计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2021.20201041
58(6): 1155-1165, 2021

高性能众核处理器申威 26010

胡向东 柯希明 尹 飞 张 新 马永飞 颜世云 马 超
(上海高性能集成电路设计中心 上海 201204)
(huxdisme@vip.sina.com)

Shenwei-26010: A High-Performance Many-Core Processor

Hu Xiangdong, Ke Ximing, Yin Fei, Zhao Xin, Ma Yongfei, Yan Shiyun, and Ma Chao
(*Shanghai High-Performance Integrated Circuit Design Center*, *Shanghai* 201204)

Abstract Based on the multi-core processor Shenwei 1600, the high-performance many-core processor Shenwei 26010 adopts SoC (system on chip) technology, and integrates 4 computing-control cores and 256 computing cores in a single chip. It adopts a 64-bit RISC (reduced instruction set computer) instruction set designed with an original design, and supports 256-bit SIMD (single instruction multiple data) integer and floating-point vector-acceleration operations. Its peak performance for double precision floating-point operations reaches 3.168TFLOPS. Shenwei 26010 processor is manufactured using 28 nm process technology. The die area of the chip is more than 500 mm^2, and the 260 cores of the chip can run stably with a frequency of 1.5 GHz. Shenwei 26010 processor adopts a variety of low power-consumption designs on the architecture level, the microarchitecture level, and the circuit level, and thus, leading to a peak energy-efficiency-ratio of 10.559GFLOPS/W. Notably, both the operating frequency and the energy-efficiency-ratio of the chip are higher than those of the worldwide contemporary processor products. Through the technical innovations of high frequency design, stable reliability design and yield design, Shenwei 26010 has effectively solved the issues of high frequency target, power consumption wall, stability and reliability, and yield, all of which are encountered when pursuing the goal of high-performance computing. It has been applied successfully to a 100PFLOPS supercomputer system named "Sunway TaihuLight" on a large scale, and therefore, can adequately meet the computing requirements for both scientific and engineering applications.

Key words Shenwei instruction set; computation-control core; computing core; low power design; energy-efficiency-ration

摘 要 申威 26010 高性能众核处理器在多核处理器申威 1600 基础上,采用片上系统(system on chip, SoC)技术,在单芯片内集成 4 个运算控制核心和 256 个运算核心,采用自主设计的 64 位申威 RISC (reduced instruction set computer)指令系统,支持 256 位 SIMD(single instruction multiple data)整数和浮点向量加速运算,单芯片双精度浮点峰值性能达 3.168TFLOPS.申威 26010 处理器基于 28 nm 工艺流片,芯片 die 面积超过 500 mm^2,芯片 260 个核心稳定运行频率达 1.5 GHz.申威 26010 处理器从结构级、微结构级到电路级,综合采用多种低功耗设计技术,峰值能效比达 10.559GFLOPS/W.芯片运行频率和能效比均超过同时期国际同类型处理器.申威 26010 通过在高频率设计、稳定可靠性设计和成品率设计等方面的技术创新,有效解决了芯片在实现高性能目标中所遇到的高频率目标、功耗墙、稳定可靠性和成品率等难题,成功大规模应用于国产 10 万万亿次超级计算机系统"神威·太湖之光",有效满足了科学与工程应用的计算需求.

收稿日期:2020-12-21;**修回日期**:2021-04-26
基金项目:"核高基"国家科技重大专项基金项目(2013ZX01028-001-001)
This work was supported by the National Science and Technology Major Projects of Hegaoji (2013ZX01028-001-001).

关键词 申威指令集;运算控制核心;运算核心;低功耗设计;能效比

中图法分类号 TP338

为了满足国产超级计算机研制对国产高性能CPU(central processing unit)的迫切需求,“十一五”期间,在“核高基”国家科技重大专项的支持下,申威处理器研发团队完成了高性能多核CPU芯片申威1600的研发[1],申威1600被成功应用于第一台全部基于国产CPU芯片构建的国产千万亿次超级计算机系统“神威·蓝光”.“十二五”期间,申威研发团队继续在国家“核高基”重大专项的支持下,成功完成了高性能众核处理器申威26010的研发.为了在性能和稳定可靠性等方面满足构建国产新一代超级计算机系统的需求,芯片研发团队在高性能多核处理器申威1600研发成果和技术基础上,突破了芯片结构设计、低功耗设计、稳定可靠性和成品率设计等多个方面的关键技术,最终于2014年完成芯片研制,并大规模应用于国产10万万亿次计算机系统——“神威·太湖之光”,该系统从2016年6月开始连续4次蝉联全球超级计算机排行榜Top500冠军,基于该系统的应用课题2次斩获超级计算应用最高奖——“戈登·贝尔奖”.

申威26010芯片采用片上系统(system on chip, SoC)技术,片上集成了4个运算控制核心和256个运算核心,以及4路128位DDR3存储访问接口和8通路PCI-E3.0等I/O接口.该芯片采用28 nm工艺流片,晶体管数量达到50亿,die面积超过500 mm^2,已接近芯片代工生产极限.处理器核心工作频率达到1.5 GHz,双精度浮点峰值性能达3.168TFLOPS,峰值功耗近300W.要实现芯片的性能、功耗和稳定可靠性等多个方面技术指标,芯片研发在结构和微结构、正确性、低功耗、稳定可靠性和成品率等方面遇到了巨大的挑战,本文主要阐述应对这些挑战的设计方法.

1 结构与组成

1.1 总体结构

申威26010处理器采用分布共享SoC芯片架构[2-3],全芯片共集成了4个运算控制核心和256个运算核心,以及4路128 b的DDR3存储器控制接口和8通路PCI-E3.0等I/O接口,总体结构如图1所示.申威26010片上包含4个核组、1个系统接口

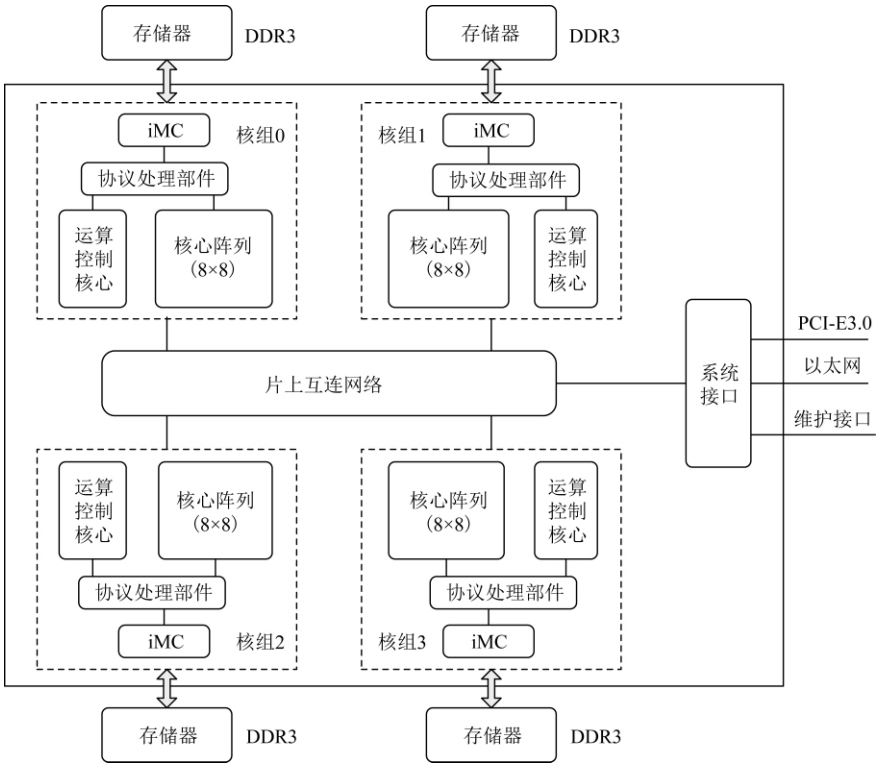

Fig. 1 The overall structure of Shenwei 26010

图1 申威26010总体结构图

和 1 套片上网络.每个核组包含 1 个运算控制核心、1 个 8×8 的运算核心阵列和 1 个协议处理部件及智能存储器访问控制接口 iMC;系统接口连接 PCI-E3.0 和以太网等 I/O 接口;片上网络实现 4 个核组和系统接口之间的互连.

自主指令集是国产处理器冲破国外同行业的技术封锁和知识产权壁垒的基础,申威 26010 处理器的 2 类核心采用申威自主 64 b 的 RISC 指令集,运算控制核心和运算核心的基础指令集保持兼容,支持 8 b,16 b,32 b 和 64 b 整数运算、单精度和双精度浮点运算,并根据高性能应用需求进行了扩展:2 类核心均支持 256 b 的 SIMD 扩展指令,支持整数和浮点的短向量操作,使得运算控制核心每个时钟周期最快可以完成 16 个双精度浮点运算,运算核心每个时钟周期最快可以完成 8 个双精度浮点运算.

1.2 运算控制核心和运算核心

芯片集成的运算控制核心负责芯片资源管理,提供各种系统服务功能,并承担系统中无法并行化的应用程序段的执行,因此对该核心的管理功能和计算性能要求均很高.申威 26010 的运算控制核心由指令流水线、运算流水线、访存流水线和 2 级 Cache 等部分组成.采用 4 译码 7 发射指令流水线结构,支持同时发射 5 条整数类指令(含访存指令)和 2 条浮点类指令,支持指令预取、转移预测、寄存器更名、乱序发射、乱序执行和推测执行.运算流水线包含 5 条整数流水线、2 条支持 256 b 的 SIMD 指令的浮点流水线以及对应的寄存器文件.访存流水线处理访存指令,实现对存储器空间和 I/O 空间的访问,控制数据 Cache 的访问.每个核心集成了容量均为 32 KB 的一级指令 Cache 和一级数据 Cache,以及指令和数据共享的 512 KB 二级 Cache.运算控制核心的总体结构如图 2 所示.

芯片集成的运算核心主要承担计算任务,由指令流水线、运算流水线、访存流水线、16 KB 一级指令 Cache 和 64 KB 可重构局部数据存储器等部分组成.运算核心指令流水线采用 2 译码 2 发射结构,支持乱序发射、乱序执行和乱序退出.运算流水线包含 2 条运算流水线,其中 1 条运算流水线支持 256 b 的 SIMD 指令,支持整数和浮点的短向量加速计算,另一条为整数运算流水线,支持 32 b 和 64 b 整数算术运算、逻辑运算、移位运算以及访存地址的计算等,2 条运算流水线共享 1 个寄存器文件.访存流水线处理访存指令,实现对存储器空间的访问,并控制可重构局部数据存储器的访问.根据应用需要,可将核心局部数据存储器重构成软硬件协同 Cache 结构.

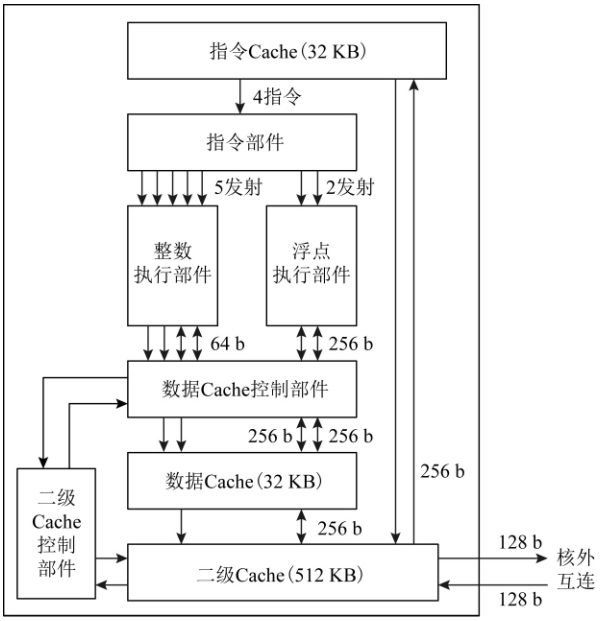

Fig. 2 The structure of the computation-controlling core

图 2 运算控制核心结构图

芯片的 2 类核心通过支持 256 b 的单指令流多数据流 SIMD 指令,支持整数和浮点的短向量操作,实现单条指令同时对多个不同数据完成相同操作,实现核心内的数据级并行;2 类核心实现的超标量结构支持核心内的指令级并行处理;核组内的不同核心之间和核组之间支持线程级或进程级等更高层次的并行处理.基于芯片支持的多粒度多层次并行处理功能,使得在 1.5 GHz 工作频率下单个运算控制核心的双精度浮点峰值性能达到 24GFLOPS,单个运算核心的双精度浮点峰值性能达到 12GFLOPS,芯片集成的 260 个核心提供的双精度浮点峰值性能可达 3.168TFLOPS.

1.3 片上存储结构

芯片集成的运算核心采用了局部数据存储器技术,每个运算核心的局部数据存储器可由软件完成数据的缓存管理,不同管理方式可同时存在并支持局部数据存储器容量的动态划分,充分结合硬件的高效性和软件的灵活性,降低芯片实现开销并满足应用对存储的需要.

运算核心的指令存储器采用 Cache 结构,硬件支持对一级指令 Cache 的指令脱靶进行合并,提高了存储总线带宽的利用率.运算核心阵列集成了更大容量的共享二级指令 Cache,进一步提高了具有局部性的指令访问命中率,降低指令脱靶访问延迟,并且减少指令脱靶对主存储器的频繁访问.

为支持片上存储的高效使用和数据在运算核心

中的灵活分配,运算核心在能够直接访问主存空间的同时,采用了多模式数据流传输技术,支持数据在核心局部数据存储器和主存间的批量带跨步的异步数据传输,实现计算与访存的并行.每个存储访问接口还实现了智能访存优化算法,优化算法可以依据不同课题的访存特征对访存请求进行访问优化,以有效提高存储带宽的使用效率.

申威 26010 核组的运算核心阵列还实现了基于预约调度的传输总线技术,多个运算核心的访存行为由集中控制器进行统一管理,多核心复用的总线资源按照效率优先兼顾公平的算法进行节拍级调度和分配,充分保证运算核心的服务质量,提升了访存效率.

总之,申威 26010 处理器的片上存储结构有效利用了片上资源,缓解了访存墙问题.

2 正确性验证

高性能处理器的正确性验证至今仍是一个业界难题,而申威 26010 处理器设计规模庞大、结构复杂,内部包含 4 个运算控制核心、256 个运算核心、4 路高带宽 DDR3 存储控制接口等众多功能模块,组成了一个逻辑极其复杂的片上系统.申威 26010 还包含核心、核组和芯片等多个设计层次,较多的设计层次使得片内运行控制更加复杂,逻辑信号传递路径越深,传递过程中的各种组合情况越复杂,设计错误隐藏也越深,验证难度越大.这个复杂的片上系统对正确性验证提出了严峻的挑战,如果仅仅采用传统处理器验证方法,难以在有限的研发周期内完成芯片的验证工作,为此,芯片验证团队在借鉴以往验证经验的基础上,主要采用了 3 种技术方法:

1) 综合采用多种验证手段.申威 26010 芯片综合采用了模拟验证[4-5]、硬件仿真加速器验证[6-7]、FPGA 实物验证[8] 和形式验证[9] 等多种验证方法.模拟验证作为一种传统的验证方法,可观性好,错误定位快,但其验证速度随着验证对象规模的增大而降低,由于申威 26010 在设计的模块级和部件级规模相对较小,主要采用该方法来进行验证,取得了较好验证效果;硬件仿真加速器验证的验证速度可以比模拟验证快很多,而且可观性好,验证过程中的信号状态可以全程跟踪,错误定位便捷,用于验证的中后期,芯片有了基本正确性以后,在核心以上层次支撑操作系统及应用程序等较大规模测试程序的验证,申威 26010 的硬件仿真加速器验证环境上几乎发现了全部软硬件接口相关的设计错误,取得了很好的验证效果;FPGA 实物验证的验证速度比硬件仿真加速验证更快,主要用于在核心以上层次支撑大量应用级测试程序的验证,申威 26010 基于自研的单核、单核组、多核组和全片等多种不同规模的 FPGA 验证平台,实现了多个层次在应用级的快速验证,有效加快了芯片的错误收敛速度;形式验证在申威 26010 中主要用于 RTL 设计与后端物理实现之间的等价性验证.

2) 采用层次化的验证策略.针对申威 26010 的层次化结构和芯片规模超大特性,将芯片的正确性验证分为模块级、部件级和芯片级 3 个层次,开发以白盒、黑盒和灰盒测试理论指导下的基于约束的随机激励、基于断言的定向激励以及多元化事务激励、场景激励,分解激励开发和验证难度,满足不同层次验证环境对运行速度和验证资源的需求.模块级运行速度快,资源用量少,侧重白盒焦点验证,在信号层面开发各种激励确保底层模块验证覆盖率.部件级运行速度较快,验证资源用量中等,侧重在协议层面开发激励,既包含白盒焦点验证和灰盒验证,也含有黑盒自动化验证.芯片级运行速度慢,验证资源用量大,侧重于在指令序列等软件可见状态层面构建自动化验证环境进行黑盒方式验证.

3) 构建可重构芯片级验证环境.可重构芯片级验证环境支持多种参数化配置,使得验证人员能够根据不同的验证需求,自由灵活地构建芯片级验证环境,较好地解决了验证覆盖率和模拟仿真速度之间的矛盾,也较好地解决了验证规模与运算资源之间的矛盾,取得了很好的验证效果.申威 26010 的可重构芯片级验证环境如图 3 所示.该环境支持芯片中的核组数量可配置,可以配置芯片的核组数量为 1～4 个,支持单核组中运算核心数量可配置,可以配置的运算核心数量为 1～64 个;支持对各核组内的运算核心阵列中的真、伪运算核心进行替换,其中伪运算核心是一个运算核心接口模型,伪运算核心模型的接口行为与真实核心完全一致,但其设计规模远小于真实运算核心;支持对各核组中的访存接口进行多种配置,包括使用真实的设计模型、虚拟存储器接口模型等;支持对芯片中的 PCI-E 和以太网接口进行配置,可选择芯片 RTL 模型中是否包含这 2 个接口.

申威 26010 通过综合采用多种验证方法,以及多层次、多规模的验证,发挥各种验证方法的优势,从不同验证层次和验证视角实现交叉验证和优势

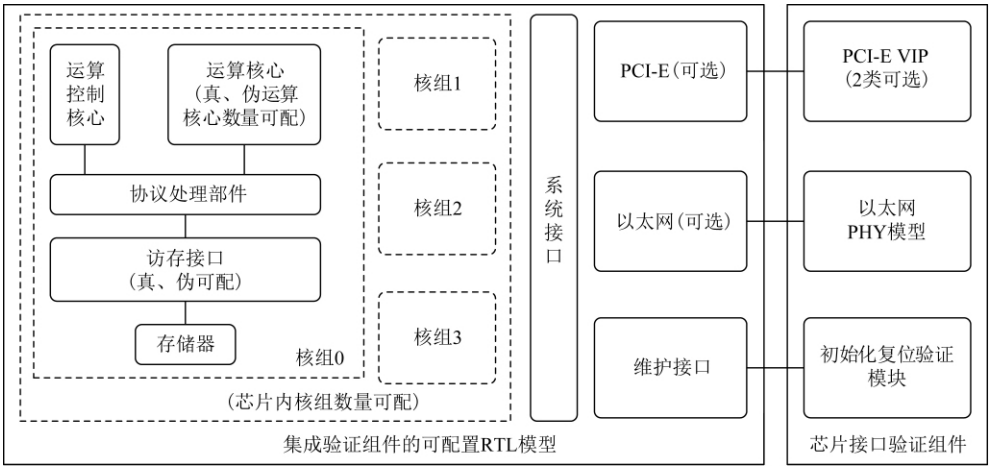

Fig. 3 Reconfigurable chip-level verification environment

图 3 可重构芯片级验证环境

互补,最终取得了很好的验证效果,实现了一次流片成功的目标.

3 物理实现

申威 26010 规模庞大,芯片尺寸已接近生产极限,这对物理实现来说是个极大的挑战,为了完成如此大规模芯片的物理设计且实现高性能的目标,本芯片采用高可复用层次化物理设计、高性能时钟系统设计和定制综合混合设计等方法,基于 28 nm 工艺实现了 1.5 GHz 的频率指标.

3.1 高可复用层次化物理设计

层次化物理设计方法是实现超大规模芯片设计的基础,该方法实现了物理设计并行化,提高了后端设计团队在统一平台上分工协作的效率,同时层次化的设计可以缩小模块的设计规模,减轻设计及检查分析对计算资源需求的压力,缩短设计优化的周期,从而可以通过增加优化迭代的次数,取得更好的设计优化效果.本芯片采用的高可复用物理设计方法支持电路和版图的层次化设计,同时支持静态时序分析、功耗分析、等价性验证和可靠性分析等层次化的检查分析,从而高效地实现了申威 26010 这款极大规模芯片的物理设计.

申威 26010 物理实现上分为核心、核组和芯片 3 个全局层次,采用自顶向下的策略,以全片 Floorplan 设计、全局电地设计以及全局时钟设计为主导,根据芯片总体要求和信号连接关系,依次确定芯片、核组和核心的面积和各层次模块的相对位置关系,制定时钟网络的实现方案,给出各层次顶层的设计资源和设计约束,实现芯片的总体布局和规划.各模块在顶层模块给予的设计约束下进行设计和优化,并将结果依次反馈给上一层次进行调整优化,实现自底向上的反馈回路.层次化的设计中采用了高可复用性的策略,功能模块和缓存模块设计好后进行 IP 化处理,给核心层进行复用,核心层固化后在核组层进行复用,在芯片层对核组进行复用,实现了高效的层次化设计.

3.2 高性能时钟系统设计

全芯片包含了多种不同频率时钟,包括:控制核心时钟、运算核心时钟、存控时钟、PCI-E 时钟、全局时钟、接口及维护时钟等.其中全局时钟频率达到 1.4 GHz,控制核心和运算核心的频率均达到 1.5 GHz.不同时钟在分布范围、时钟偏斜和时钟功耗上有不同的指标要求,需要根据它们的特点分别采用不同的设计方法:

1) 对于运算核心时钟、控制核心时钟和全局时钟 3 种高频率且分布范围广的时钟,采用“全局+局部”2 层的时钟设计结构,分层次进行低偏斜时钟设计;为增强抗 OCV(on-chip variation)的能力,全局时钟采用对称 H-tree 型结构,实现时钟从源头到各终点传播延时的精准控制.在模块局部时钟设计中,直接采用“大驱动+MESH”的方式直连到各时序单元,确保时钟信号传播的低延时和低偏斜.一个运算核组的时钟分布如图 4 所示.

2) 对于分布范围较小或频率较低的其他时钟按照平衡时钟树的方式进行单层时钟结构设计,在满足设计性能的同时也大大降低了设计复杂度.

通过上述设计方法,申威 26010 的各高频时钟

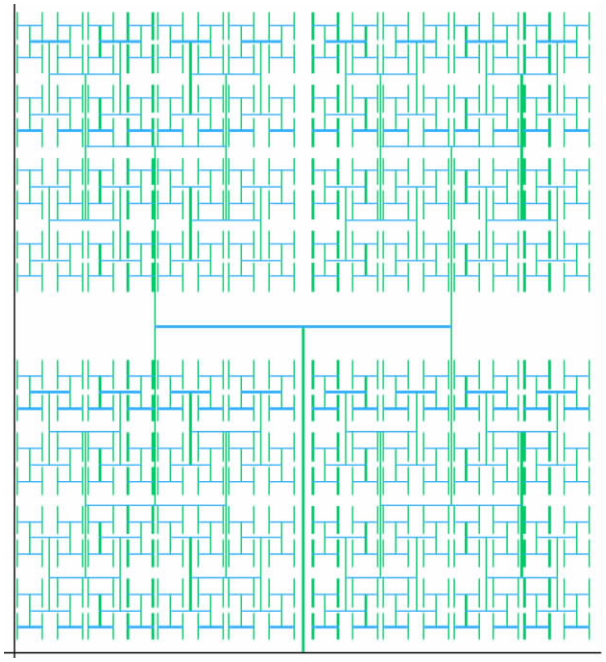

Fig. 4 Clock network distribution of an computing core group

图 4 一个运算核组的时钟网络分布图

全片分布最大偏斜均控制在 10 ps 以内,时钟占空比达到 49.85%～50.15%,经流片测试各时钟均可以稳定运行在设计频率下,达到了设计目标.

3.3 高性能定制设计

申威 26010 芯片 2 类核心的逻辑非常复杂,为了达到频率设计目标,采用了多种定制设计技术:

1) 全局通路设计

在全局芯片布局设计时优先考虑关键时序通路的设计,尽可能缩短其物理长度.此外在全局布线的金属资源选取上,也将传播速度较快的高层金属尽量向关键通路倾斜,确保关键通路的时序可以满足设计要求.

2) 定制存储器设计

访存路径一直是处理器的关键路径所在,需要进一步提升片上 SRAM 阵列的访存速度.商用的存储器综合工具(Memory Compiler)已无法满足存储器的频率要求,申威 26010 处理器内部主要 Cache 阵列均为定制实现,包含单端口和双端口阵列.定制存储器采用了容偏差灵敏放大器设计、高速译码器设计和自定时电路等关键技术[10],速度比基于商用工具生成的存储器快 27%～37%.

此外寄存器文件也是关键路径所在,由于读写端口众多,综合实现方法无法有效地布通走线,且时序难以达到指标,申威 26010 中的 5 读 5 写和 7 读 4 写寄存器文件均为定制设计,采用了自研多端口 bitcell(存储单元)、高速译码电路和多米诺读出电路等关键技术,最终满足了寄存器文件的频率设计要求.

3) 高性能时钟树定制设计

为尽可能降低时钟偏斜、降低时钟延时和增强其抗 OCV 的能力,全芯片 3 个主要高频时钟均采用定制设计方式实现,时钟主干采用定制 H-tree 时钟树结构,时钟的一级驱动单元、二级驱动单元及门控驱动单元均采用定制实现,确保整个时钟树设计具备低传播延时和低传播偏斜的特性.

通过这 3 种技术手段,芯片最终可以稳定运行在 1.5 GHz,工作频率高于国际上同期同类芯片,使芯片性能达到了设计预期,双精度浮点峰值性能达到了 3.168TFLOPS 的设计指标.

3.4 高性能综合设计

为提高设计效率,芯片的大部分控制与运算逻辑模块均采用了综合设计方法来实现,在传统商用综合设计流程的基础上,芯片开发团队根据芯片的特点自行定制开发了多项自动化功能,例如:自动填充物理信息的逻辑综合功能、关键逻辑自动打包聚集功能、根据时序自动调整并优化关键路径权重功能、自动创建定制 Mesh 时钟树功能、对关键路径或指定路径优先进行布线功能、自动在大反转电流单元两侧插入去耦电容功能[11]、集成时序分析及时序自动优化功能、集成设计规则检查及自动修复功能等.通过对综合流程的深度定制化开发,大大提高了综合设计质量和效率,模块级设计频率较标准商用流程提升 15%～20%,布线错误率下降 90%,极大地提高了设计的效率和质量.

4 低功耗设计

随着晶体管数量的增加和工作频率的提高,降低处理器的功耗变得越来越重要[12-13].申威 26010 在实现高性能的同时,从结构级、微结构级到电路级,综合采用多层次功耗优化技术来降低处理器的功耗.

1) 结构级低功耗设计

申威 26010 在结构级采用的低功耗设计技术有:

①申威 26010 的结构设计思想是通过集成众多核心来提升性能,适当降低单核心最高工作频率的要求,避免过高工作频率带来功耗的快速上升,从而有效地提升了芯片的能效比.

② 支持多种形态的工作模式.包括深度睡眠、浅睡眠和低功耗运行模式.对较长时间无工作负载的核心,可控制使其处于极低工作频率的深度睡眠状态,最大限度降低运行功耗;对短时间无工作负载的核心,特殊的停机指令可使核心处于浅睡眠状态,杜绝核心绝大多数信号的翻转从而降低功耗;对运行速度要求较低的应用程序,可以动态调整指令发射速度,达到降低运行功耗的目的.

③ 多频率设计.在满足性能需求前提下,仅核心采用最高工作频率,互连部件、存储控制器和系统接口则采取较低的工作频率,降低运行功耗.

2) 微结构级低功耗设计

申威 26010 在微结构级采用的低功耗设计技术有:

① 功能部件动态配置.采用动态切割方式,支持不同层次的部件切割,以降低功耗.一是核心级,可以根据应用需求的核心数量,将不使用的核心断开,使其处于极低工作频率状态;二是部件级,对浮点部件或 SIMD 运算部件,在运行无浮点操作或无 SIMD 运算的应用时,可动态关闭浮点部件或 SIMD 部件的时钟,降低核心的运行功耗.

② 多端口存储器设计.Cache 存储器设计采用"虚拟多端口"技术来减少物理端口数量,既降低功耗,也有效降低芯片面积.其中运算控制核心的指令 Cache 和二级 Cache 都采用物理单端口存储器,虚拟实现双端口功能,数据 Cache 则采用双端口存储器实现了虚拟三端口的功能.

③ I/O 低功耗支持.DDR3 存储器接口和 PCI-E 接口都支持低功耗模式,在没有访问请求时,可自动处于低功耗状态.

3) 电路级低功耗设计

申威 26010 在电路级采用的低功耗设计技术有:

① 采用多层次多粒度的门控时钟方式.降低平均运行功耗,细粒度控制可在模块内部实现对一定数量的触发器进行控制,粗粒度控制可在模块级、核心级和核组级进行时钟控制,从而实现不同工作模式下降低功耗的目标.同时采取动态功耗分析和电压降分析,通过布局优化和放置片上电容,避免门控时钟在降低功耗的同时造成动态电压降影响电路工作的稳定性.

② 采用多阈值晶体管混合设计.以常规阈值晶体管为主体进行设计,用速度最快的低阈值晶体管进行关键时序路径的优化,这样在满足设计频率目标前提下,尽可能采用高阈值晶体管来优化漏电功耗.通过此设计策略,在申威 26010 的 50 亿晶体管中,低阈值晶体管数量仅占 1.97%,使得常温下漏电功耗仅为 12 W.

5 可靠性设计

申威 26010 在使用中根据运行课题的不同,芯片的实际功耗往往会在几十瓦到几百瓦之间来回波动.频繁的大幅度功耗波动给芯片的稳定可靠性带来了严峻的挑战.为了确保芯片可以在实际系统中稳定运行,申威 26010 从结构设计到物理设计综合采用了多种高可靠性设计方法,有效地降低了功耗波动对电源网络系统的影响,确保了芯片在实际系统中的稳定工作.芯片稳定可靠性设计所采用的关键技术方法有:

1) 电地网络强化设计.在各运算核心和运算控制核心上均采用 BUMP 垂直供电技术,确保各部分的充足供电;采用自顶向下每层均垂直交叉打孔的网格状方式进行连接;除相互电地隔离的区域外,所有模块电地均在芯片顶层连在一起,构成一张统一完整的大网,确保电源网络的强壮性.

2) 电地网络隔离设计.同时对于不同核心区域的电地进行物理隔离,避免功耗波动导致的电压波动相互影响.

3) 去耦电容的按需使用.通过设计流程优化确保各大功耗单元周围插入去耦电容单元[14],减少电源波动.

4) 片上时钟变化平滑过渡的控制方法.在芯片整体或局部部件进行时钟频率提升或降低时,按照预设的部件粒度和时间间隔进行频率的变化,使得芯片内部时钟频率变化时功耗按梯度变化,有效降低功耗波动给芯片可靠运行带来的风险.

5) 片上存储器采取容工艺偏差自调节设计方法.在芯片运行过程中实时感知工艺参数的变化,并根据工艺参数的变化情况自动调整存储器电路的相关参数,以有效容忍制造工艺偏差,提高电路运行的稳定可靠性.

6 成品率设计

越大的芯片面积会导致更大的工艺偏差和更高的制造缺陷概率,从而会导致部分芯片出现性能或功能上的问题,降低芯片成品率.申威 26010 在设计

时采用了多种提升成品率的技术方法,主要采取的技术方法有:

1) 容偏差存储器设计

由于 SRAM 晶体管占了芯片总晶体管数的 40%,而且存储单元采用最小尺寸设计,所以 SRAM 阵列是对芯片成品率影响最大的部分.片内主要 SRAM 存储器均采用定制设计实现,在设计时采用了容偏差存储器设计来确保在大的工艺偏差下仍然能正常工作,同时采用了多种修复策略来消除制造缺陷对成品率的影响.

容偏差存储器主要采用容偏差灵敏放大器设计和自定时电路 2 种关键技术.在灵敏放大器电路中,包含有互补反相器对和对称的放电通路,需要很好的匹配才能保证逻辑的可靠性,目前常用的灵敏放大器分为电压型和电流型 2 种,通过蒙特卡罗仿真对比[15],电压型灵敏放大器具有更好的速度和稳定性,所以设计采用了电压型结构.

在版图设计时为了保证严格的对称性,首先采用半边设计方式,然后 X 轴镜像调用以保证对称性.同时对敏感器件采用中心对称的设计方式和金属线屏蔽.放大管采用大于 2 倍最小管长的方式减小失配(mismatch),并采用非最小规则进行版图设计.蒙特卡洛仿真表明电压差为 10 mV 时良率为 98.17%,电压差为 20 mV 时灵敏放大器良率已达到 100%,具有很高的容偏差能力.

工艺偏差将导致灵敏放大器差分输入端的电压差发生变化,直接影响灵敏放大器的可靠性,继而影响整个存储器的可靠性.为了改善灵敏放大器开启时间的控制,本设计采用了自定时灵敏放大器设计,能够自适应地调节开启时间,确保 SRAM 正常工作.

当有效电压差达到预定值时,灵敏放大器开启较高的正确放大概率,但是由于工艺偏差的影响,每一块 SRAM 的开启时间都有所不同,失配严重的将会引起功能错误.为了解决这个问题,灵敏放大器的信号延时部分采用了 4 级可选延时结构,阵列会根据存储器内建自测试的结果来自动选择开启的时间,首先选择最快开启档位,如果存储器失效,则依次降低开启档位直到存储器正常工作,在确保功能的前提下实现最高的性能.

此外,为了消除制造缺陷对成品率的影响,采用了多维度冗余自修复策略,对 32 KB 的一级 Cache 采用了列冗余修复策略,对容量较大的 512 KB 二级 Cache 采用了行列冗余同时修复的策略.结合 BIST 测试算法,实现了自测试自修复的功能,经测试在标准电压下 SRAM 良率达到 100%,频率达到设计目标.

2) 选取合理的时序余量(margin)

在设计分析时添加适当的时序 margin 是确保芯片在一定工艺偏差下仍可以达到工作频率的重要手段.过大的 margin 会导致严重的功耗问题,过小的 margin 则有可能导致性能不达标甚至出错,因此如何设定合理的时序分析 margin 就成为了关键.

申威 26010 在设计时会先对全局关键时钟信号进行全 Corner 偏差仿真,再根据不同关键路径间的局部相对关系和位置来确定每组路径具体的时序分析标准及 margin 的设定.这样既可以通过添加合理的时序 margin 来对抗可能的工艺偏差,也可以避免对其他无关通路的过约束设计.

3) 可制造性设计优化

为进一步减少因工艺制造偏差所导致的芯片失效,全片在物理实现时均采用了"多孔、宽线"的可制造性优化方法,即在不影响性能和布局布线的前提下,尽量采用更宽的互连线,金属层之间尽量用多通孔来代替单通孔,尽可能地减少因制造失效导致的芯片功能错误.

通过这 3 种技术手段,申威 26010 芯片的量产成品率达到了 50%以上,对于如此大规模的芯片已是非常高的成品率.

7 测试与应用

申威 26010 工作频率达到 1.5 GHz,在此频率下,全芯片双精度浮点峰值速度达到 3.168TFLOPS,整数峰值速度达到 3.522TOPS;实测单处理器 LINPACK 效率为 80.10%,持续性能为 2.538TFLOPS;实测 HPL-DGEMM 效率为 97.558%,性能为 3.091 TFLOPS,芯片峰值运行功耗为 292.7 W,性能功耗比为 10.559GFLOPS/W.申威 26010 芯片于 2014 年底设计定型,各项指标均达到了同期国际领先水平[16],申威 26010 与 2011～2015 年国际上峰值性能超过 1TFLOPS 的 CPU/GPU 之间的对比,如表 1 所示.

国家"863 计划"支持研制的"神威·太湖之光"超级计算机系统,全部采用"申威 26010"处理器.该系统共集成了 40 960 颗申威 26010 处理器,系统峰值运算速度达到 125.43PFLOPS,实测 Linpack 效率达到 74.1%,成为世界上率先突破每秒 10 亿亿次

的超级计算机系统,也是我国全部采用国产处理器的超级计算机首次位居世界第一,系统连续 4 次蝉联超级计算机 TOP500 排行榜冠军,打破了美国对我国超级计算机处理器芯片禁运的封锁.在此系统上完成的"千万核可扩展大气动力学全隐式模拟"和"非线性大地震模拟"2 项应用获得高性能计算机应用领域的最高奖"戈登·贝尔奖".同时,系统应用覆盖海洋、金融、气候、航天、新药、材料等 10 多个应用领域,为国民经济、国防、科研的发展产生了强大的推动作用,取得了显著的社会效益.

Table 1 Comparison Between Shenwei 26010 and CPU/GPU with Performance Exceeding 1TFLOPS

表 1 申威 26010 与国际上性能超过 1TFLOPS 的 CPU/GPU 对比

处理器相关信息及参数	GCN (HD7970-En)	Kepler-GK110 (Tesla K20X)	Xeon Phi (5110P)	Xeon Phi2	申威 26010
制造商	AMD	NVIDIA	Intel	Intel	
处理器类型	GPU	GPU	众核 CPU	众核 CPU	异构众核 CPU
工艺	28 nm	28 nm	22 nm 专用工艺	14 nm 专用工艺	28 nm
晶体管数/亿	43.1	71	50	72	50
芯片面积/mm^2	365	550	350	700	超过 500
频率/GHz	1	0.732	1.05	1.3	1.5
计算单元或核心数量	32 计算单元	2688 CUDA 核心	60 核心	72 核心	260 核心
峰值功耗/W	220	235	225		292.7
双精度峰值/TFLOPS	1.01	1.312	1.01	3	3.168
完成日期	2011.12	2012.05	2012.11	2015.11	2014.12
$\frac{峰值性能}{功耗}$/(GFLOPS·W^{-1})	4.59	5.58	4.49		10.559
芯片计算密度/(GFLOPS·mm^{-2})	2.77	2.39	2.89	4.29	5.49

8 结束语

申威 26010 众核处理器是我国第一款 64 b 高性能通用众核处理器,实现国产处理器"从多核到众核""从每秒千亿次到万亿次"的跨越.申威 26010 采用多粒度并行处理的 SoC 芯片架构,综合采用多种正确性验证方法,采用多种低功耗设计与管理技术,物理设计采用定制与逻辑综合相结合和容工艺偏差设计等多种设计方法,尤其在芯片的频率设计、稳定可靠性设计和成品率设计方面采取了一系列的技术方法,最终同时实现了芯片的高性能、高能效比和高稳定可靠性目标,并在高性能计算领域实现应用的突破.由此可见,虽然国产高性能处理器起步较晚,生产工艺较低,但通过架构和物理设计等方面的创新,完全有能力在特定应用领域实现有效应用.但与国际主流高性能众核处理器相比,申威 26010 众核处理器在配套的软件应用生态链建设方面还需要进一步完善,以进一步提升芯片的适应性.

继申威 26010 之后,申威众核处理器一直在探索和创新中发展,随着工艺和设计能力的提升,针对性能和效率提升这些问题采取了诸多创新性手段.在提高运算性能方面,进一步提高芯片工作频率,将运算核组数量从 4 个增加到 5~8 个,扩展运算核心内的 SIMD 宽度到 512 b,并针对应用需求完善 SIMD 指令集;在高能效设计方面,创新低功耗设计方法与功耗管理措施,将申威众核处理器的峰值功耗和运行功耗控制在合理的范围之内.最后在访存与通信带宽优化方面,使用 DDR4 或者 HBM 和 PCI-E4.0 等新型存储器及高速互连接口,使众核处理器的计算、访存和通信性能达到更优的平衡,提升实际应用效率.

参 考 文 献

[1] Hu Xiangdong, Yang Jianxin, Zhu Ying. Shenwei-1600: A high-performance multi-core microprocessor [J]. Scientia Sinica Informationis, 2015, 45(4): 513-522 (in Chinese)
(胡向东, 杨剑新, 朱英. 高性能多核处理器申威 1600 [J]. 中国科学: 信息科学, 2015, 45(4): 513-522)

[2] Hart J, Butler S, Cho H, et al. 3.6 GHz 16-core SPARC SoC processor in 28 nm [C] //Proc of IEEE Solid-State Circuits Conf Digest of Technical Papers. Piscataway, NJ: IEEE, 2013: 48-50

[3] Konstadinidis G K, Li H P, Schumacher F, et al. SPARC M7: A 20 nm 32-Core 64MB L3 cache processor [J]. IEEE Journal of Solid-State Circuits, 2015, 51(1): 79-91

[4] Bryant R E. A methodology for hardware verification based on logic simulation [J]. Journal of the ACM, 1991, 38 (2): 299-328

[5] Taylor S, Quinn M, Brown D, et al. Functional verification of a multiple-issue, out-of-order, superscalar Alpha processor-the DEC Alpha 21264 microprocessor [C] //Proc of the 35th Design Automation Conf (DAC'98). Piscataway, NJ: IEEE, 1998: 638-643

[6] Zhang Hang, Shen Haihua. Function verification of Godson2 processor [J]. Journal of Computer Research and Development, 2006, 43(6): 974-979 (in Chinese)
(张珩, 沈海华. 龙芯 2 号微处理器的功能验证[J]. 计算机研究与发展, 2006, 43(6): 974-979)

[7] Schubert K D, Roesner W, Ludden J M, et al. Functional verification of the IBM POWER7 microprocessor and POWER7 multiprocessor systems [J]. IBM Journal of Research & Development, 2011, 55(3): 10-17

[8] Zhu Ying, Chen Cheng, Xu Xiaohong, et al. Creation of FPGA verification platform for a high performance multiple-core microprocessor [J]. Journal of Computer Research and Development, 2014, 51(6): 1295-1303 (in Chinese)
(朱英, 陈诚, 许晓红, 等. 一款多核处理器 FPGA 验证平台的设计与实现[J]. 计算机研究与发展, 2014, 51(6): 1295-1303)

[9] Ludden J M, Roesner W, Heiling G M, et al. Functional verification of the POWER4 microprocessor and POWER4 multiprocessor systems [J]. IBM Journal of Reseatch & Development, 2002, 46(1): 53-76

[10] Hashimoto T, Kawabe Y, Hara M, et al. An adaptive clocking control circuit with 7.5% frequency gain for SPARC processors [J]. IEEE Journal of Solid-State Circuits, 2017, 53(4): 1028-1037

[11] Shimazaki K, Okumura T. A minimum decap allocation technique based on simultaneous switching for nanoscale SoC [C] //Proc of IEEE Custom Integrated Circuits Conf. Piscataway, NJ: IEEE, 2009: 21-24

[12] Venkat K, Jeffrey B, Georgios K, et al. Fine-grained adaptive power management of the SPARC M7 processor [C] //Proc of IEEE Solid-State Circuits Conf Digest of Technical Papers. Piscataway, NJ: IEEE, 2015: 74-76

[13] Burgess B, Cohen B, Denman M, et al. BOBCAT: AMD's low-power x86 processor [J]. IEEE Micro, 2011, 31(2): 16-25

[14] Popovich M, Friedman E G, Secareanu R M, et al. Efficient placement of distributed on-chip decoupling capacitors in nanoscale ICs [C] //Proc of IEEE Int Conf on Computer-Aided Design. Piscataway, NJ: IEEE, 2007: 811-816

[15] Shareef B, Doncker E D, Kapenga J. Monte Carlo simulations on Intel Xeon Phi: Offload and native mode [C] //Proc of 2015 IEEE High Performance Extreme Computing Conf (HPEC). Piscataway, NJ: IEEE, 2015: 2-6

[16] Krithika B, Keerthana N. Comparison of Intel processor with AMD processor with green computing [C] //Proc of Int Conf on Green Computing, Communication and Conservation of Energy. Piscataway, NJ: IEEE, 2013: 737-742

Hu Xiangdong, born in 1964. Master, senior engineer. Senior member of CCF. His main research interests include high performance computer architecture, VLSI design & verification.

胡向东, 1964 年生. 硕士, 高级工程师, CCF 高级会员. 主要研究方向为高性能计算机体系结构、超大规模集成电路设计与验证.

Ke Ximing, born in 1967. Master, senior engineer. His main research interests include high performance computer architecture and integrated circuit hardware implementation.

柯希明, 1967 年生. 硕士, 高级工程师. 主要研究方向为高性能计算机体系结构和集成电路硬件实现.

Yin Fei, born in 1978. Master, senior engineer. Her main research interests include high performance computer architecture, multi-core processor architecture and instruction set architecture.

尹 飞, 1978 年生. 硕士, 高级工程师. 主要研究方向为高性能计算机体系结构、多核处理器架构和指令集架构.

Zhang Xin, born in 1982. Master, engineer. His main research interests include design flow development, logical synthesis and physical implementation.

张 新, 1982 年生. 硕士, 工程师. 主要研究方向为设计流程开发、逻辑综合以及物理实现.

Ma Yongfei, born in 1981. Master, engineer. His main research interests include VLSI design & verification.

马永飞, 1981 生. 硕士, 工程师. 主要研究方向为超大规模集成电路设计与验证.

Yan Shiyun, born in 1981. Master, senior engineer. His main research interests include high performance computer architecture, VLSI design & verification.

颜世云,1981 年生.硕士,高级工程师.主要研究方向为高性能计算机体系结构、超大规模集成电路设计与验证.

Ma Chao, born in 1988. PhD, engineer. His main research interests include high-performance interconnect network, VLSI design and low power design.

马　超,1988 年生.博士,工程师.主要研究方向为高性能互连网络、超大规模集成电路设计和低功耗设计.

《计算机研究与发展》2019 年论文高被引 TOP10

排名	论文信息
1	施巍松，张星洲，王一帆，张庆阳. **边缘计算:现状与展望**[J]. 计算机研究与发展，2019，56(1)：69-89 Shi Weisong, Zhang Xingzhou, Wang Yifan, Zhang Qingyang. Edge Computing: State-of-the-Art and Future Directions [J]. Journal of Computer Research and Development, 2019, 56(1): 69-89
2	黄继鹏，史颖欢，高阳. **面向小目标的多尺度 Faster-RCNN 检测算法**[J]. 计算机研究与发展，2019，56(2)：319-327 Huang Jipeng, Shi Yinghuan, Gao Yang. Multi-Scale Faster-RCNN Algorithm for Small Object Detection [J]. Journal of Computer Research and Development, 2019, 56(2): 319-327
3	彭宇新，綦金玮，黄鑫. **多媒体内容理解的研究现状与展望**[J]. 计算机研究与发展，2019，56(1)：183-208 Peng Yuxin, Qi Jinwei, Huang Xin. Current Research Status and Prospects on Multimedia Content Understanding [J]. Journal of Computer Research and Development, 2019, 56(1): 183-208
4	郑庆华，董博，钱步月，田锋，魏笔凡，张未展，刘均. **智慧教育研究现状与发展趋势**[J]. 计算机研究与发展，2019，56(1)：209-224 Zheng Qinghua, Dong Bo, Qian Buyue, Tian Feng, Wei Bifan, Zhang Weizhan, Liu Jun. The State of the Art and Future Tendency of Smart Education [J]. Journal of Computer Research and Development, 2019, 56(1): 209-224
5	夏清，李帅，郝爱民，赵沁平. **基于深度学习的数字几何处理与分析技术研究进展**[J]. 计算机研究与发展，2019，56(1)：155-182 Xia Qing, Li Shuai, Hao Aimin, Zhao Qinping. Deep Learning for Digital Geometry Processing and Analysis: A Review [J]. Journal of Computer Research and Development, 2019, 56(1): 155-182
6	纪守领，李进锋，杜天宇，李博. **机器学习模型可解释性方法、应用与安全研究综述**[J]. 计算机研究与发展，2019，56(10)：2071-2096 Ji Shouling, Li Jinfeng, Du Tianyu, Li Bo. Survey on Techniques, Applications and Security of Machine Learning Interpretability [J]. Journal of Computer Research and Development, 2019, 56(10): 2071-2096
7	任家东，刘新倩，王倩，何海涛，赵小林. **基于 KNN 离群点检测和随机森林的多层入侵检测方法**[J]. 计算机研究与发展，2019，56(3)：566-575 Ren Jiadong, Liu Xinqian, Wang Qian, He Haitao, Zhao Xiaolin. An Multi-Level Intrusion Detection Method Based on KNN Outlier Detection and Random Forests [J]. Journal of Computer Research and Development, 2019, 56(3): 566-575
8	赵志远，王建华，徐开勇，郭松辉. **面向云存储的支持完全外包属性基加密方案**[J]. 计算机研究与发展，2019，56(2)：442-452 Zhao Zhiyuan, Wang Jianhua, Xu Kaiyong, Guo Songhui. Fully Outsourced Attribute-Based Encryption with Verifiability for Cloud Storage [J]. Journal of Computer Research and Development, 2019, 56(2): 442-452
9	陈游旻，陆游游，罗圣美，舒继武. **基于 RDMA 的分布式存储系统研究综述**[J]. 计算机研究与发展，2019，56(2)：227-239 Chen Youmin, Lu Youyou, Luo Shengmei, Shu Jiwu. Survey on RDMA-Based Distributed Storage Systems [J]. Journal of Computer Research and Development, 2019, 56(2): 227-239
10	赵洪科，吴李康，李徵，张兮，刘淇，陈恩红. **基于深度神经网络结构的互联网金融市场动态预测**[J]. 计算机研究与发展，2019，56(8)：1621-1631 Zhao Hongke, Wu Likang, Li Zhi, Zhang Xi, Liu Qi, Chen Enhong. Predicting the Dynamics in Internet Finance Based on Deep Neural Network Structure [J]. Journal of Computer Research and Development, 2019, 56(8): 1621-1631

数据来源:CSCD,中国知网;统计日期:2020 年 12 月 15 日