

ストレージ転送技術について

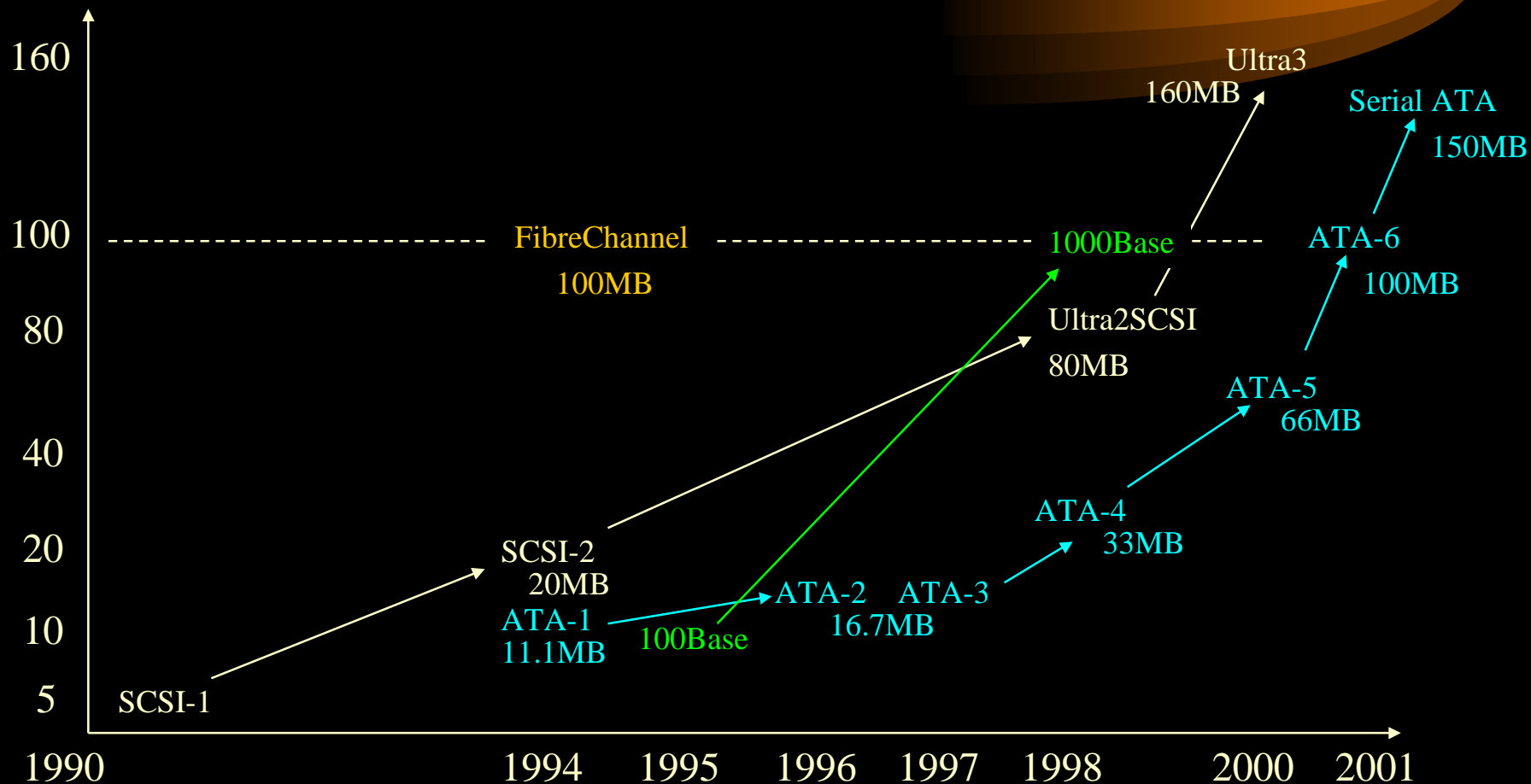


日本ヒューレット・パッカー

和田 明史

akifumi_wada@hp.com

ストレージ転送速度の歴史



SCSI

- SCSIは次のように定義されている
 - 物理レイヤと転送プロトコル(ケーブル、インターフェースなどで、速度はここで決まる)
 - SPI(SCSI on parallel interface)
 - FCP(SCSI on Fibre Channel interface)他
 - 命令セット
 - SPC, SPC-1, SPC-2, SPC-3他
- 通常SCSIといった場合、パラレルインターフェースのものを指す

SCSI

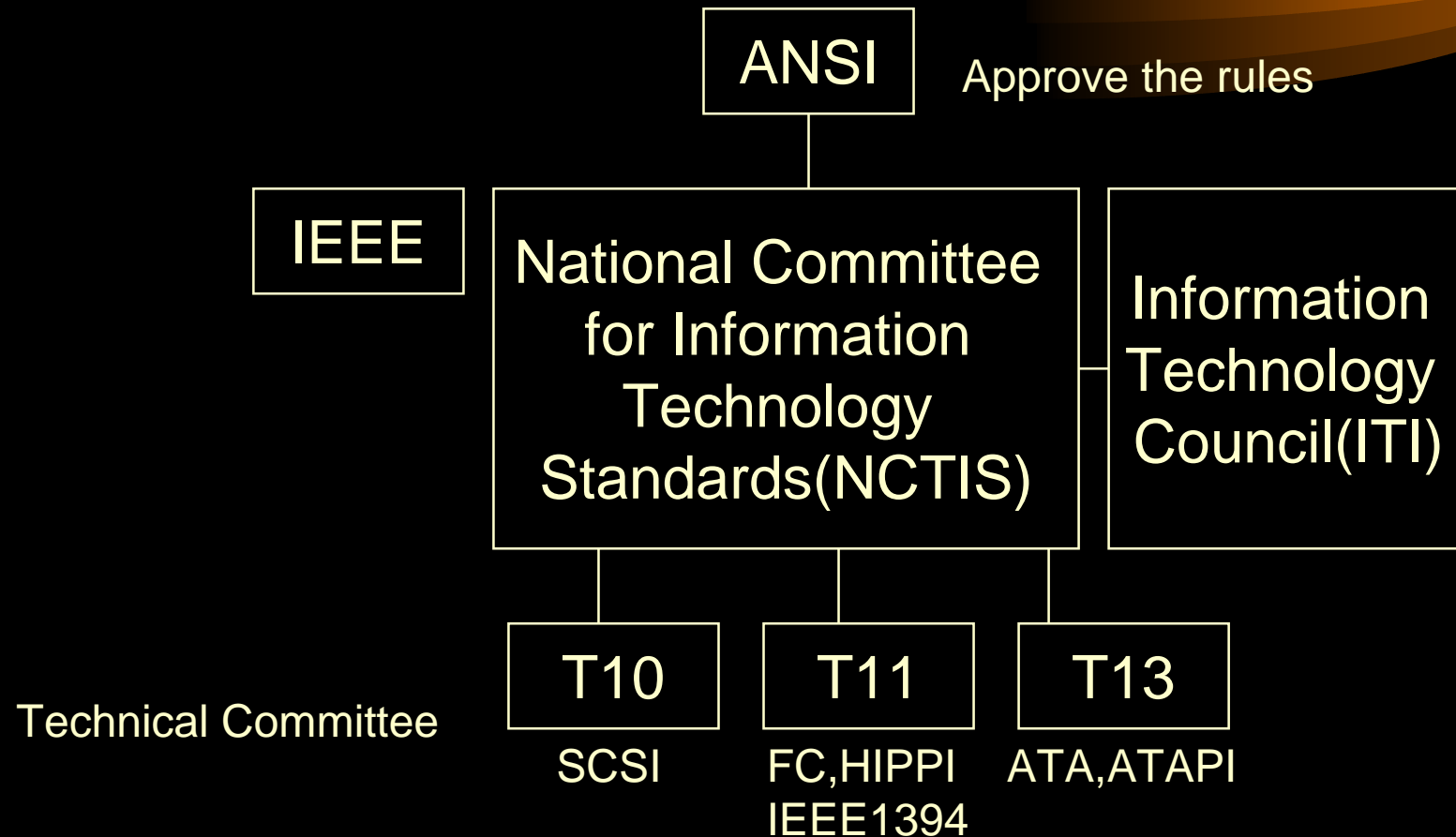
SCSI Device-type Specific Command Sets
Block Command, Enclosure Service, Stream Command, etc

Shared Command Set (For All SCSI devices)
SCSI-3 Primary Command Standard

Transport Protocols
FCP, *SCSI Parallel*, SCSI Serial Bus, etc

Physical interconnects
FC-PH, *SCSI Parallel*, IEEE 1394, etc

SCSI,ATA 標準化団体について



SCSI on Parallel Interface

	<i>Bus Speed</i>	<i>Bus Bit Width</i>	<i>Bus Clock</i>	<i>Cable length (single-end)</i>	<i>Cable length (Differential)</i>	<i>Max Devices</i>
<i>SCSI-1</i>	<i>5MB/s</i>	<i>8bit</i>	<i>5Mhz</i>	<i>6m</i>	<i>25m (HVD)</i>	<i>8</i>
<i>Fast SCSI</i>	<i>10MB/s</i>	<i>8bit</i>	<i>10Mhz</i>	<i>3m</i>	<i>25m (HVD)</i>	<i>8</i>
<i>Fast Wide SCSI</i>	<i>20MB/s</i>	<i>16bit</i>	<i>10Mhz</i>	<i>3m</i>	<i>25m (HVD)</i>	<i>16</i>
<i>Ultra SCSI</i>	<i>20MB/s</i>	<i>8bit</i>	<i>20Mhz</i>	<i>1.5/3m</i>	<i>25m (HVD)</i>	<i>4/8 (SE/DF)</i>
<i>Wide Ultra SCSI</i>	<i>40MB/s</i>	<i>16bit</i>	<i>20Mhz</i>	<i>No support</i>	<i>12/25m (LVD/HVD)</i>	<i>16</i>
<i>Ultra2 SCSI</i>	<i>40MB/s</i>	<i>8bit</i>	<i>40Mhz</i>	<i>No support</i>	<i>12/25m (LVD/HVD)</i>	<i>8</i>
<i>Wide Ultra2 SCSI</i>	<i>80MB/s</i>	<i>16bit</i>	<i>40Mhz</i>	<i>No support</i>	<i>12m (LVD only)</i>	<i>16</i>
<i>Ultra3 SCSI</i>	<i>160MB/s</i>	<i>16bit</i>	<i>80Mhz</i>	<i>No support</i>	<i>12m (LVD only)</i>	<i>16</i>

SCSI on Parallel Interface

- ストレージ接続の標準
 - ANSIによる認可のある標準
 - 完全に準拠していない製品も出てくる (Ultra160/m)
 - 比較的安価で、互換性が高く、入手しやすい
 - 性能が良い
- パラレルケーブルの問題は避けられない
 - ケーブル長と取り回し
 - 光ケーブルに比べてノイズに弱い

ATA(AT Attachment)

- AT(互換)機に接続するインターフェース、
プロトコル(命令セット)の規格
 - ANSIによる認可が通って標準となる
- パラレルインターフェースのみ定義されている(40ピンと80ピン)
- SCSIに比べて簡素な定義
- 速度の向上は著しい
 - Ultra ATA 33/66/100, ATA Raid

ATA(AT Attachment)

	<i>Bus Speed</i>	<i>Bus bit width</i>	<i>Bus Cycle</i>	<i>Cable length</i>	<i>Max Devices</i>
<i>PIO Mode0</i>	3.3MB/s	8(40pin)	1.66Mhz	18in.(46cm)	2 per channel
<i>PIO Mode1</i>	5.2MB/s	8(40pin)	2.61Mhz	//	//
<i>PIO Mode2</i>	8.3MB/s	8(40pin)	4.16Mhz	//	//
<i>PIO Mode3</i>	11.1MB/s	8(40pin)	5.55MHz	//	//
<i>PIO Mode4</i>	16.7MB/s	16(40pin)	8.33Mhz	//	//
<i>Singleword DMA 0</i>	2.1MB/s	16(40pin)	1.04Mhz	//	//
<i>Singleword DMA 1</i>	4.2MB/s	16(40pin)	2.08Mhz	//	//
<i>Singleword DMA 2</i>	8.3MB/s	16(40pin)	4.16Mhz	//	//
<i>Multiword DMA 0</i>	4.2MB/s	16(40pin)	2.08Mhz	//	//
<i>Multiword DMA 1</i>	13.3MB/s	16(40pin)	6.66Mhz	//	//
<i>Multiword DMA 2</i>	16.7MB/s	16(40pin)	8.33Mhz	//	//
<i>Ultra DMA 0</i>	16.7MB/s	16(40pin)	8.33Mhz	//	//
<i>Ultra DMA 1</i>	25.0MB/s	16(40pin)	12.5Mhz	//	//
<i>Ultra DMA 2</i>	33.3MB/s	16(40pin)	16.6Mhz	//	//
<i>Ultra DMA 3</i>	44.4MB/s	16(40pin)	22.2Mhz	//	//
<i>Ultra DMA 4</i>	66.6MB/s	16(80pin)	33.3Mhz	//	//
<i>Ultra DMA 5</i>	100MB/s	16(80pin)	50Mhz	//	//

ATAとSCSI on parallel interface

- ATAがSCSIに勝っている点
 - 定義が簡素で実装が簡単
 - ユーザーにとっては、より安価なデバイス
- ATAがSCSIに劣っている点
 - ケーブル長および接続デバイス数
 - サポートしているデバイス(命令セット)
 - コマンドキューイング(深さ、オーダー)
 - 転送エラー補正(CRCチェック)

次世代のデバイス

- SCSI/parallel interface
 - Ultra320(320MB/sec,SCSI-4)
- ATA
 - ATA/ATAPI-6(48bit addressing)
 - 28bit addressingから拡張、SCSIは32bit addressing
 - Serial ATA(150MB/sec-300MB/sec)
- 高速デバイスの開発に伴う高速バスの開発が不可欠(PCI-Xなど)

ローカルストレージから 共有ストレージへ

- SCSIやATAはいずれもローカルストレージで本体に繋げて使うものであるのに対し、コンピュータの普及と歩調を合わせるように、データの共有あるいは遠隔アクセスを可能にする技術が開発された
- NFS (Sun)
- NETBIOS (Microsoft)
- CIFS/SMB over TCP (Microsoft)

ローカルストレージから 共有ストレージへ

- LAN(10/100/Gbit Ethernet/FDDI)、WAN (ATM/Frame Relayなど)で利用可能
 - TCP/IPやUDP/IPに依存している
- SCSIやATAと違ってファイル単位で転送する(SCSIやATAはブロック単位)
- ネットワーク機器の低廉化及び高速化にともない急速に普及(ワークスタイルやシステム環境に変化)

ローカルストレージから 共有ストレージへ

- ネットワークによる共有ストレージは、サポートしている(動作する)プラットフォームが多数に渡り、異機種混合環境でも利用可能
- 既存インフラストラクチャでデータがやり取りできる
- ftpやrcpと違ってOSから透過的にアクセス可能

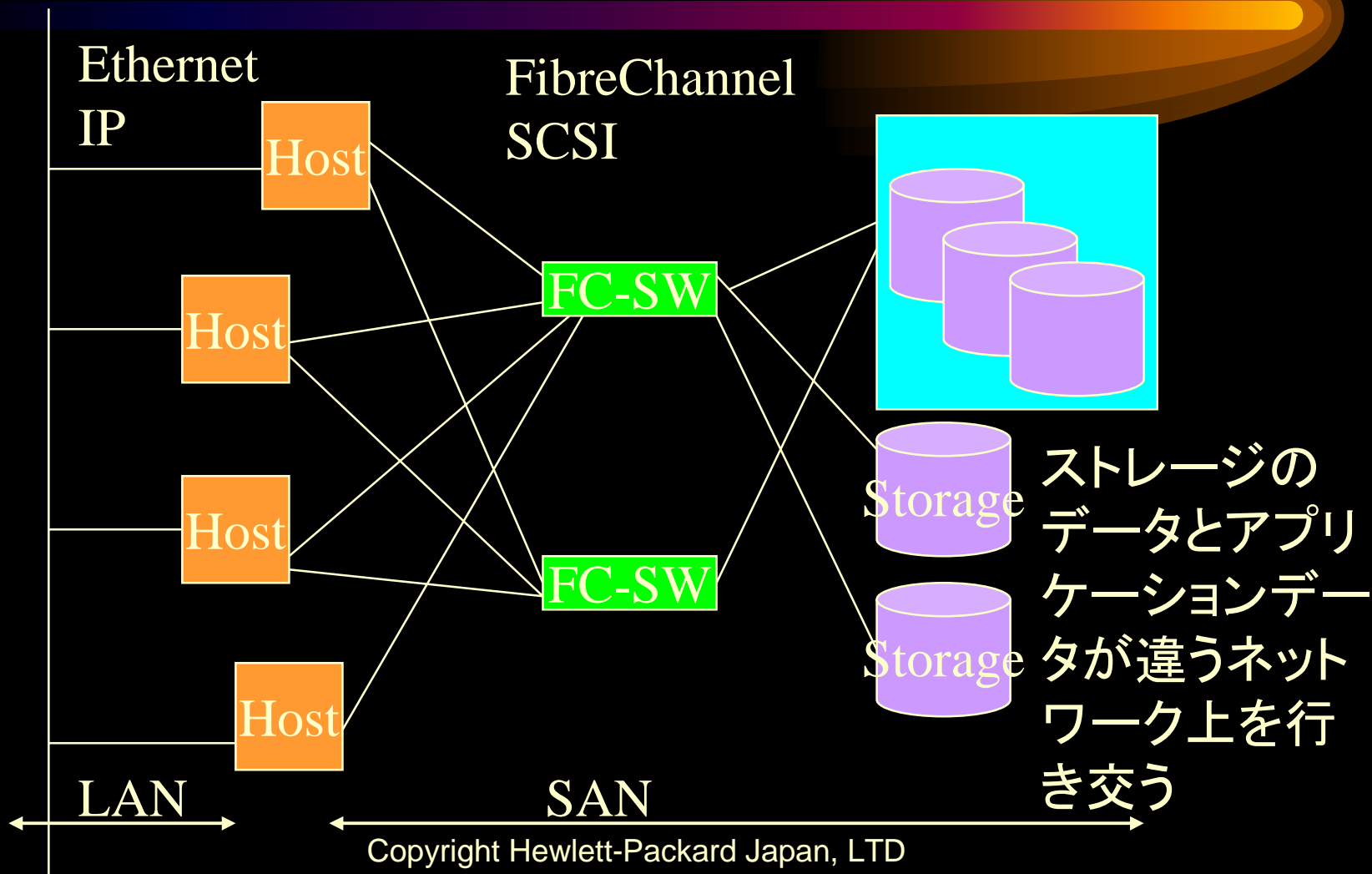
ローカルストレージから 共有ストレージへ

- アプリケーションデータとストレージのデータが同一のネットワークを行き交う
 - 帯域幅の不足
- ファイルシステムの共有
 - オーバーヘッドが大きい

ストレージエリアネットワーク

- 一般にSANと呼ぶものは、
 - ファイバーチャネルで構成された
 - LANとは別に設けられている
 - (通常は)SCSIコマンドをやりとりする
 - any-to-any (P2P)の接続が可能なネットワーク
- SCSI over FibreChannelと呼ぶべきもの
- ローカルストレージと同じように扱える

ストレージエリアネットワーク



ストレージエリアネットワーク

- 通常ファイバーチャネルで構成されているため、ファイバーチャネルの利点をすべて享受できる
 - 100MB/secのバンド幅
 - 500mあるいは10kmまでのケーブル長
 - 127あるいは1677万のデバイス接続
 - リブートを伴わない機器の追加（一部OS不可）
 - テープデバイスの共有

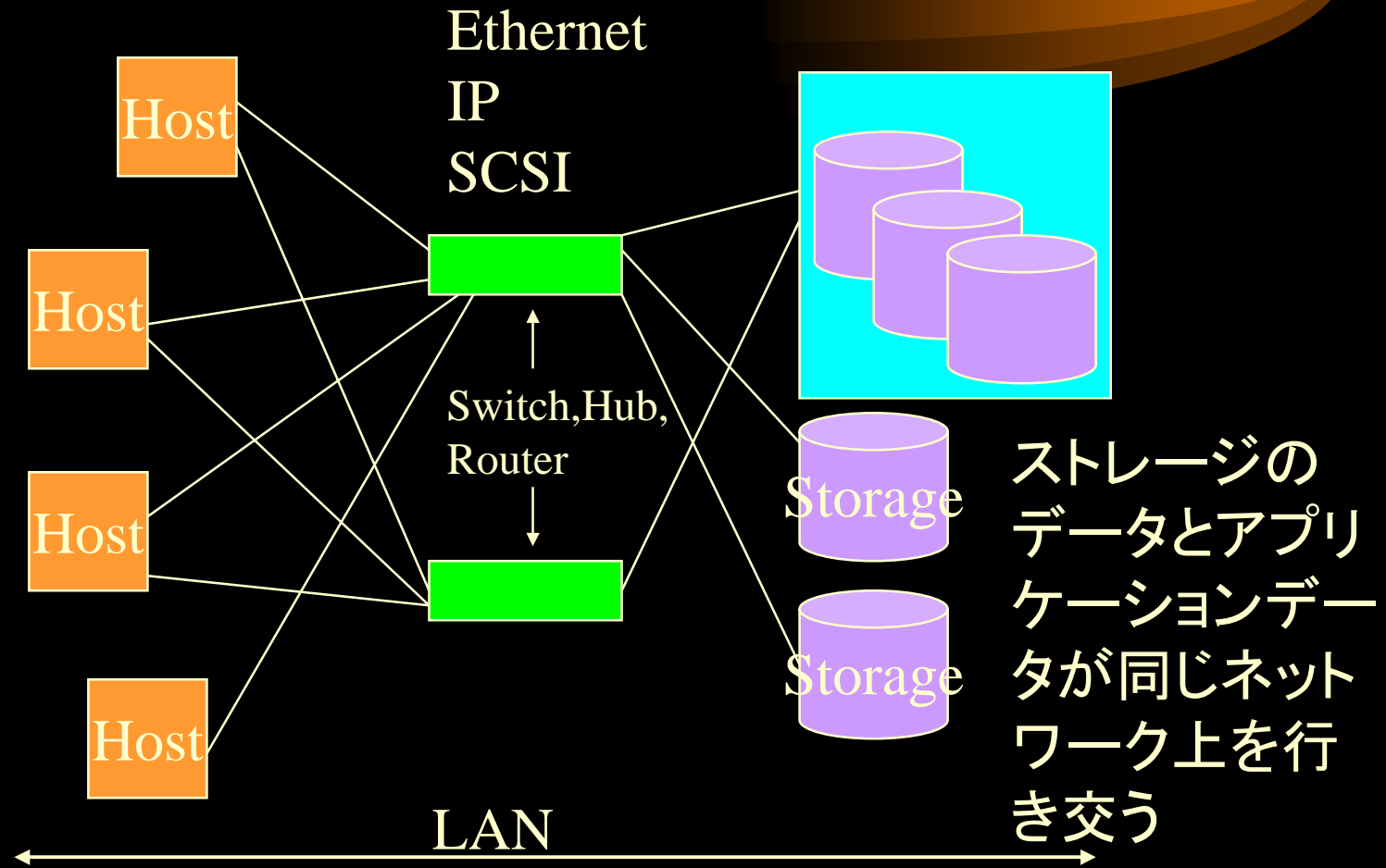
ストレージエリアネットワーク

- 市場が小さく機器が非常に高価
- SCSIやATA、LAN/WANに比べてベンダーに依存している部分が多い(互換性が低い)
- SCSI命令を直接扱うので異なるOS間で同一デバイスを共有するのは難しい
- ファイバーチャネルは距離に対する拡張性がATMなどよりも低い
 - コンバータ(ダイレクタ)の使用

iSCSI

- iSCSI = SCSI over TCP/IP
- SCSI命令をIPネットワーク上でやり取りする規格
- LAN/WANのインフラを利用してSANと同等(以上)のことに実現する
- 既存インフラストラクチャが利用できるので、投資が小さくて済む(はず)
- IETFにより標準化作業中

iSCSI



ストレージの
データとアプリ
ケーションデー
タが同じネット
ワーク上を行
き交う

iSCSI

- Gigabit Ethernetの普及によりLANの世界では高速
- IPベースのWANで利用できる
- 既設ケーブルを再利用できる
- NFS/CIFSよりもオーバーヘッドが小さい
- SCSI over FCでは出来ないルーティングが可能 (IPを用いたセキュリティポリシー)

iSCSIの課題

- パフォーマンス
 - TCP/IPスタックは通常カーネルで処理されるため、CPUの消費が激しい
 - IPを処理するカスタムチップが現在もっとも有望
 - Ethernetではペイロードサイズが1500バイトと小さいためSCSI over FCよりも効率が悪い（FCは最大2112バイトある）
 - NFS/CIFS < iSCSI over Ethernet < SCSI over FC
 - Jumbo Frame (MTU size GbE:9000, FC:65535)

iSCSIの課題

- レイテンシ
 - IPネットワークを経由するためミリ秒の単位で遅延が発生する
 - ブロックI/Oが要求するパフォーマンスを充足できない可能性が大きい
- ベスト・エフォート
 - パケット損失によるデータの遅延

次世代のテクノロジー

- ファイバーチャネル
 - 2Gbps(200MB/sec)-4Gbps(400MB/sec)-10Gbps
- ギガビット・イーサネット
 - 10Gbps
- InfiniBand
 - 2.5Gbps-30Gbps
- DWDM技術による広帯域WANサービス
 - 160Gbps-1.6Tbps

まとめ

- 企業向け(サーバー向け)ストレージという点でSCSIはATAに比べ優位
- 高速(広帯域)化は次世代バスが不可欠
 - PCI-X, InfiniBand他
- ネットワーク型ストレージとしてはNAS(NFS/CIFS)、SAN(FibreChannel)、iSCSIがあるがどれかが支配的になることは当分の間ない

- NCTIS Technical Committee
 - T10(SCSI) <http://www.t10.org/>
 - T11(FC) <http://www.t11.org/>
 - T13(ATA) <http://www.t13.org/>
- Internet Engineering Task Force(iSCSI)
 - <http://www.ietf.org/>
- IEEE(Ethernet)
 - <http://www.ieee.org/>