文章编号：1674-2974（2023）08-0141-06 DOI：10.16339/j.cnki.hdxbzkb.2023289

一种高吞吐低延迟片上互连网络路由器

李晋文†，申慧毅，齐树波
（国防科技大学 计算机学院，湖南 长沙 410073）

摘 要：本文提出了一种用于片上互连网络的低延迟高吞吐量动态虚拟输出队列路由器，该路由器可以利用前瞻路由计算和虚拟输出队列方案将路由器延迟减低到两个周期.仿真结果表明，与虫孔路由器和虚通道路由器相比，4×4 网格上的网络吞吐量分别提高了 46.9% 和 28.6%，并且在相同输入加速比下，性能比双缓冲虚通道路由器要高 1.9%.在随机合成流量下，片上网络的零负载延迟也分别降低了 25.6% 和 41%.设计实现结果表明，路由器的工作频率可以达到 2.5 GHz.

关键词：片上网络；路由器；吞吐量；延迟
中图分类号：TN913.3 **文献标志码**：A

A High-throughpur Low-latency Router for On-chip Interconnect Networks

LI Jinwen†，SHEN Huiyi，QI Shubo
（School of Computer Science，National University of Defense Technology，Changsha 410073，China）

Abstract：A low-latency high-throughput Dynamic Virtual Output Queues Router for On-chip interconnect networks is proposed in this paper，which can reduce the router latency to two cycles by leveraging look-ahead routing computation and virtual output queues scheme. The simulation results show that，compared with the wormhole router and virtual-channel router，the network throughput on a 4×4 mesh increases by up to 46.9% and 28.6%，respectively，and outperforms doubled buffer virtual channel by 1.9% under the same input speedup. Under random synthetic traffic，the zero-load-latency of the network on chip is also reduced by 25.6% and 41%，respectively. Synthesis results indicate the frequency of router can reach 2.5 GHz.

Key words：on-chip network；router；throughput；latency

随着半导体技术的飞速发展，越来越多的处理器核（多核和众核）集成在单个芯片上，而随着 MOS 管尺寸的不断缩小，门级电路延迟在不断缩小，全局互连线的延迟相对于 MOS 管延迟还在不断增加.微

* **收稿日期**：2022-11-03
基金项目：HPCL 国家重点实验室基金项目（202101-02）；国家自然科学基金资助项目（60873212），National Natural Science Foundation of China（60873212）
作者简介：李晋文（1975—），男，山西武乡人，国防科技大学研究员，博士
† 通信联系人，E-mail：lijinwen@sina.com

处理器体系结构设计的重点正在从以提高计算为中心的单核能力设计转向以互连通信为中心的多核设计.由于互连延迟可预测、设计复杂度比较低、易扩展性和结构规整,片上网络已成为CMP和MPSoC中片上众核互连最有前途的选择[1].其中2D mesh互连网络已广泛应用于许多原型芯片,如Intel 80核Teraflop、Tilera 64核和TRIPS[2-4].

片上网络的概念来源于多处理器间互连网络,但实际与多芯片间互连网络有着许多不同的特点.最重要的一点,芯片内互连线和引脚比芯片间网络中的互连线和引脚资源更丰富[1].然而,片上网络中缓冲buffer容量不足.网络的延迟对实际多核的计算性能有很大影响.当路由器的每跳延迟从一个周期增加到五个周期时,全系统的性能将下降10%[5].基准的虚通道路由器的流水线级数为4.近年来,业界提出了几种新型架构的低延迟路由器,包括推测虚通道路由器[6]、采用虫孔交换的两虚通道结构路由器[7]、混合电路交换路由器[5]、带bundle的两周期路由器[8]、组合型两周期路由器[9]、无缓存片上路由器[10]、基于时间序列开关分配路由器[11]以及关键路径延迟只有35个FO4[12]的单周期路由器(FO4是指一个反相器驱动四个相同尺寸反相器产生的延迟,高性能微处理器的周期一般约为20个FO4).

缓冲buffer的实现对互连网络的性能至关重要.缓冲buffer可以用寄存器或SRAM来实现.在芯片中,通常缓冲buffer的容量相对较小,因此使用低延迟的寄存器实现更为有利,而使用SRAM会存在较大的地址译码延迟以及存储阵列访问延迟,这些延迟与全局位线相关;此外还能节省位线预充电功耗[13].在标准的虚通道路由器中,每个虚通道都需要自带缓冲buffer,一个虚通道无法使用其他虚通道的缓冲buffer[14].DAMQ路由器设立了5个缓冲buffer队列,每个队列对应一个虚通道,多出的一个队列作为共享缓冲buffer,一个报文flit从到达到离开路由器需要3个时钟周期[15].VichaR路由器能够根据数据流量(traffic)来调节和分配每个物理通道的虚通道和缓冲buffer数量,并使用复杂的VC控制表来管理报文flit,能够有效提高缓冲buffer的使用效率,其缺点是路由器延迟会达到四个时钟周期.当路由器中发生拥塞时,无论是采用基于信用还是基于开关的流控策略,通道流水线中的缓冲buffer都不能用于缓冲flit.iDEAL路由器提出用中继器(repeater)电路来缓冲flit报文[16],然而中继器存在较大漏流问题,会导致不可靠.

本文提出了一种新型的两周期路由器——动态虚通道输出队列路由器(DVOQR),采用多端口缓冲buffer和虚拟输出队列来消除虚通道路由器中的分配站(allocation stage).采用Ready/Valid握手机制来控制路由器之间的flit流,在这种策略下,流水线通道中的存储器可以用于缓冲flit报文.

本文其余部分组织如下,第1节介绍了路由器的微架构.第2节给出了路由器的具体设计实现.第3节分析了模拟结果.最后,第4节对本文工作进行了简要总结.

1 路由器微架构

1.1 DVOQR路由器微架构

本文提出了一种新型动态虚通道输出队列路由器(DVOQR),其微架构如图1所示.路由器包括P个输入端口和P个输出端口.对于二维mesh网络,P=5;一个端口连接到本地处理器(核),其他端口连接到相邻路由器.输入单元由三个主要模块组成:集中动态缓冲器(Unified Dynamic Buffer,UDB)、集中动态缓冲分配器(Unified Dynamic Buffer Allocation,UDBA)、P个虚拟输出地址队列(Virtual Ouput Address Queue,以下简称VOAQ).输出端口包括一个P选1的仲裁器和一个P输入的多路复用器.由多个flit组成1个数据报文,存储在同一FIFO队列中,路由到同一输出端口.每个输入端口有P个FIFO队列,它们共享一个UDB并各自带一个私有的VOAQ.每个FIFO中flit的地址存储在虚拟输出地址队列(VOAQ)中.这样一来,就可以有效消除队列头阻塞(HOL)延迟问题[17].

芯片间网络路由器中的缓冲buffer一般使用SRAM来实现.大容量的多端口SRAM存储器由于需要较大的面积开销、较高的功耗和访问延迟而难以实现,而使用小容量的寄存器来实现多端口缓冲器buffer要容易得多.受片上资源的限制,UDB用低延迟的多端口寄存器实现,具有1个写端口和P个读端口.每个读端口对应1个FIFO队列.尽管使用多个端口会导致面积开销增加,但可以消除虚通道路由器流水线的分配站.

连接到输出端口的CDB,由CDB控制器和两项

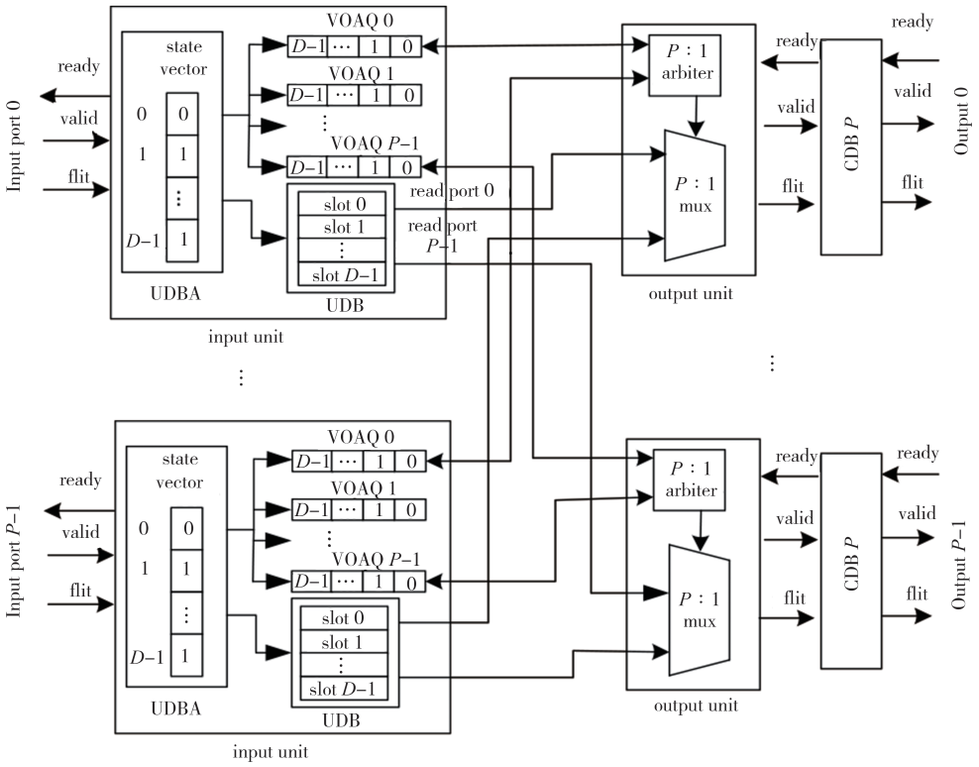


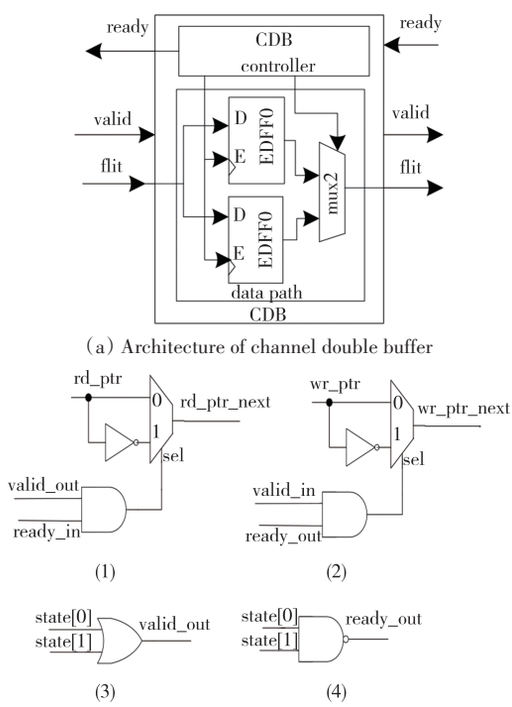
图 1　DVOQR 路由器微架构

Fig.1　Microarchitecture of DVOQR

寄存器组成,如图 2(a)所示.其中一个寄存器负责接收来自路由器的 flit,而另一个寄存器负责将 flit 发送到下一个路由器,一收一发.在下一个周期中,两个寄存器交换收发功能.因此 CDB 可以同时接收和发送 flit,可以避免流水线产生气泡.图 2(b)给出了 CDB 控制器的实现电路.state[1:0]表征两个寄存器的状态.读指针 rd_ptr 对应发送寄存器,写指针 wr_ptr 对应接收寄存器.当路由器之间的线延迟超过一个时钟周期时,可以插入多个 CDB.

UDBA 用于为队列分配时隙或释放空时隙.使用状态向量来跟踪所有时隙的状态,1 表示时隙可用.当时隙分配给 flit 时,相应的位将被清掉.采用固定优先级仲裁器以简化分配逻辑,最低可用时隙将被分配最高的优先级.

设计了四个物理 VOAQ 来缓存同一队列中的 flit.当某一个 flit 注入 UDB 时,UDBA 负责将分配给它的时隙号写入对应的 VOAQ,该 VOAQ 还会保存该报文的路由信息以及 flit 类型.在 UDB 读操作之前,需要首先从 VOAQ 中读出 UDB 中 flit 的地址,这将增加 UDB 的访问延迟.本文设计了一种新颖的移位 FIFO,可以有效减少 UDB 的读延迟.图 3 给出了 VOAQ 的微架构,使用 one-hot 向量来指向 FIFO 的尾



(a) Architecture of channel double buffer

(b) Channel double buffer controller

图 2　通道的双缓存控制器

Fig.2　Channel double buffer controller

部,而第一项指向 FIFO 的头部.尾向量的宽度比 UDB 的深度 D 要大 1.当 tail_vector[0]为 1 时,FIFO 为空;而 tail_vector[D]等于 1 时,FIFO 为满.当头数

据离开队列时,VOAQ中的其他数据将向前移一位,而tail_vector将进行右移.当新数据到达时,数据将被添加到VOAQ的尾部,并且tail_vector左移1位.当新数据在同一时钟周期内到达和离开时,tail_vector将不发生移位.

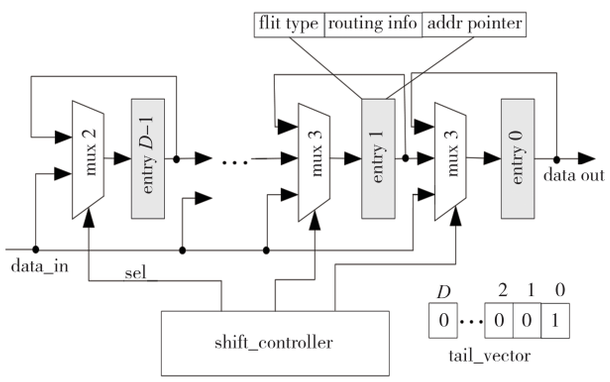


图3 VOAQ的微架构

Fig.3 Microarchitecture of virtual ouput address queue

DVOQR中的交换分配单元使用P个round-robin仲裁器实现.交换分配单元只需要一级仲裁,即可实现最大匹配,从而提高路由器吞吐量并降低分配延迟.

1.2 DVOQR流水线设计

DVOQR路由器的流水线由两站组成:flit交换站(Flit Switch,FS)和链路传输站(Link Traversal,LT).

FS站:完成交叉开关分配、前瞻路由计算、UDB读操作和Crossbar传输.其中交叉开关分配、前瞻路由计算和UDB读操作能够并行.当VOAQ的第一项是head flit报文片时,会为目的仲裁器产生一个请求信号.同时,发送VOAQ中的flit地址到UDB,启动读操作,根据报文的路由信息,采用维序路由算法进行路由的前瞻计算.如果请求未被批准,将在下一个周期中重试,而不需要再次读取flit报文.

LT站:在这一站中,flit通过物理链路发送并写入UDB,并根据FS站的前瞻路由计算结果,将分配给flit的地址写入VOAQ中.

1.3 流控机制

DVOQR使用了一种新的流控机制,称为ready-valid握手机制(handshake).ready输出表示UDB/CDB有可用的存储来接收flit报文.valid信号标识当前的flit报文是有效的.当ready和valid信号在同一个周期内有效时,说明flit报文已经提交.当下一级路由器发生拥塞时,链路上流水线中的CDB可以缓冲flit报文,这等效于增加了缓冲buffer容量.基于维序路由算法,这种流控机制可以有效避免死锁.

2 设计实现

基于RTL设计实现了用于片上2D mesh网络的DVOQR路由器,数据位宽128位,带有16项UDB,评估了路由器的性能和功耗,综合生成门级网表,并对时序进行了详细的分析.FS站和LT站的关键路径延迟分别为400 ps(11.4 FO4)和252 ps(7.2 FO4),该工艺下的FO4为35 ps.表1给出了路由器中各功能部件的面积和功耗.

表1 路由器中各功能部件的面积和功耗

Tab.1 Area and power consumption of each functional component

模块	组合逻辑面积/$(\mu m)^2$	时序逻辑面积/$(\mu m)^2$	总面积/$(\mu m)^2$	功耗/mW	数量/个
UDB	18 945	31 475	50 420	58.8	5
VOAQ	2 496	31 68	5 664	7.5	20
input port	29 731	44 093	73 824	89.3	5
output port	1 510	113	1 623	0.603	5
CDB	2 236	3 065	5 301	12.1	5
router	167,385	221,595	403,740	507.5	1

3 模拟结果

3.1 模拟方法

本文采用随机人工合成流量模型评估互连网络的性能.表2给出了模拟实验的参数设置.采用周期精确模拟器Booksim[14]来评估虫孔路由器(Wormhole Router,WH)和虚通道路由器(Virtual-channel Router ,VC).本文使用Verilog HDL设计实现了DVOQR的RTL模型.测试程序采用随机通讯的合成程序,进行了仿真模拟,预热时间为1万个时钟周期,测量时间为10万个时钟周期.

3.2 模拟结果分析

3.2.1 不同缓冲容量的影响

图4为带16项UDB的DVOQR路由器在随机流量负载下的平均延迟曲线.虫孔路由器和虚信道路由器中的输入缓冲buffer数量为16~64 flit.与其他两种路由器相比,DVOQR的吞吐量分别增加了33.2%和12%,而其他路由器缓冲buffer的容量是DVOQR的3倍.因此,DVOQR可以更有效地使用输入缓冲器.其中,三种路由器的零负载延迟分别为10.4、14.0和17.7.

表 2　模拟参数设置

Tab.2　Simulation parameter settings

network	4×4 mesh
路由算法	dimension-order routing
报文长度	four flits
流量注入	Bernoulli process
DVOQR路由器	two-stage pipeline, the depth of UDB is 16 for VOQ_16
虫孔路由器(WH)	three-stage pipeline, the depth of buffer is 16 for WH_16.
虚通道路由器(VC)	four-stage pipeline, the channel number is 4 and the depth of buffer in channel is 8 for VC_4×8.

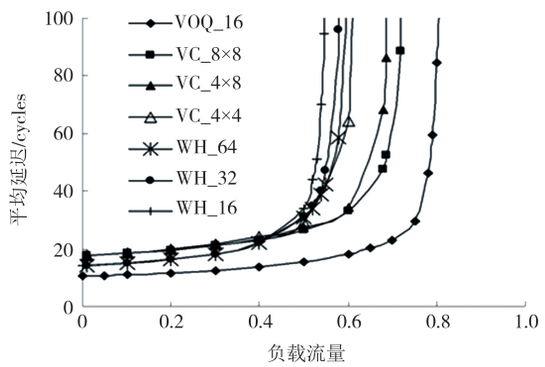


图 4　不同 buffer 容量的 DVOQR 路由器平均延迟

Fig.4　Average latency of DVOQR with different buffer capacities

3.2.2　相同输入加速比

UDB 有四个读端口,因此 DVOQR 的输入加速比是 4.图 5 给出了在随机流量负载相同输入加速比时的平均延迟曲线.与 VC_4×4 和 VC_4×8 相比,VOQ_16 的吞吐率分别增加 17.6% 和 1.9%,而 VC_8×8 和 VC_8×16 的吞吐率分别比 VOQ_16 要高 2.9% 和 7.5%.DVOQR 吞吐率比双缓冲虚通道路由器要高 1.9%.在相同的输入加速比下,采用动态缓冲 buffer 分配只需要一半的 buffer 容量就能达到相同的吞吐率.

3.2.3　UDB 深度的影响

图 6 给出了随机流量下 DVOQR 网络性能与 UDB 深度的相关性.2 项 UDB 的网络饱和点约为 50%,16 项 UDB 的饱和点可达到 82.4%.当 UDB 的深度大于 8 时,吞吐率的增加随着 UDB 深度的增加速度放缓.当注入流量小于 0.4 时,采用不同深度 UDB 的平均延迟几乎是相同的.可以根据网络流量打开或关闭一部分 UDB,这样可以有效减少缓冲 buffer 的漏流功耗.事实上,缓冲 buffer 产生的漏流功耗是整个 NoC 路由器漏流功耗的最主要来源.

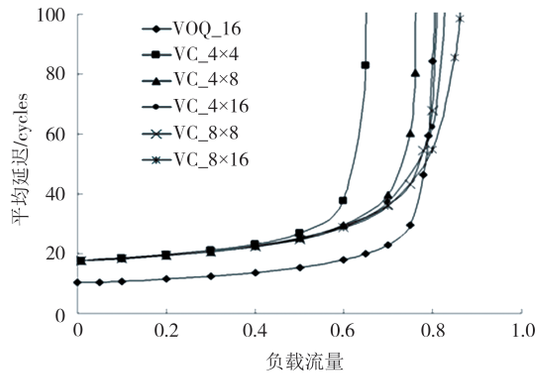


图 5　相同输入加速比下 DVOQR 平均延迟

Fig.5　Average latency of DVOQR under the same input acceleration ratio

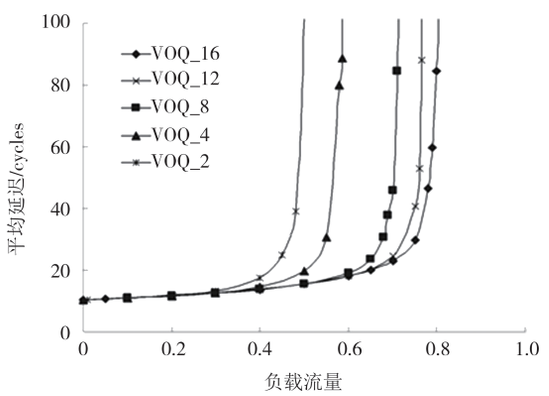


图 6　不同深度 UDB 的 DVOQR 的平均延迟

Fig.6　Average latency of DVOQR with different UDB

3.2.4　报文长度的影响

图 7 给出了随机流量下带 16 项 UDB 的 DVOQR 平均延迟与数据报文长度的关系,报文长度为 2~32 个 flit.吞吐率随着报文长度的增加而降低.报文长度为 32 flit 和 2 flit 网络的饱和点分别为 57.5% 和 87.5%.报文长度进一步增加将导致阻塞,因此需要占用更多的物理通道,而且竞争增加将导致更大的延迟.

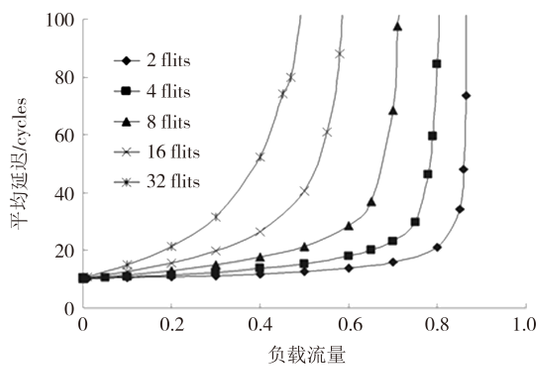


图 7　对应不同报文长度下 DVOQR 平均延迟

Fig.7　Average latency of DVOQR under different message lengths

4 结 论

本文提出了一种基于ready-valid握手流控策略的两级流水线片上互连网络路由器,该路由器采用维序路由可以避免死锁.与虫孔路由器和虚通道路由器相比,4×4 mesh网络中的网络吞吐量分别提高了46.9%和28.6%,并且在相同的输入加速比下,DVOQR路由器比双缓冲虚通道路由器性能提高了1.9%.综合结果表明,路由器的时钟频率可达2.5 GHz.

参考文献

[1] DALLY W J, TOWLES B. Route packets, not wires: on-chip interconnection networks [C]//Proceedings of the 38th Design Automation Conference . Las Vegas, NV, USA: IEEE, 2005: 684-689.

[2] VANGAL S, HOWARD J, RUHL G, et al. An 80-tile 1.28TFLOPS network-on-chip in 65nm CMOS [C]//2007 IEEE International Solid-State Circuits Conference. San Francisco, CA, USA: IEEE, 2007: 98-589.

[3] GRATZ P, KIM C, SANKARALINGAM K, et al. On-chip interconnection networks of the TRIPS chip[J]. IEEE Micro, 2007, 27(5): 41-50.

[4] WENTZLAFF D, GRIFFIN P, HOFFMANN H, et al. On-chip interconnection architecture of the tile processor[J]. IEEE Micro, 2007, 27(5): 15-31.

[5] JERGER N E, LIPASTI M, PEH L S. Circuit-switched coherence [J]. IEEE Computer Architecture Letters, 2007, 6(1): 5-8.

[6] PEH L S, DALLY W J. A delay model and speculative architecture for pipelined routers [C]//Proceedings HPCA Seventh International Symposium on High-Performance Computer Architecture. Monterrey, Mexico: IEEE, 2002: 255-266.

[7] 胡哲琨,陈杰.消息传递型片上多核系统的设计[J].湖南大学学报(自然科学版),2013,40(8):102-109.
HU Z K, CHEN J. Design of a message-passing multi-core system [J]. Journal of Hunan University (Natural Sciences), 2013, 40(8): 102-109.(in Chinese)

[8] KUMARY A, KUNDUZ P, SINGHX A P, et al. A 4.6Tbits/s 3.6GHz single-cycle NoC router with a novel switch allocator in 65nm CMOS [C]//2007 25th International Conference on Computer Design. Lake Tahoe, CA, USA: IEEE, 2008: 63-70.

[9] TIWARI V, KHARE K, SHANDILYA S. An efficient 4×4 mesh structure with a combination of two NoC router architecture [J]. International Journal of Sensors, Wireless Communication and Control, 2021, 11(2): 169-180.

[10] CHIOU S Y. Bufferless routing algorithms: a survey [J]. Advances in Computational Sciences and Technology, 2018, 11(5): 381-386.

[11] 李存禄,董德尊,吴际,等.低延迟路由器中高效开关分配机制的实现与评测[J].湖南大学学报(自然科学版),2015,42(4):78-84.
LI C L, DONG D Z, WU J, et al. Design and implementation of efficient switching in low-latency router [J]. Journal of Hunan University (Natural Sciences), 2015, 42(4): 78-84.(in Chinese)

[12] MULLINS R, WEST A, MOORE S. The design and implementation of a low-latency on-chip network [C]//Proceedings of the 2006 Asia and South Pacific Design Automation Conference. New York: ACM, 2006: 164-169.

[13] HU J C, 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.

[14] MULLINS R, WEST A, MOORE S. The design and implementation of a low-latency on-chip network [C]//Proceedings of the 2006 Asia and South Pacific Design Automation Conference. New York: ACM, 2006: 164-169.

[15] TAMIR Y, FRAZIER G L. High-performance multiqueue buffers for VLSI communication switches [C]//[1988] The 15th Annual International Symposium on Computer Architecture. Honolulu, HI, USA: IEEE, 2002: 343-354.

[16] KODI A, SARATHY A, LOURI A. Design of adaptive communication channel buffers for low-power area-efficient network-on-chip architecture [C]//Proceedings of the 3rd ACM/IEEE Symposium on Architecture for Networking and Communications Systems. New York: ACM, 2007: 47-56.

[17] KAROL M, HLUCHYJ M, MORGAN S. Input versus output queueing on a space-division packet switch[J]. IEEE Transactions on Communications, 1987, 35(12): 1347-1356.