特集

HDDのインタフェース技術

HDD Interface Technology

あらまし

ハードディスクドライブ（HDD）のインタフェースは，システムの高速転送要求に応えるためシリアル化されている。その主要なものとしてはファイバチャネル（FC），シリアルATA（SATA），Serial Attached SCSI（SAS）がある。これまでHDDの市場セグメントと使用されるシリアルインタフェースは対応していた。最近はコスト低減の要求から使用するシリアルインタフェースが選択されている。その結果，HDDの市場セグメントと従来のシリアルインタフェースとの対応も崩れてきた。例えばエンタプライズのHDDとして，従来ローエンド用であったSATAと大容量のHDD本体を組み合わせた製品が登場している。

本稿では，これらシリアルインタフェースの特徴と，インタフェースのシリアル化に対応したHDDのコントローラ部分におけるインタフェース技術の概要を説明し，今後の課題について論ずる。

Abstract

To enable faster data transfer, the interface of hard disk drives (HDDs) has been changed from a parallel interface to a serial interface. Currently, the main interfaces being used are the Fibre Channel (FC), Serial ATA (SATA), and Serial Attached SCSI (SAS) interfaces. Until now, the HDD market segments corresponded to the serial interfaces that were available. However, because of the recent demands for cost reduction, interfaces are now being used for different segments. As a result, the HDD market segments have become fractionalized, and the correspondence between market segments and serial interfaces has been lost. For example, some enterprise HDDs now use a conventional low-end SATA. This paper describes the features of these serial interfaces and the interface technology of the HDD controller that enables serialization of the interfaces. It also discusses the technological problems of future HDD interfaces and their possible solutions.

河本正和（かわもと　まさかず）

テクノロジ開発統括部LSI開発部
所属
現在，先行技術調査に従事。

ま え が き

コンピュータシステムにおけるあらゆるインタフェースはパラレル転送からシリアル転送へと変化している。この変化の最大の要因はインタフェースの転送を高速化する要求である。ハードディスクドライブ（以下，HDD）のインタフェースも例外ではない。HDDのシリアルインタフェースとしては1990年代後半に1 Gbps（10ビット＝1バイト，100 Mバイト/秒）の転送速度のファイバチャネル（FC）[1]が実用化された。その後2003年には1.5 GbpsのシリアルATA（SATA）[2]が，2004年には3 GbpsのSerial Attached SCSI（SAS）[3]が製品化されている。富士通はFC HDDを1998年に製品化し，SATAやSAS HDDは業界全体の中で早い時期から製品を開発しお客様に提供している（**表-1**）。

本稿ではこれらシリアルインタフェースの特徴と，インタフェースのシリアル化に対応したHDDのコントローラのインタフェース技術を説明し，今後の課題について論ずる。

HDDに使用されているインタフェース

HDDの使用場所や用途は多岐にわたり，それぞれに適したインタフェースが開発されてきた。これまでにデスクトップPCの内蔵HDD用としてATA（AT Attachment）インタフェース[4]が，サーバや大型ストレージシステム用としてSCSI（Small Computer System Interface）インタフェース[5]が開発され使用されてきた。今これらのパラレルインタフェースは，高速化の要求に応えるために，SCSIはFCやSASに，ATAはSATAに置き換えられつつある。さらにHDDの用途そのものも拡大しているため，新しい用途に適したインタフェースも開発されている。主なHDDインタフェースを表-1に示す。また現在開発中のCE-ATA[6]を含む4種類のシリアルインタフェースについてその特徴を以下に記述する。

● FC

FCは高速転送というデータチャネルの特性と接続の多様性というネットワークの特徴を併せ持つインタフェースである。リンクの転送速度としては1 Gbpsで始まり2 Gbpsのあと現在4 Gbpsが製品化され，さらに8 Gbpsが検討されている。このリンクは同時双方向通信であり，送信と受信が独立して動作する。FC装置の接続形態（トポロジ）を**図-1**に示す。

トポロジは，1対1，ファブリックを中心にしたスター，ループなどを基本とし，それらを組み合わせることができる。メッシュは複数のルートがあるので負荷分散や障害回避などが可能となる。

一般にシリアルインタフェースのプロトコルは階層構造になっている。これについては**表-2**を参照願いたい。

FCは最上位層のユーザプロトコル層でSCSIコマ

表-1　HDDインタフェース

インタフェース名	ATA（IDE）	SATA	SATA-2	SCSI	SAS	FC
コマンド	ATA	ATA	ATA/ATAPI	SCSI	SCSI	SCSI
量産時期	1980	2002	2005	1986	2005	1996
データ転送速度	66，100 Mバイト/秒	1.5 Gbps	1.5，3 Gbps	320 Mバイト/秒	3，6 Gbps	1，2，4，8 Gbps
ケーブル長	0.5 m	1 m	7 m	25/12 m	0.5 m/10 m	30 m/10 km
コネクタピン数	40 pin	7＋15 pin	7＋15 pin	80 pin	7＋15 pin	40 pin
装置接続台数	2	1	1	16	16 256	126
トポロジ	ストリング	1対1	1対1	ストリング	スター	ループ（FC-AL）
複数イニシエータ	不可	不可	不可	可	可	可
アービトレーション	なし	なし	なし	あり	あり	あり
ポート障害の影響	ほかの装置に影響あり	影響なし	影響なし	ほかの装置に影響あり	影響なし（EXP）	影響あり 影響なし（S/W）
インタフェースの特性と適用システム	筐体内 PC/家電製品	筐体内 PC/家電製品 ATA置換え	筐体内 PC/階層型ストレージ	筐体内および筐体外 サーバ/RAID装置	筐体内/外 サーバ/RAID装置 SCSI置換え	筐体内/外 JBOD/SBOD構成 大容量RAID装置

ンドプロトコルやIPなど複数の種類のプロトコルを単一のFCネットワークでサポートすることができる。HDD接続ではSCSIコマンドを使用する。このようにFCは多機能であるが，そのためにFCインタフェースは複雑にならざるを得ない。これを軽減する目的で，HDD接続にFCを使用する場合，FCの機能のうちの一部を選んで使用している。この選択された機能仕様のことをプロファイルと呼んでいる。代表的なものとしてPLDA[7]やFLA[8]がある。

FCは高性能なストレージシステムで多数のHDDを接続するために使用される。このようなシステムでは同時に高信頼性が要求されるため，FCインタフェースを持つHDDの本体（媒体やヘッド，それを駆動する機構とその制御回路など）も，高性能・高信頼性のものが使用されてきた。

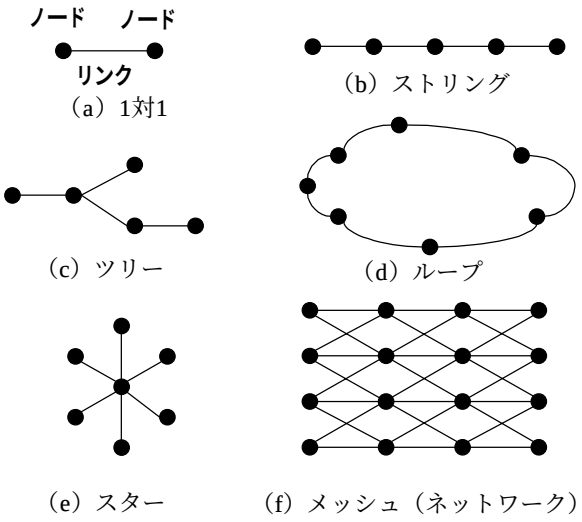

図-1　いろいろなトポロジ
Fig.1-Topology model.

● SAS

SASはシンプルなSATAのリンクとFCに類似した階層制御プロトコルを組み合わせたHDD専用インタフェースである。これについては表-2を参照願いたい。SASのリンクは3 Gbpsの双方向で，トポロジは1対1あるいはエキスパンダを中心にしたスター接続の2種類である。SASのリンクはリンク層のプリミティブの一部をSATAと共用しているのでSATA装置とリンクやエキスパンダを共用できる。ディスク制御にはSCSIコマンドセットを使用する。SASはHDD専用として初めから機能を限定しているので，FCよりもシンプルなインタフェースになっている。SASインタフェースと組み合わせるHDD本体はSCSIと同じ高性能・高信頼のものが使用されている。

● SATA

SATAはパラレルATAインタフェースをシリアル化することのみを目的に開発された。1対1接続のみとしてドライブのアドレス指定機能を持たず，機能を限定したシンプルさが特徴である。1.5 GbpsのSATAリンクはプリミティブ信号によるインターロック制御を行う（後述）ので，フレームの送信と受信を同時に行うことはできない。ディスク制御にはATAコマンドセットを使用する。

このように，本来SATAインタフェースはデスク

表-2　プロトコルの階層構造

インタフェース名および制御階層名	SCSI	FC	SAS	ATA	SATA
コマンド層	SCSI-3コマンド	SCSI-3コマンド	SCSI-3コマンド	ATAコマンド	ATAコマンド
マッピング層	—	SCSIコマンドマッピング	SCSIコマンドマッピング	—	—
プロトコル層	バスフェーズ バスシーケンス バスコンディション	プリミティブ フレーム サービス フロー制御	フレーム	インタフェース動作 レジスタ定義 転送モード定義	フレーム（FIS）
リンク層	信号線 バスタイミング	8B/10B コード	8B/10Bコード OOB信号 プリミティブ アドレスプレーム	信号線 バスタイミング	8B/10Bコード スクランブル プリミティブ
物理層	コネクタ ケーブル 電気的特性	コネクタ ケーブル 電気的特性	コネクタ ケーブル 電気的特性	コネクタ ケーブル 電気的特性	コネクタ ケーブル 電気的特性 OOB信号

トップやノートPCの内蔵HDD用に開発されてきたATAインタフェースを置き換えるものとして開発され，製品化された。しかし最近では後述するようにSATAインタフェースはエンタプライズでも使用され始めている。

● CE-ATA

CE-ATAはシリアルデータとクロックを別々の信号線で送るMMCリンク[9]を用いて1対1接続をする。データと別にクロックを送るので受信側のクロック再生のためのPLL（Phase Locked Loop）回路などが不要となり，その分回路はシンプルになる。ディスク制御にはATAコマンドセットの一部を使用する。

ここまでHDDのインタフェースとHDD本体の用途は性格が対応しているものとして説明してきた。しかし最近になってこの関係を見直す必要がでてきた。

従来，高性能で高信頼を要求されるストレージシステム用のHDDとしては，高性能のFCインタフェースと高性能・高信頼のHDD本体が組み合わされてきた。

お客様による容量増加の要求がますます強まる中，最近のストレージシステムにはビットコストの低減，スペース効率の改善，および消費電力の低減が従来以上に求められている。この要求に応えるため，データの特性に応じHDDを使い分ける方法がとられ始めている。具体例としてミッションクリティカルなデータは高性能のFC HDDに配置することで，FC HDDの処理の速さや24時間連続稼働の高負荷でも安定して動作する高信頼性を実現し，逆に大容量であるがアクセス頻度の低いデータは，少々低速であっても大容量のSATA HDDに配置する。つまり，データの特性に合ったHDDを選択することでストレージシステムの全体最適化を図る方法である。これは性能が異なるHDDで構成されるという意味で階層（Tier）型ストレージと呼ばれている。

上記の場合，SATAとFCはインタフェースの互換性がないので大容量のSATA HDDをストレージシステム内のFCネットワークに接続するにはインタフェースの変換装置が必要になる。この変換装置はコストアップや性能低下の原因となる。そこで変換を不要にするために大容量HDD本体にFCインタフェースを持たせたHDDが開発されている。またHDDのインタフェースはSATAのままでFCフレームにSATAのフレームを埋め込むFC-SATA仕様が検討されている。

こうして本来デスクトップ用のSATAインタフェースがエンタプライズでも使用されるようになり，従来のSATAはデスクトップ用という分類では済まなくなってきた。今後は，HDDのインタフェースについては，インタフェースの特性とHDD本体の特性の両面から考える必要がある。以下，この両者を区別して新しい用途やマーケットの要求に最適なインタフェース，およびそれを実現する上で考慮すべき点と問題となる事柄について論じることにする。

マーケットセグメントとインタフェース

HDDはその使用用途に従ってマーケットセグメントに分類される。このセグメント分類はHDDの備えるべき要件やHDDの性格分類という意味を持っている。HDDは例えば次の六つのセグメントに分類される。各セグメントの主要なHDDのシリアルインタフェースをカッコ内に付記する。

・エンタプライズ（FC）
・サーバ（SAS）
・デスクトップ（SATA）
・モバイル（SATA）
・コンシューマ（据置き）（SATA）
・コンシューマ（モバイル）（CE-ATA）

新しい用途により新しいセグメントが生まれたときはもとより，既存のセグメントであってもこれまでにない新しい使用方法が生まれたときには，従来のインタフェースやHDD本体との整合性を見直さなければならない。先に例を挙げたエンタプライズストレージシステムにおける大容量SATA HDDはその好例である。この例の，低速・大容量のSATA HDDは実はデスクトップで使用されるものと必ずしも同じではない。たとえSATAインタフェースであっても，SATAと組み合わされている大容量のHDD本体はエンタプライズシステムの条件下における信頼性要求を満たすことが必要だからである。

複数のホストシステムが複数のストレージシステムを共用するために，FCのようなネットワーク機能を備えたインタフェースが用いられる。このネットワークはSAN（Storage Area Network）と呼ば

れる。

ストレージシステムの内部構造がRAID（Redundant Array of Inexpensive Disks）ではなくJBOD/SBOD（Just aBunch Of Disks/Switched Bunch Of Disks）であれば，ホストがストレージシステム内のHDDを直接アクセスするので，HDDのインタフェースとしてはストレージシステムの外部インタフェースと同じFCが適している。

逆にRAIDシステムの場合はRAIDコントローラのバッファ上にユーザデータがいったん展開され，RAIDシステム内でデータは再配置される。つまり，ホストとHDDのデータアクセスは独立しており，ホストから直接HDDをアクセスすることはない。したがって，HDDインタフェースはホスト-ストレージシステム間のインタフェースとは関係なく，RAIDコントローラとその配下のHDD群の閉じたネットワークの制御に適したインタフェースであればよい。

このように，ストレージシステムのアーキテクチャによって最適なHDDインタフェースは変わる。

例えばSMB（Small to Medium-sized Business）向けの小型RAIDシステムでは，装置導入時のコストが低く，将来システムを拡張しながら性能を改善できることが要求される。この場合，デスクトップHDDとエンタプライズHDDの混在が必要になり，システム規模がそれほど大きくないことから，サブシステム内のインタフェースとしてはSASを，導入時のドライブとしてはSASで動作できるSATA HDDが使用される。

一方，階層型を含む大型のRAIDシステムでは，要求される性能，容量，HDD数から，内部インタフェースはFCが使用される。階層型では大容量SATAドライブ群は専用のロッカーに収容されインタフェース変換器を介して内部FCネットワークに接続される。

サーバやPCがHDDを内蔵する場合，HDDが接続されるホストは一つであり，複雑なネットワーク機能は不要である。このようなシステムでは従来から使用されているドライブのソフトの資産継承という観点からインタフェースは決められる。SCSI HDDが使用されていた場合はSAS HDDが，またATA HDDを使用していた場合はSATA HDDが使用される。

シリアルインタフェースの特徴

本章では，シリアルインタフェースの特徴について，とくにHDDに使用する場合に考慮すべき点と問題となる事柄について論ずる。

● ケーブル・信号・接続形態

HDDのシリアルインタフェースは電気信号のケーブルやコネクタにより信号伝送路（リンク）を形成する。

このリンクと装置（ノード）の様々な接続形態をトポロジと呼んでいる。これには図-1で示すような種類がある。またどのトポロジが各シリアルインタフェースで使用されるかについては表-1に示す。

実際のシリアルリンクは送信機と受信機がケーブルを挟んで対向している。

(1) 1対1接続では，自ノードの送信機が相手ノードの受信機とつながり，相手ノードの送信機が自ノードの受信機につながっている。こうして双方向の信号送受を行う。

(2) ループ接続では，三つ以上のノードがリンクによってループ状に結合される。各ノードに着目すれば，一つのノードの受信機はデータを送ってくる上流のノードの送信機と接続し，送信機は下流のノードの受信機と接続している。つまり一つのノードの送信機と受信機はそれぞれ別のノードと対向している。ノードの内部には受信機で受けたデータを送信機に送る論理回路があって，データをノードからノードへ次々にリレーしていく構造となる。ノード内の論理回路で結合されている送信機と受信機の1組をポートと呼ぶ。ノードには一つ以上のポートがあり得る。

(3) スター接続では，スターの中心となるノードには複数のポートがあり，それぞれのポートがほかのノードのポートとつながる。中心となるノードのポートは中心ノード内にある切換え回路に接続され，ほかのポートとフレームを交換する。切換え回路はポートとポートの接続ルートを維持するルートスイッチ型やフレームを任意のポート間で随時交換するフレームスイッチ型などがある。

(4) ストリングやツリーおよびメッシュはスターの組合せと考えることができる。

リンク上の信号レベルやデータ符号化方法の規約をリンク層と呼ぶ。シリアルインタフェースは1組のペア信号線で送受するビット列で，クロックやデータおよび制御信号などのすべての情報を伝達する。これは特殊なビットパターンやその配列に制御上の意味を持たせることで実現している。FC・SAS・SATAのいずれも8B/10Bコード(10)を使用している。

● **データ転送の速度限界**

SCSIやATAなどパラレルインタフェースは複数のデータとクロック，またはデータストローブ信号を並列に転送する。このような並列転送において，信号間のスキュー（信号線間のタイミングのずれ）やクロストークなどの影響によりパラレルインタフェースの高速化に限界が生じる。実際ATAインタフェースでは100 Mバイト/秒，SCSIインタフェースでは320 Mバイト/秒で開発は停止している。シリアル伝送ではスキューは原理的に発生しないので，クロック周波数を上げることにより高速化を図ることができる。

● **シリアルの高速化の技術**

シリアルインタフェースのデータ転送速度を上げるためにクロックを高速化する場合，実用化の限界を決定しているのはトランシーバやその周辺回路の動作周波数とリンクの特性，とくに周波数による減衰特性とその補正技術の有効性である。前者は送信機を内蔵するLSIの動作周波数とテクノロジ，つまりCMOSの場合なら設計ルールによる素子のチャネルのサイズに依存する。HDDでは上記のインタフェースの技術の向上と，媒体上のデータの書込み・読出し周波数の向上に相まってインタフェースのデータ転送速度が高速化されてきた。この遷移について，HDDインタフェースのロードマップを**図**-2に示す。

● **シリアルインタフェースの信号品質**

HDDのシリアル信号線は，ケーブルとしては光ファイバではなく2本の銅線をより合わせたものを使用する。これは差動信号を伝播（ぱ）するのに適している。またHDDと結合するコネクタをプリント板に直に取り付けたものを使用する場合，プリントパターンによるマイクロストリップラインで伝送路を形成する。

伝送路の品質として10の12乗ビットに1ビットの誤り率（BER：Bit Error Rate）がすべてのシリア

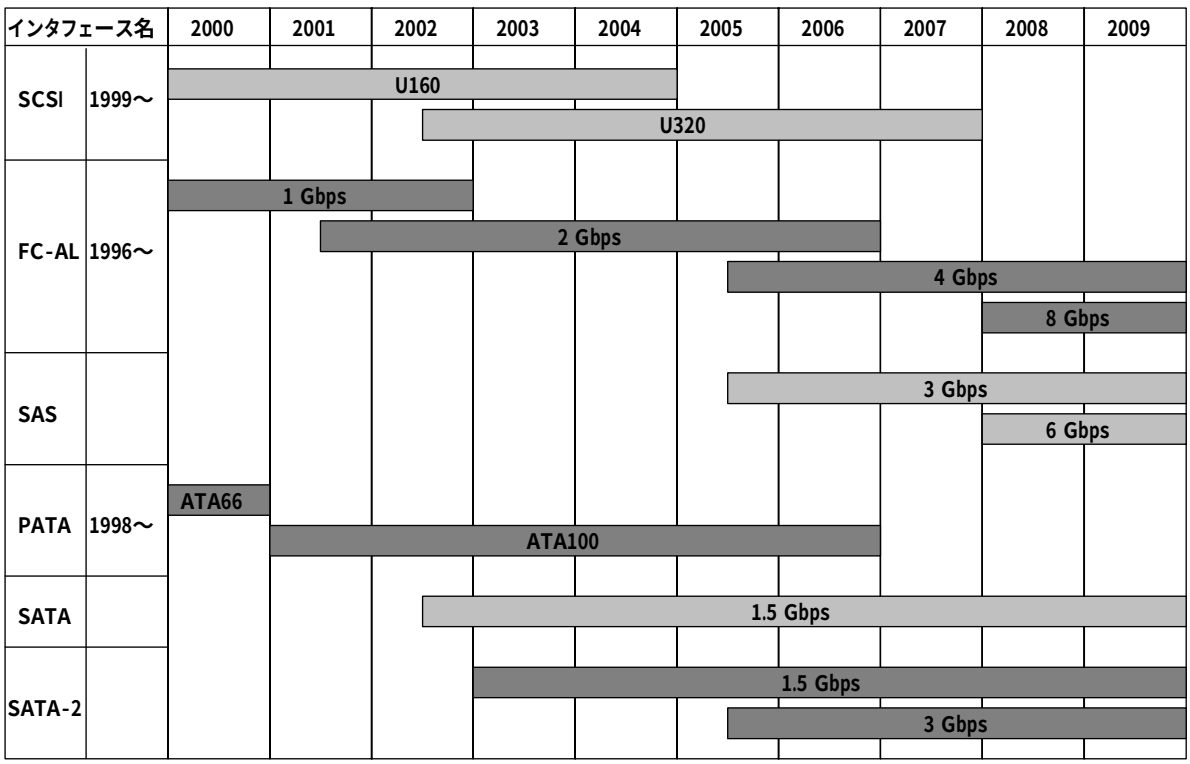

図-2　HDDインタフェースのロードマップ
Fig.2-Roadmap of HDD interface.

ルインタフェースで規定されている。実際の製品ではこれよりはるかに低いエラー率（1/100-1/1000）が要求される。このBERはジッタと密接な関係がある[11]

ジッタとは信号の変移が本来起こるべき時間から前後にずれる現象である。この原因は様々あり，合成されて現れる。原因の代表的なものとしては，回路の熱雑音や伝送路の反射，回路外からの電気ノイズなどがある。これらは物理的に避けられないものや部品の特性や設計の巧拙によりある程度避けることができるものもある。また，発生してしまったジッタに対してレシーバがどこまでエラーしないかという，レシーバのジッタ耐性は信号路における品質の劣化を補うので，伝送系全体の能力を決める一要素と言える。

● インタフェースの互換性

一般に，インタフェースで装置を接続するときに，それぞれの装置の設計者によるインタフェース規約の解釈，あるいは規約に記述されていない条件における動作が一致していなければ処理に矛盾が発生し，連結動作に悪影響を及ぼす。この矛盾の存在は実際に装置を接続して検証される。装置間に処理の矛盾がなければ，接続動作が可能になる。これを相互運用性（Interoperability）という。FCやSASについてはUNH（University of New Hampshire）のIOL（InterOperability Laboratory）などの機関が接続試験を実施したり，各メーカが装置を持ち寄って接続試験を行うプラグフェスタ（Plugfest）と呼ばれるイベントで，接続動作確認を行ったりしている。富士通は早くからこれらに参加し，製品の高い相互運用性を維持している。

● コマンドセット

シリアルインタフェースは表-2に示すように階層制御になっていて，最上層のコマンド層でSCSIやATAコマンドセットを使用する。この階層ではシリアルもパラレルも同じコマンドセットを使用する。FCとSASも物理層からマッピング層までは異なるが同じSCSIコマンド[12]を使用する。この意味でコマンド層を論理インタフェース，マッピング層以下を物理インタフェースと考えることもできる。

システムインタフェースとして開発されたSCSIコマンドセットとPCの内蔵ドライブインタフェースとして開発されたATAコマンドセット[4]にはこの本来の用途から生じた機能上の違いがある。例えばセクタのバイト長がSCSIでは可変仕様であるが，ATAでは512バイトの固定である。このようなコマンドセットの機能の差はHDDの機能の差となっている。

FCやSASと，SATA HDDを混在して使用する場合，SCSIとATAの機能差をどのように吸収するかが問題になる。これまでは個々に対応していたが，現在NCITS T10/T13で変換仕様を標準化する作業が行われている[13]

● データ保全

ストレージシステムのデータ保全とは，最終的にホストのメモリ上にあるユーザデータを誤りなくHDDの媒体上に記録したり読み出したりすることである。ここではインタフェースとHDDドライブ内のデータ保全について，シリアルインタフェースではどのように強化されているかについて説明する。

シリアルインタフェースでは8ビットのデータを10ビットコード化して送る。ノイズなどによりこのコード内のビットが変化すると高い確率で無効なコードになるのでこれをチェックする。フレーム内のコードの場合は無効コードのチェックに加えてフレームに付加されたCRC（Cyclic Redundancy Check）によりデータの正当性チェックを行う。またフレームのヘッダ部分には各フレームに付加したシーケンシャルな番号が埋め込んであるのでこれをチェックしてフレームの消失を検出している。

HDC（Hard Disk Controller）内部のユーザデータの移動については隙間なくパリティビットやCRCなどによるチェックを行っている。また媒体上のデータにはLBA（Logical Block Address）を含んだECC（Error Correction Code）を付加してデータの誤りを検出および訂正することにより，目的のセクタであることを確認している。これらのチェックはすべてハードウェアで実施している。

シリアルインタフェースの接続性の機能強化によって1台のRAIDコントローラに接続されるHDD数は増加する。RAIDシステムではユーザデータの分散配置と集合再構成を行うが，装置台数の増加とともにデータの配置は複雑になる。このデータ配置で発生する誤動作を検出するためにデータに付番し，これをストレージシステムの要所要所で確認することにより，配置のミスとともに誤動作位置も確定す

る方法としてEnd to End[14]の仕様の規格化が進められている。富士通はこの規格化に参加している。

● **セキュリティ機能**

シリアルインタフェースがネットワークの特質の一つとして接続の自由度を拡大し，その結果大規模なストレージネットワークが可能になった。これに伴い，接続されている装置に不当なアクセスが行われる可能性がでてきた。そこで，アドレスを基にしたゾーンに装置群を論理的に分割し，アクセスをゾーン内に限定することで不正アクセスを抑止する方法が開発された。FCではスタートポロジの中心に位置するファブリックがこれに接続されるすべての装置のアドレスを管理しているので，ゾーンによる排他制御もファブリックが行う。SASではエキスパンダの物理的ポート位置にアドレスを付与し，これを基準にしたアクセス制限が検討されている。

上記は各インタフェースに固有のアクセス制限であるが，インタフェースの種類に依存しない方法が検討されている。TCG（Trusted Computing Group）[15]が作成しているTrusted Peripheral（TPer）がそれである。これは公開鍵を用いて暗号化した装置固有のIDによって装置の認証を行うことにより，装置接続やアクセスの可否を制御しようとするものである。富士通はこの標準化作業にも参加している。

HDDのコントローラ

ここからは，これまで述べてきたシリアルインタフェースの特徴と問題点に対処するHDDコントローラのインタフェース技術について論ずる。

本章ではHDDコントローラ全体について簡単に説明した後，ホストインタフェース制御部分について，シンプルな機能のSATAと，多くの機能を持ったFCの実際の構造を比較しながら説明する。

HDDのコントローラは，大きく二つの機能に分けられる。一つはホストから送られるコマンドを解読し，要求されたデータを転送したり，状態を報告したりするホストインタフェースの制御である。もう一つはヘッドの位置決めや媒体の回転制御および媒体上のデータフォーマットの制御などのドライブ制御である。

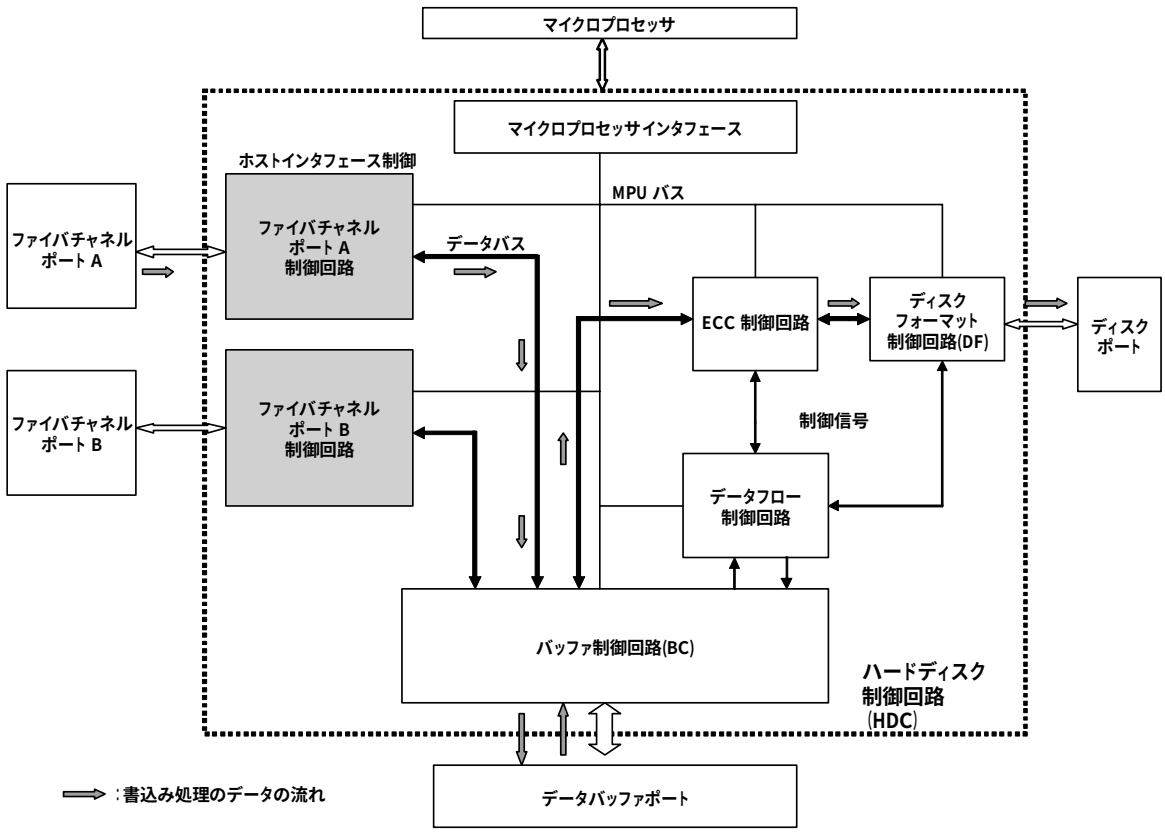

図-3　FCのHDD制御
Fig.3-HDD controller for FC.

● コントローラの構造

FCのHDDコントローラ全体のブロック図を**図**-3に示す。

コントローラの核となるHDC周辺には，コネクタやインピーダンス整合回路から成るポート，制御用のマイクロプロセッサと制御メモリ，数Mバイトから十数Mバイトの容量のデータバッファ，ディスクポート側には，ディスク媒体からのデータの読出し書込みを行うリードチャネル，リード・ライトヘッドの位置決めやスピンドルモータの回転を制御するサーボコントローラ，ディスクの機構などがある。これらは図では一部省略している。

HDC内には，2組のFCホストインタフェース制御回路とデータバッファの制御回路（BC），ECCの生成や検査およびエラー訂正データの生成を行うECCエンジン，媒体上のデータのフォーマットを制御する回路（DF）がある。図中に書込み処理のデータの流れを示す。

ホストからFCインタフェースにより送られたデータはFCインタフェース制御回路を通してデータバッファにいったん書き込まれる。一定の量のデータがデータバッファに蓄えられたらこのデータを読み出し，ECCを生成付加し，ディスクフォーマット制御DFで媒体上のフォーマットに整えてディスクポートに送出する。このあとデータを媒体上に書き込む。読出しはほぼ逆順になる。

ここではFCのHDCを例に挙げたが，ほかのインタフェースであってもコントローラの基本的な構成は同じである。ここでホストインタフェース制御部が2組あるのは，デュアルポートだからである。シングルポートではホストインタフェース制御部分は1組だけである。

● ホストインタフェース制御部

FCホストインタフェース制御部の内部を**図**-4に示す。FCの特徴は受信と送信が完全に独立していることと，両者の間にループトポロジ用の制御回路を持っていることである。

初めに受信側のデータの流れを説明する。受信機（RX）で受け取ったシリアル信号からPLLによりクロックを生成し，このクロックによってデータ列を再生する。データ列の特定のパターンからキャラクタの切れ目を検出してインタフェース上の10ビットキャラクタごとに切り分け，10B/8Bデコードを行って8ビット長のデータコードと制御コードを検

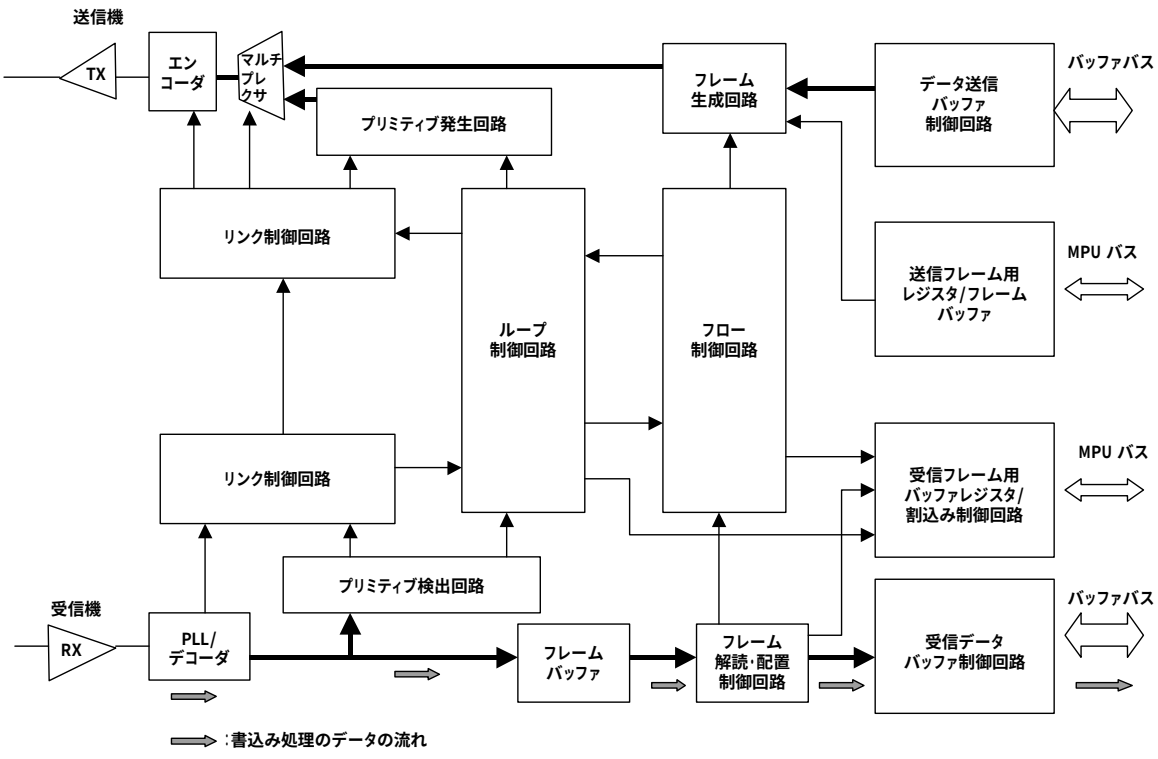

図-4　FCのインタフェース制御回路
Fig.4-Interface control logic for FC.

出する。ここまでがPLL/Decoderによるキャラクタ操作である。つぎに制御コードと3個のデータコード列（4バイト・1ワード）の組合せで表示されるプリミティブを検出する。各プリミティブには固有の意味が付与されているので，データフレームの開始と終了，フレームを受信できる状態にあることなどを検出できる。こうして検出したフレームはフレームバッファに置かれ，フレームデコーダによって正常性のチェックや分類を行う。ユーザデータのフレームはバッファ制御回路を通ってバッファに書き込まれる。

つぎに送信側について説明する。バッファコントロールによりバッファから読み出されたデータは送信フレーム用レジスタに用意されたフレームヘッダ用の制御データと合わせてフレームジェネレータでデータフレームに整えられてエンコーダに送られる。ここで8B/10B変換を施した後，シリアルに配列され送信機（TX）で送出される。

フローコントロールはホストからのフレーム受信可能状態を保持し，許容された数だけのフレームを送出するよう制御する。

リンクコントロールは受信側のプリミティブや信号の状態を監視してリンクの状態制御を行う。送信側のリンクコントロールはフレーム間のギャップに制御用のプリミティブを挿入する。

FCとの比較のためにSATAのホストインタフェース部分を**図**-5に示す。

SATAの受信側と送信側のデータの流れはFCとほぼ同じであるが制御部分は一つしかない。これはSATAの受信と送信が連動しているからである。例えばSATAのフレーム（FIS）を受信している間，送信側はフレーム受信中を表すR_IP（Receive In Progress）プリミティブを送信し続ける。このように送信と受信は一つの処理をプリミティブを確認しあいながら一体となって実行する。このため，プリミティブやフレームを送受する制御回路の状態遷移は1種類となり，回路も一つになる。またフレームの転送もデータをバッファから読み出すか書き込むかのどちらか一つの動作なのでバッファ制御回路も一つになっている。さらに1対1接続のみなので，FCのようなループ制御は持たない。

FCインタフェースの課題

本章では，HDDのコントローラに使用されるFCインタフェースについて技術的課題とその解決方法を述べる。

FCでは，少ない送受信機数で沢山のHDDを接続するために考案されたFC-AL（Fibre Channel Arbitrated Loop）(1)というループトポロジを採用している。このループトポロジを実現するためにはFCポート内部にループトポロジに固有なコントロールを持つ必要がある。これがループコントロールである（図-3）。

● ループコントロール

ループ内の各ポートはループ上のデータが流れてくる側（上流）のポートから送られてくるプリミ

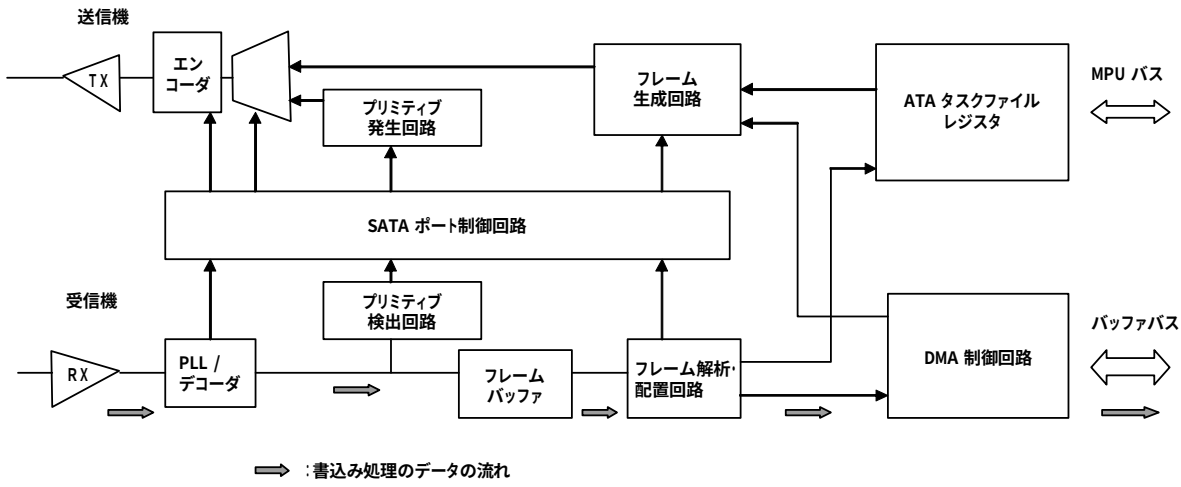

図-5　SATAホストインタフェース制御回路
Fig.5-Host interface control logic for SATA.

ティブをループコントロールが解読し，受け取ったプリミティブをそのまま回送するか，あるいは自身で新たに生成したプリミティブに置き換えるか判断する。そして送るべきプリミティブを下流に送り出す。このようにしてループ全体でプリミティブをリレー転送することでループ上のすべてのポートがループの状態やほかのポートの要求を共通に認識し，連携してループ制御を行っている。

ループ上の各ポートは電源投入後，ループイニシャライズ処理によりループ動作の初期化を行い，このときに各ポートのループアドレスを決定する。つぎにこのループアドレスを埋め込んだプリミティブをほかのポートと交換しながらループを一時的に占有するための競合（Loop Arbitration）の制御を行い，競合に参加したポートの中から一つのポートがループ占有権を得る。占有権を得たポートは相手ポートを選んで1対1結合状態を確立し，結合した2ポート間でフレームやプリミティブの交換を行う。この結合状態が確立している間，ほかのポートはフレームを中継し，受信側と送信側のクロックの速度差を吸収する必要が生じたらフレームのギャップにプリミティブを挿入したり，取り除いたりする。結合処理を終えたポートは結合状態を解除する。

このように，複数のポートが連携して動作するので異なるメーカのHDDが混在するシステムでは動作のタイミングも含め相互運用性が厳密に求められる。FCのプラグフェスタはこのために企画されたと言える。

プリミティブやフレームの流れに着目すると，ループコントロールは上流からのフレームをポートが自身で取り込むかあるいは下流に流すか，さらに自分のフレームを下流に流すかというフレームルータの役割をしていると見ることができる。またループ内の全ポートが連携してループアービトレーションを行うときに，均等サービスを実現していることも特筆されるべきことである。

● バッファの転送能力配分の最適化

以下にFCの場合の，データバッファを使用する回路を列挙し，主要な回路が要求するデータ転送速度の例を示す。

・FCインタフェース425 Mバイト/秒×2（4.25 Gbps×2）
・媒体　リード・ライト100 Mバイト/秒
・ECC生成・訂正100 Mバイト/秒
・MPUアクセス
・RAID用Exclusive OR演算100 Mバイト/秒
・フォーマット制御定数テーブルアクセス
・ダイナミックメモリのリフレッシュ

上記の例から明らかなようにバッファの負荷は最大で1150 Mバイト/秒以上となる。

ところが，バッファの転送能力は，一般にこれよりも小さいのが普通である。例えばFCのコントローラの内部バスに合わせて4バイト幅とし，部品入手の容易性からDDR RAMを150 MHzで動作させるとする。これは瞬間転送で1200 Mバイト/秒の転送能力があるが，アドレスやリード・ライトの切換えで平均速度は800 Mバイト/秒程度になる。これでは二つのFCインタフェースが4 Gbpsで同時に動作するにも不足していることになる。

不足対策として，バッファの転送能力を増大するにはメモリのデータ幅を拡張したり，より高速のメモリを使用したりする方法がある。しかし，コントローラ内のバス幅や動作クロックを変えることは装置全体への影響が極めて大きいので困難である。

そこで，限られたバッファのデータ転送能力をこれらの処理にどのように振り向けるか，調整することによってバッファ転送能力分配の最適化を行う。まず，上記のバッファの使用要求の性格はそれぞれ異なっている。媒体のリード・ライトは途切れることなくサービスしないとリード・ライトの中断による回転待ちが発生して性能が著しく低下する。したがって，無条件で必要なだけの転送能力を媒体リード・ライトに割り当てる。

FCインタフェースは，フレーム転送中は一時停止できないので，これも途切れることなくサービスしなければならない。ただし，データ転送の開始や継続は制御可能なのでファームのフロー制御による負荷調整ができる。

ECCの要求はバッファ上のデータの誤り訂正のためであり発生頻度も低く，時間的制約も緩い。ほかの要求には時間的制約があるが負荷としては軽い。

現在使用している転送能力配分の最適化方法は，つぎのとおりである。まず，2段階のラウンドロビンアービタ（巡回サービス型競合調停）を使用して，バッファによるサービスの優先順位を2段階に分ける。つぎに各データ経路にあるFIFO（First In

First Out）の大きさとデータ転送速度に合わせてバッファサービス1回あたりのデータ転送量をプログラムで調整する。こうしてFCインタフェースの速度や媒体リード・ライトの速度に応じて転送能力の配分を調整し，限られたバッファ転送能力を効果的に使用している。

● インタフェースの高速化対応技術

FCインタフェースの高速化により伝送路の信号品質が悪化することに対処する必要がある。

リンクの高速化はクロックを早くすることにより実現する。このときに問題になるのが受信機のジッタ耐性である。耐性向上のため，PLL回路によるクロック再生は周波数変動に対応する反面，外来ノイズに影響を受けやすい。ノイズ対策として，高速クロックのサンプリングによるデータ再生方式を採用している。

同じく，伝送路の周波数特性を補償するためのプリエンファシス（送信側）やイコライザ（受信側）を使用しているが，伝送路の特性はお客様の実装条件で大きく変わるのが常である。そのため，補正特性の調整をプログラミング可能にし，幅広いお客様のシステムに適合するようにしている。

今後の課題

システムメーカやエンドユーザからのHDDへの要求には，高性能化や大容量化，省電力化や静粛性など様々ある。ここではインタフェースに関連する市場要求とその実現のための課題について論じる。

すでに階層型ストレージとして製品化されているように，ストレージシステム向けのHDDは高性能化の方向と大容量化の方向に分化していくと考えられる。このとき両タイプのHDDにはストレージシステムの内部インタフェースに合わせて複数のインタフェースのサポートが要求されている。これはどのシリアルインタフェースのデータ転送も高速であるために，従来の用途別の範囲を越えて使用されるためと考えられる。この要求に応える効率的な開発方法はインタフェース制御部分をモジュール化して開発の重複を避けて効率化することと，コマンドレベルの互換性を同時に実現する方法が考えられる。

またそれぞれのインタフェースには更なる転送速度の向上が求められている。これについてはFCは4 Gbpsから8 Gbpsへ，SASは3 Gbpsから6 Gbpsへ，SATAは1.5 Gbpsから3 Gbpsへそれぞれ高速化する。これは回路の動作クロックを2倍にすることで対応できるが，デュアルポートの場合は伝送路の2倍と2ポート分の2倍の計4倍の要求が発生し，バッファの転送能力がボトルネックになりかねない。前述したように，転送能力配分の最適化を図っていく必要がある。

さらに，高速化に伴う伝送路の影響を最適化するために，伝送路の特性判定により最適化する適応型が検討されている。規格化の動向に注目しながら適宜対応することが必要になると考えられる。

現在，コンシューマ向けのHDDとしてCE-ATAインタフェースが仕様化[9]されている。これはATA系コマンドの一部分に限定したCE-ATA系コマンドを使用する。このことは新しい用途に最適なドライブを提供するためにドライブの価格・性能の観点から機能が見直されているということである。さらに，使用場所や目的に合わせて仕様化段階から最適化されようとしている。このインタフェースはHDDシリアルインタフェースのシンプル化の極致として注目する必要がある。

また，シリアルインタフェースの持つネットワーク機能により多種多様なHDDが一つのネットワークに混在することが可能になり，RAIDシステムでは複雑なデータ配置やデータ転送ルートを取る可能性はますます増大する。このような傾向に対して，単にデータ転送ルートだけでなくシステム全体としてデータ保全機能をどう強化するかが課題になる。この意味でEnd-To-End Data Protection機能[11]およびドライブのセキュリティ機能としてTCGの仕様[12]の規格化の意味は増大しつつある。

む　す　び

本稿では，現在のHDDのコントローラに使用されているシリアルインタフェースと，今後HDDインタフェースに使用される技術課題とその解決方法について紹介した。

新しい用途やマーケットの要求に最適なドライブを供給することがお客様から求められている。富士通は，従来の技術的蓄積を最大限に活用してお客様ニーズに適合した安定な製品をタイムリに開発することによりこれに応えていこうとしている。

参 考 文 献

（1） X3T11/Project1133D：Fibre Channel Arbitrated Loop（FC-AL-2）. Rev 7.0, 1999/4/1.

（2） Serial ATA WG：High Speed Serialized AT Attachment. Rev 1.0, 2000/11/15.

（3）X3T10/Project1601D：Serial Attached SCSI-1.1. Rev 9, 2005/3/18.

（4） X3T13/Project1532D：AT Attachment with Packet Interface-7 volume1-Register Delivered Command Set, Logical Register Set, Logical Register set. ATA/ATA-7V1, 2004/4/21.

（5） SCSI Parallel Interface-4 (SPI-4). Rev 10, 2002/05/06.
http://www.t10.org/ftp/t10/drafts/spi4/spi4r10.pdf

（6） CE-ATA WG：CE-ATA Storage Interface Specification. Rev 1.001, 2005/6/14.

（7） X3T11/Project1162DT：Fibre Channel Private Loop SCSI Direct Attach (FC-PLDA). Rev 2.1, 1997/9/22.

（8） NCITS T11/Project 1235-DT/Rev 2.7：FABLIC LOOP ATTACHMENT (FC-FLA). Rev 2.7, 1997.8.12.

（9） MMCA Technical Committee：The Multi Media Card. Ver 3.31, 2003/3, p.10.

（10）A. X. Widmer and P. A. Franaszek：A DC-balanced, Partitioned-Block, 8B/10B Transmission Code. *IBM Jounal of Reseach and Development*, Vol. 27, No.5, p.440-451（1983）.

（11）INCITS T11.2/ Project1316-DT/Rev 12.1：Fiber Channel – Methodologies for Jitter and Signal Quality Specification – MJSQ. Rev 12.1, 2003.12.7, p.26.

（12）X3T10：SCSI Block Commands-2 (SBC-2). Rev 16, 2004/11/13.
http://www.t10.org/ftp/t10/drafts/sbc2/sbc2r16.pdf

（13）INCITS T10 SAT Working Group：SCSI/ATA Translation (SAT). Rev 4, 2005.5.17, p.xv Foreword.

（14）Jim Coomes：SBC 32 Byte Commands for End-to-End Data Protection. Rev 7, 2004/4/21.
http://www.t10.org/ftp/t10/document.03/03-307r7.pdf

（15）TCG Peripherals Work Group：TPer & MCTP Requirements. Ver 1.0 Rev 0.03, 2004/8/13.