

科学研究費助成事業 研究成果報告書

平成 27 年 6 月 12 日現在

機関番号：12601

研究種目：基盤研究(A)

研究期間：2011～2014

課題番号：23240005

研究課題名(和文)汎用自動チューニング機構を実現するためのソフトウェア基盤の研究

研究課題名(英文)Research on Software Foundations for General Purpose Automatic Tuning Mechanisms

研究代表者

須田 礼仁(SUDA, Reiji)

東京大学・情報理工学系研究科・教授

研究者番号：40251392

交付決定額(研究期間全体)：(直接経費) 31,500,000円

研究成果の概要(和文)：自動チューニング(AT)を汎用的技術として確立することをめざし、以下の4領域にて研究を行った。数理領域では、AT数理コアライブラリATMathCoreLibを開発・公開、また性能相関を扱う数理手法を開発した。プログラミング領域では、上記ライブラリをAT記述言語ppOpen-ATに組み込み、またコード配置の偶発性による性能変化を打ち消す手法を提案。システム領域では汎用的Auto-Tunerの構成法提示、また性能情報取得を軽量化した。アプリケーション領域では行列計算・大規模探索のAT、上記ライブラリのアプリに適用した。このように4領域を連携しつつ集中的に研究、ATの汎用技術化に向けて成果を挙げた。

研究成果の概要(英文)：Aiming to establish Automatic Tuning (AT) as a general technology, we researched in the following four domains. In Math Domain, we developed and released AT mathematical core library ATMathCoreLib, and developed mathematical methods for AT with performance correlation. In Programming Domain, we combined ATMathCoreLib with AT description language ppOpen-AT, and proposed a method to cancel performance variation from code placements. In System Domain, we proposed architecture of general purpose Auto-Tuner, and created light-weight performance data collection mechanism. In Application Domain, we apply AT to matrix computations and large scale search, and apply ATMathCoreLib to some applications. Thus we researched on four domains in focus and in collaboration, and achieved research results toward general purpose AT methodology.

研究分野：ハイパフォーマンスコンピューティング

キーワード：自動チューニング 性能モデリング 逐次実験計画 自動チューニング記述言語 コンパイラ技術 軽量化 行列計算 大規模探索

1. 研究開始当初の背景

自動チューニングは、ソフトウェアに性能に影響する要素を可変性として埋め込み、その可変性を調整しつつ実際の計算環境やデータのもとで計算を行い、性能を測定することにより、良好な性能を達成する調整パラメータを選択するというパラダイムである。従来手法では性能をモデル化して適切なパラメータを設定するということが多く行われていたが、ハードウェアの進歩の速さ、並列計算機における性能要因の複雑さ、実行時にしか与えられないデータに依存する性能の変動などから、最大限に性能を引き出すためには自動チューニングが必須であると考えられている。

本研究を開始した当初では、自動チューニング技術への理解と期待が広がりつつある状況であった。先行していた事例として、密行列計算ライブラリ ATLAS、高速フーリエ変換ライブラリ FFTW、疎行列計算ライブラリ XAbclib などがあった。

研究代表者および本研究課題に参加するメンバーは、特定領域や科研費、JST CREST において自動チューニング研究を推進してきた日本を代表する研究者であり、国際的にも国際ワークショップ iWAPT を主催したり、Springer から英語書籍を出すなど、世界をリードしてきた。

しかし、自動チューニングの研究は事例ベースで行われていて、汎用技術としての研究は遅れていた。

2. 研究の目的

前述のような背景のもとで、本課題では、自動チューニングを汎用的なソフトウェア技術として確立することをめざし、そのために必要な研究領域を「数理」「プログラミング」「システム」「アプリケーション」の4つと定め、これら4領域を集中的にかつ連携しつつ進めることにより、自動チューニング研究の飛躍的推進と汎用技術への高度化を目指した。

「プログラミング」領域は、自動チューニングのためのプログラミング記述手法の研究と、記述されたプログラムを実行コードへとコンパイルする変換手法の研究を含む。高性能な変種を生成するコンパイラ技法と、自動チューニングのためのプログラミング環境という2つの面から、汎用自動チューニング技術の開発にアプローチした。

「数理」領域は、自動チューニングのための性能分析、性能モデリング、実験計画、パラメータ最適化を含む。研究代表者が本課題以前から取り組んでいるテーマであるが、複数のパラメータが相互に影響する場合や、並列自動チューニングなどにおける自動チューニングの数理手法の開発を目指した。

「システム」領域には、自動チューニング

機構を支える種々の仕組みが含まれる。重要なテーマの一つは性能測定と、その記憶、ならびに必要な性能情報を抽出する仕組みである。また、プログラミング、数理、システム、アプリケーションを接続し、自動チューニングのサイクルをまわす機構が必要である。この分野は特に遅れており、このような課題の突破口を開くことを目指した。

「アプリケーション」領域は、それぞれの重要なアプリケーションにおける自動チューニングの適用である。ひとつには、各分野のアプリケーションにおける自動チューニング機構への要請がどのようなものかを分析すること、もうひとつには、他の領域で開発された手法をアプリケーションに適用することである。

これら4領域の研究を設定すること自身に新規性があるが、それぞれの分野を深め、しかも相互に連携して進めることにより、全体として自動チューニング研究を深く広く発展させることを目指した。

3. 研究の方法

本課題に取り組むため、これまで自動チューニング研究において我が国を代表する研究者が集まり、チームを構成して研究体制を構築した。チーム内は4つの領域にわかれ、それぞれの研究を進めるとともに、相互連携の取り組みも進めた。

全体の統括と、数理領域は、研究代表者である須田礼仁(東京大学)が担当した。これに研究協力者である電気通信大学名誉教授弓場敏嗣の助言を得て、本課題全体を推進した。

プログラミング領域では、佐藤周行(東京大学)を中心として、東京大学の片桐孝洋、京都大学から九州工科大学に異動した八杉昌宏が連携して研究を推進した。

システム領域では、今村俊幸(電気通信大学 理化学研究所)が中心として、鴨志田良和(東京大学)と大澤範高(千葉大学)が参加して研究を進めた。ただし鴨志田は途中で事情により離脱した。

アプリケーション領域では、山本有作(神戸大学 電気通信大学)を中心に、美添一樹(東京工業大学)と藤井昭宏(工学院大学)、玉田嘉紀(東京大学)が参加して研究を推進した。

具体的な研究内容は成果のほうで説明するが、チームとしての研究推進等は以下のように進めてきた。1年に1回、本課題全体のミーティングを開催し、各年度の目標を設定し、2年度目からは実績を報告、意見を交換しあった。またこれとは別にマイクロワークショップとして、自動チューニングに関連する科研費と合同で、合宿形式の研究会を行い、最新の研究成果の情報交換と、意見の交換を行った。これらは研究の方向性を確認し、連携して推進することに寄与した。このほか、

4つの領域内や、領域間のミーティングは適宜実施した。また、多くのメンバーが参加しているクローズドな研究会「自動チューニング研究会」の場も活用して、密接な研究連携を進めた。

組織的な対外発信も行った。国内の関連学会である「日本応用数理学会年会」、「計算工学講演会」、また海外では隔年で行われる「SIAM PP」および「SIAM CSE」において、自動チューニングを研究している他の科研費やプロジェクトと共同してセッションをオーガナイズし、定期的に情報発信を行った。また我々自身が主催するものとして、2006年より開催している国際ワークショップ「iWAPT (international Workshop on Automatic Performance Tuning)」を2011年からは本課題にて継続して実施し、国際的な自動チューニング研究の発展に寄与した。2011年にはシンガポール、2012年には神戸、2013年にはバルセロナ、2014年にはユーロ圏で実施した。また自動チューニング関連の科研費やプロジェクトと合同して、「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 (ATTA)」を毎年12月に東京都内で開催し、企業等を含む国内の関係者に広く情報発信をした。さらに国際ワークショップとして「ATAT-HPSC (Conference on Advanced Topics and Auto Tuning in High Performance Scientific Computing)」を台湾大学の協力を得て2013年から毎年台湾を会場に開催し、台湾の自動チューニングへの取り組みを支援した。また、ホームページを立ち上げて、本課題とその取り組み、および成果を広く情報発信する取り組みを行った。

4. 研究成果

数理領域では、自動チューニングのための数理手法とそのライブラリ化を実施した。我々はベイズ統計と逐次実験計画に基づく自動チューニングの数理手法を提案している。本課題では、従来手法に加えて次の拡張を行った。

まず、「選択枝変更コストを含むオンライン自動チューニング」では、チューニングパラメタを切り替え際に発生するオーバーヘッドを考慮したうえで性能を最適化する手法を提案した。ここで考察した事例は疎行列データ構造であり、切り替えオーバーヘッドは疎行列データ構造の新たな作成に伴うオーバーヘッドである。このオーバーヘッドを明示的に扱うことにより、未測定の選択枝を試す回数が減り、全体としてのコストが減少した。

このあと、従来手法を大きく発展させる、関連のある自動チューニングの性能モデルの研究を進展させた。関連とは、異なるパラメタ値であっても似たような性能の傾向を示すことであるが、定量的にうまく扱うことが難しかった。しかし、我々の手法は統計学

に基づいており、性能モデルに誤差が含まれていてもそれを適切かつ定量的に扱い、自動チューニングの手法を実現できる特徴がある。最初の例題として、行列積における行列サイズパラメタを考察した。行列積はアンローリングにより多数の等価コードが生成できる利点がある。行列サイズは、大きさによりキャッシュの影響が異なり性能が変わることと、2の冪や2の冪を因子として含むサイズではキャッシュライン衝突のために性能が低下することが知られている。我々は、空間統計に用いられるkrigingを中心とする手法を用い、variogramを推定した。これを約20種類の計算機における行列積性能データに当てはめたところ、ちょうど空間統計で想定されるようなvariogramが得られた。このことから、krigingモデルにより関連のある性能情報を活用できることを示した。

関連のある性能の第2の応用として、計算機の消費電力のチューニングを行った。CMOSトランジスタのリーク電流は温度によって変わり、同じ計算をしても実行時の温度によって消費電力が異なる。このため、温度の低いときに測定した選択枝は見かけ上電力効率が高く見えてしまう。これの対処方法は2つであり、1つめは各温度で独立にチューニングして最適な選択枝を選ぶこと、2つめは電力の温度依存成分を差し引くことである。しかし、いずれの手法でも、温度が自由に制御できないために、低温や高温の電力が容易に測定できず、モデル化ができるだけの十分な情報は一般に得られない。しかし我々が提案しているオンライン自動チューニングでは、手持ちのデータの範囲で最適と考えられる選択枝を自動的に選ぶことができるため、実際に測定できない温度についてはモデルを構築せずに最適化が可能である。これにより、エネルギー最小の選択枝を適切に選ぶことができることを示した。

また、これら自動チューニングのための数理手法のうち汎用性のある手法を自動チューニング数理コアライブラリATMathCoreLibとして構築、公開し、仕組みと使用方法の説明を行った。最初は汎用性の高いライブラリとして公開したが、背景の数理がわからないと利用するのが難しかったため、より使い方を制約した容易なインタフェースを提供した。

プログラミング領域では、自動チューニング記述言語ABCLibScript / ppOpen-ATをアプリケーション領域で開発されているAMGソルバに適用し、既存のアプリケーションに容易に自動チューニング機構が取り込めることを実証した。また、ABCLibScriptに数理領域で開発されているATMathCoreLibを組み込み、電力最適化に適用して、適切な自動チューニングが領域を連携して実施できることを示した。また、自動チューニングにおいて問題となりつつある微小な変換にとまらぬ性能変化について詳細に分析した。こ

ード配置のみが異なる複数のプログラムを生成し、その性能測定結果を統計処理することで、コード配置の偶発性を打ち消す手法を提案した。加えて、データ再配置により局所性を改善する手法を実用的な処理系にて評価した。また、コンパイラ技術の発展として、インライン展開の解析、条件文の解析の厳密化を行った。また、CPU/GPU における性能モデルの詳細な解析を行った。

システム領域では、汎用的な Auto-Tuner を構成するための汎用的なフレームワークを構成するためのソフトウェア構成法を研究した。この Auto-Tuner は汎用ツールとの API を規定し、適切なデータ交換を行うことで高度な自動チューニングを実現する。またこの成果を GPGPU の計算プログラミングである CUDA により密行列計算を行う CUDA-BLAS に適用し、実証した。

また、並列ソフトウェアにおける性能情報の抽出の軽量化に取り組んだ。情報取得自体の軽量化と、変化の少ない部分の収集頻度を下げることにより、1 回あたり平均 1 ミリ秒以下で情報取得することが可能となった。既存手法の 10 倍から 100 倍ほどの高速化に相当する。またその応用として、システムノイズによるアプリケーションの性能低下について、原因となるプロセスを推定し、その影響を定量化する研究に取り組んだ。また、フロントエンドノードと計算ノードのアーキテクチャが同一であることを前提としているインストール時自動チューニング機能付きソフトウェアを全く変更せずに、フロントエンドノードと計算ノードのアーキテクチャが異なるシステムに移植する仕組みを構築した。また、並列オンライン自動チューニングで必要となるパラメタサーベイ型のジョブ実行を効率的に行うために、疑似的な動的プロセス生成ミドルウェアの設計とプロトタイプ実装、性能評価を行った。

アプリケーション領域では、第一原理分子動力学法における固有値計算について、数理論域で開発した自動チューニングのモデルである 4DAC で分析した。また、特異値分解アルゴリズムに、数理論域で開発した ATMathCoreLib を適用し、適切なブロックサイズが決定できることを示した。また、特異値分解における CPU/GPU の負荷分散最適化について研究し、最大 4 倍の性能向上を達成した。複数アルゴリズムが異なる性能パラメタを持つ場合の性能モデルの構築法を提案し、多項式を用いた性能モデルにおける最適次数の決定法について数値実験に基づく調査を行った。また、ハウスホルダー法に基づく QR 分解について、従来アルゴリズムと TSQR をモデル化し、列数が小さいときは TSQR 法、大きいときには従来法が優位となることを示した。またモデルと実測がよく一致することを示した。また、流体解析における代数的マルチグリッド (AMG) ライブラリの自動チューニングを研究した。デフォルトのパラメ

タ設定からパラメタの種類ごとに順次最適化した。

また、今回新たに自動チューニングの対象として、並列大規模グラフ探索を取り上げた。これは行列計算などに比べアルゴリズムが複雑で、調整パラメタが多い傾向にある。従来はコア数 1,200 程度まで高速化が得られていたが、これを 4,800 コアまで拡張した。また、囲碁をターゲットとして性能に大きく影響するパラメタを分析し、5 つ程度を特定した。また、バイオインフォマティクスにもターゲットを広げた。遺伝子ネットワーク推定アルゴリズムの並列計算の研究を進め、多様な環境で高い性能を得る実装を開発した。

これらの成果により、自動チューニングのプログラミング、システム、数理、アプリケーションの各領域の研究を深めることができたとともに、領域をまたがる研究成果を挙げることにより、我々の提案する汎用自動チューニングの構成方式が有効であることを示唆することができた。

また、これらの研究成果は、情報処理学会 HPC 研究会等で個別に発表したほか、計算工学講演会 (2011~2014 年)、日本応用数理学会年会 (2011~2014 年)、SIAM PP (2012, 2014 年)、SIAM CSE (2013, 2015 年) および EASIAM (2012 年) では関係者が独自のオーガナイズドセッションやミニシンポジウムを持ち、自動チューニングの研究成果をプレゼンスをもって報告した。さらに、関連課題と共同で国際ワークショップ iWAPT (international Workshop on Automatic Performance Tuning) を主催し、2011 年はシンガポール、2012 年は神戸、2013 年はバルセロナ、2014 年は米国ユージーンで開催した。また国内向けには「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」を関連研究課題と共同で毎年 12 月に東京大学で開催し、企業の研究者を含む多くの参加者に情報を発信した。また、台湾の研究者と連携を深め、2013 年から毎年 2~3 月に国立台湾大学において国際会議 ATAT-HPSC (Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing) を開催し、欧米からも研究者を招聘して情報発信および国際的な研究振興に寄与した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 19 件)

Chapa Martell Mario Alberto, SATO, Hiroyuki, **Linear Performance-Breakdown Model: A Framework for GPU kernel programs performance analysis**, Int'l J. Networking and Computing, Vol 5, No 1, 86--104, 2015. 査読有

<http://www.ijnc.org/index.php/ijnc/>

[article/view/98](#)

須田礼仁, **自動チューニング: 数理解的手法によるソフトウェア高性能化**, 次世代計算科学の基盤技術とその展開, 京都大学数理解析研究所講義録 1848 pp. 1-14, Aug. 2013. 査読なし
<http://www.kurims.kyoto-u.ac.jp/~kyodo/kokyuroku/contents/pdf/1848-01.pdf>

藤井昭宏, **SMAC 法による流体解析を対象とした AMG ライブラリの自動チューニング方式**, 電子情報通信学会論文誌 D, Vol. J96-D No.5, pp. 1321-1329. 2013 年 5 月 査読有

http://ci.nii.ac.jp/els/110009603596.pdf?id=ART0010063949&type=pdf&lang=jp&host=cinii&order_no=&ppv_type=0&lang_sw=&no=1433824447&cp=

深谷 猛, 山本有作, 張紹良, **動的計画法を用いたブロックハウスホルダ QR 分解アルゴリズムの性能最適化**, 情報処理学会論文誌コンピューティングシステム (ACS), Vol. 4, No. 4, pp. 146-157, 2011. 査読有

[学会発表](計 6 7 件)

長島聖児, 深谷猛, 山本有作, 横川三津夫, **オンライン自動チューニングのための性能モデルの構築法 ~ 正方行列の特異値分解を例にして ~**, 日本応用数学会 2015 年研究部会連合発表会, 明治大学中野キャンパス, 2015 年 3 月 6 日.

Takahiro Katagiri, Cheng Luo, Reiji Suda, Shoichi Hirasawa, Satoshi Ohshima, **Energy Optimization for Scientific Programs Using Auto-tuning Language ppOpen-AT**, Proceedings of IEEE MCSoc2013, pp.123-128, 東京・NII (Best Paper Award)

鴨志田良和, **Simplifying Install-time Auto-Tuning for Cross-Compilation Environments by Program Execution Forwarding**, 18th IEEE International Conference on Parallel and Distributed Systems (Poster) 2012 年 12 月 17 日 シンガポール

須田礼仁, **自動チューニングのための相関モデル: 行列積における行列サイズパラメタ**, 情報処理学会研究報告, Vol. 2012-HPC-135, No. 21, 第 135 回 HPC 研究会, 鳥取, Aug. 2nd, 2012.

Toshiyuki Imamura, **ASPEN-K2: Automatic-tuning and Stabilization for the Performance of CUDA BLAS Level 2 Kernels**, 15th SIAM Conference on Parallel Processing for Scientific Computing, 2012. 2. 15,

Savannah, Georgia, USA.

[図書](計 2 件)

片桐孝洋, **スパコンプログラミング入門 - 並列処理と MPI の学習** - 東京大学出版社, ISBN978-4-13-062453-4, 2013 年 3 月 12 日

美添 一樹, 山下 宏 著, 松原 仁 編, **コンピュータ囲碁 モンテカルロ法の理論と実践**, 共立出版, ISBN 978-4-320-12327-4.

[産業財産権]

出願状況 (計 0 件)

取得状況 (計 0 件)

[その他]

ホームページ

http://sudalab.is.s.u-tokyo.ac.jp/~reiji/kiban_a/index.html

6. 研究組織

(1) 研究代表者

須田 礼仁 (SUDA, Reiji)

東京大学・情報理工学系研究科・教授

研究者番号: 40251392

(2) 研究分担者

佐藤 周行 (SATO, Hiroyuki)

東京大学・情報基盤センター・准教授

研究者番号: 20225999

山本 有作 (YAMAMOTO, Yusaku)

電気通信大学・情報理工学研究科・教授

研究者番号: 20362288

今村 俊幸 (IMAMURA, Toshiyuki)

独立行政法人理化学研究所・計算科学研究機構・チームリーダー

研究者番号: 60361838

美添 一樹 (YOSHIZOE, Kazuki)

東京工業大学・情報理工学研究科・研究員

研究者番号: 80449115

(3) 連携研究者

片桐 孝洋 (KATAGIRI, Takahiro)

東京大学・情報基盤センター・准教授

研究者番号: 40345434

藤井 昭宏 (FUJII, Akihiro)

工学院大学・情報学部・講師

研究者番号: 10383986

(4) 研究協力者

弓場 敏嗣 (YUBA, Toshitsugu)

八杉 昌宏 (YASUGI, Masahiro)

大澤 範高 (OSAWA, Noritaka)