小分子无相位辅助场量子蒙特卡罗方法基准测试

Zoran Sukurma，[1,2]Martin Schlipf，[3]Moritz Humer，[1,2]Amir Taheridehkordi，[1]和 Georg Kresse[1,3]1)维也纳大学物理学院与计算材料科学中心，*Kolingasse 14-16，A-1090* 维也纳，奥地利2)维也纳大学物理学院与维也纳物理学博士生院，*Boltzmanngasse 5，A-1090* 维也纳，奥地利3)*VASP Software GmbH，Sensengasse 8，1090* 维也纳，奥地利

（日期：2023年3月9日）

我们报告了 ph-AFQMC（无相位辅助场量子蒙特卡罗）的可扩展 Fortran 实现，并展示了其出色的性能以及随系统规模增长的良好标度特性。此外，我们研究了对无相位近似的修改，这些修改有助于减少 ph-AFQMC 中常见的过度相关问题。我们将该方法应用于 HEAT 集中的 26 种分子、苯分子和水团簇。对于 HEAT 集中的分子，我们发现总能量的平均绝对偏差为 1.15 千卡 / 摩尔，接近化学精度。对于苯分子，尽管改进算法使用的是单 Slater 行列式试探波函数，其精度仍与采用 400 个斯莱特行列式的原始无相位方案相同。尽管有这些改进，我们发现氰基、一氧化碳和一氟氧分子存在系统误差，需要通过更精确的试探波函数加以解决。对于水团簇，我们发现 ph-AFQMC 得出的结合能非常准确，与 CCSD(T) 的差异通常小于 0.5 千卡/摩尔。

arXiv:2303.04256v1 [physics.chem-ph] 7 Mar 2023

I. 引言

密度泛函理论（DFT）通过将多体薛定谔方程映射为单粒子问题来近似求解该方程，现已成为量子化学和凝聚态物理中的常规方法。[1–3] 尽管其形式简单，且能够描述包含数百个原子的系统，DFT 并不总是足够精确，无法作为通用解决方案。拉伸的 H_2 分子是 DFT 局限性的一个著名例子，因为其结果在定性上是错误的。[4] 关键问题在于通过近似交换-关联（xc）能量泛函处理电子-电子关联效应。[5] Perdew 建立了这样一个类比：开发更先进的交换-关联泛函以达到化学精度，就像攀登雅各布天梯一样。[6] 更高的阶梯可带来更高的精度，但代价是更高的计算成本和更低的可迁移性。基于机器学习的新型交换-关联势[7,8] 试图解决精度问题，但仍然难以克服较低的可迁移性。机器学习泛函能否在未来几年取代传统泛函，仍有待观察。机器学习交换-关联势以及机器学习力场[9,10] 都需要高度精确的参考数据。此外，更精确的方法对于校准新方法，或研究 DFT 通常表现不佳的强关联材料，也同样重要。

正因如此，在过去 50 年间，人们发展出了一整套相关一致性方法。[11,12] 在给定基组内，全组态相互作用（FCI）能够提供精确解[13]，但其计算规模呈指数增长。因此，它只能用于规模较小的体系（最多包含 10^{10} 个斯莱特行列式）。通过对福克空间进行巧妙划分，人们发展出了一类称为选定组态相互作用的方法[14–16]，以及其随机版本——全组态相互作用量子蒙特卡罗（FCIQMC）。[17–19] 这些方法使得在保持传统 FCI 精度的同时，能够处理最多包含 10^{30} 个斯莱特行列式的福克空间，同时仍保持传统 FCI 的精度。然而，它们的计算规模仍呈弱指数增长，因此难以应用于更为现实的体系。

除了组态相互作用方法外，最突出的计算方法都基于耦合簇展开。其中最受欢迎的是被称为“黄金标准”的耦合簇单激发、双激发和微扰三激发（CCSD(T)），[20,21] 它被广泛用于处理分子，并且已知在静态关联效应较弱的体系中能给出极佳的结果。小分子的高精度基准数据通常通过耦合簇展开计算。例如，一系列聚焦于高精度外推从头算热化学（HEAT）的论文，在耦合簇展开中纳入了直至全五重激发的电子-电子关联。[22–24] 对于 HEAT 集中的分子，这些结果被认为与全组态相互作用的结果一样准确。更重要的是，HEAT 研究还表明，单独使用 CCSD(T) 的精度不足以达到化学精度（< 1 kcal/mol），但在许多相关情况下已接近这一精度。此外，CCSD(T) 的计算复杂度随体系规模的七次方增长，并且在存在强静态关联效应时表现不佳，例如键解离。这些缺点凸显了开发替代方法的必要性。

密度矩阵重整化群（DMRG）是一种变分方法，广泛用于处理具有强关联效应的系统。[25–27] 它利用矩阵乘积态编码一个空间维度上的局域性，并降低组态相互作用展开的指数标度。对于低维量子系统（如量子点以及沿一个维度延展的分子），DMRG 因其高效且准确而成为首选方法。尽管 DMRG 适用于任意系统，但其约 30 个活性轨道的限制，使其无法成为适用于更通用、更紧凑的分子或固体的通用工具。

量子蒙特卡罗方法，例如变分蒙特卡罗[28] (VMC)和扩散蒙特卡罗[29–31] (DMC)，其标度随系统尺寸呈低阶多项式增长，并且现

已成为数百个电.子体系的常规计算方法。然而，Nemec 及其合作者[32] 指出，使用 DMC 达到化学精度并非易事。他们报告称，对于 G2 集中的 55 个分子，原子化能的平均绝对偏差为 3.2kcal/mol。[33] 此外，VMC 和 DMC 需要准确的多电子波函数模型，包括 Slater 行列式的线性组合、Jastrow 因子、回流波函数等。最后，近期结合深度学习与 VMC 的研究已成为解决量子多体问题的一种有前景的方法。[34–36]

在本研究中，我们采用辅助场量子蒙特卡罗（AFQMC）方法。AFQMC 是一种成熟的投影蒙特卡罗方法，已成功应用于多种领域。[37–45] 最初的 AFQMC 被构建为一种采用 Metropolis 算法的路径积分方法。[46–49] 对于没有严重费米子符号问题的系统，该方法表现良好。由于几乎所有一般的费米子系统都存在符号问题，因此将其重新表述为开放式随机游走，进而发展出约束路径（cp）-AFQMC，并在模型哈密顿量方面取得了显著成功。[50,51] 为了处理一般的多体哈密顿量，Zhang 及其同事提出了无相位（ph）-AFQMC。[52] ph-AFQMC 已成功应用于采用高斯型轨道（GTO）的孤立分子，以及采用平面波基组（PWs）的固体。[53] 近年来，随着该方法日益普及，ph-AFQMC 的应用范围也不断扩大。[37–45] 该方法广受欢迎的原因之一是，ph-AFQMC 的计算量随系统大小呈四次标度。这种标度源于局域能量计算，并可通过张量缩并、[54,55] 密度拟合技术[56,57] 或平面波基组进一步降低。[53] 与 DMC 相比，AFQMC 得到的能量与其他传统量子化学方法完全一致，因此可以直接比较相关能。

控制费米子相位问题对于保证辅助场量子蒙特卡罗方法的准确性至关重要。在 ph-AFQMC 中，无相位近似对行走者权重施加约束，[52,58,59] 类似于扩散蒙特卡罗中的固定节点近似。[29] 无相位近似可能引入一些未受控的误差，通常会导致高估相关能。在 ph-AFQMC 中，系统地改进试探波函数可以减小无相位近似误差。当试探波函数趋近于精确基态波函数时，ph-AFQMC 能量也趋近于精确基态能量，无相位误差随之消失。通常，试探波函数由 Hartree-Fock（HF）或 Kohn-Sham（KS）轨道构成的单个斯莱特行列式组成。它们之所以特别受欢迎，是因为所有量子化学和固体物理代码都能轻松获得这些轨道，同时还可避免更复杂试探波函数所引入的额外计算成本。近期研究聚焦于多行列式试探波函数[60–64]。例如，Mahajan等使用了 10^4个斯莱特行列式，而与单个斯莱特行列式的情况相比，ph-AFQMC 计算的总成本仅增加了 3 倍。

在本研究中，我们采用一种更简单的方法。我们将仔细考察，对权重更新进行修改是否能够减少对单行列式试探波函数的超越需求。通过这些修改，ph-AFQMC 能够为 HEAT 集提供接近化学精度的结果。Bomble等[23] 在双ζ基组（cc-pVDZ）下提供了高度精确的 CCSDTQP 分子能量。在本工作中，我们将 ph-AFQMC 能量的准确性与这一小分子参考数据集进行对比基准测试。Borda等近期在 G1 测试集上开展了类似研究。[65]

本文其余部分安排如下：第二节简要回顾 ph-AFQMC 方法及其提出的改进，随后在第三节介绍实现细节。第四节中，我们展示并讨论 ph-AFQMC 的不同应用。最后，第五节总结我们的研究结果，并探讨未来的发展方向。

II. 辅助场量子蒙特卡罗形式体系

本节简要介绍辅助场量子蒙特卡罗形式体系。关于该理论更详细的综述，感兴趣的读者可参阅相关综述文章。[52,58,59] 考虑在任意正交归一单粒子基组中表示的完整多体哈密顿量：

$$\begin{aligned}\hat{H} &= \hat{H}_1 + \frac{1}{2}\sum_g \hat{L}_g^2 \\ &= \sum_{pq} h_{pq}\,\hat{a}_p^\dagger \hat{a}_q + \frac{1}{2}\sum_g \sum_{pq} L_{g,pq}\,\hat{a}_p^\dagger \hat{a}_q \sum_{rs} L_{g,rs}\,\hat{a}_r^\dagger \hat{a}_s,\end{aligned} \tag{1}$$

其中，$\hat{a}_p^\dagger$ 和$\hat{a}_q$ 分别为费米子产生算符和湮灭算符。单粒子哈密顿量矩阵元h_{pq} 包含所有单体项。双体哈密顿量可写为单体算符$\hat{L}_g$的平方和。在高斯基组中，我们可以通过对电子排斥积分（电子排斥积分）的迭代 Cholesky 分解[66,67] 得到这些算符。指标p、q、r和s 遍历N 个基函数，指标g 遍历N_g 个乔列斯基向量。通常，$N_g \approx 10N$。为简便起见，我们忽略自旋指标。

精确的多体基态波函数 $|\Phi_0\rangle$可由初始态 $|\Psi_I\rangle$经长时间虚时间传播提取得到。

$$|\Phi_0\rangle = \lim_{n\to\infty}\left[e^{-\tau\left(\hat{H}-E_0\right)}\right]^n |\Psi_I\rangle\,, \tag{2}$$

其中，τ 为虚时间步长，E_0 为估算的基态能量。该方程在时间步无限小的极限下是精确的。在本研究中，哈特里–福克轨道构成初始波函数。

为高效处理电子-电子相互作用，采用 Hubbard-Stratanovich（HS）变换[68,69]

$$e^{-\frac{\tau}{2}\Sigma_g \hat{L}_g^2} = \int \mathrm{d}x^{N_g}\; p(x_g)\, e^{i\sqrt{\tau}\Sigma_g x_g \hat{L}_g} + \mathcal{O}(\tau^2)\,, \tag{3}$$

其中，$p(x_g)$ 是标准正态分布，x_g 是从该分布中抽取的随机数。换言之，

Hubbard-Stratonovich 变换将实际的相互作用粒子系统映射为一个与随机场耦合的非相互作用粒子系统。由于电子排斥积分是正定的，随机场是纯虚数，也就是说，可以定义一个¯复值有效哈密顿量h，其形式为

$$\bar{h}^w_{pq} = h_{pq} - \frac{i}{\sqrt{\tau}}\sum_g x^w_g L_{g,pq}, \tag{4}$$

其中，上标w强调为每个行走者抽取独立的随机场。由N_w个行走者组成的系综表示精确的多体基态，其中每个行走者由以下元素表示：实值权重W、相位 θ和单个斯莱特行列式 $|\Psi\rangle$。因此，精确的多体基态波函数可近似表示为

$$|\Phi_0\rangle = \left(\sum_{kw} W^w_k e^{i\theta^w_k}\right)^{-1} \sum_{kw} W^w_k e^{i\theta^w_k} \frac{|\Psi^w_k\rangle}{\langle\Psi_{\mathrm{T}}|\Psi^w_k\rangle}, \tag{5}$$

其中，索引$|k$ 表示时间步，索引$|w$ 表示行走者。 $|\Psi_{\mathrm{T}}\rangle$表示试探波函数。行走者的初始化方式如下：

$$W^w_0 = 1; \qquad \theta^w_0 = 0; \qquad |\Psi^w_0\rangle = |\Psi_{\mathrm{I}}\rangle. \tag{6}$$

行走者的运动方程为

$$|\Psi^w_{k+1}\rangle = e^{-\tau\bar{h}^w}|\Psi^w_k\rangle, \tag{7}$$

$$W^w_{k+1}e^{i\theta^w_{k+1}} = W^w_k e^{i\theta^w_k}\frac{\langle\Psi_{\mathrm{T}}|\Psi^w_{k+1}\rangle}{\langle\Psi_{\mathrm{T}}|\Psi^w_k\rangle}. \tag{8}$$

在 AFQMC传播过程中，这个近似的基态波函数可用于获取系统的可观测量。最简单、也可能是最重要的可观测量——总能量——定义为哈密顿量的混合期望值

$$E_0 = \frac{\langle\Psi_{\mathrm{T}}|\hat{H}|\Phi_0\rangle}{\langle\Psi_{\mathrm{T}}|\Phi_0\rangle} \approx \frac{\sum_{kw} W^w_k e^{i\theta^w_k} E_{\mathrm{loc}}(\Psi^w_k)}{\sum_{kw} W^w_k e^{i\theta^w_k}}, \tag{9}$$

其中，局域能量估计量$E_{\mathrm{loc}}(\Psi^w)$使用广义维克定理[70]计算

$$\begin{aligned} E_{\mathrm{loc}}(\Psi^w) &= \frac{\langle\Psi_{\mathrm{T}}|\hat{H}|\Psi^w\rangle}{\langle\Psi_{\mathrm{T}}|\Psi^w\rangle} \\ &= \sum_{pq} h_{pq}G^w_{pq} + \frac{1}{2}\sum_g\sum_{pqrs} L_{g,pq}L_{g,rs}(G^w_{pq}G^w_{rs} - G^w_{ps}G^w_{rq}). \end{aligned} \tag{10}$$

单体约化密度矩阵G^w_{pq}定义为

$$G^w_{pq} \equiv G_{pq}(\Psi^w) = \frac{\langle\Psi_{\mathrm{T}}|\hat{a}^\dagger_p\hat{a}_q|\Psi^w\rangle}{\langle\Psi_{\mathrm{T}}|\Psi^w\rangle} = \left[\Psi^w(\Psi^\dagger_{\mathrm{T}}\Psi^w)^{-1}\Psi^\dagger_{\mathrm{T}}\right]_{qp}. \tag{11}$$

这里， Ψ_{T} 和 Ψ^w 表示N_e 个占据轨道，它们属于N 个基函数。方程（7、8、9、10）定义了自由传播（fp）辅助场量子蒙特卡罗。

A. 平均场扣除

乔列斯基向量$L_{g,pq}$ 对角分量的绝对值减小，会使辅助场量子蒙特卡罗中的涨落减小，从而获得更好的统计结果。为此，引入一个位移

$$\bar{L}_g = \langle\hat{L}_g\rangle_{\mathrm{T}} = \frac{\langle\Psi_{\mathrm{T}}|\hat{L}_g|\Psi_{\mathrm{T}}\rangle}{\langle\Psi_{\mathrm{T}}|\Psi_{\mathrm{T}}\rangle} \tag{12}$$

加到二体哈密顿量中

$$\hat{H}_2 = \frac{1}{2}\sum_g\left(\hat{L}_g - \bar{L}_g\right)^2 + \sum_g \bar{L}_g\hat{L}_g - \frac{1}{2}\sum_g \bar{L}^2_g\,. \tag{13}$$

这里，我们将第二项识别为与试探波函数相对应的经典哈特里势，并将其加到单体哈密顿量中。最后一项是与试探波函数相对应的哈特里能。平均场扣除不会增加 AFQMC 传播的计算成本。

B. 重要性采样

在 AFQMC 中引入重要性函数会使式 (3) 中的高斯概率密度发生偏移。这种偏移称为“力偏置”。将力偏置选为

$$f^w_g = -i\sqrt{\tau}\,\frac{\langle\Psi_{\mathrm{T}}|\hat{L}_g - \bar{L}_g|\Psi^w\rangle}{\langle\Psi_{\mathrm{T}}|\Psi^w\rangle}. \tag{14}$$

可使随机场中的统计涨落最小化。[58]应用平均场扣除和重要性采样后，有效哈密顿量（式 (4)）变为

$$\bar{h}^w_{pq} = h_{pq} - \frac{i}{\sqrt{\tau}}\sum_g (x^w_g - f^w_g)(L_{g,pq} - \bar{L}_g\delta_{pq}). \tag{15}$$

力偏置是依赖于行走者的量，必须在每个时间步为每个行走者进行计算。力偏置对于成功开展AFQMC模拟至关重要，然而，计算力偏置会给辅助场量子蒙特卡罗过程带来相当大的计算开销，甚至可能成为计算瓶颈。[59]

C. 无相位近似

fp-AFQMC 在形式上是精确的，但存在费米子相位问题。该问题体现在能量计算的表达式中（式（9））。经过平衡时间后， $|\Psi_{\mathrm{I}}\rangle$中包含的激发态会指数衰减，行走者将均匀分布于整个复平面。因此，式（9）中的分母趋于零，整个表达式变得无定义。为规避费米子相位问题，人们引入了无相位近似。

由于式（15）给出的有效哈密顿量是复数值的，轨道也会变成复数值。在扩散蒙特卡罗中，$|\Phi\rangle$ 和 $-|\Phi\rangle$ 都是薛定谔方程的有效解。类似地，在辅助场量子蒙特卡罗中，如果 $|\Phi\rangle$是有效解，那么对于任意 $\theta \in [0, 2\pi)$，$e^{i\theta}|\Phi\rangle$ 也是有效解。为防止辅助场量子蒙特卡罗模拟中的噪声呈指数增长，Zhang等[52] 引入了无相位近似

$$W_{k+1}^w = W_k^w \left| \frac{\langle \Psi_T | \Psi_{k+1}^w \rangle}{\langle \Psi_T | \Psi_k^w \rangle} I^w \right| \max\left(0, \cos(\Delta\theta)\right), \tag{16}$$

$$\theta_k^w = 0, \tag{17}$$

其中，重要性采样因子 I^w 和相位变化 $\Delta\theta$ 定义为

$$I^w = \exp\left[\sum_g x_g^w f_g^w - \frac{1}{2} f_g^w f_g^w\right], \tag{18}$$

$$\Delta\theta = \mathfrak{Im} \ln \frac{\langle \Psi_T | \Psi_{k+1}^w \rangle}{\langle \Psi_T | \Psi_k^w \rangle} \approx \mathfrak{Im} \sum_g x_g^w f_g^w. \tag{19}$$

更新方程 (16) 也可以改写为

$$W_{k+1}^w = W_k^w \max\left(0, \mathfrak{Re} \frac{\langle \Psi_T | \Psi_{k+1}^w \rangle}{\langle \Psi_T | \Psi_k^w \rangle}\right) |I^w|. \tag{20}$$

Zhang 等人[52,58] 指出，不同的近似会得到相同的期望值，但标准差略有不同。

在本研究中，我们对这一说法进行了更定量的考察。为此，我们修改了总能量计算表达式（方程 (9)）

$$E_0 = \frac{\sum_{kw} W_k^w \cos(\theta_k^w) E_{\mathrm{loc}}(\Psi_k^w)}{\sum_{kw} W_k^w \cos(\theta_k^w)} \tag{21}$$

以及行走者权重的更新过程

$$W_{k+1}^w e^{i\theta_{k+1}^w} = W_k^w e^{i\theta_k^w} \frac{\langle \Psi_T | \Psi_{k+1}^w \rangle}{\langle \Psi_T | \Psi_k^w \rangle} \mathfrak{Re} I^w \,. \tag{22}$$

本文将这一新方法称为 ph*-AFQMC。与原方法相比，它有两个关键区别：首先，我们保留行走者权重的复数性质，并使用其实部进行种群控制以及计算物理可观测量，例如能量。当 $|\theta_k^w| \geq \frac{\pi}{2}$ 时，行走者会被明确删除。其次，我们使用重要性权重的实部，而不是绝对值。我们认为只有实部才有意义，而使用绝对值会增大带有虚部的权重。这可能导致 ph-AFQMC 系统性地过度关联。我们将在第四节表明，ph*-AFQMC 得到的相关能绝对值系统性地较小。

III. 实现细节

本节介绍我们在 QMCFort 中实现的辅助场量子蒙特卡罗。[71] QMCFort 使用 Fortran 编写，并采用（多线程）BLAS/LAPACK 实现快速线性代数运算，并使用消息传递接口和 OpenMP 进行并行化。除辅助场量子蒙特卡罗外，QMCFort 还实现了限制性、限制性开壳层和非限制性 Hartree-Fock（RHF、ROHF 和 UHF）以及二阶 Møller–Plesset 微扰理论。电子排斥积分（ERIs）采用 McMuchie-Davidson 方案计算，[12] 并分解为乔列斯基向量。因此，QMCFort 可作为独立工具，用于设置和运行辅助场量子蒙特卡罗模拟。不过，只有其中的辅助场量子蒙特卡罗部分针对大规模计算进行了高度优化。基于这一原因，我们还将 QMCFort 与 PySCF 接口，用于孤立体系；并与 VASP 接口，用于扩展体系。[72] [2] 本节余下部分将详细介绍辅助场量子蒙特卡罗流程中计算密集部分的实现，以及 QMCFort 与 VASP 之间的接口。

A. 有效哈密顿量

在每个时间步，对每个行走者，我们根据式 (15) 构建有效哈密顿量。计算量最大的部分是将移位后的随机场与乔列斯基向量 $\sum_g (x_g^w - f_g^w) \cdot L_{g,pq}$进行卷积。将轨道指标 p 和 q 视为一个组合指标后，我们可将这一缩并转换为矩阵乘法，其计算量随系统规模呈立方标度（$\sim N^2 N_g N_w$）。通常，每个 MPI 进程处理 10 到 100 个行走者，并同时为它们计算这一缩并。尽管交换能的计算（第 III D 节）具有更差的标度，但有效哈密顿量的构建往往是辅助场量子蒙特卡罗中计算开销最大的操作，因为每个时间步都需要进行这项计算。

B. AFQMC传播

有效哈密顿量根据式 (7) 传播轨道。特罗特-铃木传播子中的矩阵指数[73,74] 用六阶泰勒展开近似。这需要多次将有效哈密顿量作用于轨道，即执行形式如下的矩阵乘法

$$\Psi_{pi}^w \to \bar{\Psi}_{pi}^w = \sum_q \bar{h}_{pq}^w \Psi_{qi}^w \,. \tag{23}$$

这些按$N^2 N_e$的比例缩放。

C. 力偏置

力偏置的计算（见式（14））等价于哈特里势和能量的计算。其计算复杂度随系统规模呈立方增长，即 $N N_g N_e N_w$。与有效哈密顿量（第三节 A）类似，力偏置是针对所有行走者同时计算的。为使其更清晰，我们首先定义收缩乔列斯基向量

$$\mathscr{L}_{g,pi} = \sum_q L_{g,pq} \Psi_{T,qi}, \tag{24}$$

然后，力偏置为

$$f_g^w = \sum_{pi} \mathscr{L}_{g,pi} \Psi_{pi}^w \,. \tag{25}$$

在这里，我们可以将 $\{p,\ i\}$ 视为一个单独的指标，并通过矩阵乘法计算这一缩并。我们使用力偏置来计算行走者的哈特里能量 Ψ^w

$$E_{\mathrm{H}}(\Psi^w) = \frac{1}{2} \sum_g f_g^w f_g^w . \tag{26}$$

D. 交换能

在 AFQMC 的所有单项计算中，交换能的计算所需的计算成本最高。尽管式（10）的朴素实现的标度为 $N^3 N_g N_w$，但我们通过计算中间数组 $\alpha_{g,ij}^w$

$$\alpha_{g,ij}^w = \sum_p \mathscr{L}_{g,pi} \Psi_{pj}^w \,. \tag{27}$$

行走者 Ψ^w 的交换能为

$$E_{\mathrm{x}}(\Psi^w) = -\frac{1}{2} \sum_g \sum_{ij} \alpha_{g,ij}^w \alpha_{g,ji}^w \,. \tag{28}$$

构造中间数组 $\alpha_{g,ij}^w$ 的四次标度（$N N_g N_e^2 N_w$）过程也被转化为矩阵-矩阵乘法。为获得最佳性能，会同时处理多个行走者，每批约 5–10 个。比如，每隔 10 个时间步才计算一次交换能也可以。由于需要对相互独立的局域能进行采样，这不会显著影响统计；另一种做法是分块平均。由 α 张量计算交换能的标度为 $N_g N_e^2$，但与本节讨论的其他任务相比，其计算开销可以忽略不计。

E. 其他任务

除了上述操作外，AFQMC流程中还有几个部分值得一提。由于AFQMC传播子是非酉的，我们会通过对行走者矩阵进行QR分解，定期对行走者重新正交化 Ψ^w。我们会控制行走者群体，以避免权重过小或过大。我们选择了梳状方法[75]，，因为它能使行走者总数保持不变。

F. 性能

表 I 总结了计算量较大的任务所对应的操作次数。假设行走者数量与系统规模无关，

表 I. AFQMC流程中计算密集部分每个行走者每个时间步的操作次数及其渐近标度。N 表示基函数的数量，N_g表示乔列斯基向量的数量，N_e 表示系统中的占据态数量。

Task	操作次数	渐近标度
有效哈密顿量	$N^2 N_g$	$\mathcal{O}(N^3)$
AFQMC传播	$N^2 N_e$	$\mathcal{O}(N^3)$
力偏置	$N N_g N_e$	$\mathcal{O}$(氮^3)
交换能	$N N_g N_e^2 + N_g N_e^2$	$\mathcal{O}$(氮^4)

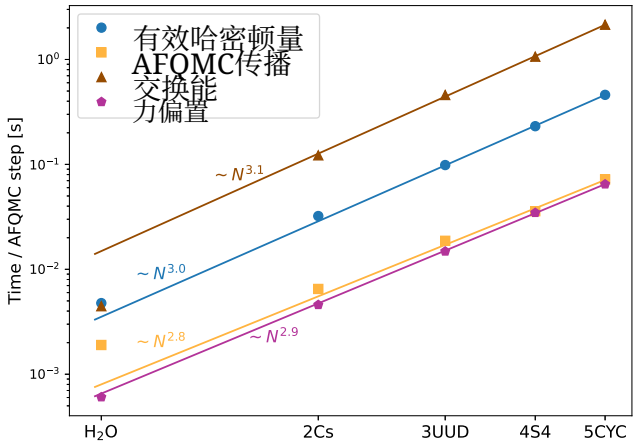


图 1. 在不同规模的水团簇上测得的计算密集型操作每个辅助场量子蒙特卡罗步骤的耗时。辅助场量子蒙特卡罗计算在配备 4800 个行走者的双路 Intel Skylake Platinum 8174 处理器上进行，使用了 48 个消息传递接口进程。计时结果取 200 个辅助场量子蒙特卡罗步骤的平均值。

我们列出了这些任务随系统规模增大时的渐近标度。我们对不同尺寸的水团簇进行了渐近行为评估，最多包含 5 个水分子。计算在一台双路 Intel Skylake Platinum 8174 处理器上完成，每个处理器有 24 个核心，标称基准频率为 3.1 GHz。我们对 200 个辅助场量子蒙特卡罗步的计时结果取平均，每一步在 48 个消息传递接口进程上处理 4 800 个行走者。

图 1 展示了计算量较大的任务在每个 AFQMC 步骤中的耗时。为估算各项任务的有效标度指数，我们对数据点进行幂函数拟合：$f(x) = Cx^\beta$。正如预期，构建有效哈密顿量、AFQMC传播和力偏置计算均呈立方标度。测得的交换能指数 $\beta = 3.1$ 与预期指数 $\beta = 4$ 并不完全一致。随着系统规模增大，我们观察到交换评估的每核性能有所提升（图 2）。这实际上降低了标度指数。尽管如此，交换评估仍是相关计算任务中性能最低的一项。如图 2 所示，大多数其他任务的运行效率较高，性能达到每核 20–35 吉 FLOPS。一个显著的例外是 AFQMC传播，在大型系统中其性能超过

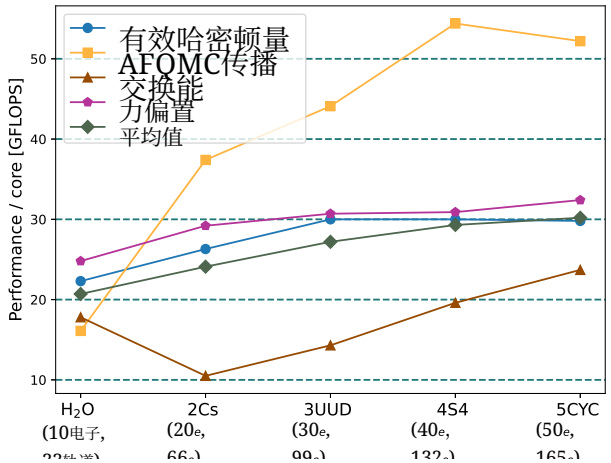


图 2. 在不同大小的水团簇上测得的 AFQMC 流程中计算密集型操作的单个 CPU 核心性能。x轴上的数对表示电子数和轨道数。在单个 CPU 节点（双路 Intel Skylake Platinum 8174）上运行 AFQMC 代码，平均可达到 1–1.5 TFLOPS（每个核心 30 GFLOPS，绿色线）。

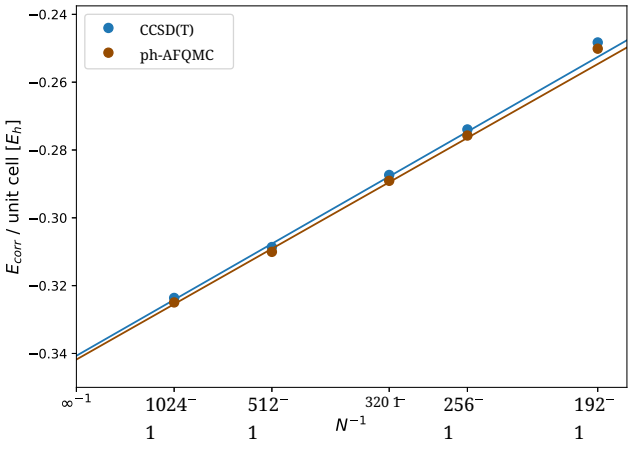


图 3. 金刚石的 CCSD(T) 和 ph-AFQMC 相关能随正则轨道数量变化的收敛行为。使用前三个点（1024^{-1}、512^{-1} 和 320^{-1}）进行完全基组线性外推。计算采用能量截断为 700 eV 的 2x2x2 金刚石超胞。

每核 50 吉 FLOPS，十分接近峰值性能。

G. 与 VASP 的接口

VASP 提供一组单电子平均场轨道$\{\phi_p(\mathbf{r})\}$。基于这些轨道，我们计算单粒子哈密顿量矩阵元 h_{pq} 和乔列斯基向量 $L_{g,\,pq}$ （更多详情请参见附录）。我们将矩阵元和乔列斯基向量导出到文件中，再用 QMCFort 读取。随后，扩展体系的辅助场量子蒙特卡罗计算等同于使用高斯基组对孤立分子进行的计算。

例如，我们计算金刚石的 2x2x2 超胞。我们研究辅助场量子蒙特卡罗和 CCSD(T) 相关能量随正则轨道数的收敛情况。CCSD(T) 能量使用 Cc4s 代码计算。[76] 采用的晶格常数为a = 3.567 Å，能量截断为$E_{\text{cut}} = 700$ eV。此外，我们使用能量截断E^{χ} cut = 500 eV 对库仑核进行截断。VASP 使用投影缀加波势[3,77]，所用势称为 C_GW，其价电子组态为 $2s2p$。该势的芯半径为$r_{\text{core}} = 1.5$ a.u.；其局域势等同于 d赝势，并使用了 4 个投影子（两个s投影子，半径为$r_{\text{cut}} = 1.2$ a.u.；两个p投影子，半径为$r_{\text{cut}} = 1.5$ a.u.）。

图 3 展示了 ph-AFQMC 和 CCSD(T) 的相关能随正则轨道数量变化的收敛情况。轨道数量从较小的基组（含 64 个电子和 192 个轨道，即 64e、192o）变化到相对较大的基组（含 64 个电子和 1024 个轨道，即 64e、1024o）。通过数据点拟合出一种 $1/N$ 行为。ph-AFQMC 和 CCSD(T) 曲线几乎无法区分，这表明这两种方法对周期体系给出的能量一致。关于用于扩展体系的 AFQMC 计算的更多细节，将在后续发表的论文中报告。[78]

为比较 CCSD(T) 和 ph-AFQMC 的计算成本，我们使用 256 条和 1024 条能带进行了计算。对于 256 条能带的计算，CCSD(T) 在单个计算节点（2x AMD Epyc（Milan））上耗时 20 分钟，而 ph-AFQMC 耗时 11 小时。然而，对于 1024 条能带的计算，CCSD(T) 在 4 个计算节点（2x AMD Epyc（Milan））上耗时 24 小时，而 ph-AFQMC 耗时 48 小时。我们测得的 CCSD(T) 和 ph-AFQMC 计算耗时表明，辅助场量子蒙特卡罗方法具有良好的标度特性；随着轨道数量增加、系统复杂度提高，其优势也愈发显著。

IV. 结果与讨论

在接下来的四个小节中，我们将报告：（i）HEAT 集中 26 个分子的总分子能量；（ii）使用 cc-pVDZ 基组计算的苯分子的 AFQMC 能量；以及（iii）水团簇的 AFQMC 结合能。

A. HEAT 分子

我们采用冻结芯近似，并使用参考文献23（HEAT）中的几何结构，计算了 HEAT 集中 26 个分子的总能量。所有计算均采用双ζ基组（cc-pVDZ）。我们使用 QMCFort 代码计算乔列斯基向量、参考 Hartree-Fock（RHF/UHF）波函数以及所有 AFQMC 能量。我们

表 II. HEAT 集[22] 分子在不同理论水平下的总分子能量（单位：哈特里）：ph-AFQMC[52], ph*-AFQMC、CCSD(T)、CCSDTQP[23] 和 HCI[14,15]。所有能量均使用 cc-pVDZ 基组计算。表格最后三行分别列出了均方根偏差（RMSD）、平均有符号偏差（MSD）和平均绝对偏差（MAD）。

分子	ph-AFQMC	ph*-AFQMC	CCSD(T)	CCSDTQP	HCI
H_2	−1.16363(2)	−1.16338(3)	−1.163426	−1.163426	−1.163426
CH	−38.37764(6)	−38.37705(7)	−38.379655	−38.380241	
CH_2	−39.04181(5)	−39.04150(6)	−39.041196	−39.04165	
NH	−55.09087(5)	−55.09077(7)	−55.097448	−55.0917	
CH_3	−39.71651(6)	−39.71592(7)	−39.715543	−39.71607	
NH_2	−55.73265(6)	−55.73222(8)	−55.732506	−55.733076	
OH	−75.55854(6)	−75.55837(8)	−75.559233	−75.559689	
HF	−100.22933(7)	−100.2290(1)	−100.228131	−100.228622	
H_2O	−76.24249(8)	−76.24167(9)	−76.241018	−76.241649	−76.241637
NH_3	−56.40334(7)	−56.4024(1)	−56.401913	−56.402517	
C_2H	−76.3999(2)	−76.3986(2)	−76.398558	−76.401217	
CN	−92.4997(2)	−92.4973(2)	−92.488695	−92.49276	−92.492788
C_2H_2	−77.1118(2)	−77.1092(2)	−77.109249	−77.110678	
CO	−113.0594(2)	−113.0567(2)	−113.054431	−113.055892	
HCN	−93.1912(2)	−93.1881(2)	−93.188321	−93.18991	
N_2	−109.2781(2)	−109.2750(2)	−109.275298	−109.277012	−109.277005
HCO	−113.5778(2)	−113.5757(2)	−113.575706	−113.577384	
CF	−137.4764(1)	−137.4754(2)	−137.474848	−137.476019	
NO	−129.5970(2)	−129.5945(2)	−129.597778	−129.599737	
HNO	−130.1736(2)	−130.1697(2)	−130.170989	−130.172906	
O_2	−149.9792(2)	−149.9787(2)	−149.985684	−149.987773	
HO_2	−150.5606(2)	−150.5603(2)	−150.558481	−150.56038	
OF	−174.5004(2)	−174.4998(2)	−174.497924	−174.50009	
H_2O_2	−151.1963(2)	−151.1940(2)	−151.19363	−151.195266	
F_2	−199.0968(2)	−199.0965(2)	−199.097448	−199.099328	
CO_2	−188.1561(2)	−188.1525(2)	−188.147429		−188.149551
均方根偏差（单位：mE_h）	2.89	2.73	1.65		
平均有符号偏差（单位：mE_h）	−0.38	1.17	1.39		
平均绝对偏差（单位：mE_h）	1.85	1.84	1.39		

使用 PySCF 软件包[72] 计算 CCSD(T) 能量。使用 Dice代码[14,15] 计算热浴组态相互作用能量。为排除 AFQMC 中可能的系统误差来源，我们（i）以较为保守的阈值 10^{-8} 截断乔列斯基向量，以确保对分子电子排斥积分的表示具有足够的精度；（ii）选择相对较小的时间步长 $0.002E_h^{-1}$，以避免显著的时间步长误差；（iii）对系统进行 40000 个时间步的平衡。为确保统计结果可靠，我们传播 6000 个行走者，直至预测的分子能量的均值标准误低于 0.2mE_h。

表 II 比较了 ph-AFQMC 和 ph*-AFQMC 的分子能量与“金标准”CCSD(T) 的结果，以及包含变分三重、四重和五重激发的更精确耦合簇展开（CCSDTQP）的结果。由于后者实际上已收敛至全组态相互作用极限，我们将其用作参考值。参考文献23中缺少 CO_2 分子的 CCSDTQP 值，因此我们采用热浴组态相互作用（HCI）的结果。[14] 为验证 HCI 方法的准确性，我们选取了另外几个分子（H_2, H_2O、氰基、N_2），发现它们与 CCSDTQP 值的最大差异为 0.01 mE_h，远低于 AFQMC 的统计误差。与这些参考值相比，我们发现 ph-AFQMC 的精度相近

ph*-AFQMC 方法的均方根偏差（RMSD）分别为 2.89 mE_h 和 2.73 mE_h。作为参照，CCSD(T) 对这些分子的 RMSD 为 1.65 mE_h 。平均绝对偏差（MAD）分别为 1.39 mE_h（CCSD(T)）、1.85 mE_h（ph-AFQMC）和 1.84 mE_h（ph*-AFQMC），表明不同方法之间的差异较小，这说明一些离群值拉低了 AFQMC 的整体表现。平均有符号偏差（MSD）为 1.39 mE_h（CCSD(T)）和 1.17 mE_h（ph*-AFQMC），表明存在相关作用不足；相比之下，ph-AFQMC 平均存在 −0.38mE_h 的相关作用过度。改进方法（ph*-AFMQC）的结果也更接近 CCSD(T)：其 RMSD 为 2.66 mE_h ，而 ph-AFQMC 的 RMSD 为 3.67 mE_h。

接下来，我们根据图4中的结果图示进行更详细的分析。首先关注使用更广泛的 CCSD(T) 方法。图4显示，与参考值相比，CCSD(T) 系统性地低估了电子相关能。最常见的偏差约为 1 kcal/mol（即 1.59 mE_h）。偏差最大的情况是氰基分子，偏差为 4.1 mE_h。偏差较大的原因可能是 UHF波函数中存在较大的自旋污染。[23]

ph-AFQMC 和 ph*-AFQMC 的结果呈现出相似的

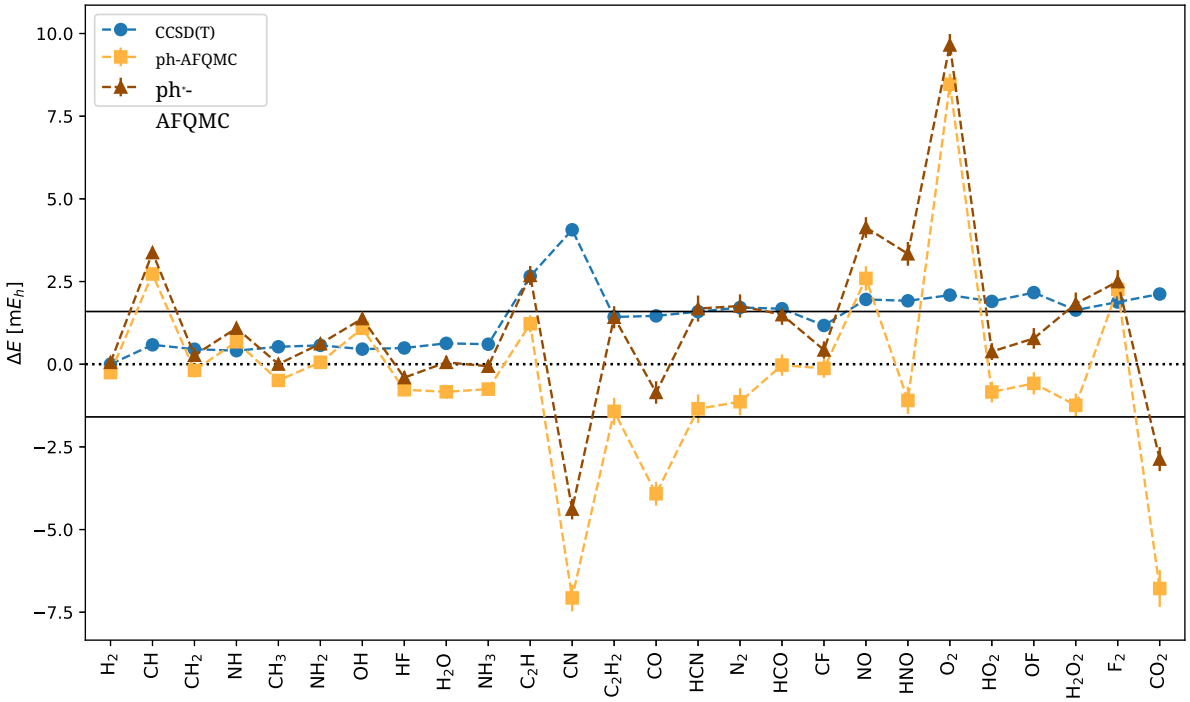


图4. CCSD(T)、ph-AFQMC 和 ph*-AFQMC 相对于 HEAT 集参考 CCSDTQP 能量的总能量差 ΔE。计算采用 cc-pVDZ 基组和冻结芯近似，仅包含最后一个占据壳层。试探波函数为单个 RHF/UHF 斯莱特行列式。黑色实线表示化学精度范围（ΔE < 1千卡/摩尔）

趋势。O_2 分子是辅助场量子蒙特卡罗方法中表现最差的情形，欠相关达到 8.6mE_h。解决这一问题需要采用多行列式试探波函数，例如通过在开壳层子空间中进行完备活性空间自洽场（CASSCF）计算得到。开壳层原子中类似的问题可通过使用多行列式 CASSCF 试探波函数或采用对称性恢复技术来解决。[59] 从统计中排除 O_2 后，ph-AFQMC 和 ph*-AFQMC 方法的均方根偏差分别降至 2.41 和 2.11mE_h。

与 O_2 分子不同，CN、CO 和 CO_2 分子的 ph-AFQMC 结果显示出显著的相关作用过度（CN 分子最高达 7mE_h）。这些偏差源于固定节点误差。我们改进的方法（ph*-AFQMC）系统性地提高了能量，相较于 ph-AFQMC 尤其如此。值得注意的是，ph-AFQMC 相关作用过度越明显的情况，两者的差异也越大。因此，CN、CO 和 CO_2 分子的固定节点误差显著降低，但残余误差仍然较大。更好的试探波函数有望进一步降低固定节点误差。

Borda 等人[65] 使用 G1 测试集进行了类似的基准测试，该测试集与 HEAT 集部分重叠。他们还使用带冻结芯近似的 cc-pVDZ 基组，并将通过 QMCPACK 软件包计算的 AFQMC 总能量与 CCSDTQ 能量进行了比较。我们使用 UHF 试探波函数，而他们在 AFQMC 计算中使用 ROHF 试探波函数。图 5 显示了 ph-AFQMC 相对于参考能量的偏差。总体而言，两种 ph-AFQMC 计算所得的大多数能量彼此相差不到 1 mE_h。其余情况下，除 CN 分子外，我们的结果似乎更接近参考结果。这些差异可能的原因包括：(i) 本研究中的统计误差和系统误差较小；(ii) 分子几何结构不同；(iii) 试探波函数不同。最后一点对 CN 分子尤为重要，因为 UHF 波函数存在较大的自旋污染，因此 ROHF 波函数是更合适的试探波函数。

接下来，我们通过比较 ph*-AFQMC 与 CCSD(T) 在 cc-pVDZ 基组和规模约大六倍的 aug-cc-pVQZ 基组下的总能量差异，研究基组的影响。后者包含芯电子，克服了较小基组中采用的冻结芯近似。图 6 显示，两种基组下的能量差异极为相似。三个显著例外是 F_2，O_2，和羟基，其偏差略大于 1 kcal/mol。这一结果有两方面的重要意义：首先，使用 cc-pVDZ 基组比较 ph*-AFQMC 与 CCSD(T) 已足够准确。从完备基组极限的计算中几乎无法得出任何新结论。其次，ph*-AFQMC 能够以 CCSD(T) 方法的精度准确描述芯电子效应。

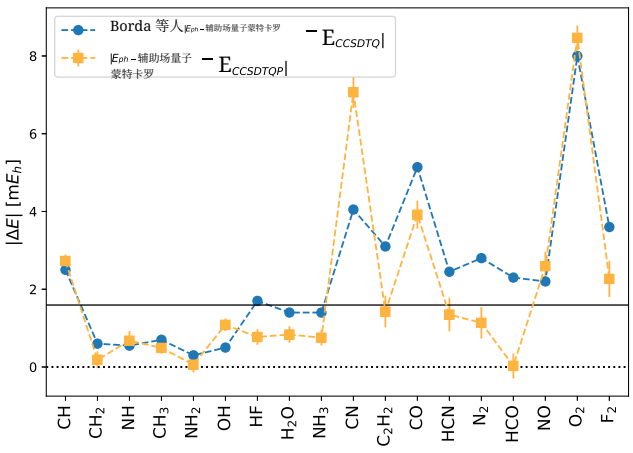


图 5. G1 集和 HEAT 集中重叠分子的 ph-AFMQC（QMCPACK）与 CCSDTQ（见参考文献 65）以及 ph-AFQMC（QMCFort）与 CCSDTQP 之间的总能量绝对差值 $|\Delta E|$。

B. 苯分子

采用双ζ基组（cc-pVDZ）的苯分子基态是一个很有意义的基准测试体系，可用于评估最先进的相关一致性方法，因为它是能够直接通过 FCI 对角化处理的最大体系之一。采用 cc-pVDZ 基组时，苯分子包含 108 个轨道和 30 个电子（采用冻结芯近似）。

Eriksen 等人[79] 对苯分子进行了盲测，比较了十种不同的方法，包括耦合簇展开方法、不同的选定组态相互作用方法以及量子蒙特卡罗方法。作者认为，全耦合簇约化（FCCR）和多体 FCI 展开（MBE-FCI）基本上得到了 $-863.0\,\mathrm{m}E_\mathrm{h}$ 的精确相关能。Lee 等人[43] 在这项研究中补充了采用两种不同试探波函数得到的 ph-AFQMC 结果：RHF 波函数和 CASSCF(6,6) 多组态波函数。ph-AFQMC+RHF 高估了相关能 3.1 $\mathrm{m}E_\mathrm{h}$，而 ph-AFQMC+CAS(6,6) 则多计算了 1.2 $\mathrm{m}E_\mathrm{h}$ 的相关能。为求完整，我们还补充了 CCSD(T) 结果，其相关能低估了 3.5 $\mathrm{m}E_\mathrm{h}$。我们的 ph-AFQMC+RHF 结果与使用 QMCPACK 代码计算的 ph-AFQMC+RHF 结果相当。[80,81] 这进一步验证了我们在 QMCFort 中的 AFQMC 实现。与 CN、CO 类似，CO_2, ph*-AFQMC+RHF 大幅降低了相关能的绝对值，导致能量的相关程度低估了 1.3 $\mathrm{m}E_\mathrm{h}$。在这种情况下，ph*-AFQMC+RHF 的表现似乎与 ph-AFQMC+CAS(6,6}相近。不过，更好的试探波函数有望提高准确性，更重要的是，能得到更可控的结果。图 7 展示了所有方法及其对应的相关能。

C. 水团簇

人们投入了大量精力来开发能够忠实表征体相水性质的经验模型。事实证明，好的模型还必须能够充分描述地球大气层中存在的小型水团簇。Temeslo等人 收集了不同大小水团簇的代表性结构。[82] 他们研究了哪些贡献对于获得准确的形成能至关重要。Motta 等人 报告称，AFQMC 和 CCSD(T) 的偏差大于统计涨落。[54] 在此，我们重新考察这些团簇，以审视 AFQMC 对水的计算准确性。

我们估算了含$n \leq 5$ H_2O 个分子的最稳定水团簇的结合能

$$E_b(n\mathrm{H_2O}) = E(n\mathrm{H_2O}) - nE(\mathrm{H_2O})\,. \tag{29}$$

这里，$E(n\mathrm{H_2O})$ 是团簇的总能量，而$E(\mathrm{H_2O})$ 是水分子的基态能量。团簇几何结构取自参考文献82。我们使用 aug-cc-pVDZ 基组，通过 MP2 优化了单个水分子，得到键长 0.96593 Å 和键角 103.866°。为与参考文献54比较结果，我们采用重原子增强基组（O 原子使用 aug-cc-pVDZ，H 原子使用 cc-pVDZ）和全电子波函数。PySCF 程序包[72] 计算了 CCSD(T) 相关能，QMCFort 则计算了限制性 Hartree-Fock、MP2 和辅助场量子蒙特卡罗的相关能。我们以 $10^{-6}E_\mathrm{h}$ 为阈值截断电子排斥积分的 Cholesky 分解。我们以时间步长 $0.01E_\mathrm{h}^{-1}$传播了 4 800 个行走者，共 140 000 步。每 100 步评估一次交换能。对于最大的团簇，我们使用了数量减半的行走者。此设置使结合能中的系统误差保持在 0.05 kcal/mol 以内。

表 III 列出了参考文献54中的 RHF、MP2、CCSD(T)、ph-AFQMC、ph*- AFQMC 和 AFQMC 结合能。尽管 RHF 结合能与相关一致性方法的结果存在显著差异，但其他所有方法的结果均在化学精度范围内一致。与此前的 AFQMC 结果相比，我们的 AFQMC 数值与 MP2 和 CCSD(T) 数值更加吻合。参考数据中较大的统计误差可能部分解释了 AFQMC 结合能之间的差异。然而，参考 AFQMC 结合能存在明显的相关作用不足，这也表明其中存在系统误差。系统误差的一个可能来源是参考文献54中电子排斥积分（$10^{-4}E_\mathrm{h}$）进行 Cholesky 分解时采用了较宽松的阈值。

为检验 ph-AFQMC 和 ph*-AFQMC 的尺寸一致性，我们将 3UUD 团簇中的第二个和第三个分子分别沿x 和y 方向移动 8 Å，使团簇解离为单个分子。我们分别计算了拉伸后团簇的总能量以及各个水分子的总能量。研究结果汇总于表 IV。考虑统计涨落后，所有片段的总能量相同，且等于拉伸后团簇总能量的三分之一。这表明 ph-AFQMC 和 ph*-AFQMC 具有尺寸一致性。

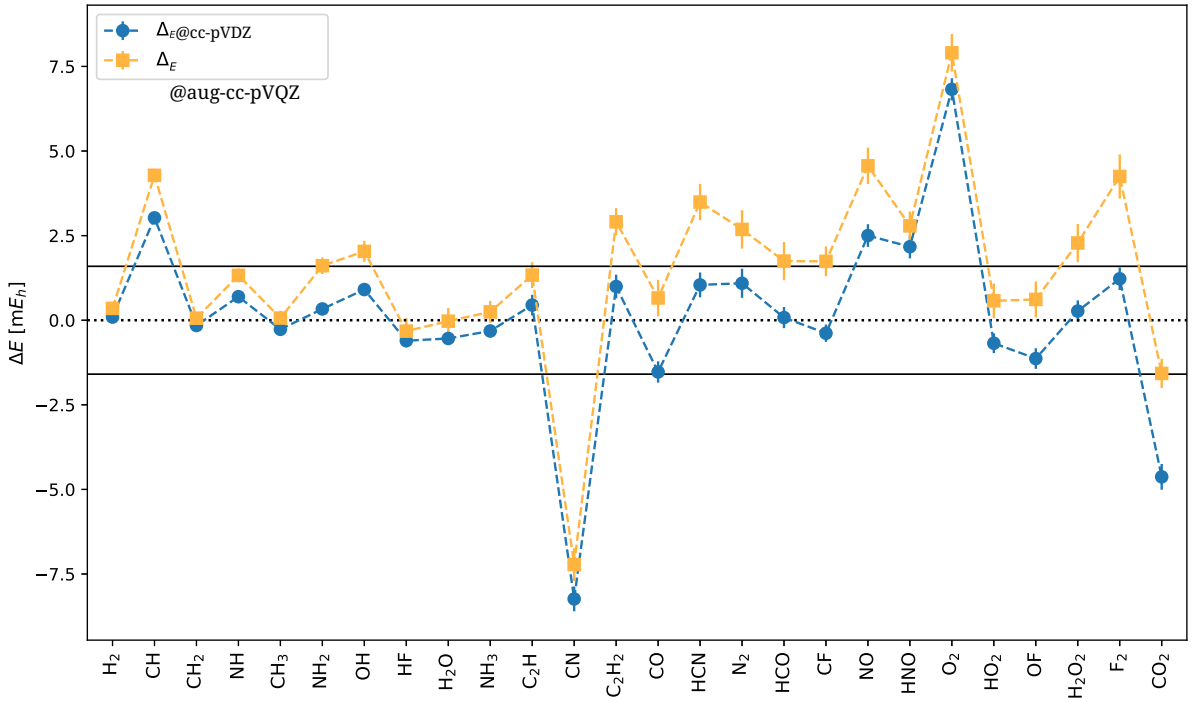


图 6. HEAT 集中 ph*-AFQMC 与 CCSD(T) 的总能量差 ΔE。cc-pVDZ 计算采用冻结芯近似；aug-cc-pVQZ 计算包含所有电子。闭壳层和开壳层分子分别采用单个 RHF 和 UHF 斯莱特行列式作为试探波函数。黑色实线表示化学精度（ΔE < 1千卡/摩尔）。

表 III. 使用重原子增强 cc-pVDZ 基组和全电子波函数计算的四种最稳定水团簇的结合能，这些团簇最多含有 5 个 H_2O 分子。文中报告了 RHF 、二阶 Møller–Plesset 微扰理论、 CCSD(T) 以及不同 ph-AFQMC 方法计算的数值，单位为千卡 / 摩尔。

簇	RHF	MP2	CCSD(T)	ph-AFQMC	ph*-AFQMC	ph-AFQMC[54]
2Cs	−3.82	−5.22	−5.18	−5.17(5)	−5.06(7)	−5.11(31)
3UUD	−10.52	−15.83	−15.62	−15.67(9)	−15.68(9)	−14.78(64)
4S4	−19.00	−28.36	−27.87	−28.12(10)	−28.11(10)	−26.49(46)
5CYC	−25.30	−37.48	−36.78	−37.14(28)	−37.31(28)	−36.27(59)

表 IV。拉伸的 3UUD 团簇及其片段的辅助场量子蒙特卡罗能量，以及结合能表明，ph-AFQMC 和 ph*-AFQMC 均具有尺寸一致性（所有数值均以 E_h 给出）。

簇	ph-AFQMC	ph*-AFQMC
3UUD	−228.82750(12)	−228.82448(12)
1-H_2O	−76.27576(4)	−76.27475(4)
2-H_2O	−76.27580(4)	−76.27477(4)
3-H_2O	−76.27585(4)	−76.27481(4)
E_b	0.00009(14)	0.00015(14)

五、结论

无相位辅助场量子蒙特卡罗因其高精度、低多项式标度（N^3 –N^4）以及在量子化学和凝聚态物理中的适用性而日益受到青睐。以 DFT 或哈特里–福克解为起点，它可以被视为一种自然的扩展这些方法，在保持类似标度的同时实现更高精度。

我们介绍了辅助场量子蒙特卡罗（AFQMC）的 Fortran 实现 QMCFort，它能够在中央处理器上高效执行大规模计算。该代码采用消息传递接口、OpenMP 和并行基础线性代数子程序库进行并行化，在典型系统上接近峰值性能（见图 2）。QMCFort 可以独立运行 AFQMC 模拟，也可以通过与 VASP 和 PySCF 的接口获取双电子中间量的乔列斯基向量。

我们使用 QMCFort，将 ph-AFQMC 方法的精度与“金标准”CCSD(T) 以及精度更高的 CCSDTQP 方法进行了比较。为此，我们计算了 HEAT 集中 26 个分子、苯分子和水团簇的 ph-AFQMC 能量。对于 HEAT 集，与 CCSDTQP 的结果相比，ph-AFQMC 的平均绝对偏差（MAD）为 1.85 mE_h (1.15 kcal/mol)。四个分子（CN、CO、CO_2, O_2）表现不佳，是造成这一偏差的主要原因。这类离群值显然令人担忧，因为在没有参考计算的情况下很难识别它们，

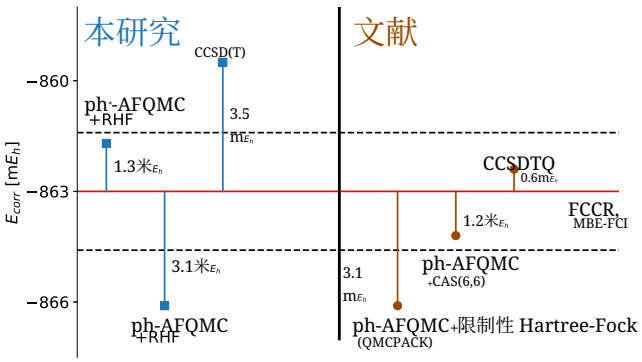


图 7. 使用 cc-pVDZ 基组和冻结芯近似计算的苯分子的相关能。ph-AFQMC(RHF)、ph*-AFQMC(RHF) 的相关能由 QMCFort 代码计算，而 CCSD(T) 能量由 PySCF 代码计算。[72] 其他数值取自参考文献 43 和 79。红线表示 $-863.0\,\mathrm{m}E_\mathrm{h}$ 的参考相关能，黑色虚线表示化学精度的范围（1 kcal/mol）。所有能量值均以 $\mathrm{m}E_\mathrm{h}$ 为单位。

出来。此外，对于 HEAT 集，CCSD(T) 无疑更加稳健；即使基态波函数涉及显著的双激发，它也能达到很高的精度。离群值的出现以及辅助场量子蒙特卡罗方法表现不佳的一种可能解释是，该方法无法准确处理强双激发。这表明，采用单行列式试探波函数的辅助场量子蒙特卡罗方法，只能处理缺乏强静态关联效应的弱关联体系，即使考虑双激发空间也是如此。

对于 HEAT 集中的其余分子，ph-AFQMC 的表现与 CCSD(T) 相当或更好。我们对无相位近似进行了修改——ph*-AFQMC——以至少部分克服 ph-AFQMC 中常见的相关作用过度问题。对于 CN、CO 和 CO_2分子，相关作用过度的影响尤为显著，改进后的方法确实显著改善了能量。对于苯分子，ph*-AFQMC 得出的相关能比全组态相互作用参考值高 $1.3\mathrm{m}E_\mathrm{h}$。这一结果明显优于 ph-AFQMC，且精度与采用多行列式 CAS(6,6) 试探波函数的 ph-AFQMC 相当。最后，水团簇的 AFQMC 结合能与 CCSD(T) 和二阶 Møller–Plesset 微扰理论的结合能相差不超过 0.5 kcal/mol。对于这四个水团簇，我们目前的结果总体上比以往的 AFQMC 结果更接近 CCSD(T) 结果，证实了我们的观点：采用单个斯莱特行列式的辅助场量子蒙特卡罗方法尤其适用于弱相关分子和材料。

ph*-AFQMC 和 ph-AFQMC 都是处理费米子符号问题以及噪声指数增长的相当临时性的方法。我们认为，目前的工作并未 conclusively 证明 ph* 近似优于 ph 近似。然而，目前的工作清楚地表明，即使对无相位近似进行微小改动，也会影响最终结果。尽管我们未能找到一种能够更普遍地改善结果的近似方法，但我们认为，有必要进一步开展这一方向的研究。

总而言之，ph-AFQMC 是一种前景广阔且可靠的方法，具有较高的精度和合理的计算成本。如果本研究所考察的分子具有弱相关性，该方法几乎可以达到化学精度。我们计划通过算法改进并采用更准确的试探波函数，进一步提升 QMCFort 的精度。

VI. 致谢

谨此感谢奥地利科学基金会（FWF）通过项目 P 33440 提供资助。所有计算均在 VSC4 / VSC5（维也纳科学集群）上完成。

作者声明

利益冲突

作者声明不存在任何利益冲突。

数据可用性

支持本研究结论的数据可在本文中获取。

附录：与 VASP 的接口

考虑一组平均场轨道 $\{\phi_p(\mathbf{r})\}$。单粒子哈密顿量矩阵元的计算方式如下：

$$h_{pq} = \int \mathrm{d}^3(\mathbf{r})\ \phi_p^*(\mathbf{r})\hat{H}_1\phi_q(\mathbf{r})\ , \tag{A.1}$$

其中，$\hat{H}_1$ 包含动能以及电子与原子核之间的库仑吸引作用。为了获得与乔列斯基向量$L_{g,pq}$等价的对象，我们引入双轨道密度

$$\rho_{\mathbf{G},pq} = \frac{1}{\sqrt{\Omega}} \int \mathrm{d}^3(\mathbf{r})\ e^{i\mathbf{G}\mathbf{r}}\ \phi_p^*(\mathbf{r})\phi_q(\mathbf{r})\ , \tag{A.2}$$

其中，$\{\mathbf{G}\}$ 是由截断能$\mathbf{G}^2/2 \le E_\mathrm{cut}$ 定义的倒格矢集合，Ω 是系统的体积。乔列斯基向量与双轨道密度之间的关系为

$$\sum_g^{N_g} L_{g,pq}L_{g,rs} = \sum_\mathbf{G}^{N_\mathbf{G}} \frac{4\pi}{\mathbf{G}^2}\ \rho_{\mathbf{G},pq}\rho_{\mathbf{G},rs}\ , \tag{A.3}$$

其中，$N_g \leq \min(N_{\mathbf{G}}, N^2)$。在参考文献78中，我们引入了一种截断奇异值分解来获得乔列斯基向量

$$L_{g,pq} = \mathrm{SVD}\left(\frac{\sqrt{4\pi}}{|\mathbf{G}|}\rho_{\mathbf{G},pq}\right) . \tag{A.4}$$

[1]W. Kohn 和 L. J. Sham，《物理评论》140，A1133（1965年）。[2]G. Kresse 和 J. Furthmüller，《物理评论B》54，11169（1996年）。[3]G. Kresse 和 D. Joubert，《物理评论B》59，1758（1999年）。[4]M.-C. Kim、H. Park、S. Son、E. Sim 和 K. Burke，《物理化学快报》6，3802 2015 26722874（），PubMed文献编号：。[5]A. J. Cohen、P. Mori-Sánchez 和 W. Yang，《科学》321，792（2008年）。[6]J. P. Perdew 和 K. Schmidt，AIP会议论文集577，1（2001年），https://aip.scitation.org/doi/pdf/10.1063/1.1390175。[7]R. Nagai、R. Akashi 和 O. Sugino，NPJ计算材料6（2020年），10.1038/s41524-020-0310-0。[8]R. Pederson、B. Kalita 和 K. Burke，自然综述·物理4，357（2022年）。[9]J. Behler 和 M. Parrinello，《物理评论快报》98，146401（2007年）。[10]M. S. Chen、J. Lee、H.-Z. Ye、T. C. Berkelbach、D. R. Reichman 和 T. E. Markland，《化学理论与计算杂志》0，null（0），PubMed文献编号：36730728。[11]A. Szabo 和 N. S. Ostlund，《现代量子化学：高级电子结构理论导论》（Dover出版社，纽约，美国，1996年）。[12]T. Helgaker、P. Jørgensen 和 J. Olsen，《分子电子结构理论》（John Wiley & Sons，有限公司，奇切斯特，2000年）。[13]N. Handy，《化学物理快报》74，280（1980年）。[14]A. A. Holmes、N. M. Tubman 和 C. J. Umrigar，《化学理论与计算杂志》12，3674（2016年），PubMed文献编号：27428771。[15]S. Sharma、A. A. Holmes、G. Jeanmairet、A. Alavi 和 C. J. Umrigar，《化学理论与计算杂志》13，1595（2017年），PubMed文献编号：28263594。[16]M. Dash、S. Moroni、A. Scemama 和 C. Filippi，《化学理论与计算杂志》14，4176（2018年），PubMed文献编号：29953810。[17]G. H. Booth、A. J. Thom 和 A. Alavi，《化学物理学杂志》131，054106（2009年）。[18]D. Cleland、G. H. Booth 和 A. Alavi，《化学物理学杂志》132，041103（2010年）。[19]K. Ghanem、K. Guther 和 A. Alavi，《化学物理学杂志》153，224115（2020年）。[20]G. D. Purvis 和 R. J. Bartlett，《化学物理学杂志》76，1910（1982年）。[21]R. J. Bartlett 和 M. Musiał，《现代物理评论》79，291（2007年）。[22]A. Tajti、P. G. Szalay、A. G. Császár、M. Kállay、J. Gauss、E. F. Valeev、B. A. Flowers、J. Vázquez 和 J. F. Stanton，《化学物理学杂志》121，11599（2004年）。[23]Y. J. Bomble、J. F. Stanton、M. Kállay 和 J. Gauss，《化学物理学杂志》123，054101 2005（）。[24]M. E. Harding、J. Vázquez、B. Ruscic、A. K. Wilson、J. Gauss 和 J. F. Stanton，《化学物理学杂志》128，114111（2008年）。[25]S. R. White 和 R. L. Martin，《化学物理学杂志》110，4127（1999年）。[26]U. Schollwöck，《现代物理评论》77，259（2005年）。[27]G. K.-L. Chan 和 S. Sharma，《物理化学年度评论》62，465（2011年），PubMed文献编号：21219144。[28]W. L. McMillan，《物理评论》138，A442（1965年）。[29]J. B. Anderson，《化学物理学杂志》65，4121（1976年）。[30]D. Ceperley、G. Chester 和 M. Kalos，《物理评论B》16，3081（1977年）。[31]W. Foulkes、L. Mitas、R. Needs 和 G. Rajagopal，《现代物理评论》73，33（2001年）。[32]N. Nemec、M. D. Towler 和 R. J. Needs，《化学物理学杂志》132，034111（2010年）。[33]L. A. Curtiss、K. Raghavachari、G. W. Trucks 和 J. A. Pople，《化学物理学杂志》94，7221（1991年）。[34]D. Pfau、J. S. Spencer、A. G. D. G. Matthews 和 W. M. C. Foulkes，《物理评论研究》2，033429（2020年）。[35]J. S. Spencer、D. Pfau、A. Botev 和 W. M. Foulkes，“更优、更快速的费米子神经网络”，（2020年），arXiv:2011。07125 [physics.comp-ph]。[36]L. Gerard、M. Scherbela、P. Marquetand 和 P. Grohs，“利用深度学习求解薛定谔方程的金标准解：我们需要多少物理知识？”（2022年）。[37]W. A. Al-Saidi、S. Zhang 和 H. Krakauer，《化学物理学杂志》127，144101（2007年）。[38]W. Purwanto、S. Zhang 和 H. Krakauer，《化学物理学杂志》130，094107（2009年）。[39]W. Purwanto、H. Krakauer、Y. Virgus 和 S. Zhang，《化学物理学杂志》135，164105 2011( )。[40]Y. Virgus、W. Purwanto、H. Krakauer 和 S. Zhang，物理评论B 86，241406 2012( )。[41]W. Purwanto、S. Zhang 和 H. Krakauer，《化学物理学杂志》142，064302 (2015)。[42]M. Motta、D. M. Ceperley、G. K.-L. Chan、J. A. Gomez、E. Gull、S. Guo、C. A. Jiménez-Hoyos、T. N. Lan、J. Li、F. Ma、A. J. Millis、N. V. Prokof’ev、U. Ray、G. E. Scuseria、S. Sorella、E. M. Stoudenmire、Q. Sun、I. S. Tupit- syn、S. R. White、D. Zgid 和 S. Zhang（西蒙斯多电子问题合作组织），《物理评论X辑》7，031059 (2017)。[43]J. Lee、F. D. Malone 和 D. R. Reichman，《化学物理学杂志》153，126101 (2020)。[44]J. Lee、F. D. Malone 和 M. A. Morales，《化学理论与计算杂志》16，3019 2020 32283932 ( )，PubMed文献编号：。[45]B. Rudshteyn、J. L. Weber、D. Coskun、P. A. Devlaminck、S. Zhang、D. R. Reichman、J. Shee 和 R. A. Friesner，《化学理论与计算杂志》18，2845 (2022)，PubMed文献编号：35377642。[46]J. E. Hirsch，物理评论B 31，4403 (1985)。[47]G. Sugiyama 和 S. Koonin，《物理年刊》168，1 (1986)。[48]S. R. White、D. J. Scalapino、R. L. Sugar、E. Y. Loh、J. E. Gubernatis 和 R. T. Scalettar，物理评论B 40，506 (1989)。[49]S. Sorella、S. Baroni、R. Car 和 M. Parrinello，《欧洲物理快报》8，663 (1989)。[50]S. Zhang、J. Carlson 和 J. E. Gubernatis，《物理评论快报》74，3652 (1995)。[51]S. Zhang、J. Carlson 和 J. Gubernatis，物理评论B 55，7464 (1997)。[52]S. Zhang 和 H. Krakauer，《物理评论快报》90，136401 (2003)。[53]M. Suewattana、W. Purwanto、S. Zhang、H. Krakauer 和 E. J. Walter，物理评论B 75，245123 (2007)。[54]M. Motta、J. Shee、S. Zhang 和 G. K.-L. Chan，《化学理论与计算杂志》15，3510 (2019)，PubMed文献编号：31091103。[55]J. Lee 和 D. R. Reichman，《化学物理学杂志》153，044131 (2020)。[56]F. D. Malone、S. Zhang 和 M. A. Morales，《化学理论与计算杂志》15，256 2019( )。[57]J. L. Weber、H. Vuong、P. A. Devlaminck、J. Shee、J. Lee、D. R. Reich- man 和 R. A. Friesner，《化学理论与计算杂志》18，3447 (2022)，PubMed文献编号：35507769。[58]S. Zhang，关联物质中的涌现现象：由于利希研究中心和德国模拟科学研究生院在于利希研究中心组织的秋季学校，2013年9月23日至27日；《关联电子秋季学校讲义（2013）》3 (2013)。[59]M. Motta 和 S. Zhang，《WIREs计算与分子科学》8，e1364 (2018)，https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1364。[60]H. Shi 和 S. Zhang，物理评论B 88，125132 (2013)。[61]C.-C. Chang、B. M. Rubenstein 和 M. A. Morales，物理评论B 94，235144 2016( )。[62]A. Mahajan 和 S. Sharma，《化学理论与计算杂志》17，4786 (2021)，PubMed文献编号：34232637。[63]A. Mahajan、J. Lee 和 S. Sharma，《化学物理学杂志》156，174111 (2022)。[64]W. J. Huggins、B. A. O’ Gorman、N. C. Rubin、D. R. Reichman、R. Babbush 和 J. Lee，《自然》603，416 (2022)。[65]E. J. Landinez Borda、J. Gomez 和 M. A. Morales，《化学物理学杂志》150，074105 2019( )。[66]N. H. F. Beebe 和 J. Linderberg，《国际量子化学杂志》12，683 (1977)，https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560120408。[67]H. Koch、A. Sánchez de Merás 和 T. B. Pedersen，《化学物理学杂志》118，9481 2003( )。[68]J. Hubbard，《物理评论快报》3，77 (1959)。[69]R. Stratonovich，载于《苏联物理学文献》，第2卷（1957年），第416页。[70]R. Balian 和 E. Brezin，《新物理学杂志B辑》64，37 (1969)。[71]如有合理需求，可向作者索取 QMCFort。[72]Q. Sun、T. C. Berkelbach、N. S. Blunt、G. H. Booth、S. Guo、Z. Li、J. Liu、J. D. McClain、E. R. Sayfutyarova、S. Sharma、S. Wouters 和 G. K.-L. Chan，《WIREs计算与分子科学》8，e1340 (2018)。[73]H. Trotter，10，545 (1959)。[74]M. Suziki，56，1454 (1976)。[75]M. Calandra Buonaura 和 S. Sorella，物理评论B 57，11446 (1998)。[76]F. Hummel、T. Tsatsoulis 和 A. Grüneis，《化学物理学杂志》146，124105 (2017)。

[77]P. E. Blöchl，物理评论B**50**，17953 (1994)。[78]A. Taheridehkordi，“采用投影缀加波方法的无相位辅助场量子蒙特卡罗：在固体中的应用”，(2023)，准备中的手稿。[79]J. J. Eriksen、T. A. Anderson、J. E. Deustua、K. Ghanem、D. Hait、M. R. Hoffmann、S. Lee、D. S. Levine、I. Magoulas、J. Shen、N. M. Tubman、K. B. Whaley、E. Xu、Y. Yao、N. Zhang、A. Alavi、G. K.-L. Chan、M. Head-Gordon、W. Liu、P. Piecuch、S. Sharma、S. L. Ten-no、C. J. Umrigar 和 J. Gauss，《物理化学快报》**11**，8922 (2020)，PubMed文献编号：33022176。[80] P. R. C. Kent、A. Annaberdiyev、A. Benali、M. C. Bennett、E. J. Landinez Borda、P. Doak、H. Hao、K. D. Jordan、J. T. Krogel、I. Kylän- pää、J. Lee、Y. Luo、F. D. Malone、C. A. Melton、L. Mitas、M. A. Morales、E. Neuscamman、F. A. Reboredo、B. Rubenstein、K. Saritas、S. Upadhyay，

G. Wang, S. Zhang, and L. Zhao，《化学物理学杂志》**152**, 174105（2020年）。[81]J. Kim, A. D. Baczewski, T. D. Beaudet, A. Benali, M. C. Bennett, M. A. Berrill, N. S. Blunt, E. J. L. Borda, M. Casula, D. M. Ceperley, S. Chiesa, B. K. Clark, R. C. Clay, K. T. Delaney, M. Dewing, K. P. Esler, H. Hao, O. Heinonen, P. R. C. Kent, J. T. Krogel, I. Kylänpää, Y. W. Li, M. G. Lopez, Y. Luo, F. D. Malone, R. M. Martin, A. Mathuriya, J. McMinis, C. A. Melton, L. Mitas, M. A. Morales, E. Neuscamman, W. D. Parker, S. D. P. Flores, N. A. Romero, B. M. Rubenstein, J. A. R. Shea, H. Shin, L. Shulenburger, A. F. Tillack, J. P. Townsend, N. M. Tubman, B. V. D. Goetz, J. E. Vincent, D. C. Yang, Y. Yang, S. Zhang, and L. Zhao，《物理学：凝聚态》**30**, 195901（2018年）。[82]B. Temelso, K. A. Archer, and G. C. Shields，《物理化学杂志 A》**115**, 12034（2011年），PubMed文献编号: 21910428，https://doi.org/10.1021/jp2069489。