

# 计算化学软件使用经验

张宝花  
中国科学院超级计算中心  
2012.11.09

# 提纲

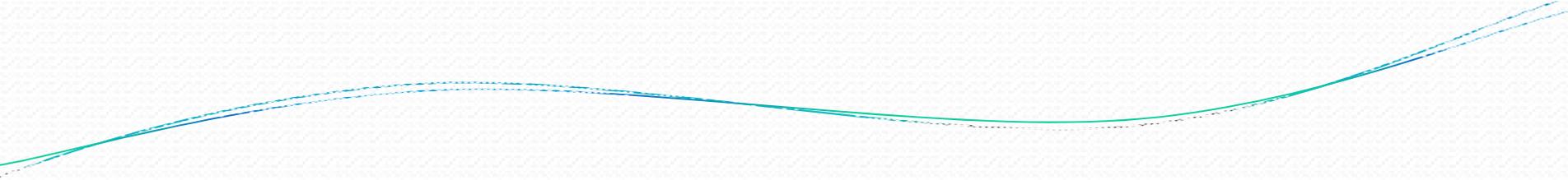
- 量子化学软件
- 分子模拟软件
- Gaussian软件
- VASP软件
- MS软件——DMOL<sub>3</sub>和CASTEP
- 总结

# 量子化学软件

- 量子化学——应用量子化学的基本原理和方法研究化学问题。
  - 基于第一性原理，主要处理体系电子效应相关
  - 分子轨道理论、价键理论、密度泛函理论
  - 研究分子结构、化学反应、分子性质
  - 常见量子化学软件：Gaussian、NWChem、VASP、Q-Chem、ADF、Turbomole、Molpro、Gamess等。
  - 计算特点：对CPU、内存、IO等需求比较大，大量的集合通信

# 分子模拟软件

- 分子模拟——依靠牛顿力学来模拟分子体系的运动
  - 基于经典牛顿方程，不考虑电子效应，应用不同的力场
  - 主要处理大分子体系：动力学行为、宏观性质
  - 考虑不同系综的计算：NPT, NVE, NVT等
  - 常见的分子动力学软件：Gromacs、NAMD、AMBER、LAMMPS、Charmm等。
  - 计算特点：对CPU和IO需求较大，大量的点到点通信。



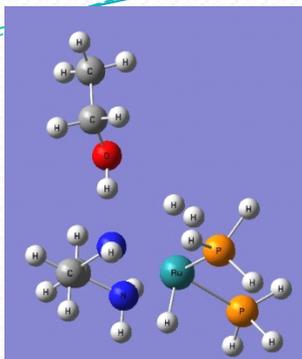
# Gaussian软件

# 概述

- **Gaussian**是一个功能强大的量子化学综合软件包。其可执行程序可在不同型号的大型计算机，超级计算机，工作站和个人计算机上运行，并相应有不同的版本。
- 最新版本：G09 C01
- 特点优势：处理反应机理和过渡态、计算分子性质
- 程序设计时考虑到使用者的需要。所有的标准输入采用自由格式和助记代号，程序自动提供输入数据的合理默认选项，计算结果的输出中含有许多解释性的说明。程序另外提供许多选项指令让有经验的用户更改默认的选项，并提供用户个人程序连接Gaussian 的接口。

# 计算特点

- 并行策略分为节点间openmp并行和节点间的linda并行。
- 对计算能力的需求很大，单个节点配备越多的CPU将提供越强大的计算能力。
- 对内存要求较高（例如CIS，CCSD等计算），节点内存比较大，计算效率高。
- 在计算过程中要写出大量临时文件，I/O要求高
- 网络要求低延迟通信



```
%chk=A_INT1-H2-1.chk
# opt iop(1/8=5) b3lyp/geneCP
```

Title Card Required

0 1

坐标和键长、键角、二面角等参数。。。。。

C N P O H o

6-31+G\*\*

\*\*\*\*

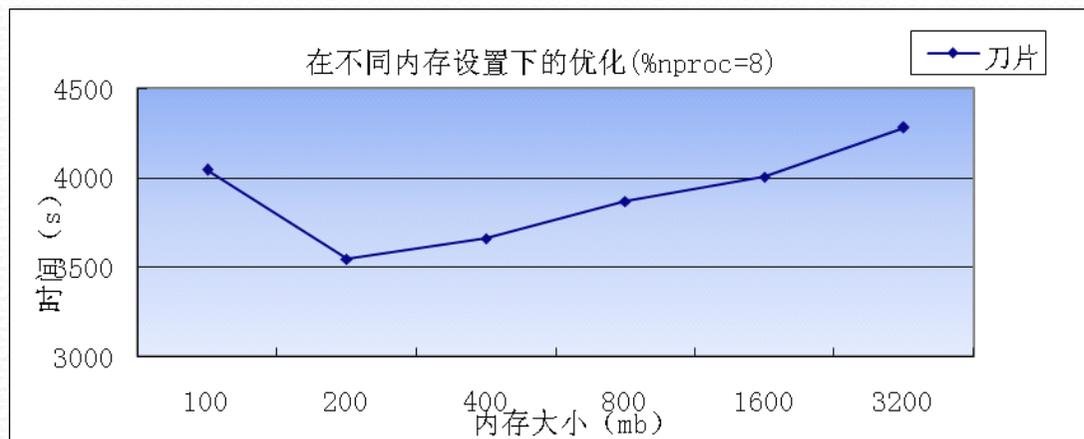
Ru o

lanl2dz

\*\*\*\*

Ru o

lanl2dz



胖节点: 2套SGI Altix 4700, Intel Itanium 2 处理器, NUMA结构, 主频1.66GHz, 内存2.5TB, 188颗处理器, 376个CPU核心

刀片节点: 1140个两颗四核Xeon处理器E4540,主频3.00GHz, 内存32GB

## Gaussian 程序的结构

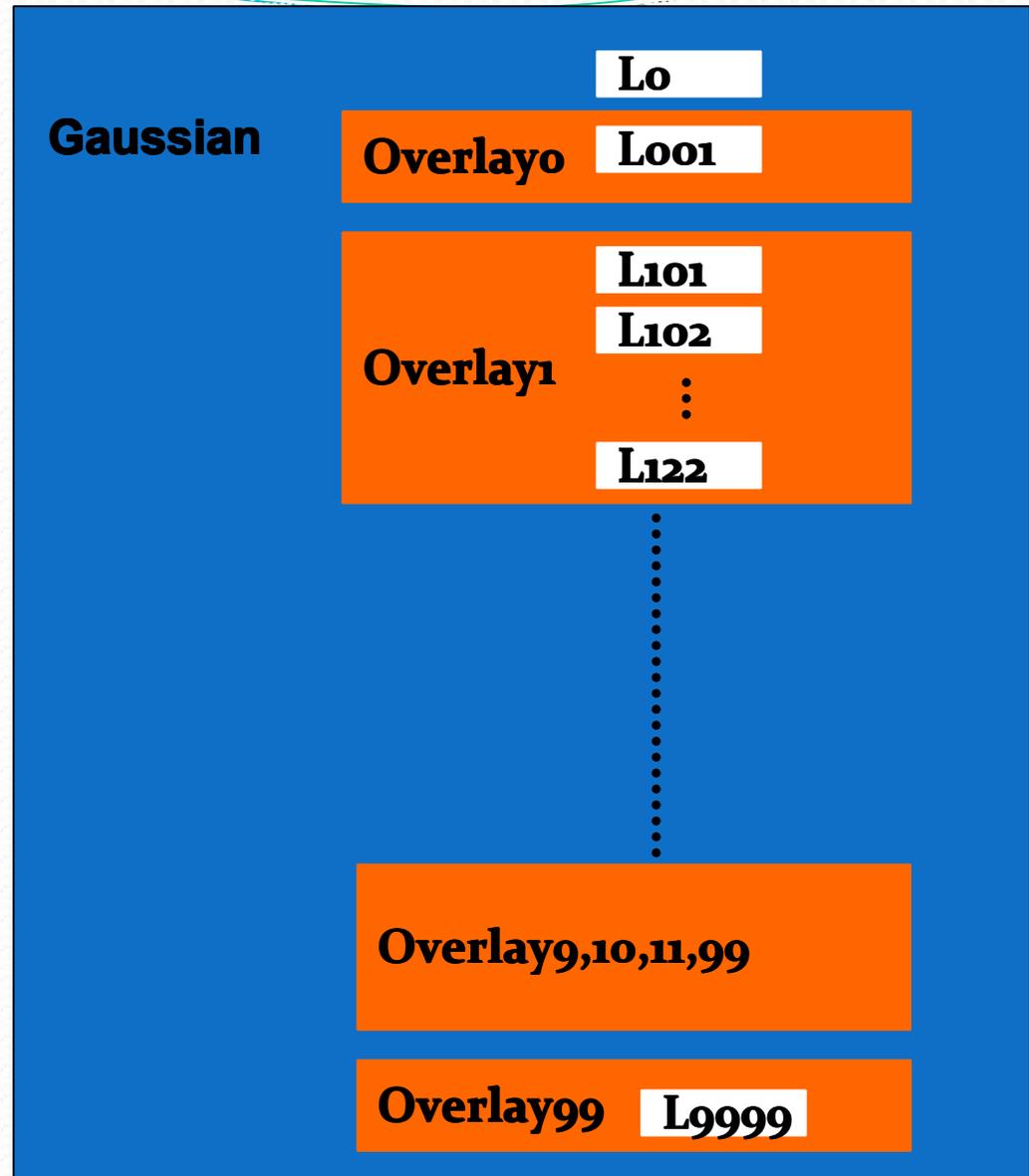
**link0:** 初始化程序，控制**overlay**

**link1:** 读入并处理**Route Section**，建立要执行的**link**列表

**link9999:** 终止计算

其中部分程序可以并行执行，部分程序只能串行执行

几个比较消耗CPU的Link 都实现了并行（如L502， L701、 L913）



## 资源使用

越是精确的计算，就越需要大的资源，就越昂贵。一般的，体系的大小用总基组函数数量(N)表示，也有的方法取决于占据轨道和非占据轨道的基组函数数量(O和V)。

下表列出不同方法理论上的资源消耗( $N^4=N^4$ )

	正常的			实际消耗	
	CPU	内存	磁盘	CPU	磁盘
传统SCF	$N^4$	$N^2$	$N^4$	$N^{3.5}$	$N^{3.5}$
直接SCF	$N^4$	$N^2$	-	$N^{2.7}$	$N^2$
MP2能量					
传统	$ON^4$	$N^2$	$N^4$	$ON^4$	$N^4$
直接	$ON^4$	$OVN$	-	$O^2N^3$	$N^2$
半直接	$ON^4$	$N^2$	$VN^2$	$O^2N^3$	$VN^2$
MP2梯度					
传统	$ON^4$	$N^2$	$N^4$	$ON^4$	$N^4$
直接	$ON^4$	$N^3$	-	$O^2N^3$	$N^2$
半直接	$ON^4$	$N^2$	$N^3$	$O^2N^3$	$N^3$
MP4, QCISD(T)	$O^3V^4$	$N^2$	$N^4$	$O^3V^4$	$N^4$
全CI (CPU)	$((O+V)!/O!V!)^2$				

对于小的体系，其计算成本差别是不大的，只有处理大的体系时，其差别才显现出来。

# 使用经验——计算设置

- Route.Rou修改缺省使用核数，内存大小。
- 节点内采用openmp并行，效率较高。节点间采用Linda并行，效率不高。建议采取节点内并行计算。
- 读写文件指定
  - 优先使用当前作业所在文件系统。
  - 可以跨文件系统存储读写文件（o3不可以）
- 内存并非给得越多越好，需要综合考虑到计算的需要和硬件水平，最有效率的方法是根据作业类型估算所需要内存的大小。可以使用

“freqmem natoms nbasis r|u c|d functions”估算内存。

Functions可以使用%kjob=301来得到

一般的出错都是由于内存不够或RWF文件空间不够引起的。  
级别越高的算法对内存和REF文件空间的要求就越高

# 使用经验——快速得到所需结果

- 一般情况：
  1. 做足homework，构建合理输入文件，选取合适的方法基组。  
(问题规模会随所选取的粒子的种类数目、基组大小以及算法收敛快慢等多种因素的变化而变化)
  2. 坐标形式对于效率没有影响。对称性比较高的分子，用矩阵输入比较容易保持对称性
  3. 控制输出标准，可以简化输出，节省时间
  4. 建立检查点文件
- 关于优化：
  1. 先小基组优化，再大基组优化
  2. 对于大体系，先固定优化一部分结构，再放开优化
  3. 在结构出现震荡时，取出最低能量结构重新计算力场数，往往很快得到优化结果。
  4. 有些体系收敛不要求4个YES，如大体系（势能面比较平缓）只要求前两项收敛即可。

# 使用经验——快速得到所需结果

- 关于频率计算：

1. 在优化的结构基础上进行，并且计算采用的方法和基组必须一致，否则没有意义。
2. 所需内存一般为优化的2倍。

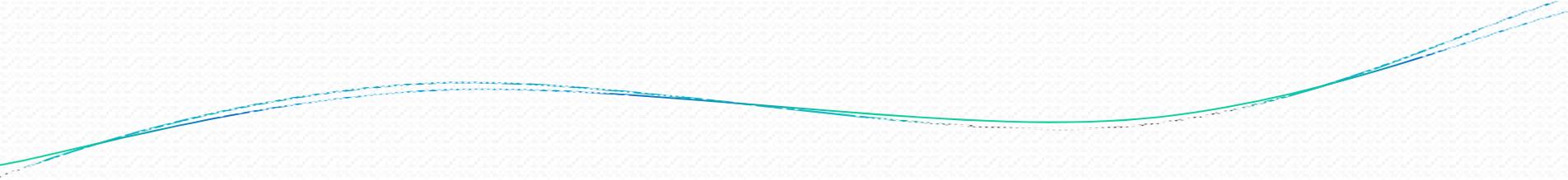
- 关于过渡态：

1. 合理选取方法得到合适的过渡态。(TS、QST<sub>2</sub>、QST<sub>3</sub>)

如TS法：使用势能面扫描法或者控制主反应坐标法找到与过渡态结构相近的初始结构。势能面扫描代价>控制主反应坐标法。

QST<sub>2</sub>法：两个输入结构不必须为优化的稳定点（无需花时间对叠合结构优化），关键在于确保两个结构叠合出来的结构最接近过渡态。

2. 优化得到过渡态结构后，一定要进行IRC验证。IRC验证走不到反应物或产物构型时，可以取结构做优化判断。



# VASP软件

# 概述

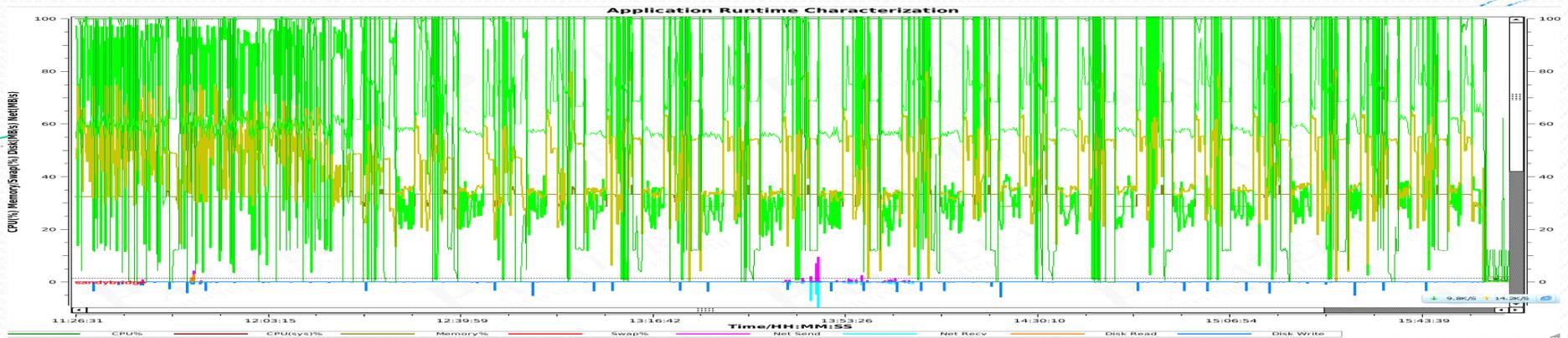
- VASP是Vienna Ab-initio Simulation Package的缩写,它是使用赝势和平面波基组,进行从头量子力学分子动力学计算的软件包
- 最新版本: 5.3.2
- 功能:
  - 采用周期性边界条件(或超原胞模型)处理原子、分子、团簇、纳米线(或管)、薄膜、晶体、准晶和无定性材料,以及表面体系和固体
  - 计算材料的结构参数(键长,键角,晶格常数,原子位置等)和构型
  - 计算材料的状态方程和力学性质(体弹性模量和弹性常数)
  - 计算材料的电子结构(能级、电荷密度分布、能带、电子态密度和ELF)
  - 计算材料的光学性质
  - 计算材料的磁学性质
  - 计算材料的晶格动力学性质(声子谱等)
  - 表面体系的模拟(重构、表面态和STM模拟)
  - 从头分子动力学模拟
  - 计算材料的激发态(GW准粒子修正)
- 计算特点:
  - CPU、内存要求高。特别是内存容量和内存带宽对性能影响较大
  - 要求低通信延迟,其磁盘I/O相对较少

# 软件安装

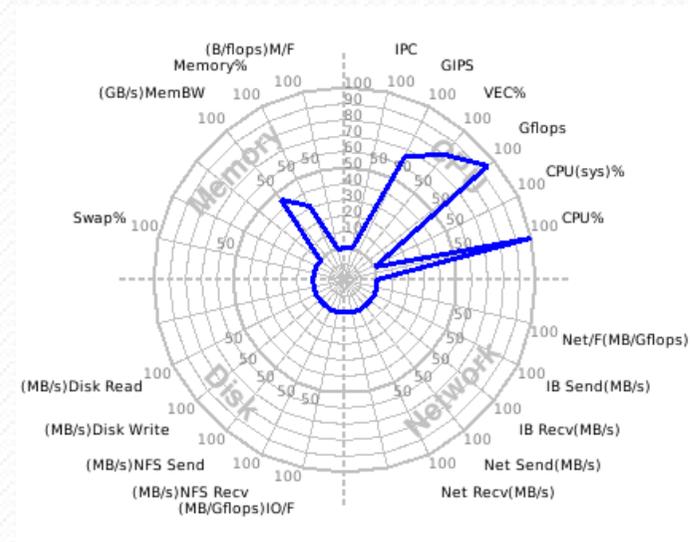
- 编译器, BLAS, fftw, mpi, vasp&vtst
- 步骤:
  - 安装包解压
  - `cd vasp/src`
  - 修改makefile, make 编译出vasp可执行文件
- Trick:
  - 推荐Intel编译器, 采用goto-blas, intelmpi/openmpi获取更好性能。
  - 选择合适的优化指令 (`-O3`, `-xSSEX`等)

# 应用特征分析

组成	选配方案
Workload	CoSb <sub>3</sub> 超胞方钴矿 (Atoms:864, bands:3110)
VASP	5.2
CPU	Intel Xeon处理器
内存	64GB DDR3 1600MHz
OS	Linux



Swap	-nan%	CPU	100.00 %	Net Send	0.00 MB/s
		Real Perf	105.00 Gflops		
		Peak Perf	345.60 Gflops		
		Real/Peak	30.38 %		
Memory	32.0%	X87	0.02 %		
Mem Read	45.24GB/s	VEC	100.00 %		
Mem Write	17.93GB/s	AVX	14.66 %		
		CPI	0.86		
		GIPS	57.73		
		LLCM	99.01 %	Net Recv	0.00 MB/s
Disk Read	0.00 MB/s	Disk Write	0.05 MB/s		
NFS Send	0.00 MB/s	NFS Recv	0.00 MB/s		



CPU%	CPU(sys)%	Memory%	Swap%	Net Send(MB/s)	Net Recv(MB/s)	Disk Read(MB/s)	Disk Write(MB/s)	IB Send(MB/s)	IB Recv(MB/s)	NFS Send(MB/s)	NFS Recv(MB/s)	GPU%	Power(W)	Gflops
99.06	0.22	31.4	0.00	0.03	0.03	0.00	0.09	N/A	N/A	0.04	0.00	N/A	N/A	95.00

MemBW(GB/s)	VEC%	AVX%	X87%	CPI	GIPS	LLCM%	F/M(flops/B)	F/Net(Gflop/ME)	F/IO(Gflop/MB)	F/Disk(Gflop/M)	F/NFS(Gflop/M)
44.18	80.77	31.21	0.03	0.78	65.91	82.98	2.15	10251.25	684.10	1039.99	1999.11

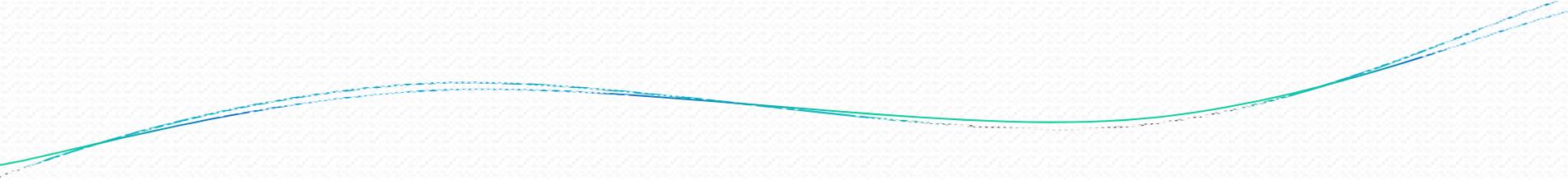
## 使用建议

$\text{Memory} = \text{NKDIM} * \text{NBANDS} * \text{NPLMV} * 16 / 10 (\text{Byte})$

NKDIM: K点数目 NBANDS: 能带数目 NPLMV: 平面波数目

- 输入文件INCAR文件:

**NPAR:** 并行计算band的节点数, 这个选项可能会大幅增加内存的需求。建议设置等于一个节点的核心数



# Material Studio软件包

## 概述

- 能够提供分子模拟、材料设计以及化学信息学和生物信息学全面解决方案的高度模块化的集成产品。
- 最新版本：6.0
- MS软件能使任何研究者达到与世界一流研究部门相一致的材料模拟的能力。模拟的内容包括了催化剂、聚合物、固体及表面、晶体与衍射、化学反应等材料 and 化学研究领域的主要课题。
- 灵活的Client-Server结构。其核心模块Visualizer运行于客户端PC，计算模块（如Discover，DMol<sub>3</sub>，CASTEP等）运行于服务器端
- SCCAS购买Visualizer，**DMOL<sub>3</sub>**，**CASTEP**，ONETEP，Discover

# CASTEP and Dmol3

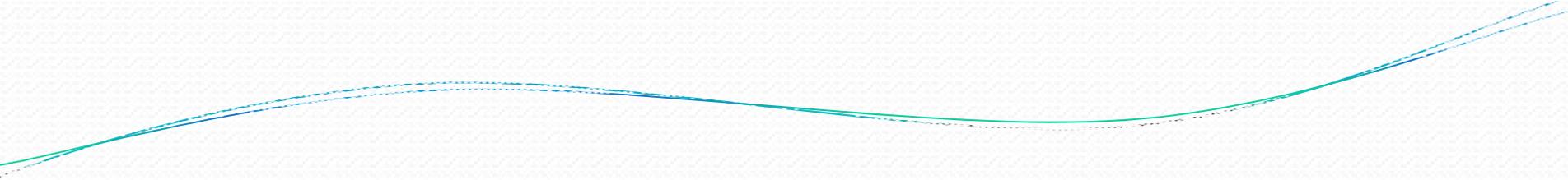
- **CASTEP**: 先进的量子力学程序，广泛应用于陶瓷、半导体、金属等多种材料，可研究：晶体材料的性质（半导体、陶瓷、金属、分子筛等）、表面和表面重构的性质、表面化学、电子结构（能带及态密度）、晶体的光学性质、点缺陷性质（如空位、间隙或取代掺杂）、扩展缺陷（晶粒间界、位错）、体系的三维电荷密度及波函数等。
- **DMOL3**: 独特的密度泛函（DFT）量子力学程序，是唯一的可以模拟气相、溶液、表面及固体等过程及性质的商业化量子力学程序，应用于化学、材料、化工、固体物理等许多领域。可用于研究均相催化、多相催化、分子反应、分子结构等，也可预测溶解度、蒸气压、配分函数、熔解热、混合热等性质。

## 一些经验

- CASTEP与VASP计算特点相近，计算量与平面波数量的3次方成正比。
- DMol3：对内存和I/O要求比较高。计算产生大量的临时文件。需要考虑临时文件的存放问题。

# 总结

- 关于软件安装
  - 细读readme或者install文件，选取合适的编译器，并行库，数学库等。
  - 安装后测试，调优
  - 软件是人写的，都有或多或少的缺陷，排除安装错误，修改源码或者联系软件开发团队。
- 关于软件使用
  - 根据计算体系及特点选取合适的软件，模型的选定和建构非常重要。
  - 进行收敛性测试非常必要。对于人为选择参数，要进行一番测试以显示即便再增加精细度，也不会再改变我们要下的科学判断或结论。



**Thank you for your attention !**