科学计算环境初步

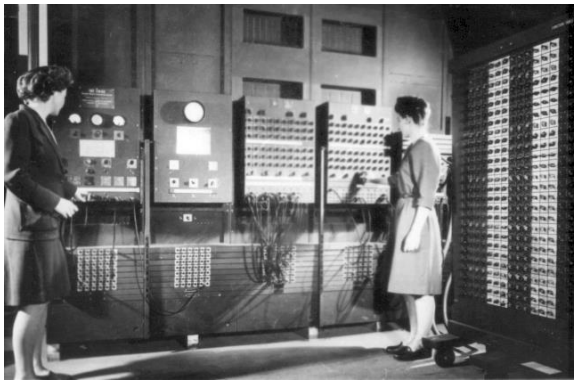
李震宇 (USTC)

Outline

- 硬件与操作系统
- 常用软件
  - 文本编辑：vim、emacs
  - 图形图像：gimp、gnuplot
- 作业管理系统
- 编程与软件
  - 编程语言 (脚本：bash、python；编译：Fortran)
  - 编译工具 (make)
  - 数学库
- 软件包
  - VASP编译

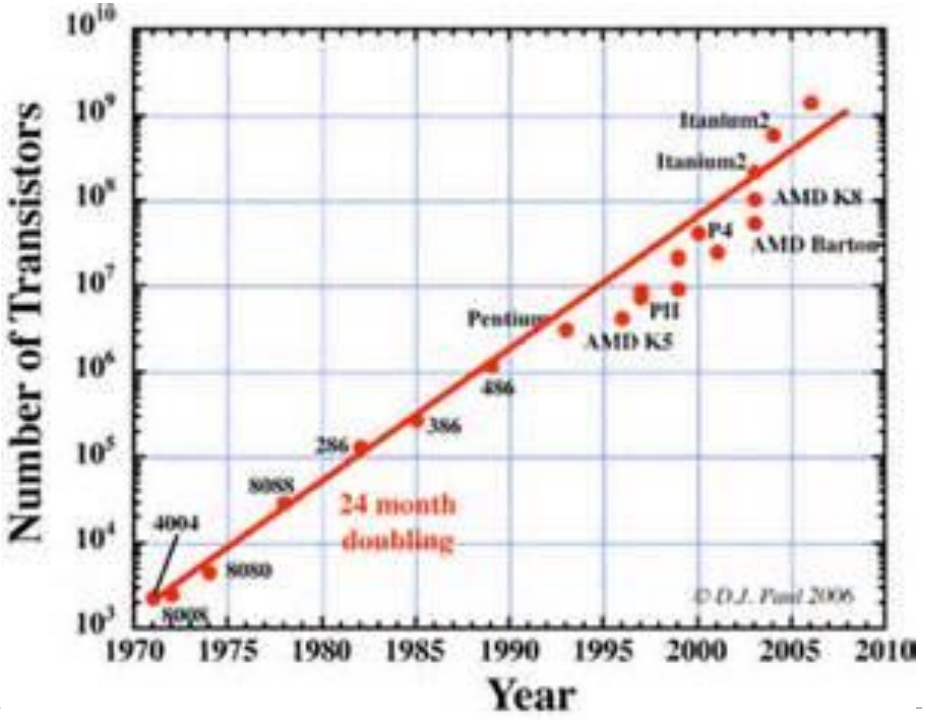
计算机基本组成

- A computer is a general purpose device that can be programmed to carry out a finite set of arithmetic or logical operations (from wikipedia).
- central processing unit (CPU)
- memory (RAM and ROM)
- peripheral devices



CPU与摩尔定律

- 复杂指令集CISC (x86, Pentium、Core、Xeon)
- 精简指令集RISC (PowerPC、SPARC、MIPS、Alpha)
- 显式并行指令计算EPIC (ia64, Itanium)



超级计算机

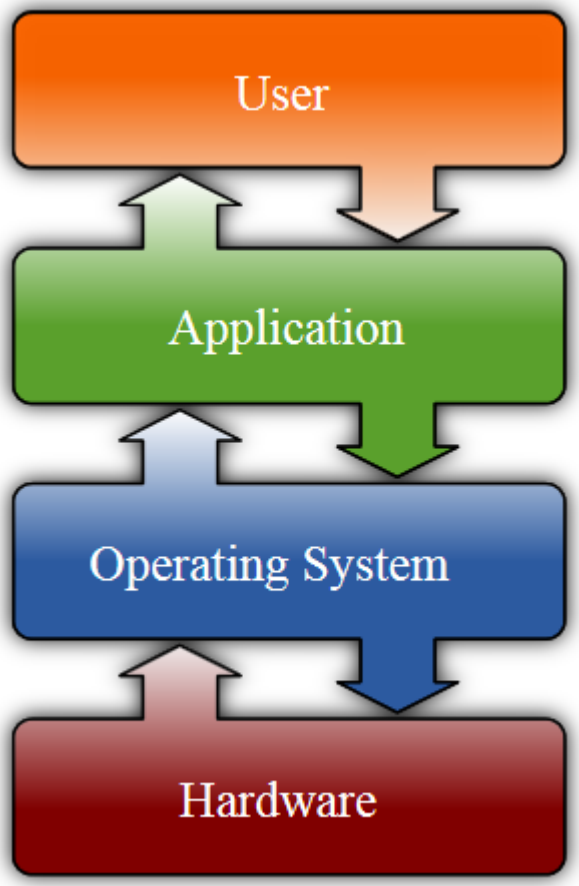
- 从大型机到集群系统
- 在科学计算中应用广泛
- 能效，FLOPS/Watt

Blue Gene/P

Year	Supercomputer	Peak speed	Location
2008	IBM Roadrunner	1.026 PFLOPS	Los Alamos, USA
		1.105 PFLOPS	
2009	Cray Jaguar	1.759 PFLOPS	Oak Ridge, USA
2010	Tianhe-IA	2.566 PFLOPS	Tianjin, China
2011	Fujitsu K computer	10.51 PFLOPS	Kobe, Japan
2012	Cray Titan	17.59 PFLOPS	Oak Ridge, USA
2013	NUDT Tianhe-2	33.86 PFLOPS	Guangzhou, China

操作系统

- 进程管理 (Processing management)
- 内存管理 (Memory management)
- 文件系统 (File system)
- 网络通信 (Networking)
- 安全机制 (Security)
- 用户界面 (User interface)
- 驱动程序 (Device drivers)



常见操作系统

- Windows

- Mac OS
  - Mach 内核：System vX.X, Mac OS 8, Mac OS 9
  - Darwin 内核(BSD UNIX based)：Mac OS X
- UNIX
  - AIX, HP/UX, Solaris, FreeBSD, OpenBSD, NetBSD, …
- GNU/Linux
  - GNU's Not UNIX (Richard Stallman)
  - GNU GPL 版权许可下的内核(Linus Torvalds)
  - Debian, Ubuntu, Fedora, Red hat, SUSE, …

Tux

Linux操作系统常用命令

- Bash (Bourne-Again SHell)
- 文件：pwd, cd, ls, find, cp, scp, ln, mkdir, mv, rm, touch, tar, chgrp, chmod, chown, df, du, mount
- 文本：cat, less, more, tail, diff, grep, awk, sed, vi
- 作业：bg, fg, kill, nohup, sudo, source, ps, top, whereis, which, man
- 用户：passwd, export, quota, su, who
- 其他：expr, set, ping, traceroute, ssh
- 输入输出重定向与管道：>, >>, n>&m, <, |

XWindow系统

- X窗口系统是一种以位图方式显示的软件窗口系统。
- 采用C/S的架构模型，具有“网络透明性”。
  - export DISPLAY=host:0
- XFree86 → X.org (X11R7.7)
- 桌面环境
  - KDE(Qt)
  - GNOME(gtk+)
- Windows下的X server
  - Cygwin/X、Xming
- X程序
  - xterm

常见网络互联方案

- 局域网：Ethernet over twisted pair (10BASE-T, 100BASE-TX, and 1000BASE-T)
- 节点间互联
  - Ethernet (10GBASE-X, 40GBASE-X, 100GBASE-X)
  - Infiniband (aggregate links)

	SDR	DDR	QDR	FDR-10	FDR	EDR
1X	2 Gb/s	4 Gb/s	8 Gb/s	10.3125 Gb/s	13.64 Gb/s	25 Gb/s
4X	8 Gb/s	16 Gb/s	32 Gb/s	41.25 Gb/s	54.54 Gb/s	100 Gb/s
12X	24 Gb/s	48 Gb/s	96 Gb/s	123.75 Gb/s	163.64 Gb/s	300 Gb/s

*actual data rate, not signaling rate
**run *ibhosts* and *ibstat* for info

  - Myrinet (Myrinet 2000, Myri 10G)

网络服务：telnet/rsh/ssh

- Telnet：TCP/IP协议族的一项，是Internet远端登录服务的标准协议和主要方式，常使用23号端口。
- rsh：以另外的用户身份甚至在另外的计算机上远程使用shell
- SSH(Secure Shell)：创建在应用层和传输层基础上的一项安全协议，为计算机上的Shell提供安全高效的传输和使用环境，常使用22号端口。
  - OpenSSH为其开放源代码实现。
  - ssh-keygen -t rsa
  - cp ~/.ssh/id_rsa.pub ~/.ssh/authorized_keys
  - Windows客户端程序：putty

网络服务：ftp/scp

- FTP (File Transfer Protocol)：TCP网络上文件传输协议。
  - 匿名服务：anonymous
  - 主动模式：客户端从一个任意的非特权端口N (N>1024)连接到服务器的命令端口(21)，然后开始监听N+1端口，并发送FTP命令“port N+1”。服务器从数据端口(20)连接到客户端N+1端口。
  - 被动模式：客户端打开任意两个端口(N > 1024和N+1)。从端口N连接服务器的端口21，提交PASV命令。服务器开启数据端口(P > 1024)，并发送PORT P命令给客户端。客户端发起从N+1到P的连接。
- SCP (Secure copy)：基于ssh协议的安全文件传输。
  - Windows客户端程序：WinSCP

vim文本编辑

- 从vi发展出来的一个文本编辑器(Vi IMproved)
- VIM许可证协议，兼容GPL，包含一些慈善条款
- 模式
  - 普通模式(启动默认模式)
  - 插入模式(按a或i进入)
  - 可视模式(操纵文本块)
  - 命令行模式(“:”键执行命令，“/”或“?”键搜索)
- 主要优点
  - 简洁、高效、普及
  - 让“QWERTY”键盘用户将手指保持在默认键位上，使编辑时手指移动更少。

emacs文本编辑

- 源自Editor MACroS，宏编辑器
- GNU Emacs和XEmacs
- 使用lisp作为扩展语言
- 模式：基于文本类型
- 主要优点：多功能、高可扩展性、高可定制性
- 编辑器之战
  - Emacs是“一个伟大的操作系统，只缺个体面的编辑器”
  - “Escape Meta Alt Control Shift”或者“EMACS Makes Any Computer Slow”
  - vi有两个模式：“不停地哔哔叫”和“搞砸一切”。

Gimp图像处理

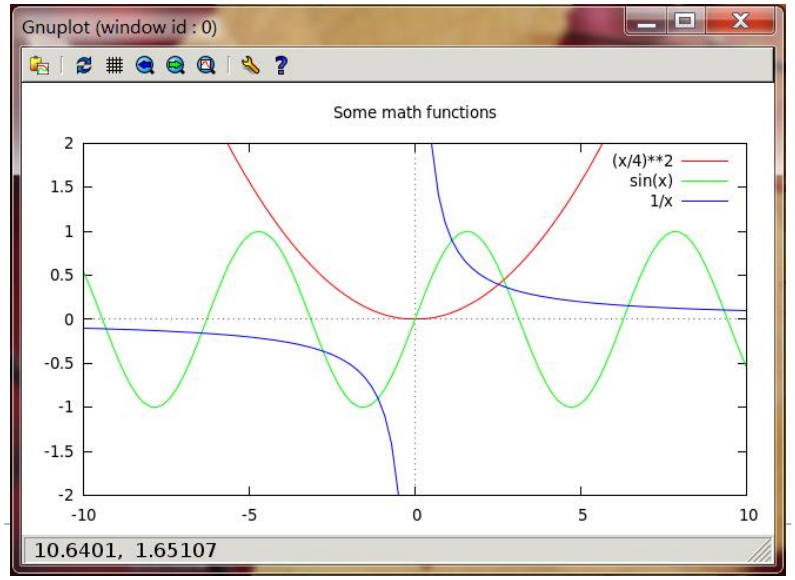
- GNU Image Manipulation Program：一款跨平台开放源代码图像处理自由软件。
- 支持GNU/Linux、MS Windows、Mac OS X等平台。
- 能够实现多种图像处理方面的要求，包括照片润饰、图像合成和创建图像

- <<写给大家看的设计书>> Robin Williams
  - 对齐、亲密性、重复、对比

gnuplot绘图

- 一个基于命令行的产生函数、数据和拟合的二维/三维图的软件(不是基于GPL发布)。

```
gnuplot> set title "Some math functions"
gnuplot> set xrange [-10:10]
gnuplot> set yrange [-2:2]
gnuplot> set zeroaxis
gnuplot> plot (x/4)**2, sin(x), 1/x
```



作业管理系统

- 作业排队提交、优先级管理、动态负载平衡、日志与统计分析。。。
- Portable Batch System (PBS)
  - OpenPBS、Torque、PBS Pro
  - qsub -N name -q queue -l nodes=4:ppn=12 job.pbs
  - qstat, qdel, pbsnodes
  - pbs_sched → Maui
- IBM Platform LSF (Load Sharing Facility)
  - bsub –n 8 –q queue < job.lsf
  - bjobs, bhist, bkill, bqueues, bhosts

编程语言

- 脚本语言
  - 解释执行，开发效率优先
  - 自动批处理、Web编程、文字处理、“胶水”语言
  - bash, lisp, javascript, awk, sed, perl, python…
- 编译语言
  - 编写、编译、链接、运行 (edit-compile-link-run)
  - 运行效率优先
  - C, Basic, Pascal, Fortran
- 数据结构和算法
- 面向对象程序设计(Object-oriented programming)
  - 类、对象、方法、封装、继承、多态(Polymorphism)

Python语言

- 哲学：优雅、明确、简单，反对“总是有多种方法来做同一件事情”(perl)
- 使用缩进(4个空格)来定义语句块
- 完全面向对象，函数、模块、数字、字符串都是对象。
- Python 2.7 → Python 3.0
- Numpy、Scipy
- matplotlib、pylab (an alternative to MATLAB)

C语言

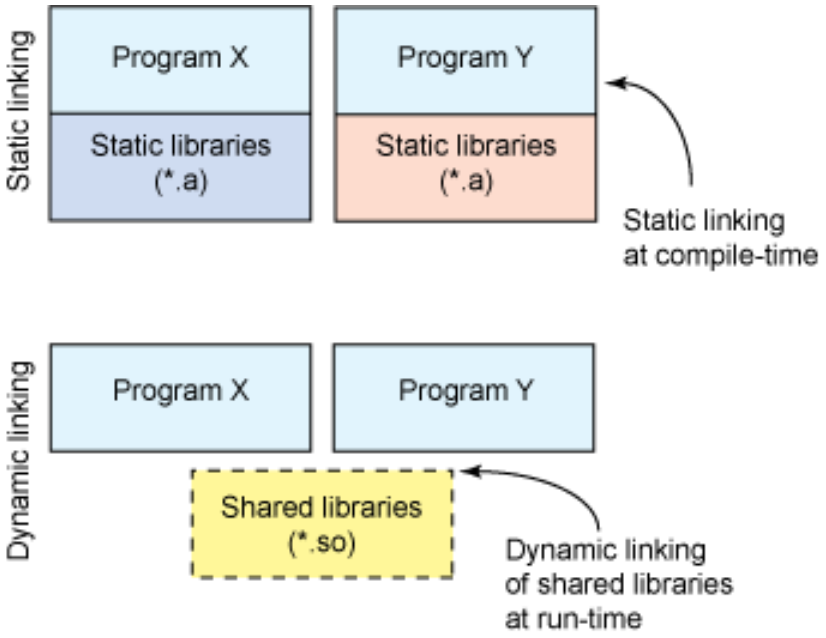
- 最流行的计算机编程语言
- 接近硬件底层，已取代汇编语言用来编写Linux内核
- 编译预处理(C preprocessor)
  - #include <stdio.h>
  - #define PI 3.14159
  - 条件编译：#if, #ifdef, #ifndef, #else, #elif, #endif

GCC

- GNU Compiler Collection
- 前端：产生Abstract syntax tree
  - 支持Ada, C, C++, Fortran, Java, Objective C, Go…
- 中间层：优化
- 后端：产生机器代码
  - 支持Alpha, ia32(x86), x86-64, ia64, MIPS, SPARC…

Libraries

- 库是用于开发软件的子程序集合
- 静态链接：在链接时将库的内容加入到可执行程序中
  - /path/libmylib.a
  - -L/path -lmylib
- 动态链接：在可执行文件装载时或运行时，由操作系统的装载程序加载库
  - /path/libmylib.so.version
  - ln –sf … /path.libmylib.so
  - ldd vasp
  - ldconfig -p
    - /etc/ld.so.config
    - LD_LIBRARY_PATH



GNU binutils

- as:汇编器
- ld:连接器
- ar:维护静态库文件，可做创建、修改和取出的操作
  - ar rcs libmylib.a mylib.o
- nm:显示目标文件内的符号
- objdump:显示目标文件的相关信息，亦可反汇编
- ranlib:产生静态库的索引

GNU Make

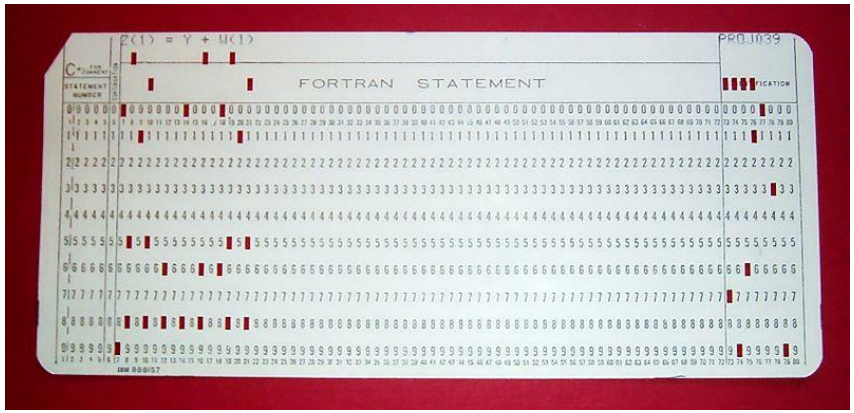
- 通过读取“makefile”的文件，自动化建构软件
- 在非UNIX环境中，通常使用IDE
- makefile语法
  - MACRO = definition
  - target [target ...]: [component ...]

    -$(RM) target

    @echo “starting…”
  - 后缀规则 from.to: ; COMMAND… (. SUFFIXES: .from .to)
  - 模式规则 %.o: %.c ; COMMAND…
  - $@ (目标文件), $%(目标静态库成员), $<(第一个依赖文件), $*(模式规则中的茎), $?(比目标文件更新的依赖文件), $^(所有依赖文件), $+(同上，含重复)

Fortran语言

- **For**mula **Tran**slation，第一个高级编程语言
- 广泛应用于数值计算，积累大量优秀源程序
- 矩阵元素列优先
- FORTRAN 77→Fortran 90/95→Fortran 2003/2008
  - 固定列格式与自由格式
  - OOP支持
- 编译器
  - PGI, Absoft, Intel
  - gfortran from gcc, g95



Intel Fortran Compiler

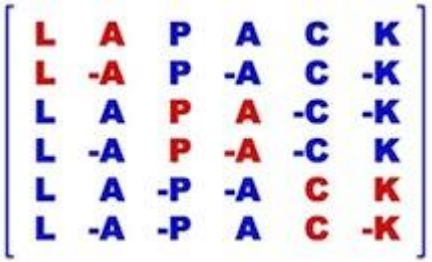
- 下载Intel® Composer XE for Linux非商业版本
- 运行install.sh安装
  - 输入email收到的序列号
  - 选择安装如下组件
    - Intel Fortran Compiler XE 13.1 Update 3 on Intel(R) 64
    - Intel Math Kernel Library 11.0 Update 5 on Intel(R) 64
- 设置环境变量
  - source /opt/…/bin/compilervars.sh intel64
  - source /opt/…/mkl/bin/mklvars.sh intel64
- 类似方法可以安装Intel C/C++编译器

BLAS数学库

- Basic Linear Algebra Subprograms
  - Level 1: scalar, vector and vector-vector operations
  - Level 2: matrix-vector operations
  - Level 3: matrix-matrix operations
- 命名规则
  - S (real), D (double precision), C (complex), Z(complex*16)
  - **GE**neral, **G**eneral **B**and, **SY**mmetric, **HE**rmitian, **Tr**iangular
  - DGEMM: $\alpha op(A)op(B)+\beta C$, $op(X)=X, X^T, X^H$; C-m×n
- 不同的实现版本
  - 硬件厂商优化：MKL from intel
  - 第三方优化：GotoBLAS/OpenBLAS, ATLAS (automatically tuned)

LAPACK数学库

- 针对解线性方程组、最小二乘、本征值、奇异值等问题的Fortran90 (from v3.2, 2008)程序库
- 高效运行在基于缓存的体系结构
- driver, computational, and auxiliary routines
- 大量计算基于对BLAS的调用。辅助函数ILAENV可以返回最佳块尺寸。如果返回值为1，不采用块运算，调用Level 2的BLAS函数而不是Level 3

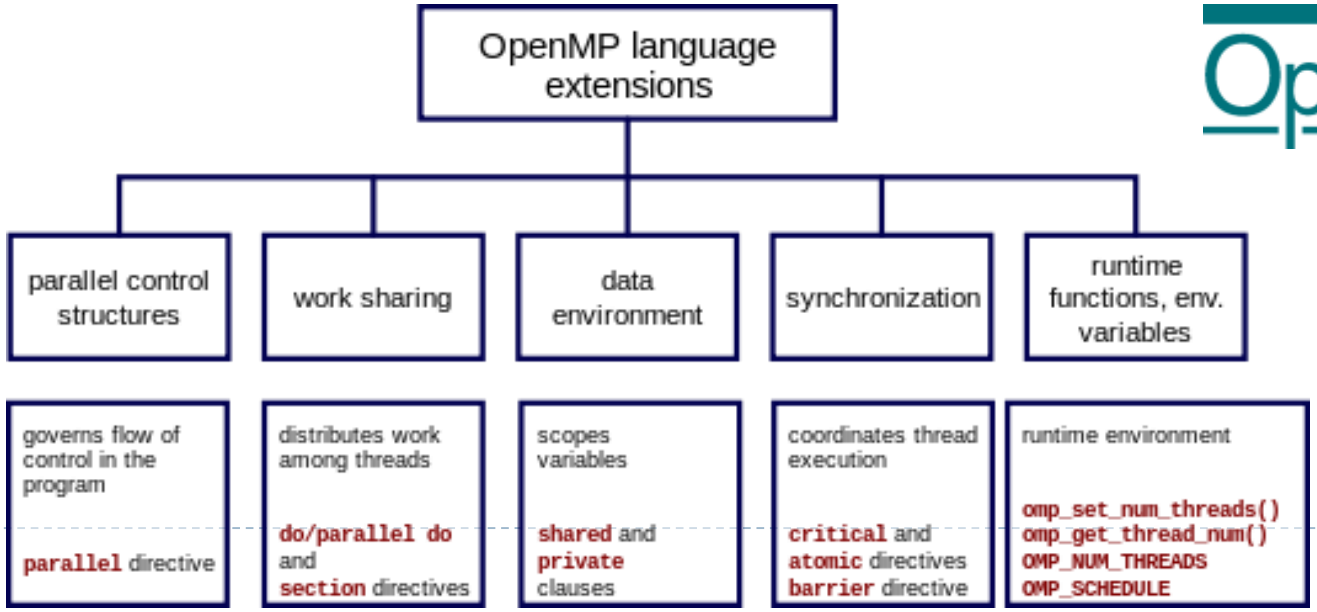


FFTw数学库

- 快速傅立叶变换：$O(n^2) \rightarrow O(n \log n)$
- Fastest Fourier Transform in the West (MIT)
- 用C语言编写，但有Fortran接口
- 高效的原因
  - 包含多种算法，对一个给定的尺寸，可以通过这些算法的组合运用(a plan)获得高的性能
  - 通过代码产生器(genfft in OCaml)产生高度优化的代码
  - 通过显式的分治策略从多级存储硬件架构中获得好处

OpenMP

- Open Multi-Processing，一套支持跨平台共享内存方式的多线程并发的编程API，支持C,C++和Fortran
- 通过在源代码中加入专用的预处理指令，由编译器自动将程序进行并行化
  - #pragma omp <directive> [clause[[,] clause] ...]
- 环境变量OMP_NUM_THREADS



MPI

- Message Passing Interface
- 大多数信息传递接口的实现为库，不需要编译器支持。
- 基本概念
  - Communicator objects connect groups of processes in the MPI session.
  - Point-to-point: MPI_Send
  - Collective: MPI_Bcast, MPI_Reduce
  - Derived datatypes: MPI_INT, MPI_CHAR, MPI_DOUBLE

Open-MPI安装

- 运行configuration进行配置
  - ./configure --prefix=/usr/local/openmpi CC=icc CXX=icpc F77=ifort FC=ifort
- 编译安装
  - make; sudo make install
- 设置环境变量
  - export PATH=/usr/local/openmpi/bin:${PATH}
  - export LD_LIBRARY_PATH=/usr/local/openmpi/lib:${LD_LIBRARY_PATH}
- 查看安装情况
  - ompi_info | grep “MCA btl”

Open-MPI

- Modular Component Architecture (MCA)
  - Framework, Component, Module
- 三层模型
  - OMPI – the MPI layer
    - allocator, btl (Byte Transfer Layer), …
  - ORTE – the run-time layer (Run-Time Environment)
    - errmgr, filem, ras (resource allocation system), …
  - OPAL – the operating system/platform layer (Portable Access Layer)
    - backtrace, crs (checkpoint and restart service), timer, …
- 运行
  - mpirun -np 8 -machinefile hosts --mca btl openib,self,sm vasp.x

Blacs/PBLAS/ScaLAPACK

- Basic Linear Algebra Communication Subprograms， 支持CMMD, MPI, MPL, NX, PVM
- Process grid, contexts, and ID-less communication
- Parallel BLAS: PDGEMM
- ScaLAPACK
  - Initialize the process grid (SL_INIT)
  - Distribute the matrix on the process grid (DESCINIT)
  - Call ScaLAPACK routine
  - Release the process grid (BLACS_GRIDEXIT)

异构化体系并行计算

- Graphics Processing Unit (GPU)
  - Compute Unified Device Architecture (CUDA) by Nvidia
    - C语言编译器(nvcc)
    - CUDA FFT、BLAS库
    - CUDA runtime驱动
  - Open Computing Language (OpenCL)
    - Intel, IBM, AMD, Nvidia, Apple, Sumsang, ARM…
- Intel集成众核(Many Integrated Core, MIC)架构
  - 至强融核(Xeon Phi)协处理器
  - 更小的内核和硬件线程，以及更宽的矢量单元
  - 仍然基于x86架构，应用软件移植方便。通过重核来实现粗粒度并行，通过指令级并行实现细粒度并行。

VASP简介

- Vienna Ab-initio Simulation Package
- Total energy calculation
- Geometry optimization
- Molecular dynamics
- Transition state locatior
- Lattice dynamics
- Chemical shift
- Core level shift
- Excited state
- …

VASP编译

- make vasp.5.lib
- make vasp.5.3, based on makefile.linux_ifc_P4 shipped with VASP
  - **sequential:** ifort, MKL(blas, lapack, fftw)
  - **parallel:** mpif90(ifort), MKL(scalapack)
  - **G-only:** -DwNGZhalf

- 上机实践：编译串行单k点版本VASP程序，测试CACHE等编译参数对计算速度的影响