Hybrid Josephson-CMOS Random Access Memory with Interfacing to Josephson Digital Circuits

The work during this period brought our efforts to realize a 64-kbit hybrid Josephson-CMOS memory for 4K operation to a successful conclusion. The memory inputs and outputs are at the several-millivolt level. The inputs are first amplified to 60 mV using Suzuki stacks; these signals drive CMOS comparators, the outputs of which are at volt level, as required by the commercially manufactured 65 nm CMOS memory array. Final designs were fabricated and tested. The access time is less than 400 ps and the read power dissipation is less than 10 mW. A side
Hybrid Josephson-CMOS Random Access Memory with Interfacing to Josephson Digital Circuits

ABSTRACT
The work during this period brought our efforts to realize a 64-kbit hybrid Josephson-CMOS memory for 4K operation to a successful conclusion. The memory inputs and outputs are at the several-millivolt level. The inputs are first amplified to 60 mV using Suzuki stacks; these signals drive CMOS comparators, the outputs of which are at volt level, as required by the commercially manufactured 65 nm CMOS memory array. Final designs were fabricated and tested. The access time is less than 400 ps and the read power dissipation is less than 10 mW. A side project was performed to evaluate the concept for a novel wholly Josephson memory proposed by a Japanese memory expert. We found that the proposed memory cell has inadequate design margins. We realized the other goal of the project, which was to analyze the factors limiting Josephson memories. Presentations for the Applied Superconductivity Conference (ASC) in October 2012 were prepared and presented. Publications resulted. We also brought together an international group of 30 cryogenic memory researchers in a post-ASC workshop at UC Berkeley for more discussions. And presentations were made at the Superconductive Electronics Conference (ISEC) in July 2013.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received  Paper


08/27/2013  5.00 Lizhen Zheng, Stephen Whiteley, Hoki Kim, Jaewoo Kim, Xiaofan Meng, Thomas Ortlepp, Theodore Van Duzer. 64-kb Hybrid Josephson-CMOS 4 Kelvin RAM With400 ps Access Time and 12 mW Read Power, IEEE Transactions on Applied Superconductivity, (06 2013): 1400104. doi:


TOTAL:  4

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)
Number of Papers published in non peer-reviewed journals:

(c) Presentations

"Multiplexer for CMOS-to-Josephson interfaces"
Lizhen Zheng, Stephen R. Whiteley, Hoki Kim, Daniel Wei, Thomas Ortlepp and Theodore Van Duzer, ISEC, Washington, DC 2011


The following talks were given at the Post-ASC Memory Workshop:
1. "Introduction and Architecture Considerations," T. Van Duzer

"Power dissipation and access time of a 256-bit Josephson random access memory based on vortex transitional memory cells," T. Ortlepp and T. Van Duzer, ISEC, July 7-11,2013, Cambridge, MA

Number of Presentations: 7.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

TOTAL:
Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received  Paper

08/22/2013  2.00 Thomas Ortlepp. Vortex transitional superconductor random access memory (VTS-RAM), NOT published Material presented in ISEC'13 ( )

TOTAL:  1

Number of Manuscripts:

Books

Received  Paper

TOTAL:

Patents Submitted

Patents Awarded

Awards

0

Graduate Students

<table>
<thead>
<tr>
<th>NAME</th>
<th>PERCENT SUPPORTED</th>
<th>Discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniel Wei</td>
<td>0.50</td>
<td></td>
</tr>
</tbody>
</table>

FTE Equivalent: 0.50

Total Number: 1

Names of Post Doctorates
Names of Faculty Supported

<table>
<thead>
<tr>
<th>NAME</th>
<th>PERCENT_SUPPORTED</th>
<th>National Academy Member</th>
<th>FTE Equivalent</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theodore Van Duzer</td>
<td>0.10</td>
<td>Yes</td>
<td>0.10</td>
<td>1</td>
</tr>
</tbody>
</table>

Names of Under Graduate students supported

<table>
<thead>
<tr>
<th>NAME</th>
<th>PERCENT_SUPPORTED</th>
<th>FTE Equivalent</th>
<th>Total Number</th>
</tr>
</thead>
</table>

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period.

- The number of undergraduates funded by this agreement who graduated during this period: 0.00
- The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields: 0.00
- The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields: 0.00
- Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale): 0.00
- Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering: 0.00
- The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense: 0.00
- The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

<table>
<thead>
<tr>
<th>NAME</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniel Wei</td>
<td>1</td>
</tr>
</tbody>
</table>

Names of personnel receiving PHDs

<table>
<thead>
<tr>
<th>NAME</th>
<th>Total Number</th>
</tr>
</thead>
</table>

Names of other research staff
<table>
<thead>
<tr>
<th>NAME</th>
<th>PERCENT_SUPPORTED</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lizhen Zheng</td>
<td>0.50</td>
</tr>
<tr>
<td>Stephen Whiteley</td>
<td>0.10</td>
</tr>
<tr>
<td>Hoki Kim</td>
<td>0.10</td>
</tr>
<tr>
<td>Xiaofan Meng</td>
<td>0.20</td>
</tr>
</tbody>
</table>

**FTE Equivalent:** 0.90

**Total Number:** 4

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress
STATEMENT OF THE PROBLEM STUDIED

The lack of a fast memory with adequate capacity has been a long-standing problem in superconductor digital technology. 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 [1]. 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 only 4 kb and was published in 1995 [2]. 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 somewhat. Also, ideas of memory cells using ferromagnetic material to shrink the cells are being studied.

We have taken a radically different approach [3] 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 (see Attachment for figures) 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 array (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 by TSMC so one can expect that if one cell works, almost all cells will work. (We tested 32 cells with nearly identical results.) This has allowed us to make a 64-kb memory and will facilitate scaling to larger capacities.

STATE OF THE HYBRID MEMORY PROJECT AT THE START IN 2010

The first stage of the hybrid memory is the Josephson amplifier using stacks of Josephson junctions called “Suzuki stacks” (Fig. 2) to raise the signal level from a few millivolts to several tens of millivolts.[4] The concept had been in the literature since 1988 but implementations had been plagued with multi-level outputs and narrow margins. We started research into the causes and solutions for these problems that extended through the first two years of the project and ended with a robust design procedure for Suzuki stacks that has been published. [5]

The second stage is a CMOS amplifier/comparator the output of which is at volt level as needed to drive the memory array. Earlier years of the project attempted to use the circuits proposed by the original paper. For a variety of reasons, we had been unable to achieve robust behavior. At the beginning of the current grant period, we chose to use a comparator circuit similar to the sense amplifiers used in CMOS memories to detect output voltages. Over the following years several different circuits were considered including one with dynamic offset cancellation that was designed for room temperature operation at high frequencies up to 4 GHz.[6] This was developed by another group in our department; therefore, we could get convenient access to the layout and to the designers.

Our earlier CMOS technologies for the project had been 250 nm and 180 nm; now we were moving to 65 nm under the TSMC shuttle program. Experience with the 180 nm circuits showed that, because of the frozen-out substrate, charging occurred under the circuits leading to unstable circuit performance. Therefore, we incorporated some layout procedures used in similar SOI circuits to avoid the charging problems. This turned out for unknown reasons to be unnecessary in the TSMC 65 nm circuits.

In our earlier work with 250 nm and 180 nm CMOS, we could use compact 3T and 4T dynamic memory cells to store charge because they did not leak. The 65 nm cells do leak so we switched to an 8T SRAM cell. Such cells have received so much attention that they have evolved into very compact components and memories based on them are physically small. The memory output must be multiplexed into a 16-bit wide word for transmission to the processor. In the earlier CMOS generations, the CMOS multiplexer was very slow so we devised a superconducting multiplexer. In considering applying this circuit to 65 nm technology for a 64 kb memory, we found that it was not faster than the CMOS multiplexer so it was abandoned. The superconducting multiplexer may find use in memories of much larger capacities but this has not been evaluated.

RESEARCH ON THE HYBRID MEMORY COMPONENTS UNDER THIS GRANT

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 the Josephson-based Suzuki stacks with a four-junction logic (4JL) Josephson input gate [7]. Before this grant started, there was no clear design strategy for making Suzuki stacks with robust behavior. There were issues with narrow margins and multi-level switching.
The schematic of the basic Suzuki stack is shown in Fig. 2 in the Attachment. In the memory application there are several parallel stacks, one for each bit in the input word. It was challenging to power the stacks without the switching of one affecting the margins of the others and with minimum power dissipation. We also did an electromagnetic study to see if the switching of one in the set could produce an electromagnetic pulse that could switch another stack in the set. (It couldn’t.) We studied the effect of the input circuit and found the use of a 4JL logic gate provided a good buffer input. The nature of the output circuit of the Suzuki stack is critical; done incorrectly there can be a feedback into the stack causing incorrect switching. All these factors affect the design margins of the stack. We designed and verified Suzuki stacks with 40 mV, 60 mV and 80 mV output and 4JL gate drivers. Our design rules are

• critical current and number of junctions depends on the desired output current; \( I_c \) should be about twice the desired output current,
• proper isolation between 16 individual stacks requires a supply voltage which is about ten times the output voltage,
• the ground-plane under the SS must be removed to enable very small ground capacitance,
• the layout needs to be designed to minimize the load capacitance on the SS,
• current overshoot during switching is prevented by proper design of the load network.

We chose the 60 mV stacks (24 junctions in each leg of the stack) to use in the system to assure robust switching of the following CMOS comparators. The details of this work were published. [5]

A clocked CMOS comparator decides if the stack output voltage is above (logic 1) or below (logic 0) a given reference voltage, which is typically at the midpoint of the input signal. The comparator output is a 1 V full swing digital signal for the following CMOS memory. A level shifter between the stack and the CMOS comparator shifts near-ground stack output signal level to the input voltage level required by the comparator. Fig. 3 shows the block diagram of the hybrid interface. A 60 mV stack is employed in the system for the best compromise of power and interface performance. The measured delay of the 60 mV stack with RC loading is 47 ps. The best interface performance was achieved using a dynamic offset comparator that was designed for room temperature operation [6] but which worked well at 4 K. Research on an improved version of the dynamic offset comparator aimed at reducing the power dissipation were carried out, but the timing of the last TSMC run did not allow its evaluation.

CMOS MEMORY ARCHITECTURE

The 64kb (256 word lines x 256 bit lines) CMOS memory is organized as a 4 x 16k memory block and provides 16 bit data access. Each 16k memory block (128 Word Line x 128 Bit Line) is divided into 8 x 2k micro memory array (128 word lines x 16 bit lines). The micro array structure is the basic memory array form and only one array (1/32) is accessed (Fig. 4).

To achieve low active power, a partial block decoding scheme and global word line scheme are adopted. A global word line for each 16k memory block drives 8 micro arrays and the local wordline of one selected micro array is activated. To reduce loading of the predecoded address, the number of global word line drivers is reduced from 128 to 64. Instead, another predecoded address line doubling as a local word line is used at each micro array (Fig. 5).

Each micro array has a local word line driver and read/write circuit. One activated micro array sends/receives 16 bit data. 16 bit data from each micro array are connected to 16 common data bus and multiplexed by block decoding address (Fig 6). The micro array is composed of a 128 x16 memory cell array, read word line driver, write word line driver and local read/write column mux. The decoded signals (bsel_r/w_e/o) are combined with global word line enabled read word line and write word line.

The layout of the memory showing the locations of the tested cells is in Fig. 7.

BIT-LINE READ-OUT CIRCUIT

At the output of the memory array, the selected memory cell provides a current that drives a superconductor detector for which we use a 4JL logic gate [7]. There are 16 of these detectors powered in parallel; this requires careful design to avoid interaction through the powering circuit while minimizing the power.

The main design criteria for this project are memory capacity, read access time and read power consumption. Although the overall hybrid concept appears straight-forward, achieving stable, robust operation with acceptable access time and power dissipation required a high level of design expertise. The complicating factors are: achieving stable, robust operation of the Josephson array amplifiers (Suzuki stacks) while minimizing power dissipation and maximizing speed; designing the CMOS comparators (amplifiers) with high sensitivity, minimum delay, low noise, and low operating power; designing the CMOS memory array for minimum delay and power dissipation, and providing reliable fast, low-power detectors to amplify the data.
currents from the memory cells.

TECHNOLOGIES

The Josephson chips were fabricated by Hypres in their 4.5 kA/cm² four metal layer niobium process. The chip size is 5 mm x 5 mm to fit our high frequency American Superconductor test probe.

Our CMOS chips were fabricated in TSMC 65 nm Bulk CMOS Mixed-Signal RF General Purpose Plus Low K CU process. 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-fold reduction in metal wire resistance compared with room temperature. 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 measurements 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. 8 shows the single-stage inverter delay extracted from a 409-stage ring oscillator frequency measurement at 4 K and RT versus supply voltage. 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 in order to achieve short memory access time.

The chips are wire-bonded together with 0.55 mm gold wires. An effort was made to bump bond the chips together in order to minimize the impedance of the connections. It was ultimately defeated by the metallurgy. The Hypres chips are provided with suitable metal layers on the pads to allow for low temperature solder bonding. However, our TSMC chips come from a wafer used for many different projects so it was not possible to get the pads on our chips coated by them with metal layers suitable for low-temperature solder. We tried a scheme to coat the pads on individual chips but it is too difficult to get reliable results. In the end we used wire bonds; for the very short bonds we used, the bonds have no observable effect up to the 1 GHz of our measurements.

SUMMARY OF THE MOST IMPORTANT RESULTS

Our hybrid memory consists of a 5 mm × 5 mm Hypres 4.5 kA/cm² chip and a piggy-backed 2 mm × 1.5 mm TSMC 65 nm CMOS chip. The two chips are connected by short (550 µm) wire bonds. Fig. 9 shows the chip assembly. Preamplifier Suzuki stacks and the current sensing 4JL gates are seen on the Josephson chip. The CMOS comparators and the 900 µm × 400 µm CMOS memory core are seen on the CMOS chip. In addition to the input signal 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 generated 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.

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. We tested 32 different bit cell locations including the cells with fastest and slowest read access time. (Fig.7) In all functionality tests, the hybrid memory operates correctly. The tested chips are not pre-selected.

Fig. 10 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 pass through 4JL gates to generate identical output waveforms. The delay from RD_Mon to DR0 is the memory read access time (not including the 4JL gate delay, which is a few picoseconds). 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. This is our shortest measured access time. The delays of the components of the CMOS memory core and the 4JL gate contributing to the total read delay are simulated and listed in Table I (see Attachment) The 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 readout are involved.

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 in a corresponding fully accessible memory (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 that 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.

BIBLIOGRAPHY


TECHNOLOGY TRANSFER

This period saw the culmination of our years of effort to make a hybrid Josephson-CMOS memory with sufficiently small access time and power dissipation to be considered for application in a superconductive computer. Using 65 nm CMOS technology in a 64 kbit memory array we demonstrated access time of 400 ps with a read power of 12 mW at 1 GHz operation.

We reported the overall system and the important interface circuits between the Josephson circuits and the CMOS memory array at the Applied Superconductivity Conference (ASC) in Portland, OR in October 2012. We organized a memory workshop at UC Berkeley following the ASC for further clarifying discussions.

We proposed a one-year study to evaluate the prospects for a 1 Mbit memory using a memory array made in advanced (22 nm) CMOS technology. Our hybrid memory will scale easily to larger memory capacities. The number of interface circuits, which are the main sources of power dissipation scale logarithmically with memory capacity. The access time follows a similar slow increase with memory capacity.

We have created an archive of our research so that if further work were funded the details of our research will be available.
Attachment for

Scientific Progress section

Hybrid Josephson-CMOS Random Access Memory with Interfacing to Josephson Digital Circuits

Fig. 1 Block diagram of the hybrid memory system
Fig. 2 Schematic diagram of the Suzuki stack.

Fig. 3. Block diagram of the hybrid interface circuit.
Fig. 4 Floor plan of 64k (256 x 256) memory and final dimensions

Fig. 5 Global word line scheme (64 global word line, local word line)
Fig. 6 128 WLs x 16 BLs Micro array.

Fig. 7 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. 8 Single-stage inverter delay in a 409-stage ring oscillator

Fig. 9 Wire-boded chip pair: 2 x 1.5 mm² CMOS chip on top of 5 x 5 mm² Josephson chip.
Fig. 10. Block diagram for read access time test scheme. Delay from RD_Mon to DR0 is the system read access time.

TABLE I

SUMMARY OF DELAY AND READ POWER AT 1 GHz OF THE 64-KB HYBRID MEMORY SYSTEM AND COMPONENTS

<table>
<thead>
<tr>
<th>Component</th>
<th>Delay (ps)</th>
<th>Read Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suzuki Stacks (x 14)</td>
<td>47</td>
<td>2.3</td>
</tr>
<tr>
<td>CMOS comparators (x 14)</td>
<td>167</td>
<td>6.4</td>
</tr>
<tr>
<td>CMOS decoder + World lines</td>
<td>135</td>
<td>2.47</td>
</tr>
<tr>
<td>Memory cells + Bit lines (x 16)</td>
<td>48</td>
<td>0.13</td>
</tr>
<tr>
<td>4JLs (x 16)</td>
<td>10</td>
<td>0.448</td>
</tr>
<tr>
<td>Total</td>
<td>402</td>
<td>11.88</td>
</tr>
</tbody>
</table>