Multiresolution Methods for Solving Electronic Structure

George Fann, Robert Harrison
Oak Ridge National Laboratory and
University of Tennessee - Knoxville

Gregory Beylkin
University of Colorado

Joint JUSTIPEN-LACM Meeting, 2007

The funding

- This work is funded by the U.S. Department of Energy, the Office of Advanced Science Computing, SciDAC SAPP Program, the base math program, the division of Basic Energy Science, Office of Science, under contract DE-AC05-00OR22725 with Oak Ridge National Laboratory. This research was performed in part using
  - the Molecular Science Computing Facility in the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory under contract DE-AC06-76RLO 1830 with Battelle Memorial Institute,
  - resources of the National Energy Scientific Computing Center which is supported by the Office of Energy Research of the U.S. Department of Energy under contract DE-AC03-76SF0098,
  - and the National Leadership Computing Facility at Oak Ridge National Laboratory under contract DE-AC05-00OR22725 .

MADNESS

Multiresolution
Adaptive
Numerical
Scientific
Simulation

Essential techniques for fast computation

- Multiresolution

$$V_0 \subset V_1 \subset \cdots \subset V_n$$

$$V_n = V_0 + \left(V_1 - V_0\right) + \cdots + \left(V_n - V_{n-1}\right)$$

- Low-separation rank

$$f(x_1, \ldots, x_d) = \sum_{l=1}^{M} \sigma_l \prod_{i=1}^{d} f_i^{(l)}(x_i) + O(\varepsilon)$$

$$\left\| f_i^{(l)} \right\|_2 = 1 \qquad \sigma_l > 0$$

- Low-operator rank

$$A = \sum_{\mu=1}^{r} \underline{u}_\mu \sigma_\mu \underline{v}_\mu^T + O(\varepsilon)$$

$$\sigma_\mu > 0 \qquad \underline{v}_\mu^T . \underline{v}_\lambda = \underline{u}_\mu^T . \underline{u}_\lambda = \delta_{\mu\lambda}$$

How to “think” multiresolution

- Consider a ladder of function spaces

$$V_0 \subset V_1 \subset V_2 \subset \cdots \subset V_n$$

  – E.g., increasing quality atomic basis sets, or finer resolution grids, …

- Telescoping series

$$V_n = V_0 + (V_1 - V_0) + (V_2 - V_1) + \cdots + (V_n - V_{n-1})$$

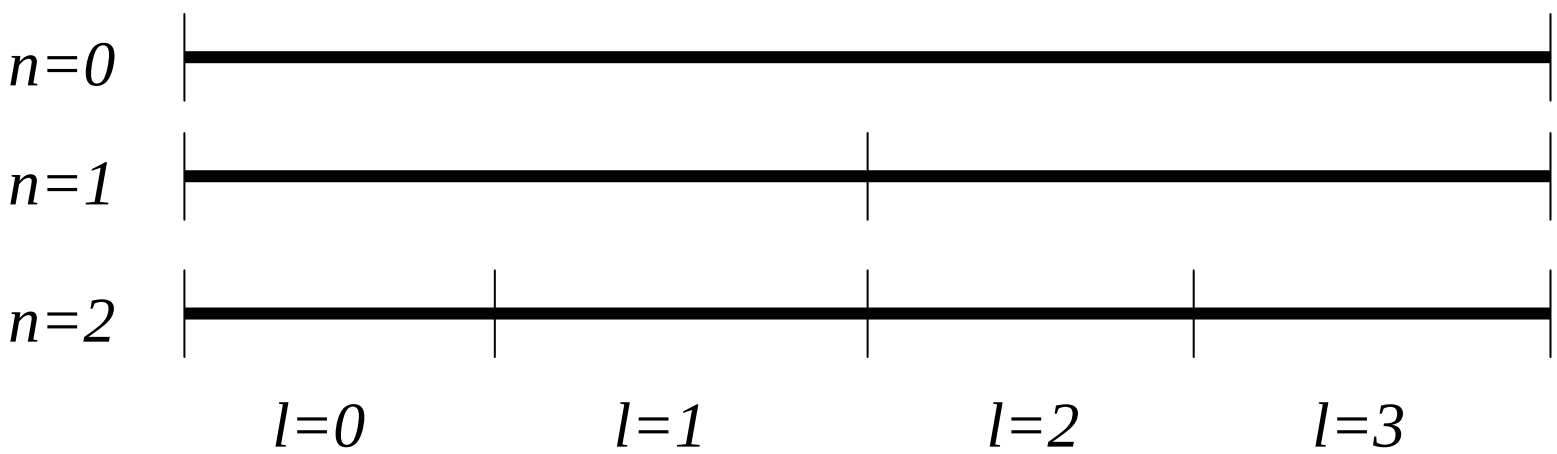
  – Instead of using the most accurate representation, use the difference between successive approximations

  – Representation on V_0 small/dense; differences sparse

  – Computationally efficient; possible insights

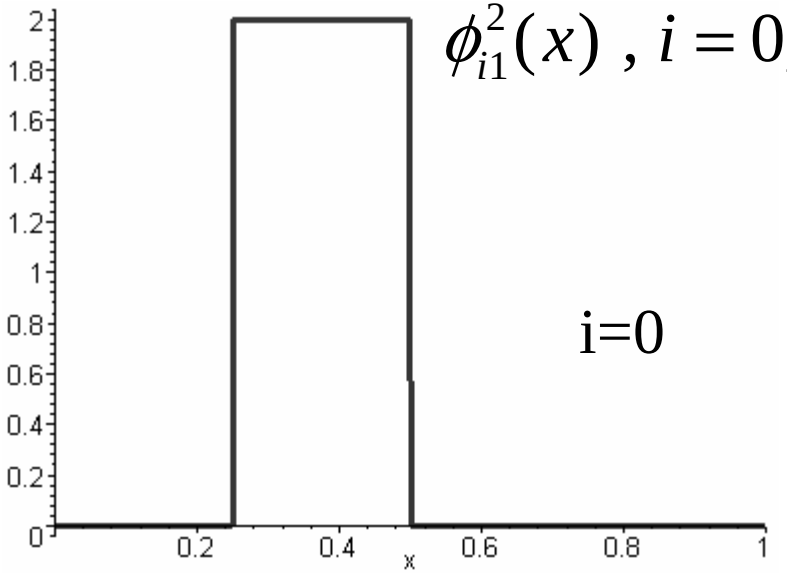
Scaling Function Basis

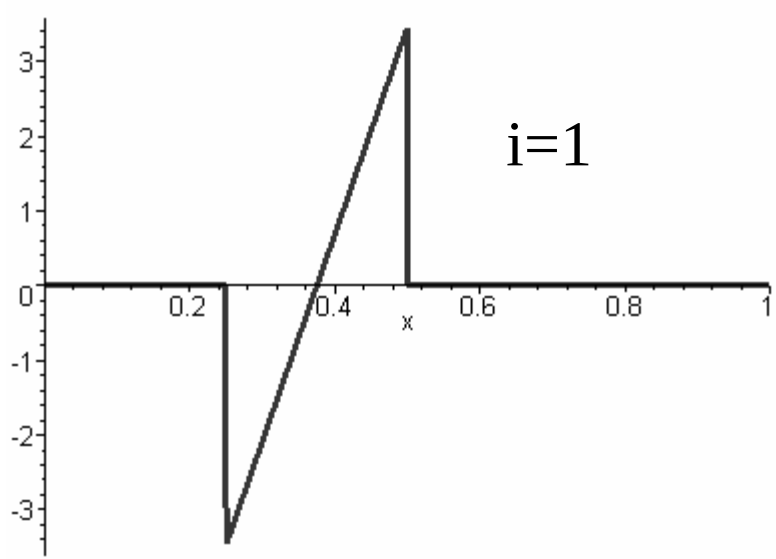
- Divide domain into 2^n pieces (level n)
  - Adaptive sub-division (local refinement)
  - l^{th} sub-interval $[l*2^{-n},(l+1)*2^{-n}]$ $l=0,\ldots,n-1$
- In each sub-interval define a polynomial basis
  - First k Legendre polynomials $\phi_i(x) = \sqrt{2i+1}P_i(2x-1)$
  - Orthonormal, disjoint support $\phi_{il}^n(x) = 2^{n/2}\phi_i(2^n x - l)$

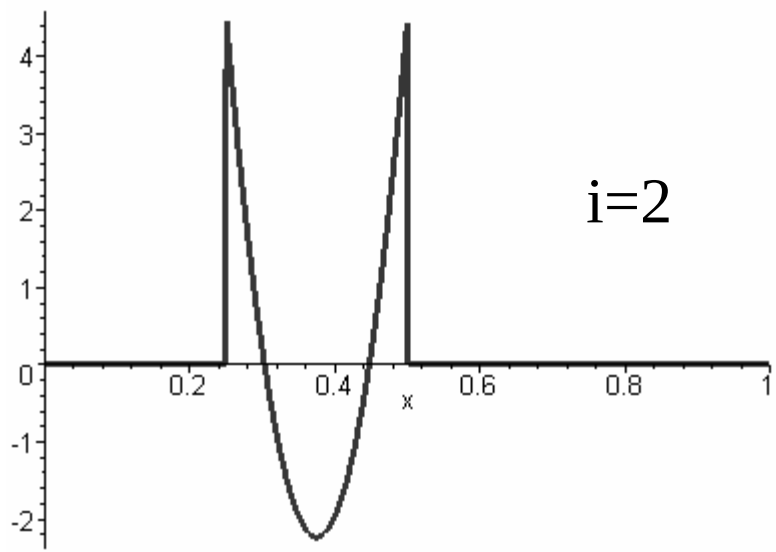


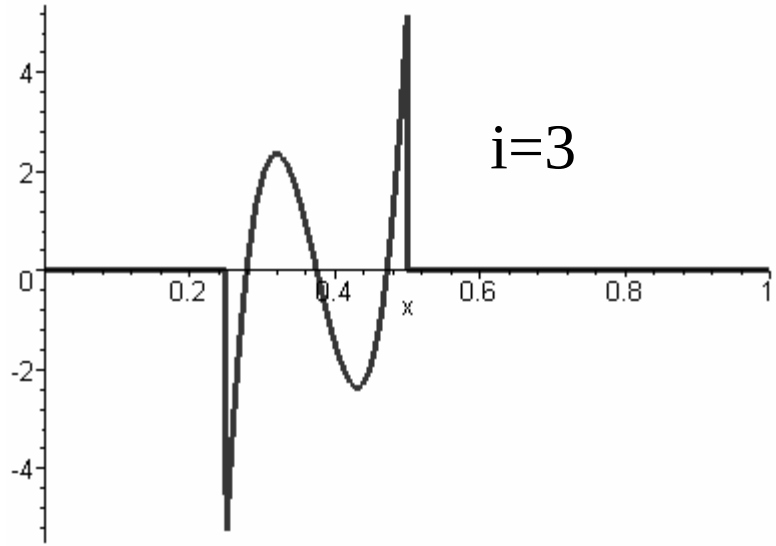
Scaling Function Basis - III

$\phi_{i1}^{2}(x)$, $i = 0,\ldots,3$









Multiwavelet Basis

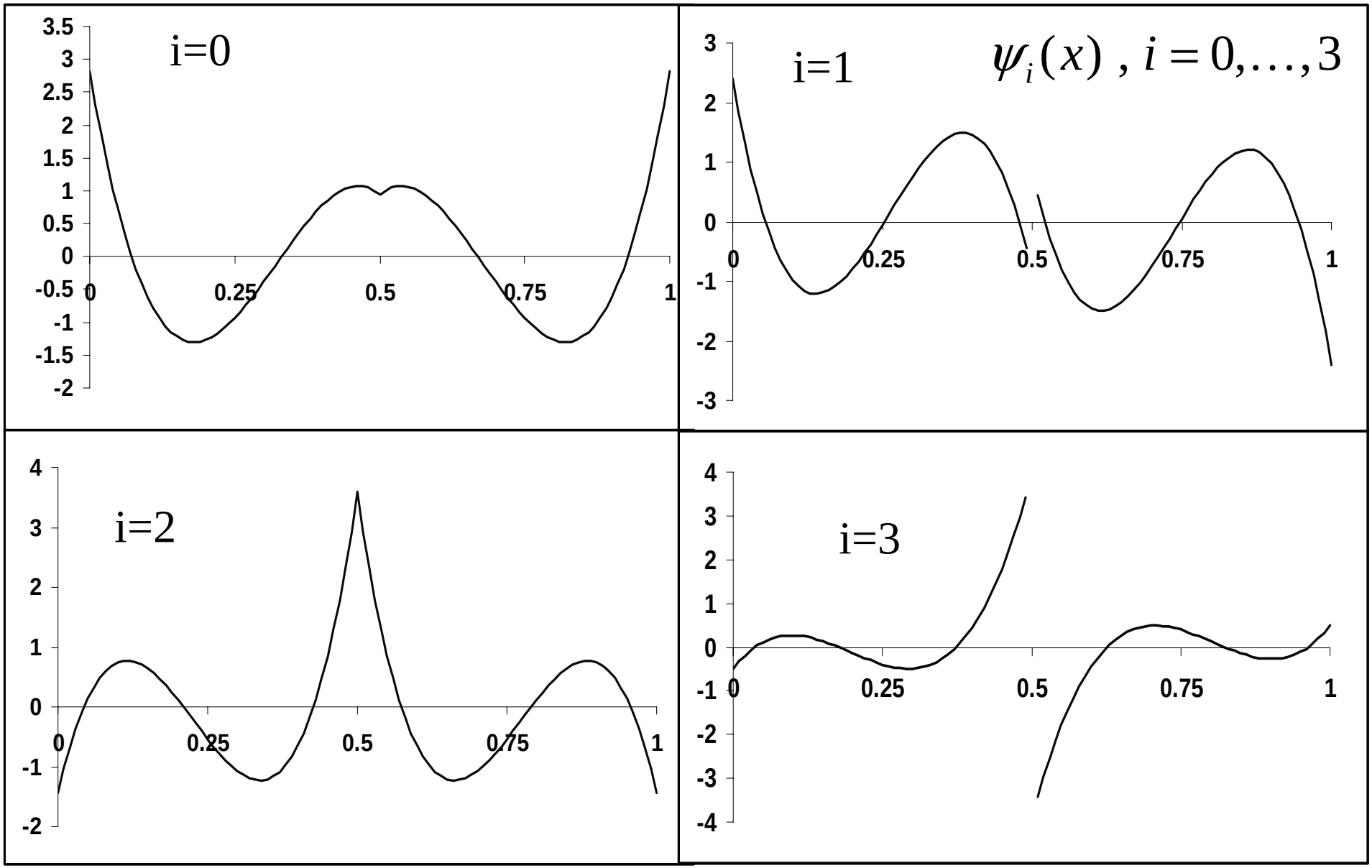
- Space of polynomials on level n is V_n
- Wavelets - an orthonormal basis to span $W_n = V_{n+1} - V_n$
- Currently use Alpert's basis
- Vanishing moments
  - Critically important property
  - Since W_n is orthogonal to V_n the first k moments of functions in W_n vanish, i.e.,

$$\int x^j \psi_i(x)dx = 0,\ j = 0,\ldots,k-1$$

- Sparse representations of many physically important kernels

Multiwavelet Basis - III

$k=4$



Adaptive Refinement

- To satisfy the global error condition

$$\left\| f - f^{n} \right\|_2 \le \varepsilon \left\| f \right\|_2$$

- Truncate according to

$$\left\| d_l^n \right\|_2 \le 2^{-n/2} \varepsilon \left\| f \right\|_2$$

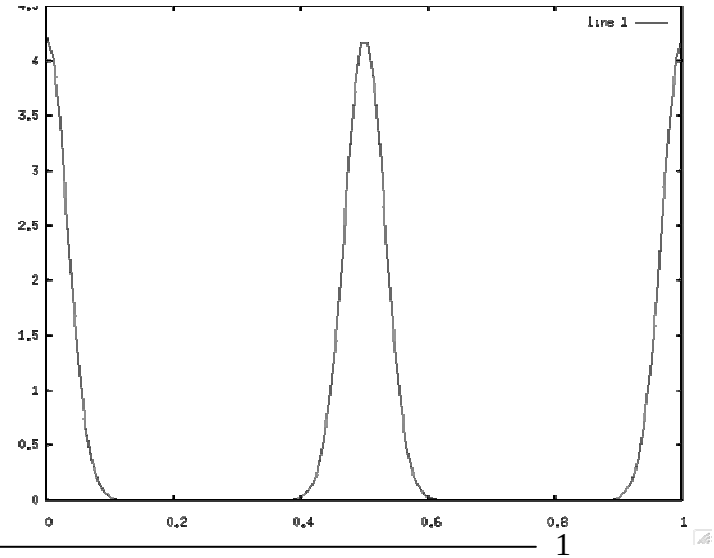
- This is rather conservative – usually use

$$\left\| d_l^n \right\|_2 \le \varepsilon$$

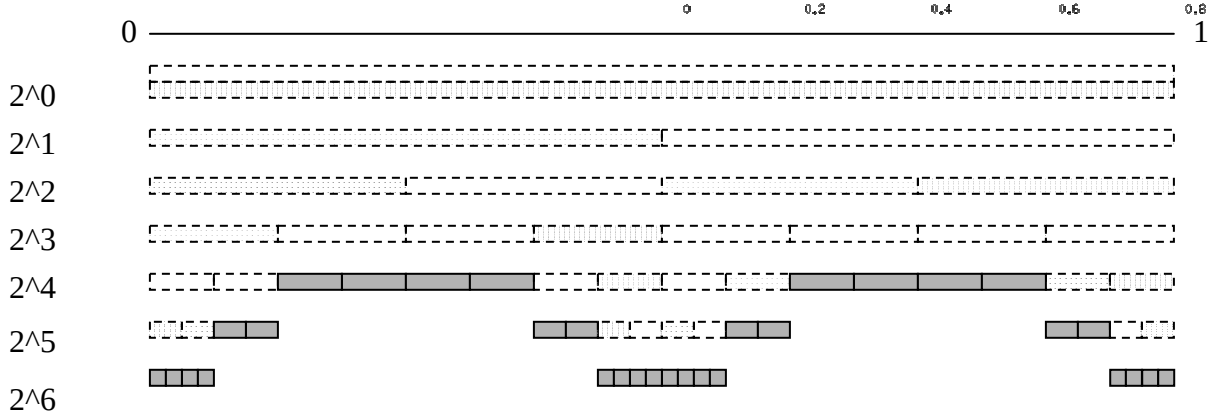
1-D example

$$f(x) = \left(\frac{2a}{\pi}\right)^{\frac{1}{4}} e^{-a(x-1)^2} + \left(\frac{2a}{\pi}\right)^{\frac{1}{4}} e^{-a(x-0.5)^2} + \left(\frac{2a}{\pi}\right)^{\frac{1}{4}} e^{-a(x)^2}$$

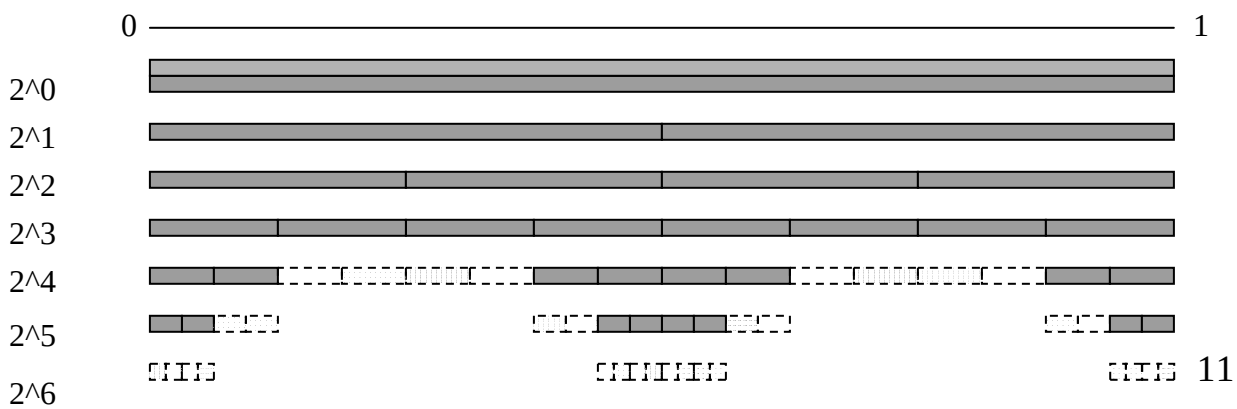
with k = 6 (#basis fn) and accuracy threshold = 1e-4



Refinement tree in the *scaling function basis*



Refinement tree in the *multi-wavelet basis*



Non-standard form:Telescoping Series

- P_n an orthogonal projection into V_n
- Q_n an orthogonal projection into $W_n = V_{n+1} - V_n$

$$P_n + Q_n = P_{n+1}$$

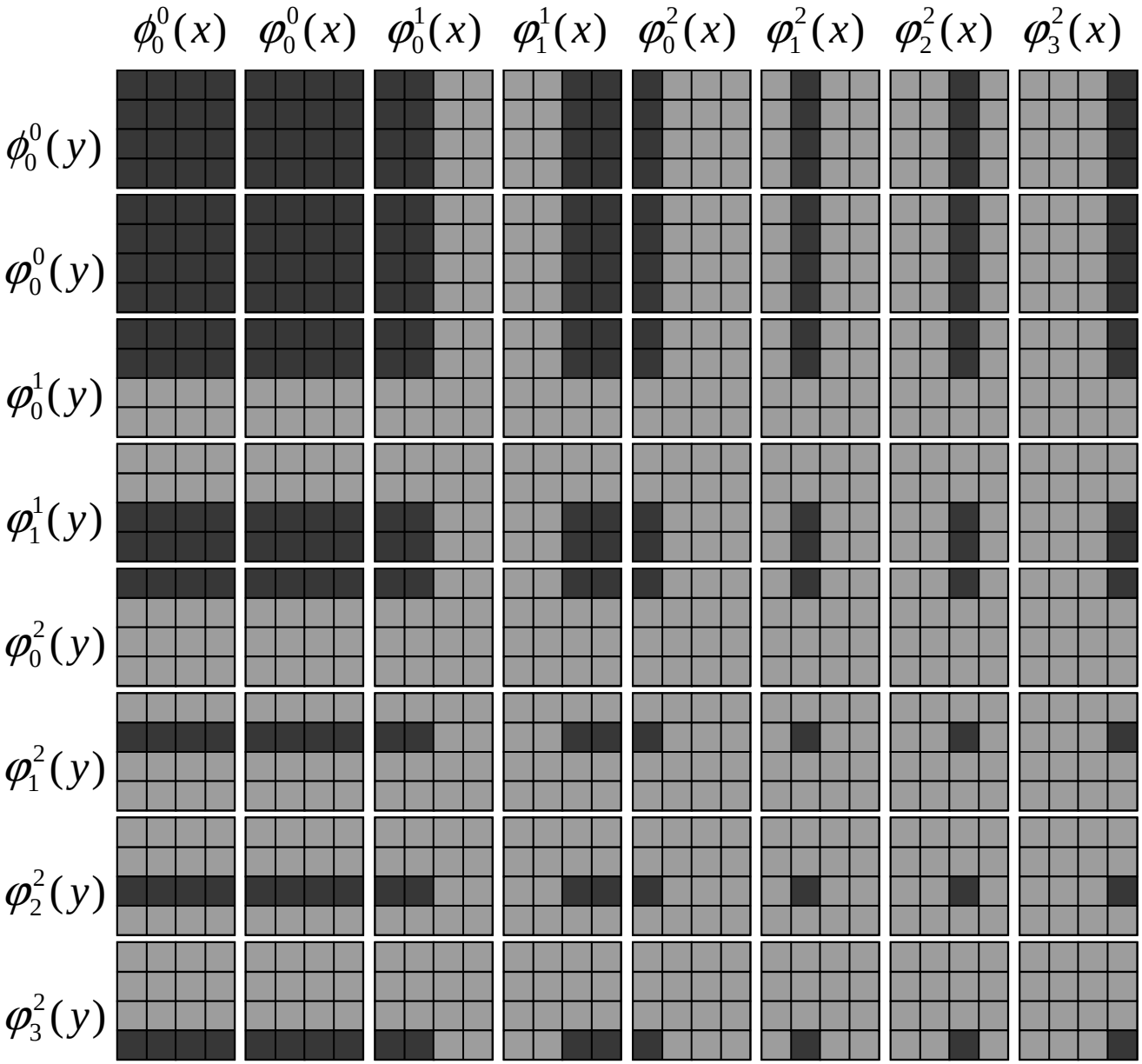
- Consider the projection of an operator T

$$\begin{aligned}
T_n &= P_n T P_n \\
&= \left(P_{n-1} + Q_{n-1}\right) T \left(P_{n-1} + Q_{n-1}\right) \\
&= P_{n-1} T P_{n-1} + Q_{n-1} T Q_{n-1} + Q_{n-1} T P_{n-1} + P_{n-1} T Q_{n-1} \\
&= T_{n-1} + A_{n-1} + B_{n-1} + C_{n-1} \\
&= T_0 + \sum_{\eta=0}^{n-1} A_\eta + B_\eta + C_\eta
\end{aligned}$$

Non-standard form of operators

- Standard form
  - Matrix elements between different length scales
  - Not very efficient on modern computers with deep memory hierarchies
  - Potentially *O(Nlog N)* non-zero terms
- Non-standard form
  - No matrix elements between length scales
  - *O(N)* terms
  - Act on (modified) non-standard form of functions
  - Easy to compute
  - Toeplitz matrix for translationally invariant operators
  - Enables true local refinement of functions

Standard-Form basis in 2-D



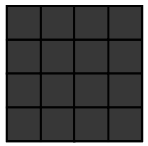
Tensor product basis on level 3 (level 3 scaling functions == level 2 wavelets)

Red indicates the support of the function

Non-standard-Form basis in 2-D

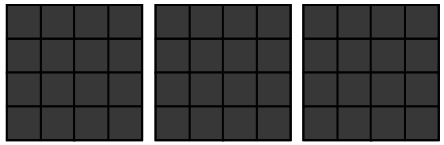
Red indicates function support

$n=0$



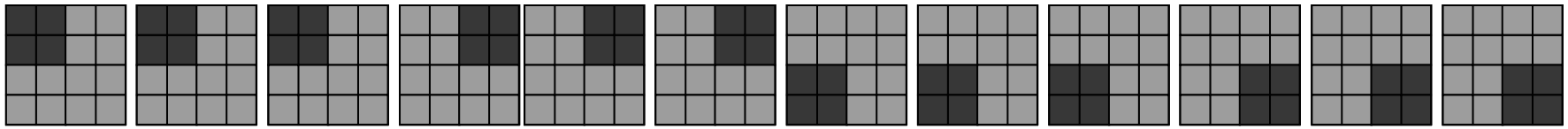
$$\left\{\phi_0^0(x)\phi_0^0(y)\right\}$$

$n=0$

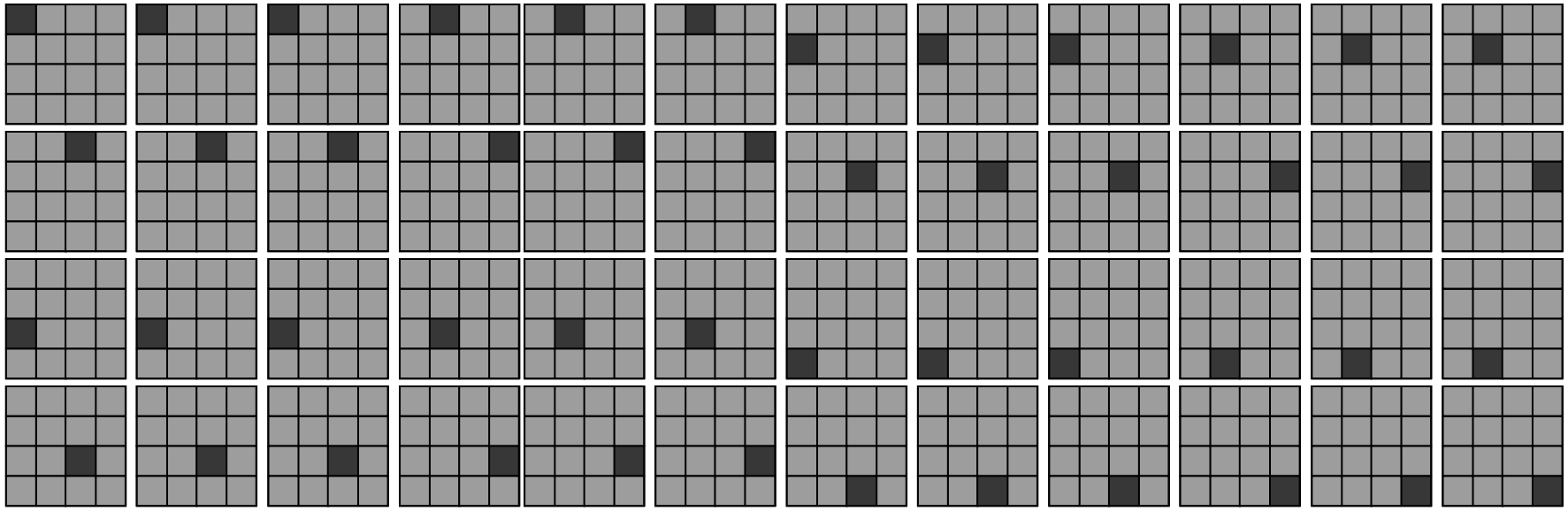


$$\left\{\phi_l^n(x)\varphi_m^n(y),\varphi_l^n(x)\phi_m^n(y),\varphi_l^n(x)\varphi_l^n(y)\right\} \quad l,m=0,..,2^n-1$$

$n=1$



$n=2$



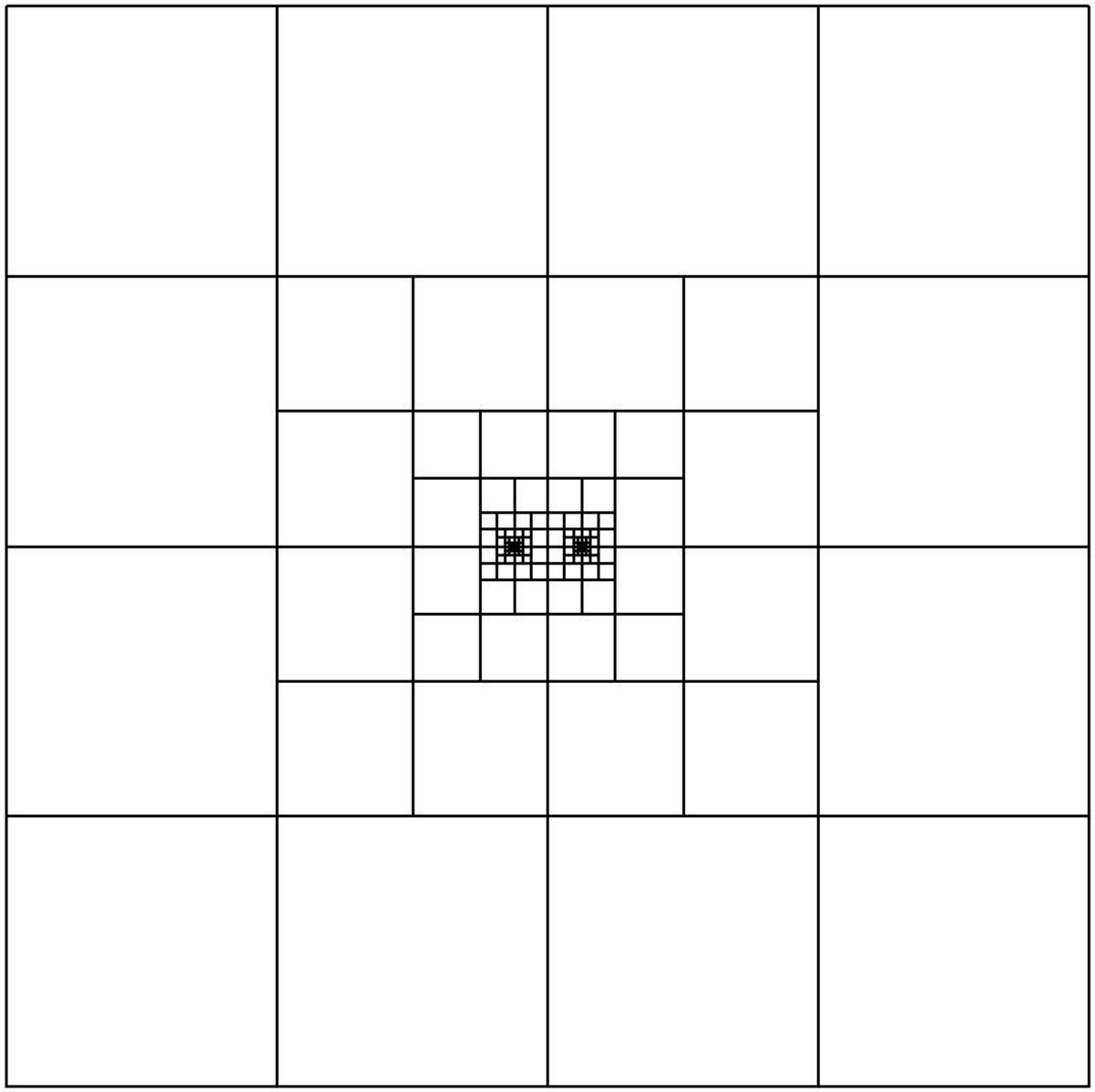
Computation of the gradient

$$\rho(\underline{r}) = \sum_{\underline{i}=0}^{k-1} s_{\underline{i}\,0}^{0} \phi_{\underline{i}\,0}^{0}(\underline{r}) + \sum_{n'=0}^{n-1} \sum_{\underline{l}}^{2^{n'}-1} \sum_{\underline{i}=0}^{k-1} d_{\underline{i}\,\underline{l}}^{n'} \psi_{\underline{i}\,\underline{l}}^{n'}(\underline{r})$$

$$\frac{\partial V}{\partial q}(\underline{r}) = \sum_{\underline{i}=0}^{k-1} r_{\underline{i}\,0}^{0} \phi_{\underline{i}\,0}^{0}(\underline{r}) + \sum_{n'=0}^{n-1} \sum_{\underline{l}}^{2^{n'}-1} \sum_{\underline{i}=0}^{k-1} c_{\underline{i}\,\underline{l}}^{n'} \psi_{\underline{i}\,\underline{l}}^{n'}(\underline{r})$$

$$\int \frac{\partial V(\underline{r})}{\partial q} \rho(\underline{r}) d\underline{r} = \sum_{\underline{i}=0}^{k-1} r_{\underline{i}\,0}^{0} s_{\underline{i}\,0}^{0} + \sum_{n'=0}^{n-1} \sum_{\underline{l}}^{2^{n'}-1} \sum_{\underline{i}=0}^{k-1} c_{\underline{i}\,\underline{l}}^{n'} d_{\underline{i}\,\underline{l}}^{n'}$$

Since the basis is orthonormal, the integral is a simple inner product



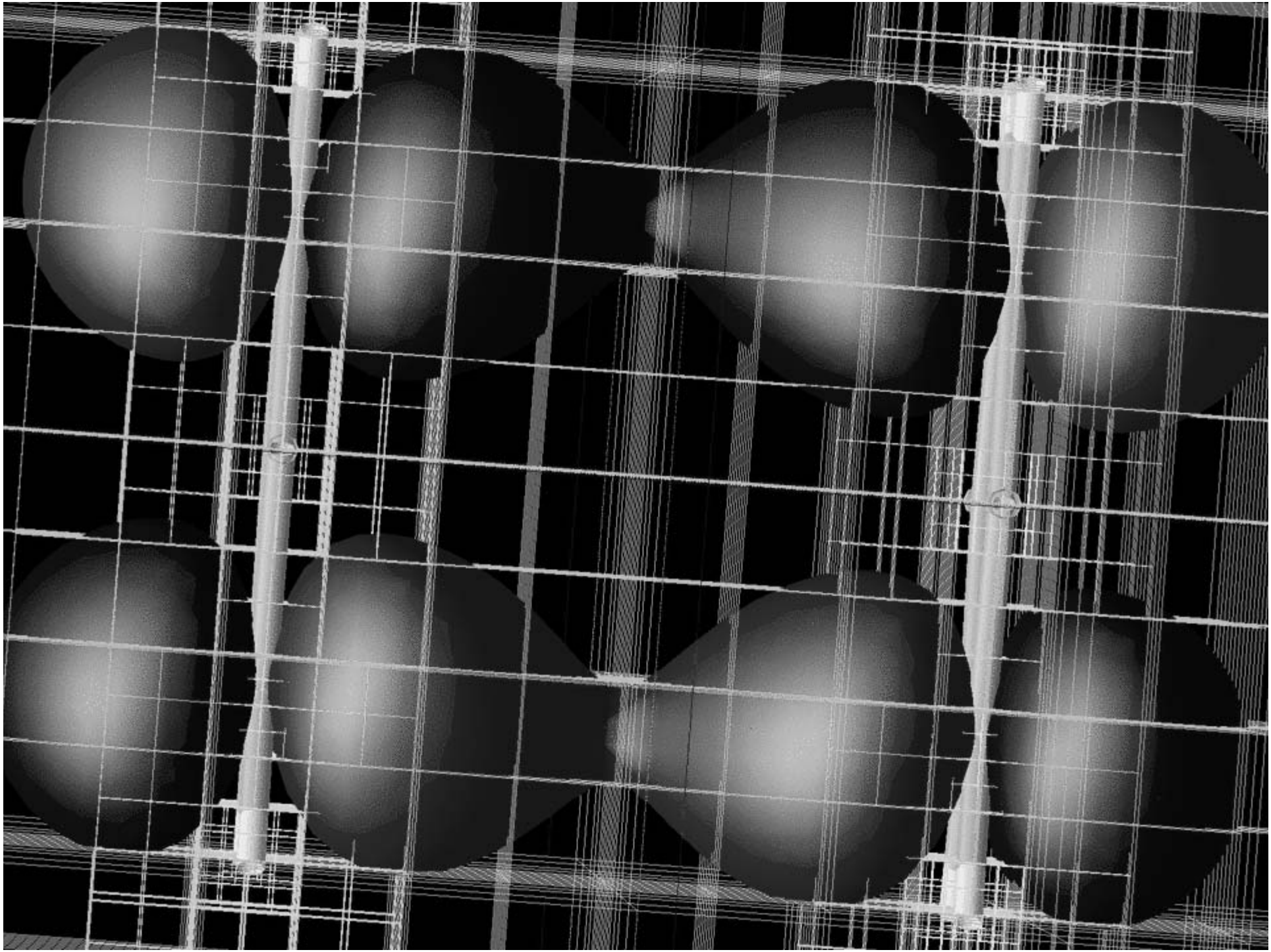
- Slice thru grid used to represent the nuclear potential for H_2 using k=7 to a precision of 10^{-5}.

- Automatically adapts – it does not know a priori where the nuclei are.

- Nuclei at dyadic points on level 5 – refinement stops at level 8

- If were at non-dyadic points refinement continues (to level ??) but the precision is still guaranteed.

- In future will unevenly subdivide boxes to force nuclei to dyadic points.



Integral operators in 3D

$$T * f = \int ds K(r-s) f(s)$$

- Non-standard-form matrix elements easy to evaluate from compressed form of kernel *K(x)*
- Application in 1-d is fairly efficient
  - $O(N_{box}k^2)$ operations

$$r_{ii'}^{nl}$$

- In 3-d seems to need $O(N_{box}k^6)$ operations
  - Prohibitively expensive

$$r_{ii',jj',kk'}^{n\underline{l}}$$

- Separated form
  - Beylkin, Cramer, Mohlenkamp, Monzon
  - $O(N_{box}k^4)$ or even $O(N_{box}k^3)$ in 3D

$$r_{ii',jj',kk'}^{nl} = \sum_i \omega_i X_{ii'}^{nl_x} Y_{ii'}^{nl_y} Z_{ii'}^{nl_z}$$

Integral Formulation

- E.g., used by Kalos, 1962

$$\left(-\tfrac{1}{2}\nabla^2 + V\right)\Psi = E\Psi$$

$$\Psi = -2\left(-\nabla^2 - 2E\right)^{-1} V\Psi$$

$$= -2G * \left(V\Psi\right)$$

$$\left(G * f\right)(r) = \int ds \frac{e^{-k|r-s|}}{4\pi|r-s|} f(s) \text{ in 3D ; } k^2 = -2E$$

Converges as a fixed-point iteration without preconditioning for the lowest eigen function.

Separated form for integral operators

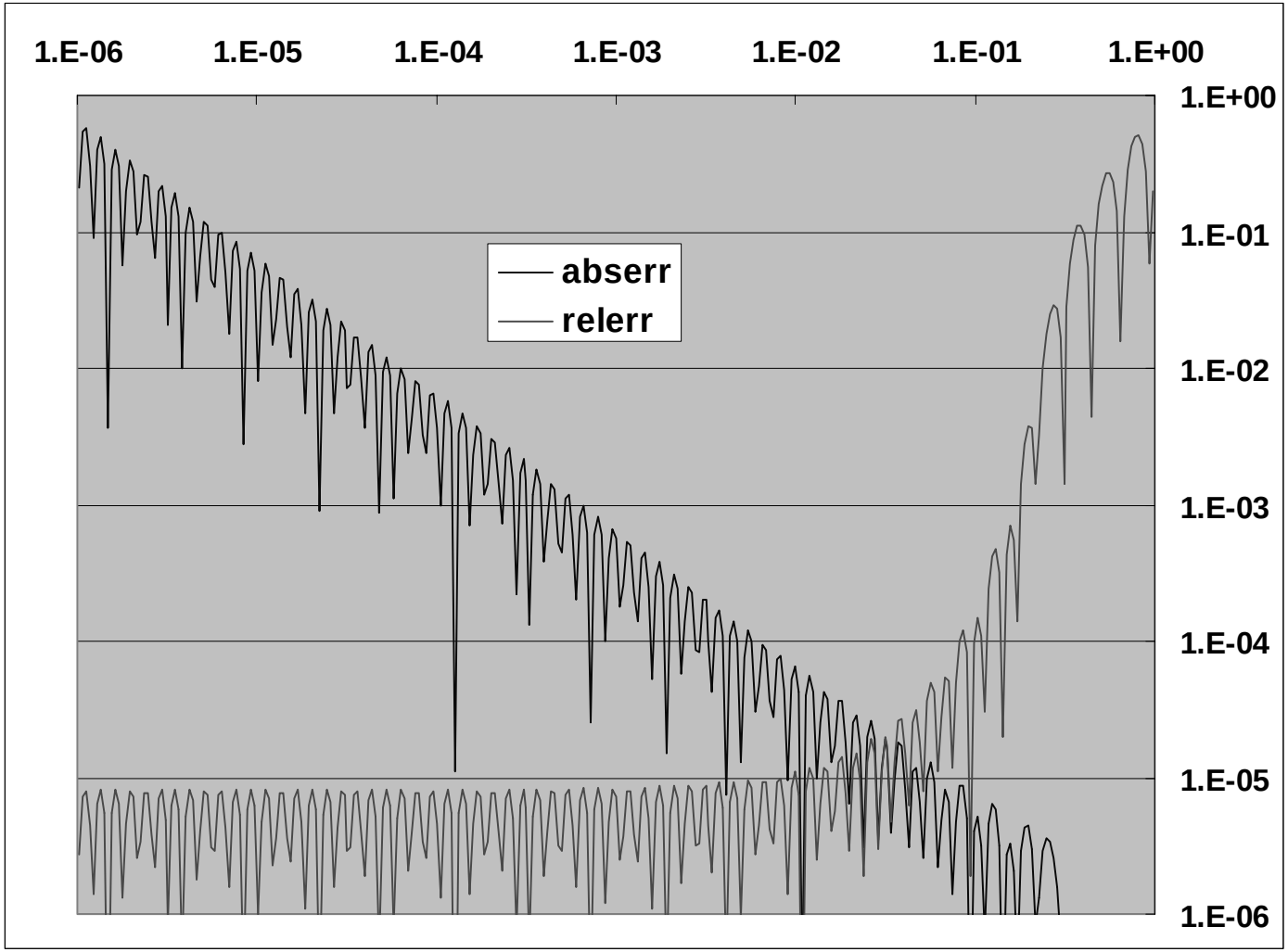
$$T * f = \int ds K(r-s) f(s)$$

- Approach in current prototype code
  - Represent the kernel over a finite range as a sum of Gaussians

$$K(r) = \sum_i \omega_i e^{-t_i r^2} + O(\varepsilon)$$

  - Only need compute 1D transition matrices (X,Y,Z)
  - SVD the 1-D operators (low rank away from singularity)
  - Apply most efficient choice of low/full rank 1-D operator
  - Working on better algorithms

Separated form for bound-state Helmholtz kernel



exp(-30*r)/(4πr) expanded as 36 Gaussians over [1e-6,1] accurate to about 1e-5 relative error (until value is tiny due to exp. decay).

Logarithmic dependence on precision and range

Separated form for integral operators

$$T * f = \int ds K(r-s) f(s)$$

- Approach in current prototype code
  - Represent the kernel over a finite range as a sum of Gaussians

$$K(r) = \sum_i \omega_i e^{-t_i r^2} + O(\varepsilon)$$

  - Only need compute 1D transition matrices (X,Y,Z)
  - SVD the 1-D operators (low rank away from singularity)
  - Apply most efficient choice of low/full rank 1-D operator
  - Even better algorithms not yet implemented

Integral operators in 3D

$$T * f = \int ds K(r-s) f(s)$$

- Non-standard-form matrix elements easy to evaluate from compressed form of kernel *K(x)*
- Application in 1-d is fairly efficient
  - $O(N_{box}k^2)$ operations

$$r_{ii'}^{nl}$$

- In 3-d seems to need $O(N_{box}k^6)$ operations
  - Prohibitively expensive

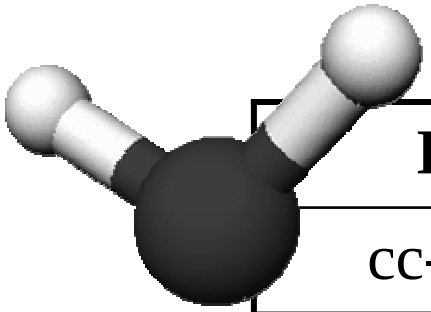
$$r_{ii',jj',kk'}^{n\underline{l}}$$

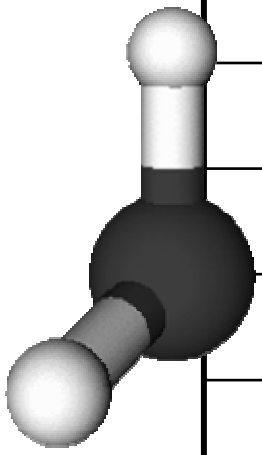
- Separated form
  - Beylkin, Cramer, Mohlenkamp, Monzon
  - $O(N_{box}k^4)$ or even $O(N_{box}k^3)$ in 3D

$$r_{ii',jj',kk'}^{nl} = \sum_i \omega_i X_{ii'}^{nl_x} Y_{ii'}^{nl_y} Z_{ii'}^{nl_z}$$

Water dimer LDA

aug-cc-pVTZ geometry, kcal/mol.

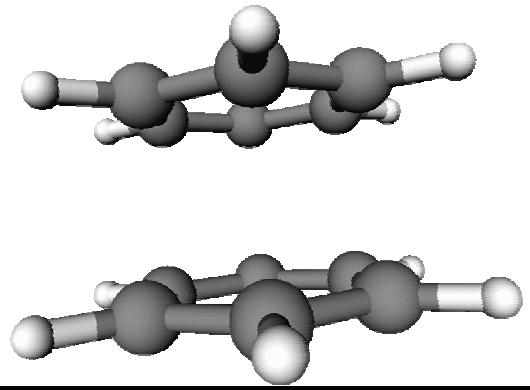




Basis	Uncorrected	BSSE	Corrected
cc-pVDZ	-11.733	-3.958	-7.775
cc-pVTZ	-9.464	-1.654	-7.810
cc-pVQZ	-8.708	-0.821	-7.888
aug-cc-pVDZ	-8.187	-0.382	-7.805
aug-cc-pVTZ	-7.992	-0.086	-7.906
aug-cc-pVQZ	-7.995	-0.054	-7.941
$\varepsilon=10^{-3}$	-6.483		
$\varepsilon=10^{-5}$	-7.932		
$\varepsilon=10^{-7}$	-7.943		

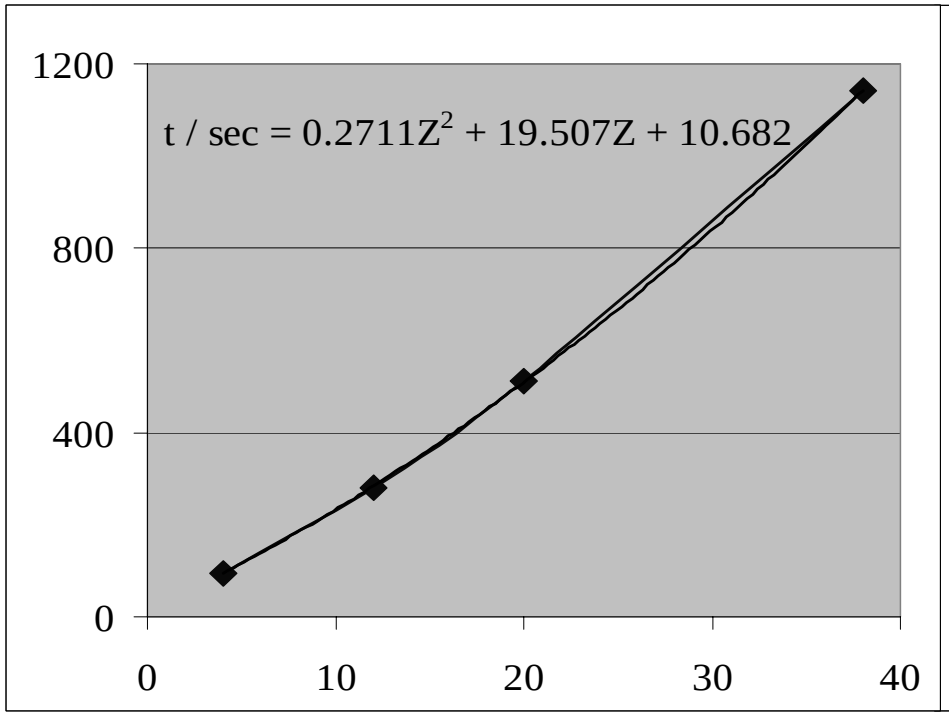
Benzene dimer LDA

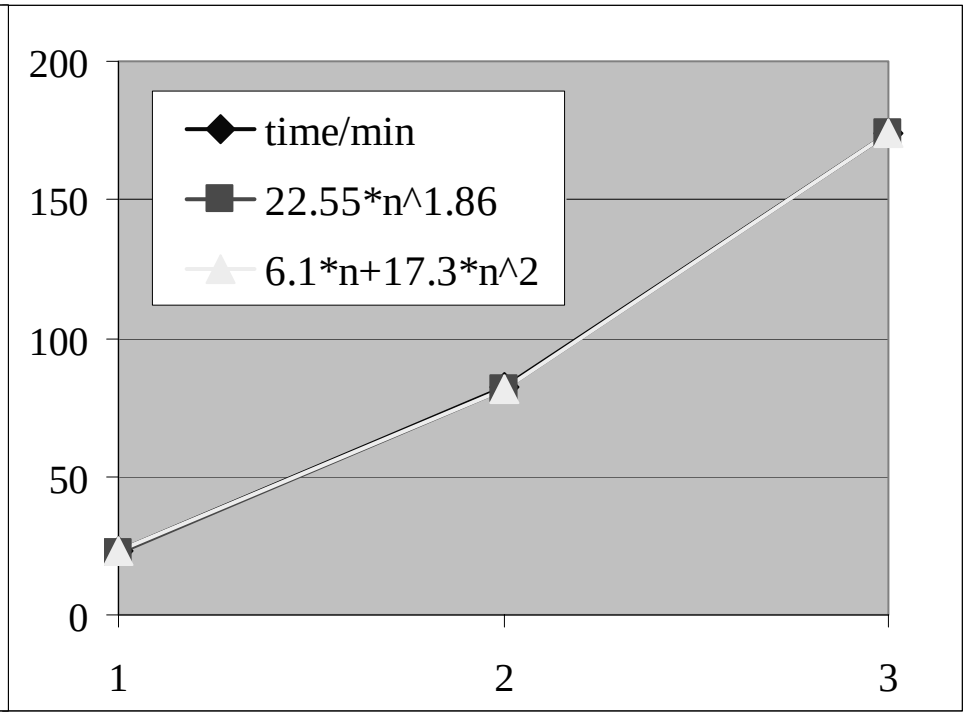
aug-cc-pVDZ geometry, kcal/mol.



Basis	Uncorrected	BSSE	Corrected
cc-pVDZ	-1.506	-1.035	-0.471
cc-pVTZ	-1.271	-0.387	-0.884
cc-pVQZ	-1.074	-0.193	-0.881
aug-cc-pVDZ	-1.722	-0.698	-1.024
aug-cc-pVTZ	-1.159	-0.193	-0.966
$\varepsilon=10^{-5}$	-0.872		
$\varepsilon=10^{-7}$	-0.956		
$\varepsilon=10^{-9}$	-0.956		

LDA scaling with Z and system size (energy $\varepsilon=10^{-5}$)

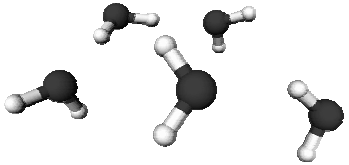


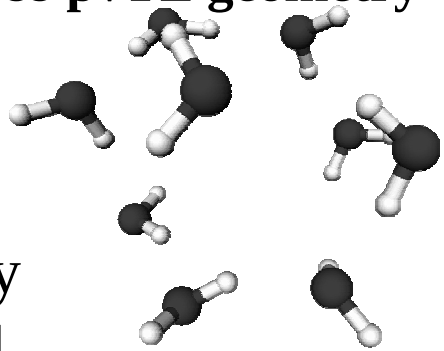


Alkali earth atoms Z=4,12,20,38

$(C_6H_6)_n$ MP2 aug-cc-pvTZ geometry

$(H_2O)_n$ n=5,9 … t = O($n^{1.1}$)



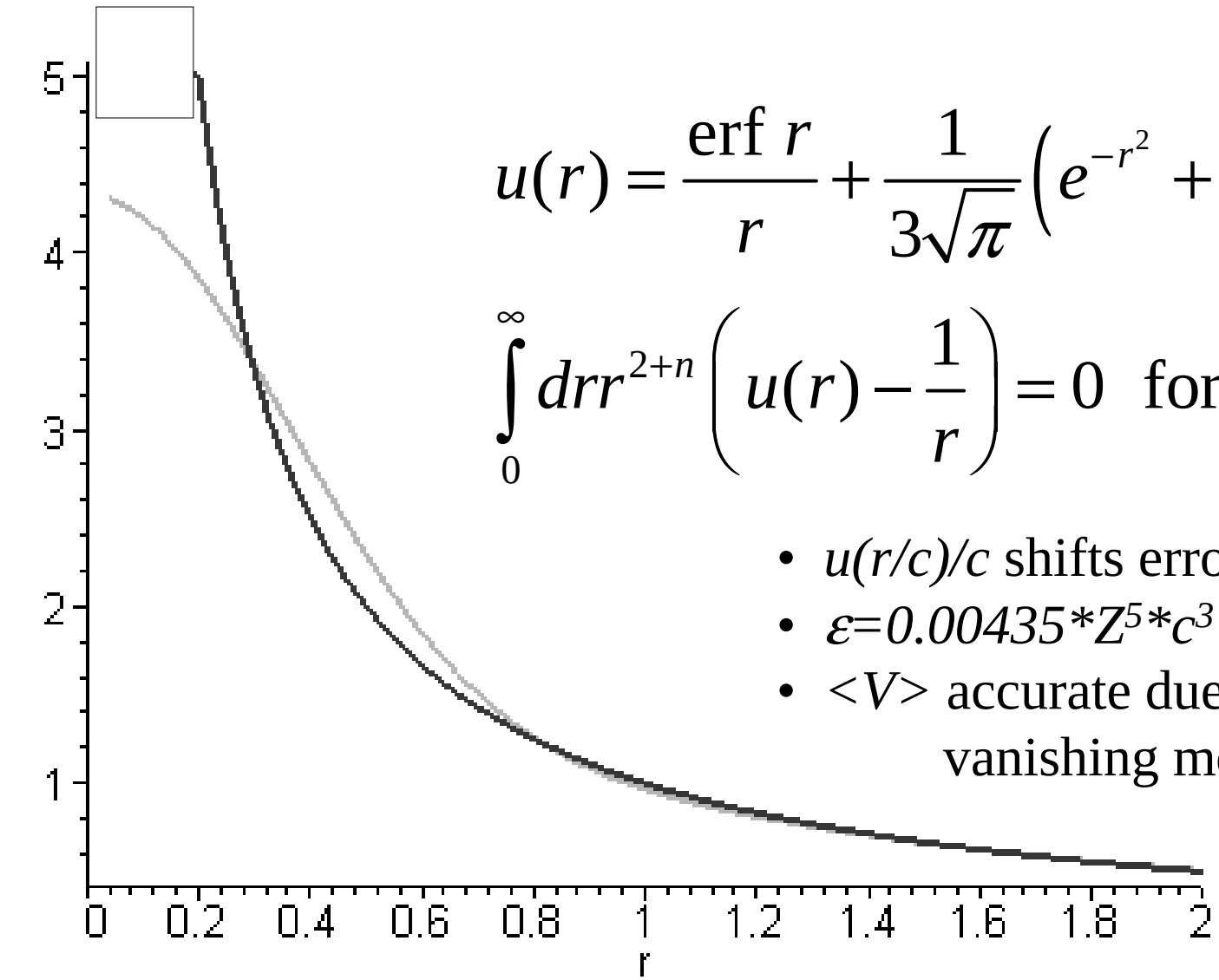


Stacked benzene – MOs are delocalized by symmetry

Water cluster – MOs are asymptotically localized

(long tail is smooth so is inexpensively treated)

Smoothed Nuclear Potential

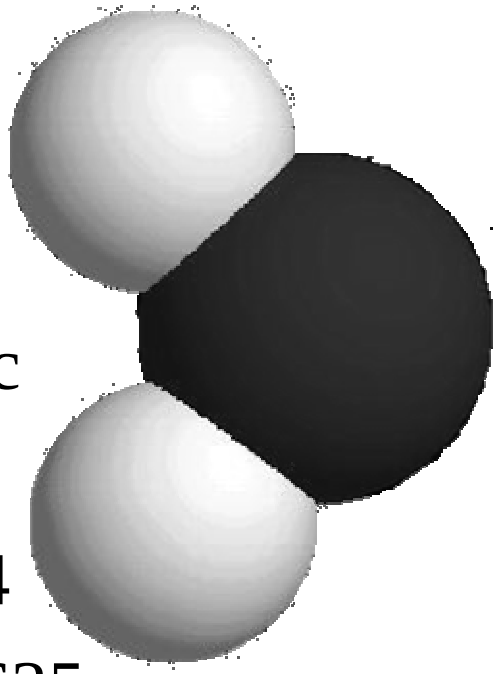


$$u(r) = \frac{\operatorname{erf} r}{r} + \frac{1}{3\sqrt{\pi}}\left(e^{-r^2} + 16e^{-4r^2}\right)$$

$$\int_0^\infty dr r^{2+n}\left(u(r) - \frac{1}{r}\right) = 0 \quad \text{for} \quad n = 0,1,2$$

- $u(r/c)/c$ shifts error to $r<c$
- $\varepsilon = 0.00435 * Z^5 * c^3$
- $\langle V \rangle$ accurate due to vanishing moments

Translational Invariance



	Dyadic	Non-dyadic
10^{-3}	-75.9139	-75.9139
10^{-5}	-75.913564	-75.913564
10^{-7}	-75.91355634	-75.91355635

- Uncontracted aug-cc-pVQZ –75.913002
- Solving with e=1e-3, 1e-5, 1e-7 (k=7,9,11)
- Demonstrates translation invariance and that forcing to dyadic points is only an optimization and does not change the obtained precision.
- Average orbital sizes 1.6Mb, 8Mb, 56Mb

High-precision Hartree-Fock geometry for water

- Pahl and Handy Mol. Phys. 100 (2002) 3199
  - Plane waves + polynomials for the core
  - Finite box (L=18) requires extrapolation
  - Estimated error 3μH, 1e-5 Angstrom
- k=11, conv.tol=1e-8, ε=1e-9, L=40
  - Max. gradient = 3e-8, RMS step=5e-8
  - Difference to Pahl 10μH, 4e-6 Angstrom, 0.0012

Basis	**OH**	**HOH**	**Energy**
k=11	0.939594	106.3375	-76.06818006
Pahl	0.939598	106.3387	-76.068170
cc-pVQZ	0.93980	106.329	-76.066676

He atom Tests, 6-D

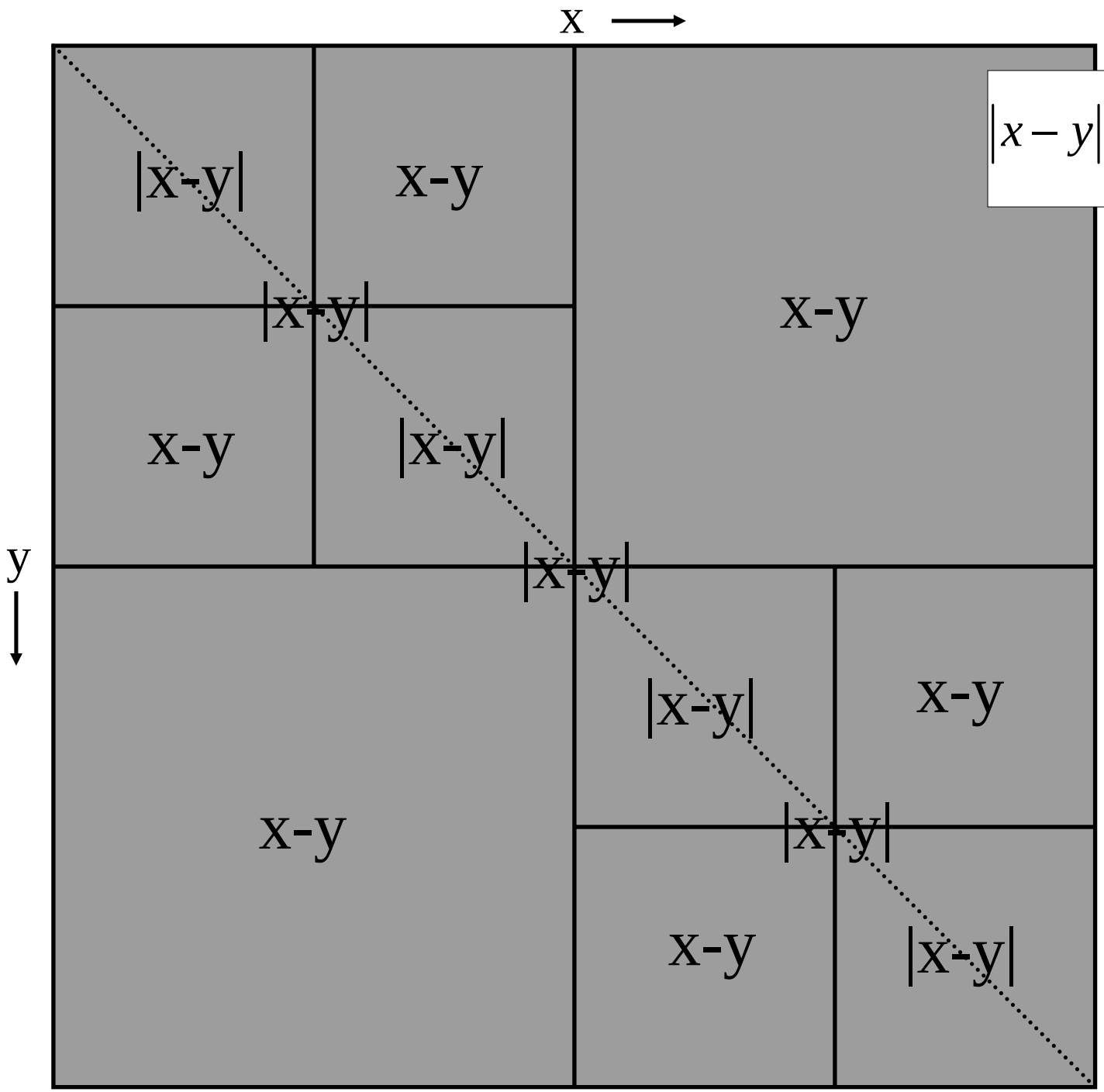
- Compute expectation value of the total Hamiltonian for He atom with a trial wavefunction in 6-D multiresolution
  - 5-th order multiwavelet bases
  - truncation threshold: 1e-3

$$\Psi\left(r_1, r_2\right) = \left(1 + c\, r_{12}\right) e^{-\xi\left(r_1 + r_2\right)}$$

		$\langle\Psi\|H\|\Psi\rangle$	$\langle\Psi\|V_{nuc}\|\Psi\rangle$	$\langle\Psi\|1/r_{12}\|\Psi\rangle$
c=0.50, ξ=2.0	6D MRA	–2.873 37	– 7.189 83	1.022 19
	Analytic	–2.876 58	– 7.189 87	1.022 15
c=0.00, ξ=2.0	6D MRA	–2.748 40	– 7.999 86	1.250 04
	Analytic	–2.750 00	– 8.000 00	1.250 00

Kinetic energy is currently a problem for the prototype – will be resolved with better refinement criterion

- in 3D there is no problem; in 6D should get identical result at C=0



$$|x-y| = \sum_{\mu=1}^{r} f_\mu(x) g_\mu(y)$$

r = separation rank

In 3D, ideally must be one box removed from the diagonal

Diagonal box has full rank

Boxes touching diagonal (face, edge, or corner) have increasingly low rank

Away from diagonal $r = O(-\log \varepsilon)$

Timings and sizes

- Wavefunction (k=4, ε=1e-3, deepest n=5)
  - (c=0.5) 0.9GB
  - (c=0.0) 42MB
- R12*Wavefunction
  - (c=0.5) 1.8GB memory + 0.5h CPU
- Kinetic*Wavefuction
  - (c=0.5) 0.8h CPU
  - (c=0.0) 50s CPU
- Vnuc*Wavefunction
  - (c=0.5) 0.5h CPU
  - (c=0.0) 133s CPU

Anticipate 10+x speedup possible with better implementation and algorithms.

But, more general systems will require deeper refinement and higher-order wavelets.

Also looking at fast algorithms for Gaussian geminal form

High-level composition using functions and operators

- Conventional quant. chem. uses explicitly indexed sparse arrays of matrix elements
  - Complex, tedious and error prone
- Python classes for Function and Operator
  - in 1,2,3,6 and general dimensions
  - wide range of operations

$$H\varphi = -\frac{1}{2}\nabla^2\varphi + V\varphi$$

$$J(r) = G * \rho = \int \frac{\rho(s)}{|r-s|}\,ds$$

```
Hpsi = -0.5*Delsq*psi+ V*psi
J = Coulomb.apply(rho)
```

- All with guaranteed speed and precision

Summary

- Multiresolution provides a general framework for computational chemistry
  - Accurate and efficient with a very small code
  - Multiwavelets provide high-order convergence and accommodate singularities
- Readily accessible by students and researchers
  - Familiar orthonormal basis (Legendre polynomials)
  - Compression and reconstruction (c.f., FFT)
  - Fast integral operators (c.f., FMM)
- Separated form for operators and functions
  - Critical for efficient computation in higher dimension
- Expect speed competitive to Gaussians in near future
  - Optimal separated forms for kernels, multi-scale non-linear solver, better implementation

Current and Future Work

- Operators, functions (i.e. for Dirac) over complex numbers
- High dimensional 6+ for MADNESS
- Better basis for more compact approximation of functions and operators
- Time-dependent solvers
- Hypersingular?

Example:
Future Separated Representation

$$H = \sum_{j=1}^{N} (-\Delta_j + \| x_j \|^2)$$

$$H^{-1} = \sum_{m=1}^{M} \hat{w}_m \prod_{j=1}^{N} e^{-\tau_m \|x_j\|^2} e^{-\sigma_m \|x_j - y_j\|^2} e^{-\tau_m \|y_j\|^2}$$