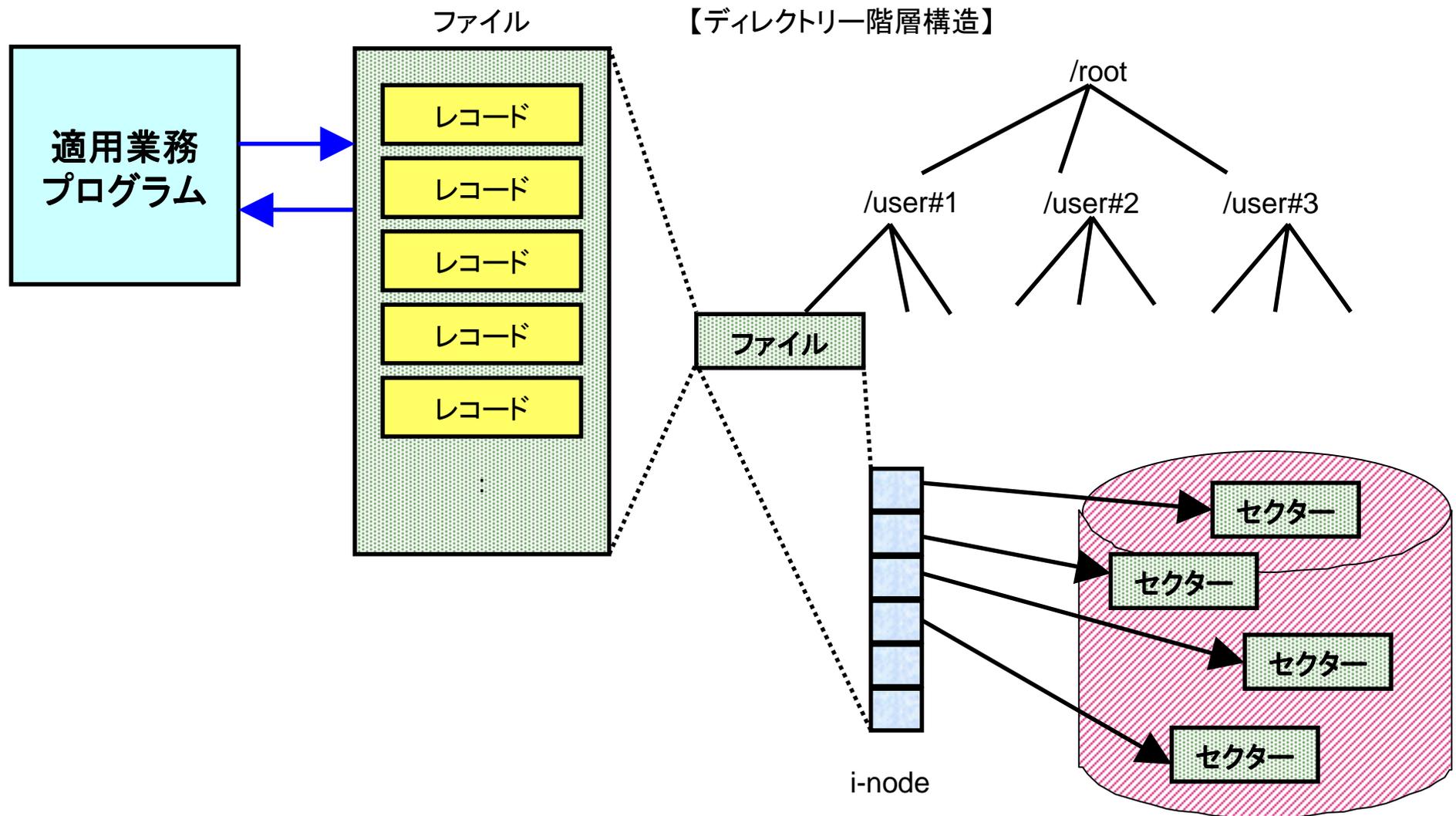
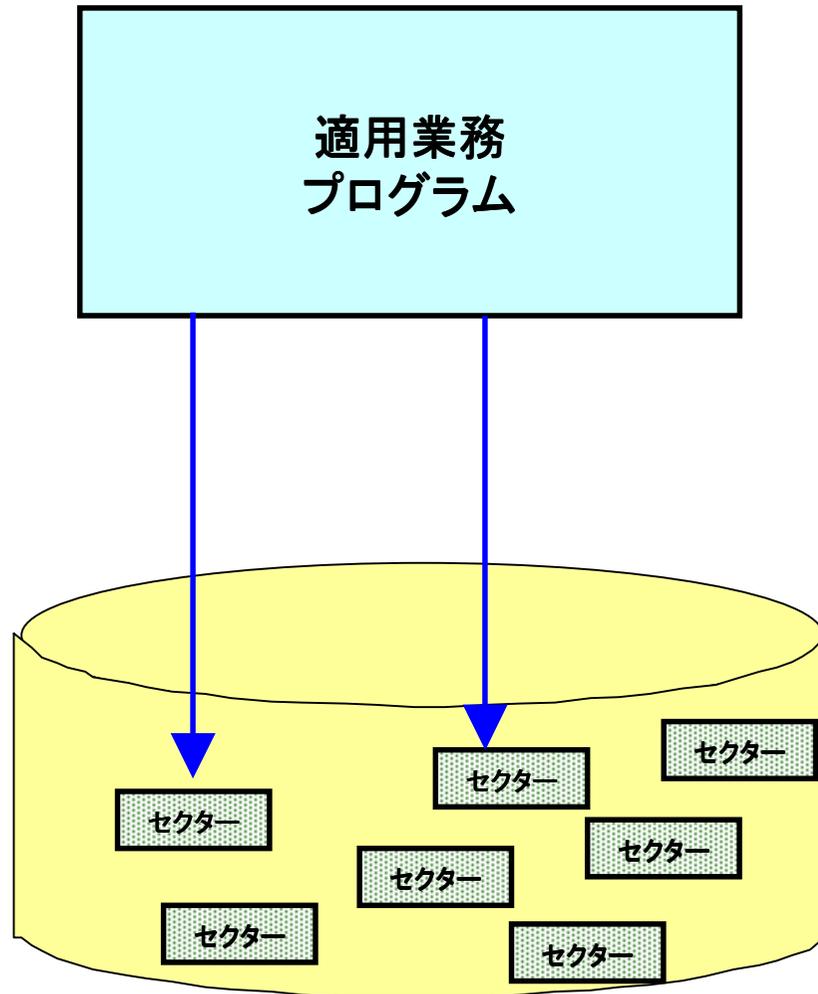

ファイルとファイル・システム

ファイルとファイル・システム

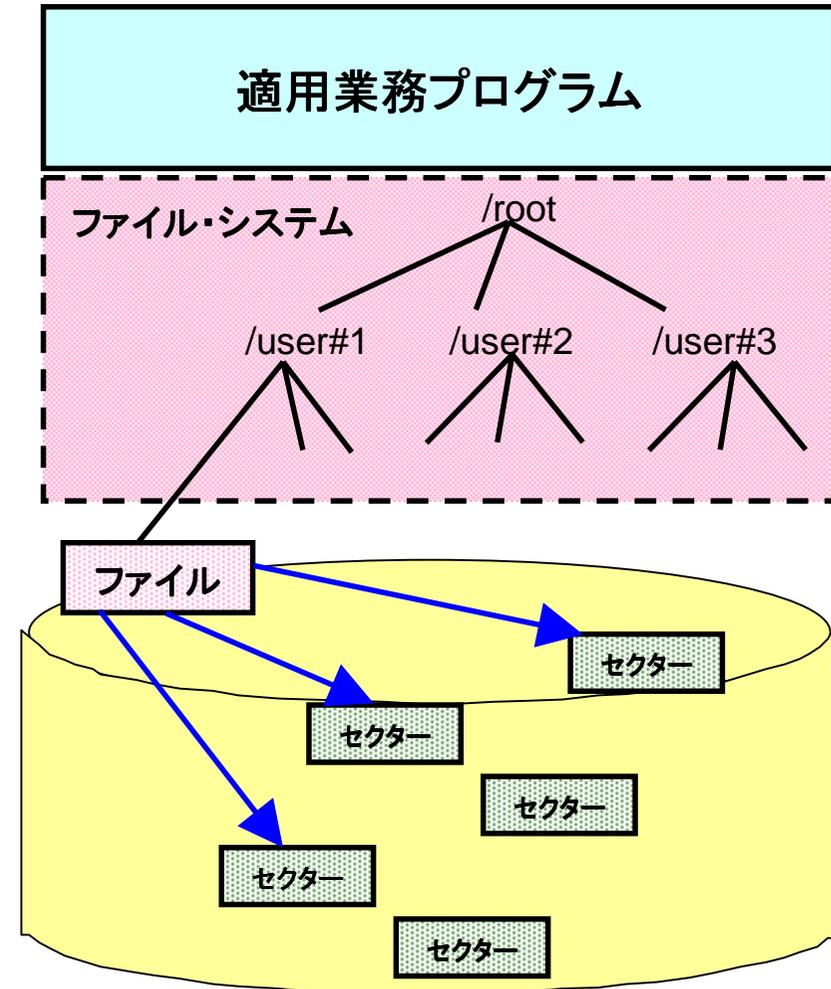


ロー・デバイス・アクセスとファイル・システム・アクセス

【ロー・デバイス・アクセス】

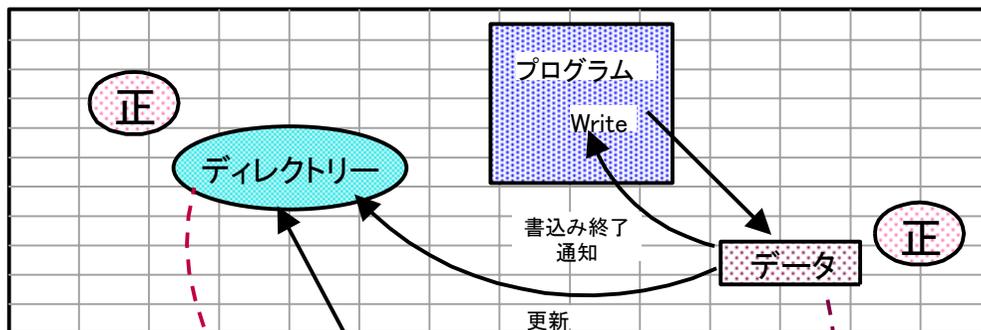


【ファイル・システム・アクセス】

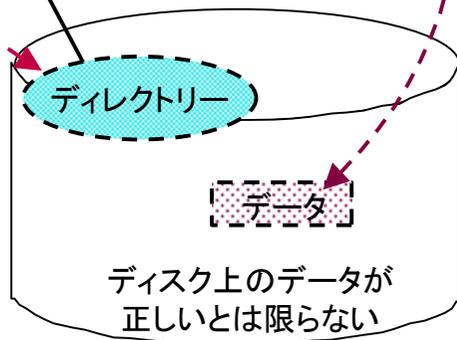


オープン系ファイル・システムのI/O処理の特徴

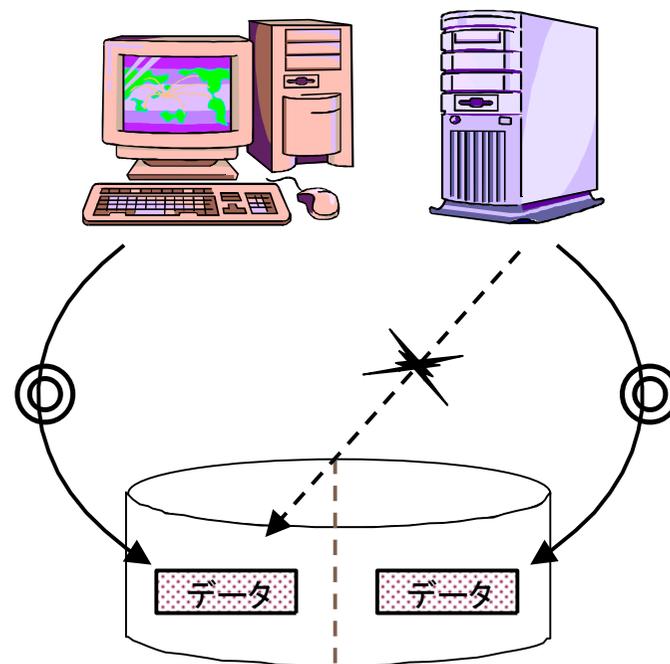
サーバーの主記憶メモリー



非同期更新
(後で書く)



非同期更新
(後で書く)



オープン系システムでは
ディスク共用によるデータ共用が困難

SCSI 規格

SCSI (Small Computer System Interface)

- SCSIとは、コンピュータと補助記憶装置を接続するための規格
 - スカジィ
 - HDDを接続する標準的なインタフェース
 - 対象は、HDDのみではなく、フロッピーディスク、テープ装置、CD、DVD、スキャナ、プリンタ、コンピュータ・システムなど、幅広い

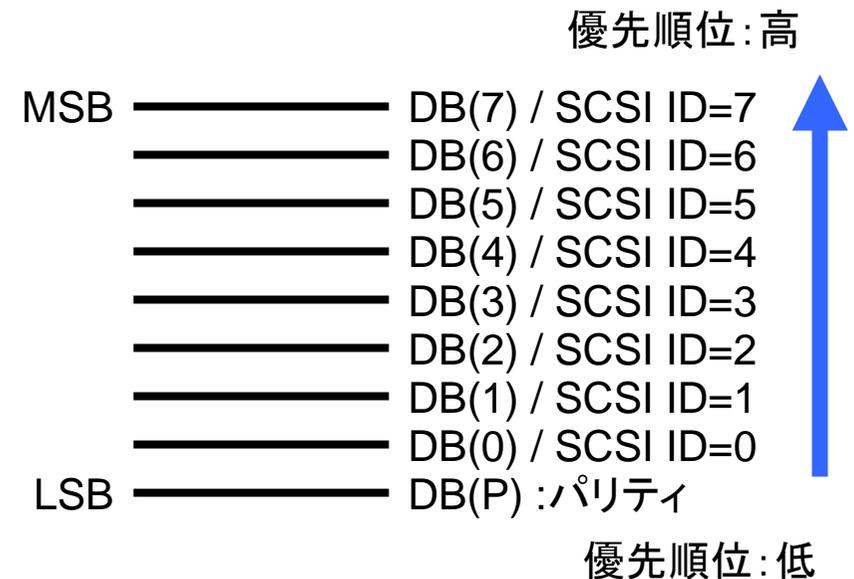
- バス型

- 平行・インタフェース

- データ巾: 8+1 (パリティ)
- データ線とアドレス線 (SCSI ID 指定) が共通
- バス使用権の優先順位が固定

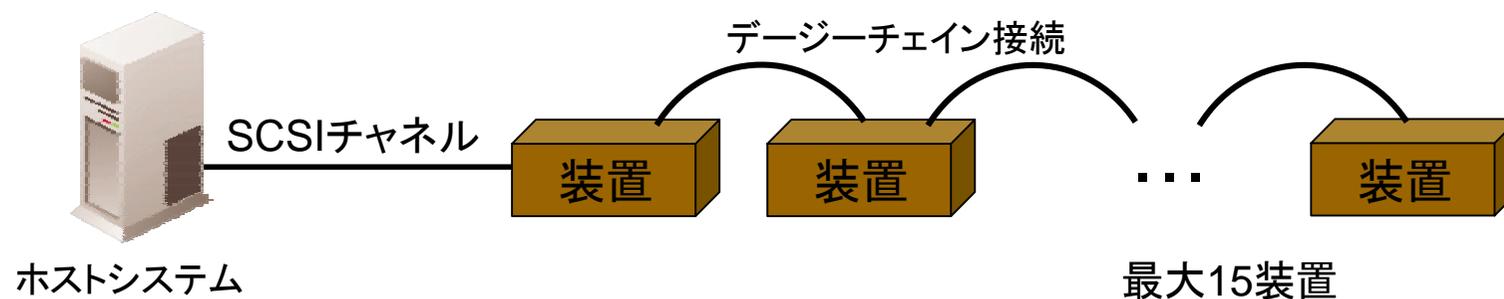
- ANSI NCITS (旧 X3 委員会) T10 部会

- SCSI-2 : ANSI X3.131-1994
- SCSI-3 : ANSI X3.270-1996 他

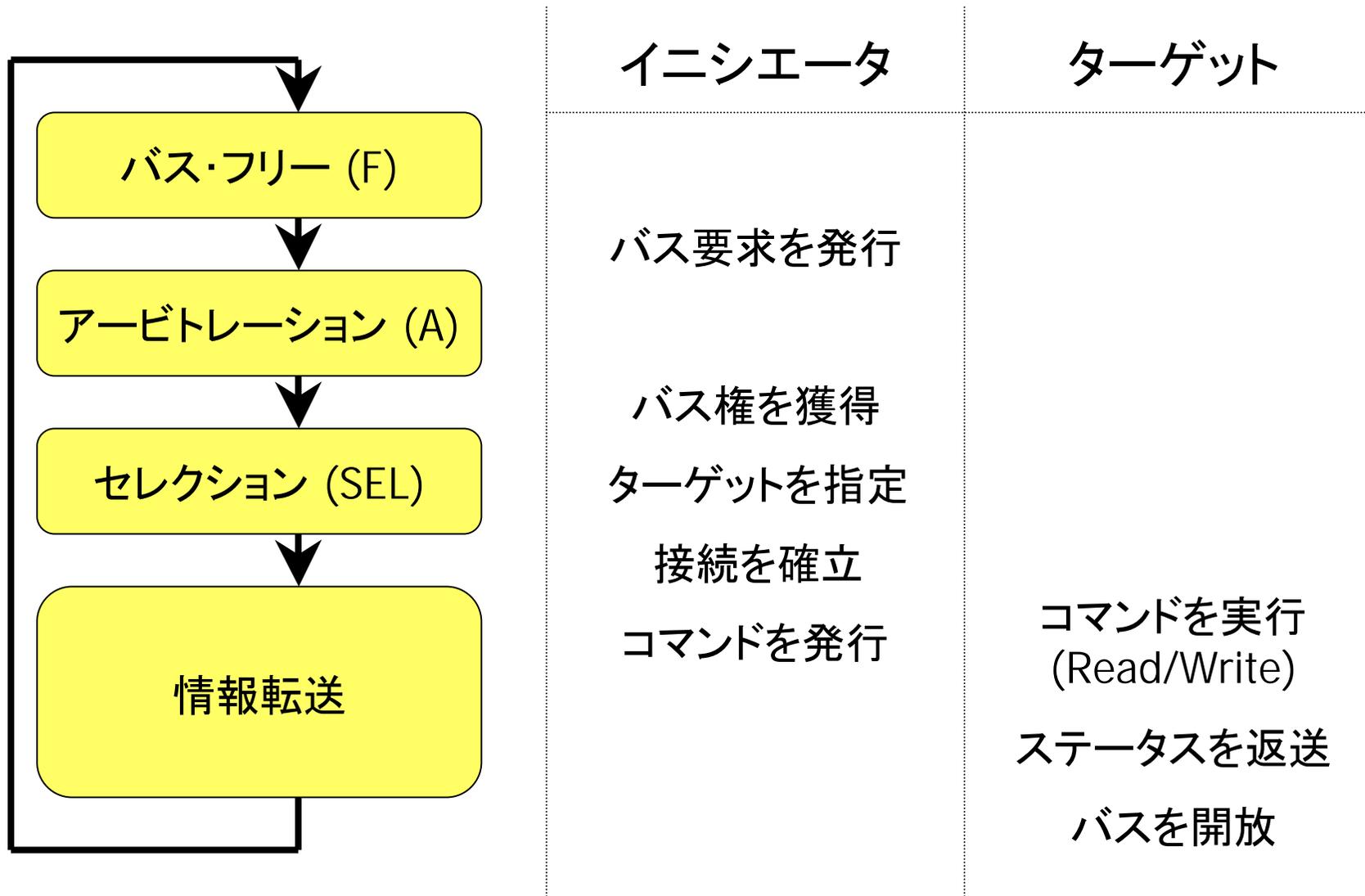


SCSIデバイスと接続形態

- イニシエータ・デバイス
 - コマンドの発行
 - ホスト (HBA)
- ターゲット・デバイス
 - コマンドに対する応答
 - HDD
- イニシエータとターゲットの間で1:1の伝送
 - 他のデバイスはバスが開放されるのを待つ



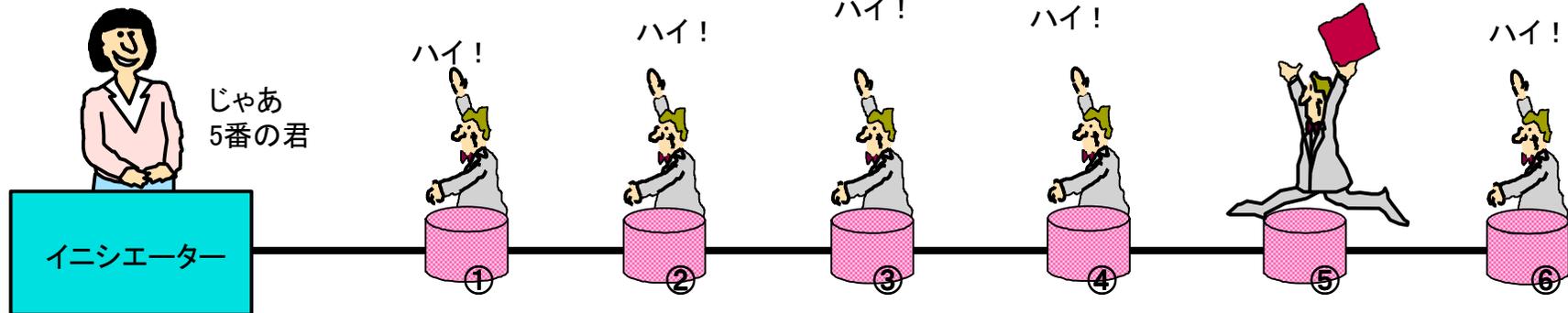
SCSIデバイスの動作概要



アービトレーション

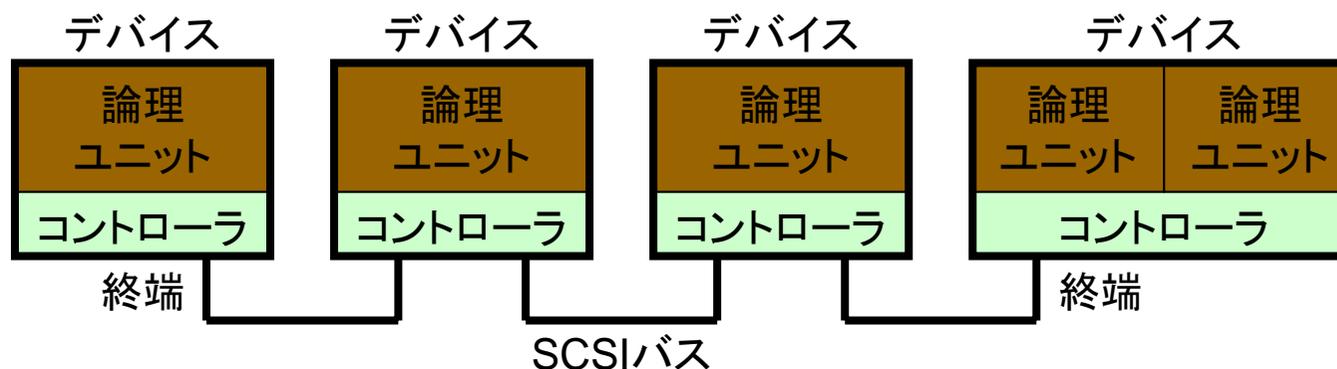
- 次のサービス提供のための、選定作業
 - イニシエーターは次にサービスを提供するデバイスを選定するために、アービトレーションを行う
 - デバイス全てに対し、サービス要求があるかどうかを尋ね、優先順位に従って次にバスの使用权を得るデバイスを選定する
 - SCSIではデバイス番号によって優先順位の高さが決まる

次は誰を指しましょうかしら



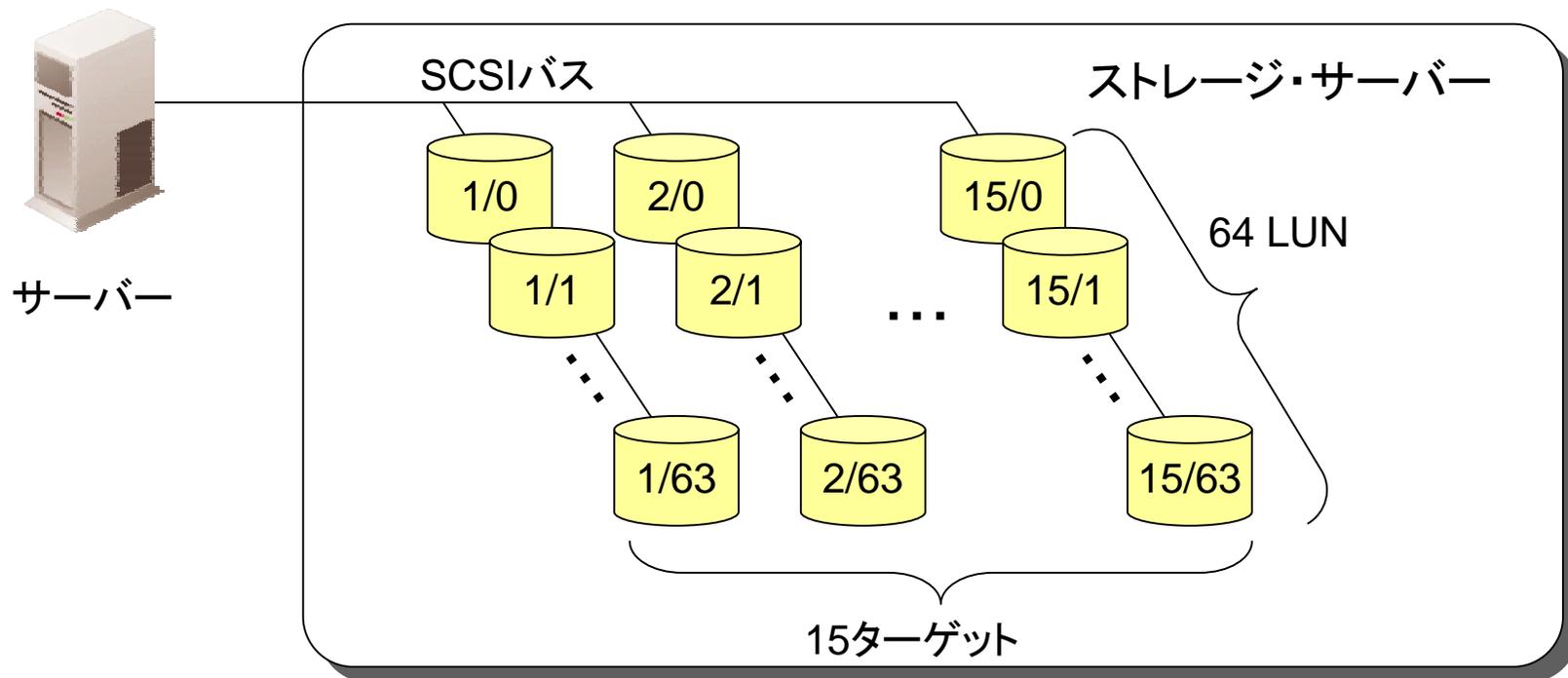
SCSIデバイスの接続

- SCSIデバイス=コントローラ + 論理ユニット
 - コントローラ：インターフェース
 - 論理ユニット(Logical Unit)：装置そのもの
 - 8コントローラ/バス
 - 8論理ユニット/コントローラ
 - デイジーチェーン接続



LUN (Logical Unit Number)

- ストレージ・サーバー内の各ディスクはLUNとしてサーバーから認識される
 - LUNは物理SCSI HDDイメージで認識される
 - ターゲット数とLUN数の最大値はサーバー側のOSに依存
 - ポートあたり最大960ディスク(15ターゲット×64 LUN)

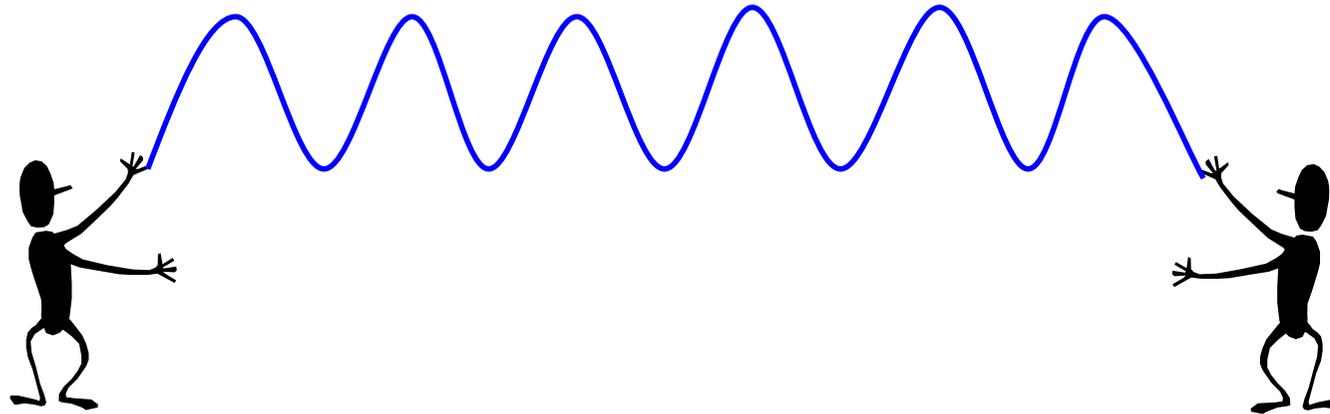


SCSIの進化: SCSI-2

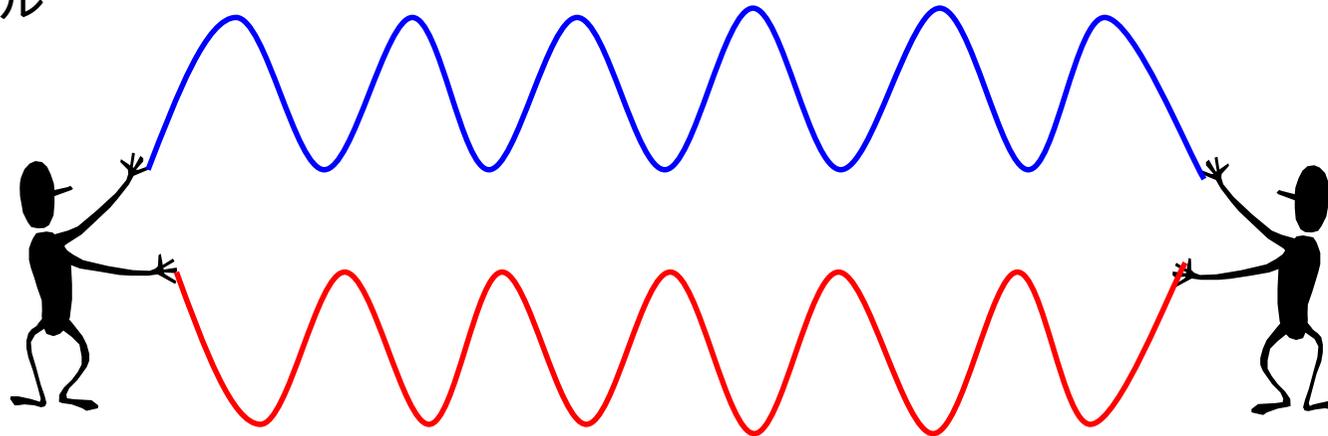
- コンピュータ本体とハード・ディスクなどの記憶装置の接続に用いられるSCSI規格の一つで、SCSI-1を改良した第二世代の規格群
 - 1994年にアメリカ規格協会(ANSI)が標準化
- 2種類の規格
 - 「Fast SCSI」(Fast Narrow SCSI)
 - 8ビット幅
 - 10MB/s
 - 最大接続台数は8台
 - 「Fast Wide SCSI」
 - 16ビット幅
 - 20MB/s
 - 最大接続台数は16台
 - 最大ケーブル長はどちらもシングルエンド駆動で3m、ディファレンシャル駆動で12m(LVD)～25m(HVD)である。
- SCSI-2からは、高速なデータ転送の妨げとなる回線終端での信号の反射を抑えるため、終端に取り付けるターミネータはアクティブ型が必須となった

シングル・エンドとディファレンシャル

■ シングル・エンド



■ ディファレンシャル

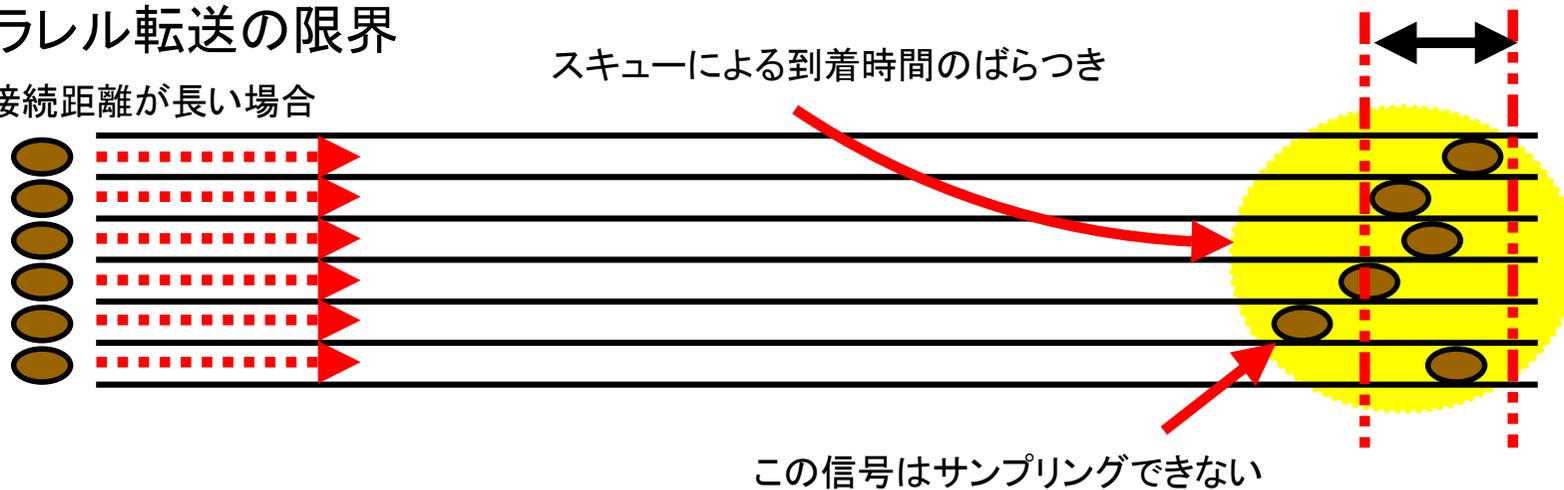


パラレルSCSIの限界

- 高速化が進んでいますが以下の欠点がある
 - 短い接続距離
 - 限られた装置数
 - 制限された拡張性
 - 単一経路のアクセス

パラレル転送の限界

接続距離が長い場合



SCSIの進化: SCSI-3

- コンピュータ本体とハード・ディスクなどの記憶装置の接続に用いられるSCSI規格の一つで、SCSI-2を改良・拡張した第三世代の規格群
- パラレル転送方式を用いていた従来のSCSI規格群(パラレルSCSI)の延長に当たるUltra SCSI等の仕様に加え、Fibre Channel、SSA、IEEE 1394など、シリアル転送方式を採用したシリアルSCSI規格群を新たに制定
- SCSI-3アーキテクチャモデル
 - コマンド・レベル
 - デバイス個別のコマンド
 - 共通コマンド (SCSI-3 Primary Command : SPC)
 - プロトコル・レベル
 - デバイスを接続するプロトコル
 - インターコネクト・レベル
 - 物理媒体
 - シリアル・インタフェースの追加



パラレルSCSI-3規格

- Ultra SCSI規格
 - 「Ultra SCSI」
 - 8ビット幅、20MB/s
 - 「Wide Ultra SCSI」
 - 16ビット幅、40MB/s
 - 駆動方式は従来のシングルエンド駆動のままだったため、最大接続長が8台接続時で1.5mまで、4台接続時で3mまでと短くなってしまった
- Ultra2 SCSI規格
 - 「Ultra2 SCSI」「Wide Ultra2 SCSI」では、駆動方式をディファレンシャル方式に改め、それぞれ40MB/s、80MB/sという高い転送レートを維持しながら、最大接続長を12m(LVD)～25m(HVD)に伸ばすことに成功
- Ultra3 SCSI規格
 - 転送速度を160MB/sに高めた「Ultra3 SCSI」(Ultra160 SCSI)や、320MB/sの「Ultra320 SCSI」などの仕様が策定されている

SCSI規格：LVDとHVD

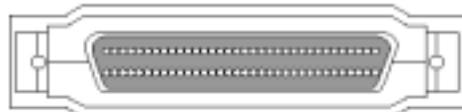


SCSIインタフェース	SCSIバス転送速度	説明
HVD (High Voltage Differential)	40MB/sec	+と-の2本の信号線でデータを転送する方式である 信号線への供給電圧は5Vを使用します
LVD (Low Voltage Differential)	80MB/sec	+と-の2本の信号線でデータを転送する方式である 信号線への供給電圧は3.3Vを使用します
SE (Single-Ended)		1本の信号線でデータを転送する方式である

注) テープドライブ自体の転送速度はドライブによって決まります

注) 同一SCSIバス上に、LVD/SEおよびHVD SCSIアダプタ、テープドライブ、ターミネータを混在しないでください。
これらの機器が破壊されるおそれがある。

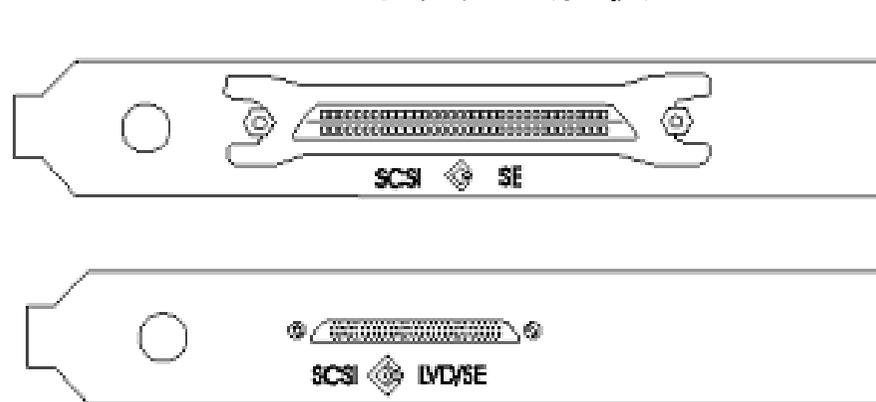
高密度/ハーフピッチ68ピン
HD68 (High Density 68 pin)



0.8mm VHDCI

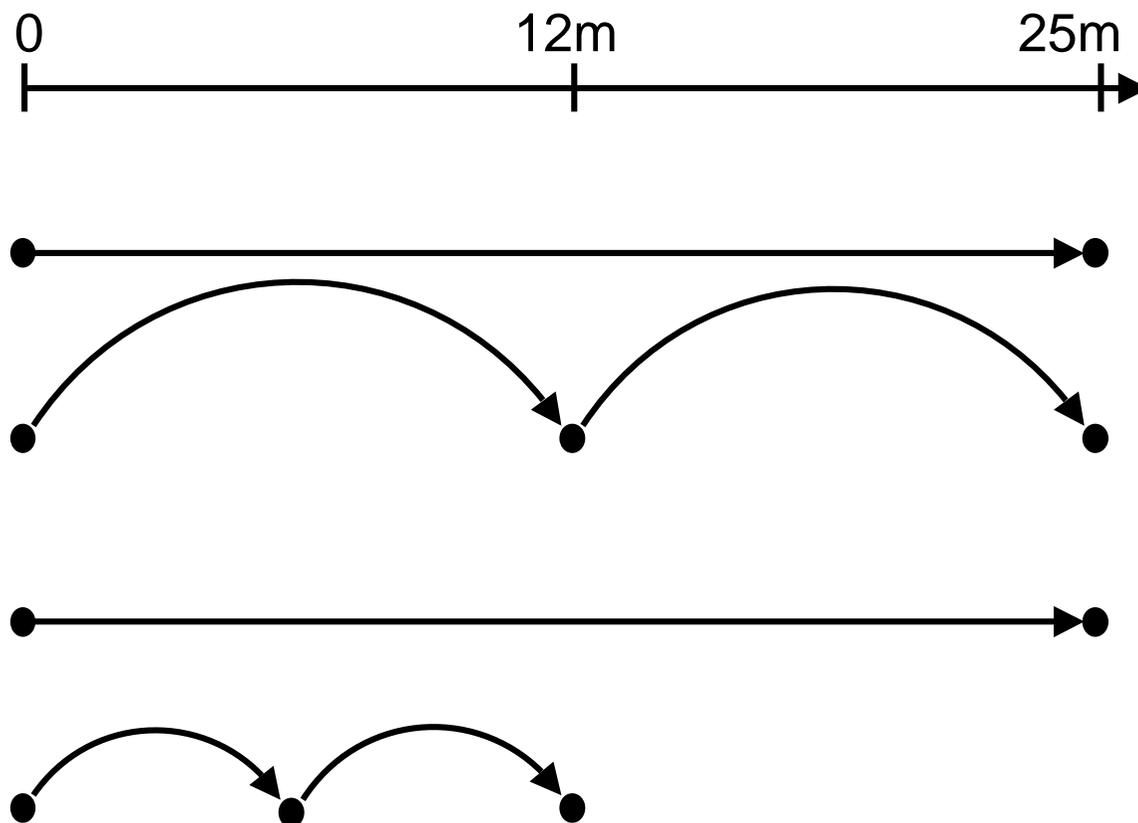


SCSIコネクタ形状



SCSIケーブル長の制限事項

- HVDの場合
 - Point-to-point接続
 - 全長25mまで
 - マルチドロップ接続
 - 全長25mまで
- LVDの場合
 - Point-to-point接続
 - 全長25mまで
 - マルチドロップ接続
 - 全長12mまで



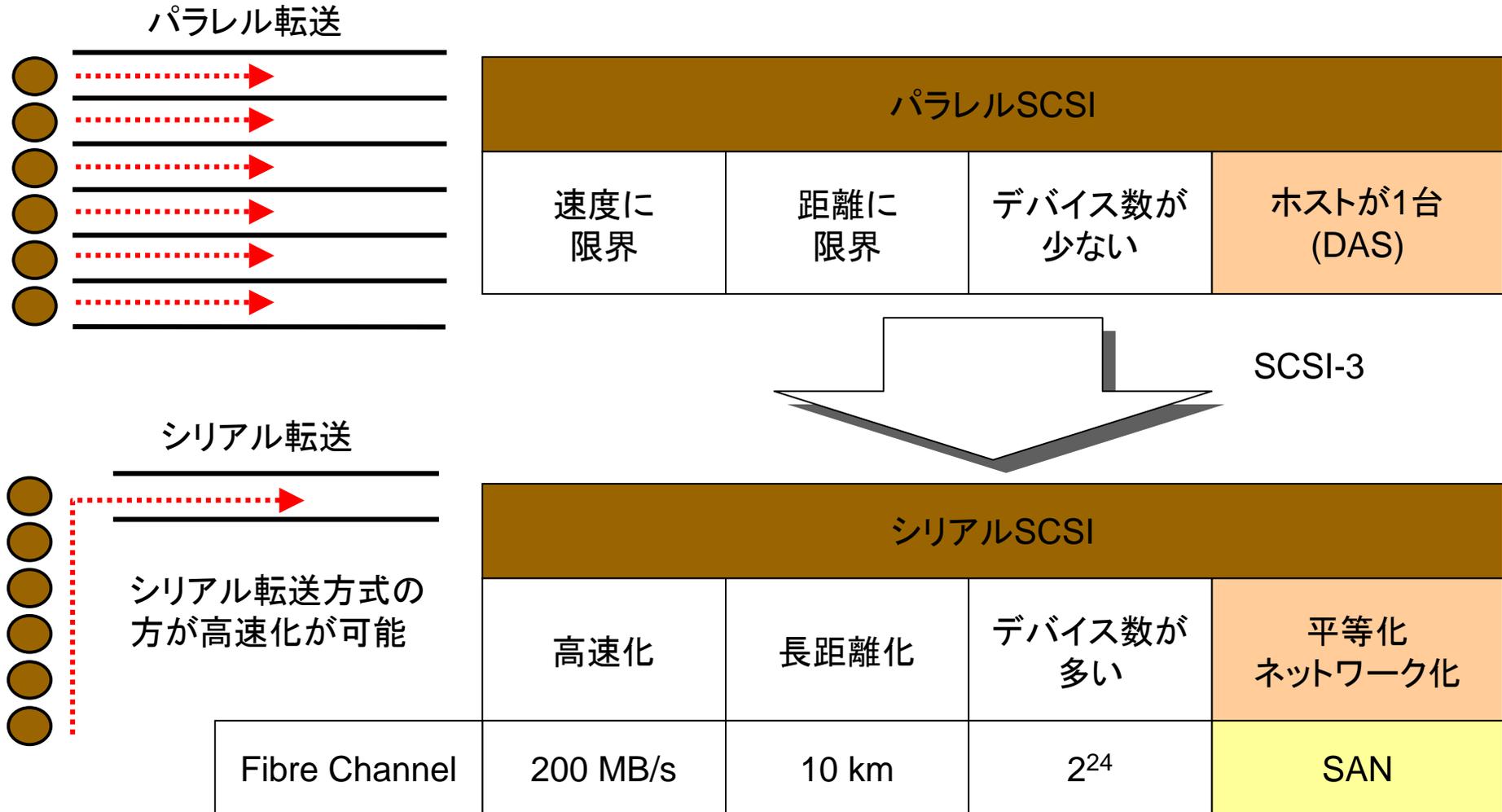
SCSI規格のまとめ

	Pケーブル (68)	Qケーブル (68)	データ巾 兼 SCSI ID数	転送モード					
				非同期	同期	Fast	Ultra	Ultra2	Ultra3
Narrow	○	-	8	1.5MB/s	5MB/s	10MB/s	20MB/s	40MB/s	80MB/s
Wide	○	-	16	-	10MB/s	20MB/s	40MB/s	80MB/s	160MB/s

※ コマンドやステータスは常に非同期転送モードで伝送される

ファイバーチャネル SCSI over Fibre Channel

シリアルSCSIの登場



Fibre Channel規格

- 高速シリアル・インターフェイス・テクノロジー
 - 送信と受信の2本の転送路を使用
- いくつかのオープンな標準から構成
 - メディアと物理インターフェイス(光ファイバーと銅線)
 - データ転送、リンク・サービス、信号プロトコル
 - 上位レベル・プロトコルのマッピング(異なるコマンド・セット: SCSI, FICON, HIPPI, IP, IPI-3, ATM, 他)
- FCP上のSCSIプロトコルは、従来のパラレルバス上のSCSIプロトコルと同一



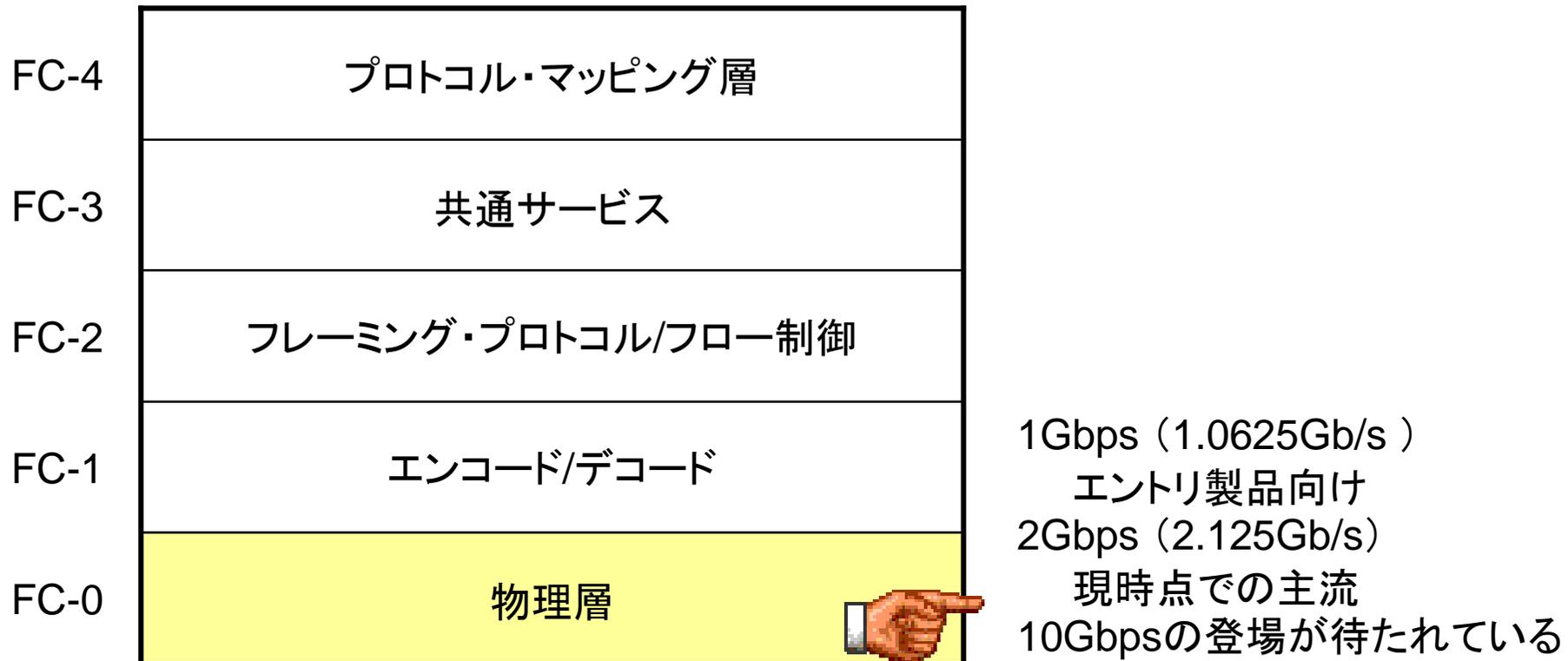
FCP とは

- ストレージエリアネットワーク(SAN)を可能にする基礎の技術
 - 高速データ転送
 - 100-200MB/sec(全二重)
 - 200-400MB/sec(全二重)
 - 最長10kmの接続距離
 - 銅線:30m
 - ショートウェーブファイバーケーブル:500m
 - ロングウェーブファー場ケーブル:10km
 - リピーターを用いることで最大100 Kmまで延長可能
 - 大規模で拡張性のある構成のサポート
 - Point-to-point
 - Arbitrated Loop
 - ファブリックスイッチ
 - コントローラあたりのデバイス数の増加
 - FC-ALで最大127台まで
 - ファブリックスイッチで最大1600万台まで
 - ネットワーク可能
 - パラレルSCSIはサーバーとストレージを接続する専用のパラレルバス
 - ホット・プラグ可能

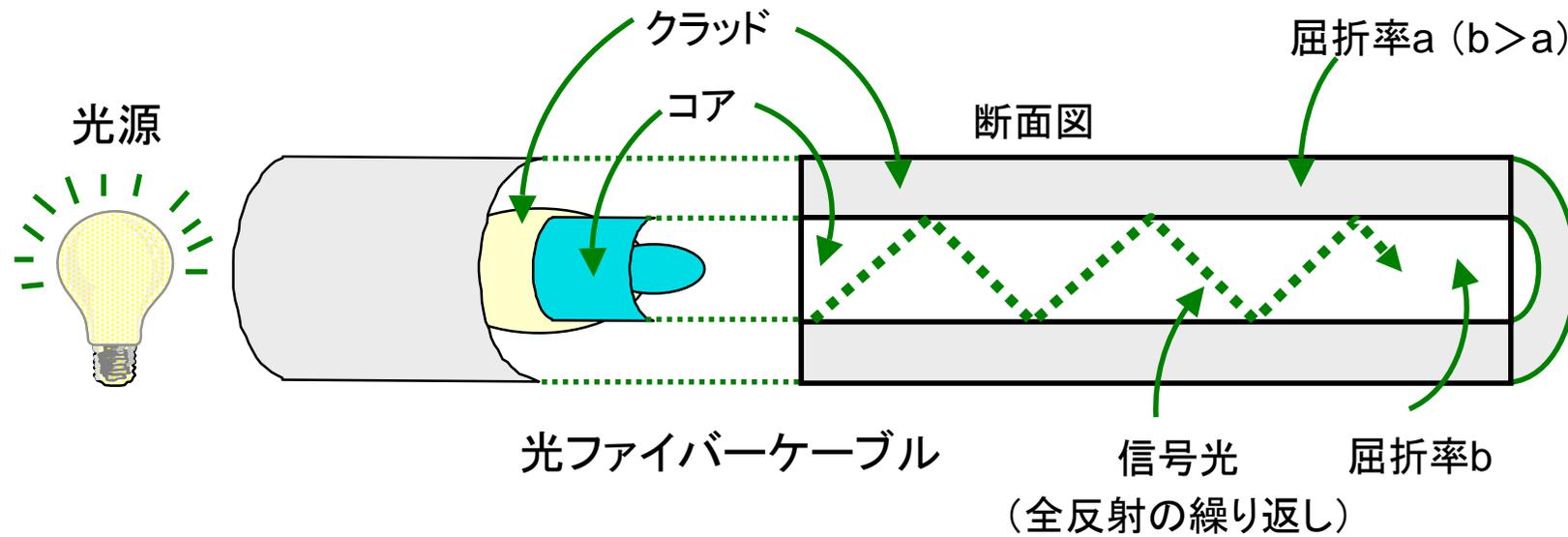
ファイバーチャネル規格:FC-0

ケーブル、コネクタ、トランスミッター、レシーバーなどの媒体の物理的な特性を規定

- 光メディア、または、銅メディア上で動作
- 銅メディアはファイバーチャネルディスクドライブで主に使用されている

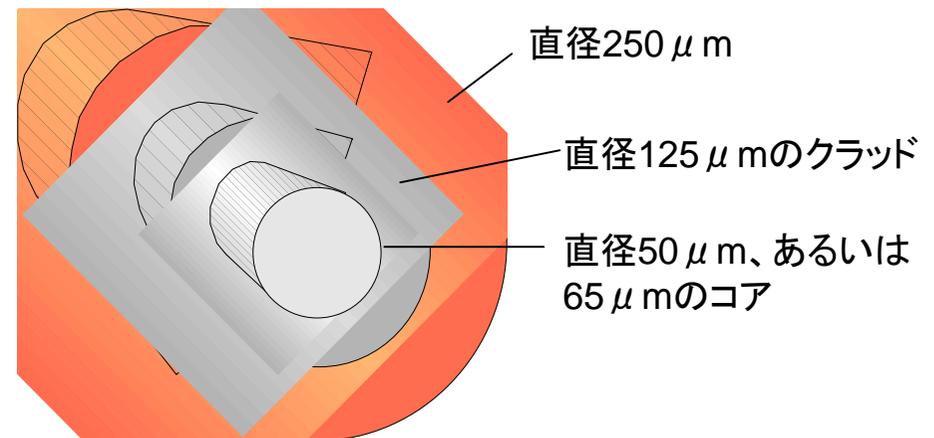
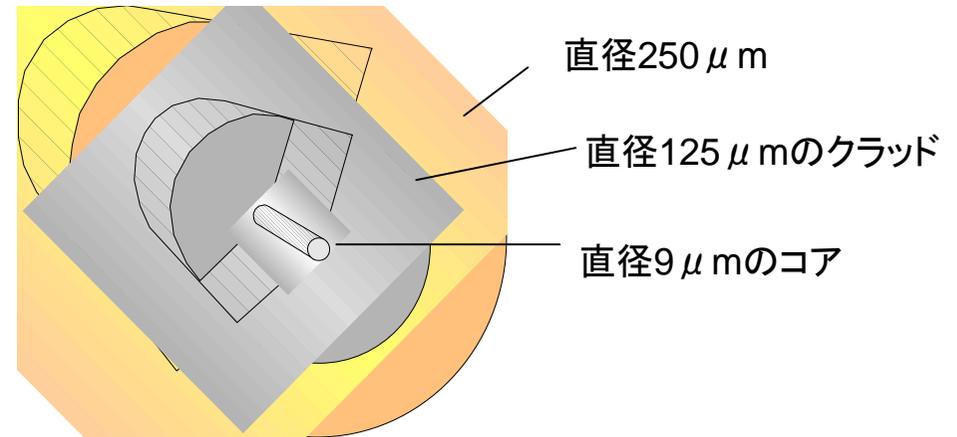


光ファイバーの構造



ファイバーの種類

- シングル・モード・ファイバー (SMF)
 - 直径 $10\mu\text{m}$ 以下の1つのモードのみを伝送するファイバー
 - ケーブル中を進んでいくレーザー光は1つのモードしか存在しないので分散を起こすことなく光パルスを高速に伝えられ長距離伝送ができる
 - ファイバーの材料には純度の高い石英が使用されており折り曲げに弱く高い加工技術も必要
 - LXレーザー
- マルチ・モード・ファイバー (MMF)
 - 直径 $50\mu\text{m}$ 、あるいは、 $62.5\mu\text{m}$ の異なるモードが混在するファイバー
 - 伝送距離も短く低速伝送
 - 材料としてプラスチックを使っているので安価でかつ折り曲げにも強く加工しやすい
 - LED、SXレーザー



ファイバーケーブル・コネクタ

■ SCコネクタ

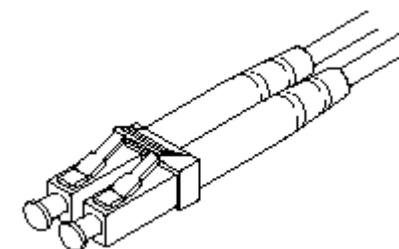
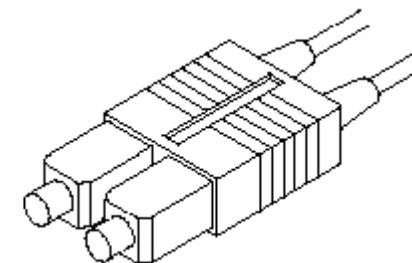
- Gigabit EthernetやFibre Channelで 사용되는標準的なコネクタ
 - IEC 60874-14規格
- 光ファイバー1芯毎に1つのコネクタが必要
 - 送信用(TX)、受信用(RX)の計2本のケーブルを接続

■ LCコネクタ

- Gigabit EthernetやFibre Channelで 사용되는小型コネクタ
- PUSH/PULL着脱可能

■ MT-RJコネクタ

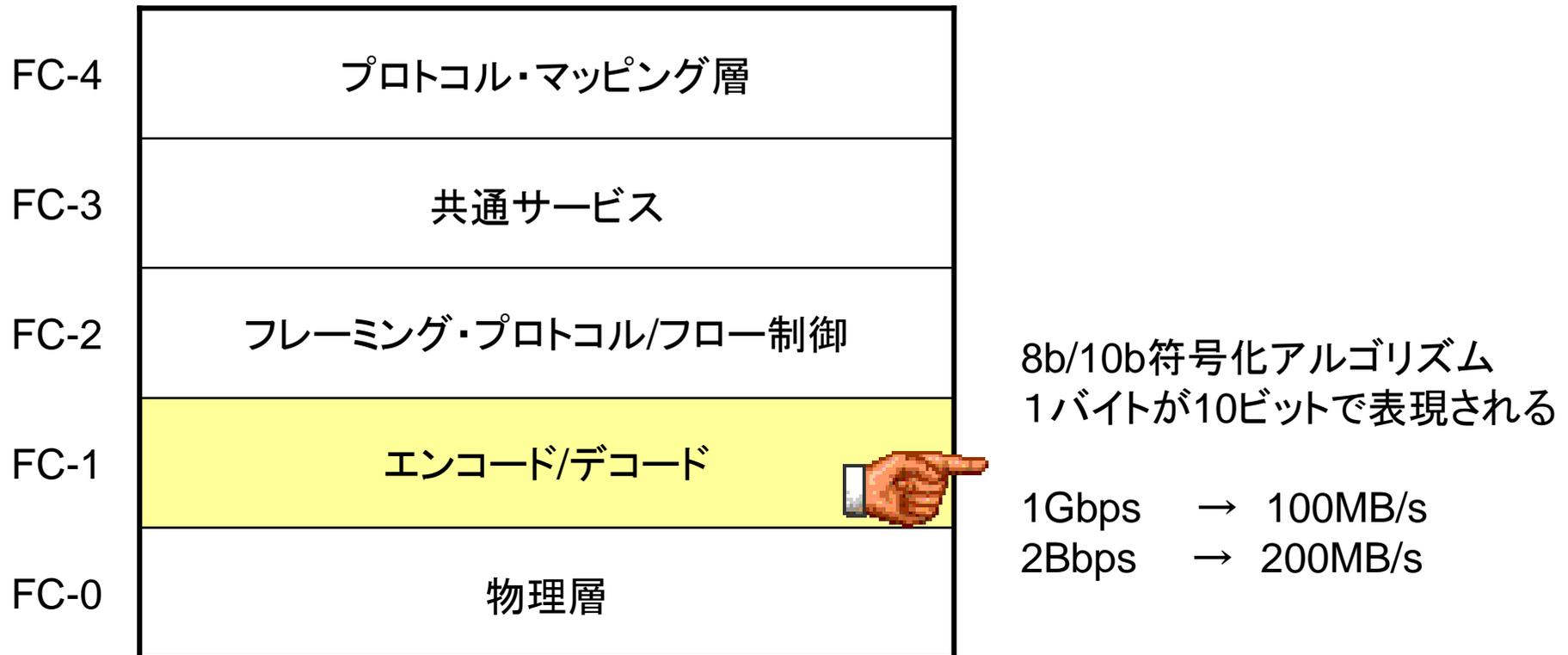
- Gigabit Ethernetで 사용되는小型コネクタ
- PUSH/PULL着脱可能
- 2芯ケーブルを使用するため1つのコネクタで送信・受信可能



ファイバーチャネル規格:FC-1

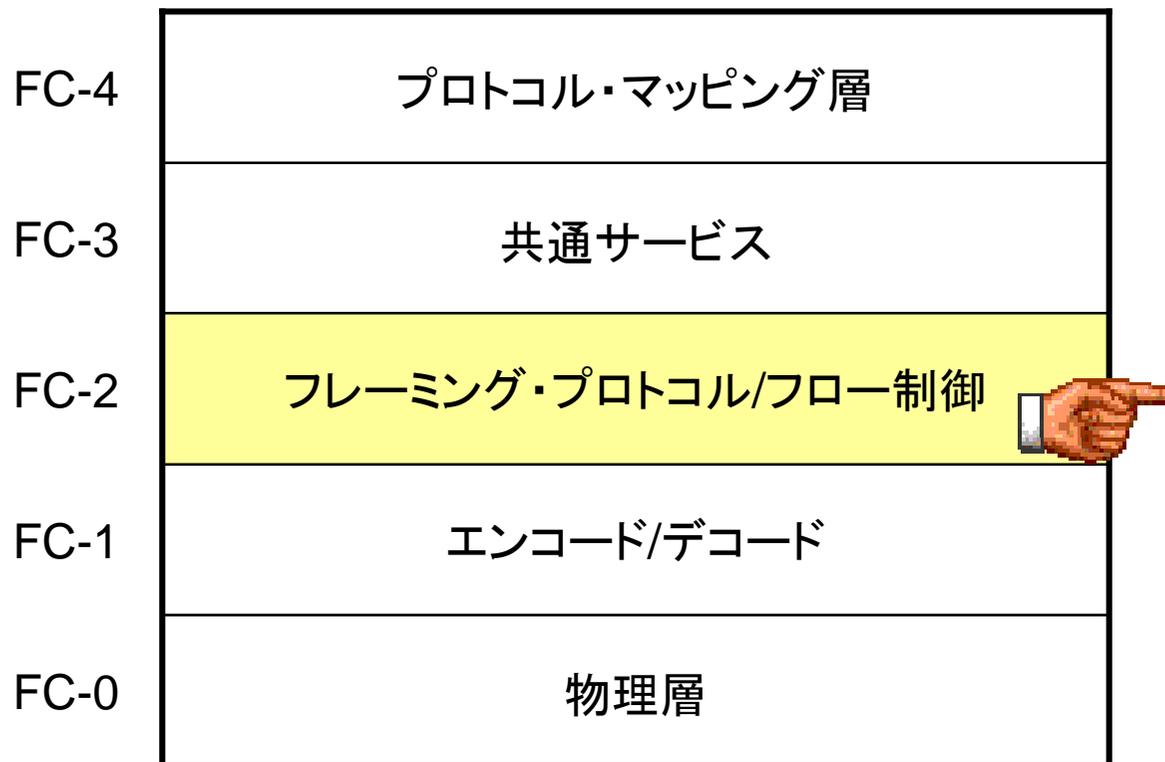
エンコーディング・スキームを定義

データ転送のための同期を取るために使用される



ファイバーチャネル規格:FC-2

フレーム構造とバイト順を含む信号プロトコルを定義
データは、フレーム単位で転送される
フレームは最大2112バイトの可変長



ファイバーチャネルのデータ構造

SOF 4バイト	フレーム・ヘッダ 24バイト	データ・フィールド 最長2112バイト		CRC 4バイト	EOF 4バイト
		オプション・ヘッダ 64バイト	ペイロード 2048バイト		

フレーム・フォーマット

Word	Bits 31-24	Bits 23-16	Bits 15-8	Bits 7-0
0	Destination ID (Address)			
1	Source ID (Address)			
2	フレーム制御			
3	シーケンスID		シーケン・スカウント	
4	Originator Exchange ID		Responder Exchange ID	
5				

SOF Start Of Frame
 EOF End Of Frame

サービス・クラス

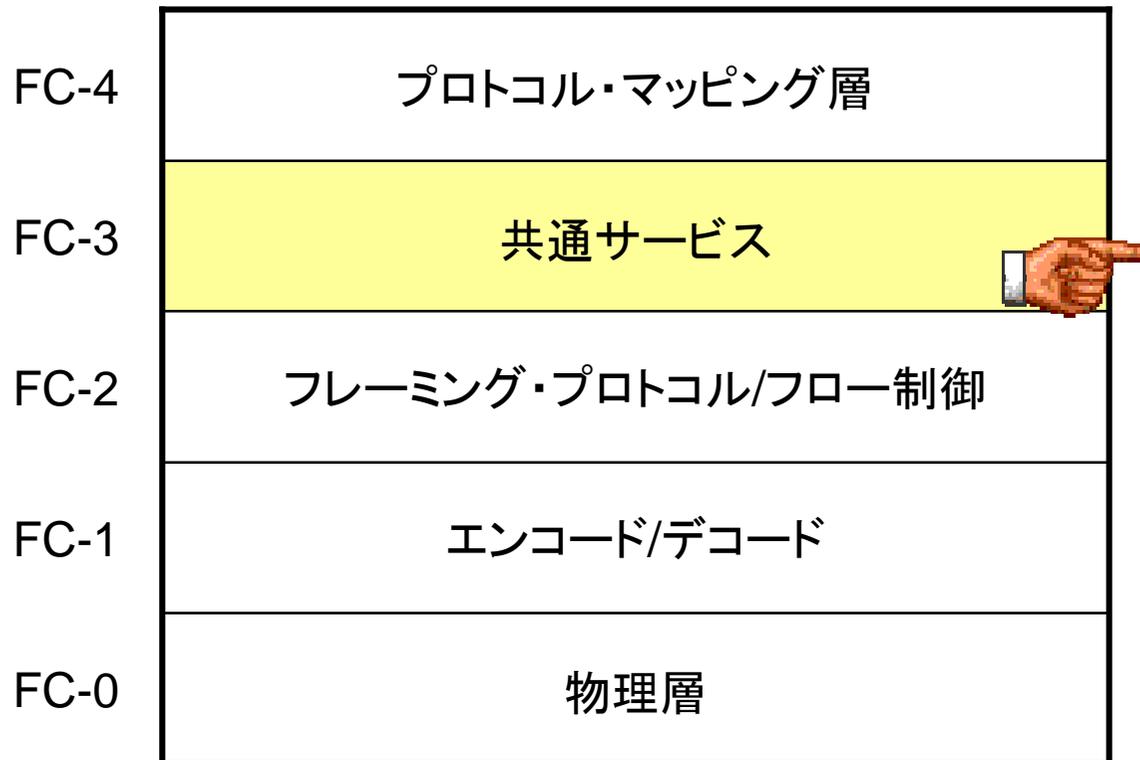
- フロー制御はサービス・クラスにより規定される
- 現在最も利用されているのは以下の3種類
 - クラス1
 - 確認応答ありのコネクション型サービス
 - チャネルプロトコルの機能
 - 送信されるフレーム毎にACKフレームが送り返される
 - 通信するデバイスが帯域幅の全体を使用
 - クラス2
 - 確認応答ありのコネクションレスサービス
 - フレームはスイッチに伝送され、スイッチの都合の良いときに送信
 - 利用可能な帯域幅をデバイスで共有
 - クラス3
 - 確認応答なしのコネクションレスサービス
 - SAN上で最も頻繁に使用されるサービスクラス
 - トラフィックが少ないときは帯域幅をフルに使用、トラフィックが多い時は帯域幅を共有

ファイバーチャネル規格:FC-3

ファイバーチャネルの一般サービス層

ネームサーバーなど

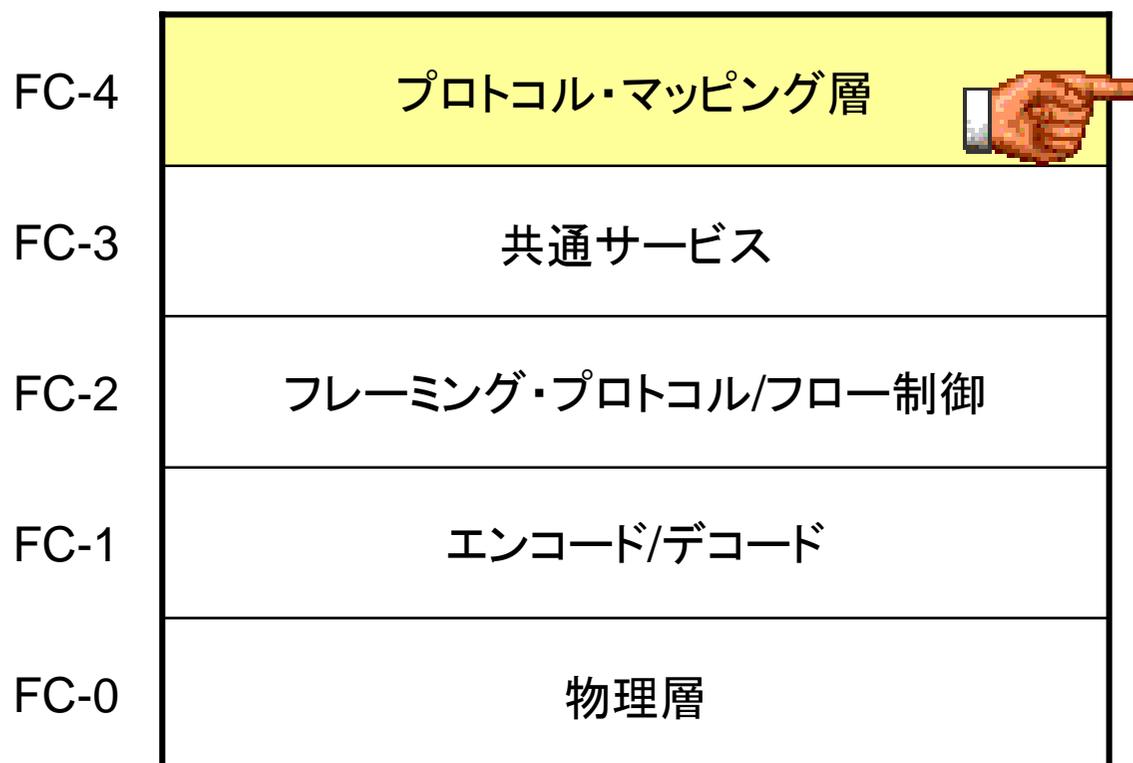
現時点では使用されていない



Fibre Channel上位層プロトコル

ファイバーチャネルのULPマッピング層

SCSI、IP、HIPPI、IPI-3、ATMなどのULP(Upper Level Protocol)をファイバーチャネルを通じて伝送するための規格

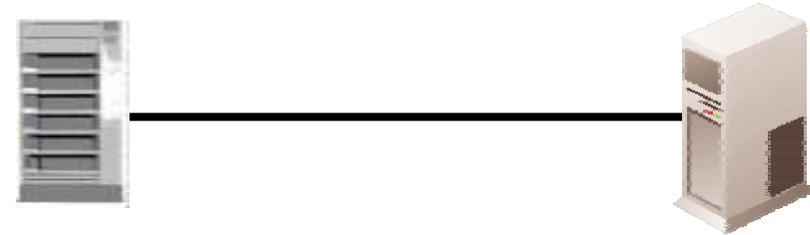


SCSI-FCP

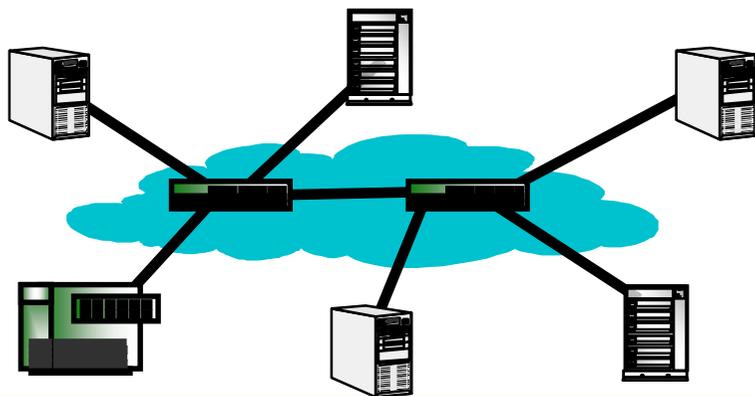
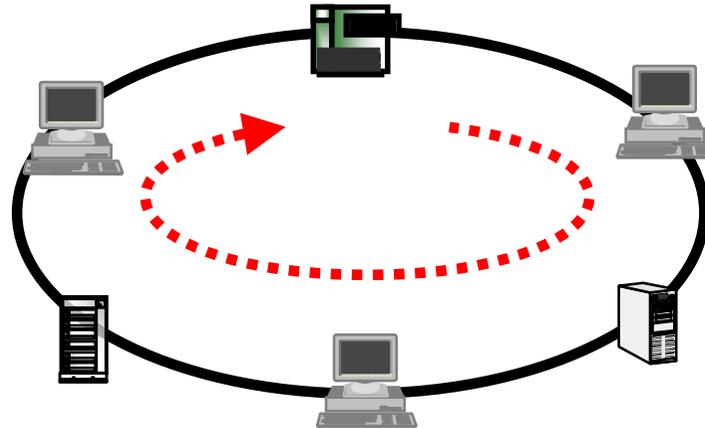
SCSIフレームをファイバーチャネルプロトコルにカプセル化する規格

Fibre Channelの3つの接続方式

Point to Point



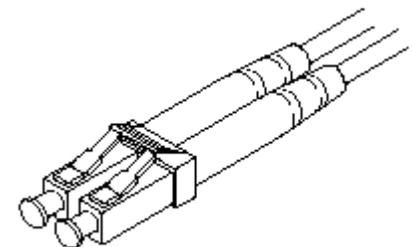
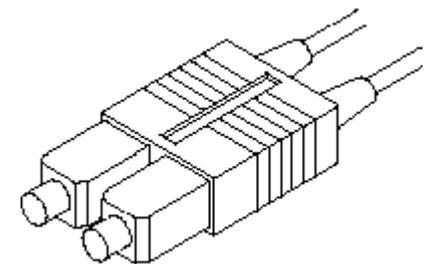
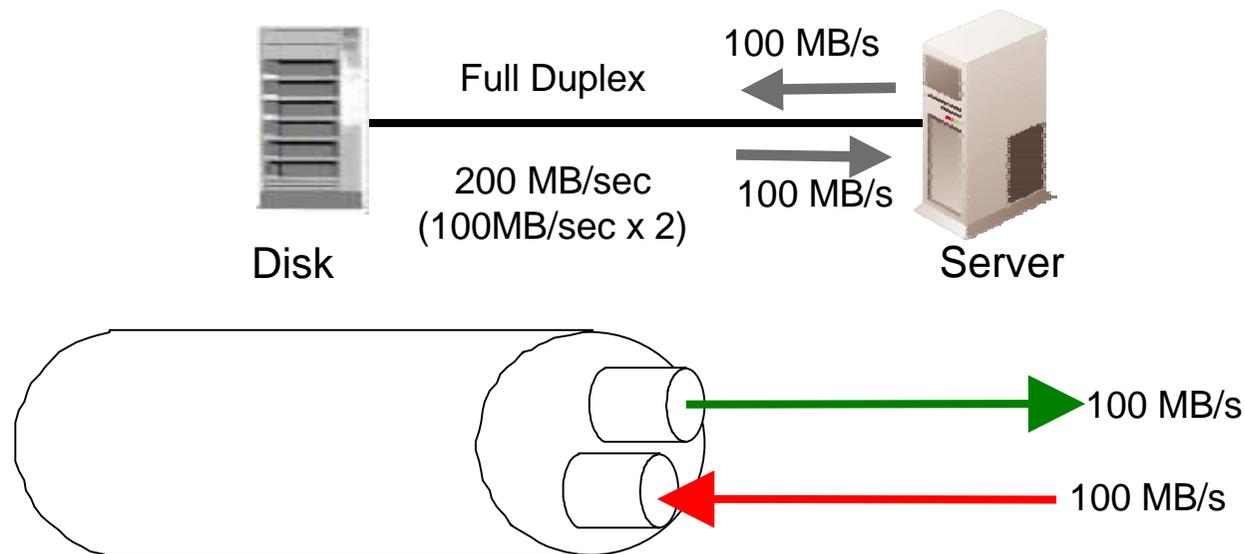
Arbitrated Loop



Switched Fabric

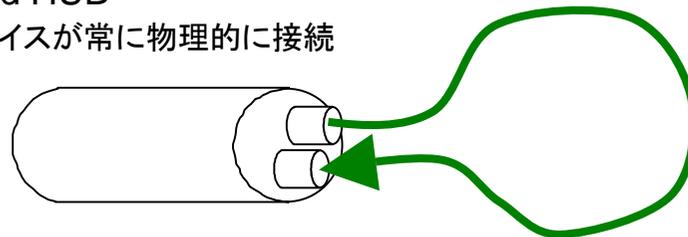
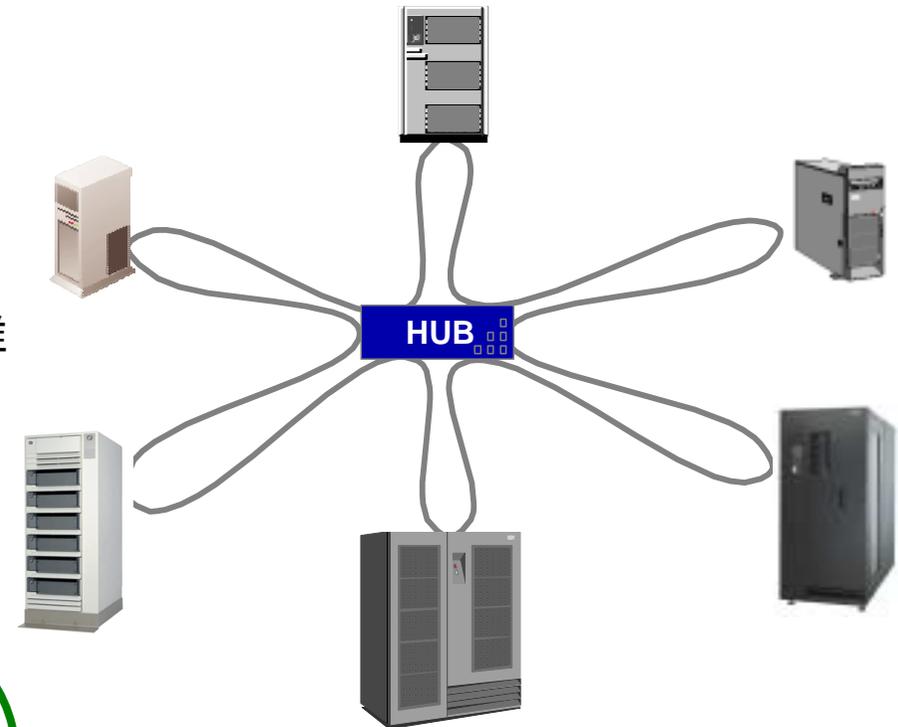
Point to Point

- デバイスをスイッチに直結するために主に使用される
 - 一般にイニシエータデバイスとターゲットデバイスがPoint-to-point型で接続されることはほとんどない
- アドレス指定なし
 - 送信先は常に相手側
- 初期化ルーチンは非常に簡単



FC-AL (Fibre Channel Arbitrated Loop)

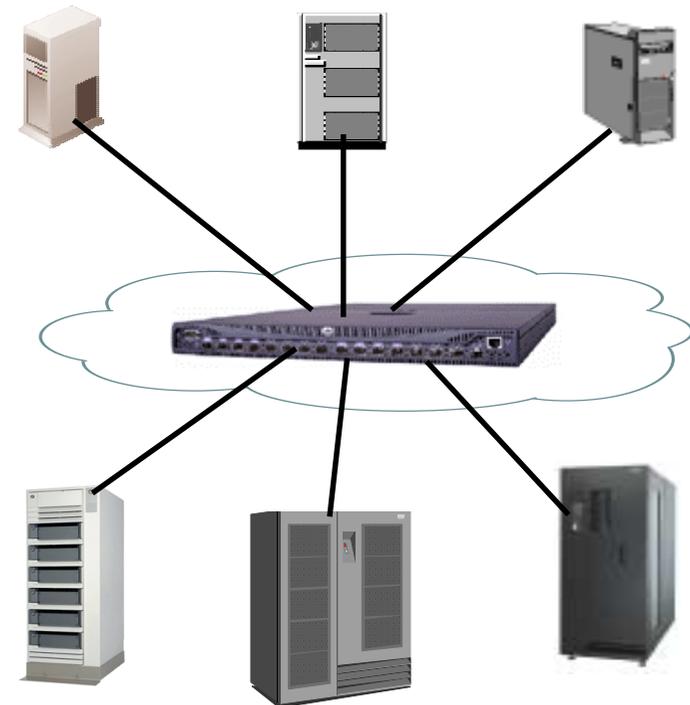
- スイッチなしで127台までのデバイスを接続可能
- ループ上の全てのデバイスが1本のファイバーチャネルの帯域を共有
 - リング状にファイバーを1方向へ接続
- 8ビットのAL_PA (Arbitrated Loop Physical Address) によってデバイスを識別
- ハブとケーブルによる接続
- 性能は以下の要素に依存
 - ノード数
 - 接続距離
 - 負荷
- 2重ループを構成することで高可用性を実現
- ホットプラグ対応だが、初期化プロセスは非常に複雑
 - ループ内の全てのデバイスにAL_PAを割り当て
 - ループへのノードの追加、除去はループリセットが必要
- 2種類のHUB装置が提供されている
 - Managed HUB
 - デバイスから有効なデータが送られるまで接続しない
 - Unmanaged HUB
 - 全デバイスが常に物理的に接続



リングは常に1方向へ送信

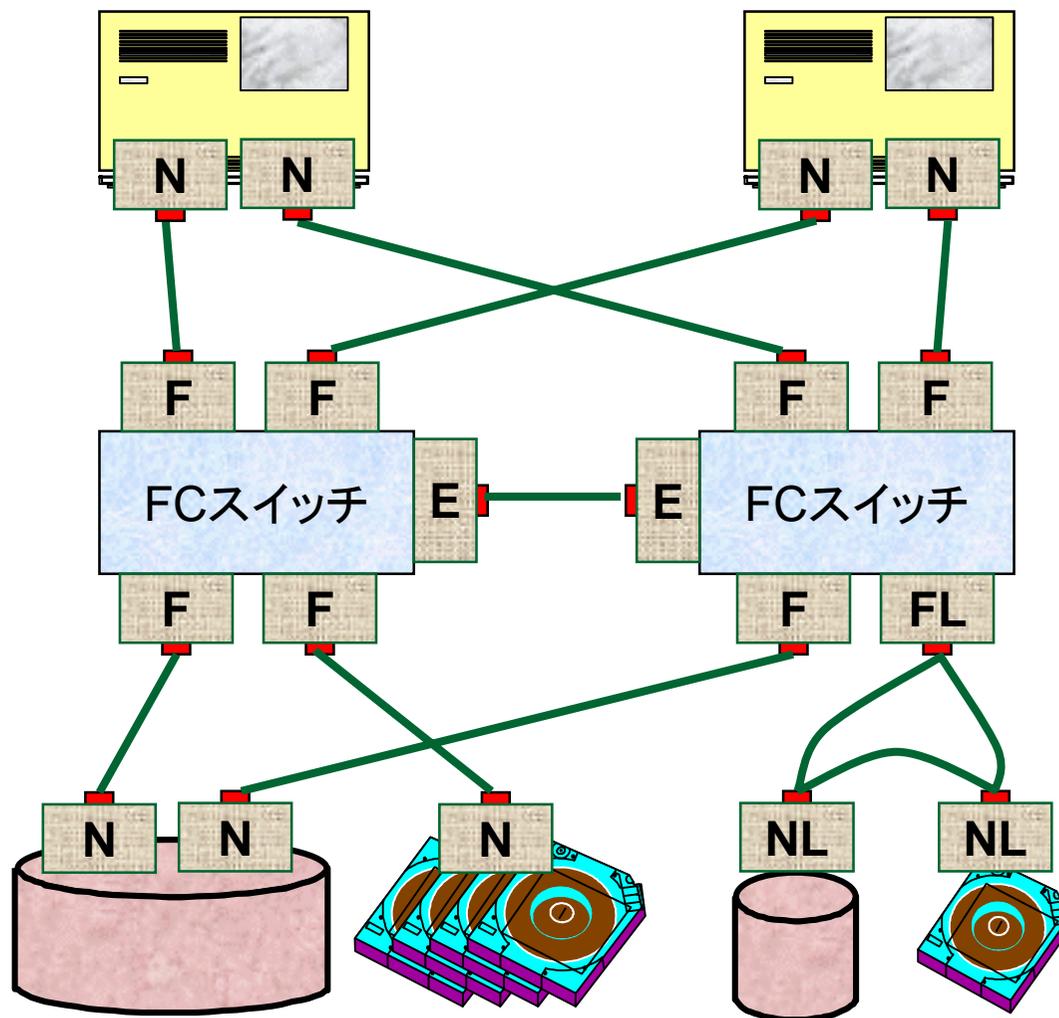
Fibre Channel Switched Fabric

- 高いデータ転送バンド幅
 - ノード当たり100-200 MB/secのバンド幅
 - 双方向転送
- ドメイン内に最大1600万ノードが接続可能
- 接続待ちは経路と負荷状態に依存
- リンクの接続距離は最長10km
- スケーラブルでフレキシブルな再構成
- Fabricの管理が必要



FC で利用されるポートの名称と種類

- F-port
 - スイッチにあり、ストレージ装置、またはサーバーのHBAとの接続に利用
- E-port
 - スイッチにあり、スイッチ間接続に利用
- N-port
 - サーバーのHBAやストレージ装置にあるポート
- NL-port
 - サーバーのHBAやストレージ装置にあるポートで、ループ・トポロジーで利用する
- FL-port
 - スイッチにあり、ループ・トポロジーで利用する



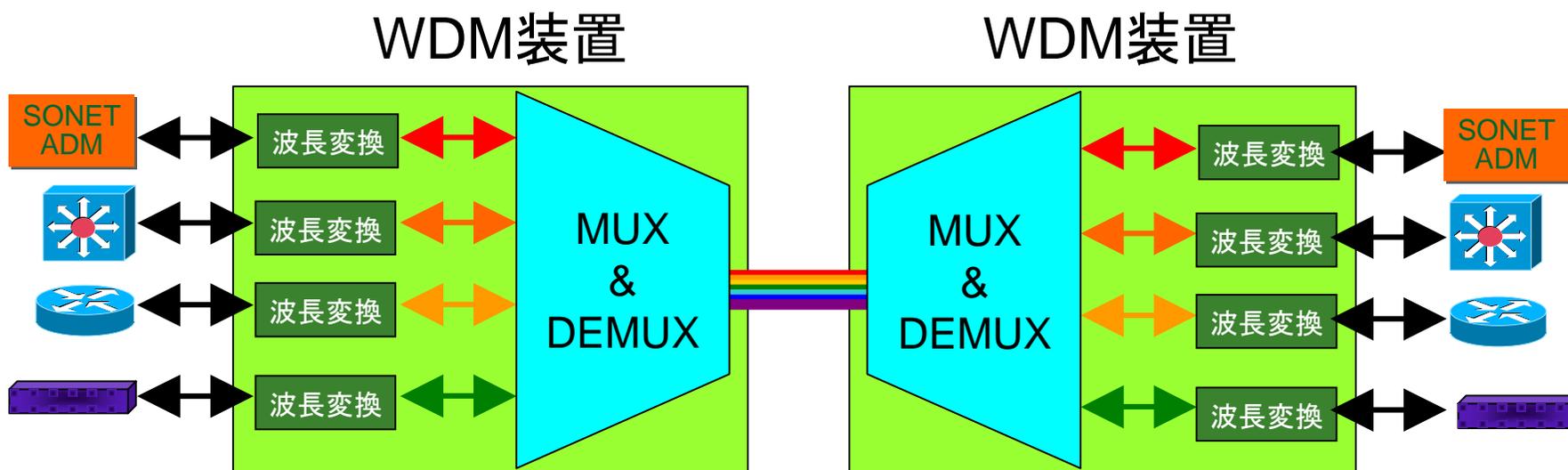
Fibre Channel HBA

- HBA (Host Bus Adapter)
 - サーバーをファイバーチャネル接続するために必須のアダプタカード
 - サーバーの各種バス毎に提供される
 - 現在主流はPCIやPCI-Xバス規格
 - 各OS用のデバイスドライバが必要
 - OSからはSCSIボードとして認識される
 - プロトコル処理などほとんどの処理はカード上で実行されるのでCPU使用率が低い
 - 送信ポートと受信ポートの1対
 - 多重ポート対応HBAも提供されているが、HBAレベルの冗長性を確保するには個別のHBAを用意する必要がある
- 製品例: QLogic QLA2300シリーズ
 - 64bit 66/133MHz PCI-X対応 (標準32/64bit 33/66MHz PCI互換)
 - ファイバーチャネル転送速度最大400MB/s (Full Duplex)
 - ファイバーチャネルビットレート (1Gb/2Gb) 自動切替
 - FC-ALプライベート及びパブリックループ/Point-to-Point/スイッチファブリック接続サポート
 - FLポート及びFポートによるファブリックログイン サポート

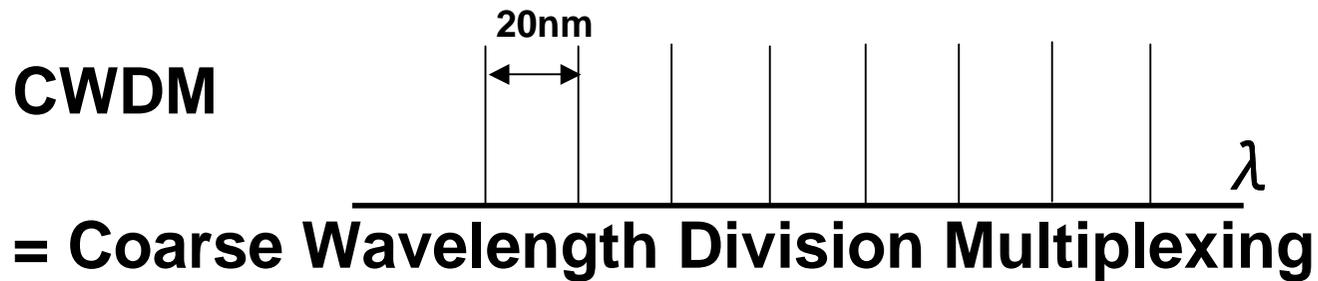
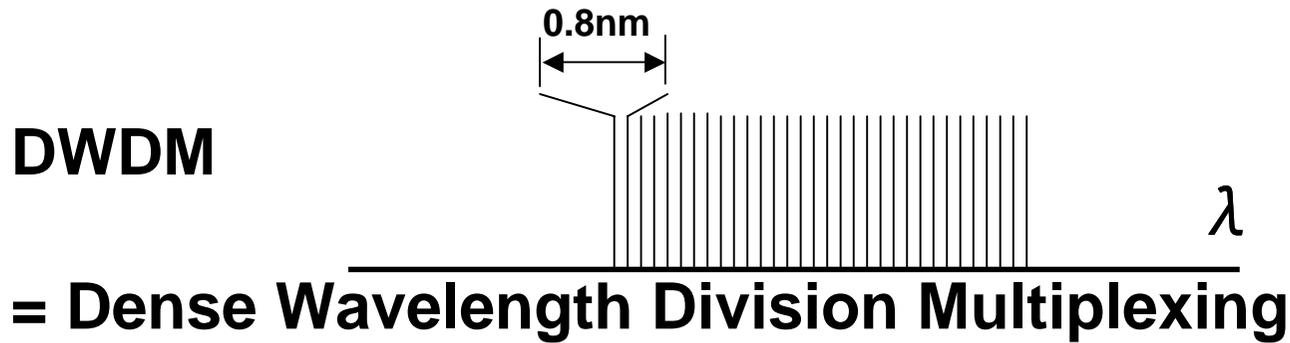


リピータ/WDM装置

- ファイバーチャネル・リピータは、ファイバーチャネルの接続距離制限である10kmを越えるために使用される
- WDM(Wavelength Division Multiplexing)
 - 複数の波長の光信号を合波して1本の光ファイバーに挿入
 - 1本の光ファイバーを伝播してきた複数の波長の光信号を各波長ごとに分波



CWDMとDWDM



	CWDM	DWDM
多重数/伝送容量	4~16 λ / 4~16Gbps	64~128 λ / 1Tbps~
波長間隔/設置環境	波長間隔が広い(10~60nm) 光学系部品が低コスト	波長間隔が狭い(0.8nm) 光学系部品が高コスト
伝送距離/用途	10~50kmの中・短距離向け 都市内の拠点間通信	数100kmの長距離向け 都市間や国内基幹の通信

テープ装置の基礎

テープ装置選択の基準

- 信頼性
 - 信頼性の低いテープ装置を使用したのではテープバックアップの意味がありません
- 容量
 - バックアップ対象のストレージ容量に見合ったテープ容量を確保
 - 1巻あたりの容量
 - テープ・ライブラリー装置を使用し、複数テープを使用したバックアップ
- 読み書き速度
 - バックアップ・ウィンドウ内にバックアップが終了するのか
 - 転送速度は
 - 並列化(複数ドライブの使用)は可能か

代表的なテープ規格

- LTO
 - Linier Tape Open
 - Seagate、HP、IBMが作成した規格
 - オープン・システムにおける業界の標準になりつつある
 - LTO-G1、LTO-G2の2種類がある

- DLT / Super DLT
 - LTOが出るまで、オープン・システムにおける事実上の業界標準テープ 規格
 - Quantam社がドライブを一括製造
 - DLT、Super DLTなどがある

- その他
 - AIT, PetaSite、4mm/8mm DAT、QIC

テープ記録方式の比較

ヘリカル・スキャン(Herical Scan)

- 回転ヘッド
- トラックはテープに対して傾いている
- 高密度化が容易(メディアのサイズが小さい傾向)
- Start/Stopのパフォーマンスが良くない
- テープの走行方向は終始一方向
- メディア、ヘッドが比較的劣化しやすい



8mm,
4mm
DAT
DTF
AIT
PetaSite
など

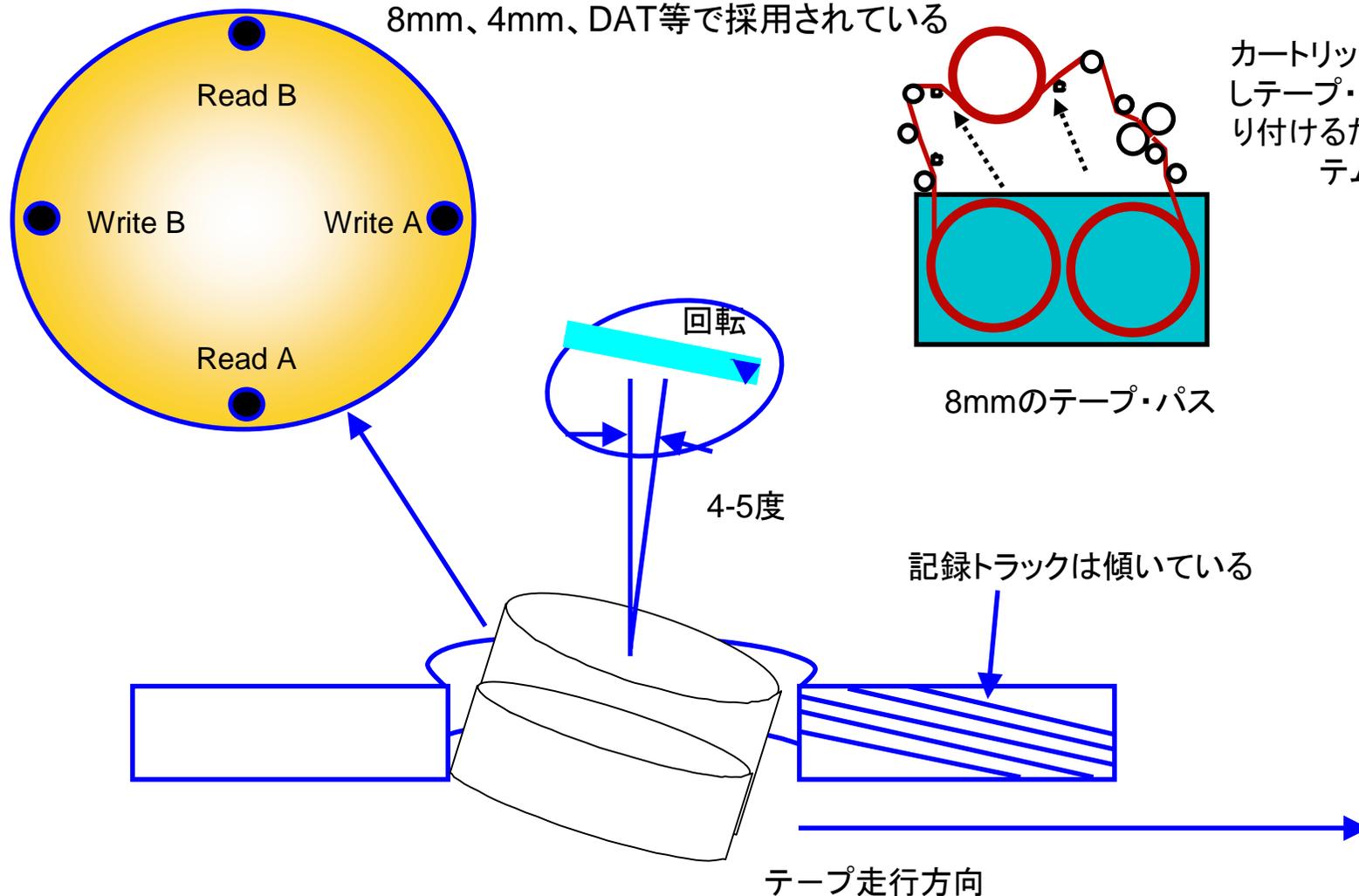
ロンギチューディナル(Longitudinal)

- 固定ヘッド(垂直移動のみ)
- トラックはテープの走行方向にに対して水平
- 高密度(メディアのサイズが比較的大きい)
- Start/Stopのパフォーマンスが良い
- (多くの場合)テープの走行方向は双方向化可能
- メディア、ヘッドが比較的劣化しにくい



DLT
IBM 3570
IBM 3480
IBM 3490
IBM 3590
LTO
など

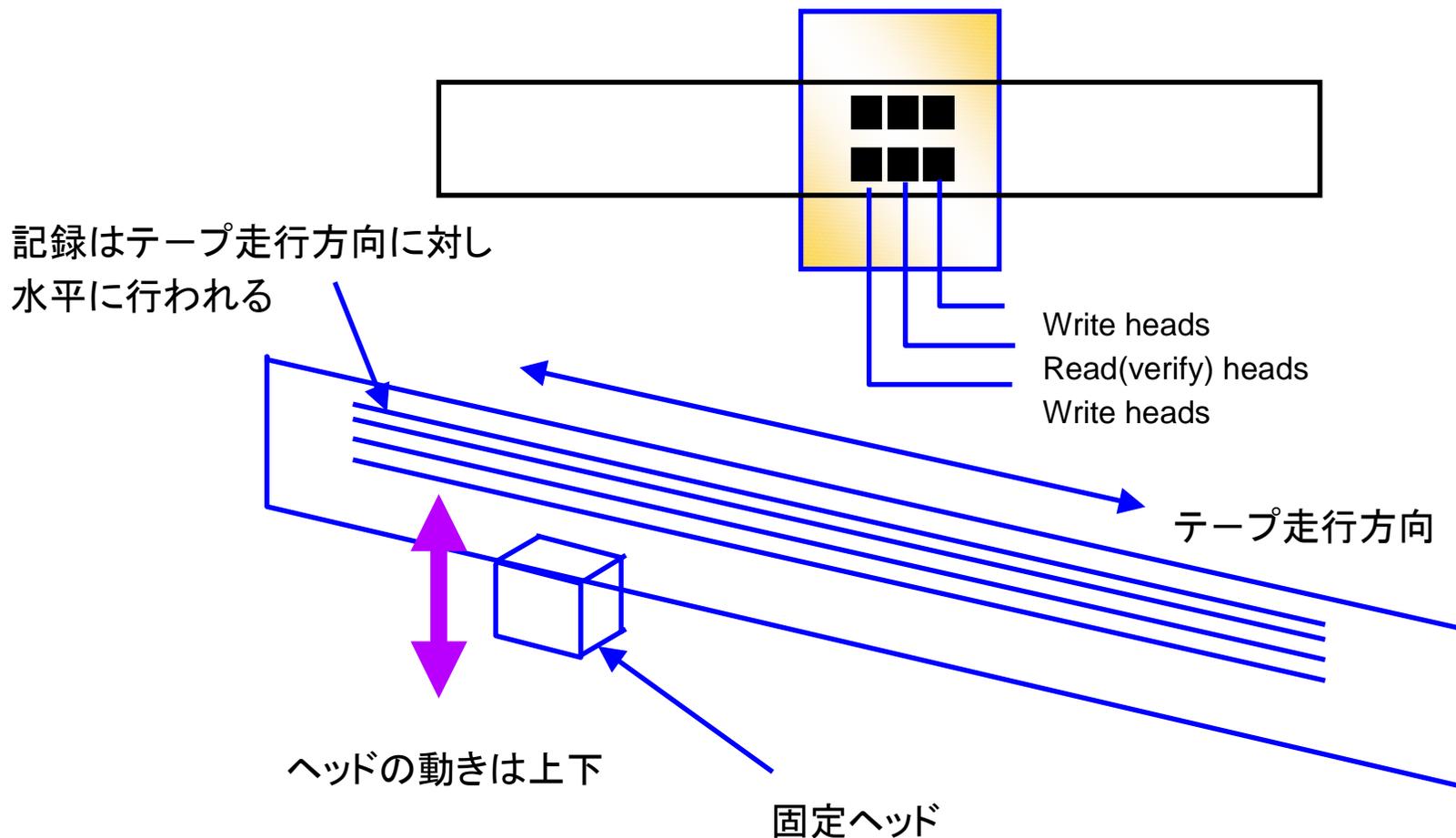
ヘリカル・スキャン方式



カートリッジからテープを引き出しテープ・メディアをヘッドにこすり付けるためローディング・システムが複雑になる

メディアとヘッドやローラーの接触が多いため、テープ・メディアの傷みが早い

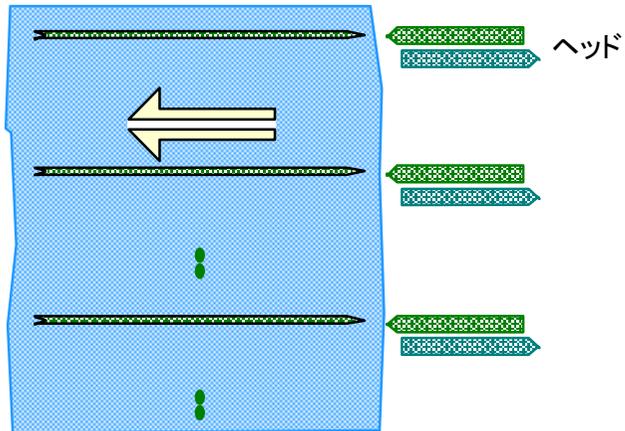
ロングチューディナル方式 (LTOで採用)



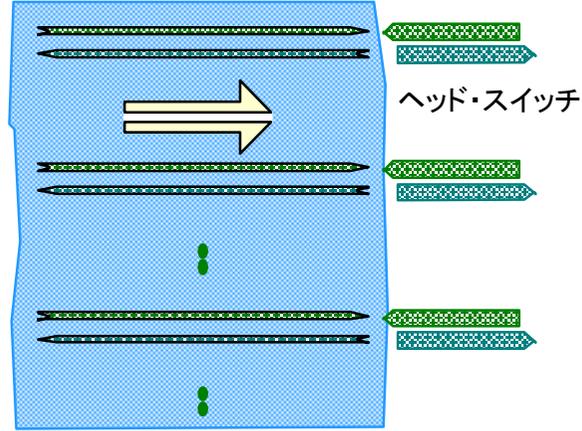
メディアとヘッドやローラーの接触が少ない、耐久性の高いテープ装置が開発可能

マルチトラックロングチューディナル記録方式

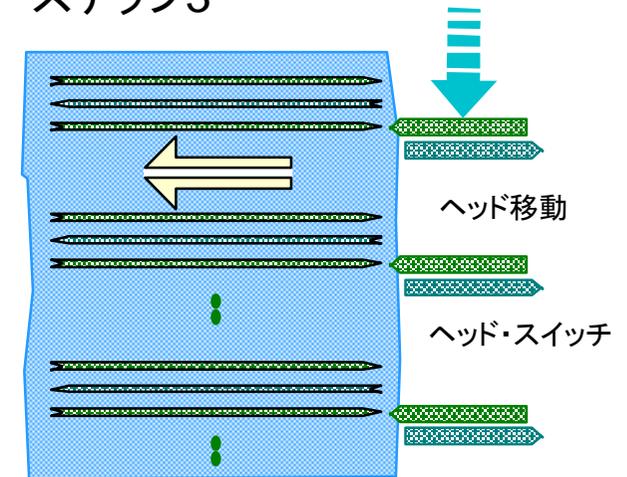
ステップ1



ステップ2



ステップ3

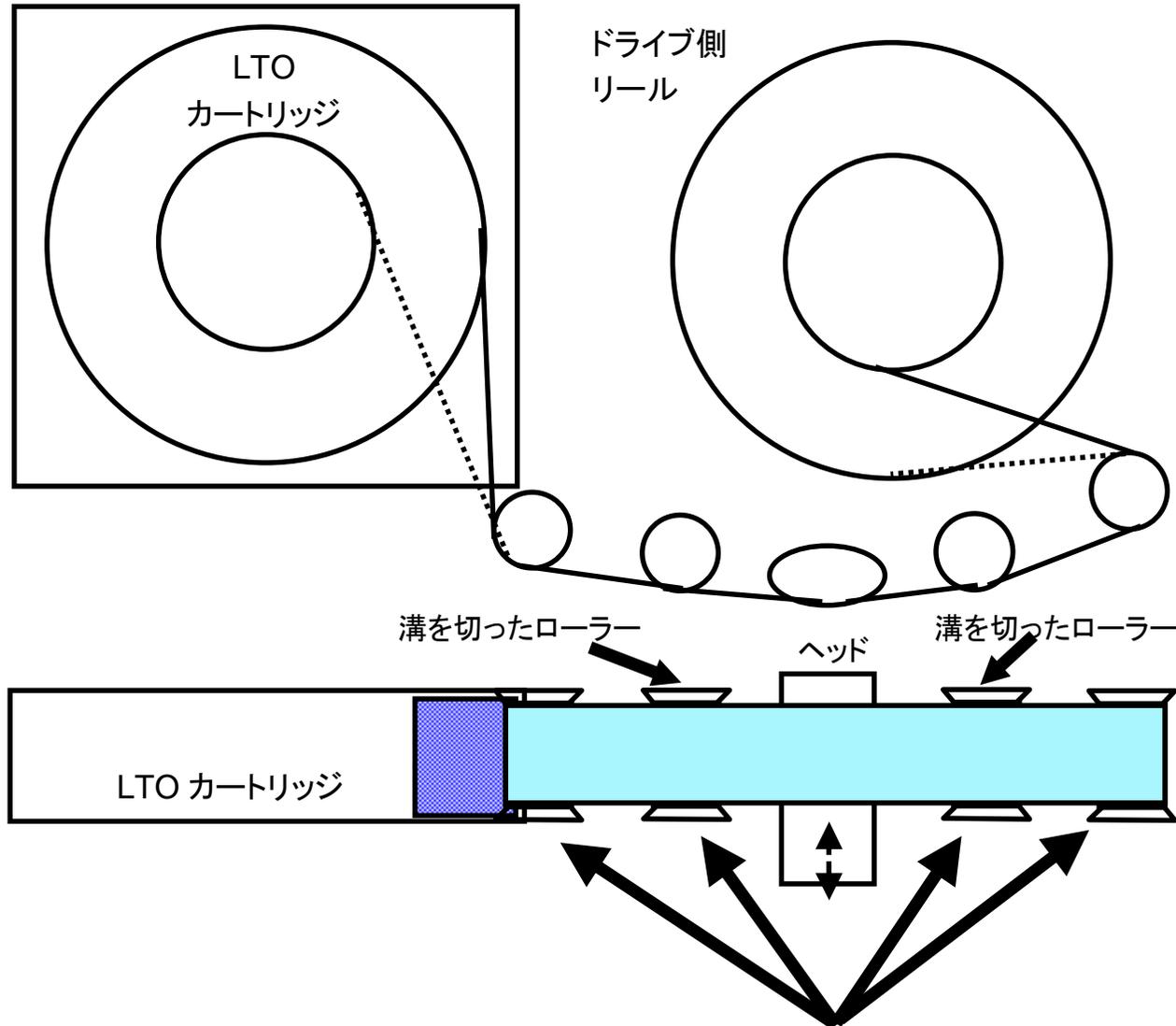


- 1時点では、一方方向にのみ16トラックでライト
- 両方向にリード／ライト可能
- ヘッドの移動によりトラックを移動させる
- IBM 3480 : 18トラック (200MB/巻)
- IBM 3590-E : 256トラック (30GB/巻)
- IBM LTO Ultrium: 384トラック (100GB/巻)
- IBM LTO Ultrium 2: 512トラック (200GB/巻)

特長

- 高信頼性
 - シンプルなテープ・パス
 - ヘッド、メディアのダメージを軽減
- 高速なストップ・スタート処理
- 記録容量を増加

LTOの先進技術：Surface Control Guiding



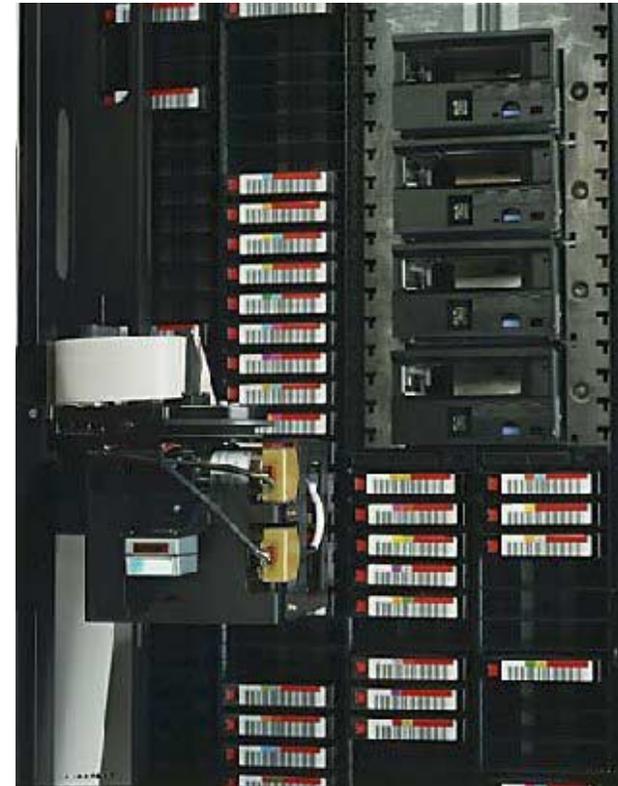
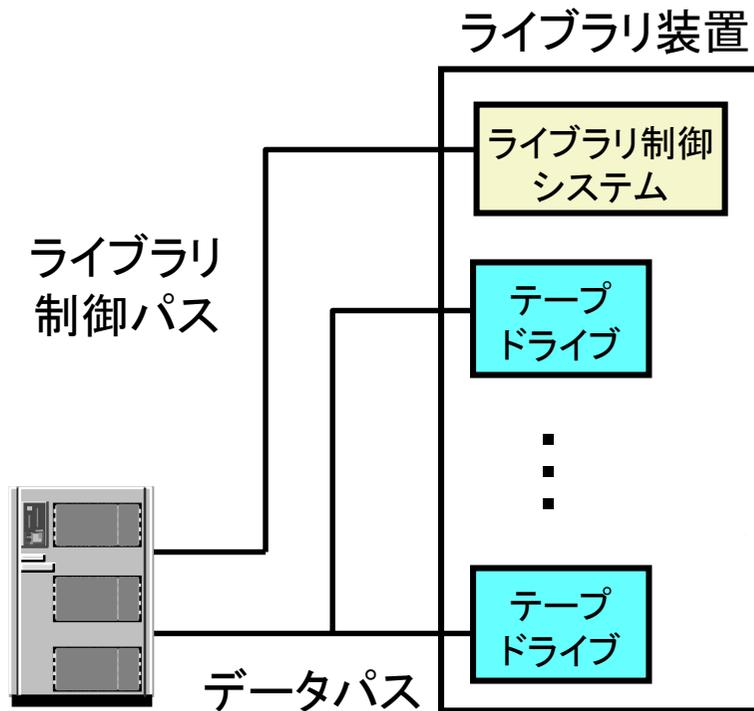
- テープ表面を利用してテープ走行のずれを修正
- 旧来の方法 (Edge Guiding) と比較して、テープ・エッジへの負荷やダメージを大幅に削減。
- テープ走行の高速化と高密度化に大きく貢献



ローランド・アプリケーション・ガイド
日本アイ・ビー・エム株式会社

テープ・ライブラリー装置の概要

- 複数のテープドライブとライブラリー制御システム接続から構成



テープ装置の接続インタフェース

- SCSI LVD
 - サーバーのローカルバックアップやバックアップサーバー直付け
 - IAサーバーで主に使用されているSCSIインタフェース
- SCSI HVD
 - サーバーのローカルバックアップやバックアップサーバー直付け
 - UNIXサーバーで主に使用されているSCSIインタフェース
- ファイバーチャネル
 - 主にSAN接続
 - 主にテープ・ライブラリー装置向け

ちょっと休憩

