IEEE/CSC & ESAS EUROPEAN SUPERCONDUCTIVITY NEWS FORUM, No. 22, October/November 2012.
The published version of this preprint appeared in IEEE Transactions on Applied Superconductivity 23, 1700504 (June 2013).

64-kb Hybrid Josephson-CMOS 4 Kelvin RAM with 400 ps Access Time and 12 mW Read Power

Theodore Van Duzer, *Life Fellow, IEEE,* Lizhen Zheng, *Member, IEEE,* Stephen R. Whiteley, *Member, IEEE,* Hoki Kim, Jaewoo Kim, Xiaofan Meng, and Thomas Ortlepp

***Abstract*—We have designed, simulated, fabricated and tested a 64-kb hybrid Josephson-CMOS memory using a 5 mm x 5 mm Josephson interface chip and a 2.0 mm x 1.5 mm CMOS chip. The Josephson chip uses the Hypres 4.5 kA/cm^2 niobium technology and the CMOS chip is made using the TSMC 65 nm technology. The chips are connected using short wire bonds in a piggy-back package. The chip sizes and pad layouts have been constrained to allow testing in our wideband American Cryoprobe Model BCP-2 test probe in order to measure ultra-short delays. The test signals of 5 mV amplitude are chosen to represent the signals that would be supplied to the memory in a digital computing or signal processing system. Each input signal is first amplified in a four-junction logic (4JL) gate driving a Suzuki stack, which, in turn, drives a highly sensitive CMOS comparator that raises the signal to volt level. Such amplifiers are provided for the address, data, read and write inputs to the CMOS memory. Output currents from the memory cells are detected by ultra-fast 4JL gates providing 5 mV output signals; an equivalent arrangement was used for the delay tests. The overall read delay is the access time, which we find to be about 400 ps. We extrapolate from the measured and calculated power dissipation in this partially accessed 64-kb memory to a fully accessed 64-kb memory and find the expected overall read power dissipation to be about 12 mW for operation at 1 GHz.**

***Index Terms*— Hybrid memory, access time, high-speed interface, 4 K CMOS**

I. Introduction

The lack of a fast memory with adequate capacity has been a long-standing problem in superconductor digital technology [1]. Since the field started in the late 1960s, there have been a number of projects aimed at making various types of digital systems, especially computation. Famously, IBM had a very large project that was ultimately abandoned in 1983 because there was no evident way to provide memory on the scale needed [2]. Work toward a digital computer in Japan that started before 1983 continued with improved fabrication technology. Several memory ideas were tried. Almost all were based on the idea of persistent current in a superconductor loop to store a magnetic flux quantum. The largest fully functional 4 K memory had a capacity of 4 kb and was published in 1995 [3]. The memory cells in such wholly superconductor memories are rather large in order to store a flux quantum. Potentially, improved foundries could ameliorate this situation. Also, ideas of memory cells using magnetic material to shrink the cells are being studied.

We have taken a radically different approach [4]-[7] in which we take advantage of the ever improving complementary metal oxide semiconductor (CMOS) technology. There are large numbers of designers and excellent foundries for CMOS technology and a commercially driven motivation for improvement.

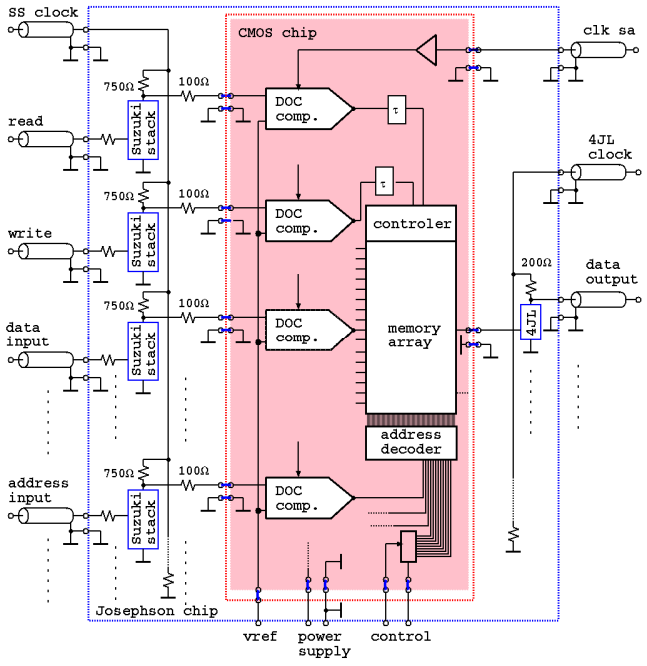

Fig. 1. Block diagram of the hybrid Josephson-CMOS RAM system. On the CMOS chip are the memory core and comparators for memory input amplification. On the Josephson chip are the Suzuki stacks as input signal preamplifiers and 4JL gates as memory current sensors. The 4 K hybrid memory with all 5 mV IOs are compatible with superconductor electronics.

Fig. 1 is a block diagram of the hybrid Josephson-CMOS random access memory (RAM) system. The core of our system is a 64-kb CMOS static random access memory (SRAM) which is designed professionally in 65 nm CMOS technology to achieve minimum delay and power dissipation, taking account of the 4 K properties of CMOS devices and circuits. There is a high level of perfection in fabrication so

Manuscript received October 9, 2012, revised November 15, 2012. This work was supported by IARPA under ARO W911NF-10-1-0120.

The authors are with the EECS Dept., University of California, Berkeley, CA 94720, USA, email: vanduzer@eecs.berkeley.edu.

The published version of this preprint appeared in IEEE Transactions on Applied Superconductivity 23, 1700504 (June 2013).

one can expect that if one cell works, all cells will work. This has allowed us to make a 64-kb memory and will facilitate scaling to larger capacities.

A key part of our memory system is the amplification of the millivolt superconductor logic signals to the volt level signals needed by the CMOS memory. This is done in two stages. The millivolt superconductor logic signals are amplified to 60 mV using a Suzuki stack (SS) [8] with a four-junction logic (4JL) gate [9]. This drives a very sensitive CMOS comparator to produce the required volt level signals. The selected memory cell provides an output current that drives a superconductor detector.

The main design aspects for this project are memory capacity, read access time and read power consumption. We present the results of simulations and measurements on a fully functional 64-kb memory operating at 4 K as well as values of access time and power dissipation. This 64-kb memory is the first fully functional 4 K memory of capacity beyond 4 kb.

II. 4 K CMOS Devices

Our CMOS chips were fabricated in TSMC 65 nm bulk CMOS Mixed signal RF General Purpose Plus Low K CU process. [10] Device test chips including individual transistors, metal wires, resistors, and ring oscillators were fabricated and measured to characterize the 4 K CMOS devices. We observed a four times reduction in metal wire resistance compared with room temperature (RT). This results in a significant reduction of the RC delay of the long wires in the memory layout at 4 K. The transistor threshold voltage is increased by about 0.2 V. Both IV characteristic and ring oscillator output jitter measurement confirm this fact. The device leakage current is reduced by at least three orders of magnitude. In ring oscillator measurements, we observed that the output frequency increased by 17% to 30% at 4 K versus RT. Fig. 2 shows the single stage inverter delay extracted from a 409-stage ring oscillator frequency measurement at 4 K and RT.

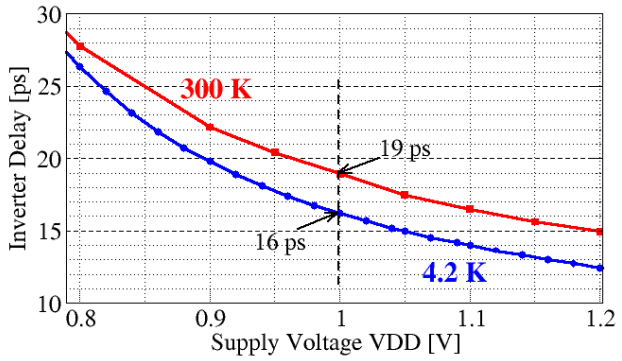

Fig. 2. Comparison of single-stage 65 nm inverter delay versus supply voltage at 4 K and RT.

To compensate for the threshold voltage increase at 4 K, we chose TSMC's "low threshold transistors" for our design. And due to the significantly reduced leakage current, we are able to choose the high-performance CMOS process instead of the much slower low leakage CMOS process to achieve short memory access time.

III. Hybrid Interface Circuits

A Suzuki stack serves as preamplifier for the following CMOS comparator in the system. Suzuki stacks [8] with 40 mV, 60 mV and 80 mV output and 4JL gate drivers [9] have been designed and verified substantially [11]. A clocked CMOS comparator [12] decides if the SS output voltage is above or below a given reference voltage, which is typically at the midpoint of the input signal. It regenerates the 1V full swing digital signal for the following CMOS memory. A level shifter between the SS and the CMOS comparator shifts near-ground SS output signal level to the input voltage level required by the comparator.

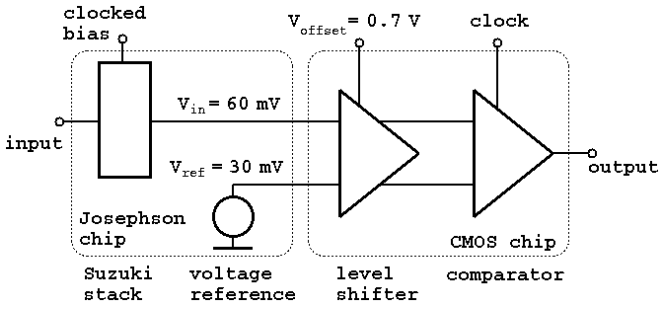

Fig. 3. Block diagram of the hybrid interface circuit.

Fig. 3 shows the block diagram of the hybrid interface. 60 mV SS is employed in the system for the best compromise of power and interface performance. The measured delay of the 60 mV SS with RC loading is 47 ps. The measured power consumption of one 60 mV SS is 165 µW in the system configuration. The main performance figures of the CMOS comparators are measured delay of about 170 ps including the level shifter, the measured uncertainty range in the decision (or gray zone) of 2-6 mV and the calculated power consumption of about 460 µW [13]. The interface circuit was extensively tested and it provides a measured system bit-error-rate below 10^{-12}.

IV. CMOS Memory

The complete 64-kb CMOS SRAM is composed of memory array, block and row decoders, global and local word line drivers, control logic and output multiplexer. It has 12-bit address inputs, separate read and write command inputs, 16-bit data inputs, and 16-bit data outputs.

Fig. 4 shows the layout of our CMOS SRAM which is optimized for low power consumption and short read access. In the center is the main decoder and global word line drivers. The memory is divided into four 16-kb (128 word lines x 128 bit lines) macros. Each macro has its own global world line drivers. Each macro is further divided into eight 2-kb (128 word lines x 16 bit lines) micro array blocks. Each block has its own control logic for block selection, input and output data multiplexing, local word line decoder and local world line driver. For each read/write operation, only one 2-kb block is

The published version of this preprint appeared in IEEE Transactions on Applied Superconductivity 23, 1700504 (June 2013).

active, 16 memory cells on the selected row in the block are activated. The critical path of the read access is modeled in spice simulation considering the 4 K device characteristics to achieve optimized power and delay. The total read access time is 183 ps in simulation and power is 2.6 mW in simulation.

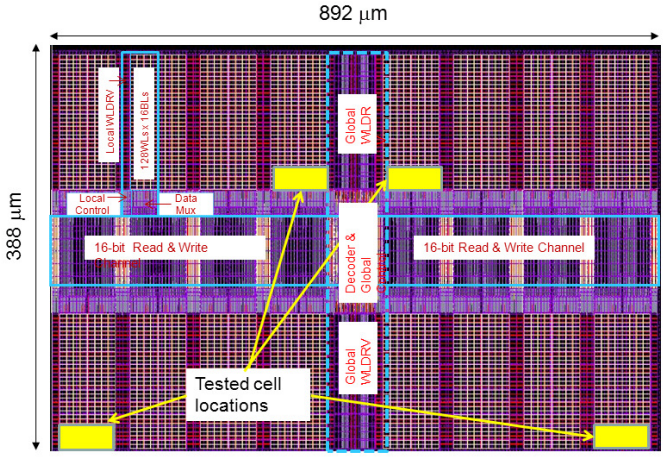

Fig. 4. Layout of the 64-kb CMOS SRAM. The center components include the decoder, global word line drivers (WLDRV) and the data bus. The rest are the memory arrays with local control and local WLDRV.

Fig. 5 (a) shows the schematic of the 8-transistor memory cell in this project. This memory cell read and write are separate, providing wide power supply (VDD) margin. The four cross-coupled transistors in the center form a bi-stable flip-flop, representing state "0" or "1". To write, the write word line enable signal WWL turns high and the complementary input data WBL/WBLB set the memory cell state. When WBL is "1", the inputs set the memory state to "1"; node X remains high even after WWL is disabled. The stored state can be read out by enabling the read word line RWL. For V(X) = "1", an output current I_{out} flows to the read bit line RBL. When V(X) = 0, no I_{out} is generated. Fig. 5 (b) shows the layout. The memory size is 1.4 µm x 2.4 µm, which is about 1000 times smaller than a Josephson memory cell [14]. This CMOS memory cell layout is not the minimum size due to the extra well and substrate contacts inserted to prevent transistor threshold voltage hysteresis as a precaution. It is confirmed experimentally that such hysteresis effect does not exist in 65 nm TSMC process we are using. So the memory size can be further reduced and the related power and access time will also be reduced, especially for a large capacity memory such as 1-Mb or more.

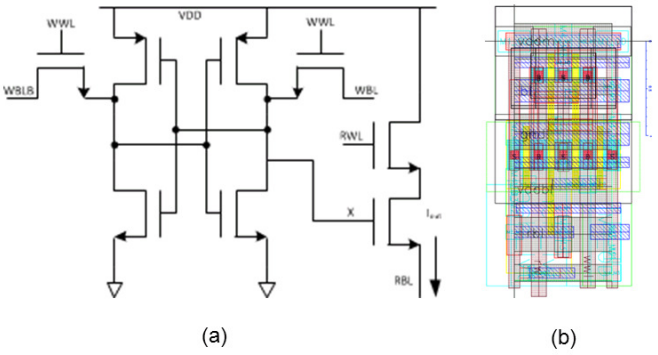

Fig. 5. (a) Schematics and (b) layout of a 8-transistor memory cell. The memory cell size is 1.4 µm x 2.4 µm.

V. System Measurement

Our hybrid memory consists of a 5 mm x 5 mm Hypres 4.5 kA/cm^2 chip and a piggy-backed 2 mm x 1.5 mm TSMC 65 nm CMOS chip. The two chips are connected by short wire bonds (550 µm long). Fig. 6 shows a photograph of the chip assembly. Preamplifier Suzuki stacks and the current sensing 4JL gates are marked on the Josephson chip. The CMOS comparators and the 900 µm x 400 µm CMOS memory core are marked on the CMOS chip. Beside input signals lines, we have separate power supplies for the Josephson circuits and the CMOS circuits. All interface amplifiers receive a high-speed external clock signal. We generate the clock and all input signals with a commercial HP 70843A error performance analyzer and an HP data generator system E2903A. The data outputs as well as the timing reference signal are observed using a Tektronix 11801A sampling oscilloscope. All input and output signals of the hybrid memory are signals with 5 mV amplitude.

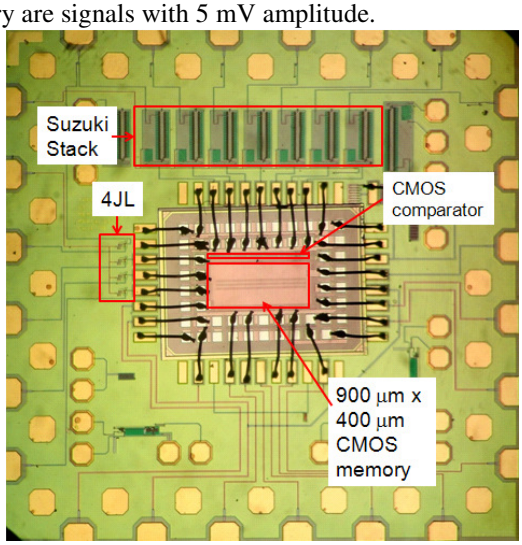

Fig. 6. Micrograph of a wire-bonded 64-kb hybrid memory.

The published version of this preprint appeared in IEEE Transactions on Applied Superconductivity 23, 1700504 (June 2013).

For functionality testing, we apply a sequence of write and read signals. Various bit combinations are written to different memory locations. We intentionally change addresses each cycle and our memory features single cycle read and write operation. Fig. 7 shows an example of the test pattern used for the functional testing. We tested 32 different bit cell locations including the cells with fastest and slowest read access time. In all functionality tests, the hybrid memory operates correctly. The tested chips are not pre-selected.

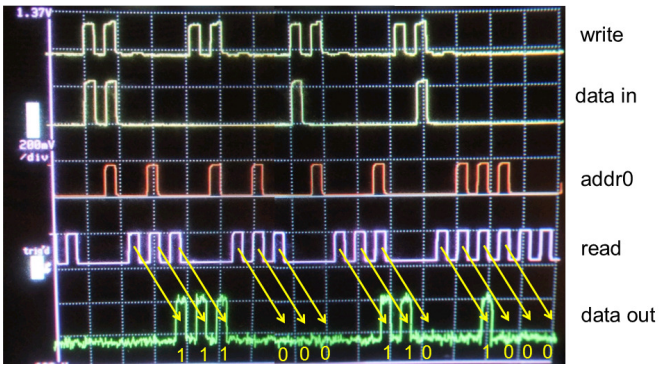

Fig. 7. Waveforms of one correct operation of the hybrid memory.

Fig. 8 shows the test scheme for the memory access time measurement. The read signal (RD) is split on the Josephson chip to generate a timing reference signal. The reference signal as well as the data output are passing 4JL gates to generate identical output waveforms. The delay from RD_Mon to DR0 is the memory read access time (not

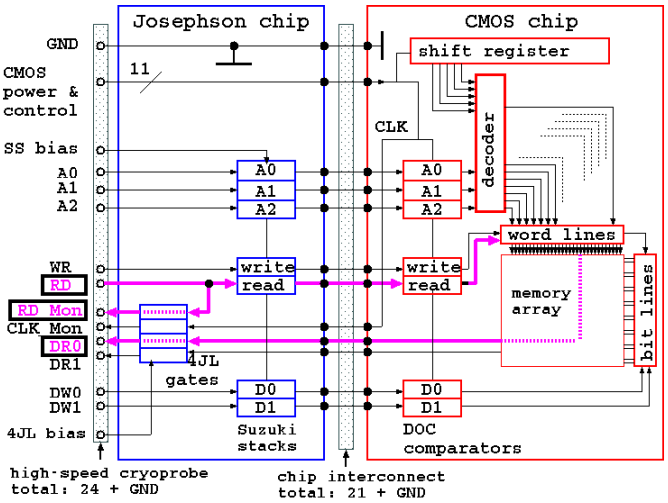

Fig. 8. Block diagram for read access time test scheme. Delay from RD_Mon to DR0 is the system read access time.

With supply voltage VDD=1.0 V chosen for the memory, the read access time ranged from 390 ps to 430 ps among the 32 cells tested. With 10% VDD increase from 1 V to 1.1 V, a 20% power increase, the access time can be reduced from 390 ps to 320 ps as shown in Fig. 9. This is our shortest measured access time. The delays of the CMOS memory components contributing to the total read delay are simulated and listed in table I. Compared to read delay, write delay includes the same decode and word line delay. Write delay is expected to be a few tens of ps shorter than the read access time since no read bit line charging and 4JL read out are involved.

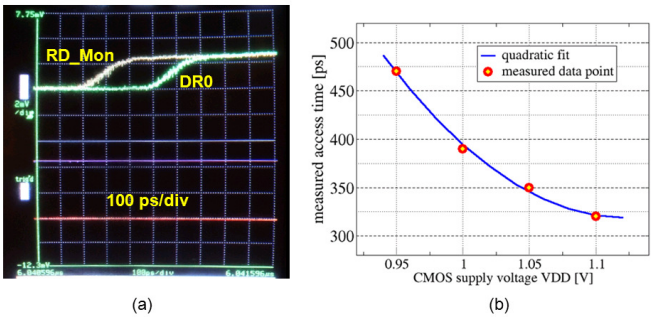

Fig. 9. (a) Measurement of 320 ps access time. (b) Access time dependence on supply voltage VDD.

We also measured the system power consumption of our current partially accessed memory system on the test chip. It contains only seven hybrid interface channels out of total 30 input channels due to high-speed test limitation. We then calculate the power consumption of a fully accessible 64-kb hybrid memory. We summarize the measured and calculated read delay and read power consumption of a completed 64-kb RAM system and its components at 1 GHz with 1 V supply voltage in Table I. For the read power calculation, we count only 14 hybrid interface channels (12 addresses, read and write). For the write power, we need to add 16 input data hybrid interface channels but no bit line output current. The total write power estimated for the 64-kb RAM system is about 21 mW. The power consumption values of the CMOS memory components in Table I are obtained from simulations.

From Table I, we can see for the 64-kb hybrid RAM, the overhead of power and delay from the hybrid input interface is considerable. However, the percentage of overhead will be reduced with increased RAM capacity. Only 4 additional address lines are needed to expand a 64-kb RAM to a 1-Mb RAM.

TABLE I
SUMMARY OF READ DELAY AND READ POWER AT 1 GHZ OF THE 64-KB HYBRID RAM SYSTEM AND COMPONENTS

Circuit Components	Delay (ps)	Read Power (mW)
Suzuki stacks (x 14)	47	2.3
CMOS comparators (x 14)	167	6.4
CMOS decoder + Word lines	135	2.47
Memory cells + Bit lines (x 16)	48	0.13
4JLs (x 16)	10	0.448
Total	407	11.88

VI. CONCLUSION

We have developed and successfully demonstrated a 64-kb hybrid Josephson-CMOS 4 K memory. The key components are high-speed low-power hybrid interfaces and a 64-kb CMOS memory core optimized for 4 K operation. This

memory system is fully compatible with Josephson electronics and provides a read access time of 390-430 ps with only 12 mW read power consumption and 21 mW write power at 1 GHz. The majority of the power consumption comes from the interface circuits. This makes the implementation of larger memories even more attractive and scaling to 1-Mb appears straight forward.

This 64-kb memory is the first fully functional Josephson compatible 4 K memory of capacity beyond 4 kb. The hybrid Josephson-CMOS 4 K memory is an attractive near term solution for the Josephson memory needs in high performance computing [1], [15].

ACKNOWLEDGMENT

This work was supported by IARPA under ARO W911NF-10-1-0120.

The authors are very appreciative of the support of the TSMC University Shuttle Program and are very thankful to Sang Dhong and Osamu Takahashi of TSMC for making the shuttle runs possible. We could not have achieved the results stated in this paper without their very generous support.

We thank Elad Alon, Lingkai Kong, and Yue Yang for very helpful discussions and for contributing their DOC comparator design for our evaluation.

REFERENCES

[1] F. Bedard, N. Welker, G. R. Cotter, M. A. Escavage, and J. T. Pinkston, "Superconducting Technology Assessment," *National Security Agency, Office of Corporate Assessments*, 2005.

[2] S. M. Faris, W.H. Henkels, E.A. Valsamakis, and H.H. Zappe, "Basic design of a Josephson technology cache memory," *IBM J. Res. Dev.*, vol. 24**,** pp. 143–154, 1980.

[3] S. Nagasawa, Y. Hashimoto, H. Numata, and S. Tahara, "A 380 ps, 9.5 mW Josephson 4-Kbit RAM Operated at a High Bit Yield," *IEEE Trans. Appl.Supercond.***,** vol.5, no. 2, pp. 2447-2452, 1995.

[4] Q. Liu, T. Van Duzer, X. Meng, S. R. Whiteley, K. Fujiwara, T. Tomida, K. Tokuda and N. Yoshikawa, "Simulation and measurements on a 64-kbit hybrid Josephson-CMOS memory," *IEEE Trans. Appl. Supercond.,* vol. 15, no. 2, pp. 415-418, Jun. 2005.

[5] N. Yoshikawa, T. Tomida, M. Tokuda, Q. Liu, X. Meng, S. R. Whiteley, and T. Van Duzer, "Characterization of 4 K CMOS Devices and Circuits for Hybrid Josephson-CMOS Systems," *IEEE Trans. Appl. Supercond.,* vol. 15, no. 2, pp. 267-271, Jun. 2005.

[6] Q. Liu, K Fujiwara, X. Meng, S. R. Whiteley, T. Van Duzer, N. Yoshikawa, Y. Thakahashi, T. Hikida and N. Kawai, "Latency and power measurements on a 64-kb hybrid Josephson-CMOS memory," *IEEE Trans. Appl. Supercond.,* vol. 17, no. 2, pp. 526-529, Jun. 2007.

[7] K. Fujiwara, Q. Liu, T. Van Duzer, X. Meng, and N. Yoshikawa, "New Delay-Time Measurements on a 64-kb Josephson-CMOS Hybrid Memory With a 600-ps Access Time," *IEEE Trans. Appl. Supercond.,* vol. 20, no. 1, pp. 14-19, Feb. 2010.

[8] H. Suzuki, A. Inoue, T. Imamura, and S. Hasuo, "A Josephson driver to interface Josephson junctions to semiconductor transistors," in *IEDM Tech. Dig.*, 1988, pp. 290–293.

[9] H. Nakagawa, E. Sogawa, and S. Kosaka, "Operating Characteristics of Josephson Four-Junction Logic (4JL) Gate," *Japanese Journal of Applied Physics*, vol. 21, no. 4, pp. L198-L200, 1982.

[10] www.tsmc.com/english/dedicatedFoundry/technology/65nm.htm

[11] T. Ortlepp, L. Zheng, S. Whiteley, and T. Van Duzer, "Design guidelines for Suzuki stacks as reliable high-speed Josephson voltage drivers," submitted for publication.

[12] L. Kong, Y. Lu, and E. Alon, "A multi-GHz area-efficient comparator with dynamic offset cancellation," in *Custom Integrated Circuits Conference (CICC),* 2011, pp. 1–4.

[13] T. Ortlepp, Stephen R. Whiteley, L. Zheng and T. Van Duzer, "High-speed hybrid superconductor-to-semiconductor interface circuit with ultra-low power consumption," *IEEE Trans. Appl. Supercond.,*vol 23, Digital Object Identifier 10.1109/TASC.2012.2227918, 2013.

[14] S. Nagasawa, H. Hasegawa, T. Hashimoto, H. Suzuki, K. Miyahara, and Y. Enomoto, "Superconducting Latching/SFQ Hybrid RAM," *IEEE Trans. Appl. Supercond.,* vol. 11, no. 1, pp. 533-536, Mar. 2001.

[15] D. S. Holmes, A. L. Ripple, M. A. Manheimer, "Energy-Efficient superconducting computing – power budgets and requirements," *IEEE Trans. Appl. Supercond.,***23** submitted for publication.