

1.2 本文

1.2.1 光電ハイブリッドスイッチシステムの研究開発について

2020年初頭からの世界的なパンデミックは、リモートワークなど、私たちのライフスタイルに大きな変化をもたらした。デジタル社会でのライフスタイルの変化は、そのインフラであるデータセンタにも大きな負荷をもたらしていることは周知のとおりである。本研究開発プロジェクトでは、まえがきでも述べた通り、爆発的増加が予想されるデータセンタでの消費電力削減を目的として行ったものである。近い将来のデータセンタのアーキテクチャは、CPUやメモリ、ストレージなどのデータ処理資源をひとまとめに配置したディスタグリゲーション・アーキテクチャと呼ばれる構成に移行すると予想される。そこでは、データ処理の負荷に応じて、適切な計算資源の選択を効率的に行うことが求められる。すなわち、データセンタ内における計算資源群を結ぶネットワークの高速化と、その消費電力の低減が不可欠である。これまでのデータセンタは、電気スイッチを多数接続した”leaf-spine”と呼ばれるネットワーク構成が一般的であり、接続される計算機資源のラック数は1000にもなると言われている。光産業技術振興協会の試算によれば、2030年にはディスタグリゲーション型のデータセンタではサーバやストレージもさることながら、データ処理資源間の接続に使われるネットワークの消費電力が占める割合は3/4にも上がることが示されている（図1.1）。

データセンタの電力削減のため、本プロジェクトではその消費電力が非常に低い光スイッチの活用に着目する。光スイッチは消費電力が低いという利点を有する反面、その規模と切り替え速度との間にトレードオフがあることが知られており、これまでデータセンタに要求される性能・機能を満たすことができず、データセンタ内のネットワーク用には実用的に用いられてこなかった。すなわち、従来のデータセンタにおいて唯一の選択肢である電気スイッチと同等の機能（高速なルーティング、メモリへのパケットの蓄積）などを提供することが光スイッチには困難であり、データセンタへの適用検討は学会等での報告が若干あるのみにとどまっていた。本プロジェクトでは、データセンタにおけるデータトラフィックの特性に着目し、従来の電気スイッチと光スイッチを組み合わせる新しいアーキテクチャのネットワーク、すなわち本プロジェクトのタイトルである「光電ハイブリッドスイッチシステム」を採用することで、上述の光スイッチの欠点を問題としないデータセンタスイッチシステムの実現を目指してきた。

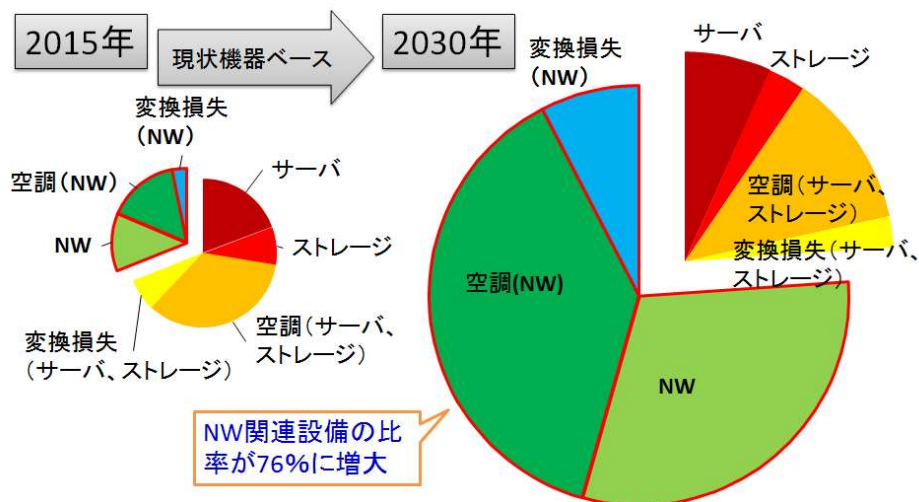


図 1.1 データセンタにおける消費電力予測の推移

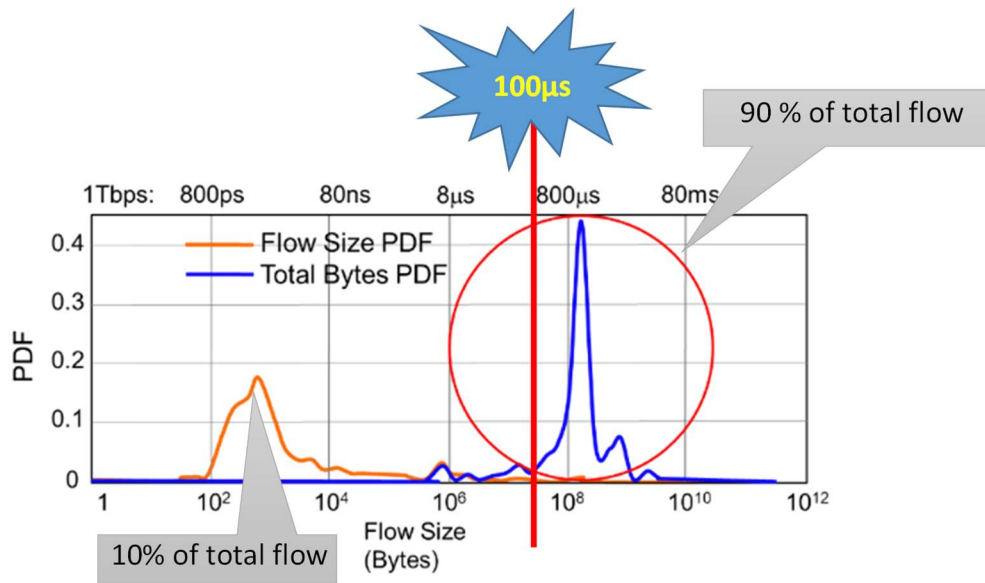


図 1.2 データセンタにおけるトラフィックのフローサイズ、トータル長に対する発生確率の分布

このデータセンタにおけるデータトラフィックの特性とは、図 1.2 に示すトラフィックのフローサイズ、トータル長に対する発生確率の分布である。図 1.2 において横軸はトラフィックのフローサイズやそのフローのトータルの持続時間であり、縦軸はその発生確率を示す。オレンジの線に着目するとフローサイズが 10^3 バイト程度の長さのものが高い発生確率を示すが、そのトータルのバイト数は非常に少量であることがわかる。一方で、 10^8 バイト程度のフローはごくたまにしか発生しないが、そのトータルの量は全体のフローの 90% にも及ぶことがわかる。すなわちデータセンタにおけるネットワークトラフィックは、持続時間が短く頻発する **mice** と持続時間が長くたまにしか発生しない **elephant** に分けられる。本プロジェクトでは、この性質に着目し、すべてのトラフィックを光スイッチで処理せず、電気スイッチと光スイッチを適材適所で活用する。すなわち、**mice** フローは電気スイッチで **elephant** フローは光スイッチへと割り振ることで現実的なデータセンタ内ネットワークの実現を目指す。では、これまでにデータセンタスイッチとして実用に供されてこなかった光スイッチはどのようなものであるべきか？この問いに対する答えは図 1.2 に示されている。すなわち、図 1.2

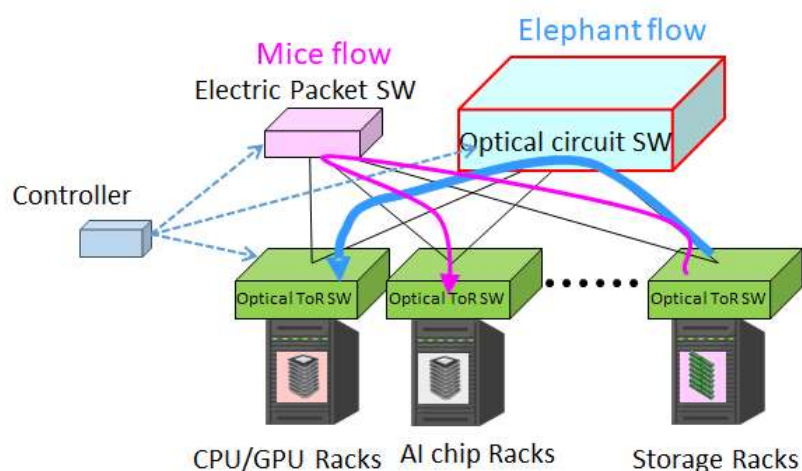


図 1.3 ディスアグリゲーション型データセンタに向けた光電ハイブリッドスイッチシステム

からわかる重要な事実として、mice フローと elephant フローは、およそ 100 μ s オーダの持続時間を境として分けられることがわかる。したがって、データセンタにおける光スイッチは、100 μ s オーダの切り替え速度を有し、かつ 1000 台程度のラック間を接続する規模であること、を要件として定義してもよい。

上述の前提をもとに、本プロジェクトで開発する光電ハイブリッドスイッチの概略を図 1.3 に示す。ディスアグリゲーション型データセンタでは CPU や GPU、AI 処理専用のラック、またはストレージ専用ラックなどが相互に接続される。ラック間は 100, 200, 400Gbps などの高速のリンクで接続されることが予想される。図 1.3 の光電ハイブリッドスイッチは、ラック間接続のネットワークを Electrical packet SW (以下 EPS)と示した電気の packets スイッチと Optical circuit SW(以下 OCS)と示した光スイッチ(光交換スイッチ)から構成される。前者の EPS としては、従来の電気スイッチを用いる。後者の OCS には、

- 1000 台程度のラック接続可能であること
- 切り替え速度が 100 μ s オーダであること

を要件として規定する。先に述べた通り、光スイッチの規模と切り替え速度はトレードオフの関係にあり、両者の積が性能指数として定義できる。図 1.4 は過去に報告された光スイッチの規模と切り替

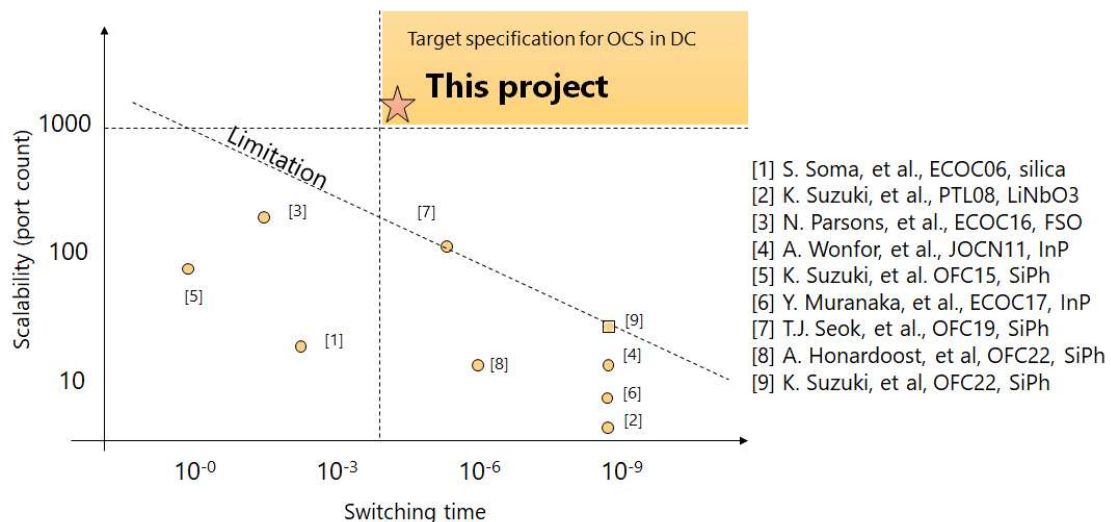


図 1.4 過去の光スイッチの規模と切り替え速度の報告例、および本プロジェクトの位置づけ

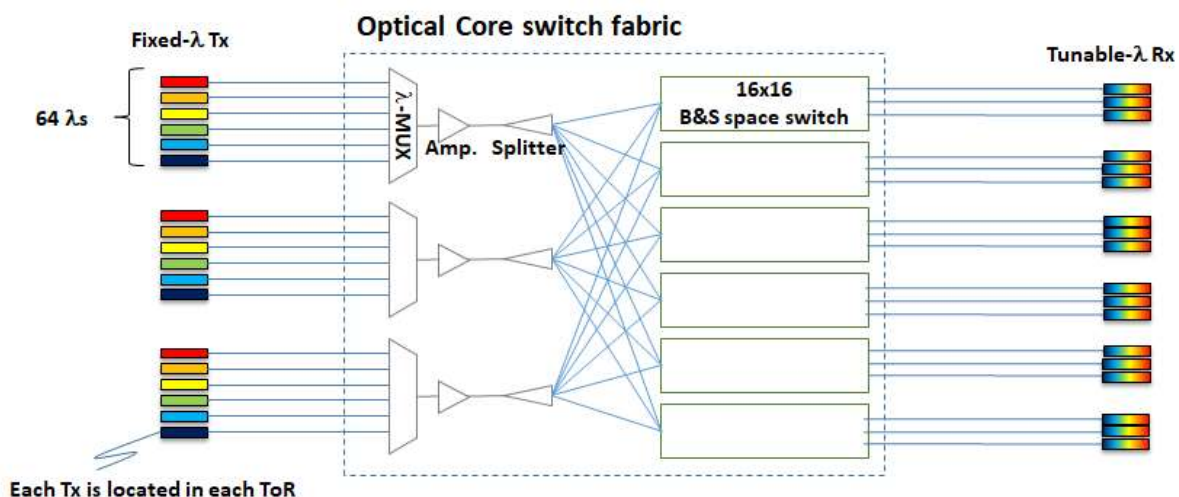


図 1.5 波長と空間を組み合わせた光スイッチの構成

え速度の報告例であるが、光電ハイブリッドスイッチシステムにおける光スイッチとしてはこの性能指数を大きく更新するものが必要である。したがって、光電ハイブリッドスイッチシステムの実現に向けては、既存の光スイッチでは不十分であり新たな開発が必要である。科学革命の構造でトーマスクーンが述べているように科学は通常科学とパラダイムシフトを引き起こす革新科学に分類される。本プロジェクトで要求される切り替え速度は既出の光スイッチで可能な領域に存在するのに対して、規模は過去の報告例を一桁程度上回るものであり、パラダイムシフトが要求される。パラダイムシフトを引き起こす 1000 ポート規模の光スイッチを実現するために、本プロジェクトでは光の特性に着目した波長と空間を組み合わせた光スイッチ構成を採用する。図 1.5 は波長と空間を組み合わせた光スイッチの構成である。光は電磁波でありその振動周波数 f に対して、 $\lambda = c/f$ (c は真空中の光速) なる波長を有する。すなわち光信号の周波数に応じて光の色が異なる。図 1.5 の構成では異なる波長の送信器を各ラックに設置し、特定の色がラックに対応するように設定する。後述するように 64 波長を一つの群として定義し、16 個の群を一つの光コアスイッチに送信する。光コアスイッチでは、ある群に含まれる信号を λ -MUX で波長合波したのちに、光アンプで増幅し、その後スプリッタで 64 分岐する。分岐した波長多重信号のうち一つは 16x16 B&S switch (B&S は broadcast and select の略) と呼ばれる空間スイッチに入力される。当該空間スイッチには、同様に 16 個の波長群をそれぞれ入力される。空間スイッチの内部構成はブロードキャスト (分配) — セレクト (選択) 構成であり、その 16 個のうちの出力の一つには、16 個の波長群のうちの一つが選択的に出力される。受信器は送信器と同様に各ラックに設置されるが、受信器は 64 波の信号のうちの一つを選択的に受信する波長選択機能を有する。したがって、受信器は 16 個の波長群に含まれる 64 波の信号のうちの一つを任意に選択して受信することができる。上述の光スイッチの規模は 64 波長の波長群を 16 群活用するので、1024 ポートの規模を達成可能である。光通信では $\lambda = 1.5 \mu\text{m}$ の C バンドと呼ばれる波長領域の信号を用いることが一般的であり、4.8THz の光周波数帯域幅を有する。この波長帯では、光ファイバをはじめ半導体レーザや受光素子、波長合分波器や光スイッチなどの基本的なデバイスが商用化されており、本プロジェクトのデータセンタネットワークの実現に好適である。前述のように 200~400Gbps の信号を想定する場合、デジタルコヒーレント信号では 64Gbaud 程度のシンボルレートであるため、75GHz 程度の波長間隔が適切である。この場合 4.8THz の光周波数帯域には 64 波長の配置が可能である。また、上述の構成では、64 波長の信号のうちの一つを受信機で選択する必要があるが、波長選択機能としてコヒーレント検波による波長選択性を活用する。一般にデジタルコヒーレント方式は C バンドで実用化されており、この観点でも C バンドの光信号でネットワークを構成することが好ましい。さらにデジタルコヒーレント方式を用いる利点としては、受信感度の高さが挙げられる。図 1.5 に示したように、波長と空間を組み合わせた光スイッチの構成では、一つの信号を 1/1000 に分岐する必要があるため、受信感度の高さもデジタルコヒーレント方式の利点である。加えて、デジタルコヒーレント方式では、変調多値度を変化させることで波長配置の変更なく伝送容量を可変できる点も、利点の一つである。さらに波長と空間を組み合わせる光スイッチは、B&S Switch と波長群を増設することで、データセンタの規模(ラック数)に応じた構成が実現可能である点も付記する。

もう一つの要件である 100 μs の切り替え速度実現への開発要素は、図 1.5 の構成を考慮しなければならない。図 1.5 の構成において、16x16 B&S Switch が 100 μs で空間切り替えを行う必要があり、波長選択を行う受信器が 100 μs で波長選択を行う必要がある。さらに受信器ではバースト的に発生する信号を受信するためのバースト信号処理プロセッサも不可欠である。これら、100 μs の波長選択と受

信を行う信号処理プロセッサは、各ラックに設定される Top of Rack(ToR)スイッチである。本プロジェクトでは、光電ハイブリッドスイッチシステムを実現するため、これらのデバイス開発およびそれらを用いた検証実験を行った。その詳細は後述の各節で説明する。

本プロジェクトの最終的なゴールは、1Tbps の伝送システムの基盤技術を目指すものであるが、本研究期間内での CMOS 微細化技術の進展及び部品の入手性を考慮し、原理実証は 200Gbps システムにて実施した。さらに、当初は研究開発終了後に 200Gbps システムの事業化を推進するとともに、次ステップである 400G システムを参加組合員にて自主的に開発を進めることを計画したが、400 Gbps 市場が当初予想よりも早く立ち上がりつつあることを考慮し、400G システムの開発を前倒して実施した。

最後に本節に本プロジェクトの効果に関して、光電ハイブリッドスイッチシステムの導入障壁およびプロジェクトの眼目である電力削減効果について説明する。

本プロジェクトの光電ハイブリッドスイッチシステムは新しいネットワーク構成であり、既存の構成からのスイッチング（乗り換え）コストを考慮すべきである。表 1.1 は、従来の電気スイッチのみを用いたシステムと光電ハイブリッドスイッチを用いたシステムの装置規模に関して試算したものである。従来の電気スイッチのみによるシステムを 1 としたとき、光電ハイブリッドスイッチシステムでは、スイッチシステムの帯域はその 10%が電気スイッチに、90%が光スイッチに割り振られる。これは、図 1.2 に説明したように、mice が 10%、elephant が 90%のフローとして存在することに対応する。また、従来の EPS のシステムでは規模を拡大するために数 10 ポート規模のものを多数接続し 1000 ポート規模を実現する必要がある。光電ハイブリッドスイッチシステムでは、EPS で処理すべきトラフィックが従来の 1/10 であるため、EPS を構成するスイッチシリコンをより多数のポートで共有することが可能である。したがって、必要とされる EPS の台数、また EPS 間を接続するファイバリンク、リンクを接続するトランシーバの数をそれぞれ 1/10, 1/20 と大幅に削減できる。光スイッチ側のパスについては、EPS と異なり一層のネットワークであるため、光スイッチは究極には 1 台あればよい。リンクの数も従来の EPS と異なり 1 層で構わないので 1/5 程度に削減可能である。

本プロジェクトの目標は、基本計画の最終目標に基づき、2030 年のデータセンタにおいて事業開始時点に比べて 100 倍の電力効率を目指すものである。事業開始時点における技術である電気スイッチのみを用いたネットワークと比較した場合、光電ハイブリッドスイッチシステムでは上述の通り電気スイッチは 1/10 の規模で十分であり、残りの 90%のトラフィックは光スイッチで処理される。光スイッチは圧倒的に消費電力が小さく、本開発で実現を目指す ToR スwitchでの消費電力を考慮してもと、トータルでは 60~70%の電力削減効果、すなわち 2 倍以上の電力削減を期待できる。さら

表 1.1 従来の電気スイッチのみを用いたシステムと光電ハイブリッドスイッチを用いたシステムの装置規模に関する試算

	EPS only system	Photonic-electronic hybrid switching system	
		Electrical	Optical
Bandwidth	1	0.1	0.9
No. switch boxes	1	0.1	0.01 (2000x2000)
No. optical links (TRxs)	1	0.05	0.2

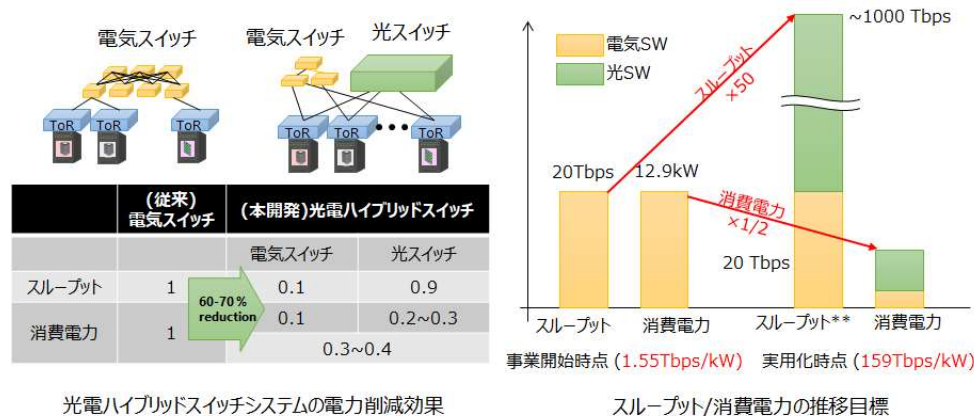


図 1.6 本プロジェクトの光電ハイブリッドスイッチシステムによる電力効率改善効果の試算

に、電気スイッチのスループットのロードマップを参照すると、事業開始時点(2018年)では20Tbps程度、さらに2030年まで外挿して考慮すると1000Tbps級までのスループット、すなわち50倍の性能が要求されると見積もられる。したがって、性能あたりの電力効率としては、2倍×50倍の100倍の電力削減効果を見込む。従来技術では1000Tbps級のスループットを電気スイッチで実現する必要があるが、光スイッチは光電変換を行わないため、電力はリンク速度によらないことが大きな利点であることを付記する。

以降に、

研究開発項目①-1「光電ハイブリッドスイッチ制御技術」

研究開発項目①-2「光電ハイブリッドスイッチ制御アルゴリズム」

研究開発項目①-3「光スイッチネットワークアーキテクチャの最適化の研究開発」

研究開発項目①-3-1「光スイッチネットワークアーキテクチャの研究開発」

研究開発項目①-3-2「光スイッチ要素機能部品・システム高性能化の研究開発」

研究開発項目②「光 ToR スwitchの研究開発」および②-1「光波長送受信器の研究開発」

研究開発項目②-2「バースト多値プロセッサの研究開発」

研究開発項目③「光コアスイッチの研究開発」.

研究開発項目④「国際標準化」

研究開発項目⑤「光電ハイブリッドスイッチシステム実証」

の個別テーマごとの、研究開発成果について述べる。