社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2007－HPC－111 (32)
2007／8／2

次世代光インターコネクト上でのMPIアプリケーションの評価

滝澤 真一朗[†] 遠藤 敏夫[†] 松岡 聡[†,††]

将来の数万プロセッサ規模のシステムでは，全ノードを高バンド幅で全対全接続するネットワークはコストや電力消費の問題で構築が難しい．この問題を解決するため，低バンド幅電気パケットネットワークと光サーキットネットワークの双方を活用するネットワークを提案する．光ネットワークは電気スイッチをまたぐ離れたノードとの通信にのみサプリメンタルに使用する．この環境でのMPIアプリケーション通信は，光回線に接続されているプロセスが通信パターンから構築したトポロジに沿って，他プロセスのスイッチをまたぐ通信をフォワードことで実現する．提案ネットワーク，提案通信手法をNas Parallel BenchmarksのMGで評価した結果，電気ネットワークだけを用いた場合よりプロセス間距離が短くなり，実行性能向上が見込めることが確認できた．

Anarysis of MPI Applications over Next Generation Optical Interconnect

Shin'ichiro Takizawa,[†] Toshio Endo[†] and Satoshi Matsuoka[†,††]

For the future tens of thousands of processors systems, it is difficult to construct interconnects which fully connect all nodes with high bandwidth due to cost and power consumption. We propose a network which utilizes both fully-connected low bandwidth electronic packet switched network and optical circuit switched network. Optical network is supplimentally used only when a node communicates with nodes in other packet switches. MPI application runs on this environment in such manner that processes connect to optical circuits forward other processes' messages that cross packet switches, in accordance with a topology constructed from communication pattern. As a result of evaluations, our proposal achieves lower inter-process distance than electronic network.

1. はじめに

将来のペタスケール時代のスーパーコンピュータは，シングルプロセッサコアのクロック上昇率の頭打ちのため，マルチコアプロセッサを数千から数十万搭載した高度に並列化されたシステムとなりうる．そのようなシステムのノード間インターコネクトとして，過去のスーパーコンピュータやクラスタシステムで用いられていた，パケット交換方式を採用するクロスバーやFat Treeなどの全体全接続ネットワークは，コスト面・性能面において現実的ではなくなる．この問題は現状のシステムでも確認できる．例えば，Blue Gene/Lは65536個のプロセッサを接続数の少ない3Dトーラスネットワークに接続し，各プロセッサがメッセージをフォワーディングすることにより離れたプロセスとの通信を実現している[1]．また，東京工業大学のTSUBAME Grid Cluster[2]では全対全接続を維持しているものの，上流のバンド幅が低い構成になっている．

この問題の解決策として，次世代インターコネクトとして，各ノードを安価な低バンド幅電気パケットネットワークと，高バンド幅光サーキットネットワークの双方に接続するネットワークが提案されている[3),4]．これらネットワークではサイズの小さいメッセージは電気ネットワークで送信され，サイズの大きいメッセージは光ネットワークを用いて，ノード間で回線を確立した上で送信される．光回線確立には数ミリ秒時間を要するが，これら手法では通信パターンの解析，予測を行い，通信が起こる前にあらかじめ回線を確立する方法が提案されている．高価な高バンド幅パケットネットワークを使用せず，高価で高消費電力なOEO（Optical-Electrical-Optical）変換機が必要ない光サーキットネットワークを用いるため，低コストで実装できるメリットがある．一方，デメリットとして大容量メッセージを全対全で交換する場合に性能低下が見込まれる．しかしMPI並列アプリケーショ

† 東京工業大学
Tokyo Institute of Technology
†† 国立情報学研究所
National Institute of Informatics

ンの多くは通信に局所性があり，各プロセスは総プロセス数に対してはるかに少ない数の相手としか通信をしない．さらに，集団通信の多くは小容量メッセージの交換に使用されていることが報告されている[5]．そのため，通信の局所性を満たすようにプロセス配置，あるいは光回線確立の管理・スケジューリングを行なえば，性能を維持したままコストを削減することができる．しかし，これら既存研究では光ネットワーク部の規模が大きくなるため，大規模環境の構築は現実的ではない．

本研究では，安価な低バンド幅電気パケットネットワークと光サーキットネットワークの双方を活用したノード間インターコネクトを提案する．提案ネットワークでは，計算ノードは電気ネットワークと光ネットワークにそれぞれ1つずつNICを持ち，光ネットワークは，電気ネットワークにおいてスイッチをまたぐ大容量通信が起こる場合にのみ，サプリメンタルに利用する．また，バンド幅の低い電気ネットワークの上流リンクの利用を減らし，かつ，光ネットワークの利用頻度を減らすために，ノード間でメッセージのフォワーディングを行なう．提案ネットワーク上でMPIアプリケーションにおける通信相手との距離を評価したところ，電気スイッチ間を接続するのに十分な少ない光回線数を割り当てることで，電気ネットワークだけを用いた場合より短距離で通信を行えることを確認した．

2. 関連研究

光ネットワークを用いたグリッド・クラスタ環境でのMPIアプリケーションの評価を行なっている既存研究があり，本章ではそれらで用いられているネットワークトポロジとその利用法について述べる．

Barkerらは，各ノードが低バンド幅電気パケットネットワークに1つのNICを，光サーキットネットワークに多数のNICを持つネットワークを提案している[3]．この環境での一対一通信はまず電気ネットワークを用いて開始される．そして，トラヒックを監視しつつ，通信データ量が増加した場合には，光回線を確立し光ネットワーク上で通信を行なう．2ノードのいずれかでも全光NICを使いきっている場合には，LRUルールに従い古い回線を解放し，新規に回線を確立する．一方，集団通信は電気ネットワーク上でのみ行われる．本研究とは，各ノードに光ネットワークに多数のポートを持たせる点と，集団通信を電気ネットワーク上だけで行なっている点が異なる．

Kamilらは低バンド幅電気パケットネットワークと光サーキットネットワークを組み合わせたハイブリッドネットワークHFASTを提案している[4]．HFASTは電気ネットワークと計算ノードの間にコネクションプールとして光ネットワークをはさむ構成になっている．従来のネットワークを用いた場合には局所性のあるノード同士の通信を最適化するためにはタスクマイグレーションが用いられていたが，HFASTを用いれば光回線の割り当て問題となり，軽量な通信最適化が可能である．この研究では帯域遅延積以上のデータ転送にのみHFASTネットワークを用い，小容量データ転送や集団通信には別の低バンド幅電気ネットワークを用いる．2つの分断された電気ネットワークを用いている点と，集団通信は電気ネットワーク上だけで行う点が本研究とは異なる．

光ネットワーク環境でのMPIアプリケーション性能評価として，Kimらはサイト間を光ネットワークで接続した光グリッド環境と，シングルクラスタ環境での実行性能の比較評価を行なっている[6]．光ネットワークを用いると，IP通信に比べ通信遅延のばらつきが小さくなるため，MPI_Barrierなど同期を取る際にprocess skewが起こりにくくなると述べている．MPIアプリケーションは，光グリッドを用いると最大で倍性能が向上すると述べているが，光グリッド環境のノード数とシングルクラスタ環境のノード数が異なること，アプリケーションの通信量を一切考慮していないことより，フェアな評価とは言えず，更なる詳細な評価が必要である．また，実験に用いられた光グリッドも2サイトを接続したものと小規模な環境である．

井本らは計算ノードが直接光ネットワークに接続された環境でのMPIライブラリの実装を行なっている[7]．このライブラリは，リングトポロジ型光ネットワーク上に構成された共有メモリインターフェースを介して通信を行なう．しかし，リングネットワーク上の通信は単一トークンパッシング方式であり，1度に1プロセスしか通信を行えないため，Ethernetを用いた場合より実行性能が落ちている．この性能低下はMPI_Alltoallのように，全体全でメッセージを交換し合う場合に特に顕著に現れると考えられる．光ネットワークしか利用しない点で本研究とは異なる．

3. サプリメンタル光ネットワークの提案

3.1 サプリメンタル光ネットワーク

将来の高度に並列化された並列計算機のノード間インターコネクトとして，図1に示すネットワーク環境を提案する．計算ノードは電気パケットネットワークと光サーキットネットワークにそれぞれ1つずつNICを持つ．電気ネットワークはTreeネットワークのような全対全接続である必要があるが，上流リンクのバンド幅は低くてよい．光ネットワークには光バーストスイッチング技術を用い，各スイッチはメッシュ上に接続されており，数ミリ秒単位での回線切り替えが可能とする．各ノードは光ネットワークへは1つしかNICを持たないので，回線交換方式の性質上，同時に接続

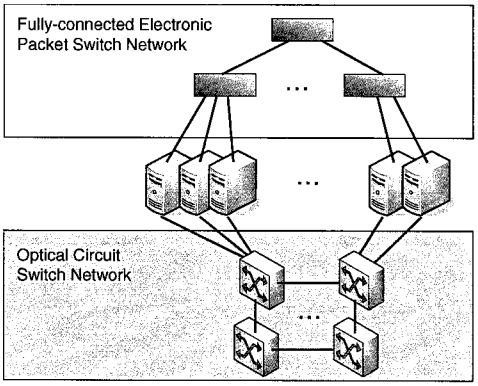

図 1 提案ネットワーク

できる通信相手は 1 ノードに限られる．しかし，4 章で具体的に述べるように，光ネットワークは電気ネットワークの補完として使用するため，接続数が 1 でも問題はない．また構築コストの都合から，回線確立ミス（呼損）が起こりうる波長数の少ない安価なネットワークを構築しても機能する．

提案ネットワークは安価な機材のみで構築できるので，数万ノードからなる大規模環境にも適応できる．また既存環境を，光ネットワークを追加するだけで，容易に拡張することも可能である．

3.2 他ネットワークとの比較

提案手法と 2 章で紹介した関連研究で用いられているネットワークとの比較を，全対全接続性，電気ネットワークのポート数，光ネットワークのポート数，光ネットワークにおける呼損の可否の 4 項目について行い，結果を表 1 にまとめた．表中の N はノード数を，n は電気スイッチ数を，k はノードあるいはスイッチ 1 つが持つ光 NIC 数を表す．ここで，ネットワークのポート数はネットワークとノード，あるいは別のネットワークを接続可能なスイッチにおけるポート数の合計を意味する．また，「光バックエンド」は Kim らの光グリッド環境や，電気ネットワークのバックエンドに光ネットワークを使用する手法を一般化したネットワークを表し，「全光ネットワーク」は井本らのリング型光ネットワークのなどを一般化したネットワークを表す．ここでは回線交換方式の光ネットワークを用いたとして比較を行なった．

本提案や Barker ら，Kamil らの提案ネットワークでは，全対全接続の電気パケットネットワークを用いるため，全対全接続性がある．光バックエンドや全光ネットワークでは，電気スイッチやノードに搭載する光 NIC 数に応じて，全対全接続性は決まる．しかし，金銭的コストやノード構成により搭載できる NIC 数に制限があるため，全対全接続を実現するのは難しい．

本提案，Barker ら提案，光バックエンドネットワークにおいては単一電気ネットワークを用いるので，電気ネットワークのポート数はノード数 N に等しい．一方，Kamil ら提案のネットワークは，2 つの電気ネットワークを用いるため，そのポート数は最大で $2N$ になる．

提案ネットワークでは各ノードは 1 つの光 NIC を持つので，光ネットワークのポート数はノード数 N に等しくなる．一方，Barker ら提案では各ノードが複数光 NIC を持つこと許可しているため，ポート数は kN になる．Kamil ら提案の HFAST では，光ネットワークは N 個のノードを最大 N 個の電気ポートに接続するため，最大 $2N$ のポート数となる．光バックエンド，全光ネットワークにおいては，搭載する NIC 数に応じてポート数が決まる．

最後に呼損について，提案ネットワークにおいては光回線が確立できない場合でも電気ネットワークで通信が行えるので問題が無い．同様なことは Barker ら，Kamil ら提案にも当てはまる．しかし，光バックエンド，全光ネットワークにおいては主な通信を光ネットワーク上で行なうため，呼損の発生は通信不能，大幅な性能低下につながる．特に光バックエンドの場合には，スイッチ以下の全ノードが通信できなくなる問題が起こる．

以上より，ポート数が少なく，光ネットワークの呼損の影響が少ない提案手法は，他手法に対し安価に構築可能である．

4. 提案ネットワークにおける MPI プロセス通信

4.1 ネットワーク使用方法

提案ネットワーク上で MPI アプリケーションを実行する際には，以下の条件にマッチする通信のみ光ネットワークを使用する．

- 一定サイズ以上のメッセージ交換を行なう場合
- 電気パケットネットワークでスイッチをまたいだ通信が起こる場合

1 つ目の条件は，高バンド幅の光ネットワークを有効活用するために，サイズの小さいメッセージは従来どおりの電気ネットワークで転送するという方針である．メッセージサイズの閾値は，光ネットワークの帯域遅延積以上のサイズとする．2 つ目の条件は，1 つ目の条件を満たす通信において，バンド幅の低い電気ネットワークの上流リンクは用いずに，通信ペア間で光回線を確立し，光ネットワーク上で通信を行なうという方針である．この 2 つの条件から，提案ネットワークにおける通信は以下にまとめられる．

- サイズの小さい一対一通信，集団通信はパケット通信のみで行なう
- サイズの大きい一対一通信，集団通信には光回線通信も用いる

表 1　提案ネットワークと関連研究で提案されているネットワークの比較

ネットワーク	全対全接続	電気ネットワークポート数	光ネットワークポート数	呼損の可否
提案ネットワーク	あり	N	N	可
Barker ら提案	あり	N	kN	可
Kamil ら提案	あり	最大 $2N$	最大 $2N$	可
光バックエンド	構成依存	N	kn	否
全光ネットワーク	構成依存	0	kN	否

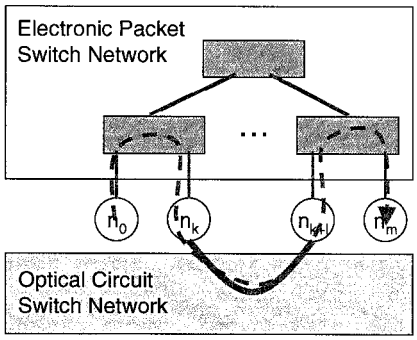

図 2　提案ネットワークにおけるフォワーディング例

- 電気パケットスイッチ内通信で済む場合には光ネットワークは使用しない

このように提案手法では，バンド幅の低い電気ネットワークの上流リンクのショートカットとして，サプリメンタルに光ネットワークを使用する．

理想としては，電気スイッチをまたいだ全てのプロセス間通信は光ネットワークを用いて行ないたいが，光ネットワークの規模や他アプリケーションの光回線利用状況により，十分な光回線を用意できない場合もある．そこで，アプリケーションの通信パターンを基に，利用可能な光回線だけを用い，あらかじめプロセス間で通信トポロジを構築し，その上でのフォワーディングテーブルを作成し，各プロセスがメッセージをフォワードする手法を用いる．例えば，図 2 において，ノード n_k とノード n_{k+l} だけが光回線で接続されている場合に，ノード n_0 がノード n_m に大容量メッセージを送信する場合には，図中の点線で表される矢印に従いメッセージをフォワードする．また，光回線数が足りず，他のスイッチ下ノードと光回線で接続されない孤立スイッチができてしまった場合には，電気スイッチの上流リンクで通信を行なう．このようにあらかじめ光パスを確立し，フォワーディングテーブルを作ることにより，アプリケーション実行中に時間のかかる光回線切り替えを行なわずに済む．しかし，アプリケーションの通信パターンを取得する必要があるので，実現するためには事前実行をするか，繰り返し処理を行なうアプリケーションである必要がある．

4.2　フォワーディングテーブル作成アルゴリズム

フォワーディングテーブル作成アルゴリズムを 2 種類提案する．どちらのアルゴリズムも本質的にはプロセスをグルーピングし，グループを超えた通信に光回線を使用する方式である．また，アルゴリズムの前提として，各プロセスは電気ネットワークのトポロジ情報を取得できるとする．

最初に，プロセスの電気スイッチ配置によりグルーピングを行なう Switch Partitioning 方式を提案する．プロセスは電気ネットワークトポロジを知っているので，どのプロセスとの通信がスイッチをまたぐか識別できる．そこで，スイッチをまたぐ通信を行なうプロセス間に光回線を割り当てる．少ない光回線数でも全スイッチ間を接続できるように，ラウンドロビン方式でスイッチ間に回線を割り当てる．このとき，同じスイッチ間通信を行なうプロセスペアに対しては，ランク値の小さいプロセスから順に回線を割り当てることにした．

このように構築した電気・光混合ネットワーク上でバンド幅を基準とした距離ベクトル型アルゴリズムを用い，フォワーディングテーブルを作成する．すなわち，高バンド幅の光回線に接続されたプロセスは他プロセスのスイッチをまたぐ通信を中継することになる．しかし，この方法では離れたスイッチ下のプロセスとの通信が多くなり，光回線に接続されたプロセスに負荷が集中する問題がある．

そこで，アプリケーションの一対一通信トポロジによりプロセスをグルーピングする Communication Partitioning 方式を提案する．この方式の目的は，通信を行なうプロセス同士を同一スイッチ下に配置することにある．通信量の多いノード間リンクを切断することでスイッチ数にプロセスグループを分割し，切断されたリンクに光回線を割り当てる．光回線割り当て方式，フォワーディングテーブル作成方式は最初の手法と同様である．この方式を採用するためには，あらかじめ通信トポロジにマッチするようにプロセスを配置するか，プロセスマイグレーションや MPI ランク値の再割り当てを行なう必要がある．

5.　評　　価

提案ネットワークを用いた場合のプロセス間距離について評価を行なう．Nas Parallel Benchmarks の MG，クラス C，プロセス数 64 を実行して通信パターンを抽出し，図 3 に示す 64 ノードの提案ネットワー

クに適用した場合の距離を計算により求めた．この環境では，電気ネットワークの4機の下流スイッチ下に16ノードが1Gbpsで接続されており，上流スイッチと4Gbpsで接続されている．また，光ネットワークの1回線のバンド幅はプロセスが実行されるノードのバス速度に律速されるとし，PCI Express x16のバンド幅32Gbpsとした．1ノード1MPIプロセスとした．

この環境でのプロセス間通信は，下流電気スイッチ内通信（$Comm_e$），電気スイッチをまたぐ通信（$Comm_{es}$），光回線通信（$Comm_o$），光回線によるフォワーディング通信（$Comm_{fk}$，kはフォワーディング回数），の4パターンある．ノード・下流電気スイッチ間リンク距離をL_{el}，電気スイッチ間リンク距離をL_{eu}，光回線に接続されたノード間リンク距離をL_oとした場合，各通信でのプロセス間距離は以下のように定義できる．

$$Comm_e = L_{el}$$
$$Comm_{es} = L_{el} + L_{eu} + L_{el}$$
$$Comm_o = L_o$$
$$Comm_{fk} = L_{el} + (L_o + L_{el}) \times k$$

クラスタのようにノードが密接した環境を想定しているので，電気ネットワーク，光ネットワーク共に通信遅延は十分小さいとし，対照的に大きく異なるバンド幅を距離の尺度とした．以降の評価では各リンクの距離をバンド幅の逆数を取り，それぞれ$L_{el} = 32$，$L_{eu} = 8$，$L_o = 1$とし，光回線通信ノードペア数（光回線数，以下ペア数）を変化させたときの平均距離，最大距離を求める．

提案2手法と，電気ネットワークだけを用いた場合の比較を行なう．Switch Partitioning方式（S方式）と，電気ネットワークだけの場合には，MPIランク値は図中の左のプロセスから順に割り振る．提案2手法の場合には，8192バイト以上のメッセージのみ光ネットワークを使用するとした．メッセージをフィルタすることにより，MGの各プロセスの一対一通信相手数は6になる．Communication Partitioning方式（C方式）では，このフィルタされた通信グラフを通信量の多いプロセス間リンクを切断することにより4等分し，スイッチ以下に配置する．そのため，MPIランク値の順はグラフ分割に基づき，ランダムになる．

各プロセスの6つの一対一通信相手との平均・最大距離の変移を図4に示す．この図における平均は，通信回数による加重平均を表す．図中凡例のS，C，Eはそれぞれ S方式，C方式，電気ネットワークだけを用いた場合を表す．S方式，C方式の場合は上流電気ネットワークは用いないため，ペア数が2以下の場合には4機のスイッチを接続することができないので結果がない．S方式では，ペア数が4以上になると最大距離がフォワーディング1回の理論上最短の$Comm_{f1} = 65$を達成する．これはMGの特性で，図のようなトポロジの場合，表2に示すように，各スイッチ内プロセスは他の2つのスイッチ内プロセスとしか通信せず，4ペアあれば全ノード間通信が行えるようにスイッチ間を接続できるためである．一方，C方式の場合にはランク値が分散しているため，スイッチ間が全対全に接続される必要があり，理論最短達成には6ペア必要になる．すなわち，光回線数が6以上あればフォワーディングによる最大距離$Comm_{f1}$が65になり，$Comm_{es}$の72より短くなるため，平均距離も短くなり，本手法を用いた場合の実行性能向上が見込める．一方，回線数が6未満の場合には繰り返しフォワーディングされるため距離が長くなる．

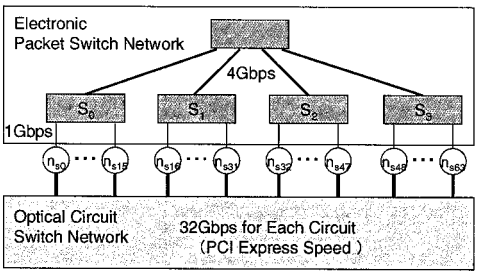

図3 想定環境

次に，全プロセスとの通信距離の変移を図5に示す．これは提案ネットワーク上で，一対一通信の通信パターンを基に構築したトポロジ，フォワーディングテーブルを用いて，大容量集団通信を実行する際の通信性能に影響する．集団通信を実現するため，スイッチ間を接続できないペア数が3未満の場合には，上流電気ネットワークも使用した距離を計算した．S方式の最大距離がペア数に関係なく総じて他手法の最大距離より大きいのは，MGのスイッチ間接続の特性による．表2に示されている通り，各スイッチ下プロセスは特定のスイッチ下プロセスとは一対一通信を行なわないため，それらスイッチ内プロセス間には光回線が割り当てられない．そのため，それらスイッチ間で通信を行なう場合には2回フォワーディングする必要があり，最大距離が$Comm_{f2} = 98$となる．一方のC方式では，6ペア以上あればスイッチ間を全対全で接続できるので，最大距離は一対一通信の場合と同じ$Comm_{f1}$である．すなわち，光回線数が6以上であればS方式，C方式共に平均距離は電気ネットワークの平均距離以下になるため，本手法は有効である．しかし実際には，集団通信の性能は種類やアルゴリズムによって大きく変わるため，今後詳細に検討する予定である．

6. おわりに

本研究では計算機間インターコネクトとして，電気パケットネットワークと光サーキットネットワーク

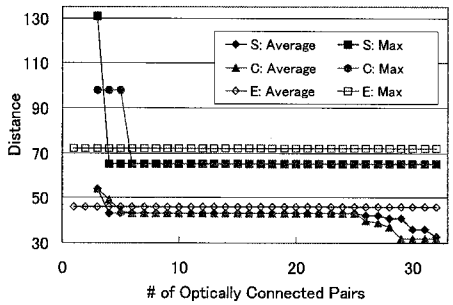

図 4　NPB MG の一対一通信相手との距離

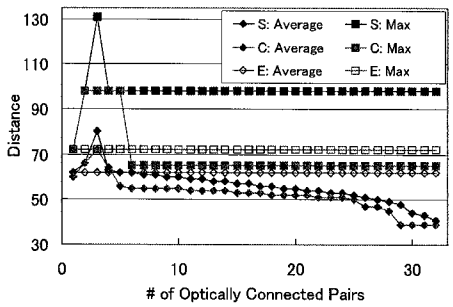

図 5　NPB MG の全対全通信相手との距離

表 2　S 方式における NPB MG のスイッチ間通信数

Switch ID	S_0	S_1	S_2	S_3
S_0	–	16	0	16
S_1	16	–	16	0
S_2	0	16	–	16
S_3	16	0	16	–

の双方を活用するネットワークを提案した．電気ネットワークの上流リンクのバンド幅は高い必要がなく，ノードあたりの光 NIC 数も 1 枚だけなので，既存環境を容易に安価に拡張できる．さらに提案ネットワーク上での MPI アプリケーションの通信の局所性を利用した通信アルゴリズムを提案した．このアルゴリズムを用いることで，少ない光回線数で，電気ネットワークだけを用いた場合よりプロセス間の平均距離が短くなり，実行性能向上が見込めることが確認できた．

しかし現時点ではまだ評価は十分でなく，さまざまなアプリケーションの通信パターンを分析し，それらが提案ネットワークに適応可能か調査するとともに，以下の項目を今後の課題として考えている．

- プロセス数の多い，より大規模な環境を用いた性能評価
- スイッチや計算ノードにおける，混雑や輻輳を考慮した通信性能評価
- 提案環境での MPI アプリケーション実行性能をシミュレーションにより評価
- Barker らや，Kamil らの提案するネットワークやその他光ネットワークを利用した環境上でのアプリケーション実行性能との比較

謝辞　本研究の一部は科学研究費補助金特定領域研究 (18049028) の補助による．

参 考 文 献

1) Davis, K., Hoisie, A., Johnson, G., Kerbyson, D. J., Lang, M., Pakin, S. and Petrini, F.: A Performance and Scalability Analysis of the BlueGene/L Architecture, *SC '04: Proceedings of the 2004 ACM/IEEE conference on Supercomputing*, Washington, DC, USA, IEEE Computer Society, p. 41 (2004).

2) 松岡 聡: TSUBAME の飛翔：ペタスケールへ向けた「みんなのスパコン」の構想, 情報処理学会研究報告　2006-HPC-107 (pp37-42 July 31) (2006).

3) Barker, K. J., Benner, A., Hoare, R., Hoisie, A., Jones, A. K., Kerbyson, D. J., Li, D., Melhem, R., Rajamony, R., Schenfeld, E., Shao, S., Stunkel, C. and Walker, P.: On the Feasibility of Optical Circuit Switching for High Performance Computing Systems, *SC '05: Proceedings of the 2005 ACM/IEEE conference on Supercomputing*, Washington, DC, USA, IEEE Computer Society, p. 16 (2005).

4) Kamil, S., Pinar, A., Gunter, D., Lijewski, M., Oliker, L. and Shalf, J.: Reconfigurable Hybrid Interconnection for Static and Dynamic Scientific Applications, *ACM International Conference on Computing Frontiers* (2007).

5) Shalf, J., Kamil, S., Oliker, L. and Skinner, D.: Analyzing Ultra-Scale Application Communication Requirements for a Reconfigurable Hybrid Interconnect, *Proceedings of the ACM/IEEE SC 2005 Conference* (2005).

6) Kim, D., Jin, H.-W., Jeong, K., Lee, J. and Noh, M.: Performance Measurement and Analysis of High-Performance Parallel Applications over Lambda Grid, *The 9th International Conference on Advanced Communication Technology*, Vol. 1, pp. 792–796 (2007).

7) Imoto, M., Taniguchi, E., Baba, K. and Murata, M.: Implementation and evaluation of MPI library with Globus toolkit for establishing λ computing environment, *Proceedings of 6th Asia-Pacific Symposium on Information and Telecommunication Technologies*, pp. 421–426 (2005).