

Superconductor Technologies for Extreme Computing

Arnold Silver

Workshop on Frontiers of Extreme Computing

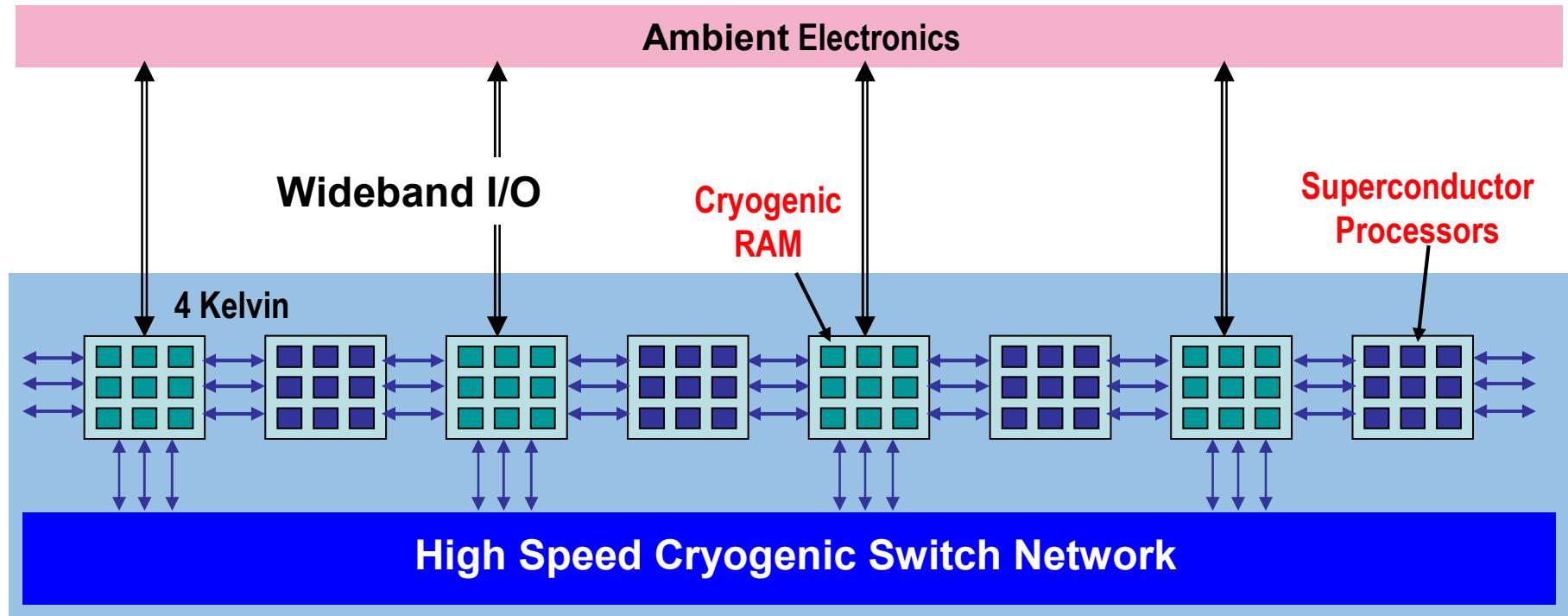
Monday, October 24, 2005

Santa Cruz, CA

Outline

- **Introduction**
- **Single Flux Quantum (SFQ) Technology**
- **State-of-the-Art**
- **Prospects**
- **Quantum Computing**
- **Summary**

Notional Diagram of a Superconductor Processor



- Superconductor processors communicate with local cryogenic RAM and with the cryogenic switch network.
- Cryogenic RAM communicates via wideband I/O with ambient electronics.

Early Technology Limited

- **Early superconductor logic was voltage-latching**
 - Voltage state data
 - AC power required
 - **Speed limited** by RC load and reset time (~GHz)
- **Single Flux Quantum (SFQ) is latest generation.**
 - Current/Flux state data
 - SFQ pulses transfer data
 - DC powered
 - Higher speed (~100 GHz)
- **Incremental progress on DoD contracts.**
 - Small annual budgets
 - Focus on small circuit demos
 - Minimal infrastructure investment

SFQ Features

- Quantum-mechanical devices
- An “electronics technology”
- High speed and ultra-low on-chip power dissipation
 - Fastest, lowest power digital logic
 - ≥ 100 GHz clock expected
 - \sim nW/gate/GHz expected
- Wideband communication on-chip and inter-chip
 - Superconducting transmission lines
 - Low-loss
 - Low-dispersion
 - Impedance matched
 - 60 GHz data transfer demonstrated with negligible cross-talk

Comparison of a 12 GFLOPS SFQ and CMOS chip

40 kgate SFQ chip	50 GHz clock	2 mW	Plus 0.8 W cooling power
2 Mgate CMOS chip	1 GHz clock	80 W	Also requires cooling

Some Issues Need To Be Addressed

- **Present disadvantages**
 - Low chip density and production maturity
 - Inadequate cryogenic RAM
 - Cryogenic cooling
 - Cryogenic - ambient I/O
- **Density and maturity will increase with better VLSI**
- **Promising candidates for cryogenic RAM**
 - Hybrid superconductor-CMOS
 - Hybrid superconductor-MRAM
 - SFQ RAM
- **Cryogenics is an enabler for low power**
- **Options for wideband I/O exist**

Technology Overview

➤ Basic technology

- Josephson tunnel junctions and SQUIDs
- SFQ logic gates
- SFQ transmitters-receivers
- Cryogenic memory
- Superconducting films produce microstrip and stripline transmission lines
 - Zero-resistance at dc (no ohmic loss)
 - Low-loss, low-dispersion at MMW frequencies
 - Impedance-matched
 - Wideband

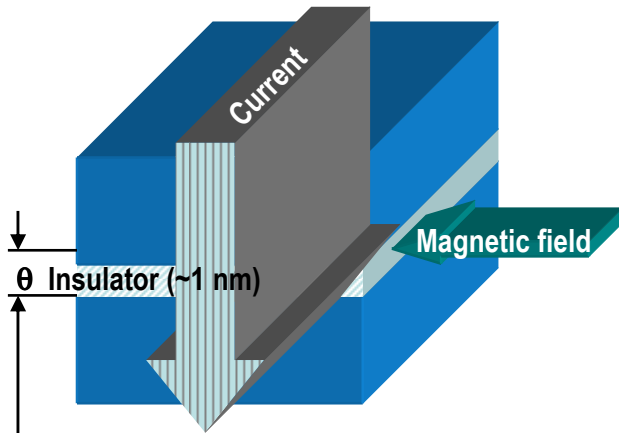
➤ Enabling technologies

- Advanced VLSI foundry
- Superconducting multi-chip modules
- Wideband I/O technologies
 - Optical fiber
 - Electrical ribbon cable
 - Cryogenic LNAs

Comparison of SFQ - CMOS Functions

Function	CMOS	SFQ
Basic Switch	<ul style="list-style-type: none"> ▪ Transistor 	<ul style="list-style-type: none"> ▪ Josephson tunnel junction (a 2 terminal device)
Data Format	<ul style="list-style-type: none"> ▪ Voltage level 	<ul style="list-style-type: none"> ▪ Identical picosecond (current) pulses
Speed Test	<ul style="list-style-type: none"> ▪ Ring oscillator 	<ul style="list-style-type: none"> ▪ Asynchronous flip-flop, static divider ▪ 770 GHz achieved ▪ 1,000 GHz expected
Data Transfer	<ul style="list-style-type: none"> ▪ Voltage data bus ▪ RC delay with power dissipation 	<ul style="list-style-type: none"> ▪ “Ballistic” transfer at $\sim 100 \mu\text{m/ps}$ in nearly lossless and dispersion-free passive transmission lines (PTL)
Clock Distribution	<ul style="list-style-type: none"> ▪ Voltage clock bus 	<ul style="list-style-type: none"> ▪ Clock pulse regeneration and ballistic transfer at $\sim 100 \mu\text{m/ps}$ in nearly lossless and dispersion-free PTLs
Logic Switch	<ul style="list-style-type: none"> ▪ Complementary transistor pair 	<ul style="list-style-type: none"> ▪ Two-junction comparator
Bit Storage	<ul style="list-style-type: none"> ▪ Charge on a capacitor 	<ul style="list-style-type: none"> ▪ Current in a lossless inductor
Fan-In, Fan-Out	<ul style="list-style-type: none"> ▪ Large 	<ul style="list-style-type: none"> ▪ Small
Power	<ul style="list-style-type: none"> ▪ Volt levels 	<ul style="list-style-type: none"> ▪ Millivolt levels
Power Distribution	<ul style="list-style-type: none"> ▪ Ohmic power bus 	<ul style="list-style-type: none"> ▪ Lossless superconducting wiring
Noise	<ul style="list-style-type: none"> ▪ $\geq 300 \text{ K}$ thermal noise 	<ul style="list-style-type: none"> ▪ 4 K thermal noise that enables low power operation

Josephson Tunnel Junction



$$J = J_C \sin \theta$$

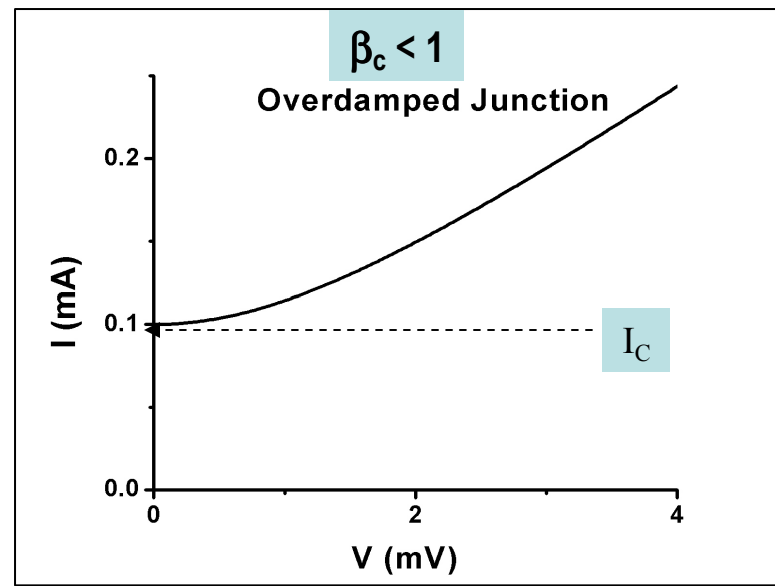
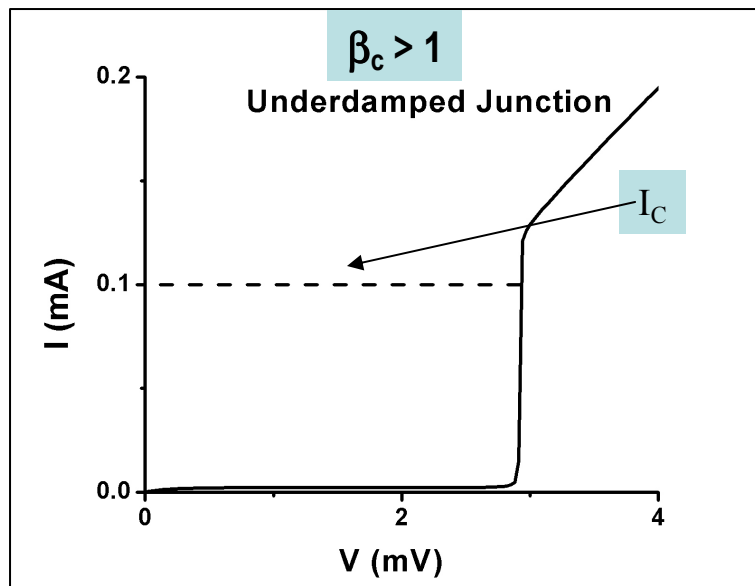
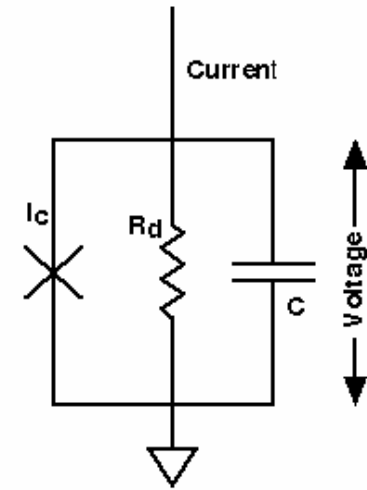
$$V = \frac{h}{2e} f$$

$$f = \frac{1}{2\pi} \frac{d\theta}{dt}$$

$$\frac{h}{2e} \equiv \Phi_0 \cong 2.07 \text{ mV} \cdot \text{ps}$$

Damping Parameter

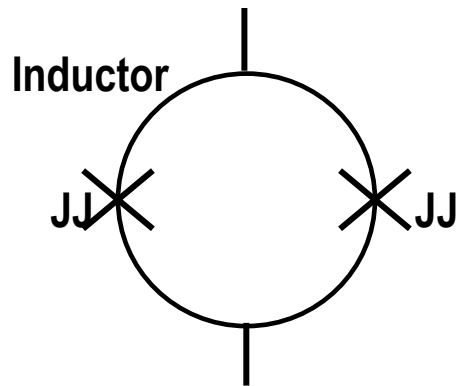
$$\beta_c = \frac{2\pi}{\Phi_0} (I_C R_d) (R_d C)$$



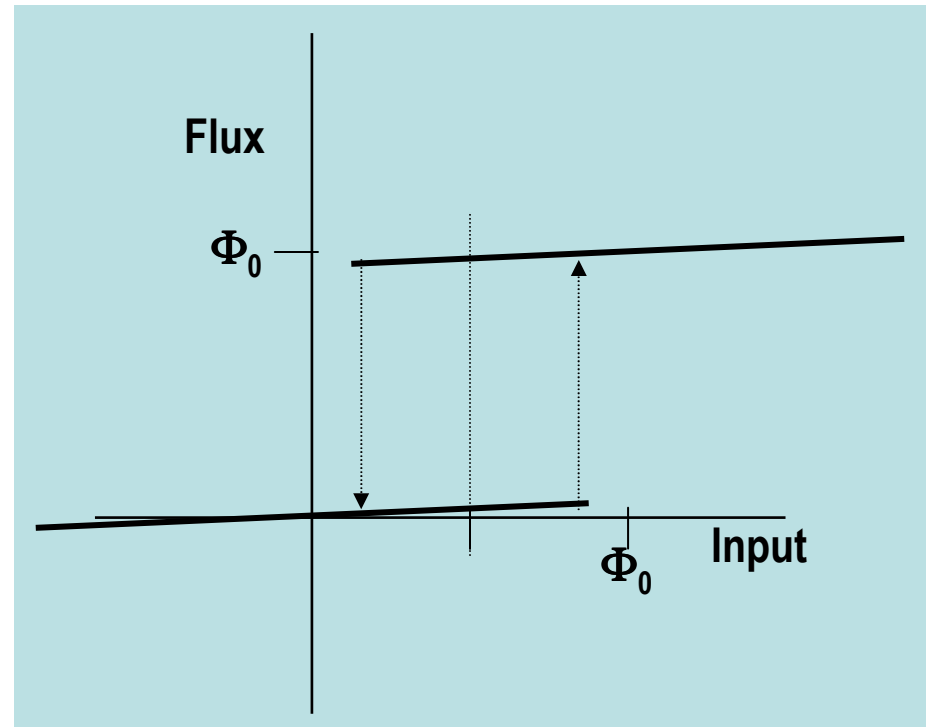
SQUIDS Are Basic SFQ Elements

- Combine flux quantization with the non-linear Josephson effects
- Store flux quantum or transmit SFQ pulse

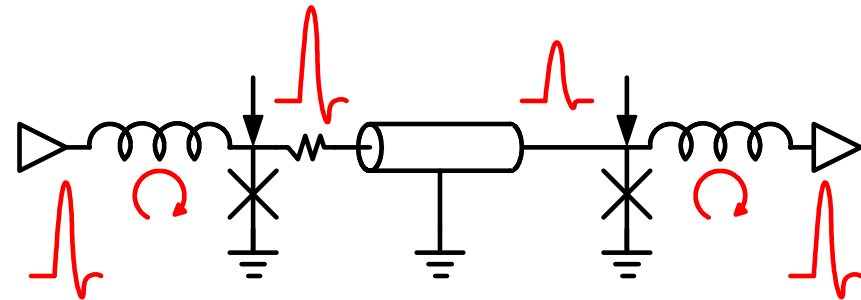
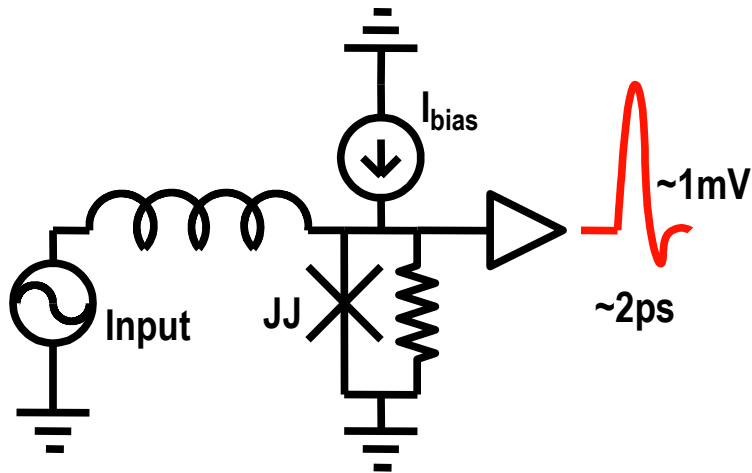
$$2\pi \frac{L i_{\text{circ}}}{\Phi_0} + \sum_{\text{junctions}} \theta_{JJ} = 2\pi k; \quad k = \text{integer}$$



Double JJ (dc) SQUID



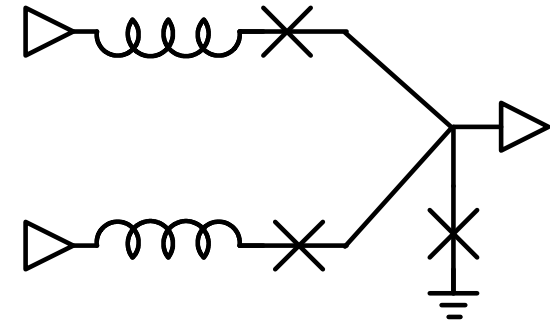
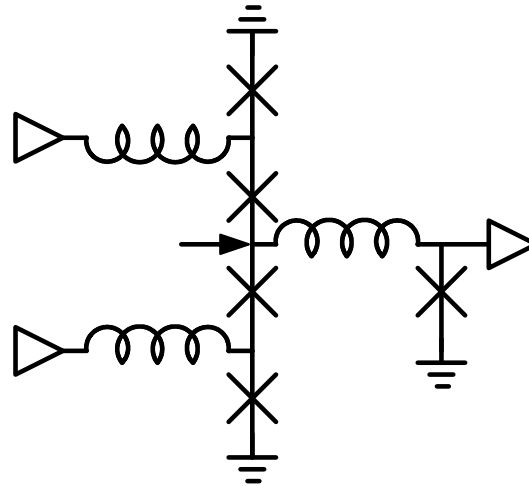
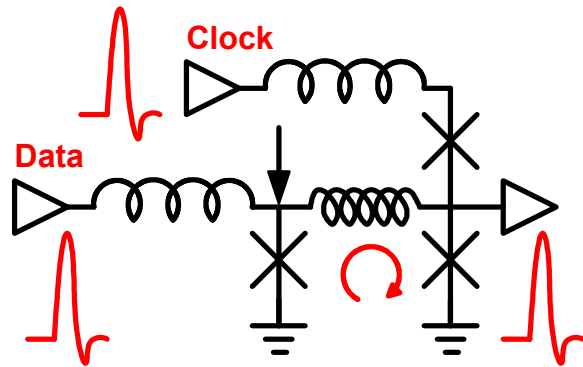
SFQ Is A Current Based Technology



- When $(Input + I_{bias})$ exceeds JJ critical current I_c , JJ “flips”, producing an SFQ pulse.
- Area of the pulse is $\Phi_0 = 2.067 \text{ mV-ps}$
- Pulse width shrinks as J_c increases
- SFQ logic is based on counting single flux quanta

- SFQ pulses propagate along impedance-matched passive transmission line (PTL) at the speed of light in the line ($\sim c/3$).
- Multiple pulses can propagate in PTL simultaneously in both directions.

SFQ Gates



Data Latch (DFF)

- SFQ pulse is stored in a larger-inductance loop
- Clock pulse reads out stored SFQ
- If no data is stored, clock pulse escapes through the top junction

“OR” Gate (merger)

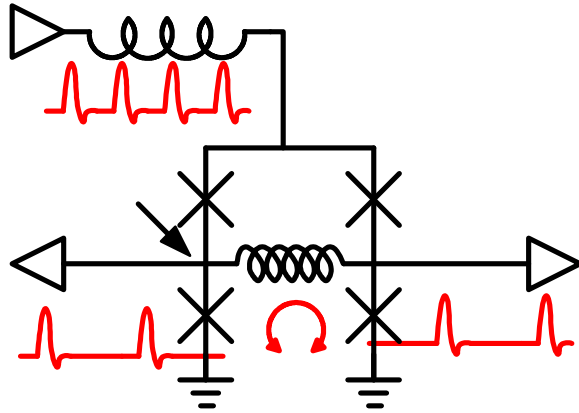
- Pulses from both inputs propagate to the output

“AND” Gate

- Two pulses arriving “simultaneously” switch output junction
- DFF in each input produces clocked AND gate

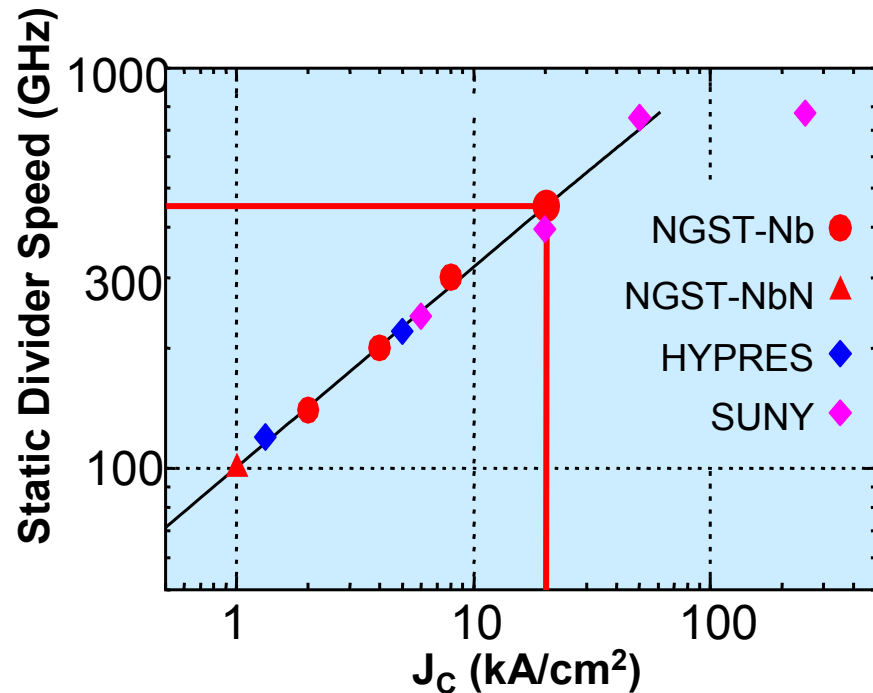
- PTLs transmit clock and data signals
- Average number of junctions per gate is 10

SFQ Is The Fastest Digital Technology

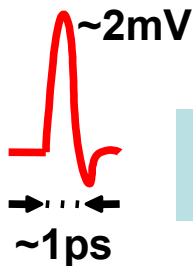


Toggle Flip-Flop – Static Frequency Divider

- Benchmark of SFQ circuit performance
- Maximum frequency scales with J_C



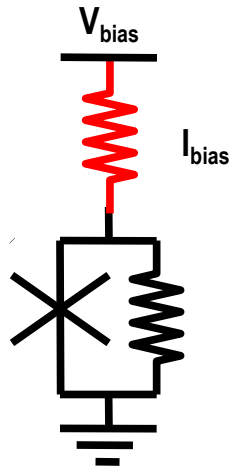
- Measured dc to **446 GHz** static divider
- **770 GHz** demonstrated in experiment



Picosecond SFQ pulses can encode **terabits per second**.

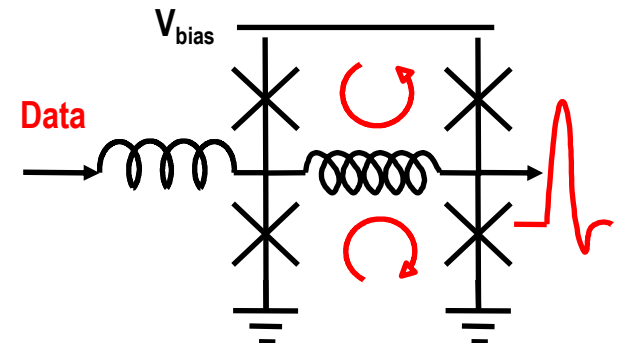
SFQ Is The Lowest Power Digital Technology

- One SFQ pulse dissipates $I_C \Phi_0$ in shunt resistor
 - For $I_C = 100 \mu\text{A} \Rightarrow 2 \times 10^{-19}$ Joule ($\sim 1\text{eV}$)
 - ~ 5 junctions switch in single logic operation
 - $1 \text{ nW/gate/GHz} \Rightarrow 100 \text{ nW/gate at } 100 \text{ GHz}$



- Static power dissipation in bias resistors: I^2R
- For $I_C = 100 \mu\text{A}$ biased at $0.7 I_C$
 - Typical $V_{\text{bias}} = 2 \text{ mV}$ (to maximize bias margin)
 - 140 nW/JJ , 1400 nW/gate is 23 X the dynamic power

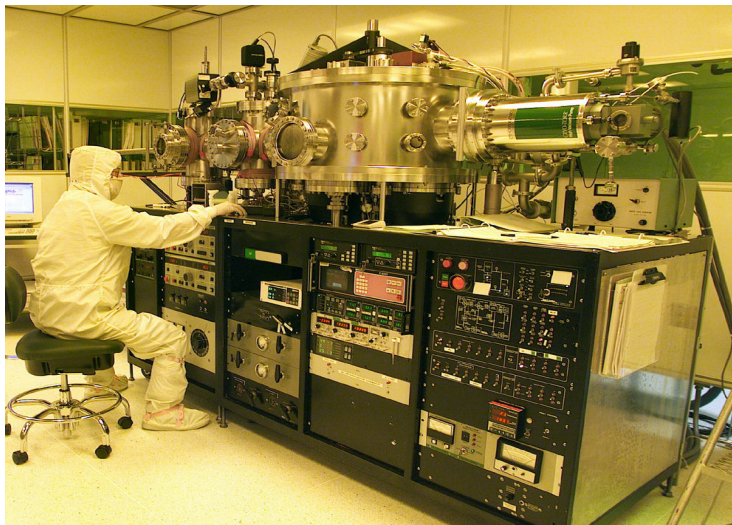
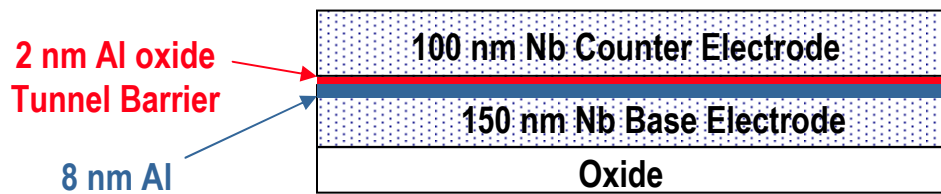
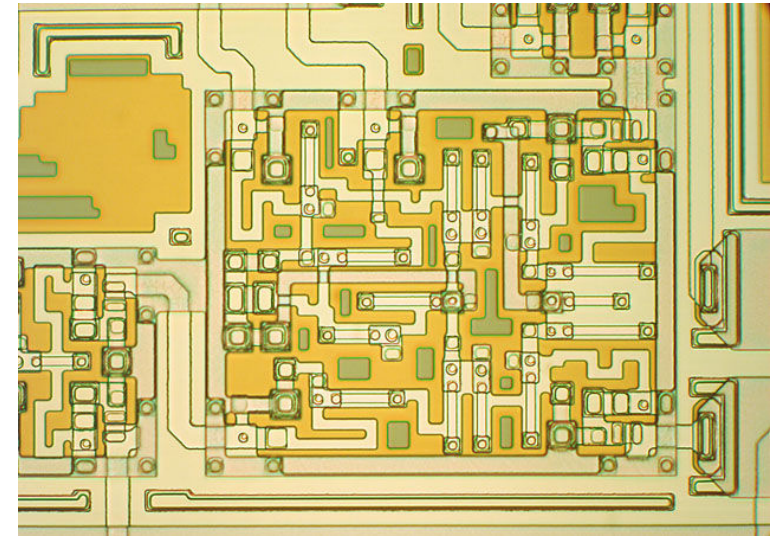
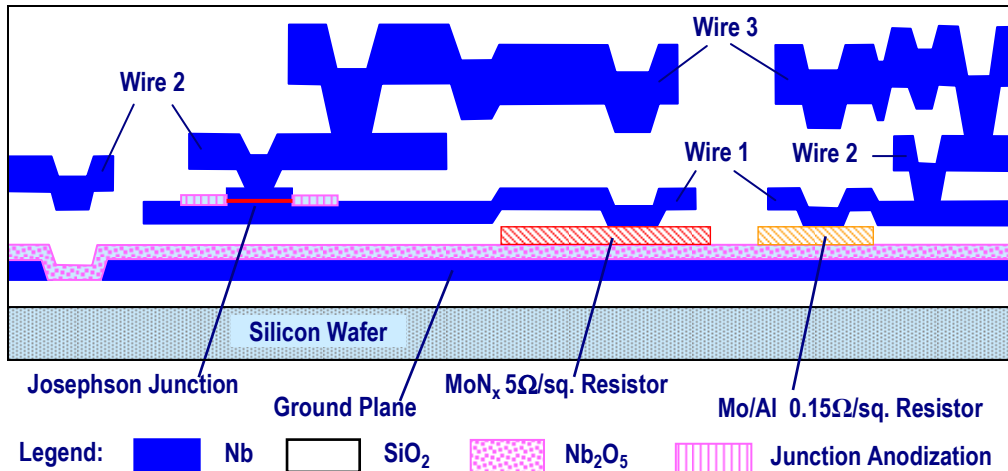
- Voltage-biased SFQ gates will eliminate bias resistors and static power dissipation
 - Self-clocked complementary logic
 - Incorporates clock distribution circuitry
 - $V_{\text{bias}} = \Phi_0 F_{\text{Clock}}$



SFQ Digital ICs Have Been Developed

- **First SFQ circuit (~ 1977) was a dc to SFQ converter integrated with toggle flip-flops to form a binary counter.**
- **Extensive development of SFQ logic did not occur until after 1990.**
- **Advanced SFQ logic was developed on HTMT FLUX.**
 - Architecture
 - Design tools
 - LSI fabrication
 - Logic
 - High data-rate on-chip communications
 - Inter-chip communications
 - Vector registers
 - Microprocessor logic chip

Superconductor IC Fabrication Is Simpler Than CMOS

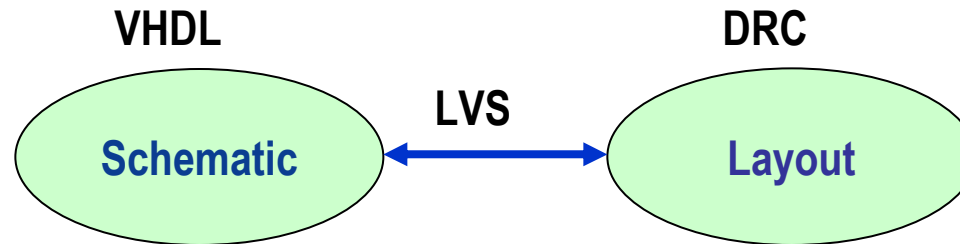


- Oxidized silicon wafers (100-mm)
 1. Deposit films (Nb trilayer, Nb wires, resistors, and oxide)
 2. Mask (g-line, i-line photolithography or e-beam)
 3. Etch (dry etch, typical gases are SF_6 , $\text{CHF}_3 + \text{O}_2$, CF_4)
 4. Repeated 14 to 15 times
- No implants, diffusions, high temperature steps
- Trilayer deposition forms Josephson tunnel junction
- All layers are deposited *in-situ*
- Al is passively oxidized *in-situ* at room temperature
- 1 μm minimum feature, 2.6 μm wire pitch
- Throughput limited by deposition tools

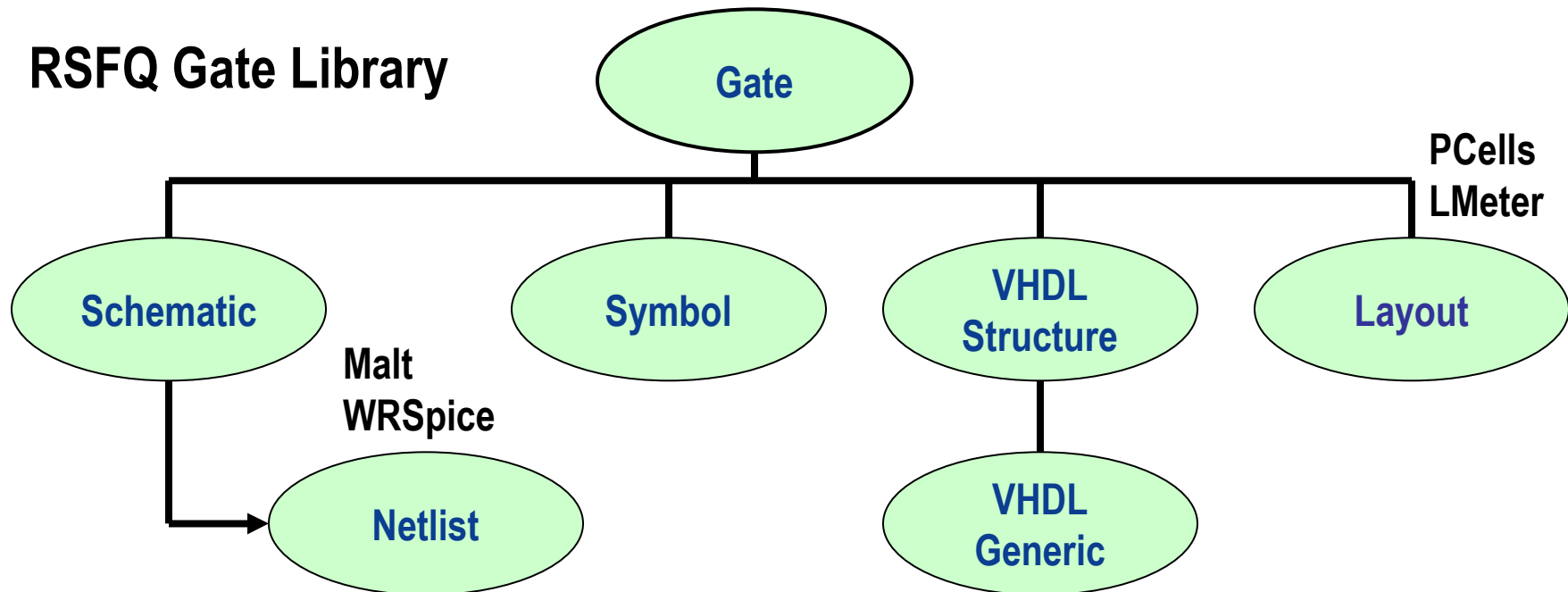
Cadence-based SFQ Design Flow (NGST)

Is similar to Semiconductor Design

Logic Synthesis & Verification



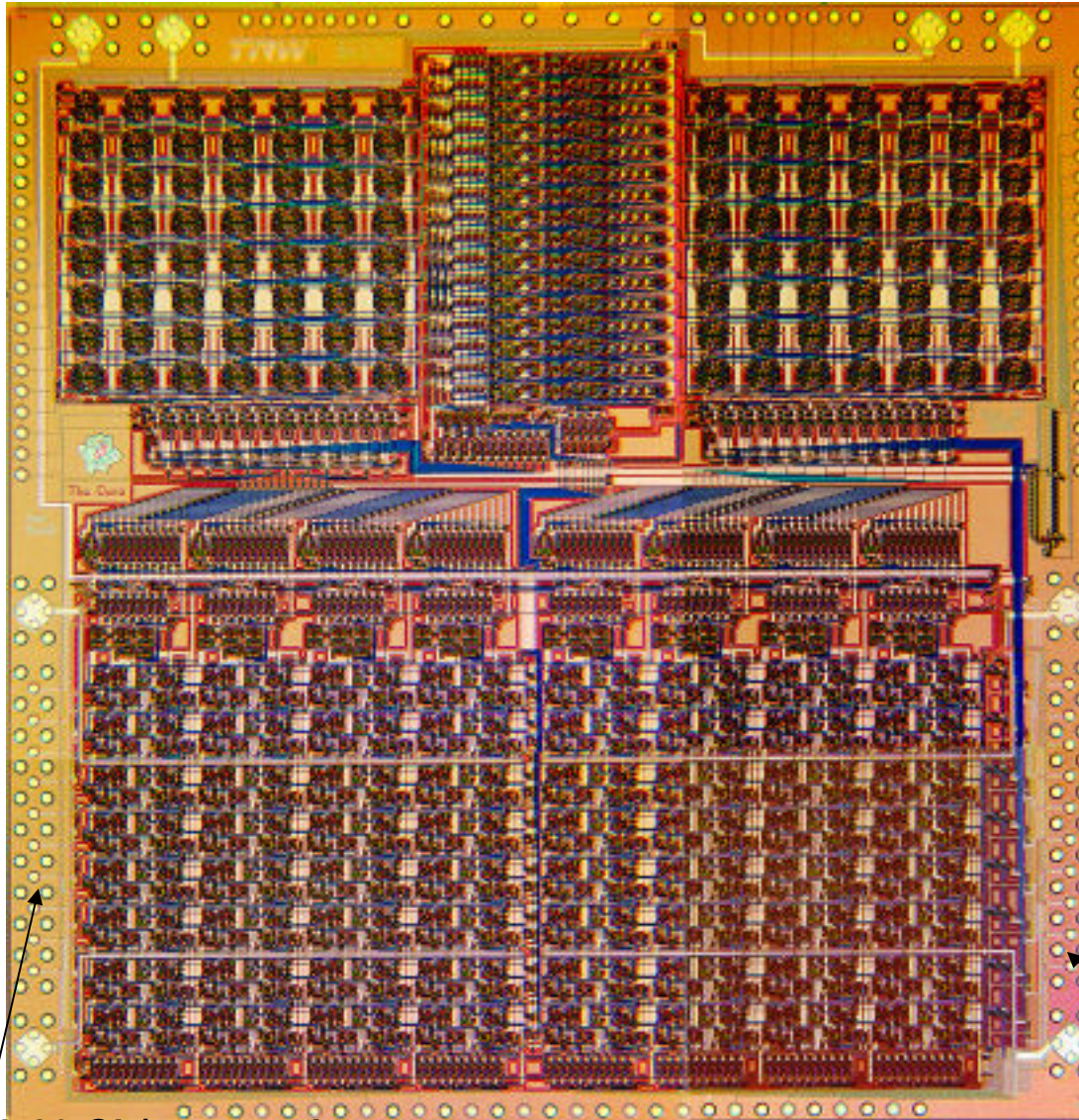
RSFQ Gate Library



Complex Chips Have Been Reported

Function	Complexity	Speed	Cell Library	Organizations
FLUX-1. 8-bit μP prototype. 25 30-bit-dual-op instructions.	63 K Junctions. 10.3 mm x 10.6 mm.	Designed for 20 GHz. Not tested.	Yes. Incorporates drivers/receivers for PTL.	Northrop Grumman, Stony Brook, JPL
CORE1α10. 8-bit bit-serial μP. 7 8-bit instructions.	7 K Junctions. 3.4 mm x 3.2 mm.	21 GHz local clock. 1 GHz system clock. Fully functional.	Yes. Gates connected by JTLs and/or PTLs	ISTEC-SRL, Nagoya U., Yokohama National U.
MAC and Prefilter for programmable pass-band A/D converter.	6 K–11 K Junctions. 5 mm x 5 mm.	20 GHz design	Yes. Gates connected by parameterized JTLs and/or PTLs	Northrop Grumman
A/D converter	6 K Junctions.	19.6 GHz.	?	Hypres
Digital receiver	12 K Junctions.	12 GHz.	?	Hypres
FIFO buffer memory	4K bit. 2.6 mm x 2.5 mm	32 bits tested at 40 GHz.	No	Northrop Grumman
X-bar switch	128 x 128 switch. 32 x 32 module.	2.5 Gbps.	No	NSA, Northrop Grumman
SFQ X-bar switch	32 x 32 module.	40 Gbps.	No	Northrop Grumman

FLUX-1 Microprocessor Chip

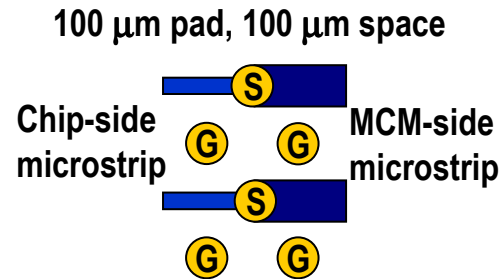
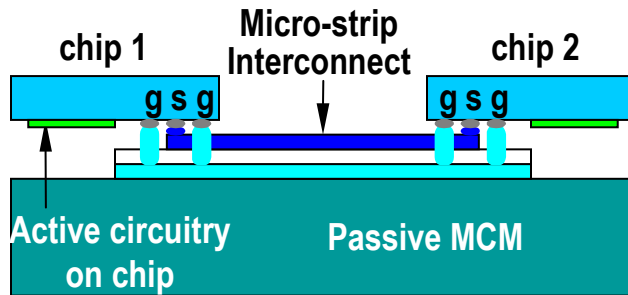


- Objective to demonstrate of 5K Gate SFQ chip operating at 20 GHz
- 8-bit microprocessor design
- 1-cm chip
- 8 - 20 Gb/s transmitters, receivers
- FLUX-1 chip redesigned, fabricated, partially tested
- 1.75 μm , 4 kA/cm² junction Nb technology
- 20 GHz internal clock
- 5 GByte/sec inter-chip data transfer limited by μP architecture
- Scan path diagnostics included
- 63 K junctions, 5 Kgate equivalent
- Power dissipation \sim 9 mW @ 4.5K
- 40 GOPS peak computational capability (8-bits @ 20-GHz clock)
- Fabricated in TRW 4 kA/cm² process in 2002

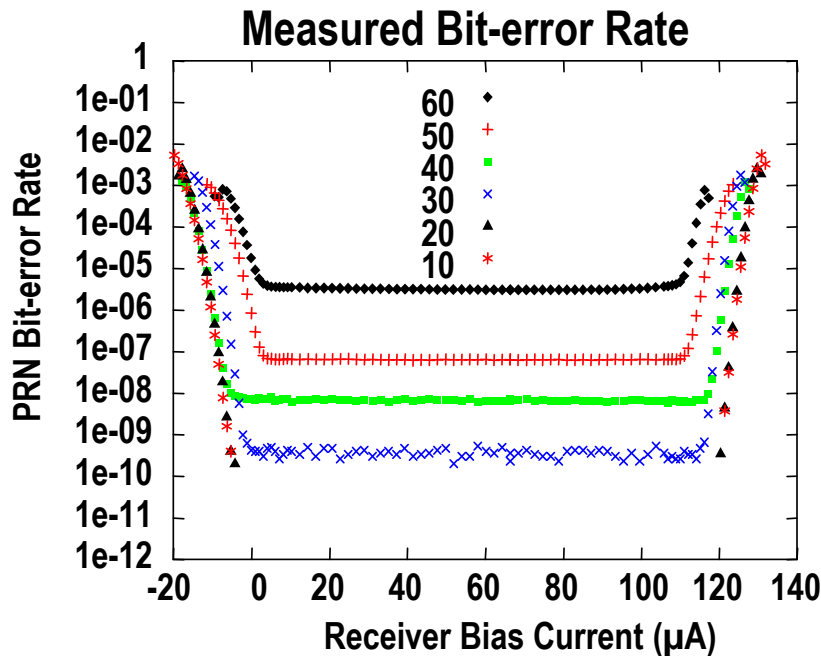
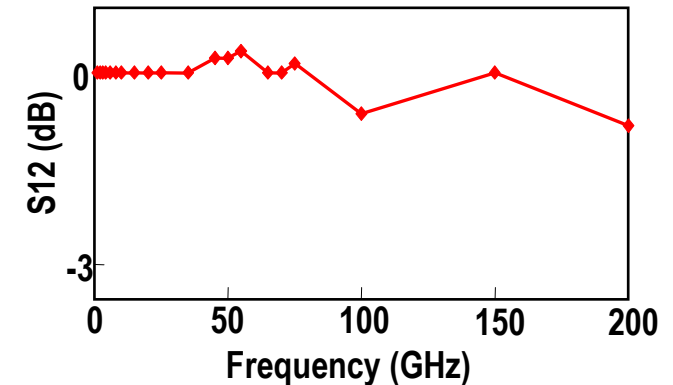
8-20 Gb/s transmitters

8-20 Gb/s receivers

60 GHz Interconnect Demonstrated



Chip-to-MCM Pad Optimization



- MCM Nb stripline wiring is low loss, wideband
- High density, low impedance solder bump arrays
- Ultra-low power driver-receiver enables high data rate communications
- SFQ data format enables multiple bits in transmission line simultaneously, increases throughput
- Demonstrated to 60 Gb/s through 2 solder bumps, 4 Ω resistor, and 4 Ω transmission lines on chip and MCM
- Timing errors produced BER floor above 30 Gb/s

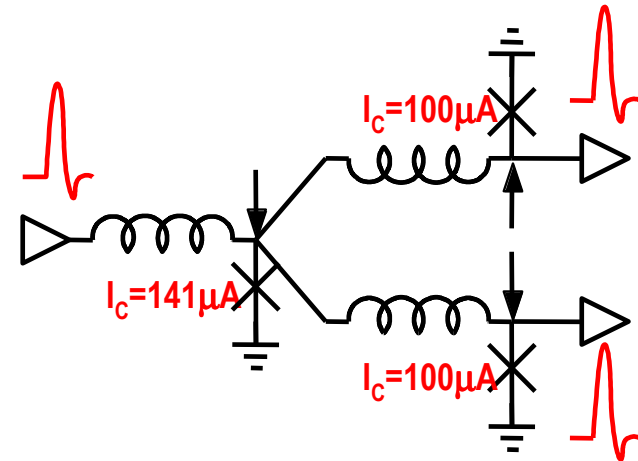
SFQ Faces Challenges of 100+ GHz Technologies

➤ Low power

- Low fan-out, need **“pulse splitting”**:
 - JTL provides current amplification
 - Amplified pulse can drive two JTLs
- All connections are point-to-point
- Fast, large RAM is hard to make

➤ High speed

- **No global clock**
 - Clock and data pulses are considered to be the same
 - Need to consider asynchronous/delay insensitive/self-timed/micropipelined
- On-chip latencies can reach many clock cycles
 - 10 ps clock period in PTL corresponds to 2 mm length
 - Pulse splitting adds latency



➤ On the cutting edge

- No truly automated place-and-route yet
- Off-the-shelf CAD tools need to be heavily customized
- Efficient gate library approach has to be refined

➤ Requirement for wideband I/O to ambient RAM

Improved Chip Performance Feasible

➤ Improve parameters by orders-of-magnitude

- + Increase junction and gate density
- + Increase clock frequency
- + Increase junction speed to 1,000 GHz by increasing $J_C \geq 100 \text{ kA/cm}^2$
- + Increase chip yield
- Reduce power dissipation to SFQ switching dissipation level
- Reduce bias current

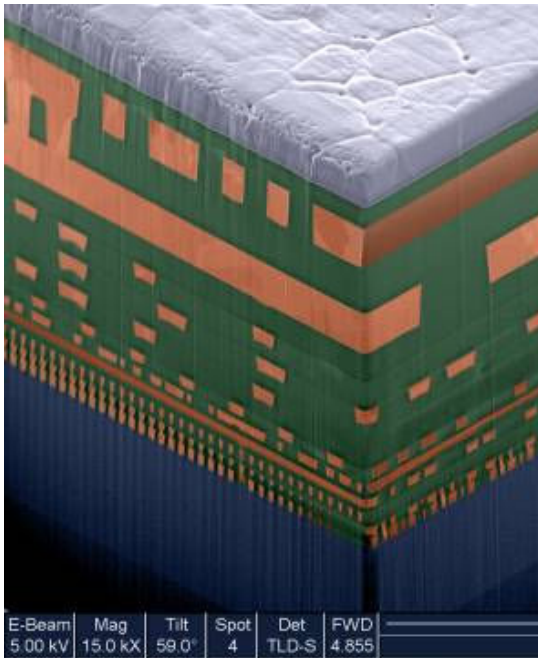
➤ Establish foundry following CMOS practice

- Lithography at 250-180 nm; 90-60 nm
- $J_C > 20 \text{ kA/cm}^2$; $\geq 100 \text{ kA/cm}^2$
- Add superconducting layers 7-9; > 20
- Vertically separate power and data transmission from gates
- Achieve $\geq 1\text{M}$ junctions/cm² ($\geq 10^5$ gates); 100-250M junctions/cm² (10-25M gates)
- Increase clock to 50 GHz; ≥ 100 GHz

➤ Improve CAD tools and methods

- May need to improve physical models for junctions with higher J_C
- Shorten development time

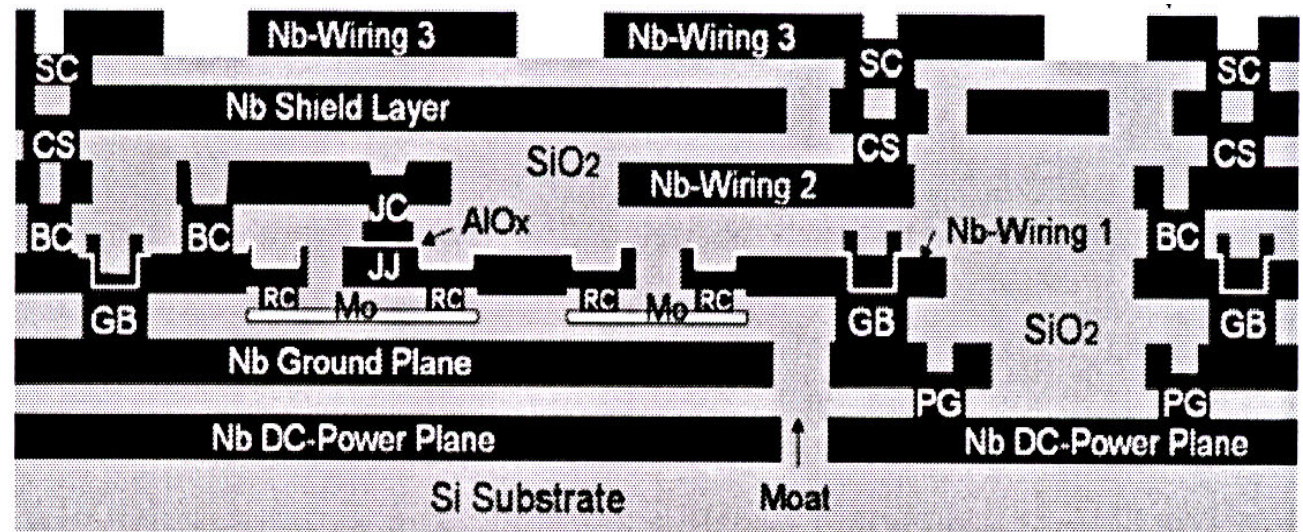
Density Is Increased by Adding Wiring Layers



IBM 90-nm Server-Class CMOS process

Fully-Planarized, 6-Metal Process (Proposed by ISTECSRL, Japan, Nagasawa et al., 2003)

- More metal layers are essential to increase chip density
- Vertically isolate power and communications lines from active devices
- Superconducting ground planes are excellent shields
- Full planarization and competitive lithography



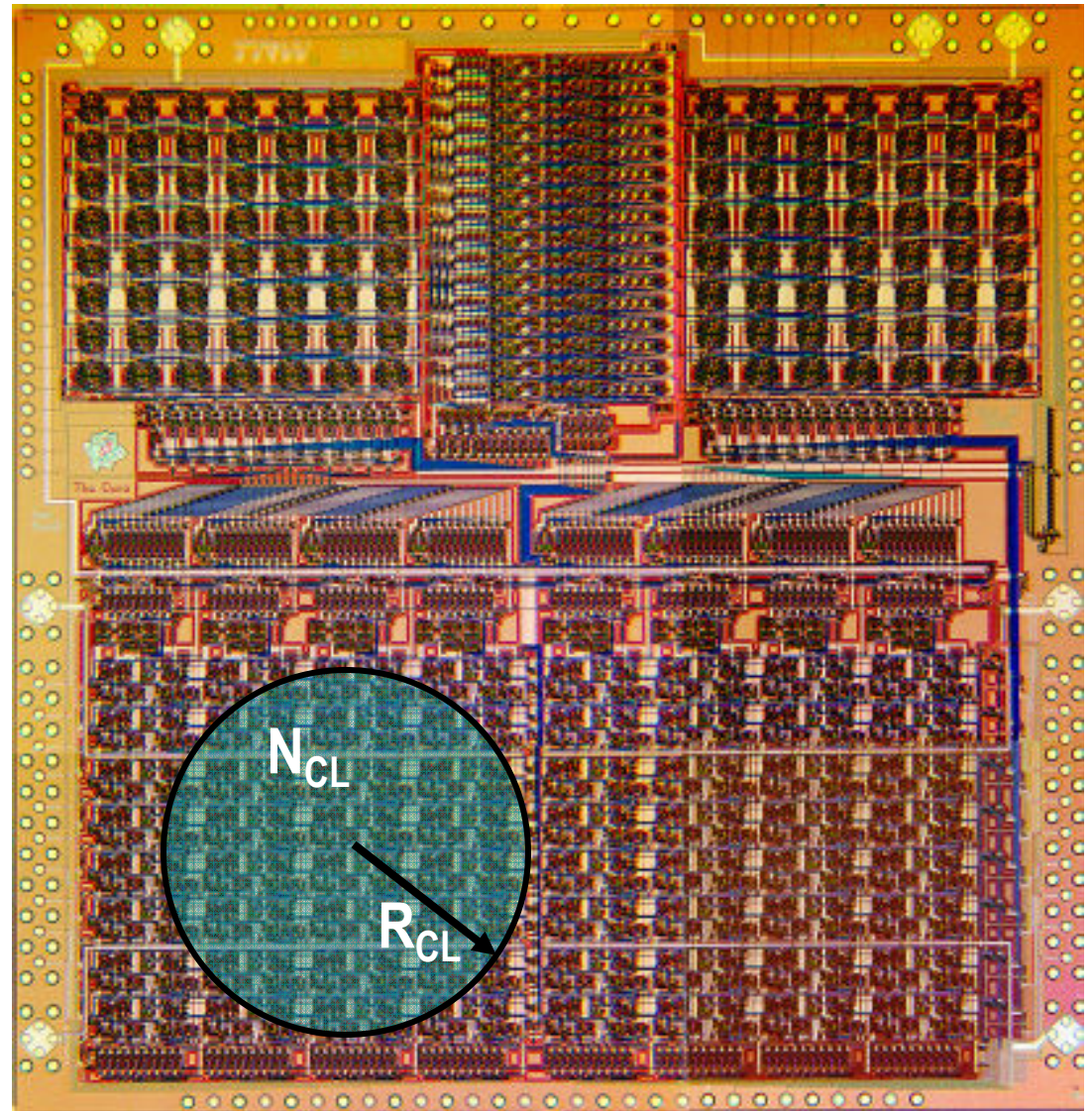
SFQ Technology Projections

	Before 2004	2010	Beyond 2010
Technology Projections			
Technology Node	1 μm	250 - 180 nm	90 nm or better
Current Density	8 kA/cm ²	50 kA/cm ²	> 100 kA/cm ²
Superconducting Layers	4	7 - 8	~ 20
New Process Elements	NA	Full Planarization	<ul style="list-style-type: none"> ▪ Alternate barriers ▪ Additional junction trilayers ▪ Vertical resistors and inductors
Power	$I_C V_{\text{bias}}$	Reduced Bias Voltage	<ul style="list-style-type: none"> ▪ CMOS-like ▪ Reduced I_C
Projected Chip Characteristics			
Junction Density	60 k/cm ²	2 - 5 M/cm ²	100-250 M/cm ²
Clock Frequency	< 20 GHz	50 - 100 GHz	100 - 250 GHz
Power	0.2 μW /Junction	8 nW/GHz/Junction	0.4 nW/GHz/Junction
	Increased Clock Frequency		Increased Density
Process Improvement	<ul style="list-style-type: none"> ▪ Smaller junction with higher J_C 		<ul style="list-style-type: none"> ▪ Smaller line pitch ▪ Greater vertical integration
Benefits	<ul style="list-style-type: none"> ▪ Faster circuits ▪ Larger signals 		<ul style="list-style-type: none"> ▪ More gates/cm² ▪ Reduced on-chip latency
Potential Disadvantages	<ul style="list-style-type: none"> ▪ Possibly larger spreads ▪ Increased system latency 		<ul style="list-style-type: none"> ▪ Potentially lower yield

Latency is measured in clock ticks

Gate Access Within Clock Period Is Important

- Clock radius (R_{CL}) is maximum distance data can travel within a clock period.
- N_{CL} is number of gates within a clock radius.
- Clock radius is limited by time-of-flight and the clock frequency.
- Increasing gate density is essential to increasing effectiveness.



Density Is Key To Gate Access

		Clock (GHz)	25	50	100	200	250
		Clock Radius (mm)	4	2	1	0.5	0.4
		Clock Area (mm ²)	50	12.6	3.14	0.79	0.5
Density (JJs/cm ²)	Density (Gates/mm ²)	Number of Gates Within Clock Radius (N _{CL})					
5 K	5	250	63	16	4	2.5	
60 K	60	3 K	750	190	47	30	
1 M	1 K	50 K	13 K	3.1 K	790	500	
5 M	5 K	250 K	63 K	16 K	4 K	2.5 K	
30 M	30 K	1.5 M	380 K	94 K	24 K	15 K	
100 M	100 K	5 M	1.3 M	310 K	79 K	50 K	
250 M	250 K	12.5 M	3.1 M	790 K	200 K	130 K	

Clock radius assumed to be 1/2 of time-of-flight.

High-End SFQ Computing Engine

2005

- **Not feasible**
~ 100 chips per processor
0.5 M processor chips, ~ 10^9 gates

2010

- ~ 10 chips per processor
40 K processor chips, ~ 10^9 gates

After 2010

- ~ 10 to 20 processors per chip
400 processor chips, including embedded memory

Applications to Quantum Computing

- **Quantum computing is being investigated using superconducting qubits.**
- **Flux-based superconducting qubits are physically similar to SFQ devices.**
- **SFQ circuits are best candidates to control/read superconducting qubits at millikelvin temperatures.**

Summary

- **SFQ needs major engineering development in chip technology if it is going to be a player in high-end computing.**
- **The engineering requirements are understood and a development plan defined.**
- **Prospects are exciting and achievable.**