

「PCクラスタワークショップ in 大阪2015」

# PEZY Computing

Peta/Exa/Zetta/Yotta ... PetaFLOPS and beyond

民間ベンチャー企業による、小規模  
スーパーコンピュータ開発の取り組み

2015年 2月20日

株式会社PEZY Computing

UltraMemory株式会社

株式会社ExaScaler

創業者 齊藤 元章

# PEZYグループの構成

創業:2010年  
社員数:11名

## 株式会社PEZY Computing (ペジーコンピューティング)

- ・独自メニーコア・プロセッサ開発
- ・同汎用PCIeボード開発
- ・同独自システムボード開発
- ・同アプリケーション開発
- ・半導体2.5次元実装技術開発
- ・ウェハ極薄化応用技術開発

創業:2014年  
社員数:9名

## 株式会社ExaScaler (エクサスケラー)

- ・液浸冷却技術開発
- ・HPC液浸システム開発
- ・液浸スパコンシステム開発
- ・液浸冷却水槽販売
- ・液浸冷却システム販売
- ・液浸冷却用ボード類販売

創業:2013年  
社員数:26名

## UltraMemory株式会社 (ウルトラメモリ)

- ・超広帯域独自DRAM開発
- ・DRAM積層技術開発
- ・磁界結合メモリIF開発
- ・ウェハ極薄化応用技術開発
- ・広帯域、高速DRAM開発
- ・最先端汎用DRAM受託開発

PEZY Computing社で開発するメニーコア・プロセッサ、  
UltraMemory社で開発する超広帯域積層カスタムDRAM、  
ExaScaler社で開発する液浸冷却システムを組み合わせ、  
最終システムとしてHPC、スーパーコンピュータ製品を開発  
すると共に、各社の要素技術と製品を個別に展開します。

# これまでのプロセッサ等開発実績

創業者2名による合計13種類の大規模プロセッサ等開発履歴(現法人以前の11種)

極めて小規模な開発グループながらも、平均で1年半に1デザインの開発ペースで1995年から20年間に渡るプロセッサ等の開発を経験。

医療システム系法人内でのプロセッサ開発は、2005年以降の開発が鈍化。好不況の波が小さいとされる医療業界でも、収益悪化によって真っ先に開発費を削られるのがプロセッサ開発であった。過去3回の憂き目を見て、プロセッサ開発に完全に特化した事業体設立の必然性を痛感する。

Processor	Year	Process	Die Size (mm)	Clock	Gates	Architecture	Core number	FLOPS	Power	Memory
Version 1.0	1997	600nm	8.0*8.0	50MHz	1.2M	VLIW+SIMD	1 Core/8 ALU	Fixed Point	6W	SDR
Version 1.5	1999	350nm	7.3*7.3	80MHz	1.5M	VLIW+SIMD	1 Core/8 ALU	Fixed Point	3W	SDR
3DVR Version 1.0	1999	350nm	13.65*13.65	133MHz	0.8M	Hardwired Pipeline	2 Pipeline	-	32W	DDR
Version 2.0	2001	250nm	8.1*8.1	80MHz	1.8M	VLIW+SIMD	1 Core/8 ALU	160M	2W	SDR
3DVR Version 2.0	2001	160nm	15.6*15.6	250MHz	3.2M	Hardwired Pipeline	4 Pipeline	-	20W	DDR
Version 2.0 shrink	2003	180nm	6.5*6.5	167MHz	1.8M	VLIW+SIMD	1 Core/8 ALU	333M	1W	SDR
Version 2.5	2003	180nm	6.5*6.5	167MHz	2M	VLIW+SIMD	1 Core/8 ALU	333M	2W	DDR
DBF Version 1.0	2003	180nm	11.5*9.6	40MHz	2.5M	Hardwired Pipeline	-	-	10W	-
Version 3.0	2005	130nm	16.5*12.0	333MHz	34M	RISC+VLIW+SIMD	8 Core/40 ALU	13.3G	19W	DDR
Version 3.0 B	2005	130nm	9.5*12.0	250MHz	20M	VLIW+SIMD	1 Core/8 ALU	8G	6W	DDR
3DVR Version 3.0	2008	130nm	10.5*10.5	333MHz	5.5M	Hardwired Pipeline	2 Pipeline	-	10W	DDR2

# 現法人でのプロセッサ開発と予定

創業者2名による合計13種類の大規模プロセッサ等開発履歴(現法人以降)

2010年に設立したPEZY Computingでは、事業の状況に関わらずに、2年毎に1世代のプロセッサ開発の速度を維持したいと考えています。

開発中のPEZY-SC2では、一気に4倍となる4,096コアの集積を実現する計画です。

PEZY-SC3以降では、集積度を更に上げ、高価なIPを再利用して、最先端プロセスを早期に利用する目的から、Multi-Dieプロセッサ構成を計画します。

今後も世代毎に搭載するコア数を2倍ずつ増やして、最終的には10kコアを大きく超える集積度を目指します。

その過程で、HMCやHBMでも不足することが確実なメモリ帯域を十分確保するため、独自のメモリIFと、低消費電力・超広帯域・大容量を実現する積層メモリを、グループ内で開発して行きます。

Processor	Year	Process	Die Size (mm)	Clock	Gates	Architecture	Core number	FLOPS Double/Single	Power	Memory
PEZY-1	2012	40nm	21.0*16.8	533MHz	220M	RISC+SMT (MIMD)	512 Core	166/333G	35W	DDR3/Wide IO
PEZY-SC	2014	28nm	21.1 *19.5	733MHz	580M	RISC+SMT (MIMD)	1,024 Core	1.5/3.0G	70W	DDR4/Custom Ultra-Wide IO
PEZY-SC2 (under development)	2016	14/16nm	TBD	1GHz	2.4G+	RISC+SMT (MIMD)	4,096 Core	8.2/16.4T	100W	HMC or HBM/ Custom TCI Stacked DRAM

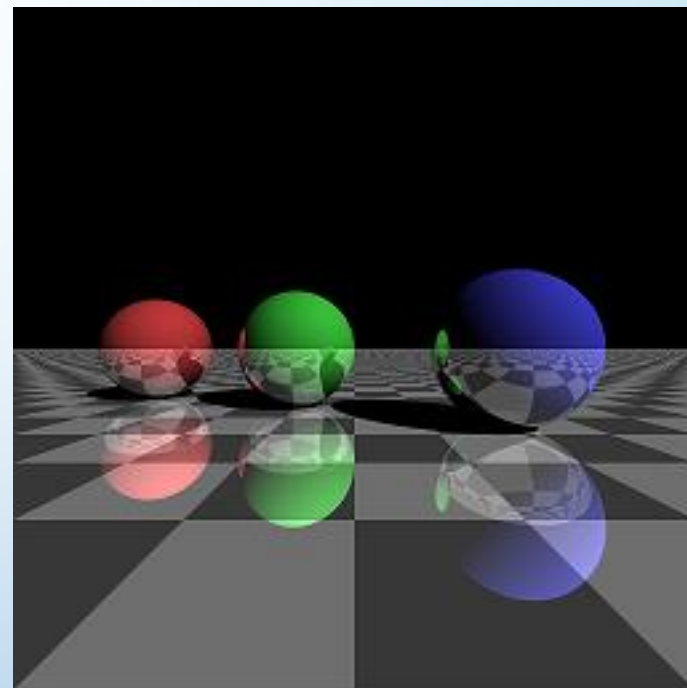
# 想定するアプリケーション

## 産業用アプリケーション

産業用高解像度画像装置: 信号処理、再構成処理、画像表示・処理・解析  
次世代データセンター用ビッグデータ処理  
ネットワーク・無線通信機器用信号処理  
リアルタイム・レイトレーシング (右図)  
地質学情報表示・解析、その他

## スーパーコンピューティング

流体解析、MD計算、QCD計算  
化合物解析、新物質探索  
遺伝子解析、タンパク質解析、代謝解析  
核融合実験解析・シミュレーション  
気象予測、地震予測、防災・減災、その他



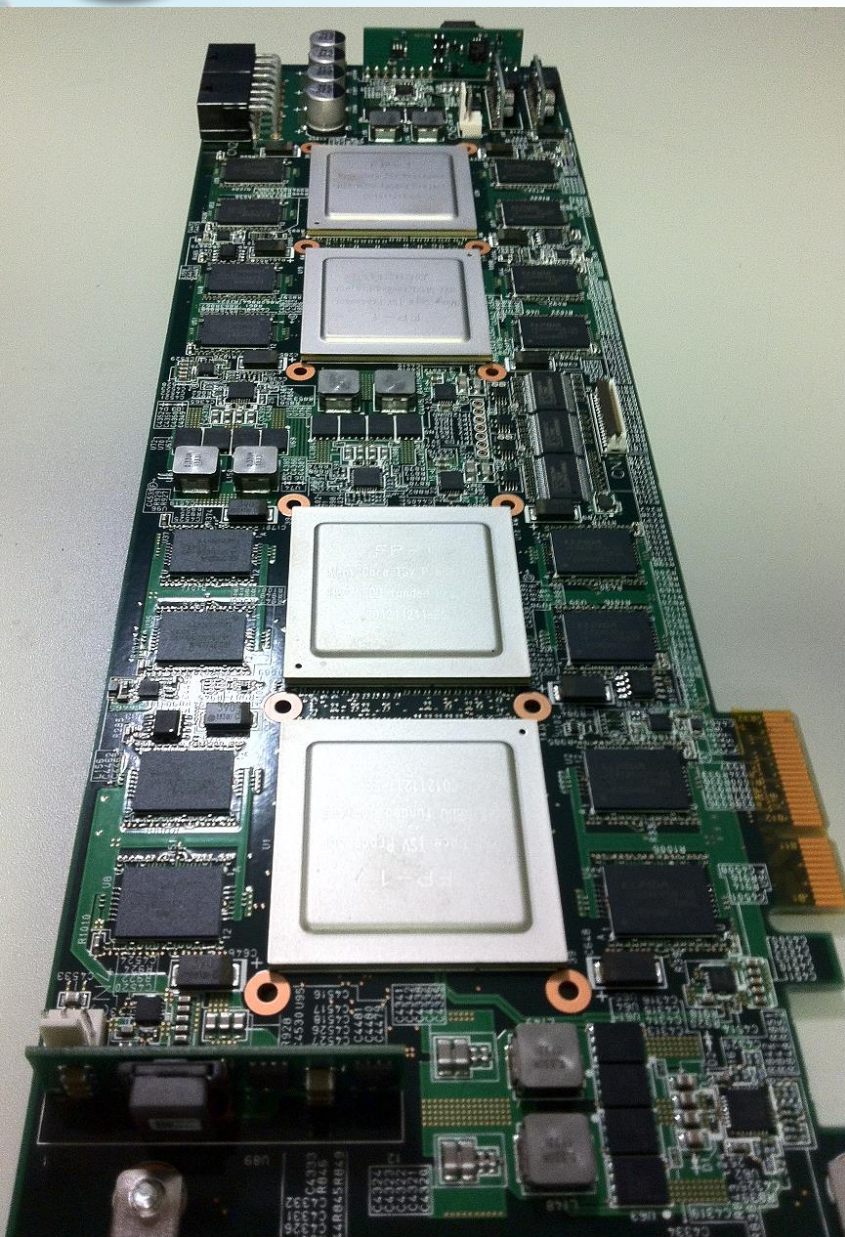
## 医療系アプリケーション

創薬、次世代自動診断、治療効果判定、予後予測、新疾患発見システム  
診断装置への搭載: 小型・次世代超音波装置、CT装置、MRI装置  
画像処理システム: 3次元・4次元 Volume Rendering、形状認識、自動診断

## クラウドコンピューティング、AI

各種

# マルチプロセッサ (PEZY-1) ボード



## 64GBのメモリを搭載した、汎用PCIeボード

大容量積層TSVメモリ(社外)を世界で初めて採用したQuad Processorボード

総プロセッサコア数	:2,048
総メモリ容量	:64GB
総メモリ帯域	:200GB/sec.
総演算性能	:2.56TFLOPS
消費電力	:180Watt

# PEZY-1採用事例

「PEZY-1」は、超音波装置の信号処理、画像処理用プロセッサとして、国内大手医療機器メーカーの新型ポータブル超音波装置に採用済み。据置き型装置でしか得られなかった機能、画質、性能をポータブル装置で実現し、電池駆動によりポータブル機ながら1時間の稼働時間を確保。

今後、研究室レベルでは既に関済済みの新機能で、ポータブル型の超音波装置本体には実装が不可能とされてきた機能群を順次実装していく予定。



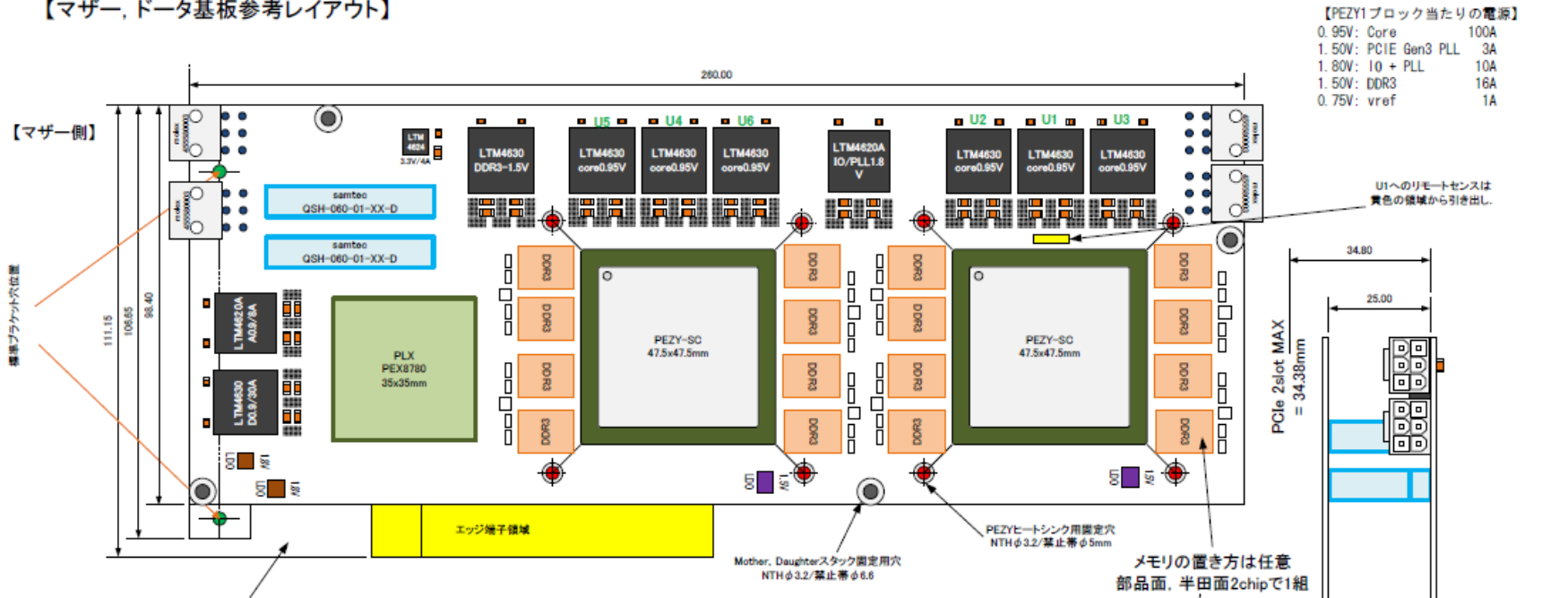
# 完成した「PEZY-SC」の仕様

Processor		PEZY-1	PEZY-SC
製造プロセス		TSMC 40G (40nm)	<b>TSMC 28HPM (28nm)</b>
	ダイサイズ	335mm <sup>2</sup>	412mm <sup>2</sup>
コア性能	動作周波数	533MHz	<b>733MHz</b>
	キャッシュ	L1: 128KB, L2: 1MB, L3: 無し	L1: 1MB, L2: 4MB, <b>L3: 8MB</b>
周辺回路	動作周波数	66MHz	66MHz
IPs	内蔵CPU (管理・デバッグ用)	ARM926 x 2 Cache L1: 16KB*2, L2: None	ARM926 x 2 Cache L1: 32KB*2, <b>L2: 64KB</b>
	PCIe	PCIe Gen2 x 4Lane 6Port (2GB/s x 6 = 12GB/s)	PCIe Gen3 x 8Lane 4Port (8GB/s x 4 = <b>32GB/s</b> )
	DDR	DDR3 64bit 1,333MHz 4Port (10.6GB/s x 4 = 42.6GB/s)	DDR4 64bit 2,400MHz 8Port (19.2GB/s x 8 = <b>153.6GB/s</b> )
コア (PE) 数		512 PE + 64 PE (redundancy)	<b>1,024 PE</b>
演算性能		533G Flops (単精度浮動小数点)	3.0T Flops (単精度浮動小数点)
		266G Flops (倍精度浮動小数点)	<b>1.5T Flops</b> (倍精度浮動小数点)
消費電力		50W (Leak: 20W, Dynamic: 30W)	60W (Leak: 10W, Dynamic: 50W) <b>46W@533MHz</b> (PEZY-1以下)
パッケージ	DDR版	40*40mm (1,517pin)	47.5*47.5mm (2,112pin)



# Dual「PEZY-SC」ボード (DDR3版)

【マザー、ドータ基板参考レイアウト】



【PEZY1ブロック当たりの電源】

0.95V: Core	100A
1.50V: PCIe Gen3 PLL	3A
1.80V: IO + PLL	10A
1.50V: DDR3	16A
0.75V: vref	1A

【ドータ側】

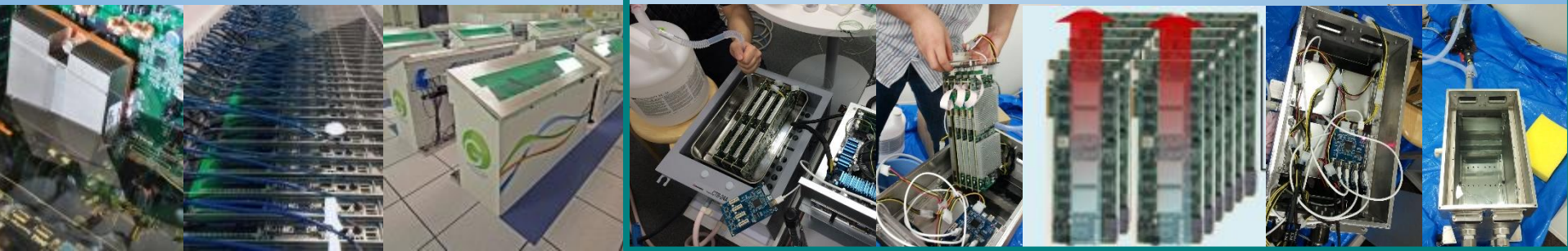


# 独自の「液浸冷却」システムを開発

ベンチャー企業として大規模システムを開発することは到底不可能で、必然、極力小規模なシステムを開発して性能を追求するしかなかった。その観点から「液浸冷却」しか選択肢は無く、当初は油浸冷却を検討。最終的に体積効率を極大化する目的で、不活性液体であるフッ化炭素で、気化熱による冷却に用いられる低沸点品ではなく、174度の高沸点品を液相のみで使用する方法でも、非常に高い熱密度の冷却が可能であることを確認し、主要半導体表面温度が30度台での運用に目処をつけることが出来た。(気化熱による冷却は、2次冷却の低効率から断念。)

1) 同液浸冷却手法、2) 液相と気相双方を混在させるハイブリッド型液浸冷却手法、3) 液浸槽内温度分布を最小化しつつ高い保守を容易にするための液浸槽構造、4) 液浸冷却に特化したモジュールとキャリアボードによる基板構成に関する基本特許を申請して、各種開発を推進。

当社内の独自手法による液浸冷却基礎実験の様子

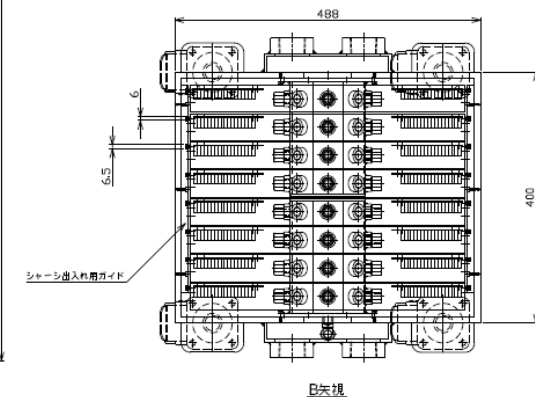
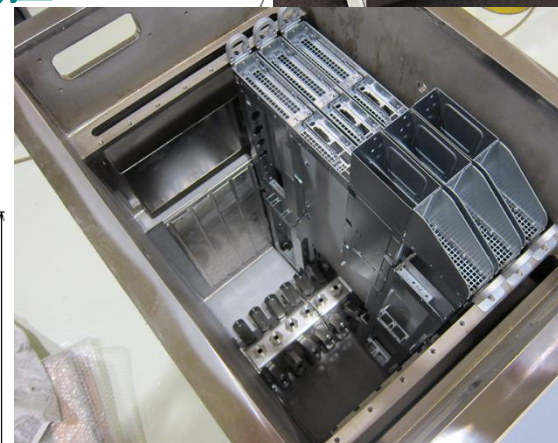
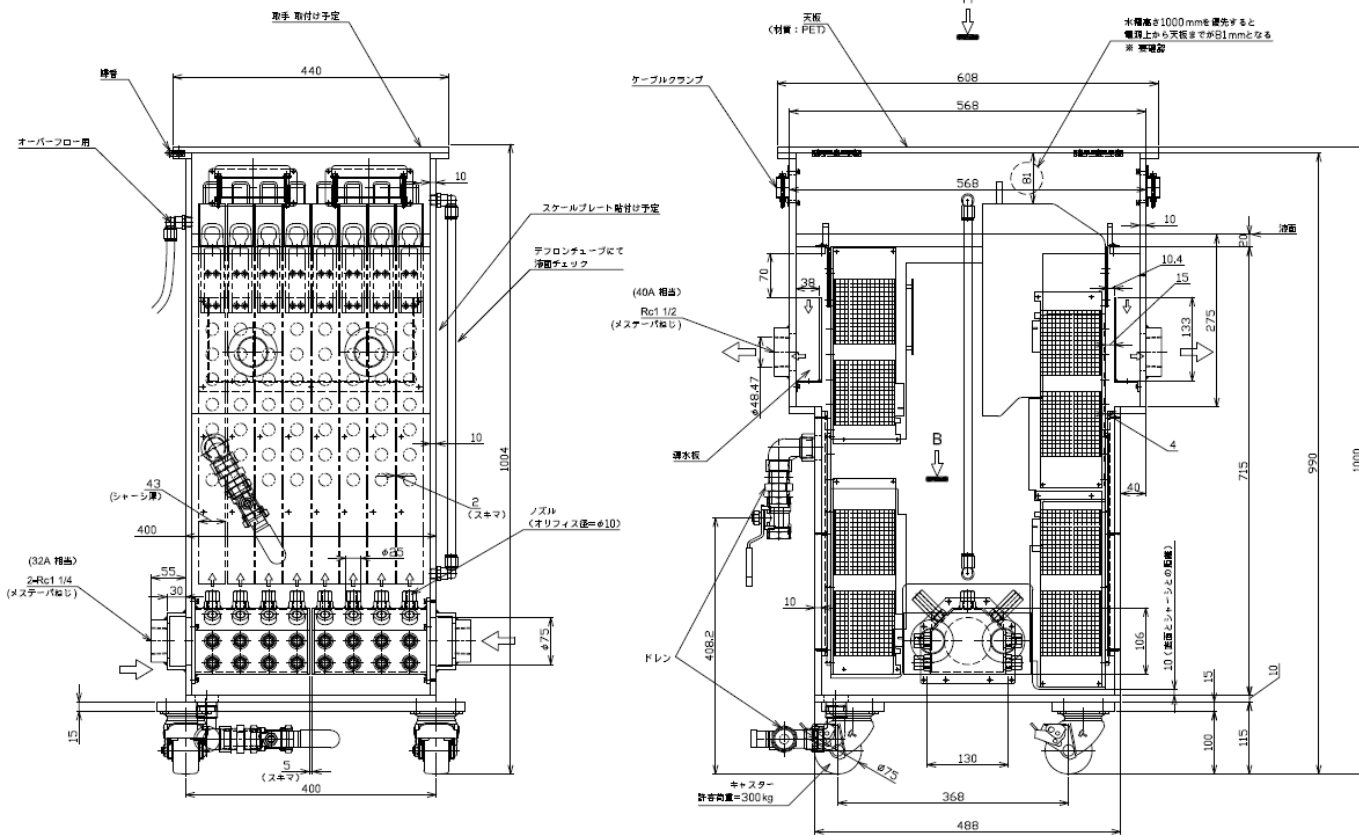
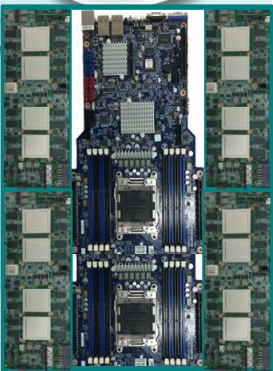


# 液浸冷却スパコン「ExaScaler-1.0」

液浸冷却によるGREEN500(2014年11月)申請用システム  
独自開発した液浸槽「ESLC-8」を4台使用して構成

目標) HPL性能: 240TFLOPSを30kW以下で実現 (8GFLOPS/W)  
結果) HPL性能: 191TFLOPS/38.6kW = 4.95GFLOPS/Wに留まる

フッ化炭素浸漬型の液浸槽(独自開発) + 外部冷却装置  
本構成を4セット使用し、理論最大値395TFLOPS



# 液浸槽「ESLC-8」を用いた「ExaScaler-1.0」



# 高エネルギー加速器研究機構「Siren (睡蓮)」



# 「Top500」と「Green500」の結果

10月31日までに高エネルギー加速器研究機構様に搬入と設置を終えさせて頂いた、理論性能395TFLOPSの「Suiren(睡蓮)」が稼働を開始しています。実質稼働日数が限られていましたが、10月31日締切りの「Top500」で369位(178.1TFLOPS)、11月14日締切りの「Green500」では4.95GFLOPS/W(186TFLOPS時)を計測して2位(4.95GFLOPS/W)に認定されました。

AMD社の最新GPGPU「FirePro S9150」を採用し、4か月の最適化期間を確保して世界1位となったドイツのGSI Helmholtz Center (HPL全演算時間平均で5.27GFLOPS/W)には及ばなかったものの、今後の最適化を進めて同等以上の数値を目指します。



# 「ExaScaler-1.0」開発の実際(1)

## ExaScaler-1.0開発成功の理由:

- HPL実装は、理化学研究所の牧野淳一郎先生に、有難くも公開されたGRAPE-DR用コードを御提供頂いた上、多々御指導と御助力を頂いた。牧野チーム所属の似鳥啓吾先生にも、コード最適化に御参画頂いた。
- 設置先である高エネルギー加速器研究機構様と同計算科学センターの石川正先生から、多大な御配慮と御支援を頂くことが出来た。
- Green500申請用の電力測定には、経験豊富な東京大学平木研究室の平木敬先生を始めとする皆様に、御協力を仰ぐことが出来た。
- 社外の協力会社(液浸槽製作、冷却装置製作、基板製造・アセンブリ、各種部品と資材の調達)各社様の献身的な御協力を頂いた。
- 8月12日にESが上がって来たばかりの大規模プロセッサ「PEZY-SC」のデバッグと立ち上げを、僅か1か月間で完了することが出来た。
- PEZY-SCと液浸冷却以外の新規開発要素を極力排して工数を最小化し、PEZY-SCの開発を除いた全てを7か月の期間に収めることが出来た。
- 10月初頭から11月前半まで、社員に相当の頻度での徹夜作業を敢行して貰い、設置サイトに泊まり込んでの集中的な開発を完遂出来た。

# 「ExaScaler-1.0」開発の実際(2)

## 数々の至らぬ点と反省:

- 液浸冷却の有効性と多大なメリットは実験で十分に確認できたものの、「液相のみの冷却」による単純な構成しか間に合わせられなかった。
- 空冷前提に設計された19インチラック用ブレードサーバーを流用。
- 液浸冷却の大きな利点の体積の最小化を、十分に実現出来なかった。
- マザーボードデザインの制約から、PEZY-SC間通信は採用出来ず。
- PEZY-SCは完全なES品で、パラメータ特性が全くの未知数であった。
- 消費電力特性なども未知数で、パッケージ設計が十分でなかった。
- 調達時期と入手性の問題から、PEZY-SC用メモリは高速、低消費電力のDDR4品ではなく、DDR3品を採用せざるを得なかった。
- PEZY-SCモジュールボードもES版で、駆動電圧を下げられなかった。
- 折角、貴重なHPLコードを牧野淳一郎先生から御提供頂きながらも、実装時間が足りずにホスト(Xeon)での処理が少なくなってしまう。
- PCIe Switch (80 lane品)の消費電力も予想以上に大きかった。
- 基板上のDC-DCの容量が不足し、また変換効率が予想を下回った。



# 開発を開始した「PEZY-SC2」

Processor		PEZY-SC	PEZY-SC2
製造プロセス		TSMC 28HPM (28nm)	<b>TBD (14-16nm FinFET)</b>
	ダイサイズ	412mm <sup>2</sup>	400-500mm <sup>2</sup>
コア性能	動作周波数	733MHz	<b><u>1GHz</u></b>
	キャッシュ	L1: 1MB, L2: 4MB, L3: 8MB	<b>50MB in total (TBD)</b>
周辺回路	動作周波数	66MHz	66MHz
IPs	内蔵CPU	ARM926 x 2 管理・デバッグ用	<b>64bit CPU (TBD)</b> <b>汎用演算用</b>
	PCIe	PCIe Gen3 x 8Lane 4Port (8GB/s x 4 = 32GB/s)	PCIe Gen3/4 x 8Lane 6Port <b>(48-96GB/s)</b>
	DRAM	DDR4 64bit 2,400MHz 8Port (19.2GB/s x 8 = 153.6GB/s)	Custom Stacked DRAM 8Port <b>(500GB/s x 8 = 4.0TB/s)</b>
コア (PE) 数		1,024 PE	<b><u>4,096 PE</u></b>
演算性能		3.0T Flops (単精度浮動小数点) 1.5T Flops (倍精度浮動小数点)	16.4T Flops (単精度浮動小数点) <b><u>8.2T Flops</u></b> (倍精度浮動小数点)
消費電力		60W (Leak: 10W, Dynamic: 50W)	100W (Leak: 10W, Dynamic: 90W)
パッケージ		47.5*47.5mm (2,112pin)	Multi-Die Package (TBD)

# 「スカンク・ワークス」の開発手法

## ステルス

Ben R. Rich SKUNK WORKS

スカンク・ワークスの秘密

ベン・R・リッチ[著] 増田興司[訳]

## 戦闘機



講談社

座右の書:「ステルス戦闘機(スカンク・ワークスの秘密)、1997年(絶版)、Ben R. Rich著」

製造段階で数千人規模の作業となる最先端の超高度戦闘機技術開発も、主要な開発は10人程度の小さなチームで短期間に、しかも矢継ぎ早に行われた事実に衝撃を受ける。

(書籍紹介文)

例のない創造型組織の責任者が明かす成功のノウハウ

数々の最新鋭軍用機を開発した知られざる[創造型組織]の全て

スカンク・ワークス(ロッキード先進開発計画)はF-104・U-2・SR-71・F-117Aなどのハイテク機を短期間・低コストで開発

〈少数精鋭・独立・秘密〉を第一義にした例のない組織の責任者がその全貌を明かす