https://doi.org/10.1038/s44335-026-00106-z

Material optimization of interconnects and transistors for energy-efficient and high-performance cryogenic GAA technology

Check for updates

Albi Mema[1] ✉, Anirban Kar[1], Andrea Padovani[2], Luca Vandelli[3], Munehiro Tada[4], Takahisa Tanaka[4], Luca Larcher[3], Gaurav Thareja[3] & Hussam Amrouch[1] ✉

Unprecedented computational demand is reshaping the landscape of energy-efficient and high-performance computing. As Moore's law approaches its physical limits, continued progress to meet this demand can no longer rely solely on transistor scaling. Cryogenic operation of CMOS circuits offers a promising path forward, enhancing the subthreshold slope (SS) of transistors while enabling certain materials to exhibit superconductivity. This dual advantage positions device- and material-level optimizations as a compelling solution for next-generation cryogenic computing. In this work, we demonstrate that unlocking the full potential of cryogenic CMOS requires strategic transistor and interconnect-level optimizations. We show that (1) material optimization for tuning transistor work function is vital for translating SS improvements into a significant increase in ON current, while (2) exploiting superconducting interconnect materials at cryogenic temperatures is essential to eliminate resistive parasitics and overcome interconnect bottlenecks. To quantify these benefits, we employ a cryogenic-aware 2 nm gate-all-around nanosheet digital design flow, incorporating calibrated device models along with a full standard cell library. Our evaluation results targeting the compute fabric of the AI accelerator circuit demonstrate that our optimizations provide a 3× increase in TOPS, 7.7× increase in TOPS/W, and 23× improvement in efficiency TOPS^2/W.

Emerging computational paradigms are pushing computing requirements beyond the limits of Moore's law. In response, cryogenic complementary metal-oxide-semiconductor (CMOS) circuits are emerging as promising enablers for next-generation computing, with applications spanning high-performance systems, artificial intelligence (AI), quantum control, and space engineering[1]. Cryogenic CMOS originally emerged from quantum computing, where the need for circuits operating below 10 K to reliably control qubits is unavoidable, especially as the number of qubits scales up[2]. Meanwhile, the rapid evolution of AI fuels unprecedented demand for more efficient and powerful high-performance computing systems. These systems must deliver increasing levels of computational throughput, pushing conventional CMOS technology toward its physical and thermal limits[3]. This growing disparity between computational demand and attainable performance calls for new paradigms capable of overcoming the barriers of conventional CMOS.

The benefits of cryogenic CMOS cannot be realized without careful device- and material-level optimizations. Transistor characteristics, interconnect behavior, variability, and design methodologies all shift significantly at cryogenic temperatures, necessitating dedicated research to bridge the gap between device physics and large-scale system implementations. Achieving meaningful performance gains at system-level requires coordinated optimization across multiple abstraction layers. In particular, device- and interconnect-level enhancements must be optimized to ensure these improvements translate into system-level benefits. Beyond the physical layer, standard cell libraries and memory macros must be characterized and interconnect models updated for operation at cryogenic temperatures to ensure consistent timing closure and power consumption estimation. Such cross-domain optimization is essential to fully exploit the potential of cryogenic CMOS for scalable, energy-efficient high-performance computing architectures.

Among the most promising device candidates for sustaining Moore's law and meeting the demands of cryogenic and high-performance workloads is the gate-all-around (GAA) nanosheet transistor. GAA transistors are expected to replace FinFET technology as the next leading transistor

[1]TUM School of Computation, Information and Technology; Chair of AI Processor Design; Munich Institute of Robotics and Machine Intelligence, Technical University of Munich, Munich, Germany. [2]Department of Engineering Sciences and Methods, University of Modena and Reggio Emilia, Modena, Italy. [3]Applied Materials, Santa Clara, CA, USA. [4]Keio University, Tokyo, Japan. ✉e-mail: a.mema@tum.de; amrouch@tum.de

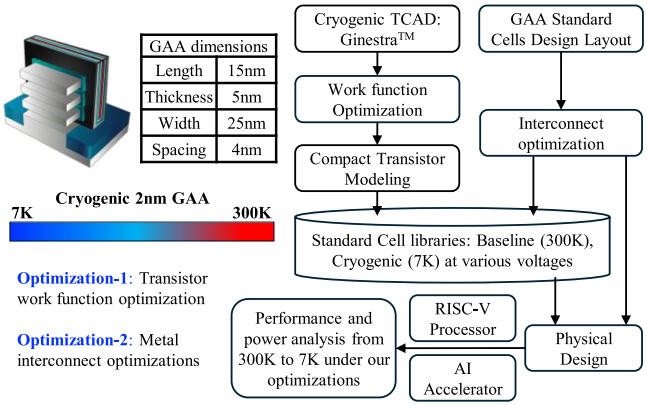


Fig. 1 | Material-to-system evaluation framework for cryogenic 2 nm gate-all-around (GAA) technology. The framework illustrates a complete register-transfer-level to GDSII (RTL-to-GDS) design flow that integrates device, circuit, and system-level modeling under cryogenic conditions. At the device level, Ginestra™ technology computer-aided design (TCAD)[9] simulations are calibrated against measurement data to capture quantum effects and leakage mechanisms. These results are used to generate compact transistor models incorporating temperature-dependent behavior and work-function optimization for iso-I_{OFF} operation. In parallel, standard cells are designed, laid out, and characterized across voltage and temperature to generate timing and power libraries. Interconnect optimization is applied by modifying material properties to reflect cryogenic behavior, including superconductivity effects. The resulting models and libraries are then used in chip implementations, including a RISC-V processor and AI accelerator, enabling extraction of performance, power, and energy efficiency metrics. This framework enables systematic propagation of material- and device-level optimizations to system-level evaluation.

architecture[4]. By surrounding the channel with the gate on all sides, GAA devices offer superior electrostatic control with increased current drive and reduced leakage[5]. This extended gate control enables better channel modulation, allowing lower threshold voltage and an improved subthreshold slope. The stackable nanosheet geometry offers an additional degree of freedom in modulating the channel width without affecting area.

To quantitatively evaluate the benefits of the optimizations at device- and material-level, an implementation methodology is employed. The flow of this methodology begins with register-transfer-level (RTL) descriptions of the target system and proceeds through logic synthesis, physical design, and sign-off to obtain realistic post-implementation metrics. This flow, commonly referred to as RTL-to-GDS, enables accurate extraction of timing, power, and area metrics while capturing the influence of cryogenic-aware device and interconnect modifications. The analysis includes both standard cell logic and memory macros, reflecting the composition of modern high-performance computing architectures.

In this work, we investigate the optimization of GAA nanosheet transistors alongside superconducting interconnects operating at cryogenic temperatures. We propose targeted device and material-level modifications, providing a comprehensive evaluation of their implications on circuit and system performance. These optimizations are not treated in isolation, but are concurrently evaluated and propagated across the full design stack, enabling their combined system-level impact to be quantified. While prior works have explored these techniques individually, their combined impact across device, circuit, and full-system level has not been systematically evaluated within a consistent RTL-to-GDS evaluation flow. The analysis spans from device-level to system-level implementation, ensuring that the effects of device and interconnect enhancements are accurately reflected in post-layout performance metrics. This approach enables the quantification of performance, power, and area benefits under cryogenic operation. The framework is shown in Fig. 1. Although comprehensive system-level evaluations at cryogenic temperatures exist in the literature[6–8], they show results using standard copper interconnects and FinFET architectures. Our evaluation extends prior system-level studies by targeting not only a general-purpose processor, but also a massively parallel 64 × 64 MAC AI accelerator.

Results

Material-to-system evaluation framework

Bridging material and device innovations to system-level performance requires a rigorous evaluation methodology that spans the entire technology stack. This methodology follows standard industry RTL-to-GDS design practices, but is extended to incorporate cryogenic-aware device and interconnect modeling. Material-to-system evaluation integrates transistor-level models, interconnect properties, and standard cell designs at the lowest level. On top of this, standard cell-level metrics are extracted and integrated into the full-system benchmarking flows, together with interconnect characteristics. Such a framework allows the impact of device physics, material properties, and technology optimizations to be translated into architectural metrics such as throughput, latency, and energy efficiency. Such cross-layer evaluation is particularly critical at cryogenic temperatures, where both transistor behavior and interconnect properties deviate significantly from room temperature trends. By embedding cryogenic device and interconnect modifications into RTL-to-GDS design flows, it becomes possible to quantify how low-temperature benefits manifest at system-level. This enables a realistic assessment of design trade-offs, ensuring that improvements in device physics can be systematically linked to system-level performance gains. In this work, the methodology is used to evaluate the combined impact of device- and interconnect-level optimizations across the full design stack.

Cryogenic-aware GAA transistors

Transistor behavior at cryogenic temperatures differs substantially from room temperature operation, as fundamental material properties are strongly temperature-dependent. Before performing any transistor-level optimization, it is essential to establish a simulation setup capable of capturing variations in material properties between room and cryogenic temperatures. For this purpose, p- and n-type 2 nm nanosheet field-effect transistors (n-NSFET and p-NSFET) are simulated using the Ginestra™[9] technology computer-aided design (TCAD) tool.

To accurately model transistor behavior, quantum effects such as trap-assisted tunneling (TAT) and quantum-corrected transport are included alongside classical equations. TAT, caused by defect-assisted tunneling through the bandgap, manifests as variations in gate current for both n- and p-NSFETs. As shown in Fig. 2, Ginestra™ self-consistently solves Schrödinger and semiclassical equations, ensuring quantum-accurate predictions of leakage and current density. Additionally, electron density near the drain is corrected using the quantum potential $\nabla G_{n/p}$ (Fig. 3). After incorporating these quantum effects, reliability mechanisms–bias temperature instability (BTI), hot-carrier injection (HCI), and self-heating (SHE)–are also modeled, as they remain key degradation sources for GAA transistors even under cryogenic operation[10].

Calibration against experimental measurements[11] is essential to ensure that simulated device behavior matches physical transistors across temperatures ranging from room to cryogenic. Model parameters in Ginestra™ are tuned to closely reproduce measured data. Metrics such as gate current, total gate capacitance (C_{GG}), and conductance are used as calibration targets across different gate-source voltages (V_{GS}). As shown in Fig. 4, simulated results closely match measurements. Special attention is paid to gate current across a wide temperature range, which allows for a more accurate characterization of gate dielectric traps, including determining their atomic nature[11]. As shown in Fig. 5a–b, Ginestra™ simulations reproduce experimental data with high fidelity[11].

To clarify the origin of the calibration data, we note that the 2 nm GAA structure was modeled by combining structural information and electrical data from multiple industrial sources. Specifically, the geometrical features of the 3D GAA architecture are reconstructed based on TechInsights data from a state-of-the-art commercial GAA technology node, though specific details remain undisclosed due to confidentiality constraints. Regarding

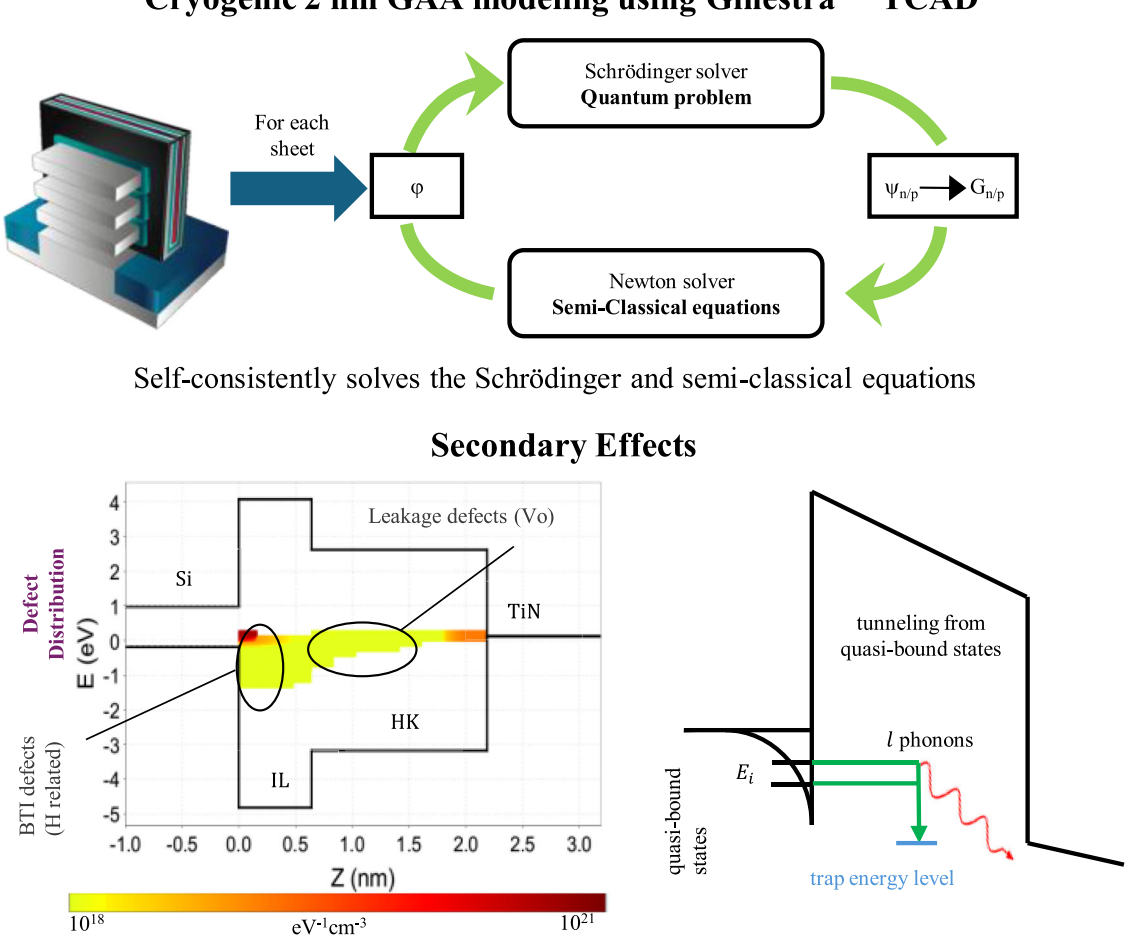


Fig. 2 | Modeling of dielectric defects and trap-assisted tunneling in cryogenic GAA transistors using Ginestra™ TCAD. Atomic-scale defects within the gate dielectric, including leakage-related oxygen vacancies and BTI-induced traps, are modeled through their energy distributions and interactions with carriers. Multi-phonon trap-assisted tunneling (TAT) is captured via transitions between quasi-bound channel states and defect levels within the dielectric. The framework self-consistently solves the Schrödinger equation to obtain quantized carrier states and the semiclassical transport equations to model carrier flow. This coupled approach enables accurate prediction of gate leakage currents across temperatures, forming the basis for reliable cryogenic device modeling and subsequent compact model extraction.

electrical calibration, gate leakage, capacitance, and conductance characteristics used to extract dielectric trap properties are obtained from planar SiO_2/HfO_2 devices. In contrast, transistor transfer characteristics (I_d-V_g) are calibrated using measurements from state-of-the-art GAA technologies at room temperature. Both these sources of data are combined to create a consistent cryogenic Ginestra™ TCAD model.

Cryogenic-aware transistor compact modeling

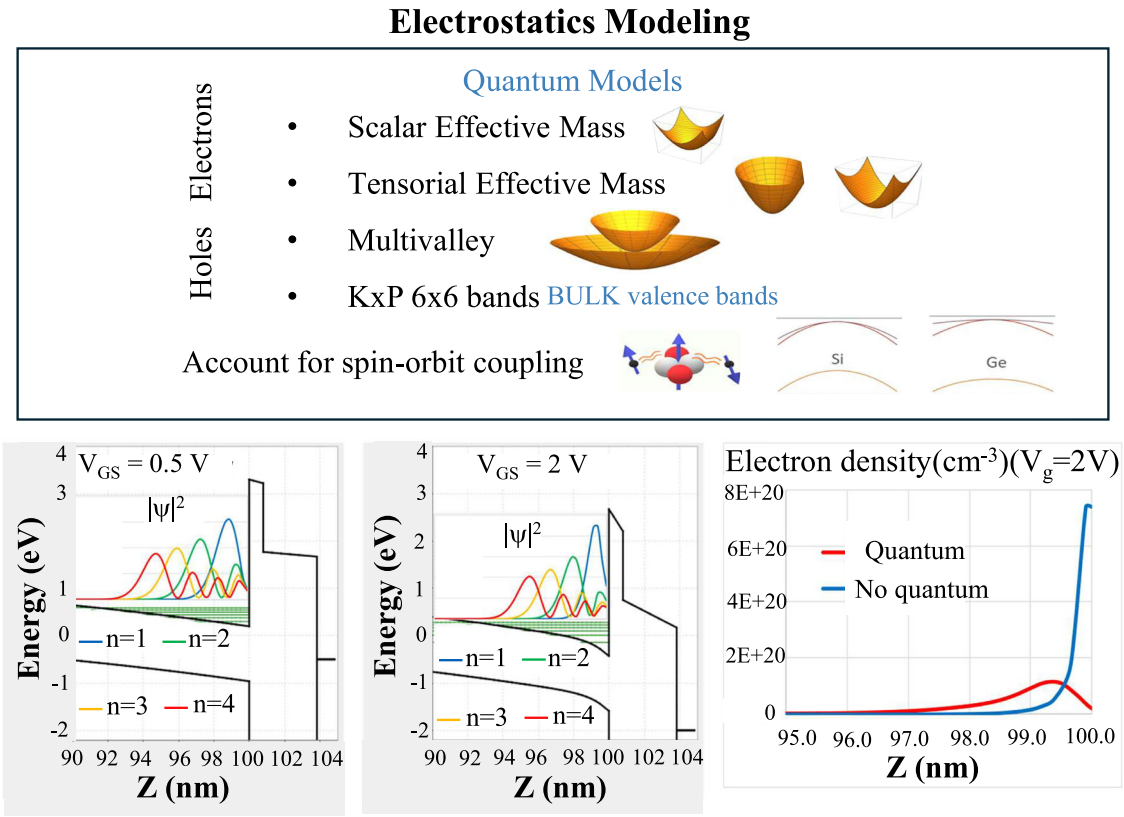
The industry-standard compact model for GAA transistors, BSIM-CMG[12,13], is a physics-based model that represents and captures the transistor's electrical behavior, enabling fast, accurate circuit-level simulations. Our calibration begins at 300 K to extract baseline parameters related to geometry, threshold voltage, mobility, and resistance. At cryogenic temperatures, several physical effects–such as nonlinear subthreshold slope due to band-tail states, threshold voltage shift, and altered carrier mobility–become prominent. These are captured by introducing temperature-dependent corrections to the subthreshold swing, threshold voltage, and mobility models. Additional parameters are incorporated to model enhanced Coulomb and surface roughness scattering, nonlinear velocity saturation, and modified drain-induced barrier lowering. Temperature coefficients for the effective drain-source voltage and carrier velocity were also refined to ensure smooth transition across operating regimes[14–16]. The resulting model shows close agreement with TCAD simulations calibrated to experimental data, accurately reproducing device behavior down to 7 K.

To ensure consistency, I_d-V_g characteristics across a range of drain, gate, and source biases are extracted from Ginestra™ TCAD simulations. The Ginestra™ TCAD model has been calibrated against experimentally measured gate-stack characteristics, establishing a consistent link between measurement, device-level modeling, and compact model extraction. As shown in Fig. 5c–d, model parameters are systematically tuned to match these reference characteristics, ensuring that the extended compact model closely tracks TCAD behavior. The results of the calibration are stored in the model card, which contains the set of parameters necessary for the compact BSIM-CMG model to operate correctly.

The cryogenic-aware BSIM-CMG model enables us to seamlessly integrate cryogenic-aware device characteristics into larger-scale circuit analysis. As a result, circuit and later system designers can accurately estimate the performance, power, area (PPA), and reliability trade-offs of nanosheet transistors in cryogenic environments.

Cryogenic-aware material optimizations

Two key optimizations are investigated in this work to unlock the performance and energy efficiency gains offered by cryogenic operation. The first targets transistor behavior at cryogenic temperatures, while the second



Quantum-corrected semi-classical transport model

$$\boldsymbol{J}_n = q\mu_n(\mathbf{F},n,\mathbf{r})(\mathbf{F}+\nabla G_n)n + qD_n(\mathbf{F},n,\mathbf{r})\nabla n - \frac{3}{2}qD_n(\mathbf{F},n,\mathbf{r})\nabla\ln(m_n^*)$$

$$\boldsymbol{J}_p = q\mu_p(\mathbf{F},p,\mathbf{r})(\mathbf{F}+\nabla G_p)p + qD_p(\mathbf{F},p,\mathbf{r})\nabla p - \frac{3}{2}qD_p(\mathbf{F},p,\mathbf{r})\nabla\ln(m_p^*)$$

Fig. 3 | Quantum-corrected transport modeling in cryogenic GAA nanosheet transistors. The corrective quantum potential $\nabla G_{n/p}$, derived from the Schrödinger equation, captures carrier confinement and quantization effects in the channel and is incorporated into the semiclassical drift-diffusion transport framework. The model combines quantum-mechanical solutions for carrier wavefunctions and subband energies with classical transport equations to enable self-consistent simulation. The plots illustrate electron density and energy profiles with and without quantum correction, highlighting the impact of quantization on carrier distribution. This approach enables more accurate estimation of drain current and transport behavior in deeply scaled nanosheet devices, particularly under cryogenic conditions.

leverages the unique advantages of superconducting materials that emerge only at low cryogenic temperatures.

Optimization-1

To fully exploit the steeper subthreshold slope (SS) induced by cryogenic operation, precise tuning of the work functions of n-NSFETs and p-NSFETs is essential. Without this adjustment, the threshold voltage (V_T) increase observed at low temperatures leads to a sharp reduction in I_{OFF} and consequently in I_{ON}, negating potential performance benefits. To preserve the performance advantage, V_T is shifted negatively through work-function optimization so that I_{OFF} remains constant between 300 K and 7 K—a condition known as iso-I_{OFF}, where the transistor V_T at cryogenic conditions is tuned until the off-state leakage current matches that of room temperature. This ensures that the improved SS directly translates into higher I_{ON}.

Our analysis shows that a 260 mV negative shift in V_T is required to achieve iso-I_{OFF} condition (Fig. 5d). However, existing 2 nm GAA nodes enable only a 200 mV shift through conventional gate-stack engineering[17]. To bridge this gap, further V_T adjustment is achieved using novel metal-gate materials: TiAlC combined with La-based dipoles for n-NSFETs, and TiN/W combined with Al-based dipoles for p-NSFETs[18–20]. Unlike prior studies[6,7] that assume ideal threshold voltage shifts, this work uses physically realizable work-function engineering solutions. While prior works such as[8] explore dipole-based tuning, we employ different material choices and propagate their impact through compact modeling to full-system evaluation. Regarding on-die variation, the work-function tuning is treated as a systematic technology-level V_T adjustment, while random threshold-voltage mismatch is assumed comparable to baseline GAA technology and is not explicitly modeled in this study. The resulting model card is then used in circuit and system simulations as outlined in Fig. 7.

Optimization-2

Cryogenic operation also reduces the resistance of metal interconnects, which play a dominant role in system-level delay and power. All interconnect layers are therefore optimized by replacing conventional copper (Cu) with niobium (Nb) (Fig. 6a). As shown in Fig. 6b, niobium transitions into a superconducting state below 7.8 K[21], exhibiting effectively zero electrical resistance. In parasitic extraction, all Nb layers below the critical temperature are assigned zero sheet resistance, consistent with operation within the superconducting regime. This eliminates resistive parasitics that typically limit signal propagation and contribute to dynamic power dissipation, enabling notable performance gains–particularly in large-scale

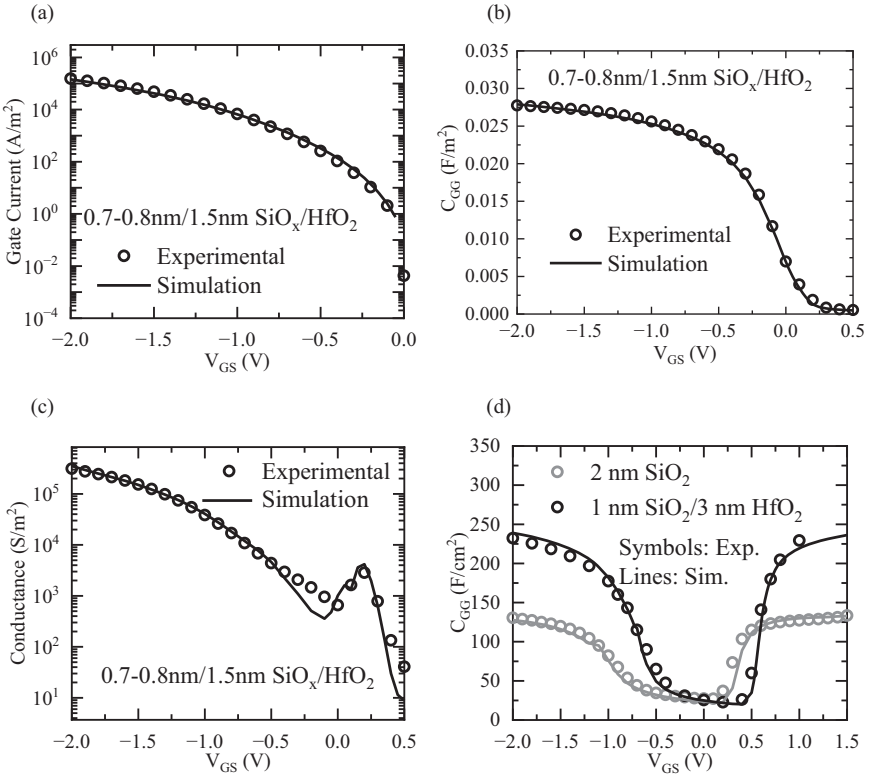


Fig. 4 | Experimental validation of Ginestra™ TCAD models for nanosheet GAA devices across key electrical characteristics. Symbols denote experimental measurements, while solid lines represent TCAD simulation results. **a** Gate current as a function of gate-source voltage (V_{GS}) for SiO_2 /HfO_2 gate stacks, demonstrating accurate modeling of leakage behavior. **b** Total gate capacitance (C_{GG}) as a function of V_{GS}, where C_{GG} denotes the total gate capacitance per unit area. **c** Conductance as a function of V_{GS}, validating charge transport behavior across bias conditions. **d** Comparison of C_{GG} for different oxide thicknesses and material stacks, highlighting the ability of the model to capture variations in dielectric properties. Across all panels, the close agreement between simulations and measurements confirms that the calibrated TCAD model accurately captures both leakage and capacitance characteristics, providing a reliable basis for compact model extraction and subsequent circuit-level analysis.

systems with long interconnect paths and high fan-out nets. While prior research in refs. [21,22] has already successfully demonstrated the fabrication and basic inverter-chain using Nb interconnects, our work builds upon their observations to perform a full system-level evaluation.

Niobium (Nb) interconnects are compatible with advanced CMOS manufacturing, as they can be fabricated using standard 300 mm processes including sputtering, ArF immersion lithography, and dry etching. Unlike copper, Nb exhibits strong resistance to diffusion into dielectric layers, eliminating the need for barrier or adhesion layers (e.g., Ta/TaN) and simplifying integration while reducing line resistance at scaled nodes. Although current demonstrations are based on larger feature sizes, a phased integration strategy is envisioned: near-term deployment in semi-global and global interconnect layers to maximize superconductivity benefits, followed by long-term extension to local routing layers as scalability to sub-50 nm dimensions is established. In this work, all interconnect layers are modeled under the assumption of Nb-based metallization: M0-M2 for standard-cell internal routing, M1-M6 for digital place-and-route in the RISC-V and accelerator, and M0-M6 for the SRAM macro.

It is important to note that superconducting Nb-based interconnects are subject to critical current (I_c) limitations, beyond which superconductivity breaks down. Measured data from 100 nm wide, 50 nm thick Nb interconnects[21,22] demonstrates an I_c of approximately 3.2 mA. The typical wire width for this node is around 24 nm. At this width, the critical current of Nb interconnects is in the sub-mA range. This is close to the current that a transistor can supply and therefore close attention must be paid to ensure that Nb interconnects do not conduct current above I_c, at which point they would lose the superconducting properties. In this work, all evaluations assume operation within the superconducting regime, where currents remain below the critical threshold. This work does not impose an explicit per-net current constraint; rather, sub-critical operation can be achieved by constraining the physical implementation of the design, including prohibiting high-drive strength cell usage during synthesis and subsequent optimization steps of the digital flow and ensuring appropriate power distribution network (PDN) design.

In the superconducting regime, long interconnects may exhibit ringing and waveform overshoot, particularly for fast-switching nets such as clock or global routing signals. Conventional EDA timing analysis, which relies on RC-based models, does not fully capture these inductive effects; therefore, critical nets require selective SPICE-level validation. To mitigate potential signal integrity issues, standard physical design techniques such as layer hopping (introducing controlled resistance through vias) and the insertion of high-resistance discharge paths (e.g., reverse-biased diodes) can be employed. While these measures introduce non-idealities compared to the zero-resistance assumption, they provide a practical approach to maintain robust operation within existing design flows, leveraging techniques widely used in VLSI physical design for antenna violation mitigation; however, their detailed implementation is beyond the scope of this work.

To rigorously assess the impact of niobium on both cell- and system-level parasitics, interconnect resistance data in the 2 nm nanosheet PDK is updated. Parasitic extraction is then performed during the interconnect optimization stage (Fig. 7), followed by recharacterization of standard cells to capture niobium's impact on delay and energy consumption. The updated interconnect data are also used in the physical design and static timing analysis (STA) tools to ensure accurate evaluation of this optimization's benefits.

Cryogenic-aware standard cell libraries

The next stage of our device-to-system evaluation pipeline focuses on developing standard cells, the fundamental building blocks of digital systems. Each standard cell consists of n- and p-NSFETs arranged in a complementary configuration, with interconnects linking the source, drain, and gate terminals to perform binary logic operations. These cells form the foundation of larger and more complex circuits such as processors, memories, and AI accelerators. Evaluating their behavior at cryogenic temperatures is essential to connect device- and material-level optimizations with system-level performance and efficiency gains.

To this end, we design, layout, and characterize standard cells using a 2 nm nanosheet PDK. The library includes a wide range of 71 combinational

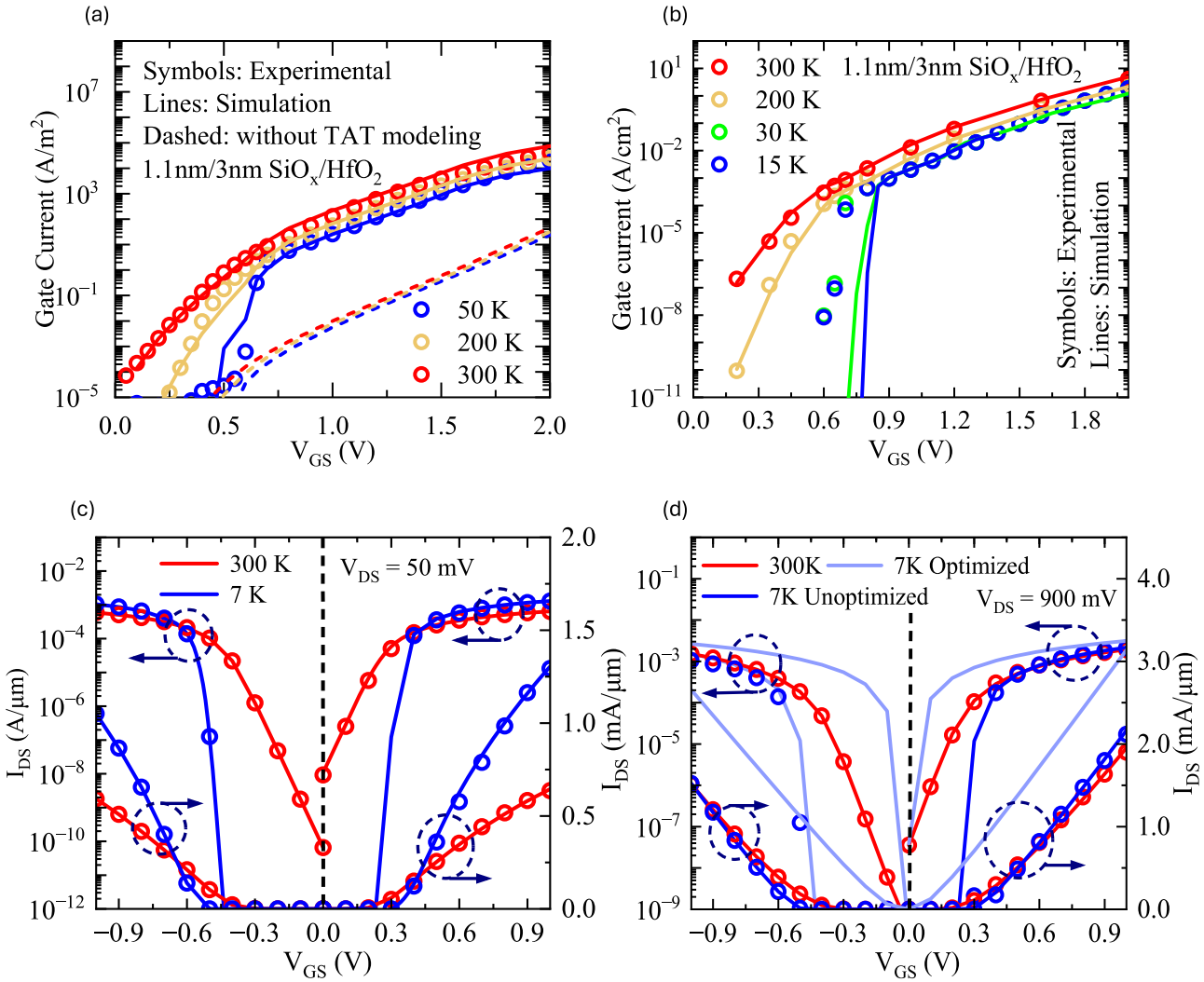


Fig. 5 | Validation of trap-assisted tunneling and compact model calibration for cryogenic GAA nanosheet transistors. a Gate current as a function of gate-source voltage (V_{GS}), comparing experimental measurements (symbols) with Ginestra™ TCAD simulations (solid lines)[11]. Dashed lines indicate simulations without trap-assisted tunneling (TAT), demonstrating that neglecting TAT significantly underestimates gate leakage current. **b** Measured and simulated gate current at different temperatures down to cryogenic conditions with TAT modeling, confirming that the TAT model accurately captures temperature-dependent leakage behavior[11]. **c** Transfer characteristics (I_{DS}-V_{GS}) at low drain-source voltage (V_{DS} = 50 mV), where symbols represent TCAD results and solid lines denote the calibrated BSIM-CMG compact model, showing close agreement. **d** Transfer characteristics at high drain-source voltage (V_{DS} = 900 mV), illustrating iso-I_{OFF} operation, where threshold voltage is shifted to maintain constant off-state current between 300 K and 7 K. This results in increased drive current (I_{ON}) at cryogenic temperature. Together, these results validate both the physical leakage modeling and the accuracy of the compact model used for circuit-level simulations.

and sequential cells, from simple gates such as inverters, NANDs, and NORs (Fig. 8b) to more complex elements like latches and flip-flops. Following physical design, parasitic extraction is performed on each cell using an industry-standard sign-off tool to generate accurate post-layout netlists. These netlists capture parasitic resistance and capacitance values at both the transistor level and interconnect level under cryogenic and room-temperature conditions.

During characterization, each standard cell is systematically evaluated across all input signal slews and output load capacitances. This exhaustive process ensures that all metrics relevant to digital operation, such as area, propagation delay, output transition time, dynamic power, and leakage power, are captured. The results are compiled into the "Liberty" format, enabling high-level RTL descriptions to be synthesized into physically implementable logic (Fig. 7). Characterization spans a wide range of temperatures and supply voltages, from 300 K to 7 K and from 0.9 V down to 0.3 V. The calibrated model card of the optimized transistors generates standard cell behavior that accurately reflects the calibrated device and internal interconnect behavior. Fully compatible with EDA tools, our cryogenic-aware libraries enable accurate timing and power sign-off for complex circuits such as processors and AI accelerators.

Memory-level analysis

In high-performance systems, on-chip memories are a critical component of system-level design. Implementing memories using sequential logic results in large area and power overhead. Instead, static random access memories (SRAMs) are more area- and energy-efficient. Among the different types of SRAM cells, the 6-transistor (6T) SRAM remains the most widely adopted. It is favored due to the small area footprint, high speed and maturity over multiple design nodes.

A 6T SRAM cell consists of two cross-coupled inverters forming a positive feedback loop for data storage, and two access transistors whose gates are controlled by a common signal named wordline. The access transistors enable the internal storage nodes to connect to bitlines during read and write operations. Wordlines select entire rows of cells, forming words, while bitlines connect cells column-wise, defining the word width (number of bits per word). Peripheral circuits like sense amplifiers, wordline

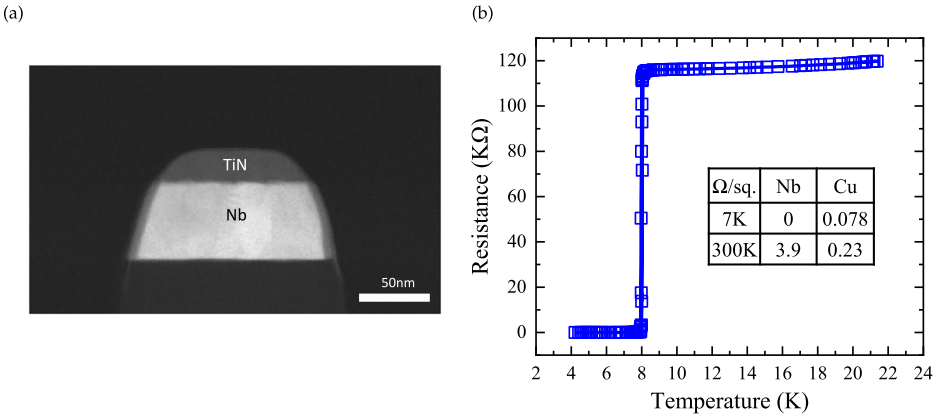


Fig. 6 | Structural view and electrical characterization of niobium (Nb) interconnects for cryogenic operation. a High-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) image of the fabricated Nb interconnect structure, showing the metal line geometry and integration within the dielectric stack. **b** Sheet resistance as a function of temperature for Nb and copper (Cu) interconnects. The horizontal axis represents temperature, while the vertical axis represents resistance in kilo-ohms per square. The plotted curves indicate that Nb undergoes a transition to a superconducting state below approximately 7.8 K, where its resistance approaches zero, while Cu retains finite resistance. The inset table summarizes representative sheet resistance values at 7 K and 300 K for both materials. This transition eliminates resistive parasitics in interconnects, enabling improved signal propagation and reduced power dissipation in cryogenic CMOS circuits.

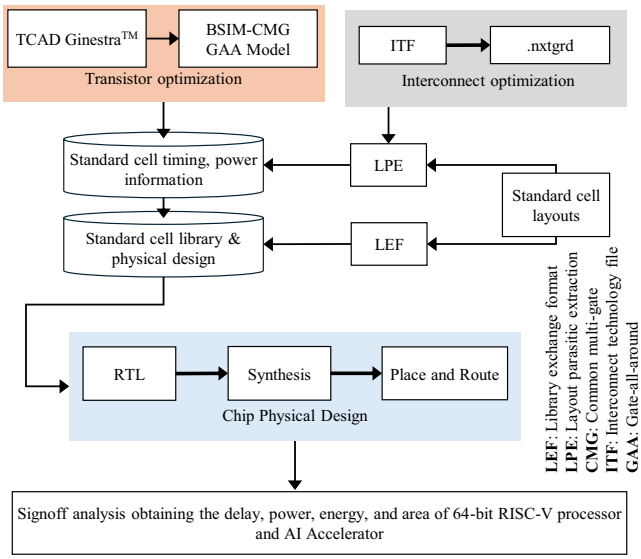


Fig. 7 | Complete RTL-to-GDS material-to-system evaluation flow for cryogenic GAA technology. The framework illustrates the integration of device, interconnect, and system-level modeling within a unified design flow. Transistor-level optimization is performed using Ginestra™ TCAD and translated into a calibrated BSIM-CMG (common multi-gate) compact model. In parallel, interconnect optimization is incorporated through modified interconnect technology files (ITF), enabling accurate modeling of parasitic effects. Standard cell layouts are generated and characterized to produce timing and power libraries, represented in Library Exchange Format (LEF). The register-transfer-level (RTL) design is synthesized and implemented through place-and-route to obtain the chip-level physical design. Layout parasitic extraction (LPE) is then applied to capture interconnect effects. Finally, sign-off analysis is performed to extract delay, power, energy, and area metrics for complex systems such as processors and AI accelerators. This flow enables consistent propagation of cryogenic device and interconnect optimizations to system-level performance evaluation.

drivers, and write drivers assist in read and write operations via the bitlines. Wordline drivers activate and disable the shared wordlines selecting a specific word.

We design and layout the 6T SRAM cells and peripheral circuits using the 2 nm nanosheet PDK. A custom memory compiler is developed to generate the array and place the periphery circuits as well as extract the post-layout netlist. Finally, a mixed-mode simulation testbench is used to validate the memory functionality and extract delay and power metrics. Iso-I_{OFF} cryogenic operation of SRAM at advanced technology nodes has been studied in detail[8,23], demonstrating improvements in delay and power at reduced supply voltages. However, applying threshold voltage tuning through work-function engineering (Opt-1) introduces significant trade-offs in SRAM stability in terms of static noise margin (SNM)[8,23]. In high-density cells, Hold and Read SNM have been shown to plummet by 50% and 87%, respectively, under *iso-*I_{OFF} conditions. Furthermore, this tuning reduces the slope of the inverter voltage transfer characteristics, degrading the normalized mean SNM from 0.413 to 0.287 and rendering the memory entirely inoperable when scaling the supply voltage to 0.2 V. As a result, such optimizations can compromise reliable SRAM operation. For this reason, our analysis of 6T SRAM-based arrays focuses exclusively on interconnect optimization (Opt-2), where performance benefits can be achieved without adversely impacting cell stability.

System-level analysis

In the subsequent stage, the RTL description of system-level design is synthesized into its physical implementation, utilizing both the cryogenic-aware libraries and optimized interconnect. This implementation enables evaluation of system-level performance and energy metrics. Two complex digital systems are selected: a 64-bit RISC-V (Rocket) processor[24], widely adopted as an open-source CPU baseline, and the compute fabric of an AI accelerator comprising a 64 × 64 array of multiply-accumulate (MAC) units enabling 4096 parallel operations (Fig. 10c). These case studies encompass both general-purpose processor workloads and massively parallel arithmetic accelerators, demonstrating the benefits of the proposed optimizations across diverse design classes. The sequential state of both designs and the register file of the RISC-V core are implemented using flip-flops from the standard cell library. Cache and other memory structures, usually implemented as SRAM macros, are treated as black-box, thus not contributing to the reported metrics. Memory macros are not present in the implemented designs. The analysis of memory is performed separately, as described in the previous subsection.

We set the baseline cell library as the library without optimizations and generate baseline gate-level netlists and physical implementation based on the delay and power information it contains. This baseline design is then used to extract and evaluate the metrics of both optimizations. This ensures that automatic optimizations introduced by EDA tools do not bias the comparative analysis.

Subsequently, comprehensive timing and power evaluation is performed on the baseline implementation using cryogenic-aware libraries for both optimizations across various supply voltages. Delay, power, and area are extracted during this stage. As the transistors operate under iso-I_{OFF} conditions at 7 K, leakage power remains constant.

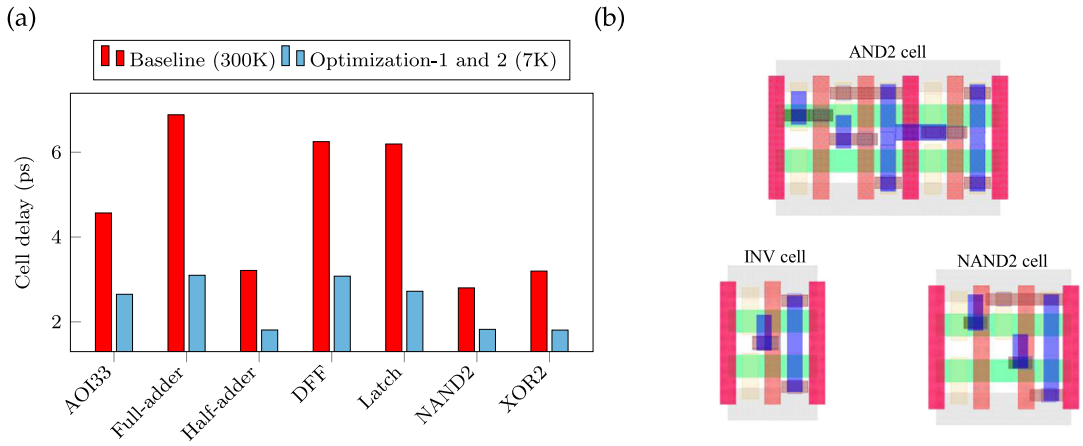


Fig. 8 | Impact of cryogenic operation on standard cell delay and layout implementation in 2 nm GAA technology. a Post-layout delay comparison of representative standard cells, including inverter (INV), NAND2, XOR2, latch, and D flip-flop (DFF), under baseline conditions at 300 K and optimized cryogenic conditions at 7 K. The bars represent delay values extracted from SPICE simulations and stored in standard cell libraries. The reduction in delay is primarily attributed to transistor-level optimization (Opt-1), as standard cell interconnects are limited to short local routing layers and are therefore minimally affected by interconnect optimization (Opt-2). **b** Example layouts of 3.5 track (3.5T) INV, NAND2, and AND2 standard cells implemented in a 2 nm GAA process design kit, illustrating the physical arrangement of nanosheet transistors and local interconnects. The AND2 cell is constructed by combining NAND2 and inverter structures. These results demonstrate that cryogenic-aware transistor optimization significantly improves standard cell performance while maintaining standard layout design practices.

Evaluating cryogenic effects on standard cells

The cryogenic cell library shows on average 1.9× reduction in cell delay when compared to the baseline. Standard cell layouts use only short interconnects confined to M0–M2 routing layers. Because these local wires contribute minimally to delay and power metrics, the benefit of optimization-2 is negligible at this level. As a result, the improvements observed in Fig. 8a are primarily driven by optimization-1.

Evaluating cryogenic effects on memory

The generated SRAM array, as shown in Fig. 9a consists of 64 words with each word having a length of 32 bits. To ensure that wordline drivers are placed in close proximity to the SRAM cells and therefore minimize routing delay, the array is divided into two banks of 64 wordline, 16-bitline sub-arrays. A total of 64 wordline drivers and 32 sense amplifiers, along with bitline precharge circuits, multiplexers, and write drivers that enable read and write operations are included in the SRAM array macro. During the design of the memory macro, interconnect layers M0 to M6 are used to ensure correct connectivity under density constraints of the SRAM array.

The SRAM macro heavily relies on long interconnects spanning from layers M0 to M6, making it more sensitive to interconnect optimization. As discussed previously, applying transistor-level optimization (Opt-1) introduces stability trade-offs in SRAM cells[8,23]. Therefore, this evaluation focuses exclusively on optimization-2 (interconnect optimization). The evaluation of the SRAM memory reveals improvements in both delay and power performance under optimization-2 compared to the baseline design. As shown in the figure, the normalized average SRAM delay is reduced by 18% and 19% when performing binary 1 and binary 0 write operations. This demonstrates that the proposed optimization effectively enhances write speed. In terms of power consumption, the average write and read power are reduced by 32% and 26%, respectively, thereby highlighting substantial power savings. Overall, optimization-2 achieves notable improvements in write performance and energy efficiency, making it highly beneficial for memory subsystems targeting low-power and high-speed applications.

Evaluating cryogenic effects on RISC-V processor

The RISC-V processor serves as a full-system benchmark where both transistor-level (Opt-1) and interconnect-level (Opt-2) optimizations are simultaneously present. The design makes extensive use of all six routing layers, allowing combined cryogenic improvements to be captured at the chip level. In Fig. 10a, we show the RISC-V processor results for room and cryogenic temperatures under varying supply voltages. The maximum processor delay, also known as the critical path delay, represents the longest combinational path between two registers. The baseline configuration at 300 K exhibits the highest delay across the voltage range. For voltages below 0.5 V, library characterization fails as essential cells, such as flip-flops, cannot be characterized. At 7 K, the non-optimized transistor shows a slight delay improvement at higher voltages. However, due to increased V_T at cryogenic temperatures, the minimum operating voltage rises to 0.6 V, slightly higher than the baseline.

The RISC-V using GAA transistors under optimization-1 shows a minimum and maximum reduction in delay by 0.15 ns and 0.42 ns, respectively, compared to the baseline. The smallest reduction is observed at high voltage supply values, and the highest reduction at low voltage levels. This highlights the importance of a cryogenic-aware transistor. In addition, these optimized transistors enable the standard cells to operate down to 0.3 V. Delay remains almost constant across the voltage range, displaying only a slight increase at low power supply voltage values. Optimization-2 offers an additional average delay reduction of 0.115 ns. Nevertheless, the improvements are not as substantial as those offered by optimization-1.

As shown in the bottom figure of Fig. 10a, the energy efficiency, quantified using the energy-delay product (EDP), follows trends similar to delay, where the baseline at 300 K performs worst, and the case where both optimizations are considered yields the best results. The most energy-efficient design conditions are located where EDP reaches its minimum value. The baseline result at 300 K shows a minimum EDP at 0.6 V. Because of the increase in V_T for the baseline at 7 K, the minimum shifted to 0.7 V. For both Optimizations 1 and 2, EDP decreases monotonically with voltage, showing no distinct minimum. In both cases, the most efficient operation voltage is 0.3 V.

Evaluating cryogenic effects on AI accelerator

As shown in Fig. 10d, the results of the AI accelerator compute fabric demonstrate 3× improvement in operational frequency, reflected in tera operations per second (TOPS) metric, translating to significantly faster execution of workloads under the evaluated cryogenic conditions. A 7.7× increase is observed in energy efficiency, measured in TOPS/W. We have observed 23× boost in the compound performance-energy efficiency metric $TOPS^2/W$, a metric used to highlight high-performance and energy-efficient AI accelerators. These figures characterize the MAC compute fabric and exclude on-chip memory and data movement.

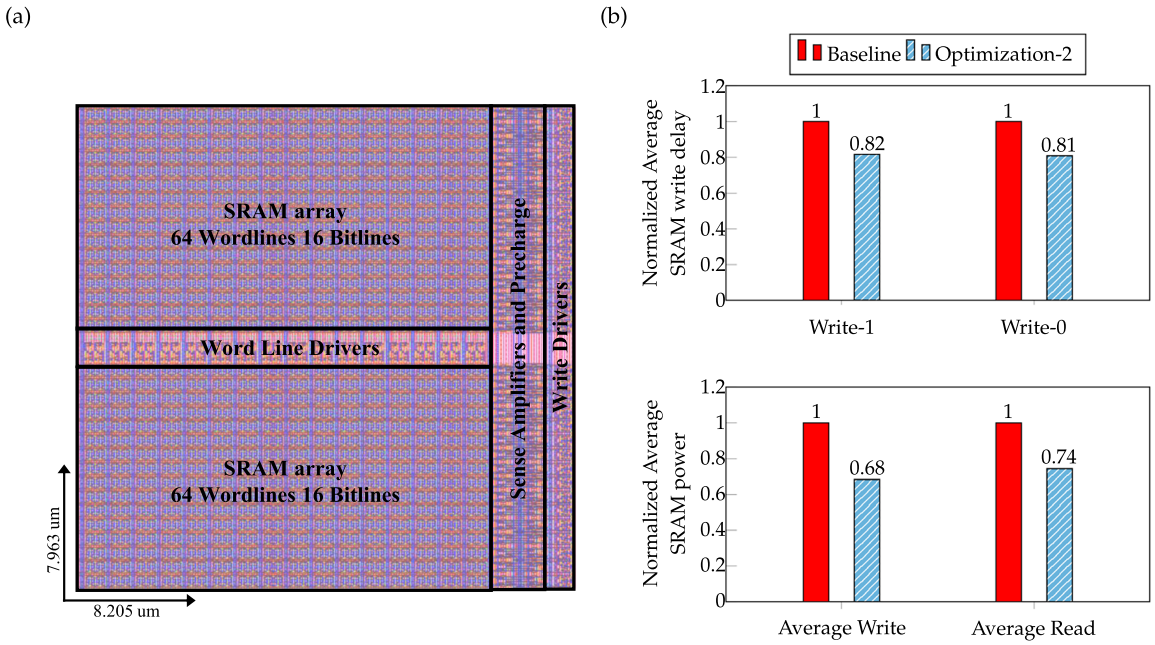


Fig. 9 | SRAM macro layout and performance improvement through interconnect optimization under cryogenic operation. a Layout of the 6-transistor (6T) SRAM macro consisting of 64 wordlines and 32 bitlines, including peripheral circuits such as wordline drivers, sense amplifiers, precharge circuits, and write drivers. The physical dimensions of the macro are indicated, and the layout highlights the use of multiple interconnect layers to route signals across the array. **b** Normalized delay and power comparison between the baseline design and interconnect-optimized design (Opt-2). The horizontal axis distinguishes between write-1 and write-0 operations for delay, and average write and read operations for power. The vertical axis represents normalized values, where the baseline is set to 1. Bars correspond to baseline and optimized cases. The results show reduced delay and power under optimization, primarily due to decreased interconnect resistance in long routing paths. These findings demonstrate that interconnect optimization significantly enhances SRAM performance and energy efficiency without modifying the transistor structure.

Evaluating cryogenic effects on self-heating

The increase in I_{ON} enabled by the steeper SS directly translates to a lower operating voltage at the same or better performance levels, as observed in the RISC-V case. In turn, this significantly enhances reliability, as SHE and defect-generation mechanisms exhibit an exponential dependence on supply voltage. Our analysis, supported by advanced multi-physics Ginestra™ simulations, demonstrates an approximately 80% reduction in SHE-induced ΔT as shown in Fig. 10b. This substantial mitigation of local temperature rise alleviates thermal stress on the channel and surrounding materials, slowing down defect generation and propagation. It also reduces the risk of thermal hotspots propagating to the interconnects, helping maintain Nb wires within the superconducting regime. Collectively, these improvements ensure that CMOS circuits operating under cryogenic conditions are not only higher performing and more energy-efficient, but also inherently more reliable.

Discussion

This work investigates whether device- and material-level optimizations can unlock the full potential of cryogenic CMOS for future high-performance and AI computing systems. Our findings confirm that both transistor work function tuning and superconducting interconnect integration are critical enablers for realizing substantial performance and energy-efficiency gains at low temperatures. Importantly, these improvements arise from exploiting fundamental physical advantages of cryogenic operation–namely, steeper subthreshold slopes and the emergence of superconductivity.

The methodology employed in this work establishes a rigorous, quantitative bridge between device physics and system-level performance. By calibrating multi-physics TCAD Ginestra™ models against experimental data, extending compact models for cryogenic conditions, and validating complete digital systems through RTL-to-GDS implementation, we apply an end-to-end evaluation framework that can be reused across future optimization strategies. This framework provides a practical pathway to systematically assess material, device, and interconnect innovations in a unified simulation-to-design flow, enabling cross-layer co-optimization from device- to system-level. All results assume operation within the superconducting regime by means of limited current densities and reduced self-heating. Although the system- and memory-level analyses are currently presented separately, in an actual circuit both must coexist on the same die. In this case, Optimization-1 must be omitted from the SRAM region and applied only to the standard-cell logic region. We believe this is achievable from a manufacturing perspective using the same multi-work-function patterning already used to provide multi-threshold-voltage (multi-Vt) transistors on silicon, at the cost of additional masks. This technique is already employed at the granularity of individual standard cells; the logic-versus-SRAM partition required here is substantially coarser.

Relative to previous studies that focused on either device characterization or small-scale circuit operation at cryogenic temperatures, our results extend the analysis to full RTL-to-GDS implementations, including a 64-bit RISC-V processor and a parallel AI accelerator. The observed improvements in performance and energy efficiency, combined with the intrinsic reliability benefits of reduced self-heating, position cryogenic CMOS as a scalable and sustainable approach for extending digital scaling beyond the limits of Moore's law. All reported TOPS/W and TOPS^2/W values are on-chip electrical metrics extracted at the device and interconnect level and do not include the input power of the cryogenic refrigeration system. In a full deployment, the wall-plug efficiency would additionally depend on cryocooler overhead, which at these temperatures is significant but is often amortized in environments where the cooling infrastructure is already provisioned (e.g., quantum control). The gains reported here should therefore be read as the electrical efficiency of the circuit, not end-to-end system efficiency.

Methods

Device-level modeling

Device-level simulations were carried out using Ginestra™, a state-of-the-art semiconductor device simulation platform capable of modeling the

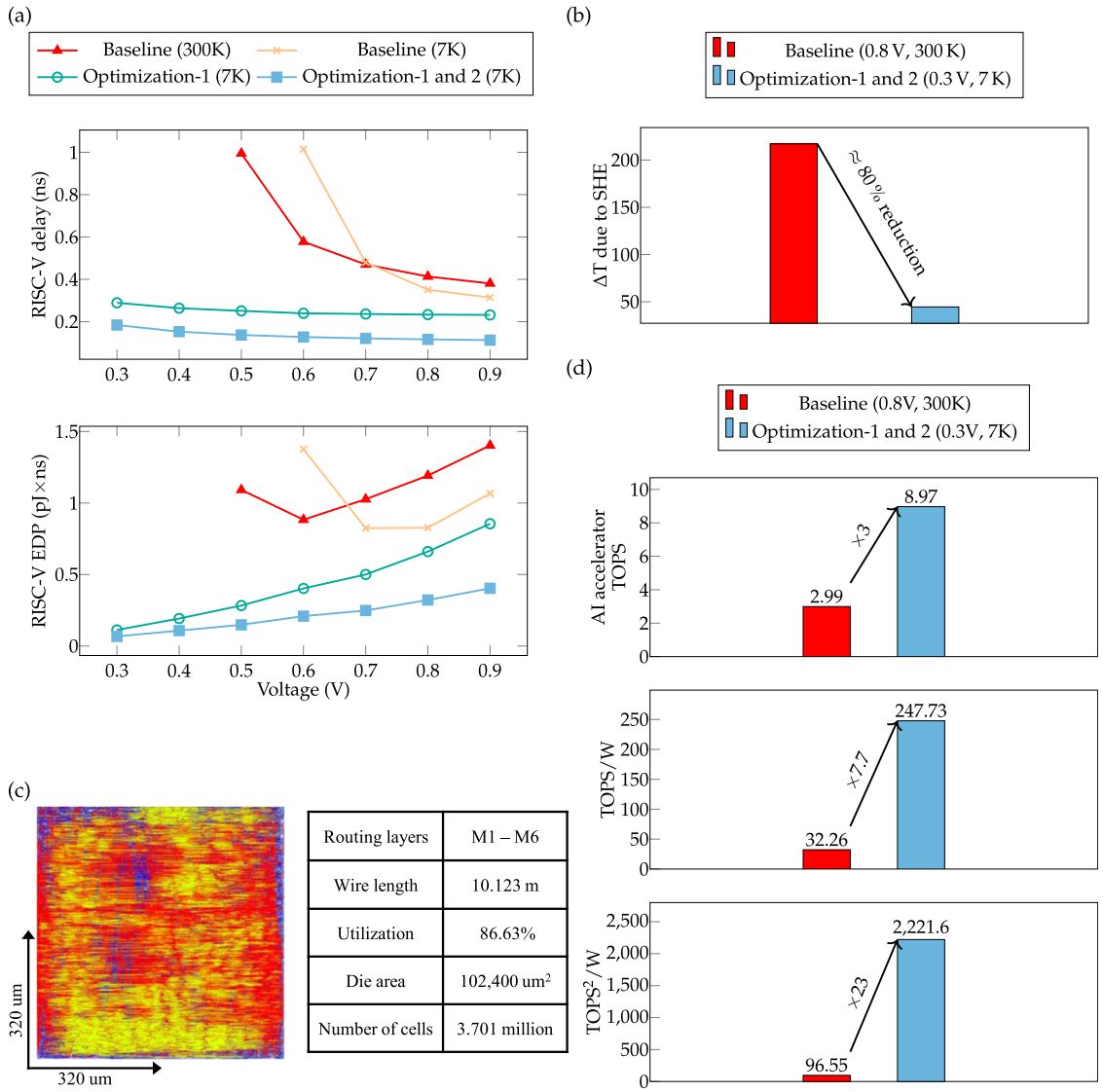


Fig. 10 | System-level performance, energy efficiency, and self-heating improvements under cryogenic transistor and interconnect optimization. **a** Delay and energy-delay product (EDP) of a 64-bit RISC-V processor as a function of supply voltage for baseline operation at 300 K, baseline at 7 K, transistor optimization (Opt-1), and combined transistor and interconnect optimization (Opt-1 and Opt-2) at 7 K. The curves show that cryogenic-aware transistor optimization enables operation at lower voltages and reduces delay and EDP. **b** Channel self-heating, represented as temperature rise (ΔT), comparing baseline operation at 300 K at 0.8 V and optimized operation at 7 K at 0.3 V, demonstrating a significant reduction in self-heating under cryogenic conditions. **c** Physical layout of the compute fabric of the AI accelerator, indicating routing layers, total wire length, utilization, die area, and number of standard cells. **d** Performance and energy efficiency metrics of the AI accelerator compute fabric, including tera operations per second (TOPS), TOPS per watt (TOPS/W), and the combined metric $TOPS^2/W$. Bars represent baseline operation at 300 K and optimized operation at 7 K, showing improvements of 3× in performance, 7.7× in energy efficiency, and 23× in the combined metric. These results demonstrate that combined device and interconnect optimization enables substantial system-level gains in cryogenic CMOS.

electrical behavior of both conventional and emerging technologies through a detailed physics-based framework. The tool captures charge and ion transport, charge trapping, degradation mechanisms, and material modifications within the simulated stacks. Defects are represented as discrete entities characterized by two fundamental parameters–the thermal ionization energy (E_{th}) and the relaxation energy (E_{rel})–which define their atomic structure[11,25] and enable simulation of their interactions with charge carriers, material bands, electrodes, and with each other.

Charge trapping and transport are modeled by simultaneously considering multiple physical mechanisms, including direct and Fowler-Nordheim tunneling, trap-assisted tunneling (TAT), drift-diffusion in conduction, valence, and defect bands, and thermionic emission. The dominant mechanism depends on external conditions such as temperature, applied bias, and the device state. Capture and emission times are computed within the multi-phonon framework, incorporating carrier-phonon coupling and lattice relaxation processes[11,26].

The transport equations are solved self-consistently together with Poisson's equation, taking into account defect charge states and occupation, voltage polarity, and the detailed band profile of the material stack. Self-heating is modeled by solving Fourier's heat equation, with thermal resistances at the source, drain, gate, and substrate contacts evaluated using an approach similar to that in refs. [27,28].

Transistor compact modeling

Transistor behavior was modeled using an extended BSIM-CMG compact model calibrated through Synopsys HSPICE. Key parameters governing the transistor behavior were swept systematically, and the error between TCAD Ginestra™ and HSPICE outputs was minimized using a stepwise

refinement of the parameter set. The same calibration procedure was applied using TCAD Ginestra™ data generated at cryogenic temperatures (down to 7 K), ensuring that temperature-dependent effects were captured directly through parameter tuning rather than extrapolation.

Interconnect optimization

To enable accurate interconnect optimization simulations at cryogenic temperatures, the interconnect technology file (.itf) used by Synopsys StarRC was modified for superconductivity. StarRC was then used to generate parasitic capacitance and resistance tables. The modified .itf assigns zero resistivity to superconducting layers, so that StarRC produces capacitance-only parasitics for the Nb interconnects below 7.8 K.

Standard cell and memory design

Standard cells and memory 6T-SRAM cells with their periphery circuits were designed and simulated in Synopsys Custom Compiler and PrimeSim design environment. The respective layouts were created in Custom Compiler while meeting all design-rule and connectivity checks. Parasitics for these layouts were extracted using StarRC. Synopsys PrimeLib was then used to characterize the standard cells at room and cryogenic temperatures, producing libraries that reflect the transistor-level optimizations. Regarding the memory instance, a dedicated mixed-mode simulation framework was developed to extract delay and power results.

System-level implementation and analysis

At the system level, the RTL design was synthesized into gate-level netlists using Synopsys Design Compiler and implemented as a physical design using Synopsys ICC2, which outputs the placed-and-routed GDSII file. To aid in the place and route of the design, a technology file was developed for this PDK. The technology file contains information on the routing layers and was used by ICC2 to perform placement and routing. Interconnect parasitics were extracted from this GDSII file using StarRC together with the previously generated parasitic tables. Synopsys PrimeTime is then used to perform sign-off static timing analysis, using the extracted parasitic netlists from different interconnect optimizations and the temperature-dependent libraries generated under different transistor optimizations to obtain final power and delay numbers.

Data availability

The datasets generated and/or analyzed during the current study are not publicly available due to confidentiality restrictions and export control rules imposed when working on advanced 2 nm gate-all-around technologies but are available from the corresponding author on reasonable request.

Code availability

The code generated and/or analyzed during the current study is not publicly available due to confidentiality restrictions and export control rules imposed when working on advanced 2 nm gate-all-around technology and the proprietary commercial software (Ginestra™ and Synopsys EDA tools) used in this work, but is available from the corresponding author on reasonable request.

Received: 24 November 2025; Accepted: 23 September 2026;
Published online: 02 November 2026

References

1. Charbon, E. et al. Cryo-CMOS for quantum computing. In *2016 IEEE International Electron Devices Meeting (IEDM)*, 13.5.1–13.5.4 https://doi.org/10.1109/IEDM.2016.7838410 (2016).
2. Devoret, M. H. & Schoelkopf, R. J. Superconducting circuits for quantum information: an outlook. *Science* **339**, 1169–1174 https://doi.org/10.1126/science.1231930 (2013).
3. Radamson, H. H. et al. State-of-the-art and future perspectives in advanced CMOS technology. *Nanomaterials* **10** https://doi.org/10.3390/nano10081555 (2020).
4. Loubet, N. et al. Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET. In *2017 Symposium on VLSI Technology*, T230–T231 https://doi.org/10.23919/VLSIT.2017.7998183 (2017).
5. Nagy, D. et al. Benchmarking of FinFET, nanosheet, and nanowire FET architectures for future technology nodes. *IEEE Access* **8**, 53196–53202 https://doi.org/10.1109/ACCESS.2020.2980925 (2020).
6. Saligram, R., Prasad, D., Pietromonaco, D., Raychowdhury, A. & Cline, B. A 64-bit ARM CPU at cryogenic temperatures: Design technology co-optimization for power and performance. In *2021 IEEE Custom Integrated Circuits Conference (CICC)* https://doi.org/10.1109/CICC51472.2021.9431559 (2021).
7. Prasad, D. et al. Cryo-computing for infrastructure applications: a technology-to-microarchitecture co-optimization study. In *2022 International Electron Devices Meeting (IEDM)* https://doi.org/10.1109/IEDM45625.2022.10019436 (2022).
8. Saligram, R., Raychowdhury, A. & Datta, S. The future is frozen: cryogenic CMOS for high-performance computing. *Chip* **3**, 100082 https://doi.org/10.1016/j.chip.2023.100082 (2024).
9. Applied Materials Ginestra™. https://www.appliedmaterials.com/us/en/semiconductor/solutions-and-software/software-solutions/ginestra-simulation-platform.html (2022).
10. Kar, A., Klemme, F., Chauhan, Y. S. & Amrouch, H. On the severity of self-heating in FDSOI at cryogenic temperatures: In-depth analysis from transistors to full processor. In *2024 IEEE International Reliability Physics Symposium (IRPS)*, 1–6 https://doi.org/10.1109/IRPS48228.2024.10529429 (2024).
11. Vandelli, L. et al. A physical model of the temperature dependence of the current through SiO_2/HfO_2 stacks. *IEEE Trans. Electron Devices* **58**, 2878–2887 https://doi.org/10.1109/TED.2011.2158825 (2011).
12. BSIM-CMG Group. https://bsim.berkeley.edu/models/bsimcmg/ (2024).
13. Duarte, J. P. et al. BSIM-CMG: Standard FinFET compact model for advanced circuit design. In *ESSCIRC Conference 2015—41st European Solid-State Circuits Conference (ESSCIRC)*, 196–201 https://doi.org/10.1109/ESSCIRC.2015.7313862 (2015).
14. Pahwa, G., Kushwaha, P., Dasgupta, A., Salahuddin, S. & Hu, C. Compact modeling of temperature effects in FDSOI and FinFET devices down to cryogenic temperatures. *IEEE Trans. Electron Devices* **68**, 4223–4230 https://doi.org/10.1109/TED.2021.3097971 (2021).
15. Kar, A., Parihar, S. S., Klemme, F., Chauhan, Y. S. & Amrouch, H. Benchmarking cryogenic circuits using 5 nm FinFETs for quantum processing. In *2025 IEEE International Symposium on Circuits and Systems (ISCAS)*, 1–5 https://doi.org/10.1109/ISCAS56072.2025.11044153 (2025).
16. Parihar, S. S. et al. Cryogenic CMOS for quantum processing: 5-nm FinFET-based SRAM arrays at 10 K. *IEEE Trans. Circuits Syst. I Regul. Pap.* **70**, 3089–3102 https://doi.org/10.1109/TCSI.2023.3278351 (2023).
17. Yeap, G. et al. 2 nm platform technology featuring energy-efficient nanosheet transistors and interconnects co-optimized with 3DIC for AI, HPC and mobile SoC applications. In *2024 IEEE International Electron Devices Meeting (IEDM)*, 1–4 https://doi.org/10.1109/IEDM50854.2024.10873475 (2024).
18. Manabe, K. et al. Comprehensive analysis of determining factors of the effective work function of TiAlC/TiN metal gates for advanced gate-all-around CMOS integration. *Jpn. J. Appl. Phys.* **64**, 01SP28 https://doi.org/10.35848/1347-4065/ada9f8 (2025).
19. Wu, Z. et al. $LaSiO_x$- and Al_2O_3-inserted low-temperature gate-stacks for improved BTI reliability in 3-D sequential integration. *IEEE Trans. Electron Devices* **69**, 915–921 https://doi.org/10.1109/TED.2022.3141983 (2022).
20. Huang, Z. et al. Dipole engineering at HfO_2/SiO_2 interface by ultra-thin Al_2O_3 to modulate flat-band voltage for cryogenic temperatures.

Vacuum **234**, 114073 https://doi.org/10.1016/j.vacuum.2025.114073 (2025).
21. Iguchi, N. et al. 200nm-pitched, superconducting Nb interconnects for cryo-CMOS application. In *2024 IEEE International Interconnect Technology Conference (IITC)*, 1–3 https://doi.org/10.1109/IITC61274.2024.10732217 (2024).
22. Numata, H. et al. Superconducting Nb interconnects for cryo-CMOS and superconducting digital logic applications. *Jpn. J. Appl. Phys.* **63**, 04SP73 https://doi.org/10.35848/1347-4065/ad37c1 (2024).
23. Parihar, S. S., Thomann, S., Pahwa, G., Chauhan, Y. S. & Amrouch, H. 5nm FinFET cryogenic SRAM evaluation for quantum computing. In *2023 Device Research Conference (DRC)*, 1–2 https://doi.org/10.1109/DRC58590.2023.10186959 (2023).
24. Berkeley Architecture Research - Chipyard. https://chipyard.readthedocs.io/.
25. Padovani, A. et al. A sensitivity map-based approach to profile defects in MIM capacitors from *I* - *V*, *C* - *V*, and *G* - *V* measurements. *IEEE Trans. Electron Devices* **66**, 1892–1898 https://doi.org/10.1109/TED.2019.2900030 (2019).
26. Zhang, M., Huo, Z., Yu, Z., Liu, J. & Liu, M. Unification of three multiphonon trap-assisted tunneling mechanisms. *J. Appl. Phys.* **110**, 114108 https://doi.org/10.1063/1.3662195 (2011).
27. Yoo, C., Chang, J., Seon, Y., Kim, H. & Jeon, J. Analysis of self-heating effects in multi-nanosheet FET considering bottom isolation and package options. *IEEE Trans. Electron Devices* **69**, 1524–1531 https://doi.org/10.1109/TED.2022.3141327 (2022).
28. Yan, J.-Y. et al. Thermal resistance modeling of back-end interconnect and intrinsic FinFETs, and transient simulation of inverters with capacitive loading effects. In *2016 IEEE International Electron Devices Meeting (IEDM)*, 35.6.1–35.6.4 https://doi.org/10.1109/IEDM.2016.7838550 (2016).

Acknowledgements

We thank Wei-Ji Chao, Paul Genssler, S. Mansour, Noor Elabyad, and Jonas Bader from TUM for their support in standard cells, RTL and SRAM memory.

Author contributions

A.M. performed the analysis at the standard cell library and chip level. A.K. calibrated the BSIM-CMG transistor model. A.P. and L.V. calibrated the TCAD Ginestra™ devices. M.T. and T.T. fabricated and measured the niobium interconnect material. L.L. and G.T. guided the device and interconnect modeling and simulations. H.A. conceived the idea and supervised the overall work from the device to the system level. All authors contributed to the writing of the manuscript and provided feedback.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Albi Mema or Hussam Amrouch.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026