A Cryo-CMOS Digital Cell Library for Quantum Computing Applications

E. Schriek, F. Sebastiano, *Senior Member, IEEE*, and E. Charbon, *Fellow, IEEE*

***Abstract*—We present a digital cell library optimized for 4.2 K to create controllers that keep quantum processors coherent and entangled. The library, implemented on a standard 40-nm CMOS technology, was employed in the creation of the first 4.2 K RISC-V processor. It has achieved a minimum supply voltage of 590 mV, energy-delay product of 37 fJ/MHz, and maximum operating frequency of 740 MHz, all at 4.2 K in continuous operation. These results have been obtained from stand-alone characterization, successfully executing small C programs/benchmarks at 4.2 K. The overall performance of the library compares well against the state-of-the-art libraries designed for room temperature. In particular, we compared the performance of the proposed library against a foundry supplied library for the same process in several combinational benchmark circuits, showing significant improvements in power dissipation and frequency of operation.**

***Index Terms*—CMOS library, cryo-CMOS, cryogenic electronics, quantum computing, RISC-V.**

I. Introduction

Quantum computing promises a tremendous speedup over conventional von Neumann architectures. The core of a quantum computer is an array of quantum bits (qubits) that need to be controlled by classical circuits. Today, classical controllers for qubit arrays are designed to operate at room temperature (RT) to enable the highest possible flexibility during experimentation. However, only a limited number of qubits can be controlled in this manner due to the complexity of RT controllers, both in hardware and wiring. To support scaling of qubit arrays, we have proposed the use of CMOS systems operating at deep-cryogenic temperatures (cryo-CMOS) [1]. At these temperatures, typically 1–4.2 K, the available cooling power limits the power budget of such electronics to a few Watts, thus calling for low-power (LP) system design, where LP in this context means "as low as possible." In this letter, we present an LP cell library designed to operate from RT down to deep-cryogenic temperatures, where we exploit the improved device characteristics to enhance digital performance operating in super- and sub-threshold regimes.

II. Low Voltage, Low Power in Deep-Cryogeny

Prior research has demonstrated that the absolute theoretical lower limit for the supply voltage of CMOS digital circuits at RT (300 K) is $V_{DD,\min} \approx 2(kT/q)\ln(2) = 36$ mV [2]. At these voltages, MOS transistors in a CMOS circuit are operating in subthreshold, where the following equation holds:

$$I_{DS} = I_0 \frac{W}{L} e^{\frac{V_{GS}-V_{TH}}{nv_t}} \left(1 - e^{-\frac{V_{DS}}{v_t}}\right);\ I_0 = \mu_0 C_{ox} \frac{W}{L}(n-1)v_t^2 \quad (1)$$

where n is the subthreshold slope (SS) factor and $v_t = kT/q$, while the other parameters are well known in solid-state device engineering.

At deep-cryogenic temperatures (<30 K), MOSFET threshold voltage V_{TH} increases by about 100 mV, mobility doubles, and SS is well below the 60-mV/decade limit imposed by RT operation, even though n increases by an order of magnitude (SS $= \ln(10)nv_t$) [3], [4]. Other parameters, such as transistor mismatch, can substantially impact design margins and yield, especially in the sub-/near-threshold region of operation due to the steepness of SS [5].

The net effect in subthreshold regimes is a decrease of leakage currents, as a byproduct of the decreased SS, by orders of magnitude, implying a significant increase in the $I_{\mathrm{ON}}/I_{\mathrm{OFF}}$ ratio, as shown in the measurement of Fig. 1. Assuming an ideal SS factor $n = 1$, at 4.2 K, according to well established RT models, one could theoretically achieve $V_{DD,\min} \approx 2\ln(2)v_t = 0.48$ mV, which appears to be a significant design opportunity in minimizing power consumption, by aggressively reducing V_{TH} thanks to the significantly decreased leakage current.

However, at 4.2 K, the consensus is that $n \approx 34.9$ [3], [4]. Thus, this fundamental limit is actually $V_{DD,\min} \approx 2.47$ mV. Additional nonidealities include reverse short-channel effect (RSCE) and inverse narrow-width effect (INWE). Both effects substantially modulate the threshold voltage. RSCE was shown to have a significantly reduced effect down to 30 K [6], due to the temperature dependence of the bulk Fermi potential, which is less sensitive to doping concentration at low temperatures. For INWE, the effect is significantly larger for the pMOS, which results in an imbalance with nMOS devices, thus leading to a significant increase in $V_{DD,\min}$. However, to the best of our knowledge, no study of INWE below 77 K, and in particular at 4.2 K, exists.

Finally, latch-up has been found to be unpredictable in deep-cryogenic operation. Latch-up immunity typically improves at temperatures lower than RT, thanks to higher base–emitter voltages and lower current gain of parasitic bipolar transistors. However, substrate and well resistance increase significantly and shallow-level impact ionization (SLII), a mechanism for carrier generation, emerges below 50 K [7]. The presence of this mechanism invalidates the classical condition for latch-up at RT, thus appearing to deteriorate the latch-up immunity at deep-cryogenic temperature (4.2 K) with respect temperatures above 50 K.

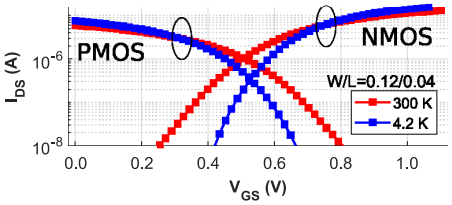

Fig. 1. I_{DS} versus V_{GS} for standard 40-nm CMOS transistors at 300 K and 4.2 K.

Manuscript received May 17, 2020; revised July 7, 2020; accepted August 5, 2020. Date of publication August 18, 2020; date of current version September 9, 2020. This article was approved by Associate Editor Sorin Voinigescu. This work was supported in part by Intel Corporation. *(Corresponding author: E. Charbon.)*

E. Schriek was with the Department of Quantum and Computer Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands. He is now with GrAI Matter Labs, 5656 AG Eindhoven, The Netherlands.

F. Sebastiano is with the Department of Quantum and Computer Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands, and also with the QuTech, Delft University of Technology, 2628 CD Delft, The Netherlands.

E. Charbon is with the Faculty of Engineering, EPFL, 2002 Neuchâtel, Switzerland, and also with the Kavli Institute of Nanoscience Delft, 2628 CD Delft, The Netherlands (e-mail: edoardo.charbon@epfl.ch).

Digital Object Identifier 10.1109/LSSC.2020.3017705

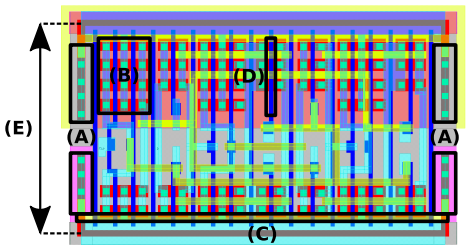

Fig. 2. Proposed cell layout features highlighted on D-flip-flop: (A) integrated well taps on cell boundary, (B) INWE aware sizing, (C) secondary power rails for back-biasing, (D) exclusively minimum length transistors, and (E) increased mismatch aware cell sizing.

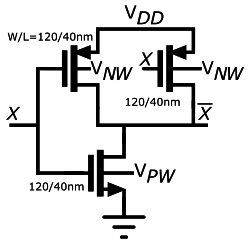

Fig. 3. Schematic of the 1X strength inverter cell.

TABLE I
IMPLEMENTED CELLS AS A PROOF-OF-CONCEPT LIBRARY

Cell Name	Logic Style	Strengths	Fan-in
Inverter	Static	1/2/4X	1
Buffer	Static	1/2/4X	1
Nand-2	Static	1/2X	2
And-Or-Invert-2	Static	1/2X	3
D Flip-Flop, No Reset	Transmission Gate	1/2X	1
Multiplexer-2	Transmission Gate	1X	2
Level-Shifter (LH/HL)	Static	1X/1X	1/1
Power Gate	Header	N/A	N/A
Decap	Cross-Coupled	N/A	N/A

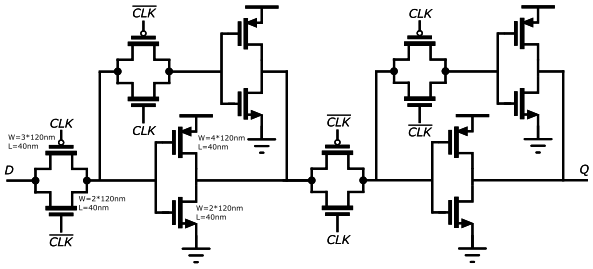

Fig. 4. D-flip-flop with specific sizing implementing measures (A), (B), (C), (D), and (E), as described in the text.

III. DESIGN OF COOLIB AND COOLRISC-V

In conventional LP and in subthreshold LP design, RSCE and INWE are often exploited to reduce V_{TH}, SS, and the input capacitance. In addition, different choices in cell sizing, drive strength, gate fan-in, or logic style are generally adopted. Sometimes, back-biasing is added to contain leakage power by applying a reverse body-bias. Based on this knowledge and supported by extensive device and parasitics measurements, we created a CMOS library designed to operate in moderate subthreshold regimes at deep-cryogenic temperatures. The library, denominated cooLib, was fabricated in standard 40-nm CMOS technology and tested continuously from 300 K down to 4.2 K in a liquid He (LHe) cryostat.

In the library we adopted the following methodology.

1) Create extensive substrate contacts and well-taps, so as to minimize the chance of latch-up at 4.2 K.
2) Resize the transistors widths to a minimum, to reduce V_{TH} by exploiting INWE modulation. nMOS/pMOS imbalance is tackled by "stacking" multiple minimum width transistors.
3) Add secondary power rails to enable forward back-biasing for nMOS (V_{PW}) and pMOS (V_{NW}) devices independently, so as to compensate for an increase of V_{TH} at 4.2. To support this feature, a triple-well process is selected. In addition, use low-V_{TH} transistors.
4) Minimize the length of transistors (in contrast to conventional RT subthreshold standard cell design, where a nonminimal length is typically chosen to exploit RSCE).
5) When useful, make the layout aware of mismatch by increasing the overall height of the cells.
6) Choices in drive strength and fan-in are kept to the smallest possible values, as is common for the state-of-the-art in 300-K standard cells.
7) Certain basic logic functions are omitted, such as NOR, as typical 2/1 sizing in the subthreshold does not hold and therefore requiring excessive pMOS sizing to achieve balance.

All these measures are exemplified in Figs. 2 and 3, where transistor "stacking" for the pMOS is used. Table I lists the cells implemented in cooLib. Fig. 4 shows how sizing was used to implement measures (A), (B), (C), (D), and (E) in the table. Measures (F) and (G) were applied to the whole library and cannot be shown in the schematics. We used both static and transmission gate CMOS logic styles, as well as header, and cross-coupled techniques. Next, we embedded cooLib as a library in a design flow employing a suite of standard EDA tools from various vendors. The purpose was the design of more complex components, based on cooLib and relevant for quantum-computing applications, such as signal-processing functions and microprocessors.

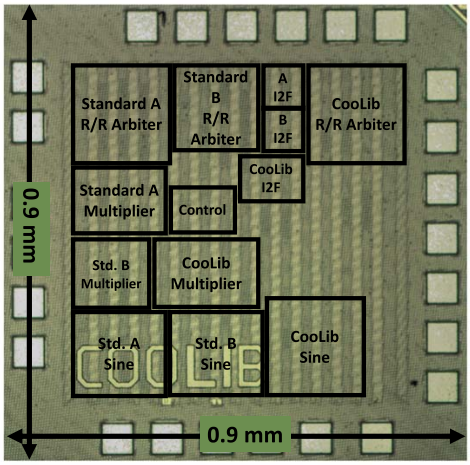

Fig. 5. Micrograph of the test-chip fabricated in a standard 40-nm CMOS process, implementing cooLib. In the insets, the implemented functions are indicated (see text and Table III).

Two chips were fabricated: the first, shown in Fig. 5, implemented a benchmark set derived from the EPFL combinational benchmark suite [8]. Each of the benchmarks was implemented in three versions: one with the proposed cells, an implementation with foundry-supplied cells restricted to cells listed in Table I (LP-A in the following), and an implementation with the unrestricted foundry-supplied cells (LP-B in the following). The second chip, "CoolRISC-V," shown in Fig. 6, is a proof-of-concept of a fully functional microprocessor based on the

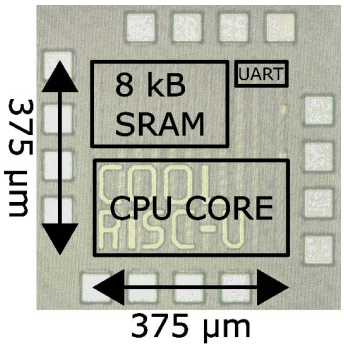

Fig. 6. Micrograph of the RISC-V chip fabricated in a standard 40-nm CMOS process.

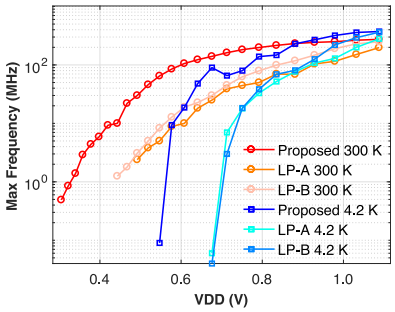

Fig. 7. F_{MAX} for the 16×16 multiplier.

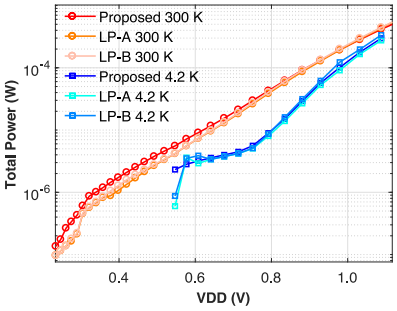

Fig. 8. Total power at 100 kHz for the multiplier.

TABLE II
COMPARISON WITH STATE-OF-THE-ART

	This Work		ESSCIRC'18 [8]	JSSC'17 [9]
Technology	40-nm CMOS		28-nm FDSOI	40-nm CMOS
Architecture	RV32IM		RV32IM	Cortex-M0
Temperature	4.2 K	300 K	300 K	300 K
F_{MAX}	740 MHz @ 1.2 V	475 MHz @ 1.2 V	-	-
F @ V_{MIN}	9 MHz @ 0.59 V	3.2 MHz @ 0.3 V	1 MHz @ 0.25 V	0.8 MHz @ 0.2 V
Min. E_{CORE}	8.90 pJ	5.80 pJ	4.18 pJ	8.80 pJ
Min. EDP_{CORE}	0.037 pJ/MHz	0.061 pJ/MHz	0.075 pJ/MHz	0.6 pJ/MHz

RISC-V architecture [9]. This chip was based on the standard RISC-V instruction set; it was implemented using standard automatic place & route tools where we selected CooLib's standard cells.

IV. RESULTS AND DISCUSSION

We created a setup that could be used to test the maximum operating frequency F_{MAX} and power dissipation at 300 K and progressively down to 4.2 K. Tables II and III list the results of the

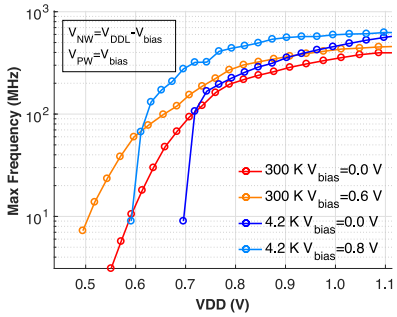

Fig. 9. F_{MAX} for the RISC-V processor.

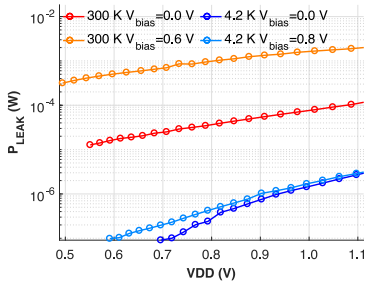

Fig. 10. Leakage power in the RISC-V.

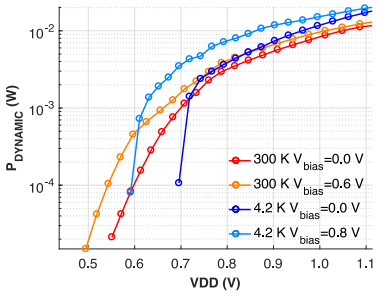

Fig. 11. Dynamic power in the RISC-V.

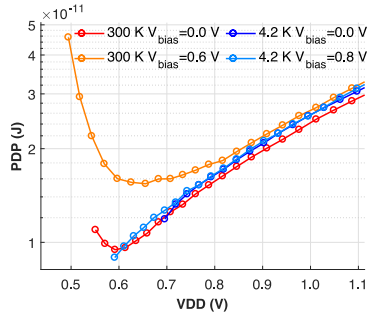

Fig. 12. Power delay product in the RISC-V.

cell library comparison. As an example, F_{MAX} and the dissipated power for one of the benchmark cells are shown in Figs. 7 and 8, respectively, to demonstrate proper operation down to 4.2 K. Let us define the minimum supply voltage where the logic is still function as $V_{DD,op}$. The proposed cells allow a reduction of $V_{DD,op}$ by 0.14 V without forward bias V_{BIAS}. Consequently, a significant improvement of F_{MAX} was obtained. In turn, this allows for a power reduction with quadratic scaling, as a lower supply voltage is required to achieve equal frequency.

Figs. 9–15 outline the RISC-V performance. At 4.2 K, an additional 0.1-V reduction in supply voltage can be achieved with forward

TABLE III
PERFORMANCE COMPARISON WITH A CONVENTIONAL FOUNDRY SUPPLIED LP STANDARD CELL LIBRARY. IMPLEMENTATION A: 40-NM LP RESTRICTED TO CELLS DEFINED IN TABLE I. IMPLEMENTATION B: LP WITHOUT CELL RESTRICTION. THE REPORTED P_{AVG} NUMBERS ARE MEASURED AT THE CORRESPONDING $V_{DD,OP}$

		$V_{DD,OP}$ [V]			F_{MAX} @ 0.6 V [MHz]			F_{MAX} @ 0.7 V [MHz]			P_{AVG} @ 100 kHz [μW]		
Benchmark	Temp.	Pro-posed	A	B	Pro-posed	A	B	Pro-posed	A	B	Pro-posed	A	B
16X16 Multiplier	4.2 K	0.54	0.68	0.68	16.3	-	-	74.2	4.6	2.0	2.34	3.76	3.88
	300 K	0.3	0.49	0.44	100.4	9.7	17.4	145.2	34.0	39.7	0.61	2.68	1.92
Sine	4.2 K	0.58	0.68	0.68	1.95	-	-	20.9	2.2	1.1	3.91	4.00	4.46
	300 K	0.39	0.34	0.39	15.2	9.2	9.5	29.6	25.5	26.4	3.46	2.11	2.18
Int-to-Float	4.2 K	0.54	0.68	0.68	51.4	-	-	178.1	42.5	11.7	0.80	1.57	3.41
	300 K	0.24	0.38	0.36	118.5	75.2	73.44	174.2	191.9	158.0	0.08	0.55	0.28
Round-Robin Arbiter	4.2 K	0.58	0.68	0.68	21.6	-	-	46.6	2.0	1.8	7.89	11.56	11.64
	300 K	0.32	0.32	0.31	33.7	10.0	34.7	59.9	37.3	80.8	1.72	3.20	2.25

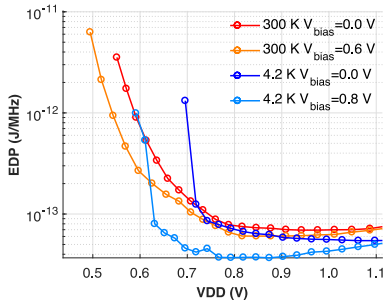

Fig. 13. Energy delay product in the RISC-V.

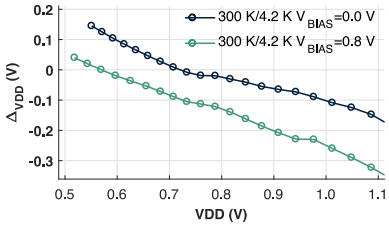

Fig. 14. ΔVDD required to achieve equal F_{MAX}.

bias. The maximum forward-bias voltage is higher at 4.2 K, thanks to the source-bulk diode requiring an additional 0.2 V to start conducting on top of the typical 0.7 V at 300 K. The increase in carrier mobility resulted in an F_{MAX} increase of about 1.6× for $V_{DD} \gg V_{\mathrm{TH}}$. As V_{DD} approaches V_{TH}, a crossover in F_{MAX} and dynamic power P_{DYN} is observed, as predicted by the $I_{DS} - V_{GS}$-relationship of Fig. 1. After this crossover, both F_{MAX} and P_{DYN} appear to follow the SS, which effectively raises $V_{DD,op}$ by almost 2× when compared to that of 300 K. While the reduction in leakage power matched expectations, the insensitivity of leakage current to forward biasing at 4.2 K is noteworthy. Lower leakage results in a lower minimum-energy point achieved at a lower V_{DD} (Fig. 10). The improvements in delay and leakage energy, together, allow for a very low energy-delay product for maximum forward biasing at 4.2 K (Fig. 13). Finally, when operating at the same F_{MAX}, considerable reductions in power consumption were obtained at 4.2 K, compared to 300 K (Fig. 15), thanks to a significant decrease of V_{DD} (Fig. 14). Furthermore, this improvement increases with forward biasing.

The theoretical fundamental limit of $V_{DD,\mathrm{min}} \approx 2.47$mV at 4.2 K appears to be impractical, as the minimum achieved $V_{DD,op}$ is many orders of magnitude larger. The steep SS, quickly reducing I_{DS} as voltage drops at 4.2 K shows to be the main showstopper, as F_{MAX} drops very quickly below the MHz range. Further ways of decreasing V_{TH} are necessary.

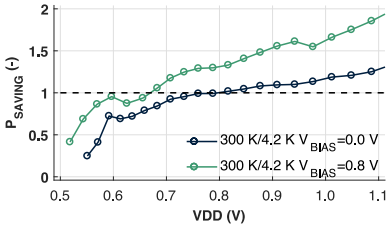

Fig. 15. Average power improvements for equal F_{MAX}.

Although additional improvements could be achieved by adopting FD-SOI to further reduce V_{TH}, the proposed standard cells optimized for 4.2-K operation allowed the microprocessor to be competitive with the state-of-the-art LP implementations (Table II), with a superior energy-delay product, thus demonstrating the suitability of the approach for the classical cryogenic hardware required in quantum computers.

REFERENCES

[1] E. Charbon *et al.*, "Cryo-CMOS for quantum computing," in *Proc. IEEE Int. Electron Device Meeting*, San Francisco, CA, USA, 2016, pp. 343–346.
[2] R. M. Swanson and J. D. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *IEEE J. Solid-State Circuits*, vol. 7, no. 2, pp. 146–153, Apr. 1972.
[3] 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.
[4] A. Beckers, F. Jazaeri, H. Bohuslavskyi, L. Hutin, S. De Franceschi, and C. Enz, "Characterization and modeling of 28-nm FDSOI CMOS technology down to cryogenic temperatures," *Solid-State Electron.*, vol. 159, pp. 106–115, Sep. 2019.
[5] P. A. T'Hart, J. V. Dijk, M. Babaie, E. Charbon, A. Vladimirescu, and F. Sebastiano, "Characterization and model validation of mismatch in nanometer CMOS at cryogenic temperatures," in *Proc. IEEE Eur. Solid-State Circuits Conf.*, Dresden, Germany, 2018, pp. 246–249.
[6] B. Szelag, F. Balestra, and G. Ghibaudo, "Comprehensive analysis of reverse short-channel effect in silicon MOSFETs from low-temperature operation," *IEEE Electron Device Lett.*, vol. 19, no. 12, pp. 511–513, Dec. 1998.
[7] J. Morris, P. Prabhat, J. Myers, and A. Yakovlev, "Unconventional layout techniques for a high performance, low variability subthreshold standard cell library," in *Proc. IEEE Comput. Soc. Annu. Symp. VLSI (ISVLSI)*, 2017, pp. 19–24.
[8] L. Amaru, P.-E. Gaillardon, and G. De Micheli, "The EPFL combinational benchmark suite," in *Proc. Int. Workshop Logic Synth. (IWLS)*, 2015, pp. 1–5.
[9] R. Uytterhoeven and W. Dehaene, "A sub 10 pJ/cycle over a 2 to 200 MHz performance range RISC-V microprocessor in 28 nm FDSOI," in *Proc. IEEE Eur. Solid-State Circuits Conf.*, Dresden, Germany, 2018, pp. 236–239.