

分级环片上网络互连

王 炜^{1),2)} 乔 林¹⁾ 杨广文¹⁾ 汤志忠¹⁾

¹⁾(清华大学计算机科学与技术系 北京 100084)

²⁾(解放军信息工程大学信息工程学院计算机科学与技术系 郑州 450002)

摘 要 在大规模、超大规模片上互连网络中,因为二维互连方式的性能较差而使多维互连方式成为可选方案之一.文中首先基于区域划分设计了一种分级环互连结构,分析了其静态互连特性,然后基于卡诺图编码设计了一种分级环互连的路由结构以及寻径方法,在均匀通信模式测试了不同的分级环级联链路缓冲区设置方法下网络的性能,详细分析了按照等比序列设置分级环级联链路缓冲区时分级环互连方式的动态网络特性,最后根据互连性能与 Mesh 等二维片上互连方式比较的结果,给出了分级环互连方式的使用场合.实验结果表明,虽然在较小规模网络中性能较差,但是分级环互连方式能以较低的成本、较高的性能实现大规模、超大规模片上网络的互连,其中单环分级互连方式在较低网络负载下综合性能更好,而双环分级互连方式则具有更大的网络负载能力,在较高网络负载下性能更好.

关键词 片上多处理器;片上网络;分级环;性能分析

中图法分类号 TP393

DOI 号: 10.3724/SP.J.1016.2010.00326

A Kind of Hierarchical Ring Interconnection Networks-on-Chip

WANG Wei^{1),2)} QIAO Lin¹⁾ YANG Guang-Wen¹⁾ TANG Zhi-Zhong¹⁾

¹⁾(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

²⁾(Department of Computer Science and Technology, Institute of Information Engineering,
PLA Information Engineering University, Zhengzhou 450002)

Abstract The interconnection performance of the 2-D Networks-on-Chip is so poor in the large and very large scales that the multi-dimensional interconnections become one of the choices to achieve better interconnection performance. This paper gives a design of a kind of hierarchical ring Networks-on-Chip based on the regional partition and analyses the static characteristics of the hierarchical rings. Then it proposes a kind of the hierarchical ring router and the routing algorithm based on the Karnaugh map coding. It tests the characteristics with different sets of the buffers in the hierarchical router nodes under the uniform random communication pattern and gives a detailed analysis of the dynamic characteristics of the hierarchical rings with the buffers in the hierarchical router nodes set in accordance with the geometric sequence under the global uniform random communication pattern. Finally, it points out the cases when the hierarchical rings work well according to the comparison of the characteristics of the hierarchical ring interconnections and some of the 2-D Network-on-Chip interconnections such as Mesh. The results show that, although working poor in small scales, the hierarchical rings can interconnect the large and very large scales Networks-on-Chip with lower cost and higher performance. Comparatively, the sin-

收稿日期:2009-07-12;最终修改稿收到日期:2009-09-04. 本课题得到国家自然科学基金(60773149)、国家“八六三”高技术研究发展计划项目基金(2006AA01A101,2008AA01Z108)及国家“九七三”重点基础研究发展规划项目基金(2007CB310900)资助. 王 炜,男,1975年生,博士研究生,讲师,主要研究方向为计算机系统结构、片上多处理器与片上网络. E-mail: wwwei-05@mails. tsinghua. edu. cn. 乔 林,男,1972年生,博士,副教授,主要研究方向为计算机系统结构、并行编译与优化技术. 杨广文,男,1963年生,教授,博士生导师,主要研究领域为计算机系统结构、并行与分布计算. 汤志忠,男,1946年生,教授,博士生导师,主要研究领域为计算机系统结构、指令级并行编译技术.

gle-hierarchical ring interconnection works better under lower loads and the double-hierarchical ring interconnection has larger load capability and works better under higher loads.

Keywords chip multiprocessor; networks-on-chip; hierarchical ring; performance analysis

1 引言

随着计算机技术的不断发展,片上多处理器(Chip Multiprocessor, CMP)逐渐成为主流, CMP 中互连网络(即片上网络 Networks-on-Chip, NoC)及其通信逐渐成为影响系统性能的重要因素. 一般来说,高吞吐量、低延迟为片上网络追求的重要目标^[1-2],而由于物理空间的限制,布局布线复杂度、功耗、芯片面积也是片上网络设计时需要考虑的重要因素^[2],因此要求片上网络使用尽量简单的物理结构、通信协议和控制(路由)机制.

基本的二维 NoC 主要有总线与线性阵列、Mesh 网格、环与带弦环等及其变形结构. 文献[1]简要分析了总线与线性阵列的网络特性,文献[3-4]基于全局均匀随机模型分别分析了如图 1 所示的几种片上二维网络的性能,我们另一篇文章则基于局部均匀随机模型分析了这些结构的性能. 分析结果表明,在网络规模较大时单独直接使用这些结构均不能高效、快速地实现 NoC 的互连^[1,3-4],因此,需要考虑新的互连方法,例如将基本的二维 NoC 向高维扩展、组合使用多种基本拓扑结构以及设计新型互连结构等.

本文首先基于区域划分设计了一种分级环互连结构,分析了其静态互连特性,然后基于卡诺图编码设计了一种分级环互连的路由结构以及寻径方法,基于均匀通信模式测试了不同的级联链路路由结点缓冲区设置方法下网络的性能,详细分析了按照等比序列设置级联链路路由结点缓冲区时分级环互连方式的动态网络特性,最后根据互连性能与 Mesh 等二维片上互连方式比较的结果,给出了分级环互连方式的使用场合.

2 分级环互连结构

最基本的环状互连是线性阵列互连的一种改进,它将线性阵列首尾两个端点连结起来,有效降低了网络直径^[1,5]. 但如结点数目较多,网络直径仍然比较大.

一种降低网络直径的方法是使用多环. 文献[3]分析了如图 1(d)和(e)所示的多环相切直连和多环相切的类 Torus 结构的性能. 实验数据表明:这种多环相切方式的互连效率并不理想^[3].

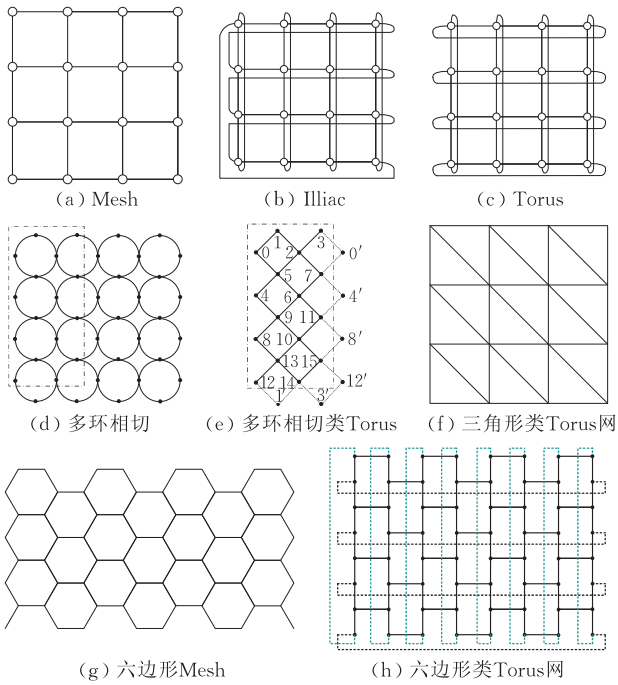


图 1 二维 Mesh 及其部分变形结构

另一种降低环网网络直径的方法是使用带弦环^[3,5]. 随着弦数量的增加,带弦环网络直径显著减小,而且网络吞吐量也明显增加. 但是,随着弦的增加,网络结点的度也增大,需要更多的连线资源和芯片面积,功耗也相应增加. 另外,弦数量的上升也增加了布局布线的复杂度.

因为在很多情况下,通信往往具有局部性:只有少数通信需要在相邻较远的结点间进行,更多的通信是在相邻较近的结点间进行^[6-8]①,这就意味着可以将局部区域的结点连接到一起,然后通过逐级互连从而构成分级结构^[1,9],为此将若干个结点连接到一个较小的环上(我们称这些小环构成的网络为第 1 级网络,以此类推),然后将若干个小环连接到一个较大的环上,以此类推,直到将所有结点连接到

① NAS Parallel Benchmarks: ProActive implementations. http://proactive.inria.fr/index.php?page=nas_benchmarks, 2007

一起. 这种结构中由于环上的结点数目较少, 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结点之间通过级联的方法进行通信, 所以延迟也较小, 对于一个 $n \times n (n = 2^r, r \geq 2)$ 网络, 平均寻径距离为 $2r - 1$, 等分带宽均为 2.

假设每 α 个结点连接成一个环, 每 α 个较小的环连接成一个较大的环, 则完成网络连接需要将网络分成 $2\log_\alpha n$ 级, 一共需要 $n^2 + \sum_{i=1}^{2\log_\alpha n-1} \frac{n^2}{\alpha^i}$ 条链路, 其中用于级联的链路共 $\sum_{i=1}^{2\log_\alpha n-1} \frac{n^2}{\alpha^i}$ 条. 当 $\alpha = 4$ 时即构成了

如图 2(a) 所示的分级环互连结构. 为便于后面的寻径操作, 对结点进行卡诺图编码. 一个 n 位码字的完整卡诺图编码序列共有 2^n 个编码, 而且任意两个相邻的编码之间有且仅有一个码字不同. 对所有结点的纵横坐标均进行卡诺图编码后, 如图 2(a) 所示的分级环互连结构可以描述为:

对于任何一点 $S(S_x, S_y)$, 令 $S_x = x_k \cdots x_2 x_1$, $S_y = y_k \cdots y_2 y_1$, $x_0 = y_0 = 1$, 则当且仅当 $x_i = y_i = 1$, $i = 0, 1, 2, \cdots, p - 1$ 时有点 $D(D_x, D_y)$ 与 S 相连, 其中

$$\begin{cases} D_x = x_k \cdots x_{p+1} \overline{x_p} x_{p-1} \cdots x_2 x_1 \\ D_y = y_k \cdots y_2 y_1 \end{cases}$$

或

$$\begin{cases} D_x = x_k \cdots x_2 x_1 \\ D_y = y_k \cdots y_{p+1} \overline{y_p} y_{p-1} \cdots y_2 y_1 \end{cases}$$

若某一结点处在第 i 级网络中, 那么第 $j (0 < j < i)$ 级网络必有且仅有一个环通过该结点. 称与之相连的链路最高处于第 k 层网络的结点为第 k 层结点, 则第 $k (0 < k < 2\log_\alpha n)$ 层结点共有 $\frac{n^2(\alpha - 1)}{\alpha^k}$ 个, 其结点度为 $2k$; 第 $2\log_\alpha n$ 层结点共有 α 个, 其结点度为 $4\log_\alpha n$.

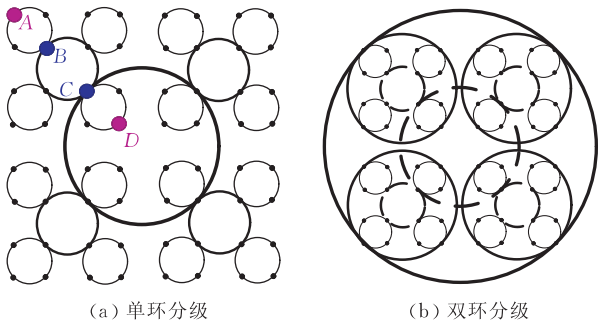


图 2 分级环互连

在如图 2(a) 所示的分级环互连结构中, 因为有级联环的存在, B 和 C 之间的寻径距离近了, 但是 A 和 D 之间的寻径距离仍然较远. 当然, 也可以使级

联环经过 A 和 D 而不经 B 和 C , 此时的网络互连关系是这样的:

对于任何一点 $S(S_x, S_y)$, 令 $S_x = x_k \cdots x_2 x_1$, $S_y = y_k \cdots y_2 y_1$, $x_0 = y_0 = 0$, 则当且仅当 $x_i = y_i = 0$, $i = 0, 1, 2, \cdots, p - 1$ 时有点 $D(D_x, D_y)$ 与 S 相连, 其中

$$\begin{cases} D_x = x_k \cdots x_{p+1} \overline{x_p} x_{p-1} \cdots x_2 x_1 \\ D_y = y_k \cdots y_2 y_1 \end{cases}$$

或

$$\begin{cases} D_x = x_k \cdots x_2 x_1 \\ D_y = y_k \cdots y_{p+1} \overline{y_p} y_{p-1} \cdots y_2 y_1 \end{cases}$$

为了进一步减少结点间的寻径距离, 可以使用两套级联环, 如图 2(b) 所示分别将类似 A 和 D 、 B 和 C 的结点级联起来, 这样就构成了双环分级互连方式, 其等分带宽为 4. 使用这种双环分级互连方式连接上述 $n \times n (n = 2^r, r \geq 2)$ 网络, 平均寻径距离为 r . 仍然假设每 α 个结点连接成一个环, 每 α 个较小的环连接成一个较大的环, 则非级联链路数不变, 级联链路数变为单环分级互连方式的 2 倍, 因此共需链路 $n^2 + 2 \sum_{i=1}^{2\log_\alpha n-1} \frac{n^2}{\alpha^i}$ 条; 每一层结点的度也不变, 只是各层结点的数目发生变化: 第 $k (0 < k < 2\log_\alpha n)$ 层结点共有 $\frac{n^2(\alpha - 2)}{\alpha^k}$ 个, 而第 $2\log_\alpha n$ 层结点共有 2α 个.

3 分级环互连路由与寻径方法

和文献[3-4]一样, 我们仍然采用虫蚀寻径以包为单位分片传输数据, 在不阻塞的情况下, 每个时钟周期数据片向前传送一跳, 并使用超时丢弃策略解决死锁或活锁问题[3-4].

3.1 分级环互连路由结构

多级环互连中, 仅仅处于第 1 级环上的路由结点只需要连接其第 1 级环上的“左”、“右”路由结点和本地结点, 因此其内部路由结构与文献[3-4]中讨论的路由结构相同即可. 值得注意的是, 与文献[3-4]中讨论的一样, 由于路由结点除了连接其它路由结点外, 还需要连接到本地结点, 因此, 所有路由结点(包括下面讨论的处于其它层上的路由结点)的链路数应该在前面讨论的基础上加 1[3].

多级环互连中级联点上的路由结点需要完成多维互连, 因此级联点的路由结构与文献[3-4]中所使用的路由结构有所不同. 图 3 所示的是多级环的路由结点及其连接, 其中图 3(a) 的路由结点 C 就是图 2(a) 中的路由结点 C , 图 3(b) 给出的是路由结点 C 内部的结构.

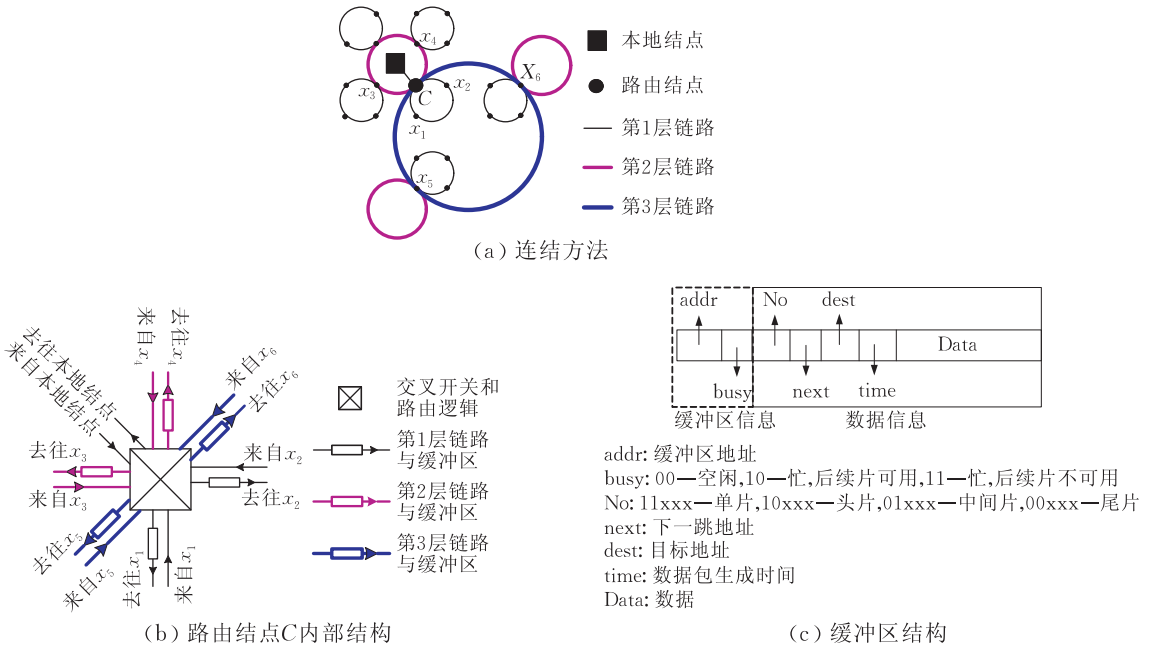


图 3 多级环网路由结点及其连接

为了降低网络功耗,减少片上网络的芯片面积并提高网络传送速度,仍然采用出端缓冲片上路由结构^[3],所有结点间连接均通过一入一出两条单向链路完成,结点内部则通过交叉开关实现同时多对链路相连.与本地结点连接的链路不设置缓冲区,数据缓冲由本地结点内部完成.与其它路由结点相连的链路上采用出端缓冲的方式,下一时刻当前方不阻塞时,被传送的数据片将由下一路由结点的交叉开关直接送到某出端缓冲区.缓冲区结构如图 3(c)所示,与文献[3-4]中一样,仍然是将所有出端缓冲区和本地结点统一编址.

一般而言,层次互连结构中的级联链路将承担比非级联链路更多的通信传送,而且级联层次越高,链路上的通信越多.为了避免级联链路成为系统的瓶颈,需要增强级联链路的通信传送能力.

最简单而且有效的方法是如图 3(a)所示增加级联链路的带宽,相应地需要增加不同级联链路的出端缓冲区的个数,这就意味着较高层路由结点中需要更大规模的交叉开关和寻径逻辑.交叉开关的硬件复杂度为 $O(n^2, \omega)$, 其中, n 为输入/输出数量, ω 为带宽^[10-12].而且当 n 超过 16 以后,不仅交叉开关过于复杂,而且不能保证在一拍之内完成输入/输出之间的互连^[5],因此,不能将级联链路/缓冲区数量设置过大.为此考虑 4 种级联链路/缓冲区设置模式:

模式 A:所有级别均设置 1 个链路/缓冲区,在 $4 \times 4, 8 \times 8, 16 \times 16, 32 \times 32$ 网络里交叉开关最大规

模分别为 $5 \times 5, 7 \times 7, 9 \times 9, 11 \times 11$;

模式 B:不同级按照等差序列设置,从低到高各层分别设置 1、2、3、4…个链路/缓冲区,在 $4 \times 4, 8 \times 8, 16 \times 16, 32 \times 32$ 网络里交叉开关最大规模分别为 $7 \times 7, 13 \times 13, 21 \times 21$ 和 31×31 , 其中 $16 \times 16, 32 \times 32$ 网络里分别有 4 个和 16 个规模大于 16×16 的交叉开关;

模式 C:不同级按照等比序列设置,从低到高各层分别设置 1、2、4、8…个链路/缓冲区,在 $4 \times 4, 8 \times 8, 16 \times 16, 32 \times 32$ 网络里交叉开关最大规模分别为 $7 \times 7, 15 \times 15, 31 \times 31, 63 \times 63$, 其中 $16 \times 16, 32 \times 32$ 网络里分别有 4 个和 16 个规模大于 16×16 的交叉开关;

模式 D:不同级按照本级 $1/4$ 区域结点规模设置,从低到高各层分别设置 1、4、16、64…个链路/缓冲区,在 $4 \times 4, 8 \times 8, 16 \times 16, 32 \times 32$ 网络里交叉开关最大规模分别为 $11 \times 11, 43 \times 43, 171 \times 171, 683 \times 683$, 其中规模超过 16×16 的交叉开关分别有 0 个、4 个、16 个和 64 个.

模式 D 方式下在 8×8 及其以上规模网络中均有为数不少的交叉开关过于复杂,因此不予考虑;而模式 B 和模式 C 方式下只有 16×16 和 32×32 网络中有极少数比较复杂的交叉开关,其所占的比例较小,可以考虑.后面的实验数据表明,模式 C 方式下网络性能基本满足需要,可以不考虑更复杂的设置方法.

表 1 分别给出了 $\alpha=4$ 时分级环互连以及其它

互连结构的部分静态性能。显然,在较复杂的交叉开关处的寻径与数据传送时间将有所延长,但是可以

采取某些方法使得仍然比结点间的延迟相比小得多^[13],为简便计暂时不考虑。

表 1 分级环互连($\alpha=4$ 时)部分静态网络特性与其它互连方式的比较

规模	结构	等分带宽	网络直径/ 平均寻径距离	平均结点度/ 最大结点度	各种分层链路数/缓冲区数设置下的成本					
					每层一个		等差序列		等比序列	
					链路/缓冲区	交叉开关	链路/缓冲区	交叉开关	链路/缓冲区	交叉开关
4×4	Tri-Torus	16	2 / 1. 6	6 / 6	48	784	—	—	—	—
	Mesh	4	6 / 2. 67	3 / 4	24	264	—	—	—	—
	Illiac	8	3 / 2	4 / 4	32	400	—	—	—	—
	H-ring(s)	2	6 / 2. 93	2. 5 / 4	20	208	24	304	24	304
	H-ring(d)	4	4 / 2. 33	3 / 4	24	272	32	464	32	464
8×8	Tri-Torus	32	6 / 3. 17	6 / 6	192	3136	—	—	—	—
	Mesh	8	14 / 5. 33	3. 5 / 4	112	1320	—	—	—	—
	Illiac	16	7 / 4	4 / 4	128	1600	—	—	—	—
	H-ring(s)	2	10 / 4. 76	2. 56 / 6	84	928	108	1696	112	1920
	H-ring(d)	4	6 / 3. 73	3. 25 / 6	104	1280	152	2816	160	3264
16×16	Tri-Torus	64	14 / 6. 46	6 / 6	768	12544	—	—	—	—
	Mesh	16	30 / 10. 67	3. 75 / 4	480	5832	—	—	—	—
	Illiac	32	15 / 8	4 / 4	512	6400	—	—	—	—
	H-ring(s)	2	14 / 6. 70	2. 66 / 8	340	3840	448	7872	480	10624
	H-ring(d)	4	8 / 5. 44	3. 31 / 8	424	5376	640	13440	704	18944
32×32	Tri-Torus	128	30 / 13. 11	6 / 6	3072	50176	—	—	—	—
	Mesh	32	62 / 21. 33	3. 88 / 4	1984	24456	—	—	—	—
	Illiac	64	31 / 16	4 / 4	2048	25600	—	—	—	—
	H-ring(s)	2	18 / 8. 68	2. 66 / 10	1364	15520	1760	33568	1840	54528
	H-ring(d)	4	10 / 7. 18	3. 33 / 10	1704	21824	2496	57920	2656	99840

数据显示,分级环互连方式的等分带宽较小,而且等分带宽与网络规模无关,这表明分级环互连方式的性能瓶颈在级联链路上,而且网络规模越大,瓶颈作用越明显。

4×4 网络中,分级环互连方式的网络直径和平均寻径距离与 Mesh 方式相当,但是随着网络规模的增加,分级环互连方式的网络直径和平均寻径距离相比较其它几种结构越来越小,在 32×32 网络中甚至低于三角形网眼环网,这表明,分级环互连方式更适合较大规模网络。

和 Mesh 网相似,随着网络规模的增加,分级环互连方式的平均结点度逐渐上升,但是与 Mesh 网中最大结点度保持不变不同,分级环互连方式的最大结点度也不断增加,但是即使在 32×32 网络中分级环互连方式的平均结点也低于 Mesh 方式,这表明分级环互连方式中不同结点的度差异更大。

在各种规模网络中,无论联结链路结点的链路和缓冲区按照上述何种方式设置,单环分级互连方式所需的链路和缓冲区数都不多于 Mesh 方式、双环互连方式则不多于三角形网眼环网方式,表明分级环互连方式在链路和缓冲区上所需的成本低于对应的互连结构,这就使得分级环互连方式的布局布线复杂度较低,而且系统功耗较低、网络传送速度较快。另一方面,在网络规模较小时分级环互连方式用于交叉开关的总成本不高于对应的互连结构,但是

随着网络规模的增加,当使用等差或等比序列模式设置级联链路有结点链路/缓冲区时,分级环中用于交叉开关的总成本逐渐超过对应的互连结构,这主要是因为随着规模的增加,较高层联结链路所需交叉开关过于复杂的缘故,而那少数几个特别复杂的交叉开关花费了交叉开关总成本的相当部分,当然这个问题也可以通过某些技术予以解决^[13]。

值得注意的是,双环互连方式相比单环方式虽然多了一套级联链路,但是其平均寻径距离的优势并不明显,而且该优势在所有规模网络下几乎保持不变,而其所需的链路数/缓冲区数与用于交叉开关的总硬件成本却远大于单环方式,这表明,整体相对而言双环方式并不优于单环方式。

综上对网络静态特性的分析可知,通过合理分配级联链路宽度和缓冲区数量,在较大规模网络中分级环互连方式特别是单环分级互连方式将获得更小的网络延迟、更少的互连链路与更低的网络布局布线复杂度、更低的功耗,而这些是以局部更复杂的交叉开关和整体更高的交叉开关成本作为代价的。

3.2 基于卡诺图编码的分级环寻径方法

多级环互连仍使用同文献[3-4]一样的通信协议,寻径操作由寻径逻辑根据包头中目标地址自动完成。寻径逻辑在进行寻径操作时,总是根据当前包头结点的位置和目的结点的位置来确定下一条的位置。进行卡诺图编码后,若两个结点的横坐标、纵坐

标的最高位分别相等,则两个结点同属于一个 $1/4$ 区域,若两个结点的横坐标、纵坐标的次高位也分别相等,则两个结点同属于上述 $1/4$ 区域中的一个小 $1/4$ 区域,以此类推.当两个结点同属于一个 $1/4$ 区域时,只需要在该 $1/4$ 区域内寻径即可,相当于寻径时将那些分别相等的较高位地址舍弃,从而降低了寻径的维度.

若两个结点的横坐标或纵坐标的最高位不相等,则两个结点不属于一个 $1/4$ 区域,此时需要判断在当前结点所在的 $1/4$ 区域内实现两个 $1/4$ 区域级

联的路由结点的位置.若此时的结点恰好是那个级联结点,只需要将当前的(也即“舍弃了”部分高位码字的)纵横坐标中最高位与目的结点最高位不同的那一个(如果两个都不相同,任取其一)取反然后拼接上前面“舍弃”的部分高位码字即得到了下一跳的坐标;若此时的结点不是那个需要的级联结点,则将那个需要的级联结点作为暂时的目的结点,重新进行网络寻径操作,直到找到下一跳的坐标.

这种基于卡诺图编码的路由寻径方法快速、有效而且控制逻辑简单,很容易在片上实现.

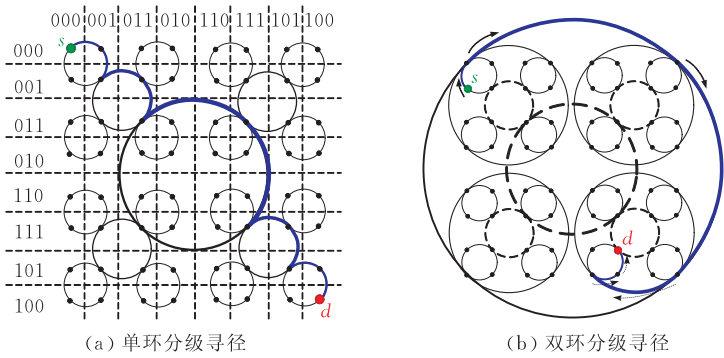


图 4 基于卡诺图编码的分级环寻径

图 4 分别给出了单环和双环分级互连中使用上述寻径方法的一条寻径线路.需要注意的是,双环分级中实现不同 $1/4$ 区域级联的点有两个,分别位于内环和外环上,因此在寻径的时候需要根据源和目的的位置关系做出选择.

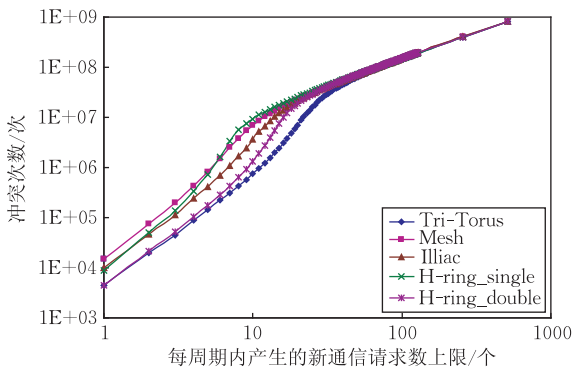
4 分级环互连网络特性

本文同文献[3-4]一样,通过对同一通信请求序列在不同结构网络中的模拟传输比较分级环互连结构与其它互连结构的网络性能.另外的实验结果表明,全局均匀随机通信模式是局部均匀随机通信模式的特例,全局均匀随机通信模式下不同结构网络

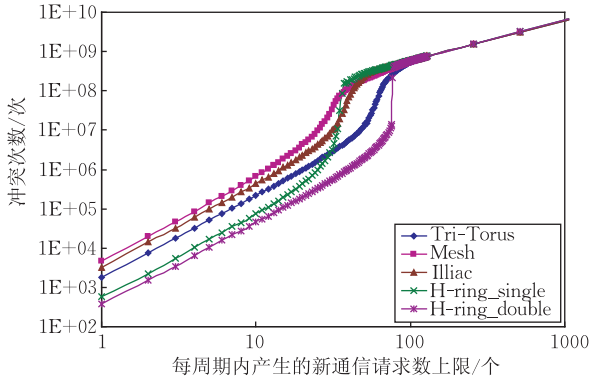
性能之间的相互关系能够完全代表局部均匀随机通信模式下的相互关系,因此,下面的实验中主要讨论全局均匀随机通信模式下的网络性能.另外,文献[3-4]的实验数据显示,相比较六边形网眼 Mesh 及其变形结构、多环相切及其回绕结构而言,四边形、三角形网眼 Mesh 网络及其变形结构具有更好的规模扩展性和网络负载能力,同时网络综合性能也较好,而 Torus 结构和 Illiac 结构的网络性能非常近似,因此,下面的实验主要将分级环互连结构与 Mesh、Illiac 和三角形网眼 Mesh 结构做比较.

4.1 网络冲突与通信平均延迟

图 5 和图 6 分别给出了不同规模各种结构的网络冲突总数和网络通信平均延迟.实验数据显示,当



(a) 8×8 网络



(b) 32×32 网络

图 5 网络冲突总数

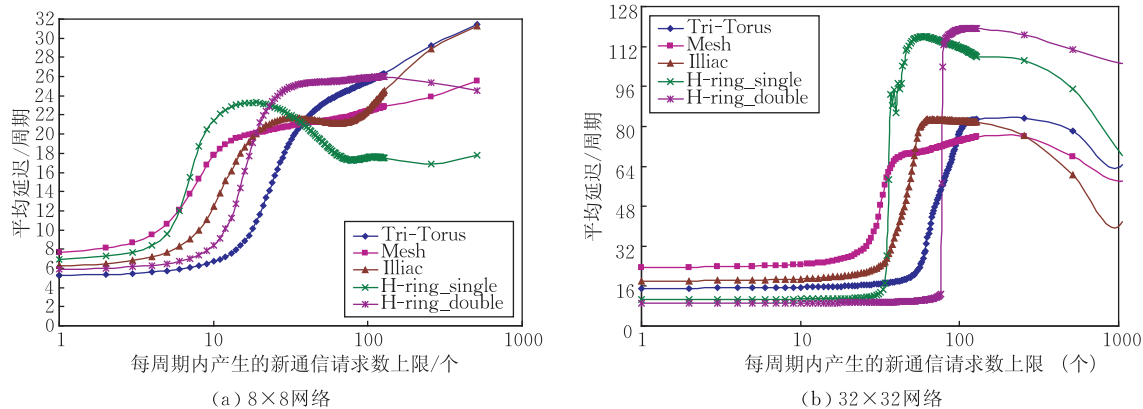


图 6 网络通信平均延迟

网络规模较小(包括 8×8)时,单环分级互连方式下网络的冲突总数不低于 Illiac 方式,而双环分级互连方式不低于三角形网眼环网方式,对应地,它们的网络通信平均延迟也分别不低于其它两种方式;但是当网络规模增大时,分级环互连方式下的网络冲突总数变得比三角形网眼环网方式还要少,相应地,其网络通信平均延迟也最小.这表明,相对而言,分级环互连方式更适合于较大规模网络的连接与通信.

两种分级环互连方式中,单环比双环方式网络冲突总数多而网络通信平均延迟高,这显然是因为单环分级方式的链路数较少的原因.但是,在较低网

络负载下二者性能相差不大.

4.2 网络通信传送完成率与负载能力

图 7 所示的是不同规模各种结构网络通信传送完成率的情况.数据显示,单环分级互连方式的通信传送完成率低于双环方式,但是在较小规模下相差不大.在较小规模中,单环分级互连方式的通信传送完成率最低,但随着网络规模的增加,其通信传送完成率逐渐变得与 Mesh 结构的相当;而双环互连方式下的通信传送完成率则从低于三角形网眼环网方式到超过三角形网眼环网方式.这同样表明分级环互连方式更适合于较大规模网络.

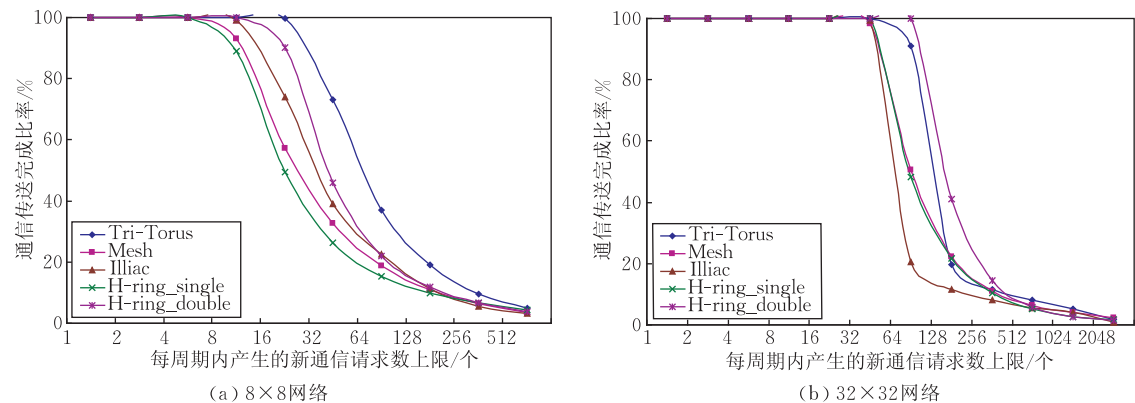


图 7 网络通信传送完成率

如文献[3-4]一样,本文使用平均每一时刻网络中允许的新通信请求的最大值来表示网络负载,并定义网络通信传送完成率不低于 95%时网络所能承受的最大负载为**网络的理想负载**,它同时也是平均延迟随着网络负载增加而增加率最大时的网络负载;定义使网络通信传送完成率不低于 80%时网络所能承受的最大负载为**网络的有效负载**,它同时也是网络通信传送完成率随着网络负载增加而降低率最大时的网络负载.图 8 给出了各结构的理想负载

能力.

数据显示,当网络规模逐渐增加时,单环分级互连方式的理想负载逐渐从低于 Mesh 方式增加到高于 Mesh 方式,但总体上都与 Mesh 方式相当,而双环方式的则逐渐从远低于三角形网眼环网方式增加到高于三角形网眼环网方式,这表明,分级环互连方式更适合较大规模网络,而且其负载能力的规模扩展性要好于其它结构.另外,单环方式的负载能力低于双环方式,而且随着网络规模的增加,二者负载能

力的差别越来越大,这表明,相对而言双环方式更适合于高通信负载网络的互连.

图 9 给出的是各种规模下所有网络均工作在 Mesh 网的理想负载下时网络平均延迟的情况. 图中显示,在较小规模网络中,单环分级互连方式的平均延迟大于 Mesh 方式而双环方式则稍大于三角形网眼环网方式,但是在较大规模网络中,它们分别低于对应的互连结构,而且网络规模越大,其差距越大,这进一步表明分级环互连方式相比较而言更适于较大规模网络的互连,而且网络规模越大,其相对性能提升越多.

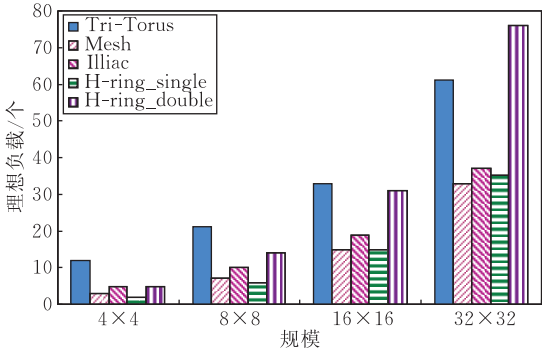


图 8 不同片上网络的理想负载能力

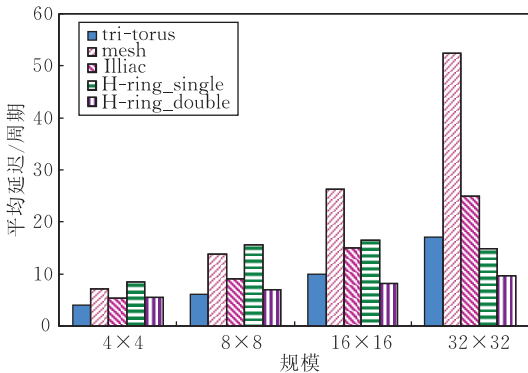


图 9 工作在 Mesh 理想负载下的网络平均延迟

5 结论和下一步的工作

本文设计了基于二维区域划分的分级环片上网络互连结构和基于卡诺图编码的分级环路由结构与寻径方法,面向全局均匀随机通信模式详细分析了分级环互连的网络特性.

与 Mesh 等通过二维平面方式进行互连不同,分级环通过级联方式将网络中的结点按区域逐层连结起来,不仅节省了链路,而且降低了网络直径和平均寻径距离. 在分级环互连中,级联链路成为影响网络性能的主要因素,实验数据表明,只需按照等比序

列设置各层路由结点中的缓冲区数量即可以获得较好的网络性能. 由于增加的链路和缓冲区较少,分级环方式互连成本仍然较低.

由于网络规模较小时与 Mesh 等二维互连方式相比分级环互连方式节省的互连成本有限、网络直径以及平均寻径距离的减少有限,而级联方式又一定程度上增加了网络冲突产生的可能,因此在较小规模下分级环方式的互连性能较差;随着网络规模的增加,分级环互连方式下网络直径和平均寻径距离的减小越来越明显,逐渐抵消了由于级联而导致的网络冲突增加的因素,从而使得分级环方式的互连性能越来越好,这表明,分级环互连方式更适合大规模超大规模片上网络.

特别指出的是,以上结果是基于全局均匀通信模式分析得到的,基于局部均匀通信模式的实验数据表明,局部通信概率越高,分级环互连方式的相对性能越好.

与单环分级互连相比,双环分级互连具有更大的网络负载能力和更低的单位延迟负载能力,因此当网络负载较大时,双环方式的综合性能会更好;但是,在较低的网络负载下,双环方式各方面的性能与单环方式差别不大,而双环方式所需的缓冲区数量远多于单环方式、其布局布线复杂度也远高于单环方式,因此,在网络负载较低时,单环方式是更好的选择.

以上的分析都是假设限定区域内的通信分布是均匀随机的,它能够分析某一类网络应用在特定互连结构中的平均性能. 为了分析具体应用在相关互连结构下的性能,下一步将使用真实的片上网络通信分析、验证网络性能.

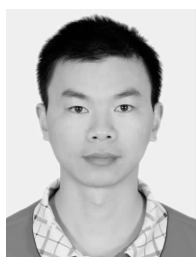
参 考 文 献

[1] Wang Wei, Tang Zhi-Zhong, Qiao Lin. Summarize of the Inter-connection in Chip Multiprocessor. Computer Science, 2008, 35(9): 7-8(in Chinese)
(王炜,汤志忠,乔林. 片上多处理器互连技术综述. 计算机科学, 2008, 35(9): 7-8)

[2] Pande P P, Jones M, Ivanov A, Saleh R. Performance evaluation and design trade-offs for network-on-chip Interconnect architectures. IEEE Transactions on Computers, 2005, 54(8): 1025-1040

[3] Wang Wei, Qiao Lin, Yang Guang-Wen, Tang Zhi-Zhong. Performance analysis of the 2-D networks-on-chip. Journal of Computer Research and Development, 2009, 46(10): 1601-

- 1611(in Chinese)
(王伟, 乔林, 杨广文, 汤志忠. 二维片上网络互连性能分析. 计算机研究与发展, 2009, 46(10): 1601-1611)
- [4] Wang Wei, Qiao Lin, Yang Guang-Wen, Tang Zhi-Zhong. Performance analysis of the extended 2-D grid interconnection networks-on-chip. *Journal of Tsinghua University (Science and Technology)*, 2010, 50(1): 161-164(in Chinese)
(王伟, 乔林, 杨广文, 汤志忠. 扩展二维网格片上互连性能分析. 清华大学学报(自然科学版), 2010, 50(1): 161-164)
- [5] Zheng Wei-Min, Tang Zhi-Zhong. *Computer Architecture*. 2nd Edition. Beijing: Tsinghua University Press, 2006 (in Chinese)
(郑伟民, 汤志忠, 计算机系统结构. 第2版. 北京: 清华大学出版社, 2006)
- [6] Ahmad F, Yuan Xin. Communication characteristics in the NAS parallel benchmarks//*Proceedings of the International Conference on Parallel and Distributed Computing Systems (PDCS'02)*. Cambridge, USA, 2002. IASTED/ACTA Press, 2002: 724-729
- [7] Vetter J S, Mueller F. Communication characteristics of large-scale scientific applications for contemporary cluster architectures. *Journal of Parallel and Distributed Computing*, 2003, 63(9): 853-865
- [8] Kim JunSeong, Lilja D J. Characterization of communication patterns in message-passing parallel scientific application programs//*Proceedings of the Network-Based Parallel Computing: Communication, Architecture, and Applications (CANPC'98)*. Las Vegas, Nevada, USA, 1998: 202-216
- [9] Manjikian N. Prototyping a hierarchical ring interconnect for system-on-chip multiprocessor//*Proceedings of the 2nd Annual IEEE Northeast Workshop on Circuits and Systems (NEWCAS 2004)*, 2004: 85-88
- [10] Denneau M M. The Yorktown simulation engine//*Proceedings of the 19th Design Automation Conference*. Las Vegas, Nevada, 1982: 55-59
- [11] Broomell G, Heath J R. An integrated-circuit crossbar switch system design//*Proceedings of the 4th International Conference on Distributed Computing Systems*. San Francisco, California, USA, 1984: 278-287
- [12] Barber F E et al. A 64×17 non-blocking crosspoint design//*Proceedings of the 31st International Solid State Circuit Conference*. San Francisco, CA, 1988: 116-117
- [13] Hoare R R, Zhu Ding, Jones A K. A two-stage hardware scheduler combining greedy and optimal scheduling. *Journal of Parallel and Distributed Computing*, 2008, 68(1): 1437-1451



WANG Wei, born in 1975, Ph. D. candidate. His research interests include computer architecture, chip-multiprocessor and networks-on-chip.

His research interests include computer architecture, optimizing compiler techniques for parallelism.

YANG Guang-Wen, born in 1963, professor, Ph. D. supervisor. His research interests include computer architecture, parallel and distributed computing.

TANG Zhi-Zhong, born in 1946, professor, Ph. D. supervisor. His research interests include computer architecture, compiler techniques for instruction level parallelism.

QIAO Lin, born in 1972, associate professor, Ph. D.