An Energy Efficient Set Associative Cache Algorithm

S.Subha
SITE
VIT, India

ABSTRACT
Set associative caches have fixed ways. Entire cache is enabled during cache operation. This paper proposes cache architecture mapping cache line to fixed cache way of mapped set. The address is mapped to set as in conventional set associative cache. The tag value of the mapped line is divided into blocks of size of number of cache ways. The average of maximum and minimum frequency of this division is the mapped way. The proposed model is simulated with SPEC2K benchmarks. The average memory access time degradation of 3.8% is seen over traditional set associative cache. The energy saving of 49% is observed in proposed model.

General Terms
Algorithms, Cache memories, Computer Architecture

Keywords
Average memory access time, Energy savings, Set associative cache

1. INTRODUCTION
Caches are of three kinds - direct mapped, set associative and fully associative[1,2]. A cache block is placed in fixed set in direct mapped cache. It occupies any of w fixed ways in w-way set associative cache. A cache block can be placed in any of n blocks in fully associative cache of n blocks. All cache ways are enabled in set associative cache during address mapping. Algorithms to enable selective cache ways are proposed in literature. An algorithm to selectively disable cache ways during modest activity resulting in energy saving for small degradation in performance is proposed in [3]. A method to predict a way and access directly in set associative cache is proposed in [5]. Energy is saved in this operation. Partial tag comparison of mapped set is proposed in [6] to save energy. The authors in [4] propose a method to perform tag comparison before cache access leading to single-cycle hit time. The author in [7] proposes an algorithm to enable cache ways selectively using next access time using profiling. If a match is found with next access time, the way is enabled else all cache ways of mapped set is enabled to place the line respecting the replacement policy of the cache.

This paper proposes a cache architecture enabling one cache way in set associative cache architecture. An address is mapped into specific way. The line is mapped to a set as in set associative cache. The tag portion of the address is divided into blocks of size of number of cache ways. The average of most frequent way and least frequent way is the mapped way. This algorithm is simulated with SPEC2K benchmarks. Energy saving of 49% with degradation of 3.8% in average memory access time over traditional set associative cache of same size is observed.

The rest of paper is organized as follows. Section 2 gives motivation, section 3 proposed model, section 4 mathematical model of proposed model, section 5 simulation, section 6 conclusion, section 7 acknowledgments followed by references.

2. MOTIVATION
Consider 2-way set associative cache of four sets of 32 bytes line size. Consider the address trace 1024, 2048, 4096, 1345, 2098, 3456, 100010. These are mapped to sets 0,0,0,1,2,0,2 respectively in traditional set associative cache. All the set ways are enabled in this mapping. Assume the cache operates in two modes-high power mode and low power mode. Let the energy consumed be denoted as E_{high} and E_{low} respectively. The energy consumed for this address trace is equal to

$$7*4*2*E_{high} \quad (1)$$

If E_{high}=10J the total energy consumed is 560J. Next consider the following algorithm for address a. Let each address be 32 bits wide with block size of 32bytes.

1. Compute setnumber = a %4, tagnumber = a div 4.

2. do {

Compute

x=tagnumber % 4

tagnumber = tagnumber div 4

Increment frequency[x]

} until (tagnumber !=0)

3. Let max=maximum(frequency[i]), i=1,2

min = minimum(frequency[i]), i=1,2

Choose i=(max+min)/2. Map the address to way- i.

4. Stop.

According to above algorithm the addresses are mapped to way zero of sets 0,0,0,1,2,0,2 respectively. One way is enabled in this mapping. The total energy consumed is given by

$$7*[2*4*E_{low}+E_{high}] \quad (2)$$

Assume E_{high}=10J and E_{low}=5J. The energy consumed in proposed algorithm from (2) is 7*[2*4*5+5]J=315J. An improvement of 43% in energy is observed. This is the motivation of this paper.

3. PROPOSED MODEL

Consider w-way set associative cache of S sets in level one of a system. Consider address a. Consider the following algorithm for address mapping of a.

1. Compute setnumber = a %S, tagnumber = a div S.

2. do {

Compute

x=tagnumber % w

tagnumber = tagnumber div w

Increment frequency[x]

} until (tagnumber !=0)

3. Let max=maximum(frequency[i]), i=1,2,..,w

min = minimum(frequency[i]), i=1,2,..,w

Choose i=(max+min)/2. Map the address to way- i.

4. Stop.

The proposed model is depicted in Fig. 1

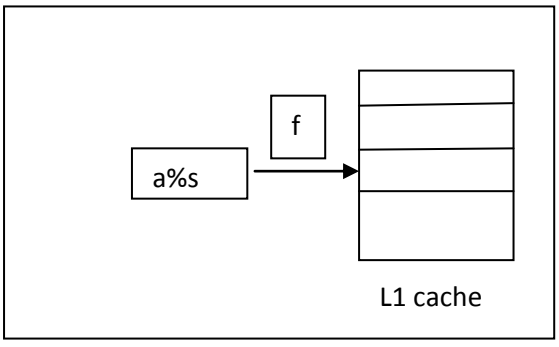

Figure1. Architecture of proposed model. An address a is mapped to a%s set. The function f maps to way.

4. MATHEMATICAL ANALYSIS OF PROPOSED MODEL

Consider a cache system with two cache levels L_1 and L_2. Let level one cache be w_1-way set associative cache of S_1 sets, level two w_2-way set associative cache of S_2 sets. Let the cache operate in two modes - high energy mode and low energy mode. Let the energy consumed be denoted as E_{high} and E_{low} respectively. Consider address trace of R references. Let h_1, m_1, h_2, m_2 denote number of level one hits, number of level one miss, number of level two hits, number of level two miss in proposed model respectively. Denote traditional set associative cache as C_{trad} and proposed model as C_{prop}.The energy consumed in traditional set associative cache is

$$E(C_{trad}) = (w_1S_1 + w_2S_2)RE_{high} \quad (3)$$

The term in (3) indicates that all the cache ways are enabled during cache operation.

Consider the cache model proposed in section 3. The energy consumed in this case is

$$E(C_{prop}) = (w_1S_1 + w_2S_2)RE_{low} + R(E_{high} - E_{low}) + (R - h_1)(w_2)(E_{high} - E_{low}) + E_{cs} \quad (4)$$

The first term is the energy consumed by cache system in non-operative mode. All accesses enable one level one cache way. This is indicated by second term. For level one misses, the level two cache is accessed. The level two cache is accessed during level one miss. This is the third term. The proposed model requires hardware to collect the frequencies, calculate the maximum and minimum frequency, the average of two. This consumes energy given by fourth term for this cache sub-system.

An improvement in energy consumption is observed if

$$(w_1S_1 + w_2S_2)RE_{high} >$$

$$(w_1S_1 + w_2S_2)RE_{low} + R(E_{high} - E_{low}) + (R - h_1)(w_2)(E_{high} - E_{low}) + E_{cs} \quad (5)$$

Let $H_1, H_2, T_1, T_2, T_{12}, M$ be level one hit time, level two hit time, transfer time between level one and level two, miss penalty respectively in traditional set associative cache. The average memory access time (AMAT) is given by

$$AMAT(C_{trad}) = \frac{1}{R}\begin{pmatrix} H_1T_1 + H_2(T_1 + T_2 + T_{12}) + \\ (R - H_1 - H_2)M \end{pmatrix} \quad (6)$$

Let t_1, t_2, t_{12}, m be level one hit time, level two hit time, transfer time between level one and level two, miss penalty respectively in proposed set associative cache. The AMAT is given by

$$AMAT(C_{prop}) = \frac{1}{R}\begin{pmatrix} h_1t_1 + h_2(t_1 + t_2 + t_{12}) + \\ (R - h_1 - h_2)m \end{pmatrix} \quad (7)$$

An improvement in AMAT is observed if

$$\frac{1}{R}(H_1T_1 + H_2(T_1 + T_2 + T_{12}) + (R - H_1 - H_2)M)$$

>

$$\frac{1}{R}(h_1t_1 + h_2(t_1 + t_2 + t_{12}) + (R - h_1 - h_2)m) \quad (8)$$

5. SIMULATION

The proposed model is simulated with SPEC2K benchmarks. Addresses from SPEC2K benchmarks were collected using Simplescalar Toolkit. Routines in C language were written to simulate the proposed model. The parameters for simulation are shown in Table 1. A two level cache model is assumed for simulation. Level one cache simulates the proposed model. Level two cache is inclusive set associative cache. The average memory access time is shown in Fig. 2. As seen from Fig. 2 the AMAT of proposed model degrades by 3.8% with traditional set associative. The energy consumption for proposed model is calculated as shown in Table 2. The address is assumed to be 32 bits wide. Energy during non-operation is 5J per way. Energy during operation is 10J per way. The cache size is same as for proposed model for traditional set associative cache. The energy required for additional hardware is assumed to be 5J per address. As seen from Table 2 energy saving of 49% is observed.

Table1 Simulation Parameters

S.No	Parameter	Value
1	L1 cache size	1024 sets
2	L1 associativity	64
3	L2 cache size	2048 sets
4	L2 associativity	128
5	Line size	32 bytes
6	L1 access time	3 cycles
7	L2 access time	12 cycles
8	L1 to L2 access time	20 cycles
9	Miss penalty	65 cycles

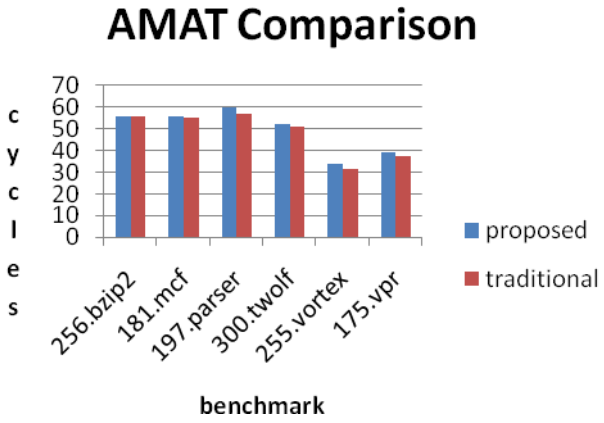

Figure2. AMAT comparison

Table 2 Energy Comparison

name	Energy	Energy(trad)	%improv
256.bzip2	7.65924E+11	1.53132E+12	49.98292
181.mcf	6496847320	12989235200	49.98283
197.parser	53174967630	1.06309E+11	49.98085
300.twolf	7396472650	14788198400	49.98395
255.vortex	19740458380	39472332800	49.98913
175.vpr	14349375240	28691660800	49.98765
average	1.44514E+11	2.88929E+11	49.98456

6. CONCLUSION

A set associative cache architecture enabling one way in mapped set is proposed in this paper. Mathematical model for the proposed model is developed. The proposed model is simulated with SPEC2K benchmarks. An energy saving of 49% with 3.8% increase in AMAT over traditional set associative cache is observed. The proposed model can be extended to include various measures of dispersion to map to cache way.

7. ACKNOWLEDGMENTS

The author expresses thanks to SantaClara University, CA, USA for providing SPEC2K benchmarks and Simplescalar Toolkit.

8. REFERENCES

[1] Alan Jay Smith, Cache Memories, Computing Surveys, Vol.14, No.3, September 1982, pages 473-530

[2] David .A. Patterson, John L Hennessey: Computer System Architecture : A Quantitative Approach, 3rd edition, Morgan Kaufmann Publishers Inc., 2003

[3] David.H. Albonesi: Selective Cache Ways: On Demand Cache Resource Allocation, Proceedings of International Symposium on Microarchitecture, 1999, pages 248-259

[4] Jian Chen, Ruihua Peng, Yuzhuo Fu, Low Power Set Associative Cache with Single-Cycle Partial Tag Comparison, Proceedings of ASICON 2005, pp. 73-77

[5] Michael D. Powell, Amit Agarwal, T. N. Vijaykumar, Babak Falsafi, and Kaushik Roy, Reducing Set-Associative Cache Energy via Way-prediction and Selective Direct-Mapping, Proceedings of MICRO, 2001

[6] RuiMin, Zhiyong Xu, Yiming Hu, Wen-ben Jone, Partial tag comparison: a new technology for power-efficient set-associative cache designs, Proceedings of 17th International Conference on VLSI Design, 2004, pp. 183-188

[7] S.Subha, A Set Associative Cache Model with Energy Saving, Proceedings of ITNG, 2008, pp. 1249-1250

.