



中山大學

SUN YAT-SEN UNIVERSITY

超级计算机原理与操作

超级计算机的性能评测

吴迪

中山大学计算机学院

2023年春季

计算机性能

计算机性能指标

- **峰值性能 (Peak Performance)**
 - 理论上超级计算机的硬件资源可以完成操作的最大速率（如每秒执行的操作总数）
 - 峰值性能由**时钟频率和体系结构**决定。
- **持续性能 (Sustained Performance)**
 - 超级计算机在运行应用程序时达到的实际或真实性能
 - **持续性能 \leq 峰值性能**
- **基准测试程序**是用于评测持续性能的特定程序

性能退化 (SLOW)

- **饥饿 (Starvation)**
 - 缺少负载，或者负载不能均匀分布
- **延迟 (Latency)**
 - 访存延迟、数据传输延迟、流水线执行延迟
- **开销 (Overhead)**
 - 和计算无关的额外工作量（如任务调度、资源管理）
- **等待 (Waiting)**
 - 对共享资源的等待

性能提升

- **提升高性能计算机的性能主要方法：**
 - 硬件扩展
 - 并行算法
 - 性能监控
 - 工作与数据分发
 - 任务粒度控制
 - 其他

大纲

- **基准评测介绍**
- **计算性能评测集**
- **IO性能评测集**
- **网络性能评测集**
- **能耗评测集**
- **应用评测**

基准评测介绍

- 性能基准评测就是从**基准测试程序和测试规范**的角度评价和预测系统的性能。
- 基准评测：
 - 可以帮助机构确定所需**采购**的超级计算机
 - 可以指导制造商高性能计算机系统的设计方向
 - 是探索HPC 趋势的**重要历史记录**

常见基准评测分类

- 计算性能评测集：Linpack, HPCG, Graph500.
- IO性能评测集：MDTest, IOR, IO500.
- 网络性能评测集：IMB, OSU Benchmark.
- 能耗评测集：Green 500.
- 应用评测集：Miniapplication、戈登·贝尔奖.

计算性能评测集

Linpack

- 用来度量系统的**浮点计算能力**。
- Linpack性能是指**求解双精度线性代数方程组时所达到的实际性能**
 - 包括Linpack100, Linpack1000, HPL
 - Linpack100 和 Linpack1000 分别求解规模为 100 阶和 1000 阶的线性代数方程组。
 - HPL是针对现代的并行计算机的评测基准。
- 评测指标：**每秒浮点运算次数 (flops)**

HPL

- HPL 具体为求一个 n 维的线性方程组的解：

$$Ax = b; A \in \mathbb{R}^{n \times n}; x, b \in \mathbb{R}^n$$

- 当系数矩阵A完成了LU分解后，方程组 $Ax = b$ 就可以化为 $L(Ux) = b$ ：

$$Ax = L(Ux) = b$$

- 等价于求解两个方程组 $Ly = b$ 和 $Ux = y$
 - 通常来说A的逆是不容易求的。
 - 而将A分解为LU的形式，单位下三角矩阵L的逆和上三角矩阵U的逆是容易求的
 - 因此很容易可以求出y和x，运算量将非常小

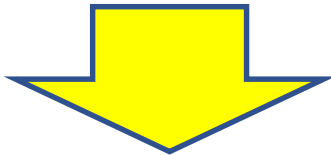
HPL

- HPL 是国际超级计算系统 TOP500 的重要依据。

2020 年 11 月 TOP500 榜单

排名	名称	国家	处理器核数	Rmax(TFlop/s)	Rpeak(TFlop/s)	功率 (千瓦)
1	Fugaku	日本	7,630,848	442,010.0	537,212.0	29,899
2	Summit	美国	2,414,592	148,600.0	200,794.9	10,096
3	Sierra	美国	1,572,480	94,640.0	125,712.0	7,438
4	神威·太湖之光	中国	10,649,600	93,014.6	125,435.9	15,371
5	Selence	美国	555,520	63,460.0	79,215.0	2646
6	天河-2A	中国	4,981,760	61,444.5	100,678.7	18,482

HPCG

- 实际应用中存在大量用偏微分方程建模，稀疏计算和不规则的访存模式
 - 无法用Linpac衡量
- 
- HPCG (high performance conjugate gradient,高性能共轭梯度法):
 - 作为Linpac的补充，是求解稀疏矩阵方程组的一种迭代算法

HPCG

- **HPCG模拟三维热力学运动问题，从而转化为求解离散的三维偏微分方程模型问题**
 - 使用局部对称高斯塞德尔预条件子的预处理共轭梯度法
 - 主要数据为**对称正定稀疏矩阵**

HPCG

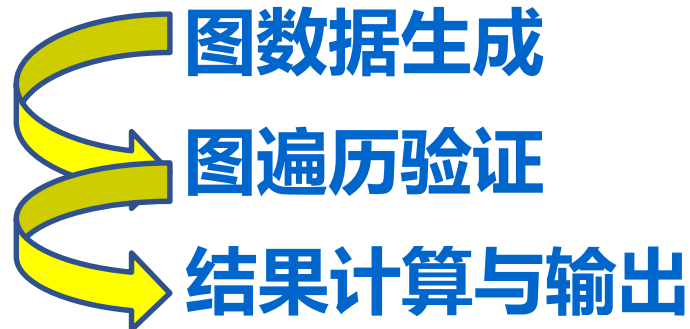
2020 年 11 月 HPCG TOP500 榜单

排名	TOP500 排名	名称	国家	处理器核 数	Rmax(TF lop/s)	HPCG(TFlo p/s)
1	1	Fugaku	日本	7,630,848	442,010.0	16004.50
2	2	Summit	美国	2,414,592	148,600.0	2925.75
3	3	Sierra	美国	1,572,480	94,640.0	1795.67
4	5	Selence	美国	555,520	63,460.0	1622.51
5	7	JUWELS	德国	449,280	44,120.0	1275.36
6	10	Damma m-7	沙特阿拉伯	672,520	22,400.0	881.40

Graph500

- Graph500用来衡量计算机在处理**数据密集型应用**的能力。
 - 利用**图遍历**中广度优先搜索（BFS）或单源点最短路径（SSSP）算法。

- Graph500评测流程：



- 评测指标：**TEPS**（每秒遍历的边数）

Graph500

- Graph500 旨在提高对复杂数据的认识，它**强调系统的通信子系统**，而不再专注于双精度浮点数。

2020 年 11 月 Graph500榜单

排名	名称	国家	处理器核数	节点数	规模	GTEPS
1	Fugaku	日本	7,630,848	158976	41	102956
2	神威·太湖之光	中国	10599680	40768	40	23755.7
3	TOKI-SORA	日本	276480	5760	36	10813
4	OLCF Summit	美国	86016	2048	40	7665.7
5	SuperMUC-NG	德国	196608	4096	39	6279.47
6	NERSC Cori-1024 haswell partition	美国	32768	1024	37	2562.16

IO性能评测集

MDTest

- MDTest 用于评估文件系统的**元数据性能**的**基准评测**。
 - 该程序通过在一组机器（通常是 HPC 集群中的计算节点）上**并行创建、统计和删除目录树和文件树**来评测 IO 性能。
- **评测指标：每秒操作数（OP/秒）**

IOR

- IOR可用于使用**多种IO 接口** (例如 POSIX, MPI-IO, HDF5 等) 和**访问模式**来测试**并行文件系统的性能**。
 - 通过接收参数, 在客户端上产生特定的工作负载从而测试系统的 IO 性能并输出评测结果。
 - 评测结果中**带宽**是通过传输的数据量除以停止时间戳与开始时间的差值得到。

IO500

- 由于测试方法、工具、参数甚至测试步骤的先后顺序不同，不同厂商发布的 IO 性能测试结果具有**很大的差异性**
 - IO500 最终分数是**IOR分数**和**MDTest分数**的几何平均值
 - IO500可以对高性能存储系统进行标准的测试和比较

IO500

2020 年 11 月 Graph500榜单

排名	名称	文件系统	得分	BW(GIB/S)	MD(KIOP/S)
1	Pengcheng Cloudbrain-II on Atlas 900	MadFS	7043.99	1475.75	33622.19
2	Wolf	DAOS	1792.98	371.67	8649.57
3	WekaIO on AWS	WekaIO Matrix	938.95	174.74	5045.33
4	Frontera	DAOS	763.80	78.31	7449.56
5	Presque	DAOS	537.31	108.19	2668.57
6	Tianhe-2E	Lustre	453.68	209.43	982.78

网络性能评测集

IMB

- 用于评估HPC集群在不同消息粒度下**节点间点对点、全局通信的效率**
 - **点对点通信**：评测测试的是两个进程间的消息传递，包括了 Ping-Pong 和 Ping-Ping 测试
 - **全局通信**：评测测试的是全局负载下消息的收发效率，包括 Sendrecv 和 Exchange测试

OSU Benchmark

- 程序生成不同规模的数据，并执行各种不同模式的**MPI通信**，测试各种通信模式的**带宽和延迟**
 - 由Ohio State University提供
 - 分为**点对点通信**和**组通信**两种形式。

能耗评测集

Green500

- **Green500提供高性能计算机的能耗排名。**
 - 评测指标：使用 **PPW** (performance per watt) **每瓦特**性能作为其指标来对能源效率进行排名。

$$\text{PPW} = \frac{\text{Performance}}{\text{Power}}$$

Green500

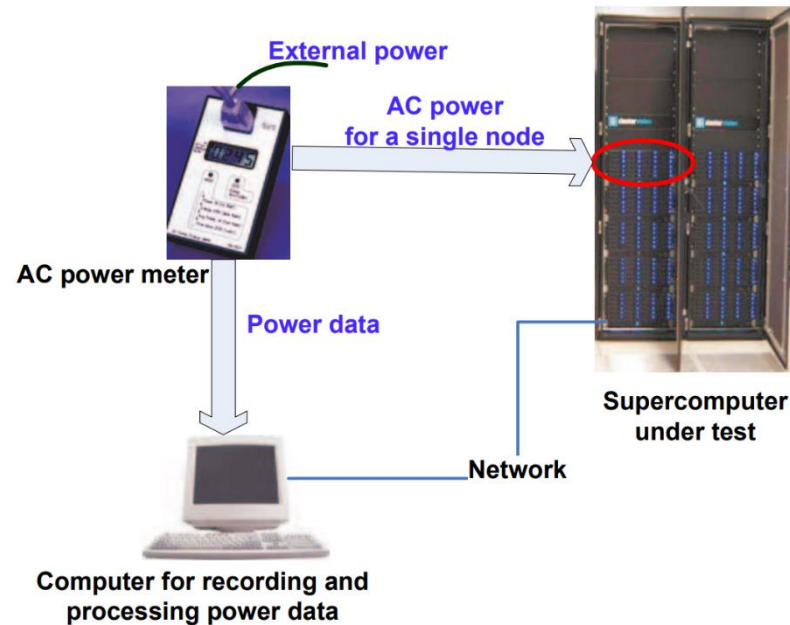


图2-7 超级计算机上单个单元的功率测量图

- $\text{GFLOPS Per Watt} = \frac{R_{\max}(\text{in GFLOPS})}{\bar{P}(R_{\max})(\text{in Watt})}$
- $\bar{P}(R_{\max}) = N \cdot \bar{P}_{\text{unit}}(R_{\max})$

Green500

2020 年 11 月 Green500榜单

排名	名称	国家	处理器核数	Rmax(Tflop/s)	功率 (千瓦)	PPW (GFLOPS /Watt)
1	DGX SuperPOD	美国	19,840	2,356.0	90	26.195
2	MN-3	日本	1664	1,652.9	65	26.039
3	JUWELS	德国	449,280	44,120.0	1,764	25.008
4	Spartan2	法国	23,040	2,566.0	106	24.262
5	Selence	美国	555,520	63,460.0	2646	23.983
6	A64FX prototype	日本	36,864	1,999.5	118	16.876

应用评测集

Miniapplication

- 评估超级计算机对于**动态应用程序**的性能
 - 是真实应用程序的更小版本
- Mantevo 集包含大量应用领域的**开源 Miniapplication**
 - 例如：MiniAMRda, MiniFE, MiniGhost, MiniMD...

戈登贝尔奖

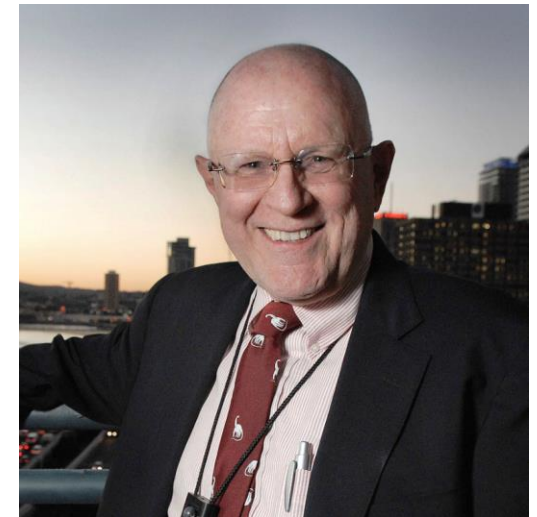
- **戈登贝尔奖 (ACM Gordon Bell Prize) : 超算界诺贝尔奖**
 - 美国计算机协会设立于1987年，每年颁发
 - 是一种超级电脑应用软件设计奖，奖金象征性1万美元，由Gordon Bell提供
 - 通常会由**当年前500排行名列前茅的超级电脑系统之上所跑的的应用软件**获得
- **奖项通常分为：**
 - 最高性能奖 (Peak Performance)
 - 最高性价比奖 (Price/Performance)
 - 特别奖 (Special Achievement)

Gordon Bell

- **戈登·贝尔 (Gordon Bell)**

- 1934年8月19日出生于美国密苏里州的柯克斯维尔
- 1956年，获MIT电子工程学士学位。
- 1957年，获MIT电子工程硕士学位。
- 1960年-1983年，在DEC任副总裁，负责研发。
- 1983年7月，合伙创办核心(Encore)计算机公司。
- 1986年，全美科学基金会(NSF)计算机及信息科学和工程助理主任。
- 1991年-1995年，担任微软公司顾问。
- 1993年，获WPI名誉博士学位。
- 1995年8月-至今，微软湾区研究中心高级研究员

- **作为DEC的技术灵魂，构思、设计和主持开发的超级计算机PDP-4，PDP-5，PDP-6，PDP-8，PDP-10及PDP-11**



中国第一次获得戈登贝尔奖

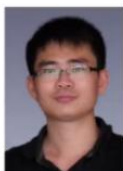
- 2016年中国第一次获得戈登贝尔奖：
 - 神威·太湖之光超级电脑上的“**全球大气非静力云分辨模拟**”应用软件得奖，使用了超过一千万个核来完成一次气候的数值仿真
 - 从2007年开始相关研究，到2011年冲奖团队正式成立，到2012年第一次尝试冲奖，再到2016年正式获奖



2020年戈登贝尔奖

- 2020年11月该奖项颁给了一支由中美科学家组成的研究团队，
 - 因“结合分子建模、机器学习和高性能计算相关方法，将具有从头算精度的分子动力学模拟的极限提升至**1亿个原子规模**”获奖
 - 获奖的8人团队中，有7张中国面孔
 - 该团队将这一工作在美国超算Summit机器上全机运行，模拟分别实现了双精度**91PFlops**、混合单精度**162PFlops**和混合半精度275PFlops的峰值性能

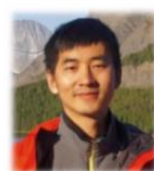
The Team



Weile Jia @ UCB



Han Wang @ IAPCM



Mohan Chen @ PKU



Denghui Lu @ PKU



Lin Lin @ UCB



Weinan E @ Princeton



Roberto Car @ Princeton



Linfeng Zhang @ Princeton

小结

- **基准评测是从基准测试程序和测试规范的角度评价和预测系统的性能。**
- **高性能计算领域上的基准评测通常包括：**
 - 计算性能的评测（Linpack、HPCG、Graph500）
 - IO 性能的评测（MDTest、IOR 和 IO500）
 - 网络性能的评测（IMB 和 OSU Benchmark）
 - 能耗方面的评测（Green500）
 - 应用评测（Miniapplication、戈登·贝尔奖）

参考文献

- Dongarra J J, Luszczek P, Petit A. The linpack benchmark: past, present and future. Concurrency and Computation: practice and experience, 2003, 15(9):803–820.
- Dongarra J J, Moler C B, Bunch J R, et al. LINPACK users' guide. SIAM, 1979.
- Dongarra J, Heroux M A, Luszczek P. High-performance conjugate-gradient benchmark: A new metric for ranking high-performance computing systems. The International Journal of High Performance Computing Applications, 2016, 30(1):3–10.
- Murphy R C, Wheeler K B, Barrett B W, et al. Introducing the graph 500. Cray Users Group (CUG), 2010, 19:45–74.
- Kunkel J, Lofstead G F, Bent J. The virtual institute for i/o and the io-500. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2017.
- Bell G, Bailey D H, Dongarra J, et al. A look back on 30 years of the gordon bell prize. The International Journal of High Performance Computing Applications, 2017, 31(6):469–484.