

基于 Load 重用的低功耗数据 Cache 设计

李 泉 泉^{1,2}, 薛 志 远^{1,2}, 张 铁 军², 王 东 辉², 侯 朝 焕²

(1 中国科学院大学, 北京 100049; 2 中国科学院 声学研究所, 北京 100190)

摘 要: 针对嵌入式处理器中数据 Cache 功耗显著的特点, 提出了一种基于 Load 重用的低功耗数据 Cache 设计方法. 通过保存 Load 指令从数据 Cache 中取回的数据, 实现了随后 Load 指令对该数据的重新使用, 从而减少了数据 Cache 的访问次数, 有效降低了数据 Cache 的功耗. 在 SuperV_EF01 DSP 上的实验结果显示, 采用该方法后, 在处理器性能没有损失的情况下, 数据 Cache 功耗平均降低 29.48%, 面积仅增加 0.64%.

关键词: Load 重用; 低功耗; 数据 Cache; 嵌入式处理器

中图分类号: TP302

文献标识码: A

文章编号: 1000-7180(2014)06-0005-03

Low Power Data Cache Design Based on Load Reuse

LI Quan-quan^{1,2}, XUE Zhi-yuan^{1,2}, ZHANG Tie-jun², WANG Dong-hui², HOU Chao-huan²

(1 University of Chinese Academy of Sciences, Beijing 100049, China;

2 Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Data cache consumes a large amount of energy in embedded processor. This paper proposes a low power design method of data cache based on load reuse to save the data cache power consumption. The load reuse unit saves the loaded data from the data cache and the saved data could be reused by later loads. This approach can reduce a majority of data cache accesses, thus saving the data cache power consumption significantly. Experimental results of SuperV_EF01 DSP show that, in comparison with traditional data cache, this approach could save 29.48% of data cache power consumption, with only 0.64% of data cache area increasing and no performance degradation.

Key words: load reuse; low power; data cache; embedded processor

1 引言

随着集成电路技术的发展, 处理器的性能和集成度都得到了很大的提高, 由此带来的功耗问题也日益严重. 作为弥合处理器内核与主存之间速度差距的重要部件, 数据 Cache 由于具有很高的访问频率和很大的面积而功耗显著, 如文献[1]中的一款现代高端嵌入式处理器中, 有 28.5% 的功耗来自片上数据 Cache. 因此, 有效降低数据 Cache 的功耗, 对于降低整个处理器的功耗至关重要.

近年来出现了很多结构级低功耗 Cache 设计技术. Phased Cache^[2]将标志存储器的访问和数据存

储器的访问分为两个阶段进行, 消除了无关路数据存储器访问功耗, 但是该方法增加了访存延时, 降低了处理器的性能. Way-Predicting Cache^[3]通过记录程序运行过程中的 Cache 访问信息预测将要访问的路, 避免了无关路的访问, 但是预测失败会导致额外的标志存储器和数据存储器并行访问. Filter Cache^[4]通过在处理器内核和 Cache 之间增加一个容量较小的零级缓存来减少 Cache 的访问次数, 但是零级缓存缺失时仍然需要访问 Cache, 引起了额外的访存延时. 文献[5]中提出了一种基于超窄数据的低功耗数据 Cache 设计方法, 能够从动态和静态两个方面节省 Cache 的功耗, 但是该方法需要对

收稿日期: 2013-08-03; 修回日期: 2013-09-27

基金项目: 国家“九七三”计划项目(2009CB320202)

Cache 存储体进行修改。

针对现有方法存在的问题,本文提出一种基于 Load 重用的低功耗数据 Cache 设计方法.该方法在处理器性能没有损失的情况下,以较小的硬件代价减少了 Load 指令对数据 Cache 的访问次数,可有效降低数据 Cache 的功耗。

2 Load 重用分析

对于数据位宽为 N -bit 的数据 Cache,一次 Load 操作能够从数据 Cache 中读出 N -bit 的数据.如果随后的 Load 指令能够重新使用该数据,则可以减少 Load 指令对数据 Cache 的访问次数.图 1 中显示了部分 Powerstone 基准程序^[6]在数据 Cache 位宽为 128 bit 的情况下可重用 Load 指令占总的 Load 指令的比例,其中 blit 程序中可重用 Load 指令占总的 Load 指令比例最大,为 74.57%;ucbqsrt 程序中可重用 Load 指令占总的 Load 指令比例最小,为 22.75%;所有测试程序中可重用 Load 指令占总的 Load 指令的比例平均为 41.49%.如果能够消除这些可重用 Load 指令对数据 Cache 的访问,则可以有效降低数据 Cache 的功耗。

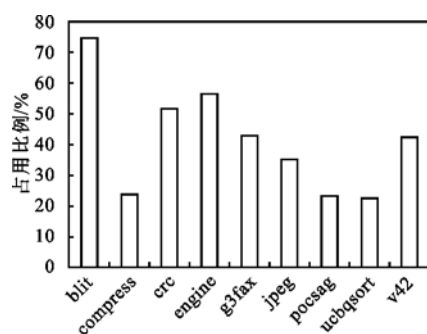


图 1 可重用 Load 指令的比例

3 数据 Cache 结构设计

为了减少 Load 指令对数据 Cache 的访问,本文提出了一种基于 Load 重用的低功耗数据 Cache 结构,如图 2 所示.其中数据 Cache 位宽为 128 bit,访存地址位宽为 32 bit.它在传统数据 Cache 结构基础上增加了一个 Load 重用单元. Load 重用单元主要包括一个 128 bit 的数据寄存器、一个 28 bit 的地址寄存器、一个比较器以及一些简单逻辑电路;其中数据寄存器用来保存 Load 指令从数据 Cache 中取回的数据,地址寄存器用来保存该数据对应的地址。

Load 重用单元的工作过程如图 3 所示.在程序执行过程中遇到 Load 指令时,Load 重用单元会将

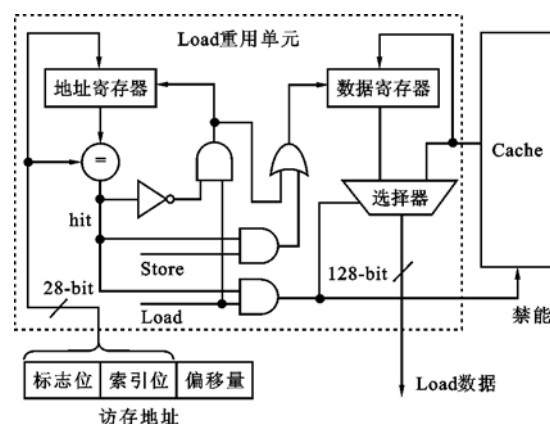


图 2 基于 Load 重用的数据 Cache 结构

访存地址中的高 28 bit 与地址寄存器中的值进行比较;如果相等,则表示地址寄存器命中,Load 部件直接从数据寄存器中读取数据而不需要访问数据 Cache;如果不相等,则 Load 部件从数据 Cache 中读取数据,Load 重用单元会将该访存地址中的高 28 bit 存入地址寄存器中并将该 Load 指令从数据 Cache 中取回的数据存入数据寄存器中.如果遇到 Store 指令且访存地址中的高 28 bit 命中地址寄存器,则 Store 部件会将数据同时写入数据 Cache 和数据寄存器中的对应位,图 2 中没有画出 Store 指令对应的数据通路.如果该访存地址中的高 28 bit 没有命中地址寄存器,则不需要修改数据寄存器。

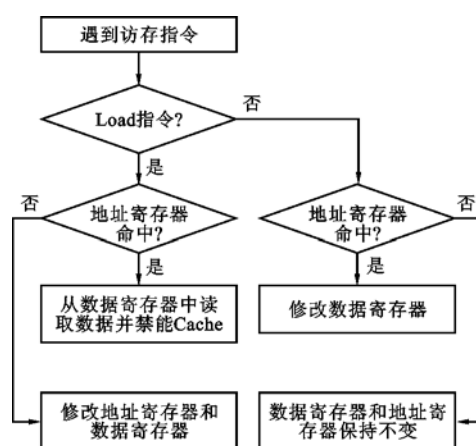


图 3 Load 重用单元工作流程图

在基于 Load 重用的取数过程中,后续 Load 指令可以充分利用前面 Load 指令从数据 Cache 中取回的数据,有效过滤掉了不必要的数据 Cache 访问,从而降低了数据 Cache 的功耗。

4 数据 Cache 功耗分析

传统数据 Cache 结构消耗的能量可以近似为式 (1), 其中 N_1 为出现数据 Cache 读命中的次数, N_2

为出现数据 Cache 读缺失的次数, N_3 为出现数据 Cache 写命中的次数, N_4 为出现数据 Cache 写缺失的次数, E_{DC_r} 为读一次数据 Cache 平均消耗的能量, E_{Miss_r} 为出现一次数据 Cache 读缺失平均消耗的能量, E_{DC_w} 为写一次数据 Cache 平均消耗的能量, E_{Miss_w} 为出现一次数据 Cache 写缺失平均消耗的能量。

$$E_{DC} = N_1 \times E_{DC_r} + N_2 \times E_{Miss_r} + N_3 \times E_{DC_w} + N_4 \times E_{Miss_w} \quad (1)$$

本文提出的数据 Cache 结构消耗的能量可以近似为式(2), 其中 N_5 为加入 Load 重用单元后出现数据 Cache 读命中的次数, E_{LR} 为 Load 重用单元消耗的能量。本文提出的数据 Cache 结构相对传统数据 Cache 结构获得的能量节省比例可以表示为式(3)。由于 Load 重用单元的硬件结构很简单, 因此 E_{LR} 的值很小, μ 主要受 N_5 的影响, 即读数据 Cache 的次数越少, 获得的能量节省比例越大。

$$E_{DC_new} = N_5 \times E_{DC_r} + N_2 \times E_{Miss_r} + N_3 \times E_{DC_w} + N_4 \times E_{Miss_w} + E_{LR} \quad (2)$$

$$\begin{aligned} \mu &= \frac{E_{DC} - E_{DC_new}}{E_{DC}} \\ &= \frac{(N_1 - N_5) \times E_{DC_r} - E_{LR}}{E_{DC}} \end{aligned} \quad (3)$$

5 实验结果

以 SuperV_EF01^[7] 数字信号处理器 (digital signal processor, DSP) 为基础, 设计实现了本文提出的基于 Load 重用的低功耗数据 Cache 结构。SuperV_EF01 DSP 是中国科学院声学研究所研发的一款面向高性能嵌入式应用的、具有自主设计指令集的 32 位 DSP, 采用四发射超长指令字结构, 包含向量协处理器和浮点协处理器, 工作频率为 500 MHz; 它的片内包含一个数据位宽为 128 bit、容量为 8 kB 的直接映射方式的指令 Cache 和一个数据位宽为 128 bit、容量为 16 kB 的两路组相联映射方式的数据 Cache。实验所用的测试程序来自于 Powerstone 基准程序, 其访存指令占总指令的比例如图 4 所示。

在 SMIC 90 nm 工艺下采用 Design Compiler 对加入 Load 重用单元的 SuperV_EF01 DSP 设计文件进行综合。使用 VCS 对综合后生成的网表文件进行测试, 得到各个测试程序的执行时间和开关活动性统计文件。基于开关活动性统计文件, PrimeTime PX 可以分析出数据 Cache 的功耗。以传统数

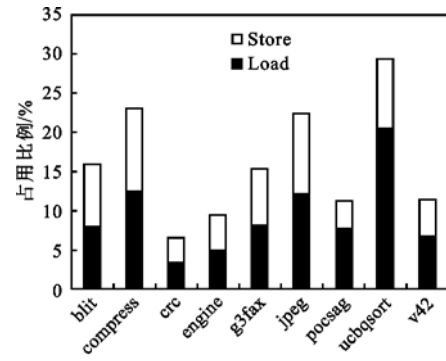


图4 访存指令占总指令的比例

据 Cache 结构下的测试结果为基准, 对采用本文方法后的测试结果进行归一化, 如图 5 所示。采用本文方法后, 运行 blit 程序时获得的数据 Cache 功耗节省比例最大, 为 49.88%; 运行 compress 程序时获得的数据 Cache 功耗节省比例最小, 为 17.04%; 运行所有测试程序时获得的数据 Cache 功耗节省比例平均为 29.48%。采用本文方法后, 测试程序的执行时间没有发生变化, 这主要是由于 Load 重用单元只是改变了 Load 部件的取数通道, 对 Load 指令的取数次数和数据 Cache 的缺失次数并没有影响。

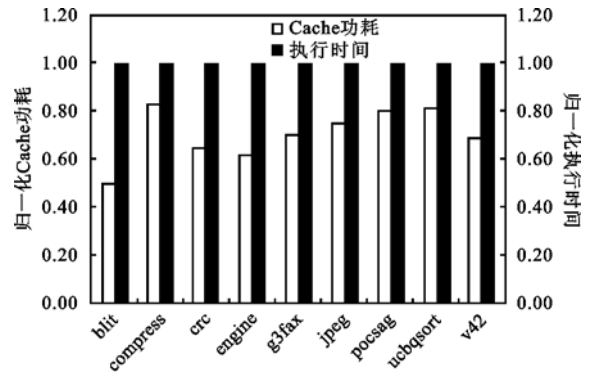


图5 对功耗和性能的影响

表 1 中显示了本文方法对数据 Cache 面积的影响。与传统数据 Cache 相比, 其面积仅增加 0.64%, 硬件开销很小。

表 1 对面积的影响

传统数据 Cache 面积/ μm^2	542 055.26
本文结构面积/ μm^2	545 518.07
增加率/%	0.64

6 结束语

本文根据 Load 指令对数据 Cache 的访问特性, 提出了一种基于 Load 重用的低功耗数据 Cache 设计方法。该方法通过重新使用 Load 指令从数据

(下转第 11 页)

参考文献:

- [1] 刘敏钰,吴咏,武卫国. 无线传感网络(WSN)研究[J]. 微电子学与计算机,2005,22(7):58-61.
- [2] Hu L, Evans D. Localization for mobile sensor networks [C] //Proceedings of 10th International Conference on Mobile Computing and Networking. Philadelphia: Association for Computing Machinery, 2004: 45-57.
- [3] Wang W, Zhu Q. RSS-based monte carlo localisation for mobile sensor networks [J]. IET Communications, 2008, 2(5):673-681.
- [4] Bao Ying, Chen Junfeng, Shi Zhiguo, et al. New real-time resampling algorithm for particle filters [C] // Wireless Communications & Signal Processing. Nanjing,2009:1-5.
- [5] Fross D, Langer J, Fross A, et al. Hardware implementation of a particle filter for location estimation[C] //2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN). Korea; Busan, 2010: 1-6.
- [6] Adiprawita W, Ahmad A S, Sembiring J, et al. New resampling algorithm for particle filter localization for mobile robot with 3 ultrasonic sonar sensor[C]//Electrical Engineering and Informatics (ICEEI). Bandung, Indonesia,2011:1-6.
- [7] Bolic M, Athalya A, Djuric P M, et al. Generic hardware architecture for sampling and resampling in particle filters [J]. EURASIP Journal on Applied Signal Processing, 2005(17):2888-2902.
- [8] Winkler G. Image analysis random fields and dynamic monte carlo methods[M]. Berlin, Heidelberg; Springer-Verlag, 1995.
- [9] 束礼宝,宋克柱,王砚方. 伪随机数发生器的FPGA实现与研究[J]. 电路与系统学报,2003,8(3):121-124.
- [10] He T, Huang C D, Blum B M, et al. Range-free localization schemes for large scale sensor networks[C] //Proceedings of the Ninth Annual International Conference on Mobile Computing and networking. San Diego: Association for Computing Machinery, 2003: 81- 95.
- [11] 洪少华,史治国,陈抗生. 用于纯方位跟踪的简化粒子滤波算法及其硬件实现[J]. 电子与信息学报, 2009, 31(1): 96-100.

作者简介:

雷鑑铭 男,(1974-),博士,副教授.研究方向为 WBAN 关键技术研究及硬件系统、神经接口及脑机接口关键技术及电路设计、数模混合超大规模集成电路设计及工艺。

(上接第7页)

Cache 中取回的数据来减少 Load 指令对数据 Cache 的访问次数,从而有效降低数据 Cache 的功耗.在 SuperV_EF01 DSP 上的实验结果显示本文方法在处理器性能没有损失的情况下可以平均节省 29.48% 的数据 Cache 功耗.此外,本文方法的硬件实现简单,非常适合于低功耗嵌入式处理器设计.

参考文献:

- [1] Aragon J L, Veidenbaum A V. Optimizing CAM-based instruction cache designs for low-power embedded systems[J]. Journal of Systems Architecture, 2008, 54(12):1155-1163.
- [2] Hasegawa A, Kawasaki I, Yamada K, et al. SH3: high code density, low power[J]. IEEE Micro, 1995, 15(6): 11-19.
- [3] Powell M D, Agarwal A, Vijaykumar T N, et al. Reducing set-associative cache energy via way-prediction and selective direct-mapping[C] // Proceedings of the 34th International Symposium on Microarchitecture, Austin, TX, United States: IEEE Computer Society Press, 2001: 54-65.
- [4] Kin J, Gupta M, Mangione-Smith W H. Filtering memory references to increase energy efficiency[J]. IEEE Transactions on Computers, 2000, 49(1): 1-15.
- [5] 马志强,季振洲,胡铭曾. 基于超窄数据的低功耗数据 Cache 方案[J]. 计算机研究与发展, 2007, 44(5): 775-781.
- [6] Malik A, Moyer B, Cermak D. A low power unified cache architecture providing power and performance flexibility[C] // Proceedings of the 2000 International Symposium on Low Power Electronics and Design. Rapallo, Italy: ACM Press, 2000: 241-243.
- [7] 王琪,高瑛珂,华斯亮,等. 可复用微处理器片上调试功能的设计与实现[J]. 计算机辅助设计与图形学学报, 2012, 24(10): 1369-1374.

作者简介:

李泉泉 男,(1987-),博士研究生.研究方向为嵌入式处理器体系结构、数字集成电路设计与验证。