



# Pooneh Compression: A Simple Last Level Cache Compression for CMPs

**Pooneh Safayenikoo**

Supervisor: Dr. Ismail Akturk

University of Missouri

YArch '19

# Outline



- Introduction
- Related Work
- Proposed Method
- Evaluation Methodology
- Future Work



- The main memory is the most power-hungry component in multi/many-core CMPs
  - Speed and bandwidth gap between processors and main memories
  - This gap can be handled by cache
- The demand for last level cache capacity keep increasing
  - Cache compression method can be a promising solution
- Compression challenges
  - (1) Compression/decompression overhead (latency and energy)
  - (2) The cache organization based on variable size data block



- FPC (Frequent Pattern Compression) [1], BDI (Base-Delta-Immediate) [2], LZ (Lempel-Ziv) [3], BPC (Bit Plane Compression) [4], C-Pack [5] and SC2 [6]
  - Complex hardware
  - Low compression ratio
  - High latency for decompression
- Dictionary Sharing [7] encodes a cache block with a dictionary
  - Distinct 4-byte chunks of a cache block
  - However, it cannot compress if there are more than 16 distinct values among 64 chunks

# Key Data Patterns



- Zero Values

0x00000000	0x00000000	0x00000000	0x00000000	0x00000000	0x00000000	0x00000000	0x00000000
------------	------------	------------	------------	------------	------------	------------	------------

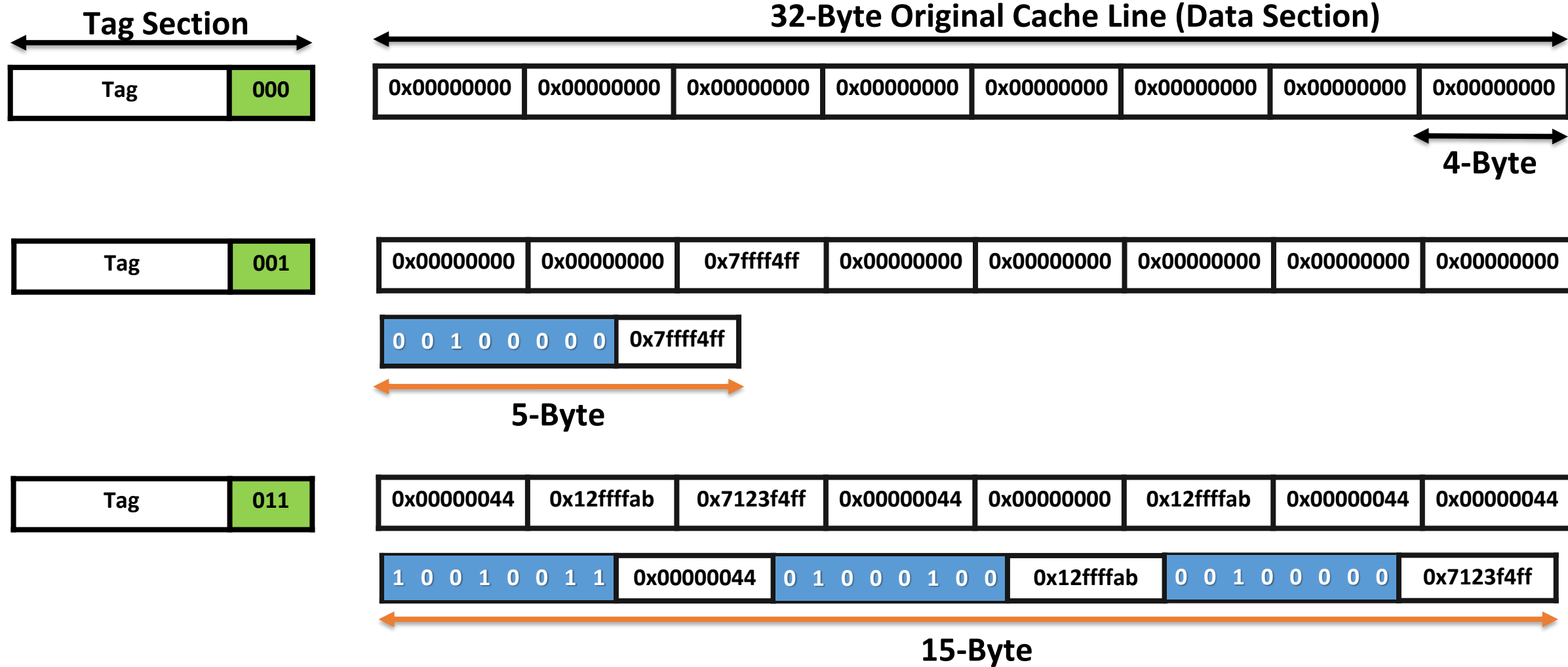
- Repeated Values

0xcdcccc3d	0xf000003f	0xf000803f	0xcdcccc3d	0xf000003f	0xf000003f	0xcdcccc3d	0xf000803f
------------	------------	------------	------------	------------	------------	------------	------------

- Few Non-zero Values (not repeating)

0x00000000	0x6666e63e	0x00000000	0x00000000	0x00000000	0x00000000	0x00000000	0x00000000
------------	------------	------------	------------	------------	------------	------------	------------

# Overview of the Proposed Method





- Evaluate our proposed method with GEM5 [8] simulator
- PARSEC and SPEC benchmarks and choosing the proper ROI
- Overhead analysis (area, power, and latency)
  - Design Compiler, Chisel
- Sensitivity analysis
  - Associativity
  - Replacement policy
  - Cache size
  - etc



- Profile a wide range of applications to get more insight on why the zeros and repetitive patterns manifest and how they can be exploited
- Compare with prior compression schemes
- Design a novel cache architecture that integrates Pooneh to compress blocks of varying granularities
- Explore this compression technique beyond just the last level cache

**Thank you!**





# References



- [1] A. R. Alameldeen and D. A. Wood, “Frequent Pattern Compression: A significance-based compression scheme for L2 caches,” Technical Report 1500, Computer Sciences Department, University of Wisconsin-Madison, Tech. Rep., 2004.
- [2] G. Pekhimenko, V. Seshadri, O. Mutlu, P. Gibbons, M. Kozuch, and T. Mowry, “Base-delta-immediatecompression: practical data compression for on-chip caches,” in Proceedings of the 21st International Conference on Parallel Architectures and Compilation Techniques, 2012, pp. 377–388.
- [3] J. Ziv and A. Lempel, “A universal algorithm for sequential data compression,” IEEE Transactions on Information Theory, vol. 23, no. 3, 1977, pp. 337–343.
- [4] J. Kim, M. Sullivan, E. Choukse, and M. Erez, “Bit-Plane Compression: Transforming Data for Better Compression in Many-Core Architectures,” in ISCA, 2016, pp. 329–340.
- [5] X. Chen, L. Yang, R. Dick, L. Shang, and H. Lekatsa, “CPACK: a high-performance microprocessor cache compression algorithm,” in IEEE Educational Activities Department vol. 18, 2010, pp. 1196–1208.
- [6] A. Arelakis and P. Stenstrom, “SC2: A Statistical Compression Cache Scheme,” in ISCA, 2014, pp. 145–156.
- [7] B. Panda, A. Sez nec, “Dictionary sharing: An efficient cache compression scheme for compressed caches.” In Microarchitecture (MICRO), 2016, pp. 1-12.
- [8] N. Binkert, B. Beckmann, G. Black, S. K. Reinhardt, A. Saidi, A. Basu, J. Hestness, D. R. Hower, T. Krishna, and S. Sardashti, “The gem5 simulator,” ACM SIGARCH Computer Architecture News, vol. 39, 2011, pp. 1-7.