






Received 8 June 2023; revised 20 July 2023; accepted 27 July 2023; date of publication 1 August 2023;
date of current version 30 August 2023.

Digital Object Identifier 10.1109/TQE.2023.3300833

Cryogenic Embedded System to Support Quantum Computing: From 5-nm FinFET to Full Processor

PAUL R. GENSSLER¹  (Member, IEEE), FLORIAN KLEMME¹  (Member, IEEE),
SHIVENDRA SINGH PARIHAR^{1,2} (Member, IEEE),
SEBASTIAN BRANDHOFER³ (Graduate Student Member, IEEE),
GIRISH PAHWA⁴  (Member, IEEE), ILIA POLIAN³  (Senior Member, IEEE),
YOGESH SINGH CHAUHAN²  (Fellow, IEEE),
AND HUSSAM AMROUCH^{1,5,6} (Member, IEEE)

¹Chair of Semiconductor Test and Reliability (STAR), University of Stuttgart, 70174 Stuttgart, Germany

²Nanolab, Department of Electrical Engineering, IIT Kanpur, Kanpur 208016, India

³Chair of Hardware Oriented Computer Science, University of Stuttgart, 70174 Stuttgart, Germany

⁴Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720 USA

⁵Chair of AI Processor Design, Technical University Munich, 80333 Munich, Germany

⁶Munich Institute of Robotics and Machine Intelligence, Technical University Munich, 80333 Munich, Germany

Corresponding author: Paul R. Genssler (e-mail: genssler@iti.uni-stuttgart.de).

This work was supported in part by the Carl Zeiss Foundation, in part by the Ministry of Economic Affairs, Labour and Tourism Baden Württemberg, in the frame of the Competence Center Quantum Computing Baden-Württemberg under Project “QORA,” in part by Advantest as part of the Graduate School “Intelligent Methods for Test and Reliability” (GS-IMTR) at the University of Stuttgart, and in part by the German Research Foundation (DFG) “Open Access Publication Funding/2023-2024/University of Stuttgart” under Grant 512689491.

ABSTRACT Quantum computing can enable novel algorithms infeasible for classical computers. For example, new material synthesis and drug optimization could benefit if quantum computers offered more quantum bits (qubits). One obstacle for scaling up quantum computers is the connection between their cryogenic qubits at temperatures between a few millikelvin and a few kelvin (depending on qubit type) and the classical processing system on chip (SoC) at room temperature (300 K). Through this connection, outside heat leaks to the qubits and can disrupt their state. Hence, moving the SoC into the cryogenic part eliminates this heat leakage. However, the cooling capacity is limited, requiring a low-power SoC, which, at the same time, has to classify qubit measurements under a tight time constraint. In this work, we explore for the first time if an off-the-shelf SoC is a plausible option for such a task. Our analysis starts with measurements of state-of-the-art 5-nm fin-shaped field-effect transistors (FinFETs) at 10 and 300 K. Then, we calibrate a transistor compact model and create two standard cell libraries, one for each temperature. We perform synthesis and physical layout of a RISC-V SoC at 300 K and analyze its performance at 10 K. Our simulations show that the SoC at 10 K is plausible but lacks the performance to process more than a few thousand qubits under the time constraint.

INDEX TERMS Cryogenic CMOS, 5-nm fin-shaped field-effect transistor (FinFET), hyperdimensional computing, machine learning classification, quantum computing, system on chip (SoC).

I. INTRODUCTION

Quantum computing is a pertinent avenue of resolution for challenging computational problems that supersede the realm of classical computing. Such problems include, but are not limited to, new material synthesis, drug optimization [1], and integer factorization [2]. All of these demand a huge quantity of high-fidelity quantum bits (qubits). This calls

for quantum computer upscaling, realized through specialized CMOS-based compute circuits. Such circuits can beneficially bridge the gap between the quantum and classical domains by directly processing information obtained from the qubits. CMOS-based circuits can perform the necessary classification to digitize the readout and many other essential tasks.

A. INEVITABLE NEED FOR CRYOGENIC CIRCUITS

Control circuits operating at room temperature (i.e., 300 K) pose a considerable constraint on quantum computers, particularly given that each qubit might require individual control [3]. Conventionally, qubits operate at close to absolute zero (e.g., 10 mK) since they retain their superimposed state for longer at cryogenic temperatures. However, this conjures an input–output bottleneck, as highlighted in a recent experiment that shows that isolating and controlling merely 53 qubits require enormous structural overheads [3], [4]. This includes 200 wideband coaxial cables, 45 microwave circulators, and a rack of electronic circuits. Notwithstanding the isolation, a heat flux leakage can occur from the control circuits to the qubits placed within the cryogenic system. This is spurred by a temperature gradient between 300 and 0.1 K induced at the two ends of each wire that can put the entire quantum system at peril.

Moreover, the short qubit decoherence time, ranging from ns to ms, exacerbates existing timing constraints. Notably, the qubit is acutely sensitive to noise and heat. The timing constraints are further stressed by huge latencies caused by lengthy cables. Thus, to overcome the above and effectively realize quantum computer up scaling, CMOS circuits must operate at cryogenic temperatures, i.e., in the vicinity of the qubits. This ensures the unimpeded and coherent functioning of numerous qubits.

B. KEY CHALLENGES BEHIND CRYOGENIC CIRCUITS

Operating CMOS circuits at cryogenic temperatures is riddled with a unique set of challenges pertaining to power optimization. This is because the circuits face power dissipation constraints at low temperatures. If existing power constraints are unaddressed, the resulting heat can affect not only the state of the qubits but also lead to the worst-case outcome of qubit destruction. Hence, the power constraints have the highest priority, even higher than the achievable clock frequency.

Circuits operating at cryogenic temperatures require an extremely tight power budget. The control circuits strictly function within a power upper limit of only 100 mW at a temperature of 10 K, further lowering to 10 mW at 0.1 K [5]. In addition, computations have to be fast enough to satisfy the time constraints dictated by the short qubit coherence time. Given these caveats, a sophisticated and reliable CMOS-based circuit operating at cryogenic temperatures has to express the characteristic features of processing qubit information: 1) extremely rapidly and 2) at ultralow power.

C. NEED FOR A CRYOGENIC SYSTEM ON CHIP (SOC)

An SoC combines the control and readout circuitry for the manipulation and measurement of qubits with a general-purpose processor. The addition of this processor enables the execution of arbitrary software codes, removing the dependency on dedicated hardware for every task or connections with the 300 K domain. This advantage was recently recognized by Intel and they demonstrated the first cryogenic

SoC for quantum computing [6]. However, their focus was on the implementation of the qubit circuitry, and they did not evaluate the capabilities of the included processor. Thus, it is unclear what processing can be done under the strict power budget that can be spared at cryogenic temperatures. Classification of the quantum measurements is not the only task for the classical processing part of the circuit. A general-purpose processor with its high flexibility is required to enable crucial tasks, such as running calibration protocols, loading the next quantum computation, and improve the runtime of popular quantum computing paradigms relying on classical processing, such as dynamic circuits [7] or variational quantum algorithms [8], [9]. Ultimately, to achieve fully error-corrected quantum computers, complex quantum error correction protocols have to be executed.

Dedicated hardware solutions for each of these tasks are costly and slow to develop. In the fast-paced field of quantum computing, the hardware could be outdated before it was even deployed. The processor included in [6] is much more flexible, but might miss an important instruction or be limited in the available memory. A new cryogenic SoC would have to be designed. Off-the-shelf SoCs, designed for room temperature use, are available in a wide range of specifications and capabilities and could quickly be swapped in and out, depending on the requirements of the tasks. However, the question arises if deploying such SoC is even plausible in a quantum system, with power consumption (i.e., heat dissipation) and processing speed being critical. Both are impacted as a CMOS transistor at 10 K exhibits different power and timing characteristics.

D. NEED FOR A CRYOGENIC-AWARE TRANSISTOR COMPACT MODEL

State-of-the-art SPICE models do not capture the indisputable influence of unconventionally low temperatures on the physics of the semiconductor transistors. The underlying changes are marked by a decrease of leakage current and transistor subthreshold swing (SS), and an increase of carrier mobility and transistor threshold voltage. Thus, SPICE models are ill-equipped to account for the aforementioned fundamental processes at cryogenic temperatures, and research in this direction is presently infantile. Without a cryogenic-aware transistor compact model, not only correct SPICE simulations are not possible, but standard cell library characterization (which is indispensable for creating cryogenic-aware cell libraries for logic synthesis) is also not possible.

E. OUR MAIN CONTRIBUTIONS WITHIN THIS WORK

We are the first to explore the plausibility of deploying an off-the-shelf SoC at cryogenic temperatures for the classification of the quantum measurements. Fig. 1 provides an overview of our work and serves as an outline. To enable our exploration, we first measure the characteristics of a state-of-the-art 5-nm fin-shaped field-effect transistor (FinFET) transistor at room temperature (300 K) and at cryogenic temperature (10 K). Using those measurements, we calibrate the modified cryogenic-aware Berkeley Short-channel IGFET Model

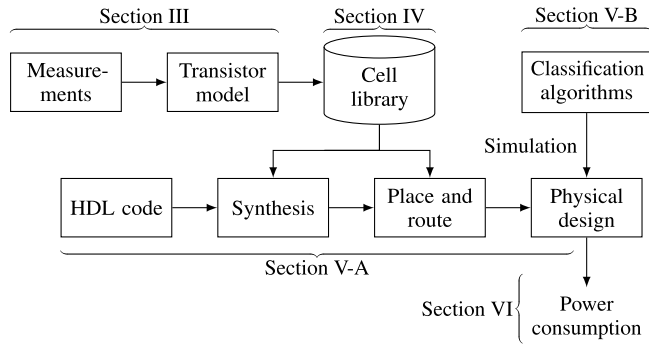


FIGURE 1. Outline of this article and an overview of our modeling. We cover the whole stack, from physical transistor measurements through cell library characterization to the physical design simulation of the full system on chip.

common multigate (BSIM-CMG) transistor compact model to reproduce the measurements. Two standard cell libraries are characterized by employing this new compact model. We perform logic synthesis and physical design of a RISC-V SoC with the 300 K standard cell library as a baseline for the off-the-shelf system. Then, we perform power and timing analysis using the 10 K library to explore the impact of this significant change in temperature on the SoC. Finally, we simulate the execution of two classification algorithms to answer the question if an off-the-shelf SoC can classify the qubit measurements under tight power and time constraints.

II. OPERATION OF QUANTUM COMPUTERS

An n -qubit quantum computer can store, manipulate, and measure an n -qubit quantum state $|\psi\rangle$ defined as

$$|\psi\rangle = \sum_{x \in \{0,1\}^n} \alpha_x |x\rangle \quad (1)$$

where $|x\rangle$ are basis states and α_x are complex probability amplitudes whose modulus squared sum up to one [10]. Upon measuring $|\psi\rangle$, the bit string x , corresponding to the basis state $|x\rangle$, is read with probability $|\alpha_x|^2$. The target quantum state must, therefore, be prepared and measured repeatedly to obtain a precise distribution of the measurement bit strings. When using a quantum computer to solve problems, such as integer factorization [2] or the simulation of a molecule [1], one or multiple quantum states are prepared and measured sequentially to estimate the desired solution.

A quantum computer must typically be calibrated before it can start manipulating and measuring quantum states. During calibration, the quantum state manipulation primitives are fine-tuned to the prevailing operational parameters of the quantum computer, and a classifier is trained that maps the electrical signal of the quantum computer's measurement apparatus to its corresponding bit value [11], [12]. For the IBM quantum computers based on superconducting qubits, typically a boxcar integrator is used to project the measurement signal onto the I/Q plane where a single-qubit measurement results in a complex number that is represented as an in-phase and quadrature component [11], [12]. As seen in Fig. 2, the

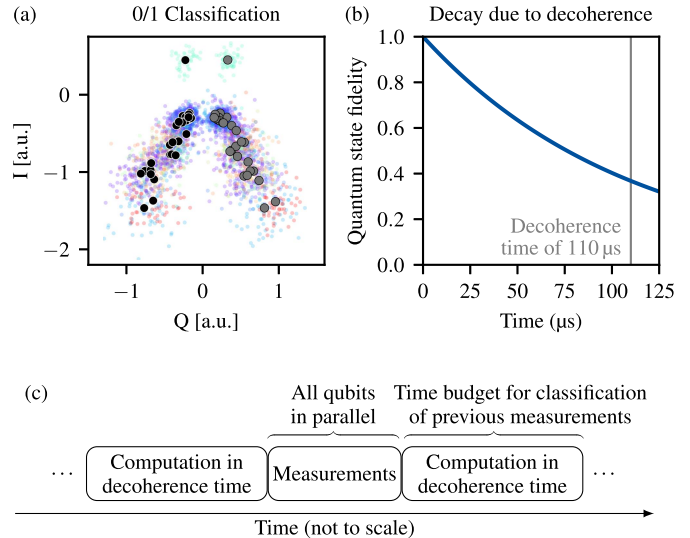


FIGURE 2. (a) Our measurements of the phase and amplitude of electrical signals from 27 qubits of an IBM Falcon quantum processor in the I/Q plane. For each qubit, a pair of black and gray dots represents the mean I/Q values for the measurement signal corresponding to the two qubit states, which is obtained during a calibration phase. The colored dots are measurements that are classified as 0 or 1 based on their proximity to that qubit's black or gray dots, respectively. The challenge is the (b) drop in quantum state fidelity limiting the runtime budget to the decoherence time of the IBM Falcon quantum processor. (c) New quantum computation is started after measuring the qubits. Hence, classifying the latest measurements has to be completed within the decoherence time at the latest to not bottleneck the overall quantum computer.

measurement signals of the i th qubit are generally in close proximity in the I/Q plane if they correspond to the same measurement outcome. The measurement classifier is trained by the data obtained through preparing and measuring each qubit individually in the $|0\rangle$ and $|1\rangle$ basis state while ignoring the remaining qubits. After calibration, the quantum state manipulation primitives and the measurement classifier are available for quantum computations.

IBM Quantum currently offers the largest quantum computer based on 127 superconducting qubits over the cloud via their qiskit framework [13] and plans to build a quantum computer with over 4000 qubits by 2025 [14]. The classified and also the raw I/Q plane measurement data of quantum computations are accessible through the qiskit framework and used in this work.

III. CRYOGENIC CMOS TRANSISTORS

Operating a CMOS transistor at cryogenic temperatures offers multiple advantages. A smaller SS, lower leakage current, higher mobile charge carriers' mobility, reduction in thermal noise, and parasitic resistances are a few to name. The smaller SS leads to near-ideal steep switching and reduction in over-the-barrier charge carriers' transport results in a lower OFF-state current I_{OFF} and higher mobility due to lower carrier scattering results in a higher ON-state current I_{ON} . These remarkable improvements in transistor characteristics are not new and have been an active area of research since the early 1980s [15], [16]. However, cryogenic

temperature also results in some challenges, such as a higher threshold voltage V_{th} of the transistor, carrier freeze-out in the substrate, and kink in the drain current [15], [16]. Ongoing scaling of the CMOS technology driven by Moore's law has reduced the minimum feature size to 5 nm. These reduced dimensions result in a higher mismatch between the electrical characteristics of the two identical transistors fabricated on the same chip. Mismatch in transistor characteristics and V_{th} increase at cryogenic temperature are major challenges faced by circuit designers and affect the circuit design significantly [17].

Authors in [18] and [19] showed that transistors fabricated using 160 and 40 nm bulk CMOS technologies result in an almost equal amount of performance improvement. However, the 40 nm technology with higher gate control and improved short-channel effects outperforms the older generation technologies at both 300 and 4 K. The authors in [20] and [21] showed that FinFETs from both 14- and 10-nm technologies can offer a significant power reduction while operating at cryogenic temperatures for a similar speed. A detailed study on the impact of ionized donor impurity on 10 nm technology node-based transistors suggests that direct transport through individual dopants results in a higher leakage current [22]. However, in our measurements, we have not observed the impact of resonant tunneling due to ionized dopants. Although the authors in [20] and [21] reported the FinFETs cryogenic characterization, these studies were limited up to 77 K. Han et al. [23] presented the 16-nm FinFET cryogenic characterization from 2.5 to 300 K. In our previous work [24], we have characterized 5-nm FinFET technology at 300 and 10 K.

A. TRANSISTOR COMPACT MODEL CALIBRATION AND VALIDATION FOR 5-NM FINFET TECHNOLOGY

As described in our previous work [24], we obtain the process-dependent model parameters, such as doping, oxide thickness, and gate material work function after setting the appropriate simulation environment. The subthreshold behavior of the transistors is affected by interface trap charges and source–drain coupling. Subthreshold characteristics of the measured FinFETs at room temperature (300 K) are captured by BSIM-CMG [25] model parameters for work function (PHIG), interface traps (CIT), and coupling capacitance between source/drain and channel (CDSC). The low-field mobility U_0 and field-dependent mobility degradation parameters (i.e., UA, UD, EU, and ETAMOB) are extracted from the transfer characteristics (drain–source current I_{DS} and gate voltage V_G) when the transistor operates at a low V_{DS} and moderate inversion (see Fig. 3). Subsequently, series resistances model parameters (RSW, RDW, RSWMIN, and RDWMIN) from the strong-inversion regime (higher V_G) are also obtained.

To capture the impact of drain-induced barrier lowering, we use the ETA0, PDIBL2, and CDSCD model parameters. Model parameter optimization is achieved by observing the I_{DS} – V_G (see Fig. 3) characteristics at lower and higher V_{DS} .

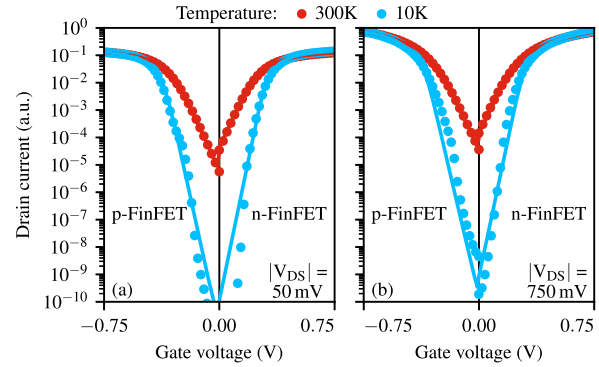


FIGURE 3. Transfer characteristics of p- and n-FinFET for 10 and 300 K in (a) linear ($V_{DS} = 50$ mV) and (b) saturation ($V_{DS} = 750$ mV). Symbols and lines show the data from measurement and calibrated model simulation, respectively [24].

As V_{DS} increases, carrier velocity begins to saturate. With a further increase in V_{DS} , transfer (I_{DS} – V_G) and output characteristic (I_{DS} – V_{DS}) show a slight increase in drain current. This is realized through the velocity saturation model parameters VSAT, VSAT1, MEXP, and KSATIV. At higher V_{DS} and V_G , the impact of velocity saturation and channel length modulation is captured by minimizing the error between measurement and simulation data of I_{DS} – V_G and I_{DS} – V_{DS} .

Metal–oxide semiconductor (MOS) transistor performance at cryogenic temperatures (10 K) improves with the reduction in carrier scattering [26]. Cryogenic operations result in a very small electron concentration in the conduction band at the same V_G because of Fermi–Dirac statistics (probability of finding an electron in conduction band reduces drastically with reduction in temperature), and there are simply not enough high energy electrons to climb the barrier, which reduces over-the-barrier transport. This decrease results in a huge improvement in SS and reduction of I_{OFF} . This causes a drastic change in the fundamental characteristics of semiconductor transistors at cryogenic temperatures, relative to 300 K. Some dominant effects at cryogenic temperatures are as follows: nonlinear temperature dependence in SS characteristics, increase in V_{th} , surface roughness scattering, Coulomb scattering, and nonlinear velocity saturation effect [26], [27], [28]. For example, our measurements show a 47% and 39% increase in V_{th} for n-channel FinFet and p-channel FinFET at cryogenic temperatures. Nevertheless, V_{th} remains at a low value due to the ultra-low- V_{th} transistors. To account for these effects in SPICE simulations of FinFETs, we use the model equations presented in [26] along with the industry-standard BSIM-CMG compact model [25]. It describes the behavior of a transistor through the underlying physics-based models that take carefully into account many necessary aspects, such as temperature dependency, short-channel effects, and quantum confinement, among others. This allows for an accurate and detailed modeling with which experimental measurements can be reproduced.

As the existing BSIM-CMG model is based on Maxwell–Boltzmann statistics, we use it along with the modifications

presented in [26]. For electron density calculation, this model captures the impact of Fermi–Dirac statistics from 300 K to cryogenic temperatures. The effective density of states, surface potential, and charges are highly temperature dependent, and thus, we first obtain an effective temperature at cryogenic temperatures [26]. The nonlinearity in the SS is caused by the band-tail effect [27], [28]. Recently, source-to-drain tunneling has been proposed as a possible mechanism for the SS saturation at low temperatures and could be the major cause of higher leakage current [29]. In this work, the impact of the band-tail effect and traps on SS and V_{th} is captured using T0, D0, KT11, KT12, and TVTH model parameters [26].

Peak mobility is enhanced as the temperature-dependent lattice vibration decreases at cryogenic temperatures and the thermal velocity of the charge carriers decreases. The effective mobility of charge carriers with lower thermal velocities decreases at higher vertical fields due to increased surface roughness scattering. Through the optimization of temperature coefficients for Coulomb scattering (UD1 and UD2), and the temperature coefficients for phonon/surface roughness scattering UA1, UA2, and EU1, the impact of Coulomb and surface roughness scattering is accounted for in the mobility model. Additional model parameters are used to obtain the nonlinear temperature dependency of the velocity saturation and pinch-off voltage. Those parameters include the effective drain-to-source voltage (V_{dseff}) smoothing TMEXP, its temperature coefficients TMEXP1 and TMEXP2, the temperature coefficients for the saturation velocity (AT, AT1, and AT2), and KSATIVT, KSATIVT1, and KSATIVT2 to model the temperature dependence of the channel pinch-off effect. The model is validated against experimental data, as shown in Fig. 3. Intrinsic randomness of the measurements is observed at lower V_G and is the likely cause of discrepancies between the simulated and measured results at lower current.

IV. CRYOGENIC-AWARE STANDARD CELL LIBRARIES

Standard cell libraries bridge the gap between the modeling of individual transistors and the design of complex digital circuits. They are indispensable for the indispensable for the electronic design automation (EDA) tool flows to perform logic synthesis. In the standard cell characterization process, our calibrated transistor model is placed into a wide range of standard cells and simulated using a commercial SPICE simulator. The resulting figures of merit are then collected to build standard cell libraries that are fully compatible with the existing commercial EDA tool flows to seamlessly perform logic synthesis, timing signoff, and power signoff.

A. STANDARD CELL LIBRARY CHARACTERIZATION PROCESS

Fig. 4 depicts the design flow from transistor model extraction up to the generation of standard cell libraries. The inputs to the characterization flow are highlighted in blue color. Both, the extended BSIM-CMG compact model and the calibration data obtained in Section III, are inputs to the characterization flow and are vital for the SPICE

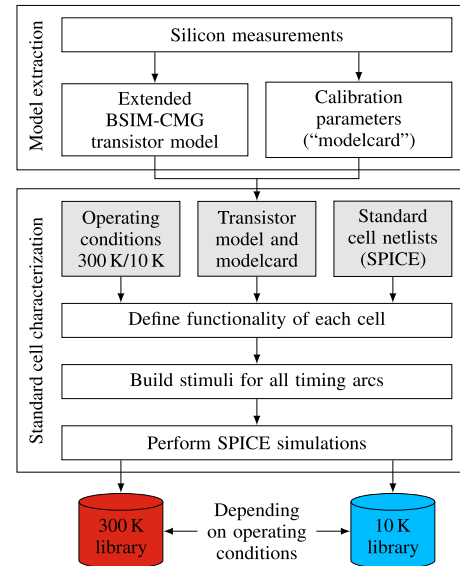


FIGURE 4. Overview of the design flow from transistor modeling to standard cell libraries. Inputs to the characterization flow are shaded in gray, and outputs are highlighted in red (300 K) and blue (10 K) color. The temperature can be adjusted as part of the operating conditions, allowing us to build libraries for different temperature scenarios.

simulations. During standard cell library characterization, the only parameter changed in the compact model is the number of fins, which acts as a current multiplier [25]. Self-heating is considered in the compact model but has only a negligible impact on the standard cells at 10 K. It is noteworthy that self-heating effects even decrease in advanced bulk FinFET devices at cryogenic temperatures [21]. In contrast, they increase in SOI technologies [30]. Besides the transistor model, SPICE netlists for a wide range of combinational and sequential standard cells are provided to the characterization flow. In this work, we obtain 200 different standard cells from the open-source ASAP7 PDK [31], including parasitic resistances and capacitances. The cells are designed for a 7-nm technology node and, thus, geometrically very close to our 5-nm transistor model.

When integrated into a larger circuit, a single cell can exhibit very different behavior based on its experienced timing arcs, input signal slews, and output load capacitances. To obtain an adequate model of each standard cell for a wide range of conditions, each cell is characterized under 7×7 slew-load combinations for all possible timing arcs the cell can experience. The required preprocessing steps, including the generation of stimuli and SPICE decks, are carried out by the commercial characterization tool flow *Synopsys PrimeLib*. The subsequent SPICE simulations are performed by *Synopsys PrimeSim*, collecting over 1×10^6 measurements that are eventually gathered in the resulting standard cell libraries. These measurements include delay and transition times for each timing arc, pin capacitances, switching energy, and leakage power. The resulting standard cell libraries are generated in the industry-standard *Liberty* format making them usable in most established EDA tools.

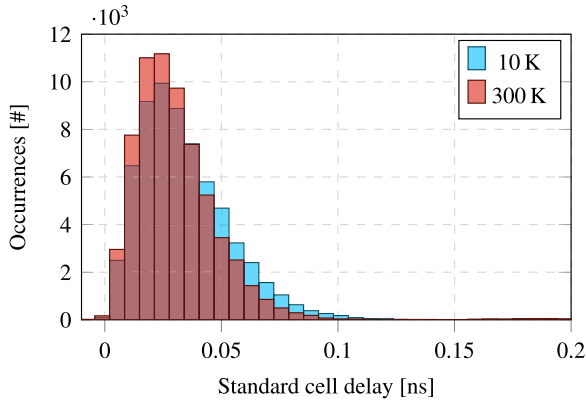


FIGURE 5. Histogram shows the delays across all 200 cells in the standard cell library. The large overlap of the histograms for 300 and 10 K demonstrates that the delay is only slightly increased at cryogenic temperatures. A benefit of the low temperatures is the significant reduction of leakage power, rendering it almost negligible (not shown).

B. IMPACT OF CRYOGENIC TEMPERATURES ON THE STANDARD CELL CHARACTERISTICS

With a transistor model calibrated for a wide temperature range, the corresponding standard cell libraries can be generated by adjusting the temperature in the operating conditions. In this work, we generate standard cell libraries for operation at room temperature (300 K) and cryogenic temperature (10 K). With the standard cell libraries at hand, the impact of cryogenic temperatures can be explored at the cell and circuit level.

Fig. 5 shows a histogram of all delays occurring in the standard cell libraries. The histograms span data from all cells and conditions stored in the library, giving a holistic picture of the technology under 300 and 10 K in red and blue bars, respectively. Although slight differences are observable, the histograms overlap to a large degree, indicating only minor differences in delay when operated at cryogenic temperature. In addition, average dynamic power is reduced slightly for some cells and increased for others. Most importantly, *leakage power* is reduced dramatically at 10 K. This behavior can be explained by the electrical characteristics of the extracted transistor model. While temperature merely shows an impact on the I_{ON} of the transistor, I_{OFF} is reduced by multiple orders of magnitude when operating at cryogenic temperatures, as shown in Fig. 3.

V. CRYOGENIC SOC AND APPLICATIONS

To evaluate the plausibility of a full SoC, the whole SoC is synthesized and placed with the room temperature library. Then, the cryogenic-aware standard cell libraries are employed to analyze power and timing at 10 K.

A. SOC DESIGN FLOW

The employed SoC is a fully functional system, including a RISC-V CPU core, caches, and periphery like a memory controller. A single five-stage in-order Rocket CPU [32] is combined with a split L1 cache for data and instructions, each with 16 and a shared L2 cache of 512. The hardware description language (HDL) code is created with the assistance

of the Chipyard framework [33]. Then, a commercial synthesis tool is employed to create a gate-level netlist. At this stage, the previously described 300 K standard cell library is employed. The gate-level netlist is then fed to a commercial place and route tool in combination with the 300 K standard cell libraries.

Static random-access memory (SRAM) arrays, the core building block of L1 and L2 caches among others, are provided through the ASAP7 PDK [31] as internet protocol (IP) cores. However, these IP cores only include the physical size and timing but not their power consumption. We add the missing power values based on our previous work [24]. In [24], we have modeled SRAM cells and peripheral circuitry, such as sense amplifiers and write drivers, based on the same calibrated BSIM-CMG transistor compact model at 300 and 10 K. This enables a complete power estimation for all the SRAM utilized by the SoC. Read and write accesses as well as hold and leakage are included. Quantum computing specific peripheries, such as signal generators, are not included because of their specificity to the physical implementation of the quantum system. The focus of this work is on the impact of a full SoC on the cooling and time budget.

B. QUANTUM MEASUREMENT CLASSIFICATION

To evaluate the dynamic power consumption accurately, two classification algorithms are implemented in C-Code and simulated. While more complex algorithms promise a higher accuracy [34], they are also more computationally demanding. To estimate a baseline, two simpler classification algorithms are selected in this work. First, k-nearest neighbors algorithm (KNN) is a nonparametric clustering method [35]. The calibration phase is performed offline and returns the center points for each qubit, as shown in Fig. 2 and described in Section II. After qubit measurement, the distances of the new data point to its qubit’s centers in the IQ plane are calculated. The label (0 or 1) of the nearest center is returned as the result. In this work, the Euclidean distance d is computed between a center (x_C, y_C) and the measurement (x_M, y_M) with (2).

$$d = \sqrt{(x_M - x_C)^2 + (y_M - y_C)^2} \quad (2)$$

After the distances to the two centers for 0 and 1 are computed, they are compared and the closest selected as the classification result. The computation of (2) can be optimized because the square root is a linear operation. In other words, the radicand will be larger for a longer distance and, thus, comparing the radicands is sufficient. Hence, the computationally expensive square root operation is unnecessary and removed.

Hyperdimensional computing (HDC) is a machine learning method based on large vectors: hypervectors [36]. The components of the vectors can be simple bits, making its implementation very light weight, e.g., the bind operation \oplus is a binary XOR. A point $P = (x_P, y_P)$ is *encoded* into a hypervector with (3) employing the item hypervectors \vec{x}_P and

TABLE 1. Full SoC is Synthesized for 300 K and 0 ns Clock Period

| Temperature | Critical path delay | Clock frequency |
|-------------|---------------------|-----------------|
| 300 K | 1.04 ns | 960 MHz |
| 10 K | 1.09 ns | 917 MHz |

The timing analysis is conducted with Synopsys PrimeTime for 300 and 10 K. The difference is less than 10% because ION is only slightly affected and, thus, the delay of the standard cells is similar, as discussed in Section IV-B.

\vec{y}_P for its quantized x and y values.

$$\vec{P} = \vec{x}_P \oplus \vec{y}_P \tag{3}$$

Such item hypervectors are constant and generated once during the program compilation. A size of 128 bits per hypervector is sufficient, and a total of 32 are created to cover the x and y value range. Because each bit in a hypervector is independent, each 128-bit HDC operation can be split into two 64-bit instructions for the 64-bit RISC-V SoC.

For each qubit, the center points from the calibration phase are encoded using (3) into \vec{C}_0 and \vec{C}_1 . After the measurement phase, each measurement is first quantized and encoded into \vec{M} . Then, the Hamming distances to its \vec{C}_0 and \vec{C}_1 are computed, and the closest selected as the result. The Hamming distance is the popcount of $\vec{C} \oplus \vec{M}$. This computation can be partially simplified from two to one XOR operations, as shown in the following:

$$\begin{aligned} d_0 &= \text{popcount}(\vec{C}_0 \oplus \vec{M}) \\ d_0 &= \text{popcount}(\vec{C}_0 \oplus \vec{x}_M \oplus \vec{y}_M) \\ d_0 &= \text{popcount}(\vec{x}_{C_0 \oplus M} \oplus \vec{y}_M). \end{aligned} \tag{4}$$

Since these hypervectors are themselves the products of XOR operations, their order can be rearranged. Instead of computing the XOR of \vec{x}_M and \vec{C} every time, $\vec{x}_{C_0 \oplus M}$ is precomputed and replaces the item hypervectors for the x component. A drawback is the doubling of the memory consumption of the executable to store $\vec{C}_0 \oplus \vec{x}_M$ and $\vec{C}_1 \oplus \vec{x}_M$. Because of the few item hypervectors and small dimension of 128 bits, the memory footprint is increased by only 256 bytes.

VI. EVALUATION AND COMPARISONS

A. PROCESSOR TIMING ANALYSIS

As a baseline, the physical design of the SoC is synthesized with the 300 K library. The clock period is set to 0 ns to force the EDA tools to optimize as much as possible. The reported critical path delay (worst-case slack) from the tools determines the possible operating frequency. Guardbands, e.g., for process variation, are assumed to be equal at both temperature corners, effectively nullifying themselves in a comparison. Hence, they are not considered in this work.

The gate-level netlist representing the physical design of the SoC is provided to a commercial static timing analysis tool. The timing analysis is repeated with both libraries and the results are reported in Table 1. At 300 K, the critical path has a length of 1.04 ns, which corresponds to a clock frequency of 960 MHz. As shown in Fig. 5, some standard

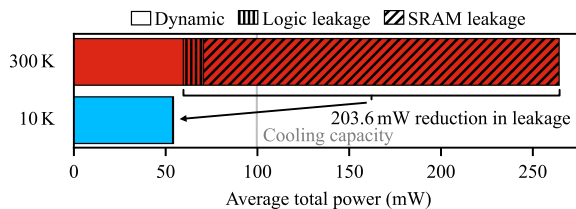


FIGURE 6. Average power consumption of the KNN for quantum measurement classification. The dynamic power at cryogenic temperatures is reduced by 10% from 63.5 to 57.4 mW. However, the major contributor is the leakage from SRAM, which is suppressed and reduced to only 0.48 mW at 10 K. This large reduction makes the SoC feasible given a cooling capacity of 100 mW.

cells are slower at 10 K, which leads to an increase in the critical path delay to 1.09 ns or 917 MHz. This represents a 4.6% slowdown. Fig. 3 shows that the I_{ON} of n-FinFET and p-FinFET transistors at 10 K is similar to 300 K. Therefore, the switching delay of the transistors is similar, thus the propagation delay of the cells and, thus, the hold times of the circuit are not impacted.

B. PROCESSOR POWER ANALYSIS

Power analysis is often done with statistical switching activities, e.g., 20% of all cells are activated per cycle. However, such statistical switching activities do not reflect the actual power consumption because, for simpler tasks, such as classifying a measurement, only parts of the SoC have to be engaged. Therefore, the two classification algorithms for quantum measurements are simulated with the gate-level netlist of the physical design. The actual switching activity numbers are extracted from these simulations and, in combination with the physical design, processed by *Cadence Voltus* to calculate the average power consumption of the whole SoC.

Since the two algorithms for classification represent less demanding workloads, the Dhrystone benchmark [37] is also simulated to report a general average. The dynamic power consumption at both temperatures is similar, as shown in Fig. 6. At 300 K, standard cells for logic contribute about 11 mW to the leakage power, whereas the 581 total on-chip SRAM contribute 193 mW. Operating at nominal supply voltage combined with ultralow- V_{th} transistors results in such a high SRAM leakage, which is in line with other works [38]. In addition, the short channel effect and quantum tunneling are other key causes. Hence, the SoC would be infeasible for a cryogenic system given the limited cooling capacity of 100 mW [5]. However, the significant reduction of the leakage current of transistors when operated at cryogenic temperatures is reflected at circuit and SoC level. At 10 K, the leakage from logic and SRAM can almost be neglected with 0.48 mW, a reduction by 99.76%. Consequently, the SoC becomes feasible for a cryogenic system and demonstrating that on-chip memories can be enlarged for systems at 10 K.

TABLE 2. Average Clock Cycles to Classify One Measurement

| Method | 20 qubits | 400 qubits |
|--------|-----------|------------|
| KNN | 41.5 | 72.8 |
| HDC | 184.8 | 242.4 |

More qubits result in more cache misses increasing the number of clock cycles.

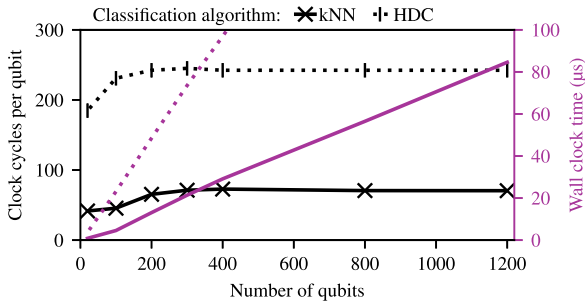


FIGURE 7. With an increase in the number of qubits, the time to classify all of them through a KNN becomes more important. The maximum quantum computation duration, i.e., the decoherence time, for the investigated IBM Falcon quantum processor is about 110 μ s. However, the SoC (clocked at 1000 MHz) would be tasked with other workloads as well, rendering it a bottleneck for systems with hundreds or thousands of qubits. The popcount operation for HDC requires too many cycles to be competitive.

C. QUANTUM MEASUREMENT CLASSIFICATION: EXECUTION TIME ANALYSIS

The execution times of the two classification algorithms for quantum measurements are evaluated. Table 2 shows the average number of clock cycles needed for a classification of a single measurement. Although HDC comprises simpler binary and logical instructions, it is 3.3 \times slower than the distance computations with floating point calculations. The main contributor is the lack of a popcount instruction in the RISC-V instruction set architecture, which is essential for the HDC similarity computation. Hardware support would reduce the computation time significantly.

While the time for a single KNN classification is small, the challenge arises from scaling up the quantum system to thousands of qubits that have to be classified within a given time frame. We assume here that this time frame is set by the maximum duration of a continuous quantum computation. Therefore, to not stall the quantum computer and let the classification become a bottleneck, the data processing has to be faster than the given time frame, as shown in Fig. 2. This time frame is determined by the decoherence time of the quantum computer, which specifies the maximum time in which a quantum state can retain its properties. Our experiments on the IBM Falcon quantum processor report this time is around 110 μ s. However, typically users strive toward shorter quantum computation durations to minimize the error from the exponential decay due to decoherence. Hence, the numbers given in Fig. 7 portray a best-case scenario in which the full decoherence time is available for classification and no other tasks have to be performed by the processing system. Such other tasks include loading the next quantum computation,

providing the confidence intervals for the different solutions, or executing quantum error correction protocols, among others. Thus, the SoC has to perform other tasks and cannot be fully occupied with classifying measurement results.

VII. PERSPECTIVE AND DISCUSSION

Currently, quantum computing is only possible at cryogenic temperatures. Cooling any computing system to such low temperatures is challenging and heat dissipation, i.e., power consumption, must be limited to not overwhelm the cooling. This work shows that it is easily possible to deploy an off-the-shelf system developed for room temperatures at cryogenic temperatures. The timing is impacted only marginally and is within expected guard bands. Power consumption is even reduced, especially through the significant reduction in leakage power, demonstrating the plausibility of a cryogenic SoC for quantum measurement classification. Further power reduction could be achieved by work function engineering at the transistor level, supply voltage reduction, alternative SRAM designs, or clock and power gating at the circuit level. However, scaling to large quantum systems with thousands of qubits is still challenging for off-the-shelf SoCs.

The evaluated RISC-V SoC becomes a bottleneck for classifying the quantum measurements for about 1500 qubits while consuming half of the available cooling budget. Other hardware components, such as signal generators and analog-to-digital converters, have to be cooled as well, opening a new field of potential co-optimizations. On the one hand, exceeding the cooling budget and increasing the temperature of the quantum system increases the error rate. On the other hand, faster processing enables more repetitions of the same quantum computations to overcome erroneous computations or more sophisticated error correction algorithms. Furthermore, heat transfer is comparatively slow, creating the potential for short but high-power processing bursts followed by a low-power idle phase without impacting the qubits. Such tradeoffs and power management strategies can be explored and experimentally evaluated with flexible, software-controlled SoCs more efficiently and faster than with fixed hardware implementations. This work shows that off-the-shelf SoCs are plausible at cryogenic temperatures and that there is no need for dedicated chips only to explore such tradeoffs, error correction algorithms, and power management strategies.

Integrating a cryogenic SoC into the quantum computer setup offers indeed a wide range of possibilities and the flexibility to explore several improvements that can have a huge impact on the throughput, result quality, and duration of quantum computations. For hybrid quantum-classical algorithms, such as the quantum approximate optimization algorithm or the variational quantum eigensolver, an integrated SoC decreases the data movement and would, thus, allow for more optimization steps given a specified runtime budget leading to higher quality results. Furthermore, the time required for the calibration of the quantum computer would decrease and, thus, increase the throughput further.

Potentially, it would also allow to include more data, e.g., to classify qubit measurements with a higher precision. In the near-term, an integrated cryogenic SoC would also be able to reduce the runtime requirements of dynamic circuits [7] and also apply error mitigation algorithms on-the-fly, further improving the throughput. In the future, a range of quantum error correction protocols could be evaluated more thoroughly, reducing the time required to achieve fully error-corrected quantum computers.

This work shows that processing a large number of qubits is not feasible with a regular processor but will require hardware support. Dedicated SoCs, such as [6], already include dedicated hardware for quantum measurement classification. Nevertheless, this dedicated hardware is fixed after the design and cannot be improved or replaced without a redesign of the SoC. Hence, an SRAM-based field-programmable gate array (FPGA) fabric could be an interesting addition to SoC. The SRAM's leakage power is very low at 10 K, and FPGAs offer a large degree of flexibility yet consume comparatively little power. Similar to the exploration of different power management strategies and quantum error correction methods outlined above, the FPGA fabric can be reconfigured to select between a high-power low-latency or a low-power high-latency classification algorithm, depending on the complexity and error-robustness of the intended quantum computation.

VIII. CONCLUSION

In this work, we explored for the first time how SoC implemented with a cutting-edge 5 nm technology at room temperature would behave at cryogenic temperatures. Such a general-purpose system can not only classify quantum measurements of the qubits but support other tasks, such as quantum error correction as well. Further, by analyzing an off-the-shelf design aimed at room temperature operations, already existing hardware can be deployed quickly. The SoC's power consumption is reduced and can fit within the 100 mW cooling capacity. We have shown that the timing of the system is comparable to room temperature and the significant reduction of the leakage power enables SoCs with large on-chip memories.

ACKNOWLEDGMENT

The authors would like to thank the Central Research Facility IIT Delhi for facilitating the cryogenic characterization of FinFETs and Munazza Sayed for proofreading the manuscript. The authors would also like to thank IBM for the use of IBM Quantum Services for this work.

REFERENCES

- [1] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, "Elucidating reaction mechanisms on quantum computers," *Proc. Nat. Acad. Sci.*, vol. 114, no. 29, pp. 7555–7560, 2017, doi: [10.1073/pnas.1619152114](https://doi.org/10.1073/pnas.1619152114).
- [2] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in *Proc. 35th Annu. Symp. Found. Comput. Sci.*, 1994, pp. 124–134, doi: [10.1109/SFCS.1994.365700](https://doi.org/10.1109/SFCS.1994.365700).
- [3] S. J. Pauka et al., "A cryogenic CMOS chip for generating control signals for multiple qubits," *Nature Electron.*, vol. 4, no. 1, pp. 64–70, Jan. 2021, doi: [10.1038/s41928-020-00528-y](https://doi.org/10.1038/s41928-020-00528-y).
- [4] F. Arute et al., "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, Oct. 2019, doi: [10.1038/s41586-019-1666-5](https://doi.org/10.1038/s41586-019-1666-5).
- [5] F. Sebastiano et al., "Cryogenic CMOS interfaces for quantum devices," in *Proc. 7th IEEE Int. Workshop Adv. Sensors Interfaces*, 2017, pp. 59–62, doi: [10.1109/TWASI.2017.797421](https://doi.org/10.1109/TWASI.2017.797421).
- [6] J. Park et al., "A fully integrated cryo-CMOS SoC for state manipulation, readout, and high-speed gate pulsing of spin qubits," *IEEE J. Solid-State Circuits*, vol. 56, no. 11, pp. 3289–3306, Nov. 2021, doi: [10.1109/JSSC.2021.3115988](https://doi.org/10.1109/JSSC.2021.3115988).
- [7] S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, and Z. Nazario, "The future of quantum computing with superconducting qubits," *J. Appl. Phys.*, vol. 132, no. 16, Art. no. 160902, 2022, doi: [10.1063/5.0082975](https://doi.org/10.1063/5.0082975).
- [8] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," 2014, *arXiv:1411.4028*, doi: [10.48550/arXiv.1411.4028](https://doi.org/10.48550/arXiv.1411.4028).
- [9] A. Peruzzo et al., "A variational eigenvalue solver on a photonic quantum processor," *Nature Commun.*, vol. 5, no. 1, 2014, Art. no. 4213, doi: [10.1038/ncomms5213](https://doi.org/10.1038/ncomms5213).
- [10] M. A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*. Cambridge, MA, USA: Cambridge Univ. Press 2002, doi: [10.1017/CBO9780511976667](https://doi.org/10.1017/CBO9780511976667).
- [11] C. Tornow, N. Kanazawa, W. E. Shanks, and D. J. Egger, "Minimum quantum run-time characterization and calibration via restless measurements with dynamic repetition rates," *Phys. Rev. Appl.*, vol. 17, 2022, Jun. 2022, Art. no. 064061, doi: [10.1103/PhysRevApplied.17.064061](https://doi.org/10.1103/PhysRevApplied.17.064061).
- [12] T. Alexander et al., "Qiskit pulse: Programming quantum computers through the cloud with pulses," *Quantum Sci. Technol.*, vol. 5, no. 4, 2020, Art. no. 044006, doi: [10.1088/2058-9565/aba404](https://doi.org/10.1088/2058-9565/aba404).
- [13] M. S. Anis et al., "Qiskit: An open-source framework for quantum computing," 2021. [Online]. Available: <https://zenodo.org/record/2562111>
- [14] "IBM Quantum Roadmap," Accessed: Oct. 1, 2023. [Online]. Available: <https://www.ibm.com/quantum/roadmap>
- [15] A. Kamgar and R. Johnston, "Delay times in Si MOSFETS in the 4.2–400 K temperature range," *Solid-State Electron.*, vol. 26, no. 4, pp. 291–294, 1983, doi: [10.1016/0038-1101\(83\)90125-9](https://doi.org/10.1016/0038-1101(83)90125-9).
- [16] H. Hanamura, M. Aoki, T. Masuhara, O. Minato, Y. Sakai, and T. Hayashida, "Operation of bulk CMOS devices at very low temperatures," *IEEE J. Solid-State Circuits*, vol. 21, no. 3, pp. 484–490, Jun. 1986, doi: [10.1109/JSSC.1986.1052555](https://doi.org/10.1109/JSSC.1986.1052555).
- [17] P. A. t Hart, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Subthreshold mismatch in nanometer CMOS at cryogenic temperatures," in *Proc. 49th Eur. Solid-State Device Res. Conf.*, 2019, pp. 98–101, doi: [10.1109/ESSDERC.2019.8901745](https://doi.org/10.1109/ESSDERC.2019.8901745).
- [18] R. M. Incandela, L. Song, H. Homulle, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and compact modeling of nanometer CMOS transistors at deep-cryogenic temperatures," *IEEE J. Electron Devices Soc.*, vol. 6, pp. 996–1006, 2018, doi: [10.1109/JEDS.2018.2821763](https://doi.org/10.1109/JEDS.2018.2821763).
- [19] H. Homulle, "Cryogenic electronics for the read-out of quantum processors," Ph.D. dissertation, Delft Univ. Technol., 2628 CD Delft, Netherlands, 2019.
- [20] A. Chabane et al., "Cryogenic characterization and modeling of 14 nm bulk FinFET technology," in *Proc. IEEE 47th Eur. Solid State Circuits Conf.*, 2021, pp. 67–70, doi: [10.1109/ESSCIRC53450.2021.9567802](https://doi.org/10.1109/ESSCIRC53450.2021.9567802).
- [21] H. L. Chiang et al., "Cold CMOS as a power-performance-reliability booster for advanced FinFETs," in *Proc. IEEE Symp. VLSI Technol.*, 2020, pp. 1–2, doi: [10.1109/VLSITechnology18217.2020.9265065](https://doi.org/10.1109/VLSITechnology18217.2020.9265065).
- [22] M. Pierre, R. Wacquez, X. Jehl, M. Sanquer, M. Vinet, and O. Cueto, "Single-donor ionization energies in a nanoscale CMOS channel," *Nature Nanotechnol.*, vol. 5, no. 2, pp. 133–137, 2010, doi: [10.1038/nnano.2009.373](https://doi.org/10.1038/nnano.2009.373).
- [23] H.-C. Han, F. Jazaeri, A. D'Amico, A. Baschiroto, E. Charbon, and C. Enz, "Cryogenic characterization of 16 nm FinFET technology for quantum computing," in *Proc. IEEE 47th Eur. Solid State Circuits Conf.*, 2021, pp. 71–74, doi: [10.1109/ESSCIRC53450.2021.9567747](https://doi.org/10.1109/ESSCIRC53450.2021.9567747).

- [24] S. S. Parihar, V. M. Van Santen, S. Thomann, G. Pahwa, Y. S. Chauhan, and H. Amrouch, "Cryogenic CMOS for quantum processing: 5-nm finfet-based sram arrays at 10 k," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 70, no. 8, pp. 3089–3102, Aug. 2023, doi: [10.1109/TCSL.2023.3278351](https://doi.org/10.1109/TCSL.2023.3278351).
- [25] BSIM-CMG Technical Manual. [Online]. Available: <http://bsim.berkeley.edu/models/bsimcmg/>
- [26] G. Pahwa, P. Kushwaha, A. Dasgupta, S. Salahuddin, and C. Hu, "Compact modeling of temperature effects in FDSOI and FinFET devices down to cryogenic temperatures," *IEEE Trans. Electron Devices*, vol. 68, no. 9, pp. 4223–4230, 2021, doi: [10.1109/LED.2021.3097971](https://doi.org/10.1109/LED.2021.3097971).
- [27] H. Bohuslavskiy et al., "Cryogenic subthreshold swing saturation in FD-SOI MOSFETs described with band broadening," *IEEE Electron Device Lett.*, vol. 40, no. 5, pp. 784–787, May 2019, doi: [10.1109/LED.2019.2903111](https://doi.org/10.1109/LED.2019.2903111).
- [28] A. Beckers, F. Jazaeri, and C. Enz, "Theoretical limit of low temperature subthreshold swing in field-effect transistors," *IEEE Electron Device Lett.*, vol. 41, no. 2, pp. 276–279, Feb. 2020, doi: [10.1109/LED.2019.2963379](https://doi.org/10.1109/LED.2019.2963379).
- [29] H.-C. Han, H.-L. Chiang, I. P. Radu, and C. Enz, "Analytical modeling of source-to-drain tunneling current down to cryogenic temperatures," *IEEE Electron Device Lett.*, vol. 44, no. 5, pp. 717–720, May 2023, doi: [10.1109/LED.2023.3254592](https://doi.org/10.1109/LED.2023.3254592).
- [30] K. Triantopoulos et al., "Self-heating effect in FDSOI transistors down to cryogenic operation at 4.2 k," *IEEE Trans. Electron Devices*, vol. 66, no. 8, pp. 3498–3505, Aug. 2019, doi: [10.1109/TED.2019.2919924](https://doi.org/10.1109/TED.2019.2919924).
- [31] V. Vashishtha, M. Vangala, and L. T. Clark, "ASAP7 predictive design kit development and cell design technology co-optimization: Invited paper," in *Proc. IEEE/ACM Int. Conf. Comput.-Aided Des.*, 2017, pp. 992–998, doi: [10.1109/ICCAD.2017.8203889](https://doi.org/10.1109/ICCAD.2017.8203889).
- [32] K. Asanovic et al., "The rocket chip generator," Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, Tech. Rep. UCB/Eecs-2016-17, vol. 4, 2016. [Online]. Available: <https://www2.eecs.berkeley.edu/Pubs/TechRpts/2016/Eecs-2016-17.html>
- [33] A. Amid et al., "Chipyard: Integrated design, simulation, and implementation framework for custom SoCs," *IEEE Micro*, vol. 40, no. 4, pp. 10–21, Jul./Aug. 2020, doi: [10.1109/MM.2020.2996616](https://doi.org/10.1109/MM.2020.2996616).
- [34] E. Magesan, J. M. Gambetta, A. D. Córcoles, and J. M. Chow, "Machine learning for discriminating quantum measurement trajectories and improving readout," *Phys. Rev. Lett.*, vol. 114, Art. no. 200501, May 2015, doi: [10.1103/PhysRevLett.114.200501](https://doi.org/10.1103/PhysRevLett.114.200501).
- [35] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: Consistency properties," *Int. Stat. Rev./Revue Int. Stat.*, vol. 57, no. 3, pp. 238–247, 1989, doi: [10.2307/1403797](https://doi.org/10.2307/1403797).
- [36] P. Kanerva, "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors," *Cogn. Comput.*, vol. 1, no. 2, pp. 139–159, 2009, doi: [10.1007/s12559-009-9009-8](https://doi.org/10.1007/s12559-009-9009-8).
- [37] R. P. Weicker, "Dhrystone: A synthetic systems programming benchmark," *Commun. ACM*, vol. 27, no. 10, pp. 1013–1030, 1984, doi: [10.1145/358274.358283](https://doi.org/10.1145/358274.358283).
- [38] A. Shafaei, S. Chen, Y. Wang, and M. Pedram, "A cross-layer framework for designing and optimizing deeply-scaled FinFET-based sram cells under process variations," in *Proc. 20th Asia South Pacific Des. Automat. Conf.*, 2015, pp. 75–80, doi: [10.1109/ASPDAC.2015.7058984](https://doi.org/10.1109/ASPDAC.2015.7058984).



Paul R. Genssler (Member, IEEE) received the Dipl.Inf. (M.Sc.) degree in computer science from TU Dresden, Dresden, Germany, in 2017. He is currently working toward the Ph.D. degree in computer engineering with Semiconductor Test and Reliability (STAR) Chair, Computer Science and Electrical Engineering Faculty, University of Stuttgart, Stuttgart, Germany.

In 2018, he started his Ph.D. research with Chair for Embedded Systems (CES), Karlsruhe Institute of Technology, Karlsruhe, Germany. His

research interests include emerging technologies, system architecture, and emerging brain-inspired methods for IC test and beyond. He has served as a Reviewer for the IEEE Internet of Things Journal.



Florian Klemme (Member, IEEE) received the B.Sc. degree in system integration from the University of Applied Sciences Bremerhaven, Bremerhaven, Germany, in 2014, and the M.Sc. degree in computer science from the Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2018. He is working toward the Ph.D. degree in computer engineering with the Chair of Semiconductor Test and Reliability (STAR), University of Stuttgart, Stuttgart, Germany.

He is currently a Doctoral Researcher with the Chair STAR, University of Stuttgart. His research interests include cell library characterization and machine learning techniques in electronic design automation and computer-aided design, specifically toward the reliability of transistors, and integrated circuits in advanced technology nodes.



Shivendra Singh Parihar (Member, IEEE) is working toward the Ph.D. degree in microelectronics and very large scale integration (VLSI) with the Indian Institute of Technology Kanpur, Kanpur, India.

He is currently with the Chair of Semiconductor Test and Reliability (STAR), University of Stuttgart, Stuttgart, Germany as a Research Scholar. His research focuses on the characterization and compact modeling of advanced CMOS technologies for circuit design.



Sebastian Brandhofer (Graduate Student Member, IEEE) received the M.Sc. degree in computer science from the University of Stuttgart, Stuttgart, Germany, in 2017.

In 2018, he started his Ph.D. research with the Chair of Hardware Oriented Computer Science (HOCOS), University of Stuttgart, in 2018. His research interests include quantum computing, quantum circuit optimization, and computer-aided design.

Mr. Brandhofer was the recipient of the IBM Quantum Open Science Prize in 2021.



Girish Pahwa (Member, IEEE) received the M.Tech. and Ph.D. degrees in electrical engineering from the Indian Institute of Technology (IIT) Kanpur, Kanpur, India, in 2020.

He is currently an Assistant Professional Researcher with the Department of Electrical Engineering and Computer Sciences (EECS), University of California (UC) Berkeley, Berkeley, CA, USA. He is also the Executive Director of the Berkeley Device Modeling Center (BDMC), UC Berkeley. From 2020 to 2021, he was a Postdoctoral Researcher with the Department of EECS, UC Berkeley.

He is a Co-Developer of industry-standard BSIM-CMG, BSIM-IMG, BSIM-BULK, BSIM-SOI, and BSIM4 models. He has also developed the first industry-standard models for cryogenic FinFET and FDSOI FETs for quantum computing and cold electronics applications. He has authored or coauthored more than 50 technical publications in prominent journals and conferences in the field of device modeling and simulation. His research interests include the modeling and simulation of nanoscale devices and device circuit co-design, and optimization of emerging transistor technologies with a special emphasis on ferroelectric devices.

Dr. Pahwa was the recipient of the IEEE Electron Devices Society Early Career Award in 2022, Outstanding Ph.D. Thesis Award from IIT Kanpur in 2020, and Best Paper Award at the IEEE International Conference on Emerging Electronics (ICEE), Mumbai, India, in 2016. He is a Reviewer for several reputed journals.



Ilia Polian (Senior Member, IEEE) received the Diploma M.Sc. and Ph.D. degrees in computer science from the University of Freiburg, Germany, in 1999 and 2003, respectively.

He is currently a Full Professor and the Director of the Institute for Computer Architecture and Computer Engineering, University of Stuttgart, Stuttgart, Germany. He has coauthored more than 200 scientific publications. His scientific research interests include hardware-oriented security, emerging architectures, test methods,

and quantum computing.

Dr. Polian was the recipient of two Best Paper Awards.



Yogesh Singh Chauhan (Fellow, IEEE) received the Ph.D. degree in microelectronics engineering from EPFL Switzerland, in 2007.

He is a Class of 1984 Chair Professor with the Indian Institute of Technology Kanpur, Kanpur, India. He was with Semiconductor Research and Development Center, IBM, Bangalore, India, during 2007–2010, Tokyo Institute of Technology, Tokyo, Japan, in 2010, University of California Berkeley, Berkeley, CA, USA, during 2010–2012, and ST Microelectronics, Geneva, Switzerland,

during 2003–2004. He is the Developer of several industry standard models: ASM-GaN-HEMT model, BSIM-BULK (formerly BSIM6), BSIM-CMG, BSIM-IMG, BSIM4, and BSIM-SOI models. His research group is involved in developing compact models for GaN transistors, Fin-FET, nanosheet/gate-all-around FETs, FDSOI transistors, negative capacitance FETs, and 2-D FETs. He has authored or coauthored more than 300 papers in international journals and conferences. His research interests include characterization, modeling, and simulation of semiconductor devices.

Prof. Chauhan is the Fellow of National Academy of Engineering. He is the Editor of IEEE TRANSACTIONS ON ELECTRON DEVICES and Distinguished Lecturer of the IEEE Electron Devices Society. He is the Chair of IEEE-EDS Compact Modeling Committee. He is the Founding Chairperson of IEEE Electron Devices Society U.P. chapter and Chairman-Elect of IEEE U.P. section. He was the recipient of Ramanujan Fellowship in 2012, IBM Faculty Award in 2013, P. K. Kelkar Fellowship in 2015, CNR Rao Faculty Award, and Humboldt fellowship and Swarnajayanti fellowship in 2018. He was on the technical program committees of IEEE International Electron Devices Meeting (IEDM), IEEE International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), IEEE European Solid-State Device Research Conference (ESSDERC), IEEE Electron Devices Technology and Manufacturing (EDTM), and IEEE International Conference on VLSI Design and International Conference on Embedded Systems.



Hussam Amrouch (Member, IEEE) received the Ph.D. (summa cum laude) degree in computer engineering from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, in 2015.

He is currently a Professor heading the Chair of AI Processor Design, Technical University of Munich (TUM), Munich, Germany. He is also working on brain-inspired computing with the Munich Institute of Robotics and Machine Intelligence (MIRMI), Munich, and is the Head of the Semiconductor Test and Reliability (STAR) research group within the University of Stuttgart, Stuttgart, Germany. He was a Research Group Leader with the KIT, where he was leading the research efforts in building dependable embedded systems. He has authored or coauthored more than 210 publications in multidisciplinary research areas (including around 90 journals) across the entire computing stack, starting from semiconductor physics to circuit design all the way up to computer-aided design and computer architecture. His research interests include design for reliability and testing from device physics to systems, machine learning for CAD, HW security, approximate computing, and emerging technologies with a special focus on ferroelectric devices.

Dr. Amrouch is the Editor of the *Nature Scientific Reports* journal. He was the recipient of eight HIPEAC Paper Awards and three best paper nominations at top EDA conferences: DAC'16, DAC'17, and DATE'17 for his work on reliability. He was on the technical program committees of many major EDA conferences, such as DAC, ASP-DAC, and ICCAD, and is a Reviewer in many top journals, such as *Nature Electronics*, IEEE TRANSACTIONS ON ELECTRON DEVICES, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: REGULAR PAPERS, IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION, IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, and IEEE TRANSACTIONS ON COMPUTERS. His research in HW security and reliability have been funded by the German Research Foundation (DFG), Advantest Corporation, and the U.S. Office of Naval Research (ONR).

Dr. Amrouch is the Editor of the *Nature Scientific Reports* journal. He was the recipient of eight HIPEAC Paper Awards and three best paper nominations at top EDA conferences: DAC'16, DAC'17, and DATE'17 for his work on reliability. He was on the technical program committees of many major EDA conferences, such as DAC, ASP-DAC, and ICCAD, and is a Reviewer in many top journals, such as *Nature Electronics*, IEEE TRANSACTIONS ON ELECTRON DEVICES, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: REGULAR PAPERS, IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION, IEEE TRANSACTIONS ON COMPUTER-AIDED DESIGN OF INTEGRATED CIRCUITS AND SYSTEMS, and IEEE TRANSACTIONS ON COMPUTERS. His research in HW security and reliability have been funded by the German Research Foundation (DFG), Advantest Corporation, and the U.S. Office of Naval Research (ONR).