



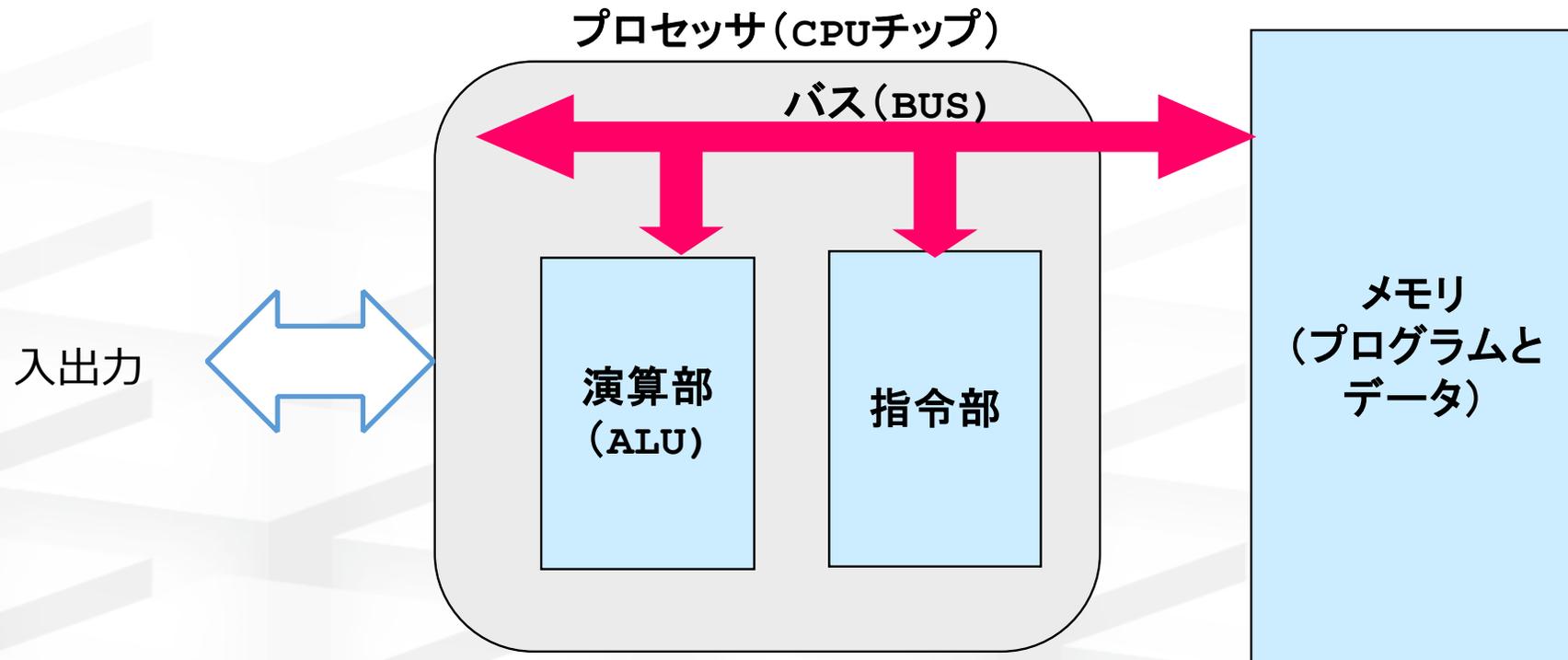
スーパーコンピュータ のベンチマークについて

理化学研究所 計算科学研究センター 副センター長
フラッグシップ2020 副プロジェクトリーダー
アーキテクチャ開発チーム・チームリーダー
佐藤 三久

計算機の原理について、少し勉強してみましょう。

コンピュータの基本的な仕組み

- いまのコンピュータ（ノイマン型）は、プロセッサ（CPU、コア）とメモリでできている。
 - メモリはプログラムやデータを格納する場所
 - プロセッサはそのメモリからプログラムやデータを読み出して、プログラムを実行しています。
 - 指令する部分：プログラムを解釈（？）して指令する
 - 演算する部分：足し算や掛け算をする部分



プログラム

● コンピュータに対する指令書

- メモリに書かれていて、CPUはメモリから一つずつ読みだして、実行する
- メモリは、データをいれておく箱のようなもの。番地という番号がついている。
- レジスタ：一時的な計算場所



1. レジスタを0にする。
2. 100番地からデータを読み出す
3. レジスタに加算
4. 101番地からデータを読み出す
5. レジスタに加算
6. 102番地からデータを読み出す
7. レジスタに加算
8.

レジスタ [27]

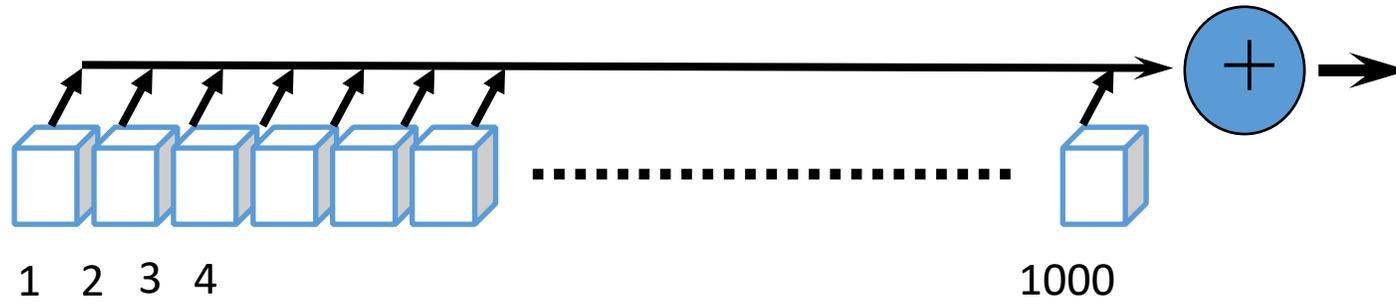
10 14 3



番地 100 101 102

クロックという
信号を基準として、
1ステップずつ実行

1000個の数の合計を計算する場合は…



- プログラムは、1000個のメモリの箱にある数を順番に読み出して、加算すればよい。
- 1つの数を取り出して、加算するのに、1秒とすると、1000個の数だと、1000秒かかる。
- 実際のコンピュータでは、数十ナノ秒（十億分の1秒）で一つの加算ができるので、十数マイクロ秒（100万分の1秒）で計算できる！

計算機を速くする方法

① 動作を速くする。

- クロックを速くする

(PCのプロセッサは2~3GHzの周波数)

- 速いトランジスタ (回路) をつかう

- マイクロプロセッサ

一つのチップの中にCPUが全部はいったコンピュータ。PCに使われ、今のマイクロプロセッサは、昔のスパコンよりもはやい!



Intel の
プロセッサ

コンピュータを速くするには、…

② コンピュータの中を工夫する

- 一度に、たくさんの命令を実行できるようにする
- メモリの読み出しを速くする… など。

- ベクトル型スーパーコンピュータ

科学技術計算によく使われる行列演算を得意とする計算機

1990年代には、日本のスーパーコンピュータはこのタイプで全世界で広く使われていた。いまのマイクロプロセッサにも同じような機能がある。



Fujitsu VPP500



Fujitsu VPP5000



NEC SX-4



NEC SX-5

富岳のプロセッサA64FXは？

- Armの命令セット⇒スマートフォンのプロセッサと同じ命令セット
原理的には、同じプログラムが動作する。
- SIMDと呼ばれる命令セット⇒技術的にはベクトルスパコンと類似
我が国のスパコンの経験が生きている
- HBM (high bandwidth memory)という高速のメモリを持つ

ベクトルコンピュータの技術



VPP5000

「富岳」の技術



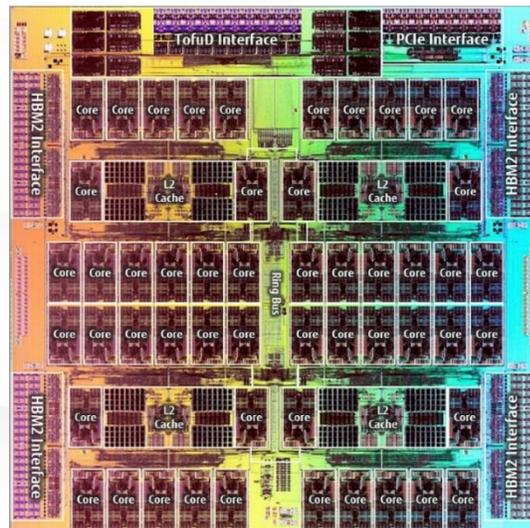
スマートフォンの技術



コンピュータを速くするには、...

③ たくさんのコンピュータを同時に使う。

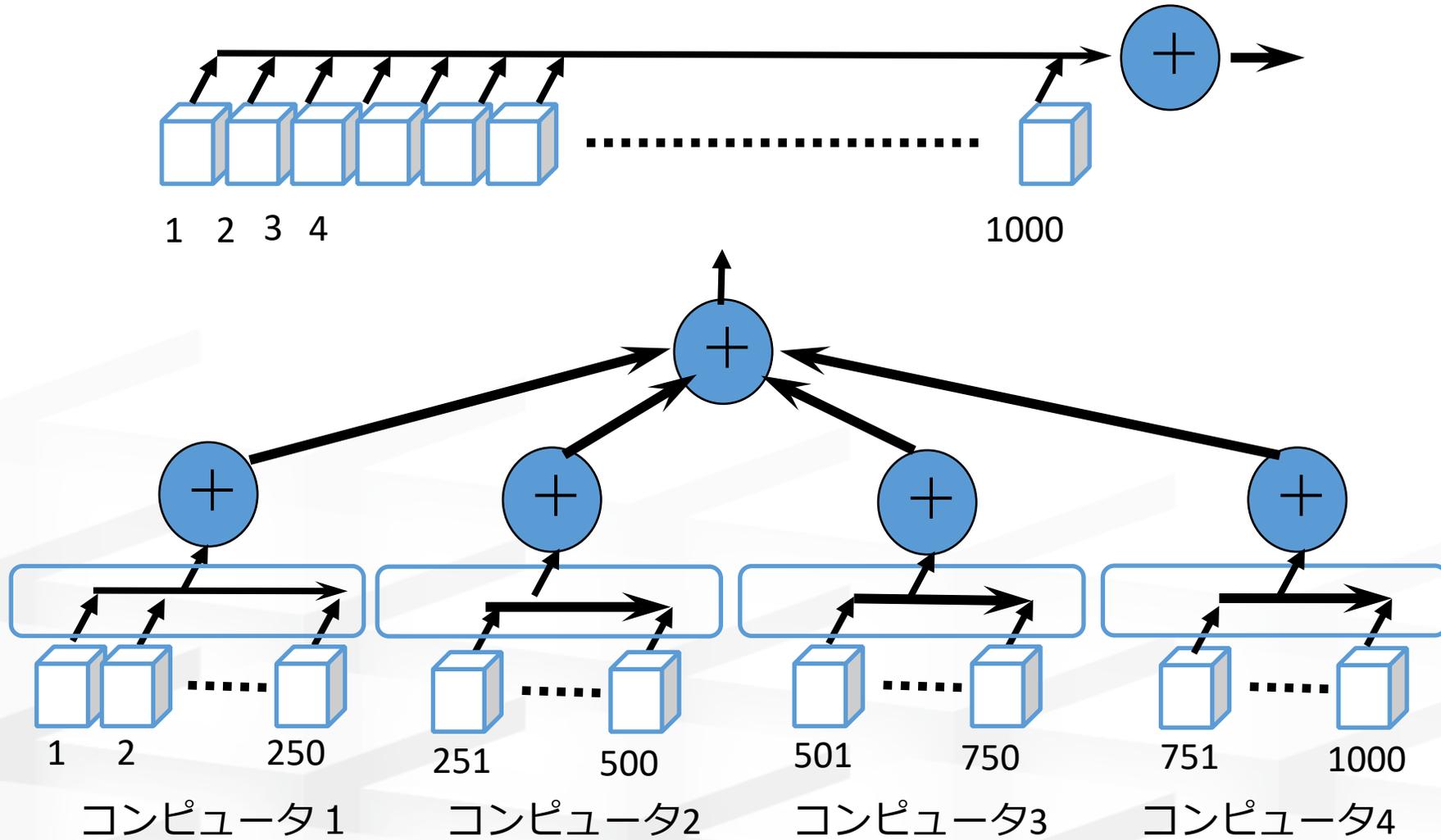
- 並列コンピュータ
- 今のスパコンの主流はこれ！
- PCでも、スマホでも2, 3個のコンピュータ (コア) がはいつている



富岳のプロセッサ
A64FXは、48個のコンピュータ (コア)
が内蔵されている

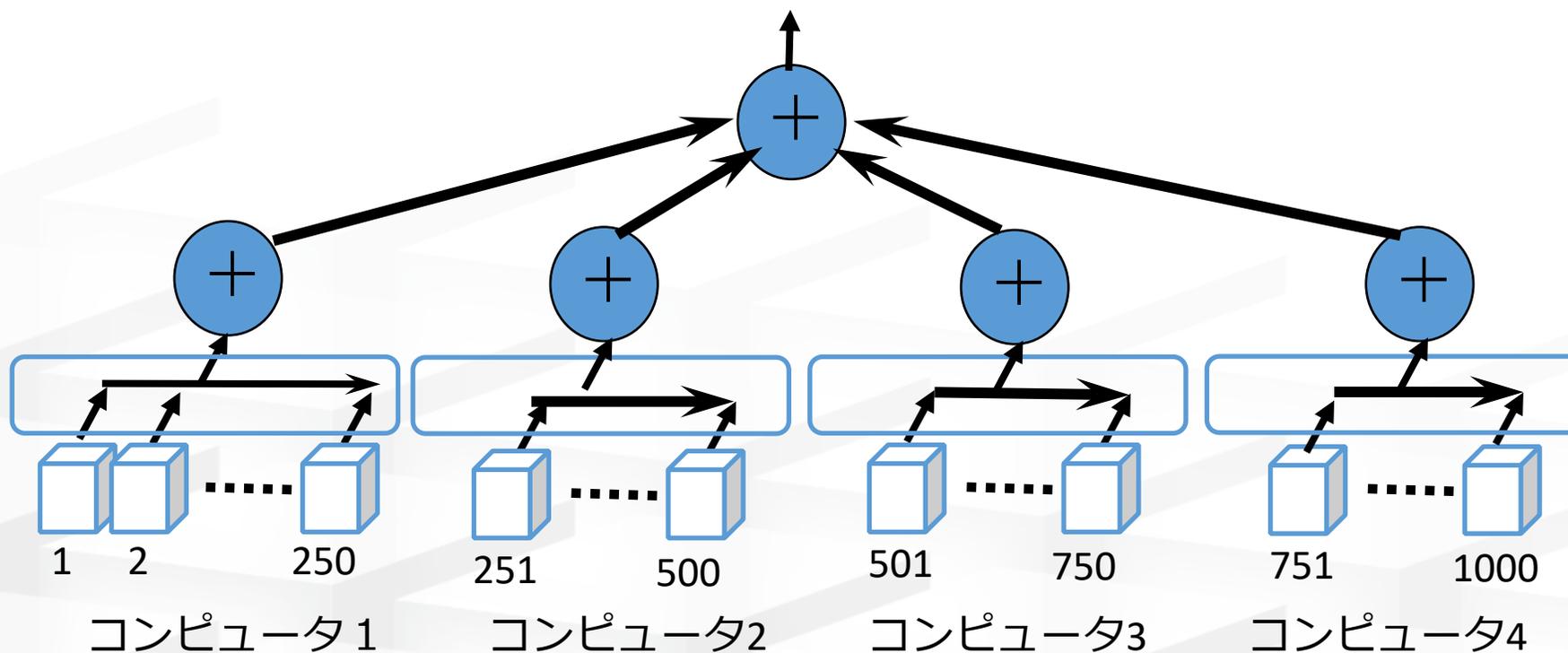
並列計算とは？ なぜ、速くなるのか？

4つのコンピュータで1000個の数を足す場合...



4つのコンピュータで1000個の数を足す場合…

- 1つの数を取り出して、加算するのに、1秒とすると、1000個の数だと、1000秒かかる。
- 4つのコンピュータを使えば、それぞれのコンピュータで250個の数字をたせばいいので、250秒 + α ができる。大体、4倍！



スパコンはコンピュータをたくさん繋げたもの！



並列計算機



「京」から「富岳」へ



「京」から「富岳」への進歩

	京	富岳	比
ノードのコア数	8	48	
半導体技術 (nm)	45	7	
コア当たりの性能(GFLOPS)	16	64(70)	4(4.4)
ノード当たりの性能(TFLOPS) 倍精度(64bits)	0.128	3.072 (3.379)	24 (26.4)
ノード当たりの性能(TFLOPS) 単精度(32bits)	0.128	6.144 (6.758)	48 (52.8)
ノードのメモリバンド幅(GB/s)	64	1024	16
ラック当たりのノード数	96	384	4
ラックの性能(TFLOPS)	12.3	1036.8	84
システム全体のノード数	82,944	158,976	1.9
コア数	663,552	7,630,848	11.5
システム性能 (PFLOPS) 上：倍精度、下：単精度	10.6	488 (537) 977(1070)	42.3(52.2) 84.6(104.4)

カッコ内は、ブースト時

ポスト「京」の性能について

ポスト「京」の開発目標

- ・最大で「京」の100倍のアプリケーション実効性能※1
- ・消費電力 30~40MW (「京」は12.7MW)

ポスト「京」のターゲットアプリケーション※4実効性能

(数値は、「京」の性能との比較)

「京」とポスト「京」の性能比較

	ポスト「京」※2	「京」
理論演算性能	400 PFlops以上 (対「京」比:約34倍以上)	11.3 Pflops
総メモリバンド幅 ※3	150 PB/sec以上 (対「京」比:約29倍以上)	5,184TB/sec

- ※1 ハードウェアの性能向上とアプリケーションのアルゴリズムの改良効果を合わせて演算性能を比較するもの。
- ※2 ポスト「京」に搭載されるCPUの性能(理論演算性能2.7 TFlops以上、メモリバンド幅1,024GB/sec)、搭載数(15万個以上)から推定。
- ※3 単位時間当たりどれだけのデータをメモリからCPUに転送できるかの値。

(参考)

- ※4 ポスト「京」では、5分野から9つの主たるターゲットアプリケーションを選定。
- ※5 総合科学技術・イノベーション会議 評価専門調査会 第2回評価検討会(平成26年10月28日)の資料より抜粋。
- ※6 Genomon以外試作機での測定値を元に推計。試作機1ノード(1CPU)を使ってアプリケーションの一部を実行した時間から推定。
- ※7 CSTI報告時に想定していたアプリケーションのバージョンが更新され、問題設定が変更されているため比較できないが、1日あたりのゲノム情報解析の検体数は2,000検体以上であり目標(1,000検体以上)をクリアしている。

分野	重点課題	2014年時点の目標性能※5	現時点の性能見込み※6	想定プログラム
社会の健康長寿の実現	①生体分子システムの機能制御による革新的創薬基盤の構築	100倍	<u>125倍以上</u>	GENESIS
	②個別化・予防医療を支援する統合計算生命科学	- ※7	8倍以上	Genomon
防災・環境問題	③地震・津波による複合災害の統合的予測システムの構築	15倍	45倍以上	GAMERA
	④観測ビッグデータを活用した気象と地球環境の予測の高度化	75倍	<u>120倍以上</u>	NICAM+LETKF
エネルギー問題	⑤エネルギーの高効率な創出、変換・貯蔵、利用の新規基盤技術の開発	40倍	40倍以上	NTChem
	⑥革新的グリーンエネルギーシステムの実用化	15倍	35倍以上	Adventure
産業競争力の強化	⑦次世代の産業を支える新機能デバイス・高性能材料の創成	35倍	30倍以上	RSDFT
	⑧近未来型ものづくりを先導する革新的設計・製造プロセスの開発	20倍	25倍以上	FFB
基礎科学の発展	⑨宇宙の基本法則と進化の解明	50倍	25倍以上	LQCD

コンピュータ(スパコン) の性能は どうやって測るのか

そもそも、コンピュータの性能はどうやって測るの？

- **演算能力はどうやって「表す」のか**

- **1秒あたりの（理論）演算可能回数**

- 数値は「浮動小数点」という形式で格納されている。
- MFLOPS: Millions of Floating Point OperationS. (1秒間に 10^6 回の浮動小数点処理)
- GFLOPS : 10^9 回, TFLOPS : 10^{12} 回, PFLOPS : 10^{15} 回、EFLOPS …
- 理論ピーク性能 = 常に演算器が使えた場合の性能
 - 演算器の数 x クロックあたりの演算回数

これだと
つながっていない
てもいい？

- **並列コンピュータの場合は**

一台の1秒あたりの演算可能回数 x 台数

- では、演算能力はどうやって「測る」のか
- **Top500**
- **Green500**
- **HPCG**
- **Graph500**
- 他にもいろいろあるけど、…
 - IO, AI/DL, SPEC ….

Top 500

世界1位とか
というのはこれ！

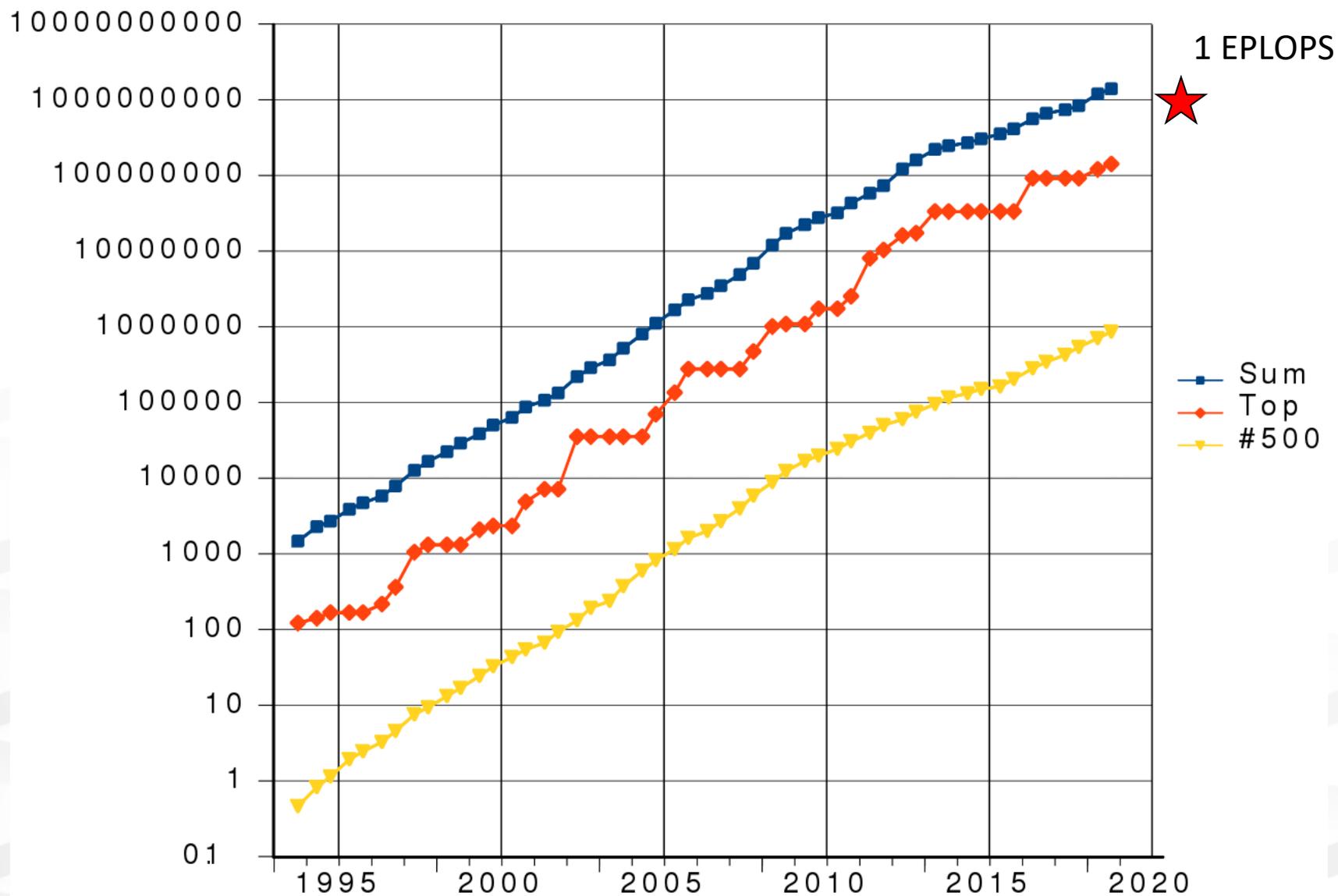


- LINPACK(密行列の直接解法)と言われるベンチマークプログラムの性能を計測するベンチマーク
- Top500リストは、LINPACKの実行性能を指標として世界で最も高速なコンピュータシステムの上位500位までを定期的にランク付けし、評価するプロジェクト。1993年に発足し、スーパーコンピュータのランキングを年2回(6月、11月)発表している。
- 数十年のデータがあるため、スパコン性能の予測にも使われている。<http://www.top500.org/>
- 多くの科学技術計算や産業アプリケーションで使用される浮動小数点数の演算能力を測ることを目的としている。
- 本ベンチマークで高い性能を出すためには、搭載メモリ限界の問題規模で測定する必要があり、数時間にわたる実行が不可欠。高いスコアは計算能力と同時にシステムが故障なく長時間稼働できることも意味している。
- 近年、実際のアプリケーションの性能との乖離が指摘されている。

Rank	Site	Computer/Year Vendor	Cores	R _{max}	R _{peak}	Power
1	RIKEN Advanced Institute for Computational Science (AICS) Japan	K computer, SPARC64 VIIIfx 2.0GHz, Tofu interconnect / 2011 Fujitsu	548352	8162.00	8773.63	9898.56
2	National Supercomputing Center in Tianjin China	Tianhe-1A - NUDT TH MPP, X5670 2.93GHz 6C, NVIDIA GPU, FT-1000 8C / 2010 NUDT	186368	2566.00	4701.00	4040.00
3	DOE/SC/Oak Ridge National Laboratory United States	Jaguar - Cray XT5-HE Opteron 6-core 2.6 GHz / 2009 Cray Inc.	224162	1759.00	2331.00	6950.60
4	National Supercomputing Centre in Shenzhen (NSCS) China	Nebulae - Dawning TC3600 Blade, Intel X5650, NVidia Tesla C2050 GPU / 2010 Dawning	120640	1271.00	2984.30	2580.00
5	GSIC Center, Tokyo Institute of Technology Japan	TSUBAME 2.0 - HP Xeon 6C X5670, Nv Linux/Windows / 2011 NEC/HP				
6	DOE/NNSA/LANL/SNL United States	Cielo - Cray XE6 8-c Cray Inc.				
7	NASA/Ames Research Center/NAS United States	Pleiades - SGI Altix Xeon HT QC 3.0/Xe Infiniband / 2011 SGI				
8	DOE/SC/LBNL/NERSC United States	Hopper - Cray XE6 Cray Inc.				
9	Commissariat a l'Energie Atomique (CEA) France	Tera-100 - Bull bullx S6010/S6030 / 2010 Bull SA				
10	DOE/NNSA/LANL United States	Roadrunner - Blade Cluster, PowerXCell 1.8 GHz, Voltaire Inf IBM				



これからのスパコンの性能が予測できる？



最先端のスパコンを作る時の問題は、…

- **いまのスパコンの性能は、並列処理から**
 - つまり、性能 = プロセッサの性能 × 台数
- **すなわち、コンピュータ数**
- **ということは、性能は結合するコンピュータの数を増やせばいい**
⇒ **が、電力が限界**
- **2008年から、Top500に電力消費量を表示するようになった**
 - これからのスパコンは
電力が大切

たとえば、現在のデスクトップPCの電力は、**100W**で、**100GFLOPS** (1GF/W) これで、並列システムをつくと…

1TF = 1000W

10PF = 10MW (京なみ)

1EF = 1000MW

- Top500リストにランキングされているシステムのうちで、電力当たりの性能を競うベンチマーク
- 500位以内に入るためには、数PFLOPSのシステムでなければならない。
- 富岳のプロトタイプで、2019年11月に 16.876GF/Wで 1位を獲得！汎用プロセッサながら、GPUを越える電力性能を実証。

Green500, Nov. 2019

A64FX prototype –
Fujitsu A64FX 48C 2GHz
ranked **#1** on the list

768x general purpose A64FX
CPU w/o accelerators

- 1.9995 PFLOPS @ HPL, 84.75%
- 16.876 GF/W
- Power quality level 2

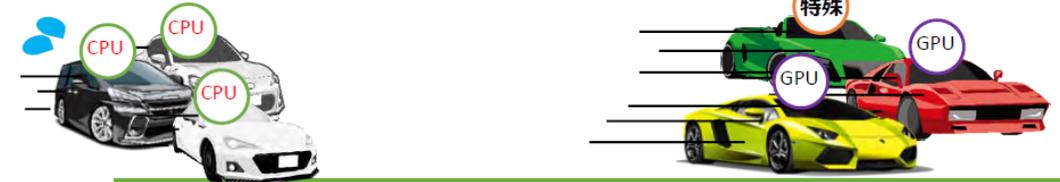
Rank	System	Cores	Max (TFlops)	Power (kW)	Power Efficiency (GFlops/watt)
1	A64FX prototype – Fujitsu A64FX, Fujitsu A64FX 48C 2GHz, 768x general purpose A64FX CPU w/o accelerators	36,864	1,999.5	118	16.876
2	NA-1 – Dell/Oracle/2.0 Xeon D, 1371 14C 1.30GHz, Intel/land, NVIDIA M100, IBM PowerPC/IBM, PEZY Computing / Executer Inc. PEZY Computing K.K. Japan	1,271,640	1,333.2	60	16.254
3	AMOS – IBM Power System AC920, IBM POWER9 20C 3.45GHz, Dual-raid Mellanox EDR Infiniband, NVIDIA M100, IBM PowerPC/IBM, PEZY Computing / Executer Inc. PEZY Computing K.K. Japan	130,000	8,045.0	510	15.771
4	Saber – IBM Power System AC920, IBM POWER9 20C 3.45GHz, Intel/land EDR, Mellanox Tesla P100 SXM2, IBM PowerPC/IBM, PEZY Computing / Executer Inc. PEZY Computing K.K. Japan	23,262	1,444.0	94	15.374
5	Summit – IBM Power System AC920, IBM POWER9 20C 3.45GHz, NVIDIA Tesla V100 SXM2, Dual-raid Mellanox EDR, Intel/land, IBM PowerPC/IBM, PEZY Computing / Executer Inc. PEZY Computing K.K. Japan	2,414,792	148,400.0	10,794	14.719

FUJITSU CONFIDENTIAL

Copyright 2019 FUJITSU LIMITED

Green500試作機1位獲得。何がすごいのか？

● これまでのGreen500



GPU搭載スパコンや特殊スパコンが上位を独占

● 今回のGreen500

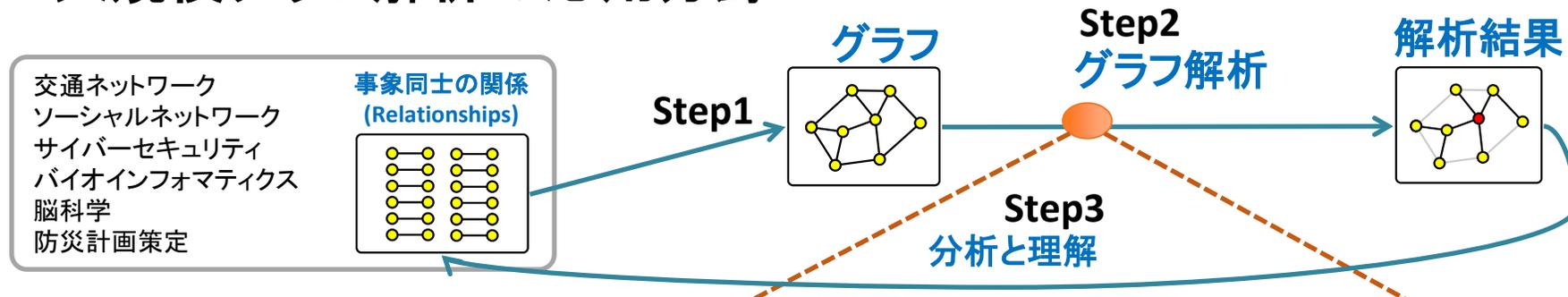


初めて汎用CPU搭載スパコンが世界1位に！

- 疎な係数行列から構成される連立一次方程式を解く計算手法である共役勾配法 (conjugate gradient method) を用いた新たなベンチマーク・プログラム
 - HPCG = High Performance Conjugate Gradient solver
- 実際のアプリケーションで求められる性能要件との乖離から、**産業利用など実際のアプリケーションでよく使われるCG法のプログラムで性能を評価するために提案された。**
 - 自動車や飛行機の空力設計、構造計算、…
- **演算性能だけでなく、メモリアクセス性能がベンチマークの結果を大きく左右する。**
- 2014年6月のISC14で世界の主要なスーパーコンピュータ15システムでの測定結果の発表を経て、同年11月に米国ニューオーリンズで開催されたHPCに関する国際会議SC14から正式なランキングが発表されている。
 - 京は、Nov. 2014 – Nov. 2015で、1位

- 超大規模グラフの探索能力で計算機を評価するベンチマーク
 - 実社会における複雑な現象は、大規模なグラフ（頂点と枝によりデータ間の関連性を示したもの）として表現される場合が多いため、コンピュータによる高速なグラフ解析が必要とされる
 - 2つの性能指標がある。「京」「富岳」では、主にBFSを評価。
 - **BFS(Breadth First Search):** スタート頂点に繋がっている全ての辺を調べて、繋がっている全ての頂点を知り、次はそれらの頂点から繋がっている全ての辺を調べるという風にして、グラフの全ての頂点に到達する経路を見つけるまでの計算時間を測る。辺の総数を処理時間で割ってTEPS(Traversed Edges Per Second)を求める
 - **SSSP(Single Source Shortest Path):** スタート頂点から、全ての頂点への最短の経路を求めるというベンチマーク。なお、SSSPではそれぞれの辺は長さを持っており、経路に含まれる辺の長さの合計が経路の長さとなる。
 - 演算能力だけでなく、メモリ性能、ネットワーク性能が重要！
 - 「京」はJun 2014 1位、Nov 2014 2位、Jun 2015-Jun 2019 1位。
 - Green Graph500もある。グラフ処理の場合の省電力性能は TEPS / Watt (電力性能比)を計測。

大規模グラフ解析の応用分野



- 並列グラフ探索 (幅優先探索 : **BFS**)
- 最適化 (最短路 : **SSSP**, 最大フロー, 最小費用フロー)
- クラスタリング (グラフ分割, コミュニティ抽出)

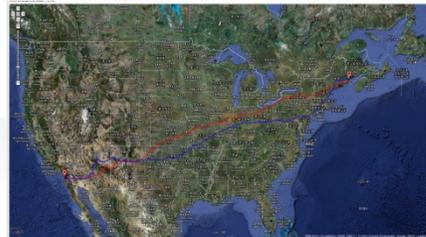


ソーシャルネットワーク

6,160万点 & 14億7千万枝

全米道路ネットワーク

2,400万点 & 5800万枝



サイバーセキュリティ

150億/日 のアクセスログ



ニューラル・ネットワーク

890億点 & 100 兆枝

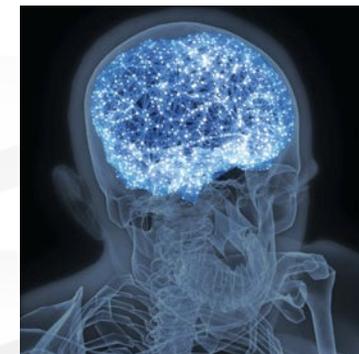


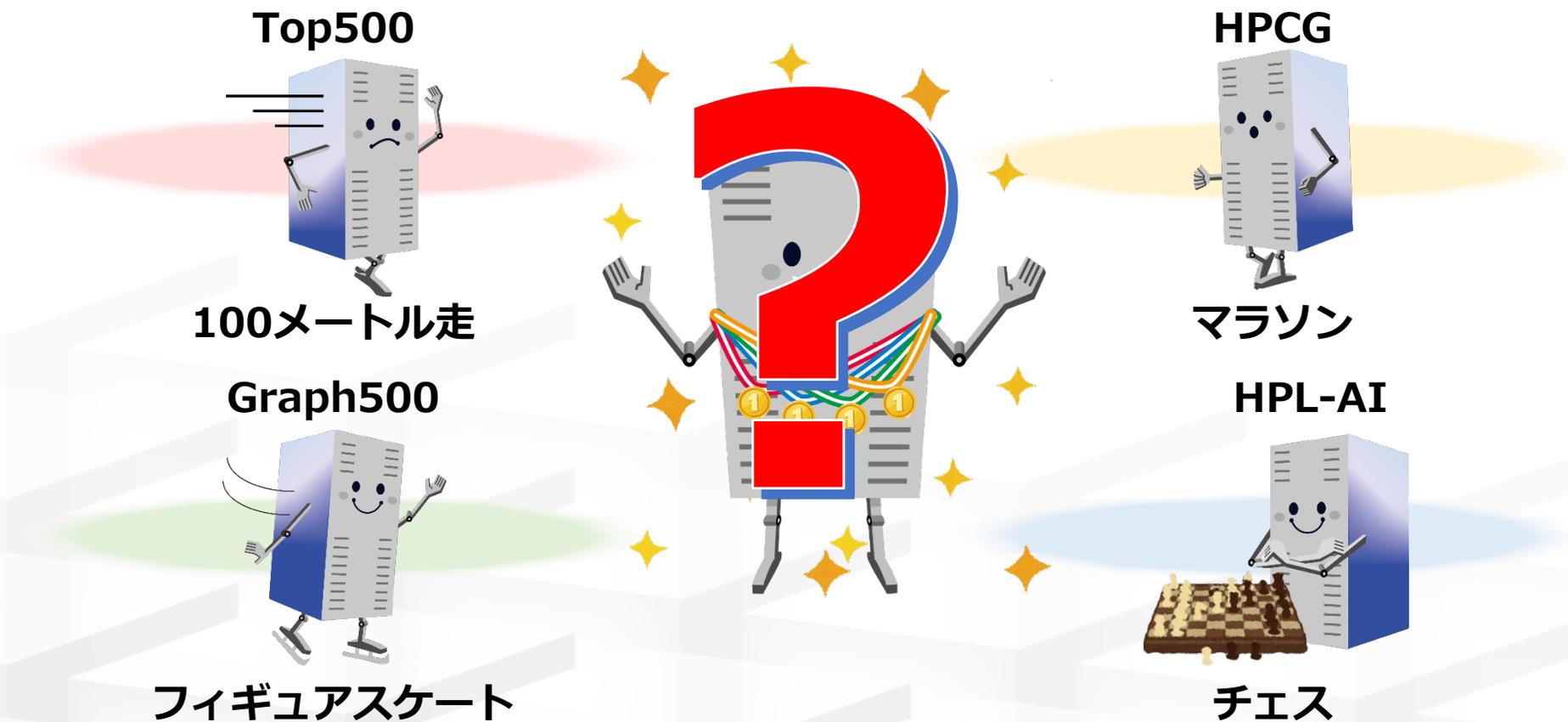
Image: Illustration by Mirko Ilic

グラフ解析の利用方法と応用分野

- **低精度計算を用いることを認めたHPL** (High-Performance LINPACK) の性能を計測するベンチマーク
 - 倍精度計算よりも計算精度が劣ってしまうため、引き続き反復改良と呼ばれる技術で倍精度計算と同等の精度にすることを求めている。
 - 従来のHPLは、倍精度演算 (10進で16桁の浮動小数点数) のみで計算することがルールに定められていた。
- 近年、GPUや人工知能向けの専用チップで低精度演算 (10進で5桁、もしくは10桁) の演算器を多数搭載し、高性能化をする計算機が多数現れており、ディープラーニングなどのAI処理に使われているが、これらの性能演算性能がTop500に反映されない問題から、提案された。
- **間接的に、低精度演算での演算能力を評価することで、ディープラーニングなどのAI処理の性能を評価することを目的とする。**
- HPL-AIは2019年11月にルールが公表されたため、今回が初めてのベンチマークランキングの発表となる。

「富岳」は様々なランキングで世界トップレベル

- それぞれのランキングは評価する性能が全く異なる。
評価する性能を競技で例えると…



様々な用途・計算方法で世界トップレベルの性能
汎用性の高い万能スーパーコンピュータ

おわりに：ベンチマークとスパコンのトレンド

- ベンチマークの結果だけでなく、実際のアプリでの性能が重要！
- Top500から、近年のスパコンの進歩の停滞が指摘されている。
 - 性能の伸びが、2012年あたりから、これまでの年率1.9倍から1.2倍に。
- 一方で、Top500は主に演算性能のみで、実際のアプリの性能を反映していないという意見から、HPCGやGraph500での評価にも興味が集まっている。
- システムの性能の伸びに比べ、プロセッサの性能は伸びていない。
 - 性能は、プロセッサの個数の増加（大規模化）、アクセラレータ（メニ・コアを含む）による。
 - プロセッサ自体の性能は伸びが鈍っている。ムーアの法則の終焉・スローダウン
- 大規模化により、電力性能の重要性が顕著になっている
- 2021年-2022年には、USのエクサスパコンが登場か。
 - USは、GPUを使ったシステムとなる見込み。

ご清聴 ありがとうございました。