PC サーバの入出力インタフェース動向

日本電気（株）クライアント・サーバ事業部
阿部 晋樹　s-abe@cd.jp.nec.com
上野 伸二　s-ueno@bc.jp.nec.com

PC サーバのアーキテクチャが，基幹業務サーバやテレコムの市場まで広がりつつある．それに伴い PC サーバの入出力インタフェースに変化が現れてきている☆1．ここ 10 年でマイクロプロセッサの動作周波数は 30 倍以上の高速化を成し遂げてきた．同様に外部の入出力装置であるハードディスクのインタフェースも約 25 倍の高速化，ネットワークについては約 100 倍の高速化を達成してきた．市場の広がりとともに，外部機器の高速化は続き，今後 10 年間で 50 倍以上の高速化が要求される見通しである．一方で，PC サーバの入出力インタフェースは，1993 年頃から採用されてきた PCI バスが継続して利用されているように高速化が足踏みをしている．近年になって，この状況を打破するように，さまざまなアプローチが検討されはじめている．本稿では PC サーバの入出力インタフェースのさまざまな技術動向を紹介し，それらの目指す方向について考察していきたい．

◆次世代入出力インタフェースの必要性

図 -1 は弊社 PC サーバである Express5800 の代表的なシステム構成である．複数個のインテル社「Xeon プロセッサ」を 400MHz のフロントサイド・バスを介して ServerWorks 社「GC-HE」というチップセットに接続し，大容量データを高速に処理している．入出力インタフェースとしては，64bit 幅 /100MHz で動作する PCI-X バスと 32bit 幅 /33MHz で動作する PCI バスをサポートしている．近年の PC サーバは，システム稼働中でも PCI ボードの交換が可能な PCI Hot plug 技術に対応することでシステムの高可用性を実現し，さらに，約 1GByte/ 秒を実現できる PCI-X バスを複数本分搭載することで大規模なシステムに対応している．

この PCI-X の登場によって，Gbit 級のイーサネットやファイバー・チャネルのホスト・バス・アダプタに対応できるようになった．しかし，ネットワーク・スピードはさらに加速し，10Gbit 級のイーサネットやファイバー・チャネル・インタフェースの技術がすでに見えてきている．しかし，PCI バスから PCI-X バスへの進化で採用したような動作周波数を上げバス幅を広げていく高速化アプローチでは，バスを制御する LSI のピン数が増え，ボードの配線レイアウトも厳しくなり実現するコストがかかる．

一方で，サーバ利用の側面からは，インターネット普及により，従来サーバが取り扱ってきたトランザクション処理だけでなく，メディアや音声データなどのストリーミング処理への対応も必要となってきている．さらには，インターネット経由で飛来する予測できない処理が増加するにつれて，サーバの処理能力をスケーラブルに拡張できるような対応も必要となってきている．従来の PCI バスや PCI-X バスでは，単純なバス・プロトコルであったため，画像データのように大量データ転送を行いながら，音声データのように一定のバンド幅を確保していくようにデータの種類に応じたバンド幅の調整やスケーラブルな拡張を容易に実現できなかった．

今後の PC サーバにおいては，これらを低コストに解

☆1 本稿では，サーバ内で入出力ボードを接続する拡張スロットのインタフェース仕様を，PC サーバの入出力インタフェースと狭義に定義する．

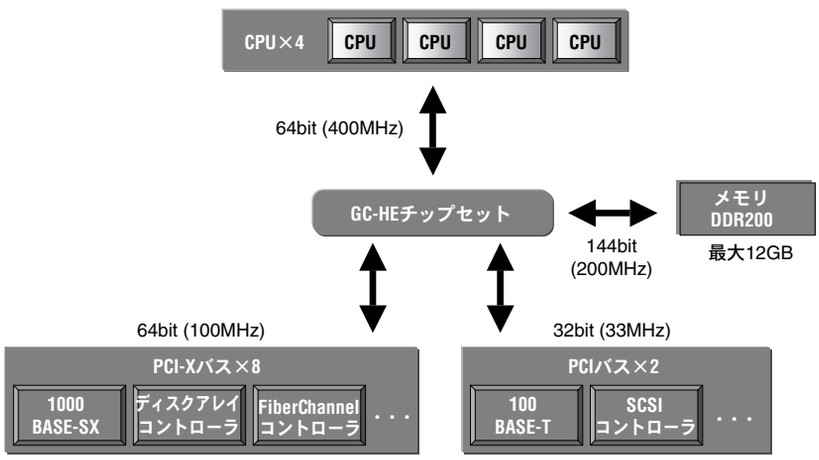

図-1　NEC Express5800 サーバのシステムアーキテクチャ例

名称	接続形態	クロック伝送方式	PCI インフラとの S/W 互換性
PCI32/33	バス接続	単一クロック	あり
PCI-X2.0		ソースシンクロナス	
HyperTransport	PtoP 接続（パケット転送）		
PCI-Express		データにクロックを多重化	
InfiniBand			なし

表-1　次世代入出力インタフェースの特徴区分

決できる入出力インタフェースを備えていく必要があり，さまざまなアプローチが検討されている．PCI バスの標準化団体である PCISIG でまとめている PCI-X2.0 や PCI-Express，米 AMD 社が中心となって提唱してきた HyperTransport，主要なコンピュータベンダが推進してきた InfiniBand などが従来の PCI バスに変わる技術として着目されている．これらの技術の相違点について次の章から取り上げてみたい．

◆次世代入出インタフェースの相違点

すでに使われている PCI バスと，次世代入出力インタフェースの PCI-X2.0 および PCI-Express，HyperTransport，InfiniBand の相違をまずは大きく捉えてみたい．これらのカテゴリ分けの方法は見方によるが，**表-1** に示すように区分できる．まずは，プロトコルの前提となる物理的な接続方法．これは，データ送信者の調停を必要とするバス接続か，全二重ネットワーク技術を取り入れたポイント－ポイント接続かによって 2 種類に区分できる．これは，別な視点で分類すれば，送信先や送受信制御に必要な情報をデータと独立した伝送路として設けるバス接続と，ネットワーク同様にヘッダとしてパケット化するポイント－ポイント接続と区分することもできる．次に，送受信者間のクロックの取り扱い方法で 3 種に区分できる．1 つは送信者と受信者の間で，同一クロックにより制御する方法，もう 1 つは送信するデータと同一線路方向にクロックを送信するソース・シンクロナス技術で，クロックとデータの遅延格差を縮め転送クロックを向上させる方法，残るはデータ送信回路でデータ線路にクロックを埋め込み，受信側でクロックを抽出することで，さらに高速な伝送を実現する方法である．最後に，PCI バスを代替する技術か否かの区分がある．移行期を考慮するならば，代替技術とは PCI バスとの互換性維持の技術と定義できるだろう．ここでは，ハードウェアとしての互換性維持は，すでに伝送方式が異なることから論じる意味がなく，従来の PCI カードで培われてきたソフトウェア・インフラの観点で，その互換性を維持しているか否かの区分になる．

以上の分類で，取り上げた入出力インタフェースは 5 種類に区別され特徴づけられる．PCI および PCI-X2.0 と残り 3 種は，バス接続か全二重のポイント－ポイント接続かで区分できる．ポイント－ポイント接続の方

名称	PCI	PCI-X	PCI-X2.0
動作周波数（MHz）	33/66	66/100/133	266/533
データ幅（bit）	32/64		
バンド幅（Byte/秒）	133M～533M	～1G	～4.2G
IO 信号電圧	5V/3.3V	3.3V	1.5V
スロット数	4スロット @33MHz 2スロット @66MHz	1スロット @133MHz 2スロット @100MHz 4スロット @66MHz	1スロット
トランザクション方式	インターロック	スプリット	

表-2　PCI・PCI-X・PCI-X2.0 の比較

は，一般には動作周波数を容易に上げられるが，バス接続に比べ接続ノード側での制御ハードウェアが複雑になる．クロックの取り扱いに関しては従来の PCI バスを除いてデータとともにクロックを送信する．PCI-Express や InfiniBand はデータにクロックを埋め込むシリアル・インタフェースであるが，他はソース・シンクロナス技術で伝送する．シリアル・インタフェースは，クロックとデータ間の遅延格差の課題が残るソース・シンクロナス転送に比べて一般に動作周波数を上げやすい．また，クロック線を必要としないため，伝送線路の信号数も少ない．このため，PCI-Express や InfiniBand は送受信二線で 2.5Gbps を実現している．第 3 の区分である PCI インフラとのソフトウェア互換性については，InfiniBand とそれ以外で分かれている．InfiniBand では PCI バスの代替というよりは，機器間を接続するようなシステム I/O インタフェースとして使用されることを念頭に置いている．InfiniBand 以外の方式は，コンピュータ内部の LSI 間インタフェースや入出力スロットとして現状の PCI バスの代替になっていくのではないかと予想される．

しかし，特徴の長所・短所から，これらのどれが今後の次世代入出力インタフェースとして主導権を握るのか，あるいは，棲み分けていくかを見通すのは難しい．なぜなら，技術だけでなく推進しているメーカの位置づけや実現時期に大きく左右するからである．次節からは，各々のインタフェースについて，やや詳細な特徴と業界の動向について概説していく．

《PCI バスと PCI-X2.0 の相違点》

PCI-X2.0 は，サーバ向けチップセットを開発している ServerWorks 社をはじめ多くのサーバ向け部品を開発している会社から賛同を得ている．これは従来の PCI バス技術の延長を強く意識したインタフェースだからであり，PCI バス移行のアプローチは LSI を開発・提供する会社にとって導入しやすいからといえる．一方で，PCI-X2.0 は高速化を追求し，バス・プロトコルが，かなり改善されている．**表-2** で，従来の PCI バスと PCI-X，PCI-X2.0 との比較を行った．PCI-X2.0 においては，バス接続を基本とした複数のバス・マスタを許すプロトコルを踏襲するものの，PCI-X（133MHz）と同様に物理接続上は 1 対 1 接続になっている．つまり，送受信の切り替わりがある半二重伝送になる．

従来の PCI バスは読み出し要求（リード）時には，データ応答があるまでバスを解放しないか，繰り返しの読み出し要求を行うプロトコルであった．PCI-X，PCI-X2.0 においてはスプリット・トランザクション方式と呼ばれるバス解放の仕組みを採用した．データ応答を待たずに最大 8 個までの読み出し要求を発行できるため，バスバンド幅の効率的な使用が可能になる．さらに，一度に送受信できるデータ・ブロックの大きさも拡大，電気的特性も大幅に改善させている．その上，従来 PCI バスとの共存ができるようハードウェアおよびソフトウェア互換性を持たせているのが大きな特徴である．PCI-X2.0 は PCI-X に比べてデータ転送レートをソース・シンクロナス技術で 2 倍／4 倍にするとともに，ECC を追加できる機能を追加して，高性能だけでなく，信頼性機能を高めてサーバ向けになっている．現在，2003 年下半期のリリースに向け，PCISIG において仕様を検討中である．PCI-X2.0 を推進しているメーカによれば，ソフトウェア互換性はもちろんのことハードウェア互換性を維持できる入出力インタフェースが重要という．一方で，次節以降で紹介するインタフェースは，従来 PCI バスとはハードウェア互換性を意識していない．むしろ，サーバに限らずネットワーク機器，パソコンやモバイルでも利用できる次世代入出力インタフェースを目指している．

	InfiniBand	PCI-Express	HyperTransport
バス方式	シリアル	シリアル	パラレル
データ線数	2，8，24（差動）	2，4，16，32，64（差動）	2，4，8，16，32
クロック	2.5GHz 埋め込み	2.5GHz 埋め込み	200 ～ 800MHz DDR x 1 ～ 4 ソース・シンクロナス
データ・バンド数	2Gbps ～ 24Gbps	2Gbps ～ 64Gbps	800Mbps ～ 51.2Gbps
トポロジ	・Point-to-Point ・Star（Switch）	・Point-to-Point ・Star（Switch）	・Point-to-Point ・Daisy Chain ・Star（Switch）
適用領域	システム I/O インタフェース	LSI 間接続，外部拡張カード	LSI 間接続
接続メディア	銅線，光ファイバ・ケーブル	プリント配線基板，ケーブルの予定あり	プリント配線基板
PCI 互換性	非互換	ソフトウェア互換	ソフトウェア互換

＊三者はいずれも全二重のインタフェースであり，表ではいずれも片方向のみの値を示している．

表 -3　InfiniBand/PCI-Express/HyperTransport の比較

《PCI バス·ハードウェア非互換の次世代入出力インタフェース》

InfiniBand および PCI-Express，HyperTransport の特徴を**表 -3** で比較する．三者ともパケット化によるポイント－ポイント接続のインタフェースであることが大きな特徴である．ここでは 3 方式がどのような意図で開発され，またこの先どこへ向かっていくのかを概説する．

サーバ市場においては，限界の見えている PCI バスをより高速な入出力インタフェースへ移行させようとする動きはかなり以前から存在しており，1998 年には Intel が中心となって NGIO（Next Generation I/O）と呼ばれるポイント－ポイントのシリアル・インタフェースの I/O 技術が開発された．またその一方で IBM などが中心となり同じくシリアル・インタフェースの Future I/O とよばれるインタフェースも開発されていた．両方式とも PCI バス欠点の完全な克服を謳い，理想的な入出力インタフェースを目指したこともあり，結果として PCI バスとはハードウェアはもちろん，ソフトウェアに関してもまったく互換性のないものであった．最終的に両者は業界の声に押され 1999 年に InfiniBand として統合された．本来 PCI バスの置換を狙っていた InfiniBand ではあったが，統合されたことにより，さらに入出力インタフェースの理想像を追う形となり，バーチャル·チャネルによる QoS 制御，アドレスの IPv6 採用，独自のマネージメント方式など，広範囲にわたる検討がなされ，多くの機能追加が行われた．ハードウェアおよびソフトウェアによるサポートが必要ではあるが，イーサネット，SCSI，ファイバー・チャネル，ミリネットなど転送パケットはすべて InfiniBand パケットのペイロード・データとして取り扱うことが可能であるといえ，その適用範囲は非常に広い．

InfiniBand 発足当初は PCI バスの置換という入出力インタフェースのパラダイム・シフトが期待されたが，PCI バスとのソフトウェア非互換であることから，現在はコンピュータ間を接続するインタフェースという位置に落ち着いたように見受けられる．

2000 年には仕様書 1 版がリリースされ，InfiniBand 技術を搭載した機器もすでに市場に登場している．また PCI-X バスを介した InfiniBand ホスト・バス・アダプタの他，PCI-Express や HyperTransport を介した InfiniBand ブリッジなどが計画されており，サーバ用システム I/O ネットワークのインタフェースとしての利用も期待されている．

PCI-Express は 2002 年に PCISIG により仕様化されたポイント－ポイント接続のシリアル・インタフェースであり，3GIO（3rd Generation I/O）の名前で Intel が中心となって開発が行われた．PCI-Express はコンピュータ内部の LSI 間接続を念頭に開発され，また，PCI バスの置換を強く意識しており PCI バスとのソフトウェア互換を保っている．さらに，PCI-Express 対応ソフトウェアにより PCI バスを超えるさまざまな機能を使用することが可能であり，将来的にはソフトウェアに関しても PCI バス互換から PCI-Express 固有の使用環境へ徐々に移行させようとする意図が伺える．PCI-Express 固有機能としては，他のコンピュータ間同士の接続を可能とするプロトコルや，バーチャル・チャネルを用いた QoS によるバンド幅制御など，いくつか InfiniBand との類似点が見られる．ただし，両者のターゲットの違いは明確であり InfiniBand はサーバによるシステム I/O ネットワークとして光ケーブルなどを介した遠距離間の接続

も視野に入れており，一方 PCI-Express は PC サーバからモバイル PC までをも含む汎用的な LSI 間接続インタフェース，または PCI バスや AGP バスの代替となる外部拡張カードのインタフェースとして位置付けられる．PCI-Express はボードの物理形状に関しても既存 PCI を意識したものとなっており，カードのサイズおよびブラケットの仕様はほぼ PCI カードと同一である．

PCI-Express は仕様が策定されたばかりであり，実際に搭載製品が市場に登場するのは 2004 年頃と思われる．この際，CPU/ メモリ・サブシステムと I/O サブシステム間の接続は PCI-Express が採用され，PCI バスは PCI-Express を PCI または PCI-X バスへ変換するブリッジによって提供されることになるであろう．デスクトップ PC においては，過去に ISA バス・スロットと PCI バス･スロットが共存し徐々に ISA が姿を消したように，いずれはすべて PCI-Express スロットのみに移行していくかもしれない．しかしサーバにおいては，PCI-X2.0 採用の選択肢もあり今後の動向を注目したい．

HyperTransport は 1997 年から AMD によって開発が進められていた入出力インタフェース技術である．2001 年には非営利組織 HyperTransport Technology Consortium が設立，仕様の普及が図られ，デスクトップ PC やゲーム機，組み込み機器においてすでに市場に多くの搭載製品が出回っている．

HyperTransport は前二者同様パケット化によるポイント－ポイント接続であることに変わりないが，シリアルではなく，ソース・シンクロナスを採用している．同じソース･シンクロナスを採用する PCI-X2.0 との大きな違いは，データ線を受信専用，送信専用に分離し全二重とし，高速転送を可能とするためにデータ線のピン数を減らし，8 本のデータ線あたり 1 本のクロックとしたことである．

複数のデータ線を持つバスを高速化すると各データ線間の時間的なずれ（スキュー）が問題となる．高速になればなるほど，受信側での各データ線に対するスキューの許容量が小さくなり，これを押さえ込むために高度な技術が必要となり結果的にコストアップに繋がる．InfiniBand や PCI-Express ではシリアル・バスを採用しクロックをデータに埋め込んで送信することによりこの問題を回避している．ただしシリアル・バスで高いバンド幅を達成するためには高い周波数で送信を行う必要があり，PCI-Express では 2.5GHz のクロックをデータに埋め込んでいる．これに対し HyperTransport では PCI-Express より遅い 200 ～ 800MHz のクロックを使用しているが，複数のデータ線を有するパラレルバスを用いることにより高バンド幅を得る方式を採っている．また 1 本のクロック線に対しデータ線を 8 本までとし，スキュー問題の緩和を図っている．プリント配線基板への実装は，PCI-Express のような高速シリアル方式より，HyperTransport のようなバス方式の方が多くの実績があり，またボードへの実装が容易でもあることから一部領域で普及が進んでいる要因ともなっている．

HyperTransport はあくまでもコンピュータ内部の LSI 間の接続であり，仕様上デージー･チェーンをサポートすることから，外部拡張カードなどのフォームファクタは規定されていない．また I/O の入出力インタフェースへの適用にとどまらず，今年登場する AMD の CPU Opteron では CPU 間のインタフェースとしても利用される予定である．

一方で，今日までネットワーク機器や組み込み系デバイス業界では PCI を共通バスとして使用してきたが，その代替として HyperTransport が採用され始めている．PC 業界だけでなくネットワーク業界，組み込み系デバイス業界からある程度支持されていることが伺え，PCI-Express とは違った利用範囲で棲み分けが行われている．

以上のように，本稿では PC サーバの入出力インタフェースの相違点を挙げながら次世代入出力インタフェースの動向を簡単に紹介してきた．大きな流れは，パケット化などのネットワークの技術を取り込んできていることである．それにつれ，バス技術の特徴であった簡素なプロトコルを捨て，ネットワーク制御のような複雑なプロトコルへと推移しつつある．今後の半導体の急速な発展によって，プロトコル複雑化に伴うハードウェア規模へのインパクトは解消されてくるだろう．一方で，プロトコルの複雑化に伴い，メーカ間での相互接続評価が従来にまして重要となるであろう．よって，多くの入出力機器メーカの賛同を得るとともに相互接続評価方法も確立した技術が PC サーバの次世代入出力インタフェースとして主導権を握ることになると思われる．

参考文献
1）InfiniBand Trade Association Home Page, http://www.infinibandta.org/
2）PCI-SIG PCI Express Specification Page, http://www.pcisig.com/specifications/pciexpress
3）HyperTransport Consortium White Paper page, http://www.hypertransport.org/tech_whitepapers.html
（平成 15 年 1 月 12 日受付）