rise/fall time is related to the propagation delay; it is approximated as twice the propagation delay (2.48).

The value of C (output capacitance) must take into account the capacitance of the coupling capacitors. For the initial design, capacitive loading effects were designed with 2 pF load capacitance, $C_{L.}$ The value of C also included the output capacitance of the inverter. It is observed from the PDK that the intrinsic capacitors of the NMOS and PMOS transistor are much smaller than 2 pF and C can thus be approximated as equal to C_L. Using the values given here and the threshold voltage as given in Table 3.2, the on-resistance is calculated to be

$$R_{on,N} = \frac{\tau_{PHL}}{C\left\{\ln\left[4\left(\frac{V_{0H}-V_{tn}}{V_{0H}+V_{0L}}\right)-1\right]+\frac{2V_{tn}}{V_{0H}-V_{tn}}\right\}}. \qquad (4.1)$$
$$= 638.2\,\Omega$$

The on-resistance can also be expressed as

$$R_{on,N} = \frac{1}{K_n(V_{0H}-V_{tn})}, \qquad (4.2)$$

with K_n given by

$$K_n = K_n^{'}\left(\frac{W}{L}\right)_n, \qquad (4.3)$$

where $K_n^{'}$ = 309.8 µA/V^2 is a process transconductance parameter; W and L are design parameters which are computed based on the rise time and the load capacitance. The transistor design equations must account for the effects of velocity saturation due to technology scaling. The on-resistance is then modified with (2.22),

$$R^{'}_{on,N} = \frac{R_{on,N}}{\kappa(V)}. \qquad (4.4)$$

By using (2.22) together with the process parameters in chapter 3 with V_{DS} = 1.2 V and L = 160 nm,

$$\kappa(V) = 0.36\,. \tag{4.5}$$

The scaling factor value is 0.36 or 64 % less compared to an ideal value where there is no saturation. For the experimental setup, the clock generator is not measured independently (the clock generator and CPs were integrated into the fabricated IC), thus probing into its functionality at an experimental level is limited to the overall performance of the CPs. The aspect ratio of an NMOS is

$$\left(\frac{W}{L}\right)_n = \frac{\kappa(V)}{K_n' R_{on,N}\left(V_H - V_{TN}\right)}\,. \tag{4.6}$$
$$= 2.36$$

An integer value of 3 was used for the transistor ratio. A length of 160 nm was used and a width of 480 nm (for a width-to-length ratio of 3). For an inversion point of V_I = ½V_{DD} and given that $V_{tn} \approx V_{tp}$ (see process parameters for short channel in Table 3.2), $K_p = K_n$. Therefore,

$$\left(\frac{W}{L}\right)_p = \frac{K_n'}{K_p'}\left(\frac{W}{L}\right)_n\,. \tag{4.7}$$
$$= 6$$

Thus, the width of the PMOS is 2.88 µm. The parasitic capacitances depend on the device ratios; these parameters are computed to verify that their effects are negligible compared to the load capacitance. The parasitic capacitors that will have an effect are those connected between the input and the output of the inverter. For NMOS and PMOS, the contributing capacitors are C_{GD} and C_{BD} for each transistor. Since the transistor operates in the saturation region, the C_{GB} capacitors make no contribution even though they are connected between the input and output. The approximated values have been computed as follows:

$$C_{GD} = C_{ol}, \tag{4.8}$$

$$C_{BD} = C_{jD}. \tag{4.9}$$

The total parasitic capacitance is

$$\begin{aligned} C_{PAR} &= C_{ol,p} + C_{ol,n} + C_{jD,p} + C_{jD,n} \\ &\approx \left[C_A(V)L_nW_n + 2C_P(V)(L_n + W_n)\right] + \left[C_A(V)L_pW_p + 2C_P(V)(L_p + W_p)\right] \\ &= 5.89\, fF \end{aligned} \tag{4.10}$$

The value resulting from (4.10) was computed using the process parameters as presented in chapter 3 and (3.7) - (3.8). Thus, the parasitic capacitance is much smaller than the 2 pF load capacitance. The NM was computed by first determining the following parameters [32]:

$$\begin{aligned} V_{IH} &= \frac{2K_R(V_{DD} - V_{TN} + V_{TP})}{(K_R - 1)\sqrt{1 + 3K_R}} - \frac{(V_{DD} - K_R V_{TN} + V_{TP})}{K_R - 1}, \\ &= 0.84\,\text{V} \end{aligned} \tag{4.11}$$

$$\begin{aligned} V_{OL} &= \frac{V_{IH}(K_R + 1) - V_{DD} - K_R V_{TN} - V_{TP}}{2K_R}, \\ &= 0.14\,\text{V} \end{aligned} \tag{4.12}$$

$$\begin{aligned} V_{IL} &= \frac{2\sqrt{K_R}(V_{DD} - V_{TN} + V_{TP})}{(K_R - 1)\sqrt{3 + K_R}} - \frac{(V_{DD} - K_R V_{TN} + V_{TP})}{K_R - 1}, \\ &= 0.57\,\text{V} \end{aligned} \tag{4.13}$$

$$\begin{aligned} V_{OH} &= \frac{V_{IL}(K_R + 1) + V_{DD} - K_R V_{TN} - V_{TP}}{2}. \\ &= 1.127\,\text{V} \end{aligned} \tag{4.14}$$

Note that the first letter of the subscript represents either input or output (*I* or *O*). Each output parameter depends on the equivalent input parameter; low input results in high output and vice versa. The K_R parameter is given by

$$K_R = \frac{K_N}{K_P} = 6.6 . \tag{4.15}$$

The low NM is

$$NM_L = V_{IL} - V_{OL} = 0.43\,\text{V} \tag{4.16}$$

and the high NM is given by

$$NM_H = V_{OH} - V_{IH} = 0.29\,\text{V} . \tag{4.17}$$

The CMOS inverter has negligible static power dissipation, since the off-resistance is sufficiently large and as a result, there is no DC current. The dynamic power dissipation for a 20 MHz clock and load capacitor of 2 pF is computed as

$$p_D = CV_{DD}^2 f = 57.6\,\mu W . \tag{4.18}$$

The inverter transfer function is shown in Figure 4.1. The two points, (V_{IL}, V_{OH}) and (V_{IH}, V_{OL}), are determined when the slope is 1.

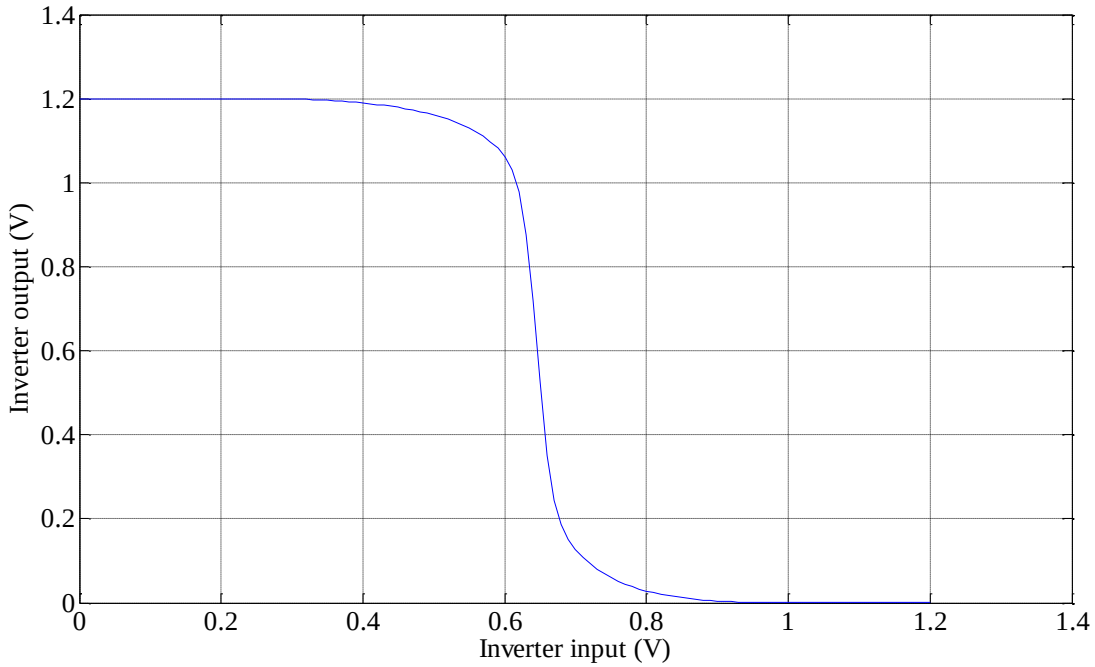


Figure 4.1. Transfer characteristic curve of an inverter.

From Figure 4.1, V_{IL} = 0.59, V_{OH} = 1.1 V, V_{IH} = 0.72 V and V_{OL} = 0.1 V. Using these values, the NMs are computed as

$$NM_L = V_{IL} - V_{OL} = 0.49\,\text{V} \tag{4.19}$$

$$NM_H = V_{OH} - V_{IH} = 0.38\,\text{V} \tag{4.20}$$

.

The theoretical values for NM_L and NM_H are 0.43 V and 0.29 V, respectively. The results show correlation between the theoretical models and simulations, thus this verifies the validity of the PDKs supplied by the foundry. The higher NM is higher by 31 %; the inverter was designed for an inversion voltage (midpoint of output voltage) of 0.6 V and the voltage achieved is 0.62 V. Thus, this has contributed to the higher NM as computed from the simulation results. The rise time of the inverter, in other words, the time from 90 % (1.08 V) to 10 % (0.12 V) of the output signal, is determined from Figure 4.2.

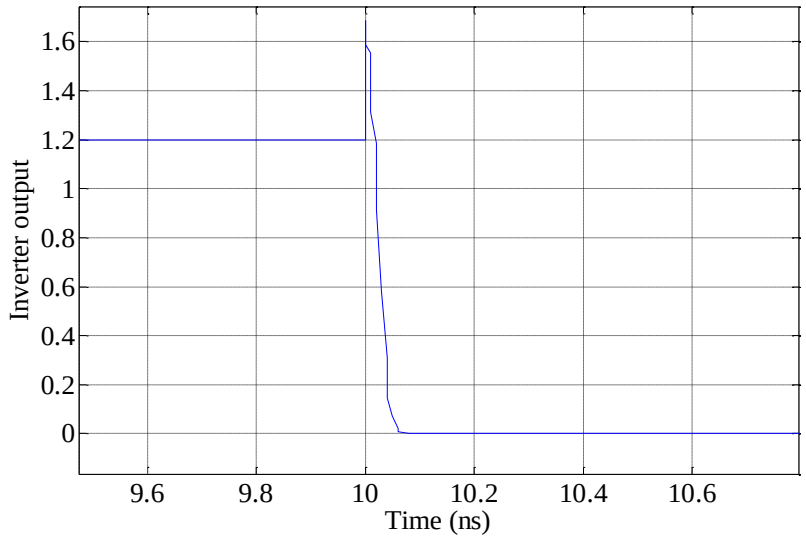


Figure 4.2. Inverter rise time is measured from the output vs time transient response.

The rise time in Figure 4.2 is 30 ps. A rise time of 2 ns was used as a maximum allowable value and the simulated value is lower by a factor of 100. It is desired to keep the rise time as small as possible: the small rise time is suitable to avoid charge sharing and to improve the CPs' performance. The operation of the inverter is verified in Figure 4.3, where the input and equivalent output are shown.

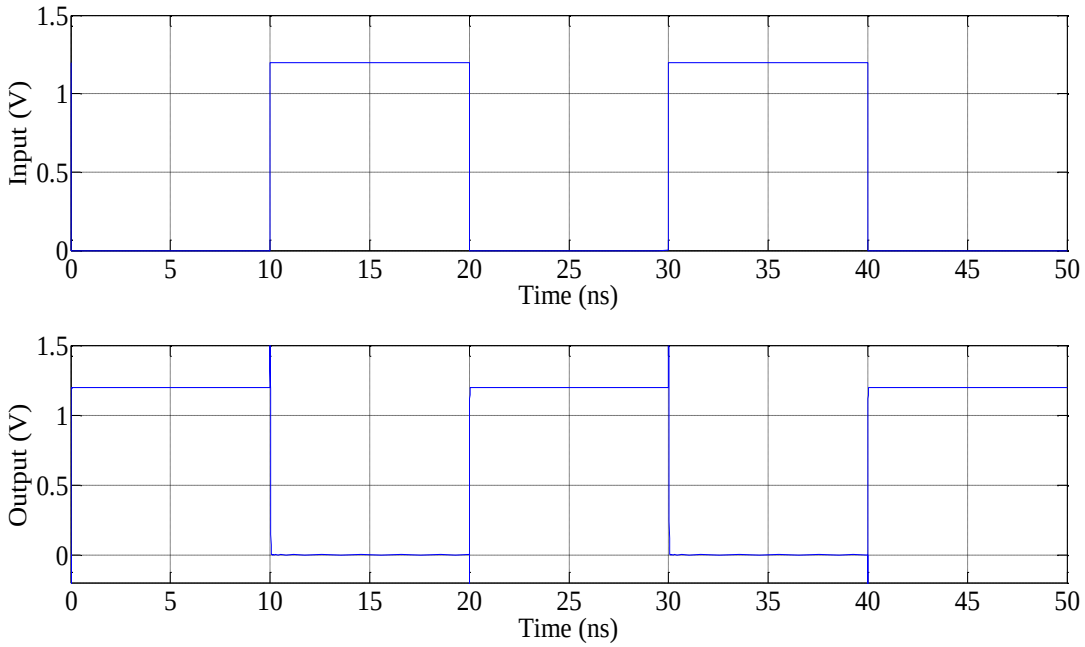


Figure 4.3. Response of an inverter; the input and output signals are shown.

The results in Figure 4.3 for an inverter show the input signal and the output response signal, which is the inverted input signal. As can be seen in Figure 4-3, there is a small overshoot on the falling edge, but the fall/rise times of 30 ps provide a good response time to minimise charge sharing. The schematic diagram for the inverter is given in Figure 4.4.

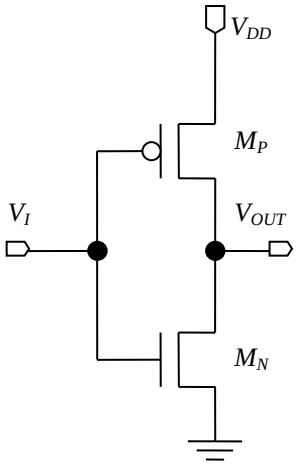


Figure 4.4. Inverter schematic.

The design of the NOR gate is based on the inverter design; the NOR gate schematic is derived from Figure 4.4. The analogy is that the NOR gate must have the same delay as the inverter. Thus, the on-resistance of the NOR must be the same as that of an inverter. To achieve this, the PMOS transistor ratio must be twice that of an inverter; the ratio for NMOS stays the same. For the NOR gate,

$$\left(\frac{W}{L}\right)_{p,NOR} = 2\left(\frac{W}{L}\right)_{p,INV}. \qquad (4.21)$$
$$= 12$$

The NOR gate is characterised based on the inverter. The rise time for the NOR is shown in Figure 4.5 and it is measured between 10 and 90 % of the output voltage. The value is 0.15 ns (150 ps), which is higher than the rise time achieved for an inverter.

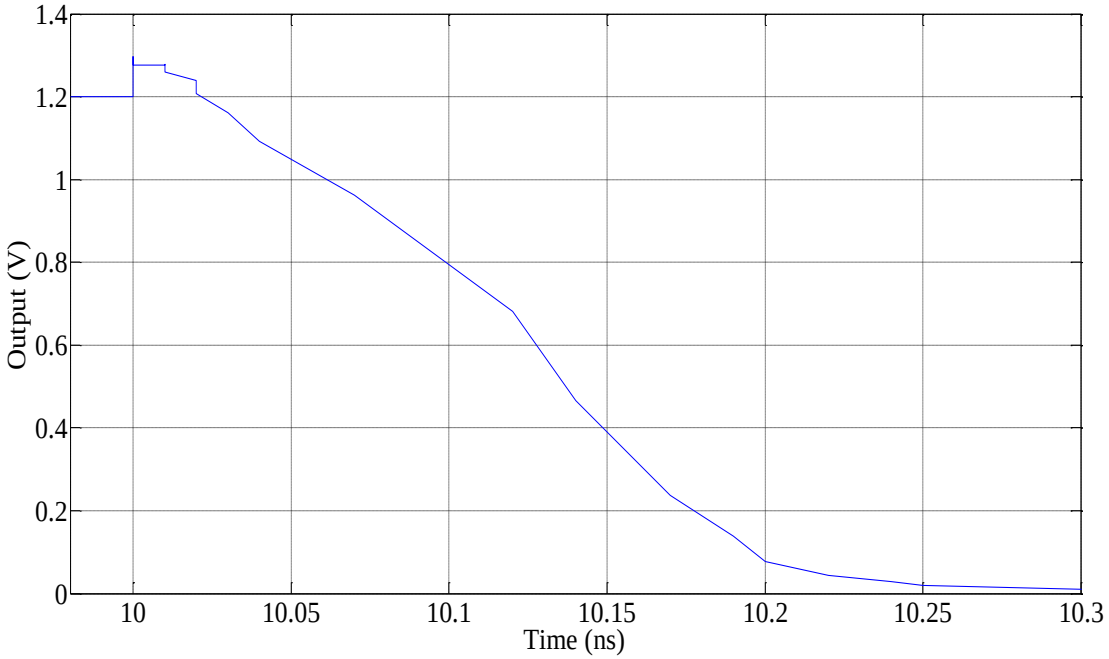


Figure 4.5. Transient response of NOR gate used to measure rise time.

The increase in rise time in Figure 4.5 is caused by the increased time constant at the output node. Although the output resistance is equivalent, the output capacitance has increased in response to more transistors connected at the output node. The operation of the NOR gate was simulated and the two inputs and the resulting output are shown in Figure 4.6.

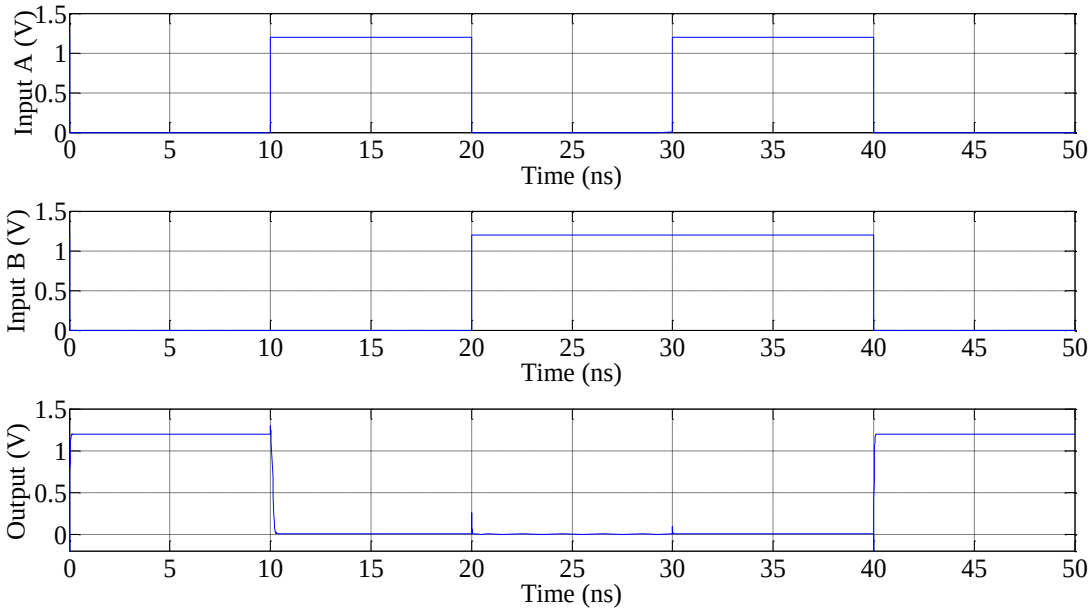


Figure 4.6. Two input signals and an output signal for a NOR gate.

The difference in response time between an inverter and NOR gate is noticeable in Figure 4.6. The rise/fall time of the NOR gate is much longer when compared to Figure 4.5 showing that of the inverter. The inverter and NOR gate circuits are used as building blocks for the two-phase clock generator. A clock generator uses an inverter and two NOR gates, hence the performance of both the inverter and NOR gate has a direct influence on the clock generator. The NOR schematic used for simulation is given in Figure 4.7.

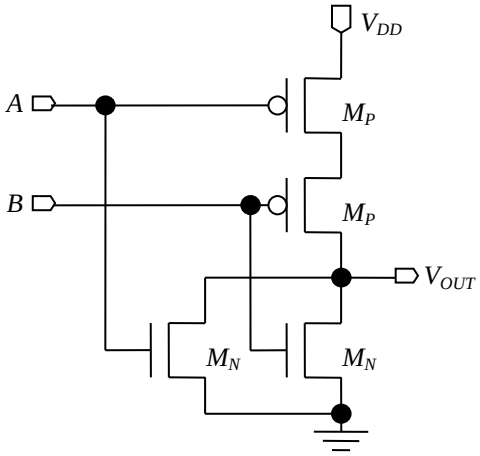


Figure 4.7. NOR gate schematic.

The circuit diagram for an inverter in given in Figure 4.4 and can be compared to Figure 4.7; it is noted that a NOR gate is a modified version of an inverter, with two NFETs connected in parallel. The values calculated for the inverter and NOR gate are given in Table 4.1.

Table 4.1. Summary of parameters calculated for inverter and NOR gate

Parameter	Inverter	NOR	Units
$(W/L)_n$	3	3	-
$(W/L)_p$	6	12	-
Length	160	160	nm
C (load)	2	2	pF
Rise/Fall time	2	2	ns
$R_{on,N}$	638.2	638.2	Ω
C_{PAR}	5.89	-	fF
P_D (dissipation)	57.6	-	µW
NM_L	2	-	V
NM_H	0.29	-	V

The other parameters, provided in Table 4.1, as calculated for an inverter, have a small variation and are approximately the same for the NOR gate. The schematic for two-phase non-overlapping clock signals using an inverter and the NOR gates is shown in Figure 4.8.

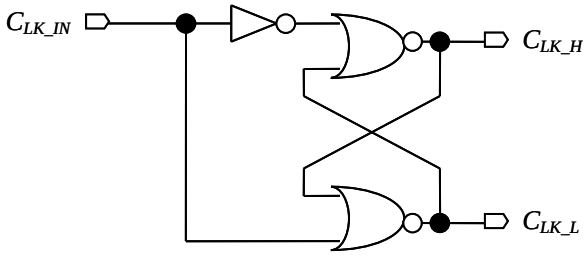


Figure 4.8. Two-phase clock generator.

The two-phase clock generator in Figure 4.8 provides non-overlapping clock signals to a CP circuit. The rise time must be small enough to avoid charge sharing for the CP circuit. The two-phase non-overlapping clock signals with the equivalent input clock signal are shown in Figure 4.9. The maximum and minimum output voltages are 1.2 and 0 V, respectively.

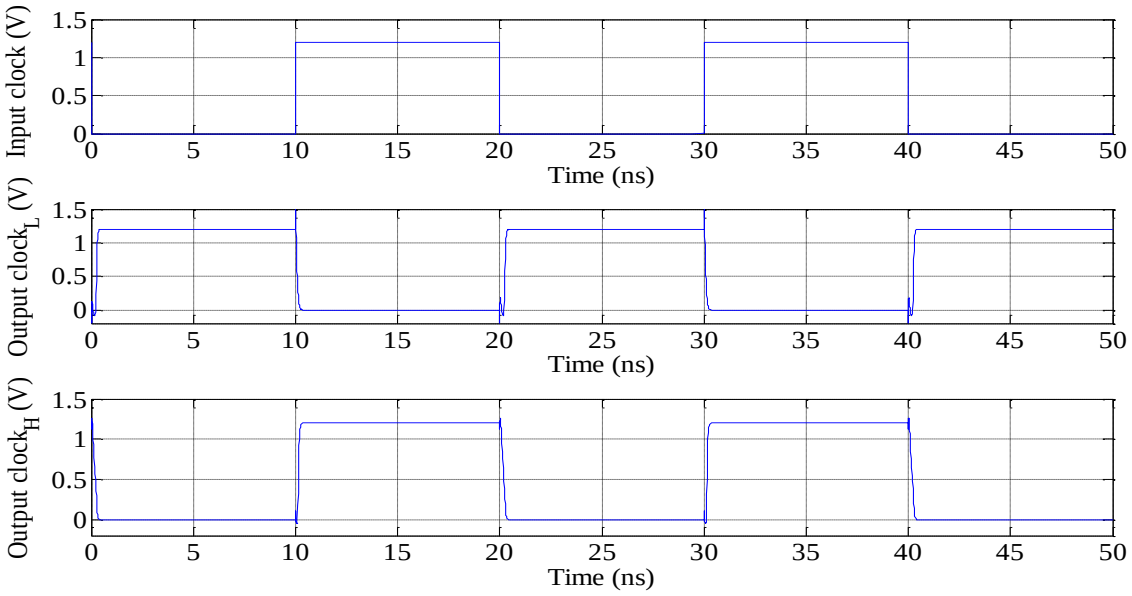


Figure 4.9. Input of a clock generator and the generated two-phase non-overlapping signals.

The generated clock signals are non-overlapping, as shown in Figure 4.9, and this is desired for efficient charge transfer, as well as to reduce charge-sharing effects for the CP circuits. Shorter rise time was achieved (approximately equal to the NOR gate rise/fall time) and this favours the requirement for high-frequency applications. A shorter rise time reduces the effects of charge sharing and maximises the efficiency of charge transfer to boost the output voltage. The clock generator was designed to enable the experimental design of CPs and thus, to evaluate CPs based on the secondary research questions, layout designs are presented in

Appendix B. The simulation results presented here correlate with the requirements to integrate the clock generator and CPs for experimental testing, as integrity is maintained after integration without any negative loading effects.

4.3 CHARGE PUMPS

The questions on CP devices relate to output loading and scaling effects. The reduced current due to scaling is less likely to provide the required voltage and current to program the FG devices. The experimental design was developed for evaluation of CPs' capability to program FG devices. The setup is such that external loads can be connected to the output, either capacitive or resistive. The output voltage is measured with and without the output load. The output power is also computed from the measured voltage and based on the load applied. Thus, power consumption and output voltages were the parameters to be measured experimentally and evaluated. Three different structures of CP have been designed for experimental measurements: the Dickson CP and two different voltage doubler CPs.

The CP design can be optimised for either chip size area or for power consumption. Although the chip size is also of critical importance for low supply application, power consumption has a higher priority. The chip size is compensated for by the scaling of process technology and that has enabled more focus to be placed on functional performance rather than area optimisation. The optimal number of stages is given in (2.81); the design choice and considerations are discussed here.

The output voltage is determined by the FG transistor and is not a specific value. Depending on the priority of the FG transistor in terms of programming efficiency and programming time, the voltage required is different. The size and oxide thickness of the FG also determine the minimum required voltage. Electron injection and tunnelling programming methods have a direct influence on charge retention in the FG, charge trapping in the oxide and on the overall reliability of the FG devices owing to the imperfections introduced during programming cycles. The targeted maximum output voltage is 5 V for programming FG devices that have been considered. Note that

$$x = 4.166\,, \tag{4.22}$$

since V_{DD} = 1.2 V, which is limited by process technology. In most applications, the CPs are designed for a specific targeted output current. In such cases, the stage capacitance is determined by (2.78). Thus, for a specific output voltage, the stage capacitance can be calculated. For programming FG transistors, the output voltage is specified, but not the output current. The parasitic capacitance is usually much smaller than the stage capacitance. As a design choice, α was chosen to be 0.1, a typical value for the ratio of parasitic-to-stage capacitance [38], [50]. This is verified by computing the total parasitic capacitance of the IC design. Using the experimental design values, the number of stages is

$$N_{opt} = 4.12\,. \tag{4.23}$$

The number of stages chosen must be an integer and four stages were used. There are three reasons for choosing four rather than five. The value calculated above is based on maximum targeted voltage, where a high value for parasitic capacitance was used. Furthermore, most CPs studied in the literature use four stages and this work can thus be benchmarked based on related work. Note that if the number of stages calculated was higher than four, then the design choices would have been re-evaluated by choosing the number of stages as four and calculating other parameters. The theoretical CP efficiency is

$$\begin{aligned} n_p &= \frac{x}{N+1+\alpha N^2 / N+1-x}\,. \\ &= 0.6 \end{aligned} \tag{4.24}$$

The efficiency calculated is about 60 %. The value above will be considered as minimum efficiency for the following reasons: 1) the number of stages calculated was rounded down and thus the maximum output voltage will be reduced from 5 to 4.8 V. The efficiency can be adjusted to 63 % by using the correct value of output voltage. 2) The parasitic capacitance was chosen to be a “worst-case” value taking into consideration the stage capacitance, i.e. α = 0.1. By using this value, the efficiency that can be achieved is 70 %. Thus, 60 % was used as a reference for minimum achievable efficiency. The design modifications are discussed with the

simulations results presented in this chapter. Note that the design does not depend on the actual value of the stage capacitance, thus the stage capacitance can be modified to adjust the coupling ratio to achieve the desired efficiency. As discussed earlier, a value of 2 pF was chosen for stage capacitance.

The transistor ratios are determined based on the output resistance. The output resistance is given by

$$R_{OUT} = \frac{N}{fC}, \\ = 100\,\text{k}\Omega \tag{4.25}$$

which is much higher than the on-resistance.

The design discussed is based on a basic structure: the Dickson CP. The designs developed in this dissertation are the Dickson CP and a voltage doubler. The design for both structures was made equivalent to enable comparison. The number of stages, input voltage and operating frequency are equal. To compare the Dickson and voltage doubler, the time constants of these devices must be equal. Since the structure of the voltage doubler has two capacitors per stage compared to the one of the Dickson CP, the stage capacitors of the voltage doubler must be twice that of Dickson CP. Thus, to ensure equal time constants, the device ratio for the Dickson CP is twice the ratios for the voltage doubler, i.e.

$$\left(\frac{W}{L}\right)_{Dickson} = 2\left(\frac{W}{L}\right)_{Doubler}. \\ = 20 \tag{4.26}$$

A value of 20 is therefore considered to be the design choice. The length (L) of the Dickson CP is 480 nm and for the voltage doubler the length is 180 nm, while the width of the Dickson CP is 10 µm and that of the voltage doubler is 1.8 µm. Other differences between the two structures will be discussed in the simulation section, including the performance and power analysis. The Dickson CP circuit is given in Figure 4.10.

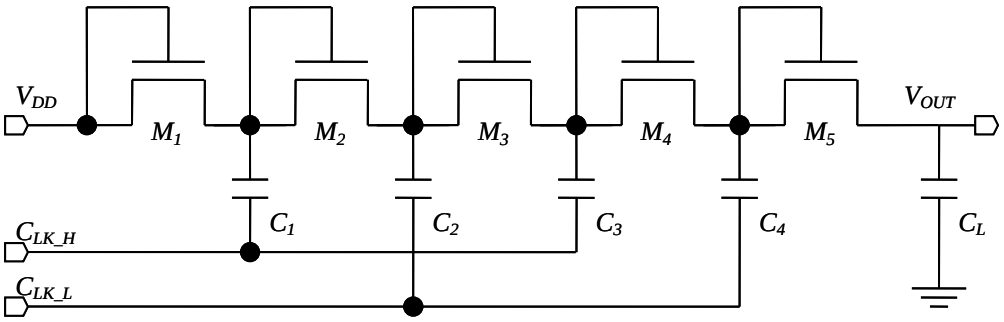


Figure 4.10. Dickson charge pump schematic.

The input clock signals (C_{LK_H}, C_{LK_L}) in Figure 4.10 are from the non-overlapping clock generator sub-system and NFET type transistors were used. The input voltage (V_{DD}) is limited by the process used to 1.2 V. The voltage doubler circuit is shown in Figure 4.11.

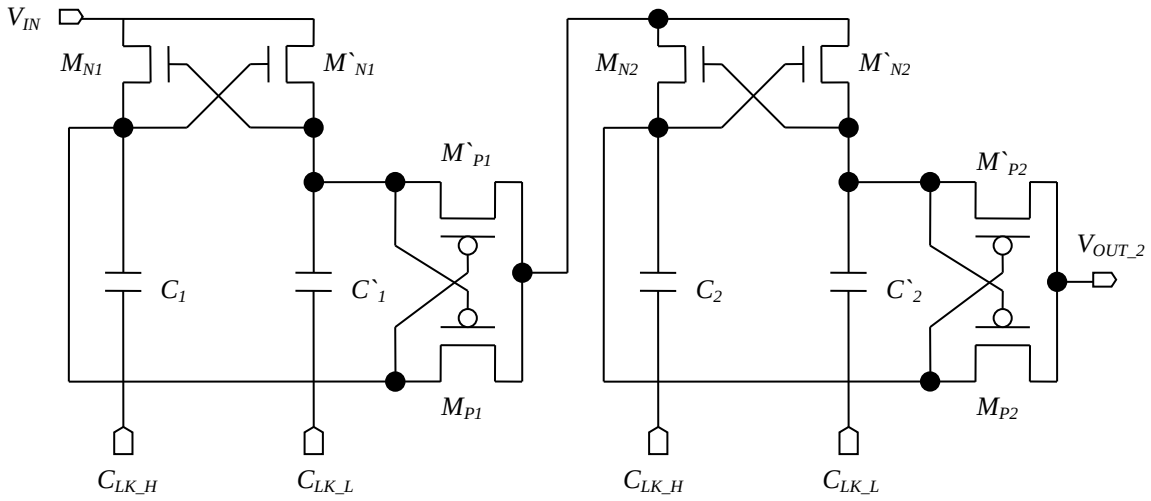


Figure 4.11. Voltage doubler CP used for simulations.

The input clock signal and the input voltage are supplied by the signal generator. The structures for the voltage doubler CP consist of four transistors (two NFET and two PFET) per stage compared to one transistor per stage for the Dickson CP. The number of capacitors per stage is double in comparison to the Dickson CP; Figure 4.11 shows the first two stages of the four stages implemented for the voltage doubler CP. The device parameters for the Dickson CP and voltage doubler are given in Table 4.2.

Table 4.2. Summary of parameters calculated for Dickson pump and voltage doubler

Parameter	Dickson	Voltage doubler	Units
$(W/L)_n$	20	10	-
$(W/L)_p$	-	10	-
L	480	180	nm
C (stage)	2	2	pF

The voltage doubler was designed based on CPs for comparison purposes and the results are provided in Table 4.2; the supply voltage for both CPs is 1.2 V. The results given in Figure 4.12 demonstrate the dependence of the Dickson CP on the clock frequency.

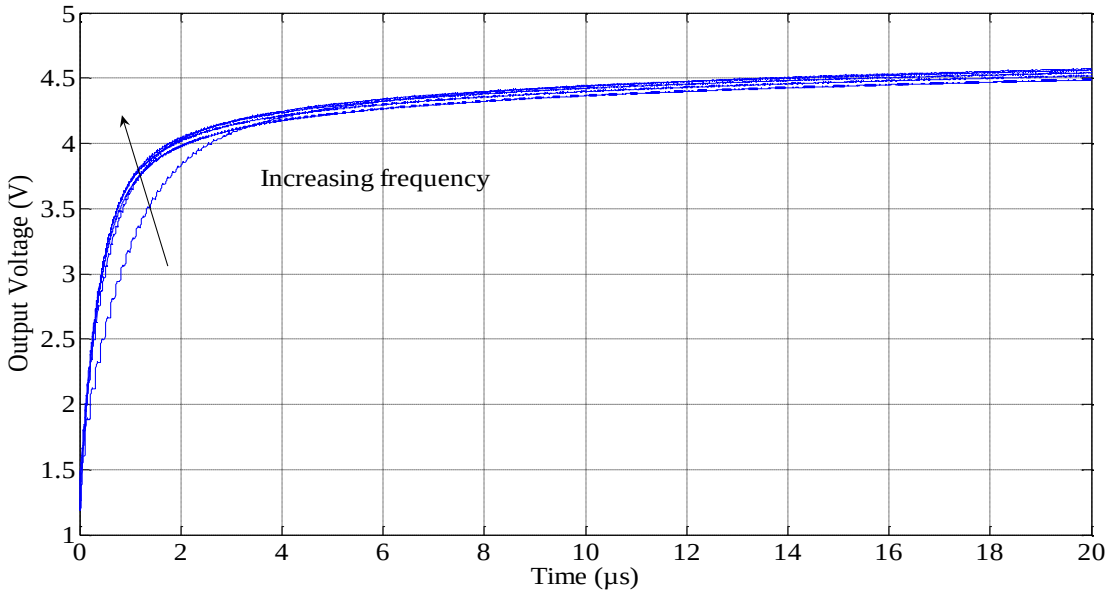


Figure 4.12. f_1 = 10 MHz, f_2 = 20 MHz, f_3 = 30 Mz, f_4 = 40 MHz and f_5 = 50 MHz. Transistor length and width, L = 480 µm, W = 10 µm, n = 1. These results are for the Dickson CP. Size of the stage capacitors; L = 4.8 µm, W = 10 µm and m = 4 (2.01 pF). Size of load capacitor; L = 4.8 µm, W = 10 µm and m = 4 (2.01 pF). MIM capacitors were used.

The frequency was varied from 10 to 50 MHz in steps of 10 MHz and the simulation results show less dependence on the clock frequency. The maximum output voltage achieved from

the four stages is 4.5 V, as shown in Figure 4.12, which should ideally be 4.8 V for a four-stage CP. Based on mathematical analysis, the output voltage should be less by the sum of the threshold voltage of all four stages, i.e. 4.8 V – 4 × 0.43 V = 3.08 V. These results are therefore inconsistent, as 4.5 V was achieved in simulation. The Dickson CP suffers significantly from loading effects, especially as a result of resistive loading.

The static CTS CP is greatly dependent on clock frequency. For a variation of 10 to 50 MHz, the output voltage varies from about 1.8 to 2.6 V. The output voltage achieved is also less than that of the Dickson CP. This type of result was not expected, given the results reviewed in the literature study. Possible reasons include high dependence on clock frequency and the stage capacitors. During simulation, capacitors were varied between low (0.5 pF) and high (µF) values. Thus, simultaneous optimisation of stage capacitors and the clock frequency is required. Another possible reason is the process fabrication required, which also limits the output voltage. The limitations are those that require the bulk of the NMOS to be connected to ground (which has a less limiting effect; it limits the minimum voltage) and the bulk of the PMOS to be connected to V_{DD} (which has severe effects of limiting the maximum CP output voltage). To enable validations for limits imposed by the bulk terminal on a PMOS, two voltage doubler devices were fabricated: one with the bulk connected to V_{DD} and another with the bulk terminal connected to the next stage.

The voltage doubler simulations with the bulk connected to the next stage are shown in Figure 4.13.

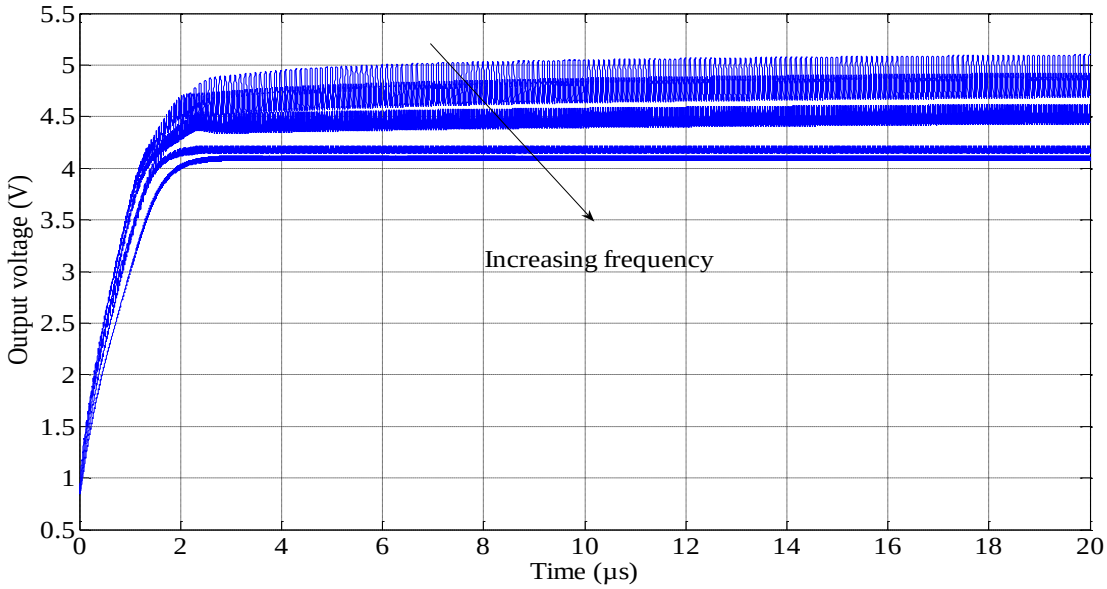


Figure 4.13. f_1 = 10 MHz, f_2 = 20 MHz, f_3 = 30 Mz, f_4 = 40 MHz and f_5 = 50 MHz. Transistor length and width, L = 480 nm, W = 10 µm, n = 1. These results are for the voltage doubler with the bulk of the PMOS connected to the higher voltage of the next stage. Size of the stage capacitors; L = 4.8 µm, W = 10 µm and m = 4 (2.01 pF). Size of load capacitor; L = 4.8 µm, W = 10 µm and m = 4 (2.01 pF). MIM capacitors were used.

The results in Figure 4.13 show that the output voltage varies between 4.1 and 5 V for a frequency of 10 to 50 MHz in steps of 10 MHz. The two graphs of lower voltage are for 40 and 50 MHz, thus at higher frequency the ripple voltage (desired) is reduced, as well as the output voltage (undesired). Because of process fabrication specifications, the bulk of the PMOS transistor must be connected to V_{DD} and the simulation results in Figure 4.14 show the limitation imposed by this requirement.

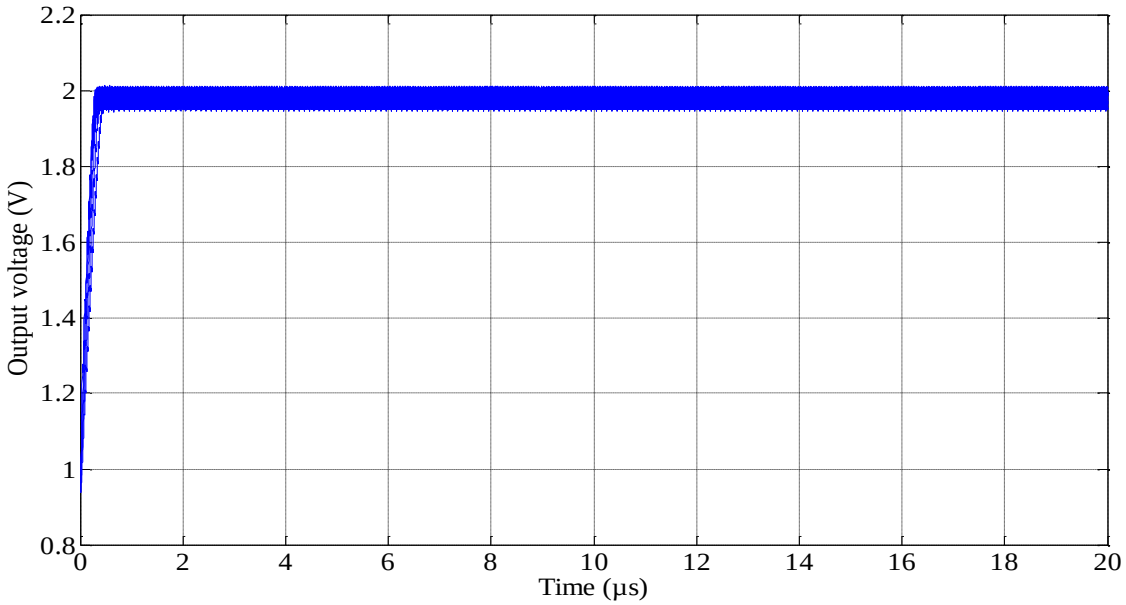


Figure 4.14. f_1 = 10 MHz, f_2 = 20 MHz, f_3 = 30 Mz, f_4 = 40 MHz and f_5 = 50 MHz. Transistor length and width, L = 480 nm, W = 10 µm, n = 1. These results are for the final voltage doubler with bulk connected to V_{DD}. Size of the stage capacitors; L = 4.8 µm, W = 10 µm and m = 4 (2.01 pF). Size of load capacitor; L = 4.8 µm, W = 10 µm and m = 4 (2.01 pF). MIM capacitors were used.

As can be seen in Figure 4.14, the output voltage saturates at 2 V and this limit is imposed by having the bulk connected to V_{DD} and not to the next stage. Because the source is not tied to the bulk of the PFET, the increase in the bulk-to-source voltage therefore limits the output voltage. Based on fabrication or process technology requirements, the voltage doubler is compromised and its performance is poorer than that of the Dickson CP, which is also unexpected, as discussed in the literature study (chapter 2). The comparison of the Dickson CP and the two voltage doublers is shown in Figure 4.15.

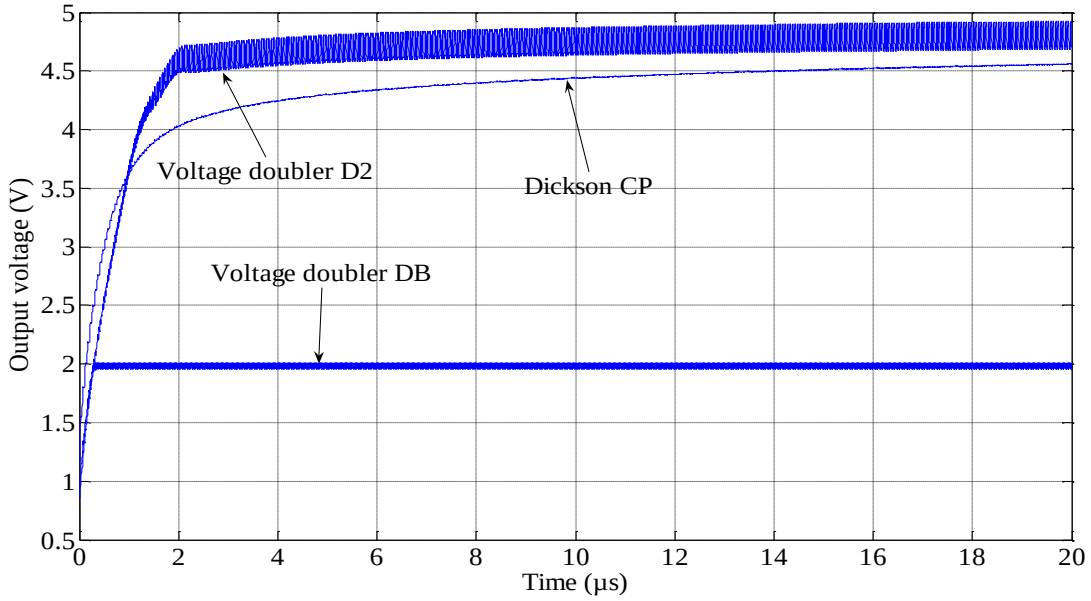


Figure 4.15. Simulations of Dickson CP and Voltage doubler at 20 MHz. Parameters for Dickson CP: C_{stage} = 500 fF, C_{load} = 3.5 pF, W = 10 µm and L = 480 nm. Voltage doubler parameters: L = 120 nm, W = 1.2 µm, C_{stage} = 5 pF, C_{load} = 15 pF. Input supply voltage is 1.2 V.

The two CPs in Figure 4.15 were both simulated for four stages. The Dickson CP achieved a maximum voltage of 4.5 V and the voltage doubler achieved 4.8 V. Two structures of the voltage doubler were simulated: the actual structure and the structure modified to comply with the fabrication requirements. The actual structure produced 4.8 V (Figure 4.15) and had its bulk terminal connected to the next high voltage stage, which from the second stage was higher than V_{DD}. The modified stage had bulk terminals for all stages connected to V_{DD} and thus its output voltage was limited to 2 V (Figure 4.15).

Device scaling causes current leakages and charge losses through the thin oxide thickness. Process variation and parasitic parameters have a negative effect on highly scaled devices. The simulation results presented here show that CPs can still provide the required programming voltage. The settling time is shorter and would thus enable faster device programming. The evaluation of output drive capability and loading effects are required to validate the hypothesis. The Dickson CP and voltage doubler CP were used to assess different structures and the layout designs are presented in Appendix B. The Dickson CP output voltage is limited by the threshold voltage and the voltage doubler CP was developed to overcome this. The experimental measurements are presented in Section 5.3.1.

4.4 FLOATING GATES

The experimental setup design for FG devices must provide ways to measure charge retention capabilities, switching characteristics and the effects of high frequencies in order to address the research questions. The dominant process parameter, which is a limiting factor that creates uncertainties for charge retention, is the oxide thickness. There must be a balance between the programming voltages and the programming time, or else the field across the oxide thickness produces charge leakages or the device may be over-programmed/erased and get trapped in one state. The extreme states of the FG devices can result in a short circuit between the FG and the channel or it could result in an absolute isolation of charge between the FG and the channel, thus prohibiting the device from being either programmed or erased. Some of the charges can be trapped in the oxide and could result in substrate conduction due to the increased oxide field [46].

The switching characteristics (determined by insertion and isolation loss parameters) are evaluated on the basis of two programmed states, the injection and tunnelling states. The experimental setup design enabled drain sweep and gate sweep in order to measure the drain current after programming the devices in either of the states. The high-frequency experiments were measured by a VNA which generates *S*-parameters. The FG devices must be programmed and then measured afterwards in both states separately.

The FG charge changes as the device is programmed. The threshold voltage also changes accordingly [11].

$$\Delta V_{FG} = \frac{\Delta Q_{FG}}{C_T}. \tag{4.27}$$

The input to the gate terminal is coupled through a capacitor. The coupled voltage is determined from the following expression,

$$\begin{aligned} \Delta V_{FG} &= \Delta V_{CG} \frac{C_{IN}}{C_T}, \\ &= V_{eff\_th} \end{aligned} \tag{4.28}$$

where V_{CG} is the control gate voltage and V_{eff_th} is the effective change in threshold voltage. There is interdependence between Q_{FG} and V_{FG}, thus V_{FG} is determined iteratively. A threshold voltage is extracted experimentally from I_{DS} versus V_{GS} and $I_{DS}^{½}$ versus $V_{GS,}$ for linear and saturation operation, respectively. A simulation schematic for programming/erasing an FG device is shown in Figure 4.16.

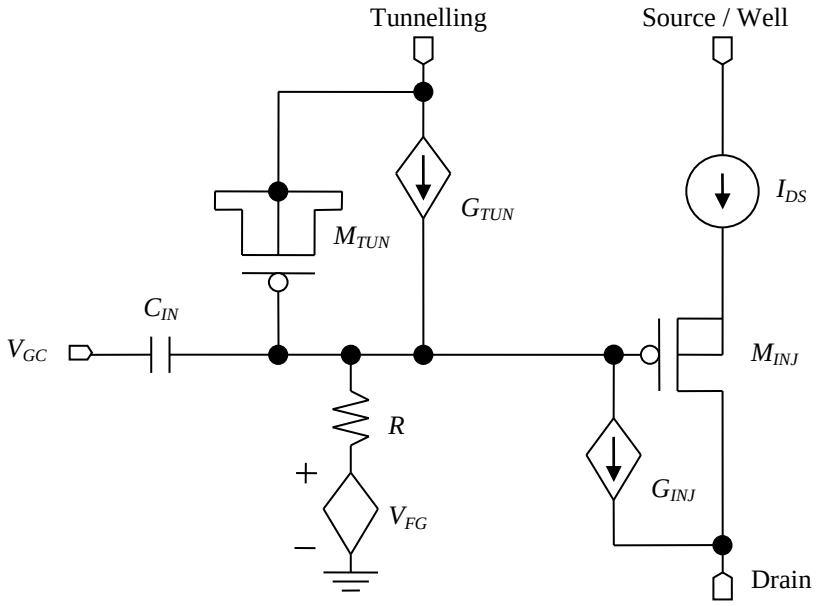


Figure 4.16. Simulation model with dependent voltage source (for DC convergence) and current sources (for injection and tunnelling) for FG device.

Note that the gate voltage in Figure 4.16 is coupled through the input capacitance and is directly proportional to the control gate voltage by the coupling ratio. Thus, I_{DS} versus V_{GC} (control gate voltage) is used for extraction and the results are modified with the coupling ratio constant. The charge can be evaluated iteratively as presented in [11].

The source, drain, bulk and FG potentials are input parameters to the model; the source current is also an input parameter. Another important parameter for FG devices in switching applications is the on-resistance,

$$r_{on} = \frac{v_{ds}}{i_{on}}, \tag{4.29}$$

where i_{on} is the drain/source current measured at the output terminal and v_{ds} is computed from the input parameters. The initial charge or voltage of the FG is unknown and dependent on the layout structure. In addition, the programming of FG is dependent on FN tunnelling and CHEI, which are modelled with fit parameters that are determined experimentally. No extraction experiment was done before the final fabrication of the FG devices. The extraction therefore had to be done based on the final measurements of the fabricated devices. The LVS was only done for CP devices and it was not possible for FG devices because of the floating node. The following FG devices were fabricated for characterisation purposes, Table 4.3:

Table 4.3. Summary of parameters for the fabricated FG devices.

Parameter	FG 1	FG 2	FG 3	FG 4
L (nm)	350	350	180	180
W (µm)	10	2	1.5	8
C_{TUN} (µm^2)	1 x 1	1 x 1	0.8 x 0.8	0.8 x 0.8
C_{IN} (µm^2)	10 x 10	10 x 10	8 x 8	8 x 8

All the devices in Table 4.3 were fabricated together with the Dickson CP and two voltage doublers. Two of the FG devices in Table 4.3 have a length of 0.35 µm, which is equivalent to the length of devices that were used at GT for experimental measurements. Thus, these devices were fabricated to enable measurement and to compare the results with those acquired at GT. The other two devices in Table 4.3 have a length of 180 nm, which will enable measurement for high scaling and further characterisation of high frequencies. The experimental measurements are presented in Section 5.3.2 and the layout designs are presented in Appendix B.

4.5 CONCLUSION

The design of the sub-systems has been presented in this chapter and verified with the simulation results. The Dickson CP and voltage doubler are the two CPs that have been fabricated. The clock generator has been integrated with the CPs for final fabrication and four different FG devices have also been fabricated for prototyping. The practical measurements are presented in chapter 5. The scaling of CMOS leads to charge leakages for FG devices and reduced supply voltage demands more effort from CPs to have sufficient voltage for programming the FG devices. Thus, the fabricated devices are used to determine the effects of charge leakage, as well as the voltage range and stability of CPs. Furthermore, suitability for mm-wave frequency has been tested.

The simulation results for the clock generator, Dickson CP and voltage doubler have achieved what was expected based on theoretical results and literature findings. The models for simulating FG devices depend on fit parameters that are extracted from measured results. Simulations were therefore completed after determining the fit parameters experimentally to validate the models used. The FG measured results are presented in Section 5.3.2.

CHAPTER 5: MEASUREMENT RESULTS

5.1 MEASURED RESULTS

The practical measurements are presented and discussed in this chapter. The results are divided into two sections: measurements of FG devices conducted at GT using pre-existing prototypes (not developed as part of this work) and measurements of the fabricated IC developed to validate the hypothesis.

5.2 GT MEASUREMENTS FOR FG DEVICES

This section presents the measurements of FG devices that were tested at GT. The following measurements are discussed:

a) Gate sweep after charge injection
b) Gate sweep after tunnelling the devices
c) Injection and tunnelling time characteristics.

The effects of tunnelling and charge injection were analysed after programming the devices and performing a gate sweep experiment. The output current was measured against the input voltage and the parameters were extracted from the graphical representation of the results. The measurements obtained at GT provided the basis for validating FG devices and were used as reference for the design of the fabricated devices. The experiments that were completed enabled practical insight: understanding the process/procedure of evaluating and characterising FG devices. The 0.35 µm technology node was used for the samples measured at GT and these devices formed the basis for the 130 nm technology node which was used for the fabricated devices.

The injection is measured based on three varying parameters: gate voltage, gate-to-drain voltage and injection time. Tunnelling characteristics are measured based on varying the tunnelling voltage and also the tunnelling time. The programming and erasing time, equivalent

to injection and tunnelling, are not measured directly. The time variation results are, however, extrapolated from the gate sweep results, which are time-dependent. The FG transistors which were tested have a structure as given in Figure 5.1.

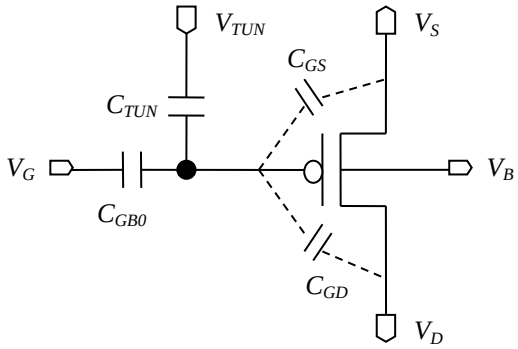


Figure 5.1. Floating gate transistor schematic. It is formed by connecting an input capacitor and tunnelling capacitor to a *p*-channel FET. The gate is not connected to any pin and thus forms a floating gate for the whole structure.

The FG transistor structure in Figure 5.1 is implemented physically by having a common gate for all three devices. The C_{TUN} and V_{TUN} are used for tunnelling the devices, while C_{IN} and V_{IN} are used for injection. For tunnelling, the voltage conditions are: $V_S = V_D = V_G = 0$; V_{TUN} is the only variable. The injection voltage conditions are $V_D = V_{TUN} = 0$; V_G and V_S are variables. The capacitors used are MOSFET capacitors formed by shorting the drain, source and bulk of a *p*-channel FET.

The results for different FG devices measured in GT are presented here: parameter values are presented in Table 3.3. The four devices used for measurements in GT differ in width or input capacitance. Devices 8 and 10 have the same width of 10 µm. The input capacitance for device 8 has an area of 10 x 10 µm^2, while that of device 10 is 1 x 1 µm^2. Device 71 and device 73 have a width of 2 µm and the input capacitance is different: 1 x 1 µm^2 for device 71 and 10 x 10 µm^2 for device 73. The two devices with a smaller width have a smaller range for the drain current; this can be seen by comparing the semi-log (drain) plots. The detailed results are presented in Appendix A and discussed in Sections 5.2.1 and 5.2.2.

5.2.1 Tunnelling results

The input capacitance affects the tunnelling time; the smaller the capacitance, the shorter the time. For example, the tunnelling time for device 8 is 0 – 5 s and that for device 10 is 0 – 1 s. The required tunnelling voltage is also smaller for device 10, 6.5 V compared to 8.5 V for device 8. Devices 8 and 10 have a range of 1.25 – 1.9 dB, while devices 71 and 73 have a range of 0.5 dB. Thus, the smaller device width limits the output current. This is directly related to the amount of charge that can be stored in the FG. A larger area (W x L) can store more charge and achieve higher drain current.

The tunnelling time is shorter for small input capacitance (device 10 and 71), 0 – 1 s. The input capacitance also affects the sub-threshold voltage; devices 10 and 71 have a sub-threshold of 1.7 – 2 V, and devices 8 and 73 have a sub-threshold between 0.7 – 0.9 V. Thus, a higher sub-threshold is required for the smaller input capacitance. The sub-threshold voltage determines the voltage for optimum injection efficiency for given programming conditions. The injection efficiency therefore does not depend on the sub-threshold voltage only; the bias conditions for drain-to-source voltage and the gate voltage affect the injection efficiency. The other effects for the devices considered will be discussed after presenting device injection results in the next section.

5.2.2 Injection results

Device injection reduces the FG threshold voltage and thus increases the drain current. The devices (8 and 10) with a width of 10 µm have a maximum current between 3.1 and 3.4 mA, while the devices with a width of 2 µm have a maximum current of 0.6 mA. Thus, the device area limits the amount of charge in the FG, which has a direct influence on the drain current. For all devices, it can be noticed that as the gate and drain-to-source voltage increases, the output current also increases. For smaller width, the current saturates at about 0.6 mA and for larger width it saturates at about 3.4 mA.

The effects of input capacitance are more visible at lower bias voltage for gate voltage and drain-to-source voltage. For a width of 10 µm, a drain of 0.25 mA is achieved for device 8 (1 x 1 μm^2 input capacitor) at the lowest bias condition and 2 mA is achieved for device 10 (10 x 10 μm^2). Thus, for both injection and tunnelling, smaller input capacitance results in faster charge transfer, when either adding or removing charge from the FG. For mm-wave

applications, device programming should be faster, thus it is desired that tunnelling and injection methods must be fast. The same results can be noticed from devices 71 and 73, where 0.33 and 0.15 mA were achieved at the lowest bias condition.

5.3 MEASUREMENTS FOR FABRICATED IC

In Section 1.2, four secondary research questions were presented to support the steps towards validating the hypothesis. A measurable outcome has been associated with each question, as shown in Figure 5.2.

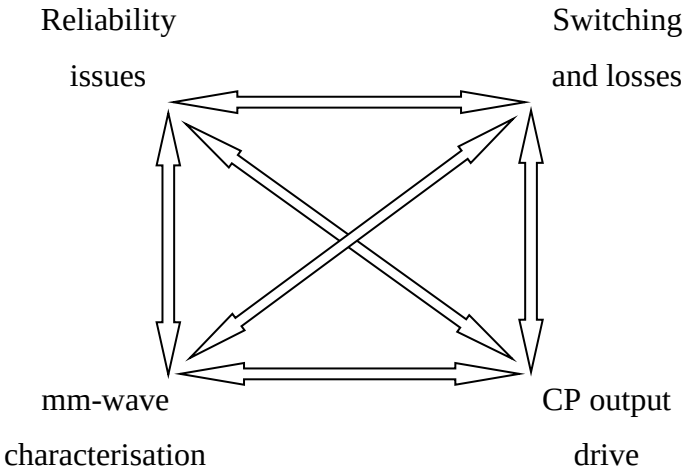


Figure 5.2. Measurable outcomes were derived from secondary questions and there is interdependency between the measurable outcomes.

The outcomes presented in Figure 5.2 are discussed below, while detailed experimental tests and procedures are provided in Table 3.4 and Table 3.5.

Reliability issues:

The reliability is measured based on charge retention capability, endurance against programming cycles, losses due to defects, process variation and charge leakages. A detailed discussion is presented in Section 2.7.1. The oxide thickness is reduced with technology scaling and the probability of charge leakages thus increases with device scaling. The oxide thickness, given by (2.31), has more effects on the reliability of FG devices. This is discussed in Section 2.4.2.1. There is a trade-off between the programming time and voltage. To avoid

stressing the devices, low voltages can be applied for a long time, but high voltages must be applied for only a short time. The oxide thickness has an effect on the injection efficiency and also determines the required programming voltages.

Switching and losses:

The insertion and isolation losses are determined by the amount of charge that can be transferred in the FG. The charge in the FG affects the threshold voltage and consequently the output current. The losses are determined by the ratio of either the currents or resistance of the ON and OFF states. For good switching performance, the ratio must be high, i.e. there must be sufficient difference between the two states.

CP output drive:

The CPs are used to program the FG devices. The voltage and current output drive capability must be sufficient to program FG devices and must not suffer from loading effects. The CPs are integrated with a two-phase clock driver and the output voltage is measured on different loads; current can also be derived from a known load applied at the output. The stage and load capacitors are in the same order, but (usually) the load capacitor is higher by a factor of two, as the output voltage is higher than the stage voltages. The stage capacitors must be much higher than the parasitic capacitors to reduce the parasitic effects, i.e. $C_{STAGE} >> C_{PAR}$. The stage capacitors are in the pF range.

mm-wave characterisation:

To meet the frequency bandwidth for mm-wave applications, highly scaled devices are required, but the transistor parameters (including parasitic and threshold voltage) are not scaling with the same ratio as the device dimensions (as discussed in Section 2.4.2.1) and this results in saturation velocity. Modelling complexity for process variation, interconnections and parasitic effects increase with a frequency increase, making it even more complicated to account for these parameters in mm-wave applications. There is also a demand for specialised equipment to reduce or minimise effects introduced by the measurement equipment, connectors and cables. Section 3.6.2 discusses the tests and the fabricated prototype used to evaluate the hypothesis.

5.3.1 Charge pumps

The first device measured was the voltage doubler with the bulk connected to the next stage. After a successfully captured measurement, the voltage doubler provided an output voltage of 1 V. This was lower than the expected output, in which, for simulation results, 4.8 V was achieved. In evaluating the possible reasons, a second device with the bulk connected to V_{DD} was measured. The output voltage achieved was 0.7 V, which was also much lower than the expected value of 2 V as achieved in simulations results. The last device to be measured was the Dickson CP. At first, the output voltage was 2.96 V, lower than 4.5 V. Further experimenting was done on the Dickson CP after 24 hours to assess the difference between simulation and measured results. While experimenting and testing against different conditions, the input voltage was increased to exceed 1.2 V in anticipation of the output voltage also increasing, but such experimenting (which may be attributed to human error) violated the maximum limit of the device, which is 1.2 V. The device output voltage dropped to 1.5 V and a further drop to 0.6 V was recorded after also increasing the amplitude to the clock input signal to exceed 1.2 V.

The simulation results were investigated. Since an input voltage of 1.2 V was applied to a four-stage Dickson CP, an ideal output voltage would be 4.8 V. The expected output voltage is $4.8 – 4 \times 0.43$ V = 3.08 V, with a 0.43 V drop per stage due to the threshold voltage. Thus, the measured results were correlating with the theoretical results by achieving 2.96 V with only a 0.12 V difference, which may be attributed to process variations and voltage drop across interconnections. The simulation results proved to be inconsistent with the actual theoretical results, as the simulation output was not reduced by the threshold voltages, as expected. There was a 0.3 V drop due to threshold voltages in simulation results compared to 1.72 V from the theoretical analysis. The reasons for this may be the PDKs provided by the foundry for the fabrication process that was used for the simulation environment settings. It was not possible to investigate further experimentally, as all the CP devices are integrated with a clock generator: the experimental setup design did not enable measurement of the clock generator and CP devices separately. The clock generator and CPs are interfaced on-chip without additional pins to probe or check subsystem-by-subsystem. The separate subsystems were not included in the fabricated chip owing to limited chip space in the MPW run. Thus, either the clock generator or the CPs were damaged, maybe even both.

5.3.2 Floating gates

The gate and drain sweep for all four FG devices were performed. The gate and drain sweep measurements for all four FG devices were taken before either injection or tunnelling. Similar measurements were also taken after programming the devices. Tunnelling was done first with 5 V and then injection was done with 5.5 V, as 5 V was insufficient. In all cases, the devices were also measured after 24 hours following either injection or tunnelling to determine possible overnight losses. The devices were left for 10 hours and measured again to determine possible charge losses. The second tunnelling was done immediately after the first and device 4 was declared faulty after the first injection, as it did not respond to further programming.

For the experimental measurements given below, a comparison is made to the measurements taken with the device in its initial state from fabrication (*INT*), after tunnelling the device from the initial state ($TUN5_1$), after electron injection following the first tunnelling (*INJ55*), 10 hours after injection (*10H*) and then after tunnelling following injection ($TUN55_2$). The compared measurements are derived from Appendix C, which presents detailed results based on different control gate voltages for drain voltage sweep measurements and different drain-to-source voltages for gate voltage sweep measurements. The DC programming conditions are presented in Table 5.1.

TABLE 5.1
THE PROGRAMMING CONDITIONS FOR TUNNELLING AND ELECTRON INJECTION.

Parameter	Tunnelling	Injection
V_{CG} (V)	0	5
V_{TUN} (V)	5	0
V_S (V)	0	5
V_D (V)	0	0
Pulse width (s)	15	15

The tunnelling and injection voltage was 5 V (as provided in Table 5.1), but this voltage was also increased to 5.5 V in certain conditions. The programming results are presented in this Section. The initial charge after fabrication is unknown and different for all devices, thus there is uncertainty about the results achieved by the measurements before programming the devices used as reference.

Device 4

The measured results for device 4 gate sweep are presented in Figure 5.3 and Figure 5.4.

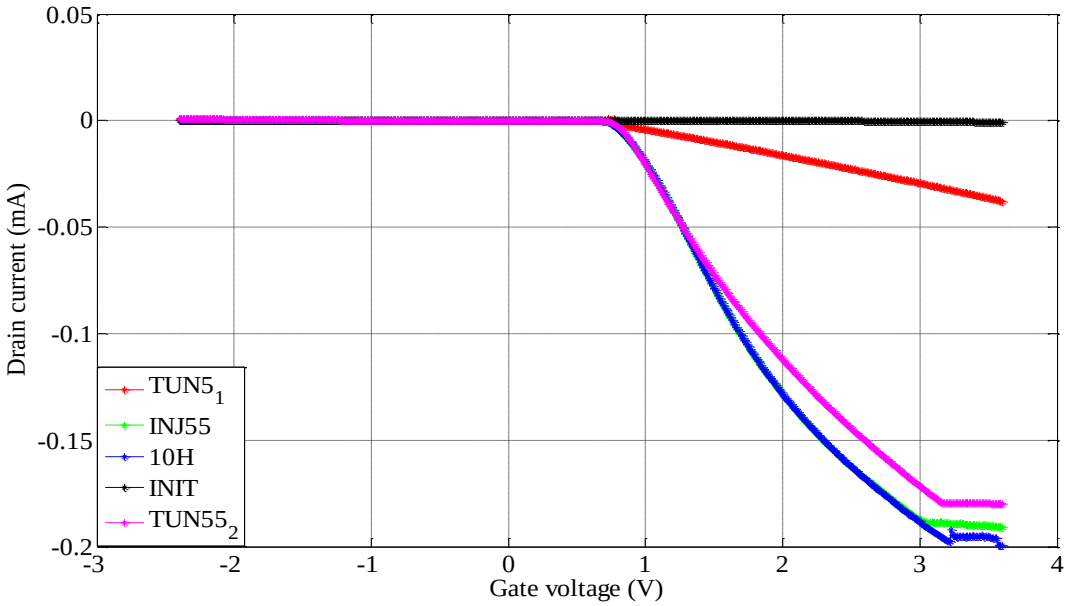


Figure 5.3. Gate voltage sweep comparison for device 4.

The fabricated device was measured directly from fabrication and the PCB test board without either tunnelling or injection, denoted by the *INIT* legend in Figure 5.3. After the first tunnelling (*TUN5*$_1$), the threshold voltage dropped from about 2.3 V to about 0.7 V; as shown in Figure 5.4. There was almost no change in threshold voltage after injection (*INJ55*), but there was an increase in the drain current. The device was measured again 10 hours after injection (*10H*), but there was no change for gate voltages below 3 V and a slight change above 3 V. Tunnelling was done following the injection (*TUN55*$_2$); although there is a visible change in current, the variation is within 0.025 mA. This value is much smaller compared to the difference between the first tunnelling (*TUN5*$_1$) and the injection (*INJ55*), which is about

0.19 mA. Also note that following the tunnelling, the device did not return to the same level of the first tunnelling.

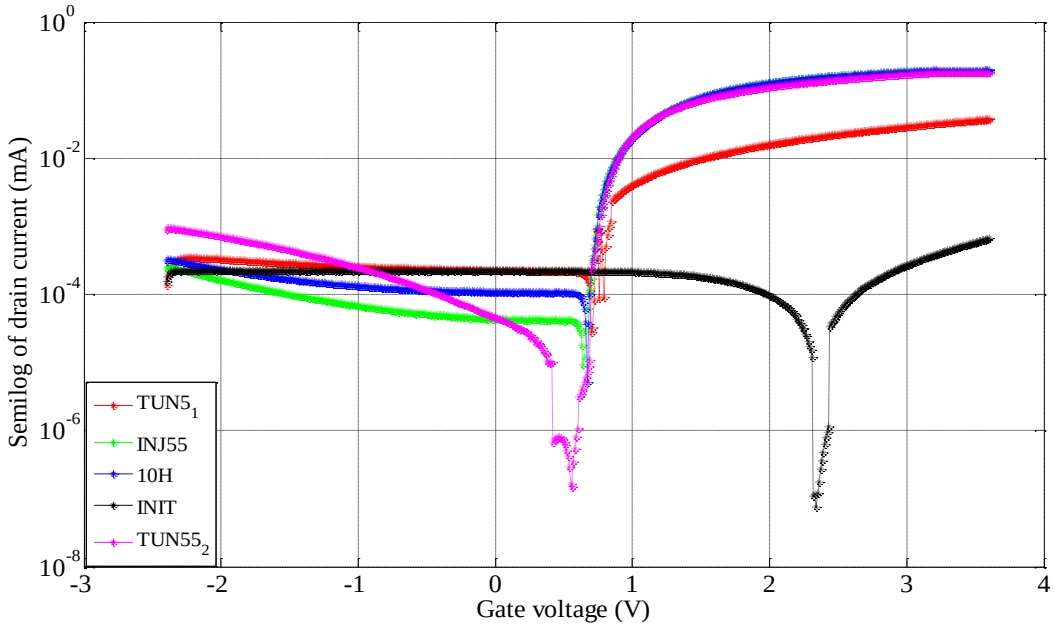


Figure 5.4. Gate voltage sweep comparison for device 4, with logarithmic current scale.

The threshold voltage is reduced by tunnelling and increased by injection. Following the first tunnelling, the threshold voltage was reduced (Figure 5.4), but subsequent programming with either injection or tunnelling did not have an effect on the threshold voltage. The gate sweep measurements are consistent with the drain sweep measurements presented in Figure 5.5.

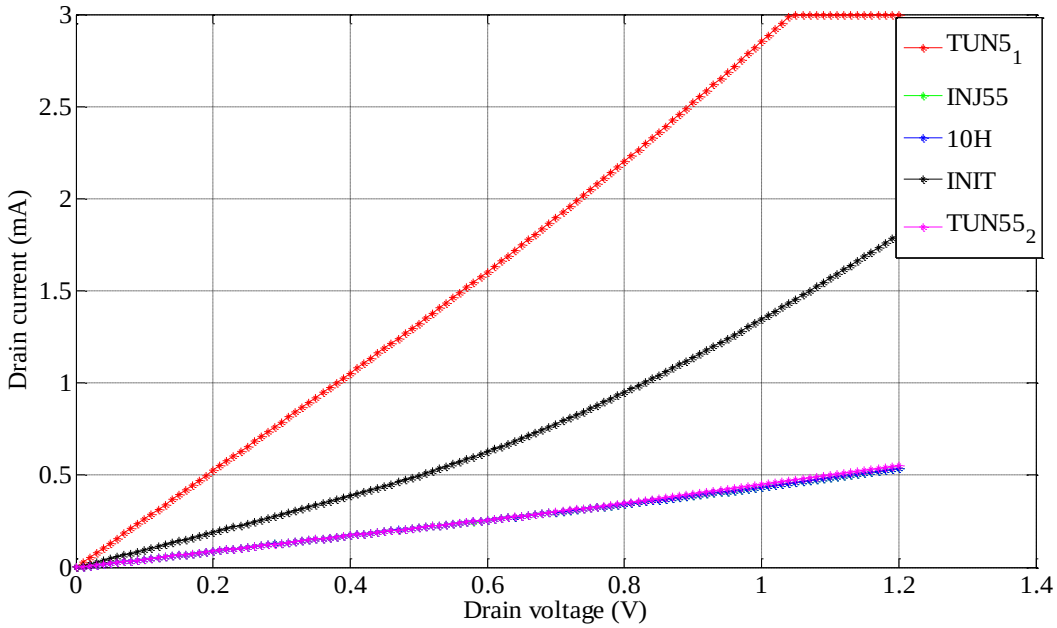


Figure 5.5. Drain voltage sweep comparison for device 4.

The first tunnelling reduced the threshold voltage and thus resulted in an increase in the drain current (*TUN5$_1$*), as shown in Figure 5.5. Although the gate sweep measurements show no change in threshold voltage, the drain voltage sweep does reveal that, following injection (*INJ55* and *10H*), the drain current was reduced, which translates to an increase in threshold voltage. This corresponds to what is expected; injection increases threshold voltage and thus reduces the output current. The electron tunnelling after injection was supposed to reduce the threshold voltage and increase the current. A second tunnelling was done, but did not have an effect. Subsequent results were taken and the tunnelling voltage was increased from 5 V to 5.5 V and then to 6 V, but the responses did not change. Another injection was then done and still no change was recorded. The device was thus trapped and neither further injection nor tunnelling programming was possible.

Device 3

The gate voltage sweep measurements for device 3 are shown in Figure 5.6 and Figure 5.7.

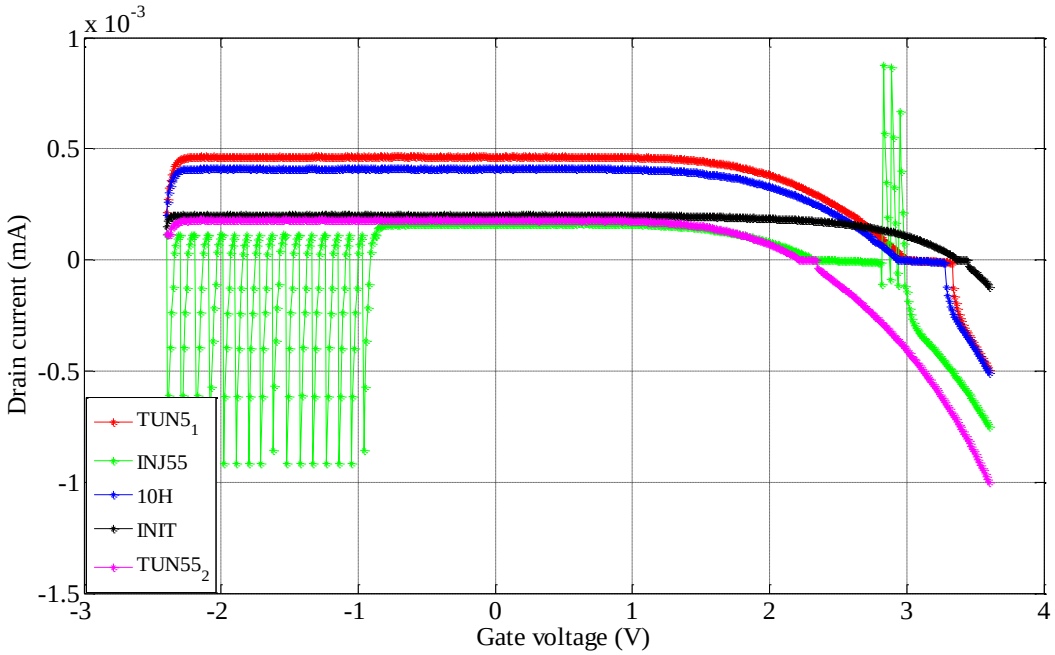


Figure 5.6. Gate voltage sweep comparison for device 3.

The initial device threshold (*INIT*) is higher than the other thresholds of either tunnelled or injected states; as shown in Figure 5.6. After the first tunnelling ($TUN5_1$), the threshold voltage was reduced and the current increased as shown in Figure 5.7. After injection (*INJ55*), the current was reduced, the threshold voltage was also reduced; the opposite was expected as predicted by GT measurements and the literature review (further discussion in Section 5.4).

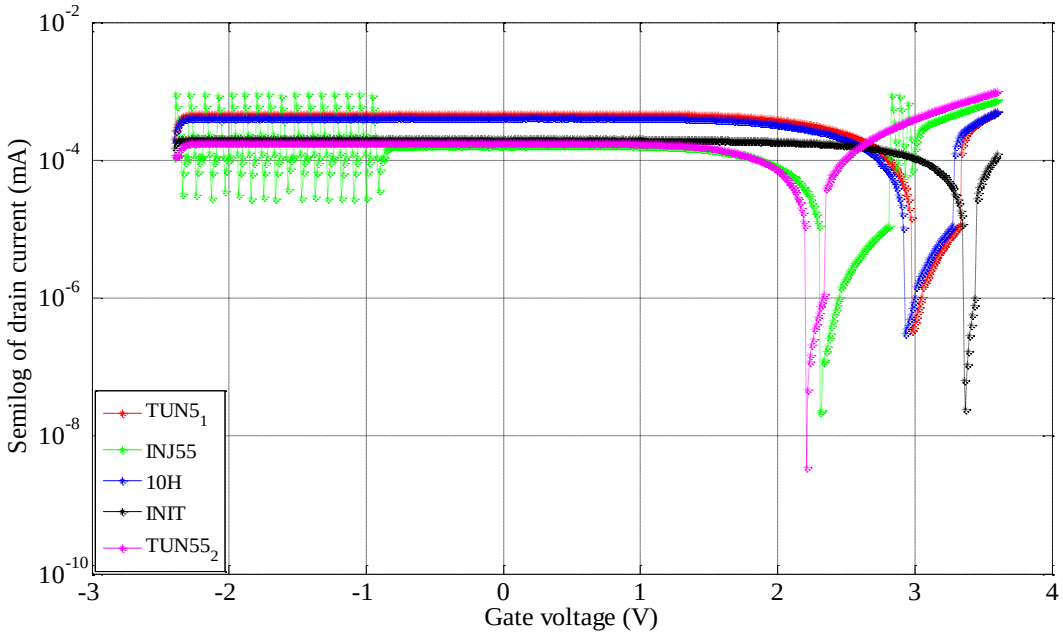


Figure 5.7. Gate voltage sweep comparison for device 3, with logarithmic current scale.

The measurements done 10 hours after injection (*10H*) present other undesirable effects (Figure 5.6 and Figure 5.7). The results (*10H*) were supposed to be similar to the injection results (*INJ55*), with a slight variation as in the case of Device 4 (Figure 5.4 and Figure 5.5). The results were similar to the first tunnelling case ($TUN5_1$), which could then be related to possible charge leakages. FG2 has the smallest transistor area and is thus likely to be affected by process variation and leakage currents, compared to the other devices presented here. Because of the smaller area, the amount of charge that can be transferred is smaller and that is why small variations have a higher influence. It is also interesting to note that tunnelling measurements after injection ($TUN55_2$) have reduced the threshold voltage, as is expected when tunnelling a device. To enhance the interpretation further, the drain sweep measurements were also done and are presented in Figure 5.8.

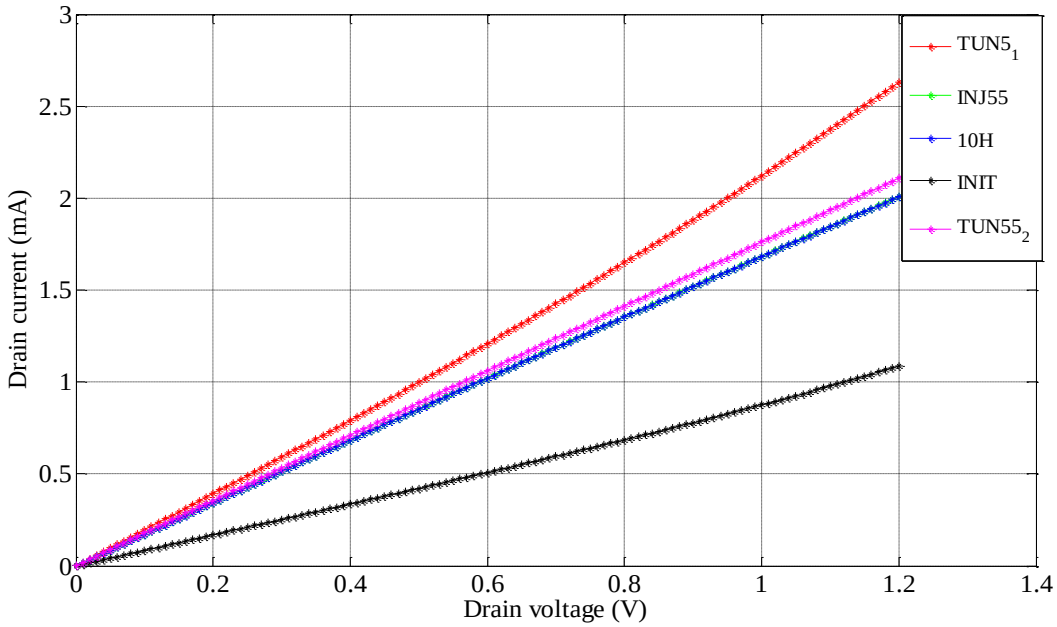


Figure 5.8. Drain voltage sweep comparison for device 3.

The tunnelling measurements (*TUN5$_1$*) show an increase in drain current from 1.1 mA to about 2.7 mA from the initial device state as received from fabrication (*INIT*); as shown in Figure 5.8. Injection has resulted in a decrease in current (from 2.7 to 2 mA), reflecting an increase in threshold voltage. The tunnelling after injection resulted in a slight increase in the drain current of about 0.1 mA. Further discussion is presented in Section 5.4.

Device 2

The measurement results for the gate sweep of device 2 are shown in Figure 5.9 and Figure 5.10.

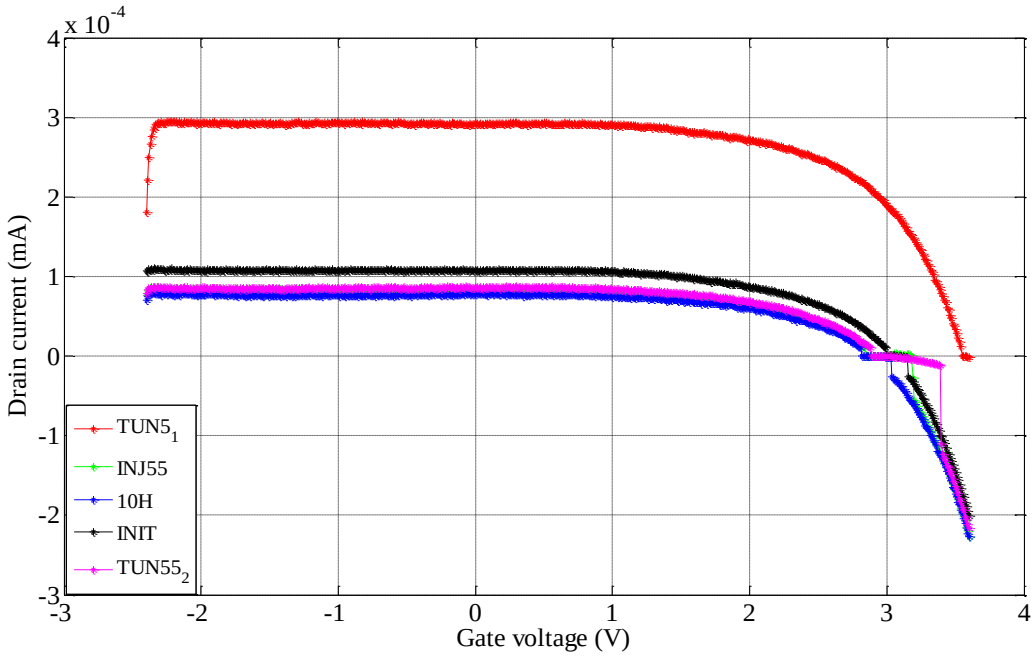


Figure 5.9. Gate voltage sweep comparison for device 2.

The initial state measurements (*INIT*), as shown in Figure 5.9, have a threshold voltage lower than the first tunnelling results ($TUN5_1$). This is rather different compared to device 3 and 4 presented here. The first tunnelling increased the threshold voltage, while injection (*INJ55*) reduced it to within 0.1 V compared to the initial state threshold, as shown in Figure 5.10. The tunnelling and injection provided opposite results with regard to threshold voltage changes. Even further tunnelling ($TUN55_2$) had almost no visible effects, as there was only the slightest change compared to the injection results.

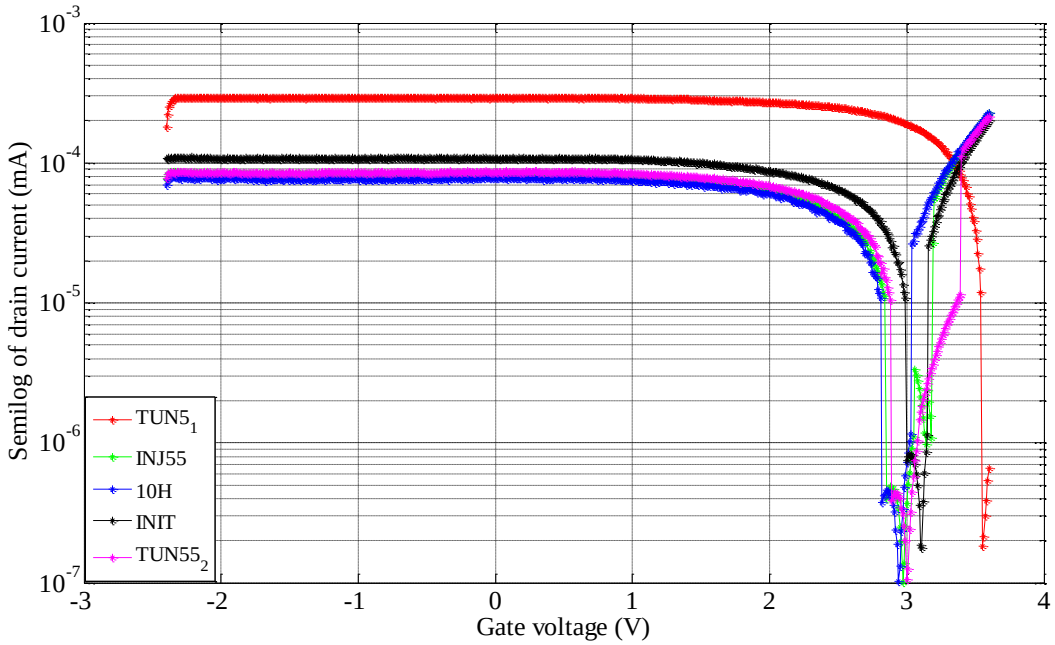


Figure 5.10. Gate voltage sweep comparison for device 2, with logarithmic current scale.

The results in Figure 5.10 show that the threshold voltage was varied between 2.9 and 3.6 V, with the initial state threshold voltage of 3.1 V. FG 3 has a longer gate length and the total variation is 0.7 V (between minimum and maximum values); this value is lower compared to FG 1 and FG 2. The contributing factor is the fact that the FG 3 threshold voltage shifted to a higher value after first tunnelling, which is opposite to results achieved in FG 1 and FG 2. Since reference is based on the initial measurement, there is uncertainty because the initial charge is unknown for these devices. The device dimensions and also the coupling capacitor dimensions are different. Owing to the shared floating inter-poly with different total area, the initial charge for each device is expected to be different and thus to cause different initial threshold voltages. The measurements taken immediately after injection (*INJ55*) and the measurements taken 10 hours after injection (*10H*) have a direct correlation and no variation. The drain sweep measurements are presented in Figure 5.11.

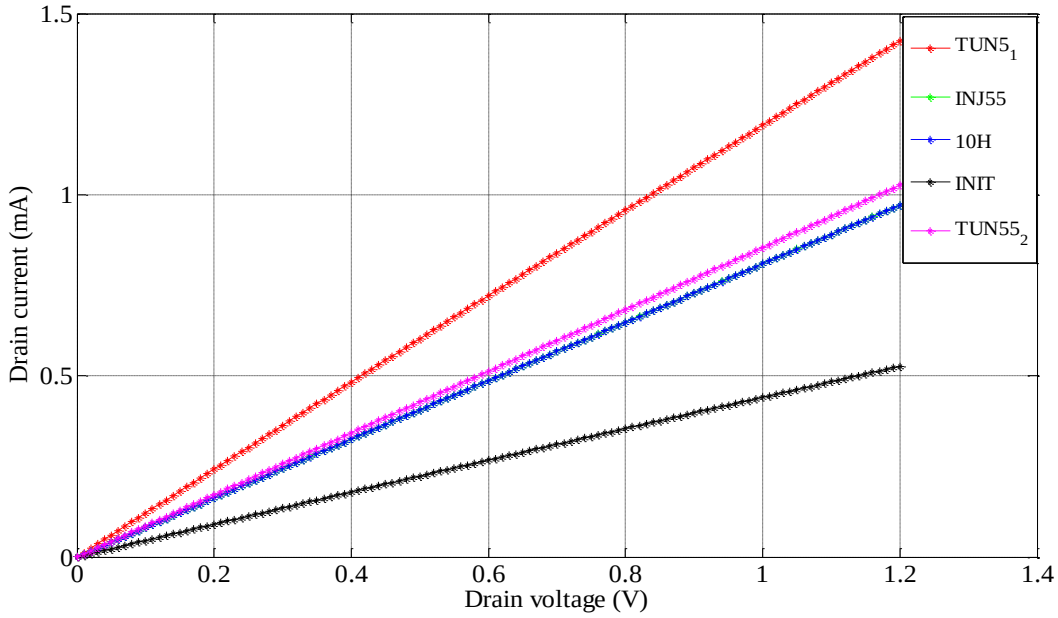


Figure 5.11. Drain voltage sweep comparison for device 2.

The results in Figure 5.11 show that the initial state had the lowest drain current of 0.51 mA at a drain voltage of 1.2 V. After tunnelling, the current increased to 1.4 mA, meaning there was a reduction in threshold voltage due to tunnelling. The injection results (*INJ55* and *10H*) overlap each other and have a highest current of 0.98 mA, which shows that threshold voltage was increased due to injection, as expected. Further tunnelling caused a very slight increase in current to achieve 1.01 mA. Further discussion is presented in Section 5.4.

Device 1

The measurements captured for devices 2, 3 and 4, as presented here, were not taken for device 1. Device 1 was found to be faulty before injection and tunnelling measurements, even though tunnelling and injection were also done for device 1, but there were no changes. The measurements for drain sweep of device 1 for different gate voltages are presented in Figure 5.12.

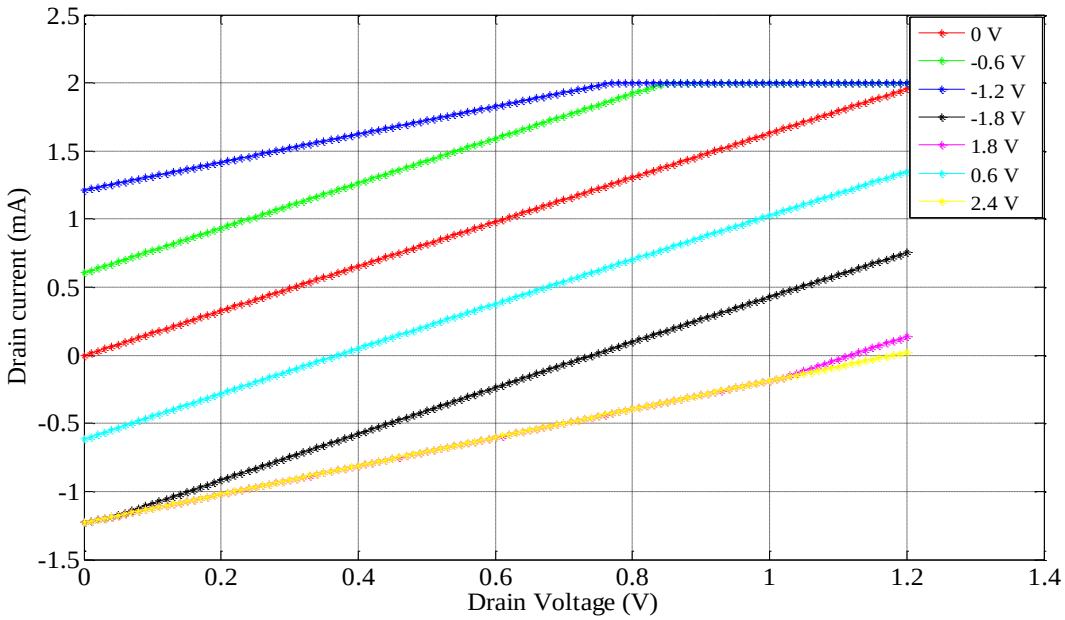


Figure 5.12. Drain voltage sweep of device 1 for different control gate voltages.

As expected, the drain current is at 0 mA for a control gate voltage of 0 V, as reflected in Figure 5.12. The relevance of the results presented is not related to the absolute applied control gate voltage, but rather the response of the drain current relative to the control gate voltage. Since the FG devices have a control gate terminal coupled through a control capacitor, it means that the DC signal is isolated. Thus, the drain sweep results should be the same irrespective of the control gate voltage applied. It was, however, clear that this was not the case and the inability of the device to be either tunnelled or injected proved that device 1 was faulty.

5.4 CONCLUSION

This chapter presented the practical measurement results for FG devices. The injection and tunnelling characteristics were measured with four different devices, which differed in width or area of input capacitance. The results showed that as the device width was reduced, the achievable output current was also reduced. Thus, when larger currents are required, a larger device width must be used. The effect is that more die area is taken up by these large devices, thus an area penalty is paid for high power applications.

The input capacitance affects the programming time of the devices and has less effect on the achievable output current. The capacitance effect on the output current is only visible at low bias voltages. A smaller input capacitance reduces the programming time for both injecting and tunnelling the FG devices. The technology scaling results in smaller devices and thus, smaller output currents. For highly scaled devices, current leakage is another concern for FG because of reduced oxide thickness. This decreases the retention time of FG devices, since charge cannot be maintained for longer periods. As formulated in Section 1.2, current limits and leakages are dependent variables used to evaluate the hypothesis. High current leakages reduce the reliability of FG devices and could thus prohibit the use of FG devices for mm-wave applications.

The maximum measured drain current was 3 mA for device 4, 2.7 mA for device 3 and 1.4 mA for device 2. Devices 3 and 4 have a gate length of 180 nm, while the gate length of devices 1 and 2 is 0.35 µm. Based on this, it was expected that devices 1 and 2 would achieve higher drain currents compared to devices 3 and 4. Although the results show device 2 to have a drain current lower than that of devices 3 and 4, it should not be translated or interpreted as the highest possible output current that can be achieved. The measurement results were all obtained under similar conditions, but devices with smaller widths have a lower amount of charge that can be transferred and therefore also require less programming time and lower programming voltages. It was due to the precautions taken during the measurements that an experiment with higher programming voltages and longer duration was not carried out to avoid damaging the devices, but owing to human error, some known and unknown precautions

were not accounted for. Eventually the devices were damaged after overheating experienced while soldering SMA connectors to enable VNA measurements.

It is also a challenge to design and simulate for FG devices, as the initial charge and consequently the threshold voltage are unknown and cannot be predicted accurately. It was also visible from the measurements that the initial charge was different for the fabricated devices. For devices 2 and 3, the initial charge resulted in the minimum drain current, but for device 4 the initial charge resulted in drain current that was between the injection and tunnelling states.

The FG devices were also measured 24 hours after either tunnelling or injection, although this was a short duration to determine retention capability as part of the desired parameters to address the research questions. There were some slight variations in certain cases, even in such a short period. This further highlights the reliability issues as emanating from the research questions in respect of long periods required for charge retention. The difference between the maximum currents measured for both injection and tunnelling represents the values used to determine the isolation and insertion losses. The ratio and difference between injection and tunnelling are thus important measures in addressing the research questions.

The frequency analyses were supposed to be measured using a VNA, which requires an SMA connector to be soldered to the devices. While soldering the SMA connectors to the devices, the devices were damaged (by the soldering temperature) and further measurements were impossible to capture with integrity. Although initial measurements for the VNA and baking were done, this was before the devices were found to be faulty and the measurements were consequently unreliable for further analysis.

Although it was possible to obtain measurements on CP devices, it was not possible to investigate the damaged or faulty devices experimentally. The main reason was that the experimental setup design did not allow the clock generator and CP device to be measured separately. The foundry maximum specified limits were exceeded (conditions attributed to human error) and this proved costly, as it resulted in damage to the devices. The state-of-the

art imposed limits and demanded advanced equipment for experimental setup and careful testing.

The dimensions of the fabricated devices are presented in Table 4-3. Device 4 has an area of 1.44 μm^2, device 3 has an area of 0.27 μm^2 and device 2 has an area of 0.7 μm^2. The ratio between the current measured after tunnelling (higher – ON state for switching) and after injection (lower – OFF state for switching) is 1.5 (1.5 mA and 1 mA), 1.35 (2.7 mA and 2 mA) and 6 (3 mA and 0.5 mA) for devices 2, 3 and 4, respectively. The ratio is higher for devices which have a larger area; this relates to the amount of charge that can be contained in the FG. The suitability of FG devices for switching applications depends on the ratio of the ON and OFF states. It was expected that the larger the FG area, the higher the ratio would be. To validate the hypothesis for switching applications further, it was also desired to characterise the associated losses for insertion and isolation. For this, device 4 was superior compared to devices 2 and 3, because device 4 achieved the highest current in the ON state and the lowest current in the OFF state. Device 3 displayed higher isolation losses due to higher OFF state current and device 2 displayed higher insertion losses due to lower ON state current.

The reliability issues are dominated by the oxide thickness of the transistor, which contributes to charge leakages: shorter transistor length causes more uncertainties. This was experienced for devices 3 and 4; the devices' length is 180 nm. For device 4, tunnelling was performed and measurements taken, but when the device was measured after 24 hours, similar measurements were not obtained. The device returned to the state it was in before tunnelling. After tunnelling the device with the same voltage, but for a longer period, the charge was maintained and similar results were achieved after 24 hours. For device 3, this was experienced after injection and obtaining similar results with 10 hours' difference between measurements, but it was only visible for the gate voltage sweep results. It can also be noted from the gate voltage sweep experimental results that the threshold voltage showed no variation for subsequent programming; the significant change only happened when the devices were programmed from their initial state. Even though this was the case, the drain voltage sweep measurements showed the actual effects expected for drain current changes after programming the devices. The conclusion based on this is that even though for the *p*-type transistor there is no change in

the threshold voltage, current does change. The behaviour of having no significant change in the threshold voltage, but a noticeable change in the drain current is interesting to observe and poses a challenge for further investigative work in future. The drain current changes correlate with the expected behaviour of the devices: tunnelling increases current and injection reduces current. The experimental results further raise the uncertainties caused by the oxide thickness and the unknown initial charge in the FG after fabrication. For all devices, the initial charge was different. The temperature tests or baking of the device was also done, but after discovering that the devices had been damaged (before baking), these results were void and are not presented here, as no conclusion can be drawn from the defective devices.

CHAPTER 6: CONCLUSION

6.1 INTRODUCTION

This chapter summarises the research findings based on the literature, simulations and experimental work completed in this dissertation. The research work was focused on evaluating the suitability of FG devices for mm-wave application and thus, characterising the FG devices. Although the scope of work covered may be relevant in a number of applications, switching application was of interest in this research.

The critical evaluation of the hypothesis is presented, as well as the challenges and limitations experienced while conducting this research. The results obtained were not complete to characterise the FG devices fully, but the results achieved suggested that FG devices could potentially be used in mm-wave for switching applications. Thus, the last section of the chapter suggested future work for improvements and further investigation.

6.2 CRITICAL HYPOTHESIS EVALUATION

The scientific research process presented in Figure 1.1 shows the scientific-community cycle and the hypothetico-deductive cycle, which when combined form a complete cycle for hypothesis validation/verification. The hypothetico-deductive cycle was completed using the research methodology as presented in Section 3.2 based on the literature study, mathematical analysis, CAD verification for simulations and IC prototyping for testing. There are two possible outcomes from this cycle: the hypothesis must either be redefined or consistency must be achieved. For the convenience of the reader, the hypothesis is repeated here:

If the current limits (or ratios) of the ON/OFF states are affected by device scaling required for mm-wave applications, then the EEPROMs' switching performance would not be suitable for mm-wave applications.

Further investigation into the research problem resulted in refining and thus addressing secondary research questions of the hypothesis. The secondary questions raised the issues of charge retention/leakages, CP output drive capability, FG mm-wave frequency behaviour and FG switching characteristics. The secondary research questions are presented in Section 1.2. The following are important aspects and outcomes of the deductive cycle:

- The literature study on different structures and alternatives for FG and CP devices is evaluated and discussed in chapter 2. The underlying theories in context and mathematical expressions are also presented.
- The circuit designs ("experiments"), together with CAD verification results, are presented in chapter 4 and the results of the prototypes are presented in chapter 5. The research methodology, experimental requirements/equipment and other resources used to develop prototypes are presented in chapter 3.
- The fundamental behaviour of FG devices is highly dependent on charge retention. It is desired to maintain the programmed state of the FG. Failure to achieve this is caused by charge leakages. Charge retention was tested over a window between 10 and 30 hours. In multiple testing, there was an instance where charge retention failed after programming a device. Most leakages occur very slowly and the effects can be identified after years. Other leakages occur as a result of failures in the oxide and some in response to conductions through the substrate. However, the failure that was observed experimentally was due to inadequate programming of the FG device. After sufficient programming time had been applied, the failure did not occur again after programming.
- FG devices require voltages higher than the supply voltages in order to apply either tunnelling or injection programming. Thus, CPs were evaluated for suitability to program FG devices successfully using the 130 nm process technology transistors. Different structures of CPs were discussed, but only the Dickson CP and voltage doubler CP were prototyped for comparison and further investigation. Initial measurements of the Dickson CP yielded results that correlated with what was expected, but further evaluation was prohibited by device failure, which resulted from exceeding the device voltage limits during the experimental measurements. Voltage

doubler measurements were below the minimum expected values by 50 % and no improvement was achieved under different experimental conditions.

- The difference between the programming states of the FG device determines the switching characteristics. The threshold voltage of FG is determined by the amount of charge in the FG. Devices with different lengths and widths were fabricated and evaluated. It was found that the larger the FG transistor area, the more charge in the FG could be transferred, hence the higher the ratio (of output drain current) between the extreme states of the FG.

After the hypothetico-deductive cycle, the scientific-community cycle was followed to confirm existing theories or identify new theories of the hypothesis. In cases of competing theories, a selection was made and justified for selection of a specific application; trade-offs were assessed and weighted. The following are important aspects, outcomes and underlying theories of the scientific-community cycle:

- Technology scaling is not linear with all the transistor parameters. Different parameters scale with different ratios; physical limits do also constrain the scaling of threshold voltage and the oxide thickness. As a result of non-linear scaling across different parameters, velocity saturation occurs and causes conditions where transistor square law design equations are no longer applicable to estimate the transistor behaviour fully. Thus, design equations that factor the effects of velocity saturation were used.
- In theory, ideal insertion and isolation losses of a transistor switch are zero, but in practice ideal conditions are never achieved. A non-ideal requirement is to minimise the losses by ensuring that there is minimum/negligible current flow in the isolation state and that there is minimum/negligible conduction losses during the insertion state. Although minimising losses is essential, in memory programming applications the ratio/difference between the extreme states is a more critical parameter than the actual values. However, for overall system performance and in applications where there are power constraint requirements, the actual values do also play a critical role and must be optimised.
- Measurements were conducted at GT with 0.35 µm process technology and the fabricated device was based on 130 nm process technology. Although an attempt was

made to change the transistor sizes of the 130 nm process technology to be equivalent to the 0.35 µm process technology (for comparison purposes), further investigation revealed that oxide thickness plays a critical role in the programming and reliability of FG devices. A direct comparison was not possible, as the oxide scales down with technology scaling and it was therefore different for the 0.35 µm and 130 nm process technologies. Increasing device sizes has the minimum effect on the actual behaviour based on oxide thickness, but it is effective to minimise process variation effects. Charge trapping in the oxide increases the probability of leakages and substrate conduction; the reliability issue of the oxide creates uncertainties in the behaviour of FG devices. Uncertainties are also introduced during the fabrication process and the initial charge in the FG after fabrication is consequently unknown.

Because of challenges and limitations, other theories were not evaluated or verified in this work; mm-wave frequency measurements, device lifetime testing and endurance tests, predictions and deductions are given, based on what was achieved in this work. Section 6.2 discusses the challenges and limitations encountered, while Section 6.3 discusses suggested future work for further investigation to verify other theories.

6.3 CHALLENGES AND LIMITATIONS

- The functional behaviour of an FG device depends on the initial charge, but the initial charge is unknown and the fabrication process has an effect on the final charge. In the FG devices measured, the initial charge was different for all four devices. It is therefore difficult to set up a simulation environment to predict FG behaviour. Another challenge is the fact that the FG does not have a DC path, so the simulation environment regards this as a floating node.
- The injection and tunnelling models are based on statistical modelling and require experimental extraction to determine the fit parameters. In view of the fabrication sponsorship process whereby one run is sponsored per year, it is costly to obtain multiple prototypes in different fabrication runs to provide for initial testing for fit parameters. Engineering judgement and decisions were based on literature findings and

estimated values were used for tunnelling and injection programming voltages and times.

- Because of the limited sponsored wafer size and the limit imposed by bondpads on the number of external connections from the IC, the clock driver and CPs circuits were integrated and fabricated as a combined system. A problem or limitation was experienced during debugging when the devices did not work according to expectations. Thus, it was impossible to determine without doubt whether the malfunctioning device was the clock driver or the CP circuit. Even for the fabricated devices, the number of samples provided was limited and this reduced the flexibility of experimenting above or to the extreme values, since no duplicate of any sample was available.
- High-frequency characteristics are greatly affected by connectors, cables used during measurements, interconnections, wirebonding and bondpads. Although the effects introduced by additional components can be modelled and de-embedded from the measured results, all models are complex and still do not guarantee 100 % accuracy. By using on-chip probing, the use of bondpads, interconnections, connectors and wirebonding can be eliminated or minimised. The unavailability of on-chip probing consequently imposed a limitation, as other secondary effects could have been eliminated.
- Other delimitations and key assumptions are discussed in Section 1.6.

6.4 SUGGESTED FUTURE WORK

- Design for testing: preparing for best measurement setup to enable and improve accuracy and integrity of the results. Saving chip area by including more circuits (or subsystems) may be costly, as certain measurement setups may not be conducted. For example, the bondpads spacing is determined by the probe sizes for on-chip probing and the space needed for PCB design wirebonding. Thus, a layout design developed with an end goal to enable measurements may overcome certain constraints during testing.
- If possible, different prototypes for DC and AC measurements must be developed, as the setup and equipment requirements vary. Other measurements may be compromised

by trying to have prototypes that accommodate both AC and DC measurements. For example, AC circuits may require DC blocking capacitors, which could prohibit DC measurements. An additional bondpad may be used to access the node before blocking the capacitor, but this could influence the AC measurements. For VNA measurements, matching and de-embedding circuits are included to optimise AC measurements. It is recommended that equipment limitations be taken into account during experimental setup design: for example, a parameter analyser and VNA cannot be used for either tunnelling or injection programming, as these items of equipment have limited current, which is insufficient for programming.

- Probability studies for tunnelling and injection are essential to improve on the programming efficiency. Unfortunately the fit parameters for modelling tunnelling and injection are determined experimentally for better accuracy. This may require two fabrication runs, where the first can be used to determine fit parameters and the second can then be developed based on the extracted parameters and improved models.
- Interconnections, connectors, cables and calibration of equipment must be studied thoroughly to develop de-embedding circuits and internal circuits and also to enable off-chip measurements that may affect results. The use of on-chip probing can greatly reduce the impact of external effects introduced by circuits and equipment used to enable off-chip measurements. The development of a test board must also minimise interferences introduced by its components during measurements.
- Although there is great uncertainty about the initial charge in the FG device, the window of uncertainty can be improved by integrating the FG models with the layout structure or design. Other uncertainties are due to process variations and have to be investigated further in simulations; corner simulations are typically used to evaluate the extreme values. The PDK provided by the foundry did not include models to support corner simulations, but Monte Carlo simulation models are provided as part of the PDK.
- Integrating CPs and clock generator circuits introduced a limitation for debugging the failure of the IC. If the CP was developed separately and an off-the-shelf clock generator used, it would have enabled further practical evaluation of the CP. The clock generator could also have been included as a separate sub-system to evaluate it separately.

- Even though there were limitations on having all practical measurements conducted to evaluate or confirm other theories, the results achieved in this work provided a positive framework, as the study highlighted that 130 nm process technology can be used to develop FG devices. This suggests that it is essential to undertake further studies related to this work to characterise and optimise the use of FG devices fully as switching components.
- A decision was made to assess the CP devices beyond the foundry-specified limits. This human error or misjudgement caused device failure. The device failure prevented capturing of other planned measurements required for further characterisation of the devices. The decision to exceed the limits was made while investigating the uncertainties of the measurements. In future, even though there are uncertainties in the measurement, it is recommended that all other planned measurements be conducted before experimenting or investigating the uncertainties that require measurements outside the initial experimental setup design.
- While soldering SMA connectors, the devices where unintentionally stressed to higher temperatures, which resulted in FG devices' failure. Nevertheless, this may not be regarded as an absolute human error, as soldering was not accounted for or anticipated, hence its effect was not known while conducting the measurements. As a recommendation, it is essential to test one device at a time when conducting additional experiments which involve conditions not accounted for in the initial experimental setup design. Unfortunately, in the case of this dissertation, all devices were soldered and the effects were determined when assessing results.

REFERENCES

[1] G. Avenier, M. Diop, P. Chevalier, G. Troillard, N. Loubet, J. Bouvier, L. Depoyan, N. Derrier, M. Buczko, C. Leyris, S. Boret, S. Montusclat, A. Margain, S. Pruvost, S. T. Nicolson, K. H. K. Yau, N. Revil, D. Gloria, D. Dutartre, S. P. Voinigescu and A. Chantre, "0.13 µm SiGe BiCMOS Technology Fully Dedicated to mm-Wave Applications, *IEEE Journal of Solid-State Circuits,* vol. 44, no. 9, pp. 2312-2321, September 2009.

[2] H. Zhang, Y. Zhao and A. Doboli, "A Scalable σ-Space Based Methodology for Modeling Process Parameter Variations in Analog Circuits," *Microelectronics Journal,* vol. 39, no. 12, pp. 1786-1796, December 2008.

[3] D. Sylvester, K. Agarwal and S. Shah, "Variability in Nanometer CMOS: Impact, Analysis, and Minimization" *Integration, The VLSI Journal,* vol. 41, no. 3, pp. 319-339, May 2008.

[4] T. Yao, M. Q. Gordon, K. K. W. Tang, H. H. K. Yao, M. Yang, P. Schvan and S. P. Voinigescu, "Algorithmic Design of CMOS LNAs and PAs for 60 GHz Radio," *IEEE Journal of Solid-State Circuits,* vol. 42, no. 5, pp. 1044-1057, May 2007.

[5] M. Racanelli and P. Kempf, "SiGe BiCMOS Technology for RF Circuit Applications," *IEEE Transactions on Electron Devices,* vol. 52, no. 7, pp. 1259-1270, July 2005.

[6] S. Seok, N. Rolland and P. A. Rolland, "A Novel Packaging Method Using Wafer-level BCB Polymer Bonding and Glass Wet-etching for RF Applications," *Sensors and Actuators A: Physical,* vol. 147, no. 2, pp. 677-682, October 2008.

[7] K. Lee, J. Chun and K. Kwon, "A Low Power CMOS Compatible Embedded EEPROM for Passive RFID Tag," *Microelectronics Journal,* vol. 41, no. 10, pp. 662-668, October 2010.

[8] A. Hajimiri, "mm-wave Silicon ICs: Challenges and Opportunities," *Proceedings: IEEE Custom Integrated Circuit Conference,* San Jose, 16-19 September 2007, pp. 741-747.

[9] D. Foty, "Next-Generation Wireless Networks: A New World, From Bottom to Top," *Proceedings: IEEE AFRICON,* Windhoek, 26-28 September 2007, pp. 1-9.

[10] A. Conte, G. L. Guidice, G. Palumbo and A. Signorello, "A High-Performance Very Low-Voltage Current Sense Amplifier for Nonvolatile Memories," *IEEE Journal of Solid-State Circuits,* vol. 40, no. 2, pp. 507-514, February 2005.

[11] J. R. Raguet, R. Laffont, R. Bouchakour, V. Bidal, A. Regnier and J.M. Mirabel, "New EEPROM Concept for Single Bit Operation," *Solid-State Electronics,* vol. 52, no. 10, pp. 1525-1529, October 2008.

[12] S. Atsumi, A. Umezawa, T. Tanzawa, T. Taura, H. Shiga, Y. Takano, T. Miyaba, M. Matsui, H. Watanabe, K. Isobe, S. Kitamura, S. Yamada, M. Saito, S. Mori and T. Watanabe, "A Channel-Erasing 1.8-V-Only 32-Mb NOR Flash EEPROM with a Bitline Direct Sensing Scheme," *IEEE Journal of Solid-State Circuit,* vol. 35, no. 11, pp. 1648-1654, November 2000.

[13] R. E. Barnett and L. Jin, "An EEPROM Programming Controller for Passive UHF RFID Transponders With Gate Clock Regulation Loop and Current Surge Control," *IEEE Journal Solid-State Circuits,* vol. 43, no. 8, pp. 1808-1815, August 2008.

[14] M. K. Zalalutdinov, J. D. Cross, J. W. Baldwin, B. R. Ilic, W. Zhou, B. H. Houston and J. M. Parpia, "CMOS-Integrated RF MEMS Resonators," *Journal of Microelectromechanical Systems,* vol. 19, no. 4, pp. 807-815, August 2010.

[15] T. S. Hall, C. M. Twingg, J. D. Gray, P. Hasler and D. V. Anderson, "Large-Scale Field-Programmable Analog Arrays for Analog Signal Processing," *IEEE Transactions on Circuits and Systems-I: Regular Paper,* vol. 52, no. 11, pp. 2298-2307, November 2005.

[16] G. K. Fedder, "Structured design of integrated MEMS," *Proceedings: IEEE Microelectromechanical Systems Conference*, Orlando, 17-21 January 1999, pp. 1-8.

[17] D. G. Crnkovic, "Theory of science," Mälardalen Real-Time Research Centre, Mälardalen University, Technical Report (ISRN) MDH-MRTC-64/2001-1-SE, September 2001.

[18] W. Chen, *The VLSI Handbook*, 2nd Ed. Taylor and Francis Group, LLC, 2007.

[19] B. C. Mabuza, S. Sinha and D. Foty, "On-chip Impedance Tuning for mm-wave Applications," *Proceedings: South African Conference on Semi and Superconductor Technology (SACSST),* Stellenbosch, 8-9 April 2009, pp. 49-53.

[20] A. Sebastiani, R. Piagge, A. Modelli and G. Ghindini, "High-K Dielectric for Inter-poly applications in Non Volatile Memories," *Microelectronics Reliability*, vol. 47, no. 4-7, pp. 598-601, April-May 2007.

[21] H. Rücker, B. Heinemann, W. Winkler, R. Barth, J. Borngräber, J. Drews, G. G. Fischer, A. Fox, T. Grabolla, U. Haak, D. Knoll, F. Korndörfer, A. Mai, S. Marschmeyer, P. Schley, D. Schmidt, J. Schmidt, M. A. Schubert, K, Schulz, B. Tillack, D. Wolansky and Y. Yamamoto, "A 0.13 µm SiGe BiCMOS Technology Featuring f_T/f_{max} of 240/330 GHz and Gate Delays Below 3 ps," *IEEE Journal of Solid-State Circuits,* vol. 45, no. 9, pp. 1678-1686, September 2010.

[22] S. T. Nicolson, K. H. K. Yau, S. Pruvost, V. Danelon, P. Chevalier, P. Garcia, A. Chantre, B. Sautreuil and S. P. Voinigescu, "A Low-Voltage SiGe BiCMOS 77-GHz Automotive Radar Chipset," *IEEE Transactions on Microwave Theory and Techniques,* vol. 56, no. 5, pp. 1092-1104, May 2008.

[23] C. Kothandaraman, S. K. Iyer and S. S. Iyer, "Electrically Programmable Fuse (eFuse) Using Electromigration in Silicides," *IEEE Electron Device Letters,* vol. 23, no. 9, pp. 523-525, September 2002.

[24] M. Alavi, M. Bohr, J. Hicks, M. Denham, A. Cassens, D. Douglas and M. Tsai, "A PROM Element Based on Salicide Agglomeration of Poly Fuses in a CMOS Logic Process," *Proceedings: IEEE International Electron Devices Meeting*, Washington DC, 7-12 December 1997, pp. 855-858.

[25] E. G. Gebreselasie, S. H. Voldaman, Z. X. He, D. Coolbaugh, R. M. Rassel, T. Kirihata, A. Paganini, C. G. Cox, S. A. S. Onge, J. S. Dunn, R. E. Halbach and J. M. Lukaitis, "Electrically Programmable Fuses for Analog and Mixed Signal Applications in Silicon Germanium BiCMOS Technologies," *Proceedings: IEEE Bipolar/BiCMOS Circuits and Technology Meeting,* Boston, 30 September - 2 October 2007, pp. 238-241.

[26] D. Maier-Schneider, S. Kolb, B. Winkler and W. M. Werner, "Novel Surface-Micromachined Low-Power Fuses for On-Chip Calibration," *Sensors and Actuators. A, Physical,* vol. 97-98, pp. 173-178, April 2002.

[27] W. R. Tonti, J. A. Fifield, J. Higgins, W. H. Guthrie, W. Berry and C. Narayan, "Product Specific Sub-Micron e-Fuse Reliability and Design Qualification," Proceedings: *IEEE International Integrated Reliability Workshop Final Report at IBM Eng and Technol. Service,* S. Late Tahoe, 20 – 23 October 2003, pp. 36-40.

[28] T. Vähä-Heikkilä, "MEMS Tuning and Matching Circuits, and Millimeter Wave On-Wafer Measurements", *VTT Publications,* pp. 1-86, 2006.

[29] A. Di Bartolomeo, H. Rücker, P. Schley, A. Fox, S. Lischke and K. Na, "A Single-Poly EEPROM Cell for Embedded Memory Applications," *Solid-State Electronics,* vol. 53, no. 6, pp. 644-648, June 2009.

[30] C. Lin and C. Sun, "A Single-Poly EEPROM Cell Structure Compatible to Standard CMOS Process," *Solid-State Electronics*, vol. 51, no. 6, pp. 888-893, May 2007.

[31] H. Sung, T. F. Lei, T. Hsu, S. W. Wang, Y. Kao, Y. Lin and C. S. Wang, "Novel Single-Poly EEPROM with Damascene Control-Gate Structure," *IEEE Electron Device Letters*, vol. 26, no. 10, pp. 770-772, October 2005.

[32] J. M. Rabaey, A. Chandrakasan and B. Nikolic., "The CMOS Inverter", *Digital Integrated Circuits: A Design Perspective*, 2nd Ed. Upper Saddle River, 2003.

[33] P. Pavan, R. Bez, P. Olovo and E. Zanoni, "Flash Memory Cells – an Overview," *Proceedings of the IEEE*, vol. 85, no. 8, pp. 1248-1271, August 1997.

[34] A. Regnier, R. Laffont, R. Bouchakour and J. M. Mirabel, "A New Architecture of EEPROM for High Density and High Reliability Application," *Proceedings: IEEE Non-Volatile Memory Technology Symposium*, Marseille, 15-17 November 2004, pp. 143-148.

[35] R. Bouchakour, N. Harabech, P. Canet, J. M. Mirabel, Ph. Boivin and O. Pizzuto, "A New Physical-Based Compact Model of Floating-Gate EEPROM Cells," *Journal of Non-Crystalline Solid*, vol. 280, no. 1-3, pp. 122-126, February 2001.

[36] Y. L. Chiou, J. P. Gambino and M. Mohammad, "Determination of the Fowler-Nordheim Tunnelling Parameters from the Fowler-Nordheim Plot," *Solid-State Electronics*, vol. 45, no. 10, pp. 1787-1791, October 2001.

[37] Ih. Chen and C. W. Teng, "A Quantitative Physical Model for the Band-to-Band Tunnelling-Induced Substrate Hot Electron Injection in MOS Devices," *Proceedings: IEEE Transaction on Electron Devices*, vol. 39, no. 7, pp. 1646-1651, July 1992.

[38] G. Palumbo, D. Pappalardo and M. Gaibotti, "Charge-Pump Circuits: Power-Consumption Optimization," *IEEE Transaction on Circuits and Systems I: Fundamental Theory and Applications*, vol. 49, no. 11, pp. 1535-1542, November 2002.

[39] D. Baderna, A. Cabrini, M. Pasotti and G. Torelli, "Power Efficiency Evaluation in Dickson and Voltage Doubler Charge Pump Topologies," *Microelectronics Journal*, vol. 37, no. 10, pp. 1128-1135, October 2006.

[40] J. Wu and K. Chang, "MOS Charge Pumps for Low-Voltage Operation," *IEEE Journal of Solid-State Circuits*, vol. 33, no. 4, pp. 592-597, April 1998.

[41] A. Kurokawa, T, Sato, T. Kanamoto and M. Hashimoto, "Interconnect Modeling: A Physical Design Perspective," *IEEE Transactions on Electron Devices*, vol. 56, no. 9, pp. 1840-1851, September 2009.

[42] H. Aziza, E. Bergeret, A. Perez and J-M. Portal, "An Efficient Model to Evaluate the Impact of Design Parameters on Charge Pump Circuits' Performances: Application on RFID Circuits," *Proceedings: International Conference on Design and Technology of Integrated System in Nanoscale Era*, Rabat, 2-7 September 2007, pp. 108-113.

[43] D. Lui, X. Zou, F. Zhang and M. Deng, "Circuits and Implementation of a Low-Power Embedded EEPROM Memory," *Proceedings: 8th International Conference on Solid-State and Integrated Circuits Technology*, Shanghai, 23-26 October 2006, pp. 836-838.

[44] H. A. C. Tilmans, "MEMS Components for Wireless Communications," *Proceedings: EURO Sensors*, Prague, 15-18 September 2002, pp. 1-34 (on CD-ROM only).

[45] X. Wang, D. Wu, F. Qiao, P. Zhu, K. Li, L. Pan and R. Zhou, "A high efficiency CMOS charge pump for low voltage operation," *Proceedings: IEEE 8th International Conference on ASIC*, Changsha, 20-23 October 2009, pp. 320-323.

[46] S. S. Chung, S. T. Liaw, C. M. Yih, Z. H. Ho, C. J. Lin, D. S. Kuo and M. S. Liang, "N-channel Versus P-channel Flash EEPROM - Which One has Better Reliabilities," *Proceedings: IEEE International Reliability Physics Symposium*, Orlando, 30 April – 3 May 2001, pp. 67-72.

[47] R. J. Richards and H. J. De Los Santos, "MEMS for RF/Microwave Wireless Applications: The Next Wave," *Microwave Journal*, vol. 44, no. 3, pp. 20,24,28,32-41, March 2001.

[48] K. Rahimi, C. Diorio, C. Hernandez and M. D. Brockhausen, "A Simulation Model for Floating-Gate MOS Synapse Transistors," *Proceedings: IEEE International Symposium on Circuits and Systems*, Phoenix-Scottsdale, 26-29 May, 2002, pp 532-535.

[49] J. Gray, R. Robucci and P. Hasler, "The Design and Simulation Model of an Analog Floating-Gate Computational Element for use in Large-Scale Analog Reconfigurable Systems," *Proceedings: IEEE International Symposium on Circuits and Systems*, Knoxville, 10-13 August, 2008, pp 253-256.

[50] S. A. Bhalerao, A. V. Chaudhary and R. M. Patrikar, "A CMOS Low Voltage Charge Pump" *Proceedings: International Conference on VLSI design*, Bangalore, 6-7 January, 2007, pp. 941-946.

[51] R. Chawla, F. Adil, G. Serrano and P. E. Hasler, "Programmable G_m-*C* Filters Using Floating-Gate Operational Transconductances Amplifiers," *IEEE Transactions on Circuits and Systems-I: Regular Paper*, vol. 54, no. 3, pp. 481-491, March 2007.

[52] P. C. Brady, H. Lo, G. J. Serrano, F. Adil, M. R. Kucic, P. E. Hasler, D. V. Anderson, A. W. Pereira, "Floating-Gate Reference Circuit," US Patent No. 7 034 603, April 25, 2006.

APPENDIX A: GT MEASUREMENTS

A.1 TUNNELLING

The work presented here is for measurements done at GT. Samples that were provided by GT and fabricated were used for experimental purposes and to acquire experience with measurements relating to FG devices. The design models, layouts and the PCB designs remain the intellectual property of GT, refer to [15] and [51]-[52] for GT research work. Based on NDA with GT, the results provided here are not supposed to be published as the author's primary work.

A.1.1 Device 8

Tunnelling the devices increases the threshold voltage and thus reduces drain current. Thus, after tunnelling the devices, the measurements represent the minimum current that has been achieved for a particular tunnelling condition. All the results were measured using gate sweep from 0 to 2.4 V, which is limited by the TSMC process technology, 0.35 µm [15], and the value was chosen to be below the limit of over-stressing the devices used during the measurements.

The results in Figure A.1 were measured for device 8. The tunnelling voltage was varied between 0 and 5 V with an increment of 0.5 s. As the tunnelling time is increased, the initial current, when gate voltage is zero, is reduced. A logarithmic presentation is used to determine the changes in the sub-threshold and also to measure the range of the sub-threshold. The tunnelling voltage used was 8.5 V.

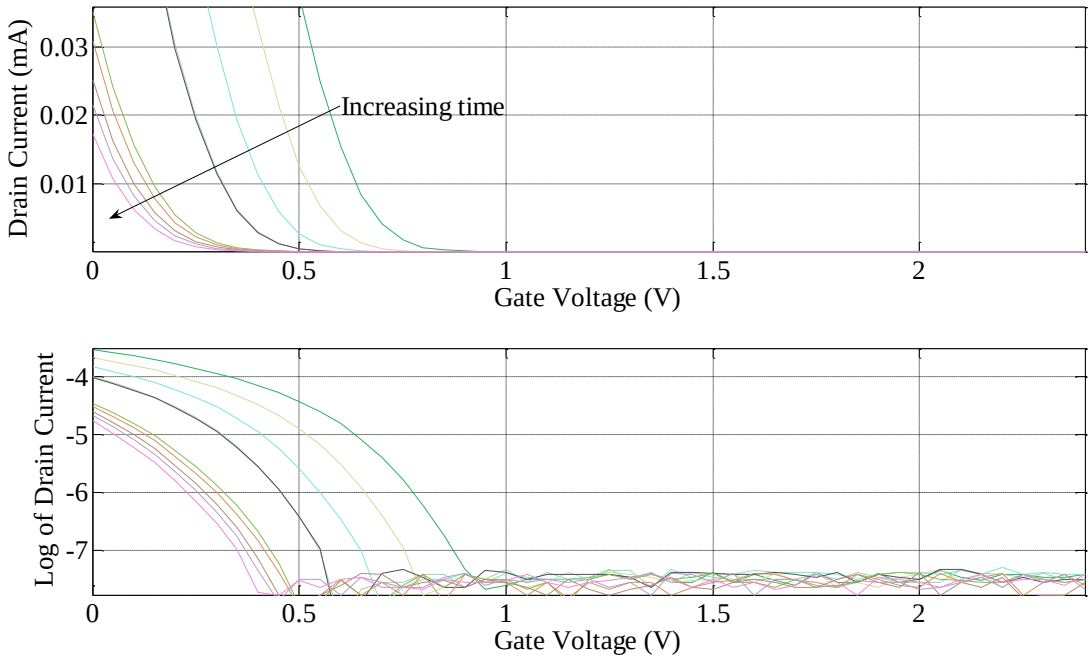


Figure A.1. Gate voltage sweep of device 8 after tunnelling voltage with 8.5 V. The tunnelling time was varied from 0 to 5 s incremented by 0.5 s. The graph illustrates the effects of the tunnelling duration.

The range of drain current in Figure A.1 for different tunnelling times is between 16 and > 35 µA. It was impossible to measure the actual maximum value because of the asymptotic behaviour of the results. As can be seen in Figure A.1, as the gate voltage is reduced the current increases exponentially. The logarithmic range is between |3.5| and |4.75|, with a window of 1.25. The sub-threshold voltage for the two extreme cases, minimum and maximum currents, is 0.4 and 0.7 V, respectively. Thus, in this case the device performs in a similar manner as when tunnelling was not done, as a result of short tunnelling time.

A.1.2 Device 10

The results in Figure A.2 were measured for device 10. The tunnelling voltage was varied between 0 and 1 s with an increment of 0.1 s. As the tunnelling time increased, the initial current, where the gate voltage was zero, was reduced.

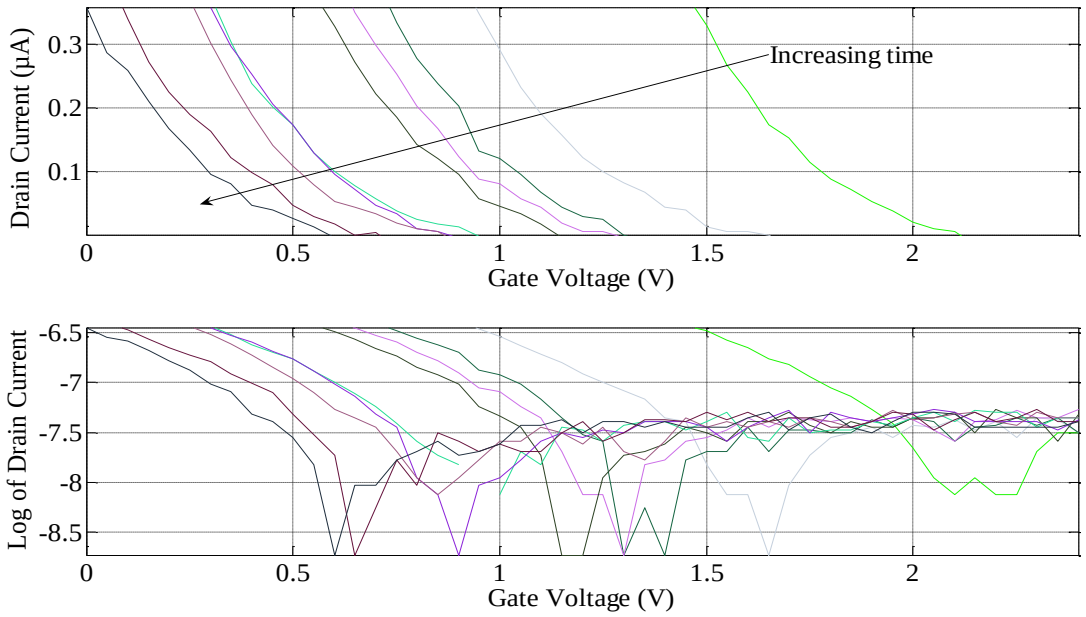


Figure A.2. Gate voltage sweep of device 10 after tunnelling voltage with 6.5 V. The tunnelling time was varied from 0 to 1 s incremented by 0.1 s. The graph illustrates the effects of the tunnelling duration.

The range of drain current in Figure A.2 for different tunnelling times is between 0.35 and > 0.9 µA. It was impossible to measure the actual maximum value; as can be seen in Figure A.2, as the gate voltage was reduced the current increased exponentially. The logarithmic range is between |4.6| and |6.5|, with a window of 1.9. The sub-threshold voltage for the two extreme cases, minimum and maximum currents, is 0.6 and 2 V, respectively. For limited tunnelling time, the device operates in the ON state.

A.1.3 Device 71

The results in Figure A.3 were measured for device 71. Tunnelling voltage was varied between 0 and 1 s with an increment of 0.1 s. As the tunnelling time increased, the initial current, when the gate voltage was zero, was reduced.

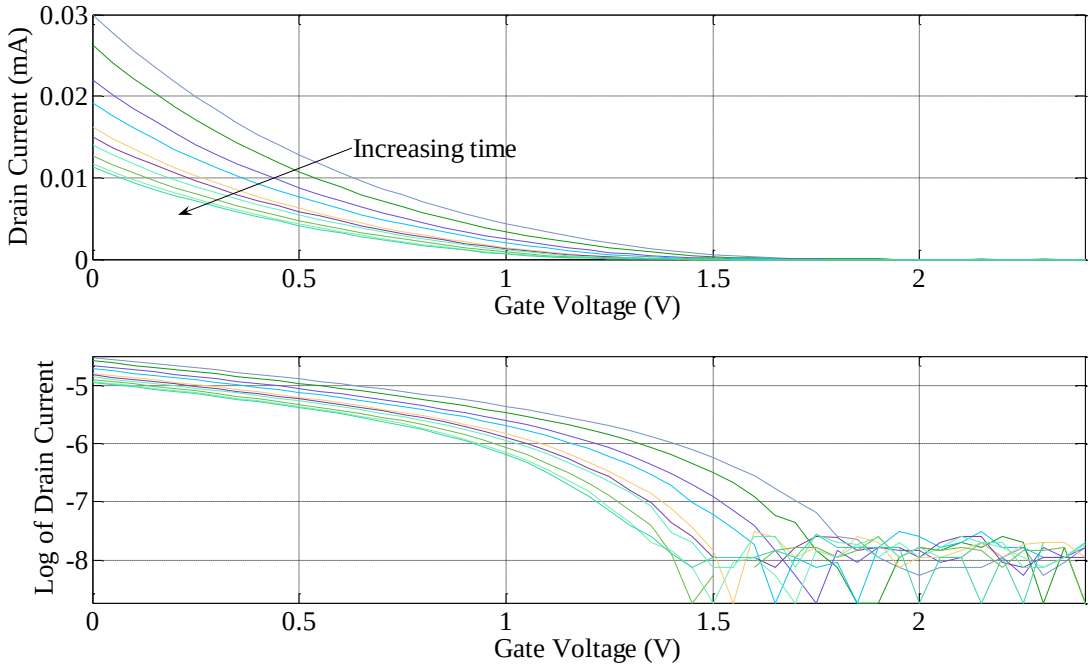


Figure A.3. Gate voltage sweep of device 71 after tunnelling voltage with 7.5 V. The tunnelling time was varied from 0 to 1 s incremented by 0.1 s. The graph illustrates the effects of the tunnelling duration.

The range of drain current in Figure A.3 for different tunnelling times is between 12 and 30 µA. The logarithmic range is between |4.5| and |5|, with a window of 0.5. The sub-threshold voltages for the two extreme cases, minimum and maximum currents, are 1.4 and 1.7 V, respectively.

A.1.4 Device 73

The results in Figure A.4 were measured for device 73. The tunnelling voltage was varied with an increment of 0.5 s between 0 and 5 s. As the tunnelling time increased, the initial current, were gate voltage was zero, was reduced.

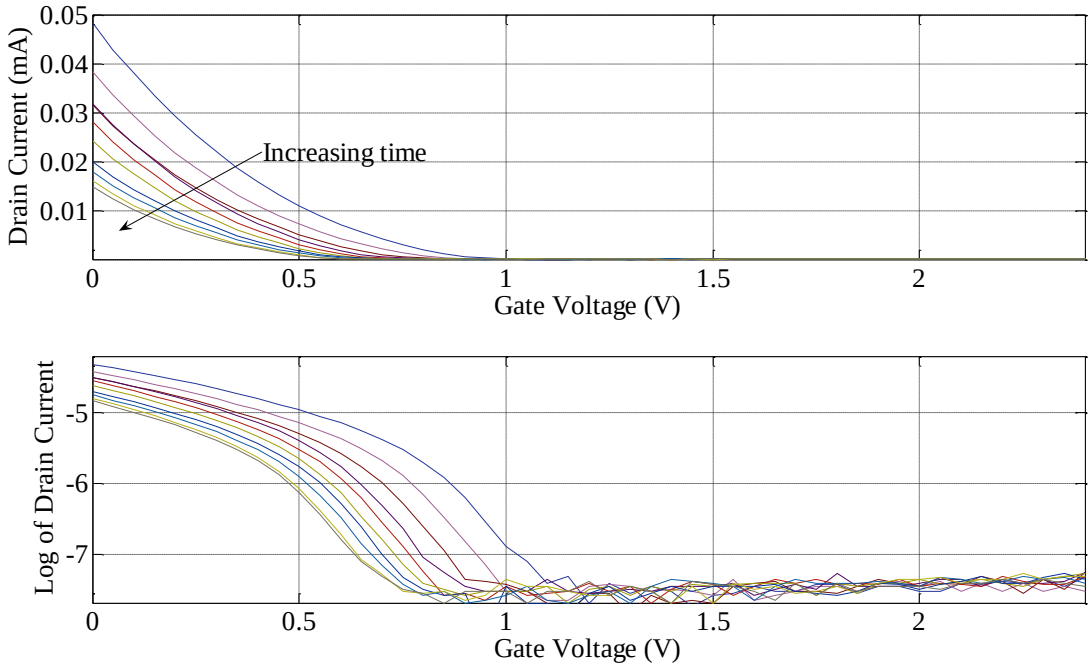


Figure A.4. Gate voltage sweep of device 73 after tunnelling voltage with 8.5 V. The tunnelling time was varied from 0 to 5 s incremented by 0.5 s. The graph illustrates the effects of the tunnelling duration.

The range of drain current in Figure A.4 for different tunnelling times is between 13 and 48 µA. The logarithmic range is between |4.3| and |4.7|, with a window of 0.4. The sub-threshold voltages for the two extreme cases, minimum and maximum currents, are 0.6 and 0.9 V, respectively.

A.2 INJECTION

A.2.1 Device 8

The injection of FG devices reduces the threshold voltage and the transistor can therefore easily be turned ON. Unlike tunnelling, which has only tunnelling voltage and time as variables for measurement, injection depends on the source-to–drain voltage, the gate voltage and time. The source-to-drain voltages used are 5, 6 and 7 V, while the gate voltages used are 1.6 and 3.2 V. The injection time was varied over the same range for all devices, thus the performance is evaluated based on the variation of drain-to-source and gate voltages. For each measurement condition, the injection time is increased logarithmically; as the injection time

increases the drain current also increases. The results for 5 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.5

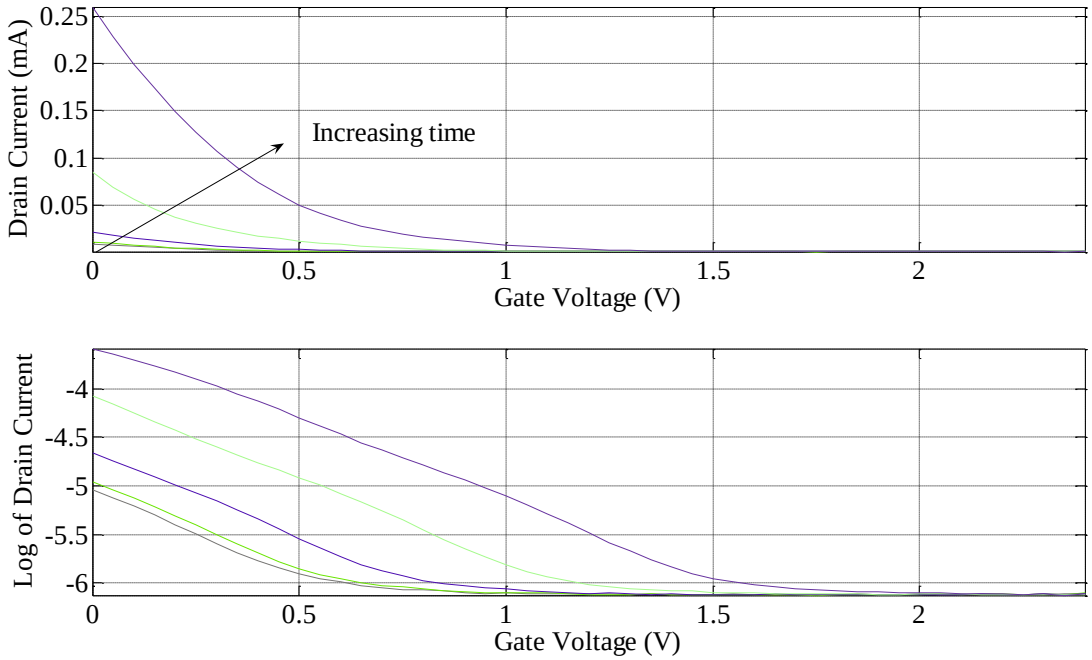


Figure A.5. Gate voltage sweep of device 8 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 5 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.5 the minimum and maximum drain currents are 20 and 250 µA, respectively. The effective gate sweep range for minimum current is between 0.6 and 1.5 V. The results for 5 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.6.

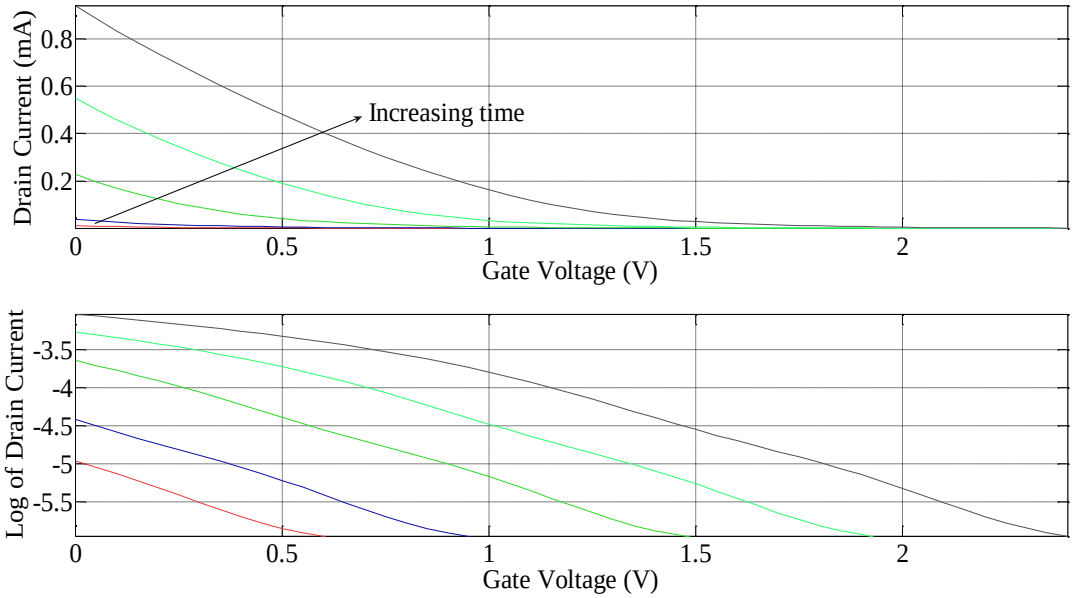


Figure A.6. Gate voltage sweep of device 8 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 5 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 20 and 900 µA, respectively. The effective gate sweep range for minimum current is between 0.6 and 2.4 V; as shown in Figure A.6. The results in Figure A.5 and Figure A.6 are based on 5 V for drain–to-source voltage. The results for 6 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.7.

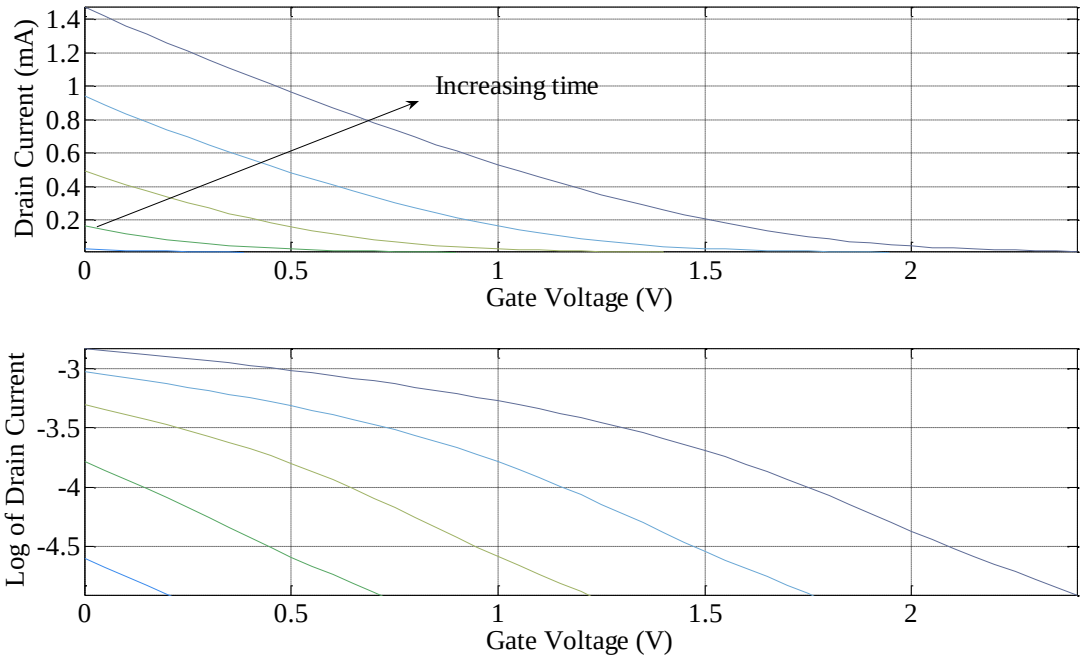


Figure A.7. Gate voltage sweep of device 8 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 6 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.7 the minimum and maximum drain currents are 20 µA and 1.42 mA, respectively. The effective gate sweep range for minimum current is between 0.2 and 2.4 V. The drain current has increased substantially by increasing the drain-to-source voltage. This dependency and the effects of FG device sizes are discussed at the end of this section. The results for 6 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.8.

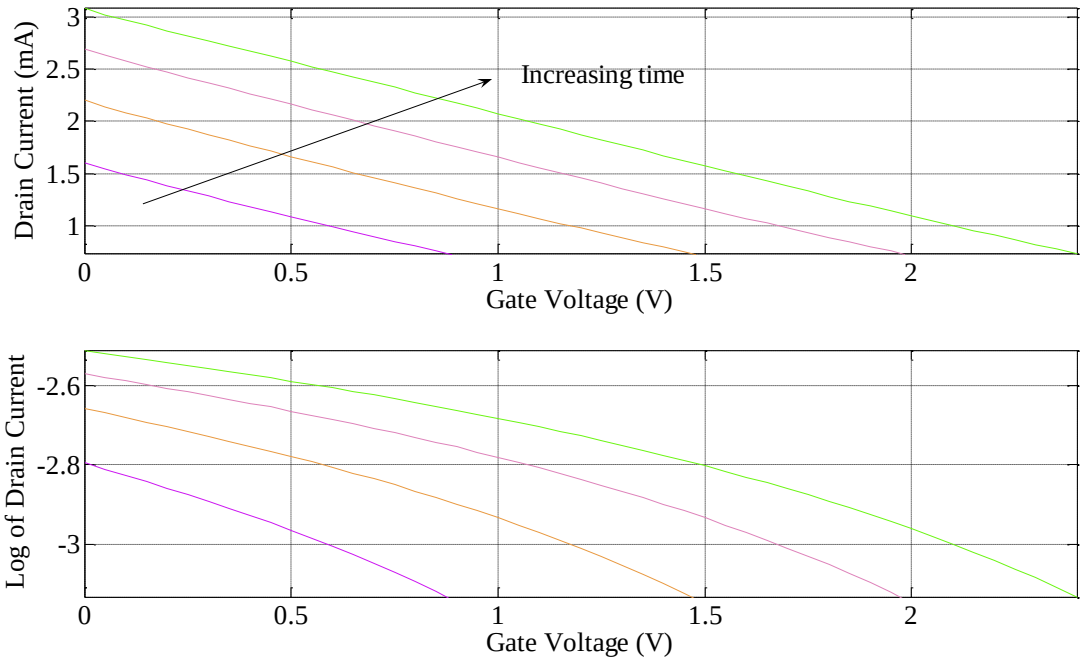


Figure A.8. Gate voltage sweep of device 8 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 6 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 1.6 mA and 3.1 mA, respectively. The effective gate sweep range for minimum current is between 0.2 and 2.4 V; as shown in Figure A.8. The results for 7 V drain–to-source voltage with 1.6 V gate voltage are presented in Figure A.9.

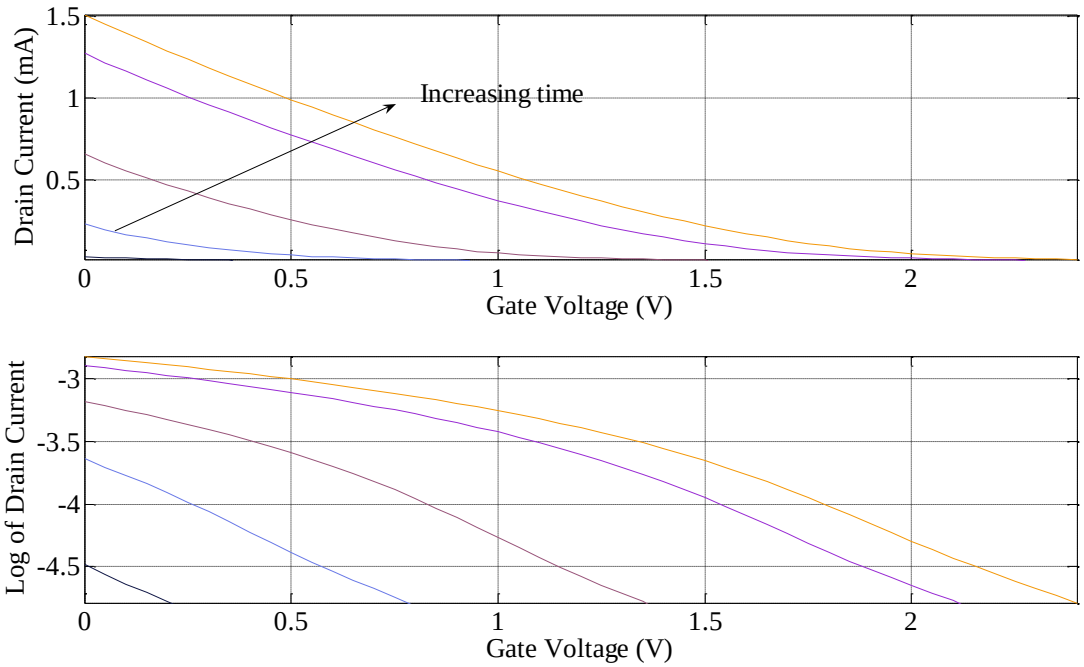


Figure A.9. Gate voltage sweep of device 8 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 7 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.9 the minimum and maximum drain currents are 200 µA and 1.5 mA, respectively. The effective gate sweep range for minimum current is between 0.2 and 2.4 V. The results for 7 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.10.

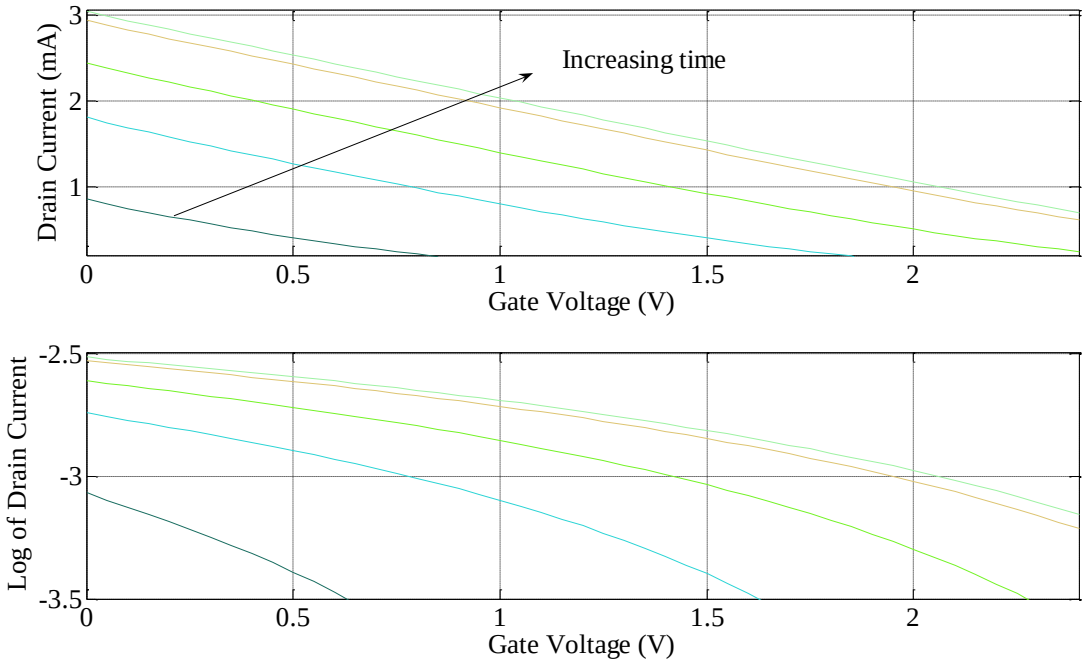


Figure A.10. Gate voltage sweep of device 8 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 7 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 900 µA and 3.1 mA, respectively. The effective gate sweep range for minimum current is between 0.6 and 2.4 V; as shown in Figure A.10. The next section presents the results for the other three devices used for measurements. The last section evaluates the results achieved and identifies the difference in performance based on the testing conditions.

A.2.2 Device 10

The results for 5 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.11.

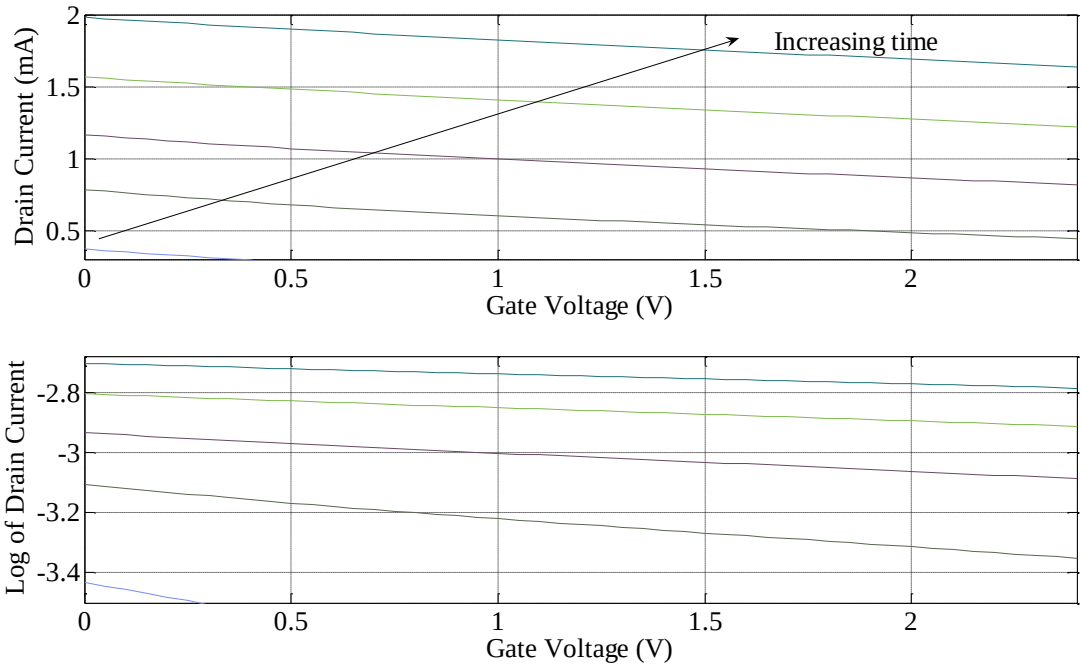


Figure A.11. Gate voltage sweep of device 10 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 5 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.11 the minimum and maximum drain currents are 380 µA and 2 mA, respectively. The effective gate sweep range for minimum current is between 0.3 and 2.4 V. The results for 5 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.12.

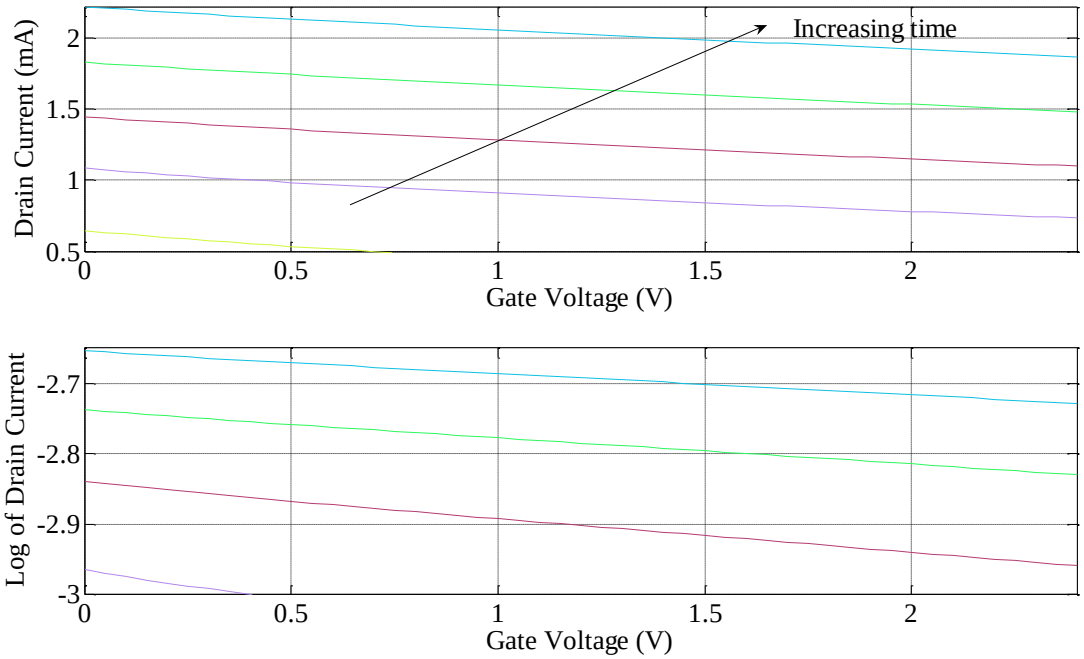


Figure A.12. Gate voltage sweep of device 10 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 5 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 600 µA and 2.2 mA, respectively. The effective gate sweep range for minimum current is between 0.4 and 2.4 V; as shown in Figure A.12. The results for 6 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.13.

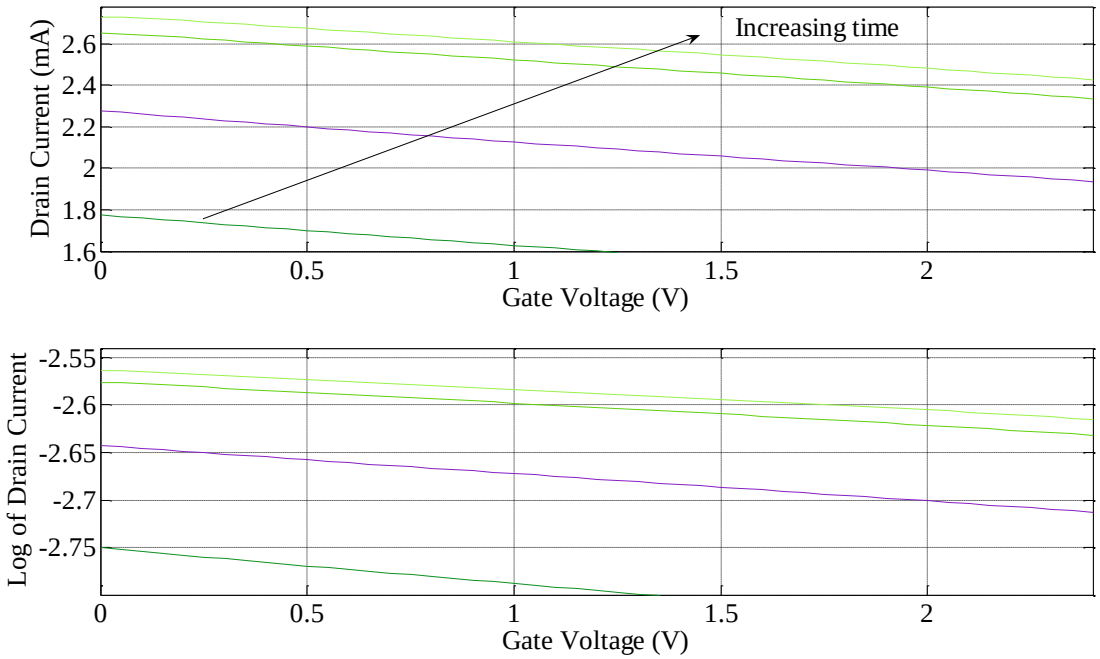


Figure A.13. Gate voltage sweep of device 10 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 6 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.13 the minimum and maximum drain currents are 1.78 mA and 2.75 mA, respectively. The effective gate sweep range for minimum current is between 1.3 and 2.4 V. The results for 6 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.14.

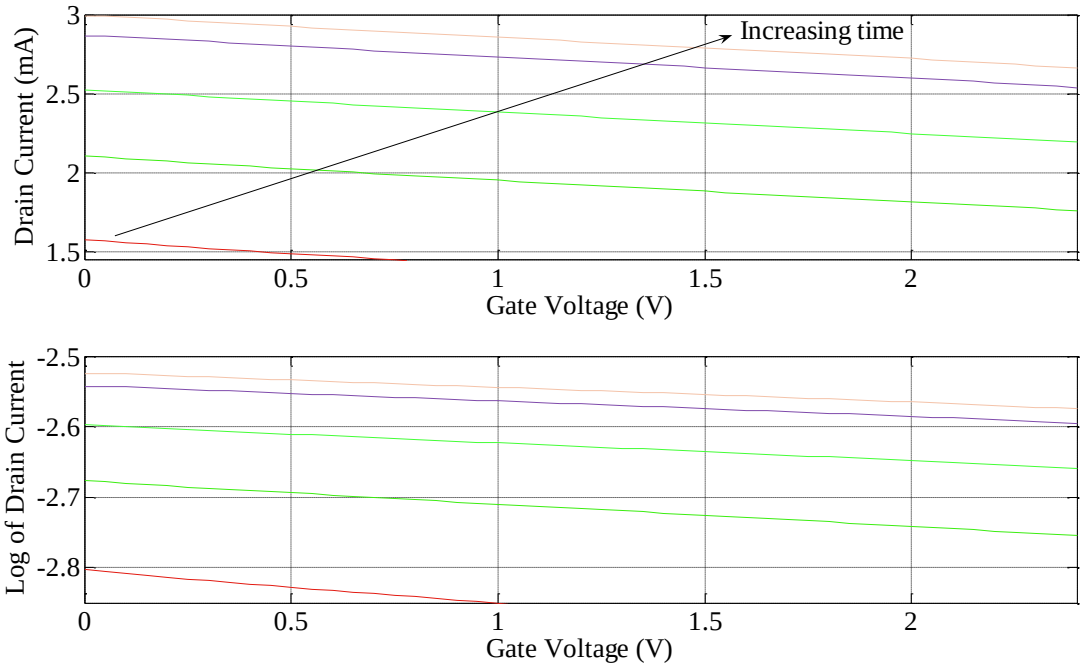


Figure A.14. Gate voltage sweep of device 10 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 6 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 1.6 mA and 3 mA, respectively. The effective gate sweep range for minimum current is between 0.6 and 2.4 V; as shown in Figure A.14. The results for 7 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.15.

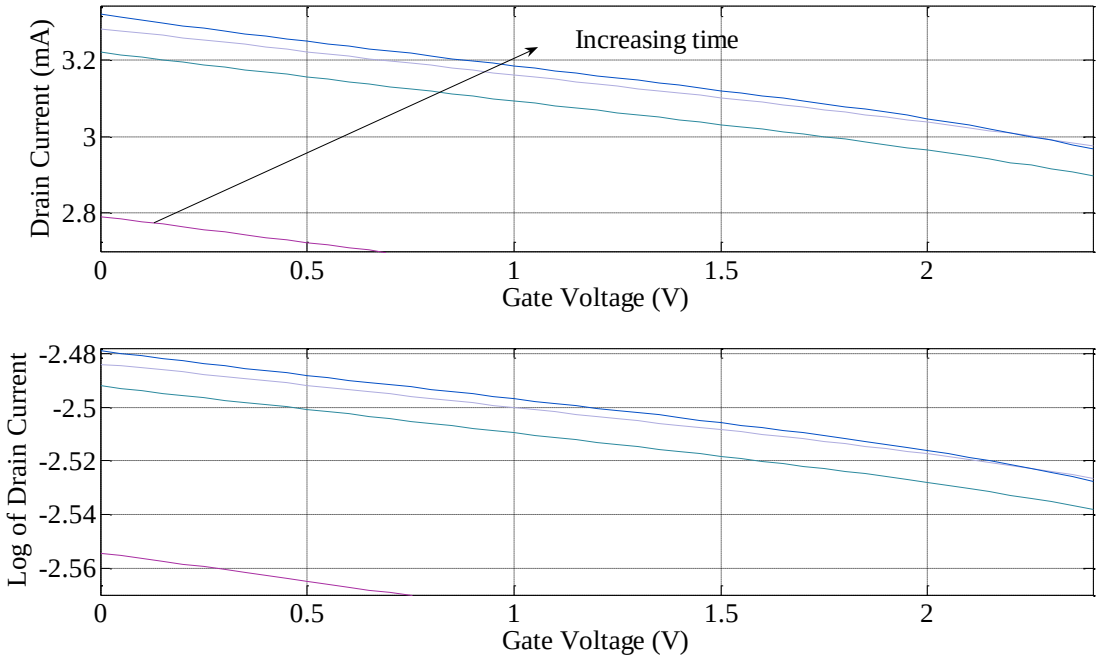


Figure A.15. Gate voltage sweep of device 10 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 7 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.15 the minimum and maximum drain currents are 2.78 mA and 3.34 mA, respectively. The effective gate sweep range for minimum current is between 0.75 and 2.4 V. The results for 7 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.16.

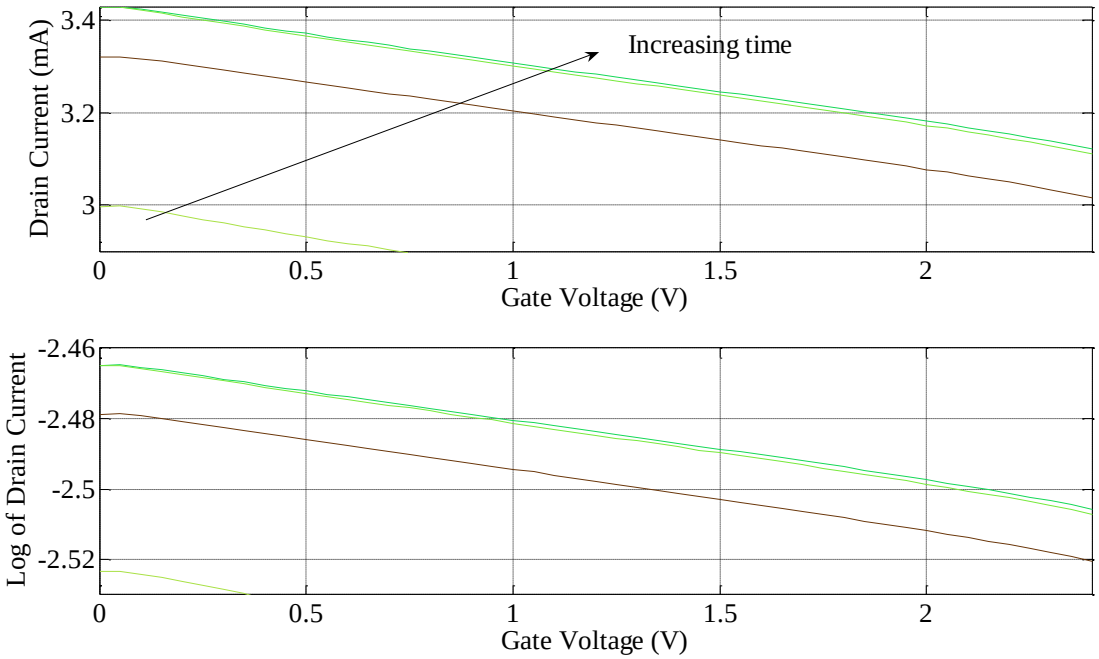


Figure A.16. Gate voltage sweep of device 10 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 7 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 3 mA and 3.42 mA, respectively. The effective gate sweep range for minimum current is between 0.3 and 2.4 V; as shown in Figure A.16. In the results presented for devices 8 and 10, as the drain-to-source voltage and gate voltage increase, the drain current at zero gate sweep voltage increases. Devices 8 and 10 have a width of 10 µm and the next two devices (71 and 73) have a width of 2 µm. The difference between devices 8 and 10 is the control capacitance (device 8 has a length and width of 10 µm, while for device 10 the length and width are 1 µm).

A.2.3 Device 71

The results for 5 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.17.

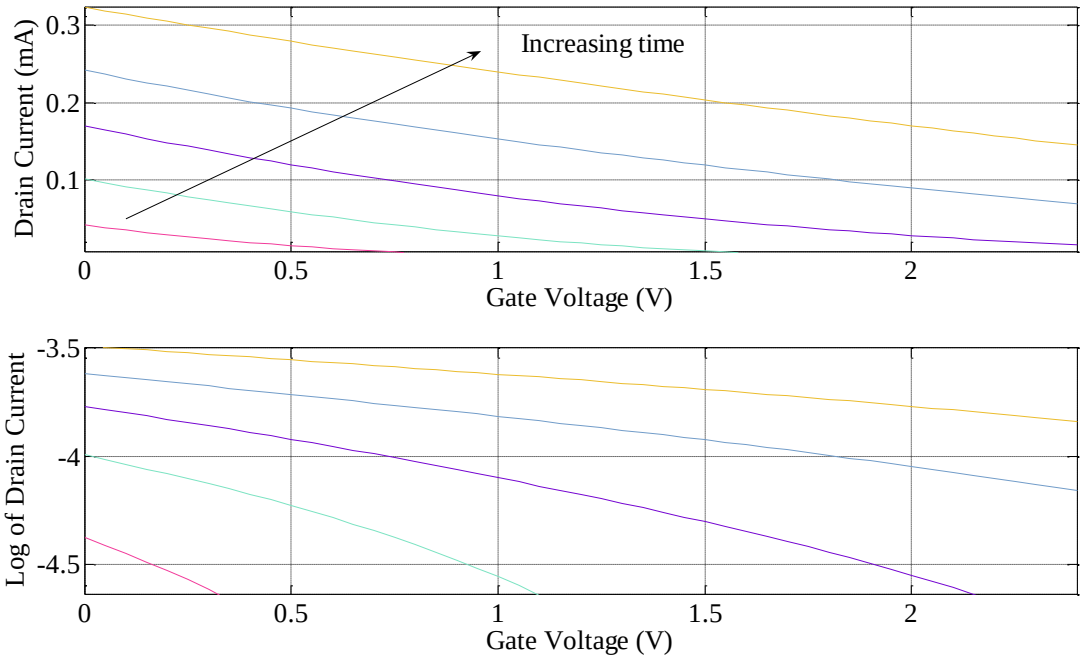


Figure A.17. Gate voltage sweep of device 71 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 5 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.17, the minimum and maximum drain currents are 50 µA and 300 µA, respectively. The effective gate sweep range for minimum current is between 0.3 and 2.4 V. The results for 5 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.18

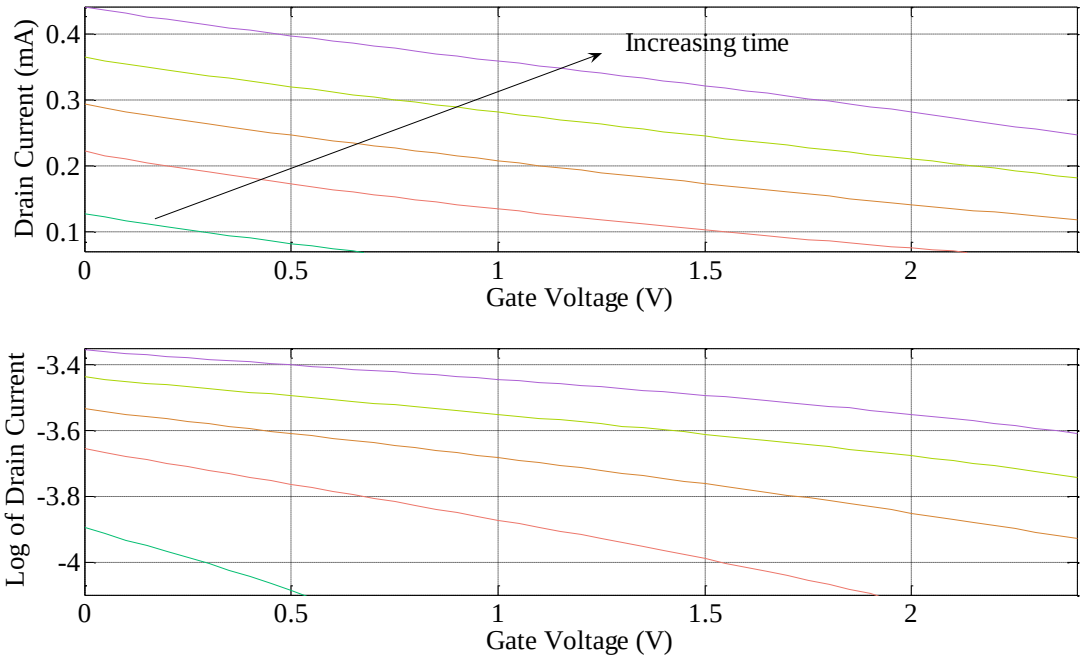


Figure A.18. Gate voltage sweep of device 71 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 5 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 130 µA and 500 µA, respectively. The effective gate sweep range for minimum current is between 0.52 and 2.4 V; as shown in Figure A.18. The results for 6 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.19.

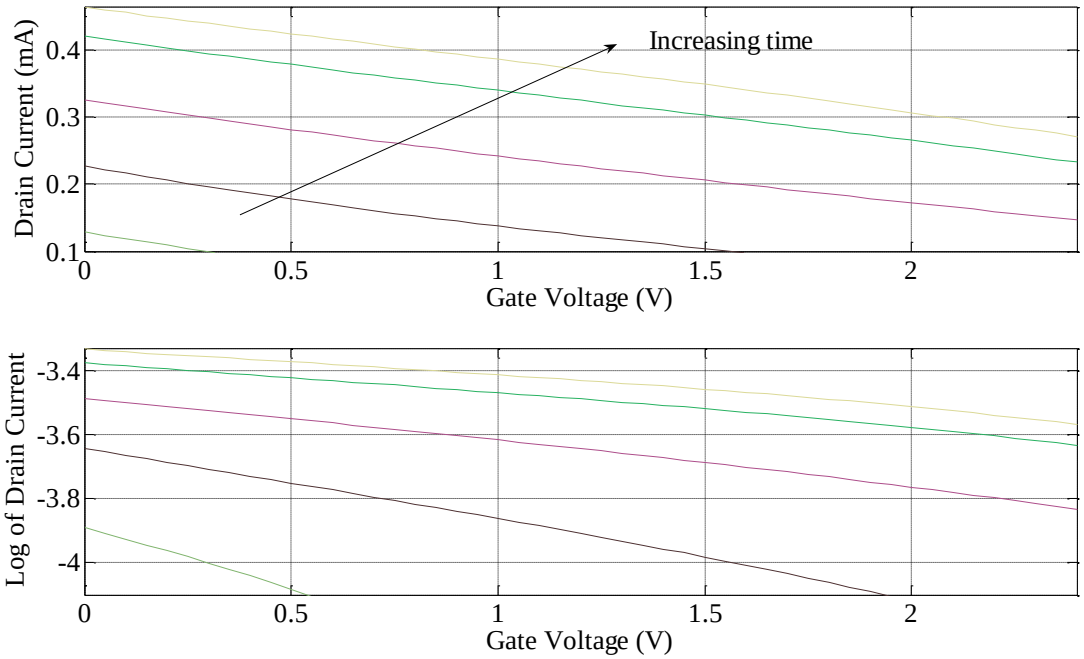


Figure A.19. Gate voltage sweep of device 71 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 6 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.19 the minimum and maximum drain currents are 130 µA and 460 µA, respectively. The effective gate sweep range for minimum current is between 0.6 and 2.4 V. The results for 6 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.20.

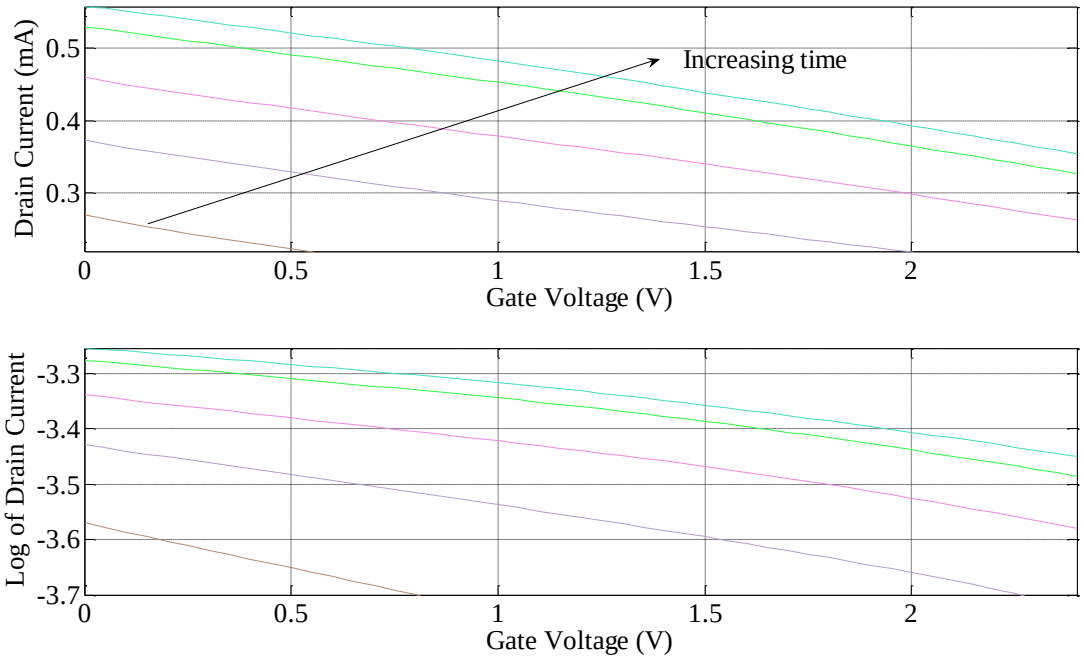


Figure A.20. Gate voltage sweep of device 71 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 6 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 280 µA and 580 µA, respectively. The effective gate sweep range for minimum current is between 0.8 and 2.4 V; as shown in Figure A.20. The results for 7 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.21.

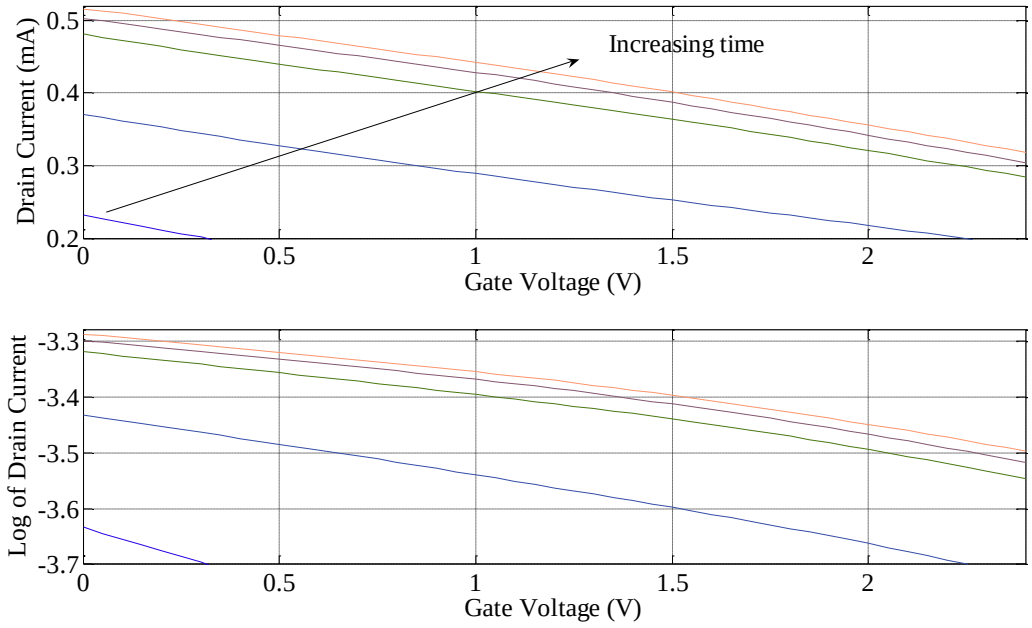


Figure A.21. Gate voltage sweep of device 71 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 7 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.21 the minimum and maximum drain currents are 240 µA and 530 µA, respectively. The effective gate sweep range for minimum current is between 0.3 and 2.4 V. The results for 7 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.22.

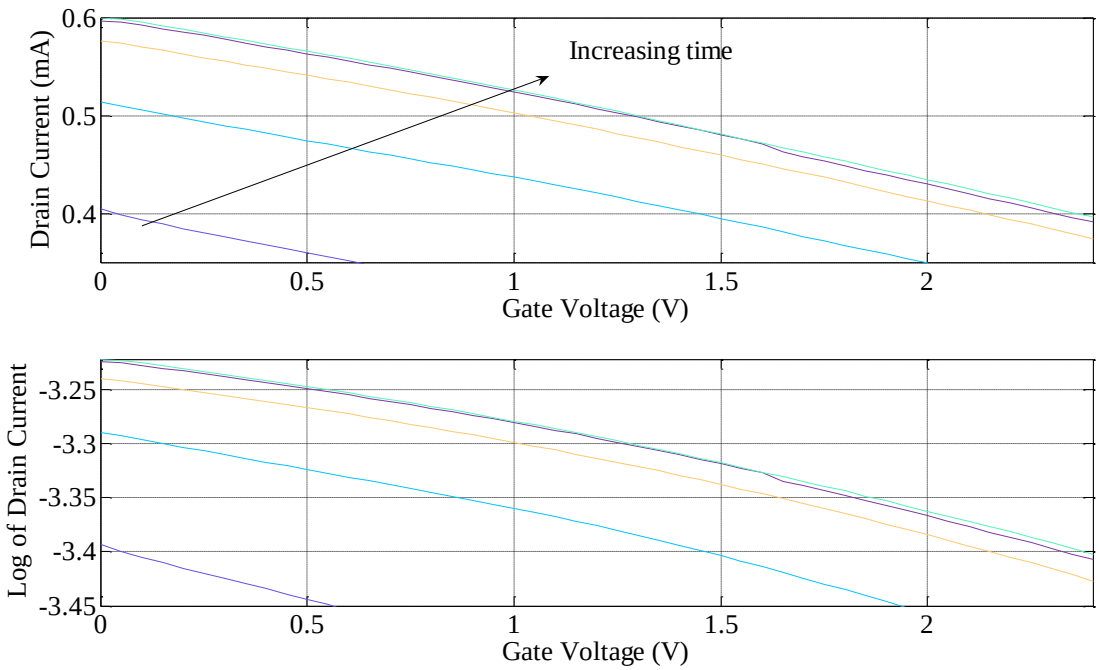


Figure A.22. Gate sweep voltage of device 71 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 7 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 400 µA and 600 µA, respectively. The effective gate sweep voltage range for minimum current is between 0.6 and 2.4 V; as shown in Figure A.22. The currents achieved with device 71 are low compared to devices 8 and 10. Device 71 has a smaller width and the measured results show high dependency on the FG transistor width. A detailed discussion is presented in Section 5.2.2.

A.2.4 Device 73

The results for 5 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.23.

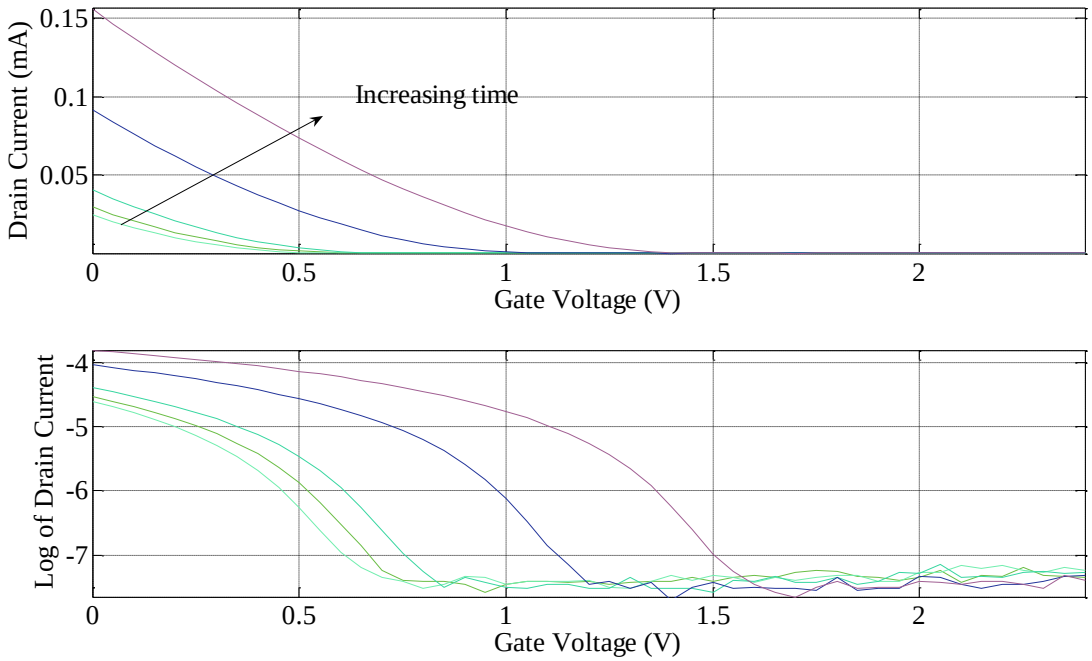


Figure A.23. Gate voltage sweep of device 73 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 5 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.23 the minimum and maximum drain currents are 20 µA and 150 µA, respectively. The effective gate sweep range for minimum current is between 0.6 and 1.5 V. The results for 5 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.24.

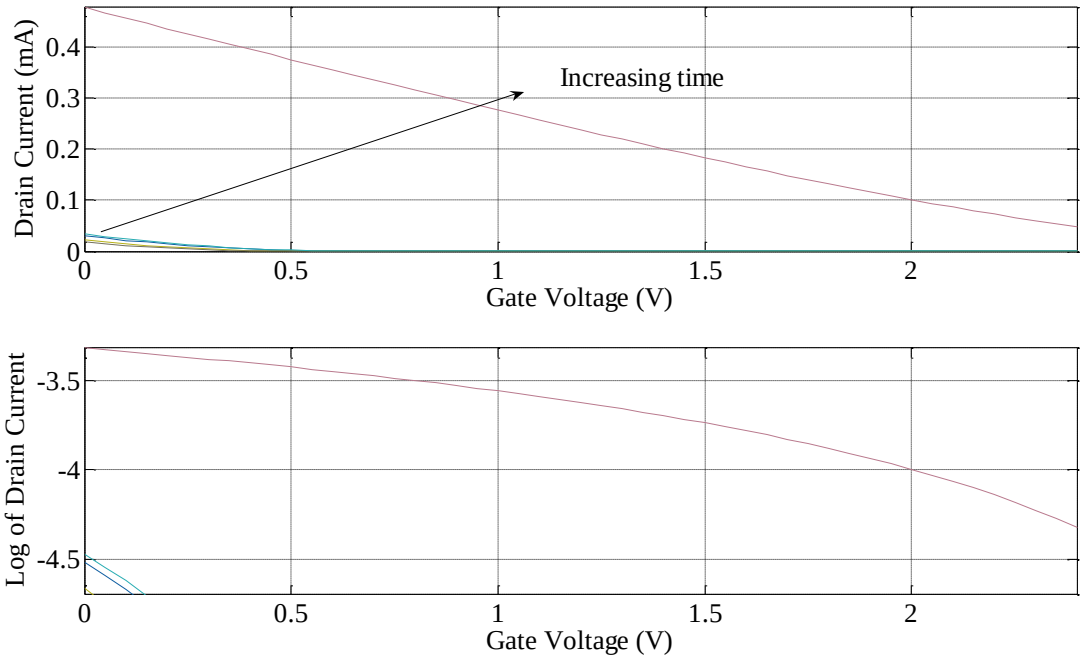


Figure A.24. Gate voltage sweep of device 73 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 5 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 20 µA and 500 µA, respectively. The effective gate sweep range for minimum current is between 0.2 and 2.4 V; as shown in Figure A.24. The results for 6 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.25.

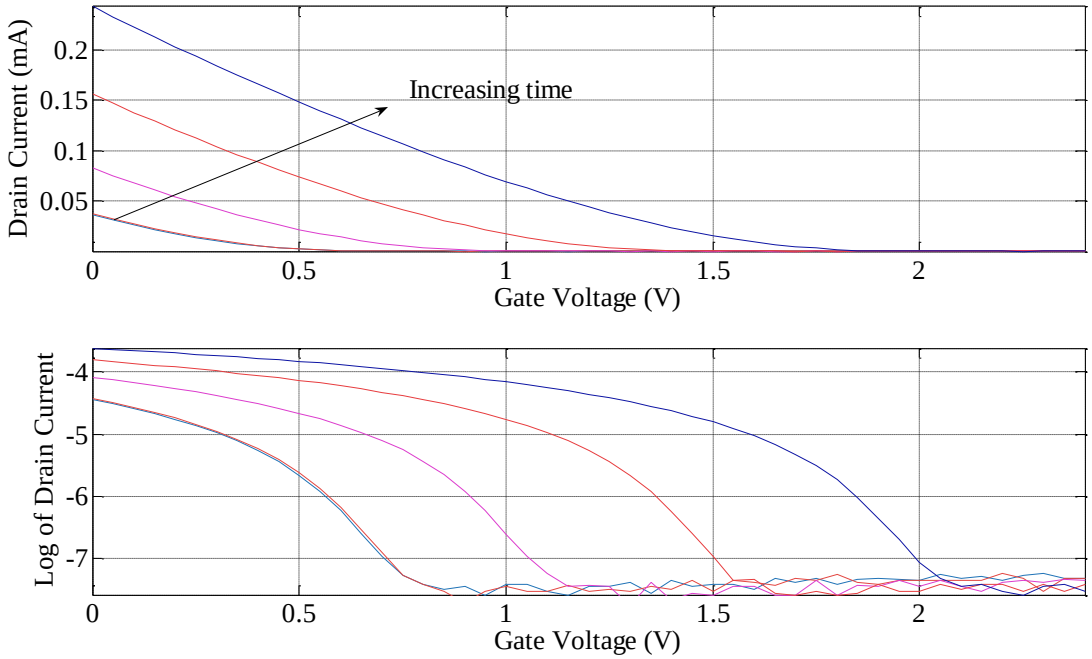


Figure A.25. Gate voltage sweep of device 73 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 6 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.25 the minimum and maximum drain currents are 50 µA and 300 µA, respectively. The effective gate sweep range for minimum current is between 0.75 and 2.1 V. The results for 6 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.26

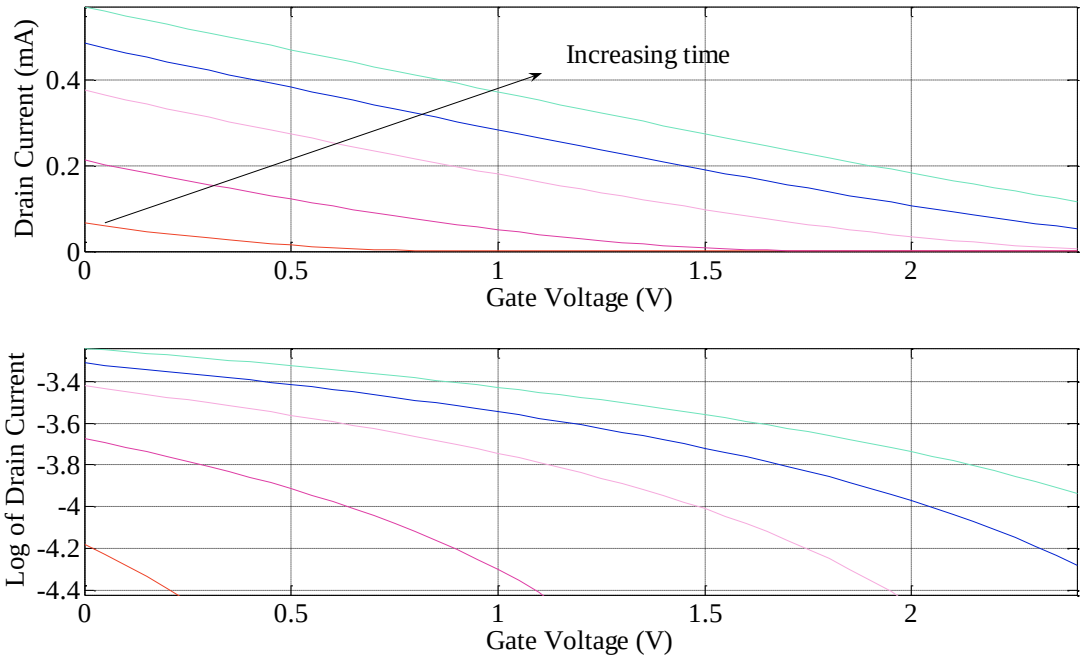


Figure A.26. Gate voltage sweep of device 73 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 6 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 80 µA and 580 µA, respectively. The effective gate sweep range for minimum current is between 0.3 and 2.4 V, as shown in Figure A.26. The results for 7 V drain-to-source voltage with 1.6 V gate voltage are presented in Figure A.27.

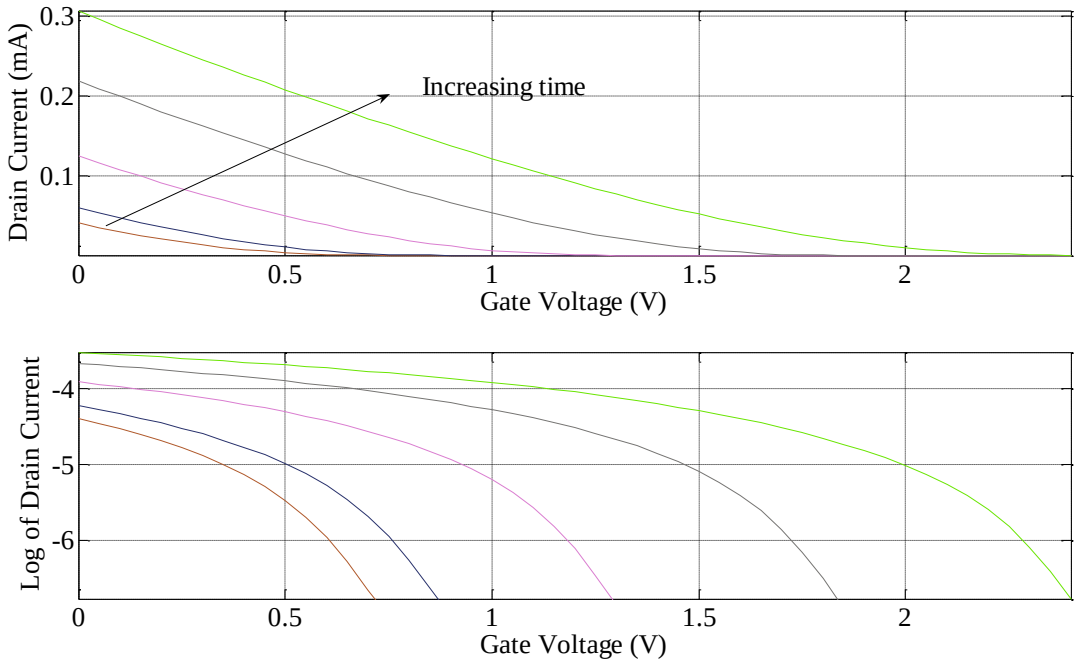


Figure A.27. Gate voltage sweep of device 73 after injection. The gate voltage is 1.6 V and the source-to-drain voltage is 7 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

In Figure A.27 the minimum and maximum drain currents are 40 µA and 300 µA, respectively. The effective gate sweep range for minimum current is between 0.7 and 2.4 V. The results for 7 V drain-to-source voltage with 3.2 V gate voltage are presented in Figure A.28.

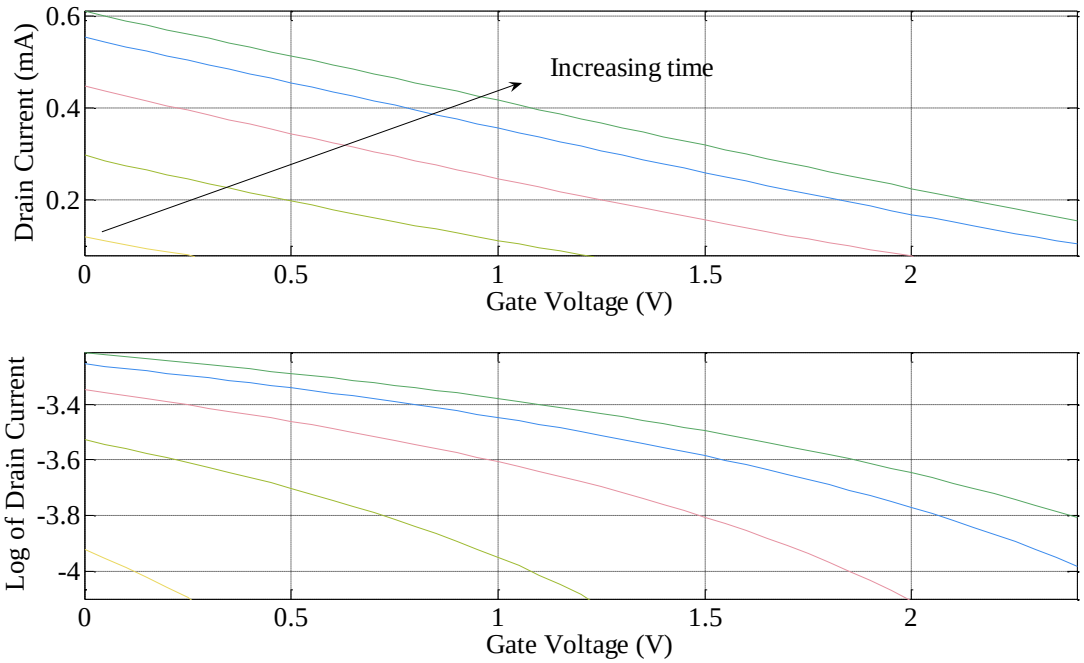


Figure A.28. Gate voltage sweep of device 73 after injection. The gate voltage is 3.2 V and the source-to-drain voltage is 7 V. The injection time was varied logarithmically from 10 to 10^5 s. The graphs illustrate different injection times.

The minimum and maximum drain currents are 110 µA and 600 µA, respectively. The effective gate sweep range for minimum current is between 0.3 and 2.4 V; as shown in Figure A.28. The drain current for device 73 is in the µA range, which is comparable with device 71, but lower than the drain currents of devices 8 and 10, which are in the mA range. The discussion and conclusions of the results presented here are in the main part of this dissertation, Section 5.2.

APPENDIX B: CIRCUIT LAYOUTS

In this appendix, the layouts schematics used for the fabricated devices are presented. The two-phase generators were integrated with the Dickson CP and voltage doubler CPs, thus they were not fabricated as sub-systems to enable independent measurements for the clock generator (Figure B.1) and the CPs (Figures B.2 to B.4). The layouts for FG devices are presented through Figures B.5 to B.8.

Two-phase clock generator

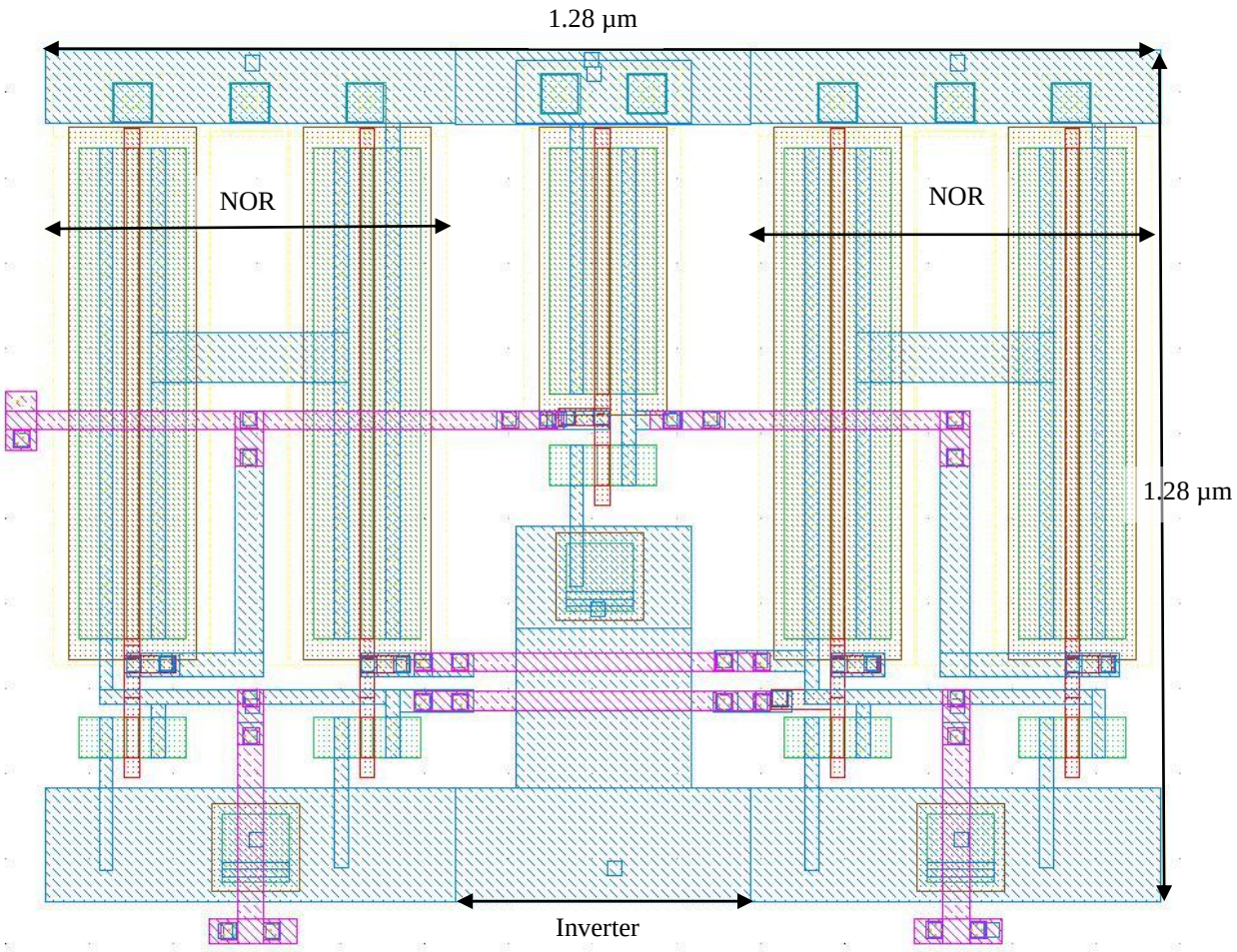


Figure B.1. A two-phase clock layout; two NOR gates and an inverter.

Dickson CP

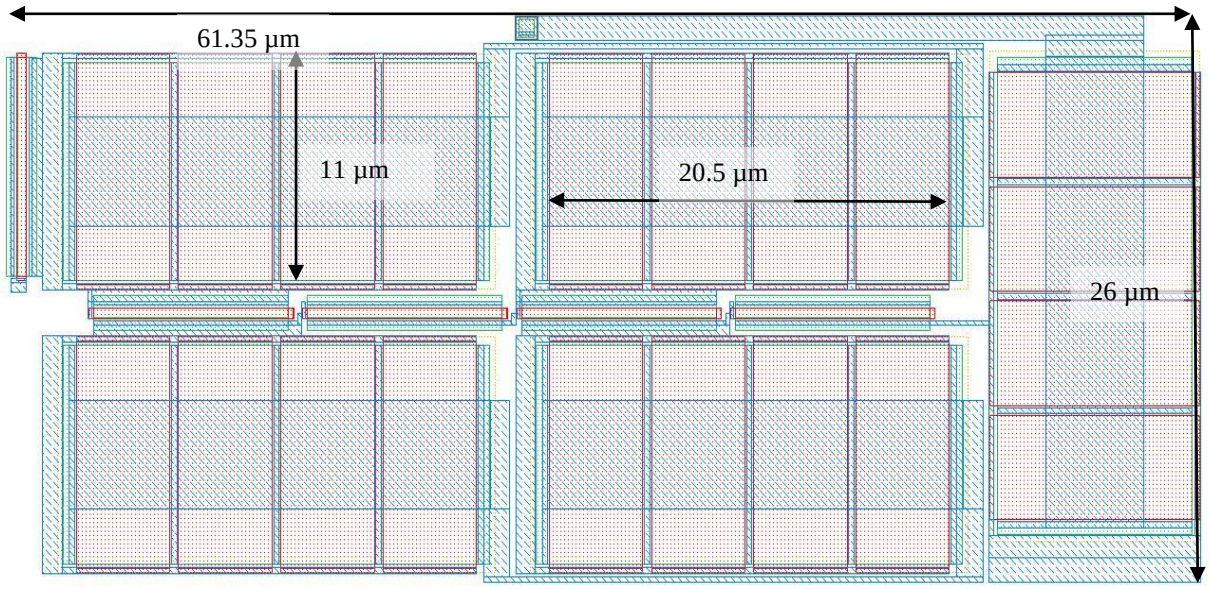


Figure B.2. A Dickson CP. A NFET is used with W = 10 µm and L = 480 nm; stage capacitor - W = 10 µm and L = 4.8 µm. Load capacitor - W = 10 µm and L = 4.8 µm (2 pF).

Voltage doubler CPs

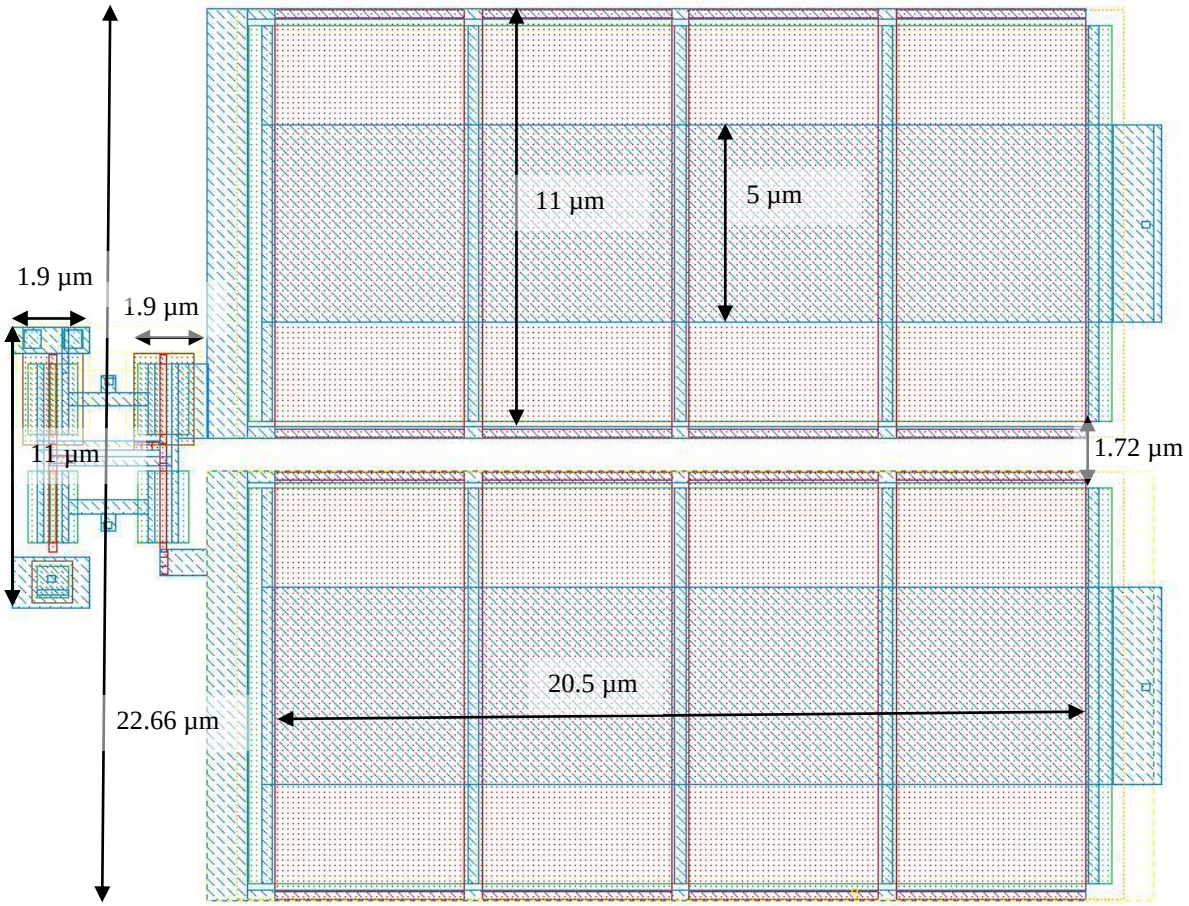


Figure B.3. A layout for the voltage doubler CP stage. NFET W = 1.8 µm and L = 180 nm; PFET W = 1.8 µm and L = 180 nm; stage capacitor - W = 10 µm and L = 4.8 µm (2 pF).

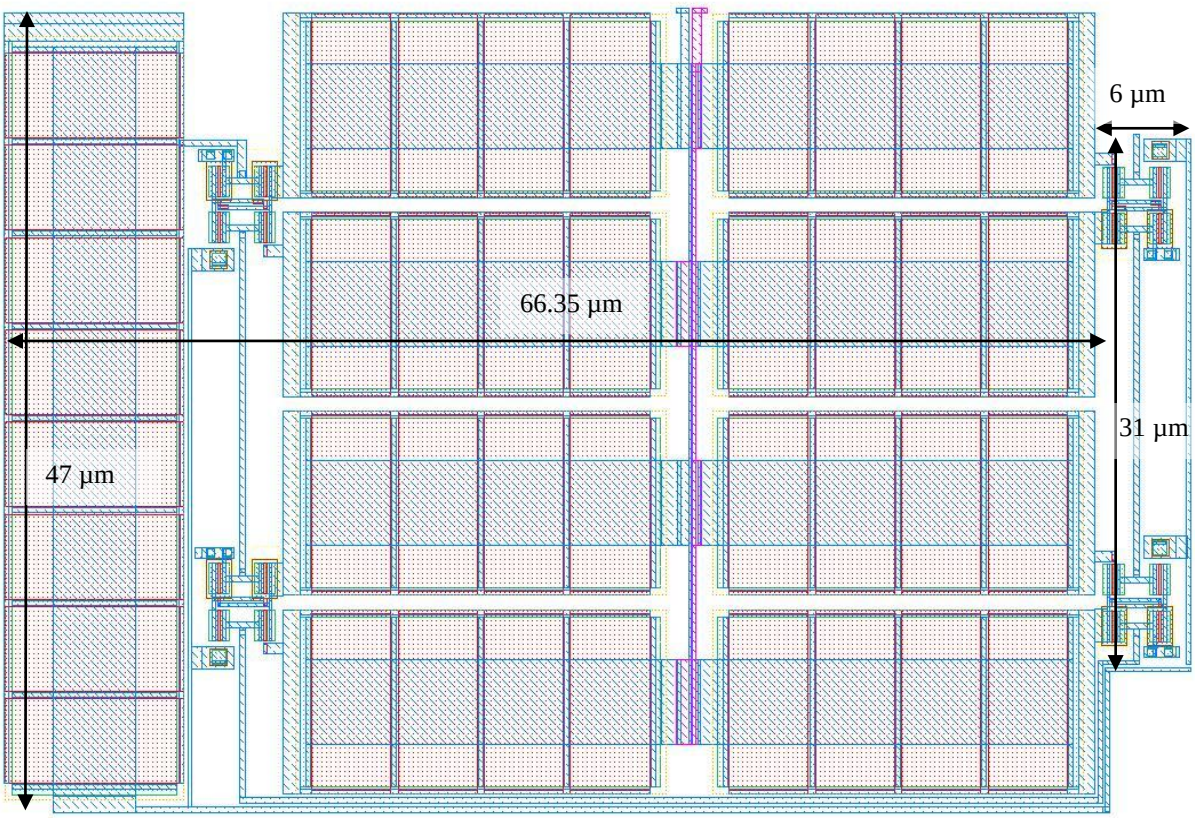


Figure B.4. Four-stage voltage doubler CP with bulk connected to next stage. NFET W = 1.8 µm and L = 180 nm; PFET W = 1.8 µm and L = 180 nm; stage capacitor W = 10 µm and L = 4.8 µm; load capacitor W = 10 µm and L = 5 µm (2 pF).

Floating gates

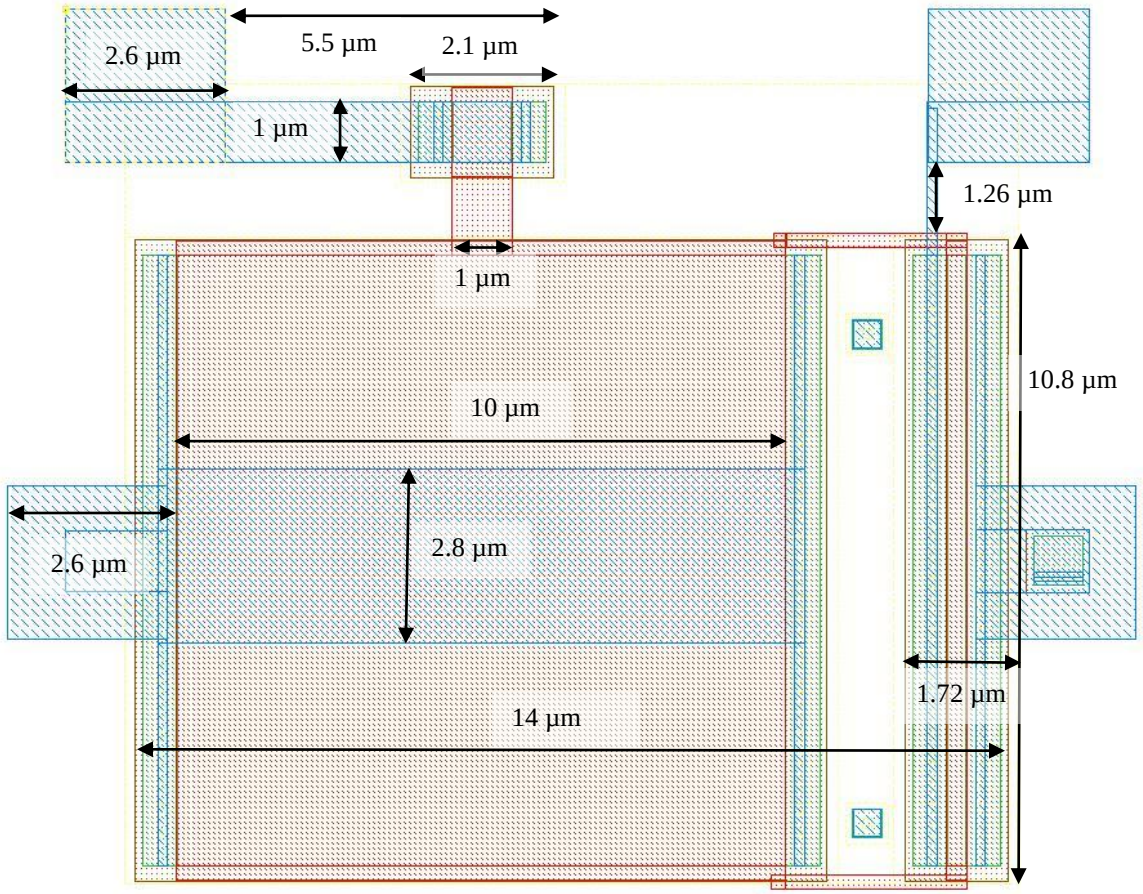


Figure B.5. FG device 1. Tunnelling capacitor W = 1 µm and L = 1 µm; control capacitor W = 10 µm and L = 10 µm; PFET transistor W = 10 µm and L = 350 nm.

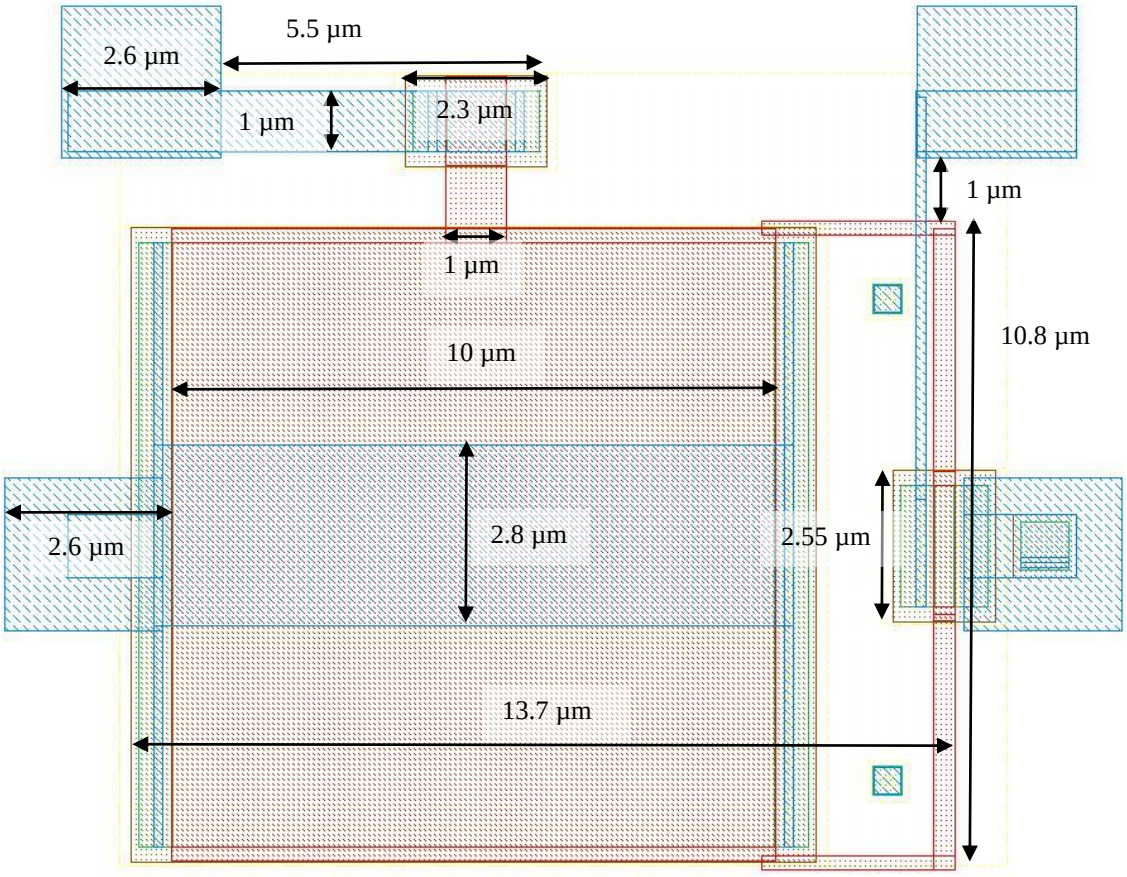


Figure B.6. FG device 2. Tunnelling capacitor W = 1 µm and L = 1 µm; control capacitor W = 10 µm and L = 10 µm; PFET transistor W = 2 µm and L = 350 nm.

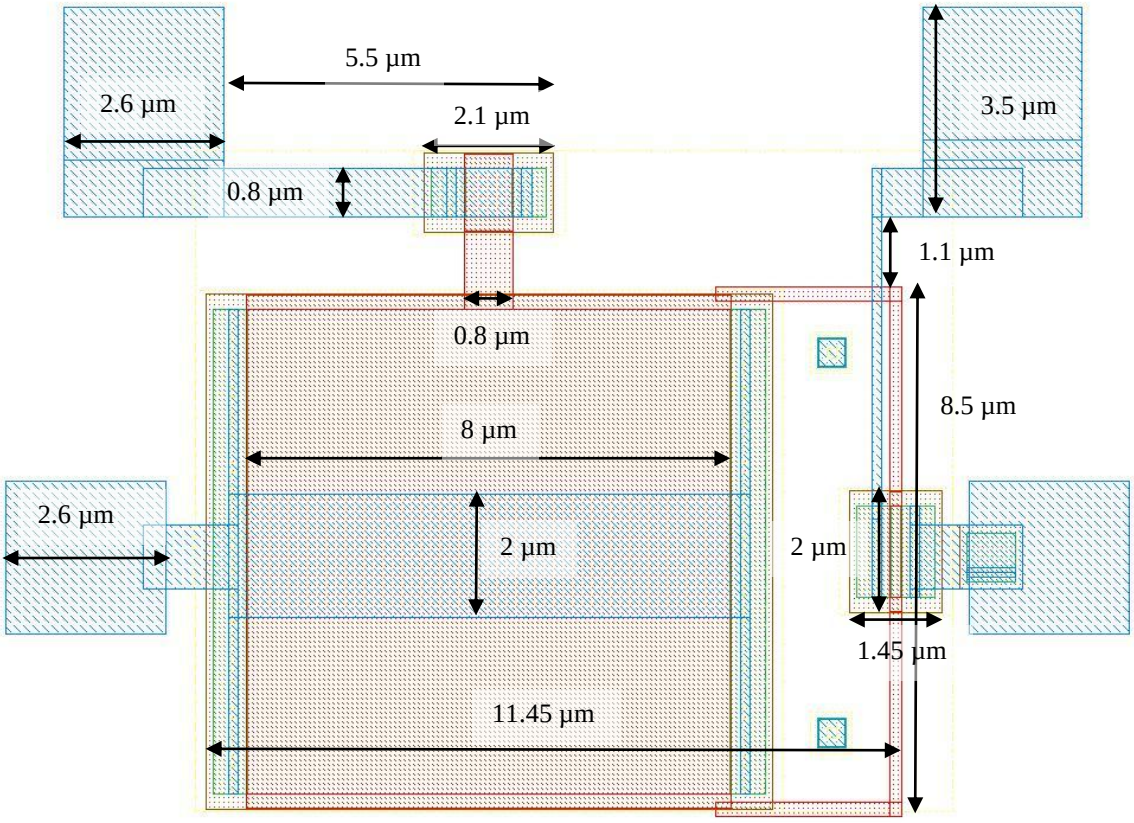


Figure B.7. FG device 3. Tunnelling capacitor W = 0.8 µm and L = 0.8 µm; control capacitor W = 8 µm and L = 8 µm; PFET transistor W = 1.5 µm and L = 180 nm.

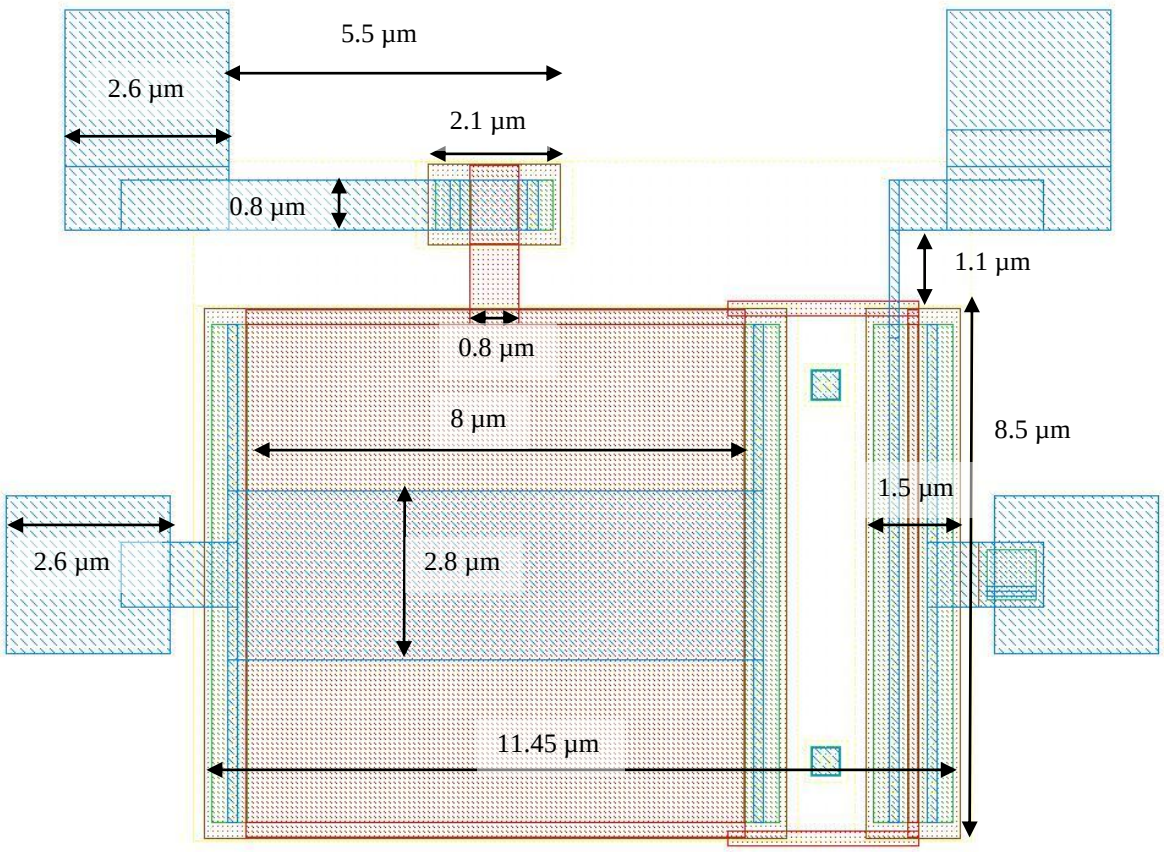


Figure B.8. FG device 4. Tunnelling capacitor W = 0.8 µm and L = 0.8 µm; control capacitor W = 8 µm and L = 8 µm; PFET transistor W = 8 µm and L = 180 nm.

APPENDIX C: MEASURED RESULTS

C.1 FG DEVICE 4

The results of drain voltage sweep for device 4 before programming are presented in Figure C.1.

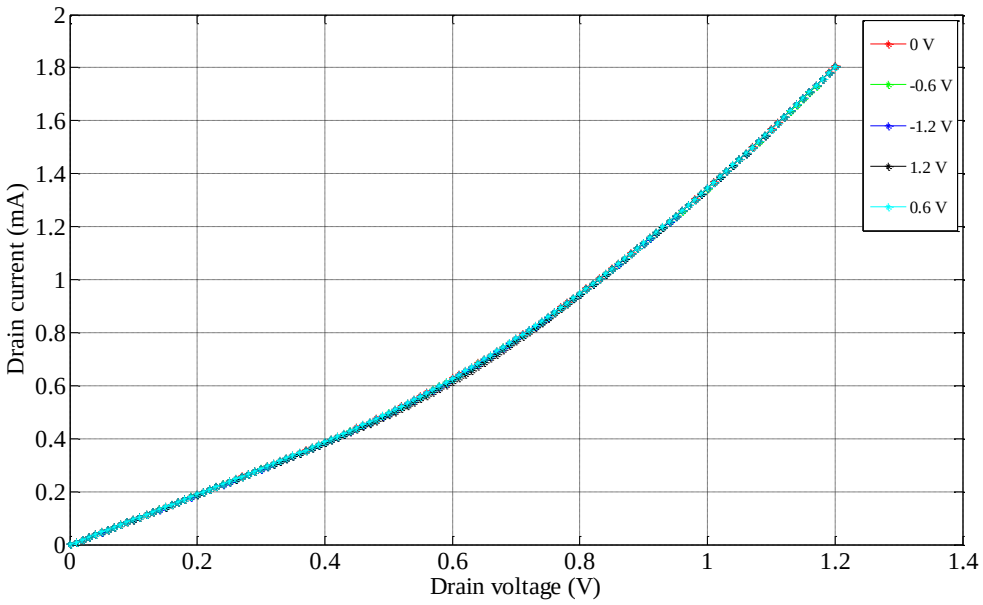


Figure C.1. Device 4 drain voltage sweep before programming the device.

The drain voltage sweep was done for different control gate voltages as shown by the legends in Figure C.1. The results are independent of the control gate voltage, since the control gate capacitor blocks the DC path. The maximum drain current is 1.8 mA. The gate voltage sweep measurements are presented in Figure C.2 and Figure C.3 for different drain-to-source voltages as represented by the legends.

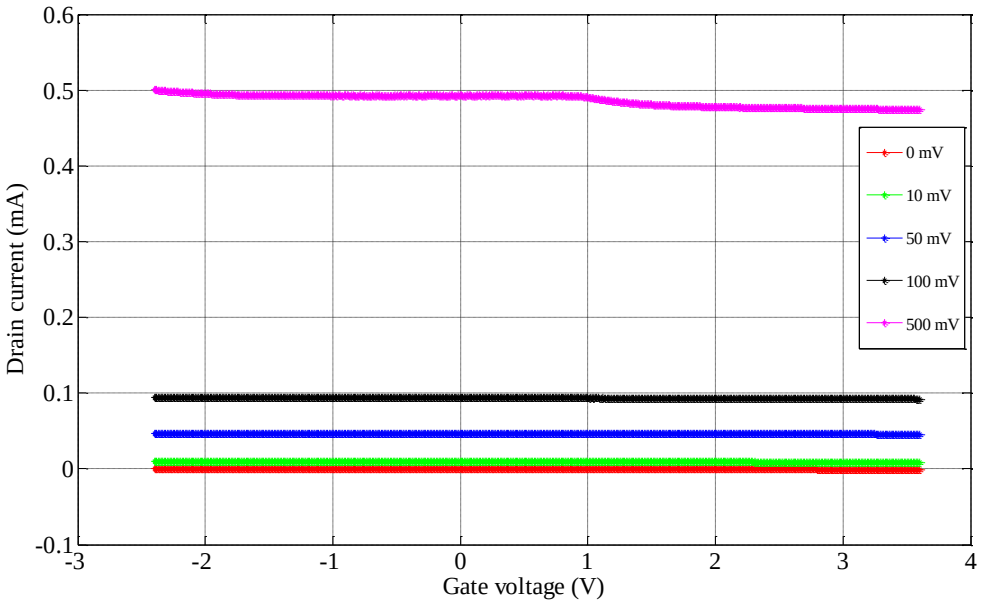


Figure C.2. Device 4 gate voltage sweep before programming.

As the drain-to-source voltage increases, the drain current also increases; as shown in Figure C.2. For a drain-to-source voltage between 0 and 100 mV, the drain current varies from 0 to 0.1 mA. Increasing the drain-to-source voltage from 100 to 500 mV (by factor of 5) results in an increased drain current by a factor of 5 from 0.1 to 0.5 mA. The results in Figure C.2 show that the drain current is approximately constant over the gate voltage sweeping range and it varies based on the drain-to-source voltage. The threshold voltage behaviour is dominantly visible for low drain-to-source voltages from the semilog graph in Figure C.3.

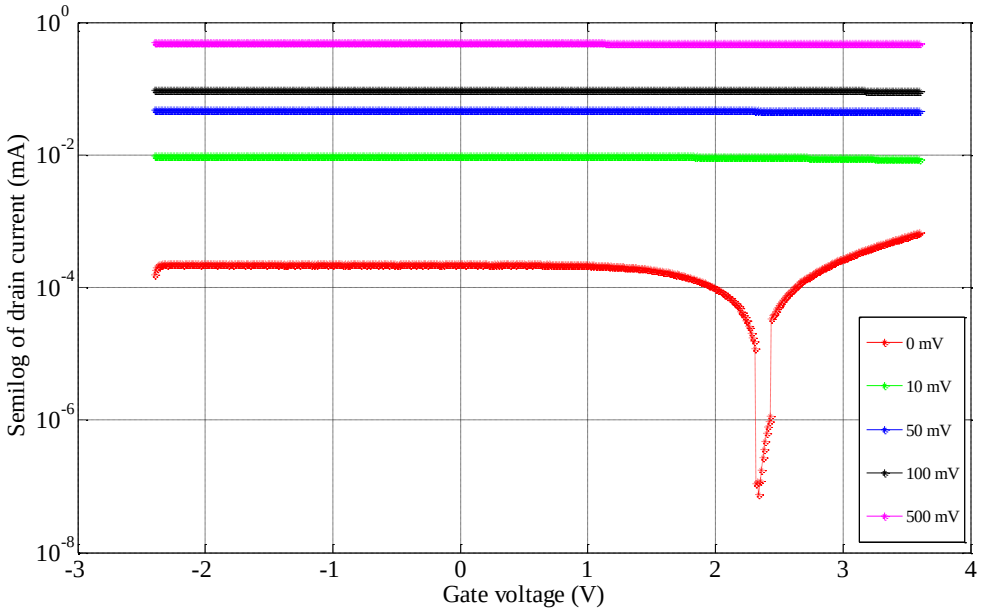


Figure C.3. Device 4 gate voltage sweep before programming with logarithmic scale.

The threshold voltage in Figure C.3 is 2.3 V. It can be noted that for voltages below threshold (ignoring the device behaviour around the threshold) the drain current is constant at its minimum level and above threshold it starts increasing linearly. The drain voltage sweep measurements after tunnelling are shown in Figure C.4.

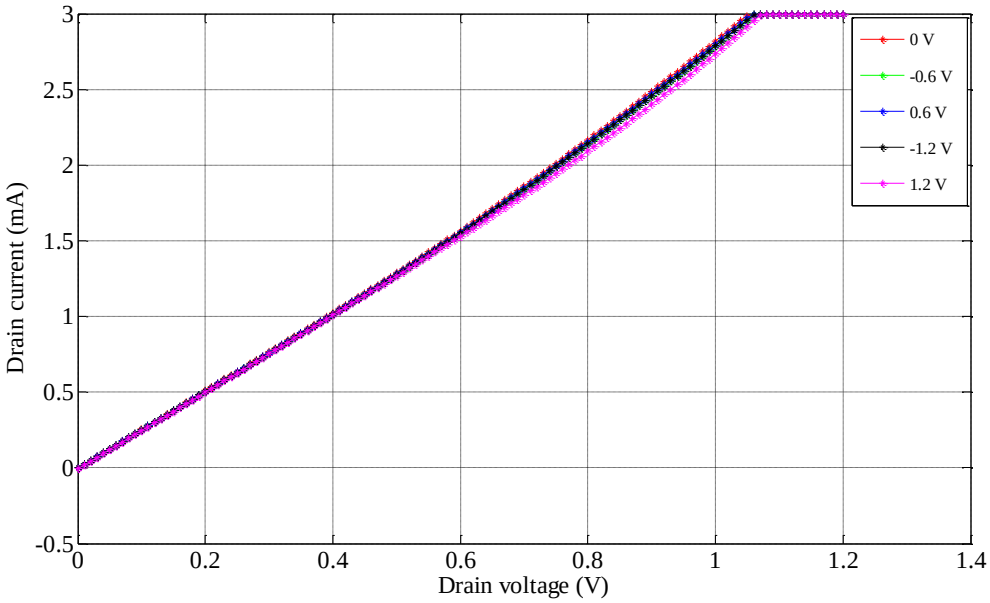


Figure C.4. Device 4 drain voltage sweep after tunnelling with 5 V.

The drain current increased from 1.8 mA (before programming as shown in Figure C.1) to a saturation level of 3 mA after tunnelling (Figure C.4). The drain current shows a small variation with respect to control gate voltage. The results for gate voltage sweep after tunnelling are shown in Figure C.5.

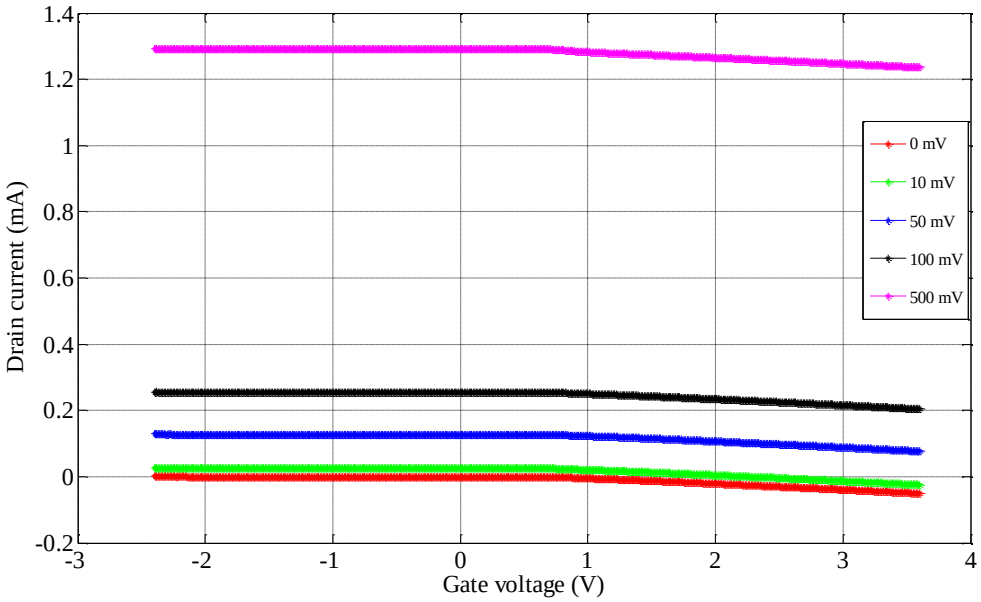


Figure C.5. Device 4 gate voltage sweep after tunnelling.

After tunnelling, the drain current increased by a factor ranging between 2.3 and 2.6 for different drain-to-source voltage. The maximum current is 1.3 mA (Figure C.5) compared to 0.5 mA (Figure C.2) before tunnelling the device. The change in threshold voltage can be interpreted from the semilog scale in Figure C.6.

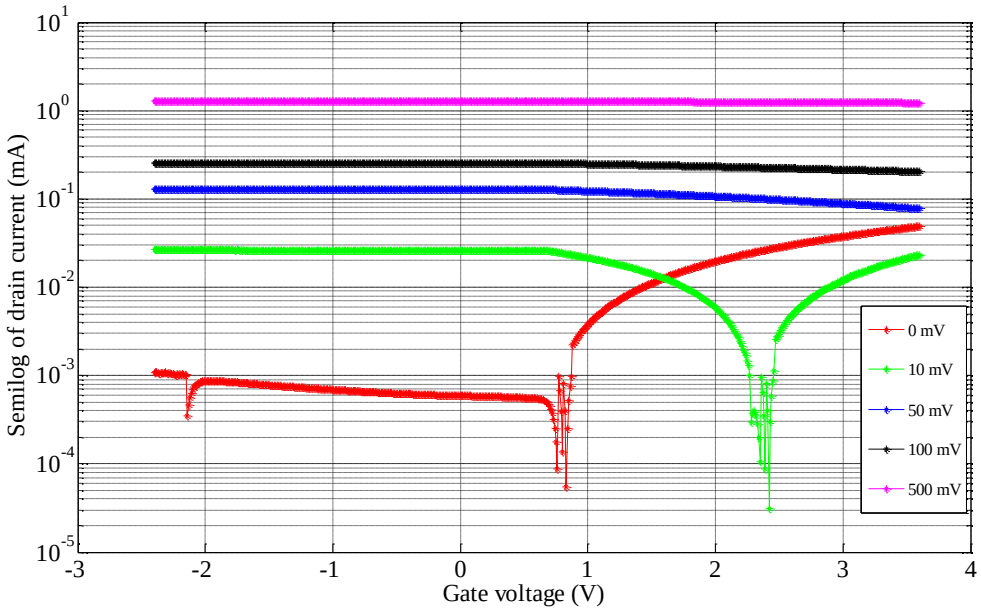


Figure C.6, Device 4 gate voltage sweep after tunnelling with logarithmic scale.

The threshold voltages are 1.75 and 2.4 V for 0 and 10 mV, respectively (Figure C.6). Thus, the threshold voltage was reduced by at least 0.55 V from 2.3 V achieved before programming the device (Figure C.3). The gate voltage sweep measurements after injection are shown in Figure C.7.

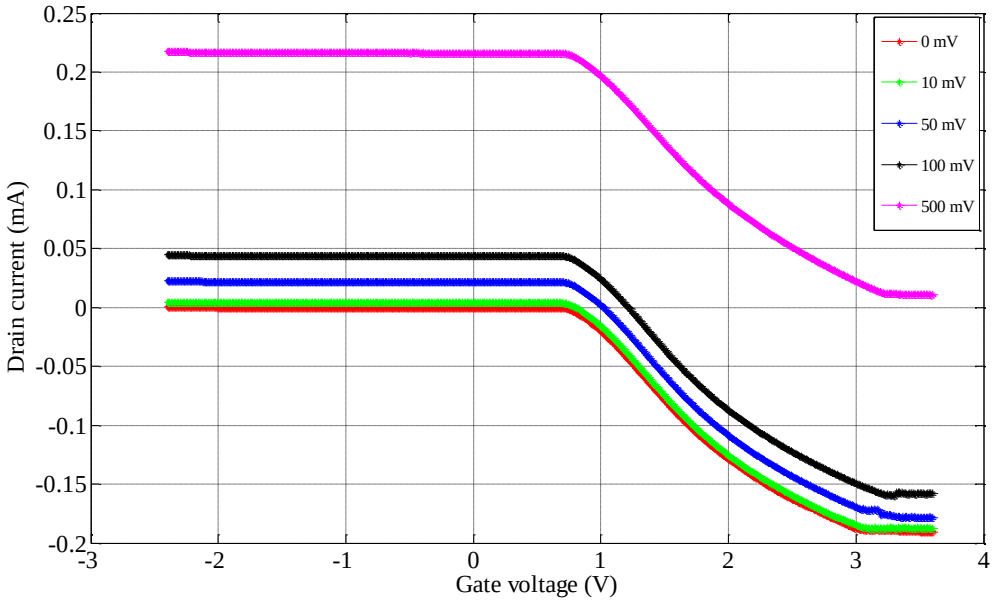


Figure C.7. Device 4 gate voltage sweep after injection.

The maximum drain current decreased from 0.5 mA (as measured after tunnelling) to 0.22 mA after injection programming as shown in Figure C.7. The drain current for different drain-to-source voltage was reduced on average by a factor of 2 and the threshold voltage changes can be noted from Figure C.8.

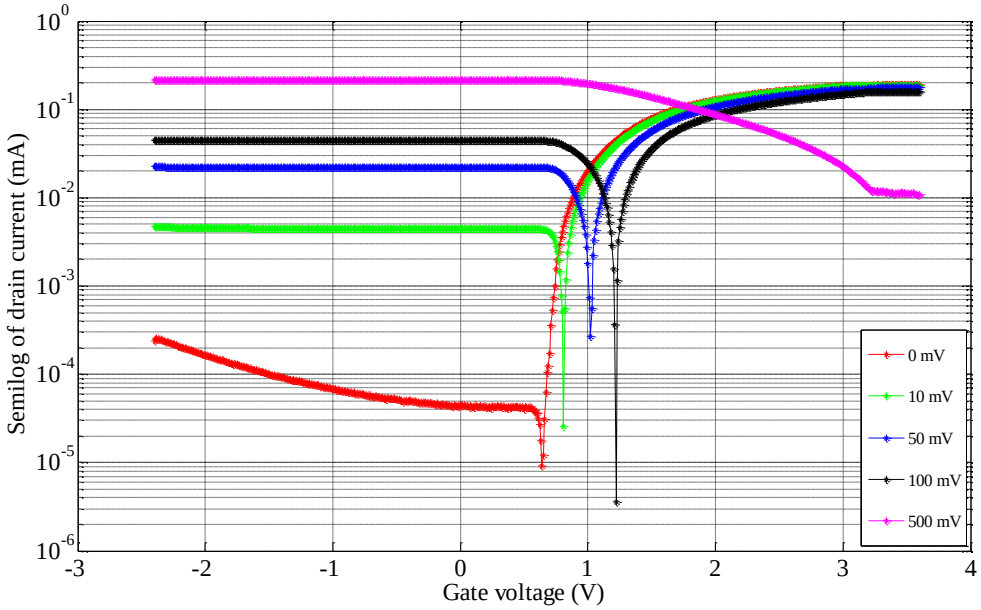


Figure C.8. Device 4 gate voltage sweep after injection with logarithmic scale.

The threshold voltage varies from 0.73 to 1.27 V, as shown in Figure C.8. The threshold before programming was 2.3 V and the threshold was between 0.8 and 2.3 V after tunnelling. Tunnelling reduces the threshold voltage and injection increases the threshold voltage, but based on the results achieved, injection decreased the threshold voltage. The initial charge after fabrication is unknown and different for all devices, thus there is uncertainty about the results achieved by the measurements before programming the devices as reference.

C.2 FG DEVICE 3

The results of drain voltage sweep for device 3 before programming are presented in Figure C.9.

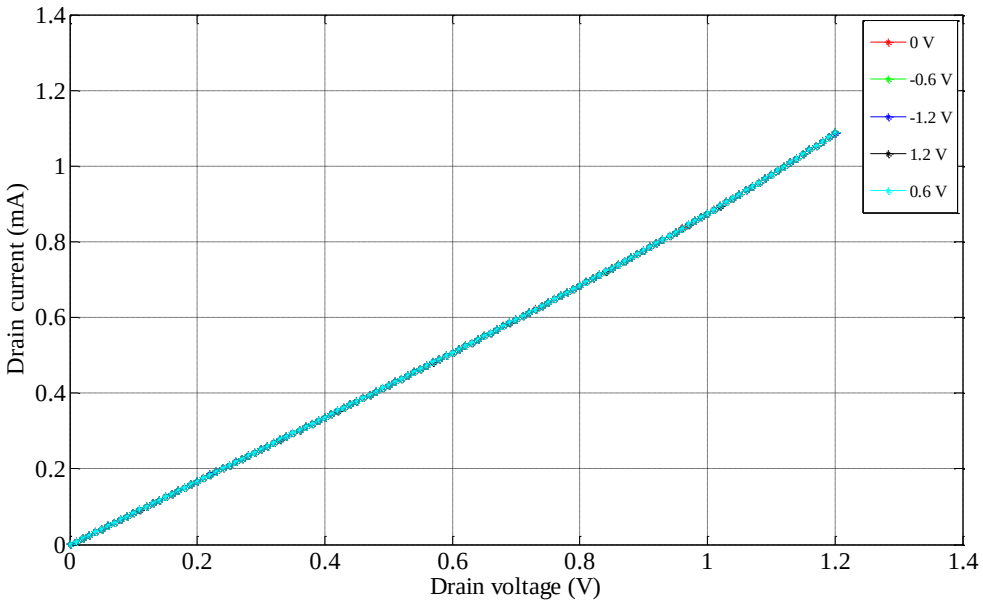


Figure C.9. Device 3 drain voltage sweep before programming the device.

The drain voltage sweep was done for different control gate voltages as shown by the legends in Figure C.9. The results are independent of the control gate voltage since the control gate capacitor blocks the DC path. The maximum drain current is 1.1 mA. The gate voltage sweep measurements are presented in Figure C.10 and Figure C.11 for different drain-to-source voltages as represented by the legends.

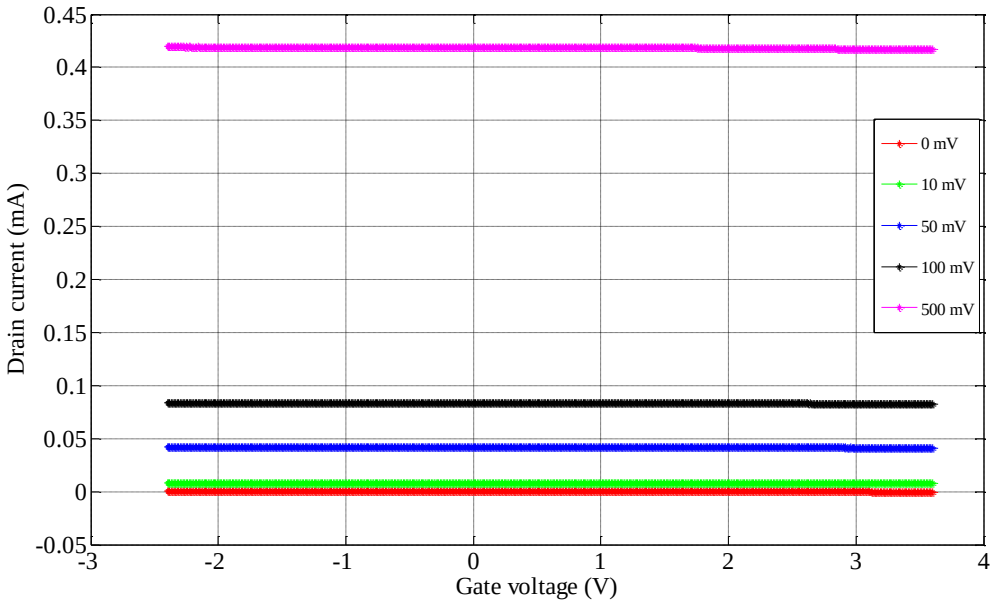


Figure C.10. Device 3 gate voltage sweep before programming the device.

As the drain-to-source voltage increases, the drain current also increases; as shown in Figure C.10. For a drain-to-source voltage between 0 and 100 mV, the drain current varies from 0 to 0.85 mA. Increasing the drain-to-source voltage from 100 to 500 mV (by a factor of five) results in an increased drain current by a factor of 5 from 0.85 to 0.42 mA. The results in Figure C.10 show that the drain current is approximately constant over the gate voltage sweeping range and varies based on the drain-to-source voltage. The threshold voltage behaviour is dominantly visible for low drain-to-source voltages from the semilog graph in Figure C.11.

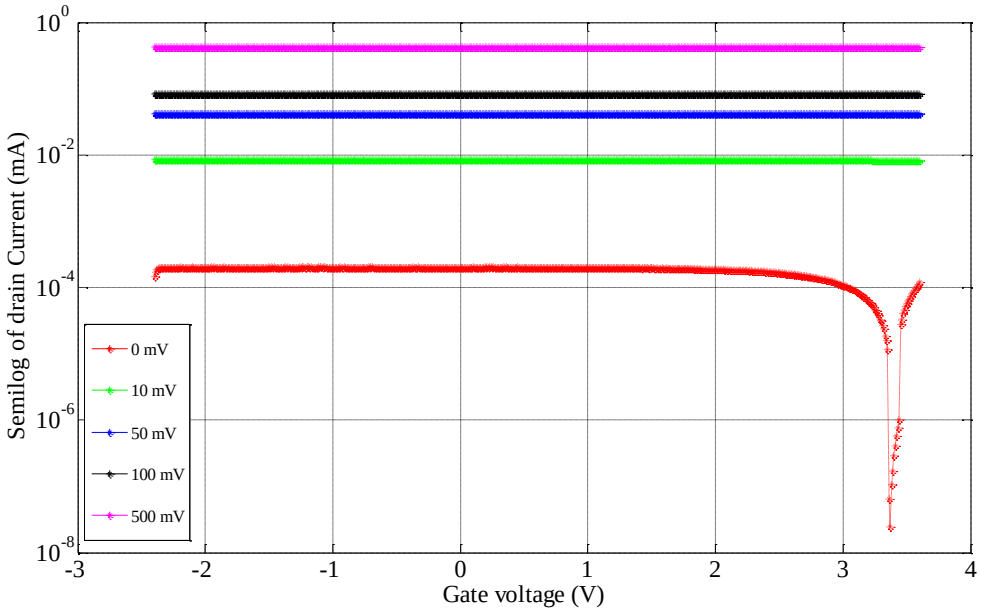


Figure C.11. Device 3 gate voltage sweep before programming the device with logarithmic scale.

The threshold voltage in Figure C.11 is 3.3 V. It can be noted that for voltages below threshold (ignoring the device behaviour around the threshold) the drain current is constant at its minimum level and above threshold it starts increasing linearly. The drain voltage sweep measurements after tunnelling are shown in Figure C.12.

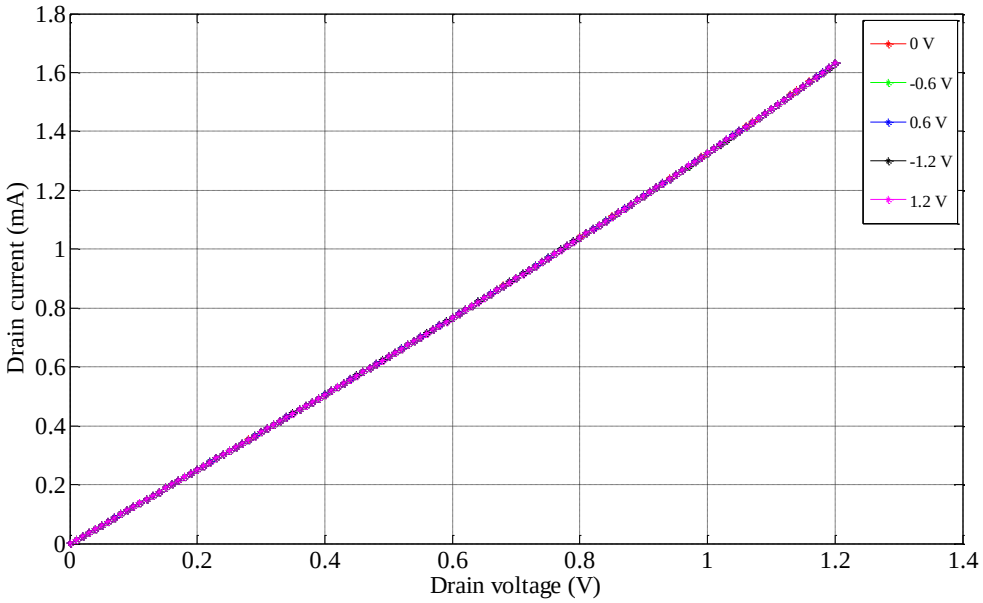


Figure C.12. Device 3 drain voltage sweep after tunnelling.

The drain current increased from 1.1 mA (before programming as shown in Figure C.9) to 1.61 mA after tunnelling (Figure C.12). The drain current shows a small variation with respect to control gate voltage. The results for gate voltage sweep after tunnelling are shown in Figure C.13.

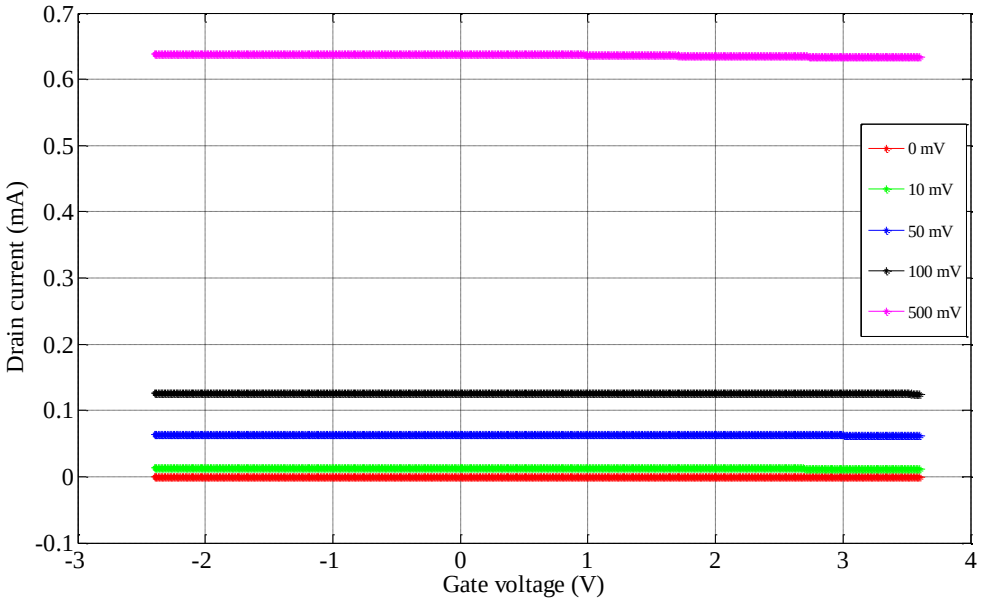


Figure C.13. Device 3 gate voltage sweep after tunnelling.

After tunnelling, the drain current increased by a maximum factor of 1.5 for a different drain-to-source voltage. The maximum current is 0.63 mA (Figure C.13) compared to 0.42 mA (Figure C.10) before tunnelling the device. The change in threshold voltage can be interpreted from the semilog scale in Figure C.14.

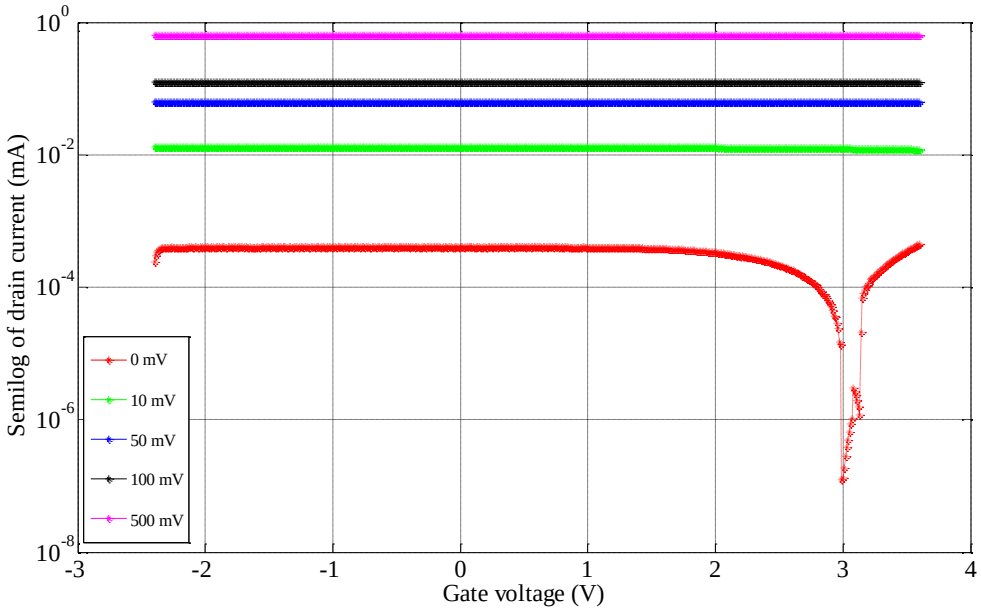


Figure C.14. Device 3 gate voltage sweep after tunnelling with logarithmic scale.

The threshold voltage is 3 V for 0 mV, as shown in Figure C.14. Thus, the threshold voltage was reduced by at least 0.3 V from 3.3 V achieved before programming the device (Figure C.11). The gate voltage sweep measurements after injection are shown in Figure C.15.

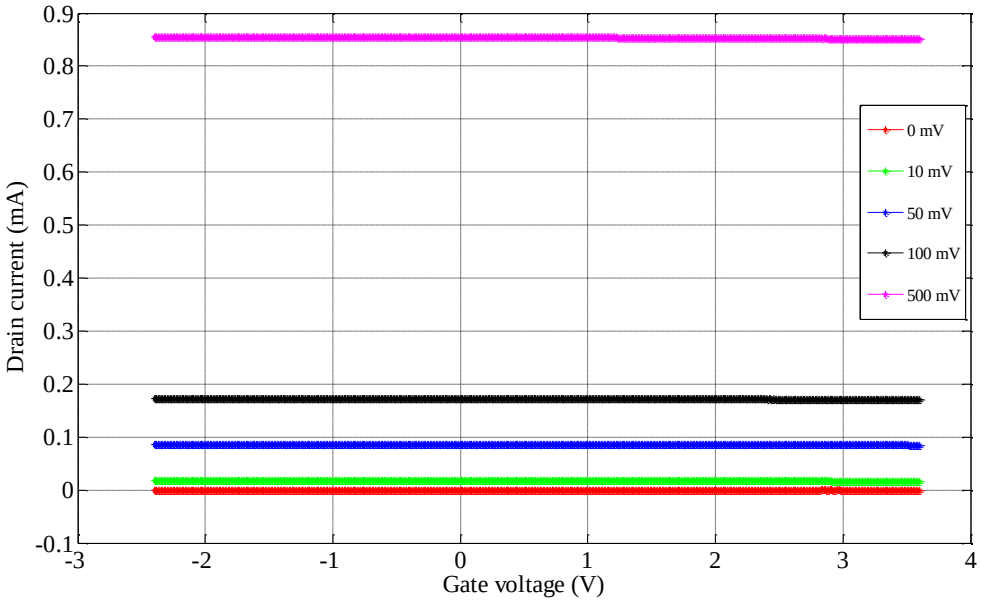


Figure C.15. Device 3 gate voltage sweep after injection.

The maximum drain current increased from 0.63 mA (as measured after tunnelling) to 0.86 mA (factor of 1.36) after injection programming as shown in Figure C.15. The drain current is supposed to decrease after injection as in the case of device 4, but the unknown initial charge has an effect on the results, as there is no neutral charge reference. The threshold voltage changes can be noted from Figure C.16.

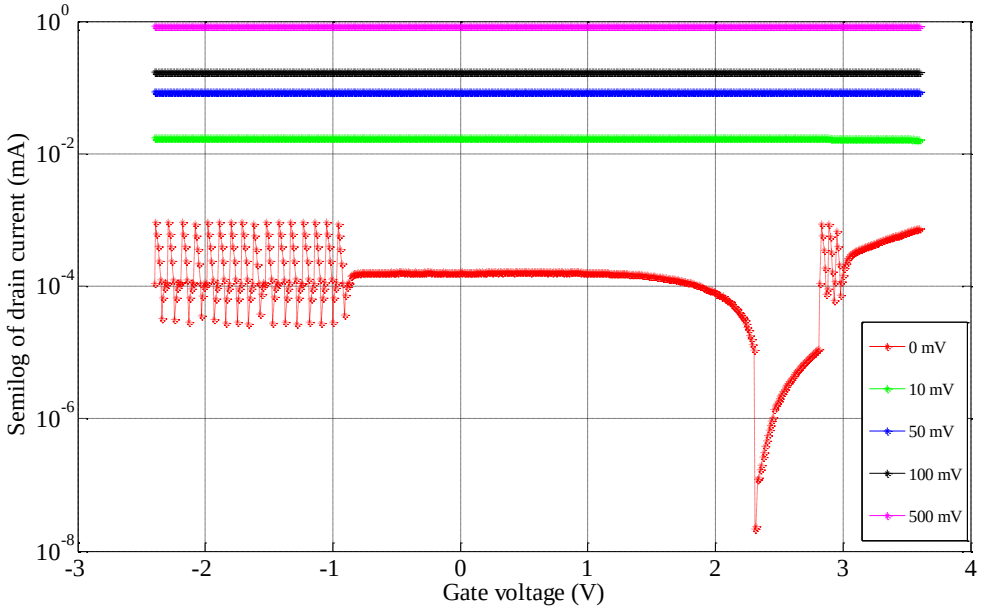


Figure C.16. Device 3 gate voltage sweep after injection with logarithmic scale.

The threshold voltage is 2.2 V, as shown in Figure C.16. The threshold before programming was 3.3 V and after tunnelling the threshold was 3 V. Tunnelling reduces the threshold voltage and injection increases the threshold voltage, but based on the results achieved, injection decreased the threshold voltage. The initial charge after fabrication is unknown and different for all devices, thus there is uncertainty about the results achieved by the measurements before programming the devices as reference.

C.3 FG DEVICE 2

The results of drain voltage sweep for device 2 before programming are presented in Figure C.17.

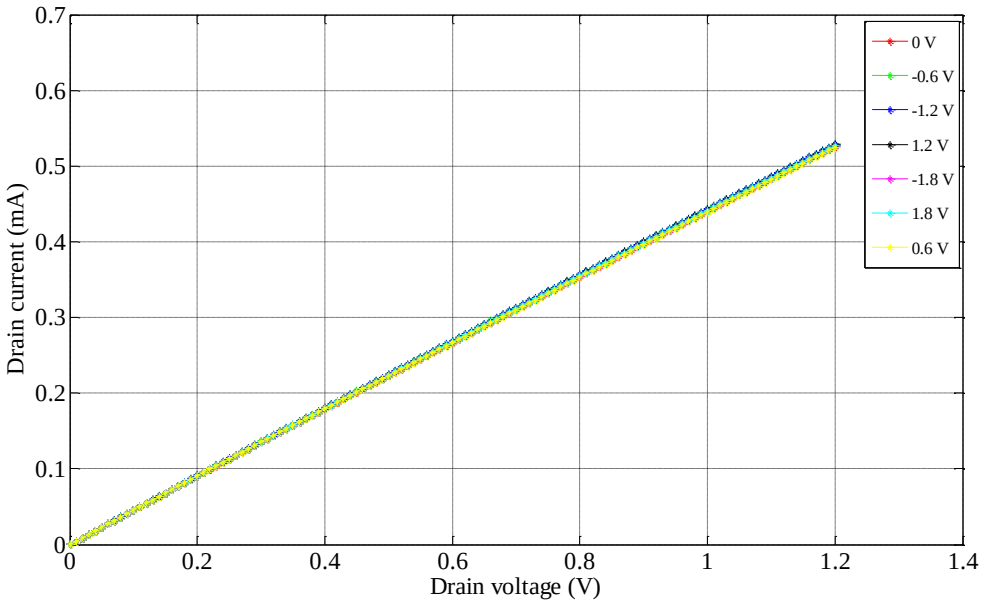


Figure C.17. Device 2 drain voltage sweep before programming the device.

The drain voltage sweep was done for different control gate voltages as shown by the legends in Figure C.17. The results are independent of the control gate voltage, since the control gate capacitor blocks the DC path. The maximum drain current is 0.52 mA. The gate voltage sweep measurements are presented in Figure C.18 and Figure C.19 for different drain-to-source voltages as represented by the legends.

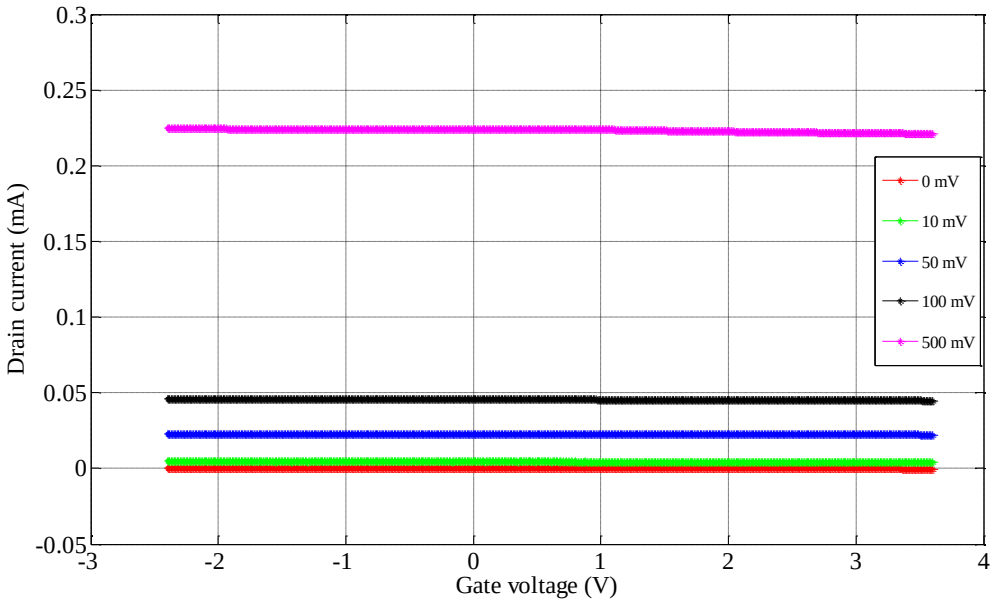


Figure C.18. Device 2 gate voltage sweep before programming the device.

As the drain-to-source voltage increases, the drain current also increases; as shown in Figure C.18. For a drain-to-source voltage between 0 and 100 mV, the drain current varies from 0 to 0.05 mA. Increasing the drain-to-source voltage from 100 to 500 mV (by factor of 5) results in an increased drain current by a factor of 5 (from 0.05 to 0.23 mA). The results in Figure C.18 show that the drain current is approximately constant over the gate voltage sweeping range and varies based on the drain-to-source voltage. The threshold voltage behaviour is dominantly visible for low drain-to-source voltages from the semilog graph in Figure C.19.

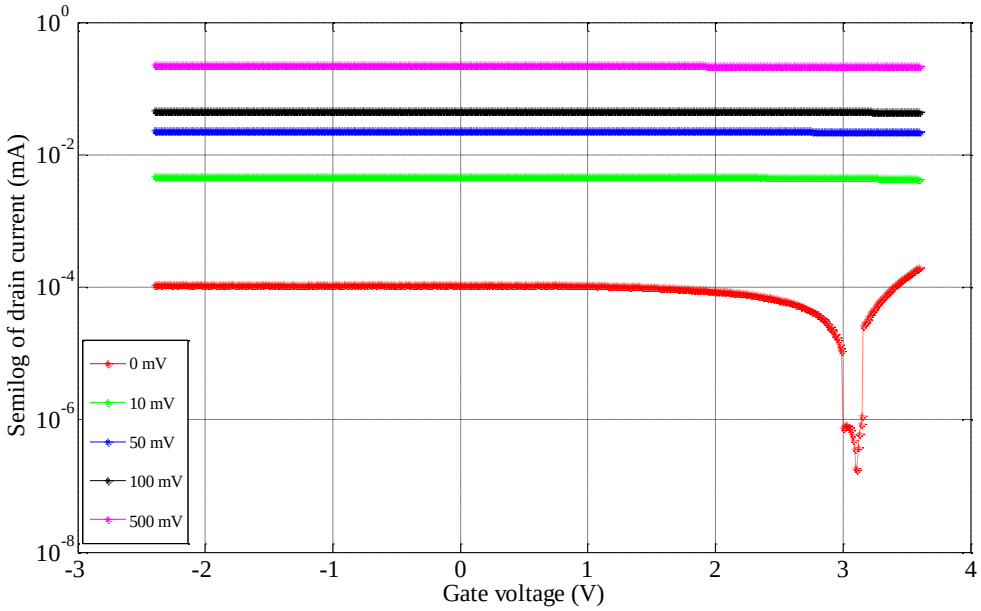


Figure C.19. Device 2 gate voltage sweep before programming the device with logarithmic scale.

The threshold voltage in Figure C.19 is 3.1 V. It can be noted that for voltages below threshold (ignoring the device behaviour around the threshold) the drain current is constant at its minimum level and above threshold it starts increasing linearly. The drain voltage sweep measurements after tunnelling are shown in Figure C.20.

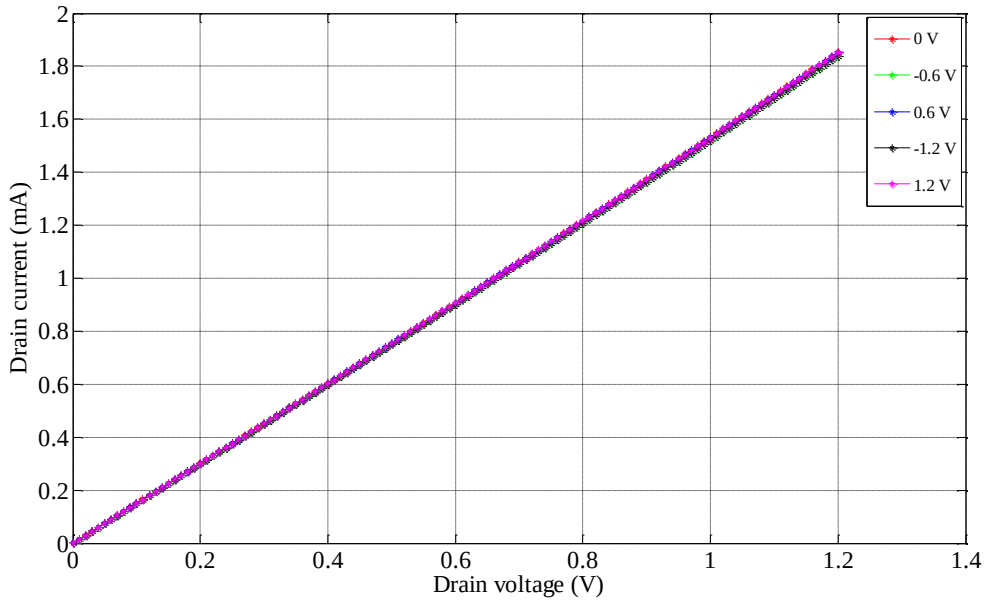


Figure C.20. Device 2 drain voltage sweep after tunnelling.

The drain current increased from 0.52 mA (before programming as shown in Figure C.17) to a saturation level of 1.82 mA after tunnelling (Figure C.20). The drain current shows a small variation with respect to the control gate voltage. The results for the gate voltage sweep after tunnelling are shown in Figure C.21.

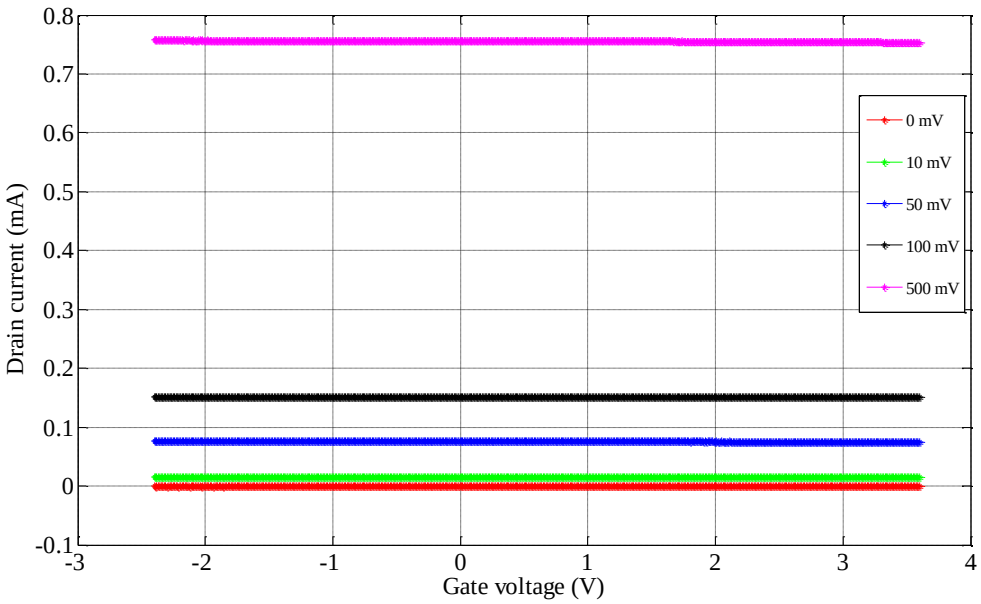


Figure C.21. Device 2 gate voltage sweep after tunnelling.

After tunnelling, the maximum drain current for different drain-to-source voltages increased by a factor of 3, from 0.23 mA (Figure C.18) to 0.76 mA (Figure C.21) before tunnelling the device. The change in threshold voltage can be interpreted from the semilog scale in Figure C.22.

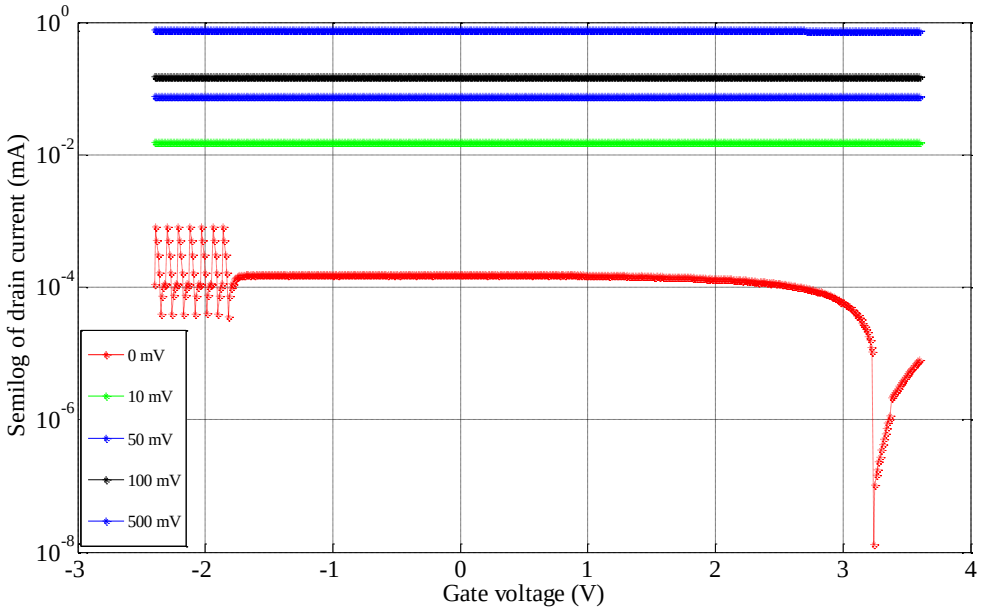


Figure C.22. Device 2 gate voltage sweep after tunnelling with logarithmic scale

The threshold voltage is 3.2 V for 0 mV, as shown in Figure C.22. Thus, the threshold voltage was increased slightly by 0.1 V from 3.1 V achieved before programming the device (Figure C.19). The gate voltage sweep measurements after injection are shown in Figure C.23.

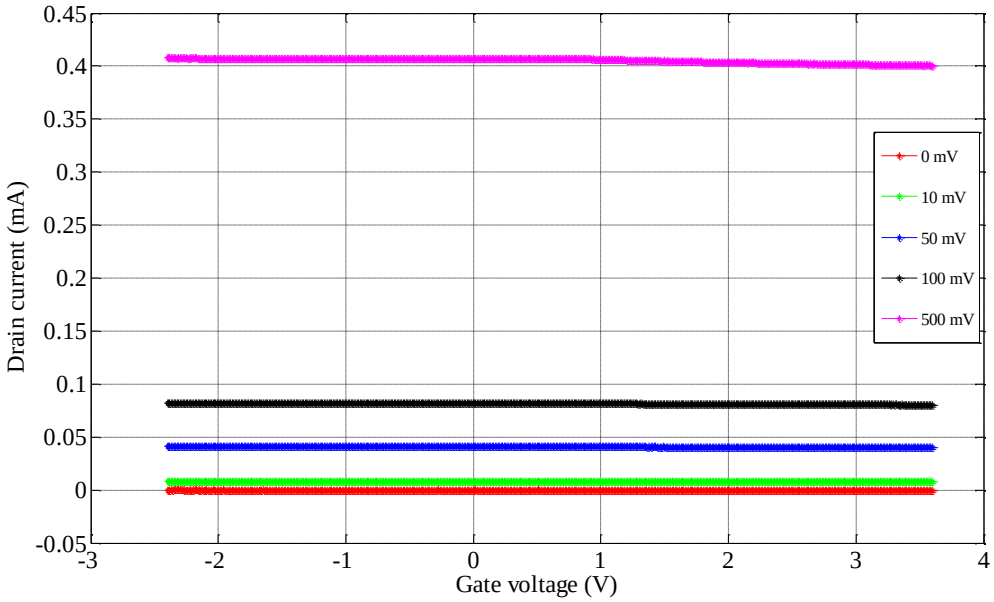


Figure C.23. Device 2 gate voltage sweep after injection.

The maximum drain current decreased from 0.76 mA (as measured after tunnelling) to 0.4 mA after injection programming, as shown in Figure C.23. Thus, the maximum drain current for different drain-to-source voltages was reduced by a factor of 1.9 and the threshold voltage changes are can be noted from Figure C.24.

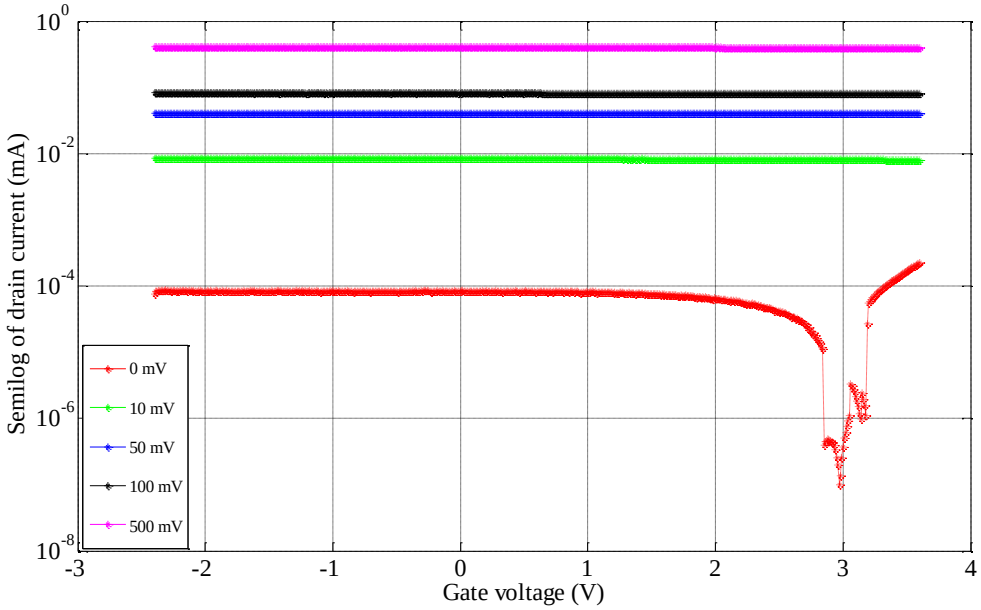


Figure C.24. Device 2 gate voltage sweep after injection with logarithmic scale.

The threshold voltage is 3 V, as shown in Figure C.24. The threshold was 3.1 V before programming and 3.2 V after tunnelling. Tunnelling reduces the threshold voltage and injection increases the threshold voltage, but based on the results achieved, injection decreased the threshold voltage. The initial charge after fabrication is unknown and different for all devices, thus there is uncertainty about the results achieved by the measurements before programming the devices as reference. The graphical results presented here are for individual devices. Section 5.3.2 presents comparison results and the discussions thereof.