

2009
ANNUAL REPORT



2009 科研成果年报



中国科学院计算技术研究所



中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

2009 科研成果年报

二〇一〇年九月
中国科学院计算技术研究所科研处编

编委主任：李国杰

编委副主任：孙凝晖

编委：李锦涛 隋雪青 张光辉 徐志伟

王 凡 胡伟武

主编：王兆其

责任编辑：罗瑞丽 吴金凤

编辑：洪学海 杨 宁 杨 佳 王 琴

李 丹 高 文

通讯员：张睿超 陆 京 杨 欣 程秀云

戴 媛 李晓宇 田卫平 雷 俊

任 菲 赵静蕊 王 晴 曲 芃

所长致辞



过去几十年，我国计算机领域的科研工作大多数还是属于“填补空白”性质，原始创新较少。计算所作为国立研究所，只有走到技术最前沿的位置，才能真正成为技术的引领者。“十二五”期间是我国信息科技的战略转型期，通过 5-10 年的努力，预计新的重大成果将会在 2015-2020 年内陆续涌现。

对计算机学科而言，科研的最大动力来自经济和社会的强大需求。衡量一个国立研究所是否卓越，不仅仅要看发表文章的水平 and 引用率、在国际学术界的影响，更主要的要看对国家、社会和企业的贡献和影响。我们要把改变我国信息产业发展方式的宏观需求化解为本学科本课题组具体科研方向的选择，让科研成果真正地能为国家所用。

从今年开始，计算所每年将出版一本《科研成果年报》，目的是让用户和企业更加了解计算所的科研工作，加速技术转移，从企业的反馈意见中也可以发现更多的需求。同时也希望通过成果的交流，听取同行的批评建议，进一步改进技术。《科研成果年报》是交流的窗口，也是合作的纽带，将使计算所的科研工作更有效地融入全社会的创新价值链。

目录

战略规划.....	1
年度科研概况	3
科研成果与主要进展	6
重大科研任务进展.....	6
龙芯 3A 处理器.....	6
龙芯 3B 处理器.....	8
曙光 6000 千万亿次高效能计算机.....	9
延长摩尔定律的微处理芯片新原理、新结构与新方法研究..	11
面向蛋白质科学的高性能计算研究.....	14
研究方向科研进展.....	16
获奖成果	24
数字电路实速检测和故障诊断技术及其应用	24
自修复处理器的 3S 新原理、新方法与新结构	26
鉴定 / 验收的代表性成果.....	28
曙光 5000A 高效能计算机	28
语义资源空间的模型、理论、方法和工具	30
“天玑”网络舆情监测系统	32
传感器网络.....	34
支持大型公共设施安全问题研究的虚拟现实技术与系统 ..	36
数字内容集成分发平台.....	38
科研基地建设	40
信息内容安全技术国家工程实验室建设进展	40
学术活动.....	42
学位论文.....	44
全国百篇优秀博士学位论文简介	44
中国科学院优秀博士学位论文简介	48
学术论文.....	51
获 2009 年计算所优秀论文一等奖论文摘要.....	52
2009 年发表学术论文检索统计报告	60

战略规划

——计算所实施“创新 2020”的总体考虑

未来十年是计算所的战略转型期，科研规划的总体发展思路是：以计算设备和处理方法为核心，致力于做强中国计算机产业；成为保卫国家网域安全的“蓝军”；着重发展使得信息技术更泛在、更智能的战略先导技术；加强与健康、环境等国家重大需求相关的交叉研究。

计算所依据中科院发布的《中国至 2050 年信息科技发展路线图》提出的 U-INS (Universal, User-oriented, Ubiquitous Information Network Systems) 战略体系，制定研究所“十二五”发展战略规划，进一步凝练科技发展目标。在惠及全民 (Universal) 方向上，以低功耗高性能芯片 (CPU)、高通量计算机 (HTC)、云计算软件、海量存储等进行布局，为网络大规模普及的后端处理等提供核心技术；在无所不在

(Ubiquitous) 的方向上，以传感网、物联网、后 IP 网、LTE 无线通信等进行布局，为泛在网络提供基础设施；在以用户为中心 (User-centric) 的方向上，以低成本终端技术、为个性化服务的智能技术等进行布局，为实现以用户为中心的网络提供技术基础。以上布局相互支撑和交叉，共同促进泛在信息网络体系的建立。高效能高通量服务平台作为泛在网时代的计算所各个学科方向创新的载体。

计算所未来科研产出按引领重大产业、满足国家重大需求、促进学科发展三个模块进行总体布局，即为建立和做强国家自主可控的信息产业，提供处理器、高性能计算机、存储等核心装备，为实现惠及我国 8 亿人的信息化做出不可替代的贡献；为构建可信的国家网络空间，提供网络关防、



网络靶场、舆情分析系统等关键系统，为保障我国国家网络安全和社会稳定做出历史性的贡献；面向未来无处不在的信息网络世界，提供使得人人上网、物联网、信息应用更智能的关键技术，为我国信息化与工业化的更紧密融合做出实质性贡献。

在“十二五”期间，计算所将重大产出规划在低功耗万亿次通用处理器、千万级吞吐量的高通量计算机（HTC）、国家网络空间安全保障平台、国家网络靶场、LTE（第四代）无线通信基带芯片与协议栈等五个重要方向上。

到 2015 年，计算所不但要继续成为我国高端计算机系统和处理器芯片研制的“火车头”，具有中国特色的信息化的核心技术提供者，还要成为我国信息关防和舆情分析核心技术的骨干研究力量及网络、通信、智能等信息技术重要方向的引领者，成为我国信息技术与其他科学融合交叉前沿研究的重要基地，成为受国内外同行尊重的国际先进的研究所。

在管理上，计算所“十二五”期间的发展将以稳定发展、适度扩张为主线，在管理上要向规范的“精细管理”迈进，着力提高管理的技术含量，提高管理人员的能力水平，以实现战略性管理为目标，管理思路和管理水平要走在时代的前列。重点是创造“从优秀到卓越”的文化氛围，扎实引进卓越高端人才，建立“人人是才”的理念和有序流动的机制；完成北京市环保园计算所新园区的主体建设（112 亩土地，9 万平方米建筑面积），彻底解决长期以来计算所园区建设受限制的困扰，完善网络型研究所和高效的技术转移体系。

经过 10 ~ 20 年努力，计算所将以实际的产出和影响向全国人民和全世界证明：中科院计算所是我国创新和产业价值链上不可或缺和不可替代的一个环节，将成为信息领域企业界公认的技术源头。经过 10 ~ 20 年努力，计算所将实现从优秀到卓越的历史性跨越，成为成果、人才、管理、环境等方面国际一流的研究所。

年度科研概况

计算所创建于 1956 年，是我国第一个专门从事计算机科学技术综合性研究的学术机构，是中国科学院知识创新工程首批试点单位之一，也是创新三期研究所综合配套改革首批七个试点所之一。经过近六十年的努力，目前已发展成为由本部核心和建在上海、苏州、宁波、东莞、肇庆、台州、顺德、烟台等地的若干个分部 / 分所组成的网络型研究所。

计算所的主要学科方向为计算机系统结构、网络科学与技术、智能信息处理，主要发展目标是在计算机科学的基础理论研究和应用基础研究方面达到世界一流水平。

2009 年，计算所在认真执行中国科学院知识创新工程三期规划方案和面向 2020 年的战略规划的基础上，继续召开每年春秋两季的战略规划会和战略研讨会，在科研方向上坚持做目标明确的应用基础研究和共性关键技术(核心技术)研究。

计算所本部按照学科方向分为三个研究部，

计算所组织机构包括四个重点实验室、七个研究中心。三个研究部分别为：计算机系统研究部、网络科学与技术研究部和智能信息处理研究部；四个重点实验室包括中科院计算机系统结构重点实验室、网络重点实验室、中科院智能信息处理重点实验室以及前瞻研究实验室；七个研究中心包括：高性能计算机研究中心、微处理器研究中心、网络技术研究中心、信息安全研究中心、普适计算研究中心、集成应用中心、先进无线技术联合研究中心。

计算机系统研究部主要从事与计算机系统相关的研究，如计算机体系结构、高性能计算机系统、微处理器、编译和测试等；网络科学与技术研究部主要研究网络科学的基础理论体系、新一代网络通信 / 互联标准与关键技术、网络体系结构与系统软件平台、惠及大众的低成本网络服务软件以及网络与信息安全关键技术与系统；智能信息处理研究部主要从事智能信息处理相关的基础理论



研究和技术前沿的探索性、创新性研究；前瞻研究实验室主要从事基础性、前瞻性和交叉学科的研究。

此外，我所建有国家发展和改革委员会部署的一个国家工程实验室，即信息内容安全技术国家工程实验室；科技部部署的两个国家工程中心，即国家高性能计算机工程技术研究中心和国家并行计算机工程技术研究中心。

截至2009年底，计算所拥有在职职工702人，流动人员80人（博士后与客座访问学者），离退休人员945人。在职人员中，共有科研人员453人、科技支撑人员128人，包括中国科学院院士1人、中国工程院院士2人；正高级专业技术人员55人（研究员44人）、副高级专业技术人员193人

（副研究员98人）；中国科学院“百人计划”入选者11人、“国家杰出青年科学基金”获得者3人。国内外合作学者中，包括中国科学院“爱因斯坦”讲席教授1人，中国科学院创新团队1个（5名国外知名专家），中科院“特聘研究员”3人，客座研究员/客座副研究员110人。全所进创新岗位人员共有277人。

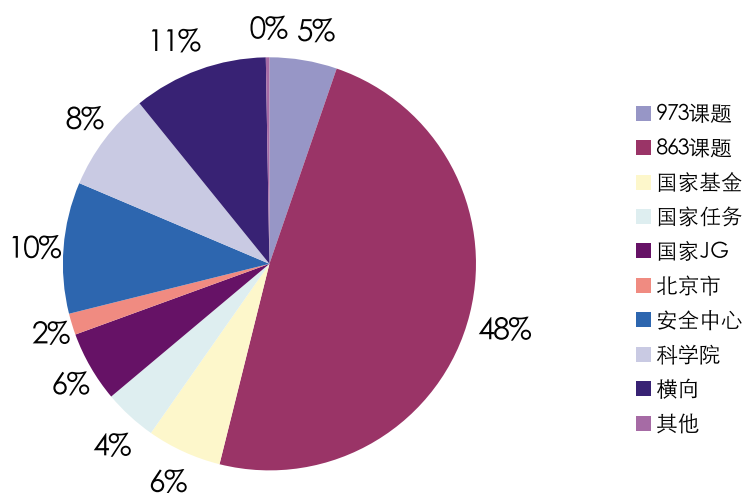
2009年，计算所共有在研项目369项。当年新立项目课题164项，包括中国高技术研究发展计划（863）项目11项、科技支撑计划项目3项、国际合作项目1项；国家自然科学基金创新研究群体项目1项、杰出青年基金项目1项、重点项目1项、联合基金1项、面上和青年基金25项；北京市创意文化产业项目1项；中国科学院知识创新工程重要方向性项目3项、对外合作重点项目1项。另外还有院地合作等其它项目。年度到款达3.2亿元。

2009年，计算所共取得科技成果31项。其中，863计划重大成果“曙光5000A”系统峰值运算速度达到233.5万亿次，Linpack值为180.6万亿次，在2008年11月发布的第32届TOP500排行榜上位列第10，是当时除美国以外世界上最快的高性能计算机，在2009年4月召开的科技成果鉴定会上受到鉴定委员会的高度评价。全年共发表学术论文562篇，其中期刊论文178篇，会议报告384篇；被SCI收录85篇，被CPCI收录212篇，被EI收录466篇，被中国科学引文数据库收录210篇；申请专利142项，授权专利152项。

2009年，计算所实现横向开发项目4821.75

课题类别	09年课题数	09年课题立项总额
863课题	11	10962.50
国家基金	36	1919.00
国家任务	10	1960.00
国家JG	5	1871.51
北京市	6	96.00
安全中心	33	4433.78
科学院	13	1605.00
研究所	10	202.50
横向	34	2259.94
非课题	6	251.50
合计	164	25561.73

2007-2009年立项课题结构



2009 年度到款比例

万元，通过分部转移成功项目收入 1833.43 万元，现有 11 家控股、参股公司，8 个分部 / 分所，从事科技开发工作的人员 731 名，实现销售收入 67035.23 万元、利税 5763.5 万元。高性能计算机、龙芯技术、网络安全、企业信息化平台、凤芯、AVS、网格、网络计算机、云计算等技术项目成功孵化，尤其高性能计算机曙光系列服务器、龙芯产品正逐渐形成规模产业化，大龙芯公司已落户北京。目前已经在长三角、珠三角、环渤海地区建立苏州、上海、肇庆、宁波、东莞、台州、

顺德、烟台共 8 个分部 / 分所，得到地方政府以股本金以及项目资金支持近 3 亿元。计算所的近百项技术通过分部辐射到当地中小企业，对当地企业核心竞争力的增强和产品利润的提升起了关键作用，带动当地超过 100 亿元的经济和社会效益。

2009 年有三位外籍教授依托计算所被聘请为“中国科学院外国专家特聘研究员”，分别是 Branka Vucetic、Paolo lenne Lopez 和 Olivier Temam 教授。

科研成果与主要进展

重大科研任务进展

龙芯 3A 处理器

2009 年计算所在研的一些重大科研项目取得重要进展，如龙芯 3A 芯片流片成功、曙光 6000 高效能计算机突破了一批关键技术等。计算所部署的各科研方向也取得了新进展。

2009 年，计算所共取得科技成果 31 项。其中，863 计划重大成果“曙光 5000A”以系统峰值运算速度 233.5 万亿次，Linpack 值 180.6 万亿次位列第 32 届 TOP500 排行榜第 10，是当时除美国以外世界上最快的高性能计算机。在 2009 年 4 月召开的曙光 5000A 科技成果鉴定会上，该成果得到了鉴定委员会的高度评价。

此外，计算所在 2009 年加大了科研基地建设力度，信息内容安全技术国家工程实验室建设进展顺利。

龙芯 3A 处理器是 863 重点项目课题“四核龙芯 3 号处理器设计”（课题编号为 2008AA010901，起止时间是 2008 年 1 月至 2010 年 12 月）的研究成果，同时也是核高基课题“高性能多核 CPU 研发及应用”的前期应



龙芯 3A 芯片

用研究平台。

2009 年龙芯 3A 取得了重要的阶段性成果，于 10 月份流片成功。这是国内第一款高性能多核通用处理器芯片，也是国内第一款采用 65 纳米工艺设计的通用处理器芯片。

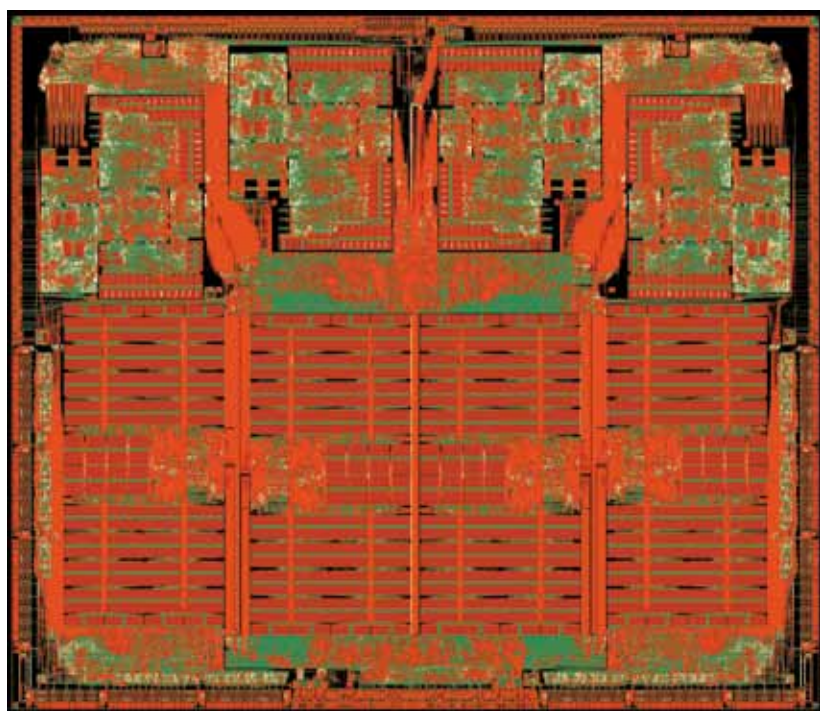
龙芯 3A 采用意法 65 纳米工艺设计，集成 4 个龙芯 464 处理器核（64 位四发射 RISC 处理器核）、4MB 共享二级 Cache、2 个 64 位的 DDR2/3 内存控制器、2 个 16 位 HT1.0 接口，最高主频 1GHz，最大功耗 10 瓦。龙芯 3A 的芯片面积为 174 平方毫米，晶体管数目为 4.25 亿个。

龙芯 3A 的每个核包含 2 个浮点乘加部件，因此峰值双精度浮点运算能力为每秒

16GFLOPS，SPEC CPU2000 Rate 的分值超过 20。龙芯 3A 的 DDR2/3 控制器的最高工作频率为 400MHz，HT 接口的工作频率为 800MHz，因此内存带宽为 12.8GB/s，IO 带宽为 6.4GB/s。

龙芯 3A 的创新之处有：可伸缩的多核体系结构；可配置的片上存储；层次化的 Cache 一致性协议；X86 二进制翻译的硬件加速。

龙芯 3A 的相关研究成果发表在著名国际会议和期刊上：处理器结构相关的工作发表在 IEEE Micro 2009 和 Hotchips 2008；验证相关的工作发表于 HPCA 2009；IO 优化相关的工作发表在 HPCA 2010；X86 二进制翻译加速的相关工作发表于 ICCD 2009。



龙芯 3A 芯片版图

重大科研任务进展

龙芯 3B 处理器

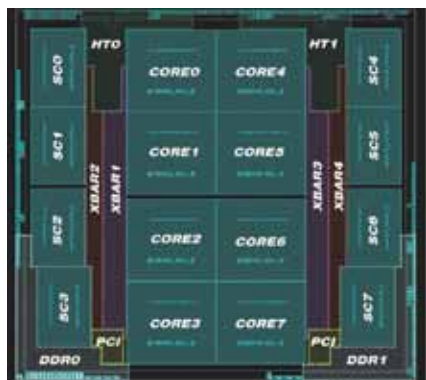
龙芯 3B 是核高基项目课题“高性能多核 CPU 研发与应用”(项目编号 2009ZX01028-002-003)的第一款芯片,主要应用于曙光 6000 千万亿次高性能计算机和专用信息处理系统。

龙芯 3B 采用意法 65 纳米工艺设计,主频 1GHz,功耗 25 瓦,片内集成 8 个 64 位的龙芯向量处理器核,4MB 的二级 Cache,2 个 DDR2/3 内存控制器,2 个 16 位的 HT1.0 接口,1 个 PCI/PCI-X 接口和 1 个 LPC 接口。

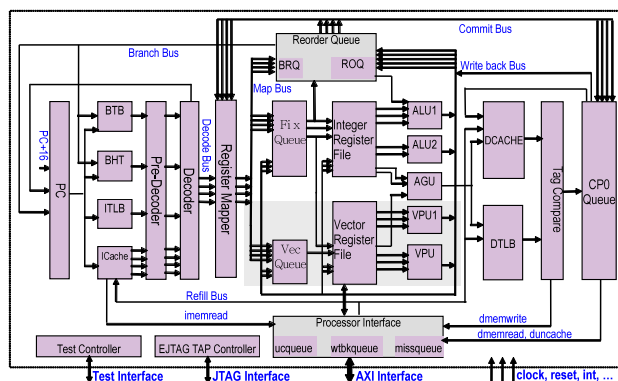
龙芯 3B 的最大特色是龙芯向量处理器核的设计。该向量处理器核将龙芯 464 通用处理器

核的浮点部件和浮点寄存器堆替换为 2 个 256 位的向量处理部件和 1 个 128x256 位的向量寄存器堆,使用龙芯 3B 在 1GHz 的工作频率下峰值双精度浮点计算能力达到每秒 128GFLOPS,8000 颗龙芯 3B 即可达到每秒千万亿次的计算能力。

龙芯 3B 设置了 300 多条专用的向量处理指令,并建立了基于构令的全流水数据通路,使得高密度的计算和高吞吐的数据传输能够协调一致。在龙芯 3B 平台上,Linpack 的理论效率达到 90% 以上,FFT 的效率达到 87%。



龙芯 3B 芯片布局



向量核结构框图

重大科研任务进展

曙光 6000 千万亿次高效能计算机

曙光 6000 千万亿次高效能计算机系统研制是由中国科学院计算技术研究所、曙光信息产业有限公司等单位共同承担的国家 863 计划“高效能计算机与网格服务环境”重大专项研究课题，起止时间是 2009 年 1 月至 2010 年 12 月。

曙光 6000 系统将包括两个部分，一部分为 HTC 分区，即服务分区，将采用基于 X86 处理器的星群架构，以满足用户大量商业化应用及服务运行需求；另一部分为 HPC 分区，即计算分区，将采用基于异构处理器的超并行（Hyper Parallel Processing, HPP）体系结构。计算和服务分区共享基础架构、网络存储和机群软件。系统中有 3 套互连网络，分别是：数据网 / 存储网（40Gb InfiniBand 网）、全局同步网（3.125Gb 专用网）、管理网（千兆以太网）。

HTC 分区主要由曙光 TC3600 刀片式服务器节点组成。TC3600 是全球首款完全符合 SSI 刀片标准的服务器系统，支持所有 SSI 兼容的刀片，具有高速网络交换模块、低速网络

交换模块、InfiniBand 子卡和系统管理模块等多种可灵活选配的模块。TC3600 为 10U 10 片结构，在 10U 高度的机箱中，最大支持 10 片热插拔刀片，还同时支持 10 个热插拔 IO 扩展刀片。TC3600 采用了多个部件冗余设计技术，具有更好的可靠性、可用性，完全能够为企业用户、数据中心等提供高效率的不间断服务。

HPP 是中科院计算所为了同时应对千万亿次高性能计算和效用计算的挑战而提出的新型体系结构。HPP 的主要技术特点包括：具有芯片级、板级、节点级三级并行结构，提供多通道并发核到核通信机制；计算节点内具有全局地址空间，支持多种编程模型，提高了系统并行度；超并行节点 - HPP 节点是包含 X86 和龙芯处理器的异构系统，具有单一操作系统映像。其中，以 X86 处理器作为管理单元，运行操作系统服务，实现对主流软件栈的兼容；以低功耗的龙芯处理器作为计算单元，运行计算任务，从而解决功耗问题。HPP 节点具有高计

算密度和高功耗密度特点，降低了系统的空间需求。

目前，中科院计算所正在加紧研制该系统，TC3600 刀片服务器节点已经完成了设计开发

并正在测试，HPP 系统控制器等重要部件正在研制中。预计 HTC 分区将于 2010 年 5 月完成，随后交付国家华南超级计算中心使用，HPC 分区也将于 2010 年底问世。



曙光 6000（星云）系统

重大科研任务进展

延长摩尔定律的微处理芯片新原理、新结构与新方法研究

“延长摩尔定律的微处理芯片新原理、新结构与新方法研究”（项目编号为 2005CB321600）是计算所李国杰院士作为首席科学家承担的国家重点基础研究发展计划项目（973 项目），起止时间为 2005 年 12 月至 2010 年 12 月。2009 年取得的进展如下：

1. 完成了可扩展的 16 结点 Godson-T 处理器的设计、万亿次原型芯片 TGAP 的结构设计、逻辑设计和高性能低功耗多核时钟设计，以及利用片上存储实现独立的 I/O DMA Cache 的研究。

2. 开发了对龙芯平台上的可执行程序进行动态优化的系统。进行了栈变量提升、基于内存池的动态优化、共享 Cache 的动态划分等研究。提出了流敏感、上下文敏感的指针分析技术，解决需要大量存储空间问题的自适应任务创建策略，以及解决二进制定态优化中的非对齐访存问题的基于剖面分析和异常处理相结合的处理策略。

3. 提出了高速低功耗浮点乘加器的优化设计、

结构级自适应门控电源设计，以及降低测试功耗技术的方法。提出重叠传输周期提高吞吐率和一种用电源门控降低空闲阶段功耗的方法。完成了 TGAP 原型芯片高性能低功耗多核时钟设计。

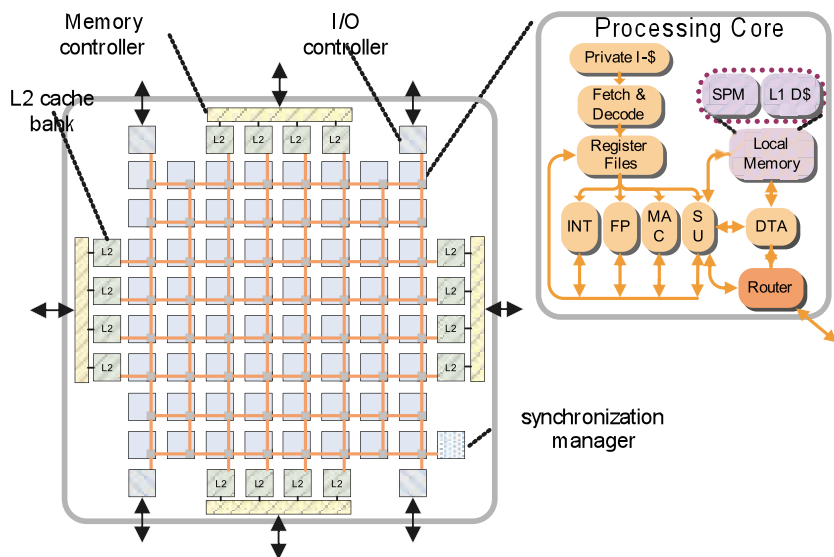
4. 提出了存储器内建自修复方法、众核处理器自修复和任务调度方法、基于间歇故障脆弱因子的可靠性评估方法、考虑老化效应与软错误的统一故障检测方法和缓解老化效应的向量控制方法，以及考虑工艺偏差的关键通路选择及其测试生成方法等，并采用体系结构脆弱因子评估基于 3S 技术可靠性设计与性能之间的优化关系，将可靠性设计技术集成到原型芯片 Godson-T 中。

5. 提出了处理器的事务级建模方法及其验证技术、避免非功能状态的时延测试生成算法和多测试源共存情况下的测试时间优化方法。进行了 Godson-T 芯片的可测试性设计和验证。

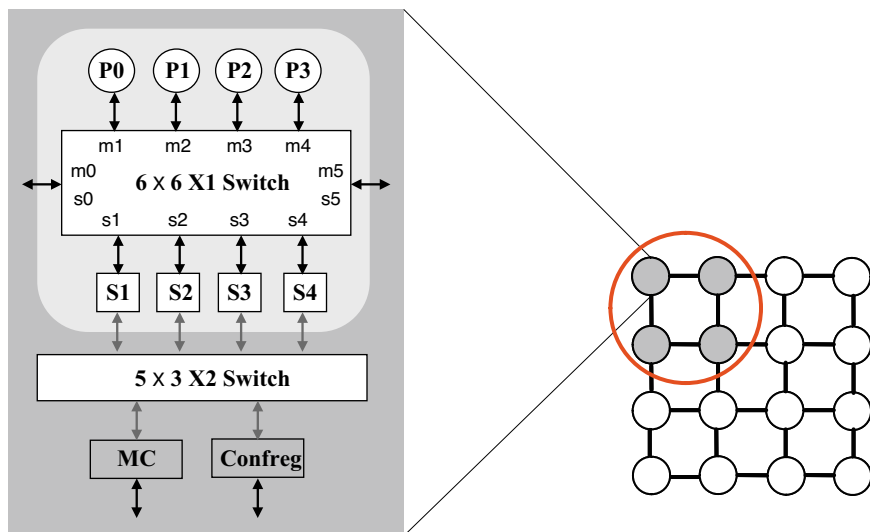
6. 本年度共发表或录用国际学术论文 58 篇，国内学术论文 21 篇；获得授权国家发明专利 4

项，申请受理国家发明专利 20 项。相关研究成果发表在 IEEE Micro、IEEE Trans. 等国际期刊，ISCA、HPCA、ICCD、CGO、DATE、Euro-

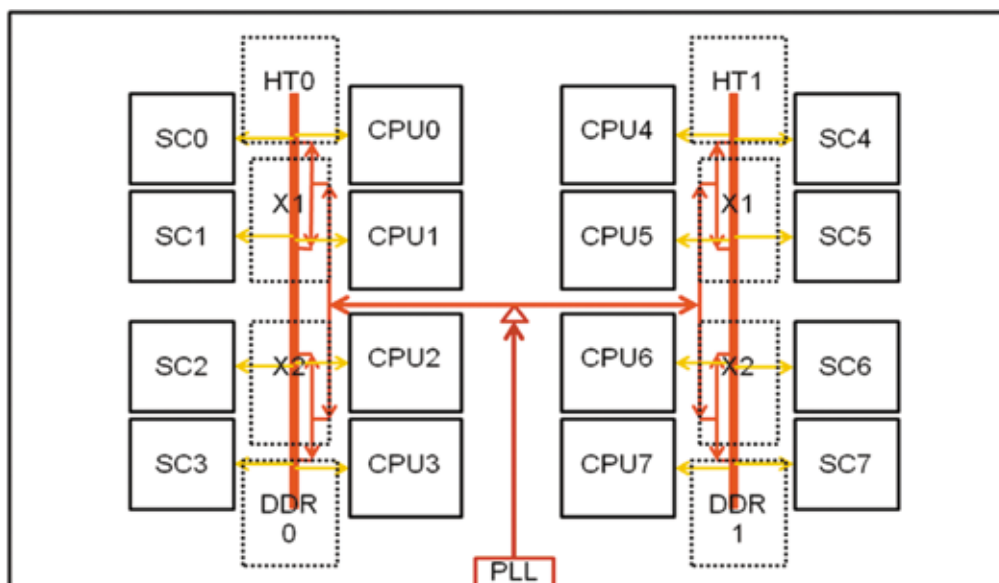
Par 等国际会议，以及 JCST、《计算机学报》《计算机研究与发展》《软件学报》和《计算机辅助设计与图形学学报》等国内主办的学术刊物。



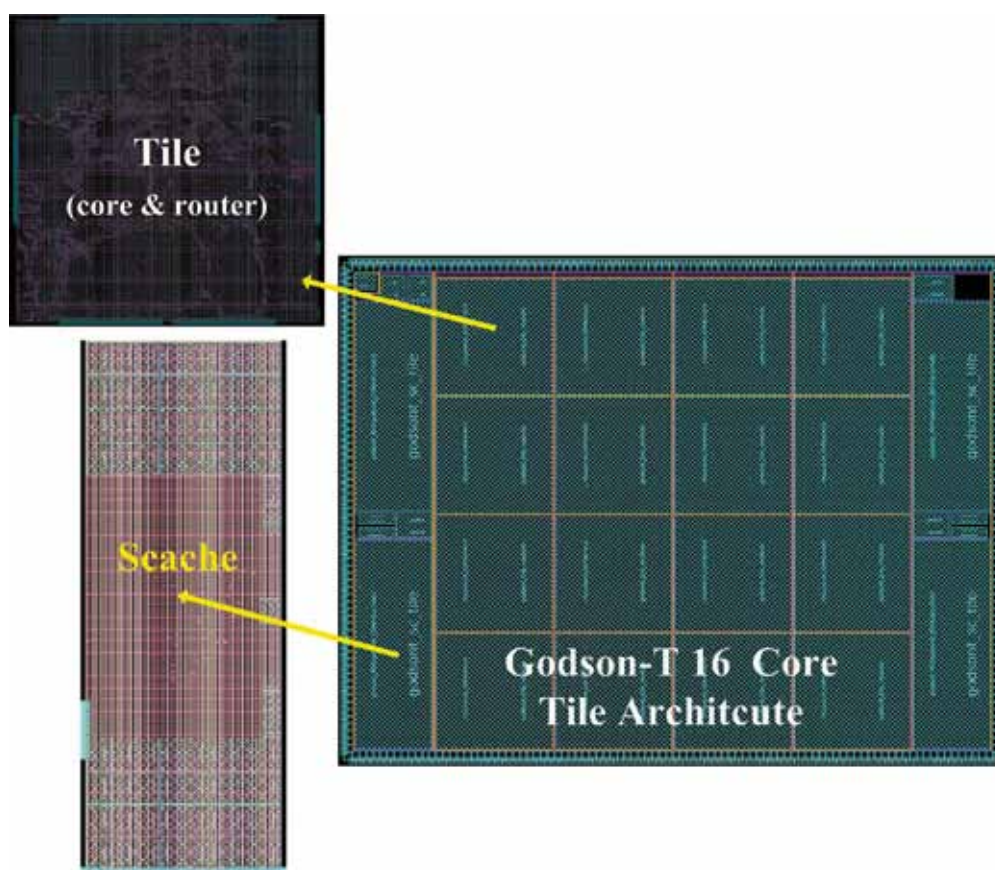
Godson-T 众核芯片结构示意图



万亿次原型芯片 TGAP 的结构示意图



TGAP 的全局时钟网络拓扑结构图



Godson-T 原型芯片顶层版图



重大科研任务进展

面向蛋白质科学的高性能计算研究

计算所牵头承担的中国科学院知识创新工程重大项目“面向蛋白质科学的高性能计算研究”(项目编号为 KGCX1-YW-13), 起止时间是 2008 年 1 月至 2010 年 12 月, 合作单位有上海生命科学研究院、中科院生物物理研究所、中科院计算机网络信息中心、中科院软件研究所等。计算所承担了其中的三个课题, 分别是: 课题三——超龙一号高性能计算机系统, 课题四——蛋白质修饰鉴定的算法研究与软件开发, 课题六——冷冻电镜二维图像处理方法与三维重构研究。截至 2009 年年底, 计算所承担的各课题取得了如下一些重要的阶段性成果:

1. 课题三: 形成了采用 HPP 体系结构、具有异构特点的超龙一号技术框架; 完成了 HPP 系统控制器设计; 完成了全局同步网络交换机原型设计; 完成了一款 HPP 原型系统; 开展了可重构计算研究, 实现了基于 FPGA 的动态可重构加速

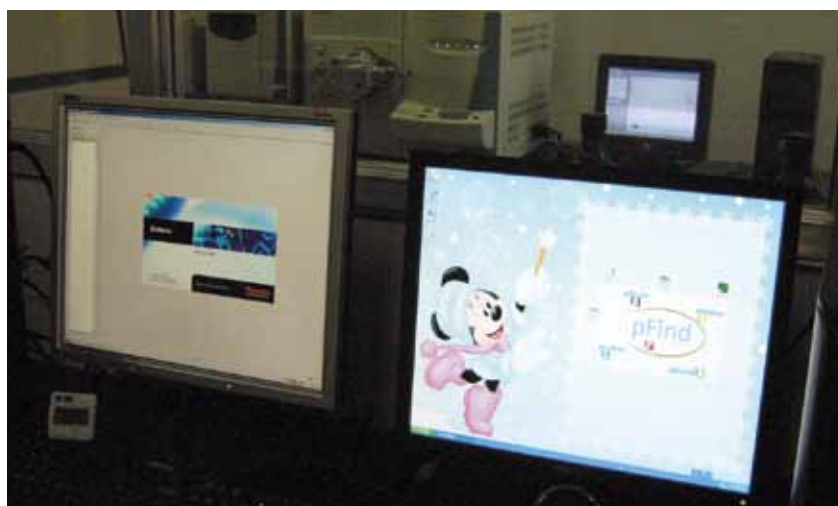
卡; 实现了基于 GPU 的加速计算; 对 Eman 的核心 Kernel 进行了加速测试, 已经验证了实现的正确性和有效性。

2. 课题四: 开发了蛋白质鉴定软件 pFind2.2 并行版, 实验表明, 其加速比效果大大超过了 SEQUEST、Mascot 和 X!Tandem 的现有报道加速比效果; 研究了基于谱图聚类的修饰检测算法, 成功检测到了蛋白质修饰; 建立了核心岩藻糖修饰肽谱库系统等。上述成果已投稿国际期刊。另外, 进行了 pFind 软件推广, 为使用单位新安装 pFind 系统 26 套, 包括 2 套并行版, 升级 43 套。

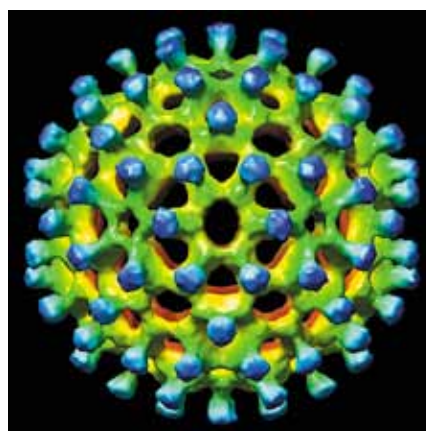
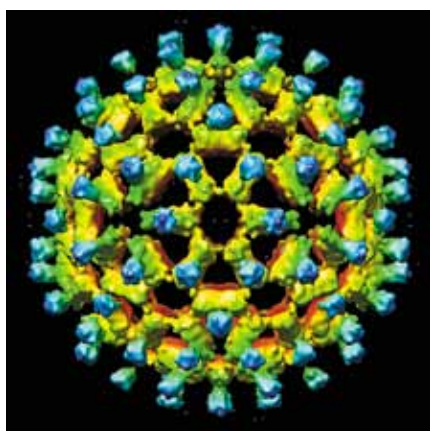
3. 课题六: 开发了冷冻电镜单颗粒识别软件 Picker, 颗粒挑选识别率达到 95.6%, 明显优于当前国外的同类软件; 开发了冷冻电镜三维重构并行软件 ParaEMAN, 在曙光 5000A 高效能计算系统(1024 个处理单元)上得到了

接近 60% 的并行效率。目前这些软件已经在精度和速度上赶超了国外流行软件，并且已经在中科院生物物理所应用；研究了新的冷冻电镜

三维重构算法，得到国际同行的高度评价；在国内外知名学术期刊和会议上发表了 8 篇学术论文。



pFind 系统与国外商业软件在中科院动物所质谱室蛋白质鉴定服务中同台竞技



三维重构结果精度对比（左：EMAN 的结果，右：我们系统的重构结果）



研究方向进展

研究方向进展

VLSI 测试、验证与可靠性设计

以保障芯片质量和成品率、提高计算机系统可靠性为目标,深入研究 VLSI 测试、验证与可靠性设计方法学和关键技术。2009 年主要进展有:在测试与验证方面,提出了基于错误屏蔽概率的静态可观测性分析方法、基于游标延迟线的片上通路时延测量电路结构、考虑工艺偏差的关键通路选择和测试生成方法、协同优化捕获与移位功耗的测试向量 \times 位填充方法等;在可靠性设计方面,提出了众核处理器的自修复和任务调度技术、基于间歇故障脆弱因子的可靠性评估方法、面向任意故障模型的多故障诊断方法和基于怀疑窗的后硅调试机制等;在 16 核原型芯片 Godson-T 中集成应用了验证、测试与可靠性设计相关技术。上述部分成果发表于 ISCA2010、DATE2010 和 VTS2010 上。

编译

计算所的编译团队是国内最著名的编译研究

队伍。2009 年,该团队在面向曙光 6000 的编程环境中,对 UPC 语言进行了层次并行性扩展 (HUPC),以支持现代并行体系结构中日益加深的存储层次;面向龙芯的编译系统及工具链(国家“核高基”重大专项项目)的研发取得了良好的进展,龙芯编译系统的性能和二进制翻译系统效率均已达到或接近项目最终指标。该团队研究了迭代编译中输入集对编译优化的影响;提出了一种基于指向层次的高效的流敏感上下文敏感指针分析,可有效分析百万行规模的 C 语言程序;提出了一种动态检测方法,以较小的代价检测出 MySQL、Apache 等大型程序中潜在的一些违反顺序一致性的错误;提出了多种动态存储优化策略,有效提高数据访问效率;提出了一种自适应的任务创建策略 Adaptive TC,有效改进了现有的 work-stealing 技术;提出了一种面向普通程序员的基于函数的并行编程方法,有效提高了程序开发效率。上述部分成果发表于 PLDI2010、CGO09 和 CGO10 上。

高性能体系结构

完成了访存 trace 采集工具 HMTT2.0 的研制, 可以支持对 DDR2 内存访问地址的长时间实时采集和分析。并基于此工具开展了 DMA cache 的研究工作。完成了基于 GPU 的蛋白质冷冻电镜重构算法库 GPU-EMAN 的实现和优化。对基于多核和众核结构的图分析、稀疏矩阵向量乘等算法优化取得进展。完成了一种基于 Linux 的非对称内核结构操作系统 GenerOS 的设计和实现, 可以将多核系统分为应用核和多个不同功能的 OS 专用核。完成了并行模拟器框架 SimK 的设计开发, 基于此开发的模拟器最大可以进行 1024 核的 HPP 全系统模拟。该框架可以支持 SMP 和集群两级并行, 在 SMP 下可以获得超线性加速比, 在基于 IB 的集群上可以获得 50% 以上的效率。

操作系统

研究基于虚拟化技术的能力服务计算相关理论, 设计并实现了能力流动架构及一系列局部、全局能力流动算法。为降低系统能耗, 从操作系统角度构建了绿色云计算模型—LogPCM 模型。研制具有高可靠、可信、扁平化的分布式虚拟机监控器系统 (DVMM), 解决支持多核处理器的跨节点资源动态流动问题, 包括设计并实现了内存、磁盘、I/O 的跨节点流动机制。围绕 863 课题“分布式 I/O 资源虚拟化技术研究”, 实现了完整的分布式虚拟机 I/O 架构及其中的关键技术。围绕基金委重点项目“基于虚拟机架构的可信计算环境与可信软件设计”, 设计了可信的虚拟平台架构, 并实现了其中的可信远程 I/O、可信接入控制等关键

技术。基于上述工作设计并实现了 TRAINBOW 原型系统。

机器学习与数据挖掘

2009 年获得国家自然科学基金重点项目“WEB 搜索与挖掘的新理论与方法”, 负责人: 北京大学李晓明、计算所何清; 较好地完成了国家 863 计划课题“基于感知机理的智能信息处理技术”, 2009 年 11 月 29 日顺利通过验收; 受中国移动研究院委托, 组织开发完成了基于云计算的并行数据挖掘平台用于电信 TB 级实际数据的挖掘, 实现了高性能、低成本的数据挖掘, 这是基于云计算的数据挖掘在中国的最早的实践。现正在开展数据挖掘云服务研究开发。

知识网络

大规模分布式互联网络上的资源有效管理始终是一个重要的研究课题。知识网络组面向未来网络互联环境, 围绕规范组织、语义互联和智能聚融这三个科学问题开展系统的基础研究工作, 提出资源空间模型、语义链网络模型等基础模型和方法。针对信息物理社会的互联网络资源管理开展了前瞻性研究工作, 探讨了交互语义及其在信息物理社会互联环境中的内在影响机制, 提出了未来信息物理社会的基础框架模型, 为进一步开展信息物理社会中的资源高效管理研究奠定了基础。

自然语言处理

2009 年, 自然语言处理组在论文发表、国际

评测和推广应用等方面均取得了一定的进展。研究组在顶级国际会议 ACL 和 EMNLP 上共发表 8 篇论文,在论文质量和数量上均位居国际科研机构前列。在国际影响力最大的 NIST 机器翻译评测中,研究组在汉译英任务中总成绩名列第 3,这是亚洲科研机构取得的历史最好成绩。研究组开发的专利翻译、移动设备翻译和汉韩新闻翻译等系统均获得企业用户的认可和称赞,表明研究组的统计机器翻译技术正逐步走向实用化。

智能科学

圆满完成国家 863 计划课题“软件自治愈与自恢复技术”,在基于模型的故障诊断和自治计算方面进行了深入的研究,研制了网构软件自治管理器 eHealer 系统,具有重要的应用前景,并已获得软件著作权;提出一种心智模型 CAM,它由意识、记忆和高级认知功能构成,可能为新型的类脑计算机研究提供一种理论基础。

网络舆情研究

该方向以海量互联网信息的获取、挖掘和处理为研究目标,2009 年度的主要进展包括:针对论坛、博客、新闻提出了高效的可扩展的获取算法;在网络信息多层次舆情特征提取方面,提出了垃圾过滤算法,基于网络资源内在特性与人工规则相结合的元信息抽取和识别算法;大规模文本流的先行分类和层次化聚类算法的研究;用户行为日志的信息挖掘算法的研究;信息网络建模与网络社区分析以及网络评论的倾向性分析算法的研究等;上述部分成果在 ACM SigIR2009、

ACM WWW2009、ECIR2009、NAACL2009、ACL2009、CCIR2009、计算机学报等会议和杂志上发表;部分关键技术应用到 TREC2009 的 blog 任务和 web 检索任务中,取得了不错的效果,部分成熟的关键技术集成到天玑舆情系统中,已在政府、金融、教育等行业取得了良好的应用效果。

网络计算系统软件

主要进展是:在大规模分布式网络计算系统体系结构方面,分别面向网格计算和个人网络计算的需求,提出了网格的层次化虚拟组织体系结构,以及面向个人的跨域资源受控共享体系结构模型,解决网络资源组织、访问控制和应用构造三方面的问题。围绕网格软件基准测试目标,综合科学计算、数据共享、流程处理、网格文件存储等网格系统典型应用场景,在考虑请求到达时间分布和负载组成的基础上,提出了基于无穷队列的基准负载模型并据此实现基准测试集和工具。在面向个人的分布式数据管理方面,提出一种不依赖于任何单一服务的个人互联网存储方案,在保证个人数据可靠性的同时,提高“冷数据”的可用性。通过分析 5 万名 ADSL 用户的上网记录,从中发现用户的社会性行为对网页访问概率分布有一定的影响,一个网站的社会性越强,这个网站的网页访问概率偏离 Zipf 分布的可能性越大。在系统软件的性能优化方面,针对大规模数据处理需求,提出了基于列存储的分布式数据存储结构,使存储空间节省 25%,数据处理性能提高 30%,其开源实现已应用于 Facebook 公司。

此外，在作业调度优化和分布式并行查询优化方面均产生了阶段性成果。上述研究的相关成果发表在 JoSC, PDCAT 等刊物上。

众核设计

围绕超龙计划等战略研究目标，结合生命科学计算、网络多媒体和安全等战略应用，研究片上众核并行处理器设计方法和可扩展可重构处理器设计原理等挑战性前瞻问题。目前的主要研究进展有：已经建立 64 核众核 Godson-T 模拟平台、性能监视平台、以及 Runtime 系统和编译系统等相关工具链；开展了众核处理器的结构优化及应用程序优化等关键问题研究；已经建立 Godson-T RTL 和 FPGA 模拟验证平台，并开展了具有模拟功能的研究性流片；已开展具有高度可扩展性的单片多线程处理器结构研究。

跨媒体计算

研究网络环境下大规模跨媒体信息智能计算的理论与关键技术，着力解决其在大规模网络多媒体应用中所面临的精度、速度以及数据等瓶颈问题。在局部视觉特征、语义概念映射、自适应反馈、大规模路径剪枝等方面提出了一系列有效方法，在国际著名学术期刊上发表论文 5 篇，独立实现了国际顶级会议 SIGIR 长文的突破，在视频检索国际评测 TRECVID 和国际 WEB 服务比赛中取得了第一名的优异成绩。提出了基于片上并行处理的视频计算方法，获得了大规模视频检测的性能 / 功耗比的显著提升，相关成果在国家广电总局等单位取得良好应用效果。

虚拟现实

探讨了正常情况下环境因素对人类行为的影响，提出了一种复杂环境中群体疏散模型及其并行求解算法，并将其应用于上海世博会客流疏导方案模拟。在火与烟雾的建模方面，提出了一个燃烧现象的可计算物理模型以及燃烧蔓延模型，比传统的粒子模型能产生更加逼真的视觉效果。探讨了人体运动的实时跟踪与控制问题，解决了虚拟环境中“用户化身”和“虚拟化身”协同工作的难题，并将其应用于酒泉航天测试发射仿真训练中，获总装成果鉴定。

生物信息学

生物信息学研究组 2009 年继续在计算蛋白组学这个新方向上，从序列 - 结构 - 功能三个方面开展研究工作。在蛋白质序列鉴定方面，研究了基于质谱数据的蛋白质及其修饰鉴定和定量分析算法，完成了蛋白质鉴定软件 pFind 单机新版及集群并行版，在国内外推广应用。在蛋白质三维结构预测方面，设计了蛋白质结构预测的高效采样算法，为生物学家提供计算服务，并应用于癌症早期诊断技术研究；研究了艾滋病 / 肺结核的传播趋势及其防治措施。在蛋白质冷冻电镜三维重构方面，完成了冷冻电镜单颗粒图像识别软件 Picker、三维重构软件 ParaEMAN；研究了新的三维重构算法。在复杂生物网络方面，通过构建分析双色生物网络，结合医学实验数据，预测了非编码 RNA 基因功能及在疾病中相关的作用。部分研究成果已发表在 ISMB、Bioinformatics、Molecular and Cellular Proteomics、BMC

Bioinformatics、BMC systems biology、Theoretical Computer Science 等知名会议和期刊上。

传感网

该方向密切结合国家重要应用需求，坚持理论和应用相结合的发展模式，研究传感网的基础科学问题和关键技术。拥有从智能感知、网络传输到网络支撑管理等各层次系统化的核心技术，以及文物保护、水资源、环保等实用化系统应用成果。2004 年起，承担和参与了二十余项传感器网络相关国家科研任务，包括 NSFC、CNGI、863 探索导向及目标导向、973 项目、国家科技支撑项目等。目前承担和参与六项“新一代宽带无线移动通信网”国家科技重大专项的传感器网络研发和产业化项目。

移动互联网

2009 年度集中于移动互联网网络架构及其关键技术研发。在架构方面，提出了无线局域网与蜂窝移动通信网络的融合构架；在融合网络移动管理方面，提出了支持移动节点在异种网络间高性能、高灵活性移动的切换算法和兼容 IPv4/IPv6 的通用位置管理协议，并实现了双栈移动 IP 和代理移动 IP 系统；在异构网络联合资源优化方面，深入研究了多宿多流并发传输关键技术，结合无线网络 MAC 调度、频谱分配、速率控制等资源分配创新算法以达到网络资源的充分利用。另外，通过与联通、联想等公司的密切合作，积极尝试上述关键技术在业界的实际应用，并研发综合业

务服务平台，进行实际试运营。

网络测量

网络测量是通过测量网络、业务与用户行为，建立网络与业务的科学模型，提出网络资源优化方法，为构建可测、可管、可控的 IP 网络奠定基础。课题组开展了面向业务感知的网络测量算法与实现技术、大规模网络测量部署策略与建模分析方法、流量控制与资源优化机制等三方面的研究，进展如下：(1) 提出了超高速链路数据包捕获与自适应非线性抽样方法，设计了精确抽样与快速业务流匹配算法，实现了精确高效的新型业务流量识别。(2) 研究了网络、资源分布与新型业务行为描述基础理论，在实际网络测量的基础上，提出了网络与业务行为模型化方法，获取全网多维指标，刻画精确的网络与业务行为特征。(3) 研究了通信子网与资源子网的形式化表达理论，提出了高效的网络流量控制与网络资源优化方法，解决 P2P 资源高效组织与定位。

网络路由监测分析

本课题组致力于网络监测分析的基础理论和关键技术研究，并在此基础上研制具有自主知识产权的相关系统，为新一代可运维、可管理、高可信的互联网与业务系统提供支撑手段。2009 年度，在国家自然科学基金项目 - 网络路由策略一致性验证方法研究中，提出了面向 ISP 的域间路由策略推断模型 RPIM，提出了一种基于规则的前缀劫持检测方法 RSPD，并基于公网路由数据进行了有效验证。在计算所科研平台支撑

基金项目—面向大规模网络仿真应用的网络组件平台中，设计并开发了网络路由模拟仿真平台 RealEnv-Plan，通过模拟实际网络中的报文交互、选路策略及路由过程，可真实地再现实际网络状况，向上层提供策略验证、故障定位、协议评估等功能。在支持国家信息化建设的相关产品研究与开发上，成功研制了网络路由流量监测分析平台 TrafficXpert，将网络中两大主要元素——路由和流量进行有机结合，并成功在实际网络中部署。

网络可信控制

课题组对分布式网络和无线网络环境下的可信控制技术研究，发表学术论文 20 余篇，申请国家发明专利 8 项。在分布式网络可信控制方面，对 P2P 网络的全局信任模型及分布式信任控制开展了相关研究，研发了可信任 P2P 应用系统，解决 P2P 系统中的“节点自私性”及“不可靠服务和欺诈”两大典型问题，大大提高 P2P 网络中共享资源的可信度、可用性和利用率。该系统即将在教育网上进行百万用户量级的示范应用。在无线网络环境下的可信控制方面，研发了可信无线 mesh 网络实验平台，结合博弈论和模糊理论，研究无线 Mesh 网络环境下的动态信任模型，解决无线资源分配、路由等方面的节点自私性和恶意攻击问题，构建可信可控的无线 Mesh 网络。

网络新媒体研究

项目组以“十一五”国家科技支撑计划重大

项目——数字音像内容集成分发平台（于 2009 年 5 月顺利通过科技部项目验收）为研发基础，面向数字内容服务行业的新需求，构建可管可控、具有版权保护功能的数字内容集成与分发系统，实现基于 IP 网络的数字内容存储和传播，为终端消费者提供便捷高效的数字内容服务。项目组针对融合 CDN 与 P2P 特征的数字内容快速分发、数字内容的网络化版权管理、P2P 网络的流量规划与控制、P2P 应用的通用缓存服务等关键技术进行了深入研究，形成多项具有创新性和自主知识产权的技术成果，其中发表论文 42 篇，提交发明专利申请 18 项，获得发明专利 2 项，获得软件著作权登记 15 项。

信息安全

信息安全中心 2009 年继续发扬在网络信息安全方向满足国家重大战略需求的优势，研制的重要系统在工信部、广电总局得到应用，在奥运安保、国庆 60 周年等重大敏感时期做出了实际贡献，获得了中央领导的肯定和赞扬；在网络信息安全方向的研究方面，基于 973 项目和 863 重点项目的支持，在正则表达式匹配、数据流挖掘、网络异常大流识别算法等方面取得了突破，在顶级国际会议上发表了多篇论文；同时，推进了信息内容安全技术国家工程实验室建设，为建设网络信息安全研究长效机制和环境作出了贡献。

嵌入式系统

在国家 863 计划重点项目以及自然科学基金等支持下，重点开展了泛在设备互联技术、情境

感知技术和基于位置的服务技术的研究。完成了《信息设备资源共享协同服务（IGRS）设备语义描述》标准框架草案 1 项、提交国家实时定位标准草案 1 项；开发了面向 Windows 和 Linux 平台的泛在设备互联中间件系统；研制了支持 IGRS 和 UPnP 设备互联的桥接器原型产品；研制了多超声波的高精度定位节点和低功耗便携式 WiFi 定位标签等产品原型以及基于低功耗 WiFi 节点的温室大棚远程监测系统。

感知计算

重点开展了基于低成本传感的新型感知交互技术与系统的研究，该研究基于传统嵌入式平台，集成加速度计、陀螺仪等低成本传感器件，并配以自主研发的低功耗高性能智能识别算法，能够方便地以手持或可穿戴方式，采集多模运动动作或人体行为数据。通过采用坐标自映射以及人机界面协同控制等技术，能实现在三维空间感知手势动作意图，从而完成以更加自然便捷的方式进行交互，同时通过可穿戴方式对诸如摔倒等异常行为进行感知识别。关键技术及系统可通过行业定制广泛应用于白板教育、家庭娱乐、老人监护等领域。

实时信号处理

重点开展了面向星载 SAR 等实时信号处理技术的研究，完成了高速信号源与记录器的高速通路的联试、I/O 与 PE 的硬件研制以及仿真信号的软件和成像软件。研制用于实时成像处理的 GPU 桌面基础系统。完成了中科院卫星遥感地面站委

托项目 Radarsat1 成像处理软件的研制合同。完成了光学遥感卫星遥感图像的云判算法及实时实行的研究工作。

自然人机交互

重点开展了残疾人信息无障碍人机交互技术及智能终端方面的研究，研发了面向盲人上网的服务器端语音推送系统，该系统采用智能网页分析技术和多种人机交互技术，可帮助用户快速定位到感兴趣的内容。研发了符合盲人输入习惯的计算机盲文点字输入装置，可接收盲文点字输入并将其自动转化为汉字，准确率 94% 以上。在上述技术的基础上研发低成本便携式盲人计算机原型系统，该系统在 2009 年 5 月“科技助残行动计划”研讨会上做了展示，受到了领导和专家的好评。

先进无线技术

LTE 基带处理芯片项目 2009 年的主要进展如下：在基带算法设计方面，完成了 LTE 终端物理层基带处理算法的设计和浮点模型建模的工作，同时完成了部分算法的定点化等工作。在芯片设计方面，完成了 LTE 终端基带处理芯片的架构设计，实现了包括 Turbo 译码，Viterbi 译码和速率匹配等功能在内的外接收机以及包括混合基 FFT、IFFT、上行参考信号生成、信道编码等功能在内的上行发射机硬件加速器的设计，同时完成了包括加解密算法模块等在内的部分 SoC IP 的设计。在原型验证系统方面，完成了基于 FPGA 验证平台的硬件设计、FPGA 基础模块设计和集成，已

经开始进行基带处理芯片硬件 IP 的验证工作。在协议栈方面，完成了面向 LTE 终端系统的协议栈软件系统，该软件系统可稳定运行于基于 LS232 200MHz 的验证平台和基于 ARM9 200MHz 的验证平台上。针对 LTE 基带处理芯片申请相关专利 8 篇，发表学术论文 4 篇。

空间计算技术

在 863 目标导向项目“基于先进计算架构的高性能空间分析中间件”和“基于新软硬件体系结构的网络地图服务器关键技术”支持下，对中大规模以时间域中空间数据及其关联的自然和社会属性数据为研究对象，初步形成拓扑分析、叠加分析、网络分析、缓冲分析、地统分析等的高性能计算体系，并在变异空间数据计算方面形成了十余项专利和一套高性能空间计算中间件。目前正在进行点空间数据智能化分析技术的研究，以期望解决点要素，如固定位置传感器和基础设施，非固定位置的 RFID 物品和移动目标等的信息管理、数据同化和模拟及智能化处理技术。

软件集成与服务计算

研究如何有效地支持开放环境下 IT 资源的无缝集成、有序共享与即时协同，让人们能“多、快、好、省”地构建分布式行业应用。已经形成了自主知识产权的 VINCA 服务集成方法和核心技术以及 VINCA 服务集成套件，初步形成专利和软件登记群。通过专利等成果转化，对产业已经产生实质性贡献。研究成果辐射到北京市产业联盟并在

国家科技信息资源中心和 20 个省级科技信息服务节点部署应用了具有分布式架构的“科技信息资源联动系统”。

存储系统

2009 年主要在海量文件系统和块存储设备系统两个方面继续深入研究，主要进展为：1) 在文件系统方面，研究并实现了符合并行网络文件系统 NFSv4.1 标准的海量文件系统，并在此基础上，研究并实现了文件级快照，可以针对文件或目录完成秒级高频度快照；文件级冗余技术，支持 RAID1 与 RAID5 级别；2) 在块存储设备方面，主要实现并完善了高性能的按需分配存储系统 ASD，分配粒度从原有系统的 16MB 做到目前的 128KB，而保持等同性能；研究并实现了网络冗余存储系统 BW-RAID 原型，能够实现多个存储节点间的数据冗余，并通过异步冗余计算较显著地优化冗余访问性能。

获奖成果

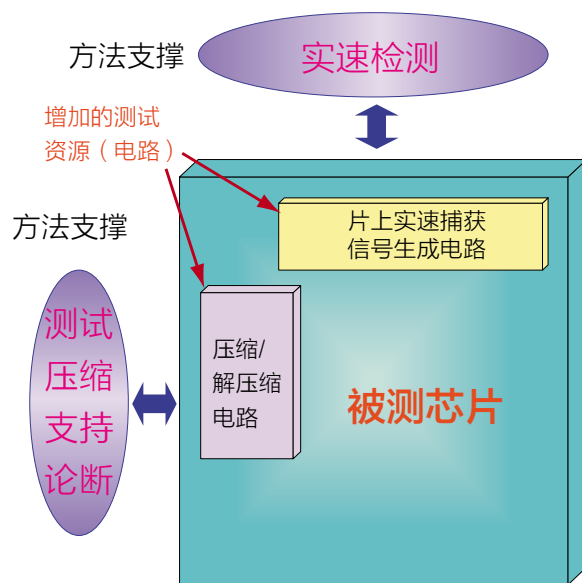
数字电路实速检测和故障诊断技术及其应用

“数字电路实速检测和故障诊断技术及其应用”获 2008 年度北京市科学技术奖三等奖。主要完成人：李晓维、李华伟、韩银和、胡瑜、徐勇军、闵应骅。

该成果属于集成电路技术、信息技术领域，适用于超大规模集成 [VLSI] 电路和系统芯片 [SOC] 的实速检测与故障诊断，尤其是在低频测试仪上检测高频数字电路。

主要技术内容有：提出的实速检测片上捕获信号生成方法，用于交流扫描测试的快速信号产生电路可重配置，具有电路结构简单、延迟小、且和待测电路无关的特性，可实速捕获测试响应，满足了使用低速测试仪检测高频数字电路的需求。提出的集成电路全速电流测试方法，通过测量电路变化时消耗的平均电流来检测 IDDQ 和 IDDT 可测故障，可检测那些用电压测试方法无法检测的故障。提出的支持故障诊断的测试响应压缩技术，利用卷积变换提供空间和时间维度上

的测试压缩，单输出确保了最大压缩率；针对响应中错误位分布特点，能够定位 3 位及以下错误位信息。提出的测试流程优化技术，设计了测试项目划分及快速测试流程调度方法，而不依赖耗



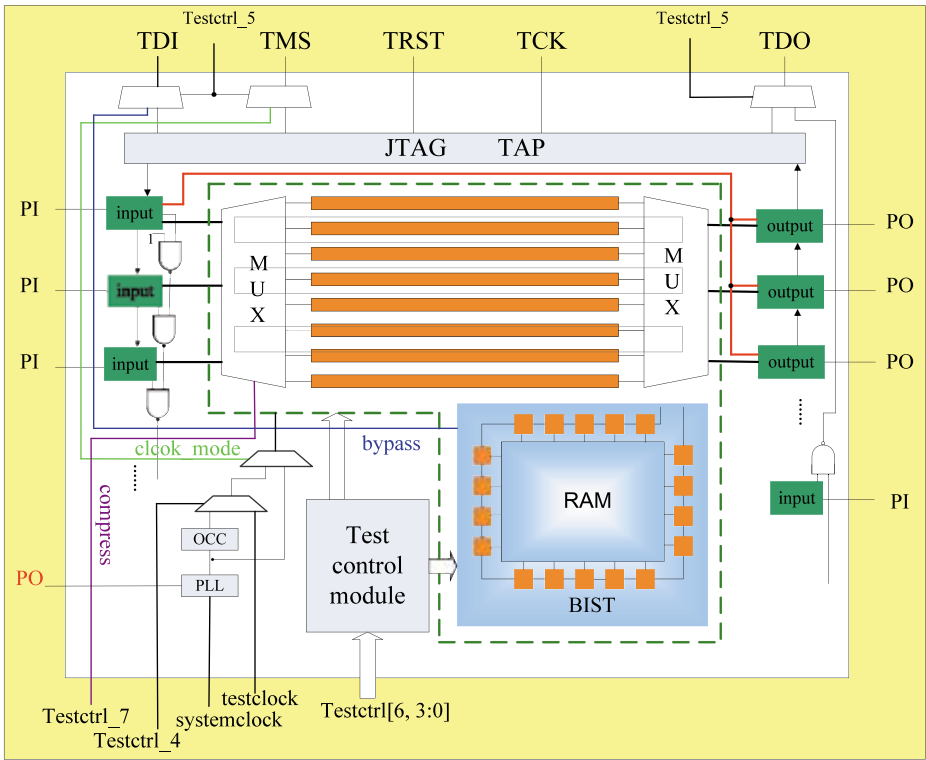
数字电路实速检测和故障诊断技术

时的动态规划方法；同时，测试流程中集成了自诊断和诊断优化技术，诊断电路可完成自身诊断，保证了诊断的精确度。

基于以上核心发明点，设计了面向国产 50MHz 数字测试设备的 VCD-VCT 全流程测试程序开发软件。在数字集成电路测试设备、高性能处理器芯片验证测试、低功耗微处理器芯片批量测试、以及与测试相关的 EDA 领域得到了应用。使用本专利技术在 400MHz 以下的测试

仪器上，为 1GHz 以上的高频数字电路提供了高故障覆盖率实速检测技术，打破了国外厂商的技术垄断和西方国家对我国测试设备的出口限制。

本研究成果已获国家发明专利 10 项、已受理国家发明专利 13 项，计算机软件著作权 17 项，发表学术论文 30 篇；推动了数字电路实速检测技术的发展以及故障诊断技术在芯片设计和验证测试中的应用。



实速检测和测试压缩技术在龙芯 2F 可测性设计中的应用

获奖成果

自修复处理器的 3S 新原理、新方法与新结构

该项目获得了 2009 年度中国质量协会质量技术奖可靠性管理优秀项目奖，主要完成人：李晓维、胡瑜、吕涛、张磊、李华伟、韩银和等。中国质量协会质量技术奖由中国质量协会于 2005 年设立，经科技部国家科学技术奖励工作办公室批准，其目的是表彰在质量技术领域做出突出贡献的单位或个人，促进我国质量技术进步，提高我国整体质量管理水平、创新能力和竞争能力。

在国家重点基础研究发展计划（973）课题“高性能处理芯片的可靠性设计”、国家自然科学基金课题“嵌入式存储器内建自修复方法研究”和“微处理器自修复设计基础技术研究”资助下，课题组取得了一系列研究成果：提出了自修复处理器的自测试、自诊断和自修复（简称 3S）新原理、新方法与新结构，在领域著名国际学术会议 ITC、DATE、VTS 和学术期刊 IEEE Transactions on VLSI Systems 等发表学术

论文 9 篇；突破了一批关键技术，申请国家发明专利 6 项；开发了多个自主版权软件工具，申请软件登记 2 项；研制的自修复处理器芯片已用于无线传感网络（WSN）节点。主要成果包括：

（一）基于内容可寻址技术的单核处理器自修复方法。利用内容可寻址存储器细化冗余资源粒度，提高修复率和冗余资源利用率，简化了故障诊断和冗余分配算法。自修复方法兼具灵活性与高效率，在获得相同修复效率时，冗余资源和控制电路面积开销显著低于国际上已有自修复方法。

（二）基于拓扑重构技术的“N+M”多核处理器自修复方法。该方法为 N 核处理器芯片提供 M 个冗余核，如图 1 所示，利用拓扑重构屏蔽底层物理拓扑的多样性，通过求解二次分配 NP 完全问题，为操作系统和程序员提供优化的统一虚拟拓扑，从而提高芯片的成品率。

（三）自修复处理器在 WSN 节点中的应用。

基于 3S（自测试、自诊断、自修复）原理和方法，已成功研制出一款单核自修复处理器 SRP。SRP 处理器兼容 AVR RISC 指令集，能够容忍 20 个位和 10 个字的固定型故障，已应用于宁波中科集成电路设计中心研制的 WSN 节点控制器，如图 2(a) 所示；图 2(b) 显示了 SRP 节点故障

注入、自测试、自诊断与自修复过程。随着集成电路制造工艺的发展，晶体管的失效将成为常态，自修复微处理器的 3S 新原理、新方法与新结构，能够有效地提高芯片成品率与可靠性，是实现低成本信息化的重要技术手段，将在消费电子和安全关键等领域发挥重要作用。

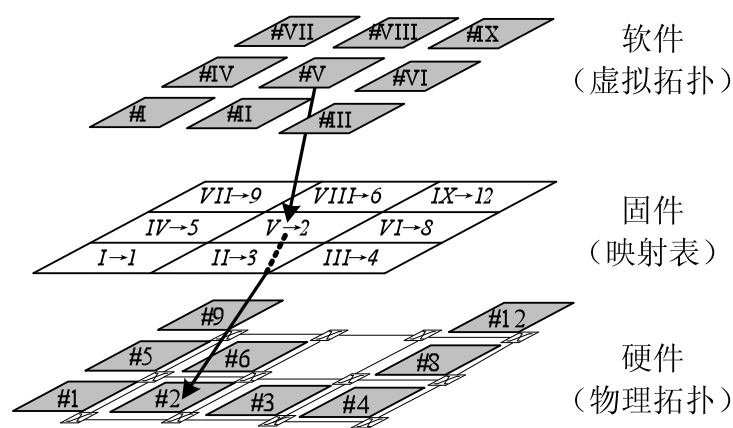
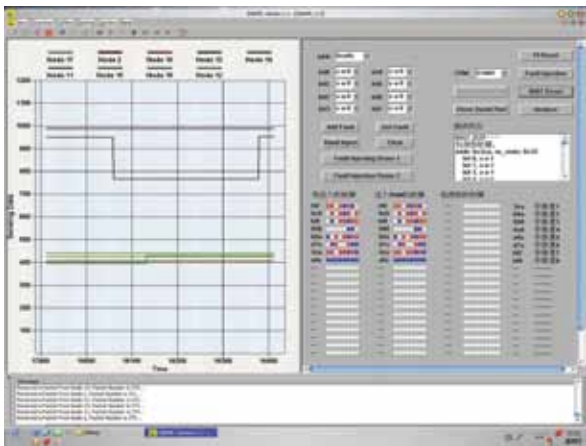


图 1: 基于拓扑重构技术的“N+M”多核处理器自修复方法



(a). SRP 节点



(b). SRP 节点的故障注入、自测试、自诊断与自修复过程

图 2 : SRP 处理器在 WSN 节点中的应用

鉴定 / 验收的代表性成果

曙光 5000A 高效能计算机

“曙光 5000A 高效能计算机”（课题编号为 2006AA01A102，以下简称曙光 5000A）是国家 863 计划信息领域重大项目“高效能计算机及网格服务环境”的研究成果，于 2009 年 4 月 16 日通过了中国科学院组织的科技成果鉴定。曙光 5000A 系统峰值运算速度 233.5 万亿次，Linpack 值 180.6 万亿次，系统效率为 77.34%，MPI 延迟 1.6 μ s，系统峰值功耗 992 千瓦，在 2008 年 11 月发布的第 32 届 TOP500 排行榜上位列第十，是当时除美国以外世界上最快的高性能计算机系统。

曙光 5000A 采用了基于多核 CPU 的星群体系结构，能够兼容运行大量的商业应用；采用了低功耗 CPU 和高密度刀片设计，在降低系统功耗的同时提高了计算的密度、节省了空间；刀片机箱内置高性能的 IB 交换模块，提升了系统效率，增强了系统的可靠性并降低了成本；封闭式水冷机柜的设计极大地提高了系统的冷却效率，降低了环境噪音。这些技术的综合运用，使得曙光

5000A 在实现了高性能的同时，还实现了高密度、高可靠、低功耗、低成本、易管理等目标。

鉴定委员会一致认为：曙光 5000A 兼顾了提高技术和产业化水平，是对高效能计算机快速可持续发展道路的成功探索，总体上处于当前国际



曙光 5000A 高效能计算机系统

先进技术水平，其中性能功耗比、性能密度比在基于通用 CPU 的百万亿次系统中具有世界领先水平，为研制低成本、低功耗、高密度、高效能的千万亿次计算机奠定了技术基础。

曙光 5000A 于 2009 年 5 月中旬运抵用户单位上海超级计算中心并被命名为“魔方”。经过几

个月的紧张安装、调试和测试，系统于 10 月初正式对外开放运行，并一举成为了世界上最大的公用计算服务平台。曙光 5000A 的部署极大地提高了上海超算的计算能力，为广大的科研、教育、商业、工业用户提供了强有力的计算工具，为科技创新提供了重要的基础设施保障。



TOP500 证书

鉴定 / 验收的代表性成果

语义资源空间的模型、理论、方法和工具

本成果由国家 973 计划“语义网格的基础理论、模型与方法研究”项目课题“语义资源空间的模型、理论、方法和工具”（课题编号为 2003CB317001）资助完成，2009 年 9 月通过验收。

研究面向未来互联环境，围绕规范组织、语义互联和智能聚融这三个科学问题研究大规模分布式异构动态网络互联环境下新型资源（信息、知识和服务）空间模型及其管理方法。

分类和关系是人类管理资源的基本方法。从这一观点出发，提出了资源空间模型、语义链网络模型、知识流模型以及大规模分布式网络上资源发布和共享方法。形成了系统的理论、模型和方法。

针对异构资源描述、组织和查询问题，资源空间模型利用分类语义对异构资源进行抽象和组织，建立多维分类语义信息描述模型以及高效多维分类查询索引，支持基于分类语义信息的异构资源查询。分类语义信息是描述异构信息的基础，

具有良好的语义适应性和可扩展性。资源空间模型可以应用于具有多种不同语义结构的领域，如软件设计，本体知识表达，多维数据管理，Web 多侧面信息管理等等。提供这样一个统一、简洁的语义信息描述模型可以减少大规模信息系统的信息异构性和不兼容性，提高信息的共享效率和效果。资源空间模型理论涉及资源空间的范式、完整性、表达能力、查询语言、查询效率、操作完备性、高效存贮机制、自组织对等资源空间模型、概率资源空间模型等。模型在机械故障信息管理、欧盟第六框架 GREDIA 项目和加拿大科学院电子健康等项目得到应用。国外专家公开发表评价认为资源空间模型是网络资源管理的一项突破。

语义链网络模型 (Semantic Link Network) 分为三个层次。第一个层次 SLN1.0 是对超链网的扩展，采用了简洁、可扩展的图模型来描述资源间的关系信息。通过引入关系规则，语义链网络可以描述常用的关系结构并支持关系推理和查询。与其它模型不同，语义链网络的特色在于其

动态性、演化性和可推理性。第二个层次 SLN2.0 是一个可扩展、自组织的数据模型。第三个层次 SLN3.0 是跨空间的信息物理社会语义链网络。语义链网络的语义是动态涌现的,而且操作可影响全局。提出了拓扑中心性、语义社区发现方法以及基于多样性和简单性的语义涌现规律。研究成果得到国内外学者在网页预取等方面的应用。

资源空间模型和语义链网络模型相结合,可以提供更为灵活的资源描述、管理和查询的统一模型。在资源空间模型和语义链网络的基础上,课题研究了一系列大规模分布式网络基础平台模

型和算法,以支持大规模、分布式网络环境下资源的查询和管理。探索了分散式动态知识共享的方法。最后,课题研究提出了未来信息物理社会和交互语义等前瞻模型和框架。

研究成果出版英文专著 2 部。在 Artificial Intelligence, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Parallel and Distributed Systems, ACM Transactions on Internet Technology, Communications of the ACM, Computer, Journal of American Society for

Information Science and Technology 等著名国际刊物发表论文 100 余篇。在国际会议做大会特邀报告 10 次。受邀请与 Tim Berners-Lee 等专家一同在著名国际学术刊物上发表邀请论文。课题验收专家组认为:“研究成果得到 Web 发明人 Tim Berners-Lee、语义网倡导者 J. Hender 和网格先驱 I.Foster 等著名专家, Science 和 IEEE Transactions on Knowledge and Data Engineering 等著名刊物,及 WWW 等著名国际会议的引用和评论 1494 次,得到多个国家和领域的采纳和应用,取得了好的效果。研究成果达到国际领先水平。”



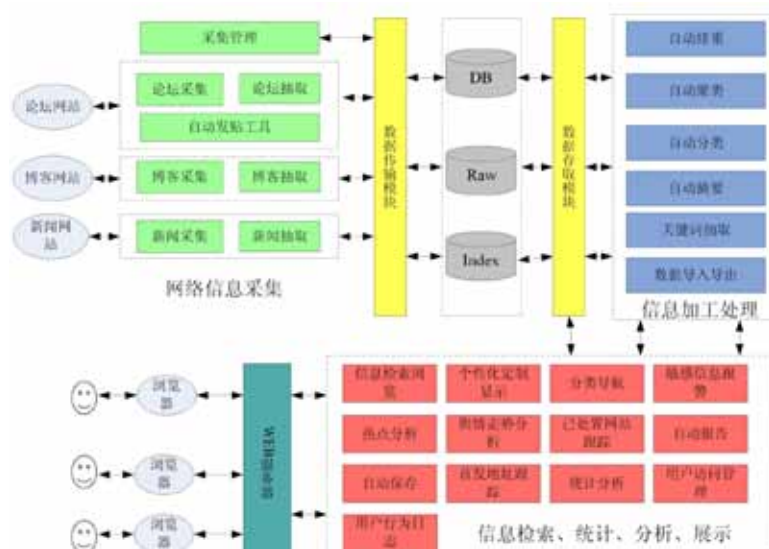
鉴定 / 验收的代表性成果

“天玑”网络舆情监测系统

天玑网络舆情监测系统由 973 课题“大规模文本内容计算”（课题编号为 2004CB318109，2009 年 9 月通过验收）、863 课题“网络文本的倾向性分析”（课题编号为 2007AA01Z441，2009 年 12 月结题）和“网络话题的发现、传播及动态演化特征分析技术”（课题编号为 2007AA01Z438，2009 年 12 月结题）资助完成，融专业搜索、信息抽取、文本分析和智能

数据挖掘等技术于一体，能够从 BBS 论坛、新闻网站、博客等多种信息通道中实时采集和提取舆情信息，跟踪舆情动向，监测事态的宏观分布与变化趋势，为政府部门、企业乃至个人用户提供热点专题监控、个性化信息推送、综合智能搜索等服务，并可以图文并茂的形式提供客观、完整、准确的舆情监测报告。

天玑网络舆情监测系统具有强大的舆情信息



天玑网络舆情监测系统总体架构

获取能力：能够覆盖与舆情相关的多个主流舆情通道；提供对大规模信息源的高性能采集；具有强大的网页获取和解析能力；采用高效、精确、配置方便的信息抽取模型，可有效提取页面中的舆情信息。



天玑网络舆情监测系统界面

系统还具有高效的舆情分析与处理能力：支持数据增量更新的高性能实时索引检索功能；高效的文章排重和转载识别算法；基于自然语言处理的关键词提取和自动文摘技术；对特定话题信息实现自动搜集和全方位跟踪、溯源，提供话题的时间、空间分布及趋势分析；根据网络信息的关注情况自动发现热点舆情话题，并对其进行分析统计和跟踪；提供被监测信息源的一站式、全方位的监控和浏览。

系统架构具有良好的可扩展性，可根据需求灵活定制，已在政府、金融、教育等行业用户取得了良好的应用效果。

天玑网络舆情系统能够为政府相关管理部门掌控网络舆情态势提供支撑，提升决策管理能力和处置突发事件的能力，对建立和谐社会、保障国家信息安全具有重大的社会效益。该成果还可以辐射到社会经济的多个领域，为企业、行业等提供舆情技术和服

务，具有极大的社会经济价值。天玑系统已经广泛应用于工信部、广电总局等关键部门，成为了舆情掌控的重要工具，为维护社会稳定、保障国家网络空间安全等发挥了实际作用。其中，部分系统已经稳定运行一年多，在多次重大政治任务中发挥了关键作用；在广电总局的应用也取得了良好的效果，满足了用户的需要，得到相关部门的好评；系统在中国证监会的应用，推动证监会的舆情监控工作取得了长足的进步，为维护资本市场的稳定，保护中小投资者的利益提供了有效的支持。并且，中国证监会建设的网络信息监控系统获得了 2009 年度证券期货业科学技术奖二等奖。



相关技术组件（ICTCLAS 分词组件、Firtex 检索平台、CoolFish P2P 平台）

鉴定 / 验收的代表性成果

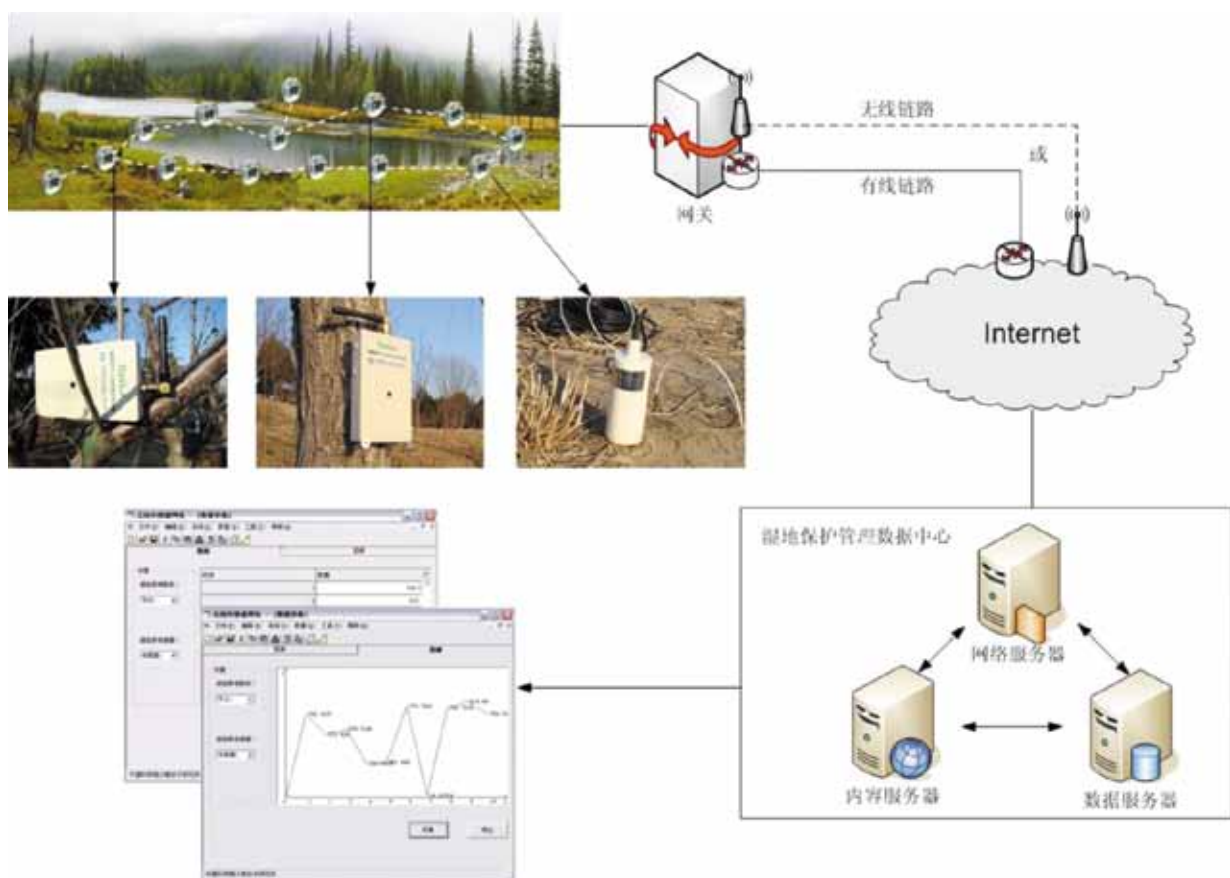
传感器网络

在 863 计划目标导向类课题“无线传感器网络的节点技术”（课题编号为 2006AA01Z215）的资助下，计算所成功开发了面向环境保护应用的传感器网络系统，包括具有数据采集功能的传感节点、用于网络传输的汇聚节点、用于网络协议转换的网关节点以及功能丰富的系统软件等。该课题于 2009 年 12 月通过验收。环境监测是无线传感网络领域一项典型的实际应用，通过具有传感、处理、计算、存储和无线通信多种能力的智能节点来完成复杂的以应用目标为需求导向的环境监测工作。通过节点的分布式协作进行自适应的组网和通信，还能实现全自动的数据统计采样、数据处理、数据融合、查询式监控、远程动态功能升级以及不同优先等级报警等先进的监控措施，从而形成一套完整的面向环境保护监控系统的传感器网络技术解决方案与实施系统。该基于传感器网络技术的环境监测系统已经达到实用

化、产业化的阶段。

该课题成果已应用于故宫博物馆文物保护环境监测，博物院内的管理人员可以对特定环境参数进行远程实时访问。系统安装布设及维护具有方便灵活，不破坏古建筑，不影响文物展示，扩展性强等特点。课题成果还应用于无锡水资源及环境保护物联网应用示范工程。传感网环境监测技术在监测规模、监测地域、监测准确度、监测灵活程度等方面都有传统监控手段难以比拟的优势。本系统解决方案也可以扩展到大多数需要进行多点网络化数据采集的系统，对军事、工业和科学研究等诸多重要领域都有着深远的影响。

随着当前物联网技术的发展，作为物联网技术的重要组成部分，该基于传感器网络技术的环境监测系统将会有更广阔的发展空间，创造更多的经济效益。



鉴定 / 验收的代表性成果

支持大型公共设施安全问题的虚拟现实技术与系统

“支持大型公共设施安全问题研究的虚拟现实技术与系统”由十一五 863 目标导向类课题“支持大型公共设施安全问题研究的虚拟现实系统研发”(2006AA01Z336)和国家科技支撑计划“世博科技专项”课题“2010 年上海世博会人流疏导应急预案模拟关键技术研究与服务应用”(2009BAK43B38)资助完成。共发表论文 24 篇, 申请专利 12 项(其中 7 项已经获得授权), 登记软件著作权 9 项。

技术特色在于: 首先, 提出了一种基于连续模型的高密度群体模拟技术, 能够模拟客流高峰期, 大规模、高密度虚拟群体的运动行为, 模拟结果接近实测数据, 误差小于国外主流商用软件; 其次, 定义了一种语义模型, 能够描述多层、室内外结合、包含各种公共设施的复杂场景, 使得虚拟群体能够有

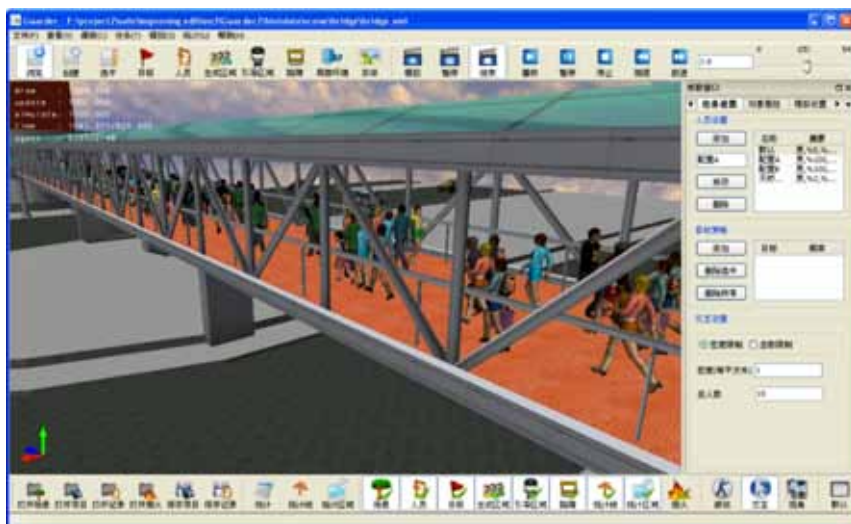
效地在复杂场景(城市地铁站、火车站、写字楼群)中完成不同的任务目标; 最后, 结合高性能计算技术, 提出了一种超大规模(5 万-100 万)群体运动的并行计算方法, 能够有效地实现 5 平方公里范围内 100 万规模群体运动的并行模拟。此外, 系统还支持复杂三维场景的快速建模, 火灾、烟雾等自然灾害的模拟。

该成果可以广泛应用于人员疏散预案的制定、分析与评估, 安全人员的培训、任务布置, 大型活动人流规划方案的制定、分析与评估, 公共设施建筑设计的安全性能分析, 公共安全事故的事后分析等方面。

该成果已经在多个重要场所与活动的安全分析与预案演练中实际应用, 包括: 2010 年上海世博会高峰期客流压力测试与人群疏散模拟、2008 年北京奥运会主体育场人群疏散

策略分析、北京西客站高峰期人流规划与疏导、北京地铁奥林匹克公园站高峰客流等问题研究；为上述场所的人群安全疏导与管理提

供了有力的技术支撑和科技保障，获得了应用部门的一致好评，产生了一系列良好的社会效益。



Guarder 系统界面



模拟某地铁站人群疏散

鉴定 / 验收的代表性成果

数字内容集成分发平台

本成果是在“十一五”国家科技支撑计划重大项目现代服务业共性技术支撑体系与应用示范工程课题“数字音像内容集成分发平台”(课题编号为2006BAH02A11)的资助下完成的,于2009年5月通过了验收。课题组面向数字音像服务行业的新需求,历经近三年的研发,对数字音像网络发行产业中的关键技术给出了具有自主知识产权的技术解决方案,开发了易于交互与交换、具有版权保护功能的数字音像内容集成与分发平台,实现了高质高效、开放集成和资源共享的数字音像内容集成、分发和交换服务,并开展了积极有效的示范应用。

(1) 数字音像内容集成分发平台是面向多家内容运营商、内容提供商、终端制造商以及消费者的统一集成分发平台,由数字音像内容集成管理系统、数字音像内容互联和交换系统、数字音像内容网络分发系统、数字音像内容版权保护系统、网络发行智能终端系统等五个有机联系、协同工作的系统架构而成,实现基于IP网络的数字音像内容存储和传播,为终端消

费者提供便捷高效的数字音像内容服务,并应用数字版权管理体系保障各方的合法权益。

(2) 课题针对数字音像内容网络出版发行中的若干关键技术,如内容快速网络分发、内容互联和交换、海量内容集成管理、内容网络版权保护与双向认证授权、智能终端的构件化定制等技术进行深入研究和实现,形成多项具有创新性和自主知识产权的技术成果。

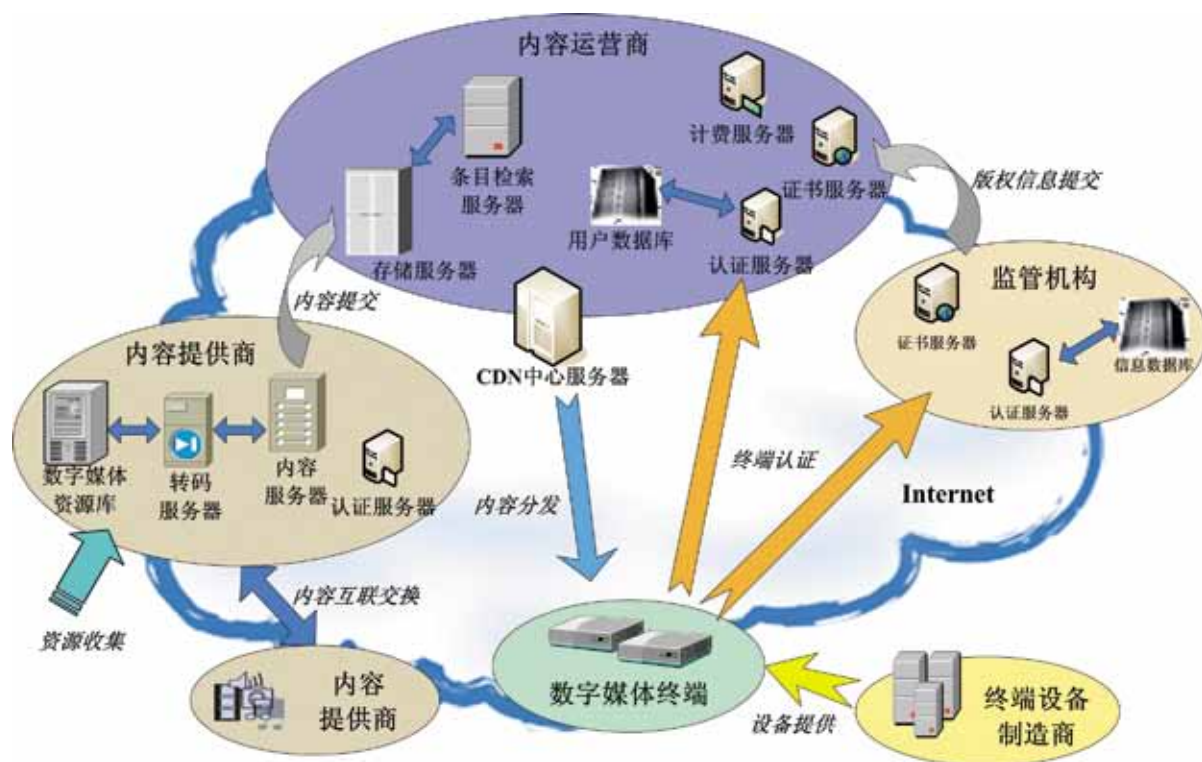
(3) 以数字音像内容集成分发平台为基础,课题积极探索新型数字音像内容出版发行的商业模式,并在相关行业机构和多家企业进行试应用,示范单位对平台系统研发产品和服务质量均表示满意。通过课题成果的积极示范应用,平台可用性、可靠性、稳定提供服务的能力得到有效检验,数字媒体内容的网络出版发行商业模式得到初步实施和验证,用户体验良好。

数字音像内容集成分发平台可应用于数字媒体网络出版和发行行业,实现数字音像内容的集成与分发服务。该平台融合产业链中内容提供商、内容运营商、网络运营商、监管机构

和消费者等多种角色，提供统一开放、标准化的、可管可控的、资源共享和联合认证的综合平台，并为广大互联网用户提供高质、高效的正版数字媒体下载和在线视听服务，有效保护正版音视频资源的版权信息，解决产业链中各环节的诸多问题。通过制定数字音像内容网络发行智能终端接入接口规范、第三方监管接口规范等

标准化接口规范，使更多的内容运营商、终端提供商可以基于此平台完成数字音像内容可管可控的发行、传播和使用，形成开放互联的平台环境。

本成果已获国家发明专利 2 项、已受理国家发明专利 18 项，获得计算机软件著作权 15 项，出版专著 1 部，发表学术论文 42 篇。





科研基地建设

信息内容安全技术国家工程实验室建设进展

1、工程实验室建设定位

信息内容安全技术国家工程实验室主要任务是围绕国家网络信息内容安全的重大需求，以基于网络的信息检测分析与安全管理为重点，开展信息内容安全基础理论和网络信息获取、分析、挖掘、安全对抗等核心技术研究，进行大规模模拟试验和真实环境下的持续验证，通过对信息内容安全理论与技术的研究，形成一系列具有自主知识产权的信息内容安全基础理论与关键技术，培养相关人才；承担国家级网络信息内容安全重大课题；同时积极进行产学研的合作与交流，为信息内容安全产业界提供新技术的辐射。

2、工程实验室建设内容

工程实验室实行理事会领导下的实验室主任负责制，首届理事长单位由中国科学院计算技术研究所担任，其它理事单位有：国家计算机网络与信息安全管理中心、哈尔滨工业大学、北京邮

电大学、深圳市任子行网络技术有限公司、中国科学院计划财务局、中国科学院高技术研究与发展局。李国杰院士任理事长。实验室主任为方滨兴院士，常务副主任为郭莉正研级高工。工程实验室的主要建设内容有：开展信息内容安全基础理论与关键技术的研究；建设网络舆情分析与预警研究验证平台、海量网络音视频信息研究验证平台、互联网真实网络流实验测试平台以及互联网模拟和仿真平台等支撑平台；进行研发办公区与会议培训区的工程改造。

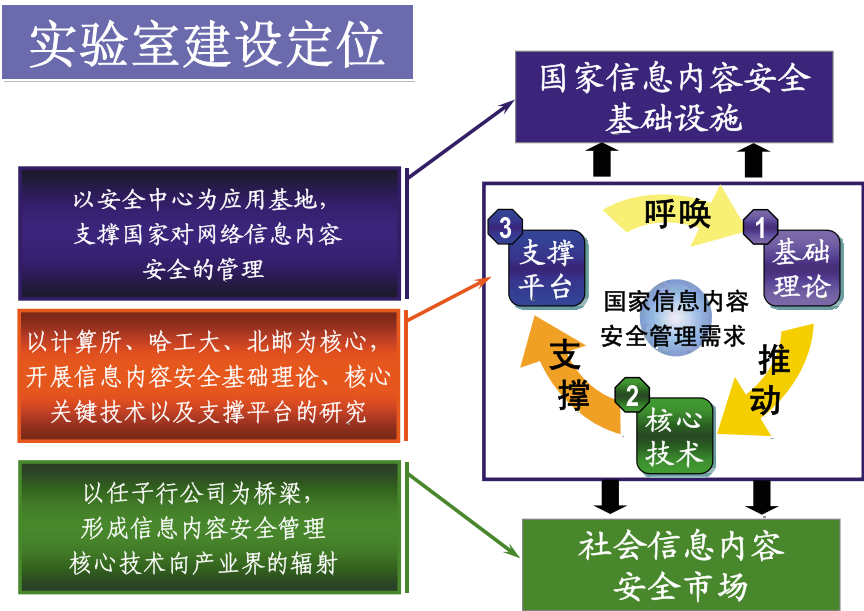
3、工程实验室建设进展

2009 年是工程实验室建设期的第一年，重点开展了基础环境投资建设。已经完成大部分的基建改造，包括计算所新楼十层机房改造，以及办公区改造等；四个平台建设完成过半，其中所内建设的三个平台（网络舆情分析与预警研究验证平台、海量网络音视频信息研究验证平台、互

联网模拟和仿真平台）已完成设备购置 1073 万，占总计划的 46%；国家计算机网络与信息安全管理中心的互联网真实网络流实验测试平台已完成设备购置 1078 万，占总计划的 42%。同时工程实验室在学科建设方面也取得了一些成果：有 29

份相关专利被受理申请，软件著作权登记 14 份，发表相关论文 87 篇，有 3 个项目获奖。

工程实验室在基础设施与学科建设、高技术研究产出方面都取得了较好成绩，建设进展顺利。



信息内容安全技术国家工程实验室建设定位

学术活动

2009 年, 我所学术交流活动依然很频繁。全年共有 149 人次参加国际性学术会议, 共举办了 11 次所级学术报告和 2 次长期出访人员归国学术报告。此外, 还举办了第五届语义、知识与网格国际会议和 Hadoop 2009 技术沙龙等。

第五届语义、知识与网格国际学术会议 SKG (International Conference on Semantics, Knowledge and Grid, <http://www.culturegrid.net/SKG2009/>) 于 2009 年 10 月 12 至 14 日在珠海度假村酒店隆重举行。来自十多个国家和地区的一百多位研究人员开展了卓有成效的学术交流。

SKG 系列国际会议是在中国科学院和国家重大基础研究计划 (973 计划) 项目“语义网格的基础理论、模型与方法”支持下发起的第一个跨越三个研究领域 (Semantics, Knowledge and Future interconnection environment) 的国际学术会议, 目前已成功举办 5 届。SKG2009 的论文集由 IEEE 出版, 往届会议论文集均被 EI 收录, 会后将选择高质量论文在国际学术刊物上出版专集。

Hadoop 是 Apache 组织下的一个活跃的开

源项目, 提供了一套名为 HDFS 的分布式文件系统以及支持 Map-Reduce 的计算框架, 得到了产业界和研究界的共同关注。Hadoop 技术现已在大批互联网公司得到成功应用, 并且国内外研究界在其上进行了广泛的研究。

为了更好的发扬开源精神, 扩大 Hadoop 开源项目在中国的影响, 在科研院所和企业界推广 Hadoop 技术, 促进 Hadoop 技术爱好者之间的交流, 在成功举办前两次 Hadoop 技术沙龙的基础上, 于 2009 年 11 月 15 日在中国科学院计算技术研究所召开 Hadoop in China 2009 大会。会议主办方为计算所。大会邀请了来自 Yahoo!、百度、Facebook、中移动研究院、Google、Cloudera、EMC、VMWare 以及计算所等的专家做报告并进行讨论。大会免费注册人员约 300 人, 其中外籍人士 10 人。

2009 年 12 月 4 日到 5 日, 我所主办了主题为“面向多核时代的编译技术”的研讨会。来自美国明尼苏达大学的 Pen-Chung Yew 教授、Intel 公司的 Tin-Fook Ngai 研究员、中国科学院计算技术研究所的乔如良研究员围绕多核时代的

编译技术发展分别做了主题报告。会议期间，来自中国科学院计算技术研究所、清华大学、北京大学、复旦大学、中国科学技术大学、北京理工大学、北京科技大学、武汉大学、上海交通大学、

国防科技大学、江南计算研究所等单位的同志分别进行了报告和交流，有约二十家单位参加了本次研讨会，几乎囊括了国内所有建有编译研究队伍的单位。



丹麦 Aalborg 大学 C. Jensen 教授在 SKG2009 会议上做大会特邀报告



新加坡国立大学 Beng Chin OOI 教授在 SKG2009 会议上做大会特邀报告

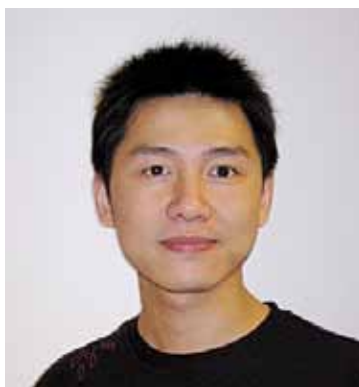


中科院计算所诸葛海研究员在 SKG2009 会议上做专题讲座

学位论文

2009 年计算所共有 145 人取得硕士学位，其中优秀硕士学位论文 69 篇；有 93 人取得博士学位，其中优秀博士学位论文 58 篇。2002 级硕博连读生王嘉平所完成的题为《基于数据驱动的表面质感建模与绘制》的博士学位论文，被评为 2009 年全国百篇优秀博士学位论文。2002 级博士研究生季向阳完成的题为《可伸缩视频编码研究》的博士学位论文，被评为 2009 年中国科学院优秀博士学位论文。

全国百篇优秀博士学位论文简介



王嘉平，男，1980 年 3 月出生，2004 年 7 月师从中国科学院计算技术研究所博士生导师沈向洋教授，于 2007 年 8 月获博士学位。

邮箱: e_boris2002@hotmail.com

中文摘要

真实世界中物体所呈现出丰富的外观取决于物体材质和光的复杂交互作用。作为真实感图形学的核心课题之一，表面质感建模和绘制重点研

究如何有效、真实地描述物体“材质” (Material)，即物体材料与光的交互作用，使得可以在计算机中绘制出真实的物体光照效果。同时，真实材质数据的测量、模型重建，以及如何使用材质模型进行绘制也是表面材质建模与绘制的重点研究问题。完备性、有效性和易用性是表面质感建模的三个目标。完备性是指模型的表达能力，包括能

够表达的光照和视觉效果种类,逼真程度,以及能够描述的真实世界材质的范围等。有效性主要考量模型的参数个数、状态空间的维度以及模型存储空间的复杂度。易用性是指模型的测量、重构以及绘制的难易程度,包括对真实材质进行测量的难易程度和所需条件,和对各种真实感绘制流水线的支持程度等。

虽然近年来表面质感建模与绘制的研究已有长足的发展,但是复杂表面质感的建模,特别是对具有复杂几何和反射属性的粗糙表面材质,半透明材质,以及时变材质的建模还面临着很大挑战。一方面,由于这些材质中材料与光交互作用的物理机制非常复杂,千变万化,甚至具体过程仍然是未知的,使得基于物理的仿真模型很难满足完备性和有效性的要求。另一方面,这些材质所引起的光影效果和光照方向、视线方向、以及入射和出射表面位置相关,构成了八维的双向表面散射反射分布函数(BSSRDF)或其高维简化表达,使得数据采集困难,模型数据量大,无法满足有效性和易用性的要求。

本论文针对这些挑战,在该领域的三个前沿课题:时变材质建模、粗糙表面材质建模以及半透明材质建模上开展了研究,分别提出了新的材质模型与建模方法,以及高效实用的测量方法。相关成果发表于计算机图形学最高水平的期刊ACM Transactions on Graphics以及图形学最高水平会议ACM SIGGRAPH上。论文的主要内容和创新点如下:

1. 基于“表观流形”(Appearance

Manifold)的时变材质模型

自然界大部分材质都会受外界环境影响而随时间发生变化。这些变化同时导致表面反射属性及其分布的变化。由于这一过程内在的物理化学机制非常复杂和多样,使得基于物理的建模方法很难具有通用性。同时,由于这一变化过程可能历时非常漫长,使得直接测量整个时变序列建模方法非常困难。本论文提出了时变材质的“表观流形”(Appearance Manifold)模型及其建模算法。表观流形模型揭示了时变材质在时间上变化是渐变的这一基本特性,从而将各种时变材质都表达为高维反射属性空间中的低维流形。在建模算法中,则利用时变材质同一时刻在空间上变化包含处于不同时间阶段样本这一重要特性,引入流形学习(Manifold Learning)的思想,构造高维空间中样本的紧邻拓扑关系。通过分析各个空间样本的相对测地距离以建立不同位置上的表面质感样本在变化趋势上的一致性关系,从而揭示各点上表面质感样本的相对时间关系。继而根据这个关系来推导时变表面质感的空间上和时间上的演化过程。为此论文提出了新的帧间一致性纹理合成技术(Frame-Coherent Texture Synthesis),成功从二维纹理样本在时间变化一致性的约束下合成三维的时变纹理序列,即具有高真实感的表面质感变化序列。

表观流形模型是图形学中第一个适用于不同时变材质的通用表观模型。表观流形的分析方法不依赖于时变过程的物理化学机制,而是直接对外观样本进行分析建立模型,具有良好的完备性。论文将模型成功应用于多种不同类型时变材

质,如植物枯黄,金属氧化,石头磨光等不同的时变材质等,获得了高度真实的效果。同时模型的建立过程避免了测量时间序列样本,而是通过流形学习技术从静态样本的空间分布中逆向求解时间序列,使得模型测量的时间不依赖于时变过程的速度,从而极大缩短了数据采集的时间,具有很好的易用性。该模型由于只需要一个时间点上的静态样本数据就可以预测时变表面质感的变化过程,使得正/逆模拟真实世界中物体的风化过程成为可能,在文物保护,文化遗产再现等方面也具有现实的应用价值。该项工作发表在2006年的国际图形学年会 (ACM SIGGRAPH) 上,收录于国际图形学领域最高水平的期刊 ACM Transactions on Graphics (ACM TOG, IF 4.02), 并已申请国际专利,受理并公开。论文发表后,已有来自 ACM TOG、IEEE TVCG 等高水平国际期刊论文引用 11 次。

2. 基于微面元合成 (Microfacet Synthesis) 的粗糙表面材质建模方法

六维的表面变化双向反射分布函数 (Spatially Varying Bidirectional Reflectance Distribution Function) 在图形学中被用于描述不透明粗糙表面上反射属性。自 1999 年,国际学者 (Dana et al) 在 ACM TOG 上首次发表双向反射分布函数的测量技术以来,国际图形学界大多采用多视点多光照的测量系统。这种建模和测量方式设备昂贵,数据量大,易用性和有效性较差。国际学者 (Gardner et al) 于 2003 年在 ACM TOG 上发表了易用的基于线性光源的测量系统,但只能采

集各向同性 (Isotropic) 的表面变化双向反射分布函数。本论文提出了一个基于微面元的样本合成算法用于建模表面变化双向反射分布函数。该方法采用基于微面元 (Microfacet) 的表面反射模型,利用表面各点反射属性的相关性和冗余性,通过单一视点下的四维反射数据样本来重构并合成表面在任意视点下的六维表面变化双向反射分布函数,并能够很好地处理各向异性 (Anisotropic) 的双向反射分布函数。论文进一步提出了基于空间剪枝、样本空间聚类 and 局部欧氏空间线性重建的加速算法,将合成算法复杂度从 $O(N^2R)$ 降至 $O(N \log(NR))$ 。基于微面元的模型将模型数据量从 $O(N^6)$ 降低为 $O(N^4)$, 提高了模型方法的有效性。基于微面元合成建模方法则可以从单一视点的采样数据建模并重构表面反射属性,从而避免了图像配准 (Image Registration) 操作,降低了测量高分辨率材质难度,充分利用成像系统的解析度,因而具有极佳的易用性。基于该建模算法,论文同时提出了一个低成本的测量系统用于获得高质量表面反射属性数据。该项工作发表在 ACM SIGGRAPH 2008, 收录于 ACM TOG, 并已申请国际专利,正在审批中。

3. 基于扩散方程的通用异质半透明材质模型

本论文提出基于扩散方程 (Diffusion Equation) 的通用次表面散射模型,用于异质半透明材质的建模、测量以及实时绘制。模型采用偏微分方程简洁有效地描述了异质半透明物体内部的次表面散射。基于该模型,论文提出了逆向求解建模算法 (Inverse Diffusion Equation), 实现

了图形学中首个异质半透明材质次表面散射属性测量与重构算法及其 GPU 并行实现,通过异质半透明物体外观的稀疏测量数据有效地恢复物体内部的材质属性与空间分布。基于该模型,论文同时提出了首个支持不同光照条件和实时材质编辑的次表面散射实时绘制算法。为了支持可形变异质半透明物体的实时绘制,本文提出了 Polygrid 体网格模型,该体网格模型具有一致 6 连接的拓扑结构并可以描述任意形状的物体。这一结构使扩散方程在各个结点上能够有一致的计算方式,便于并行算法的高效实现。本文提出的异质半透明材质模型优于国际学者 (Jensen et al) 于 2001 年在 ACM TOG 上提出的同质 (Homogeneous) 半透明材质 Dipole 模型,而可用于描述异质及同质半透明材质,具有更好的完备性。模型利用物体内部空间材质属性和分布通过扩散方程快速计算次表面散射,与基于表面采样的表达与绘制方式相比,数据量从 $O(N^4)$ 降低为 $O(N^3)$,具有良好的有效性。而在易用性上,论文所提出的建模方法突破了国际学者 (Peers et al) 于 2006 年发表于 ACM TOG 的基于密集光束测量系统,提出了基于稀疏光场的测量系统,大大减少了测量过程中的图像采集数量。同时绘制算法复杂度从 $O(N^4)$ 降低为 $O(N^3)$,并适于 GPU 并行实现。该项工作发表在 ACM TOG 上,并已申请国际专利,正在审批中。

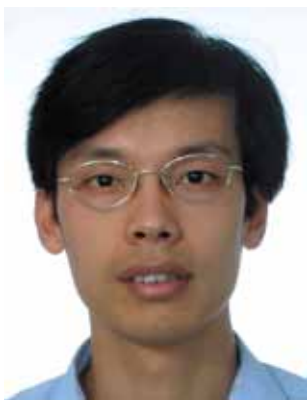
4. 准同质 (Quasi-homogeneous) 半透明材质模型

准同质半透明材质是一类由很多具有相似材

质属性的颗粒单元构成的材质,如面包,海绵等。准同质半透明既有复杂的几何细微结构 (粗糙表面) 又有复杂的次表面光线散射过程 (半透明),因而其材质建模也成为一难题。本文提出了一个准同质半透明材质的多尺度材质模型。该模型基于本文对准同质半透明材料次表面散射特性的一个重要发现:其在局部,由于材质的异构性,次表面散射是各向异性的。而整体上,光子在穿越有一定长度且均匀分布的异质材料时,其轨迹在统计意义上和穿越同质材料类似。据此本文建立了全局和局部分解的光线传输模型,有效的将材质上八维的 BSSRDF 光线传输函数表达为一系列低维的全局和局部传输模型,并提出了该模型相应的测量与重构算法。从而可以准确有效的描述准同质材质中全局光线传播,以及光线通过表面的细微结构传入、传出物体的情况。该模型不需要费时的次表面光线传输的物理仿真,就计算出射光强的分布,以实现快速绘制。实验证实模型可真实有效的模拟对准同质半透明材质中光传输过程,生成真实的绘制结果,具有良好的有效性。本模型是计算机图形学领域第一个可以从真实物体上测量并重构的 BSSRDF 材质模型。该项工作已发表在 ACM SIGGRAPH 2005,收录在同年的 ACM TOG 里,并已获国际专利 (US 7,312,797)。迄今论文已有来自 ACM TOG、IEEE TVCG, CGF 等高水平国际期刊上的论文引用 12 次。

关键词: 真实感绘制, 实时绘制, 双向纹理函数, 着色模型, 双向反射分布函数, 次表面散射, 时变材质, 扩散方程, 自然现象

中国科学院优秀博士学位 论文简介



季向阳，男，1976 年 9 月出生，2002 年 9 月师从中国科学院计算技术研究所赵德斌教授、高文教授，于 2008 年 7 月获博士学位。
邮箱: xiangyangji@hotmail.com

中文摘要

在基于互联网的新数字媒体环境下，越来越多的视频内容在互联网上进行传递。然而，Internet 由于其自身设计和构造的特点，并不具有服务质量（QoS）的保障，时延和丢包率等也是随时间动态变化的；此外，网络传输视频信息的内容与形式非常丰富，使用不同设备的不同用户对不同的视频内容也会有不同服务质量的要求。因此，面向互联网的视频服务给流化视频编码技术提出了新的问题。一个基本方法是使编码生成的码流具有高度可伸缩性，能够根据网络条件的变化和接收端的要求随时对码流进行调整。同时，这些可伸缩功能可以很好地提供内容的层式保护、多通道内容的制作与分发、灵活的视频会议、移

动商务、视频点播系统、多点的监控系统和端到端的无线传输等。

本文对可伸缩视频编码中时域、空域与质量可伸缩以及它们的组合编码进行了深入的研究，提出了高效、简单、鲁棒的可伸缩编码方法。主要包括：

1. 在时域可伸缩编码方面，研究了高效的 B 帧编码技术。提出了两种高效的双向预测模式，即对称预测模式与直接预测模式。这两项技术均已被 AVS 接受。第一项技术用于替代传统双向预测模式，此模式只需要一个运动矢量被编码，在假设运动近似匀速的条件下，另一个运动矢量可以由它推导出来。这种模式在预测精度与运动矢量信息编码所需比特数之间取得有效的平衡，从而可以获取更好的率失真性能。第二项技术是时域直接模式，它不但解决了 AVS 标准中 B 帧与 P 帧都只有两个参考帧缓冲区限制下如何正确得到

参考索引的问题,而且在保持高精度时域直接模式运动矢量的推导中避免了除法运算。

2. 提出了一种低复杂度的块级运动矢量精度自适应 B 帧编码方法,这种方法可以使 B 帧双向预测块自适应地选择运动矢量精度来降低运动补偿插值的复杂度。根据运动补偿线性插值的性质,对双向预测块前后向亮度预测信息的运动补偿插值进一步地合并,从而再一次降低基于子像素的双向预测块的运动补偿复杂度。此外,本文还为双向预测块提出了一种简单的色度运动补偿插值方法。实验结果表明,这种方法可以在与现有方法保持相当编码效率情况下,极大地降低 B 帧解码复杂度,甚至比 P 帧有更低的运动补偿复杂度。

3. 在质量可伸缩视频编码方面,提出了一种对多个质量增强层进行不同漏预测的质量精细可伸缩视频编码方法。首先假设基本层码流和部分增强层码流可用情况下,对基于两个漏因子的漏预测方法给出了误差传递理论分析。基于这种分析,提出了如何有效地产生用于增强层预测的信息。通过选择合适的漏因子,可以有效地将增强层的信息引入预测环内,从而使得这种方法在达到高效编码的同时又能有效地抑制漂移误差。进一步,为了降低解码复杂度,提出了变换域系数缩放方法来解决部分增强层的多次重构问题。另外,提出了一种编码优化方法来改善多 FGS 层编码的漂移误差问题。实验结果表明,与现有的质量精细可伸缩视频编码方法相比,所提出的方法可以有效地抑制漂移误差,因此极大地提高了编码效率。

4. 提出了应用于低延时传输的基于小波与漏

预测的空域-质量组合的可伸缩视频编码。现有的基于分层编码的方法支持空域-质量组合的可伸缩视频编码,但通常不能有效地去除不同空域分辨率层间的冗余,这往往使得高分辨率层的编码效率不够理想。在所提出的方法中,通过小波变换来实现空域分辨率可伸缩以保证消耗在低分辨率图像上的比特能够全部为高分辨率层使用。通过空-频域运动补偿相结合的漏预测方法来实现对编码效率与漂移误差的控制。另外,根据不同误差传递率的应用需求,给出了最优漏因子选择的理论指导。

5. 提出了一种基本层嵌入 H.264/AVC 编码的三维小波全方位可伸缩视频编码方案。三维小波视频编码通过开环运动补偿机制来实现其视频流的高度可伸缩性,因此往往很难在特定码率点达到最优的率失真性能,尤其是在低码率点,由于过多的比特数花费在运动信息编码上,不能产生令人满意的视频质量。一种有效的方法是对最低分辨率的基本层利用 H.264/AVC 编码来有效提高低码率点视频的质量。一方面,采用优化的基本层编码结构,增强层可以充分利用基本层的重构信息;另一方面,通过在时域小波变换中引入一种帧内模式编码,可以有效抑制“t+2D”方案中低分辨率视频解码的运动非对齐空域混叠效应。

6. 提出了一种用于 H.264/AVC SVC 中分级 B 帧编码时整帧丢失的错误隐藏算法。作为 H.264/AVC 可伸缩部分的扩展,可伸缩编码(H.264/AVC SVC)为生成的码流提供了灵活的时域、空域和质量的可伸缩性。但是,在不可靠网络中传

输时,如果出现信息包丢失,即使是可伸缩的视频码流,解码后的质量仍然会很差。在我们提出的算法中,通过利用相邻视频图像的时域运动关系,可以简单有效地推导丢失帧的运动信息,从而可以在正确接收的前向和后向参考图像上进行运动补偿来恢复丢失帧。实验结果表明,当发生信息包丢失时,与 H.264/AVC SVC 中使用的错误隐

藏算法相比,本文提出的错误隐藏算法可以明显地提高解码视频的主、客观质量。该算法也可以应用于带可分级 B 帧编码的 H.264/AVC 码流的错误隐藏,同时也可以扩展到三维小波编码中时域高频子带的错误隐藏。

关键词: 视频编码; 可伸缩编码; 中国音视频编码标准 (AVS); B 帧预测; 错误隐藏



2009 年 12 月 29 日,中国科学院 2009 年度优秀博士学位论文、院长奖、冠名奖学金及优秀导师奖颁奖会在中科院院部机关举行,图为会后两位获奖者与中国科学院常务副院长、中国科学院研究生院院长白春礼院士合影留念。(左起:季向阳,白春礼院士,王嘉平)

学术论文

2009 年发表论文 562 篇，其中期刊论文 178 篇，会议论文 384 篇。被 SCI 收录 85 篇，被 CPCI 收录 212 篇，被 EI 收录 466 篇，被中国科学引文数据库收录 210 篇。

第三届中国科学院计算技术研究所优秀论文奖和优秀著作奖共评出优秀论文奖一等奖九篇，优秀论文奖二等奖十九篇；优秀著作奖二等奖一部。

获奖作品第一作者名单如下：

优秀论文奖一等奖第一作者名单：

张鹏、胡明清、李如豹、胡伟武、邢晶、樊莉亚、徐扬、李建军、谢术富

优秀论文奖二等奖第一作者名单：

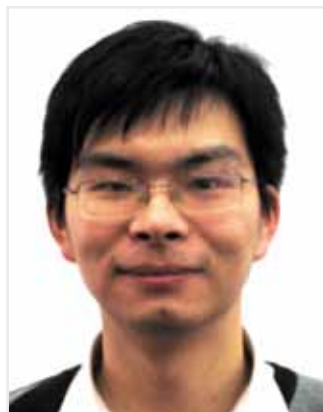
孟晓烜、郭嘉丰、沈华伟、张国清、王桂玲、陈云霄、王洋、李波、夏添、范东睿、虎嵩林、宋莹、谭光明、张磊、诸葛海、刘洋、郑伟、刘延伟、王达

优秀著作奖二等奖作者名单：张玉军

获 2009 年计算所优秀论文 一等奖论文摘要

1. Peng Zhang, Xingquan Zhu, Li Guo. Mining data streams with labeled and unlabeled training examples. In Proc. Ninth IEEE International Conference on Data Mining, Miami, USA, 2009, pp.627–636.

第一作者：张鹏，zhangpeng@ict.ac.cn



In this paper, we propose a framework to build prediction models from data streams which contain both labeled and unlabeled examples. We argue that due to the increasing data collection ability but limited resources for labeling, stream data collected at hand may only have a small number of labeled examples, whereas a large portion of data remain unlabeled but can be beneficial for learning. Unleashing the full potential of the unlabeled instances

for stream data mining is, however, a significant challenge, consider that even fully labeled data streams may suffer from the concept drifting, and inappropriate uses of the unlabeled samples may only make the problem even worse. To build prediction models, we first categorize the stream data into four different categories, each of which corresponds to the situation where concept drifting may or may not exist in the labeled and unlabeled data. After that, we propose

a relational k-means based transfer semi-supervised SVM learning framework (RK-TS3VM), which intends to leverage labeled and unlabeled samples to build prediction models. Experimental results and

comparisons on both synthetic and real-world data streams demonstrate that the proposed framework is able to help build prediction models more accurate than other simple approaches can offer.

2. Mingqing Hu, Yiqiang Chen, and James Tin-Yau Kwok. Building sparse multiple-kernel SVM classifiers. IEEE Transaction on Neural Networks; 2009, 20(5): 827–839.

第一作者: 胡明清, humingqing@ict.ac.cn



The support vector machines (SVMs) have been very successful in many machine learning problems. However, they can be slow during testing because of the possibly large number of support vectors obtained. Recently, Wu et al. (2005) proposed a sparse formulation that restricts the SVM to use a small number of expansion vectors. In this paper, we further extend this idea by integrating with techniques from multiple-kernel learning (MKL). The kernel function in this sparse SVM formulation no longer

needs to be fixed but can be automatically learned as a linear combination of kernels. Two formulations of such sparse multiple-kernel classifiers are proposed. The first one is based on a convex combination of the given base kernels, while the second one uses a convex combination of the so-called “equivalent” kernels. Empirically, the second formulation is particularly competitive. Experiments on a large number of toy and real-world data sets show that the resultant classifier is compact and

accurate, and can also be easily trained by simply alternating linear program and standard SVM solver. Index Terms—

Gradient projection, kernel methods, multiple-kernel learning (MKL), sparsity, support vector machine (SVM).

3. Rubao Lee, Zhiwei Xu. Exploiting stream request locality to improve query throughput of a data integration system. *IEEE Transactions on Computers*, 58(10): 1356–1368.

第一作者：李如豹, lirubao@software.ict.ac.cn



This paper focuses on the problem of improving throughput of distributed query processing in an RDBMS-based data integration system. Although a buffer pool can be used in an RDBMS to cache disk pages in memory to reduce disk accesses, it cannot be used for data integration queries since its foundation, the memory-disk hierarchy, does not exist. The lack of a data sharing mechanism limits system throughput because unnecessary data requests increase burden on data sources and redundant resultant data transfers waste network bandwidth. To address the problem, we present a new

technique called request window, which can detect and exploit data sharing opportunities among concurrent queries. Request window exploits a new stream request locality which reflects common query interests among independent users in a short time period. The existence of such a locality makes it possible to collect a group of related data requests and process them as a batch by request window. Evaluation on a PostgreSQL-based data integration system shows that request window can significantly increase system throughput when running a distributed TPC-H workload.

4. Weiwu Hu, Jian Wang, Xiang Gao, Yunji Chen, Qi Liu, Guojie Li. GODSON-3: A scalable multicore risc processor with x86 emulation. IEEE Micro, 2009, 29(2):17-29.

第一作者: 胡伟武, hww@ict.ac.cn



The godson-3 microprocessor aims at high-throughput server applications, high-performance scientific computing, and high-end embedded applications. It offers a scalable network on chip, hardware support

for x86 emulation, and a reconfigurable architecture. The four-core godson-3 chip is fabricated with 65-nm cmos technology. Eight- and 16-core godson-3 chips are in development.

5. Jing Xing, Jin Xiong, Ninghui Sun, Jie Ma. Adaptive and scalable metadata management to support a trillion files. In Proc. SC2009.

第一作者: 邢晶, xingjing@ncic.ac.cn



Nowadays more and more applications require file systems to efficiently maintain million or more files. How to provide high access performance with such a huge number

offiles and such large directories is a big challenge for cluster file systems. Limited by static directory structures, existing file systems will be prohibitively inefficient for this use. To

address this problem, we present a scalable and adaptive metadata management system which aims to maintain a trillion files efficiently. Firstly, our system exploits an adaptive two-level directory partitioning based on extendible hashing to manage very large directories. Secondly, our system utilizes fine-grained parallel processing within a directory and greatly improves performance of file creation or deletion. Thirdly, our system uses multiple-layered metadata cache management which improves memory utilization on the servers. And finally, our system uses a dynamic load-

balance mechanism based on consistent hashing which enables our system to scale up and down easily.

Our performance results on 32 metadata servers show that our user-level prototype implementation can create more than 74 thousand files per second and can get more than 270 thousand files attributes per second in a single directory with 100 million files. Moreover, it delivers a peak throughput of more than 60 thousand file creates/second in a single directory with 1 billion files.

6. Liya Fan, Fa Zhang, Gongming Wang. A framework to refine particle clusters produced by EMAN. *Bioinformatics*, 2009, 25(12), pp.276-280.

第一作者: 樊莉亚, fanliya@ict.ac.cn



Motivation: EMAN is one of the most popular software packages for single particle reconstruction. But the particle clusters produced during its model refining stage are of low qualities. We attempt to refine

the particle clusters by more accurately determining orientations of particles, and thereby achieving higher resolutions of consequent 3D structures.

Results: A particle reclustering framework

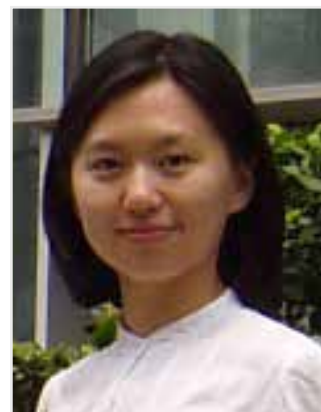
(PRF) is introduced, which consists of three components. Each of them is responsible for one of the basic tasks of PRF: normalization, threshold determination and reclustering. Our implementation is also described and proved to meet the constraints proposed by PRF. Experiments revealed that our implementation improved resolutions of

consequent structures for most cases, but only a little extra execution time was incurred. Therefore, it is practical to incorporate PRF in EMAN to improve qualities of generated 3D structures.

Availability and Implementation: Implementation of our algorithm is available upon request from the authors.

7. Yang Xu, Gareth Jones, Bin Wang. Query dependent pseudo relevance feedback based on wikipedia. In Proc. SIGIR09, Boston, USA, 2009.

第一作者: 徐扬, xuyang@ict.ac.cn



Pseudo-relevance feedback (PRF) via query-expansion has been proven to be effective in many information retrieval(IR) tasks. In most existing work, the top-ranked documents from an initial search are assumed to be relevant and used for PRF. One problem with this approach is that one or more of the top retrieved documents may be non-relevant, which can introduce noise into the feedback process. Besides, existing methods generally

do not take into account the significantly different types of queries that are often entered into an IR system. Intuitively, Wikipedia can be seen as a large, manually edited document collection which could be exploited to improve document retrieval effectiveness within PRF. It is not obvious how we might best utilize information from Wikipedia in PRF, and to date, the potential of Wikipedia for this task has been largely unexplored. In our

work, we present a systematic exploration of the utilization of Wikipedia in PRF for query dependent expansion. Specifically, we classify TREC topics into three categories based on Wikipedia: 1) entity queries, 2) ambiguous queries, and 3) broader queries. We propose and study the effectiveness of three methods for expansion term selection, each modeling the Wikipedia based pseudo-relevance information from a different perspective. We

incorporate the expansion terms into the original query and use language modeling IR to evaluate these methods. Experiments on four TREC test collections, including the large web collection GOV2, show that retrieval performance of each type of query can be improved. In addition, we demonstrate that the proposed method out-performs the baseline relevance model in terms of precision and robustness.

8. Jianjun Li, Chenggang Wu, Wei-Chung Hsu.
An evaluation of misaligned data access handling
mechanisms in dynamic binary translation systems. In
Proc. IEEE/ACM CGO 2009, pp.180–189.

第一作者: 李建军, lijianjun@ict.ac.cn



Binary translation (BT) has been an important approach to migrate application software across instruction set architectures (ISAs). Some architectures, such as X86, allow misaligned data accesses (MDAs), while most modern architectures have the alignment restriction that requires data to be aligned in memory on natural boundaries. In a binary translation system,

where the source ISA allows MDA and the target ISA does not, memory operations must be carefully translated to satisfy the alignment restriction. Naive translation will cause frequent misaligned data access traps to occur at runtime on the target machine, and severely slow down the migrated application. This paper evaluates different approaches in handling MDA in

binary translation systems. It also proposes a new mechanism to deal with MD As. Measurements based on SPECCPU2000

and CPU2006 benchmark show that the proposed approach can significantly outperform existing methods.

9. Shufu Xie, ShiguangShan, XilinChen, XinMeng, WenGao. Learned local gabor patterns for face representation and recognition. Signal Processing, 2009, 89(12): 23–33.

第一作者: 谢术富, sfxie@jdl.ac.cn



In this paper, we propose Learned Local Gabor Patterns(LLGP) for face representation and recognition. The proposed method is based on Gabor feature and the concept of texton, and defines the feature cliques which appear frequently in Gabor features as the basic patterns. Different from Local Binary Patterns(LBP) whose patterns are predefined, the local patterns in our approach are learned from the patch set, which is constructed by sampling patches from Gabor filtered face images. Thus, the patterns in our approach are face-specific and desirable for face perception tasks. Based on these learned patterns, each facial image is converted into multiple pattern maps and the

block-based histograms of these patterns are concatenated together to form the representation of the face image. In addition, we propose an effective weighting strategy to enhance the performances, which makes use of the discriminative powers of different facial parts as well as different patterns. The proposed approach is evaluated on two face databases: FERET and CAS-PEAL-R1. Extensive experimental results and comparisons with existing methods show the effectiveness of the LLGP representation method and the weighting strategy. Especially, heterogeneous testing results show that the LLGP codebook has very impressive generalizability for unseen data.

2009 年发表学术论文 检索统计报告

本统计报告的国际论文数据取自国际上颇具影响的检索工具:《科学引文检索》(Science Citation Index, SCI)、《工程索引》(The Engineering Index, EI)、《科技会议录索引》(Index to Scientific and Technical Proceeding, ISTP)。本统计报告的国内论文数据取自《中国科学引文数据库》(Chinese Science Citation Database, CSCD)。2008 年 ISI Web of Knowledge 平台进行了升级, ISTP 更名为 CPCI (Conference Proceedings Citation Indexes), 在 ISI Web of Science (WOS) 数据库可以同时检索到 SCI 和 CPCI 的数据及相互引用关系。因此, 本报告也将给出 WOS (SCI 和 ISTP) 收录引用的统计结果。

本统计报告的检索日期是 2010 年 9 月。

基本情况

根据统计, 2009 年计算所发表论文总计被 SCI 收录 85 篇, 被 CPCI 收录 212 篇, 被 EI 收录 466 篇, 被中国科学引文数据库收录 210 篇。当年被中国科学引文数据库引用篇数为 12 篇, 总被引用次数为 15 次; 当年 WOS 总被引用数为 59 次, 其中被引用数最高的是

11 次, 这篇文章的作者是 Zhuge, H (Zhuge, Hai), 文章标题: Communities and Emerging Semantics in Semantic Link Network: Discovery and Learning, 来源出版物是 IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 21 (6): 785-799 JUN 2009。

表 1 2009 年计算所发表论文的收录引用总体情况

SCI 收录 论文数	CPCI 收录 论文数	WOS 收录 论文被引用数	EI 收录 论文数	CSCD 收录 论文数	CSCD 收录 论文被引用数
85	212	59	466	210	15

刊载来源分析

SCI 收录计算所 2009 年发表论文刊载来源情况如表 2 所示。本统计只列出了刊载来源论文

数较多或刊载来源影响因子较高的论文情况。其中影响因子为 2009 版的期刊引用报告（JCR）数据。

表 2 SCI 收录计算所 2009 年发表论文刊载来源情况

刊载来源	影响因子	收录论文数
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY	0.632	6
SIGNAL PROCESSING-IMAGE COMMUNICATION	0.873	4
CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE	1.004	3
EXPERT SYSTEMS WITH APPLICATIONS	2.908	3
INFORMATION SCIENCES	3.291	3
BMC BIOINFORMATICS	3.428	2
CHINESE JOURNAL OF ELECTRONICS	0.156	2
COMPUTERS & MATHEMATICS WITH APPLICATIONS	1.192	2
IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING	2.285	2
IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS	1.733	2
IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS	1.010	2
IET COMMUNICATIONS	0.751	2
JOURNAL OF SYSTEMS ENGINEERING AND ELECTRONICS	0.269	2
NEUROCOMPUTING	1.440	2
PATTERN RECOGNITION LETTERS	1.303	2

SIMULATION MODELLING PRACTICE AND THEORY	0.799	2
THEORETICAL COMPUTER SCIENCE	0.943	2
ANALYTICA CHIMICA ACTA	3.757	1
BIOINFORMATICS	4.926	1
BIOLOGY DIRECT	3.322	1
BMC GENOMICS	3.759	1
BMC STRUCTURAL BIOLOGY	2.790	1
BMC SYSTEMS BIOLOGY	4.064	1
COMPUTERS & EDUCATION	2.059	1
EPL	2.893	1
IEEE MICRO	3.205	1
IEEE SIGNAL PROCESSING LETTERS	1.173	1
IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING	1.471	1
IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY	2.548	1
IEEE TRANSACTIONS ON COMPUTERS	1.822	1
IEEE TRANSACTIONS ON CONSUMER ELECTRONICS	0.942	1
IEEE TRANSACTIONS ON MAGNETICS	1.061	1
IEEE TRANSACTIONS ON MULTIMEDIA	1.822	1
IEEE TRANSACTIONS ON NEURAL NETWORKS	2.889	1
INTERNATIONAL JOURNAL OF INNOVATIVE COMPUTING INFORMATION AND CONTROL	2.932	1
JOURNAL OF INFORMATION SCIENCE	1.706	1
JOURNAL OF STATISTICAL MECHANICS-THEORY AND EXPERIMENT	2.670	1
MOLECULAR & CELLULAR PROTEOMICS	8.791	1
NEW JOURNAL OF PHYSICS	3.312	1
PATTERN RECOGNITION	2.554	1
RAPID COMMUNICATIONS IN MASS SPECTROMETRY	2.695	1

CPCI 收录计算所 2009 年发表的论文刊载来源情况如表 3 所示。这里列出了刊载来源论文数大于 4 篇的会议名称。从中可以看出, CPCI 收录的计算所发表论文较多的会议有: IEEE 网络、体系结构与存储国际会议 (IEEE INTERNATIONAL CONFERENCE ON NETWORKING, ARCHITECTURE, AND STORAGE), IEEE 多媒体国际会议与展览 (IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO), 网格与协同计算国际会议 (INTERNATIONAL CONFERENCE ON GRID AND COOPERATIVE COMPUTING), IEEE 全球电信会议 (IEEE GLOBAL TELECOMMUNICATIONS CONFERENCE) 等。

表 3 CPCI 收录计算所 2009 年发表的会议论文刊载来源情况

会议名称	收录 论文数
IEEE INTERNATIONAL CONFERENCE ON NETWORKING, ARCHITECTURE, AND STORAGE	16
IEEE INTERNATIONAL CONFERENCE ON MULTIMEDIA AND EXPO	11
8TH INTERNATIONAL CONFERENCE ON GRID AND COOPERATIVE COMPUTING	10
IEEE GLOBAL TELECOMMUNICATIONS CONFERENCE (GLOBECOM 08)	10
IEEE INTERNATIONAL SYMPOSIUM ON PARALLEL AND DISTRIBUTED PROCESSING WITH APPLICATIONS	9
17TH INTERNATIONAL CONFERENCE ON GEOINFORMATICS	8
15TH PACIFIC RIM INTERNATIONAL SYMPOSIUM ON DEPENDABLE COMPUTING	5
IEEE INTERNATIONAL CONFERENCE ON CLUSTER COMPUTING (CLUSTER 2009)	5
IEEE/WIC/ACM INTERNATIONAL CONFERENCES ON WEB INTELLIGENCE (WI)/ INTELLIGENT AGENT TECHNOLOGIES (IAT)	5
16TH IEEE INTERNATIONAL CONFERENCE ON IMAGE PROCESSING	4
IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (ICC 2009)	4

EI 收录计算所 2009 年发表论文刊载来源情况如表 4 所示。这里列出了刊载来源 EI 收录论文数大于 9 篇的期刊或会议名称。从中可以看出, EI 收录的计算所发表论文数大于 30 篇的期刊或会议有: 计算机科学讲义

(Lecture Notes In Computer Science) (包括 人工智能讲义 (Lecture Notes In Artificial Intelligence) 以及生物信息学讲义 (Lecture Notes In Bioinformatics)) 和《计算机研究与发展》等。

表 4 EI 收录计算所 2009 年发表论文刊载来源情况

期刊或会议名称	EI 收录 论文数	期刊 / 会议
Lecture Notes In Computer Science (包 括 Lecture Notes In Artificial Intelligence 以及 Lecture Notes In Bioinformatics)	37	会议
计算机研究与发展	25	期刊
2009 IEEE International Conference On Networking, Architecture	16	会议
高技术通讯	14	期刊
计算机辅助设计与图形学学报	12	期刊
计算机学报	12	期刊
Proceedings Of The International Conference On Parallel And Distributed Systems - Icpads	10	会议
软件学报	9	期刊
通信学报	9	期刊



中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

地址：北京市海淀区中关村
科学院南路 6 号
邮编：100190
电话：(010) 62601116
传真：(010) 62567724
[http: //www.ict.ac.cn](http://www.ict.ac.cn)