

KSSOLV-GPU 第一性原理计算软件简介

一、软件介绍

KSSOLV-GPU 基于 MATLAB 解释性语言的开源第一性原理计算软件包 KSSOLV 开发的 GPU 加速版本。KSSOLV-GPU 仿照主流开源第一性原理计算软件 Quantum ESPRESSO，使用 MATLAB 语言编写，类似 Python，均为解释型语言，学习成本较低，易于读写，而且通过 GPU 加速，快速求解第一性原理密度泛函理论 KS (Kohn-Sham) 方程。在掌握相关知识的基础上，可以借此习得主流第一性原理计算程序的理论和方法，无需安装，即刻便可运行，KSSOLV 是用 MATLAB 编写的串行代码，受到计算速度的限制，而在 KSSOLV-GPU 中调用 GPU 对 KSSOLV 代码进行加速，实现了个人 PC 端达到主流计算软件的计算速度，可以在课堂上为学生演示量子力学计算。量子力学的教学具有抽象性、复杂性等特点，传统的纯理论教学方式不利于学生的学习和兴趣的激发，借助 MATLAB 平台 KSSOLV-GPU 可以开展项目式和仿真式相结合的教学模式。MATLAB 还提供了丰富的图形表现方法，使得量子力学密度泛函理论第一性原理可以方便地、多样的实现可视化，增强学生对理论知识的理解，提高学习兴趣，保证学习效果，同时增强学生理论到实践的知识转化能力。KSSOLV-GPU 是一套既可以代替主流计算软件科研，又适用于教学和科研的科教一体化计算程序，可以应用到前沿交叉科学研究和教学中去。

1. KSSOLV-GPU 的教学意义

高校物理化学专业的本科生或低年级研究生，在步入科研生涯之前或开展科研之初，有必要学习第一性原理计算的理论以及相关的软件。而在教育意义上，第一性原理计算软件，是综合了数学、物理和化学理论知识的代码实现，可以搭建起理论与实验之间的桥梁，也能帮助学生更好地整理自己的知识体系。然而，如果将第一性原理计算程序搬入课堂，教授学生使用计算软件的门槛高，难度大。

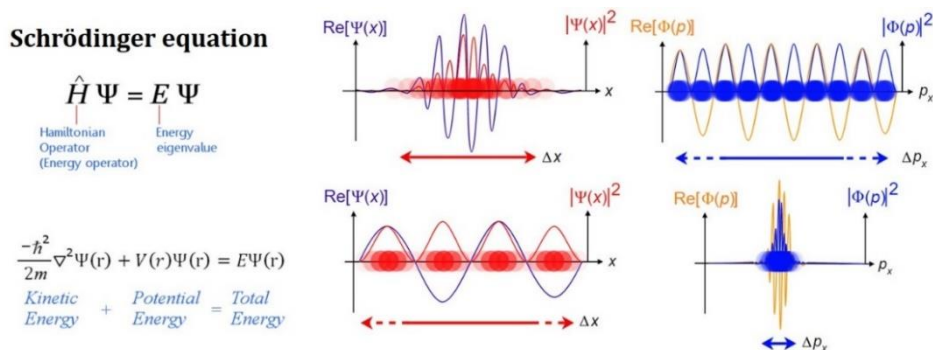


图 1.1 薛定谔方程

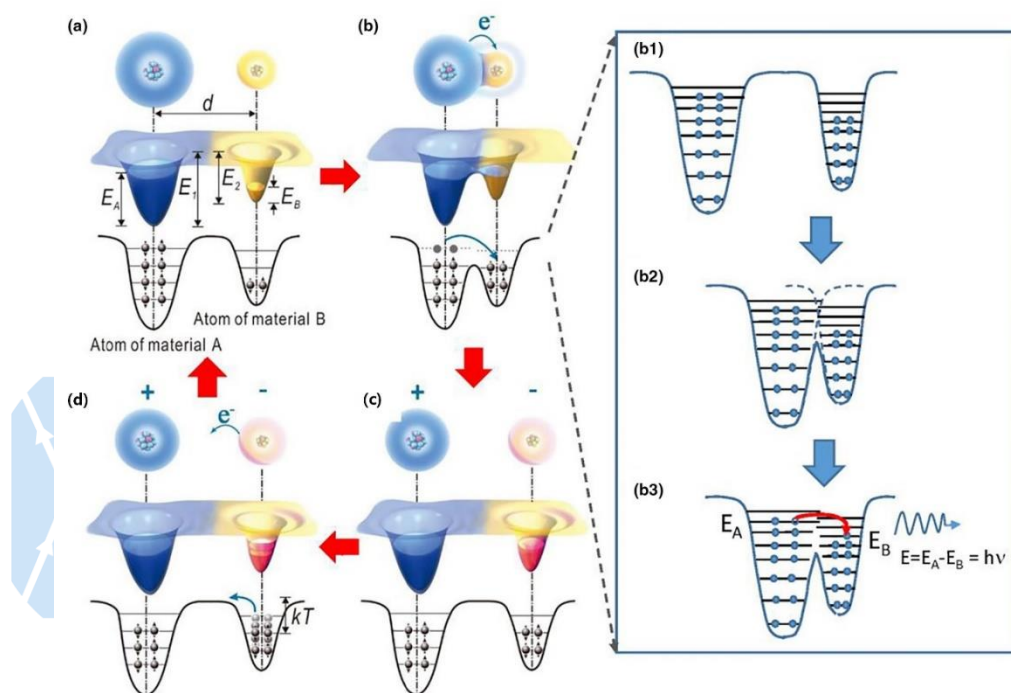


图 1.2 微观电子相互作用模型

目前，用于科研的第一性原理量子力学的主流 DFT 计算程序，大多为欧美国家编写的 Fortran/C++ 代码（VASP、Quantum ESPRESSO 和 SIESTA）。其源代码可读性并不好，且在加入 CPU-MPI 和 GPU-CUDA 并行计算后，更加难以理解，不利于初学者学习和使用。此外，主流的第一性原理计算程序往往需要在 Linux 环境下配置运行，而学生个人电脑中还是 Windows 系统居多。即使安装了 Linux 虚拟机，主流 DFT 计算程序的编译安装过程也非常复杂。而多数物理和化学系的学生并没有很强的计算机基础，大多本科生只学过基础的 C 语言，对 Fortran 和 C++、Linux 系统软件编译过程不甚了解。因此，学习难度较大，门槛较高。目前国内并没有一款适用于教学和科研、可在个人笔记本电脑运行的第一性原理

软件。

KSSOLV-GPU 采用多种先进的对角化算法 (LOBPCG 和 Davidson 等), 在 MATLAB 中实现了标准平面波基组电子结构计算, 并且加入了基于线性响应理论 (Linear-Response) 的 TDDFT 激发态算法以及基于 HSE06 的杂化泛函计算。KSSOLV-GPU 仿照主流计算软件 Quantum ESPRESSO, 使用 MATLAB 语言编写, 类似 Python, 均为解释型语言, 学习成本较低, 更加易于读写。在掌握相关知识的基础上, 更容易习得主流第一性原理计算程序的理论和方法。MATLAB 具有内置编译器, 且在主流的 Windows 系统下就能有很好的表现, 下载后无需安装, 即刻便可运行, 而在配备显卡的笔记本电脑或个人台式机上, KSSOLV-GPU 可以在较短时间内完成中等体系的理论计算和可视化。学生无需掌握复杂的编译原理, 从而降低了学习第一性原理软件的门槛。MATLAB 还提供了丰富的图形表现方法, 使得量子力学密度泛函理论第一性原理可以方便地、多样的实现可视化, 增强学生对理论知识的理解, 提高学习兴趣。

软件包 KSSOLV-GPU 还可以为本科生、研究生量子力学的创新型教学提供有力的平台。量子力学的教学具有抽象性、复杂性等特点, 传统的纯理论教学方式不利于学生的学习和兴趣的激发, 借助 KSSOLV-GPU 平台可以开展项目式和仿真式相结合的教学模式。指导学生以具体的课题实现为目标, 通过数值仿真和系统仿真理解理论知识, 这样不仅丰富教学的形式, 更重要的是可以培养学生的学习兴趣, 保证学习效果, 同时增强学生理论到实践的知识转化能力。此外, MATLAB 平台还有强大的并行计算能力, 而量子力学中又涉及到大量的计算问题, 所以借助 KSSOLV-GPU 平台, 还可以培养学生的计算思维和并行计算能力, 为以后专业课学习和科研能力的培养打下扎实的计算基础。

2. KSSOLV-GPU 的科研意义

除了教学以外, 我们还希望 KSSOLV-GPU 能成为研究人员可行可用的第一性原理计算工具。KSSOLV 是用 MATLAB 编写的串行代码, 受到计算速度的限制, 在实际科研中, 效率会大打折扣。而 KSSOLV-GPU 能够在个人 PC 端达到主流计算软件的计算速度, 提高研究效率。得益于 MATLAB 的开发环境, 研究人员可以直接使用 NVIDIA® GPU 来加速 AI、深度学习或其他计算密集型分析。使用

MATLAB 及其 Parallel Computing Toolbox™, 可以直接在 MATLAB 中调用 NVIDIA GPU, 有 500 多个内置函数可供使用。利用 GPUArray、特殊数组类型和并行化数值算法等高级别构造, 无需进行底层的 CUDA 编程即可对 MATLAB 应用程序进行并行化, 从而降低理论计算研究者对大型计算机的依赖, 减少研究者的计算成本。MATLAB 不仅提供本地多核的并行计算和仿真, 同时支持大规模的集群计算和仿真, 研究人员无需额外开发投入、可借助学校超算平台已部署的 MATLAB 集群多 GPU 运算环境, 获得计算速度进一步提升的空间。

对于第一性原理计算的算法开发者, KSSOLV-GPU 具有其他软件所没有的优势。KSSOLV-GPU 基于 MATLAB 环境, 语法简单, 线性代数功能强大, 在实现 DFT 计算中的算法和公式时, 可以极大程度降低开发难度和代码量。同时, MATLAB 自带成熟的 profiler 工具, 可以便捷地探查已完成的代码, 帮助研究者有针对性地对代码进行分析和优化。研究人员可以先在 KSSOLV-GPU 中用较少的代码量编写出程序的新功能, 将运行结果与主流计算软件 (VASP 和 Quantum ESPRESSO) 做对比, 确保结果的正确性后, 再编写为 Fortran/C++ 代码, 这样可以大大减少代码开发的工作量。

二、软件功能

KSSLOV 利用密度泛函理论近似求解薛定谔方程得到能量、原子力等信息, 能够使用局域密度近似 (LDA)、广义梯度近似 (GGA)、杂化泛函 (HSE06) 近似等方法求解 Kohn-Sham 方程。KSSLOV 能利用 MATLAB 的 GPU 加速工具进行 GPU 加速, 其能够采用 GPU 加速的计算功能包括:

2.1 基本功能

1. 基态的电子结构计算

能量、原子力、电荷密度

2. 交换关联泛函

LDA、GGA、HSE06

3. 几何优化

原子弛豫

4. 高性能计算

5. 优异的并行性能 Profiling 热点分析工具

2.2 特色功能

1. 激发态的电子结构计算

含时密度泛函理论 TDDFT, 线性响应的含时密度理论 LR-TDDFT

2. GW 计算

格林函数与屏蔽相互作用的自能展开, 利用多体微扰理论进行激发态计算

3. CPU-GPU 异构并行加速计算

采用 MATLAB, GPU 加速功能, 调用 CUDA 进行图形处理器计算

4. 独创的杂化泛函加速算法计算功能

ACE (自适应压缩交换算符)、ISDF (插值可分密度拟合)、PC-DIIS (投影子空间迭代求逆)

2.3 教学演示功能

1. 基态电子结构信息

2. 激发态电子结构信息

3. 光谱计算

4. 电子态密度 (DOS)
5. 晶体能带结构 (BAND)
6. 表面能计算
7. 化学形成能计算
8. 结构优化 (RELAX)
9. 化学反应路径与反应能量
10. 量子力学科普 (中学生)
11. 第一性原理计算学习 (大学生)
12. 新方法新理论的实现和分析 (科研人员)

三、性能测试

KSSOLV 是一套采用 MATLAB 语言编写、基于平面波基组来求解 Kohn-Sham 密度泛函理论(DFT)问题的软件工具集。在 DFT 计算中,最昂贵的部分通常是自洽场(SCF)中 Kohn-Sham 哈密顿量的对角化。为了使个人电脑能够进行中等规模(约百原子)的 DFT 计算,我们利用 MATLAB 内置的并行计算工具箱,提出了一种 CPU-GPU 混合方法,以加速 KSSOLV 中的迭代对角化算法。我们将 KSSOLV-GPU 在 NVIDIA RTX3090、V100、A100 三种 GPU 上的性能与 KSSOLV 在常规 CPU 上的性能进行比较。

测试一：常用算法性能测试

测试内容

在 RTX3090、V100 和 A100 三种不同类型的 GPU 上测试了矩阵乘法(mtimes)、快速傅里叶变换(fft)、元素积(次数)、水平级联(horzcat)、矩阵除法(mrdivide)和特征值(eig)的性能,并与 MATLAB 中的 CPU 进行了比较,用加速比来表示每个 GPU 的加速能力。

测试条件

硬件	软件	测试内容
CPU: Intel (R) Xeon (R) CPU E5-2698v4	MATLAB R2021a	矩阵乘法(mtimes)、快速傅里叶变换(fft)、元素积(次数)、水平级联(horzcat)、矩阵除法(mrdivide)和特征值(eig)
GPU: RTX3090		
GPU: V100		
GPU: A100		

测试结果

	mtimes	fft	times	horzcat	mrdivide	eig
A100	387.8	168.6	251.2	343.5	109.4	13.6
V100	137.0	85.5	134.8	171.8	50.0	11.9
RTX3090	11.7	35.9	155.2	191.7	12.5	9.9

表 0-1 选取矩阵乘法(mtimes)、快速傅里叶变换(fft)、元素乘积(times)、水平级联(horzcat)、矩阵除法(mrdivide)和特征值(eig) MATLAB 运算在测试矩阵全维上的最大 GPU 加速比。加速比由 CPU 的运算时间除以 GPU 运算时间得到。

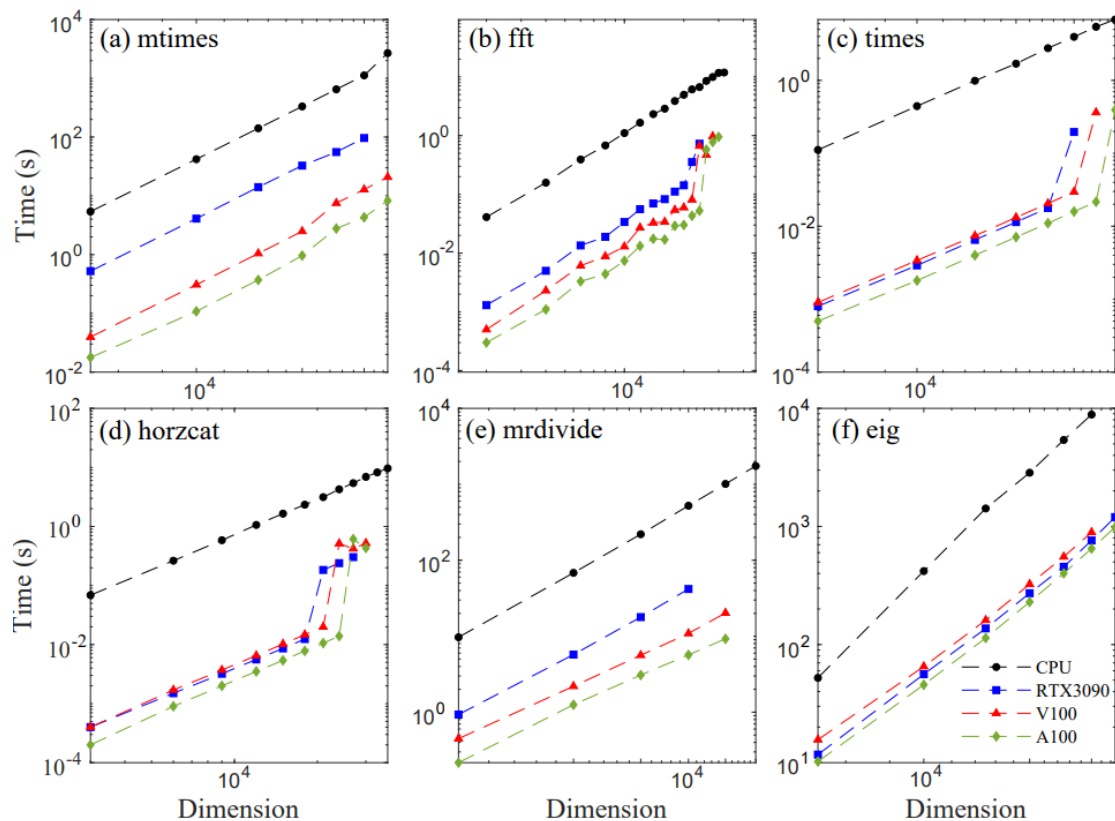


图 0-1 (a)矩阵乘法(mtimes)、(b)快速傅里叶变换(fft)、(c)元素乘积(times)、(d)水平级联(horzcat)、(e)矩阵除法(mrdivide)和(f)特征值(eig)随测试矩阵维数增加,在 MATLAB 上,三种不同类型的 GPU 上分别与 CPU 的运算性能进行比较。

可以看到,不同的操作在 GPU 上都有相当大的加速比。对于 V100 和 RTX3090 GPU, RTX3090 作为最新的消费级 GPU 在 eig、horzcat 和 time 操作上的性能与 V100 相当甚至更

好,但在 fft 上的性能只是 V100 的一半,在 mtimes 和 mrdivide 操作上的性能分别是 V100 的 1/10 和 1/5 左右。因此,可以在此基础上推断出对 KSSOLV 迭代对角化算法 GPU 加速性能良好。

测试二：使用 GPU 测试各对角化算法的性能

测试内容

测试随 Si 原子数的增加,在 CPU 和 GPU 上相同算法的计算性能和扩展性。

测试条件

硬件	软件	测试内容
CPU: Intel (R) Xeon (R) CPU E5 - 2698v4	MATLAB R2021a	在 CPU 和 GPU 上分别进行 LOBPCG、Davidson、Chebyfilt、OMM 标度对角化计算, Si 原子数目从 8 到 128, 参数 Ecut 取 20.0 Ha, 统计分析其加速时间和加速比。
GPU: RTX3090		
GPU: V100		
GPU: A100		

测试结果

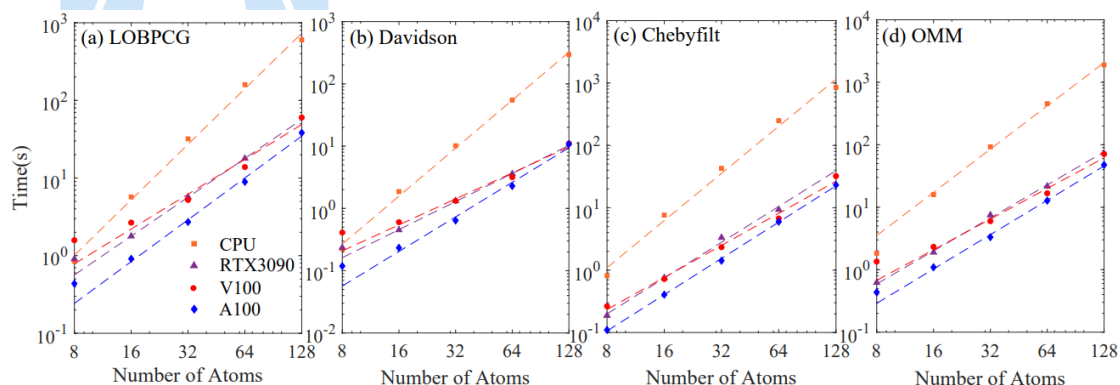


图 0-1 取 Ecut = 20.0Ha, 在 CPU 和 GPU 上进行不同的对角化运算, 时间消耗随着 Si 原子数目的增加而增加。(a) 在 CPU 和 GPU 上进行 LOBPCG 标度对角化计算; (b) 在 CPU 和 GPU 上进行 Davidson 标度对角化计算; (c) 在 CPU 和 GPU 上进行 Chebyfilt 标度对角化计算; (d) 在 CPU 和 GPU 上进行 OMM 标度对角化计算。

	LOBPCG	Davidson	Chebyfilt	OMM
CPU	2.37	2.56	2.51	2.30
RTX3090	1.66	1.50	1.91	1.76
V100	1.49	1.29	1.71	1.63
A100	1.79	1.84	1.93	1.83

表 0-1 取 Ecut = 20.0Ha, 在 CPU、RTX3090、V100 和 A100 上分别对 LOBPCG、Davidson、Chebyfilt 和 OMM 对角化算法的散点图进行线性拟合得到标度。

在不同 GPU 上测试了不同算法的加速比，结果如图 0-2 所示。加速比（speedup ratio）的计算公式为：

$$\text{speedup ratio} = \frac{\text{time of CPU}}{\text{time of GPU}}$$

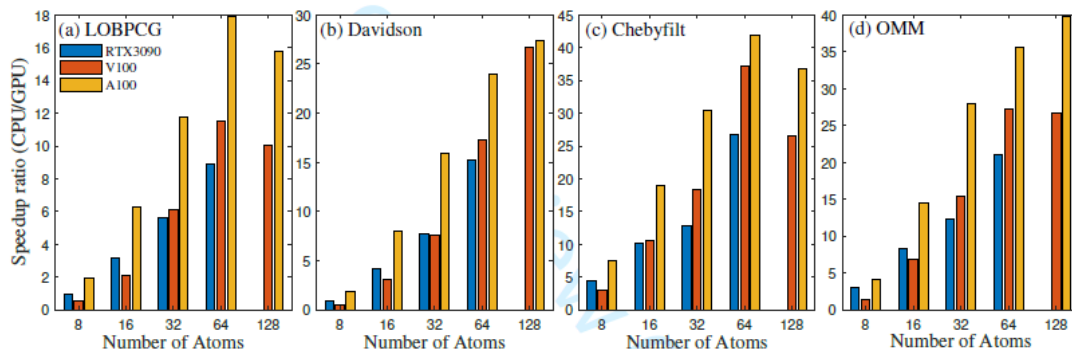


图 0-2 取 $E_{\text{cut}} = 20.0\text{Ha}$ ，与 CPU 相比，通过 GPU 加速后，不同对角化算法的加速比随着 Si 原子数目的增加而增大；(a) LOBPCG 在不同 GPU 上的加速比；(b) Davidson 在不同 GPU 上的加速比；(c) Chebyfilt 在不同 GPU 上的加速比；(d) OMM 在不同 GPU 上的加速比。

	LOBPCG	Davidson	Chebyfilt	OMM
RTX3090	8.94	15.28	26.82	21.01
V100	11.52	26.71	37.10	27.15
A100	17.89	27.38	41.81	39.77

表 0-2 LOBPCG、Davidson、Chebyfilt 和 OMM 对角化算法在 RTX3090、V100 和 A100 上的最大加速比

从测试的结果可以发现，所有对角化算法的执行时间随着原子数目的增加而变长，GPU 的加速效果明显。通过 GPU 加速，可以降低对角化过程的标度，通过 GPU 进行多线程并行计算，使得计算过程更快。

测试三：取不同 E_{cut} 值的性能

测试内容

测试随 Si 原子数的增加，且选取不同 E_{cut} 值时，在 CPU 和 GPU 上相同算法的计算性能和扩展性。

测试条件

硬件	软件	测试内容
CPU: Intel (R) Xeon (R) CPU E5 - 2698v4	MATLAB R2021a	在 CPU 和 GPU 上分别进行 LOBPCG、Davidson、Chebyfilt、OMM 标度对角化计算，Si 原子数目从 8 到

		128, 参数 Ecut 取 5、10、15、20、30 Ha, 统计分析其加速时间和加速比。
--	--	---

测试结果

(1) LOBPCG

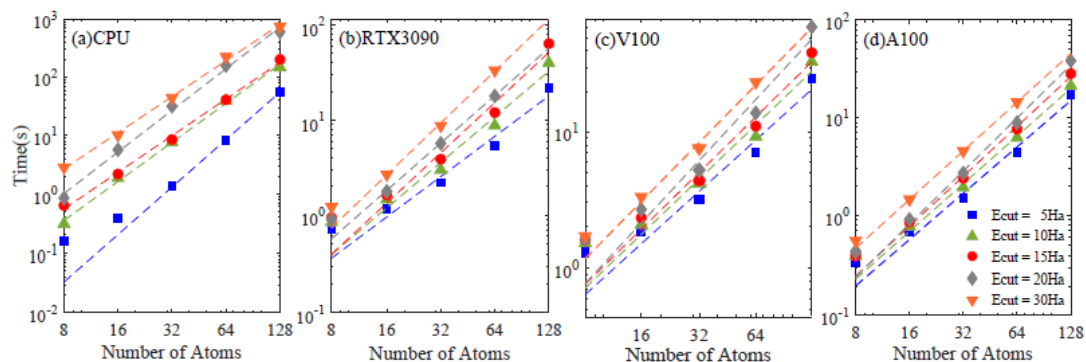


图 0-1 取不同的 Ecut、在不同的 CPU、CPU 下, LOBPCG 的时间消耗随着原子数目增加的变化:(a)通过 CPU 进行 LOBPCG 对角化计算; (b)通过 RTX3090 进行 LOBPCG 对角化计算; (c)通过 V100 进行 LOBPCG 对角化计算; (d)通过 A100 进行 LOBPCG 对角化计算。

(2) Davidson

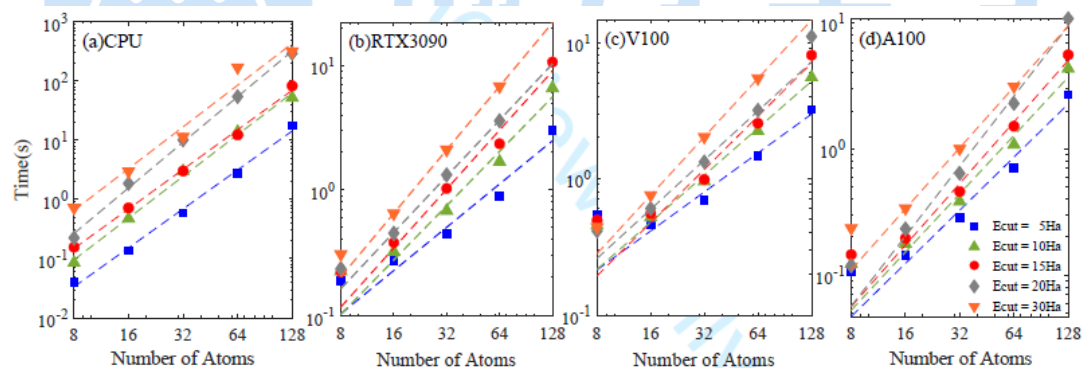


图 0-1 取不同的 Ecut、在不同的 CPU、CPU 下, Davidson 的时间消耗随着原子数目增加的变化:(a)通过 CPU 进行 Davidson 对角化计算; (b)通过 RTX3090 进行 Davidson 对角化计算; (c)通过 V100 进行 Davidson 对角化计算; (d)通过 A100 进行 Davidson 对角化计算。

(3) Chebyfilt

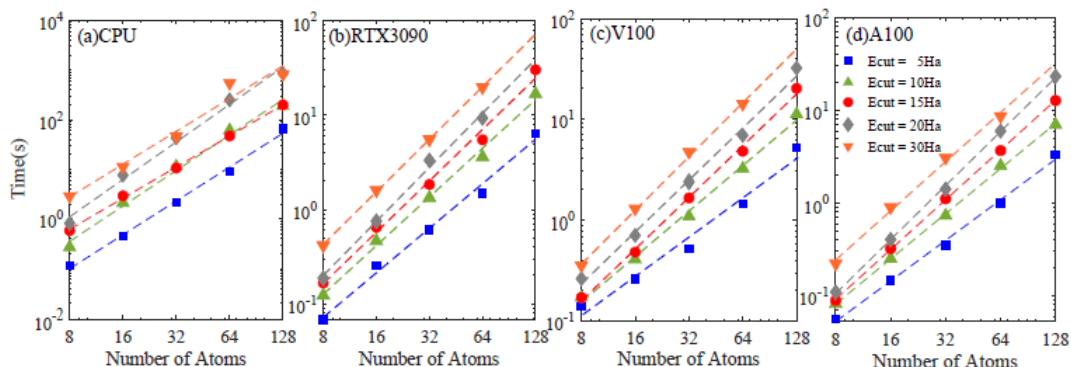


图 0-1 取不同的 Ecut、在不同的 CPU、GPU 下，Chebyfilt 的时间消耗随着原子数目增加的变化：(a) 通过 CPU 进行 Chebyfilt 对角化计算；(b) 通过 RTX3090 进行 Chebyfilt 对角化计算；(c) 通过 V100 进行 Chebyfilt 对角化计算；(d) 通过 A100 进行 Chebyfilt 对角化计算

(4) OMM

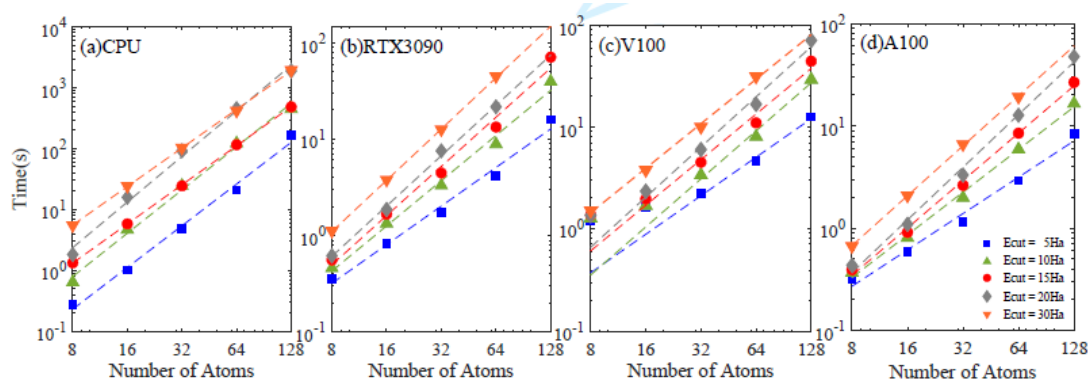


图 0-1 取不同的 Ecut、在不同的 CPU、GPU 下，OMM 的时间消耗随着原子数目增加的变化：(a) 通过 CPU 进行 OMM 对角化计算；(b) 通过 RTX3090 进行 OMM 对角化计算；(c) 通过 V100 进行 OMM 对角化计算；(d) 通过 A100 进行 OMM 对角化计算。

计算结果

	LOBPCG					Davidson					Chebyfilt					OMM				
Ecut(Ha)	5.0	10.0	15.0	20.0	30.0	5.0	10.0	15.0	20.0	30.0	5.0	10.0	15.0	20.0	30.0	5.0	10.0	15.0	20.0	30.0
CPU	2.68	2.23	2.09	2.38	2.06	2.19	2.34	2.22	2.56	2.34	2.27	2.38	2.08	2.51	2.20	2.27	2.36	2.13	2.49	2.10
RTX3090	1.40	1.58	1.75	1.66	1.82	1.14	1.44	1.56	1.50	1.69	1.57	1.71	1.81	1.91	1.85	1.33	1.56	1.70	1.73	1.77
V100	1.25	1.32	1.36	1.49	1.41	0.95	1.15	1.30	1.20	1.43	1.28	1.49	1.70	1.71	1.77	1.25	1.56	1.48	1.63	1.47
A100	1.55	1.61	1.68	1.79	1.66	1.40	1.53	1.62	1.84	1.60	1.45	1.62	1.78	1.93	1.76	1.18	1.38	1.54	1.71	1.62

表 0-1 不同对角化算法在不同 Ecut 的 CPU 和 GPU 上的标度

测试结论

通过在 CPU 和 GPU 上测试不同对角化算法的标度和加速比，与基于 CPU 计算相比，GPU 拥有强大的加速性能和更低的基态计算规模。尤其是性价比较好的最新消费级 GPU 卡

RTX3090，其计算性能不输于数据中心级 V100，充分说明可以借助 MATLAB 强大的可视化能力和优秀的消费级 GPU 卡来实现 KSSOLV-GPU 的加速计算和可视化，从而得以在个人电脑上快速完成第一性原理计算和可视化分析，降低计算材料科学的门槛。

四、开发者信息

- 开发人：胡伟、杨金龙、秦新明、李杰岚、万凌云、焦诗哲、张振林等
- 开发单位：中国科学技术大学
- 编程语言：MATLAB
- 开源类型：部分开源
- 版本：1.0
- 邮箱：zlzustc@mail.ustc.edu.cn、whuustc@ustc.edu.cn



瀚海量子
Hanhai Quantum