

技術文件

# 企業 SSD 介面比較

## 序言

PCI Express (PCIe) 是同時用於用戶端和企業運算應用程式的一般用途匯流排介面。現有的大量儲存介面 (SATA、SAS) 會透過連接到 PCIe 介面的主機介面卡連接到主機電腦。

SATA 介面設計為硬碟機 (HDD) 介面，SAS 介面則設計為裝置介面和儲存子系統介面/基礎架構兼具。由於硬碟機與系統需求都已有相當發展，需要更快速的介面與新功能，SATA 和 SAS 介面已經過了好幾次的修改。

固態硬碟機 (SSD) 很快就為這些介面增添了可觀的新效能需求，因為 SSD 的資料傳輸速率已從過去每秒數十至數百 MB，變成了目前的每秒數千 MB。除了資料傳輸速率的提升之外，由於 SSD 去除了機械動作，因此也增加了本身可以執行的每秒輸入與輸出作業 (IOPS) 數目。

此一發展使得各方需要能夠更完善地實現現有標準，同時加強現有的介面標準，以管理新的效能需求並保持與現有系統架構的相容性。

此文件將討論不同的介面，並對照發生的各種效能與相容性權衡的情況。

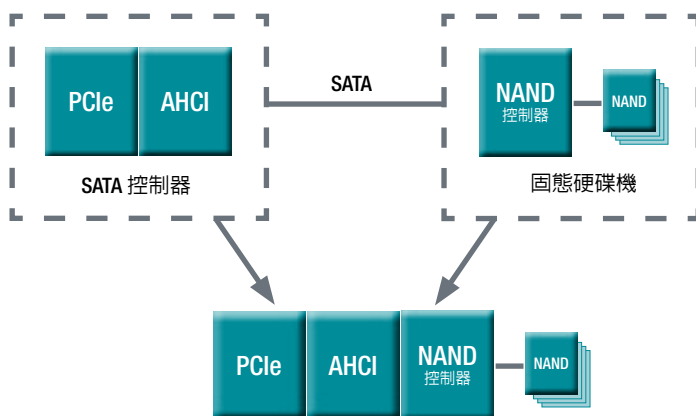
# 企業 SSD 介面比較



## SATA 介面

SATA 是專為透過連接線或印刷電路板 (PCB) 佈線進行之點對點連線設計的低成本介面。主機連線會連至進階主機控制器介面 (AHCI)，此介面一般位於主機晶片組，功能是做為 PCIe 匯流排上的主機介面卡。此介面有一些設計問題，可能會造成每個指令 1 $\mu$ s (或更長時間) 的匯流排佔用情形。對於速度為 10 $\mu$ s 完成 4KB 傳輸的硬碟機而言，這並不是重要問題，但 SSD 可於 2 $\mu$ s (或更短時間) 傳輸 4KB 的資料—因此佔用情形會變得嚴重，且使得 SATA 介面無法真正發揮高效能大量儲存介面的功能。

SATA 仍然很適合用做主要考量因素為成本 (而非效能) 的低成本 SSD 介面。SATA 架構也能夠整合到管理 SATA 指令集的主機介面卡，而不需實際包含實體 SATA 介面 (PHY) (圖 1)。



來源：Seagate Technology, 2011 年  
圖 1：架構合併

## SAS 介面

SAS 也是序列介面，它會透過主機介面卡連接到主機，但存在一些明顯的差異，使其適合做為 SSD 介面：

- 硬體經常性費用較低
- 傳輸速率更快
- 並連埠
- 有效的控制器驅動程式介面

此外，SAS 包括 (SATA 中沒有的) 可提升連接介面裝置之可靠度與可用性的功能：

- 健全的序列通訊協定
- 多重主機支援
- 端對端資料完整性
- 雙連接埠功能
- 高度並行與整合

## 硬體經常性費用較低

沒有與 SATA AHCI 控制器同等級的 SAS 通用主機介面。相反地，有多個供應商在以效能為主要考量的 SAS 主機介面卡市場中相互競爭 - 除了個別硬碟機介面，也包括整合多重硬碟機主軸傳輸速率以達更佳傳輸速度的多種 RAID 系統介面。此外，SAS 主機介面卡是專為管理較高效能 SSD 和 HDD (例如使用短程尋軌的 15K RPM 硬碟機) 而設計的。由於硬體主機介面卡和管理該主機介面卡的裝置驅動程式是設計為整體系統，專為 SSD 提供的新設計已可供使用，且進一步提升了傳輸速率和 IOPS。

## 傳輸速率更快

SAS 連接埠目前支援高達 6Gb/秒的資料傳輸速率。LSI 和 PMC-Sierra 等公司目前正在抽樣試驗支援 12 Gb/秒資料傳輸速率以及 2 百萬次以上 IOPS 的發展中設計，未來速度甚至可能達到 24 Gb/秒。

## 並連埠

並連埠的概念是 SAS 架構所固有的，亦即整合多個連結以在一或多部主機與裝置之間提供多個同步路徑。目前的 SAS 硬碟機接頭為硬碟機定義了兩個連接埠。基於設計理由，目前的硬碟機不支援並連埠—僅支援雙埠，其中每個連接埠都具有防止進行並連埠組態的不同 SAS 位址。

接受的 SAS-3 (12 Gb/秒) 設計提案可以將硬碟機上的連接埠數目增加為四個，所有連接埠都能連接到相同網域，或兩兩成對連接到不同的網域。極少 SSD 能支援雙連接埠裝置上的雙埠，並同時支援並連埠。

# 企業 SSD 介面比較



## 健全的序列通訊協定

SAS 序列通訊協定以提供序列傳輸端和接收端的校準。如此可以透過補足通道長度、抗阻失配和符際干擾，增進連接線或背板上的訊號品質。SAS 序列通訊協定也會管理硬體層級的錯誤偵測和重新傳輸。如此可以更快復原訊號斷續問題。

## 多重主機支援

SAS 介面和交換結構能讓多個主機存取相同的裝置。這項功能可以用來管理主機故障以及資料路徑故障，以改善資料可用性。

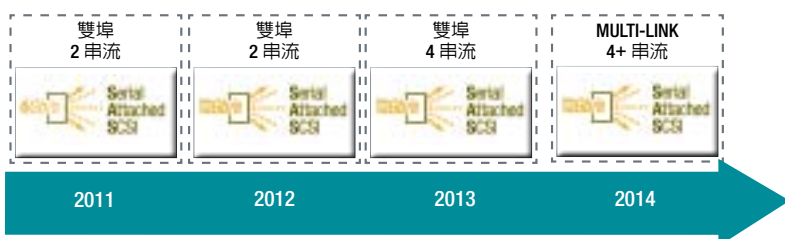
## 端對端資料完整性

從資料建立於主機資料緩衝區到跨 PCIe 介面與 SAS 介面傳輸，直到儲存於裝置並再次讀取和傳輸到主機資料緩衝區的過程中，SAS 介面可以透過資料周期備援檢查 (CRC) 確認資料完整性。如此可以在應用程式到 RAID 控制器的整個途徑以及裝置上設置多個檢查點。此功能有時亦稱為防護資訊 (PI)。

## 雙連接埠功能

SAS 目標裝置支援雙連接埠作業。如此可讓您建立兩個容錯網域 (fault domain) 並提供更高的可用性。即使其中一條連接埠路徑發生故障，而無法沿該途徑進行存取，仍可以使用第二個連接埠存取裝置。

以往 Seagate 曾經帶動市場的介面採用。Seagate 目前正與 SCSI 商業協會 (STA) 及其他業界領導者合作運用部署廣泛的現有 SAS 基礎架構 (圖 2)。表 1 說明 12 Gb/秒 SAS 與 MULTI-LINK 對系統建置商和一般使用者組織有何助益。



來源：Seagate Technology, 2011 年  
圖 2：SAS 介面發展

## PCI Express 介面

PCI Express (PCIe) 是系統中將週邊裝置連接到主機處理器，並通過記憶體控制器到達記憶體結構的基本介面。先前所述的 SATA 和 SAS 介面會透過 PCIe 介面 (或主機介面卡) 連接到主機處理器和記憶體。

表 1：SAS 的優點與建議 Multit-Link 加強功能 SAS

表 1：SAS 的優點與建議 Multit-Link 加強功能 SAS	
多重連結 (BW)	X4 (4x600 MB/秒)
可用電源	25W (2.5 吋)
總等待時間	極低
多重主機通訊協定	是
高可用性	是 (雙埠)
擴充性	極佳
經驗證的健全通訊協定堆疊	是
可維修的熱抽換式	是
與現有的管理軟體相容	是

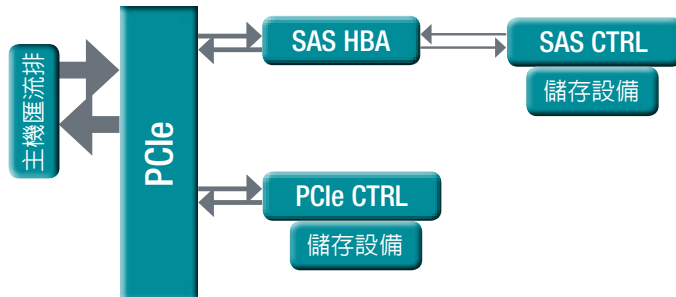
來源：Seagate Technology, 2011 年

PCIe 介面是原始 PCI 介面序列式實作，能夠在週邊設備和主機處理器/記憶體之間提供並列式位址/資料連線。PCIe 介面會跨一或多條通道 (每個通道皆由一個傳輸和一個接收序列式介面組成) 通訊。可以使用多達 32 個通道將主機連接到裝置。每個通道上的序列式資料傳輸速率會視執行的 PCIe 標準版本而定，最新版本為 3.0 且資料傳輸速率約為 1 GB/秒。

對於 1U 伺服器，PCIe 介面是設計為在 (用戶端) 主機板上運用單一連接器，或在 (伺服器) 主機板上使用具有兩個連接器的右側角度轉接器。也可以使用接線系統 (雖然較少使用)。2U、4U 或 7U 伺服器有更多個 PCIe 插槽，類似用戶端實作。PCIe 規格也會使用傳輸端 (和接收端) 校準來配合配置抗阻變化調整，但會當作較 SAS 短的傳輸通道。

PCIe 交換器可以容納單根 I/O 虛擬化 (SR-IOV) 和多根 I/O 虛擬化 (MR-IOV) — 它們是用來在具有單一或多部主機的虛擬 (Hypervisor) 系統中改善控制器效能的方法。SR-IOV 正開始在介面卡上普及運用；然而，VMware 可能尚無法從中獲益。介面卡一般不支援 MR-IOV。

# 企業 SSD 介面比較



來源：Seagate Technology，2011 年  
圖 3：SAS 介面發展

使用 PCIe 介面連接的儲存裝置會透過直接登錄連線，或是透過會經由額外接線或背板型介面連接到裝置的主機介面卡。

目前兩種架構都有多項不同實作。SATA 會在需要不同 AHCI 驅動程式但對應到相容 IDE 舊式實作的 Intel 或 AMD AHCI 系統晶片組 (南橋) 中使用主機匯流排配接卡實作。這些介面也會實作不同的 RAID 管理功能。

SAS 有多個 HBA 供應商提供額外擴充器和 RAID 控制器，這些項目都使用專利的裝置驅動程式和 BIOS，以滿足各種效能和配置需求。

PCIe 驅動程式控制器介面是以 NVM Express 規格以及提議的 SCSI over PCIe (SOP) 規格實作。

前述的 SATA 整合架構是 PCIe 直接登錄連線的另一個例子。

## PCIe SSD 現今情況

目前市場上有兩種主要的 PCIe SSD 類型：原生與整合商。原生控制器會連接主機 PCIe 匯流排，然後直接控制多個快閃記憶體匯流排。這些匯流排一般會使用製造商專利且僅適用於特定裝置的軟體介面。部分上述實作會將位址轉譯和其他功能的負擔加諸於主機 CPU 和記憶體。當裝置用於工作量較高的情況下時，如此會連帶造成應用程式系統資源降低。此外，這些獨特的硬碟機和硬體組合在市場上屬於新產品，有時容易發生不穩定，因為它們的產業生態系統仍在發展中。

整合商機型需要使用不同的設計方法。此方法會使用連接多個 SAS 或 SATA SSD 的現有 SAS 或 SATA RAID 控制器。這些項目會一同封裝於單一 PCIe 卡。RAID 控制器會整合多個裝置的效能以提供高效能層級。這些設計以現有的經驗證企業級硬體和軟體介面為開發基礎，非常穩定和成熟。此外，這些設計使用可執行位址轉譯和其他功能的智慧型控制器，能讓您透過應用程式充分使用系統 CPU 週期和記憶體，即使處於較大 I/O 工作量的情況下也一樣。

## PCIe SSD 的未來發展

SOP 和 NVMe 方法在架構方面很類似。然而，NVMe 是由產業工作團體進行開發，而 SOP 則是由經認可開放式標準論壇進行開發的。NVMe 僅適用於揮發記憶體裝置，而 SOP 則適用於主機匯流排配接卡和具備橋接不同 SOP 裝置之功能的 RAID 控制器。此外，SOP 會大幅運用現有產業架構和功能，而 NVMe 則使用極為有限的新指令集與佇列介面。

## 介面優勢與問題

每項前述的儲存架構都各有優點與問題。視整體系統設計而定，使用特定架構的優點可能會凌駕於該架構的相關問題，要做出恰當的決策有賴於徹底的分析。該決策也必須將與現有系統設計的相容性納入考量。

例如，要更新配備現有 2.5 吋 SATA HDD 與 SSD 的筆記型電腦系統，只能使用相同實體大小的 SSD 以及相同 (或更新) 的 SATA 介面。在此案例中，SSD 的速度會有所限制；超過現有主機 SATA 介面速度將不會加入系統效能。

在類似的情況下，使用短程尋軌 15K-RPM SAS HDD 儲存資料庫索引的企業伺服器可以使用 SAS SSD 升級，如此將會增進整體系統效能，但當一些其他系統因素成為新瓶頸 (CPU、記憶體、網路、配接卡等) 時，即達到效能增加限制。

在新的系統架構中，加入固態儲存設備可以大幅提升系統效能，但僅限於系統架構其餘部分能夠容納增加之資料傳輸速率與資料頻寬的範圍內。要增加 SSD 的資料傳輸速率也需要為裝置提供更高電源，且不論 SSD 安裝於何處，都需要更佳的散熱功能。

# 企業 SSD 介面比較



表 2：本地與整合商 PCIe SSD 比較

	原生	整合商
指令/傳輸	專利 (主機/主控體中使用 FTL)	SCSI 或 SATA (多重 SSD，內建控制器)
委員會	無	無
標準型	否	是
快閃記憶體效能	高	高
CPU 超載	高	低
簡短佇列等待時間	極低	低
深層佇列等待時間	中	低
使用個案擴充	否	是 (RAID、HBA 等)
成熟度	發展中	以經過驗證的產業架構為基礎
企業功能設定 (防護資訊、安全性、管理等)	否	取決於實作

1 FTL：快閃記憶體轉換層

來源：Seagate Technology, 2011 年

表 3：SOP 與 NVMe PCIe SSD 比較

	SOP <sup>1</sup>	NVMe <sup>2</sup>
指令/傳輸	SOP/PQP <sup>3</sup> (控制器中使用 FTL)	NVMe/NVMe (控制器中使用 FTL)
委員會	T10/INCITS <sup>4</sup>	業界工作團體
標準型	是 (ANSI/ISO)	否
快閃記憶體效能	極高	極高
CPU 超載	低	低
簡短佇列等待時間	極低	極低
深層佇列等待時間	低	低
使用個案擴充	是 (RAID、HBA 等)	否 (僅限 NVM)
成熟度	以經過驗證的產業架構為基礎	TBD
企業功能設定 (防護資訊、安全性、管理等)	完整支援	有限責任保固

<sup>1</sup>SOP：SCSI over PCI Express

<sup>2</sup>NVMe：非揮發性記憶體聯盟

<sup>3</sup>PCle 排序介面

<sup>4</sup>INCITS：資訊技術標準國際委員會

來源：Seagate Technology, 2011 年

另一個因素是作業系統裝置驅動程式可用時機和這些新 SSD 介面的 BIOS 支援，以及軟體起始可靠度。

## 介面與快閃固態硬碟機等待時間現況

有許多關於哪些因素會增加等待時間，以及這些因素對應用程式效能實際影響程度的錯誤觀念。就這個層面來看時，著重於整體而非部分情況是很重要的。

導致 SSD 等待時間的重要原因是快閃記憶體本身。SLC 存取時間是 25µs+；MLC 存取時間是 50µs+，兩者都假定沒有存取競爭。隨著佇列深度增加，快閃記憶體存取競爭就可能大幅增加等待時間。

一旦快閃記憶體開始存取，其他對相同記憶體的要求就必須等候。八個快閃記憶體顆粒 (flash die) 共用通用匯流排，如此會導致記憶體顆粒需要等候使用匯流排。內部管理活動會增加額外等待時間 (位址轉譯、空間回收、平均抹寫儲存區塊等)。

另一個層面是作業系統，不論存取通訊協定以及交互連接情形如何，都會增加等待時間。這包括了檔案系統、磁碟管理員、類別驅動程式以及內容切換執行時間。

對應用程式而言，通訊協定以及交互連接的差異對於等待時間的影響微乎其微 (不到 1 毫秒)。

[www.seagate.com](http://www.seagate.com)

美洲地區  
亞太地區  
歐洲、中東和非洲

Seagate Technology LLC 10200 South De Anza Boulevard, Cupertino, California 95014, United States, +1 408 658 1000  
Seagate Singapore International Headquarters Pte. Ltd. 7000 Ang Mo Kio Avenue 5, Singapore 569877, +65 6485 3888  
Seagate Technology SAS 16-18, rue du Dôme, 92100 Boulogne-Billancourt, France, +33 1 41 86 10 00

© 2012 Seagate Technology LLC. 版權所有。在美國印製。Seagate、Seagate Technology 和 Wave 標誌是 Seagate Technology LLC 在美國和/或其他國家的註冊商標。其他商標或註冊商標均為其個別擁有者的財產。Seagate 得隨時變更產品供應項目或規格，恕不另行通知。TP625.1-1203TW，2012 年 3 月