

SS研 科学技術計算分科会
アクセラレータ技術の現状と今後
～HPCとアクセラレータ～

2008年10月22日

村上和彰

murakami@i.kyushu-u.ac.jp

国立大学法人九州大学 教授

SS研 会長

概要

高性能科学技術計算(HPC)とアクセラレータとの関係は歴史が長い。ベクトル処理もアクセラレータの一種であり、かつ、その元祖的存在である。ベクトル処理が時間軸方向のデータレベル並列処理だったものを空間軸方向に置き換えたものが現在主流となっているSIMD処理であり、CELLやClearSpeedがこれに該当する。

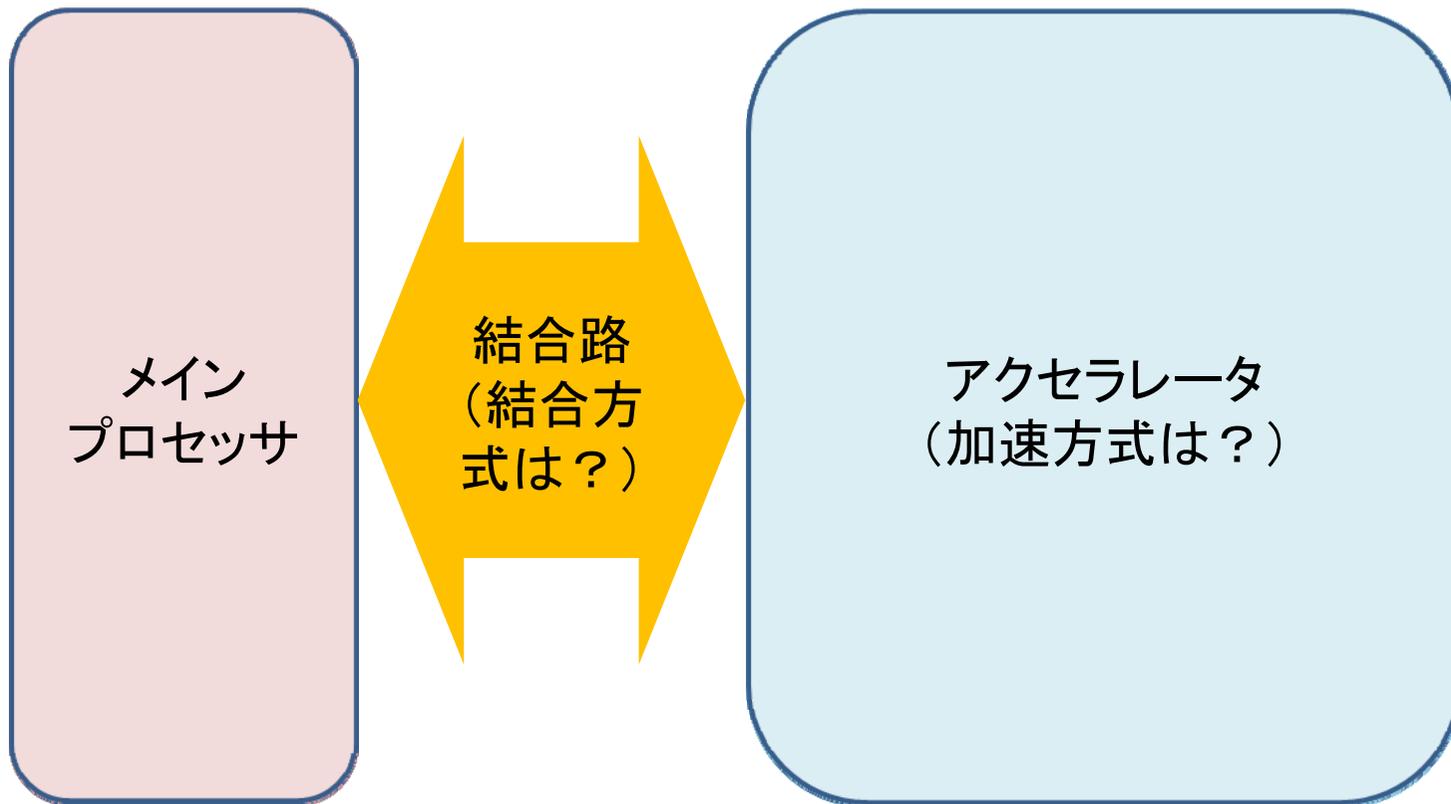
一方、信号処理の世界もアクセラレータとの付き合いは古い。DSP然り、最近ではDAPDNAのように数百個の演算器を2次元配列に配置したものも登場している。

世の中のテクノロジードライバーがコンシューマーエレクトロニクスに移行している昨今、今後のHPC業界におけるアクセラレータの進む方向性について議論する。

アクセラレータ (accelerator) とは？

- Wikipedia
 - Hardware accelerator: An additional unit of hardware to perform some function faster than is possible in software running on the normal CPU.
- むらかみの定義
 - 主たるプロセッサにおける処理の一部を当該プロセッサに代わって実行することで、当該プロセッサ単独実行に比べて全体性能を向上(=加速)させることを可能とする(主たるプロセッサとは異なるアーキテクチャの)プロセッサまたはハードウェア

アクセラレータとは？



アクセラレータ & 採用システム実例一覧

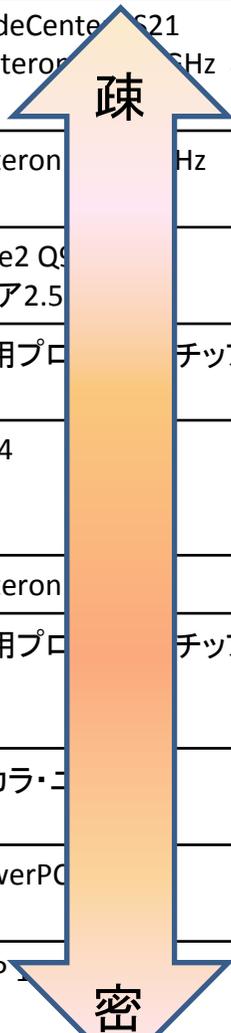


採用システム	メインプロセッサ	結合方式	アクセラレータ	加速方式
IBM Roadrunner	BladeCenter LS21 (Opteron DC 1.8GHz × 2)	ボード間: デュアル・ギガビット・イーサネット	BladeCenter QS22 (PowerXCell 8i 3.2GHz × 2)	4-way SIMD × 8-way マルチコア
東工大TSUBAME	Opteron DC 2.4GHz	ボード間: PCI-X	ClearSpeed Advance CSX600 PCI-X Board	96-way SIMD
DELL Precision T7400	Core2 Q9300 4コア2.5GHz	チップ間: PCI-Express	Nvidia GeForce 9800 1.4GHz	8-way SIMD × 16-way マルチコア
GRAPE	汎用プロセッサ・チップ	チップ間: 共有バス	重力計算専用プロセッサ GRAPE	並列演算パイプライン
九州大学EHPC/Eric	SH-4	チップ間: SH-4バス	二電子積分計算専用プロセッサEric	初期積分計算専用コア × 1 + 漸化計算専用コア × 4
CRAY XD1	Opteron	チップ間:	FPGA	ハードウェア処理
CREST SFQ-RDP	汎用プロセッサ・チップ	チップ間: 共有バス	九州大学LSRDP(再構成可能大規模データパス80GHz) MCM × 4	2次元FPUアレイ (1024FPU/MCM)
NEC SX-9	スカラ・ユニット	チップ内:	ベクトル・パイプライン × 8	並列ベクトル処理
IBM CELL/B.E.	PowerPC 4GHz	チップ内: 共有リングバス (EIB)	SPE 4GHz × 8	4-Way SIMD
IPFLEX DAPDNA-II	DAP 166MHz	チップ内: 共有バス	DNA 166MHz	2次元ALUアレイ (168ALU)
TI OMAP 3530	ARM Coretex-A8	チップ内: 共有バス	TMS320C64x DSP Core	VLIW DSP

アクセラレータ & 採用システム実例一覧



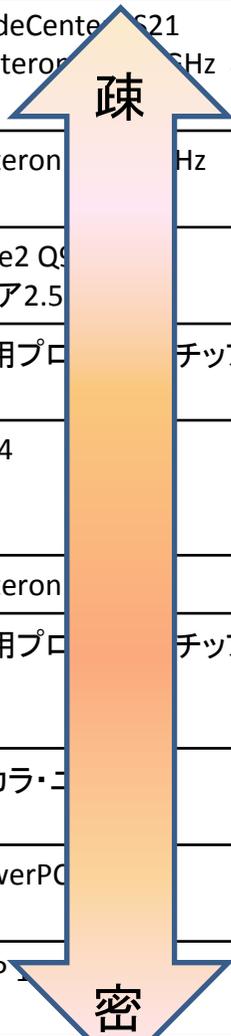
採用システム	メインプロセッサ	結合方式	アクセラレータ	加速方式
IBM Roadrunner	BladeCenter QS21 (Opteron 2.6GHz × 2)	ボード間: デュアル・ギガビット・イーサネット	BladeCenter QS22 (PowerXCell 8i 3.2GHz × 2)	4-way SIMD × 8-way マルチコア
東工大TSUBAME	Opteron 2.6GHz	ボード間: PCI-X	ClearSpeed Advance CSX600 PCI-X Board	96-way SIMD
DELL Precision T7400	Core2 Duo 4コア2.5GHz	チップ間: PCI-Express	Nvidia GeForce 9800 1.4GHz	8-way SIMD × 16-way マルチコア
GRAPE	汎用プロセッサ	チップ間: 共有バス	重力計算専用プロセッサ GRAPE	並列演算パイプライン
九州大学EHPC/Eric	SH-4	チップ間: SH-4バス	二電子積分計算専用プロセッサEric	初期積分計算専用コア × 1 + 漸化計算専用コア × 4
CRAY XD1	Opteron	チップ間:	FPGA	ハードウェア処理
CREST SFQ-RDP	汎用プロセッサ	チップ間: 共有バス	九州大学LSRDP(再構成可能大規模データパス 80GHz) MCM × 4	2次元FPUアレイ (1024FPU/MCM)
NEC SX-9	スカラー・コア	チップ内:	ベクトル・パイプライン × 8	並列ベクトル処理
IBM CELL/B.E.	PowerPC	チップ内: 共有リングバス (EIB)	SPE 4GHz × 8	4-Way SIMD
IPFLEX DAPDNA-II	DAP DNA	チップ内: 共有バス	DNA 166MHz	2次元ALUアレイ (168ALU)
TI OMAP 3530	ARM Coretex-A8	チップ内: 共有バス	TMS320C64x DSP Core	VLIW DSP



アクセラレータ & 採用システム実例一覧



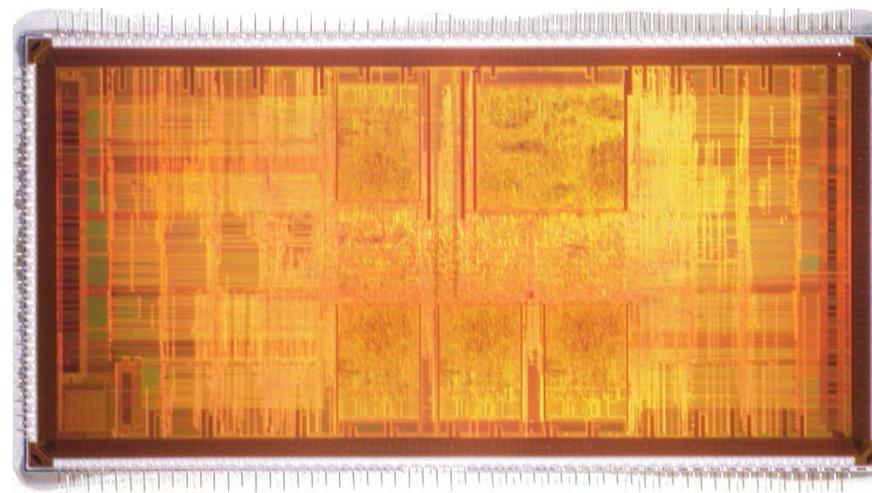
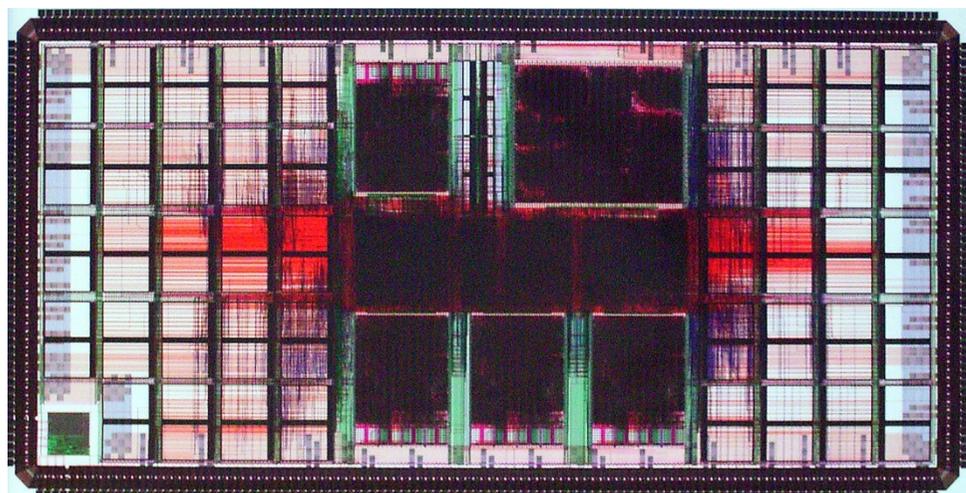
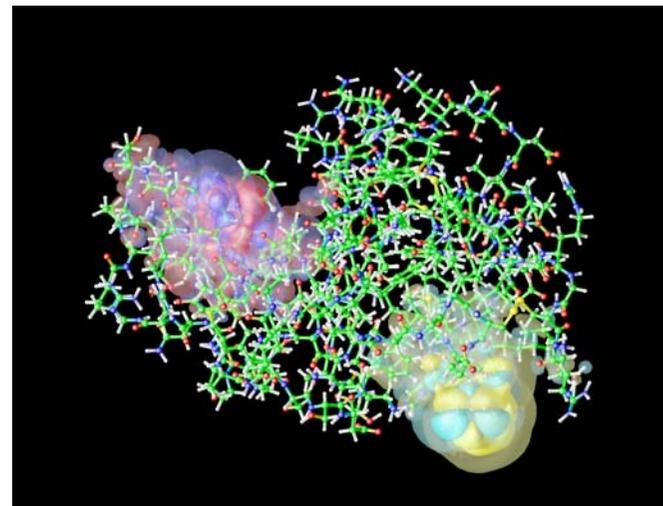
採用システム	メインプロセッサ	結合方式	アクセラレータ	加速方式
IBM Roadrunner	BladeCenter QS21 (Opteron 2.6GHz × 2)	ボード間: デュアル・ギガビット・イーサネット	BladeCenter QS22 (PowerXCell 8i 3.2GHz × 2)	4-way SIMD × 8-way マルチコア
東工大TSUBAME	Opteron 2.6GHz	ボード間: PCI-X	ClearSpeed Advance CSX600 PCI-X Board	96-way SIMD
DELL Precision T7400	Core2 Duo 4コア2.5GHz	チップ間: PCI-Express	Nvidia GeForce 9800 1.4GHz	8-way SIMD × 16-way マルチコア
GRAPE	汎用プロセッサ	チップ間: 共有バス	重力計算専用プロセッサ GRAPE	並列演算パイプライン
九州大学EHPC/Eric	SH-4	チップ間: SH-4バス	二電子積分計算専用プロセッサEric	初期積分計算専用コア × 1 + 漸化計算専用コア × 4
CRAY XD1	Opteron	チップ間:	FPGA	ハードウェア処理
CREST SFQ-RDP	汎用プロセッサ	チップ間: 共有バス	九州大学LSRDP(再構成可能大規模データパス 80GHz) MCM × 4	2次元FPUアレイ (1024FPU/MCM)
NEC SX-9	スカラー・コア	チップ内:	ベクトル・パイプライン × 8	並列ベクトル処理
IBM CELL/B.E.	PowerPC	チップ内: 共有リングバス (EIB)	SPE 4GHz × 8	4-Way SIMD
IPFLEX DAPDNA-II	DAP DNA	チップ内: 共有バス	DNA 166MHz	2次元ALUアレイ (168ALU)
TI OMAP 3530	ARM Coretex-A8	チップ内: 共有バス	TMS320C64x DSP Core	VLIW DSP



九州大学EHPC/Eric ～Ericチップ～

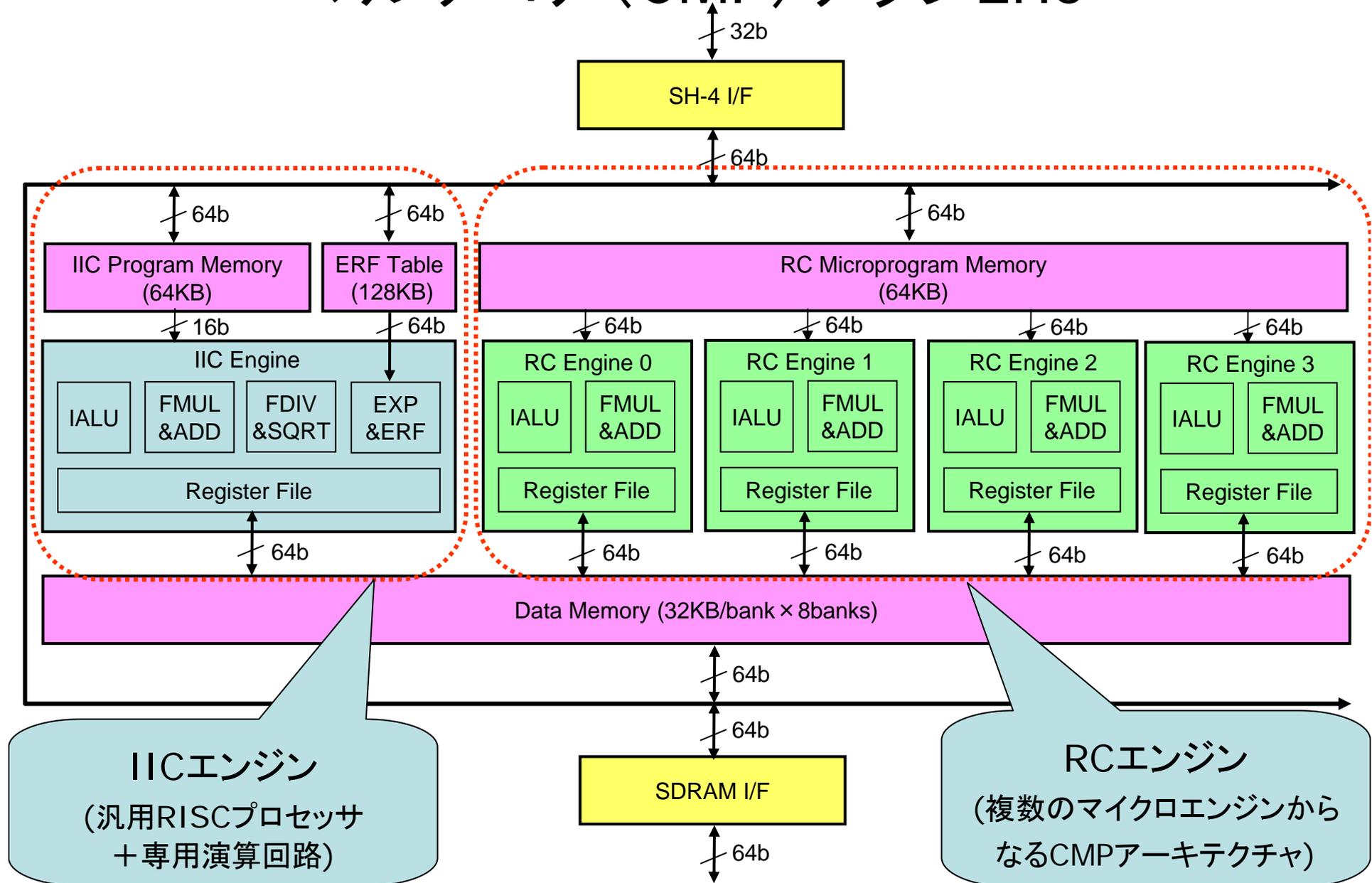
Eric: (世界初の)二電子積分計算専用LSI

- 仕様
 - プロセス: TSMC 0.13um, 6層, Cu配線
 - チップサイズ: 5 x 10mm²
 - 論理ゲート: 4M
 - メモリ: 704KB
 - パッケージ: セラミックPGA 257PIN, 50.8mm□
- 動作テスト結果
 - 動作周波数200MHzでテスト→歩留まり約80%
- 消費電力
 - 2.1W

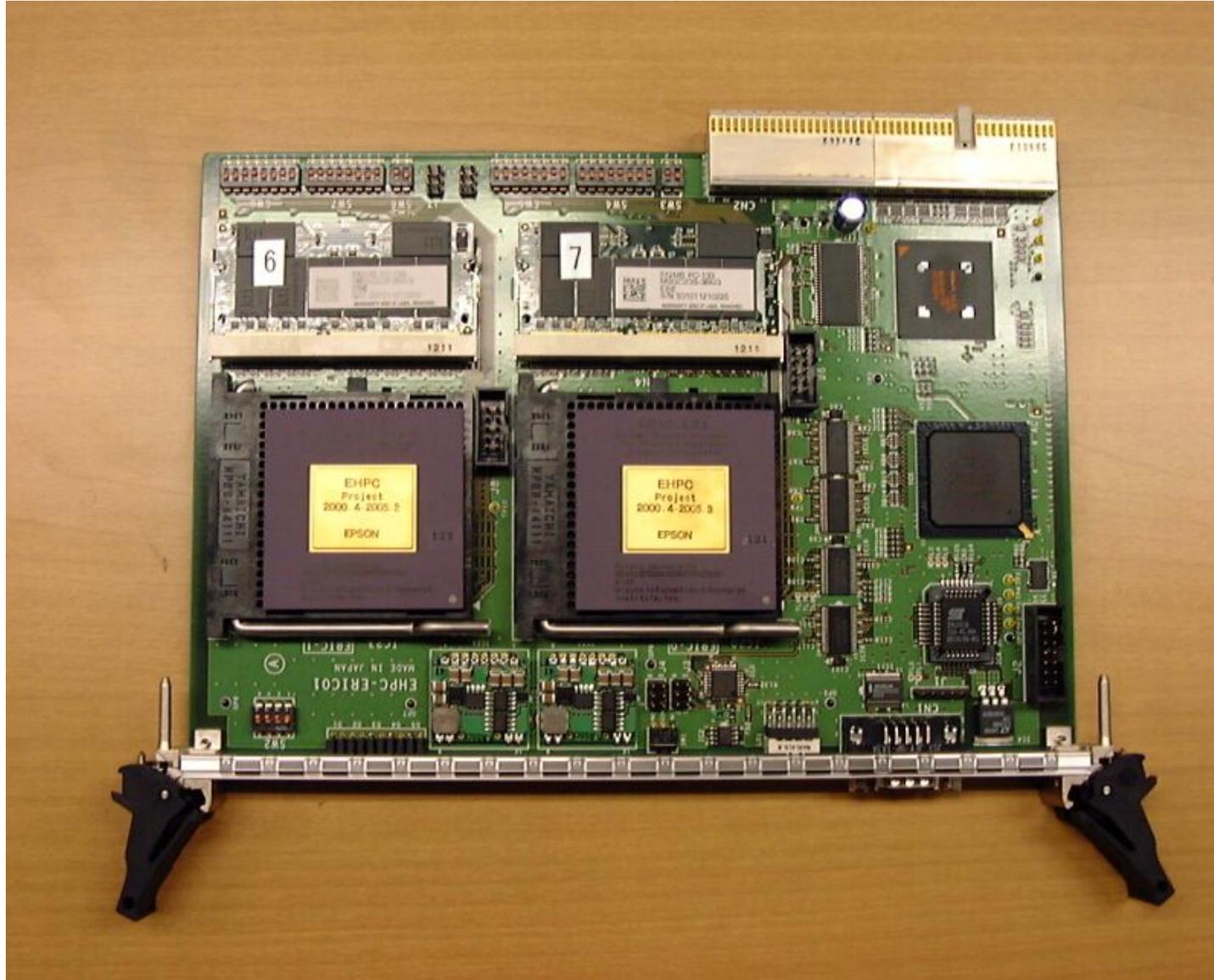


九州大学EHPC/Eric

～マルチコア(CMP)チップEric～



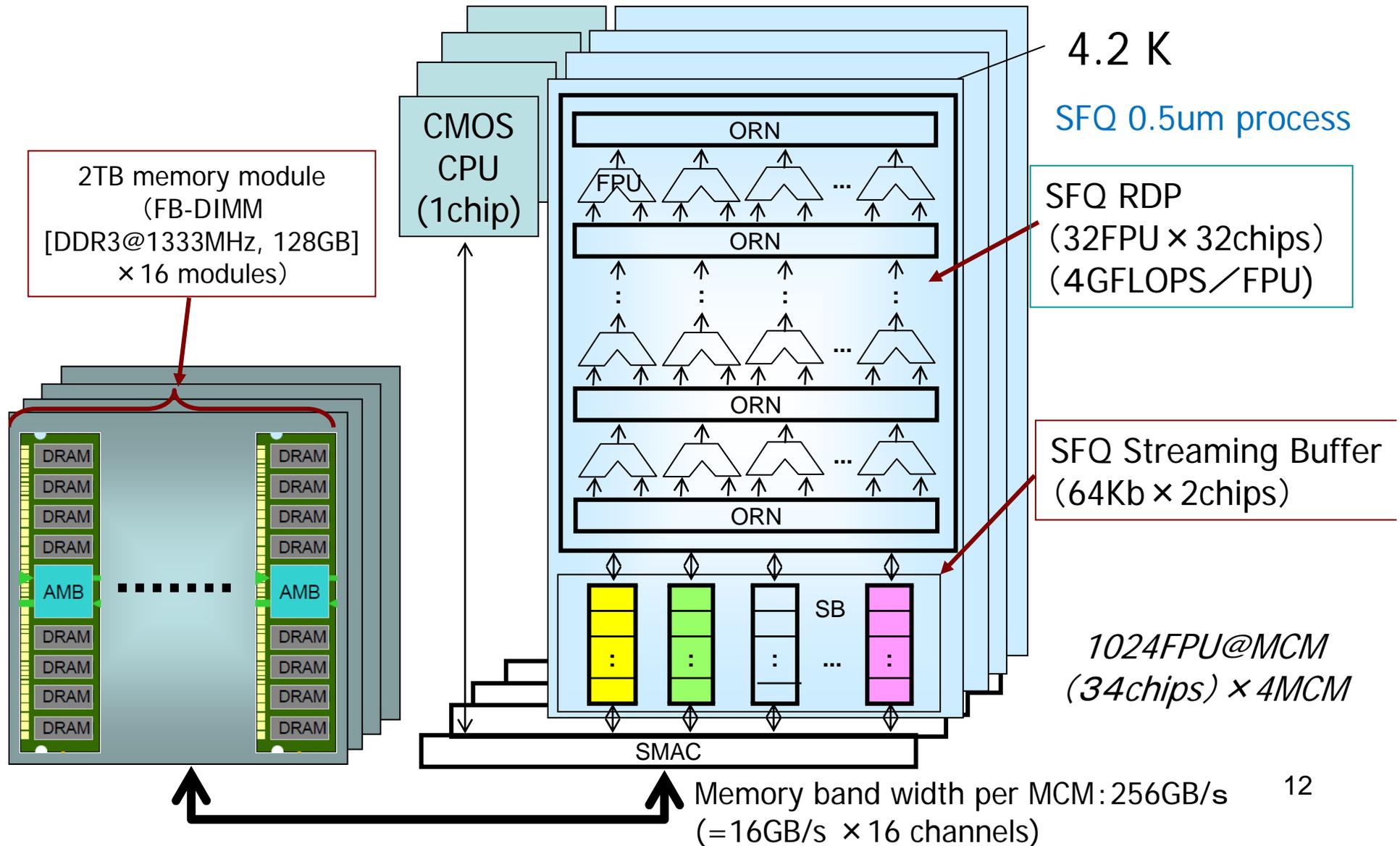
九州大学EHPC/Eric ～EHPCボード(SH-4+Eric × 2)～



九州大学EHPC/Eric ～EHPCシステム～

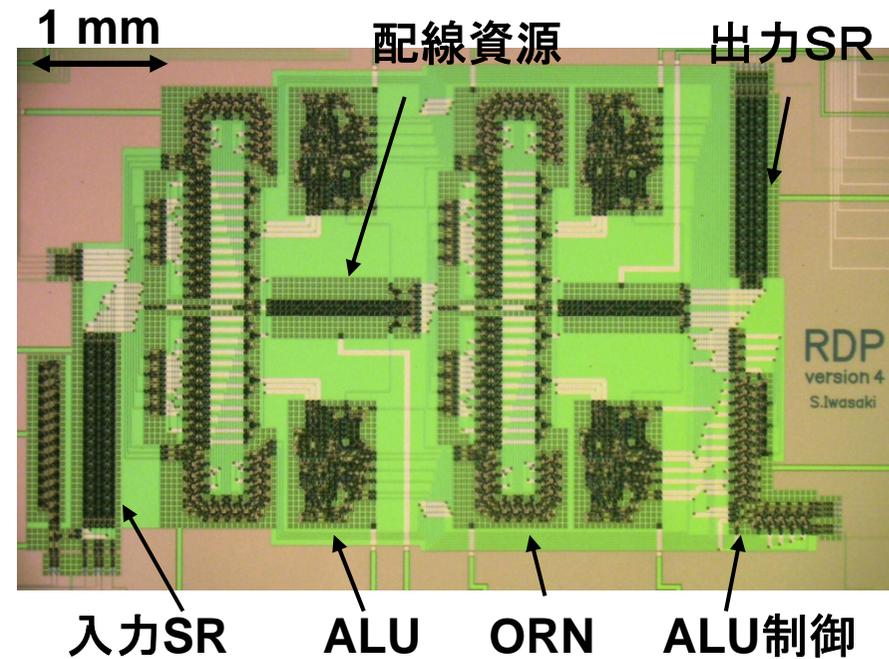
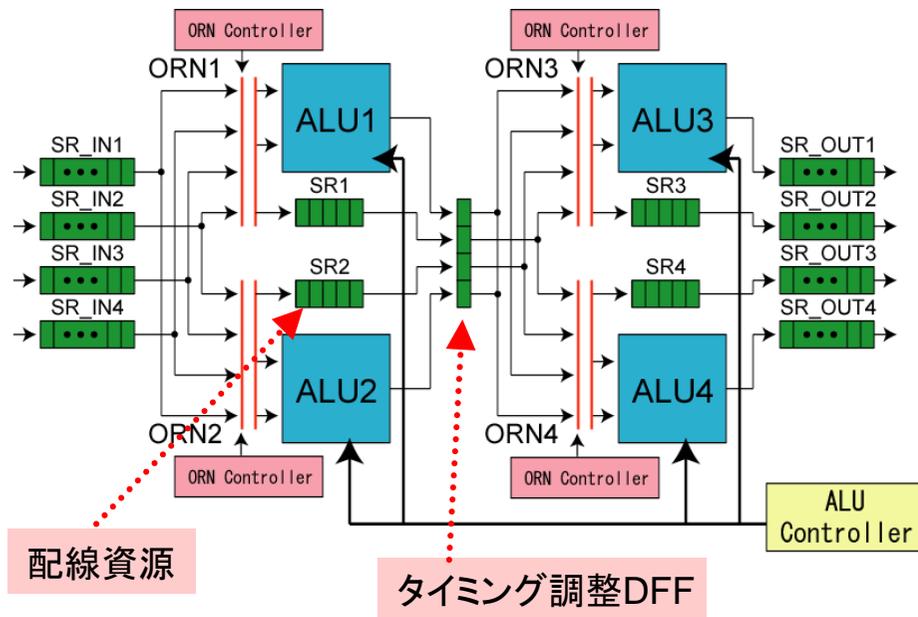


CREST SFQ-RDP



CREST SFQ-RDP

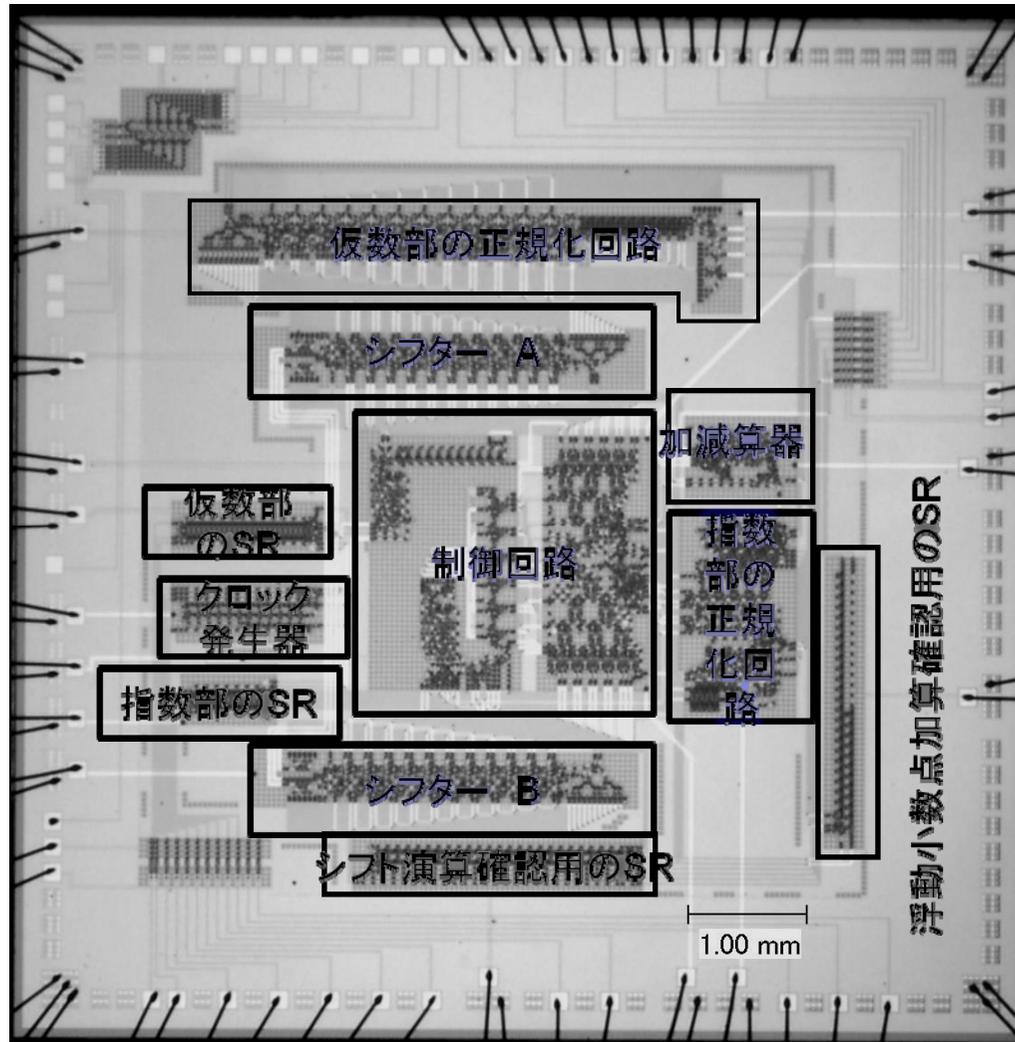
~2x2 SFQ-RDP~



設計周波数: 25 GHz、バイアス電流: 1.286 A
 回路面積: 5.90 x 3.68 mm²、接合数: 10839

CREST SFQ-RDP

～半精度浮動小数点加算器～

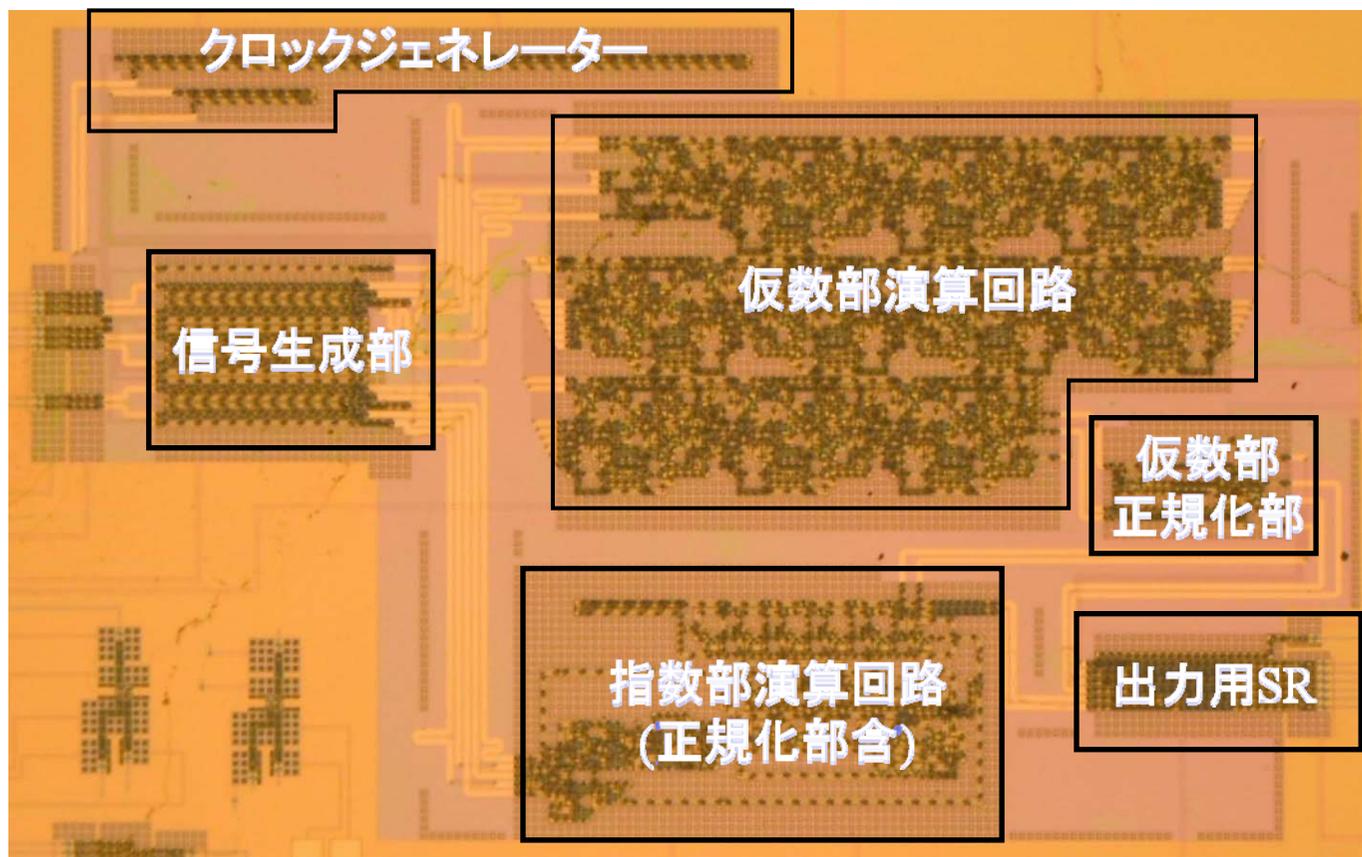


SRL Nb 2.5 kA/cm²
standard process

設計周波数 : 20 GHz
性能 : 1 GFLOPs
接合数 : 10404 JJs
消費電力 : 3.1 mW
回路面積 : 5.86 × 5.72 mm²

CREST SFQ-RDP

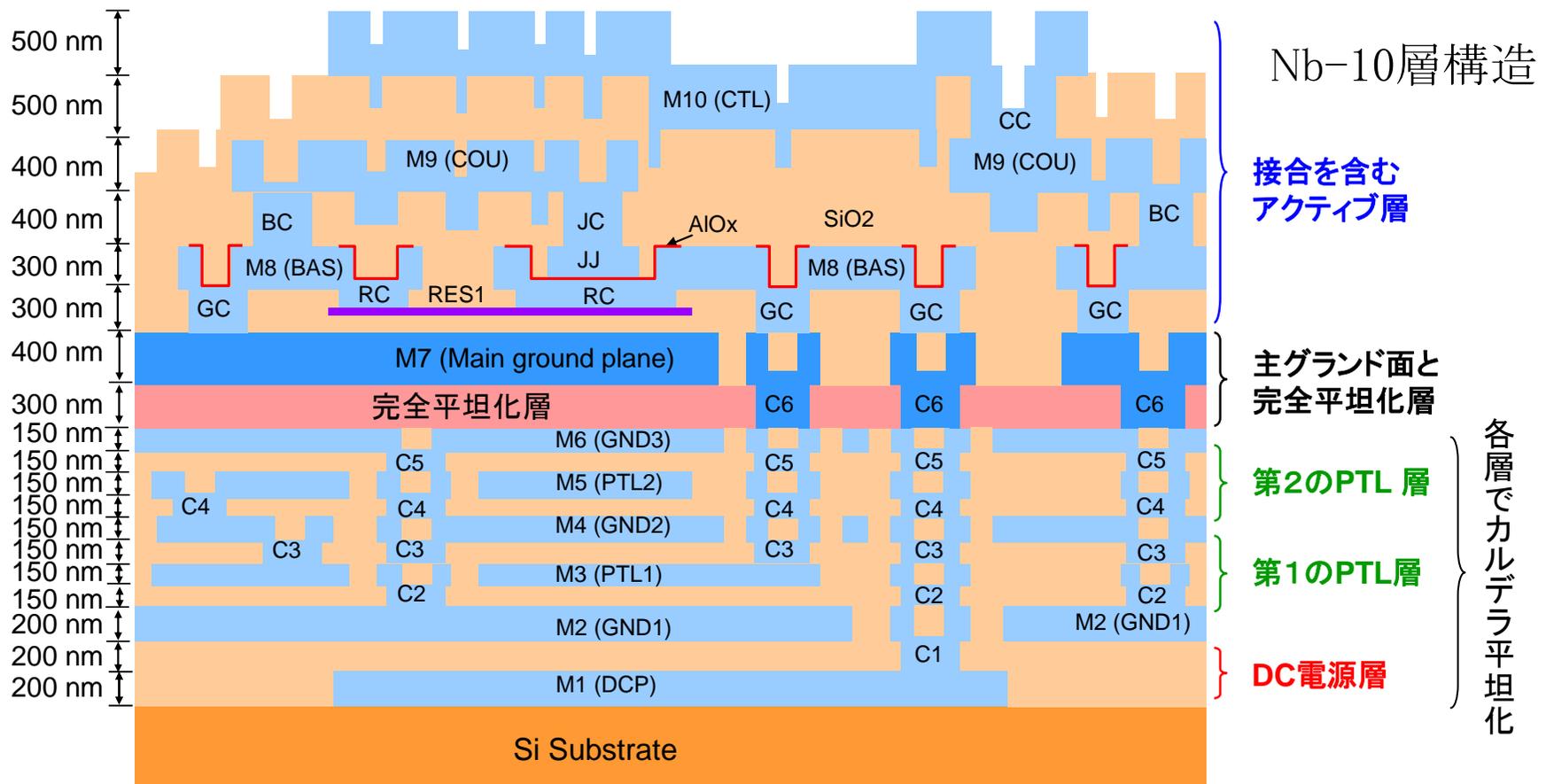
～半精度浮動小数点乗算器～



SRL Nb 2.5 kA/cm²
standard process

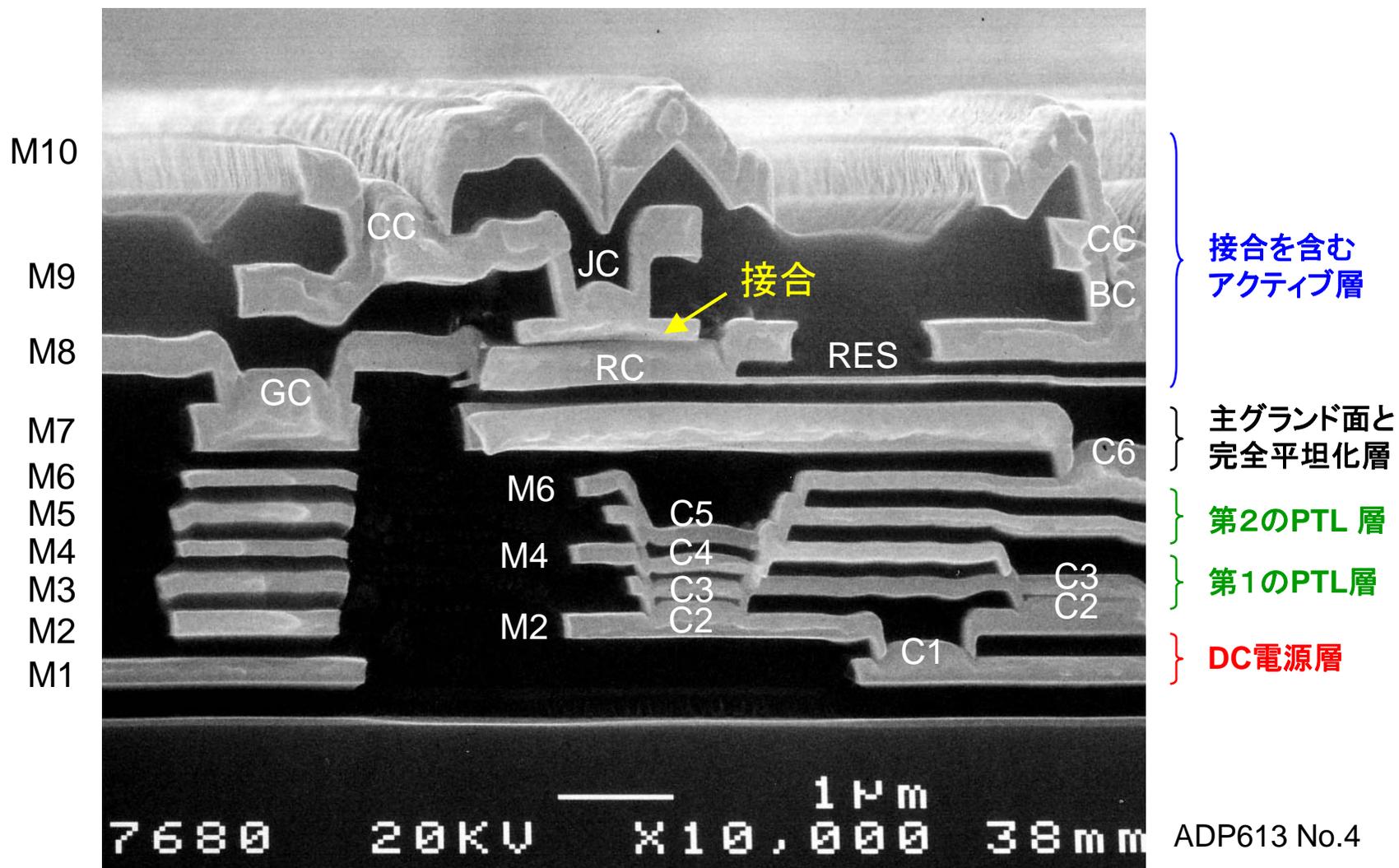
CREST SFQ-RDP

～1 μ m多層配線プロセス～



CREST SFQ-RDP

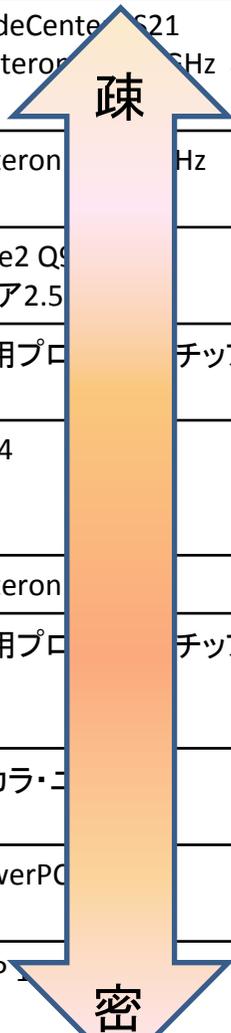
～新ニオブ10層構造の断面SEM写真～



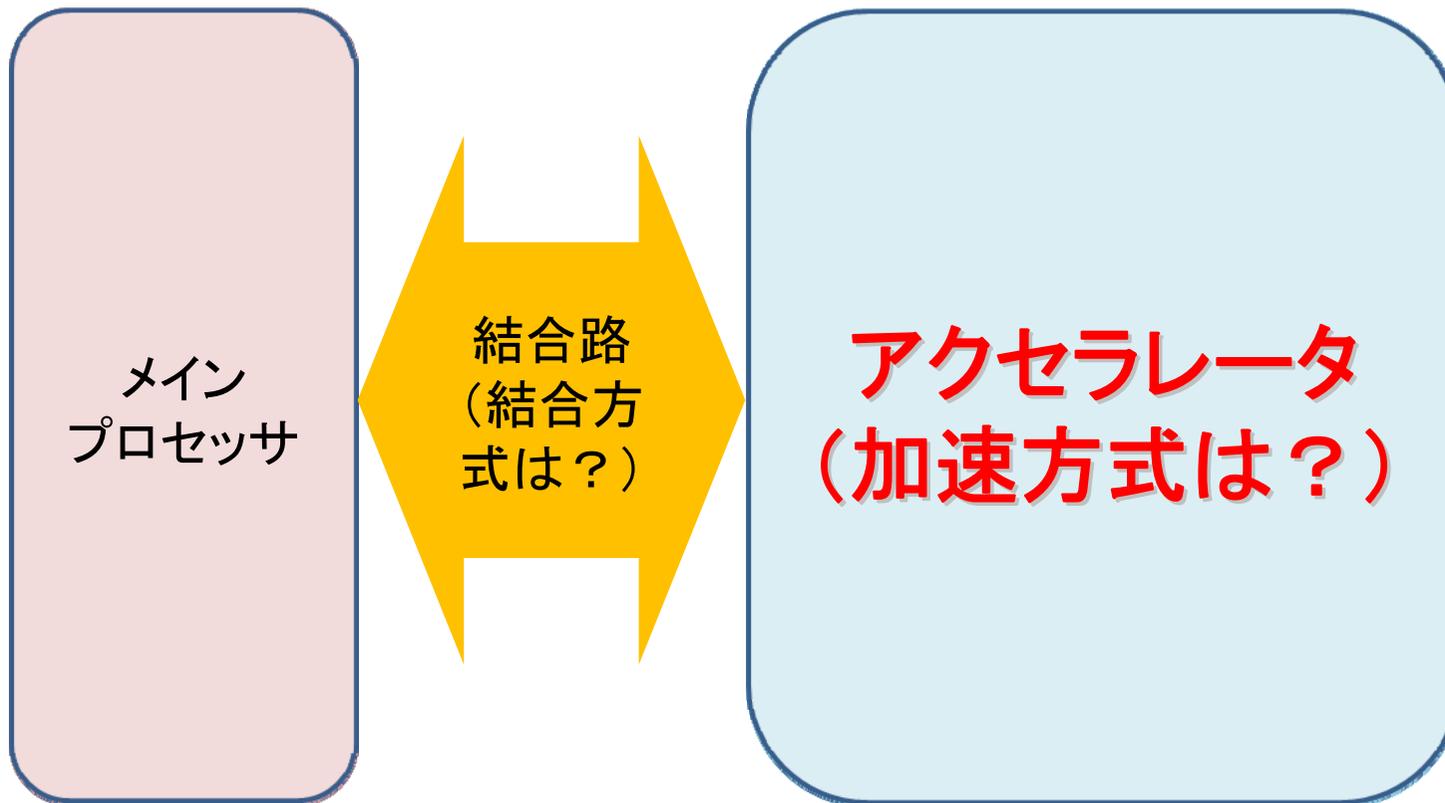
アクセラレータ & 採用システム実例一覧



採用システム	メインプロセッサ	結合方式	アクセラレータ	加速方式
IBM Roadrunner	BladeCenter QS21 (Opteron 2.6GHz × 2)	ボード間: デュアル・ギガビット・イーサネット	BladeCenter QS22 (PowerXCell 8i 3.2GHz × 2)	4-way SIMD × 8-way マルチコア
東工大TSUBAME	Opteron 2.6GHz	ボード間: PCI-X	ClearSpeed Advance CSX600 PCI-X Board	96-way SIMD
DELL Precision T7400	Core2 Duo T7400 4コア2.5GHz	チップ間: PCI-Express	Nvidia GeForce 9800 1.4GHz	8-way SIMD × 16-way マルチコア
GRAPE	汎用プロセッサ	チップ間: 共有バス	重力計算専用プロセッサ GRAPE	並列演算パイプライン
九州大学EHPC/Eric	SH-4	チップ間: SH-4バス	二電子積分計算専用プロセッサEric	初期積分計算専用コア × 1 + 漸化計算専用コア × 4
CRAY XD1	Opteron	チップ間:	FPGA	ハードウェア処理
CREST SFQ-RDP	汎用プロセッサ	チップ間: 共有バス	九州大学LSRDP(再構成可能大規模データパス 80GHz) MCM × 4	2次元FPUアレイ (1024FPU/MCM)
NEC SX-9	スカラー・コア	チップ内:	ベクトル・パイプライン × 8	並列ベクトル処理
IBM CELL/B.E.	PowerPC	チップ内: 共有リングバス (EIB)	SPE 4GHz × 8	4-Way SIMD
IPFLEX DAPDNA-II	DAPDNA-II	チップ内: 共有バス	DNA 166MHz	2次元ALUアレイ (168ALU)
TI OMAP 3530	ARM Coretex-A8	チップ内: 共有バス	TMS320C64x DSP Core	VLIW DSP



アクセラレータとは？



アクセラレータ & 採用システム実例一覧



採用システム	メインプロセッサ	結合方式	アクセラレータ	加速方式
IBM Roadrunner	BladeCenter LS21 (Opteron DC 1.8GHz × 2)	ボード間: デュアル・ギガビット・イーサネット	BladeCenter QS22 (PowerXCell 8i 3.2GHz × 2)	4-way SIMD × 8-way マルチコア
東工大TSUBAME	Opteron DC 2.4GHz	ボード間: PCI-X	ClearSpeed Advance CSX600 PCI-X Board	96-way SIMD
DELL Precision T7400	Core2 Q9300 4コア2.5GHz	チップ間: PCI-Express	Nvidia GeForce 9800 1.4GHz	8-way SIMD × 16-way マルチコア
GRAPE	汎用プロセッサ・チップ	チップ間: 共有バス	重力計算専用プロセッサ GRAPE	並列演算パイプライン
九州大学EHPC/Eric	SH-4	チップ間: SH-4バス	二電子積分計算専用プロセッサEric	初期積分計算専用コア × 1 + 漸化計算専用コア × 4
CRAY XD1	Opteron	チップ間:	FPGA	ハードウェア処理
CREST SFQ-RDP	汎用プロセッサ・チップ	チップ間: 共有バス	九州大学LSRDP(再構成可能大規模データパス80GHz) MCM × 4	2次元FPUアレイ (1024FPU/MCM)
NEC SX-9	スカラ・ユニット	チップ内:	ベクトル・パイプライン × 8	並列ベクトル処理
IBM CELL/B.E.	PowerPC 4GHz	チップ内: 共有リングバス (EIB)	SPE 4GHz × 8	4-Way SIMD
IPFLEX DAPDNA-II	DAP 166MHz	チップ内: 共有バス	DNA 166MHz	2次元ALUアレイ (168ALU)
TI OMAP 3530	ARM Coretex-A8	チップ内: 共有バス	TMS320C64x DSP Core	VLIW DSP

加速方式一覽(1/2)

方針	方式	活用対象	メインプロセッサ	アクセラレータ	
基礎体力の強化	クロック周波数の向上		✓✓✓		
	メモリアクセスレイテンシの短縮		✓✓✓		
	メモリバンド幅の向上	クロック周波数向上		✓✓✓	
		信号線数拡大		✓	✓
		メモリサイズの拡大		✓✓✓	
	実行命令数の削減	命令セットアーキテクチャの最適化		✓	✓✓✓
カスタム命令		特徴的・固定的処理 データ依存関係		✓✓✓	
SIMD演算命令		データレベル並列性	✓	✓✓✓	
ベクトル演算命令		データレベル並列性		✓✓	
コンパイラ最適化機能の強化			✓✓	✓	
並列化	データレベル並列処理	データレベル並列性	✓	✓✓✓	
		SIMD演算	空間的並列性	✓	✓✓✓
		ベクトル演算	時間的並列性		✓✓
	命令レベル並列処理	命令レベル並列性		✓✓✓	
		命令パイプライン処理	時間的並列性	✓✓✓	
		スーパースカラ処理	空間的並列性	✓✓✓	

加速方式一覽(2/2)

方針	方式	活用対象	メインプロセッサ	アクセラレータ
並列化(続き)	スレッド/タスクレベル並列処理	粗粒度並列性	✓✓✓	✓✓✓
	マルチスレッディング		✓✓	
	マルチコア/メニーコア		✓✓✓	✓✓✓
カスタム化	ASIP(特定用途向けプロセッサ)	特徴的・固定的処理		✓✓
	構成可能プロセッサ			✓
	再構成可能プロセッサ			✓
	ハードウェア処理	特徴的・固定的処理		✓✓
	布線論理			✓✓
	再構成可能ハードウェア			✓
階層化	階層化メモリ	参照の局所性	✓✓✓	✓✓
	スクラッチパッドメモリ			✓✓
	キャッシュメモリ		✓✓✓	
	仮想記憶		✓✓✓	
投機化	先行制御	動的振舞の規則性	✓✓✓	
	命令プリフェッチ	参照の連続性	✓✓✓	
	データプリフェッチ	参照の規則性	✓✓	
	予測制御	動的振舞の偏り	✓✓✓	
	分岐予測		✓✓✓	
	値予測		✓	

アクセラレータ向き主要加速方式

方針	方式	活用対象	メインプロセッサ	アクセラレータ
実行命令数の削減	命令セットアーキテクチャの最適化		✓	✓✓✓
	カスタム命令	特徴的・固定的処理 データ依存関係		✓✓✓
	SIMD演算命令	データレベル並列性	✓	✓✓✓
	ベクトル演算命令	データレベル並列性		✓✓
	コンパイラ最適化機能の強化		✓✓	✓
並列化	データレベル並列処理	データレベル並列性	✓	✓✓✓
	SIMD演算	空間的並列性	✓	✓✓✓
	ベクトル演算	時間的並列性		✓✓
	命令レベル並列処理	命令レベル並列性	✓✓✓	
	スレッド／タスクレベル並列処理	粗粒度並列性	✓✓✓	✓✓✓
	マルチスレッディング		✓✓	
	マルチコア／メニーコア		✓✓✓	✓✓✓
カスタム化	ASIP(特定用途向けプロセッサ)	特徴的・固定的処理		✓✓
	構成可能プロセッサ			✓
	再構成可能プロセッサ			✓
	ハードウェア処理	特徴的・固定的処理		✓✓
	布線論理			✓✓
	再構成可能ハードウェア			✓

アクセラレータとは？

メイン
プロセッサ

結合路
(結合方
式は？)

アクセラレータ
(加速方式は？)

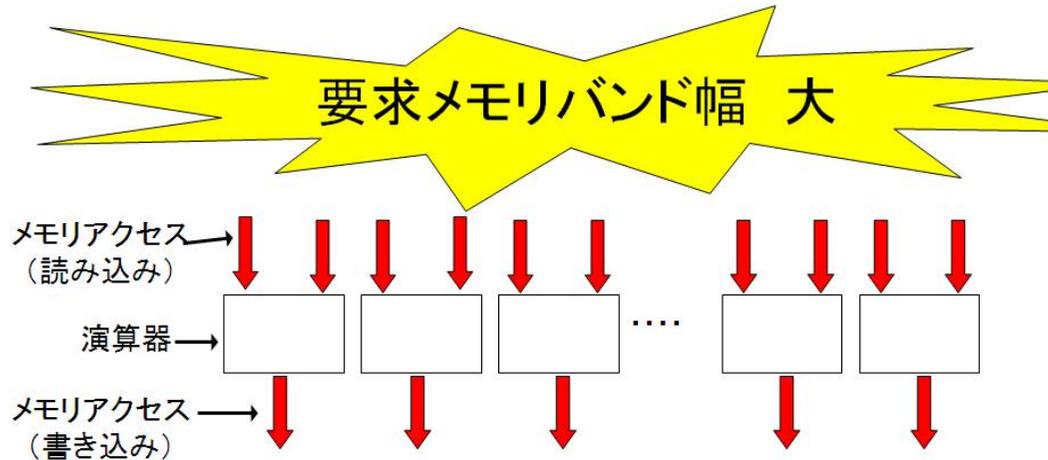
しかし、並列処理度に比例して増加するメモリアクセス頻度(メモリプレッシャー、要求メモリバンド幅)にどう対処するか？

- コア内はデータレベル並列性を活用したカスタム化
- または、汎用性を重視してデータレベル並列処理(SIMD演算)のみ
- 上記と直交する形でマルチコア化

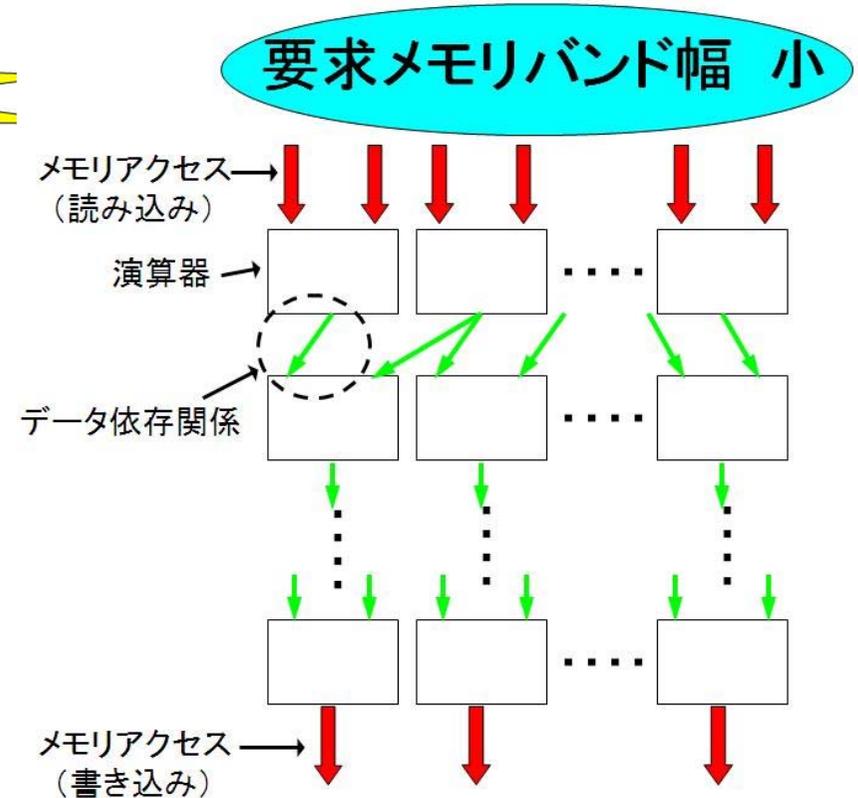
アクセラレータの課題

～増大する要求メモリバンド幅にどう対応するか？～

従来のアクセラレータの場合
(SIMD演算、ベクトル演算)

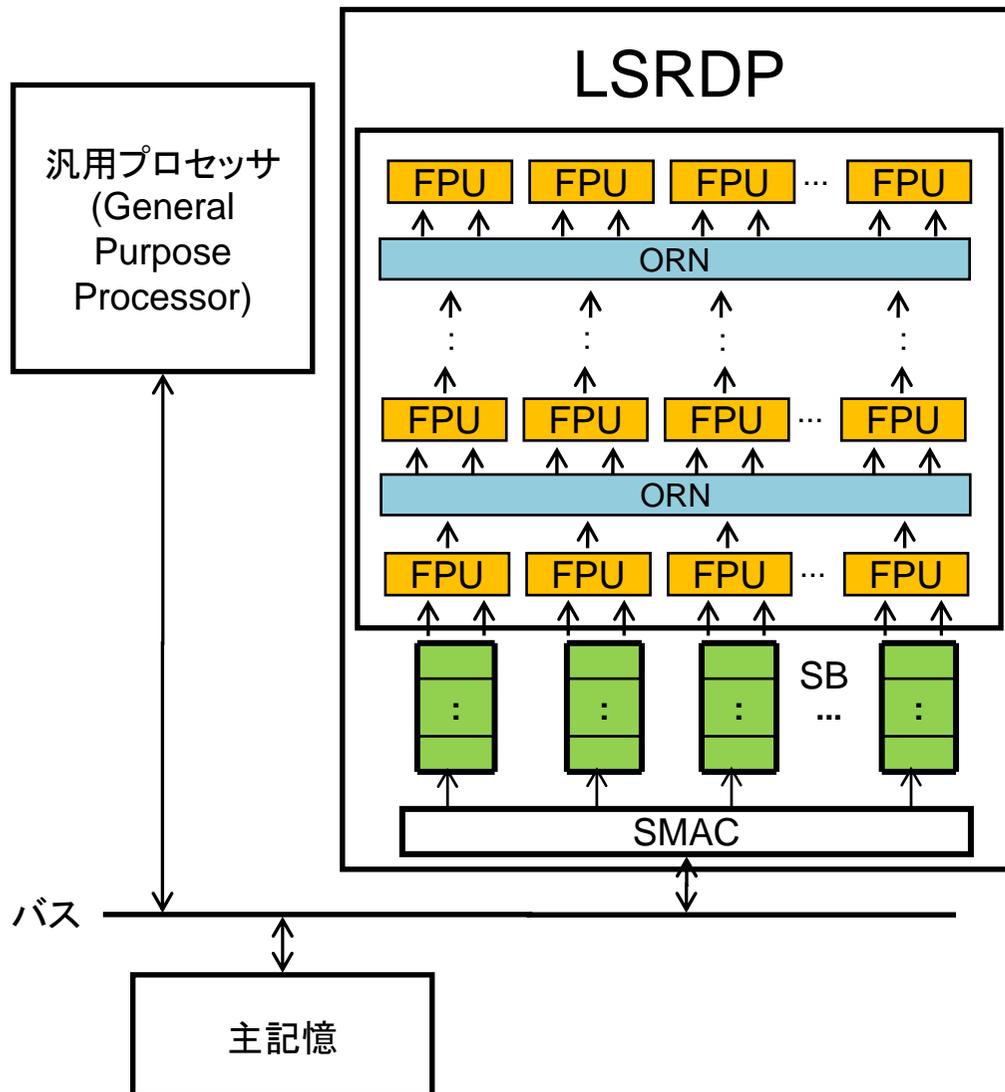


大規模データパス (LSRDP)
の場合



大規模再構成可能データパス

(LSRDP: Large Scale Reconfigurable Data Path)



- 多数の演算器 (FPU: Floating-Point Unit) とそれらを相互接続する網 (ORN: Operand Routing Network) を搭載し、
 - FPUで行う演算内容
 - ORN上のFPU間接続関係を再構成可能としたデータパス

データレベル並列性とデータ依存関係を同時に活用することにより、要求メモリバンド幅を抑えつつ高い演算性能を実現！

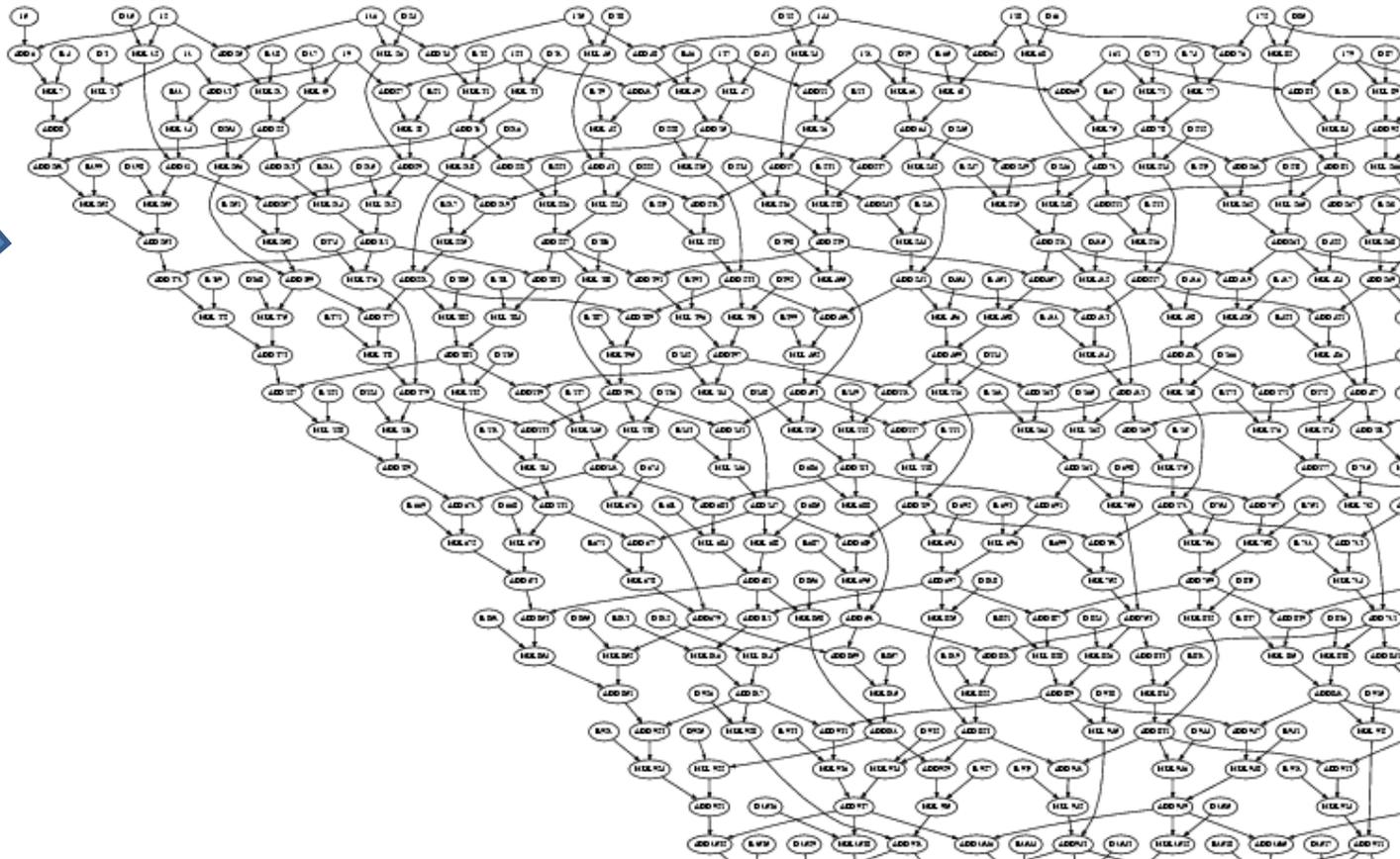
アプリ例：1次元時間発展熱伝導方程式

$$T(x_i, t_{j+1})$$

$$= T(x_i, t_j) + A * [T(x_{i-1}, t_j) + T(x_{i+1}, t_j) - 2T(x_i, t_j)] / (\Delta x)^2$$

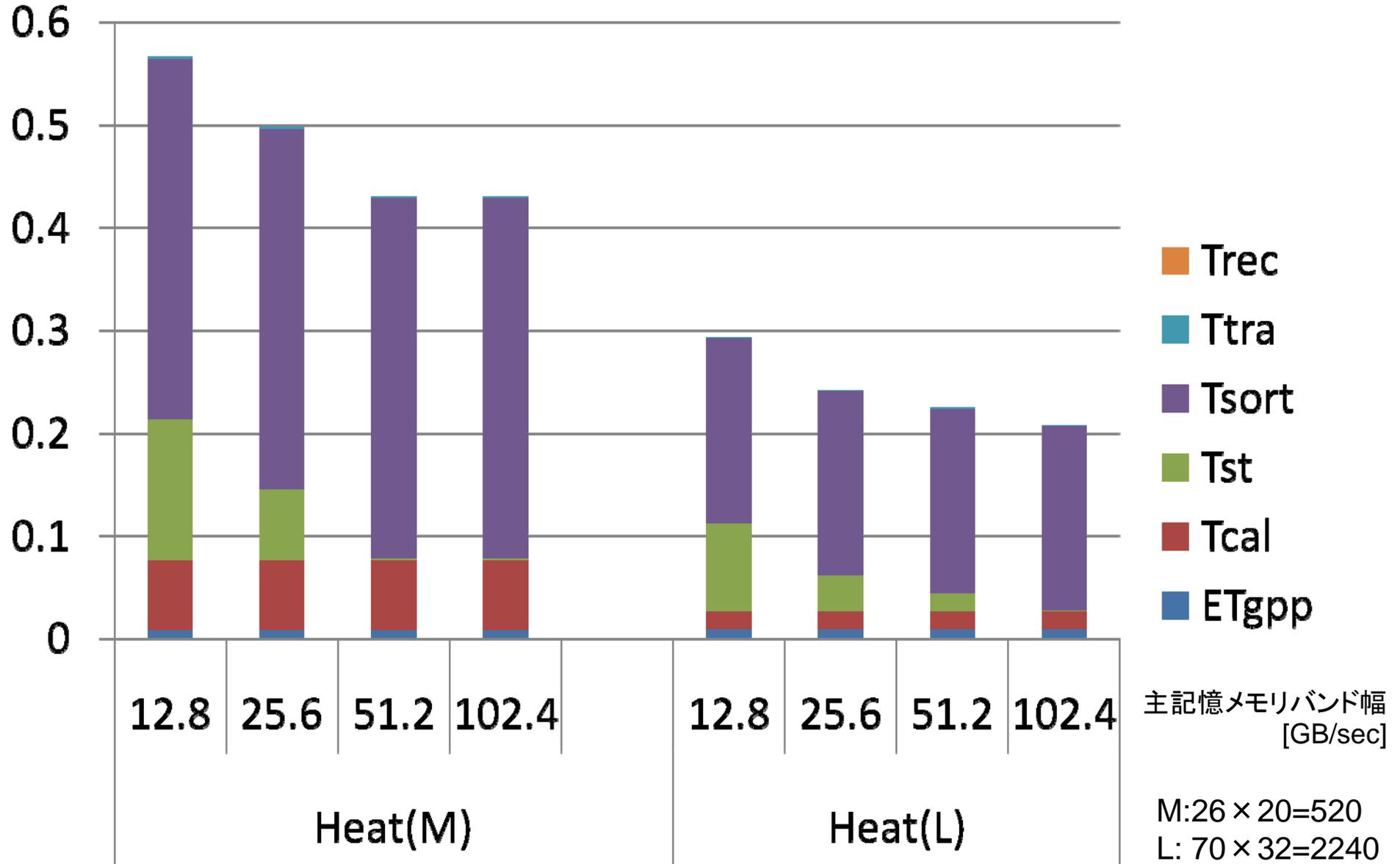
$$= D * T(x_i, t_j) + B * [T(x_{i-1}, t_j) + T(x_{i+1}, t_j)]$$

32 × 32 LSRDP : 32入力、16出力、728 FPU (add, mul)、364即値



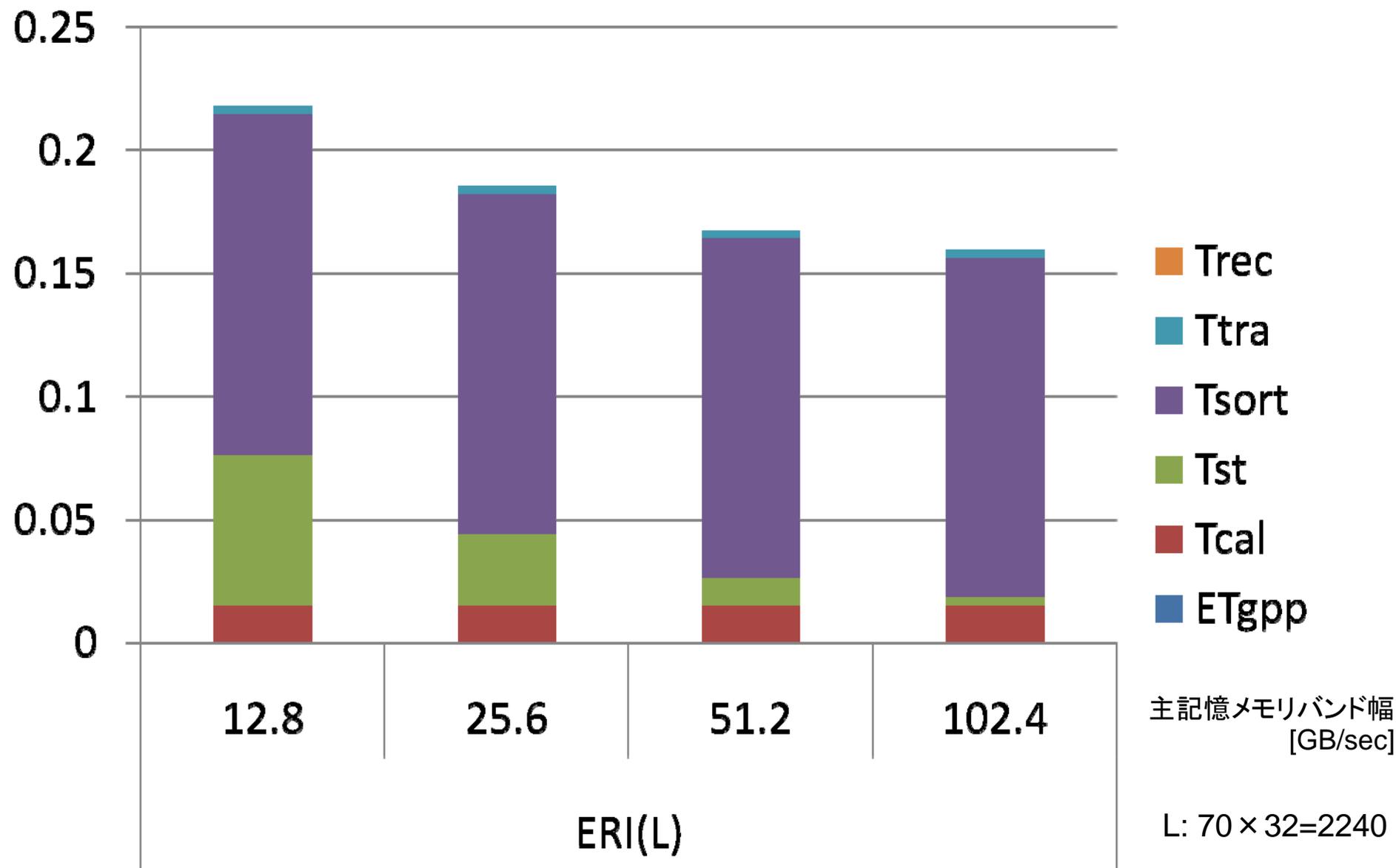
アプリ例：1次元時間発展熱伝導方程式

正規化した実行時間(3.2GHz SimpleScalarの実行時間を1)



アプリ例: 二電子積分計算

正規化した実行時間 (3.2GHz SimpleScalarの実行時間を1)

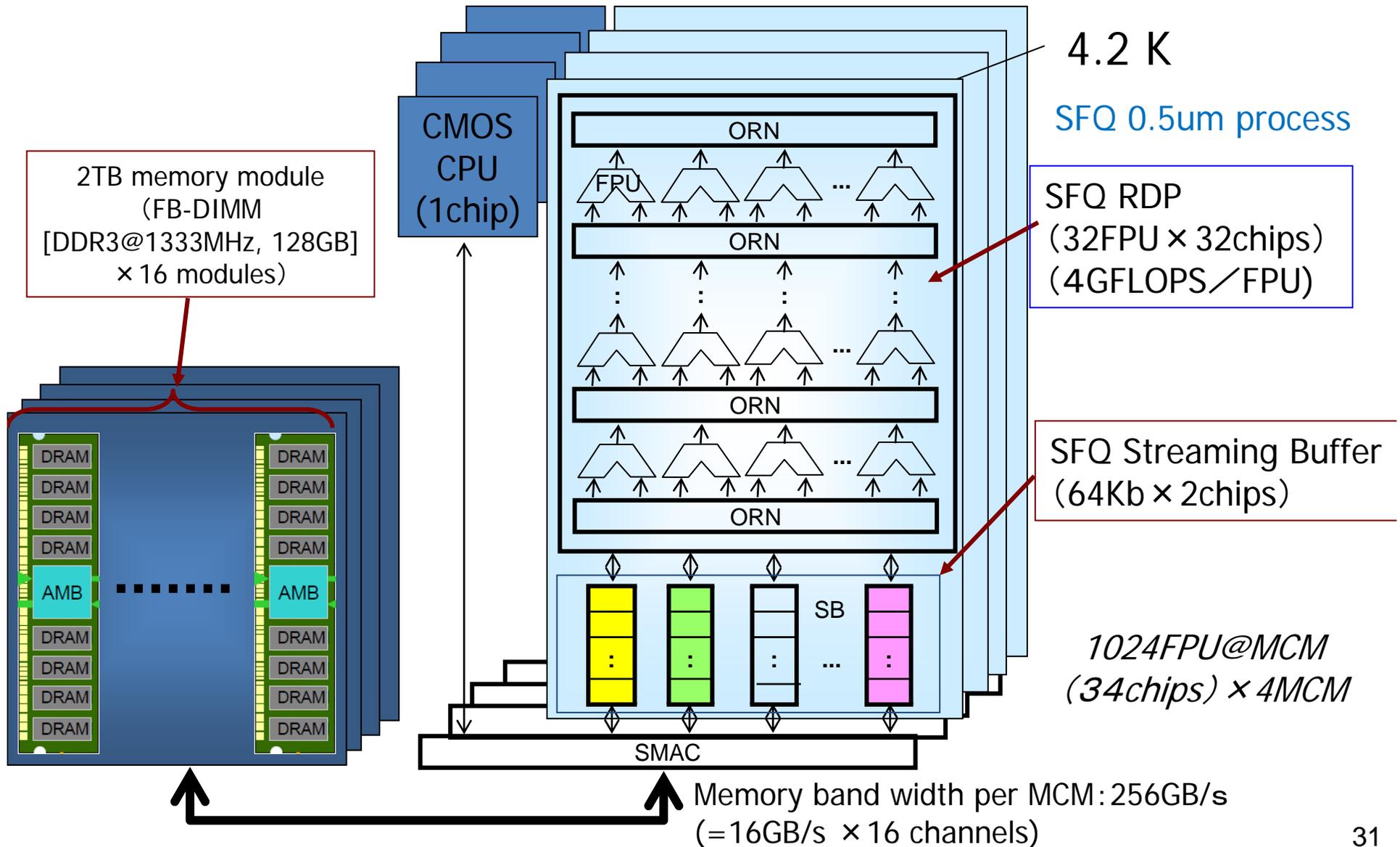


アクセラレータ & 採用システム実例一覧



採用システム	メインプロセッサ	結合方式	アクセラレータ	加速方式
IBM Roadrunner	BladeCenter LS21 (Opteron DC 1.8GHz × 2)	ボード間: デュアル・ギガビット・イーサネット	BladeCenter QS22 (PowerXCell 8i 3.2GHz × 2)	4-way SIMD × 8-way マルチコア
東工大TSUBAME	Opteron DC 2.4GHz	ボード間: PCI-X	ClearSpeed Advance CSX600 PCI-X Board	96-way SIMD
DELL Precision T7400	Core2 Q9300 4コア2.5GHz	チップ間: PCI-Express	Nvidia GeForce 9800 1.4GHz	8-way SIMD × 16-way マルチコア
GRAPE	汎用プロセッサ・チップ	チップ間: 共有バス	重力計算専用プロセッサ GRAPE	並列演算パイプライン
九州大学EHPC/Eric	SH-4	チップ間: SH-4バス	二電子積分計算専用プロセッサEric	初期積分計算専用コア × 1 + 漸化計算専用コア × 4
CRAY XD1	Opteron	チップ間:	FPGA	ハードウェア処理
CREST SFQ-RDP	汎用プロセッサ・チップ	チップ間: 共有バス	九州大学LSRDP(再構成可能大規模データパス80GHz)MCM × 4	2次元FPUアレイ (1024FPU/MCM)
NEC SX-9	スカラ・ユニット	チップ内:	ベクトル・パイプライン × 8	並列ベクトル処理
IBM CELL/B.E.	PowerPC 4GHz	チップ内: 共有リングバス (EIB)	SPE 4GHz × 8	4-Way SIMD
IPFLEX DAPDNA-II	DAP 166MHz	チップ内: 共有バス	DNA 166MHz	2次元ALUアレイ (168ALU)
TI OMAP 3530	ARM Coretex-A8	チップ内: 共有バス	TMS320C64x DSP Core	VLIW DSP

CREST SFQ-RDP



HPCとアクセラレータの今後

