

高性能・高信頼な数値計算手法と応用の新展開(2)

JHPCN

● 研究の意義

- I) 疎行列ソルバ、H行列演算等の代表的な数値計算アルゴリズム、各アプリケーション（地震学、大気海洋科学、構造力学、流体力学、電磁気学等）について、メモリアクセス最適化及び分散メモリ通信最適化に着目し、安定で高性能な手法を各システムに実装し、消費電力の効果を検証しつつ、変動精度演算を実用的レベルまで引き上げることを目指す。
- II) 疎行列ソルバ、H行列、大規模言語モデル(LLM)などの演算に加えて、基本線形計算カーネル群（BLAS、テンソル演算、疎行列・ベクトル積）を対象として、実用的な精度保証と混合精度演算のアプリケーション拡大を目指す。I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演算精度を自動チューニング（Auto-tuning, AT）技術により、動的に制御する手法の確立を目指す。

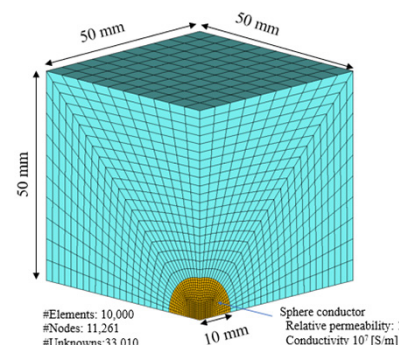
● 主な構成員と研究計画

- ① 疎行列ソルバ（主担当：岩下・中島・藤田、担当：市村・深谷・星野・河合・八代・荒川・Wellein・Basermann）
- ② H行列・テンソルトレインとLLM（主担当：横田、担当：岩下・伊田・星野・椋木・大島・河合）
- ③ 精度保証（主担当：荻田、担当：片桐・尾崎・中島・椋木・寺尾・内野・Marques）
- ④ 自動チューニングと計算機システム（主担当：片桐、担当：大島・坂本・塙・近藤）

● 研究内容紹介

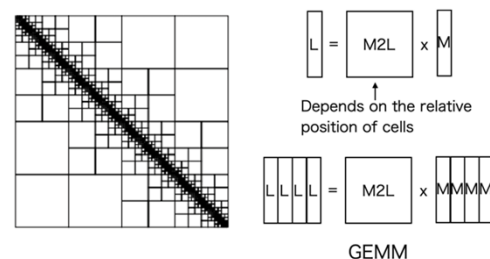
● ①疎行列ソルバ

- (1)Residual Replacement等による低精度・混合精度型前処理付き反復法の安定化：様々な材料定数分布とメッシュ幅分布を持つ対象モデルから得られる連立一次方程式を低精度・混合精度型前処理付き反復法で解く場合のResidual Replacement等の手法の調査・研究をする。
- (2)係数行列固有値分布の推定：最大200K自由度程度の比較的小規模な問題について、疎行列・密行列を対象としたライブラリ等を使用し固有値分布を算出し、実問題における大規模疎行列の固有値分布を推定する手法を検討する。
- (3)IC分解に関する固有値による分析を実応用問題へ展開：これまでに提案・発表したILUB前処理を電磁場有限要素解析で評価する。（図1(a)）
- (4)反復改良法に基づく混合精度型の線形ソルバに関する研究：これまでに調査されていない内部ソルバを用いた場合の調査・研究をすることで、汎用性を高める。



(a)性能評価に用いる電磁解析のモデル

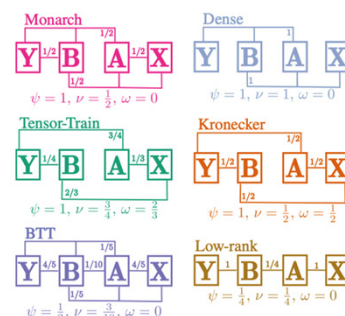
図1：①疎行列ソルバの研究内容



(a) FMMのTensor Core実装の高速化

● ②H行列・テンソルトレインとLLM

- (1) FMMのTensor Core実装の高速化：本研究では、FMMにおけるM2Lの計算をTensor Coreを用いて実装し、大幅な高速化を図る。また、M2Lを行列積の形で記述したFMMはH2行列のメモリをO(1)に低減する方法と見做すことができる。（図2(a)）
- (2) LLMの行列近似を用いた高速化：近年、密行列積の計算コストを低ランク近似、Tensor Train(TT)、Block Tensor Train (BTT)、Monarch行列などを用いて低減する手法が提案されているが、その近似が実際の学習に及ぼす影響を十分に調べた例は少ない。本研究では、これらの密行列積の高速近似手法をTensor Core上で実装し、混合精度演算・行列近似がもたらす学習精度への影響を調べる。（図2(b)）

(b) LLMの行列近似を用いた高速化
図2：②H行列・テンソルトレインとLLMの研究内容