

基于拥塞和热点感知的低延时片上网络路由器设计

刘炎华¹, 石世领¹, 孙海燕¹, 孙 玲¹, 黄新明^{1,2}

(1 南通大学电子信息学院, 江苏 南通 226019; 2 伍斯特理工学院电气工程与计算机系, 伍斯特 01609)

摘 要: 为了改善片上网络路由器的通信延时, 提高路由器的可靠性, 通过改良路由器架构、优化路由器关键路径, 结合设计的基于拥塞和热点感知路由算法, 提出了一种低延时片上网络路由器设计方法。相同配置下的综合结果表明本文设计的路由器与 CONNECT 路由器相比具有工作速度快、传输延时低的优点; Hotspot 负载模型下的测试结果显示该路由算法明显优于传统的超前 XY 路由算法。

关键词: 片上网络; 路由器; 低延时; 热点

中图分类号: TP302

文献标识码: A

文章编号: 1000—7180(2018)06—0122—06

Congestion and Hotspot Aware Router Design for Low Latency Network-on-Chip

LIU Yan-hua¹, SHI Shi-ling¹, SUN Hai-yan¹, SUN Ling¹, HUANG Xin-ming^{1,2}

(1 School of Electronic Information, Nantong University, Nantong 226019;

2 Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester 01609)

Abstract: To decrease the communication latency and improve the reliability of the routers in the Network-on-Chip, a low latency router based on congestion- and hotspot-aware routing algorithm is proposed by building a novel virtual channel router architecture and optimizing its critical path. The experimental results show that the proposed router can achieve faster operating frequency and lower transmission latency when compared with CONNECT router. Under Hotspot traffic, our routing algorithm is obviously superior to the look-ahead XY routing.

Key words: network on chip; router; low latency; hotspot

1 引言

随着半导体工艺技术进步和系统多核化发展, 多核芯片的片内通信技术一直是人们关注的焦点之一。与传统总线相比, 片上网络(Network on Chip, NoC)具有可扩展性强、灵活性高等特点, 被认为是解决多核互连的有效方案^[1]。而路由器作为提供 NoC 通信的实体, 其架构设计对片上网络的性能至关重要。

通常, 一个性能优良的路由器在达到较高工作频率的同时, 还要实现尽可能低的单跳传输延时(数据包通过单个路由器的延时)和多跳传输延时

(数据包完成传输过程的延时)以满足数据的低延时传输特性。然而, 这两方面的要求往往相互牵制。此外, 在高负载情况下的 NoC 网络中, 某些路由器可能由于长时间处理大量的数据包而形成热点(局部发热过高)的情况, 影响了系统的可靠性。

为了减少路由器的单跳通信延时, 基于虚通道的片上网络路由器在过去十年得到了许多改进。

文献[2]实现了将超前路由与投机性方法结合使用的方案, 这样超前路由计算、虚通道分配与交叉开关分配可并行处理, 流水线架构缩减到两级。

文献[3]提出了一种快速虚通道的路由器架构, 微片通过该路径可以直接旁路掉源、目的节点间的

收稿日期: 2017—09—07; 修回日期: 2017—11—01

基金项目: 国家自然科学基金面上项目(61571246); 江苏省高校自然科学研究重大项目(16KJA510006); 江苏省科技厅重点研发计划项目(BE2016007—2); 南通市科技应用基础研究计划项目(GY12015004)

中间路由器,但是这种方法不能很好的适应相邻路由器间的通信.

文献[4]通过并行处理虚通道分配、交叉开关分配和数据传输阶段,实现了一个两级流水的虚通道路由器架构.但是这种两级流水的路由器架构,其关键路径延时长,且由于虚通道只能存储相应类型的数据包,网络容易陷入拥塞状态.

文献[5]以一种非投机性方式实现了虚通道与交叉开关并行分配的路由架构,路由器流水线降为两级.但是,这种路由器的关键路径延时较大,且虚通道不能同时存储多个数据包,利用率不高.

上述研究的路由器采用不同方法,缩短了路由器的流水级数,但它们都没有考虑对路由器关键路径的优化,适用范围局限.此外,上述路由器仅使用维序路由算法,缺乏应对不同网络负载和减少热点状况的能力,增加了数据在网络中传输时的等待时间.因此,本文目标在于提高路由器适应性的同时,进一步优化其路径延时.

2 路由器架构

本文在现有相关研究基础上,提出了一种可配置的低延时片上网络路由器设计方案,设计出一种

基于拥塞与热点感知技术的两级流水低延时片上网络路由器,旨在兼顾路由器工作频率的同时,减少路由器的单跳传输延时和多跳传输延时.

通过分析典型的虚通道路由器发现,典型虚通道路由器中虚通道分配模块电路结构复杂,关键路径延时高,且对系统性能没有显著提升.因此本文设计了一种虚通道/交叉开关分配策略,剔除了虚通道分配阶段,缩减了路由器流水级数.

此外,设计采用了超前路由计算技术,进一步缩减路由器流水线到两级,并且通过增加路由器拥塞与热点感知功能,增强了路由器对网络不同情形的自适应性.

图 1 给出了本文设计的改进后的路由器顶层架构,该架构路由器由 P 个虚通道(对应 2D-Mesh 网络架构, P 的数量可配置为 5)、 P 个输入端口模块、 P 个超前路由计算模块、1 个下游虚通道状态记录模块、1 个热点检测模块、1 个交叉开关分配模块和 1 个交叉开关组成.为了验证所设计的路由器性能,本文定义路由器端口数为 5、虚通道数为 4,且用其组成了一个 2D-Mesh 5×5 的片上网络,该片上网络使用的微片格式如图 2 所示.路由器的数据传输过程及各主要模块的设计原理详述如下.

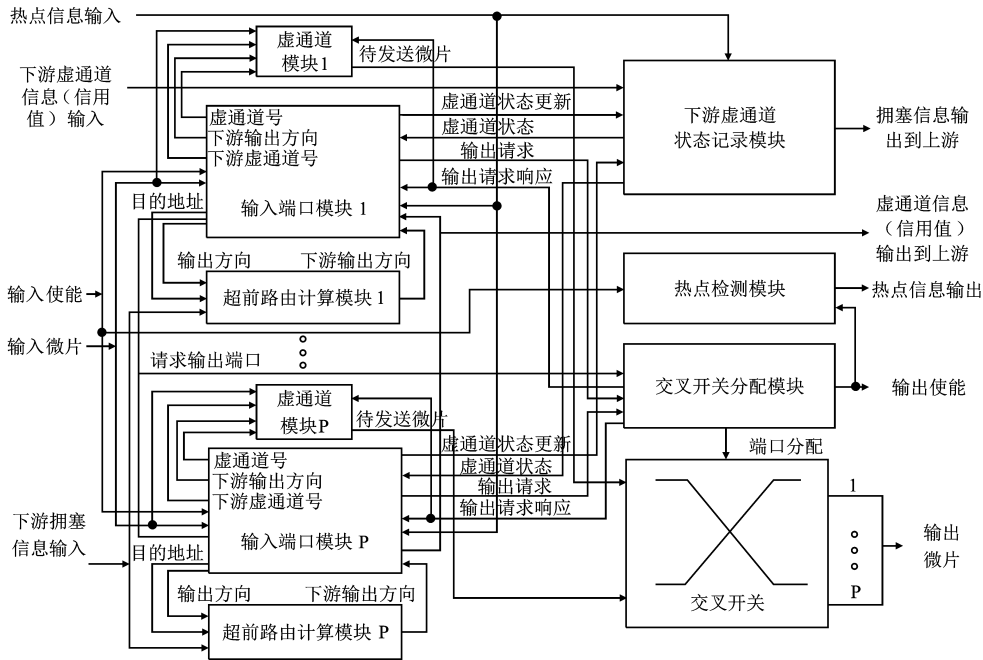


图 1 片上网络路由器架构

2.1 路由器工作流程

本设计中,微片的传输阶段如下.
第一阶段 接收微片并存储.当微片到达某一路由器时,输入端口模块首先根据微片中的标识字

段将微片暂存到对应的虚通道中.
第二阶段 完成超前路由计算并判断输出条件.当缓存在虚通道中的微片继续向下传输时,需要完成以下工作:

		37	36	35:32	31:29	28:26	25:23	22:20	19:17	16:0
头微片	2b10	头标识	尾标识	虚通道号	输出方向	X_目的结点	Y_目的结点	X_源结点	Y_源结点	未使用
数据微片	2'b00	37	36	35:32	31:0 数据负载					
		头标识	尾标识	虚通道号						
尾微片	2'b01	37	36	35:32	31:0 数据负载					
		头标识	尾标识	虚通道号						

图 2 微片格式

(1)下游虚通道状态记录模块为每个输出端口提供一个下游路由器中可供使用的候选虚通道;

(2)超前路由计算模块根据接收到的信息计算出该微片到达下游路由器的输出方向;

(3)对暂存在虚通道中的微片进行请求屏蔽.

第三阶段 输出权限仲裁. 交叉开关分配模块对输出请求进行仲裁.

第四阶段 更新状态信息. 下游虚通道状态记录模块更新下游路由器的虚通道使用状态, 同时热点检测模块对当前路由器的热点状态进行更新.

第五阶段 输出微片. 输入端口模块将待传输微片送到交叉开关, 然后交叉开关根据端口分配信号将各个微片传输到相应的输出端口.

2.2 热点模型

每个微片通过路由器和链路到达下游路由器都会消耗一定的能量. 根据 2.1 描述的微片传输过程可知, 其能耗主要包括:接收微片能耗 $E_{接收微片}$, 路由计算能耗 $E_{路由计算}$, 请求屏蔽能耗 $E_{请求屏蔽}$, 开关分配能耗 $E_{开关分配}$, 状态记录能耗 $E_{状态记录}$ 和发送微片能耗 $E_{发送微片}$. 相比其他操作, 路由计算阶段所消耗的能量可忽略. 因此路由器能量消耗评估如下式所示.

$$E_{微片} = E_{接收微片} + E_{请求屏蔽} + E_{开关分配} + E_{状态记录} + E_{发送微片} \quad (1)$$

式中 $E_{微片}$ 代表每个微片通过一个路由器所消耗的能量.

路由器的发热量则与其处理的数据量相关, 如果一个路由器长时间处于忙碌状态, 该节点的温度就会大大高于网络中其他节点的温度, 造成节点老化甚至损坏. 本文将此路由节点定义为热点. 按照这种定义, 本文提出了一种热点模型:

$$HT \geq Th \quad (2)$$

式中, HT 表示路由器在 T 时间内所消耗的能量; Th 表示路由器在 T 时间内允许消耗的能量.

这样, 路由器 r 在 T 时间内所产生的热量 Hr 可描述为:

$$Hr = \sum_{i=1}^T (E_{微片} \times Nr) \quad (3)$$

式中, Nr 表示路由器 r 在单位时间内发送的微片数. 若路由器 r 在时间 T 内发送的微片数大于该时间内其允许处理的最大微片数 Nth , 则若该路由器为热点, 须满足条件如下式:

$$\sum_{i=1}^T Nr \geq Nth \quad (4)$$

式中, $Nth = Th/E_{微片}$. 为此, 本文相比典型虚通道路由器, 增加了一个热点检测模块(如图 1 所示), 该模块不会对路由器流水线产生影响.

2.3 虚通道/交叉开关分配

与典型虚通道路由器的分配策略不同, 本设计中虚通道与交叉开关的分配过程如图 3. 输入端口模块向交叉开关分配模块发送输出请求; 同时下游虚通道状态记录模块为请求的输出端口提供一个候选虚通道; 当头微片赢得交叉开关分配的输出口后, 候选虚通道也被分配给该头微片使用. 这种分配策略须确保微片所请求的下游路由器中存在可用虚通道, 为此, 本设计在微片输出请求发送到交叉开关分配模块前, 采取了请求屏蔽技术. 即微片拟经过的下游路由器中虚通道无缓存或者没有可用虚通道, 则该微片输出请求直接被屏蔽. 此外, 如果下游路由器为热点, 输出请求同样被屏蔽, 可见, 请求屏蔽技术还有效缓解了某个路由器长时间处于热点的状况.

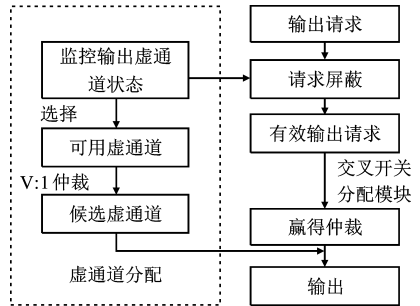


图 3 下游虚通道状态记录模块

虚通道分配过程完成后, 头微片获得的虚通道

会一直为其后的数据微片和尾微片保留,而当尾微片传输结束,意味着当前数据包已经发送完毕,该虚通道就变为可用状态,此时其会被分配给其他数据包使用。所以相比现有路由器,本文中的虚通道使用方法允许虚通道同时存储多个数据包。

为进一步减少路由器的单跳传输延时,本设计还优化了路由器的关键路径。图4中,状态选择、请求屏蔽和交叉开关分配组成了路由器的关键路径。为了降低请求屏蔽所增加的延时,本设计提前一个时钟将指定的下游虚通道状态保存起来,而在下一时钟直接读取该状态到请求屏蔽模块,这就消除了请求屏蔽过程引起的状态选择延时。而在交叉开关分配模块中,根据已有研究结果^[6],选择了拥有更低关键路径延时、面积和功耗开销的输入优先分离分配器。并且,设计还选用了一种快速仲裁器来^[7]进一步减少分配延时。相比典型虚通道路由器,本设计虽然增加了请求屏蔽模块,但通过上述优化却有效减小了整个关键路径延时。

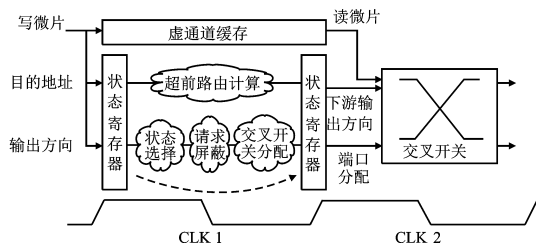


图4 路由器关键传输路径示意图

3 拥塞与热点感知的路由算法

路由算法在很大程度上影响着片上网络路由器的性能,现有的低延时片上网络路由器大都使用维序路由算法,难以应对网络中的不同负载任务。另外,传统自适应路由算法一般将下游路由器中空闲虚通道数、虚通道缓存数或者输出端口请求数作为拥塞标准,然而这些评判标准仅仅反映了下游路由器中虚通道的当前状态。当网络长时间处于繁忙状态时,路由器有极大可能成为热点。因此,本设计将热点参数也作为路由指导信息,该信息与拥塞信息相结合,能更好地指导微片的传输,减少热点的产生,增加系统的可靠性。

为了实现本文所提的路由算法,在路由器架构中添加了用于拥塞和热点感知的控制逻辑,这些控制逻辑的引入并不会增加路由器的流水阶段,也没有加重路由器关键路径的延时。设计在超前XY路由算法的基础上,实现了一种基于拥塞与热点感知

的超前路由算法,该算法降低了数据包的多跳传输延时,并能平衡网络负载,减少网络热点产生,算法伪代码如下所示:

输入. 源路由器的坐标 (X_S, Y_S) , 目的路由器的坐标 (X_D, Y_D) , 表示East、South、West和North方向路由器内虚通道使用程度的 C_E, C_S, C_W 和 C_N , 以及表示各方向热点信息的 H_E, H_S, H_W 和 H_N 。

输出. 输出方向 EOutputE。

EStep1. 路由器接收到微片, 计算 $\Delta X = X_D - X_S$, $\Delta Y = Y_D - Y_S$ 。

Step2. 如果 $\Delta X = 0, \Delta Y = 0$, 则 Local 为输出方向; 否则, 执行下一步。

Step3. 如果 $\Delta Y = 0$, 若 $\Delta X > 0$, 则 East 为输出方向; 若 $\Delta X < 0$, 则 West 为输出方向; 如果 $\Delta X = 0$, 若 $\Delta Y > 0$, 则 South 为输出方向; 若 $\Delta Y < 0$, 则 North 为输出方向; 否则, 执行下一步。

Step4. 如果 $\Delta X > 0, \Delta Y > 0$, 则 East 或 South 为候选方向; 否则, 执行步骤 5。

Step4. 1. 比较 H_E 与 H_S , 若 $H_E = 1, H_S = 0$, 则 South 为输出方向; 若 $H_E = 0, H_S = 1$, 则 East 为输出方向; 否则, 执行下一步。

Step4. 2. 若 $H_E = H_S$, 比较 C_E 与 C_S 。当 $C_E = C_S$ 时, 任意选择 East 或 South 为输出方向; 当 $C_E > C_S$ 时, 选择 South 为输出方向; 当 $C_E < C_S$ 时, 选择 East 为输出方向。

Step5. 如果 $\Delta X > 0, \Delta Y < 0$, 则 East 或 North 为候选方向; 否则, 执行步骤 6。

Step5. 1. 比较 H_E 与 H_N , 若 $H_E = 1, H_N = 0$, 则 North 为输出方向; 若 $H_E = 0, H_N = 1$, 则 East 为输出方向; 否则, 执行下一步。

Step5. 2. 若 $H_E = H_N$, 比较 C_E 与 C_N 。当 $C_E = C_N$ 时, 任意选择 East 或 North 为输出方向; 当 $C_E > C_N$ 时, 选择 North 为输出方向; 当 $C_E < C_N$ 时, 选择 East 为输出方向。

Step6. 如果 $\Delta X < 0, \Delta Y > 0$, 则 West 或 South 为候选方向; 否则, 执行步骤 7。

Step6. 1. 比较 H_W 与 H_S , 若 $H_W = 1, H_S = 0$, 则 South 为输出方向; 若 $H_W = 0, H_S = 1$, 则 West 为输出方向; 否则, 执行下一步。

Step6. 2. 若 $H_W = H_S$, 比较 C_W 与 C_S 。当 $C_W = C_S$ 时, 任意选择 West 或 South 为输出方向; 当 $C_W > C_S$ 时, 选择 South 为输出方向; 当 $C_W < C_S$ 时, 选择 West 为输出方向。

Step7. 如果 $\Delta X < 0, \Delta Y < 0$, 则 West 或 North 为候选方向。

Step7. 1. 比较 H_W 与 H_N , 若 $H_W = 1, H_N = 0$, 则 North 为输出方向; 若 $H_W = 0, H_N = 1$, 则 West 为输出方向; 否则, 执行下一步。

Step7. 2. 若 $H_W = H_N$, 比较 C_W 与 C_N 。当 $C_W = C_N$ 时, 任意选择 West 或 North 为输出方向; 当 $C_W > C_N$ 时, 选择 North 为输出方向; 当 $C_W < C_N$ 时, 选择 West 为输出方向。

4 实验结果与分析

本文基于 Verilog HDL 实现了所设计的片上网络路由器,通过 Modelsim 软件完成了路由器的功能验证,并在 Quartus II 软件环境下选用型号为 EP3SL340F1760C2 的 FPGA 进行综合与性能分析。

本文另选取了两款低延时片上网络路由器作为对比:① CONNECT 是面向 FPGA 所设计的一款开源片上网络软核^[8];② 文献[9]实现的一款针对 FPGA 的通用低延时片上网络路由器。

为了准确地评估三种路由器的性能指标,分别将它们按如表 1 所示的参数进行配置,由于路由器②不支持虚通道技术,将三者分两组进行对比。表 2 列出了①和③在 5×5 网络规模下的硬件开销和最大工作频率(EP3SL340 上总共可用资源 ALUTs: 270 400, Registers: 270 400, Memory bits: 16 662 528)。

表 1 mesh NoC 配置基准参数

不同网络配置	①	②	③Our design
Topology	mesh	mesh	mesh
Virtual Channel	yes	no	yes
Numbers of VCs per port	4	no	4
Input VC buffer size	4	no	4
Router pipeline latency	1 cycle	2 cycles	2 cycles
Payload Width	32bits	32bits	32bits

表 2 5×5 规格下路由器①与③的硬件性能

Router	Logic Utilization	ALUT/Register/ Memory	Fmax/MHz (0℃/85℃)
①	40%	51541/76901/0	94/87
③	28%	56579/35052/57120	200/185

从表 2 可以看出,通过对所设计路由器关键路径进行了一定的优化,本文路由器的的工作频率与路由器①大约维持着一个两倍的关系。这样,本文路由器的单跳延时与路由器①相比基本维持。而本文所设计的路由器,依靠其两级流水线的优势,吞吐量与①相比有极大的提升。另外路由器①的虚通道模块使用的是 FPGA 中的 LAB 资源实现的,造成其总资源消耗多,而本文路由器利用 FPGA 中的大量存储资源来实现虚通道缓存,所以整体资源消耗更平均,有利于系统的扩展。

图 5 给出了将本文设计的路由器配置为与文献[9]相同情形时路由器的综合性能结果。由图可见,当端口数为 5 时,两者的最大工作频率相当;但当

端口数为 7 时,本文设计的路由器最大工作频率提高了约 10%。可见,与文献[9]相比,本文设计能灵活地适用于不同网络规格下,可扩展性强。此外,文献[9]不支持虚通道技术。

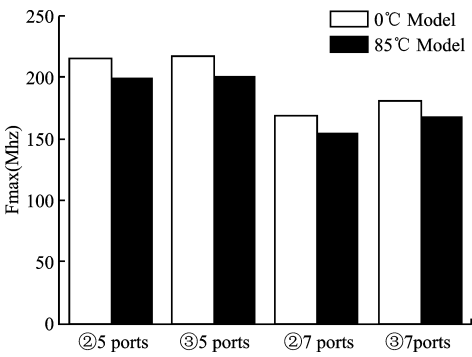


图 5 路由器②和③在不同工作温度下的最大工作频率

在分析了单个路由器性能的基础上,利用本文设计的路由器进一步构建了一个 5x5 规格的片上网络。作为比较,还设计了一个基于传统超前 XY 路由算法实现的路由器,并构建了同样规格的片上网络。为了对设计进行有效的性能评估,设置在 Hotspot 非均匀负载模型来对片上网络进行测试,测试结果分别如图 6 和图 7 所示。从图 6 可以看出,本文设计的路由算法在非均匀负载下的延时性能优于超前 XY 路由算法。当注入率为 42%时,其延时减少了 25%。观察图 7 可见,在注入率为 42%情况下,本文设计的路由算法几乎能避免热点产生,大幅度改善了片上网络中的热点状况。

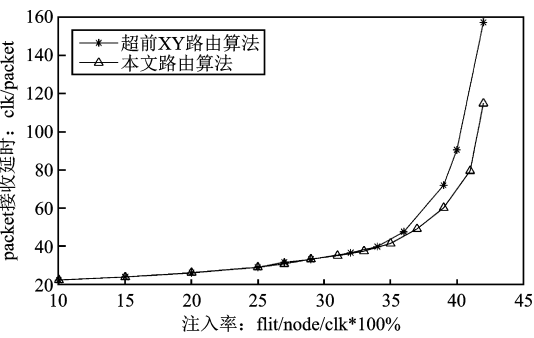


图 6 Hotspot 负载模型下接收延时

5 结束语

为了提高设计的灵活性和可扩展性,本文实现了一种高度可配置的虚通道路由器模型。通过设计一种新的虚通道/交叉开关分配方式,使路由器能够支持在较高的频率运行、同时拥有低单跳传输延时的特性;所设计的请求屏蔽技术不仅能提高对虚通道的使用率,还可以暂停微片发送到热点,增加

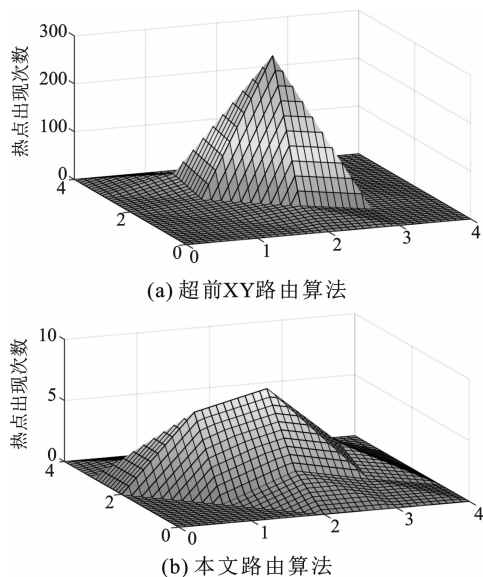


图7 两种算法在 Hotspot 负载模型下的热点分布情况

系统的可靠性;此外,为了降低热点出现频率,本文还提出了一种基于拥塞和热点感知的超前路由算法,使微片在传输时能选择拥塞程度更小的非热点路径,减少微片的多跳传输延时,降低路由器变成热点的情况。这些特性使得所设计的片上网络能在合理的资源消耗范围内,工作在更高的频率,当网络的注入率上升时,系统能在改善热点情况的情况下,提供更低的通信延时。

参考文献:

- [1] Kreider M, Bar R, Beck D, et al. Open borders for system-on-a-chip buses: A wire format for connecting large physics controls[J]. Physical Review Special Topics-accelerators and Beams, 2012, 15(8):1-10.
- [2] Mullins R, West A and Moore S. The Design and Implement of a Low Latency On Chip Network[C] // Proceedings of the Asia and South Pacific Design Automation Conference. Los Alamitos: IEEE Computer Society Press, 2006: 164-169.
- [3] Krishna T, Kumar A, et al. NoC with Near-Ideal Express Virtual Channels Using Global-Line Communication[C] // Proceedings of the 16th Annual Symposium on High-Performance Interconnects. Los Alamitos:

IEEE Computer Society Press, 2008: 11-20.

- [4] 张庆利,刘成,肖立伊,等. 支持确定性/自适应路由的低延时片上路由器设计[J]. 计算机辅助设计与图形学学报, 2009, 21(12): 1706-1714.
- [5] Ye L, McCanny J, Sezer S. Exploring Virtual-Channel Architecture in FPGA based Networks-on-Chip[C] // Proceedings of the 24th IEEE International System on Chip Conference. Los Alamitos: IEEE Computer Society Press, 2011: 302-307.
- [6] Becker D U. Efficient microarchitecture for network-on-chip routers [D]. California: Stanford University, 2012.
- [7] Dimitrakopoulos G, Chrysos N, Galanopoulos K. Fast arbiters for on-chip network switches[C] // Proceedings of the 26th IEEE International Conference on Computer Design. Los Alamitos: IEEE Computer Society Press, 2008: 664-670.
- [8] Papamichael M K, Hoe J C. The CONNECT Network-on-Chip Generator [J]. Computer, 2015, 48(12): 72-79.
- [9] Ye L, McCanny J, Sezer S. Generic Low-Latency NoC Router Architecture for FPGA Computing Systems [C] // Proceedings of the 21st International Conference on Field Programmable Logic and Applications. Los Alamitos: IEEE Computer Society Press, 2011: 82-89.

作者简介:

刘炎华 男,(1980-),博士,副教授.研究方向为高性能SoC/NoC体系结构设计及测试。

石世领 男,(1990-),硕士研究生.研究方向为片上网络路由器设计与优化。

孙海燕 男,(1977-),博士,副研究员,主要研究方向为系统级封装设计与测试。

孙玲(通讯作者) 女,(1976-),博士,教授,硕士生导师.研究方向为VLSI电路设计、集成电路封装与测试。

E-mail:sun. l@ntu. edu. cn.

黄新明 男,(1974-),博士,教授,博士生导师,主要研究为电路与系统、无线通信、可重构计算、视频图像处理和云计算安全。