Research on Compression Hardware for Energy Reduction in Cache Memory

Shunsuke Kawamura (0610028)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 7, 2008

Keywords: energy reduction, cache memory, data compression, floating-point data.

1 Introduction

In recent years, reduction of energy has been an important subject, since the energy consunption of a processor is increasing. Especially, leakage current is increasing as the process rule of transistors becomes deeper. The current which flows continuously even when the transistor is switched off is called leakage current. In the high-performance processor in which cache memory is increasing, energy reduction of cache memory attracts attention.

In this research, the energy-reducing system[1, 2] by voltage control and data compression of cache memory is reexamined. I investigate the optimal compression algorithm taking notice of the floating point data compression which has not been verified in particular.

2 Related works

2.1 Gated-Vdd[3]

Cache Decay[4] which uses Gated-Vdd[3] is mentioned as a method of reducing the energy of cache memory. This technique inserts a Gated-Vdd

transistor with a high threshold between the SRAM cell and GND, and energy is reduced by turning off the power supply to the cell.

2.2 The cache block compression algorithm in integer data[1, 2]

When the power supply of a cache block is turned off, the data in the cache block are lost, and the performance penalty by the increase in cache misses may occur. To solve this problem, there is the technique of reducing energy by compressing the data stored in a cache block, and turning off the power supply to the portion which was vacant after compression, preventing data loss.

The system tries compression for the data stored in the L2 cache memory. The data is stored into the L2 cache memory in the compressed form if the data are compressed into smaller than a half of the block size. Energy is reduced by turning off the power supply to the vacated space using Gated-Vdd. Data loss by turning off the power supply of the cache block is avoided by storing the data in the compressed form.

In literature[2], four compression algorithms were investigeted. In the simulation results of SPECint95, X-RL algorithm is most efficient and reduces energy by about 27% on average compared with the non-compressing execution.

3 Data compression for floating-point data

In literature[1, 2], the technique has not been verified for floating point data. Therefore, it is necessary to investigate the validity for floating point data, and an algorithm suitable for floating point data.

In this research, the energy-reducing system[1, 2] by voltage control and data compression of cache memory is reexamined. The optimal compression algorithm is investigated by taking notice of the floating point data compression which has not been verified in particular. Secondary (L2) cache is a target of the energy-reducing technique.

In this research, the FPC(Floating-Point Compression) algorithm[6] is newly introduced in order to compress the floating point data. The FPC algorithm compresses floating-point values by using two context prediction

mechanisms(FCM[7], DFCM[8]). Target value before the one of the compression value goes into FCM and DFCM. The predicted value is outputted by the value prediction based on the previous date. XOR with the predicted values is performed, and the one with longer leading-zero is chosen. Compression data is obtained by appending the non-zero part and the code which indictes how many zero bits are cut off.

4 Evaluation

I evaluate the results of simulations. SPECfp95 benchmark[9] was used as the programs for evaluation. The rate of energy reduction, the number of blocks per compression size and execution clock cycles were shown for each compression algorithms for every benchmark program. As a result of the experiment, the X-RL algorithm reduced the energy by about 14% on average compared with the non-compressing execution. The average execution speed was the increase in 5% with the X-RL algorithm which was able to carry out energy reduction most.

5 Conclusion

In this research, the energy-reducing system by voltage control and data compression of cache memory was reexamined. This research aimed at low power consumption of L2 cache by the optimal compression algorithm for the floating point data. Moreover, in addition to four conventional algorithms, the Floating-Point Compression algorithm was added and verified. For SPECfp95 benchmark, two or more data compression algorithms were used, and the energy reduction rate was evaluated.

References

[1] A.Matsuda. "Research on power consumption reduction that uses data compression", Japan Advanced Institute of Science and Technology Master thesis 2006.

[2] T.Kawahara. "Research on cache block compression algorithm for reducing power consumption", Japan Advanced Institute of Science and Technology Master thesis 2007.

[3] M.Powell, S.Yang, B.Falsafi, K.Roy, T.N.Vijaykumar. "Gated-Vdd: A Circuit Technique to Reduce Leakage in Deep-Submicron Cache Memories", Proc. of ISLPED, pp.90–95, 2000.

[4] S.Kaxiras, Z.Hu, M.Martonosi. "Cache Decay: Exploiting Generational Behavior to Reduce Cache Leakage Power", Proc. of ISCA, pp.240–251, 2001.

[5] M.Kjelso, M.Gooch, S.Jones. "Design and Performance of a Main Memory Hardware Data Compressor", Proc.of EuroMicro, pp.423–430, 1996.

[6] M.Burtscher, P.Ratanaworabhan. "High Throughput Compression of Double-Precision Floating-Point Data", Data Compression Conference, pp.293–302. 2007.

[7] Y. Sazeides and J. E. Smith. "The Predictability of Data Values", Proc. of 30th International Symposium on Microarchitecture, pp.248–258. 1997.

[8] B. Goeman, H. Vandierendonck and K. Bosschere. "Differential FCM: Increasing Value Prediction Accuracy by Improving Table Usage Efficiency", Proc. of HPCA, pp.207–216. 2001.

[9] Standard Performance Evaluation Corporation. 'SPEC CFP95 Benchmarks', http://www.spec.org/cpu95/CFP95/.