東京大学情報基盤センター［ノーズ］

nodes

vol.5

特集

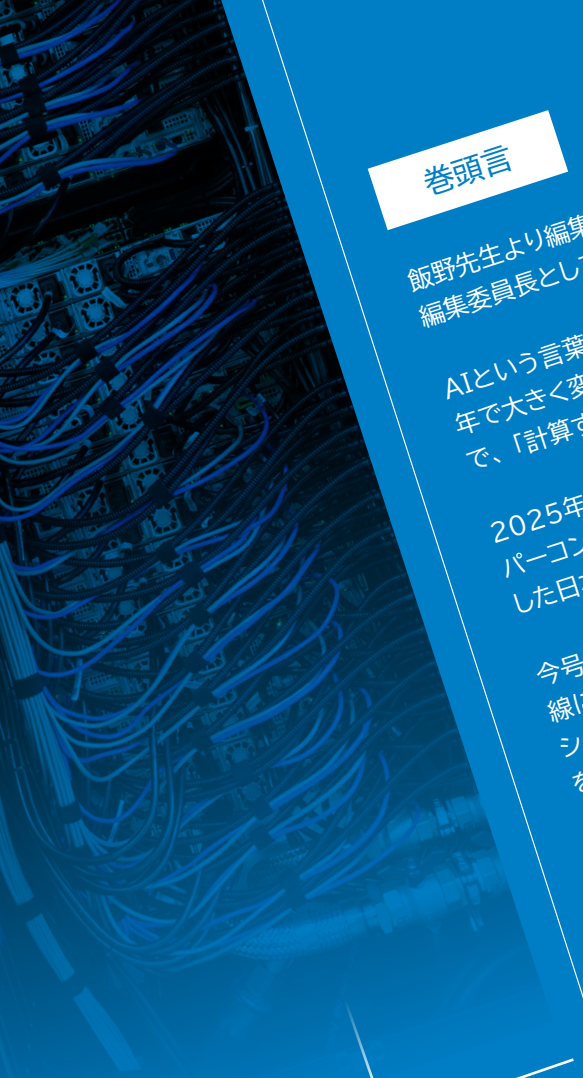
AI時代のスパコンと情報科学

連載 nodesの光明

AI研究を推進させる「UTokyo Azureサービス」の導入と運用

連載 飛翔するnodes

Miyabi-GのGPUとCPUを使いこなし気象・気候シミュレータ「SP-MIROC」を高速化



巻頭言

飯野先生より編集委員長を引き継ぎました、永井佑紀です。
編集委員長として最初のnodesをお届けします。

AIという言葉を耳にしない日はないほど、研究や社会の風景はこの数年で大きく変わりました。機械学習が新しい発見の手法として広がる中で、「計算すること」の意味そのものが広がっているのを感じます。

2025年、東京大学情報基盤センターは筑波大学と共同で、新スーパーコンピュータ「Miyabi」の運用を開始しました。AIとHPCを融合した日本最大級の計算基盤として、多様な分野の研究を支えています。

今号の特集では、Miyabiの概要とその背景に加え、AI×HPCの最前線に立つ研究者のみなさんの声をお届けします。銀河進化シミュレーションや気象・気候研究など、計算機の進化がもたらす新しい可能性をご紹介します。

また「nodesの光明」や「飛翔するnodes」では、センターのサービスや研究の現場を取り上げました。情報基盤を支える多くの“nodes”の姿を感じていただければ幸いです。

創刊第5号のnodesを、どうぞお楽しみください。　（永井佑紀）

nodes

「ノーズ」vol.5

編集・発行　東京大学情報基盤センター
編集委員長　永井佑紀
編集委員　大山智也、川瀬純也、下川辺隆史、関谷貴之、中村遼、本田匠、有馬和美、大林由尚

編集・執筆協力　株式会社インプレス
取材・執筆　大原雄介
撮影　井上勝也（井上勝也写真事務所、p.10～14）
阪口貴紀（p.8～9、p.15）
鹿野宏（株式会社Lab、表紙、p.3～5）

デザイン　株式会社ワックスグラフィックス
印刷　株式会社広済堂ネクスト

Contents

特集　AI時代のスパコンと情報科学　03

連載　nodesの光明　10
AI研究を推進させる「UTokyo Azureサービス」の導入と運用

連載　飛翔するnodes　12
Miyabi-GのGPUとCPUを使いこなし気象・気候シミュレータ「SP-MIROC」を高速化

nodesのひろがり　14

特集

AI時代のスパコンと情報科学

東大と筑波大が共同構築した最新スパコン「Miyabi」。倍精度演算性能80.1PFLOPSという日本トップクラスの性能を持ち、2025年1月から運用が開始されました。 AIの急速な進歩に対し、このスパコンがどのように使われているのか、今後使われるようになるのか、またAIを活用した情報科学の現在と未来について見ていきましょう。

スーパーコンピュータ「Miyabi」日本最大級のAI×HPCプラットフォーム

最先端共同HPC基盤施設（JCAHPC）は2013年、筑波大学と東京大学が協定を結び、筑波大学計算科学研究センターと東京大学情報基盤センターが共同で設置した施設です。地理的に言えば筑波大学と東京大学の中間あたりに位置する、千葉県柏市にある東京大学の柏キャンパスで2025年から運用を開始したスーパーコンピュータであるMiyabiについて、筑波大学の建部修見教授、東京大学の塙敏博教授、住元真司特任教授の3人に話を伺いました。

大学の枠を超え、共同調達で「日本一」を目指す

図1は筑波大学と東京大学のスーパーコンピュータの歴史です。2000年代後半、各大学で別々に仕様を定めてスーパーコンピュータを導入するのではなく、共通仕様にして大学主導でスーパーコンピュータを設計、導入する動きがありました。T2K、筑波大学と東京大学、京都大学の3学の頭文字を取ったものですが、このT2K仕様のものが2008年から導入されました。この時は仕様こそ同じでも、設置場所はそれぞれの大学で個別でした。この動きをさらに踏み込み、ふたつの大学で共同で予算を出し合うことで日本一のスーパーコンピュータを狙おうと2013年に筑波大と東大が連携することになりました。京都大学が外れているのは、ひとつは距離的に離れていること、もうひとつは調達時期が京都大学だけ違っていて、共同調達ができなかったことに起因します。

こうした経緯で、筑波大学と東京大学で共同で調達したJCAHPCの第1世代になるのがOakforest-PACSというシステムです。これはIntelのXeon Phi、Knights Landingというコード名のCPUを使った8,208ノードのシステムを、やはりIntelのOmni-Pathというインターコネクトで接続した構成で、性能だけではなく電力効率の向上も狙ったものでした。このOakforest-PACSは2022年3月に運用が終了しています。

空白の3年を経て、GPUメインの第2世代へ移行

実はOakforest-PACSの運用が終了する前から、第2世代も協力してやりましょう、という話は出ていました。その第2世代ですが、実行性能と電力効率を考えるとGPUをメインに据えるのが妥当だろうとなったものの、CPUでないと対応できないアプリ

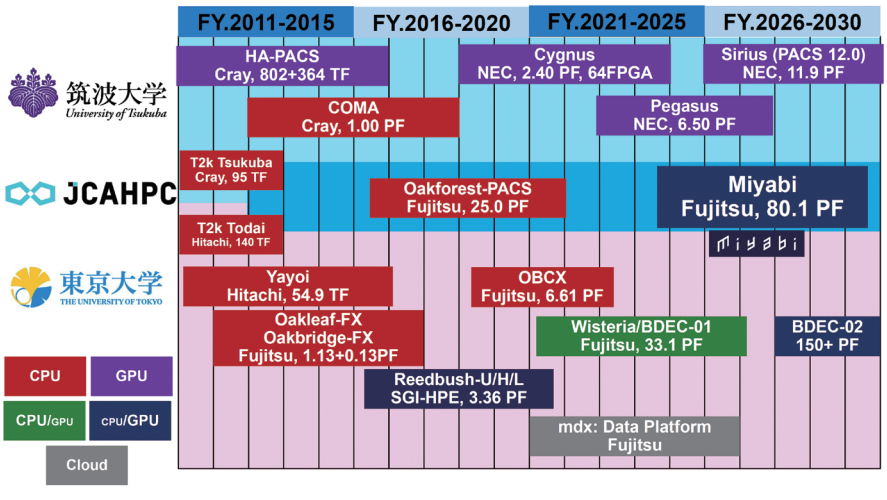


図1 筑波大学と東京大学のスーパーコンピュータの歴史
（計画段階を含む）

筑波大学
計算科学研究センター 教授
建部修見

専門は高性能計算および計算機システム。東京大学大学院理学系研究科情報科学専攻博士課程修了、博士（理学）。独立行政法人産業技術総合研究所 主任研究員などを経て、2015年より現職。

東京大学
情報基盤センター 教授
塙敏博

専門はインタコネクトおよび高性能計算システム。慶應義塾大学大学院理工学研究科計算機科学専攻博士課程修了、博士（工学）。筑波大学計算科学研究センター 准教授などを経て、2020年より現職。

東京大学
情報基盤センター 特任教授
住元真司

専門は計算機システム。慶應義塾大学大学院理工学研究科卒。富士通株式会社次世代TC開発本部シニアアーキテクトなどを経て、2022年より現職。

ケーションを使っているユーザーも一定数存在することから、GPUをメインにしたMiyabi-Gと、CPUのみで構成されるMiyabi-Cの2種類のシステムを組み合わせる形になりました（図2）。ただMiyabiが運用を始めたのは2025年ですから、Oakforest-PACSの運用終了から3年ほど空いてしまいました。これはMiyabiの導入のためにOakforest-PACSの設備の撤去が必要だったこと、Miyabiに適した製品が入手可能になる時期まで少々時間がかかったことなどもありますが、逆にその期間を使ってユーザの既存コードのGPU対応に充てることができました。

アーキテクチャ刷新とGrace Hopper採択の決め手

Miyabiの構成と申しますか、どのGPUを選ぶかという問題に関しては、各GPUベンダーにベンチマークを事前にお願いして、その結果を基に判断することになりました。さまざまな検討を経て、最終的にNVIDIA NVLink-C2CによるCPU-GPUの密な連携が可能なNVIDIA Grace Hopper(図3)が最適と考え、最終的にこのような仕様で富士通から納入されました。

これと並行して、既存のCPUを利用したアプリケーションをGPUに移行させる作業も必要になってきます。既存のユーザーは、Fortran+OpenMPという形で記述されているアプリケーションを多数お持ちだったりします。これをOpenACCに移植する作業ではNVIDIAにご協力いただきました。

ただ、OpenACCを使ってしまうと今度は他の構成に移行しにくいという問題がありまして、現在なるべくニュートラルなプログラムの書き方というのを検討しているところです。

話を戻すと、2023年11月に第2世代のシステムの導入を決定したという発表をしたのですが、その時点では構成が完全に決まっていました。発表前から移行作業をどんどん進めるとともにベンダーのサポートも得られたのですが、当時はコロナによるサプライチェーンの影響が残っており、余裕を持ってそのような計画で進めました。従来のx86からArmにアーキテクチャが変わることは、問題がなかったわけではありませんが、GPUに移行すること自体に比べれば大きな問題ではありませんでした。移行前にはいろいろ問題があったわけですが、幸いNVIDIAの全面的な協力を得られたおかげで、稼働段階ではそうした問題が全部解決しています。

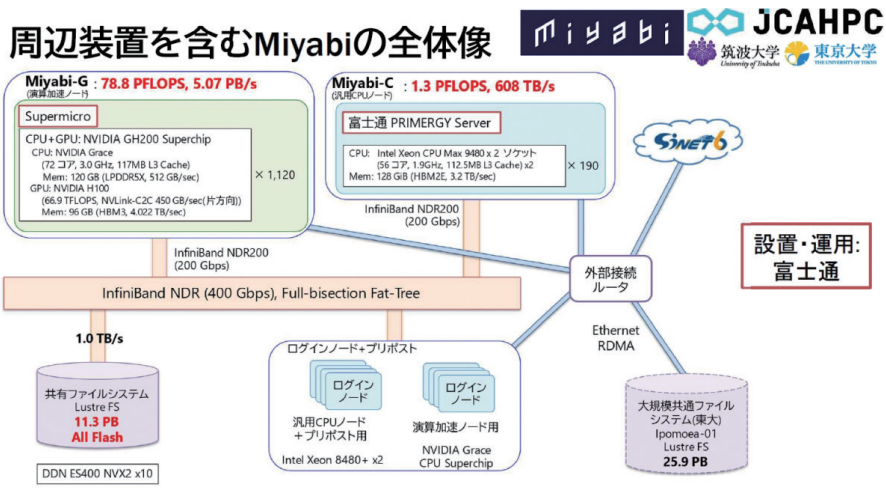


図2 Miyabiのシステム構成

3,000人が活用する運用ルールとHPCIの枠組み

運用を始めて最初の数か月は移行作業のせいか徐々にという具合でしたが、現在はかなり使われています。ノード数は1,120ですが、これをほぼすべて使えるのは月に1回あるHPCチャレンジという時だけで、これは公募ベースで選択しています。他は256ノードまでだと最大24時間、128ノード以下だと最大48時間という時間制限を設けてありまして、あとはユーザーがジョブスケジューラーに自分のジョブを投げるとスケジューリングされて実行されるという形です。使われ方には時期によって多少ムラがありまして、年度末が近くなると80 ～ 90%まで負荷が上がり、ジョブの待ち時間も長くなりがちです。

ユーザーは約3,000人ですが、半分以上は学外の方です。利用にあたってはプロジェクトに対して課金が発生します。学生などの場合は、プロジェクトに参加して使ってもらう形ですね。他にも、HPCIプログラムというものがありまして、もともとは理化学研究所のスパコン「京」などに合わせて設けられたもので、プロジェクトを公募し、審査に合格すると無償で利用できます。Miyabiも現在この対象に含まれていて、他に筑波大学では学際共同利用プログラム、東京大学ではJHPCNプログラムで公募ベースで無償利用が可能です。学術利用だけでなく、企業の研究者も含め、いろいろな形でMiyabiを利用できます。

AI活用が切り開くスパコンの未来

実際のMiyabiの使われ方を見ていると、当初こそHPC系が多かったものの、最近はAIが増えています。具体的に何をやっているのかを追跡できているわけではないのですが、Pythonの稼働時間が増えているというのはAIを使っている傍証になります。

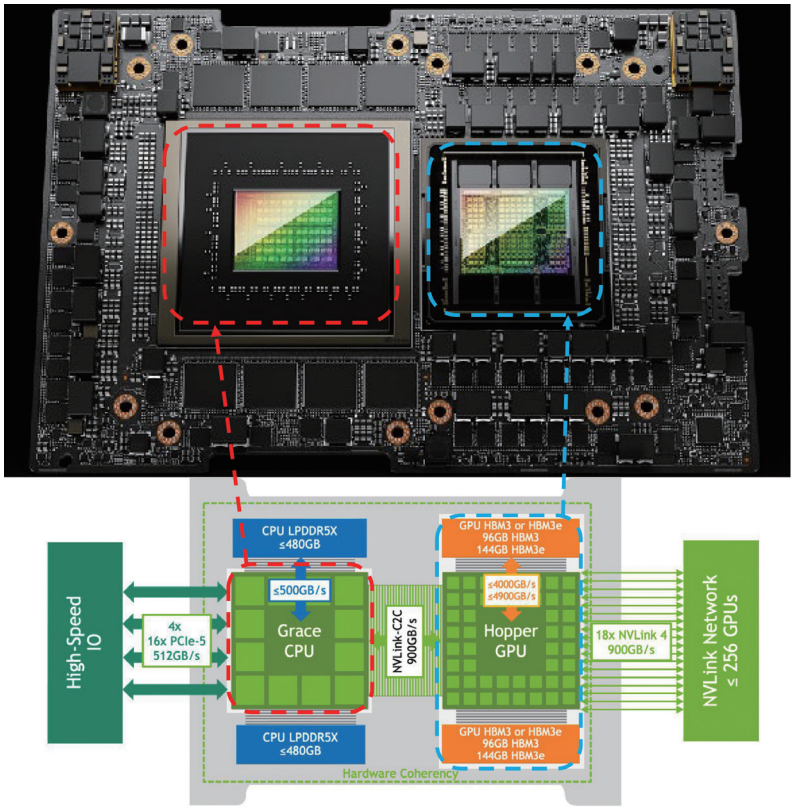


図3 NVIDIA Grace Hopper Superchipの論理的概要
出典：NVIDIA

Miyabiに関しては、基本的には現在の構成のまま運用終了まで変わらないと思います。この先のAI需要の高まりに対して、後継のシステムでは一層の対応が必要になりそうです。富岳NEXTが2030年ですから、NVIDIAのGPUで言えばRubinか、その次のFeynman世代ということになりますか。Miyabiの後継も同じような話になると思うのですが、ひとつ言えるのはFP64（*1）の能力は今後あまり増えない傾向にある事でしょうか。現在は低精度の演算器を効率よく用いて高い精度の演算が可能なアルゴリズムの研究が進められています。NVIDIA Tensorコア（*2）を使ってFP64と同等の精度でずっと高速に演算が可能というアルゴリズムがすでにありまして、今はそれがどこまでちゃんと適用できるかをみんなで調べているところですね。これがうまくいくようであれば、全面的にそういう方向にシフトしていくのかもしれません。

今後はAI for Scienceの深化に加え、JHPC-quantumプロジェクトや量子イノベーションイニシアティブ協議会（https://qii.jp/）へのサービス提供を通じたQC-HPC連携にも注力します。スパコンと量子計算を統合するハイブリッド運用により、創薬や材料開発などで従来の限界を超える革新の創出を目指しています。多様な計算資源が有機的に結合することで、科学技術の変革を牽引する基盤となります。Miyabiがこの架け橋として、新たな科学の発展に貢献していくことを期待しています。

参考資料

次世代HPC・AI 研究開発支援センター（HAIRDESC）
https://hairdesc.jp/

（*1）倍精度浮動小数点演算。有効数字15桁。
（*2）NVIDIA GPUの中で行列積和演算を行う回路。AI処理で多用される行列演算を効率的に実行する。

AIを用いた銀河進化シミュレーション

Miyabiを利用すると、どんな研究が可能になるのか？ 実際にMiyabiの大規模HPCチャレンジに採択されたテーマとして、理化学研究所の平島敬也博士に「AIを用いた銀河進化シミュレーション」のお話を伺った。

理化学研究所
数理創造研究センター
数理基礎部門 基礎科学特別研究員
平島敬也

超新星爆発をどのように扱うかが課題

私が行っている仕事は、銀河のシミュレーションを高い解像度で行おうというものです。銀河全体の進化、究極的には我々が住んでいるこの太陽系や銀河がどうやって生まれてきたのかを調べようというプロジェクトです。シミュレーションそのものは銀河全体を空間的に格子分けし、ある時点におけるそれぞれの格子内部の動きを計算したのち、格子同士の相互作用を計算するというプロセスを繰り返しますが、実はいくつか問題があります。

もっとも難しいのは、恒星が爆発する超新星爆発をシミュレーション中で取り扱う際です。超新星爆発が起こると衝撃波が発生してそれが周囲に影響を及ぼすわけですが、この爆発現象というのは100年～1,000年ぐらいで終わってしまいます。一方銀河全体を扱う場合、銀河が1回転するのに1億年くらいかかるので、タイムスケールが5～6桁近く違うことになります。これを従来のシミュレーションで取り扱う場合、例えば100ノードのマシンで並列計算をしていたとして、その超新星爆発が起きた場所を取り扱っているのは1ノードしかないので、そのノードが超新星爆発の影響を計算している間、他の99ノードは止まってしまうんですね。タイムスケールのギャップの大きさがこの問題につながるわけです。

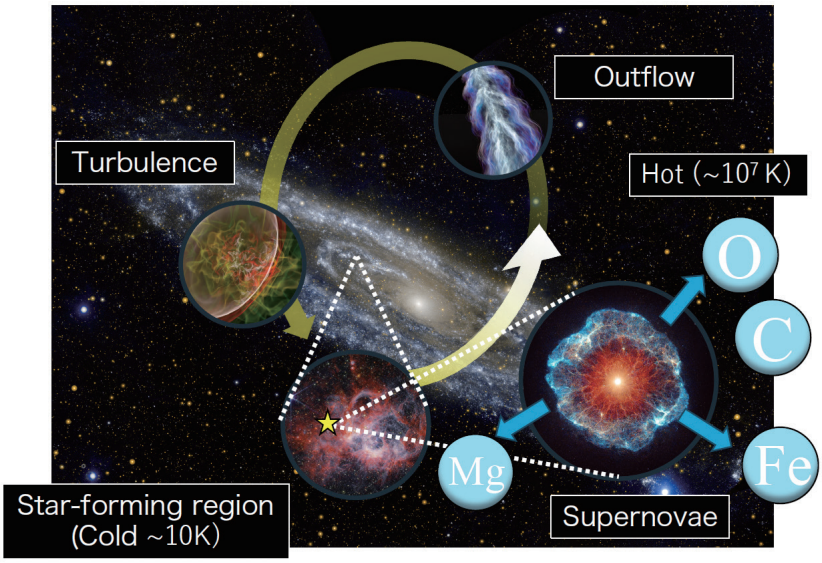


図1 銀河内でのガスの循環。超新星爆発は銀河の成長に大きな影響を与えるが、これまでシミュレーションで取り扱うのが難しかった

出典：Hirashima et al., SC 25 (2025)（画像提供：NASA/JPL-Caltech, ESA, CSA, STScI）

ひとつひとつの星をそれぞれ別に扱う

また、銀河全体の進化を正確に捉えるために、太陽のようなひとつひとつの恒星が周囲に与える影響を調べようとすると、非常に高い分解能が求められます。従来のシミュレーションでは計算負荷を抑えるため、約1,000個の星をひとつの「点」としてまとめて扱う手法が主流でした。また、計算上の大きなボトルネックとなる「超新星爆発の初期段階」については、球対称構造を仮定した物理モデルを用いることで、計算の簡略化を図っていました。

つまり、個別の星をシミュレートするには、これまでの1,000倍という極めて高い解像度が必要になります。この目標に対し、単純に粒子数や並列度を増やすだけでは対応できません。計算密度が高まると並列化効率が著しく低下し、計算時間が年単位に膨れ上がってしまうからです。1,000倍の高解像度化を実現するには、この並列化の壁

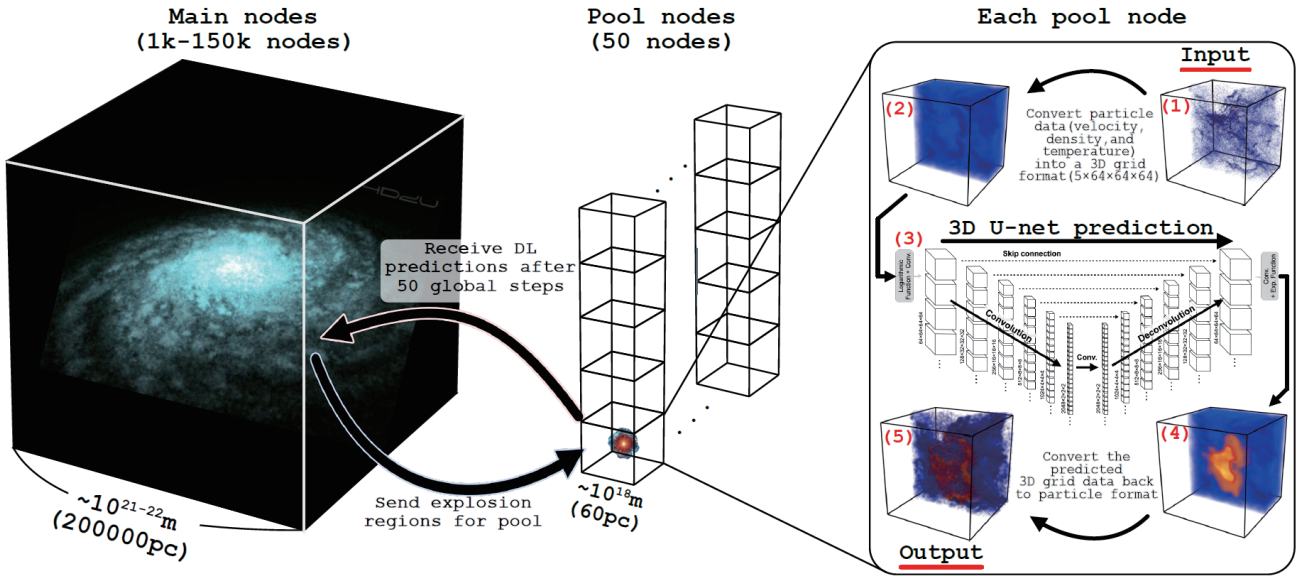


図2 開発した銀河シミュレーションの概念図

出典：Hirashima et al., SC 25 (2025)（画像提供：©2010 Takaaki Takeda, Junichi Baba, Takayuki Saitoh, 4D2U Project, NAOJ）

を打破する新たな計算手法の確立が不可欠です。

我々のグループでは今回、サロゲートモデル（複雑なシミュレーション挙動を機械学習などで予測する）を利用することにしました。先ほどの話に戻りますが、銀河系全体のみ、あるいは超新星爆発のみであれば、これまでも高精度・高効率の計算は可能だったわけです。問題はこの両者を組み合わせることが難しいという点でした。

まず超新星爆発の話をします。超新星爆発が発生すると、その周囲のガスの動きなどを予測するというAIサロゲートモデルを開発しました。このモデルは、爆発が起こると、時間変化に伴いシェル（超新星残骸：爆発によって放出された物質が周囲の星間物質を掃き集めながら膨張することで形成される）がどう膨張していくかを推論するものです。この過程は、以前は流体力学・熱力学の計算などを利用して行っていました。このサロゲートモデルに爆発したというシグナルを送ると、10万年後にはシェルが広がり高温の領域がこう形成され、その際の速度はこんな感じ、というものが出力されます。そして、この推論結果をシミュレーション側に戻すことで、計算時間を大幅に短縮できるようになりました。

現在はCPUで処理
将来はGPUがメインに

当初、このサロゲートモデルを組み込んだシミュレーションは富岳で動作させました。これは富岳のシステム全体、約15万ノードを利用して星や星間ガスなどを示す3,000億個の粒子の計算を行うことで、天の川銀河に属する星のひとつひとつを個別にシミュレーションした初のケースです。

この富岳での動作により、星1個単位の解像度による銀河系シミュレーションの実現可能性を示せたわけですが、だからと言って富岳を1か月とか2か月の間、独占できるわけではありません。ですので、渦状腕構造（渦巻銀河の中心から外側に向かって渦を巻いて伸びる腕状の構造）の解析や銀河内の元素の循環、太陽系や生命の材料の起源解明といった実際のサイエンスに貢献するためには、もう少し小規模なシステムでもこのシミュレーションがきちんと実行できるようにする必要があります。

Miyabiや現在議論されている富岳NEXTは、いずれもGPUがメインとなるシステムです。ですのでCPUで動くコードをGPU側に持ってくることになります。我々のグループでは、惑星形成や粒子の相互作用の計算に利用でき、複数のプラットフォームをサポートできるライブラリも開発してきたため、書き換えも効率的に行えます。

さらに、通信が大きなボトルネックになっている問題もあります。Miyabiで1ノードを利用した時の計算効率は50%近くでしたが、1,024ノードだと10%台まで低下しました。ノード数が増えるとどうしても悪化します。将来的には、こうした効率をもう少し改善して、Miyabiで3か月や半年でシミュレーションを終わらせることで、サイエンス的に有意義な結果が出せるような状況に持っていきたいですね。

平島敬也／専門は天文学、機械学習、HPC。東京大学大学院理学系研究科天文学専攻博士課程修了、博士（理学）。日本学術振興会特別研究員DC1などを経て、2025年より現職。

深く学ぶには

The First Star-by-star N-body/Hydrodynamics Simulation of Our Galaxy Coupling with a Surrogate Model
https://dl.acm.org/doi/10.1145/3712285.3759866

気象分野における HPCとAIの 融合の方向性

国立環境研究所
環境情報部
研究情報室
室長代行
八代尚

東京大学
情報基盤センター
学際情報科学研究部門
准教授
本田匠

GPUを中核としたMiyabiの構成は、HPCの世界にAIを組み合わせることを可能にした。ただしそのためには、使う側の対応も必要だ。気象シミュレーションの世界にどうAIを取り込んでいくかについて、八代尚博士と本田匠准教授にお話を伺った。

気象シミュレーションは メモリ帯域依存

これまでの気象シミュレーションについて簡単に説明すると、メモリ帯域に非常に依存したものです。流体系の計算では、ある1点の計算を行うにあたり近傍の値を全部取り込む必要があります。このため大量のメモリアクセスが発生します。例えば初代の地球シミュレータはBF比（Bandwidth/FLOPS比：メモリ帯域と演算性能の比）が4と非常に広帯域だったのですが、2代目になると2.5に、京コンピュータはギリギリ1でしたが、富岳は0.3ぐらいまで下がっています。結果としてモデルの演算効率（理論性能に対する実性能比)は落ち続けています。

実は気象シミュレーションというのは非常に学際的なプログラムの集まりでして、アルゴリズムが1種類ではないのです。1点に対するステンシル計算をするだけではなく、大気の放射を上層から順に解いていくとか、雲の相変化量を計算するとか、すべてアルゴリズムが異なり、それぞれに最適化の方法も違います。おそらく科学系のシミュレーションの中では、一番しんどい部類かと思います。

それでも京コンピュータや富岳の時も、いろいろ工夫することで何とか性能を出すことができました。地球シミュレータから京コンピュータになる時が約100倍、京コンピュータから富岳に移るのも100倍以上の性能になりました。ただ京から富岳では、ハードウェアの性能だと40倍程度の向上しかありません。しかしモデル全体の精度をFP64からFP32に落とすことで、約1.6倍の高速化ができました。

さらにメモリ転送だけしているような部分を極力減らすようにブラッシュアップを積み重ねることで、何とかトータル100倍以上を達成できました。

ただ、そろそろやれることがなくなってきました。次の富岳Nextにしても FP64の性能は富岳の数倍程度でしかありませんし、コアそのものの性能は上がらないかむしろ落ちていて、ただコアの数はずっと多い。ですので処理の並列度を引き上げないといけないのですが、それもやはり限度があります。

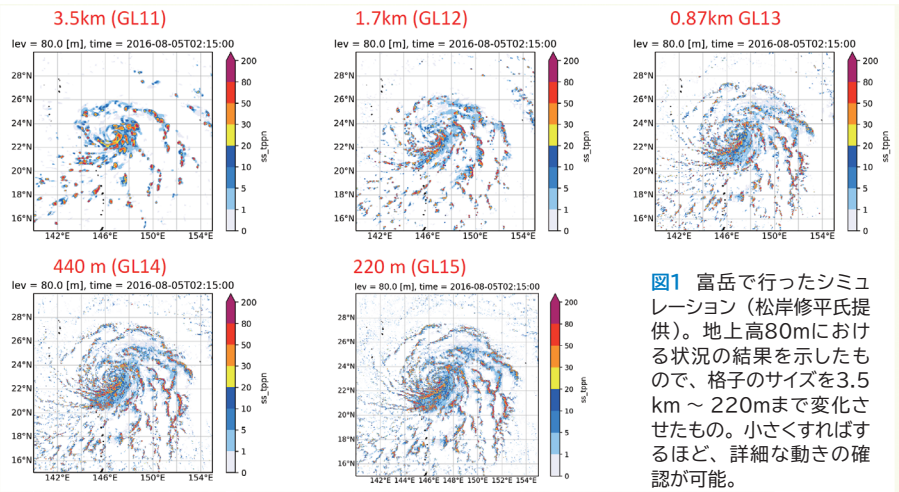


図1　富岳で行ったシミュレーション（松岸修平氏提供）。地上高80mにおける状況の結果を示したもので、格子のサイズを3.5 km ～ 220mまで変化させたもの。小さくすればするほど、詳細な動きの確認が可能。

アルゴリズムの計算を 推論で代替できるか

実は富岳への移植を始める前から、気候モデルのGPUへの移植を進めています。ただ、これをMiyabiやその前のWisteria-A（Aquarius）で動かすと劇的に速くなるわけではない。要するにGPUに搭載されているHBMメモリの帯域で性能が決まってしまっている。ソフトウェアの最適化も富岳への移植の際にやり切った感がある。ではこれ以上性能を上げるにはどうす

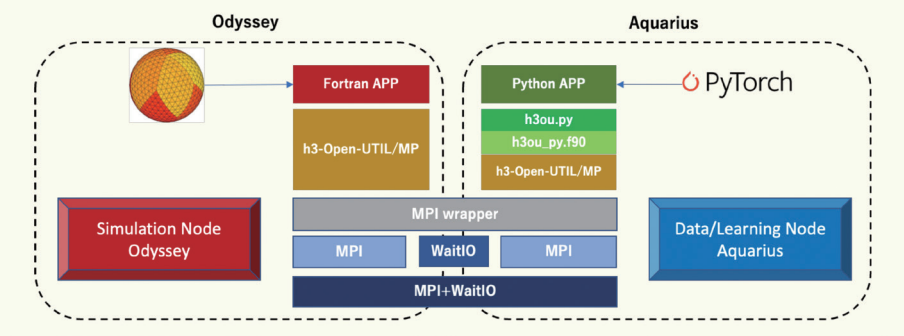


図2 Wisteriaを利用した際には、CPUノードであるOdysseyで従来のFortranで記述したシミュレーションを稼働させ、GPUノードであるAquariusでPythonで記述されたサロゲートモデルのその場学習と、実行試験を行った。異なるスパコンシステム、異なるアーキテクチャのノード、異なるプログラミング言語の間の連携は、東京大学が中心となって開発したカップリングライブラリと通信ライブラリを用いている。
出典：Arakawa, T., Yashiro, H., Sumimoto, S., Yamazaki, K., Nakajima, K., AI for Science on Modular Supercomputers: Performance Evaluation of Coupled Atmospheric-AI Workflows, SCA/HPCAsia 2026 (Poster)

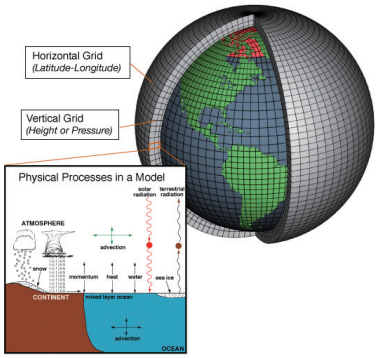


図3 単純に気象と言っても、実際にはさまざまなプロセスの影響をかなり受けるため、それぞれの影響を個別のモデルを利用してシミュレーションし、結果をまとめる必要がある。
画像提供：NOAA

るか？ ここで出てくるのがAIです。

先程も触れましたが、我々の気象モデルはさまざまなアルゴリズムに基づく計算を組み合わせているわけですが、その個々のアルゴリズムの計算をサロゲートモデル（p.7参照）を利用して、推論で代替できたらいいね、という話です。問題は、そこでどの程度の精度を担保できるのか、それをどう検証してゆくのか。それと、あまりブラックボックスの比率を増やしてしまうと、我々は何を研究しているのか分からなくなってしまいますので、全部をブラックボックス化したくはありません。ですので実際には、精緻ですが遅い(計算ベースの)モデルと、(サロゲートモデルによる)模倣するだけですが高速なモデルの両方を使い分けることが必要だろう、と考えています。

ただAIのモデルの弱点は外挿です。学習したものではない、その外側にある状況をうまく予測できないのですが、気候シミュレーションはそういう外挿が多い分野です。現在進行形で進む地球温暖化に伴い、気候は急速に変化しています。この1,000年で人類が経験したことのないような現象を外挿しようとした場合のフルサロゲーションは危険です。現在、向こう5日間とか2週間とかならそれなりの精度で計算できますが、100年後とかですとおかしな結果しか出ないでしょう。実は最近のAI気象モデルのほとんどが欧米の気象機関が作成した同じデータを教師データに用いていまして、これで構築したモデルを使うことで、計算をした場合と比べて1,000倍ぐらい高速になります。ただ、現行の手法で予測できるのは限界があり、その先になると方法を変えないといけません。

シミュレーションにAIを入れる形へ

そこで現在取り組んでいるのは、コンポーネントベースのサロゲーションです。大気の放射の部分だけとか、雲の微細な部分とかを、元のサブモデルの良いところも悪いところもひっくるめて忠実に再現するモデルを構築し、全体計算の中でのそれらの組み合わせによって複合効果やフィードバックを表現しようというやり方です。気象モデルなどは一般にデータ同化（シミュレーション結果を現実のデータと比較、パラメータを修正してより現実のデータに近づける）という手法で予測精度を高めているのですが、サロゲートモデルの学習のプロセスの中にデータ同化を含めてフィードバックすることで今後精度を高めることも考えています。

シミュレーションを行うためには初期データが重要です。試験的に運用されている次世代気象レーダーがあるのですが、この観測をデータ同化した初期値を使うと、ゲリラ豪雨を捉えるシミュレーションが可能になるかもしれません。これまでは得られた観測のデータ同化によってシミュレーションの初期値を準備していたわけですが、最近はデータ同化を経ずに観測データからAIで直接予測を行うアプローチも研究されています。

AIの学習データはシミュレーションとデータ同化から作られており、その意味でAIはシミュレーションに依存しています。今後、シミュレーションに頼らないAIも開発されていくかもしれません。AIが観測データから自分で学習データを作り、その学習データによって自身をさらに強化し、より良い学習データを作る。つまり、AIがシミュレーションに頼らずに自己完結するフィードバックを回すわけです。

まだまだやることが山積みですが、まずはサロゲートモデルのもう少し広範な利用や、データ同化にAIを活用する方法を模索しながら、従来コードのGPUへ向けた最適化などを進めていきたいと考えています。

八代尚／専門は大気科学、高性能計算、大気化学。東北大学大学院理学研究科地球物理学専攻修了、博士（理学）、理化学研究所計算科学研究センター 研究員などを経て、2019年から現職。

本田匠／専門は気象学・データ同化。九州大学大学院理学府地球惑星科学専攻博士課程修了、博士（理学）。理化学研究所計算科学研究センターデータ同化研究チーム 基礎科学特別研究員などを経て、2024年から現職。

2024年10月、Microsoftから東京大学にAzureを利用するためのクレジットが寄付されました。これにより、東京大学では一定量まで無償でMicrosoftのAzureというクラウドサービスを利用できる制度が現在運用されており、研究や授業に活用されています。ここでは、UTokyo Azureと名付けられたこの制度の目的や使い方などについて、東京大学情報システム本部の中村誠講師にお話を伺いました。

nodesの光明：情報基盤センターサービスの裏側

AI研究を推進させる「UTokyo Azure

UTokyo Azureを自由に使ってほしい

UTokyo Azureは、2023年8月にMicrosoftと東京大学の間でGX（Green Transformation）とD&I（Diversity&Inclusion）、AI研究の推進に向けた連携に関する基本合意書が締結されたことに端を発します。この時は東京大学の藤井輝夫総長とMicrosoftのサティア・ナデラCEOが面談して連携が決まったのですが、その後2024年4月に行われた日米首脳会談の中で、今後5年間Microsoftが東京大学と慶應義塾大学、カーネギーメロン大学のAI研究に対し、それぞれ1,000万ドルの資金を提供することが決定され、2024年10月から2029年9月まで、Microsoft Azureのクレジットが東京大学に対して寄付されることになりました。

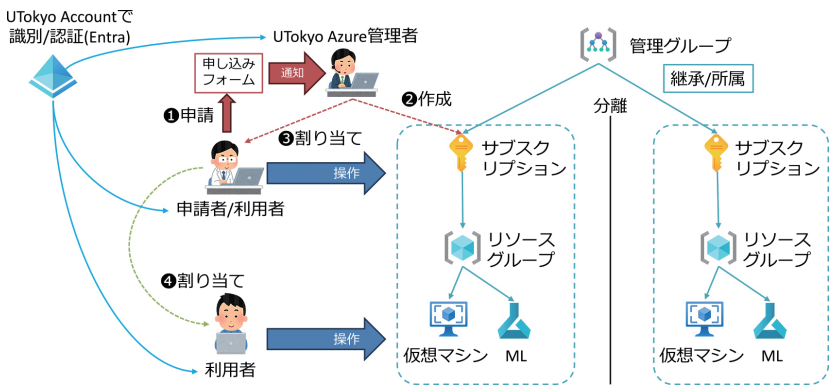
そんなUTokyo Azureですが、東大としては自由に使っていただきたいというスタンスです。GXにしてもD&IにしてもAIにしても、今時のITインフラを活用して、何か新しいことを生み出してみましょうという側面が強いと思うんですよ。現状で申し上げれば、AIはもう道具になってきていて、皆さんお使いになられる。多分AIの研究をなさっている、あるいはAIを利用して何らかの研究を行っておられる方はUTokyo Azureを沢山使われていると思うのですが、もう少し気軽に使っていただきたいということで、間口を広く取っています。研究計画に従ってという使い方よりも、まずは使っていただいて、実際にうまくいったら、それを発展的に研究してもらえばいいと。

中村誠／専門は計算機システム・ネットワーク。東京大学大学院修了後、現職。

Azureで提供されているサービスはそのまま使える

実のところ、UTokyo Azureといってもユーザーから見ると、通常のMicr

図1 ユーザー登録の流れ



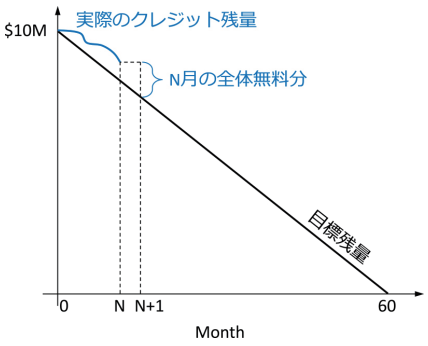


図2 クレジット管理のイメージ

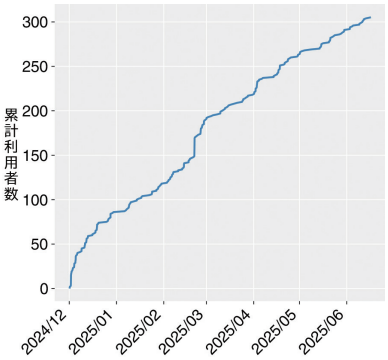


図3 第1回のAzure利用説明会は2025年2月後半に実施された

サービス」の導入と運用

osoft Azureとまったく同じです。自分で仮想マシンをデプロイ（配置）して、というプリミティブな使い方はもちろん可能ですが、最近ですとAzureの側に色々なサービスが最初から用意されています。仮想マシンであればLinuxの各ディストリビューションやWindows系OSが用意されていますし、AI系で言えばGPTなどはすぐに利用可能になっています。

とくにOpenAIのサービスに関しては、MicrosoftがOpenAIと契約をしている関係で、最新のモデルもすぐに利用できます。ほかにもAzureで提供されているストレージやファイルシステムのサービスなどもそのまま利用可能です。基本的な方針として利用範囲を狭めるような運用はしませんが、利用料が高額になってしまいそうなものに関しては、利用を制限するなど、一定の調整を行う場合があります。

我々情報システム本部がUTokyo Azureをどう運用しているのかという話になりますと、正確には情報基盤センターと情報システム本部と共同で運用を行っているのですが、利用者は、利用料金が発生した際に支払いの責任が取れる教職員に限らせていただいています。ただ、自分の研究室の学生に使わせることは可能で、この場合は指導教員が支払いに責任を持つ格好ですね。登録されたら、UTokyo Accountという東京大学の情報システムを利用する際に使われる全学的なアカウントがあるのですが、このUTokyo AccountにUTokyo Azureのサブスクリプションをひも付けする形で利用権を付与するので、あとはお好きなように使えます（図1）。

利用説明会を通じて うまい使い方を共有

もうひとつ、クレジットの分配とその動向の監視という作業があります。先に触れましたが、総額1,000万ドル分のクレジットを5年間、60カ月で使う形になるわけですが、これをユーザーに割り振るにあたって、不公平が起きないような制度設計を行いました。ただ、中には一時的に大量の計算資源を必要とするケースもあるので、そうした場合には柔軟に対応できるように配慮しています。その上でクレジットの残量管理を行っています。

図2がその構図ですが、黒い線がクレジットの残量目標、青い線が実績ですね。クレジットの残量が目標よりも多い場合にはそもそも利用者への課金が発生しないわけですが、それでも余った場合には翌月に繰り越し、次の月の無料利用分が増えます。逆に目標残量を下回る場合には、無料利用分を超過した分の課金が利用者に発生し、それが全体のクレジットに充当されるような形にしています。

こんな形で、ユーザーがどれだけAzureのリソースを消費しているのか、というのは追跡できているんですが、その先の「どんな使い方をしているのか」に関しては、追い切れていません。サブスクリプションは2025年6月の時点で305、現時点（2026年1月）ですとおそらく400ぐらい発行しています。割合ですと情報科学や情報工学分野が全体の4分の1を占めるくらいで多いのですが、分野そのもので言えば57に渡っており、惑星科学や天文学などの理学分野、電気電子、原子力、航空宇宙などの工学分野だけでなく、文学や法学、心理学といった人文や社会科学系の分野でもAzureの利用が試行されています。ただ、こうした分野でAzureがどう使われているのかまでは把握できていません。今後、利用説明会を行う際にヒアリングしていければと考えています。

この利用説明会ですが、2025年度は2回実施しました。Azureを使うことに慣れておられる先生ばかりではないので、使い方であるとか、あるいはどんなことができるのかということを認知していただくのが目的なのですが、実際にサービス開始からのサブスクリプションの数を追いかけていると、説明会を行った直後にサブスクリプションが急増していまして（図3）、やはり説明会は重要と考えています。

また、こうした説明会の中ですでに使われている先生に、Azureのうまい使い方を披露していただき、それをみんなで共有していくことが一番重要と考えています。まだ日付などは決まっていないのですが、2026年も2月か3月に説明会を開催する予定です。この説明会などを通して、UTokyo Azureのリソースを先生方の研究にうまく役立てていただきたいというのが管理を行う側からの希望です。

飛翔するnodes：研究ピックアップ

2025年8月、東京大学情報基盤センター スーパーコンピューティング研究部門の山崎一哉助教の「GPU化済の気象コードにおけるホストCPUコアの有効活用」という論文が、情報処理学会の2025年度コンピュータサイエンス領域奨励賞を受賞しました。この論文について、山崎助教にお伺いしました。

Miyabi-GのGPUとCPUを使いこなし気象・気候シミュレータ「SP-MIROC」を高速化

東京大学情報基盤センター
スーパーコンピューティング研究部門
助教
山崎一哉

気象シミュレーションは膨大な計算が必要

気象庁が線状降水帯の予測に利用するような詳細な気象シミュレータというのは、空間を南北1km、高さはそれ以上に細かいグリッドで区切り、細かい雲の動きなどを秒単位で、グリッド同士の相互作用も含めて計算していくので、猛烈に時間がかかります。短期間の予測だとこうしたものが必要になるわけですが、地球温暖化の予測など、もっと長期間のシミュレーションだとここまで細かい格子での計算は不要です。グリッドのサイズも100kmなどと大きくなり、数十分といった長めのステップで計算を行います。こうした長期間のシミュレーションでは、局地的な現象は計算できない代わりに、大域的なエネルギー収支は極力厳密に再現できるよう工夫されています。

今回私の開発したSP-MIROCというシミュレータは、もともと全球気候モデルのMIROC6という、大気・海洋・陸地・雪氷などの変化を考慮して、流体力学・力学・化学・物理学・生物学などの方程式を用いて地球の気候を再現し、気候の変化を表現するモデルと、SCALE-RMという特定の領域の大気の状態を細かく表現するモデルの連携を行う、ハイブリッド型のものです。こうしたハイブリッド型のシミュレーションは、地球全体とまでは言いませんが、熱帯での対流活動の振る舞いや細かい雲の動きとか、そうした研究に向いています。

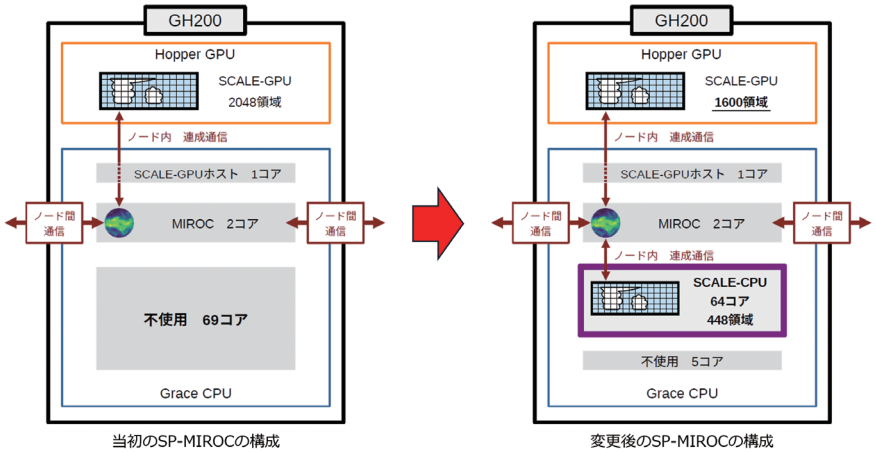
SP-MIROCの場合、まず128kmで区切ったグリッドに対し、600秒ステップでシミュレーションの計算を行います。これはMIROC6による、いわゆる粗いシミュレーションですね。ただし、このひとつのグリッドをさらに2kmに細かく分割し、その中は12秒という短い時間間隔で計算するのですが、ここをSCALE-RMで行います。グリッド同士の相互作用は粗いシミュレーション側で行っているのでSCALE-RM側では省くことができ、またSCALE-RM側では次元軸をひとつ落とすという工夫で計算量を減らしていますが、それでもMIROC6と比較して2～3桁は計算量が多くなります。

このMIROC6による粗いシミュレーションと、SCALE-RMによる個々のグリッド内の細かなシミュレーションは同一の時間軸で進んでいきます。つまり、粗いシミュレーションで1ステップ600秒分の計算をするのと並行して、SCALE-RMの細かいシミュレーションも600秒分に相当する50ステップの計算を進めないといけません。計算量が大きくなる理由はここにもあります。

Grace CPUを活用することで性能向上

このSP-MIROCは、東大と筑波大共同のMiyabiというスーパーコンピュータ上で実行するものです。Miyabiの中にはMiyabi-Gという演算加速ノードがあり、これはNVIDIAのGH200という、Grace CPUとHopper GPUがひとつのパッケージになったものをベースとしたシステムですが、このMiyabi-GのHopper

図1 これまで使っていなかったGrace CPUの活用で15～25%の性能向上を達成



GPUを主に利用していました。一般にこうしたCPUとGPUの組み合わせは、CPUとGPUの通信に待ち時間が発生するため、当初SP-MIROCでは、図1の左側のようにSCALEの重い計算を全部Hopper GPUに行わせ、72コアのGrace CPUの方は、計算が軽いMIROC部分用などに3コアが稼働しているだけでほぼ遊んでいました。ところが、遊んでいたGrace CPUを活用するようにしたことで（図1右側）、15 ～ 25%の性能向上が達成できたというのが今回の論文の骨子です。

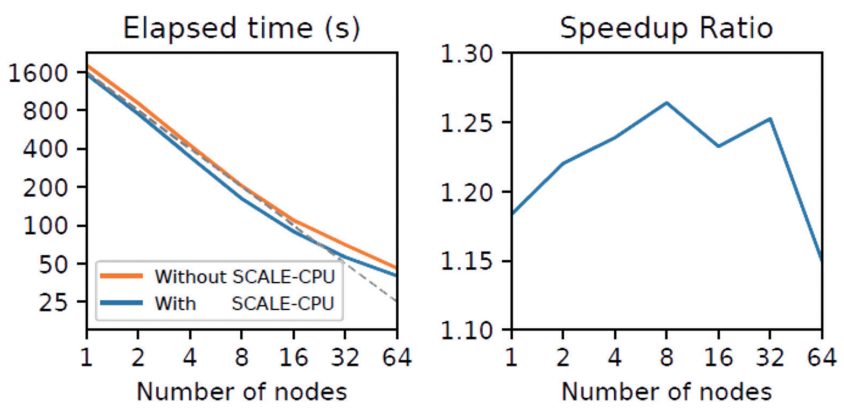
今回は全部で8,192通りのシミュレーションを4つのノードに分散させる形で、ノードひとつあたり2,048通りの計算をさせていますが、これを1,600と448に分散させ、GPUとCPUでこれを処理する形になります。この比率ですが、GPU側の計算処理が終わるちょっと前にCPU側の計算処理を終わらせて、GPUからの演算結果を受け取って後処理ができる、というタイミングになるように調整した結果です。そのため、ほぼこれがSCALE-RMの処理におけるCPUとGPUの性能比ということになります。

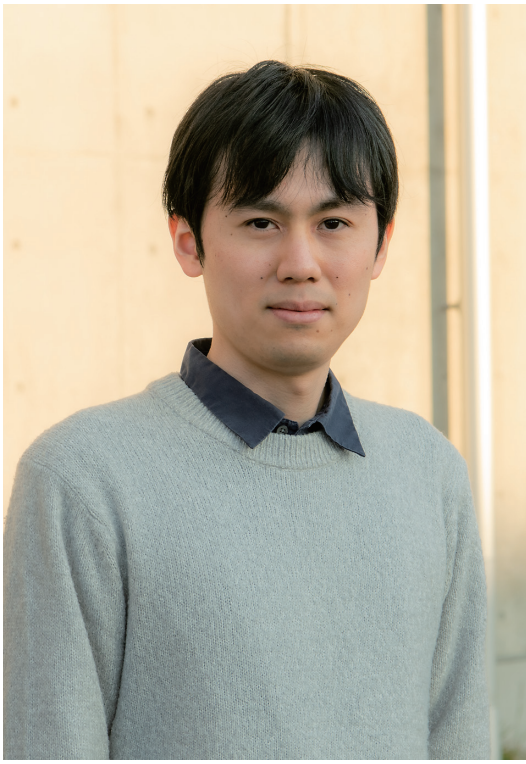
計算性能の理論値を比較すると、Hopper GPUとGrace CPUのベクトル倍精度演算性能は10倍の開きがあります。またメモリ帯域は8倍の性能差があります。にもかかわらず、SP-MIROCを実行した場合、CPUとGPUの性能比は448:1,600、約4倍の性能差ということになります。SCALE-RMでは煩雑にメモリアクセスを行う必要があるので、性能は演算性能というよりもメモリ帯域がボトルネックになるのですが、SCALE-RMでは扱うべきデータの量がそれほど多くないため、キャッシュにかなりの部分が収まってしまうようなのです。このため実際にはキャッシュのアクセス速度がボトルネックになっており、これが4倍程度の性能差という結果になり、結果としてトータルでの演算性能を25%ほど引き上げることができました。

消費エネルギーの最適化も今後のミッション

もっともこの25%の性能向上という結果は最良の条件での数字です。図2は同時に利用するノード数を増やしていった場合の所要時間と、その際の性能向上率を示したものです。左は純粋に所要時間、右はCPUを利用した場合の性能を、しない場合と比較したものになります。先に示した25%というのは8ノードで構成した場合の数字であり、16ノードや32ノードでもそれなりに性能は出ますが、64ノードだと急激に効率が落ちているのが分かります。また1 ～ 2ノードの場合も性能向上率は芳しくありません。

図2 同時に利用するノード数を増やした場合の所要時間とその際の性能向上率





山崎一哉／専門は気象学（数値モデリング・メソ気象・衛星画像解析）、高性能計算（シミュレーションにおけるGPUの活用）。東京大学大学院修了後、2023年より現職。

また消費電力で言えば、CPUを利用しない場合は1ノードあたり600W以内で収まっているのですが、CPUも併用すると、消費電力が最大800W程度まで増えます。計算自体が高速化されるので、計算全体で積算した消費電力量はここまで大幅には増えませんが、それでも、1 ～ 32ノードの利用時はトータルの消費電力量が5 ～ 10%増加しています。これはCPUの電力効率がGPUに劣るためで、この消費電力の効率が下がるという問題も今後の課題と言えます。

現在の構成ですと、1日分の計算が1ノードで1,600秒、32ノードで1分前後で処理できていますが、実際には何年にもおよぶ影響といったものを計算するので、もっと高速化できる方法を検討していきたいと思います。また、今回は演算時間を最小化するという試みでしたが、今後は消費エネルギーも含めて最適化できる方向を探っていきたいと考えています。

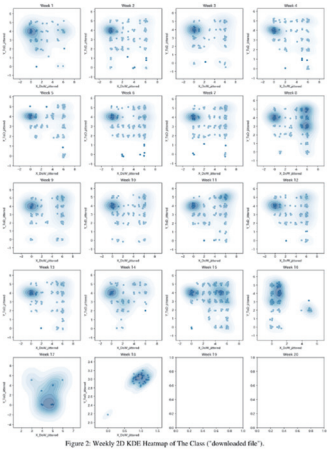
学習の「見える化」を目指して

大山智也（情報メディア教育研究部門 講師）

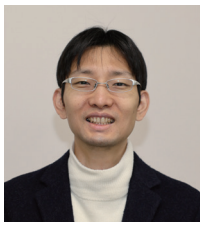
ラーニング・アナリティクス（LA）という言葉をご存じでしょうか？ 主にオンラインで提供される教材や、学習用アプリから得られるログを解析し、授業からの脱落の兆候や、高い学業成績につながる学習行動を見つけようとする取り組みです。オンラインでの学習が当たり前になった今日、LAを通じて学習者を支援したり、教育内容を改善しようという動きは盛んになっています。一方で、日本の高等教育ではこうした取り組みが進んでいるとは言い難い状況で、本学も例外ではありません。

そこで、情報メディア教育研究部門では、学習管理システム「UTOL」を運用しながら、日々蓄積されている学習者の操作ログを分析し、有用な知見を見つけ出す試みに着手しています。とはいえ、まずはデータ基盤構築やポリシー策定に取り掛かっている段階で、こうした安全な仕組みづくりを経て、ようやくスタートラインに立てるところです。

まだ青写真ですが、学生が自身の学習活動状況をモニタリングしたり、講師が担当クラスの学生の授業への関与度合いをチェックできるツールの開発なども視野に入れています。客観的な指標により気付きを得ながら学ぶ／教えられる未来を目指します。



学習行動の可視化の例：横軸は曜日（授業日← →課題提出日）、縦軸は時間帯。週によって活動の集中する時間帯が異なり、例えば8週目では提出日時間際での活動が目立つ。引用元：Ohyama, T. & Mitsuishi, T. (2024), https://www.learntechlib.org/primary/p/224558/.



大山智也／専門は行動データサイエンス、学習分析など。2020年3月に筑波大学システム情報工学研究科博士後期課程を修了、博士（社会工学）。JSPS DC2／PD、筑波大学システム情報系、東北大学データ駆動科学・教育研究センターでの助教を経て、2024年4月より現職。

nodesのひろば

学生と作る、ユーザのための”ひとつの窓口”

―金子亮大（情報システム本部 学術専門職員）

情報システム本部の金子です。いきなりですが、情報システム本部は情報基盤センターとは別の組織です。とはいえ、その違いを説明できる人は学内にも多くないでしょう。実際、利用者にとってこの違いは大きな問題ではありません。

情報システム本部は全学共通の「UTokyo Account」や、Microsoft 365、Zoomなどを提供しています。一方、例えばUTokyo Accountで利用できるGoogle Workspaceは情報基盤センターの担当です。歴史的な経緯で分かれていますが、利用者に必要なのは情報が一元的にまとまっていることです。この考え方から、ワンストップで情報提供とサポートを行うポータルサイト「utelecon」が2020年に生まれました。

私の主な業務のひとつは、このutelecの運営のお手伝いです。特に重要なのが、運営に参加してくれる学生アルバイトの皆さんとのコミュニケーションです。現在150名以上の学生が、記事作成やサポート対応に日々尽力しており、私は働きやすい環境の整備やスキル向上のための支援として、個々の状況を踏まえた勤務ルール作成や、ライティング研修などに取り組んでいます。学生の皆さんとの協働を大切にしながら、uteleconの活動の価値をさらに高めていきたいと考えています。

半年に一度のutelecon関係者の交流会の様子。この日は学生・職員・教員が入り混じって「未来の東京大学はどうあるべきか」というテーマで大いに語り合いました。

金子亮大／専門は人文地理学、地理情報システム（GIS）。東京大学大学院総合文化研究科在学中にutelecon の活動に参加したことが縁で、2025年4月より現職。趣味は観世流の能の謡と仕舞の稽古。

AI時代に挑む研究情報の統合 —— 古宇田光（情報基盤センター長付 高度学術専門職員／プリンシパルURA）

古宇田光／1988年よりNECで材料・デバイスの研究とマネジメントに従事。2010年より東京大学物性研究所で国家プロジェクトである「京」「富岳」「元素戦略」などの事業のマネジメントを担当。2019年より東京大学リサーチ・アドミニストレーター推進室で研究力強化関連業務を実施。

大学には、研究や教育を進めるだけでなく、新しい分野を切り開き、地球環境や社会の課題を解決する役割があります。こうした課題は、ひとつの研究分野だけでは対応できません。本学には多様な研究部門があり、AI時代に必要な情報も豊富です。しかし、部門ごとに情報が分かれていて、うまくつなぐのは簡単ではありません。特に、公開できない情報や本学独自の情報を安全に扱うには、着実なルールや体制とシステム整備が必要です。

私はリサーチ・アドミニストレーター推進室で、研究情報を集めるシステム開発を研究推進部と共に進めています。並行して、集めた情報を安全に蓄積・共有し、活用するシステムを開発するには、情報基盤センターの力が欠かせません。そこで、2024年度から情報基盤センターの兼務となり、情報システム部や図書館とともに、教員と職員が協働する一回り大きな開発体制の構築を支援しました。現在はその事務局として、情報を安全着実に蓄積し分類するシステム、検索や分析で価値を生み出すシステムの開発に携わっています。この挑戦を通じて本学の情報から新しい価値を生み出し、社会に貢献できることを心から楽しみにしています。

しまなみ海道「来島海峡大橋」。橋のように研究情報をつなぎ未来への道を拓く。

がり

コンピュータが学習進化する？ −医療も宇宙もその先へ —— 髙玉圭樹（学際情報科学研究部門 教授）

髙玉圭樹／専門は進化計算、機械学習、マルチエージェントシステムに関する研究と応用（医療・ヘルスケア、宇宙・交通）。1998年東京大学大学院博士課程修了、博士（工学）。国際電気通信基礎技術研究所（ATR）、東京工業大学講師、電気通信大学助教授、教授を経て、2024年より現職。GECCO 2018のGeneral Chair。著書に「マルチエージェント学習 −相互作用の謎に迫る−」など。

我々を取り囲むマクロな世界も個々のミクロな体調も日々変化しているため、それらの変化にスマートに対応することが求められています。しかし、「今日の最適解」は「明日の最適解」とは限らない（むしろ異なることの方が多い）ため、過去のデータを用いて学習する深層学習や生成AIでは限界があります。そこで、これらのブラックボックス問題に対処するAI技術（学習進化技術）を研究しています。

左は人を対象、右は人工物を対象とした研究の応用例を示しており、それらに必要となる進化計算・機械学習に基づくブラックボックス最適化を探究しています。

例えば、「どうすれば健康になり、日中の生産性を向上できるのか？」の答えは不明ですが、介護施設で生活する高齢者、会社で働く従業員にとって欠かせない問題です。そこで各人のライフスタイルを設計し、パフォーマンスの最適化や、睡眠状態を改善することで癌や新型コロナの予防を目指しています。また、人工物の場合、「予期せぬことが起こる宇宙や災害にどう対処すればよいか？」の答えも不明ですが、どの事態でも宇宙機や物資輸送は機能し続けないといけません。そこで、プログラムが進化する人工衛星や、航空機・船・車の輸送方法を探究しています。2024年には、JAXAと共同で開発した宇宙探査機SLIMを打ち上げ、月面着陸に成功しました。

nodes ［ノーズ］vol.5

2026年3月発行

編集・発行

東京大学情報基盤センター

〒277-0882 千葉県柏市柏の葉6-2-3 柏II キャンパス内
https://www.itc.u-tokyo.ac.jp
©Information Technology Center, The University of Tokyo
ISSN 2436-9543（Print）、ISSN 2436-9551（Online）

本誌へのご感想、ご要望をお寄せください。
印刷版の送付のご要望もこちらで承ります。
https://www.itc.u-tokyo.ac.jp/public/nodes/

［表紙写真］
最先端共同HPC基盤施設（JCAHPC）の新スーパーコンピュータシステム Miyabi

［裏表紙写真］
x86ベースのMiyabi-Cが収納されているラック

東京大学情報基盤センターの最新情報を随時発信中！

Facebook　Instagram　X　YouTube