EUCAS 2017
Geneva, Switzerland
Sep 19, 2017

2EO1-01

Execution of Stored Program in a Single-Flux-Quantum Microprocessor with Embedded Memories at 50 GHz

Akira Fujimaki, Ryo Sato[1], Yuki Hatanaka[1], Yuichi Matsui, Kazuyoshi Takagi[2], Naofumi Takagi[2], Masamitsu Tanaka[1]

[1]Nagoya University, [2]Kyoto University

Acknowledgment

This work was supported by JST ALCA and JSPS KAKENHI (Grant Numbers 16H02340, 26220904, and 16H02796), and the VLSI Design and Education Center of the University of Tokyo, in collaboration with Cadence Design Systems, Inc. The circuits were fabricated in CRAVITY of AIST.

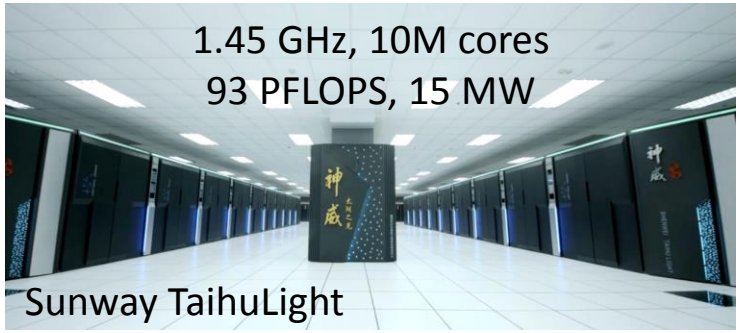
CRAVITY AIST

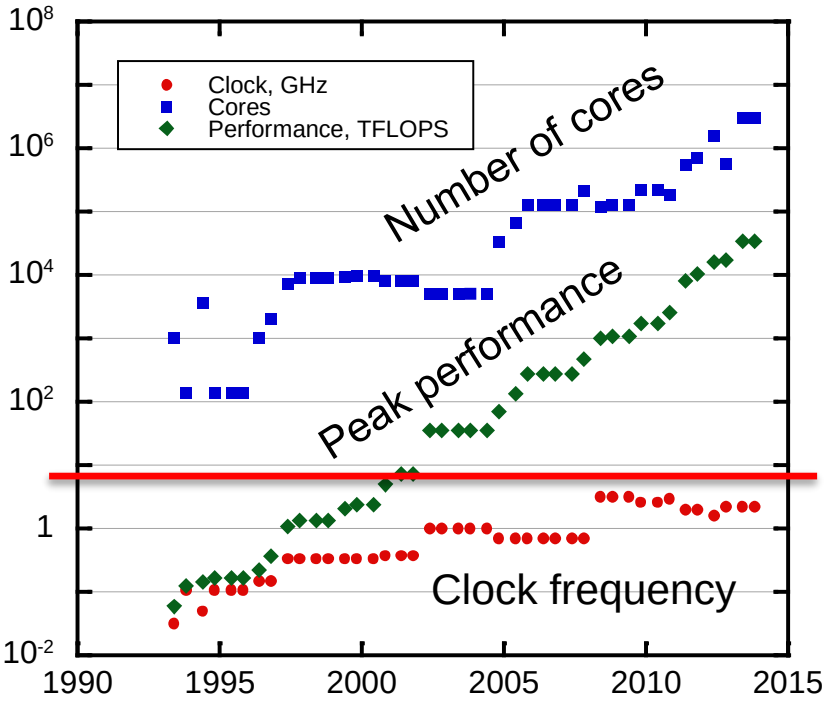
Outline

- History of development of SFQ microprocessors
- Execution of stored programs
  - First demonstration except Si technology
- Bit parallel ALU for lowered latency
- Other issues for practical SFQ microprocessors
- Summary

"Power-Wall Problem" in Large-Scale Computing

- Supercomputers
  - 10–20 MW
    - Use many MPU cores operating at 1–2 GHz
  - Requirements for 10^{18} FLOPS: 1 GW power & 1 km^2 area
- Data centers
  - 1.3% of all electricity use for the world (2010) [1]



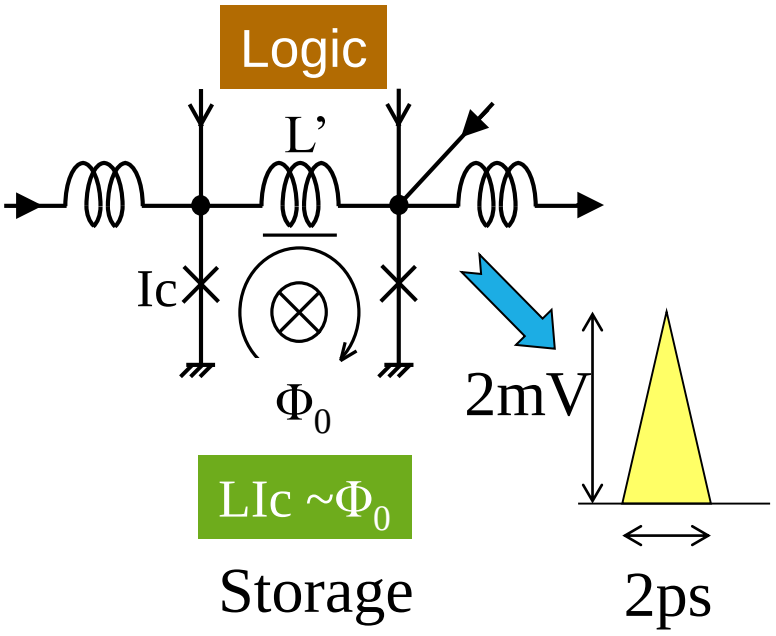


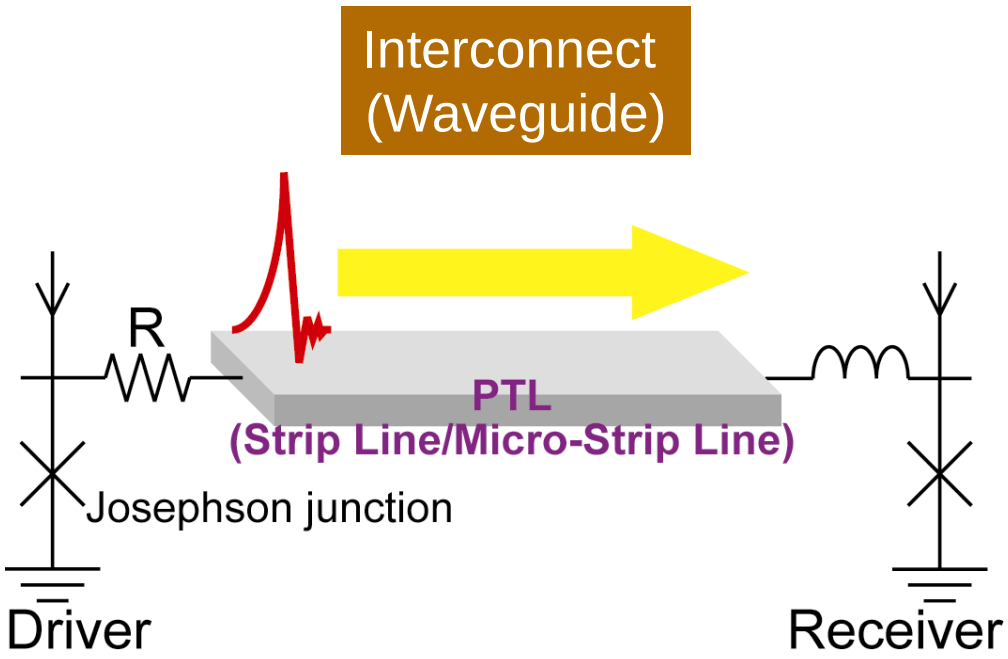
New Technology for energy-efficient high-performance computing is desired.

[1] J. G. Koomey, www.analyticspress.com/datacenters.html (2011)

Based on www.top500.org

Special Features of SFQ Circuits



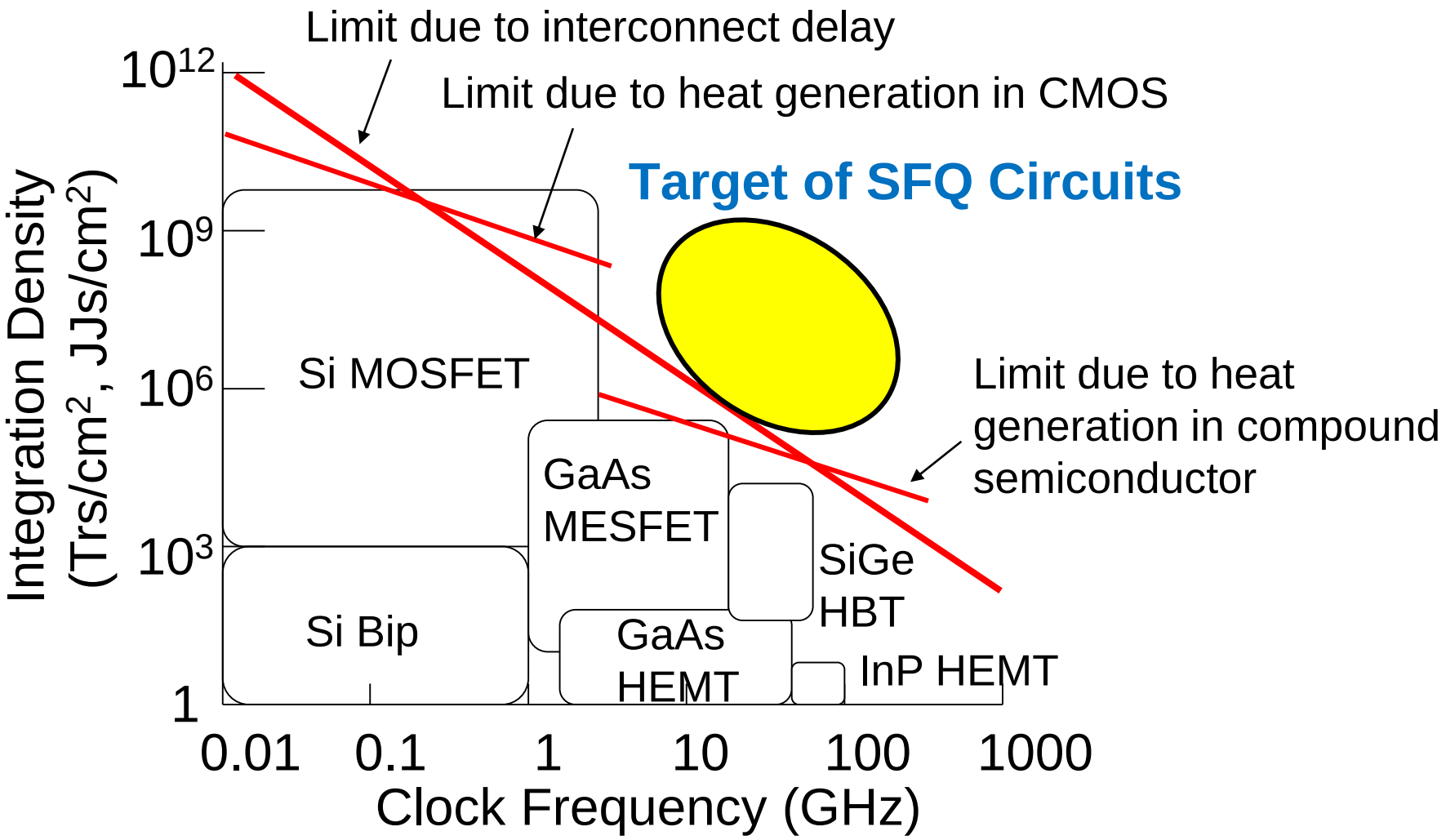


- Signal propagation at the speed of light with small distortion in interconnects based on waveguides.
- No recharge process both in logic operation and interconnects.
- Scaling law

➡ High-speed & low power

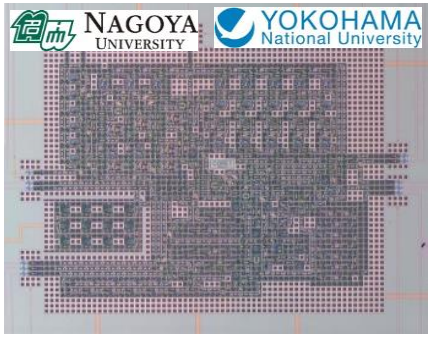
➡ Suitable to LSIs

Appealing Feature of SFQ Circuits

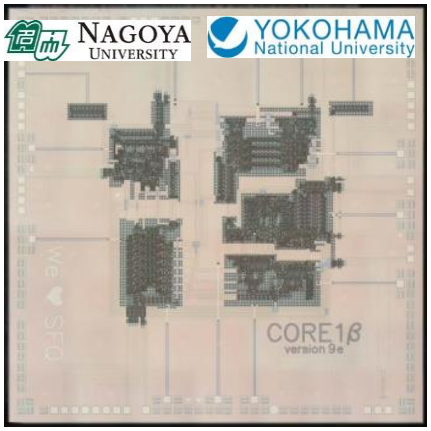


Development of RSFQ Microprocessors

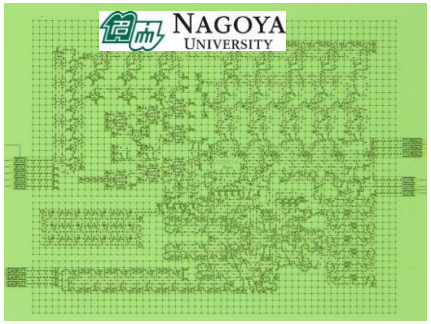
- CORE: simple, bit-serial processing
  - Ease circuit complexity. Use hardware (& energy) efficiently.
  - High-frequency clock operations (15–100 GHz)



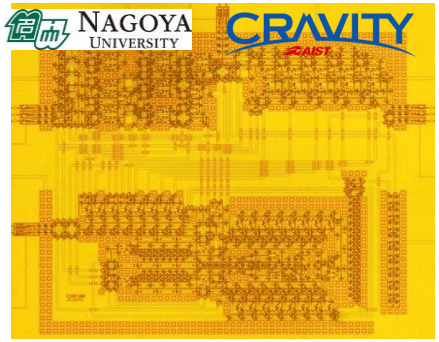
CORE1α v5 (2003)
4999 JJs, 15 GHz
167 MIPS, 1.6 mW
J_c=2.5 kA/cm^2
Prototype



CORE1β v9e (2006)
10955 JJs, 25 GHz
1400 MOPS, 3.3 mW
J_c=2.5 kA/cm^2
Pipeline processing



CORE1α LV (2013)
3869 JJs, 35 GHz
400 MIPS, 0.23 mW
J_c=10 kA/cm^2
Energy efficiency



CORE100 (2015)
3073 JJs, 100 GHz
800 MIPS, 1.0 mW
J_c=20 kA/cm^2
Ultrahigh frequency

Main Issues Left for Practical Microprocessors

- **Execution of meaningful programs stored in cryogenic memory** with energy-efficient SFQ circuits
- High-frequency operation of bit-parallel processing for small latency
- Energy-efficient power supply for dc-powered SFQ circuits
- Large capacity memory and voltage amplifiers with small footprints

Comparison of Energy-Efficient Circuits

- Typical case in 10 kA/cm^2 Process; may vary in targets.
- AC circuits require smaller but high-frequency power currents that are identical to operating speeds.

Easy to be replaced

	RSFQ (Base)	LR-bias LV-RSFQ	ERSFQ eSFQ	RQL	AQFP
Power supply	DC	DC	DC	AC	AC
Static power	1	0.1x	0	0	0
Dynamic power	1	~1x	1.1–2x	2x	<< 0.1x
Clock frequency	50 GHz	20–50 GHz	20–50 GHz	10–20 GHz	5 GHz
Power (per 1k JJs)	2.5 mW (2.5 mV, 1 A)	0.2 mW (0.2 mV, 1 A)	0.1 mW (0.1 mV, 1 A)	20 μW (0.6 mA, 50 Ω)	25 nW (20 μA, 50 Ω)
Gate Energy	100 aJ	5–10 aJ	2–4 aJ	0.5–1 aJ	0.01 aJ

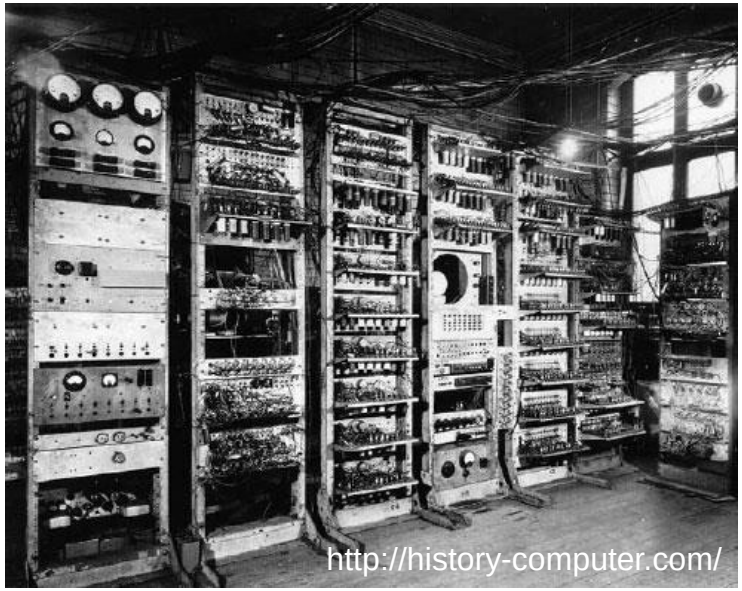
[LR-bias] T. Nishidai et al. *Physica C* **445-448** (2006) 1029; [LV-RSFQ] M. Tanaka et al. *Jpn. J. Appl. Phys.* **51** (2012) 053102; [ERSFQ] D. E. Kirichenko et al, *IEEE Trans. Appl. Supercond.* **21** (2011) 776; [eSFQ] M. Volkmann et al. *Supercond. Sci. Technol.* **26** (2013) 015002; [RQL] Q. P. Herr et al. *J. Appl. Phys.* **109** (2011) 103903; [AQFP] Takeuchi et al. *J. Appl. Phys*. **115** (2014) 103910.

Main Issues Left for Practical Microprocessors

- **Execution of meaningful programs stored in cryogenic memory** with energy-efficient SFQ circuits
- High-frequency operation of bit-parallel processing for small latency
- Energy-efficient power supply for dc-powered SFQ circuits
- Large capacity memory and voltage amplifiers with small footprints

Next Big Milestone

- Bit-serial microprocessors integrated with sufficient RAMs **to demonstrate "stored-program computing"**. We hoped that the stored programs included the program for finding the greatest divisor, which had been demonstrated with the first stored-program computer.



Manchester Small-Scale Experimental Machine (1948)
(First stored-program computer)

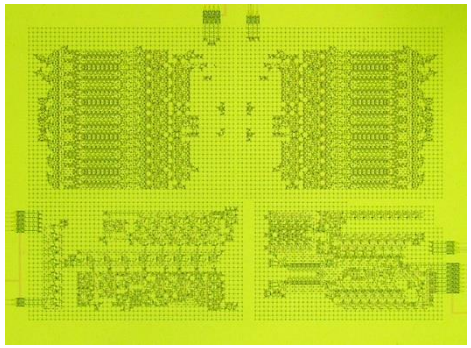
✓ 5.2 × 2.2 m, 1000 kg
✓ 550 vacuum tubes
✓ 3.5 kW

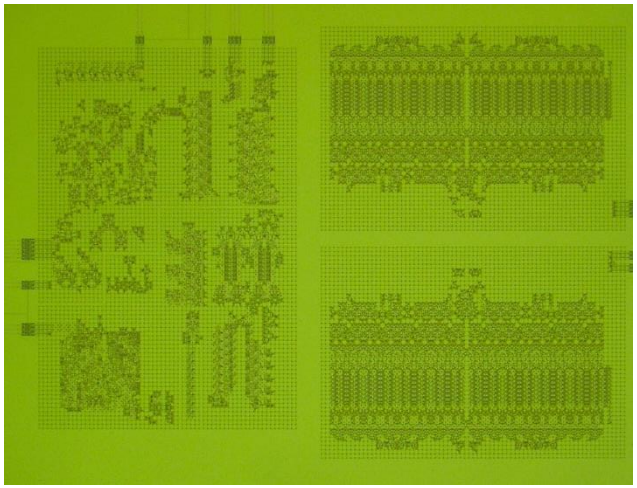
F. C. Williams and T. Kilburn,
"Electronic Digital Computers"
Nature **162** (1948) 487

CORE e Series

CRAVITY AIST

1 mm



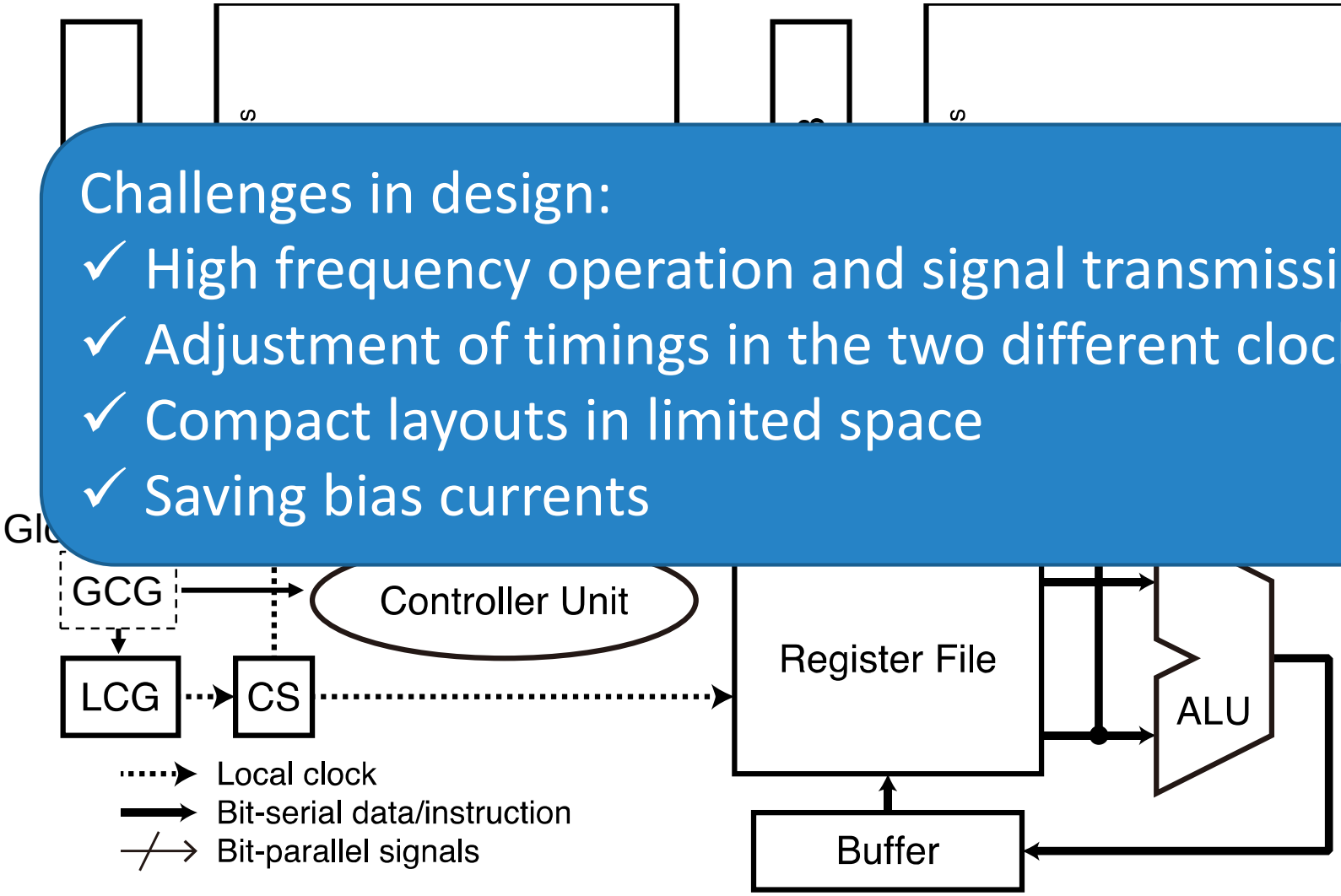


	CORE e2 (v5h)	CORE e4 (v5)
# of Registers	2	4
# of Instructions	13	20
Memories	2 x 128 bits	2 x 256 bits
Performance	333 MIPS (50 GHz bit-serial operation; 2 GHz clock / 6 cycle per instruction)	
Power	2.52 mW	4.57 mW
JJ Count	10603	20330
Feature	Minimal microarchitecture design with reduced RAMs	Full-featured CORE e MPU with a rich instruction set

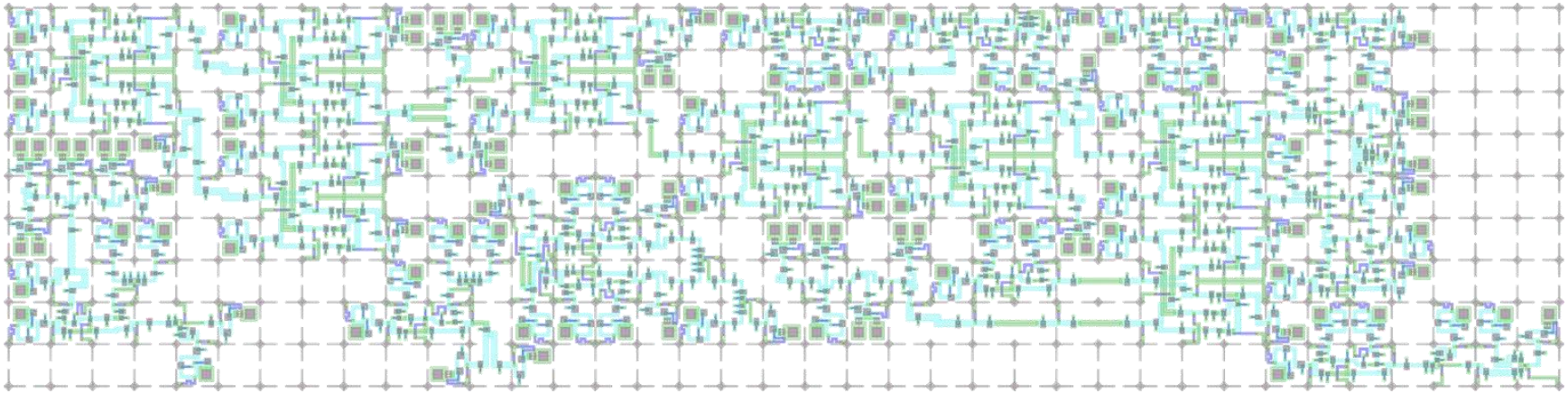
Microarchitecture (e2)

Challenges in design:

- ✓ High frequency operation and signal transmission
- ✓ Adjustment of timings in the two different clocks
- ✓ Compact layouts in limited space
- ✓ Saving bias currents



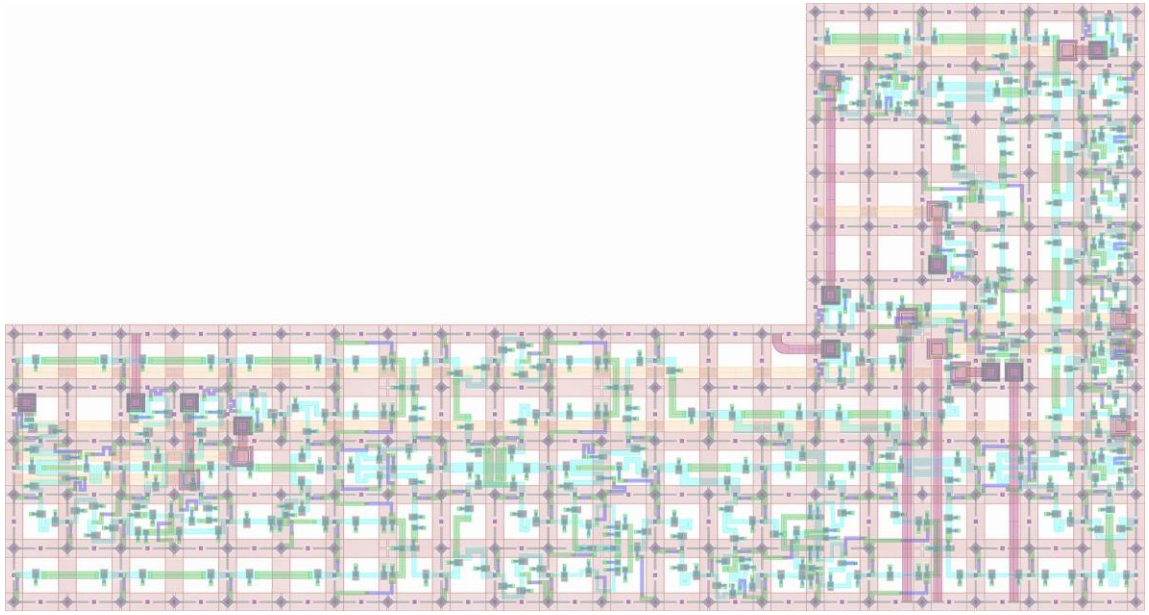
Controller Unit

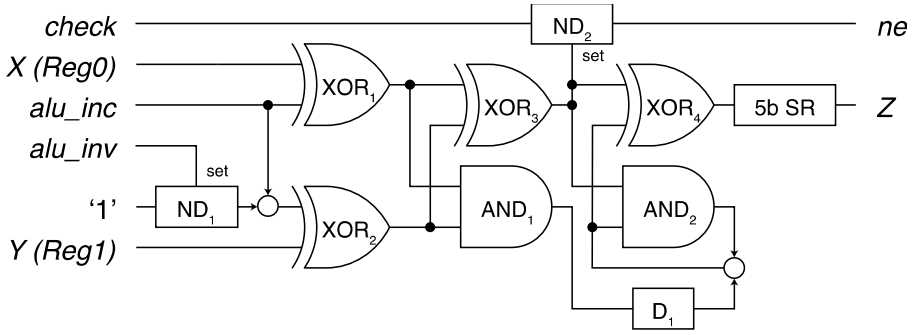


- **We carefully assigned opcode.**
  - **op7 and op6: the primary opcode**
  - **op5: a flag of ALU operation**
  - **op4: corresponds to reg1_read or a flag for conditional branch.**
  - **op3: a flag of indirect-addressing memory access.**
  - **op2, op1, op0: correspond to reg0_read, alu_inv, and alu_inc**

- BDD node count: 19
- 763 JJs, 213 μW
- 0.300 mm^2, 275k JJ/cm^2

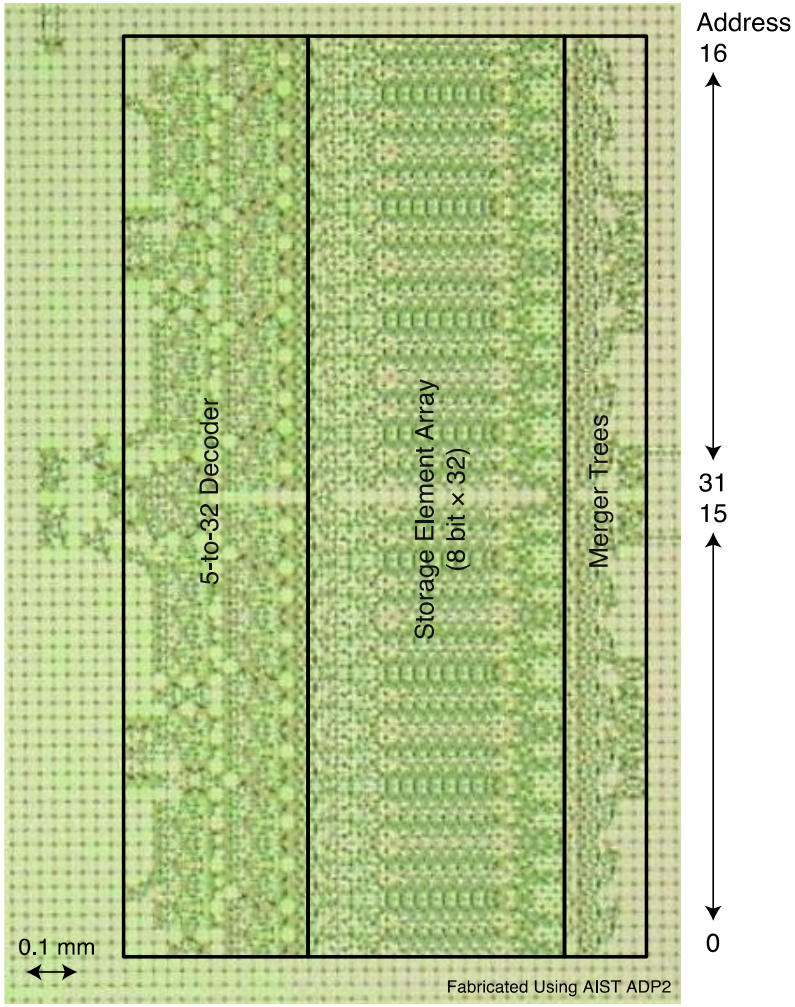
CORE e2 ALU

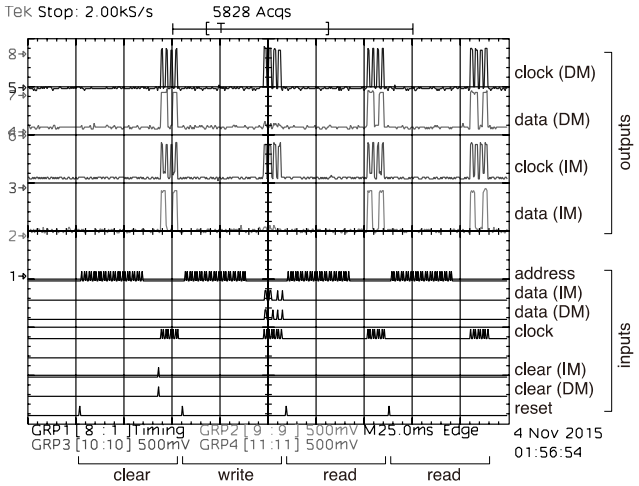




- Supporting Operations
  - ✓ ADD, SUB, MV, INC, DEC
  - ✓ Comparison (NE, LT)
- 348 JJs, 98.3 µW
- 0.13 mm^2, 274k JJ/cm^2

256-bit Shift-Register Memory

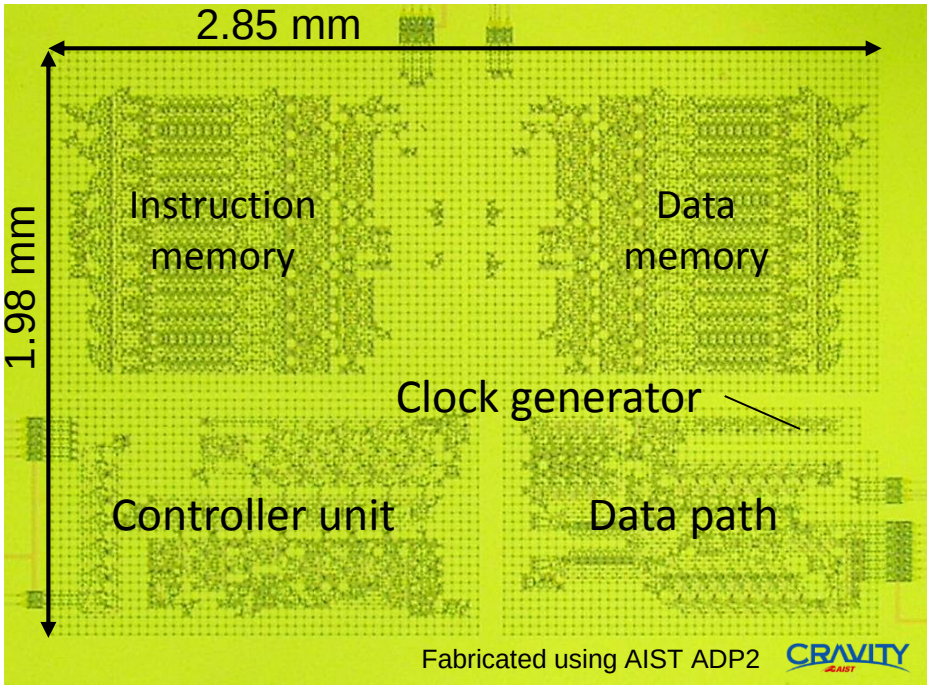


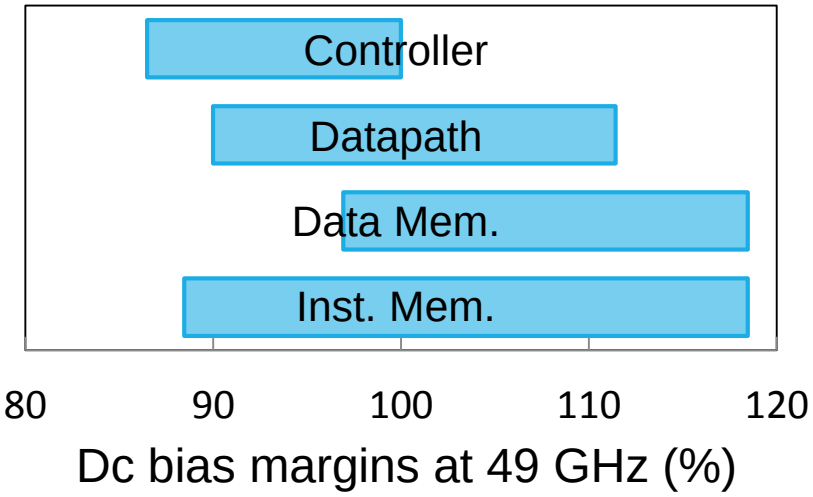


- 6597 JJs, 1.41 mW (Min. I_c = 70 μA)
- 1.38 mm × 1.95 mm
- 300k JJs/cm^2
- Access time (at 50 GHz)
  - Decode: 60 ps
  - Readout: 285 ps
  - Erase: 222 ps

M. Tanaka et al, *IEEE TAS* **26** (2016) 1301005.

CORE e2 Chip (128-bit memories)





- JJs count: 10603
- Bias current: 1007 mA
- Power: 2.52 mW
- Area: 5.64 mm^2

Test Programs

- Boundary: Small-scale programs written **within 16 lines**
  - **sum**
    - Calculate 1 + 2 + … + N
  - **total**
    - Calculate the sum of an array
  - **idiv**
    - Integer division
  - **aliq**
    - Compute the greatest divisor (Demonstrated program in SSEM)
  - **gcd**
    - Euclidean Algorithm to find the greatest common divisor (GCD) (Algorithm described in BC 300)

Inst.	Name	Instruction Meaning
HLT	Halt	Stop.
ADD	Add	Reg0 ← Reg0 + Reg1
SUB	Subtract	Reg0 ← Reg0 - Reg1
MV	Move	Reg0 ← Reg1
INC	Increment	Reg0 ← Reg0 + 1
DEC	Decrement	Reg0 ← Reg0 - 1
SKNE	Skip if not equal	if (Reg0 != Reg1) PC ← PC + 2
SKLT	Skip if less than	if (Reg0 < Reg1) PC ← PC + 2
LDR	Load register	Reg1 ← DM[Reg1]
STR	Store register	DM[Reg1] ← Reg0
JMP	Jump	PC ← addr
LD	Load	Reg1 ← DM[addr]
ST	Store	DM[addr] ← Reg0

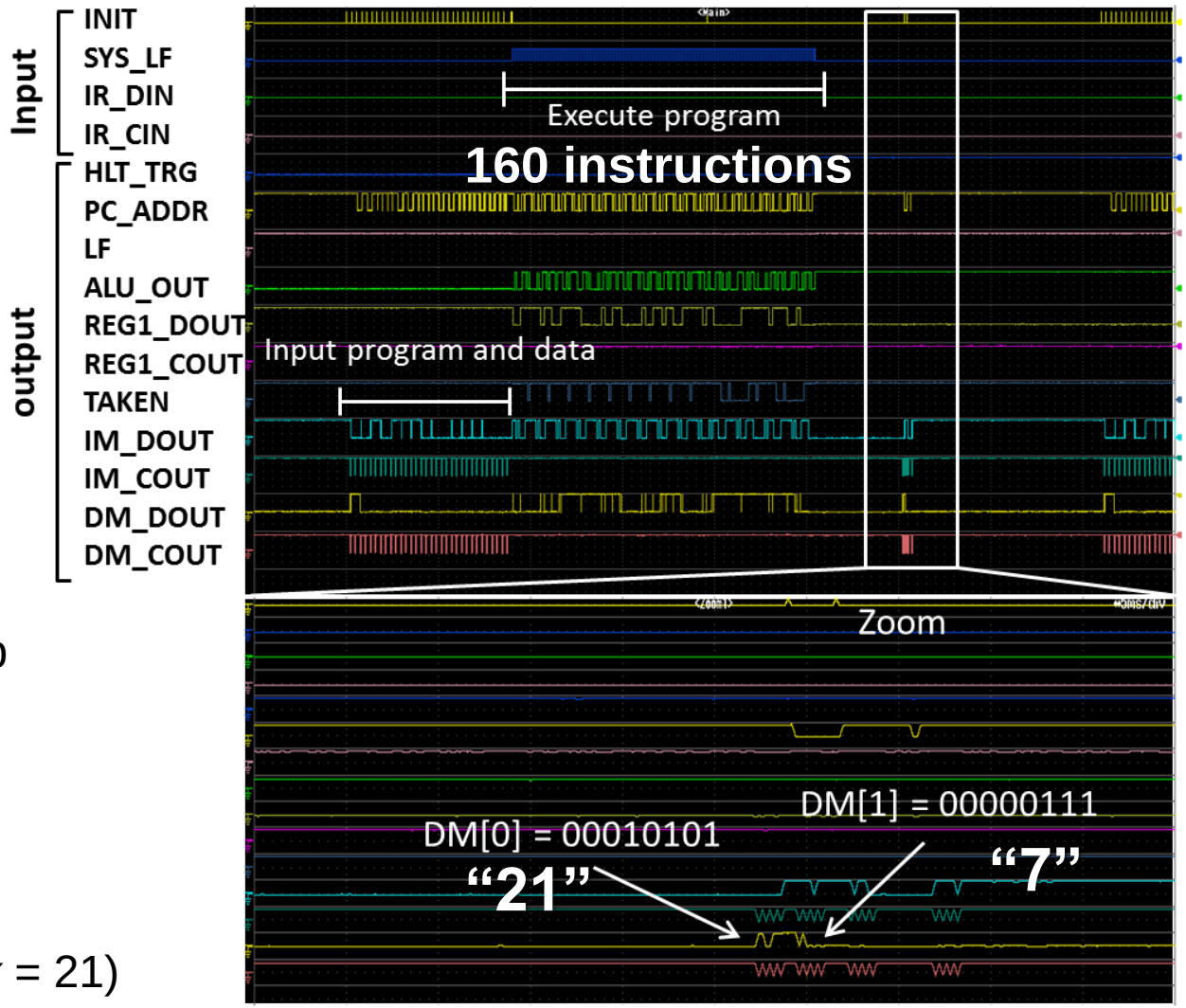
All the programs have successfully been demonstrated.

Program Demonstration (Aliquot X=21)

IM		DM	
00:	LD 00	00:	X: input
01:	MV	01:	Y: result
02:	DEC	02:	
03:	ST 01	03:	
04:	LD 00	04:	
05:	MV	05:	
06:	LD 01	06:	
07:	SUB	07:	
08:	SKNE	08:	
09:	HLT	09:	
0a:	SKLT	0a:	
0b:	JMP 07	0b:	
0c:	LD 01	0c:	
0d:	JMP 01	0d:	
0e:		0e:	
0f:		0f:	

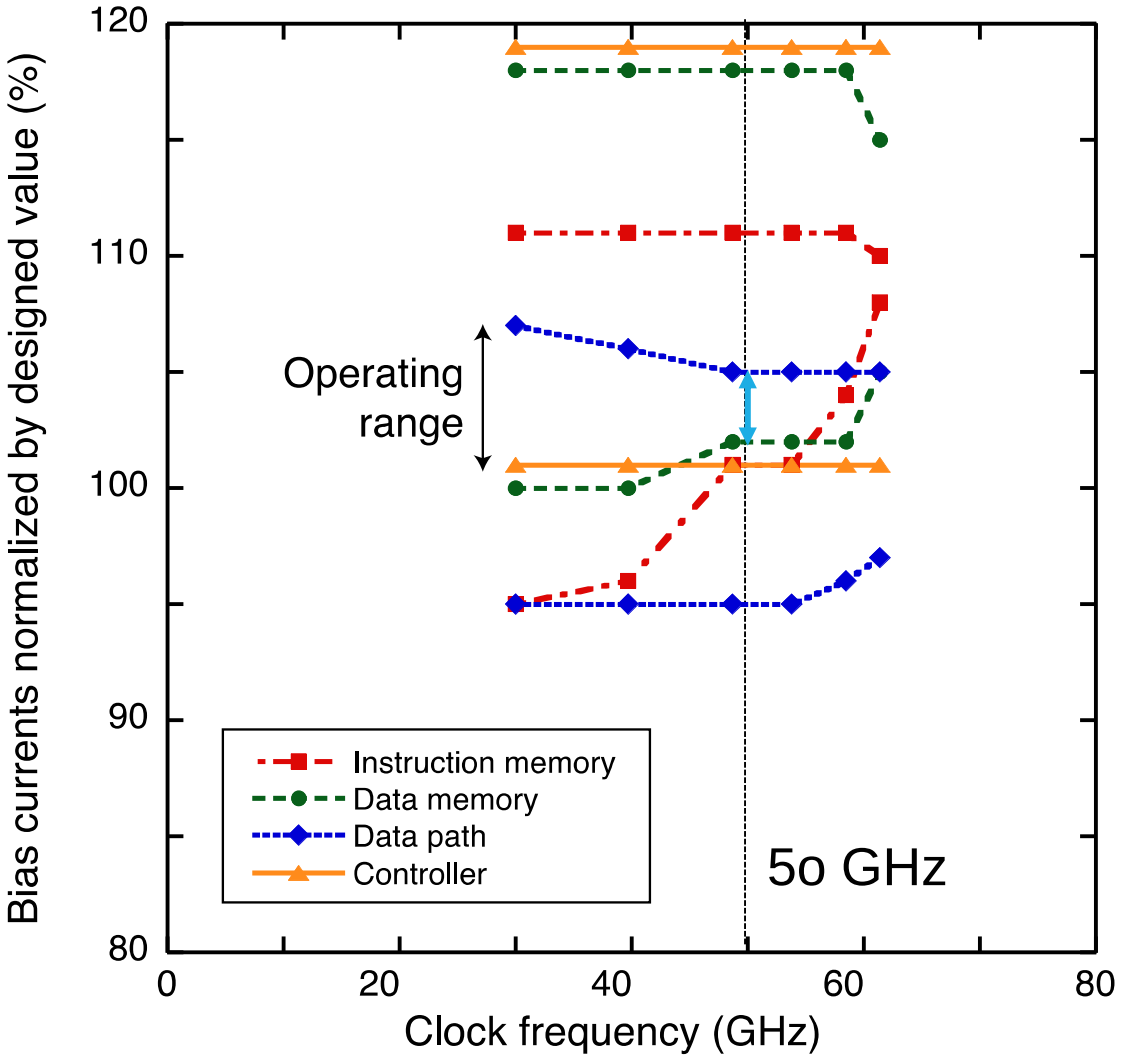
Double loop

Find greatest divisor of x (x = 21)



The third demonstration of the Aliquot program; Vacuum tube (1948), Si, and SFQ.

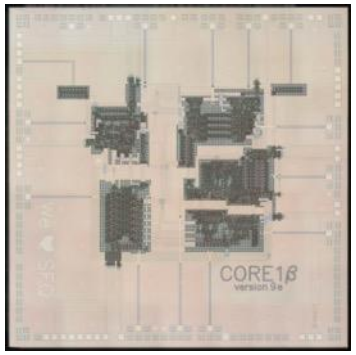
Frequency Dependence



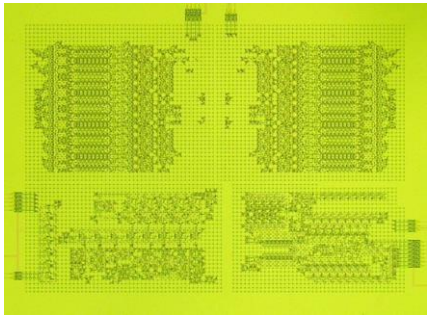
- ✓ Successfully executed programs composed of up to 200 instructions.
- ✓ Obtained comparable bias margins for five different test programs.
- ✓ **Overlap margin was 5% at 50 GHz**
- ✓ **Maximum frequency of bit-serial operation was 61 GHz**

Next Phase

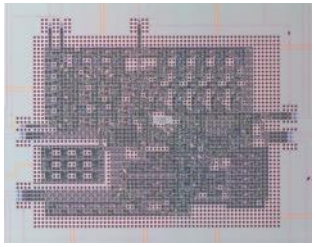
Bit-Serial (Complexity-Reduced)



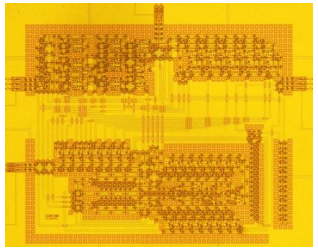
CORE1β MPU (2006)
25 GHz, 10955 JJs



CORE e2 MPU (2014–2016)
50 GHz, 10000–20000+ JJs



CORE1α MPU (2003)
15 GHz, 4999 JJs



CORE100 MPU (2015)
100 GHz, 3073 JJs

High energy-efficiency

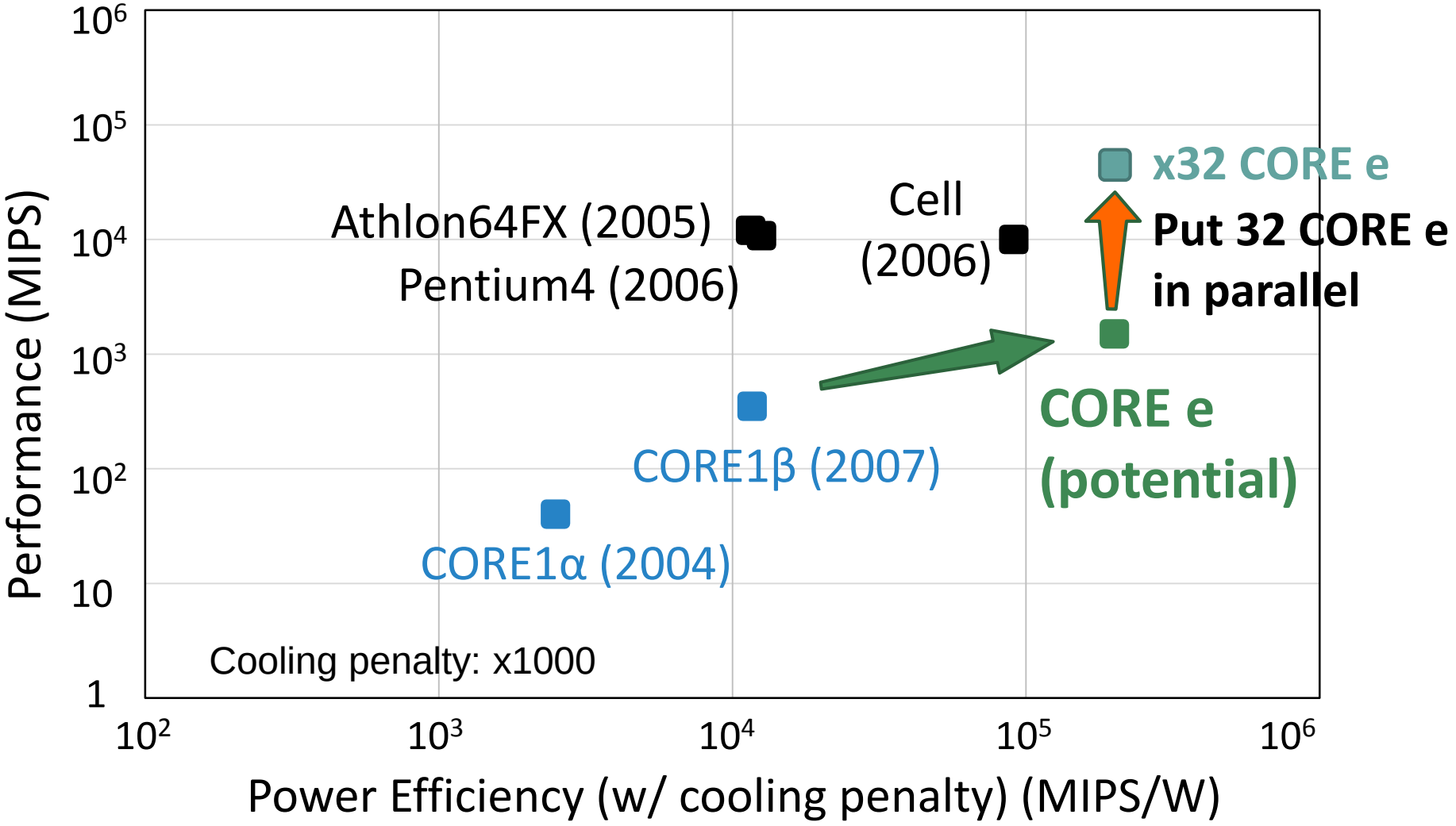
Manycore (Many-CORE e)

Low-Latency

Bit-Parallel, Ultrapipeline

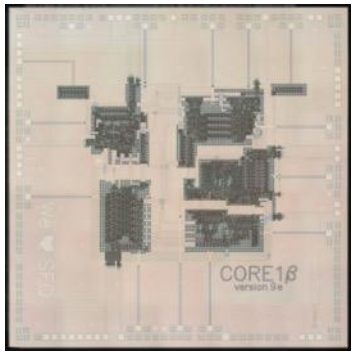
Prospect of Performance and Efficiency

(32 bit parallel or compatible)

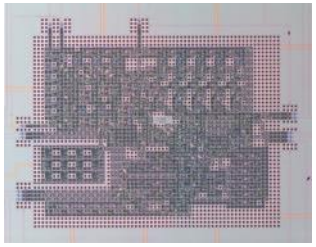


Next Phase

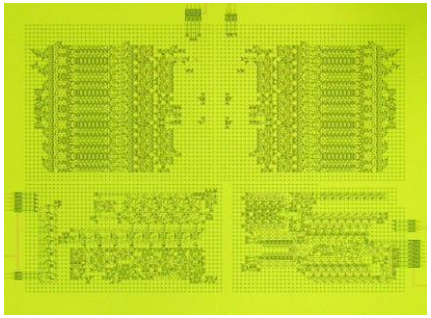
Bit-Serial (Complexity-Reduced)



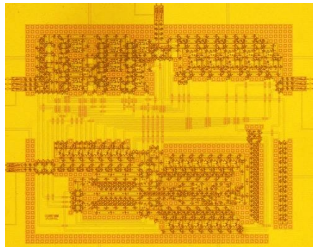
CORE1β MPU (2006)
25 GHz, 10955 JJs



CORE1α MPU (2003)
15 GHz, 4999 JJs



CORE e2 MPU (2014–2016)
50 GHz, 10000–20000+ JJs



CORE100 MPU (2015)
100 GHz, 3073 JJs

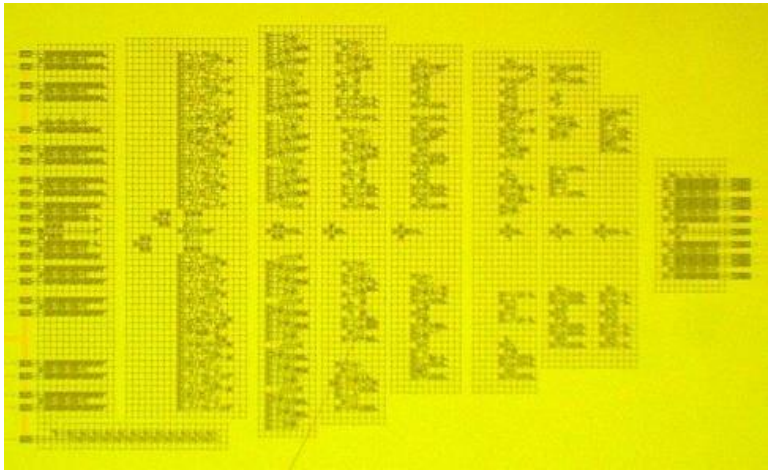
High energy-efficiency

Manycore (Many-COREe)

Low-Latency

Bit-Parallel, Ultrapipeline

Bit-Parallel Gate-Level Pipelined ALU



4868 JJs
50 GIPS
1.4 mW
36000 GIPS/W

Collaboration with Prof. Inoue, Kyushu Univ.

Specification

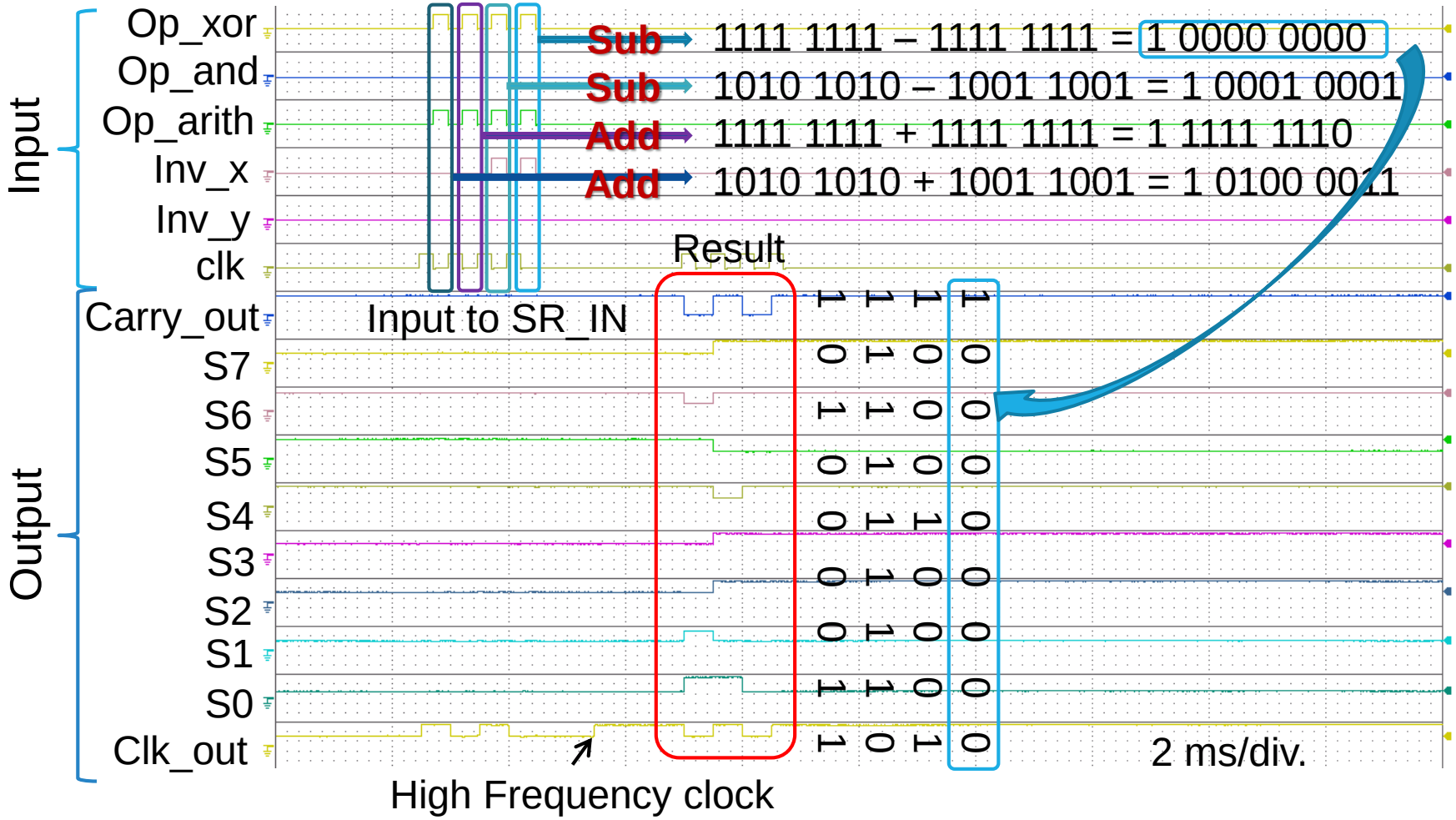
- ✓ Target frequency: 50 GHz
- ✓ Datapath: **Ultra pipelining (gate-level pipelining)**
- ✓ Bitwidth: 8 bits
- ✓ Functions: ADD, SUB, AND, OR, XOR, NOR, etc.

Based on Brent-Kung adder

- Minimum number of logic gates (w/o D flip-flops)
- Sparse wiring tracks
- Small fanouts (Max. 3)
- Maximum logic depth

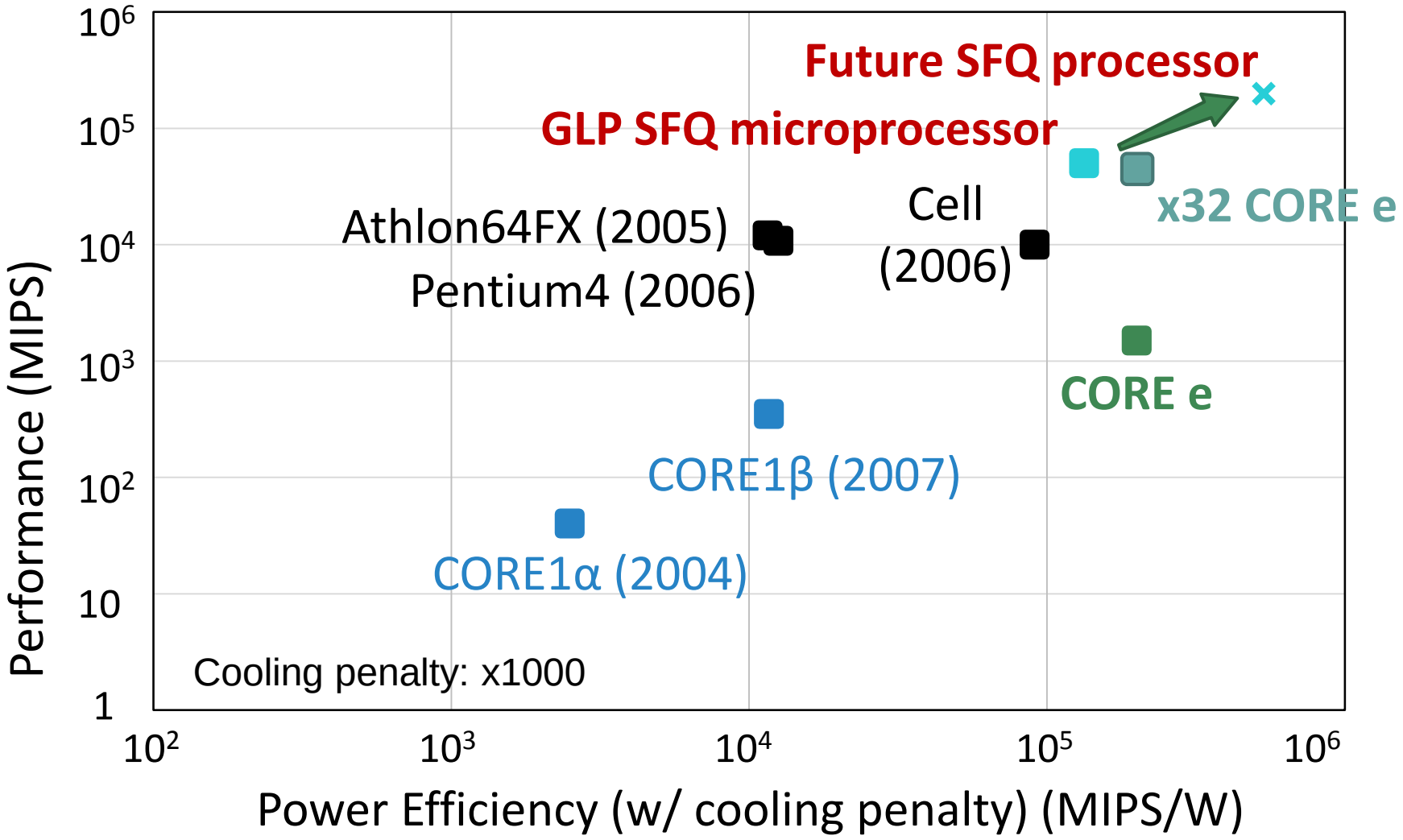
R. Brent and H. Kung, *IEEE Trans. Comput.* **G31** (1982) 260

Addition/subtraction in Parallel ALU



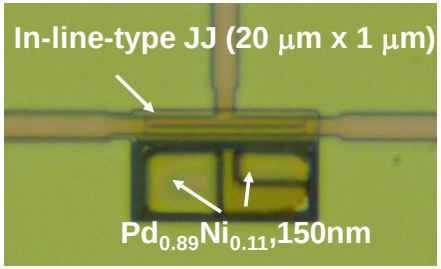
Prospect of Performance and Efficiency

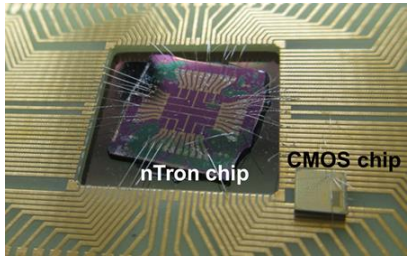
(32-bit parallel or compatible)



Main Issues Left for Practical Microprocessors

- Execution of meaningful programs stored in cryogenic memory with energy-efficient SFQ circuits
  - ✓ **resolved**
- High-frequency operation of bit-parallel processing for small latency
  - ✓ **resolved**
- Energy-efficient power supply for dc-powered SFQ circuits
  - ✓ **resolved by using superconducting diodes**
- Large capacity memory and voltage amplifiers with small footprints
  - ✓ **resolved by nano-cryotrons and CMOS memory**



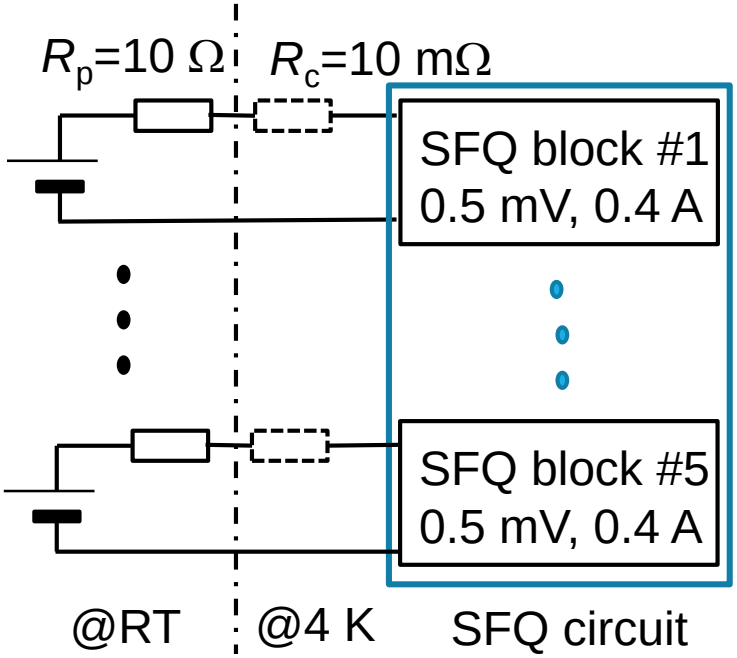


Summary

- Classical RSFQ circuits have matured over the decades.
- Programs stored in embedded memories have been demonstrated at 50 GHz, which is the first demonstration except Si technology
- Eight-bit-parallel processing has been executed at 50 GHz with the gate-level-pipelining technique.
- By introducing new technologies such as superconducting diodes or nano-cryotrons (nTrons), the issues for the practical applications are resolved.
- SFQ technologies should be combined with quantum information processing technologies for the next-generation supercomputers.

Thank you for your attention

AC/DC Converter for DC-Powered SFQ Circuits

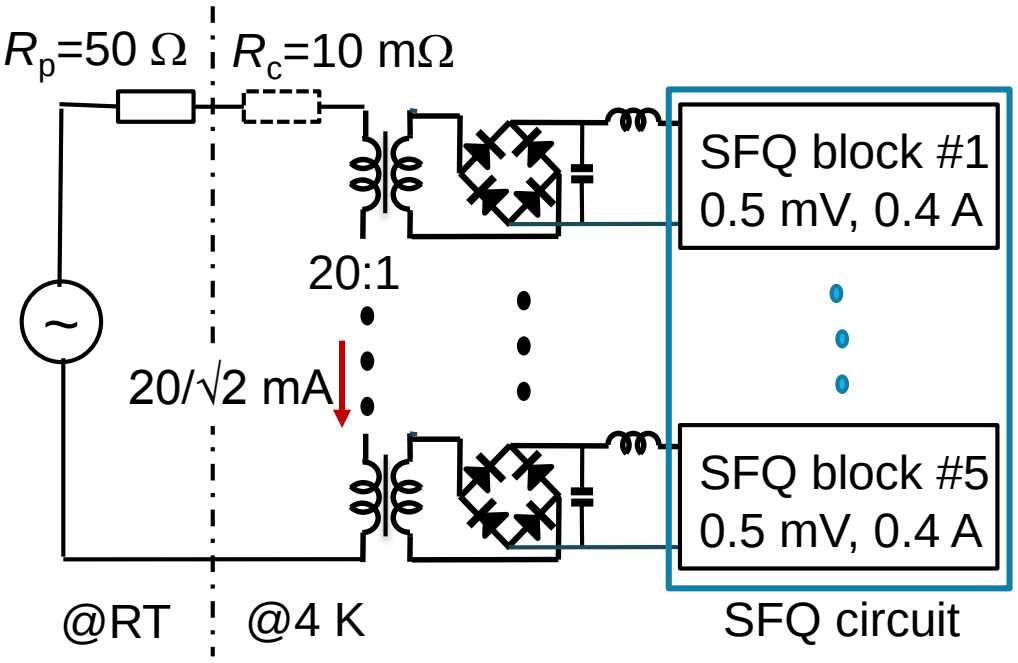


P_{supply}=8 W

P_{SFQ}=1 mW

$P_{contact}$=8 mW

Present



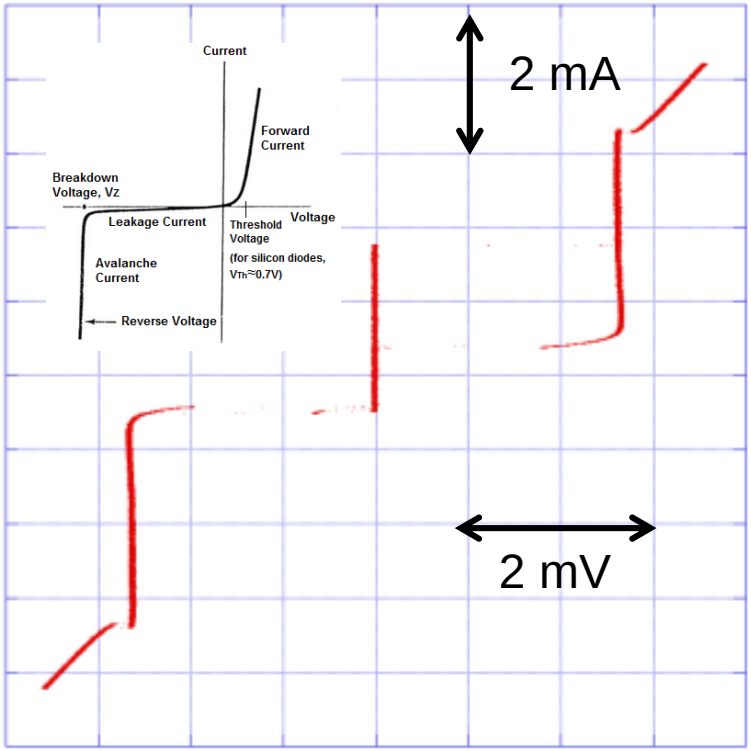
P_{supply}=10 mW $P_{contact}$≈0

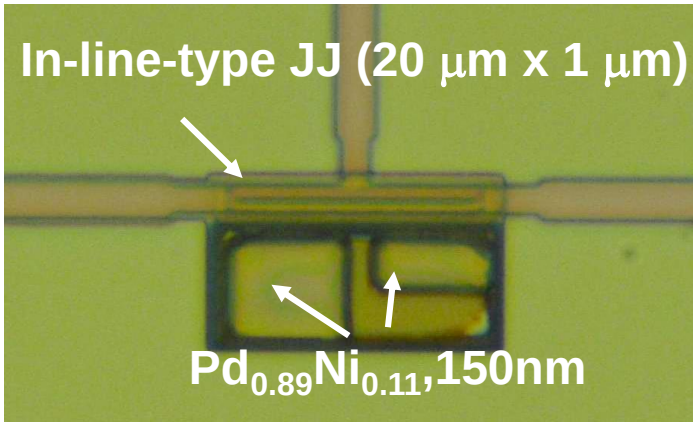
P_{SFQ}=1 mW

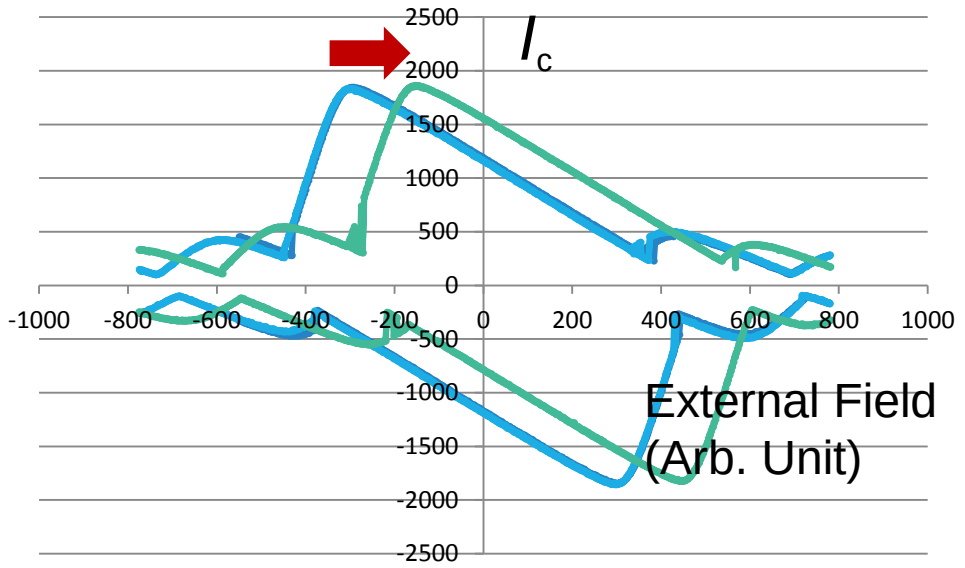
After introduction of superconducting diodes

AC/DC converter is essential for DC-powered SFQ circuits.

Superconducting Diode Based on Residual Magnetization

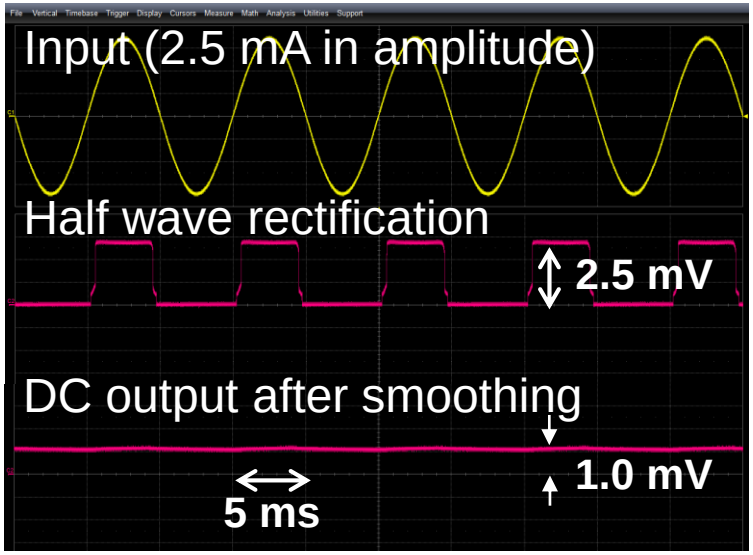


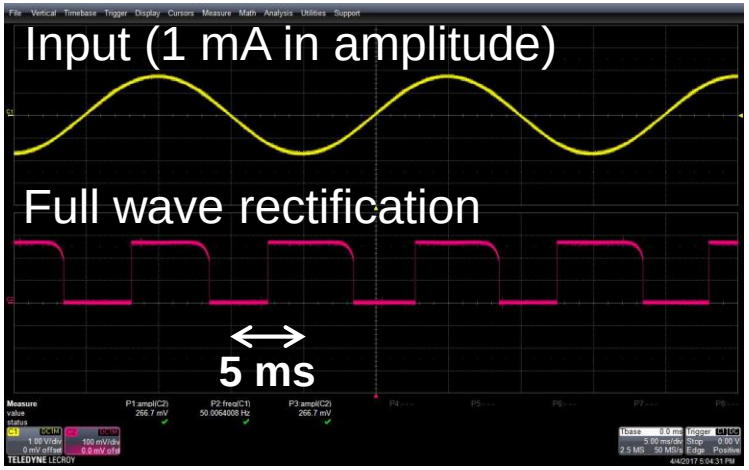




- A diode with V_{th}=0 is obtained.
- Critical currents can be controlled.

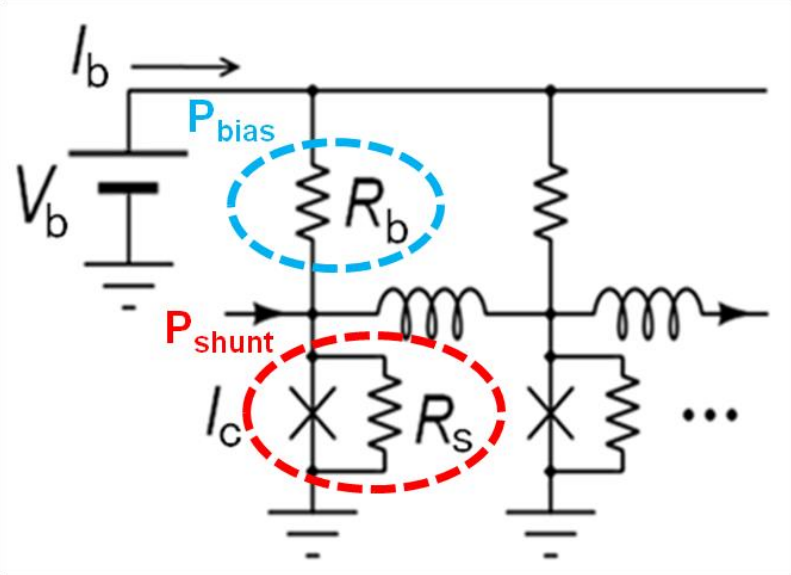
Rectification with Superconducting Diodes





We can control DC output voltages by changing the phase of the switching. This might open superconducting power electronics.

Issue for Energy-Efficiency



R_b is used for providing a constant current to each Josephson junction.

Power consumption at R_b
(Static power consumption)

$$P_{\text{bias}} = \frac{V_b^2}{R_b} \approx 0.7 I_c V_b$$

Example: DFF
$P_{\text{bias}} = 1.8\mu\text{W}$

Power consumption at R_s
(Dynamic power consumption)

$$P_{\text{shunt}} = f I_c \Phi_0$$

f: operating frequency

Example: DFF
$P_{\text{shunt}} = 36n\text{W}$

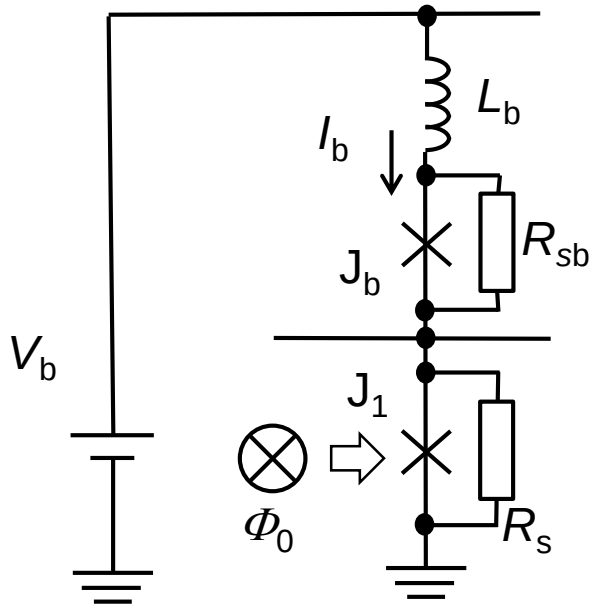
Typically, $I_c \Phi_0 \approx 2 \times 10^{-19} (J)$

Necessity for eliminating static power consumption.

DC-Powered Energy-Efficient SFQ Circuits

Bias resistors are replaced with inductors and junctions.

ERSFQ circuit (Hypres)



D. E. Kirichenko, et al., IEEE Trans. Appl. Supercond., **21**, 776(2011).

Advantage

- The base of design has been established because resources obtained from the RSFQ circuits can be used.
- PTLs can be used as interconnects.
- Possibly suitable to higher density because no mutual coupling is used.

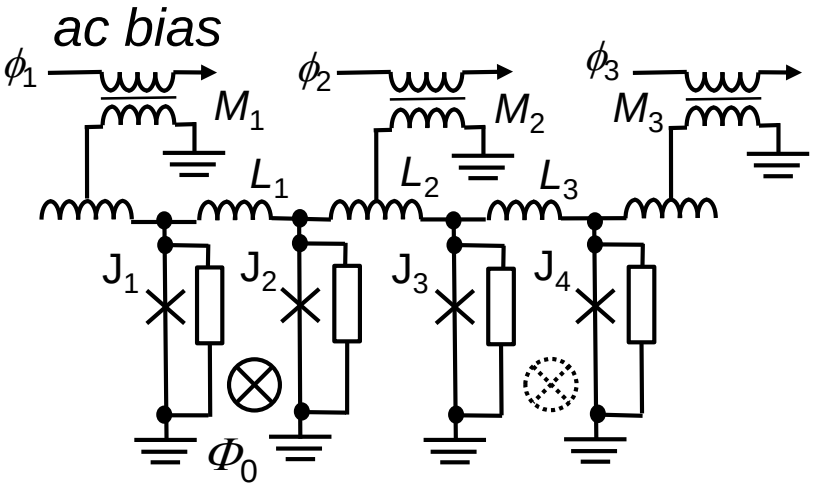
Disadvantage

- Difficult to make energy-efficient voltage supply around 0.1 mV.

AC-Powered Energy-Efficient SFQ Circuits

Circuits are driven by AC currents provided via transformers.

Example
Reciprocal Quantum Logic
(Northrop Grumman)



Q. P. Herr, et al., J. Appl. Phys., **109**, 103903 (2011).

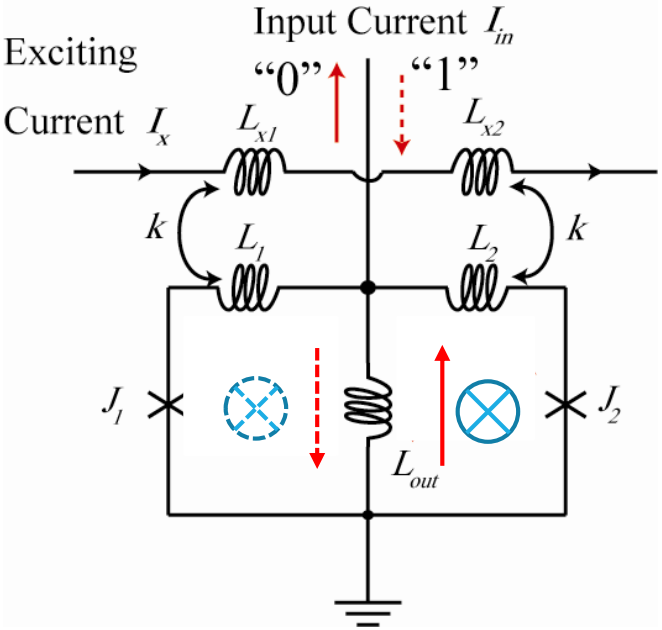
Advantage
- Provided AC currents are used as clock signals.
- NOT logic is easy to be made.
- The above means the RQL can be made up of smaller number of junctions.

Disadvantage
- Transformers are needed for all the gates, indicating downsizing to sub-micron scale is difficult.
- High-frequency design technique is essential for operation.

AC-Powered Energy-Efficient SFQ Circuits

Circuits are driven by AC currents provided via transformers.

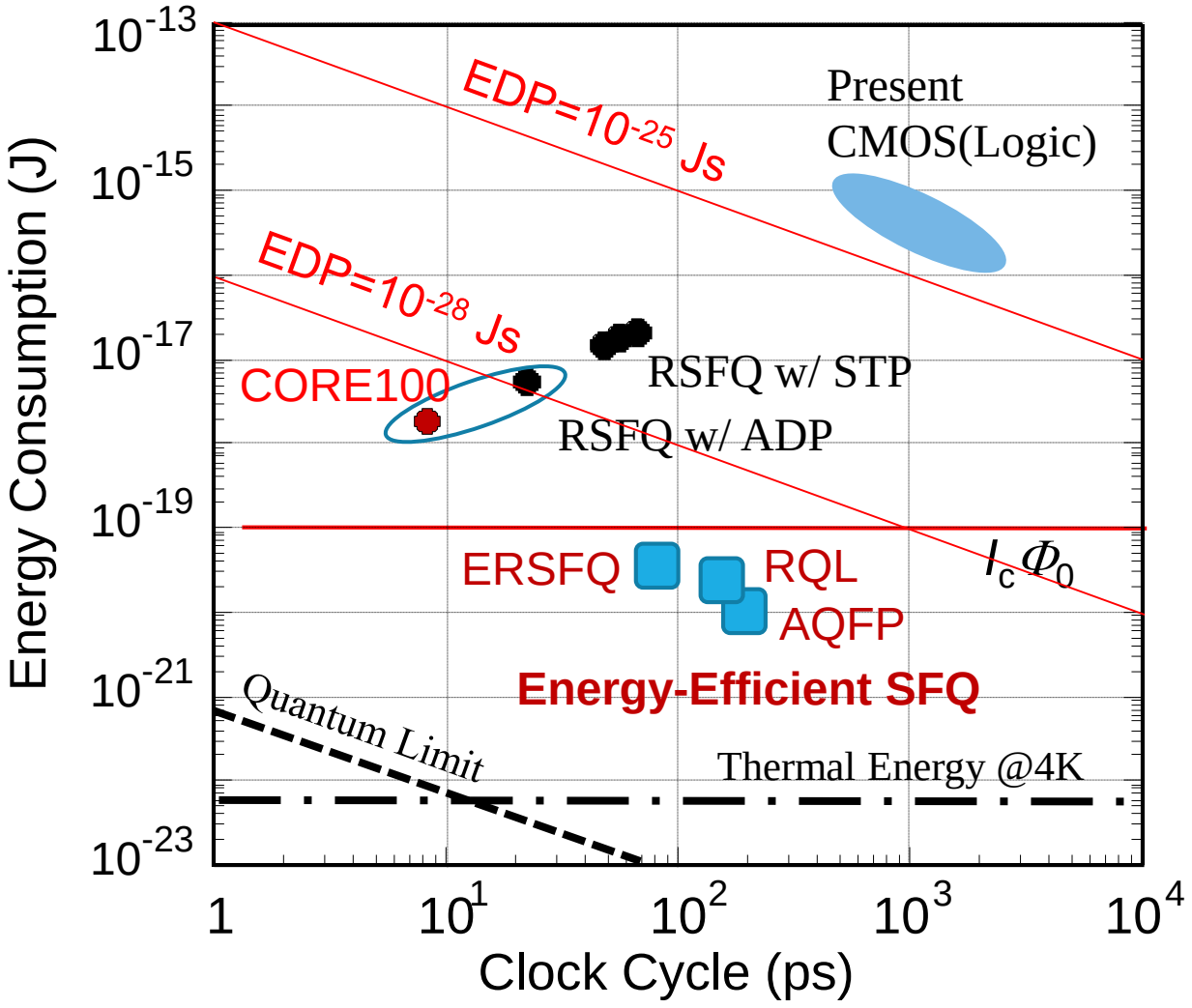
Example
Adiabatic Quantum Flux Parametron
(Yokohama Nat'l Univ.)



Advantage
- Very small energy consumption because of no phase jump in switching.
- All the logic operations are achieved based on a single 'majority' gate, leading to the robustness to the process variation.

Disadvantage
- Operating frequency is relatively low.
- Difficult to make long interconnects.
- DC offset currents are needed for

Energy-Efficiency in Integrated Circuits

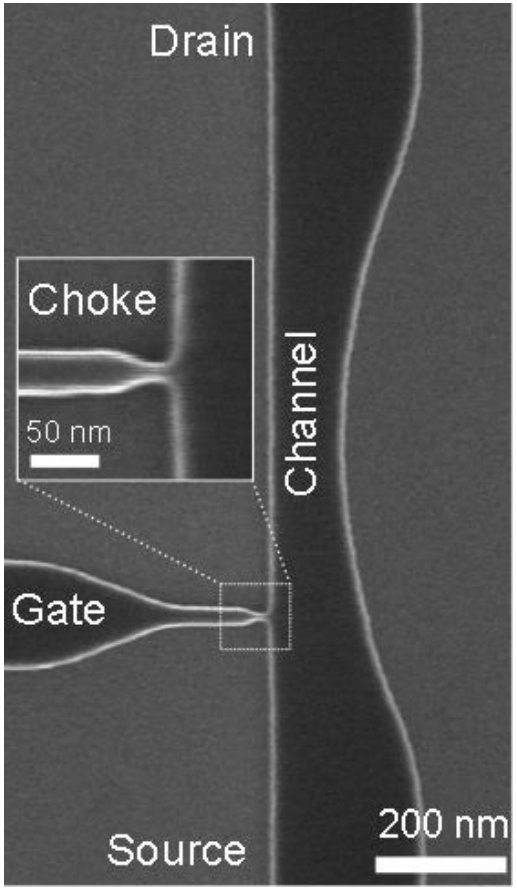


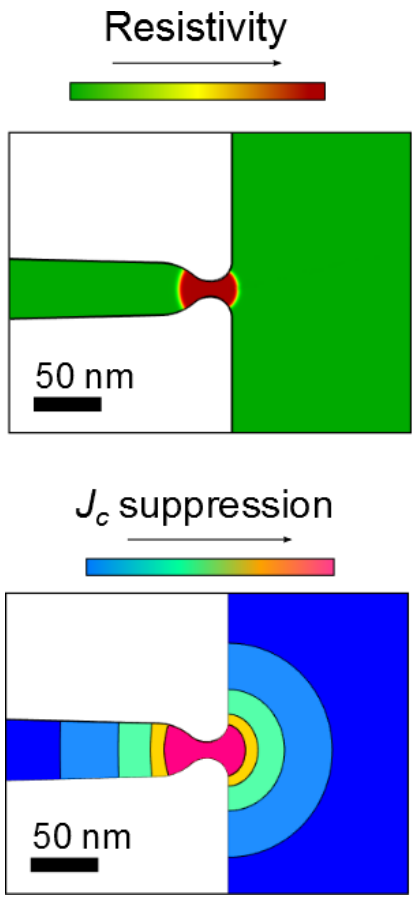
$$\text{Energy Consumption} = \frac{\text{Total power x Clk cycle}}{\text{Number of devices}}$$

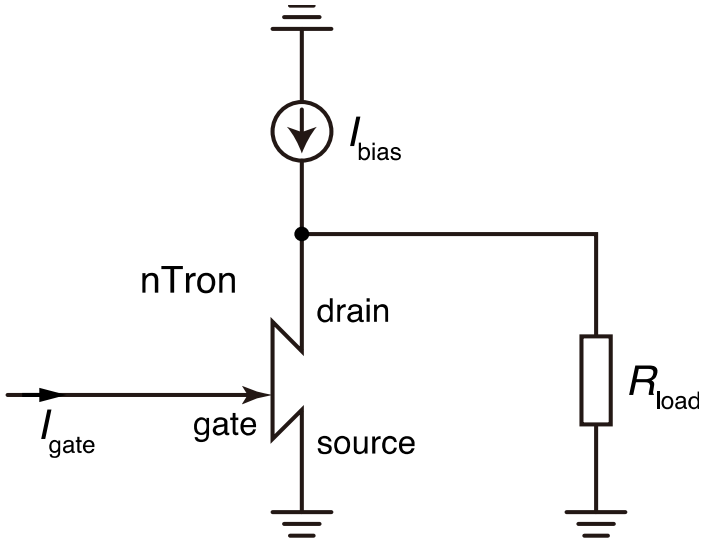
STP: AIST 2.5-kA/cm^2 Nb/AlO$_x$/Nb Standard Integrated Circuit Process.

ADP: AIST 10-kA/cm^2 Nb/AlO$_x$/Nb Advanced Integrated Circuit Process

Nanowire Cryotron (nTron)





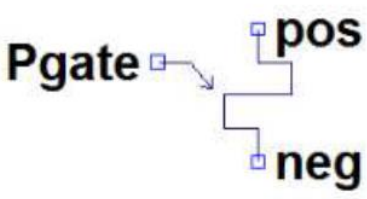


- ✓ Fabricated by a single NbN layer
- ✓ Switched by thermal assisting
- ✓ High output voltage (Sub-V)
- ✓ High-impedance (kΩ range)
- ✓ ~ 100 ps, 10^{-18} J/bit

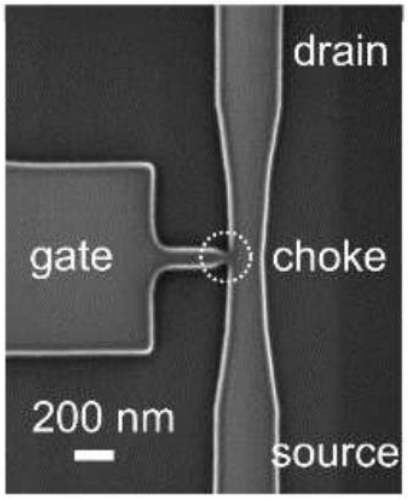
A. N. McCaughan and K. K. Berggren, *Nano Lett.* **14** (2014)

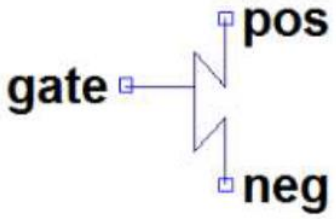
nTron Family (MIT)

Single nanowire SNSPD

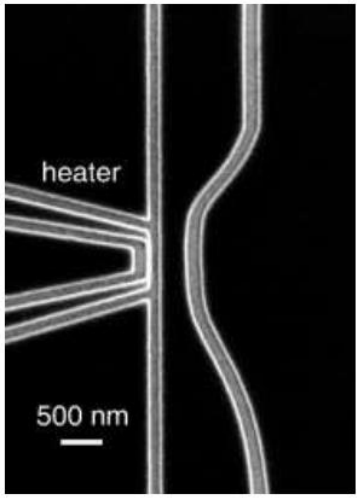


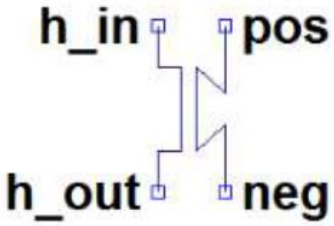
NanoCryotron (nTron)



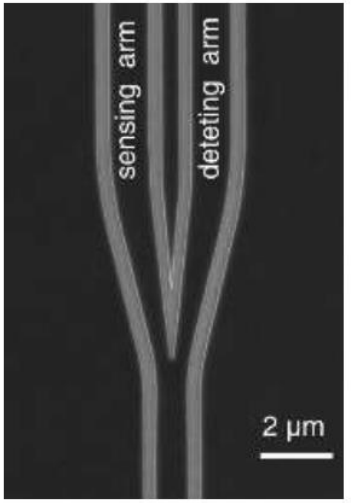


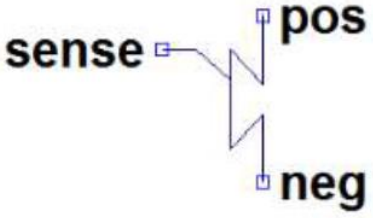
Gate isolated cryotron (hTron)





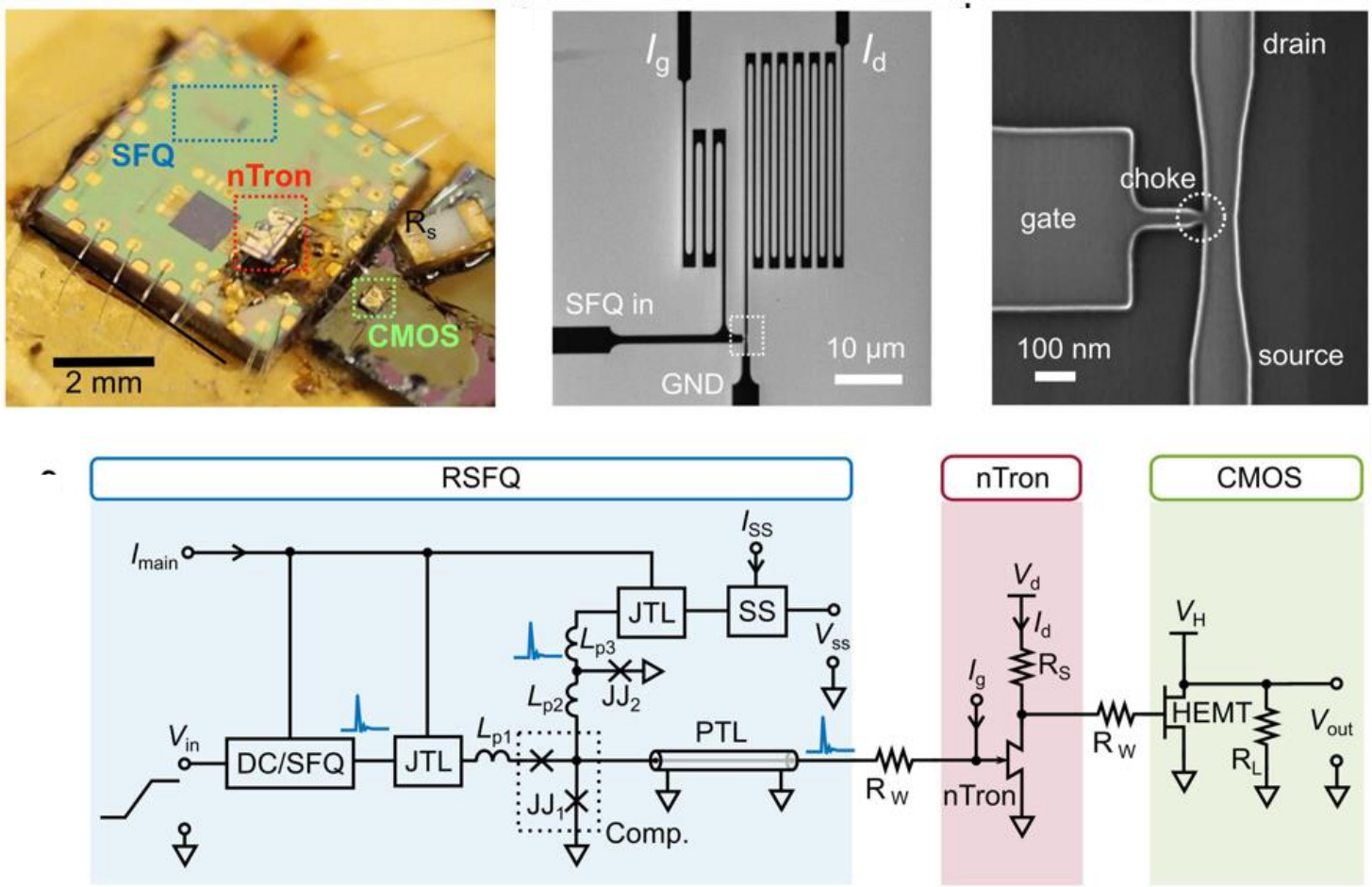
Current crowding cryotron (yTron)





Courtesy of Dr. Zhao (MIT)

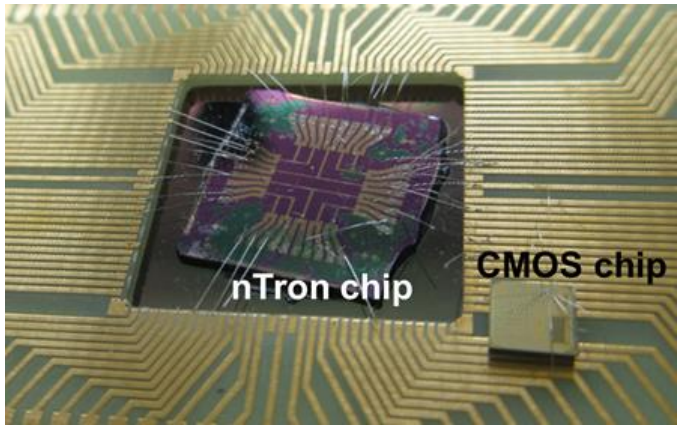
Demo. of nTron for Driving Semicon. Tr

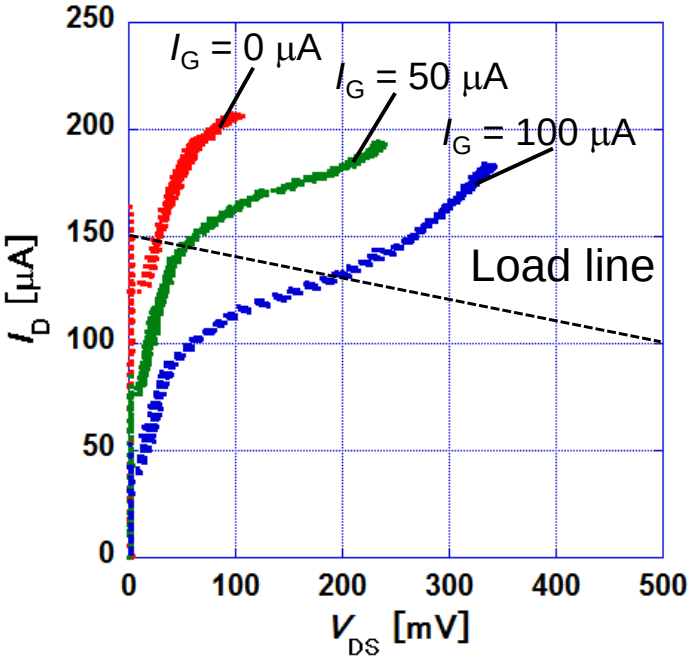


nTron can serve as a voltage amplifier needed between SFQ circuits and semiconductor circuits.

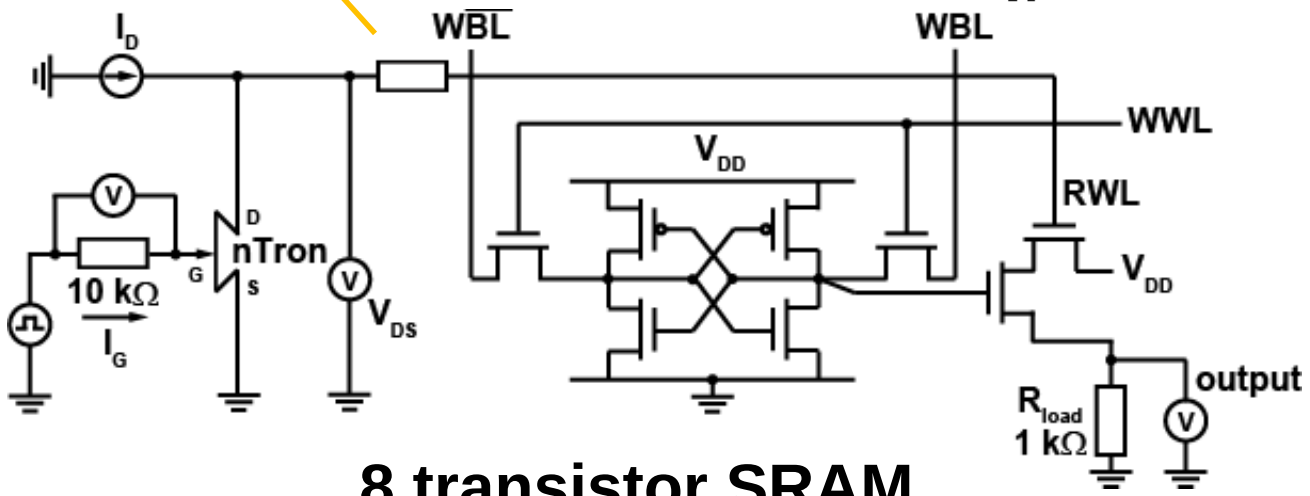
Q.-Y. Zhao et al, *Supercond. Sci. Technol.* **30** (2017)

NbTiN nTron + CMOS memory cell



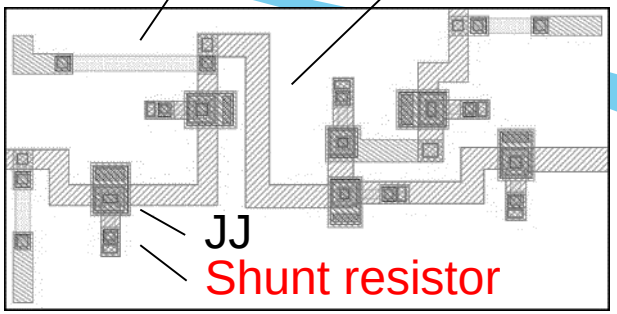


Al wire bonds



8 transistor SRAM

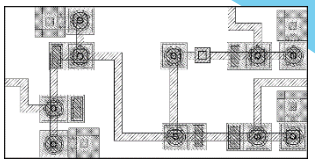
Issues for Larger-Scale Integration



D Flip-Flop: 40 x 80 μm
(Assuming Min. I_c = 50 μA)

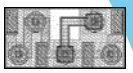
- Shunt-resistor-free JJs
- ERSFQ/eSFQ

(Elimination of resistors)



20 x 40 μm

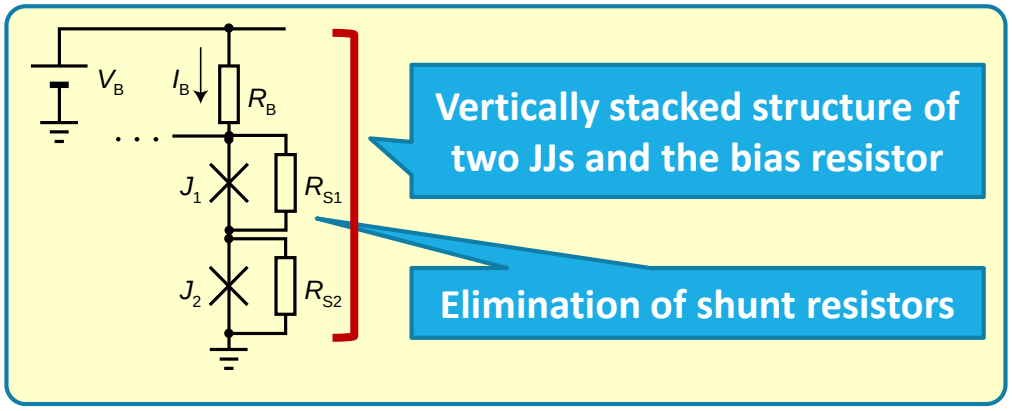
- High Sheet Inductance (NbN, etc.)
- JJ Stack



8 x 16 μm

- Equipment update

2 x 4 μm



Line and space: 1μm

0.25μm