7 RELATED WORK

A number of studies have been carried out recently regarding characterization of SPEC CPU2017 workloads, however, to the best of our knowledge, this paper presents the first systematic study of the memory behavior of the SPEC CPU2017 suite.

SPEC CPU2017 Characterization: Bucek et al. [8] present an overview of CPU2017 suite and discuss its reportable execution. Limaye and Adegbija [21] use hardware performance counter statistics to characterize SPEC CPU2017 applications with respect to several metrics such as instruction distribution, execution performance, branch and cache behaviors. They also utilize Principal Components Analysis [10] and hierarchical clustering to identify subsets of the suite. Similarly, Panda et al. [24] characterize the CPU2017 benchmarks using *perf*, and leverage statistical techniques to identify cross application redundancies and propose subsets of the entire suite, by classifying multiple benchmarks with similar behaviors into a single subset. Further, they also provide a detailed evaluation of the representativeness of the subsets. Amaral et al. [5] propose the Alberta Workloads for the SPEC CPU2017 benchmark suite hoping to improve the performance evaluation of techniques that rely on any type of learning, for example the formal Feedback-Directed Optimization (FDO). Additionally, in order to ameliorate large simulation times, Wu et al. [26] analyze the program behavior and consequently propose simulation points [12] for the suite.

Memory Characterization of Workloads: Jaleel [16] determined the memory system requirements of workloads from SPEC CPU2000 and CPU2006 using binary instrumentation. Henning [15] discussed the memory footprints of CPU2006 workloads, while Gove [11] analysed their working set sizes. Bienia et al. [7] present memory behavior of PARSEC benchmark suite. John et al. [18] discusses a taxonomy of workload characterization techniques.

8 CONCLUSION

In this paper, we provide the first, comprehensive characterization of the memory behavior of the SPEC CPU2017 benchmark suite. Our working set analysis shows that many workloads have a working set much higher than 32 MB (maximum cache size assumed in our experiments), implying the continued importance of cache hierarchies for benchmark performance. We also show that Rate benchmarks, both INT and FP, still have main memory consumption well below 900 MB, which was target memory footprint for CPU2006. Almost 90% of the workloads have main memory consumption below 5 GB, with the average across the suite being 1.82 GB. However, workloads have extremely varying peak memory bandwidth usage, with some benchmarks requiring as little as 0.2 MB/s, to others utilizing upto 2.3 GB/s.

In addition, our experiments have revealed some interesting results with respect to dynamic instruction counts and distributions. The average instruction count for SPEC CPU2017 workloads is 22.19 trillion, which is an order of magnitude higher than the SPEC CPU2006. In addition, we find that FP benchmarks typically have much higher compute requirements: on average, FP workloads carry out three times the number of arithmetic operations as compared to INT workloads.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers and shepherd for their useful comments and feedback. This work was supported in part by SERB grant (ECR/2017/000887), as well as Indian Institute of Technology Gandhinagar and Ashoka University, Sonipat.

REFERENCES

[1] [n. d.]. PS man-page. http://man7.org/linux/man-pages/man1/ps.1.html.

[2] [n. d.]. SPEC CPU2017 Base Compilation Flags. https://www.spec.org/cpu2017/Docs/runrules.html#rule_2.3.

[3] [n. d.]. SPEC CPU2017 Documentation. https://www.spec.org/cpu2017/Docs.

[4] Junwhan Ahn, Sungjoo Yoo, Onur Mutlu, and Kiyoung Choi. 2015. PIM-enabled instructions: a low-overhead, locality-aware processing-in-memory architecture. In *Computer Architecture (ISCA), 2015 ACM/IEEE 42nd Annual International Symposium on.* IEEE, 336–348.

[5] J. N. Amaral et al. 2018. The Alberta Workloads for the SPEC CPU 2017 Benchmark Suite. In *Proceedings of ISPASS.*

[6] Manu Awasthi et al. 2012. Managing data placement in memory systems with multiple memory controllers. *International Journal of Parallel Programming* 40, 1 (2012).

[7] Christian Bienia et al. 2008. The PARSEC benchmark suite: Characterization and architectural implications. In *Proceedings of PACT.*

[8] James Bucek, Klaus-Dieter Lange, et al. 2018. SPEC CPU2017: Next-Generation Compute Benchmark. In *Companion of the 2018 ACM/SPEC International Conference on Performance Engineering.* ACM, 41–42.

[9] Arnaldo Carvalho De Melo. 2010. The new linux perf tools. In *Slides from Linux Kongress*, Vol. 18.

[10] George H Dunteman. 1989. *Principal components analysis.* Number 69. Sage.

[11] Darryl Gove. 2007. CPU2006 working set size. *ACM SIGARCH Computer Architecture News* 35, 1 (2007), 90–96.

[12] Greg Hamerly, Erez Perelman, and Brad Calder. 2004. How to use simpoint to pick simulation points. *ACM SIGMETRICS Performance Evaluation Review* 31, 4 (2004), 25–30.

[13] Hasan Hassan et al. 2016. ChargeCache: Reducing DRAM latency by exploiting row access locality. In *Proceedings of HPCA.*

[14] John L. Henning. 2006. SPEC CPU2006 Benchmark Descriptions. *SIGARCH Comput. Archit. News* 34, 4 (2006).

[15] John L. Henning. 2007. SPEC CPU2006 Memory Footprint. *SIGARCH Comput. Archit. News* 35, 1 (March 2007), 84–89. https://doi.org/10.1145/1241601.1241618

[16] Aamer Jaleel. 2010. Memory characterization of workloads using instrumentation-driven simulation. *Web Copy: http://www. glue. umd. edu/ajaleel/workload* (2010).

[17] Youngbin Jin, Mustafa Shihab, and Myoungsoo Jung. 2014. Area, power, and latency considerations of STT-MRAM to substitute for main memory. In *Proc. ISCA.*

[18] Lizy Kurian John, Purnima Vasudevan, and Jyotsna Sabarinathan. 1999. Workload characterization: Motivation, goals and methodology. In *Workload Characterization: Methodology and Case Studies. Based on the First Workshop on Workload Characterization.* IEEE, 3–14.

[19] Emre Kültürsay et al. 2013. Evaluating STT-RAM as an energy-efficient main memory alternative. In *Proceedings of ISPASS.*

[20] Benjamin C Lee, Engin Ipek, Onur Mutlu, and Doug Burger. 2009. Architecting phase change memory as a scalable dram alternative. In *ACM SIGARCH Computer Architecture News*, Vol. 37. ACM, 2–13.

[21] A. Limaye and T. Adegbija. 2018. A Workload Characterization of the SPEC CPU2017 Benchmark Suite. In *Proceedings of ISPASS).*

[22] Jamie Liu et al. 2012. RAIDR: Retention-aware intelligent DRAM refresh. In *ACM SIGARCH Computer Architecture News*, Vol. 40.

[23] Chi-Keung Luk et al. 2005. Pin: Building Customized Program Analysis Tools with Dynamic Instrumentation. *SIGPLAN Not.* 40, 6 (June 2005).

[24] R. Panda et al. 2018. Wait of a Decade: Did SPEC CPU 2017 Broaden the Performance Horizon?. In *Proceedings of HPCA.*

[25] Aniruddha N. Udipi et al. 2010. Rethinking DRAM Design and Organization for Energy-constrained Multi-cores. In *Proceedings of ISCA.*

[26] Qinzhe Wu et al. 2018. Hot Regions in SPEC CPU2017. (2018).

[27] Xiaoxia Wu et al. 2009. Hybrid cache architecture with disparate memory technologies. In *ACM SIGARCH computer architecture news*, Vol. 37.

[28] Wangyuan Zhang and Tao Li. 2009. Exploring phase change memory and 3D die-stacking for power/thermal friendly, fast and durable memory architectures. In *Proceedings of PACT.*