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

Editorial:

Design and application of new storage systems

Guangyan ZHANG[‡1], Dan FENG[2], Keqin LI[3], Zili SHAO[4], Nong XIAO[5], Jin XIONG[6], Weimin ZHENG[1]

[1]*Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*
[2]*School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China*
[3]*Department of Computer Science, State University of New York, NY 12561, USA*
[4]*Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China*
[5]*School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510000, China*
[6]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China*
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

https://doi.org/10.1631/FITEE.2310000

A storage system is the core of a computer, and plays an important role in the sustainable development of emerging strategic industries, such as artificial intelligence, big data, cloud computing, and the Internet of Things. Storage stack access is a major factor restricting the performance of data-intensive systems because of the increasing performance of processors and network devices. Recently, new storage devices have attracted wide attention due to their ability to break the "memory wall." These devices include block-addressable flash memory, byte-addressable nonvolatile random access memory (NVRAM), in-memory computing devices, and large-capacity optical storage. Continuous innovation in algorithms, software designs, and hardware is necessary to build large-scale storage systems with high throughput, low access latency, and high data reliability. It can address challenges in building larger-scale and higher-performance systems with more complex structures. Further, it can boost the experience in building and applying relevant systems and accelerate developing the big data processing.

‡ Corresponding author
ORCID: Guangyan ZHANG, https://orcid.org/0000-0002-3480-5902

Researchers have been trying to solve the "memory wall" problem and enhance the corresponding software and hardware ecosystem; consequently, much progress has been made in the design and application of new storage systems, including but not limited to the following:

1. New storage devices continue to be designed and optimized, e.g., open-channel solid-state drives (SSDs), byte-addressable NVRAM, and in-memory computing devices. In addition, some simulators, emulators, and software-defined device development platforms are presented to promote the design and optimization of new storage devices.

2. Existing storage software systems, initially designed for hard drive disks or traditional SSDs, cannot use the performance potential of emerging storage devices completely. New file systems, storage management software, not only structured query language (no-SQL) databases, and key components are designed for emerging storage devices.

3. Studies have employed new storage devices to accelerate application solving (e.g., combinatorial optimization problem) and boost the performance of traditional storage systems (e.g., hard disk drive (HDD) based erasure-coded storage).

In this context, the journal *Frontiers of Information Technology & Electronic Engineering* has organized a special feature on the design and application of new storage systems. This special feature covers assistant design tools, various storage software, approaches, and related applications for new storage devices. In addition, this feature intends to provide a review of advances and future research directions in new storage systems. Seven papers have been selected for this feature after a rigorous review process, including one review article and six research articles.

Guangyan ZHANG et al. presented a comprehensive literature review of the design and application of open-channel SSDs from five key metrics: throughput, latency, lifetime, performance isolation, and resource use. The study first introduced the open-channel SSDs from the aspects of physical layout, properties of the flash translation layer, and design of interfaces. It explained the performance advantages and further optimization opportunities in designing and applying open-channel SSDs. Then, the methodologies were discussed in detail to leverage the performance benefits of open-channel SSDs in designing interfaces, co-designing flash translation layers, exploiting internal parallelism, and optimizing input/output (I/O) scheduling and garbage collection. The paper discussed the challenges in this area to bridge the gap between theoretical study and practical implementation. Further, the study explored potential future development to demonstrate the benefits of open-channel SSDs.

Despite the rapidly developing SSD features in the market, research on flash firmware remained mostly simulation-based due to the lack of a realistic and extensible SSD development platform. Zili SHAO et al. proposed SoftSSD, a software-defined SSD development platform for rapid flash firmware prototyping. The core of SoftSSD is a novel framework with an event-driven programming model. New flash translation layer (FTL) algorithms can be applied using the programming model and integrated into full-featured flash firmware directly. SoftSSD has been implemented with real hardware and evaluated with real application workloads. Experiments revealed that SoftSSD can achieve good performance, observability, and extensibility. SoftSSD has been open-sourced for public access.

The emergence of new hardware, e.g., persistent memory (PM) and smart network interface (SmartNIC), has brought new opportunities to file system design. However, using the features of PM and SmartNIC is challenging. Yitian YANG and Youyou LU designed and implemented a local file system called NICFS that applied the high throughput and byte addressability of PM and the processing power of SmartNIC to improve file system performance and reduce host CPU use. A series of experiments verified the system performance, scalability, and effectiveness of each part of the design.

PM file systems achieve high performance by exploiting the advanced features of PMs, including non-volatility, byte addressability, and dynamic random access memory (DRAM) like performance. However, these PMs suffer from limited write endurance. Existing space management strategies in PM file systems induce a severely unbalanced wear problem, quickly damaging the underlying PMs. Duo LIU et al. proposed an efficient wear-leveling-aware multigrained allocator called WMAlloc. Moreover, a bitmap-based multi-heap tree (BMT) was proposed to enhance WMAlloc, by avoiding the recursive split and inefficient heap searches. This significantly reduced the overhead of space management while achieving better wear-leveling of underlying PMs. The results from extensive experiments further validated the effectiveness of WMAlloc.

Extendible hashing is an effective way to manage large-scale data and improve the efficiency of storage systems. Tao CAI et al. designed NEHASH, a high-concurrency extendible hashing for non-volatile memory (NVM), which uses a multilevel hash directory with lazy expansion to improve the concurrency and efficiency of extendible hashing. The study optimized the management strategy of hash directories and buckets and distributed them between DRAM and NVM. NEHASH achieved higher read and write throughput in a multithreaded environment than the existing extendible hashing schemes.

Erasure coding (EC) has better storage efficiency but higher update overhead and repair costs than replication. In addition, concurrent updates produce consistency and reliability challenges in EC applications. Yaofeng TU et al. introduced an erasure-coded storage system called decoupled data updating and coding (DDUC), which uses PM to implement a lightweight logging mechanism and decouples data updating and EC encoding. Further, a data placement policy was proposed that combines replication and parity blocks. This addressed the data reliability reduction caused by concurrent updates while ensuring high concurrency by saving temporary redundant blocks of data at the checksum node.

Combinatorial optimization problems are critical and common but are NP-hard and difficult to solve. The chaotic simulated annealing algorithm effectively solves the combinatorial optimization problems. However, general computing platforms cannot execute it efficiently. Guangyu SUN et al. proposed a software–hardware co-optimization scheme. First, the algorithm implementation was modified to be more hardware-friendly while maintaining effectiveness. Then, a hardware architecture called COPPER was designed for in-memory computing using the memristor. COPPER can run the chaotic simulated annealing algorithm efficiently and significantly improve the speed and energy efficiency.

Overall, a broad spectrum of current research topics relevant to the design and application of new storage systems is covered in this special feature. These included new types of storage devices and software-defined device development platforms, file systems designed for new storage devices, storage allocators in file systems, extendible hashing for NVM, and applications for new devices. This collection of diverse but interconnected topics may benefit those interested in new storage systems or related areas.

Finally, we thank all the authors for their support and valuable contributions to this special feature. We are especially grateful to all the reviewers for their insightful comments and helpful suggestions to all the submissions.

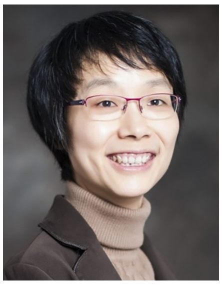
Guangyan ZHANG is an associate professor and PhD advisor in the Department of Computer Science and Technology at Tsinghua University. He received his BS and MS degrees in computer science from Jilin University, Changchun, China, in 2000 and 2003, respectively, and his PhD degree in computer science and technology from Tsinghua University, Beijing, China, in 2008. He is a professional member of ACM and a distinguished member of CCF. His research interests include big data computing, network storage, and distributed systems.

Dan FENG is a professor and dean of the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST), Wuhan, China. She received her BE, ME, and PhD degrees in computer science and technology from HUST in 1991, 1994, and 1997, respectively. She has more than 200 publications in major journals and international conferences, including *IEEE TC*, *IEEE TPDS*, *IEEE TCAD*, *ACM-TOS*, FAST, USENIX ATC, ISCA, EuroSys, HPDC, SC, ICS, DAC, and DATE. She has served as a reviewer of multiple journals, including *IEEE TC* and *IEEE TPDS*, and in program committees of multiple international conferences, including ATC 2023, FAST 2022, SC 2011 & 2013, MSST 2012 & 2015, and SRDS 2020. She is a member of IEEE and ACM. Her research interests include computer architecture, non-volatile memory technology, distributed and parallel file systems, and massive storage systems.

Keqin LI is a SUNY Distinguished Professor of Computer Science at the State University of New York. He is also a National Distinguished Professor at Hunan University, China. He has authored or co-authored over 900 journal articles, book chapters, and refereed conference papers. He has received several best paper awards. He is among the world's top five most influential scientists in parallel and distributed computing in terms of single-year and career-long impacts based on a composite indicator of the Scopus citation database. He is a fellow of AAAS, IEEE, and AAIA, and a member of Academia Europe (academician of the Academy of Europe). His research interests include cloud computing, fog and mobile edge computing, energy-efficient computing and communication, embedded and cyber-physical systems, heterogeneous computing systems, big data computing, and high-performance computing.

Zili SHAO is a professor in the Department of Computer Science and Engineering at The Chinese University of Hong Kong, Hong Kong, China. He received his BE degree in electronic mechanics from the University of Electronic Science and Technology of China, Chengdu, China, in 1995, and his MS and PhD degrees from the Department of Computer Science, University of Texas at Dallas, Dallas, TX, USA, in 2003 and 2005, respectively. His research interests include big data and storage systems, embedded software and systems, and related industrial applications.

Nong XIAO is a professor at the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China. He received his BS and PhD degrees in computer science from the College of Computer, National University of Defense Technology (NUDT), Changsha, China, in 1990 and 1996, respectively. His research interests include network parallel computing, large-scale storage systems, and computer architecture.



Jin XIONG is a professor at the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China. She received her PhD degree from the University of Chinese Academy of Sciences, Beijing, China, in 2006. She is a member of ACM and IEEE and a senior member of CCF. Her research interests include storage and file systems.

Weimin ZHENG is a professor in the Department of Computer Science and Technology, Tsinghua University, Beijing, China. He received his MS degree from Tsinghua University in 1982. He is an academician of Chinese Academy of Engineering. His research covers distributed computing, compiler techniques, and network storage.