

極低温超電導コンピュータ・アーキテクチャの開拓 ～古典から量子の世界へ～

井上こうじ

九州大学

副学長

大学院システム情報科学研究院 教授

システムLSI研究センター センター長

量子コンピューティングシステム研究センター長

日本・エジプト科学技術連携センター長

発表内容

コンピューティング
技術の現状

超伝導古典
コンピューティング

超伝導量子
コンピューティング

まとめ

発表内容

コンピューティング
技術の現状

超伝導古典
コンピューティング

超伝導量子
コンピューティング

まとめ

コンピュータの進化

1946年 世界最初の電子計算機ENIAC

設置面積：約167 m², デバイス：真空管17,468本, 性能：5,000回の加減算/秒, 消費電力150KW

<https://ja.wikipedia.org/wiki/ENIAC>

面積1/20万, デバイス数100万倍, 性能 25×10^6 倍, 消費電力1/500
劇的な境界条件の変化がコンピュータの姿と応用（そして社会までも）を変える！

2017年 Tesla V100 GPU (Voltaマイクロアーキテクチャ) NVIDIA

ダイ：約815mm², トランジスタ：約210億個, 性能：125テラFLOP/秒, 消費電力300W

コンピュータの発展が止まる! ～半導体微細化の終焉：ポストムーア時代へ～

— 1970 — 1980 — 1990 — 2000 — 2010 — 2020 — 2030 — 2040 →

主な
応用

計算指向：電卓、弾道計算、メインフレーム処理、など

メディア処理指向：静止画処理、動画像処理、音声処理、など

社会応用指向 (Society 5.0)：ビッグデータ処理、AI処理、最適化問題、など

世界初プロセッサ
4004 誕生

最新プロセッサ
Core-i9

半導体の微細化が
コンピュータの発展を支えた時代
(ムーアの時代)

トランジスタ数：2,400

80万倍

>20億

動作周波数：100 KHz

5万倍

~5GHz

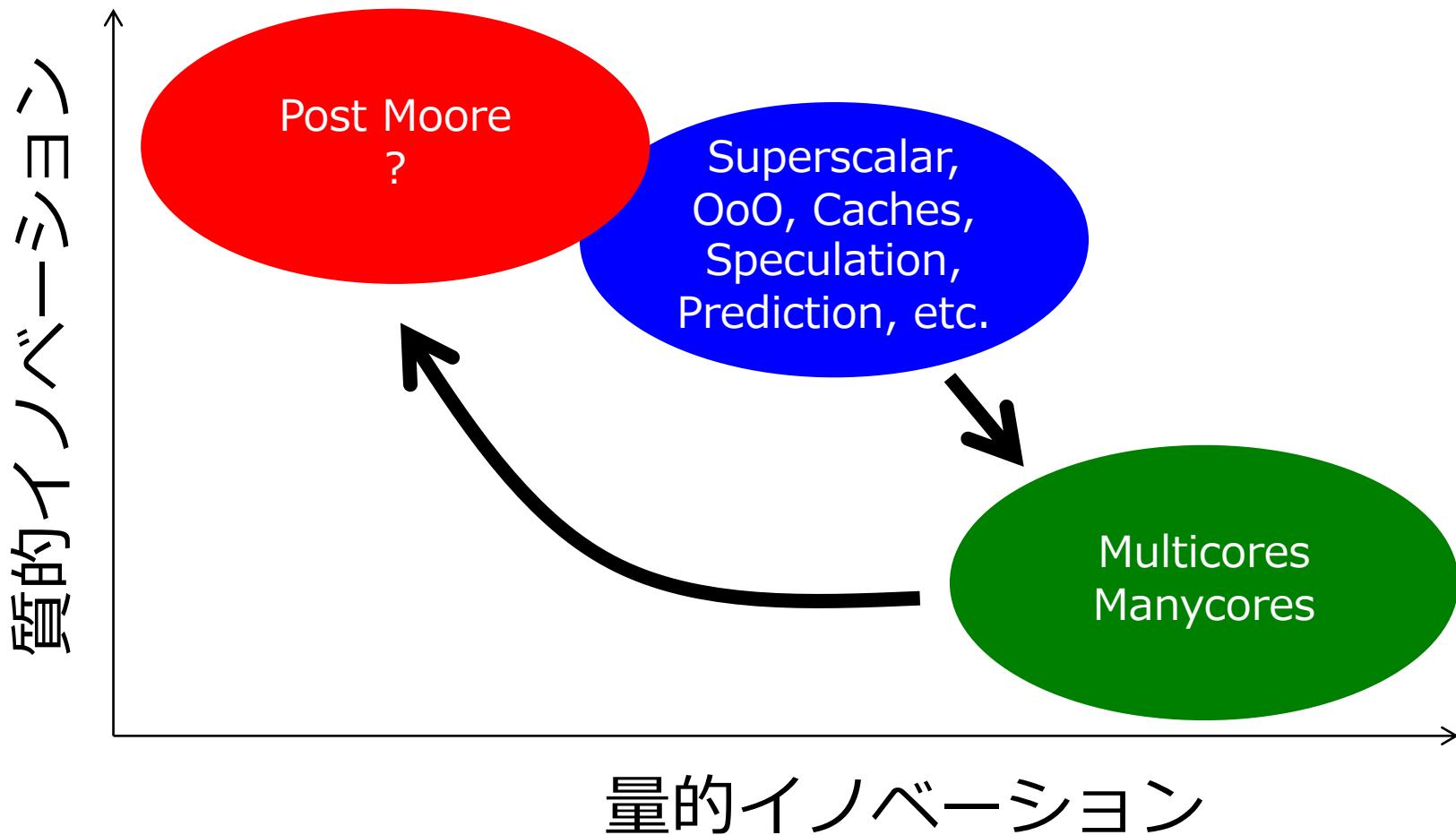


ポスト
ムーア時代

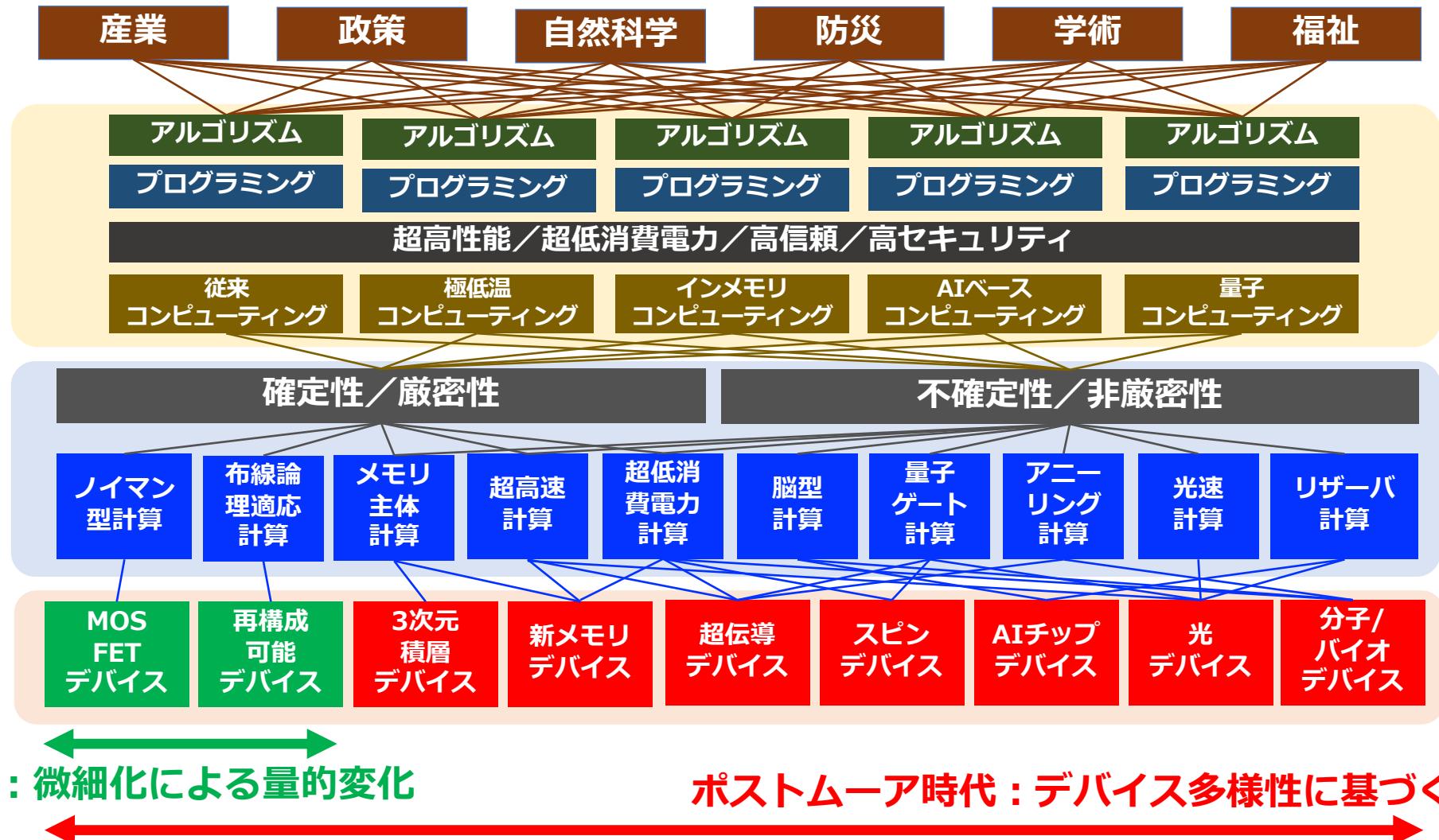
新奇デバイス活用
に新たな道を！
DS × CS

微細化限界

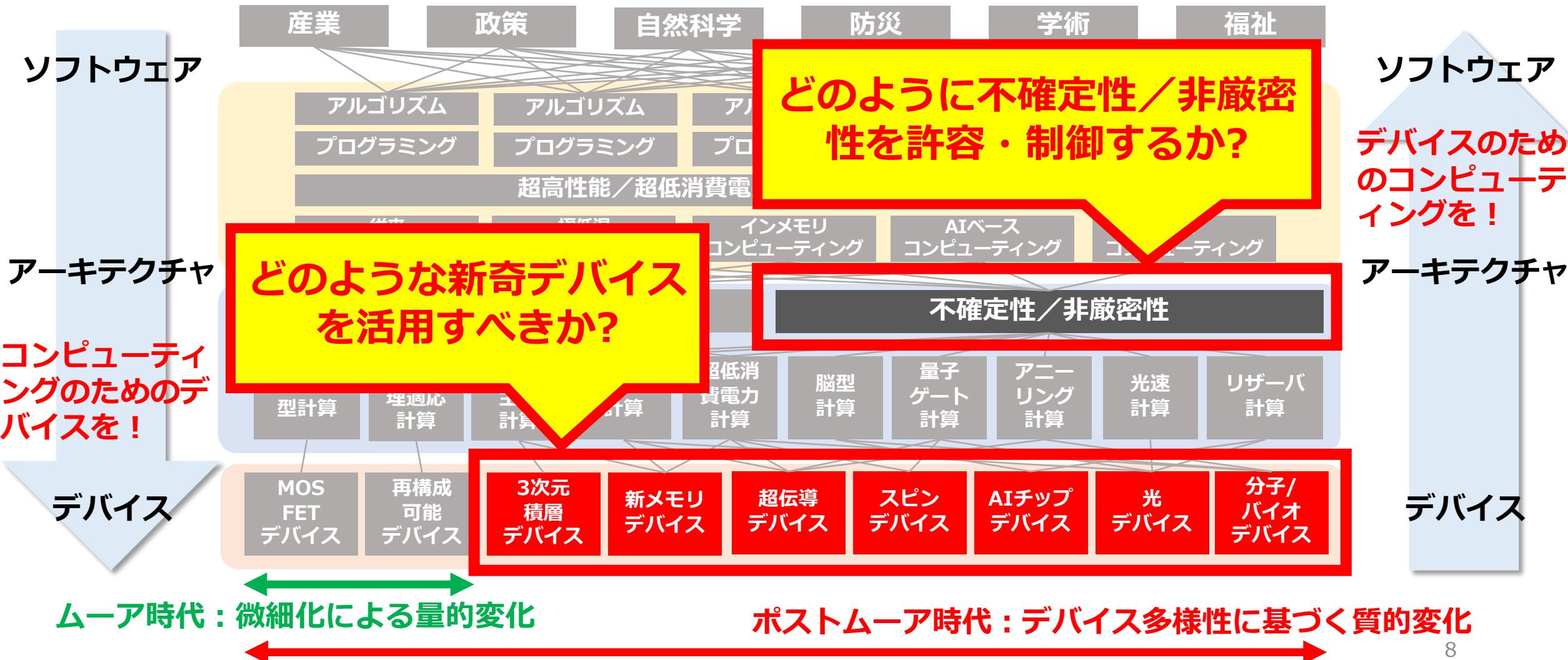
量的变化から質的变化へ



新奇デバイス活用による 「量的→質的アプローチ」のパラダイムシフト!



ポストムーア型コンピュータ構成法を確立するためには?



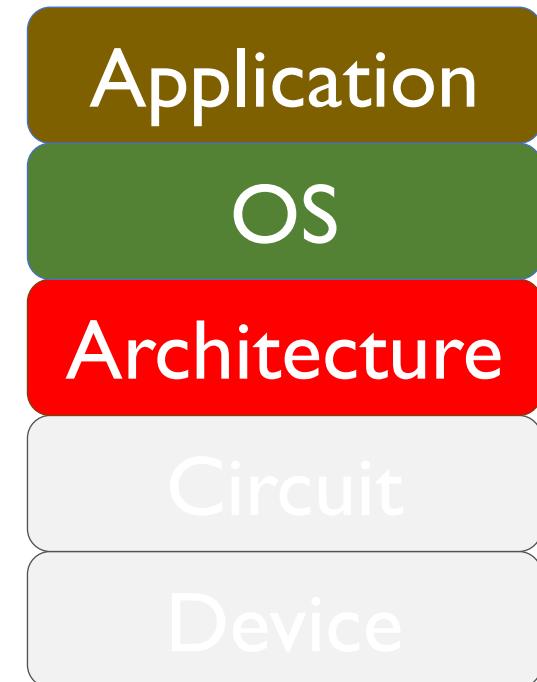
DS×CS これからの研究

これまでの研究

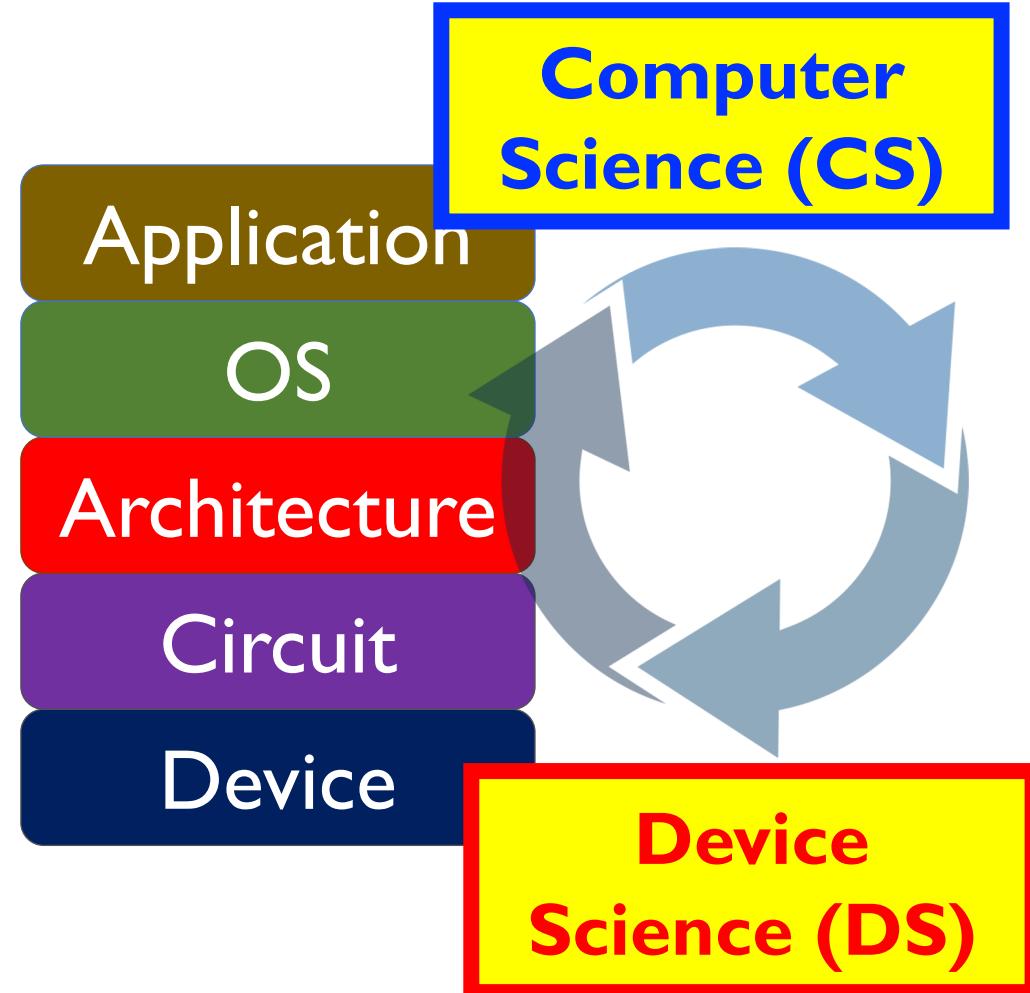


新奇Device

Device Science



Computer Science



Device
Science (DS)

発表内容

コンピューティング
技術の現状

超伝導古典
コンピューティング

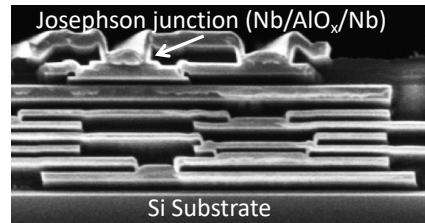
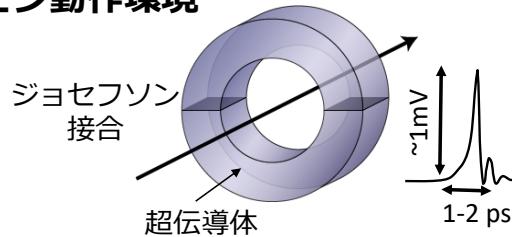
超伝導量子
コンピューティング

まとめ

单一磁束量子(SFQ)回路を用いた 極低温コンピューティング

SFQ回路の特徴

4ケルビン動作環境



世界的競争の現状

— 2010 — 2015 — 2020 —→

消費電力1/1,000
光と同程度の高速性

超微弱
SFQパルス (装置は米国・中国に3世代遅れ)
国内製造プロセス

世界の動向 (技術ロードマップ/学術)

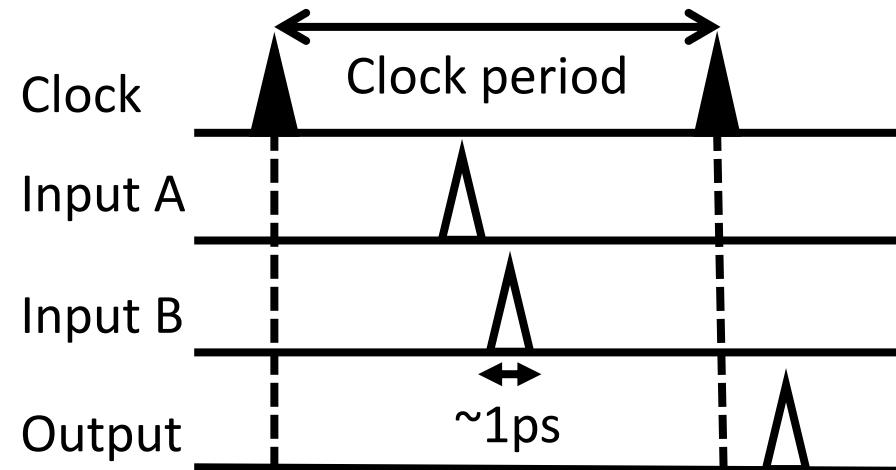
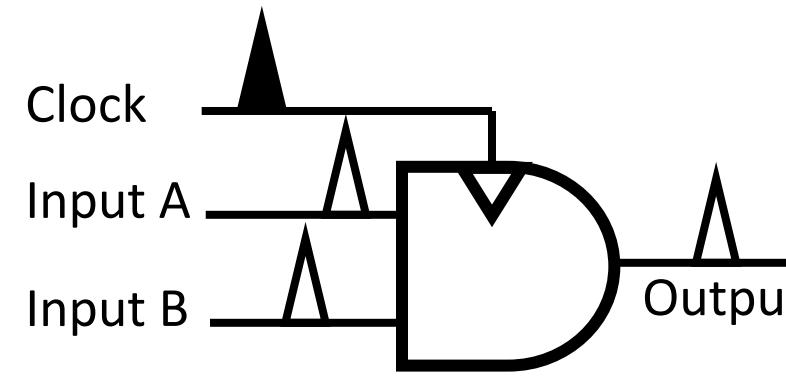
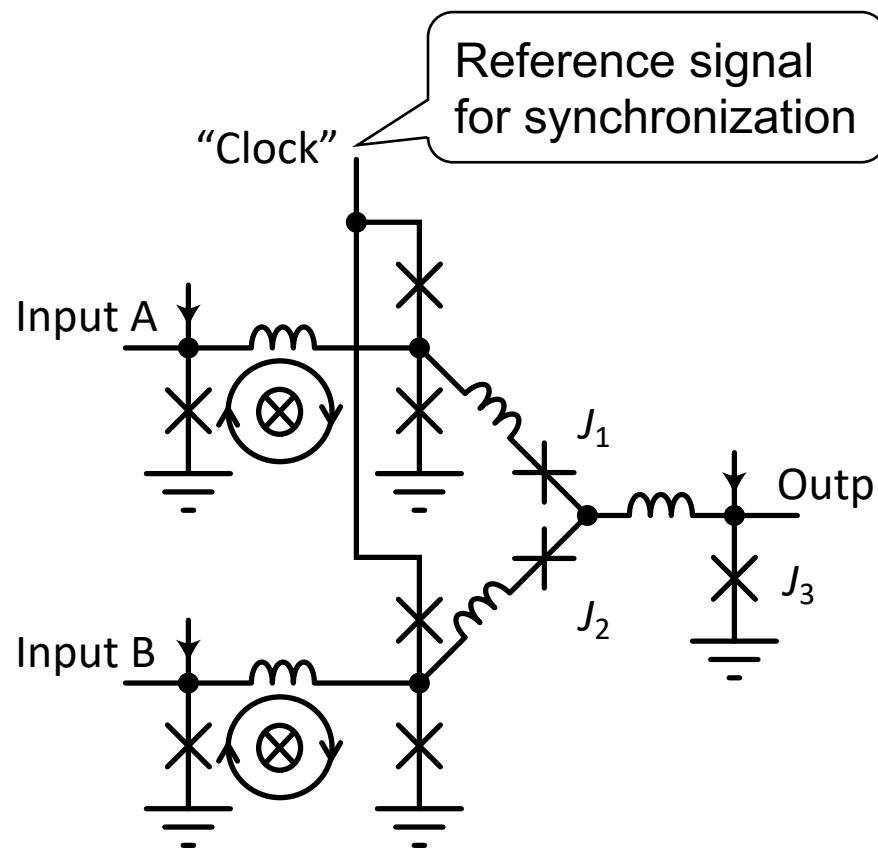
SFQアーキテクチャ国際WS

iARPA(米国高等研究開発局),
NSA(米国物理化学研究所),
米国立研究所, 米国各大学, 九大

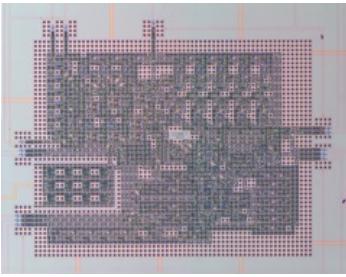
<https://www.iarpa.gov/index.php/research-programs/c3>
<https://www.scmp.com/news/china/society/article/2161390/can-china-build-us145-million-superconducting-computer-will>

SFQ ロジックゲート

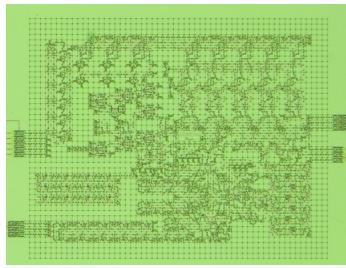
- 回路駆動のための「駆動パルス（Clock）」信号が必要
- 各論理ゲートはラッチ機能を有する



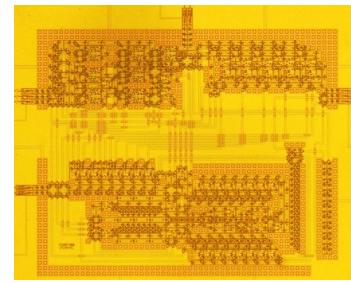
SFQマイクロプロセッサの歴史



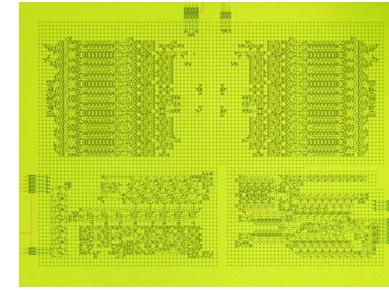
CORE1 α (2003)
4999 JJs, 15 GHz
167 MIPS, 1.6 mW



CORE1 α LV (2013)
3869 JJs, 35 GHz
400 MIPS, 0.23 mW



CORE100 (2015)
3073 JJs, 100 GHz
800 MIPS, 1.0 mW



CORE e2 v5h (2016)
10603 JJs, 50 GHz
333 MIPS, 2.5 mW

Energy-
efficiency

Ultra High-
Frequency

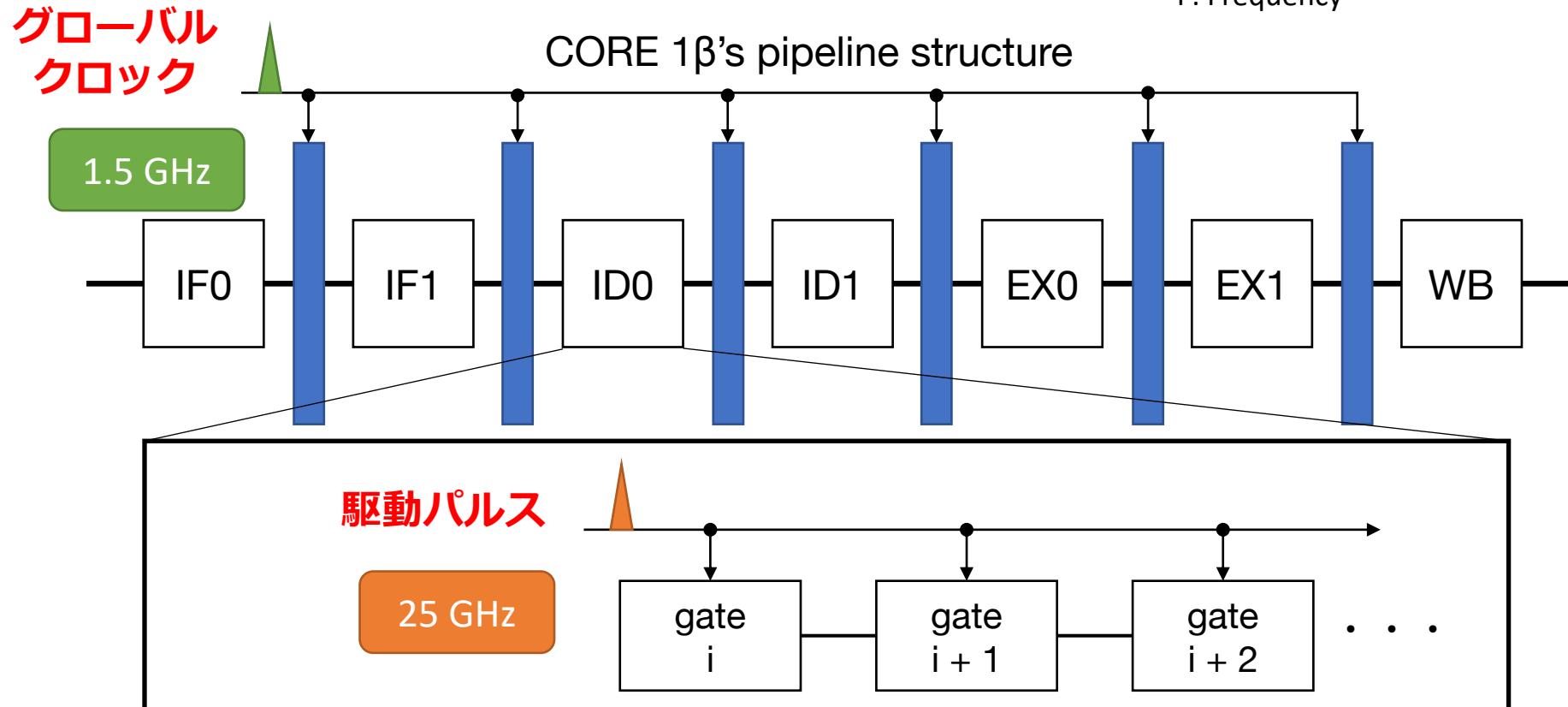
Stored-program
computing demo

SFQプロセッサの問題 → ディバイバス特性を考慮したアーキ?

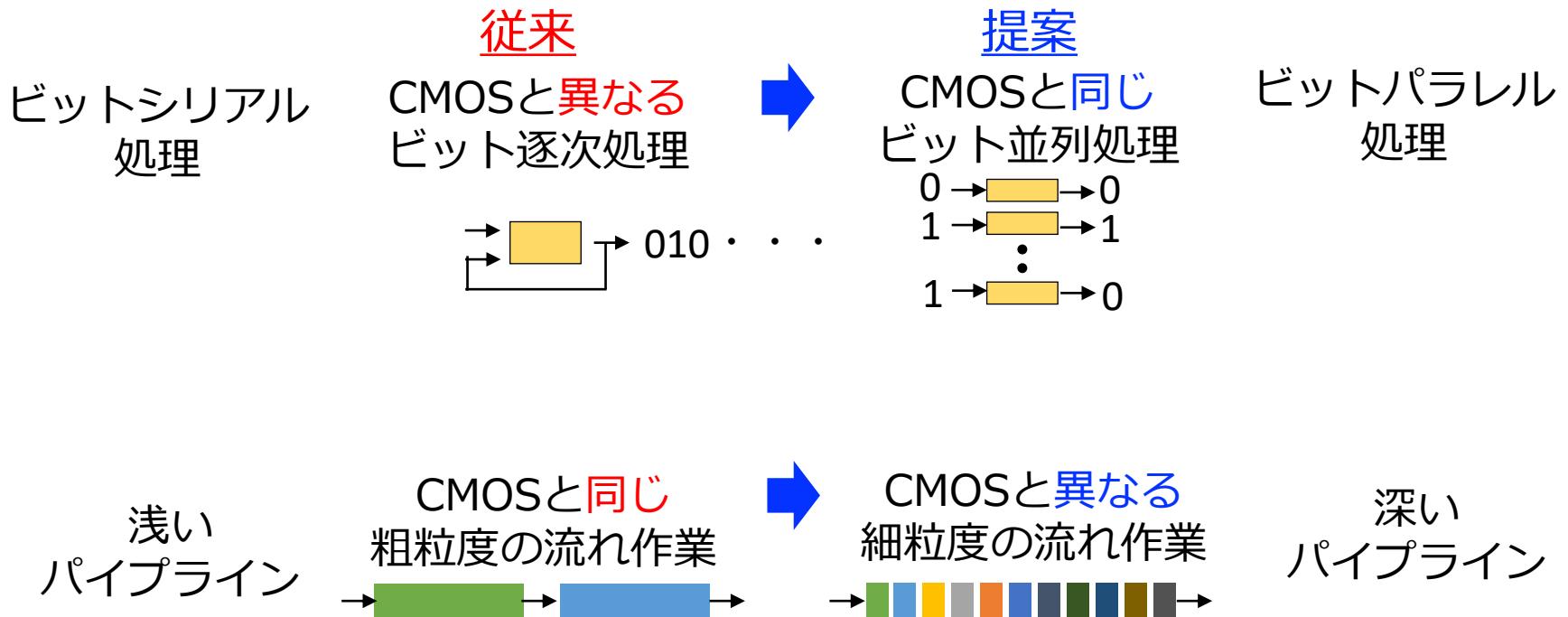
性能 = 1 / 実行時間

$$\text{実行時間} = \text{IC} \times \text{CPI} \times 1/\text{F}$$

IC: Instruction Count
CPI: Clock cycles Per Instruction
F: Frequency



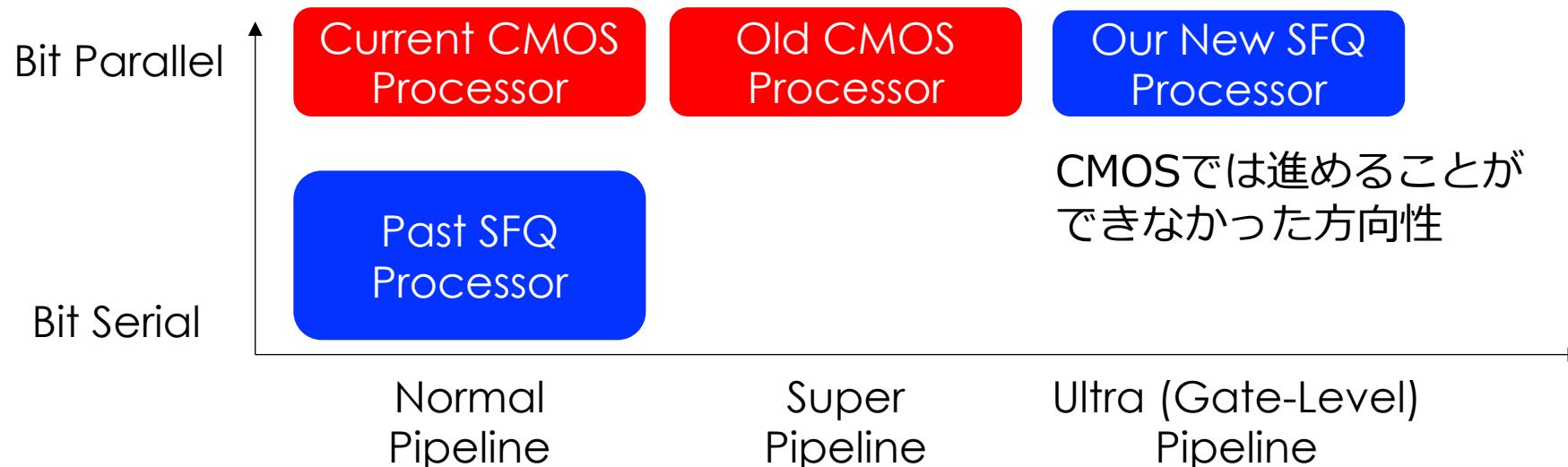
SFQ向けアーキテクチャを考える!



ビットパラレル型ゲートレベルパイプライン・アーキテクチャを考案!

我々のアプローチ

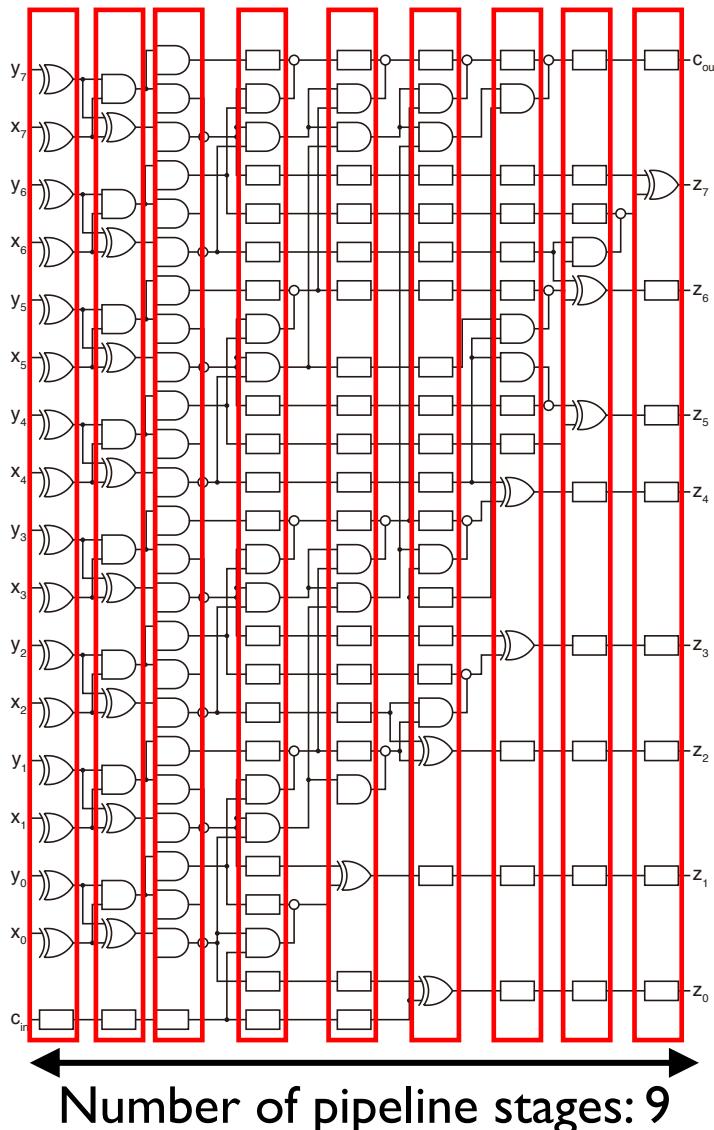
	現行 CMOS マイクロプロセッサ	SFQ マイクロプロセッサ	
		CORE e2[2]	提案
命令パイプライン の深さ	浅い (数～十数段)	浅い (6段)	深いゲートレベル命令パイプライン
データパスビット 幅	ビットパラレル	ビットシリアル	ビットパラレル
パイプライン ハザードの対処	OoO実行等によるハ ザード回避	パイプライン ストール	細粒度マルチスレッディ ングによるストール隠蔽



最初の意見

そんなの高速で動くわけがない！

やってみよう！8-bit Bit-Parallel ALU



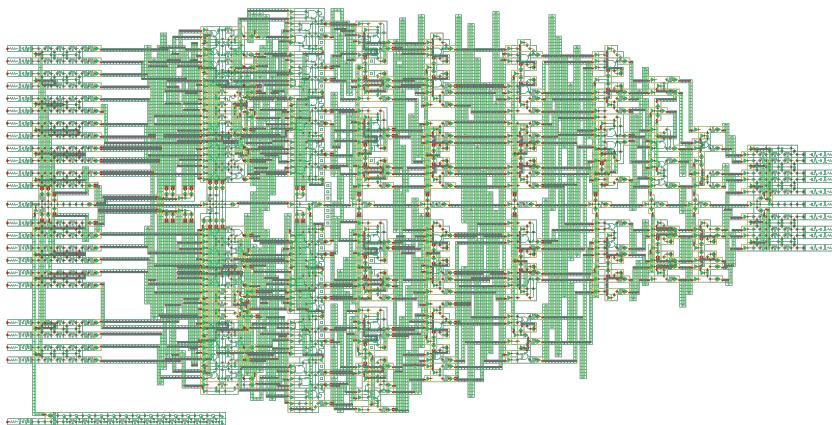
- ✓ Target frequency: 50 GHz
- ✓ Gate-level pipelining
- ✓ Functions: ADD, SUB, AND OR, XOR, NOR, etc.
- ✓ Data length: 8 bits

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

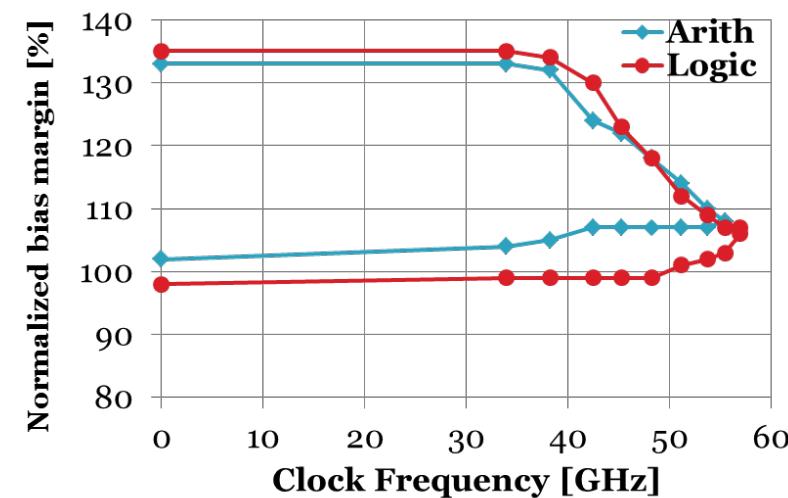
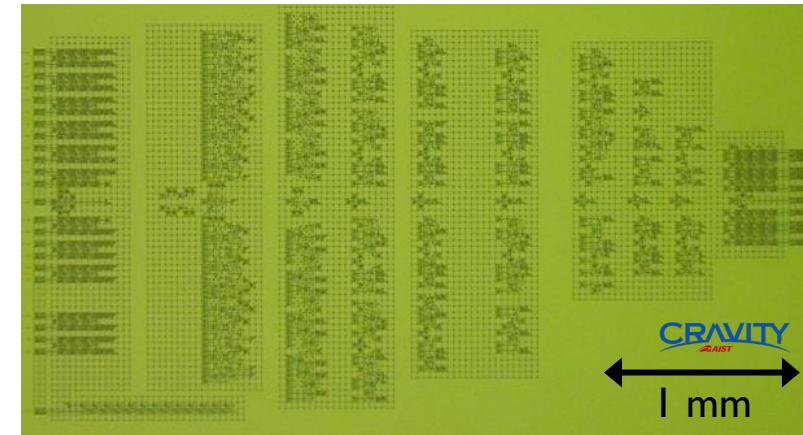
It Works!

**ISLPED'17 Design Contest
Honorable Mentions**

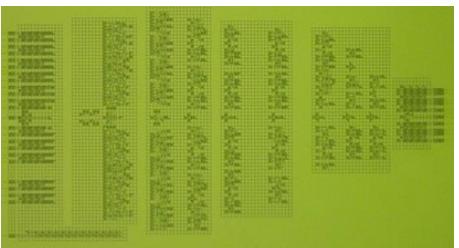
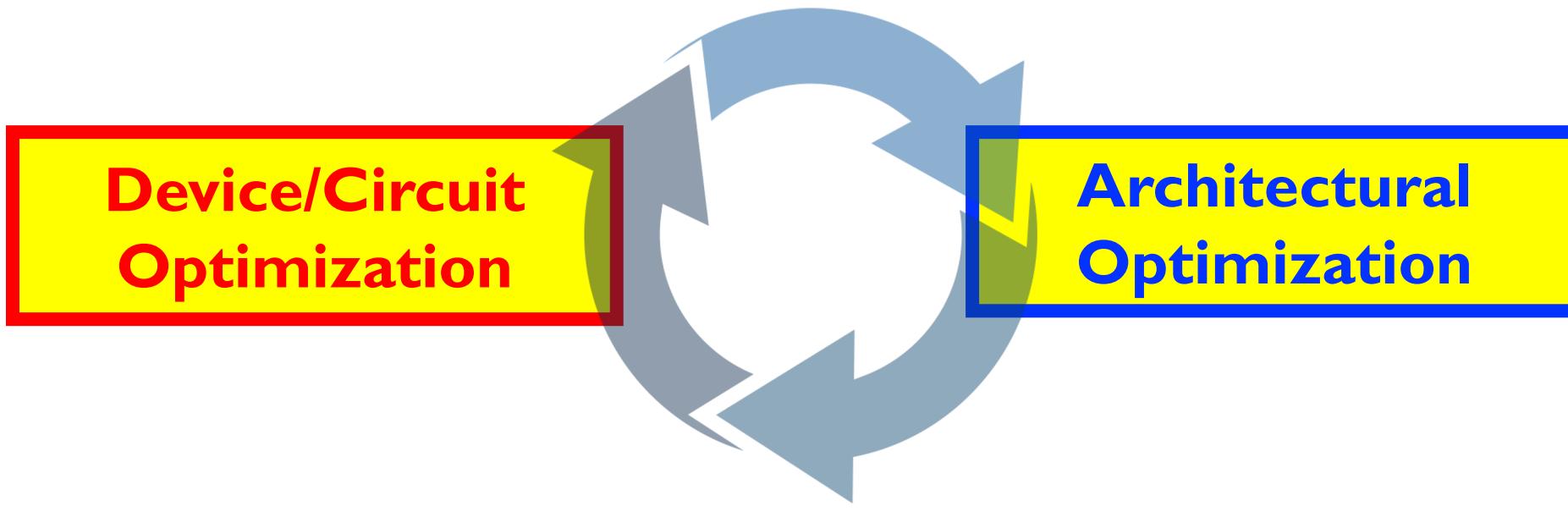


1.6 mW, 56 GHz 8-bit ALU
~35 TOPS/W

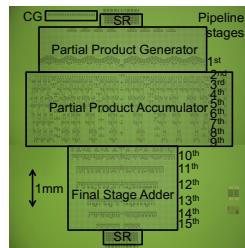
→ Next design achieved
112 TOPS/W



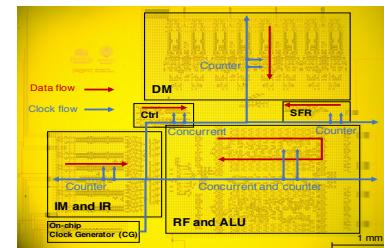
Our Approach & Outcome



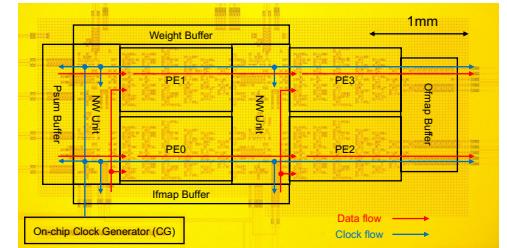
56GHz 1.6mW ALU
ISLPED'17 Design Contest
Honorable Mention



48GHz 5.6mW Multiplier
ISSCC'19
SilkRoad Award



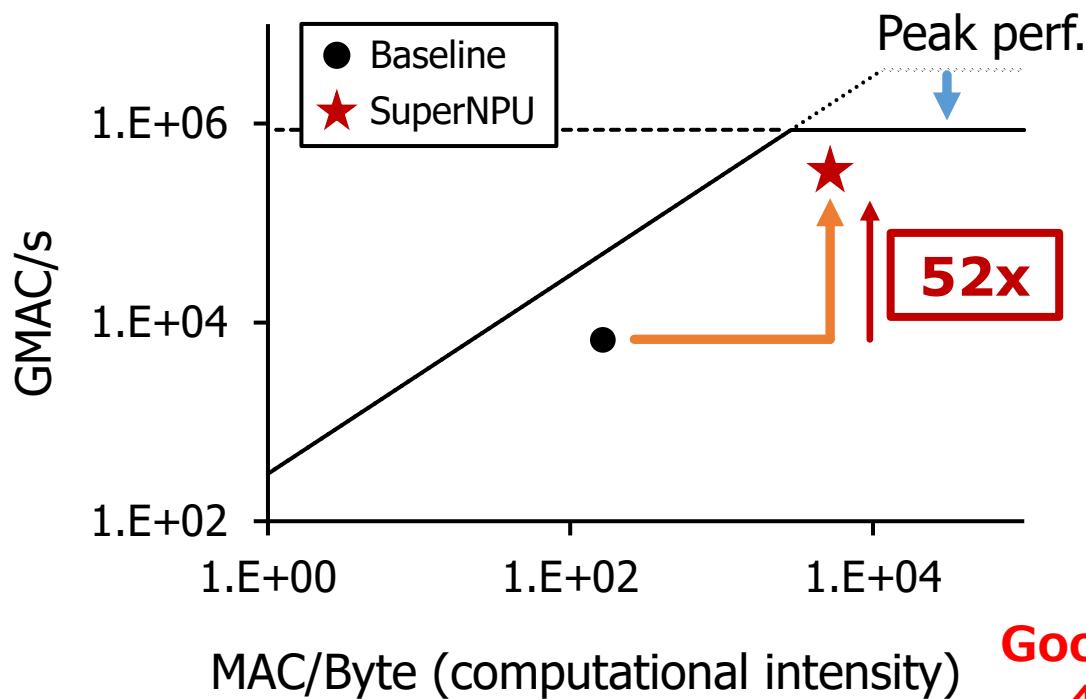
32GHz 6.2mW Processor
VLSI Symposium'20
Selected as a featured paper



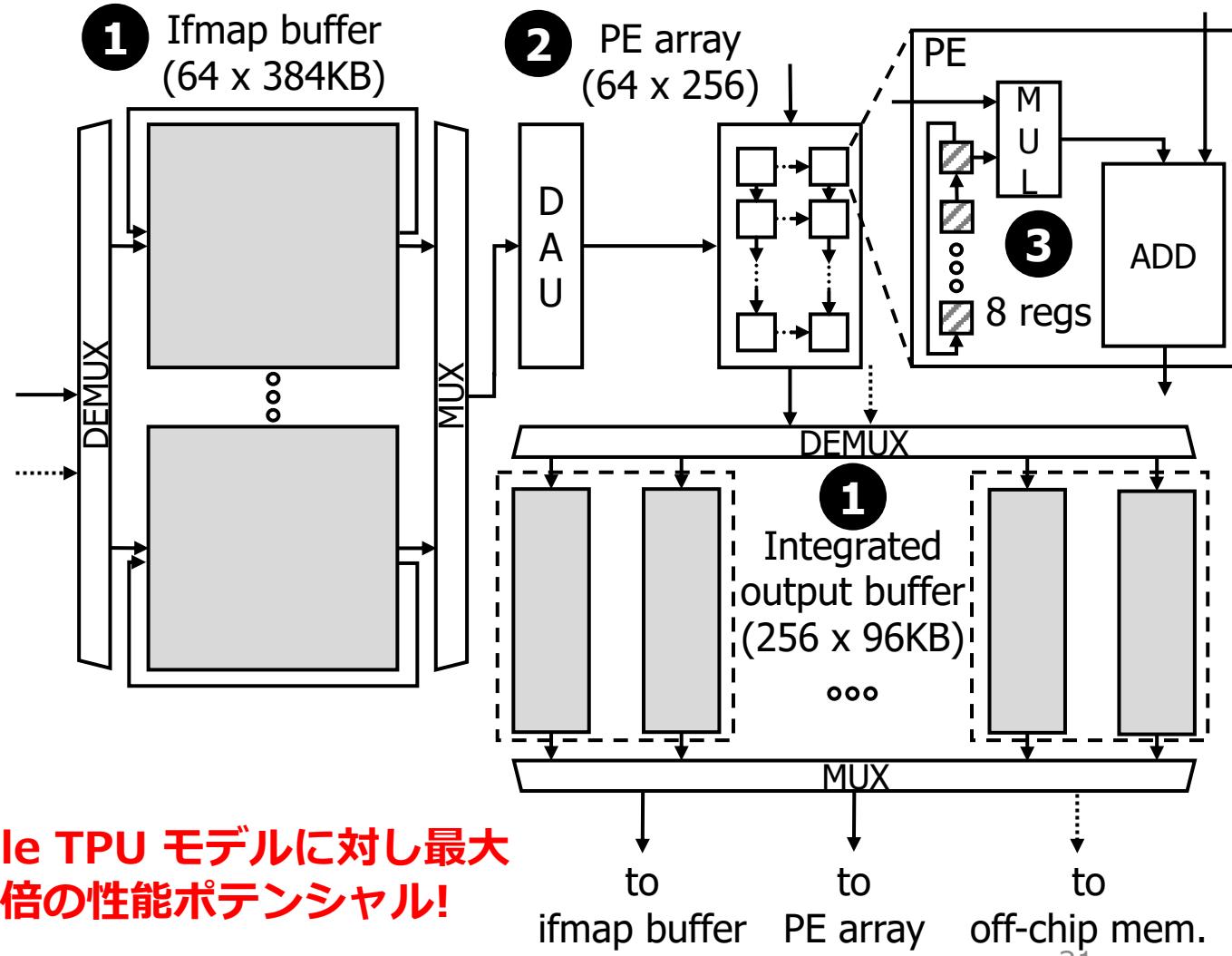
50GHz AI Accelerator
MICRO'20

超伝導AIコンピューティング：SuperNPU

- ① バッファ分割
- ② HW資源リバランス
- ③ 重みレジスタ数増加



Google TPU モデルに対し最大
40倍の性能ポテンシャル!



2020年世界TOP12のアーキテクチャ研究に選出（日本初）



How did the road to new adventure in Superconductor Neural Network Accelerator happen?

1: What is Superconductor SFQ?

Moore's Law, doubling the number of transistors every two years, has driven the evolution of computer systems. Unfortunately, it is reaching its limits. This is marking the beginning of the so-called post-Moore's era, where new emerging devices, and superconductor single-flux-quantum (SFQ) logic, are promising candidates. A superconductor ring can be used to compose a superconductor SFQ logic device by adding Josephson junctions (JJ) to the ring to control the magnetic flux quantum. It fundamentally changes the way logic is composed, making it possible to achieve extremely low-latency computation.

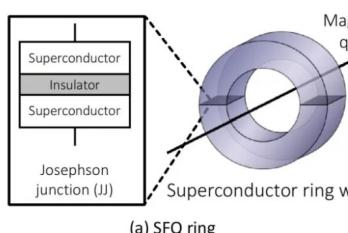


Figure 1: Superconductor SFQ ring

2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)

SuperNPU: An Extremely Fast Neural Processing Unit Using Superconducting Logic Devices

Koki Ishida^{1*}, Ilkwon Byun^{2*}, Ikki Nagaoka³, Kosuke Fukumitsu¹, Masamitsu Tanaka³, Satoshi Kawakami¹, Teruo Tanimoto¹, Takatsugu Ono¹, Jangwoo Kim², and Koji Inoue^{1†}

¹Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University
{koki.ishida, kosuke.fukumitsu, satoshi.kawakami, teruo.tanimoto, takatsugu.ono, koji.inoue}@cpc.ait.kyushu-u.ac.jp

²Department of Electrical and Computer Engineering, Seoul National University
{ik.byun, jangwoo}@snu.ac.kr

³Department of Electronics, Nagoya University
nagaoka@super.nuee.nagoya-u.ac.jp, masami_t@nagoya-u.jp

Abstract—Superconductor single-flux-quantum (SFQ) logic family has been recognized as a highly promising solution for the post-Moore's era, thanks to its ultra-fast and low-power switching

Among several candidates, superconductor SFQ logic family [3], [4] is a highly promising solution thanks to its ultra-fast speed and low power consumption at 4 K. The SEO tech-

IEEE
micro
The magazine for chip and silicon systems designers



6 GUEST EDITOR'S INTRODUCTION

Top Picks From the 2020 Computer Architecture Conferences

Daniel A. Jiménez

10

les

Behind MLPerf: Driving AI Inference Performance

Maia Reddi, Christine Cheng, David Kanter, Peter Mattson, Michael Muller, and Carole-Jean Wu

19

les

Superconductor Computing for Neural Networks

Koki Ishida, Ilkwon Byun, Ikki Nagaoka, Kosuke Fukumitsu, Masamitsu Tanaka, Satoshi Kawakami, Teruo Tanimoto, Takatsugu Ono, Jangwoo Kim, and Koji Inoue

27

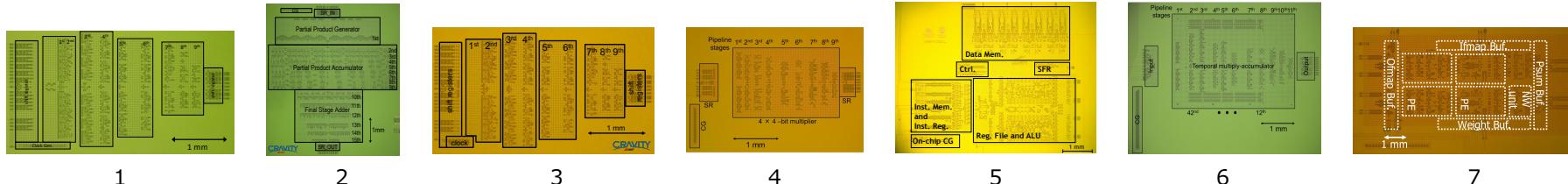
les

Leaking Secrets Through Compressed Caches

Po-An Tsai, Andres Sanchez, Christopher W. Fletcher, and Daniel Sanchez

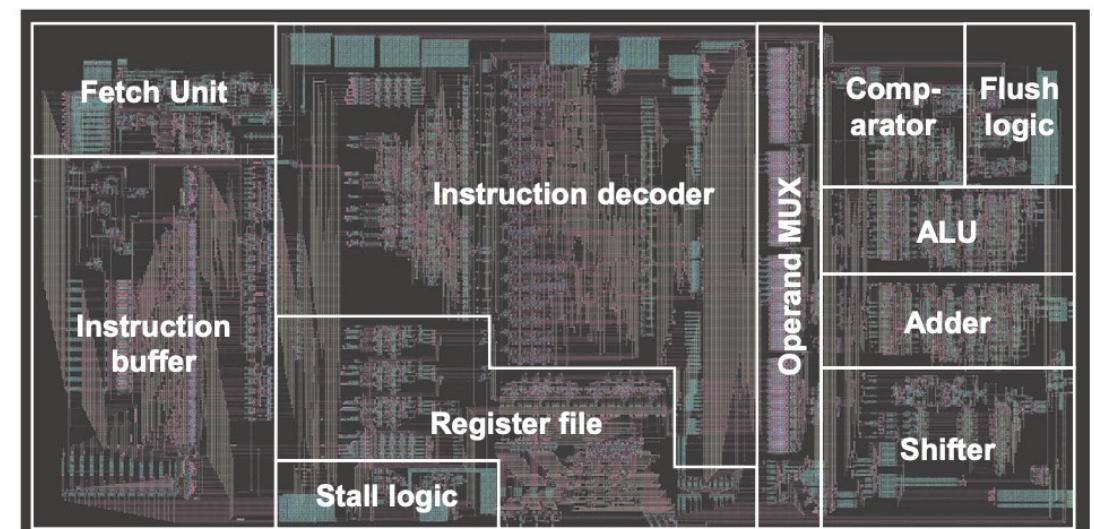
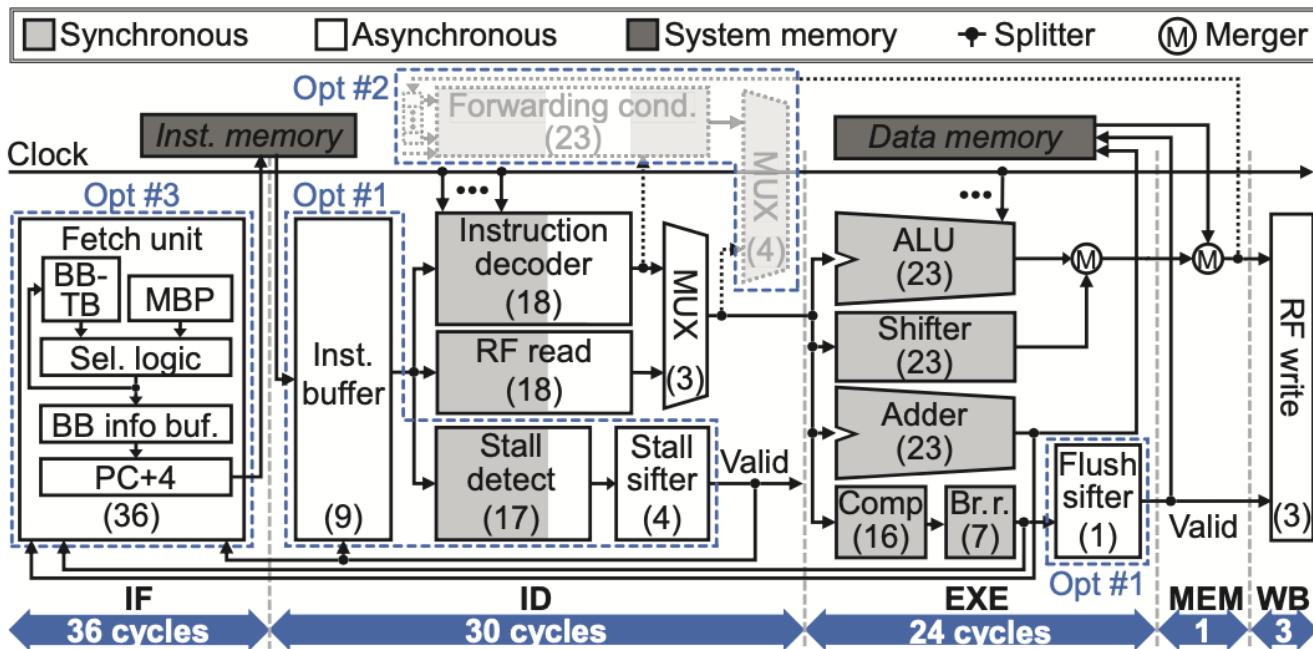
世界で最初に実証する!

Fabricated Chip	Purpose	Frequency [GHz]	Power [mW]	Efficiency [TOPS/W]	#of JJs	Year
1: 8-bit ALU	First demo. of gate-level pipeline	56	1.6	40	4,846	2017
2: 8-bit array-type multiplier	large-scale circuit design	48	5.6	8.5	20,251	2018
3: low voltage 8-bit ALU	0.5mV low-voltage operation	30	0.276	109	7,451	2019
4: low-voltage 4-bit multiplier	large-scale low-voltage operation	51	0.134	381	4,498	2019
5: 4-bit microprocessor	large-scale datapath	32	6.5	2.5	25,403	2019
6: low-voltage 4-bit MAC	basic function for AI acceleration	38	0.366	104	9,739	2020
7: 2x2 systolic PE array	prototype of <i>SuperNPU</i>	34	0.711	382	9,263	2021



State-of-the-art Designs

Junhyuk Choi et al., “SuperCore: An Ultra-Fast Superconducting Processor For Cryogenic Applications,” MICRO 2024. (to appear)



発表内容

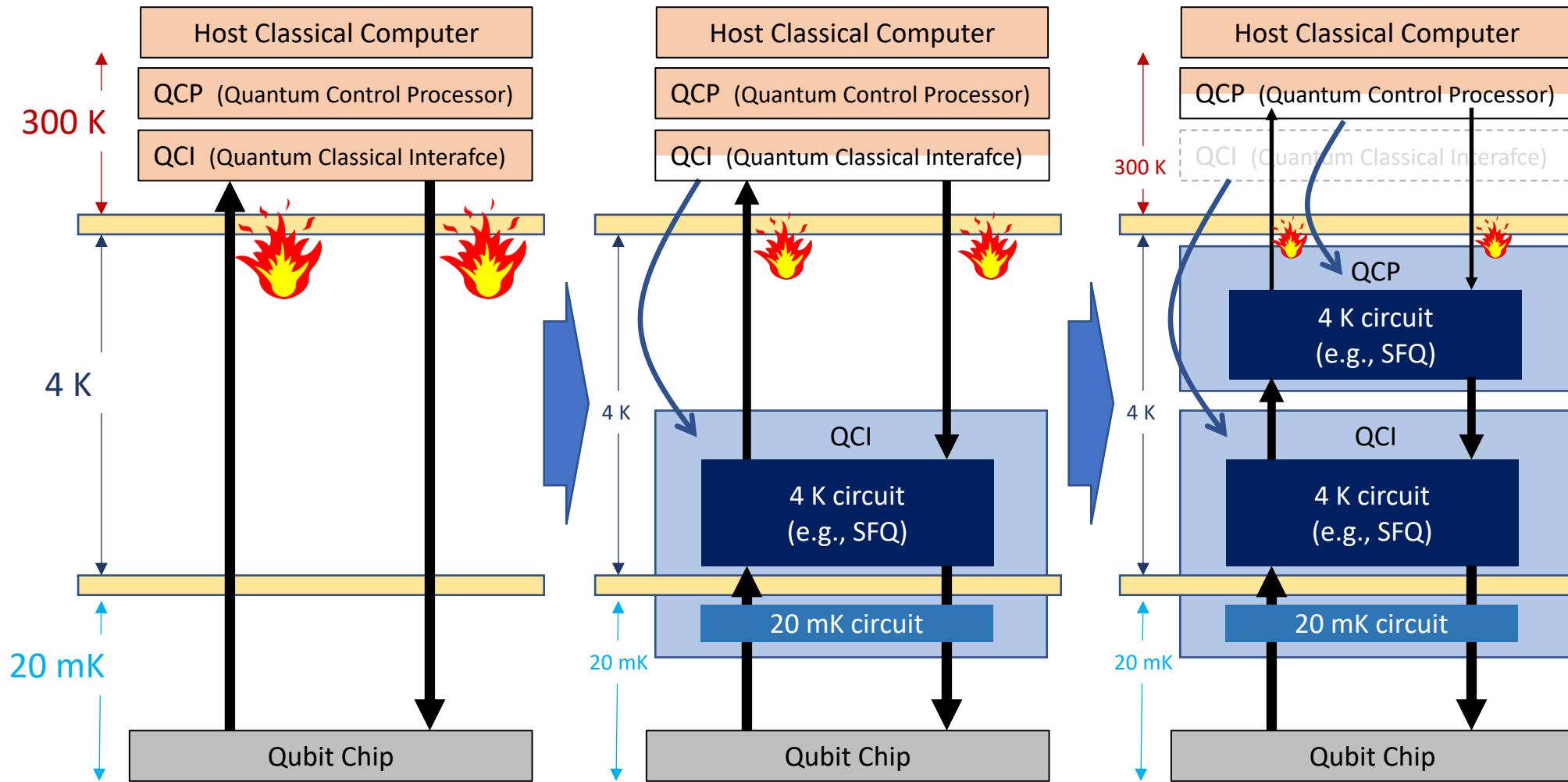
コンピューティング
技術の現状

超伝導古典
コンピューティング

超伝導量子
コンピューティング

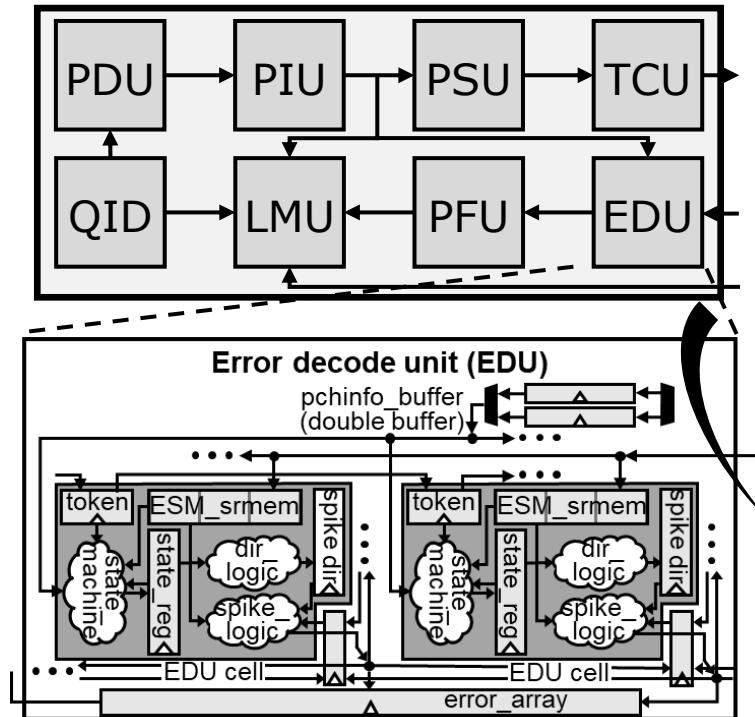
まとめ

なぜ超伝導量子コンピュータでSFQなのか？



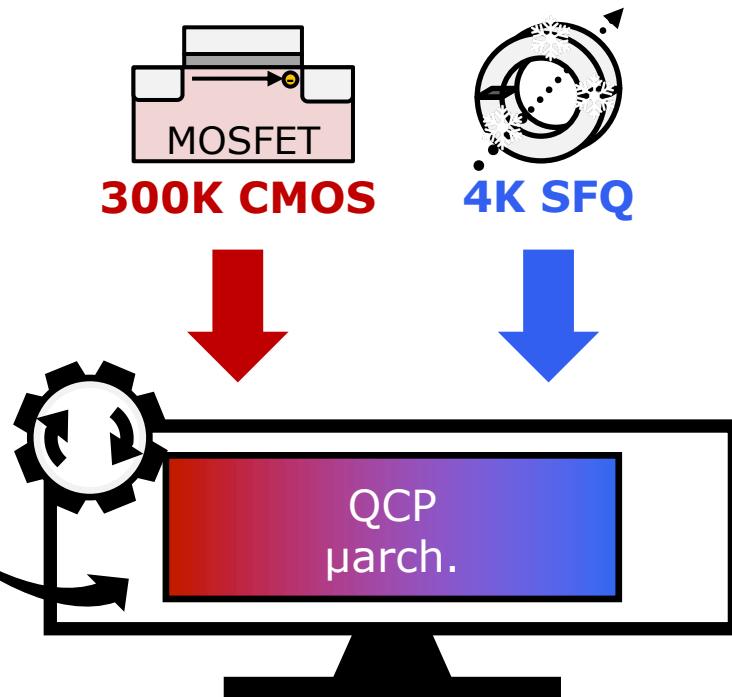
XQsim: Research Overview

Full QCP μarchitecture



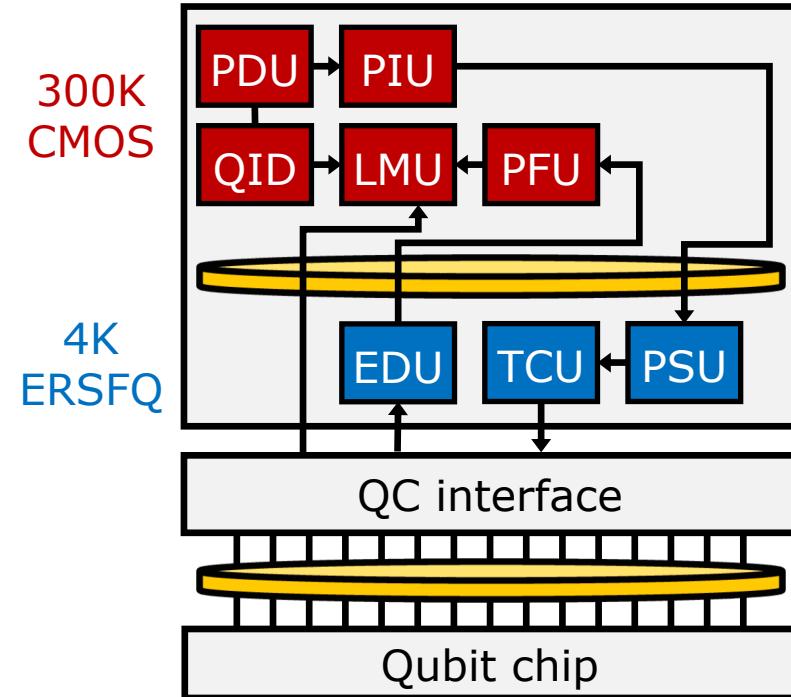
Detailed RTL
impl. & validation

QCP modeling tool



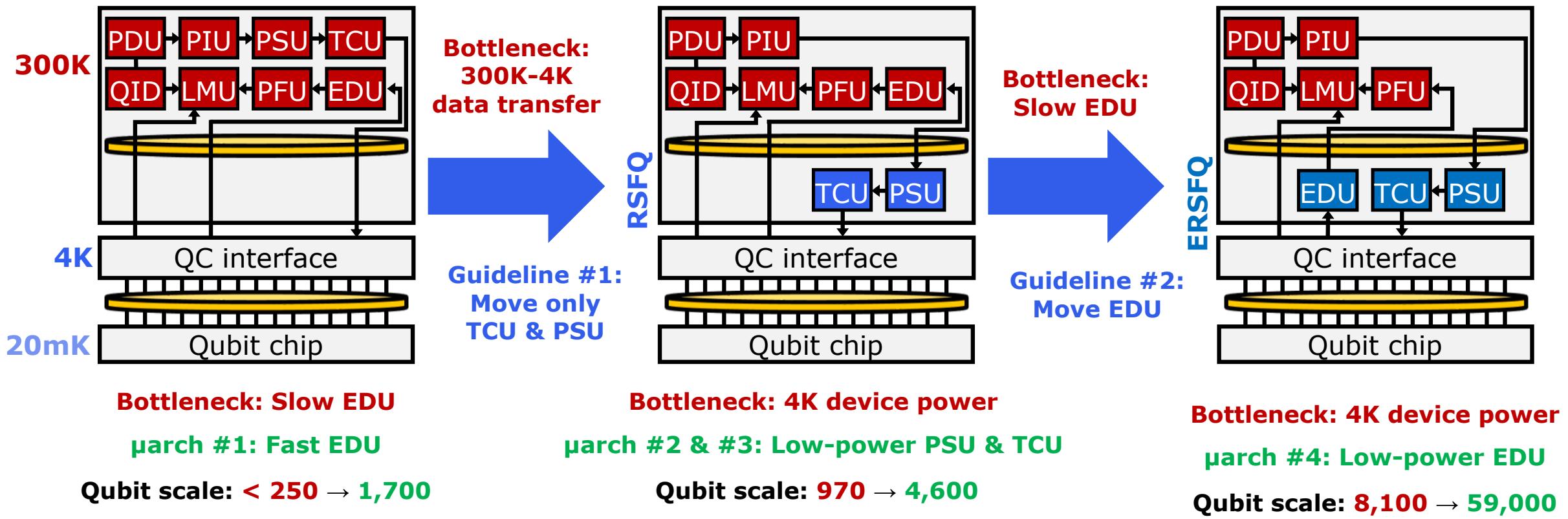
Cross-technology
modeling & simulation

10+K qubit QCP arch.



Temp. & Tech. & Arch.
optimizations²⁷

SFQ回路による量子誤り耐性により量子ビットのスケーラビリティを向上できる!



量子ビット数の増加にはSFQ技術の導入が有効であることを定量的に評価

発表内容

コンピューティング
技術の現状

超伝導古典
コンピューティング

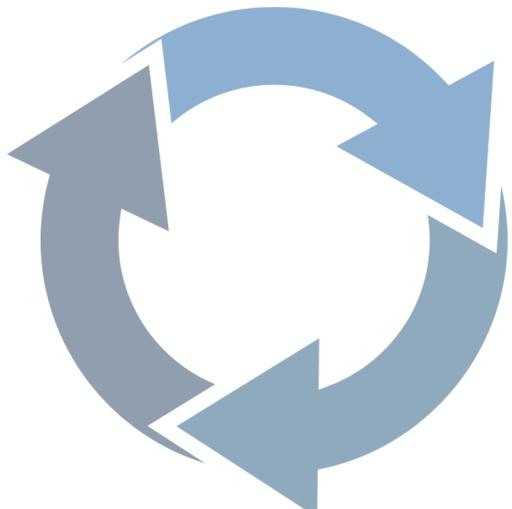
超伝導量子
コンピューティング

まとめ

次世代コンピューティング

新規デバイスの利点/欠点を最大化/隠蔽するアーキテクチャ創成

デバイス設計での各種制約からの開放
→アグレッシブな新奇デバイス創成へ！

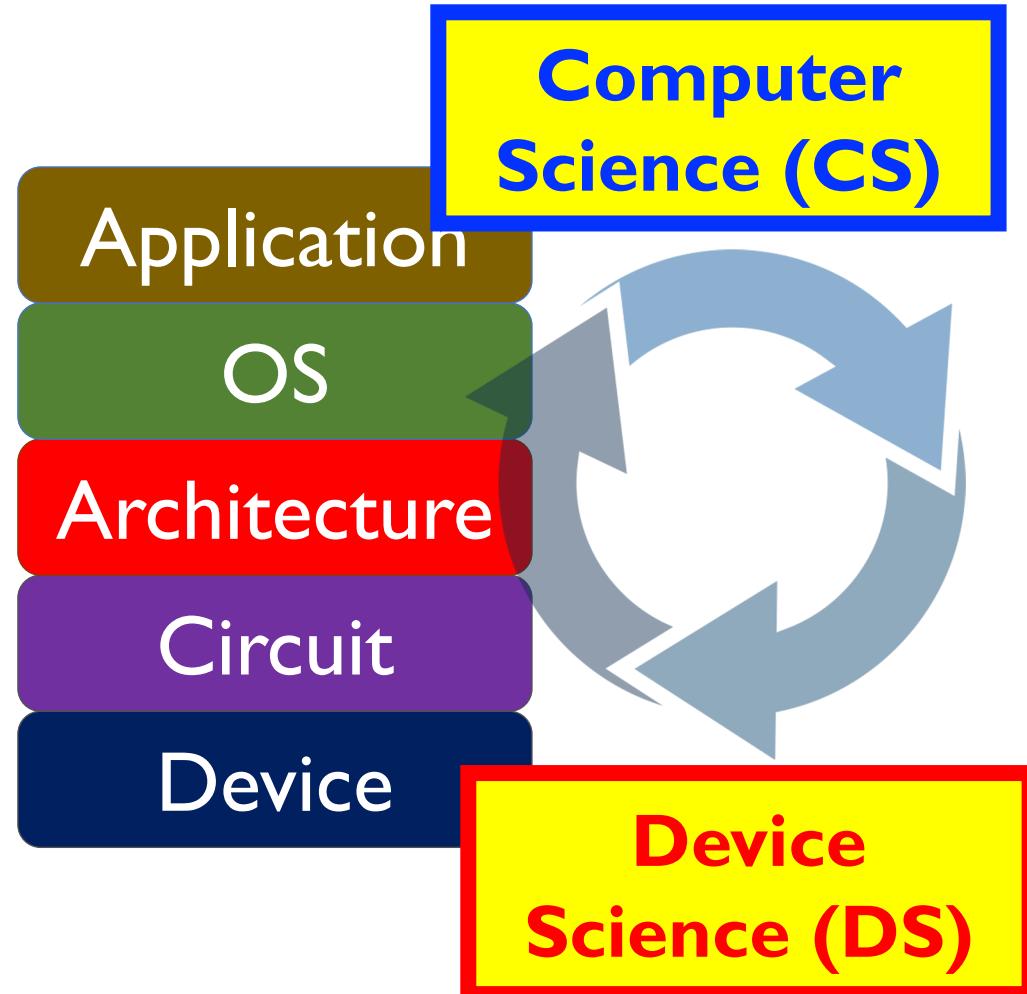


コンピューティングのための新規デバイス創成

ピークなデバイス特性による新たなトレードオフ
→新アーキテクチャ創成へ！

CS×DS

これからの研究



「コンピューティング力は国力」である!

- ・社会情報基盤として「コンピューティング力」を維持・強化し続けることは極めて重要
- ・コンピューティング技術の開拓は継続が必要
 - ・社会が求めるアプリケーションが変われば、必要となるコンピューティング力が変わる
 - ・新たな材料・デバイス技術が発見されれば、適切なコンピューティング方式が変わる
 - ・求められる制約が変われば、コンピュータシステムのあるべき姿が変わる
 - ・このような変化を世界にさきがけて察知・予測し、最先端コンピューティング技術を開拓し続けなければならない
- ・そのためには、デバイス技術だけではなく、設計開発技術（ツール）、ソフトウェア技術、応用展開技術、を含めた「CS×DSサイクル」のコンセプトに基づきコンピュータを発展させ続けなければならない

Acknowledgments

This work was supported by JST-Mirai Program Grant Number JPMJMI18E1, JSPS KAKENHI Grant Numbers JP19H01105, JP18H05211, JP18J21274, JP22H05000, JST Moonshot R&D Grant Number JPMJMS2067. The circuit is designed with the support by VDEC of the University of Tokyo in collaboration with Cadence Design Systems, Inc., and fabricated in the CRAVITY of AIST. We also appreciate the support from National Research Foundation of Korea (NRF) grant funded by the Korean Government (NRF-2019R1A5A1027055, NRF-2020M3H6A1084857).