様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ６ 月 １９ 日現在

機関番号：１７４０１
研究種目：基盤研究(C)（一般）
研究期間：2017～2019
課題番号：１７Ｋ０００８２

研究課題名（和文）大規模並列計算システム向け低遅延ネットワーク・トポロジに関する研究

研究課題名（英文）A study on low delay network topology for a large scale parallel computing system

研究代表者
飯田　全広（IIDA, Masahiro）

熊本大学・大学院先端科学研究部（工）・教授

研究者番号：７０３６３５１２

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：　本研究課題では，Degree/Diameter Programの研究成果をOrder/Degree Programに接続するためのアルゴリズム開発を行った．手動でCayleyグラフをODPに適用する方法によってODPが解けることが示され，その成果をIEICE論文誌で報告した．また，グラフ処理アクセラレータは，グラフをストリームで処理するためのデータ構造の提案し，IEICE RECONF研究会にて報告した．グラフ処理アクセラレータの開発において，平均最短経路長ASPLの計算が課題であると明確になったが，その解決には至らなかった．しかし，この検討で，シリアル概略加算器を発明し，特許を出願した．

研究成果の学術的意義や社会的意義
グラフ処理に対する新しいアプローチとして，本研究課題で実施した内容は学術的な意義が高い．また，研究過程で生まれたシリアル概略加算器は，グラフ処理のみならず様々な処理に適用できる可能性があることから，学術的，社会的意義は計り知れない．

研究成果の概要（英文）：　In this work, we developed an algorithm to connect the research results of the Degree/Diameter Program to the Order/Degree Program. It is shown that the ODP can be solved by manually applying the Cayley graph to the ODP, and the results are reported in the IEICE journal. In addition, the graph processing accelerator proposed a data structure for processing graphs in streams, which was presented at the IEICE Technical Committee on Reconfigurable Systems (RECONF). In the development of the graph processing accelerator, it became clear that the calculation of the mean shortest path length ASPL was a problem, but it was not solved. However, with this consideration, a serial approximate adder was invented and a patent was applied.

研究分野：コンピュータシステム

キーワード：グラフ処理　概略計算　FPGA　リコンフィギャラブルシステム

１．研究開始当初の背景

「京」をはじめとするスーパーコンピュータ（以下，スパコン）は，日本の科学技術を支える重要な基盤技術である．特にスパコンを用いたシミュレーションは，「理論」「実験」と並ぶ第三の科学技術の手法として益々重要度は上がっている．スパコンは数千個から数万個の計算ノードが相互に接続された大規模の並列計算システムであり，このような大規模システムの高性能化は，各ノードの基本性能を向上するだけでは達成できない．数万の計算ノードを接続するネットワークでは，通信遅延が性能に大きな影響を与えるからである．すなわち，次世代スパコンでは通信遅延の小さい計算ノード間の相互接続方式が求められている．

一般的なスパコン等のネットワークは，メッシュやトーラスなどの直接網やファットツリーなどの間接網が使われるが，ノード数が多くなると任意のノード間の通信遅延は劇的に増加する．一方，大規模並列処理が必要なアプリケーション，例えばビッグデータ解析やディープラーニングなどは，小さなデータがネットワーク上を頻繁に通信されるため，低遅延のネットワーク求められている．従来のメッシュやトーラスのネットワークのトポロジは，ノード数（頂点数）が増えるとグラフの直径，ノード間の平均距離は共に急激に増加する．例えば，**n** 次元トーラスなら，次数は **2n**，直径はノード数の **n** 乗根に比例する．したがって，ノード間の通信遅延も必然的に増加する．同じノード数ならランダムトポロジが直径，平均距離ともに小さいことが知られているが，実装上は困難な側面を持っている．このように大規模な並列計算システムの低遅延ネットワークを実現するためには，以下の技術課題を解決しなければならない．

1) 任意のノード数，所望の直径でノード間の平均距離が最小のネットワークの設計方法，
2) 1)に加え，動的なノードの増減に対応するネットワークの制御方法，
3) ランダムトポロジに近い実現可能なネットワークの実装方法．

これまで **FPGA** の配線構造の研究において，グラフ理論の見地からスモールワールド・ネットワークを応用して配線遅延の最小化する研究を行ってきた．この研究は上記の課題と同様に論理ブロック（ノード）間の結ぶ配線構造をグラフとしてとらえ，そのノード間の平均距離を縮めることで実現した．また，最近では国立情報学研究所主催の小直径グラフ探索コンペ“ **Graph Golf** ”**[1]**において，**2015** 年には入賞，**2016** 年は **Widest Improvement** 部門で第一位を受賞し，**Deepest Improvement** 部門でも第二位となり，両部門で好成績を収めた．このコンペは与えられたノード数，次数を満たす中で直径がなるべく小さい無向グラフを見つける課題であり，本研究課題解決の基礎となる．これらの研究を踏まえ，さらに発展させることで課題解決にあたる．

[1] “Graph Golf: The Order/Degree Problem Competition,”http://research.nii.ac.jp/graphgolf/

２．研究の目的

本研究の目的は，大規模並列計算システムのネットワーク設計手法を確立することにある．

３．研究の方法

グラフ理論では，与えられた次数・直径を満たす中でノード数が最大のグラフを見つける次数・直径問題**(degree/diameter problem; DDP)**に取り組まれており，多くの成果が蓄積されている．しかし，大規模並列システム向けのネットワークでは，ノード数（頂点数）や次数は実装上の制約や予算，電力などの外的要因で決定され，**DDP** の成果を直接トポロジとして採用することできない場合が多い．ネットワーク設計者は，**DDP** とは逆に与えられたノード数・次数を満たす中で，直径・平均距離が小さいグラフを見つけたい．この問題は頂点数・次数問題**(order/degree problem; ODP)**と名付けられている．しかし，**ODP** はまだ研究が始まったばかりで，有効な解決策が見出されていない．

本研究では，以下の３つ項目について研究期間内に実現する．

1) **DDP** の研究成果を **ODP** に接続するためのアルゴリズム開発
2) ネットワーク設計のためのグラフ処理アクセラレータの開発
3) 上記の成果を統合した設計環境の構築と導出されたグラフの評価・実証実験

1)のアルゴリズム開発は，**DDP** で見つかっているグラフを **ODP** に転用するアルゴリズムの開発である．すでに **2016** 年度の“ **Graph Golf** ”においてその雛型が完成しており，その効果は確認されている．このアルゴリズムをエンハンスして完成度を高めると共に実用レベルに引き上げる．具体的には，以下の手順で行う．

1. **Brown Graph** や **Cayley Graph** から指定の直径，ノード数，次数が近いグラフを生成
2. ノード数の過不足を調整（多いときは削除，少ないときは追加）し，リンクを張る
3. 直径，平均距離を求める
4. 指定の直径，ノード数，次数になるまで **1.**から **3.**を繰り返す
5. **2-Opt** 最適化で解を改善する

この中で **1.**は **DDP** の成果（知見）である．**2.**は本研究のポイントであり，どのような指標を

基に追加ノードからリンクを追加するか，またはどのようなアルゴリズムで実施するかを初年度に見極める. 3.は計算時間がかかる部分であり, 次の研究項目で解決を目指す. 5.は Brute force approach であるが，より一段の改善が見込めるかどうかの判断のために実施する．ここで解が更新されるということは，アルゴリズム的にまだ改善の余地があることになる．したがって，ここでの結果をフィードバックさせることでアルゴリズムをブラッシュアップする．

2)のアクセラレータ開発は，様々なグラフ操作を行う場合，通常の PC などでは処理時間がかかり過ぎるため，その解決策として開発する．グラフ探索では，グラフの作成，ノードの追加・削除，リンクの追加削除，ノード間距離の計算，直径の計算，平均距離の計算など様々な処理を行う．これらの処理は，一般的に不規則な制御フローやメモリアクセスパターンを持ち，ビット操作も多いために並列化による高速化が困難である．したがって，マルチコア CPU や GPU 等には向かない処理である．実際，ODP のグラフの探索では数ヶ月間マルチコア CPU を動かし続けなければならないケースもあり，実際のネットワーク設計では高速化を実現できなければ実用化できない．そこで研究代表者が持つ FPGA に関する研究実績を活かしてグラフ処理アクセラレータを開発する．ここではプロセッサに対して 2 桁以上の高速化を目指す．

RGP の中核をなすグラフ処理エンジンは，シストリックアレイを用いたストリーム処理でグラフ操作を行う．これはメモリアクセスを最小化するためと並列にグラフ操作を可能にするためである．図１に RPG システムの基本構成を示す．グラフ処理エンジンは，複数のグラフ処理ユニットでシストリックアレイを構成し，BDD 表現のグラフを読込み，パイプライン的に並列処理を行う．これらの回路は FPGA 上に実装される. FPGA ボードはサーバマシンに複数枚搭載でき，サーバマシンはネットワークを介して拡張できるように設計する．

RPG システムのソフトウェアインタフェースは，Python の NetworkX ライブラリをディスパッチすることで実現し，Python コードをアクセラレーション可能にする．こうすることで Python を用いてグラフ処理プログラムを記述すれば，通常のサーバでも FPGA ボード搭載のサーバでも動作することができる．

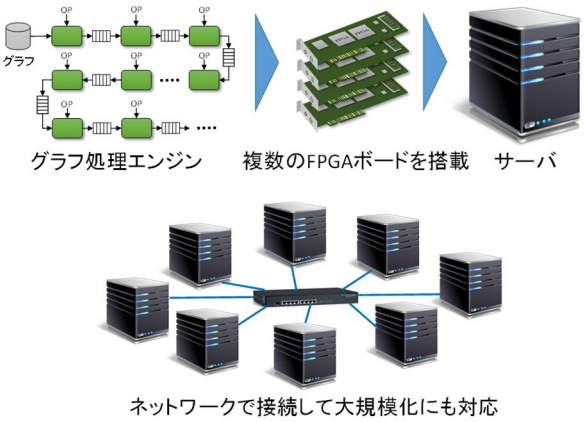


図１RGP システムの基本構成

3)の実証実験では, ODP の解である任意のグラフを大規模並列計算システムのネットワークに適用した場合の性能評価を行う．また，ランダムトポロジに近いネットワークのルーティング問題や動的ネットワークの再構成問題，ラック配置最適化（二次割当て問題）など解決すべき問題は多くある．本研究では，これらの問題に対して完全なランダムグラフではなくスモールワールド性を持つグラフを用いて解決の糸口を見つける．スモールワールド性を持つグラフは，ランダムグラフと同等の直径・平均距離を実現しつつ，ある程度の規則性を持つために，ルーティング問題，ラック問題に対しては有効な手法となりえる．

４．研究成果

1) DDP の研究成果を ODP に接続するためのアルゴリズム開発

平成 29 年度は Degree/Diameter Program（DDP)の研究成果を Order/Degree Program (ODP)に接続するためのアルゴリズム開発を目標に実施してきた．手動ではあるが Cayley グラフを ODP に適用する方法によって ODP が解けることが示され，Graph Golf 2017 において，"General Graph Widest Improvement Award"と"Grid Graph Deepest Improvement Award"の二賞を受賞することができた．この成果は，電子情報通信学会の論文誌(ED Special Issue: Parallel and Distributed Computing and Networking)に投稿した．一方，もう一つの研究家課題であるグラフ処理アクセラレータは，グラフをストリームで処理するためのデータ構造の提案を行った．提案データ構造は，ZDD を修正して利用した．この結果については，電子情報通信学会 RECONF 研究会にて報告した(IEICE Techinical Report RECONF2017-38, FPGA を用いたグラフストリーム処理の一検討，松崎，尼崎，飯田，久我，末吉).

また，平成 30 年度（2018 年度）は，Graph Golf 2018 にて，General Graph Deepest Improvement Awards を受賞した．これまでのグラフ探索の結果は，電子情報通信学会の論文誌(Order Adjustment Approach using Cayley Graphs for Order/Degree Problem, Kitasuka, Matsuzaki, Iida, IEICE TRANSACTIONS on Information and Systems, Vol.101, No.12, pp.2908-2915)として刊行した.

2) ネットワーク設計のためのグラフ処理アクセラレータの開発

これらの結果から, ボトルネックとなる平均最短経路長 ASPL（average shortest path length）

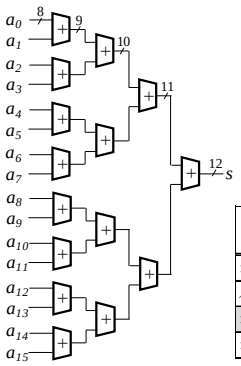
の計算を如何に早くするかが課題であると明確になった．そこで，平成 30 年度（2018 年度）は，グラフ処理のアクセラレーションを行う題材として，ASPL の計算を高速化するために，近似計算のハードウェア・アルゴリズムの検討と評価を行った．

平成 31 年度（2019 年度）は，この近似計算アルゴリズムを ASPL 計算に適応する実験を実施する予定であったが，残念ながら，グラフアクセラレータとして完成させることができなかった．しかし，この研究を通じて考案した概略計算方式は，特許（特願 2019-211774）として申請中である．なお，この発明は，本研究課題ではうまく活用できなかったが，別件の研究課題に適用をすすめている．

本発明は，ビットシリアルデータの上位ビットから受け取り，データの全ビットを受け取る前に加算を開始し，順次，概略計算結果を出力する演算回路である．このとき，演算結果には誤差が生じるが，本発明では，その誤差の範囲が確定できるため，許容誤差範囲であれば，短いレイテンシで結果を得られることが特徴である．従来の方法として，(1)パラレルの概略加算器，(2)概略計算ではないシリアル加算器，(3)本発明であるシリアルの概略加算器を比較する．

(1)のパラレルの概略加算器の ESA は，クリティカルパスになるキャリーの伝搬を切って，セグメント化することで正確性を犠牲にしても高速化を実現する加算器である．また，LOA は，加算の下位ビットの方は全体の誤差に対して影響度か低いことを利用して，下位の方を加算の代わりに OR ゲートにすることで ESA と同じ効果を得る加算器である．どちらも精度と演算速度のトレードオフを諮る方式である．本発明はシリアル演算で概略計算できる点が異なる．

(2)の従来のシリアル加算器は，先に説明した通り下位ビットから入力し，下位ビットから出力するシリアル加算器である．上位ビットがわからないので最終的な加算結果がすべての入力ビットが来るまでわからない．本発明は，上位ビットから加算を行うので，最初から加算結果の大体の値が出力でき，ある一定の誤差を含むが，入力ビットが増えるにしたがって加算結果が正確になっていく．



(b) N=16の回路構成例

(a) アプリケーション例

$$s = \sum_{i=0}^{N} a_i$$

(c) 各方式の計算時間（クロックサイクル数）

Mビットの数値N個の合計を求める場合の計算時間	N	16				32				64				128			
	M	8	16	32	64	8	16	32	64	8	16	32	64	8	16	32	64
シリアル加算器	$M+2\log_2 N$	16	24	40	72	18	26	42	74	20	28	44	76	22	30	46	78
パラレル加算器	$M+\log_2 N$	12	20	36	68	13	21	37	69	14	22	38	70	15	23	39	71
シリアル概略加算器（最短）	$\log_2 N$	4	4	4	4	5	5	5	5	6	6	6	6	7	7	7	7
シリアル概略加算器（最終）	$M+2\log_2 N$	16	24	40	72	18	26	42	74	20	28	44	76	22	30	46	78

図 2 アプリケーション例と計算時間の比較

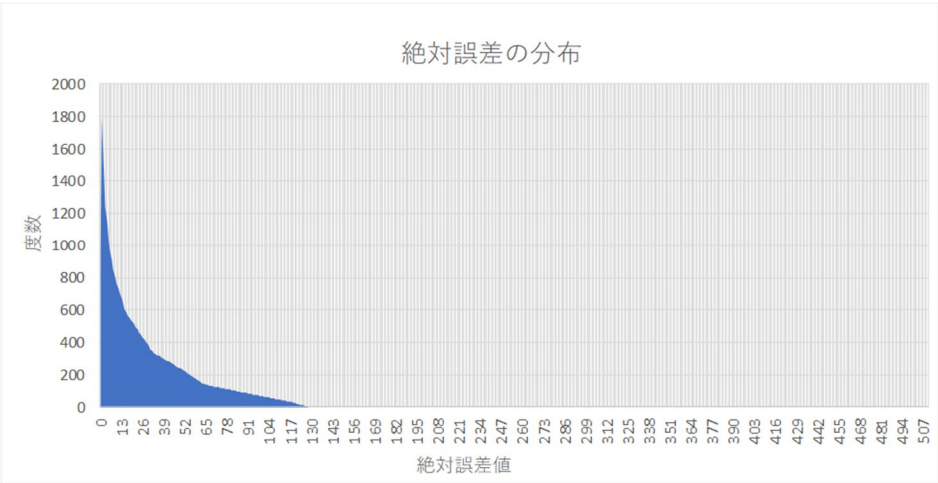


図 3 絶対誤差の分布（８ビット加算の場合）

次に，これらの３方式について，図に示した総和を取るアプリケーションを例に計算時間を比較する．図の(a)に示すように，N 個のデータの総和をツリー状に構成した加算器を用いて計算する．入力 *a0*～*a15* の 16 個のデータの回路図例を図の(b)に示す．(c)には一般化して各入力データのビット数を M，データの個数を N として総和を求める計算時間（ク

ロック数）を算出している．シリアル加算器は，$M + 2\log_2 N$クロックで計算を完了する．パラレル加算器は，格段の計算は１クロックで計算できるとすると$M + \log_2 N$クロックである．それに対して，本発明は，最初のデータが出てくるのには，最短$\log_2 N$クロックしかかからず，すべてのデータの計算が終了するのはシリアル加算器と同じ$M + 2\log_2 N$クロックである．これらからわかる通り，本発明のシリアル概略加算器は，圧倒的に少ないクロック数で概算計算を出力することができる．

図 3 は，2 つの 8 ビットの数を本発明のシリアル概略加算器で最後まで計算したときの絶対誤差の分布である．8 ビットの値は 0～255 まで表現できるので，2 つの 8 ビットの数値の加算結果は，0～510 の値の範囲に入る．この結果は，510 までの絶対誤差値がいくつ出現したか，すなわち，度数をグラフ化している．ここからわかるように，本発明の概略加算器は，絶対誤差が小さい値に多く分布していることから計算結果への影響は小さいといえる．また，絶対誤差の最大値は 128 で，最小値は 0 である．

本発明の後に解決すべき課題は，（１）計算精度の制御方式，（２）乗算の概略計算の２点ある．（１）の計算精度の制御とは，現在の回路構成は入力された１ビットのみを用いて計算している．しかし，これを２ビット，３ビットと増やすことで概略計算の誤差が小さくなることがわかっている．したがって，入力値のビット数を制御して計算精度を制御する回路方式を確立する必要がある．（２）の乗算の概略計算は，一般的に乗算は時間がかかる計算であり，上位桁から計算が難しい演算である．しかし，アプリケーションの多くは乗算と加算をセットで用いる積演算のような計算が多く，実用化のためには乗算の概略計算方式を確立する必要がある．本研究課題では，これらの課題の解決に至らず，グラフアクセラレータとして完成させることができなかった．

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 T Kitasuka, T Matsuzaki, M Iida	４．巻 101(12)
２．論文標題 Order Adjustment Approach Using Cayley Graphs for the Order/Degree Problem	５．発行年 2018年
３．雑誌名 IEICE TRANSACTIONS on Information and Systems	６．最初と最後の頁 2908-2915
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transinf.2018PAP0008	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 松崎貴之，尼崎太樹，飯田全広，久我守弘，末吉敏則	４．巻 117
２．論文標題 FPGAを用いたグラフストリーム処理の一検討	５．発行年 2017年
３．雑誌名 信学技報，vol．117，no．279，RECONF2017-38，pp．7-12，2017年11月．	６．最初と最後の頁 7-12
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計1件（うち招待講演　0件／うち国際学会　0件）

１．発表者名 松崎貴之
２．発表標題 FPGAを用いたグラフストリーム処理の一検討
３．学会等名 デザインガイア2017 -VLSI設計の新しい大地-
４．発表年 2017年

〔図書〕　計0件

〔出願〕　計1件

産業財産権の名称 演算装置、及び演算方法	発明者 飯田全広，尼崎太樹，古賀大顕	権利者 同左
産業財産権の種類、番号 特許、特願2019-211774	出願年 2019年	国内・外国の別 国内

〔取得〕　計0件

〔その他〕

－

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
連携研究者	北須賀　輝明 (Kitasuka Teruaki) (70343332)	広島大学・工学研究院・准教授 (15401)	