前　言

我们高兴地向读者推出本刊“计算机体系结构前沿技术”专题！本专题收录的6篇文章既包含不同技术领域和方向的综述，也包含了具体技术的发明和介绍.

自旋转移力矩随机存储器(STT-RAM)是一种新型非易失性随机存储器.STT-RAM与传统的静态随机存储器(SRAM)分别在写入速度、集成密度、制造成本、功耗、使用寿命等方面各有优势和劣势，而综合这2种随机存储器件而形成的SRAM和STT-RAM混合缓冲存储器是一种重要的缓存结构.然而在具有不同特性的缓存器件的同一个缓存结构中如何分配它们的使用比例是一个对性能、功耗、使用寿命均具有影响的重要问题.范浩等作者的文章“一种基于强化学习的混合缓存能耗优化与评价”提出一种基于强化学习的技术，解决待缓存的数据在这2种缓存器件中的分配问题.文章提出缓存访问请求的写入强度的概念，并提出利用强化学习的方法根据写入强度及数据复用信息获得不同分配下的能耗信息，从而给出数据在SRAM和STT-RAM之间的分配策略.实验结果证明了作者所提出的技术的有效性.

卷积神经网络是深度学习中的一类重要工具，已经在计算机图形、图像、计算机视觉、语言处理与机器翻译等获得广泛的应用，而Winograd快速卷积算法因能有效降低卷积神经网络中卷积操作的计算复杂度而受到广泛关注；另一方面，多核处理器是一类重要的体系结构，以多核处理器构造的计算平台承载着大量的卷积神经网络相关的应用任务.因此面向多核处理器系统的Winograd快速卷积算法的高效率实现是有重要价值的研究课题.王庆林等作者的“面向飞腾多核处理器的Winograd快速卷积算法优化”介绍了面向我国自行研制的飞腾多核处理器的Winograd快速卷积算法的优化实现技术.文章提出了一种高性能并行Winograd快速卷积算法，作者把Winograd算法分解为不同的计算部分，设计了相应部分的数据操作及与之相应的数据布局，以并行的方式实现Winograd快速卷积计算.文章对所提出的技术进行了实验测试，实验结果表明了文章所提出的技术可以有效地提高算法访存和计算速度.

图计算广泛地应用于社会计算、网络计算、智能信息处理等各种领域中，而广度优先搜索是图计算中的一个重要的算法，也是考察计算系统性能的典型算法之一.优化图的广度优先搜索算法是提高计算性能的重要研究问题.张承龙等作者的“面向高通量计算机的图算法优化技术”探讨高通量计算机中图广度优先搜索算法的优化技术.文章综合介绍了图广度优先搜索中算法优化的各种技术，并针对图广度优先搜索算法执行中的访存效率问题，提出了基于高效的位图访问的冗余访存消除技术，以及通过算法执行过程的优化提高访存局部性的技术，从而减少了缓存访问频次，提高缓存的访问效率.文章对所提出技术在计算性能、功耗方面进行了实验，实验结果表明了文章所提出的技术的有效性.

针对图计算开展定制硬件结构设计也是当前的一个研究前沿技术方向.现场可编程门阵列(FPGA)图计算加速器是一种重要的图计算加速方向.然而,FPGA 的硬件编程是一项耗时费力的困难任务.高效的 FPGA 编程与开发环境是帮助底层编程人员提高 FPGA 开发与应用效率的重要工具.郭进阳等作者的文章“FPGA 图计算的编程与开发环境:综述和探索”介绍 FPGA 编程环境技术与系统的研究现状及重要问题.文章从编程模型、高层次综合、编程语言以及程序接口等方面介绍了构造良好的 FPGA 图计算编程环境的关键技术及发展状况,并结合应用领域的特征,分析了该领域存在的挑战性问题和未来的研究方向.

大数据计算获得日益广泛的应用,成为当代计算系统的主要应用,而基于 Spark 计算框架的系统成为重要的大数据计算平台.由于 Spark 是一种基于内存计算的框架,应用程序访存行为特征的信息对于程序的优化实现有重要的价值.鉴于当前 Spark 访存行为分析工具大多从同一层次上获取程序访存行为信息,许丹亚等作者的“基于 Spark 的大数据访存行为跨层分析工具”贯穿 Spark 层、JVM 层和 OS 层多个层次,建立应用程序语义和底层物理内存信息之间的联系,构造了 Spark 访存行为跨层分析工具 SMTT.文章介绍了 SMTT 的结构、内存追踪方案、不同层次之间语义关联的实现方式以及系统的访存行为分析方法.作者进行了分析工具 SMTT 的应用实例的实验,结果表明 SMTT 可以为获取 Spark 程序访存行为信息提供有效的工具,从而为 Spark 系统优化提供有力的手段.

通用图形处理器(GPGPU)以能够支持并发线程的执行而提供高性能的处理能力而获得了日益广泛的应用.然而,GPGPU 在支持不规则访存程序的执行中会因为缓存争用而降低其并发执行的效率.GPGPU 缓存子系统的优化问题是提高 GPGPU 性能的关键技术问题.张军等作者的“通用图形处理器缓存子系统性能优化方法综述”综合介绍了 GPGPU 访存子系统优化的技术途径.文章从线程级并行性(TLP)调节、访存顺序调整、数据通量增强、末级缓存(LLC)优化以及基于非易失存储器(NVM)的缓存结构等方面介绍了 GPGPU 缓存子系统优化方法,分析了优化技术的发展现状,并讨论了值得研究的挑战性问题和技术发展方向.

本专题征文发出后得到研究人员的积极响应,踊跃投稿.我们感谢广大作者对本刊的大力支持! 我们感谢审稿专家对于稿件的认真审查以及中肯的意见和建议! 希望本专题的出版对计算机领域的研究、工程开发人员、教育工作者和研究生产生有益的启发和参考作用! 我们感谢本刊编委会和编辑部开辟这样一个介绍体系结构前沿技术的专题,促进体系结构前沿技术的研究和知识传播,感谢他们对本专题的支持和和辛勤的工作!

刘志勇 研究员 中国科学院计算技术研究所
窦 勇 教 授 国防科技大学

2020 年 5 月