Adaptive diagonal basis sets for electronic structure

SciCADE 2024

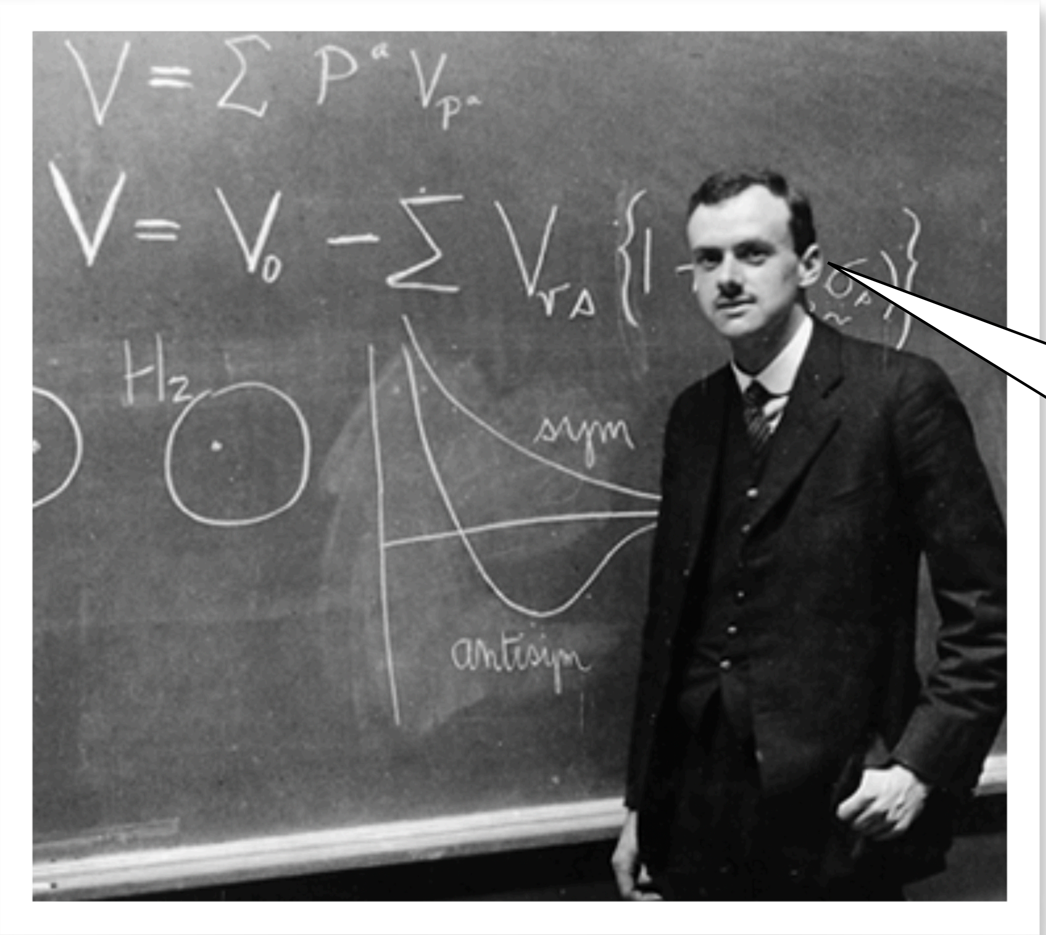
Numerical Methods for Quantum Many-Body Problems

July 17, 2024

Michael Lindsey

UC Berkeley

Electronic structure



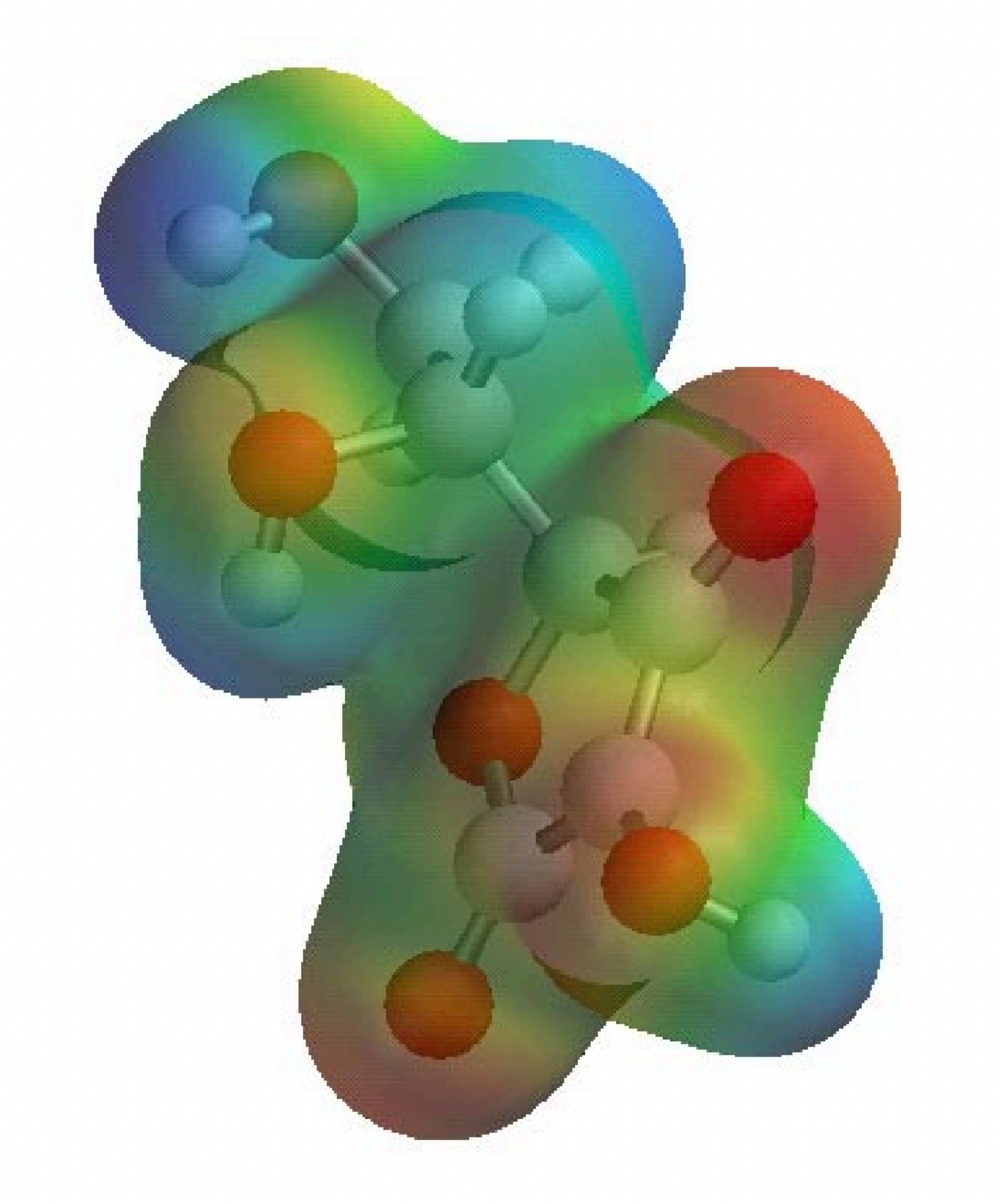
Schrödinger equation (1926)

$$i\hbar\,\dot{\Psi}(t) = \hat{H}\,\Psi(t)$$

> The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.
>
> **Paul Dirac (1929)**

Electronic structure

- In electronic structure, we consider nuclei with fixed positions and solve for the many-body ground state of the electrons
    - This is the lowest eigenfunction of a Schrödinger operator $\hat{H}$
- An oracle that can evaluate the map **{ nuclear coordinates } → { ground state energy }** suffices to determine:
    - equilibrium molecular geometries
    - molecular dynamics (in the **Born-Oppenheimer approximation**)
- 1998 Nobel Prize in Chemistry to **W. Kohn** (*density functional theory*) and **J. Pople** (*quantum chemistry basis sets*)



Electron density and electrostatic potential of the vitamin C molecule
(Source: nobelprize.org)

Many-body Schrödinger equation

- Electrons are fermions, so the wavefunction $\Psi = \Psi(x_1, \dots, x_N)$ is an antisymmetric function of *N* (number of electrons) space variables
- Can interpret $|\Psi(x_1, \dots, x_N)|^2$ as the probability density of finding the *N* electrons in these positions
- The marginal distribution $\rho(x) := N \int |\Psi(x, x_2, \dots, x_N)|^2 dx_2 \cdots dx_N$ is called the **electron density**

Many-body Schrödinger operator

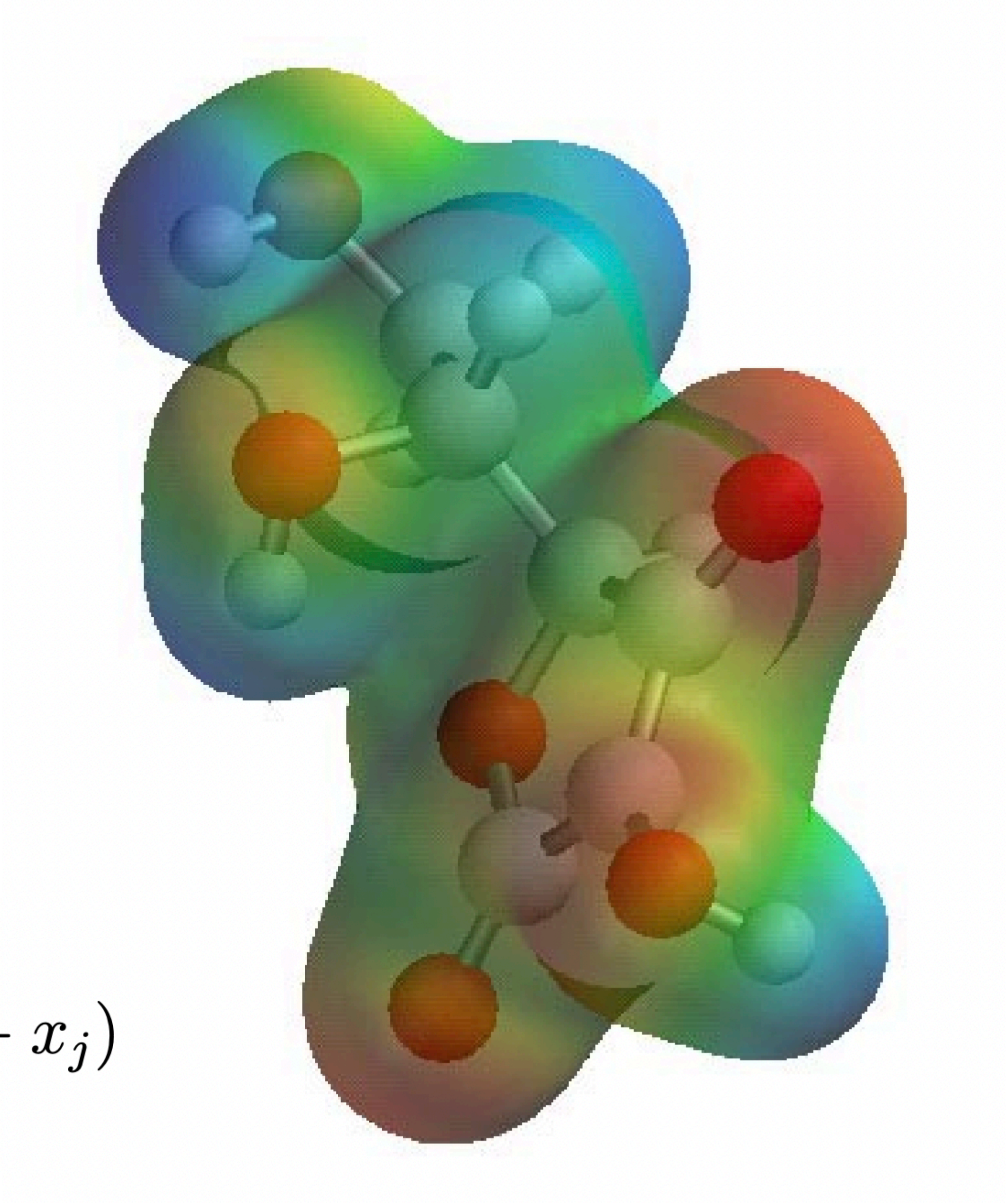
$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N} \Delta_{x_i} + \sum_{i=1}^{N} V^{(1)}(x_i) + \frac{1}{2}\sum_{i \neq j} V^{(2)}(x_i - x_j)$$

External potential

$$V^{(1)}(x) = -\sum_{I} \frac{Z_I}{|x - R_I|}$$

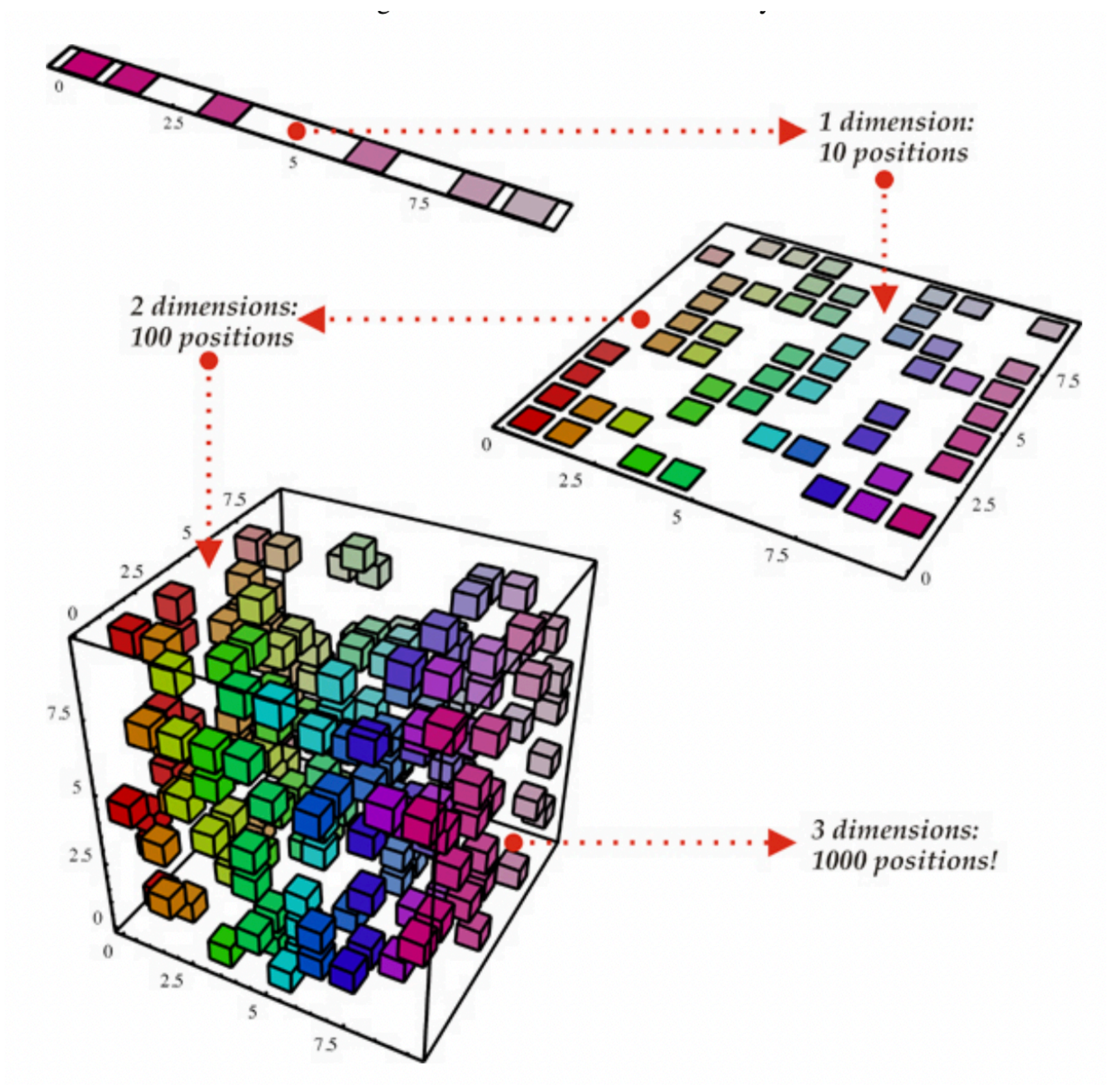
Electron-electron potential

$$V^{(2)}(x_1, x_2) = \frac{1}{|x_1 - x_2|}$$



Electron density and electrostatic potential of the vitamin C molecule (Source: nobelprize.org)

Discretization and curse of dimensionality



Curse of dimensionality
(Source: Bengio, 2015)

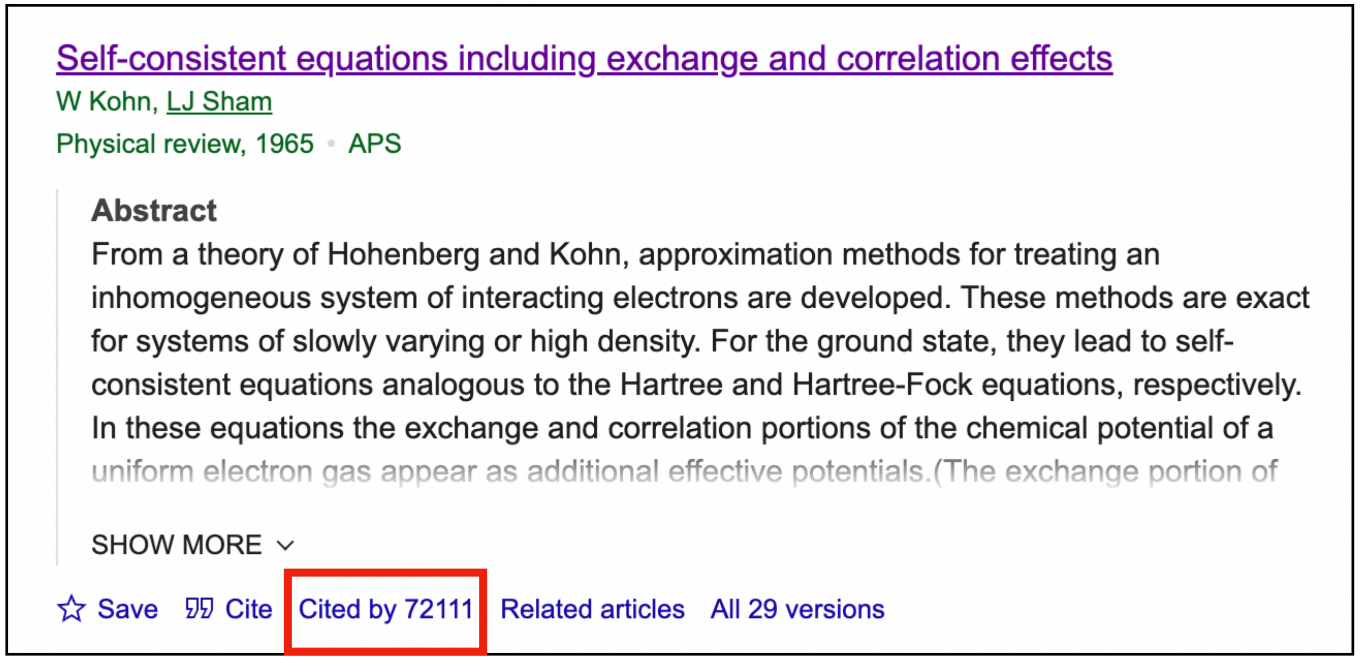
- The wavefunction $\Psi = \Psi(x_1, \ldots, x_N)$ is a function of 3N scalar variables!
  - Intractable to discretize directly (*curse of dimensionality*)
- We need a framework for discretizing each individual space variable, but we also need a framework for avoiding exponential complexity in *N*
  - This talk is about the ***former*** task, achieving structure that ***makes the latter task easier***

Single-particle methods

- **Single-particle methods** avoid the curse of dimensionality by considering the simplest possible antisymmetric state, a **Slater determinant**:

$$\Psi = \psi_1 \wedge \psi_2 \wedge \cdots \wedge \psi_N \qquad \Psi(x_1, \ldots, x_N) = \det([\psi_i(x_j)])$$

- The functions $\psi_k(x)$ are obtained as the lowest *N* eigenfunctions of an *effective one-body operator* $\hat{H}_{\text{eff}}$ acting on functions of a ***single space variable***
  - $\hat{H}_{\text{eff}} = \hat{H}_{\text{eff}}[\psi_1, \ldots, \psi_N]$ depends on the eigenstates → **nonlinear eigenvalue problem**
- **Hartree-Fock approximation** finds lowest-energy Slater determinant
- **Density functional theory (DFT)** expresses $\hat{H}_{\text{eff}} = \hat{H}_{\text{eff}}[\rho]$ where ρ is the electron density of the Slater determinant



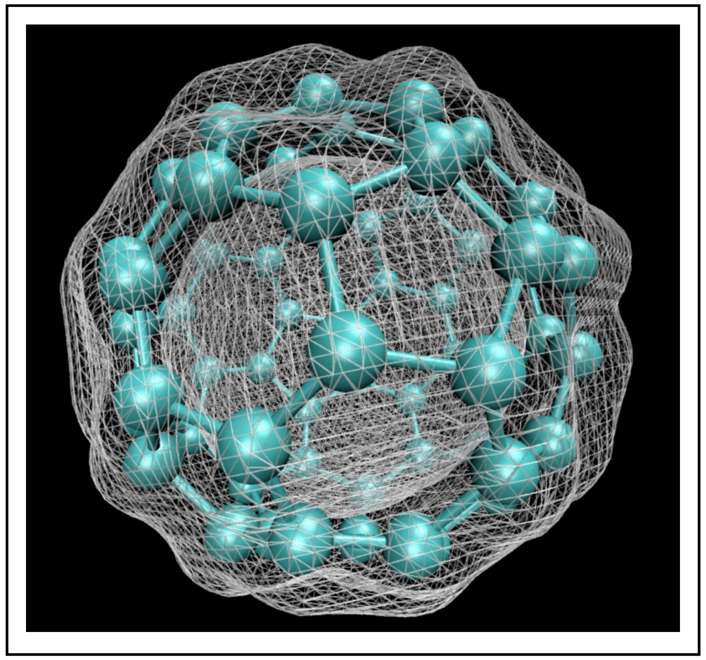
Self-consistent equations including exchange and correlation effects

W Kohn, LJ Sham

Physical review, 1965 · APS

Abstract

From a theory of Hohenberg and Kohn, approximation methods for treating an inhomogeneous system of interacting electrons are developed. These methods are exact for systems of slowly varying or high density. For the ground state, they lead to self-consistent equations analogous to the Hartree and Hartree-Fock equations, respectively. In these equations the exchange and correlation portions of the chemical potential of a uniform electron gas appear as additional effective potentials.(The exchange portion of

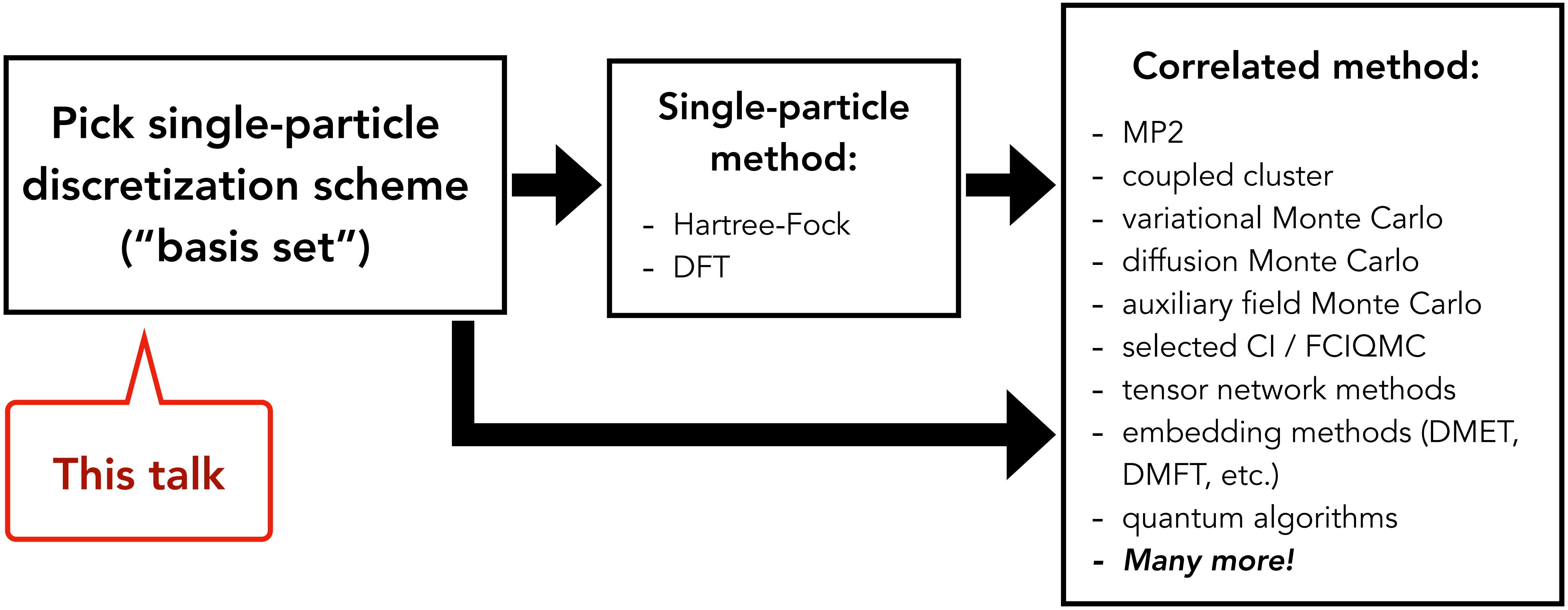


Electron density of C_{60} molecule computed via DFT

Source: Isaac Tamblyn

Quantum chemistry pipeline

- DFT is the most efficient and widely used quantum chemistry method
- **New frontier:** *strongly correlated systems* where DFT fails



Pick single-particle discretization scheme (“basis set”)

This talk

Single-particle method:

- Hartree-Fock
- DFT

Correlated method:

- MP2
- coupled cluster
- variational Monte Carlo
- diffusion Monte Carlo
- auxiliary field Monte Carlo
- selected CI / FCIQMC
- tensor network methods
- embedding methods (DMET, DMFT, etc.)
- quantum algorithms
- ***Many more!***

Basis sets

- First, choose a **single-particle basis**
    - Truncated orthonormal basis $\{\phi_i\}_{i=1}^M$ for $L^2(\mathbb{R}^3)$
- Second, consider the antisymmetric tensor product basis for $\bigwedge_{i=1}^N L^2(\mathbb{R}^3)$

$$|\Phi_I\rangle = \phi_{i_1} \wedge \phi_{i_2} \wedge \cdots \wedge \phi_{i_N}, \quad I = (i_1, \ldots, i_N), \quad 1 \le i_1 < i_2 < \cdots < i_N \le M$$

- Third, construct Galerkin projection:

$$H_{IJ} = \langle \Phi_I | \hat{H} | \Phi_J \rangle$$

- Finally, in principle, diagonalize $H = (H_{IJ})$
- Clever bookkeeping **(second quantization)** rewrites the truncated Hamiltonian as

$$\sum_{i,j=1}^{M} h_{ij}\, a_i^\dagger a_j + \frac{1}{2} \sum_{i,j,k,l=1}^{M} v_{ijkl}\, a_i^\dagger a_j^\dagger a_l\, a_k$$

- Here the **creation and annihilation operators** $a_i^\dagger$ and a_i carry the interpretation of adding and removing an electron from *i*-th basis state

Basis sets

- Tensors (h_{ij}) *and* (v_{ijkl}) completely encode the **quantum chemistry Hamiltonian**

$$\sum_{i,j=1}^{M} h_{ij}\, a_i^\dagger a_j + \frac{1}{2} \sum_{i,j,k,l=1}^{M} v_{ijkl}\, a_i^\dagger a_j^\dagger a_l\, a_k$$

- Even though we have discretized the continuum, we still have exponentially large dimension with respect to number of particle *N*
- However, the **1-electron and 2-electron integrals** can still be computed tractably:

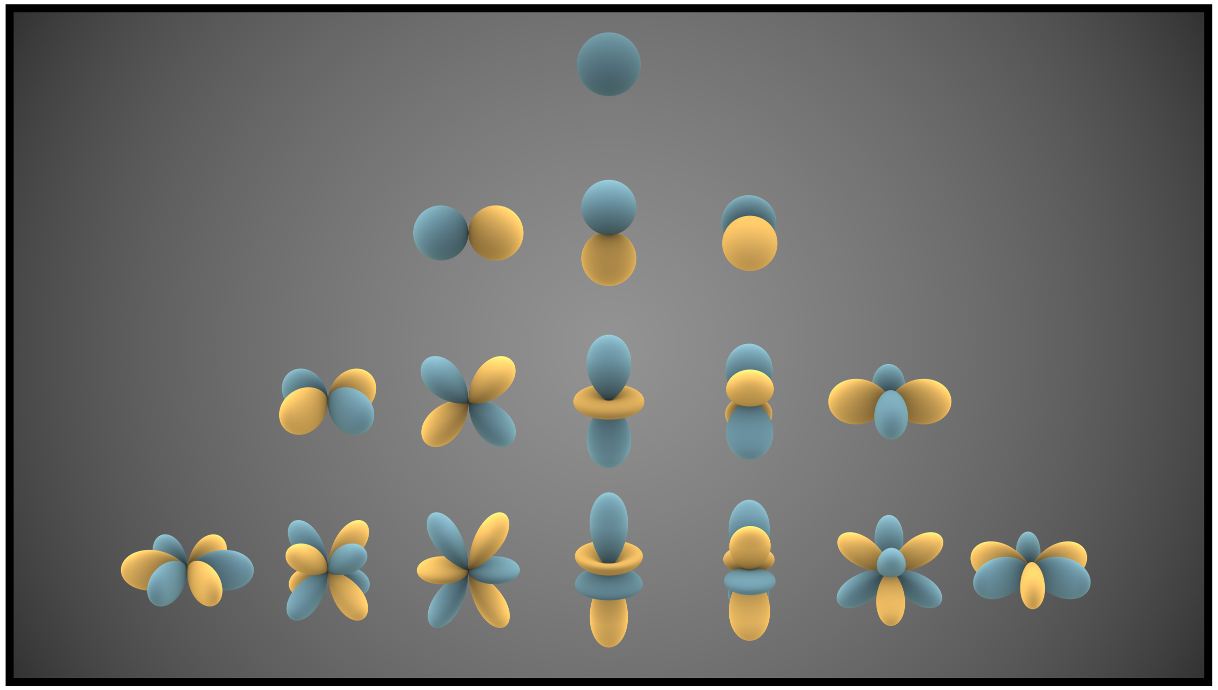
$$h_{ij} = \int \phi_i(x)\, [-\Delta + V^{(1)}(x)]\, \phi_j(x)\, dx$$

$$v_{ijkl} = \int \phi_i(x)\, \phi_j(x)\, V^{(2)}(x-y)\, \phi_k(y)\, \phi_l(y)\, dx\, dy$$

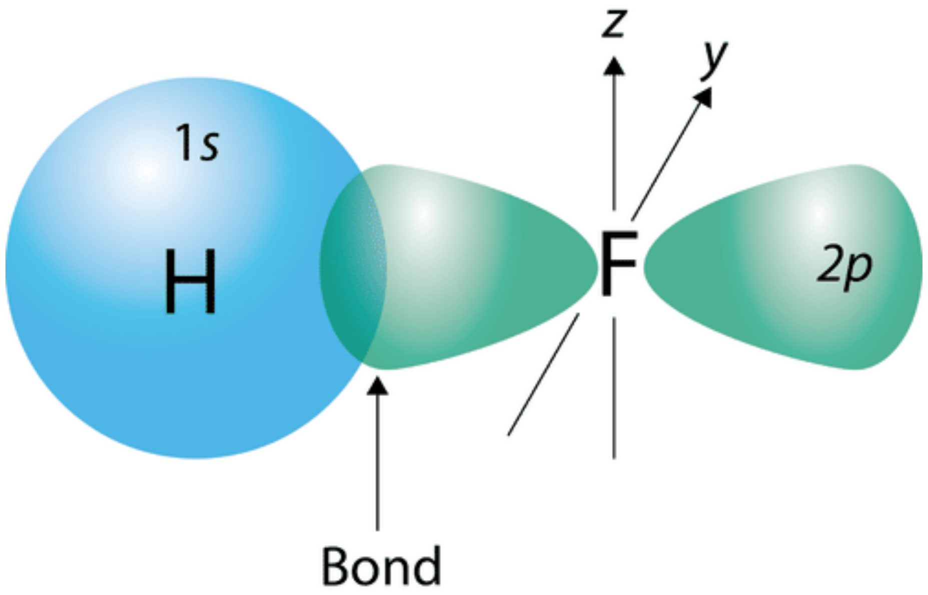
- Most quantum chemistry methods can handle any Hamiltonian presented in this form, you just need to hand them these integrals!

Quantum chemistry basis sets

- Since the early days of computational quantum chemistry **(cf. Pople)**, the standard single-particle basis sets are called **Gaussian-type orbitals** (**GTOs**)
- These in turn are inspired by **Slater-type orbitals** (**STOs**) which are based on eigenfunctions of the single-particle Schrödinger operator for the Hydrogen atom
  - Spherical harmonics times radius-dependent term of the form $e^{-\alpha r}$ (nonsmooth at the origin)
- Radial dependence with linear combinations of $e^{-\alpha_p r}$ radial dependencies → **GTOs**
  - Enables exact evaluation of 1-electron and 2-electron integrals



Spherical harmonics: *s*, *p*, *d*, and *f* orbitals



Molecular orbitals (eigenfunctions of effective single-particle operator) are linear combinations of atomic orbitals

Complete basis set limit

- GTOs can often work ***surprisingly*** well, but it is hard to reach the **complete basis set limit** (complete removal of error due to discretization)
  - Not a very systematic expansion away from atomic cores, and moreover computational scaling of storing v_{ijkl} is $\sim M^4$
  - Tensor compression strategies such as **density fitting** [Dunlap (2000)] and **tensor hypercontraction** [Hohenstein et al (2012)] can achieve cubic and quadratic memory scaling, with advantages depending on the downstream usage in a correlated calculation

- **Tradeoff:** highly adaptive to problem geometry, but almost no exploitable structure in the quantum chemistry Hamiltonian
  - Can we construct a basis that is **both adaptive** and **structured**

Why not something more systematic?

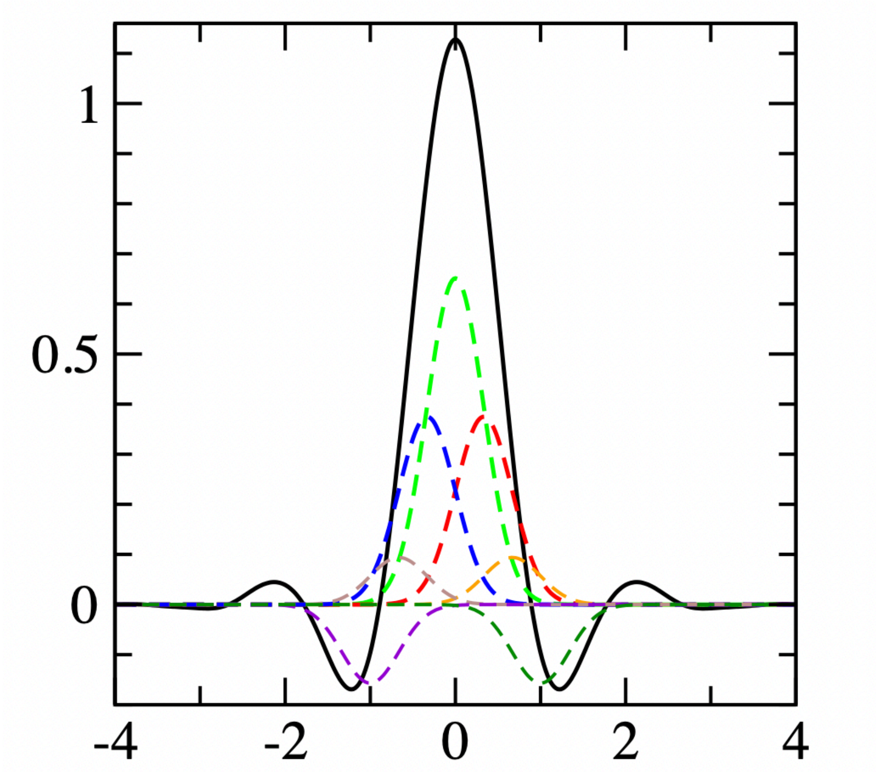
- Standard question from applied mathematicians: why not something more systematic, as in classical numerical PDEs?
  - Finite differences, finite elements, discontinuous Galerkin, Fourier basis (“planewaves”), multiresolution/wavelet approaches, …
- All of these approaches have shown success for single-particle methods, which resemble classical numerical PDE, and in fact planewaves are often used in the physics community
- Complicated set of competing desiderata for correlated methods:
  - **Need:** Galerkin formulation … ❌ *finite difference*
  - **Need:** orthogonality … ❌ *finite elements*
  - **Need:** high order of accuracy … ❌ *finite elements*
  - **Need:** differentiability … ❌ *discontinuous Galerkin, most multiresolution approaches*
  - **Need:** fast integrals … ❌ *finite elements, most multires approaches*
  - **Want:** adaptivity … ❌ *planewaves*
  - **Want:** structured 2-electron integrals … ❌ *GTOs, most multires approaches*
- Planewaves and GTOs satisfy needs but not all wants. ***Can we get everything we want?***

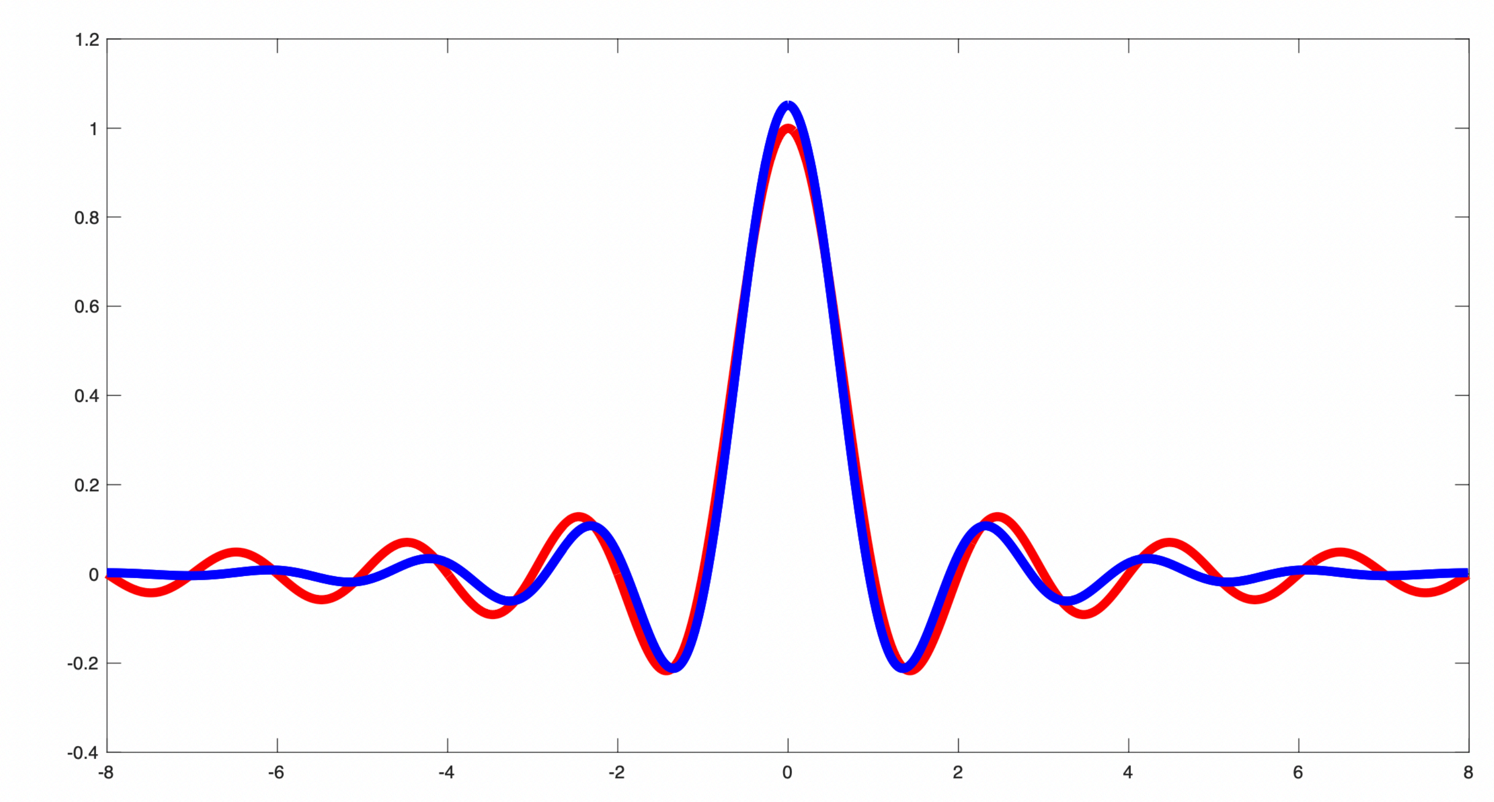
Diagonal basis sets

- One structured desideratum: ϕ_i **acts like a delta function** $w_i\,\delta_{x_i}$ when integrated against smooth functions (i.e., polynomials up to some order or bandlimited functions):

$$\int f(x)\,\phi_i(x)\,dx = f(x_i)\,w_i, \quad w_i := \int \phi_i(x)\,dx$$

- Hard to balance with completeness and orthogonality, but it can be done
  - **Sinc basis** (dual to planewaves), **Meyer wavelets** [cf. Daubechies book], **gausslets** [White (2017)]





Above: example **Gausslet**

Left: **Sinc function (red)** and **Meyer scaling function (blue)**

Fourier transform interpretation: *flat near zero frequency*. Either:

- exactly flat on interval, or
- flat Taylor series up to some order

Diagonal basis sets

- Formally apply with $u\,\phi_j$ in place of f

$$U_{ij} := \int \phi_i(x)\,u(x)\,\phi_j(x)\,dx \quad = \quad u(x_i)\,\phi_j(x_i)\,w_i \quad = \quad u(x_i)\int \phi_i(x)\,\phi_j(x)\,dx \quad = \quad u(x_i)\delta_{ij}$$

- Galerkin projection matrix of a diagonal operator is apparently diagonal!
    - But this derivation is not justified!
- Imagine we consider $\psi = \sum_j c_j \phi_j$ that is smooth compared to the resolution of the diagonal basis, then the coefficients of $u\,\psi$ in the $\{\phi_i\}$ basis are $\sum_j U_{ij}\,c_j$

$$\sum_j U_{ij}\,c_j \quad = \quad \int \phi_i(x)\,u(x)\,\psi(x)\,dx \quad = \quad u(x_i)\,\psi(x_i)\,w_i \quad = \quad u(x_i)\int \phi_i(x)\,\psi(x)\,dx \quad = \quad u(x_i)\,c_i$$

- Therefore $U = (U_{ij})$ **acts** like a diagonal matrix on the smooth part of $\mathrm{span}\{\phi_i\}$, which is where the wavefunctions of interest live anyway (the kinetic energy operator will keep us in the low-frequency subspace!)
- Taking $U = (U_{ij})$ to be diagonal is called the **diagonal approximation** and it can be viewed as a **pseudospectral method** [cf. nice textbook of Boyd (2000)]

Diagonal basis sets

- Similar reasoning applied to two-electron integrals yields diagonal structure

$$v_{ijkl} = v_{ik}\,\delta_{ij}\,\delta_{kl}$$

- In turn the quantum chemistry Hamiltonian has the simplified form :

$$\sum_{i,j=1}^{M} h_{ij}\,a_i^{\dagger} a_j + \frac{1}{2}\sum_{i,j=1}^{M} v_{ij}\,\hat{n}_i\,\hat{n}_j$$

  - Here $\hat{n}_i = a_i^{\dagger} a_i$ are the "number operators," which all commute
  - We retain the diagonal structure of the two-body interaction in our discretization
- My motivation for studying diagonal basis sets comes from a reduction of complexity in correlated methods that I am interested in studying:
  - variational Monte Carlo, auxiliary field Monte Carlo [work in progress]
  - tensor network methods [White and Stoudenmire (2019)], quantum algorithms [McClean et al (2020)]
- **Current drawback:** diagonal basis sets are grid-based and *not adaptive to molecular geometry*, so the size M of the basis set must be quite large (like planewaves)

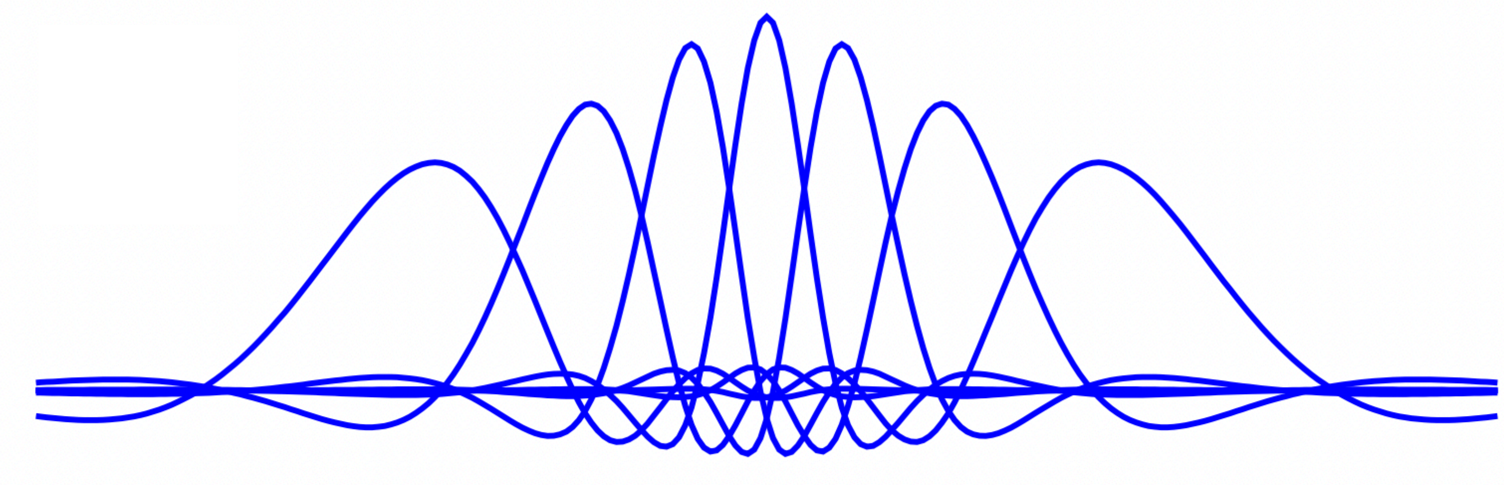
Method 1

Fast and spectrally accurate construction of adaptive diagonal basis sets for electronic structure, arXiv:2407.06171

Collaboration with **Sandeep Sharma** (UC Boulder)

Deformed diagonal basis sets

- **Idea:** deform a grid-based construction while preserving diagonality



- Given a diagonal basis set $\{\phi_i\}$ and an invertible deformation $T : \mathbb{R}^d \to \mathbb{R}^d$, it is not difficult to formally define a deformed diagonal basis set $\{\eta_i\}$

$$\eta_i(x) := \phi_i(T(x))\sqrt{J(x)}, \qquad J(x) := \det DT(x)$$

- However, there are two nontrivial challenges:
  - How to compute the deformation?
  - How to compute the 1- and 2-electron integrals?

Monge-Ampère equation

- Some related work [Gygi, *PRB* (1992); Gygi, *Eur. Phys. Lett.* (1993)] constructs heuristic adaptive deformations based on atom centers, though limited to DFT and it is difficult to guarantee invertibility
- We really want to **match a prescribed density** $\rho(x)$ of basis functions
  - Say on a box $[0,1]^3$
- The deformation T solves a **Monge-Ampère equation (MAE)**

$$\det DT(x) = \rho(x)$$

- Various applications, including optimal transport [cf. textbooks of Villani]
  - Solution is nonunique!
- We want a fast and spectrally accurate solver such that we can evaluate our approximate solution T and DT analytically (in order to compute 1- and 2-electron integrals later with high precision)
  - Many many existing approaches, usually to find unique solution $T = \nabla\phi$, such as [Benamou and Brenier (2000), Cuturi (2013), **ML** and Rubinstein (2017)], none quite adequate to our purposes

Knothe-Rosenblatt transport

- There is a unique solution of triangular type, called the **Knothe-Rosenblatt transport** [Rosenblatt (1952), Knothe (1957)] :

$$T(x, y, z) = (\, T_1(x), T_2(x, y), T_3(x, y, z)\,)$$

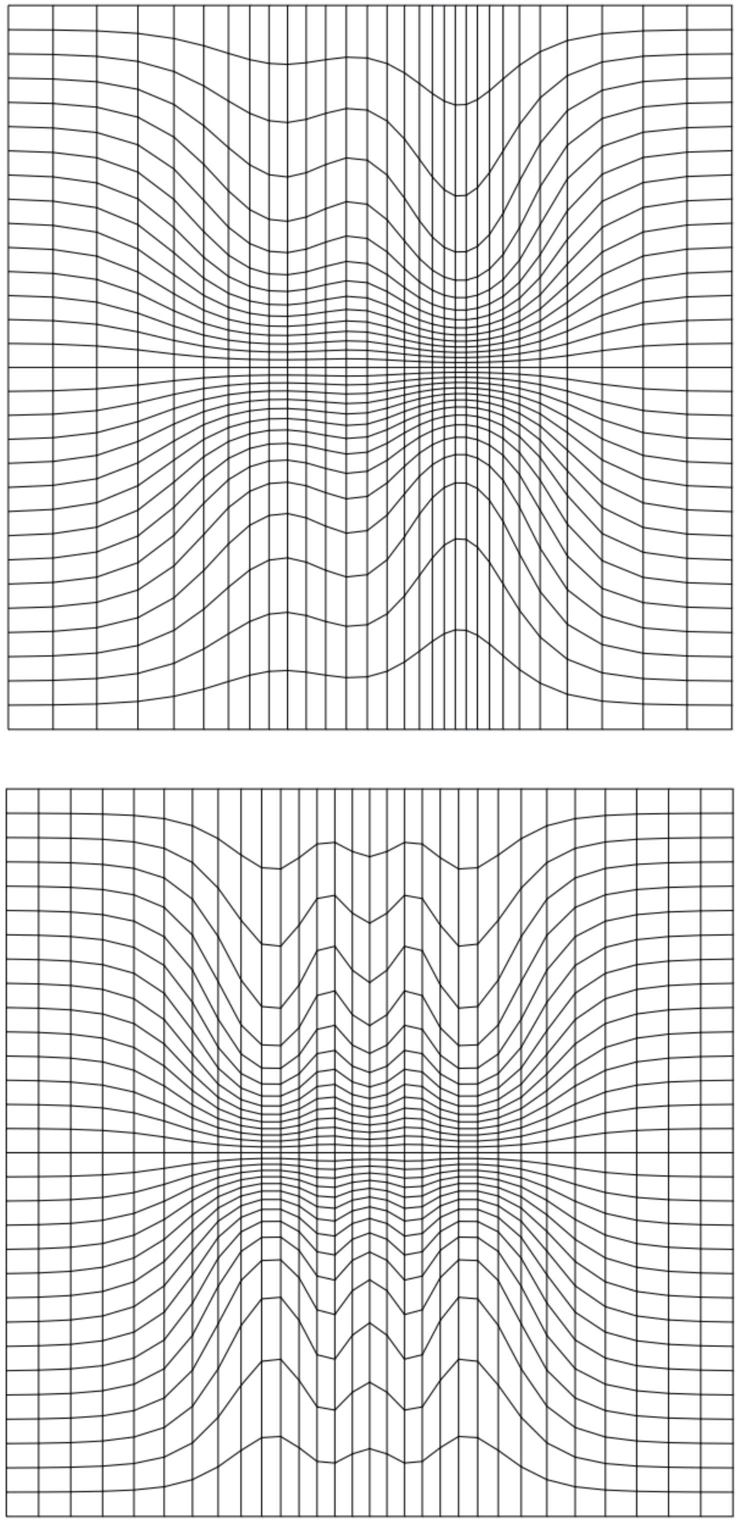
- DT is upper triangular and the map T can be computed semianalytically coordinate-by-coordinate by computing antiderivatives of conditional marginal distributions $\rho_1(x)$, $\rho_2(y\,|\,x)$, and $\rho_3(z\,|\,x, y)$
- **Observation:** we can analytically solve for the K-R transport if ρ is a sum of separable functions, such as a multivariate Chebyshev expansion

$$\rho(x) = \sum_{\alpha} c^{\alpha} g^{\alpha}(x), \qquad g^{\alpha}(x) = \prod_{i=1}^{d} g_i^{\alpha}(x_i)$$

Cyclic Knothe-Rosenblatt flow

Problem:

K-R transport treats coordinates asymmetrically, yields awkward distortion



Pictured: *y-z* and *x-z* coordinate lines of KR transport for a density concentrated at three `atom centers' in the *x-y* plane

Solution:

Cycle roles of coordinates in

$$T = T^{(N)} \circ \cdots \circ T^{(1)}$$

where each $T^{(n)}$ is only a small deformation, solving a MAE:

$$\det(DT^{(n)}) \propto g^{(n)}$$

The right-hand sides pull back to the reference

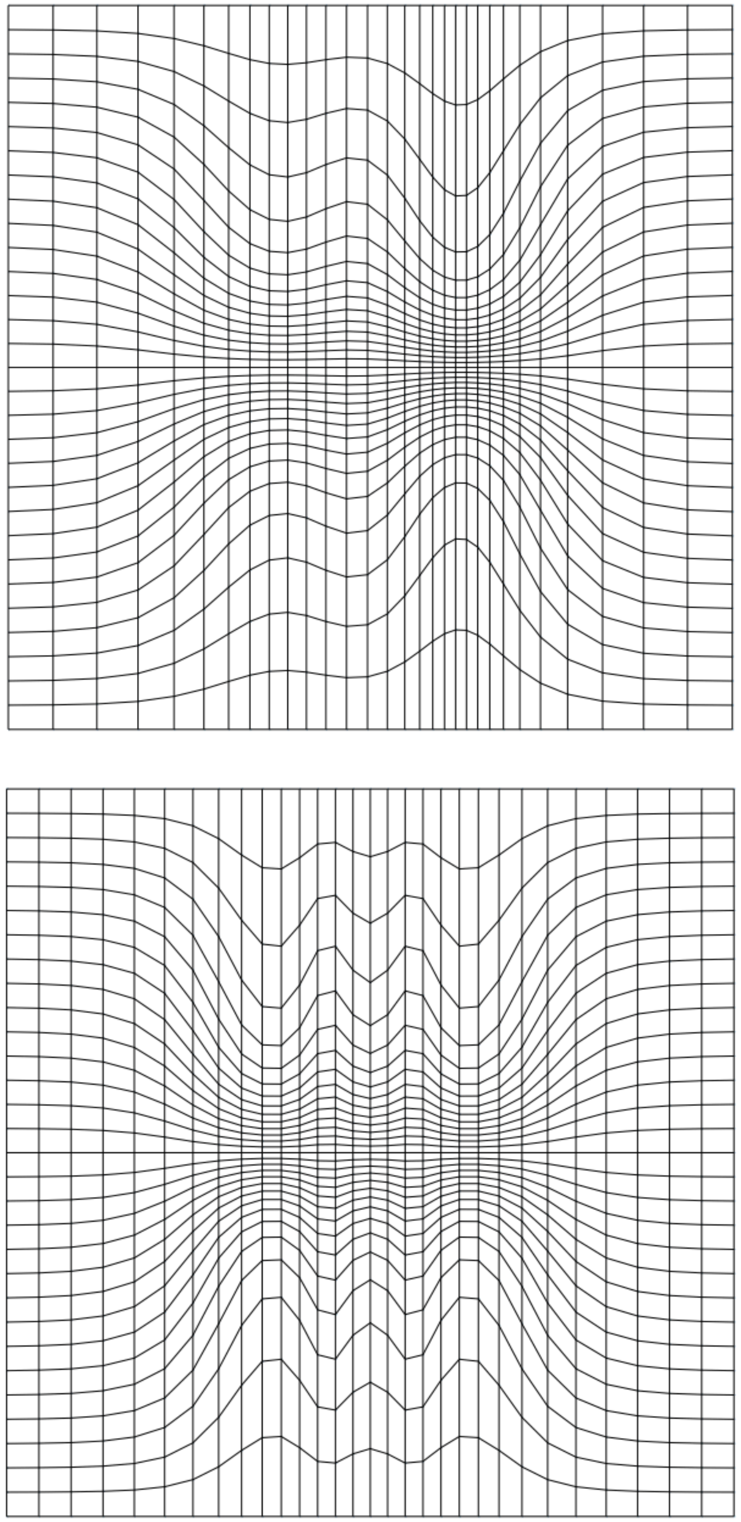
$$g^{(n)} \propto \rho^{1/N} \circ [T^{(n-1)} \circ \cdots \circ T^{(1)}]^{-1}$$

Use recursion

$$g^{(n)} \propto g^{(n-1)} \circ \left[T^{(n-1)} \right]^{-1}$$

to fit each $g^{(n)}$ with a Chebyshev interpolation

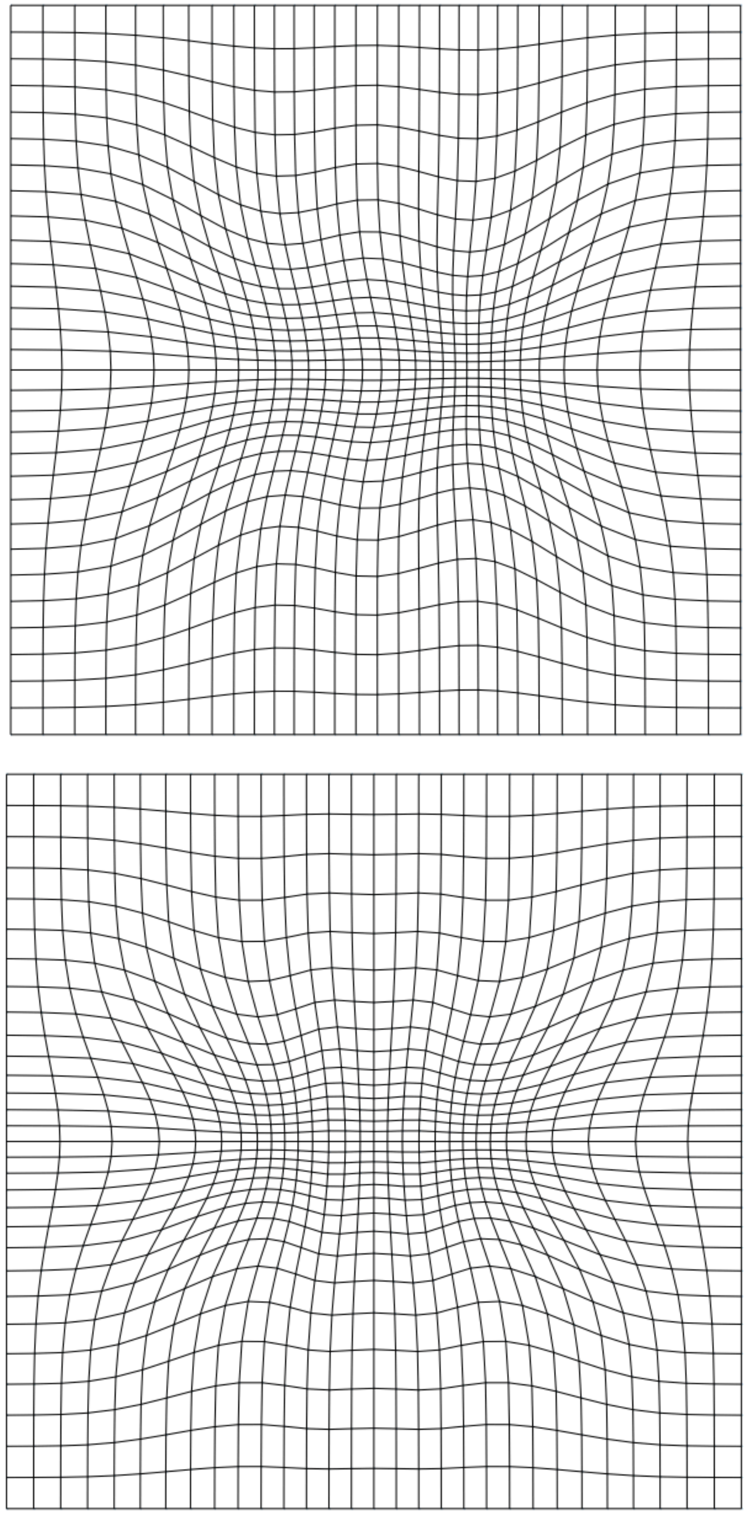
Cyclic Knothe-Rosenblatt flow



Obtain better deformation!

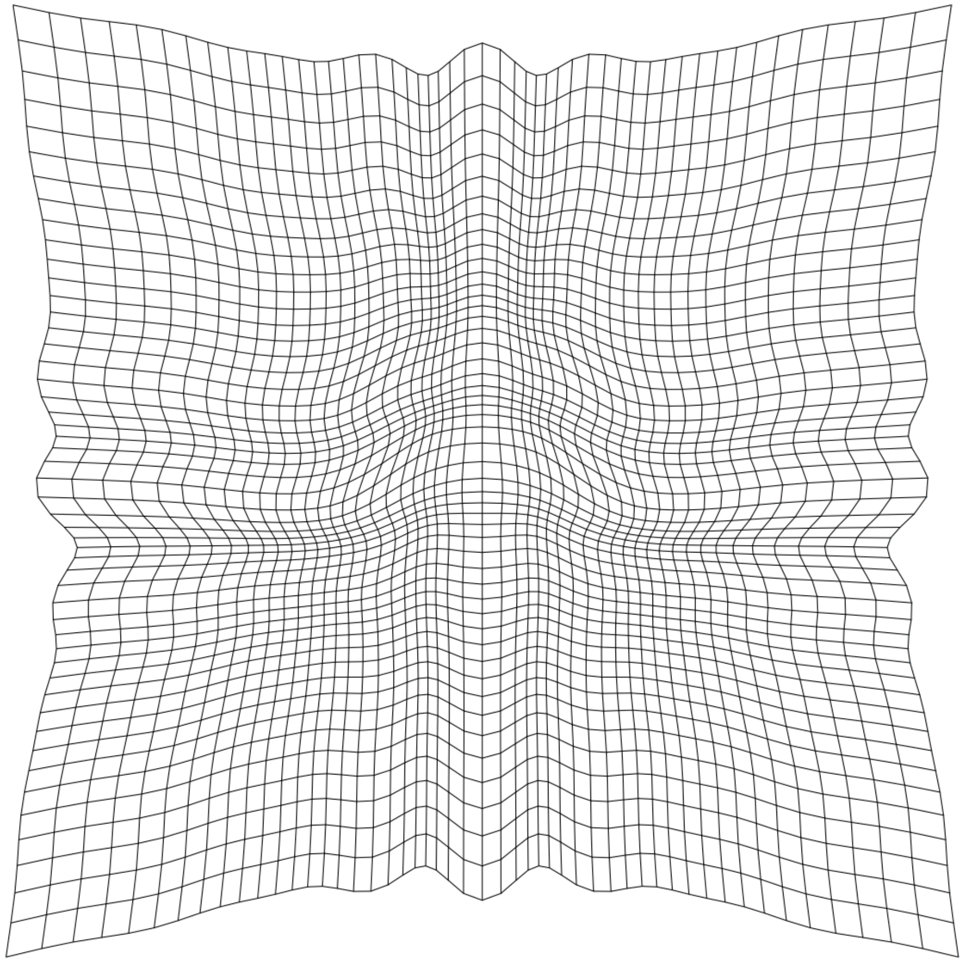
Skipping numerical details (NUFFTs for almost linear scaling performance, extension to periodic case, etc.)

Pictured: same coordinate lines for KR transport (left) and cyclic KR flow (right)



Self-consistent inverse formulation

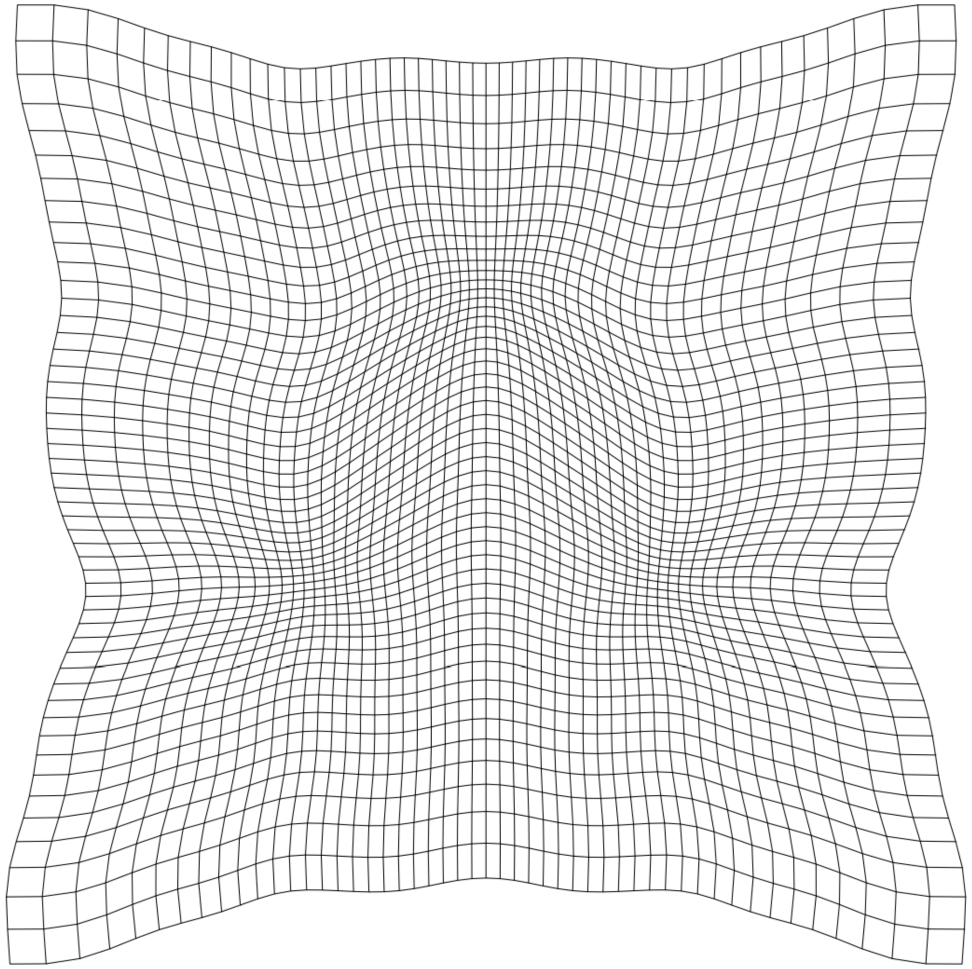
- One more complication: for sharper deformations, we apparently require a finer discretization in order to resolve all the features of $\rho(x)$
  - In fact, our approach can be viewed as a collocation method for the MAE
  - Significant aliasing error if the grid cannot keep up with the deformation
- Instead, consider computing the inverse map $S(y) = T^{-1}(y)$, which solves the inverse MAE $\det DS(y) = 1/\rho(S(y))$
- Can be reduced to a sequence of MAEs with fixed RHS via self-consistent iteration (cf. **ML** and Rubinstein (2017))
  - Determine $S^{(k+1)}$ given $S^{(k)}$ by solving $\det DS^{(k+1)}(x) = 1/\rho(S^{(k)}(x))$



Pictured: cutout of *x*-*y* coordinate lines of cyclic KR flow for a very sharp deformation (over 3 orders magnitude of variation in the target density). Visible aliasing error is present for this discretization and is expensive to remove.

Self-consistent inverse formulation

- One more complication: for sharper deformations, we apparently require a finer discretization in order to resolve all the features of $\rho(x)$
  - In fact, our approach can be viewed as a collocation method for the MAE
  - Significant aliasing error if the grid cannot keep up with the deformation
- Instead, consider computing the inverse map $S(y) = T^{-1}(y)$, which solves the inverse MAE $\det DS(y) = 1/\rho(S(y))$
- Can be reduced to a sequence of MAE's with fixed RHS via self-consistent iteration (cf. **ML** and Rubinstein (2017))
  - Determine $S^{(k+1)}$ given $S^{(k)}$ by solving $\det DS^{(k+1)}(x) = 1/\rho(S^{(k)}(x))$



Pictured: aliasing error is removed using the self-consistent inverse formulation (same computational grid size).

Pseudospectral approximation

- Now we have a deformation and a diagonal basis

$$\eta_i(x) := \phi_i(T(x))\sqrt{J(x)}, \qquad J(x) := \det DT(x)$$

- We choose $\{\phi_i\}$ to be a basis of periodic sinc functions, ρ to be electron density
- Need to compute 1-electron and 2-electron integrals
- Due to diagonal approximation, only need to compute diagonal terms for 1- and 2-body potentials
- **Kinetic matrix** (matrix of Laplacian operator in our single-particle basis)

$$\mathbf{T}_{ij} = -\int \eta_i(x)\Delta\eta_j(x)\,dx$$

- By plugging in **pseudospectrally accurate** approximation $\eta_i(x) \approx \phi_i(T(x))\sqrt{J(x_i)}$ we can perform matvecs by $\mathbf{T}$ in $O(M\log M)$ time using FFTs
- All other key operations can be expressed in terms of these matvecs

Pseudospectral approximation

Also need diagonal nuclear integrals

$$\int \frac{\eta_i^2(x)}{|x - R_I|}\,dx \quad \longrightarrow \quad \int \frac{\eta_i(x)}{|x - R_I|}\,dx = -\Delta^{-1}\eta_i(R_I)$$

And diagonal electron repulsion integrals

$$\mathbf{v}_{ik} = \int\int \eta_i^2(x)\frac{1}{|x-y|}\eta_k^2(y)\,dx\,dy \quad \longrightarrow \quad w^{-2}\sqrt{J_iJ_k}\int\int \eta_i(x)\frac{1}{|x-y|}\eta_k(y)\,dx\,dy$$

$$= w^{-2}\sqrt{J_iJ_k}\left[-\int \eta_i(x)\Delta^{-1}\eta_k(x)\,dx\right]$$

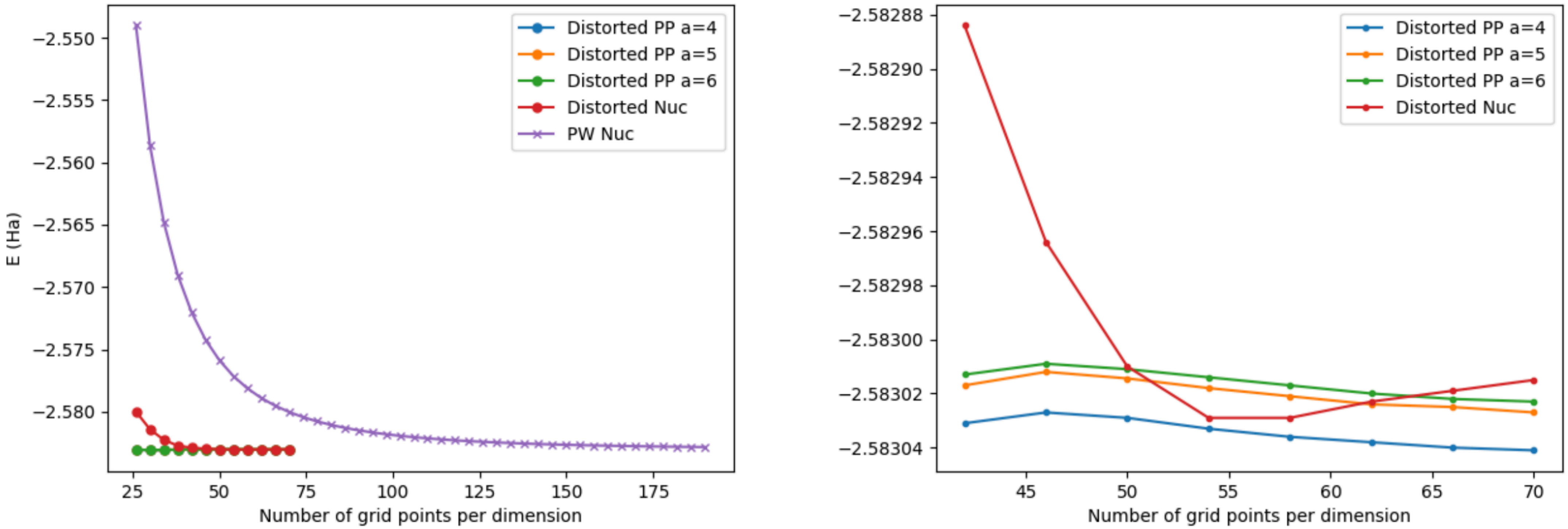
We have used the pseudospectral substitution

$$\eta_i\eta_j \approx w^{-1}\sqrt{J_i}\,\eta_i$$

Meanwhile $\mathbf{T}^{-1}$ approximates the matrix of $-\Delta^{-1}$ in our basis $\{\eta_i\}$

- Hence nuclear and electron repulsion integrals can be constructed in terms of linear solves in $\mathbf{T}$ (fast matvecs available, so fast inversion as well via CG)

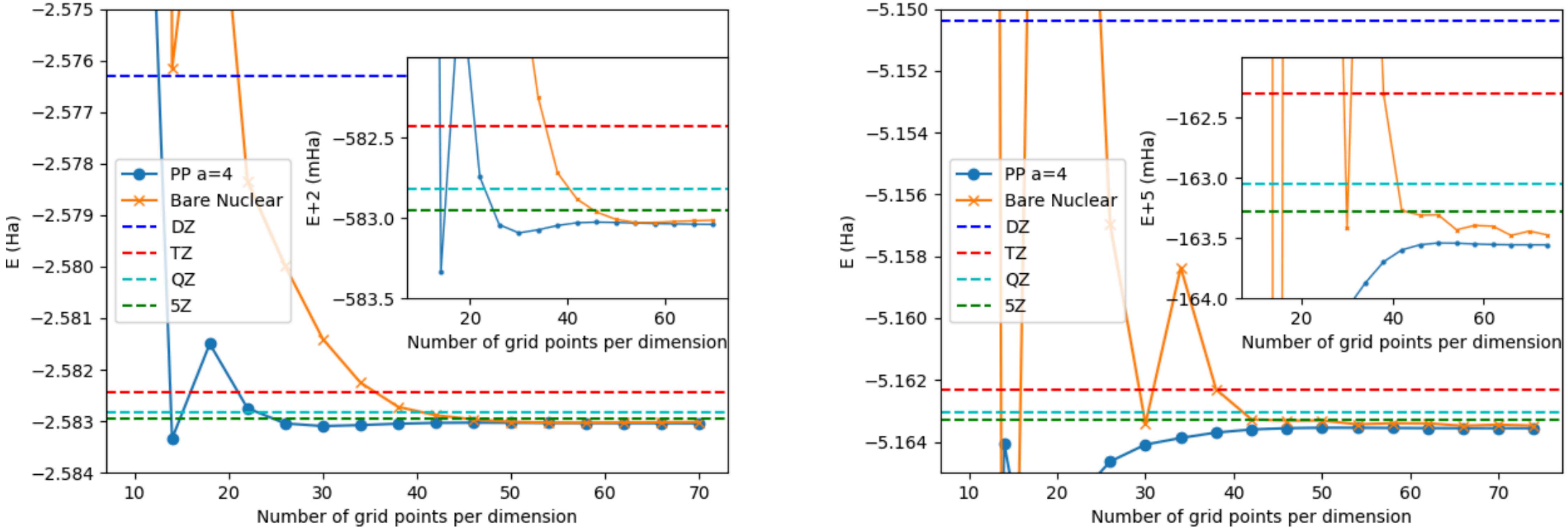
Results: helium atom



Pictured: energy convergence of Hartree-Fock approximation for helium atom, compared to planewaves. We use the smooth all-electron pseudopotential (PP) of Gygi (2023) as an alternative to dealing with the bare nuclear potential directly.

See the preprint for details of almost-linear scaling HF implementation, specification of basis function density, etc.

Results: helium dimer



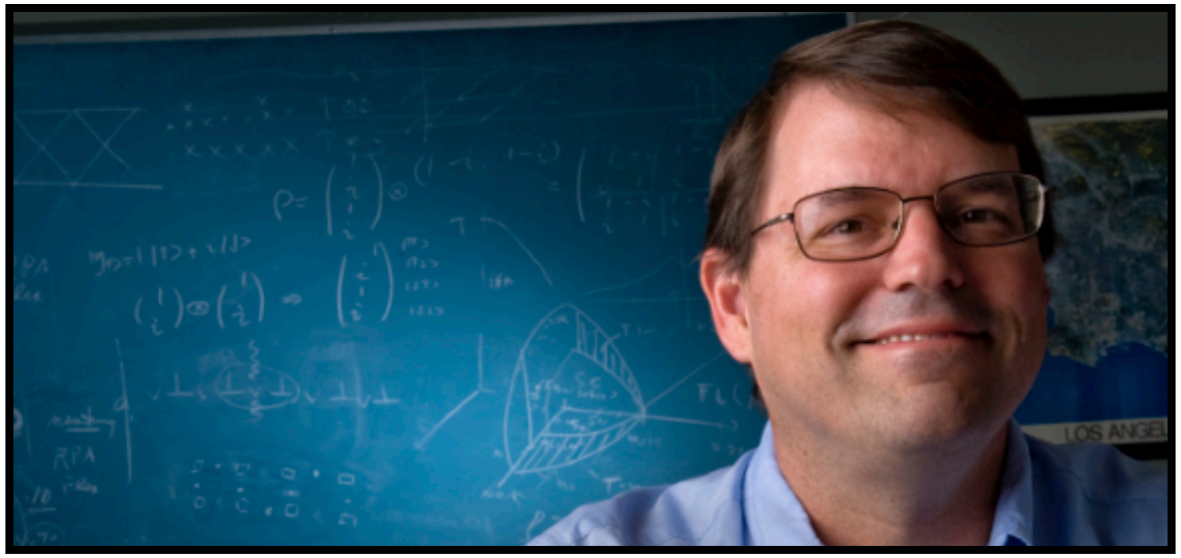
Pictured: energy convergence of Hartree-Fock approximation for helium dimer, compared to standard GTO basis sets. Note that not many more additional basis functions are needed compared to helium atom.

Other molecules are considered in the preprint!

Method 2

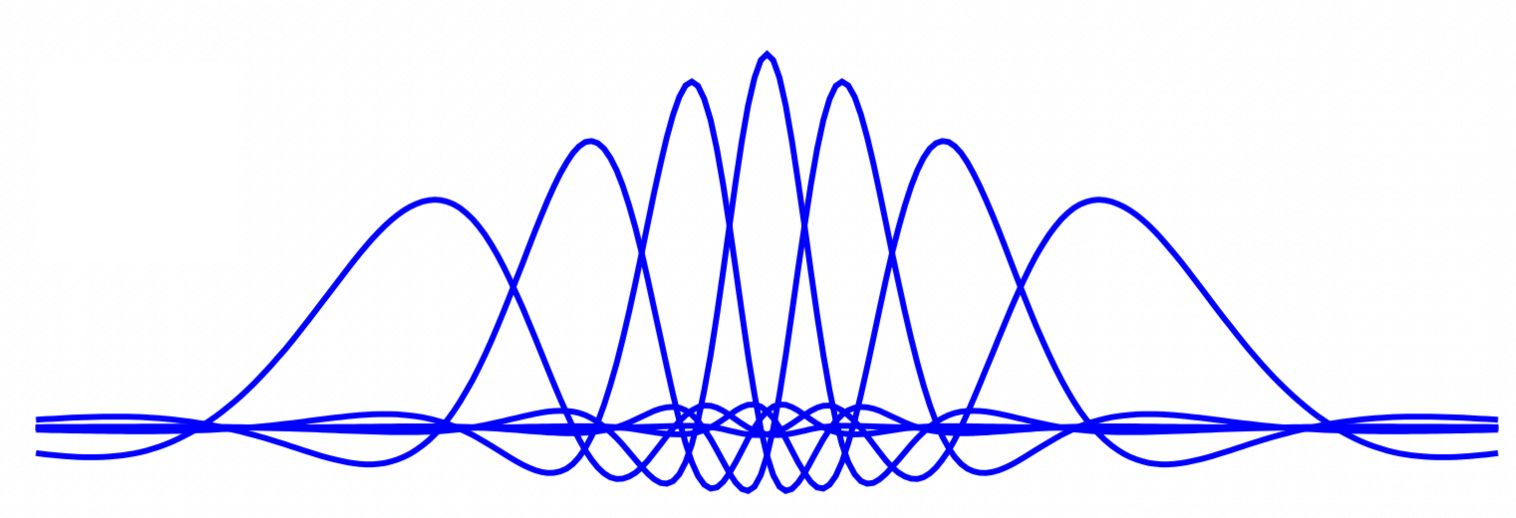
Nested gausslet basis sets, J. Chem. Phys. (2023)

Collaboration with **Steve White** (UC Irvine)



1D diagonal basis sets

- **Motivation:** four competing properties for a basis set $\{\phi_i\}$ of 1D variable
  - **Completeness (C):** $\{\phi_i\}$ can represent polynomials up to some order
  - **Orthonormality (O):** $\{\phi_i\}$ are orthonormal
  - **Moment (M):** ϕ_i integrates like the delta function $w_i\,\delta_{x_i}$ against polynomials up to some order — this is the ***diagonal property***!
  - **X-diagonalization (X):** the matrix of the diagonal operator *x* in the $\{\phi_i\}$ basis is diagonal, i.e., $\int \phi_i(x)\, x\, \phi_j(x)\, dx \;=\; x_i\,\delta_{ij}$



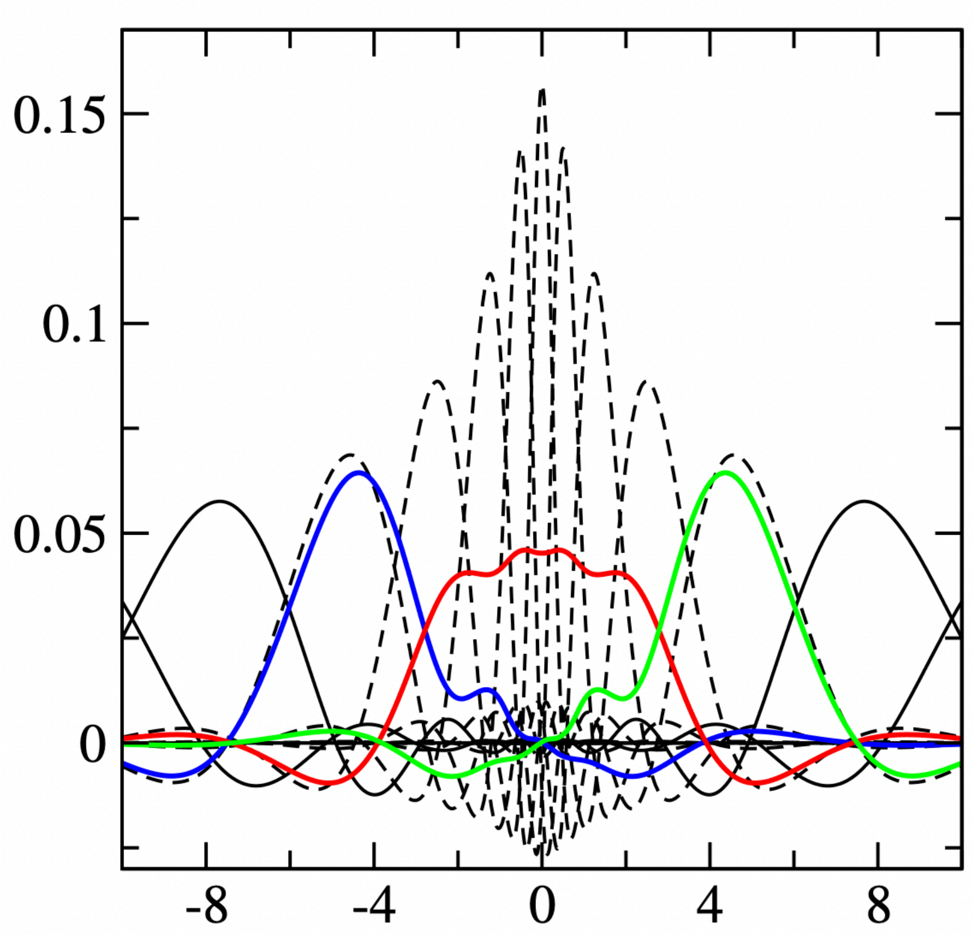
COMX theorem

- **Theorem** (Steve White and M.L.): if **C**, **O**, and **X,** then **M**.
  - More specifically: **C** up to order *p* yields **M** up to order *p*+1.

- Think of **C** as a property of the **subspace** spanned by $\{\phi_i\}$
  - A unique **O**rthogonal and **X**-diagonal basis spanning this subspace can be constructed by diagonalizing the *x* operator on the subspace
  - Thus **O** and **X** come for free
  - Theorem implies **M** (diagonal property) comes for free as well from this construction

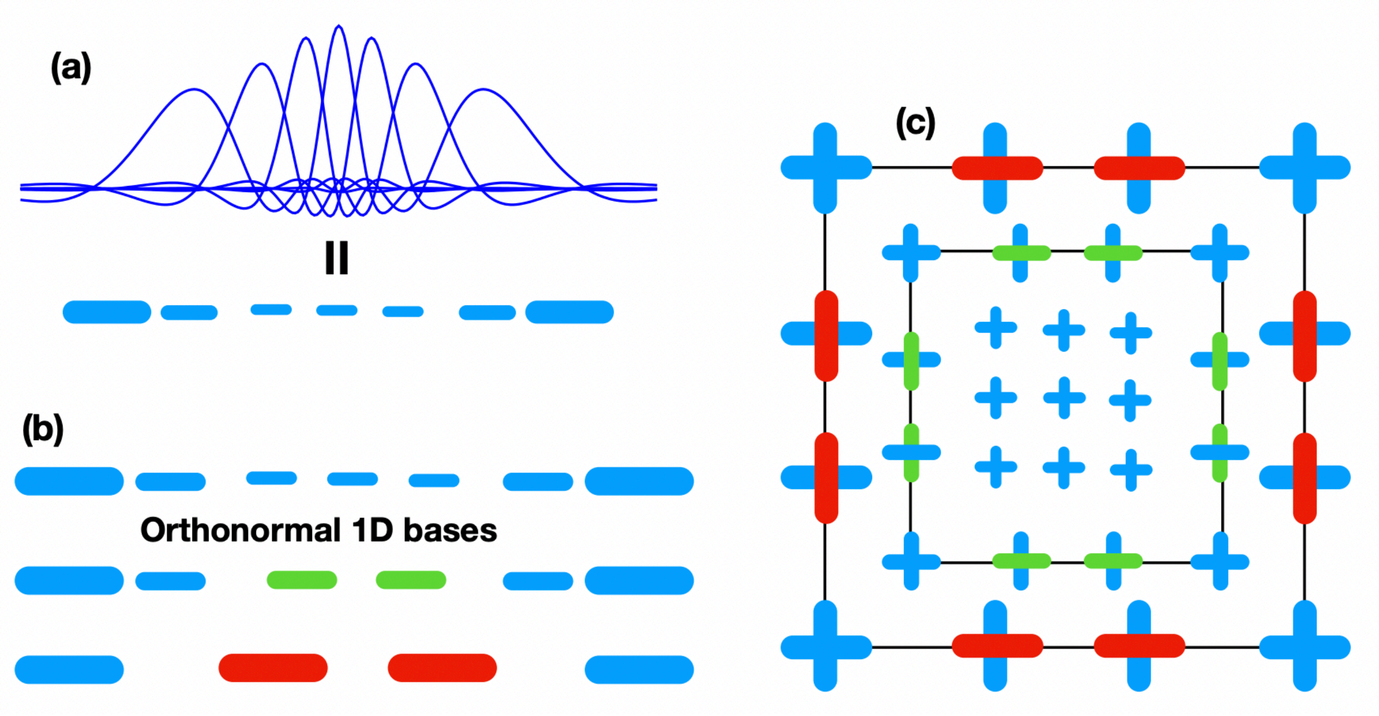
COMX theorem

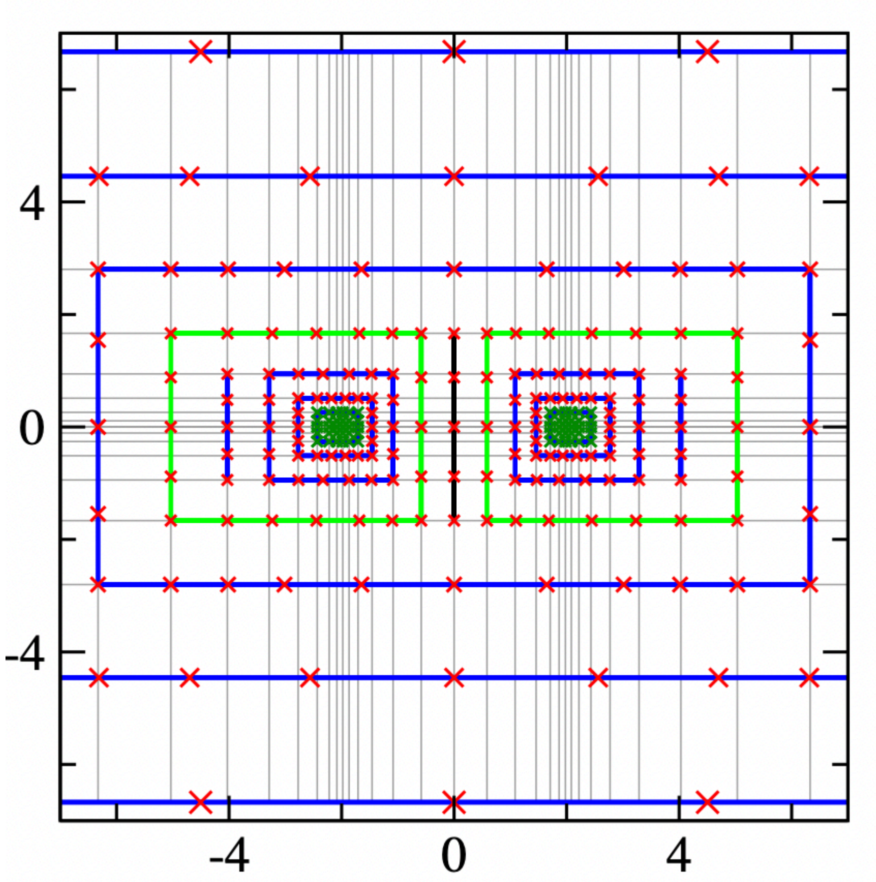
- **Implication:** it is easy to construct COMX (diagonal) basis sets in 1D with desired spans
- Start with 1D diagonal basis, coarsen the basis in some region, then X-diagonalize to get a COMX basis



Nested gausslets

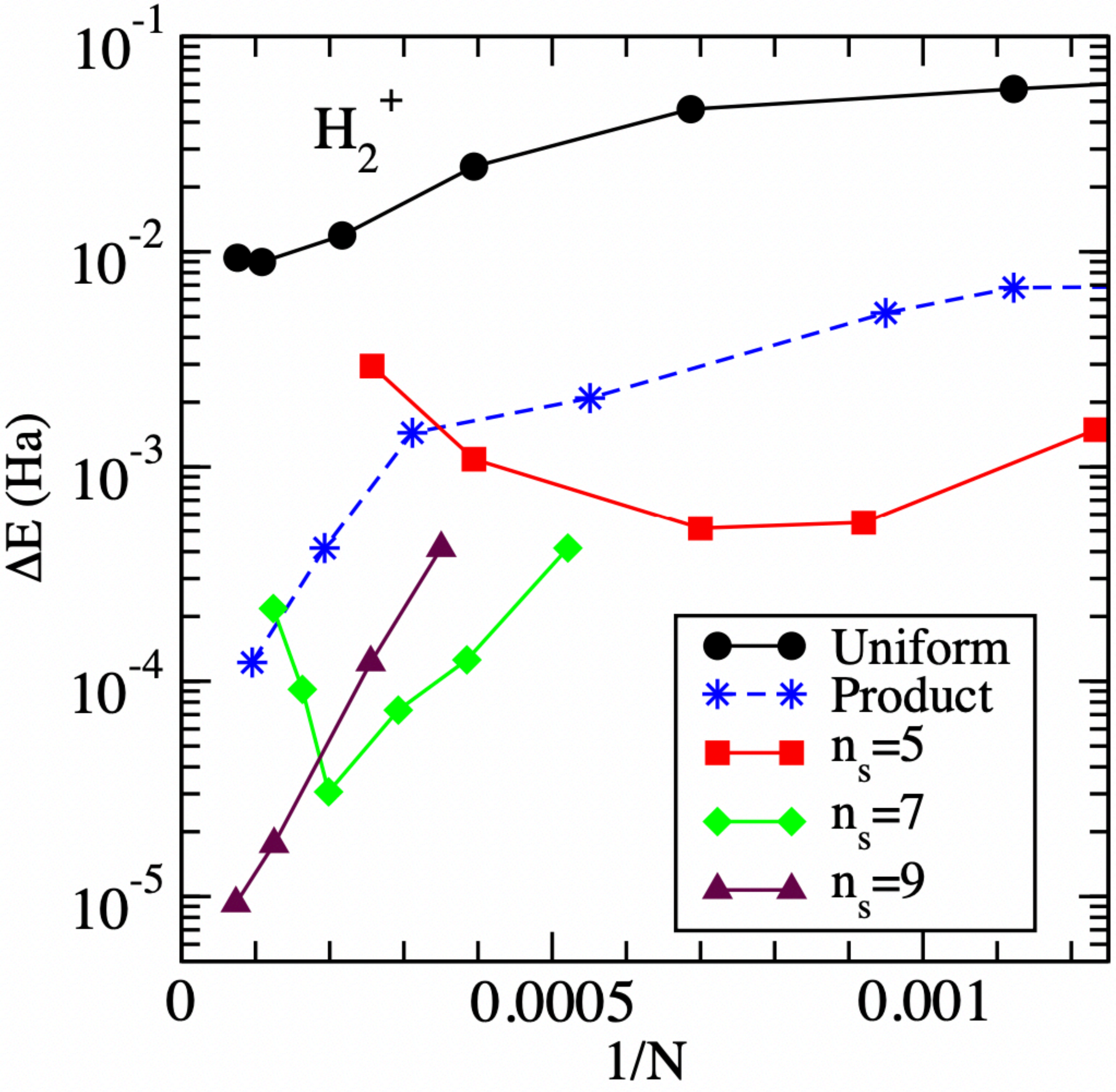
- **Nested gausslets (NG):** intricate construction of product functions based on taking a 1D COMX backbone and coarsening inner functions to produce more 1D COMX bases at different levels



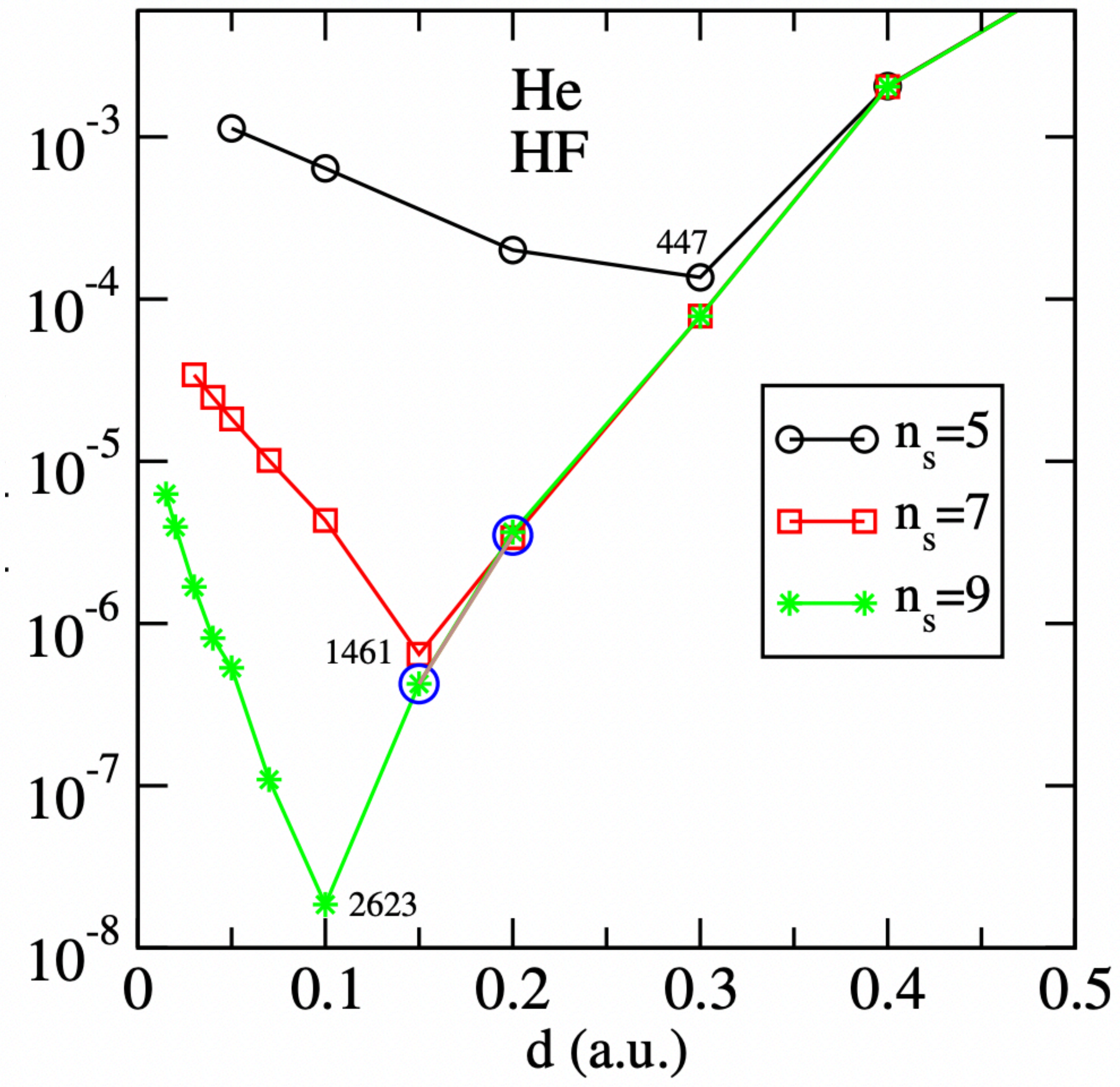


- Highly efficient (remove unnecessary basis functions near the edge)
- Can be combined with **GTOs** while still preserving the diagonal approximation!

Nested gausslets



Energy error for H_2 positive ion in NG basis

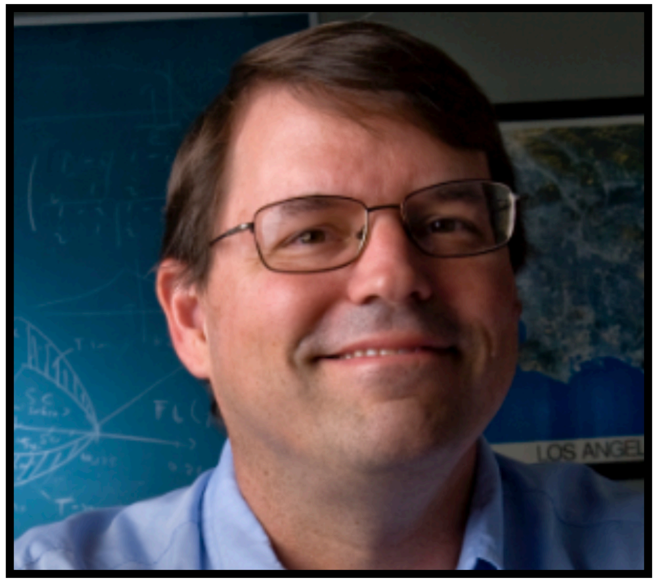


Hartree-Fock energy error for Helium atom in NG/GTO basis

Conclusions

- **Mass transport + pseudospectral approximation:**
  - Fully automatic procedure
  - Allows implicit (matrix-free) access to 1- and 2-electron scaling, which can permit almost linear scaling in *M* in several downstream methods
- **Nested gausslets:**
  - Extremely compact basis achieving high accuracy
  - So far construction is not automatic for all geometries
- **Ongoing** (joint with both collaborators at right):
  - Applying the COMX theorem to get a fully automatic construction
  - Algorithmic developments of AFQMC and VMC to exploit diagonal bases
- Thank you for your attention!

Sandeep Sharma
UC Boulder



Steve White
UC Irvine