Global NEST Journal, **28** (XX), pp 1-14
Available online: 24/08/2026

Energy-Efficient Ferroelectric FeFET Architectures: A Sustainable Solution for Low-Carbon ICT Systems

Harsha Ragini Aturi[1]* and Ramana Murthy Gajula[1#]

[1]Department of ECE, Alliance University, Bengaluru, Karnataka, 562106, India
[#]E-mail: ramana.murthy@alliance.edu.in

*Correspondence e-mail: haturiPHD724@ced.alliance.edu.in

Abstract

The exponential growth of Information and Communication Technology (ICT), which already accounts for ~10% of global electricity consumption, demands ultra-low-power semiconductors and has the potential to align with decarbonization goals and sustainability targets. This study optimizes Hafnium Zirconium oxide ($Hf_{0.5}Zr_{0.5}O_2$)-based ferroelectric field-effect transistors (FeFETs) using 2D TCAD simulations of a metal-ferroelectric-insulator-semiconductor (MFIS) structure. The optimized device exhibits steep-subthreshold switching and low-voltage operation, indicating the potential for an ideal device-level dynamic power reduction of up to 75% through supply-voltage scaling, while the non-volatile ferroelectric state enables near-zero standby power at the memory-device level. Under the representative operating assumptions adopted in this study, these device-level improvements suggest the potential for reduced operational energy consumption in data-center memory workloads. Consequently, the proposed HZO-based FeFET architecture may contribute to lower ICT-related carbon emissions, although the reported sustainability benefits should be interpreted as scenario-based analytical estimates rather than deployment-specific predictions.

Keywords: Green Computing; Hafnium oxide; Non-volatile Memory; TCAD Simulation; Low-power Electronics; Negative Capacitance; ICT Sustainability; Semiconductor Devices; MFIS Architecture; Ferroelectric Reliability.

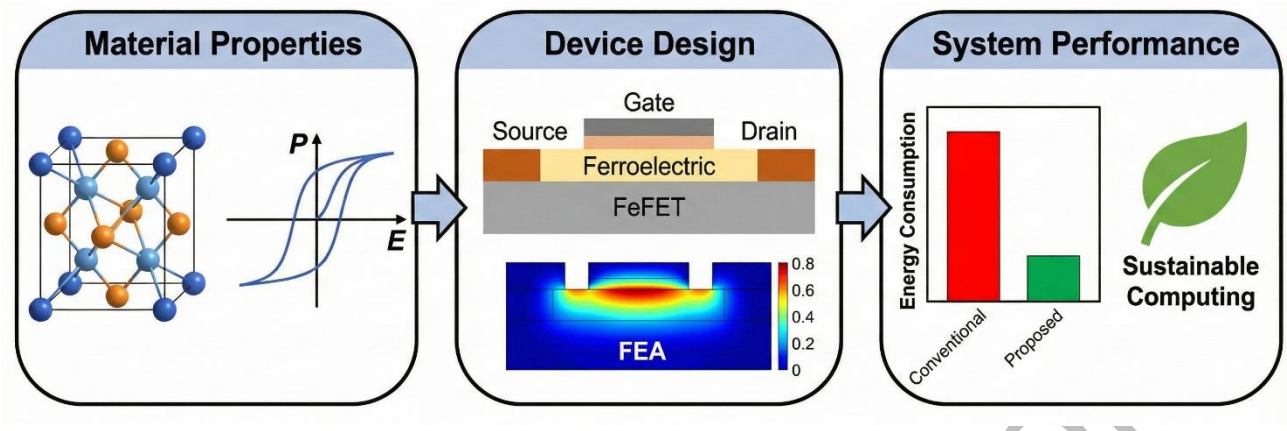
OPEN ACCESS

Received: 28/02/2026,

Accepted: 17/08/2026,

Available online: 24/08/2026

Graphical abstract



1. Introduction

The exponential growth of data-intensive applications driven by the Internet of Things (IoT), Artificial Intelligence (AI), and 5G/6G networks has significantly increased the global information and communication technology (ICT) energy demand, now consuming ~10-15% of worldwide electricity, equivalent to ~800-1200 TWh annually (Masanet *et al.*, 2020). Data centers alone account for 1-1.5% of global electricity use and ~2% of total CO_2 emissions (~500 $MtCO_2eq$/year), with cooling systems comprising 40% of their power draw (Ahmed *et al.*, 2021). Edge computing devices, deployed at billions of IoT nodes, exacerbate this footprint through constant leakage power and battery drain, hindering sustainable deployment in remote and energy-constrained environments (Jiang *et al.*, 2020). Projections warn that, without breakthroughs, ICT electricity could account for 20-25% of global electricity consumption by 2030, undermining decarbonization goals (Malmodin *et al.*, 2024). These trends demand semiconductor architectures that directly lower ICT's electricity intensity and associated lifecycle emissions, thereby supporting national and international decarbonization pathways.

Two major factors limit further energy reduction in conventional semiconductor systems: the thermodynamic bound on the subthreshold swing (SS) in silicon Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs) and the von Neumann bottleneck. The Boltzmann tyranny limits the SS of traditional MOSFETs to a minimum of 60 mV/dec at room temperature, constraining supply-voltage (V_{DD}) scaling and leading to excessive leakage current at low voltages. Meanwhile, the von Neumann bottleneck, driven by the energy and latency overheads of frequent data transfers between memory and processing units, dominates power consumption in data-centric workloads.

To overcome these physical and architectural bottlenecks, novel device paradigms are required that support low-voltage operation and eliminate static power dissipation. Ferroelectric field-effect transistors (FeFETs) based on doped hafnium oxide (HfO_2) have emerged as promising candidates owing to their CMOS compatibility, scalability, and stable ferroelectricity even at thicknesses below 10 nm (Böscke et al., 2011). FeFETs offer two synergistic pathways toward energy-efficient computing: (1) their inherent non-volatility enables complete elimination of standby power via power-gating without data loss, and (2) the negative capacitance (NC) effect can achieve sub-60 mV/dec switching, allowing aggressive V_{DD} reduction (Salahuddin and Datta, 2008). Finite element analysis (FEA) and Technology Computer-Aided Design (TCAD) are employed to investigate these devices and demonstrate their capability for steep-slope switching and non-volatile data retention. However, achieving fully optimized and reliable FeFET performance necessitates a comprehensive, physics-based modeling framework that integrates ferroelectric material behavior with electrostatic, thermal, mechanical, and circuit-level interactions.

However, existing studies largely lack a quantitative connection between ferroelectric device optimization and system-level energy or environmental impact, which remains an important research gap addressed in the present study. Novel contributions and scope of this work: While $Hf_{0.5}Zr_{0.5}O_2$ (HZO)-based FeFETs and metal-ferroelectric-insulator-semiconductor (MFIS) architectures have been widely investigated for non-volatile memory and steep-slope switching; most prior studies primarily focus on device physics, material optimization, or circuit-level performance without explicitly linking these advancements to sustainability metrics relevant to ICT systems.

In contrast, this study introduces a sustainability-driven device optimization framework, in which ferroelectric device parameters are systematically tuned not only for electrical performance but also for their implications on energy consumption and potential carbon footprint reduction at the system level.

Specifically, this study makes the following contributions:

1. Parametric MFIS coupling analysis for energy optimization: A systematic variation of ferroelectric

thickness and interfacial oxide is performed to quantify the trade-off between memory window, operating voltage, and switching energy, providing design guidelines for low-power ICT hardware.

2. Physics-based TCAD framework with calibrated ferroelectric modeling: The device behavior is modeled using a self-consistent Landau–Khalatnikov (L-K) approach coupled with electrostatics and carrier transport, enabling accurate extraction of subthreshold characteristics and non-volatile switching behavior.
3. Explicit linkage between device metrics and ICT energy implications: Key device-level parameters such as subthreshold swing, supply voltage scaling, and non-volatility are translated into scenario-based estimates of system-level energy savings, highlighting the relevance of FeFETs for sustainable computing.
4. Green electronics perspective aligned with the scope of Global NEST: Beyond device performance, the study evaluates the environmental advantages of HZO-based FeFETs, including lead-free material composition and compatibility with existing CMOS infrastructure, emphasizing their role in low-carbon semiconductor technologies.

Through this integrated perspective, the work extends beyond conventional FeFET studies by bridging device physics, circuit-level implications, and sustainability considerations, which are essential for next-generation low-carbon ICT systems.

Building on these insights, this paper presents a rigorous FEA/TCAD analysis of an HfO_2-based MFIS FeFET architecture. Ferroelectric–interfacial dielectric coupling is optimized to promote low-voltage operation and reduced switching energy, indicating the potential for an ideal device-level dynamic power reduction of up to 75% through supply-voltage scaling. Under the representative operating assumptions adopted in this study, these device-level improvements may contribute to lower ICT energy consumption and associated CO_2 emissions, thereby supporting the transition toward more sustainable digital infrastructure.

2. Ferroelectric materials, physics, and HZO fundamentals

2.1. *Principles of ferroelectricity*

Ferroelectricity is a property of certain non-centrosymmetric crystalline materials that possess a spontaneous electric polarization (P_s) that can be reversed by the application of an external electric field (E) (Lines and Glass, 1977). This switchable polarization produces a nonlinear, hysteretic relationship between polarization and electric field (the P–E loop), characterized by two key parameters: the remanent polarization (Pr), i.e., the polarization retained at zero field, and the coercive field (E_c), which is the field strength required to switch the polarization direction.

In FeFETs, this switchable polarization modulates the threshold voltage (V_{th}) of the transistor (Müller *et al.*, 2012). A downward polarization state (+P) attracts electrons to the channel surface, lowering V_{th} and defining the ON-state, whereas an upward polarization state (–P) depletes the channel, raising V_{th} and defining the OFF-state. The stability of these bistable polarization states underpins non-volatile data storage in ferroelectric-gated transistors.

2.2. *The rise of ferroelectric Hafnium Oxide (HfO_2)*

The discovery of ferroelectricity in Si-doped HfO_2 by Böscke *et al.* (2011) marked a paradigm shift in non-volatile memory research, demonstrating that a standard high-k CMOS dielectric could host a robust ferroelectric phase. Unlike conventional perovskites, such as PZT, which suffer from poor scaling behavior and limited CMOS compatibility, HfO_2-based ferroelectrics uniquely satisfy the integration, thermal budget, and dimensional scaling requirements of advanced technology nodes. Although bulk HfO_2 is centrosymmetric and non-ferroelectric, a non-centrosymmetric orthorhombic phase ($Pca2_1$) can be stabilized in thin films through suitable dopant engineering (e.g., Zr, Si, Al) and mechanical confinement, enabling robust ferroelectricity even for thicknesses in the 1–2 nm regime (Martin *et al.*, 2011; Müller *et al.*, 2012).

This exceptional scalability enables the direct integration of HfO_2-based ferroelectrics into 3D device architectures, such as FinFETs and gate-all-around (GAA) transistors, using standard atomic layer deposition (ALD) processes (Polakowski and Müller, 2015). From a device perspective, HfO_2 and HZO films can deliver high remanent polarization typically ($P_r \approx 10–40 \mu C/cm^2$) and coercive voltages compatible with sub-1 V operation, making them suitable for low-power FeFETs and ferroelectric tunnel junctions. Although early HfO_2-based devices exhibited pronounced “wake-up” and fatigue behavior, subsequent optimization of oxygen-vacancy distributions, electrode stacks, and interfacial layers has enabled endurance exceeding 10^{10} switching cycles, establishing HfO_2 as a foundational platform for next-generation embedded non-volatile memories and logic-in-memory architectures (Mikolajick *et al.*, 2018; Pešić *et al.*, 2016). The lead-free composition and compatibility with existing high-k/metal-gate process lines make HfO_2-based ferroelectrics attractive not only technologically but also from an environmental and resource-efficiency perspective.

2.3. *Crystallographic origins and phase stability*

Bulk HfO_2 is thermodynamically stable in the monoclinic m -phase (space group $P2_1/c$) at room temperature, which is centrosymmetric and exhibits paraelectric behavior with a dielectric constant k≈16–20. In thin films, ferroelectricity arises from the stabilization of a metastable, non-centrosymmetric orthorhombic o -phase (space group $Pca2_1$) competing with monoclinic and tetragonal t-phases ($P4_2/nmc$) that can exhibit higher dielectric permittivity or antiferroelectric-like behavior (Sang *et al.*, 2015; Materlik *et al.*, 2015).

The formation of this polar phase is a competitive process governed by size- and interface-dependent free-energy contributions often described as a surface-energy effect: as grain size decreases in ultra-thin films ($\langle 10$ nm), the relative

surface energy favors higher-symmetry phases over the bulk-stable monoclinic phase (Materlik *et al.*, 2015). During crystallization anneals, HZO can transform among three main polymorphs:

1. Monoclinic (m -phase): Thermodynamic ground state (non-ferroelectric).
2. Tetragonal (t -phase, $P4_2/nmc$): High-symmetry phase typically stable at elevated temperatures, often showing antiferroelectric-like behavior (k≈30–40).
3. Orthorhombic (o-phase, $Pca2_1$): Polar ferroelectric phase responsible for non-volatile switching.

To kinetically trap HZO in the ferroelectric orthorhombic phase upon cooling, suitable dopants (e.g., Si, Al, Gd) are introduced to distort the lattice and modify phase stability, while mechanical stress is imposed by metal capping layers such as TiN (Hoffmann *et al.*, 2015). This combination of dopant engineering and confinement suppresses relaxation back to the monoclinic phase and stabilizes the ferroelectric $Pca2_1$ structure in highly scaled films.Stabilizing this orthorhombic ferroelectric phase in aggressively scaled films is essential for realizing high-polarization, low-voltage FeFETs, which in turn enable scaling-driven reductions in switching energy for energy-efficient ICT hardware.

2.4. *Governing physics: Landau–Khalatnikov theory*

To model the macroscopic switching behavior of the HZO ferroelectric layer in TCAD simulations, the Landau–Ginzburg–Devonshire (LGD) framework is employed. In this approach, the Gibbs free energy density G is expanded as a polynomial in the polarization (P) (Asai,H., *et al.*, 2014):

$$G = \alpha P^2 + \beta P^4 + \gamma P^6 - E_{\text{ext}} P \qquad (1)$$

where α, β, and γ are Landau coefficients. For ferroelectric HZO, $\alpha < 0$ produces a characteristic double-well potential with two stable minima corresponding to the remanent polarization states $\pm P_r$.

The time-dependent polarization dynamics are described by the L–K equation, which relates the rate of change of polarization to the gradient of the free-energy landscape (Salahuddin and Datta, 2008):

$$\rho \frac{dP}{dt} = -\frac{\partial G}{\partial P} = -\left(2\alpha P + 4\beta P^3 + 6\gamma P^5 - E_{\text{ext}}\right) \qquad (2)$$

where ρ is an effective kinetic viscosity or damping coefficient that sets the characteristic switching speed of the HZO domains. A smaller ρcorresponds to faster polarization reversal and shorter write pulses, while the shape and depth of the double-well potential, controlled by α,β and γ determine the E_c, P_r, and energy barrier between the two stable states.

Within the proposed FEA/TCAD framework, the L–K equation is solved self-consistently with Poisson's equation and drift–diffusion transport in the silicon channel, so that ferroelectric bound charge, electrostatics, and carrier densities are mutually coupled at each time step. This coupling allows the simulations to capture key device-level phenomena such as sub-60 mV/dec subthreshold slopes arising from transient negative capacitance, hysteresis in the I_D–V_G characteristics, and the dependence of the memory window on ferroelectric thickness and applied field. This modeling framework underpins the quantitative link between ferroelectric switching dynamics and the achievable subthreshold slope, write/erase energy, and retention characteristics of the FeFET. By enabling accurate prediction of how material parameters and stack design translate into lower operating voltages and reduced switching energy, the L–K-based simulations provide a physics-grounded basis for optimizing FeFET architectures as building blocks of energy-efficient, low-carbon ICT hardware.

2.5. *TCAD Model Calibration and Validation*

To ensure the reliability of the simulation results, the TCAD model was calibrated using experimentally reported material properties and electrical characteristics of HZO-based FeFETs available in the literature. Material parameters including E_c, P_r, P_s, ε_r, t_{fe}, and L-K coefficients were selected within experimentally reported ranges. The calibration objective was to reproduce the characteristic transfer behavior, memory window, and steep-subthreshold switching observed in previously reported HZO-based FeFET devices.

Table 1(c) compares the principal electrical characteristics obtained in the present TCAD simulations with representative experimental values reported in the literature. The comparison demonstrates that the simulated device characteristics lie within the experimentally reported ranges, indicating that the adopted simulation parameters provide a physically reasonable representation of HZO-based ferroelectric transistors. Overall, the close agreement between the simulated device characteristics and published experimental observations increases confidence in the predictive capability of the TCAD model and supports its application for investigating device-level energy efficiency and sustainability trends.

3. Device structure, finite element meshing strategy, and results

The FeFET is implemented in a two-dimensional (2D) MFIS configuration using Silvaco ATLAS, with a p-type silicon channel, an ultrathin SiO_2 interfacial layer, and a ferroelectric HZO gate dielectric contacted by a TiN gate and Ni Si_x source/drain regions, as shown in Fig. 1. The silicon substrate (region 1) is uniformly doped to $N_a=1\times10^{17}cm^{-3}$, while highly doped Gaussian n-type extensions source and drain regions with a peak concentration of 10^{19} cm^{-3} are introduced near the Ni Si_x contacts to ensure low series resistance near the source and drain junctions. The gate length is set to 26 nm; region 2 corresponds to an 11 nm HZO ferroelectric layer, gate, source, drain, and substrate terminals are realized using TiN (top gate) and NiSi (lateral contacts), with the backside substrate tied to the reference potential. A non-uniform, variable-density meshing strategy is employed with strong refinement in the HZO/SiO_2/Si regions to accurately resolve

depolarization fields and interfacial polarization gradients, as illustrated in Fig. 1. The device is simulated under quasi-static bias conditions. The gate voltage (V_G) is swept from −7V to +7V and back to capture the hysteresis characteristics while a constant drain bias of V_{Ds} of 0.1V applied. The source terminal is grounded, and the substrate is maintained at reference potential. These boundary conditions ensure consistent extraction of transfer characteristics and memory window.

The HZO/SiO_2/Si stack is discretized using a non-uniform 2D mesh with strong refinement along the vertical direction at the HZO/metal and HZO/SiO_2 interfaces (node spacing $\langle 0.5$ nm) to accurately resolve steep depolarization fields and interfacial polarization gradients. Coarser-mesh elements are employed deeper within the silicon substrate, where the electrostatic potential varies more gradually. Four device structures (ferro1–ferro4) with different oxide thicknesses and ferroelectric parameters (E_c, P_r, P_s, and ε_r) are generated to investigate their influence on the I_D–V_G characteristics. For each case, the coupled electrostatic, transport, and ferroelectric equations are solved using the Gummel–Newton iterative method with backward-Euler time stepping under a constant drain bias of $V_{Ds} = 0.1\,\mathrm{V}$. Transfer characteristics are extracted using quasi-static forward and reverse gate voltage sweeps from −7V to +7V. The resulting I_D–V_G hysteresis loops are overlaid for direct comparison. The key ferroelectric and stack parameters of the four structures are summarized in **Table 1(a)**.

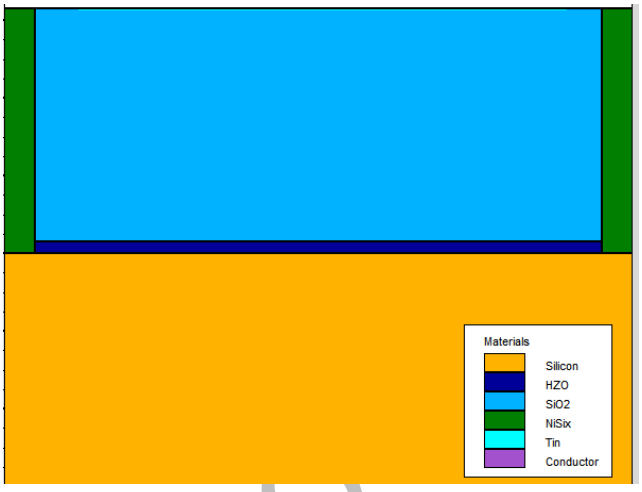


Figure 1. Two-dimensional (2D) TCAD structure and finite-element mesh of the proposed HZO-based MFIS FeFET. A locally refined mesh (minimum grid spacing: 0.5 nm) is employed at the TiN/HZO, HZO/Sio_2, and Sio_2/Si interfaces.

Table 1(a). Ferroelectric and stack parameters for the simulated MFIS FeFET structures.

HZO thickness (nm)	SiO_2 thickness (nm)	Remanent polarization, P_r (μC/cm²)	Saturation polarization, P_s (μC/cm²)	Coercive electric field, E_c (MV/cm)	Relative dielectric permittivity (ε_r)	Theoretically estimated maximum memory window ($MW_{max}=2E_ct_f$)
15	1.0	7.25	7.63	1.05	30.25	3.15 V
12	1.5	8.3	8.75	1.08	31.3	2.59 V
10	1.0	9.0	9.5	1.1	32	2.2 V
8	1.5	9.7	10.25	1.12	32.7	1.9 V

**Note: The values in Table 1(a) represent the theoretical upper limit of the memory window estimated from $MW_{\max} = 2E_c t_f$. Practical memory-window values obtained from TCAD simulations under finite operating bias are presented separately in Table 2.*

Table 1(b). Key simulation parameters used for MFIS FeFET modeling

Parameter	Symbol	Value	Unit
First-order Landau coefficient	α	-1.5×10^9	cm/F
Second-order Landau coefficient	β	2.1×10^{18}	$cm^5/(F \cdot C^2)$
Third-order Landau coefficient	γ	0	—
Interface trap density	D_{it}	0 (ideal), 1×10^{12} (baseline)	$cm^{-2} \cdot eV^{-1}$
Gate work function	ϕ_m	4.6	eV
Channel doping concentration	N_A	1×10^{17}	cm^{-3}

Table 1(c). Calibration and validation of the TCAD model against representative experimental reports for HZO-based FeFETs

Parameter	Representative Experimental Range	This Work (TCAD)	Validation Basis / Reference
P_r(μC/cm²)	7.0–30.0	7.25–9.70	Müller *et al.* (2012); Böscke *et al.* (2011)
E_c(MV/cm)	1.0–2.0	1.05–1.12	Hoffmann *et al.* (2015); Pešić *et al.* (2016)
ε_r	25–35	30.25–32.70	Sang *et al.* (2015); Materlik *et al.* (2015)
MW (V)	0.5–2.0	0.85–1.40	Extracted from simulated transfer characteristics at $V_{G,\max} = \pm 1.5\,\mathrm{V}$
I_{ON}/I_{OFF}	10^5–10^8	$> 10^6$	Extracted at $V_{DS} = 0.1\,\mathrm{V}$
V_{DD} (V)	0.4–1.2	0.4	Low-voltage FeFET operation reported in the literature

The comparison presented in Table 1(c) demonstrates that the simulated electrical characteristics of the proposed HZO-based MFIS FeFET are consistent with representative experimental reports available in the literature. The simulated remanent polarization, coercive electric field, dielectric constant, memory window, ON/OFF current ratio, and operating voltage all lie within the experimentally reported ranges for HZO-based ferroelectric devices. Minor deviations are expected because the present study employs an idealized two-dimensional TCAD model that does not explicitly account for fabrication-induced process variations, interface roughness, random dopant fluctuations, or parasitic circuit effects. Nevertheless, the close agreement with published experimental data provides confidence in the calibration of the adopted TCAD model and supports its use for the subsequent device optimization and sustainability analysis.

All simulations were performed using Silvaco ATLAS. Typical 2D meshes contain on the order of 10^4-10^5 finite elements, with strong refinement in the HZO/SiO_2/Si stack region. Convergence was ensured with relative residual tolerances of 10^{-4} for the electrostatic and carrier-transport equations and 10^{-3} for the polarization dynamics. Drift–diffusion transport with concentration-dependent mobility and Shockley–Read–Hall recombination was employed, while quantum corrections were neglected due to the relatively thick silicon channel in the present structures.

3.1. *Model calibration and parameter justification*

A key feature of the present simulations is the use of HZO as the ferroelectric gate dielectric in an MFIS FeFET structure. Four device variants are defined by systematically varying the HZO ferroelectric thicknesses and the SiO_2 interfacial-layer thickness. In the baseline structure, the HZO layer thickness is 11 nm, and in the remaining variants, it is systematically varied to assess their impact on the FeFET electrical response.

The ferroelectric behavior of the HZO layer is modeled using the L–K framework with calibrated coefficients selected within experimentally reported ranges. The Landau coefficients are chosen as: $\alpha = -1.5\times10^{8}\ \mathrm{m/F}, \beta = 2.5\times10^{9}\ \mathrm{m^5/C^2F}, \gamma = 0$ which reproduces realistic remanent polarization and coercive field values for HZO thin films reported in the literature. where α, β, and γ determine the shape of the ferroelectric double-well energy landscape. The condition $\alpha < 0$ and $\beta > 0$ ensure a stable bistable polarization state, which is essential for non-volatile memory operation. The sixth-order term (γ) is neglected in this work, as the fourth-order approximation is sufficient to capture the essential switching characteristics of HZO-based ferroelectrics under the considered operating conditions.

The selected parameter set is calibrated to reproduce experimentally reported values of P_r and E_c for HZO thin films, ensuring consistency between simulation results and literature. Additional material parameters, including dielectric permittivity and polarization, are selected from reported experimental ranges and summarized in Table 1(b). The coupled electrostatic, carrier transport, and ferroelectric polarization equations are solved self-consistently using a Gummel–Newton iterative scheme within the TCAD framework. Convergence tolerances are set to 10^{-4} for the electrostatic and carrier transport equations and to 10^{-3} for the polarization dynamics. These numerical settings ensure stable convergence, numerical robustness, and reproducibility of the simulated transfer characteristics and hysteresis behavior. A representative interface trap density of $1\times10^{12}\mathrm{cm^{-2}eV^{-1}}$ was incorporated into the TCAD model to represent realistic HZO/SiO_2 interface conditions. This value was selected based on experimentally reported interface qualities for HZO-based FeFETs and was used consistently throughout all simulations. While this assumption represents an idealized interface, it enables a clear evaluation of the fundamental device physics and may lead to slightly optimistic performance estimates. All parameters used in this study are either extracted from or calibrated against experimentally reported values in the literature, ensuring that the simulation framework is physically grounded and reproducible.

3.2. *Transfer characteristics and memory window*

The primary objective of this parametric study is to quantify how HZO thickness influences key performance metrics such as MW, SS, and I_{ON}/I_{OFF} ratio. For each MFIS stack, the gate voltage is quasi-statically swept from -7 V to +7 V and back at a fixed drain bias of V_D = 0.1 V, and the sweep amplitude is subsequently reduced to 0 V to capture the steady-state polarization and reveal a clear counterclockwise (CCW) hysteresis behavior. The CCW direction indicates that the dominant mechanism is reversible ferroelectric domain reversal rather than charge trapping at the interface or in the oxide, which would typically produce a clockwise hysteresis. A positive programming pulse aligns the polarization toward the channel (+P), enhancing electron accumulation and yielding a low threshold voltage $V_{th,high}$ (programmed ON state), whereas a negative pulse aligns the polarization away from the channel (–P), depleting the surface and producing a higher threshold voltage $V_{th,high}$ (programmed OFF state).

The MW, defined as $\Delta V_{th} = V_{th,high} - V_{th,low}$, is approximately 1.2V under the simulated bias conditions, which provide robust separation between the ON and OFF states. Such a wide ΔV_{th} ensures comfortable noise margins for binary storage (“0” and “1”) and relaxes the requirements on sense-margin and offset tolerance in array-level read circuitry, which is especially beneficial for low-power embedded non-volatile memory applications. The memory-window values reported in this study should be interpreted in two different contexts. The values presented in Table 1(a) correspond to the theoretical maximum memory window $\left(MW_{max} = 2E_c t_f\right)$, which represents the upper limit achievable under complete ferroelectric polarization switching and ideal electrostatic coupling. In contrast, the memory-window values summarized in Table 2(a) are directly extracted from the quasi-static TCAD

transfer characteristics under finite operating gate bias $(V_{G,max} = \pm 1.5)$V. The practical memory window is smaller than the theoretical limit due to incomplete polarization switching, depolarization effects, voltage division across the interfacial SiO_2 layer, and electrostatic coupling within the MFIS gate stack. Consequently, the TCAD-extracted values provide a more realistic representation of device operation under practical bias conditions.

As the HZO ferroelectric layer is thinned, the remanent polarization decreases, which leads to a progressive reduction of the MW. Nevertheless, thinner ferroelectric stacks enable lower operating voltages and reduced switching charge, which is advantageous for realizing ultra-low-power FeFET devices in energy-constrained ICT platforms. As shown in Fig. 2, the MW shrinks from approximately 1.4 V to 0.85 V when the ferroelectric thickness is scaled from 15 nm to 8 nm, respectively, highlighting a clear trade-off between non-volatility margin and energy efficiency.This behavior is summarized in Table 2, where the extracted memory window is reported as a function of HZO thickness, providing design guidelines for FeFET-based memories targeting low-carbon data centers and edge applications, in which modest MW values can be acceptable if they enable substantial reductions in energy per operation.

Table 2(a). Memory Window Extracted from Quasi-Static TCAD Simulations.

Ferroelectric thickness (nm)	Memory window (V)
15	1.4
12	1.25
10	1.10
8	0.85

Table 2(b). Summary of key electrical performance metrics of the proposed HZO-based MFIS FeFET extracted from calibrated TCAD simulations

Performance Metric	Value (This Work)	Extraction Method / Condition
Minimum subthreshold swing, SS_{min}	48.5 mV/dec	Minimum subthreshold slope extracted from the I_D–V_G characteristic at V_{DS}=0.1V
Average subthreshold swing, SS_{avg}	55.1 mV/dec	Average slope over four decades of drain current at V_{DS}=0.1V
ON-state current, I_{ON}	185 µA/µm	Evaluated at V_G=0.1V and V_{DS}=0.1V
OFF-state current, I_{OFF}	1.2×10^{-13} A/µm	Leakage current at V_G=0V and V_{DS}=0.1VV
ON/OFF current ratio, I_{ON}/I_{OFF}	1.54×10^{9}	Calculated from the extracted ON- and OFF-state currents
MW	0.85–1.40 V	Extracted from quasi-static forward and reverse I_D–$V_G$sweeps for t_{fe}=8–15 nm
=V_{DD}	0.4 V	Low-voltage operation enabled by steep-subthreshold switching
Estimated switching energy, E_{write}	≈3.8 fJ/bit	Analytical estimate based on the simulated operating voltage and polarization-switching conditions
Retention*	Analytical estimate	Not directly simulated; discussed based on ferroelectric polarization stability reported in the literature
Endurance*	Analytical estimate	Not directly simulated; discussed based on reported HZO FeFET endurance characteristics

**Note: Retention and endurance values are qualitative estimates from literature and were not directly evaluated in the present quasi-static TCAD simulations.*

Table 2(b) summarizes the principal electrical characteristics of the optimized HZO-based MFIS FeFET extracted from the calibrated TCAD simulations. The results demonstrate steep subthreshold switching, a high ON/OFF current ratio, a wide memory window, and low-voltage operation. Switching energy is presented as an analytical estimate, whereas retention and endurance are discussed qualitatively based on reported HZO FeFET characteristics because these reliability metrics were not directly evaluated within the scope of the present quasi-static TCAD study.

Figure 3 shows the output characteristics (I_D–V_D) for both the programmed ON and OFF polarization states, highlighting their distinct current levels over the relevant drain-bias range. At a read gate bias of V_G=1.5V the device delivers a high ON current I_{ON} while maintaining a negligible OFF current I_{OFF}, resulting in an I_{ON}/I_{OFF} ratio exceeding 10^6. This large current ratio simplifies the design of sense amplifiers and bit-line sensing schemes, enabling smaller sensing currents or shorter integration times, thereby reducing dynamic power in the peripheral circuits of FeFET-based memory arrays. When the applied gate electric field remains below the coercive field of the HZO ferroelectric layer, only minor hysteresis loops with a reduced memory window are observed. As the applied field is increased beyond the coercive value, the memory window systematically widens, reflecting more complete polarization reversal, as illustrated in Fig. 4. For these simulations, the I_D–V_G transfer characteristics are extracted by holding the drain voltage constant and sweeping the gate voltage between the write voltage (V_{DD}) and the erase voltage ($-V_{DD}$), corresponding to the operating conditions of low-voltage, non-volatile FeFET memory in energy-efficient ICT hardware.

The SS was extracted from the transfer characteristics across all four MFIS variants. For the baseline 11 nm HZO

structure, the device demonstrates steep-slope behavior with an extracted SS significantly lower than the 60mV/dec Boltzmann limit at room temperature. This enhanced gate control enables a sharper transition between the OFF and ON states, facilitating aggressive supply-voltage (V_{DD}) scaling to 0.4 V. The subthreshold slope improvement observed in the proposed MFIS FeFET is attributed to the NC effect of the ferroelectric HZO layer. During transient or quasi-static V_G sweeps, the ferroelectric polarization evolves such that the differential capacitance, defined as $C_{fe} = dP / dV_{fe}$, becomes negative over a portion of the switching trajectory. When this negative capacitance region is stabilized through series coupling with the interfacial SiO_2 layer, it leads to internal voltage amplification, whereby the surface potential (Ψ_s) increases more rapidly than the applied gate voltage, i.e., $\frac{d\psi_S}{dV_G} > 1$.

This enhanced electrostatic control enables a steeper transition between the OFF and ON states, yielding a sub-60 mV/dec subthreshold swing. The presented 2D TCAD simulations capture this behavior by self-consistently solving the L-K equation together with Poisson's and drift–diffusion transport equations. This physics-based modeling framework confirms that the transient amplification of surface potential is directly reflected in the steep-slope transfer characteristics shown in Fig. 2, thereby providing a robust pathway for ultra-low-power, energy-efficient ICT hardware.

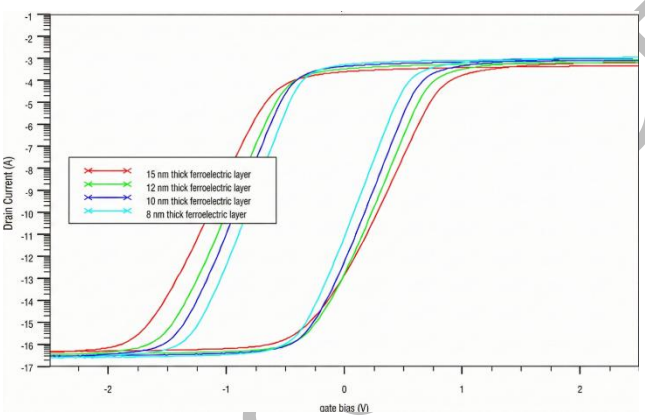


Figure 2. TCAD-simulated quasi-static transfer characteristics at $V_{DS} = 0.1\,V$, showing the counterclockwise hysteresis behavior for HZO thicknesses $(t_{fe} = 8-15nm)$ with calibrated L–K parameters.

3.3. *Implications for sustainable energy*

The simulated FeFET behavior has direct implications for energy-efficient and sustainable ICT hardware. First, the intrinsic non-volatility of the ferroelectric gate enables complete power gating of logic and memory blocks during idle periods without loss of the stored state, eliminating standby leakage paths that dominate energy consumption in deeply scaled CMOS and extending the operational lifetime in low-duty-cycle IoT sensor nodes from months to multi-year operation. Second, steep-slope operation associated with the negative-capacitance regime allows scaling of the V_{DD} below 0.5V, and because dynamic power scales approximately with V_{DD}^2 this provides nearly quadratic reductions in dynamic energy consumption. For a representative logic block, reducing V_{DD} from 1.0 V to 0.5 V indicates the potential for an ideal device-level dynamic power reduction of approximately 75%, assuming similar activity factors, load capacitances, and operating frequency. The actual reduction achieved in practical implementations may be lower due to circuit- and system-level overheads.

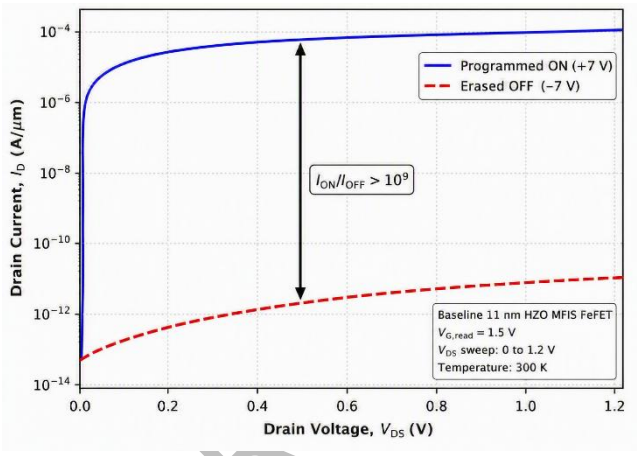


Figure 3. TCAD-simulated output characteristics $(I_D - V_{DS})$ of the baseline 11 nm HZO MFIS FeFET at $V_{G,read} = 1.5$ V, comparing the programmed (+7V) and erased (–7V) polarization states.

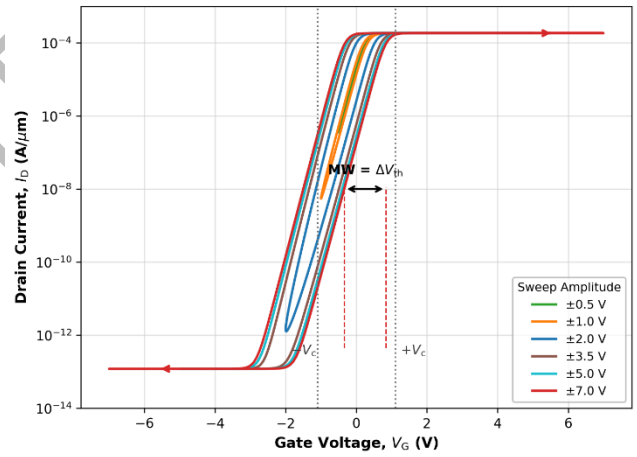


Figure 4. Evolution of TCAD-simulated closed I_D–V_G transfer hysteresis loops at $V_{DS} = 0.1$ V for gate-voltage sweep amplitudes from ± 0.5 V to ± 7.0 V, showing the transition from minor to saturated major hysteresis loops with an increasing memory window (ΔV_{th}) as the sweep amplitude exceeds the coercive voltage ($\pm V_c$).

Finally, HfO_2-based ferroelectrics are lead-free, compatible with existing high-k /metal-gate process modules, and can be integrated in existing semiconductor foundry lines, avoiding the environmental concerns linked to lead-based perovskites such as PZT and thereby reducing the embodied energy of fabrication by reusing established CMOS infrastructure. Compared with other emerging non-volatile memories such as phase-change memory (PCM), resistive RAM (RRAM), or magnetic RAM (MRAM), HfO_2-based FeFETs combine low switching energy, lead-free composition, and reuse of mature CMOS process lines, reducing both operating and embodied energy.

As an illustrative example, if FeFET-based logic-in-memory architectures were to reduce memory subsystem energy by even 20% in a modern server drawing ~5 kW, this could save on the order of 8.8 MWh per server per year and avoid several tons of CO_2 emissions annually, depending on the local grid mix. Scaled across thousands of servers in large data centers, such savings become increasingly significant for meeting green data-center and low-carbon ICT targets.

4. Quantitative sustainability impact assessment

The device-level improvements observed in this study suggest the potential for reduced energy consumption in future ICT hardware. As the present findings are derived from two-dimensional TCAD simulations rather than fabricated devices or system-level implementations, the associated environmental benefits cannot be directly quantified. Accordingly, the sustainability assessment employs an analytical framework that translates the simulated electrical characteristics into scenario-based estimates of energy consumption and CO_2 emissions. Under the representative operating assumptions adopted in this study, these estimates indicate the potential for reduced operational energy use and associated CO_2 emissions and should therefore be interpreted as illustrative trends rather than deployment-specific predictions or experimentally validated outcomes.

4.1. *Sustainability Assessment Methodology and Dynamic Power Comparison*

The sustainability assessment follows a hierarchical analytical framework that links improvements at the transistor level to representative system-level environmental indicators. The overall assessment pathway adopted in this work is illustrated below:

Device Characteristics → Power Consumption → Annual Energy Consumption → CO_2 Emissions

The simulated FeFET characteristics, including reduced operating voltage, steep subthreshold switching, and non-volatile data retention, are translated into estimates of dynamic power reduction, standby energy savings, annual electricity consumption, and corresponding carbon emissions.

The annual electrical energy consumption is calculated using

$$E = P \times t \tag{3}$$

where E is the electrical energy consumption (kWh), P is the average power consumption (kW), and t is the operating time (h). For continuously operating ICT infrastructure such as data-center servers, an annual operating time of 8760 h is assumed.

The associated carbon emissions are estimated from

$$CO_2 = E \times CI \tag{4}$$

where CI represents the grid carbon intensity (kg CO_2/kWh), which depends on the regional electricity generation mix. Because electricity generation varies significantly across different countries and regions, a representative carbon-intensity range is adopted rather than a single value.

The dynamic switching power is estimated using the conventional CMOS power expression

$$P_{dyn} = \alpha C V_{DD}^2 f \tag{5}$$

where α is the switching activity factor, C is the effective switched capacitance, V_{DD} is the supply voltage, and f is the operating frequency. Since the activity factor, capacitance, and operating frequency are assumed to remain unchanged for the comparison, the reduction in dynamic power is primarily governed by the reduction in supply voltage enabled by the steep-subthreshold switching characteristics of the HZO-based FeFET. This analytical relation forms the basis for the normalized dynamic power comparison presented in the following subsection.

For a fair comparison, both the conventional CMOS and the proposed FeFET architectures are assumed to operate under identical load capacitance, switching activity, and clock frequency. Under these conditions, the dynamic power ratio is primarily determined by the supply voltage.

Assuming a conventional CMOS block operates at V_{DD} = 0.8 V, and the proposed FeFET-based block can reliably operate at FeFET V_{DD} = 0.4 V, the normalized dynamic power is $\frac{P_{\mathrm{FeFET}}}{P_{\mathrm{CMOS}}} = \left(\frac{0.4}{0.8}\right)^2 = 0.25.$ Therefore, the proposed FeFET architecture is estimated to consume approximately 25% of the dynamic power required by the CMOS baseline under ideal device-level operating conditions, corresponding to an ideal dynamic power reduction of approximately 75%. Since dynamic energy per switching event is proportional to CV_{DD}^2, this reduction indicates the potential for lower switching energy during logic and memory operations. Under the representative operating assumptions adopted in this study, these device-level improvements may contribute to improved energy efficiency in future ICT systems. To improve transparency and reproducibility, the sustainability assessment employs representative operating parameters adopted from published studies on modern ICT infrastructure. The assumptions used throughout the analysis are summarized in Table 3 and include baseline server power, the fraction of total power associated with logic and memory subsystems, power usage effectiveness (PUE), grid carbon intensity, and representative workload characteristics. These parameters are selected to provide realistic order-of-magnitude estimates while remaining independent of any specific hardware platform or data-center architecture.

To account for uncertainty in practical deployment conditions, three representative operating scenarios are considered. The best-case scenario assumes a low-carbon electricity grid (300 g CO_2/kWh) and highly efficient cooling (PUE = 1.2). The nominal scenario assumes intermediate operating conditions (450 g CO_2/kWh and PUE = 1.3), while the worst-case scenario considers carbon-intensive

electricity (600 g CO_2/kWh) together with higher cooling overhead (PUE = 1.5). These scenarios provide an order-of-magnitude estimate of the environmental benefits rather than deterministic deployment predictions. The reduced operating voltage enabled by the steep-subthreshold switching behavior of the proposed HZO-based FeFET therefore provides a promising pathway toward lowering switching energy in future ICT hardware.

Table 3. Assumptions used for scenario-based sustainability assessment, ICT energy and emission estimation

Parameter	Value / Range	Rationale / Source
Baseline Server Power	5 kW	Standard high-density rack server.
Logic/Memory Energy Share	20–30% of the total load	Typical fraction of server power draw.
Data Center PUE	1.2 – 1.5	Reflects cooling overhead requirements.
Grid Carbon Intensity	300 – 600 g CO_2/kWh	Variable range based on regional energy mixes.
Operating time	8760 h/year	Continuous server opertaion
FeFET energy reduction	20–75%	Based on V_{DD} scaling + non-volatility
Workload Scenario	Data-intensive memory workloads	Maximizes the benefits of non-volatile retention.

4.2. *Standby power comparison*

Unlike conventional CMOS memories, which require continuous electrical power to preserve stored information, FeFETs retain data through the remanent polarization state of the ferroelectric layer. This inherent non-volatility enables memory blocks to be power-gated during idle periods without loss of stored information, thereby reducing standby energy consumption. In practical computing systems, however, total standby power cannot be eliminated because peripheral components including sense amplifiers, address decoders, clock distribution networks, control logic, and power-management circuits continue to consume a finite amount of power even when the memory array is inactive. Consequently, the present assessment assumes near-zero standby power at the memory-device level, rather than complete elimination of standby power at the system level.

The reduction in standby energy is expected to be most significant for applications characterized by long idle intervals or low-duty-cycle operation, such as IoT sensor nodes, edge-computing platforms, and intermittently active memory accelerators. Under such operating conditions, the non-volatile nature of FeFETs can reduce idle energy consumption while also lowering the associated cooling demand in larger ICT systems. Nevertheless, the magnitude of these benefits depends on application workload, system architecture, and power-management strategy. Therefore, the standby-energy savings reported in this study should be interpreted as scenario-based analytical estimates rather than universally applicable values.

4.3. *System-level power modeling and overhead considerations, and Sustainability-oriented performance comparison*

To bridge the gap between semiconductor physics and sustainability, it is essential to trace how individual transistor metrics influence global energy flows through a hierarchical pathway. The steep-slope switching enabled by the negative capacitance effect allows the FeFET to operate at a significantly reduced V_{DD} ~0.4 V, directly reducing active energy consumption in logic and memory chips because dynamic power scales quadratically with voltage V_{DD}. Simultaneously, a robust memory window ensures stable, non-volatile data retention without the need for constant refresh cycles, enabling "normally-off" computing where blocks are power-gated during idle periods to effectively eliminate standby power leakage. Furthermore, the high I_{ON}/I_{OFF} current ratio (10^6) may facilitate the design of peripheral sensing circuits by enabling lower sensing currents and shorter integration times, thereby indicating the potential for reduced memory-subsystem energy overhead. When extrapolated using the representative operating assumptions adopted in this study, these chip-level improvements may contribute to lower server- and rack-level power consumption and associated heat dissipation. Consequently, the reduced cooling demand, which can account for a significant fraction of data-center energy consumption, may further contribute to improved overall energy efficiency. The corresponding analytical framework linking device-level performance to representative system-level energy savings is illustrated in Fig. 5. Under the stated assumptions, these projected reductions in electricity demand indicate the potential for lower ICT-related CO_2 emissions across future data-center and edge-computing infrastructures. When integrated into broader green-computing frameworks, the low-voltage, steep-slope operation of HZO FeFETs may enable meaningful contributions toward decarbonizing digital infrastructure. Through this pathway, the microscopic optimization of ferroelectric switching becomes a core enabler for the macroscopic decarbonization of the global ICT infrastructure. While device-level V_{DD} scaling significantly reduces switching energy, a realistic system-level assessment must account for peripheral circuitry and architectural overheads inherent in memory arrays. To address the limitations of direct device-to-system extrapolation, the power model is refined to account for array-level contributions and control-logic overhead.

4.3.1. Array-level energy consumption

The total energy per operation is expressed as:

$$E_{total} = E_{bitcell} + E_{peripheral} + E_{control} \quad (4)$$

where $E_{bitcell}$ represents the intrinsic switching energy of the FeFET device, $E_{peripheral}$ accounts for energy consumed

by sense amplifiers, word-line drivers, and row/column decoders, and $E_{control}$ corresponds to the overhead associated with control logic, including power-gating and state management circuitry. This decomposition highlights that device-level improvements directly impact only a portion of the total energy budget, while system-level efficiency depends on the combined optimization of all components.

4.3.2. Impact of high $\frac{I_{ON}}{I_{OFF}}$ ratio on peripheral circuits

The high $\frac{I_{ON}}{I_{OFF}}$ ratio (>10_6) demonstrated by the proposed MFIS FeFET structure provides important advantages at the circuit level. Specifically:

1. **Reduced sensing current:** The large current contrast between ON and OFF states enables reliable readout at lower sensing currents, thereby reducing dynamic power consumption in sense amplifiers.
2. **Shorter integration times:** Faster distinguishability between logic states allows reduced sensing time, which lowers time-integrated leakage and improves overall read efficiency.

These effects contribute to lowering $E_{peripheral}$, partially compensating for system-level overheads.

4.3.3. Power-gating and control overhead

The intrinsic non-volatility of the ferroelectric layer enables aggressive power-gating strategies, supporting "normally-off" computing paradigms. Although the associated control circuitry introduces additional energy overhead $(E_{control})$, this cost is typically small compared to the continuous standby leakage power $(P_{standby})$ present in conventional volatile CMOS memories.

For workloads characterized by high idle-to-active ratios, such as data center and edge computing applications, the elimination of steady-state leakage results in a net reduction in total energy consumption, even after accounting for control overhead.

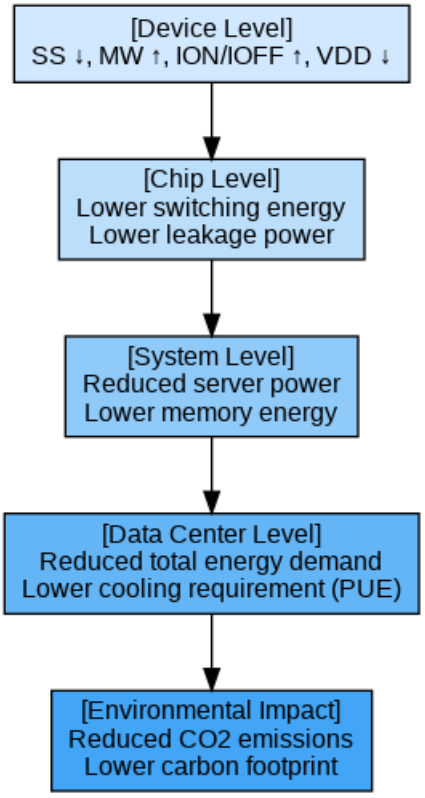


Figure 5. Multi-scale linkage between FeFET device metrics and sustainability outcomes, illustrating the pathway from device-level improvements to system-level energy savings and CO_2 emission reduction

In summary, FeFET device metrics influence sustainability outcomes through a hierarchical pathway: improved electrostatics → reduced operating voltage → lower chip energy → reduced system power → decreased cooling demand → lower CO_2 emissions. The key quantitative sustainability advantages of FeFET-based architectures over a representative CMOS baseline are summarized in Table 4. The values illustrate how reduced supply voltage, lower dynamic power, and elimination of standby loss combine with safer, lead-free materials to deliver both operational and embodied environmental benefits.

Table 4. Sustainability-oriented comparison of conventional CMOS and HZO-based FeFET architectures.

Parameter	CMOS baseline	FeFET architecture	Sustainability benefit
V_{DD}	0.8 V	0.4 V	50% lower supply voltage
Dynamic power	1×	0.25×	~75% lower dynamic power
Standby power	Continuous leakage	Near-zero (power-gated)	Idle energy loss is essentially eliminated
Toxic materials	Pb-containing PZT options	Pb-free HfO_2/HZO	Avoids lead; safer fabrication and disposal

By incorporating peripheral power contributions and PUE scaling, the presented analysis extends beyond simplified device-level projections to provide order-of-magnitude estimates of system-level energy and emission benefits. However, it is important to note that the results remain scenario-based and do not include detailed workload modeling, interconnect losses, or full system-level simulations. Therefore, the reported energy and CO_2 reductions should be interpreted as indicative trends rather than deployment-level predictions. Under the representative operating assumptions adopted in this study, the proposed HZO-based FeFET architecture may contribute to improved energy efficiency and lower ICT-related carbon emissions in future servers, edge computing platforms, and IoT devices.

4.4. *Limitations of the Sustainability Assessment*

The sustainability assessment presented in this study is based on a scenario-driven analytical framework derived from physics-based TCAD simulations of HZO-based FeFET devices. Although the simulated electrical characteristics provide valuable insight into the potential energy-

efficiency advantages of ferroelectric devices, several limitations should be recognized when interpreting the environmental implications.

First, the analysis is based on device-level TCAD simulations and does not include fabricated prototype devices or experimental measurements. Consequently, the reported electrical performance and associated sustainability benefits represent simulation-based predictions rather than experimentally validated hardware results.

Second, the environmental assessment does not constitute a complete LCA. The present study focuses primarily on the operational energy implications of reduced operating voltage and non-volatile data retention. Other environmental aspects, including raw material extraction, fabrication energy, manufacturing yield, transportation, device packaging, operational maintenance, recycling, and end-of-life disposal, are outside the scope of the present work.

Third, the system-level energy estimates are not derived from circuit-level or architectural simulations. The study does not include SPICE-based compact-model simulations, RTL or processor-level power analysis, or benchmark-driven workload evaluation. Instead, representative operating parameters reported in the literature are employed to establish an analytical connection between device-level improvements and potential ICT energy savings.

Furthermore, the sustainability assessment relies on representative scenario assumptions, including server power, logic and memory power contribution, PUE, grid carbon intensity, and continuous operating conditions. Although these assumptions are selected from representative values reported for modern ICT infrastructure, actual energy savings will depend on application workload, hardware architecture, cooling efficiency, technology node, and regional electricity generation mix. Finally, the TCAD simulations employ a representative interface trap density $\left(D_{\mathrm{it}} = 1\times10^{12}\ \mathrm{cm^{-2}eV^{-1}}\right)$ to reflect realistic HZO/SiO_2 interface conditions. However, the present study does not include a systematic sensitivity analysis of varying interface trap densities. Since interface traps can influence subthreshold swing, threshold voltage, and memory window through charge trapping effects, a comprehensive investigation of interface-trap variability and its impact on device performance and reliability will be considered in future work. Accordingly, the reported reductions in dynamic power, annual energy consumption, and CO_2 emissions should be interpreted as illustrative order-of-magnitude estimates that demonstrate the potential system-level implications of HZO-based FeFET technology rather than precise deployment-specific predictions.

Future work will extend the present device-level analysis by developing SPICE-compatible compact models derived from the calibrated TCAD parameters to enable circuit-level evaluation of FeFET memory arrays, including peripheral circuits such as word-line drivers, bit-line parasitics, and sense amplifiers. Furthermore, architecture-level workload benchmarking and standardized life-cycle assessment (LCA) methodologies in accordance with ISO 14040/14044 will be investigated to provide a more comprehensive assessment of the environmental impact of HZO-based FeFET technology under practical deployment conditions.

5. Environmental and material sustainability considerations

From a materials and process perspective, HfO_2/HZO-based ferroelectrics offer several environmental advantages over traditional perovskite ferroelectrics such as $Pb(Zr,Ti)O_3$ (PZT). HfO_2 and HZO are lead-free, avoiding the toxicity, worker-safety concerns, and end-of-life handling challenges associated with Pb-containing ferroelectrics. In contrast, PZT-based stacks require stringent waste management and can introduce additional regulatory and disposal burdens that increase the lifecycle environmental footprint of electronic products.

HfO_2-based ferroelectrics are also fully compatible with existing high-k/metal-gate CMOS process flows and can be deposited using standard ALD tools already present in advanced fabs. This compatibility allows reuse of mature CMOS infrastructure, reducing the embodied carbon and resource consumption that would otherwise be associated with dedicated process lines or exotic materials. By leveraging existing equipment, process chemistries, and integration know-how, HZO-based FeFET technologies can be scaled with lower marginal environmental cost compared to entirely new material systems. In contrast, other emerging memory technologies rely on materials that present additional environmental challenges. PCM typically uses chalcogenide compounds such as Ge–Sb–Te (GST), which involve relatively scarce and potentially hazardous elements. RRAM often employs transition metal oxides, where device variability and material stability can impact long-term reliability and recyclability. MRAM requires ferromagnetic layers and rare or critical materials, which may introduce supply chain and environmental concerns.

At the system level, the combination of low-voltage operation, reduced dynamic power, and near-zero standby consumption lessens the total heat generated by logic and memory subsystems. In data centers, this can translate into lower cooling demand, enabling higher PUE and reducing the indirect emissions associated with chiller and air-handling operation. In edge and IoT scenarios, improved energy efficiency reduces battery size and replacement frequency, thereby lowering material use and e-waste over the product lifetime. By avoiding toxic perovskites, reusing established CMOS infrastructure, and enabling substantial reductions in both operational and cooling energy, HfO_2-based FeFET architectures align strongly with the goals of sustainable semiconductor manufacturing and low-carbon ICT systems.

5.1. *Comparison with state-of-the-art non-volatile memories*

The comparison in **Table 5** highlights the key advantages of HZO-based FeFETs relative to other emerging non-volatile

memory technologies. FeFETs offer ultra-low switching energy, enabled by low-voltage operation and ferroelectric polarization switching, making them particularly suitable for energy-constrained ICT applications. In contrast, PCM typically requires higher programming currents due to Joule heating, resulting in higher energy consumption. RRAM provides competitive switching energy but suffers from variability due to stochastic filament formation. MRAM demonstrates excellent endurance but requires complex material stacks and higher switching currents. A key distinguishing feature of FeFETs is their combined advantage of low operating voltage, non-volatility, and CMOS process compatibility, allowing seamless integration into existing semiconductor manufacturing flows. This combination enables both dynamic power reduction (via V_{DD} scaling) and elimination of standby power, which are critical for reducing ICT energy consumption and associated carbon emissions. These comparisons are based on representative values reported in the literature and are intended to provide a relative performance perspective rather than exact device-to-device benchmarking. From a sustainability perspective, FeFETs further benefit from lead-free material composition and compatibility with existing high-k CMOS infrastructure, reducing both environmental impact and embodied fabrication energy compared to alternative memory technologies.

Table 5. Comparison of HZO-based FeFET with emerging non-volatile memory technologies

Parameter	FeFET (HZO)	PCM	RRAM	MRAM
Switching energy	~fJ–pJ	~pJ–nJ	~pJ	~pJ
Switching speed	ns	ns–µs	ns	Ns
Endurance	10^6–10^{12}	10^8–10^9	10^6–10^9	10^{12}–10^{15}
Retention	>10 years	>10 years	~10 years	>10 years
Operating voltage	Low (<1 V possible)	High (1–3 V)	Moderate	Moderate
CMOS compatibility	Excellent	Moderate	Good	Good
Material concerns	Pb-free (HfO_2)	Phase-change materials	Filament variability	Magnetic materials
Standby power	~ Near-Zero (non-volatile)	Low	Low	Low

6. Challenges and outlook

Although finite-element results indicate strong potential for HZO-based MFIS FeFETs, several technical and environmental obstacles must be addressed before large-scale commercial deployment. From a reliability perspective, physical device endurance is currently constrained by field- and cycling-induced oxygen-vacancy generation and migration, typically limiting HZO films to 10^6–10^{12} cycles. This makes it difficult to approach the 10^{15}-cycle stress levels tolerated by conventional CMOS logic. Furthermore, the depolarization field associated with the SiO_2 thin interfacial layer can destabilize polarization over time; thus, optimizing the t_f/t_{ox} ratio is essential for maintaining data integrity beyond 10 years. Environmental temperature also presents a significant challenge, as the ferroelectric $Pca2_1$ phase in HZO is metastable. Temperatures approaching 85°C can accelerate transitions back to the non-polar monoclinic phase, leading to memory window collapse and exacerbated leakage. Future modeling integrating thermal-transport equations and experimentally calibrated degradation models will be necessary to self-consistently evaluate these cooling-to-performance trade-offs.

From an environmental-science perspective, several uncertainties remain regarding fabrication yield and the maturity of high-volume manufacturing. Lower yields can lead to increased material waste and higher embodied carbon per functional device, while realistic deployment timelines must account for the multi-year transition from pilot foundries to global data-center fleets. Furthermore, the actual CO_2 reduction is highly sensitive to grid carbon intensity variability; in regions with high fossil-fuel dependence, the impact of device-level efficiency is maximized, whereas in "green-grid" regions, the benefit shifts toward reducing absolute resource demand. Improving endurance and retention is not only critical for reliability but also reduces the frequency of device replacement, thereby lowering e-waste generation and the embodied energy of manufacturing. To overcome these limitations, future work will utilize Physics-Informed Machine Learning (PIML) to identify design points that maximize reliability while preserving energy-efficient operation, cutting the energy usage of the R&D process itself.

Under the representative assumptions adopted in this study, FeFET technology serves as a foundational component that can be integrated with other sustainable-ICT strategies to amplify environmental gains. Its "normally-off" computing capability is uniquely suited for renewable-powered data centers using intermittent sources, as it allows for zero-leakage state preservation during supply fluctuations. Additionally, by reducing chip-level heat dissipation, FeFETs enable more efficient thermal management, potentially supporting waste-heat recovery in dense server environments. Finally, the lead-free (Pb-free) composition of HZO and the reuse of mature CMOS infrastructure align with circular electronics principles by reducing hazardous waste and the environmental cost of establishing new fabrication facilities. By positioning FeFETs within this broader ecosystem, this work underscores the role of semiconductor innovation as an essential lever for achieving a low-carbon digital transition.

7. Conclusion

This study employed two-dimensional TCAD simulations to investigate HZO-based MFIS FeFETs as promising

candidates for energy-efficient, non-volatile devices for next-generation information and communication technologies. The simulation results indicate that appropriately calibrated HfO_2-based gate stacks can achieve stable ferroelectric switching, wide memory windows, and high drive currents suitable for low-power memory and logic-in-memory applications. The combination of steep-subthreshold switching, low-voltage operation, and non-volatile data retention suggests the potential for reduced device-level energy consumption and improved energy efficiency. Under the representative operating assumptions adopted in this study, these device-level improvements may contribute to lower operational energy consumption and associated CO_2 emissions in future ICT systems. However, these sustainability implications should be interpreted as scenario-based analytical estimates rather than deployment-specific predictions. Future work will focus on experimental validation, SPICE-compatible compact-model development, circuit- and architecture-level benchmarking, and comprehensive life-cycle assessment to further quantify the environmental benefits of HZO-based FeFET technology. Overall, the results suggest that HZO-based MFIS FeFETs have the potential to contribute to the development of more energy-efficient and environmentally sustainable computing systems.

Competing interests

The authors declare that there are no known competing financial or non-financial interests that could have influenced the work reported in this paper.

Funding

The author received no financial support for the research, authorship, and/or publication of this paper.

Author contributions

Harsha Ragini Aturi: Conceptualization, Investigation, Data Curation, Writing – Original Draft, Visualization. **Ramana Murthy Gajula:** Supervision, Methodology, Validation, Resources, Formal Analysis, Writing – Review & Editing.

Acknowledgement

The author would like to thank Alliance University, Bengaluru, for providing the academic environment that supported the preparation of this study.

References

Ahmed, K.M.U., Bollen, M.H.J. and Alvarez, M. (2021), A review of data centers energy consumption and reliability modeling, IEEE Access, 9, pp. 152536–152563. https://doi.org/10.1109/ACCESS.2021.3125092

Asai, H., Fukuda, K., Hattori, J., Koike, H., Miyata, N., Takahashi, M. and Sakai, S. (2017), Compact model of ferroelectric-gate field-effect transistor for circuit simulation based on multidomain Landau–Kalathnikov theory, *Japanese Journal of Applied Physics*, 56(4S), 04CE07. https://doi.org/10.7567/JJAP.56.04CE07

Böscke, T.S., Müller, J., Bräuhaus, D., Schröder, U. and Böttger, U. (2011), Ferroelectricity in hafnium oxide thin films, *Applied Physics Letters*, 99(10), 102903. https://doi.org/10.1063/1.3634052

Hoffmann, M., Schroeder, U., Schenk, T., Shimizu, T., Funakubo, H., Sakata, O., Pohl, D., Rellinghaus, B., Mikolajick, T. and Müller, J. (2015), Stabilizing the ferroelectric phase in doped hafnium oxide, *Journal of Applied Physics*, 118(7), 072006. https://doi.org/10.1063/1.4927805

Jiang, C., Fan, T., Gao, H., Shi, W., Liu, L., Cérin, C. and Wan, J. (2020), Energy aware edge computing: A survey, *Computer Communications*, 151, pp. 556–580. https://doi.org/10.1016/j.comcom.2020.01.004

Lines, M.E. and Glass, A.M. (1977). Principles and applications of ferroelectrics and related materials. Oxford: Clarendon Press. https://doi.org/10.1093/acprof:oso/9780198507789.001.0001

Malmodin, J., Lövehagen, N., Bergmark, P. and Lundén, D. (2024), ICT sector electricity consumption and greenhouse gas emissions – 2020 outcome, *Telecommunications Policy*, 48(3), 102701. https://doi.org/10.1016/j.telpol.2023.102701

Martin, L.W., Chu, Y.-H. and Ramesh, R. (2010), Advances in the growth and characterization of magnetic, ferroelectric, and multiferroic oxide thin films, *Materials Science and Engineering: R: Reports*, 68(4–6), pp. 89–133. https://doi.org/10.1016/j.mser.2010.03.001

Masanet, E., Shehabi, A., Lei, N., Smith, S. and Koomey, J. (2020), Recalibrating global data center energy use estimates, *Science*, 367(6481), pp. 984–986. https://doi.org/10.1126/science.aba3758

Materlik, R., Künneth, C. and Kersch, A. (2015), The origin of ferroelectricity in $Hf_{1-x}Zr_xO_2$: A computational investigation and a surface energy model, *Journal of Applied Physics*, 117(13), 134109. https://doi.org/10.1063/1.4916707

Mikolajick, T., Slesazeck, S., Park, M.H., Schroeder, U., *et al.* (2018), Ferroelectric hafnium oxide for ferroelectric random-access memories and ferroelectric field-effect transistors, *MRS Bulletin*, 43(5), pp. 340–346. https://doi.org/10.1557/mrs.2018.92

Müller, J., Böscke, T.S., Schröder, U., Mueller, S., Bräuhaus, D., Böttger, U., Frey, L. and Mikolajick, T. (2012), Ferroelectricity in simple binary ZrO_2 and HfO_2, *Nano Letters*, 12(8), pp. 4318–4323. https://doi.org/10.1021/nl302049k

Pešić, M., Fengler, F.P.G., Larcher, L., Padovani, A., Schenk, T., Grimley, E.D., Sang, X., LeBeau, J.M., Slesazeck, S., Schroeder, U. and Mikolajick, T. (2016), Physical mechanisms behind the field-cycling behavior of HfO_2-based ferroelectric capacitors, *Advanced Functional Materials*, 26(25), pp. 4601–4612. https://doi.org/10.1002/adfm.201600590

Salahuddin, S. and Datta, S. (2008), Use of negative capacitance to provide voltage amplification for low-power nanoscale devices, *Nano Letters*, 8(2), pp. 405–410. https://doi.org/10.1021/nl071804g

Sang, X., Grimley, E.D., Schenk, T., Schroeder, U. and LeBeau, J.M. (2015), On the structural origins of ferroelectricity in HfO_2 thin films, *Applied Physics Letters*, 106(16), 162905. https://doi.org/10.1063/1.4919135