

基于拓扑划分的片上网络快速映射算法

邓植^① 顾华玺^{*①} 杨银堂^② 李慧^①

^①(西安电子科技大学 ISN 国家重点实验室 西安 710071)

^②(西安电子科技大学微电子学院 西安 710071)

摘要: 该文针对片上网络建立了以能耗和流量均衡为优化目标的映射模型, 提出一种基于拓扑划分的快速映射算法(TPBMAP)。该算法不仅考虑芯片的布局特性从而产生规整的拓扑, 还采用虚拟 IP 核技术修正通信核图以完成 IP 核和网络节点数不等的映射; 通过引入以流量均衡为目标的优化模型同时将通信量大的 IP 核映射到拓扑边缘区域, 有效地降低了网络中心的流量; 采用迭代的拓扑划分方法以及将通信量大的 IP 核映射到网络相邻位置, 可快速完成低能耗映射。仿真结果表明, 相比现有算法, 该文提出的算法在映射速度、全网能耗以及网络中心流量等方面有较大优势。

关键词: 片上网络; 映射优化; 拓扑划分

中图分类号: TP301.6

文献标识码: A

文章编号: 1009-5896(2011)12-3028-07

DOI: 10.3724/SP.J.1146.2011.00422

A Fast Topology Partition Based Mapping Algorithm for Network-on-Chip (NoC)

Deng Zhi^① Gu Hua-xi^① Yang Yin-tang^② Li Hui^①

^①(State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China)

^②(School of Microelectronic, Xidian University, Xi'an 710071, China)

Abstract: A fast mapping algorithm, named Topology Partition Based MAPping (TPBMAP), is proposed considering the energy consumption and traffic balance for the Network-on-Chip (NoC). The proposed algorithm not only generates topology automatically by taking the characteristic of chip layout into consideration, but also achieves the mapping when the number of network nodes is not equal to that of IP cores, using the technology of virtual IP core. Introducing the optimized model, the goal of which is the balance of traffic, the IP cores with heavy traffic are mapped to the edge region to reduce effectively traffic in the center of network. TPBMAP uses iteration of topology partition and makes the IP cores map to the nodes with neighboring in each iteration process, in order to complete rapidly mapping with low energy. According to the experimental results, the proposed algorithm can achieve mapping result with a fast rate, low energy consumption and balanced traffic.

Key words: Network-on-Chip (NoC); Mapping optimization; Topology partition

1 引言

随着半导体技术工艺的发展, 单个芯片上可集成的 IP (Intellectual Property) 核数量不断增加, 这就使得将一个系统集成到单个芯片上成为可能, 最终片上系统(System-on-Chip, SoC)应运而生。随着系统中逻辑门、计算及存储单元尺寸的逐渐减小及系统集成度的进一步增加, 现有的总线结构在时延、吞吐量、功耗以及可扩展性等方面面临着巨大的挑

战, 为此设计一种新的体系结构以解决众多 IP 核间的互连就成了研究的关键所在, 由此出现了片上网络(Network-on-Chip, NoC)^[1-3]。NoC 已成为当前研究的热点^[4]。

映射优化是片上网络设计的关键问题之一, 它是在给定通信图、拓扑以及 IP 核库的基础上, 将每个任务分配到合适的 IP 核上并安排每个 IP 核的任务执行顺序及在 NoC 拓扑中的位置^[5]。映射的主要任务就是将 IP 核或任务模块高效地分配到网络节点上, 使各种应用得以顺利地顺利完成。不同的映射结果对系统的能耗、时延及执行时间有着重要的影响, 通常衡量一个映射性能好坏的主要指标有能耗、时延、热量均衡以及服务质量(Quality of Service, QoS)等^[6]。

2011-05-05 收到, 2011-08-15 改回

国家自然科学基金(60803038, 60725415, 61070046), 中央高校基本业务费项目(K50510010010), 高等学校学科创新引智计划(B08038)以及国家重点实验室专项基金(ISN1104001)资助课题

*通信作者: 顾华玺 hxgu@xidian.edu.cn

映射问题属于二次分配问题 (quadratic assignment problem) 的范畴, 它已被证明是 NP-complete 问题^[7], 其搜索尺度随搜索空间增长呈阶乘递增, 在有限时间和空间中大规模精确求解相当复杂, 常采用启发式算法进行近似优化求解^[8-11]。Morgan 等人^[8]采用遗传算法(GA)完成以面积和时延为优化目标的 IP 核映射优化, 将多目标优化采用加权归一的方法进行单目标优化, 后采用遗传算法求解。通过实验仿真得到该方法在面积和时延方面具有良好的性能。Wang 等人^[9]为避免映射算法过早陷入局部最优, 提出了混沌扰动的离散粒子群优化(DPSO)算法完成以能耗和时延为优化目标、从任务集到 IP 核及 IP 核到网络节点的两阶段映射。实验证明该方法在能耗和时延映射上比遗传算法有更好的性能。Hu 等人^[10]采用分支定界算法完成从 IP 核到网络节点的能耗最小化映射。Dong 等人^[11]采用神经网络算法完成 IP 核的映射, 通过实验仿真得出该方法是一种高性能、低能耗的映射方法。以上这些方法虽然可以通过大量的迭代得到较优解, 但通常是以高的时间复杂度为代价, 难以应用到大规模快速的 IP 核映射之中; 同时这些方法不能保证在短时间内获得低能耗的映射结果, 更未考虑在常见网状拓扑中网络中心热点的产生对网络的吞吐、时延以及可靠性等性能的严重影响。

本文针对网状拓扑片上网络, 分析并提出了能耗模型和流量均衡模型, 给出了一种基于拓扑划分的片上网络快速 IP 核映射算法。该算法首先考虑到芯片特性而生成尽可能规整的网络拓扑, 并采用虚拟 IP 核技术对通信核图进行修正以完成 IP 核和网络节点数目不相等情况下的映射, 同时算法不仅能快速地将 IP 核映射到网络节点上, 还可将通信量大的节点映射到彼此相邻的网络边缘区域, 既保证了全网通信的低能耗又避免了网络中心热点的产生。

2 问题描述与数学模型

定义 1 设有向图 $G(T, C)$, 其中 $t_i \in T$ 为顶点集合, 有向边 $c_{ij} = \langle t_i, t_j \rangle \in C$, c_{ij} 的边权重 v_{ij} , 当 T 代表 IP 核集, c_{ij} 表示从 IP 核 t_i 到 t_j 信息流方向, v_{ij} 表示从 IP 核 t_i 到 t_j 的通信量, 称有向图 $G(T, C)$ 为通信核图, 记为 CCG。

定义 2 设有向图 $G(N, P)$, 其中 $n_i \in N$ 为顶点集合, 有向边 $p_{ij} = \langle n_i, n_j \rangle \in P$, p_{ij} 的边权重 e_{ij} , 当 N 代表网络节点集, p_{ij} 表示从节点 n_i 到 n_j 的物理信道, e_{ij} 表示从节点 n_i 到 n_j 发送 1 bit 数据所消耗的平均能耗, 称有向图 $G(N, P)$ 为拓扑架构图, 记为 TAG。

2.1 能耗模型

Hu 等人在文献[12]中建立了片上网络的能耗模型为

$$E_{\text{bit}} = E_{S_{\text{bit}}} + E_{B_{\text{bit}}} + E_{W_{\text{bit}}} + E_{L_{\text{bit}}} \quad (1)$$

其中 E_{bit} 为单位比特能耗, 即通过 1 bit 数据的能耗, $E_{S_{\text{bit}}}$, $E_{B_{\text{bit}}}$, $E_{W_{\text{bit}}}$ 和 $E_{L_{\text{bit}}}$ 分别表示在路由器中的交叉开关、缓存、路由器内部互连线路以及网络链路上所消耗的能量。

由于在 NoC 中节点间链路相对于路由器内部的链路要长的多, 且片上多采用寄存器微结构单元, $E_{B_{\text{bit}}}$ 和 $E_{W_{\text{bit}}}$ 相比较式(1)中的其他两项要小得多, 因此在文献[12]中作者认为片上网络能耗模型可近似为

$$E_{\text{bit}} = E_{S_{\text{bit}}} + E_{L_{\text{bit}}} \quad (2)$$

由式(2)可得, 从源节点 n_i 到目的节点 n_j 发送 1 bit 数据的能耗为

$$E_{\text{bit}}^{n_i, n_j} = H \times E_{S_{\text{bit}}} + (H + 1) \times E_{L_{\text{bit}}} \quad (3)$$

其中 H 为从源节点 n_i 到目的节点 n_j 所经过的跳数, H 常采用 Manhattan 距离。

2.2 流量均衡模型

网络中心热点的产生会影响网络的吞吐、时延等性能, 尤其是在网状拓扑的片上网络设计中, 网络中心热点产生的不均衡热量会严重影响到整个芯片的可靠性, 因此流量均衡是片上网络优化设计的重要方面。网络的流量均衡状况不仅与该节点通过的流量大小有关, 同时还与节点在网络中位置有关。通过网络中每个节点的流量与全网平均流量越接近, 则该网络均衡性能就越好; 热点区域位于网络中心对网络的影响远远超过位于网络边缘区域的影响。本文提出一种衡量网络流量均衡状况的模型为

$$\text{Btraf} = \frac{1}{m} \sum_{i=1}^m (\text{Dtraf}_{n_i} \times \text{factor}_{n_i}) \quad (4)$$

其中 Dtraf_{n_i} 为节点 n_i 的流量与全网流量均值的差, factor_{n_i} 为节点 n_i 的影响因子, 即该节点的流量对全网流量均衡的影响因子, 它与该节点在网络中的位置有关。

$$\text{Dtraf}_{n_i} = |\text{traf}_{n_i} - \text{atraf}| \quad (5)$$

其中 $|x|$ 表示变量 x 的绝对值, traf_{n_i} 为通过网络节点 n_i 的流量大小, atraf 表示全网流量均值, $\text{atraf} = \frac{\sum_{i=1}^m \text{traf}_{n_i}}{m}$, $1 \leq i \leq m$, m 为网络节点总数。

节点 n_i 的影响因子与该节点位置的关系为

$$\text{factor}_{n_i} = e^{-\text{dis}(n_i, n_c)} \quad (6)$$

其中 $\text{dis}(n_i, n_c)$ 为节点 n_i 与网络中心点 n_c 的距离, 若

在 2D 网状拓扑中,用一组 2 维坐标 (x_i, y_i) 表示节点 n_i , 则 $\text{dis}(n_i, n_c) = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$ 。

基于式(4)-式(6), 则流量均衡模型可表示为

$$\text{Btra} = \frac{1}{m} \sum_{i=1}^m \left(\left| \text{traf}_{n_i} - \frac{\sum_{i=1}^m \text{traf}_{n_i}}{m} \right| \times e^{-\text{dis}(n_i, n_c)} \right) \quad (7)$$

其中 traf_{n_i} 为通过网络节点 n_i 的流量, $\text{dis}(n_i, n_c) = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$, m 为网络节点总数。网络流量越均衡, 热点区域距离网络中心越远, 则 Btra 值就越小。

2.3 映射模型

基于以上定义及模型, 映射问题就可以转化为寻找一个映射函数 $T \xrightarrow{\text{map}()} N$, 使其满足如下优化问题:

$$\min \left\{ \text{Energy} = \sum_{\forall c_i, j} v_{ij} \times e_{\text{map}(t_i), \text{map}(t_j)} \right\} \quad (8)$$

$$\min \left\{ \text{Btra} = \sum_{i=1}^m \left(\left| \text{traf}_{n_i} - \frac{\sum_{i=1}^m \text{traf}_{n_i}}{m} \right| \times e^{-\text{dis}(n_i, n_c)} \right) \right\} \quad (9)$$

$$\text{s.t. } \forall t_i \in T, \text{map}(t_i) \in N \quad (10)$$

$$\forall t_i \neq t_j \in T, \text{map}(t_i) \neq \text{map}(t_j) \in N \quad (11)$$

其中式(8)中 v_{ij} 表示从 IP 核 t_i 到 t_j 的通信量, $e_{\text{map}(t_i), \text{map}(t_j)} = H \times E_{S_{\text{bit}}} + (H + 1) \times E_{L_{\text{bit}}}$, H 为网络节点 $\text{map}(t_i)$ 和 $\text{map}(t_j)$ 的通信距离, $\text{map}(t_i)$ 为 IP 核 t_i 映射到对应网络的节点, 其与映射结果有关。式(10)和式(11)表示一个 IP 核只能映射到一个网络节点, T 为 IP 核集, N 为网络节点集。

3 基于拓扑划分的 NoC 映射

基于拓扑划分映射(Topology Partition Based MAPping, TPBMAP)算法主要包括如下几个阶段: 首先根据通信核图中 IP 核数目产生尽可能规整的网络拓扑, 并根据拓扑中节点数目采用虚拟 IP 核技术修正通信核图; 其次对网络拓扑进行划分并按划分结果对通信核图进行相应数目的 IP 核划分; 最后通过计算子拓扑区域的能耗, 并将能耗最小的位置作为当前 IP 核的映射位置进而完成子拓扑区域映射。

3.1 拓扑产生

根据通信核图中 IP 核数目 k 生成的网络拓扑 $\text{row} \times \text{col}$ 应该满足条件 $\text{row} \times \text{col} \in \left\{ \left\lfloor \sqrt{k} \right\rfloor \times \left\lfloor \sqrt{k} \right\rfloor, \left\lfloor \sqrt{k} \right\rfloor \times \left(\left\lfloor \sqrt{k} \right\rfloor + 1 \right), \left\lfloor \sqrt{k} \right\rfloor \times \left(\left\lfloor \sqrt{k} \right\rfloor + 2 \right), \dots, \left\lfloor \sqrt{k} \right\rfloor \times \left\lfloor \sqrt{k} \right\rfloor \right\}$, 并

使得产生的拓扑节点数目应大于等于 IP 核数目且尽可能使二者相等, 即满足 $\min\{\text{row} \times \text{col} - k\}$ 及 $(\text{row} \times \text{col}) - k \geq 0$, 其中 row 为拓扑行数, col 为拓扑列数。 $\lfloor x \rfloor$ 为变量 x 的向下取整, $\lceil x \rceil$ 为变量 x 的向上取整。

3.2 通信核图修正

根据拓扑节点数目 m 对通信核图进行修正包括如下步骤(设通信核图中 IP 核数目为 k):

步骤1 若网络节点数目 m 大于 IP 核数目 k , 则在通信核图中增加 $m - k$ 个虚拟 IP 核, 并假设这些虚拟 IP 核与其它 IP 核的通信量为 0, 以保证通信核图中 IP 核数目和网络节点数目相等。否则不对通信核图修正;

步骤2 若在步骤 1 中对通信核图进行修正, 则待到所有 IP 核映射完毕删除步骤 1 中的虚拟 IP 核。

3.3 拓扑划分

设拓扑区域为 $\text{row} \times \text{col}$, row 为行数, col 为列数。拓扑划分包括如下步骤:

步骤1 判断当前划分次数, 若为奇数次则进行行列划分, 否则进行行划分;

步骤2 若拓扑的行(列)数为奇数, 则分为左, 中, 右(上, 中, 下)3 个子拓扑平面, 其拓扑参数分别为 $\text{row} \times \lfloor \text{col}/2 \rfloor$, $\text{row} \times 1$, $\text{row} \times \lfloor \text{col}/2 \rfloor$ ($\lfloor \text{row}/2 \rfloor \times \text{col}$, $1 \times \text{col}$, $\lfloor \text{row}/2 \rfloor \times \text{col}$); 若为偶数则分为左, 右(上, 下)2 个子拓扑平面, 其拓扑参数为 $\text{row} \times \lfloor \text{col}/2 \rfloor$, $\text{row} \times \lfloor \text{col}/2 \rfloor$ ($\lfloor \text{row}/2 \rfloor \times \text{col}$, $\lfloor \text{row}/2 \rfloor \times \text{col}$);

步骤3 重复划分拓扑, 直到划分的子集中网络节点数目小于等于 2 个。

3.4 通信核图划分

对当前通信核图进行相应数目的 IP 核划分, 使得划分的 IP 核间通信量最小, 并标记此次划分结果到每个 IP 核中。通信量最小的 IP 核划分按照以下步骤进行:

步骤1 根据通信核图构造通信关系矩阵 \mathbf{W} , 并计算矩阵 \mathbf{W} 对应的特征值及特征向量;

步骤2 对特征值进行有序排列, 按照拓扑划分结果对特征值进行相应数目划分;

步骤3 按特征值划分结果寻找对应的 IP 核, 并标记划分结果到每个 IP 核中。

3.5 子区域映射

子区域映射包括如下步骤:

步骤1 在通信核图中选定通信的起始 IP 核, 若在通信核图中还有与起始 IP 核具有相同标记的 IP 核, 则随机选择一种匹配方式将这两个 IP 核映射到该标记所对应的子区域;

步骤2 从通信核图的未映射 IP 核中选择当前欲映射 IP 核, 选择的优先级为先考虑未映射 IP 核与已映射 IP 核是否具有通信关系, 再考虑未映射 IP 核与已映射 IP 核间的通信量大小, 最后从未映射 IP 核中随机选择一个 IP 核作为当前欲映射 IP 核;

步骤3 选择与当前欲映射 IP 核具有相同标记向量的 IP 核, 并计算这两个 IP 核在其对应区域不同位置的通信能耗, 将能耗最小处的位置作为 IP 核的最终映射位置;

步骤4 标记本次映射结果到每个 IP 核中, 返回步骤 2, 直到所有的 IP 核映射。

3.6 算法伪代码

为详细说明 TPBMAP, 本文给出了算法伪代码如表 1 所示。

3.7 复杂度分析

假设有一个具有 k 个 IP 核的通信核图映射到 $m(m \geq k)$ 个节点的网状拓扑上, 则拓扑产生时间复杂度为 $O(1)$, 通信核图修正时间复杂度为 $O(1)$, 网络拓扑及通信核图划分复杂度至多为 $O(\log_2^m)$, 子区域映射复杂度为 $O(m)$, 因此基于拓扑划分的 NoC 映射算法时间复杂度为 $O(m^2)$ 。

4 实验及结果分析

为验证提出的 TPBMAP 的性能, 本文分别对 TPBMAP 和基于遗传算法(GA)的映射^[13]分别进行了实验仿真。在通信核图选取时, 本文采用当前仿真中普遍使用的 16 核 Video Object Plane Decoder (VOPD)和 12 核 MPEG-4 解码器^[14-16], 如图 1 所示。仿真采用 Pentium(R) Dual-Core CPU E6500 @2.93 GHz, 2.93 GHz, 1.96 GB 内存的 PC 机上的

表 1 算法伪代码

Topology Partition Based MAPping (TPBMAP)	
输入:	通信核图(CCG), 拓扑架构图(TAG)
输出:	映射结果
(1)	if (topology parameters are not given)
(2)	generate topology (); //拓扑产生
(3)	endif
(4)	if (number of nodes is not equal to IPs)
(5)	modify communication core graph (); //通信核图修正
(6)	endif
(7)	initialize Cnum = 1; //初始化划分次数为 1
(8)	while (number of IP cores in same flag>2) //若划分子集中 IP 核数目大于 2 则重复执行
(9)	if (mod (Cnum,2) ==1)
(10)	cut NoC topology architecture into 3 parts; //拓扑 3 划分
(11)	cut the communication core graph; //通信核图划分
(12)	flag result to IP core; //标记划分结果到 IP 核中
(13)	else
(14)	cut NoC topology architecture into 2 parts; //拓扑 2 划分
(15)	cut the communication core graph; //通信核图划分
(16)	flag result to IP cores; //标记划分结果到 IP 核中
(17)	endif
(18)	Cnum = Cnum + 1; //划分数加 1
(19)	endwhile
(20)	compute energy with different location (); //子区域映射

MATLAB 软件完成。在能耗计算时, 链路和路由器能耗分别为 0.7066 nJ 和 0.9334 nJ, 时钟频率为 1 GHz, 0.18 μm CMOS 工艺^[17]。

我们首先对不同映射算法在图 1 所示的两种具体应用通信核图的能耗进行了仿真, 如表 2 所示,

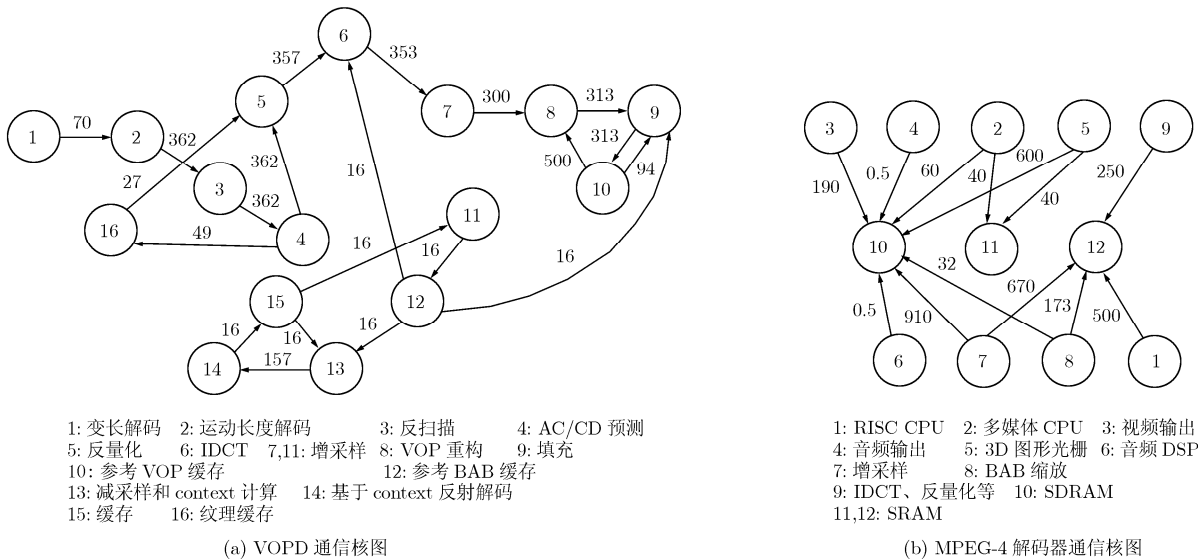


图 1 应用具体通信核图(MB/s)

表2 基于GA的映射算法与TPBMAP在能耗和运行时间的对比

	基于GA的映射算法						TPBMAP
	迭代次数	1	5	10	50	1000	
VOPD	能耗(mJ)	14265.56	11708.80	11582.52	10936.36	10936.36	9506.30
	运行时间(s)	0.047	0.188	0.250	0.641	9.297	0.0179
MPEG-4 解码器	能耗(mJ)	11206.68	9822.52	9646.22	8669.60	8669.60	9078.80
	运行时间(s)	0.047	0.188	0.219	0.516	7.141	0.0186

分别为 TPBMAP 和基于 GA 映射的运行时间及能耗对比分析表,表中分别记录了基于 GA 映射的第 1, 5, 10, 50 和 1000 次迭代的时间及其最优能耗值。从表 2 可以看出在相同的运行时间内,在 VOPD 应用下基于 GA 的映射能耗值在 14265.56 mJ 以上,而本文提出的 TPBMAP 仅为 9506.30 mJ,能耗降低了 33.4%;在 MPEG-4 解码器应用下,在相同运行时间内 GA 映射的能耗值在 11206.68 mJ 以上,而本文提出的映射方法仅为 9078.80 mJ,降低了 19.0%。在两种映射算法得到相同的能耗的情况下,基于 GA 的映射至少需要迭代 10 次,运行时间为 TPBMAP 的 10 倍多。由基于 GA 映射的迭代性能可以看出,由于遗传算法过早地陷入局部最优,以至于在后来相当长的运行时间内都没有更优解的产生。

通过不同映射结果的流量仿真,我们得到了不同映射结果下的空间流量分布状况,如表 3 和图 2、图 3 所示。图 2 为不同映射算法将 VOPD 通信核图映射到 4×4 2D 网状拓扑的两种结果下,采用 XY 维序路由的网络流量空间分布示意图。可以看出在 GA 映射结果下网络中心流量较大,且明显地产生了热点区域,而本文提出的 TPBMAP 将流量大的 IP 核映射到了网络边缘,网络中热点也得到了有效缓解。通过仿真计算,基于 GA 的映射结果的流量均衡值为 110.9,而本文提出的映射结果的流量均衡值为 83.6。由表 3 可以得到在 TPBMAP 映射结果下,网络最大流量为 1129 MB/s,距离网络中心为 1.6 倍单位跳数,而 GA 结果的最大流量为 1524 MB/s,且距网络中心较近,仅 0.7 倍单位跳数。由 GA 产生的映射结果的全网平均流量为 549.5 MB/s,

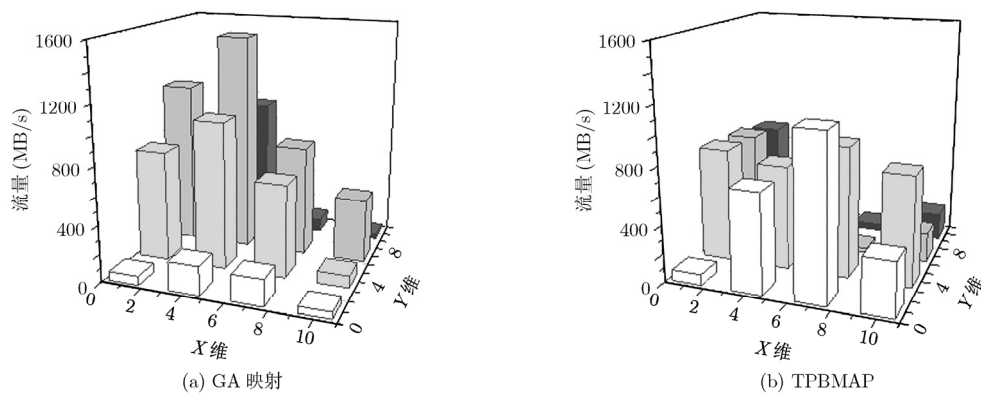


图2 VOPD下不同映射结果的流量空间分布

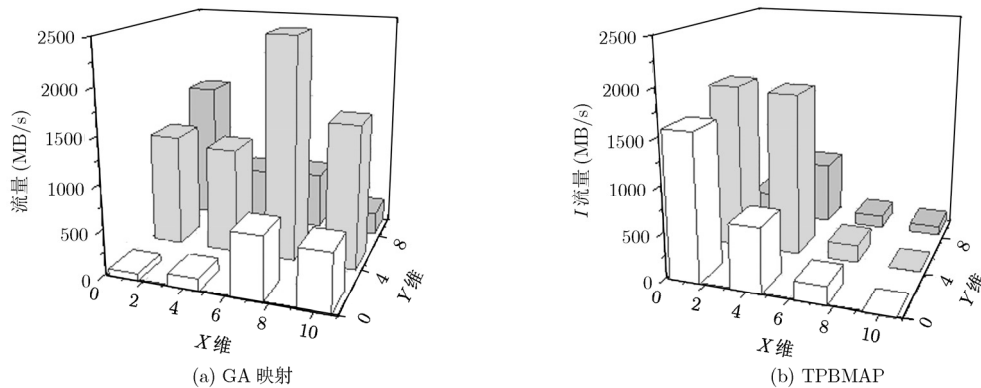


图3 MPEG-4 解码器下不同映射结果的流量空间分布

表 3 基于 GA 的映射算法与 TPBMAP 在空间流量分布性能对比 (MB/s)

应用	映射方法	网络最大流量			网络中心流量	全网平均流量	网络最大值 与均值之差
		最大值	位置	与网络中心距离(跳数)			
VOPD	TPBMAP	1129	(1,3)	1.6	433.8	496.9	632.1
	GA	1524	(3,2)	0.7	994.0	549.5	974.5
MPEG-4 解码器	TPBMAP	1830	(2,1)	1.5	991.8	625.7	1204.3
	GA	2463	(2,3)	0.5	1816.5	927.2	1534.8

而本文提出的 TPBMAP 映射结果平均流量仅为 496.9 MB/s。

图 3 为分别采用 TPBMAP 和基于 GA 的映射算法将 MPEG-4 解码器映射至 4×3 2D 网状拓扑映射结果, 在采用 XY 维序路由的网络流量空间分布示意图。可以看出, 在本文提出的 TPBMAP 映射结果下网络中心的流量得到有效降低, 流量均衡效果要好于基于 GA 的映射结果。通过仿真分析, 在基于 GA 映射结果的流量均衡值为 175.6, 而本文提出的映射结果的流量均衡值为 165.4。由表 3 可以看出, 在 MPEG-4 解码器应用下, TPBMAP 映射结果的最大流量为 1830 MB/s, 距离网络中心较远为 1.5 倍单位跳数, 而基于 GA 的映射结果的网络最大流量为 2463 MB/s, 距离网络中心仅为 0.5 倍单位跳数, 且该映射结果的全网平均流量为 927.2 MB/s, 是 TPBMMAP 映射结果网络平均流量的 1.5 倍。

5 结论

映射优化问题是片上网络研究的一项关键技术, 它的模型建立及其求解性能影响着片上网络最终的性能。本文针对片上网络设计平台, 根据芯片布局特性生成尽可能规整的拓扑, 并采用虚拟 IP 核技术修正通信核图, 以完成 IP 核和网络节点不等的映射, 采用拓扑划分、通信核图核划分及最小能耗的子区域映射, 从而得到一种快速、低能耗和流量均衡的 IP 核映射算法。仿真结果表明, 与基于遗传算法的映射相比, 本文提出的映射算法不仅可以快速得到映射结果, 在相同运行时间能耗值可降低 33.4%, 并且在流量均衡方面也有更好的性能。

参 考 文 献

- [1] 葛芬, 吴宁. 面向特定应用的片上网络低能耗拓扑生成方法[J]. 系统工程与电子技术, 2010, 32(8): 1754-1759.
Ge Fen and Wu Ning. Low energy topology generation approach for application-specific network on chip[J]. *System Engineering and Electronics*, 2010, 32(8): 1754-1759.
- [2] 王坚, 李玉柏, 蒋勇男. 片上网络通信性能分析建模与缓存分

- 配优化算法[J]. 电子与信息学报, 2009, 31(5): 1059-1062.
Wang Jian, Li Yu-bai, and Jiang Yong-nan. Communication performance analytical model and buffer allocation optimizing algorithm for Network-on-Chip[J]. *Journal of Electronics & Information Technology*, 2009, 31(5): 1059-1062.
- [3] 常政威, 熊光泽, 桑楠, 等. 基于电压岛的能量和可靠性感知 NoC 映射[J]. 计算机辅助设计与图形学学报, 2009, 21(1): 19-26.
Chang Zheng-wei, Xiong Guang-ze, Sang Nan, et al. Energy- and reliability-aware mapping for NoC implemented with voltage islands[J]. *Journal of Computer-aided Design & Computer Graphics*, 2009, 21(1): 19-26.
- [4] 董文箫, 沈海斌, 全励, 等. 基于遗传算法的片上网络低能耗映射[J]. 浙江大学学报(理学版), 2010, 37(6): 654-656, 669.
Dong Wen-xiao, Shen Hai-bin, Quan Li, et al. Power-aware mapping based on genetic algorithm for Network-on-Chip[J]. *Journal of Zhejiang University(Science Edition)*, 2010, 37(6): 654-656, 669.
- [5] 岳培培, 刘建, Anjum S, 等. NoC 映射问题中的列举路径分配算法[J]. 电子科技大学学报, 2008, 37(1): 54-57.
Yue Pei-pei, Liu Jian, Anjum S, et al. Enumeration-based path allocation algorithm in NoC mapping[J]. *Journal of University of Electronic Science and Technology of China*, 2008, 37(1): 54-57.
- [6] 陈庚生, 陈亦欧, 胡剑浩. 面向实时数字信号系统关键链路延时的 NoC 映射方法研究[J]. 电子与信息学报, 2010, 32(7): 1638-1643.
Chen Geng-sheng, Chen Yi-ou, and Hu Jian-hao. A novel critical delay-aware mapping method for real-time digital signal systems with NoC platform[J]. *Journal of Electronics & Information Technology*, 2010, 32(7): 1638-1643.
- [7] Sartaj S and Teofilo G. P-complete approximation problems[J]. *Journal of Association for Computing Machinery*, 1976, 23(3): 555-565.
- [8] Morgan A A, Elmiligi H, El-Kharashi M W, et al. Multi-objective optimization for Networks-on-Chip architectures using Genetic Algorithms[C]. IEEE International Symposium on Circuits and Systems (ISCAS), Paris, May 2010: 3725-3728.

- [9] Wang Lei and Ling Xiang. Energy-and latency-aware NoC mapping based on chaos discrete particle swarm optimization[C]. International Conference on Communications and Mobile Computing (CMC), Shenzhen, April 2010, 1: 263–268.
- [10] Hu Jing-cao and Marculescu R. Energy-and performance-aware mapping for regular NoC architectures[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2005, 24(4): 551–562.
- [11] Dong Yi-ping, Wang Yang, Li Zhen, *et al.* High performance and low latency mapping for neural network into network on chip architecture[C]. IEEE 8th International Conference on ASIC(ASICON), Changsha, October 2009: 891–894.
- [12] Hu Jing-cao and Marculescu R. Exploiting the routing flexibility for energy/performance aware mapping of regular NoC architectures[C]. Design, Automation and Test in Europe Conference and Exhibition, Munich, March 2003: 688–693.
- [13] Tang L and Kumar S. A two-step genetic algorithm for mapping task graphs to a network on chip architecture[C]. Euromicro Symposium on Digital System Design, Turkey, September 2003: 180–187.
- [14] Chen Yan-cang, Xie Lun-guo, and Li Jin-wen. An energy-aware heuristic constructive mapping algorithm for Network on Chip[C]. IEEE 8th International Conference on ASIC(ASICON), Changsha, October 2009: 101–104.
- [15] Tornero R, Sterrantino V, Palesi M, *et al.* A multi-objective strategy for concurrent mapping and routing in networks on chip[C]. IEEE International Symposium on Parallel & Distributed Processing(IPDPS), Rome, May 2009: 1–8.
- [16] Van Der Tol E B and Jaspers E G T. Mapping of MPEG-4 decoding on a flexible architecture platform[C]. SPIE-Media Processors, 2002, 4674: 1–13.
- [17] Chen Xuning and Peh L. Leakage power modeling and optimization in interconnection networks[C]. The International Symposium on Low Power Electronics and Design(ISLPED), Seoul, August 2003: 90–95.
- 邓 植: 男, 1986 年生, 硕士生, 研究方向为片上网络以及映射算法.
- 顾华玺: 男, 1977 年生, 博士, 副教授, 研究方向为网络技术、片上网络以及光互连等.
- 杨银堂: 男, 1962 年生, 教授, 博士生导师, 研究方向为 VLSI 技术、NoC 体系架构等.
- 李 慧: 女, 1988 年生, 硕士生, 研究方向为光片上网络.