**REFERENCES**

[1] N. Agrawal, William J. Bolosky, John R. Douceur, and Jacob R. Lorch. A Five-Year Study of File-System metadata. In *FAST’07*, Feb. 2007.

[2] A. Anand, S. Sen, A. Krioukov, F. Popovici, A. Akella, Andrea C. Arpaci-Dusseau, Remzi H. Arpaci-Dusseau, and S. Banerjee. Avoiding File System Micromanagement with Range Writes. In *OSDI’08*, Dec. 2008.

[3] A. Batsakis, R. Burns, A. Kanevsky, J. Lentini, and T. Talpey. AWOL: An Adaptive Write Optimizations Layer. In *FAST’08*, Feb. 2008.

[4] P. Carns, K. Harms, W. Allcock, C. Bacon, S. Lang, R. Latham, and R. Ross. Understanding and Improving Computational Science Storage Access through Continuous Characterization. *ACM Transactions on* *Storage*, 7(3):1–26, 2011.

[5] F. Chen, T. Luo, and X. Zhang. CAFTL: A Content-Aware Flash Translation Layer Enhancing the Lifespan of Flash Memory based Solid State Drives. In *FAST’11*, pages 77–90, Feb. 2011.

[6] A. T. Clements, I. Ahmad, M. Vilayannur, and J. Li. Decentralized Deduplication in SAN Cluster File Systems. In *USENIX ATC’09*, Jun. 2009.

[7] L. Costa, S. Al-Kiswany, R. Lopes, and M. Ripeanu. Assessing Data Deduplication trade-offs from an Energy Perspective. In *ERSS’11*, Jul. 2011.

[8] A. El-Shimi, R. Kalach, A. Kumar, A. Oltean, J. Li, and S. Sengupta. Primary Data Deduplication - Large Scale Study and System Design. In *USENIX ATC’12*, Jun. 2012.

[9] FIU traces. http://iotta.snia.org/traces/390.

[10] D. Frey, A. Kermarrec, and K. Kloudas. Probabilistic Deduplication for Cluster-Based Storage Systems. In *SOCC’12*, Nov. 2012.

[11] M. Fu, D. Feng, Y. Hua, X. He, Z. Chen, W. Xia, F. Huang, and Q. Liu. Accelerating Restore and Garbage Collection in Deduplication-based Backup Systems via Exploiting Historical Information. In *USENIX’14*, Jun. 2014.

[12] Garth Gibson. Storage at Exascale: Some Thoughts from Panasas CTO. Interview. May 2011.

[13] B. S. Gill, M. Ko, B. Debnath, and W. Belluomini. STOW: A Spatially and Temporally Optimized Write Cache Algorithm. In *USENIX ATC’09*, Jun. 2009.

[14] A. Gupta, R. Pisolkar, B. Urgaonkar, and A. Sivasubramaniam. Leveraging Value Locality in Optimizing NAND Flash-based SSDs. In *FAST’11*, Feb. 2011.

[15] Y. Hu and Q. Yang. DCD - Disk Caching Disk: A New Approach for Boosting I/O Performance. In *ISCA’96*, May 1996.

[16] Y. Hua and X. Liu. Scheduling Heterogeneous Flows with Delay-aware Deduplication for Avionics Applications. *IEEE Transactions on Parallel* *and Distributed Systems*, 23(9):1790–1802, 2012.

[17] K. Jinand and E. L. Miller. The Effectiveness of Deduplication on Virtual Machine Disk Images. In *SYSTOR’09*, pages 1–12, May 2009.

[18] Stephanie Jones. Online De-duplication in a Log-Structured File System for Primary Storage. Technical Report UCSC-SSRC-11-03, University of California Santa Cruz. May 2011.

[19] S. Kiswany, M. Ripeanu, S. S. Vazhkudai, and A. Gharaibeh. STDCHK: A Checkpoint Storage System for Desktop Grid Computing. In *ICDCS’08*, Jun. 2008.

[20] R. Koller and R. Rangaswami. I/O Deduplication: Utilizing Content Similarity to Improve I/O Performance. In *FAST’10*, pages 1–14, Feb. 2010.

[21] C. Li, P. Shilane, F. Douglis, H. Shim, S. Smaldone, and G. Wallace. Nitro: A Capacity-Optimized SSD Cache for Primary Storage. In *USENIX’14*, Jun. 2014.

[22] M. Lillibridge, K. Eshghi, and D. Bhagwat. Improving Restore Speed for Backup Systems that Use Inline Chunk-Based Deduplication. In *FAST’13*, Feb. 2013.

[23] J. Lofstead, M. Polte, G. Gibson, S. Klasky, K. Schwan, R. Oldfield, M. Wolf, and Q. Liu. Six Degrees of Scientific Data: Reading Patterns for Extreme Scale Science IO. In *HPDC’11*, Jun. 2011.

[24] B. Mao, H. Jiang, S. Wu, Y. Fu, and L. Tian. SAR: SSD Assisted Restore Optimization for Deduplication-based Storage Systems in the Cloud. In *NAS’12*, Jun. 2012.

[25] B. Mao, H. Jiang, S. Wu, Y. Fu, and L. Tian. Read Performance Optimization for Deduplication-based Storage Systems in the Cloud. *ACM Transactions on Storage*, 10(2):1–22, 2014.

[26] N. Megiddo and D. Modha. Arc: A self-tuning, low overhead replacement cache. In *FAST’03*, Mar. 2003.

[27] D. Meister, J. Kaiser, A. Brinkmann, T. Cortes, M. Kuhn, and J. Kunkel. A Study on Data Deduplication in HPC Storage Systems. In *SC’12*, Nov. 2012.

[28] J. Menon. A Performance Comparison of RAID-5 and Log-Structured Arrays. In *HPDC’95*, pages 167–178, Aug. 1995.

[29] D. T. Meyer and W. J. Bolosky. A Study of Practical Deduplication. In *FAST’11*, Feb. 2011.

[30] Y. Oh, J. Choi, D. Lee, and Sam H. Noh. Caching less for better performance: Balancing cache size and update cost of flash memory cache in hybrid storage systems. In *FAST’12*, Feb. 2012.

[31] R. Patterson, G. Gibson, E. Ginting, D. Stodolsky, and J. Zelenka. Informed prefetching and caching. In *SOSP’95*, Dec. 1995.

[32] E. Pinheiro, W.-D. Weber, and L. A. Barroso. Failure Trends in a Large Disk Drive Population. In *FAST’07*, pages 17–29, Feb. 2007.

[33] E. Rozier and W. Sanders. A Framework for Efficient Evaluation of the Fault Tolerance of Deduplicated Storage Systems. In *DSN’12*, Jun. 2012.

[34] S. Savage and J. Wilkes. AFRAID: A Frequently Redundant Array of Independent Disks. In *USENIX ATC’96*, Jan. 1996.

[35] J. Schindler, S. Shete, and Keith A. Smith. Improving Throughput for Small Disk Requests with Proximal I/O. In *FAST’11*, Feb. 2011.

[36] B. Schroeder and G. Gibson. Disk Failures in the Real World: What Does an MTTF of 1,000,000 Hours Mean to You? In *FAST’07*, pages 1–16, Feb. 2007.

[37] K. Srinivasan, T. Bisson, G. Goodson, and K. Voruganti. iDedup: Latency-aware, Inline Data Deduplication for Primary Storage. In *FAST’12*, Feb. 2012.

[38] D. Stodolsky, G. Gibson, and M. Holland. Parity Logging Overcoming the Small Write Problem in Redundant Disk Arrays. In *ISCA’93*, pages 64–75, May 1993.

[39] J. Tucci. Cloud + Big Data = Massive Change, Massive Opportunity, Keynote Address at UW CSE 2011-12 Annual Industrial Affiliates Meeting. Oct 2010.

[40] V. Vasudevan, M. Kaminsky, and David G. Andersen. Using Vector Interfaces to Deliver Millions of IOPS from a Networked Key-value Storage Server. In *SOCC’12*, Oct. 2012.

[41] R. Villars, C. Olofson, and M. Eastwood. Big Data: What It Is and Why You Should Care, White Paper, IDC. 2011.

[42] S. Wu, H. Jiang, D. Feng, L. Tian, and B. Mao. WorkOut: I/O Workloa Outsourcing for Boosting the RAID Reconstruction Performance. In *FAST’09*, Feb. 2009.

[43] S. Wu, H. Jiang, and B. Mao. IDO: Intelligent Data Outsourcing with Improved RAID Reconstruction Performance in Large-Scale Data Centers. In *LISA’12*, Dec. 2012.

[44] W. Xia, H. Jiang, D. Feng, and Y. Hua. Similarity and Locality based Indexing for High Performance Data Deduplication. *IEEE Transactions* *on Computers, Accepted*, 2014.

[45] C. Zhang, X. Yu, A. Krishnamurthy, and Randolph Y. Wang. Configuring and Scheduling an Eager-Writing Disk Array for a Transaction Processing Workload. In *FAST’02*, Jan. 2002.

[46] X. Zhang, Z. Huo, J. Ma, and D. Meng. Exploiting Data Deduplication to Accelerate Live Virtual Machine Migration. In *Cluster’10*, pages 88–96, Sep. 2010.

[47] Y. Zhang and V. Prabhakaran. DADA: Duplication Aware Disk Array. In *FAST’11 (Poster)*, Feb. 2011.