Advanced mathematical strategies to speed up energy-efficient microelectronic device modeling

Nia Maheshwari

SULI Summer 2024

Abstract

Ferroelectric-based transistors can exhibit negative capacitance, a property that allows for lower operating voltages. FerroX is a 3D simulation framework that enables investigation of the energy dynamics within these structures. FerroX performs differential-equation based calculations, and is currently limited by inadequate time integration capabilities. A successful implementation of the *SUite of Nonlinear and DIfferential/ALgebraic equation Solvers* (SUNDIALS) [1, 2] support doubled the time step length, and helped overcome this restriction. These improvements to FerroX introduce higher resolution, more realistic microelectronics modeling– key steps towards developing ultra-low power devices.

Background and Introduction

Moore's Law predicts that the density of transistors in integrated circuits doubles every two years, a pattern that manufacturers of Information and Communications Technology (ICT) have successfully followed in the past. Within the last decade, this practice has begun to drive energy consumption. As visualized in Fig. 1(b), ICTs are predicted to account for approximately one quarter of energy consumption by 2030. This reveals that the current structure of transistors is unsustainable, and their inefficiency poses a limitation to expanding global computation capability. Engineering transistors that are energy efficient is one strategy that can help surpass this restriction by reducing the power consumption of ICTs. Achieving this entails overcoming the Boltzmann Tyranny, a distribution that defines the minimum voltage required to operate a transistor.

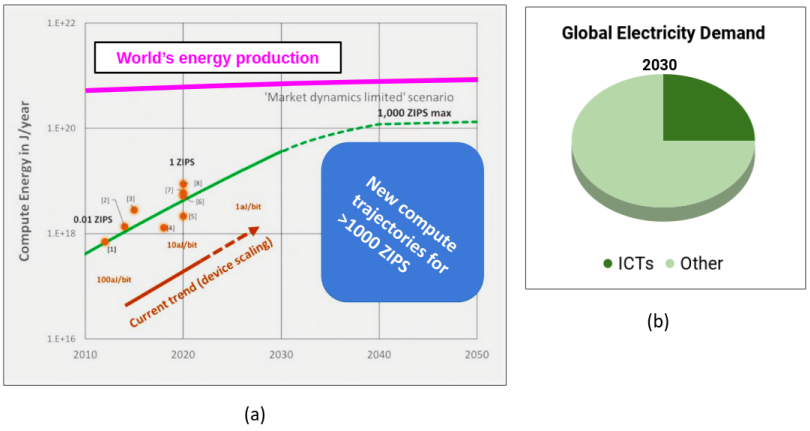

Figure 1: *(a) Energy inefficiency of computers is the limiting factor in future global computing capacity. Image credit: Semiconductor Research Corporation, Semiconductor Industry Association. (b) With the current structure of transistors, Information and Communication Technology (ICT) is predicted to account for nearly one quarter of all energy consumption by 2030.*

First introduced by Salahuddin et al. in 2008, replacing the standard insulator in a transistor gate stack structure with a ferroelectric, as illustrated in Fig. 2(b), can enable lower power operation. [3] When experiencing spontaneous polarization, ferroelectric material exhibits the unique phenomenon of negative capacitance. Negative capacitance can amplify the flow of charge across the semiconductor channel, allowing a transistor to reach the constant on-off threshold for a lower applied voltage. As shown in Fig. 2(a), Negative Capacitance Field Effect Transistors (NCFETs) require a smaller voltage than Field Effect Transistors (FETs), which supports that ferroelectric material can be a revolutionary substance for enabling a lower operating voltage in transistors.

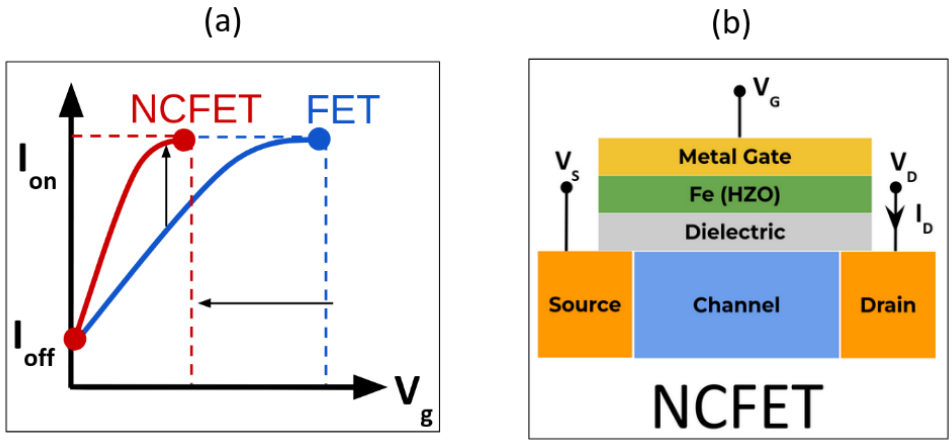

Figure 2: *(a) This graph illustrates that Negative Capacitance Field Effect Transistor (NCFET) requires lower voltage than Field Effect Transistor (FET) to overcome on-off threshold. (b) A simplified NCFET diagram, depicting the gate stack, semiconductor channel, source, and drain components. Kumar et al. 2024.*

To better understand this physical property and its enhancing conditions, we utilize FerroX- a scalable, massively parallel, 3D phase-field simulation framework. FerroX enables us to investigate the polarization of charge across ferroelectric heterostructures through the coupling of multiple numerical schemes. Following the description of Kumar et al. [4], FerroX is a massively-parallel, scalable simulation framework that self-consistently solves the time-dependent Ginzburg Landau (TDGL) equation for ferroelectric polarization, Poisson's equation for electric potential, and the charge equation for carrier densities in semiconductor regions, as depicted in a typical time step through Fig. 3(a).

Currently, FerroX utilizes first order Forward Euler and second order Trapezoid schemes to numerically integrate the time-dependent Ginzburg-Landau equation for ferroelectric polarization. These schemes are explicit and impose a strict stability criterion on the maximum allowed time step size. The goal of this work is to incorporate SUNDIALS library in FerroX to perform time-integration steps. SUNDIALS contains a number of robust and advanced time integrators which can be incorporated into existing simulation codes, such as FerroX. The central discussion of this paper is related to the scope for improvement in FerroX performance, with a focus on minimizing the time taken by integration computations in particular.

Materials and Methods

Overview of FerroX

Collectively, these equations model the complex physics of multi-material stacks, including metal-ferroelectric-insulator-metal (MFIM) and metal-ferroelectric-insulator-semiconductor-metal (MFISM) structures, the latter of which is illustrated in a simplified representation in Fig. 3(b).

The TDGL equation is responsible for characterizing the polarization within ferroelectrics. Within this equation, the P = (P_x, P_y, P_z) and r = (x, y, z) terms represent the electric polarization and spatial vectors, respectively; T is the kinetic or viscosity coefficient, and $(\frac{\delta F}{\delta P(r,t)})$ is the primary factor for the progres-

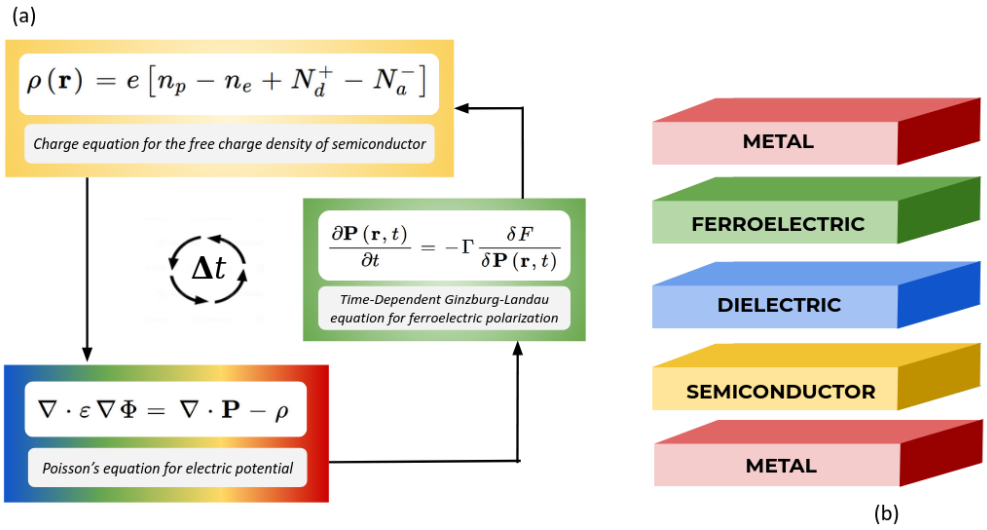

Figure 3: *The scalable, massively parallel, 3D simulation program FerroX models the complex physical properties of multi-material stacks by self-consistently solving the equation system. (a) A visualization of one time step within the coupled equation system. (b) A typical multi-material stack.*

sion of the system. Poisson's equation determines the distribution of electric potential under conditions of electrostatic equilibrium; here, ϵ is a spatially-varying permittivity and ρ is the total free charge density within the semiconductor layer. For MFIM structures, this term can be toggled to zero due to the lack of a semiconductor material; however, simulations involving MFISM structures rely on the charge equation to compute charge density– a variable dependent on the local distribution of electric potential within the semiconductor region. The e term is the elementary charge, and n_p $(\mathbf{r})$, n_e $(\mathbf{r})$, N_d^+ $(\mathbf{r})$, and N_a^- $(\mathbf{r})$ are of densities of holes, electrons, ionized donors, and acceptors at spatial location $\mathbf{r}$ [4]. In situations where the semiconductor layer is composed of undoped silicon, the terms $N_d^+(\mathbf{r})$ and $N_a^-(\mathbf{r})$ are negligible [4].

Fig. 3(a) illustrates a typical time step of FerroX. A typical time step of FerroX is shown in Fig. 3(a). We initialize polarization $\mathbf{P}$ using a uniformly distributed random numbers and compute the corresponding distributions of potential (ϕ) and charge density (ρ) iteratively until self-consistency is reached. At each time step we solve the TDGL equation to update $\mathbf{P}$ and use the updated P to compute ϕ and ρ. The Poisson solver used in our framework provides second order accuracy in space. The temporal integrators for the TDGL equation are configured to produce either first or second-order convergence in time, depending on a run time input parameter.

Implementing SUNDIALS

Each computation requires a number of time steps of a certain length, which can determine the overall time of one simulation. The goal of this project is to determine how to incorporate advanced SUNDIALS features in order to lengthen the maximum supported time step such that simulation runtime is optimized. Integrating SUNDIALS into the FerroX source code followed the simple structure of an if-else code block. To grant SUNDIALS a supplementary role within the FerroX code, our conditional implementation of the library would be controlled by an input parameter that functions equivalent to a boolean variable.

Results and Discussion

Using FerroX, we explore the dynamics of domain walls, polarization switching, and the negative capacitance effect in an MFIM stack. The device consists of a 5-nm-thick HZO and 4-nm-thick Al2O3 as the ferroelectric and dielectric layers, respectively. Lateral dimensions along the x and y axes are 32 nanometers.

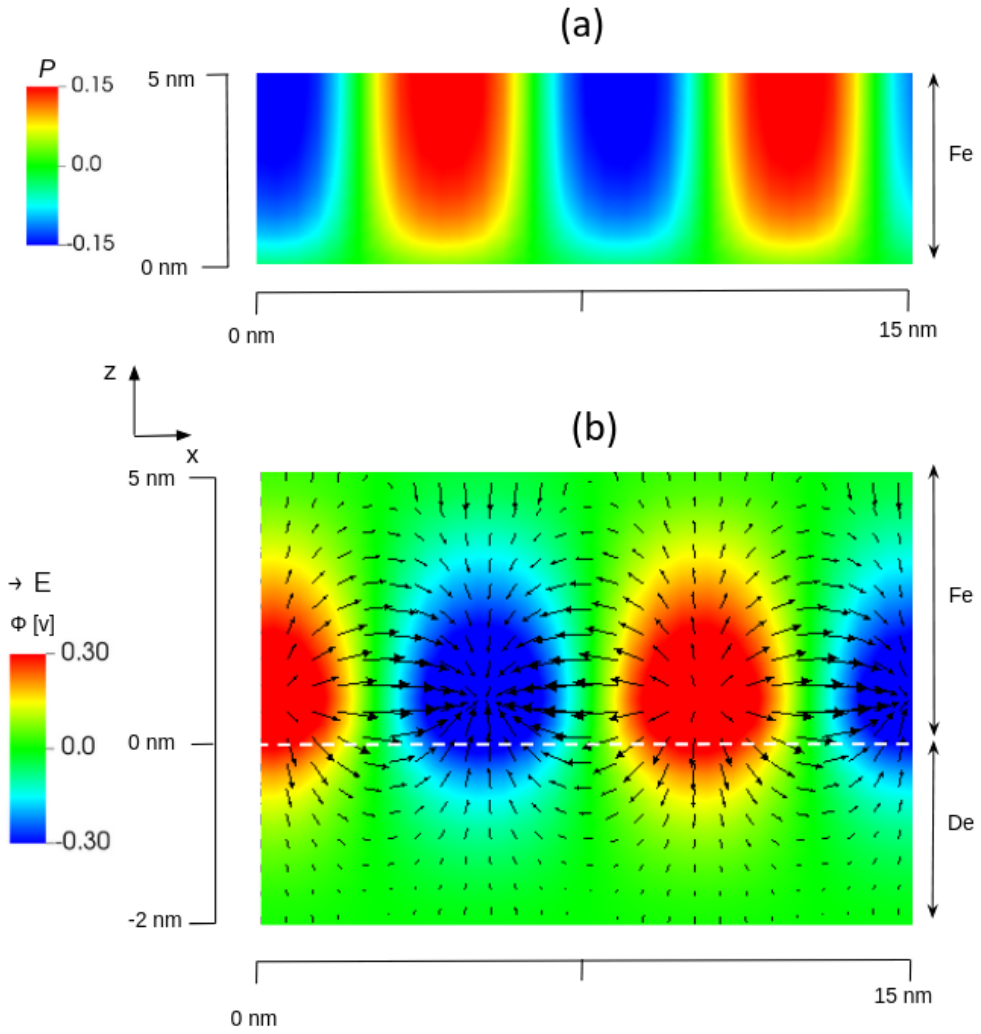

Figure 4: *Stable Simulation Result as Viewed through VisIt Software. (a) A depiction of the polarization (P(x,y)) across the ferroelectric layer. (b) Potential profile* ($\phi(x, y)$) *across the ferroelectric and dielectric interfaces, along with the electric field, illustrating* 180° *domain formation.*

Applied-voltage induces polarization switching due to the dynamics of domain walls which results in the negative capacitance effect in MFIM stacks. We investigate multi-domain formation in FE by considering an applied voltage V_{app} of 0 V across the stack. Once initialized, the polarization P in the ferroelectric layer evolves towards equilibrium with minimized total energy. The depolarization field, generated at the HZO-Al2O3 interface, which acts opposite to the polarization and results in an increase in the depolarization energy is compensated by the formation of 180° domains of alternate positive and negative P values of approximately equal size as shown in Fig. 4(a). The non-homogeneity in the FE-DE interface potential manifests into a spatially varying potential profile which shows maxima and minima corresponding to negative and positive P values as shown in Fig. 4(b). Corresponding electric field vectors are shown as arrows.

Establishing Baseline FerroX Performance

The FerroX problem set up is described through an input parameters file that defines the geometry of the heterostructure to be simulated, as well as additional numerical and physical parameters. The time step (dt) and number of time steps (nsteps) parameters were the focus of the initial set of tests. The product of dt and nsteps was maintained across tests, so dt was adjusted first (increased by 1 fs, or 1e-13 s) and followed

by the nsteps variable as appropriate. The maximum stable time step supported by each internal FerroX method is listed in Table 1.

Testing SUNDIALS Methods

In order to understand and compare the individual capabilities of FerroX with and without external libraries, we followed a similar procedure with the SUNDIALS package. After incorporating SUNDIALS into the source code and makefile, we incrementally modified the dt and nsteps variables until the simulation failed. This procedure allowed us to assess the time integration capabilities of FerroX both before and after enabling SUNDIALS. In asserting the benchmark performance for internal and native methods, we can better understand the magnitude of improvement yielded by implementing advanced SUNDIALS techniques.

Internal FerroX Method	**Maximum Supported Time Step (s)**
First Order	4.0e-13
Second Order	4.0e-13
Native Runge Kutta Method	
Forward Euler	4.0e-13
Trapezoid	4.0e-13
SSPRK3	5.0e-13
RK4	5.0e-13

Table 1: *Establishing the baseline maximum stable time step supported by internal FerroX and AMReX-based native runge-kutta methods.*

Verification and Validation

Before exploring novel SUNDIALS methods, the four native AMReX methods were validated through convergence tests, with a focus on each method's respective order of accuracy. Since the FerroX system is based on partial differential equations (PDEs), its behavior is dictated by three primary traits: consistency, stability, and convergence. Consistency represents how well a numerical method can approximate the initial PDE. Stability, for explicit numerical schemes, defines a particular criterion value past which the scheme no longer follows the expected error development; however, implicit schemes are unconditionally stable- they lack a stability criterion. This particular set of tests for FerroX were centered on investigating the error development, or convergence.

Input parameters represent a physical stack structure that is 32 nm on each side. The stack material composition is 25% semiconductor material (Si), 25% dielectric (SiO2), and 50% ferroelectric material (HZO). In our exploration of the dynamics of domain walls, polarization switching, and the negative capacitance effect, the polarization P is set to be uniform along one of the lateral dimensions (y) in a 3D FerroX simulation. This neutralizes the contribution from gradients in P and ϕ along the y direction, narrowing the scope of the problem to two dimensions within the x-z plane.

Our method utilized the L_2 norm to compute the error of the function from coarse to medium and medium to fine time resolution values, as depicted in Equation 2 and Equation 3 respectively. Both error values were used to compute the order of accuracy with the convergence rate, or Equation 1.

$$\text{Rate} = \log_2 \left(\frac{E_c^m}{E_m^f} \right) \tag{1}$$

$$E_c^m = \sqrt{\frac{1}{N_c} \sum_{i,j,k} |\phi_m - \phi_c|^2} \tag{2}$$

$$E_m^f = \sqrt{\frac{1}{N_m} \sum_{i,j,k} |\phi_f - \phi_m|^2} \tag{3}$$

The convergence tests were set up with coarse, medium, and fine time resolution values identical to those in the Numerical Convergence Study as demonstrated by Kumar et al. These tests utilize the AMReX-based *ComparePlotfiles* function, which returns a norm that represents the error from the coarse (dt = 5.0e-14) to medium (dt = 2.5e-14) time step and another that represents the error from the medium (dt = 2.5e-14) to fine (dt = 1.25e-14) time step. In an ideal scenario, the Rate approximates to either 1, 2, 3, or 4 for Forward Euler, Trapezoid, SSPR3, and RK4 respectively, depending on which time integration method is examined. As shown in Table 2, the convergence rates closely approximate each method's corresponding order of accuracy.

Method	**Order**	E_c^m	E_m^f	**Rate**
Forward Euler	1	4.01e-8	1.99e-8	1.005
Trapezoid	2	7.89e-10	1.82e-10	2.114
SSPRK3	3	4.27e-11	4.86e-12	3.135
RK4	4	2.08e-12	1.17e-13	4.150

Table 2: *Verification and validation of Native Explicit Runge-Kutta methods through order of accuracy tests that yield convergence rates.*

Exploring Advanced Methods

The SUNDIALS interface offers several different categories of time integration methods, including Explicit Runge-Kutta (ERK), Diagonally Implicit Runge-Kutta (DIRK), and Multi-rate functions. Table 3 shows examples of Diagonally Implicit Runge-Kutta (DIRK) methods with their maximum time step, and an evident doubling of the allowable time step. DIRK methods in particular yielded significant results, as we observe the *ARKODE_BILLINGTON_3_3_2* technique facilitating the doubling of the supported time step in FerroX. Although this is a significantly larger time step when compared with the capabilities of internal FerroX methods, the net performance benefit was offset by an extensive runtime– an observation that encourages the further investigation of multi-rate functions in cooperation with DIRK methods to reduce computational time.

Diagonally Implicit Runge Kutta (DIRK)	**Time Step**	**Runtime**
ARKODE_IMPLICIT_TRAPEZOID_2_2	6.0e-13	700
ARKODE_BILLINGTON_3_3_2	9.0e-13	630
ARKODE_IMPLICIT_MIDPOINT_1_2	6.0e-13	600

Table 3: *Maximum time step supported by advanced methods that are a part of the SUNDIALS interface.*

Sensitivity Analysis

To fully leverage the capabilities of SUNDIALS, we performed a series of sensitivity tests that reveal how the development of each variable individually affects the overall simulation stability. The purpose of sensitivity testing enabled us to determine the variables within a time integration equation that pose the greatest restriction to time step length. In the context of FerroX, sensitivity testing enables us to divide the TGDL equation into elements that are better suited for multi-rate SUNDIALS time integration functions rather than internally present methods.

Filtering the equation can improve the efficiency of the code by allocating the variables that develop faster than others to a more fitting time integration method. This procedure essentially allows the code to

'divide and conquer' complex equations in a way that can increase the maximum stable time step value. The method consisted of systematically setting different variables of the TDGL equation's right-hand-side (RHS) to zero individually as well as in combination, in order to determine which variables updated faster than the rest. This included the terms that represent the bulk Landau free energy, gradient energy, and electric energy of the applied electric field. Failed simulations entailed unreasonably large or small output values, and a 'checkerboard' image as visualized through VisIt. Successful simulations depicted repeating patterns, as shown in Fig. 4(a). The sensitivity testing enabled pinpointing the grad variables, g44 and g11 as the quickest developing terms. Multi-rate functions will be particularly useful in this scenario, where different components of the system have been observed to evolve at significantly different rates.

Conclusions and Future Work

Successfully implementing SUNDIALS support into the source code doubled the allowed maximum time step with DIRK methods. However, utilizing advanced functions alone (i.e. DIRK) did not yield significant results due to the extensive computational time of each step. Future work will focus on refactoring the source code to incorporate multi-rate functions from SUNDIALS. In allocating finer time resolution for rapidly developing terms and coarser resolution for others, FerroX performance can reap the benefit of a lengthened time step.

Acknowledgements

My mentors, Prabhat Kumar and Andy Nonaka, for involving me in their breakthrough research on low-power technology. This work was supported in part by the US Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program.

References

[1] D. J. Gardner, D. R. Reynolds, C. S. Woodward, C. J. Balos, Enabling new flexibility in the SUNDIALS suite of nonlinear and differential/algebraic equation solvers, ACM Transactions on Mathematical Software (TOMS) (2022). doi:10.1145/3539801.

[2] A. C. Hindmarsh, P. N. Brown, K. E. Grant, S. L. Lee, R. Serban, D. E. Shumaker, C. S. Woodward, SUNDIALS: Suite of nonlinear and differential/algebraic equation solvers, ACM Transactions on Mathematical Software (TOMS) 31 (3) (2005) 363–396. doi:10.1145/1089014.1089020.

[3] S. Salahuddin, S. Datta, Use of negative capacitance to provide voltage amplification for low power nanoscale devices, Nano Letters 8 (2) (2008) 405–410, pMID: 18052402. arXiv:https://doi.org/10.1021/nl071804g, doi:10.1021/nl071804g.
URL `https://doi.org/10.1021/nl071804g`

[4] P. Kumar, A. Nonaka, R. Jambunathan, G. Pahwa, S. Salahuddin, Z. Yao, Ferrox: A gpu-accelerated, 3d phase-field simulation framework for modeling ferroelectric devices, Computer Physics Communications 290 (2023) 108757.