

片上多核处理器的结构级功耗建模与优化技术研究*

张戈** 胡伟武 黄琨 曾洪博 王君

中国科学院计算技术研究所 微处理器中心, 中国科学院计算机系统结构重点实验室, 北京 100190

摘要 功耗是导致片上多核处理器出现故障的重要诱因, 也是片上多核处理器设计的重要制约因素. 如何降低多核处理器的功耗并提高处理器能量效率, 具有很大的研究意义与探索空间. 文中主要从体系结构设计者的角度, 并结合电路实现, 研究并总结纳米级工艺下片上多核处理器的功耗建模与评估方法, 及其不同构件的低功耗优化技术. 通过提出创新高效的多核处理器结构级功耗评估方法及其模拟平台, 提高多核结构功耗模拟的准确性与灵活性, 并以此为依托, 开展处理器核、片上网络、片上存储及其一致性协议的各方面优化, 寻求提高多核处理器功耗有效性的微体系结构, 为国产多核处理器的低功耗设计提供一定借鉴与参考.

关键词 片上多核处理器 功耗建模 功耗评估 低功耗结构

半导体工艺的迅速发展使微处理器的速度和集成度越来越高, 同时处理器的功耗与散发的热量也呈指数级增长, 每三年处理器的功耗密度就能翻一番. 主频至上的计算机处理器设计技术正在终结, 性能功耗比继性能价格比之后正在成为计算机的主要设计指标, 低功耗和热优化设计已经成为微处理器研究中的核心问题. 目前, 越来越多的处理器朝着单片多核处理器 (chip multi-processor, 简称 CMP)^[1] 的方向发展, 在片上多核处理器设计中, 功耗有效的结构设计是决定这些处理器是否具有可扩展潜力以及好的性能功耗比的重要因素之一, 因此已经有越来越多的设计者开始关注多核处理器的功耗评估和功耗优化方法.

片内多核处理器结构的主要设计思想是通过简化超标量结构, 将多个相对简单的超标量处理器核集成到一个芯片上, 这样可以充分利用工艺进步带来的芯片面积上数以亿计的晶体管的增加, 同时避免连线延迟的影响, 并充分开发线程级并行性 (thread-level parallelism, TLP), 提高处理器吞吐

量. 出于上述原因, 片内多核已经势不可当地成为了当今通用处理器的主流发展方向. 目前已经有大量的基于多核结构的处理器芯片面世, 如 IBM Power5^[2], Intel Montecito^[3], Sun Niagara^[4], CELL 处理器^[5] 等. 从一定程度上讲, 多核结构有效缓解了单处理器核功耗的爆炸性增长, 但是, 功耗问题在多核处理器设计中并没有得到根本解决, 并且其特有的多核心结构决定了多核处理器的功耗研究仍是一个至关重要的课题. 另外, 在纳米级工艺条件下 (90 nm 以下), 一些新的功耗问题, 如漏电功耗等使得片上多核的功耗问题更加突出. 因此, 多核处理器更加需要良好的功耗解决方案, 低功耗设计也成为了贯穿多核处理器各个层次设计的主要考虑因素之一.

从体系结构角度看, 多核处理器主要包括处理器核、核间互连以及片上存储体 (cache) 三个主要部分. 由此, 片上多核处理器的体系结构低功耗研究可以围绕功耗评估, 处理器核功耗优化, 片上网络功耗优化以及片上缓存功耗优化这四个方面展开,

2009-02-22 收稿, 2009-08-09 收修改稿

* 国家重点基础研究发展计划 (批准号: 2005CB321603)、国家高技术研究发展计划 (批准号: 2009AA01Z125) 和国家自然科学基金 (批准号: 60803029) 资助项目

E-mail: gzhang@ict.ac.cn

其中行之有效的功耗评估方法是开展多核处理器低功耗设计的基础。

1 多核功耗建模与评估

正确有效的功耗建模与评估方法是有效开展低功耗设计的基础。从功耗来源上看, CMOS 电路产生的功耗主要分为动态功耗和静态功耗。随着工艺进入纳米级工艺以及晶体管数目的增长, 漏电功耗的问题在多核处理器中已经越来越严重。从功耗评估方法上看, 功耗评估研究的两个焦点问题是精度和效率问题。通常来讲, 功耗评估的抽象层次越高, 其分析精度就越差, 但其分析效率和功耗优化技术的成效就越好。反之, 抽象层次越低, 其分析精度就越高, 但其分析效率和功耗优化技术的成效就越差。在片上多核处理器的设计中, 由于处理器各个层次的设计都受到功耗因素的制约, 不同构件间又存在相互影响, 能否在设计早期就对功耗进行完整和准确的测量, 并对不同的设计方案作出评价, 具有非常重要的意义。

鉴于结构级功耗评估的重要性, 国内外已有很多学者和机构对此进行了研究。在针对单处理器核的功耗评估方面, 应用较为广泛的结构级功耗模拟器就包括 Wattch^[6], SimplePower^[7], Sim-Panalyzer^[8], PowerTimer^[9], Hot-Leakage^[10], XTREM^[11], Sesame^[12]等。在上述模拟器中, 以 Wattch 最具有代表性, 也是目前学术界应用最多的功耗模拟器之一, 它的基本做法是按照功能把处理器分成不同的模块, 对每个模块基于分析或经验赋予相应的功耗代价, 并与性能模拟器相结合, 根据模块被访问的次数计算出相应功耗。另外, 随着多核处理器和片上系统的流行, 针对片上路由和互连线的高层模拟器也应运而生, 如 Orion^[13], INTACTE^[14]等。但是, 由于结构级功耗模拟的固有难度与不准确性, 目前的处理器结构级功耗分析工具也只限于在学术领域的一些研究应用, 还难以见到商业化的产品。

结构级功耗模拟的主要问题在于无法对处理器模块的实际电路特性(如晶体管尺寸、物理连线、实际工艺参数以及电路形式等)进行准确的刻画, 因此其功耗模拟误差较大。以目前应用最广泛的 Wattch 为例, 它把处理器中的大部分结构归结为 RAM/CAM 两类结构, 这种刻画方式过于抽象,

不能很好地反应真实处理器设计中的情况; 其次, 对于处理器中普遍存在的大量随机控制逻辑, 由于其表现形式不规则, 也无法通过一个简单的抽象模型进行刻画。除此之外, 面对新工艺与新结构, 片上多核处理器的功耗建模还存在以下几个新的挑战。

1.1 纳米级工艺条件下的功耗新问题

工艺进步使得片上多核处理器可以集成更多的晶体管, 目前已经有越来越多的处理器采用 65 和 45 nm 工艺, 如正在研制中的龙芯 3 号多核处理器^[15]就采用了国际上较为先进的 65 nm 工艺。在这种工艺条件下, 工艺偏差对性能和功耗的影响加大, 芯片的电压降(IR-drop)以及在片波动(on-chip variation)问题严重, 片上连线的数目和宽度都显著增加, 连线功耗占处理器总功耗的比重不断上升(如有的处理器中连线功耗已占到处理器总动态功耗的 50%以上^[16], 并且研究表明这种趋势在多核处理器环境下将更加突出^[17])。与此同时, 多核处理器中漏电功耗的比重也越来越大, 在工艺特征尺寸小于 90 nm 条件下广泛采用了多阈值电压工艺, 因此必须慎重评估上述各个方面对多核处理器设计的影响, 结合准确的电路与工艺参数进行功耗计算。

1.2 多核处理器中结构变化对功耗的影响

多核处理器所集成的单个处理器核必须保持良好的低功耗特性, 因此不再像传统单处理器那样, 为了一味地追求性能而采用过于复杂的结构, 转而在设计中尽量使结构简洁有效, 更加倾向于采用以往的嵌入式处理器内核的结构, 因此以往的高性能通用处理器的功耗评估模型会显得不太适应。另外, 对多核处理器中新出现的互连部件, 如交叉开关(crossbar)、路由器(router)、缓冲器(buffer)等, 也都需要进行单独的功耗建模。尤其是对于核间长距离的互连线, 需要进行细致准确的功耗刻画。

1.3 物理实现方法对功耗的影响

对于同一种结构, 其不同的物理实现方法会造成完全不同的功耗结果, 如龙芯 2F 处理器核与龙芯 2E 处理器核^[18]采用了完全相同的体系结构, 但通过物理设计中的优化措施使得龙芯 2F 处理器的平均功耗降低了 30%—40%, 因此不同的物理实现方法对处理器实际功耗有着很大的影响。在多核处

理器中,处理器核为了保持良好的低功耗特性,往往采取了更加复杂的功耗管理与低功耗设计手段,如动态频率、电压调节,低功耗任务调度,睡眠、休眠各种低功耗模式等,并且在电路设计方法上,多核处理器开始由传统高能耗的全定制动态电路设计,转向为以静态电路和标准单元为主的低功耗电路设计方法。在这种情况下,结构级功耗评估方法如果不能考虑物理实现的具体变化,其功耗评估的结果将变得更加不可信。

因此,一个好的结构级多核功耗评估模型应该能够考虑到上述各方面因素,正确评估结构、电路与工艺的各方面变化对功耗评估结果带来的影响,从而为结构设计者提供准确可信的功耗参考数据,为多核结构的功耗优化提供有效指导与服务。

2 基于物理反标的功耗评估方法与平台

基于上述分析,我们在研究过程中,提出一种新的结构级功耗评估方法和应用平台,其基本思想是:把处理器的结构研究与电路研究紧密结合,通过对多核处理器中各种基本功能模块进行参数化电路实现和功耗特征提取,形成标准的功耗库模型,使处理器电路的物理信息能够灵活而准确地反馈至上层模拟工具,从而使得传统结构级功耗模拟方法中不好估计的物理参数(如晶体管尺寸、模块内连线以至漏电功耗、面积、时序等)都具有真实而详细的数据,从而大大提高结构级功耗模拟的准确

性。

图1为基于物理反标的多核处理器性能、功耗模拟平台,该功耗模拟平台包括高层模拟和底层实现两个部分,二者通过基于物理反标的功耗模型库进行调用,能够快速准确地对一个多核处理器的不同微体系结构进行功耗建模与评估。其中底层实现部分包括对处理器各种基本块的电路建模与参数化RTL的描述,通过自动综合与功耗计算EDA平台,能够对各个电路基本块的功耗数据进行自动计算,并反标至功耗模型库。高层模拟部分主要包括一个多核性能模拟器和一个多核功耗模拟器,多核性能模拟器以龙芯2号处理器核和片上网络为设计基础,功耗模拟器通过接收性能模拟器的微体系参数、模块翻转信息,并调用功耗模型库的功耗建模数据,完成对处理器运行程序过程中的实时功耗计算。这种基于物理反标的功耗建模与评估流程,可以有效地利用上层结构模拟的快速性和底层功耗建模的准确性,一定程度上解决了多核功耗模拟中的一些关键问题。该平台的初步实现在文献[18]中具有更加详细地阐述,目前已完成设计与初步校准,并在以龙芯3号多核芯片的物理设计中做进一步校准与完善。

在建模内容上,我们上述提出的功耗模拟平台充分考虑了片上多核结构的新特征,因此建模内容更加真实,能够准确地反映未来片上多核系统的功耗特性,并进行不同多核结构的功耗模拟。在我们

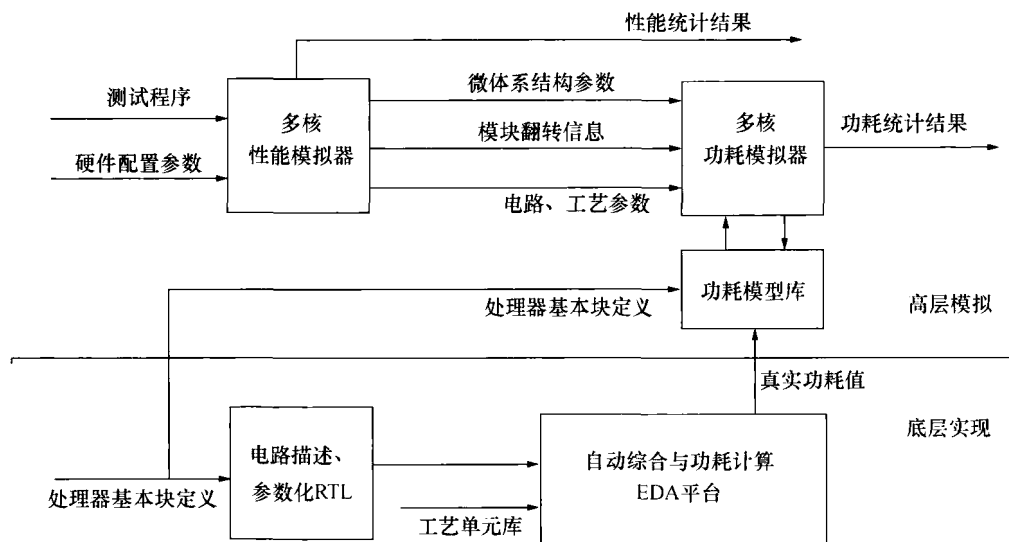


图1 基于物理反标的多核处理器性能、功耗模拟平台

的多核功耗模拟器中,不但包括了动态功耗的评估,还包括了漏电功耗和电路面积的评估.对处理器核间的网络互连部件,如路由器、交叉开关,仲裁器,以及核间长距离的互连线等,该模拟平台提出了专门的功耗与延迟计算模型进行刻画,从而能够针对目前流行的片上网络结构进行全面而准确地模拟.

表 1 总结了基于物理反标的功耗评估技术与传统功耗评估技术的比较.传统结构级功耗评估方法由于缺乏实际电路信息而不能对功耗做出准确的估计,而电路级的功耗分析方法虽然精度较高,但速度和灵活性较差.与国外相关研究工作相比,该平台通过建模方式、建模对象与建模内容的新思路和新方法,一定程度上提高了结构级功耗评估的准确性和实用性.我们通过以龙芯 2E 为建模对象的实验分析和比较,初步表明该模拟器设计方法正确,速度较快,可信度高,其功耗建模得到的误差比传统结构级模拟器的功耗评估误差明显要小^[19].

表 1 基于物理反标的功耗评估技术与传统功耗评估技术的比较

	传统结构级功耗评估技术	传统门级/电路级功耗评估技术	基于物理反标的功耗评估技术
建模方法	基于大模块的结构抽象	基于 SPICE 或门级网表仿真	基于物理反标的多层次、细粒度建模
建模对象	单核处理器(动态电路为主的高功耗结构)	小规模功能电路	多核处理器(静态单元设计为主的低功耗结构)
灵活性	好(可灵活地评估不同的处理器结构)	差(必须依赖电路或门级实现)	好(可灵活地评估不同的多核处理器结构)
仿真速度	快(几百 k 指令/s)	很慢	快(几百 k 指令/s)
准确度	差	好	较好

基于上述模拟平台,我们可开展多核处理器的一系列结构评估和低功耗设计工作,目前已取得了一定成果与应用,包括:(i)针对并行应用程序行为,提出了一种动态电压、频率调度算法,可以有效降低多核处理器中每个处理器核的平均运行功耗.(ii)针对多核处理器片上网络的数据传输特性,提出一种零有效的片上网络电路结构,提高片上网络通信的功耗效率.同时对片上网络中的缓冲器设计做进一步研究,提出了功耗有效的片上网络缓冲器设计.(iii)针对多核处理器中的缓存一致性问题,提出了一种能够降低功耗开销的缓存一致性协

议,提高了缓存一致性协议的效率与功耗有效性.

3 处理器核功耗优化与功耗管理

就像提高处理器性能需要从算法、系统、结构、电路、工艺等多个层次来考虑一样,降低功耗也需要从上述多个层次不同角度进行系统研究.例如,可以通过设计新的算法减少程序执行过程中运算的次数以减少能量消耗;可以由编译优化减少访存次数来减少功耗;可以通过操作系统关闭当前不用的功能部件及处理器核,或者动态调整处理器核的时钟频率及电压达到控制功耗的目的;可以重新设计总线的编码方式,减少同时跳变的信号个数来降低功耗;可以在逻辑和电路层次上通过门控时钟、门控电源、双阈值电压、动态电压变换、半频率时钟、异步逻辑等方法降低功耗;可以通过工艺进步减少耦合电容以及降低漏电来降低功耗等.

对于片上多核处理器而言,以往单核处理器的多数低功耗技术都可以继续应用于片上多核处理器.但与传统高性能单处理器设计中所不同的是,片上多核处理器的每个处理器核不再倾向于采用高复杂度、高功耗的处理器核,而是转而通过在片上集成更多的复杂性较低的处理器核,以利于通过开发线程级并行性来提升性能,降低功耗.另一方面,在片上多核处理器中,功耗管理变得更加重要.由于片上多核的每个处理器核是相对独立并通过片上网络松散耦合的,每个处理器核运行完全独立的指令流(multi-programming)或者运行通过同步操作协同运行的指令流(parallel programming).在这种情况下,程序本身在不同时段的运行特性(指令级并行度、访存操作密度、二级缓存失效率等)不同,或者不同处理器核到达同步点的时间先后不同,各个处理器核在同一时刻的繁忙状况也不同.利用程序本身不同阶段执行特性的差别或者并行程序各线程不精确同步的特点,可以单独对各个处理器核进行变压变频控制,降低功耗.

图 2 给出了我们在多核功耗与性能模拟平台上运行 Ocean 并行程序时,其中一个处理器核上的 IPC 以及定点队列、访存队列、失效队列的利用率的变化情况(每隔 10000 拍统计一次).从图中揭示的 IPC 和队列利用率等体系结构信息的变化趋势可以看出:从整体来看,处理器核在不同时期的性能

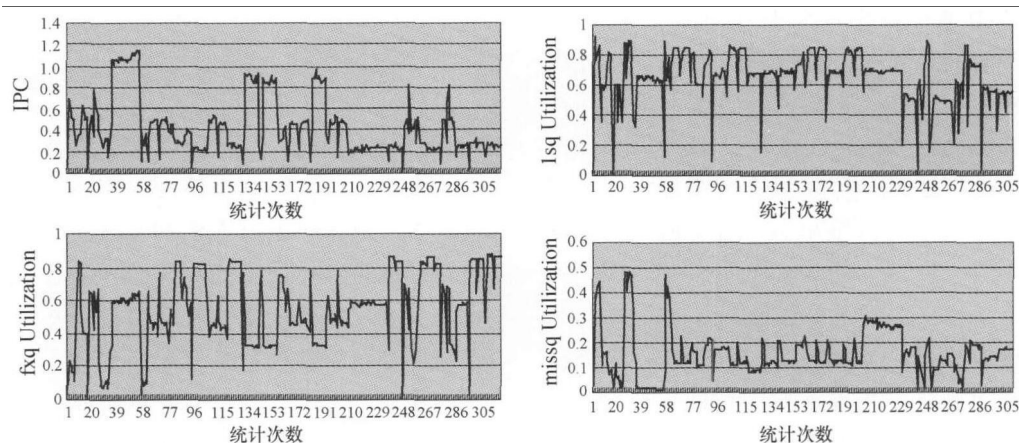


图2 处理器核运行并行负载程序时的体系结构参数变化情况

变化得很剧烈, 因此对于每个处理器核实施动态电压、频率缩放(dynamic voltage & frequency scaling, DVFS)算法拥有相当大的功耗优化空间. 另一方面, 仔细观察数个时间间隔的IPC等参数变化情况, 则可以发现这些参数变化相对较平稳, 这一点则为DVFS的预测算法提供了可能性: 可以利用前几个时间间隔内的处理器核的IPC等体系结构的变化情况, 推测未来时间间隔内性能的变化可能性, 以此来决定何时和如何作DVFS.

由于在多核处理器中应用动态电压、频率调节的重要性的应用潜力, 近年来开始有不少工作针对多核处理器DVFS等功耗管理方法进行了研究. 文献[20]研究了在一个众核(many core)体系结构上运行的并行程序在给定性能约束条件下的功耗优化问题. 优化空间包括对处理器核个数的调节以及芯片整体的DVFS技术. 文章提出了一种启发式搜索算法减少功耗优化的搜索空间, 适用于程序的执行具有重复性的并行程序, 并需要事先对程序特性进行统计. 文献[21]把处理器核的个数作为参数, 从理论和实验上分析了在相同性能下的功耗和相同功耗下的性能, 认为在给定的性能目标下并行计算能够带来很大的功耗节省空间. 文献[22]基于栅障(Barrier)机制实现高级配置与电源接口(advanced configuration and power interface, ACPI)电源管理规范, 提高并行程序的功耗有效性. 其基本思想是当某个处理器核过早地到达栅障时, 可以让其进入睡眠状态以节省无谓的功耗开销. 由于进出睡眠状态的切换操作功耗开销也较大, 在处理器核到达栅障

之前降低其工作电压和频率, 使它与其他处理器核能同时到达栅障, 可能会更好地节省功耗. 文献[23]提出使用相同指令集异构多核的方法降低功耗. 该方法限定任意时刻只能有一个处理器核处于活跃状态, 因此根据程序的当前运行特性, 动态地在这些不同复杂性的处理器核间进行切换, 可以达到提高性能并降低功耗的目的, 但这种结构仅适用于串行程序.

时钟网络是处理器的重要组成部分, 龙芯3号的设计实现了模块级的变频, 每个处理器核及模块时钟可以运行于独立的频率, 可以十分方便地根据程序行为和温度传感器的反馈对各模块动态地调节频率. 同时为了满足时序需要和减轻在片波动(on chip variation)的影响, 时钟设计通过全局mesh结构与局部H树相结合的方式, 并在局部时钟分布上采用细粒度门控时钟进一步控制时钟动态功耗. 另外, 如何在不同频率的时钟异步通信方式上进行高效的同步设计也是我们的研究内容之一^[24]. 此外, 以前的时钟网络设计基本上都是全局同步的设计, 但是由于硅片的增大和片上晶体管数量指数型增加, 使得要设计出满足低扭斜和低抖动要求的全局同步的时钟网络已经越来越困难. 因此, 在片上多核处理器中有必要将目光转向全局异步、局部同步(globally asynchronous locally synchronous, GALS)的设计^[25], 其优点是不需要设计满足全局同步要求, 约束极为苛刻的低扭斜和低抖动全局时钟网络, 使各模块的时钟设计变得简单.

基于全局同步、局部异步的时钟设计, 我们在

多核功耗、性能评估平台上,模拟了处理器多个处理器核之间运行在不同的时钟频率以及电压下的情况,提出了一种基于投票选择的动态电压/频率调节算法,其基本思想是通过对每个处理器核的 IPC、定点队列、浮点队列以及访存失效队列利用率的变化趋势的研究,按照投票策略组织这些信息来动态调节各处理器核的电压和频率,并通过延迟决定算法过滤不准确的信息毛刺,从而有效降低多核处理器运行时的能量消耗并提高其能量效率.具体算法如下:每个处理器核以当前时间间隔的 IPC 变化趋势为一组,定点队列和浮点队列的变化趋势为一组,失效队列变化趋势的相反数为一组,三组数据采用投票选择算法,即当三组数据中有两组或以上的变化趋势显示下一时间间隔需要调节电压和频率

则记录下来且累加,若不需要调节则累减,当连续做出调节决定的次数超过阈值时则真正地调节电压和频率.我们采用上述方法对多核处理器运行 SPLASH-2 并行基准测试,并对其程序的性能、功耗进行统计,表 2 给出了我们模拟的多核结构参数和 DVFS 算法的阈值参数,图 3 则给出了实验结果.根据实验结果统计,处理器功耗平均节省 23.4%,而性能下降 9.9%,能量-延迟积由此可以降低 14.9%.同时,在研究中我们发现,并行程序由于片上互连网络的原因,处理器的访存行为与串行程序相比有更大的不确定性,处理器核等待访存结果的时间通常相对更长,这样可以给 DVFS 算法更大的利用空间,所以相对于串行程序, DVFS 算法用在并行程序上具有更好的效果.

表 2 多核模拟器中的结构参数和 DVFS 算法中的阈值参数

处理器核数量	16	处理器的 DVFS 参数	值(参考 Intel XScale)
单处理器核参数	值	电压改变范围/V	0.65—1.20
分支预测器	Gshare,9 位 GHR,4096 项 PHT,128 项 BTB,两路组相联,随机替换	频率改变范围/GHz	0.250—1.0
流水线宽度	4:4:5:5:4	频率改变速度/ns·MHz ⁻¹	49.1
取指:译码:发射:写回:提交		同步开销	1.0 GHz 时钟周期的 30%(300ps)
功能部件	2 个定点部件,1 个访存部件,2 个浮点部件	DVFS 算法的阈值参数	值
定点发射队列	16	Interval	10000
浮点发射队列	16	IPC DeviationThreshold/%	10
存储访问队列	32	Fxq DeviationThreshold/%	10
失效访问队列	8	Ftq DeviationThreshold/%	10
指令提交队列	64	Missq DeviationThreshold/%	10
Instruction Cache	64 kB,4 路组相联,随机替换	SaturateCounter	6
Data Cache	64 kB,4 路组相联,随机替换	FreqChange/%	12.5
Level 2 Cache	512 kB,4 路组相联,随机替换		

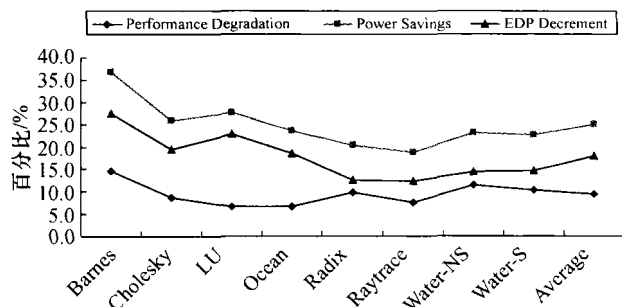


图 3 16 核处理器运行 SPLASH-2 程序的测试结果

4 片上网络功耗优化

随着处理器设计进入纳米级工艺与多核时代,片上互联线的数目、宽度和延迟都显著增加,信号跨越整个芯片需要多个周期,造成片上通信和存储体访问时延的不均匀性.在互连方式上,片上网络(networks on chip-NoC)代替总线连接处理器核和片上缓存,具有较高的吞吐率和较好的可扩展性,已经成为片上多核处理器互连结构的事实标准^[26].

但片上网络同时也引入了功耗开销,例如在 Alpha 21364 中,互连网络的功耗占据了整个处理器总功耗的 25%^[27],而 MIT 的 Raw 处理器中的片上互连网络占据了处理器总功耗的 36%,其中每个本地节点上的 Router 的功耗占据节点功耗的 40%^[28],因此研究片上网络的功耗有效性对提高整个多核处理器的功耗效率有着重要意义。

片上网络的结构如图 4 所示,主要包括 4 个组件:输入缓冲(input buffer)、仲裁器(arbiter)、交叉开关(crossbar)和链接(link)。图中的仲裁器由路由电路(routing logic)、仲裁电路(arbitration logic)和流控制电路(flow control logic)组成,流控制电路还接收下行网络传送过来的 Credit 信息;图中的队列为输入缓冲,片上网络有多少方向的输入通道,就有多少队列项;交叉开关逻辑将输入缓冲中的数据包转发给输出通道;而链接则是连接不同路由电路之间的数据线,即输入通道和输出通道。

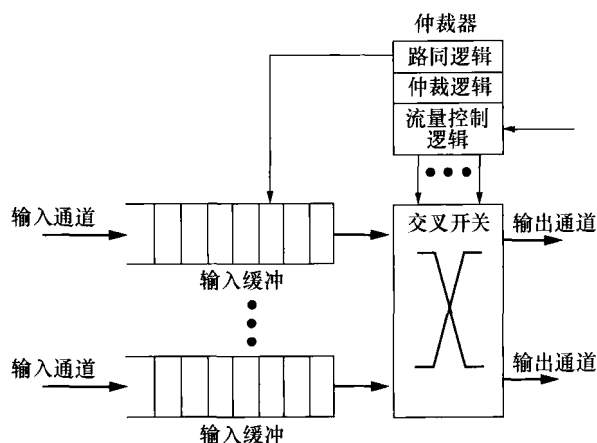


图 4 片上网络结构示意图

影响片上网络功耗的因素很多,主要可以分为两个方面:微体系结构和注入流量。微体系结构主要决定单个消息在片上网络中传递的功耗,而注入流量则决定片上网络中消息的密度和总量。片上多核处理器一般遵循共享存储编程模型,且共享片上二级缓存也成为设计主流^[29]。每个处理器核一级缓存失效引发到共享二级缓存的请求与应答以及缓存一致性协议要求的核间数据传输和无效请求与应答形成片上网络中的主要流量;在分片式片上多处理器结构中,处理器核发出的访存请求映射到二级缓存块的方式也会导致消息在片上网络里传输的跳

数及拥塞程度不同,形成不同的功耗。从部件组成看,仲裁器由于自身所占用的晶体管资源很少,功耗很小,几乎可以忽略不计;而其他组件都占用大量晶体管资源和金属线资源,且翻转率很高,因此功耗也是相当可观的。

可以从不同角度降低片上网络功耗,如微体系结构与互连电路的选择,缓存一致性协议的优化,以及访存地址到片上网络节点的映射方法等。文献[30]针对片上网络路由电路提出了 3 个低功耗技术:分段交叉开关(segment crossbar)、捷径交叉开关(cut-through crossbar)和直写缓冲(write-through buffer)。文献[31]基于片上网络数据传输宽度不断增加而控制传输宽度保持不变的假设,指出控制传输将造成绝大部分传输线的无谓翻转,提出把片上网络 FIFO 从物理上划分为控制信号与数据信号两部分,并根据消息的性质采用部分预充电,从而降低功耗。文献[32]将动态电压调整技术应用于片上网络的互连线,采用基于历史信息的动态电压调整策略,根据连线的占用状况调整连线的电压和频率。文献[33]基于分片式片上多处理器和共享二级缓存,指出若一级缓存失效引起的片上网络请求由一个远程二级缓存块应答,消息的延迟和引起的功耗都将增大。该文提出受损复原(victim replication)技术和受损迁移(victim migration)技术,有选择地把一级缓存的替换块备份在本地二级缓存块中,从而减少远程访问带来的片上网络延迟和功耗开销。

我们基于第三节中提出的多核处理器的功耗、性能模拟平台,对多核处理器片上网络传输信号的特点进行了一定分析,通过把片上网络上传输的信号分为三类:(i)数据信号,主要是一级指令缓存和一级数据缓存失效后或脏块替换时与二级缓存之间交换的数据内容;(ii)地址信号,主要是一级缓存失效时访问相应的二级缓存的物理地址;(iii)控制信号,用来控制地址和数据信号的传输。对上述三类信号进行实验和分析后发现,互连网络传输中信号为 0 的可能性大大高于信号为 1 的可能性,这主要是由于在传输数据和地址中存在很多数值较小的立即数,基址偏移、标志数、数组下标、循环变量、同步变量等引起。图 5 表示了在我们的多核处理器模拟平台上运行 SPLASH-2 时片上网络传输信号的 0—1 百分比情况。

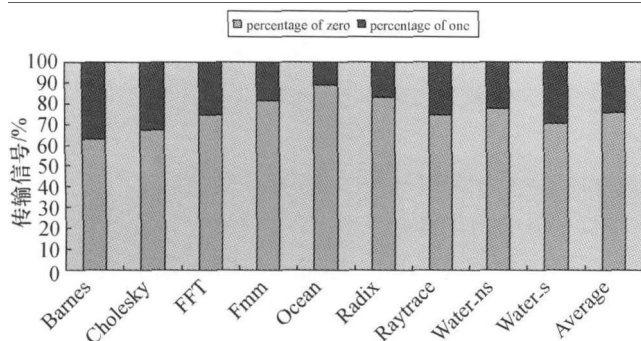


图 5 多核处理器运行 SPLASH-2 时 NoC 上传输信号的 0—1 百分比

由此我们指出,采用“零有效”的片上互连网络微体系结构及其电路实现,能够显著提高互连网络的功耗效率.如图 6 所示,我们利用动态电路的形式,对传统片上网络中的三种主要部件:输入缓冲器、链接、交叉开关结构进行改进,当这些部件的传输或输出结果为 0 时不需要进行放电,从而有效地提高了功耗效率.我们在文献[34, 35]中,则进一步针对片上网络中的输入输出缓冲,进行了不同宽度、深度下片上网络功耗、性能的模拟,提出了适用于片上网络的缓冲器设计,并同时提出了提高电路设计可靠性的方法.

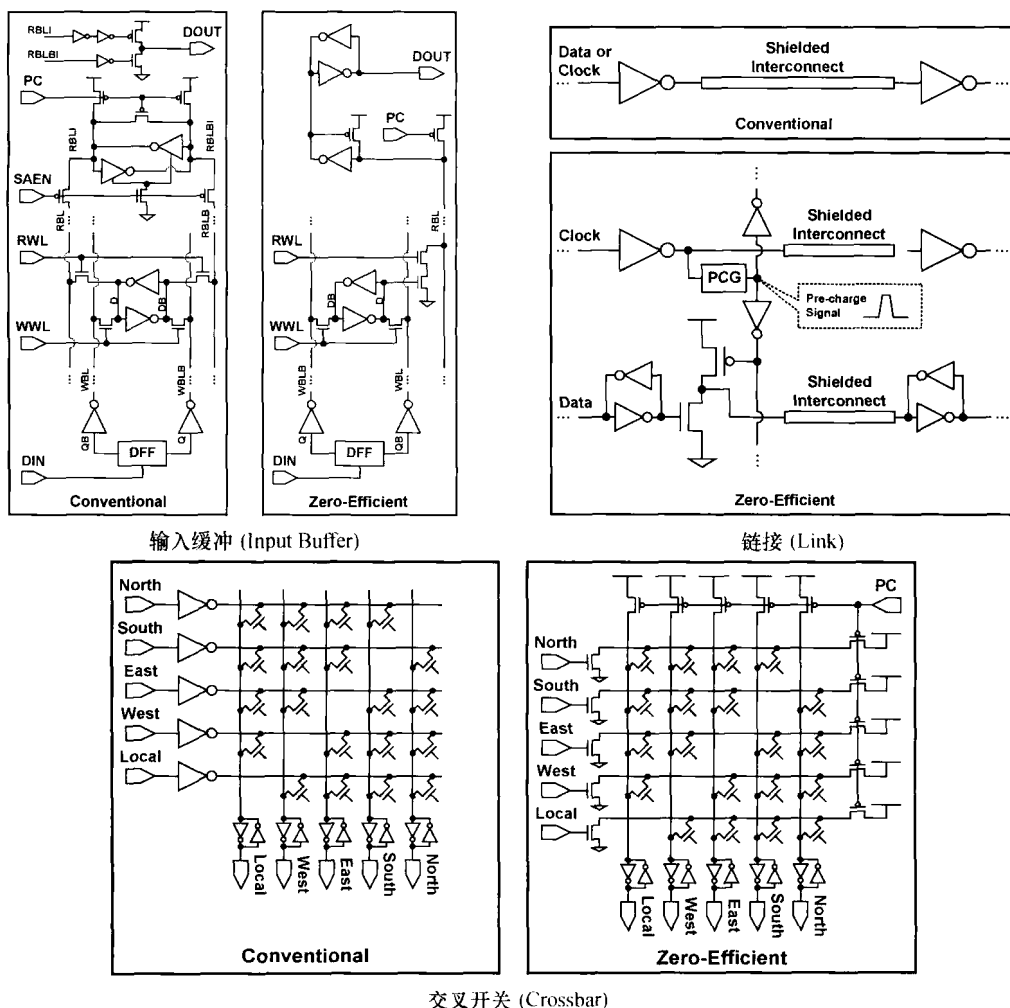


图 6 传统片上网络结构与“零有效”的片上网络结构对比

5 片上缓存与缓存一致性功耗优化

工艺的进步使得片上集成更多的晶体管成为可能,这对于多核处理器设计尤为重要。在纳米级工艺条件下,可以利用丰富的片上资源,把更多的存储单元(cache)集成到片内,显著提高存储访问的带宽。由于在片上多核处理器中,多个进程/线程同时运行需要更大的片上缓存容纳所有进程/线程的工作集,且会造成更为严重的资源冲突,极大地增加了片上缓存的访问压力,而片外存储体的访问延迟和功耗都非常大(文献[36]指出,片外存储体访问功耗可达片上缓存访问功耗的35倍)。因此如何减少片外存储体访问并降低片上缓存功耗,也是片上多核处理器当前研究的热点。

已有大量工作针对传统缓存结构进行了功耗方面的优化,其中如何对缓存进行分块以及如何减少平均访问次数是结构级研究的主要思路。目前,高性能处理器中缓存大都采用多路组相联结构,在多核处理器设计中尤其如此,因此如何降低多路组相联缓存的并行访问功耗尤其重要。传统缓存结构中一般同时访问多路的标签(tag)与数据(data)部分,如图7(a)所示,在保证性能的同时也浪费了过多

的能量。为了解决传统多路组相联缓存消耗过多不必要功耗的缺陷,文献[37]中提出了 phased cache,首先进行标签比较决定访问哪一路,然后再访问该路数据,如图7(b)所示。但这样延长了缓存的访问时间。为了能同时取得高性能与低功耗的优点,文献[38]中提出了针对多路组相联缓存结构的路预测方法,每次访问缓存之前需要访问路预测表格,预测当前访问所在的路。基于传统路预测结构,并结合指令访问顺序性很强的特点,我们在龙芯 GS232 CPU IP 中提出并实现了一种新的路预测方法来降低指令缓存的功耗,如图7(c)所示,称之为快速路预测结构。该结构将路预测信息存储在 cache 块或 BTB(分支目标地址缓存)与 RAS(返回地址栈)中,在取当前指令的同时得到下次取指所需要的路预测信息,能够有效解决传统路预测方法中路预测信息与缓存访问串行的缺陷,有更高的路预测准确率,而且路预测准确率受组相联程度的影响也很小。实验表明,与并行结构相比,采用该结构的4路组相联指令缓存,对 SPEC CPU 2000 基准测试程序可获得平均95%以上的路预测准确率,访问功耗平均降低73.3%,而 cache 访问时间仅增加2.31%。

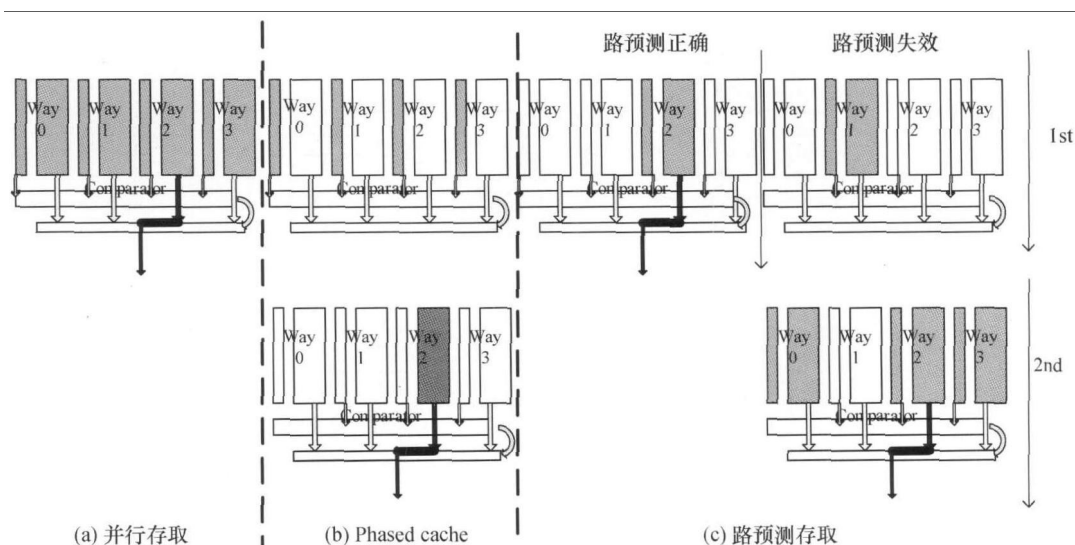


图7 并行访问 cache(a), phased cache(b)与路预测 cache(c)行为比较

在如何降低片上多处理器的二级缓存功耗方面,文献[29]比较了私有二级缓存和基于地址散列的分散共享二级缓存,提出了“基于处理器的共享

二级缓存分组”(shared processor-based split L2)技术,每个处理器核都根据动态运行特性,分配到若干二级缓存分组。该技术介于私有二级缓存和基

于地址散列的分散共享二级缓存之间, 但无论处理器核分配到的二级缓存分组的查询还是其他二级缓存分组的查询都需要多个分组同时进行, 这会产生一定的开销. 文献[39]指出, 在互相竞争的进程的缓存失效率的导数相等时, 这些进程的总缓存失效率数将达到最小, 并采用较为简单的硬件跟踪缓存失效率导数的一阶近似, 在不同进程间合理分配二级缓存份额, 减少高功耗的片外存储体访问操作. 文献[40]指出, 传统的 LRU 替换策略会使得局部性差的进程侵吞其他进程的共享二级缓存份额, 增加二级缓存上的冲突, 因此采用改进的 LRU 策略, 限制局部性差的进程的共享二级缓存份额, 可以减少二级缓存的失效次数. 该文通过在体系结构级提供二级缓存份额寄存器, 并由操作系统根据进程实时运行信息配置各进程的二级缓存份额, 达到硬件简单和操作系统宏观控制的设计目标. 文献[41]提出了一种新颖的从访存请求到分散共享的二级缓存块的映射方法. 通过将从访存请求到二级缓存块的映射与页映射绑定的方法, 在页映射转换出物理页号的同时, 也转换出该物理页对应的二级缓存块号, 通过合理的映射, 使得访存请求尽可能多地由靠近处理器核的二级缓存块应答, 同时也降低了片上网络的延迟和功耗.

由片上多级缓存所带来的片上缓存一致性是片上多核处理器具有的另一个重要特征. 传统的缓存一致性协议主要从降低延迟, 提高性能入手, 并不

十分关心缓存一致性协议对功耗带来的影响, 但随着片上多核数目的增长和功耗问题的日益严重, 从功耗优化的角度来考虑缓存一致性协议设计, 对芯片系统结构设计人员提出了新的问题与研究角度. 缓存一致性的功耗研究需要借助一个能够灵活而详细地进行功耗、性能分析的多核模拟平台, 通过提出的多核功耗、性能模拟平台, 可以方便地研究缓存一致性协议所带来的功耗开销问题, 并对其进行各方面的优化. 基于上述思路, 我们提出了功耗感知的片上缓存一致性协议^[42], 降低缓存一致性协议在多核处理器中的功耗开销.

将片上网络的特性暴露给一致性协议将会带来一些功耗优化的机会. 在实验过程中, 我们观察到当目录在收到读写请求后查询缓存块状态时, 可能会发现存在某些处理器核的一级缓存中持有共享数据备份或者独占地持有数据备份, 并且它们到请求者的距离比从目录(即是 home 结点)到请求者的距离要短. 在获知这些处理器核在片上网络中位置的情况下, 一致性协议可以将请求转发给其中离请求者最近的共享者(或者独占数据持有者). 数据由此处理器核中的一级缓存送回给请求者, 而不是由较远 home 结点中的二级缓存体提供. 图 8 表示了对于读写失效的情况下改进的缓存一致性协议的工作过程, 其中 R, H, S 分别代表远程请求节点、宿主节点以及较近的共享数据的节点. 通过改进的缓存一致性协议, 可以有效地降低片上访存的功耗.

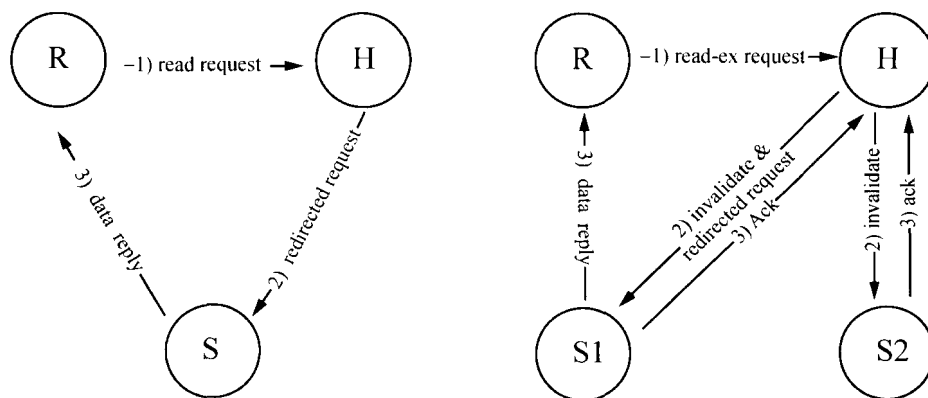


图 8 基于功耗优化的片上缓存一致性协议

不难看出, 在未来的研究工作中, 如何提高访存效率并降低片上访存功耗是体系结构研究的长期

热点, 而提高访存效率与降低功耗总体上是相一致的, 因此能够有效提高访存效率的方法一般来讲同

样也有益于降低访存功耗。

6 总结

多核处理器的低功耗设计已成为学术界和工业界研究的共同热点,对于片上多核处理器的不同设计模块和设计层次,都存在行之有效的降低功耗的方法,而这些方法又可能是互相牵制,互相影响的。因此多核处理器设计需要贯穿系统结构到电路工艺的各方面的丰富知识,尤其是需要结构设计和物理设计的结合,才能在设计早期对一个处理器设计做出正确的评估,从而做出优化的多核架构的选择。本文从多核处理器的结构级功耗评估方法,以及处理器核功耗优化与功耗管理、片上网络优化、片上缓存与缓存一致性功耗优化几个方面入手,分别阐述了多核处理器中功耗评估与优化设计中的一些问题与方法,并重点介绍了基于物理反标的多核功耗模拟平台,以及在该平台上开展的一系列相关研究与应用。本文工作以龙芯多核处理器设计为研究背景,提出多核处理器的低功耗设计与功耗管理解决方案,希望能够为未来国产处理器的设计方法提供一定借鉴与参考。

参考文献

- Hammond L, Nayfeh BA, Olukotun K. A single-chip multiprocessor. *IEEE Computer*, 1997, 30(9): 79—85
- Kalla R, Sinharoy B, Tendler JM. IBM Power5 Chip: A dual-core multithreaded processor. *IEEE Micro*, 2004, 24(2): 40—47
- Naffziger S, Stackhouse B, et al. The implementation of a 2-core, multi-threaded itanium family processor. *IEEE Journal of Solid-State Circuits*, 2006, 41(1): 197—209
- Kongetira P, Aingaran K, Olukotun K. Niagara: A 32-Way multithreaded sparc processor. *IEEE Micro*, 2005, 25(2): 21—29
- Pham D, Behnen E, et al. The design and implementation of a first-generation CELL processor. In: *Proc of the 2005 IEEE International Solid State Circuits Conference*, San Francisco, CA, USA, February 2005, 45—50
- Brooks D, Tiwari V, Martonosi M. Wattch: A framework for architectural-level power analysis and optimizations. In: *Proceedings of the 27th Annual International Symposium on Computer Architecture*, June 2000, 83—94
- Ye W, Vijaykrishnan N, Kandemir M, et al. The design and use of simplepower: A cycle-accurate energy estimation tool. *Proceedings of 38th Design Automation Conference*, June 2000, 340—345
- Project TSPM. Sim-Panalyzer2.0 _ Reference Manual. Tech Report. University of Michigan, the University of Colorado, 2004
- Brooks D, Bose P, Srinivasan V, et al. New methodology for early-stage, microarchitecture-level power-performance analysis of microprocessors. *IBM Journal of Research and Development*, Sep-Nov 2003
- Zhang Y, Parikh D, Sankaranarayanan K, et al. Hotleakage: A temperature-aware model of subthreshold and gate leakage for architects. Tech Report. CS-2003-05. Department of Computer Science, University of Virginia, 2003
- Gilberto Contreras, Margaret Martonosi, Jinzhan Peng, et al. XTREM: A power simulator for the Intel XScale® core. *ACM SIGPLAN Notices*, 2004, 39(7): 115—125
- van Stralen P, Pimentel AD. Signature-based microprocessor power modeling for rapid system-level design space exploration. *IEEE/ACM/IFIP Workshop on Embedded Systems for Real-Time Multimedia*, Oct. 2007, 33—38
- Wang HS, Zhu XP, Peh LS, et al. Orion: A power-performance simulator for interconnection networks. *Proceedings of the 35th Annual IEEE/ACM International Symposium on Microarchitecture*, 2002, 294—305
- Rahul Nagpal, Arvind Madan, Amrutur Bhardwaj, et al. INTACTE: An interconnect area, delay, and energy estimation tool for microarchitectural explorations. *Proceedings of the 2007 International Conference on Compilers, Architecture, and Synthesis for Embedded Systems*. Salzburg, Austria. 2007, 238—247
- Hu WW, Wang J, Gao X, et al. GODSON-3: A scalable multicore RISC processor with X86 emulation. *IEEE Micro*, 2009, 29(2): 17—29
- Magen N, Kolodny A, Weiser U, et al. Interconnect-power dissipation in a microprocessor. In: *Proceedings of the 2004 International Workshop on System Level Interconnect Prediction*, 2004, 7—13
- Kumar R, Zyuban V, Tullsen DM. Interconnections in multi-core architectures: Understanding mechanisms, overheads and scaling. In: *Proceedings of 32nd International Symposium on Computer Architecture*, 2005, 408—419
- Hu WW, Zhao JY, Zhong SQ, et al. Implementing a 1GHz four-issue out-of-order execution microprocessor in a standard cell ASIC methodology. *Journal of Computer Science and Technology*, 2007, 22(1): 1—14
- 黄琨, 张戈, 王君. 基于物理反标的高性能处理器功耗建模与评估方法. *计算机辅助设计与图形学学报*, 2007, 11, 19(11): 1471—1475
- Li J, Martinez JF. Dynamic power-performance adaptation of

- parallel computation on chip multiprocessors. In: Proceedings of The 12th International Symposium on High-Performance Computer Architecture, 2006, 77—87
- 21 Li J, Martinez JF. Power-performance implications of thread-level parallelism on chip multiprocessors. IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2005), March 2005, 124—134
- 22 Li J, Martinez J, Huang M. The thrifty barrier: Energy efficient synchronization in shared-memory multiprocessors. In: Proceedings of International Symposium on High-Performance Computer Architecture (HPCA), 2004, 14—23
- 23 Kumar KI, Farkas NP, Jouppi P. et al. Single-ISA heterogeneous multi-core architectures: The potential for processor power reduction. In: Proceedings of the 36th Annual IEEE/ACM International Symposium on Microarchitecture, 2003, 81—92
- 24 Fan QF, Zhang G, Hu WW. A synchronized variable frequency clock scheme in chip multiprocessors. IEEE International Symposium on Circuits and Systems (ISCAS 08), May 2008, 3410—3413
- 25 Anoop Iyer, Diana Marculescu. Power and performance of globally asynchronous locally synchronous processors. In: Proceedings of ISCA, 2002, 158—168
- 26 William JD, Brian T. Route packets, not wires: On-chip interconnection networks. In: DAC'01: Proceedings of the 38th conference on Design automation, 2001, 684—689
- 27 Wang HS, Peh LS, Malik S. Power-driven design of router microarchitectures in on-chip networks. In: Proceedings of the 36th International Symposium on Microarchitecture, San Diego, CA: IEEE Computer Society, November 2003, 105—116
- 28 Kim JS, Taylor MB, Miller J et al. Energy characterization of a tiled architecture processor with on-chip networks. In: Proceedings of the International Symposium on Low Power Electronics and Design, 2003, 424—427
- 29 Liu C, Sivasubramaniam A, Kandemir M. Organizing the last line of defense before hitting the memory wall for CMPs. In: 10th IEEE Symposium on High-Performance Computer Architecture, Feb. 2004, 176—185
- 30 Wang HS, Peh LS, Jha N. Power-driven design of router microarchitectures in on-chip networks, In: International Symposium On Microarchitecture(MICRO'03). Nov. 2003, 105—116
- 31 Banerjee S, Dutt N. FIFO Power Optimization for On-Chip Networks. TR 03—40, December 19, 2003, 187—191
- 32 Shang L, Peh LS, Jha NK. Dynamic Voltage scaling with links for power optimization of interconnection networks. In: Proceedings of HPCA'03, Anaheim, CA. 2003, 91—102
- 33 Zhang M. Latency reduction techniques in chip multiprocessor cache systems. PhD Thesis at the Massachusetts Institute of Technology, January 2006
- 34 Wang J, Huang K, Zhang G, et al. Energy-efficient input buffer design using data-transition oriented model. International Symposium on Integrated Circuits (ISIC07), Sept. 2007, 244—247
- 35 Wang J, Zeng HB, Hu WW, et al. Zero-efficient buffer design for reliable network-on-chip in tiled chip-multi-processor. IEEE Design Automation and Test in Europe (DATE), Munich, Germany, 2008, 03, 792—795
- 36 Moreshet T, Bahar RI, Herlihy M. Energy-aware microprocessor synchronization: Transactional memory vs. locks. In: WMPI, 2006, 47—54
- 37 Hasegawa A, Kawasaki I, Yamada K, et al. SH3: High code density, low power. IEEE Micro, August, 1995, 15(16): 11—19
- 38 Inoue K, Ishihara T, Murakami K. Way-predicting set-associative cache for high performance and low energy consumption. International Symposium on Low Power Electronics and Design (ISLPED), August 1999, 273—275
- 39 Suh GE, Rudolph L, Devadas S. Dynamic partitioning of shared cache memory. Journal of Supercomputing, 2004, 28(1): 7—26
- 40 Moravan MJ, Bobba J, Moore KE, et al. Supporting nested transactional memory in LogTM. In: International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), October 2006, 359—370
- 41 Jin L, Lee H, Cho S. A flexible data to L2 cache mapping approach for future multicore processors. In: Proc. Workshop Memory Systems Performance and Correctness (MSPC), Oct. 2006, 92—101
- 42 Zeng HB, Wang J, Zhang G, et al. An interconnect-aware power efficient cache coherence protocol for CMPs. IEEE Parallel and Distributed Processing Symposium(IPDPS), 2008, 03