プロセッサ間通信向け超高速インタコネクト技術

Ultra-High-Speed Interconnect Technology for Processor Communication

● 土肥義康　● Samir Parikh　● 尾形祐紀　● 小柳洋一

あらまし

クラウドを構成するサーバやストレージシステムの性能を向上させるためには，システムをつなぐインタコネクトの高バンド幅化が不可欠である。富士通は既にUNIXサーバ「SPARC M10」のCPU間通信において，1信号線あたり14.5 Gbpsのデータ転送を実現するCMOS高速インタコネクトを開発，製品化しているが，更なる高バンド幅化を実現するため，1信号線あたり32 Gbpsの高速データ伝送を行うプロセッサ間インタコネクトの研究を行っている。高速化の実現に当たり，送信回路で高周波動作が必要のない通信方式，受信回路で広帯域のロス補償を行うイコライザ回路や，高精度なサンプルクロック生成が必要のないデータ受信方式を搭載し，32 Gbpsで30 dB以上のロス補償機能を有する高速インタコネクトを開発した。これら高速インタコネクト技術を，CPUチップに搭載することにより，サーバシステムの全体性能を2倍以上向上させることが可能となる。

本稿では，28 nmテクノロジーで試作を行った，これらの超高速インタコネクトの新規技術を紹介する。

Abstract

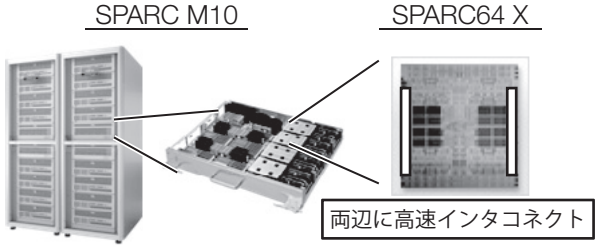
In order to improve the performance of storage systems and servers that make up the cloud, it is essential to have high-bandwidth interconnects that connect systems. Fujitsu has already marketed a CMOS high-speed interconnect product that works between CPUs for the UNIX server (SPARC M10). Its data rate per wire is 14.5 Gb/s. And Fujistu has recently researched a CMOS interconnect that can operate at over 32 Gb/s per lane to achieve a higher data rate. We researched some new techniques for high-speed interconnects such as a data interleaved driver in a transmitter, wideband loss compensation equalizer, and a clock and data recovery system using a data interpolator. When these technologies are implemented in a CPU chip, we can expect to double the performance of server systems. This paper introduces the features of ultra-high-speed interconnect technologies for 32 Gb/s serial interconnects implemented using 28 nm CMOS technology.

ま え が き

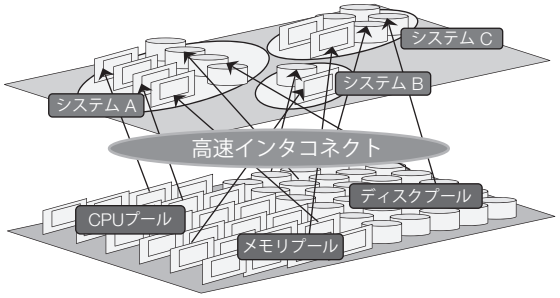
クラウドコンピューティングを支えるデータセンターなどに向けてサーバのデータ処理能力向上が一層求められている。そのような中で，年々続く半導体プロセスの微細化や，1チップに複数のCPUコアを搭載するマルチコアCPU構造の適用により，1チップで処理可能なデータ量が爆発的に大きくなっている。CPUで処理されたデータは，チップ内部の入出力回路を介して外部と送受信を行うが，チップの信号線数には制限があるため，バスを用いて通信速度を向上させることが困難である。このため，1信号線あたりの通信容量を極限まで高めた（10 Gbps以上）高速インタコネクトが要求されている｛**図-1**（a)｝。高速インタコネクトはCPUチップ間のみではなく，資源プール化アーキテクチャーにより仮想化されたCPU，メモリやストレージを接続するためにも用いられ，適用範囲が拡大している｛図-1（b)｝。デバイス間に必要な通信量は3年で2倍に増加することに比して，CMOSテクノロジーの性能向上速度は年率5％程度であり，その乖離が年々大きくなっている。富士通は既にUNIXサーバ「SPARC M10」[(1)]のCPU間通信において，1信号線あたり14.5 Gbpsのデータ転送を実現するCMOS高速インタコネクトを開発，製品化している。次世代のサーバに対しては，要求されるバンド幅は2倍以上となることが予想され，更なる研究開発が必要となっている。

高速インタコネクトの概要と課題

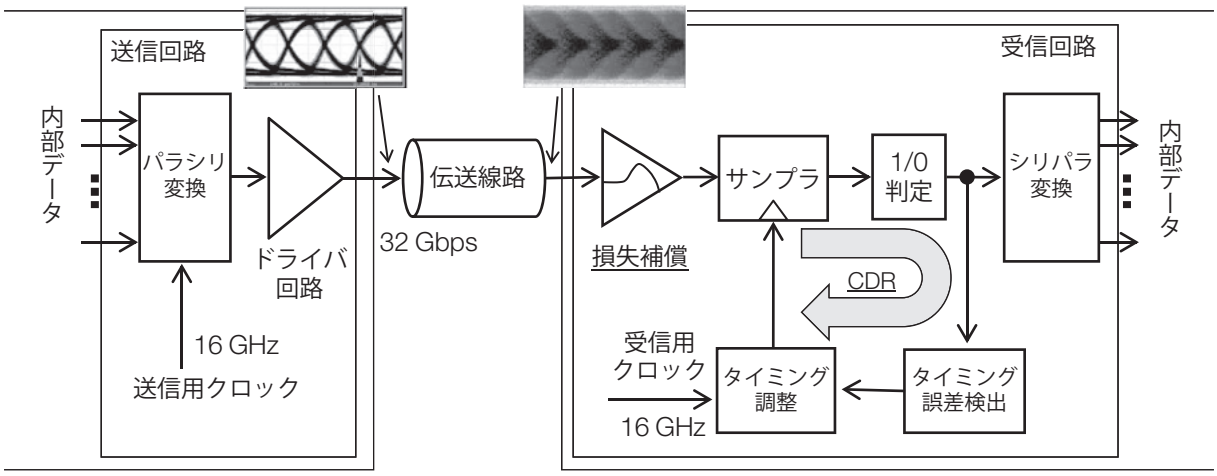
システムをつなぐ高速インタコネクトはCPUチップ同士のデータ通信を行うために同チップに併載される。例えばSPARC64 Xは，複数のCPUコアをチップ上に配し，CPUのデータ入出力のため，チップの左右両辺に高速インタコネクトマクロを搭載する｛図-1（a)｝。高速インタコネクトは，



（a）CPU



（b）資源プール化アーキテクチャー



（c）ブロック図

図-1　高速インタコネクトの概要

CPUなどから生成される内部データのパラレルーシリアル（パラシリ）変換を行い伝送線路に送出する送信回路と，送出された信号を入力とし，入力データのシリアルーパラレル（シリパラ）変換を行い，再びCPUに内部データとして1/0信号を送出する受信回路から構成される｛図-1（c)｝。送信回路と受信回路は，パッケージ，プリント配線基板やケーブルなどで構成される伝送線路を介して接続される。伝送線路は材質や形状に依存する損失を有しており，伝搬する信号の周波数成分が高いほど信号品質が劣化する。このため，送信回路から出力された1/0信号は受信回路に到達するまでに，データの1/0の判定が困難な波形となる。これを符号間干渉（ISI：Inter Symbol Interference）と呼ぶ。また，送信回路と受信回路は，別々のクロック源に同期動作しており，クロック源固有の周波数差や揺らぎによるタイミング誤差を有する。このため，受信回路は，信号の劣化を補償する損失補償機能と，受信回路入力データから送信回路のタイミング情報を検出し，動作タイミングを調整することで，受信回路と送信回路の同期動作を行う機能（CDR：Clock and Data Recovery）を持つ。

32 Gbpsなど，将来の更なるデータレートの高速化に向けては，送受信回路の高速化だけが課題ではない。プリント基板の電気配線などの伝送線路起因の信号品質劣化が顕著となるため，それを補償する損失補償回路の高性能化が必要不可欠である。また，受信回路のタイミング同期の精度も高速化に比例して高める必要がある。例えば32 Gbpsのインタコネクトを実現するためには，1 ps（10^{-12}秒）以下の調整精度が必要とされる。

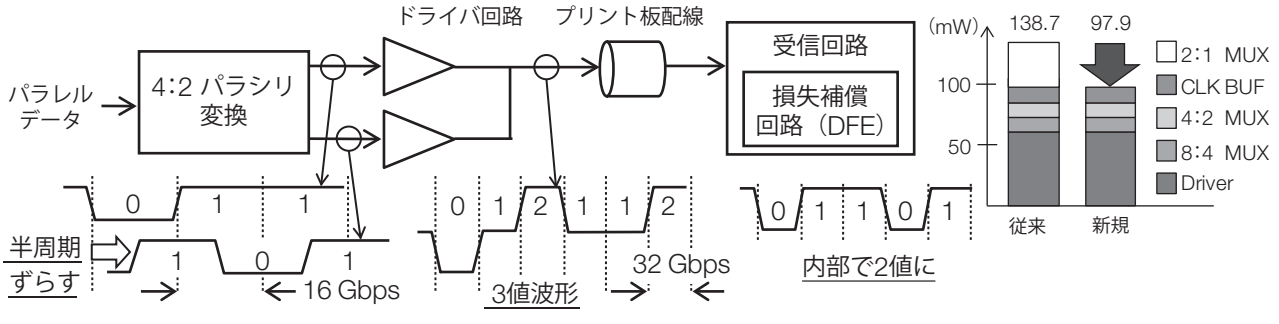
本稿では，次世代プロセッサ間通信において，32 Gbpsを実現する三つのCMOS高速インタコネクト技術について紹介する。

インタリーブドライバ回路

送信回路は，CPUなどから出力される複数の低速なパラレルデータを一つの高速なシリアル信号に変換を行った後，伝送線路用のドライバ回路を用いて，受信回路に対して出力を行う。32 Gbpsのシリアルデータを生成するには，最終段の2:1パラシリ変換回路は，16 GHzの高速クロックにより動作し，送信回路の中で最も高速動作が必要な回路である。このため，2:1パラシリ変換回路の回路実現が非常に難しく，送信回路全体の高速化のボトルネックとなっている。更には，回路は高速で動作するほど消費電力が大きくなるため，2:1パラシリ変換回路の消費電力は送信回路の中で大きな割合を占めることとなる。

そこで，既存の送信回路から2:1パラシリ変換回路を排し，4:2パラシリ変換回路の半周期ずれた二つの出力信号を，ドライバ回路で電圧方向に加算するインタリーブ方式を考案した（**図-2**）。[2] 従来方式は時間方向への信号圧縮であるが，インタリーブ方式は電圧（縦）方向の信号圧縮である。このような信号圧縮方式は，一般に振幅変調（AM：Amplitude Modulation）と呼ばれる。

従来方式では，2:1パラシリ変換回路の入力信号を生成する4:2パラシリ変換回路はタイミングを合わせた2信号を出力するが，本方式は，4:2変換回路出力で半周期ずらした2信号を生成し，それらを入力とするそれぞれのドライバ回路の出力を合成



DFE：Decision Feedback Equalizer

図-2　インタリーブドライバ回路

した信号をチップ外部に出力する通信方式である。この出力信号は2信号が電圧方向に合成されるため，3値信号として伝送線路を伝搬し受信回路に到達する。対向する受信回路は，信号損失を補償する機能として，一般的にDFE（Decision Feedback Equalizer）を搭載する。送信された3値信号は，隣接する1/0信号が1:1の同じ電力割合で相互干渉した信号であるため，伝送線路で発生する信号損失によるISIと同義とみなすことが可能である。もともとDFEは伝送線路のロス起因のISI除去を行うことが主機能であるが，この3値信号を符号間干渉とみなすことにより，信号から重畳する隣接信号成分を除去し，元の2値信号を復元することも同様に可能である。

本技術により，高速動作が必要な2:1変換回路を不要とする新規送信回路方式を実現し，32 Gbpsの高速動作と従来比電力70％を達成した。

データ補間回路

受信回路のCDRとは，入力信号の時間方向のタイミング誤差を検出し，その誤差を補正して，正しく入力信号の1/0判定を行う機能である。送信回路と受信回路の間には，微小な周波数誤差，および周波数揺らぎが存在するため，1/0信号の判定を行うための位相が連続的に変化する。このため，CDR機能の搭載は受信回路にとって不可欠である。従来方式では，1/0判定を行った入力データを用いて，タイミング誤差を検出し，入力信号をサンプリングするクロックのタイミングを自動的に調整する。タイミング誤差の検出は，連続する二つのデータ中心（D[n]，D[n＋1]）と，その二つのデータ遷移中（B[n]）の論理値を比較することで行われる。例えば，D[n]＝0，D[n＋1]＝1でB[n]＝0の場合，サンプルクロックが入力データに対して早いと判断し，サンプルクロックの位相を遅くなるように調整する。1/0判定はこのように位相調整されたサンプルクロックを用いて，データ中心の信号振幅をサンプリングして行うのが一般的である。しかし，バンド幅要求の拡大に伴い1ビットあたりの時間が世代ごとに1/2に短くなっており，調整回路の高精度化が課題となっている。例えば，32 Gbpsの1/0の1ビットの時間は31.25 psであり，サンプルクロックの調整精度は，1ビットあたり64分割の分解能（0.5 ps以下）が要求される。サンプルクロック調整回路は，高速動作かつ高精度のタイミング調整機能の実現が必要となり，受信回路全体の高速化のボトルネックとなる。

今回開発したデータ補間回路を用いたCDR[(3)]は，サンプルクロックの位相調整機能を排し，送信回路とは非同期となるサンプルクロックで動作する（**図-3**）。データ補間回路は，データサンプラ（gm）と可変容量を用いたスイッチドキャパシタから構成される。データサンプラは送信側と非同期に駆動され，サンプリングした電圧を可変容量の比率を調整したスイッチドキャパシタを用いて，データ中心の電圧値をチャージシェアリングの補間により復元する。チャージシェアリングとは，複数の容量間でそれぞれに蓄積された電荷Qを，複数の容量の総容量によって共有することである。例えば，二つの個別の容量（C_1，C_2）のそれぞれにおいて，電圧（V_1，V_2）が印可されている場合，蓄積されている電荷（Q_1，Q_2）は$Q_1=C_1V_1$と$Q_2=C_2V_2$と表される。この二つの容量を接続すると，総電荷量$Q=Q_1+Q_2$，総容量はC_1+C_2となり，接続後新たに発生する容量の電圧Vは，$V=(C_1V_1+C_2V_2)/(C_1+C_2)$となる。チャージ

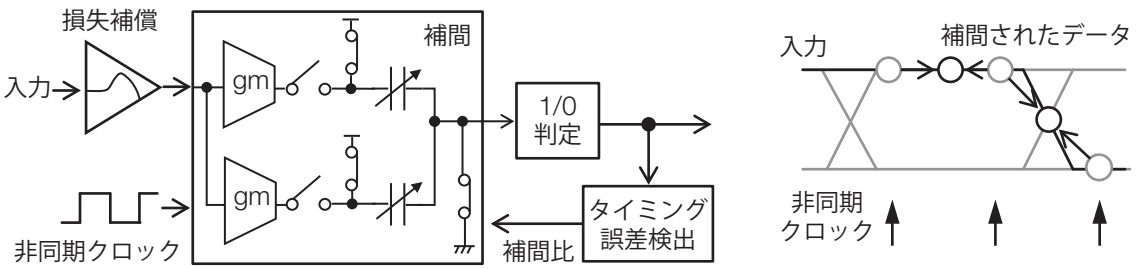


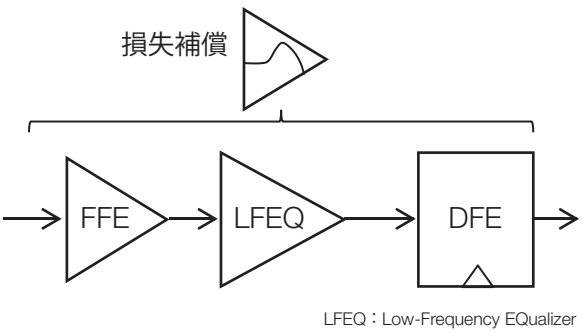
図-3　データ補間回路

シェアリングを行う際の容量比は，サンプルクロックタイミング調整と同様にデータ中心と遷移の情報から検出することが可能である。スイッチドキャパシタによる電圧補間に必要な分解能は1ビットあたり64階調であり，クロックタイミングの調整方式の精度と同等であるが，電圧値に変換すると，その分解能は3 mV程度である。この分解能は，類似した技術を用いるアナログ－デジタル変換器（ADC）の技術分野においては，とりたてて高分解能ではなく，電圧方向の高精度化を実現することは容易である。

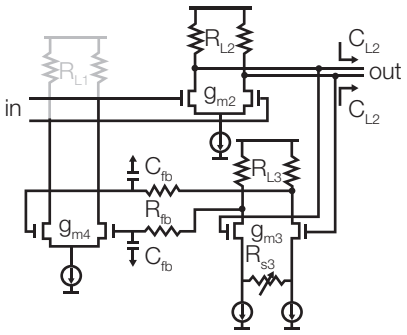
本技術により，受信回路では高速，高精度クロック生成を行う必要がなくなり，32 Gbps以上の更なるデータレートの高速化が可能となる。

広帯域ロス補償

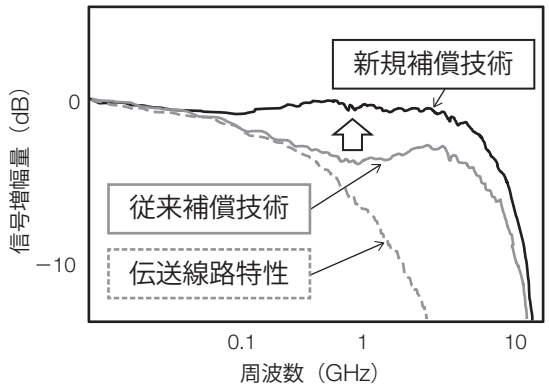
伝送線路で発生するISIを除去し，受信に必要な信号品質を再生するために，受信回路ではFFE（Feed Forward Equalizer）や，DFEといった等化回路が搭載される。伝送線路損失は高周波になるほど増大するため，信号の高速化に対して大きな課題となる。更に送受信信号の高速化により，全体の周波数が上がったため，これまで低周波で損失が小さいと考えていた信号成分の損失を無視することが困難になってきた。**図-4**（c）グレーの線は，32 Gbpsを想定したときの従来補償技術を用いた場合の信号増幅量を表している。信号増幅量は0のとき，損失が存在せず，信号品質が良く，負の値が大きくなるほど，損失が大きく信号品質が劣化し，データ列の1/0判定が困難になる。解析の結果，所望の高周波の信号成分の復元を行うだけでは損失補償は不十分であり，低周波から1 GHzまでの信号増幅量のロールオフによる歪が受信信号品質の劣化に大きく影響していることが判明した。本送受信方式においては，送信回路から伝送線路を介してFFE，LFEQ（Low-Frequency



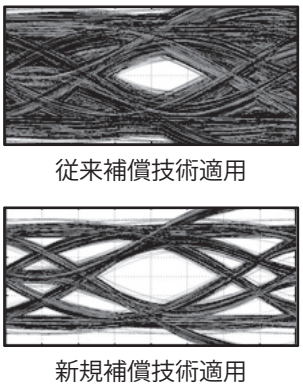
（a）新規ロス補償技術



（b）LFEQ回路図

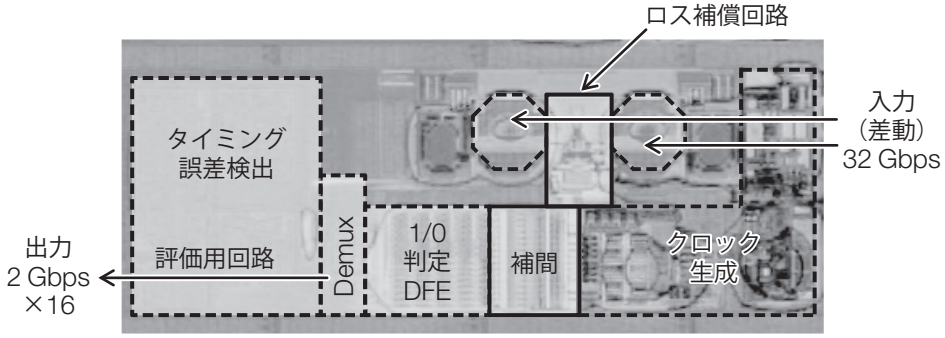


（c）周波数特性

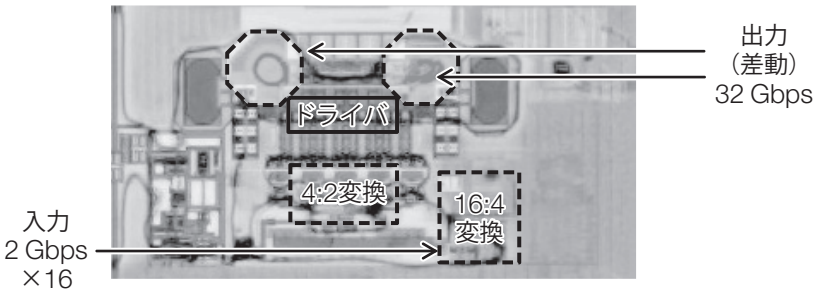


（d）改善効果

図-4　広帯域ロス補償回路



（a）受信回路



（b）送信回路

図-5　チップ写真

Equalizer）までの，DFE手前までの増幅度は通信速度の1/4の周波数まで平らであることが必要である。新規補償技術では，従来の高周波損失補償に加え，LFEQを用いて低周波領域の損失まで補償する技術を開発した｛図-4（a)｝。図-4（b）はLFEQの実回路図である。この技術により低周波から高周波の広い周波数領域にわたり，伝送線路の損失を補償することが可能となった｛(図-4（c）黒の実線｝。[4] 図-4（d）は時間領域のシミュレーションによる従来技術と新規技術の信号品質の比較である。従来技術に対して，信号品質を約2倍改善することが可能となった。

本技術の導入により，32 Gbps実機評価においても，要求される誤り率10^{-12}で，サーバ間プロセッサ通信に必要なバックプレーンの80 cmの高速信号伝送を可能とした。28 nm CMOS 標準プロセスを用いて実装された，三つの高速インタコネクト技術を搭載する送受信回路のチップを**図-5**に示す。

む　す　び

本稿では，サーバ間あるいはCPU間で高速のデータ転送を実現する32 Gbps超高速CMOSインタコネクト技術を紹介した。高周波域と低周波域のロス補償を実現する損失補償回路によって，プリント基板80 cm以上の伝送線路においても大容量のデータ転送を行うことが可能となった。また，インタリーブドライバ回路やデータ補間回路を設けることにより，高速で高精度な要素回路を排除し，低消費電力化や設計性の向上が可能となった。高速インタコネクト技術をCPUに搭載することにより，CPU間通信速度を向上させ，サーバシステムの全体性能を2倍以上向上させることが可能となる。今後，次世代サーバやスーパーコンピュータなどの性能向上に貢献していく。

参考文献

（1）R. Kan et al.：A 10th Generation 16-Core SPARC64 Processor for Mission-Critical UNIX Server. ISSCC Dig. Tech. Paper，p.60-61，Feb. 2013.
（2）Y. Ogata et al.：32 Gb/s 28 nm CMOS Time-Interleaved Transmitter Compatible with NRZ Receiver with DFE. ISSCC Dig. Tech Paper, p.40-41,

Feb. 2013.
（3） Y. Doi et al.：32 Gb/s Data-Interpolator Receiver with 2-Tap DFE in 28 nm CMOS. ISSCC Dig. Tech. Paper, p.36-37, Feb. 2013.
（4） S. Parikh et al.：A 32 Gb/s Wireline Receiver with a Low-Frequency Equalizer, CTLE and 2-Tap DFE in 28 nm CMOS. ISSCC Dig. Tech. Paper, p.28-29, Feb. 2013.

著者紹介

土肥義康（どい　よしやす）
ICTシステム研究所サーバテクノロジ研究部 所属
現在，サーバ向け高速インタコネクト関連の研究に従事。

尾形祐紀（おがた　ゆうき）
ICTシステム研究所サーバテクノロジ研究部 所属
現在，サーバ向け高速インタコネクト関連の研究に従事。

Samir Parikh
米国富士通研究所 所属
現在，サーバ向け高速インタコネクト関連の研究に従事。

小柳洋一（こやなぎ　よういち）
ICTシステム研究所サーバテクノロジ研究部 所属
現在，サーバ向け高速インタコネクト関連の研究に従事。