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

PAUL R. GENSSLER[1], (Member, IEEE), FLORIAN KLEMME[1], (Member, IEEE), SHIVENDRA SINGH PARIHAR[1,2], (Member, IEEE), SEBASTIAN BRANDHOFER[3], (Member, IEEE), GIRISH PAHWA[4], (Member, IEEE), ILIA POLIAN[3], (Senior Member, IEEE), YOGESH SINGH CHAUHAN[2], (Fellow, IEEE), and HUSSAM AMROUCH[1,5,6], (Member, IEEE)

[1]Chair of Semiconductor Test and Reliability (STAR), University of Stuttgart, Stuttgart, Germany. (email: {genssler, klemme, pariharsa}@iti.uni-stuttgart.de)
[2]Nanolab, Department of Electrical Engineering, IIT Kanpur, Kanpur, India. (email: chauhan@iitk.ac.in)
[3]Chair of Hardware Oriented Computer Science, University of Stuttgart, Stuttgart, Germany. (email: {sebastian.brandhofer, ilia.polian}@iti.uni-stuttgart.de)
[4] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, USA. (email: pahwa@berkeley.edu)
[5]Chair of AI Processor Design, Technical University Munich (TUM), Munich, Germany. (email: amrouch@tum.de)
[6]Munich Institute of Robotics and Machine Intelligence (MIRMI), Munich, Germany.

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

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 a few (milli)kelvin and the traditional 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 FinFETs at 10 K 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 FinFET, System on Chip, SoC, Quantum computing, Machine learning classification, Hyperdimensional computing

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 up-scaling, 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.

The 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].
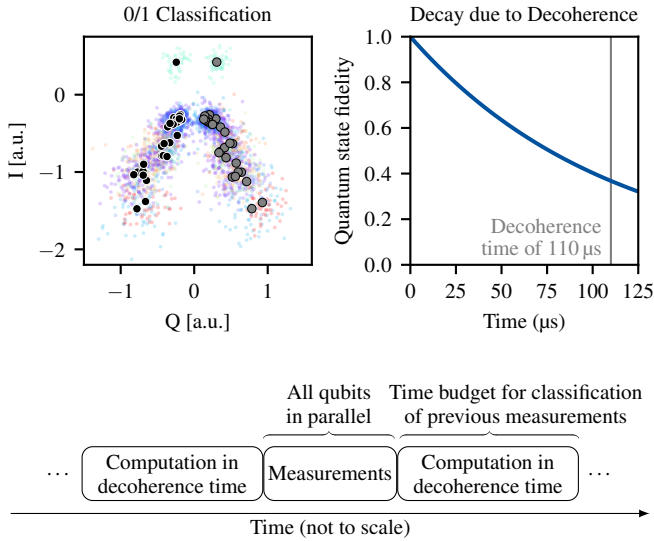
Figure 1. Our measurements from the investigated IBM Falcon quantum processor with 27 qubits. The different colors represent measurements of the different qubits. The black and gray larger dots represent the center of a qubit and are obtained during a calibration phase. Each colored dot has to be classified into 0 (black centers) or 1 (gray centers). The runtime budget of the measurement classification is limited by the decoherence time of the IBM Falcon quantum processor. A 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.

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 requires enormous structural overheads [3], [4]. This includes 200 wide-band coaxial cables, 45 microwave circulators and a rack of electronic circuits. The isolation notwithstanding, 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 K 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.

The 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]. Additionally, 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 a) extremely rapidly and b) at ultra-low power.

The 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 like running calibration protocols, loading the next quantum computation, 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 by [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.

The 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 sub-threshold swing (SS), and an increase of
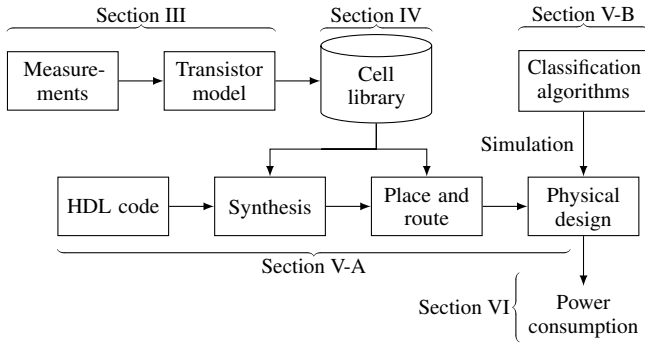
Figure 2. Outline of this paper and an overview of our modeling. We cover the whole stack, from physical transistor measurements through cell library characterization to the physical design simulation of the full system on chip.

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.

Our main contributions within this work are as follows: 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. 2 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 FinFET transistor at room temperature (300 K) and at cryogenic temperature (10 K). Using those measurements, we calibrate the modified cryogenic-aware 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

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

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 [11], 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 [12], [13]. For the IBM quantum computers based on superconducting qubits, typically a boxcar integrator is used to project the measurement signal onto the IQ plane where a single-qubit measurement results in a complex number that is represented as an in-phase and quadrature component [12], [13]. As seen in Fig. 1, the measurement signals of the *i*-th qubit are generally in close proximity in the IQ 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 [14] and plans to build a quantum computer with over four thousand qubits by 2025 [15]. The classified and also the raw IQ 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 [16], [17]. 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 [16], [17]. 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 [18].

Authors in [19], [20] showed that transistors fabricated

using 160 nm 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 outperform the older generation technologies at both 300 K and 4 K. [21], [22] showed that FinFETs from both 14 nm and 10 nm technologies can offer a significant power reduction while operating at cryogenic temperatures for a similar speed. Although, [21], [22] reported the FinFETs cryogenic characterization, these studies were limited up to 77 K. [23] presented the 16 nm FinFET cryogenic characterization from 2.5 K to 300 K. In our previous work [24], we have characterized 5 nm FinFET technology at 300 K 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 sub-threshold behavior of the transistors is affected by interface trap charges and source-drain coupling. Sub-threshold 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 U0 and field-dependent mobility degradation parameters (i.e., UA, UD, EU, and ETAMOB) are extracted from the transfer characteristics (drain source current I_{DS} - gate voltage V_G) when the transistor operates at a low V_{DS} and moderate inversion (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 (Fig. 3) characteristics at lower and higher V_{DS}. 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}.

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]–[28]. 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, 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 non-linearity in the SS is caused by the band-tail effect [27], [28]. 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 an optimization of temperature coefficients for Coulomb scattering (UD1 and UD2), 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

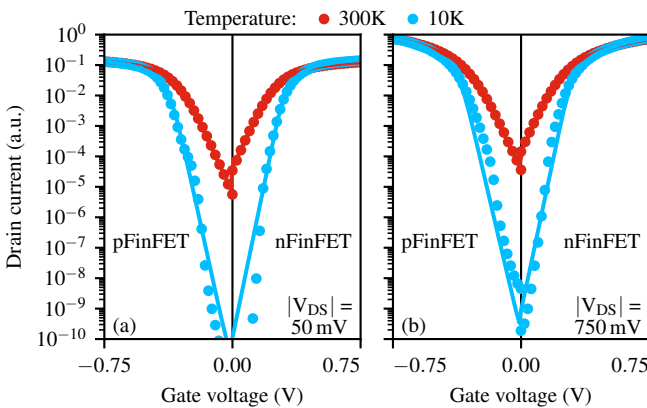**Figure 3.** Transfer characteristics of p- and n-FinFET for 10 K and 300 K in (a) linear (V_{DS} = 50 mV) and (b) saturation (V_{DS} = 750 mV). Symbols and lines show the data from measurement and calibrated model simulation, respectively [24].
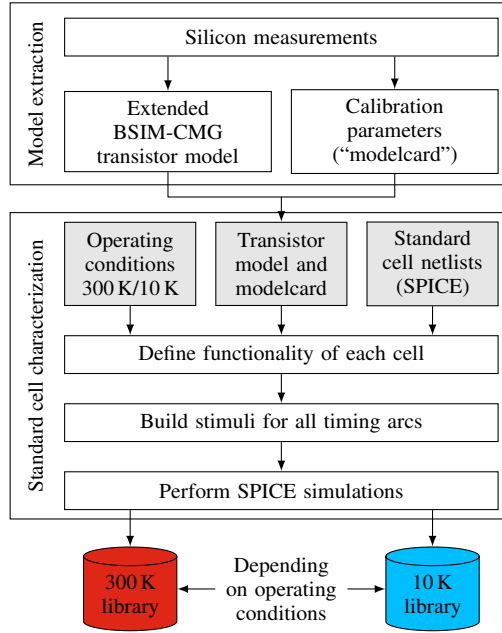
Figure 4. Overview of the design flow from transistor modeling to standard cell libraries. Inputs to the characterization flow are shaded in gray, outputs are highlighted in red (300 K) and blue (10 K) color. The temperature can be adjusted as part of the operating conditions, allowing us to build libraries for different temperature scenarios.

Figure 5. The histogram shows the delays across all 200 cells in the standard cell library. The large overlap of the histograms for 300 K 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).

and pinch-off voltage. Those parameters include the effective drain to source voltage (V_{dseff}) smoothing TMEXP, its temperature coefficients TMEXP1, 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 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. THE 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
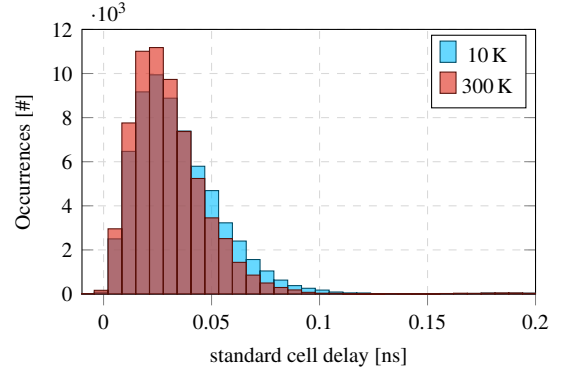
and are vital for the SPICE simulations. Besides the transistor model, SPICE netlists for a wide range of combinational and sequential standard cells are provided to the characterization flow. In this work, we obtain 200 different standard cells from the open source ASAP7 PDK [29] including parasitic 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 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.

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 K 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 5-stage in-order Rocket CPU [30] is combined with a split L1 cache for data and instructions, each with 16 KB and a shared L2 cache of 512 KB. The hardware description language (HDL) code is created with the assistance of the Chipyard framework [31]. 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.

SRAM arrays, the core building block of L1 and L2 caches among others, are provided through the ASAP7 PDK [29] as 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 K 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 [32], 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 non-parametric clustering method [33]. The calibration phase is performed offline and returns the center points for each qubit, as shown in Fig. 1 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 Equation (2).

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

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 Equation (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 [34]. 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 Equation (3) employing the item hypervectors $\vec{x}_P$ and $\vec{y}_P$ for its quantized x and y values.

$$\vec{P} = \vec{x}_P \oplus \vec{y}_P \qquad (3)$$

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 Equation (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 Equation (4).

$$
\begin{aligned}
d_0 &= popcount(\ \vec{C}_0 \oplus \vec{M}) \\
d_0 &= popcount(\ \vec{C}_0 \oplus \vec{x}_M \oplus \vec{y}_M) \qquad (4) \\
d_0 &= popcount(\ \vec{x}_{C0 \oplus M}\ \oplus \vec{y}_M)
\end{aligned}
$$

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, $\vec{x}_{C0 \oplus M}$ is pre-computed 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 Tab. 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 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 nFinFET and pFinFET 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 [35] 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 KB total on-chip SRAM contribute 193 mW. 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.

C. QUANTUM MEASUREMENT CLASSIFICATION: EXECUTION TIME ANALYSIS

The execution times of the two classification algorithms for quantum measurements are evaluated. Tab. 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 × 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. 1. 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 µs. However, typically users strive towards 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,

Table 1. The full SoC is synthesized for 300 K and 0 ns clock period. The timing analysis is conducted with Synopsys PrimeTime for 300 K and 10 K. The difference is less than 10 % because I_{ON} is only slightly affected and thus the delay of the standard cells is similar, as discussed in Section IV-B.

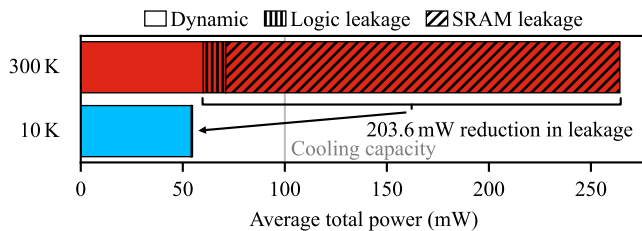| Temperature	Critical path delay	Clock frequency
300 K	1.04 ns	960 MHz
10 K	1.09 ns	917 MHz

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 mW 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.

Table 2. Average clock cycles to classify one measurement. More qubits result in more cache misses increasing the number of clock cycles.

Method	20 qubits	400 qubits
KNN	41.5	72.8
HDC	184.8	242.4

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. 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 (QAOA) or the variational quantum eigensolver (VQE), 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 to enable 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 an 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 an 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.
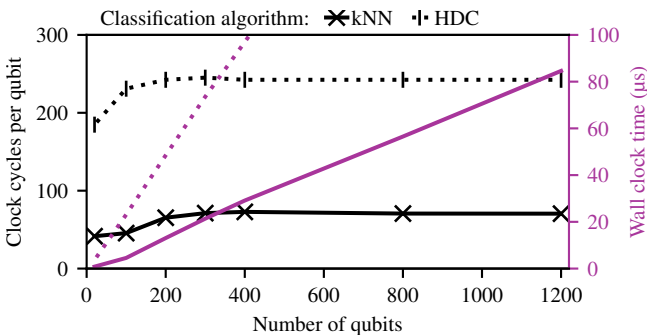

Figure 7. With an increase in the number of qubits, the time to classify all of them through a KNN becomes more important. The maximum quantum computation duration, i.e. the decoherence time, for the investigated IBM Falcon quantum processor is about 110 μs. However, the SoC (clocked at 1000 MHz) would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

References

[1] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, "Elucidating reaction mechanisms on quantum computers," Proceedings of the National Academy of Sciences, vol. 114, no. 29, pp. 7555–7560, 2017.

[2] P. W. Shor, "Algorithms for quantum computation: Discrete logarithms and factoring," in Proceedings 35th annual symposium on foundations of computer science, 1994, pp. 124–134.

[3] S. J. Pauka, K. Das, R. Kalra, A. Moini, Y. Yang, M. Trainer, A. Bousquet, C. Cantaloube, N. Dick, G. C. Gardner, M. J. Manfra, and D. J. Reilly, "A cryogenic CMOS chip for generating control signals for multiple qubits," Nature Electronics, vol. 4, no. 1, pp. 64–70, 2021.

[4] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, et al., "Quantum supremacy using a programmable superconducting processor," Nature, vol. 574, no. 7779, pp. 505–510, 2019.

[5] F. Sebastiano, H. A. Homulle, J. P. van Dijk, R. M. Incandela, B. Patra, M. Mehrpoo, M. Babaie, A. Vladimirescu, and E. Charbon, "Cryogenic CMOS interfaces for quantum devices," in 2017 7th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), 2017, pp. 59–62.

[6] J. Park, S. Subramanian, L. Lampert, T. Mladenov, I. Klotchkov, D. J. Kurian, E. Juarez-Hernandez, B. P. Esparza, S. R. Kale, A. B. K. T., S. P. Premaratne, T. F. Watson, S. Suzuki, M. Rahman, J. B. Timbadiya, et al., "A fully integrated cryo-CMOS SoC for state manipulation, readout, and high-speed gate pulsing of spin qubits," IEEE Journal of Solid-State Circuits, vol. 56, no. 11, pp. 3289–3306, 2021.

[7] S. Bravyi, O. Dial, J. M. Gambetta, D. Gil, and Z. Nazario, "The future of quantum computing with superconducting qubits," Journal of Applied Physics, vol. 132, no. 16, p. 160 902, 2022.

[8] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," arXiv preprint arXiv:1411.4028, 2014.

[9] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'brien, "A variational eigenvalue solver on a photonic quantum processor," Nature Communications, vol. 5, no. 1, pp. 1–7, 2014.

[10] M. A. Nielsen and I. Chuang, Quantum computation and quantum information, 2002.

[11] M. Reiher, N. Wiebe, K. M. Svore, D. Wecker, and M. Troyer, "Elucidating reaction mechanisms on quantum computers," Proceedings of the national academy of sciences, vol. 114, no. 29, pp. 7555–7560, 2017.

[12] 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, p. 064 061, 6 2022.

[13] T. Alexander, N. Kanazawa, D. J. Egger, L. Capelluto, C. J. Wood, A. Javadi-Abhari, and D. C. McKay, "Qiskit pulse: Programming quantum computers through the cloud with pulses," Quantum Science and Technology, vol. 5, no. 4, p. 044 006, 2020.

[14] A-tA-v, M. S. ANIS, Abby-Mitchell, H. Abraham, AduOffei, R. Agarwal, G. Agliardi, M. Aharoni, V. Ajith, I. Y. Akhalwaya, G. Aleksandrowicz, T. Alexander, M. Amy, S. Anagolum, Anthony-Gandon, et al., Qiskit: An open-source framework for quantum computing, 2021.

[15] IBM Quantum Roadmap, https://www.ibm.com/quantum/roadmap.

[16] A. Kamgar and R. Johnston, "Delay times in Si MOSFETS in the 4.2–400 K temperature range," Solid-State Electronics, vol. 26, no. 4, pp. 291–294, 1983.

[17] H. Hanamura, M. Aoki, T. Masuhara, O. Minato, Y. Sakai, and T. Hayashida, "Operation of bulk CMOS devices at very low temperatures," IEEE Journal of Solid-State Circuits, vol. 21, no. 3, pp. 484–490, 1986.

[18] P. A. t Hart, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Subthreshold mismatch in nanometer CMOS at cryogenic temperatures," in ESSDERC 2019 - 49th European Solid-State Device Research Conference (ESSDERC), 2019, pp. 98–101.

[19] 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 Journal of the Electron Devices Society, vol. 6, pp. 996–1006, 2018.

[20] H. Homulle, "Cryogenic electronics for the read-out of quantum processors," Ph.D. dissertation, Delft University of Technology, Mekelweg 5, 2628 CD Delft, Netherlands, 2019.

[21] A. Chabane, M. Prathapan, P. Mueller, E. Cha, P. A. Francese, M. Kossel, T. Morf, and C. Zota, "Cryogenic characterization and modeling of 14 nm bulk FinFET technology," in ESSCIRC 2021 - IEEE 47th European Solid State Circuits Conference (ESSCIRC), 2021, pp. 67–70.

[22] H. L. Chiang, T. C. Chen, J. F. Wang, S. Mukhopadhyay, W. K. Lee, C. L. Chen, W. S. Khwa, B. Pulicherla, P. J. Liao, K. W. Su, K. F. Yu, T. Wang, H. S. P. Wong, C. H. Diaz, and J. Cai, "Cold CMOS as a power-performance-reliability booster for advanced FinFETs," in 2020 IEEE Symposium on VLSI Technology, 2020, pp. 1–2.

[23] H.-C. Han, F. Jazaeri, A. D'Amico, A. Baschirotto, E. Charbon, and C. Enz, "Cryogenic characterization of 16 nm FinFET technology for quantum computing," in ESSCIRC 2021 - IEEE 47th European Solid State Circuits Conference (ESSCIRC), 2021, pp. 71–74.

[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 Transactions on Circuits and Systems I: Regular Papers, pp. 1–14, 2023.

[25] BSIM-CMG Technical Manual.

[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 Transactions on Electron Devices, vol. 68, no. 9, pp. 4223–4230, 2021.

[27] H. Bohuslavskyi, A. G. M. Jansen, S. Barraud, V. Barral, M. Cassé, L. Le Guevel, X. Jehl, L. Hutin, B. Bertrand, G. Billiot, G. Pillonnet, F. Arnaud, P. Galy, S. De Franceschi, M. Vinet, et al., "Cryogenic subthreshold swing saturation in FD-SOI MOSFETs described with band broadening," IEEE Electron Device Letters, vol. 40, no. 5, pp. 784–787, 2019.

[28] A. Beckers, F. Jazaeri, and C. Enz, "Theoretical limit of low temperature subthreshold swing in field-effect transistors," IEEE Electron Device Letters, vol. 41, no. 2, pp. 276–279, 2020.

[29] V. Vashishtha, M. Vangala, and L. T. Clark, "ASAP7 predictive design kit development and cell design technology co-optimization: Invited paper," in 2017 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), 2017, pp. 992–998.

[30] K. Asanovic, R. Avizienis, J. Bachrach, S. Beamer, D. Biancolin, C. Celio, H. Cook, D. Dabbelt, J. Hauser, A. Izraelevitz, et al., "The rocket chip generator," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2016-17, vol. 4, 2016.

[31] A. Amid, D. Biancolin, A. Gonzalez, D. Grubb, S. Karandikar, H. Liew, A. Magyar, H. Mao, A. Ou, N. Pemberton, P. Rigge, C. Schmidt, J. Wright, J. Zhao, Y. S. Shao, et al., "Chipyard: Integrated design, simulation, and implementation framework for custom SoCs," IEEE Micro, vol. 40, no. 4, pp. 10–21, 2020.

[32] 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, p. 200 501, 20 2015.

[33] E. Fix and J. L. Hodges, "Discriminatory analysis. nonparametric discrimination: Consistency properties," International Statistical Review/Revue Internationale de Statistique, vol. 57, no. 3, pp. 238–247, 1989.

[34] P. Kanerva, "Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors," Cognitive computation, vol. 1, no. 2, pp. 139–159, 2009.

[35] R. P. Weicker, "Dhrystone: A synthetic systems programming benchmark," Commun. ACM, vol. 27, no. 10, pp. 1013–1030, 1984.

PAUL R. GENSSLER (M'21) received the Dipl. Inf degree (M.Sc.) in computer science in 2017 at TU Dresden, Germany. In 2018 he started his PhD research at the Chair for Embedded Systems (CES) at Karlsruhe Institute of Technology, Germany. Since 2020 he continues his PhD at the Semiconductor Test and Reliability (STAR) chair within the Computer Science, Electrical Engineering Faculty at the University of Stuttgart. His research interests include emerging technologies, system architecture, and emerging brain-inspired methods for IC test and beyond.

FLORIAN KLEMME (M'20) is a Doctoral Researcher at the Chair of Semiconductor Test and Reliability (STAR), University of Stuttgart. He received the B.Sc. in System Integration from the University of Applied Sciences Bremerhaven, Germany, in 2014 and the M.Sc. in Computer Science from the Karlsruhe Institute of Technology, Germany, in 2018. He is currently working towards the Ph.D. degree at the Chair of Semiconductor Test and Reliability, University of Stuttgart. His research interests include cell library characterization and machine learning techniques in electronic design automation and computer-aided design, specifically towards the reliability of transistors and integrated circuits in advanced technology nodes.

SHIVENDRA SINGH PARIHAR is a doctoral student at the Indian Institute of Technology, Kanpur, Uttar Pradesh, India. Currently, he is associated with the Chair of Semiconductor Test and Reliability (STAR) at the University of Stuttgart, Germany as a research scholar. His primary research interests are characterization and compact modeling of advanced CMOS technologies for circuit design.

SEBASTIAN BRANDHOFER received the M.Sc. degree in computer science from the University of Stuttgart, Germany in 2017. He started his PhD research at the Chair of Hardware Oriented Computer Science (HOCOS) at the University of Stuttgart in 2018. His research interests include quantum computing, quantum circuit optimization and computer-aided design. He won the IBM Quantum Open Science Prize in 2021.

GIRISH PAHWA is an assistant professional researcher in the department of electrical engineering and computer sciences (EECS) at the University of California (UC) Berkeley. He is also the executive director of the Berkeley Device Modeling Center (BDMC) at UC Berkeley. Prior to this, he worked as a postdoctoral researcher at EECS, UC Berkeley, from 2020-2021. He received his Ph.D. and M.Tech. Degrees in electrical engineering from the Indian Institute of Technology (IIT) Kanpur in 2020. His research primarily focuses on 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. 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 over 50 technical publications in prominent journals and conferences in the field of device modeling and simulation. He is the recipient of the IEEE EDS early career award in 2022, the outstanding Ph.D. thesis award from IIT Kanpur in 2020, and the best paper award at the IEEE International Conference on Emerging Electronics (ICEE), Mumbai, India, in 2016. He is a member of IEEE and serves as the reviewer of several reputed journals.

ILIA POLIAN received the Diploma M.Sc. and Ph.D. degrees from the University of Freiburg, Germany, in 1999 and 2003, respectively. He is a Full Professor and the Director of the Institute for Computer Architecture and Computer Engineering, University of Stuttgart, Germany. He coauthored over 200 scientific publications. His scientific interests include hardware-oriented security, emerging architectures, test methods, and quantum computing. He received two Best Paper Awards.

YOGESH SINGH CHAUHAN is a 'Class of 1984' Chair professor at Indian Institute of Technology Kanpur, India. He was with Semiconductor Research & Development Center at IBM Bangalore during 2007 – 2010; Tokyo Institute of Technology in 2010; University of California Berkeley during 2010-2012; and ST Microelectronics 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, FinFET, Nanosheet/Gate-All-Around FETs, FDSOI transistors, Negative Capacitance FETs and 2D FETs. His research interests are characterization, modeling, and simulation of semiconductor devices. He is the Fellow of IEEE and Indian 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 has published more than 300 papers in international journals and conferences. He received Ramanujan fellowship in 2012, IBM faculty award in 2013 and P. K. Kelkar fellowship in 2015, CNR Rao faculty award, Humboldt fellowship and Swarnajayanti fellowship in 2018. He has served in 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 (S'11-M'15) is Professor heading the Chair of AI Processor Design within the Technical University of Munich (TUM). He is, additionally, with the Munich Institute of Robotics and Machine Intelligence (MIRMI) in Germany and is the head of the Semiconductor Test and Reliability (STAR) research group within the University of Stuttgart, Germany. Prior to that, he was a Research Group Leader at the Karlsruhe Institute of Technology (KIT) where he was leading the research efforts in building dependable embedded systems. He currently serves as Editor at the Nature Scientific Reports Journal. He received his Ph.D. degree with the highest distinction (Summa cum laude) from KIT in 2015. His main research interests are 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. He holds 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 has served in the technical program committees of many major EDA conferences such as DAC, ASP-DAC, ICCAD, etc., and as a reviewer in many top journals like Nature Electronics, T-ED, TCAS-I, TVLSI, TCAD, TC, etc. He has more than 210 publications in multidisciplinary research areas (including 86 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 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).