jh200037-NAH

高性能・変動精度・高信頼性数値解析手法とその応用

中島研吾（東大情報基盤センター）

概要　本研究は，最先端のスパコン向けに開発された高性能数値アルゴリズムに対して，半精度から倍精度，倍々精度までの広範囲をカバーする変動精度演算を適用し，精度保証，そのための自動チューニング手法を開発する。開発された手法を様々なアプリケーションに適用することで，低精度を中心とした変動精度演算の科学技術シミュレーションへの有効性を検証する。開発したアルゴリズム，アプリケーションの消費電力の直接測定によって，各計算の特性と低精度演算の有効性を消費電力の観点から検討する。

1. 共同研究に関する情報

(1) 共同研究を実施した拠点名

北海道大学　東京大学　東京工業大学
名古屋大学　九州大学

(2) 共同研究分野

超大規模数値計算系応用分野

(3) 参加研究者の役割分担

- 中島研吾（東大）（代表）全体統括・高性能アルゴリズム
- 市村　強（東大）（副代表）高性能アルゴリズム・地震シミュレーション
- 横田理央（東工大）（副代表）高性能アルゴリズム・精度保証
- Sameer Deshmukh, 大友 広幸, Peter Spalthoff,Qianjiang Ma , Muhammad Ridwan Apriansyah（東工大）H 行列精度保証
- 岩下武史（北大）高性能アルゴリズム
- 深谷　猛（北大）高性能アルゴリズム
- 池原 紘太（北大）混合精度アルゴリズム
- 塙　敏博（東大）性能最適化・電力測定・FPGA
- 伊田明弘（東大）高性能アルゴリズム
- 星野哲也（東大）高性能アルゴリズム
- 坂本龍一（東大）電力測定・AT
- 有間英志（東大）電力測定・AT
- 古村孝志（東大）地震シミュレーション
- 藤田航平（東大）高性能アルゴリズム
- 近藤正章（東大）精度保証・電力測定・AT
- 奥田洋司（東大）工学シミュレーション
- 森田直樹（東大）工学シミュレーション
- 荻田武史（東女大）精度保証・AT
- 田中一成（早大）精度保証・AT
- 尾崎克久（芝工大）精度保証・AT
- 片桐孝洋（名大）精度保証・AT
- 山梨祥平（名大）非構造ステンシル問題 AT
- 八代　尚（環境研）大気シミュレーション
- 河合直聡（理研）高性能アルゴリズム
- 井上弘士（九大）精度保証・電力測定・AT
- 荒川　隆（RIST）大気シミュレーション
- 成瀬　彰（NVIDIA）精度保証・AT
- 堀越将司（インテル）精度保証・AT
- 大島聡史（名大）高性能アルゴリズム
- Gerhard Wellein (FAU Erlangen-Nürnberg, Germany)高性能アルゴリズム
- Achim Basermann (DLR, Germany) 高性能アルゴリズム
- Osni Marques (LBNL,USA) 高性能アルゴリズム, AT

2. 研究の目的と意義

エクサスケールシステムにおける高性能数値アルゴリズム実現には，メモリ・ネットワークの階層の深化に対応した通信最適化（Serial，Parallel）

とともに省電力・省エネルギー（以下「省電力」）に向けた検討が必要である。Approximate Computing（S. Mittal, A Survey of Techniques for Approximate Computing, ACM CSUR 48-4, 2016）は，低精度演算の積極的活用により計算時間短縮，消費電力削減を図る試みであり，従来は画像認識等の計算精度の要求されない分野を対象としていたが，昨今は数値計算において半精度から四倍精度まで演算精度を動的に変動させる変動精度（Transprecision）の研究が進められている。数値計算による近似解（数値解）は様々な計算誤差を含み，計算結果の信頼性の観点から，数値解の正しさを数学的に保証する必要があり，低精度・変動精度使用時，悪条件問題には重要であるが，実問題で現れる大規模疎行列・H 行列への応用例はほとんどない。本研究では，JHPCN システム群の中で消費電力当たり計算性能（GFLOPS/W 値）の高いシステムを主たるターゲットとして，以下を実施する：

- 疎行列演算，H 行列演算，ステンシル演算等の代表的数値アルゴリズム，各アプリケーション（地震，大気科学，量子科学，構造力学）について，Serial・Parallel 通信最適化に着目した高性能最適化手法を各システムにおいて実装し，低精度演算・変動精度演算について検討し，消費電力を測定する。
- 疎行列演算や H 行列演算を対象として，特に悪条件問題における実用的な精度保証法を確立する。更に前項の各アルゴリズム，アプリケーションについて所望の結果精度達成という条件下で，計算時間や消費電力を最小化する最適な演算精度を自動チューニング技術によって動的に制御する手法を確立する。
- 本研究によって開発された高性能・変動精度・高信頼性数値解法を，自動チューニング機構を有するアプリケーション開発・実行環境 ppOpen-HPC（JST-CREST 2011-2018，https://github.com/Post-Peta-Crest/ppOpenHPC）及び（計算＋データ＋学習）融合を実現する革新的ソフトウェア基盤 h3-Open-BDEC（科研費基盤 S 2019-2023，http://nkl.cc.u-tokyo.ac.jp/h3-Open-BDEC/）に実装し，東大 Oakforest-PACS（OFP），東大 Reedbush（RBH，RBL），東大 Oakbridge-CX（OBCX），東工大 Tsubame-3（TSB3）で公開する。将来的には「富岳」も含む他のセンターのスパコンへの導入も視野に入れる。

本研究は，最先端のスパコン向けに開発された高性能数値アルゴリズムに対して，半精度から倍精度，倍々精度までの広範囲をカバーする変動精度演算を適用し，精度保証，そのための自動チューニング手法を開発する試みとしては初めてのものであり，開発された手法を様々なアプリケーションに適用することで，低精度を中心とした変動精度演算の科学技術シミュレーションへの有効性を検証できる。開発したアルゴリズム，アプリケーションの消費電力の直接測定によって，各計算の特性と低精度演算の有効性を消費電力の観点から検討可能となる。

3. 当拠点公募型研究として実施した意義

JHPCN は多様な計算機環境を備え，実用的なシステムとして最も GFLOPS/W 値の高い OFP, RBH, RBL，OBCX，Tsubame-3 等の大規模システムを有し，本研究の目指す高性能・変動精度・高信頼性数値解法の研究には最適である。RBL，OBCX では「ノード固定」における設定カスタマイズにより，個別ノードの消費電力測定が可能である。JHPCN は様々な分野の専門家を擁し，本研究のような学際的研究を推進する体制を容易に構築でき，北大，東大，東工大，名大，九大各センターから様々な分野の研究者が参加している。JHPCN 各センターはオープンソースソフトウェア活用に積極的であり，本研究の成果を公開，各センターのスパコンにデプロイし，講習会等の普及活動を協力して行い，利用者拡大，ソフトウェアの更なる改良が可能となる。

4. 前年度までに得られた研究成果の概要

本研究では，ステンシル計算（①構造格子，②半構造格子），疎行列演算（③一般行列，④悪条件問題，⑤Adaptive CG），⑥H 行列，⑦精度保証，⑧消費電力測定，⑨自動チューニング手法，の各項目についての研究開発を実施する。本研究は 2018 年度から，3 年計画として実施し，上記①～⑨の項目について，研究開発成果を学会等で発表する他，ISC18・19，SC18・19 等の国際会議の展示ブースでも紹介した。2018 年度の段階で低精度・混合／変動精度演算をアルゴリズム，最適化，精度保証，消費電力まで含めて扱った研究事例はほとんどなかったが，SC19（2019 年 11 月）では，低精度・混合／変動精度演算を数値計算に取り入れた研究事例が多数見られた。2019 年度までに得られた知見は以下の通りである：

- 従来，倍精度演算が適用されていた科学技術計算に単精度・半精度・混合演算を適用し，反復改良法（Iterative Refinement）併用により，同等の精度の計算結果がより短時間で得られる場合がある。一般に，計算時間短縮に比例して消費エネルギー（Joule）は減少。
- 悪条件問題では，低精度演算では正解が得られない場合がある。特に半精度演算は変数の範囲が限定されるため注意が必要であり，精度の要求されない反復法前処理等に適用するべきである。
- 従来の M 疎行列向け精度保証手法（T. Ogita 他, 2001）は，相対誤差上限の見積もりが厳しめであったため，より現実的な手法を開発し，悪条件問題への有効性が示された。
- 演算精度の影響は，問題規模・疎行列格納手法，アーキテクチャによって多様である。
- 局所的に演算精度を変更する手法の開発に着手し，有効性が示された。

① 実問題向け精度保証手法の開発（疎行列）

係数行列 A が M 行列性を持つ場合，高速な精度保証法〔Ogita et al., Computing, 2001〕を適用可能である。不均質場におけるポアソン方程式を解く ICCG 法〔6〕への適用事例では，条件数が大きくなると誤差限界が過大評価になっている可能性がある。本年度は〔Ogita et al., Computing, 2002〕に基づく手法を開発し， Reedbush-U（1 ノード）を使用して，昨年度と同様の条件で数値実験を実施した。図 1 に示すように，提案方式によって誤差限界の過大評価が大幅に抑制できている。また，近似解と精度保証の計算時間は同程度である。

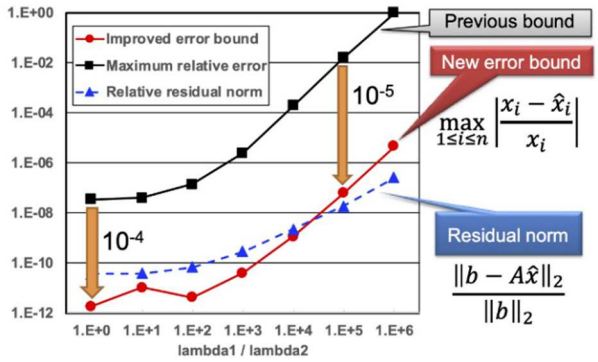

図 1 条件（λ_1/λ_2）毎の最大相対誤差（実線）と相対残差ノルム（破線）

② FP21 適用

比較的条件の良い問題では，倍精度演算（FP64）のかわりに単精度（FP32）を使うことによって計算時間を短縮できる場合があるが, 半精度（FP16）は有効桁数が 3 桁程度のため，実問題に使用することは困難である。本研究では FP16 と FP32 の中間的な特性を持つデータ型である FP21 を定義し，有限要素解析における前処理において活用した。FP21 は符号 1 ビット・指数部 8 ビット・仮数部 12 ビットの合計 21 ビットからなり，FP32 (符号 1 ビット・指数部 8 ビット・仮数部 23 ビットの合計 32 ビット)と指数部のビット数が同一となる。FP21 は FP32 と同じレンジを持つもののデータサイズが 1/1.5 となるため，メモリバンド幅ネックのカーネルにおいては実行時間が 1/1.5 になると期待される。なお，現在の主要な計算機においては FP21 の演算器はサポートされていないため，変数のメモリへの格納時のみに FP21 を使い, 計算時には FP21 を FP32 に変換して演算する。本研究では上記の FP21 と共に，人工知能を用いた前処理方法を開発

することで，V100 GPU ベースの計算機において従来ソルバー比で約 4 倍の高速化を実現した。

③ H 行列向け手法の高速圧縮手法の開発

H 行列計算の中で最も工夫を要するのは，密行列を低ランク行列に圧縮する部分である。圧縮には一般的に randomized SVD が用いられ，その内部カーネルは複数の QR 分解から構成される。昨年度は多くの小さな QR 分解を GPU 上で並列に行う batched QR 分解を開発したが，本年度はこれを TSQR（Tall & Skinny）に拡張し，任意の大きさの行列の randomized SVD を行うための QR 分解の枠組みを開発した。NVIDIA V100 上で FP32 を用いた場合の計算時間は cuSOLVER による QR 分解と比較して最大 2.17 倍高速となった。

④ 問題規模・行列格納手法と演算精度の関係

倍精度演算（FP64）と比較すると単精度演算（FP32）では計算時間が減少するが，アーキテクチャ，最適化の程度によってその影響は様々である。本年度研究では，ポアソン方程式を解く ICCG 法に対して，OFP（Intel Xeon Phi），OBCX（Intel CLX）1 ノードにおいて様々な疎行列格納形式，問題規模に対して計算を実施した。

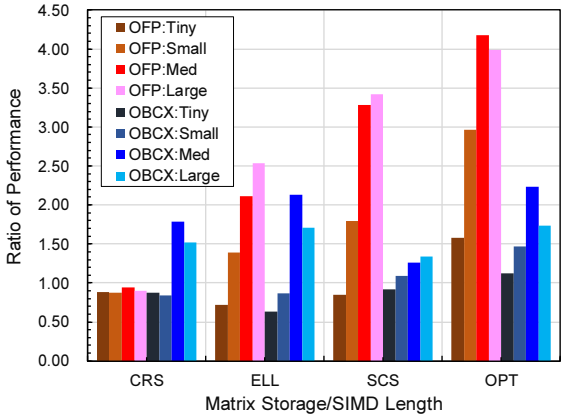

図 2　アーキテクチャ・問題規模・疎行列格納法と演算精度の効果

図 2 は，CRS・倍精度の計算効率を 1.00 とした場合の単精度における計算効率である（CRS：Compressed Row Storage，ELL：Ellpack-Itpack，SCS：SELL-C-σ，OPT：SCS を更に最適化したもの）。OFP ではベクトル化が不十分な場合には単精度による計算時間短縮効果は得られない。OFP，OBCX 共に，問題規模が大きく，データ移動量が増加するほど単精度による時間短縮の効果が顕著である。最適精度選択のための自動チューニング手法策定には，これらの項目を考慮する必要がある。

⑤ 局所的な変動精度演算適用

2018 年 7 月に出席した国際会議 WCCM では，問題の条件に応じて局所的に演算精度を変動させる可能性について指摘された。本年度は，大規模分散並列問題で局所的・動的に演算精度を変動，動的に負荷分散を適用する手法の開発を進めている。まず，領域分割に基づく並列有限要素法において，領域毎に異なる精度を適用した場合の検証を，並列有限要素法による三次元構造解析における加法シュワルツ法に基づく ICCG 法ソルバーに対して実施した。規則形状を 8 分割した場合，6 領域については倍精度，2 領域については混合精度（前処理・加法シュワルツ法部分のみ単精度）の疎行列ソルバーを適用し，均質問題において，全 8 領域倍精度の場合と同じ反復回数で同じ解を得た。

5.　今年度の研究成果の詳細

本研究では，4. で示した 9 項目について研究開発を実施した。ここでは紙面の都合上代表的な研究事例，すなわち，疎行列演算，H 行列演算，精度保証，自動チューニング，変動精度・任意精度の研究事例について紹介する。

①疎行列関連（1）：前処理付き反復法における混合精度演算〔4,5,6,8,9,10,13,15,20,21,22,23,26〕

不均質場における三次元熱伝導方程式を ICCG 法で解く場合において，表 1 に示すように，各演算部分を倍精度（FP64），単精度（FP32），半精度（FP16）を適用した場合，熱伝導率比λ_1/λ_2（行列の条件数に比例）との関係について，富士通 FX700（A64FX）1 ノード，48 コアを使用した検討を実施した。前処理に半精度演算を適用した D-H，S-

H はλ_1/λ_2=10^6の場合は収束解が得られていない他は, D-S, D-H は D-D と比較して計算誤差は無い。一方で S-S，S-H はλ_1/λ_2=10^3 の以下の場合は誤差も 1%以下であるがλ_1/λ_2が増加すると急速に誤差が増加する。図 3 は，各ケースにおける無次元化した反復回数であり，D-S，D-H では解が得られる範囲では，反復回数は D-D の場合と変化していないことがわかる。S-H は S-S とほぼ同様である。

表 1　混合精度演算における設定

	反復法主要部	前処理	前処理部分ベクトル
D-D	倍精度	倍精度	倍精度
D-S	倍精度	単精度	単精度
D-H	倍精度	半精度	単精度
S-S	単精度	単精度	単精度
S-H	単精度	半精度	単精度

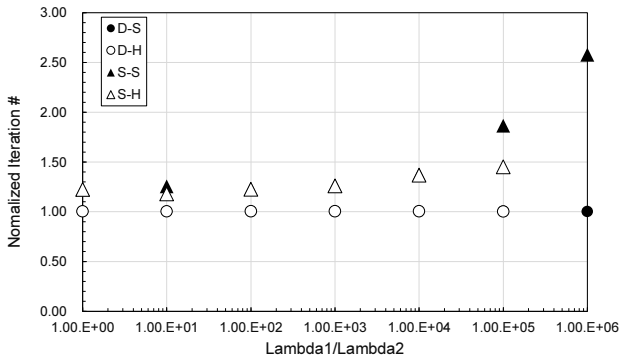

図 3　D-D の反復回数で無次元化した反復回数

②疎行列関連（2）: SIMD・メニーコア計算機用の精度混合疎行列ソルバー〔1〕

地震解析等で使われる低次の非構造格子有限要素法において，ソルバー本体は倍精度・前処理は単精度を使う精度混合演算により計算コストを削減，行列ベクトル積カーネル中のランダムアクセスの一部を連続アクセスに変換しコア間の競合を低コストで解決する方法を提案することで，SIMD・メニーコア CPU の能力を引き出す手法を開発した。Oakforest-PACS（OFP）において，従来手法比で 3.99 倍の高速化を実現した。

③疎行列関連（3）: Parareal 法への低精度演算・混合精度演算適用〔12,16,18〕

時間並列計算法の一つである Parereal 法において予測・修正計算に使用される SOR 前処理付き BiCGStab 法について，従来は倍精度のみ適用していたが，前処理に単精度演算を適用し，反復回数は増加するものの，名古屋大「不老」Type I サブシステムによる予備評価では，前処理 1 回当たりの実行時間を 47%削減できることが示された。更に適切な型変換タイミングの検討，半精度演算適用などの研究開発を実施した。

④TensorCore 上での低ランク近似〔2,3〕

TensorCore への数値データの入力は半精度（FP16）である必要があり，このため単精度行列積を行う場合は入力行列を半精度へキャストする必要があり計算精度が劣化する。本研究ではこの半精度へキャストする際に失われる仮数部を別の半精度変数で保持する。この補正計算では，失われる仮数部を計算する際にアンダーフローを起こす確率が高い。このため，同時に指数部の調整を行うことでアンダーフローを抑制し，この補正計算自体の精度の向上を行った。また, TensorCore の内部では足し込みが単精度を用いて行われるが，その際の丸めの方法が通常の単精度とは異なるなるため, これも精度劣化の原因となる。このため，中間報告では TensorCore を用いた行列積を 5-6 桁の有効数字の精度で行なうことができたが，単精度の誤差と比べると大きい結果となった。

最終報告では，この丸めの問題に対処するため TensorCore の足し込みを用いない方法を導入し，TensorCore を用いた行列積において単精度と全く同じ精度を得ることに成功した。図 4 に単精度の cuBLAS, TensorCore を用いた cuBLAS, TensorCore+精度補正の 3 通りの行列積における

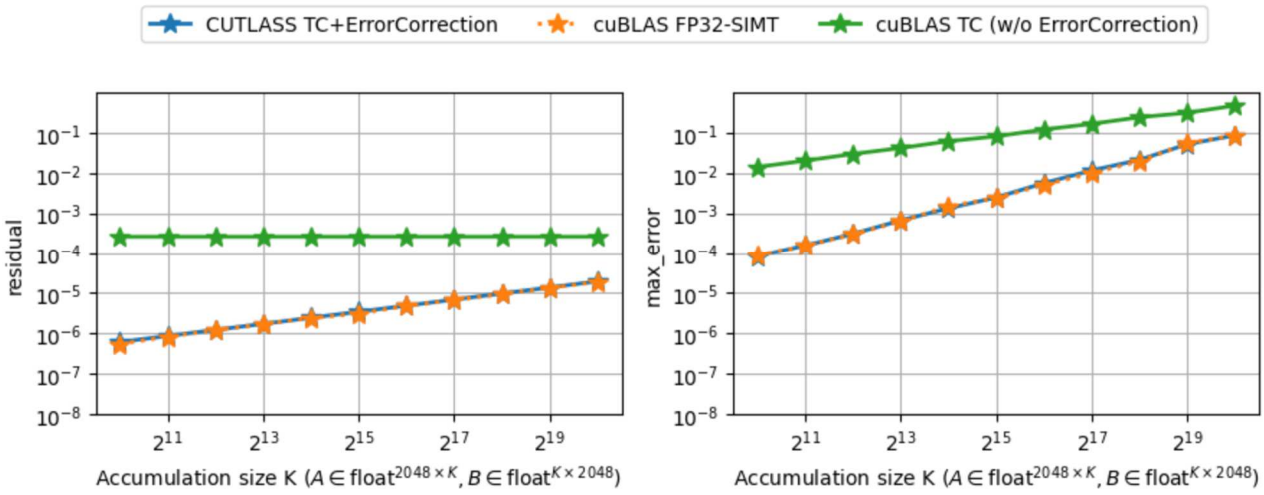

図 4 単精度の cuBLAS(青), TensorCore を用いた cuBLAS(緑), TensorCore+精度補正(橙)の比較

誤差を示す。これは，中間報告で示した(1)補正項の追加，(2)アンダーフロー抑制のための正規化，に加え，今回初めて行った(3)足し込みの抑制，の３つがそろって初めて実現できる。また，TensorCore+精度補正は，単精度の cuBLAS と比べて約半分の計算時間で実現できることが分かった。

⑤精度保証関連〔4,6,17〕

2019 年度に開発した疎行列ソルバー向け精度保証手法を単精度演算による不均質熱伝導場に対する ICCG 法に適用した。精度保証部分は，λ_1/λ_2=10^4 で破綻する。

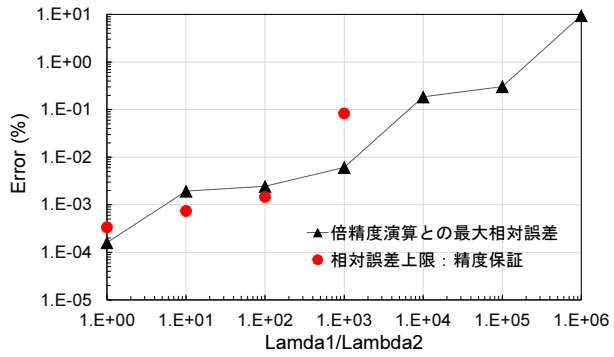

図 5 単精度演算時の各λ_1/λ_2における精度保証から得られた相対誤算上限と倍精度演算との最大相対誤差

図 5 に示すように，λ_1/λ_2=10^3 までは倍精度演算との最大相対誤差と精度保証から得られる相対誤差の上限はほぼ一致しており，本手法が単精度演算にも有効であることが示された。

更に本手法を分散メモリ環境向けに拡張し，Oakforest-PACS（OFP），8 ノード，128 ノードを使用し，不均質場における並列多重格子前処理付き CG 法によるポアソン方程式求解プロセスに適用を実施した（表 2，図 6）。

表 2 多重格子法による精度保証実施ケース〔6〕

	Problem Size	Environment for Computing
Small	2,097,152 (=128^3)	• 1 node of OFP • 8 MPI processes • HB 8×8
Large	536,870,912 (=1,024×1,024×512)	• 128 nodes of OFP • 1,024 MPI processes • HB 8×8

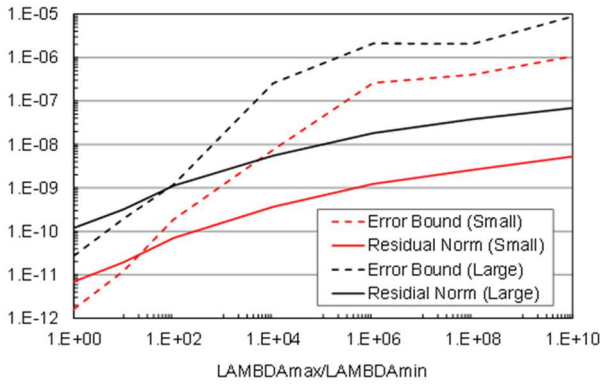

図 6 不均質場における並列多重格子前処理付き CG 法によるポアソン方程式求解プロセスについて精度保証から得られた相対誤算上限と倍精度演算との最大相対誤差（表 2 参照）

⑥ AT 関連：AT 言語 ppOpen-AT による混合精度演算の AT〔14,19〕

ppOpen-AT により生成される混合精度演算コード（「最適化候補」と呼ぶ）のうち，どの最適化候補を使用するかを性能パラメタにする AT 方式を提案した。対象箇所の実行時間の計測し，ユーザからの精度要求を考慮した上で，どの最適化候補を利用するかを決定する AT 機能を提案した。

本提案では，(1)演算対象の演算精度を変化させる；(2) ユーザによりあらかじめ提示された演算誤差以下となる；という条件で，最高速となる最適化候補の組み合わせを選ぶ AT 機能を実現することを目的にする。以下の指定を前提とする：

1)ソフトウェア開発者による演算精度要求指定：基準演算に対する相対的な演算精度劣化の許容値（許容相対誤差）を事前に与える。たとえば，倍精度演算結果に対し，相対誤差で 1e-10 以下にする等，の指定である。

2) 最適化対象：基準となる演算精度に対し，より低い精度の演算（もしくは，より高い精度の演算）による混合精度演算により，1)の要求を満たし，最高速となる組み合わせを選ぶ。

スーパーコンピュータ「不老」TypeIサブシステム(Fujitsu PRIMEHPC FX1000)を用いて，倍精度(DP)演算と単精度(SP)演算の混合精度演算の予備評価を行った。「不老」TypeIサブシステムは，ノード当たり，DP 演算は 3.3792 TFLOPS，SP 演算は 6.7584 TFLOPS である。全球雲解像モデル NICAM (Non-hydrostatic ICosahedral Atmospheric Model)(非静力学 正二十面体 大気モデル)の，雲の微小な物理演算を行うカーネル physicskernel_microphysics に含まれるサブルーチン mp_nsw6 を評価対象とした。

本年度前期からの知見に基づき最終報告では，ユーザが任意のプログラムの部分に混合精度演算の AT 適用場所をユーザが与える方式を提案し，予備評価を行った。結果を表 3 に示す。

表 3 スーパーコンピュータ「不老」TypeIサブシステムでの混合精度実行の速度向上率と演算精度

演算精度の組み合わせ [グループ番号]	速度向上率	演算精度 (QV)
すべて DP	1.00	1.172E-12
すべて SP	2.28	2.099E-08
28	**1.35**	**6.902E-12**
29	1.33	6.899E-12

表 3 の結果により，すべて SP にすれば 2.28 倍の速度綱領を得るが，すべて DP の時の精度（NICAM での QV 変数）に対して約 4 桁の大幅な精度劣化を招くことがわかる。そこで，グループ 28 の混合精度にすれば，ほぼ同一の桁数の演算精度で，かつ 6 倍ほどの精度劣化を許容できれば，1.35 倍の速度向上を享受することができる。

以上のことにより，実アプリケーションにおいて，本提案の AT の効果があることを明らかにした。

⑦ 変動精度・任意精度への取り組み（1）：CPU への FP21・FP42 実装〔25〕

一般的に FP64 が用いられる ICCG 法の IC 前処理に対して，汎用 CPU 上での FP16 や FP32 に加えて，IEEE754 に定義されていない FP21，FP42 の実用性について評価を行った。結果，問題の条件に依存するが，低精度でも十分な収束性を示し，計算時間短縮に寄与することが確認できた。

低精度演算の実アプリケーションへの適用は SIMD 演算の効率化やメモリ転送量の削減による計算時間短縮の効果が期待できるが，同時に演算精度の低下による影響も評価する必要がある。本研究では ICCG 法での低精度演算の実用性について，IC 前処理の行列やベクトルを低精度化し，検討を行った。IC 前処理の演算精度による ICCG 法全体の収束性への影響は比較的小さく，低精度化による効果が得やすい。ただし，悪条件問題を解く場合には前処理行列で係数を表現できなくなるなど，精度による影響が発生する。実際に構造格子での構造解析の問題を ICCG 法で解く場合の，IC 前処理の低精度化による効果を図 7 に示す。構造解析の問題ではポアソン比を変化させることで

問題の条件が大きく変化する。本評価ではポアソン比を 0.3～0.49 まで変化させ，IC 前処理の行列部のみに FP21,32,42,64 を適用した場合の反復回数および計算時間を計測した。評価では OBCX スーパーコンピュータの 1 ノードを使用した。係数行列の格納形式は CRS 形式である。なお，中間報告時に課題としていた FP21，FP42 の使用によるオーバーヘッドに関しては，OBCX 上では 0.4%程度と十分に小さいことを確認している。ポアソン比が 0.3～0.43 の間では演算精度の収束性に対する影響は小さく，低精度なデータ型ほど計算時間短縮の効果が大きくなった。0.44～0.49 の間では FP21 の収束性悪化が目立ち，FP32 の方か高速であった。本評価では FP42 の優位性が見いだせなかったが，非構造格子の問題など，より悪条件の問題では FP42 の効果が得られる条件があると考える。

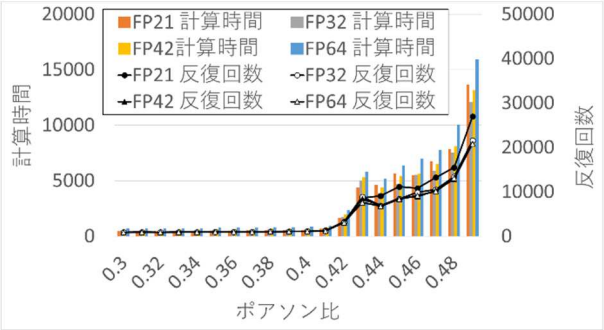

図 7　構造解析問題での低精度化の効果

次に，A64FX プロセッサを搭載した FX700 上で，FP16 を含めた評価結果を図 8 に示す。本評価で使用した問題は，構造格子上で差分法を用いて離散化したポアソン方程式を対象としている。解くべき問題の条件は一定とし，純粋に低精度化による計算時間短縮の効果のみを計測した。行列の格納形式は SIMD 化の効果を得やすい Sell-C-σ形式で係数行列を用いており，また行列だけでなくベクトルの精度も変化させた。図 8 に示すように，行列およびベクトルの低精度化により計算時間が短縮されており，行列，ベクトルともに FP64 で計算した場合と比較して，行列を FP16，ベクトルを FP32 で格納した場合には，IC 前処理部のみで 42.8%，ICCG 法全体で 16.9%の計算時間短縮を達成した。FP16 などのより低精度なデータ型の使用による計算時間短縮への影響や当問題での FP21，FP42 の効果は今後評価していく。

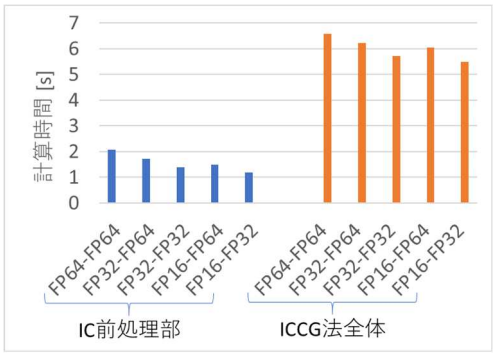

図 8　ポアソン方程式での低精度化の効果

⑧変動精度・任意精度への取り組み（2）：FPGA〔7,24〕

CPU, GPU では，あらかじめ演算器として用意された FP64, FP32, FP16 しか利用できず，任意精度の実現にはこれらを組み合わせエミュレーションする必要がある。FPGA を用いることで，任意精度の浮動小数点演算をハードウェアとして実現可能である。そこで，任意精度の演算器をハードウェア記述言語(HDL)または高位合成(HLS)によって実現する。一方，アプリケーションから FPGA に演算オフロードを実現するために，現状では OpenCL によるフレームワークを利用する必要があり，これらの演算器をライブラリとして実現し OpenCL から wrapper により呼び出す必要がある。

本年度は，IEEE754 浮動小数点表現を拡張した任意精度の演算器について，Verilog HDL および SystemVerilog による実装と，C++による実装を試み，Intel Stratix10 GX FPGA 上で性能を比較した。その結果，C++による実装においても低コストで HDL に匹敵する演算性能が得られた。

さらに，行列積およびステンシル計算のベンチマークにこれらの任意精度演算器を適用し，期待通りの計算結果が得られていることを確認すると同時に，浮動小数点ビット数を削減した場合には演算器の搭載数を増加させることができ，全体として高い性能が得られることを示した。本研究に

ついては，2021 年 7 月に開催される xSIG 2021 での発表，および Best Master's Student Award の受賞が決定している〔7〕。

6. 今年度の進捗状況と今後の展望

本研究は，計算科学・計算機科学・数値アルゴリズム分野の研究者の協力のもと，様々な数値アルゴリズムの最新アーキテクチャに向けた最適化，低精度・変動精度導入による高速化，消費電力節約，および精度保証に基づく最適精度選択のための自動チューニング選択手法の確立と実アプリケーションでの検証を目指したものである。本研究は 2018 年度に 3 年計画で開始し，4. で示した 9 項目（ステンシル計算（2 項目），疎行列演算（3 項目），H 行列，精度保証，消費電力測定，自動チューニング手法）について研究開発を実施し，2020 年度は最終年度にあたっていた。

当初計画は，2018 年度：各アルゴリズム・アプリケーション最適化，多様な演算精度適用，消費電力測定，2019 年度：精度保証手法確立，2020 年度：自動チューニング手法確立，であった。

2020 年度は各アルゴリズムの，演算精度，最適化，アーキテクチャの他，問題規模も考慮して消費電力・エネルギーへの体系的な影響評価を継続して実施する他，低精度・混合／変動精度演算に関する研究開発を，これまでの研究成果を元に継続して実施し，自動チューニング手法確立を目指す。精度保証手法については，疎行列演算に加えて，H 行列向け手法の研究開発も実施する。各センターの保有する NVIDIA V100，東大情報基盤センターに 2019 年度末迄に導入される富士通 FX700（A64FX（富岳））クラスタ等を使用して，新アーキテクチャ向け検討も実施する。特に富士通 FX700 クラスタを使用した，半精度演算（FP16）のフィージビリティースタディを様々な手法，アプリケーションについて重点的に実施する予定であった。

全体的に研究として順調に進捗しており，特に疎行列演算については，FX700 クラスタを使用した半精度演算（FP16）のフィージビリティースタディ，FP21・FP42 の実装，FPGA による実装については大きな進展があり，国内外で高い評価を得ている。また，2019 年度に提案した疎行列演算向け精度保証手法を分散環境に拡張し，検証を実施し，成果を国際会議で発表する予定である。

一方，H 行列演算向けの精度保証手法，最適精度選択のための自動チューニング手法の確立，については予備的検討に留まり，あまり研究を進展させることができなかった。本研究は 2021 年度から新たに 3 年計画で第 2 フェーズに入るが，これらの未達成項目については継続して取り組む予定である。

当初目標に対する達成度としては，上記①～⑥，⑧の各項目において，ほぼ 100%，ただ⑦精度保証については 70%，⑨自動チューニングについては 50%である。大きな目標としていた自動チューニング手法の確立が達成できなかったこと，ライブラリ等の公開があまり進まなかったこともあり，全体としての達成度は 3 年間で 80%である。

最適演算精度選択のための自動チューニングには機械学習の知見を併用する。そのために，行列の性質（固有値分布，条件数等）を短時間で精度良く算出する手法の研究開発を実施するが，このような手法については，2019 年度に予備的検討を実施済である。

7. 研究業績一覧

(1) 学術論文（査読有り）

[1] K. Fujita, M. Horikoshi, T. Ichimura, L. Meadows, K. Nakajima, M. Hori, L. Maddegedara, Development of element-by-element kernel algorithms in unstructured finite-element solvers for many-core wide-SIMD CPUs: Application to earthquake simulation, Journal of Computational Science, Volume 45, 2020, 101174, ISSN 1877-7503, https://doi.org/10.1016/j.jocs.2020.101174

(2) 国際会議プロシーディングス（査読有り）

[2] S. Deshmukh, R. Yokota, Distributed Memory Task-Based Block Low Rank Direct Solver, ISC

High Performance 2020 (Research Poster), June 2020

[3] H. Ootomo, R. Yokota, Randomized SVD on TensorCores, ISC High Performance 2020, (Research Poster), June 2020

[4] K. Nakajima, T. Iwashita, H. Yashiro, T. Shimokawabe, H. Matsuba, H. Nagao, T. Ogita, T Katagiri, h3-Open-BDEC: Innovative Software Platform for Scientific Computing in the Exascale Era, ISC High Performance 2020, (Project Poster), June 2020

[5] Nakajima, K., Gerofi, B., Ishikawa, Y., Horikoshi, M., Efficient Parallel Multigrid Solver on Intel Xeon Phi Cluster, IXPUG (Intel Extreme Performance Users Group) HPC Asia 2021, 2021

[6] Nakajima, K., Ogita, T., Kawai, M., Efficient Parallel Multigrid Methods on Manycore Clusters with Double/Single Precision Computing, IEEE Proceedings of the 16th International Workshop on Automatic Performance Tuning (iWAPT 2021) in conjunction with 35th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2021), 2021（in press）

[7] T. Hara, T. Hanawa, Transprecision Calculation Platform Offloaded on FPGA, Proceedings of xSIG 2021, 2021 (in press) (**Best Master's Student Award**)

(3) 国際会議発表（査読無し）

[8] Nakajima, K., Innovative Methods for Scientific Computing in the Exascale Era by Integrations of (Simulation+Data+Learning), MS14: Physics-based Simulation of Earthquake Hazards with HPC & HQC , COMPSAFE 2020（The 3rd International Confernce on Computational Engineering & Science for Safety and Environmental Problems）

[9] Nakajima, K., h3-Open-BDEC: Innovative Software Platform for Scientific Computing in the Exascale Era, Session IV: Toward Effective & Efficient Next-Generation HPC Software Ecosystems, The SOS 24 Virtual Conference (March 17, 2021, On-Line)（**招待講演**）

[10] Nakajima, K., Matsuba, K., Hanawa, T., Furumura, T., Tsuruoka, H., Nagao, H., Integration of 3D Earthquake Simulation & Real-Time Data Assimilation on h3-Open-BDEC, 2021 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT in HPSC) (Online, March 2021)

[11] T. Fukaya, Exploiting Lower Precision Computing in the GMRES(m) Method, 2021 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT in HPSC) (Online, March 2021)

[12] S. Ohshima, Effectiveness of Low-/Mixed-Precision Computation on Parareal Method, 2021 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT in HPSC) (Online, March 2021)

(4) 国内会議発表（査読無し）

[13] 中島研吾，坂本龍一，星野哲也，有間英志，塙敏博，近藤正章，低精度演算とアプリケーション性能，情報処理学会研究報告（2020-HPC-174-5），2020

[14] 片桐孝洋，演算精度と実行時間を考慮する自動チューニング方式とppOpen-ATへの実装について，第25回計算工学講演会（オンライン，2020年6月）

[15] 中島研吾，岩下武史，八代尚，下川辺隆史，松葉浩也，長尾大道，荻田武史，片桐孝洋，（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，第25回計算工学講演会（オンライン，2020年6月）

[16] 大島聡史，飯塚幹夫，小野謙二，Parareal法における低精度計算・混合精度計算の活用について，第25回計算工学講演会（オンライン，2020年6月）

[17] 中島研吾，荻田武史，塙敏博，河合直聡，伊田明弘，星野哲也，低精度・混合精度演算による高性能・高信頼性疎行列ソルバー，情報処理学会研究報告（2020-HPC-175-2），2020

[18] 大島聡史，飯塚幹夫，小野謙二，Parareal法における低精度演算・混合精度演算の活用，日本応用数理学会2020年度年会（オンライン，2020年9月）

[19] 片桐孝洋，山梨祥平，八代尚，大島聡史，永井亨，自動チューニング言語ppOpen-ATによる混合精度演算の最適化機能について，日本応用数理学会2020年度年会（オンライン，2020年9月）

[20] 中島研吾，メニィクラスタ向け並列多重格子法，日本応用数理学会2020年度年会（オンライン，2020年9月）

[21] 中島研吾，メニィクラスタ向け並列多重格子法,，情報処理学会研究報告（2020-HPC-176-6），2020

[22] 中島研吾，Balazs Gerofi，石川裕，堀越将司，メニィコアクラスタ向け並列多重格子法の最適化，日本応用数理学会「行列・固有値問題の解法とその応用」研究部会 第 30 回研究会（オンライン，2020 年 12 月 7 日）

[23] 中島研吾，塙敏博，下川辺隆史，坂本龍一，有間英志，星野哲也，伊田明弘，三木洋平，河合直聡，芝隼人，Society 5.0 を実現する BDEC システム，大学 ICT 推進協議会 2020 年度年大会（AXIES 2020）（オンライン，2020 年 12 月 11 日）

[24] 原 忠辰，塙 敏博， FPGAによる変動精度演算に向けた実装方法の検討，情報処理学会研究報告（2020-HPC-177-11），2020

[25] 河合直聡，中島研吾，FP21及びFP42を使用した不完全コレスキー分解前処理，情報処理学会研究報告（2020-HPC-177-21），2020

[26] 中島研吾，岩下武史，八代尚，下川辺隆史，長尾大道，荻田武史，片桐孝洋，松葉浩也，（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，第12回 自動チューニング技術の現状と応用に関するシンポジウム（ATTA2020）（オンライン，2019年12月25日）

[27] 田中一成，中尾 充宏: 不連続拡散係数を持つ３次元ポアソン方程式の解に対する事前誤差評価，日本応用数理学会2021年研究部会連合発表会（オンライン，2021年3月）

[28] 深谷 猛，岩下 武史，GMRES(m)法における行列データの低精度化に関する検討，日本応用数理学会2020年度研究部会連合発表会（オンライン，2021年3月）

(5) その他（特許，プレス発表，著書等）

なし