次世代サーバ内高速信号伝送技術

Research on Next-Generation Server Interconnects

● William W. Walker　● 日高康雄

あらまし

本稿では富士通研究所で行っているサーバ内の電気および光伝送技術の研究について紹介する。まず，サーバ内伝送に要求される伝送速度が年々高速化の傾向にある社会的背景に触れた後，電気伝送技術としてチャネル損失を補正する「等化」という技術について説明する。そして，この等化技術により25 Gbpsという高速伝送を電気的に実現する際の課題と，低消費電力化を図りつつ40 Gbpsを光伝送技術で達成する際の課題を示す。

Abstract

This paper explains the need for continuous bandwidth improvement in computer server interconnects, and the work that is being done at Fujitsu Laboratories to provide the bandwidth, both electrically and, in the future, optically. It explains concepts of frequency dependent channel loss and equalization, and the hurdles to be overcome to reach 25 Gbps per lane using electrical HSIO and 40 Gbps per lane using optical HSIO while improving energy efficiency.

ま　え　が　き

1994年に始まったインターネットの商業化は，今日に至るまで通信需要の指数関数的な増加をもたらしてきた。[1] 最近耳にする「ビデオ・オン・デマンド」，「スマートフォン」や「ソーシャルネットワーク」などはこういったインターネット通信の成長を牽引している一例である。データセンターではこの膨大な量の通信をサーバを用いて処理している。富士通研究所ではサーバ内のバックプレーンを介して通信を行う高速入出力回路（HSIO：High-Speed Input/Output）の研究開発を行っている。

本稿では，次世代のHSIOの設計における困難な課題について述べた後，電気伝送から光伝送への移行が低消費電力化と高速化の両立には不可欠であることを説明する。

広帯域化の問題

図-1はHSIOのロードマップを示している。横軸は年を表し，左縦軸は1レーンあたりの伝送レートをGbps（1秒あたり10億ビット）の単位で表している。「レーン（Lane）」は電気通信の場合，一対の差動配線であり，光通信の場合は1本の光ファイバか導波路である。右縦軸は1ビット伝送するために消費されるエネルギー，pJ/bitを表している。p（ピコ）は1兆分の一という非常に小さい値で，J（ジュール）は，例えば，1リットルのガソリンは3200万Jのエネルギーを含んでいる。このように，1レーンあたりの伝送レートを向上させつつ，1ビット伝送するために必要なエネルギーを低下させることが求められている。伝送レートを向上させることを広帯域化という。

電気伝送における広帯域化の問題を山で起きる分散とこだまの現象を例えに説明してみよう。高い山の峰の上にいる登山家，Joeが隣の山の峰に友人を見つけ，「Hello，I'm Joe ！」と呼びかけたとする（**図-2**）。

ところが友人が最初に耳にするのはかすかな「Hello」，続いて「Hello」のこだまと重なった「I'm」，そして「I'm」のこだまと「Hello」の二度目のこだまと重なってほとんど聞き取れなくなった「Joe」となる。Joeが同じように呼びかけ続けるとこだまが重なって友人にとっては全く聞き取れない騒音となってしまう。友人が聞き取れるようにするには，Joeは前に言った言葉のこだまが聞こえなくなるまで待ち，こだまと次の言葉が重ならないくらいにゆっくり話さなければならない。すると，一つの文を話し終わるのに長い時間がかかってしまう。この情報を伝達する速度のことを「帯域」と言い，相手が聞き取れるようにするには，帯域を落とす必要がある。

音声ではなく電気信号で，「1」と「0」という二つだけの語彙を使っているところが異なるにしても，これがまさにHSIO回路の設計者が日々奮闘し

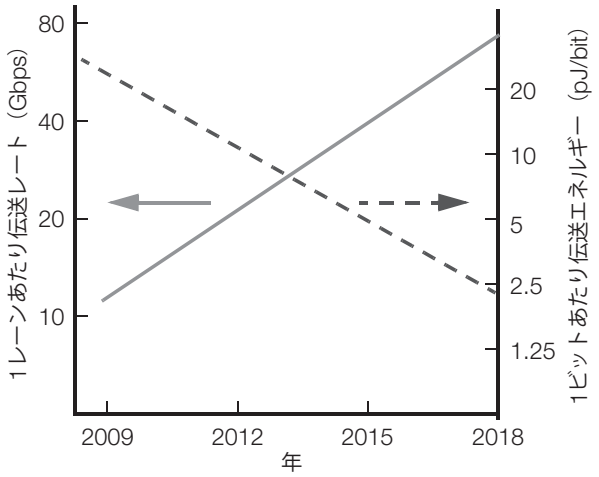

図-1　高速入出力回路のロードマップ
Fig.1-HSIO roadmap.

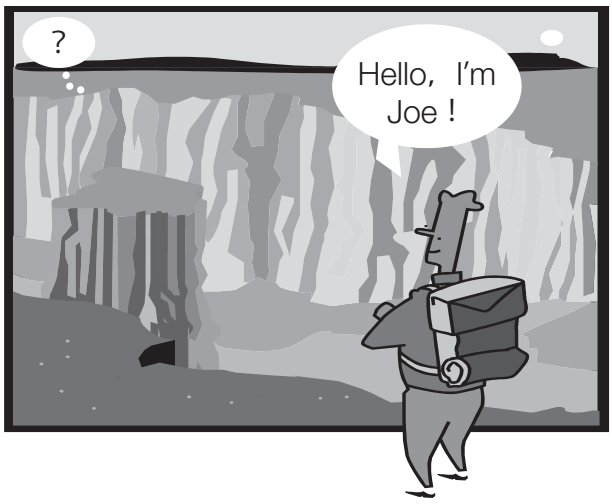

図-2　山で起きる分散とこだまの現象
Fig.2-Problem of signal dispersion and echo.

ている現象である。なお，通信の分野では，語彙に含まれる単語のことを符号（Symbol）といい，複雑な通信方式では，「0，1，2，3」のような符号集合を用いる場合もある。Joeの符号集合には，すべての英単語が含まれている。情報を伝達する媒体を「伝送路（Channel）」と呼ぶ。Joeの「伝送路」は渓谷で囲まれた空気であるが，電気伝送で用いられる伝送路は，プリント基板（PCB）上の銅線（「トレース（Trace）」と呼ばれる）と，PCBの逆の面や内部の層にある別のトレースとの間の接続（「ビア（Via）」と呼ばれる）によって構成されている。

分散と反射

Joeの声が友人に届くときには小さくなったのは空気中のあらゆる方向に声が分散したからである。電気信号の場合は電磁界の放射とトレースの抵抗が信号を分散させる。トレースの抵抗は配線を太くすれば低く抑えられるように思えるが，その分，単位面積あたりに引けるトレースの本数が減ってしまうので，効果的ではない。トレースの間隔が2倍になると，何らかの方法でトレースあたりの信号速度を2倍にしない限り，帯域は半分になってしまう。今日我々が使っている10 Gbps程度の高速伝送になると，「表皮効果」と呼ばれる物理現象によって，電流は導体の表面付近にしか流れない。周波数が2倍になると，この表皮効果によってトレースの抵抗が増加するため，そもそも抵抗を減らすために配線を太くした意味がなくなってしまう。このように，周波数が高くなるにつれ，信号の減衰が大きくなるため，伝送速度を単純に速めることはできない。

さらに悪いことに，ある周波数fで伝送される「1」と「0」の符号列は，実は，広い範囲の周波数成分を含んでおり，符号列が伝送路を伝播する間に，その周波数成分の内の高周波成分が低周波成分に比べて大きく減衰し，受信端では信号波形が鈍ってしまう。この波形鈍りは隣接符号にも影響を及ぼすため，「符号間干渉（ISI：Inter Symbol Interference）」を引き起こす。これはこだまとは似ているようで異なる現象である。

電気的なこだまは「反射」と呼ばれ，ビアのようにトレースが不連続な箇所を信号が通過する際に生じる。不連続点があると信号の一部がこだまとして跳ね返り，そのこだまが別の不連続点を見つけるとそこで再び跳ね返り，受信回路に向けて，元の信号をこだまが追いかけていくという現象が起こる。これはJoeの声が峰から峰へ反射を繰り返して友人の耳にたどり着いたのと同様の現象である。多重反響が起こると友人がJoeの言っていることが全く分からなかったように，電気信号も受信端で受信しづらくなる。なお，電気信号のこだま（反射）は，元信号の微分（変化量）であることを付け加えておく。

図-3はデジタル信号が前述した周波数依存の減

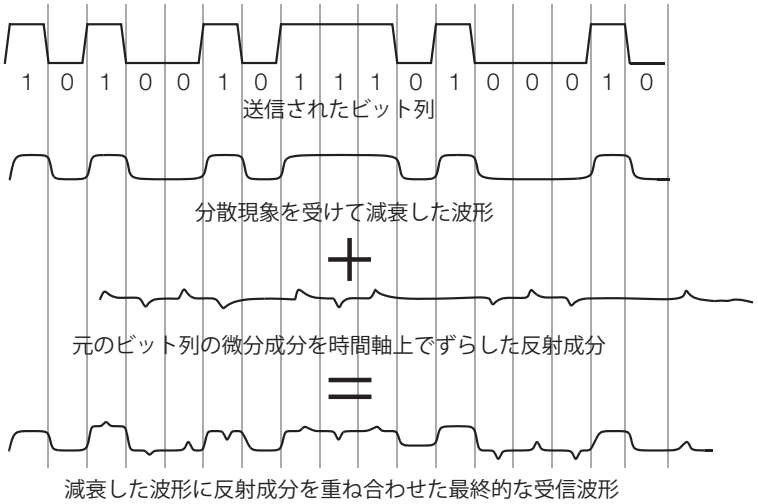

図-3　デジタル信号伝送で起きる分散と反射の現象
Fig.3-Problem of dispersion and echo in digital world.

衰と反射の影響を受けた状況を示している。一番上の波形は送信された元のビット列であり，二番目の波形は同じビット列が反射はないが損失のある伝送路を通過した後の減衰した波形を示している。元のビット列の微分成分を時間軸上でずらした微弱な反射成分が三番目の波形であり，最終的に受信回路に届くのは減衰した波形に微弱な反射成分を重ね合わせた一番下の歪(ひず)んだ波形となる。

上で述べたように，サーバ内伝送で扱われる符号は「1」と「0」のみで構成されているため，「等化」と呼ばれる極めて簡単な回路技術を用いて「符号間干渉」や「反射」を部分的に取り除くことができる。

等　化　技　術

信号は送信から受信の間に存在する媒体を伝送する内に減衰してしまい，減衰の具合は周波数成分によって異なる。

等化技術とは，特定の周波数成分を増幅または減衰させることで，望まれざる物理的な効果を打ち消すための回路技術のことである。等化技術は長い歴史を持っている。例えば，低音を減衰させることで，レコード盤の溝を近接させ，1枚のレコードに長時間の音楽を記録できるようにと開発された技術がLPレコードである。記録時に減衰させた低音は，再生時に増幅することによって，自然な低音が再現されるようになっている。

著者らが開発した最先端のバックプレーン用のHSIO回路[(2)]においては以下の三つの等化技術が用いられている（**図-4**）。

(1) プレエンファシス（Pre-emphasis）

信号を伝送する前に，直前または直後に伝送するビットの一部をあらかじめ加算もしくは減算してから伝送することで，符号間干渉を取り除く等化技術である。

(2) 連続時間線形等化器（CTLE）

受信器内に構成された高周波成分を増幅するアナログフィルタであり，やはり符号間干渉を抑制することができる。図-4を見ると分かるように，伝送路の周波数依存の減衰特性と逆の特性を持つように設計されている。

(3) デシジョン・フィードバック等化器（DFE）

直前に受信した符号が「1」か「0」であるかに従って，現在の符号が「1」か「0」を判定する閾値を設定する等化回路。符号間干渉の影響で直前の符号が「1」の場合，現在のビットの電圧が上昇してしまう場合，その分論理判定閾値を高く設定することで，符号間干渉を差し引く効果がある。図-4ではDFEは単純なスイッチとして表されている。

DFEは，原理的には過去のどの時刻に受信した符号であっても，その符号間干渉を差し引くことができるので，反射の影響を低減することも可能である。しかし，DFE回路で反射の影響をキャンセルするには多数のタップ（符号一つ分の回路のこと）を備える必要があり，回路面積および消費電力が著しく増えるため，実際は現実的ではない。より現実的な対策は，PCBを注意深く設計し，反射を低減させることである。

富士通研究所では，3次元電磁界シミュレータを用いてPCBのビアの改良を行っている。周囲の配線や電源層を注意深く設計することで，配線層間のスムーズな電気信号の移動を実現でき，信号の反射を極力抑え，DFEのタップ数を減らし，低消費電力化を実現することができる。また，周波

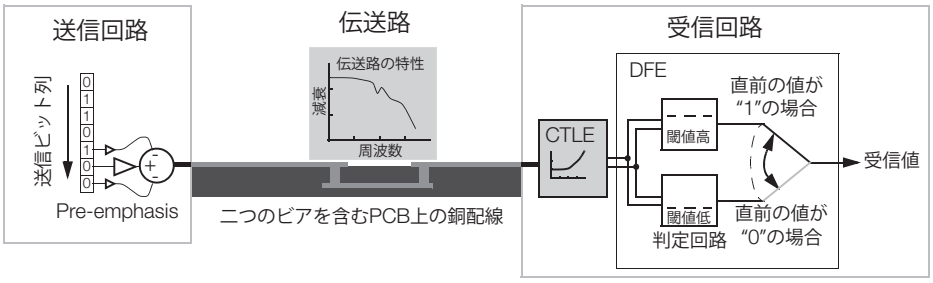

図-4　電気伝送路と高速入出力回路
Fig.4-Complete HSIO link.

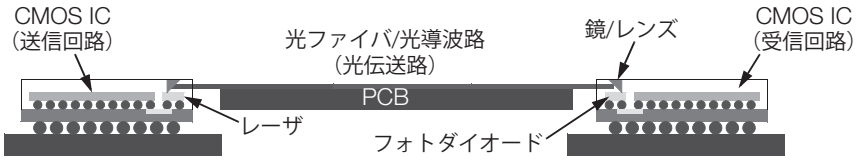

図-5　光伝送路と高速入出力回路
Fig.5-Optical HSIO link.

数依存の減衰の少ない材料を使ったPCBの研究も行っている。これらの技術を用いて，2012年までに1レーンあたり25 Gbps以上の伝送速度を実現できるだろう。

将来の高速伝送技術

電気伝送の速度向上はいずれ物理的な限界に達するであろう。従来，半導体プロセスの微細化とそれに伴う電源電圧の低下の恩恵を受けて，電気伝送の消費電力は年々減少の一途をたどってきた。しかし，半導体プロセスの微細化は，物理的にも技術的にも限界に近づいている。一方，上で論じたように，電気伝送を高速化するには伝送路の問題を解決する必要があり，より複雑な回路を使わざるを得なくなってきている。その次はどうしたら良いのであろうか？

光伝送は，電気伝送と較べて100メートル以上の長距離伝送では圧倒的に有利で，10から100メートルの距離の伝送ではほぼ同等な状態である。2009年にノーベル賞を受賞したチャールズ・カオ氏の功績[(3)]もあり，光ファイバは非常に低い損失を実現することができるようになった。この光伝送をバックプレーンに応用する技術が次に注目されている研究課題である。光伝送をバックプレーン上で実現できれば等化器をなくして回路を簡略化でき，消費電力を減らすことができる。

光バックプレーンの開発に当たっての課題を以下に示す。

(1) 垂直共振器面発光レーザ（VCSEL：Vertical Cavity Surface Emitting Laser）を用いた光源は，レーザ発光のために大きな電力を必要とする。しかし，伝送速度が高速化するにつれ，1ビットあたりの電力が減少するため，深刻な問題ではなくなってきている。

(2) 光素子，光パッケージ，および光伝送チャネルのコストがバックプレーンのコストに追加される。

(3) 一つのサーバには数百から数千の高密度の伝送路が必要であるが，現状の電気光変換モジュールは大き過ぎる。

(4) 従来，高価な化合物半導体技術を使わなければ，レーザ光を信号で変調したり，フォトダイオードから信号を復調したりすることができなかった。

富士通研究所ではマイクロプロセッサ向けの標準的なCMOS半導体技術を使用した電気光変換回路の開発にも取り組んでいる。この技術を用いれば，将来，マイクロプロセッサに光送受信ポートを搭載することも可能になる。著者らは既に40 Gbps/レーンの通信分野向けトランスポンダをCMOS半導体技術で実現可能であることを実証した。[(4)] 著者らは現在，この新しいCMOS回路技術と，高密度，低コストな光素子を同じパッケージに統合する技術を組み合わせることで，この技術をサーバのバックプレーン向けに適用しようとしている（**図-5**）。

む　す　び

指数関数的に拡大しているインターネットの通信需要を支えるために，コンピュータ内の伝送速度も指数関数的に向上させることが要求されている。富士通研究所は最先端のCMOSプロセス，高度な適応等化回路，そしてバックプレーン伝送路の設計技術を活用した，伝送速度10 Gbps，電力効率20 pJ/bitという世界最先端の高速伝送技術を保有しており，[(2)] 開発ロードマップでの優位を維持すべく，25 Gbpsの伝送技術も開発中である。また，伝送速度40 Gbps以上かつ電力効率5 pJ/bit以下の光伝送技術に向けた研究も始まっている。

参考文献

(1) Minnesota Internet Traffic Studies. http://www.dtc.umn.edu/mints

(2) Y. Hidaka et al.：A 4-Channel 10.3Gb/s Transceiver With Adaptive Phase Equalizer for 4-to-41dB Loss PCB Channel. Intl. Solid State Circuits Conference, Feb. 2011.

(3) K. C. Kao et al.：Dielectric-fibre waveguides for optical frequencies. *Proc. Of the Institution of Electrical Engineers*, Vol.113, No.7, p.1151-1158 (1966).

(4) N. Nedovic et al.：A 3-Watt 39.8 - 44.6 Gb/s Dual-Mode SFI5.2 SerDes Chip Set in 65nm CMOS. *IEEE Journal of Solid State Circuits*, Vol.45, No.10, p.2016-2029 (2010).

著者紹介

William W. Walker

米国富士通研究所 所属

現在，高性能CMOSメモリ回路，高速電気伝送I/O回路，および高速光伝送I/O回路の研究に従事。

日高康雄（ひだか　やすお）

米国富士通研究所 所属

現在，主としてサーバシステムの高速電気伝送I/O回路の研究開発に従事。