Energy-Efficient Superconducting Computing—Power Budgets and Requirements

D. Scott Holmes, *Senior Member, IEEE*, Andrew L. Ripple, and Marc A. Manheimer

***Abstract*—Large-scale computing system characteristics vary by application class, but power and energy use has become a major problem for all classes. Superconducting computing may be able to serve the needs of these systems significantly better than conventional technology. Recent developments in single flux quantum circuit technology for digital logic include variants with greatly improved energy efficiency. Concepts were investigated for computing systems capable of performance in the range from 1 to 1000 PFLOP/s. The concept systems were constrained to use existing commercial cryogenic refrigerators and Nb superconducting technology. In order to meet the performance goals, cache and main memory capable of operating at cryogenic temperatures will be required. Superconducting computing is shown to be potentially competitive on the basis of power and energy efficiency if key component technologies can meet specific goals. Potential advantages of superconducting computing are identified as well as areas requiring further development.**

***Index Terms*—Rapid single flux quantum (RSFQ), single flux quantum, supercomputers, superconducting integrated circuits, superconducting logic circuits.**

I. Introduction

POWER AND energy use by large-scale computing systems is a significant and growing problem. A 2010 study by Bronk, *et al.* projected that U.S. data center energy use would rise from 72 to 176 TWh between 2009 and 2020, assuming no constraints on energy availability [1]. The potential benefit to the U.S. of technology that reduces energy requirements by a factor of 10 is on the order of $15 billion annually by the year 2020, assuming an energy cost of 0.1 $/kWh. Note that this counts only the benefit of energy savings and does not include the potential economic benefits resulting from increased data center operation or savings due to reduced construction costs.

The growth of large, centralized computing facilities is being driven by several factors including cloud computing, support of mobile devices, Internet traffic growth, and computation-intensive applications. Classes of large-scale computing systems include supercomputers, data centers and special purpose machines.

Manuscript received October 10, 2012; accepted January 25, 2013. Date of publication February 4, 2013; date of current version March 10, 2013.

D. S. Holmes is with Booz Allen Hamilton, McLean, VA 22102 USA (e-mail: Holmes_Douglas@bah.com).

A. L. Ripple was with Booz Allen Hamilton, McLean, VA 22102 USA. He is now with RipTronix LLC, Lovettsville, VA 20180 USA (e-mail: andrewripple@yahoo.com).

M. A. Manheimer is with the Laboratory for Physical Sciences, College Park, MD 20740 USA (e-mail: marc@lps.umd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASC.2013.2244634

Supercomputers are also called high performance or high end systems. Information about the top 500 supercomputers is readily available [2], [3]. The total power demand of the TOP500 in 2012 was ~0.25 GW. Current efforts to improve the energy efficiency of supercomputers include DARPA and DoE programs with performance goals of 1 EFLOP/s for 20 MW by about 2020. Note that the FLOP metric is based on Linpack, double-precision floating point operations and 1 EFLOP/s is 10^{18} floating point operations per second.

Data Centers worldwide number roughly 500 000 in 2011 and draw an estimated 31 GW of electric power [4]–[6]. Information about data centers is harder to find as there is no comprehensive list and much of the information is not public. Exceptions include colocation data centers [7], which are available for hire and include about 5% of data centers by number, and the Open Compute Project led by Facebook [8].

Conventional computing technology based on CMOS (Complementary Metal-Oxide-Semiconductor) switching devices and normal metal interconnects may not be able to increase energy efficiency fast enough to keep up with increasing demand for computing. Superconducting computing makes use of low temperature phenomena with potential advantages. Josephson junctions are superconducting devices based on the Josephson effect; they switch quickly (~1 ps) and dissipate little energy per switch ($< 10^{-19}$ J). Single flux quantum (SFQ) circuits produce small current pulses that travel at about 1/3 the speed of light, c. Superconducting passive transmission lines (PTL) are also able to transmit the pulses with extremely low losses. Superconducting computing circuits typically operate at temperatures below 10 K. Large computing systems typically require cooling, so refrigeration is acceptable.

Performance as measured by computation throughput, not energy efficiency, was the dominant computing need until recently, so the interest in superconducting electronics was based on processor speed. Rapid Single Flux Quantum (RSFQ) technology developed in the 1980s uses resistors to regulate bias currents to the Josephson junctions [9] and has been used to demonstrate circuits operating up to 770 GHz [10]. In RSFQ, static power dissipation in the bias resistors is about 100 times greater than the dynamic power dissipation in the Josephson junctions and associated damping resistors. Even including refrigeration, RSFQ circuits did not require more power than conventional technologies, which allowed some advantages when performance was the dominant need. Outside of niche applications, computers based on RSFQ have not been competitive due to lack of adequate: 1) SFQ-compatible

TABLE I
SYSTEM POWER BUDGETS FOR SUPERCONDUCTING SUPERCOMPUTERS

	System Performance (PFLOP/s)			
	1	10	100	1,000
(quantity) Refrigeration system manufacturer, model	(1) SHI, SRDK-415D-F50	(2) SHI, GM-JT CG310SLCR	(1) Linde, LR70 helium reliquefier	(2) Linde, LR280 helium reliquefier
• Cost	• 45 k$	• 320 k$	• 2 M$	• 6 M$
• Refrigeration capacity	• 1.5 W at 4.2 K	• 10 W at 4.3 K	• 100 W at 4.4 K	• 1,020 W at 4.4 K
• Efficiency	• 5000 W/W	• 1280 W/W	• 450 W/W	• 395 W/W
Power, refrigeration (kW)	7.5	12.8	45	400
Power, other (kW)	7.5	27.2	155	1,600
Total system power (kW)	**15**	**40**	**200**	**2,000**
Computation efficiency (GFLOP/J)	67	250	500	500

memory, 2) data interconnects between low- and room-temperature environments, and 3) fabrication capability for superconducting electronics.

Recent developments in SFQ technology include variants with greatly improved energy efficiency such as reciprocal quantum logic (RQL) [11]–[13], efficient SFQ (ERSFQ, eSFQ) [14]–[16], LR-RSFQ [17], [18] and low-voltage RSFQ [19], [20]. Reductions in power and energy have been demonstrated by over a factor of 100 relative to conventional RSFQ. Prospects for cryogenic memories have also improved recently with the development of memory elements combining features of Josephson junctions and magnetic random access memory (MRAM) [21]–[27]. The ability to operate both logic and memory within the cold environment, rather than with the main memory out at room temperature, decreases demands on the interconnects to room temperature to the point that engineering solutions can be found.

Recent work by Nagy *et al.* [28] indicates that the Moore's Law trajectory of the semiconductor industry, with its doubling of components in an integrated circuit about every two years, is a consequence of exponentially growing revenues and not just time. To take off in a similar manner, superconducting electronics needs significant markets to fund its development. Energy-efficient large-scale computing and intermediate, related products such as routers and network switches may be the market opportunity that has been needed.

II. SYSTEM REQUIREMENTS

A. System Ranges

Superconducting computers are being evaluated for potential energy efficiency benefits relative to conventional technology. The benefit of an energy-saving technology scales as the number of systems times the energy savings per system. Supercomputing was selected for this feasibility study as less memory is required than for data center applications. Ranges of computer systems studied were set as follows for parameters including performance, computation efficiency, and architecture.

1) Performance (FLOP/s): Superconducting technology requires refrigeration and the efficiency of cryogenic refrigerators improves with size, thus favoring large-scale systems. If the performance of the top 500 supercomputers continues to follow historical exponential growth curves, by 2020 the range can be expected to span from roughly 10 to 1000 PFLOP/s [2]. The number of computer systems at the low end grows rapidly, especially for data centers, although these systems typically are not rated using a FLOP/s metric. Our estimate is that systems of interest range down about another order of magnitude. The performance range studied was thus from 1 to 1000 PFLOP/s. Note that 1 PFLOP/s was the target for the Hybrid Technology Multi-Threaded (HTMT) project which included the design for an RSFQ-based processor over a decade ago [29], [30].

2) Efficiency (FLOP/J): The most efficient supercomputer at the time of this writing achieves 2 GFLOP/J [3]. The DoE goal of 1 EFLOP/s for 20 MW equates to 50 GFLOP/J. An attractive goal would be an overall system energy efficiency at least 10 times better: 500 GFLOP/J.

3) Architecture: Computing system design (architecture) affects programmability and the classes and range of work a computer can perform well. For example, Murphy [31] shows that applications dominated by integer operations are significantly more sensitive to memory latency and data rate than traditional supercomputer applications dominated by floating point operations. FLOP/s is known to be an imperfect metric, but has been the most common metric for supercomputers and thus has been used commonly in reported architectural metrics. Desirable architectural metrics for supercomputers designed for floating-point-intensive applications include:

- Main memory: 0.1 to 1 B s/FLOP
- Main memory latency (access time): < 100 cycles
- Main memory data access rate: 1 B/FLOP
- Input/Output data rate: 10^{-5} to10^{-3} B/FLOP
- Parallelism: fewer processors is generally better

The input/output specification assumes a conventional supercomputer architecture with the data storage system outside the main portion of the computer. For such systems the input/output specification includes data storage access.

Achieving the main memory latency goal requires that the main memory be very close to the logic. For a superconducting computer, this requires the main memory to be co-located with the logic the cold space.

B. System Power Budgets

Power budgets for systems based on commercial, 60 Hz cryogenic refrigerators are shown in Table I. Refrigerator efficiency improves by a factor of about 10 over this range and contributes significantly to the ability to reach our overall system efficiency

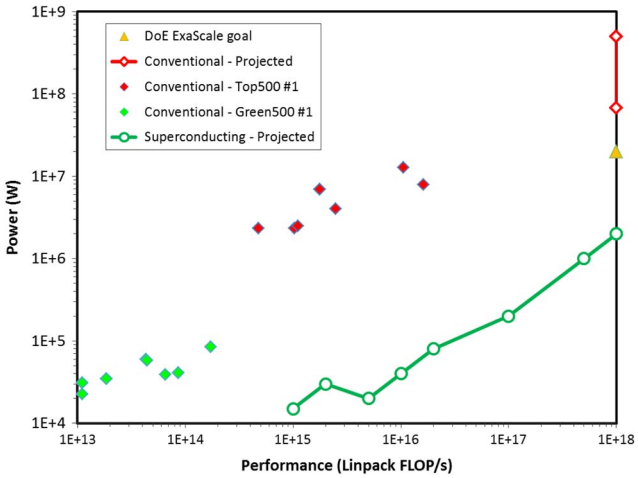

Fig. 1. Power versus performance for conventional [2], [3] and superconducting [Table I] computer systems.

goal with the larger computer systems. All of the refrigeration systems are commercially available. "Non-refrigerated" system components includes room temperature interconnect drivers and receivers, power supplies, and storage memory. A potential problem is that storage memory provided by disk drives currently does not improve in energy efficiency with capacity at the PB scale needed by systems in this performance range. Solid state storage class memory (SCM) is promising, but still evolving. Getting economies of scale for non-refrigerated components will be a challenge that requires study and possible development of more energy efficient technologies.

Power for the projected superconducting computer systems is plotted as a function of performance in Fig. 1. Also plotted for comparison are conventional systems, both existing and projected. "Conventional—Projected" shows the range of estimates for the ExaScale study Strawman system [32], [33].

The power dissipation in a conventional supercomputer typically is divided roughly evenly between logic, memory and interconnects. A refrigerated system would also require power to remove heat that leaks into the low-temperature environment by conduction down structural and communications links, by convection through gasses and liquids, and by thermal radiation. The refrigeration heat load budget targets for components were set at 30% for each of logic, memory, and interconnects and at 10% for heat leaks. The range for each component budget could vary from these targets by about a factor of 3. For example, the memory target is 30% of the refrigeration heat load budget, but could range from 10% to 90%.

III. Feasibility Assessment

The possibility of achieving the system goals and power budgets was assessed for the major component groupings.

A. Logic

The energy dissipated when switching a Josephson junction is nearly $I_c\Phi_0$ where I_c is the critical current of the junction and Φ_0 is the flux quantum $(2.07 \times 10^{15}\ \mathrm{V \cdot s})$, a fundamental constant [34]. The minimum energy must be sufficiently large to avoid unintentional switching due to thermal noise. For a single junction with a dc bias and a desired bit error rate (BER) $< 10^{-30}$, the minimum critical current is about 100 μA [34]. The minimum energy per switch is thus about 2×10^{-19} J.

There are prospects for reducing the critical current and thus the power and energy. Klein and Mukherjee [35] found that the phases of Josephson junctions in circuits are not completely independent and thermal fluctuations must switch the combined circuit to cause a spurious switching. In this case the effective critical current can be greater than I_c of a single junction. The net effect might be to reduce the minimum critical current by about a factor of 2. The use of phase-shifting elements has also been shown to significantly reduce the influence of thermal noise on circuit behavior [36].

RQL uses ac biasing and 2 pulses to signify a digital "1" and is found to dissipate about 1/3 of $I_c\Phi_0$ per switch [13]. RQL has also been found to be more stable against thermal errors. Reduction of the minimum critical current by a factor of 4 to 10 appears possible [11].

A simplistic estimate of the power required to compute 1 PFLOP/s using RQL circuits with $I_c = 25$ μA, gate counts for CMOS architectures, and JJ activity factor α_F is:

$$\begin{aligned} P &= (\text{FLOP/s})(\text{gate/FLOP})(\text{JJ/gate})(\text{energy/JJ})\alpha_F \\ &= (10^{15})(5000)(6)(1.7 \times 10^{-20}\ \text{J})(0.5) \\ &= 0.26\ \text{W at } 4\,\text{K}. \end{aligned} \tag{1}$$

From Table I the maximum heat load for a 1 PFLOP/s computer is 1.5 W at 4.2 K and 30% of this is 0.45 W, which indicates that RQL might be able to meet the efficiency requirement. Other SFQ variations such as eSFQ might also prove suitable. Even further reductions in energy and power might be possible using reversible or adiabatic computing circuits, although they have not yet been demonstrated and are likely to be much slower.

The number of Josephson junctions per processor used in the feasibility study was 20 million, which is greater than the 7.2 million estimated for the HTMT SPELL processor [37] to allow for somewhat more complexity and flexibility.

B. Memory

Low-temperature memories are not as well developed as logic circuits. The magnetic flux threading a superconducting loop in steady state is quantized and thus can be used to provide the physical basis for a digital memory element. The absence or presence of a flux quantum in the loop represents binary "0" or "1". Superconducting memory cells have one or more Josephson junctions in the loop to control and sense the number of flux quanta present. The largest demonstrated superconducting random-access memory (RAM) is only 4 Kibit (4096 bits) [38], [39]. Until very recently, SFQ memories used less energy-efficient RSFQ technology and the smallest cell size was 15 μm $\times$ 15 μm—much larger than the cell sizes now achievable using current CMOS technology.

The power dissipated at 4 K by 1 PB of memory can be estimated from the design for a 64-Kibit RSFQ memory chip [38] as

$$(10^{15}\ \text{B})(8\ \text{bit/B})(7\times 10^{-4}\ \text{W/chip})/(2^{16}\ \text{bit/chip}) = 85\ \text{MW}. \quad (2)$$

Available superconducting memories based on RSFQ technology clearly dissipate too much power for use as memory in the 4 K environment of petascale computer systems. Significant improvements are possible using the new, more energy-efficient SFQ technologies and finer lithography. Still, improvements in energy efficiency by factors of millions will be a challenge.

Superconducting memories seem most promising for register and cache memory on the processor chip where speed is extremely important. Significant improvements in physical density and energy efficiency will be required even for these applications.

CMOS memories designed and fabricated to operate at 4 K can be integrated with SFQ circuits. Hybrid Josephson-CMOS memories up to 64 Kibit have been developed and tested [40], [41]. While these are larger and denser than purely superconducting memories built to date, the power and energy dissipation is too large to serve in an exascale superconducting computer.

The search for more suitable memories has begun and some concepts are promising, however none have been demonstrated.

C. Interconnects and Heat Leak

Heat leak between the warm and cold portions of a superconducting computer is an important consideration. Locating the main memory within the cold space greatly decreases, but does not eliminate the problem.

Input from room temperatures to SFQ circuits is conceptually easy as the output from a single-mode optical fiber can be used to directly drive a Josephson junction, but development has been limited [42]. More developed optical input schemes require intermediate components such as a photodiode [43].

Output from SFQ circuits to room temperature systems is the biggest interconnect challenge. SFQ circuits produce signals of a few mV whereas drivers for coaxial or optical outputs typically require volt-level signals. One approach is to send small signals from the 4 K environment to semiconducting amplifiers operating at an intermediate temperature stage [44].

On-chip and chip-to-chip, superconducting passive transmission lines (PTL) provide an effective means to move SFQ pulses within the cold portion of the computer system [45], [46]. Transmitting a bit requires 2 Josephson junctions to switch, one on each end of the PTL. The power to access cache memory is estimated as 2.2 to 18 mW per PFLOP/s, using RQL with an average of 1 pulse per bit, $I_c = 100\ \mu$A, and $E_{sw} = I_c\Phi_0/3$ per switch, and 2 times the cache data access rate ranges given in Table IV. Similarly, the power to access main memory is estimated as 0.5 to 9 mW per PFLOP/s, using double SFQ pulses with an average of 2 repeaters in each transmission path, and 2 times the main memory data access rate ranges given in Table IV. The high number of each range was used for the estimates in Table II.

TABLE II
INTERCONNECT POWER DISSIPATION AT 4 K

	System Performance (PFLOP/s)			
	1	10	100	1,000
I/O data rate [a] (Tbit/s)	0.8	8	80	800
• Channels [b], 20 Gbit/s	• 80	• 800	• 8,000	• 80,000
Power leads	[c]	[c]	[c]	[c]
Input data	[c]	[c]	[c]	[c]
Cache memory access	18 mW	180 mW	1.8 W	18 W
Main memory access	9 mW	90 mW	0.9 W	9 W
Output data	10 mW	100 mW	1.0 W	10 W
• Drivers, eSFQ-to-DC [d]	• 0.024	• 0.24	• 0.002	• 0.024
• Ribbon cable to 40 K	• 8.3	• 83	• 0.83	• 8.3
• VCSEL array at 40 K [d]	• 0 [e]	• 0 [e]	• 0 [e]	• 0 [e]
Interconnects, total	**0.1 W**	**1 W**	**10 W**	**100 W**
I/O budget	0.4 W	3 W	30 W	300 W

[a] Specified using the mid-range I/O data rate (10^{-4} B/FLOP)·(8 bit/B).
[b] Channel capacity is 2 times the specified I/O data rate.
[c] No estimate made. [d] [47].
[e] Vertical-cavity surface-emitting laser (VCSEL) heat load is less than refrigerator intermediate stage capacity, so no effect on 4 K capacity.

Supplying electrical power to a cryogenic environment with minimal heat leak is well studied and well developed. Most challenging are leads to carry high currents or high frequencies (GHz).

Heat leak is estimated to be about 0.1 W for a 1 PFLOP/s system with optical fiber inputs, eSFQ-to-DC output drivers with Cu ribbon cable to semiconductor amplifiers at 40 K, and superconducting PTL on and between chips. To keep the additional heat leak down power leads from exceeding the power budget will require methods such as ac biasing, as used in RQL, or recycling of dc current. Table II summarizes the estimated heat loads due to interconnects.

D. Chip Area

1) Logic: The required chip area was estimated as the number of processors times the number of Josephson junctions (JJ) per processor (20 million) divided by the JJ area density (1×10^{10} m^{-2}; see discussion in Section V-B). The estimated chip area per processor is thus 2×10^{-3} m^2 (20 cm^2).

2) Memory: An average density of 900 GB/m^2 is estimated for a technology with a bit cell size of 360 nm × 360 nm, which is $4F^2$ with a feature size $F = 180$ nm. Multiple junction layers may be required to achieve such an average bit density. The technology for such a bit cell has yet to be demonstrated.

3) Interconnect: The logic chip area was increased by 5% to account for Input/Output drivers and receivers.

E. Packaging and System Volume

The volume required depends on the packaging scheme. High-density packaging schemes were reviewed for non-cryogenic chips in the ExaScale study [32] (see sections 6.6.1, 7.1.5). The design concept for this feasibility study is to use multi-chip modules (MCM) based on carrier substrates with superconducting Nb wiring, but no active circuits. Stacks of

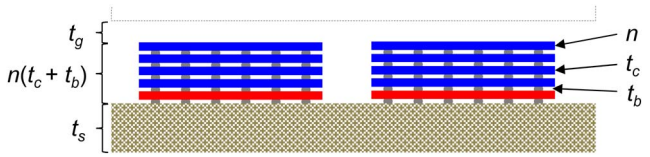

Fig. 2. Multichip module (MCM) packaging concept.

processor or memory chips would be bump bonded onto each carrier substrate as shown in Fig. 2.

The relationship between packaged volume, V, and active circuit area, A_a, is

$$\left(\frac{V}{A_a}\right)_{MCM} = [t_g + t_s + n(t_c + t_b)]\,(f \cdot n)^{-1} \qquad (3)$$

where:

t_g 2 to 10 mm; vertical gap between stacked MCMs
t_s 0.5 to 1 mm; thickness of the MCM substrate
t_c 50 to 200 μm; thickness of a circuit chip
t_b 10 to 30 μm; thickness added by bump bonds
n 1 to 5; number of chips in a stack
f 0.4 to 0.6; effective in-plane area fraction covered by chips, including MCM-to-MCM horizontal spacing

The range for $(V/A_a)_{MCM}$ using these numbers is 0.001 m to 0.030 m.

Table III includes component and total cryostat volumes for the computer systems studied.

The technology required to build large superconducting computing systems is not yet fully available. Key areas requiring development are discussed in following sections.

TABLE III
CIRCUIT AREA AND CRYOSTAT VOLUME RANGES

	System Performance (PFLOP/s)			
	1	10	100	1,000
Logic, processors [a]	40,200	402,000	4,020,000	40,200,000
• circuit area [b] (m^2)	80.4	804	8,040	80,400
• MCM volume (m^3)	0.08-2.4	0.8-24	8-240	80-2,400
Memory [c] (PB)	1	10	100	1,000
• circuit area (m^2)	1.1e+3	1.1e+4	1.1e+5	1.1e+6
• MCM volume (m^3)	1.1-33	11-330	110-3,300	1,100-33,000
Interconnect channels [d]	80	800	8,000	80,000
• circuit area (m^2)	1.2	12	120	1,200
• MCM volume (m^3)	0.002-0.06	0.02-0.6	0.2-6	2-60
Other [e] (m^3)	0.1-3.3	1-33	10-330	100-3,300
System volume				
• high (m^3)	39	390	3,900	39,000
• low (m^3)	1.3	13	130	1,300

[a] RQL, I_c = 25 μA, 8.3 GHz [48].
[b] 2×10^7 Josephson junctions per processor, 10^{10} JJ/m^2.
[c] Josephson MRAM [48], 1 byte per FLOP/s, 900 GB/m^2.
[d] Specifications in Table II, circuit area is 5% of logic area.
[e] Cryostat + structure estimated as 10% of total MCM volume.

TABLE IV
MEMORY GOALS FOR SUPERCONDUCTING COMPUTING (SC) BY MEMORY TYPE

Memory	Data rate (B/FLOP)		Per processor [a]		Total for 1 PFLOP/s	
type	SC	Exascale	SC	Exascale	SC	Exascale
Register	8 to 24	8	1 KiB to 4 KiB	8 KiB [b]	2.5 MB to 4,100 MB	1,360 MB
Cache (L1-L3)	1 to 8	3.5	128 KiB to 1 MiB	192 + ? KiB	0.3 GB to 1,050 GB	37.2 GB
Main	0.05 to 1	0.16	?	16 GB [c]	0.1 to 1 PB	0.0036 PB

"Exascale" refers to the Exascale strawman design [32].
[a] SC design: 2 FPU/processor, Exascale design: 4 FPU/processor.
[b] 8 KiB = (4 FPU/processor)(128 registers/FPU)(128 bit/register)/(8 bit/B).
[c] Local to processor.

IV. MEMORY REQUIREMENTS

Large quantities of data must be stored in memory and moved between logic circuits and memory. To minimize the energy and time required, memory is best located as close to the logic as possible. Best would be memory that operates at the same temperature as the logic. Signal compatibility is also highly desirable to avoid the cost of transforming signals.

Given the refrigeration cost to remove heat from a low temperature environment, a memory operating at 4 to 10 K must dissipate very little energy for:

- static operation (maintaining the memory state)
- dynamic operation (reads, writes)
- data movement between processor and memory
- signal transformation

Memory power is expected to be proportional to both the total quantity of memory and to the memory access rate.

A. Primary Goals

Primary memory energy goals can be developed from system energy efficiency requirements. In refrigerated systems, the available power is reduced by refrigeration overhead and heat leaks into the cold space. A reasonable target for memory is roughly 1/3 the available cooling power from Table I, with upper and lower bounds about a factor of 2 higher and 5 lower, respectively. The target is not well represented by a single number because the memory quantity and access rate are not set requirements, but can be traded off with energy and power.

1) Quantity and Access Rate: The memory goals for superconducting computers given in Table IV were developed based on the ranges for comparable systems such as those in [32] (see Table 7.7) and [49]. Higher quantities are typically more desirable.

The quantity of register and cache memory is specified in bytes per processor (B/processor) and the total for each type is $= (\text{B/processor})(\text{FLOP/s})/[(\text{cycle/s})(\text{FLOP/cycle} \cdot \text{processor})]$. Expected ranges for the other quantities in this formula are 1 to 100 GHz clock speed (cycle/s) and 1 to 4 FLOP/(cycle·processor). The number of processors thus ranges from 2.5e+3 to 1e+6 per PFLOP/s. Processing a double-precision FLOP requires 2 reads and 1 write of 64 bits (8 bytes) each. The desired total memory access rate is thus about $3 \times 8 = 24$ B/FLOP.

Ranges for memory quantities appearing in Table V were set with minimums (worst) at 1/4 the quantity in the SPELL processor design, which included 256 Kibit of cache and 8 Mibit of RAM [37]. The maximums (best) were set a factor of 2^{12} (4096) higher.

TABLE V
LOW-TEMPERATURE MEMORY GOAL RANGES

		Cache		Main	
	Unit	Best	Worst	Best	Worst
Operating *T*	K	4	10	4	10
Energy per access [a]	J/bit	6e-19	5e-17	5e-18	1e-15
Static power	W/bit	5e-19	2.5e-16	5e-19	2.5e-16
Read time	cycle	1	5	5	50
@ 10 GHz	ps	100	500	500	5,000
@ 100 GHz	ps	10	50	50	500
Write time	cycle	2	20	20	200
@ 10 GHz	ps	200	2,000	2,000	20,000
@ 100 GHz	ps	20	200	200	2,000
Latency [a]	cycle	1.3	10	10	100
@ 10 GHz	ps	133	1,000	1,000	10,000
@ 100 GHz	ps	13	100	100	1,000
Distance [b]	cycle	0.01	1	1	100
@ 10 GHz	m	1e-4	1e-2	1e-2	1
@ 100 GHz	m	1e-5	1e-3	1e-3	1e-1
Total per chip	bit	128 Mi (2^{27})	32 Ki (2^{15})	8 Gi (2^{33})	2 Mi (2^{21})
Density [c]	bit/m^2	3e+13	3e+8	2e+15	2e+10

Primary goals are in bold face.

Read and write times for memories are given in units of clock cycles and in ps at the specified clock frequency.

[a] Average computed as (2R+W)/3.

[b] Distance from logic.

[c] Including local interface circuits.

2) Access Energy: The energy per memory access (read or write) is useful when evaluating candidate memory technologies. The goal access rate from Table IV is 1 to 8 B/FLOP for cache memory and 0.05 to 1 B/FLOP for main memory. This ratio of about 10 cache to main memory accesses requires that cache accesses be more energy efficient. The cache and main memory power ranges were each set at 0.1 to 1 times the total memory power target, which was set above at 1/3 the available cooling power from Table I. The energy bounds are then calculated for a 1 PFLOP/s computer as 6×10^{-19} to 5×10^{-17} J/bit for cache memory and as 5×10^{-18} to 1×10^{-15} J/bit for main memory. Note that this lower bound is not a strict limit, but only indicates a point beyond which additional efforts to improve energy efficiency will not yield significant benefits for the overall system.

The average memory access energy depends on the ratio of reads/writes (R/W). This ratio varies by memory architecture and application, but a range of 1 to 4 is reasonable and 2 is a commonly used number for estimation purposes. The average energy $E = [(R/W)E_{\mathrm{R}} + E_{\mathrm{W}}]/[(R/W) + 1]$.

3) Static Power: Static power dissipation ranges can be developed similarly. The quantity of cache memory from Table IV is 0.3 to 1,050 GB·s/PFLOP, which is at most just 1% of the quantity of main memory. Given the relatively small quantity of cache memory, it will be ignored when calculating the static power range, but the static power range will be expected to apply to cache memory as well as main memory. We will set the upper limit for static power at half the target and the lower bound of interest at 1% of the target memory power, in which case the range is 5×10^{-19} to 2.5×10^{-16} W/bit.

4) Operating Temperature: Since data movement costs significant energy unless superconducting transmission lines are used, the memory access energy might be minimized by locating the main memory within the low temperature environment where superconducting transmission lines can operate. The temperature must be below about 70 K, depending on the superconducting properties of the material used in the transmission lines. The higher the operating temperature, the farther the memory must be located from the logic, which operates at temperatures of 4 to 5 K using presently available SFQ technology.

B. Secondary Goals

Secondary goals affect the ability of a memory technology to satisfy the primary goals and help us to select promising candidates for further development. For example, the energy to move data between logic and memory can be a significant part of the memory access energy. The logic-memory interconnect energy therefore affects the choice of memory technology and can be used to set secondary goals.

1) Time to Read or Write: The time required to read or write a memory element can be significant. For some technologies the write time is much longer than the read time. Read time is more important as the processor often must wait until the data arrives. Writing, on the other hand, does not normally require the processor to wait for completion. Typical delay times for computers based on current CMOS technology are L1 cache: 1 clock cycle; L2 cache: 10 cycles; main memory: 100 cycles. Faster read times are beneficial, as pointed out by Murphy [31] and discussed earlier. Therefore we set goal ranges for **cache** memories at 1 to 5 clock cycles to read and 2 to 20 clock cycles to write and for **main** memories at 5 to 50 clock cycles to read and 20 to 200 clock cycles to write.

2) Distance from Logic: Increased physical distances between logic and memory also introduce delays. Significant delays decrease computing speed and thus decrease computation energy efficiency because refrigeration power remains roughly constant. The significance threshold for delays in communicating with cache memory is around 0.1 clock cycle and with main memory is around 10 clock cycles. Information moves on superconducting transmission lines at speeds of about $c/3$, so a separation of 100 mm causes a delay of $\sim$1 ns, which is 10 clock cycles for a computer operating at 10 GHz. Factors which affect the distance between logic (processor chips) and main memory include operating temperatures, physical densities of each, compatibility of the fabrication processes, and chip mounting technology. The goal range will be considered to be within a factor of 10 of the significance threshold.

3) Density: Memory cost and physical volume are inversely proportional to the bit density that can be fabricated on a chip. The minimum (worst) density in the range of interest was set as the minimum memory quantity divided by the area of a 10 mm $\times$ 10 mm chip. The maximum (best) was set a factor of 10^5 higher.

C. Memory Goal Summary

Cache and main memory goal ranges are summarized in Table V. "Worst" goals are limits beyond which the overall system goals cannot be achieved. "Best" goals are in some

cases not strict limits, but rather limits beyond which minimal additional benefits are expected for the overall system.

V. Fabrication Technology Requirements

The larger the total area of integrated circuit chips required to build a given computer system, the lower the performance and the higher the initial cost, volume, and operating expense. Circuit complexity and density achievable on a single chip needs to improve greatly to make possible a supercomputer competitive on the basis of cost.

To remain competitive with other technologies, circuit density must continue to improve. Chatterjee [50] provides a concise history of the approaches used successfully by the semiconductor microelectronics industry. CMOS transistors initially benefited from Dennard scaling rules [51]. More recently the semiconductor industry has been forced to develop new materials and new device structures to keep circuit density advancing more or less according to Moore's law. The superconducting electronic industry can use similar approaches.

A. Foundry for Superconducting Electronics

Although much work has been done [52]–[59], present fabrication technology cannot produce integrated circuits with the density or number of junctions on a chip required for large-scale computing applications. For example, the largest SFQ processors reported to date have only about 22 000 junctions [60], far less than the 10 to 20 million estimated to be required.

Approaches to increase the density of superconducting integrated circuits include:

1) Feature Size: Decreasing F from 1000 nm down to 90 nm could increase density by as much as $(1000/90)^2 \cong 120$ times.

2) Superconducting Layers: Additional layers for wiring or inductors allows stacking of components.

3) JJ layers: Development of robust barrier technology could enable multiple Josephson junction layers.

4) Increase Critical Current Density, J_c: Higher J_c allows smaller junction sizes. An additional benefit is that junctions with higher J_c switch faster, allowing an increase in the operating speed or clock frequency. The increase in processing speed can be used to increase performance or to reduce the number of processors required to achieve a given performance level.

5) Self-Shunted Junctions: Nb/Al/AlOx/Nb Josephson junctions with J_c above about 1 GA/m^2 (100 kA/cm^2) do not require shunt resistors and thus require less chip area [61].

6) Barrier materials: Highly uniform and reproducible barriers are needed to produce large numbers of Josephson junctions with tightly controlled critical currents. Aluminum oxide barriers are thin, delicate, and difficult to fabricate reproducibly with J_c above about 100 MA/m^2. Other barrier materials may allow fabrication of higher-J_c junctions, which would be smaller in size [62], [63].

7) High Inductance Materials: The inductance of a superconducting stripline decreases as the conductor width approaches the thickness of the dielectric between the conductor and ground planes [64], [65]. Superconductors such as NbN have higher kinetic inductance than the Nb typically used in superconducting electronic circuits [66].

8) Magnetic Materials: Josephson junctions incorporating magnetic layers could be used to provide bias currents [67], [68] or to create memory elements [21], [22] with fewer circuit elements or less area.

B. Roadmap for Superconducting Electronics

A roadmap is needed similar to the International Technology Roadmap for Semiconductors (ITRS). Earlier roadmaps for superconducting electronics such as [30], [54]–[58] are a start, but need to be supported with models and extended for multiple technology generations. Community support and commitment are required to derive full benefit from a technology roadmap.

Models are needed to predict what can be achieved and to guide fabrication development efforts [69]. For example, a preliminary model for the area of a standard set of SFQ components created by one of the authors [70] indicates that a process capable of fabricating circuits with $I_c = 25\ \mu$A, $J_c = 100$ MA/m^2, $F = 180$ nm, and 2 inductor layers could achieve the JJ area density $(1 \times 10^{10}\ \text{m}^{-2})$ used in Section III-D to estimate system volumes.

VI. Design Tool Requirements

Electronic design automation (EDA) tools are used to develop large, complex electronic circuits. EDA is a subcategory of computer aided design (CAD). Superconducting digital circuits for large-scale computers are expected to require millions of components, for which EDA tools will be essential.

The EDA tool flow needed to design very large scale (VLS) superconducting integrated circuits (ICs) is similar to the tool flow used to design complementary metal-oxide-semiconductor (CMOS) circuits, but there are important differences. In 1999 Gaj, *et al.* published the results of a survey of EDA tools for superconducting digital electronics [71]. EDA tools identified as needed, but missing at that time included: timing optimization, automated logic synthesis, and automated layout synthesis. A re-evaluation of the current situation was motivated by the many years of subsequent work [72]–[87] and by renewed interest in development of large-scale superconducting computers.

A. Large Scale SFQ Design Flow

A survey of several practitioners (see Acknowledgements) was undertaken to determine both current processes and the tool flow desired for large scale SFQ integrated circuit design. To summarize the findings:

Semiconductor industry design flows and tools should be used when applicable to leverage investment and infrastructure as much as possible.

Some needed tools either do not exist or require significant improvement, including all of those identified as needed, but missing in 1999. Additionally, we have identified need for standard cell SFQ libraries common to all foundries, tools to automate timing analysis, perform equivalence checking and LVS (layout versus schematic).

B. Needed SFQ Design Tool Enhancements

The existing SFQ design flow is adequate for small scale SFQ design and implementation and current development activities will provide an adequate framework for medium scale circuits. The level of support needed for large scale logic design and implementation is inadequate and needs to be addressed. In order to realize large scale SFQ logic a common standard cell library must be available from the foundry to all circuit designers and the existing development tools will need to be enhanced or replaced to provide the same level of support that CMOS development tools provide.

The items that need to be enhanced or developed are:

1) Common Standard Cell Library: A standard cell library that is developed and maintained by the foundry is required for each logic family. Cadence is the tool of choice for the majority of SFQ designers; therefore a standard cell library that is supported by the Cadence tool flow is the logical choice. Standard cells optimized for worst-case timing would reduce or eliminate the need for the logic designer to tweak device parameters and would allow the logic designer to focus on the design functionality instead of the device physics. Development efforts to create this library are currently underway and need to be monitored to ensure future goals are met since this is the basic building block that is required for large scale circuit design. Each standard cell should provide:

- HDL (hardware description language) gate level simulation model,
- schematic symbol,
- full layout optimized to allow operation at the specified worst-case timing.

The HDL gate level simulation model can be used for functional and timing simulations. The model should have self-contained assertion statements that validates input transitions. This gate level simulation model is useful for small to medium scale logic design, but will significantly impact the simulator performance with large scale circuit design, hence the need for a static timing analysis tool. Regardless, the gate level simulation model is required.

2) Automated Logic Synthesis Tools: Custom tools and synthesis algorithms need to be developed to efficiently translate behavioral HDL to the appropriate SFQ logic. These tools should:

- accept a technology file or library of components for each SFQ logic family,
- produce a standard format synthesized netlist from a hardware description language,
- leverage existing CMOS synthesis tools.

3) Automated Place and Route Tools: Novel placement and routing algorithms optimized for SFQ need to be explored. Additionally, the back-end tool flow needs to be more tightly integrated and fully automated. These tools should:

- accept a technology file or library of components for each SFQ logic family,
- produce a standard post place-and-route output file,
- leverage existing CMOS placement and routing tools.

4) Automated Timing Analysis Tools: Static timing analysis tools and algorithms are needed for SFQ designs. These tools should:

- leverage existing CMOS tool capabilities,
- accept a synthesized netlist and provide rough timing analysis (this feature could be integrated into the logic synthesis tool),
- accept a place-and-route output file and provide accurate timing analysis (this feature could be integrated into the Place-and-Route tool),
- provide accurate timing files that can be used by the logic simulator.

5) Formal Verification Tools: Equivalence Checking tools and Layout-Versus-Schematic tools will be necessary once automated synthesis and layout tools have been developed. This functionality could possibly be built into the SFQ synthesis and layout tools.

6) Automated Built-In-Self-Test (BIST) Insertion Tools: BIST insertion capabilities are not a requirement for small to medium circuit design, but need to be considered during the development of SFQ tools. For example, automatic insertion of scan chains and debug logic for memories should be considered while SFQ tools and libraries are being developed so that a custom BIST insertion tool will not be needed.

C. Design Tool Conclusions

The EDA tool flow needed to design very large scale (VLS) superconducting integrated circuits is similar to the tool flow used to design complementary metal-oxide-semiconductor (CMOS) circuits, but there are important differences. In some cases, industry standard CMOS EDA tools have been successfully modified to meet the needs of small scale superconducting ASIC designers. Most important, in order of need, are tools to automate: logic synthesis, placement and routing, and timing optimization. Additionally, tools to insert built in self-test (BIST) circuits will be needed in the future to allow testing of functionally complex chips.

VII. Other Technology Requirements

A. Packaging Technology

The range of volumes for a 1 PFLOP/s system (1.3 to 39 m^3) shown in Table III clearly shows the benefit to be gained by pursuing advanced, dense packaging technologies. MCMs have been demonstrated, but greater levels of stacking are needed.

B. Interconnect Technologies

Recent work on interconnects between chips on an MCM [88] and on dc current leads [89] is promising. Further development is needed in these areas as well as for I/O between the room temperature and cryogenic environments, especially for applications requiring higher I/O data rates.

VIII. Conclusion

System power budgets were developed for a range of computer performance (1 to 1000 PFLOP/s) based on commercially

available cryogenic refrigerators. The goal seems worthy and the technologies required to build such computers appear to be available or within reach with the notable exception of memory. If suitable memories become available, significant work remains in the areas of circuit density, computer architecture, fabrication, packaging, testing, and system integration.

ACKNOWLEDGMENT

The authors thank industry and university researchers for responding to the IARPA CAD Tools Questionnaire, including: A. Fujimaki-Nagoya University; A. Inamdar, D. Gupta-HYPRES Inc.; A. Herr, O. Oberg, Q. Herr, and S. Shauck-Northrop Grumman Corp.; G. Rose-D-Wave Systems Inc.; M. Dorojevets-Stony Brook University; Nobuyuki Yoshikawa-Yokohama National University; Thomas Ortlepp-University of California, Berkeley; V. Semenov-Stony Brook University.

REFERENCES

[1] C. Bronk, A. Lingamneni, and K. Palem, "Innovation for sustainability in information and communication technologies (ICT)," James A. Baker III Inst. Public Policy, Rice Univ., Houston, TX, USA, Oct. 2010, Tech. Rep.
[2] *TOP500*. [Online]. Available: www.top500.org
[3] *The Green500 List*. [Online]. Available: www.green500.org
[4] J. Koomey, "Worldwide electricity used in data centers," *Environ. Res. Lett.*, vol. 3, no. 3, pp. 034008-1–034008-8, Jul.–Sep. 2008.
[5] J. G. Koomey, C. Belady, M. Patterson, A. Santos, and K.-D. Lange, *Assessing Trends Over Time in Performance, Costs, and Energy Use For Servers*. Oakland, CA, USA: Analytics Press, Aug. 2009.
[6] J. Koomey, *Growth in Data Center Electricity Use 2005 to 2010*. Oakland, CA: Analytics Press, Jul. 2011.
[7] Colocation Data Centers. [Online]. Available: www.datacentermap.com
[8] Open Compute Project, opencompute.org.
[9] K. K. Likharev and V. K. Semenov, "RSFQ logic/memory family: A new Josephson-junction technology for sub-terahertz-clock-frequency digital systems," *IEEE Trans. Appl. Supercond.*, vol. 1, no. 1, pp. 3–28, Mar. 1991.
[10] W. Chen, A. V. Rylyakov, V. Patel, J. E. Lukens, and K. K. Likharev, "Rapid single flux quantum T-flip flop operating up to 770 GHz," *IEEE Trans. Appl. Supercond.*, vol. 9, no. 2, pp. 3212–3215, Jun. 1999.
[11] Q. P. Herr, A. Y. Herr, O. T. Oberg, and A. G. Ioannidis, "Ultra-low-power superconductor logic," *J. Appl. Phys.*, vol. 109, no. 10, pp. 103903-1–103903-8, May 2011.
[12] O. T. Oberg, Q. P. Herr, A. G. Ioannidis, and A. Y. Herr, "Integrated power divider for superconducting digital circuits," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 571–574, Jun. 2011.
[13] A. Y. Herr, Q. P. Herr, O. T. Oberg, O. Naaman, J. X. Przybysz, P. Borodulin, and S. B. Shauck, "An 8-bit carry look-ahead adder with 150 ps latency and sub-microwatt power dissipation at 10 GHz," *J. Appl. Phys.*, vol. 113, no. 3, pp. 033911-1–033911-6, Jan. 2013.
[14] O. A. Mukhanov, "Energy-efficient single flux quantum technology," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 760–769, Jun. 2011.
[15] D. E. Kirichenko, S. Sarwana, and A. F. Kirichenko, "Zero static power dissipation biasing of RSFQ circuits," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 776–779, Jun. 2011.
[16] M. Volkmann, A. Sahu, C. Fourie, and O. Mukhanov, "Implementation of energy efficient single flux quantum (eSFQ) digital circuits with sub-aJ/bit operation," *Supercond. Sci. Technol.*, vol. 26, no. 1, pp. 015002-1–015002-28, Jan. 2013.
[17] Y. Yamanashi, T. Nishigai, and N. Yoshikawa, "Study of LR-loading technique for low-power single flux quantum circuits," *IEEE Trans. Appl. Supercond.*, vol. 17, no. 2, pp. 150–153, Jun. 2007.
[18] T. Ortlepp, O. Wetzstein, S. Engert, J. Kunert, and H. Toepfer, "Reduced power consumption in superconducting electronics," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 770–775, Jun. 2011.
[19] M. Tanaka, M. Ito, A. Kitayama, T. Kouketsu, and A. Fujimaki, "18-GHz, 4.0-aJ/bit operation of ultra-low-energy rapid single-flux-quantum shift registers," *Jpn. J. Appl. Phys.*, vol. 51, no. 5, pp. 053102-1–053102-4, May 2012.
[20] M. Tanaka, A. Kitayama, T. Koketsu, M. Ito, and A. Fujimaki, "Low-energy consumption RSFQ circuits driven by low voltages," *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, p. 1701104, Jun. 2013.
[21] K. Senapati, M. G. Blamire, and Z. H. Barber, "Spin-filter Josephson junctions," *Nat. Mater.*, vol. 10, no. 11, pp. 849–852, Sep. 2011.
[22] V. Ryazanov, V. V. Bol'ginov, D. S. Sobanin, I. V. Vernik, S. K. Tolpygo, A. M. Kadin, and O. A. Mukhanov, "Magnetic Josephson junction technology for digital and memory applications," *Phys. Proc.*, vol. 36, pp. 35–41, 2012.
[23] T. I. Larkin, V. V. Bolginov, V. S. Stolyarov, V. V. Ryazanov, I. V. Vernik, S. K. Tolpygo, and O. A. Mukhanov, "Ferromagnetic Josephson switching device with high characteristic voltage," *Appl. Phys. Lett.*, vol. 100, no. 22, pp. 222601-1–222601-5, May 2012.
[24] J. Bulzacchelli, W. Gallagher, and M. Ketchen, "Hybrid superconducting-magnetic memory cell and array," U.S. Patent 8 208 288, Jun. 26, 2012.
[25] A. Herr and Q. Herr, "Josephson magnetic random access memory system and method," U.S. Patent 8 270 209 B2, Sep. 18, 2012.
[26] O. Mukhanov, A. Kadin, I. Nevirkovets, and I. Vernik, "Superconducting devices with ferromagnetic barrier junctions," U.S. Patent Appl. 2012/0 184 445 A1, Jan. 13, 2012.
[27] I. V. Vernik, V. Bol'ginov, S. Bakurskiy, A. Golubov, M. Kupriyanov, V. Ryazanov, and O. Mukhanov, "Magnetic Josephson junctions with superconducting interlayer for cryogenic memory," *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, p. 1701208, Jun. 2013.
[28] B. Nagy, J. D. Farmer, Q. M. Bui, and J. E. Trancik, presented at the Statistical Basis for Predicting Technological Progress, Jul. 2012 arXiv:1207.1463.
[29] J. J. Dongarra and D. W. Walker, "The quest for petascale computing," *Comput. Sci. Eng.*, vol. 3, no. 3, pp. 32–39, May/Jun. 2001.
[30] F. Bedard, N. Welker, G. R. Cotter, M. A. Escavage, and J. T. Pinkston, "Superconducting technology assessment," Nat. Security Agency Office Corp. Assessments, Fort Meade, MD, USA, 2005.
[31] R. Murphy, "On the effects of memory latency and bandwidth on supercomputer application performance," in *Proc. IEEE 10th IISWC*, 2007, pp. 35–43.
[32] P. M. Kogge, "ExaScale Computing Study: Technology Challenges in Achieving Exascale Systems," CSE Dept. Tech., Univ. Notre Dame, Notre Dame, IN, USA, Rep. TR-2008-13, Sep. 28, 2008.
[33] P. Kogge, "The tops in flops," *IEEE Spectrum*, vol. 48, no. 2, pp. 48–54, Feb. 2011. [Online]. Available: http://spectrum.ieee.org/computing/hardware/nextgeneration-supercomputers/0
[34] P. Bunyk, K. Likharev, and D. Zinoviev, "RSFQ technology: Physics and devices," *Int. J. High Speed Electron. Syst.*, vol. 11, no. 1, pp. 257–306, Mar. 2001.
[35] M. Klein and A. Mukherjee, "Thermal noise induced switching of Josephson logic devices," *Appl. Phys. Lett.*, vol. 40, no. 8, pp. 744–747, Apr. 1982.
[36] O. Wetzstein, T. Ortlepp, R. Stolz, J. Kunert, H.-G. Meyer, and H. Toepfer, "Comparison of RSFQ logic cells with and without phase shifting elements by means of BER measurements," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 814–817, Jun. 2011.
[37] M. Dorojevets, P. Bunyk, D. Zinoviev, and K. Likharev, "COOL-0: Design of an RSFQ subsystem for petaflops computing," *IEEE Trans. Appl. Supercond.*, vol. 9, no. 2, pp. 3606–3614, Jun. 1999.
[38] S. Nagasawa, K. Hinode, T. Satoh, Y. Kitagawa, and M. Hidaka, "Design of all-DC-powered high-speed single flux quantum random access memory based on a pipeline structure for memory cell arrays," *Supercond. Sci. Technol.*, vol. 19, no. 5, pp. S325–S330, May 2006.
[39] S. Nagasawa, T. Satoh, K. Hinode, Y. Kitagawa, and M. Hidaka, "Yield evaluation of 10-kA/cm^2 Nb multi-layer fabrication process using conventional superconducting RAMs," *IEEE Trans. Appl. Supercond.*, vol. 17, no. 2, pp. 177–180, Jun. 2007.
[40] K. Fujiwara, L. Qingguo, T. Van Duzer, M. Xiaofan, 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. 14–20, Feb. 2010.
[41] T. Van Duzer, L. Zheng, S. R. Whiteley, H. Kim, X. Meng, and T. Ortlepp, "64-kb hybrid Josephson-CMOS 4 Kelvin RAM with 400 ps access time and 12 mW read power," *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, p. 1 700 504, Jun. 2013.
[42] Y. Arita, N. Yoshikawa, and B. Toshihiko, "Integration of optical waveguides with single flux quantum circuits," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 839–842, Jun. 2011.
[43] Y. Hashimoto, S. Yorozu, and Y. Kameda, "Development of cryopackaging and I/O technologies for high-speed superconductive digital systems," *IEICE Trans. Electron.*, vol. E91.C, no. 3, pp. 325–332, 2008.

[44] A. Inamdar, S. Rylov, S. Sarwana, and D. Gupta, "Superconducting switching amplifiers for high speed digital data links," *IEEE Trans. Appl. Supercon.*, vol. 19, no. 3, pp. 1026–1033, Jun. 2009.

[45] Q. P. Herr, M. S. Wire, and A. D. Smith, "Ballistic SFQ signal propagation on-chip and chip-to-chip," *IEEE Trans. Appl. Supercond.*, vol. 13, no. 2, pp. 463–466, Jun. 2003.

[46] S. B. Kaplan, V. Dotsenko, and D. Tolpygo, "High-speed experimental results for an adhesive-bonded superconducting multi-chip module," *IEEE Trans. Appl. Supercond.*, vol. 17, no. 2, pp. 971–974, Jun. 2007.

[47] O. Mukhanov, HYPRES Inc., Elmsford, NY, USA, private communication, 2012.

[48] Q. Herr and A. Herr, Northrop Grumman Corporation, Baltimore, MD, USA, private communication, 2012.

[49] R. Drost, C. Forrest, B. Guenin, R. Ho, A. V. Krishnamoorthy, D. Cohen, J. E. Cunningham, B. Tourancheau, A. Zingher, A. V, G. Lauterbach, and I. Sutherland, "Challenges in building a flat-bandwidth memory hierarchy for a large-scale computer with proximity communication," in *Proc. 13th Symp. High Perform. Interconnects*, Aug. 2005, pp. 13–22.

[50] P. Chatterjee, "Device scaling: The treadmill that fueled three decades of semiconductor industry growth," *IEEE Solid-State Circuits Newslett.*, vol. 12, no. 1, pp. 14–18, 2007.

[51] R. H. Dennard, F. H. Gaensslen, V. L. Rideout, E. Bassous, and A. R. LeBlanc, "Design of ion-implanted MOSFETs with very small physical dimensions," *IEEE J. Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, Oct. 1974.

[52] A. M. Kadin, C. A. Mancini, M. J. Feldman, and D. K. Brock, "Can RSFQ logic circuits be scaled to deep submicron junctions?" *IEEE Trans. Appl. Supercond.*, vol. 11, no. 1, pp. 1050–1055, Mar. 2001.

[53] P. Bunyk, "RSFQ random logic gate density scaling for the next-generation Josephson junction technology," *IEEE Trans. Appl. Supercond.*, vol. 13, no. 2, pp. 496–497, Jun. 2003.

[54] L. A. Abelson and G. L. Kerber, "Superconductor integrated circuit fabrication technology," *Proc. IEEE*, vol. 92, no. 10, pp. 1517–1533, Oct. 2004.

[55] A. Silver, P. Bunyk, A. Kleinsasser, and J. Spargo, "Vision for single flux quantum very large scale integrated technology," *Supercond. Sci. Technol.*, vol. 19, no. 5, pp. S307–S311, May 2006.

[56] H. J. M. ter Brake, F.-Im. Buchholza, G. Burnell, T. Claeson, D. Crété, P. Febvre, G. J. Gerritsma, H. Hilgenkamp, R. Humphreys, Z. Ivanov, W. Jutzi, M. I. Khabipov, J. Mannhart, H.-G. Meyer, J. Niemeyer, A. Ravex, H. Rogalla, M. Russo, J. Satchell, M. Siegel, H. Töpfer, F. H. Uhlmann, J.-C. Villégier, E. Wikborg, D. Winkler, and A. B. Zorin, "SCENET roadmap for superconductor digital electronics," *Phys. C*, vol. 439, no. 1, pp. 1–41, Jun. 2006.

[57] D. T. Yohannes, A. Inamdar, and S. K. Tolpygo, "Multi-J_c (Josephson critical current density) process for superconductor integrated circuits," *IEEE Trans. Appl. Supercond.*, vol. 19, no. 3, pp. 149–153, Jun. 2009.

[58] S. Anders *et al.*, "European roadmap on superconducting electronics: Status and perspectives," *Phys. C*, vol. 470, pp. 2079–2126, 2010.

[59] M. Tanaka, H. Akaike, A. Fujimaki, Y. Yamanashi, N. Yoshikawa, S. Nagasawa, K. Takagi, and N. Takagi, "100-GHz single-flux-quantum bit-serial adder based on 10-kA/cm^2 niobium process," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 792–796, Jun. 2011.

[60] M. Tanaka, Y. Yamanashi, N. Irie, H.-J. Park, S. Iwasaki, K. Takagi, K. Taketomi, A. Fujimaki, N. Yoshikawa, H. Terai, and S. Yorozu, "Design and implementation of a pipelined 8 bit-serial single-flux-quantum microprocessor with cache memories," *Supercond. Sci. Technol.*, vol. 20, no. 11, pp. S305–S309, Nov. 2007.

[61] V. Patel and J. E. Lukens, "Self-shunted Nb/AlO_x/Nb Josephson junctions," *IEEE Trans. Appl. Supercond.*, vol. 9, no. 2, pp. 3247–3250, Jun. 1999.

[62] G. L. Kerber, A. W. Kleinsasser, and B. Bumble, "Fabrication of submicrometer high current density Nb/Al $-$ AlN_x/Nb junctions," *IEEE Trans. Appl. Supercond.*, vol. 19, no. 3, pp. 159–166, Jun. 2009.

[63] D. Olaya, P. D. Dresselhaus, S. P. Benz, A. Herr, Q. P. Herr, A. G. Ioannidis, D. L. Miller, and A. W. Kleinsasser, "Digital circuits using self-shunted Nb/Nb_xSi_{1-x}/Nb Josephson junctions," *Appl. Phys. Lett.*, vol. 96, no. 21, pp. 213510-1–213510-3, May 2010.

[64] V. Belitsky, C. Risacher, M. Pantaleev, and V. Vassilev, "Superconducting microstrip line model studies at millimetre and sub-millimetre waves," *Int. J. Infrared Millimeter Waves*, vol. 27, no. 6, pp. 809–834, Jun. 2006.

[65] Y. Mizugaki, A. Kawai, R. Kashiwa, M. Moriya, and T. Kobayashi, "Analytical inductance calculation of superconducting stripline by use of transformation into perfect conductor model," *IEICE Trans. Electron.*, vol. E93.C, no. 4, pp. 486–488, 2010.

[66] A. J. Annunziata, D. F. Santavicca, L. Frunzio, G. Catelani, M. J. Rooks, A. Frydman, and D. E. Prober, "Tunable superconducting nanoinductors," *Nanotechnology*, vol. 21, no. 44, pp. 445202-1–445202-6, Nov. 2010.

[67] T. Ortlepp, M. O. Ariando, C. J. M. Verwijs, K. F. K. Foo, A. Andreski, H. Rogalla, F. H. Uhlmann, and H. Hilgenkamp, "RSFQ circuitry using intrinsic π-phase shifts," *IEEE Trans. Appl. Supercond.*, vol. 17, pp. 659–663, Jun. 2007.

[68] A. Andreski, "Superconducting circuits with π-shift elements," Ph.D. dissertation, Univ. Twente, Enschede, The Netherlands, Sep., 2011.

[69] International Technology Roadmap for Semiconductors (ITRS) Working Group Models. [Online]. Available: http://www.itrs.net/models.html

[70] D. S. Holmes, C. L. Lau, and M. A. Manheimer, "Modeling superconducting electronic circuit fabrication for roadmap development," unpublished.

[71] K. Gaj, Q. P. Herr, V. Adler, A. Krasniewski, E. G. Friedman, and M. J. Feldman, "Tools for the computer-aided design of multi-gigahertz superconducting digital circuits," *IEEE Trans. Appl. Supercond.*, vol. 9, no. 1, pp. 18–38, Mar. 1999.

[72] K. Gaj, Q. P. Herr, V. Adler, D. K. Brock, E. G. Friedman, and M. J. Feldman, "Toward a systematic design methodology for large multi-gigahertz rapid single flux quantum circuits," *IEEE Trans. Appl. Supercond.*, vol. 9, no. 3, pp. 4591–4606, Sep. 1999.

[73] S. Yorozu, Y. Kameda, H. Terai, A. Fujimaki, T. Yamada, and S. Tahara, "A single flux quantum standard logic cell library," *Phys. C, Supercond.*, vol. 378–381, pp. 1471–1474, Oct. 2002.

[74] C. Fourie, "A tool kit for the design of superconducting programmable gate arrays," Ph.D. dissertation, Dept. Electron. Eng., Stellenbosch Univ., Stellenbosch, South Africa, 2003.

[75] N. Yoshikawa, K. Yoda, H. Hoshina, K. Kawasaki, K. Fujiwara, F. Matsuzaki, and N. Nakajima, "Cell based design methodology for BDD SFQ logic circuits: A high speed test and feasibility for large scale circuit applications," *IEEE Trans. Appl. Supercond.*, vol. 13, no. 2, pp. 523–526, Jun. 2003.

[76] Z.-H. Zhang and B.-R. Guan, "Design and simulation of RSFQ/RISC computer system," *PIERS Online*, vol. 1, no. 5, pp. 583–586, 2005.

[77] H. Toepfer and T. Ortlepp, "Design infrastructure for rapid single flux quantum circuits," *Cryogenics*, vol. 49, no. 11, pp. 643–647, Nov. 2009.

[78] M. Tanaka, K. Obata, Y. Ito, S. Takeshima, M. Sato, K. Takagi, N. Takagi, H. Akaike, and A. Fujimaki, "Automated passive-transmission-line routing tool for single-flux-quantum circuits based on A^* algorithm," *IEICE Trans. Electron.*, vol. E93.C, no. 4, pp. 435–439, 2010.

[79] F. G. Ortmann, A. van der Merwe, H. R. Gerber, and C. J. Fourie, "A comparison of multi-criteria evaluation methods for RSFQ circuit optimization," *IEEE Trans. Appl. Supercond.*, vol. 21, no. 3, pp. 801–804, Jun. 2011.

[80] C. J. Fourie, O. Wetzstein, T. Ortlepp, and J. Kunert, "Three-dimensional multi-terminal superconductive integrated circuit inductance extraction," *Supercond. Sci. Technol.*, vol. 24, no. 12, pp. 125015-1–125015-12, Dec. 2011.

[81] F. Mehdipour, H. Honda, K. Inoue, and K. Murakami, "Hardware and software requirements for implementing a high-performance superconductivity circuits-based accelerator," in *Proc. 3rd ASQED*, 2011, pp. 229–235.

[82] Y. Kameda, S. Yorozu, and Y. Hashimoto, "Automatic single-flux-quantum (SFQ) logic synthesis method for top-down circuit design," *J. Phys., Conf. Ser.*, vol. 43, no. 1, pp. 1179–1182, Jan. 2006.

[83] N. Yoshikawa and J. Koshiyama, "Top down RSFQ logic design based on a binary decision diagram," *IEEE Trans. Appl. Supercond.*, vol. 11, no. 1, pp. 1098–1101, Mar. 2011.

[84] O. Oberg, "Superconducting logic circuits operating with reciprocal magnetic flux quanta," Ph.D. dissertation, Dept. Phys., Univ. Maryland, College Park, MD, USA, 2011.

[85] A. L. Ripple, D. S. Holmes and M. A. Manheimer, "Computer-aided design tools for superconducting integrated circuit design," unpublished.

[86] C. J. Fourie, O. Wetzstein, J. Kunert, and H.-G. Meyer, "SFQ circuits with ground plane hole-assisted inductive coupling designed with InductEx," *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, p. 1 300 705, Jun. 2013.

[87] C. J. Fourie and M. H. Volkmann, "Status of superconductor electronic circuit design software," *IEEE Trans. Appl. Supercond.*, vol. 23, no. 3, p. 1 300 205, Jun. 2013.

[88] S. Narayana, V. K. Semenov, Y. A. Polyakov, V. Dotsenko, and S. K. Tolpygo, "Design and testing of high-speed interconnects for superconducting multi-chip modules," *Supercond. Sci. Technol.*, vol. 25, pp. 105012-1–105012-16, Oct. 2012.

[89] R. J. Webber and J. Delmas, "Operation of YBCO current leads as bias lines to cryocooler-mounted 4 K superconducting electronics," *Phys. Proc.*, vol. 36, pp. 256–261, 2012.