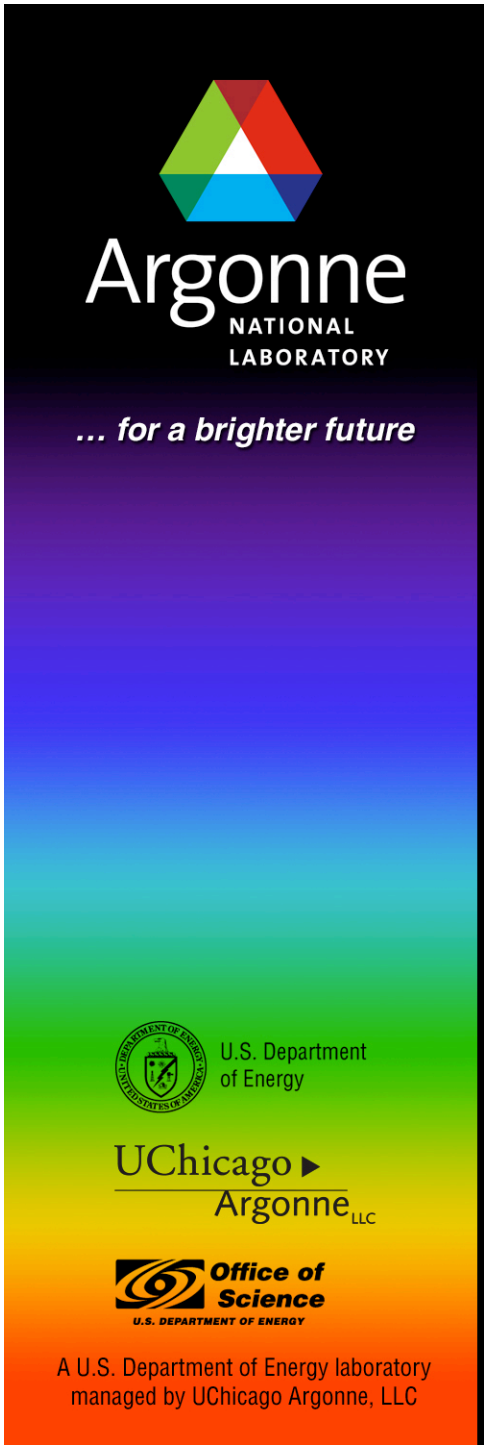


The COLUMBUS Project - General Purpose Ab Initio Quantum Chemistry

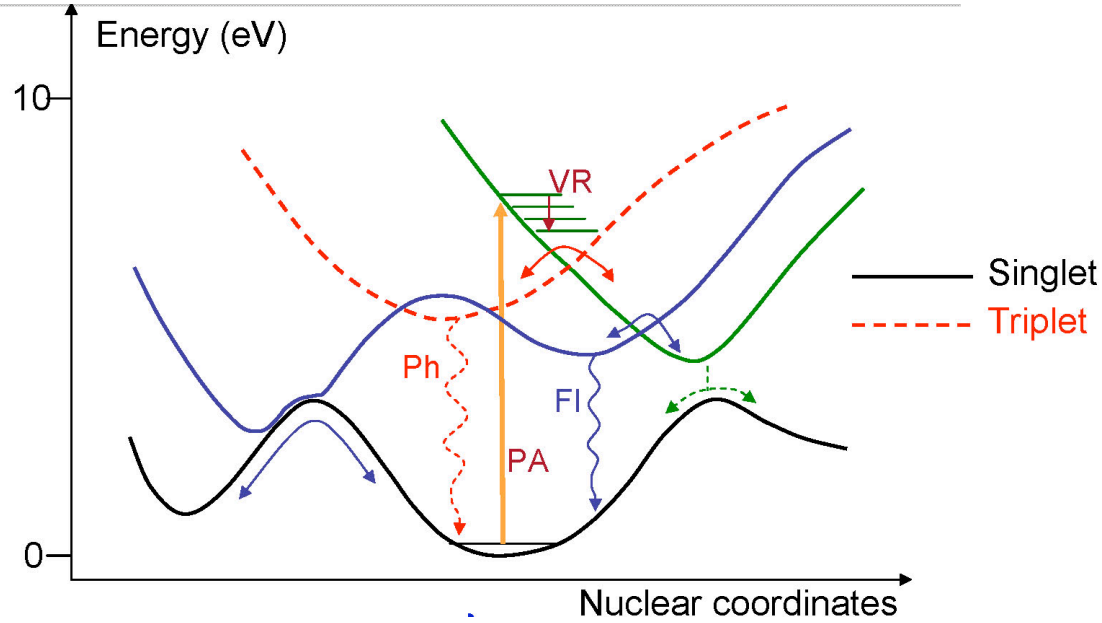
I. Background and Overview

Ron Shepard
Chemistry Division
Argonne National Laboratory

CScADS Workshop, Snowbird, Utah, July 23, 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
Fl 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)

Real Symmetric Eigenvalue Problem

- Use the iterative Davidson Method for the lowest (or lowest few) eigenpairs
- *Direct CI*: **H** is not explicitly constructed, **w=Hv** are constructed in "operator" form
- Matrix dimensions are 10^4 to 10^9
- All floating point calculations are 64-bit

Davidson Method

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 Elements

- $H_{mn} = \langle m| H^{op} |n\rangle$
- $|n\rangle = |\phi(\mathbf{r}_1)\,\sigma_1\,\phi(\mathbf{r}_2)\sigma_2 \ldots \phi(\mathbf{r}_n)\sigma_n|$ Slater Determinant
  with $\sigma_j = \alpha, \beta$

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

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

…Matrix Elements

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

- 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 *coupling coefficients*; these are sparse and are recomputed as needed

Matrix-Vector Products

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

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

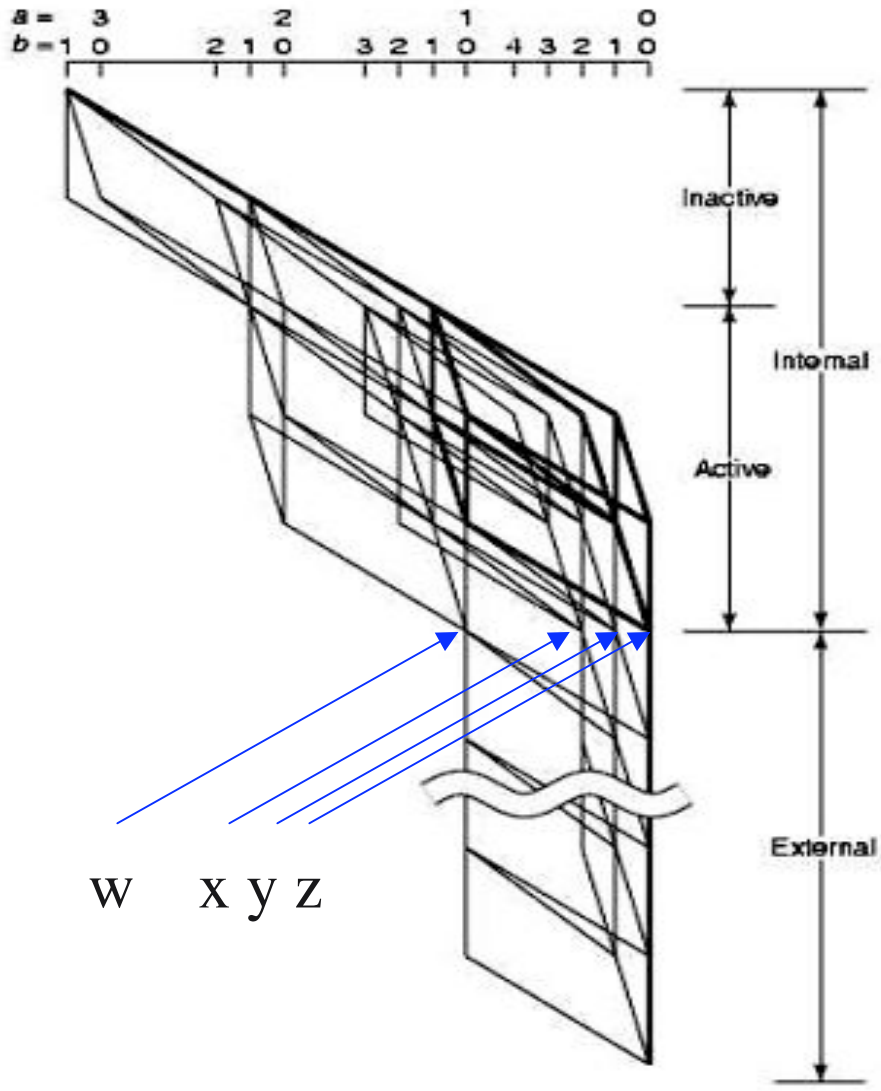
$$= \sum_{n}^{Ncsf} \sum_{p,q}^{Norb} h_{pq} \langle m | E_{pq} | n \rangle x_n + \tfrac{1}{2} \sum_{n}^{Ncsf} \sum_{p,q,r,s}^{Norb} g_{pqrs} \langle m | e_{pqrs} | n \rangle x_n$$

- The challenge is to bring together the different factors in order to compute **w** efficiently

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
- 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

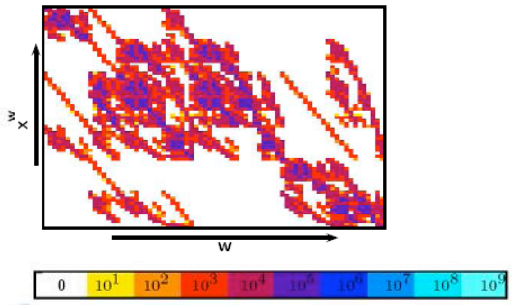
Integral Types

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

	$\underline{z}$	$\underline{y}$	$\underline{x}$	$\underline{w}$
$\underline{z}$	0	1	2	2
$\underline{y}$	1	0,2	1,3	1,3
$\underline{x}$	2	1,3	0,2,4	2
$\underline{w}$	2	1,3	2	0,2,4

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}.



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

- Total effort 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

Original Program (1980)

- Need to optimize wave functions for N_{csf}=10^5 to 10^6
- Available memory was typically 10^5 words
- Must segment the vectors, **v** and **w**, and partition the matrix **H** into subblocks, then work with one subblock at a time.

…First Parallel Program (1990)

- Networked workstations using TCGMSG
- Each matrix subblock corresponds to a compute task
- Different tasks require different resources (pay attention to load balancing)
- Same vector segmentation for all g_{pqrs} types
- g_{pqrs}, $\langle m| e_{pqrs} |n\rangle$, **w**, and **v** were stored on external shared files (file contention bottlenecks)

Current Parallel Program

- Eliminate shared file I/O by distributing data across the nodes with the GA Library
- Parallel efficiency depends on the vector segmentation and corresponding **H** subblocking
- Apply different vector segmentation for different g_{pqrs} types
- Treat upper- and lower-triangle parts separately (larger segments per task or less memory per task)
- Tasks are timed each Davidson iteration, then sorted into decreasing order and reassigned for the next iteration in order to optimize load balancing
- Capable of optimizing expansions up to N_{csf}=10^9