

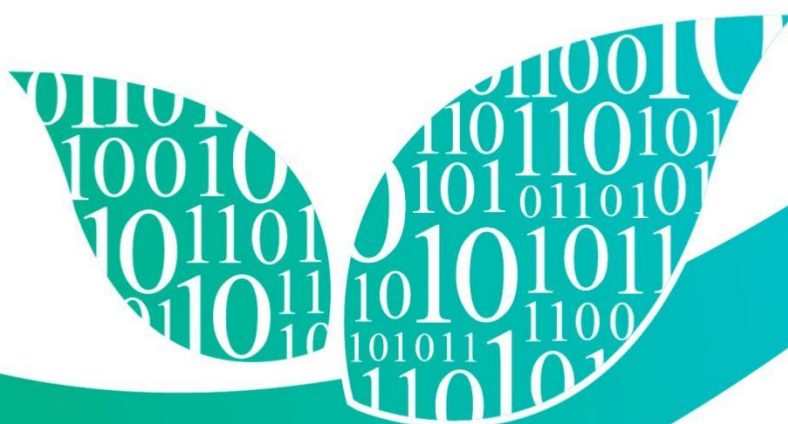


开放数据中心委员会  
Open Data Center Committee

[编号 ODCC-2022-05001]

# 企业级 SSD 技术与应用报告

Enterprise SSD Technology and Application Report



开放数据中心标准推进委员会

2022-04 发布

## 版权说明

ODCC（开放数据中心委员会）发布的各项成果，受《著作权法》保护，编制单位共同享有著作权。

转载、摘编或利用其它方式使用 ODCC 成果中的文字或者观点的，应注明来源：“开放数据中心委员会 ODCC”。

对于未经著作权人书面同意而实施的剽窃、复制、修改、销售、改编、汇编和翻译出版等侵权行为，ODCC 及有关单位将追究其法律责任，感谢各单位的配合与支持。

[www.ODCC.org.cn](http://www.ODCC.org.cn)

## 编制说明

本报告由中国信息通信研究院云计算与大数据研究所牵头撰写,在撰写过程中得到了多家单位的大力支持,在此特别感谢以下参编单位和参编人员:

参编单位(排名不分先后):

中国信息通信研究院、深圳忆联信息技术有限公司、浪潮电子信息产业股份有限公司、烽火通信科技股份有限公司、联想(北京)信息技术有限公司、西安三星电子研究所、SK 海力士、得一微电子股份有限公司、深圳市江波龙电子股份有限公司

参编人员(排名不分先后):

郭亮、李洁、王少鹏、赵精华、孙聪、李军、吕玉彬、宗岗岗、李传宝、胡振国、侯毅、豆坤、郑南植、金文珠、赵玉珠、罗挺、孔燕斌、陈林峰、冼曙光

项目经理:

赵精华 zhaojinghua@caict.ac.cn

[www.ODCC.org.cn](http://www.ODCC.org.cn)

# 前言

随着企业数字化转型的持续推进和前沿技术的加速应用，数据总量大幅增长，数字经济持续繁荣。存储是信息产业的重要基础组成之一，市场对数据存储容量、数据传输速度、硬件设备性能等各方面要求的不断提升也推动着存储技术的持续进步，需求迭代与技术迭代将共同刺激存储市场释放进一步增长空间。企业级 SSD 作为基础 IT 硬件设备，市场规模持续扩大，应用需求稳步增长。

本报告聚焦企业级 SSD 技术和应用，梳理了全球及我国企业级 SSD 市场现状、发展趋势、政策演进和标准迭代等内容，并在部件技术创新、接口性能优化、传输协议升级方面进行了深入分析，阐释了 SSD 性能、可靠性、稳定性测试方案和影响因素，绘制了全方位、全领域、多角度的企业级 SSD 产业图谱。

未来企业级 SSD 发展或将呈现以下趋势：在新兴技术方面，持续推动先进存储技术研发和应用，提升存储容量，满足数据计算要求；在绿色低碳方面，通过多措并举的方式提质降耗，从存储材料、设计规范、操作使用等多个方面，建立高能效、低功耗技术体系；在技术攻关方面，研发投入持续加大，技术水平全面提高，建立起高效异构的融合技术架构；在标准规范方面，强化标准支撑，健全评估体系，全面提升企业级 SSD 发展能力。

ODCC 于 2016 年启动企业级硬盘基准测试 BESH 项目并接受首批测试，2021 年参与贡献的 YD/T 3824《面向互联网应用的固态硬盘测试规范》和 YD/T 3825《面向互联网应用的机械硬盘测试规范》两项行业标准正式发布，目前 BESH 测试正在火热进行中。未来，BESH 项目组将继续以标准为依据，以测评为抓手，持续赋能企业级 SSD 高质量健康发展。

由于企业级 SSD 技术迭代快、涉及范围广，报告内容疏漏在所难免，敬请读者给予批评指正。如有意见或建议请联系 [dceco@caict.ac.cn](mailto:dceco@caict.ac.cn)。



# 目录

|                               |     |
|-------------------------------|-----|
| 版权说明 .....                    | I   |
| 编制说明 .....                    | II  |
| 前言 .....                      | III |
| 1. 企业级 SSD 存储总体情况 .....       | 1   |
| 1.1. 企业级 SSD 概述 .....         | 1   |
| 1.1.1. 简介 .....               | 1   |
| 1.1.2. 技术概述 .....             | 1   |
| 1.1.3. 分类方式 .....             | 2   |
| 1.2. 全球企业级 SSD 市场分析 .....     | 3   |
| 1.2.1. 市场趋势 .....             | 3   |
| 1.2.2. 产业政策 .....             | 6   |
| 1.3. 我国企业级 SSD 市场分析 .....     | 7   |
| 1.3.1. 市场趋势 .....             | 7   |
| 1.3.2. 产业政策 .....             | 8   |
| 1.4. 企业级 SSD 热点技术规范 .....     | 8   |
| 1.4.1. EDSFF 响应存储性能迭代要求 ..... | 8   |
| 1.4.2. PCIe 赋能设备数据高速传输 .....  | 10  |
| 1.4.3. NVMe 满足高速存储访问需求 .....  | 11  |
| 2. 企业级 SSD 创新技术分析 .....       | 13  |
| 2.1. 企业级 SSD 部件技术创新 .....     | 13  |

|        |                |    |
|--------|----------------|----|
| 2.1.1. | 主控             | 13 |
| 2.1.2. | 固件             | 15 |
| 2.1.3. | 闪存             | 17 |
| 2.2.   | 企业级 SSD 接口技术创新 | 20 |
| 2.2.1. | PCIe 接口        | 20 |
| 2.2.2. | SAS 接口         | 20 |
| 2.2.3. | U.2 接口         | 22 |
| 2.3.   | 企业级 SSD 传输协议创新 | 23 |
| 2.3.1. | NVMe           | 23 |
| 2.3.2. | NVMe-oF        | 23 |
| 2.3.3. | ZNS            | 25 |
| 3.     | 企业级 SSD 测试技术分析 | 26 |
| 3.1.   | 测试工具           | 26 |
| 3.2.   | 测试指标           | 27 |
| 3.2.1. | 性能测试           | 27 |
| 3.2.2. | 可靠性测试          | 28 |
| 3.2.3. | 稳定性测试          | 29 |
| 3.3.   | 影响因素           | 30 |
| 3.3.1. | 数据块尺寸          | 30 |
| 3.3.2. | 写入速率           | 31 |
| 3.3.3. | 芯片误码           | 32 |
| 3.4.   | 权威测试           | 33 |

|                          |    |
|--------------------------|----|
| 4. 企业级 SSD 产业图谱 .....    | 35 |
| 4.1. 主控厂商 .....          | 35 |
| 4.1.1. 领域图谱 .....        | 35 |
| 4.1.2. 发展现状 .....        | 35 |
| 4.1.3. 发展趋势 .....        | 36 |
| 4.2. NAND Flash 厂商 ..... | 37 |
| 4.2.1. 领域图谱 .....        | 37 |
| 4.2.2. 发展现状 .....        | 37 |
| 4.2.3. 发展趋势 .....        | 38 |
| 4.3. DRAM 厂商 .....       | 39 |
| 4.3.1. 领域图谱 .....        | 39 |
| 4.3.2. 发展现状 .....        | 39 |
| 4.3.3. 发展趋势 .....        | 40 |
| 4.4. 盘厂商 .....           | 41 |
| 4.4.1. 领域图谱 .....        | 41 |
| 4.4.2. 发展现状 .....        | 42 |
| 4.4.3. 发展趋势 .....        | 42 |
| 5. 总结与展望 .....           | 43 |

## 1. 企业级 SSD 存储总体情况

### 1.1. 企业级 SSD 概述

#### 1.1.1. 简介

企业级 SSD（enterprise solid-state drive，企业级固态硬盘）是指应用于高性能计算、边缘计算、高端存储、数据中心等各种企业级场景中的固态硬盘，具备不间断工作能力，能够处理 I/O 密集型工作负载，如数据库文件、索引日志、数据分析以及其它对性能要求较高的事务处理操作<sup>[1]</sup>。与消费级 SSD 相比，它具备更强性能、更高可靠性、更强耐用性。企业级 SSD 面向企业级用户，要求保证数据安全性、速度稳定性和长期耐用性，对产品高可靠性交付要求更严。企业级 SSD 与消费级 SSD 的情况对比如表 1 所示。

表 1 企业级 SSD 与消费级 SSD 情况对比

|         |         | 企业级 SSD  | 消费级 SSD                          |
|---------|---------|--|----------------------------------|
| 使用对象及功能 |         | 多用于 IT、动画设计等行业，在满足使用要求的同时能够保障数据安全性、速度稳定性和长期耐用性 | 个人电脑使用，能够实现高速流畅操作                |
| 关注点     |         | 注重随机性能、延迟、IO QoS 的保证以及稳定性，要求可靠性、长时间运作、高 MTBF   | 注重顺序性能、功耗、价格等因素                  |
| 性能      | 颗粒材质    | 稳定系数较高的颗粒材质                                    | 对材质要求相对较低                        |
|         | 性能稳定性   | 读写低延迟，要求 SSD 读写速度以及快速响应等                       | 初始读写性能较高，随着使用时间的增加而逐步降速          |
|         | 延迟      | 以最少的延迟量访问存储设备，要求极低延迟                           | 对于用户而言，可接受的延迟相对较长                |
| 可靠性     | 数据完整性   | 在各种恶劣环境都不能出现数据丢失                               | 只保护 FTL 数据，但不需要保护 DRAM 缓存中其他用户数据 |
|         | 断电保护配置  | 配备专用电容，突然断电时，SSD 控制器将从专用电容中获取电能                | 只能做到基础断电保护                       |
| 耐久性     | 负载周期    | 24 小时不间断访问，24x7 负载周期                           | 约 8 小时充分利用，80%左右的时间处于空闲状态        |
|         | 额定 MTBF | 大约 200 万-250 万小时                               | 大约 150 万小时                       |

#### 1.1.2. 技术概述

与个人电脑、笔记本电脑和平板电脑等设备中使用的消费级 SSD 相比，企



企业级 SSD 性能优势明显，增强服务能力和耐用性。企业级 SSD 能够在断电时为 DRAM 存储数据提供有效保护，同时采用了更强大的纠错码（ECC）技术，提供始终如一的高质量服务和期限更长的保修服务。与此同时，企业级 SSD 通过优化 NAND 闪存技术，实现比消费级 SSD 更强的耐用性。由于向芯片写入新数据时，需要经历反复的编程和擦除，导致 NAND 闪存磨损，而企业级 SSD 通过损耗均衡算法、自我修复功能和超量配置等技术，使企业级 SSD 可以更大程度保留 NAND 闪存，以备芯片磨损或出现故障时使用，提高了 NAND 闪存的续航能力。

企业级 SSD 的快速发展与相关技术的持续迭代密切相关，接口速率、存储密度等技术不断升级，带动企业级 SSD 性能稳步增强。接口技术的优化使得传输速率明显提高，满足应用的高速访问需求。SSD 相比于 HDD，最大的优势在于随机读写的能力 IOPS，HDD SATA 接口仅能达到数百量级，而 SSD PCIe 接口的 IOPS 可以提高数千倍；即使同为 SATA 接口的 SSD，也比相同接口的 HDD 的 IOPS 要高上百倍<sup>[2]</sup>。随着企业级 SSD 接口从 SATA 升级到 PCIe 的过程中，速率也从 SATA3.0 的 6GB/s 飙升到 PCIe6.0 x16 的 128GB/s。NAND 闪存密度增加可以提升企业级 SSD 性能。最早的企业级 SSD 通常使用 SLC（Single-Level Cell，一阶存储单元）NAND 闪存，即每个单元存储一位。SLC NAND 可以提供高性能和高耐用性，全生命周期内每个单元的写入次数可以达到十万次。随着 NAND 闪存技术优化，企业级 SSD 可以选用耐用性更低但成本更少、容量更大的 NAND 闪存，包括 MLC（Multi-Level Cell，二级存储单元）、TLC（Triple-Level Cell，三级存储单元）、QLC（Quad-Level Cell，四阶存储单元）和多层数 3D NAND。

### 1.1.3. 分类方式

企业级 SSD 有多种接口类型和外形规格。在接口类型方面，包括串行 ATA（Serial ATA，SATA）、串行连接 SCSI（Serial-attached SCSI，SAS）和 PCI 接口（PCI Express，PCIe），可以与中央处理器（CPU）进行数据传输；在外形规格方面，企业级 SSD 包括可与硬盘驱动器安装在同一插槽中的 2.5 英寸驱动器，HHHL（Half-height Half-length）插卡或可插入计算机 PCIe 总线的 M.2 模块，以及带有双列直插式内存模块（Dual In-line Memory Module，DIMM）芯片的小型电路板，

该电路板可以与计算机主板连接。

NVMe凭借高速率、低功耗、低时延、强兼容的特性，成为企业级SSD的主流传输协议。在速率方面，相比于SCSI和SATA传输协议，NVMe所需的CPU指令集数量减少了一半左右，将I/O命令和响应映射到主机的共享内存，支持多核处理器并行，充分利用并行数据路径，极大提升传输速度和吞吐量和，缓解CPU的压力；在时延方面，与SATA相比，NVMe协议使得数据不再需要通过控制器再中转到CPU，而是可以通过PCIe通道直接与CPU连接，延时几乎可以忽略；在功耗方面，NVMe协议采用功耗管理（Power Management，PM）命令集以及相应的功耗管理模型，加入自动功耗状态切换和动态能耗管理功能，有效降低功耗水平；在兼容性方面，NVMe协议能够匹配不同的平台和系统，无需使用相应驱动就能够正常运行。

## 1.2. 全球企业级 SSD 市场分析

### 1.2.1. 市场趋势

全球企业级 SSD 市场总体呈现上升趋势，市场规模稳步扩大，PCIe 所占份额将进一步增大。IDC 数据显示，企业级 SSD 全球出货量将从 2020 年的约 4750 万块增长到 2021 年的 5264 万块左右，年均增长率为 10.7%，预计到 2025 年，将增加到 7436 万块左右。从整体上看，2020 年到 2025 年期间，企业级 SSD 出货量始终保持上升趋势。企业级 SSD 市场主要包括 SAS SSD、SATA SSD、PCIe SSD、DDR SSD 四个类别，PCIe SSD 所占比重大幅增长。企业级 PCIe SSD 出货量份额从 2020 年的 46.9%，增长到 2025 年的 88.9%，成为企业级 SSD 的市场的主流产品；DDR SSD 的重要性在持续增高，2020 年到 2025 年期间的年均增长率高达 105%，甚至超过了 PCIe；而 SAS SSD、SATA SSD 的出货量和所占比重都在逐年下降。具体增长情况如图 1 所示。

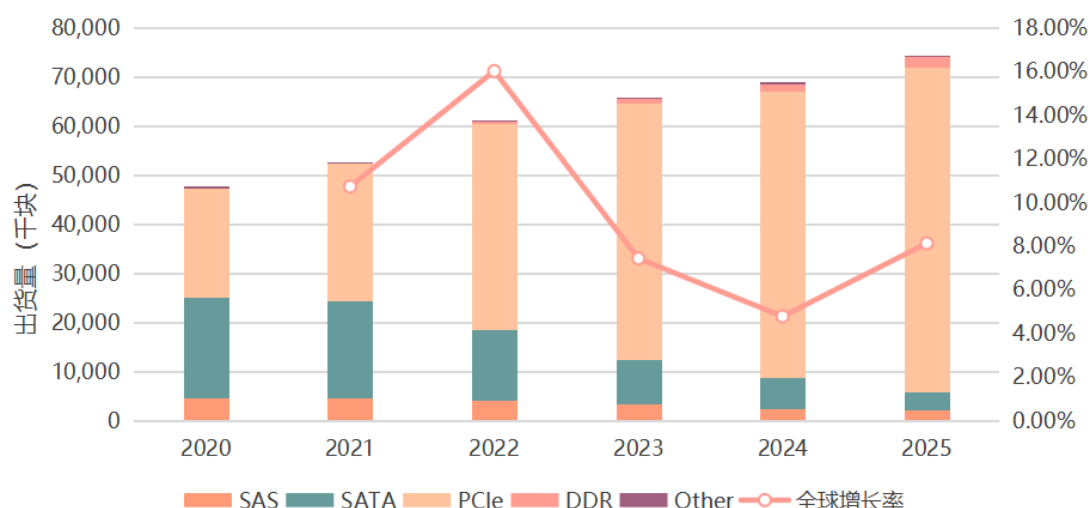


图 12 2020 年-2025 年全球企业级 SSD 出货量

数据来源：IDC

企业级 SSD 不断提升数据完整性等各性能指标的同时，使得容量成本不断下降，大容量企业级 SSD 逐渐成为市场主流。根据 IDC 统计数据，全球企业级 SSD 单位容量价格呈持续下降趋势，由 2020 年接近 0.19 美元/GB 下降至 2025 年的 0.07 美元/GB，6 年价格下降超过 63%。随着 3D NAND 技术、NVME 协议和其他固件算法的使用范围逐渐扩大，企业级 SSD 容量、运行速度，处理能力，耐久度得到巨大的飞跃。企业级 SSD 单盘容量实现持续增长，IDC 数据显示，2020 年到 2025 年期间，大容量企业级 SSD 出货量总体呈现上升趋势，1TB 以上的 SSD 出货量年复合增长率均超过 10%，容量超过 4T 的企业级 SSD 增长率超过 30%，16T 以上的 SSD 出货量在 2020 年仅为 2.5 万块，而到 2025 年将达到 316.9 万块，年复合增长率高达 125%。2020 年-2025 年全球企业级 SSD 不同容量出货量及容量成本变化情况如图 2 所示。

www.ODCC.org.cn



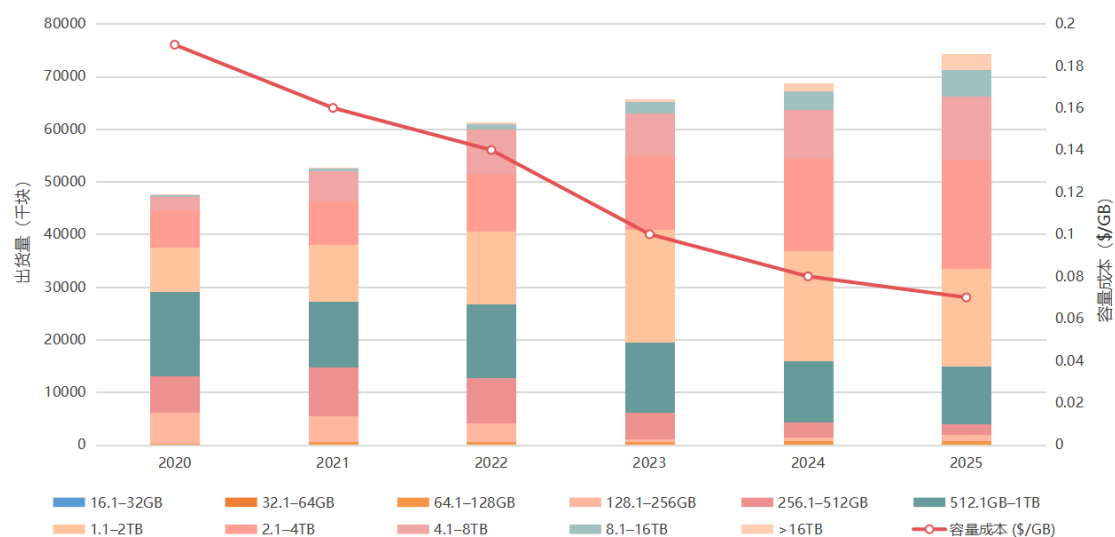


图 2 2020 年-2025 年全球企业级 SSD 不同容量出货量及容量成本

数据来源：IDC

数字化转型和云迁移的推进，2022 年企业级 SSD 的需求将继续保持强劲活力。各国政府及企业对于云端应用需求快速增长，北美云服务供应商的库存调整基本完成，开始逐步扩大存储容量，Intel 和 AMD 等企业的新一代服务器 CPU 开始量产，诸多因素并发，带动企业级 SSD 的需求增长。SK 海力士数据显示，2022 年企业级 SSD 市场需求将持续增长，连续两年的年增长率超过 30%，如图 3 所示。与此同时，对于算力需求的激增和算力基础设施的投资扩大，使得用于数据中心的企业级 SSD 所占比重持续攀升，预计 2022 年占比接近 60%，到 2023 年将达到 70% 左右。

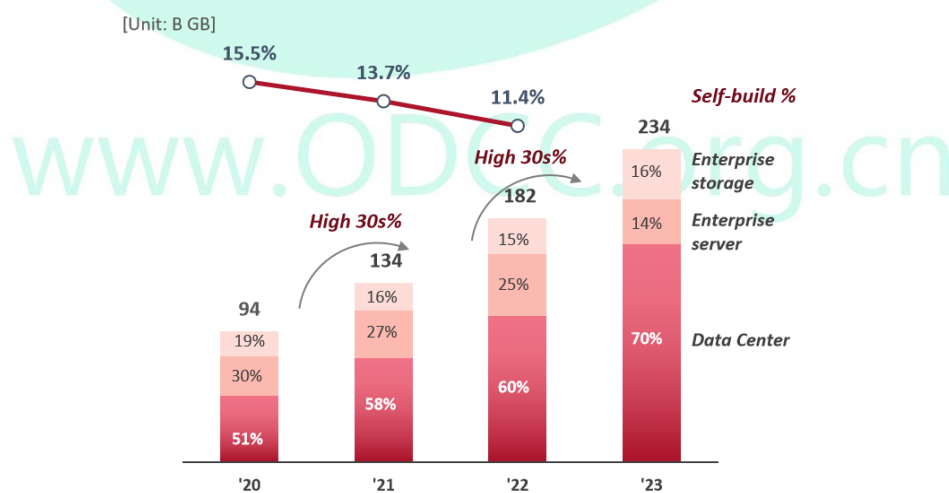


图 2 2020 年-2023 年企业级 SSD 市场需求

数据来源：SK 海力士



### 1.2.2. 产业政策

半导体产业是前沿技术领域，世界主要国家通过政策鼓励大力推动半导体产业发展。固态硬盘是半导体存储产品的重要形态，近年来世界主要国家积极发布政策推动以半导体为代表的新材料技术发展。

美国高度重视半导体存储芯片的战略价值，持续完善政策体系，不断丰富行业标准，使存储产业规范化、标准化。美国政府以政策为抓手，鼓励国内半导体厂商进行技术创新，完善专利保护法，对半导体存储相关技术进行全面的保护。除了政策支持外，美国的行业协会也积极通过商业形式加强了存储产品的测试和认证，如美国存储性能委员会（Storage Performance Council）推出了 SPC BENCHMARK 存储设备测试规范，该测试规范能够对存储设备性能进行全方面评估，该测试规范的推出有助于推动美国及全球存储产业发展。

韩国半导体产业在国内需求有限、国际竞争激烈的环境下，保持良好的发展态势和强大的产业优势。韩国政府为促进企业技术研发积极性，研发时大力投入，产出后加强保护。1994 年，韩国推出了《半导体芯片保护法》，此后，韩国政府还指定芯片产业及技术为影响国家竞争力的核心技术，致力于高度保障技术及产权。韩国的半导体产业进入全球半导体产业的第一梯队后，韩国仍继续加强政策支持力度，不仅通过“BK21”及“BK21+”等计划对大学、专业或研究所进行精准、专项支援，还在 2016 年时推出半导体希望基金，投资于半导体相关企业，旨在聚焦新技术的开发，尤其是储存新技术方面。2021 年 5 月，韩国政府正式公布“K 半导体战略”，在未来 10 年投资约 4500 亿美元，将韩国建设成全球最大的半导体制造基地。

日本半导体产业制定强有力的应对策略，着力摆脱即将衰败的泥沼。日本为重塑半导体产业霸主地位，扩大国内存储厂商芯片产能，在 2021 年先后发布《半导体数字产业战略》和《半导体产业紧急强化方案》。2021 年 6 月，日本政府对半导体、数字基础设施及数字产业做出综合部署，制定了以扩大国内半导体生产能力为目标的《半导体数字产业战略》。2021 年底，日本发布《半导体产业紧急强化方案》，吸引厂商新建半导体工厂，同时为老旧半导体厂房设备提供资金支持，致力于到 2030 年将日本半导体企业的营收提高至 2020 年的 3 倍。与此同

时，日本政府和企业合作加强合作，组建产官学联合援助协会，包括各县政府、半导体企业、大学等多家机构。

### 1.3. 我国企业级 SSD 市场分析

#### 1.3.1. 市场趋势

我国数据总量快速增长，对数据存储能力提出更高要求，带动企业级 SSD 快速发展。IDC 数据显示，到 2025 年，全球数据总量将达到 175ZB，中国数据量年复合增长率领先全球，预计到 2025 年将达到 48.6ZB，在全球数据总量中占比约 27.8%，成为数据量最大的地区，数据中心存储量占比将超过 70%。企业级 SSD 作为数据中心的重要存储介质，出货量持续增长。TrendForce 数据显示，我国企业级 SSD 快速增长，2019 年出货量不到 400 万块，到 2025 年预计将超过 1300 万片，年增长率保持在 18% 以上，发展空间广阔。2018 年-2025 年中国数据量在全球所占比重的情况如图 4 所示。

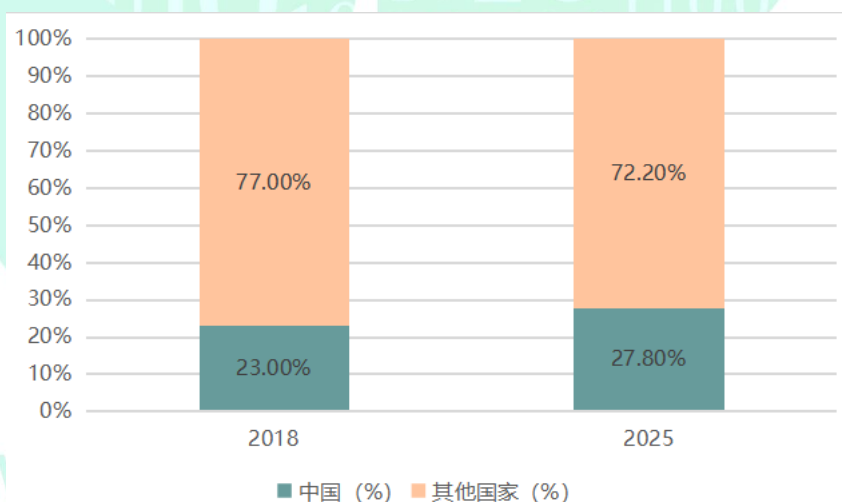


图 3 2018 年-2025 年中国数据量占比

数据来源：IDC

我国企业级 SSD 市场规模持续扩大，PCIe SSD 成为企业级 SSD 市场的主要参与者。数据显示，2019 年，我国企业级 SSD 市场规模不足 100 亿元，到 2021 年，总规模已经超过 200 亿元<sup>[3]</sup>。预计未来，随着企业投资力度的加大和新兴技术带来的市场需求激增，企业级 SSD 将拥有更大的市场空间和更强的竞争能力，

凭借速度快、零噪音、能便携、防摔震等优势，实现对企业级机械硬盘更深、更广、更多的替代。在企业级 SSD 市场中，PCIe 逐渐取代 SATA 及其他类别，成为企业端的主流产品。2019 年，PCIe 企业级 SSD 市场份额仅为 41%，在此后两年里，PCIe 总体规模和市场比重屡创新高，预计 2022 年，占比将达到 74% 以上，到 2025 年，市场比重预计将突破 90%。

### 1.3.2. 产业政策

依托数字中国建设愿景，我国政府发布多项政策支持存储产业发展，出台多项政策鼓励高端信息设备制造产业建设。2015 年国务院引发《中国制造 2025》，提出要坚持创新驱动、智能转型、强化基础、绿色发展，加快从制造大国转向制造强国，包括半导体存储在内的“新一代信息技术”被列入重点发展领域<sup>[4]</sup>。2016 年 7 月国务院发布《国家信息化发展战略纲要》，明确提出要“夯实基础设施，强化普遍服务”，将“泛在先进的基础设施”视为“信息化发展的基石”，半导体存储设备作为高性能存储的代表，同样属于信息领域关键的先进基础设施<sup>[5]</sup>。2017 年 2 月，国务院发布《战略性新兴产业重点产品和服务指导目录》（2016 版），该文件明确将固态硬盘列入“下一代信息技术产业”分类下的“下一代信息网络产业”中，同时也在“集成电路”产业中提及半导体存储器<sup>[6]</sup>。2020 年 7 月国务院发布《新时期促进集成电路产业和软件产业高质量发展的若干政策》，重申了发展集成电路产业对国民经济进步及信息化建设的重要意义，指出针对 28nm、65nm 和 130nm 芯片分别给予不同程度的减免企业所得税优惠<sup>[7]</sup>。2021 年 12 月，中央网络安全和信息化委员会印发《“十四五”国家信息化规划》，提出加快集成电路关键技术攻关，推动计算芯片、存储芯片等创新，加快关键材料研发和特色工艺突破，“存储”再次成为关注重点<sup>[8]</sup>。

## 1.4. 企业级 SSD 热点技术规范

### 1.4.1. EDSFF 响应存储性能迭代要求

企业级和数据中心 SSD 形态设计的更新迭代，适用于新一代 NVMe 硬盘规格的 EDSFF 标准应运而生，成为增强存储性能的重要依据。EDSFF（enterprise data center SSD form factor，企业级数据中心 SSD 外观尺寸）是应用于企业和数



据中心 SSD 封装尺寸的系列规范，改进了 SSD 容量、散热、供电和可扩展性，在 1U 服务器上实现 PB 级的存储能力，更可以在 2U 服务器上获得更极致的存储性能。EDSFF 建立在 NVMe 功能的基础上，与传统 SSD 相比，吞吐量增加 6 倍，延迟减少 7 倍，为客户提供了更强大的存储选项。EDSFF 作为高速多用途连接器，具有强大的生态系统，实现完全热插拔支持，采用内置 led、ESD 外壳和免托架设计，配备完全可定制的门锁，无需工具即可使用，能够用于 SSD 的普通 pinout。

**EDSFF 实现高密度和数据中心 TCO 降低灵活可选 (E1.L)，可扩展的高性能 (E1.S) 和主流 2U 服务器/存储支持 (E3)。**E1.S 面向可扩展和高灵活度的性能型存储，有五种不同厚度的选择，具有尺寸小、可扩展、部署灵活等特性，在驱动器的前面是一个有两个螺丝孔的法兰盘，增强锁结构适配；E1.L 主要针对 QLC 等高容量存储，比其他任何驱动器外形都要长得多，用于包含大量闪存的系统，适合容量为数十 TB 的 SSD；E3 适用于 3 英寸 SSD，主要针对 2U 服务器/存储中的高性能 SSD，也可以用于 1U 服务器，但相对于 1U 的 E1.S，可能存在制冷和密度分布的缺点；E3.S 的形状尺寸被拉伸了一点，以匹配 OCP NIC 3.0 标准的尺寸，是面向云服务和企业数据中心 NVMe SSD 新形态标准。E3.S 的重要特色就是针对 PCIe 5.0 乃至未来的 PCIe 6.0 做好了准备，扩展性更强，并在性能、散热、功耗等方面都做了改进和统一，支持 PCIe x4/x8 通道，非常适合大容量、高密度的全闪存存储阵列(AFA)<sup>[9]</sup>。

**EDSFF规范的突出优势使其应用范围持续扩大，为企业级SSD外形规格提供广阔的优化空间。**EDSFF在占用最小空间的同时，能够实现高密度和高功率存储，满足了数据中心对密度、功率、容量、性能和冷却的需求，企业级SSD厂商和服务设备商在平衡功耗、速度和冷却效率的同时优化存储空间，应当关注存储驱动器的外形，加快构建更高性能、更多类型的基础设施以满足EDSFF规范要求。尽管在向EDSFF过渡的过程中存在一些挑战，例如，服务器机柜需要重新设计以适应EDSFF要求，但从总体趋势看，EDSFF在未来将得到广泛部署，逐步取代2.5英寸规格和M.2外形要求，如图5所示。



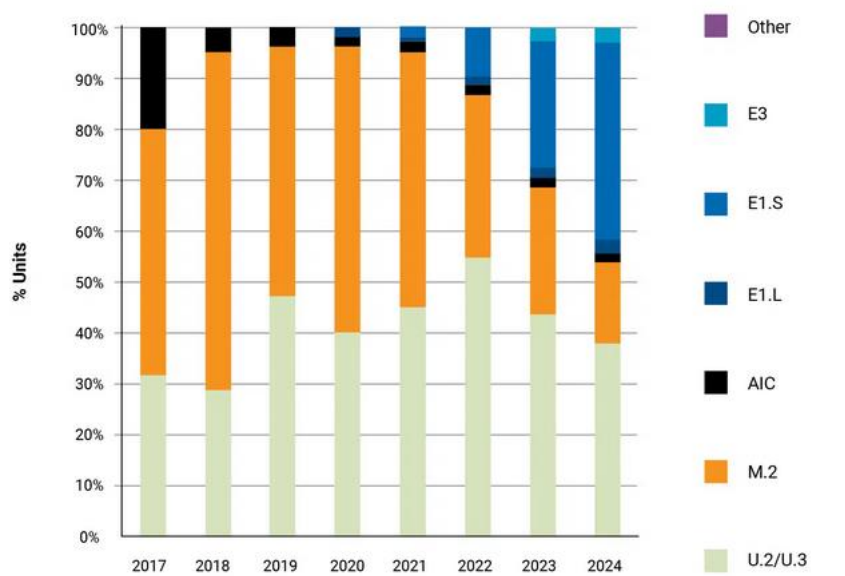


图 4 2017-2024年存储形态发展趋势

数据来源：EDSFF

#### 1.4.2. PCIe 赋能设备数据高速传输

PCI Express 规范为数据高速交互提供依据，是服务器总线提升传输速率的主要解决方案。PCI-Express (peripheral component interconnect express, PCIe) 是一种高速串行计算机扩展总线标准，于 2003 年，由 PCI-SIG (PCI Special Interest Group, PCI 特设组) 正式发布。PCIe 1.0 带宽为 8GB/s，到 2010 年的 PCIe 3.0，速度已经提升至 32GB/s，而实现规模化应用的 PCIe 4.0 速度可以达到 64GB/s。PCIe 接口有 x1、x2、x4、x8、x16 等多种配置，数字表示数据通道的数量，通道越多，数据从固态硬盘传输到主板的速度就越快。目前，PCIe 6.0 规范已经正式发布，预计最高写入速度为 11GB/s，最高读取速度为 12GB/s，PCIe 5.0 SSD 设备将在 2022 年正式上市<sup>[10]</sup>。PCIe 系列规范带宽速度如表 2 所示。

表 2 PCIe 系列规范带宽速度

| 插槽<br>宽度 | PCIe 1.0<br>(2003) | PCIe 2.0<br>(2007) | PCIe 3.0<br>(2010) | PCIe 4.0<br>(2017) | PCIe 5.0<br>(2019) | PCIe 6.0<br>(2021) |
|----------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| x1       | 0.25GB/s           | 0.5GB/s            | ~1GB/s             | ~2GB/s             | ~4GB/s             | ~8GB/s             |
| x2       | 0.5GB/s            | 1GB/s              | ~2GB/s             | ~4GB/s             | ~8GB/s             | ~16GB/s            |
| x4       | 1GB/s              | 2GB/s              | ~4GB/s             | ~8GB/s             | ~16GB/s            | ~32GB/s            |

|     |       |       |         |         |         |          |
|-----|-------|-------|---------|---------|---------|----------|
| x8  | 2GB/s | 4GB/s | ~8GB/s  | ~16GB/s | ~32GB/s | ~64GB/s  |
| x16 | 4GB/s | 8GB/s | ~16GB/s | ~32GB/s | ~64GB/s | ~128GB/s |

数据来源：PCI-SIG

**PCIe 市场规模持续扩大，服务器设备是规范迭代升级的主要受益者。**企业级 SSD 的发展使得 PCIe 在 2021 年获得重大发展机遇，并在未来 5 年内获得持续增长。服务器厂商先后推出众多支持 PCIe 4.0 的服务器主板，支持 NVMe 硬盘、GPU、网卡等相对应设备；当 PCIe 5.0 实现商用后，市场规模将得到进一步扩大，大幅度弱化存储设备在整个计算机系统性能中的性能瓶颈，NVMe 存储设备将成为受益最大的设备；到 2023 年，服务器设备能够率先实现 PCIe 6.0 应用，为人工智能（机器学习）、数据中心、物联网（IoT）、汽车、航空航天和军事等领域提供有效支撑，为高性能计算中心建立组合式基础设施架构<sup>[11]</sup>。

#### 1.4.3. NVMe 满足高速存储访问需求

**NVMe 系列标准规范了 SSD 访问接口，确保 NVMe 在网络架构、接口管理、服务器和系统管理规范化和明确化。**NVMe 是由 NVM Express 非盈利组织发布的规范，定义了主机软件如何通过 PCIe 总线与非易失性内存通信。它是所有形式（U.2、M.2、AIC、EDSFF）的 PCIe 固态硬盘（SSD）的行业标准。除了 NVMe 基本规范之外，还包括用于规范网络架构上使用 NVMe 命令的 NVMe over fabric（NVMe-oF），以及用于服务器和存储系统管理 NVMe/PCIe SSD 的 NVMe 管理接口（NVMe-MI）规范。随着使用非易失性存储器高速（NVMe）接口规范的技术从 HDD 发展到 SSD，再发展到 SCM，存储性能得到了巨大提升。通过 NVMe 访问存储介质所花时间相比以前的硬盘技术减少了 1000 倍。不同技术之间的样本搜索时间存在差异，HDD 为 2-5 毫秒，SATA SSD 的搜索时间为 0.2 毫秒，NVMe SSD 样本需花费 0.02 毫秒<sup>[12]</sup>。SCM 通常比 NVMe 闪存 SSD 快三到五倍。

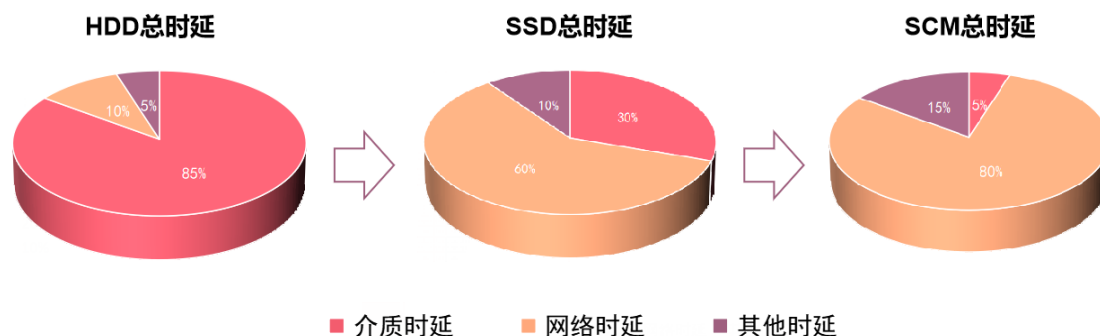


图 6 HDD 和 SSD 的端到端时延故障

NVMe 2.0 在企业级 SSD 方面进行重大调整，覆盖大量企业级用例，简化企业级存储池管理，加快推动新技术应用。采用全新存储管理机制，明确 NVM Set 和 Endurance Group 创建和管理方式。NVMe 2.0 存储管理机制能够实现灵活创建和删除 NVM Sets 与 Endurance Groups，动态管理 SSD 容量，针对静态配置产品，解决主机与驱动之间的分工需求，根据不同用例进行 SKU 配置<sup>[13]</sup>。支持 ZNS (Zoned Namespaces, 分区命名空间) 等新概念 SSD 技术。基于 NVMe 1.4a 版本，NVMe 2.0 正式引用 ZNS 技术，突破了传统区块设备模型的存储技术概念。NVMe SSD 包括 SSD 控制器、闪存空间和 PCIe 接口，闪存空间可以划分为若干个独立逻辑空间，按照 0 到 N-1 划分的逻辑空间被称为命名空间。采用 ZNS 技术的 SSD 比常规 SSD 性能更好，可以节约 10 倍左右的 OP 预留空间，节约 8 倍的 DRAM 缓存需求，同时减少设备端写入放大，优化吞吐量和延迟。目前该技术已经在企业级 SSD 得到应用。

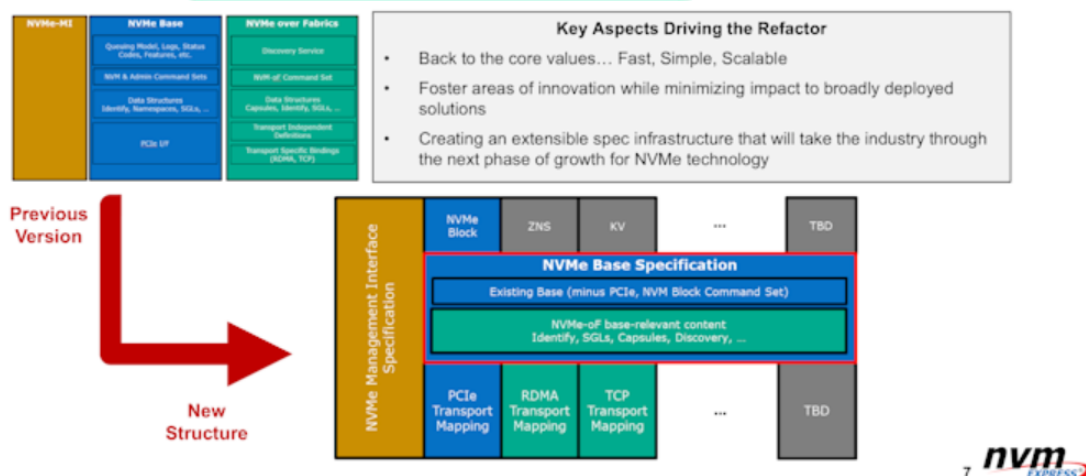


图 5 NVMe 2.0 规范调整

图片来源：NVM Express



## 2. 企业级 SSD 创新技术分析

### 2.1. 企业级 SSD 部件技术创新

#### 2.1.1. 主控

##### (1) ARM 架构

ARM 架构采用 ARM 指令集系统，根据适用范围的不同，开发差异化处理器体系结构。ARM 架构采用精简指令集（RISC）设计，通过高时钟频率、短周期执行指令集，能够有效降低芯片复杂度。Load/Store 指令架构使处理器可以根据 Load 和 Store 指令存取内存数据，RISC 架构处理器中均包含多种存放指令和数据的寄存器，ARM 架构亦是如此，ARM 处理器虽然没有采用单周期指令，但是为了便于指令解码，绝大多数指令采用定长设计，在进行运算时，CPU 只处理寄存器内数据。除了 ARM 指令集，ARM 架构还可以支持 32 位或者 16 位 Thumb 指令集，它能够提供更强的代码密度，与 ARM 指令集交互工作，进一步增强处理器性能。与 x86 架构的复杂指令集（CISC）相比，ARM 架构有效提升了常用操作的运行效率。

Arm v9 架构重点关注人工智能、安全性和专用处理，极大提升整体性能，支撑机器学习、数字信号处理、图像和语音识别等任务需求。Arm v9 支持 Aarch32 和 Aarch64 指令、NEON、加密扩展、信任区等，在 Arm v8 的基础上，新增了可缩放矢量扩展 2.0（SVE2）、Arm 机密计算架构（CCA）、跟踪和调试、事务内存扩展（TME）等功能<sup>[14]</sup>。SVE2 在部分 Arm 超级计算机内核中增加定点运算，向量长度为 128bit，最高可达 2048bit，面向专用 DSP 和 XR（增强和虚拟现实）工作负载，应用于 5G、基因学、计算机视觉等多个领域<sup>[15]</sup>；Arm 机密计算架构（CCA）包括领域管理扩展（RME）和 Arm 机密计算固件架构两个部分，建起全新的安全硬件环境，同时简化硬件设计，增强机密计算的可扩展性、再利用性和可移植性；跟踪和调试的分支记录缓冲区扩展（BRBE）提供信息分析能力，可用于调试和优化，嵌入式跟踪扩展（ETE）和跟踪缓冲区扩展（TRBE）改进了 Arm v9 的跟踪功能；事务性内存扩展（TME）支持 Arm 架构的硬件事务性内存（HTM），解决高并发写入、多线程编程问题，并通过减少锁争用导致的序列化，使得粗粒



度、线程级的并行性可以随着 CPU 数量的增加而扩展。

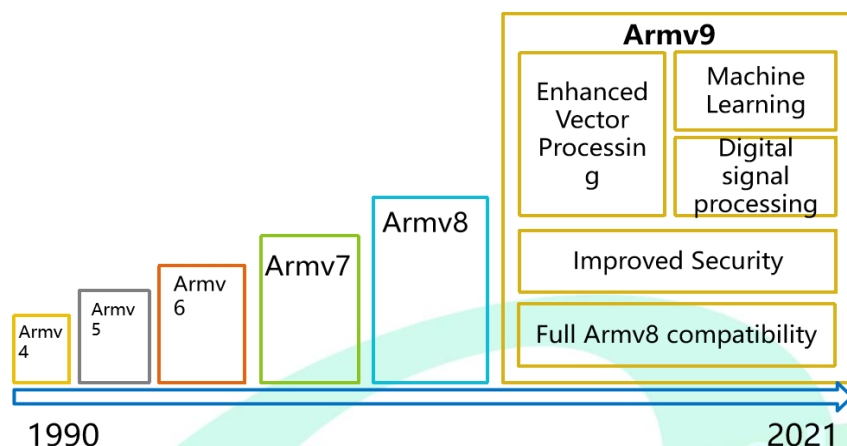


图 6 Armv9 架构优势

图片来源：ARM

## (2) Cortex 处理器

Cortex 处理器由 Cortex-A 系列、Cortex-R 系列和 Cortex-M 系列组成，面向不同的任务要求，提供差异化功能。Cortex-A 系列处理器面向承担复杂计算任务的设备，多用于企业级和数据中心 IT 设备，要想使 Cortex-A 处理器保持均衡的性能、功耗和尺寸，需要安装前端处理器，即在 Cortex-R 处理器上增加性能；Cortex-R 系列处理器针对高性能进行了优化，满足实时处理和低时延要求，是企业级 SSD、消费级 SSD 和数据中心中的常用处理器；Cortex-M 系列用于离散处理和微控制器，可以通过数据快速移动实现低时延和实时性，极低功耗的消费级控制器通常会选择此类型号，但 Cortex-M55、Cortex-M33 和 Cortex-M0+ 等处理器可以为 SSD 提供辅助<sup>[16]</sup>。Cortex 处理器系列如表 3 所示。

表 3 Cortex 处理器系列

| ARM Cortex-A   | ARM Cortex-R   | ARM Cortex-M   |
|--|--|--|
| 面向承担复杂计算任务的设备  | 针对高性能进行了优化，满足实时处理要求  | 用于离散处理和微控制器  |
| SSD 主控的典型处理器类别   |  |  |
| <ul style="list-style-type: none"> <li>✧ Cortex-A55</li> <li>✧ Cortex-A53</li> </ul> | <ul style="list-style-type: none"> <li>✧ Cortex-R82</li> <li>✧ Cortex-R8</li> <li>✧ Cortex-R5</li> </ul> | <ul style="list-style-type: none"> <li>✧ 拥有可选 ARM 自定义指令和协处理器接口 Cortex-M55 和 Cortex-M33</li> <li>✧ Cortex-M7</li> <li>✧ Cortex-M0+</li> </ul> |

**Cortex-A55 是 Arm 最强大、最高效的中端系列处理器，以 DynamIQ 技术为基础，适用于企业级 SSD 和数据中心设备等复杂计算任务。** Arm Cortex-A55 处理器采用 DynamIQ 技术、Armv8-A 扩展架构和专用机器学习指令，是第一代 CPU 应用的组成部分。Cortex-A55 使用经过重新设计的微架构系统，显著提高处理器性能，在规格和能效方面形成较强竞争力。与的 Cortex-A53 相比，它的性能提高了 18%，功率效率提高了 15%。作为可扩展的 CPU，Cortex-A55 可用于解决从边缘到云计算的各种应用场景。Cortex-A55 既可以在独立应用中实现，也可以作为 Cortex-A7x 系列的“缩小版”CPU。

**Cortex-R 系列处理器能够为企业级 SSD 主控提供有效支撑，满足实时性高、复杂性强的任务要求。**Cortex-R5 处理器为实时应用程序提供了扩展的故障遏制。Cortex-R5 处理器在 Cortex-R4 的基础上，增强了错误管理、扩展功能安全和 SoC 集成等方面的性能，可用于深度嵌入式实时系统和关键安全系统。Cortex-R8 是该系列产品中性能最强实时处理器，使用 Armv7-R 架构的 32 位指令集内核、11 级流水线和超标量无序执行，可以从单核配置扩展到四核配置，充分释放工作负载并行性，单核 Cortex-R8 也可以根据工作负载进行断电<sup>[17]</sup>。Cortex-R82 实现 64 位 Armv8-R AArch64 架构的处理器，通过 Arm Neon 技术实现机器学习负载加速，同时采用可选内存管理单元(MMU)来支持复杂操作系统(OS)，适用于数据中心、企业级 SSD 和计算存储驱动器(CSD)。

### 2.1.2. 固件

#### (1) 固件算法

**固件是增强企业级 SSD 性能的重要组成，高效高质的固件算法可以提升固件性能，实现闪存和主控的完美兼容，增强 SSD 整体性能。**固件算法由垃圾回收、损耗均衡、坏块管理、数据保持、断电恢复等多种算法组成。由于同一个物理块不能无限次擦除<sup>[18]</sup>，因此需要通过垃圾回收将现有物理块的有效数据迁移到空闲块，通过算法回收已写入或被标记删除的无效数据，进而擦除整个块数据，使其变为可用状态；损耗均衡包括动态损耗均衡和静态损耗均衡两种，动态算法是指写入新数据时，优先选择损耗少的物理块，静态算法是把冷数据迁移到损耗多的物理块上，从而释放可写空间；在坏块管理方面，坏块作为读写擦异常的物

理块，需要进行管理和标记，坏块管理算法通常采用略过策略或替换策略；在**数据保持**方面，数据存放时间太久会出现电荷泄露，物理页多次读取会对周围页数据产生干扰，需要通过数据保持算法监测物理块数据，将超过 ECC 设定阈值或物理页读取次数阈值的块数据迁移到其他物理块上；**断电恢复**针对物理页在未写完时发生的断电，断电恢复算法采用将映射表写入闪存的方法，在发生断电时可以找回之前的映射表，实现断电恢复。

## (2) 纠错算法

**纠错算法技术能力是提升企业级 SSD 整体水平的关键，增强算法研发设计能力至关重要。**LDPC (Low-Density Parity-check Codes, 低密度奇偶校验码) 编码是一种企业级 SSD 常用的软信息编码技术，能够延长使用寿命，保存更长时间数据；用户数据与校验数据比例会不同以往；与信噪比和 UBER 有关，并不是简单的固定 bit 数纠错。在 LDPC 纠错码方面，为了优化纠错性能和编解码算法效果，需要关注 LDPC 校验矩阵<sup>[19]</sup>。在设计时需要将错误平层 (Error Floor) 降低到不影响系统性能的水平，减弱编解码算法复杂性，同时控制芯片功耗、系统性能和制造成本。在编解码算法方面，由于闪存颗粒读写次数增加会导致较高的错误率，可以通过升级 LDPC 编解码模块对应的接口协议延长使用周期，并建立解码算法流程自适应调整机制，在保证功耗最少、时延最低的情况下完成解码任务。

**ECC 技术逐渐成为提升企业级 SSD 核心竞争力的有效方式和性能保障。**随着 NAND 技术的不断迭代和单元内比特数量的不断增多，使得数据错误率随之增长，提升 SSD 纠错能力和纠错算法技术能力，成为 SSD 性能提升的重要方向。ECC (Error Correcting Code, 纠错码) 技术在编码时加入冗余位，使闪存芯片在读取数据时极大减少 BER (Bit Error Ratio, 比特出错率)，对于高性能芯片具有重要意义<sup>[20]</sup>。国内厂商积极研发 ECC 技术，设计出高效译码算法，响应高吞吐、低时延、强纠错存储要求。以英韧科技为例，针对功耗、复杂度和吞吐率的差异化需求，研发出多种不同性能的解码专利算法，同时利用机器学习和人工智能技术对各种解码算法进行结构和参数优化，同时将 ECC 引擎做成了指令集的形式，能够灵活适配各种闪存颗粒。英韧的 PCIe SSD 各类主控芯片 Shasta+ 及 Rainier



系列均支持 4K LDPC，纠错能力可以完全覆盖 QLC NAND。

### (3) Telemetry

Telemetry 是一种可以远程的从物理设备或虚拟设备实时采集数据的技术。Telemetry 实现对设备、网络、协议、Overlay、业务进行多层网络健康度全面监控并评估生成可视化结果，从而实现网络故障快速定位，帮助运维人员解决网络运维问题。这种技术相比传统的网络监控方式，不仅快速，而且直观，为智能化以及精细化网络运维提供了重要的技术支撑和整体的解决方案。

SSD 产品中引入 Telemetry 属性，通过定制的 Extend SMART 日志涵盖了磨损、失败操作计数和内部异常状态相关的三大类属性。在磨损相关属性方面，磨损相关属性包含 NAND 级别的读写量，SSD 内部块的磨损程度等信息，可用于检测由磨损导致的 SSD 故障；在失败操作计数属性方面，失败操作计数属性包含 SSD 内部读写以及擦除失败等相关计数，能反映出 SSD 内部 NAND 数据存储的可靠性；在内部异常相关属性方面，内部异常相关属性包含内部温度过高，缓冲溢出等由负载量和外部环境导致的内部状态的异常，很好地反映了 SSD 内部工作状态的稳定性。

通过多个维度上对 SSD 内部状态进行监控，全面了解 SSD 内部组件的健康状况。监控更全面，对于 SSD 的各个硬件组件都有记录；记录更详细，对于温度等有更详细的记录，故障日志更详细；更贴近应用层，对于各种工作负载记录更详细。通过引入 Telemetry，能够进一步提升故障盘的预测准确度，也能大大提高数据中心云存储系统的智能运维效率。

#### 2.1.3. 闪存

##### (1) QLC 技术

**QLC SSD 大容量、低成本的特性，能够增强存储密度和服务器集成度。**QLC SSD 与 HDD 相比，具有更明显的随机读性能优势，为服务器和数据中心带来更低的读延迟。QLC SSD 大容量技术难度低于 HDD，发热少于 HDD，随机速度更快，抗震性能更强，单位体积容量密度更高。QLC SSD 适用于实时性要求更高的数据读取密集型应用，如 AI 计算、机器学习、金融数据实时分析和各种在线



大数据信息挖掘等。与此同时，QLC 与 SLC、MLC、TLC 等闪存芯片相比，具有明显的成本优势。在海量数据的应用场景中，QLC SSD 的成本竞争力为其在数据中心领域赢得更大的应用场景。存储需求的快速增长带动 QLC SSD 等大容量、高密度的新型闪存芯片不断扩大应用范围，极大地节省数据中心面积，降低运营成本。

QLC 可以为 3D NAND 提供更大的存储空间，适用于读取量大的工作负载，而 TLC 的写入性能更强，出错率更低。相较于 2D NAND，3D NAND 具有更高密度，可以实现更高的写入和擦除次数。由于 QLC NAND SSD 每个单元可以存储 4 位数据，有 16 个电压状态，但 TLC NAND SSD 只能存储 3 位数据，拥有 8 个电压状态。在读取性能方面，与 QLC NAND 相比，由于 TLC 每单元的写入位数少，因此需要更长的擦除周期和程序时长，而 QLC NAND SSD 的顺序读取速度与 TLC SSD 一样快，更适用于读取量大的工作负载。而在 3D NAND 写入性能方面，TLC SSD 要优于 QLC SSD，尽管这两种技术都采用纠错算法来保证数据的完整性，但该过程在基于 QLC 的驱动器上会花费更多的处理周期，严重影响写入性能。TLC 和 QLC 将呈现相辅相成、协同发展趋势。未来，50% 的 NAND 闪存将是 3D QLC，剩下的部分中绝大多数将由 3D TLC 完成。此外，QLC NAND 将逐渐取代部分 TLC NAND，这两项技术将一起向前发展。

## (2) SLC Cache 技术

SLC Cache 规避了高密度存储的减速问题，为 TLC SSD 或者 QLC SSD 提供写入性能、存储容量和成本优势。SLC Cache 本质是用 TLC、QLC 等颗粒模拟 SLC 的工作模式，将 TLC、QLC 等多个工作状态按照 SLC 的两个状态进行标记，只判断最高位状态，简化控制难度的同时，提高速度和耐久性。SLC Cache 的加速策略包括动态容量和固定容量两个类型。动态容量的 SLC Cache 是指根据剩余容量来规划 SLC Cache，对于大容量 SSD 可以实现大缓存冗余，但是随着硬盘占用率提高，整体性能随之下降，这对主控智能化提出更高要求，但是随着 SSD 容量增长，动态掉速问题得到极大缓解。图 9 中蓝色线条显示了 IBM 增强后 QLC 的耐用性，即在使用 SLC Cache 动态容量后，数据按照 QLC 闪存最优方式运行<sup>[21]</sup>。固定容量是根据应用场景需要，为 SSD 设置固定容量，固定容量

的 SLC Cache 每使用一部分，就会立即重新动态划分新的 SLC Cache，从而保证自始至终的高速写入。

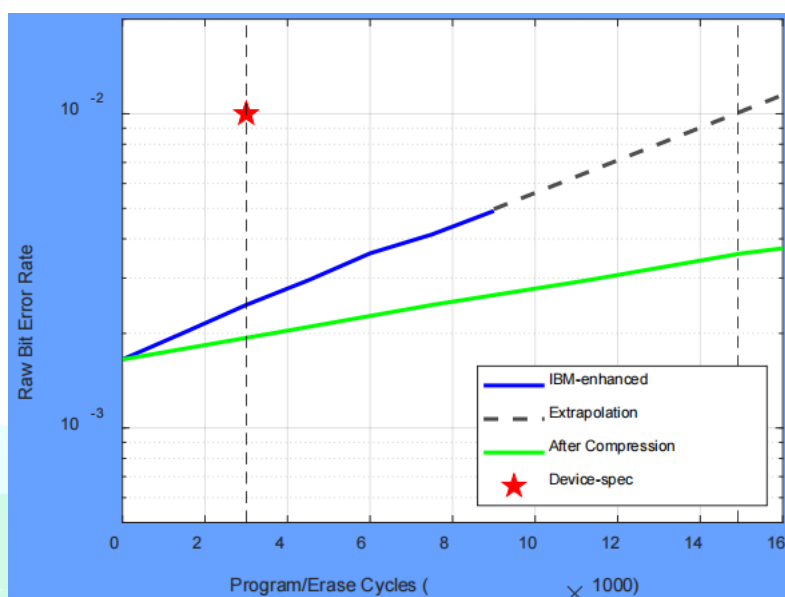


图 7 IBM FlashCore Module QLC 的动态 SLC 分配技术

图片来源：IBM

### (3) Xtacking 晶栈

**Xtacking 晶栈创新 3D NAND 闪存技术，使得速度更快、密度更高、架构更灵活。** Xtacking 晶栈可实现在 CMOS 外围电路晶圆、NAND 存储阵列晶圆两片独立的晶圆上分别加工外围电路和存储单元，当两片晶圆各自完工后，Xtacking 技术只需一个处理步骤就可通过数十亿根垂直互联通道 (VIA) 将两片晶圆键合，合并为牢固的整体，让 NAND 获取更高的 I/O 接口速度及更多的操作功能。在传统 3D NAND 架构中，外围电路约占芯片面积 20% 到 30%，Xtacking 晶栈技术将外围电路置于存储单元之上，从而实现比传统 3D NAND 更高的存储密度，芯片面积可减少约 25%。模组化的工艺充分利用存储单元和外围电路的独立加工优势，提升研发效率，缩短生产周期，为引入 NAND 外围电路以实现闪存定制化提供可能。Xtacking 晶栈 2.0 将充分利用架构优势，进一步提升 NAND 吞吐速率、提升系统级存储的综合性能、开启定制化 NAND 全新商业模式等。长江存储采用先进的 Xtacking 技术，在一个芯片里面把 NAND 存储单元和 IO 接口分开独立设计与加工，于 2020 年 4 月成功推出 128 层 TLC 和 QLC 两款产品，其接口速度达到 1600MT/s。

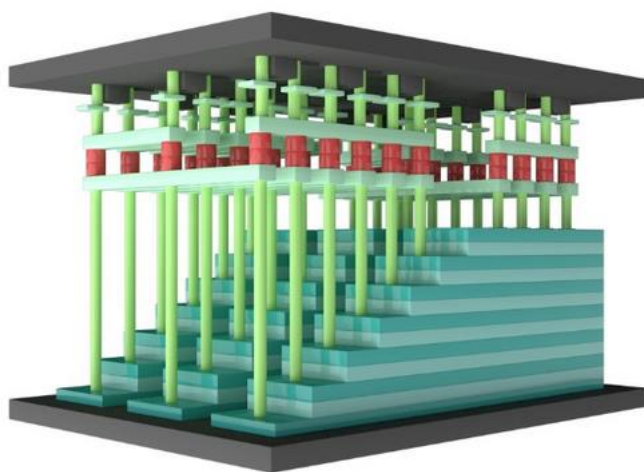


图 8 长江存储 Xtacking 技术

图片来源：长江存储

## 2.2. 企业级 SSD 接口技术创新

### 2.2.1. PCIe 接口

**PCIe 采用差分传输形式，提高抗干扰能力，支持更大规模的数据传输。**PCIe 高速串行总线作为差分传输，能够通过两根近距离传输线上传输相反、幅值相等的信号，实现对于外界净电场的辐射抵消，同时保证两条信号线在受到干扰时产生同等变化，显著增强干扰抑制能力。PCIe 连接中每条通道包含两对导线，分别用于数据包的发送和接收，并采用全双工传输设计，允许同时发送和接收数据。最小的 PCIe 连接“x1”仅有一个通道四根导线，单方向传输速率为一个时钟周期内传输 1 比特。不同 PCIe 规范所要求的传输速率不同，PCIe x16 单方向传输速度可达到 6.4GB/s。PCIe 的高速传输，能够支持 GB 级网络连接、声卡、应用等多场景要求。

### 2.2.2. SAS 接口

**SAS 接口具备更为庞大的协议体系，适应企业级 SSD 差异化应用需求。**SAS (Serial Attached SCSI, 串连式 SCSI) 接口采用电脑集线的技术，用于外围零件的数据传输，包括硬盘、CD-ROM 等，根据连接的不同设备选用相应协议进行数据传输。SAS 接口由三种传输协议组成，一是串行 SCSI 协议 (SSP)，用于和 SCSI 设备间的数据传输；二是串行 ATA 通道协议 (STP)，用于和 SATA 设备间的数



据传输；三是 SCSI 管理协议 (SMP)，用于维护和管理 SAS 设备。

SAS 技术通过双端口设计、全双工模式、后端存储网络等方式，优化企业级 SSD 操作系统效能，增强可用性、扩展性和兼容性。SAS 双端口功能是应用于企业级 SSD 的关键设计。SAS 接口的信号区和电源区之间没有缺口，并在此基础上增加了一组传输信号以实现双端口功能，为企业级的多路径应用提供支持。而 SATA 接口中间存在缺口，因此可以正常连入 SAS 背板连接器，确保 SAS 能兼容 SATA。SAS 接口双端口功能和全双工工作模式，使其在收发信号时，拥有两对差分信号对。通过 SAS 网络可以将一定量的存储盘进行互连，形成后端存储网络，涵盖内部连接线缆和外部连接线缆两大类。常用于内部的硬盘连接线缆，同时带有电源和信号线，可以直接连接硬盘；为节约空间，使用内部 Mini SAS 线缆连接器，其信号定义如表 4 和接口定义图 11 所示，最多可以支持 4 个 PHY 的互连，内部 Mini-SAS HD 接口体积则更小。SAS 外部互连线缆同样可以采用 mini-SAS 和 mini-SAS HD 进行连接，并使用支持 4PHY SAS 互连的 QSFP 连接器。

表 4 内部 Mini-SAS 线缆连接器信号定义

| 信号         | PHY 支持数 |     |     |     | PHY 匹配级别 |
|------------|---------|-----|-----|-----|----------|
|            | 1 个     | 2 个 | 3 个 | 4 个 |          |
| RX 0+      | A2      | A2  | A2  | A2  | 3 级      |
| RX 0-      | A3      | A3  | A3  | A3  |          |
| RX 1+      | 未连接     | A5  | A5  | A5  |          |
| RX 1-      | 未连接     | A6  | A6  | A6  |          |
| SIDEBAND 7 | A8      | A8  | A8  | A8  | 1 级      |
| SIDEBAND 3 | A9      | A9  | A9  | A9  |          |
| SIDEBAND 4 | A10     | A10 | A10 | A10 |          |
| SIDEBAND 5 | A11     | A11 | A11 | A11 |          |
| RX 2+      | 未连接     | 未连接 | A13 | A13 | 3 级      |
| RX 2-      | 未连接     | 未连接 | A14 | A14 |          |
| RX 3+      | 未连接     | 未连接 | 未连接 | A16 |          |
| RX 3-      | 未连接     | 未连接 | 未连接 | A17 |          |
| TX 0+      | B2      | B2  | B2  | B2  | 3 级      |
| TX 0-      | B3      | B3  | B3  | B3  |          |
| TX 1+      | 未连接     | B5  | B5  | B5  |          |
| TX 1-      | 未连接     | B6  | B6  | B6  |          |
| SIDEBAND 7 | B8      | B8  | B8  | B8  | 1 级      |
| SIDEBAND 3 | B9      | B9  | B9  | B9  |          |

|               |  |     |     |     |     |
|---------------|--|-----|-----|-----|-----|
| SIDEBAND 4    | B10  | B10 | B10 | B10 |     |
| SIDEBAND 5    | B11  | B11 | B11 | B11 |     |
| TX 0+         | 未连接  | 未连接 | B13 | B13 | 3 级 |
| TX 0-         | 未连接  | 未连接 | B14 | B14 |     |
| TX 1+         | 未连接  | 未连接 | 未连接 | B16 |     |
| TX 1-         | 未连接  | 未连接 | 未连接 | B17 |     |
| SIGNAL GROUND | A1、A4、A7、A12、A15、A18<br>B1、B4、B7、B12、B15、B18 |     |     |     | 1 级 |

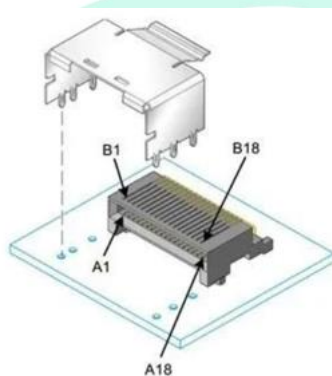


图 9 内部 Mini SAS 线缆连接器接口定义

### 2.2.3. U.2 接口

**U.2 接口具备兼容性、高速率、低延迟、低功耗优势，适用于企业级 SSD。** U.2 接口又被称为 SFF-8639（SSD Form Factor 8639），U.2 SSD 是一种高性能数据存储设备，旨在使用 SFF（small form factor，小尺寸）连接器支持 PCIe 接口，同时兼容标准化的 SAS 和 SATA 旋转磁盘和 SSD。U.2 接口能够兼容 SATA、SAS、PCIe。U.2 接口支持 1 个 SATA 接口、2 个 SAS 接口或 PCIe 企业级 SSD 硬盘中最多 4 个并行通道的 I/O，理论传输速度高达 32Gbps，而 SATA 只有 6Gbps，比 SATA 速度提高 5 倍以上。在双控制器存储阵列中，每对双通道 PCIe 链路相互独立运行，允许一个控制器出现故障，SFF-8639 连接器上已定义信号能够显示驱动器是在单端口还是双端口模式下工作。U.2 接口比 M.2 接口尺寸更大，因此最大存储容量也相应增多，同时 U.2 散热能力更强，耐受更高的工作温度，在高密度写入操作时，U.2 性能不会因为驱动器升温而降低性能。

## 2.3. 企业级 SSD 传输协议创新

### 2.3.1. NVMe

NVMe 简化了协议复杂性，显著提高了 SSD 的读写性能。NVMe 充分利用 PCIe 通道的低延时以及并行性，通过降低协议交互时延，增加协议并发能力，并且精简操作系统协议堆栈。NVMe 协议支持大规格 IO 队列和大并发，最大支持 64K 个 IO 队列，每个 IO 队列支持 64K 个 IO 并发。通过 IO 队列和 CPU 核对应，避免多个 CPU 核竞争 IO 队列，IO 流程天生免锁设计，实现更低的时延，降低 CPU 资源消耗，充分发挥 NVMe SSD 的性能。

和传统 SCSI 体系比较，NVMe 协议更为简单高效。从协议路径上看，在主机侧，NVMe 驱动协议栈去掉了传统 IO 系统中的 IO 调度层和复杂的 SCSI 协议层，换成轻量级的 NVMe 协议。从硬件传输路径上看，在存储侧，无需 SAS 控制器和 SAS Expander，SSD 通过 PCIe 总线连接，实现更低的时延，更大的带宽，更少的 CPU 资源消耗，如图 12 所示。

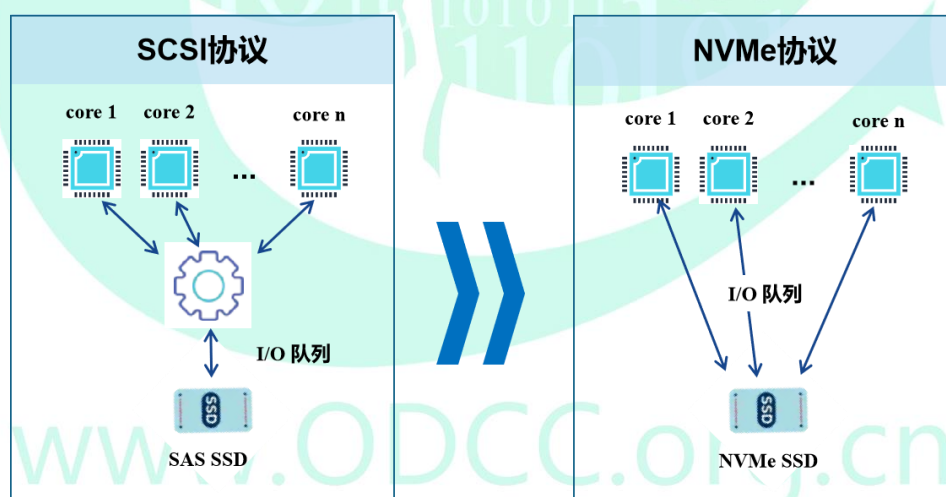


图 10 NVMe 与 SCSI I/O 队列对比

### 2.3.2. NVMe-oF

NVMe-oF 可以实现大规模计算机集群与其他多个设备互联，把 NVMe 协议在单系统中的高性能、低延迟和低协议负担优势应用于整个数据中心。NVMe-oF (NVMe over Fabrics) 协议使用基于消息的模型在主机和目标存储设备之间进行



通信, 能够将 NVMe 映射到以太网、光纤通道、RoCE 或 InfiniBand 等多个 Fabrics 传输协议中。NVMe-oF 使用 Fabrics (如 RDMA 或光纤通道) 代替 PCIe 在互联架构上进行传输映射, 使主机能够访问节点外的 NVMe SSD 资源。使用 Fabrics 的前提是端点之间发送和接收消息不需要共享内存。NVMe-oF 中的目标被称为命名空间, NVMe-oF 允许主机跨越更远的距离与存储目标通信, 同时保持超低的微秒级延时, 与 SCSI 和 SATA 协议相比, 将访问时间缩短了几个数量级。NVMe-oF 与 NVMe 一样, 能够支持 65000 个队列, 每个队列深度最多为 65000 个命令, 使得 NVMe-oF 能够在主机和驱动器之间实现高度并行的架构, 保证每个设备使用单独队列。

NVMe-oF 扩展了 NVMe 规范在 PCIe 总线上的实现, 把 NVMe 映射到多种物理网络传输通道, 实现高性能的存储设备网络共享访问。NVMe-oF 通过把 NVMe 映射到多个物理网络传输通道, 在各种物理传输层上实现高效 NVMe 命令, 达到降低存储网络协议栈处理开销并实现高并发、低时延的目的<sup>[22]</sup>。NVMe-oF 和 NVMe 之间的主要区别是传输命令的机制, NVMe-oF 相对 NVMe 的变化如图 13 所示, 基于网络远程访问 NVMe 设备, 实现 NVMe 设备高效共享访问, 借助 NVMe-oF, 可以实现 RACK 级别的 NVMe。使用 NVMe-oF, 能够在性能、扩展能力、易用性等方面形成优势。NVMe-oF 网络延迟低, 能够处理并行请求, 提高和优化性能, 包括存储阵列性能等, 减少服务器端 OS 存储堆栈的长度, 实现高速率和高可扩展性; 支持 NVMe 主机启动器与存储系统之间同时存在多条路径, 能够一次从许多主机和存储子系统发送和接收命令。

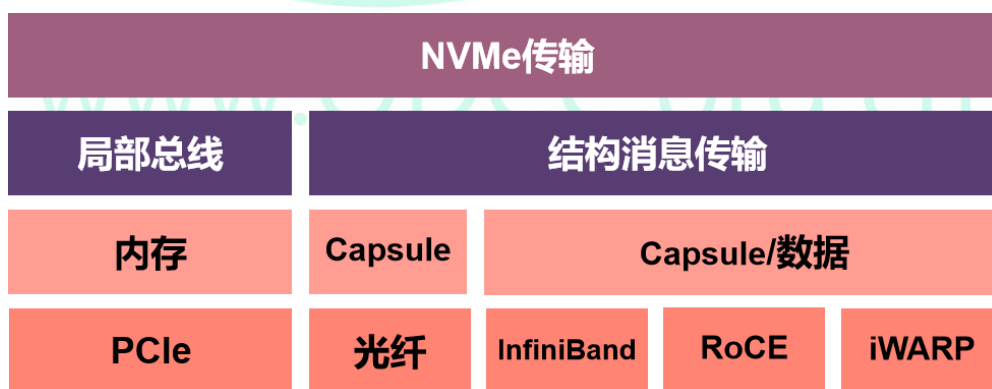


图 11 NVMe Fabrics 传输

将 NVMe-oF 与 RDMA、FC 或 TCP 一起使用, 可以形成完整的端到端

NVMe 存储解决方案，在显著提升性能的同时，通过 NVMe 实现极低时延。NVMe-oF over RDMA（NVMe/RDMA）使用 TCP 传输协议在 IP 网络上传递数据，典型的 RDMA 应用包括虚拟接口架构（Virtual Interface Architecture）、RDMA over Converged Ethernet（RoCE）、InfiniBand、omnipath 和 iWARP，其中使用最多的是 RoCE、InfiniBand 和 iWARP；NVMe over Fibre Channel（FC-NVMe）使用 16GB FC 或 32GB FC 的 HBA 和 SAN 交换机，可以将 NVMe 协议封在 FC 架构中，是 SAN 基础设施的最佳解决方案；NVMe over TCP/IP（NVMe/TCP）使用 NVMe-oF 和 TCP 协议在以太网传输数据，与 RDMA 和 FC 相比，是一种更便宜、更灵活的替代方案。



图 12 NVMe-oF 支持多种传输层协议

### 2.3.3. ZNS

Zoned Namespaces（ZNS）是从 Open Channel（开放通道）SSD 基础上发展而来的。这个概念在 NVMe 标准工作组技术提案，这个方案把 Namespace 的逻辑地址空间切分成多个 zone，因此它继承了 Open Channel SSD I/O 分离、可预测性延迟等优势。并且 ZNS 协议简化了软件架构，用户可以定制开发，所以应用范围更加广泛。

ZNS 允许根据数据的使用和访问频率对数据进行分组，并按顺序存储在 SSD 的独立区域中。无需移动和重新排列数据，ZNS SSD 可以显著减少写入操作的数量，降低驱动器的写入放大因子（WAF）——与主机系统最初指示的写入相比，

驱动器执行的实际写入量。WAF 越接近 1，SSD 的效率就越高，持续的时间就越长。



图 13 ZNS 读写方式

图片来源：三星

ZNS SSD 可以实现 WAF (Write amplification) 接近 1，相比于一般服务器 SSD 的 WAF 值在 3 到 4 之间，这是一个重大进展。这将使驱动器的使用寿命是传统的非易失性内存主机控制器接口规范 (NVMe) SSD 的四倍，使其成为服务器基础架构更环保、更可持续的解决方案。ZNS 还允许用户利用 SSD 的全部容量，消除对预留空间的需求，而预留空间要求为背景任务保留一些存储空间。这些改进使企业客户能够以更高的效率处理大数据和人工智能应用程序。

### 3. 企业级 SSD 测试技术分析

#### 3.1. 测试工具

企业级 SSD 测试时需要采用性能测试工具和读写跟踪工具。性能测试工具 iometer 是一个在 Windows 环境下测试吞吐量及响应时间的软件；fio 是一个主要在 Linux 环境下运行的工具，它也可以用于 Windows 环境，涵盖大量的底层参数，适合作为基础测试工具；iozone 是一个开源的、多操作系统的、文件系统性能测试工具，通过调用文件系统 API 以及多种文件尺寸，记录尺寸组合来模拟用户使用场景。在读写跟踪工具方面，iostat 是 Linux 开源工具 sysstat 监控块设备读写的组成部分，可以定时采集主机与块设备之间的数据流量及命令响应时间等



统计信息；block trace 是 Linux 操作系统的块设备访问跟踪工具，可以记录主机对块设备发出的命令、块地址和数据块尺寸信息，用于指令流分布和吞吐量分析；process monitor 是基于 Windows 操作系统的设备访问跟踪工具，其中的文件访问跟踪部分可以记录主机对存储设备以及文件系统发出的命令、偏移量、数据尺寸等信息，收集用于详细统计的基础数据。

## 3.2. 测试指标

### 3.2.1. 性能测试

性能测试包括基础性能测试和用户场景测试两个部分。基础性能测试用于了解企业级 SSD 的最高性能，包括顺序读写的最大带宽和随机读写的最大 IOPS，因此在顺序读写之前需要进行盘的初始化，保证初始状态与设备原始状态的相似性。在基础性能测试中，需要先测试顺序读写，而后做两个以上全部容量的预处理，再做随机读写的 IOPS 性能测试，全部读写测试需要记录温度，由于温度过高会导致固态硬盘进入超温保护状态，会出现性能降低。用户场景测试需要对企业级 SSD 做预处理，在达到稳态之后进行测试。用户场景测试需要使用 iohome 进行文件大小、数据块尺寸、数据访问方式的全面测试，根据目标客户应用场景，进行数据块尺寸分布统计以及数据方位位置统计，使用 FIO 的 bssplit 参数以及 random\_distribution 参数输入到测试用例中得到模拟客户现场应用的结果。

SSD 性能测试要从分离变量出发，根据强相关性指标选用适合的测试用例和测试方法。SSD 数据块大小和访问方式对顺序读写的性能有很大影响，Endurance/Data retention、读写次数增加、开机时间延长等因素都会增加介质误码率和纠错时间，降低使用性能；温度也会对 SSD 性能产生影响，超过设计温度范围时，读写会被强制减速；SSD 技术迭代产生的复杂度要求部件使用更复杂的底层操作，而某些技术复杂度的增加会降低 SSD 存储部件的读写性能。图 16 展示了 NVME G4 样品的顺序读写性能，横轴是测试时间，纵轴是读写带宽，从图中可以看出，NVME 固态硬盘的随机读写性能分布很广，小数据块对应的读写带宽远远低于大数据块对应的读写带宽。而从图 17 可以看出，队列深度为 1 的 128k 数据块带宽表现优于队列深度为 64 的 4k 数据块，说明数据块尺寸与队列

深度相比，数据块尺寸影响更大，即大尺寸数据块能有效提高带宽利用率。

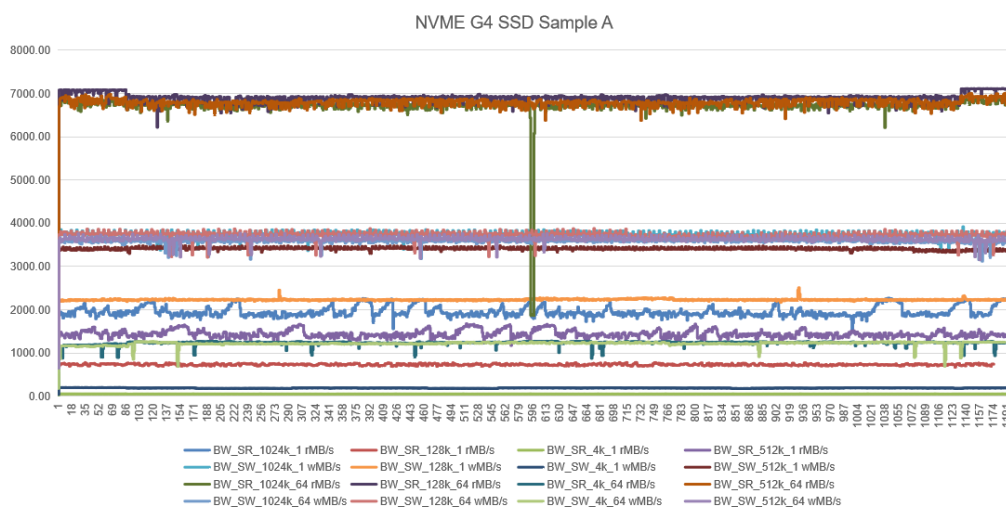


图 14 NVME G4 样品的顺序写性能

图片来源：忆联

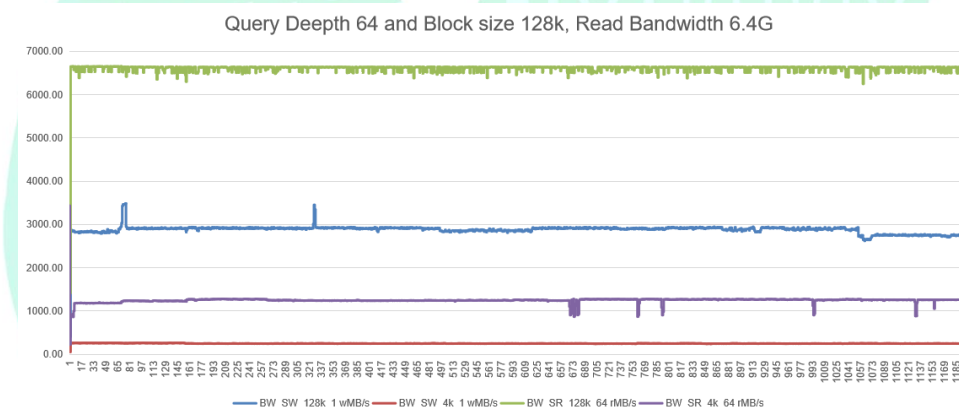


图 15 数据块尺寸与队列深度效果对比

图片来源：忆联

### 3.2.2. 可靠性测试

MTBF (Mean time between failures, 平均无故障时间) 是指相邻两次故障之间的平均工作时间, 是衡量企业级 SSD 可靠性的重要参数指标。当前市场主流企业级 SSD 的 MTBF 指标值约为 200 万小时, MTBF 越长表示可靠性越高, 保持正确工作能力越强。它反映了产品的时间质量, 是体现产品在规定时间内保持功能的一种能力, 但 MTBF 仅适用于可维修产品, 当产品的寿命服从指数分布时, 失效率的倒数表示两个失效之间的时间间隔 ( $\lambda = 1/\text{MTBF}$ )。

MTBF 主要通过实证法, 采用加速应力方式来证明产品长期可靠度, RDT

(Reliability Demonstration Test, 可靠度验证测试) 测试主要通过高温加速测试计算评估, 从测试深度、广度、持久度三个方向进行测验。深度测试是 Endurance 测试, 使用 JEDEC 标准固态硬盘耐久性工作负载, PE (Program/Erase, 擦写) 值从开始到预允许最大值验证耐久度, 即从生命周期开始到生命周期结束; 广度测试是 Quality 测试, 主要通过读、写、数据校验、Trim、Format 以及正常和异常上下电等所有用户可能的操作, 验证各种操作 Case 下的稳定性; 持久度测试是 Retention 测试, 在 SSD 闪存盘生命末期, 寿命 PE 次数达到允许的最大值之后, 投入上百块 SSD 进行 Power loss retention 测试以验证掉电后的数据保持能力。

### 3.2.3. 稳定性测试

矩阵性能测试和变长数据块混合压力测试相结合, 是测试企业级 SSD 稳定性的有效方式。SSD 性能的稳定性测试需要通过不同数据块分布、读写方式、数据块大小、队列深度以及任务数进行组合测试。采用矩阵性能测试和变长数据块混合压力测试, 是指先测试性能矩阵, 以及 Jesd219 或者自定义数据块尺寸的模拟用户现场实际操作压力, 然后再进行一次性能矩阵测试。在测试时, 抓取实时操作系统提供的带宽、IOPS、等待时间等性能数据并将结果作图, 可以得到整个测试过程中 SSD 的性能稳定性, 如图 18 所示。

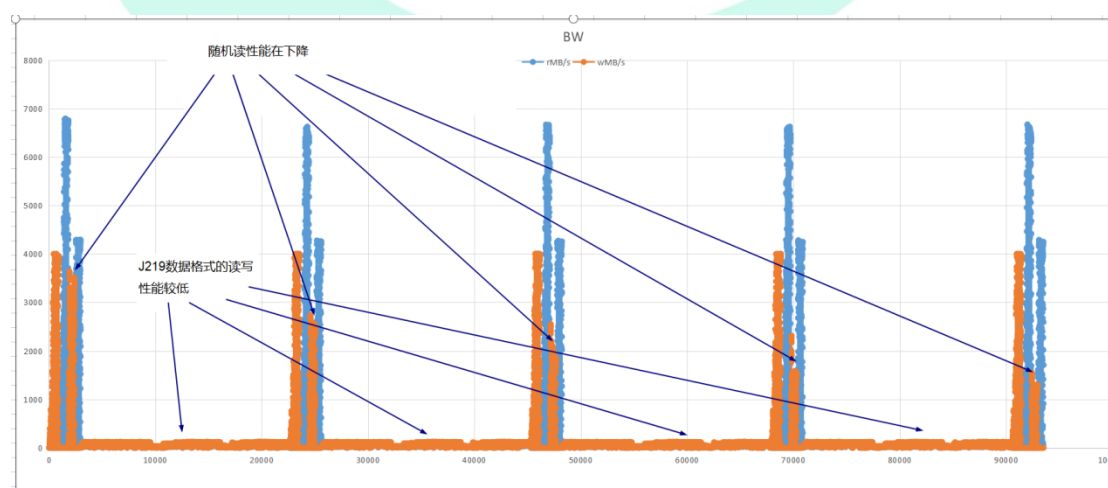


图 16 SSD 性能稳定性测试

图片来源: 忆联

完备的稳定性测试方案包括介质特性分析、寿命检测、稳定性追踪等多项内



容。需要采用智能高温控制器和 P/E Block 读写算法并行收集闪存介质的实时状态，支持 NAND 介质 High Level 指令集和 Low Level 指令集，从而完成不同生命周期阶段的稳定性测试。同时完整的稳定性测试方案还需要为用户提供自定义的介质特性控制、监测和状态数据收集服务，有效降低设备购置成本和总体拥有成本。

### 3.3. 影响因素

#### 3.3.1. 数据块尺寸

实际使用场景中的块分布与理论分布情况存在较大差异，需要采用测试工具对目标系统进行采样，使测试值尽可能接近真实值。尽管混合数据块大小读写测试可以更真实地模拟数据中心硬盘和企业级 SSD 带宽能力，但随着数据中心应用场景的增多，实际的块分布情况与 JESD219a 标准中对于块分布要求之间的差异逐渐增大。为了使测试结果更加真实可靠，可以采用 block trace、process monitor 等多种工具对目标系统进行采样，统计目标系统中实际读写过程的数据块分布情况，根据 bssplit 参数进行测试。这些工具根据获取到的主机对设备的操作情况、起始读写的块地址、读写操作的数据块大小等信息，统计一段时间内读写指令中的不同尺寸数据块分布，得到数据块尺寸列表和各尺寸数据块占比，并将这些信息输入到测试负载，就可以获得该应用场景下，目标平台上待测产品表现。图 19 通过统计某个应用软件在 4 种应用场景下的数据块分布，采用 FIO 的 bssplit 参数将块分布输入到测试数据流，得到该数据块分布在目标系统里采用待测产品的表现。

## 四种应用场景以及模拟结果

| Access     | %      | #      | Access     | %      | #      | Access     | %      | #       | Access     | %      | #       |
|------------|--------|--------|------------|--------|--------|------------|--------|---------|------------|--------|---------|
| SEQ 4K W   | 24.80% | 14,852 | SEQ 4K W   | 21.90% | 16,224 | RND 64K R  | 35.80% | 318,184 | SEQ 0.5K W | 22.70% | 429,780 |
| RND 16K W  | 17.70% | 10,599 | RND 16K W  | 14.80% | 10,970 | RND 8K R   | 18.30% | 162,171 | RND 64K R  | 14.90% | 283,247 |
| SEQ 16K W  | 14.40% | 8,634  | SEQ 16K W  | 11.70% | 8,645  | SEQ 8K R   | 11.60% | 102,630 | SEQ 8K R   | 10.20% | 193,473 |
| SEQ 0.5K W | 13.90% | 8,344  | SEQ 0.5K W | 11.40% | 8,403  | SEQ 0.5K W | 3.10%  | 27,852  | RND 8K R   | 9.50%  | 180,015 |
| SEQ 1K W   | 5.80%  | 3,461  | RND 4K W   | 10.20% | 7,522  | SEQ 512K R | 2.95%  | 26,252  | RND 4K W   | 4.90%  | 92,792  |
| RND 4K W   | 4.50%  | 2,704  | SEQ 1K W   | 4.70%  | 3,471  | RND 16K R  | 2.19%  | 19,482  | RND 8K W   | 3.80%  | 72,545  |
| RND 1K W   | 2.65%  | 1,586  | RND 8K W   | 3.70%  | 2,772  | RND 24K R  | 2.19%  | 19,422  | RND 4K R   | 2.81%  | 53,231  |
| RND 28K W  | 2.54%  | 1,522  | RND 1K W   | 2.55%  | 1,885  | RND 32K R  | 2.16%  | 19,234  | RND 16K R  | 1.99%  | 37,718  |
| SEQ 1.5K W | 2.44%  | 1,461  | RND 28K W  | 2.13%  | 1,575  | SEQ 16K R  | 2.10%  | 18,690  | SEQ 4K R   | 1.63%  | 30,842  |
| RND 8K W   | 2.32%  | 1,393  | SEQ 1.5K W | 1.98%  | 1,468  | RND 40K R  | 1.77%  | 15,733  | RND 32K R  | 1.51%  | 28,685  |
| RND 0.5K W | 1.92%  | 1,150  | RND 0.5K W | 1.84%  | 1,358  | RND 4K R   | 1.74%  | 15,429  | RND 24K R  | 1.29%  | 24,438  |
| RND 1.5K W | 1.41%  | 845    | RND 1.5K W | 1.60%  | 1,185  | RND 56K R  | 1.26%  | 11,153  | SEQ 1K W   | 1.26%  | 23,884  |
|            |        |        | RND 12K W  | 1.54%  | 1,138  | SEQ 12K R  | 1.14%  | 10,086  | RND 40K R  | 1.16%  | 22,029  |
|            |        |        |            |        |        | SEQ 24K R  | 1.04%  | 9,229   | SEQ 16K R  | 1.09%  | 20,599  |

| 场景 | 文件复制   | 数据库复制  | 数据维护   | 数据库日常  |
|----|--------|--------|--------|--------|
| 性能 | 1.6G/s | 930M/s | 3.2G/s | 2.5G/s |

图 17 四个应用场景下数据块分布以及模拟结果

图片来源：忆联

### 3.3.2. 写入速率

SSD 的理论性能是 PCIe 总线的最优值，但由于 NAND 单芯片写入能力有限，因此实际的 SSD 写入速度通常低于硬盘。SSD 要想达到理论性能值，需要保证多通道并行传输，同时单芯片写入能力达到最大速度。但实际上，虽然 PCIe 总线可以实现多通道并发，但是单任务、单队列、小数据块的测试用例无法满足多通道并行传输，与此同时，NVMe 固态硬盘是由多芯片组成的，而 NAND 芯片的写入时钟较低、写入过程复杂、单芯片写入速率有限，SSD 单芯片的物理写速度低于硬盘。因此，在多个芯片同时工作时，很难保证 SSD 的高速性，单队列、单任务状态下，SSD 表现速率比硬盘低是正常现象。NAND 单芯片的写入速率如图 20 所示。

|                           | 2D-TLC   | 3D-TLC    | 3D-QLC    |
|---------------------------|----------|-----------|-----------|
| Performance               | Poor     | Better    | Bad       |
| Plane per die             | 2        | Up to 4   | Up to 4   |
| Erase time (typ.)         | 5 ms/BLK | 10 ms/BLK | 10 ms/BLK |
| Program throughput (typ.) | 15MB/s   | 80MB/s    | 20MB/s    |
| Read throughput (typ.)    | 400MB/s  | 800MB/s   | 500MB/s   |

图 18 NAND 单芯片的写入速率

图片来源：忆联

### 3.3.3. 芯片误码

SSD 芯片内部构造的不一致性和电压存续时间都会导致误码，而误码纠正操作会影响到 SSD 性能。由于工艺过程，在 SSD 芯片内部会出现尺寸不一致性，造成读写操作出现信噪比，使得原始数据在解码过程前发生误码。进入 LDPC 解码器后，存在误码率的原始数据需要较长的解码时间，在设备上表现出命令响应时间延长和性能下降，如图 21 所示。SSD 中 NAND 芯片的数据存储要求浮电极上保持固定电压，在较高的温度下，NAND 浮电极里面的电子会快速逃逸，造成电压存续时间，芯片内表示某种状态的电压分布扩散，而扩散会表现为 NAND 芯片读取原始数据时的误码，如图 22 所示。NAND 读操作需要使用浮电极电压与参考电压进行比较，写-擦循环的增多也会导致某个具体状态对应的电压分布扩散，而某个编码所对应的电压分布扩散，会导致编码读操作出现误码。纠正误码的操作必然会影响 SSD 性能，使固态硬盘的读写性能出现抖动。

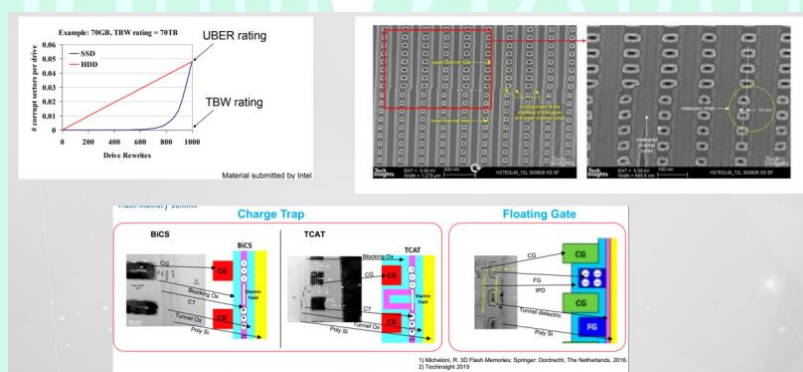


图 19 芯片内部不一致性造成误码

图片来源：忆联

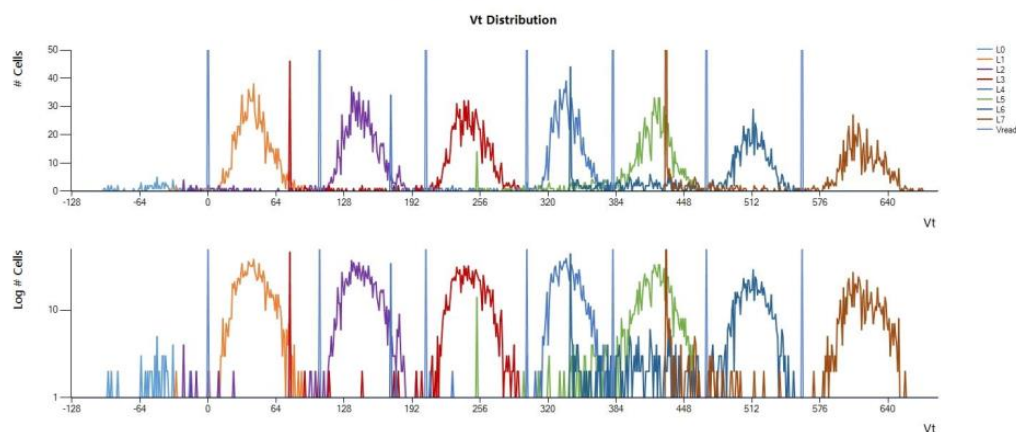


图 20 芯片内部电压分布扩散引入读误码

图片来源：忆联



### 3.4. 权威测试

权威机构进行的企业级 SSD 测试，能够提供存储类品牌在性能、稳定性、可靠性等多方面的数据参考，有效缓解了多品类、多数量企业级 SSD 产品的选择困扰。作为 ODCC 重要项目之一的企业级硬盘基准测试 BESH（Benchmark of Enterprise SSD & HDD）为用户和厂家提供了解决企业存储类产品选择困难的有效方案，以真实的测试数据为基础，为产品选型提供公正可靠的参考依据。



<http://www.odcc.org.cn/diskBenchmark.html>

BESH 测试项目持续时间长、影响范围广、参与机构多，目前已经成为业界企业级存储测试的重量级项目。2016 年，BESH 启动并接受了首批测试；2021 年，由 BESH 项目组主要参与贡献的 YD/T 3824《面向互联网应用的固态硬盘测试规范》正式发布，行业标准主要针对企业级硬盘产品的性能、稳定性、兼容性、功耗等给出了明确的测试方法。六年间，已有多家厂商参与了评测，涌现出一批优秀产品，为我国数据中心存储行业健康发展提效赋能。部分优秀产品如表 5 所示。

表 5 部分优秀产品案例

|  | 产品名称 | 企业 | 测评情况 |
|--|------|----|------|
|--|------|----|------|

|    |                          |             |                |
|----|--------------------------|-------------|----------------|
| 1  | PM9a3-E1.s               | 三星半导体       | ODCC 2021 闪耀之选 |
| 2  | Ethernet SSD             | 三星半导体       | ODCC 2021 闪耀之选 |
| 3  | PM893                    | 三星半导体       | ODCC 2021 闪耀之选 |
| 4  | DapuStor R5100           | 大普微电子       | ODCC 2021 闪耀之选 |
| 5  | D5-P5316                 | 英特尔         | ODCC 2021 闪耀之选 |
| 6  | MG09ACA18TE              | 东芝电子        | ODCC 2021 闪耀之选 |
| 7  | EXOS 20TB                | 希捷          | ODCC 2021 闪耀之选 |
| 8  | DapuStor X2900           | 大普微电子       | ODCC 2021 闪耀之选 |
| 9  | PM9A3                    | 三星半导体       | ODCC 2020 优秀产品 |
| 10 | Full Palm SSD            | 华为技术有限公司    | ODCC 2020 优秀产品 |
| 11 | Half Palm SSD            | 华为技术有限公司    | ODCC 2020 优秀产品 |
| 12 | Micron 5210 QLC SATA SSD | 美光半导体有限责任公司 | ODCC 2020 优秀产品 |
| 13 | Ultrastar DC HC550       | 西部数据        | ODCC 2020 优秀产品 |
| 14 | Ultrastar DC SN840       | 西部数据        | ODCC 2020 优秀产品 |
| 15 | 企业级 Exos X18             | 希捷          | ODCC 2020 优秀产品 |
| 16 | 企业级 Exos 2X14            | 希捷          | ODCC 2020 优秀产品 |
| 17 | PM1735                   | 三星半导体       | ODCC 2020 优秀产品 |

## 4. 企业级 SSD 产业图谱

### 4.1. 主控厂商

#### 4.1.1. 领域图谱



图 21 主控厂商领域图谱

#### 4.1.2. 发展现状

全球企业级 SSD 主控芯片持续增长，PCIe SSD 占比不断扩大。2021 年全球企业级 SSD 主控芯片出货量达到 5100 万块，相比 2020 年，增长了 13.33%，预计 2022 年仍将继续增长。企业级 SSD 中，PCIe 在 2021 年出货量达到 3140 万块，占比超过 60%，2022 年仍将持续增长；SATA 与 SAS 所占比分别为 31% 和 7%，到 2022 年，该份额将继续缩减，如图 24 所示。

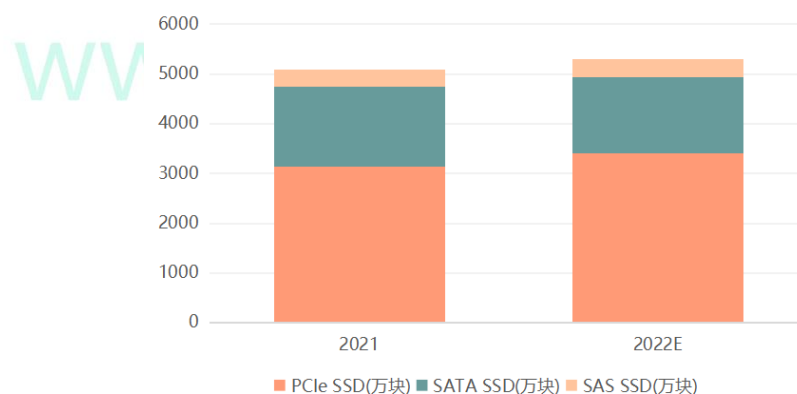


图 22 2021 年-2022 年全球企业级 SSD 主控芯片出货量



企业级 SSD 主控芯片市场国产率较低，但国内主控厂商的总体数量和研发投入正在加快增长。目前，我国 PCIe 接口 SSD 主控离量产仍存在一定差距，部分企业的关键核心技术仍依靠国外，企业级 SSD 主控市场主要被英特尔、Marvell、慧荣、三星、SandForce、群联等国际巨头企业所占领。国产主控厂商虽然面临巨大的竞争压力，但仍在积极研发生产和拓宽市场，并将自研芯片应用于存储产品中。SSD 主控芯片厂商包括三类，第一类是自研芯片并应用于产品中的企业，包括大普微电子、得一微、华澜微、忆芯科技、大唐存储等；第二类是仅从事主控芯片技术研究但不实际生产存储产品的企业，包括英韧科技和联芸科技等；第三类企业则是从第三方主控厂商购买芯片，通过自营或代工厂进行产品生产，主要包括嘉合劲威和江波龙等。

#### 4.1.3. 发展趋势

企业级 SSD 的主控芯片需要适配 PCIe Gen5 协议。5G、人工智能、云计算对数据计算提出更高要求，企业级 SSD 主控能够适应海量数据的处理要求，而 PCIe 5.0 与 PCIe 4.0 相比，带宽大幅提升，主控性能成倍增加，成为企业级 SSD 主控芯片的重要趋势。Marvell 在 2021 年发布全球首个 PCIe 5.0 NVMe SSD 主控芯片，预计其续读取速度可达 14GB/s、连续写入速度可达 9GB/s，随机读取 180 万 IOPS、随机写入 100 万 IOPS，同时实现 40% 的能效提升。微芯科技研发出企业级 PCIe 5.0 SSD 主控，支持 16 个可编程 NAND 闪存通道，提供超过 14GB/s 的吞吐量和超过 300 万次 IOPS。群联电子 2022 年发布 PCIe Gen5 的 SSD 旗舰控制芯片 PS5026，最高连续读写效能达 10000MB/s。

www.ODCC.org.cn

## 4.2. NAND Flash 厂商

### 4.2.1. 领域图谱



图 23 NAND Flash 厂商领域图谱（中国信通院整理）

### 4.2.2. 发展现状

NAND 厂商在全球企业级 SSD 市场占据绝对优势，海外巨头企业垄断 NAND Flash 市场。2021 年 NAND 在整个企业级 SSD 市场中的份额高达 87%，非 NAND 所占比重仅为 13%。三星、东芝、西部数据、海力士、美光和英特尔 6 家厂商占据全球 99% 以上的 NAND Flash 份额，中国作为全球第二大 NAND Flash 市场，芯片自给能力欠缺，长江存储等国内 NAND 企业急需缩短和国际巨头企业的技术差距。

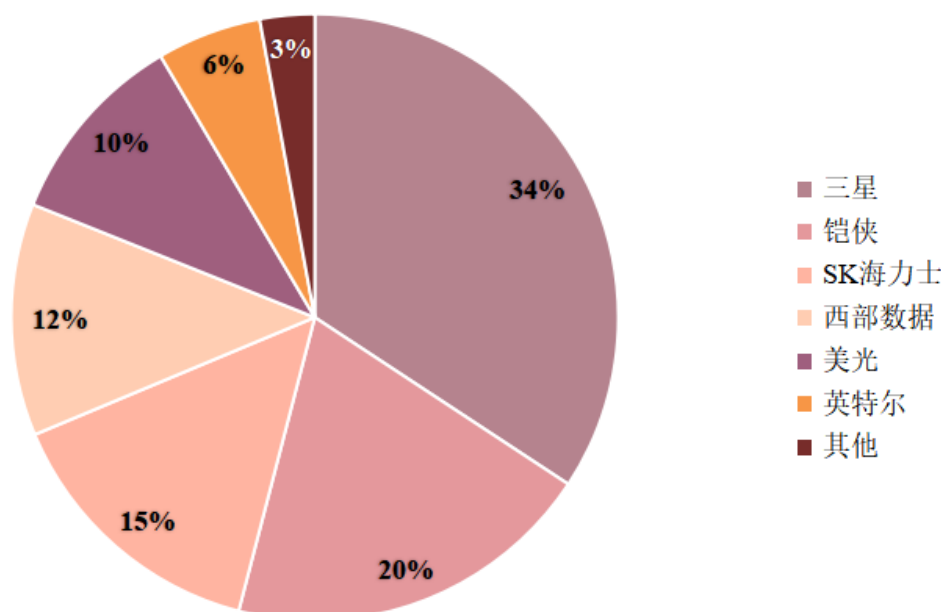


图 24 2021 年四季度 NAND Flash 各厂商市场份额（中国信通院整理）

#### 4.2.3. 发展趋势

高密度 NAND 芯片技术取得重大进展，200 层以上 NAND 芯片的出现使高密度 NAND 技术取得巨大突破。目前，主流的企业级 SSD 都已经实现了 3D NAND 技术应用，64 层芯片成为主流产品，部分 SSD 甚至达到 176 层<sup>[23]</sup>。三星预计将在 2022 年第一季度开始量产第 7 代 176 层 NAND，在 2022 年底或 2023 年上半年发布 200 层及以上堆叠的 NAND 芯片，并在 2023 年上半年开始量产；SK 海力士在 2020 年底宣布了 176 层 NAND 的消息，预计在 2023 年初推出 230 层 NAND 芯片。美光在 2020 年底发布 176 层的 NAND 芯片，铠侠正在加快研发 162 层的 3D NAND 技术。长江存储的 128 层堆叠工艺无论在容量、位密度还是 I/O 速度方面都足以与其他厂商的产品竞争，技术方面已赶上业内领跑者。

NAND 增强层数技术的同时，开始推动横向 3D、4D 闪存发展。随着堆叠层数的增加，制造闪存的成本和时间会随之提高，3D 闪存技术能够降低每比特成本，并且大幅提高每片晶圆可制造的比特数。铠侠和西部数据应用 Circuit Under Array CMOS 配置和 4Plane 操作，使得第六代 3D NAND 在写性能方面提高了近 2.4 倍，在读取延迟方面缩短了 10%。此外，I/O 性能也提高了 66%。SK 海力士预计在 2023 年初推出 4D NAND 芯片。



### 4.3. DRAM 厂商

#### 4.3.1. 领域图谱



图 25 DRAM 厂商领域图谱（中国信通院整理）

#### 4.3.2. 发展现状

当企业级 SSD 的大量数据在主控中无法处理时, DRAM 可以暂时存放数据, 从而保证数据处理效率和传输流畅性, 多用于高端芯片。DRAM 厂商总体呈现三大巨头寡头垄断, 其他企业加快发展的局面。TrendForce 数据显示, 2021 年上半年, 全球 DRAM 产值超过 433 亿美元, 各厂商持续增长, 如图 28 所示, 预计 2022 年, DRAM 产值将突破 1220 亿美元, 呈现供不应求的发展局势。三星、SK 海力士、美光三大巨头所占份额接近 95%, 出货量和销售均价同步增长。2021 年上半年, 三大 DRAM 厂商销售收入的季度增长率均超过 20%, 头部厂商三星的增长率甚至超过 30%, 如图 29 所示, 预计 2022 年, 在市场继续上扬的总体趋势下, 三大 DRAM 厂商会继续保持量价齐升的走势。

南亚科、华邦、力积电三家来自中国台湾厂商加快增长, 收入水平大幅增长, 但产能困境日益凸显。由于报价持续高涨, 2021 年上半年, 三家台厂增涨迅猛, 南亚科和华邦的季增长率超过 30%, 力积电增长率也解决 20%, 但由于产能不足, 且新工厂尚未投产, 供给能力逐渐收紧。长鑫存储作为国内唯一量产自研 DRAM 芯片的产线和公司, 正在持续增加研发投入, 提升产能水平, 吸引企业投资, 预计未来, 长鑫存储在全球的市场占有率将达到 1.5%到 2%。

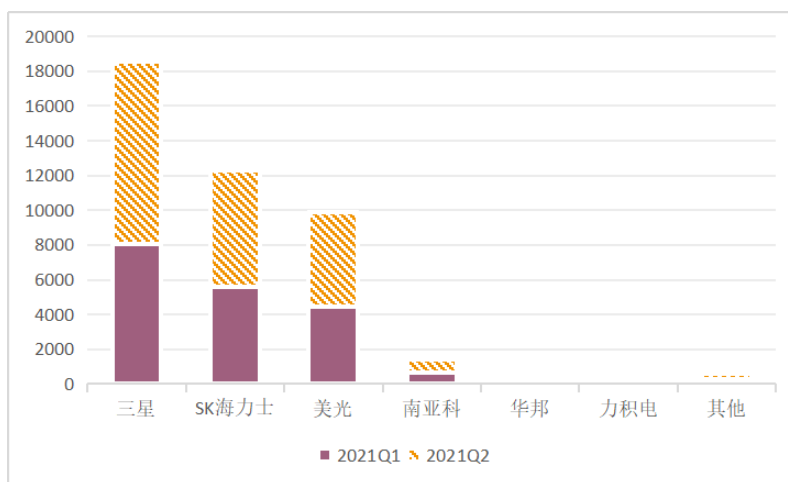


图 26 2021 年上半年 DRAM 厂商产值

数据来源: TrendForce

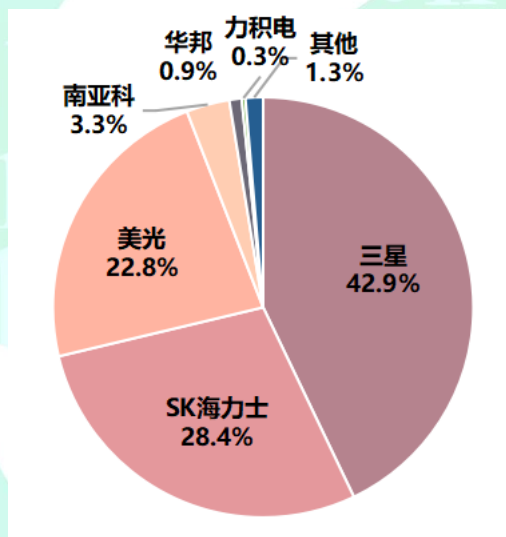


图 27 2021 年上半年 DRAM 厂商市场份额

数据来源: TrendForce

#### 4.3.3. 发展趋势

DRAM 厂商加快布局 EUV 光刻机，头部 DRAM 企业积极推进相关技术研发，提升先进制程芯片的生产制造能力。EUV 光刻机是芯片制造的关键环节，特别是 7nm 以下的高端芯片，EUV 光刻机发挥着不可替代的关键作用。SK 海力士采用 EUV 技术的第四代 10nm DRAM 已经正式量产，同时，为增加芯片产能，SK 海力士斥巨资打造新产线 M16，主要用于第四代 10nm 制程的 DRAM 记忆体生产；美光与 ASML 展开 EUV 光刻机采购谈判工作，预计在 2024 年开始采用 EUV 技术生产存储器；三星在 2020 年就已采用 EUV 技术量产 1z 纳米制程

DRAM，在 2021 年使用 EUV 设备量产 1 α nm DRAM。EUV 技术能够有效提升生产效率，缩小 DRAM 节点的设计尺寸，ASML 预计，2022 年全新的 EUV 设备出货量将达到 55 台。

DDR5 凭借高速率、低功耗的性能优势，已经在头部 DRAM 厂商的主流产品上实现应用，并逐渐成为 DRAM 产业的热门技术。最新的 DDR5 SDRAM 标准 (JESD79-5A) 能够满足云和企业数据中心应用需求，极大提高企业级 SSD 性能和能效。三星推出基于 HKMG (High-K Metal Gate) 工艺的 DDR5 内存模组，其容量可达到 512GB，在降低 13% 功耗的同时，将速度提升至 DDR4 的两倍左右，能够满足超级计算机、人工智能 (AI) 和机器学习等超高算力需求；SK 海力士开始量产基于 EUV 技术的 1 α nm 工艺 DDR5 内存芯片，核心容量达到 3GB，单个内存条容量可以达到 48GB；美光面向服务器 RDIMM 内存，推出了基于 1znm 技术的 DDR5 内存样品；AMD 预计在 2022 年，推出基于 Zen4 架构的 DDR5。国内厂商也在 DDR5 领域取得巨大突破，长鑫存储在生产时主要采用 19nm 制程工艺，计划在 2022 年推出 17 纳米工艺的 DDR5 和 LPDDR5 芯片。TechInsights 预计，在 2022 年到 2023 年，DDR5 产品将实现颠覆性进展，提供带宽更高、性能更强的产品，容量可以达到 32 GB，实现与 TSV DRAM 堆栈相结合的多裸片 DDR5 芯片。

## 4.4. 盘厂商

### 4.4.1. 领域图谱





图 30 盘厂商领域图谱（中国信通院整理）

#### 4.4.2. 发展现状

企业级 SSD 出货容量持续扩大,三星发售的企业级 SSD 容量位列市场首位,英特尔、SK 海力士、西部数据、美光、东芝出货容量优势明显。TrendFocus 数据显示,2019 年第三季度,三星和英特尔的企业级 SSD 出货容量平分秋色,在市场中所占份额均在 35%左右,而到 2021 年第三季度,全球企业级 SSD 出货总容量为 34.2EB,其中三星出货容量市场份额优势明显,占据总量的一半以上,企业级 SSD 容量超过 18EB,Intel 位居其后,所占比重约为 15.2%,SK 海力士、西部数据、铠侠、美光差距较小,预计 2022 年企业级 SSD 单盘容量将得到进一步增长。2021 年三季度各厂商企业级 SSD 出货容量情况如图 31 所示。

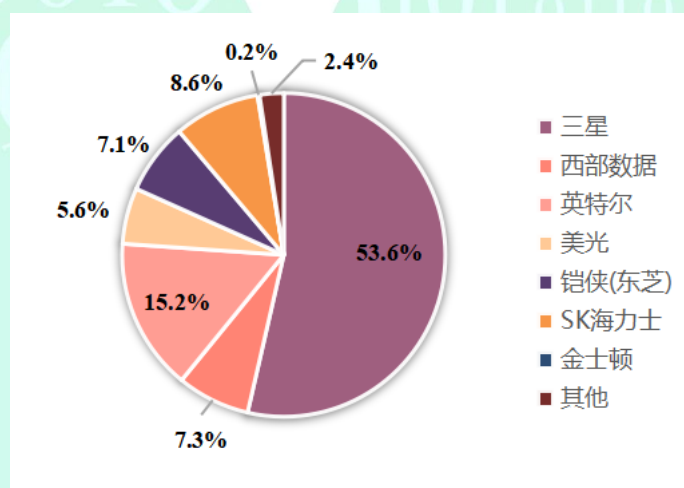


图 31 2021 年三季度企业级 SSD 厂商出货容量情况

数据来源: TrendFocus

#### 4.4.3. 发展趋势

SSD 能效持续优化,通过增强技术能力、调整接口规格、配置低耗部件等方式,SSD 盘厂商不断降低产品的功耗水平。三星企业级 SSD PM1743 采用增强的散热技术可降低碳排放量和能耗,E3.S 1/2TB 功耗分别为 20-25W、35-40W,E3.L 1TB 功耗最高仅为 40W,相比于前代产品 PM1733,能效提升超过 30%;SK 海力士发布的 PE8000 系列企业级 SSD,在支持 PCIe Gen4 的同时,保持较低的功耗水平;忆恒科技推出低功耗企业级 NVMe SSD PBlaze5 520 系列产品,

典型工作功耗仅为 9W, 待机功耗低至 4W, 并且还支持精确到每瓦的功耗控制, 满足企业对于绿色数据中心节能环保的需求。

**高性能企业级 SSD 能够为多场景、高性能、高技术应用提供更强的性能、可靠性和安全性, 是满足多元化需求的有效方式和关键趋势。**铠侠推出的企业级 PM7 SSD 系列主打高性能计算、人工智能、缓存层、金融交易与分析等用例, 读取性能较上一代提升约 20%, 容量水平可达到 30.72 TB; 西部数据推出了面向热存储的 Ultrastar DC SN840 高性能 SSD, 通过双接口、1 和 3 DW/D 耐久性以及 TCG 加密等设计, 适应 HPC、数据库、虚拟化、人工智能、5G 通讯、自动驾驶和机器学习等不同应用; 金士顿 DC1500M 采用高性能 Gen 3.0 x4 PCIe NVMe 设计, 可预测的随机读写性能和延迟, 适应高性能云服务、媒体采集、传输以及大数据应用要求。

## 5. 总结与展望

技术的不断进步推动企业级 SSD 持续迭代升级, 我国相关企业应当紧跟技术创新浪潮, 把握市场发展机遇, 重视关键核心技术攻关, 适应前沿新兴技术发展, 构建起国产化、规范化、高可靠、高性能发展体系。

**一是响应容量需求, 部署高密存储技术。**数据规模的持续攀升和算力需求的大幅增长, 企业级 SSD 需要满足海量数据计算要求, 推动高密度存储技术研发和应用成为提升存储容量的有效途径。热辅助磁记录、晶格介质存储、全息存储、DyCo5 材料存储等高密度关键存储技术持续研发创新, 以适应持续增长的数据存储容量和存储密度。热辅助磁记录技术在写入存储介质时, 通过聚焦光束对最小磁化区域进行加热, 解决少量热量对纳米级比特区域的干扰, 有效利用热量, 确保在纳米级比特单元在加热时仍保持磁化状态; 晶格介质存储通过光蚀刻微影的方式在晶格介质上划分出统一的网格磁性单元, 实现单比特占用空间少, 存储密度高; 全息存储通过消费光学和人工智能技术, 将数据存储为可重写的全息图, 借助商用高分辨率相机技术和深度学习技术, 将 HSD 设备负责性迁移到云端, 缩短读写时间, 提高实时访问效率; DyCo5 材料存储针对超高密度热辅助数据存储设备, 提出高效节能的解决方案, 使得数据写入时可实现快速磁化, 且材料耗

能更少、性能更优。

**二是重视节能降耗，建立低碳运行机制。**采用多措并举的方式，降低存储能耗，从存储材料、设计规范、操作使用等多个方面，建立高效、低功耗技术体系。使用相变材料、光敏铁电材料、氦气硬盘等存储材料能够降低大容量企业级 SSD 功耗。采用新型低功耗的钨锱碲合金相变材料，从物理上解决功耗问题，使得操作能耗与现有锑碲合金相比降低了 90%；光敏铁电材料通过输入脉冲光控制铁电材料电阻状态，不需要额外输入电流，显著降低写入储存产生的功耗；与常规的空气硬盘相比，氦气密封硬盘更适用于浸没式液冷环境，在扩大容量的同时，极大节约功耗。在架构设计时，采用低功耗数据存储和传输规范。在企业级 SSD 中应用 NVMe 协议，发挥功耗管理命令集和功耗管理模型优势；使用 DDR5 SDRAM 标准，满足云数据中心和企业数据中心应用需求，极大提高企业级 SSD 能效。对硬盘本身进行优化，在操作使用时，设置能耗控制功能，使得用户能够根据自身需求，选择是否使用节能功能，实现降耗减排。

**三是打破技术瓶颈，释放异构存储优势。**进一步提升企业级 SSD 核心技术整体水平，应当加强研发投入，攻关底层基础技术，同时充分释放异构存储优势。加大基础研究和核心技术研发投入，增强技术“硬实力”和生态“软实力”。增强主控芯片、固件算法、闪存介质的技术硬实力，提升底层硬件核心技术总体水平；企业应不断更新现有技术，构建起芯片、设备等相关产品的产学研联动生态。与此同时，采用多元异构的方式，充分释放存储芯片、算法、产品优势。通过新型超高速内外部互连技术、池化融合、重构技术等融合架构，集成不同架构单元，获取相应的架构优势，并使其能适配于复杂的、多元的数据形态与计算任务需求，适应企业级 SSD 规模扩大、计算复杂、需求多样的存储要求；提升异构存储组件适配能力，通过异构模型迁移适配和算法优化，支持进行算子定制化开发和模型适配，分析异构主控芯片性能，并针对分析结果进行适配，同时能够监控异构程序运行状态，保证事件告警零误报、零漏报。

**四是健全标准规范，构建测试评估体系。**强化标准支撑，推动高性能企业级 SSD 建设，建立完善的评估体系和机制，实现企业级 SSD 技术先进、算力赋能、绿色低碳、安全可靠。对芯片、架构、产品以及在数据中心、网络、ICT 设备、



数据服务等不同维度、不同领域应用进行测试评估,能够及时了解技术能力和应用效能,确保技术与产品高效适配;针对企业级 SSD 对于异构算力、边缘算力、智能算力、通用算力等多种算力的支持程度,进行规范化、体系化测试评估,确保 SSD 能够承载多样化算力;推动企业级 SSD 能耗评估,保障 IT 设备、数据中心等多应用场景实现低碳甚至零碳设计运行,践行低能耗、高能效建设发展要求;数据中心存储系统存储了企业大量的关键数据,必须保证这些数据安全可靠<sup>[24]</sup>,数据存储、数据传输、网络构建等运行与管理,以及 SSD 本身的研发生产和操作运行,均需要保证高度的安全可靠,要求对安全防护水平和技术防御能力进行科学化、标准化测试,保证全流程安全可靠。

## 参考文献

- [1] 吴家隐,李先绪.企业级固态硬盘最新进展及测试技术研究[J].计算机测量与控制,2020,28(06):38-41+46.
- [2] 郭亮. 面向云计算的企业级硬盘基准测试[J]. 电信网技术, 2016(10):1-4.
- [3] 艾瑞咨询. 中国企业级固态硬盘行业发展洞察[R]. 2021.
- [4] 国务院. 国务院印发《中国制造 2025》[EB/OL]. 2015[2022-03-13].  
[http://www.gov.cn/xinwen/2015-05/19/content\\_2864538.htm](http://www.gov.cn/xinwen/2015-05/19/content_2864538.htm).
- [5] 国务院. 中共中央办公厅 国务院办公厅印发《国家信息化发展战略纲要》[EB/OL]. 2016[2022-03-13]. [http://www.gov.cn/xinwen/2016-07/27/content\\_5095336.htm](http://www.gov.cn/xinwen/2016-07/27/content_5095336.htm).
- [6] 国务院. 国家发展改革委公布《战略性新兴产业重点产品和服务指导目录》2016 版 [EB/OL]. 2017[2022-03-13]. [http://www.gov.cn/xinwen/2017-02/04/content\\_5165379.htm](http://www.gov.cn/xinwen/2017-02/04/content_5165379.htm).
- [7] 国务院. 国务院关于印发新时期促进集成电路产业和 软件产业高质量发展若干政策的通知[EB/OL]. 2020[2022-03-13].  
[http://www.gov.cn/zhengce/content/2020-08/04/content\\_5532370.htm](http://www.gov.cn/zhengce/content/2020-08/04/content_5532370.htm).
- [8] 中央网络安全和信息化委员会. “十四五” 国家信息化规划 [EB/OL]. 2021[2022-03-13]. [http://www.cac.gov.cn/2021-12/27/c\\_1642205314518676.htm](http://www.cac.gov.cn/2021-12/27/c_1642205314518676.htm).

- [9] Hands J. EDSFF: Dynamic Family of Form Factors for Data Center SSDs [R]. OCP, 2020.
- [10] PCI-SIG. PCI Express 6.0 Specification[R]. 2022.
- [11] George Kimathi. PCIe 6.0 Spec Released With 64GT/s Per Lane Bandwidth[EB/OL]. 2022[2022-03-14]. <https://www.dignited.com/89481/pcie-6-0/>.
- [12] IEEE. IEEE 802 Nendica Report: Intelligent Lossless Data Center Networks[R]. New York:, 2021.
- [13] Onufryk P. NVMe 2.0 Specifications: The Next Generation of NVMe Technology[R]. SDC, 2021.
- [14] Alex Harrod. Arm's solution to the future needs of AI, security and specialized computing is v9[EB/OL]. 2021[2022-03-02]. <https://www.arm.com/company/news/2021/03/arms-answer-to-the-future-of-ai-armv9-architecture>.
- [15] ARM. Arm Storage Solution for SSD Controllers[R], 2020.
- [16] ARM. Arm Storage Solution for SSD Controllers[R]. 2020.
- [17] ARM. ARM Cortex-R Series Programmer's Guide[R]. England: ARM, 2014.
- [18] KIM S, KIM T. QLRU: NCQ-aware write buffer management algorithm for SSDs[J]. Electronics Letters, 2013, 49(17):1079-1080.
- [19] Mackay D J C, Neal R M. Near Shannon limit performance of low density parity check codes[J]. Electronics Letters, 2013, 32(6): 457-458.
- [20] 江藤良纯, 金子敏信. 纠错码及其应用[M]. 北京: 科学出版社, 2003: 183-184.
- [21] Symon A. QLC Everywhere! Even in Primary Storage![R]. IBM, 2020.
- [22] 邱奔, 王少鹏, 孙聪. 企业级存储技术与产业发展分析[J]. 中国电信业, 2021(S1):79-82.
- [23] 孙翠锋, 王祎晨. 闪存技术发展趋势及在数据中心的应用[J]. 互联网天地, 2020(12):57-60.
- [24] 郭亮, 吴美希, 王峰, 龚敏. 数据中心算力评估: 现状与机遇[J]. 信息通信技术与政策, 2021(02):79-86.



ODCC服务号



ODCC订阅号