

# 片上多处理器互连技术综述<sup>\*</sup>

王 炜 汤志忠 乔 林

(清华大学计算机科学与技术系 北京 100084)

**摘 要** 随着器件、工艺和应用技术的不断发展,片上处理器中处理器核的数目必将进一步增加,处理器芯片内部的互连及其通信成为影响处理器性能的重要因素。介绍了目前在片上多处理器中的几种典型互连方法,并简要分析了各种方法的优缺点。

**关键词** 系统设计,片上多处理器,片上网络

## Summary of the Inter-connection in Chip Multiprocessor

WANG Wei TANG Zhi-zhong QIAO Lin

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

**Abstract** With the development of the devices, technics and application technologies, the amount on the CMP will increase more and more. The inter-connection and communication will be one of the most important factors which affect the performance of the processor. This paper introduces some typical methods in the inter-connection of CMP at present briefly, and indicates the characteristic of each method.

**Keywords** System design, CMP, NOC

随着计算机技术的不断发展,片上多处理器 CMP 逐渐成为主流。而且,随着器件、工艺和应用技术的不断发展,片上处理器中处理器核的数目必将进一步增加,从目前的几个、十几个发展到几十个甚至更多。当片上多处理器中核的数目增加的时候,各个处理器核以及处理器中其它节点(如 cache)之间的互连(即片上网络)及其通信就成为影响处理器性能的重要因素了。在片上多处理器中,要求片上网络提供高速度、高吞吐量、高带宽的互连通信,同时要求用于内部互联的开销要尽量小,也就是说占用芯片的面积小、功耗低。为了达到这些要求,有必要对原有的用于多处理机间的互连方法作出必要的改进。本文简要介绍了目前在片上多处理器中的几种典型互连方法,并简要分析了各种方法的优缺点。

## 1 片上多处理器内部互连的特点

片上多处理器的内部互连方法源于多处理机间的互连方法,但是片上多处理器内部互连有其自身的特点。

首先,由于片上多处理器内部各节点之间的距离相对短得多,而且数据传输的速度也相对快很多,因此片上多处理器内部互连的速度是首先需要考虑的一个重要因素。换句话说,片上多处理器内部互连时,延迟必须足够短,否则会严重影响处理器的性能。同时,由于连接的是处理器内部的节点,因此互连带宽也必须予以单独考虑。

其次,由于连接的对象处于单个芯片内部,因此许多互连方法因为这样或那样的原因不能用于片上多处理器内部互连。例如,超立方体互连将导致芯片内部布局困难、布线过于复杂,即使是将处理器芯片进行立体布局,由于可能导致散热困难,因此一般不考虑该方法。

再次,单芯片多处理器内部所需连线的线数可以根据需要增减,而不像多处理机之间,由于进行连接接口(例如

RJ45)的限制导致线数不易改变;另一方面,由于互连逻辑中的缓冲器需要占用一定的芯片面积,并且将导致大量的功耗,因此在进行片上多处理器内部互连设计的时候,应该争取尽量少地使用缓冲器,或者使得缓冲器的大小尽量地小。

最后,由于连接对象就在芯片内部,连接的距离也比较短,因此出现传输错误的概率也较低。而且,由于受外界的干扰相对较小,因此在设计片上网络的时候,可以较少地考虑传输安全的问题。

## 2 片上多处理器主要的内部互连方法

### 2.1 基于总线方法

总线互连是最简单的互连方法之一。但是由于总线中某一时刻只允许一个源数据,因此总线中存在严重的竞争,这也导致总线的速度较慢。为了缓解这一矛盾,出现了线性阵列的方法。线性阵列中,允许不同区段同时分段使用总线,只要这若干个区段之间没有交叉重叠。

如图 1 所示,IBM Cell 处理器即采用线性阵列方法实现处理器内部各个处理器之间的互连。

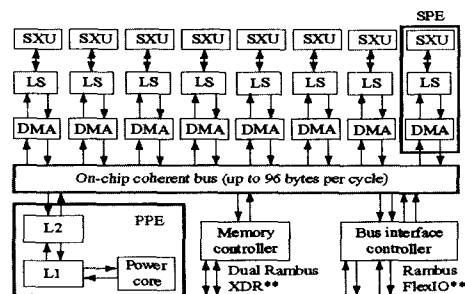


图 1 IBM Cell 处理器内部互连结构

<sup>\*</sup> 本文受国家自然科学基金项目(60573100、60773149)资助。王 炜 博士研究生,主要研究方向为片上多处理器、片上系统。

IBM Cell 处理器内部有 1 个主处理单元 PPE、8 个协处理单元 SPE。8 个协处理单元挂接在一个环状总线上,但是这条环状总线允许各个协处理单元分段使用总线。主处理单元和存储器、外部设备也挂接在这条总线上。

总线或者线性阵列的方法拓扑简单,可扩展性好,功耗低,但是由于延迟较大,竞争多,易于出现传输瓶颈。因此,当连接对象较多时,并不是一种非常好的方法。

### 2.2 环或带弦环互连

为了解决线性阵列直径大、延迟大的问题,出现了环状结构。环状结构中仍然允许分段使用,但是由于直径较大,仍然出现延迟较大的问题,于是出现了带弦环结构。当在环状结构中有选择地加入一些弦以后,随着弦数量的增加,网络直径显著减小,而且网络的吞吐量也明显增加。但是,随着网络中弦数的增加,网络中节点的度也增大了,它需要更多的连线资源和芯片面积,同时功耗也增加了。另外,网络中弦的数量增加也增加了处理器芯片内布局布线的复杂度。

为了解决上述问题,人们又提出了分级环状结构,其基本思想是将若干个节点连接到一个较小的环上,然后将若干个小环连接到一个较大的环上。以此类推,直到将所有节点连接到一起。分级环状结构中由于环中的节点数目较少,因此可以不需要弦,或者只需要少量的弦。又由于与其它环上的节点之间通过级联的方法进行通信,所以延迟也较小。

如图 2 所示,Intel 在 80 core 片上多处理器结构时即将分级环状结构作为可选择的方案之一。

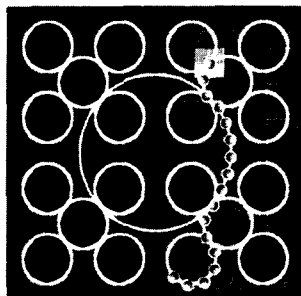


图 2 Intel 设想的一种 80 Core 处理器互连结构

在 Intel 设想的机遇环状互连的 80 core 处理器中,每 4 个处理器核通过环连接在一起,构成一个小环;每 4 个小环同样地连接,构成一个中环;最后将 4 个中环连接起来,构成一个大环。当图示中的一个处理器核要与另一个处理器核进行通信的时候,首先沿它所在的小环以最快的方式(由寻径算法决定)向上到达它所在的中环,然后以最快的方式向上到达大环;大环上根据目的处理器核的位置向下到达相应的中环,再由中环向下到达对应的小环,最后到达目的地。

这种分级的环状结构由于可扩展性好,同时延迟也较小,因此可能适合于较多处理器核之间的互连与通信。但是,这种单一的连接方式是否是最佳的选择,其实际的连接效果如何,还需要进一步研究和验证。

### 2.3 Mesh 网格互连

Mesh 网格将对象以十字网格的形式连接到一起。这种方法拓扑简单,寻径方便,而且可扩展性好,功耗也较小。但是由于所有节点在某一个方向上实际是以线性阵列的方式连接在一起的,因此网络直径较大,传输延迟也较大。为了解决这个问题,Mesh 网格做了很多改变,例如采用 Mesh 环网。

如图 3 所示,Trips 处理器内部采用了 Mesh 网格方式进行互连。

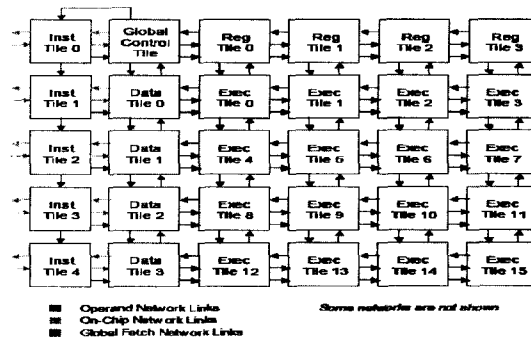


图 3 Trips 内部互连结构

在 Trips 内部,连接的对象并非处理器核一种,它把指令执行部件、数据收集部件、寄存器堆等分别置于一个 Tile 中,连接网络以网格的形式将这些 Tile 连接起来。由于执行部件之间、数据单元之间通过网格形式连接起来,因此,可以实现流水处理。

### 2.4 通过交叉开关互连

和前面的互连方法不一样,交叉开关通过硬件的方法,同时实现多对源和目的节点间的互连与通信,因此速度更快,吞吐量更大。但是,由于交叉开关方法功耗大,而且占用芯片面积也较大,因此这也是一种成本最高的方法。

图 4 所示的是 IBM Cyclops 64 系统中芯片内部的互连结构。

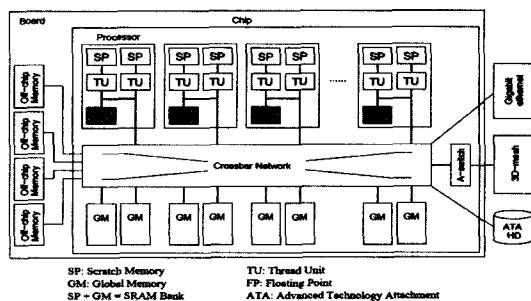


图 4 IBM Cyclops 64 芯片内部互连结构

在 IBM Cyclops 64 系统芯片中,通过一个 96×96 的交叉开关将片内的 80 个处理器、80 块分布式存储器以及片外的资源连接起来。显然这种方式下,片上网络部分将占用大量的资源,因此更多的设计者不主张采用规模过大的交叉开关。例如,在 SPIN 中,设计者通过 4×4 的交叉开关组成胖树结构,实现了 16 个节点间的互连。

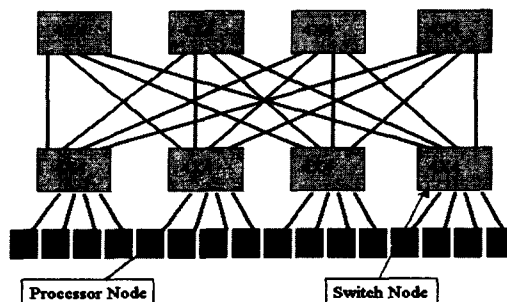


图 5 SPIN 互连结构

自动分配地址冲突问题。

(3) 互联时的 TCP 性能研究。TCP 流在 MANET 中的性能与在 Internet 中的表现是截然不同的, 在与 Internet 互联时性能恶化严重<sup>[10]</sup>。一方面拓扑的动态变化引起的丢包和重传导致了性能恶化, 另一方面拥塞控制窗口的大小在公平性与吞吐量之间很难均衡。文献[11, 12]提出了一些改善性能的方法, 对公平性及吞吐量有一定改善, 但还存在大量需要解决的难题。

(4) 混合网络的组播技术研究。MANET 中的组播技术是极其重要的功能之一, 与 Internet 互联状态下的组播协议互连和可靠组播尚无完善解决方案, 还需要进一步进行研究和改进。

(5) 网关优化配置研究。当节点规模较大、网关较多时, 优化网关位置配置成为关键问题, 需要研究如何减少或避免网关节点成为瓶颈, 把有限的带宽留给互联的数据转发。

此外, 服务质量(QoS)和安全性等问题都是自组网研究所普遍关注的, 也会在互联研究中受到关注和研究。

### 参 考 文 献

- [1] Engenstad P, Egeland G, Thanh D V. Analysis of NAT Based Internet Connectivity for Multi-Homed On-Demand Ad Hoc Networks//Proc. of Communication Networks and Distributed Systems Modeling and Simulation (CNDS 2004). January 2004
- [2] Jaewook H, Haeryong L, Jeehyeon N A, et al. Gateway Discovery and Routing in Ad Hoc Networks with NAT-based Internet Connectivity. IEEE, ETATS-UNIS. Piscataway, New

Jersey, 2004

- [3] McGee J, Karir M, Baras J S. Implementing Ad Hoc to Terrestrial Network Gateways. Wired/Wireless Internet Communications DOI 10. 1007/b95454, 2004
- [4] Sun Y, Belding-Royer E M, Perkins C E. Internet Connectivity for Mobile Ad Hoc Networks[J]. International Journal of Wireless Information Networks, 2002, 9(2)
- [5] Jonsson U, Alriksson F, Larsson T, et al. MIPMANET - mobile IP for mobile ad hoc networks [A]//Proc. MobiHOC2002 [C]. Boston, MA, USA, IEEE, 2000; 75-85
- [6] 姚尹雄, 王豪行. 一种基于中心代理节点的移动 Ad-Hoc 网络接入方法. 通信学报, 2001, 35(12): 13-16
- [7] 张凯捷, 向勇, 史美林. 移动自组网与 Internet 的多网关互连. 清华大学学报, 2007(1)
- [8] Ruiz P M, Gomez-Skarmeta A F, Groves I. The MMARP protocol for efficient support of standard ip multicast in mobile ad hoc access networks//Proceedings of the IST Mobile and Wireless Communications Summit 2003. Aveiro, Portugal, June 2003; 478-482
- [9] Ruiz P M, Galera F J, Jelger C, et al. Efficient Multicast Routing in Wireless Mesh Networks Connected to Internet//ACM International Conference Proceeding Series. vol. 138, 2006
- [10] Xu Kaixin, Bae Sang, Sungwook L, et al. TCP behavior across multihop wireless networks and the wired internet // WoW-MoM'02. September 2002
- [11] Yang Luqing, Seah W K G, Yin Qinghe. Improving fairness among TCP flows crossing wireless ad hoc and wired networks//MobiHoc'03. June 2003
- [12] ElRakabawy S M, Klemm A, Lindemann C. MANETs-Gateway adaptive pacing for TCP across multihop wireless networks and the Internet//MSWiM'06. October 2006

(上接第 8 页)

当然, 由于采用了分级互连的方法, SPIN 中的网络延迟将会比直接采用交叉开关要大, 但是这种方法将有效地降低片上网络所需的芯片面积和功耗。

**结束语** 目前, 随着处理器技术的深入发展, 片上网络的研究也不断走向深入。未来的片上网络, 应该是面向几十个、几百个节点的、易扩展、高速度、高吞吐量、低开销的。但是, 要想达到这些目标, 目前还存在一些困难。

一个困难来自性能模型。如何评价、量化测量互连结构的性能? 如果说单个的技术指标, 例如延迟、带宽、吞吐量等等, 早就有成熟的方法了, 但是如果把所有的因素都考虑进来呢? 特别是, 当我们考察某一种互连结构时, 我们需要的是整体性能, 那么如何对各个技术指标进行权衡? 还有, 随着技术的发展, 新的技术指标或者新的影响性能的因素是哪些? 如何衡量? 如何取舍?

另一个困难来自应用模型。虽然说用户对提高处理器性能的要求总是无止境的, 但是, 目前似乎很难找到更多的能用于几百个甚至上千个处理器核的应用。当任务被分配到几百个、上千个处理器核进行处理以后, 获得的性能似乎并不像我们期望的那样好, 甚至还不如不用那么多处理器核。应用模型的困难其实也影响了人们对于更大规模处理器结构研究的开展。

当然, 不论如何困难, 对于处理器技术的探索仍然在继续, 对于片上多处理器内部互连技术的研究也还在继续着。也许, 在不久的将来, 人们会看到更多更好的技术, 给社会带来更多便利。

### 参 考 文 献

- [1] Pande P P, et al. Performance Evaluation and Design Trade-offs for Network-on Chip Interconnect Architectures. IEEE Transactions on Computers, 2005, 54(8): 1025-1040
- [2] Ye T T, et al. Packetization and Routing Analysis of On-chip Multiprocessor Networks. Journal of Systems Architecture, 2004, 50(2): 81-104
- [3] Kahle J A, et al. Introduction to the cell multiprocessor. IBM J. RES. & DEV, 2005, 49(4/5): 589-604
- [4] Manjikian N. Prototyping a Hierarchical Ring Interconnect for System-on-Chip Multiprocessor//The 2nd Annual IEEE Northeast Workshop on Circuits and Systems (NEWCAS 2004). 2004; 85-88
- [5] Karim F, et al. On-chip Communication Architecture for OC-768 Network Processors//Proceedings of 38th Design Automation Conference. June 2001; 678-683
- [6] Shrout R. Intel's 80 Core Terascale Chip Explored; 4GHz clocks and more. <http://www.pcpur.com>, Feb. 2007
- [7] Dally W, et al. Route Packets, Not Wires; On-Chip Interconnection Networks//Proceedings of the 38th Design Automation Conference. June 2001; 684-689
- [8] Kumar S, et al. A network on chip architecture and design methodology//Proceedings of IEEE Computer Society Annual Symposium on VLSI. April 2002; 105-112
- [9] McDonald R, et al. TRIPS Processor Reference Manual. Version 1. 2. March 2005; 10
- [10] Guerrier P, et al. A Generic Architecture for On-chip Packet-switched Interconnections//Proceedings of Design Automation and Test in Europe. March 2000; 250-255
- [11] Zhang Ying Ping, et al. A Study of the On-chip Interconnection Network for the IBM Cyclops64 Multi-core Architecture//20th International Parallel and Distributed Processing Symposium, 2006 (IPDPS 2006). April 2006