Zhiwang YU, Runyu ZHANG, Chaoshu YANG, Shun NIE, Duo LIU, 2023. An efficient wear-leveling-aware multi-grained allocator for persistent memory file systems. *Frontiers of Information Technology & Electronic Engineering*, 24(5): 688-702. https://doi.org/10.1631/FITEE.2200468

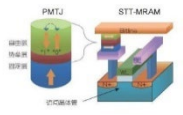
An efficient wear-leveling-aware multi-grained allocator for persistent memory file systems

Key words: File system; Persistent memory; Wear-leveling; Multi-grained allocator

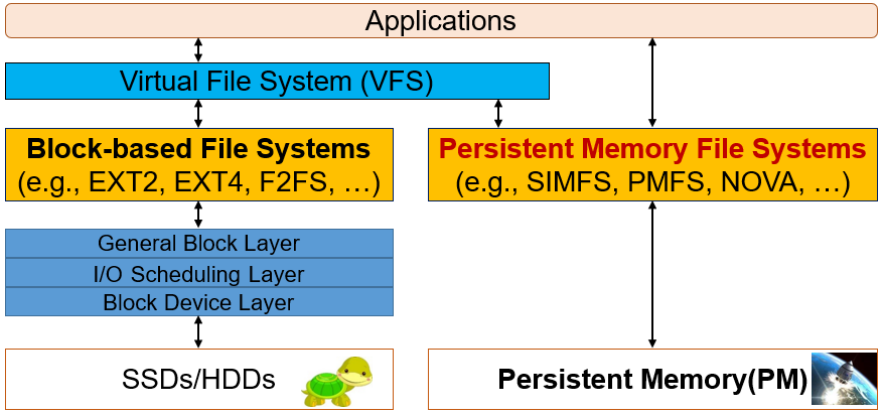
Corresponding authors: Runyu ZHANG, Chaoshu YANG
E-mail: zhangry@gzu.edu.cn, csyang@gzu.edu.cn
ORCID: https://orcid.org/0000-0003-3732-5098
https://orcid.org/0000-0002-0690-7370

Background and motivation

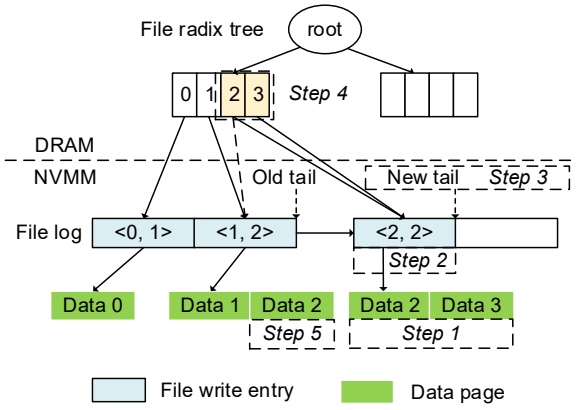
Persistent memory (PM), such as **PCM, 3D XPoint, STT-RAM, and FeRAM, has advanced features of non-volatility, byte addressability, and DRAM-like performance.**



Persistent memory file system



File access path optimization



File data organization and data consistency mechanism of NOVA

PM file systems, such as PMFS and NOVA, fully exploit the advantages of PM to **optimize file data organization, file access path, and data consistency mechanism**, making the file access throughput reach the GB/s level.

Background and motivation

Parameter	DRAM	PCM	STT-RAM	ReRAM	FeRAM	NAND Flash
Unit size (F^2)	60-100	4-12	6-50	4-10	6-40	4-6
Read latency	~10 ns	~20 ns	2-35 ns	~10 ns	40 ns	5-50 μs
Write latency	~10 ns	~100 ns	3-50 ns	~50 ns	65 ns	~500 μs
Endurance	$>10^{15}$	~10^8	~10^{15}	10^8-10^{11}	10^{14}-10^{15}	10^4-10^5

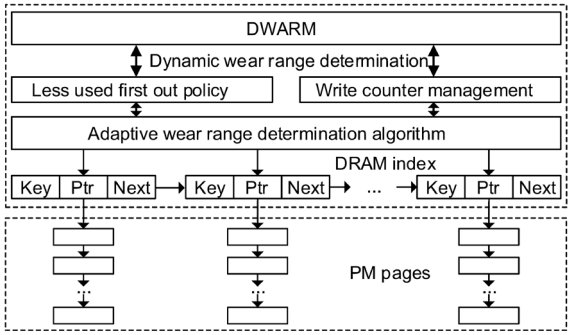
Large capacity PM

Representatives: PCM

- Low latency
- High storage density
- **Low write endurance**

However, PMs have the problem of limited write endurance. The existing space management strategies of PM file systems can easily cause "hot spots," which can damage the underlying PMs quickly.

Wear-leveling-aware allocators



Logical overview of DWARM

Wear-leveling-aware allocators of PM file systems, such as DWARM, can achieve wear-leveling of the underlying PMs to further improve their lifetime, **but may seriously degrade the performance of PM file systems.**

Method

WMAlloc algorithm

Multi-grained allocating heaps

Wear-balancing node migration

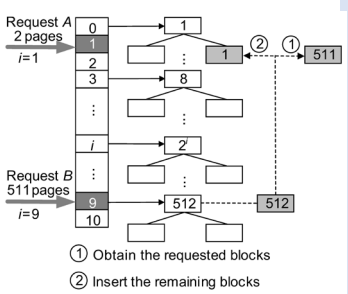
Wear-aware recycle forest

Obtain

Migrate

Insert

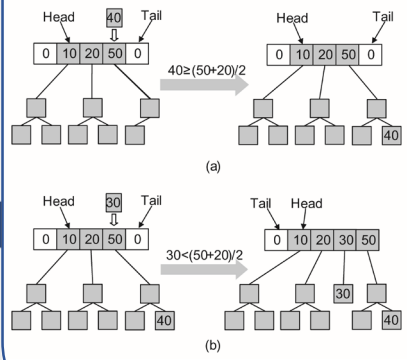
Min-heap trees



- Support multi-grained allocation
- Allocate the root node in an allocation as it has the minimum wear in the heap tree

Upload

Red-black trees



- Reclaim released blocks
- Fast block merging
- High-accuracy wear-leveling

Persistent memory

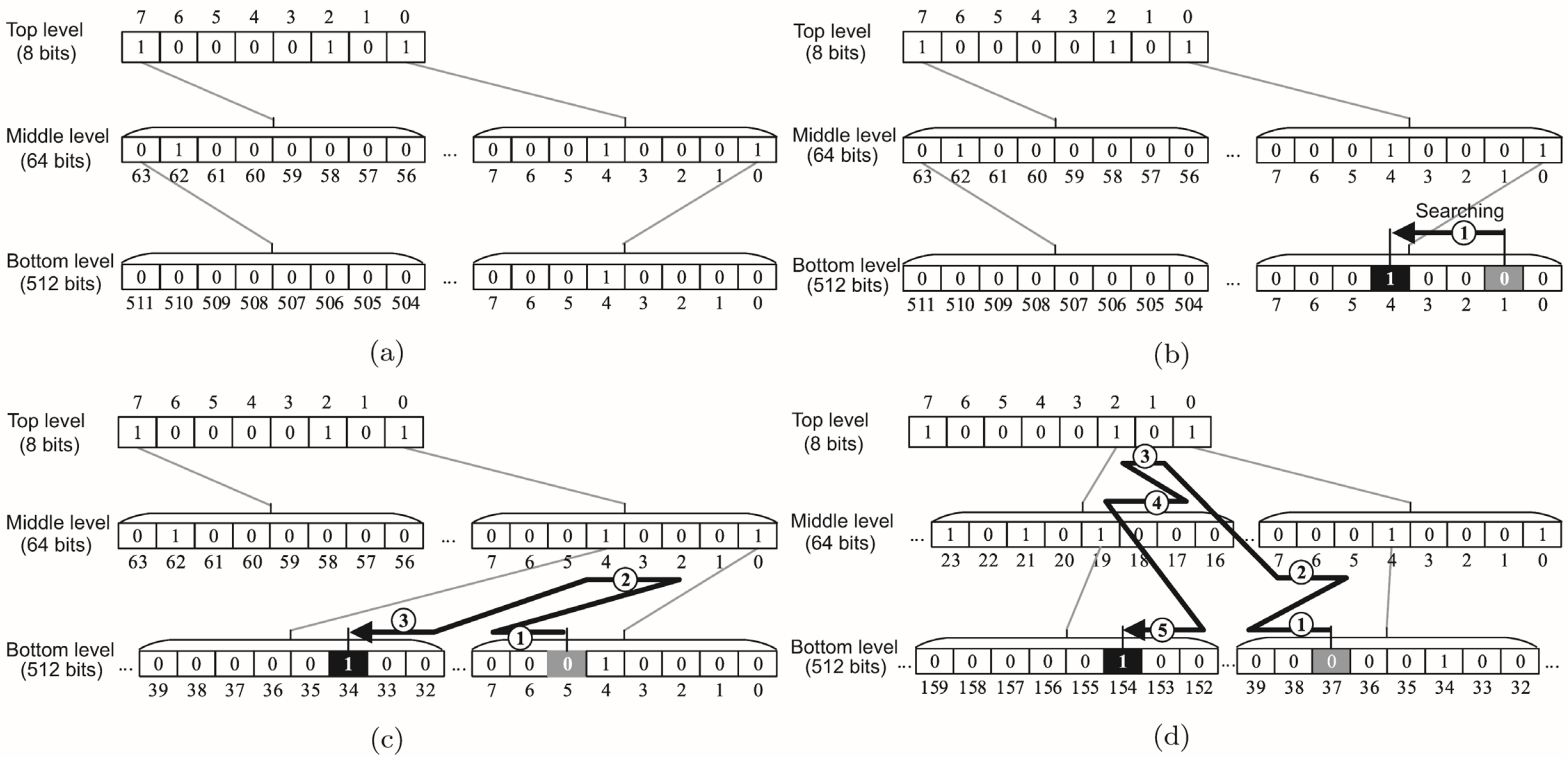
…

WMAlloc achieves high-accuracy wear-leveling of PMs while improving the performance of file systems compared with DWARM.

Method

WMAlloc still suffers from redundant overhead and heavy wear for two reasons:

- **The number of external fragmentations of WMAlloc will rapidly increase due to the inserting blocks' split.**
- **The average wear of each split sub-block needs to be recalculated.**



- **Design the Bitmap-based Multi-heap Tree (BMT) to accommodate the remaining sub-blocks and cut off the recursive split**
- **Propose the conditional wear inheritance mechanism to reduce the overhead of recalculating the average wear degree of split sub-blocks**

Major results: impact on lifetime of PM

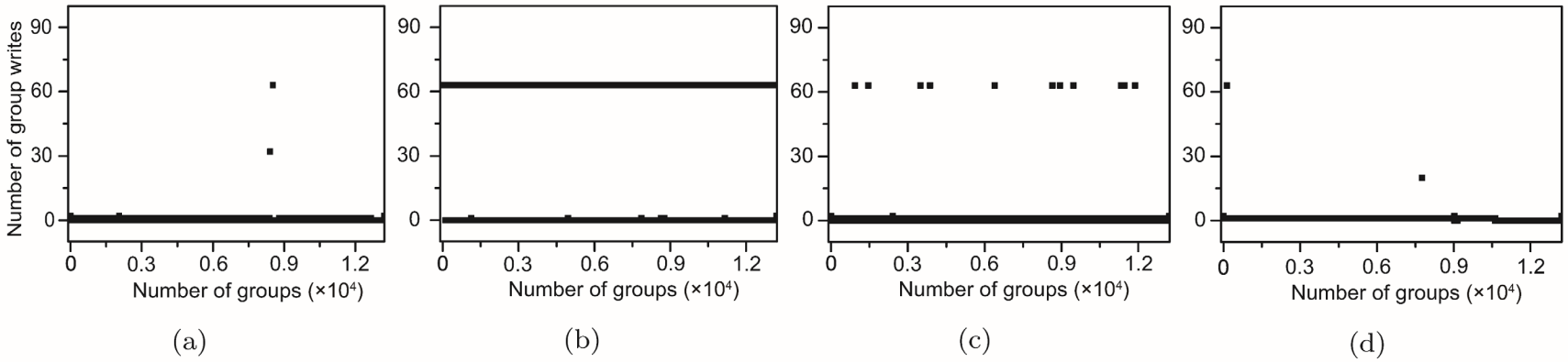


Fig. 6 The write distribution in fio: (a) NOVA; (b) DWARM; (c) WMAlloc; (d) WMAlloc-BMT

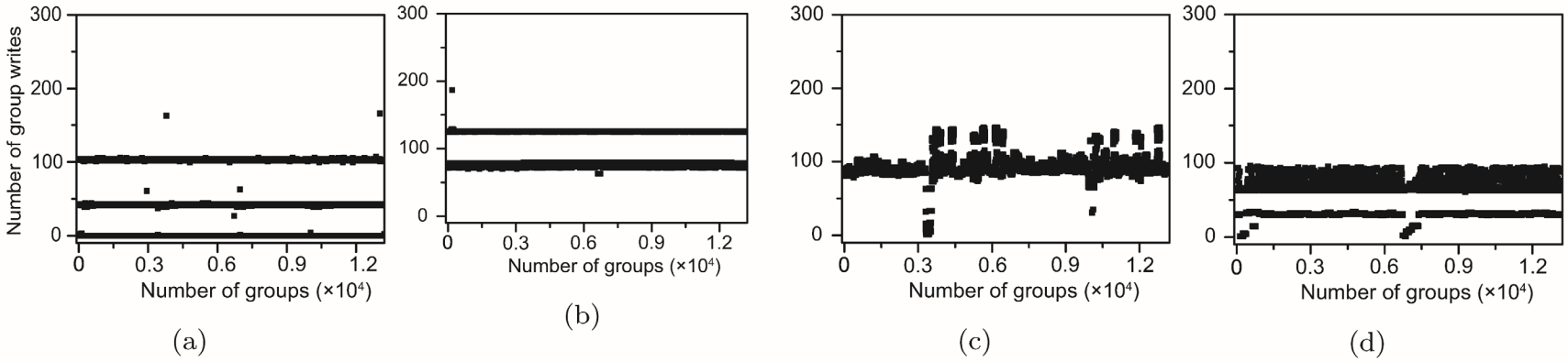


Fig. 7 The write distribution in createdelete-swing: (a) NOVA; (b) DWARM; (c) WMAlloc; (d) WMAlloc-BMT

Major results: impact on lifetime of PM

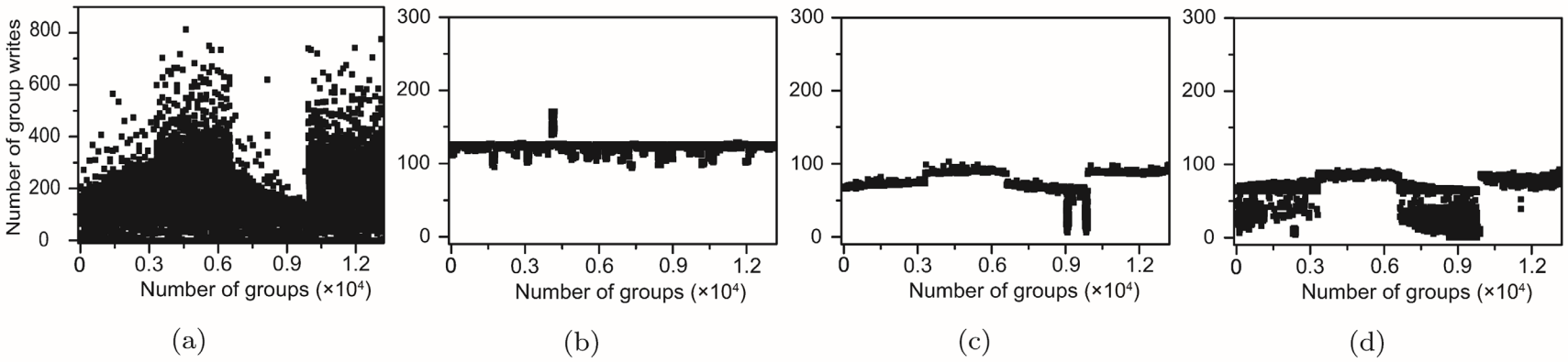


Fig. 8 The write distribution in fileserver: (a) NOVA; (b) DWARM; (c) WMAlloc; (d) WMAlloc-BMT

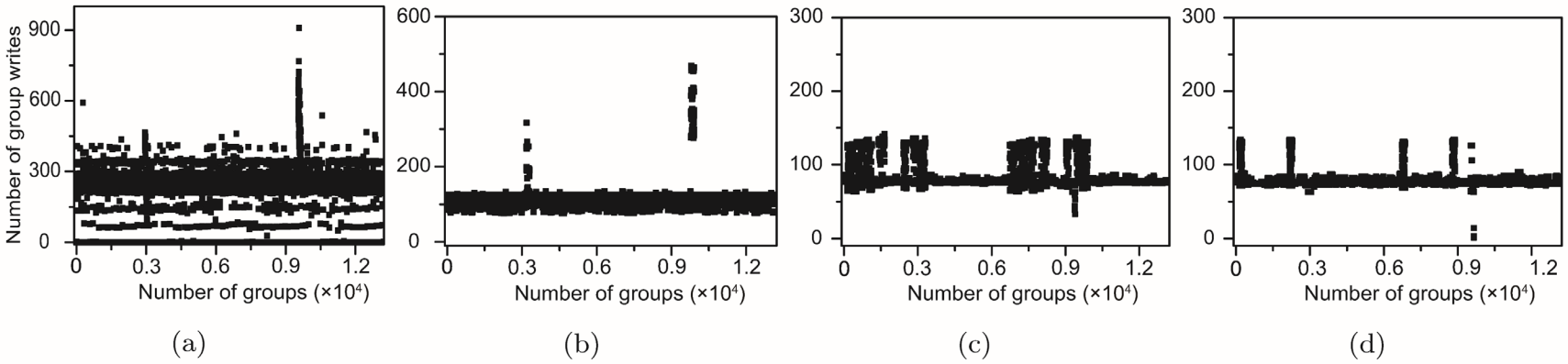


Fig. 9 The write distribution in postmark: (a) NOVA; (b) DWARM; (c) WMAlloc; (d) WMAlloc-BMT

Major results: overall performance

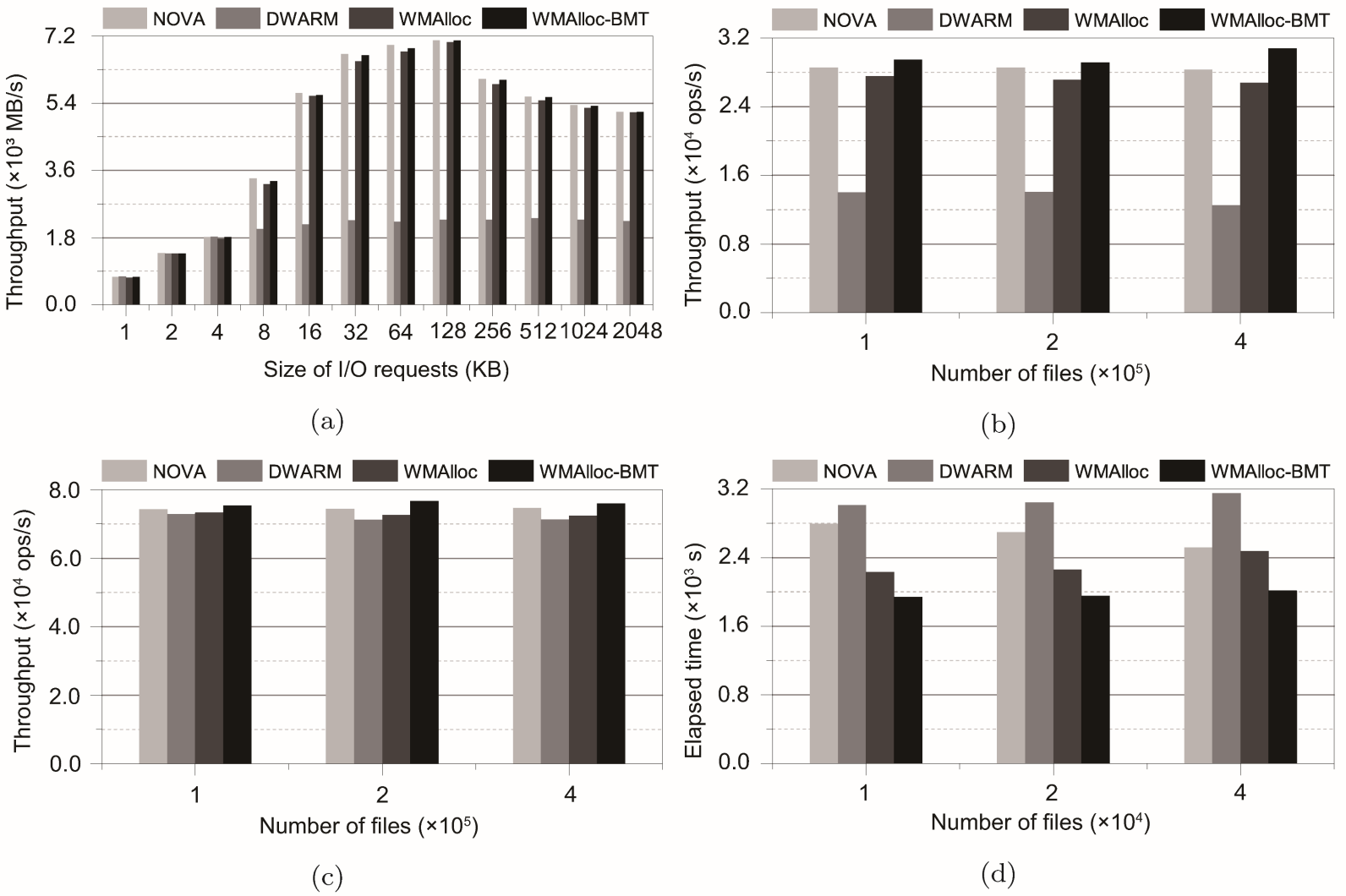


Fig. 10 The overall performance: (a) fio; (b) createdelete-swing; (c) fileserver; (d) postmark

Conclusions

Research contents

- We propose a Wear-leveling-aware Multi-grained Allocator (WMAlloc), which achieves wear-leveling of PM while improving the performance of PM file systems.
- We further propose the Bitmap-based Multi-heap Tree (BMT) to enhance the performance and wear-leveling of WMAlloc.

Experimental results

- Compared with the original NOVA and dynamic wear-aware range management (DWARM), which is the state-of-the-art wear-leveling-aware allocator of PM file systems, WMAlloc can, respectively, achieve 4.11× and 1.81× maximum write number reduction and 1.02× and 1.64× performance with four workloads on average.
- WMAlloc-BMT outperforms WMAlloc with 1.08× performance and achieves 1.17× maximum write number reduction with four workloads on average.