

特別研究報告

題目

データセンターのための 3次元オンチップ型ネットワーク構造の検討と評価

指導教員

村田 正幸 教授

報告者

池田崇栄

平成 25 年 2 月 12 日

大阪大学 基礎工学部 情報科学科

データセンターのための
3次元オンチップ型ネットワーク構造の検討と評価

池田崇栄

内容梗概

近年、情報処理におけるデータセンターの持つ役割が大きくなり、データセンターで処理されるデータ量が増加している。年々増加する大量のデータを処理するために、数万台以上のサーバーを収容する大規模データセンターが構築されるようになってきている。データセンターが大規模になるにつれ、その消費電力は大きなものとなり、低消費電力なデータセンターの構築が大きな課題となっている。データセンターの処理を低消費電力で実現する手法として、Network on Chip (NoC) の技術を用いる事により、データセンター内のサーバー間のネットワークを1チップに集約するオンチップ型データセンターの検討が進められている。オンチップ型データセンターでは、チップ上に、演算やデータ保存の機能を持つ多数のコアを配置する。オンチップ型データセンターでは、データセンターにおける多数のサーバーとサーバー間のネットワークを1チップに集約することにより、従来のデータセンターよりも消費電力を抑える事が期待できる。しかしながら、オンチップ型データセンターはコンセプトの提案と、オンチップ型データセンター内でのキャッシュの配置に関する検討がされているのみで、オンチップ型データセンターに適したネットワーク構造は明らかにされていない。そこで本報告では、オンチップ型データセンターに適したネットワーク構造の検討を行う。本報告では、データセンターに適した構造として、パケット交換スイッチと回線交換スイッチを、複数階層のチップを積層した3次元チップ上に立体的に配置することにより、低消費電力と低遅延を両立可能な3次元オンチップ型ネットワークの構造を検討する。複数階層を積層した構造である3次元オンチップ型ネットワークにおいて、階層間の接続、サーバーに該当するコアと接続するスイッチの配置階層、階層内のパケット交換スイッチ・回線交換スイッチの配置の3つの項目に分けて検討を行う。検討の結果、パケット交換スイッチから全階層の回線交換スイッチにリンクを構築し、各サーバーに該当するコアに接続されたパケット交換スイッチはすべて同一階層に配置し、それ以外の階層は回線交換スイッチのみで構成した構成が、遅延・消費電力、いずれの観点においても優れていることが分かった。また、その構成では、2次元格子型ネットワークを構成した場合と比べて、消費電力を24%削減、遅延を55%削減できることが分かった。

主な用語

NoC、データセンター、消費電力、遅延、3次元オンチップ型ネットワーク

目次

1	はじめに	6
2	関連研究	8
2.1	データセンターネットワーク	8
2.2	Network on Chip	8
3	提案するオンチップ型データセンターネットワークの検討と評価	11
3.1	オンチップ型データセンターネットワークの概要	11
3.2	検討に用いるモデル	14
3.3	検討項目	17
3.4	検討結果と考察	21
4	終わりに	43
	謝辞	44
	参考文献	45

目 次

1	オンチップ型データセンターにおけるチップ化対象	12
2	データセンター向け3次元オンチップ型ネットワーク	13
3	検討に用いる階層間の接続構成	18
4	検討に用いるサーバーとスイッチの接続構成	19
5	検討に用いる階層内の配置構成	20
6	階層間の接続構成の消費電力量による比較(平均500 bit、標準偏差20 bitの正規分布に従うトラフィックが全サーバー間に流れる場合)	25
7	階層間の接続構成の消費電力量による比較(平均500 bit、標準偏差100 bitの正規分布に従うトラフィックが全サーバー間に流れる場合)	26
8	階層間の接続構成の消費電力量による比較(特定サーバー間トラフィックの場合)	27
9	階層間の接続構成の遅延による比較(特定サーバー間トラフィックの場合)	28
10	サーバーとスイッチの接続階層の消費電力量による比較(平均500 bit、標準偏差20 bitの正規分布に従うトラフィックが全サーバー間に流れる場合)	32
11	サーバーとスイッチの接続階層の消費電力量による比較(平均500 bit、標準偏差100 bitの正規分布に従うトラフィックが全サーバー間に流れる場合)	33
12	サーバーとスイッチの接続階層の消費電力量による比較(特定サーバー間トラフィックの場合)	34
13	サーバーとスイッチの接続階層の遅延による比較(特定サーバー間トラフィックの場合)	35
14	階層内の配置構成の違いによる消費電力量による比較(平均500 bit、標準偏差20 bitの正規分布に従うトラフィックが全サーバー間に流れる場合)	39
15	階層内の配置構成の違いによる消費電力量による比較(平均500 bit、標準偏差100 bitの正規分布に従うトラフィックが全サーバー間に流れる場合)	40
16	階層内の配置構成の消費電力量による比較(特定サーバー間トラフィックの場合)	41
17	階層内の配置構成の遅延による比較(特定サーバー間トラフィックの場合)	42

表 目 次

1	階層間の接続構成の消費電力量による比較 (平均 500 bit、標準偏差 20 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)	23
2	階層間の接続構成の消費電力量による比較 (平均 500 bit、標準偏差 100 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)	23
3	階層間の接続構成の消費電力量による比較 (特定サーバー間トラフィックの場合)	23
4	階層間の接続構成の遅延による比較 (特定サーバー間トラフィックの場合) . . .	24
5	サーバーとスイッチの接続階層の消費電力量による比較 (平均 500 bit、標準偏差 20 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)	30
6	サーバーとスイッチの接続階層の消費電力量による比較 (平均 500 bit、標準偏差 100 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)	30
7	サーバーとスイッチの接続階層の消費電力量による比較 (特定サーバー間トラフィックの場合)	31
8	サーバーとスイッチの接続階層の遅延による比較 (特定サーバー間トラフィックの場合)	31
9	階層内の配置構成の違いによる消費電力量による比較 (平均 500 bit、標準偏差 20 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)	37
10	階層内の配置構成の違いによる消費電力量による比較 (平均 500 bit、標準偏差 100 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)	37
11	階層内の配置構成の消費電力量による比較 (特定サーバー間トラフィックの場合)	38
12	階層内の配置構成の遅延による比較 (特定サーバー間トラフィックの場合) . . .	38

1 はじめに

近年、クラウドコンピューティングやスマートフォン利用者の増加により、多量のデータがネットワーク上を流れ、それらのデータの蓄積・処理を行うデータセンターが重要となっている。データセンターによる処理が必要なデータ量は年々増加しており、多量のデータを処理するために、数万台のサーバーを収容する大規模なデータセンターの構築が進められている。データセンターが大規模になるにつれ、データセンター内の機器数が増え、データセンター内の消費電力が増大している。その結果、データセンターが消費する電力は、世界で消費される全電力消費の 1.1% ~1.5% を占めるようになっており [1]、ネットワーク上を流れるデータ量が増え、データセンターの大規模化が進むにつれ、この割合は更に増加すると予測されている。

データセンターでは、多量のデータを処理するために、サーバーは互いに連携してデータの処理を行っている。そのため、サーバー間を接続するネットワークがデータセンターの性能に大きな影響を与える。そのため、従来のデータセンターでは、高速なスイッチの導入や、多数の機器を用いることにより、サーバー間に十分な帯域を持つ通信を確保していた。しかしながら、データセンターの大規模化に伴い、データセンターネットワークの規模も大きくなり、その消費電力の増大が問題となっている。また、サーバーの低消費電力化に向けた検討が進んできており [2]、データセンター全体の消費電力に占めるネットワーク機器の割合はさらに大きくなると予想される。そのため、サーバー間を接続するネットワークの性能を維持しつつ、消費電力が少ないネットワークを構成する方法が検討されている [3, 4]。文献 [3] では、ネットワークを流れるトラフィック量に応じて、ネットワーク機器の電源を落とすことにより、ネットワークの消費電力を削減する手法が提案されている。また、文献 [4] においても、光回線交換スイッチと電気パケットスイッチによってデータセンター内ネットワークを構築し、より多くの電気パケットスイッチのポートの電源を落とすことができるように、トラフィック量に応じて光回線交換スイッチを設定する手法が提案されている。しかしながら、これらの手法では、トラフィック量が多い場合は、全ネットワーク機器の電源を投入する必要があり、消費電力を削減することができない。

低消費電力なデータセンターネットワークを構築する方法として、オンチップ型データセンターネットワークの検討が進められている [5]。オンチップ型データセンターネットワークでは、多数の CPU コアと、CPU コア間を結ぶネットワークを 1 チップに集約する。チップ上に構築されたネットワークは、スイッチ間の配線距離が 1mm 程度以下と著しく小さいため、小さい電力で通信が可能となる [6]。そのため、オンチップ型データセンターネットワークでは、従来のデータセンターネットワークと比べ、ネットワークの消費電力を著しく小さく抑えることができ、データセンター向けのアプリケーションを動作させる事ができる。し

かしながら、文献 [5] ではデータセンターをチップ化するコンセプトの提案のみで、具体的なネットワーク構造については検討されていない。

チップ上のネットワークは Network on Chip (NoC) と呼ばれ、チップ上にパケットスイッチや回線交換スイッチを配置することにより構成される。パケット交換スイッチを配置したネットワークでは、チップ内のコア間の通信はパケット単位あるいは、パケットをより細かな単位に分割したフリット単位で行われ、各パケットスイッチは宛先に応じて各パケット/フリットを隣接する転送先に送出する。回線交換スイッチを用いたネットワークでは、通信開始前に、通信を行うコア間を接続するように途中のスイッチを設定したのちに、通信を行う。回線交換スイッチでは、通信開始前に入力ポートと出力ポートの接続構成を決めるため、パケット交換スイッチのようなパケット単位で出口ポートを変化させることができない。しかしながら、回線交換スイッチでは、パケット交換スイッチのようなパケットごとの処理は行わないため、スイッチの消費電力・遅延ともに小さいという特徴がある。そのため、近年ではパケットスイッチと回線交換スイッチの両方のスイッチを効率的に用いた構造についても検討が進められている [6]。この構造では、回線交換スイッチを設定することにより、パケットスイッチ間の接続構成を変化させることができる。そのため、チップのアプリケーションや通信需要に応じて適切なネットワーク構造を構築することが可能となる。

また、3次元型の NoC を構築する手法についても検討が進められている [7]。3次元 NoC では、コアを立体的に配置し、平面上のみではなく、上下方向にもリンクを持つように立体的なネットワークを構成する。この構成では、平面上に全コアを配置した場合と比べ、多数のコアをより小さな面積で構築できる。そのため、コア間を結ぶリンクの長さも小さくすることができ、消費電力を削減できる。

3次元 NoC や回線交換スイッチとパケット交換スイッチを混在させた NoC を構成する技術を用いる事で、省電力で遅延の少ないオンチップ型データセンターを構築できると考えられる。しかしながら、これらの技術を用いたオンチップ型データセンターに適したネットワーク構造に関する検討は行われていない。

そこで、本報告では、3次元のネットワーク構造において、パケット交換スイッチのみではなく、回線交換スイッチも利用した、3次元オンチップ型ネットワークの検討を行う。本報告では、3次元オンチップ型データセンターネットワークの構成を、階層間の接続方法、コア間・スイッチ間の接続階層、階層内の接続構成に分けて検討を行う。そして、各項目に対して同じトラヒックを与え、消費電力と遅延を測定する事により適切な3次元オンチップ型データセンターネットワークを導く。

本報告の構成は、まず2章で、データセンターネットワークと NoC に関する関連研究について述べる。次に3章で、提案するオンチップ型データセンターネットワークについての説明と、評価結果について述べる。最後に、本報告のまとめと今後の課題について述べる。

2 関連研究

2.1 データセンターネットワーク

データセンターでは多くのサーバーが連絡してデータの処理を行なっている。そのため、サーバー間のネットワーク性能がデータセンターの処理性能に大きな影響を与える。サーバー間の通信遅延は、サーバーの連携の際に必要なデータを取得する際にかかる時間に影響を与える。そのため、アプリケーションが許容できる範囲内にサーバー間の通信遅延を抑える必要がある。

そのため、サーバー間の通信性能を確保することができるネットワーク構造に関する検討が進められてきた。文献 [8] では、ポート数の小さなスイッチのみを用いて、多数のサーバー間に十分な帯域を確保する Fatree と呼ばれるネットワーク構造が提案されている。FatTree は、ポート数が少ないスイッチを組み合わせることで POD と呼ばれる構造を構築し、POD の各ポートを最上位層に配置されたスイッチに接続することにより構築される。FatTree では、各スイッチは下位のスイッチからのリンク数と同じ本数の上位のスイッチへのリンクを持つ。そのため、いずれのスイッチもボトルネックになることなく、サーバー間に十分な帯域が確保可能となる。

しかしながら、データセンターが大規模化するにつれ、データセンターネットワークの消費電力も大きくなっており、サーバー間の通信性能の確保だけでなく、消費電力の削減も、データセンターネットワークにおける大きな課題となっている。現在、ネットワークが消費する電力は、データセンターの消費電力の 10 から 15% を占めている [9]。さらに、近年はサーバーの低消費電力化技術の開発が進んでおり、ネットワークの消費電力が占める割合も増加すると考えられる。

そこで、データセンターネットワークの消費電力を削減する技術の開発が進んでいる。文献 [3] では、サーバー間の通信需要に応じて、スイッチのポートの電源を落とすことにより、ネットワークの消費電力を削減する手法が提案されている。また、文献 [4] では、サーバー間の通信需要に応じて、低消費電力なネットワークを動的に構築する手法が提案されている。しかしながら、これらの手法では、サーバー間の通信需要が大きいと消費電力を削減することができない。そのため、サーバー間の通信需要が大きい場合にも、低消費電力で十分な帯域の通信をサーバー間に確保する手法が必要とされている。

2.2 Network on Chip

近年、半導体素子技術の進歩により、1つのチップ上に実現できる回路の規模が大きくなってきており、複数のプロセッサやメモリなどを1つのチップ上に搭載する事が可能となって

いる [10]。チップ上に搭載された複数のプロセッサやメモリは連携して動作する。そのため、プロセッサやメモリ等のチップ上のコア間での通信が必要となる。コア数が少ない場合は、バスを用いてコア間の通信路を確保することが可能である。しかしながら、バスは通信の衝突が問題となるため、コア数が多い環境に適用するのは困難である。

そこで、チップ上に回線交換スイッチやパケットスイッチを配置したネットワークを構築し、そのネットワークを通じてコア間の通信を行うネットワークオンチップ (NoC) に関する研究が進められている [11]。NoC では、チップ上にパケット交換スイッチまたは回線交換スイッチを配置することにより、ネットワークを構成する。パケット交換スイッチを配置したネットワークでは、チップ内のコア間の通信はパケット単位あるいは、パケットをより細かな単位に分割したフリット単位で行われ、パケットスイッチにおいて宛先に応じて各パケット/フリットの転送処理を行う。回線交換スイッチを用いたネットワークでは、通信開始前に、通信を行うコア間を接続するように途中のスイッチを設定したのちに、通信を行う。回線交換スイッチでは、通信開始前に入力ポートと出力ポートの接続構成を決めるため、パケット交換スイッチのようなパケット単位で各スイッチの出口ポートを変化させることができない。しかしながら、回線交換スイッチは、パケット交換スイッチのようなパケットごとの処理を行わないため、スイッチの消費電力・遅延ともに小さいという特徴がある。

そのため、近年、回線交換スイッチとパケット交換スイッチの両方を用いたネットワーク構造の検討が進められている [6, 12, 13]。この構造では、回線交換スイッチを設定することにより、パケット交換スイッチ間の接続構成を変化させることができる。つまり、アプリケーションの要求に応じて、回線交換スイッチを設定することにより、適切なパケット交換スイッチ間の接続構成を構築できる。

また、近年、チップを積層することにより、平面上のみではなく、上下方向にもスイッチを配置した 3 次元 NoC に関する検討も進められている [7]。この構成では、平面上に全コアを配置した場合と比べ、多数のコアをより小さな面積で構築できる。そのため、コア間を結ぶリンクの長さも短くすることができ、消費電力を削減できる。

NoC の技術の応用として、データセンターを 1 チップに集約したオンチップ型データセンターが提案されている [14]。オンチップ型データセンターでは、データセンター内の各サーバーに該当する機能を各コアが担当し、多数のコアが連携して動作することにより、データセンターと同様の処理を行う。データセンターを 1 チップに集約することにより、コア間の連携動作の際に発生する通信が経由するリンクの距離を著しく短くすることができ、低消費電力化が期待できる。しかしながら、オンチップ型データセンターはコンセプトの提案や、キャッシュの配置に関する検討が行われているのみであり、オンチップ型データセンターに適したネットワーク構造に関する検討は行われていない。

本報告では、パケット交換スイッチと回線交換スイッチの両スイッチを効率的に利用する、

3次元のネットワーク構造を構築するといった、上述の NoC 技術を有効に用い、オンチップ型データセンターに適したネットワーク構造の検討を行う。

3 提案するオンチップ型データセンターネットワークの検討と評価

3.1 オンチップ型データセンターネットワークの概要

データセンターは、CPU、RAM、外部記憶装置を持つサーバーと、サーバー間を接続するネットワークから構成される。このうち、CPUについては、近年研究・開発が進められているメニーコアCPUのように、各CPUをCPUコアとして搭載可能である。それに対して、RAMや外部記憶装置もチップ内に構成することはできるものの、大きな記憶容量を確保するためには、回路面積を大きくする必要がある。

そこで、本報告では、図1のように、データセンターにおけるCPU・CPU内のキャッシュと、サーバー間ネットワークをチップ上に集約し、RAMや外部記憶装置はチップ外に配置した上で、チップ上のコアからのインターフェースを通じて接続するものとする。

本報告で想定するオンチップ型データセンターネットワークを図2に示す。図2に示されるように、本報告で検討するデータセンターネットワークでは3次元のネットワーク構成を用いる。この構成では、異なる階層の同一位置に、CPUの機能を果たすコア、キャッシュの機能を果たすコアを配置し、それらを直接接続する。この直接接続されたコアとそのコアにインターフェースを通じて接続されたRAMや外部記憶装置が1台のサーバーとして機能する。このように1台のサーバーとして機能するコア同士をチップ上に並べることで、データセンターを構築する。

各サーバー間の通信は、各階層に配置されたパケット交換スイッチ、回線交換スイッチで構成されたネットワークを用いて行われる。本報告では、このオンチップ型データセンターに適したパケット交換スイッチ・回線交換スイッチのネットワーク構造について消費電力と遅延の観点から検討を行う。

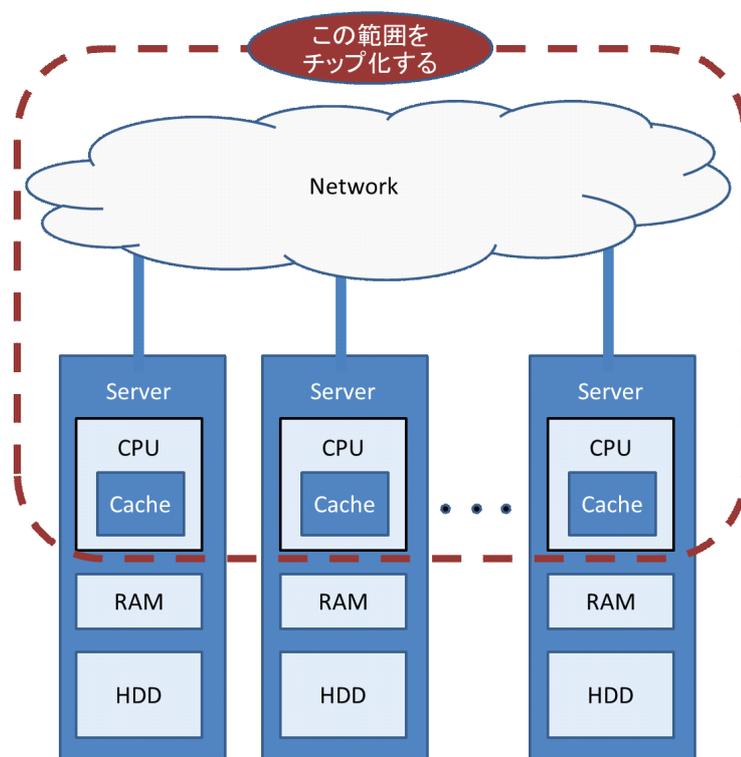


図 1: オンチップ型データセンターにおけるチップ化対象

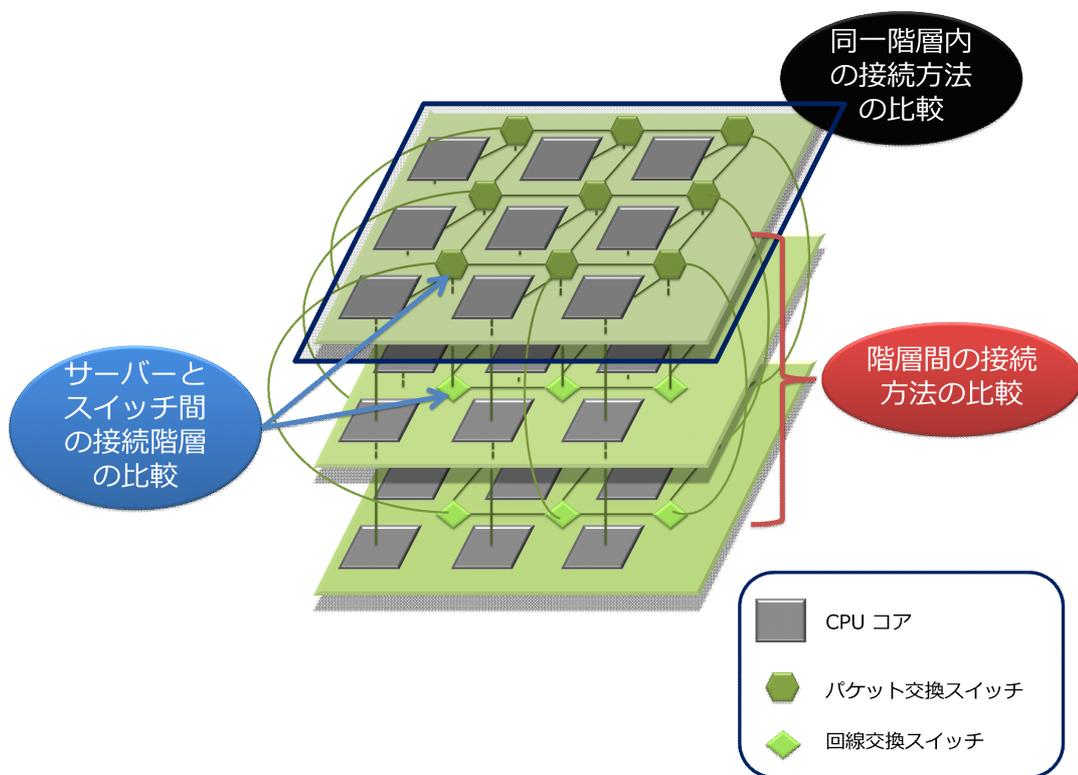


図 2: データセンター向け 3次元オンチップ型ネットワーク

3.2 検討に用いるモデル

3.2.1 電力消費モデル

チップ上に構成されたネットワークが消費する電力は、ネットワークの構造、流れるトラフィック量と、そのトラフィックに含まれるビット反転の数によって決まる。

文献 [6] では、16 bit 幅のリンクを用いて構成した 2 次元格子型の NoC において、5 割のビットが反転するトラフィックが流れた際に、回線交換スイッチ、パケット交換スイッチ、スイッチ間のリンクで消費される電力を測定し、モデル化を行っている。その結果、経由するトラフィック 1 bit あたりの消費電力は、回線交換スイッチで $0.37 \mu\text{W}$ 、パケット交換スイッチで $0.98 \mu\text{W}$ 、リンクで $(0.39 + 0.12L) \mu\text{W}$ (ただし L はリンクの長さ (mm)) であることが明らかになっている。本報告では、このモデルを用いることにより、ネットワークの消費電力を評価する。

回線交換スイッチの消費電力は、パケット交換スイッチの消費電力よりも少ない。そのため、回線交換スイッチを効率的に利用し、パケット交換スイッチを経由するトラフィック量を削減することにより、消費電力の削減が可能となる。本報告では、ネットワークの消費電力に注目するため、各コアの消費電力を含まない、各リンク、回線交換スイッチ、パケット交換スイッチで消費される電力の合計の比較を行う。

チップ上の各スイッチの消費電力やリンクの消費電力は、用いるスイッチの構成やスイッチ間のリンクの帯域によって異なる。しかしながら、どのような構成を採用した場合であっても、パケット交換スイッチは回線交換スイッチよりも複雑な処理が必要となり、その消費電力は大きくなる。そのため、各スイッチやリンクの消費電力が異なる他の構成においても、本報告の検討・考察は適用可能である。

3.2.2 遅延モデル

本報告では、各サーバー間に同時に発生するトラフィック要求を生成し、その要求されたトラフィックをすべて転送し終えるまでにかかる時間を計算することにより、ネットワーク内で発生する遅延の比較を行う。

本報告では、サーバー間に発生するトラフィックは、パケット単位に分割された後に転送されるものとする。また、各パケット交換スイッチでは、バーチャルカットスルーと呼ばれる方式を採用するものとする。バーチャルカットスルーでは、パケットはフリットと呼ばれる小さな単位に分割される。各パケットスイッチでは、フリット単位で転送を行うことにより、パケット全体が到着する前に隣接パケットスイッチへの転送を行うことができ、遅延を削減することができる。また、バーチャルカットスルーでは、各パケットスイッチはバッファを

持ち、パケットの衝突が発生した場合には、当該パケットのフリットが送出可能となるまで、バッファでパケットを保持する。

本報告で検討するネットワークは、パケット交換スイッチのみではなく、回線交換スイッチも用いている。回線交換スイッチは、事前に設定することにより、その出入口ポート間を接続する。そのため、回線交換スイッチの設定により接続されたパケット交換スイッチ間では、隣接するパケット交換スイッチへのフリットの転送と同様のフリットの転送が可能となる。そのため、回線交換スイッチを効率的に配置することにより、サーバー間の遅延を小さくすることが可能となる。

本報告では、パケット交換スイッチから隣接するパケット交換スイッチへ、あるいは、回線交換スイッチの設定により隣接関係となったパケット交換スイッチへの1フリットの転送に1クロックかかるものとして、同時に発生したトラヒック要求を転送し終わるまでのクロック数を計算し、比較を行う。パケット交換スイッチが1フリット転送する際にかかるクロック数は、パケット交換スイッチの構成により異なる。しかしながら、1フリット転送する際にかかるクロック数が変化した場合であっても、遅延の観点で評価したネットワーク構造の優劣は変化しない。そのため、本報告での議論は、異なるパケット交換スイッチを用いたネットワークにおいても適用可能である。

3.2.3 トラヒックモデル

本報告では、データセンターのサーバー間の通信を模したトラヒックを用いてネットワーク構造の評価を行う。データセンターでは、サーバー間の連携により多量のデータの処理が行われる。しかしながら、データセンターのネットワークで発生するトラヒックは、データの処理を行うアプリケーションによって異なる。本報告では、データセンターを模したトラヒックとして、以下の2種類のトラヒックを生成する。

全サーバー間通信: 全サーバー間でトラヒックが発生する。これは、全サーバーが連携して一つのデータの処理を行っている場合に相当する。

特定サーバー間通信: 特定のサーバー間でのみトラヒックが発生する。これは、データセンター内で複数の処理を並列に動作している場合に相当し、各データの処理に関係した特定のサーバー間のみで通信が発生する。

本報告では、全サーバー間通信として、平均 500 bit 標準偏差 20 bit のトラヒック需要、平均 500 bit 標準偏差 100 bit のトラヒック需要の2種類のトラヒック需要を生成した。また、特定サーバー間通信では、通信を行うサーバーをランダムに選択し、選択されたサー

バー間で 1000 bit のトラフィック需要が発生するものとし、通信を行うサーバーペアの数を変化させて評価を行った

発生させたトラフィックは、消費電力、遅延の両方に影響を与える。ただし、消費電力は各スイッチを経由するトラフィック量に比例するため、サーバー間のトラフィック量がより大きなものとなった場合であっても、本報告におけるネットワーク構造の消費電力に関する議論は適用可能である。同様に、遅延に関する議論も、サーバー間のトラフィック量がより大きなものとなった場合であっても適用可能である。

3.2.4 経路計算モデル

本報告では、消費電力を基準として定めたコア間のトラフィックの経路を用いて検討・評価を行う。

3.2.1 の電力消費モデルより、各スイッチ、リンクが消費する電力は、経由するトラフィック量に依存する。そのため、トラフィック需要の多いサーバーペア間の通信が消費電力に与える影響は大きく 1 bit あたりの消費電力がより少ない電力効率のよい経路に收容することが必要とされる。

そこで、本報告では、サーバーペアに発生している通信需要が多いものから順に経路を確定する。3.2.1 のモデルでは、各スイッチ・各リンクが使用する消費電力は、そのスイッチ・リンクを流れるトラフィック量に比例するため、各サーバーペアに対する低消費電力な経路は、各スイッチ・リンクの 1 bit あたりの消費電力を重みとしたグラフ上でダイクストラ法を用いて得ることができる。ただし、経路計算の際には、すでに設定済みの回線交換スイッチの出入口ポートは、接続されているものとして扱う。

上述の経路計算モデルでは、入力としてサーバーペア間のトラフィック需要が必要となり、集中的な経路計算が必要となる。本報告における適切なネットワーク構造の比較・検討では、効率的な経路計算が可能であるという条件のもとで、低消費電力・低遅延でサーバー間トラフィックが收容可能なネットワーク構造を明らかにすることを目的としている。しかしながら、実際のチップ上のネットワークでは、集中的な経路計算やサーバーペア間のトラフィック需要を事前に知ることは困難である。集中的な経路計算が不要で、経路計算時にトラフィック需要が不明でも動作可能な経路計算手法は今後の課題である。

3.2.5 ネットワークモデル

本報告では、3次元格子状に回線交換スイッチ、パケット交換スイッチを配置することにより、ネットワークを構成するものとする。各サーバーからは、そのサーバーを構成するいずれかのコアに直近のスイッチに接続するリンクを構築することで、ネットワークへの接続

を確保する。ただし、各サーバーは同時に複数のサーバーとの通信を行う可能性があるため、各サーバーと接続するスイッチは必ずパケット交換スイッチであるものとする。

3次元格子の上下リンクの長さは、文献[7]で構築されたネットワークと同じ、1 μm とし、同一階層内の各スイッチ間の距離は、文献[12]のネットワーク構造と同じ値である2 mmと、それよりも短い場合と長い場合である1 mmの場合、3 mmの場合について検討を行う。リンク長はリンクの消費電力に影響を与えるため、リンク長が異なるネットワークを評価することにより、リンク長が適切なネットワーク構造に与える影響を検証することが可能となる。

3.3 検討項目

本報告では、図2に示した3次元ネットワークの構成を検討する。本報告では、このネットワーク構造を、異なる階層に配置されたスイッチの接続構成、サーバーとスイッチの接続構成、階層内のスイッチの接続構成の3つの項目に分けて検討を行う。以降、各検討項目の詳細について述べる。

3.3.1 階層間の接続

階層間の接続構成の候補を図3に示す。異なる階層に属するスイッチの接続方法としては、(1) 隣接階層のスイッチ同士を接続する（隣接階層接続型）、(2) 全階層の回線交換スイッチをすべて一台のパケット交換スイッチと接続する（パケットスイッチ集約型）、(3) 隣接階層のスイッチ同士を接続した上で、全階層の回線交換スイッチと一台のパケット交換スイッチを接続する（パケットスイッチ集約+隣接階層接続型）の3種類が考えられる。

隣接階層接続型は、ポート数の多いスイッチは不要であるものの、離れた階層のスイッチを経由するためには、間にある階層のスイッチを経由する必要がある。それに対して、パケットスイッチ集約型は、全階層と接続しているパケット交換スイッチが存在する。サーバーと直接接続しているパケットスイッチを、全階層との接続を持つスイッチとすることで、各サーバーからはいずれの階層のスイッチも2ホップで到着可能であり、いずれの階層も有効に利用した通信が可能となる。パケットスイッチ集約+隣接階層接続型は隣接階層接続型のリンクとパケットスイッチ集約型のリンクを併せ持った構造であり、各サーバー間に多数の経路を確保することが可能となる。しかしながら、隣接階層接続型やパケットスイッチ集約型の2倍の階層間のリンクが必要となる。

本報告では、上記の3種類のネットワーク構造と、2次元格子状にパケット交換スイッチを配置したネットワーク構造(2次元格子型)の比較を行うことにより、適切な階層間の接続

構成を明らかにする。本報告では、第一階層はサーバーと直接接続したパケットスイッチのみを配置し、第二階層以降には、回線交換スイッチを配置するものとする。

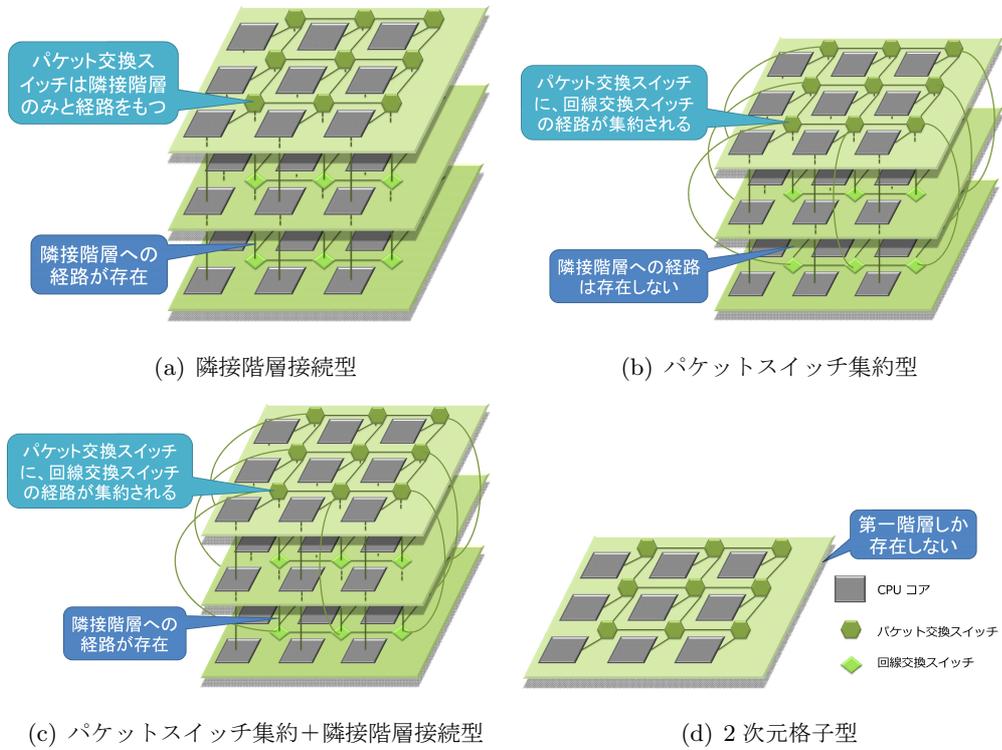


図 3: 検討に用いる階層間の接続構成

3.3.2 サーバーとスイッチの接続階層

本報告では、各サーバーはいずれかの階層において、パケット交換スイッチと接続することにより、ネットワークに接続する。そのため、図4に示されるように、全サーバーが同一階層のスイッチと接続することによりネットワークに接続する構造（接続階層集約型）、サーバーによって接続しているスイッチの階層が異なる構造（接続階層分散型）の2通りのサーバーをネットワークに接続する方法が考えられる。接続階層集約型は、全サーバーが同一階層のスイッチと接続しているため、サーバー間のホップ数を小さくすることができる。それに対して、接続階層分散型は、ホップ数は大きくなるものの、サーバー間に回線交換スイッチを経由した複数の経路を確保できるため、低消費電力な経路が見つかる確率が高くなると考えられる。

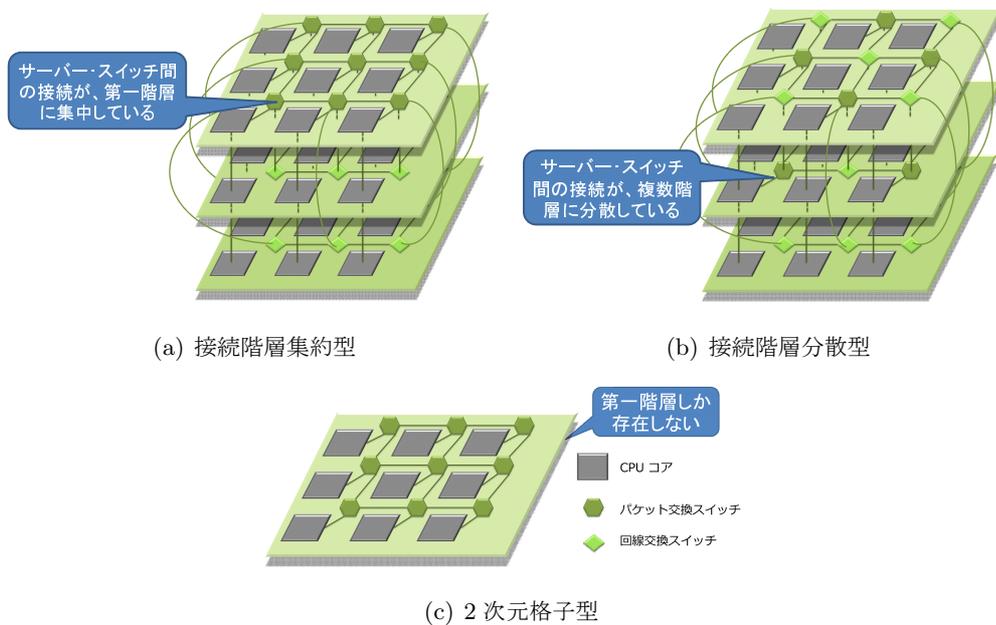


図4: 検討に用いるサーバーとスイッチの接続構成

3.3.3 階層内の配置

階層内の配置構成としては、図5に示されるように、パケット交換スイッチのみを配置した層を構成する（単一種類階層構成型）、パケット交換スイッチと回線交換スイッチを混在させる（スイッチ混在階層構成型）という2つの構成が考えられる。スイッチ混在階層構成型では、単一種類階層構成型と比べ、サーバー間の経路数が多くなるため、低消費電力な経路を発見できる可能性がある。ただし、サーバー間のホップ数や配線長は増加する。

本検討では、単一種類階層構成型とスイッチ混在階層構成型の比較を行うことで、パケット交換スイッチ同士を接続した階層があったほうがよいのか、それとも、回線交換スイッチを用いてパケットスイッチ間の経路にも冗長性を持たせた方がよいのかを明らかにする。

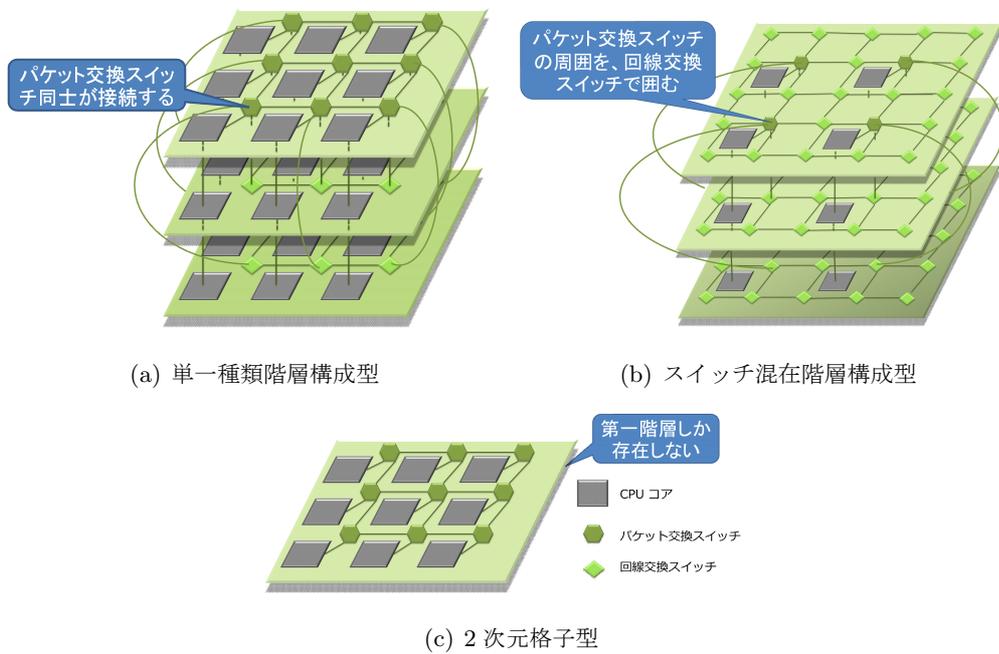


図 5: 検討に用いる階層内の配置構成

3.4 検討結果と考察

3.4.1 階層間の接続の検討

本小節では、階層間の接続構成が異なるネットワーク構成を消費電力の観点から比較を行う。表1、表2にサーバー数25、100、225台、リンク長1mm、2mm、3mmの各ネットワーク構成において全サーバー間通信を与えた場合の消費電力を示す。表3にサーバー数225台、リンク長2mmのネットワーク構成において特定サーバー間通信を与えた場合の消費電力を示す。また、全サーバー間通信の場合の消費電力の比較結果を図示したグラフを図6と図7、特定サーバー間通信の場合の消費電力の比較を図示したグラフを図8に示す。これらの図、表では、2次元格子型の場合の消費電力を100として正規化した消費電力を示す。

表1、2より、全サーバー間通信の場合、トラヒックの偏りによらず、いずれのネットワーク構造であっても、2次元格子型から削減できる電力量は1割未満である。しかしながら、特定サーバー間トラヒックの場合は、2次元格子型と比べ、より多くのトラヒック量が削減可能であり、通信を発生させたサーバーペアの数が少なければ少ないほど、削減できる消費電力は大きい。これは、通信を発生させたサーバーペアの数が少ない場合は、回線交換スイッチのみを経由した径路が見つかりやすく、経由するトラヒック1bitあたりの消費電力が大きいパケット交換スイッチを経由するトラヒック量を大きく削減できるためだと考えられる。それに対して、通信が発生するサーバーペアの数が多い場合は、すべてのサーバーペアに対して、回線交換スイッチのみを用いた経路を見つけることができず、パケット交換スイッチを経由するトラヒック量も増加してしまう。その結果、消費する電力は、2次元格子型と同程度となる。

また図6、図7より、同一階層内のリンク長が長くなるほど、2次元格子型から削減できる消費電力が少なくなっていることが分かる。これは、リンク長が長いほど、スイッチ間のリンクで消費される電力が大きくなるためである。各リンクが消費する電力が小さい場合は、各サーバーペア間のトラヒックが経由するリンクの本数を多くしてでも、経由するパケット交換スイッチを削減した方が、より消費電力を削減でき、本検討においても、そのように経由するパケット交換スイッチが少なくなるような経路が用いられている。しかしながら、リンク長が長くなり、各リンクが消費する電力が大きくなると、消費電力を小さくするためには、経由するリンク数を多くすることはできず、多くのサーバーペア間で、2次元格子型ネットワークと同様の、多数のパケットスイッチを経由した径路が選択される。

ネットワーク構成の比較を行うと、隣接階層接続型よりもパケットスイッチ集約型の方が消費電力が少ないことが分かる。これは、隣接階層接続型では、第三階層の回線交換スイッチを利用するには、第二階層の回線交換スイッチも経由する必要があるためである。そのため、第二階層の回線交換スイッチが設定によっては、第三階層の回線交換スイッチにたどり

着くことができず、第三階層以降の回線交換スイッチを効率的に利用することができない。それに対して、パケットスイッチ集約型では、各パケット交換スイッチから全階層に接続しているため、全階層の回線交換スイッチを利用することができ、より低消費電力の経路を見つけることが可能となる。

また図6～8より、パケットスイッチ集約型とパケットスイッチ集約+隣接階層接続型の消費電力は同程度となっていることがわかる。これは、いずれの階層も、パケット交換スイッチから当該階層へ直接接続されたリンクを経由することで到達可能であるためである、そのため、隣接階層間のリンクは消費電力を削減可能な経路の発見に対して限定的な効果しかない。つまり、隣接階層間のリンクは不要であると考えられる。

表4に、特定サーバー間通信を発生させた際の遅延を示す。また、この結果を図示したものを図9に示す。いずれの図表も、2次元格子型の場合の遅延を、通信を発生させたサーバーペアの数が少なければ4割以上、サーバーペアの数が多い場合でも、パケットスイッチ集約型は5割以上遅延を削減する事ができる。図より、低消費電力化を目標とした経路制御を行った場合であっても、2次元格子型と比べて、遅延を著しく削減できることが分かる。これは、回線交換スイッチを用いることにより、各サーバーペア間の通信が経由するパケット交換スイッチの数を減らすことができているためだと考えられる。その結果、より少ないクロック数で、宛先サーバーまで到達可能となる。

また、ネットワーク構造間の比較を行うと、パケットスイッチ集約型のリンクを持つ構造が、隣接階層接続型よりも遅延を抑えることができている。これは、隣接階層接続型では、パケット交換スイッチから離れた階層の回線交換スイッチを利用することができないのに対して、パケットスイッチ集約型では、いずれの階層の回線交換スイッチも利用した径路を用いることができる。その結果、パケットスイッチ集約型は、経由するパケット交換スイッチの数をより多く削減でき、遅延を削減可能である。

以上より、階層間の接続においては、特定の階層のスイッチと、他の全階層のスイッチが接続する構成が最も良いと考えられる。

表 1: 階層間の接続構成の消費電力量による比較 (平均 500 bit、標準偏差 20 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)

	1mm			2mm			3mm		
	5×5	10×10	15×15	5×5	10×10	15×15	5×5	10×10	15×15
2次元格子型	100	100	100	100	100	100	100	100	100
隣接階層接続型	98.0	96.6	94.3	98.4	97.5	95.1	98.7	97.9	95.3
パケットスイッチ集約型	97.6	93.7	90.4	97.8	94.2	91.4	98.0	94.6	91.9
パケットスイッチ集約 +隣接階層接続型	97.6	93.7	90.8	97.8	94.4	91.6	98.0	94.8	92.3

表 2: 階層間の接続構成の消費電力量による比較 (平均 500 bit、標準偏差 100 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)

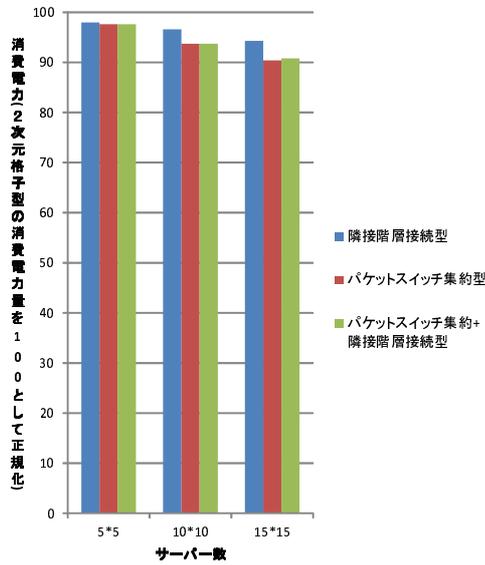
	1mm			2mm			3mm		
	5×5	10×10	15×15	5×5	10×10	15×15	5×5	10×10	15×15
2次元格子型	100	100	100	100	100	100	100	100	100
隣接階層接続型	98.0	95.8	94.3	98.2	96.0	94.9	98.5	96.3	95.3
パケットスイッチ集約型	97.6	93.6	90.4	97.8	94.1	91.3	98.0	94.5	91.9
パケットスイッチ集約 +隣接階層接続型	97.6	93.7	90.8	97.8	94.1	91.6	98.0	94.6	92.3

表 3: 階層間の接続構成の消費電力量による比較 (特定サーバー間トラフィックの場合)

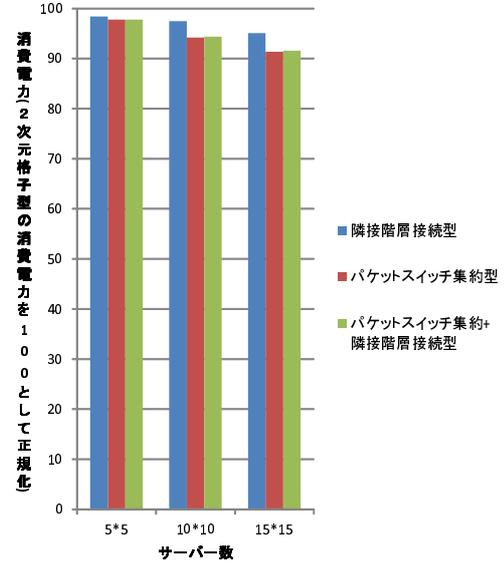
通信サーバーペア数	10	100	500	1000
2次元格子型	100	100	100	100
隣接階層接続型	76.1	84.0	91.6	93.2
パケットスイッチ集約型	76.1	78.1	86.0	88.6
パケットスイッチ集約+隣接階層接続型	76.1	78.2	86.0	88.6

表 4: 階層間の接続構成の遅延による比較 (特定サーバー間トラヒックの場合)

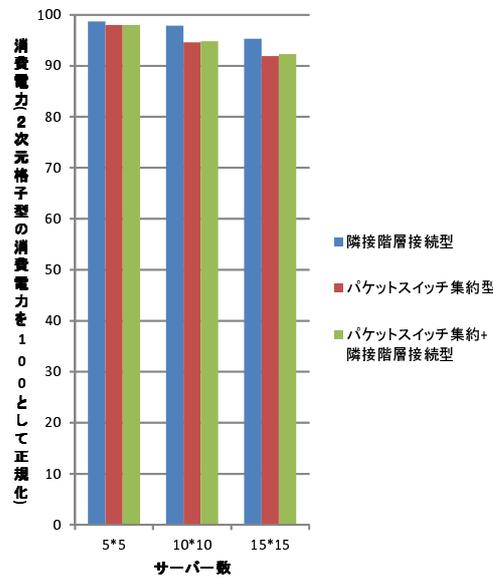
通信サーバーペア数	100	500	1000	1500
2次元格子型	100	100	100	100
隣接階層接続型	59.7	52.8	56.9	67.2
パケットスイッチ集約型	57.8	48.1	45.7	45.5
パケットスイッチ集約+隣接階層接続型	57.8	49.1	48.4	58.4



(a) リンク長 1mm

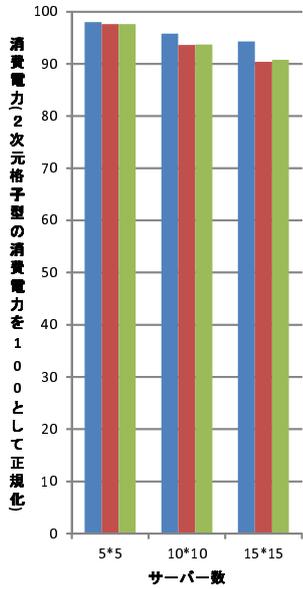


(b) リンク長 2mm

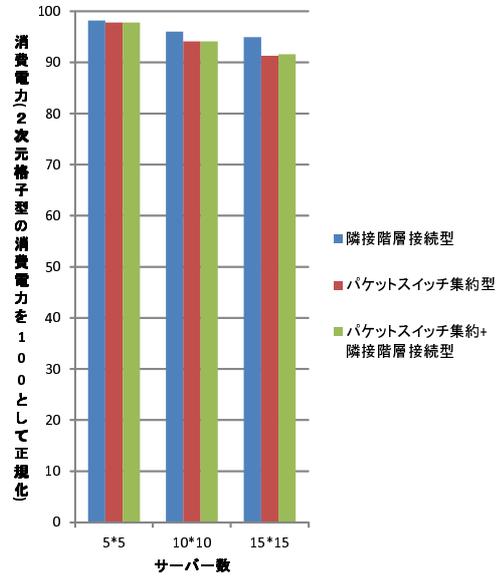


(c) リンク長 3mm

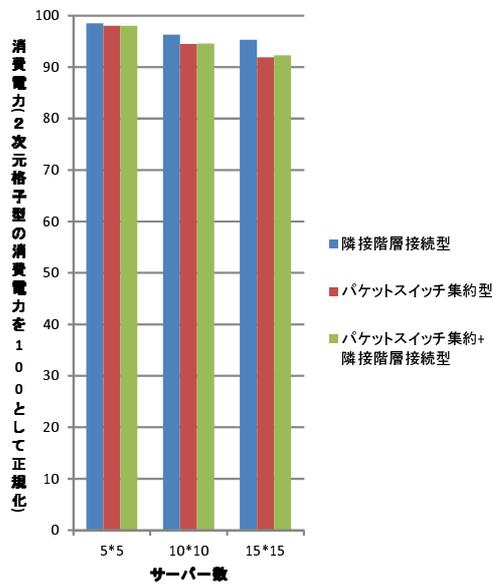
図 6: 階層間の接続構成の消費電力量による比較 (平均 500 bit、標準偏差 20 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)



(a) リンク長 1mm



(b) リンク長 2mm



(c) リンク長 3mm

図 7: 階層間の接続構成の消費電力量による比較 (平均 500 bit、標準偏差 100 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)

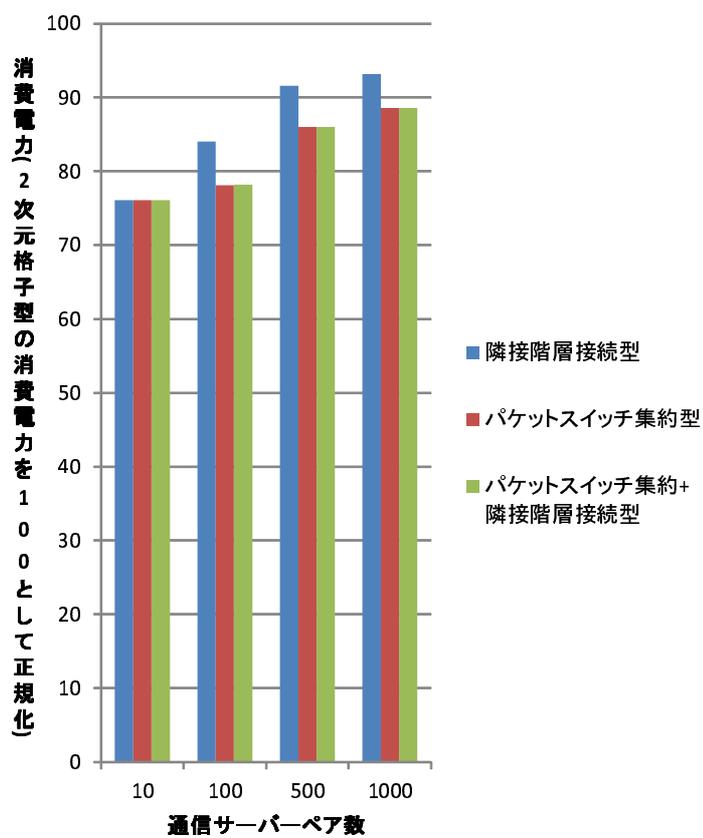


図 8: 階層間の接続構成の消費電力量による比較 (特定サーバー間トラヒックの場合)

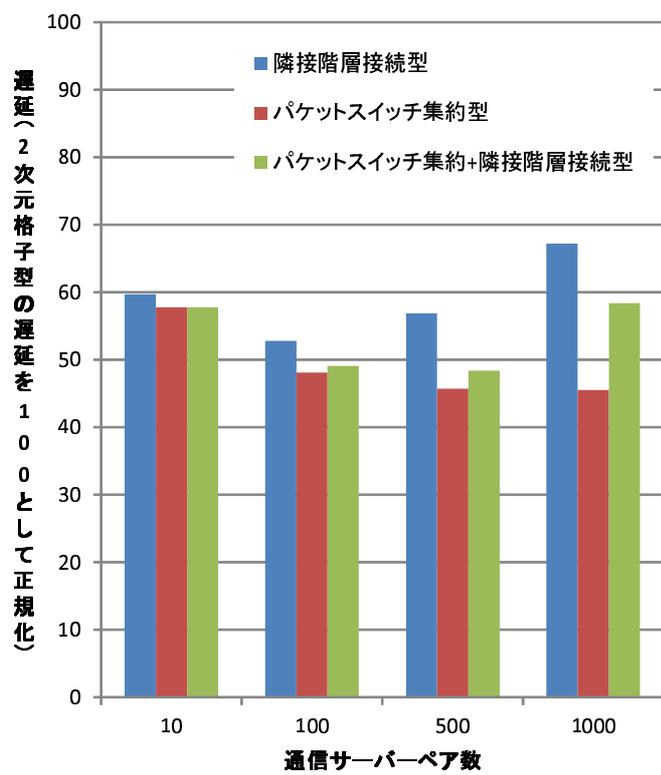


図 9: 階層間の接続構成の遅延による比較 (特定サーバー間トラヒックの場合)

3.4.2 サーバーとスイッチの接続階層の検討

本小節では、サーバーが接続しているスイッチの属する階層による消費電力・遅延の違いを評価する。3.4.1 節と同様、2次元格子型の場合の消費電力を100として消費電力の比較を行う。表5、表6にサーバー数25、100、225台、リンク長1mm、2mm、3mmの各ネットワーク構成において全サーバー間通信を与えた場合の消費電力、表7にサーバー数225台、リンク長2mmのネットワーク構成において特定サーバー間通信を与えた場合の消費電力を示す。また、全サーバー間通信の場合の消費電力の比較結果を図示したグラフを図10、図11、特定サーバー間通信の場合の消費電力の比較を図示したグラフを図12に示す。

図より、接続階層集約型は常に2次元格子型よりも低い消費電力を達成できるのに対して、接続階層分散型は2次元格子型よりも消費電力が大きくなる場合があることが分かる。これは、接続階層集約型では第一層に2次元格子型と同じ構造を持っているのに対して、接続階層分散型ではいずれの階層にも2次元格子型と同じ構造は存在しないためである。そのため、接続階層集約型は、回線交換スイッチの設定状況により、効率的な経路が見つからない場合であっても、第一層を経由することにより、2次元格子型と同じ消費電力で通信を取用することができるのに対して、接続階層分散型は、ホップ数が大きく消費電力が大きい経路を選択せざるを得ない。

表8に、特定サーバー間通信を発生させた際の遅延を示す。また、この結果を図示したものを図13に示す。いずれの図表も、2次元格子型の場合の遅延を100とした結果を示す。図より、接続階層分散型は2次元格子型よりも遅延が大きくなっていることが分かる。これは、各サーバーが接続するスイッチの階層が分散することにより、サーバー間のトラフィックが経由するスイッチの数が増大してしまうことが原因であると考えられる。

以上より、接続階層分散型は消費電力、遅延ともに2次元格子型よりも悪化する場合もあるため、サーバーと接続するスイッチは特定の階層に集約する構成が良いと考えられる。

表 5: サーバーとスイッチの接続階層の消費電力量による比較 (平均 500 bit、標準偏差 20 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)

		1mm			2mm			3mm		
		5×5	10×10	15×15	5×5	10×10	15×15	5×5	10×10	15×15
2次元格子型		100	100	100	100	100	100	100	100	100
接続階層 集約型	平均	97.6	93.7	90.4	97.8	94.2	91.4	98.0	94.6	91.9
	最小	97.2	90.7	87.4	97.5	91.2	89.5	97.8	92.4	90.0
	最大	97.8	94.8	92.7	98.0	95.7	92.6	98.3	95.2	93.1
接続階層 分散型	平均	99.5	85.8	78.6	99.5	86.8	82.3	100.8	88.5	84.0
	最小	94.2	73.5	64.2	94.3	74.1	70.5	95.0	79.2	75.1
	最大	126.2	110.9	105.2	127.9	115.5	111.9	128.8	119.4	115.8

表 6: サーバーとスイッチの接続階層の消費電力量による比較 (平均 500 bit、標準偏差 100 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)

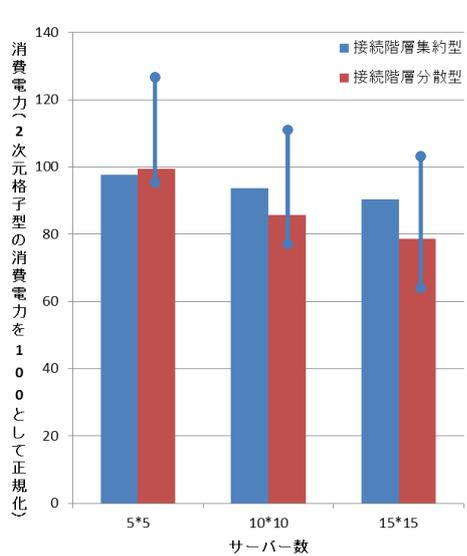
		1mm			2mm			3mm		
		5×5	10×10	15×15	5×5	10×10	15×15	5×5	10×10	15×15
2次元格子型		100	100	100	100	100	100	100	100	100
接続階層 集約型	平均	97.6	93.6	90.6	97.8	94.1	91.3	98.0	94.5	92.0
	最良時	97.4	90.9	87.4	97.6	91.5	89.3	97.8	92.4	90.0
	最悪時	97.8	94.9	91.3	98.0	95.7	92.8	98.2	96.5	93.4
接続階層 分散型	平均	99.0	84.9	77.9	99.6	86.2	81.4	101.0	88.8	83.5
	最良時	93.1	71.5	65.4	94.4	73.1	66.8	93.1	79.9	76.4
	最悪時	113.3	105.4	100.2	113.6	106.8	101.3	115.8	107.2	105.1

表 7: サーバーとスイッチの接続階層の消費電力量による比較（特定サーバー間トラフィックの場合）

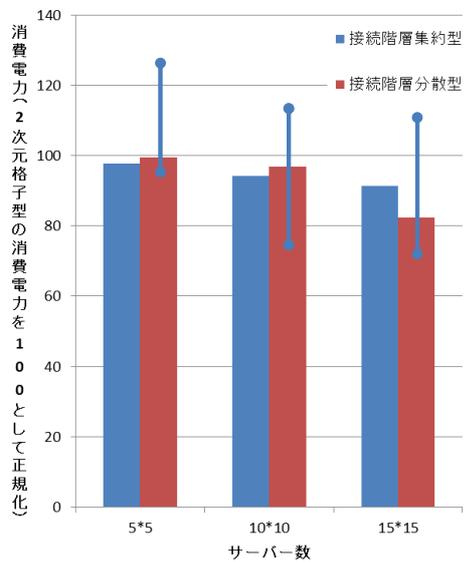
通信サーバーペア数	10	100	500	1000
2次元格子型	100	100	100	100
接続階層集約型	76.1	78.1	86.0	88.6
接続階層分散型	71.6	73.8	76.2	77.8

表 8: サーバーとスイッチの接続階層の遅延による比較（特定サーバー間トラフィックの場合）

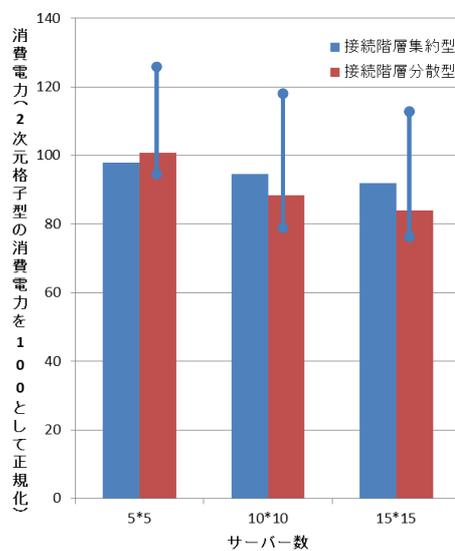
通信サーバーペア数	100	500	1000	1500
2次元格子型	100	100	100	100
接続階層集約型	57.8	48.1	45.7	45.5
接続階層分散型	115.9	110.9	150.3	144.8



(a) リンク長 1mm

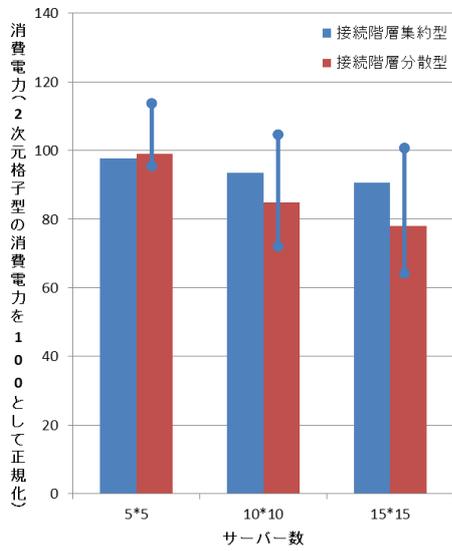


(b) リンク長 2mm

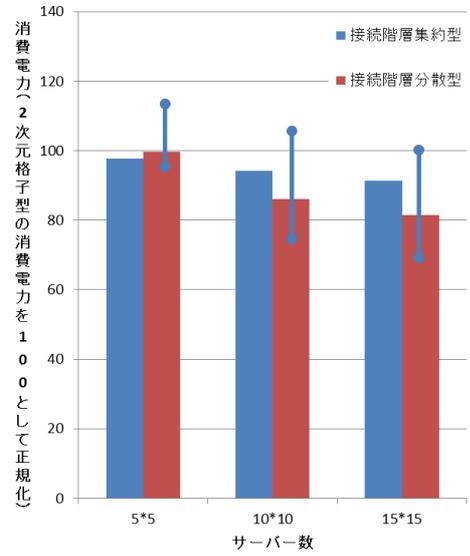


(c) リンク長 3mm

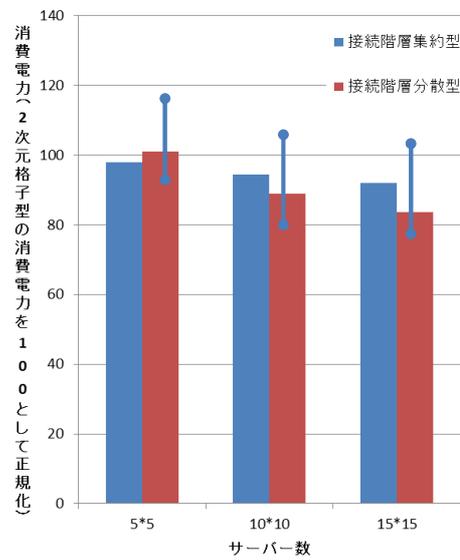
図 10: サーバーとスイッチの接続階層の消費電力量による比較 (平均 500 bit、標準偏差 20 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)



(a) リンク長 1mm



(b) リンク長 2mm



(c) リンク長 3mm

図 11: サーバーとスイッチの接続階層の消費電力量による比較 (平均 500 bit、標準偏差 100 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)

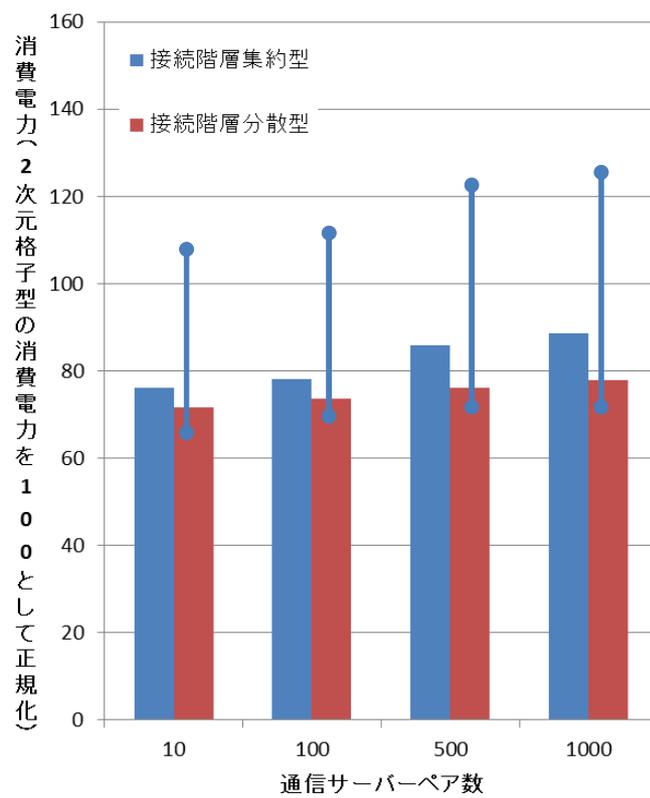


図 12: サーバーとスイッチの接続階層の消費電力量による比較 (特定サーバー間トラヒックの場合)

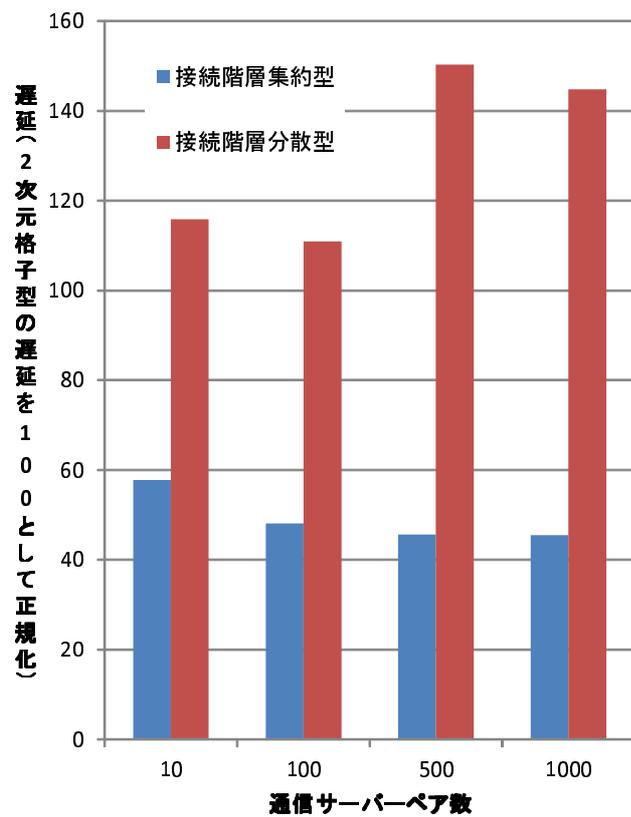


図 13: サーバーとスイッチの接続階層の遅延による比較 (特定サーバー間トラヒックの場合)

3.4.3 階層内の配置の検討

本小節では、パケット交換スイッチのみで構成した階層があった方がよいのか、パケット交換スイッチと回線交換スイッチを混在させた方がよいのかについて検討を行う。検討は、3.4.2節と同様、2次元格子型の場合の消費電力を100とした消費電力の比較により行う。表9、表10にサーバー数25、100、225台、リンク長1mm、2mm、3mmの各ネットワーク構成において全サーバー間通信を与えた場合の消費電力、表11にサーバー数225台、リンク長2mmのネットワーク構成において特定サーバー間通信を与えた場合の消費電力を示す。また、全サーバー間通信の場合の消費電力の比較結果を図示したグラフを図14と図15、特定サーバー間通信の場合の消費電力の比較を図示したグラフを図16に示す。

これらの図より、スイッチ混在階層構成型の消費電力は2次元格子型よりも悪化していることが分かる。これは、スイッチ混在階層構成型では、パケット交換スイッチの周囲を回線交換スイッチで囲む構成となっているため、サーバー間のホップ数が増大することが原因である。

表12に、特定サーバー間通信を発生させた際の遅延を示す。また、この結果を図示したものを図17に示す。いずれの図表も、2次元格子型の場合の遅延を100とした結果を示す。

これらの図より、スイッチ混在階層構成型は2次元格子型と比べて、遅延を削減できていることが分かる。これは、回線交換スイッチを設定することにより、通信量が多いパケット交換スイッチ間を直接接続し、パケット交換スイッチの処理にかかる遅延を削減できているからである。

しかしながら、スイッチ混在階層構成型よりも単一種類階層構成型の方が、より遅延を短くすることができている。これは、スイッチ混在階層構成型ではパケット交換スイッチ同士が直接接続されていないため、たとえ隣接するサーバー間の通信であっても、回線交換スイッチの設定によっては複数のパケット交換スイッチを経由する事が発生してしまうことが原因である。それに対して、単一種類階層構成型では、隣接するサーバー同士は、そのサーバーが接続するパケット交換スイッチも直接接続している。そのため、通信するサーバーペア数が増え、利用可能な回線交換スイッチがない場合でも、経由するパケット交換スイッチの数が2次元格子型と同じ経路は存在し、経由するパケット交換スイッチの数が極端に大きい経路は存在しない。その結果、単一種類階層構成型はより少ない遅延でサーバー間を接続可能となる。

このようにスイッチ混在階層構成型と比較して単一種類階層構成型の消費電力が小さく、また遅延も常に低く抑えられていることから、各層にパケット交換スイッチと回線交換スイッチを混在させた構成よりも、サーバーと接続したパケット交換スイッチのみの階層を構築し、それ以外の階層で回線交換スイッチを配置する構成の方が、オンチップ型データセン

ターに適していると考えられる。

表 9: 階層内の配置構成の違いによる消費電力量による比較 (平均 500 bit、標準偏差 20 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)

	1mm			2mm			3mm		
	5×5	10×10	15×15	5×5	10×10	15×15	5×5	10×10	15×15
2次元格子型	100	100	100	100	100	100	100	100	100
単一 種類 階層 構成型	97.6	93.7	90.4	97.8	94.2	91.4	98.0	94.6	91.9
スイッチ混在 階層 構成型	198.2	171.2	143.4	200.5	172.3	145.5	201.8	173.7	149.4

表 10: 階層内の配置構成の違いによる消費電力量による比較 (平均 500 bit、標準偏差 100 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)

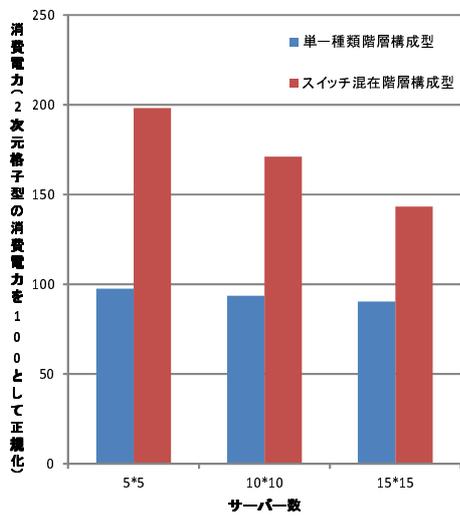
	1mm			2mm			3mm		
	5×5	10×10	15×15	5×5	10×10	15×15	5×5	10×10	15×15
2次元格子型	100	100	100	100	100	100	100	100	100
単一 種類 階層 構成型	97.6	93.6	90.6	97.8	94.1	91.3	98.0	94.5	92.0
スイッチ混在 階層 構成型	196.5	178.4	149.1	196.7	179.6	151.5	197.0	182.8	153.8

表 11: 階層内の配置構成の消費電力量による比較 (特定サーバー間トラフィックの場合)

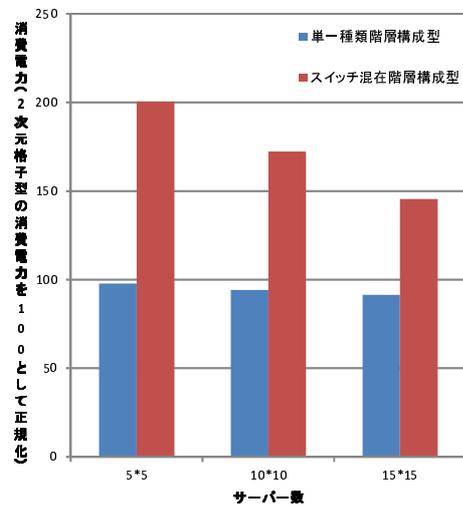
通信サーバーペア数	10	100	500	1000
2次元格子型	100	100	100	100
単一種類階層構成型	76.1	78.1	86.0	88.6
スイッチ混在階層構成型	119.4	126.7	132.3	137.1

表 12: 階層内の配置構成の遅延による比較 (特定サーバー間トラフィックの場合)

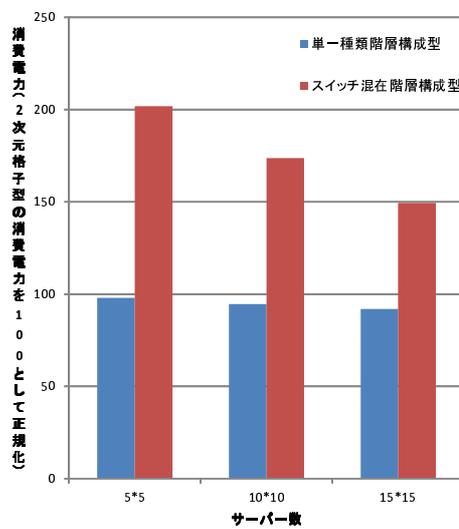
通信サーバーペア数	100	500	1000	1500
2次元格子型	100	100	100	100
単一種類階層構成型	57.8	48.1	45.7	45.5
スイッチ混在階層構成型	78.2	48.2	46.7	59.2



(a) リンク長 1mm

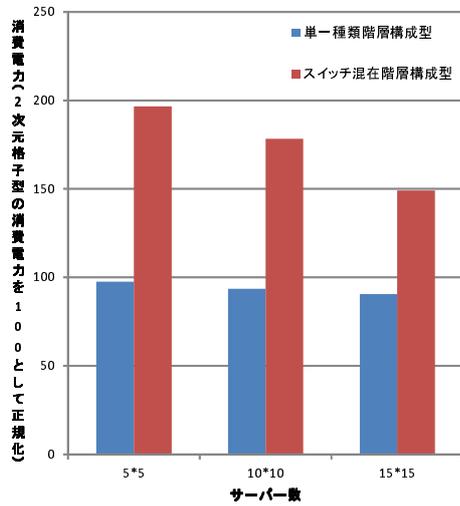


(b) リンク長 2mm

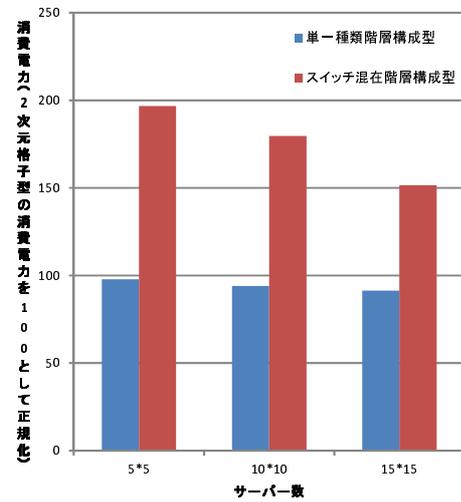


(c) リンク長 3mm

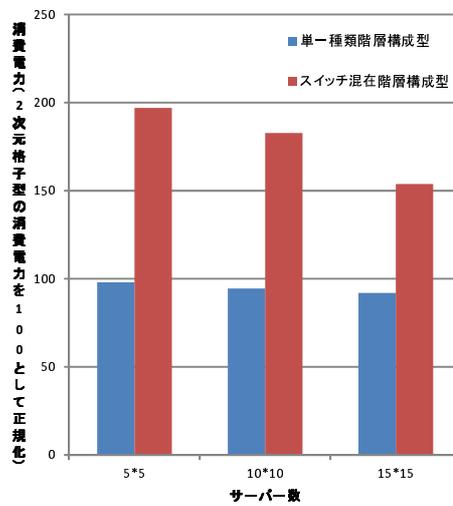
図 14: 階層内の配置構成の違いによる消費電力量による比較 (平均 500 bit、標準偏差 20 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)



(a) リンク長 1mm



(b) リンク長 2mm



(c) リンク長 3mm

図 15: 階層内の配置構成の違いによる消費電力量による比較 (平均 500 bit、標準偏差 100 bit の正規分布に従うトラフィックが全サーバー間に流れる場合)

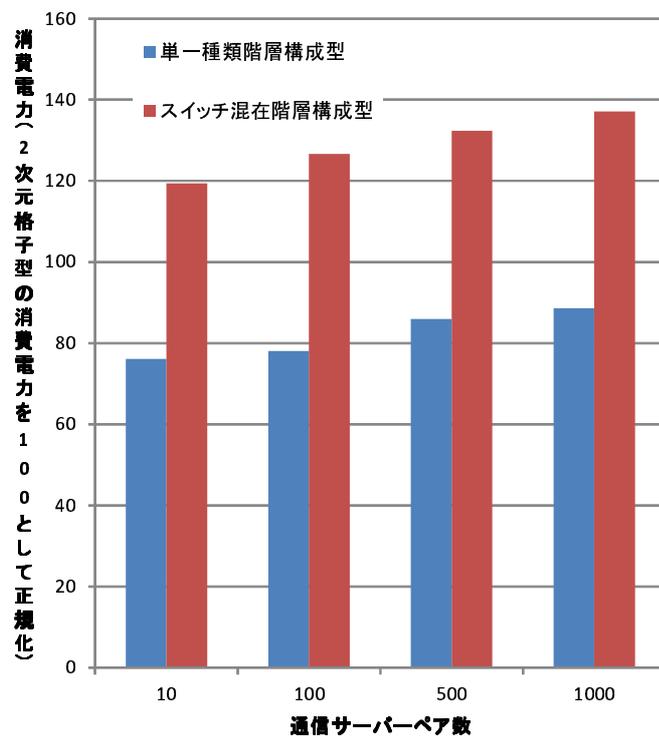


図 16: 階層内の配置構成の消費電力量による比較 (特定サーバー間トラヒックの場合)

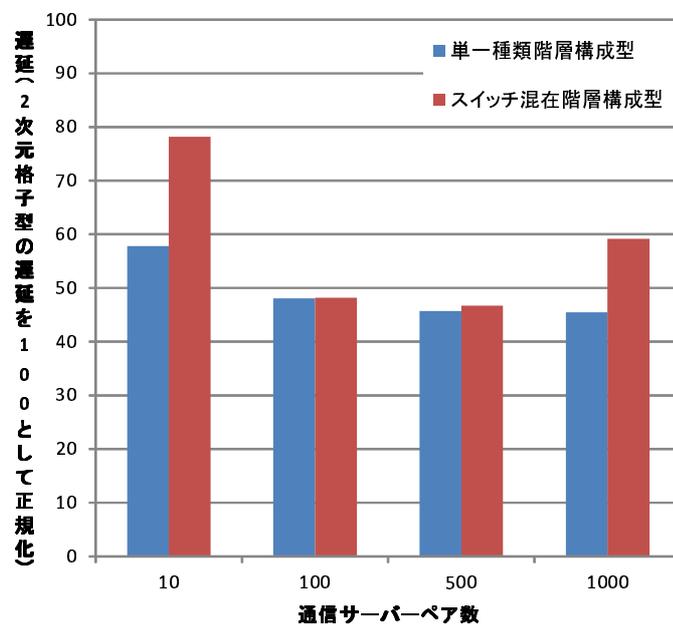


図 17: 階層内の配置構成の遅延による比較 (特定サーバー間トラヒックの場合)

4 終わりに

本報告では、データセンターのための3次元オンチップ型ネットワーク構造の構成について消費電力と遅延の観点から評価を行った。

評価の結果、各サーバーが直接接続するパケット交換スイッチを同一階層に配置した上で、同一階層内はパケット交換スイッチ、あるいは回線交換スイッチで統一し、階層間は同一位置のパケット交換スイッチと全階層の回線交換スイッチを接続する構成が、最も消費電力と遅延を削減できることが明らかになった。

また、削減できる消費電力量は通信を行うサーバの数と同一階層内のリンク長に依存し、サーバ数が増加すると、使用できる回線交換スイッチの数が減少するため、削減できる消費電力量の割合が小さくなることも明らかとなった。

今後はデータセンターのためのオンチップ型ネットワークに適した経路制御手法を検討する予定である。

謝辞

本報告を終えるにあたり、ご指導やご教授をいただきました大阪大学大学院情報科学研究科の村田正幸教授、ならびに大阪大学大学院情報科学研究科の大下裕一助教に厚くお礼申し上げます。また、奈良工業高等専門学校情報工学科の井上一成教授、大阪大学大学院情報科学研究科の若宮直紀教授ならびに大阪大学大学院情報科学研究科の荒川伸一准教授には適切な助言と指導を頂きました。心よりお礼申し上げます。さらに報告の作成にあたり、さまざまな助言を頂きました樽谷優弥氏、下間雄太氏をはじめとする村田研究室の皆様方に厚くお礼申し上げます。

参考文献

- [1] J. G. Koomey and P. D., “Growth in data center electricity use 2005 to 2010,” *The New York Times*, Aug. 2011.
- [2] D. Greenberg, P. Levis, K. Farkas, C. B. Morrey, III, and M. Neufeld, “Policies for dynamic clock scheduling,” 2000.
- [3] B. Heller, S. Seetharaman, P. Mahadevan, Y. Yiakoumis, P. Sharma, S. Banerjee, and N. McKeown, “Elastictree: Saving energy in data center networks,” in *Proceedings of USENIX*, pp. 17–32, Apr. 2010.
- [4] Y. Tarutani, Y. Ohsita, S. Arakawa, and M. Murata, “A virtual network to achieve low energy consumption in optical large-scale datacenter,” in *Proceedings of IEEE International Conference on Communication Systems*, Nov. 2012.
- [5] M. Kas, “Toward on-chip datacenters: a perspective on general trends and on-chip particulars,” *The Journal of Supercomputing*, vol. 62, pp. 214–226, Oct. 2012.
- [6] P. T. Wolkotte, G. J. M. Smit, N. Kavaldjiev, J. E. Becker, and J. Becker, “Energy model of networks-on-chip and a bus,” in *Proceedings of IEEE International Symposium on System-on-Chip*, pp. 82–85, Nov. 2005.
- [7] F. Li, C. Nicopoulos, T. Richardson, and Y. Xie, “Design and management of 3D multiprocessors using network-in-memory,” in *Proceedings of ISCA*, pp. 130–141, June 2006.
- [8] M. Al-Fares, A. Loukissas, and A. Vahdat, “A scalable, commodity data center network architecture,” vol. 38, pp. 63–74, Oct. 2008.
- [9] D. Abts and B. Felderman, “A guided tour of data-center networking,” *Communications of the ACM*, vol. 10, pp. 44–51, June 2012.
- [10] S. Borkar, “Thousand core chips : a technology perspective,” in *Proceedings of DAC*, pp. 746–749, June 2007.
- [11] T. Bjerregaard and S. Mahadevan, “A survey of research and practice of network-on-chip,” vol. 28, Mar. 2006.

- [12] M. B. Stensgaard and J. Sparso, “ReNoC: A network-on-chip architecture with reconfigurable topology,” in *Proceedings of the Second ACM/IEEE International Symposium on Networks-on-Chip*, pp. 55–64, Apr. 2008.
- [13] M. Modarressi and H. Sarbzi-Azad, *A High-performance and low-power on-chip network with reconfigurable topology*. Dynamic reconfigurable network-on-chip design: innovations for computational processing and communication, June 2010.
- [14] R. Iyer, R. Illikkal, L. Zhao, S. Makineni, D. Newell, J. Moses, and P. Apparao, “Datacenter-on-chip architectures: Tera-scale opportunities and challenges in Intel’s manufacturing environment,” *Intel Technology Journal*, vol. 11, pp. 227–237, Aug. 2007.