Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A POWER SAVING VARYING CACHE WAYS ARCHITECTURE

S. SUBHA
SITE, Vellore Institute of Technology, Vellore, India
E-mail: ssubha@rocketmail.com

ABSTRACT

Extending the cache ways in conflict sets to alternate set has been proposed in literature. Selectively enabling cache ways to save energy is proposed in literature. This paper enables occupied ways in mapped sets and alternate sets increasing the cache associativity saving energy. The model consists of variable cache ways extending the ways of conflict set to alternate set. The occupied ways are enabled. The alternate set is chosen as the set with maximum unoccupied ways. The proposed model is simulated with SPEC2K benchmarks. On an average there is improvement in power consumption by 10.6% for the chosen cache size of 256 sets with associativities of 4,8,16,32. The percentage of power improvement for L1 sizes remains same with increase in number of sets of same capacity. It increases slightly from 256x4 to 256x32 level one cache size. The average memory access time is comparable with the chosen model for the chosen cache parameters.

Keywords: *Average Memory Access, Cache Power, Variable Cache Ways, Set Associative Cache, Selective Cache Way Enabling*

1. INTRODUCTION

A cache is represented as (C, k, L) where C is the capacity with associativity k and line size L. Caches are of three kinds - direct mapped, set associative and fully associative. A line is placed in fixed position in direct mapped cache. A line can occupy any of w ways in w-way set associative cache. A line can occupy any of n blocks in fully associative cache of n blocks. The performance of cache system is denoted by average memory access time (AMAT) [1,2]. Usually the entire cache is in high power mode during operation. Many models have been proposed to reduce the cache power consumption. The AMAT can be improved by increasing the number of cache ways in set associative cache. Several algorithms have been proposed to achieve this. The author in [7] proposes algorithm to extend the cache ways to adjacent sets. This model assumes the cache operates in two power modes - high power mode and low power mode. A line is in high power model on access. The author in [5] proposes cache model to extend the ways of conflicting set to alternate set. The set with maximum vacant ways is chosen as the alternate set. The entire cache is in high power model during the cache operation. The author in [8] proposes method to decrease energy consumption in the varying cache ways model. An algorithm for line placement/replacement is proposed. The cache ways are extended to alternate set on conflict. The set with maximum vacant ways is chosen as the alternate set. The most recently accessed set is probed during address mapping. There is decrease in number of ways probed in this model saving energy assuming the cache operates in two modes high power mode and low power mode. However in all the models mentioned above the cache ways in occupied sets are enabled be it in high power mode or low power mode. A model to enable occupied cache ways is proposed in [6]. This model saves power consumption for simulated cache sizes with comparable AMAT. The gated vdd method for selective cache sets and way enabling is proposed in [9]. In this model sets are selected using index bit masking. Way sub-masking is used to select the ways. However all the selected sets have same number of enabled ways. The authors of [3] propose variable cache sets and cache ways. However there is performance degradation in this model. The author in [4] proposes cache model to enable only occupied ways. An algorithm for line placement is defined in this paper. This is achieved by introducing sequential circuit for occupied ways. There is energy savings of 10% with improvement in AMAT in this model.

This paper proposes cache model to save power

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

consumption along with variable cache ways. The cache ways are extended on conflict by choosing the cache set with maximum unoccupied ways. The occupied ways are enabled. The cache is disabled initially. The proposed model is simulated with SPEC2K benchmarks. The variable cache ways model proposed in [5] is chosen as the model for comparing the results. On an average there is improvement in power consumption by 10.6% for the chosen cache size of 256 sets with associativities of 4,8,16,32. The percentage of power improvement for L1 sizes remains same with increase in number of sets of same capacity. It increases slightly from 256x4 to 256x32 level one cache size. The average memory access time is comparable with the chosen model for the chosen cache parameters. .

The rest of the paper is organized as follows. Section 2 gives the motivation, section 3 the proposed model, section 4 the mathematical analysis of proposed model, section 5 simulations followed by section 6 conclusion and references.

2. MOTIVATION

Consider two level inclusive cache. Let level one be 2-way set associative cache of four sets. Let level two be 4-way set associative cache of eight sets. Consider the address trace 100, 200, 400, 101. Let the line size be one byte. The lines are placed in sets zero, zero for addresses 100, 200 in level one cache. For address 400 there is conflict miss in set zero. Let the line be placed in set with maximum vacant ways according to the algorithm in [5]. Choose set two to be this set. The owner bit for way-0 set-2 is set to one indicating this. The owner bits in set zero are set to zero. The line 101 is placed in set one with owner bit set to zero. All the addresses are misses in level one and level two cache. Let 10W of power be consumed per cache way during operation. The total cache power is 4*(2*4*10+4*8*10) = 1600W. The entire cache is in high energy mode during operation. Next consider the following algorithm. The cache is put in off mode initially. On line being placed in cache way, the way is enabled using sequential circuit as described in [6]. Based on this algorithm, the two ways in set zero, one way in set one and one way in set two are enabled in the above cache line placement. Assume it takes 5W for the additional circuit. The total energy consumed is (20+15+10+5)W + (4+3+2+1)*10W + 4*8*10*4W = 1430W. The first term is the power consumed by additional circuitry. For the first address the circuit is enabled for four time units giving the power as 4*5W = 20W. The rest of the values in this term are calculated accordingly. The second term is the power consumed for enabling the cache ways. For the first address the way zero of set zero is enabled for four time units and so on. The third term is the power consumed by level two cache for the entire operation. An improvement in power consumption of about 10% is observed. This is the motivation of this paper.

3. PROPOSED MODEL

Consider two level inclusive cache. Let level one be w_1 set associative cache of S_1 sets. Let level two be w_2 set associative cache of S_2 sets. The placement/replacement for level two cache is as in traditional cache of least recently used method. Consider the first level cache. The following algorithm is implemented for line placement/replacement. Each cache way has one bit vector called owner_vector. This bit indicates the ownership of the way. If the line belongs to the set, this bit is set to zero. If the line belongs to another set, it is set to one. A set can be associated with another set. A vector allot_set with one entry per set indicates the associated set number for the set. It is set to -1 by default.

1. Start
2. Compute setno = $a\%S_1$ tango = $adivS_1$.
3. If the line is found in setno, increment the number of hits, access the line and stop.
4. Check the associated set for setno. If it is defined, check for the line in the associated set. If present increment the number of hits, access the line and stop.
5. If the line is absent in the set and its associated set (if defined) increment the number of misses. If there is vacant way in the set, place the line in this way, update the owner_vector, enable this way, access the line and stop.
6. If there is vacant way in the associated set (if defined), place the line in the associated set, update owner_vector, enable this way, access the line. and stop.
7. If the set is full, and associated set is undefined for this set, choose the set with maximum vacant ways that is not allotted to be the associated set, assoc_set. Update the allot_set entry for setno to be assoc_set. Place the line in the assoc_set, enable the way, access the line and stop.
8. If there are no vacancies in the set, its assoc_set, place the line in the least

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

recently used way of setno, update the owner_vector, access the line and stop.

The algorithm places a line in mapped set if there is vacancy. It associates the set with maximum vacant ways as the alternate set assoc_set. The algorithm places the line in assoc_set. On set full for the mapped and assoc_set, the line replaces the least recently used way of the mapped set. Both the mapped set and assoc_set are accessed usually during address mapping. By enabling only the occupied ways, the power consumption is reduced.

The way is enabled using sequential circuit as defined in [6]. A line on placement enables the way it is placed. A way remains enabled once line is placed in it. The architecture of the proposed model is shown in Fig. 1.

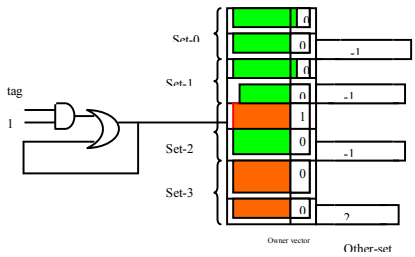

Figure 1 Architecture of Proposed System. Variable cache ways is implemented using owner_vector and other_set vectors. The logic for way enabling is shown for set-2 way-0. Similar circuits are present for ways in set-3. The other_set for set-3 is set-2. The ways in green color are disabled. The ways in red color are enabled.

4. MATHEMATICAL ANALYSIS OF PROPOSED MODEL

Consider two level inclusive cache described in section 3. Let the traditional set associative cache as described in [5] be denoted as C_{trad}. Let the proposed model be denoted as C_{prop}. Let h_1, m_1, h_2, h_o be level one hits, level one miss, level two hits, level one hits in associated set of mapped set in C_{trad}. Let t_1, t_2, t_{12}, M be level one access time, level two access time, transfer time between level one and level two, miss penalty for C_{trad}. The cache is assumed to operate in two power modes – high power mode or off mode. Let W_{high} be the power consumed in high power mode per cache way. Let there be R references. The average memory access time is given by

$$AMAT(C_{trad}) = \frac{1}{R}\begin{pmatrix} h_1 t_1 + h_o(2t_1) + \\ h_2(t_1 + t_2 + t_{12}) + \\ (R - h_1 - h_o - h_2)M \end{pmatrix} \quad (1)$$

The first term in (1) is the level one hit time in mapped set. The second term is the level one hit time in associated set. The third term is the level two hit time. The fourth term is the miss penalty.

Let H_1, H_o, H_2 be level one hits in mapped set, level one hits in alternate set of mapped set, level two hits in C_{prop}. Let T_1, T_2, T_{12}, m be level one access time, level two access time, transfer time between level one and level two, miss penalty in C_{prop}. The AMAT for this system is given by

$$AMAT(C_{prop}) = \frac{1}{R}\begin{pmatrix} H_1 T_1 + 2H_o T_1 + \\ H_2(T_1 + T_2 + T_{12}) + \\ (R - H_1 - H_o - H_2)m \end{pmatrix} \quad (2)$$

The first term in (2) is the level one hit time in mapped set. The second term is the level one hit time in alternate set of mapped set. Two accesses one to the mapped set and one to the alternate set are required in this case. The third term is level two hit time. The fourth term is the miss penalty. A performance improvement in AMAT is observed if

$$\frac{1}{R}\begin{pmatrix} h_1 t_1 + h_o(2t_1) + h_2(t_1 + t_2 + t_{12}) + \\ (R - h_1 - h_o - h_2)M \end{pmatrix} >= \frac{1}{R}\begin{pmatrix} H_1 T_1 + 2H_o T_1 + H_2(T_1 + T_2 + T_{12}) + \\ (R - H_1 - H_o - H_2)m \end{pmatrix} \quad (3)$$

Consider the power consumption. In the traditional model, the entire cache is in high power mode. The total power consumed depends on the number of ways in high energy mode. Let W_{high} be the energy consumed in high power mode per cache way. In the traditional cache, the entire cache is in high power mode. The total power consumed is given by

$$P(C_{trad}) = W_{high}(w_1 S_1 + w_2 S_2) \quad (4)$$

The first term in (4) is the power consumed by number of cache ways in level one cache. The second term is the power consumed by number of cache ways in level two cache. In the proposed model, ways on enabled on occupancy. Let w ways be enabled in level one cache during simulation of

Journal of Theoretical and Applied Information Technology
30th June 2015. Vol.76. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

R references. For level two cache the entire cache is in high power mode during cache operation. Let the power consumed by the additional circuitry be $W_{add\ln}$ .The total power consumed is given by

$$P(C_{prop}) = nW_{high} + w_2 S_2 W_{high} + W_{add\ln} \quad (5)$$

The first term in (5) is the power consumed in level one cache. The second term is the power consumed in level two cache. The third term is the power consumed by additional circuitry in level one cache. An improvement in power consumption is seen if

$$W_{high}(w_1 S_1 + w_2 S_2) >= nW_{high} + w_2 S_2 W_{high} + W_{add\ln} \quad (6)$$

5. SIMULATION

The proposed model is simulated with SPEC2K benchmarks. Addresses from SPEC2K benchmarks are collected using Simplescalar Toolkit. Routines in C language are written to simulate the proposed model. The simulation parameters are shown in Table 1. The proposed model is validated in hardware using Quartur 2. The Verilog program is simulated for representative input. The details of power consumption are derived from this simulation. The proposed model is compared with the variable cache ways model proposed in [5]. The average memory access time is shown in Fig. 2. It is assumed that access time to alternate set is equal to access time to set in this model. As seen from Fig. 2 the AMAT is comparable with the chosen variable cache ways model proposed in [5].

Table 1 Simulation Parameters

Parameter	Value
L1 size	256 sets
L1 associativity	4,8,16,32
L2 size	512 sets
L2 associativity	8,16,32,64
Line size	32 bytes
L1 access time	3 cycles
L2 access time	20 cycles
L1 to L2 transfer time	12 cycles
Miss penalty	65 cycles
Power per cache way in traditional/proposed cache	9.465mW
Power for seq. circuit for first circuit	66.90mW
Power for addln. Seq circuit	0.21mW

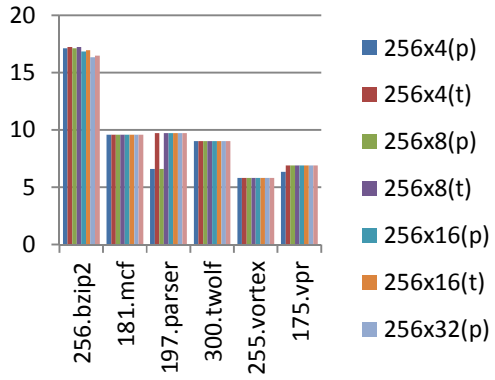

Figure 2 AMAT Comparison

The power consumed is shown in Fig. 3 and Fig. 4. As seen from Fig. 4 there is improvement in power consumption with increase in associativity for same number of sets.. The average power improvement grows almost linear with increase in associativity On an average there is improvement in power consumption by 10.6% for the chosen cache size of 256 sets with associativities of 4,8,16,32.

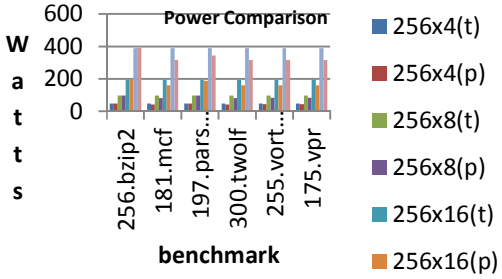

Figure 3 Power Comparison with varying associativity for fixed number of level one cache sets

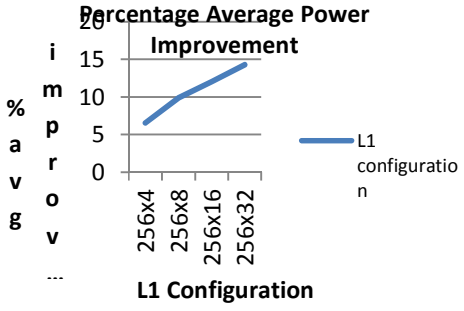

Figure 4 Percentage average power improvement for varying associativity

The power consumed for fixed cache size is simulated for varying associativities and set sizes. The cache size of 128-way set associative cache of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

64 sets was chosen for this. The average power improvement for various SPEC2K benchmarks for various set sizes and associativities are shown in Fig. 5. As seen from Fig. 5 the average percentage of power improvement remains almost constant with increase in set size.

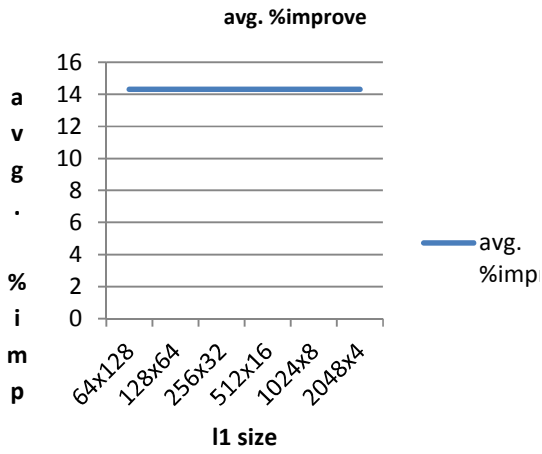

Figure 5 Average percentage power improvement for fixed level one cache size

6. CONCLUSION

A set associative cache architecture with variable ways and power saving is proposed in this paper. The lines mapped to cache set are placed in alternate set on conflict. The alternate set is chosen as set with maximum vacant ways. A set is alternate set to only one another set. The lines are placed in the mapped set or alternate set if vacancy exists. The line replaces the least recently used line of the mapped set in case of full mapped set and alternate set. A sequential circuit enables the cache line on occupancy. The lines are initially in off mode. The proposed model is simulated in first level cache of two level inclusive cache using SPEC2K benchmarks. The AMAT is comparable with the model proposed in [5]. On an average there is improvement in power consumption by 10.6% for the chosen cache size of 256 sets with associativities of 4,8,16,32. The percentage of power improvement for L1 sizes remains same with increase in number of sets of same capacity. It increases slightly from 256x4 to 256x32 level one cache size.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Alan Jay Smith, Cache Memories, Computing Surveys, Vol. 14 No.3, September 1982, pages 473-503

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

[3] K.T.Sundararajan, Timothy M.Jones, Nigel Topham, Smart Cache: A Self Adaptive Cache Architecture for Energy Efficiency, Proceedings of International Conference on Embedded Computer Systems, pp. 41-50 July 2011

[4] S. Subha, An Energy Saving Cache Algorithm, 2014 International Conference on Computational Science and Technology , 2014, ISBN: 978-1-4799-3241-2

[5] S.Subha, A Cache Architecture, Proceedings of ICACT 2013

[6] S.Subha, A Reconfigurable Cache Architecture, Proceedings of ICHPCA, 2014

[7] S.Subha, An Algorithm for Variable Cache Ways, Proceedings of ICATE 2013, January 23-25, 2013

[8] S.Subha, An Energy Saving Algorithm for Variable Cache Ways, Proceedings of ICRAIE 2014

[9] Se-Hyun Yang, Michael D Powell, Babak Falsafi, T.N.Vijaykumar, Exploiting Choice in Resizable Cache Design to Optimize Deep-Submicron Processor Energy Delay, Proceedings of HPCA, pp. 151-161,