

基于改进粒子群的 3D-Mesh CMP 片上 网络映射算法*

杨 微, 张 振, 刘怡俊

(广东工业大学 计算机学院, 广州 510006)

摘 要: 经典粒子群算法是在连续的解空间里进行的,粒子的方向、速度的确定采用的是二维空间的运算法则,不适合在 N 维空间里进行,而 3D-NOC 映射问题面对的是 N 维离散解空间。对此,提出一种改进的粒子群算法以及相应的算法评估模型。该算法通过引入非支配解(Pareto 解)的概念对粒子群算法进行改进,使得算法不仅可以对多个评估模型参数同时优化,而且可以依据实际的应用对单个评估模型参数进行优化。仿真实验结果表明,该算法在单个评估模型且同等计算规模下对比随机映射性能有显著的提升,在降低运算规模方面有很好的优化能力,且在多模型下取得的综合最优解比之相应的单独评估模型下的最优解只有很小的下降。

关键词: 3D-NOC; 片上多处理器; 映射算法; 粒子群算法; 平均网络延时模型; 最大延时模型; 功耗模型; 流量均衡模型

中图分类号: TP393.04 文献标志码: A 文章编号: 1001-3695(2013)05-1345-04
doi:10.3969/j.issn.1001-3695.2013.05.016

Improved particle swarm optimization algorithm based mapping algorithm for 3D-Mesh CMP

YANG Wei, ZHANG Zhen, LIU Yi-jun

(School of Computers, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: The classical particle swarm optimization(PSO) is carried out in continuous solution space, the orientation and the speed of the particles are determined by the algorithm of two-dimensional space in this case, which is not suitable for N dimensional space, and the 3D-NOC mapping is being used in N dimension discrete solution space. According to this problem, this paper proposed an improved particle swarm optimization algorithm and the corresponding algorithm evaluating model. It improved the algorithm particle swarm algorithm by introducing a non-dominated solution(Pareto solution) concept so that the algorithm could optimize multiple parameters of the assessment model at the same time and optimize an individual parameter of the assessment model based on the actual application. The experimental results show that, comparing to random mapping in the mapping model of the same calculation scale, the proposed algorithm has a significant performance improvement and good optimization in reducing the computing scale and the comprehensive optimal solution in a multi-model comparing separate assessment of the model, only a small decline in its performance.

Key words: 3D-NOC; CMP; mapping algorithm; particle swarm optimization algorithm; average network delay model; maximum delay model; power model; traffic equilibrium model

0 引言

随着片上多处理器(CMP)的发展,芯片内电路的复杂程度快速增加,时钟频率已经提升到极限,CPU 芯片内的线上延时大于门电路延,这种延迟逐渐变成阻碍 CPU 速度的主要负面因素。近年来,NOC 替代传统总线结构的改变虽然克服了全局延迟过长带来的信号完整性及全局同步等一系列问题,但并没有在根本上解决缩短物理连线、减小信号时延的问题。这主要是由于 2D NOC 布局条件的限制,难以保证关键部位相邻以缩短关键路径长度,而 3D 集成技术可以把不同的期间层叠起来,不仅在真正意义上缩短了连线的长度,而且克服了这种

布局的限制^[1]。与传统的 2D-NOC 相比,3D-NOC 的优势具有明显的优势。

在 3D NOC 设计中,NOC 映射算法是 NOC 设计的重要内容,文献[2]中证实片上网络映射问题对系统的功耗、延迟等性能均有很大的影响。近年来各界学者对 NOC 映射算法的研究大多是以功耗优先的映射模型设计为入口,以达到降低系统功耗和优化效率为目标。例如,面向通信能耗的 3D NOC 映射研究中提出了针对网络的通信能耗模型,以蚁群算法实现了面向通信能耗的映射^[3];基于遗传算法的 NoC 处理单元映射研究^[4]、基于拓扑划分的片上网络快速映射算法^[5]等也是这种情况。这些研究成果都在一定程度上对算法进行了优化,但是在进行算法设计时都仅考虑单个因素对映射结果的影响。

收稿日期: 2012-09-10; 修回日期: 2012-10-21 基金项目: 国家自然科学基金资助项目(61106019); 广东省省部产学研结合项目(2011B090400408, 2011A090200022)

作者简介: 杨微(1987-),女,江西抚州人,硕士研究生,主要研究方向为计算机体系结构、片上网络(wendy19872013@126.com);张振(1983-),男,硕士研究生,主要研究方向为计算机体系结构;刘怡俊(1977-),男,教授,博士,主要研究方向为计算机体系结构、集成电路设计。

而在不同的系统设计中,不仅考虑功耗、延时、流量均衡等单个因素的影响,也要综合考虑多方面因素的影响。对此,文中以粒子群算法作为本文映射算法的基础算法,提出基于改进粒子群的 3D-Mesh CMP 片上网络映射算法。该算法引入非支配解的概念,以多个评估模型为入口,改进的算法既可以对单一目标函数进行优化,也可以在多个目标函数下进行联合优化。在研究中选取 3D-Mesh 结构拓扑,由于其结构规整、布线简单,具有网络直径较短、网络各节点位置平等的优点,有利于利用灵活的任务映射算法和合适的路由策略在满足每个任务通信要求的基础上进行全网性能的优化。路由算法选择 XYZ 路由算法。

1 面向 3D-Mesh 的映射算法的目标函数

在片上网络的映射模型中,对于有 N 个任务的应用输入,对应有 $N!$ 种排列方式,构成映射模型的解空间。映射算法的目标是在解空间中寻找到一种排列映射到目标网络,使得它在平均延时、吞吐、功耗、流量均衡四个目标下达到最优,也即在对应的目标函数下取值最小。对此,以下四个模型下的目标函数作为本文提出的映射算法的评价模型。

1.1 平均延时模型

笔者对文献[6]提出的平均延时模型方案进行优化,将数据片的排队延时考虑进去,得到平均延时模型(式(1))。式中: T_{arb} 为路由计算时间, T_p 为解包打包时间, T_{swit} 为微片在路由器内部的时间, T_{link} 为信道传输时间。以 $W_{i,j}$ 为 V_i 到 V_j 的路由路径上所有路由节点的排队时间之和, V_i 到 V_j 各路路由节点的排队时间依次是 $W_{i,j}^k$ ($1 \leq k \leq H_{i,j}$)。片上网络的总延时 D 为式(2)。

$$D_{i,j} = H_{i,j} (T_{arb} + T_{swit} + T_{link}) + T_p + (B_{i,j}/w + 1) (T_{arb} + T_{swit} + T_{link}) + W_{i,j} \quad (1)$$

$$D = \sum_{i,j} D_{i,j} = (T_{arb} + T_{swit} + T_{link}) \times \sum_{i,j} H_{i,j} + \sum_{i,j} \sum_{k=1}^{H_{i,j}} W_{i,j}^k + \frac{(T_{arb} + T_{swit} + T_{link})}{w} \times \sum_{i,j} B_{i,j} + n^2 \times T_p + n^2 \times (T_{arb} + T_{swit} + T_{link}) \quad (2)$$

得到平均延时目标函数为

$$\min(\sum_{i,j} H_{i,j}) \quad (3)$$

1.2 最大延时评估模型

依据路由节点收发信息的情况,以 $\text{num}_{i,j}^k$ ($1 \leq k \leq H_{i,j}$) 为 V_i 到 V_j 路由路径上的冲突系数, t 为一个信息片在无阻塞的情况下从路由节点转发到线路传输时间,得到 V_i 到 V_j 节点延迟为

$$T_{i,j} = t \times (\sum_{k=1}^{H_{i,j}} \text{num}_{i,j}^k) \quad (4)$$

得到最大延时模型目标函数为

$$\min(\max(\sum_{k=1}^{H_{i,j}} \text{num}_{i,j}^k)) \quad (5)$$

1.3 功耗模型

笔者以文献[7]中提出的功耗模型为基础,提出基于单个节点的功耗模型,在此模型下, E_i 为路由节点 i 的功耗,片上网络总功耗 E 为所有路由节点的总功耗 $\sum E_i$ 和所有通信线路的总功耗 $\sum LE_{i,j}$ 之和, $\sum_i \text{num}_i$ ($\sum_i \text{num}_i = \sum_{i,j} (B_{i,j}/w) \times H_{i,j}$) 为路由

器处理所有微片总和。

$$E_i = (E_{\text{running}} - E_{\text{empty}}) \times (T_{\text{arb}} + T_{\text{swit}}) \times \text{num}_i + E_{\text{empty}} \quad (6)$$

$$LE_{i,j} = ef \times (B_{i,j}/w + 1) \times H_{i,j} \quad (7)$$

$$E = \sum E_i + \sum LE_{i,j} = (E_{\text{running}} - E_{\text{empty}}) \times (T_{\text{arb}} + T_{\text{swit}}) \times (\sum_i \text{num}_i) + ef \times \sum_{i,j} [(B_{i,j}/w + 1) \times H_{i,j}] + n \times E_{\text{empty}} \quad (8)$$

得到功耗模型目标函数为

$$\min(\sum_{i,j} B_{i,j} \times H_{i,j}) \quad (9)$$

1.4 流量均衡模型

结合文献[5]提出的一种衡量网络流量均衡状况的模型(见式(10)和(8))提出新的流量均衡模型。其中,令网络中转发信息片数最多的节点转发的信息片个数与平均转发量之比为 T ,设定阈值 T_{value} 。如果 $T > T_{\text{value}}$,则说明该次映射没有达到预期的防热点设计需求。由于路由器处理所有微片总和为 $\sum_i \text{num}_i$,即 $\sum_{i,j} (B_{i,j}/w) \times H_{i,j}$,由 $\sum_{i,j} (B_{i,j}/w) \times H_{i,j}/n$ 可得到网络节点的平均转发量,则映射问题在流量均衡模型下的约束条件为式(11)。

$$\text{Btraf} = \frac{1}{m} \sum_{i=1}^m (\text{dtraf}_i \times \text{factor}_i) \quad (10)$$

$$n \times \max_{i=1}^n (\text{num}_i) / \sum_{i=1}^n \text{num}_i \leq T_{\text{value}} \quad (11)$$

得到流量均衡模型的目标函数为

$$\min(\max_{i=1}^n (\text{num}_i)) \quad (12)$$

2 改进粒子群映射算法

2.1 基于改进粒子群映射算法研究

经典的粒子群算法是著名的群体智能算法,它是模仿鸟群寻找食物的过程。科学家发现:鸟群总能在较大的区域内以较快的速度寻找到食物。通过观察研究,发现鸟群具有独特的群体寻食特征,进而提出了粒子群算法[8]。在经典粒子群算法中,每个粒子代表一只鸟,最优解即所有鸟寻找的食物,每个粒子通过跟踪群体最优解和个体最优解来指导自己的搜索方向,每迭代一次,就通过比较所有粒子的适应度来产生一个群体最优解,当迭代次数达到上限或无法得到新的较优解时,该群体最优解即是粒子群算法的最终解。经典粒子群算法是在连续的解空间里进行的,且在确定粒子的方向和速度时,采用的是二维空间的运算法则,并不适合在 N 维空间进行。但是从上文可以发现在 CMP 映射模型中,所有可能的映射方式构成了一个 N 维的离散的解空间,每一种可能的映射对应这个解空间里的一个点,且单个解的每一维的值都不能相同(因为是 N 个位置的一种排列),所以经典粒子群算法不适用于 3D 片上网络映射模型。

本文提出了适合 3D-Mesh CMP 映射模型的改进粒子群算法。在改进粒子群映射算法研究中,任务序列与 CPU 核(CPU 核泛指微处理器、DSP、各种专用功能模块)采用相同的编码方式。在本算法中,解空间即为 N 个任务序列的全排列,每个粒子对应解空间的一个点,由于不同的任务映射到不同的 CPU 核,所以解空间的每个点的每一维都不相同,以两个解的相同维的不同个数作为解之间的具体,也即不同点的最小具体为 2。由于在 N 维空间中,粒子是以跳变的方式从一个点跳到另

一个点,所以算法定义最密集和最稀疏集来确定粒子跳变的方向,且每次跳变距离缩小两个单位。前面提到过片上网络映射的实质是在多目标方向的限制下在解空间找到最优解的离散空间组合优化问题,基于该特点引入 Pareto 解(非支配解)的概念:比较两个解 A_1 、 A_2 ,如果 A_1 在某个目标下优于 A_2 ,而在其他目标下比 A_2 差,则称 A_1 、 A_2 互相不支配,可同时选入非支配解集以待进一步优化;如果 A_1 在所有目标下都比 A_2 优良,则称 A_2 受 A_1 支配,此时可删除 A_2 ,继续保留 A_1 在非支配解集中。

2.2 基于改进粒子群映射算法实现

粒子探索最优解的主要过程如下:

- 初始化粒子群,为每个粒子在解空间分配一个初始位置。
- 计算每个粒子所在位置在各个模型下的目标函数值,得到本轮的非支配解集合。
- 计算非支配解集中各个点的密集度,选出最稀疏方向点和最密集方向点。为确保粒子群具有较广的解空间搜索范围,以一定的概率使一部分粒子放弃对当前优秀解的追逐而转向其他粒子未搜索到的解空间,即当前非限定解分布最稀疏的地方。
- 确定粒子下一步的探索方式,即以一定概率决定每个粒子下一步的探索方式(跟随密集点方向的局部解域探索或者选择稀疏解方向的新解域探索)。

e) 确定粒子下一步的探索方向。如果选择局部解域探索,则以解集中最密集方向点作为下一步的寻优方向点;如果以寻求新解域方式探索,则以最稀疏方向点作为下一步探索的方向点。

f) 确定下一步的最终位置。在确定的探索方向上,在探索范围内的距离确定下一步的最终位置(产生微小的扰动以避免重复点的探索,在目标位置附近随机找一个点作为下一步的最终位置)。

g) 如果指定的循环次数达到,则结束探索过程,输出当前的非限定解;否则,转到 b) 继续探索。

改进粒子群算法的整体流程如图 1 所示。

在本算法中,每个粒子按照设定的概率飞向不同的方向,即飞向已搜索解空间的密集点(本算法中用 close 变量表示)和飞向未搜索解空间的稀疏点(本算法中用 rare 变量表示)。其中 close 为非支配解集中每个点到其他点的距离和最小的点,和最大的点即为当前非支配解集最稀疏的方向点 rare。这样可以避免算法过早收敛,又能使一定数量的粒子保持探索能力,扩大在解空间的搜索范围。

计算非支配解集中的密集点、稀疏点流程如图 2 所示。

由于每一次跳变能改变两个维度的值,如果不加以控制,这将导致粒子群算法的过快收敛,会严重影响所得解。本文的改进粒子群算法采取了多种策略,既能避免算法过早收敛,又能使一定数量的粒子保持探索能力,扩大在解空间的搜索范围。

迭代的控制流程如图 3 所示。

3 实验结果与分析

笔者在 Intel^(R) Core™ 2 CPU 4300 @ 1.80 GHz 1.19 GHz,

3.50 GB 内存的 PC 机上,运用 C++ 语言在 Visual Studio 2008 环境下进行了模拟实验,对本文提出的改进粒子群算法进行了性能分析。在实验中,用片上网络任务通信(图 4)作为测试实例,以一个 $4 \times 2 \times 2$ 的 3D-Mesh 为目标映射片上网络。将所有的任务用以下编码序列表示:0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15,同时,假设映射目标 3D-Mesh 的 CPU 核心的编码序列也为 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15。

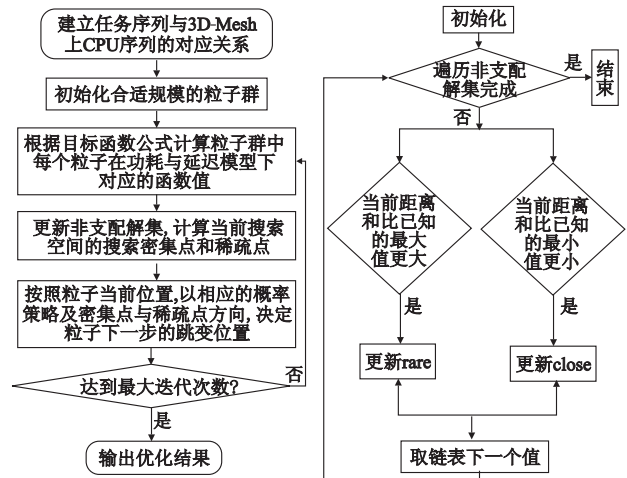


图1 基于改进粒子群算法的 Mapping 算法流程

图2 计算非支配解集中的密集点、稀疏点流程

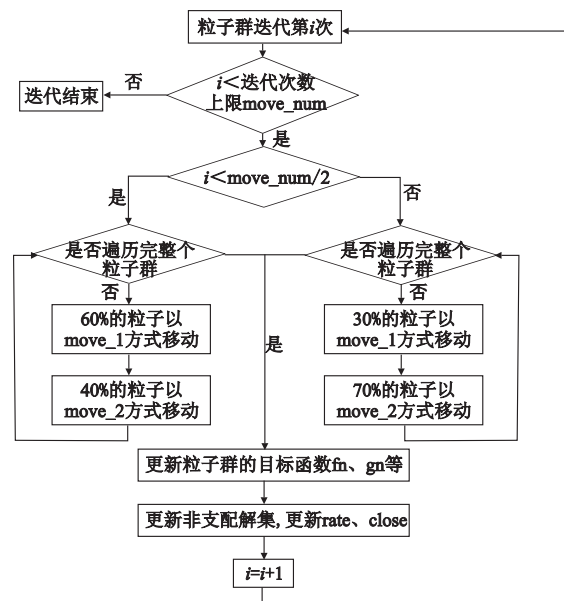
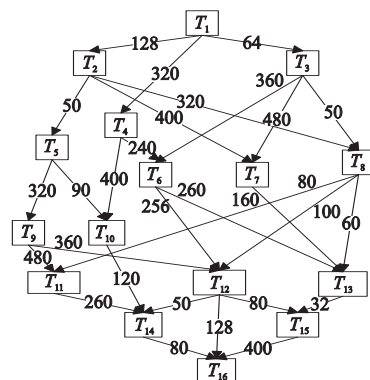


图3 迭代控制流程



3.1 面向运算规模的优化效果

首先对改进粒子群映射算法和随机单目标映射算法在各个映射模型下的性能进行了仿真实验。在算法中,选取粒子群规模定为 200,进行 100 轮搜寻循环,与同一计算规模的 20 000 次随机映射对比结果如表 1、2 所示。

表 1 同等计算规模下与随机映射平均值相比

性能参数	平均延时模型	最大延时模型	功耗模型	流量均衡模型
单目标随机映射 20 000 次平均值	224	67 458	24 743	26 765
本文多目标映射算法迭代 100 次	195	27 214	10 623	5 667
算法的优化比例/%	87	41	43	21

表 2 同等计算规模下与随机映射最好值相比

性能参数	平均延时模型	最大延时模型	功耗模型	流量均衡模型
单目标随机映射 20 000 次最好值	215	49 364	19 726	17 478
本文多目标映射算法迭代 100 次	195	27 214	10 623	5 667
算法的优化比例/%	91	55	54	32

从表 1 和 2 可以看到,本文算法在平均延时模型下的性能参数值与平均延时单目标下同等计算规模的随机映射的性能参数值相比,从 20 000 次平均值和最好值两个角度,前者分别降低到后者的 87% 和 91%;在最大延时模型下的性能参数也同样有所降低,分别降低到同等计算规模相应单目标映射模型约束下的 41% 和 55%;在功耗模型下的性能参数降低到 43% 和 54%;在流量均衡模型下的性能参数降低到 21% 和 32%。由此可以看到,本算法在同等计算规模下对比随机映射在四个映射模型下均有较显著的性能提升,共计搜索了不到 2 万个解,即可得到较好优化效果的映射问题的优良解,只占映射问题实际解规模的百万分之一,因此本 Mapping 算法在降低运算规模方面具有很好的优化能力。

3.2 面向多目标优化要求的优化效果

改进粒子群映射算法引进了非支配解的概念,可以达到同时满足多目标优化要求的目的。仿真实验得到的单独在各个模型下和综合考虑三个模型的优化解分别如表 3 所示。

表 3 算法在各个模型下的最优解

模 型	最优解															
功耗模型	4	14	2	7	9	3	6	5	12	15	1	10	13	11	0	8
平均延时模型	4	13	2	12	9	1	5	3	7	8	10	6	11	15	14	8
流量均衡模型	4	14	2	7	9	3	6	5	12	15	1	10	13	11	0	8
综合多目标模型	5	4	8	7	11	9	15	0	6	2	14	1	10	13	12	3

实验结果显示:

a) 仅考虑功耗目标的优良解并以此解作为最终映射方法,则在平均延时模型下,比之解 4 13 2 12 9 1 5 3 7 8 10 6 11 15 14 8,其平均延时比前者优化了 37%;在最大延时模型下,比之解 5 4 8 7 11 9 15 0 6 2 14 1 10 13 12 3,其最大延时参数比前者优化了 14%。

b) 仅考虑功耗目标的优良解并以此解作为最终映射方法,则在最大延迟模型下,比之解 5 4 8 7 11 9 15 0 6 2 14 1 10 13 12 3,其最大延迟参数比前者优化了 38%;在流量均衡模型下,比之解 4 14 2 7 9 3 6 5 12 15 1 10 13 11 0 8,其最热点参数值比之前者优化了 40%。

c) 同时考虑了平均延时模型、功耗模型和最大延时模型的推荐最优解为 5 4 8 7 11 9 15 0 6 2 14 1 10 13 12 3,比之流量均衡模型下的最优解 4 14 2 7 9 3 6 5 12 15 1 10 13 11,性能参数只下降了 1.8%;比之平均延时模型下的最优解 4 13 2 12 9 1 5 3 7 8 10 6 11 15 14 8,平均延时性能参数只下降了 1.5%。

由此取单一目标下的最优解为最终映射方式时,在其他评估模型下的性能可能会很差,但是改进粒子群算法可以同时考虑多个评估参数,取得综合最优解。虽然所得解不是某个评估模型下的最优解,但该映射方法在所侧重的映射评估参数下的值比之相应的单独评估模型下的最优解只有很小的下降。

4 结 束 语

近年来,源于 CMP 芯片厂商的大力推动,国际上 CMP 片上网络的发展正在加速,国内的研究还处于起步阶段,特别是面向前沿的 3D 片上网络技术的研究,成果更是稀少。本文研究了面向 3D-Mesh 片上网络上复杂通信任务向 CMP 芯片内处理核心的映射问题,提出基于改进粒子群的 3D-Mesh CMP 片上网络映射算法,并利用改进后的例子群算法,以较高的效率,正确找到多目标优化要求下的映射结果。但是本算法是针对 3D-Mesh 结构的,且研究中采用的是 XYZ 路由算法,具有一定的限制性。如何解决动态路由策略下映射评估模型的优化,将是笔者以后着重研究的课题。

参考文献:

- [1] 梁军. 粒子群算法在最优化问题中的研究[D]. 桂林: 广西师范大学, 2008.
- [2] 刘炎华, 刘静, 赖宗声, 等. 基于遗传蚁群算法的片上网络映射研究[J]. 计算机工程, 2010, 36(22): 262-264.
- [3] 李东生, 刘琪. 面向通信能耗的 3D NoC 映射研究[J]. 半导体技术, 2012, 33(7): 504-507.
- [4] 孙榕, 林正浩. 基于遗传算法的 NoC 处理单元映射研究[J]. 计算机科学, 2008, 35(4): 51-53, 84.
- [5] 邓植, 顾华玺, 杨银堂, 等. 基于拓扑划分的片上网络快速映射算法[J]. 电子与信息学报, 2011, 33(12): 3028-3034.
- [6] 杨盛光, 李丽, 高明伦, 等. 面向能耗和延时的 NOC 映射方法[J]. 电子学报, 2008, 36(5): 937-942.
- [7] WANG Hang-sheng, ZHU Xin-ping, PEH L S, et al. Orion: a power-performance simulator for interconnection networks[C]//Proc of the 35th Annual IEEE/ACM International Symposium on Microarchitecture. 2002: 294-305.
- [8] BENINI L, MICHELI G D. Networks on chips: a new SOC paradigm[J]. IEEE Computer, 2002, 35(1): 70-78.
- [9] 徐欣. 片上网络路由算法和映射算法研究[D]. 西安: 西安电子科技大学, 2010.
- [10] GUERRIER P, GREINER A. A generic architecture for on-chip packet switched interconnections[C]//Proc of Design Automation and Test in Europe. 2000: 250-256.
- [11] SIOZIOS K, ANAGNOSTOPOULOS I, SOUDRIS D. A high-level mapping algorithm targeting 3D NoC architectures with multiple VDD[C]//Proc of IEEE Computer Society Annual Symposium on ISVLSI. 2010: 444-445.
- [12] FALKENSTERN P, XIE Yuan, CHANG Yao-wen, et al. Three-dimensional integrated circuits (3D IC) floorplan and power/ground network co-synthesis[C]//Proc of the 15th Asia and South Pacific Design Automation Conference. 2010. 2010: 169-174.
- [13] CHEN H T, LIN Hong-long, WANG Zi-cheng, et al. A new architecture for power network in 3D IC[C]//Proc of Design, Automation & Test in Europe Conference & Exhibition. 2011: 1-6.
- [14] 林剑硕. NOC 映射方法和路径分配研究[D]. 上海: 上海大学, 2009.