Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

社论：

新型存储系统设计与应用

张广艳‡[1]，冯丹[2]，李克勤[3]，邵子立[4]，肖侬[5]，熊劲[6]，郑纬民[1]

[1]清华大学计算机科学与技术系，中国北京市，100084

[2]华中科技大学计算机科学与技术学院，中国武汉市，430074

[3]纽约州立大学计算机科学系，美国纽约市，12561

[4]香港中文大学计算机科学与工程系，中国香港特别行政区

[5]中山大学计算机学院，中国广州市，510000

[6]中国科学院计算技术研究所，中国北京市，100190

E-mail: gyzh@tsinghua.edu.cn; dfeng@hust.edu.cn; lik@newpaltz.edu; zilishao@cuhk.edu.hk; xiaon6@mail.sysu.edu.cn; xiongjin@ict.ac.cn; zwm-dcs@tsinghua.edu.cn

本文编译自 Zhang GY, Feng D, Li KQ, et al., 2023. Design and application of new storage systems. *Front Inform Technol Electron Eng*, 23(5):633-636. https://doi.org/10.1631/FITEE.2310000

存储系统是计算机的核心，在人工智能、大数据、云计算和物联网等新兴战略产业的可持续发展中起着重要作用。随着处理器和网络设备性能不断提高，存储软件栈成为限制数据密集型系统性能的主要因素。近年来，新型存储设备因其打破“内存墙”的能力而受到广泛关注。这些设备包括支持块寻址的闪存设备、支持字节寻址的非易失性存储器、存算一体化设备以及大容量光存储。构建高吞量、低延迟和高可靠性的大规模存储系统，需要对算法、软件设计和硬件的持续创新。这些创新可以应对大规模、高性能复杂结构系统构建中存在的挑战，还可以增加相关系统的构建和应用经验，加快大数据处理系统的开发速度。

研究人员一直致力于解决“内存墙”问题，并改进相关软硬件生态系统，从而在新型存储系统设计和应用方面取得很大进展，包括但不限于以下方面：

1. 不断推出和优化新型存储设备，例如开放通道固态硬盘、可字节寻址的非易失性存储器以及存算一体化设备。此外，陆续推出模拟器、仿真器和软件定义设备开发平台，促进了新型存储设备的设计和优化。

2. 现有存储软件系统最初为硬盘或传统固态硬盘而设计，不能充分发挥新型存储设备性能潜力。针对新型存储设备，设计了许多新型文件系统、存储管理软件、非关系型数据库以及关键组件。

3. 利用新型存储设备加速应用求解（如组合优化问题）和提升传统存储系统（如基于机械硬盘的纠删码存储）性能。

在此背景下，《信息与电子工程前沿（英文）》期刊组织了本期“新型存储系统设计与应用”专题。专题涵盖针对新型存储设备的辅助设计工具、各种存储软件、方法与相关应用，以及对新型存储系统前沿进展和未来研究方向的综述。经严格评审，选入 7 篇论文，包括 1 篇综述和 6 篇研究。

张广艳等从 5 个关键指标——吞吐量、延迟、寿命、性能隔离和资源利用率——对开放通道固

‡ 通讯作者

ORCID: 张广艳，https://orcid.org/0000-0002-3480-5902

态硬盘的设计和应用进行了全面综述。首先从物理布局、闪存转换层性质以及接口设计等方面介绍了开放通道固态硬盘，指出其性能优势和进一步提升性能的机会。然后，详细讨论了发掘开放通道固态硬盘性能的方法，包括设计接口、协同设计闪存转换层、利用内部并行性以及优化 I/O 调度和垃圾回收等。同时，讨论了将这一领域的理论研究成果应用到实际部署时面临的挑战。此外，展望了开放通道固态硬盘的发展潜能。

尽管市场上固态硬盘的特性在迅速发展，但由于缺乏真实且可扩展的固态硬盘开发平台，目前对闪存固件的研究主要基于模拟仿真。邵子力等提出一种软件定义的固态硬盘开发平台 SoftSSD，用于快速设计闪存固件原型。SoftSSD 的核心是一个具有事件驱动编程模型的新框架。可以通过编程模型来部署新的闪存转换层算法，并将其直接集成到全功能闪存固件中。SoftSSD 已在实际硬件上实现，并在真实应用场景中进行测量。实验表明，SoftSSD 可以取得良好性能、可观察性和可扩展性。SoftSSD 的开源代码已发布。

持久化内存和智能网卡这类新硬件的出现，给文件系统设计带来新机遇。然而，如何利用持久化内存和智能网卡的特性仍是一项挑战。杨倚天和陆游游设计并实现了一个名为 NICFS 的本地文件系统，该系统利用持久化内存的高吞吐量和字节寻址能力以及智能网卡的处理能力来改善文件系统性能，并减少主机 CPU 的使用。作者通过一系列实验验证了系统的性能、可扩展性和设计的每个部分的有效性。

持久化内存文件系统通过利用持久化内存的非易失性、字节可寻址性和与动态随机存取存储器相近的性能来实现高性能。然而，持久化内存文件系统存在写入持久性有限的问题。现有持久化内存文件系统中的空间管理策略会导致严重的磨损不均衡问题，迅速损坏底层持久化内存。张润宇、刘铎等提出一种高效的磨损均衡感知多粒度分配器 WMAlloc。此外，提出一种基于位图的多堆树（BMT），通过避免递归分割和低效的堆搜索来优化 WMAlloc，称作 WMAlloc-BMT，为底层持久化内存提升磨损均衡性的同时显著降低空间管理开销。文中通过大量实验验证了 WMAlloc 和 WMAlloc-BMT 的有效性。

可扩展哈希是管理大规模数据和提高存储系统效率的有效方法。蔡涛等设计了一种用于非易失性存储器的高并发可扩展哈希 NEHASH，它使用具有懒惰扩展的多层哈希目录来提高哈希目录管理的并发度和效率。该研究优化了哈希目录和哈希桶的管理策略，并将其分布在动态随机存取存储器和非易失性存储器之间。相比现有可扩展哈希方案，NEHASH 在多线程环境中实现了更高读写吞吐率。

纠删码具有更高存储效率，但与副本相比，其更新开销和修复成本也更高。此外，并发更新会使纠删码应用面临一致性和可靠性挑战。屠要峰、韩银俊等提出一种数据更新与编码解耦（DDUC）的纠删码存储系统，该系统使用持久化内存实现轻量级日志机制，并将数据更新和纠删码编码过程解耦。此外，提出一种副本和校验块相结合的数据放置策略，解决了由并发更新引起的数据可靠性降低问题，同时通过在校验和节点上保存临时冗余数据块来确保高并发性能。

组合优化问题重要且常见，但许多组合优化问题是 NP 完全的。混沌模拟退火算法有效解决了组合优化问题。然而，一般计算平台无法有效执行该算法。孙广宇等提出一种软硬件协同优化方案。首先，对算法实现作了修改，使其在保持高效率的同时对硬件更加友好。然后，设计了一种名为 COPPER 的硬件架构，使用忆阻器进行存内计算。COPPER 可有效运行混沌模拟退火算法，并显著提高计算速度和能效。

总体而言，本专题涵盖了许多与新型存储系统设计和应用相关的最新研究课题，包括新型存储设备和软件定义的设备开发平台、为新存储设备设计的文件系统、文件系统中的存储分配器、非易失存储器的可扩展散列以及新设备的应用，相信对新型存储系统及相关领域感兴趣的人员能够从中受益。

最后，我们感谢所有作者对本专题的支持和贡献。特别感谢所有评阅人对专题投稿的宝贵意见。

张广艳，专题执行主编，清华大学计算机系副教授、博士生导师。分别于 2000 和 2003 年在吉林大学获学士和硕士学位，于 2008 年在清华大学获博士学位。ACM 会员，CCF 杰出会员。研究方向包括大数据计算、网络存储和分布式系统。

冯丹，教授，博士生导师，华中科技大学计算机科学与技术学院院长。分别于 1991、1994 和 1997 年在华中科大获学士、硕士和博士学位。在 IEEE TC、IEEE TPDS、IEEE TCAD、ACM-TOS、FAST、USENIX ATC、ISCA、EuroSys、HPDC、SC、ICS、DAC、DATE 等主要期刊和会议上发表论文 200 余篇。曾任 IEEE TC、IEEE TPDS 等期刊审稿人，ATC 2023、FAST 2022、SC 2011 & 2013、MSST 2012 & 2015、SRDS 2020 等国际会议程序委员会成员。IEEE、ACM 会员。主要研究方向包括计算机体系结构、非易失性存储器技术、分布式和并行文件系统以及海量存储系统。

李克勤，纽约州立大学计算机科学杰出教授、湖南大学国家特聘教授，并行和分布式计算领域全球最具影响力的 5 位科学家之一（排名源自 Scopus 引文数据库综合指标）。亲撰、合著期刊文章、会议论文以及专著 900 余篇/部，并多次获得最佳论文奖。AAAS、IEEE、AAIA 会士，欧洲科学院院士。主要研究方向包括云计算、雾和移动边缘计算、节能计算和通信、嵌入式和信息物理系统、异构计算系统、大数据计算和高性能计算。

邵子立，香港中文大学计算机科学与工程系教授。1995 年于中国电子科技大学获学士学位，后分别于 2003 和 2005 年在美国德州大学达拉斯分校获硕士和博士学位。主要研究方向为大数据和存储系统、嵌入式软件和系统以及相关工业应用。

肖侬，中山大学计算机学院教授。分别于 1990 和 1996 年在国防科大获学士和博士学位。主要研究方向包括网络并行计算、大规模存储系统和计算机体系结构。

熊劲，中国科学院计算技术研究所教授。2006 年于中国科学院大学获博士学位。ACM、IEEE 会员，CCF 高级会员。主要研究方向包括面向云计算和云计算应用的数据中心存储系统。

郑纬民，专题主编，中国工程院院士，清华大学计算机系教授。1982 年于清华大学获硕士学位。研究方向涵盖分布式计算、编译器技术和网络存储。