Superconducting Computing and the IARPA C3 Program

Dr. D. Scott Holmes, Booz Allen Hamilton
IARPA C3 Program Technical Support

Beyond CMOS Workshop, 2016 April 4-5
Superconducting Computing and the IARPA C3 Program

• Problems in large scale computing
• Feasibility study
• Cryogenic Computing Complexity (C3) program (IARPA)
• Challenges
• Related work
• Conclusions
Problem: Increasing power requirements for conventional supercomputers
Facebook Data Center, Luleå, Sweden

- 2014 completion target
- Cost: ~760 M$
- Nearby Lule River generates 9% of Sweden's electricity (~4.23 GW)
- Average annual temperature: 1.3 °C

Specifications

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance</strong></td>
<td>27-51 PFLOP/s</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td>21-27 PB RAM</td>
</tr>
<tr>
<td></td>
<td>1900-6800 PB disk</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td>84 MW avg* (120 MW max)</td>
</tr>
<tr>
<td><strong>Space</strong></td>
<td>290,000 ft² (27,000 m²)</td>
</tr>
<tr>
<td><strong>Cooling</strong></td>
<td>~1.07 PUE</td>
</tr>
</tbody>
</table>

* estimated
Facebook Data Center, Luleå, Sweden

- 2014 completion target
- Cost: ~760 M$
- Nearby Lule River generates 9% of Sweden's electricity (~4.23 GW)
- Average annual temperature: 1.3 °C

Specifications

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>27-51 PFLOP/s</td>
</tr>
<tr>
<td>Memory</td>
<td>21-27 PB RAM</td>
</tr>
<tr>
<td></td>
<td>1900-6800 PB disk</td>
</tr>
<tr>
<td>Power</td>
<td>84 MW avg* (120 MW max)</td>
</tr>
<tr>
<td>Space</td>
<td>290,000 ft² (27,000 m²)</td>
</tr>
<tr>
<td>Cooling</td>
<td>~1.07 PUE</td>
</tr>
</tbody>
</table>

* estimated
The Problem is Power-Space-Cooling

• Upgrading a facility to more powerful computers is constrained by
  – *Power* supply capability of electric company
  – *Space* limitations
  – *Cooling* infrastructure

• Constraints on developing computers with additional processing power
  – Some estimates to reach exascale are in the hundreds of megawatts.
  – An exascale computer at 20 megawatts based on semiconducting technology will require heroic measures.
  – We will require a different technology to get beyond exascale.
Superconducting Computing and the IARPA C3 Program

- Problems in large scale computing

- Feasibility study
  - Cryogenic Computing Complexity (C3) program (IARPA)
  - Challenges
  - Related work
  - Conclusions
**Goal: Computation Efficiency**

![Graph showing computation efficiency over time](image)

- **Goals**
- **Projected**
- **Green500 #1**
- **Top500 #1**

*As of: 2012-11*

**Exascale goal (DoE), 20 MW**

**Exascale Strawman (DARPA), 68 → 500 MW**

**Titan, Sequoia, K, Japan, Tianhe-1A, China, Jaguar**

Alternative Technology Goal
Alternative Technologies

- Superconducting SFQ (single flux quantum) looks good based on switching energy-delay, but:
  - Refrigeration requires x400 to x5000 energy
  - Wiring + leakage losses dominate for other technologies

- Conclusions:
  - Full system evaluation is required for SFQ
  - Better metrics and figures-of-merit needed!

https://www.nsa.gov/research/tnw/tnw203/article2.shtml
Superconducting Computing Approach

- Low temperature operation (~4 K)
  - Allows different physics
  - Commercially available refrigeration
- Logic
  - SFQ (Single Flux Quantum)
  - Switching energy ~ $2 \times 10^{-20}$ J
- Memory
  - compatible with SFQ logic
- Interconnects
  - Superconducting in the cold space
  - Input/Output: electrical or optical
- Major energy reductions in all 3 areas!
Feasibility Study

- Superconducting supercomputer evaluation using:
  - **Near-zero energy** 4 K interconnect
  - **New** SFQ logic with no static power dissipation
  - **New** cryogenic memory ideas
  - **Electrical or optical** input and output
  - **Commercial** cryogenic refrigerators
### System Power Budgets

<table>
<thead>
<tr>
<th>Performance (PFLOP/s):</th>
<th>1</th>
<th>10</th>
<th>100</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>(qty) Refrigeration system, LHe recondensing (cost)</td>
<td>(1) SHI model SRDK-415D-F50 (45 k$)</td>
<td>(2) SHI GM-JT CG310SLCR (320 k$)</td>
<td>(1) Linde LR70 helium liquefier (2 M$)</td>
<td>(2) Linde LR280 helium liquefier (6 M$)</td>
</tr>
<tr>
<td>• Heat load at cold end *</td>
<td><strong>1.5 W @ 4.2 K</strong></td>
<td><strong>10 W @ 4.3 K</strong></td>
<td><strong>100 W @ 4.4 K</strong></td>
<td><strong>1,020 W @ 4.4 K</strong></td>
</tr>
<tr>
<td>• Cooling penalty (300 K/4 K)</td>
<td>5000 w/w</td>
<td>1280 w/w</td>
<td>450 w/w</td>
<td>395 w/w</td>
</tr>
<tr>
<td><strong>Power – cooling (60 Hz)</strong></td>
<td><strong>7.5 kW</strong></td>
<td><strong>12.8 kW</strong></td>
<td><strong>45 kW</strong></td>
<td><strong>400 kW</strong></td>
</tr>
<tr>
<td><strong>Power – everything else</strong></td>
<td><strong>7.5 kW</strong></td>
<td><strong>27.2 kW</strong></td>
<td><strong>155 kW</strong></td>
<td><strong>1,600 kW</strong></td>
</tr>
<tr>
<td><strong>Total system power</strong></td>
<td><strong>15 kW</strong></td>
<td><strong>40 kW</strong></td>
<td><strong>200 kW</strong></td>
<td><strong>2,000 kW</strong></td>
</tr>
<tr>
<td>• Computation efficiency (goal: ≥ 5 x 10^{11} FLOPS/W)</td>
<td><strong>0.7 x 10^{11} FLOPS/W</strong></td>
<td><strong>2.5 x 10^{11} FLOPS/W</strong></td>
<td><strong>5 x 10^{11} FLOPS/W</strong></td>
<td><strong>5 x 10^{11} FLOPS/W</strong></td>
</tr>
</tbody>
</table>

* typical cold end heat load budget

- Toughest problem: scaling uncooled components, primarily disk storage memory
Memory Power @ 4 K

- Cryogenic operation preferred: close to logic, so low latency
  - Superconducting interconnects for speed, low energy loss

<table>
<thead>
<tr>
<th>Performance (PFLOP/s)</th>
<th>1</th>
<th>10</th>
<th>100</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main memory (1 B-s/FLOP)</td>
<td>1 PB</td>
<td>10 PB</td>
<td>100 PB</td>
<td>1,000 PB</td>
</tr>
<tr>
<td>Power budget (@ 4 K)</td>
<td>1.5 w</td>
<td>10 w</td>
<td>100 w</td>
<td>1,000 w</td>
</tr>
<tr>
<td>• target for memory</td>
<td>0.4 W</td>
<td>3 W</td>
<td>30 W</td>
<td>300 W</td>
</tr>
<tr>
<td>Hybrid JJ-CMOS</td>
<td>19 MW</td>
<td>190 MW</td>
<td>1.9 GW</td>
<td>19 GW</td>
</tr>
<tr>
<td>(Van Duzer 2007), 20 mW/MiB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pipelined dc SFQ RAM</td>
<td>86 MW</td>
<td>860 MW</td>
<td>8.6 GW</td>
<td>86 GW</td>
</tr>
<tr>
<td>(SRL 2005), 700 μW/8 KiB</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(100 μV power supply)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Josephson MRAM</td>
<td>0.46 w</td>
<td>4.6 w</td>
<td>46 w</td>
<td>460 w</td>
</tr>
<tr>
<td>(NG proposal) 1 B/FLOP, R/W: 5/160 aJ/bit</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


**JMRAM approach might work**
### System feasibility study: power @ 4 K

<table>
<thead>
<tr>
<th>Component</th>
<th>Performance (PFLOP/s):</th>
<th>1</th>
<th>10</th>
<th>100</th>
<th>1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power budget (@ 4 K)</td>
<td></td>
<td>1.5 W</td>
<td>10 W</td>
<td>100 W</td>
<td>1,000 W</td>
</tr>
<tr>
<td>Logic (RQL, Ic = 25 μA, 8.3 GHz)</td>
<td>0.18 W 0.18 W</td>
<td>1.8 W</td>
<td>18 W</td>
<td>180 W</td>
<td></td>
</tr>
<tr>
<td>Memory (1 B/FLOP, JMRAM)</td>
<td>0.46 W 0.46 W</td>
<td>4.6 W</td>
<td>46 W</td>
<td>460 W</td>
<td></td>
</tr>
<tr>
<td>Interconnects (VCSELs @ 40 K)</td>
<td>0.1 W 0.1 W</td>
<td>1 W</td>
<td>10 W</td>
<td>100 W</td>
<td></td>
</tr>
<tr>
<td>Other (structure, radiation heat leaks)</td>
<td>0.76 W 0.76 W</td>
<td>2.6 W</td>
<td>26 W</td>
<td>260 W</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1.5 W</td>
<td>10 W</td>
<td>100 W</td>
<td>1,000 W</td>
</tr>
</tbody>
</table>

- Computation efficiency (goal: ≥ 5 x 10¹¹ FLOPS/W)
  - 0.7 x 10¹¹ FLOPS/W
  - 2.5 x 10¹¹ FLOPS/W
  - 5 x 10¹¹ FLOPS/W
  - 5 x 10¹¹ FLOPS/W

### Conclusions:
- Energy-efficient superconducting computers are possible
- Priorities: **Memory → Logic → System → Interconnects**
Superconducting computing looks promising

Exascale Power Comparison

Conventional computer at 100 MW

- Interconnect
- Memory
- Logic
- Leakage

Superconducting computer at 2 MW

Biggest win is superconducting interconnect!
## Conceptual System Comparison (~20 PFLOP/s)

<table>
<thead>
<tr>
<th></th>
<th>Titan at ORNL</th>
<th>Superconductor Supercomputer</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Performance</strong></td>
<td>17.6 PFLOP/s (#2 in world*)</td>
<td>20 PFLOP/s</td>
</tr>
<tr>
<td><strong>Memory</strong></td>
<td>710 TB (0.04 B/FLOPS)</td>
<td>5 PB (0.25 B/FLOPS)</td>
</tr>
<tr>
<td><strong>Power</strong></td>
<td>8,200 kW avg. (not included: cooling, storage memory)</td>
<td>80 kW total power (includes cooling)</td>
</tr>
<tr>
<td><strong>Space</strong></td>
<td>4,350 ft² (404 m², not including cooling)</td>
<td>~200 ft² (19 m², includes cooling)</td>
</tr>
<tr>
<td><strong>Cooling</strong></td>
<td>additional power, space and infrastructure required</td>
<td>All cooling shown</td>
</tr>
</tbody>
</table>

* TOP500, 2015-11

**Titan at ORNL**
- Performance: 17.6 PFLOP/s (#2 in world*)
- Memory: 710 TB (0.04 B/FLOPS)
- Power: 8,200 kW avg. (not included: cooling, storage memory)
- Space: 4,350 ft² (404 m², not including cooling)
- Cooling: additional power, space and infrastructure required

**Superconductor Supercomputer**
- Performance: 20 PFLOP/s
- Memory: 5 PB (0.25 B/FLOPS)
- Power: 80 kW total power (includes cooling)
- Space: ~200 ft² (19 m², includes cooling)
- Cooling: All cooling shown

* COURTESY OF THE OAK RIDGE NATIONAL LABORATORY, U.S. DEPARTMENT OF ENERGY

* COURTESY OF IARPA
Superconducting Computing and the IARPA C3 Program

• Problems in large scale computing
• Feasibility study

• Cryogenic Computing Complexity (C3) program (IARPA)
  • Challenges
  • Related work
  • Conclusions
Cryogenic Computing Complexity (C3) program goal

Develop technologies for a computer based on superconducting logic with cryogenic memory, and

Integrate a prototype that can answer these questions:

1) Can we build a superconducting computer that is capable of solving important problems?

2) Does it provide a sufficient advantage over conventional computing that we want to build it?
C3 development areas

- Approach based on:
  - *Near-zero energy* superconducting interconnect
  - *New SFQ logic* with no static power dissipation
  - *New energy efficient* cryogenic memory ideas
  - *Electrical or optical* inputs and outputs
  - *Commercial* cryogenic refrigerators

IARPA C3 program basis
Notional Prototype

<table>
<thead>
<tr>
<th>Metric</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock rate for superconducting logic</td>
<td>10 GHz</td>
</tr>
<tr>
<td>Throughput (bit-op/s)</td>
<td>$10^{13}$</td>
</tr>
<tr>
<td>Efficiency @ 4 K (bit-op/J)</td>
<td>$10^{15}$</td>
</tr>
<tr>
<td>CPU count</td>
<td>1</td>
</tr>
<tr>
<td>Word size (bit)</td>
<td>64</td>
</tr>
<tr>
<td>Parallel Accelerator count</td>
<td>2</td>
</tr>
<tr>
<td>Main Memory (B)</td>
<td>$2^{28}$</td>
</tr>
<tr>
<td>Input/Output (bit/s)</td>
<td>$10^9$</td>
</tr>
</tbody>
</table>
## Program Metrics and Goals

<table>
<thead>
<tr>
<th>Metric</th>
<th>BP</th>
<th>OP1</th>
<th>OP2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cryogenic Memory</strong></td>
<td>Memory</td>
<td>Array</td>
<td>Chip</td>
</tr>
<tr>
<td>Functional capacity (bit)*</td>
<td>1</td>
<td>2^6 ; 2^6</td>
<td>2^10 ; 2^10</td>
</tr>
<tr>
<td>Density (bit/cm^2)*</td>
<td>10^6 ; 10^5</td>
<td>5x10^6 ; 5x10^5</td>
<td>10^7 ; 10^6</td>
</tr>
<tr>
<td>Data rate, burst mode (Gbit/s)*</td>
<td>1</td>
<td>5 ; 30</td>
<td>5 ; 30</td>
</tr>
<tr>
<td>Access time, ave. (ps)*</td>
<td>10,000 ; 1,000</td>
<td>5,000 ; 400</td>
<td>5,000 ; 400</td>
</tr>
<tr>
<td>Access energy, ave. (J/bit)*</td>
<td>5x10^{-16} ; 5x10^{-17}</td>
<td>5x10^{-16} ; 5x10^{-17}</td>
<td>10^{-16} ; 10^{-17}</td>
</tr>
<tr>
<td>Logic, Comm. &amp; Systems</td>
<td>Subcircuits</td>
<td>Circuits</td>
<td>Processors</td>
</tr>
<tr>
<td>Benchmark circuits &amp; applications</td>
<td>Circuits 1</td>
<td>Circuits 2</td>
<td>Circuits 3</td>
</tr>
<tr>
<td>Complexity (JJ)</td>
<td>10^4</td>
<td>5x10^4</td>
<td>10^5</td>
</tr>
<tr>
<td>Density (JJ/cm^2)</td>
<td>10^5</td>
<td>5x10^5</td>
<td>10^6</td>
</tr>
<tr>
<td>Throughput (bit-op/s)</td>
<td>10^9</td>
<td>5x10^{10}</td>
<td>10^{11}</td>
</tr>
<tr>
<td>Efficiency @ 4 K (bit-op/J)</td>
<td>10^{16}</td>
<td>5x10^{16}</td>
<td>10^{17}</td>
</tr>
</tbody>
</table>

* Memory metrics: The first number refers to Main Memory and the second to Cache Memory.
C3 Program Organization

Phase 1
- LCS Thrust: Logic, Communications & Systems
- CM Thrust: Cryogenic Memory

Phase 2
- Logic, Communications & Systems
- Chips

Flow between Government Furnished Foundry and Performer teams:
- Design rules, test chips
- GDS files with test chip designs
- GDS files with chip designs
- Chips for test and integration
### C3 Program Schedule

<table>
<thead>
<tr>
<th>Period</th>
<th>Management Milestone</th>
<th>Month</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phase 1 Base Period (BP)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kickoff</td>
<td></td>
<td>0</td>
<td>07/29/2014</td>
</tr>
<tr>
<td>Start Date</td>
<td></td>
<td>1</td>
<td>09/01/2014</td>
</tr>
<tr>
<td>PI Meeting &amp; Workshop</td>
<td></td>
<td>6</td>
<td>02/18/2015</td>
</tr>
<tr>
<td>PI Meeting &amp; Workshop</td>
<td></td>
<td>12</td>
<td>08/11/2015</td>
</tr>
<tr>
<td>Program Review (decision on OP1)</td>
<td></td>
<td>14</td>
<td>11/05/2015</td>
</tr>
<tr>
<td><strong>Phase 1 Option Period 1 (OP1)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI Meeting &amp; Workshop</td>
<td></td>
<td>18</td>
<td>03/10/2016</td>
</tr>
<tr>
<td>PI Meeting &amp; Workshop</td>
<td></td>
<td>24</td>
<td>08/10/2016</td>
</tr>
<tr>
<td>Program Review (decision on OP2)</td>
<td></td>
<td>29</td>
<td>11/15/2016</td>
</tr>
<tr>
<td><strong>Phase 1 Option Period 2 (OP2)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI Meeting &amp; Workshop</td>
<td></td>
<td>32</td>
<td>02/15/2017</td>
</tr>
<tr>
<td>Decision on Phase 2 BAA release</td>
<td></td>
<td>34</td>
<td>04/15/2017</td>
</tr>
<tr>
<td>PI Meeting &amp; Workshop</td>
<td></td>
<td>38</td>
<td>08/15/2017</td>
</tr>
<tr>
<td>Program Review (decision on Phase 2 follow-on)</td>
<td></td>
<td>41</td>
<td>11/15/2017</td>
</tr>
</tbody>
</table>

- Marc Manheimer, Program Manager
C3 Government Team

- MIT Lincoln Laboratory
  - Superconducting electronics foundry
  - Process development

- Sandia National Laboratories
  - Failure analysis
  - Superconductivity expertise

- NIST Boulder
  - Test & evaluation of performer circuits
  - Government program support

- NASA-JPL (contract pending)
  - Tunnel junction improvement
### MIT-LL Superconducting Technology Roadmap

<table>
<thead>
<tr>
<th>Fabrication Process Attribute</th>
<th>Units</th>
<th>SFQ3ee</th>
<th>SFQ4ee</th>
<th>SFQ5ee</th>
<th>SFQ6ee</th>
<th>SFQ7ee</th>
<th>SFQ8ee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical current density</td>
<td>MA/m²</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>JJ diameter (surround)</td>
<td>nm</td>
<td>700 (500)</td>
<td>700 (500)</td>
<td>700 (300)</td>
<td>700 (300)</td>
<td>500 (200)</td>
<td>500 (200)</td>
</tr>
<tr>
<td>Nb metal layers</td>
<td>-</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Line width (space)</td>
<td>Critical layers</td>
<td>nm</td>
<td>500 (1000)</td>
<td>500 (700)</td>
<td>350 (500)</td>
<td>350 (500)</td>
<td>250 (300)</td>
</tr>
<tr>
<td></td>
<td>Other layers</td>
<td>nm</td>
<td>500 (700)</td>
<td>500 (700)</td>
<td>350 (500)</td>
<td>250 (300)</td>
<td>180 (180)</td>
</tr>
<tr>
<td>Metal thickness</td>
<td>nm</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>150</td>
</tr>
<tr>
<td>Dielectric thickness</td>
<td>nm</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>200</td>
<td>180</td>
</tr>
<tr>
<td>Resistor width (space)</td>
<td>nm</td>
<td>1000 (2000)</td>
<td>500 (700)</td>
<td>500 (700)</td>
<td>500 (700)</td>
<td>500 (500)</td>
<td>350 (350)</td>
</tr>
<tr>
<td>Shunt resistor value</td>
<td>Ω/sq</td>
<td>2</td>
<td>2</td>
<td>2 or 6</td>
<td>2 or 6</td>
<td>2 or 6</td>
<td>2 or 6</td>
</tr>
<tr>
<td>mΩ resistor</td>
<td>mΩ</td>
<td>-</td>
<td>-</td>
<td>3 - 10</td>
<td>3 - 10</td>
<td>3 - 10</td>
<td>3 - 10</td>
</tr>
<tr>
<td>High kinetic inductance layer</td>
<td>pH/sq</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Via diameter (surround)</td>
<td>nm</td>
<td>700 (500)</td>
<td>700 (500)</td>
<td>500 (350)</td>
<td>500 (350)</td>
<td>350 (250)</td>
<td>350 (200)</td>
</tr>
<tr>
<td>Via type, stacking</td>
<td>-</td>
<td>Etched, Staggered</td>
<td>Etched, Stacked (1/2)</td>
<td>Etched, Stacked (1/2)</td>
<td>Etched, Stacked (1/2)</td>
<td>Stud, Stacked</td>
<td>Stud, Stacked</td>
</tr>
<tr>
<td>Early access availability</td>
<td>-</td>
<td>2014-09</td>
<td>2015-09</td>
<td>2016-03</td>
<td>2016-09</td>
<td>2017-09</td>
<td></td>
</tr>
</tbody>
</table>

**Changes from the previous process**

- Nb/Al-AlOx/Nb JJ technology
- 200 mm Si wafers, full planarization
SFQ7ee process node (2016)

- 10 Nb layers, 250 nm minimum wire width, 300 nm spacing
- 500 nm minimum junction dia., $J_C$ target: 100 $\mu$A/$\mu$m$^2$
Logic Communications & Systems: 2 Projects

- **IBM** team (Gerald Gibson)
  - Architecture: many lightweight processors
  - ERSFQ/eSFQ circuit families (Hypres)
    - DC power, zero static power dissipation
    - “1”: single SFQ pulse; “0”: no pulse

- **Northrop Grumman** team (Quentin Herr)
  - Architecture: RISC processor
  - Reciprocal quantum logic (RQL) circuit family
    - RF clock and power (~ 10 GHz)
    - “1”: pair of reciprocal SFQ pulses; “0”: no pulses
Energy-Efficient SFQ (eSFQ, ERSFQ)

- SFQ logic families developed at Hypres
  - DC supply current, zero static power dissipation using a JJ and large inductor in series
  - Clocked by SFQ pulses distributed separately
  - ERSFQ and eSFQ can be combined in a single circuit

- Demonstration circuits:
  - ERSFQ 8-bit adder (2011)
  - 0.8 aJ/bit eSFQ circuit
415 GHz maximum operating frequency of a toggle flip-flop (TFF) circuit was demonstrated, a record for Josephson junctions with critical current density $J_c = 135$ MA/m$^2$.

ERSFQ is an energy-efficient superconducting digital circuit family, proving that both high speed and high efficiency are possible in the same circuit.
RQL 8-bit Kogge-Stone Adder

- Carry look ahead adder (CLA) core
  - Excluding: input shift registers, output amps
  - Fabricated in Hypres 45 MA/m² process

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>JJs</td>
<td>815</td>
</tr>
<tr>
<td>Ave $I_C$</td>
<td>162 $\mu$A</td>
</tr>
<tr>
<td>Clock</td>
<td>6.2 GHz</td>
</tr>
<tr>
<td>Power</td>
<td>560 nW @ 6.2 GHz with 1.2 mW clock</td>
</tr>
<tr>
<td>Energy</td>
<td>$\sim$5x10⁻⁵ pJ</td>
</tr>
<tr>
<td>Latency</td>
<td>150 ps @ 10 GHz</td>
</tr>
<tr>
<td>Area</td>
<td>$\sim$ 4 mm²</td>
</tr>
</tbody>
</table>

- Projected 64-bit adder in advanced process: 100 ps latency at 20 GHz with 1.2 mW clock

Cryogenic Memory: 2 Projects

- **Raytheon BBN** team (Tom Ohki)
  - Orthogonal STT memory (Andrew Kent, NYU)
  - Spin Hall effect memory (Robert Buhrman, Cornell U)
  - nTron drivers or memory elements (Karl Berggren, MIT)
  - eSFQ/ERSFQ logic circuits (Hypres)

- **Northrop Grumman** team (Anna Herr)
  - Spin valve devices (Nathan Newman, ASU)
  - Spin valve materials (Norman Birge, MSU)
  - Fabrication process development (Jiwei Lu, UVA)
  - RQL logic circuits (NGC)
COSTRAM: Cryogenic Orthogonal Spin Transfer RAM

- Raytheon BBN, NYU (Andy Kent), Hypres
- Memory element
  - Write current through element, not field lines
  - 3 to 4% MR
  - No thermal fluctuations required
  - <100 ps switching times
- Logic
  - ERSFQ or eSFQ (Hypres)


Optimization for Low Temperature

- OST-MRAM operates at low temperature: thermal fluctuations are not needed to initiate switching events
- Current required to write bit cell at low temperature is greatly reduced: $I_f \propto U$ for long-term stability: $U/kT > 60$

Write energy estimate: $E = I^2 R T \sim 10^{-19}$ J
Read energy estimate: $E = 10^{-20}$ J
JMRAM: Josephson MRAM

- Memory cell: Josephson junctions with magnetic tunnel barriers
  - Physics demonstrated in SFS JJs
  - Memory state: high or low $I_c$
  - Write: magnetic field switching of the “soft” magnetic layer
  - Read: Josephson effect
- No static power dissipation
- Read energy dissipated only for logical “1”
Superconducting Computing and the IARPA C3 Program

- Problems in large scale computing
- Feasibility study
- Cryogenic Computing Complexity (C3) program (IARPA)

Challenges

- Related work
- Conclusions
Challenges

• Fabrication

• Memory
  – Magnetic memory device physics and fabrication
  – Cell size

• Superconducting circuit physics
  – Coupled current and inductance
  – Low gain from JJs

• Electronic design automation (EDA) tools

• Input/Output
  – Optical needed
  – High-speed, low heat leak

• Circuit complexity, density

• Cryogenic and (non-) magnetic environment
  – Materials and packaging
  – Refrigeration

} Subject of next talk

} Future
Challenges: Memory

- Variety of device types with no clear winner
  - Magnetic spin valve (SV), spin Hall effect (SHE), spin-transfer torque (STT), ...
  - New materials: NiFeMo, NiFeCu, NiFeNb, Co/Ru/Co, [Co/Ni]_n, ...
  - New physics combining spintronics and superconductivity

- Process control
  - Low energy operation and superconductivity demand thin layers
  - Thickness (~ 1 nm) exponentially affects device parameters
  - Interfacial roughness degrades properties and increases spreads
  - Properties change from room temperature to 4 K

- Memory cell size
  - Multiple devices per cell
  - No equivalent to DRAM yet
Challenges: Superconducting Circuits

• Inductance \( (L) \) and critical current \( (I_c) \) are linked for SFQ circuits
  – SFQ: Single Flux Quantum, \( \Phi_0 = 2.07 \) fWb (mV⋅ps or mA⋅pH)
  – Flux: \( I \cdot L = \Phi = \alpha \Phi_0 \approx \Phi_0 \) \textit{required for circuit operation!}
  – Switching energy: \( E_{sw} = I_c \cdot \Phi_0 \)
  – Decreasing \( I_c \) (more energy efficient!) requires increasing \( L \)

• Low gain
  – Fan-out > 1 requires splitters
  – Low gain from Josephson junctions (3 JJs to make a 2:1 splitter)
  – Low isolation across JJs
  – Alternatives are being explored

3-terminal nanowire switch
S-F transistor

SFQ gate: set-reset (SR) flip-flop

Single Flux Quantum (SFQ)

\[ I \cdot L = \Phi_0 = 2.07 \text{ mA} \cdot \text{pH} \]

\[ \mathbf{J}_0 \]

\[ \mathbf{I}_{\text{bias}} \]

\[ \mathbf{J}_1 \]

\[ \mathbf{J}_2 \]

\[ \mathbf{J}_3 \]

\[ \mathbf{A} \]

\[ \mathbf{B} \]

\[ \mathbf{C} \]

\[ \mathbf{D} \]

\[ \mathbf{E} \]

Transfer Block

Storage Block

Decision Block

\[ \sim 1 \text{ mV} \]

\[ \sim 2 \text{ ps} \]
Challenges: EDA and Designers

• Electronic design automation (EDA) tools missing or needing improvement:
  - Inductance extraction
  - Placement and routing (inductance range)
  - Synthesis
  - Timing analysis
  - Simulation
  - PDKs (not provided by foundry!)

• SFQ designers
  - No classes in the USA
  - Few designers

“Manual design is like making lace.” (SFQ designer)
SCE EDA Tool Status

- Design, analysis, and verification tools need to be developed into a comprehensive EDA tool set specific to very large scale superconducting integrated circuits.

SCE = Superconducting Electronics
EDA = Electronic Design Automation
HDL = Hardware Description Language
P&R = Place-and-Route
DRC = Design Rule Check
LVS = Layout versus Schematic
TCAD = Technology CAD
SuperTools Program (IARPA)

- **Request for Information** (Posted 2015-01-12)
  Electronic Design Automation tools for Superconducting Electronics (EDA for SCE)

- **Proposers’ Day** (2016-02-10)

- Broad Agency Announcement (BAA) in process

- Mark Heiligman, Program Manager
Cryo-to-Room Temperature Interconnects

- Electrical interconnects allow heat into the cryogenic environment

- Optical interconnects are preferred
  - SFQ energy: $E_{SFQ} = I_c \cdot \Phi_0 = (0.1 \text{ mA})(2.07 \text{ mV} \cdot \text{ps}) = 2.1 \times 10^{-19} \text{ J} = 0.21 \text{ aJ}$
  - Photon energy: $E_{ph} = \frac{hc}{\lambda} = (6.6e-34 \text{ Js})(3e+8 \text{ m/s})/(1550 \text{ nm}) = 0.13 \text{ aJ}$
  - In is easier, as there are typically many photons per bit
  - Out is ‘IARPA hard’

- Request for Information IARPA-RFI-16-04 (posted 2016-01-28)
  Data Ingress and Egress for Cryogenic Systems

Complexity: SFQ lags semiconductors by $\sim 10^5$
Complexity: SFQ lags semiconductors by $\sim 10^5$

Exponential growth in MONEY drives Moore’s Law!

Density: SFQ lags semiconductors by $\sim 10^4$
System performance lags by $\approx 10^{10}$

- But comparison is difficult – no superconducting computer has run a LINPACK benchmark to measure performance in FLOP/s
Can superconducting computing compete?
A better way to view the relationship

CMOS

SFQ
Superconducting Computing and the IARPA C3 Program

- Problems in large scale computing
- Feasibility study
- Cryogenic Computing Complexity (C3) program (IARPA)
- Challenges

- Related work
  - Digital-RF Receiver (Hypres)
  - Quantum Annealing (D-Wave Systems)
  - Japanese Superconducting Computing Program
  - National Strategic Computing Initiative (NSCI)
  - SuperTools, Cryo-to-RT Interconnect Programs (IARPA)

- Conclusions
Digital-RF Receiver (Hypres)

- Commercial product with applications in:
  - Software-defined radio, satellite communications
- Directly digitizes RF (no analog down-conversion)
  - Ultra-wide bandwidth, multi-band, multi-carrier
- Hybrid temperature heterogeneous technology
  - Different technologies between ambient and 4 K
  - Closed-cycle cryogenic refrigerator

Quantum Annealing (D-Wave Systems)

- D-Wave® TwoX™ (2015 August 20), a commercial superconducting quantum annealing processor
- 128,000 Josephson junctions
- 1000 qubit array
- 15-20 mK operating temperature
Japanese 3-year program started

- “Superconductor Electronics System Combined with Optics and Spintronics”
- Processor goals: AQFP majority logic, 8-bit simplified RISC architecture, ~25,000 JJs, ~10 instructions
National Strategic Computing Initiative (NSCI)

Executive Order July 29, 2015

By the authority vested in me as President by the Constitution and the laws of the United States of America, and to maximize benefits of high-performance computing research, development, and deployment, it is hereby ordered as follows:

- (b) Foundational Research and Development Agencies. There are two foundational research and development agencies for the NSCI: the Intelligence Advanced Research Projects Activity (IARPA) and the National Institute of Standards and Technology (NIST).

IARPA will focus on future computing paradigms offering an alternative to standard semiconductor computing technologies. NIST will focus on measurement science to support future computing technologies. The foundational research and development agencies will coordinate with deployment agencies to enable effective transition of research and development efforts that support the wide variety of requirements across the Federal Government.
Superconducting Computing and the IARPA C3 Program

- Problems in large scale computing
- Feasibility study
- Cryogenic Computing Complexity (C3) program (IARPA)
- Challenges
- Related work

- Conclusions
Future Supercomputing Vision

- Hybrid technologies: digital (CMOS, SFQ), probabilistic, analog, neuromorphic, reversible, and quantum computing (QC) — whatever works best!
- SFQ digital platform supports multiple cryogenic technologies
- Requires optical interconnects between room temperature and cryogenic nodes

Courtesy of the Oak Ridge National Laboratory, U.S. Department of Energy
Lessons from an Emerging Technology

• Fair **metrics** are needed to evaluate alternative computing technologies
  – level the playing field to allow different technologies to compete
  – relevant lessons from hiring for diversity?

• Ramping up requires time and **resources**
  – the real Moore’s Law

• Government funding alone is not sufficient
  – cost to develop energy-efficient, large-scale computers is large
  – ramp up using smaller products and markets

• Don’t go it alone
  – use your mother elephant

• Go big or go home!
  – small improvements are not worth the effort
  – large disruptions require even larger advantages
References

Superconducting Computing in Large-Scale Hybrid Systems

D. Scott Holmes, Booz Allen Hamilton
Alan M. Kadin, Senior Member of IEEE
Mark W. Johnson, D-Wave Systems

Once focused solely on computation speed, superconducting computing is now proving useful in hybrid systems where its unique capabilities complement conventional computing technologies. Energy efficiency has become a strong motivation for developing large-scale superconducting systems.

Computer, vol. 48, Dec. 2015