

批准号：60921002

创新研究群体实施情况 工作报告

超并行高效能计算机体系结构与设计方法研究
2010年1月--2012年12月

学术带头人 李国杰



中国科学院计算技术研究所
Institute of Computing Technology, Chinese Academy of Sciences



提纲

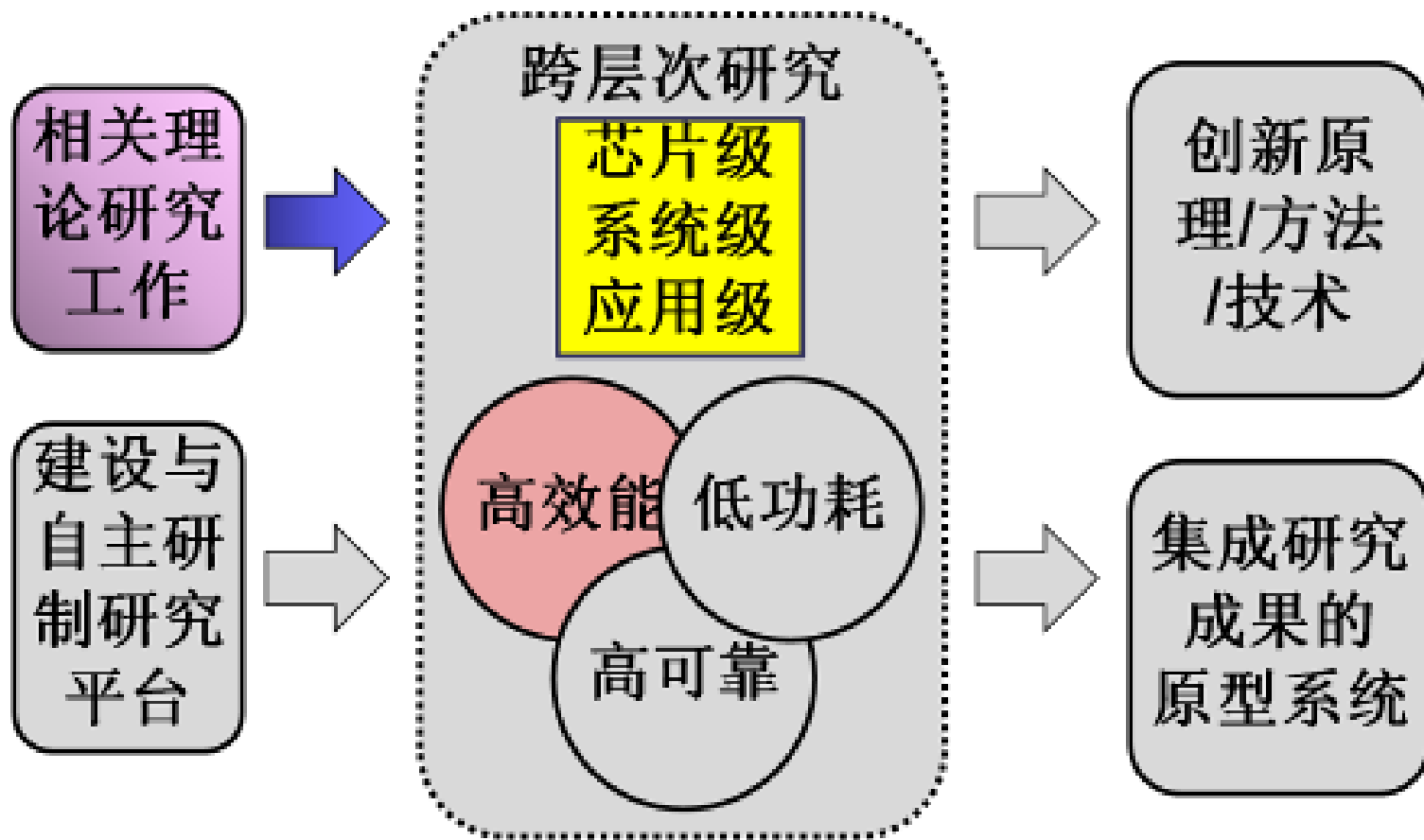
研究目标及实现情况

创新性成果及实质进展

团队建设、合作交流

后续工作计划

研究目标



以“可扩展到亿级并行度的超并行计算、低功耗、高可靠的计算机体系结构”为主要研究内容，从原理级、芯片级、系统级、应用级开展跨层级研究。

群体组成

■ 先进计算机系统方向

- 李国杰、孙凝晖、徐志伟、刘志勇、张立新、孙毓忠
陈明宇、孙晓明、詹剑锋、熊劲、侯锐、谭光明、包云岗

■ 处理器结构方向

- 胡伟武、唐志敏、范东睿、章隆兵、陈云霁

■ 编译和编程方向

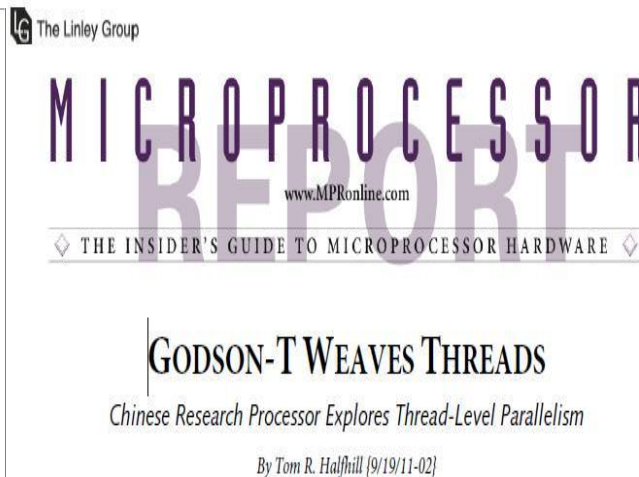
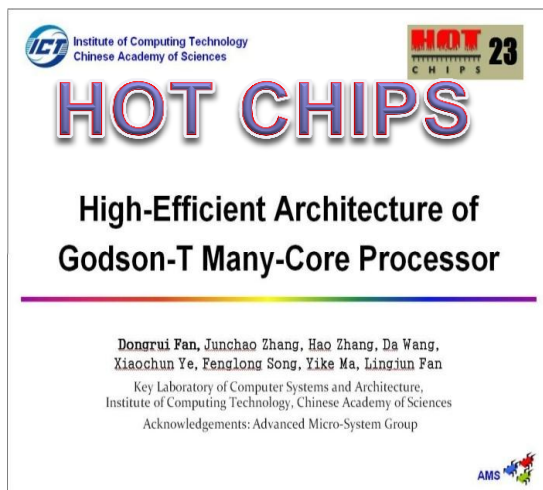
- 冯晓兵、吴承勇、武成岗、陈莉

■ 集成电路与容错方向

- 李晓维、李华伟、胡瑜、韩银和、张磊

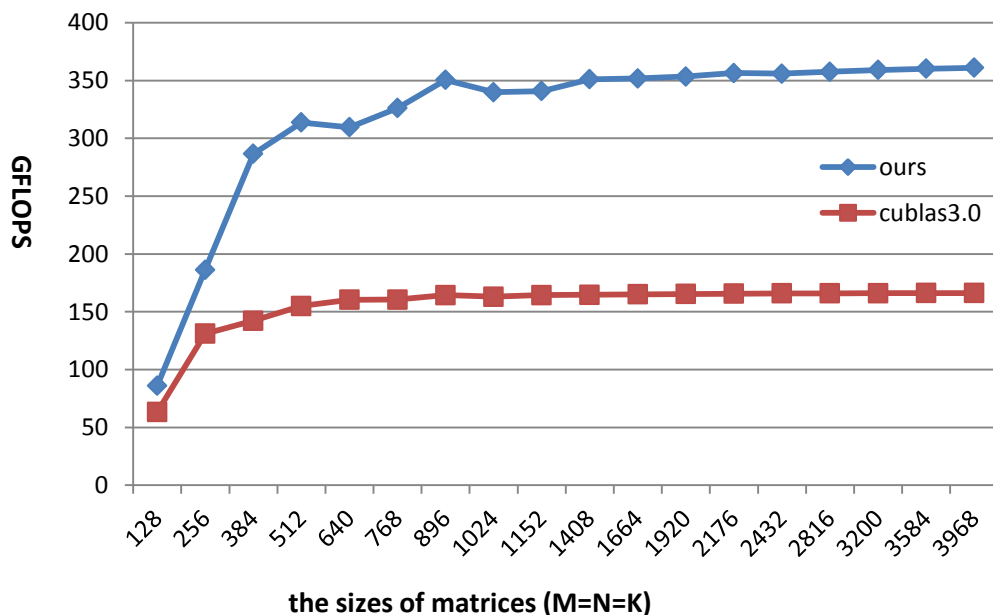
突出进展1： 并行众核处理器体系结构领域的成果与应用

- 提出新型高性能并行众核处理器体系结构
 - 研究成果应用于国家科技重大专项“核高基”——万亿次CPU的设计方案中
 - 为后续国产高性能芯片量产，提供关键技术，以实现“产学研用”的良好结合



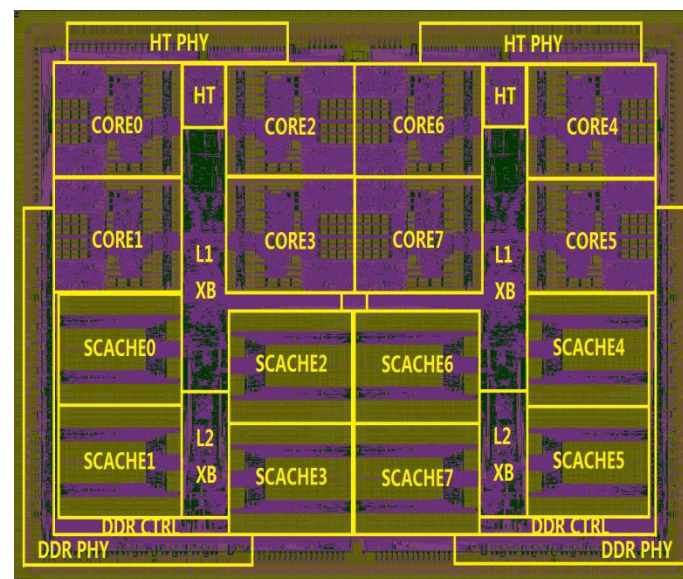
突出进展2: 超并行系统上的线性系统求解算法

- 针对CPU-GPU异构系统，提出优化稠密线性系统求解的可扩展和加速算法。
- 在曙光星云上的实现获得超过千万亿次浮点性能，同时期效率最优，总体性能提升20%，星云系统跻身世界第二的关键技术之一。
- 并用于天河1A，TSUBAME2等P级系统。
- 该论文发表于SC 2011（超级计算领域顶级会议）。



突出进展3： 八核龙芯3C处理器芯片的基础技术

- 半监督的体系结构设计空间搜索
- 基于全局时钟的并程序调试
- 高效的DMA缓存设计
- 集成8个超标量处理器核心，每核集成双256位向量部件
- 峰值运算能力192GFLOPS@1.5GHz
- 成果发表于ISCA2010、HPCA2010、IJCAI2011、ISSCC2011

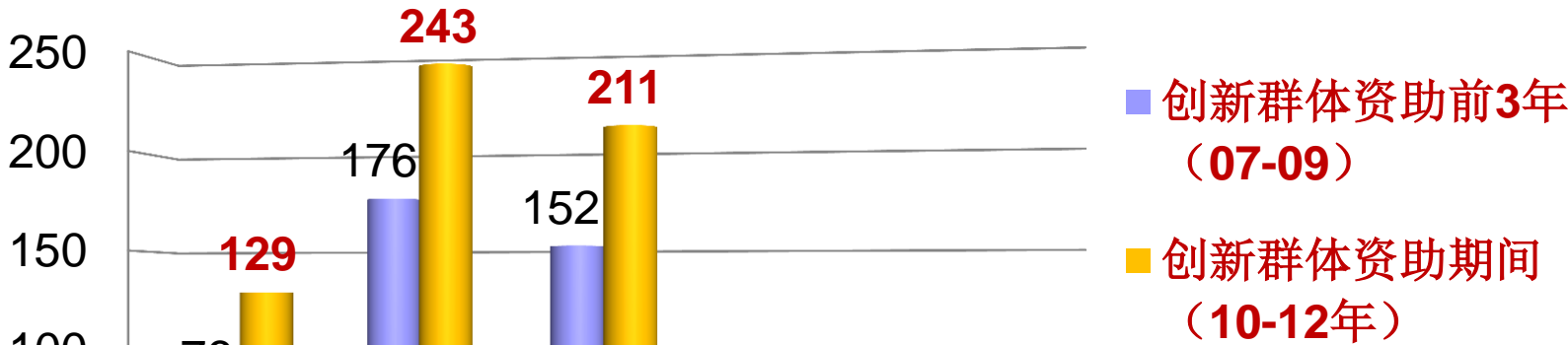


论文成果

知名期刊和领域顶级国际会议	受群体资助 发表篇数
IEEE Micro	1
IEEE Transaction on VLSI	5
IEEE Transaction on Computer	1
ACM Transactions on Architecture and Code Optimization	2
Journal of Computer Science and Technology	6
ACM/IEEE International Symposium on Computer Architecture (ISCA)	2
ACM International Conference on Supercomputing (ICS)	1
ACM/IEEE Super Computing (SC)	1
IEEE International Symposium on Code Generation and Optimization (CGO)	5
International Conference on Parallel Architectures and Compilation Techniques (PACT)	1
ACM/IEEE Design Automation Conference (DAC)	2
IEEE /ACM Conference on Design, Automation, and Test in Europe conference (DATE)	17
IEEE International Symposium on High Performance Computer Architecture (HPCA)	1
International Symposium on Microarchitecture (MICRO)	1



群体资助前后成果对比情况



会议名称	创办年份	长文录用率	大陆论文数量	本团队数量	本团队所占比例	说明
ISCA	1973	18%	4	3	75%	
HPCA	1994	18%	7	4	57%	大陆单位首次发表
MICRO	1968	15-20%	2	1	50%	
ASPLOS	1982	15-20%	1	1	100%	大陆学者首次发表双年会议

■ 在上述顶级国际会议中，本项目成员发表的论文占国内作者发表论文总数的40%左右。

国家杰出青年科学基金

1人

千人计划

1人

自然科学基金优秀青年科学基金

2人

中科院“青年创新促进会”

6人

中科院卢嘉锡青年人才奖

3人

中组部“青年拔尖人才计划”

2人

中科院优秀博士学位论文

1人

中国科学院院长奖优秀奖

1人

持续提升

委以重任

国际合作

悉心指导

团队攻关

创新群体资助

优秀人才脱颖而出

国家杰出青年基金

孙凝晖

中组部“千人计划”

张立新

优秀青年科学基金

孙晓明、陈云霁

中组部“青年拔尖人才计划”

孙晓明、陈云霁

中科院卢嘉锡青年人才奖

胡瑜、韩银和、谭光明

北京市“科技新星”

范东睿

中科院“青年创新促进会”

李华伟、胡瑜、范东睿、韩银和、谭光明、陈云霁

中科院计算所研究员

胡瑜、詹剑锋、陈云霁（29岁）

在国际学术舞台暂露头角

- **孙凝晖**研究员、**张立新**研究员、**孙毓忠**研究员、**武成岗**副研究员作为大会主席/程序主席组织体系结构重要会议 HPCA、Cluster和CGO
- 33岁的**范东睿**副研究员已成为众核处理器领域国际上知名度较高的学者，仅两年已在Supercomputer、 Hotchip等**6个**重要国际学术会议上作特邀报告
- **韩银和**副研究员担任HPCA的程序委员会委员，是国内年轻学者首次（ISCA/MICRO/HPCA以前均无）





提纲

研究目标及实现情况

创新性成果及实质进展

团队建设、合作交流

后续工作计划

取得的创新性成果

1. 高性能计算机超并行体系结构

2. 高效能计算机设计方法

3. 高并行度、低能耗的众核处理器体系结构

4. 多核处理器的测试、验证与可靠设计方法

5. 高性能计算机的并行编程和优化

代表成果1:

面向E级计算的不停顿算法级容错方法

■ 问题

- E级并行计算局部故障导致全局“停-等”将带来巨大的开销。

■ 创新点

- 提出结合基于故障**瞬时切换**、最终**结果变换**算法、后台**加速恢复**冗余的不停顿算法容错方法ABFT-HRBR, 使E级计算容错可行。

- 所提方法将故障处理开销降低 $\log(Q)$ 倍, 从 $\log(Q)$ 步减少到一步 (Q 是节点数的平方根)
- 成果在高性能领域国际会议HiPC、IPDPS上发表

代表成果2:

面向虚拟化数据中心的能力流动模型及机制

■ 问题

- 面对有服务质量和应用特征差异的大规模并发请求流，产生复杂的**资源竞争导致效能低下**

■ 创新点

- 首次提出从应用层、虚拟化层和物理层进行任务、请求和资源的协调，实现计算能力按需流动的模型和机制

- 能力流动算法提升高优先级应用性能**25%-42%**
- 成果在服务计算领域的顶级刊物《IEEE Transactions on Computing Service》和 IEEE CCGRID、Cluster会议发表

代表成果3: 指令空间重用技术

■ 问题

- 并行计算应用中，线程之间可能存在大量的指令和数据共享。消除冗余计算不仅能提高性能，还能节省能耗

■ 创新点

- 基于，提出一种可以实现指令空间重用的体系结构，以消除线程之间的冗余计算

- 所提方法可最高可消除26%的冗余计算
- 成果在微体系结构领域国际顶级会议MICRO上发表

代表成果4： 自测试自诊断自修复体系结构设计方法

- 流水线时延感知： **自测试**
 - 通路时延测量和故障通路定位： **自诊断**
 - 流水级间延迟补偿： **自修复**
- 首次在体系结构级完整的使用自测试-自诊断-自修复思路
 - 成果在领域国际顶级会议和期刊ISCA, IEEE Trans. Computer上发表

代表成果5： 迭代编译技术

■ 问题

- 迭代编译是对传统编译优化方法的有力补充，但是该技术的**实用性还很局限**

■ 创新点

1. 验证了迭代编译的**技术基础**——数据集的敏感性
2. 实现了**面向数据中心和云服务器**的迭代优化策略——IODC

- 首次在1000个数据集上对迭代编译技术进行评价
- IODC对Mapreduce应用的加速比：平均1.48，最高2.08
- 成果在编程/编译领域国际顶级会议PLDI和ASPLOS上发表

特色成果-自研平台

自研平台

Godson-T众核处理器原型芯片

支持高效能并行计算机体系结构研究的硬件模拟平台

HMTT (Hybrid Memory Trace Toolkit) 软硬件混合内存监测平台

部件级功耗监控系统

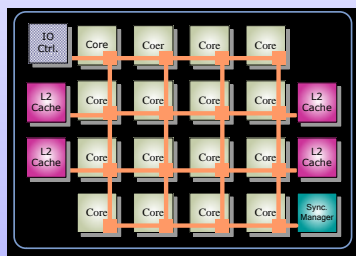
众核处理器的在线3S软硬件研究平台



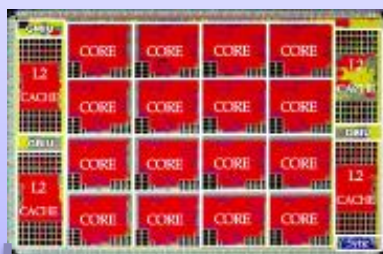
自研平台1: Godson-T众核处理器原型芯片

- 16个MIPS处理器核，4个片上共享L2-cache
- SMIC 130nm 8层金属工艺，总晶体管数目1.27亿个，面积230mm²
- 在线程模型、存储模型、同步模型和结构模型上具有一定创新性
- 在片上互连网络、片上存储系统等多个层次具有强可配置性

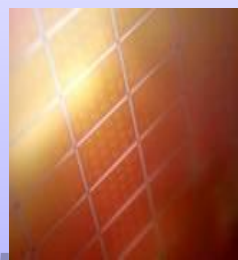
Godson-T原型芯片研制过程



RTL级系统架构



物理设计版图



晶圆



封装后芯片

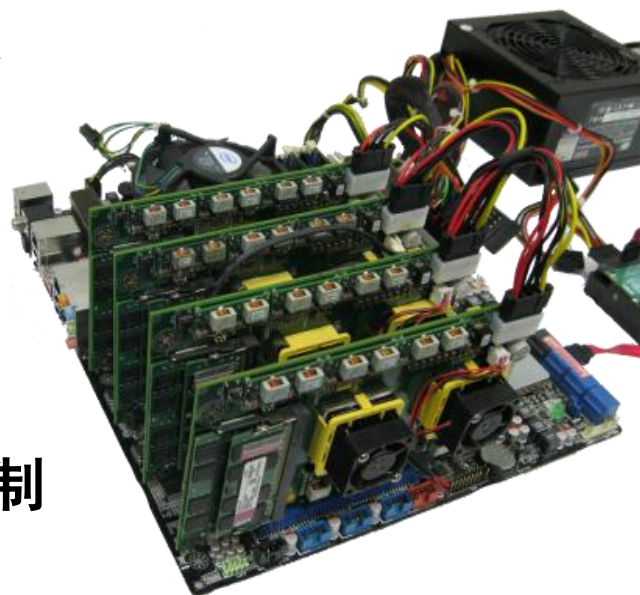


板卡系统

成果在国际顶级会议HotChips上做特邀报告

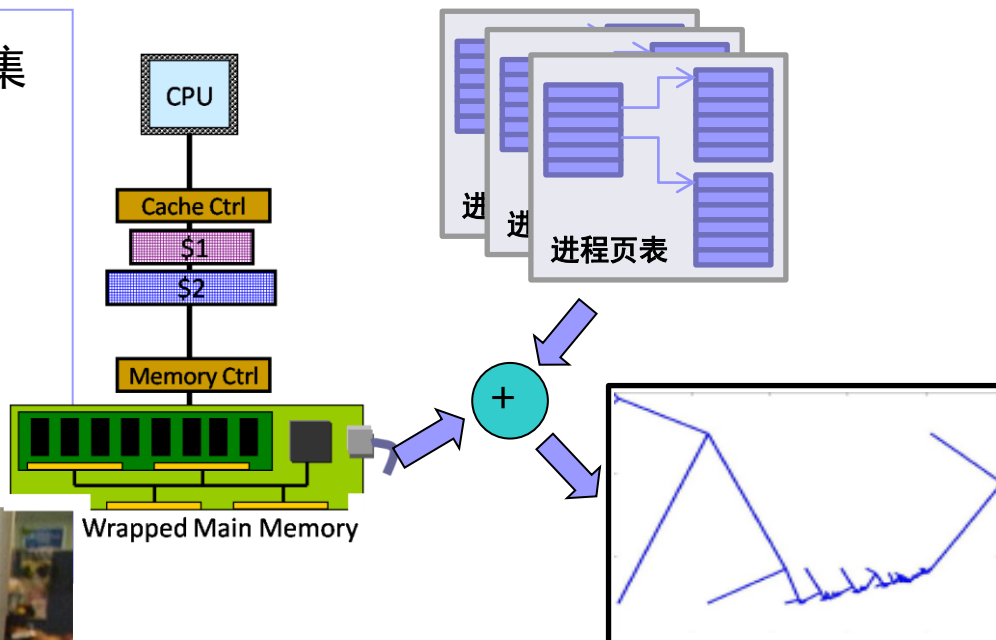
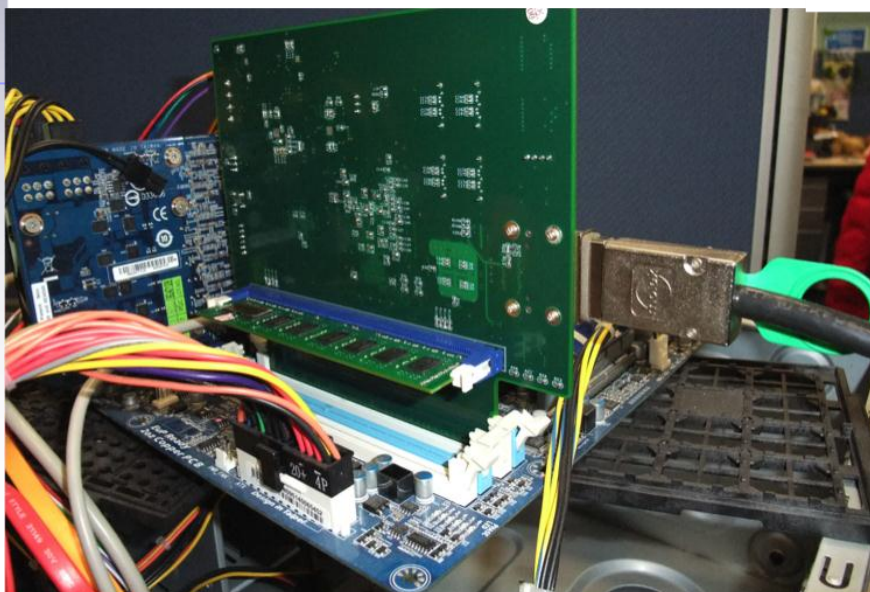
自研平台2: 基于国重芯片Godson-T的科研平台

- **Godson-T众核处理器研究平台**
 - 16个MIPS处理器核，4个片上共享L2-cache，多卡互联
- **在此平台可继续深入开展研究工作**
 - 程序性能和网络带宽及网络拥塞间的关系
 - 程序性能和LLC配置的关系
 - 系统可编程性和性能平衡等技术
- **华中科技大学基于此开发了流编译器**
- **北京理工大学基于此研究SPM/Cache机制**
- **中科院软件所基于此研究存储验证方法**



自研平台3: 低干扰访存分析系统HMTT3

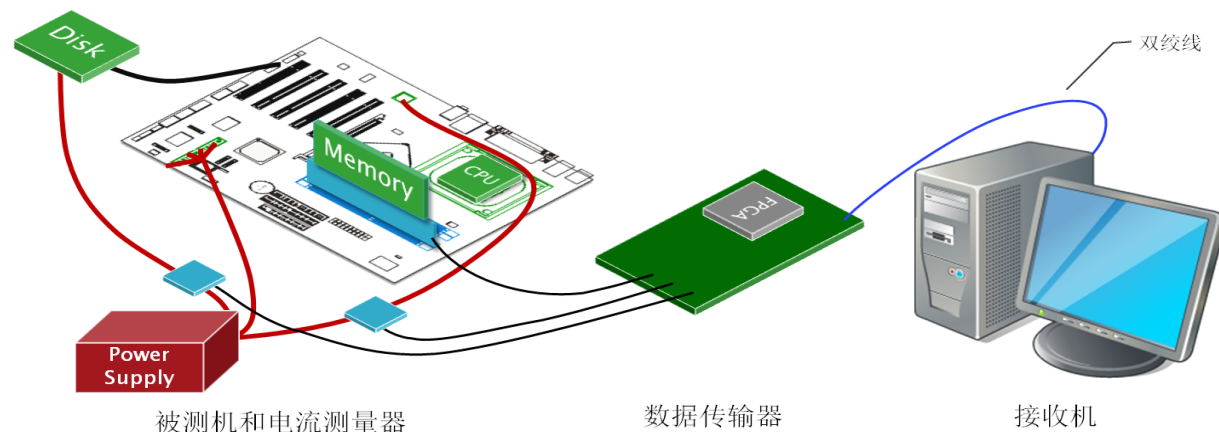
- 实时低开销真实访存地址信息采集
- 支持DDR3-800, 多种主流平台
- 采用PCIe cable后台传输, 支持600MB以上的trace采集速度
- 支持对象级粒度的语义分析



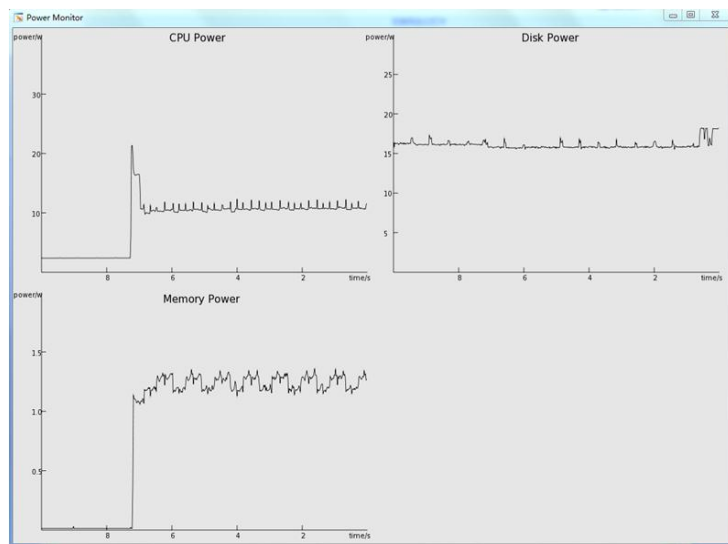
- 受到国际上内存体系结构研究领域的关注
- 与美国Stanford, CMU等大学和IBM、Intel公司, 以及法国、德国等研究机构展开合作

发表ISPASS2012, PACT2012

自研平台4: 部件级功耗分析系统



电流测量器



- 区分CPU、内存、磁盘等功能部件的功耗测量
- 每秒大于5000次采样，细粒度功耗分析
- 结合软件标注，分析短时期应用对功耗影响

发表PMP2011, HPC-China2012

自研平台5: 众核处理器的在线3S软硬件研究平台

搭载Alpha 21264、LEON3四核处理器已评估基于窄位宽的数据级冗余、动态重配置与配置回读的星载处理器可靠性设计、三模冗余/双模冗余的故障检测与隔离以及基于硬件检查点的故障恢复等可靠性设计



DATE'10最佳论文奖提名、IEEE Trans. VLSI发表，被Springer出版的英文专著引用

第 228 页 »

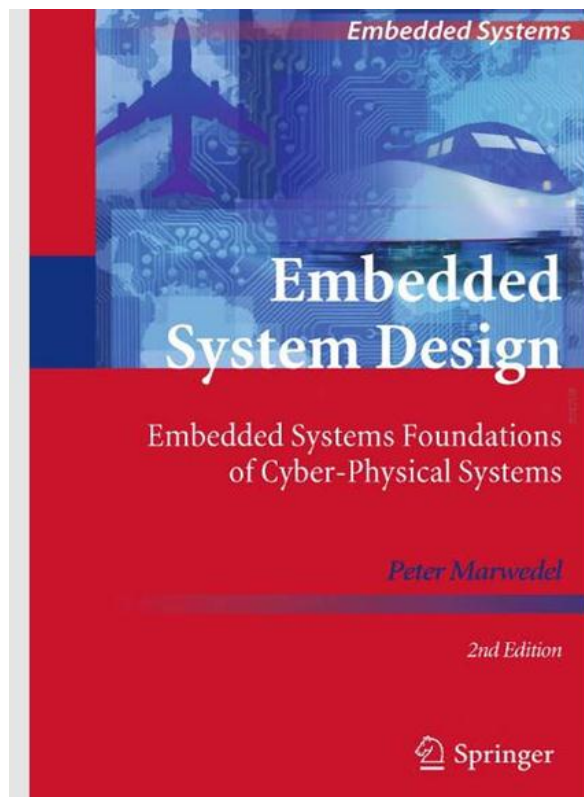
and Motet, 2002] on those areas.

There is an abundant amount of recent publications on the subject of reliability issues on system design. Examples include publications by Huang [Huang and Xu, 2010], Zhuo [Zhuo et al., 2010], and Pan [Pan et al., 2010].

5.6 Simulation

第 364 页

[Pan et al., 2010] Pan, S., Hu, Y., and Li, X. (2010). IVF: Characterizing the vulnerability of microprocessor structures to intermittent faults. ...





提纲



研究目标及实现情况



创新性成果及实质进展



团队建设、合作交流



后续工作计划



- 2011年，科技部发文（国科发基【2011】517号文件），原“中国科学院计算机系统结构重点实验室”正式获批建设“计算机体系结构国家重点实验室”，这是我国在计算机体系结构方向唯一的一个重点实验室。

与华为公司的战略合作

- 2011年9月27日，筹建中的计算机体系结构国家重点实验室与全球领先的通讯技术解决方案供应商——华为技术有限公司成立了**华为-中科院计算所联合实验室**。
- 本团队与华为公司强强合作开创了龙头企业与骨干科研院所合作的新模式，不同于一般的企业委托开发的横向合作，而是从事引领技术潮流的前瞻性研究，华为同意此项合作产生的知识产权可以辐射到其他公司。

积极而广泛的国际学术交流

● 国际学术交流

- 邀请10余人次著名国外教授、专家来访交流
- 与美国UCSB大学、瑞士EPFL、法国INRIA等研究机构开展合作



组织会议的创新群体成员

- CLUSTER-2012首次在中国
- PLDI-2012首次在中国
- HPCA-2013首次在中国
- CGO-2013首次在中国
- CTC-2012



特邀报告 (1)



会议名称: **2012 International Parallel and Distributed Processing Symposium (2011 IPDPS)**

报告时间: 2012年5月26日 地点: 上海

报告题目: **Building Billion-Threads Computer and Elastic Processor (大会Keynote)**



会议名称: **GFP 2010 : 7th IEEE International Conference on Group IV Photonics**

报告时间: 2010年9月1日 地点: 中国北京

报告题目: **More Than Moore's Law: CPU Architecture with Optical Interconnection**



会议名称: **2011 International Symposium on VLSI Design Automation & Test (2011 VLSI-DAT)**

报告时间: 2011年4月26日 地点: 台湾新竹

报告题目: **GPU-Accelerated Fault Simulation and It's New Applications**

报告人: 李华伟

特邀报告 (2)



会议名称: **2011 Hot Chips: A Symposium on High Performance Chips (2011 HotChips)**

报告时间: 2011年8月17日 地点: 美国斯坦福大学

报告题目: **High-Efficient Architecture of Godson-T Many-Core Processor**

报告人: 范东睿



会议名称: **IEEE/ACM Design Automation Conference(DAC)的Workshop “DAC Workshop on Diagnostic Services In Network-on-Chips – DSNOC”**

报告时间: 2011年6月5日 地点: San Diego

报告题目: **Resilient On-Chip Communication in MPSoC Through Network Reconfiguring**

报告人: 韩银和



提纲



研究目标及实现情况



创新性成果及实质进展



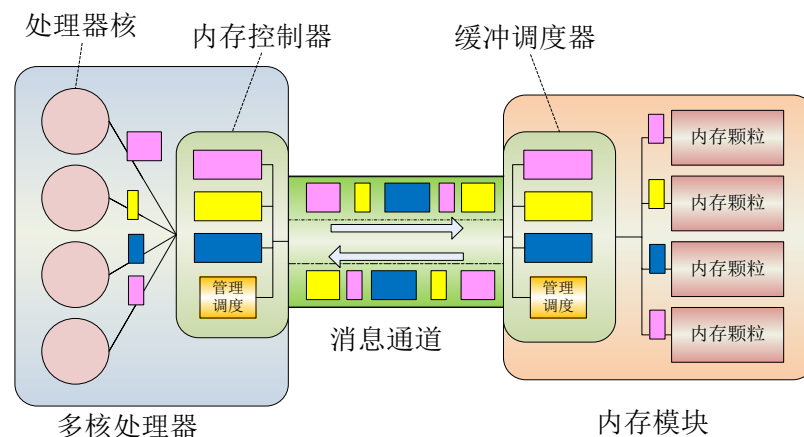
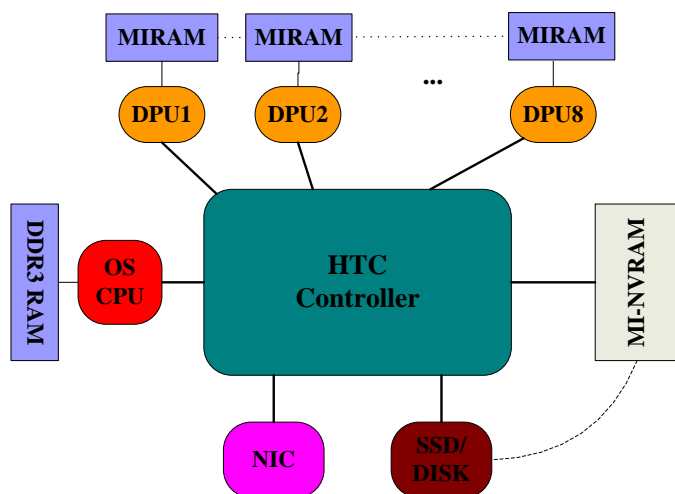
团队建设、合作交流



后续工作计划

拟开展工作1： E级高性能和高通量计算机体系结构

- 继续围绕E级高性能计算机研制的关键问题进行研究，同时关注面向高通量和数据处理的新型计算机系统，重点在访存、网络、存储等数据通道的并发性设计和全系统的低功耗设计。



拟开展工作2: 高效能计算机设计方法

- 基于现有的研究成果（基于虚拟化的可信计算平台），未来将研究**面向大数据的异构云平台**
高效能资源调度方法



拟开展工作3： 面向高通量的高能效众核处理器体系结构

■ 高吞吐

- 弱化通用核心控制
- 强化外围加速处理
- 加速器与IO紧密耦合

■ 低延迟

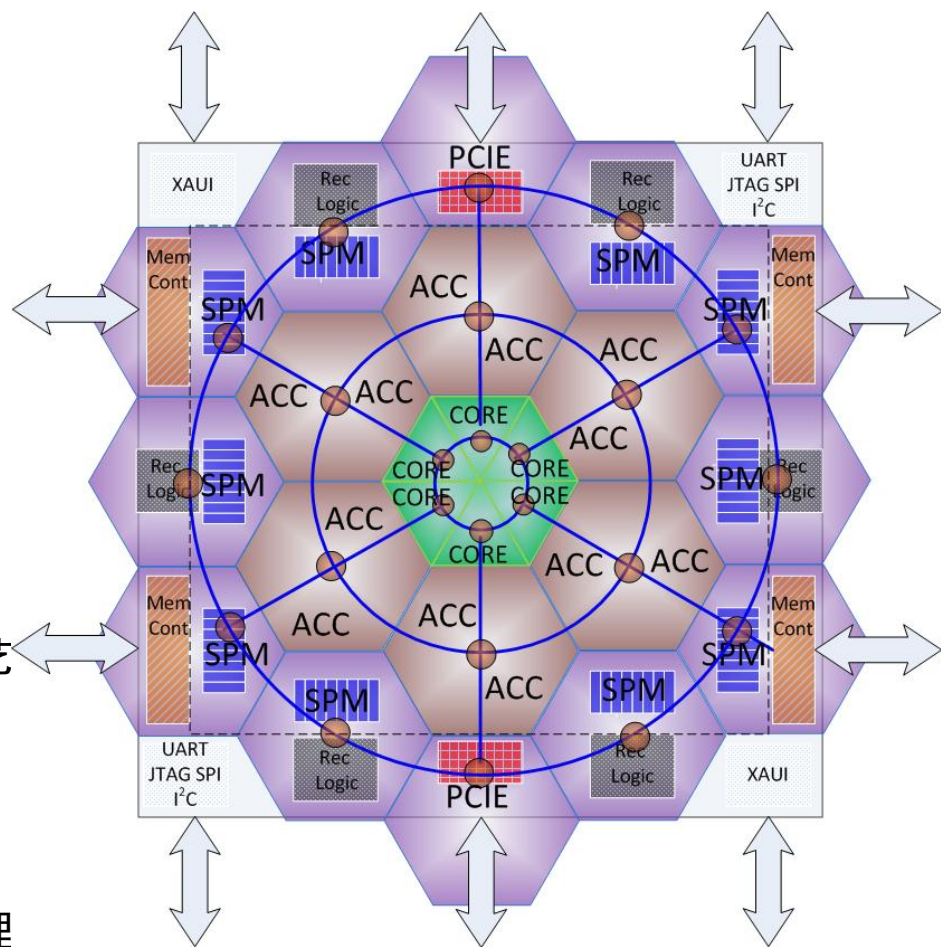
- 加速器与存储外围紧密耦合
- 减少片上数据传输延迟影响

■ 易扩展

- 易于进行同构/异构组件的扩展
- 加速器件的多层扩展组合与3D工艺扩展

■ 高实时

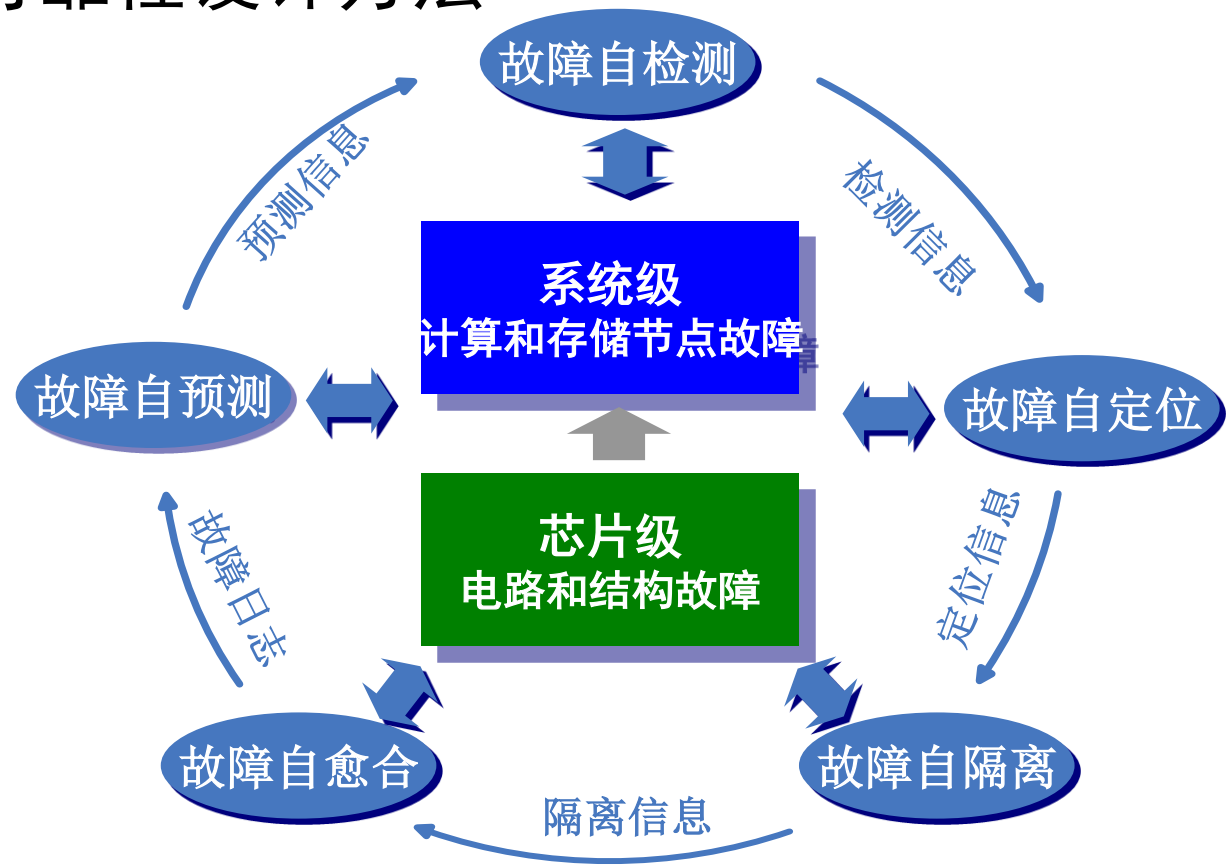
- 微结构可预测
- 片上网络和微结构可满足数据处理实时性要求



数据处理处理器Godson-D架构

拟开展工作4: 片上并行系统的测试与高可靠设计方法

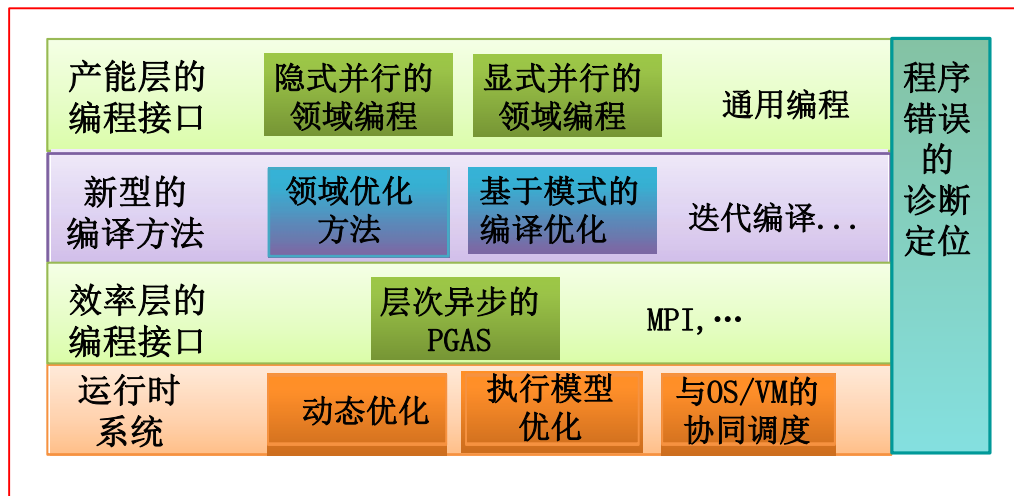
- 基于现有的研究成果（面向高性能计算的3S可靠性设计框架），未来将研究面向高通量计算的5S可靠性设计方法



拟开展工作5: 高性能计算机的并行编程和优化

- 发展现有的研究成果, 研究领域编程方法、可伸缩性好的**PGAS**语言、新型编译方法、协同的动态优化, 多层次多角度地发掘并行性

物理模拟、数据中心应用 ...

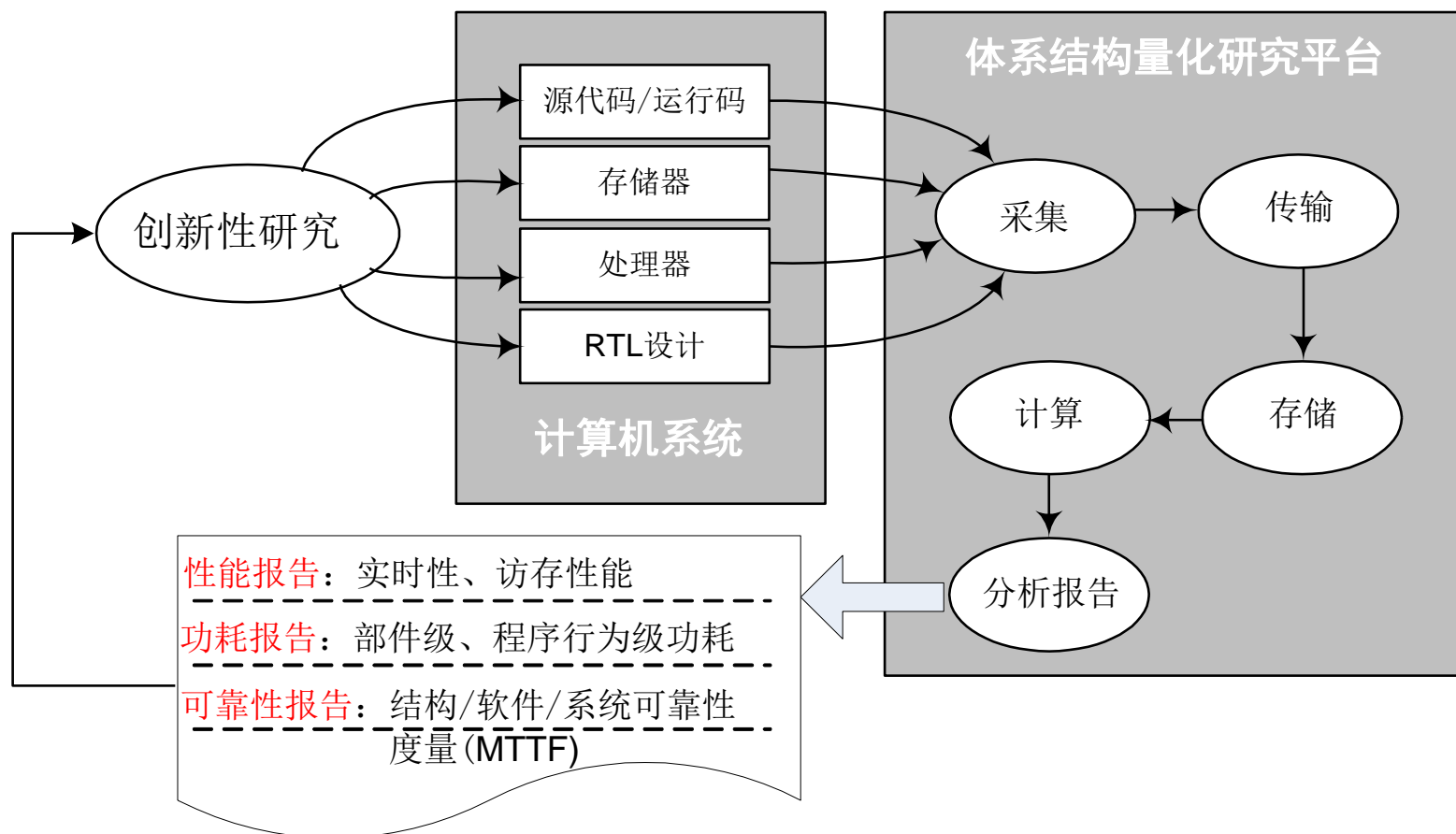


VM & 操作系统

多核、众核的集群平台

拟开展工作6: 面向计算机系统结构创新研究的量化研究平台

研制计算机系统多参数测量和综合分析平台，突破计算机系统软硬件性能、功耗和可靠性测量方面的多项关键技术，助推世界一流的突破性、变革性研究成果的诞生



创新群体二期部分量化指标

在ISCA、PLDI、MICRO、HPCA、SC、ICS、CGO、DATE等国际学术会议上发表**20篇**。

在ACM/IEEE Transactions 等高水平国际期刊上发表论文**20篇**。

申请发明专利 **100项**。

申报国家科技奖励**1-2项**。

培养3-5名计算机系统结构方向在国际学术舞台上颇为活跃的尖子人才。

本创新群体未来5-10年目标

- 单芯片**千核万线程**的处理器体系结构 and 设计方法，以及测试验证方法、自恢复方法
- **百万个处理器**芯片构成的系统的互连方法、管理方法及容错方法
- **十亿级并行度**的编程与算法设计方法及容错方法
- **ZB级大数据**的组织和处理系统设计方法及容错方法
- 高端计算如何利用新工艺、新器件（**非易失性存储器、3D集成、光互连、忆阻器**等）的最新成果

本群体追求的引领作用

从源头引领一个超过10亿美元的计算机产业

	\$1B+ Industry	Berkeley	Caltech	CERN	CMU	Illinois	MIT	Purdue	Rochest.	Stanford	Tokyo	UCLA	Utah	Wisc.
1 Timesharing														
2 Client/server														
3 Graphics														
4 Entertainment														

10年左右争取本群体在此榜上有位置！

9 VLSI design														
10 RISC processors														
11 Relational DB														
12 Parallel DB														
13 Data mining														
14 Parallel computing														
15 RAID disk arrays														
16 Portable comm.														
17 World Wide Web														
18 Speech recognition														
19 Broadband last mile														
Total	7	2	2	3	2	5	1	2	4	1	3	1	3	

Source:
Innovation
in
Information
Technology,
National
Research
Council
Press,
2003.



谢谢各位!
请批评指正!