

BDEC (Big Data & Extreme Computing)

Society 5.0実現へ向けた取組



- 東大情報基盤センターの現状
- Society 5.0とBDECシステム
- h3-Open-BDEC
- 地震シミュレーション+リアルタイムデータ同化



3システム：利用者2,600+, 学外55+%

- Reedbush (HPE, Intel BDW + NVIDIA P100 (Pascal))(本郷)

- データ解析・シミュレーション融合スーパーコンピュータ
- 3.36 PF, 2016年7月～ 2021年3月末(予定)
 - Reedbush-U (CPU only, 2020年6月30日で退役)
 - Reedbush-H (2GPU's/n), Reedbush-L (4GPU's)
- 東大ITC初GPUクラスタ (2017年3月より), DDN IME (Burst Buffer)



- Oakforest-PACS (OFP) (富士通, Intel Xeon Phi (KNL))(柏)

- JCAHPC (筑波大CCS & 東大ITC)
- 25 PF, TOP500で18位(日本3位)(2020年6月)
- Omni-Path アーキテクチャ, DDN IME (Burst Buffer)

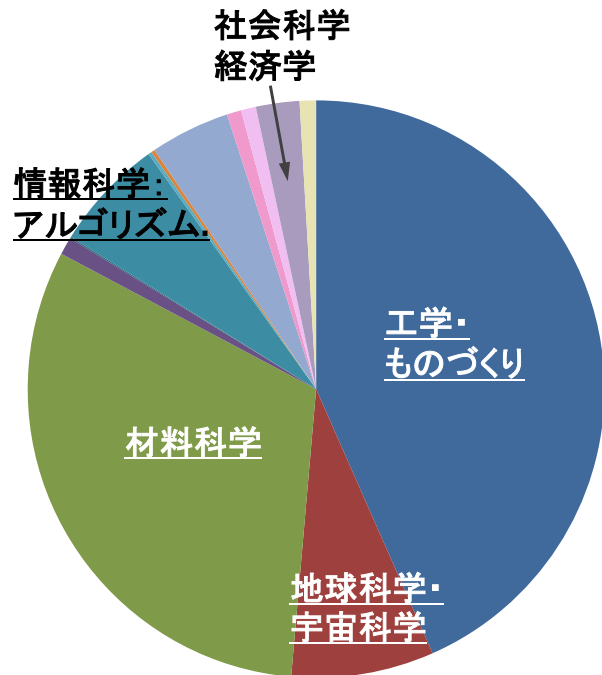


- Oakbridge-CX (OBCX) (富士通, Intel Xeon Platinum 8280)(柏)

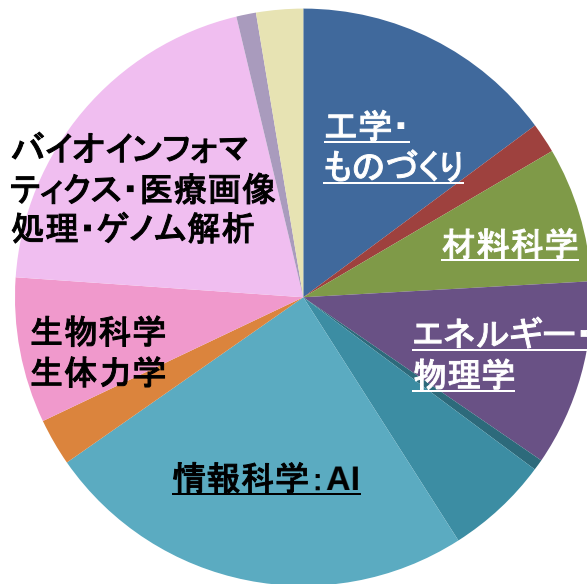
- 大規模超並列スーパーコンピュータシステム
- 6.61 PF, 2019年7月～ 2023年6月, TOP500で60位(2020年6月)
- 全1,368ノードの内128ノードにSSDを搭載



研究分野別利用CPU時間割合(2019年度)



マルチコアクラスタ
Intel BDW Only
(Reedbush-U)



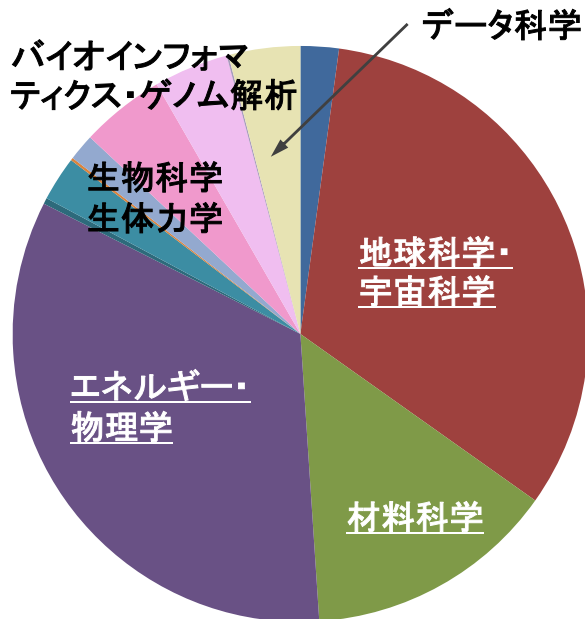
GPUクラスタ
Intel BDW + NVIDIA P100
(Reedbush-H)

- 工学・ものづくり
- 地球科学・宇宙科学
- 材料科学
- エネルギー・物理学
- 情報科学: システム
- 情報科学: アルゴリズム
- 情報科学: AI
- 教育
- 産業利用
- 生物科学・生体力学
- バイオインフォマティクス
- 社会科学・経済学
- データ科学・データ同化

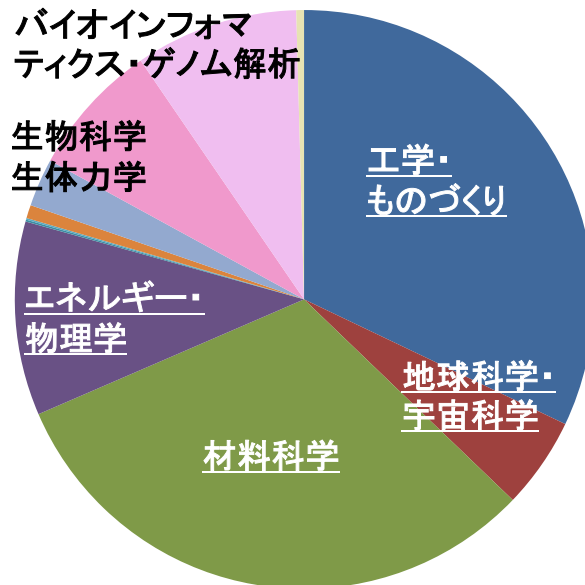


研究分野別利用CPU時間割合(2019年度)

OBCXは2019年10月～2020年9月末(12ヶ月)



メニコアクラスタ
Intel Xeon Phi
(Oakforest-PACS)



マルチコアクラスタ
Intel CLX
(Oakbridge-CX)

- 工学・ものづくり
- 地球科学・宇宙科学
- 材料科学
- エネルギー・物理学
- 情報科学:システム
- 情報科学:アルゴリズム
- 情報科学:AI
- 教育
- 産業利用
- 生物科学・生体力学
- バイオインフォマティクス
- 社会科学・経済学
- データ科学・データ同化

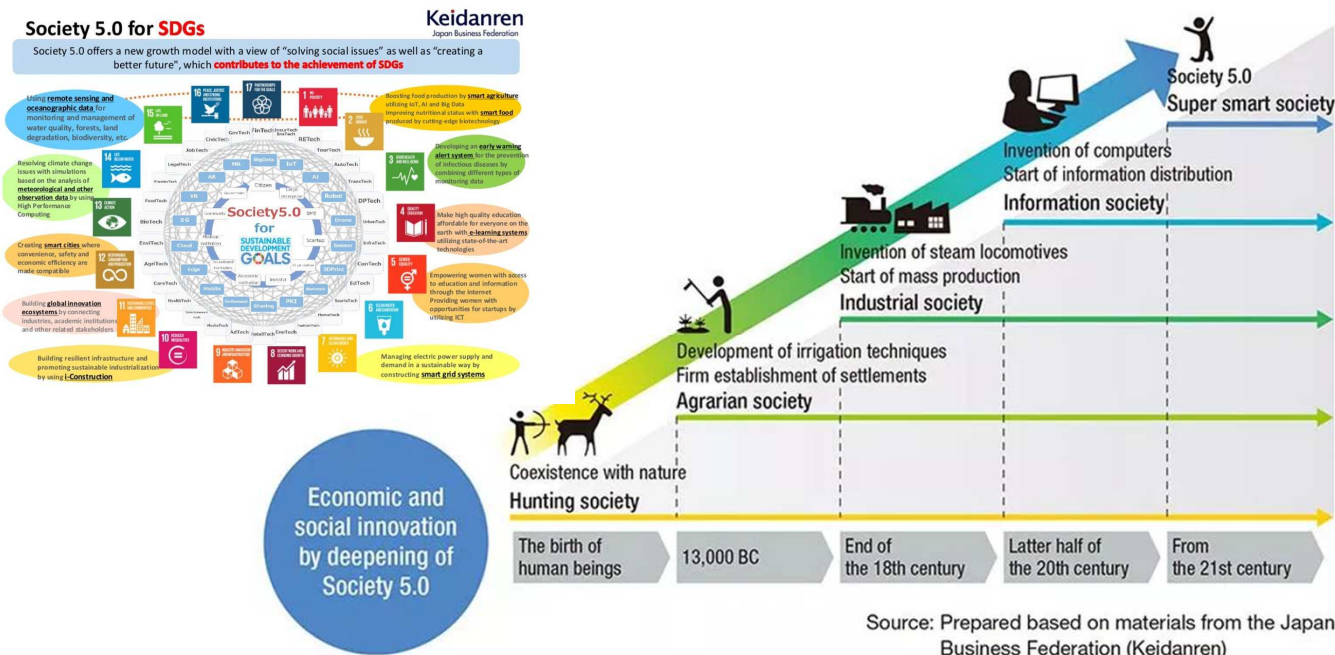


- 東大情報基盤センターの現状
- **Society 5.0とBDECシステム**
- h3-Open-BDEC
- 地震シミュレーション+リアルタイムデータ同化



Society 5.0: 日本が提唱する未来社会のコンセプト

デジタル革新・イノベーション（IoT、AI、ビッグデータ等）により、サイバー空間（仮想）とフィジカル空間（現実）を高度に融合させたシステムを構築し、経済発展と社会的課題の解決を両立する、超スマートな人間中心の社会



Source: Prepared based on materials from the Japan Business Federation (Keidanren)



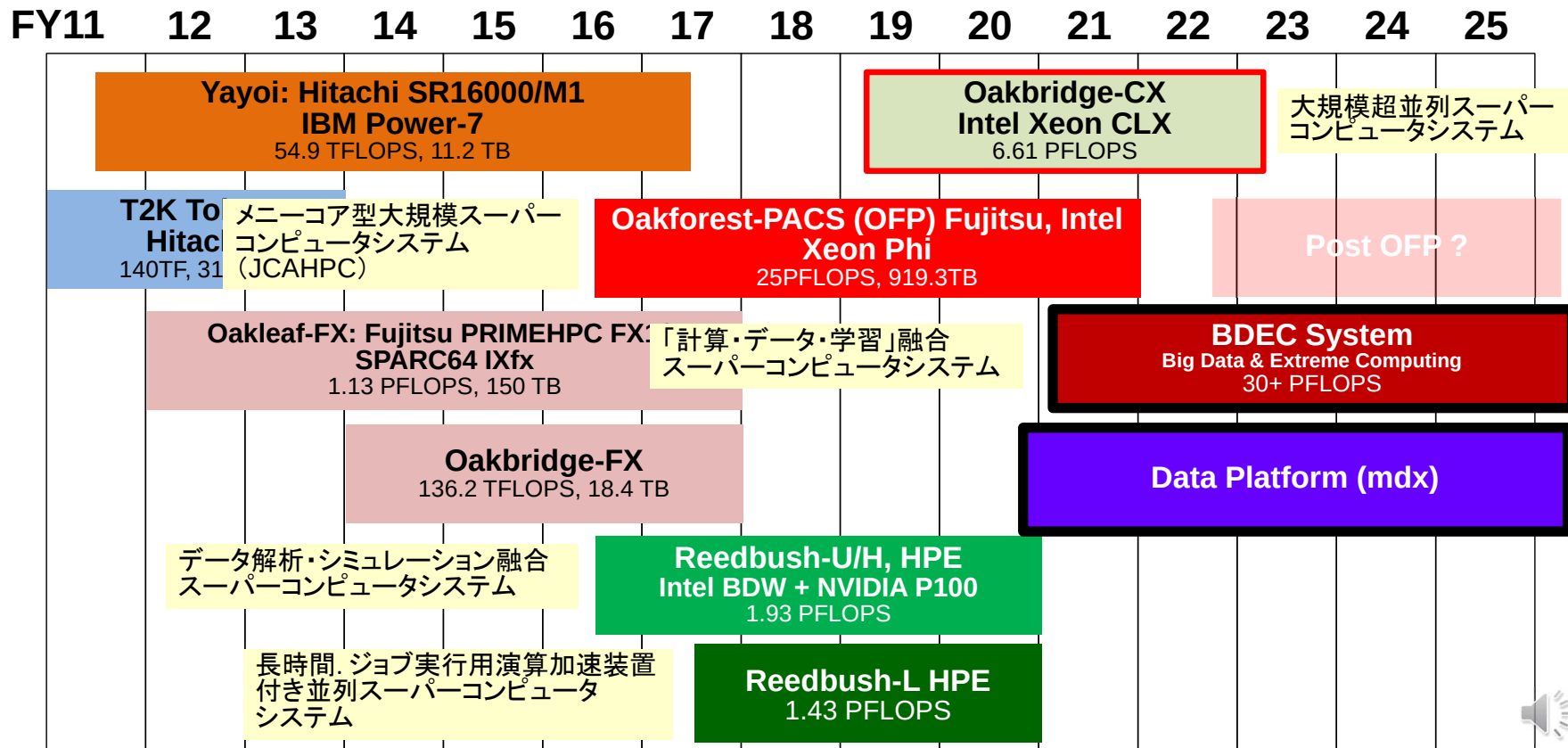
スーパーコンピューティングの今後の動向

- ワークロードの多様化
 - 計算科学, 計算工学: Simulations
 - 大規模データ解析
 - AI, 機械学習
- (シミュレーション(計算) + データ + 学習) 融合 ⇒ Society 5.0 実現に有効
 - ◆ S: シミュレーション(計算) (Simulation)
 - ◆ D: データ (Data)
 - ◆ L: 学習 (Learning)

フィジカル空間と
サイバー空間の融合
- 柏第IIキャンパスに(計算 + データ + 学習 (S+D+L)) 融合のための新しいプラットフォーム (BDEC (Big Data & Extreme Computing)) を導入予定 (2021年春)



東大情報基盤センターのスパコン



BDEC: Big Data & Extreme Computing

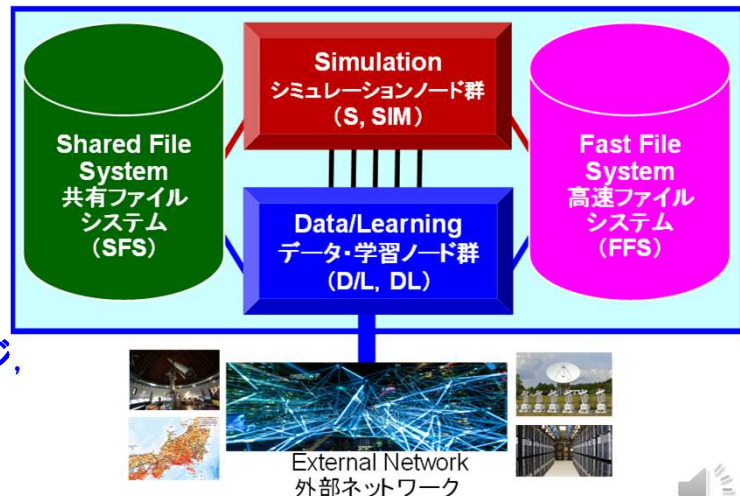
- BDEC (Big Data & Extreme Computing)

- 2021年5月(以降)運用開始予定
- 30+ PF, ~4.5 MVA(空調込み), ~360m²
- (D+L)による(S)の高度化 AI for HPC

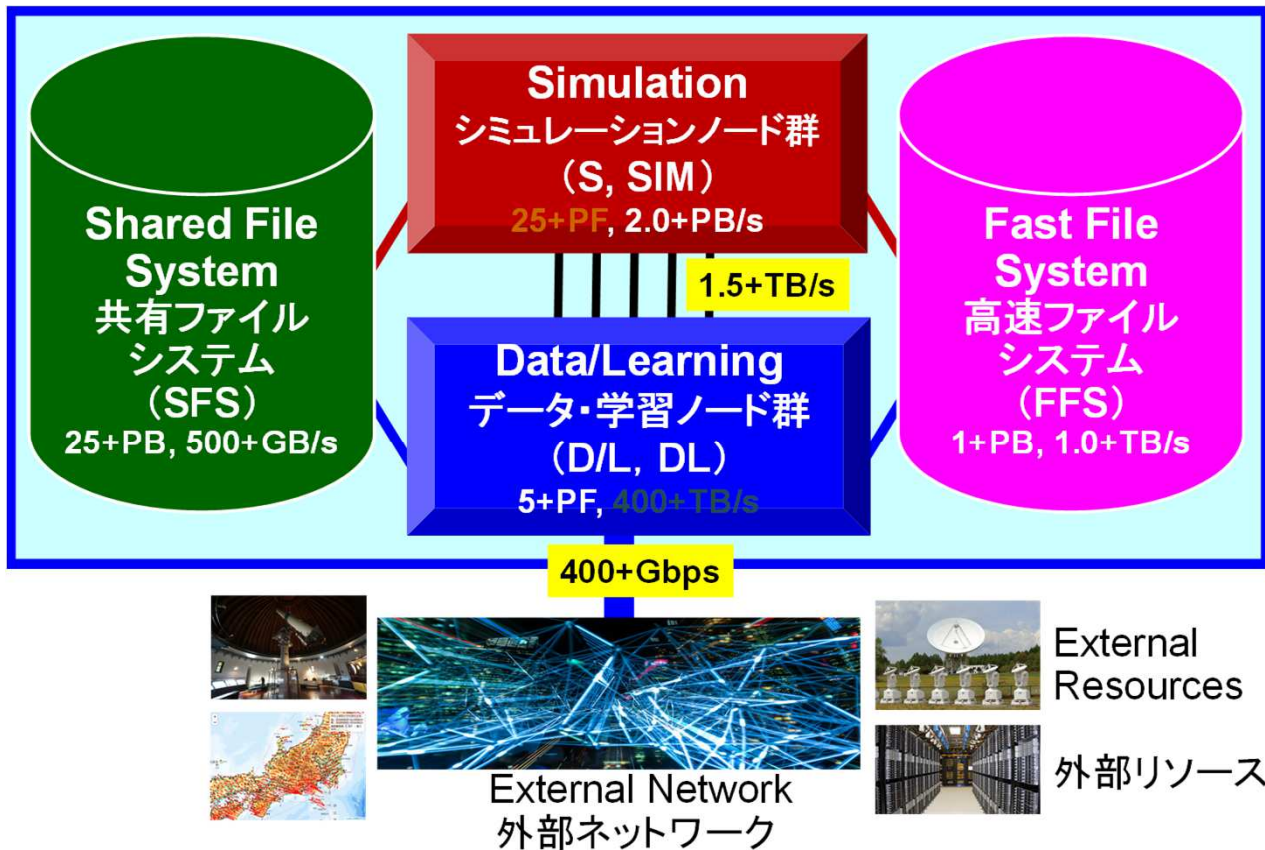
- 「シミュレーション・データ・学習(S+D+L)」融合, 2種類のノード群

- シミュレーションノード群(S, SIM)
 - 従来のスパコン
 - CPU with HBM, 25+PF, 2+PB/sec (最大9.0)
- データ・学習ノード群(D/L, DL)
 - データ解析, 機械学習
 - GPU Cluster, 5+PF, 575TB/sec
 - データ・学習ノード群の一部は外部リソース(ストレージ, サーバー, センサーネットワーク他)に直接接続
- Hierarchical, Hybrid, Heterogeneous (h3)
- ファイルシステム: 共有(大容量) + 高速

BDEC: S + D + L



BDEC: (S+D+L) 融合のためのプラットフォーム



シミュレーションノード群
Simulation Nodes
(S, SIM)

Fast File
System
(FFS)

Shared
File
System
(SFS)

データ・学習ノード群
Data/Learning
(D/L, DL)

計算科学コード

最適化されたモデル,
パラメータ

シミュレーション
ノード群, SIM

計算結果

機械学習, DDA

データ・学習ノード群
DL

観測データ

データ同化
データ解析

サーバー
ストレージ
DB
センター他



外部ネットワーク



外部
リソース群

シミュレーションノード群
Simulation Nodes
(S, SIM)

Fast File
System
(FFS)

Shared
File
System
(SFS)

データ・学習ノード群
Data/Learning
(D/L, DL)

計算科学コード

最適化されたモデル,
パラメータ

シミュレーション
ノード群, SIM

計算結果

機械学習, DDA

データ・学習ノード群
DL

観測データ

データ同化
データ解析

シミュレーションのためのモデル・パラメータのデータ
解析, AI/機械学習による最適化 (S+D+L)



- 東大情報基盤センターの現状
- Society 5.0とBDECシステム
- **h3-Open-BDEC**
 - ✓ ソフトウェアに関する研究
- 地震シミュレーション+リアルタイムデータ同化



(計算+データ+学習)融合によるエクサスケール時代の革新的シミュレーション手法(1/3)

- エクサスケール(富岳+クラス)のスパコンによる科学的発見の持続的促進のため、計算科学にデータ科学、機械学習のアイデアを導入した(計算+データ+学習(S+D+L))融合による革新的シミュレーション手法を提案
- 革新的ソフトウェア基盤「h3-Open-BDEC」の開発: 東大BDECシステム、富岳を「S+D+L」融合プラットフォームと位置づけ、スパコンの能力を最大限引き出し、最小の計算量・消費電力での計算実行を実現するために、下記2項目を中心に研究
 - 変動精度演算・精度保証・自動チューニングによる新計算原理に基づく革新的数値解法
 - 階層型データ駆動アプローチ(hDDA: Hierarchical Data Driven Approach)等に基づく革新的機械学習手法
 - Hierarchical, Hybrid, Heterogeneous $\Rightarrow$ h3



h3-Open-BDEC

「計算＋データ＋学習」融合を実現するためのソフトウェアフレームワーク

h3-Open-BDEC

変動精度演算に基づく
新計算原理

計算＋データ
＋学習

統合・通信・ユー
ティリティ

h3-Open-MATH
高性能・高信頼性・省電力
アルゴリズム (岩下)

DENSE: 密行列ソルバ,
SPARSE: 疎行列ソルバ,
LOOP-T: ループタイリング,
HACAPK: H行列ソルバ

h3-Open-VER
精度保証 (荻田)

L-DENSE: 密行列演算,
L-SPARSE: 疎行列演算, PDE: 精
度保証付PDEソルバー, NL: 非線
形問題, ITER: 反復改良法

h3-Open-AT
自動チューニング (片桐)

MEM: メモリアクセス最適化,
COMM: 通信最適化, ACC: 最適精
度, DB: データベース, LANG: AT
言語, C-GEN: コード自動生成

h3-Open-APP: 計算
計算科学アプリケーション
開発基盤 (中島)

FDM-Str: 構造格子差分法,
FEM: 有限要素法, LBM: 格
子ボルツマン法, AMR: 適応
格子, PinST: 時空間並列

h3-Open-DATA: データ
データ科学基盤 (長尾)

ASSIM: 一般のデータ同化, ADJ: ア
ジョイント法, SPM: スパースモデリ
ング, VIS: 可視化, FD: 特徴検知,
UQ: UQ, MOR: モデル縮減

h3-Open-DDA: 学習
データ駆動アプローチ基盤
(下川辺)

IF-ML: 機械学習IF, DDA: 従来型
DDA, hDDA: 階層型DDA, DB:
データベース

h3-Open-SYS
統合・制御基盤 (松葉)

EXEC: 統合実行環境,
RSCM: リソース管理, E-IO:
外部I/O, IF-DATA: 各データ形
式インタフェース

h3-Open-UTIL
並列ユーティリティ (八代)

DLB: 動的負荷分散, MP: 並列カブ
ラー, GRAPH: 階層型並列リオーダ
リング, I-IO: 内部I/O, FT: チェック
ポインティング, DBMS: DB制御

太枠で囲まれたh3-Open-APPと
h3-Open-MATH, h3-Open-DATA
とh3-Open-UTIL, h3-Open-
DATAとh3-Open-DDAは特に緊密
に連携する

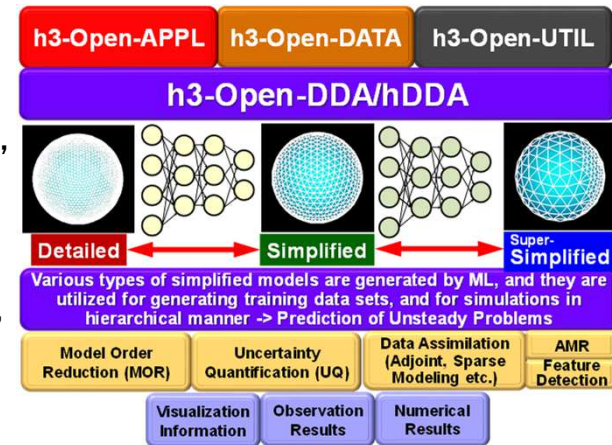
Hierarchical,
Hybrid,
Heterogeneous

Big Data &
Extreme
Computing



(計算+データ+学習)融合によるエクサスケール時代の革新的シミュレーション手法(2/3)

- 科研費基盤研究S(2019~2023年度)
 - 代表: 中島研吾(東大・情報基盤センター)
 - <http://nkl.cc.u-tokyo.ac.jp/h3-Open-BDEC/>
 - https://www.jsps.go.jp/j-grantsinaid/12_kiban/ichiran_r01/j-data/r_1_jp_19h05662.pdf
- 階層型データ駆動アプローチ:hDDA
 - シミュレーションに機械学習を適用して異なるパラメータでの解を予測するデータ駆動アプローチ(DDA)では、計算を繰り返して教師データを生成する必要がある。
 - 階層型DDA(hDDA)は、特徴検知, MOR(Model Order Reduction), UQ(Uncertainty Quantification), スパースモデリング, 適応格子等の諸機能を駆使して、計算量(メッシュ数, 粒子数)を削減した簡易モデルを、機械学習により自動生成, 教師データ生成用モデルとして利用する



階層型データ駆動アプローチ(hDDA)

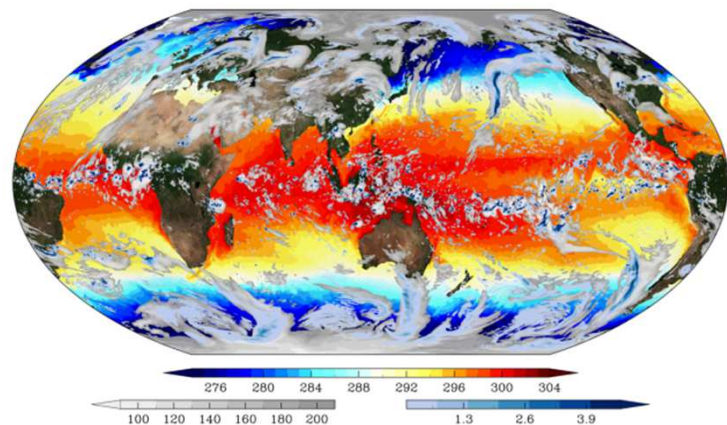
(計算+データ+学習)融合によるエクサスケール時代の革新的シミュレーション手法(3/3)

- 期待される成果と意義

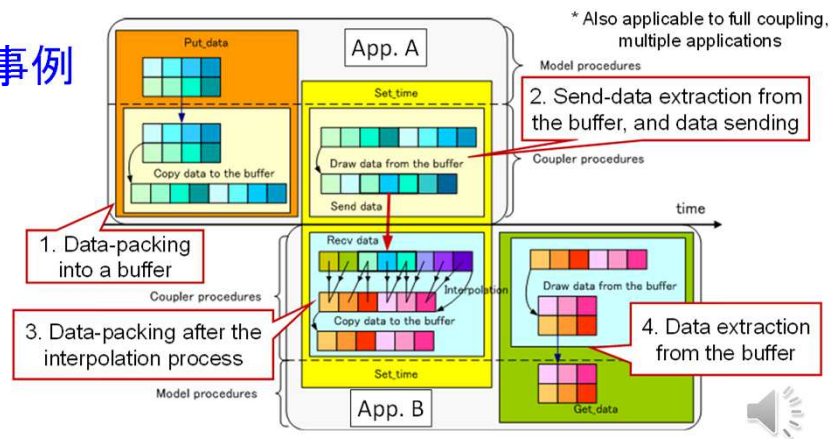
- 計算科学の専門家のみで(S+D+L)融合を容易に実現
 - 機械学習の専門家のサポートを必要としない
- ソースコード, マニュアル類も含めて一般に公開, 様々なエクサスケールシステムでの普及を目指す
 - ポスト富岳も含めたポストムーア時代への展開
- h3-Open-BDEC利用による(S+D+L)融合シミュレーションにより従来手法と同等の正確さを保ちつつ, 大幅な計算量・消費電力削減を目指す(10分の1が目標)。
- シミュレーション高度化: パラメータスタディのケース数を削減できる
- リアルタイム災害シミュレーション



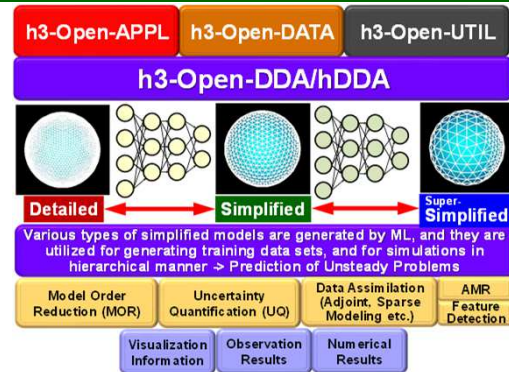
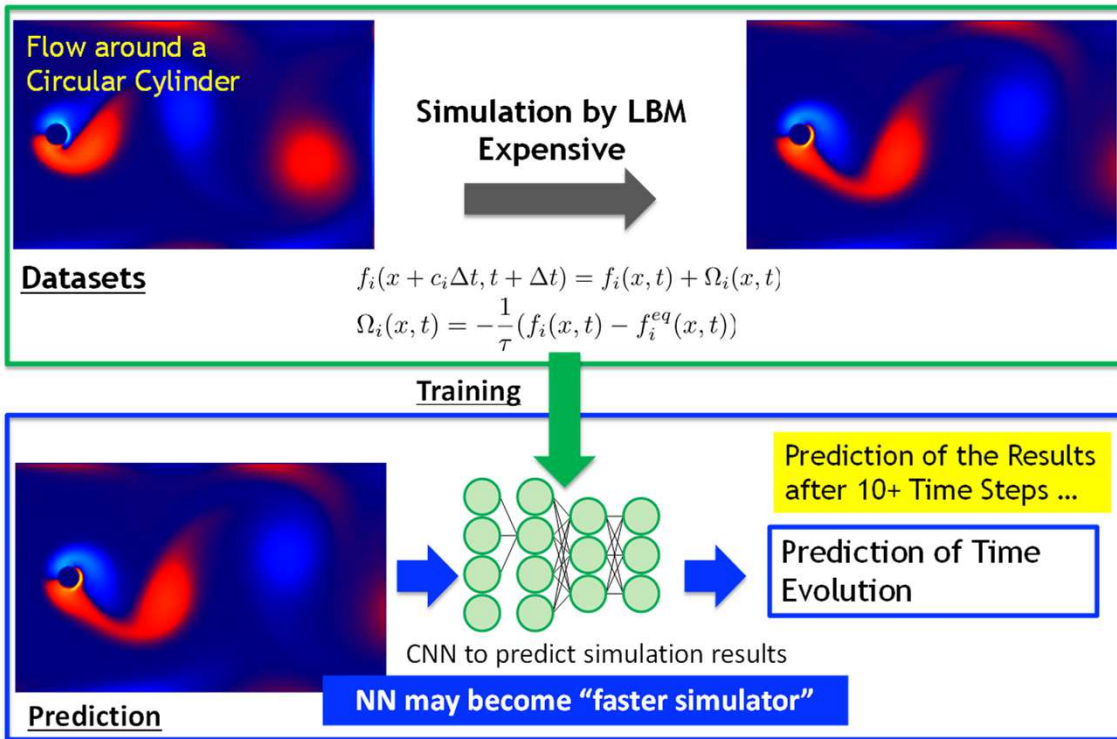
BDECにおける(S+D+L)融合



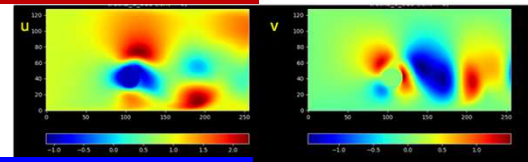
- 機械学習によるシミュレーションの高度化
 - 計算科学シミュレーションは非線形⇒多数のパラメータスタディが必要
 - 機械学習により最適パラメータが選択できれば, そのケース数を劇的に削減できる
- 大規模シミュレーション+データ同化
 - 以前から実施されている(S+D+L)の典型的事例
 - 大気海洋連成シミュレーション+データ同化
- リアルタイム災害シミュレーション
 - 地震, 洪水, 津波
 - リアルタイム観測データとの同化



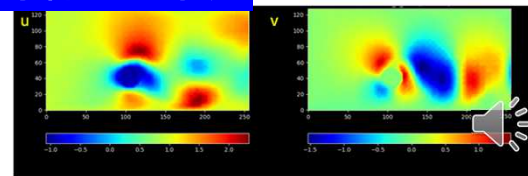
機械学習による非定常数値流体力学シミュレーション高速化 (計算+データ+学習(S+D+L))融合の実例



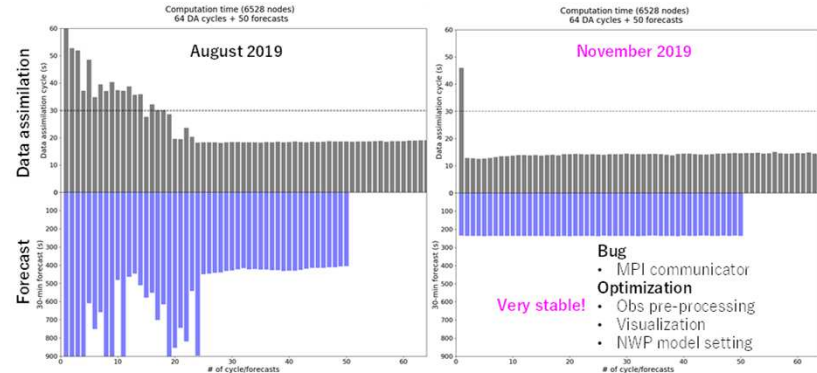
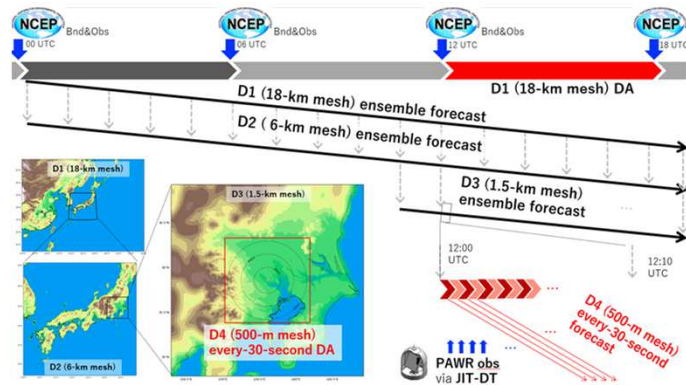
シミュレーション結果



機械学習による予測



ゲリラ豪雨予測のリアルタイム実証実験 (理化学研究所)

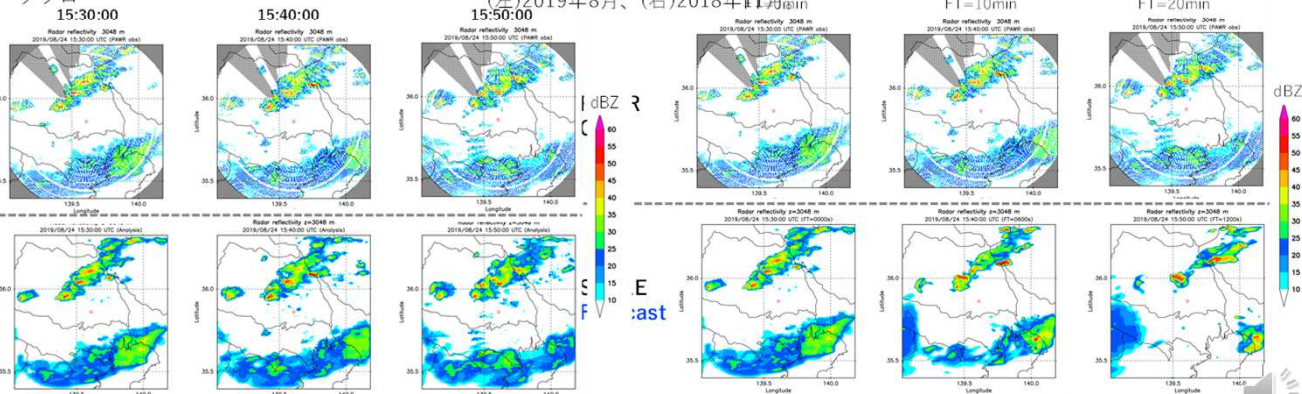


計算性能の向上。上段はデータ同化、下段は30分予報にかかった時間(秒)。
(左)2019年8月、(右)2018年11月



PAWR
Obs

SCALE-
LETKF
Analysis



〔画像提供: 三好建正博士
(理化学研究所)〕

2019年8月24日の事例についてのテスト結果。(上)レーダー観測と(下)SCALE-LETKFによる解析で得られたレーダー反射強度(dBZ)を示す。

2019年8月24日の事例についてのテスト結果。(上)レーダー観測と(下)SCALE-LETKFによる予報で得られたレーダー反射強度(dBZ)を示す。

- 東大情報基盤センターの現状
- Society 5.0とBDECシステム
- h3-Open-BDEC
- 地震シミュレーション+リアルタイムデータ同化



3システム: 利用者2,600+, 学外55+%

- Reedbush (HPE, Intel BDW + NVIDIA P100 (Pascal))(本郷)

- データ解析・シミュレーション融合スーパーコンピュータ
- 3.36 PF, 2016年7月～ 2021年3月末(予定)
 - Reedbush-U (CPU only, 2020年6月30日で退役)
 - Reedbush-H (2GPU's/n), Reedbush-L (4GPU's)
- 東大ITC初GPUクラスタ (2017年3月より), DDN IME (Burst Buffer)



- Oakforest-PACS (OFP) (富士通, Intel Xeon Phi (KNL))(柏)

- JCAHPC (筑波大CCS & 東大ITC)
- 25 PF, TOP500で18位(日本3位)(2020年6月)
- Omni-Path アーキテクチャ, DDN IME (Burst Buffer)



- Oakbridge-CX (OBCX) (富士通, Intel Xeon Platinum 8280)(柏)

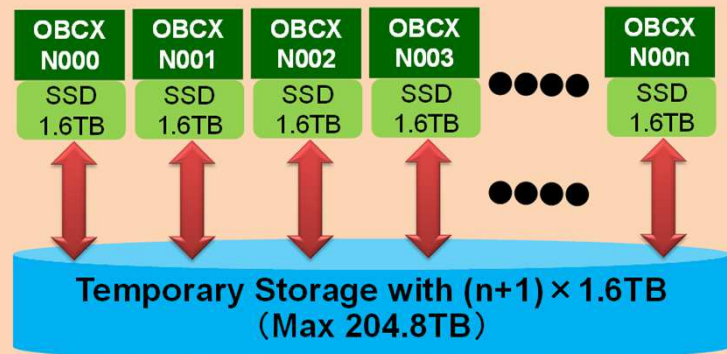
- 大規模超並列スーパーコンピュータシステム
- 6.61 PF, 2019年7月～ 2023年6月, TOP500で60位(2020年6月)
- 全1,368ノードの内128ノードにSSDを搭載



Oakbridge-CX (OBCX) : BDECに向けた実験システム

- 全1,368ノードのうち128ノードにSSD (Solid State Drive) 搭載
 - Intel SSD + BeeGFS
 - 容量: 1.6 TB/node
 - 読み書き性能: 3.20/1.32 GB/s/node
 - BeeOND (BeeGFS-on-Demand) によって合計 200+TB (128 × 1.6) の高速ファイルシステムとして使用可能
 - データ科学アプリケーション
 - ソフトウェア類も充実
 - ステージング, チェックポイント
 - 128ノードのうち16ノードはSINET経由で外部リソース(サーバー, ストレージ, センサーネットワーク)に直接接続⇒外部接続ノード

BeeGFS on Demand (BeeOND)



Total: 1,368 nodes

**128 nodes
with SSD**

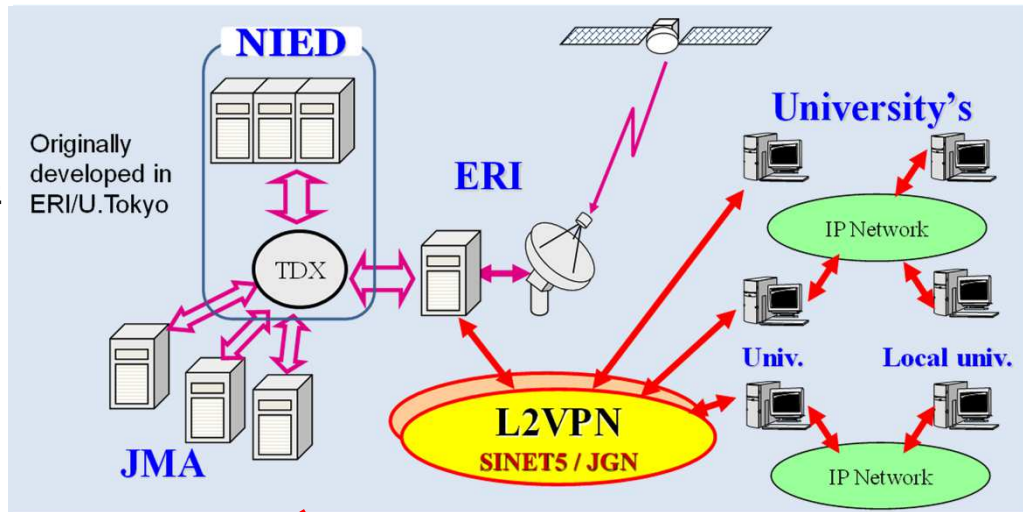
16

OBCXの16ノード(外部接続ノード)
SINET経由で外部計算機資源に直接接続,
BDECにおけるデータ・学習ノード群と同様の
役割



全国地震観測データ流通ネットワーク「JDXnet」

- 国内地震観測点の観測データ
(約2,000点, 100Hz, 3方向) を
SINET経由でリアルタイムに取得可能
 - 気象庁, 東大地震研, 防災科技研
 - 1日のデータ量: 100GB級
- OBCXの「外部接続ノード」による受信



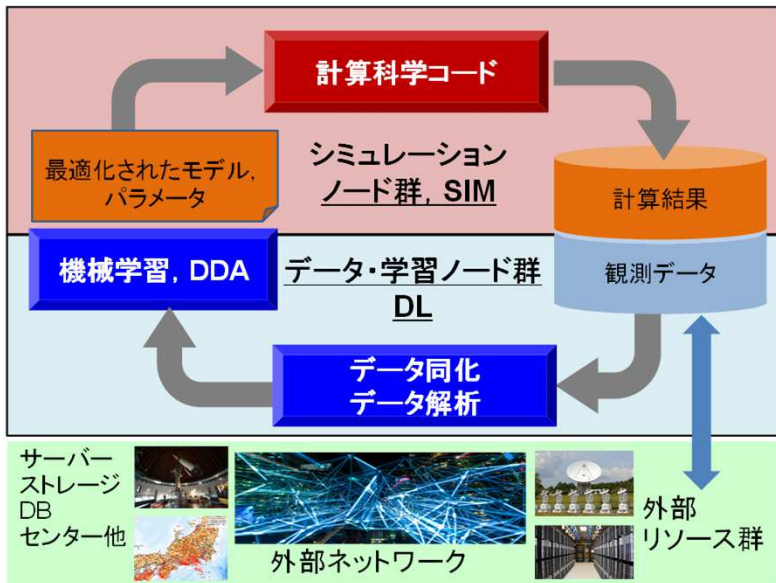
Total: 1,368 nodes

128 nodes
with SSD

16

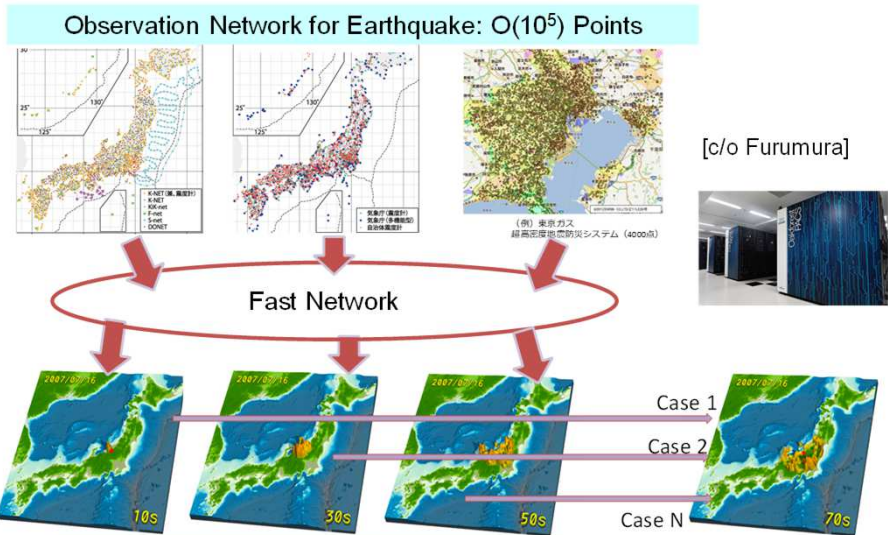
[資料提供: 鶴岡弘准教授(東大・地震研)]



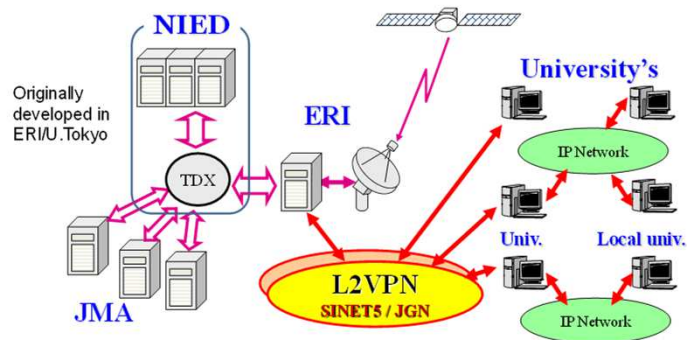


三次元地震シミュレーション+リアルタイムデータ同化/観測

JDXnetの観測データを利用したリアルタイムデータ同化/観測



[c/o Furumura]



Real-Time Data/Simulation Assimilation
Real-Time Update of Underground Model

[資料提供: 古村孝志教授
(東大・地震研)]

3D地震シミュレーション+リアルタイムデータ同化/観測

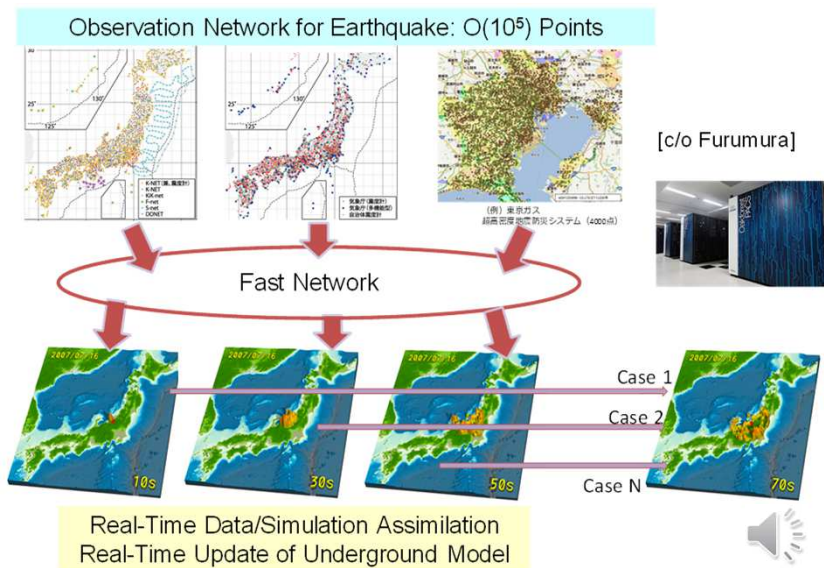
• より正確な地震波動伝播予測

- 地震波観測データを同化して得られる変位分布を初期条件として入力
- 60秒先の予測を10秒以内に計算することを目指す
- 的確な避難計画の策定



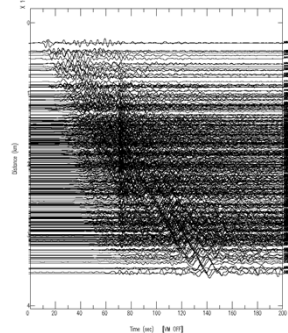
• 地下構造モデル

- 複雑, 不均質, 実はよくわかっていない
- 地震がおきるとその逆解析等をして地下構造モデルを少しずつ改良する, のが現状
- 本システムは, 普段は, 観測データ・三次元シミュレーション・データ解析を元に地下構造モデルの改良に使う予定
- 機械学習により精度高いモデル生成
- ⇒ (S+D+L) 融合へ

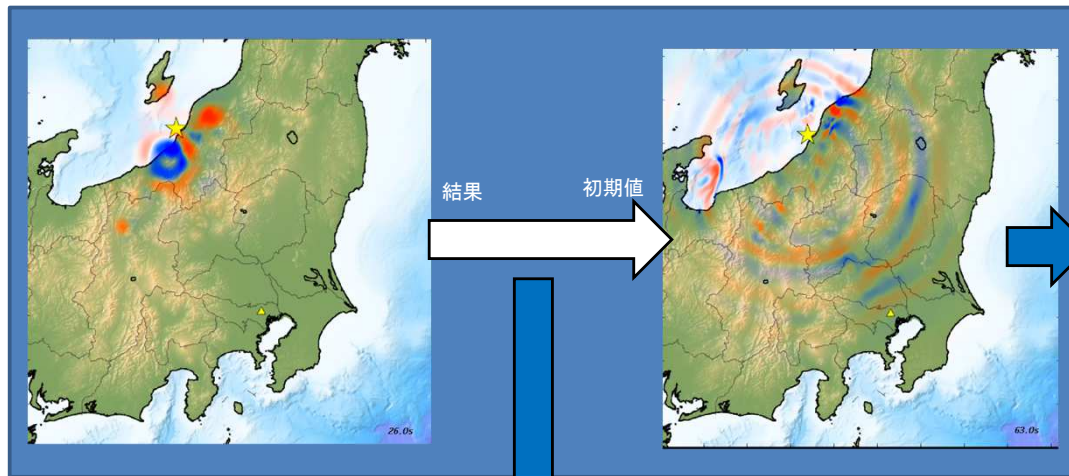
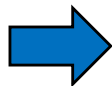


2019年度開発のシステム

① 過去に発生した大地震(2007年新潟県中越地震など)のF-net, Hi-net地震波形データをオフラインでフォーマット変換、フィルタ処理し、模擬入力データとして活用



入力



結果

初期値

結果

未来の予測
波動場

HDD

HDD

入力

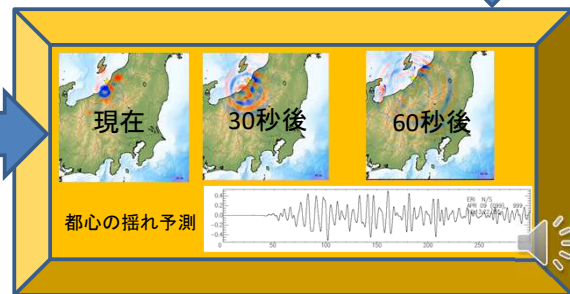
入力

Total: 1,368 nodes

128 nodes
with SSD

16

④ ディスクに保存された現在～未来の予測波動場の結果をもとに、オフラインで画像や地震波形を表示するプログラムは完成済み。



- 「Society 5.0」実現へ向けた取り組み
 - ✓東大情報基盤センターの現状
 - ✓Society 5.0とBDECシステム
 - ✓h3-Open-BDEC
 - ✓地震シミュレーション+リアルタイムデータ同化