Argonne NATIONAL LABORATORY

... for a brighter future

U.S. Department of Energy

UChicago ▶ Argonne LLC

Office of Science U.S. DEPARTMENT OF ENERGY

A U.S. Department of Energy laboratory managed by UChicago Argonne, LLC

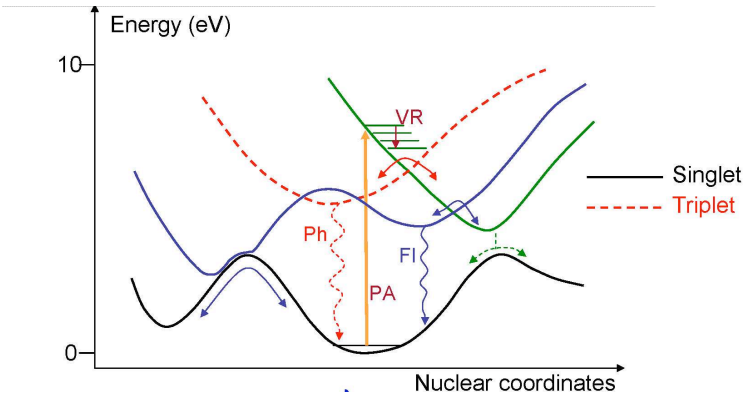
Some Algorithms in Quantum Chemistry

Ron Shepard
Chemistry Division
Argonne National Laboratory

CScADS Workshop, Snowbird, Utah, July 30, 2007

Quantum Chemistry

- Multiple Potential Energy Surfaces (Born-Oppenheimer Approximation)
- Electron Correlation (various approximations and approaches)
- Molecular Structure and Properties (ground and excited states)
- Critical Points of high-dimensional surfaces (transition states, minima, intersections, etc.)



VR vibrational relaxation
Ph phosphorescence
FI fluorescence
PA photoabsorption

conical intersection
avoided crossings
intersystem crossings

COLUMBUS

Goal: Parallel general-purpose ab initio techniques applicable to ground and excited states for arbitrary points on the PES

Applications: spectroscopy, dynamics, kinetics

History: 1980 first sequential version, 1990 first parallel version

Distribution: Open source and binary (~500K lines of Fortran+C+Perl)

Contributors: I. Shavitt (GUGA), H. Lischka (photodynamics, solvent effects, Analytic Gradients, NAC), R. Shepard (MCSCF, MRCI, Analytic Gradients, NAC, SPAM), R. Pitzer (integrals, spin-orbit CI), P. G. Szalay (Analytic Gradients, MR-AQCC, GDIIS). Th. Mueller (parallel MRCI/MR-AQCC, interoprability, spin-orbit CI) D. R. Yarkony (NAC, Conical Intersections, crossing seams), Granucci (Surface Hopping Dynamics), and many others.

http://www.univie.ac.at/columbus/

Current Support: DOE, USAF, NSF, FWF(Austria), NIC(Juelich)

Single-Point Electronic Energy and Wave Function Evaluation:

At a fixed molecular conformation and for the electronic states of interest, evaluate the expectation value

$$E(\mathbf{R}) = \frac{\left\langle \Psi_{el}(\mathbf{r};\mathbf{R}) \middle| H^{op}(\mathbf{r};\mathbf{R}) \middle| \Psi_{el}(\mathbf{r};\mathbf{R}) \right\rangle}{\left\langle \Psi_{el}(\mathbf{r};\mathbf{R}) \middle| \Psi_{el}(\mathbf{r};\mathbf{R}) \right\rangle}$$

$$H^{op} = \sum_{j}^{Nel} \frac{-\hbar^2}{2m_e} \nabla_j^2 + \sum_{j}^{Nel} \sum_{a}^{Natom} \frac{Z_e Z_a}{\left| \mathbf{r}_j - \mathbf{R}_a \right|} + \sum_{j<k}^{Nel} \frac{Z_e^2}{\left| \mathbf{r}_j - \mathbf{r}_k \right|}$$

$$\langle \ \rangle \equiv \int \int \ldots \int \ d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_{Nel}$$

The electronic wave function is expanded in a basis:

$$\left|\Psi_{el}(\mathbf{r};\mathbf{R})\right\rangle = \sum_{m}^{Ncsf} c_m \left|m(\mathbf{r};\mathbf{R})\right\rangle$$

← configuration state function

$$\left|m(\mathbf{r};\mathbf{R})\right\rangle = \sum_{\sigma_1\sigma_2\ldots\sigma_{Nel}} c_{\sigma_1\sigma_2\ldots\sigma_{Nel}} \left|\varphi_{m_1\sigma_1}(\mathbf{r}_1;\mathbf{R})\varphi_{m_2\sigma_2}(\mathbf{r}_2;\mathbf{R})\ldots\varphi_{m_{Nel}\sigma_{Nel}}(\mathbf{r}_{Nel};\mathbf{R})\right|$$

Slater determinant

$$\varphi_k(\mathbf{r};\mathbf{R}) = \sum_{\mu}^{n} C_{\mu k}(\mathbf{R})\chi_\mu(\mathbf{r};\mathbf{R}_{a(\mu)})$$

atom-centered basis functions

orbital coefficients (optimized at each conformation)

orthonormal molecular orbitals

The CSF expansion results in a real symmetric eigenvalue equation:

$$\mathbf{H}(\mathbf{R})\mathbf{c}(\mathbf{R}) = \mathbf{E}(\mathbf{R})\mathbf{c}(\mathbf{R})$$

$$H(\mathbf{R})_{mn} = \langle m|H^{op}|n\rangle = \sum_{p,q}^{Norb} h(\mathbf{R})_{pq}\langle m|E_{pq}|n\rangle + \tfrac{1}{2}\sum_{p,q,r,s}^{Norb} g(\mathbf{R})_{pqrs}\langle m|e_{pqrs}|n\rangle$$

2-particle coupling coefficient

2-electron hamiltonian integral

1-particle coupling coefficient

1-electron hamiltonian integral

Tensor Transformations of Integral Arrays:

Integral arrays as needed

$$h_{pq} = \sum_{\mu\nu} C_{\mu p} C_{\nu q} h_{\mu\nu}^{[ao]} = \left(\mathbf{C}^T \mathbf{h}^{[ao]} \mathbf{C}\right)_{pq}$$

$$h_{\mu\nu}^{[ao]} = \int \chi_\mu(\mathbf{r}) \left(-\tfrac{1}{2}\nabla^2 - \sum_a^{Natom} \frac{Z_a}{|\mathbf{r} - \mathbf{R}_a|} \right) \chi_\nu(\mathbf{r}) d\mathbf{r}$$

Integral arrays as computed

$$g_{pqrs} = \sum_{\mu\nu\lambda\kappa} C_{\mu p} C_{\nu q} C_{\lambda r} C_{\kappa s} g_{\mu\nu\lambda\kappa}^{[ao]}$$

$$g_{\mu\nu\lambda\kappa}^{[ao]} = \int\int \chi_\mu(\mathbf{r}_1) \chi_\nu(\mathbf{r}_1) \left(\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \right) \chi_\lambda(\mathbf{r}_2) \chi_\kappa(\mathbf{r}_2) d\mathbf{r}_1 d\mathbf{r}_2$$

Configuration Interaction: Real Symmetric Eigenvalue Problem

- Use the iterative Davidson Method for the lowest (or lowest few) eigenpairs
- *Direct CI*: **H** is not explicitly constructed, **w=Hx** are constructed in "operator" form
- Matrix dimensions are 10^4 to 10^{10}
- All floating point calculations are 64-bit (need 8 to 11 significant figures in the energy)

Davidson Method (simplified)

Generate an initial vector $\mathbf{x}_1$
MAINLOOP: DO n=1, NITER
  Compute and save $\mathbf{w}_n = \mathbf{H}\,\mathbf{x}_n$
  Compute the n^{th} row and column of $\mathbf{G} = \mathbf{X}^T\mathbf{H}\mathbf{X} = \mathbf{W}^T\mathbf{X}$
  Compute the subspace Ritz pair: $(\mathbf{G} - \rho\mathbf{1})\,\mathbf{c} = 0$
  Compute the residual vector $\mathbf{r} = \mathbf{W}\,\mathbf{c} - \rho\,\mathbf{X}\,\mathbf{c}$
  Check for convergence using $|\mathbf{r}|$, $\mathbf{c}$, ρ, etc.
  IF (converged) THEN
    EXIT MAINLOOP
  ELSE
    Generate a new expansion vector $\mathbf{x}_{n+1}$ from $\mathbf{r}$, ρ, $\mathbf{v}=\mathbf{X}\mathbf{c}$, etc.
  ENDIF
ENDDO MAINLOOP

Matrix-Vector Products

$$\mathbf{w} = \mathbf{H}\,\mathbf{x}$$

$$w_m = \sum_{n}^{Ncsf} H_{mn} x_n$$

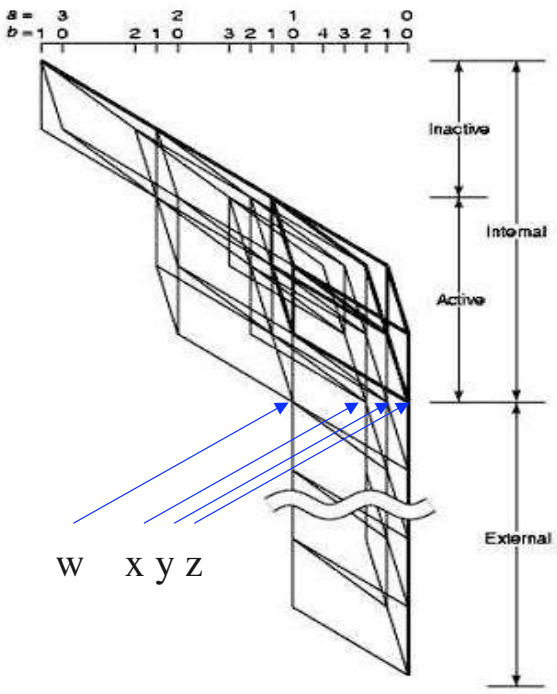
$$= \sum_{n}^{Ncsf}\sum_{p,q}^{Norb} h_{pq}\langle m|E_{pq}|n\rangle x_n + \tfrac{1}{2}\sum_{n}^{Ncsf}\sum_{p,q,r,s}^{Norb} g_{pqrs}\langle m|e_{pqrs}|n\rangle x_n$$

- The challenge is to bring together the different factors in order to compute **w** efficiently
- h_{pq} and g_{pqrs} are computed and stored as arrays (with index symmetry)
- $\langle m|E_{pq}|n\rangle$ and $\langle m|e_{pqrs}|n\rangle$ are sparse and are recomputed as needed

Coupling Coefficient Evaluation

- Graphical Unitary Group Approach (GUGA)
- Define a directed graph with nodes and arcs: *Shavitt Graph*
- Nodes correspond to spin-coupled states consisting of a subset of the total number of orbitals and electrons
- Arcs correspond to the (up to) four allowed spin couplings when an orbital is added to the graph
- Coupling coefficients are evaluated as products of "segment values" of interacting segments within the Shavitt Graph.

CSF/Walk order



← graph head

←graph tail

Integral Types

- 0: g_{pqrs}
- 1: g_{pqra}
- 2: g_{pqab}, $g_{pa,qb}$
- 3: g_{pabc}
- 4: g_{abcd}

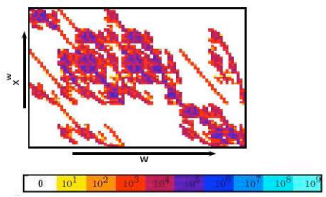
H Matrix Structure

	z	y	x	w
z	0	1	2	2
y	1	0,2	1,3	1,3
x	2	1,3	0,2,4	2
w	2	1,3	2	0,2,4

- Some integral subsets are replicated on all nodes
- Some are distributed among nodes (e.g. GA Library)

Computational Kernels and Scaling

- Depending on the block (zz, yz, wx, etc., the low-level arithmetic operations are DDOT, DAXPY, DGEMV, or DGEMM with the matrix dimension being N_{orb} or N_{ext}.



Relative cost for wx 2-external block in terms of groups of valid internal walks

- Total effort for MR-SDCI is $\sim 4^{N_{active}} N_{orb}^{6}$

This problem can use an arbitrarily large amount of computer time as the size of the molecule increases

Davidson Diagonalization Details

New Expansion Vector:

$$\mathbf{x}^{[new]} = -(\mathbf{H}\mathbf{v} - \rho\mathbf{v}) = -\mathbf{r}$$

Lanczos

$$\left(\mathbf{H}^0 - \rho\right)\mathbf{x}^{[new]} = -\mathbf{r}$$

$\mathbf{H}^0=\mathbf{D}$ $\rightarrow$Diagonal Preconditioned Residual

$$\left(\mathbf{H}^0 - \rho\right)\mathbf{x}^{[new]} = -\mathbf{r} + \varepsilon\mathbf{v}$$

RQII, IIGD, GJD, etc.

$$\mathbf{x}^{[new]} = \mathbf{Q}\bar{\mathbf{v}} \quad \text{with} \quad \bar{\mathbf{H}}\bar{\mathbf{v}} = \bar{E}\bar{\mathbf{v}}$$

$$\mathbf{P} = \mathbf{X}\mathbf{X}^T;\ \mathbf{Q} = \mathbf{1} - \mathbf{P};\ \bar{\mathbf{H}} = \mathbf{H} - \mathbf{Q}(\mathbf{H} - \mathbf{H}^0)\mathbf{Q}$$

SPAM (converging preconditoner)
J. Comp. Phys. **172**, 472 (2001)

…Davidson Diagonalization Details

Choice of v and ρ:

- Choose the lowest unconverged eigenpair (sequential convergence)
- Cycle among unconverged eigenpairs (maintains subspace faithfulness)
- Choose the approximate eigenpair that maximizes $\langle \mathbf{v}|\mathbf{v}^{ref}\rangle^2$ (vector following)
- Choose the approximate eigenpair that minimizes $|\rho - E^{ref}|$ (root homing)

Single or multiple $\mathbf{x}^{new}$ (blocked algorithms):

- This affects data reuse efficiency, task granularity, and the convergence rate

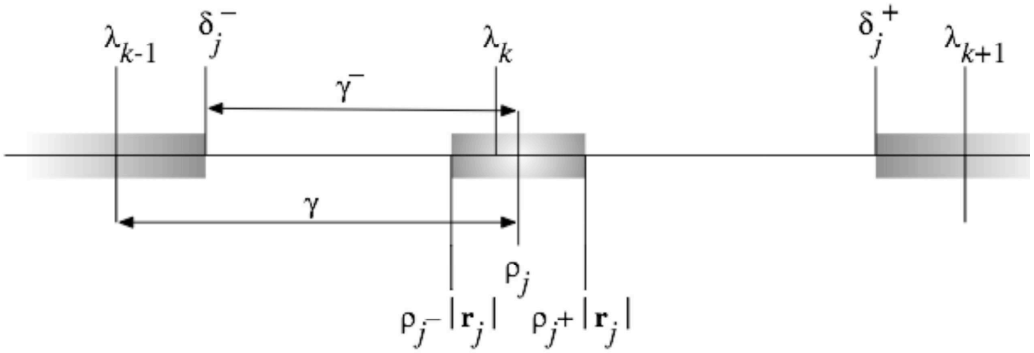
Subspace contraction (restart strategy):

- Sometimes determined entirely by storage capacity (no choice)
- Optimal max and min subspace dimensions improves efficiency (transform away useless components) by reducing memory, I/O, and communications requirements.
- Nonoptimal dimensions hurt efficiency – sometimes difficult to predict optimal strategy

…Davidson Diagonalization Details

Convergence :

- Usually do not need energies (eigenvalues) to full machine precision (8 to 11 significant figures is typical).



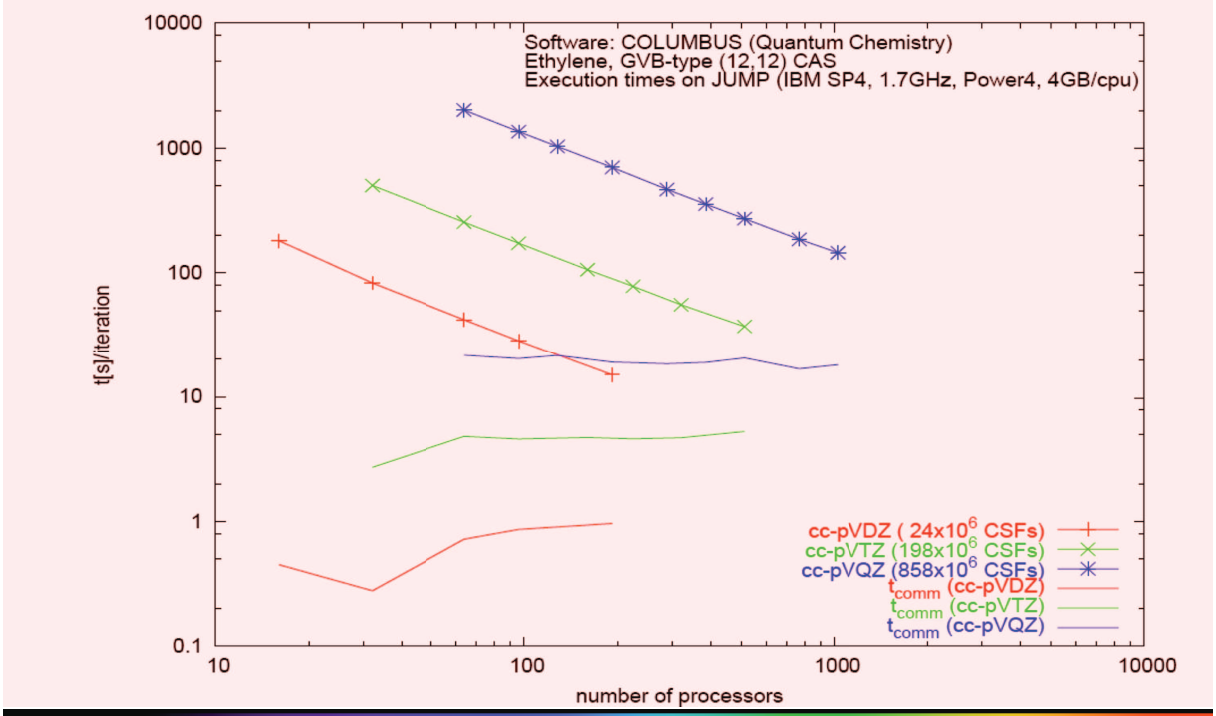
$$\frac{\mathbf{r}_j^\dagger \mathbf{M}^{-1}\mathbf{r}_j}{\gamma_j^-} \geq \frac{\mathbf{r}_j^\dagger \mathbf{M}^{-1}\mathbf{r}_j}{\gamma_j} \geq \left|\lambda_j - \rho_j\right|$$

Apply residual bound, spread bound, and gap bound recursively

Comput. Phys. Comm. **167**, 90 (2005)

Comput. Phys. Comm. **170**, 109 (2005)

COLUMBUS Scalability



General Convergence Acceleration: Direct Inversion in the Iterative Subspace (DIIS)

Want to solve $\mathbf{F}(\mathbf{x}_*) = \mathbf{0}$ with $\mathbf{F}(\mathbf{x}) \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^q$.

If $\mathbf{u}^T\mathbf{c}=1$ with $u_k=1$ for $k=1:n$, then

$$\mathbf{F}(\mathbf{x}) + \mathcal{O}\left(\|\mathbf{x}-\mathbf{x}_0\|^2\right) = \sum_{k=1}^{n} \mathbf{F}(\mathbf{x}_k)c_k = \sum_{k=1}^{n} \mathbf{e}_k c_k = \mathbf{E}\mathbf{c}$$

Least squares interpolation is given by

$$\min_{\mathbf{c}} \|\mathbf{E}\mathbf{c}\| \text{ with } \mathbf{u}^T\mathbf{c} = 1$$

$\mathbf{x}^{\text{DIIS}} = \mathbf{X}\mathbf{c}$

...DIIS

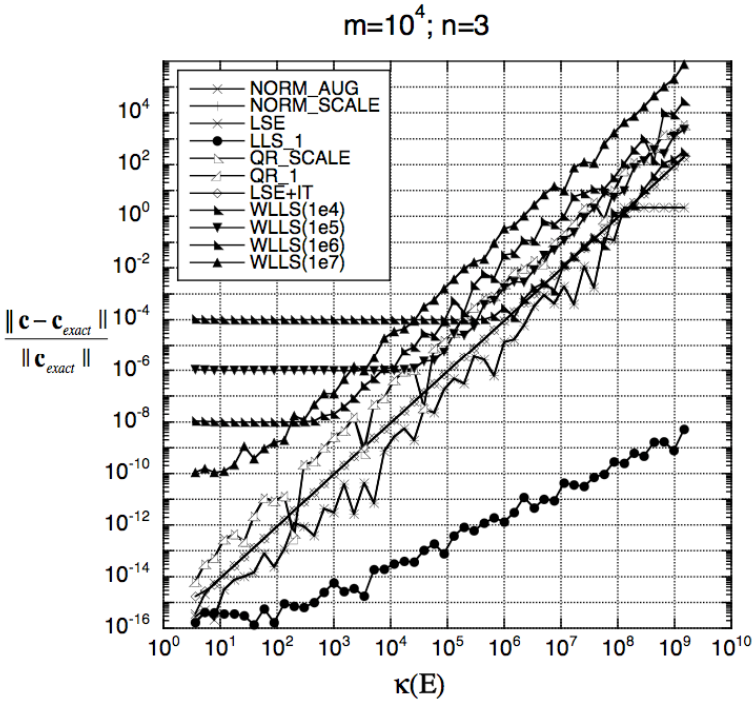
$$\begin{pmatrix} \mathbf{B} & -\mathbf{u} \\ -\mathbf{u}^T & 0 \end{pmatrix}\begin{pmatrix} \mathbf{c} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ -1 \end{pmatrix}$$ Normal equation

$$\min_{\mathbf{c}} \|\mathbf{E}\mathbf{c} - \mathbf{b}\| \text{ with } \mathbf{D}\mathbf{c} = \mathbf{d}$$ LSE

$$\min_{\tilde{\mathbf{c}}} \left\|\tilde{\mathbf{E}}\tilde{\mathbf{c}} + \mathbf{e}_n\right\|$$ Substitution and elimination

$$\min_{\mathbf{c}} \left\|\begin{pmatrix} \mathbf{E} \\ w\mathbf{u}^T \end{pmatrix}\mathbf{c} - \begin{pmatrix} \mathbf{0} \\ w \end{pmatrix}\right\|$$ Weighted least squares

…DIIS



Summary

Configuration Interaction

- Seek efficient ways to distribute **h** and **g** arrays on parallel computers
- Seek efficient and flexible iterative subspace diagonalization methods (need 8 to 11 significant figures in the eigenvalues)
- Consider contractions with $\mathbf{h}^{[AO]}$ and $\mathbf{g}^{[AO]}$ – more arithmetic operations per task, but no storage or communications requirements
- Tasks are not uniform in size – this presents load balancing challenges

Tensor Transformation

- Need efficient and general tensor transformation and tensor contraction libraries
- Must include index symmetry to reduce storage and arithmetic operation counts

Least Squares Interpolation (DIIS)

- Need robust software with $\varepsilon\kappa(\mathbf{E})$ error rather than $\varepsilon\kappa(\mathbf{E})^2$ error