Special Session: Exploring the Ultimate Limits of Adiabatic Circuits

Michael P. Frank
Cognitive & Emerging Computing Dept.
Sandia National Laboratories
Albuquerque, NM, USA
mpfrank@sandia.gov

Robert W. Brocato
RF Microsystems Dept.
Sandia National Laboratories
Albuquerque, NM, USA
orcid:0000-0001-9751-1234

Thomas M. Conte
Schools of Computer Science and Electrical & Computer Eng.
Georgia Institute of Technology
Atlanta, GA, USA
conte@gatech.edu

Alexander H. Hsia
Sandia National Laboratories
Albuquerque, NM, USA

Requiescat in pace

Anirudh Jain
School of Computer Science
Georgia Institute of Technology
Atlanta, GA, USA
anirudh.j@gatech.edu

Nancy A. Missert
Nanoscale Sciences Department
Sandia National Laboratories
Albuquerque, NM, USA
orcid:0000-0003-2082-2282

Karpur Shukla
Laboratory for Emerging Technologies
Brown University
Providence, RI, USA
orcid:0000-0002-7775-6979

Brian D. Tierney
Radiation Hard CMOS Technology Dept.
Sandia National Laboratories
Albuquerque, NM, USA
bdtiern@sandia.gov

***Abstract*—The field of adiabatic circuits is rooted in electronics know-how stretching all the way back to the 1960s and has potential applications in vastly increasing the energy efficiency of far-future computing. But now, the field is experiencing an increased level of attention in part due to its potential to reduce the vulnerability of systems to side-channel attacks that exploit, *e.g.*, unwanted EM emissions, power supply fluctuations, and so forth.**

In this context, one natural question is: Just how low can the energy dissipation from adiabatic circuits, and the associated extraneous signal emissions, be made to go? We argue that the ultimate limits of this approach lie much farther away than is commonly appreciated. Recent advances at Sandia National Laboratories in the design of fully static, fully adiabatic CMOS logic styles and high-quality energy-recovering resonant power-clock drivers offer the potential to reduce dynamic switching losses by multiple orders of magnitude, and, particularly for cryogenic applications, optimization of device structures can reduce the standby power consumption of inactive devices, and the ultimate dissipation limits of the adiabatic approach, by multiple orders of magnitude as well.

In this paper, we review the above issues, and give a preliminary overview of our group's activities towards the demonstration of groundbreaking levels of energy efficiency for semiconductor-based logic, together with a broader exploration of the ultimate limits of physically realizable techniques for approaching the theoretical ideal of perfect thermodynamic reversibility in computing, and the study of the implications of this technology direction for practical computing architectures.

Keywords—Adiabatic circuits, adiabatic logic, CMOS, physical limits of computing, low-power design, reversible computing

This work is supported by the Advanced Simulation and Computing (ASC) program at the U.S. Department of Energy's National Nuclear Security Administration (NNSA). Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for NNSA under contract DE-NA0003525. This document describes objective technical results and analysis. Any subjective views or opinions that might be expressed in this document do not necessarily represent the views of the U.S. Department of Energy or the United States Government. Approved for public release, SAND2020-8419 C.

I. Introduction

In 1961, Rolf Landauer of IBM observed that there is a fundamental physical limit to the energy efficiency of *logically irreversible* computational operations, meaning, those that lose known information to their thermal environment [1]. This observation, now enshrined as *Landauer's Principle*, is today understood (with some important qualifications) as a rigorous theorem of fundamental statistical physics and information theory [2], and is widely considered to be one of the most important results in the fundamental physics of computing.

During the 1970s, Landauer's protégé Charles Bennett realized that an alternative paradigm for computation based on *logically reversible* operations, which could in principle sidestep Landauer's efficiency limit, was theoretically coherent [3], and developed a more sophisticated understanding of the thermodynamics of computation built on this insight [4], [5]. Early concepts for engineering implementations of reversible computing (*e.g.*, [6], [7]) eventually led to Younis & Knight's discovery (in the early 1990s) that it could even be achieved for pipelined sequential CMOS, using their CRL [8] and SCRL [9] logic styles. CRL was the first complete sequential CMOS logic family that is what we call *truly, fully adiabatic*, meaning that, in an idealized, no-leakage limit, with negligible parasitic capacitances, the dissipation per logic operation can in principle be made *arbitrarily small*—even below Landauer's limit of $E_{\text{diss}} \geq kT \ln 2$ for irreversible bit erasures. Due to Landauer's Principle, we know that any physical mechanism capable of beating this limit *must* be logically reversible—although what that *means*, precisely, is misapprehended in much of the literature (see [10] for a correct account).

Note that avoiding energy dissipation *includes* avoiding energy emissions associated with side-channel attacks such as Differential Power Analysis (DPA); this led to attention being paid to the possible security applications of adiabatic circuits and reversible computing as early as 2006 [11]. However, existing implementations of adiabatic circuits typically remain far from the reversible ideal, for several reasons that we will discuss.

In 2017, an effort informally termed the *Adiabatic Circuits Feasibility Study* (ACSF) began at Sandia National Laboratories with the goal of pushing the limits of energy efficiency in physical implementations of adiabatic CMOS circuits, as a part of an effort to assess the long-term feasibility of these techniques for possible use in future high-performance computing (HPC) systems. Although the main goal of this effort is to attain more power-efficient (and ultimately, more cost-efficient) computation, improving security can be a potentially beneficial side effect, as previously mentioned.

II. Technical Approach

As mentioned above, to avoid limiting the energy efficiency (*i.e.*, to approach zero energy dissipation per operation), the physical operation of our circuits must be *truly, fully adiabatic*, which *requires* us to apply the principles of reversible computing, albeit in the generalized sense discussed in [10]. Ignoring leakage for the moment, even just to *approach* reversible, adiabatic operation in a switching circuit requires that the following fundamental "design rules" must be obeyed [12]:

1) Avoid passing current through diodes. This is due to the "diode drop," an intrinsic voltage drop that exists in a diode, which is unavoidable for fundamental thermodynamic reasons.

2) Avoid turning a switch (such as a FET) "on" when there is a (non-negligible) voltage difference V_{DS} between its drain and source terminals. Such an event would result in a substantial, non-adiabatic flow of current.

3) Avoid turning a switch "off" when there is a non-negligible current I_{DS} between drain and source, unless these nodes remain connected along an alternate path. The reasoning behind this rule is more subtle, and so it is not widely recognized, and is broken by many designs. Nevertheless, obeying this rule is essential for approaching physical reversibility [13].

Rules 2 & 3 together imply that transitions in fully adiabatic circuits must be driven by quasi-trapezoidal supply waveforms (*i.e.*, having flat tops and bottoms), since the supply voltage presented to a device must remain stable while the device is being turned on or off. But many so-called "adiabatic" circuit designs in the literature invoke sinusoidal supply waveforms, ergo aren't truly, fully adiabatic. However, to supply a flat-topped waveform resonantly (with high-quality energy recovery) presents a significant RF engineering challenge.

But, since our goal in this project is to push the limits of energy efficiency in adiabatic circuits, we are tackling the challenges head-on. Therefore, our overall technical strategy is:

1) Implement truly, fully adiabatic (ergo reversible) CMOS logic families, obeying all of the adiabatic design rules.

2) Design & implement high-quality resonant oscillators generating multiple near-ideal, flat-topped, phase-locked quasi-trapezoidal waveforms to serve as the power-clock supplies.

3) Test the adiabatic logic, driven by the resonator, measure the system power dissipation, and further refine the design.

Further, we aim to characterize the energy efficiency gains achievable in our approach across a range of available CMOS processes, to better understand which processes will afford the best possible energy efficiency using our methods.

In parallel with this core engineering effort, the scope of our effort has expanded over the last year to also include two university-based collaborations on related topics:

1) Karpur Shukla of Brown University is applying theoretical tools from modern non-equilibrum quantum thermodynamics to derive fundamental, technology-independent lower limits on energy dissipation as a function of delay for general quantum systems performing classical reversible computations. This study may help us develop new insights into the design of innovative "Beyond CMOS" device technologies for reversible computing, so as to aspirationally achieve breakthrough reductions in energy-delay product, and take the cost-efficient performance of reversible computing far beyond the limits of any possible non-reversible technology.

2) Tom Conte and Anirudh Jain of Georgia Tech aim to carry out a study of processor architectures based on reversible technology, including an assessment of the potential system-level impact of hypothetical new reversible device technologies for computational science applications.

III. Progress to Date and Early Results

A. Logic Family, Simulations, and Test Chip Layouts.

As mentioned in the previous section, our approach is based on implementing *truly, fully* adiabatic CMOS logic styles. The first of these, historically, was Younis & Knight's Charge Recovery Logic (CRL) [8]. A number of generalizations and variations on the techniques innovated by CRL were developed in subsequent literature (*e.g.*, [9], [14], [15]). We selected for our study a version called 2LAL (two-level adiabatic logic) which is effectively just a slight generalization of CRL but was reinvented by one of us (Frank) in 2000. It is an attractive target for study because of its simple clock waveforms, which have the *shortest possible* period for a fully adiabatic sequential logic family (only 4 adiabatic transition times per clock cycle), and thus, 2LAL's trapezoidal waveforms are the *closest* to an ordinary sinusoidal wave, facilitating design of resonant oscillators.

In addition, Cadence/Spectre simulations conducted at the University of Florida in 2004 for a TSMC 180 nm process suggested that 2LAL could be *extremely* efficient, capable of dissipating as little 1 eV (0.16 aJ) per device per cycle. Our initial aim in the current project was to replicate that result using contemporary device models for similar processes available today. However, we do not know if the BSIM3 MOSFET models used in the 2004 study captured leakage of those devices adequately. Considering leakage led us to new ideas for minimizing leakage in future devices, which we will discuss in the next section.

In the meantime, we decided to update the 2LAL simulation results using newer models (*e.g.*, BSIM3SOI, BSIM4) for available processes. So far, our simulation efforts have focused on the in-house 350 nm and 180 nm processes provided by the MESA fabrication facility at Sandia and a 180 nm SOI process from GlobalFoundries (GF). In these simulations, a 2LAL shift register was simulated at a 50% activity factor. At 1 MHz, using a C_L = 10 fF model interconnect capacitance, Spectre simulations of the MESA processes suggested an energy dissipation per cycle per FET of 37 aJ (230 eV) for W = 800 nm wide devices in the 350 nm process, and 6.9 aJ (43 eV) for W = 480 nm wide devices in the 180 nm process, a value roughly consistent

with corresponding data points for TSMC18 from the UF study. Meanwhile, similar simulations of the GF process using ADS suggested a minimum energy dissipation of ~80 aJ (500 eV) per cycle per FET when V_{dd} is lowered to 1.2 V.

To further validate our simulation results, we decided to lay out, fabricate, and measure energy dissipation of an actual test chip. To our knowledge, no physical chips based on CRL/2LAL have been built,[1] so this seemed worthwhile. We laid out a 720-stage 2LAL shift register in Sandia's internal 180 nm process (Figs. 1 & 2), which was taped out to a shuttle run in Aug. 2020.

One issue encountered in our simulation work was that the *dynamic* nature of logic styles such as 2LAL (and the original CRL) creates difficulties for the correct and energy-efficient operation of these circuits. In general, floating nodes are vulnerable to voltage *drift* (due to leakage) and *sag* (due to capacitive voltage-divider effects involving parasitic couplings). If, due to drift/sag, a floating node voltage diverges by ΔV from its nominal level, then, upon reconnection of that node to its reference level, there will be a sudden, non-adiabatic energy dissipation of $C(\Delta V)^2/2$. Since voltage drift tends to increase linearly with time, this leads to a component of total per-cycle energy dissipation that increases *quadratically* with clock period. In addition, drift causes logical errors at sufficiently low frequencies, or in any node that remains in the floating state for sufficiently many cycles due to a low activity factor in that part of the circuit.

Confronting these issues makes plain that a *fully static*, fully adiabatic ("*perfectly adiabatic*") logic style is required to maximize energy efficiency in adiabatic circuits. This also facilitates testing, since, in a static style, one can even pause the clocks. In June 2020, we invented one such technique, called S2LAL (for static 2LAL), which will be presented in a separate paper.

B. Power-Clock Resonator Circuit

One of the key challenges being tackled in this project is the design and development of a high-quality resonant oscillator suited to drive a fully adiabatic circuit. Our initial target is the simple 4-phase trapezoidal waveform $\hat{\phi}_i$ in Fig. 1. Early concepts for resonantly generating trapezoidal waveforms [16] were based on combining subcircuits that resonate at respective frequency components of the desired signal. For example, the first six terms of the Fourier series for $\hat{\phi}_i$ are

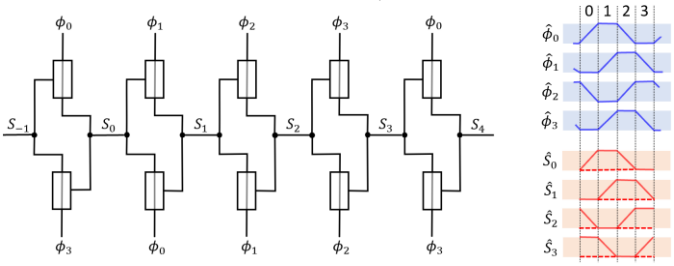

Fig. 1. Shift register schematic and timing sequence in the 2LAL logic style (a variant of CRL). Each signal x_i in the schematic is implicitly a dual-rail pair of complementary signals, an active-high signal denoted $\hat{x}_i$ and active-low $\check{x}_i$. Each rectangle denotes a parallel pair of CMOS transmission gates. A timing diagram for power-clocks $\hat{\phi}_i$ and data signals $\hat{S}_i$ is shown at right. This shift register transfers a single logical symbol (*e.g.*, logic `1`) and an identical, parallel structure would be needed to also transfer an alternate symbol (*e.g.*, logic `0`).

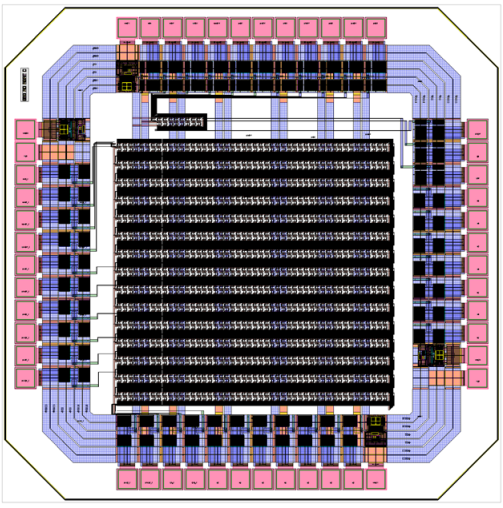

Fig. 2. Preliminary layout for the first 2LAL test chip, in Sandia's 180 nm process. At top (just inside the pad ring) is a small, 8-stage 2LAL shift register, and below it is a larger 720-stage shift register for power measurements.

$$v_{f6}(t) = V_{\mathrm{DD}}\left[\frac{1}{2} + \frac{4\sqrt{2}}{\pi^2}\left(\sin(\omega t) + \frac{\sin(3\omega t)}{3^2} - \frac{\sin(5\omega t)}{5^2} - \frac{\sin(7\omega t)}{7^2} + \frac{\sin(9\omega t)}{9^2} + \frac{\sin(11\omega t)}{11^2}\right)\right], \quad (1)$$

which together suffice to come within 2% of the ideal waveform, in theory. But, simple approaches such as those in [16] are insufficient to phase-lock the various sinusoidal components together and maintain an appropriate amplitude distribution. So, we devised an improved circuit technique that solves these problems. In simulation, Q values are ~3,000. Sample output waveforms after oscillator startup are shown in Fig. 3 below.

IV. Next Steps

Important upcoming steps for this project include:

1) Simulations, layouts and test chip fabrication for the newly invented fully static version of 2LAL (S2LAL).

2) Further development of the trapezoidal resonant oscillator circuit, including experimental verification that prototype implementations can drive the 2LAL test chip, and measurements of system power dissipation.

3) Exploration of additional available processes as implementation candidates, including MOSIS MPW offerings.

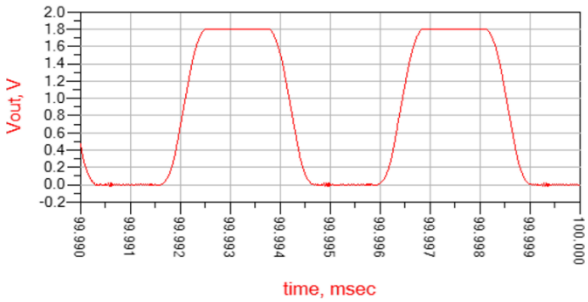

Fig. 3. Sample waveforms for Sandia's trapezoidal resonant oscillator circuit, sampled after a 1/10th-second startup period to allow the oscillator to stabilize. The operating frequency here is 230 kHz, but the same circuit design techniques should yield viable oscillator designs over a wide range of frequencies.

[1] In [14], Athas *et al.* described a shift register test chip for a logic style that is very similar to, but not exactly the same as, CRL/2LAL; also, that one was a much smaller design for a significantly older process.

In the longer term, to maximize the energy efficiency of adiabatic circuits, it will be crucial to reduce leakage. We see two potentially viable routes to this:

1) Use an older process. When properly optimized, *older* CMOS technologies can actually be *more* energy efficient than newer ones when perfectly adiabatic design styles are used and energy dissipation is aggressively minimized. This is because subthreshold leakage is much lower, since thresholds are higher, and also gate leakage is much lower, since gate oxides are thicker. These are both *exponential* effects, and thus override the *polynomial* factors in adiabatic minimum energy dissipation that result from increased C and V values.

2) Operate at cryogenic temperatures. Recent research (*e.g.*, [17]–[19]) shows that modern CMOS processes perform quite well at deep cryogenic temperatures (*e.g.*, 4 K, and perhaps lower), and moreover, subthreshold leakage is much lower than at room temperature (due to increased subthreshold slope). Due to this, device stackups can be reoptimized in several ways to dramatically reduce overall minimum energy dissipation at cryogenic temperatures, particularly when adiabatic switching is used. Although more research is needed, it is plausible that efficiency gains achievable via such methods could outweigh the substantial power overhead for cryogenic cooling ($\gtrsim$400× to go from 4 K to room temperature). Also, cryogenic operation raises the possibility of building ultra-high-Q *superconducting* resonators, or simply leaving the resonators at room temperature to reduce cooling overhead. Finally, superconducting *interconnects* offer another means to greatly improve energy efficiency of cryogenic CMOS-based systems with adiabatic switching.

V. Conclusion

Careful consideration of both the fundamental and practical limits on the levels of energy efficiency achievable by means of adiabatic circuit techniques makes it very clear that these limits are still very far from being reached. Our work is illuminating several clear and necessary steps along the path towards maximizing the energy efficiency of adiabatic CMOS:

1) Utilize truly, fully adiabatic reversible logic styles, and beyond that, fully static, "perfectly adiabatic" logic styles.

2) Develop high-Q trapezoidal resonators, which are perfectly suited for driving perfectly adiabatic logic circuits.

3) Minimize leakage aggressively, through e.g., cryogenic operation and associated reoptimization of the device design.

In the long term, as adiabatic CMOS becomes ever more efficient, there will be increasing demand for 3D fabrication processes that provide ever-more layers of active devices at an ever-lower cost per device, resulting in ever-greater levels of overall system cost efficiency, taking both manufacturing cost and power dissipation related costs into account.

Further, in the very long run, exploration of the fundamental quantum thermodynamic limits of reversible computing can lead to the discovery of completely new types of physical devices and mechanisms of operation that could lead to breakthrough levels of energy efficiency as a function of speed, perhaps many orders of magnitude beyond anything achievable with CMOS. Landauer's Principle assures us that achieving breakthroughs of such a magnitude will *require* reversible computing.

Acknowledgment

Thanks are due to Erik DeBenedictis to calling our attention to the recent line of research [17]–[19] on cryogenic CMOS.

References

[1] R. Landauer, "Irreversibility and heat generation in the computing process," IBM J. Res. Dev., vol. 5, no. 3, pp. 183–191, Jul. 1961. doi:10.1147/rd.53.0183

[2] M. Frank, "Physical foundations of Landauer's principle," in Reversible Computation, RC 2018, J. Kari and I. Ulidowski, Eds., Lecture Notes in Computer Science, vol. 11106 Cham: Springer, 2018, pp. 3–33. doi:10.1007/978-3-319-99498-7_1. See also: arXiv:1901.10327 [cs.ET]

[3] C. H. Bennett, "Logical reversibility of computation," IBM J. Res. Dev., vol. 17, no. 6, pp. 525–532, Nov. 1973. doi:10.1147/rd.176.0525

[4] C. H. Bennett, "The thermodynamics of computation—a review," Int. J. Theor. Phy., vol. 21, no. 12, pp. 905–940, Dec. 1982. doi:10.1007/BF02084158

[5] C. H. Bennett, "Notes on the history of reversible computation," IBM J. Res. Dev., vol. 32, no. 1, pp. 16–23, Jan. 1988. doi:10.1147/rd.321.0016

[6] K. Likharev, "Dynamics of some single flux quantum devices: I. Parametric quantron," IEEE Trans. Mag., vol. 13, no. 1, pp. 242–244, Jan. 1977. doi:10.1109/TMAG.1977.1059351

[7] E. F. Fredkin and T. Toffoli, "Design principles for achieving high-performance submicron digital technologies," proposal to DARPA, Nov. 1978; reprinted in Collision-Based Computing, A. Adamatzky, Ed. London: Springer, 2002, pp. 27–46. doi:10.1007/978-1-4471-0129-1_2

[8] S. G. Younis and T. F. Knight, Jr., "Practical implementation of charge recovering asymptotically zero power CMOS," in Research on Integrated Systems: Proc. 1993 Symp., C. Ebeling and G. Borriello, Eds. Cambridge: MIT Press, Feb. 1993, pp. 234–250.

[9] S. G. Younis, "Asymptotically Zero Energy Computing Using Split-Level Charge Recovery Logic," Ph.D. thesis, Massachusetts Institute of Technology, June 1994. https://dspace.mit.edu/handle/1721.1/11620

[10] M. P. Frank, "Foundations of generalized reversible computing," in Reversible Computation, RC 2017, I. Phillips and H. Rahaman, Eds., Lecture Notes in Computer Science, vol. 10301 Cham: Springer, 2017, pp. 19–34. doi:10.1007/978-3-319-59936-6_2. See also arXiv:1806.10183

[11] H. Thapliyal and M. Zwolinski, "Reversible logic to cryptographic hardware: a new paradigm," in 2006 49th IEEE Midwest Symposium on Circuits and Systems, vol. 1, IEEE, Aug. 2006, pp. 342–346. doi:10.1109/MWSCAS.2006.382067

[12] M. P. Frank, "Common Mistakes in Adiabatic Logic Design and How to Avoid Them," in Proc. Int'l. Conf. on Embedded Systems and Applications, ESA '03, CSREA Press, June 2003, pp. 216–222. http://revcomp.info/legacy/revcomp/MLPD03-Mistakes-paper.pdf

[13] M. P. Frank, "Reversibility for efficient computing," Ph.D. thesis, MIT, June 1999. https://dspace.mit.edu/ handle/1721.1/9464

[14] W. C. Athas, L. "J." Svensson, J. G. Koller, N. Tzartzanis, and E. Y.-C. Chou, "Low-Power Digital Systems Based on Adiabatic-Switching Principles," IEEE Trans. VLSI Sys., vol. 2, no. 4, pp. 398–407, Dec. 1994.

[15] V. Anantharam, M. He, K. Natarajan, H. Xie, and M. P. Frank, "Driving Fully-Adiabatic Logic Circuits Using Custom High-Q MEMS Resonators," in Proc. Int'l. Conf. on Embedded Systems and Applications, 2004, pp. 5–11. revcomp.info/legacy/revcomp/AdiaMEMS/MLPD-04.pdf

[16] S. G. Younis and T. F. Knight, "Non-dissipative rail drivers for adiabatic circuits," in Proc. Sixteenth Conf. on Advanced Research in VLSI, IEEE, Mar. 1995, pp. 404–414. doi:10.1109/ARVLSI.1995.515635

[17] A. Beckers, F. Jazaeri, and C. Enz, "Characterization and modeling of 28 nm bulk CMOS technology down to 4.2 K," IEEE J. Elec. Dev. Soc., vol. 6, pp. 1007–1018, Mar. 2018. doi:10.1109/JEDS.2018.2817458

[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. Elec. Dev. Soc., vol. 6, pp. 996–1006, Apr. 2018. doi:10.1109/JEDS.2018.2821763

[19] A. Beckers, F. Jazaeri, and C. Enz, "Cryogenic MOS transistor model," IEEE Trans. Elec. Dev., vol. 65, no. 9, pp. 3617–3625, Sep. 2018. doi:10.1109/TED.2018.2854701