

# Modeling and Design Enablement for Future Computing

*Chien-Ting Tung*



Electrical Engineering and Computer Sciences  
University of California, Berkeley

Technical Report No. UCB/EECS-2025-13

<http://www2.eecs.berkeley.edu/Pubs/TechRpts/2025/EECS-2025-13.html>

April 24, 2025

Copyright © 2025, by the author(s).  
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

Modeling and Design Enablement for Future Computing

By

Chien-Ting Tung

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Chenming Hu, Chair

Professor Sayeef Salahuddin

Professor Junqiao Wu

Spring 2025

Modeling and Design Enablement for Future Computing

Copyright © 2025  
by  
Chien-Ting Tung

## Abstract

### Modeling and Design Enablement for Future Computing by Chien-Ting Tung

Doctor of Philosophy in Engineering - Electrical Engineering and Computer  
Sciences

University of California, Berkeley  
Professor Chenming Hu, Chair

With the continued scaling of integrated circuits and the increasing number of transistors on a chip, the future IC industry faces tremendous power consumption and extreme complexity in IC design. To respond to these two challenges of future IC, I studied new devices for low-power computing and new computer-aid design with machine learning to accelerate the IC simulation. My dissertation will discuss the physics and compact modeling of new-generation memory devices for in-memory computing. Furthermore, a machine learning compact modeling framework is proposed which can accelerate the IC simulation by 5 times and more.

In Chapter 2, I will demonstrate my work on polycrystalline ferroelectric capacitor modeling. This work studied the switching dynamics of multi-grain/domain ferroelectric by calculating the accumulated switched area in a ferroelectric capacitor. The model successfully interprets the dynamics in ferroelectric material and can be used in circuit simulation for memory design. Chapter 3 extends this model to a ferroelectric tunnel junction (FTJ) with a newly developed tunneling current model. FTJ is a memristor of which resistance depends on the stored polarization. The developed model can accurately fit experimental FTJ data. Chapter 4 discusses the compact model of ferroelectric field-effect transistor (FEFET). FEFET uses ferroelectric

material as gate stack providing transistor and memory function at the same time. I developed a model that includes minor loop switching, steep switch, and inverse memory window in FEFETs with excellent fitting of the real device data. Chapter 5 discusses the physics and modeling of multi-grain/domain antiferroelectric material. Chapter 6 studies the switching dynamics of ferroelectric material at nanoscale with discrete switching behavior.

In addition to ferroelectric devices, other emerging memories are also great candidates for future computing. In Chapter 7, I show a compact model of resistive random-access memory (RRAM) which unifies the different switching mechanisms for various types of RRAMs. The model includes self-heating and disturbance effects in RRAMs and is tested with multi-level memory cell simulations. Magnetic random-access memory (MRAM) is another device I have studied. In Chapter 8, an MRAM compact model is demonstrated using a 1D Landau–Lifshitz–Gilbert (LLG) equation which keeps of physics of the full LLG equation with a speedup of 2.5 times.

From Chapter 9 to Chapter 11, I will discuss my framework of BSIM-NN, a neural network-based compact model for advanced transistors. The model uses neural networks to replace model equations of conventional transistor models. As result, it provides a generic framework for transistor modeling from FinFET, Gate-all-around (GAA) to emerging transistors such as negative capacitance FET (NCFET), which is something beyond traditional models. Moreover, the compact feature of this model can accelerate the IC simulation speed by 5 times versus the industry standard compact model. The model includes all terminal currents and charges of advanced FETs as well as the non-quasi-static effect and self-heating effect.

Chapter 12 shows a new way to do circuit simulation with physics-informed neural networks – NeuroSpice provides an alternative solution for IC simulation.

*To my family, friends, and everyone I love.*

# Content

Content.....	ii
List of Figures.....	v
Chapter 1 Introduction.....	1
1.1 Challenges of Future Electronics .....	1
1.2 Compact Model.....	3
Chapter 2 A Compact Model of Polycrystalline Ferroelectric Capacitor ....	5
2.1 Motivation .....	5
2.2 Model .....	7
2.3 Validation.....	9
2.4 Scalability.....	12
2.5 Chapter Summary .....	14
Chapter 3 A Compact Model of Metal-Ferroelectric-Insulator-Semiconductor Tunnel Junction.....	15
3.1 Motivation .....	15
3.2 Model .....	16
3.3 Discussion .....	22
3.4 Chapter Summary.....	25
Chapter 4 A Compact Model of Ferroelectric Field-effect Transistor.....	26
4.1 Motivation .....	26
4.2 Compact Model.....	28
4.3 Results .....	31
4.4 Chapter Summary.....	33
Chapter 5 A Compact Model of Antiferroelectric Capacitor.....	34
5.1 Motivation .....	34
5.2 Model .....	35
5.3 Results .....	38
5.4 Chapter Summary.....	41

Chapter 6	Theoretical Study and Modeling of Nanoscale Ferroelectric Capacitor.....	42
6.1	Motivation .....	42
6.2	Phase-field Simulation .....	44
6.3	Compact Model.....	46
6.4	Chapter Summary.....	51
Chapter 7	A Versatile Compact Model of Resistive Random-Access Memory (RRAM).....	52
7.1	Motivation .....	52
7.2	Model .....	54
7.3	Model Validation .....	57
7.4	Chapter Summary.....	65
Chapter 8	A Compact Model of Perpendicular Spin-Transfer-Torque Magnetic Tunnel Junction .....	66
8.1	Motivation .....	66
8.2	Model .....	67
8.3	Model Validation.....	72
8.4	Chapter summary .....	79
Chapter 9	BSIM-NN: Neural Network-Based BSIM Transistor Model Framework for Advanced and Emerging Technology .....	80
9.1	Motivation .....	80
9.2	Modeling Framework.....	81
9.3	Training & Inference Results.....	84
9.4	Application for Advanced Technology.....	91
9.5	Speed Comparison & Circuit Simulation .....	94
9.6	Chapter Summary.....	100
Chapter 10	Non-Quasi-Static Modeling of Neural Network-based Transistor Compact Model for Fast Transient, AC, and RF Simulations .....	101
10.1	Motivation .....	101
10.2	Model.....	102
10.3	Validation .....	105
10.4	Chapter Summary.....	110

Chapter 11	A Novel Neural Network-based Transistor Compact Model	
	Including Self-Heating .....	111
11.1	Motivation .....	111
11.2	Model.....	112
11.3	Validation .....	114
11.4	Chapter Summary .....	119
Chapter 12	NeuroSpice: IC Simulator Using Physics-Informed Neural	
	Network .....	120
12.1	Motivation .....	120
12.2	Framework.....	121
12.3	Result & Discussion .....	123
12.4	Chapter Summary .....	129
Chapter 13	Summary .....	130
13.1	Chapter Summary.....	130
13.2	Future Work .....	131
Bibliography	.....	132

## List of Figures

Fig. 1.1. The scaling trend of past, present and future electronics. Reprint with permission from [1]. .....	2
Fig. 1.2. Key device parameters and performance metrics comparing various embedded memory candidates. Reprint with permission from [1]. .....	2
Fig. 1.3. The architecture of BSIM-CMG. Reprint with permission from [2]. .....	4
Fig. 2.1. Modeled current waveforms versus time for several voltages of a 8.5 nm HZO capacitor agree with experimental data from [16] with $PR = 20 \mu C/cm^2$ , $\tau_0 = 100 ps$ , $Ea = 8.0 MV/cm$ and a $135 \Omega$ series resistance. These parameters are the same as used in [16]. © 2025 IEEE.....	9
Fig. 2.2. Fitted P-t curves of a 8.3 nm HZO capacitor from Ref. [22] with $PR = 22.9 \mu C/cm^2$ , $\tau_0 = 390 ns$ , $Ea = 1.74 MV/cm$ , $\alpha = 3.48$ , $\beta = 2.0$ and a offset voltage of -0.08 V. © 2025 IEEE.....	10
Fig. 2.3. (a) The simulation for the FE capacitor while applying a triangular waveform using the parameters extracted from Fig. 2.2. (b) Comparison between the experimental and simulated P-V curve from the voltage pulse shown in Fig. 2.3a. The difference between the two data comes from the assumption of constant permittivity. © 2025 IEEE.....	11
Fig. 2.4. The validation of the polarization accumulation by comparing the simulation results with the measurement in Ref. [22]. Pulse trains with $1 \mu s$ pulse width at different amplitudes are applied to the capacitor. © 2025 IEEE .....	12
Fig. 2.5. Discrete switching behavior for a device with $10 \eta$ groups. There are 4 states for this device different from the continuous states in large-area	

capacitors as Fig. 2.2. © 2025 IEEE.....	13
Fig. 2.6. The simulation of 10 devices with 10 $\eta$ groups which has a large device variability. © 2025 IEEE.....	13
Fig. 3.1. Schematic band diagram of an MFIS-FTJ and its equivalent circuit. © 2025 IEEE.....	17
Fig. 3.2. Comparison of our initial guess of $\psi_s$ and the $Q_s - \psi_s$ curve obtained from the analytical equation Eq. (3.1). © 2025 IEEE.....	19
Fig. 3.3. Comparison of the tunneling current model and the WKB approximation. The symbols are the WKB approximation. The lines are the compact model. For the black line, the parameters are $a = 0.52$ , $b = 0.48$ , $c = 1$ , $d = 30$ , $e = 1.8$ , and $f = 0.008$ . For the red lines the following parameters are different: (a) $d = 20$ , $e = 1.2$ ; (b) $d = 20$ , and $e = 2$ ; (c) (d): All parameters are the same. © 2025 IEEE.....	22
Fig. 3.4. Fitted results of (a) an n-type MFIS-FTJ [29] and (b) a p-type MFIS-FTJ [30]. Symbols are experimental data and lines are the simulation. For n-type, $a_+ = 0.6$ , $a_- = 0.5$ , $b_+ = 0.52$ , $b_- = 0.3$ , $c_+ = 1.2$ , $c_- = 0.7$ , $d_+ = 11$ , $d_- = 40$ , $e_+ = e_- = 3$ , and $f_+ = f_- = 0$ . For p-type, $a_+ = a_- = 0.5$ , $b_+ = -0.2$ , $b_- = -0.4$ , $c_+ = 0.8$ , $c_- = 0.22$ , $d_+ = 2$ , $d_- = 15$ , $e_+ = e_- = 1$ , and $f_+ = f_- = 0$ . © 2025 IEEE.....	23
Fig. 3.5. The write and read test of the FTJ. The inset is the sample circuit. We show the corresponding reading current and polarization to this read/write operation. © 2025 IEEE.....	24
Fig. 3.6. Transient test of this model by applying triangular waveform with varying amplitude for 1000 cycles. The simulated polarization and tunneling current are shown in (a) the 1 <sup>st</sup> cycle (b) the 1000 <sup>th</sup> cycle. © 2025 IEEE ....	25
Fig. 4.1. (a) The equivalent circuit model of FEFET and (b) the charge	

equality between FE and MOSFET. © 2025 IEEE .....	28
Fig. 4.2. Fitted $I_D$ - $V_G$ curves compared to the experimental data of (a) FDSOI-FEFET with $W=170\text{nm}$ , $L=24\text{nm}$ , $t_{\text{FE}}=10\text{nm}$ , $T=300\text{K}$ , $E_{a0+}=2.7\text{MV/cm}$ , $E_{a0-}=3.7\text{MV/cm}$ , $E_{ah+}=3.5\text{MV/cm}$ , $E_{ah-}=3.8\text{MV/cm}$ , $a=2$ , $b=-0.1$ , $c_{1,2}=1.2$ , $d_1=20$ , $d_2=500$ , $e_1=-0.5$ , and $e_2=0.4$ [52] (b) Fin-FEFET with $H_{\text{FIN}}=30\text{nm}$ , $W_{\text{FIN}}=50\text{nm}$ , $L=70\text{nm}$ , $t_{\text{FE}}=10\text{nm}$ , $E_{a0+}=4\text{MV/cm}$ , $E_{ah+}=4.6\text{MV/cm}$ [60] where symbols are the measurement and lines are the simulation. © 2025 IEEE .....	32
Fig. 4.3. (a) Fitted $I_D$ - $V_G$ curves for different minor loops of the FDSOI-FEFET [52] where symbols are the measurement and lines are the simulation with $W=170\text{nm}$ , $L=24\text{nm}$ , , $t_{\text{FE}}=10\text{nm}$ , $T=300\text{K}$ , $E_{a+}=3\text{MV/cm}$ , $E_{a-}=3.2\text{MV/cm}$ , $a=2$ , $b=-0.2$ , $c_{1,2}=1.2$ , $d_1=20$ , $d_2=500$ , $e_1=-0.3$ , $e_2=0.45$ , $m_1=600$ , $m_2=0.6$ , $m_3=1e-4$ , $m_4=1.5$ , $m_5=0.33$ , and $\sigma=0.33$ . The maximum gate sweep voltages are 1.0V, 1.2V, 1.6V and 3.0V. (b) The QV loops correspond to the IV curves. © 2025 IEEE .....	32
Fig. 5.1. (a) The schematic hysteresis loop of AFE which is divided into 4 regions: (I) $E_{\text{AFE}} < -ET -$ , (II) $-ET - \leq E_{\text{AFE}} \leq ET +$ & $P > 0$ , (III) $-ET - \leq E_{\text{AFE}} \leq ET +$ & $P < 0$ , and (IV) $E_{\text{AFE}} > ET +$ ; (b) The percentage of $Al +$ and $Ah +$ over $Al$ and $Ah$ under different $E_{\text{AFE}}$ ; (c) The schematic transient diagram of AFE. © 2025 IEEE .....	37
Fig. 5.2. Modeled hysteresis curve of a 10 nm $\text{ZrO}_2$ capacitor from [74] with $PR = 11\mu\text{C/cm}^2$ , $\tau_0 = 100\text{ps}$ , $E_a = 3.45\text{MV/cm}$ , $ET+ = ET- = 2.26\text{MV/cm}$ and $\alpha = 2.46$ . © 2025 IEEE .....	39
Fig. 5.3. Modeled hysteresis curve of a 6 nm HZO capacitor from [62] with $PR = 20\mu\text{C/cm}^2$ , $\tau_0 = 1.5\text{ns}$ , $E_a = 3.6\text{MV/cm}$ , $ET+ = ET- = 1.8\text{MV/cm}$ and $\alpha = 1.7$ . © 2025 IEEE .....	40

Fig. 5.4. Modeled hysteresis curve of a TiN/TiO <sub>2</sub> /HZO/RuO <sub>2</sub> /HfO <sub>2</sub> -stacked capacitor from [62] with $PR = 20\mu C/cm^2$ , $\tau_0 = 2ns$ , $Ea = 3.7MV/cm$ , $ET+ = 1.9MV/cm$ , $ET- = 2.3MV/cm$ and $\alpha = 1.7$ . © 2025 IEEE.....	40
Fig. 6.1. Polarization switching of a single-grain FE capacitor at 1.2 V and 1.06 V. The legend shows the polarization in the unit of C/m <sup>2</sup> . © 2025 IEEE .....	44
Fig. 6.2. Polarization switching of a 3-grain FE capacitor at 1.4 V. The legend shows the polarization in the unit of C/m <sup>2</sup> . © 2025 IEEE.....	46
Fig. 6.3. The equivalent circuit of our nanoscale FE capacitor compact model. © 2025 IEEE .....	48
Fig. 6.4. Model fitting of the single-grain FE capacitors under 3 cases: (a) 1 grain 1 nucleus $g=1e-10 m^3/F$ , (b) 1 grain 1 nucleus $g=1e-11 m^3/F$ , (c) 1 grain 2 nuclei $g=1e-10 m^3/F$ . The symbols are the COMSOL simulation, and the lines are the compact model. (d) shows the comparison between three cases. © 2025 IEEE .....	48
Fig. 6.5. Model fitting of the 3-grain FE capacitor. © 2025 IEEE.....	50
Fig. 6.6. The comparison between the COMSOL simulated PV curve and the compact model simulations using linear $Q_B$ and nonlinear $Q_B$ . © 2025 IEEE .....	50
Fig. 7.1. The schematic diagram and equivalent circuit of the conducting path in a resistive memory.....	55
Fig. 7.2. Model validation of a HfO <sub>x</sub> /AlO <sub>x</sub> bilayer RRAM [93]. The symbols are experimental data. The lines are model simulations.....	58
Fig. 7.3. Model validation with a transient measurement of a HfO <sub>x</sub> /AlO <sub>x</sub> bilayer RRAM [93]. Black and red lines are experimental data. The blue line is the model simulation with $C_{th}$ . The green dash line is the model simulation without $C_{th}$ .....	58

Fig. 7.4. Model simulated dynamics of temperature corresponding to Fig. 7.3. .....	59
Fig. 7.5. Model simulations using current ramps to several different maximum current. ....	60
Fig. 7.6. Model validation with a TiN/TiO <sub>x</sub> /HfO <sub>x</sub> /Pt planar RRAM [95] being reset to several V <sub>RESET</sub> . ....	61
Fig. 7.7. Model validation of a TiN/TiO <sub>x</sub> /HfO <sub>x</sub> /Pt planar RRAM [95] at different I <sub>COMP</sub> . The symbols are experimental data. The lines are model simulations. ....	61
Fig. 7.8. Model validation of Al-doped HfO <sub>x</sub> RRAMs [95] with different AlO <sub>x</sub> concentrations. The symbols are experimental data. The lines are model simulations. ....	62
Fig. 7.9. Model validation of a CBRAM [92]. The symbols are experimental data. The lines are model simulations.....	63
Fig. 7.10. Model validation of a Ag/Ta <sub>2</sub> O <sub>5</sub> /Pt RRAM which experienced both metal ion and oxygen vacancy switching [94]. The symbols are experimental data. The lines are model simulations.....	63
Fig. 7.11. Model validation of RRAM variation and disturbance [88]. ....	64
Fig. 7.12. 1T1R multilevel READ/WRITE simulation of a RRAM. This is 1T1R cells simulation for 40 ns with a maximum time step of 1 ps on an Intel Xeon Platinum 8260 processor. The simulation took 7.62 seconds.....	65
Fig. 8.1. Equivalent circuit of this model. The direction of the magnetic moment and the voltage polarity are shown. A separate node is used to output the state probability (Prob <sub>AP</sub> ). © 2025 IEEE .....	68
Fig. 8.2. Validation of the geometry dependence resistance model with the data (symbols) from [133]. The lines are the model results. © 2025 IEEE .	69
Fig. 8.3. Validation of the temperature dependence resistance model with the	

data (symbols) from [134]. The lines are the model results. © 2025 IEEE . 70

Fig. 8.4. Validation of this model (lines) versus the LLG model (symbols) [120] under constant write/erase voltage. The same set of parameters is used for Fig. 8.5. The LLG inputs are  $M_S=700$  emu/cm<sup>3</sup>, uniaxial anisotropy=56  $k_B T$ ,  $a=0.028$ , area= $25^2\pi$  nm, and free-layer thickness=1.4 nm. In the model, we use  $M_S=700$  emu/cm<sup>3</sup>,  $H_K=3000$  Oe,  $a=0.01$ ,  $c=0.2$ ,  $\eta_P=0.2275$ , and  $\eta_{AP}=0.214$  to fit the data. The lines are the model results. © 2025 IEEE .... 73

Fig. 8.5. Demonstration of the model predictive ability of physical parameters. The model is compared with the LLG where all the other parameters are the same as Fig. 8.4. Only free layer thickness ( $T_{FL}$ ) and  $M_S$  are changed for (a) and (b) respectively. © 2025 IEEE..... 74

Fig. 8.6. This model works under an arbitrary waveform. The validation of the tunneling current and  $m_Z$  versus LLG is shown using the same parameters as Fig. 8.4. A probability of AP state is also calculated which switches continuously between 0 and 1 and responds to the varying of applied waveform. © 2025 IEEE ..... 74

Fig. 8.7. Validation of the pulse width dependent critical current ( $I_C$ ). The lines are the model, and the symbols are the experimental data [138, 139]. Red:  $M_S=450$  emu/cm<sup>3</sup>,  $H_k=650$  Oe,  $V_F=1.17 \times 10^{-23}$  m<sup>-3</sup>,  $\alpha=1.5$ ,  $c=0.065$ ,  $\eta_P=0.183$ , and  $\eta_{AP}=0.465$ . Green:  $M_S=700$  emu/cm<sup>3</sup>,  $H_k=800$  Oe,  $V_F=4.245 \times 10^{-24}$  m<sup>-3</sup>,  $\alpha=0.03$ ,  $c=1$ ,  $\eta_P=0.183$ , and  $\eta_{AP}=0.143$  © 2025 IEEE..... 76

Fig. 8.8. Validation of the experimental RV and transient characteristics [140] with  $M_S=800$  emu/cm<sup>3</sup>,  $H_k=800$  Oe,  $V_F=1.25 \times 10^{-23}$  m<sup>-3</sup>,  $a=0.012$ ,  $c_i=0.3$ ,  $\eta_P=0.17$ , and  $\eta_{AP}=0.316$ . © 2025 IEEE..... 77

Fig. 8.9. The validation of the switching probability of the model (lines) versus experimental data (symbols) [119] with  $M_S=796$  emu/cm<sup>3</sup>,  $H_k=2000$

Oe, $V_F=1.625 \times 10^{-24} \text{ m}^{-3}$ , $\alpha=0.014$ , $c=0.6$ , and $\eta_p=0.183$ . © 2025 IEEE.....	77
Fig. 8.10. The validation of the temperature dependence model. The data comes from [141]. © 2025 IEEE .....	78
Fig. 9.1. The schematic diagram of the NN FET model, the loss function, and the Verilog-A implementation. The compact model has both IV and QV networks. The loss function includes up to second derivatives of outputs (y) and it can be adjusted by tuning the weights (a-f) of each term. Our implementation uses direct multiplication instead of loops. © 2025 IEEE .	84
Fig. 9.2. The fitted $I_D V_G$ curves at different L, $H_{FIN}$ , and EOT where the lines are the NN and symbols are the BSIM-CMG data. © 2025 IEEE .....	85
Fig. 9.3. The fitted $I_D V_D$ curves at different L, $H_{FIN}$ , and EOT where the lines are the NN and symbols are the BSIM-CMG data. © 2025 IEEE .....	86
Fig. 9.4. The fitted $I_G$ characteristics for different structures where the lines are the NN and symbols are the BSIM-CMG data. © 2025 IEEE .....	86
Fig. 9.5. CV fitting of $Q_G$ , $Q_S$ and $Q_D$ versus $V_{DS}$ for different L, $H_{FIN}$ and EOT. © 2025 IEEE .....	87
Fig. 9.6. CV fitting of $Q_G$ , $Q_S$ and $Q_D$ versus $V_{GS}$ for different L, $H_{FIN}$ and EOT. © 2025 IEEE .....	87
Fig. 9.7. $I_{ON}$ distribution relative to mean by Monte Carlo simulation for (a) $\sigma$ of $L = 0.54\text{nm}$ , (b) $\sigma$ of $H_{FIN} = 1.38\text{nm}$ , (c) $\sigma$ of $EOT = 0.04\text{nm}$ , and (d) $\sigma$ of $\Delta\phi = 0.0167\text{eV}$ . The symbols are the data generated with BSIM-CMG and the lines are the predictions of the NN model. In the parentheses, we show the error rate of $\mu$ and $\sigma$ between NN and BSIM-CMG. © 2025 IEEE.....	88
Fig. 9.8. (a) Gummel test at 4 <sup>th</sup> derivative. (b) Harmonic balance test. The slope of each line meets the theoretical prediction. © 2025 IEEE.....	89
Fig. 9.9. Thermal noise comparison between BSIM-NN and BSIM-CMG.	90
Fig. 9.10. Flicker noise comparison between BSIM-NN and BSIM-CMG.	90

Fig. 9.11. Output noise simulation validation of a common-source amplifier and a 3-stage ring oscillator. © 2025 IEEE .....	91
Fig. 9.12. The $I_D V_G$ characteristics of the NN model (lines) versus the TCAD data (symbols). © 2025 IEEE .....	92
Fig. 9.13. The $I_D V_D$ characteristics of the NN model (lines) versus the TCAD data (symbols). © 2025 IEEE .....	92
Fig. 9.14. QV and CV characteristics of the NN model (lines) versus the TCAD data (symbols). © 2025 IEEE .....	93
Fig. 9.15. The fitted $I_D V_G$ and $I_D V_D$ characteristics of NC-GAAFET where the lines are the model, and the symbols are the TCAD data. © 2025 IEEE.....	93
Fig. 9.16. Evaluation time comparison between NN and BSIM-CMG IV model. © 2025 IEEE.....	95
Fig. 9.17. (a) The inference time versus the number of hidden layers with 10 nodes per layer. (b) The inference time versus the number of nodes per layer where 3 hidden layers are set. © 2025 IEEE .....	95
Fig. 9.18. A 3-input NAND simulation with different slew rates. ....	96
Fig. 9.19. A 3-input NAND simulation with different fan outs. ....	97
Fig. 9.20. A 5-stage FO4 inverter chain simulation. The load goes from 1 to 256 inverters.....	97
Fig. 9.21. A D-flip-flop simulation.....	98
Fig. 9.22. A FO4 151-stage NAND ring oscillator simulation. ....	98
Fig. 9.23. A 32-bit ripple adder simulation. ....	99
Fig. 9.24. A SRAM READ SNM (static noise margin) variation simulation. BSIM-NN is 2 times faster than BISM-CMG. ....	99
Fig. 9.25. Speed comparison of BSIM-NN and BSIM-CMG with different benchmarking circuits. The speed boost is around 3~5X.....	100
Fig. 10.1. (a) QS IV curves of the NN model. (b) QS CV curves of the NN	

model. (c) The schematic diagram of NQS modeling flow. © 2025 IEEE	103
Fig. 10.2. (a) Model charges trained without offset included. (b) Model charges trained with offset included at $V_{DS}=0.05V$ (c) $V_{DS}=0.7V$ . © 2025 IEEE	103
Fig. 10.3. Transient simulation result when a pulse with 2ps-rise/fall time, 2ps-pulse width, and 0.7V amplitude is applied to the gate. $V_{DS}=0.05V$ & 0.7V. $I_{D,NQS}$ and $I_{S,NQS}$ are the currents without the parasitic components. © 2025 IEEE	105
Fig. 10.4. AC simulation result of a common-source amplifier with 1kW load resistance. © 2025 IEEE	106
Fig. 10.5. S-parameter simulation with a 50 W load impedance. © 2025 IEEE	107
Fig. 10.6. RF Pin-Pout simulation of the NN model and BSIM-CMG. © 2025 IEEE	107
Fig. 10.7. Simulated $f_T$ and $f_{max}$ from BSIM-CMG and the NN model. © 2025 IEEE	108
Fig. 10.8. Killer NOR gate simulation. © 2025 IEEE	109
Fig. 10.9. 151-stage NAND oscillator simulation.	109
Fig. 11.1. The subcircuit of the temperature relaxation model and the simulated $T(t)$ , $T_0'(t)$ , and $dT(t)$ . from the traditional self-heating model and the temperature relaxation model. A step function voltage is applied. © 2025 IEEE	114
Fig. 11.2. Validation of transfer characteristics at different temperatures. The lines are the NN model, and the symbols are the BSIM-CMG data. © 2025 IEEE	115
Fig. 11.3. Validation of output characteristics at different temperatures. The lines are the NN model, and the symbols are the BSIM-CMG data. © 2025	

IEEE.....	115
Fig. 11.4. Validation of CV characteristics at different temperatures. The lines are the NN model, and the symbols are the BSIM-CMG data. © 2025 IEEE .....	116
Fig. 11.5. Validation of the SH-free characteristics recovered by the temperature relaxation (TR) model. © 2025 IEEE.....	117
Fig. 11.6. Transient pulse simulations of BSIM-CMG, the NN model w/ and w/o TR model, the extracted SH-free NN with conventional SH model, and the case that $R_{TH} \neq R'_{TH}$ . © 2025 IEEE.....	118
Fig. 11.7. 1001-stage NAND oscillator simulation. ....	119
Fig. 12.1. The schematic diagram of NeuroSpice. The loss function is the DAE of the circuit and the initial condition. The neural network will update its parameters until the DAE is optimized to 0. I.C. is the initial condition. ..	121
Fig. 12.2. Example DAE of a transistor amplifier and its loss function. In this case, the KCL of each node will be solved including the current of the nonlinear device model (nmos). The time derivative is computed by the autograd function in simulation. ....	122
Fig. 12.3. The pseudocode of the device model. The MOSFET DC and switching currents are computed and returned to the loss function. The charge currents are also analytically computed with autograd. ....	122
Fig. 12.4. The simulation of a series RLC circuit with a sinusoid input. ...	124
Fig. 12.5. The simulation of a non-inverting amplifier implemented with an operational amplifier circuit.....	124
Fig. 12.6. The simulation of a transistor amplifier. ....	125
Fig. 12.7. The simulation of a CMOS inverter.....	126
Fig. 12.8. The simulation of a CMOS NAND.....	126
Fig. 12.9. The simulation of a 5-stage ring oscillator.....	127

Fig. 12.10. Polarization-voltage characteristics of a ferroelectric Landau-Khalatnikov (LK) model. It contains positive, negative, and infinite slopes.

..... 128

Fig. 12.11. The simulation of a FeRAM using the LK model..... 128

## Acknowledgment

I would like to thank my advisor, Prof. Chenming Hu. He supported my Ph.D. journal all the way. He not just cared about my academic success but about my life in Berkeley. I remember every time we met, he always checked on my body weight and he even purchased a scale for me. He is also my role model in research. I have learned a lot from his way of approaching problems, physics knowledge, group management, and presentation skills. I am always surprised by his understanding of semiconductor physics which inspired me to deepen my semiconductor knowledge. He fully supported what I wanted to do and provided me with insights on how this research would be more attractive to other people. I will miss the time we revised papers and slides together. He also gave me lots of advice on my career in both industry and academia. I am very glad I chose UC Berkeley to work with Prof. Hu. I wish to follow his footsteps and one day to make great contributions to the semiconductor industry.

I want to thank Prof. Sayeef Salahuddin to guild me through this journey. He has provided me with many insights and suggestions on ferroelectrics and machine learning. He is very knowledgeable. I have taken two of his classes and learned different aspects of understanding semiconductor and quantum physics that many of his thinking and explanations are something I would like to use in my own teaching. He also supported my job finding in academics he promoted my work in his talk and provided me the opportunity to speak at the CMC meeting.

I would like to thank my preliminary exam committee, Prof. Jeffrey Bokor and Prof. Ali Javey. Especially Prof. Bokor, he helped me a lot during my prelim preparation. Without them, I could not pass the prelim exam. This process built a solid foundation for my semiconductor knowledge. Then, I want to thank my qualifying exam committee, Prof. Chenming Hu, Prof. Sayeef Salahuddin, Prof. Jeffrey Bokor, and Prof. Junqiao Wu. Their advice provides guidelines for my following research.

Next, I want to express my gratitude to my family, my mother and father,

who provide everything I need so I can focus on my studies. They took me to explore more about Taiwan every time I went back and let me bring back many snacks.

I also want to thank my girlfriend, Ting Xiao. She has provided me with a lot of emotional support during the last few years of my Ph.D. journey. We together explore many fun places in the Bay Area and other parts of the US. Without her, my life here would be much more boring.

I want to thank my friends and lab mates in Berkeley. Girish, Ming-Yen, Jen-Hao, Avirup, Chetan, Ahtisham, Dinesh, Fredo, and Sourabh. You guys have helped me a lot with my research. Lars, Jason, Junhao, Chia-Chun, and Wei-Yang. I enjoyed the time when we discussed devices and Cal's life.

Finally, there are too many people to thank, so I'll just thank this wonderful world.

# Chapter 1 Introduction

## 1.1 Challenges of Future Electronics

Since the 1900s, the number of transistors per integrated circuit doubled every 12-24 months as Moore's law projects (Fig. 1.1) [1]. To keep scaling the devices, new technologies are introduced such as strain silicon, High-K metal gate, FinFET, Gate-all-around (GAA), and 3D integrated circuit (IC). However, as we enter the sub-5nm node, the conventional geometry scaling cannot keep up with Moore's law. People start to investigate new materials and new architecture to continue scaling. One big challenge for future IC is the memory wall. Memory wall is saying that the speed and bandwidth of memory units cannot keep up with process units. This latency limits the performance of the chip. Moreover, energy is wasted during the data movement between two units. Therefore, embedded non-volatile memory is essential to achieve low-power in-memory computing. Traditional memory such as SRAM, DRAM, and Flash are not suitable for embedded memory applications due to their power consumption (Fig. 1.2) [1]. Emerging memories with new materials become promising candidates for in-memory and neuromorphic computing. Devices like ferroelectric memory, resistive memory (RRAM), and magnetic memory (MRAM) have the advantages of low energy and power consumption, small cell area, and fast Read/Write speed to reduce the overall computing power. The first part of my dissertation will discuss the physics and compact modeling of these devices.

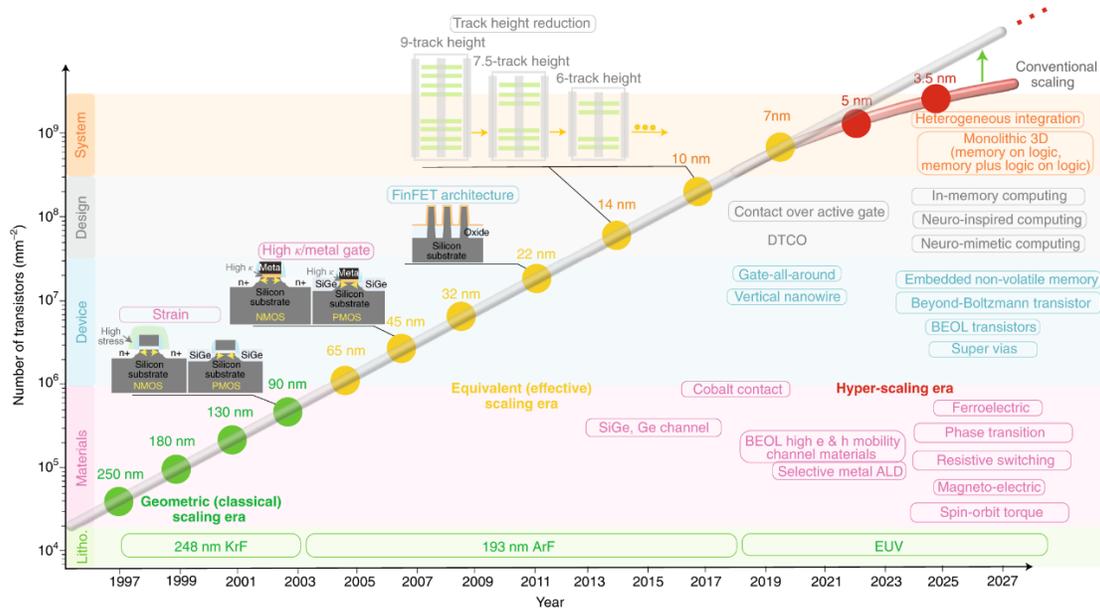


Fig. 1.1. The scaling trend of past, present and future electronics. Reprint with permission from [1].

	eSRAM	eDRAM	eFLASH	STT-MRAM	FeRAM	FeFET	PCRAM	RRAM
Cell size	120–150 $F^2$	10–30 $F^2$	10–30 $F^2$	10–30 $F^2$	10–30 $F^2$	10–30 $F^2$	10–30 $F^2$	10–30 $F^2$
Cell structure	6T	1T–1C	1T	1T–1MTJ	1T–1C	1T	1T–1PCM	1T–1R
Non-volatility	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Write voltage	<1 V	<1 V	~10 V	<1.5 V	<3 V	<4 V	<3 V	<3 V
Write energy	~fJ	~10 fJ	~100 pJ	~1 pJ	~0.1 pJ	~0.1 pJ	~10 pJ	~1 pJ
Standby power	High	Medium	Low	Low	Low	Low	Low	Low
Write speed	~1 ns	~10 ns	0.1–1 ms	~5 ns	~10 ns	~10 ns	~10 ns	~10 ns
Read speed	~1 ns	~3 ns	~10 ns	~5 ns	~10 ns	~10 ns	~10 ns	~10 ns
Endurance	$10^{16}$	$10^{16}$	$10^4$ – $10^6$	$10^{15}$	$10^{14}$	$>10^5$	$>10^{12}$	$>10^7$

Fig. 1.2. Key device parameters and performance metrics comparing various embedded memory candidates. Reprint with permission from [1].

The other challenge of future IC lies in the time of IC design. As the transistor count on a chip keeps increasing, future ICs can contain trillions of transistors. The time consumption to design such a chip will become

unacceptable with current design tools. A way to accelerate the circuit simulations for IC design is essential to enable future design. One time-consuming part of circuit simulation is the evaluation time for device models. In the second part of my dissertation, I study how to use machine learning to accelerate model evaluation by developing neural network-based transistor compact models. The complex device equations are replaced by simple matrices to speed up the IC simulation. The demonstrated speed improvement is 5 times compared to the industry standard transistor compact model which has a great impact on the future IC industry.

## 1.2 Compact Model

A major part of my dissertation is about compact models. What is the compact model? Compact models are the bridge between the foundry and the IC design house. Foundries will carefully calibrate their device technology with a compact model and embed the calibrated model in a process design kit (PDK). The PDK will be sent to design houses to represent foundries' technology that IC designers can follow. To ensure design accuracy, the compact model has to be accurate in different operation regions and geometries. Fig. 1.3 shows the architecture of the industry standard compact model of FinFET/GAA – BSIM-CMG [2]. In addition to simple IV and CV models, models for side effects and nonidealities are the keys for modern transistors. Furthermore, the model evaluation speed needs to be fast for large IC simulations. Therefore, conventional compact models are built from analytical equations with many fitting parameters for accuracy.

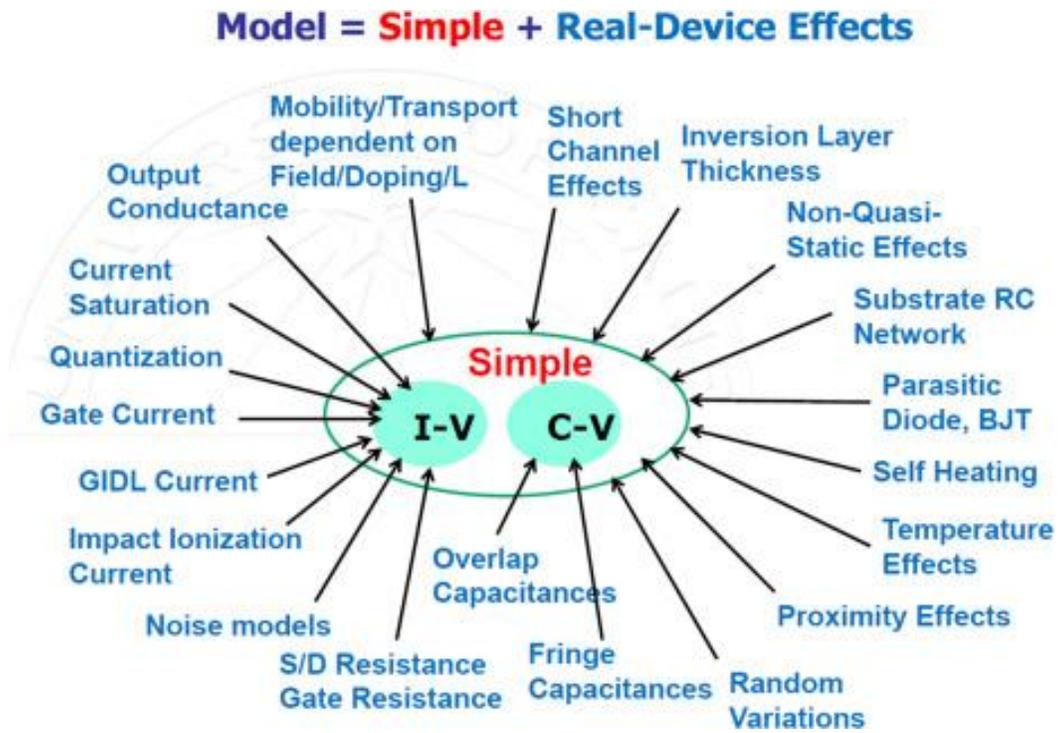


Fig. 1.3. The architecture of BSIM-CMG. Reprint with permission from [2].

## **Chapter 2 A Compact Model of Polycrystalline Ferroelectric Capacitor**

The polycrystalline thin film ferroelectric capacitor is modeled as a collection of independent grains or grain groups. Each grain or grain group is characterized by its local field-dependent switching rate, characterized by a distribution function such as Gaussian or type-2 generalized beta distribution. This computationally efficient model accurately reproduces the published experimental polarization switching waveforms and the switching current waveforms in response to all reported applied voltage waveforms. The model tracks the polarization history so that it can simulate the transition between major and minor and among minor loops as well as accumulative polarization which the conventional models cannot capture. This model is intended to simulate circuits containing ferroelectric capacitors using commercial SPICE simulators for arbitrary applied voltage waveforms. It also has the capability of simulating the discrete switching and device variability in small-area FE capacitors having a small number of grains, although there is no available experimental data to check the model accuracy in this regard. This work is published by IEEE. © 2025 IEEE. Reprinted, with permission, from [3]

### **2.1 Motivation**

The ferroelectricity in HfO<sub>2</sub> was discovered in 2011 [4, 5]. Due to its compatibility with the standard CMOS technology, ferroelectric (FE) HfO<sub>2</sub> thin film has been studied for use in promising nonvolatile memories such as the FE random access memory (FERAM), FE field effect transistor and FE tunnel junction [6, 7]. The continuous nature and accumulation properties of FEs also make them suitable for neuromorphic applications [8]. To design and analyze these memory circuits, a compact model that can capture FE

switching dynamics is therefore needed.

Several models have been proposed to describe the relation between the electric field and the switching time of the FE polarization. Kolmogorov–Avrami–Ishibashi (KAI) model is one of the widely used models among them [9, 10]. The KAI theory describes polarization switching through unrestricted growth of reverse domains nucleated from different non-interacting nucleation centers in the FE material. Thus, KAI theory is ideally suitable only for bulk FE materials and cannot describe polycrystalline FEs with multiple grains such as those based on doped HfO<sub>2</sub> [11]. Therefore, another theory called the inhomogeneous field mechanism (IFM) or nucleation-limited switching (NLS) model was proposed to model the polycrystalline FEs [12–17]. These models assume there are many small grains in the FE material and each grain has its own switching time which is determined by the local material properties. This local switching time is assumed to follow some statistical distribution.

Polarization expressions as a function of time and electric field used in these models are arrived at by assuming a constant electric field. These models have been shown to fit the experimental data of hafnium zirconium oxide (HZO) under a constant voltage pulse, but they are not able to handle arbitrary input signals. Furthermore, these models assume the polarization reversal starts from a fully polarized state. Thus, they cannot model the minor loop switching and accumulation properties [18]. Other models such as Preisach model need to use empirical interpolation and scaling of model parameters to produce partial switching, which is not physical [19]. Although with more intensive phase-field-based Landau models, the accuracy of the simulations can be improved, they are computationally expensive and therefore not very useful for the simulation of real circuits containing a large number of FE memory elements [20].

To simulate FE under arbitrary inputs, a model using Monte Carlo simulation was proposed in [21, 22]. Based on multi-grain dynamics, the FE is divided into many small physical grains and each grain has its switching probability. By introducing an accumulated time constant, the model could handle arbitrary waveforms and showed the accumulative polarization reversal [22, 23]. However, this modeling framework is computationally

intensive. The usual number of physical grains in the FE material is up to hundreds or thousands. To achieve the much higher computational efficiency required for simulating integrated circuits, this chapter presents a compact model of polycrystalline FE capacitors that can capture the switching dynamics of FE under arbitrary input waveforms in good agreement with experimental results and computationally suitable for circuit design purposes.

## 2.2 Model

I assume that the FE film consists of many crystalline grains or grain groups. Each group is characterized by a material parameter  $\eta$ , which characterizes the random variation of the local switching rates in Eq. (2.2) such as the variations in crystal grain orientation, stress, or stoichiometry. The area of the grain or grain group that has a material parameter  $\eta$  is  $A_\eta$  whose value is total film area  $A_T$  times the probability of  $\eta$ ,  $P(\eta)$ . Part of  $A_\eta$  is positively polarized and the rest is negatively polarized. The positively polarized area is denoted as  $A_{\eta+}$  and the negatively polarized area is  $A_{\eta-}$ . The positively and negatively polarized areas can increase or decrease with time under an applied voltage. However, their sum is always  $A_\eta$ . Assume that a positive voltage is applied. Previously published models and experimental studies show that the growth of  $A_{\eta+}$  follows an exponential time dependence at a fixed positive applied voltage. This suggests that the rate equation of polarization switching of the FE material is a simple first order differential equation as

$$\frac{dA_{\eta+}}{dt} = \frac{A_{\eta-} - A_{\eta+}}{\tau(t, \eta)} \text{ if } E_{FE}(t) \geq 0, \quad \frac{dA_{\eta+}}{dt} = \frac{-A_{\eta+}}{\tau(t, \eta)} \text{ if } E_{FE}(t) < 0 \quad (2.1)$$

The value of  $\eta$  is experimentally observed between 0 to 2 with a statistical distribution and a unity mean [14, 22]. Therefore, our model limits  $\eta$  from 0 to 2. It influences the switching characteristic time  $\tau(t, \eta)$  given by the Merz's law [14, 16, 17] as

$$\tau(t, \eta) = \tau_0 \exp \left[ \left( \frac{\eta E_a}{|E_{FE}(t)|} \right)^\alpha \right] \quad (2.2)$$

where  $E_a$  is the activation field which can be used as a fitting parameter to fit

experimental data or be substituted with a theoretical value,  $E_{FE}$  is the applied electric field,  $\tau_0$  is the characteristic time for a very large  $E_{FE}$ , and  $\alpha$  is a fitting parameter.

By solving Eq. (2.1), we obtain Eq. (2.3) where the integration of  $1/\tau(t, \eta)$  tracks the applied electric field.

$$\begin{aligned} A_{\eta+}(t) &= A_{\eta} - (A_{\eta} - A_{\eta+})|_{t=t_i} \exp \left[ - \left( \int_{t_i}^t \frac{1}{\tau(t', \eta)} dt' \right)^{\beta} \right] \text{ if } E_{FE}(t) \geq 0, \\ &= A_{\eta+}|_{t=t_i} \exp \left[ - \left( \int_{t_i}^t \frac{1}{\tau(t', \eta)} dt' \right)^{\beta} \right] \text{ if } E_{FE}(t) < 0, \end{aligned} \quad (2.3)$$

Eq. (2.3) is different from the NLS theory which assumes a constant electric field [10, 12]. An empirical parameter  $\beta$  is introduced to Eq. (2.3) to improve fitting with experimental data [22]. At the beginning of the simulation  $t=0$ , the initial condition for  $A_{\eta+}$  is specified by the user.  $t_1$  is the time instant when  $E_{FE}$  polarity changes for the first time. The second time when  $E_{FE}$  polarity changes is called  $t_2$ . Similarly,  $t_i$  is the  $i$ -th time when  $E_{FE}$  polarity changes.

Now, I sum up all the  $A_{\eta+}$ . I assume  $\eta$  is continuous. Thus,  $P(\eta)$  becomes  $f(\eta)d\eta$  where  $f(\eta)$  is the probability distribution function (PDF), and summation becomes an integration as Eq. (2.4).

$$A_+(t) = \sum A_{\eta+}(t) \cong \int_0^{\eta_{max}} A_T f(\eta) d\eta \times \frac{A_{\eta+}(t)}{A_{\eta}} \quad (2.4)$$

Previous experimental studies have identified two common PDFs, the Gaussian PDF [16, 17] and the generalized beta distribution of type 2 (GB2) [14, 22]. I carry out a numerical integration with 80 points to compute (4) which is more efficient as compared to [22] which uses thousands of grains to achieve the accuracy as shown in the next section. Finally, the polarization is calculated as Eq. (2.5).

$$P(t) = P_R \times \left( \frac{2A_+(t)}{A_T} - 1 \right) \quad (2.5)$$

where  $P_R$  is the remanent polarization and the switching current is obtained by the time differential of Eq. (2.5). The total current is the combination of

the switching current and the current from the background dielectric  $C_{FE}$  as Eq. (2.6).

$$I(t) = 2P_R \frac{dA_+(t)}{dt} + C_{FE} \frac{dV_{FE}}{dt} \quad (2.6)$$

## 2.3 Validation

I implement this model with Verilog-A code and validate it with different experimental data. In Fig. 2.1, we fit the experimental switching current data from [13] which is an 8.5 nm HZO capacitor. The model parameters are extracted for the given FE thickness and should be re-extracted when the thickness is changed. Here, I apply a Gaussian distribution with a mean at 1 and standard deviation,  $\sigma = 0.32$  for  $f(\eta)$ . As shown, our model can accurately capture the experimental data for different amplitudes of the applied voltage.

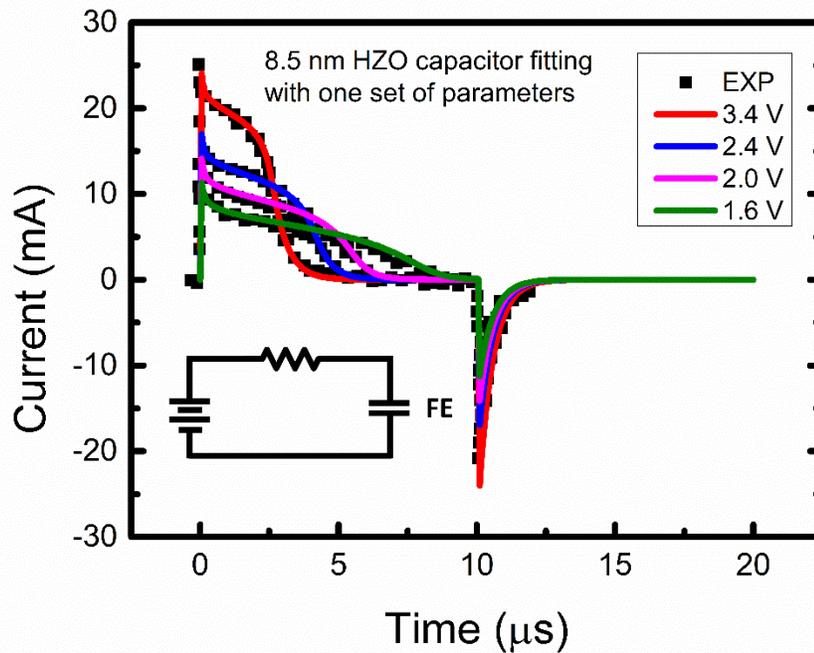


Fig. 2.1. Modeled current waveforms versus time for several voltages of a 8.5 nm HZO capacitor agree with experimental data from [16] with  $P_R = 20 \mu C/cm^2$ ,  $\tau_0 = 100 ps$ ,  $E_a = 8.0 MV/cm$  and a  $135 \Omega$  series resistance. These parameters are the same as used in [16]. © 2025 IEEE

I also validate our model with the experimental data from [18]. First, the

P-t data of an 8.3 nm HZO capacitor for different amplitudes of applied voltage pulses is fitted in Fig. 2.2 by using a GB2 PDF given by Eq. (2.7) with  $a=2.1$ ,  $b=0.99$ ,  $p=0.691$  and  $q=0.633$ .

$$GB2(\eta) = \frac{|a|b(b\eta)^{ap-1}}{B(p,q)[1+(b\eta)^a]^{p+q}} \quad (2.7)$$

where  $B(p, q)$  denotes the beta function. It shows that the model can capture the transient properties of the FE capacitor under different electric fields. Then, the same parameter set is used to test the model's capability to handle arbitrary input signals. GB2 PDF is a more general distribution and can fit the  $\eta$  distribution that the Gaussian PDF cannot handle. One should choose a suitable PDF depending on the extracted distribution from experimental data [10, 11].

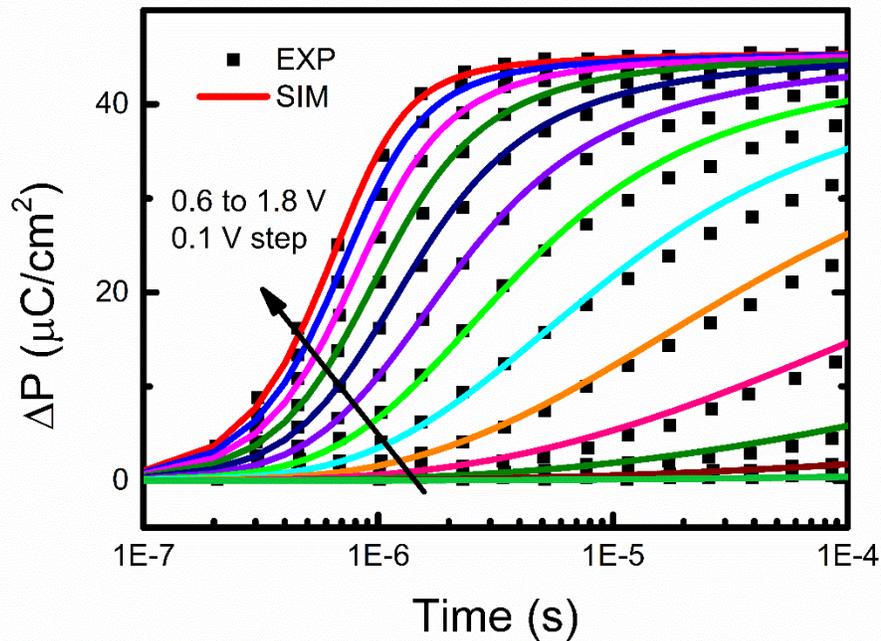


Fig. 2.2. Fitted P-t curves of a 8.3 nm HZO capacitor from Ref. [22] with  $P_R = 22.9 \mu\text{C}/\text{cm}^2$ ,  $\tau_0 = 390 \text{ ns}$ ,  $E_a = 1.74 \text{ MV}/\text{cm}$ ,  $\alpha = 3.48$ ,  $\beta = 2.0$  and a offset voltage of  $-0.08 \text{ V}$ . © 2025 IEEE

I apply the same triangular voltage waveform with varying amplitude as in [22] and shown in Fig. 2.3a. The simulated charge closely matches the

experimental data. By transforming Q-t into Q-V curves in Fig. 2.3b, I can see that our model can capture the transition between the major loop and minor loop switching and very well track the history of the FE device. The differences between the experiment and simulation could be because of a constant background permittivity assumption used in this model. Furthermore, we test the accumulative polarization property. Accumulation happens when a pulse train is applied to the FE. The polarization would gradually increase with the number of pulses. In Fig. 2.4, pulse trains with  $1 \mu\text{s}$  pulse width are applied to the FE capacitor. It shows that our simulation results can also fit the experimental data of polarization accumulation well.

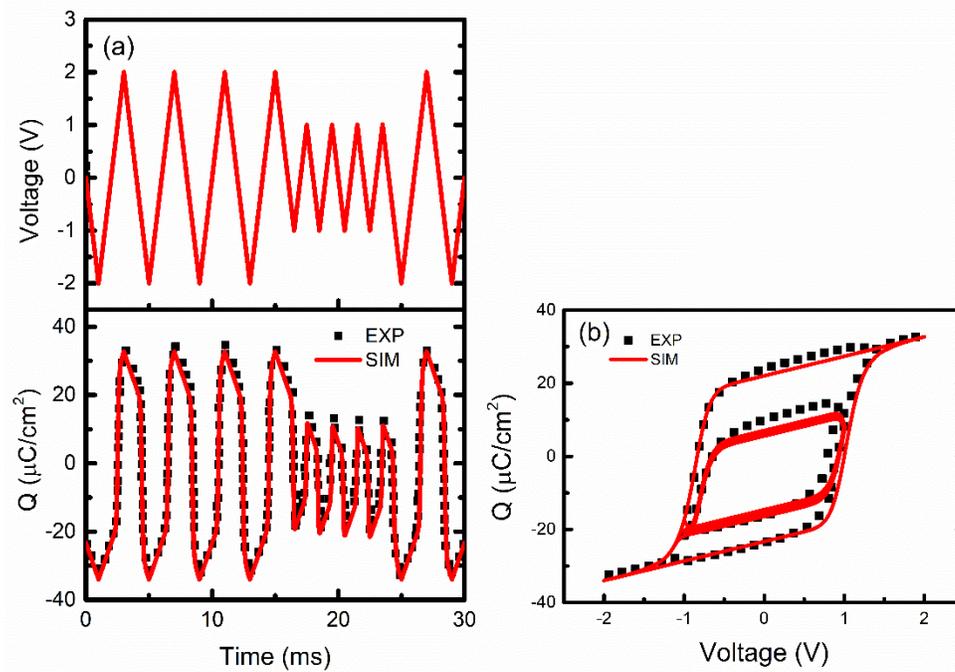


Fig. 2.3. (a) The simulation for the FE capacitor while applying a triangular waveform using the parameters extracted from Fig. 2.2. (b) Comparison between the experimental and simulated P-V curve from the voltage pulse shown in Fig. 2.3a. The difference between the two data comes from the assumption of constant permittivity. © 2025 IEEE

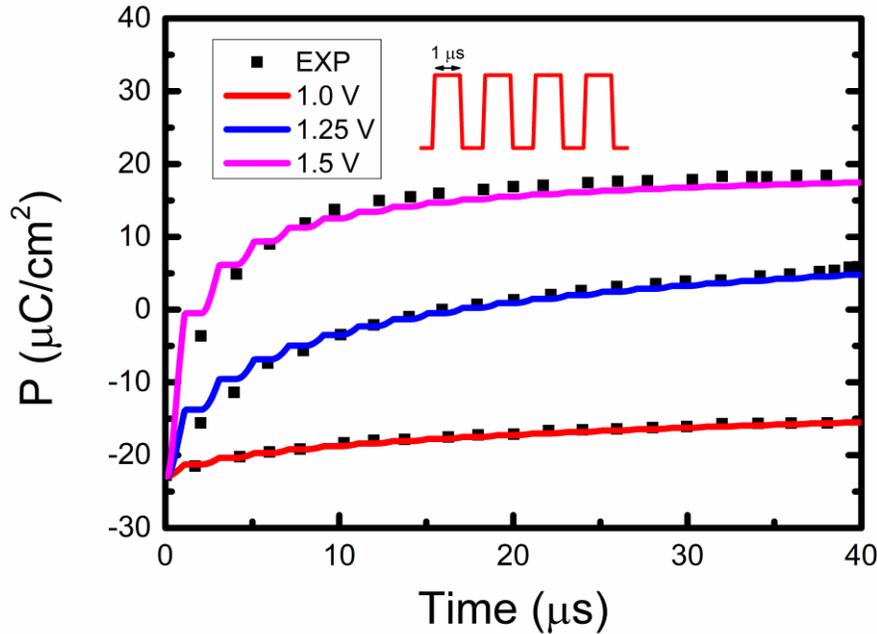


Fig. 2.4. The validation of the polarization accumulation by comparing the simulation results with the measurement in Ref. [22]. Pulse trains with  $1 \mu\text{s}$  pulse width at different amplitudes are applied to the capacitor. © 2025 IEEE

## 2.4 Scalability

I have verified this model with the large-area capacitor data. In a small-area device, the limited number of grains would make the groups discrete and cause discrete switching [21]. Normally, I use 80 groups in integration to approximate the continuous distribution of  $\eta$ . By reducing the number of  $\eta$  groups, I can model the discrete switching and device variation seen in scaled FE capacitors [17, 18, 22]. Fig. 2.5 shows discrete states of polarization in a small-area FE capacitor which contains only 10  $\eta$  groups. The switching has abrupt jumps and there are only 4 final states which is different from the continuous switching shown in Fig. 2.2. It is because the local activation fields are discrete. Small-applied voltage pulses can only switch a part of the FE no matter how long the time is. If the voltage is increased, above a certain value, it can switch to another part of the FE causing the abrupt switching state, and so on. In this case, some  $\eta$  groups might have very small  $\eta$  so basically every voltage can switch them. There are just 4  $\eta$  groups that can induce discrete

switching. I also test the variability in a small-area FE capacitor in Fig. 2.6. It shows that there is a large variation among several devices.

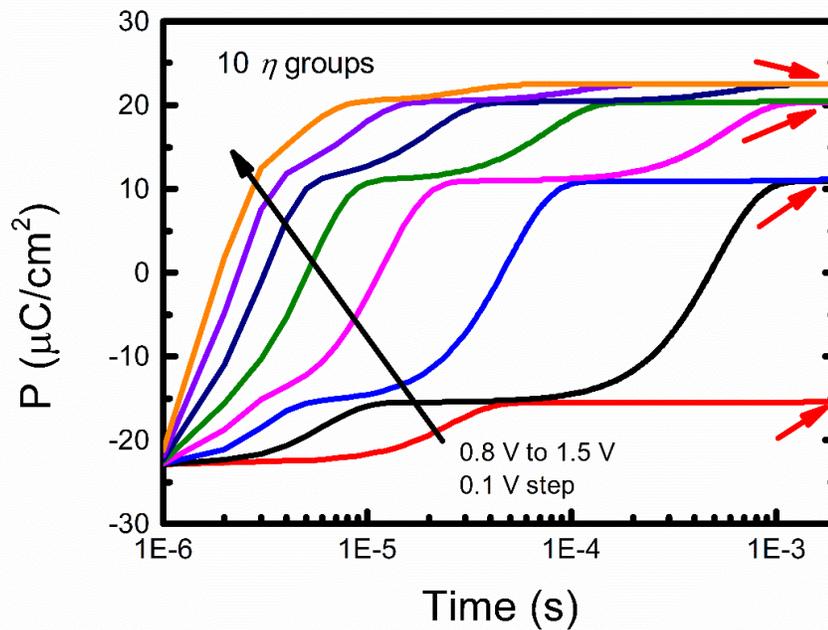


Fig. 2.5. Discrete switching behavior for a device with 10  $\eta$  groups. There are 4 states for this device different from the continuous states in large-area capacitors as Fig. 2.2. © 2025 IEEE

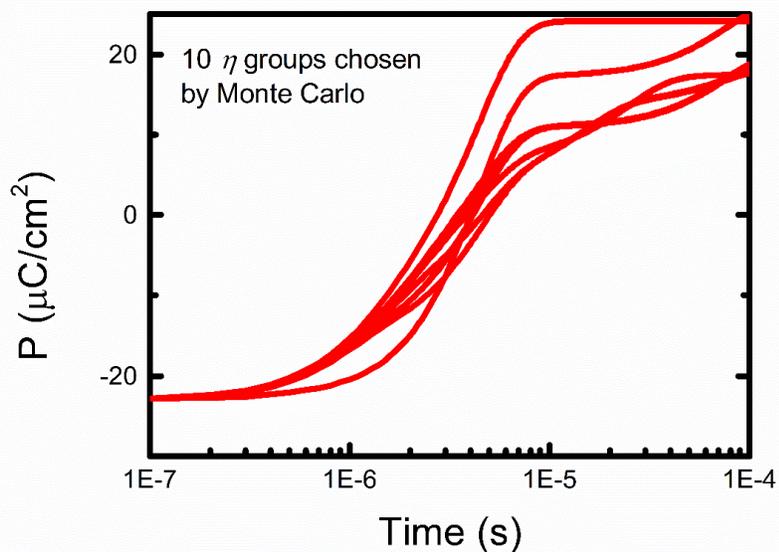


Fig. 2.6. The simulation of 10 devices with 10  $\eta$  groups which has a large device variability. © 2025 IEEE

## 2.5 Chapter Summary

I have developed a compact model of FE capacitors based on the experimental observation of polarization switching in polycrystalline materials. This model can capture the switching dynamics under arbitrary input waveforms like minor loop switching as well as polarization accumulation and is therefore suitable for modeling FE memory devices together with the write and read circuits using any circuit simulators. Although, for small-area FE capacitors, the assumptions of this model are not entirely correct, the model can still simulate discrete switching and device variations, which points out its potential for high-density FERAM simulations.

## Chapter 3 A Compact Model of Metal-Ferroelectric-Insulator-Semiconductor Tunnel Junction

I developed a compact model of metal-ferroelectric-insulator-semiconductor (MFIS) tunnel junctions. Unlike the metal-ferroelectric-metal structure with only one insulator layer, MFIS-FTJ contains two insulator layers and a semiconductor electrode. The complex structure makes it difficult to solve Poisson and charge equations self-consistently. I reported the first compact model of MFIS-FTJ to our knowledge. Previous modeling studies focused on numerical simulation, which is time-consuming and not applicable to circuit simulation. The presented compact model is suitable for commercial SPICE IC simulation. It includes a ferroelectric model that can capture polarization switching under arbitrarily applied voltage, an insulator-semiconductor model that calculates the potential profile of the MFIS stack, and an analytical tunneling current model. I demonstrate that this model can be used to simulate and fit both n-type and p-type MFIS-FTJs. This work is published by IEEE. © 2025 IEEE. Reprinted, with permission, from [24]

### 3.1 Motivation

Ferroelectric tunnel junction (FTJ) is a nonvolatile memory first proposed by Esaki *et al.* in 1971 [25]. FTJ, due to its good scalability, low-power consumption, and non-destructive reading [26], is a promising candidate for future nonvolatile memories. In 2011, discovering the ferroelectricity in HfO<sub>2</sub> [4, 5] made FTJ compatible with CMOS technology. The original FTJ is based on the metal-ferroelectric-metal (MFM) structure. The tunneling electroresistance (TER) ratio comes from the difference in screening lengths

of two electrodes causing the variation of the barrier heights in different polarization states [27, 28]. Due to the limited change in barrier heights, this structure suffers from a low TER ratio. To improve TER ratio, the metal-ferroelectric-insulator-semiconductor (MFIS) FTJ is developed. It has a much larger variation in the semiconductor surface potential than metal. A TER ratio over 30 has been reported in an n-type MFIS-FTJ with 4nm-HZO [29]. A p-type MFIS-FTJ with 1nm-HZO has also been fabricated with the on-current up to  $1 \text{ A/cm}^2$  [30]. With the growing interest in the FTJ device, a SPICE-compatible compact model of MFIS-FTJ is needed.

Although the compact model of MFM-FTJs has been studied [31, 32], there isn't a compact model of MFIS-FTJs. Unlike MFM-FTJ models that only need to consider the electron tunneling through one potential barrier, an MFIS-FTJ model needs to consider two potential barriers. Furthermore, due to the effect of the ferroelectric polarization on the band structure, the gate tunneling current model used for the HK-dielectric stack [33] cannot be applied to these devices. Previous modeling studies of MFIS-FTJs numerically calculate the energy band diagram and use the WKB approximation or the NEGF method to simulate tunneling current. This is not suitable in circuit simulations because of the significant computational time [34-37].

In this chapter, I present and demonstrate a compact model of MFIS-FTJs. Starting with the polycrystalline ferroelectric capacitor model we previously developed [38], the present model calculates ferroelectric polarization and the potential profile of the entire MFIS stack self-consistently. The potential profile is then used to calculate the tunneling current. This model has good computational efficiency and models the experimental data of MFIS-FTJ well.

## 3.2 Model

### Ferroelectric Model

An MFIS-FTJ can be seen as a ferroelectric (FE) capacitor in series with an insulator-semiconductor (IS) capacitor as shown in Fig. 3.1. To calculate the polarization of the FE capacitor, I start with the FE capacitor model

developed in the previous chapter [3]. Eq. (2.1) – (2.5) describe the model for ferroelectric polarization  $P(t)$ . The capacitor's total charge density is the sum of the polarization and the charge of background dielectric  $\epsilon_{FE}$  of the FE capacitor as Eq. (3.1). This is also the semiconductor charge density (per area),  $Q_s$ .

$$Q_s = -(P(t) + \epsilon_{FE}V_{FE}/t_{FE}) \quad (3.1)$$

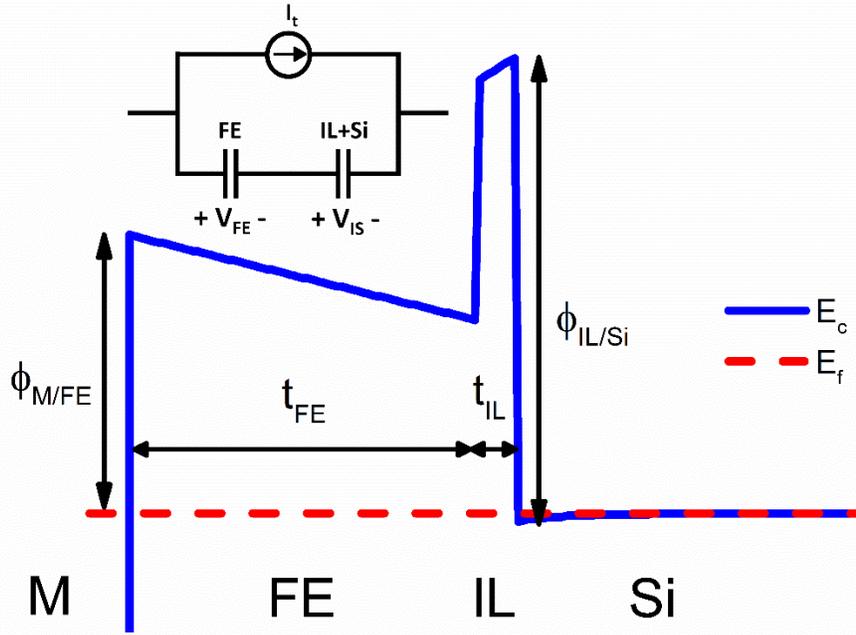


Fig. 3.1. Schematic band diagram of an MFIS-FTJ and its equivalent circuit. © 2025 IEEE

## Semiconductor Surface Potential Calculation

$Q_s$  can be used to calculate the semiconductor surface potential,  $\psi_s$  using Eq. (3.2) [39].  $N_d$  is the doping density and  $n_i$  is the intrinsic carrier density.

$$Q_s = \pm \sqrt{2\epsilon_{Si}k_B T N_d} \left[ \left( e^{\frac{q\psi_s}{k_B T}} - \frac{q\psi_s}{k_B T} - 1 \right) + \frac{n_i^2}{N_d^2} \left( e^{-\frac{q\psi_s}{k_B T}} + \frac{q\psi_s}{k_B T} - 1 \right) \right]^{0.5} \quad (3.2)$$

This equation has no explicit solution for  $\psi_s$  as a function of  $Q_s$  and must be solved iteratively. High computational speed is of prime importance for compact models for supporting the simulations of large memory circuits. Therefore, a good initial approximate solution for reducing the iteration cycles is of prime importance.

In the accumulation region, the initial approximation of  $\psi_{s0}$  is the following.

$$z = -\exp\left[-\left(\frac{Q_s}{\sqrt{2\varepsilon_{Si}k_B T N_d}}\right)^2 - 1\right] \quad (3.3a)$$

If  $z < 0$  then

$$\psi_{s0} = -\frac{k_B T}{q} \left[ W_{-1}(z) + \left(\frac{Q_s}{\sqrt{2\varepsilon_{Si}k_B T N_d}}\right)^2 + 1 \right] \times (1 + e^{\gamma Q_s}) \quad (3.3b)$$

$$W_{-1}(z) \approx \ln(-z) - \ln(-\ln(-z)) + \frac{\ln(-\ln(-z))}{\ln(-z)} \quad (3.3c)$$

If  $z \geq 0$  then

$$\psi_{s0} = \frac{k_B T}{q} \ln \left[ \left(\frac{Q_s}{\sqrt{2\varepsilon_{Si}k_B T N_d}}\right)^2 + 1 \right] \quad (3.3d)$$

In the depletion and inversion region, the initial approximation is the following.

$$z = \frac{n_i^2}{N_d^2} \exp \left[ \left(\frac{Q_s}{\sqrt{2\varepsilon_{Si}k_B T N_d}}\right)^2 + 1 \right] \quad (3.4a)$$

If  $z < 2$  then

$$\psi_{s0} = \frac{k_B T}{q} \left[ z - \left(\frac{Q_s}{\sqrt{2\varepsilon_{Si}k_B T N_d}}\right)^2 - 1 \right] \times (1 - e^{-\gamma Q_s}) \quad (3.4b)$$

If  $2 \leq z < 10^{200}$  then

$$\psi_{s0} = \frac{k_B T}{q} \left[ W(z) - \left(\frac{Q_s}{\sqrt{2\varepsilon_{Si}k_B T N_d}}\right)^2 - 1 \right] \quad (3.4c)$$

$$W(z) \approx \ln(z) - \ln(\ln(z)) + \frac{\ln(\ln(z))}{\ln(z)} \quad (3.4d)$$

If  $z \geq 10^{200}$  then

$$\psi_{s0} = -\frac{k_B T}{q} \ln \left[ \frac{N_d^2}{n_i^2} \left(\frac{Q_s}{\sqrt{2\varepsilon_{Si}k_B T N_d}}\right)^2 + \frac{N_d^2}{n_i^2} \right] \quad (3.4e)$$

$W(z)$  and  $W_{-1}(z)$  are the lambert W functions.  $\gamma$  in (3.3b) and (3.4b) are used to tune the approximation at small  $Q_s$  and improve the convergence. The condition  $z < 2$  is chosen to avoid the spike in (3.4c) if  $z$  is close to 1. The condition  $z > 10^{200}$  is introduced due to the numerical limitation of exponential in numerical computation. The value can also be adjusted depending on the simulator used. In Fig. 3.2, the initial approximation is already very close to the  $Q_s$  obtained by (3.1) using  $\psi_s$  as the input. Halley's iterative method is used to solve (3.5) to get the surface potential [40]. In most cases, the iteration converges in 2 or 3 cycles. From the semiconductor surface electrical potential and field, the voltage drops across the insulator can be computed. Finally, the terminal voltage  $V_{IS}$  of the insulator-semiconductor capacitor is the combination of the flat-band voltage  $V_{FB}$ , the insulator voltage and the semiconductor surface potential. For p-type, the equations are similar with different signs

$$V_{IS} = V_{FB} - \frac{Q_s}{C_{IL}} + \psi_s \quad (3.5)$$

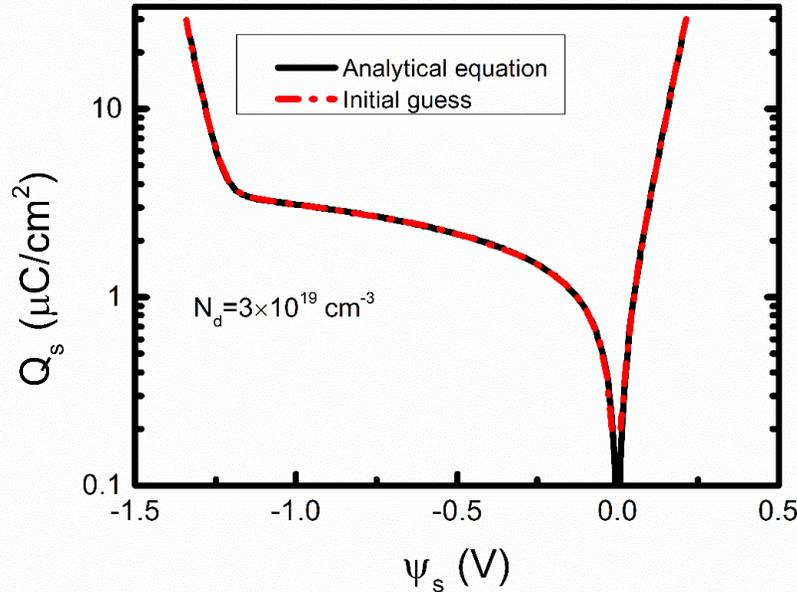


Fig. 3.2. Comparison of our initial guess of  $\psi_s$  and the  $Q_s - \psi_s$  curve obtained from the analytical equation Eq. (3.1). © 2025 IEEE

## Tunneling Current Model

To calculate the tunneling current in this compact model, I use an analytical equation Eq. (3.6). Previous studies of MFIS-FTJs often use numerical WKB approximation to simulate the tunneling current [35-37]. However, this method is not computationally efficient. Besides, WKB approximation does not consider the reflection of electron waves at the interfaces, and the band structure does not have the quantum mechanism correction. Therefore, even directly using the WKB approximation cannot produce accurate results, fitting parameters still need to be introduced. Furthermore, in addition to the conduction band electrons (CBE), valence band electrons (VBE) and holes (VBH) also need to be taken into account, which increases the computation time further [35]. In this work, we develop an analytical equation with several fitting parameters that can give the user enough flexibility to fit the experimental data of MFIS-FTJs with complex physics effects.

To simplify the model, I consider the semiconductor electrode as a metal and adapt the direct tunneling (DT) current equation from MFM-FTJ [41] to include the effect of two insulator layers. Eq. (3.6a) shows the barrier height relative to the Fermi level at the Insulator-Si and Metal-FE interfaces where  $\phi_{\text{IL/Si}}$  and  $\phi_{\text{M/FE}}$  are the band-offsets at these interfaces, and  $V_{a2}$  is the modified applied voltage to restrict the applied voltage  $V_a$  to the region of DT.  $a$  and  $b$  are both theoretically 0.5 for CBE but we make them fitting parameters to include the effect of VBE, VBH, and other quantum mechanism effects. An empirical function  $g(V_{a2})$  is also added to fit the tunneling current. This is a standard practice in compact modeling, which requires high-speed computation and agreement to less than 1% with the measured device data.

$$\phi_{\text{IL}} = \left[ \phi_{\text{IL/Si}} - \psi_s + k_B T \ln\left(\frac{N_c}{N_d}\right) \right] + q \cdot a V_{a2}, \phi_{\text{FE}} = \phi_{\text{M/FE}} - q \cdot b V_{a2} \quad (3.6a)$$

$$A_{\text{IL}} = \frac{-4t_{\text{IL}}\sqrt{2m_{\text{IL}}^*}}{3\hbar V_{\text{IL}}}, A_{\text{FE}} = \frac{-4t_{\text{FE}}\sqrt{2m_{\text{FE}}^*}}{3\hbar V_{\text{FE}}} \quad (3.6b)$$

$$T_1 = A_{\text{IL}} \left[ (\phi_{\text{IL}})^{\frac{3}{2}} - (\phi_{\text{IL}} - qV_{\text{IL}})^{\frac{3}{2}} \right] + A_{\text{FE}} \left[ (\phi_{\text{FE}} + qV_{\text{FE}})^{\frac{3}{2}} - (\phi_{\text{FE}})^{\frac{3}{2}} \right] \quad (3.6c)$$

$$T_2 = A_{\text{IL}} (\sqrt{\phi_{\text{IL}}} - \sqrt{\phi_{\text{IL}} - qV_{\text{IL}}}) + A_{\text{FE}} (\sqrt{\phi_{\text{FE}} + qV_{\text{FE}}} - \sqrt{\phi_{\text{FE}}}) \quad (3.6d)$$

$$J_{\text{DT}} = \frac{-4qm_{\text{FE}}^* \exp(T_1)}{3\pi^2\hbar^3 T_2^2} \sinh\left(\frac{3qT_2}{4} V_{a2}\right) \times g(V_{a2}) \quad (3.6e)$$

$$g(V_{a2}) = c \left[ dV_{a2}^2 + 1 + e - \sqrt{(dV_{a2}^2 + 1 - e)^2 + e(dV_{a2}^2 + 1)} \right] \times (fV_{a2}^4 + 1) \quad (3.6f)$$

I have verified this model with the simulated current using drift-diffusion and WKB approximation to calculate a bilayer MIS structure with a fixed polarization in the first layer for simplicity. Fig. 3.3 shows that this model can fit the WKB tunneling current through the bilayer structure with polarization accurately under different configurations. I can see that thinner FE, IL layers, and larger polarization would give higher currents, which agrees with [34]. For different  $N_d$ , there isn't a visible change in the positive on-current because the semiconductor is in inversion, which also agrees with [34]. The larger positive polarization will pull down the semiconductor potential more. However, for the high voltage region, the Fowler-Nordheim tunneling would appear as well as the space charge effect due to the transported carriers [42, 43]. Those effects are not considered by this compact model, which limits this model's accuracy at high-voltage regions.

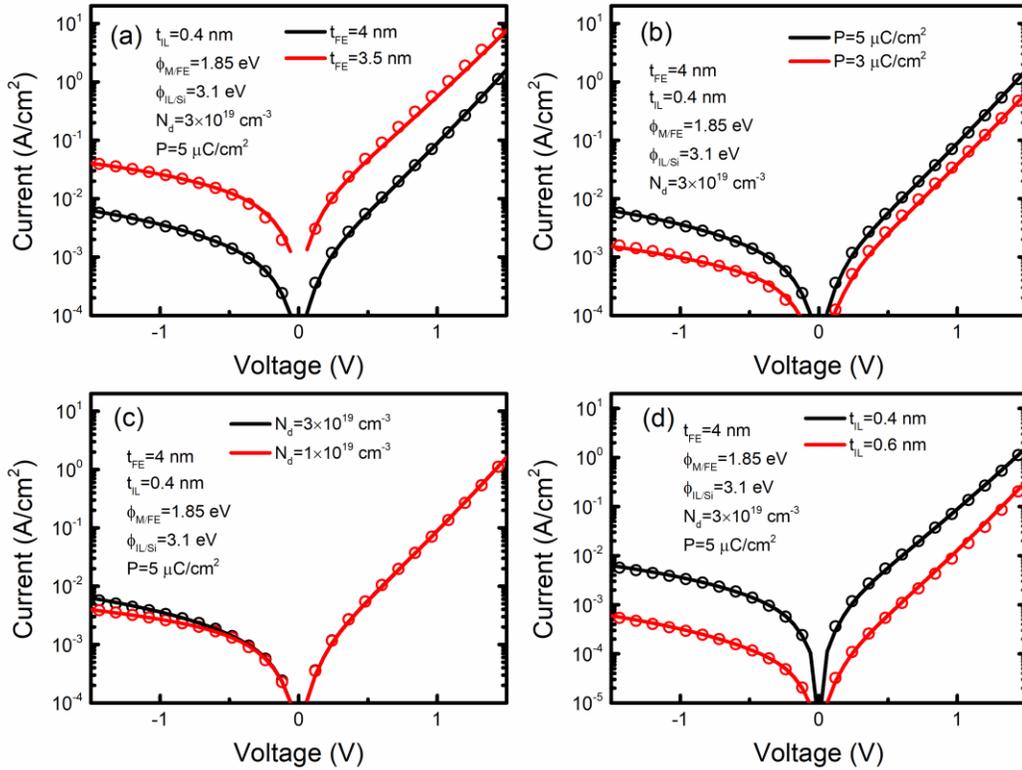


Fig. 3.3. Comparison of the tunneling current model and the WKB approximation. The symbols are the WKB approximation. The lines are the compact model. For the black line, the parameters are  $a = 0.52$ ,  $b = 0.48$ ,  $c = 1$ ,  $d = 30$ ,  $e = 1.8$ , and  $f = 0.008$ . For the red lines the following parameters are different: (a)  $d = 20$ ,  $e = 1.2$ ; (b)  $d = 20$ , and  $e = 2$ ; (c) (d): All parameters are the same. © 2025 IEEE

### 3.3 Discussion

I have implemented this compact model in Verilog-A code, the most popular programming language for compact models, and tested it on Hspice. I have modeled both n-type and p-type MFIS-FTJ experimental data using the presented MFIS-FTJ model. Fig. 3.4(a) shows the fitting of an n-type FTJ with  $t_{IL} = 0.4$  nm,  $t_{FE} = 4$  nm,  $N_d = 3 \times 10^{19}$  cm $^{-3}$  and  $P_R = 5\mu C/cm^2$  [29]; Fig. 3.4(b) shows the fitting of a p-type FTJ with  $t_{IL} = 1$  nm,  $t_{FE} = 1$  nm,  $N_a = 1 \times 10^{19}$  cm $^{-3}$  and  $P_R = 4\mu C/cm^2$  [30]. To fit these FTJ devices, the fitting parameters are treated as a function of polarization, where

I use interpolation to generate the parameters between  $+P_R$  and  $-P_R$ . I use subscript  $+$  for the value at  $+P_R$  and  $-$  for  $-P_R$  in Fig. 3.4. This compact model can be used for both n-type and p-type MFIS-FTJ regardless of whether electrons or holes are the transport carriers.

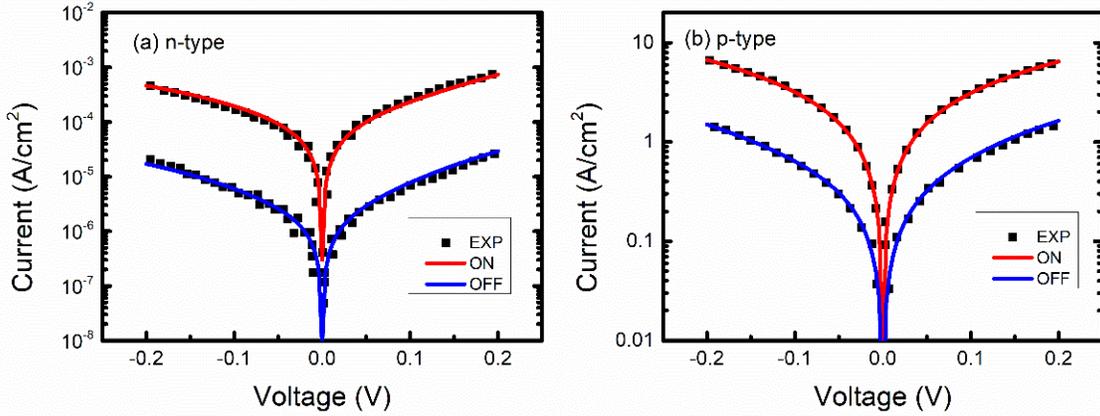


Fig. 3.4. Fitted results of (a) an n-type MFIS-FTJ [29] and (b) a p-type MFIS-FTJ [30]. Symbols are experimental data and lines are the simulation. For n-type,  $a_+ = 0.6$ ,  $a_- = 0.5$ ,  $b_+ = 0.52$ ,  $b_- = 0.3$ ,  $c_+ = 1.2$ ,  $c_- = 0.7$ ,  $d_+ = 11$ ,  $d_- = 40$ ,  $e_+ = e_- = 3$ , and  $f_+ = f_- = 0$ . For p-type,  $a_+ = a_- = 0.5$ ,  $b_+ = -0.2$ ,  $b_- = -0.4$ ,  $c_+ = 0.8$ ,  $c_- = 0.22$ ,  $d_+ = 2$ ,  $d_- = 15$ ,  $e_+ = e_- = 1$ , and  $f_+ = f_- = 0$ . © 2025 IEEE

I have tested this model for simulating MFIS-FTJ in circuits. Fig. 3.5 shows a simple circuit. To demonstrate the ability to work with arbitrary input waveforms, we apply an unusual 100ns triangular positive voltage pulse to program the FTJ to its on-state followed by a read voltage of 0.2 V, then by a negative triangular pulse to program to off state followed by a read voltage of 0.2V. Fig. 3.5 shows the device current vs time, the transient programming and read currents in the on and the off states. To further demonstrate the model's ability, in Fig. 3.6, I show the polarization switching and the tunneling current under triangular pulses of varying amplitude. The model keeps track of the polarization history well.

In addition, to test the simulation time efficiency of this compact model, I applied the waveforms of varying amplitude in Fig. 3.6a for 1000 cycles. The total runtime is only 52.27s on the Intel Xeon Gold 5115 CPU. This speed is comparable to popular transistor compact models, such as BSIM-CMG for

FinFET IC simulation, and is, therefore, suitable for the simulation of large circuits involving MFIS\_FTJ. The 52ms simulation presented in Fig. 3.6b will take hours if it is performed with TCAD.

As shown in Fig. 3.5, the polarization does not remain at its maximum value when the voltage rests at 0. It is because there is still some depolarization field across the FE layer as shown in Fig. 3.1. This leads to a decrease in TER ratio. The design rule of MFIS-FTJs should choose the FE material with a larger coercive field than the depolarization field. Here, reducing  $t_{IL}$  can help reduce the depolarization field and increase the tunneling current as shown in Fig. 3.3. The optimal design w.r.t the thickness, doping, and polarization has been studied in [34]. To reduce the depolarization field and increase the on-current as well as the TER ratio, we should reduce both  $t_{IL}$  and  $t_{FE}$  at the same time.

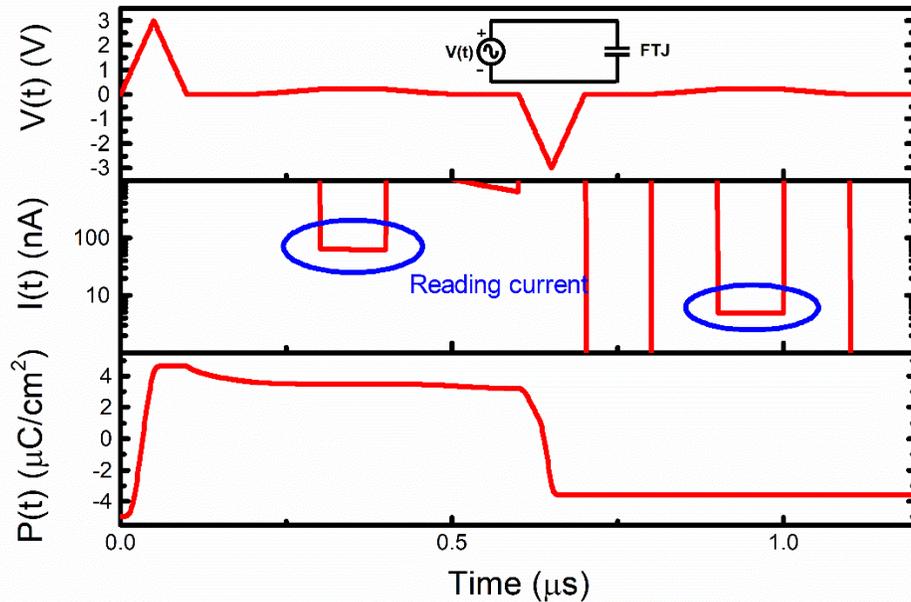


Fig. 3.5. The write and read test of the FTJ. The inset is the sample circuit. We show the corresponding reading current and polarization to this read/write operation. © 2025 IEEE

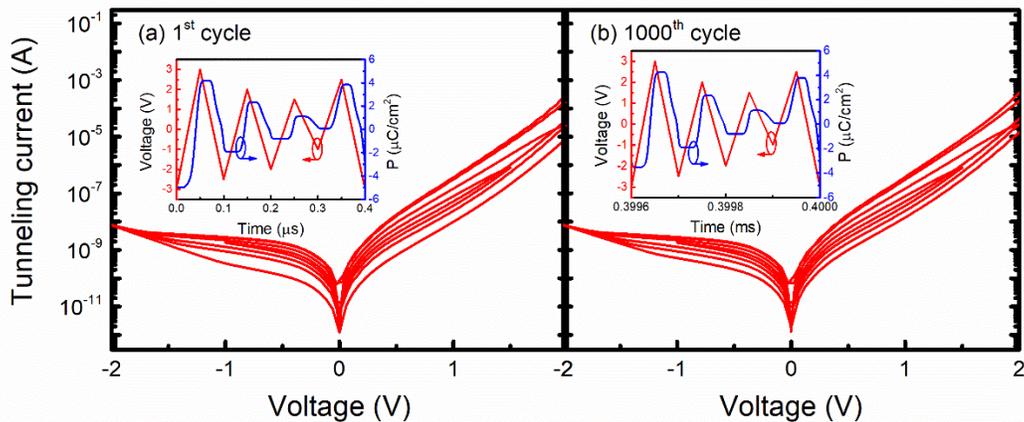


Fig. 3.6. Transient test of this model by applying triangular waveform with varying amplitude for 1000 cycles. The simulated polarization and tunneling current are shown in (a) the 1<sup>st</sup> cycle (b) the 1000<sup>th</sup> cycle. © 2025 IEEE

### 3.4 Chapter Summary

I have presented a compact model of MFIS-FTJ. It contains a time-dependent multi-grain ferroelectric model, a semiconductor surface potential model, and a tunneling current model. The model is computationally efficient for the simulation of integrated circuits because good analytical approximations are used and iterative calculation cycles are kept minimal. This model can fit both n-type and p-type FTJ experimental data well. Finally, we implemented this compact model in Verilog-A, ran with a commercial SPICE simulator, and demonstrated simulation speed comparable to MOS transistors and, therefore, its suitability for simulating large memory circuits.

## Chapter 4 A Compact Model of Ferroelectric Field-effect Transistor

In this chapter, I present a compact model of ferroelectric field-effect transistors (FEFET). The model consists of a ferroelectric (FE) capacitor model and a Berkeley Short-channel IGFET Model (BSIM), a standard SPICE MOSFET model. The FE model, similar to the nucleation-limited-switching model, is based on the statistical multidomain dynamics of FE materials. The charge equality between FE and MOSFET is satisfied through the SPICE simulator. The model reproduces the steep switching in the reverse bias region observed in experimental FEFETs and the current drop at both major loop and minor loop switching. A versatile inverted-memory-window (IMW) model can model the IMW behavior of FEFET that may be caused by charge trapping. I demonstrate that the reported model can accurately fit the published data of Fin-FEFET and FDSOI-FEFET under different bias conditions. This work is published by IEEE. © 2025 IEEE. Reprinted, with permission, from [44]

### 4.1 Motivation

Ferroelectric material has become one of the most promising candidates for future non-volatile memory (NVM) due to the discovery of ferroelectricity of thin-film  $\text{HfO}_2$  in 2011 [4]. The compatibility with the standard CMOS technology of hafnium zirconium oxide (HZO) makes it suitable for integration with our current integrated circuits (ICs). By using HZO as the gate insulator of a MOSFET, FEFET gains the most interest as an FE-based NVM. It has been demonstrated on GlobalFoundries' 22nm and 12nm Fully-depleted-silicon-on-insulator (FDSOI) processes with memory windows up to 1.5V [45]. The endurance of FEFET has been shown to exceed  $10^{10}$  cycles

[46]. The polycrystalline nature of HZO also makes FEFET possibly used for multi-bit operations [47].

For circuit design with this up-and-coming FEFET-NVM technology, the designers require a SPICE-compatible compact model of FEFET, accurately capturing its electrical characteristics. Several FEFET models have been developed previously. One of the popular FEFET models is the Preisach-based compact model, which couples the Preisach ferroelectric model with a BSIM model [19, 48]. The Preisach model is a computationally efficient model that can capture the hysteresis loop of FE. However, it is purely empirical and cannot describe the multidomain polycrystalline physics of HZO. It lacks scalability, variation, and stochasticity properties according to [23], which are important in FEFET simulations. Therefore, the nucleation-limited-switching (NLS) model is introduced, which can describe the statistical multidomain properties of HZO [12, 14, 22]. The FEFET models using the NLS framework are reported in several works [49, 50]. However, the NLS model cannot describe the steep-switching and inverted-memory-window.

In several recent experimental reports, a steep subthreshold slope ( $< 60\text{mV/dec}$ ) has been seen in FEFET [51-53]. There are many explanations for this effect: negative capacitance (NC), channel percolation or domain propagation, and so on [50, 54-56]. Several works include the percolation or domain propagation effects into their model and reproduce the steep switching characteristics; however, these works only showed the experimental trends and not a model fit to the data [50, 54, 55]. In addition to the steep switching, charge trapping is another critical issue in FEFET that may cause the IMW [52, 53]. [50, 57] include the charge trapping effect into their FEFET models using an analytical charge trapping model containing the integration over both energies and distances. Such an approach is computationally expensive and unsuitable for compact modeling for a large circuit design purpose.

In this chapter, I developed a computationally efficient compact model for FEFET using the NLS-based FE compact model I previously developed [38]. The steep-slope phenomenon as well as the IMW are captured using empirical equations that provide good fitting flexibility and computational efficiency. I demonstrate that this model can fit the experimental FEFET data

well.

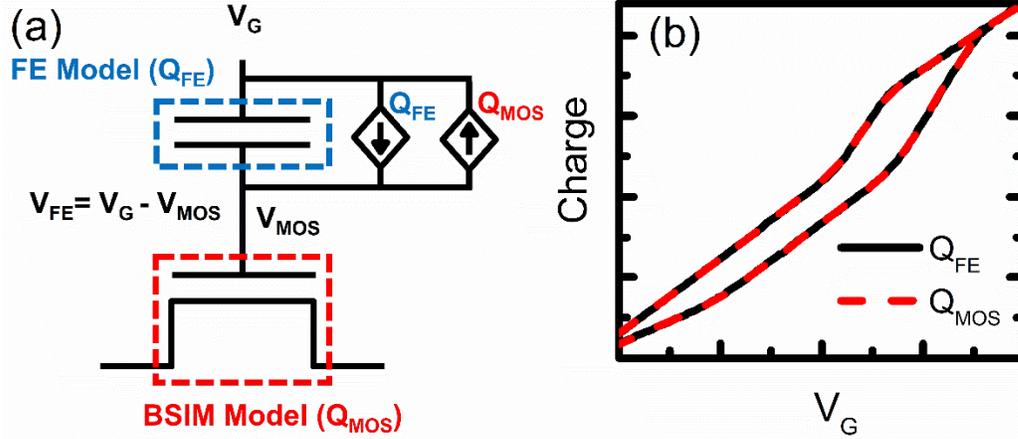


Fig. 4.1. (a) The equivalent circuit model of FEFET and (b) the charge equality between FE and MOSFET. © 2025 IEEE

## 4.2 Compact Model

### Equivalent circuit & FE model

Fig. 4.1a shows the equivalent circuit of my FEFET compact model. A FE capacitor is connected to the MOSFET gate, modeled by the BSIM model [58, 59]. The total charge on the FE capacitor ( $Q_{FE}$ ) has to equal the total charge on MOSFET ( $Q_{MOS}$ ). Since the electric field on FE varies with the channel position, directly making FE and MOSFET in series is not entirely physical. However, we choose to build the compact model in this way to achieve a good computational speed. Then, the  $Q_{MOS}$  is computed by the BSIM model, and  $Q_{FE}$  is calculated by the FE compact model in [38], rewritten here in Eq. (4.1). This FE model considers the statistical multidomain dynamics of polycrystalline FE where  $P_R$  is remanent polarization,  $E_{a+,-}$  are the activation fields at  $E_{FE} > 0$  and  $E_{FE} < 0$  respectively,  $E_{FE}$  is the electric field,  $\eta$  is the random variable describing the activation field distribution on different grains,  $P_\eta$  is the polarization for each  $\eta$  group,  $\tau_0$  is the characteristic time for a very large  $E_{FE}$ ,  $\alpha$ ,  $\beta$  and  $\gamma$  are fitting parameters,  $t_i$  is the time when  $E_{FE}$  polarity changes. I made  $E_{a+,-}$  to be  $V_{DS}$  dependent as Eq. (4.1c) to capture the variation in channel potential profile at different  $V_{DS}$  where  $E_{ah+,-}$  are the

activation fields at high  $V_{DS}$  and  $E_{a0+,-}$  are the activation fields at 0  $V_{DS}$ . The total polarization is the average of  $P_\eta$  using Eq. (4.1d), where  $\text{Prob}(\eta)$  is the probability of  $\eta$  and  $f(\eta)$  is the probability density function. The total charge is  $P_{FE}$  plus the background dielectric charge ( $\varepsilon_{FE}V_{FE}/t_{FE}$ ) times the area ( $A_{FE}$ ) where  $\varepsilon_{FE}$  is the background dielectric constant and  $t_{FE}$  is the thickness of FE.

$$\begin{aligned}
 P_\eta(t) &= P_R - (P_R - P_\eta(t_i)) \exp\left(-\left(\int_{t_i}^t \frac{1}{\tau(t')} dt'\right)^\beta\right), \\
 &\quad \text{if } E_{FE}(t) \geq 0, \\
 &= -P_R + (P_R + P_\eta(t_i)) \exp\left(-\left(\int_{t_i}^t \frac{1}{\tau(t')} dt'\right)^\beta\right), \\
 &\quad \text{if } E_{FE}(t) < 0, \quad (4.1a)
 \end{aligned}$$

$$\tau(t) = \tau_0 \exp\left(\left(\frac{\eta E_{a+,-}}{|E_{FE}(t)|}\right)^\alpha\right) \quad (4.1b)$$

$$E_{a+,-} = (E_{ah+,-} - E_{a0+,-})V_{DS}^\gamma + E_{a0+,-} \quad (4.1c)$$

$$P_{FE}(t) = \sum P_\eta(t) \text{Prob}(\eta) \cong \int_0^{\eta^{max}} P_\eta(t) f(\eta) d\eta \quad (4.1d)$$

$$Q_{FE} = A_{FE}(P_{FE} + \varepsilon_{FE}V_{FE}/t_{FE}) \quad (4.1e)$$

Since both the FE and BSIM models can only give charge as a function of voltage and not the other way around, we need to ensure the charge equality numerically. I utilize the solver in the SPICE simulator by using two current sources whose value equals  $Q_{FE}$  and  $Q_{MOS}$ , respectively. As shown in Fig. 4.1a, these two current sources form a current loop, and SPICE will automatically make sure they cancel each other. Fig. 4.1b shows a simulated  $Q_{FE}$  and  $Q_{MOS}$  using this method, from which we can see that these two charges are equal.

## Steep Switching (SS) Model

The experimental  $I_D$ - $V_G$  FEFET data from several research groups show a sub-60mV/dec slope in the reverse sweep [51-53]. This has been attributed to a transient NC (TNC) effect in [56], but it also appears under DC conditions [51], implying that it could be due to intrinsic NC. The percolation theory has also been proposed to explain the steep switching in [50, 54], where they used

an effective threshold voltage shift to produce the steep slope. Another theory is the low electric field domain propagation which says the switching rate is dominated by domain propagation at the low electric field which gives a higher switching rate [50, 55]. They modified the NLS model to show the sub-60 mV/dec slope. Due to the multiple reasons for the steep switching, we propose a phenomenology model.

I mainly adopt the idea of percolation theory, but it does not have a simple physical model and requires Monte Carlo simulation which is too computationally expansive. Thus, our SS model uses the concept of threshold voltage shift to model the sudden current drop [50, 54]. As shown in Eq. (4.3a), the internal gate voltage used in calculations ( $V_{\text{MOS}}'$ ) is the node voltage ( $V_{\text{MOS}}$ ) pulse a shift ( $\Delta V$ ).  $\Delta V$  is a function composed of  $\Delta V_1$  and  $\Delta V_2$  as Eq. (4.3b) where they are the voltage shift functions for major loop forward and reverse sweepings, respectively. The variables,  $a$ ,  $b$ ,  $c_{1,2}$ ,  $d_{1,2}$ ,  $e_{1,2}$ , and  $f_{1,2}$ , in Eq. (4.3) are empirical parameters. From percolation theory, the threshold voltage shift happens when  $P_{\text{FE}}$  is smaller than a certain value [50, 54]. Here,  $e_{1,2}$  determine where the voltage shift will happen,  $a$  and  $b$  control the smoothing between  $\Delta V_1$  and  $\Delta V_2$ , and  $c_{1,2}$ ,  $d_{1,2}$ , and  $f_{1,2}$  control the amplitude and the slope of this voltage shift. However, we have also observed a different current drop in the minor loop of experimental data (such as shown in Fig. 4.3) which is much smoother and cannot be covered by Eq. (4.3). Thus, we introduce a current multiplication factor Eq. (4.4) with  $m_1$  to  $m_5$  as the fitting parameters, which can be used to model this current decay.  $m_2$  and  $m_5$  determine the place where this current decay appears on PV plane, and the other parameters control the amplitude and smoothness of this function.

$$V_{\text{MOS}}' = V_{\text{MOS}} + \Delta V \quad (4.3a)$$

$$\Delta V = 0.5\Delta V_2 \left( \frac{\Delta V_1}{\Delta V_2} - 1 \right) \tanh[a(V_{\text{FE}} - b)] + 0.5\Delta V_2 \left( \frac{\Delta V_1}{\Delta V_2} + 1 \right) \quad (4.3b)$$

$$\Delta V_{1,2} = 0.5c_{1,2} \{ \tanh[d_{1,2}(P_{\text{FE}} - e_{1,2}P_R)] - 1 \} + f_{1,2} \quad (4.3c)$$

$$I_{\text{DS}} = I_{\text{DS}}' \times M \quad (4.4a)$$

$$M = 1 - 0.25 \{ \tanh[m_1(P_{\text{FE}} - m_2P_R)] - 1 \} \\ \times \{ (1 - m_3) \tanh[m_4(V_{\text{FE}} - m_5)] + (m_3 - 1) \} \quad (4.4b)$$

## Inverted-memory-window (IMW) Model

The IMW is caused by charge trapping (CT) [52, 53]. If we want to model CT, we need to consider the quantum mechanical tunneling and the integration over both energy and distance [50, 55, 57]. However, this is too time-consuming for a compact model. Therefore, I proposed a phenomenological IMW model that is coupled with the FE model and does not require additional computational power for the model convergence. Based on observation, the IMW happens in reverse sweeping, and the FE charge is not large enough. Therefore, the model is given by Eq. (4.5) and is activated during the reverse sweep when  $P_{FE}$  is smaller than a threshold polarization  $P_t = \sigma P_R$  where  $t_i$  is the time when reverse sweeping happens. The time constant  $\tau_{nw}$  is now determined by the corresponding electric field  $E_t$  at  $t_i$  and a new activation field  $E_{a,IMW}$  to tune the switching rate. Fig. 4.3 shows that my model can reproduce the IMW effect in  $I_D$ - $V_G$  curves and  $Q_{FE}$ - $V_G$  loop. The biggest advantage of this model is that we don't need to worry about the interaction between FE and CT which will increase the difficulty of convergence for solving the charge equality and the fitting complexity that people need to iteratively tune both models to get the good fitting.

$$P_\eta(t) = -P_R + (P_R + P_\eta(t_i)) \exp\left(-\left(\int_{t_i}^t \frac{1}{\tau_{nw}(t')} dt'\right)^\beta\right) \quad (4.5a)$$

$$\tau_{nw}(t) = \tau_0 \exp\left(\left(\frac{E_{a,IMW}}{|E_{FE}(t) - E_t|}\right)^\alpha\right) \quad (4.5b)$$

### 4.3 Results

I validated this model by fitting the published FEFET data. I modeled both Fin-FEFET and FDSOI-FEFET by coupling this FE model with BSIM-CMG [58] and BSIM-IMG [59], respectively. Fig. 4.2 shows the fitting results of an FDSOI-FEFET [52] and a Fin-FEFET [60]. The model can be used for both SS and non-SS cases at different  $V_{DS}$ . The difference in the MW at different  $V_{DS}$  is captured by Eq. (4.1c). Fig. 4.3a shows the validation with the different minor loops of the FEFET for different maximum gate voltages ( $V_{MAX}$ ) [52]. As we can see the SS is reproduced by adding a voltage shift when  $P_{FE}$  is smaller than a certain value. The charge shows a steep switching

as well as the current at the corresponding  $P_{FE}$ . For  $V_{MAX}=1.2V$ , the smoother current drop different from the SS at  $V_G=1V$  is captured by Eq. (4.4) which activates in minor loops. For IMW at  $V_{MAX}=1.0V$ , from Fig. 4.3b, we can see that the charge decreases when reverse sweeping happens and  $P_{FE}$  is smaller than  $P_t$  which produces the IMW effect.

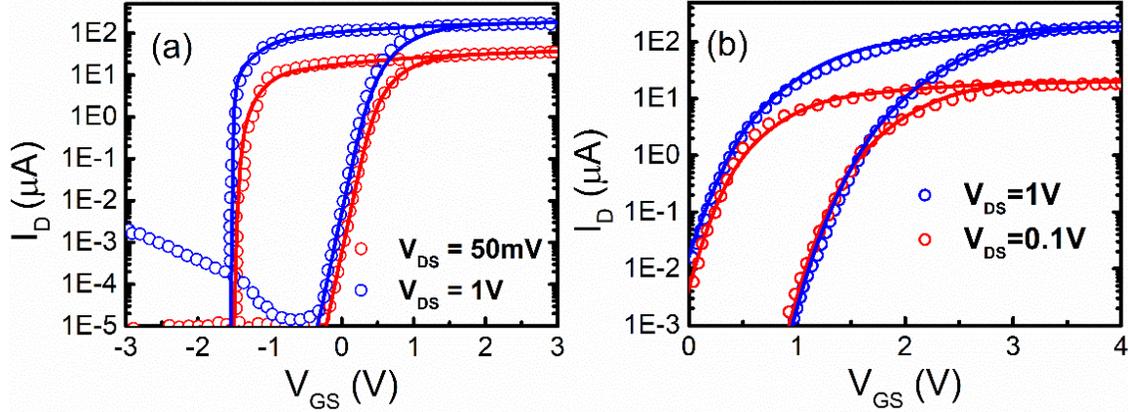


Fig. 4.2. Fitted  $I_D$ - $V_G$  curves compared to the experimental data of (a) FDSOI-FEFET with  $W=170nm$ ,  $L=24nm$ ,  $t_{FE}=10nm$ ,  $T=300K$ ,  $E_{a0+}=2.7MV/cm$ ,  $E_{a0-}=3.7MV/cm$ ,  $E_{ah+}=3.5MV/cm$ ,  $E_{ah-}=3.8MV/cm$ ,  $a=2$ ,  $b=-0.1$ ,  $c_{1,2}=1.2$ ,  $d_1=20$ ,  $d_2=500$ ,  $e_1=-0.5$ , and  $e_2=0.4$  [52] (b) Fin-FEFET with  $H_{FIN}=30nm$ ,  $W_{FIN}=50nm$ ,  $L=70nm$ ,  $t_{FE}=10nm$ ,  $E_{a0+,-}=4MV/cm$ ,  $E_{ah+,-}=4.6MV/cm$  [60] where symbols are the measurement and lines are the simulation. © 2025 IEEE

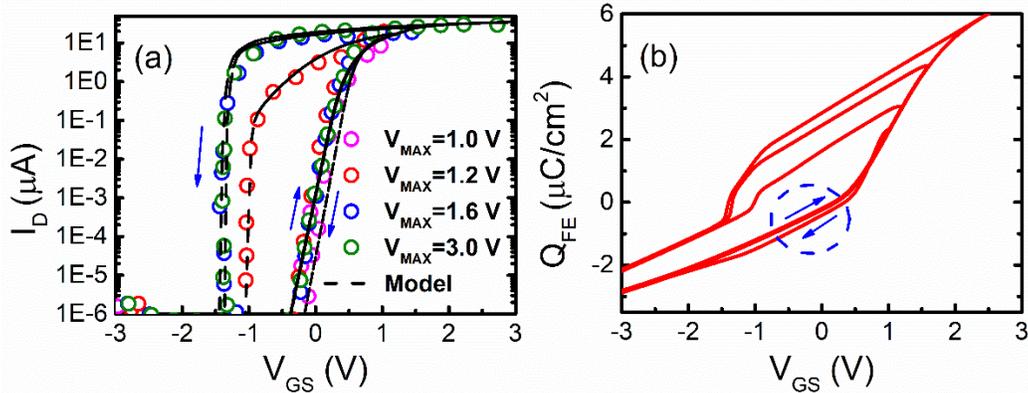


Fig. 4.3. (a) Fitted  $I_D$ - $V_G$  curves for different minor loops of the FDSOI-FEFET [52] where symbols are the measurement and lines are the simulation with  $W=170nm$ ,  $L=24nm$ ,  $t_{FE}=10nm$ ,  $T=300K$ ,  $E_{a+}=3MV/cm$ ,  $E_{a-}=3.2MV/cm$ ,  $a=2$ ,  $b=-0.2$ ,  $c_{1,2}=1.2$ ,  $d_1=20$ ,  $d_2=500$ ,  $e_1=-0.3$ ,  $e_2=0.45$ ,  $m_1=600$ ,  $m_2=0.6$ ,  $m_3=1e-4$ ,  $m_4=1.5$ ,  $m_5=0.33$ , and  $\sigma=0.33$ . The maximum

gate sweep voltages are 1.0V, 1.2V, 1.6V and 3.0V. (b) The QV loops correspond to the IV curves. © 2025 IEEE

## 4.4 Chapter Summary

I have presented a compact model of FEFET. The FE switching is modeled by a statistical multidomain switching model, which is coupled with the BSIM SPICE model with a subcircuit to achieve charge equality. To model the SS and IMW phenomena, we apply empirical models that are flexible enough to accurately fit the IV characteristics from experimental devices. We show that this model can fit Fin-FEFET and FDSOI-FEFET well at different bias conditions.

## Chapter 5 A Compact Model of Antiferroelectric Capacitor

In this chapter, I developed a compact model of antiferroelectric (AFE) capacitors. AFE material, similar to the ferroelectric (FE) material, is a good candidate for non-volatile memory applications. Unlike FE materials, no good compact models can describe the AFE materials for circuit simulation. I consider the AFE material in this study as a collection of multiple crystal groups. For each group, the polarization may switch from zero to positive or negative polarization and vice versa depending on the electric field polarity. This switching is modeled by a local field-dependent switching rate, which has a statistical distribution among the groups. I implemented this model in Verilog-A and ran it on a commercial SPICE simulator to demonstrate this model's capability to reproduce the published experimental data of the dependency of the AFE capacitor switching on the writing pulse width and voltage and the behavior of the major and minor loop are demonstrated. This work is published by IEEE. © 2025 IEEE. Reprinted, with permission, from [61].

### 5.1 Motivation

Antiferroelectric (AFE) material is a promising material for memory applications in the integrated circuit (IC) industry. Compared to FE material, experiments have shown that zirconium oxide ( $\text{ZrO}_2$ ) based AFE has higher speed and endurance compared to FE [62, 63]. Furthermore, the double hysteresis loops of AFE may potentially store two bits of data per device, doubling the storage density [64]. AFE can be applied to nonvolatile memories (NVMs) such as ferroelectric RAM (FERAM), ferroelectric tunnel junction (FTJ), and ferroelectric FET (FEFET) with better reliability [65]. AFE also shows great potential for energy storage supercapacitor applications

[66]. AFE-DRAM has been demonstrated to have a 10ns switching and 1ms retention, and high endurance with  $10^{12}$  cycles [67]. Moreover, AFE can be stacked with FE to engineer the on-current and subthreshold swing of negative capacitance FET (NCFET) [68, 69].

To model AFE, the best-known model is the Kittel’s model [70]. However, similar to the Landau–Khalatnikov (L–K) equation [20, 71, 72], simulating the multi-domain dynamics using phase-field modeling is orders of magnitude too time-consuming for IC simulation. Moreover, in HZO thin film, the polycrystalline property makes the device contain multiple grains with different material properties due to the crystal orientation, stoichiometry and stress [63, 73]. Thus, not only the multi-domain but also multi-grain physics needs to be considered. Another common model is the Preisach model [48, 74]. It uses empirical interpolation to produce the hysteresis loop, which is not physical. Therefore, I aimed to develop a computationally efficient model that considers the multi-domain/multi-grain switching physics in AFE.

Multi-grain switching has been studied in FE for many years, and one model is the nucleation-limited-switching (NLS) model [14, 16, 22]. It considers an FE capacitor made of many independent grains and each grain has a unique exponentially time-dependent switching rate. The NLS model has also been used on AFE [63, 75] but these models can only characterize (positive or negative) half of the major hysteresis loop and not any minor loops. In this paper, based on our previous model of FE capacitors [38], I constructed an AFE model that reimagines NLS switching dynamics for AFE. I showed that this model can model both the major loop and minor loop as well as the pulse-width-dependent switching characteristics well by reproducing the published experimental data [62, 74].

## 5.2 Model

In AFE, we can observe from Fig. 5.1 that the polarization has three different states:  $P_R$ ,  $-P_R$  and 0. Polarization would switch from  $-P_R$  to 0 and from 0 to  $P_R$  while voltage increases and vice versa. For each hysteresis loop, we can refer to FE domain switching where the polarization switches from nuclei and grows to the entire device [20, 72]. For example, for the left

hysteresis curve in Fig. 5.1a, the negative polarization would switch into zero polarization from some nuclei and switch the whole area when voltage increases, and the capacitor is switched from negatively polarized areas into non-polarized areas. Therefore, I approximate the total polarization into the combination of positively polarized areas  $A_+$ , negatively polarized areas  $A_-$  and non-polarized areas  $A_0$ . I first derive the switching current in region (I) in Fig. 5.1 as Eq. (5.1a) where  $E_{\text{AFE}}$  is the electric field,  $E_{T+}$  and  $E_{T-}$  are the threshold electric field, and  $P_R$  is the remanent polarization. However, directly solving three kinds of areas is numerically complex and not acceptable for a compact model. Instead, we introduce two new areas by dividing the capacitor into low-field-switching area  $A_l$  and high-field-switching area  $A_h$  where  $A_l$  starts switching under the low voltage in (II) & (III), and  $A_h$  starts switching under the higher voltage in region (I) & (IV), and  $A_l = A_h = 0.5A_T$  where  $A_T$  is the total area.  $A_l$  and  $A_h$  only consist of positively and negatively polarized areas,  $A_{l+}$ ,  $A_{l-}$ ,  $A_{h+}$ , and  $A_{h-}$ .  $A_0$  can be represented by the polarization cancellation between  $A_h$  and  $A_l$ . Fig. 5.1 shows how  $A_{l+}$  and  $A_{h+}$  varying with  $E_{\text{AFE}}$  in which we can see how polarization changes corresponding to  $A_h$  and  $A_l$ . Then, we obtain the switching current in (I) as Eq. (5.1a). Similarly, the switching current in region (IV) is shown in Eq. (5.1b). For regions (II) & (III), the derivation is simpler. Depending on the polarization  $P$ , we either calculate the decrease in  $A_{l+}$  or  $A_{l-}$  as Eq. (5.1c).

$$\begin{aligned} & I(A_0 \rightarrow A_-) + I(A_+ \rightarrow A_0) \\ \cong & I(A_{l+} \rightarrow A_{l-}) + I(A_{h+} \rightarrow A_{h-}) = 2P_R \left[ \frac{dA_{l+}}{dt} + \frac{dA_{h+}}{dt} \right] \end{aligned} \quad (5.1a)$$

$$I(A_0 \rightarrow A_+) + I(A_- \rightarrow A_0) \cong -2P_R \left[ \frac{dA_{l-}}{dt} + \frac{dA_{h-}}{dt} \right] \quad (5.1b)$$

$$\begin{aligned} I(A_+ \rightarrow A_0) &= 2P_R \frac{dA_{l+}}{dt}, \text{ if } P > 0, \\ I(A_- \rightarrow A_0) &= -2P_R \frac{dA_{l-}}{dt}, \text{ if } P < 0 \end{aligned} \quad (5.1c)$$

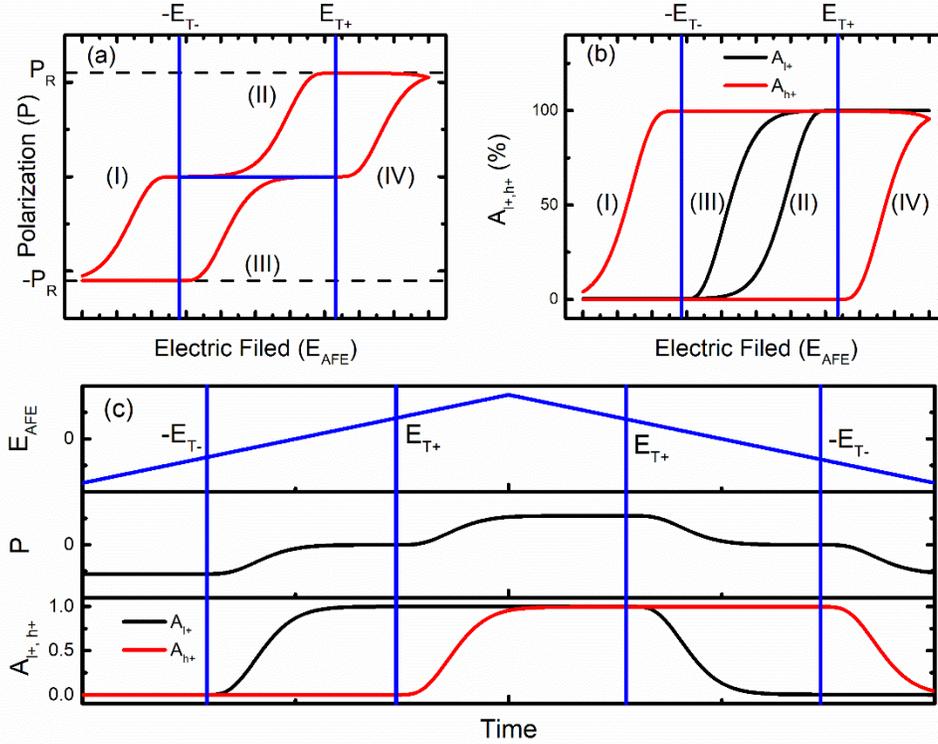


Fig. 5.1. (a) The schematic hysteresis loop of AFE which is divided into 4 regions: (I)  $E_{AFE} < -E_{T-}$ , (II)  $-E_{T-} \leq E_{AFE} \leq E_{T+}$  &  $P > 0$ , (III)  $-E_{T-} \leq E_{AFE} \leq E_{T+}$  &  $P < 0$ , and (IV)  $E_{AFE} > E_{T+}$ ; (b) The percentage of  $A_{l+}$  and  $A_{h+}$  over  $A_l$  and  $A_h$  under different  $E_{AFE}$ ; (c) The schematic transient diagram of AFE. © 2025 IEEE

After deriving the current expressions, we now set up the differential equations for each grain group.  $A_\eta$  is the total area of which the switching rate is characterized by the random variable  $\eta$  shown in Eq. (5.3).  $\eta$  has a mean of 1 and a statistical distribution  $f(\eta)$  to be determined from measurement data [38]. Here, I assumed a Gaussian distribution. Since  $A_{\eta+} = A_\eta - A_{\eta-}$ , we can write all differential equations in terms of  $A_{\eta+}$  as Eq. (5.2). The switching time constant is defined in Eq. (5.3) known as the Merz's law [76] where  $\tau_0$  is the characteristic time at a very large field,  $E_\alpha$  is the mean activation field,  $\alpha$  is a fitting parameter, and  $E_{\text{eff},l}$  is the effective electric field which equals  $E_{AFE} + E_{T-}$  at (III) & (IV) and  $E_{AFE} - E_{T+}$  at (I) & (II), and  $E_{\text{eff},h}$  equals to  $E_{AFE} - E_{T+}$  at (IV) and  $E_{AFE} + E_{T-}$  at (I) where  $E_{T+}$  and  $E_{T-}$  are fitting parameters that characterize the difference between the low-field-switching

and high-field-switching areas. We keep tracking  $A_{\eta l+}$  and  $A_{\eta h+}$  with time using the differential equations (5.2). Finally, the overall positively polarized area  $A_+$  is the summation of all  $A_{\eta+}$  where the integration is carried out numerically using Eq. (5.4). The polarization and current are calculated by Eq. (5.5) and Eq. (5.6) respectively where  $C_{\text{AFE}}$  is the background capacitance of the AFE capacitor.

$$\begin{aligned} \frac{dA_{\eta+}}{dt} &= \frac{d(A_{\eta l+} + A_{\eta h+})}{dt} = \frac{-A_{\eta l+}}{\tau_l} + \frac{-A_{\eta h+}}{\tau_h}, \text{ if } E_{\text{AFE}} < -E_{T-} \\ &= \frac{-A_{\eta l+}}{\tau_l}, \text{ if } -E_{T-} \leq E_{\text{AFE}} \leq E_{T+} \text{ \& } P > 0 \\ &= \frac{A_{\eta l-} - A_{\eta l+}}{\tau_l}, \text{ if } -E_{T-} \leq E_{\text{AFE}} \leq E_{T+} \text{ \& } P < 0 \\ &= \frac{A_{\eta l-} - A_{\eta l+}}{\tau_l} + \frac{A_{\eta h-} - A_{\eta h+}}{\tau_h}, \text{ if } E_{\text{AFE}} > E_{T+} \end{aligned} \quad (5.2)$$

$$\tau_l = \tau_0 \exp \left[ \left( \frac{\eta E_a}{|E_{\text{eff},l}|} \right)^\alpha \right], \tau_h = \tau_0 \exp \left[ \left( \frac{\eta E_a}{|E_{\text{eff},h}|} \right)^\alpha \right] \quad (5.3)$$

$$A_+(t) = \sum A_{\eta+}(t) \cong \int_0^{\eta_{\text{max}}} A_T f(\eta) d\eta \times \frac{A_{\eta+}(t)}{A_\eta} \quad (5.4)$$

$$P(t) = P_R \times \left( \frac{2A_+(t)}{A_T} - 1 \right) \quad (5.5)$$

$$I(t) = 2P_R \frac{dA_+(t)}{dt} + C_{\text{AFE}} \frac{dV_{\text{AFE}}}{dt} \quad (5.6)$$

### 5.3 Results

I implemented this model with Verilog-A and ran simulations on Hspice. To verify this compact model's accuracy, I fitted the data of a 10 nm  $\text{ZrO}_2$  capacitor subject to a triangle voltage waveform at 1 kHz with varying amplitude [74]. In parameter extraction, I first extracted  $P_R$ ,  $E_{T+}$  and  $E_{T-}$  which can be easily found as Fig. 5.1. For other parameters,  $\tau_0$  can be extracted at the high applied voltage;  $E_a$  and  $\alpha$  can be extracted by tuning the fitting of switching polarization under different voltages. The more detailed fitting process can be analogous to the FE model [16, 22]. In Fig. 5.2, my model results are in excellent agreement with the experimental data. It successfully captures the minor loop switching of AFE by tracking the

polarization history. Furthermore, we test the model with different rise/fall time. Fig. 5.3 shows the measurement and fitting of a 6 nm HZO with a rise/fall time of 50  $\mu\text{s}$  and 1  $\mu\text{s}$  [62]. When the ramp rate is fast enough,  $A_{l+}$  will not completely switch to  $A_{l-}$  and compensate  $A_{h+}$ , which let us can model the polarization retention at high speed. These demonstrations show that this model can reproduce the switching dynamics under various measurement conditions. In addition, I modeled the TiN/TiO<sub>2</sub>/HZO/RuO<sub>2</sub>/HfO<sub>2</sub> -stacked AFE [62] with 5 nm HZO and 0.5 nm TiO<sub>2</sub> in Fig. 5.4. The RuO<sub>2</sub> layer's high work function shifts the nonvolatile states from 0 V to -1V. To simulate this device, I also used the model's capability of accepting a series capacitance (TiO<sub>2</sub>) and the work function difference. Fig. 5.4 shows the fitting result.

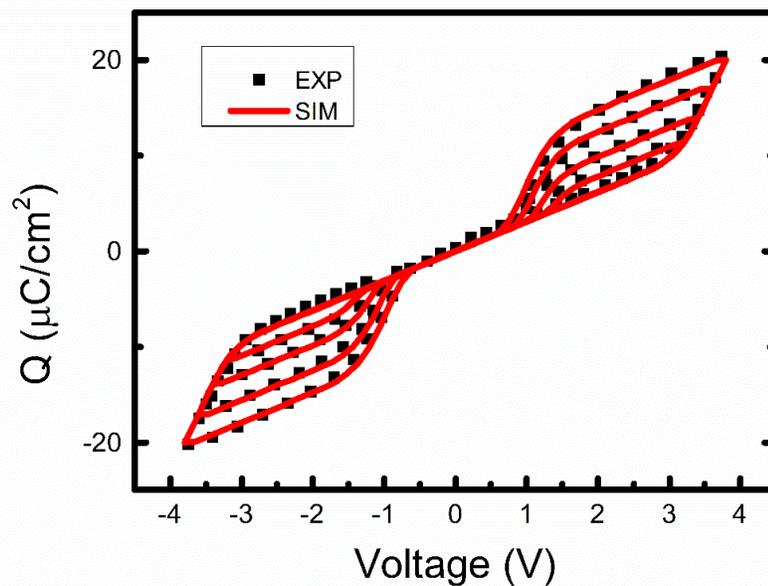


Fig. 5.2. Modeled hysteresis curve of a 10 nm ZrO<sub>2</sub> capacitor from [74] with  $P_R = 11\mu\text{C}/\text{cm}^2$ ,  $\tau_0 = 100\text{ps}$ ,  $E_a = 3.45\text{MV}/\text{cm}$ ,  $E_{T+} = E_{T-} = 2.26\text{MV}/\text{cm}$  and  $\alpha = 2.46$ . © 2025 IEEE

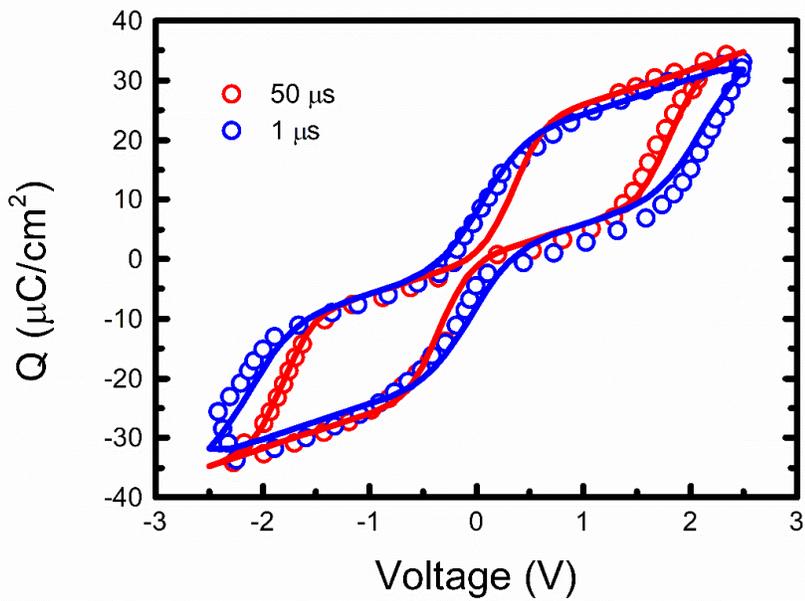


Fig. 5.3. Modeled hysteresis curve of a 6 nm HZO capacitor from [62] with  $P_R = 20\mu\text{C}/\text{cm}^2$ ,  $\tau_0 = 1.5\text{ns}$ ,  $E_a = 3.6\text{MV}/\text{cm}$ ,  $E_{T+} = E_{T-} = 1.8\text{MV}/\text{cm}$  and  $\alpha = 1.7$ . © 2025 IEEE

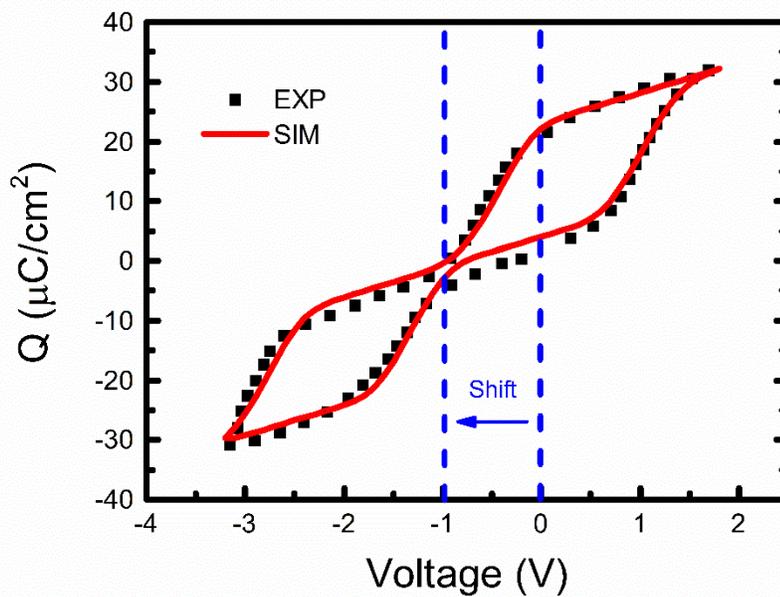


Fig. 5.4. Modeled hysteresis curve of a TiN/TiO<sub>2</sub>/HZO/RuO<sub>2</sub>/HfO<sub>2</sub>-stacked capacitor from [62] with  $P_R = 20\mu\text{C}/\text{cm}^2$ ,  $\tau_0 = 2\text{ns}$ ,  $E_a = 3.7\text{MV}/\text{cm}$ ,  $E_{T+} = 1.9\text{MV}/\text{cm}$ ,  $E_{T-} = 2.3\text{MV}/\text{cm}$  and  $\alpha = 1.7$ . © 2025 IEEE

## 5.4 Chapter Summary

I present a compact model of AFE capacitors. This model produces the AFE behaviors with the switching dynamics between positive-polarized, negative-polarized, and non-polarized states in each grain group in an AFE capacitor. This model computes the time-dependent polarization switching history through the entire simulated period of the capacitor operation. We demonstrate the model's capability to capture the major and minor loop switching as well as the speed-dependent hysteresis loop of AFE capacitors. We also show that this model can simulate the nonvolatile AFE stack, which is interested in future NVM applications.

## Chapter 6 Theoretical Study and Modeling of Nanoscale Ferroelectric Capacitor

In this chapter, I present a modeling study of nanoscale ferroelectric (FE) capacitors. I first used the phase-field simulation to study the polarization switching of a very small FE capacitor that contains only a few grains. I showed that, at higher applied voltage, the entire grain undergoes a single-domain-like switching, but, at lower applied voltage, the domain wall growth mechanism dominates due to the difference between the domain wall energies of bulk and defect nuclei. To create a compact model that includes this voltage dependence, I use a time-dependent domain switching model for each discrete grain with empirical modifications capturing the two different switching mechanisms. In addition, a voltage-dependent dielectric model is included to represent the nonlinear capacitance of the FE capacitor. I verified this compact model by fitting the results of phase-field modeling results with excellent agreement. This work is published by IEEE. © 2025 IEEE. Reprinted, with permission, from [77]

### 6.1 Motivation

Due to its compatibility with the standard CMOS technology, HZO-based FE memories become great candidates for future non-volatile memory (NVM) applications such as ferroelectric FET (FEFET), ferroelectric RAM (FERAM), and ferroelectric tunnel junction (FTJ) [7]. The polycrystalline nature of HZO provides the opportunity for multi-state operation on FE memories, which can be used in neuromorphic applications [8]. It is the multi-grain switching behavior to generate the continuous polarization states. Various experiments and modeling works have been done to study the switching behavior of FE devices [12, 14, 22, 38] using the nucleation-limited switching (NLS) model with a statistically distributed switching rate.

However, when the device area becomes very small, we can no longer describe it with a continuous distribution function. A nanoscale FE capacitor only contains a few grains, which leads to discrete polarization switching [47, 78]. The NLS model also assumes that the switching is dominated by nucleation due to the small size of the grains. The effect of domain propagation might be therefore averaged out by the statistical distribution of a large number of grains in a large-area capacitor. However, it cannot be neglected in the case of a small-area capacitor where there are only a few grains present. As the technology becomes more mature, this kind of nanoscale FE capacitor will be introduced into high-density integrated circuits (ICs). It is essential for us to study its characteristics and have a SPICE-compatible compact model for IC simulation.

Due to the lack of experimental data and the difficulty of measuring polarization-voltage characteristics of such small devices, we need to rely on numerical simulation to study the device. Phase-field modeling using time-dependent Ginzburg Landau (TDGL) theory is the most common way to study FE devices [20, 72, 79, 80]. In this chapter, I studied the properties of small-area FE by performing phase-field simulations of FE capacitors with grain sizes within 20 nm [47, 81]. I then developed a compact model to capture the electrical characteristics resulting from the TDGL model.

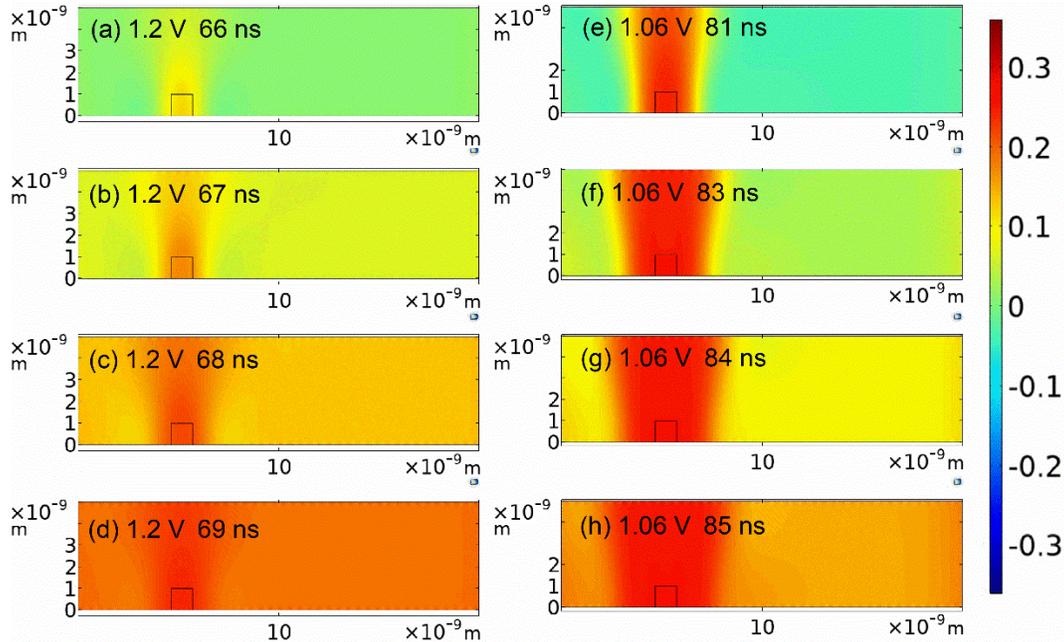


Fig. 6.1. Polarization switching of a single-grain FE capacitor at 1.2 V and 1.06 V. The legend shows the polarization in the unit of  $C/m^2$ . © 2025 IEEE

## 6.2 Phase-field Simulation

To simulate the nanoscale FE, I conducted a 2D phase-field simulation using COMSOL Multiphysics [82]. First, I studied the switching under a single grain FE capacitor with 20 nm width and 5 nm thickness. Voltage is applied on the top and bottom boundary. I set remanent polarization  $P_R=22mC/cm^2$  and coercive field  $E_C=2MV/cm^2$  [63]. Inside the capacitor, I add a 1 nm by 1nm nucleus which has a  $E_C$  smaller than the bulk value [72] shown in Fig. 6.1. Here, I set it to be  $0.7E_C$  of bulk. The coupled Poisson-TDGL simulation is set as Eq. (6.1) where  $V$  is electrical potential,  $P_j$  is the polarization component in the  $j$  direction where  $j = x$  or  $y$ ,  $\epsilon_r$  is the relative permittivity ( $= 30$ ),  $\epsilon_0$  is the vacuum permittivity,  $\rho$  is the viscosity coefficient ( $=6 \Omega m$ ),  $\alpha, \beta, \gamma$  are the Landau-Khalatnikov coefficients [71], and  $g$  is the polarization gradient coefficient ( $=1E-10 m^3/F$ ). For simplicity, I assume a dielectric relation in the  $x$  direction since, in this study, there is no voltage applied in that direction. I apply  $\vec{\nabla}P \cdot \vec{n} = 0$  for every grain boundary where  $\vec{n}$  is the normal vector [83].

$$\vec{\nabla} \cdot (-\epsilon_r \epsilon_0 \vec{\nabla} V + P) = 0 \quad (6.1a)$$

$$0 = -\epsilon_r \epsilon_0 \frac{\partial V}{\partial x} - P_x \quad (6.1b)$$

$$\rho \frac{\partial P_y}{\partial t} = -\frac{\partial V}{\partial y} - 2\alpha P_y - 4\beta P_y^3 - 6\gamma P_y^5 + 2g\nabla^2 P_y \quad (6.1c)$$

Fig. 6.1 shows the simulation results of this FE capacitor for 1.2 V and 1.06 V, respectively. I plot  $P_y$  of the entire capacitor at different simulation time instants. We can see that the polarization begins to switch at the nucleus followed by the bulk at 1.2 V. It is similar to single-domain switching since the electric field is strong enough to make the nucleus and bulk switch almost at the same time. For 1.06 V, since the applied voltage is close to the  $E_C$  of the bulk, the bulk will switch much slower than the nucleus. The domain propagation can be clearly seen in this case. It becomes a combination of domain wall switching and bulk switching. This phenomenon might be averaged out when there are a lot of grains with different sizes and material properties.

I also study the case where the device contains 3 grains in the width direction. Each grain width is set to be 16 nm [81]. I choose 3-grain because the dimensions of a short channel transistor for state-of-the-art technologies are typically around 20nm×50nm. Given the typical size of grain in HZO is 16~20nm, I think a 3-grain scenario is good enough to study these devices. For example, in [47], a 3-grain switching behavior was recently observed for a short channel FEFET. At the grain boundary, I assume there is a 0.5 nm dielectric dead layer [84, 85] that will stop the domain propagation. Moreover, in polycrystalline FE, each grain can have different FE properties [14]. Here, for simplicity, I set these grains to have different  $E_C$ , which are 1.7 MV/cm, 2.55 MV/cm and 2 MV/cm, so that it can perform the discrete switching seen in [47]. Inside each grain, I also place a nucleus the same as the one-grain case. Fig. 6.2 shows that each grain will switch at different times since they have different domain energy.

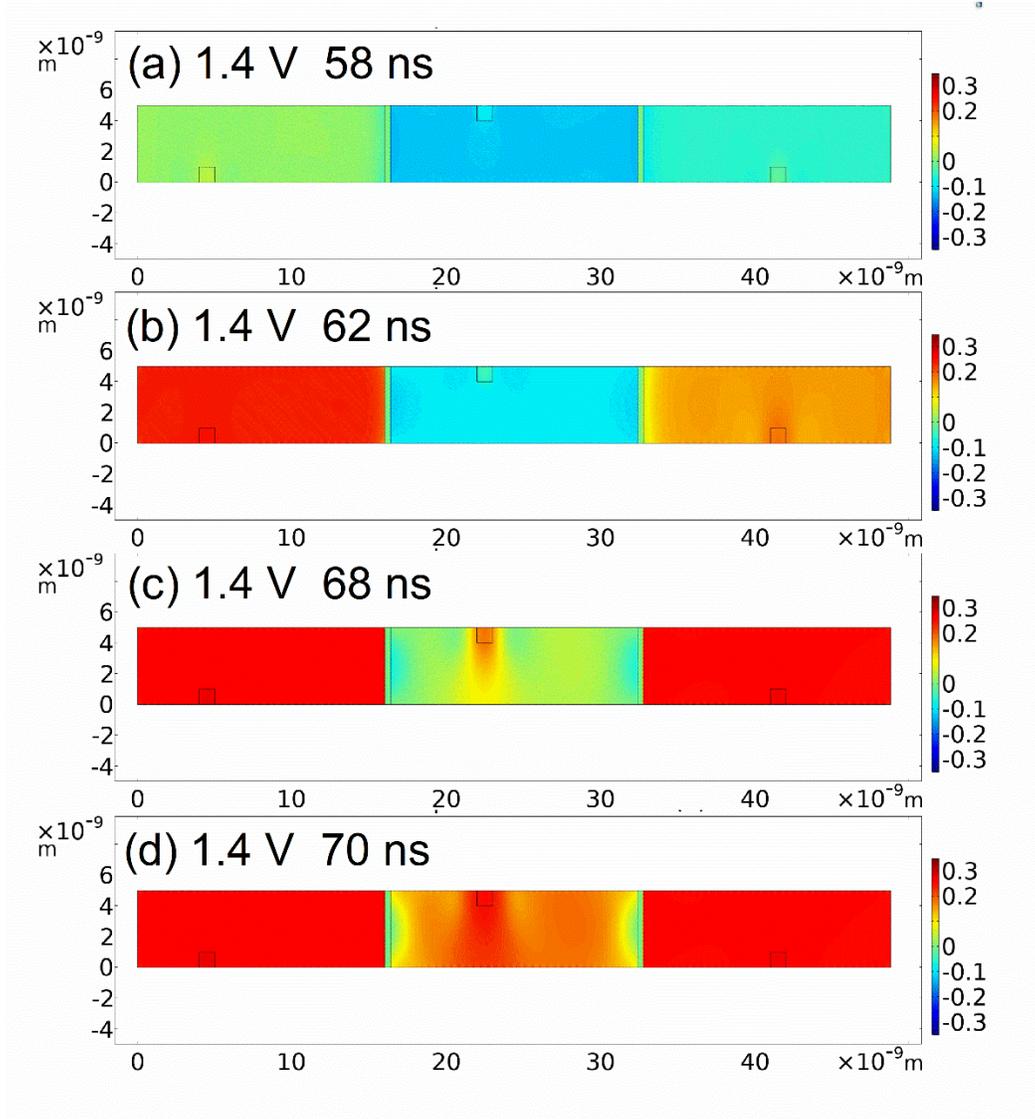


Fig. 6.2. Polarization switching of a 3-grain FE capacitor at 1.4 V. The legend shows the polarization in the unit of  $C/m^2$ . © 2025 IEEE

### 6.3 Compact Model

I now developed a compact model for nanoscale FE capacitors based on the phase-field simulation results. From Fig. 6.1 & 6.2, we can see that due to the nuclei and multi-grains, a single domain LK equation is not able to describe the system. Depending on the computational penalty, directly coding the phase-field model with Verilog-A code is not suitable for a compact model [79]. Therefore, I decided to adopt domain-switching models such as

(Kolmogorov-Avrami-Ishibashi) KAI model and (nucleation Limited Switching) NLS model [9, 10, 12]. I previously developed a compact model for large-area FE capacitors based on these models in Chapter 2 [38]. I use the same framework to develop a unified compact model valid for both the large and small-area capacitors.

In this compact model, the FE capacitor is described as a parallel combination of many small FE capacitors, each representing a grain as shown in Fig. 6.3. As stated earlier, the switching mechanism turns from homogeneous switching to domain nucleation and growth switching when the applied voltage decreases. To model this complex switching mechanism, I use Eq. (6.2) to (6.4) [38], where the subscript  $i$  means the index of the grain which starts from 1,  $P_i$  is the polarization of each grain,  $\eta$  is the fraction of the polarization of each grain to the entire capacitor,  $\tau_i$  is a time constant following the Merz's law [76],  $\tau_{0i}$  is the characteristic time constant,  $E_{ai}$  is the activation field,  $t_k$  is when the electric field changes polarity, and  $\alpha_i$  and  $\beta_i$  are fitting parameters. To capture these two switching behaviors for low and high voltages, I make  $\alpha_i$  voltage dependent as Eq. (6.5) with empirical fitting parameters ( $\alpha_{0i}$ ,  $\delta_i$ ,  $\gamma_i$  and  $V_{0i}$ ).  $\alpha_i$  will affect the rising time and the spacing between the switching curves at different voltages (Fig. 6.4). Therefore,  $\alpha_i$  value varies from high voltage to low voltage conditions. Since there are only a few grains, we can no longer use a distribution function to describe the grain variation

$$\begin{aligned} \frac{dP_i}{dt} &= \frac{-P_i + \eta_i P_R}{\tau_i(t)}, \text{ if } E_{FE}(t) \geq 0, \\ &= \frac{-P_i - \eta_i P_R}{\tau_i(t)}, \text{ if } E_{FE}(t) < 0 \end{aligned} \quad (6.2)$$

$$\tau_i(t) = \tau_{0i} \exp\left(\left(\frac{E_{ai}}{|E_{FE}(t)|}\right)^{\alpha_i}\right) \quad (6.3)$$

$$\begin{aligned} P_i(t) &= \eta_i P_R - (\eta_i P_R - P_i(t_k)) \exp\left(-\left(\int_{t_k}^t \frac{1}{\tau_i(t')} dt'\right)^{\beta_i}\right), \\ &\text{if } E_{FE}(t) \geq 0, \\ &= -\eta_i P_R + (\eta_i P_R + P_i(t_k)) \exp\left(-\left(\int_{t_k}^t \frac{1}{\tau_i(t')} dt'\right)^{\beta_i}\right), \end{aligned}$$

if  $E_{FE}(t) < 0$  (6.5)

$$\alpha_i(V_{FE}) = \alpha_{0i} - \delta_i \tanh(\gamma_i(V_{FE} - V_{0i})) \quad (6.6)$$

$$Q_B(V_{FE}) = \frac{\varepsilon_{FE0}}{t_{FE}} \left( V_{FE} + \frac{aV_{FE}}{1+b(V_{FE}+c \tanh(dP_{sw}))^2} \right) \quad (6.7)$$

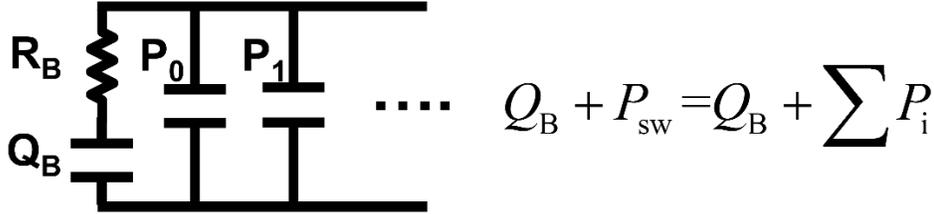


Fig. 6.3. The equivalent circuit of our nanoscale FE capacitor compact model. © 2025 IEEE

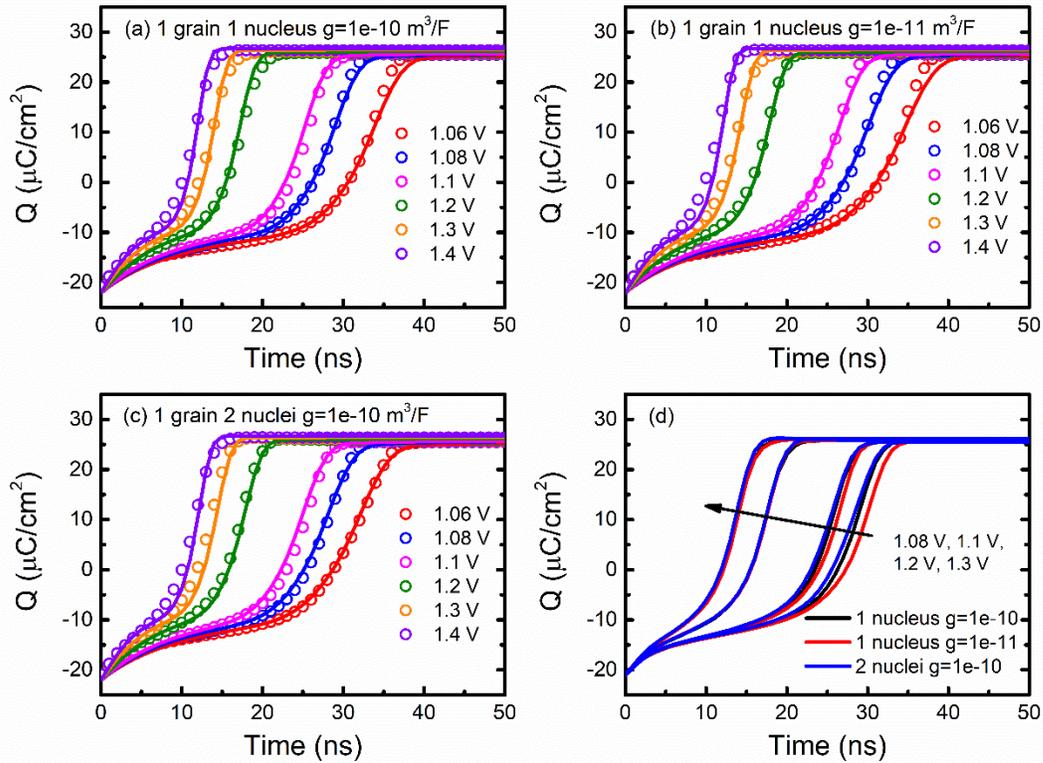


Fig. 6.4. Model fitting of the single-grain FE capacitors under 3 cases: (a) 1 grain 1 nucleus  $g=1e-10 \text{ m}^3/\text{F}$ , (b) 1 grain 1 nucleus  $g=1e-11 \text{ m}^3/\text{F}$ , (c) 1 grain 2 nuclei  $g=1e-10 \text{ m}^3/\text{F}$ . The symbols are the COMSOL simulation, and the lines are the compact model. (d) shows the comparison between three cases. © 2025 IEEE

Fig. 6.4 shows my fitting results compared to phase-field simulation for three different cases: 1 nucleus with  $g=1e-10 \text{ m}^3/\text{F}$ , 1 nucleus with  $g=1e-11 \text{ m}^3/\text{F}$  and 2 nuclei with  $g=1e-10 \text{ m}^3/\text{F}$ . We can see that polarization first rises slowly because of the background dielectric response which is not included in the NLS and KAI models. To capture that, I use the same equations Eq. (6.2) to (6.4) (with the subscript  $i=0$ ) and include the background dielectric charge  $Q_B$  as in Eq. (6.6) in this model in series with a resistance  $R_B$  where  $t_{\text{FE}}$  is the FE thickness.  $Q_B$  is expressed as a nonlinear function of  $V_{\text{FE}}$  to model the nonlinearity in the background capacitance where  $\epsilon_{\text{FE0}}$  is the nominal permittivity, and  $a$ ,  $b$ ,  $c$  and  $d$  are fitting parameters. The total charge is calculated as  $Q_{\text{sw}B}$  where  $P_{\text{sw}} = \sum P_i$  which is shown as an equivalent circuit in Fig. 6.3. This model can capture the switching under different voltages for the three cases shown in Fig. 6.4. We can also see that the larger the  $g$  and number of nuclei, the faster the switching, especially when the voltage is low.

Now, we move on to the three grains case. Since I chose a large  $E_C$  variation in my simulation, I need to use  $i=0, 1, 2, 3$  to fit the data. The result is shown in Fig. 6.5. The discrete switching comes from the  $E_C$  variation of each grain. By selecting different parameters of each  $P_i$ , the model can fit the phase-field simulation well. Then, we simulate a P-V loop with a 2.5 MHz triangular voltage wave as shown in Fig. 6.6. It shows Eq. (6.6) can fit the nonlinear background capacitance behavior well as compared to the linear charge model. The result shows a good agreement with the COMSOL simulation.

To summarize the difference between the proposed small FE capacitor model and the previous large FE capacitor model [38], first, the grain distribution is considered discrete instead of continuous due to the small number of grains; second, I include the nonlinear effect in the background dielectric rather than using a linear background capacitance; third, I modify the switching rate of this model to account for the different switching behaviors at low voltages and high voltages

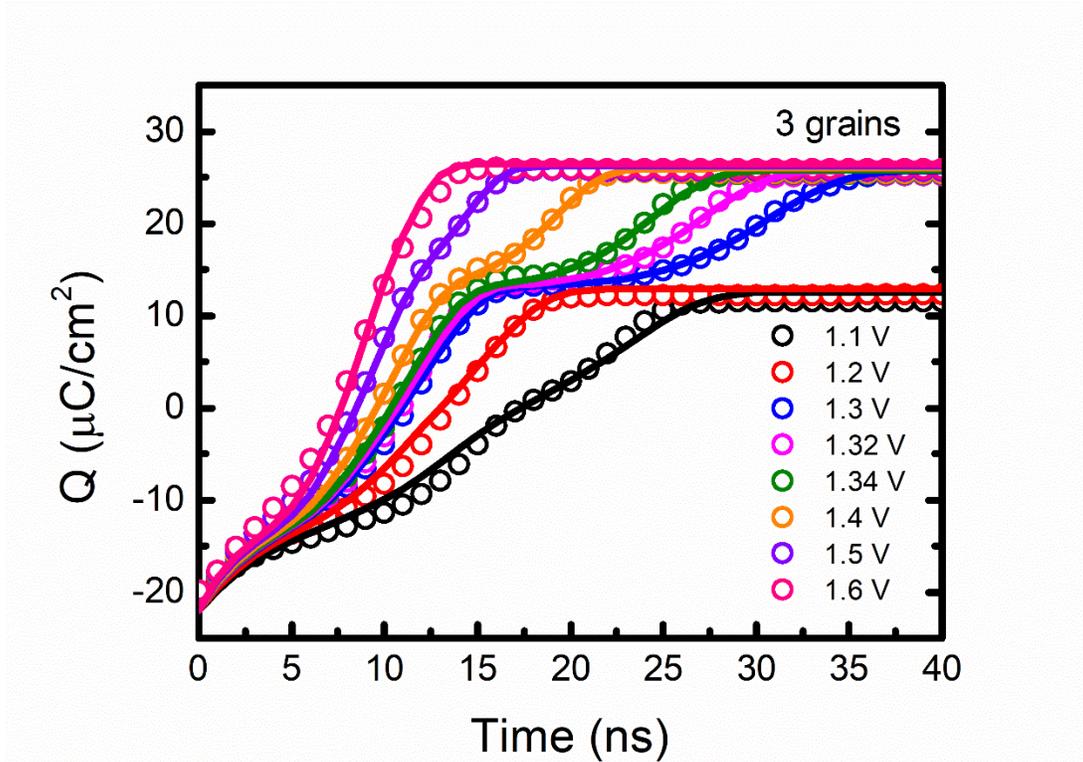


Fig. 6.5. Model fitting of the 3-grain FE capacitor. © 2025 IEEE

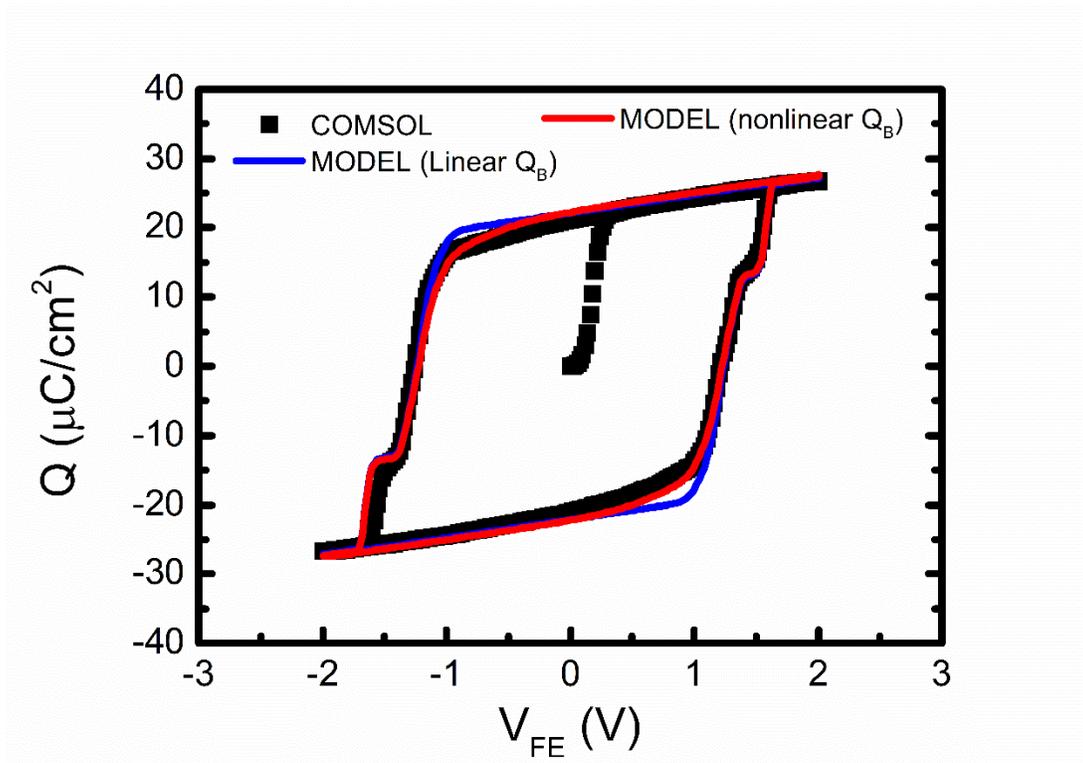


Fig. 6.6. The comparison between the COMSOL simulated PV curve and the compact model simulations using linear  $Q_B$  and nonlinear  $Q_B$ . © 2025 IEEE

## 6.4 Chapter Summary

I have analyzed the switching behavior of very small FE capacitors that contain a few ferroelectric crystal grains using phase-field simulation. The results show that, at higher voltage, the switching mechanism is almost single domain switching and, at lower voltage, it is a combination of domain wall switching and bulk spontaneous switching. I used these results to develop a compact model based on our previous large-area FE capacitor model. This model can simulate the change in the switching mechanism of the nanoscale FE capacitor at all voltages including the discrete switching of different grains and the non-linear capacitance of the FE material.

## Chapter 7 A Versatile Compact Model of Resistive Random-Access Memory (RRAM)

I present a versatile compact model for resistive random-access memory (RRAM) that can model different types of RRAM devices such as oxide-RRAM (OxRAM) and conducting-bridge-RRAM (CBRAM). The model unifies the switching mechanisms of these RRAMs into a single framework. We showcase the model's accuracy in reproducing published experimental device DC and transient characteristics of various RRAM structures. We also demonstrate the model's efficacy in capturing RRAM variability and conducting 1T1R circuit simulations.

This work is published in [86]. © 2025. This manuscript version is made available under the CC-BY-NC-ND 4.0 license

<https://creativecommons.org/licenses/by-nc-nd/4.0/>.

### 7.1 Motivation

Resistive memory is a promising technology for future nonvolatile RAM, in-memory and analog computing, and neuromorphic applications, RRAM can increase computing power by breaking the performance bottleneck of traditional architecture and therefore it reduces overall cost for large computing systems. It has the potential to offer high density, fast speed, and low energy consumption [87-90]. Very general structure of RRAM is the dielectric sandwiched between two electrodes, and the commonly used dielectrics in RRAM include hafnium oxide ( $\text{HfO}_2$ ), titanium oxide ( $\text{TiO}_2$ ), aluminum oxide ( $\text{Al}_2\text{O}_3$ ), zinc oxide ( $\text{ZnO}$ ), silicon oxide ( $\text{SiO}_2$ ), tantalum oxide ( $\text{Ta}_2\text{O}_5$ ), niobium oxide ( $\text{Nb}_2\text{O}_5$ ), and yttrium oxide ( $\text{Y}_2\text{O}_3$ ). These materials are chosen for their high dielectric constants, stability, and compatibility with existing semiconductor manufacturing processes. Each offers unique advantages, such as  $\text{HfO}_2$ 's good switching characteristics,

TiO<sub>2</sub>'s ease of integration, Al<sub>2</sub>O<sub>3</sub>'s dielectric strength, and ZnO's transparency. The selection of dielectric is critical for ensuring reliable and repeatable resistive switching behavior, which is essential for the performance and scalability of RRAM devices [91]. The switching mechanism of RRAM can be generally described as the growth and regression of a conductive path of metal or oxygen vacancies. The former is the switching mechanism of conducting-bridge RAM (CBRAM) [92] and the latter is that of oxide-based RAM (OxRAM) [93]. An RRAM can have both mechanisms working at the same time [94]. Ideally, an RRAM compact model should be robust and flexible enough to model the common characteristics of all these cases but simple and very fast for simulating large integrated circuits (IC).

There are many studies on the compact modeling of RRAM. Most of them focus on OxRAMs. Most models [95-99] model a filament growth in the thickness direction of the OxRRAM. Empirical functions are introduced to fit the RESET process in different OxRAMs such as bilayer RRAM and Al-doped HfO<sub>x</sub>-based RRAM. The models in [100-102] explicitly calculate the growth in both length and width of the filament, vacancy generation, ions hopping, recombination, and other physical phenomena. For instance, [100] provides a comprehensive model addressing both DC and AC operations of metal-oxide-based RRAM. Their model incorporates detailed physics-based mechanisms such as filament growth dynamics and ion transport, significantly increasing computational complexity and parameter extraction time. Similarly, [101] focuses on the variability and reliability aspects of RRAM design, considering factors like filament growth variations, stochastic vacancy generation, and ion hopping mechanisms. This comprehensive approach results in increased computational overhead and a more cumbersome parameter extraction process. Furthermore, [102] presents an analytical model for bipolar resistive memories and complementary resistive switches, including detailed calculations of vacancy generation, recombination processes, and filament dynamics. This model, although analytically rigorous, involves significant complexity for SPICE implementation, which is crucial for robust SPICE simulation.

Different from OxRAM, CBRAM's conduction path is mainly conducted through the metal filament [103], [104], and a few studies have been done to

empirically describe the different stages of the switching process [105], [106]. The “S. Yu and H.S.P Wong” in [107] has developed a physics-based model specific to CBRAM using the dependence of ion migration velocity on the electric field along with the vertical and lateral growth/dissolution dynamics of the metallic filament. This model considers both time-dependent transient and quasi-static switching characteristics of the Conductive Bridging RAM (CBRAM).

In this work, I developed an RRAM compact model with the goal of high flexibility and accuracy to fit different types of RRAMs with special attention paid to ease of parameter extraction, simulation efficiency, robustness, and convergence in commercial simulators.

## 7.2 Model

Fig. 7.1 shows the schematic diagram of this RRAM model, including parasitics ( $R_s$  &  $C_p$ ). I simplify the complex 3D conducting profile by consolidating these conducting paths into a single column described by  $L_H$ . The device is divided into a low-resistance region (LRR) and a high-resistance region (HRR). The LRR may contain high densities of metal ions or oxygen vacancies, while the HRR typically consists of metal-oxide or insulator. This general representation encompasses various types of RRAMs such as OxRAM and CBRAM due to their resemblances in the switching dynamics.

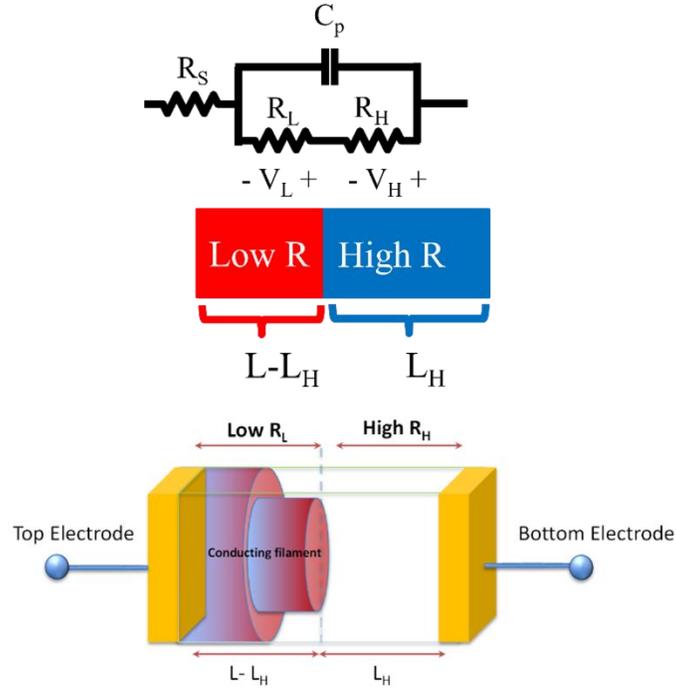


Fig. 7.1. The schematic diagram and equivalent circuit of the conducting path in a resistive memory.

During the SET, or turn-on, process, the decrease in the length of the HRR, which is also the growth of the LRR, can be described as Eq. (7.1) where  $v_S$  is a velocity factor,  $\alpha_S$  is a field factor for SET,  $q$  is the elementary charge,  $V_H$  is the voltage across  $L_H$ ,  $E_{AS}$  is the activation energy for LRR growth,  $k_B$  is the Boltzmann constant, and  $T$  is the temperature of the filament including the self-heating effect. Eq. (7.1) describes that when the given energy is above the potential barrier the ions will move to reduce  $L_H$ . As voltage, current, and  $T$  increase, the resistance suddenly drops when the electric field across  $L_H$  reaches a threshold.

$$\left. \frac{dL_H}{dt} \right|_S = -v_S \exp\left(\frac{\alpha_S q V_H(t) / L_H - E_{AS}}{k_B T}\right) \quad (7.1)$$

During the RESET or HRR growth process,  $L_H$  is modeled with Eq. (7.2a) with a similar form as Eq. (7.1), where  $v_R$  is a velocity factor,  $\alpha_R$  is the field factor for RESET, and  $E_{AR}$  is the activation energy for RESET.  $V_{\text{eff}}$  is an empirical function Eq. (7.2b) that serves as an effective voltage applied in the RESET process Eq. (7.2a) that models several switching mechanisms

happening in the RESET. These different mechanisms [100] can be viewed as voltage-driven switching that depends on a combination of  $V_R (=V_L+V_H)$  and  $V_H' (=V_H+V_{off})$ . Parameters  $a$ ,  $b$ , and  $V_{off}$  are fitting parameters.

$$\left. \frac{dL_H}{dt} \right|_R = v_R \exp\left(\frac{-\alpha_R q V_{eff}(t)/L_H - E_{AR}}{k_B T}\right) \quad (7.2a)$$

$$V_{eff}(t) = \frac{V_R(t) + aV_H'(t) + \frac{bV_R(t)V_H'(t)}{1V}}{1+a+b} \quad (7.2b)$$

Self-heating plays an important role in the switching of RRAMs [108] because the filament temperature  $T$  accelerates the SET/RESET mechanisms.  $T$  can be modeled with a first-order differential equation with a thermal capacitance  $C_{th}$  and a thermal resistance  $R_{th}$ . Eq. (7.3) can be solved by the circuit simulator. The temperatures in Eq. (7.1) and Eq. (7.2) are updated by Eq. (7.3) during each time step and self-consistent iterations by SPICE simulators (Fig. 7.3 inset). All the results consider the self-heating including DC and transient.

$$C_{th} R_{th} \frac{dT}{dt} = -T + T_{amb} + R_{th} V_R I \quad (7.3)$$

RRAM is also known for its statistical switching behavior due to the random location and movements of the ions and vacancies. This model is compatible with the published variation models. For example, users can add a random noise term to  $dL_H/dt$  such that  $dL_H/dt = dL_H/dt + \text{Noise}$  to perform stochastic simulations [95-97, 101].

The total resistance is the sum of HRR resistance and the LRR resistance. The conduction mechanism in RRAMs can be complicated. It may include direct tunneling, trap-assisted tunneling, electron hopping, Fowler–Nordheim tunneling, and so on. I model the HRR resistance with an empirical tunneling resistance equation Eq. (7.5) where  $r_H$  is the resistance factor, and  $L_0$ ,  $V_{H0}$ ,  $n_H$ ,  $n_{VH}$  are model-fitting parameters. Eq. (7.5) keeps the general physics understanding that tunneling current has an exponential dependence on distance and the resistance will be shorted when  $L_H$  becomes zero.  $n_H$  is introduced because the filament may not be an exact straight line. It is used to empirically capture those effects.

$$R_H = \frac{r_H L_H^{n_H} \exp(L_H/L_0)}{1 + (V_H/V_{H0})^{2n_{VH}}} \quad (7.5)$$

The LRR resistance is commonly assumed to be an ohmic resistance. However, from the experimental data, I found that the RESET state IV characteristics could be nonlinear (e.g., Fig. 7.10). Therefore, we include the observed nonlinearity in the LRR resistance, Eq. (7.6), where  $r_L$  is the resistance factor,  $V_L$  is the voltage across the low-resistance region, and  $V_{L0}$ ,  $n_L$ , and  $n_{VL}$  are model fitting parameters.  $n_L$  is the same as  $n_H$ , which is used to increase fitting flexibility.

$$R_L = \frac{r_L (L - L_H)^{n_L}}{1 + (V_L/V_{L0})^{2n_{VL}}} \quad (7.6)$$

To model the asymmetric IV characteristics seen in the experimental data [95], the  $R_L$  and  $R_H$  are allowed to be different in positive and negative voltage regions.

### 7.3 Model Validation

This model is coded into Verilog-A and verified with published experimental data using commercial SPICE simulators. During the simulation, SPICE solves the equivalent circuit (Fig. 7.1) (with any peripheral circuits such as a drive transistor) to find out  $L_H$  and current at each time point. Fig. 7.2 shows model calibration with a HfO<sub>x</sub>/AlO<sub>x</sub> bilayer RRAM [93]. The RRAM was operated under multilevel switching using several  $V_{RESET}$  (the maximum negative voltage sweep) and, after each RESET, it was set with a positive voltage ramp to 2.5V. Fig. 7.3 is the same device under transient pulse measurement with a -2V pulse for 500ns and a -2.3V for 50ns [109] where the model also delivers an excellent match. The dynamic of the self-heating Eq. (7.3), and Fig. 7.4 can be important in impacting the peak transient current and the time delay as shown in Fig. 7.3. In addition, I simulate the case if the RRAM is SET using current ramping as shown in Fig. 7.5 using the same parameter set as Fig. 7.2 & 7.3. Instead of abrupt switching under voltage ramping, the model predicts a negative conductivity region during current ramping, which is experimentally observed in [110].

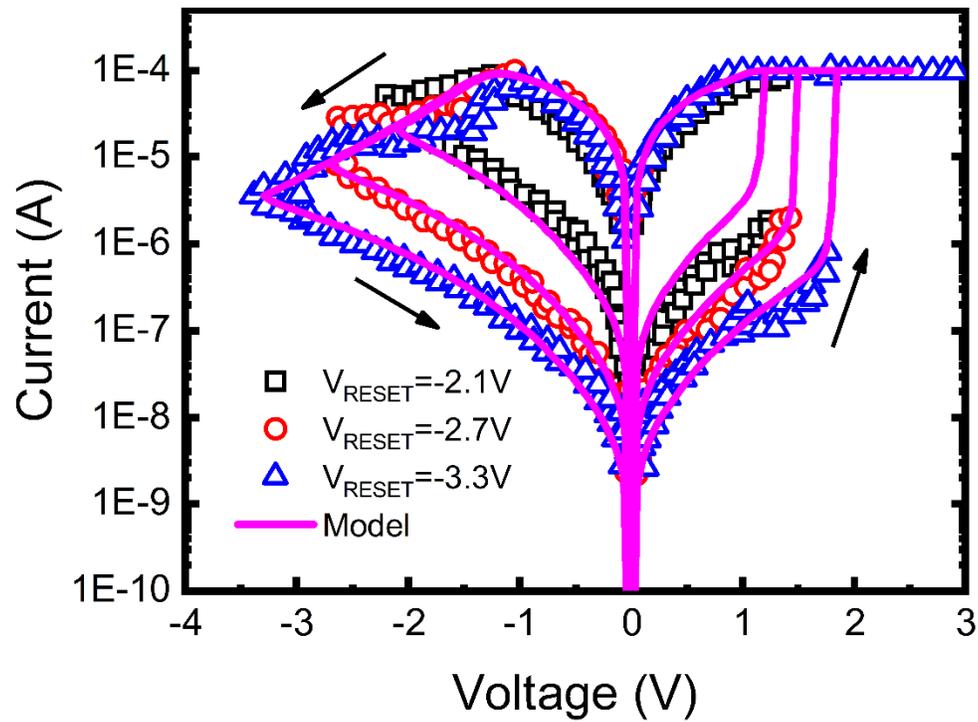


Fig. 7.2. Model validation of a HfO<sub>x</sub>/AlO<sub>x</sub> bilayer RRAM [93]. The symbols are experimental data. The lines are model simulations.

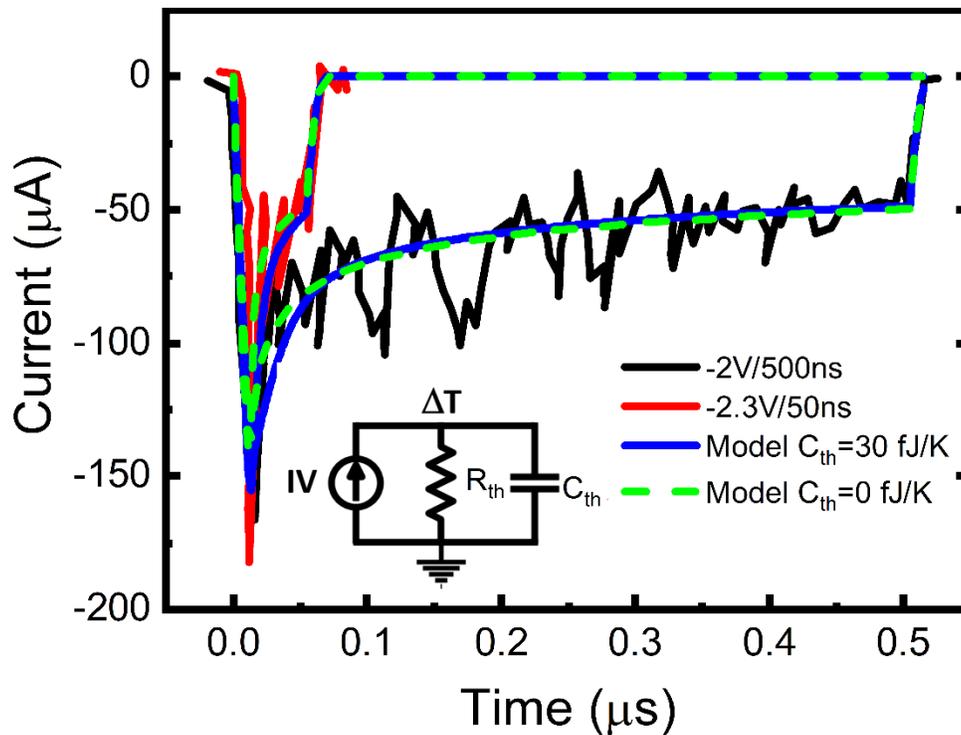


Fig. 7.3. Model validation with a transient measurement of a HfO<sub>x</sub>/AlO<sub>x</sub> bilayer RRAM

[93]. Black and red lines are experimental data. The blue line is the model simulation with  $C_{th}$ . The green dash line is the model simulation without  $C_{th}$ .

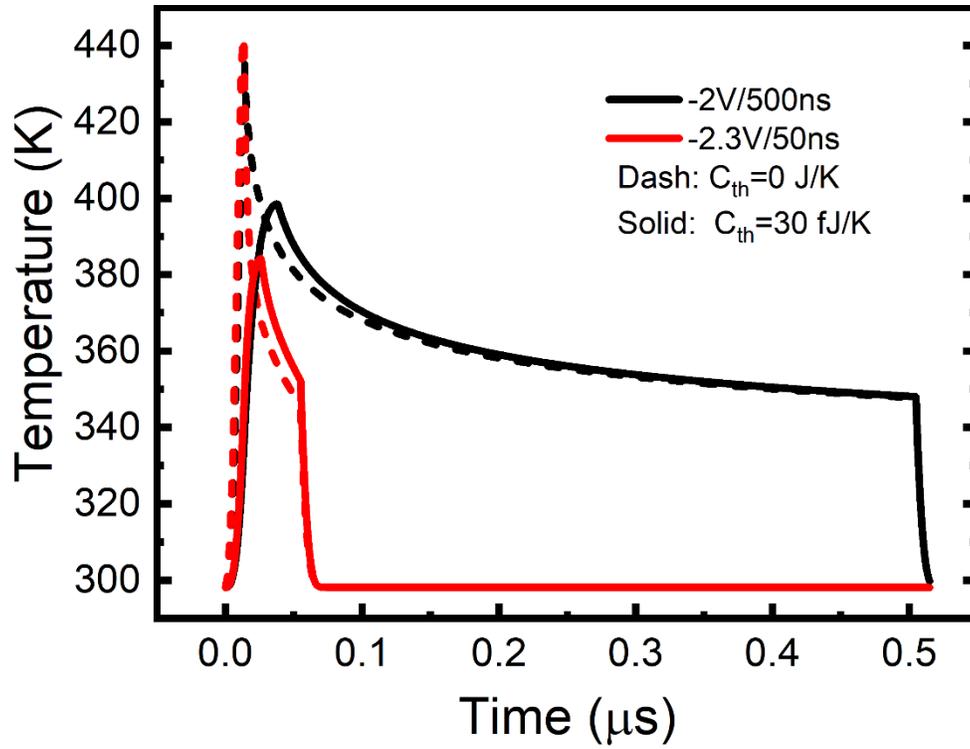


Fig. 7.4. Model simulated dynamics of temperature corresponding to Fig. 7.3.

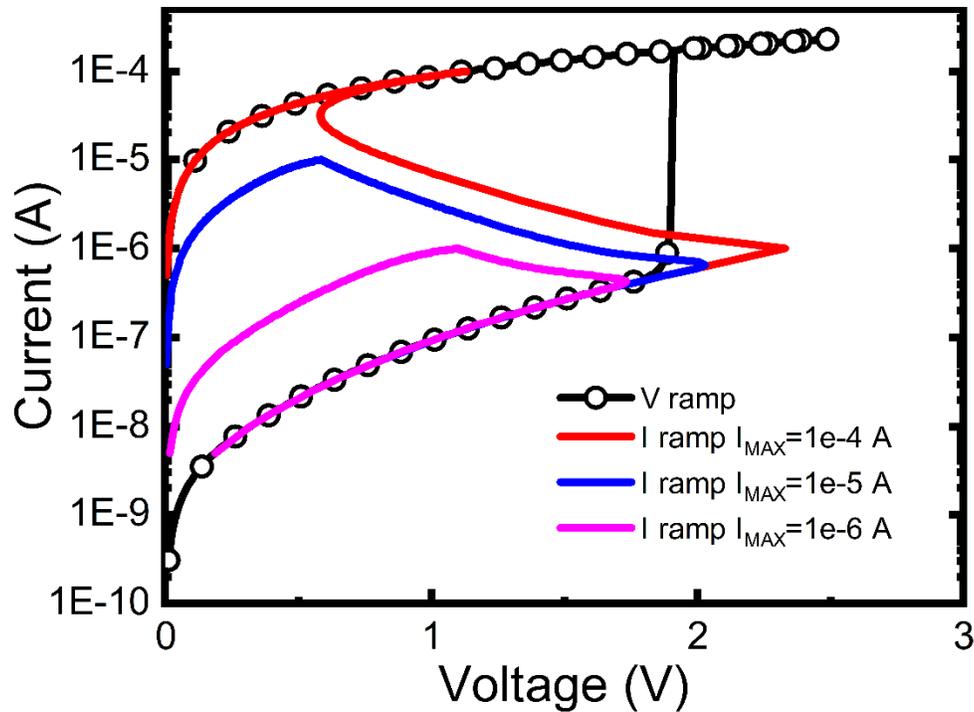


Fig. 7.5. Model simulations using current ramps to several different maximum current.

In Fig. 7.6, the model is tested on TiN/TiO<sub>x</sub>/HfO<sub>x</sub>/Pt planar RRAMs at several RESET voltages. Unlike Fig. 7.2, the device is not SET back after partial RESET. The corner of the RESET curve is fitted well by tuning (extracting) the  $a$ , and  $b$  parameters in Eq. (7.2b). Fig. 7.7 shows switching at several  $I_{\text{COMP}}$ . Using  $I_{\text{COMP}}$ , I can achieve multilevel switching in SET. To further demonstrate the model's flexibility, I verify the model on Al-doped HfO<sub>x</sub> RRAMs with different AlO<sub>x</sub> concentrations [95] in Fig. 7.8. The model fits the different shapes of RESET curves and the IV characteristics in the LR state and HR state.

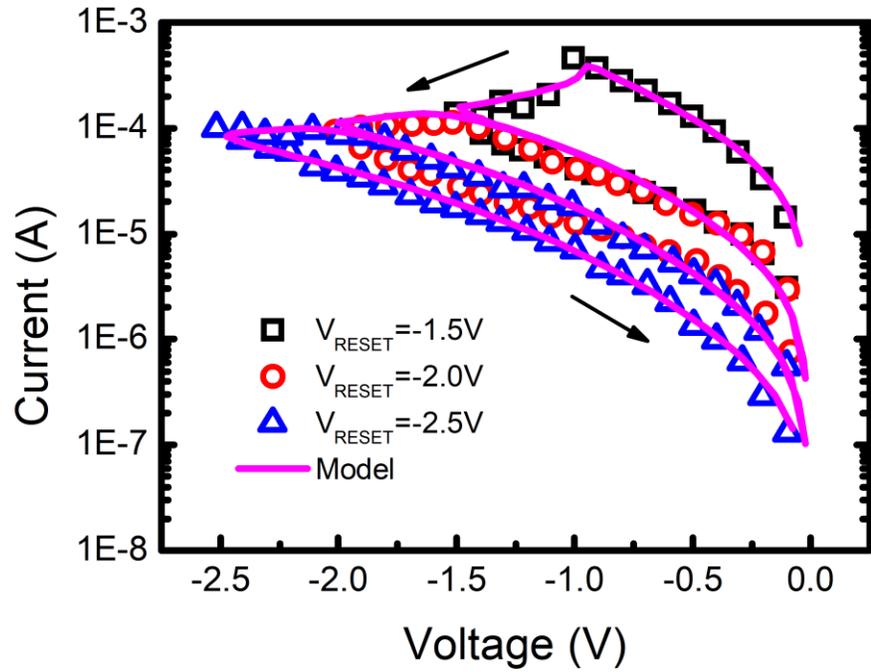


Fig. 7.6. Model validation with a TiN/TiO<sub>x</sub>/HfO<sub>x</sub>/Pt planar RRAM [95] being reset to several  $V_{\text{RESET}}$ .

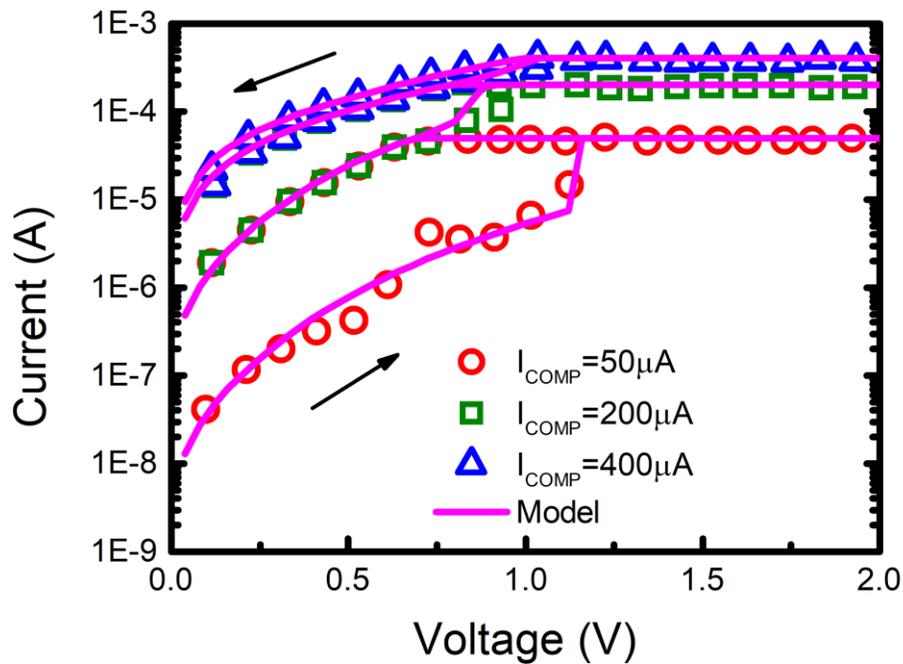


Fig. 7. 7. Model validation of a TiN/TiO<sub>x</sub>/HfO<sub>x</sub>/Pt planar RRAM [95] at different  $I_{\text{COMP}}$ . The symbols are experimental data. The lines are model simulations.

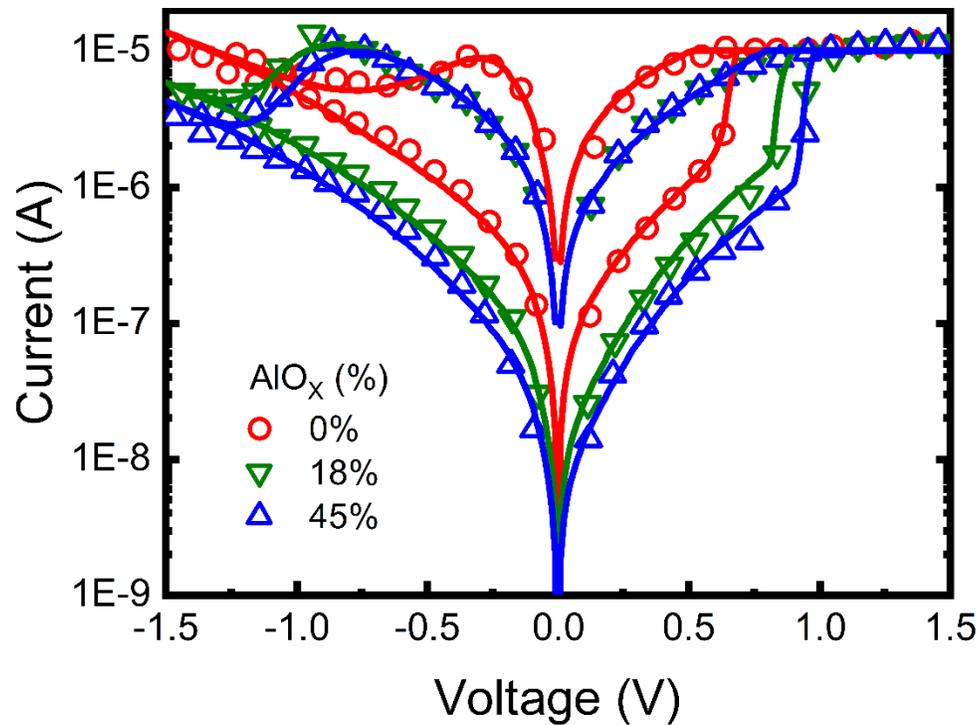


Fig. 7.8. Model validation of Al-doped HfO<sub>x</sub> RRAMs [95] with different AlO<sub>x</sub> concentrations. The symbols are experimental data. The lines are model simulations.

The model can not only be applied to OxRAMs but also CBRAMs. Fig. 7.9 shows the fit of the model to a CBRAM [92] with very good accuracy. I further demonstrate that it can also simulate the device that exhibits both ion and oxygen vacancy switching as shown in Fig. 7.10 [94] where the sharp switching in the RESET is evidence of the ion type switching coexisting with smooth oxygen vacancy switching.

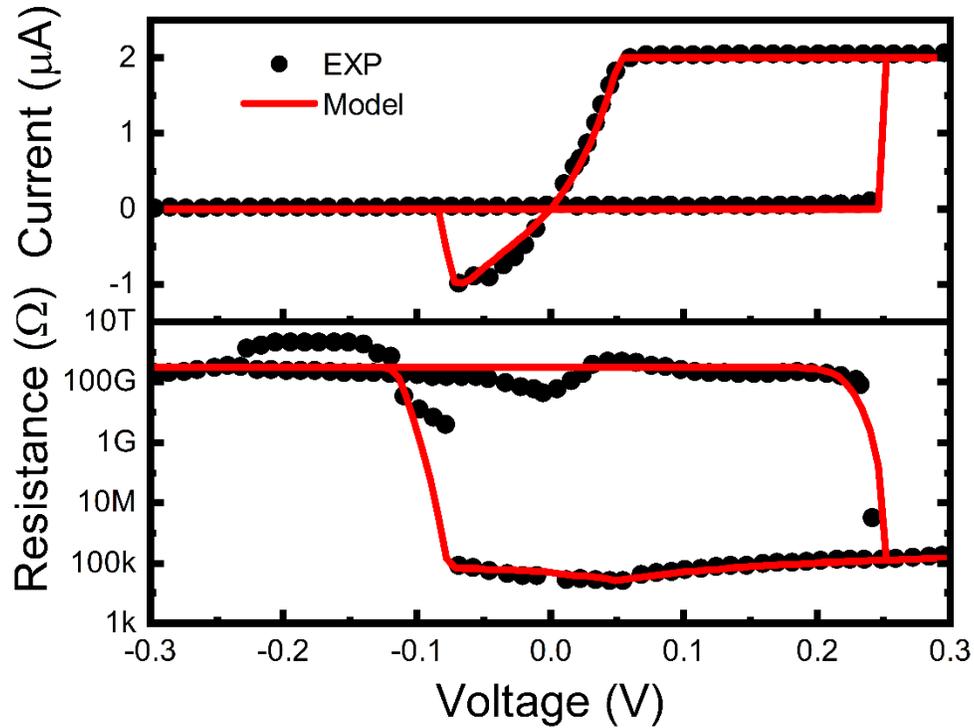


Fig. 7.9. Model validation of a CBRAM [92]. The symbols are experimental data. The lines are model simulations.

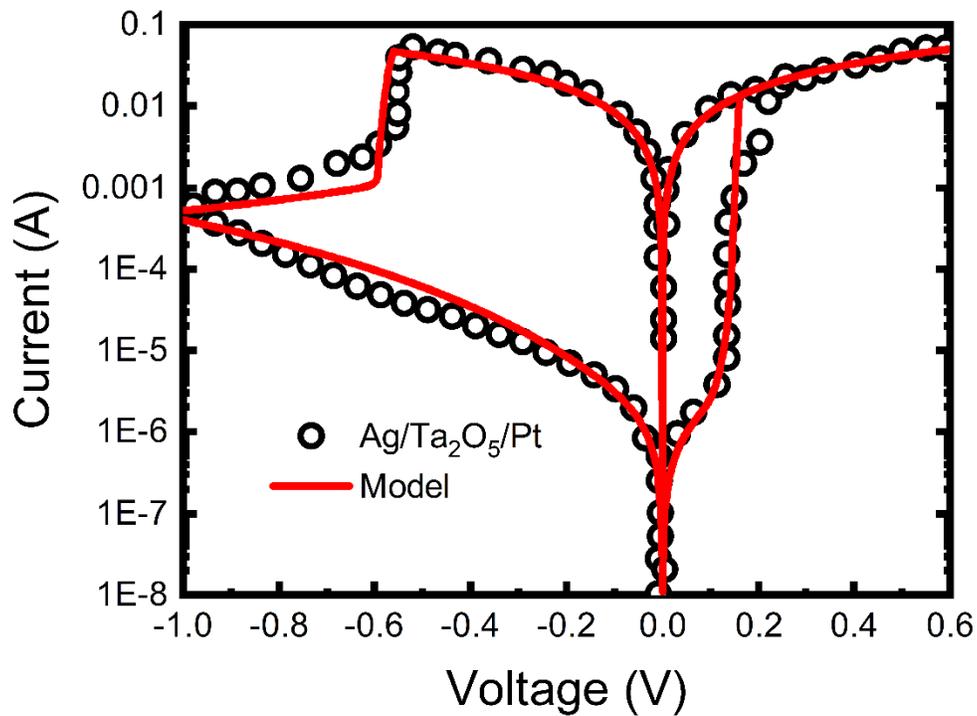


Fig. 7.10. Model validation of a Ag/Ta<sub>2</sub>O<sub>5</sub>/Pt RRAM which experienced both metal ion and oxygen vacancy switching [94]. The symbols are experimental data. The lines are

model simulations.

Users can add a noise source into the differential equations to simulate the variation in RRAM [95-97, 101]. Specifically, by adding a noise term to the equations Eq. (7.1) and Eq. (7.2), such that  $dL_H/dt = dL_H/dt + \text{Noise}$ , random fluctuations in the length of the filament are introduced. This approach allows for the generation of random noise in the filament growth dynamics. The simulation results are validated by fitting the model to experimental data from [88], as illustrated in Fig. 7.11. The inclusion of the noise source effectively captures the RRAM resistance variations, demonstrating the robustness and accuracy of our model in accounting for the real-world variability of RRAM devices. It shows that this model is compatible with the variation modeling method in [95-97, 101].

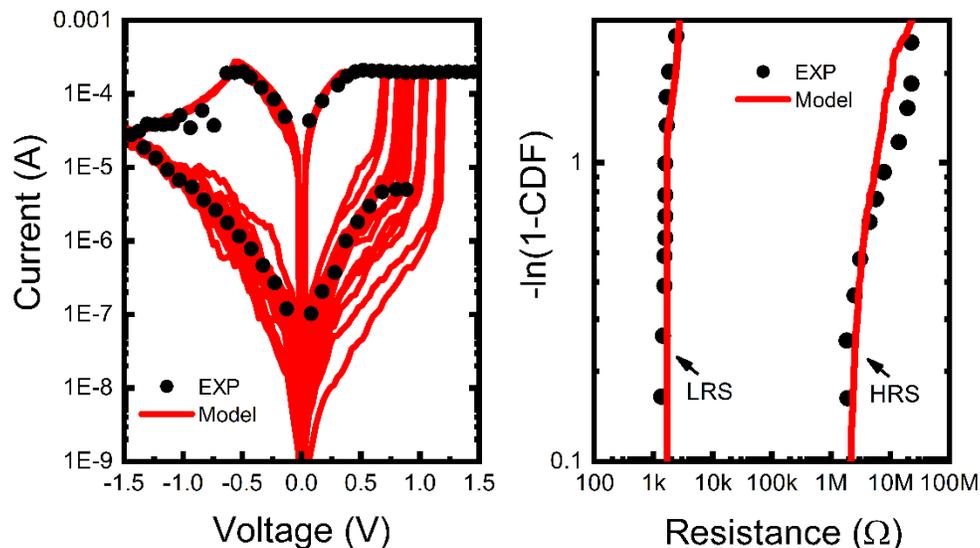


Fig. 7.11. Model validation of RRAM variation and disturbance [88].

Finally, I perform a 1T1R simulation of the calibrated RRAM model in Fig. 7.2 & 7.3 with a 7nm FinFET BSIM-CMG model [111]. A multilevel READ/WRITE operation of SET is presented in Fig. 7.12 showing the model can be integrated with the standard FET model for IC simulations.

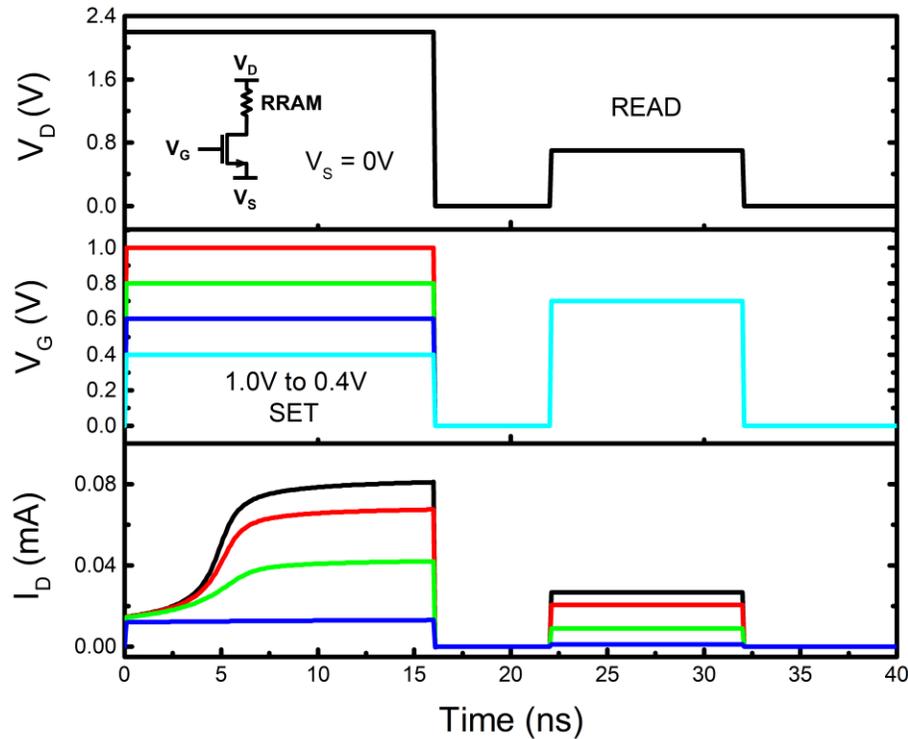


Fig. 7.12. 1T1R multilevel READ/WRITE simulation of a RRAM. This is 1T1R cells simulation for 40 ns with a maximum time step of 1 ps on an Intel Xeon Platinum 8260 processor. The simulation took 7.62 seconds.

## 7.4 Chapter Summary

I have developed a versatile, computationally efficient and robust compact model of RRAMs. It is shown to fit many published OxRAM and CBRAM devices using different material systems. The model applies to multilevel SET and RESET operations driven by voltage and current. It also contains the capability to do variation simulations and IC simulations with transistors and memory circuits.

## Chapter 8 A Compact Model of Perpendicular Spin-Transfer-Torque Magnetic Tunnel Junction

I present a new compact model of a perpendicular spin-transfer-torque (STT) magnetic tunnel junction (MTJ). Previous studies on STT-MTJs have either focused on solving the Landau–Lifshitz–Gilbert (LLG) equation or utilizing critical current-based macro models. However, the LLG approaches are too complex for large circuit simulations while the macro models fail to capture the underlying magnetization physics. In this work, I propose a semi-physical and computationally efficient compact model that accurately represents the time-dependent magnet moment and resistance. To validate this model, I compare it with various experimental data and LLG-based STT-MTJ model. The model demonstrates geometry dependence and temperature dependence. Furthermore, I develop a continuous switching probability model to effectively track the probabilities of states under arbitrary waveforms. This work is published by IEEE. © 2025 IEEE. Reprinted, with permission, from [112]

### 8.1 Motivation

Magnetic tunnel junctions (MTJs) are promising devices for non-volatile memory and in-memory computing, offering fast speed, high density, and low power consumption. An MTJ consists of a fixed magnetic layer and a free layer. In the case of spin-transfer-torque (STT)-MTJs, tunneling current is employed to switch the spin in the free layer between two states:  $0^\circ$  (parallel, P state) and  $180^\circ$  (anti-parallel, AP state) with respect to the fixed layer. This spin angle difference between the fixed and free layers alters the tunnel

magnetoresistance (TMR) [113-115]. Increasing the programming current results in faster switching speed [116, 117]. Additionally, the magnetic moment's thermal fluctuation introduces randomness in both the switching speed and probability [116-119]. To facilitate the simulation and design of large integrated circuits (ICs) incorporating STT-MTJs, there is a need for a compact model that is both computationally efficient and accurately captures all these phenomena.

Two types of SPICE models for STT-MTJs have been reported. The first type is the LLG-based model [120-124]. These models directly solve the full LLG equation, providing high accuracy in predicting magnetizations. However, they can be time-consuming for large integrated circuit (IC) simulations. The other type of model is the macro model [125-129]. These macro models are based on the concept of a critical current that varies with the duration of a current pulse. While macro models are faster in computation, they lack the ability to capture detailed information about the magnetic moment. Consequently, they become less accurate under arbitrary waveforms.

In this study, I propose a semi-physical compact model for perpendicular magnetic tunnel junctions (p-MTJs). Compared to in-plane MTJs, p-MTJs are recognized for their enhanced thermal stability and scalability [117]. This model solves the 1D magnetic moment and accurately fits the results of the LLG solver [120] under arbitrary inputs as well as experimental data. A magnetic-state probability model is also developed to continuously track the probability of P or AP states and estimate the write error rate (WER) without performing Monte Carlo simulations

## 8.2 Model

### Equivalent Circuit

Fig. 8.1 shows the equivalent circuit of the proposed model. We focus on the p-MTJ, where the magnetic moment in the fixed layer points out of the plane, and  $m_z$  is the z component of the magnetic moment in the free layer. A 1D LLG equation is used to calculate  $m_z$  and the device resistance ( $R_{\text{MTJ}}$ ). Instead of using random numbers to do stochastic simulations [119], we

calculate the switching probability as a separate output giving fast WER estimation without disturbing the tracking of the tunneling resistance.

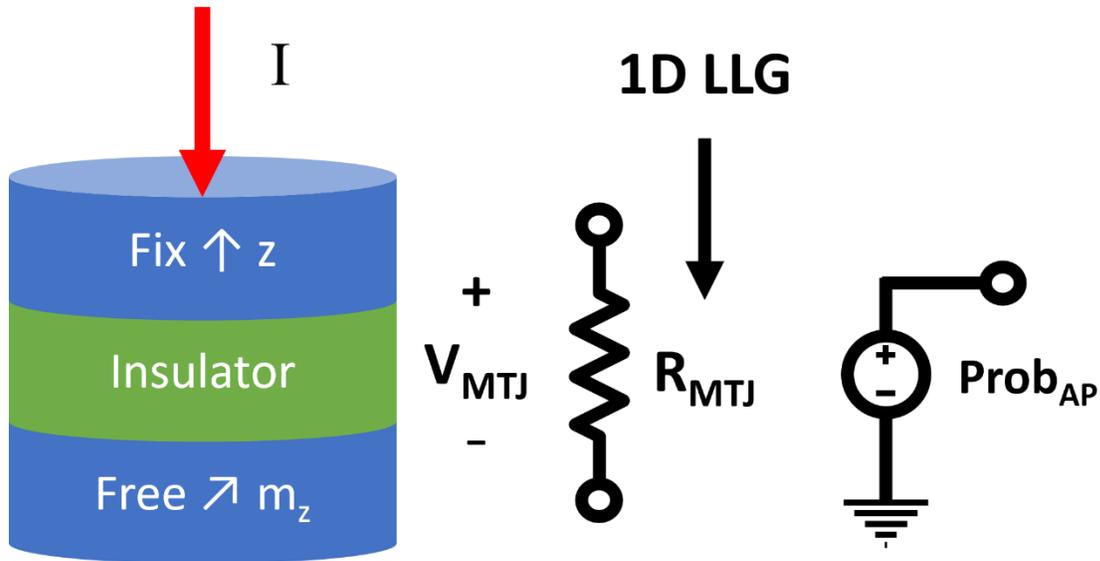


Fig. 8.1. Equivalent circuit of this model. The direction of the magnetic moment and the voltage polarity are shown. A separate node is used to output the state probability ( $\text{Prob}_{\text{AP}}$ ).

© 2025 IEEE

## Tunneling Resistance

The tunneling resistance model is shown in Eq. (8.1), where  $R_P$  is the resistance of the P state,  $\rho_P$  is the P state resistivity at  $V_{\text{MTJ}}=0$ ,  $t_{\text{ox}}$  is the tunneling layer thickness,  $\beta$  is a fitting parameter,  $A$  is the device area,  $V_1$ ,  $V_2$ ,  $n_1$ , and  $n_2$  are fitting parameters for voltage dependence. The temperature dependence is included in Eq. (8.1a) by considering the magnon model and the Fermi smearing model of  $R_P$  where  $\gamma_{1-3}$  are fitting parameters and  $E_C$  is the magnetic cutoff energy [130, 131]. TMR in Eq. (8.1e) is defined as  $(R_P - R_{\text{AP}})/R_P$ ,  $R_{\text{AP}}$  is the resistance at the AP state, and  $\text{TMR}_0$  is the TMR at  $V_{\text{MTJ}}=0$  which is a combination of magnon contribution ( $\text{TMR}_1$ ) [130, 131] and the spin polarization contribution ( $\text{TMR}_2$ ) [132].  $\gamma_{4-7}$  are the temperature fitting parameters for the spin polarization percentage,  $G_{\text{SI}}$  is the spin-independent component, and  $G_{\text{T}}$  is the prefactor for direct elastic tunneling [132].  $R_{\text{MTJ}}$  is the resistance of the MTJ such that  $I=V_{\text{MTJ}}/R_{\text{MTJ}}$ . Eq. (8.1a) considers the

thickness and area dependence of the MTJ resistance and it is verified with the data [133] shown in Fig. 8.2. The temperature dependence of  $R_p$ ,  $R_{AP}$  and TMR are verified in Fig. 8.3 [134].

$$R_{P0} = \frac{\rho_P t_{ox} \exp(\beta t_{ox}) \sin(\gamma_1 T)}{A \gamma_1 T} \left[ 1 + \gamma_2 k_B T \ln \left( \frac{k_B T}{E_C} \right) \right]^{-1} \quad (8.1a)$$

$$TMR_1 = \left( TMR_0' + 1 \right) \frac{1 + \gamma_2 k_B T \ln(k_B T / E_C)}{1 + \gamma_3 k_B T \ln(k_B T / E_C)} - 1 \quad (8.1b)$$

$$TMR_2 = \frac{2(1 - \gamma_4 T^{\gamma_5})(1 - \gamma_6 T^{\gamma_7}) TMR_0'}{TMR_0' [1 - (1 - \gamma_4 T^{\gamma_5})(1 - \gamma_6 T^{\gamma_7})] + (1 + TMR_0') \frac{G_{SI}}{G_T} + 2} \quad (8.1c)$$

$$R_P = \frac{R_{P0}}{1 + (V_{MTJ}/V_1)^{2n_1}} \quad (8.1d)$$

$$TMR = \frac{TMR_0 (= TMR_1 + TMR_2)}{1 + (V_{MTJ}/V_2)^{2n_2}} \quad (8.1e)$$

$$R_{MTJ} = \frac{2R_P(1 + TMR)}{2 + (1 + m_z)TMR} \quad (8.1f)$$

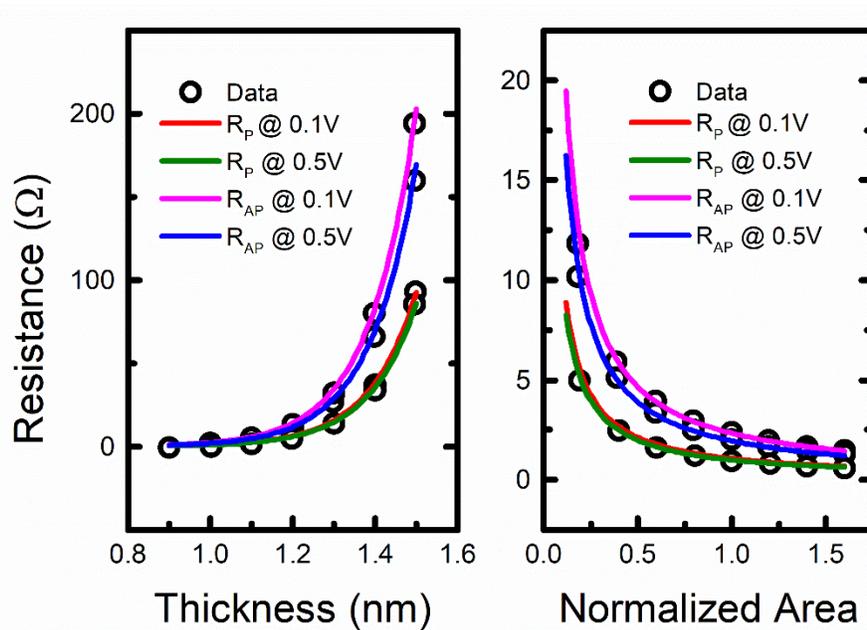


Fig. 8.2. Validation of the geometry dependence resistance model with the data (symbols) from [133]. The lines are the model results. © 2025 IEEE

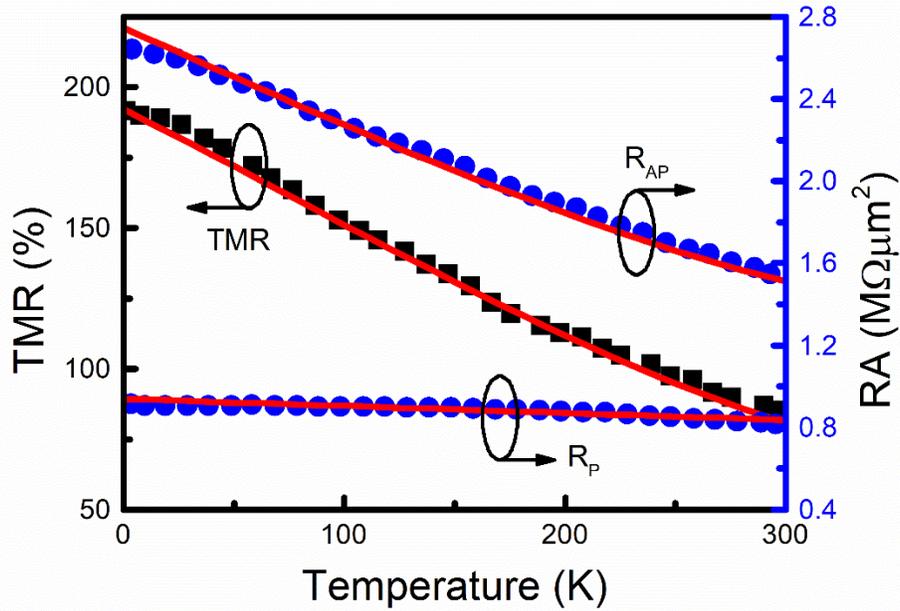


Fig. 8.3. Validation of the temperature dependence resistance model with the data (symbols) from [134]. The lines are the model results. © 2025 IEEE

## Magnetic Moment

The magnetic moment model is a modified 1D approximation of LLG [116]. Instead of assuming a constant current pulse to calculate the critical current ( $I_c$ ) [117], I solve the differential equation Eq. (8.1a) using the Verilog-A code.  $f(I)$  represents  $I/I_c$ . Since  $I_c$  could be different for  $P \rightarrow AP$  and  $AP \rightarrow P$  switching,  $I/I_c$  is smoothly connected with Eq. (8.2b) where  $I$  is the MTJ current,  $I_{c,AP}$  and  $I_{c,P}$  are the critical current for  $P \rightarrow AP$  for  $AP \rightarrow P$ , respectively, and  $\delta$  is a smoothing factor. The plus and minus signs depend on whether  $I_{c,AP}$  is larger or smaller than  $I_{c,P}$ .  $I_c$  is calculated by Eq. (8.2d) to (8.2g) where  $\alpha$  is the dumping factor,  $c_i$  is an empirical fitting parameter of critical currents,  $H_K$  is the anisotropy field,  $M_s$  is saturation magnetization,  $V_F$  is the free layer volume,  $\Delta$  is the thermal stability factor,  $t_0$  is the characteristic relaxation time, and  $k_B T$  is the thermal energy,  $\eta_i$  is the spin efficiency coefficient modeled by Eq. (8.2i) and (8.2j) which combine the temperature effects of spin polarization percentage Eq. (8.2k) and TMR [132, 135] with  $P'$  and  $PAP'$  being fitting parameters. Thermal-assisted switching is modeled with Eq.

(8.1c) [117, 118] where  $1/\tau_0$  is the attempt frequency and  $\tau_0=t_0/\sqrt{\Delta}$ . The temperature dependence of  $M_s$ ,  $H_K$  are found to be fitted with equations Eq. (8.2g) and (8.2h) [136, 137] where  $M_{s0}$ ,  $H_{K0}$  are their value at 0K, and  $M_{ST1}$ ,  $M_{ST2}$ ,  $H_{K1}$ ,  $H_{K2}$  are the fitting parameters.

$$\frac{dm_z}{dt} = \frac{1}{t_0} (f(I) - m_z)(m_z^2 - 1) \quad (8.2a)$$

$$f(I) = 0.5 \left[ \frac{I}{I_{c,AP}} + \frac{I}{I_{c,P}} \mp \sqrt{\left( \frac{I}{I_{c,AP}} - \frac{I}{I_{c,P}} \right)^2 + \delta^2 \pm \delta} \right] \quad (8.2b)$$

$$I_{c,i} = I_{c0,i} \left[ 1 - \frac{1}{\Delta} \ln \left( \frac{t}{\tau_0} \right) \right], i = P, AP \quad (8.2c)$$

$$I_{c0,i} = \frac{4qk_B T \alpha \Delta c_i}{\hbar \eta_i}, i = P, AP \quad (8.2d)$$

$$\Delta = \frac{H_K M_S V_F}{2k_B T} \quad (8.2e)$$

$$t_0 = \frac{1+\alpha^2}{1.76 \times 10^5 \alpha H_K} \quad (8.2f)$$

$$M_S = M_{S0} - M_{ST1} T^{M_{ST2}} \quad (8.2g)$$

$$H_K = H_{K0} - H_{KT1} T^{H_{KT2}} \quad (8.2h)$$

$$\eta_P = \left( \left[ -4 + \left( \frac{1}{\sqrt{P_t}} + \sqrt{P_t} \right)^3 \right]^{-1} + \frac{P_t}{2(1+P_t^2)} \right) \frac{\sqrt{TMR(TMR+2)}}{2(TMR+1)} \quad (8.2i)$$

$$\eta_{AP} = \left( \left[ -4 + 0.5 \left( \frac{1}{\sqrt{PAP_t}} + \sqrt{PAP_t} \right)^3 \right]^{-1} + \frac{PAP_t}{2(1-PAP_t^2)} \right) \frac{\sqrt{TMR(TMR+2)}}{2(TMR+1)} \quad (8.2j)$$

$$P_t = P'(1 - \gamma_4 T^{\gamma_5}), PAP_t = PAP'(1 - \gamma_6 T^{\gamma_7}) \quad (8.2k)$$

## State/Switching Probability

Due to the thermal fluctuation in the switching process and the initial magnetic angle, MTJ experiences probabilistic switching [116, 117, 119, 121].

Circuit designers need to track the switching probability of MTJs to analyze the WER. Eq. (8.3) calculates the probability of the P/AP state under any waveforms using a modified small-angle approximation of Fokker-Planck equation [116] and include both switching directions by using the function  $(2\text{Prob}_s-1)$  where  $W$  determines the switching probability through Eq. (8.3b) and  $a$  and  $b$  are integer fitting parameters to adjust the switching speed. The probability of AP state ( $\text{Prob}_{\text{AP}}$ ) is shown as Eq. (8.3c).

$$\frac{dW}{dt} = \frac{2}{t_0} \left[ (f(I) \times \begin{cases} 1, & \text{if initial state is P} \\ -1, & \text{if initial state is AP} \end{cases} + (2\text{Prob}_s - 1)^{2a+1})W + \frac{(1-2\text{Prob}_s)^{2a+1}}{\Delta} \right] \times [1 - (2\text{Prob}_s - 1)^{2b}] \quad (8.3a)$$

$$\text{Prob}_s = \exp\left(\frac{\pi^2}{4W}\right) \quad (8.3b)$$

$$\text{Prob}_{\text{AP}} = \begin{cases} \text{Prob}_s, & \text{if initial state is P} \\ 1 - \text{Prob}_s, & \text{if initial state is AP} \end{cases} \quad (8.3c)$$

### 8.3 Model Validation

This model is implemented in Verilog-A and an LLG-based STT-MTJ model [120] and tested with experimental data [119, 138]. I first test the model with an LLG model [120] by directly comparing  $m_z$ . Although the LLG model doesn't consider thermal effects, it can be insignificant since the tests are done in the nanosecond range. Fig. 8.4 shows the comparison. The model parameters  $M_s$ ,  $H_K$  and  $V_F$  are chosen to match the physical parameters in the 3D LLG simulations, and  $a$ ,  $c_i$ , and  $h_i$  are fitting parameters used to fit the model with LLG results. The parameter values are listed in the caption of Fig. 8.4. The model can capture the switching from P→AP and AP→P very well under different applied voltages ( $V_{\text{MTJ}}$ ). When the voltage increases, the current increases, which leads to a faster switching of  $m_z$ , which can also be seen in Fig. 8.4. It is also common that P→AP and AP→P switching have different current/voltage dependence that is modeled by Eq. (8.2b) with separate critical currents. The model is also predictive versus the physical device quantities. In Fig. 8.5, we present the comparison between the model and the LLG with different free layer thickness ( $T_{\text{FL}}$ ) and  $M_s$  where all the

other parameters are the same as the fitting in Fig. 8.4. It shows very good predictive ability since the equations are derived from physics.

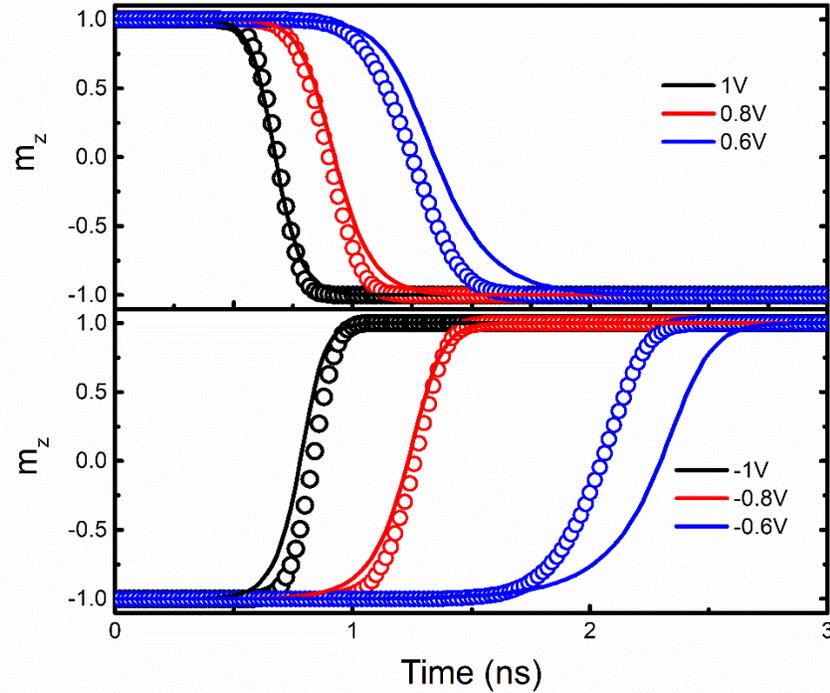


Fig. 8.4. Validation of this model (lines) versus the LLG model (symbols) [120] under constant write/erase voltage. The same set of parameters is used for Fig. 8.5. The LLG inputs are  $M_S=700 \text{ emu/cm}^3$ , uniaxial anisotropy= $56 \text{ k}_B T$ ,  $a=0.028$ ,  $\text{area}=25^2\pi \text{ nm}^2$ , and free-layer thickness= $1.4 \text{ nm}$ . In the model, we use  $M_S=700 \text{ emu/cm}^3$ ,  $H_K=3000 \text{ Oe}$ ,  $a=0.01$ ,  $c=0.2$ ,  $\eta_P=0.2275$ , and  $\eta_{AP}=0.214$  to fit the data. The lines are the model results. © 2025 IEEE

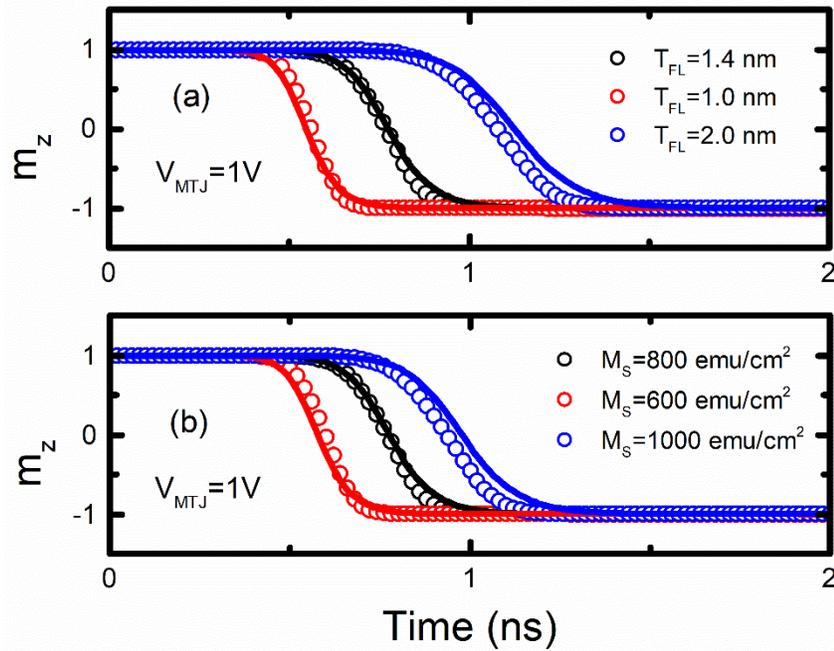


Fig. 8.5. Demonstration of the model predictive ability of physical parameters. The model is compared with the LLG where all the other parameters are the same as Fig. 8.4. Only free layer thickness ( $T_{FL}$ ) and  $M_s$  are changed for (a) and (b) respectively. © 2025 IEEE

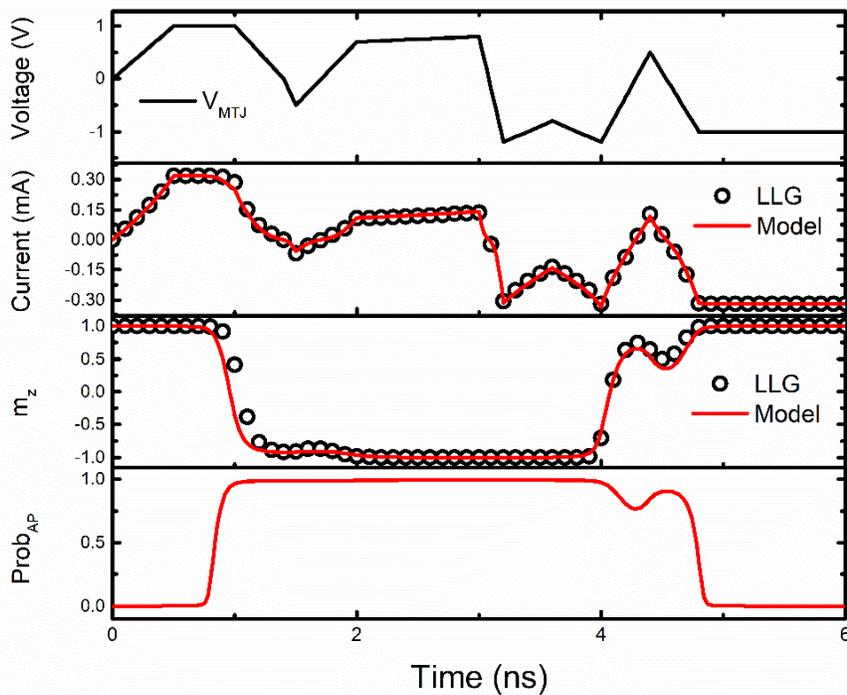


Fig. 8.6. This model works under an arbitrary waveform. The validation of the tunneling

current and  $m_z$  versus LLG is shown using the same parameters as Fig. 8.4. A probability of AP state is also calculated which switches continuously between 0 and 1 and responds to the varying of applied waveform. © 2025 IEEE

I further show that the model can work under arbitrary and complex input waveforms in Fig. 8.6. The model current and  $m_z$  show an excellent match with LLG. Moreover, this model can calculate the state probability ( $\text{Prob}_{\text{AP}}$ ) under an arbitrary waveform at the same time to provide circuit designers with the estimation of WER without stochastic simulations.

However, real devices may not be monodomain and have defects. I try to test this model with experimental data in many publications and obtained quite good fittings in all cases. Fig. 8.7 shows the fitting of the model pulse width-dependent critical current versus the experimental data under pulse measurement [138, 139]. For both data, when the pulse width is shorter than 10ns, the precessional switching dominates which is modeled by the differential equation Eq. (8.2a). As the pulse duration gets longer, the MTJ can be switched by thermal activation with smaller critical currents which is modeled by Eq. (8.2c). Moreover, I test the model with an experimental STT-MRAM read/write operation [140] shown in Fig. 8.8 with an excellent fit. The switching probability model is also verified with experimental data [119]. The result is shown in Fig. 8.9. To validate the temperature dependence model, I fit the model with a set of resistance-voltage (RV) characteristics measured at different temperatures [141] (Fig. 8.10).

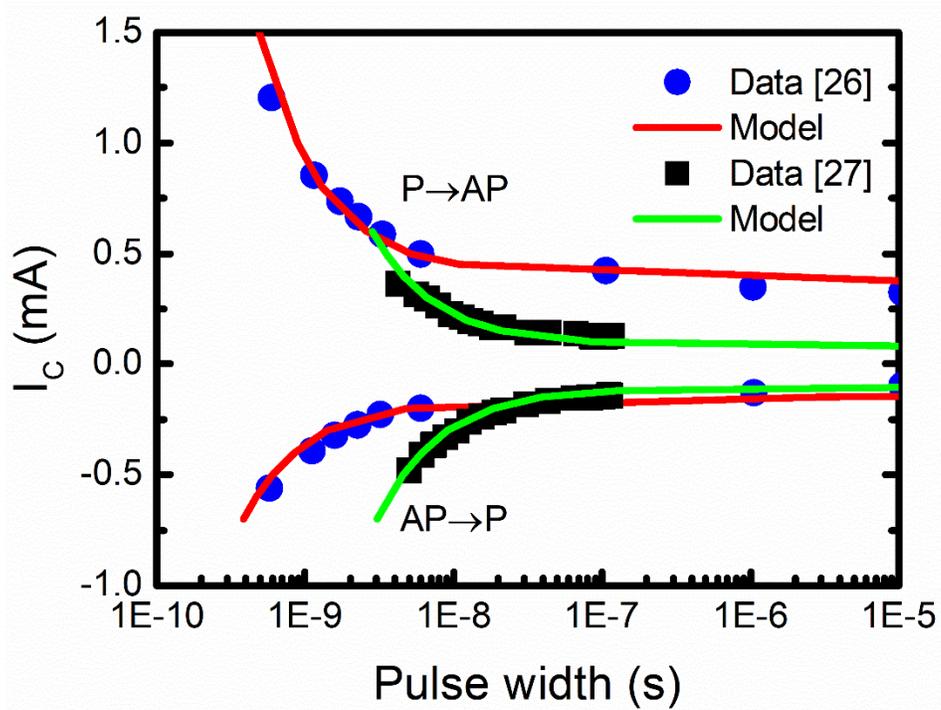


Fig. 8.7. Validation of the pulse width dependent critical current ( $I_c$ ). The lines are the model, and the symbols are the experimental data [138, 139]. Red:  $M_S=450$  emu/cm<sup>3</sup>,  $H_k=650$  Oe,  $V_F=1.17 \times 10^{-23}$  m<sup>-3</sup>,  $\alpha=1.5$ ,  $c=0.065$ ,  $\eta_P=0.183$ , and  $\eta_{AP}=0.465$ . Green:  $M_S=700$  emu/cm<sup>3</sup>,  $H_k=800$  Oe,  $V_F=4.245 \times 10^{-24}$  m<sup>-3</sup>,  $\alpha=0.03$ ,  $c=1$ ,  $\eta_P=0.183$ , and  $\eta_{AP}=0.143$  © 2025 IEEE

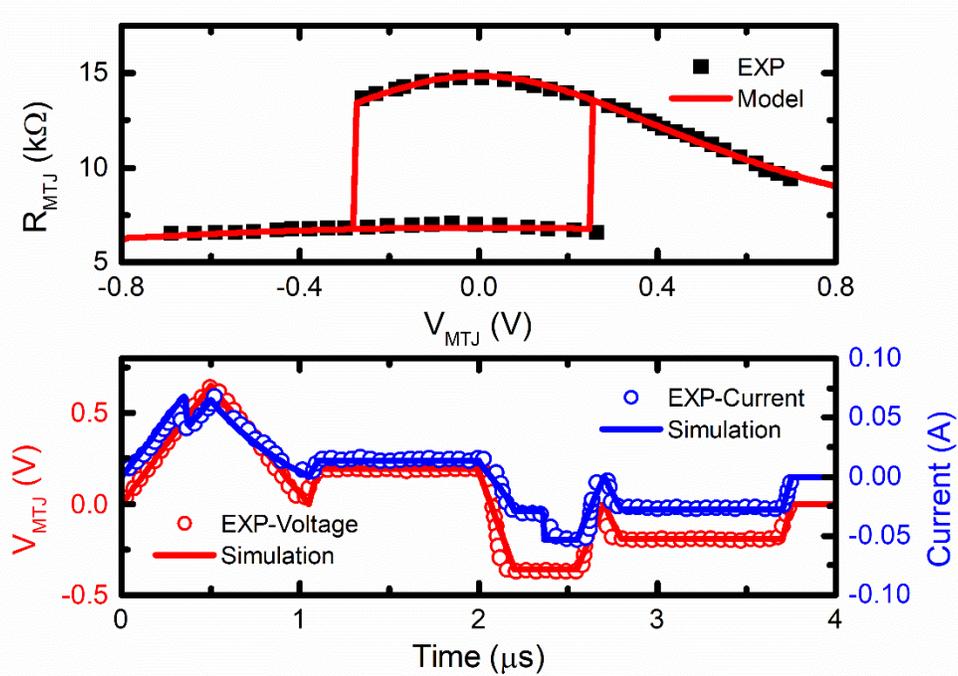


Fig. 8.8. Validation of the experimental RV and transient characteristics [140] with  $M_S=800$  emu/cm<sup>3</sup>,  $H_k=800$  Oe,  $V_F=1.25 \times 10^{-23}$  m<sup>-3</sup>,  $a=0.012$ ,  $c_i=0.3$ ,  $\eta_P=0.17$ , and  $\eta_{AP}=0.316$ . © 2025 IEEE

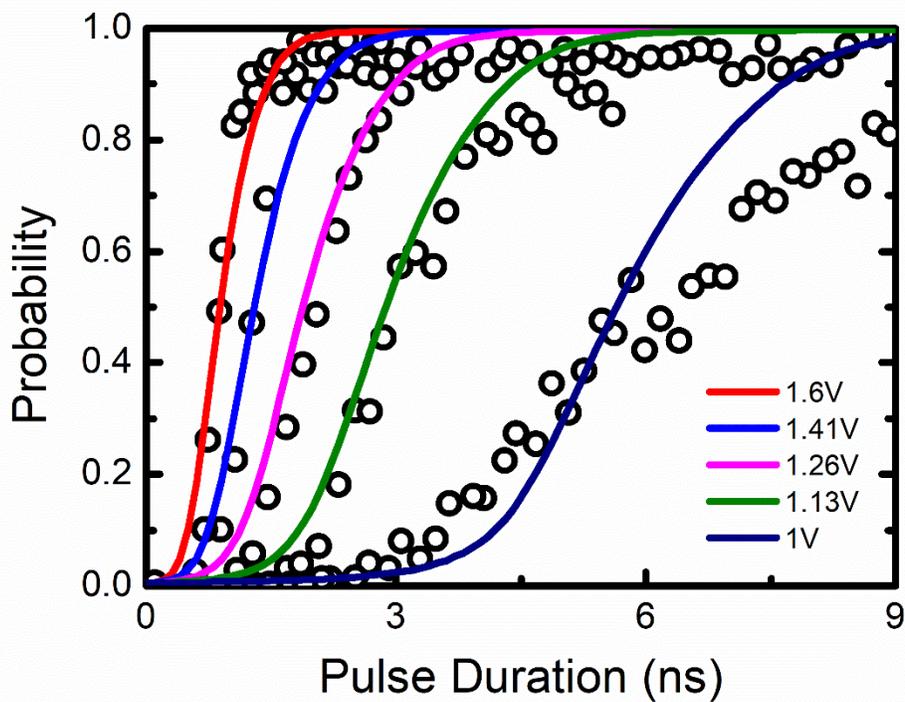


Fig. 8.9. The validation of the switching probability of the model (lines) versus

experimental data (symbols) [119] with  $M_S=796 \text{ emu/cm}^3$ ,  $H_K=2000 \text{ Oe}$ ,  $V_F=1.625 \times 10^{-24} \text{ m}^{-3}$ ,  $\alpha=0.014$ ,  $c=0.6$ , and  $\eta_p=0.183$ . © 2025 IEEE

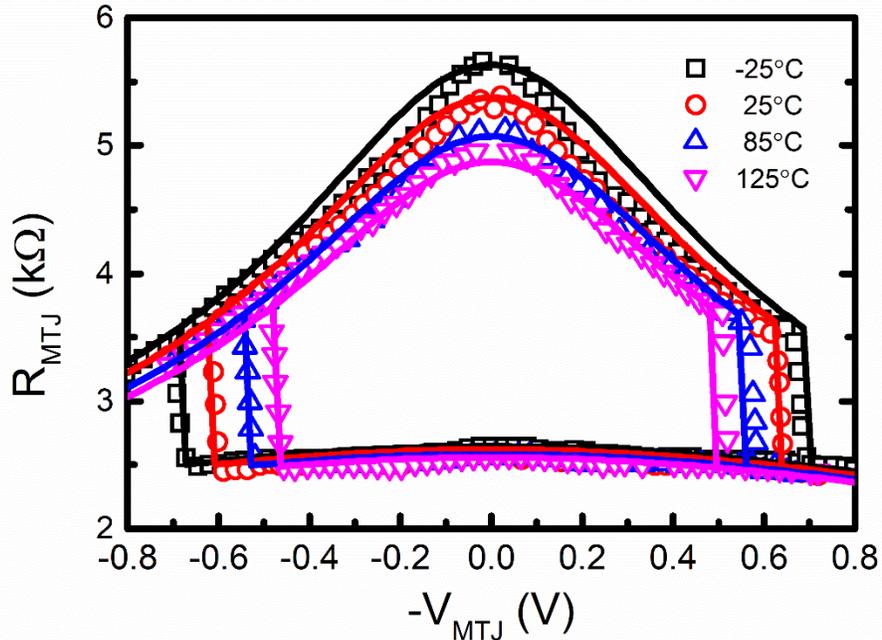


Fig. 8.10. The validation of the temperature dependence model. The data comes from [141]. © 2025 IEEE

To extract the model parameters, the users can start from extracting the resistance parameters with the RV curves. After that, the physical parameters such as  $M_S$ ,  $H_K$ , and  $V_F$  can directly come from the measured value, and  $\alpha$ ,  $c_i$ , and  $h_i$  ( $P'$ ,  $PAP'$ ) can be used to adjust the critical currents. The intermediate regime of the  $I_c$  versus pulse width plot can be tuned with  $a$ . The model users need to iteratively tune the parameters till fitting is satisfactory as commonly done in compact modeling.

Finally, I compared the model simulation speed with the LLG model [120]. The test was done using Intel Xeon Platinum 8260 with pulses simulation over 40ms and the maximum time step is 10ps on Hspice. The probability module is not included since the LLG model does not have it, but all the temperature dependences are preserved. The simulation times are 20.83s and 52.04s for this model and the LLG model, respectively. This model is 2.5 times faster than the LLG model. The speed difference could be even

more if the LLG model includes thermal effect and temperature dependence. It indicates that the proposed model can accurately model the MTJ physics and be computationally efficient at the same time.

## **8.4 Chapter summary**

I report a computationally efficient compact model of STT-MTJ which tracks the magnetization and probability of states continuously under arbitrary input waveforms. The model accuracy is validated with LLG simulation. The model allows the circuit designers to track the magnetic moment and thermal WER while designing the read/write circuits without Monte Carlo simulations. This model is implemented with an emphasis on robust convergence in Verilog-A so that it can be used efficiently with all popular SPICE circuit simulators with transistors and other circuit elements for IC design.

## **Chapter 9 BSIM-NN: Neural Network-Based BSIM Transistor Model Framework for Advanced and Emerging Technology**

I present a neural network (NN)-based transistor modeling framework BSIM-NN which includes drain, source, and gate currents and charges, and their variabilities. The training data is generated by a Berkeley Short-channel IGFET Model (BSIM) with ranges of channel lengths, widths, and oxide thicknesses. The NNs are trained to learn geometry dependence. The drain, source, and gate currents are modeled with one NN, and the charges by another NN. The NNs are trained to produce accurate variability prediction and derivatives of currents and charges. Quality and robustness tests such as Gummel symmetry, harmonic balance, and ring oscillator are performed and show excellent results. This NN modeling framework is not only useful for more Moore technologies (e.g., gate-all-around field-effect-transistor (GAAFET)) but also beyond Moore transistors (e.g., negative capacitance FET (NCFET)). This work is published by IEEE. © 2025 IEEE. Reprinted, with permission, from [142], [143], [144]

### **9.1 Motivation**

Transistor models are important to the semiconductor industry which needs fast and accurate models for circuit simulation and design optimization. Industry-standard compact models, such as the Berkeley Short-channel IGFET Model (BSIM) series of models [58, 59], use physics-based equations. Developing accurate and computationally efficient analytic equations for every complex transistor behavior, such as short channel effects and quantum effects in gate-all-around transistors [145], can be time-consuming.

Neural network (NN)-based compact models [146-148] hold the potential

to reduce the time of developing models of future new devices. The matrix multiplication nature and the ease of GPU acceleration endow NN-based compact models with the potential to reduce the time needed for model calculation during circuit simulation. Several previous works have studied using NN to model the variation in transistors. Ref. [149, 150] uses NN to predict some key merits in process variation such as  $I_{ON}$ ,  $I_{OFF}$ , and  $V_{TH}$  without modeling the entire IV characteristics. Ref. [151] uses process variation parameters as inputs to train an NN that can reproduce IV characteristics line tunnel FETs. Still, much more investigation is needed to determine whether NN-based models meet all the requirements of a practical compact model.

In this work, I present an NN-based compact model BSIM-NN of source, drain, gate currents, and charges, with variability modeling; and demonstrate its robustness for circuit simulation. Gate and drain leakage currents are included in the neural networks. Those leakages are important in evaluating circuit performance. Furthermore, source and drain charges are also included which are essential and contribute to the transient currents at source/drain terminals, which is also what BSIM does to ensure charge conservation. All currents are included in one network and all charges are included in the other network. The improved loss functions are developed to accurately train the networks with these additional outputs by considering higher order derivatives. I focus on four process variation parameters: gate length ( $L$ ), fin height ( $H_{FIN}$ ), equivalent oxide thickness (EOT) and work function difference ( $\Delta\phi$ ) of FinFETs. It is shown that the demonstrated NN model can fit the IV and CV characteristics in the presence of process variations to the higher-order derivatives of current and charge. This model can also predict the statistical distribution of the device merits when used in Monte Carlo simulations. The power of the NN model for advanced and emerging devices such as GAAFETs and NCFETs [152] is also demonstrated.

## 9.2 Modeling Framework

### Currents

To have an NN trained with the process variations of  $L$ ,  $H_{FIN}$ , EOT, and  $\Delta\phi$ , I include these parameters as inputs in NN together with  $V_{GS}$  and  $V_{DS}$ .

Fortunately, for  $\Delta\phi$ , we know from physics that the effect of  $\Delta\phi$  is equivalent to a gate voltage shift. Therefore,  $\Delta\phi$  does not need to be included among the inputs, rather it is treated as a gate voltage shift during training and inferencing as shown in Eq. (9.1). In this way, we can reduce the complexity of the NN and training time. Other variations such as fin thickness ( $T_{FIN}$ ) and temperature ( $T$ ) are not included in this work for simplicity and will be added in the next chapter. The IV NN is trained to model drain current ( $I_D$ ) and gate current ( $I_G$ ). Again, for simplicity in this study, I assume a floating-body device so the  $I_B=0$  which is mostly true for advanced transistors such as GAA. There are 3 outputs for the IV NN. The first output  $y_1$  is the transform of  $I_D$  as shown in Eq. (9.2) [153]. For  $I_G$ , we cannot easily determine its sign as  $I_D$ . To make  $I_G$  scaled by the  $\ln$  function, I separate it into positive and negative parts using the smoothing functions as Eq. (9.3) and transform them into  $y_{2p}$  and  $y_{2n}$  where  $\Delta$  is the smoothing factor and  $I_0$  is used to prevent 0. Thus, there are three outputs of this NN. The loss function is shown in Eq. (9.4), where RMS is the root-mean-square error function,  $g_m$  is the transconductance,  $g_m'$  is its derivative,  $g_{ds}$  is the output conductance,  $g_{ds}'$  is its derivative, and a to f are the coefficients for each loss. I include up to second-order derivatives to obtain the desired accuracy. To determine the weights in the loss function, I first train the NN with  $a=1$  and the others to be 0. Then, we increase the other weights one by one for each training iteration. When choosing the value of the weight, we keep the magnitude of the new loss comparable to the previous loss. Some fine-tuning of the coefficients is conducted till the accuracy meets satisfaction.

$$x = [V_{GS} - \Delta\phi, V_{DS}, H_{FIN}, L, EOT, (T_{FIN}, T...)] \quad (9.1)$$

$$I_D = V_{DS}e^{y_1}, \quad y_1 = \ln\left(\frac{I_D}{V_{DS}}\right) \quad (9.2)$$

$$y_{2p} = \ln\left(\frac{I_G}{2} + \frac{\sqrt{I_G^2 + \Delta^2}}{2} + I_0\right),$$

$$y_{2n} = \ln\left(\frac{-I_G}{2} + \frac{\sqrt{I_G^2 + \Delta^2}}{2} + I_0\right) \quad (9.3)$$

$$loss = a \cdot RMS(y_1) + b \cdot RMS(g_m) + c \cdot RMS(g_{ds}) + d \cdot RMS(g_m')$$

$$+e \cdot RMS(g_{ds}') + f \cdot RMS(y_{2p}) + f \cdot RMS(y_{2n}) \quad (9.4)$$

## Charges

For QV NN, I train  $Q_G$ ,  $Q_S$ , and  $Q_D$  all in one network and  $Q_B$  is  $-(Q_G+Q_S+Q_D)$ . Inputs are the same as Eq. (9.1) and outputs are shown in Eq. (9.5). In the loss function Eq. (9.6), I also include up to second-order derivatives to obtain good accuracy where  $a'$  to  $e'$  are the coefficients. The  $y$  in Eq. (9.6) represents  $y_{1,2,3}$  in Eq. (9.5).

$$y_{1,2,3} = Q_{G,S,D} \quad (9.5)$$

$$\begin{aligned} loss = & a' \cdot RMS(y) + b' \cdot RMS\left(\frac{\partial y}{\partial V_{GS}}\right) + c' \cdot RMS\left(\frac{\partial y}{\partial V_{DS}}\right) \\ & + d' \cdot RMS\left(\frac{\partial^2 y}{\partial V_{GS}^2}\right) + e' \cdot RMS\left(\frac{\partial^2 y}{\partial V_{DS}^2}\right) \quad (9.6) \end{aligned}$$

## Noises

Noise is important in analog and RF circuit designs. Instead of training models for noise or using complex physics-based equations to calculate noises, I use simple empirical equations to keep the model efficient. The channel thermal noise ( $S_{th}$ ) is proportional to  $g_{ds}$  at the linear region, and  $g_m$  at the saturation region [154]. I use an empirical equation Eq. (9.7) to sum the model's  $g_m$  and  $g_{ds}$  to calculate  $S_{TH}$  where  $a_{1-7}$  are fitting parameters. Flicker noise ( $S_{fn}$ ) or  $1/f$  noise is empirically modeled by Eq. (9.8) where  $b_{1-7}$  are fitting parameters. The  $S_{fn}$  model in Eq. (9.8) can adjust the subthreshold and inversion region noise separately. The gate shot noise is modeled by  $\bar{i}_g^2 = 2qI_G$ .

$$S_{th} = 4kT \left[ a_1 g_m + \frac{a_2 g_m}{1 + e^{(V_{DS} - V_{DSAT})/a_3}} + a_4 g_{ds} (1 + a_5 e^{\frac{-|V_{DS}|}{a_6 V_{GS}^{a_7}}}) \right] \quad (9.7)$$

$$S_{fn} = [(b_1 |I_D|^{b_2})^{-1} + (b_3 |I_D|^{b_4 (1 + b_5 e^{-|V_{DS}|/b_6})})^{-1}]^{-1} / f^{b_7} \quad (9.8)$$

## SPICE Implementation

The trained models are automatically coded into Verilog-A with a Python

code that we developed. The weights, biases, and matrix multiplications are coded into direct multiplications element-by-element without using any array and loop in Verilog-A as shown in Fig. 9.1. This implementation style increases the SPICE simulation speed since the array and loop are inefficient in Verilog-A. Three functions are examined in this work: sigmoid, tanh, and ISRU ( $x/\sqrt{1+x^2}$ ). Among the three activation functions, ISRU performs the best due to its simpler form and the lack of using an exponential function as shown.

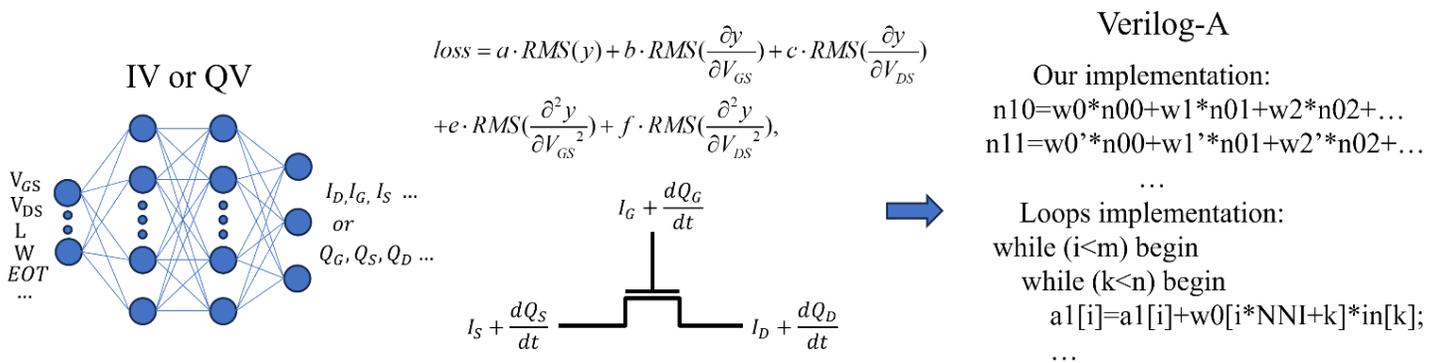


Fig. 9.1. The schematic diagram of the NN FET model, the loss function, and the Verilog-A implementation. The compact model has both IV and QV networks. The loss function includes up to second derivatives of outputs ( $y$ ) and it can be adjusted by tuning the weights (a-f) of each term. Our implementation uses direct multiplication instead of loops. © 2025 IEEE

## 9.3 Training & Inference Results

BSIM-NN is implemented with the Tensorflow package in Python. The training data is generated using a BSIM-CMG [58] model that is calibrated to the Intel 10nm-node FinFET [155] with 10 fins and  $L$  is 18nm and  $H_{\text{FIN}}$  is 46nm. I use the scaling capability of BSIM-CMG to generate data for  $L=[14, 16, 18, 20, 22, 24]$ nm,  $H_{\text{FIN}}=[38, 42, 46, 50, 54]$ nm, and  $EOT=[0.68, 0.73, 0.78, 0.83, 0.88]$ nm for training the NN to cover the range of possible device variations. The training uses 150 devices. Each device has a full  $I_D$  &  $I_G$

characteristic with  $V_{GS}$  and  $V_{DS}$  varying from  $-0.8V$  to  $0.8V$ . The importance of training the full bias spectrum is discussed in [153].

The training results are shown in Fig. 9.2 & 9.3 where I show the  $I_D$  fitting for several  $L$ ,  $H_{FIN}$ , and EOT combinations including some that are not in the training set. Both  $I_D V_G$  and  $I_D V_D$  are accurate to higher-order derivatives, especially for  $g_{ds}$  in the saturation region which is known to be difficult to model. Fig. 9.4 shows the NN modeling results of  $I_G$  for several structures and bias conditions. We can see that this modeling framework models all currents well by just using one network.

For QV NN, the same inputs are fed in with three outputs  $Q_G$ ,  $Q_S$ , and  $Q_D$ . In Fig. 9.5 & 9.6, I show the CV fitting accuracy concluding data that are not in the training data set. The NN model can fit the capacitances well for varying geometries by using just one network.

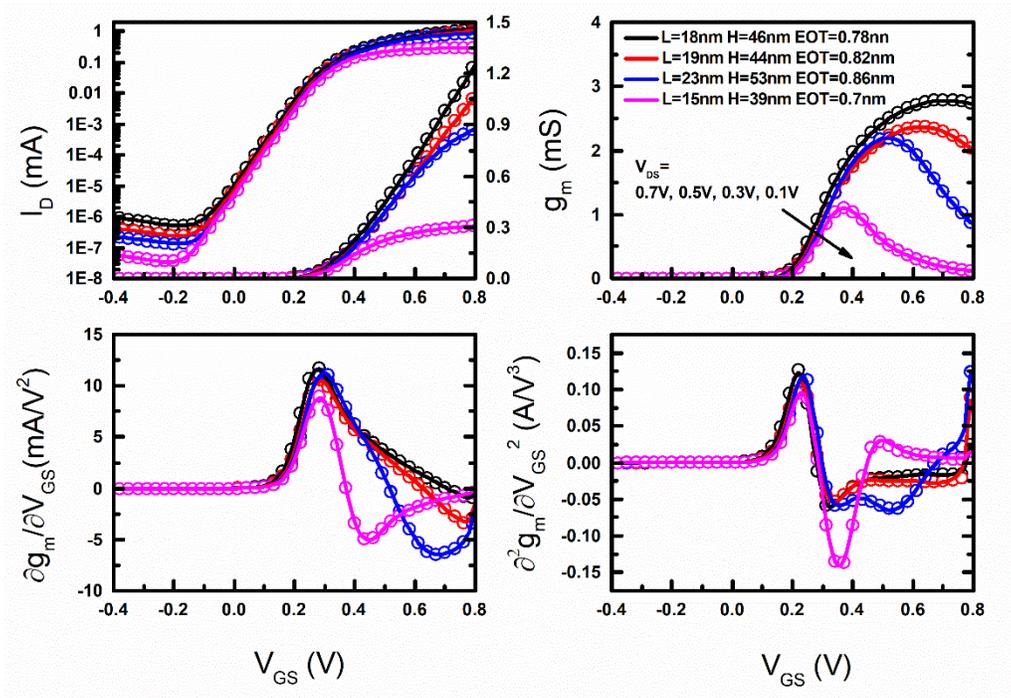


Fig. 9.2. The fitted  $I_D V_G$  curves at different  $L$ ,  $H_{FIN}$ , and EOT where the lines are the NN and symbols are the BSIM-CMG data. © 2025 IEEE

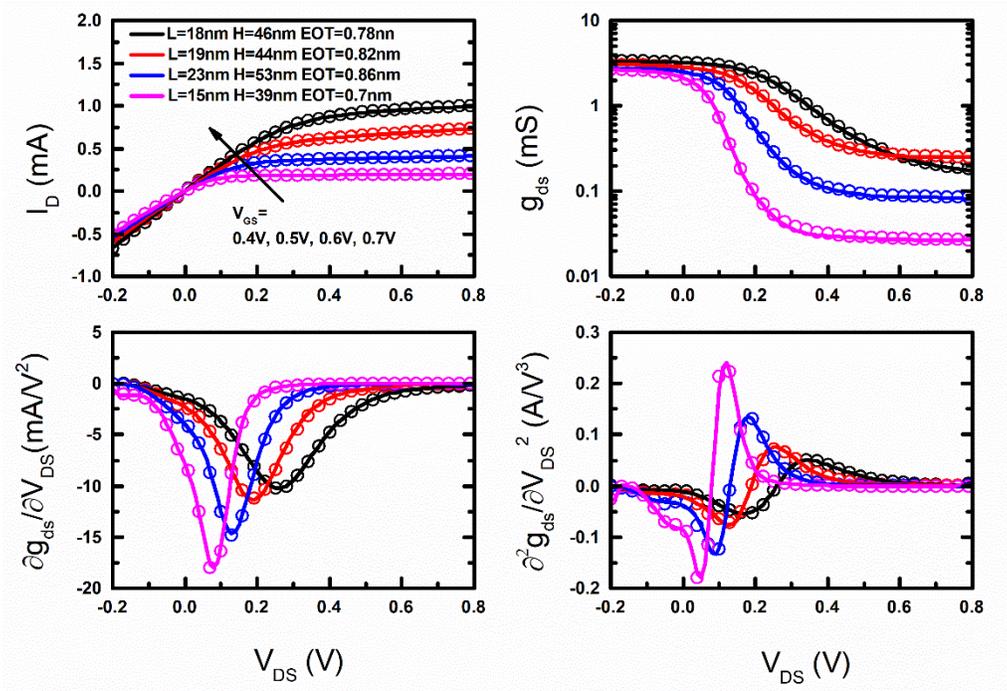


Fig. 9.3. The fitted  $I_D V_D$  curves at different  $L$ ,  $H_{FIN}$ , and  $EOT$  where the lines are the NN and symbols are the BSIM-CMG data. © 2025 IEEE

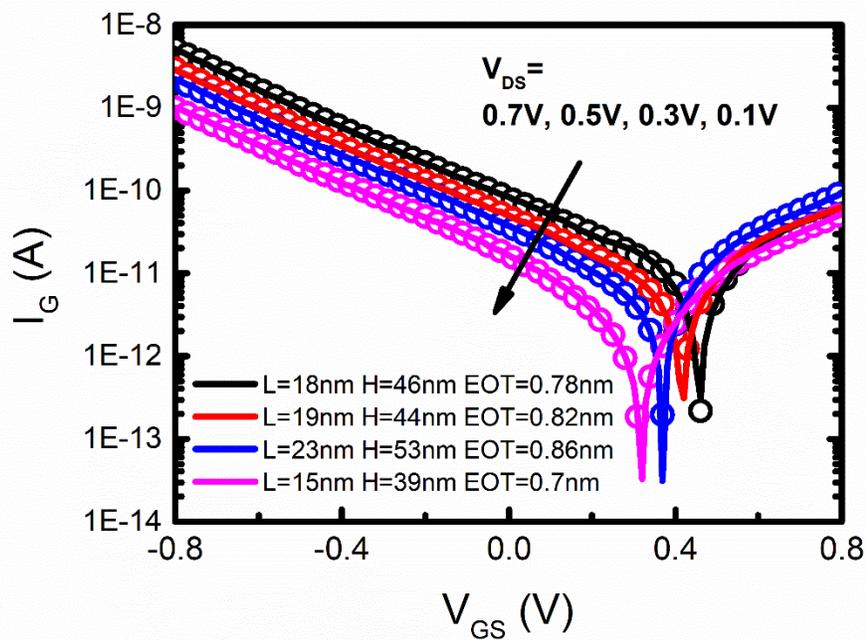


Fig. 9.4. The fitted  $I_G$  characteristics for different structures where the lines are the NN and symbols are the BSIM-CMG data. © 2025 IEEE

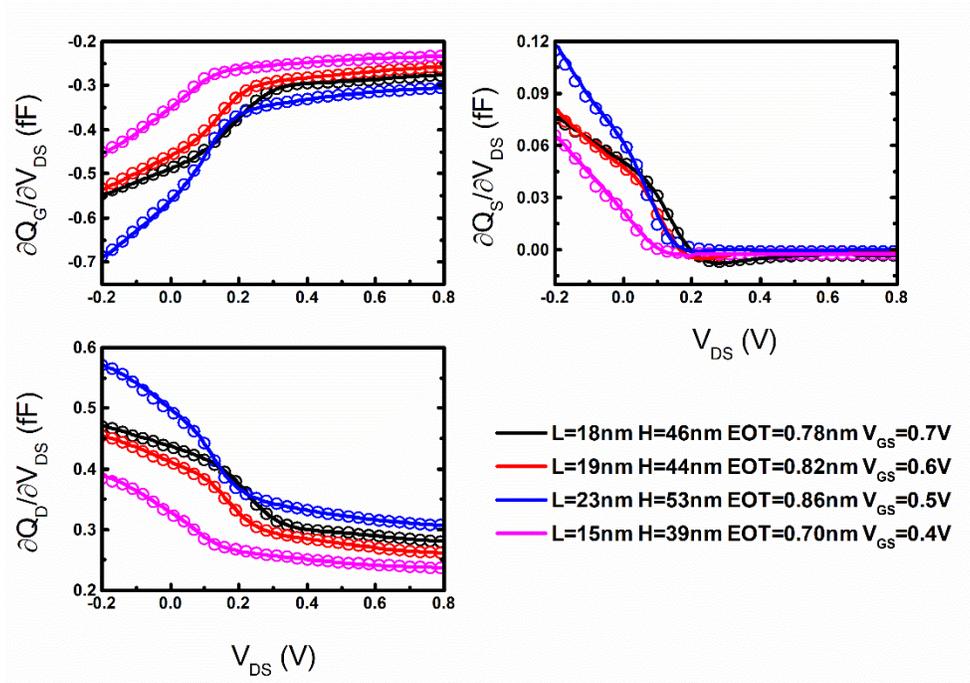


Fig. 9.5. CV fitting of  $Q_G$ ,  $Q_S$  and  $Q_D$  versus  $V_{DS}$  for different  $L$ ,  $H_{FIN}$  and EOT. © 2025 IEEE

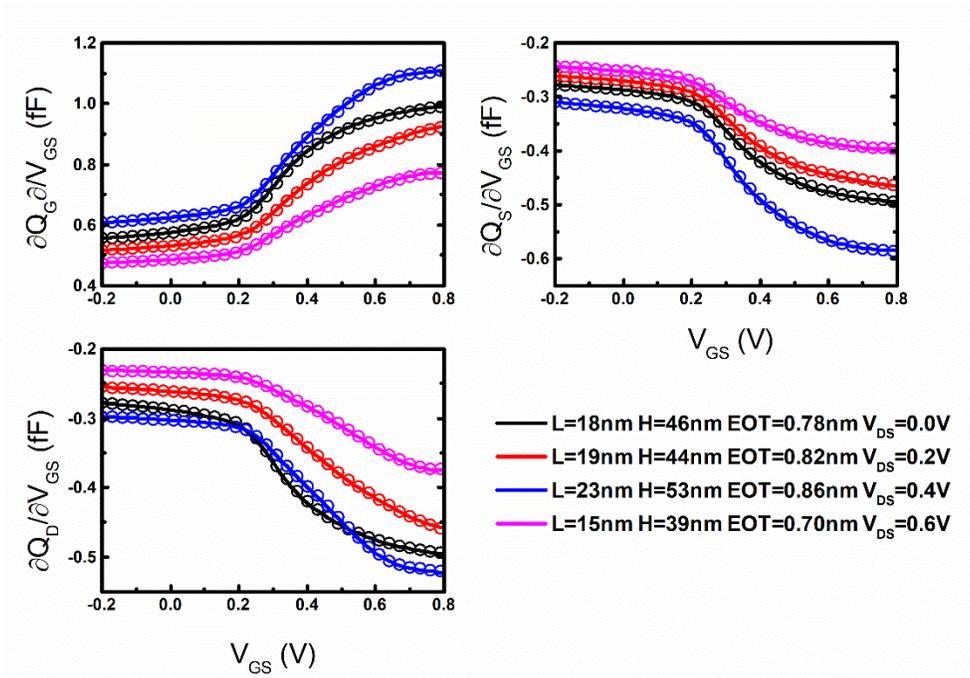


Fig. 9.6. CV fitting of  $Q_G$ ,  $Q_S$  and  $Q_D$  versus  $V_{GS}$  for different  $L$ ,  $H_{FIN}$  and EOT. © 2025 IEEE

To show the model's capability of variability modeling, I use Monte Carlo simulation to generate 2000 devices having certain variations ( $\sigma$ ) in  $L$ ,  $H_{\text{FIN}}$ ,  $EOT$ , and  $\Delta\phi$  with BSIM-CMG. Then, I extract the  $I_{\text{ON}}$  of these devices and compare them with the prediction of the NN model. Fig. 9.7 shows the  $I_{\text{ON}}$  distributions relative to the mean value. We can see that, the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of  $I_{\text{ON}}$  predicted by the NN model are in excellent agreement with those predicted by BSIM.

I also performed quality tests on the NN models. Fig. 9.8a shows the Gummel symmetry plot at 4<sup>th</sup> derivative is continuous and smooth. Because our model framework directly trains the NN using data from negative  $V_{\text{DS}}$  to positive  $V_{\text{DS}}$ , it can easily pass the Gummel test without applying smoothing functions like [146]. Therefore, this framework is applicable to devices that are inherently unsymmetric such as a MOSFET with different source and drain doping profiles. Fig. 9.8b shows the harmonic balance test result. The NN model produces the correct slope of each harmonic component.

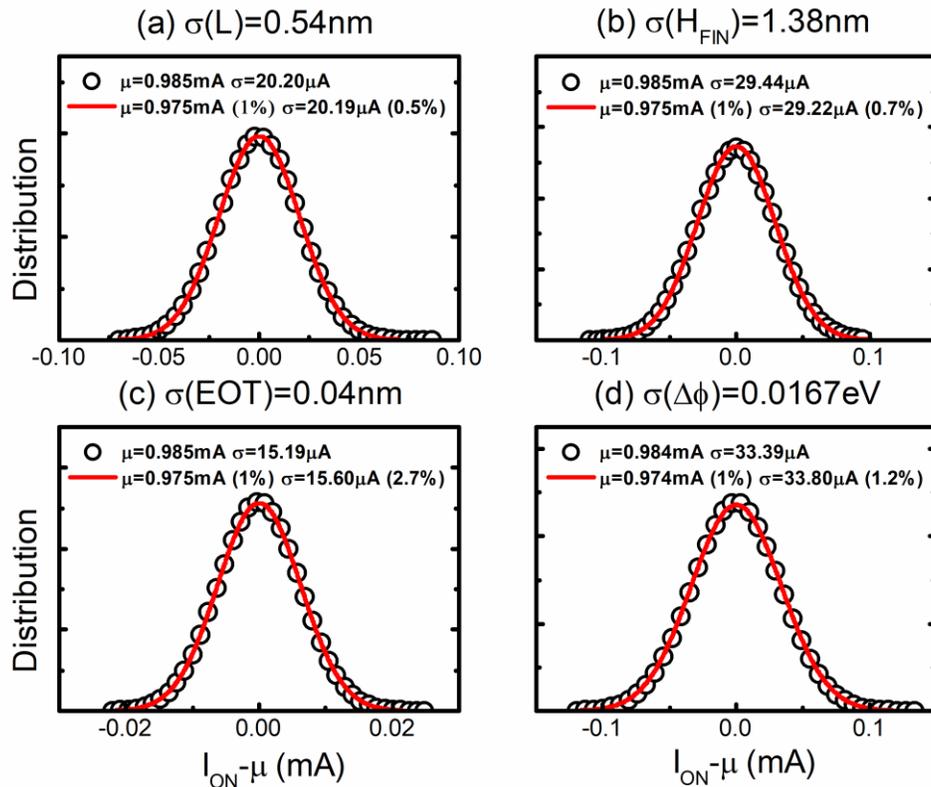


Fig. 9.7.  $I_{\text{ON}}$  distribution relative to mean by Monte Carlo simulation for (a)  $\sigma$  of  $L = 0.54 \text{ nm}$ , (b)  $\sigma$  of  $H_{\text{FIN}} = 1.38 \text{ nm}$ , (c)  $\sigma$  of  $EOT = 0.04 \text{ nm}$ , and (d)  $\sigma$  of  $\Delta\phi = 0.0167 \text{ eV}$ . The symbols are the data generated with BSIM-CMG and the lines are the predictions of the

NN model. In the parentheses, we show the error rate of  $\mu$  and  $\sigma$  between NN and BSIM-CMG. © 2025 IEEE

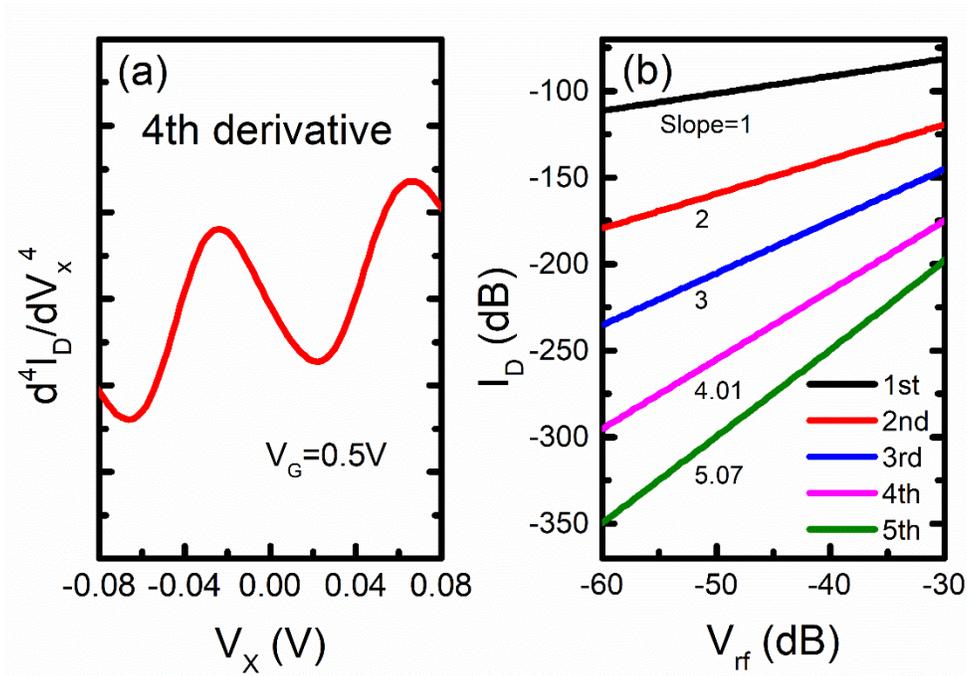


Fig. 9.8. (a) Gummel test at 4<sup>th</sup> derivative. (b) Harmonic balance test. The slope of each line meets the theoretical prediction. © 2025 IEEE

To test the noise model, the calculated thermal and flicker noises from BSIM-NN and BSIM-CMG are shown in Fig. 9.9 & 9.10. Simple empirical models can capture the noise characteristics very well. Then, the small signal noise of a common-source amplifier is tested (Fig. 9.11), and the model accurately captures the transition from flicker to thermal noise along the frequency. The phase noise of a 3-stage ring oscillator (RO) is also simulated and BSIM-NN produces the same noise spectrum as BSIM-CMG (Fig. 9.11).

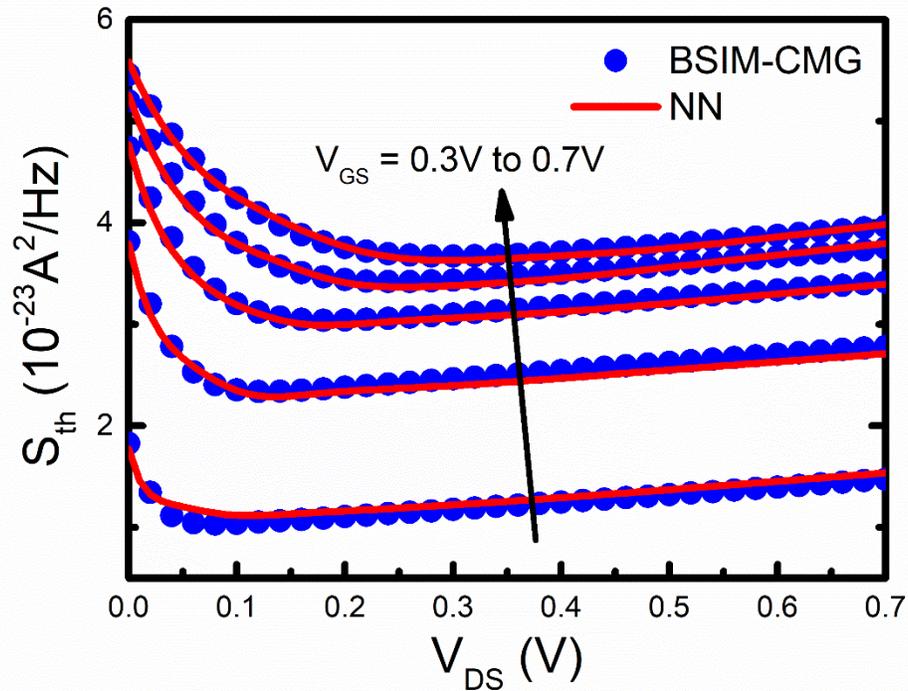


Fig. 9.9. Thermal noise comparison between BSIM-NN and BSIM-CMG.

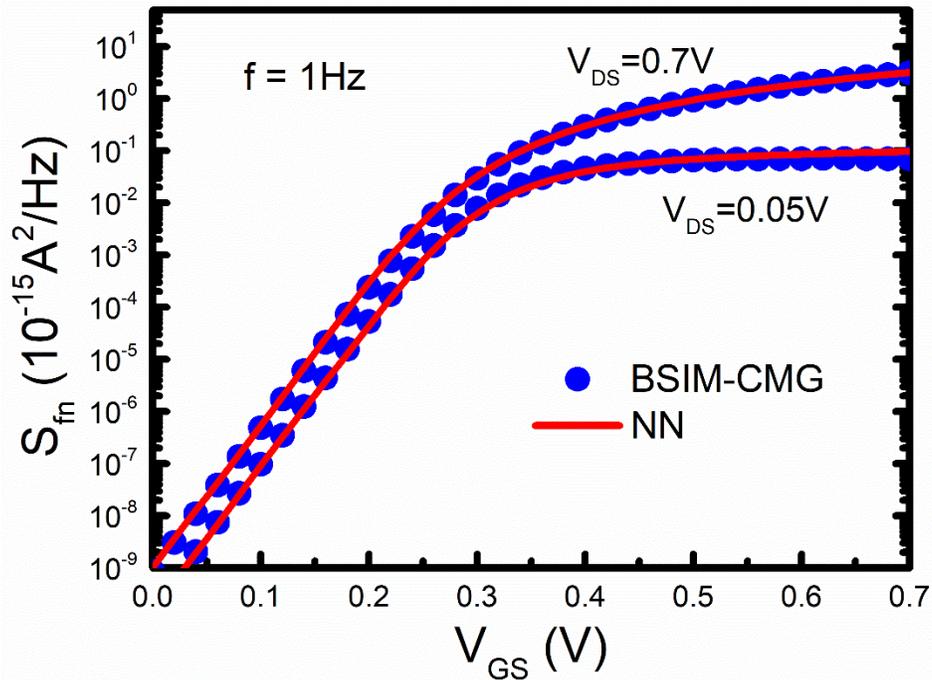


Fig. 9.10. Flicker noise comparison between BSIM-NN and BSIM-CMG.

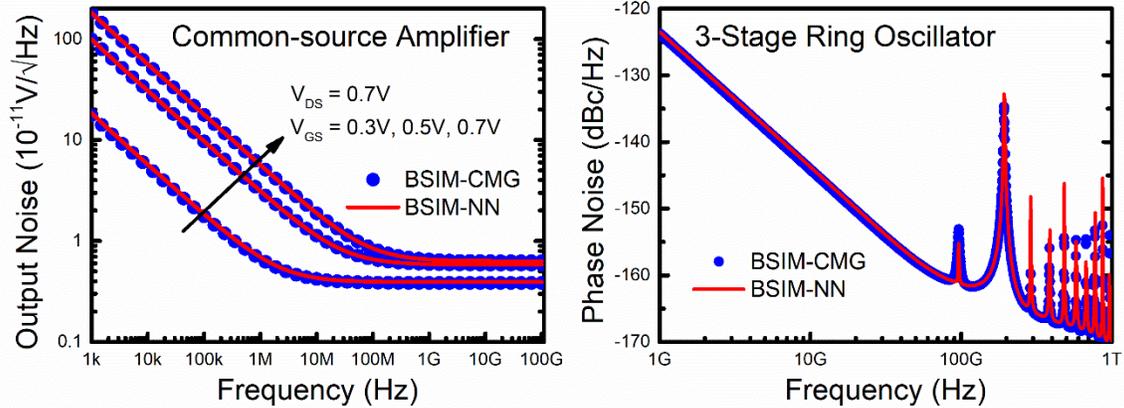


Fig. 9.11. Output noise simulation validation of a common-source amplifier and a 3-stage ring oscillator. © 2025 IEEE

## 9.4 Application for Advanced Technology

The framework can be used beyond FinFETs. I built a GAAFET with gate length=12nm, channel width=20nm, channel thickness=6nm, interfacial layer thickness=0.5nm, and oxide thickness=1.5nm. The other TCAD setup is similar to [148] including the quantum mechanics. Fig. 9.12 and 9.13 show the fitting results of transfer and output characteristics. We can see that this NN model can fit  $I_D V_G$  and  $I_D V_D$  data well up to third-order derivatives. Fig. 9.14 shows the QV and CV fittings. The NN can give accurate results of  $C_{gg}$  ( $\partial Q_G / \partial V_{GS}$ ) and  $C_{gd}$  ( $\partial Q_G / \partial V_{SD}$ ).

I have demonstrated the power of NN in modeling advanced FinFET and GAAFET. Nowadays, much research is focused on beyond Moore devices such as NCFET [152]. If we want to develop a physics-based compact model for these new devices, it usually takes a lot of time to understand physics and develop such models. By using a similar NN modeling framework, we can easily create a device model for these devices. I use TCAD to create an NC-GAAFET and train it with NN. Fig. 9.15 demonstrates the IV fitting of this device showing that the same modeling framework can also work on these emerging transistors.

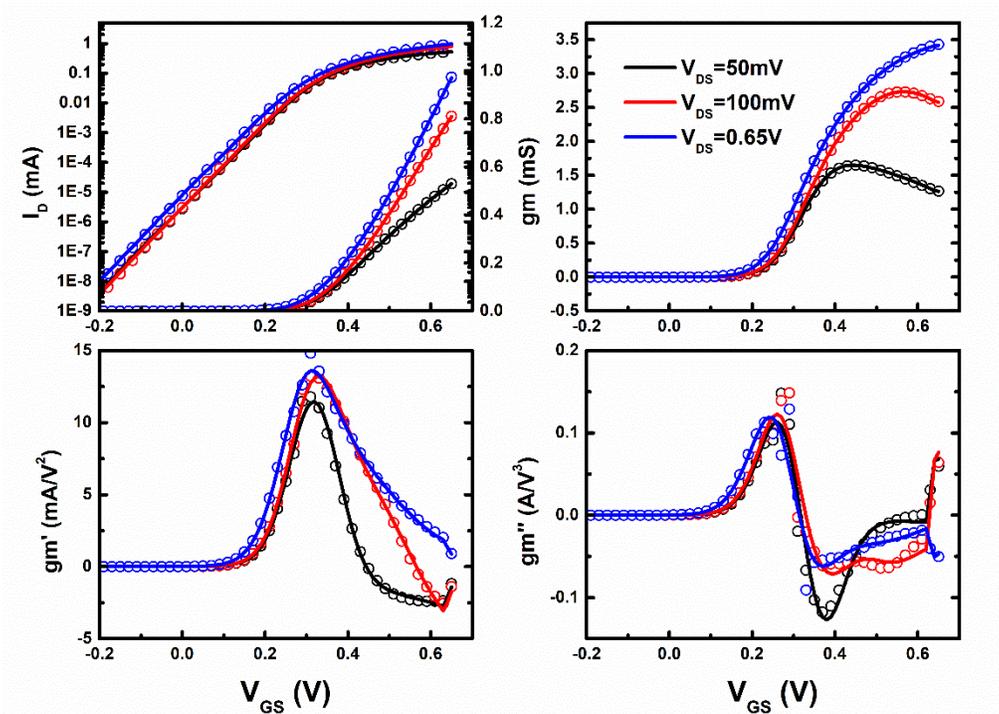


Fig. 9.12. The  $I_D V_G$  characteristics of the NN model (lines) versus the TCAD data (symbols). © 2025 IEEE

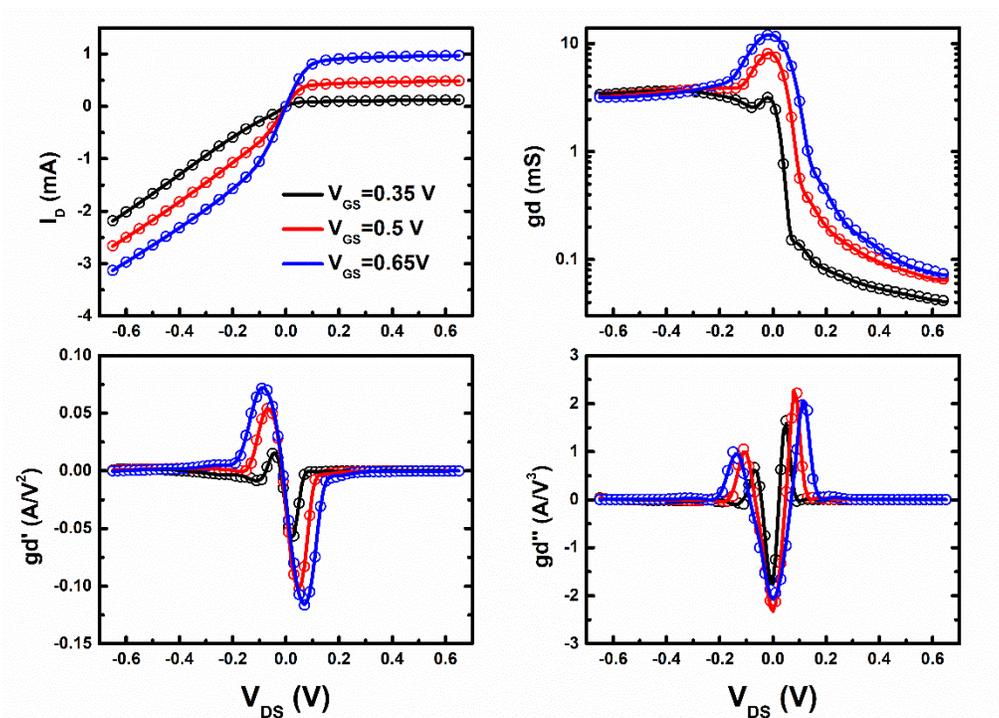


Fig. 9.13. The  $I_D V_D$  characteristics of the NN model (lines) versus the TCAD data (symbols). © 2025 IEEE

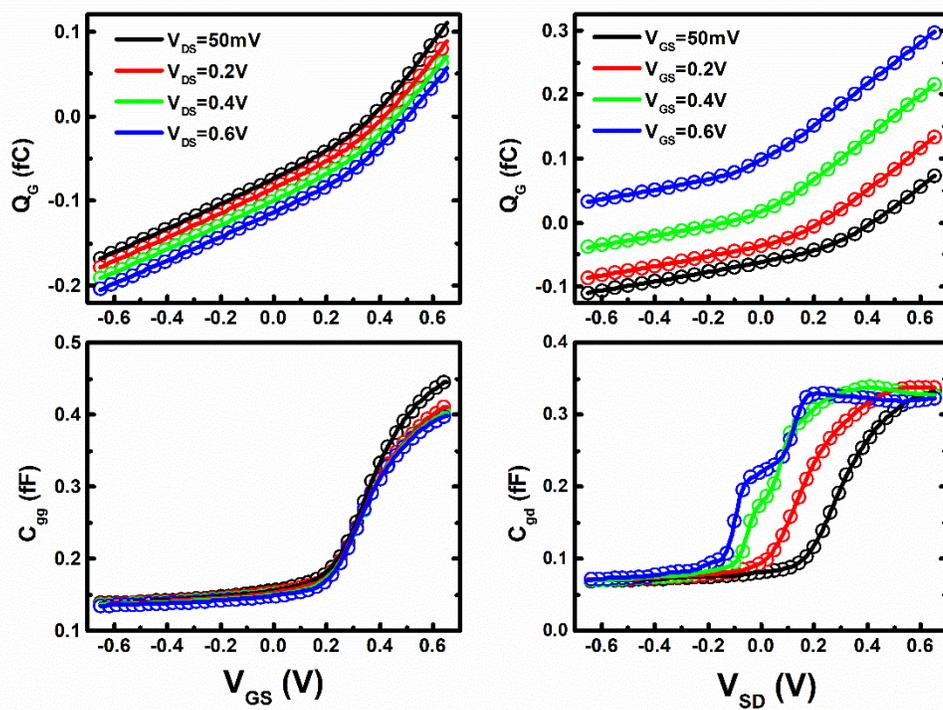


Fig. 9.14. QV and CV characteristics of the NN model (lines) versus the TCAD data (symbols). © 2025 IEEE

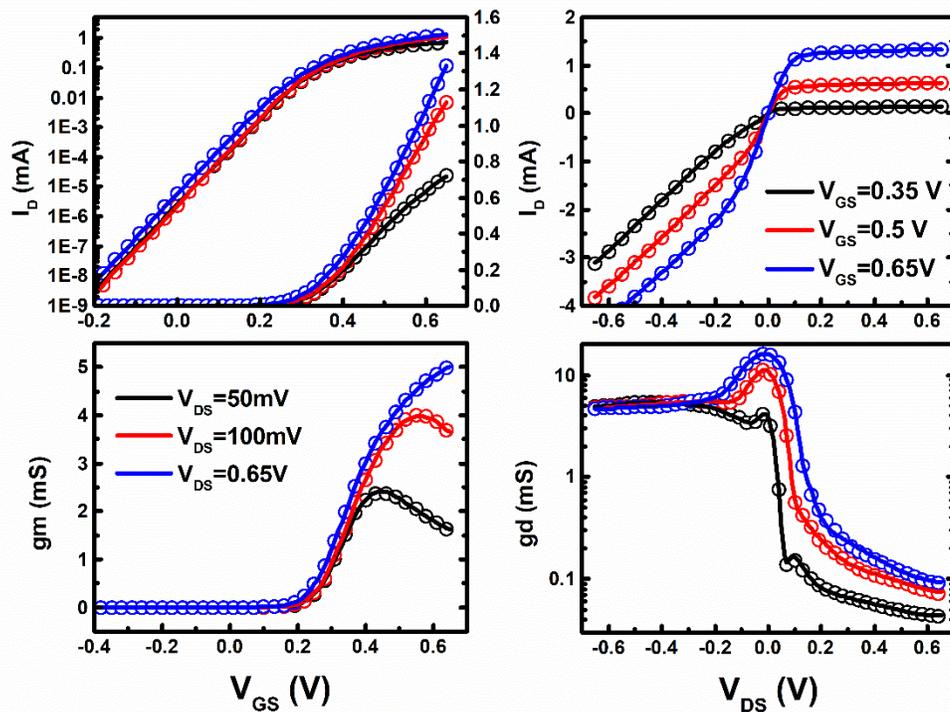


Fig. 9.15. The fitted  $I_D V_G$  and  $I_D V_D$  characteristics of NC-GAAFET where the lines are the

model, and the symbols are the TCAD data. © 2025 IEEE

## 9.5 Speed Comparison & Circuit Simulation

Both Python and SPICE are used to compare the speed of BSIM-NN and BSIM-CMG. In Python, I compared the NN IV model with the BSIM-CMG quasi-static IV model [58]. The core IV model of BSIM-CMG is coded into Python. I use NumPy and test them on Intel Xeon Platinum 8260 versus bias points. The NN model holds about 13 times speed advantage as shown in Fig. 9.16. In the case of 10 million DC points, NN takes 59.6s while BSIM-CMG takes 806s. This could be because of the reduced number of coding lines and equations or the benefit of matrix multiplication. If we can further optimize the network and use hardware acceleration such as GPU, the NN speed advantage maybe even more.

I also test the relationship between inference time and the number of hidden layers and nodes shown in Fig. 9.17. We see that both inference times increase linearly with layers and nodes. It is roughly 3s per layer and 1.1s per 10 nodes. This gives an insight that if we want to increase the complexity of NN, we should try increasing the nodes first and then increasing the layers. In any case, the NN is still much faster than conventional compact models. In addition to the model inference time, the convergence time of the circuit solver is also an important part of the total simulation time. To test the model convergence, we use the nonlinear solver in Python to solve an inverter circuit in DC simulation. 1k voltage inputs are fed into the CMOS inverter circuits composed of BSIM and NN models. It takes NN model 3.95s to process them and 20.70s for the BSIM circuit.

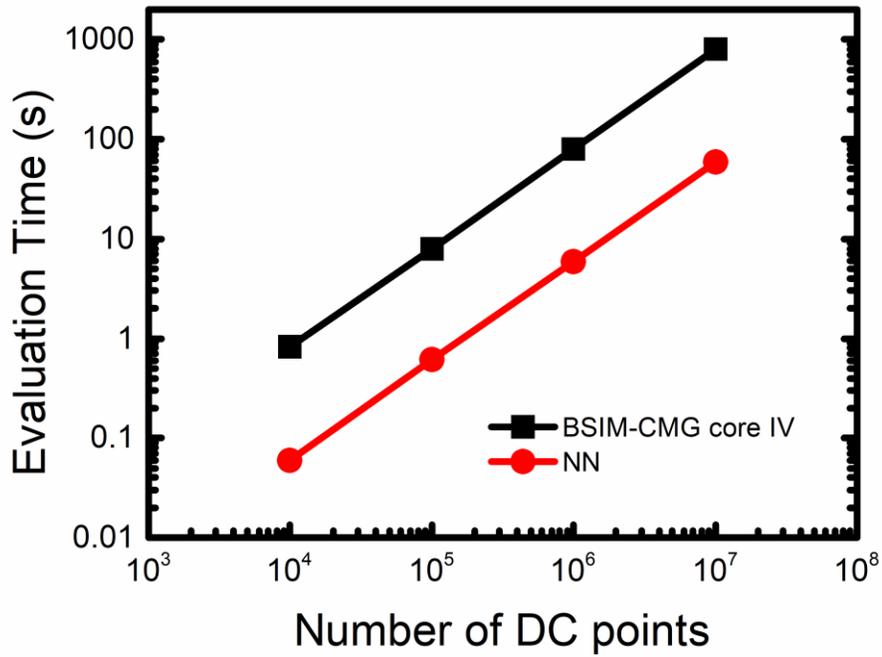


Fig. 9.16. Evaluation time comparison between NN and BSIM-CMG IV model. © 2025 IEEE

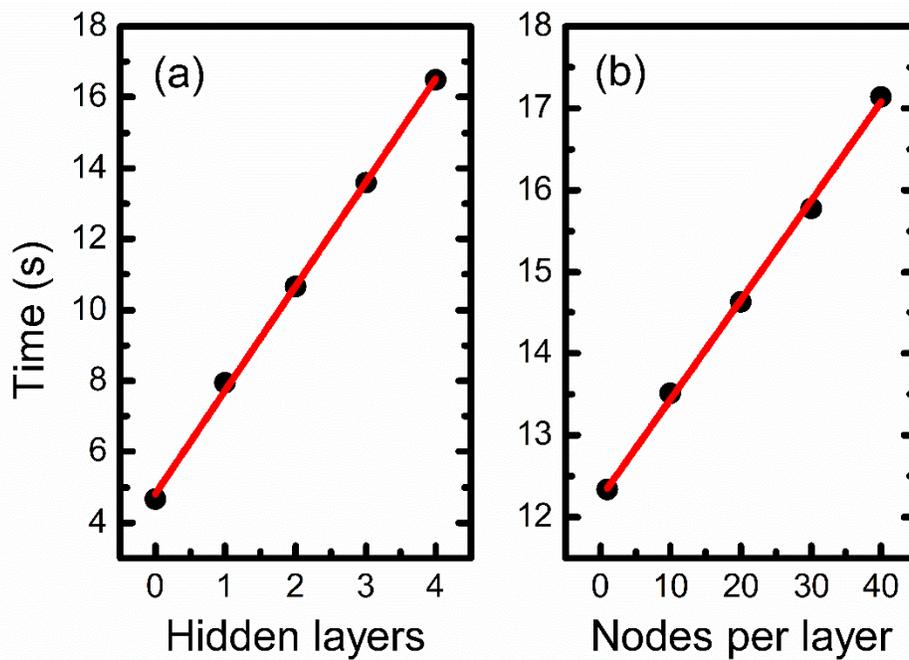


Fig. 9.17. (a) The inference time versus the number of hidden layers with 10 nodes per layer. (b) The inference time versus the number of nodes per layer where 3 hidden layers

are set. © 2025 IEEE

For SPICE, the model is coded into Verilog-A and tested with different circuit simulations. Fig. 9.18 to Fig. 9.24 shows the simulation results of a 3-input NAND, FO4 inverter chain, D-flip-flop, FO4 151-stage NAND ring oscillator, 32-bit ripple adder, and SRAM variation. All BSIM-NN results closely match the BSIM-CMG results with errors of less than 0.3%. In all of these simulations, BSIM-NN shows great speed improvement compared to BSIM-CMG. For example, in SRAM variation simulation, BSIM-NN is two times faster than BSIM-CMG. Fig. 9.25 summarizes the simulation time of different benchmarking circuits. The BSIM-NN outspeeds BSIM-CMG by 3~5 times. If matrix multiplication and GPU inference are applied to model evaluation in SPICE, the speed improvement can be even greater.

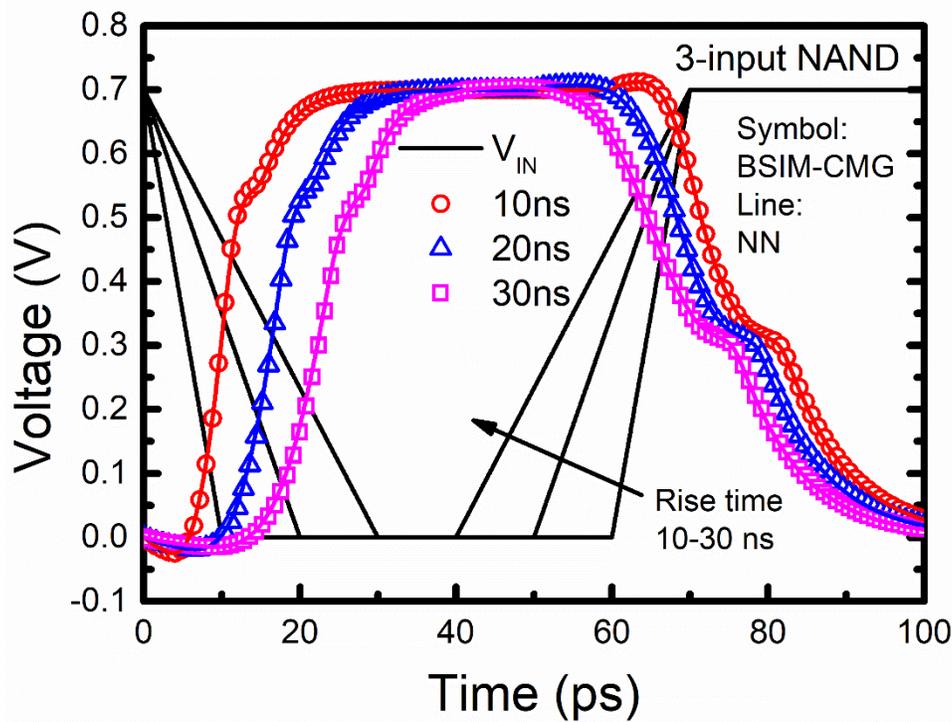


Fig. 9.18. A 3-input NAND simulation with different slew rates.

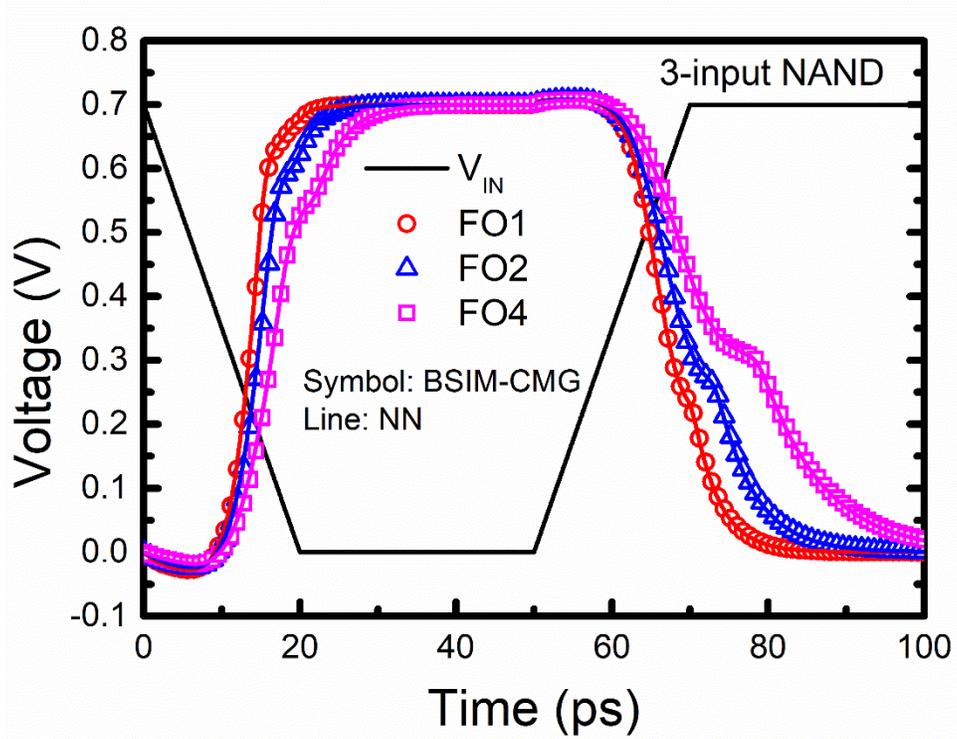


Fig. 9.19. A 3-input NAND simulation with different fan outs.

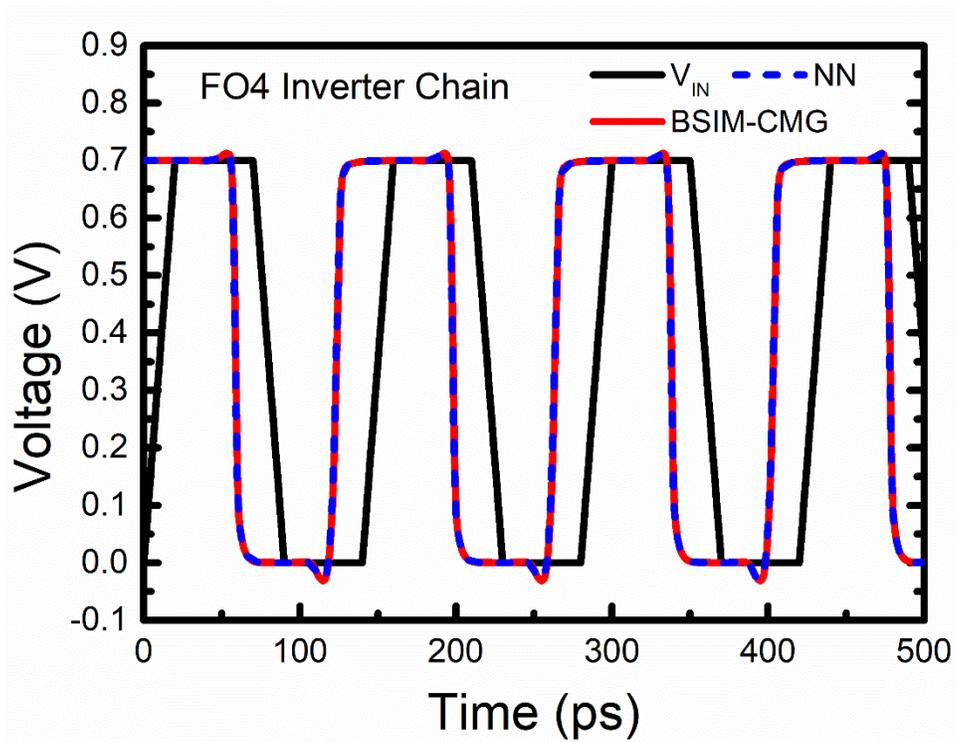


Fig. 9.20. A 5-stage FO4 inverter chain simulation. The load goes from 1 to 256 inverters.

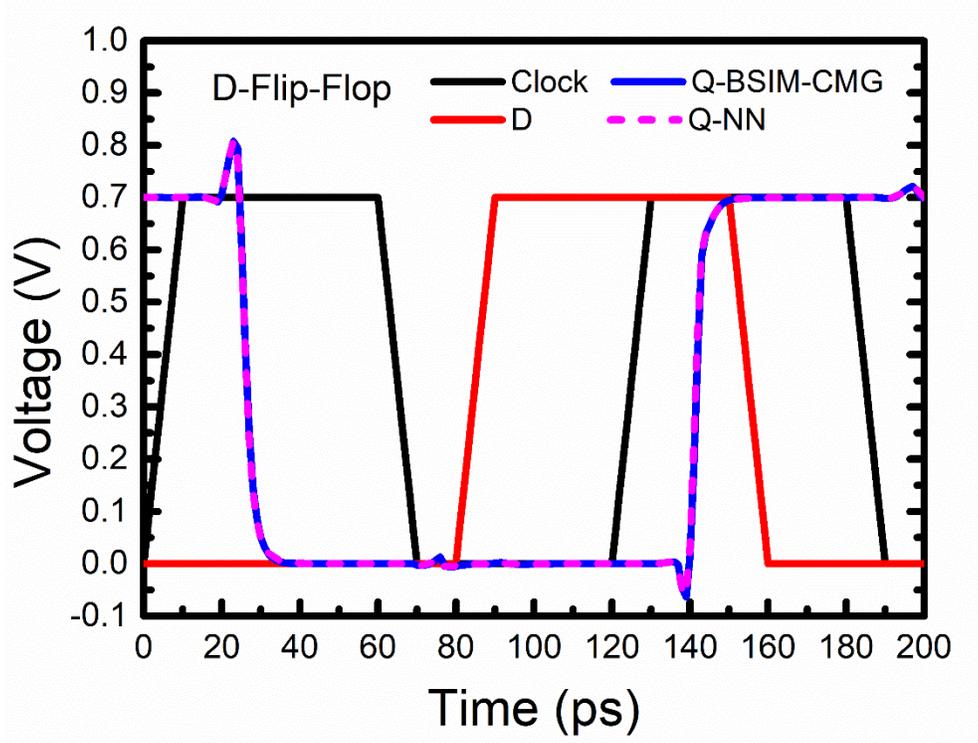


Fig. 9.21. A D-flip-flop simulation.

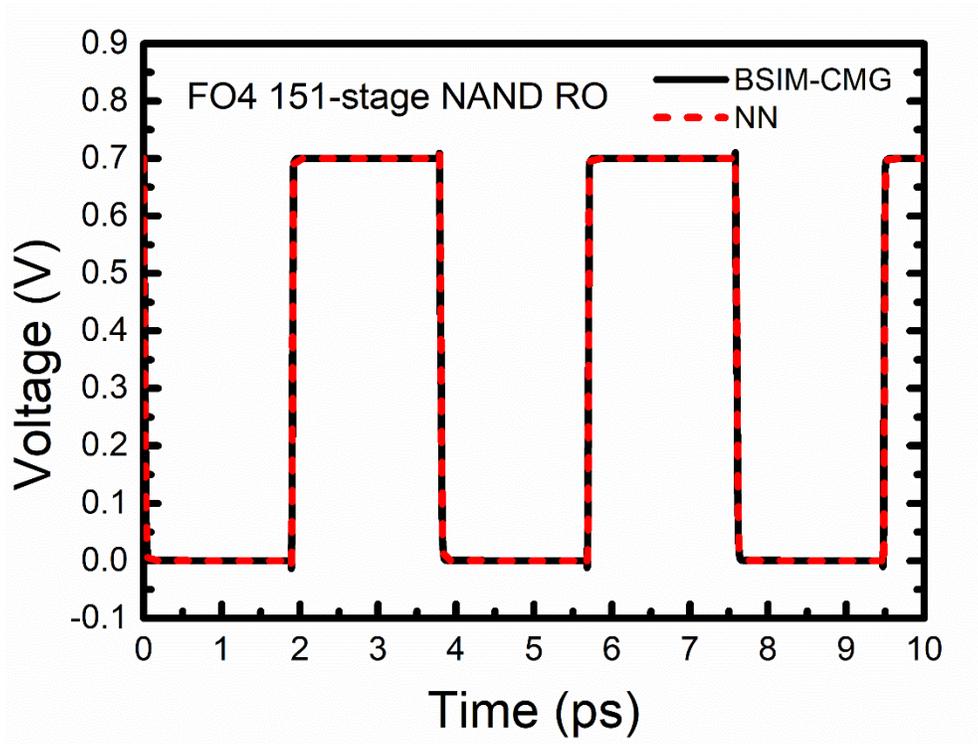


Fig. 9.22. A FO4 151-stage NAND ring oscillator simulation.

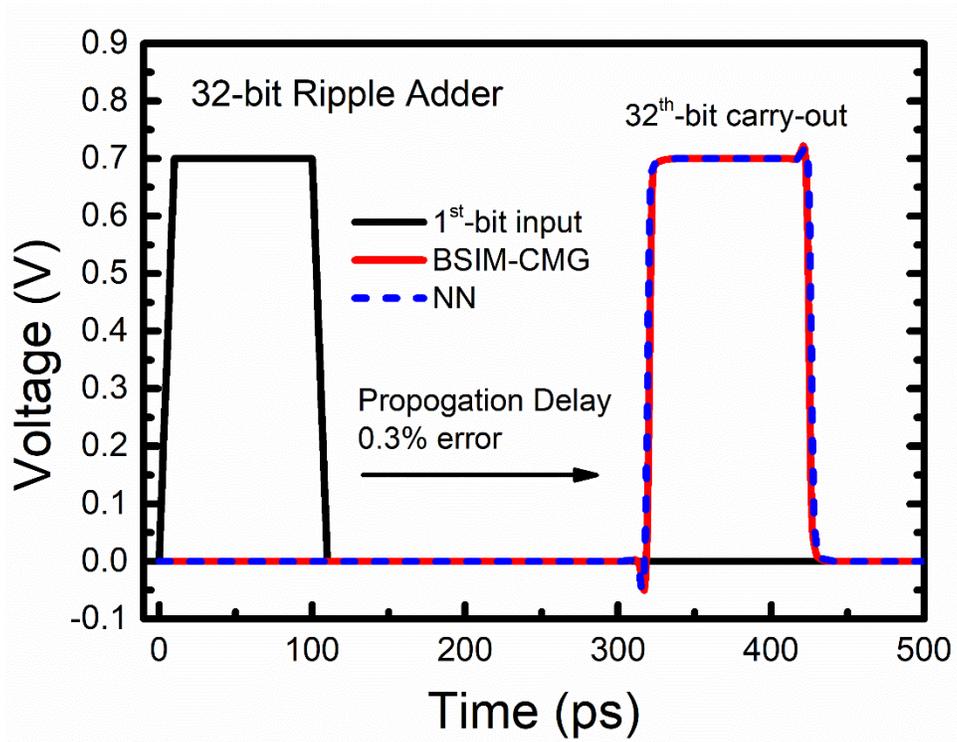


Fig. 9.23. A 32-bit ripple adder simulation.

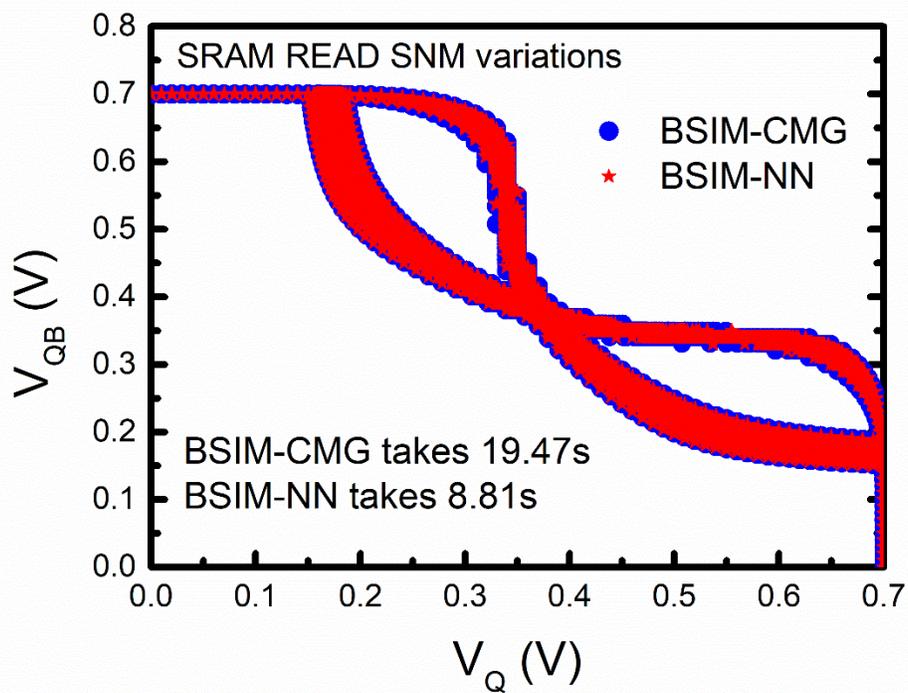


Fig. 9.24. A SRAM READ SNM (static noise margin) variation simulation. BSIM-NN is 2 times faster than BISM-CMG.

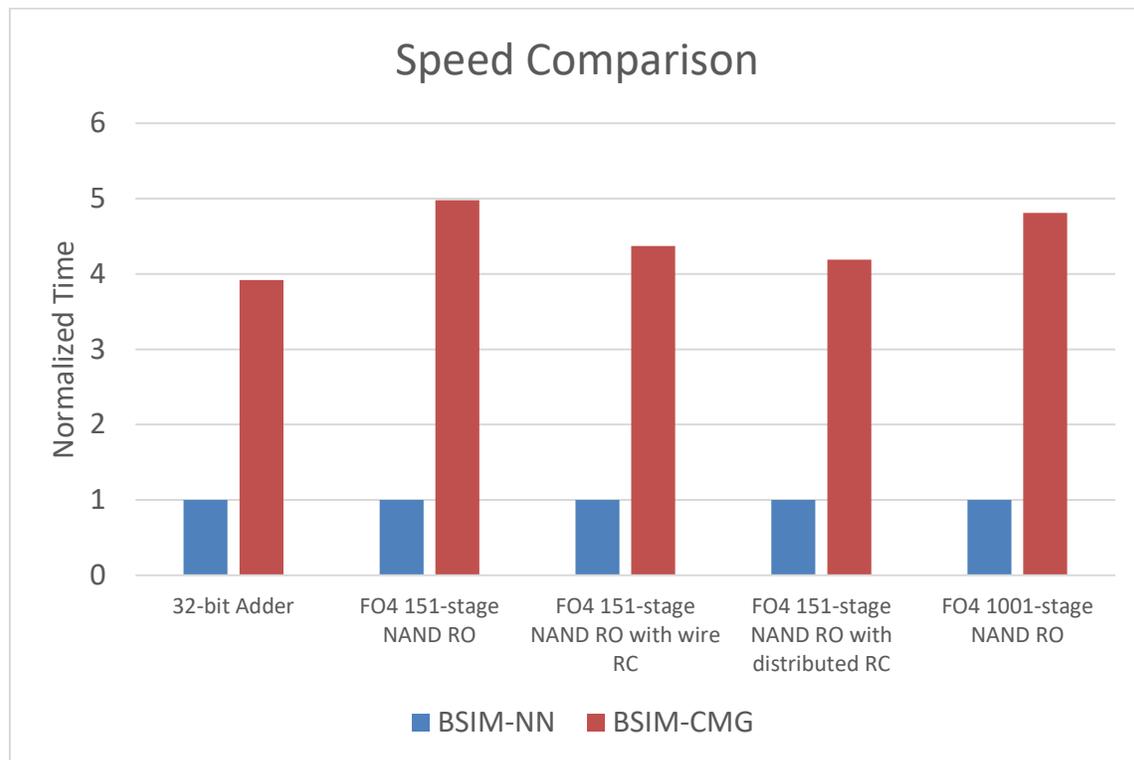


Fig. 9.25. Speed comparison of BSIM-NN and BSIM-CMG with different benchmarking circuits. The speed boost is around 3~5X.

## 9.6 Chapter Summary

I demonstrate BSIM-NN, a neural network compact model for advanced transistors such as FinFET, GAA, and NCFET. It models complex IV and CV characteristics with  $I_D$ ,  $I_G$  including leakage current and all charges ( $Q_G$ ,  $Q_S$ ,  $Q_D$ ). I demonstrate that the proposed model can accurately predict the variability of the device and give smooth and correct high-order derivatives. The activation function and Verilog-A implementation are studied to improve the circuit simulation speed. The model is tested with various circuit simulations and shows accurate results versus the standard compact model with no convergence issue. It can speed up the IC simulation time by more than 5 times compared to BSIM-CMG.

## **Chapter 10 Non-Quasi-Static Modeling of Neural Network-based Transistor Compact Model for Fast Transient, AC, and RF Simulations**

I developed a charge deficit-based non-quasi-static (NQS) model that is compatible with neural network-based transistor compact models for transient, AC, and RF simulations. The charge deficit model calculates the deficient or surplus charge in the channel to model the NQS effect. I introduce physics-based parasitic charges to extract intrinsic channel charges from trained quasi-static (QS) NN models. An improved loss function is also proposed to obtain physical charge values from capacitance-only training data. A charge deficit subcircuit is applied to calculate the NQS currents. I demonstrate the model's accuracy in transient, AC, s-parameter, and RF mix-signal simulations. The proposed model can be easily integrated with QS NN-based compact models without the loss of model efficiency. This work is published by IEEE. © 2025 IEEE. Reprinted, with permission, from [156]

### **10.1 Motivation**

Non-Quasi-Static (NQS) modeling is an important issue in transistor simulations. The Quasi-Static (QS) assumption fails when the circuit operation speed approaches and exceeds the transistor's cut-off frequency. Neural network (NN)/machine learning (ML)-based compact models developed so far are based on QS/DC data without NQS models [142, 146, 153, 157-159]. Therefore, to model transistors at high speed/frequency, there needs to be an NQS model for the NN-based compact model. In conventional compact models, several ways are proposed to model the NQS effect [160-166]. The channel segmentation approach divided the channel into several transistors [163, 164]. Each segment can be modeled by an NN model.

However, it also heavily increases the simulation time. The charge segmentation model solves the charge along the channel with spline interpolation [165, 166], which is time-consuming, requiring multiple nodes, and incompatible with NN models since many NN models do not compute the intrinsic channel charges [142, 146, 153, 157]. One simple and elegant way is the relaxation time approach (RTA) or charge deficit model which tracks the deficient or surplus charge in the channel [160-162] and is adopted by BSIM [167]. It is efficient and needs only one extra node. This method is also not directly compatible with NN models without suitable adjustment for the same reason as the charge segmentation model. In addition to including physics-based NQS models in the NN-based compact models, it is possible to build a time-dependent model directly with recurrent neural networks (RNN). RNN can model transient circuit-level performance [168, 169]. However, the training will require transient data, which is not a common practice for model fitting and may not be as reliable as DC data.

In this chapter, I demonstrate a physics-based NQS model that is compatible with NN-based transistor compact models. The parasitic charges and charge deficit model are introduced compatibly to the NN model to provide efficient modeling of the NQS effect. The model is verified with the transient, AC, and RF data from BSIM-CMG to prove the accuracy of IC simulations.

## 10.2 Model

Fig. 10.1 shows the methodology to model NQS effect with NN and the charge deficit model. First, the QS IV and QV characteristics are trained and fitted with the DC IV and CV training data by using the NN modeling methodologies in the previous chapter [142, 153]. Then, the trained QS total terminal charge is used for the NQS calculations during circuit simulations. I chose the charge deficit model which is efficient and proved to accurately fit the NQS results and can be easily integrated with the QS core device model [160-162]. The key of the charge deficit model is to calculate the time-dependent intrinsic channel charges ( $Q_i$ ) which is restricted by the transit time. Many of the NN compact modeling methods directly model the total terminal charges ( $Q_{G,S,D}$ ) or terminal capacitances [142, 146, 153, 157]. Therefore, the

problem is how to extract intrinsic charges from these black box models.

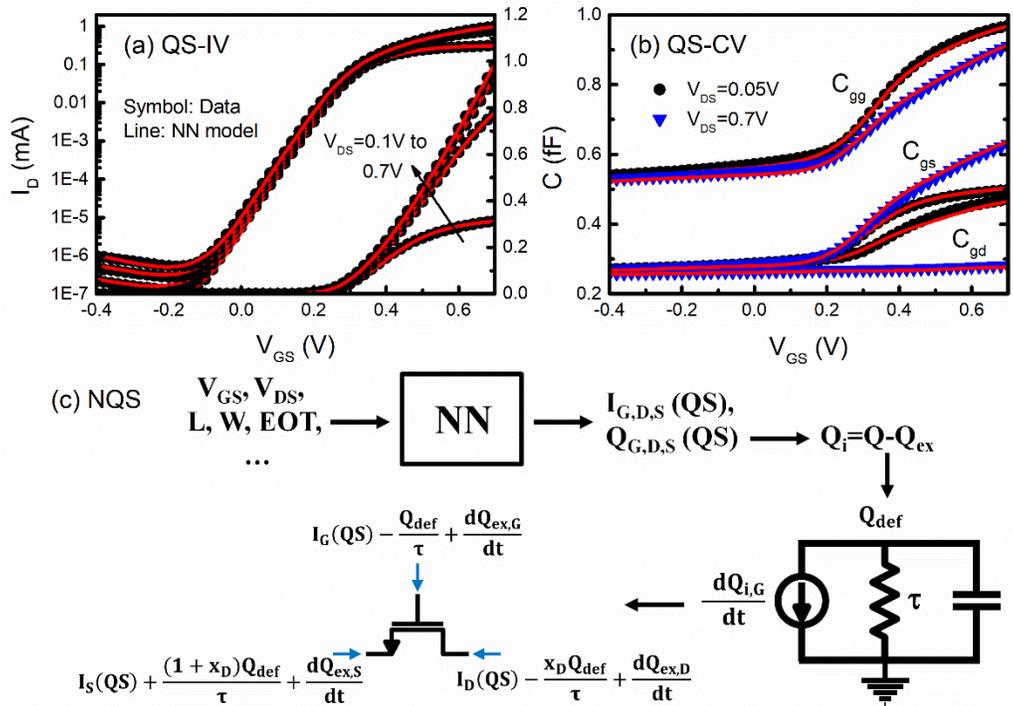


Fig. 10.1. (a) QS IV curves of the NN model. (b) QS CV curves of the NN model. (c) The schematic diagram of NQS modeling flow. © 2025 IEEE

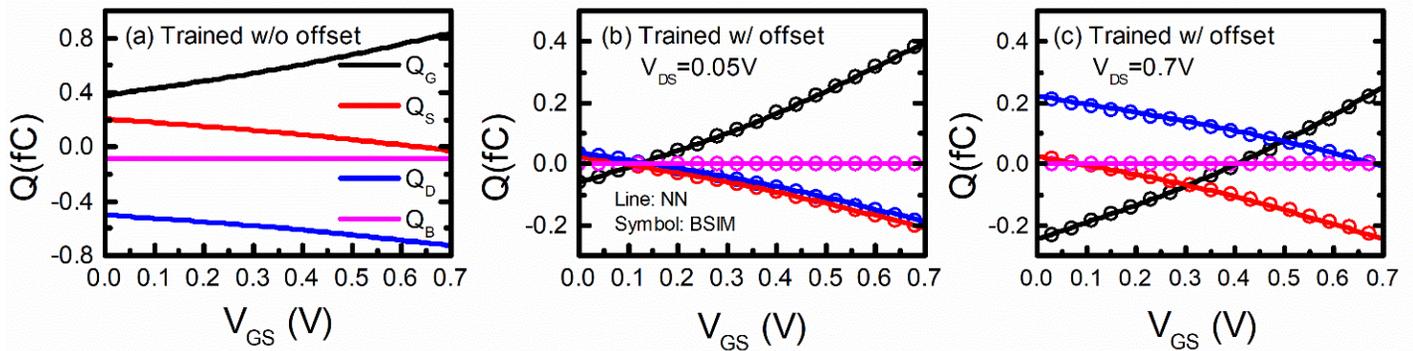


Fig. 10.2. (a) Model charges trained without offset included. (b) Model charges trained with offset included at  $V_{DS}=0.05V$  (c)  $V_{DS}=0.7V$ . © 2025 IEEE

I introduce physical parasitic charge models to extract the intrinsic

charges from our NN-based FET model [142]. The parasitic charges can be characterized as fringing ( $Q_f$ ) and overlapping charges ( $Q_{ov}$ ) modeled by physical equations [167]. Then,  $Q_i=Q-Q_{ex}$  ( $=Q_f+Q_{ov}$ ) at each terminal (G,D,S). Users will manually extract the parasitic charges with parameters such as CFS, CGSL, and CGSO in BSIM-CMG [167] to get accurate  $Q_i$ . We also need the physical terminal charge value. It is easy when the  $Q_{G,S,D}$  are available in the training data using TCAD [142, 146, 153] but it is not the case when using experimental data. Training models solely with capacitance data can cause nonphysical offset [146] and is not suitable for NQS modeling. To overcome this, we introduce offset terms ( $Q_{G,S,D0}$ ) in the loss function Eq. (10.1) to force the charges to have physical values (Fig. 10.2).  $Q_{G,S,D0}$  is the charge at  $V_{GS}=V_{DS}=0V$  which can be chosen based on the simple physical estimations, and a-e are the coefficients for tuning the loss. The value of a-e is chosen so that the magnitude of each term is comparable. It can be done by training iteratively to determine the coefficients one by one [142, 153].

$$\begin{aligned} loss = & a \cdot RMS\left(\frac{\partial Q_{G,S,D}}{\partial V_{GS}}\right) + b \cdot RMS\left(\frac{\partial Q_{G,S,D}}{\partial V_{DS}}\right) \\ & + c \cdot RMS\left(\frac{\partial^2 Q_{G,S,D}}{\partial V_{GS}^2}\right) + d \cdot RMS\left(\frac{\partial^2 Q_{G,S,D}}{\partial V_{DS}^2}\right) + e \cdot RMS(Q_{G,S,D0}) \end{aligned} \quad (10.1)$$

Finally,  $Q_i$  goes into a subcircuit shown in Fig. 10.1 to calculate the deficit charge ( $Q_{def}$ ). The terminal currents Eq. (10.2) are the DC currents ( $I_{G,D,S}$ ) plus NQS current ( $Q_{def}/\tau$ ) and displacement currents from extrinsic parasitics ( $Q_{ex,G,D,S}$ ) where  $\tau$  is the transit time, and  $x_D=Q_{i,D}/Q_{i,G}$ . Thus, the correct extraction of the  $Q_{ex}$  is important to obtain accurate transient current and frequency responses.

$$I_G(t)=I_G(QS) - \frac{Q_{def}}{\tau} + \frac{dQ_{ex,G}}{dt} \quad (10.2a)$$

$$I_D(t)=I_D(QS) - \frac{x_D Q_{def}}{\tau} + \frac{dQ_{ex,D}}{dt} \quad (10.2b)$$

$$I_S(t)=I_S(QS) + \frac{(1+x_D)Q_{def}}{\tau} + \frac{dQ_{ex,S}}{dt} \quad (10.2c)$$

This approach can be easily integrated with the NN models with high computational efficiency. It can also be simply improved when more accurate

RTA-based NQS models are developed for physics-based compact models.

### 10.3 Validation

The core QS model is trained with a calibrated BSIM-CMG model including L, W, and EOT [142] to accurately fit the DC characteristics (Fig. 10.1) and is implemented in Verilog-A. Then, I test it with BSIM-CMG having the NQS model turn-on using Hspice on Intel Xeon Platinum 8260. Fig. 10.3 shows the transient simulation of the NN-based NQS model at  $V_{DS}=0.05V$  &  $0.7V$ . A pulse is given at gate with rise/fall time=2ps, pulse width=2ps, and amplitude=0.7V. The blue curves show the QS-only NN model, and the red curves show the NN model with the NQS effect. With the introduction of the physics-based parasitic charges and NQS models, the NN model can accurately fit the transient  $I_{D,NQS}$ ,  $I_{S,NQS}$ , and  $I_G$  data where the NQS effect causes extra delay in the transient currents. In addition to the high-speed switching, the NQS effect is important in high-frequency AC responses.

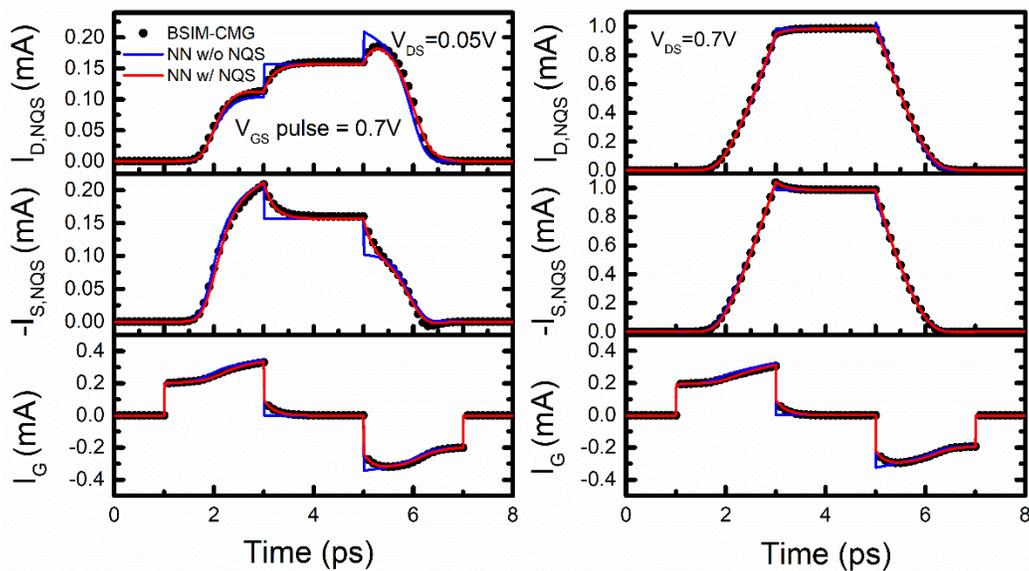


Fig. 10.3. Transient simulation result when a pulse with 2ps-rise/fall time, 2ps-pulse width, and 0.7V amplitude is applied to the gate.  $V_{DS}=0.05V$  &  $0.7V$ .  $I_{D,NQS}$  and  $I_{S,NQS}$  are the currents without the parasitic components. © 2025 IEEE

I test the AC response using a common source amplifier with  $1k\Omega$  load resistance as shown in Fig. 10.4. The NQS effect reduces the bandwidth

captured by our model. To further examine the small-signal characteristics, I conduct a s-parameter simulation. Fig. 10.5 shows the simulated s-parameters from BSIM-CMG and NN models with 8 FETs in parallel. At low frequencies, the NN with and without the NQS model shows the same characteristics. However, when the frequency increases to hundreds of GHz, the NQS NN model starts to deviate from the QS result due to the frequency-dependent capacitance and transconductance. Besides the transient and AC simulations, RF mixed signal response is evaluated. A Pin-Pout simulation is presented in Fig. 10.6 also with 8 FETs in parallel, which shows the model's capability to produce accurate RF simulation results. In Fig. 10.7, I simulate  $f_T$  and  $f_{max}$  with BSIM-CMG and the NN model where the proposed model shows excellent match.

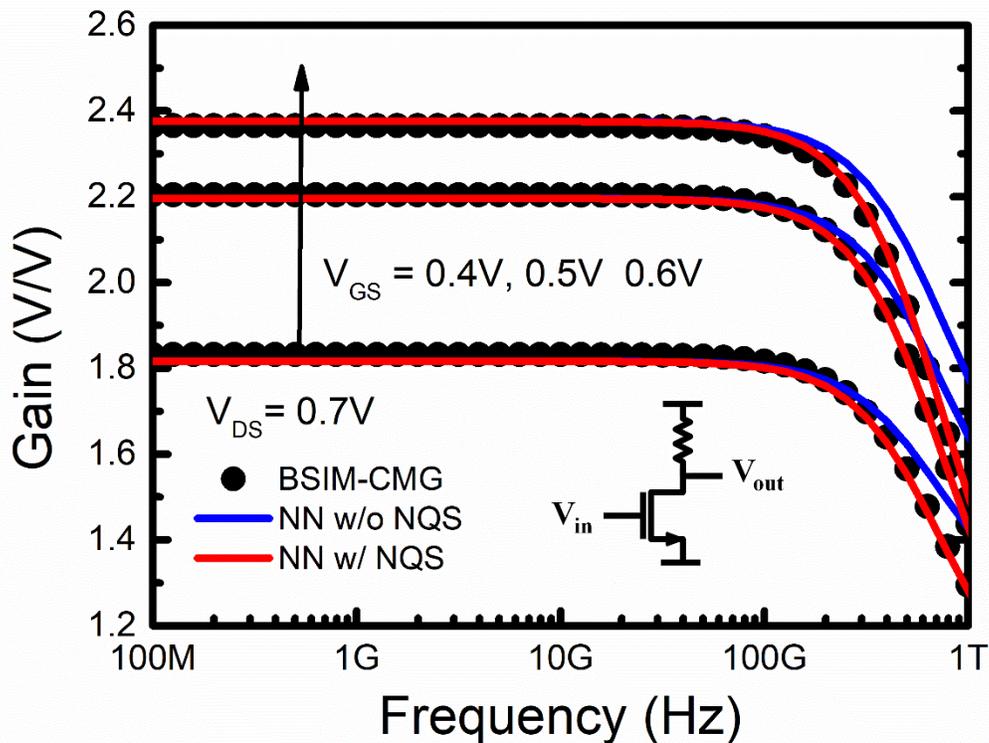


Fig. 10.4. AC simulation result of a common-source amplifier with 1kW load resistance.

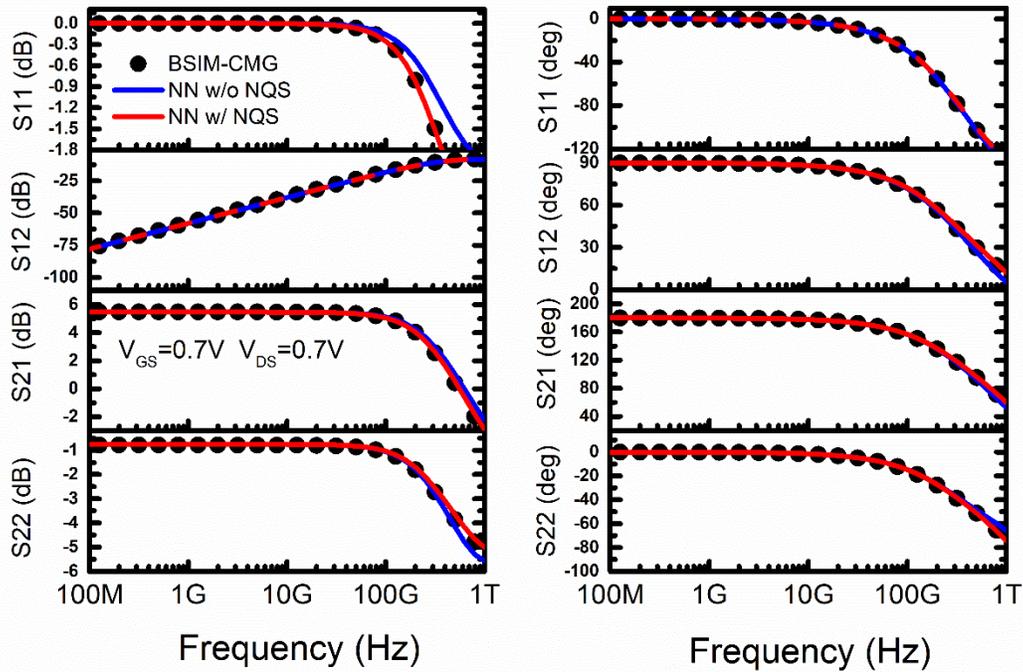


Fig. 10.5. S-parameter simulation with a 50  $\Omega$  load impedance. © 2025 IEEE

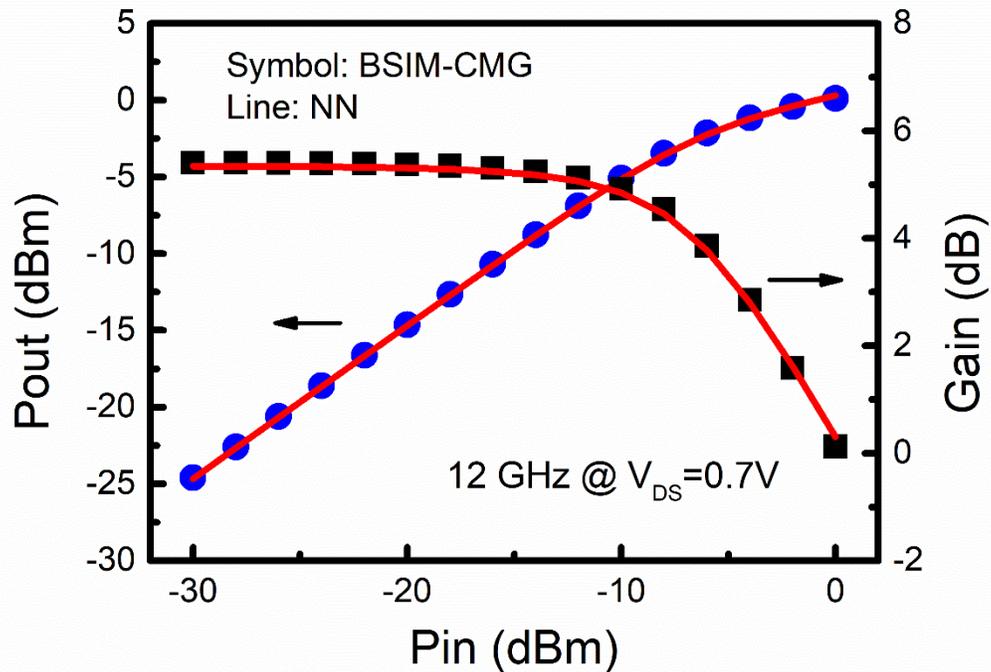


Fig. 10.6. RF Pin-Pout simulation of the NN model and BSIM-CMG. © 2025 IEEE

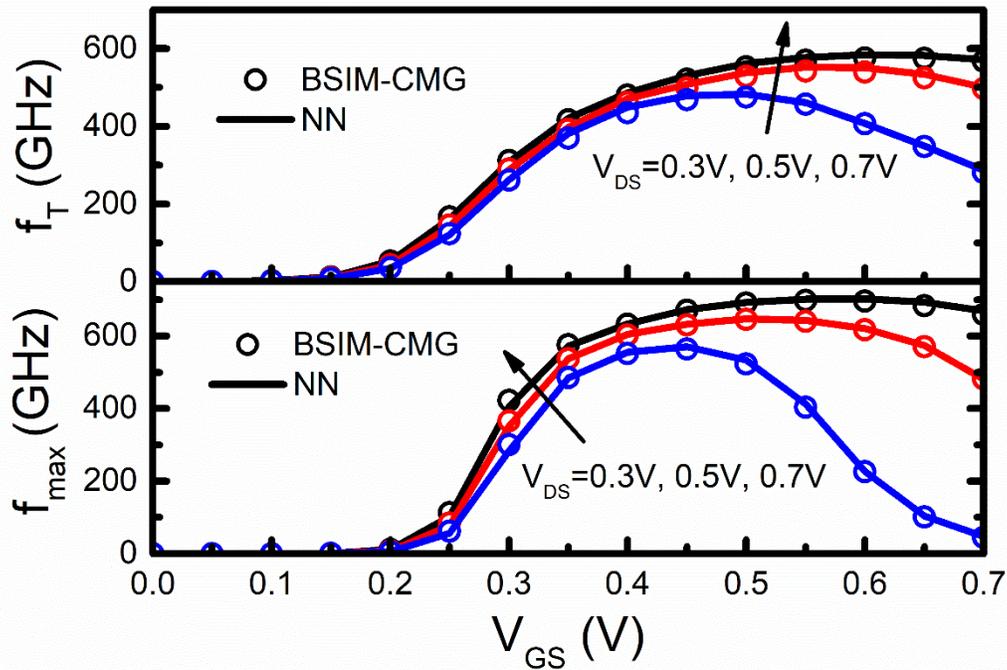


Fig. 10.7. Simulated  $f_T$  and  $f_{max}$  from BSIM-CMG and the NN model. © 2025 IEEE

Fig. 10.8 demonstrates a killer NOR gate simulation result. Without the NQS model, the  $V_X$  tends to go more negative than the NQS model predicts (green circles), which often causes unphysical results [166]. I further test the model with a 151-stage NAND oscillator (Fig. 10.9). The NQS effect changing the delay in the circuit is captured by the new NN model.

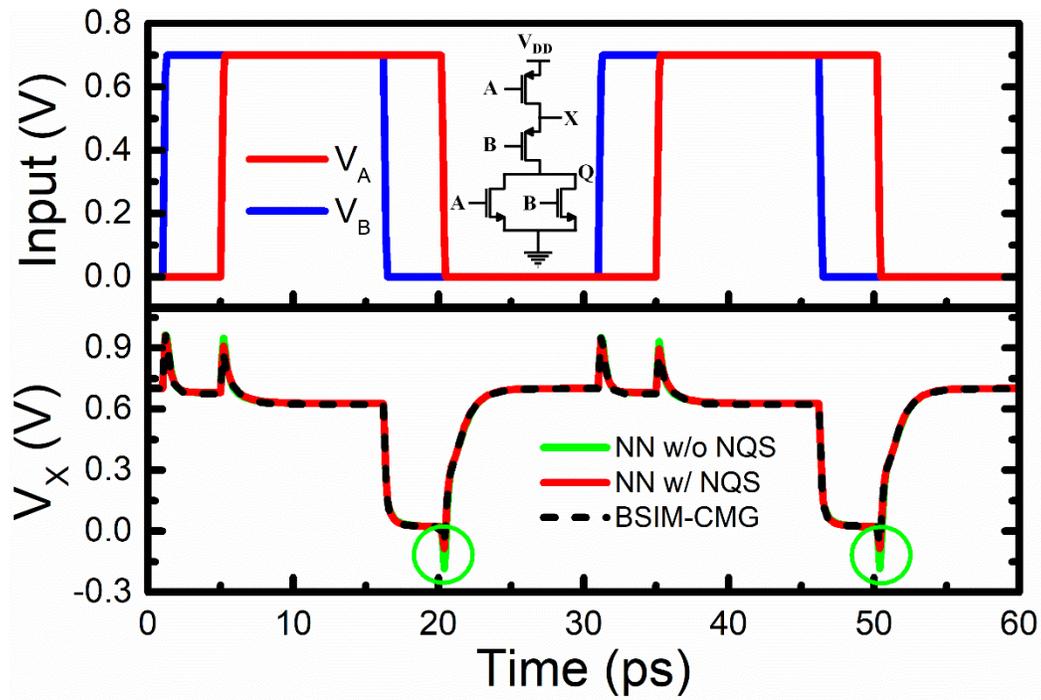


Fig. 10.8. Killer NOR gate simulation. © 2025 IEEE

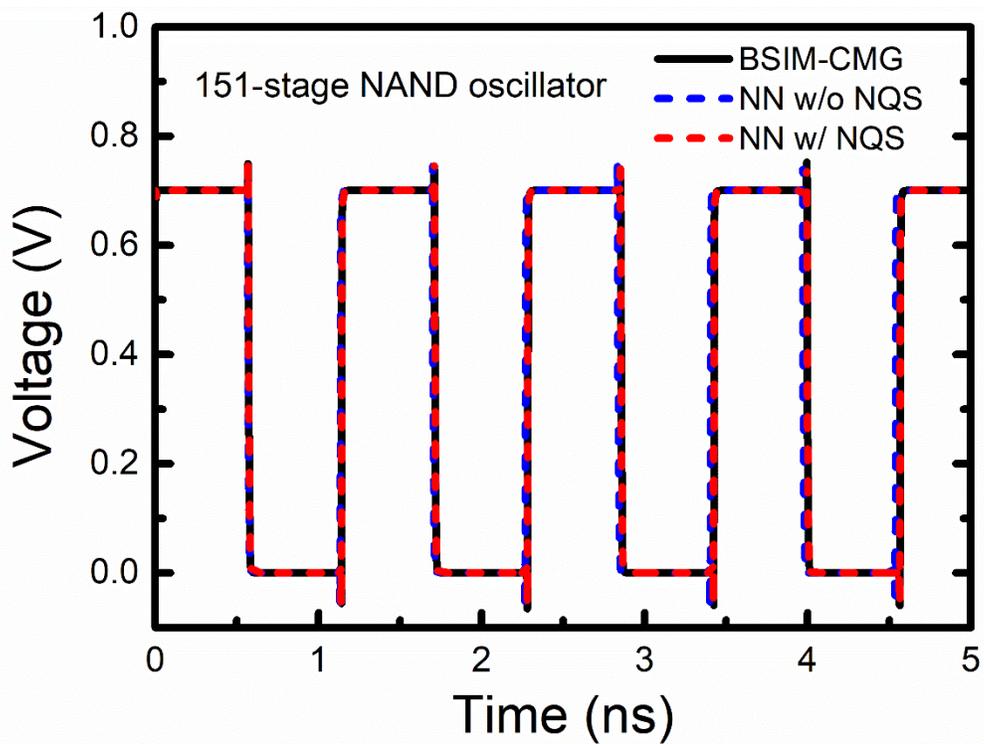


Fig. 10.9. 151-stage NAND oscillator simulation.

## 10.4 Chapter Summary

I proposed a methodology to integrate the NN-based transistor compact model with the physics-based NQS model, which can be directly applied to DC-trained NN models. Parasitic charges and offset charges are introduced to extract the intrinsic charges and the deficit charge. I showed the model's capability of doing high-speed transient, AC, and RF simulations with the same accuracy as standard compact models but with a much faster simulation speed.

## Chapter 11 A Novel Neural Network-based Transistor Compact Model Including Self-Heating

I develop a SPICE-compatible neural network-based compact model to accurately capture the temperature dependence and self-heating effects in Field Effect Transistors (FETs). The model is based on artificial neural networks with no semi-empirical temperature equations. The transfer and activation functions are optimized to improve the accuracy of the model. A new temperature relaxation model is proposed, which allows training the model using ambient temperature data without iteratively extracting the self-heating parameters. The proposed method can simply generate the ambient and dynamic self-heating characteristics for circuit simulations. The model can accurately reproduce the current-voltage (IV), capacitance-voltage (CV), and transient characteristics of FETs across a broad temperature range. This work is published by IEEE. © 2025 IEEE. Reprinted, with permission, from [170]

### 11.1 Motivation

Thermal effects play a crucial role in determining device and circuit performance [171-174]. Modern-day chips house billions of transistors, exhibiting a significant power density, operating at temperatures substantially higher than room temperature. Elevated device temperatures result in the degradation of mobility and subthreshold swing, shifts in threshold voltage, and reliability concerns. Self-heating (SH) effects contribute to a gradual increase in device temperature, thereby impacting circuit performance. With  $V_{DD}$  scaling stalled at around 0.7V and the projected rise in densities of transistors and 3D chiplet packaging, temperature and self-heating will be increasingly important to the accuracy of circuit simulation.

Conventional physics-based compact models [167, 175-177] start from

modeling the temperature dependence of physical quantities such as mobility and bandgap. Then, a separate self-heating network is used to update the temperature in the device. While extracting the model parameters, it is required to iteratively adjust the temperature and self-heating parameters. In addition to physics-based compact models, neural network (NN)/machine learning (ML)-based compact models follow the same approach to develop temperature models [159, 178-181]. These models use device temperature as input and either need to evaluate the SH subcircuit self-consistently with the NN model during the training [181], or extract/ remove the self-heating effects before the training [159, 179, 180]. These approaches require SH evaluation before or during the training, making the extraction procedure complicated.

This work proposes a new temperature and self-heating modeling approach with my NN transistor compact model [142, 153]. The proposed methodology starts with  $I(T_0)/Q(T_0)$ , not  $I(T)/Q(T)$ , where  $T$  is the device temperature and  $T_0$  is the ambient temperature.  $I(T_0)$  can simply come from DC measurements instead of pulse measurements. The model can be trained with  $I(T_0)$  and extract or recover the self-heating-free characteristics and perform the dynamic self-heating simulation with a newly developed temperature relaxation model. The extracted characteristics can also be integrated with the standard SH model for circuit simulations.

## 11.2 Model

The NN compact model is trained using DC-IV and CV characteristics including the SH effect. The model is enhanced from our previous work [142, 153] by adding  $T_0$  as an input parameter and a new transfer and activation function to improve accuracy. I train the model to be a function of  $T_0$  instead of  $T$ , such that no SH evaluation is required during training. Then, a new temperature relaxation (TR) model is used to recover the SH-free characteristics and model the SH effect after the model is trained.  $V_{GS}$ ,  $V_{DS}$ ,  $L$ ,  $W$ ,  $EOT$ , and  $T_0$  are the designated input parameters for our model. The outputs are the corresponding DC currents and charges at  $T_0$ . The Inverse Square Root Unit (ISRU) function  $ISRU(x) = \frac{x}{\sqrt{1+\beta x^2}}$  is used as an activation function. ISRU is more computationally efficient than sigmoid and tanh

implementations. The parameter  $\beta$  can be tuned to improve fitting accuracy. In the scenario where SH-free data is available, it can directly be trained using  $T$  as an input to get  $I(T)$  and  $C(T)$ . No semi-empirical temperature equations are used in the proposed model. Thus, the model is not restricted to certain predefined temperature dependencies.

To model the SH dynamic, I propose a new temperature relaxation (TR) model to map the device temperature to the corresponding ambient temperature. The traditional way to model the device temperature  $T(t)$  is  $T(t)=T_0+DT(t)$ , where  $DT(t)$  is the change in temperature due to SH. We define a temperature  $T_0'(t)$  such that  $T_0=T_0'(t)+dT(t)$  and  $T_0'(t)=T_0-dT(t)$  Eq. (11.1).  $dT_i=R_{TH}I_D V_{DS}$  Eq. (11.2) is introduced, which is the  $dT(t)$  to recover the SH-free characteristics. It is also the same as the steady-state  $DT(t)$  in the traditional model. However, the thermal resistance for the trained device ( $R_{TH}$ ) and the device for IC design ( $R'_{TH}$ ) can be different.  $DT(t)$  for the latter is  $R'_{TH}I_D V_{DS}$  and the steady-state  $dT(t)=dT_i-R'_{TH}I_D V_{DS}$  Eq. (11.2). The dynamic of  $dT(t)$  is then modeled by Eq. (11.3), where  $\tau=R'_{TH}C_{TH}$  is the thermal time constant and  $C_{TH}$  is the thermal capacitance. Fig. 11.1 shows the subcircuit for Eq. (11.2) and the simulation results. The trained DC NN model is integrated with the subcircuit in SPICE simulations.  $T_0'(t)$  is self-consistently updated during the simulations and fed to the input of the NN model to replace  $T_0$ . By mapping the input temperature to different  $T_0'(t)$ , the model tracks the corresponding  $T_0$  and  $I_D$  determined by SH. In this case,  $T_0'(t)$  decreases at  $t=0$  since no heat has accumulated yet, and  $dT(t)$  is at its maximum. It eventually relaxes to the ambient condition to get the correct DC characteristics.

$$T_0'(t) = T_0 - \delta T(t) \quad (11.1)$$

$$\delta T_i = R_{TH} I_D V_{DS}, \delta T_{DC} = (R_{TH} - R'_{TH}) I_D V_{DS} \quad (11.2)$$

$$d(\delta T_i - \delta T(t))/dt = \frac{\delta T(t) - \delta T_{DC}}{\tau} \quad (11.3)$$

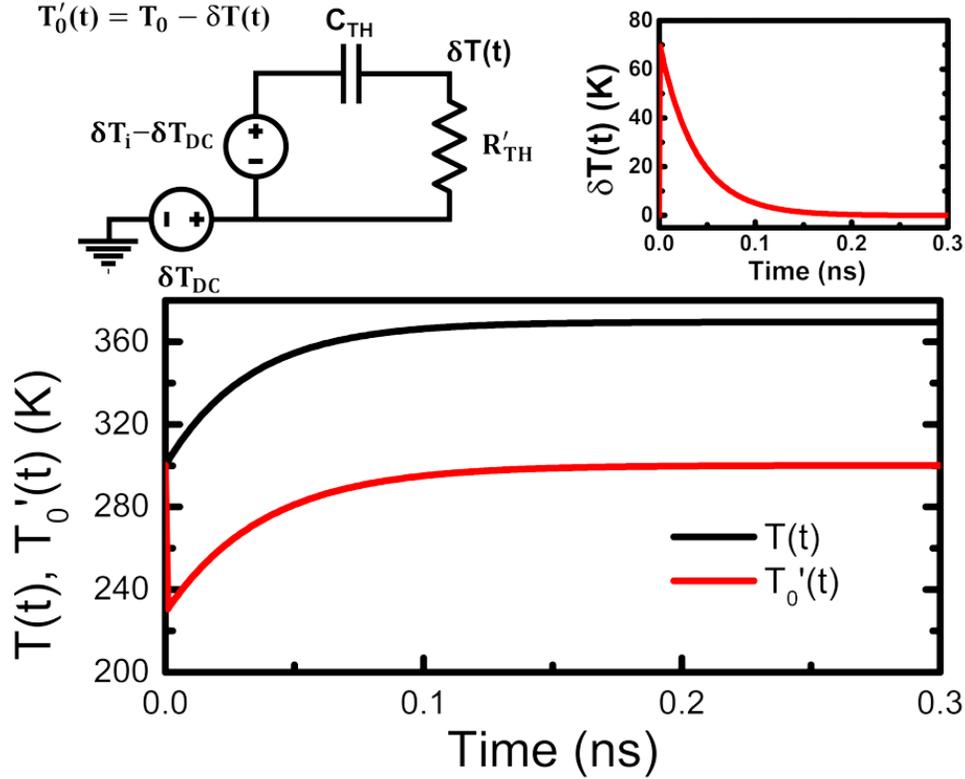


Fig. 11.1. The subcircuit of the temperature relaxation model and the simulated  $T(t)$ ,  $T'_0(t)$ , and  $dT(t)$ , from the traditional self-heating model and the temperature relaxation model. A step function voltage is applied. © 2025 IEEE

### 11.3 Validation

The model is trained using TensorFlow with the Adam optimizer and a learning rate of  $1E-3$ . Both IV and CV networks have two hidden layers with 30 and 20 neurons. It is then coded into Verilog-A for circuit simulations using HSPICE on an Intel Xeon Platinum 8260 workstation. A calibrated BSIM-CMG model card [142] is used to generate DC IV and CV data for training, validation, and testing. The model is trained from  $-50^\circ\text{C}$  to  $100^\circ\text{C}$ . Fig. 11.2 & 11.3 show the validation of  $I_D V_G$  and  $I_D V_D$  at different temperatures. It accurately models the temperature dependence of  $I_D$ ,  $g_m$ , and  $g_{ds}$  without using additional model equations. Fig. 11.4 shows the model fitting of  $C_{gg}$ ,  $C_{gs}$ ,  $C_{gd}$ , and  $C_{dd}$ . The model demonstrates excellent fitting results of these capacitances at a wide temperature range. Moreover, the  $125^\circ\text{C}$  IV and CV data is accurately modeled - showcasing the predictivity of the NN model. The

simulations for training are performed under DC conditions with self-heating in effect, and the model can generate characteristics without computing the self-heating subcircuit.

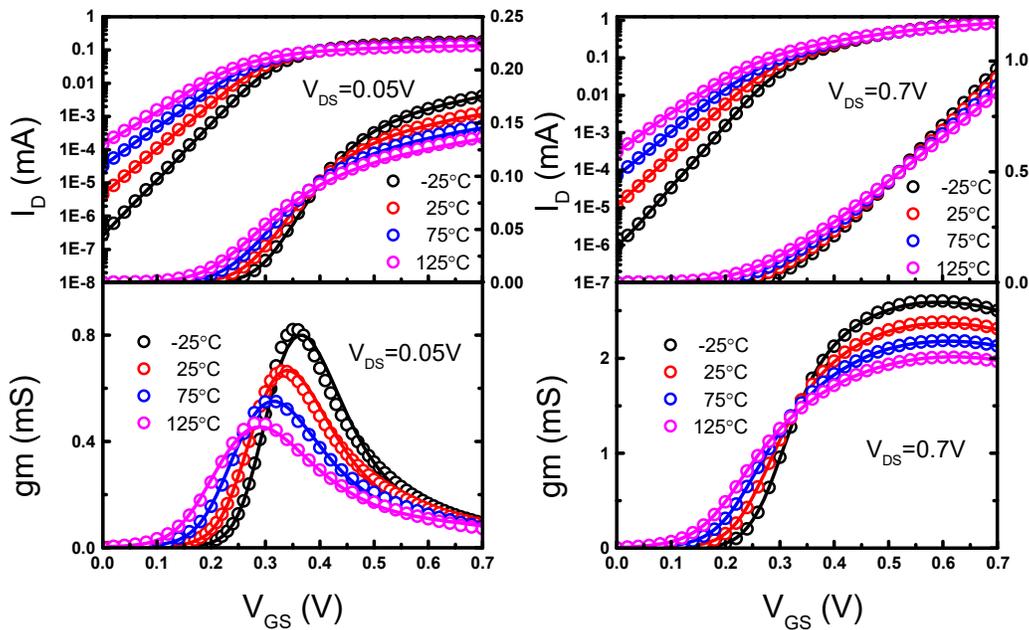


Fig. 11.2. Validation of transfer characteristics at different temperatures. The lines are the NN model, and the symbols are the BSIM-CMG data. © 2025 IEEE

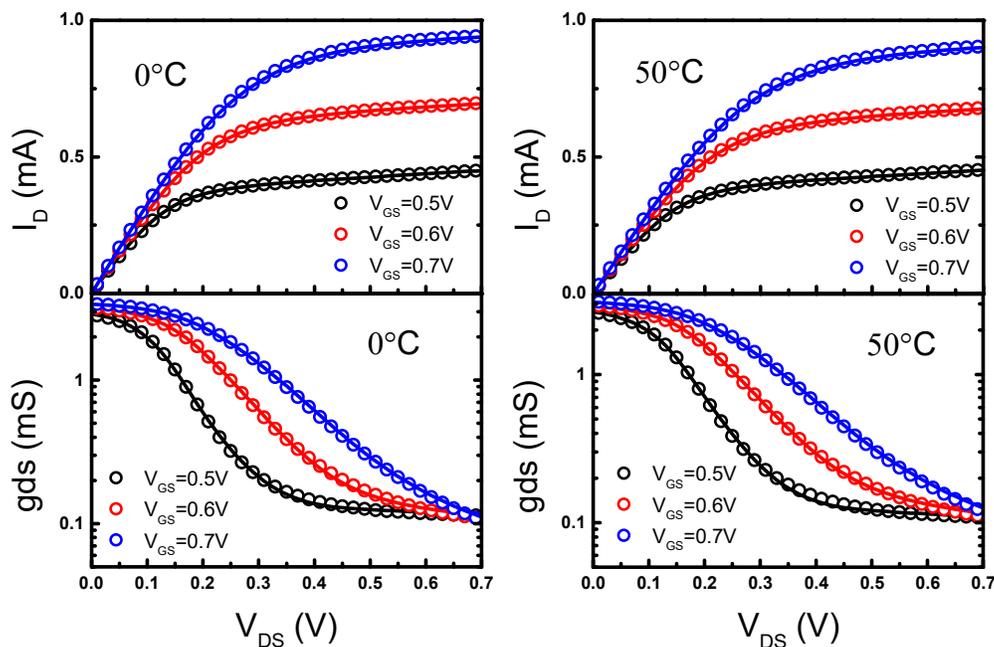


Fig. 11.3. Validation of output characteristics at different temperatures. The lines are the

NN model, and the symbols are the BSIM-CMG data. © 2025 IEEE

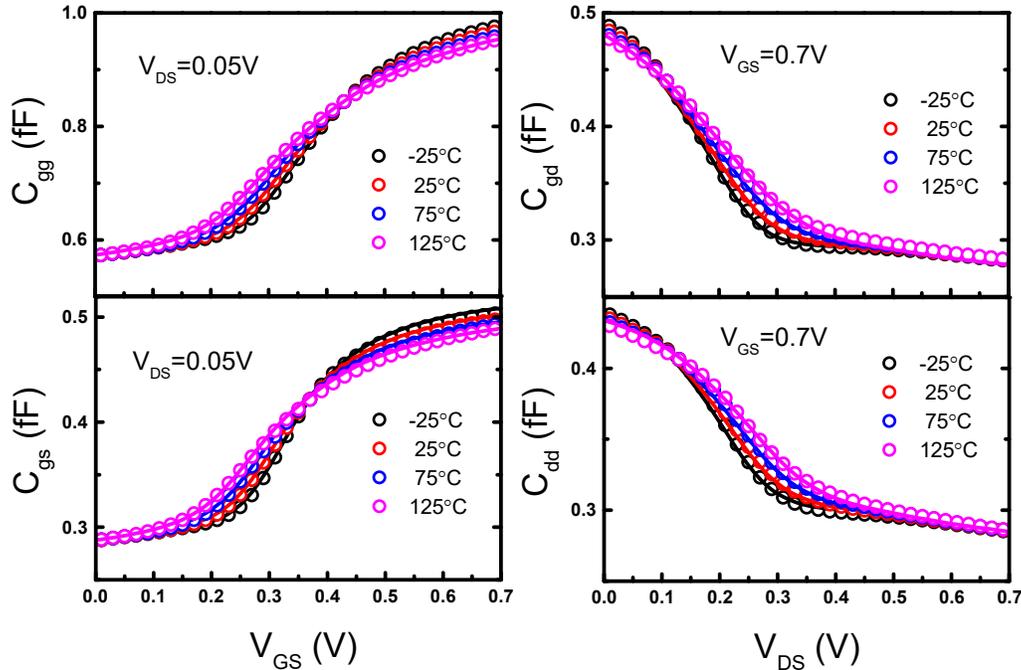


Fig. 11.4. Validation of CV characteristics at different temperatures. The lines are the NN model, and the symbols are the BSIM-CMG data. © 2025 IEEE

Then, to obtain the time-dependent self-heating characteristics, the temperature relaxation model and the subcircuit are used. In Fig. 11.5, I compare the NN model and the BSIM-CMG IV curves with the SH effect turned on and off. The proposed TR model can successfully recover the SH-free characteristics. I adjusted  $R_{TH}$ , which changes  $dT(0)$  to fit the SH-free data.  $T_0'$  is decreased and the device can be mapped to the higher current values corresponding to the case without SH. Finally, to demonstrate the proposed subcircuit for transient simulations, a pulse simulation is performed. Fig. 11.6 shows the pulse simulation results. A transistor is biased at  $V_{GS}=0.7V$  and  $V_{DS}$  pulses are applied with varying amplitudes. I show the simulation at 10 GHz, and the results generated by the NN model accurately match the output from BSIM-CMG. The  $C_{TH}$  is tuned so that the time constant of the NN model matches the testing data.  $R_{TH}$  is extracted from Fig. 11.5

which can also be obtained by fitting the peak current. With the TR model (red), the device current gradually decreases over time due to heat dissipation. Without the TR model (blue), the current will attain a steady state with no time delay. We can also use this model as an intermediate model to generate an SH-free I(T)/Q(T) model (Fig. 11.5) and combine it with the conventional SH model Eq. (11.4) (black solid). By tuning  $R'_{TH}$ , the model can capture SH and steady-state characteristics from the training data (black dash).

$$\Delta T(t) + \tau d\Delta T(t)/dt = R'_{TH} I_D V_{DS} \quad (11.4)$$

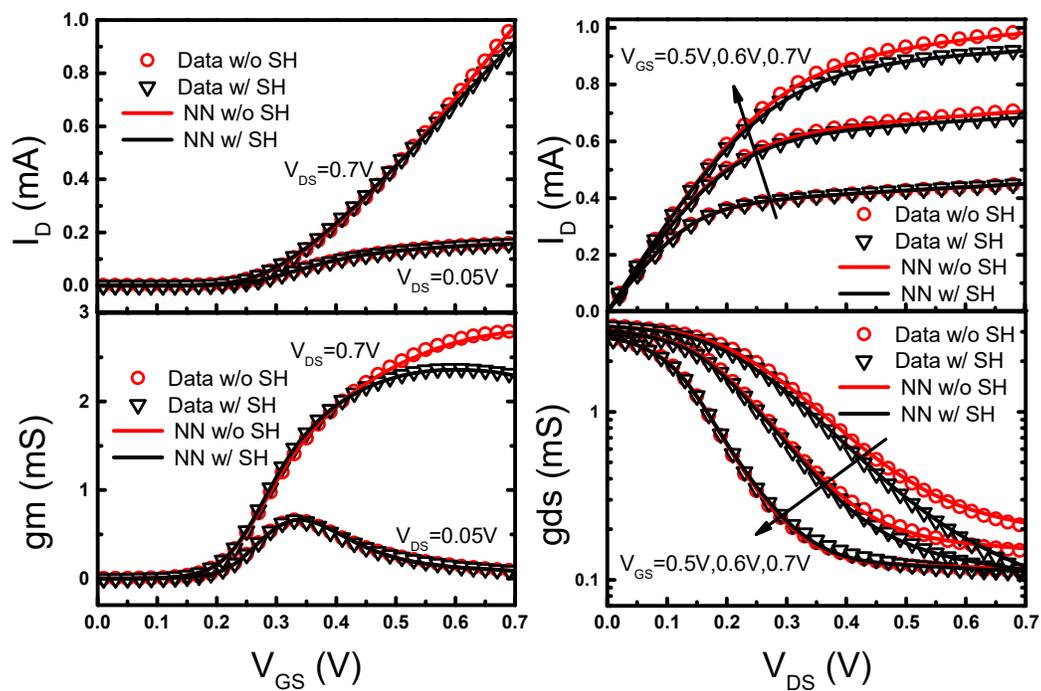


Fig. 11.5. Validation of the SH-free characteristics recovered by the temperature relaxation (TR) model. © 2025 IEEE

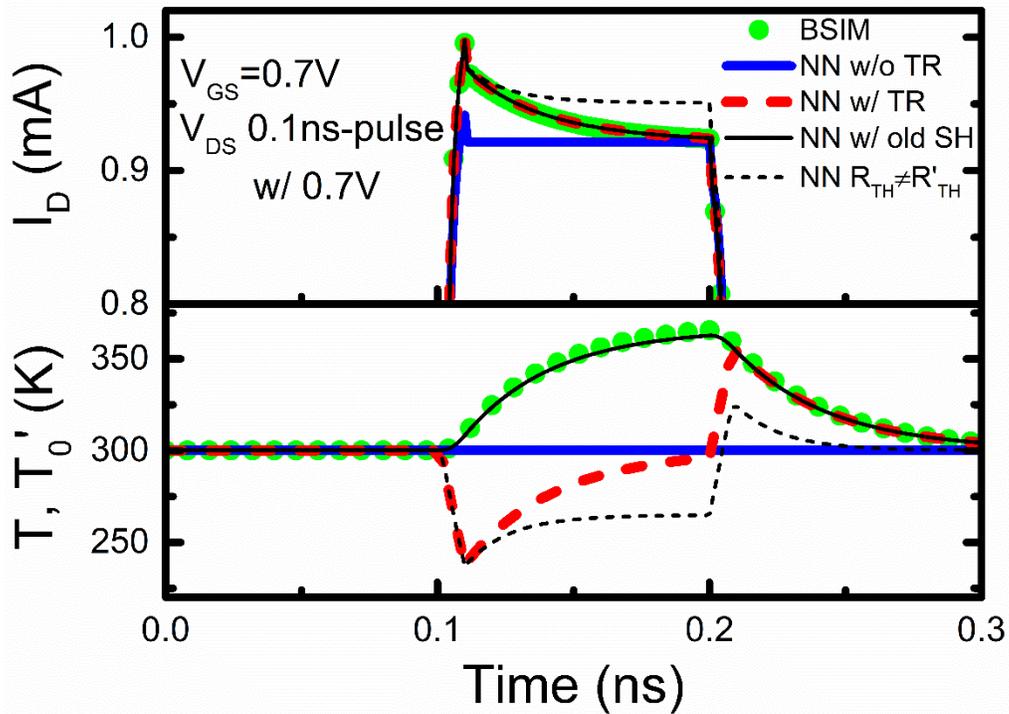


Fig. 11.6. Transient pulse simulations of BSIM-CMG, the NN model w/ and w/o TR model, the extracted SH-free NN with conventional SH model, and the case that  $R_{TH} \neq R'_{TH}$ . © 2025 IEEE

Finally, I test the model with a 1001-stage NAND oscillator (Fig. 11.7). With the TR model, the result shows a faster oscillation frequency compared to the case without TR, which matches our expectations.

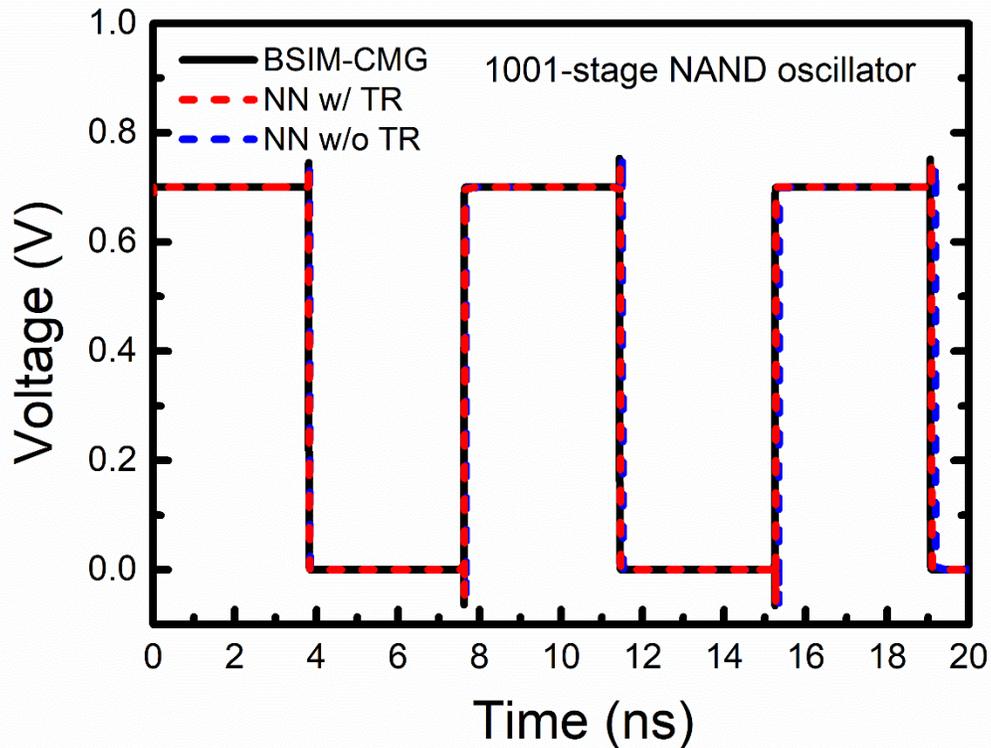


Fig. 11.7. 1001-stage NAND oscillator simulation.

## 11.4 Chapter Summary

A temperature-dependent NN transistor compact model with self-heating is proposed. It requires no predefined semi-empirical temperature functions and can model transistors across a wide temperature range. A new temperature relaxation modeling approach is implemented. The NN model can be trained using just DC characteristics without iterations in training or eliminating the SH effect during measurements. The model automatically generates DC IV and CV with the SH effect. The self-heating dynamic is modeled using the proposed temperature relaxation subcircuit with excellent accuracy. The proposed model further offers the flexibility to use the conventional self-heating subcircuit on an extracted SH-free NN model.

## Chapter 12 NeuroSpice: IC Simulator Using Physics-Informed Neural Network

In this chapter, I demonstrate the first physics-informed neural network for circuit simulation – NeuroSpice. NeuroSpice solves the differential-algebraic equations of circuits with machine learning. The differential equations are solved by minimizing the loss function residue through backpropagation. Unlike conventional SPICE using a time-discretized numerical solver, NeuroSpice solves the differential system with continuous analytical equations for the entire simulation period. I have shown that NeuroSpice can be used for circuit simulations from analog to digital applications and is also suitable for unconventional devices such as ferroelectric devices due to its flexibility.

### 12.1 Motivation

SPICE (Simulation Program with Integrated Circuit Emphasis) was developed at UC Berkeley as a general-purpose differential system solver [182]. SPICE and other simulation programs use numerical methods to solve discretized differential equations [182, 183]. As technology advanced, transistor counts have greatly increased on an IC chip. Systems across digital, analog, memory, and sensors are integrated. The growing scale of IC poses a challenge to existing circuit simulators in terms of speed and convergence.

On the other hand, artificial intelligence (AI) benefits from enhancing IC technology. AI or neural networks (NN) can learn the relationship between input and output data through training. However, NN can also be used without data using loss functions based on physics equations as a physics-informed neural network (PINN) [184, 185]. PINN solves the differential equations by minimizing the residual in the loss function. The differential equations define the loss function, and no data is needed. Several studies have demonstrated

PINN for applications such as power system simulation [186, 187] and transistor simulation [188]. So far, there is no study of PINN on IC simulations.

In this work, a PINN-based circuit simulator is proposed - NeuroSpice. NeuroSpice solves circuit differential-algebraic equations with a machine learning process. Instead of solving discretized differential algebraic equations (DAE) by time stepping, NeuroSpice finds approximate continuous analytical equations over the entire time frame. The continuous nature of PINN can potentially improve the convergence and speed of simulations. It also provides more flexibility in device modeling.

## 12.2 Framework

NeuroSpice is based on PINN as Fig. 12.1 shows. The input of NN is time ( $t$ ) and the outputs are the voltages and currents in the circuit. However, the input is not restricted to  $t$  and can be expanded to other physical quantities like space for Multiphysics systems beyond SPICE simulation. The loss function comprises circuit DAEs and the initial condition of the circuit nodes. The training optimizer will update the NN parameters from the loss. I implement the simulator with Pytorch, Adam optimizer, and Tanh activation function.

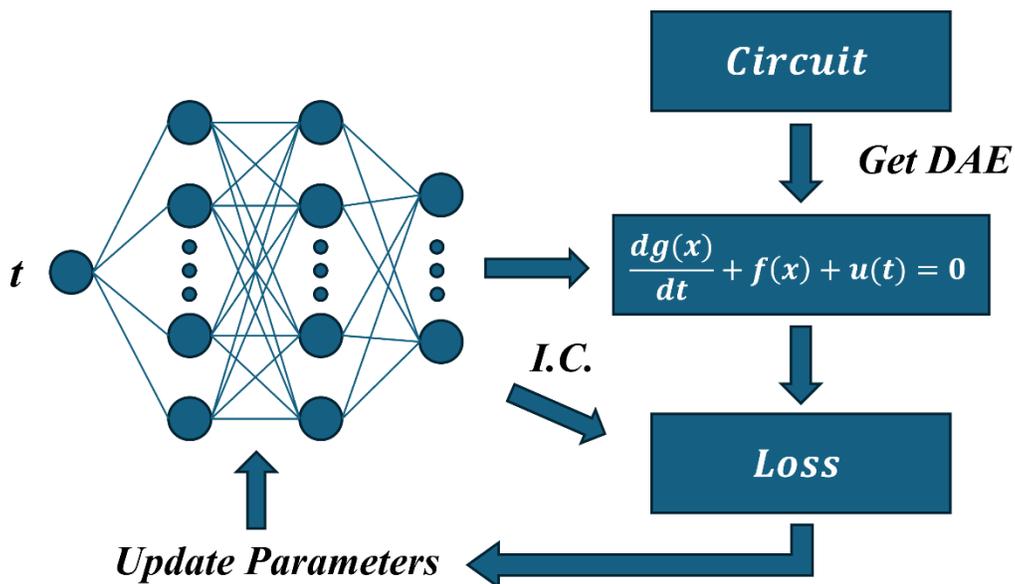


Fig. 12.1. The schematic diagram of NeuroSpice. The loss function is the DAE of the circuit and the initial condition. The neural network will update its parameters until the DAE is optimized to 0. I.C. is the initial condition.

### DAE of a Transistor Amplifier

$$L0 = \frac{v[0]}{R_S} + nmos(v[2], v[1], v[0])[2]$$

$$L1 = \frac{v[2] - V_{DD}}{R_D} + nmos(v[2], v[1], v[0])[0]$$

$$L2 = C_{in} \frac{d(v[1] - V_{IN})}{dt} + \frac{v[1]}{R_{G2}} + \frac{v[1] - V_{DD}}{R_{G1}} + nmos(v[2], v[1], v[0])[1]$$

$$DAE_{loss} = Mean(L0^2 + L1^2 + L2^2) \quad \frac{d}{dt} \text{ is computed by autograd}$$

$$Loss = \alpha \times IC_{loss} + \beta \times DAE_{loss}$$

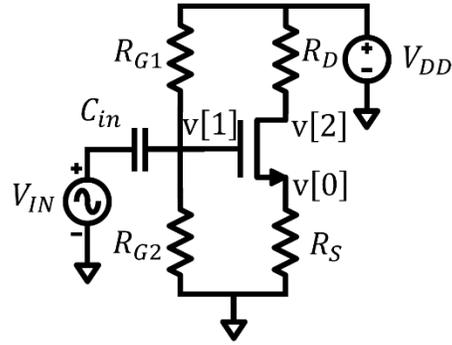


Fig. 12.2. Example DAE of a transistor amplifier and its loss function. In this case, the KCL of each node will be solved including the current of the nonlinear device model (nmos). The time derivative is computed by the autograd function in simulation.

### Device Model Pseudocode

```
class MOSFET:
```

```
    Parameter Initiatization: ...
```

```
    def current(vd, vg, vs):
```

```
        vgs = vg - vs
```

```
        vgd = vg - vd
```

```
        Id = f1(vgs, vds)
```

```
        Ig = f2(vgs, vds)
```

```
        Qg = g1(vgs, vds)
```

```
        Qd = g2(vgs, vds)
```

```
         $\frac{dQg}{dt} = \text{autograd}(Qg)$  Use autograd to compute analytical derivative
```

```
         $\frac{dQd}{dt} = \text{autograd}(Qd)$ 
```

```
        return [Id +  $\frac{dQd}{dt}$ , Ig +  $\frac{dQg}{dt}$ , -Id - Ig -  $\frac{dQd}{dt}$  -  $\frac{dQg}{dt}$ ]
```

Fig. 12.3. The pseudocode of the device model. The MOSFET DC and switching currents are computed and returned to the loss function. The charge currents are also analytically computed with autograd.

Fig. 12.2 shows an example DAE of a transistor amplifier. The DAE is formed by Kirchhoff's Current Law (KCL) for each node. The DAE involves a nonlinear device model (nmos) which is defined in Fig. 12.3. Typical device models such as BSIM (Berkeley short-channel IGFET model) [58, 142] are written in Verilog-A [189] and model users have limited control of the model evaluation in SPICE simulators. In NeuroSpice, the device model is written in Python where users can do quick prototyping. In the demonstrated model, inputs are the terminal voltages, and it returns the terminal currents. In NeuroSpice, device models can return values other than voltage and current going beyond the capability of SPICE device models. The time derivative comes from the capacitance in nmos and  $C_{in}$ . One advantage of NeuroSpice is that it uses analytical automatic differentiation (autograd) in machine learning programs such as Pytorch [190]. Therefore, there is no need for numerical differentiation. Finally, the loss is a weighted sum of DAE loss ( $DAE_{loss}$ ) and initial condition loss ( $IC_{loss}$ ) as shown in Fig. 12.2. Coefficients  $a$  and  $b$  are used to weigh these two losses so that PINN can learn to minimize both of them effectively [191]. Furthermore, between each DAE in  $DAE_{loss}$ , weights can also be applied to improve learning.

## 12.3 Result & Discussion

I test NeuroSpice functionality by simulating several circuit examples. The first is a series RLC circuit shown in Fig. 12.4. A sinusoid voltage source is applied. The simulated current ( $I_L$ ) is compared with the analytical solution derived from Laplace transform. NeuroSpice provides excellent matches. I solve the RLC with only the 1<sup>st</sup>-order derivative, which is more general in circuit analysis. In Fig. 12.5, I demonstrate a non-inverting amplifier circuit. The operational amplifier (OPAMP) has a gain of 1000 and feedback is connected to its negative terminal. The large gain of the OPAMP can cause the NN to diverge, but I adjust the weights in the loss function to prevent divergence.

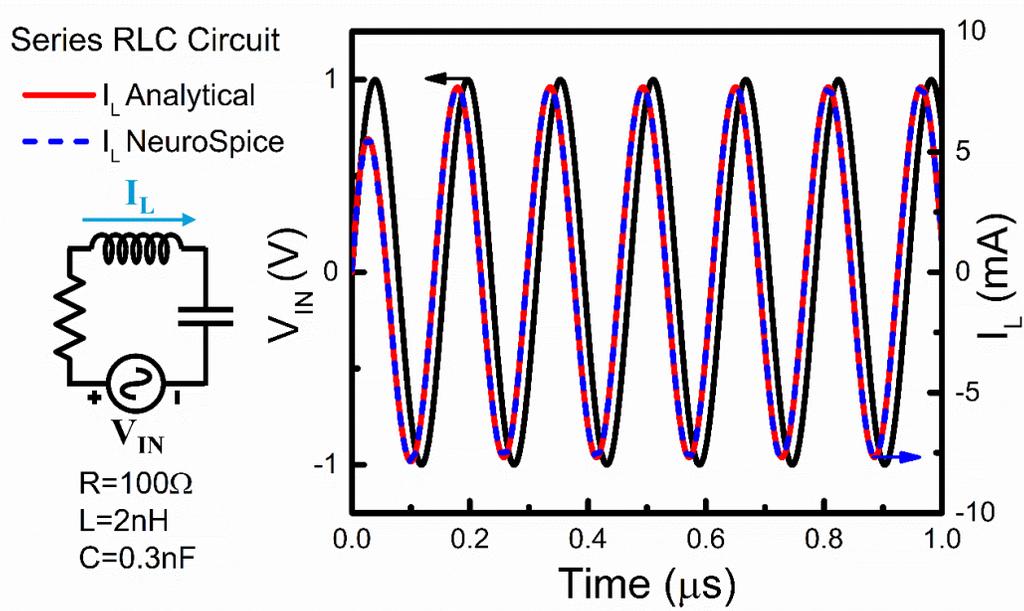


Fig. 12.4. The simulation of a series RLC circuit with a sinusoid input.

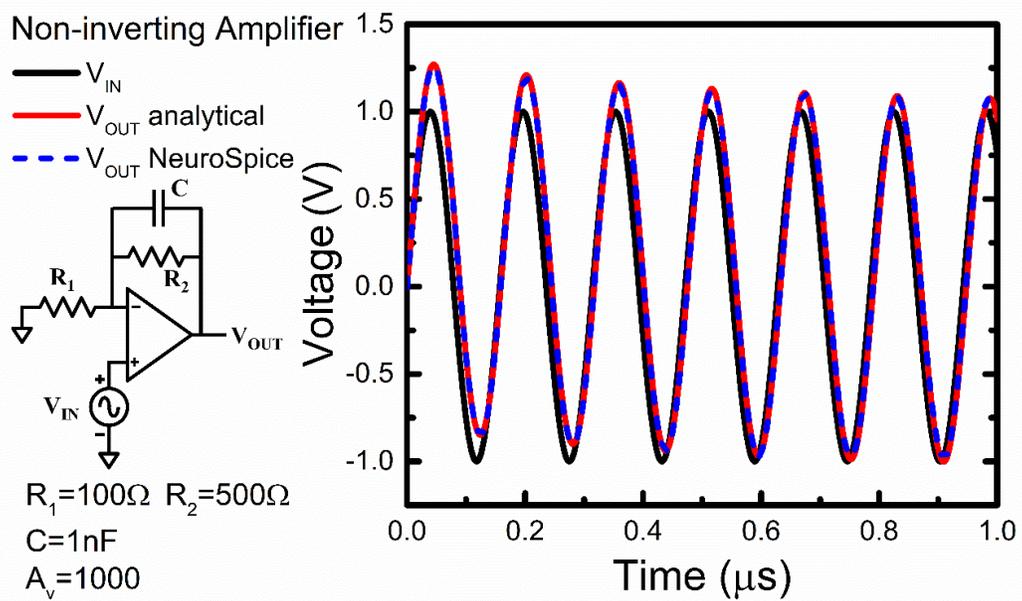


Fig. 12.5. The simulation of a non-inverting amplifier implemented with an operational amplifier circuit.

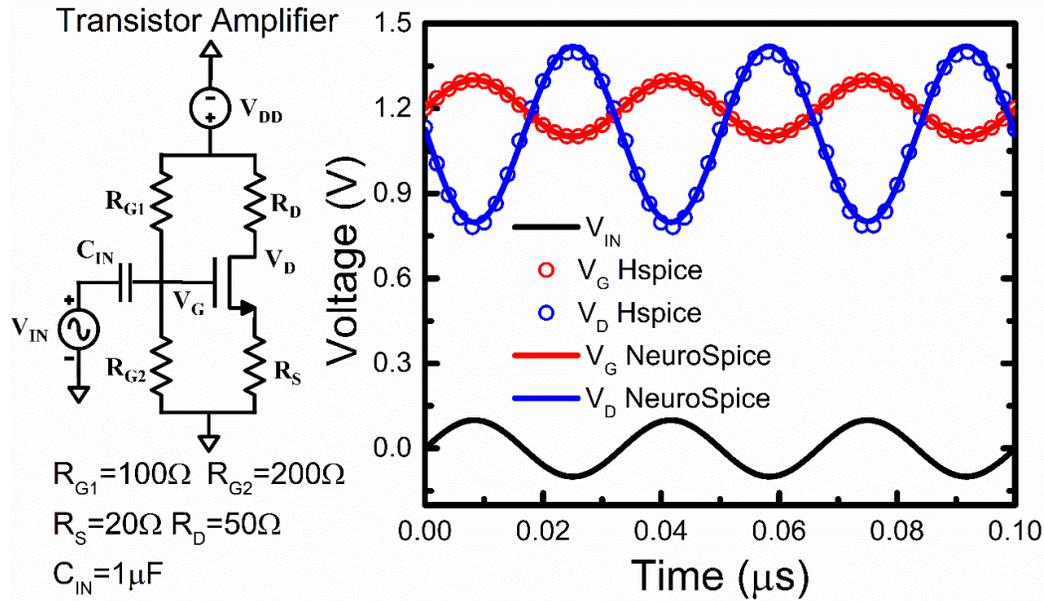


Fig. 12.6. The simulation of a transistor amplifier.

To test NeuroSpice's ability to solve nonlinear circuit elements, I simulate several example circuits using a simple MOSFET model. The simulation result of the transistor amplifier in Fig. 12.2 is shown in Fig. 12.6. Since no analytical solution is available for this circuit, Hspice is used to compare the result with NeuroSpice. The output voltage  $V_D$  and input voltage  $V_G$  from NeuroSpice match with Hspice. Next, digital circuits such as CMOS inverter and NAND are examined. These circuits involve more than one device model and often experience abrupt changes in voltages. In our study, it is important to increase the number of batches in the simulation to generate correct results. In Fig. 12.7 & 12.8, simulated  $V_{OUT}$  of CMOS and NAND are compared between NeuroSpice and Hspice showing perfect matching. Fig. 12.9 shows the test of a more complex 5-stage ring oscillator. The result indicates that NeuroSpice can simulate unstable systems like oscillators and solve DAE involving multiple nonlinear models.

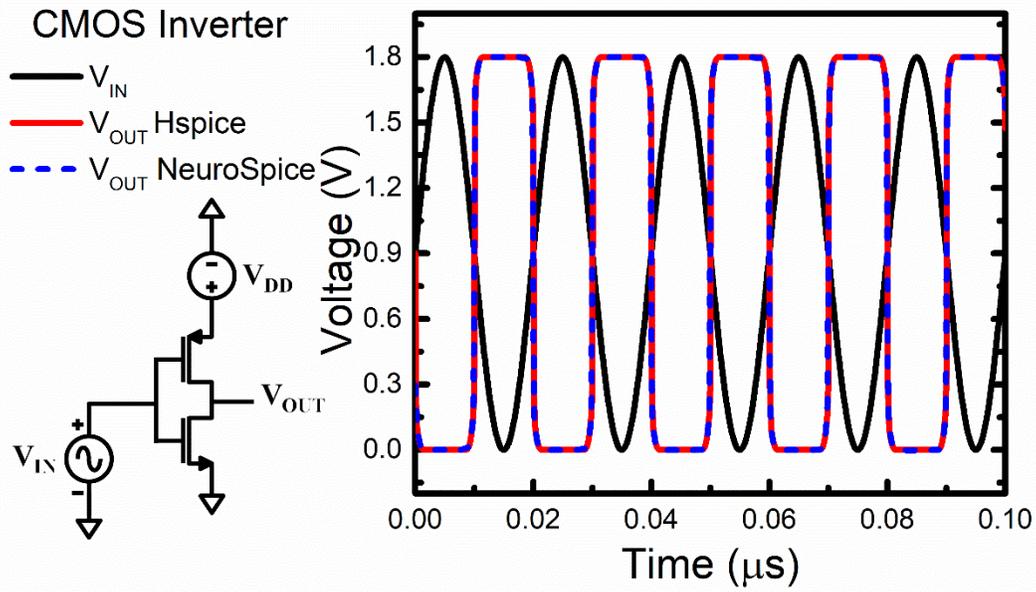


Fig. 12.7. The simulation of a CMOS inverter.

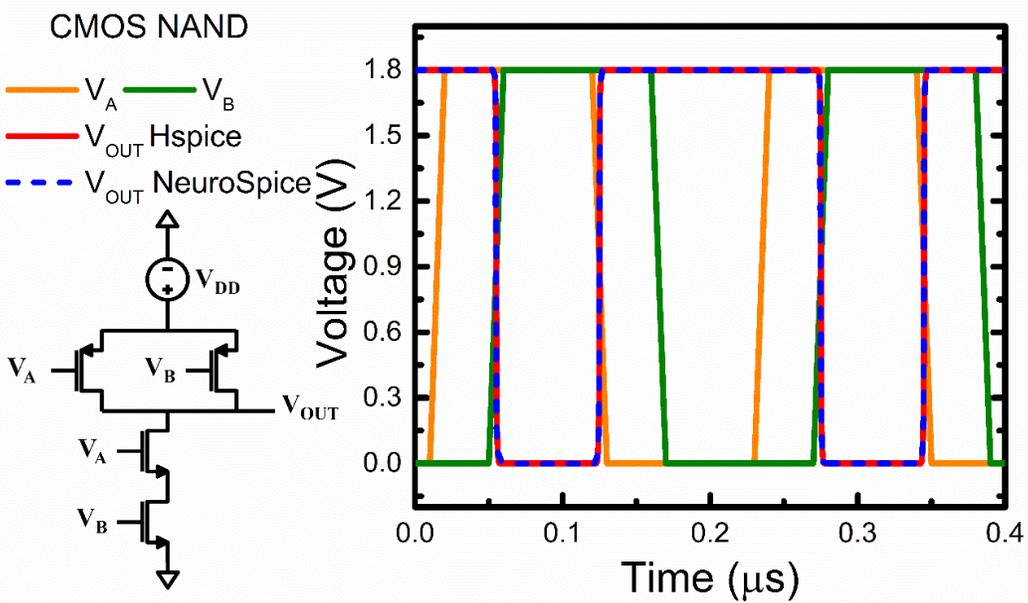


Fig. 12.8. The simulation of a CMOS NAND.

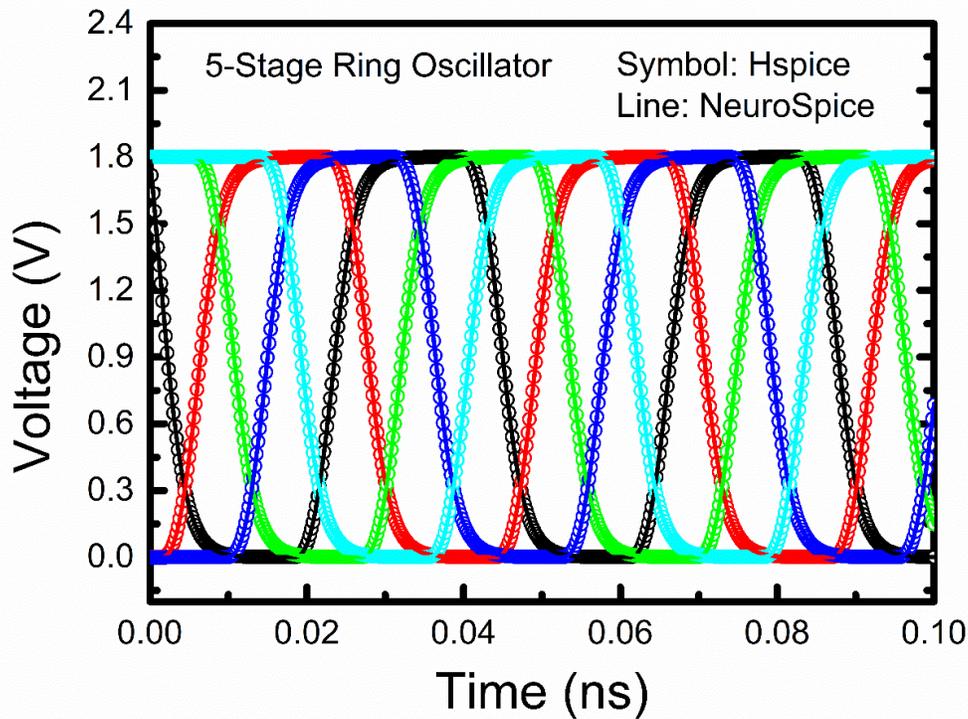


Fig. 12.9. The simulation of a 5-stage ring oscillator.

NeuroSpice is also a suitable platform for new devices and systems due to its high flexibility. For example, ferroelectric (FE) devices are a potential solution for future low-power in-memory computing [38]. A common model to describe FE material is the Landau-Khalatnikov (LK) model [71]. The LK model is a highly nonlinear model with positive, negative, and infinite slopes as shown in Fig. 12.10. Fig. 12.11 demonstrates this flexibility by simulating a FeRAM using the LK model and MOSFET model with NeuroSpice. NeuroSpice successfully models the voltage drop caused by the polarization switching in the FE capacitor.

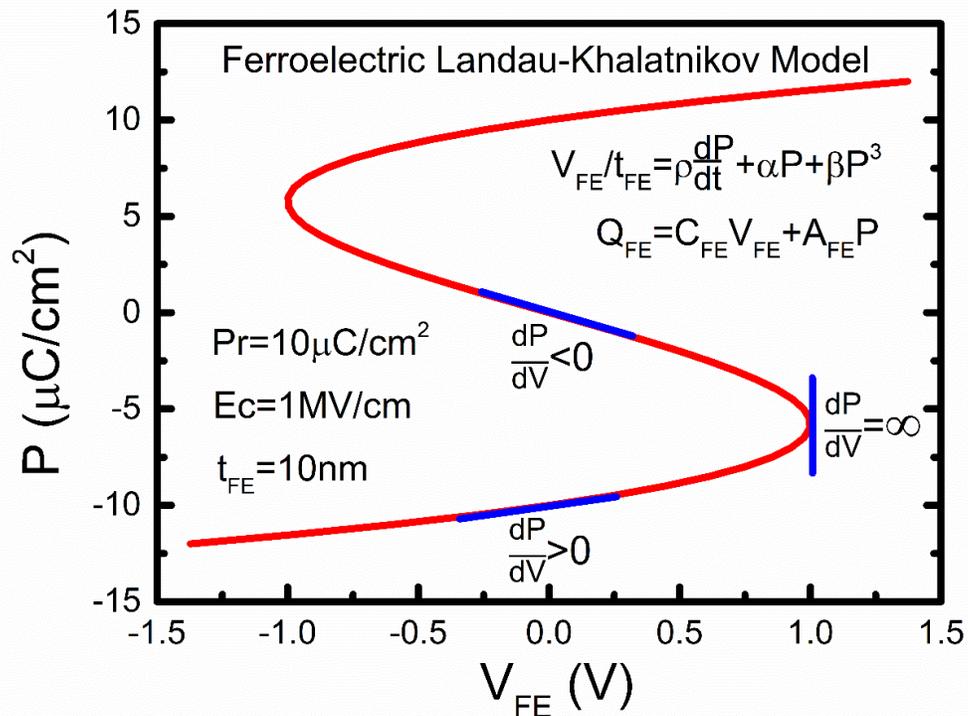


Fig. 12.10. Polarization-voltage characteristics of a ferroelectric Landau-Khalatnikov (LK) model. It contains positive, negative, and infinite slopes.

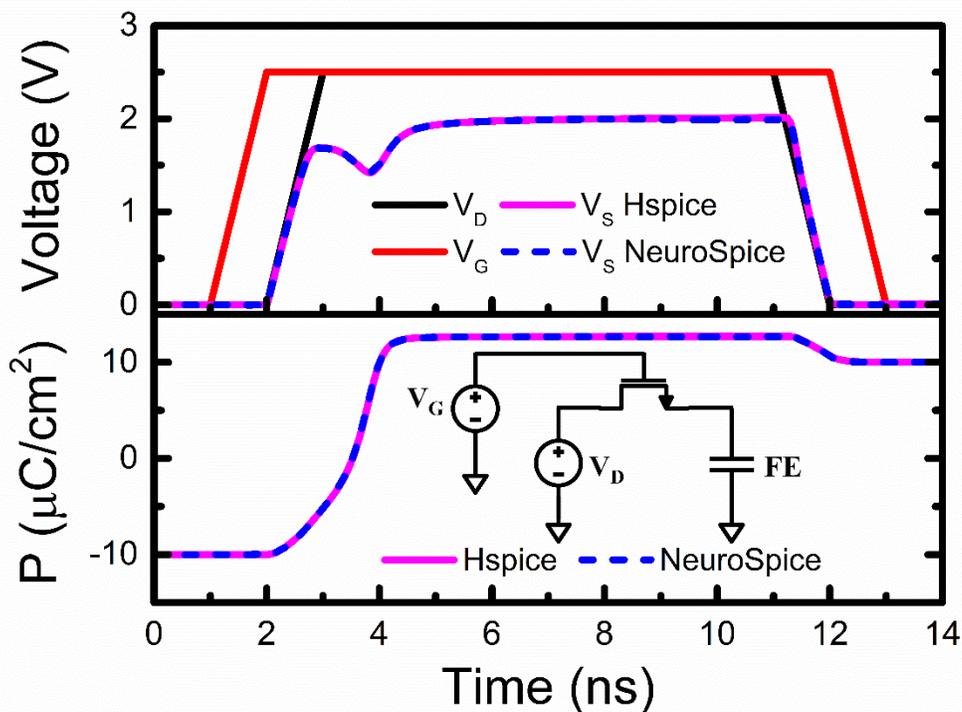


Fig. 12.11. The simulation of a FeRAM using the LK model.

## 12.4 Chapter Summary

NeuroSpice is a circuit simulator based on PINN, which goes beyond the numerical method in conventional SPICE. It can solve circuit systems with complex nonlinear devices and be applied to Multiphysics systems. The inherent differentiable nature of PINN makes it suitable for circuit inverse problems and design optimizations. Further studies are needed for network optimization, training convergence, and acceleration.

## Chapter 13 Summary

### 13.1 Chapter Summary

As technology keeps advancing and the growing demand for AI, future integrated circuits require trillions of transistors. The energy consumption of a chip becomes tremendous. The need of new devices, new architecture, and new computer-aid design are crucial. In the first part of this dissertation, I discuss the compact models of next-generation memory devices. These emerging memories can be served as in-memory computing and neuromorphic computing devices. From Chapters 2 to 6, I study the switching dynamics of ferroelectric and antiferroelectric devices including capacitors, tunnel junction memristors, and transistors. The developed model can accurately describe the multi-domain dynamics of ferroelectric devices with multi-level hysteresis which can be used for analog memory and neuromorphic circuit designs.

Other than ferroelectric devices, resistive memory (RRAM) and magnetic memory (MRAM) are also strong candidates for future computing. In Chapter 7, I demonstrate an RRAM compact model that unifies different RRAM material systems with a single set of equations. The model can simulate both metal oxide RAM (OxRAM) and conducting bridge RAM (CBRAM) and their applications for multi-level memory cells. In Chapter 8, I develop an STT-MRAM model that uses a simplified 1D Landau–Lifshitz–Gilbert (LLG) equation. The model keeps the physical predictability and is efficient in circuit simulations. A compact model for switching probability is also developed to provide fast estimation for the write error rate in the memory array without statistical simulations.

The second half of my dissertation discusses how to improve current design systems with AI. From Chapters 9 to 11, I develop a neural network-

based transistor compact model – BSIM-NN. The model includes the complete IV and CV for advanced transistors such as FinFET and GAA. The model is generic and can be simply applied to emerging technology like NCFET to enable design exploration for future devices. The model has non-quasi-static and self-heating effects for high-speed circuit simulations. Various circuits from digital, and analog to RF ICs are tested with accurate results. BSIM-NN can accelerate the IC simulation by 5 times compared to the industry standard model.

In Chapter 12, I proposed a new way to do circuit simulation with the physics-informed neural network - NeuroSpice. This method shows an alternative way to improve the convergence and speed up the IC simulation.

## **13.2 Future Work**

Compact models are the foundation of the IC design industry. With accurate and efficient compact models for future devices, designers can develop next-generation ICs with new devices. One of the future directions is using these emerging memory models to design in-memory computing and neuromorphic computing chips. The other future direction is expanding BSIM-NN with more functions including global length and width scaling, cryogenic operations, aging, and layout dependence effect. We can also study how to use machine learning and neural networks for quantum devices and silicon photonic devices. It will also be interesting to continue developing NeuroSpice to reach a real-application level.

## Bibliography

- [1] S. Salahuddin, K. Ni, and S. Datta, "The era of hyper-scaling in electronics," *Nature Electronics*, vol. 1, no. 8, pp. 442-450, 2018/08/01 2018, doi: 10.1038/s41928-018-0117-x.
- [2] Y. S. Chauhan *et al.*, "Chapter 1 - 3D thin-body FinFET and GAA: From device to compact model," in *FinFET/ GAA Modeling for IC Simulation and Design (Second Edition)*, Y. S. Chauhan *et al.* Eds.: Academic Press, 2024, pp. 1-13.
- [3] C. T. Tung, G. Pahwa, S. Salahuddin, and C. Hu, "A Compact Model of Polycrystalline Ferroelectric Capacitor," *IEEE Transactions on Electron Devices*, vol. 68, no. 10, pp. 5311-5314, 2021, doi: 10.1109/TED.2021.3100814.
- [4] T. S. Böске, J. Müller, D. Bräuhaus, U. Schröder, and U. Böttger, "Ferroelectricity in hafnium oxide thin films," *Applied Physics Letters*, vol. 99, no. 10, 2011, doi: 10.1063/1.3634052.
- [5] T. S. Böске, J. Müller, D. Bräuhaus, U. Schröder, and U. Böttger, "Ferroelectricity in hafnium oxide CMOS compatible ferroelectric field effect transistors," presented at the 2011 International Electron Devices Meeting, Washington, DC, USA, Dec., 2011.
- [6] A. I. Khan, A. Keshavarzi, and S. Datta, "The future of ferroelectric field-effect transistor technology," *Nature Electronics*, vol. 3, no. 10, pp. 588-597, 2020/10/01 2020, doi: 10.1038/s41928-020-00492-7.
- [7] A. Keshavarzi, K. Ni, W. Van Den Hoek, S. Datta, and A. Raychowdhury, "FerroElectronics for Edge Intelligence," *IEEE Micro*, vol. 40, no. 6, pp. 33-48, 1 Nov.-Dec. 2020, doi: 10.1109/mm.2020.3026667.
- [8] E. W. Kinder, C. Alessandri, P. Pandey, G. Karbasian, S. Salahuddin, and A. Seabaugh, "Partial Switching of Ferroelectrics for Synaptic Weight Storage," presented at the 2017 75th Annual Device Research Conference (DRC), South Bend, IN, USA, June, 2017.
- [9] Y. Ishibashi and Y. Takagi, "Note on Ferroelectric Domain Switching," *Journal of the Physical Society of Japan*, vol. 31, no. 2, pp. 506-510, 1971/08/15 1971, doi: 10.1143/JPSJ.31.506.
- [10] M. Avrami, "Kinetics of Phase Change. I General Theory," *The Journal of Chemical Physics*, vol. 7, no. 12, pp. 1103-1112, 1939, doi: 10.1063/1.1750380.
- [11] I. Stolichnov, L. Malin, E. Colla, A. K. Tagantsev, and N. Setter, "Microscopic aspects of the region-by-region polarization reversal kinetics of polycrystalline ferroelectric Pb(Zr,Ti)O<sub>3</sub> films," *Applied Physics Letters*, vol. 86, no. 1, 2005, doi: 10.1063/1.1845573.
- [12] A. K. Tagantsev, I. Stolichnov, N. Setter, J. S. Cross, and M. Tsukada, "Non-Kolmogorov-Avrami switching kinetics in ferroelectric thin films," *Physical*

- Review B*, vol. 66, no. 21, Dec. 2002, doi: 10.1103/PhysRevB.66.214109.
- [13] X. J. Lou, "Statistical switching kinetics of ferroelectrics," *J Phys Condens Matter*, vol. 21, no. 1, p. 012207, Jan 7 2009, doi: 10.1088/0953-8984/21/1/012207.
- [14] C. Alessandri, P. Pandey, A. Abusleme, and A. Seabaugh, "Switching Dynamics of Ferroelectric Zr-Doped HfO<sub>2</sub>," *IEEE Electron Device Letters*, vol. 39, no. 11, pp. 1780-1783, Nov. 2018, doi: 10.1109/led.2018.2872124.
- [15] N. Gong, X. Sun, H. Jiang, K. S. Chang-Liao, Q. Xia, and T. P. Ma, "Nucleation limited switching (NLS) model for HfO<sub>2</sub>-based metal-ferroelectric-metal (MFM) capacitors: Switching kinetics and retention characteristics," *Applied Physics Letters*, vol. 112, no. 26, 2018, doi: 10.1063/1.5010207.
- [16] S. D. Hyun *et al.*, "Dispersion in Ferroelectric Switching Performance of Polycrystalline Hf<sub>0.5</sub>Zr<sub>0.5</sub>O<sub>2</sub> Thin Films," *ACS Appl Mater Interfaces*, vol. 10, no. 41, pp. 35374-35384, Oct. 17 2018, doi: 10.1021/acsami.8b13173.
- [17] S. Zhukov, Y. A. Genenko, and H. von Seggern, "Experimental and theoretical investigation on polarization reversal in unfatigued lead-zirconate-titanate ceramic," *Journal of Applied Physics*, vol. 108, no. 1, 2010, doi: 10.1063/1.3380844.
- [18] H. Mulaosmanovic, T. Mikolajick, and S. Slesazeck, "Accumulative Polarization Reversal in Nanoscale Ferroelectric Transistors," *ACS Appl Mater Interfaces*, vol. 10, no. 28, pp. 23997-24002, Jul 18 2018, doi: 10.1021/acsami.8b08967.
- [19] K. Ni, M. Jerry, J. A. Smith, and S. Datta, "A Circuit Compatible Accurate Compact Model for Ferroelectric-FETs," presented at the 2018 IEEE Symposium on VLSI Technology, Honolulu, HI, USA, June, 2018.
- [20] A. K. Saha, K. Ni, S. Dutta, S. Datta, and S. Gupta, "Phase field modeling of domain dynamics and polarization accumulation in ferroelectric HZO," *Applied Physics Letters*, vol. 114, no. 20, 2019, doi: 10.1063/1.5092707.
- [21] C. Alessandri, P. Pandey, and A. C. Seabaugh, "Experimentally Validated, Predictive Monte Carlo Modeling of Ferroelectric Dynamics and Variability," presented at the 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, Dec., 2018.
- [22] C. Alessandri, P. Pandey, A. Abusleme, and A. Seabaugh, "Monte Carlo Simulation of Switching Dynamics in Polycrystalline Ferroelectric Capacitors," *IEEE Transactions on Electron Devices*, vol. 66, no. 8, pp. 3527-3534, Aug. 2019, doi: 10.1109/ted.2019.2922268.
- [23] S. Deng *et al.*, "A Comprehensive Model for Ferroelectric FET Capturing the Key Behaviors Scalability, Variation, Stochasticity, and Accumulation," presented at the 2020 IEEE Symposium on VLSI Technology, Honolulu, HI, USA, June, 2020.
- [24] C.-T. Tung, G. Pahwa, S. Salahuddin, and C. Hu, "A Compact Model of Metal–Ferroelectric–Insulator–Semiconductor Tunnel Junction," *IEEE Transactions on Electron Devices*, vol. 69, no. 1, pp. 414-418, 2022, doi: 10.1109/ted.2021.3130857.
- [25] L. Esaki, B. B. Laibowitz, and P. J. Stiles, "Polar switch," *IBM Tech. Discl. Bull.*, vol. 13, no. 8, p. 2161, 1971.
- [26] A. Chanthbouala *et al.*, "Solid-state memories based on ferroelectric tunnel junctions," *Nat Nanotechnol*, vol. 7, no. 2, pp. 101-4, Dec 4 2011, doi: 10.1038/nnano.2011.213.

- [27] M. Y. Zhuravlev, R. F. Sabirianov, S. S. Jaswal, and E. Y. Tsymlal, "Giant Electroresistance in Ferroelectric Tunnel Junctions," *Physical Review Letters*, vol. 94, no. 24, 2005, doi: 10.1103/PhysRevLett.94.246802.
- [28] J. Yoon, S. Hong, Y. W. Song, J.-H. Ahn, and S.-E. Ahn, "Understanding tunneling electroresistance effect through potential profile in Pt/Hf<sub>0.5</sub>Zr<sub>0.5</sub>O<sub>2</sub>/TiN ferroelectric tunnel junction memory," *Applied Physics Letters*, vol. 115, no. 15, 2019, doi: 10.1063/1.5119948.
- [29] M. Kobayashi, Y. Tagawa, F. Mo, T. Saraya, and T. Hiramoto, "Ferroelectric HfO<sub>2</sub> Tunnel Junction Memory With High TER and Multi-Level Operation Featuring Metal Replacement Process," *IEEE Journal of the Electron Devices Society*, vol. 7, pp. 134-139, 2019, doi: 10.1109/jeds.2018.2885932.
- [30] S. S. Cheema *et al.*, "Enhanced ferroelectricity in ultrathin films grown directly on silicon," *Nature*, vol. 580, no. 7804, pp. 478-482, Apr 2020, doi: 10.1038/s41586-020-2208-x.
- [31] Z. Wang *et al.*, "Compact modelling of ferroelectric tunnel memristor and its use for neuromorphic simulation," *Applied Physics Letters*, vol. 104, no. 5, 2014, doi: 10.1063/1.4864270.
- [32] D. Pantel and M. Alexe, "Electroresistance effects in ferroelectric tunnel barriers," *Physical Review B*, vol. 82, no. 13, 2010, doi: 10.1103/PhysRevB.82.134105.
- [33] J. C. Ranuáñez, M. J. Deen, and C.-H. Chen, "A review of gate tunneling current in MOS devices," *Microelectronics Reliability*, vol. 46, no. 12, pp. 1939-1956, 2006, doi: 10.1016/j.microrel.2005.12.006.
- [34] F. Mo, Y. Tagawa, T. Saraya, T. Hiramoto, and M. Kobayashi, "Scalability Study on Ferroelectric-HfO<sub>2</sub> Tunnel Junction Memory Based on Non-equilibrium Green Function Method with Self-consistent Potential," presented at the 2018 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2018.
- [35] P. Chang, G. Du, J. Kang, and X. Liu, "Conduction Mechanisms of Metal-Ferroelectric- Insulator-Semiconductor Tunnel Junction on N- and P-Type Semiconductor," *IEEE Electron Device Letters*, vol. 42, no. 1, pp. 118-121, 2021, doi: 10.1109/led.2020.3041515.
- [36] H.-H. Huang *et al.*, "A Comprehensive Modeling Framework for Ferroelectric Tunnel Junctions," presented at the IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2019.
- [37] P. Chang, G. Du, J. Kang, and X. Liu, "Guidelines for Ferroelectric-Semiconductor Tunnel Junction Optimization by Band Structure Engineering," *IEEE Transactions on Electron Devices*, vol. 68, no. 7, pp. 3526-3531, 2021, doi: 10.1109/ted.2021.3079881.
- [38] C. T. Tung, G. Pahwa, S. Salahuddin, and C. Hu, "A Compact Model of Polycrystalline Ferroelectric Capacitor," *IEEE Transactions on Electron Devices*, pp. 1-4, 2021, doi: 10.1109/TED.2021.3100814.
- [39] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, 2 ed. Cambridge: Cambridge University Press, 2009.
- [40] T. R. Scavo and J. B. Thoo, "On the Geometry of Halley's Method," *The American Mathematical Monthly*, vol. 102, no. 5, pp. 417-426, 1995, doi: 10.2307/2975033.
- [41] A. Gruverman *et al.*, "Tunneling Electroresistance Effect in Ferroelectric Tunnel Junctions at the Nanoscale," *Nano Letters*, vol. 9, no. 10, pp. 3539-3543,

- 2009/10/14 2009, doi: 10.1021/nl901754t.
- [42] P. Zhang, "Scaling for quantum tunneling current in nano- and subnano-scale plasmonic junctions," *Sci Rep*, vol. 5, p. 9826, May 19 2015, doi: 10.1038/srep09826.
- [43] S. Banerjee and P. Zhang, "A generalized self-consistent model for quantum tunneling current in dissimilar metal-insulator-metal junction," *AIP Advances*, vol. 9, no. 8, 2019, doi: 10.1063/1.5116204.
- [44] C. T. Tung, G. Pahwa, S. Salahuddin, and C. Hu, "A Compact Model of Ferroelectric Field-Effect Transistor," *IEEE Electron Device Letters*, vol. 43, no. 8, pp. 1363-1366, 2022, doi: 10.1109/LED.2022.3182141.
- [45] S. Dünkler *et al.*, "A FeFET based super-low-power ultra-fast embedded NVM technology for 22nm FDSOI and beyond," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2-6 Dec. 2017 2017, pp. 19.7.1-19.7.4, doi: 10.1109/IEDM.2017.8268425.
- [46] A. J. Tan *et al.*, "Ferroelectric HfO<sub>2</sub> Memory Transistors With High- $\kappa$  Interfacial Layer and Write Endurance Exceeding 10<sup>10</sup> Cycles," *IEEE Electron Device Letters*, vol. 42, no. 7, pp. 994-997, 2021, doi: 10.1109/led.2021.3083219.
- [47] H. Mulaosmanovic *et al.*, "Switching Kinetics in Nanoscale Hafnium Oxide Based Ferroelectric Field-Effect Transistors," *ACS Appl Mater Interfaces*, vol. 9, no. 4, pp. 3792-3798, Feb. 1 2017, doi: 10.1021/acsami.6b13866.
- [48] A. K. Saha and S. K. Gupta, "Modeling and Comparative Analysis of Hysteretic Ferroelectric and Anti-ferroelectric FETs," presented at the 2018 76th Device Research Conference (DRC), Santa Barbara, CA, USA, 2018.
- [49] Z. Wang *et al.*, "Depolarization Field Induced Instability of Polarization States in HfO<sub>2</sub> Based Ferroelectric FET," presented at the 2020 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2020.
- [50] Y. Xiang *et al.*, "Compact Modeling of Multidomain Ferroelectric FETs: Charge Trapping, Channel Percolation, and Nucleation-Growth Domain Dynamics," *IEEE Transactions on Electron Devices*, pp. 1-9, 2021, doi: 10.1109/ted.2021.3049761.
- [51] H. Mulaosmanovic *et al.*, "Impact of Read Operation on the Performance of HfO<sub>2</sub>-Based Ferroelectric FETs," *IEEE Electron Device Letters*, vol. 41, no. 9, pp. 1420-1423, 2020, doi: 10.1109/led.2020.3007220.
- [52] Z. Wang *et al.*, "Cryogenic characterization of a ferroelectric field-effect-transistor," *Applied Physics Letters*, vol. 116, no. 4, 2020, doi: 10.1063/1.5129692.
- [53] M. N. K. Alam *et al.*, "On the Characterization and Separation of Trapping and Ferroelectric Behavior in HfZrO FET," *IEEE Journal of the Electron Devices Society*, vol. 7, pp. 855-862, 2019, doi: 10.1109/jeds.2019.2902953.
- [54] Y. Xiang *et al.*, "Implication of Channel Percolation in Ferroelectric FETs for Threshold Voltage Shift Modeling," presented at the 2020 IEEE International Electron Devices Meeting (IEDM), 2020.
- [55] Y. Xiang *et al.*, "Physical Insights on Steep Slope FEFETs including Nucleation-Propagation and Charge Trapping," in *2019 IEEE International Electron Devices Meeting (IEDM)*, 7-11 Dec. 2019 2019, pp. 21.6.1-21.6.4, doi: 10.1109/IEDM19573.2019.8993492.
- [56] M. Kobayashi, C. Jin, and T. Hiramoto, "Comprehensive Understanding of Negative Capacitance FET From the Perspective of Transient Ferroelectric Model,"

- in *2019 IEEE 13th International Conference on ASIC (ASICON)*, 29 Oct.-1 Nov. 2019 2019, pp. 1-4, doi: 10.1109/ASICON47005.2019.8983568.
- [57] S. Deng *et al.*, "Examination of the Interplay Between Polarization Switching and Charge Trapping in Ferroelectric FET," presented at the 2020 IEEE International Electron Devices Meeting (IEDM), 2020.
- [58] J. P. Duarte *et al.*, "BSIM-CMG: Standard FinFET compact model for advanced circuit design," in *ESSCIRC Conference 2015 - 41st European Solid-State Circuits Conference (ESSCIRC)*, 14-18 Sept. 2015 2015, pp. 196-201, doi: 10.1109/ESSCIRC.2015.7313862.
- [59] S. Khandelwal *et al.*, "BSIM-IMG: A Compact Model for Ultrathin-Body SOI MOSFETs With Back-Gate Control," *IEEE Transactions on Electron Devices*, vol. 59, no. 8, pp. 2019-2026, 2012, doi: 10.1109/TED.2012.2198065.
- [60] D. D. Lu, S. De, M. A. Baig, B.-H. Qiu, and Y.-J. Lee, "Computationally efficient compact model for ferroelectric field-effect transistors to simulate the online training of neural networks," *Semiconductor Science and Technology*, vol. 35, no. 9, 2020, doi: 10.1088/1361-6641/ab9bed.
- [61] C.-T. Tung, S. Salahuddin, and C. Hu, "A Compact Model of Antiferroelectric Capacitor," *IEEE Electron Device Letters*, vol. 43, no. 2, pp. 316-318, 2022, doi: 10.1109/led.2021.3135001.
- [62] Y. Goh, J. Hwang, and S. Jeon, "Excellent Reliability and High-Speed Antiferroelectric HfZrO<sub>2</sub> Tunnel Junction by a High-Pressure Annealing Process and Built-In Bias Engineering," *ACS Appl Mater Interfaces*, vol. 12, no. 51, pp. 57539-57546, Dec 23 2020, doi: 10.1021/acsami.0c15091.
- [63] M. Si *et al.*, "Ultrafast measurements of polarization switching dynamics on ferroelectric and anti-ferroelectric hafnium zirconium oxide," *Applied Physics Letters*, vol. 115, no. 7, 2019, doi: 10.1063/1.5098786.
- [64] M. M. Vopson and X. Tan, "Four-State Anti-Ferroelectric Random Access Memory," *IEEE Electron Device Letters*, vol. 37, no. 12, pp. 1551-1554, 2016, doi: 10.1109/led.2016.2614841.
- [65] M. Pesic, U. Schroeder, S. Slesazek, and T. Mikolajick, "Comparative Study of Reliability of Ferroelectric and Anti-Ferroelectric Memories," *IEEE Transactions on Device and Materials Reliability*, vol. 18, no. 2, pp. 154-162, 2018, doi: 10.1109/tdmr.2018.2829112.
- [66] S.-H. Yi, H.-C. Lin, and M.-J. Chen, "Ultra-high energy storage density and scale-up of antiferroelectric TiO<sub>2</sub>/ZrO<sub>2</sub>/TiO<sub>2</sub> stacks for supercapacitors," *Journal of Materials Chemistry A*, vol. 9, no. 14, pp. 9081-9091, 2021, doi: 10.1039/d0ta11991a.
- [67] S.-C. Chang *et al.*, "Anti-ferroelectric HfxZr1-xO<sub>2</sub> Capacitors for High-density 3-D Embedded-DRAM," presented at the 2020 IEEE International Electron Devices Meeting (IEDM), 2020.
- [68] S.-E. Huang, P. Su, and C. Hu, "S-Curve Engineering for ON-State Performance Using Anti-Ferroelectric/Ferroelectric Stack Negative-Capacitance FinFET," *IEEE Transactions on Electron Devices*, vol. 68, no. 9, pp. 4787-4792, 2021, doi: 10.1109/ted.2021.3099090.
- [69] M. H. Lee *et al.*, "Bi-directional Sub-60mV/dec, Hysteresis-Free, Reducing Onset Voltage and High Speed Response of Ferroelectric-AntiFerroelectric

- Hf<sub>0.25</sub>Zr<sub>0.75</sub>O<sub>2</sub> Negative Capacitance FETs," presented at the 2019 IEEE International Electron Devices Meeting (IEDM), San Francisco, CA, USA, 2019.
- [70] C. Kittel, "Theory of Antiferroelectric Crystals," *Physical Review*, vol. 82, no. 5, pp. 729-732, 1951, doi: 10.1103/PhysRev.82.729.
- [71] L. D. Landau and I. M. Khalatnikov, "On the anomalous absorption of sound near a second order phase transition point," *Dokl. Akad. NaukSSSR*, vol. 96, pp. 469-472, May 1954.
- [72] S. Smith, K. Chatterjee, and S. Salahuddin, "Multidomain Phase-Field Modeling of Negative Capacitance Switching Transients," *IEEE Transactions on Electron Devices*, vol. 65, no. 1, pp. 295-298, 2018, doi: 10.1109/ted.2017.2772780.
- [73] J. Muller *et al.*, "Ferroelectricity in Simple Binary ZrO<sub>2</sub> and HfO<sub>2</sub>," *Nano Lett*, vol. 12, no. 8, pp. 4318-23, Aug 8 2012, doi: 10.1021/nl302049k.
- [74] Z. Wang, J. Hur, N. Tasneem, W. Chern, S. Yu, and A. Khan, "Extraction of Preisach model parameters for fluorite-structure ferroelectrics and antiferroelectrics," *Sci Rep*, vol. 11, no. 1, p. 12474, Jun 14 2021, doi: 10.1038/s41598-021-91492-w.
- [75] C. Liu *et al.*, "Energy storage and polarization switching kinetics of (001)-oriented Pb<sub>0.97</sub>La<sub>0.02</sub>(Zr<sub>0.95</sub>Ti<sub>0.05</sub>)O<sub>3</sub> antiferroelectric thick films," *Applied Physics Letters*, vol. 108, no. 11, 2016, doi: 10.1063/1.4944645.
- [76] W. J. Merz, "Domain Formation and Domain Wall Motions in Ferroelectric BaTiO<sub>3</sub> Single Crystals," *Physical Review*, vol. 95, no. 3, pp. 690-698, 1954, doi: 10.1103/PhysRev.95.690.
- [77] C. T. Tung, G. Pahwa, S. Salahuddin, and C. Hu, "A Compact Model of Nanoscale Ferroelectric Capacitor," *IEEE Transactions on Electron Devices*, vol. 69, no. 8, pp. 4761-4764, 2022, doi: 10.1109/TED.2022.3181573.
- [78] H. M. e. al., "Evidence of single domain switching in Hafnium Oxide based FeFETs enabler for multi-level FeFET memory cells," presented at the 2015 IEEE International Electron Devices Meeting (IEDM), Washington, DC, USA, 7-9 Dec. 2015, 2015.
- [79] C.-S. Hsu, S.-C. Chang, D. E. Nikonov, I. A. Young, and A. Naeemi, "A Theoretical Study of Multidomain Ferroelectric Switching Dynamics With a Physics-Based SPICE Circuit Model for Phase-Field Simulations," *IEEE Transactions on Electron Devices*, vol. 67, no. 7, pp. 2952-2959, 2020, doi: 10.1109/ted.2020.2990891.
- [80] A. K. Saha, M. Si, K. Ni, S. Datta, P. D. Ye, and S. K. Gupta, "Ferroelectric Thickness Dependent Domain Interactions in FEFETs for Memory and Logic: A Phase-field Model based Analysis," in *2020 IEEE International Electron Devices Meeting (IEDM)*, 12-18 Dec. 2020 2020, pp. 4.3.1-4.3.4, doi: 10.1109/IEDM13553.2020.9372099.
- [81] J. Liao *et al.*, "Grain Size Engineering of Ferroelectric Zr-doped HfO<sub>2</sub> for the Highly Scaled Devices Applications," *IEEE Electron Device Letters*, vol. 40, no. 11, pp. 1868-1871, 2019, doi: 10.1109/led.2019.2944491.
- [82] "COMSOL Multiphysics v. 5.6." <https://www.comsol.com> (accessed Nov. 1, 2021).
- [83] D. Schrade, R. Mueller, B. X. Xu, and D. Gross, "Domain evolution in ferroelectric materials: A continuum phase field model and finite element implementation,"

- Computer Methods in Applied Mechanics and Engineering*, vol. 196, no. 41-44, pp. 4365-4374, 2007, doi: 10.1016/j.cma.2007.05.010.
- [84] L. J. Sinnamon, M. M. Saad, R. M. Bowman, and J. M. Gregg, "Exploring grain size as a cause for "dead-layer" effects in thin film capacitors," *Applied Physics Letters*, vol. 81, no. 4, pp. 703-705, 2002, doi: 10.1063/1.1494837.
- [85] W. Shu, J. Wang, and T.-Y. Zhang, "Effect of grain boundary on the electromechanical response of ferroelectric polycrystals," *Journal of Applied Physics*, vol. 112, no. 6, 2012, doi: 10.1063/1.4752269.
- [86] C.-T. Tung, C. Kumar Dabhi, S. Salahuddin, and C. Hu, "A versatile compact model of resistive random-access memory (RRAM)," *Solid-State Electronics*, vol. 220, p. 108989, 2024/10/01/ 2024, doi: <https://doi.org/10.1016/j.sse.2024.108989>.
- [87] K. Tsunoda *et al.*, "Low Power and High Speed Switching of Ti-doped NiO ReRAM under the Unipolar Voltage Source of less than 3 V," in *2007 IEEE International Electron Devices Meeting*, 10-12 Dec. 2007 2007, pp. 767-770, doi: 10.1109/IEDM.2007.4419060.
- [88] H. Y. Lee *et al.*, "Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO<sub>2</sub> based RRAM," in *2008 IEEE International Electron Devices Meeting*, 15-17 Dec. 2008 2008, pp. 1-4, doi: 10.1109/IEDM.2008.4796677.
- [89] B. Govoreanu *et al.*, "10×10nm<sup>2</sup> Hf/HfO<sub>x</sub> crossbar resistive RAM with excellent performance, reliability and low-energy operation," in *2011 International Electron Devices Meeting*, 5-7 Dec. 2011 2011, pp. 31.6.1-31.6.4, doi: 10.1109/IEDM.2011.6131652.
- [90] X. P. Wang *et al.*, "Highly compact 1T-1R architecture (4F<sup>2</sup> footprint) involving fully CMOS compatible vertical GAA nano-pillar transistors and oxide-based RRAM cells exhibiting excellent NVM properties and ultra-low power operation," in *2012 International Electron Devices Meeting*, 10-13 Dec. 2012 2012, pp. 20.6.1-20.6.4, doi: 10.1109/IEDM.2012.6479082.
- [91] Y. Wu, J. Liang, S. Yu, X. Guan, and H. S. P. Wong, "Resistive switching random access memory — Materials, device, interconnects, and scaling considerations," in *2012 IEEE International Integrated Reliability Workshop Final Report*, 14-18 Oct. 2012 2012, pp. 16-21, doi: 10.1109/IIRW.2012.6468906.
- [92] M. Kund *et al.*, "Conductive bridging RAM (CBRAM): an emerging non-volatile memory technology scalable to sub 20nm," in *IEEE International Electron Devices Meeting, 2005. IEDM Technical Digest.*, 5-5 Dec. 2005 2005, pp. 754-757, doi: 10.1109/IEDM.2005.1609463.
- [93] S. Yu, Y. Wu, and H. S. P. Wong, "Investigating the switching dynamics and multilevel capability of bipolar metal oxide resistive switching memory," *Applied Physics Letters*, vol. 98, no. 10, 2011, doi: 10.1063/1.3564883.
- [94] T. S. Lee, N. J. Lee, H. Abbas, H. H. Lee, T.-S. Yoon, and C. J. Kang, "Compliance Current-Controlled Conducting Filament Formation in Tantalum Oxide-Based RRAM Devices with Different Top Electrodes," *ACS Applied Electronic Materials*, vol. 2, no. 4, pp. 1154-1161, 2020/04/28 2020, doi: 10.1021/acsaelm.0c00128.
- [95] Z. Jiang *et al.*, "A Compact Model for Metal–Oxide Resistive Random Access Memory With Experiment Verification," *IEEE Transactions on Electron Devices*, vol. 63, no. 5, pp. 1884-1892, 2016, doi: 10.1109/ted.2016.2545412.

- [96] Z. Jiang, S. Yu, Y. Wu, J. H. Engel, X. Guan, and H. S. P. Wong, "Verilog-A compact model for oxide-based resistive random access memory (RRAM)," in *2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 9-11 Sept. 2014, pp. 41-44, doi: 10.1109/SISPAD.2014.6931558.
- [97] X. Guan, S. Yu, and H. S. P. Wong, "A SPICE Compact Model of Metal Oxide Resistive Switching Memory With Variations," *IEEE Electron Device Letters*, vol. 33, no. 10, pp. 1405-1407, 2012, doi: 10.1109/led.2012.2210856.
- [98] Z. Jiang and H.-S. P. Wong, "Stanford University Resistive-Switching Random Access Memory (RRAM) Verilog-A Model," ed, 2014.
- [99] P.-Y. Chen and S. Yu, "Compact Modeling of RRAM Devices and Its Applications in 1T1R and 1S1R Array Design," *IEEE Transactions on Electron Devices*, vol. 62, no. 12, pp. 4022-4028, 2015, doi: 10.1109/ted.2015.2492421.
- [100] P. Huang *et al.*, "A Physics-Based Compact Model of Metal-Oxide-Based RRAM DC and AC Operations," *IEEE Transactions on Electron Devices*, vol. 60, no. 12, pp. 4090-4097, 2013, doi: 10.1109/ted.2013.2287755.
- [101] H. Li *et al.*, "Variation-aware, reliability-emphasized design and optimization of RRAM using SPICE model," in *2015 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 9-13 March 2015, pp. 1425-1430, doi: 10.7873/DATE.2015.0362.
- [102] S. Ambrogio, S. Balatti, D. C. Gilmer, and D. Ielmini, "Analytical Modeling of Oxide-Based Bipolar Resistive Memories and Complementary Resistive Switches," *IEEE Transactions on Electron Devices*, vol. 61, no. 7, pp. 2378-2386, 2014, doi: 10.1109/ted.2014.2325531.
- [103] R. Waser and M. Aono, "Nanoionics-based resistive switching memories," *Nature Materials*, vol. 6, no. 11, pp. 833-840, 2007/11/01 2007, doi: 10.1038/nmat2023.
- [104] R. Waser, R. Dittmann, G. Staikov, and K. Szot, "Redox-Based Resistive Switching Memories – Nanoionic Mechanisms, Prospects, and Challenges," *Advanced Materials*, vol. 21, no. 25-26, pp. 2632-2663, 2009/07/13 2009, doi: <https://doi.org/10.1002/adma.200900375>.
- [105] U. Russo, D. Kamalanathan, D. Ielmini, A. L. Lacaita, and M. N. Kozicki, "Study of Multilevel Programming in Programmable Metallization Cell (PMC) Memory," *IEEE Transactions on Electron Devices*, vol. 56, no. 5, pp. 1040-1047, 2009, doi: 10.1109/TED.2009.2016019.
- [106] D. Kamalanathan, U. Russo, D. Ielmini, and M. N. Kozicki, "Voltage-Driven On–Off Transition and Tradeoff With Program and Erase Current in Programmable Metallization Cell (PMC) Memory," *IEEE Electron Device Letters*, vol. 30, no. 5, pp. 553-555, 2009, doi: 10.1109/LED.2009.2016991.
- [107] S. Yu and H. S. P. Wong, "Compact Modeling of Conducting-Bridge Random-Access Memory (CBRAM)," *IEEE Transactions on Electron Devices*, vol. 58, no. 5, pp. 1352-1360, 2011, doi: 10.1109/TED.2011.2116120.
- [108] U. Russo, D. Ielmini, C. Cagli, and A. L. Lacaita, "Self-Accelerated Thermal Dissolution Model for Reset Programming in Unipolar Resistive-Switching Memory (RRAM) Devices," *IEEE Transactions on Electron Devices*, vol. 56, no. 2, pp. 193-200, 2009, doi: 10.1109/ted.2008.2010584.
- [109] M. Bocquet *et al.*, "Robust Compact Model for Bipolar Oxide-Based Resistive

- Switching Memories," *IEEE Transactions on Electron Devices*, vol. 61, no. 3, pp. 674-681, 2014, doi: 10.1109/ted.2013.2296793.
- [110] G. Wang *et al.*, "Impact of program/erase operation on the performances of oxide-based resistive switching memory," *Nanoscale Research Letters*, vol. 10, no. 1, p. 39, 2015/02/05 2015, doi: 10.1186/s11671-014-0721-2.
- [111] L. T. Clark *et al.*, "ASAP7: A 7-nm finFET predictive process design kit," *Microelectronics Journal*, vol. 53, pp. 105-115, 2016/07/01/ 2016, doi: <https://doi.org/10.1016/j.mejo.2016.04.006>.
- [112] C. T. Tung, A. Dasgupta, H. Agarwal, S. Salahuddin, and C. Hu, "A Compact Model of Perpendicular Spin-Transfer-Torque Magnetic Tunnel Junction," *IEEE Transactions on Electron Devices*, vol. 71, no. 1, pp. 57-61, 2024, doi: 10.1109/TED.2023.3313997.
- [113] M. Julliere, "Tunneling between ferromagnetic films," *Physics Letters A*, vol. 54, no. 3, pp. 225-226, 1975/09/08/ 1975, doi: [https://doi.org/10.1016/0375-9601\(75\)90174-7](https://doi.org/10.1016/0375-9601(75)90174-7).
- [114] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers," *Journal of Magnetism and Magnetic Materials*, vol. 159, no. 1, pp. L1-L7, 1996/06/01/ 1996, doi: [https://doi.org/10.1016/0304-8853\(96\)00062-5](https://doi.org/10.1016/0304-8853(96)00062-5).
- [115] Y. Huai, F. Albert, P. Nguyen, M. Pakala, and T. Valet, "Observation of spin-transfer switching in deep submicron-sized and low-resistance magnetic tunnel junctions," *Applied Physics Letters*, vol. 84, no. 16, pp. 3118-3120, 2004/04/19 2004, doi: 10.1063/1.1707228.
- [116] W. H. Butler *et al.*, "Switching Distributions for Perpendicular Spin-Torque Devices Within the Macrospin Approximation," *IEEE Transactions on Magnetics*, vol. 48, no. 12, pp. 4684-4700, 2012, doi: 10.1109/tmag.2012.2209122.
- [117] A. V. Khvalkovskiy *et al.*, "Erratum: Basic principles of STT-MRAM cell operation in memory arrays," *Journal of Physics D: Applied Physics*, vol. 46, no. 13, 2013, doi: 10.1088/0022-3727/46/13/139601.
- [118] R. H. Koch, J. A. Katine, and J. Z. Sun, "Time-Resolved Reversal of Spin-Transfer Switching in a Nanomagnet," *Physical Review Letters*, vol. 92, no. 8, p. 088302, 02/26/ 2004, doi: 10.1103/PhysRevLett.92.088302.
- [119] Y. Zhang *et al.*, "Electrical Modeling of Stochastic Spin Transfer Torque Writing in Magnetic Tunnel Junctions for Memory and Logic Applications," *IEEE Transactions on Magnetics*, vol. 49, no. 7, pp. 4375-4378, 2013, doi: 10.1109/tmag.2013.2242257.
- [120] X. Fong, S. H. Choday, P. Georgios, C. Augustine, and K. Roy, "Purdue Nanoelectronics Research Laboratory Magnetic Tunnel Junction Model," ed, 2014.
- [121] A. S. Roy, A. Sarkar, and S. P. Mudanai, "Compact Modeling of Magnetic Tunneling Junctions," *IEEE Transactions on Electron Devices*, vol. 63, no. 2, pp. 652-658, 2016, doi: 10.1109/ted.2015.2510368.
- [122] J.-B. Kammerer, M. Madec, and L. Hébrard, "Compact Modeling of a Magnetic Tunnel Junction—Part I: Dynamic Magnetization Model," *IEEE Transactions on Electron Devices*, vol. 57, no. 6, pp. 1408-1415, 2010, doi: 10.1109/ted.2010.2047070.
- [123] N. S. Gaul *et al.*, "A Physics based MTJ Compact Model for State-of-the-Art and Emerging STT-MRAM Failure Analysis and Yield Enhancement," in *2022 IEEE*

- International Memory Workshop (IMW)*, 15-18 May 2022 2022, pp. 1-4, doi: 10.1109/IMW52921.2022.9779246.
- [124] M. Madec, J.-B. Kammerer, and L. Hébrard, "Compact Modeling of a Magnetic Tunnel Junction—Part II: Tunneling Current Model," *IEEE Transactions on Electron Devices*, vol. 57, no. 6, pp. 1416-1424, 2010, doi: 10.1109/ted.2010.2047071.
- [125] J. D. Harms, F. Ebrahimi, X. Yao, and J.-P. Wang, "SPICE Macromodel of Spin-Torque-Transfer-Operated Magnetic Tunnel Junctions," *IEEE Transactions on Electron Devices*, vol. 57, no. 6, pp. 1425-1430, 2010, doi: 10.1109/ted.2010.2047073.
- [126] H. Lim, S. Lee, and H. Shin, "A Survey on the Modeling of Magnetic Tunnel Junctions for Circuit Simulation," *Active and Passive Electronic Components*, vol. 2016, pp. 1-12, 2016, doi: 10.1155/2016/3858621.
- [127] Z. Xu, C. Yang, M. Mao, K. B. Sutaria, C. Chakrabarti, and Y. Cao, "Compact modeling of STT-MTJ devices," *Solid-State Electronics*, vol. 102, pp. 76-81, 2014, doi: 10.1016/j.sse.2014.06.003.
- [128] A. F. Vincent, N. Locatelli, J.-O. Klein, W. S. Zhao, S. Galdin-Retailleau, and D. Querlioz, "Analytical Macrospin Modeling of the Stochastic Switching Time of Spin-Transfer Torque Devices," *IEEE Transactions on Electron Devices*, vol. 62, no. 1, pp. 164-170, 2015, doi: 10.1109/ted.2014.2372475.
- [129] G. Pahwa *et al.*, "Compact Modeling of Emerging IC Devices for Technology-Design Co-development," in *2022 International Electron Devices Meeting (IEDM)*, 3-7 Dec. 2022 2022, pp. 8.1.1-8.1.4, doi: 10.1109/IEDM45625.2022.10019433.
- [130] Y. Wang, H. Cai, L. A. B. Naviner, Y. Zhang, J. O. Klein, and W. S. Zhao, "Compact thermal modeling of spin transfer torque magnetic tunnel junction," *Microelectronics Reliability*, vol. 55, no. 9, pp. 1649-1653, 2015/08/01/ 2015, doi: <https://doi.org/10.1016/j.microrel.2015.06.029>.
- [131] V. Drewello, J. Schmalhorst, A. Thomas, and G. Reiss, "Evidence for strong magnon contribution to the TMR temperature dependence in MgO based tunnel junctions," *Physical Review B*, vol. 77, no. 1, p. 014440, 01/31/ 2008, doi: 10.1103/PhysRevB.77.014440.
- [132] L. Zhang *et al.*, "Addressing the Thermal Issues of STT-MRAM From Compact Modeling to Design Techniques," *IEEE Transactions on Nanotechnology*, vol. 17, no. 2, pp. 345-352, 2018, doi: 10.1109/tnano.2018.2803340.
- [133] J. Li, P. Ndai, A. Goel, S. Salahuddin, and K. Roy, "Design Paradigm for Robust Spin-Torque Transfer Magnetic RAM (STT MRAM) From Circuit/Architecture Perspective," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 18, no. 12, pp. 1710-1723, 2010, doi: 10.1109/TVLSI.2009.2027907.
- [134] T. Ishikawa *et al.*, "Spin-dependent tunneling characteristics of fully epitaxial magnetic tunneling junctions with a full-Heusler alloy Co<sub>2</sub>MnSi thin film and a MgO tunnel barrier," *Applied Physics Letters*, vol. 89, no. 19, p. 192505, 2006, doi: 10.1063/1.2378397.
- [135] Y. Zhang *et al.*, "Compact Modeling of Perpendicular-Anisotropy CoFeB/MgO Magnetic Tunnel Junctions," *IEEE Transactions on Electron Devices*, vol. 59, no. 3, pp. 819-826, 2012, doi: 10.1109/ted.2011.2178416.

- [136] J. G. Alzate *et al.*, "Temperature dependence of the voltage-controlled perpendicular anisotropy in nanoscale MgO|CoFeB|Ta magnetic tunnel junctions," *Applied Physics Letters*, vol. 104, no. 11, 2014, doi: 10.1063/1.4869152.
- [137] L. Yuan, S. H. Liou, and D. Wang, "Temperature dependence of magnetoresistance in magnetic tunnel junctions with different free layer structures," *Physical Review B*, vol. 73, no. 13, 2006, doi: 10.1103/PhysRevB.73.134403.
- [138] K. C. Chun, H. Zhao, J. D. Harms, T. H. Kim, J. P. Wang, and C. H. Kim, "A Scaling Roadmap and Performance Evaluation of In-Plane and Perpendicular MTJ Based STT-MRAMs for High-Density Cache Memory," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 2, pp. 598-610, 2013, doi: 10.1109/JSSC.2012.2224256.
- [139] L. Wu *et al.*, "Pinhole Defect Characterization and Fault Modeling for STT-MRAM Testing," in *2019 IEEE European Test Symposium (ETS)*, 27-31 May 2019 2019, pp. 1-6, doi: 10.1109/ETS.2019.8791518.
- [140] R. Carboni *et al.*, "A Physics-Based Compact Model of Stochastic Switching in Spin-Transfer Torque Magnetic Memory," *IEEE Transactions on Electron Devices*, vol. 66, no. 10, pp. 4176-4182, 2019, doi: 10.1109/ted.2019.2933315.
- [141] C. J. Lin *et al.*, "45nm low power CMOS logic compatible embedded STT MRAM utilizing a reverse-connection 1T/1MTJ cell," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 7-9 Dec. 2009 2009, pp. 1-4, doi: 10.1109/IEDM.2009.5424368.
- [142] C. T. Tung and C. Hu, "Neural Network-Based BSIM Transistor Model Framework: Currents, Charges, Variability, and Circuit Simulation," *IEEE Transactions on Electron Devices*, vol. 70, no. 4, pp. 2157-2160, 2023, doi: 10.1109/TED.2023.3244901.
- [143] C. T. Tung, M. Y. Kao, and C. Hu, "Neural Network-Based and Modeling With High Accuracy and Potential Model Speed," *IEEE Transactions on Electron Devices*, vol. 69, no. 11, pp. 6476-6479, 2022, doi: 10.1109/TED.2022.3208514.
- [144] C. T. Tung, S. Salahuddin, and C. Hu, "A SPICE-Compatible Neural Network Compact Model for Efficient IC Simulations," in *2024 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 24-27 Sept. 2024 2024, pp. 01-04, doi: 10.1109/SISPAD62626.2024.10733248.
- [145] A. Dasgupta *et al.*, "BSIM Compact Model of Quantum Confinement in Advanced Nanosheet FETs," *IEEE Transactions on Electron Devices*, vol. 67, no. 2, pp. 730-737, 2020, doi: 10.1109/TED.2019.2960269.
- [146] J. Wang, Y. H. Kim, J. Ryu, C. Jeong, W. Choi, and D. Kim, "Artificial Neural Network-Based Compact Modeling Methodology for Advanced Transistors," *IEEE Transactions on Electron Devices*, vol. 68, no. 3, pp. 1318-1325, 2021, doi: 10.1109/TED.2020.3048918.
- [147] M. Li, O. Irsoy, C. Cardie, and H. G. Xing, "Physics-Inspired Neural Networks for Efficient Device Compact Modeling," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 2, pp. 44-49, 2016, doi: 10.1109/JXCDC.2016.2636161.
- [148] M. Y. Kao, H. Kam, and C. Hu, "Deep-Learning-Assisted Physics-Driven MOSFET Current-Voltage Modeling," *IEEE Electron Device Letters*, pp. 1-1, 2022, doi: 10.1109/LED.2022.3168243.
- [149] K. Ko, J. K. Lee, M. Kang, J. Jeon, and H. Shin, "Prediction of Process Variation

- Effect for Ultrascaled GAA Vertical FET Devices Using a Machine Learning Approach," *IEEE Transactions on Electron Devices*, vol. 66, no. 10, pp. 4474-4477, 2019, doi: 10.1109/ted.2019.2937786.
- [150] H. Carrillo-Nunez, N. Dimitrova, A. Asenov, and V. Georgiev, "Machine Learning Approach for Predicting the Effect of Statistical Variability in Si Junctionless Nanowire Transistors," *IEEE Electron Device Letters*, vol. 40, no. 9, pp. 1366-1369, 2019, doi: 10.1109/led.2019.2931839.
- [151] C. Akbar, Y. Li, and N. Thoti, "Device-Simulation-Based Machine Learning Technique for the Characteristic of Line Tunnel Field-Effect Transistors," *IEEE Access*, vol. 10, pp. 53098-53107, 2022, doi: 10.1109/access.2022.3174685.
- [152] S. S. Cheema *et al.*, "Ultrathin ferroic HfO<sub>2</sub>-ZrO<sub>2</sub> superlattice gate stack for advanced transistors," *Nature*, vol. 604, no. 7904, pp. 65-71, 2022/04/01 2022, doi: 10.1038/s41586-022-04425-6.
- [153] C. T. Tung, M. Y. Kao, and C. Hu, "Neural Network-Based I – V and C – V Modeling With High Accuracy and Potential Model Speed," *IEEE Transactions on Electron Devices*, pp. 1-4, 2022, doi: 10.1109/TED.2022.3208514.
- [154] W. Bing, J. R. Hellums, and C. G. Sodini, "MOSFET thermal noise modeling for analog integrated circuits," *IEEE Journal of Solid-State Circuits*, vol. 29, no. 7, pp. 833-835, 1994, doi: 10.1109/4.303722.
- [155] C. Auth *et al.*, "A 10nm high performance and low-power CMOS technology featuring 3rd generation FinFET transistors, Self-Aligned Quad Patterning, contact over active gate and cobalt local interconnects," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2-6 Dec. 2017 2017, pp. 29.1.1-29.1.4, doi: 10.1109/IEDM.2017.8268472.
- [156] C. T. Tung, S. Salahuddin, and C. Hu, "Non-Quasi-Static Modeling of Neural Network-Based Transistor Compact Model for Fast Transient, AC, and RF Simulations," *IEEE Electron Device Letters*, vol. 45, no. 7, pp. 1277-1280, 2024, doi: 10.1109/LED.2024.3404404.
- [157] K. Sheelvardhan, S. Guglani, M. Ehteshamuddin, S. Roy, and A. Dasgupta, "Machine Learning Augmented Compact Modeling for Simultaneous Improvement in Computational Speed and Accuracy," *IEEE Transactions on Electron Devices*, vol. 71, no. 1, pp. 239-245, 2024, doi: 10.1109/TED.2023.3251296.
- [158] Y. S. Yang, Y. Li, and S. R. R. Kola, "A Physical-Based Artificial Neural Networks Compact Modeling Framework for Emerging FETs," *IEEE Transactions on Electron Devices*, vol. 71, no. 1, pp. 223-230, 2024, doi: 10.1109/TED.2023.3269410.
- [159] Z. Yang, A. D. Gaidhane, K. Anderson, G. Workman, and Y. Cao, "Graph-Based Compact Model (GCM) for Efficient Transistor Parameter Extraction: A Machine Learning Approach on 12 nm FinFETs," *IEEE Transactions on Electron Devices*, vol. 71, no. 1, pp. 254-262, 2024, doi: 10.1109/TED.2023.3327973.
- [160] M. Chan, K. Y. Hui, H. Chenming, and P. K. Ko, "A robust and physical BSIM3 non-quasi-static transient and AC small-signal model for circuit simulation," *IEEE Transactions on Electron Devices*, vol. 45, no. 4, pp. 834-841, 1998, doi: 10.1109/16.662788.
- [161] L. Wai-Kit, M. Chan, and P. K. Ko, "A physical approach to enhance BSIM3 NQS model for fast transient simulation," in *Proceedings 1999 IEEE Hong Kong*

- Electron Devices Meeting (Cat. No.99TH8458)*, 26-26 June 1999 1999, pp. 114-117, doi: 10.1109/HKEDM.1999.836421.
- [162] Z. Zhu, G. Gildenblat, C. C. McAndrew, and I. S. Lim, "Accurate RTA-Based Nonquasi-Static MOSFET Model for RF and Mixed-Signal Simulations," *IEEE Transactions on Electron Devices*, vol. 59, no. 5, pp. 1236-1244, 2012, doi: 10.1109/TED.2012.2186636.
- [163] M. Bucher and A. Bazigos, "An efficient channel segmentation approach for a large-signal NQS MOSFET model," *Solid-State Electronics*, vol. 52, no. 2, pp. 275-281, 2008/02/01/ 2008, doi: <https://doi.org/10.1016/j.sse.2007.08.015>.
- [164] A. J. Scholten, L. F. Tiemeijer, P. W. H. D. Vreede, and D. B. M. Klaassen, "A large signal non-quasi-static MOS model for RF circuit simulation," in *International Electron Devices Meeting 1999. Technical Digest (Cat. No.99CH36318)*, 5-8 Dec. 1999 1999, pp. 163-166, doi: 10.1109/IEDM.1999.823870.
- [165] S. Sarkar, A. S. Roy, and S. Mahapatra, "Unified large and small signal non-quasi-static model for long channel symmetric DG MOSFET," *Solid-State Electronics*, vol. 54, no. 11, pp. 1421-1429, 2010/11/01/ 2010, doi: <https://doi.org/10.1016/j.sse.2010.05.036>.
- [166] H. Wang *et al.*, "A Unified Nonquasi-Static MOSFET Model for Large-Signal and Small-Signal Simulations," *IEEE Transactions on Electron Devices*, vol. 53, no. 9, pp. 2035-2043, 2006, doi: 10.1109/TED.2005.881003.
- [167] "BSIM-CMG Technical Manual." <https://bsim.berkeley.edu/models/bsimcmg/> (accessed Sep. 1, 2021).
- [168] Z. Chen, M. Raginsky, and E. Rosenbaum, "Verilog-A compatible recurrent neural network model for transient circuit simulation," in *2017 IEEE 26th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS)*, 15-18 Oct. 2017 2017, pp. 1-3, doi: 10.1109/EPEPS.2017.8329743.
- [169] J. Xiong, A. S. Yang, M. Raginsky, and E. Rosenbaum, "Neural Networks for Transient Modeling of Circuits : Invited Paper," in *2021 ACM/IEEE 3rd Workshop on Machine Learning for CAD (MLCAD)*, 30 Aug.-3 Sept. 2021 2021, pp. 1-7, doi: 10.1109/MLCAD52597.2021.9531153.
- [170] C. T. Tung, A. Pampori, C. K. Dabhi, S. Salahuddin, and C. Hu, "A Novel Neural Network-Based Transistor Compact Model Including Self-Heating," *IEEE Electron Device Letters*, vol. 45, no. 8, pp. 1512-1515, 2024, doi: 10.1109/LED.2024.3408151.
- [171] D. Jang *et al.*, "Self-heating on bulk FinFET from 14nm down to 7nm node," in *2015 IEEE International Electron Devices Meeting (IEDM)*, 7-9 Dec. 2015 2015, pp. 11.6.1-11.6.4, doi: 10.1109/IEDM.2015.7409678.
- [172] O. Prakash, C. K. Dabhi, Y. S. Chauhan, and H. Amrouch, "Transistor Self-Heating: The Rising Challenge for Semiconductor Testing," in *2021 IEEE 39th VLSI Test Symposium (VTS)*, 25-28 April 2021 2021, pp. 1-7, doi: 10.1109/VTS50974.2021.9441002.
- [173] O. Prakash, G. Pahwa, C. K. Dabhi, Y. S. Chauhan, and H. Amrouch, "Impact of Self-Heating on Negative-Capacitance FinFET: Device-Circuit Interaction," *IEEE Transactions on Electron Devices*, vol. 68, no. 4, pp. 1420-1424, 2021, doi: 10.1109/TED.2021.3059180.

- [174] W. Ahn, C. Jiang, J. Xu, and M. A. Alam, "A new framework of physics-based compact model predicts reliability of self-heated modern ICs: FinFET, NWFET, NSHFET comparison," in *2017 IEEE International Electron Devices Meeting (IEDM)*, 2-6 Dec. 2017 2017, pp. 13.6.1-13.6.4, doi: 10.1109/IEDM.2017.8268386.
- [175] C. Gupta, R. Goel, H. Agarwal, C. Hu, and Y. S. Chauhan, "BSIM-BULK: Accurate Compact Model for Analog and RF Circuit Design," in *2019 IEEE Custom Integrated Circuits Conference (CICC)*, 14-17 April 2019 2019, pp. 1-8, doi: 10.1109/CICC.2019.8780155.
- [176] H. Agarwal *et al.*, "BSIM-HV: High-Voltage MOSFET Model Including Quasi-Saturation and Self-Heating Effect," *IEEE Transactions on Electron Devices*, vol. 66, no. 10, pp. 4258-4263, 2019, doi: 10.1109/TED.2019.2933611.
- [177] L. Zhang, D. Song, Y. Xiao, X. Lin, and M. Chan, "On the Formulation of Self-Heating Models for Circuit Simulation," *IEEE Journal of the Electron Devices Society*, vol. 6, pp. 291-297, 2018, doi: 10.1109/JEDS.2018.2801301.
- [178] Y. Kim, S. Myung, J. Ryu, C. Jeong, and D. S. Kim, "Physics-augmented Neural Compact Model for Emerging Device Technologies," in *2020 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 23 Sept.-6 Oct. 2020 2020, pp. 257-260, doi: 10.23919/SISPAD49475.2020.9241638.
- [179] J. Xu, R. Jones, S. A. Harris, T. Nielsen, and D. E. Root, "Dynamic FET model - DynaFET - for GaN transistors from NVNA active source injection measurements," in *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*, 1-6 June 2014 2014, pp. 1-3, doi: 10.1109/MWSYM.2014.6848293.
- [180] M. Li *et al.*, "A Scalable Knowledge-Based Neural Network Model for GaN HEMTs With Accurate Trapping and Self-Heating Effects Characterization," *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 9, pp. 3747-3760, 2023, doi: 10.1109/TMTT.2023.3248225.
- [181] M. A. Karami and A. Afzali-Kusha, "Adaptive Neural Network Model for SOI-MOSFET I-V Characteristic Including Self-Heating Effect," in *2006 International Conference on Microelectronics*, 16-19 Dec. 2006 2006, pp. 5-8, doi: 10.1109/ICM.2006.373640.
- [182] L. W. Nagel and D. O. Pederson, "SPICE (Simulation Program with Integrated Circuit Emphasis)," UCB/ERL M382, 1973. [Online]. Available: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/1973/22871.html>
- [183] T. Wang, A. V. Karthik, B. Wu, J. Yao, and J. Roychowdhury, "MAPP: The Berkeley Model and Algorithm Prototyping Platform," in *2015 IEEE Custom Integrated Circuits Conference (CICC)*, 28-30 Sept. 2015 2015, pp. 1-8, doi: 10.1109/CICC.2015.7338431.
- [184] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686-707, 2019/02/01/ 2019, doi: <https://doi.org/10.1016/j.jcp.2018.10.045>.
- [185] X. Meng, Z. Li, D. Zhang, and G. E. Karniadakis, "PPINN: Parareal physics-informed neural network for time-dependent PDEs," *Computer Methods in Applied Mechanics and Engineering*, vol. 370, p. 113250, 2020/10/01/ 2020, doi:

- <https://doi.org/10.1016/j.cma.2020.113250>.
- [186] J. Stiasny, B. Zhang, and S. Chatzivasileiadis, "PINNSim: A simulator for power system dynamics based on Physics-Informed Neural Networks," *Electric Power Systems Research*, vol. 235, p. 110796, 2024/10/01/ 2024, doi: <https://doi.org/10.1016/j.epsr.2024.110796>.
- [187] C. Moya and G. Lin, "DAE-PINN: a physics-informed neural network model for simulating differential algebraic equations with application to power networks," *Neural Computing and Applications*, vol. 35, no. 5, pp. 3789-3804, 2023/02/01 2023, doi: 10.1007/s00521-022-07886-y.
- [188] B. Kim and M. Shin, "A Novel Neural-Network Device Modeling Based on Physics-Informed Machine Learning," *IEEE Transactions on Electron Devices*, vol. 70, no. 11, pp. 6021-6025, 2023, doi: 10.1109/TED.2023.3316635.
- [189] C. C. McAndrew *et al.*, "Best Practices for Compact Modeling in Verilog-A," *IEEE Journal of the Electron Devices Society*, vol. 3, no. 5, pp. 383-396, 2015, doi: 10.1109/JEDS.2015.2455342.
- [190] A. Paszke *et al.*, "Automatic differentiation in pytorch," 2017.
- [191] Z. Xiang, W. Peng, X. Liu, and W. Yao, "Self-adaptive loss balanced Physics-informed neural networks," *Neurocomputing*, vol. 496, pp. 11-34, 2022/07/28/ 2022, doi: <https://doi.org/10.1016/j.neucom.2022.05.015>.