

多核+众核技术加速高性能应用

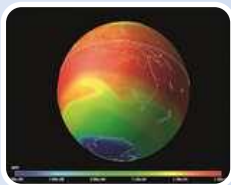
张广勇, zhangguangy@inspur.com

高性能计算应用开发工程师

浪潮-高性能服务器产品部

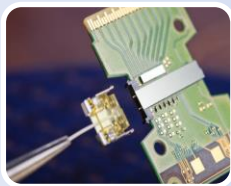
浪潮高性能进展

多核+众核技术加速高性能应用



浪潮致力于构建超算应用价值链





专用定制化超算系统设计



硬件体系结构影响软件设计方法

应用软件是否可以影响硬件体系结构设计？

如何面向特定的大规模计算应用需求进行系统的软硬件联合设计(co-design)？

如通用系统与专用系统



浪潮与高性能计算-天河一号

inspur 浪潮

性能指标:

峰值速度 (Rpeak(GFlops))	4700 万亿次 / 秒
持续速度 (Rmax (GFlops))	2566 万亿次 / 秒
共享存储总容量	2PB
全系统包含通用处理器 (CPU+GPU)	186368 个



- 参与研制的千万亿次高效能计算机系统“天河-1A”
- 在2010年11月Top500世界第一
- 浪潮完成核心高速交换主板的设计, 实现了80Gbps高速交换
- 浪潮完成计算刀片系统研发工作
- 浪潮完成8路胖节点系统设计



浪潮与高性能计算-天河二号

inspur 浪潮

2013年6月17日，top500排行榜榜首

高性能：

54.9PFLOPS（理论）

33.86PFLOPS（实测）

架构：

32000颗Ivy Bridge Xeon CPU

48000颗的Xeon Phi

承建商：

中国国防科技大学、浪潮集团





浪潮与高性能计算-神威蓝光

inspur 浪潮

国家超算济南中心神威蓝光千万亿次超级计算机



- 十一五863中大专项，由江南计算所和浪潮集团共同承担
- 峰值计算能力1.1P，其中国产处理部分1P，通用处理器部分100T
- 浪潮承担100T通用处理器部分刀片系统设计供应，采用浪潮NX560T计算刀片，2*X5675,36GB内存，500G硬盘，QDR 计算网络
- 浪潮承担2PB 高带宽海量存储系统设计供应



浪潮与高性能计算-清华大学地球模拟器

高校首个百万亿次超算，应用于全球气候变化以及清华交叉学科研究

高性能：

CPU-GPU协同加速技术，总计算性能达168万亿次每秒；

高密度：

业界最高计算密度的浪潮刀片服务器，单机柜容纳200颗最新六核处理器，9个机柜实现百万亿次超算系统

高吞吐：

计算节点之间实现40Gb/S的通信带宽，提供带宽20GB/s的总量达1000TB的海量存储系统。

高节能：

采用业界最先进的密闭式水冷机柜技术，在机柜与制冷单元LCP之间形成封闭式冷热循环通道，充分保证高密机柜35KW电功耗的散热需求



浪潮与高性能计算-上海交通大学

高校超算第一名，整体计算性能将高达到300万亿次

高性能：

CPU-GPU-MIC协同加速技术，采用最新SNB处理器构建

高密度：

业界最高计算密度的浪潮刀片服务器，单机柜容纳200颗最新六核处理器，9个机柜实现百万亿次超算系统

高吞吐：

计算节点之间实现56Gb/s的通信带宽，提供带宽20GB/s的总量达720TB的海量存储系统。

高节能：

采用业界最先进的密闭式水冷机柜技术，在机柜与制冷单元LCP之间形成封闭式冷热循环通道，充分保证高密机柜35KW电功耗的散热需求

浪潮高性能进展

多核+众核技术加速高性能应用

浪潮高性能计算应用开发团队

09年初成立，1名技术顾问，2名博士，7名硕士，5名学士

- 国内唯一具备CPU/GPU/MIC技术的专业软件开发、调优、培训、咨询团队

完成8项CPU优化项目

- 单频属性提取项目、多时窗倾角扫描

完成6项GPU/CUDA项目，正在进行2项GPU优化项目

- BLASTN、ATOM、CFD、PSTM、RNA、Scandip
- 具备从算法级、桌面软件级到集群软件的全面经验

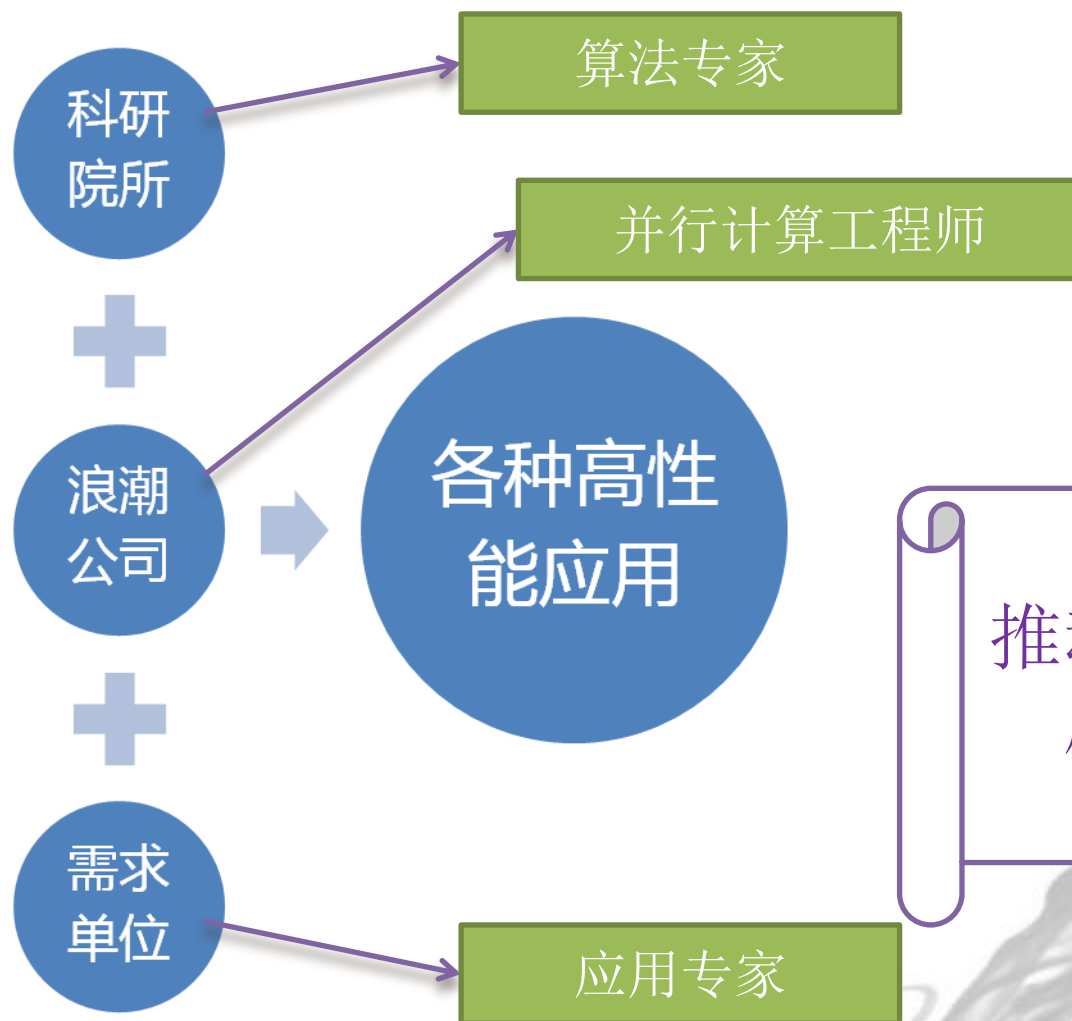
完成几十项GPU/CUDA咨询项目

- 国防科大、西工大、中科院天文台、中石化.....

完成5项MIC项目，正在进行2项MIC项目

- 石油天然气、生命科学、金融、CFD、气候气象等

合作模式

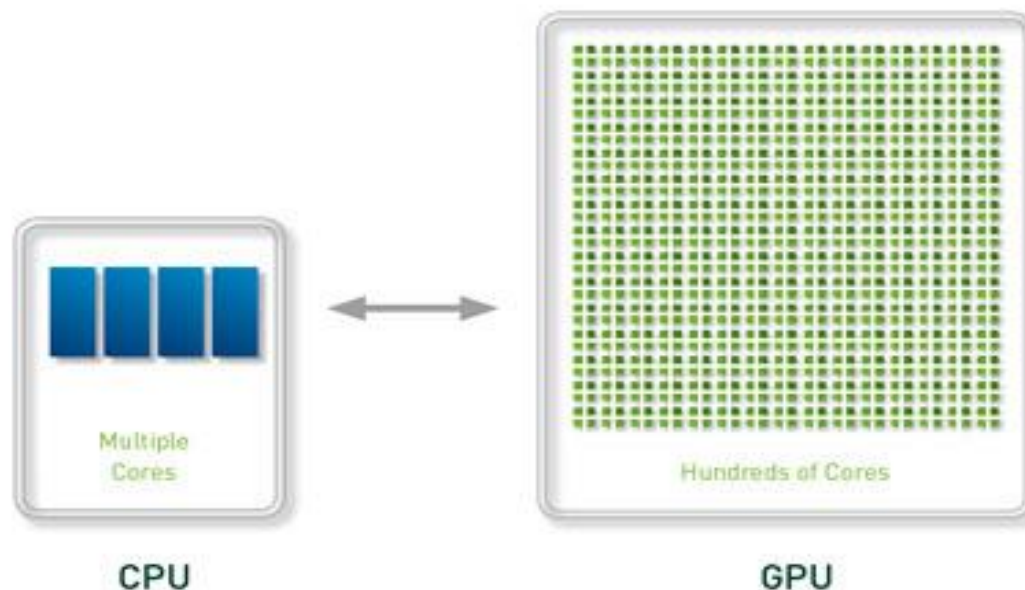


推动国产高性能计算
应用软件的发展

CPU+GPU异构协同计算

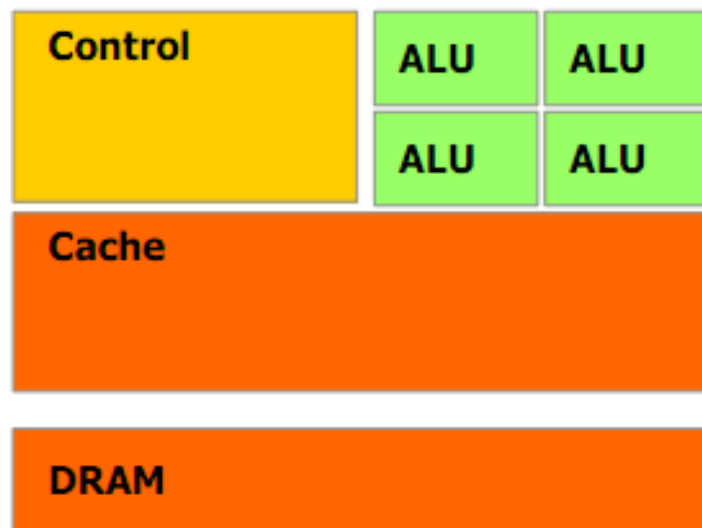
GPU计算模式

- 在异构协同处理计算模型中将CPU与GPU结合起来加以利用。应用程序的串行部分在CPU上运行，而计算任务繁重的部分则由GPU来加速。

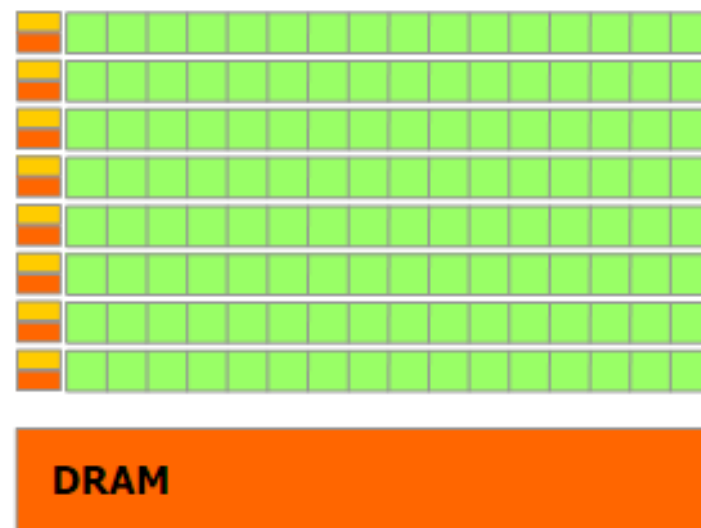


GPU的设计理念

- 更多的芯片面积用于计算单元而不是Cache
- 更简单的逻辑处理部件，但提供更高效率的轻量级线程处理
- 通过并行度的提升获得性能的提升（计算和访存）



CPU

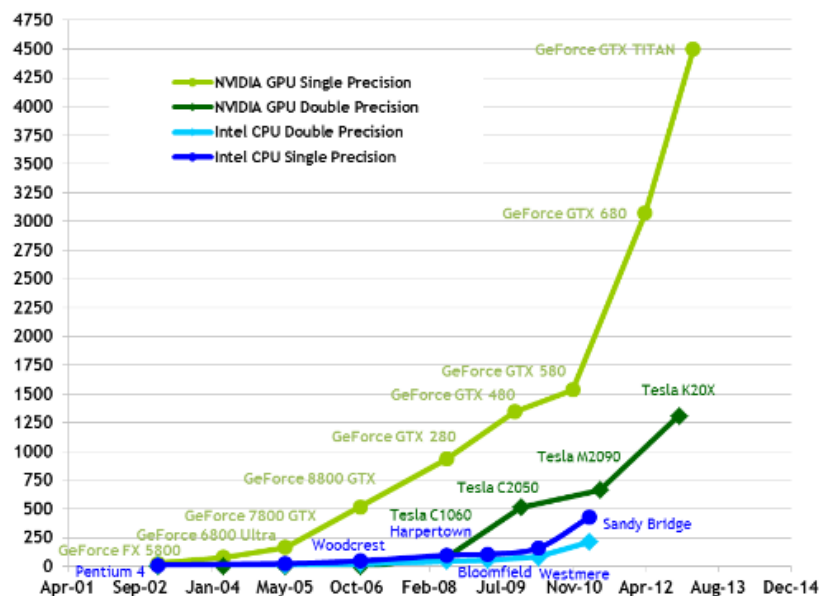


GPU

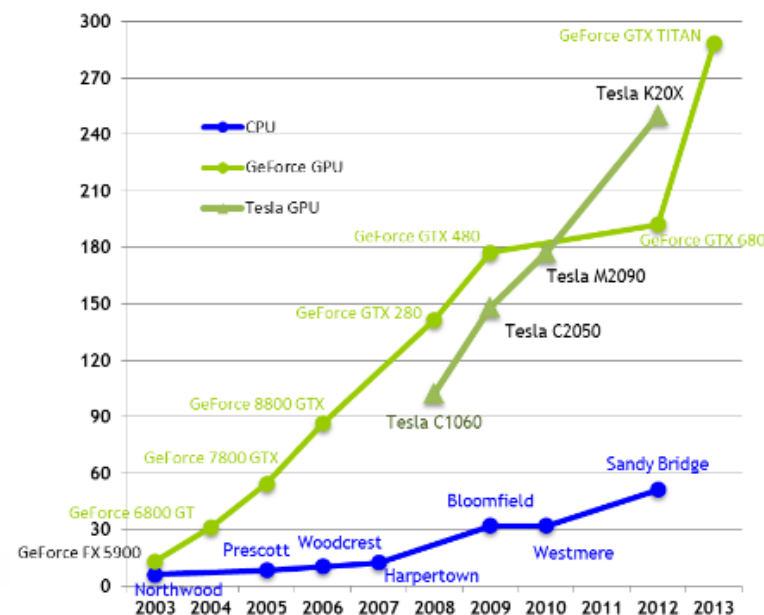
GPU的优势

- 强大的处理能力 GPU达到3Tflops/s以上
- 高带宽 290GB/s
- 低成本 Gflop/\$和Gflops/w高于CPU

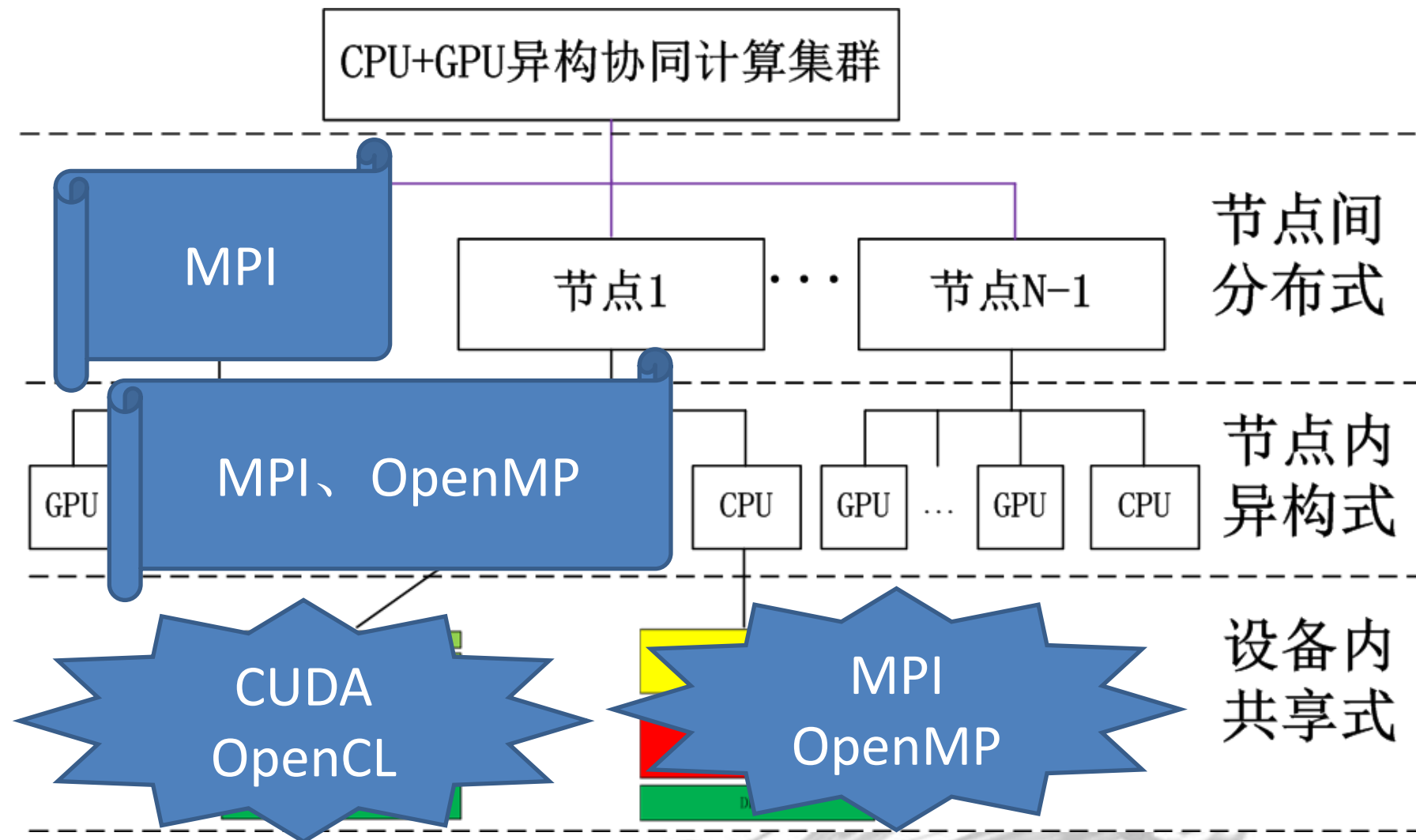
Theoretical GFLOP/s



Theoretical GB/s



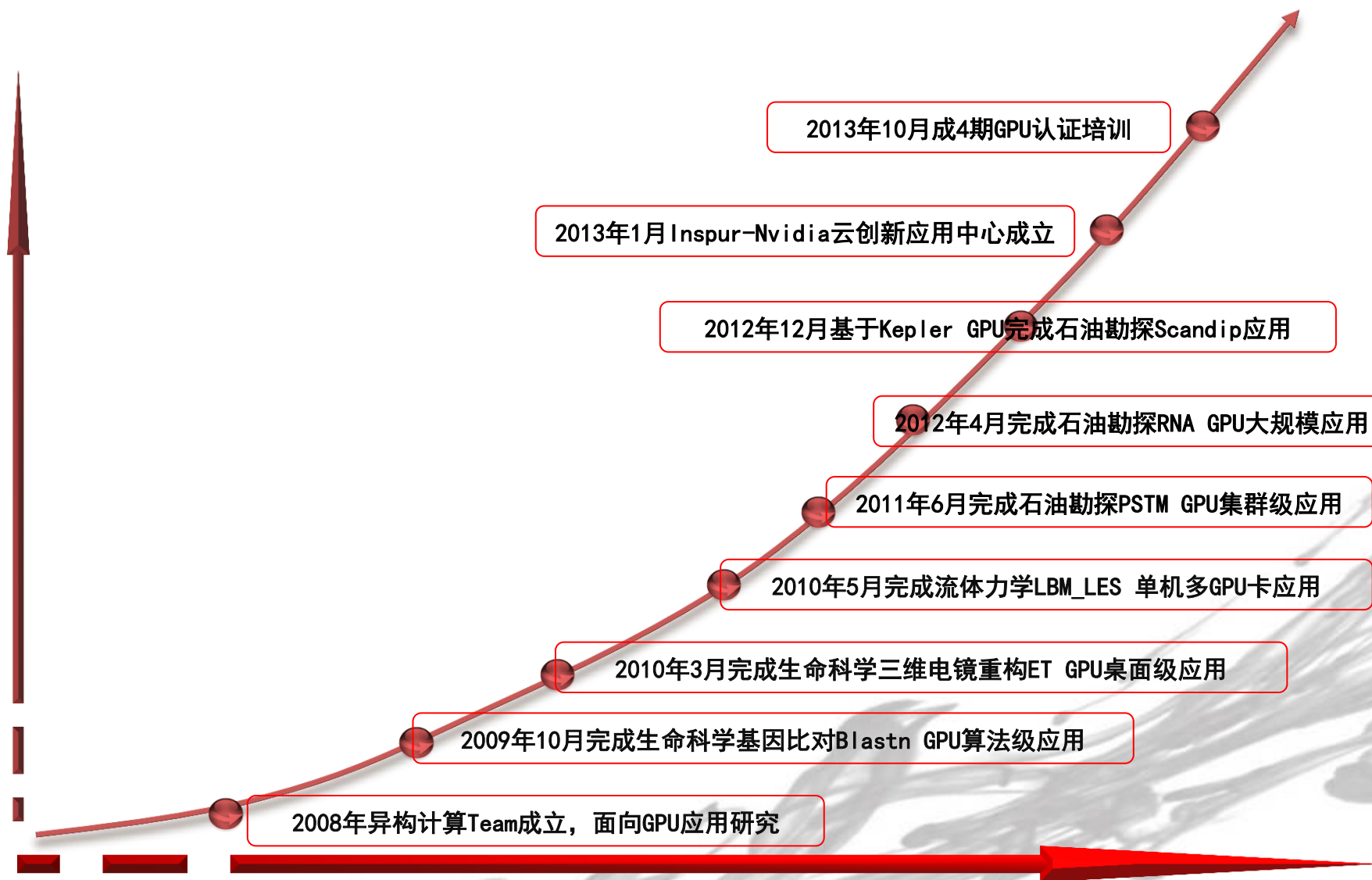
CPU+GPU计算框架



CPU+GPU计算优化技术

- 并行度
 - 多级并行：MPI+OpenMP
 - CUDA：glock、thread
- 访存优化
 - Global、shared、constant、texture、L1/L2
- 通信优化
 - CPU与GPU通信，MPI通信
- 负载优化
 - 节点间、设备间（CPU与GPU计算能力不同）、线程间

浪潮GPU技术与业务发展历程



GPU应用学术成果

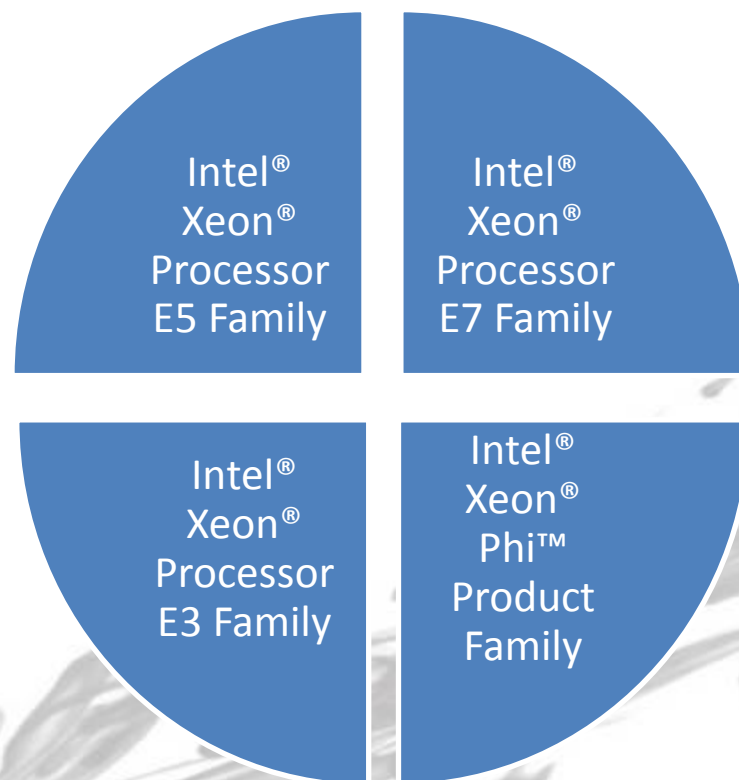
- 学术论文

《基于GPU平台的联合迭代重构算法加速》	计算机科学
《ATOM 1.0 基于 GPU的电子断层重构软件》	生物物理学报
《基于GPU实现叠前时间偏移走时计算的并行算法初探》	计算机系统应用
《一种使用GPU加速地震叠前时间偏移的方法》	微型机与应用
《使用GPU加速BLAST算法初探》	计算机工程与科学
《Implementation of a Lattice Boltzmann Method for Large Eddy Simulation on Multiple GPUs》	FGC国际会议

CPU+MIC异构协同计算

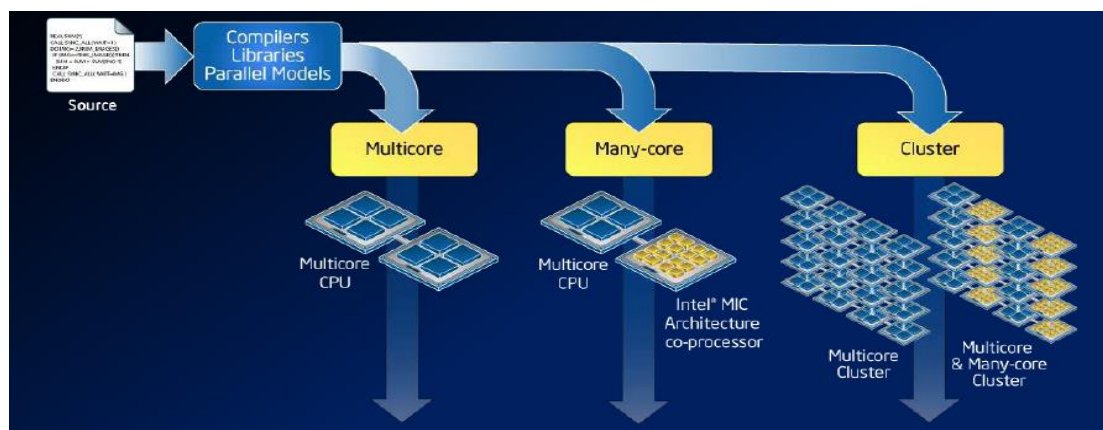
MIC介绍

- Xeon Phi是Intel® Xeon® 家族新成员
- 它的第一个正式产品
 - Knights Corner(KNC)
 - PCI-E卡
- MIC架构
 - 基于x86架构
 - 4Thread/core
 - 512位向量位宽
 - 节点方式管理（独立IP）
 - 内嵌uos
- 性能
 - 双精度>1TFlops，单精度>2TFlops



MIC技术优势

- 硬件架构（继承性）
 - 基于X86架构
- 编程语言（继承性）
 - 支持MPI/OpenMP/Pthread/OpenCL/Clik
- 编程模式（灵活性）
 - Offload
 - Native
 - Symmetric
- 工具（完整性）
 - Intel编译器、idb调试器、Vtune、MPI Trace、MKL库

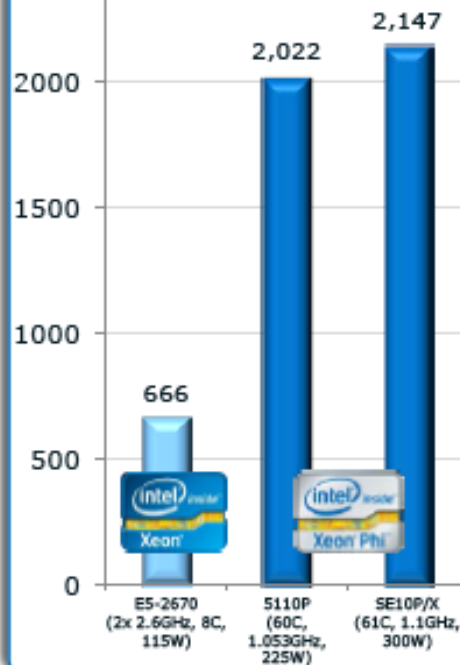


MIC VS. CPU

Single Precision (GF/s)

Up to 3.2x

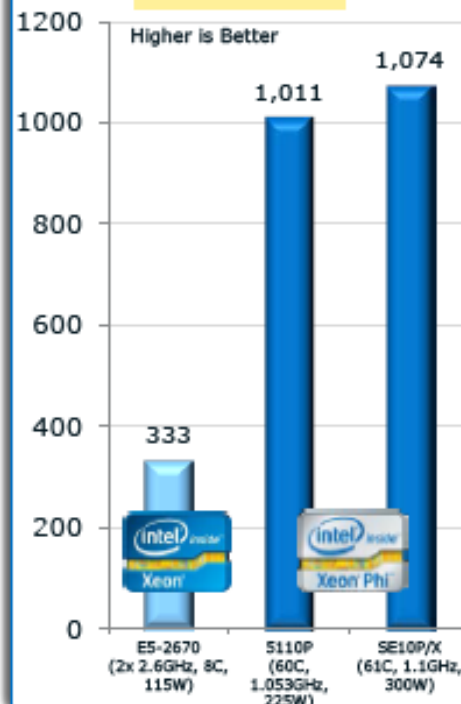
Higher is Better



Double Precision (GF/s)

Up to 3.2x

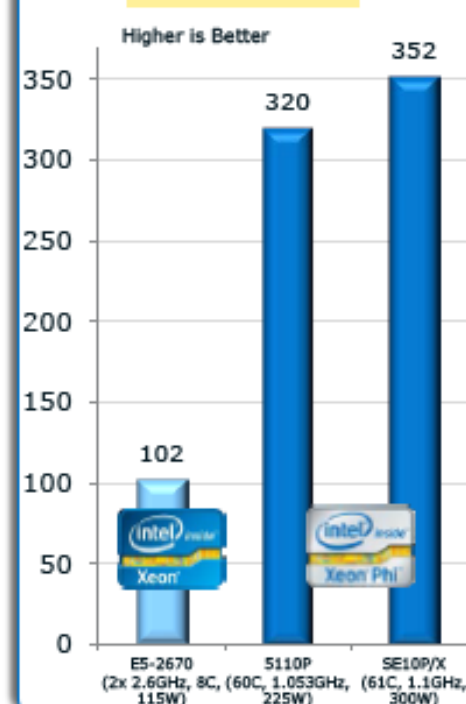
Higher is Better



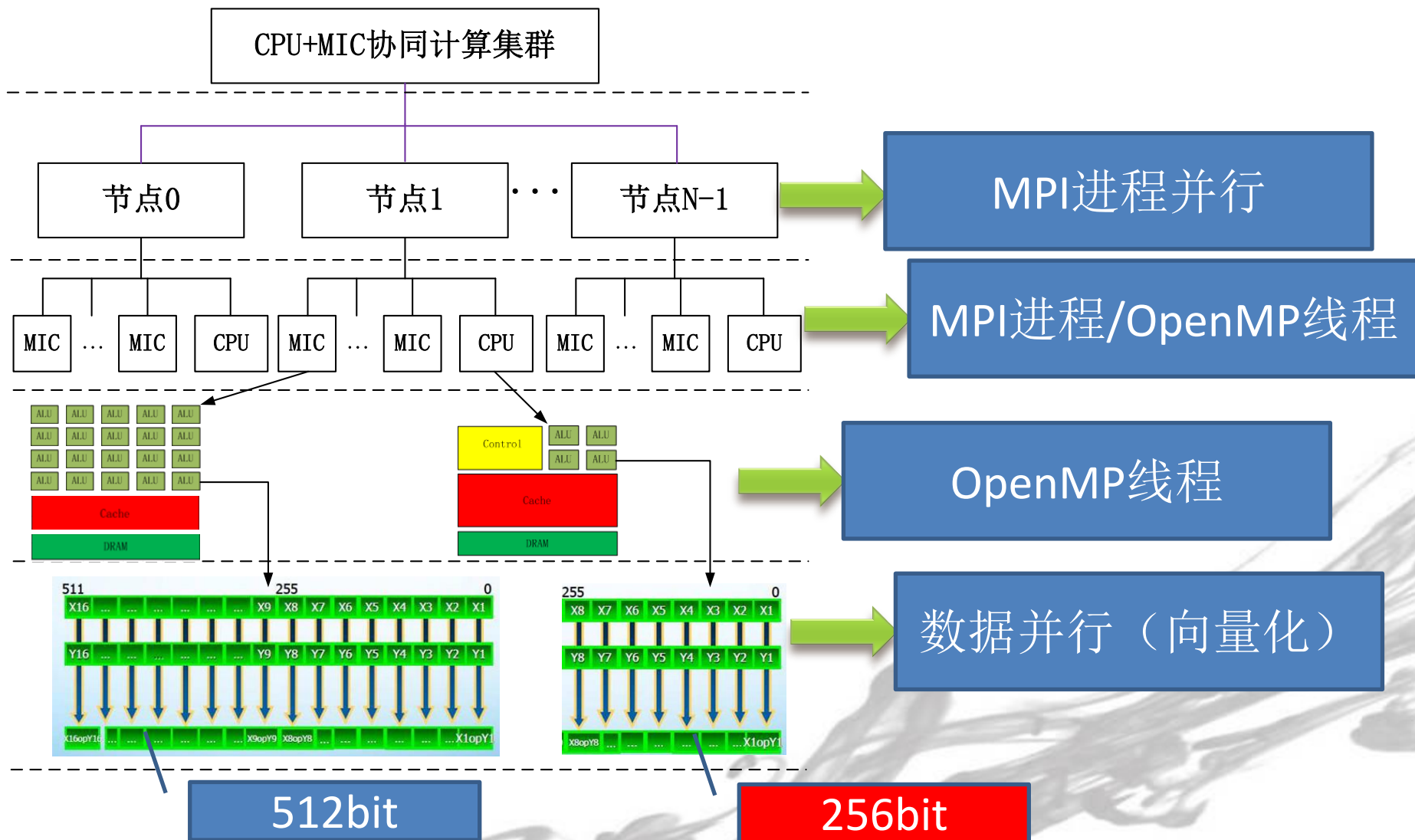
Memory Bandwidth (GB/s)

Up to 3.45x

Higher is Better



CPU+MIC异构计算架构



CPU+MIC异构计算优化

- 并行度
 - 多级并行
- 向量化计算
 - CPU: 256bit, MIC: 512bit
- 线程可扩展性
 - 几十到几百
- 访存优化
 - cache
- 通信优化
 - CPU与MIC通信, MPI通信
- 负载优化
 - 节点间、设备间、线程间

Inspur-Intel中国并行计算联合实验室



2013年9月完成基于Tianhe-2平台MIC超大规模应用实现

2013年6月-8月完成2期MIC专业认证培训

2012年11月出版《MIC高性能计算编程指南》专著

2012年9月基于KNF完成5个典型应用

2011年12月基于KNF完成4个典型应用

2011年8月实验室挂牌成立



浪潮拥有国内唯一同时具备CPU、GPU、Intel MIC技术开发能力的团队
浪潮与Intel共建“中国并行计算联合实验室”

SKA

- 简介
 - 平方公里阵SKA (Square Kilometre Array)是国际上即将建造的最大综合孔径射电望远镜。
- 进展
 - SKA 现处于建设准备阶段，已完成各关键部件概念评审；完成建设准备阶段工作任务分解和工作内容描述、正着手确定工作包，以开展关键技术攻关和样机研发。
- 浪潮工作
 - 中方科学数据处理（SDP）联盟的主要单位，将参与计算平台，数据的传输与存储，软件的加速（GPU或MIC加速），及软硬件系统的局部监控与控制的相关工作

多核+众核技术加速天文应用

- 天文应用需要超强的计算能力，CPU+GPU或CPU+MIC协同计算平台可以加速天文应用的性能，降低成本、功耗。

——高性能计算应用优化题目征集

- 共享您的应用源码，您将可能得到10倍/100倍的性能更优的多核/众核版本
 - BSDE期权优化是本次ASC13竞赛的特色应用之一，首次被拿来作为竞赛的比赛项目。
华中科技大学在本次竞赛中以性能优化**6万**倍的惊人成绩摘取“MIC优化特别奖”。



- 应用源码要求
 - <http://www.asc-events.org/ASC14/13en/Codes.php>



inspur 浪潮

谢谢！