技术资料

企业级 SSD 接口比较

简介

PCI Express (PCIe) 是一种通用的总线接口，适合于用户和企业级计算应用。现有的海量存储接口（SATA，PCIe）通过轮流连接到 PCIe 的主机适配器连接至主机中。

SATA 接口设计为一个硬盘 (HDD) 接口，SAS 接口 设计为设备和存储子系统接口/基础设施。因为 SATA 和系统要求已经发生演变，要求处理速度更快的接口和新的功能，所以 SATA 和 SAS 接口已经被多次改进。

固态硬盘 (SSD) 很快对这些接口提出了新的重大性能要求，因为 SSD 的数据速率已经从几十 MB/sec 增加到几百几千 MB/sec 。除了数据速率的增长，SSD 中缺乏机械运动也导致了每秒钟输入和输出操作的数量(IOPS) 增加，即存储设备能够实现的 IOPS 数量。

这项发展也对现有标准已经改进的实施提出了一项需求，以及增强现有接口标准，从而在保证兼容现有系统架构的同时管理新的性能需求。

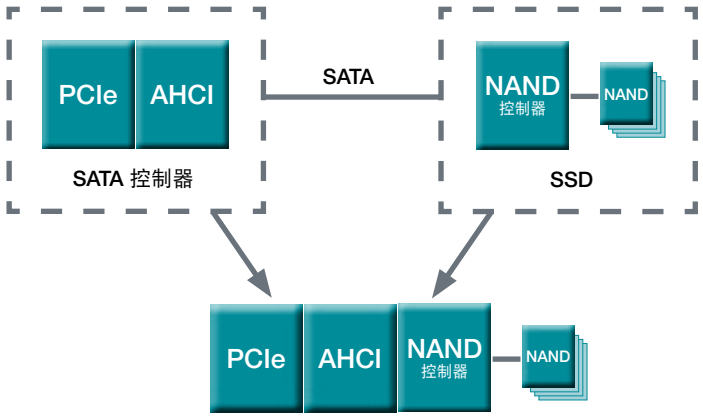
这篇文字讨论了不同的接口，并且对所遇到的各种性能和兼容性权衡进行了比较。

企业级 SSD 接口比较

SATA 接口

SATA 是为点到点连接设计的一种低成本接口，可以通过线缆或者印刷线路板 (PCB) 走线实现。主机连接面向高级主机控制器接口 (AHCI)，通常驻留于主机芯片集上面，如同主机适配器位于 PCIe 总线上面。这个接口有几个设计问题，可能使得每个指令产生 1µs 或更多总线开销。对于一个 4KB 传输需要 10µs 的 HDDs 来说，这不是主要的问题，但是 SSD 能够在 2µs 或者更少之内传输 4KB 数据——因而这个开销变得更值得注意，SATA 接口无兴趣成为一个高性能海量存储接口。

SATA 仍然适合用作低成本 SSD 接口，在这里，成本 而非性能是主要的决定因素。 SATA 架构也可以合并到一个管理 SATA 指令集的主机适配器中，并不会真正包含物理的 SATA 接口 (PHY) （图 1）。



来源：希捷公司，2011
图 1. 架构合并

SAS 接口

SAS 也是一个串联接口，通过一个主机适配器添加到主机中，但是也有明显的区别，使其成为一个适当的 SSD 接口：

- 更少的硬件开销
- 更快的传输速率
- 宽端口
- 有效的硬盘控制器接口

另外，SAS 包括 SATA 所不具备的功能，这些功能能够提升连接到接口的设备的可靠性和可用性：

- 健壮的串口协议
- 支持多主机
- 端到端数据整合
- 双端口容量
- 高度并发和聚合

更少的硬件开销

SAS 没有一个可等同于 SATA AHCI 控制器的通用主机接口。相反，SAS 主机适配器市场中有多个供应商相互竞争，在这里性能是主要的决定因素——并不仅仅是对于单个的 HDDs 接口，也面向不同的 RAID 系统，在这些系统中多个 HDD 主轴的传输速率将会聚合到一起，从而提高传输速率。另外，SAS 主机适配器是为管理高性能 SSD 和 HDDs 而设计（例如短击15K-RPM 硬盘）。因为硬件主机适配器和管理该主机适配器的硬盘被设计为一个系统，为 SSD 设计的产品经过优化，开始变得可用，并且进一步提升了传输速率以及 IPOS。

更快的传输速率

SAS 端口目前支持高达 6Gb/秒 的数据速率。一些企业，如 LSI 和 PMC-Sierra 目前正在设计支持 12Gb/s data 速率和 2 million 以上 IOPS 的样品，将来可能支持 24Gb/s。

宽端口

宽端口是 SAS 架构固有的理念——多个连接被聚合到一起，从而允许一个或多个主机与设备之间存在多个同步路径。目前的 SAS 设备连接器为设备定义两个端口。作为一种设计选择，目前的 HDDs 不支持宽端口——只有双端口，每个端口拥有不同的 SAS 地址，阻止其配置成为一个宽端口。

接受的*SAS-3*（12Gb/秒）协议允许硬盘上面的端口增加到 4 个，所有这些都连接到相同的域，或者成对连接到不同的域。非常受限的 SSD 数量可能在双端口设备上面支持宽端口（除了双端口以外）。

企业级 SSD 接口比较

健壮的串口协议

SAS 串口协议为串行传输者和接收者提供训练。通过对信道长度、阻抗失配和信号间干扰进行补偿，提升了线缆或者背板上面的信号质量。SAS 串口协议也管理硬件级别的错误检测和重传。从而使得间断信号可以迅速得到恢复。

支持多主机

SAS 接口 和交换结构允许多个主机访问相同的设备。这项功能可以用于管理主机故障，以及数据路径故障，从而提升数据的可用性。

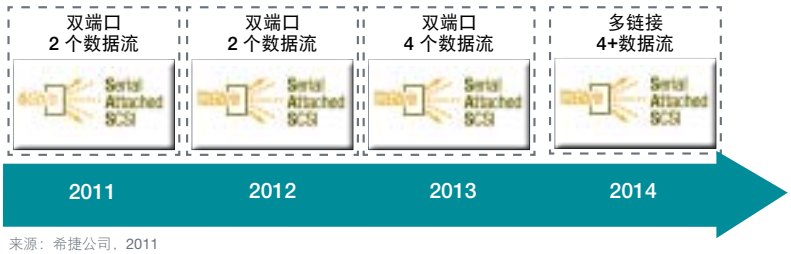
端到端数据完整性

SAS 接口 可以通过循环冗余检查 (CRC) 验证数据的完整性，从在主机数据缓存中创建数据，到经过 PCIe 接口和 SAS 接口，直到它被保存在设备上面，再次读取，并传输到主机数据缓存中。 从应用到 RAID 控制器，以及在设备上面，沿着这个路径有多个检查点。这项功能有时候称为保护信息 (PI)。

双端口性能

SAS 目标设备支持双端口操作。这样能够创建两个故障域，并提供更高的可用性。即使故障发生在一个连接到端口的路径，并且该端口阻止通过该路径访问，还是可以通过第二个端口访问该设备。

从历史的观点来说，希捷从市场中采购硬盘接口。希捷与 SCSI (STA) 同业公会以及其他行业领导者合作，充分利用已经广泛部署的 SAS 基础设施（图 2）。表 1 显示了 12Gb/秒 SAS 和多链接为系统构建者和最终用户组织带来的好处。



来源：希捷公司，2011
图 2SAS 接口演化

PCI Express 接口

PCI Express (PCIe) 是连接外围设备和主机处理器的主要接口，通过存储控制器连接到系统中的存储架构。SATA 和之前讨论的 SAS 接口都是通过 PCIe 接口（或者主机适配器）连接到主机处理器和内存中。

表 1. SAS的优势，以及提议的多链接增强 SAS	
多链接 (BW)	X4 (4x600MB/s)
可用电源	25W (2.5-inch)
总延迟	很低
多主机协议	是
高可用性	是（双端口）
可扩展性	棒极了
健壮的，被确认的协议堆栈	是
热插拔可用的	是
兼容现有的管理 SW	是

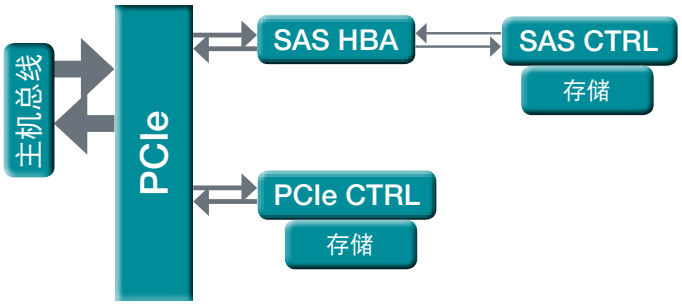
来源：希捷公司，2011

PCIe 接口是原始的 PCI 接口的串行实现，该原始接口在外围设备和主机处理器/内存之间提供并行的地址/数据连接。PCIe 接口通信是通过一条或者多条线路实现的，每条线路具有一个传输和接收串行接口。最多可以使用 32 条线路来连接主机和设备。每条线路上面的串行数据速率取决于所采用的 PCIe 标准的版本，当前的版本是 3.0，数据速率大约为 1GB/s。

对于一个 1U 用户，PCIe 接口设计为，使用一个（客户）母板上面的单一连接器，或（服务器）母板上面的两个连接器，直角适配器。布线系统也是可用的（尽管很少使用）。一个 2U、4U 或 7U 服务器具有多个 PCIe 插槽，类似于客户端应用。PCIe 规格也采用传输端（及接收端）训练以适应配置的阻抗变化，但是它目标是作为比 SAS 更短的传输信道。

PCIe 开关支持单根 I/O 虚拟化 (SR-IOV) 和多根 I/O 虚拟化 (MR-IOV)——这些方法用于提升虚拟系统（管理程序）中的控制器性能。这些虚拟带有单主机或多主机。SR-IOV 刚刚在适配器中开始成为普遍可用的；不过，VMware 可能还没有充分利用它的优势。MR-IOV 在适配器上面通常是不被支持的。

企业级 SSD 接口比较



来源：希捷公司，2011
图 3. SAS 接口演化

使用 PCIe 接口连接的存储设备通过直接注册器连接，或者通过主机适配器连接，该主机适配器接下来通过额外的线缆或者背板型接口连接到设备。

目前，这两种架构都有很多不同的实现。SATA 在系统芯片集（*南桥*）中使用主机总线适配器实现—— Intel 或 AMD AHCI——需要不同的 AHCI，但是映射到兼容的 IDE 遗产实现。这些接口也应用不同的 RAID 管理功能。

SAS 具有多个 HBAs 供应商，具有额外的扩展卡和可用的 RAID 控制器，所有这些都使用专有的设备驱动器和 BIOS 以满足不同的性能和配置需求。

PCIe 驱动控制器接口通过 NVM Express 规格和推荐的 SCSI over PCIe (SOP) 规格实现。

SATA 合并下文中描述的架构，是 PCIe 直接注册器连接的又一个实例。

今天的 PCIe SSD

当今市场上有两种主要的 PCIe SSD 类型：本地的和聚合器。本地的控制器附加到主机 PCIe 总线，然后直接控制多个闪存总线。 这些通常使用制造商专用的软件接口，并且只用于特定的设备。部分实现将地址转换的任务和其他功能放在主机 CPU 和内存中实现。因此，在设备工作负荷较重的情况下可以降低系统应用程序的资源需求。另外，相对于市场来说比较新，这些独特的硬盘和硬件组合有时会比较易于安装，因为它们的生态环境仍然处于进化过程中。

聚合器型号采用不同的设计方法。这种方法利用已有的 SATA 或 RAID SATA 控制器，多个 SAS 或 SATA SSD 附加在它上面。这些都封装在一张单独的 PCIe 卡上面。RAID 控制器集中了多个设备的性能，从而提供高性能。基于现有的经过证明的企业级硬件和软件接口，这些设计非常成熟和稳定。另外，这些设计使用智能控制器，能够实现地址转换和其他功能，允许应用程序充分利用系统 CPU 循环和内存，即使是在 I/O 工作负荷较重的情况下。

PCIe SSD 的未来

SOP 和 NVMe 两种方法的架构是类似的。不过，NVMe 正在由一个产业工作组开发，而 SOP 是由公认的开放式标准论坛开发。NVMe 仅仅以非易失内存设备的应用为目标，而 SOP 也以使用主机总线适配器和 RAID 控制器（带有不同的 SOP 设备之间的桥接功能）为宗旨。另外，SOP 着重体现现有的产业架构和功能，而 NVMe 使用一个新的受限制的指令集和排队接口。

接口优势和问题

上述的每一种存储架构都具有各自的优势，也存在一定的问题。取决于总体的系统设计，使用特定架构的好处可能比该架构带来的问题更加明显，因此，需要仔细分析才能做出适当的决定。这个决定也必须考虑兼容现有的系统设计。

例如，使用 SSD 升级一个具有 2.5-inch SATA HDD 的笔记本电脑，并且该 SSD 具有相同的物理大小和相同的（或更新）SATA 接口。在这种情况下，对于 SSD 速度有多快将会有一个限制；超出现有主机 SATA 接口速度则不会提升系统的性能。

在一个类似的情境中，一台使用短击 15K-RPM SAS HDD 存储数据索引的企业级服务器可以进行升级，使用 SAS SSD，从而提升系统的整体性能，但是只能达到其他系统因素成为新的瓶颈的程度（CPU、内存、网络、适配器等）。

在一个新的系统架构中，增加固态存储能够明显提高系统性能，但是最多只能达到系统架构的其他部分能够支持的数据速率和带宽。SSD 中的快速数据传输速率也需要有更多的电能提供给设备，并且还有更多的热量被消耗，不管 SSD 安装在哪里。

企业级 SSD 接口比较

表 2. 本地和聚合器 PCIe SSD 比较		
	本地的	聚合器
命令/传输	所有权 （FTL[1]位于主机/主存储器中） 	SCSI 或者 SATA （板卡上面有多个 SSD 和控制器） 
委员会	无	无
基于标准的	否	**是**
Flash 性能	高	高
CPU 开销	高	**低**
短队列延迟	很低	低
深队列延迟	适度的	**低**
使用案例可扩展性	否	**是**（RAID，HBA，等）
成熟	演化	**基于经过证明的产业架构**
企业级功能组 （PI，安全，管理，等）	否	**取决于实现**

闪存转换层
来源：希捷公司，2011

表 3. SOP 和 NVMe PCIe SSD 比较		
	SOP[1]	**NVMe[2]**
命令/传输	SOP/PQI[3] （制器中的 FTL 位于控制器中） 	NVMe/NVMe （制器中的 FTL 位于控制器中） 
委员会	**T10/INCITS[4]**	产业工作组
基于标准的	**是 (ANSI/ISO)**	否
Flash 性能	很高	很高
CPU 开销	低	低
短队列延迟	很低	很低
深队列延迟	低	低
使用案例可扩展性	**是**（RAID，HBA，等）	否（仅 NVM）
成熟	**基于证明产业架构**	待定
企业级功能组 （PI，安全，管理，等）	**完全支持**	有限责任

[1]SOP：SCSI over PCI 总线
[2]NVMe：非挥发性记忆联盟
[3]PCIe 排队接口
[4]INCITS：国际信息技术标准委员会
来源：希捷公司，2011

另一个因素是，支持这些新 SSD 接口的操作系统设备驱动器和 BIOS 的可用时间，以及软件的初始可靠性。

接口和 Flash SSD 延迟真相

在探讨什么因素造成延迟，以及在真实环境下如何影响应用性能这些问题时，存在很多误解。当考虑这方面的问题时，聚集于整体状况是相当重要的，而不仅仅是其中的一部分。

在 SSD 中，大部分延迟是由 flash 部分本身造成的。SLC 访问时间是 25µs+；MLC 访问时间是 50µs+，两者均假设没有访问竞争。随着队列深度增加，争夺 flash 部分的访问资源可能导致延迟明显增加。

一但一个 flash 部分开始访问，其他请求相同部分的则必须等待。多达 8 个 flash die 共享一条公用的总线，这样使得 die 一直等待，直到轮到该 die 使用总线。家务活动增加了额外的延迟（地址转换、垃圾收集、损耗平衡等）。

另外一个方面是操作系统，它添加了延迟，无论访问协议和互连如何。这些包括文件系统、卷管理器、类驱动器和上下文交换开销。

协议和互连差异对延迟的影响非常有限，应用程序可以忽略此影响。

www.seagate.com/cn

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