

# 大規模並列環境での機械学習処理とその応用 ～「富岳」におけるMLPerf HPCや津波AIへの応用～

2021.9.2

富士通株式会社  
福本 尚人

## ■ 福本 尚人

- 富士通(株) 研究本部 ICTシステム研究所 先端コンピューティングPJ  
プロジェクトマネージャー
- 高性能計算機のアプリ高速化に関する研究開発のマネジメント担当



## ■ 略歴

2006年九州大学工学部電気情報工学科、2008年同大学大学院修士修了、2012年同大学博士後期課程修了  
同年富士通(株)に入社、以来、PCクラスタ高速化技術、並列計算環境高速化技術の研究開発に従事。  
博士(工学)。  
平成30年度文部科学省科学技術賞開発部門「大規模PCクラスタ構築技術の開発」を他4名と受賞。

## ■ 関わったスーパーコンピュータ

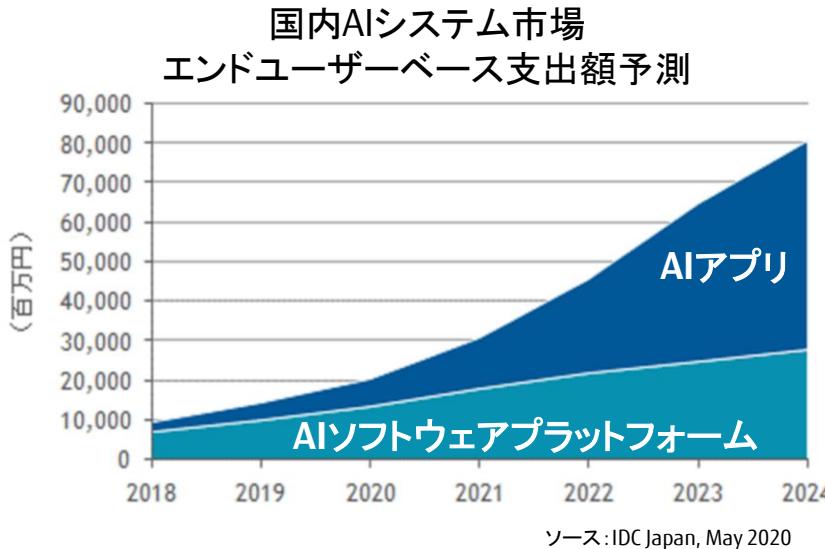
- 東大・筑波大(JCAHPC) Oakforest-PACS、産総研 ABCI: HPL高速化と計測
- 九州大学、名古屋大学、北海道大学、理化学研究所のPCクラスタ: HPL高速化と計測
- 「富岳」: AIソフトスタックの開発

# AIを中心とした スーパーコンピューティング技術の応用例

- 大規模AI計算の高速化
- A64FXシステム向けAI環境
- MLPerf HPCベンチマークの高速化
- AIを活用した津波予測

# 大規模AI計算の高速化

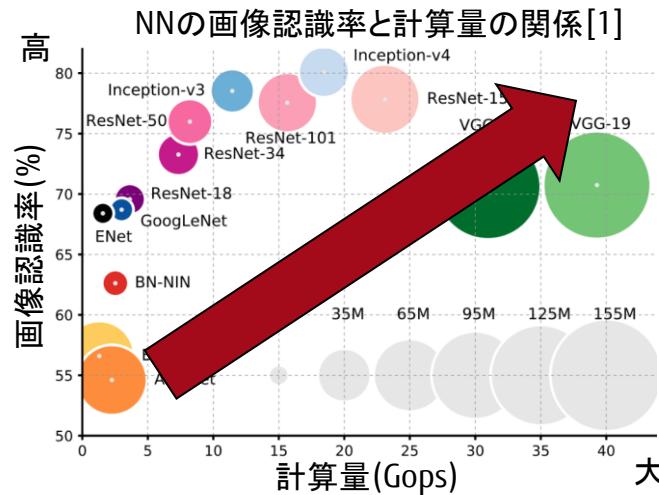
- AI技術を組み込んだアプリが急速に増加
- 様々な分野の処理にニューラルネットワーク(NN)が組み込まれる



# NN(ニューラルネットワーク)の進化と計算量の増加

FUJITSU

- 日進月歩でより精度が高いNNが提案される
- 精度が高いNNほど、計算量が増加する傾向がある



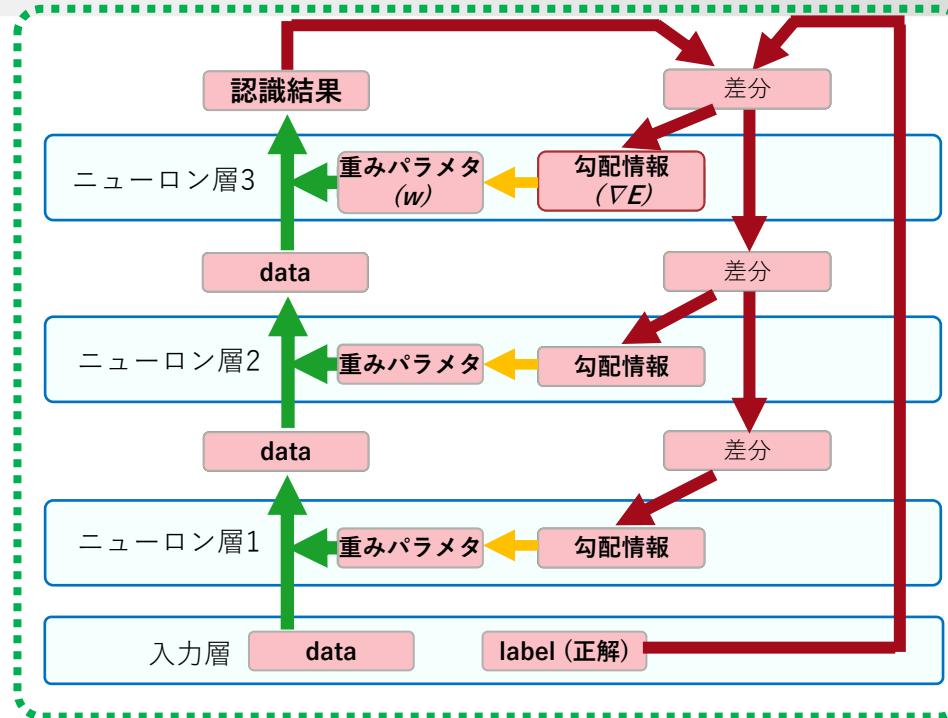
[1] A. Canziani, et al, "An Analysis of Deep Neural Network Models for Practical Applications", arxiv, 2016

# ディープラーニングの学習処理

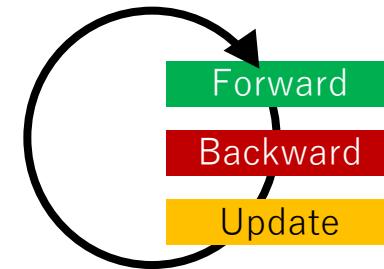
FUJITSU



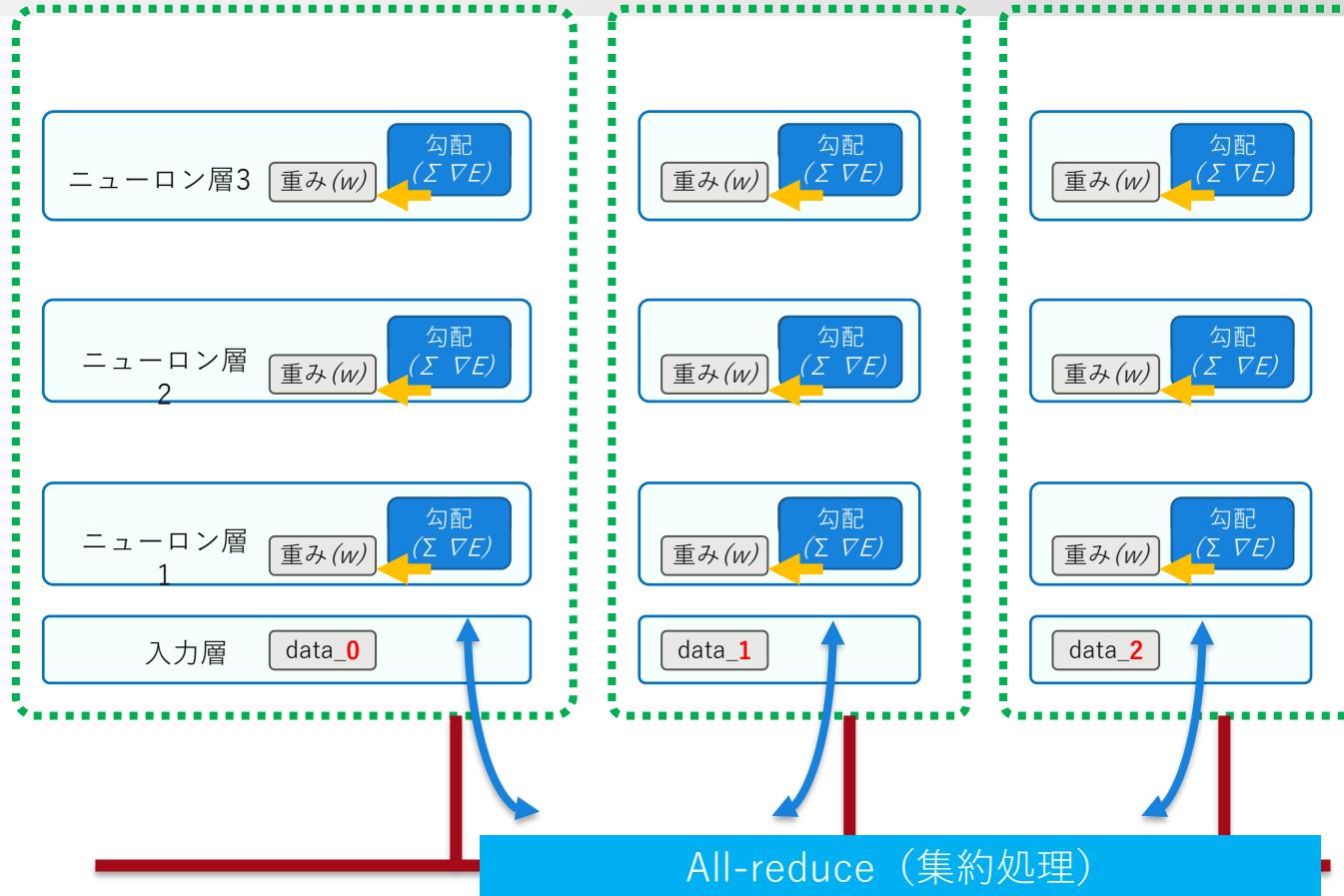
# ディープラーニングの学習処理



学習処理サイクル



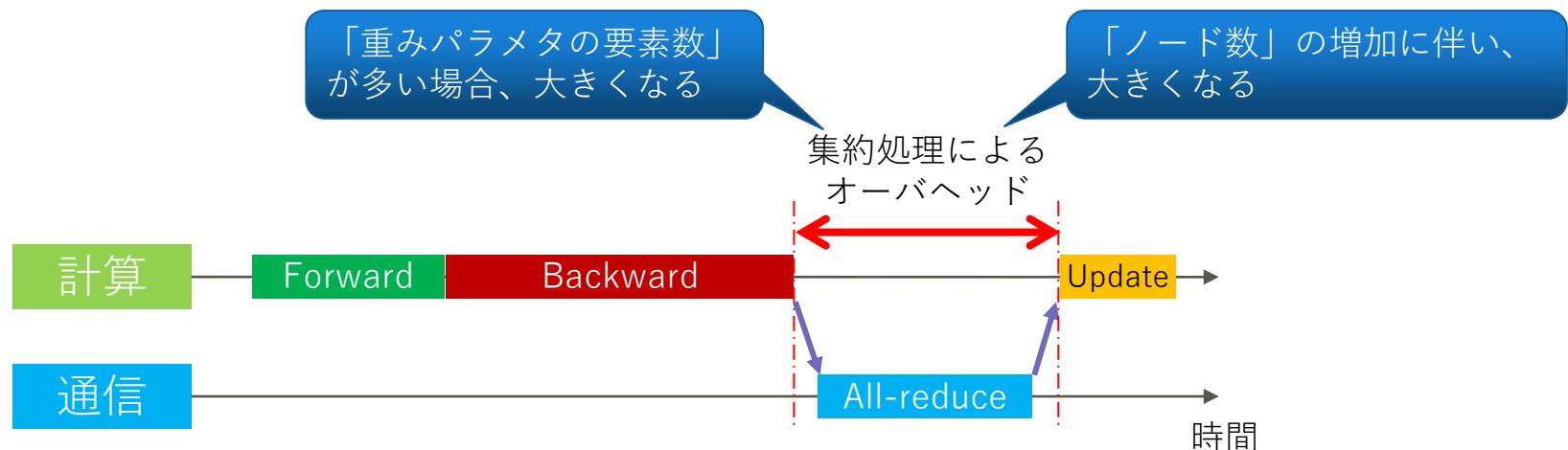
# 並列環境でのディープラーニングの学習処理



# 並列環境でのディープラーニングの学習処理の課題

FUJITSU

## ■ 通信している間、計算が止まるため、その分遅くなる



基本的な  
アイディア

- 通信時間を他の計算時間に隠蔽
- 通信時間を短縮

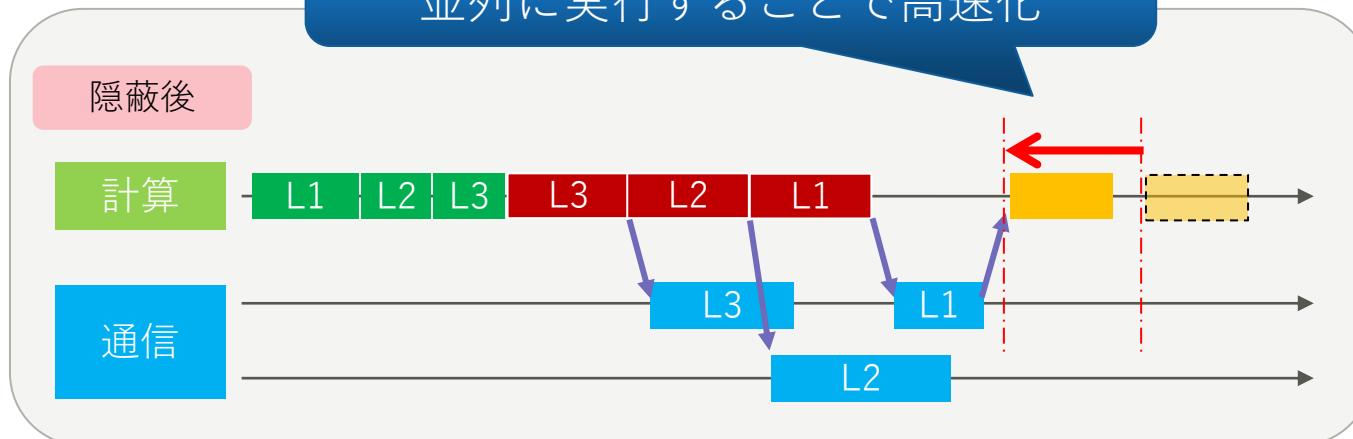
# (1) Backward処理時間への隠蔽

方法

各層のBackward処理が終わる毎に  
**層単位でAll-reduce処理**を開始する

各層のForward処理
各層のBackward処理
All-reduce処理
Update処理

Backward処理とAll-reduce処理を  
並列に実行することで高速化



## (2) Forward処理時間への隠蔽

方法

- Update処理を**分割**
- Forward処理の開始を**層毎**に判定

各層のForward処理
各層のBackward処理
All-reduce処理
Update処理

すべての層のAll-reduce処理の完了前に、  
次のForward処理を開始することで高速化

層単位でUpdate処理

計算

L3

L2

L1

L1

L2

通信

L3

L1

L2

### (3) 細分化による並列化

#### 方法

- **集約処理**を細分化して実行

- GPUからCPUメモリへのデータ転送
- ノード間のデータ転送
- Reduce演算
- CPUからGPUメモリへのデータ転送

各層のForward処理
各層のBackward処理
All-reduce処理
Update処理

All-reduceにかかる時間を短縮  
することで高速化

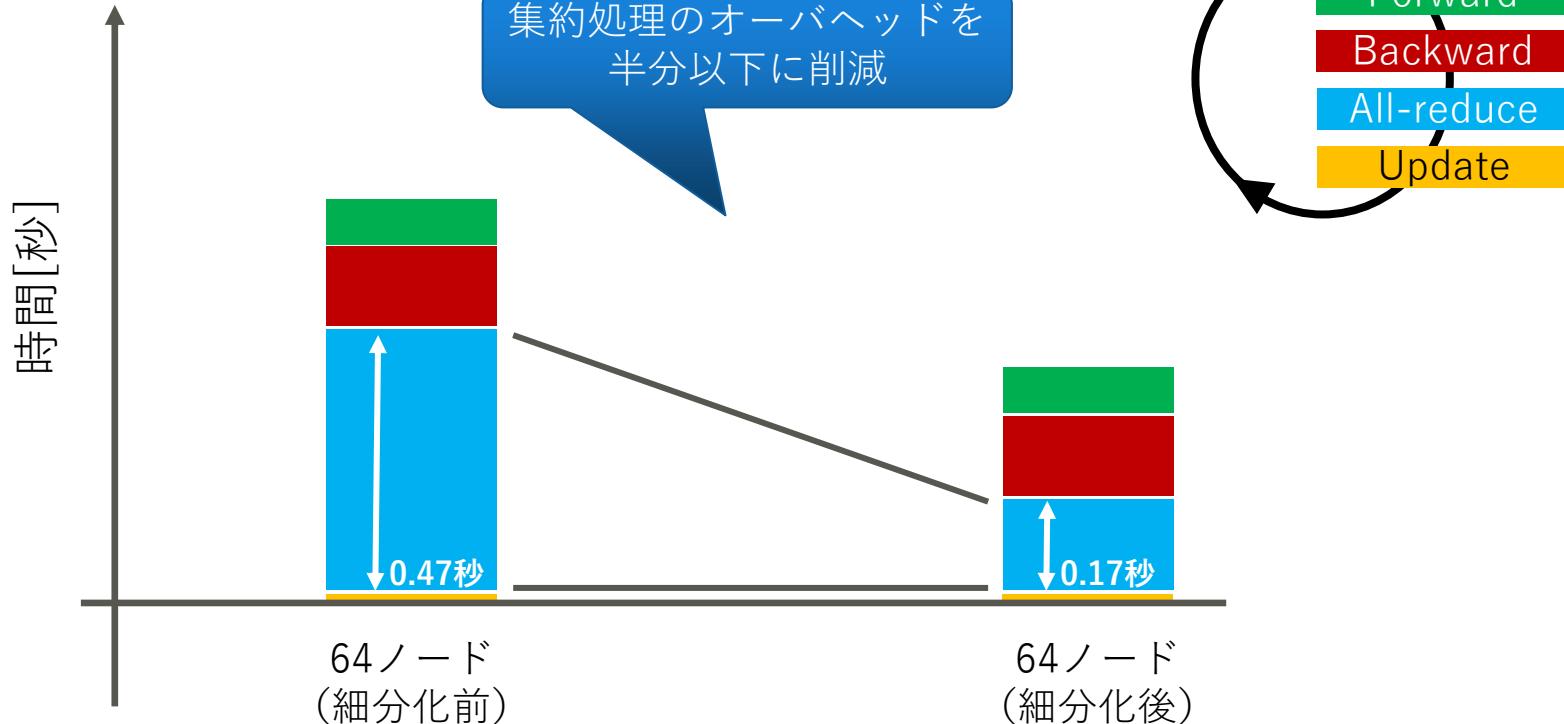
細分化後

計算

通信



# 集約処理隠蔽の効果



# 高速化の効果

- これらの高速化を突き詰めることでGPU 1基で数日かかる計算を約75秒で解く

	プロセッサ	実行時間
	Tesla P100 x 1	6.13日
Goyalら	Tesla P100 x 256	1 時間
Smithら	Full TPU Pod	30 分
秋葉ら	Tesla P100 x 1,024	15 分
Jiaら	Tesla P40 x 2,048	6.6 分
Yingら	TPU v3 x 1,024	1.8 分
三上ら	Tesla V100 x 3,456	2 分
富士通研	Tesla V100 x 2,048	1.2 分

# A64FXシステム向けの AIフレームワーク環境

# 富岳におけるAI処理の優位性

FUJITSU

## ■ 世界一の演算性能と通信性能を活かせる

### CPU単体の性能

	富岳 A64FX	Intel CPU Xeon Platinum 8280 28C 2.7GHz
コア数 (計算単位の数)	50 個	28個 1.7倍
計算速度 (倍精度演算性能)	3,379 Gflops	2,111 Gflops 1.6倍
省電力性	14.6 Gflops/W	5.84 Gflops/w 2.5倍

AIに重要な計算速度と省電力性で  
Intel CPUに対して大きな優位性

### スパコン全体の性能 Fugaku



### Summit



148 Pflops

442 Pflops

2位のSummitに対して2.9倍の演算性能  
さらに、AIに重要なサーバ間集合通信が高速

NNモデルのパラメータや学習データを探索する必要がある  
AIアプリ開発においてコンピュータの性能は非常に重要

[https://ja.wikipedia.org/wiki/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB:Summit\\_\(supercomputer\).jpg](https://ja.wikipedia.org/wiki/%E3%83%95%E3%82%A1%E3%82%A4%E3%83%AB:Summit_(supercomputer).jpg)

## ■ 業界標準のAIソフトスタックが利用可能

- TensorFlowやPyTorchといった著名なFrameworkをサポート
- AI向けの演算ライブラリは業界標準のOneDNNに富岳向け実装がある
- 富岳のAIソフトスタックはOpen Source Software (OSS)として公開

DL framework



演算ライブラリ (OneDNN)

Fugaku, FX1000

Intel server



## ■ 今後のAIソフトスタックの拡張

- 業界標準のソフトウェアをOSSコミュニティと連携して拡張
- 最新環境への更新が継続される

# MLPerf HPCベンチマークの高速化

## ■ MLPerf HPC

- HPCシミュレーションに関連する深層学習タスク
- 20年11月に初版(v0.7)がリリース

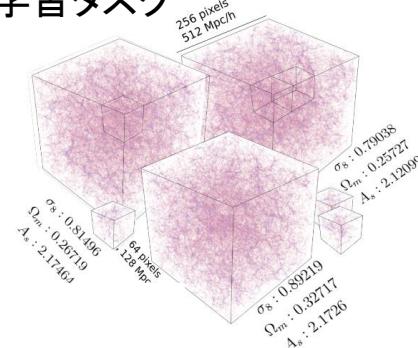
<https://mlcommons.org/en/news/mlperf-hpc-v07/>

## ■ ベンチマーク指標

- 一定精度に到達するまでの実行時間

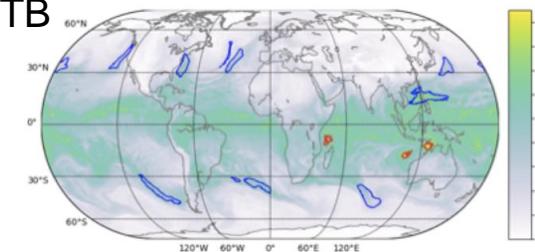
## ■ CosmoFlow

- 宇宙論的パラメータを、N体問題シミュレーションデータから推測する深層学習タスク
- データサイズ: 5.1TB



## ■ Deep CAM

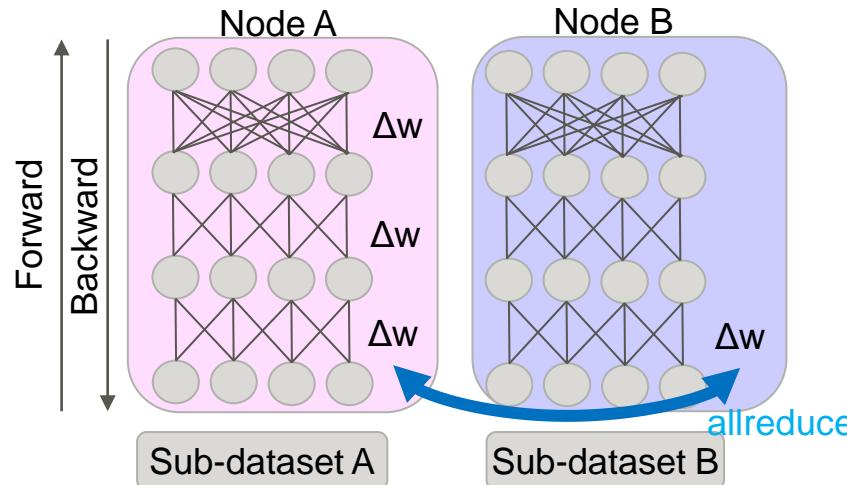
- 大規模な気候シミュレーション結果から極端な気象現象(atmospheric riverやサイクロン)を特定する深層学習タスク
- データサイズ: 8.8TB



シミュレーション結果を分析する大規模深層学習タスク

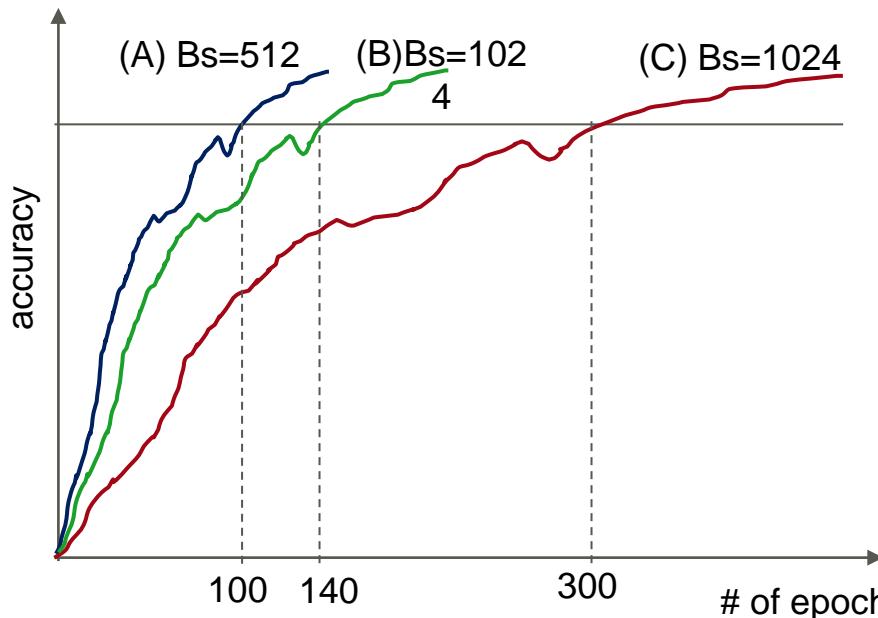
## ■ データ並列

- 通信コストは低い
- ミニバッチサイズが並列度の上限



## ■ 精度を維持したミニバッチサイズの増加

- MLPerf HPC: ある一定精度に到達するまで時間が性能指標
- バッチサイズ増加に伴うエポック数の増加は性能低下に直結



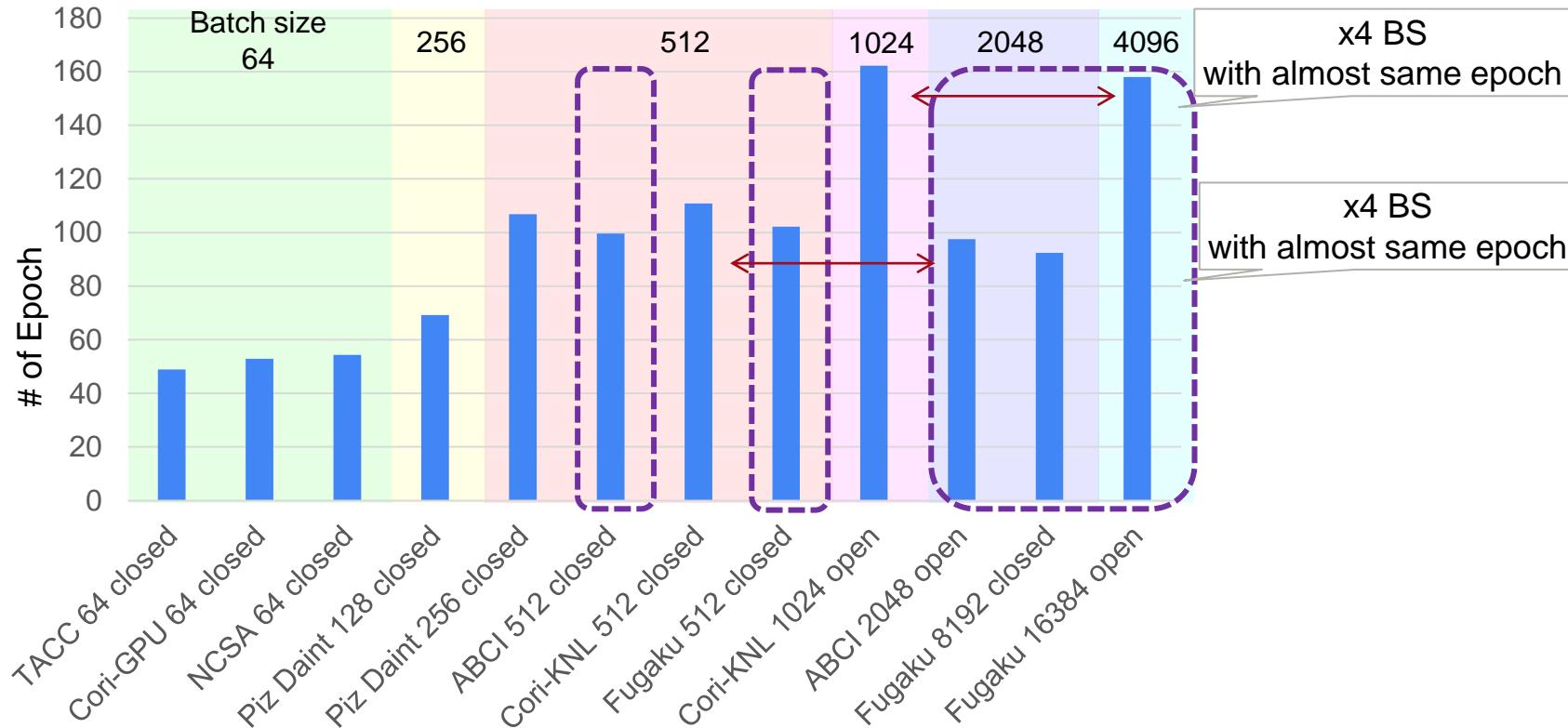
Case	BS	epoch	Elapsed (rough est.)
(A)	512	100	x1.4 100
(B)	1024	140	x3.0 70 (100*1.4 / 2)
(C)	1024	300	150 (100*3.0 / 2)

Throughput: x2, by x2 computing resources

ハイパーパラメータチューニングによるバッチサイズと精度の両立が大規模化では必須

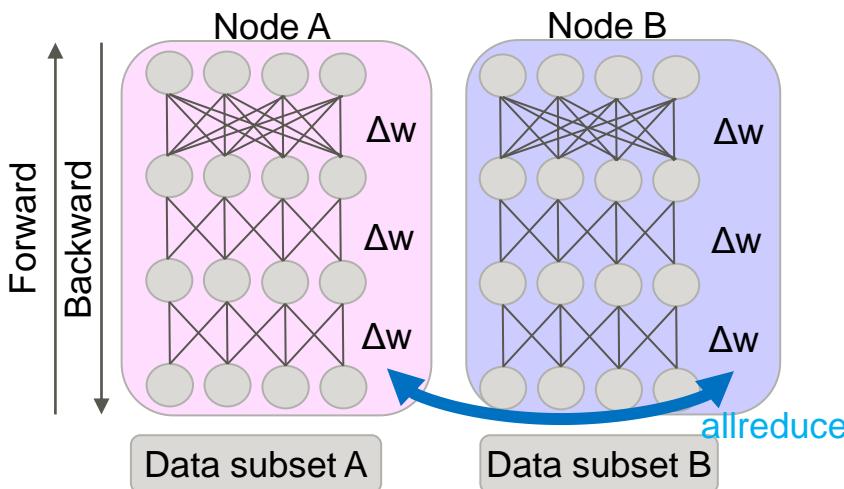
# ミニバッチサイズとエポック数

- ミニバッチサイズを増加させるとエポック数も増大する傾向に
- ABCIや「富岳」での測定では、大きなバッチサイズでもエポック数増大を抑止



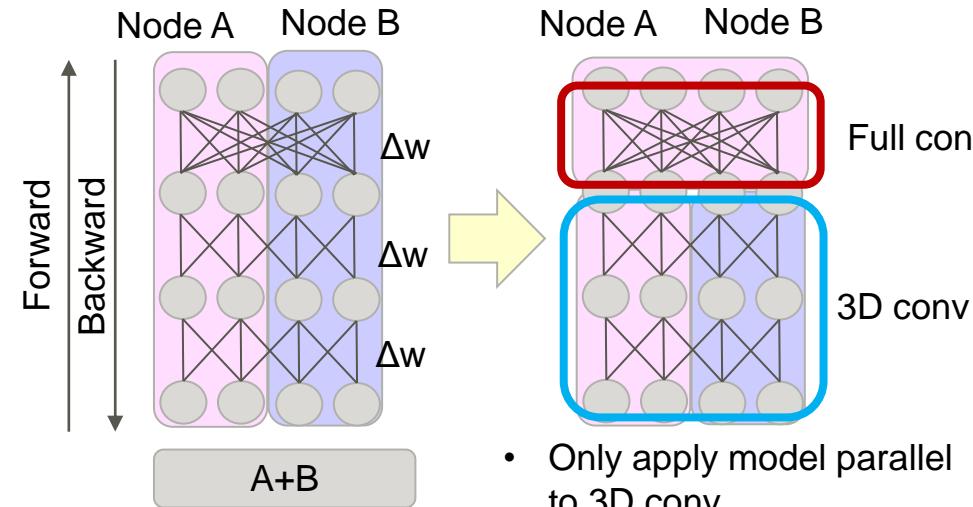
## ■ データ並列

- Allreduceによる勾配情報の交換のみ
- 性能向上は比較的容易



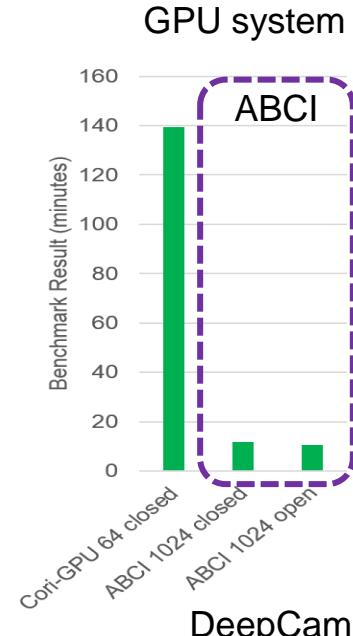
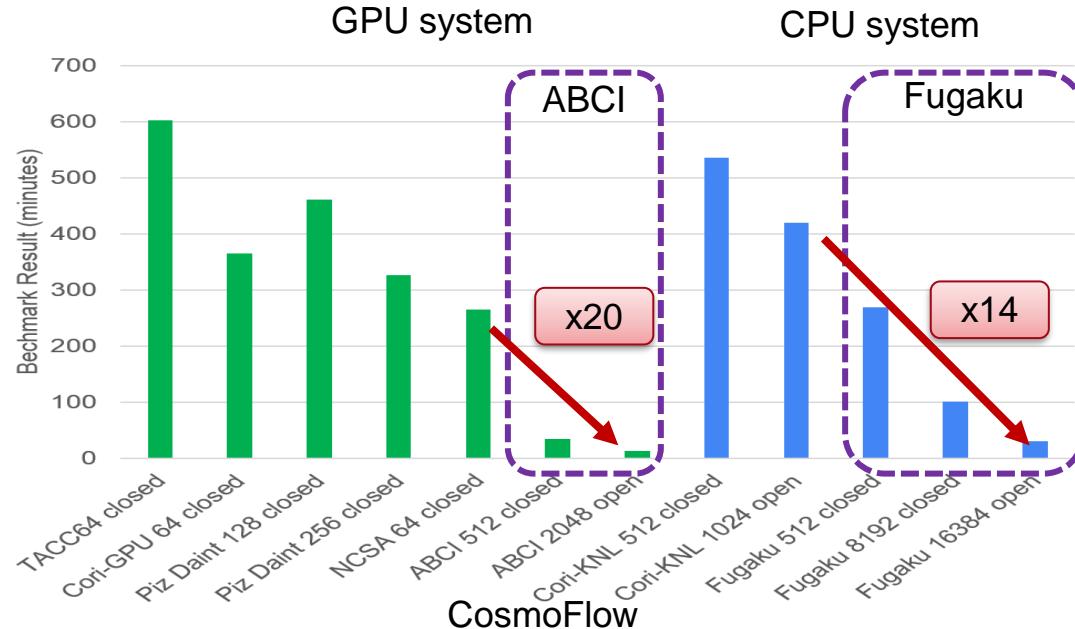
## ■ モデル並列

- ニューラルネットワーク計算そのものを並列化
- ネットワーク分割で必要となる袖交換が必要に
- 一般には性能向上が難しい



- 3D conv層に対してモデル並列化を適用、1.7倍高速に
- 単一学習タスクに対して16,384 CPUを適用(ハイブリッド並列)

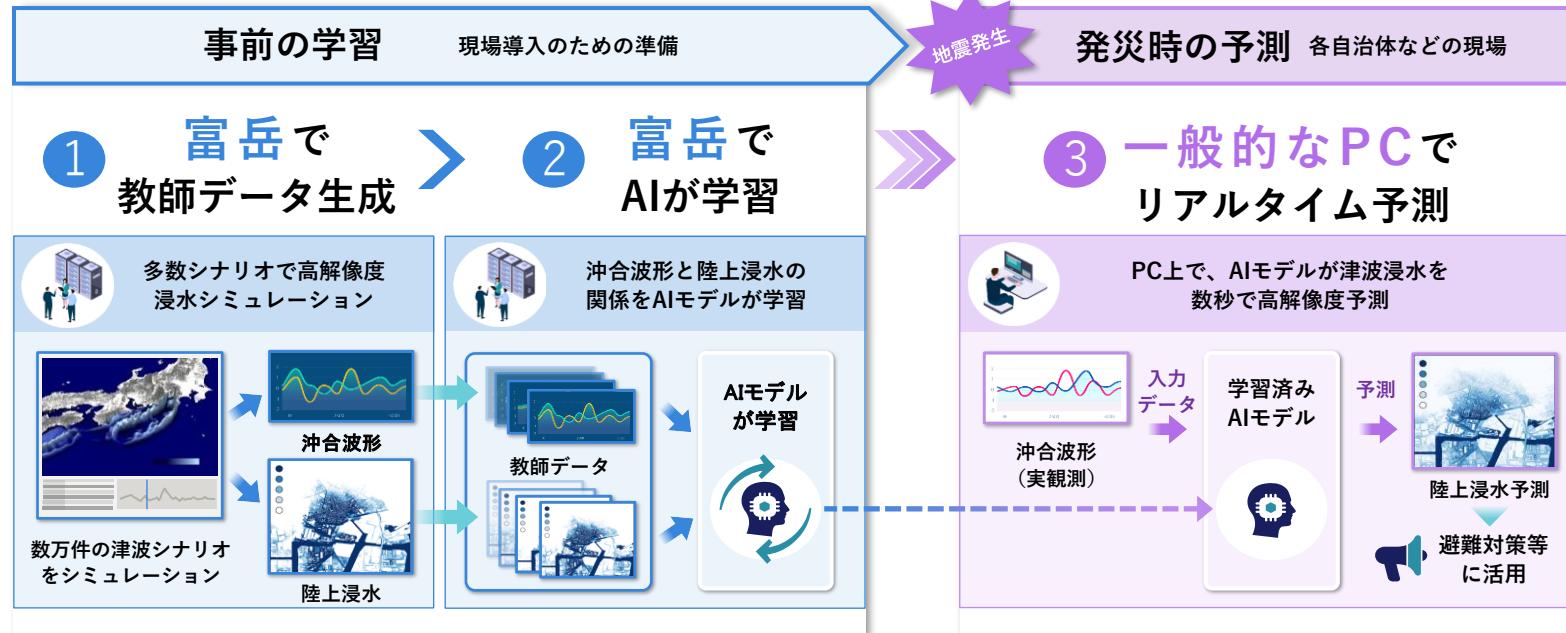
- ABCI: GPU型の2位の性能に対して20倍高速
- 「富岳」: CPU型の2位の性能に対して14倍高速



並列化技術によりシミュレーション結果の学習を大きく高速化  
但し、深層学習にはスケーラビリティ(規模)に課題

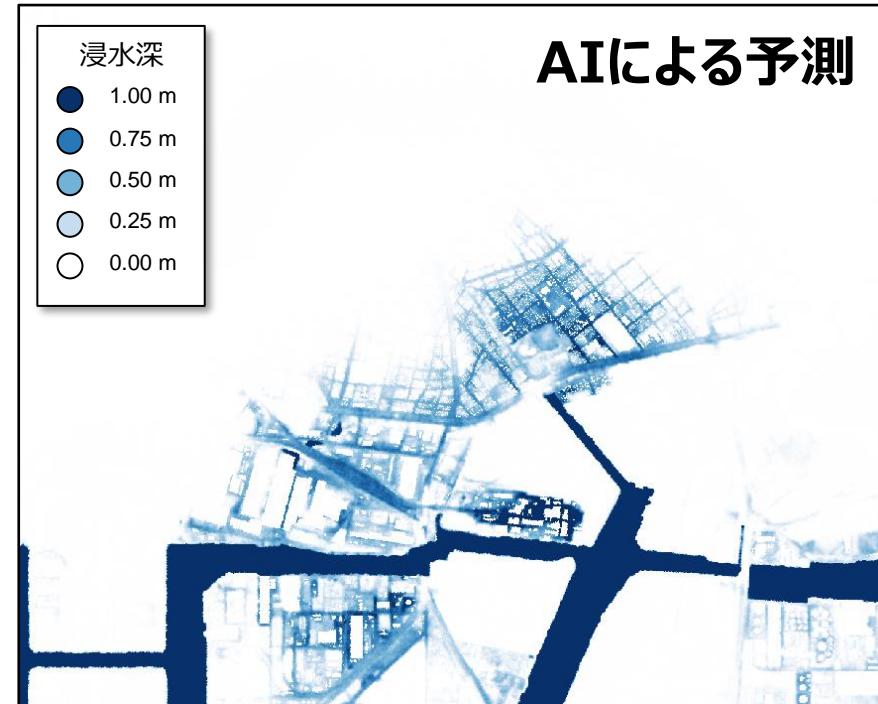
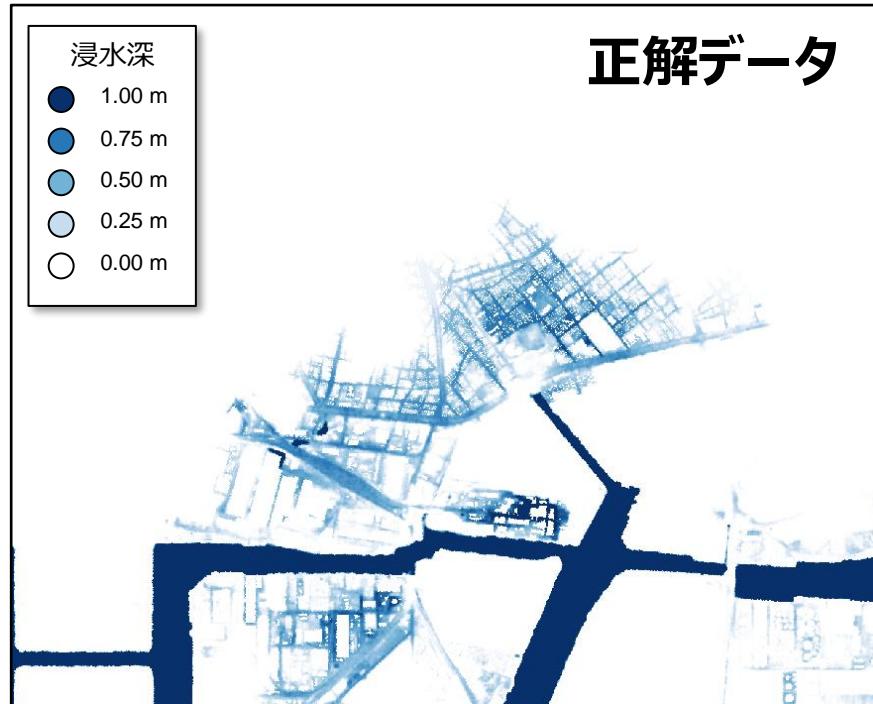
# AIを活用した津波予測 ～シミュレーションとAIの融合～

- 津波シミュレーション結果を学習したAIを生成
- 膨大な教師データを移動させず、富岳でそのまま学習
- 学習モデルを使えば、PCでも瞬時に陸上の浸水を予測



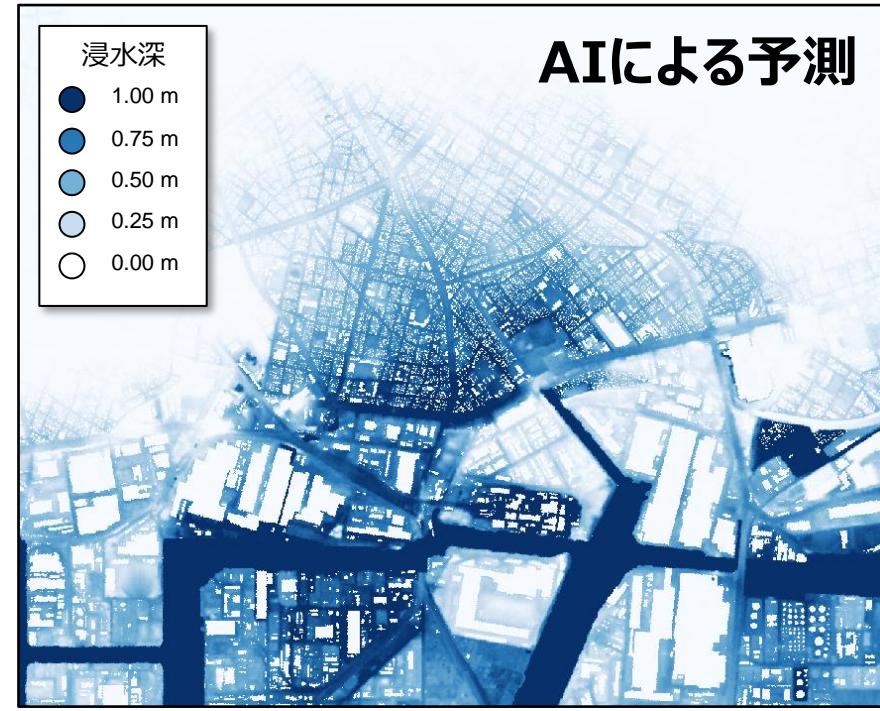
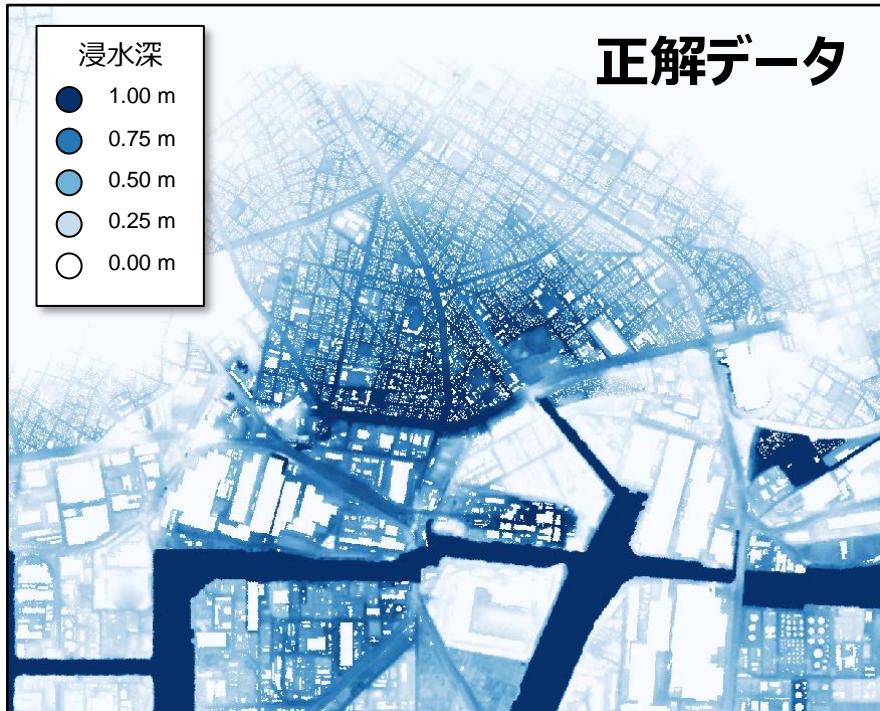
# AIによる予測と正解データの比較 ~浸水深・内閣府想定ケース~ **FUJITSU**

## ■ 正解データに十分に近い浸水深を推測



# AIによる予測と正解データの比較 ~浸水深・想定の3倍ケース~FUJITSU

## ■ 正解データに十分に近い浸水深を推測



- ディープラーニングの演算量は非常に大きい
  - ニューラルネットワークの規模とデータ量が大きいほど、精度が高くなりやすい
- ディープラーニング処理の高速化
  - 処理内容: Forward + Backward + AllReduce + Update
  - 高速化のポイント: AllReduce通信の隠蔽、大規模並列時の精度維持、適切な並列処理
  - 高速化の効果: チューニングにより非常に差ができる(例: Resnet-50, MLPerf-HPC)
- スーパーコンピュータの新しい使い方: AIとHPCの融合
  - スーパコンピューティング技術がAIを加速・進化させるドライビングフォース
  - AI技術も活用することで、シミュレーションだけでは解けない問題も解決

最先端のコンピューティング技術が新たな可能性を切り拓き、未来を創る

FUJITSU

shaping tomorrow with you