International Symposium Nov.30/Dec.3: h3-Open-BDEC:
Towards Society 5.0 by Integration of (Simulation+Data+Learning)
Information Technology Center, The University of Tokyo– Nov.30 & Dec.3, 2021

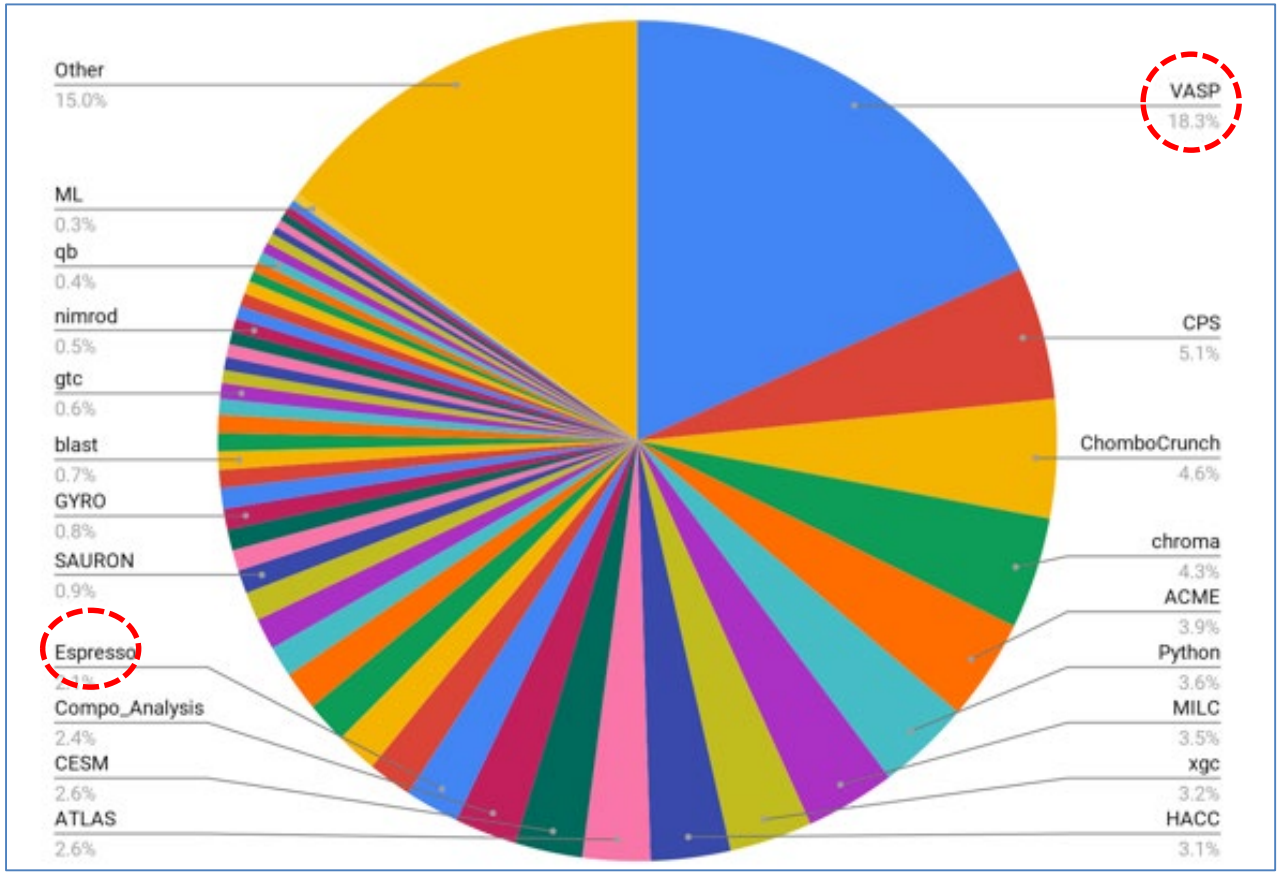
Revisiting Minimization Strategies for Solving Eigenvalue Problems

Osni Marques
Lawrence Berkeley National Laboratory
oamarques@lbl.gov

Joint work with Doru Thom Popovici, Mauro Del Ben and Andrew Canning
funding from DOE's SciDAC FASTMath

How are computer cycles used?

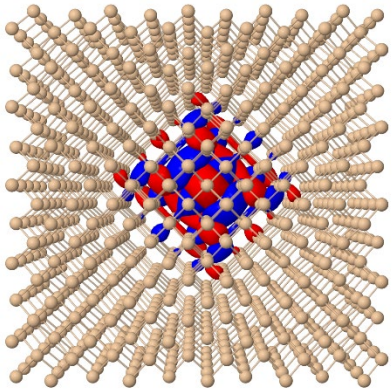
NERSC System Utilization (Aug'17 - Jul'18)

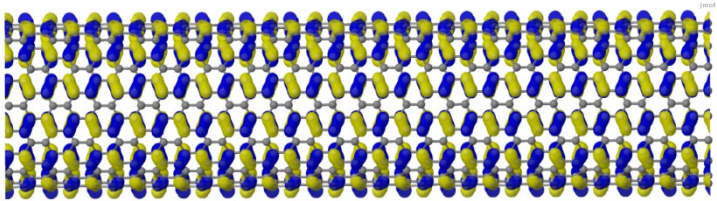


- electronic structure DFT eigenvalue problems ~ 25% of the workload
- 10 codes > 50% of the workload
- 35 codes > 75% of the workload
- Over 600 codes comprise the remaining 25% of the workload.

Electronic Structure of Materials

- Schrödinger equation: $\hat{H}\Psi = E\Psi$, $\Psi(\vec{r}_1, \ldots, \vec{r}_n)$
  - Many-particle equation
  - Very expensive to be solved (exponential)
  - Unpractical for large systems
- Density Functional Theory (DFT): $H\psi_i = E_i\psi_i$
  - Kohn and Pople, Nobel Prize in Chemistry, 1998
  - Maps the many-particle problem into a single-particle problem
  - Accurate results for structural and electronic properties of materials
  - Need to be solved self−consistently
  - $O(N^3)$ scaling with system size





Self-Consistency: Nonlinear Eigenvalue Problem

initial guess $\{\psi_i\}$

↓

calculate density

$$\rho(\vec{r}) = \sum_{i=1}^{N} |\psi_i(\vec{r})|^2$$

↓

update $H(\rho)$

↓

solve $H\psi_i = E_i\psi_i$

new set $\{\psi_i\}$

Direct Methods

- ScaLAPACK
- EigenExa
- ELPA

Iterative Methods

- only a small fraction (2-10%) of (smallest) eigenpairs is required
- limited/poor parallel performance for conventional diagonalization and/or reorthogonalization, $O(N^3)$

$$H\psi_i(r) = \left[-\frac{1}{2}\nabla^2 + V\right]\psi_i(r) = \varepsilon_i\psi_i(r)$$

$$\psi_i(r) = \sum_{j=1}^{m} c_{ji}\varphi_j(r)$$

Iterative Methods for $H\psi_i = E_i\psi_i$

- (Jacobi-)Davidson
- Locally Optimal Block Preconditioned Conjugate Gradient (LOBPCG)
- (Polynomial filtered) Lanczos
- Conjugate gradient minimization of $\psi_i^* H\psi_i$

References

- *Sameh and Wisniewski, A trace minimization algorithm for the generalized eigenvalue problem, SIAM Journal on Numerical Analysis, 19, 1243-1259 (1982)*
- *Teter, Payne, Allen, Solution of Schrodinger's equation for large systems, Phys. Rev. B 40, 12255 (1994)*
- *Mauri and Galli, Electronic-structure calculations and molecular-dynamics simulations with linear system-size scaling, Phys. Rev. B 50, 4316 (1994)*
- *Kresse and Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54, 11169 (1996)*
- *Pfrommer, Demmel and Simon, Unconstrained energy functionals for electronic structure calculations, J. Comput. Phys., 150 (1999)*
- *Saad, Chelikowsky and Shontz, Numerical methods for electronic structure calculations of materials, SIAM Review 52, 3–54 (2010)*
- *Jordan, Marsman, Kim and Kresse, Fast iterative interior eigensolver for millions of atoms, J. Comput. Phys., 231 (2012)*
- *Corsetti, The orbital minimization method for electronic structure calculations with finite-range atomic basis sets, Comput. Phys. Comm., 185 (2014)*
- *Levitt and Torrent, Parallel eigensolvers in plane-wave Density Functional Theory, Comput. Phys. Comm., 187 (2014)*

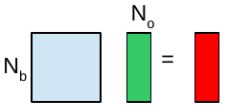
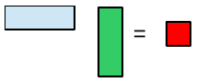
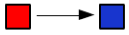
Standard (constrained) iterative CG eigensolver versus unconstrained iterative CG eigensolver

- Constrained CG method for iterative eigensolver
  - $\min_{\Psi} \mathrm{Tr}\,[\Psi^T H \Psi], \Psi = [\psi_1, \psi_2, \ldots \psi_N], \Psi^T \Psi = I$
  - CG steps followed by reorthogonalization with ScaLAPACK
  - Typically matrix size 100,000 to millions (dimension of H)
  - Operations on H and ψ_i (matrix vector for CG steps) scale well
  - Operations on small subspace scale poorly (reorthogonalization)
- Unconstrained CG method for iterative eigensolver (simplest form)
  - $\min_{\mathrm{X}} \mathrm{Tr}\,[\mathcal{S}^{-1}\mathrm{X}^T H \mathrm{X}],\ \mathcal{S} = \mathrm{X}^T\mathrm{X}, \Psi = \mathrm{X}\mathcal{S}^{-\frac{1}{2}}$
  - $\mathcal{S}^{-1} \approx (2I - \mathcal{S})$ (1st order expansion)
  - Functional has same minimum as constrained functional (trial eigenvectors orthogonal at minimum)
  - No operations on subspace matrix (scales to large core counts)
  - Convergence properties different from constrained functional

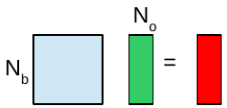
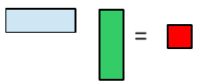
Operations for constrained iterative CG eigensolver and unconstrained iterative CG eigensolver

N_b = matrix dimension, N_o = number of eigenpairs (1-10% of N_b), p = number of processors

Constrained Solver

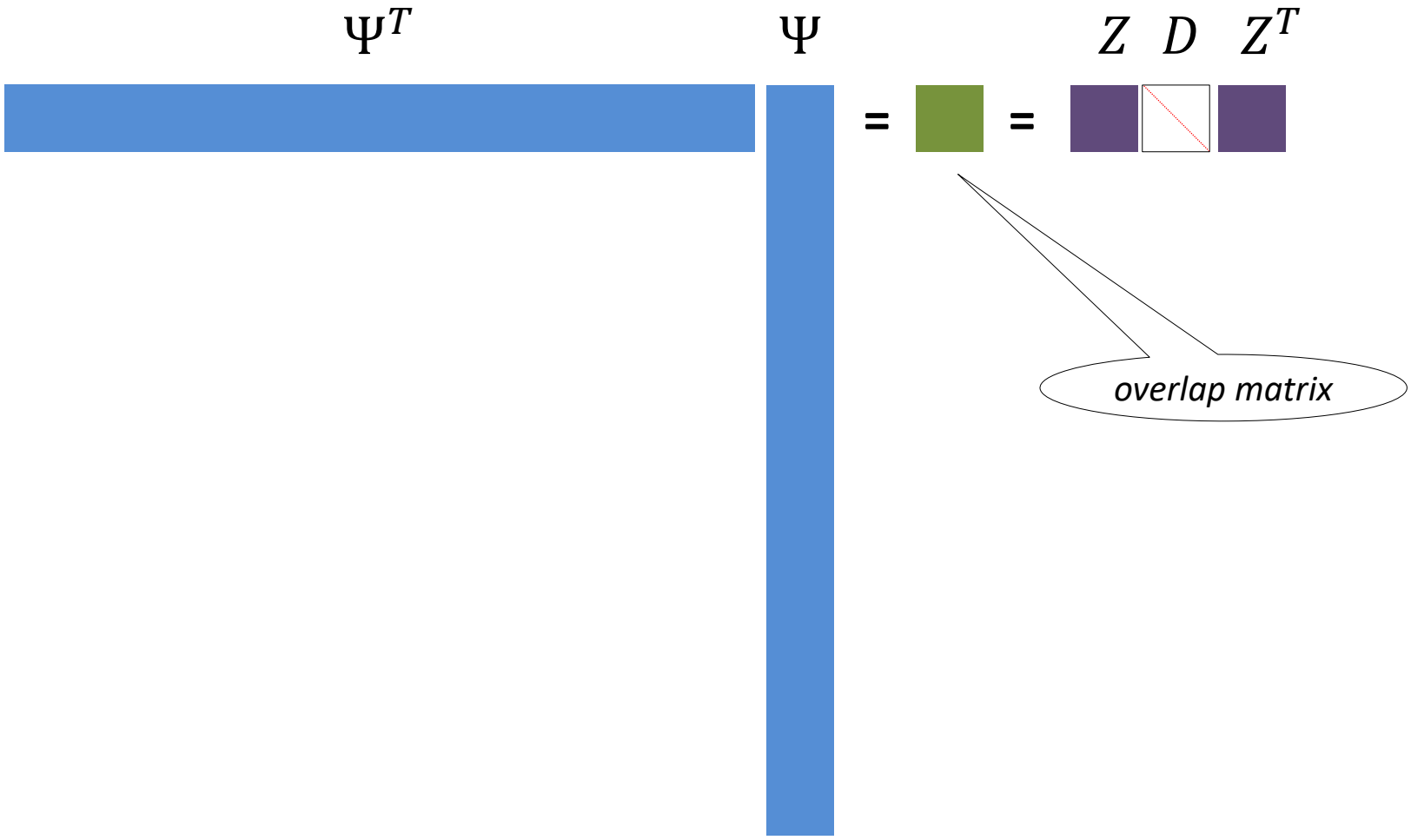
Operation	Label	Comput.	Commun.
	HX	$\frac{N_b^2 \cdot N_o}{p}$	$\frac{N_b \cdot (N_o + N_b)}{\sqrt{p}}$
	XtP	$\frac{N_b \cdot N_o^2}{p}$	$\frac{2N_b \cdot N_o}{\sqrt{p}}$
	XB	$\frac{N_b \cdot N_o^2}{p}$	$\frac{N_o \cdot (N_b + N_o)}{\sqrt{p}}$
	Transf.	$\frac{N_o^3}{p}$	$\frac{N_o^2}{\sqrt{p}}$

Unconstrained Solver

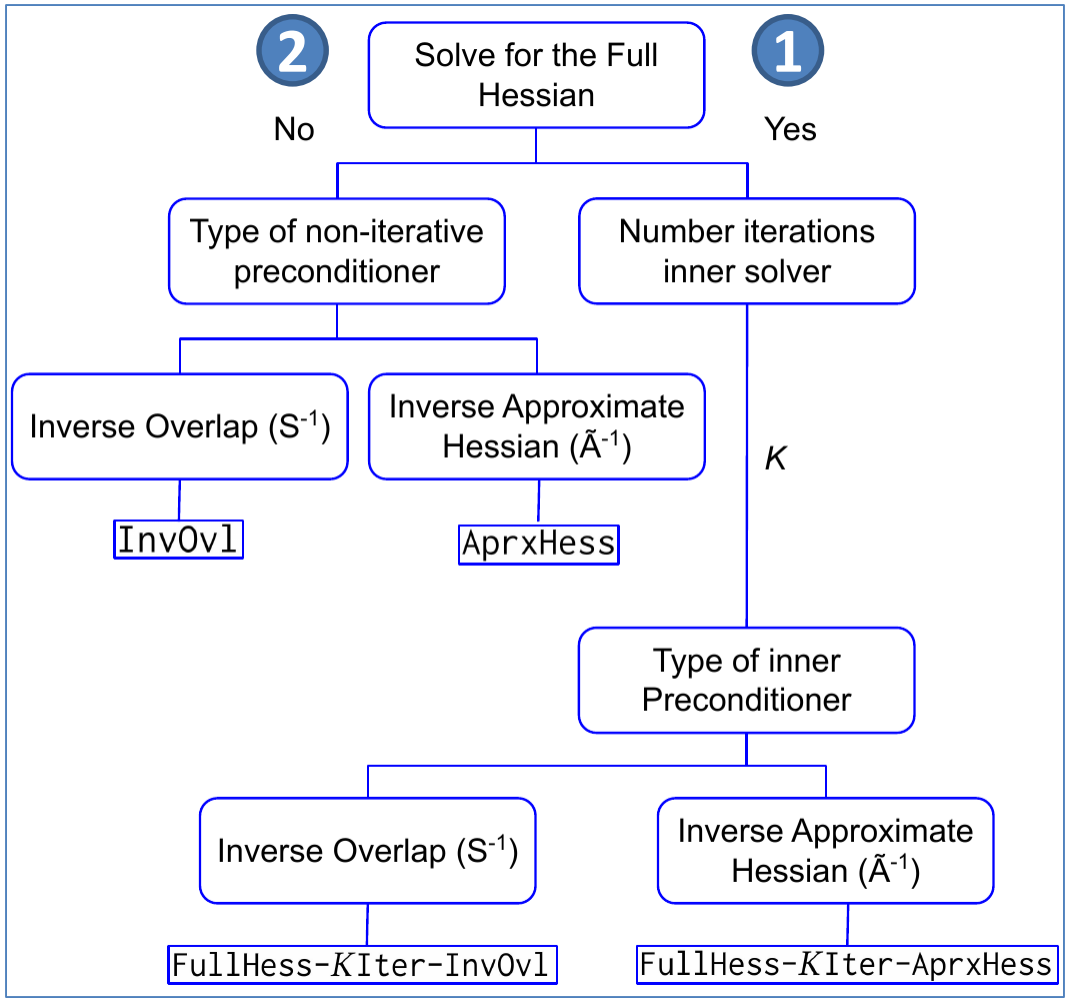
Operation	Label	Comput.	Commun.
	HX	$\frac{N_b^2 \cdot N_o}{p}$	$\frac{N_b \cdot (N_o + N_b)}{\sqrt{p}}$
	XtP	$\frac{N_b \cdot N_o^2}{p}$	$\frac{2N_b \cdot N_o}{\sqrt{p}}$
	XB	$\frac{N_b \cdot N_o^2}{p}$	$\frac{N_o \cdot (N_b + N_o)}{\sqrt{p}}$

- Important questions for constrained and unconstrained eigensolvers:
  - Convergence rate
  - Parallel scaling
  - Stability
- Unconstrained formulation can be applied to other matrices
  - Tested on Harwell-Boeing matrices

PCG for $Ax = \lambda x$: *orthogonality versus scalability*



Novel Preconditioners for PCG



$$\min_{X} \mathrm{Tr}\left[\mathcal{S}^{-1}\mathrm{X}^{T}HX\right]$$
$$\mathcal{S}^{-1} \approx (2I - \mathcal{S})$$
$$G = 4HX - 2SX\mathcal{H} - 2HX\mathcal{S}$$
$$\mathcal{H} = X^{T}HX$$

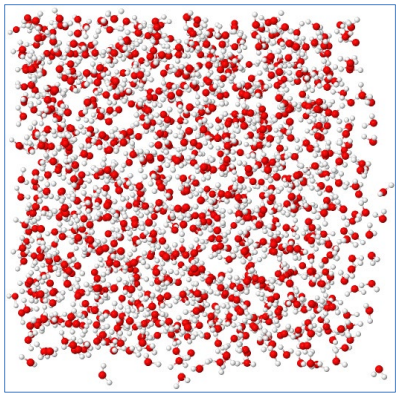
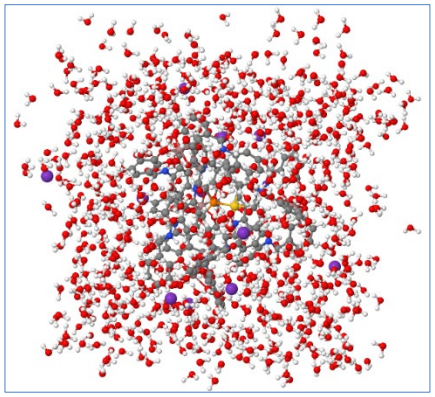
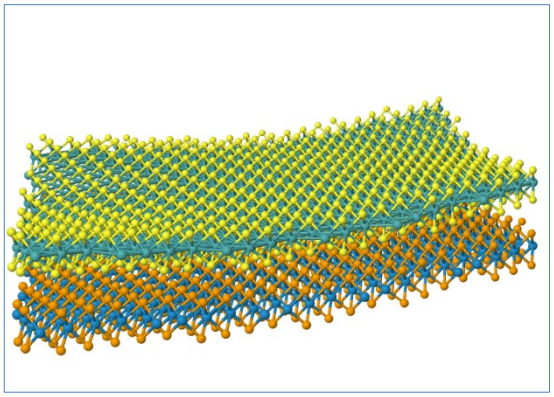
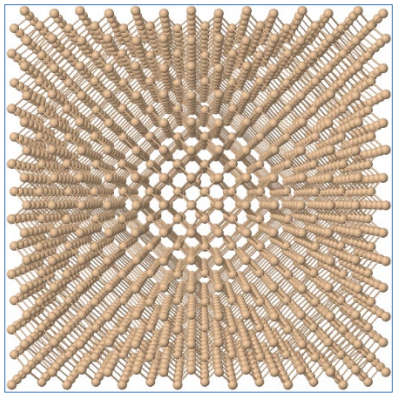
- **Option 1**: Hessian A of the unconstrained functional to precondition the gradient, $A^{-1}G$ (quasi Newton step)
  - solve $AP = G$ iteratively
  - S^{-1} or $\tilde{A}^{-1}$ as preconditioner for the inner solver (with K iterations)
  - $\tilde{A} \approx A$
- **Option 2**: use S^{-1} or $\tilde{A}^{-1}$ to precondition the unconstrained functional minimization

Numerical Experiments

- CP2K
  - quantum chemistry and solid state physics package
  - DFT using mixed Gaussian and plane waves approaches
  - Non-orthogonal basis, generalized eigenvalue problem $HC = SCE$
- Cray XC40 system (cori @ NERSC)
  - 2,388 Intel Xeon 16-core Intel Xeon Haswell
  - 9,688 68-core Intel Xeon Phi Knights Landing (KNL)
  - Hybrid MPI+OpenMP implementation
  - Intel compiler, MKL, ELPA, and LIBXSMM (latest available releases)

Del Ben, Marques and Canning, Improved Unconstrained Energy Functional Method for Eigensolvers in Electronic Structure Calculations, ICPP 2019, Kyoto, Japan.

Systems Used in the Numerical Experiments

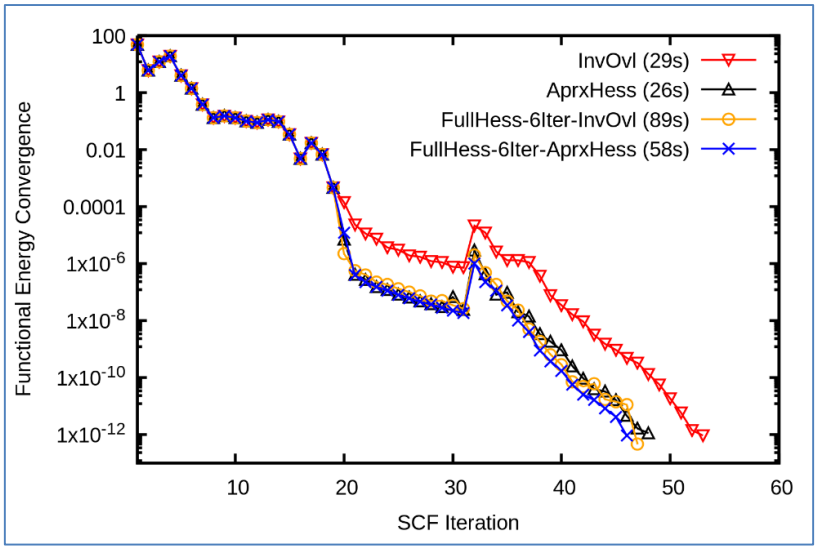
Systems used in the numerical experiments, in increasing order of "complexity" for convergence: 1024 molecules of bulk liquid water, supramolecular catalyst gold(III)-complex, bilayer of MoS_2-WSe_2, and divacancy point defect in silicon. The number of atoms range from 2,247 to 12,288.

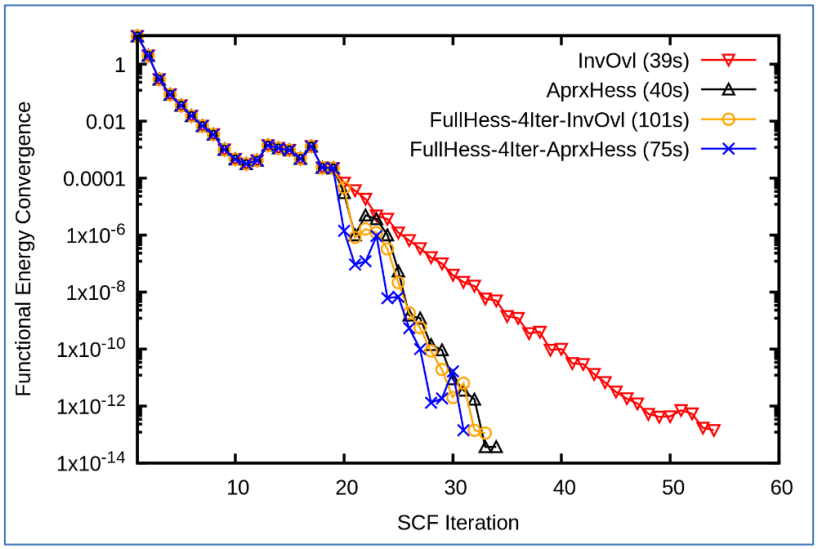
Physical and Computational Parameters

System	Label	Atoms	Basis	N_b	N_o	N_b/N_o	gap(AU)
Bulk liquid water	Water-1024	3,072	TZVP	29,696	4,096	0.14	0.128
	Water-2048	6,144	TZVP	59,392	8,192	0.14	
	Water-4096	12,288	TZVP	118,784	16,384	0.14	
Solvated catalyst complex	Complex	2,590	TZVP	26,339	3,605	0.14	0.052
MoS_2-WSe_2 bilayer	BiLayer	2,247	TZVP	51,681	9,737	0.19	0.035
Divacancy defect in silicon	SiDivac	2,742	TZVP	46,614	5,484	0.12	0.013
	SiDivac-SZV	2,742	SZV	10,968	5,484	0.5	
	SiDivac-DZVP	2,742	DZVP	35,646	5,484	0.15	
	SiDivac-TZV2P	2,742	TZV2P	79,518	5,484	0.07	

Basic physical and computational parameters of the systems employed in the numerical experiments. N_b is the basis set size, N_o is the number of eigenvectors to be computed (number of wavefunctions needed to build the electronic density), and gap is the energy difference between eigenvalues N_o and N_o+1 in atomic units (AU), the unit employed to express H.

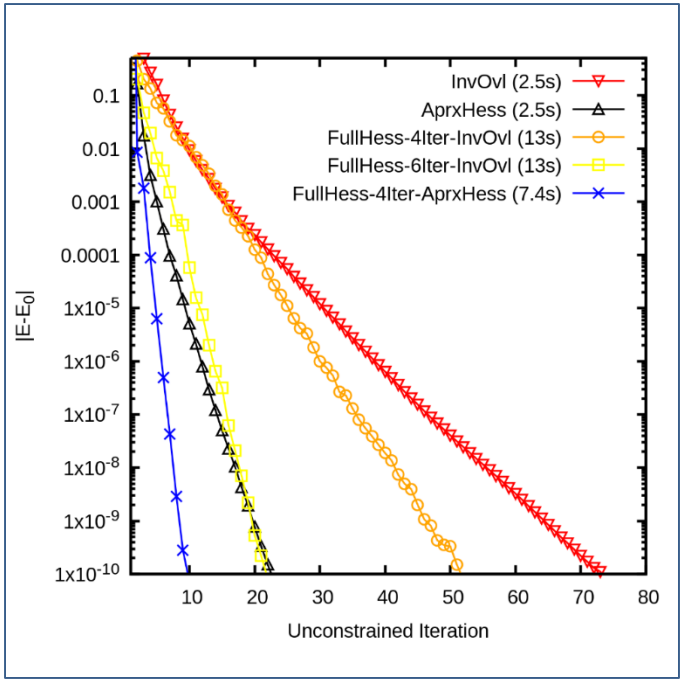
Convergence of the SCF Procedure

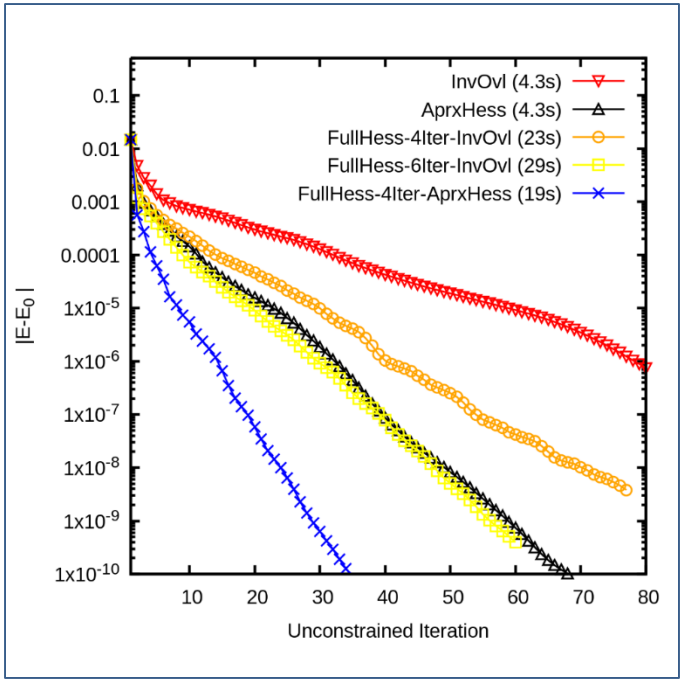




Convergence of the SCF procedure. Four setups; the average time for a single SCF step is given in parenthesis. Left: Complex. Right: SiDivac.

Convergence of the Energy

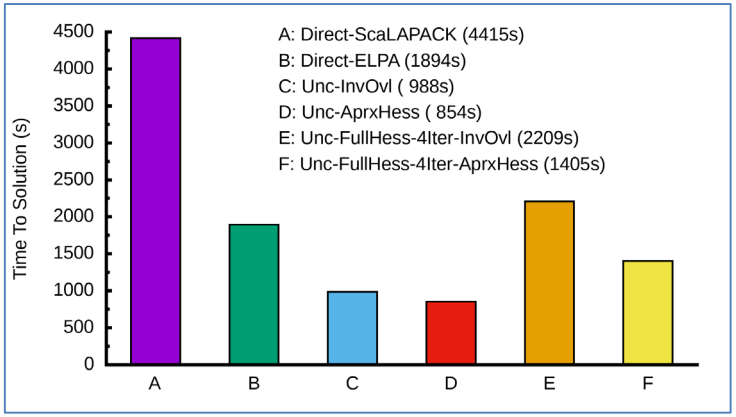




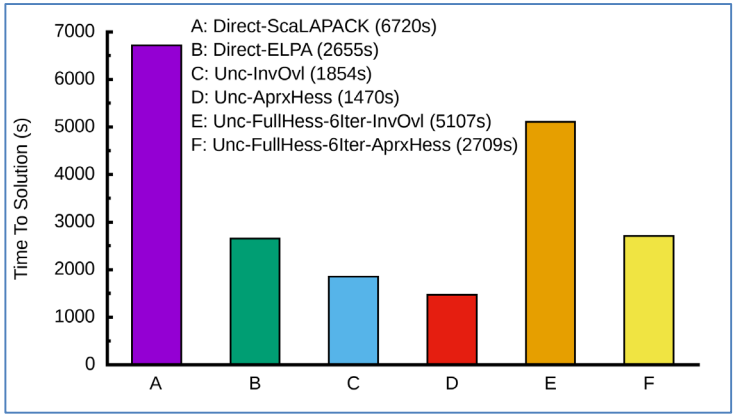
Convergence of the energy (unconstrained objective function) for a single unconstrained functional diagonalization (unconstrained subspace minimization). Five setups; the time for a single unconstrained-PCG iteration is given in parenthesis. Left: Complex. Right: SiDivac.

Time to Solution for Full SCF

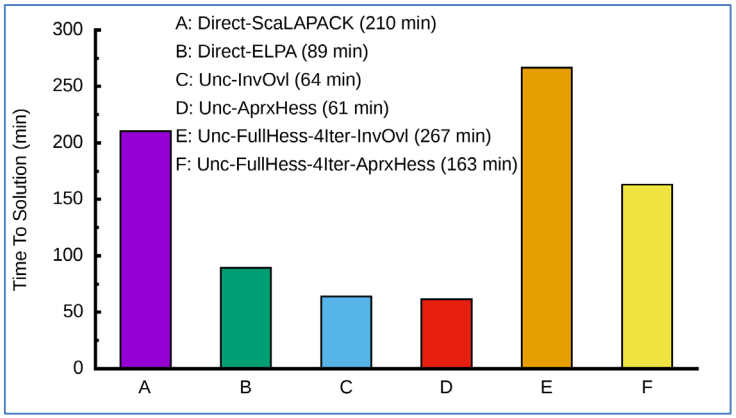
(a)



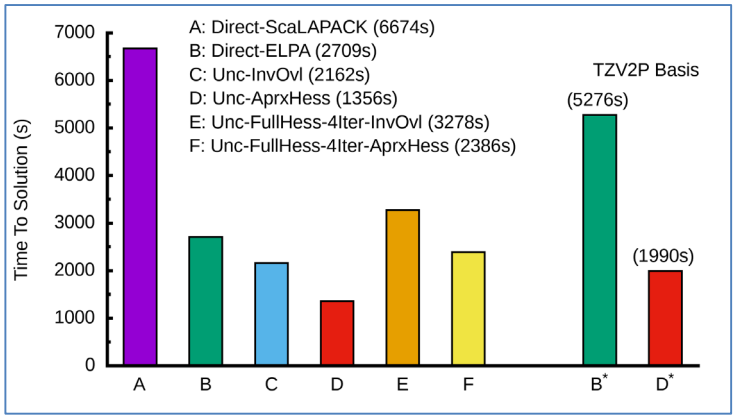
(b)



(c)

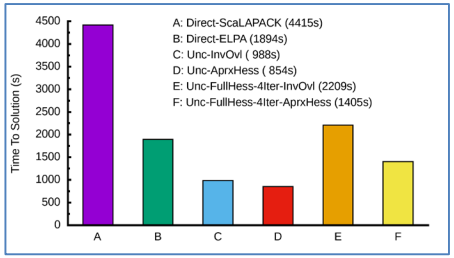


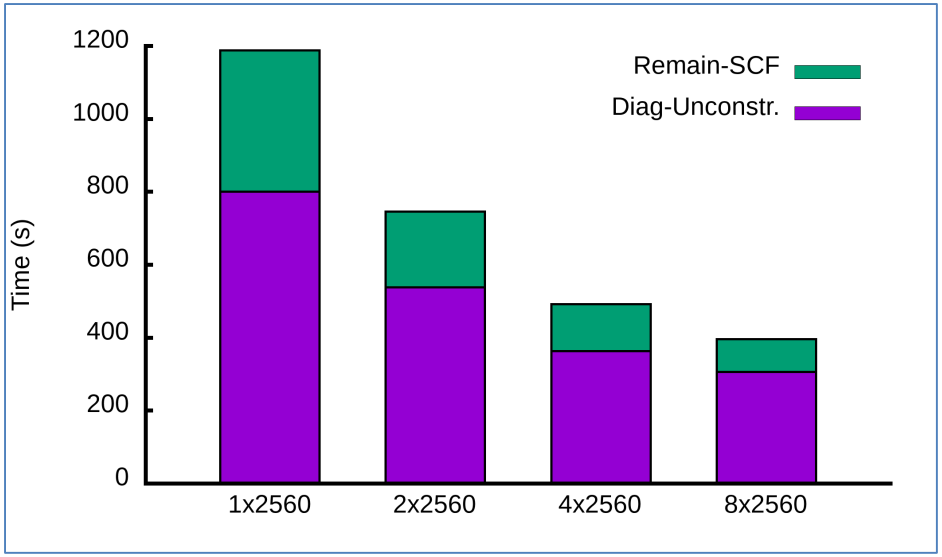
(d)

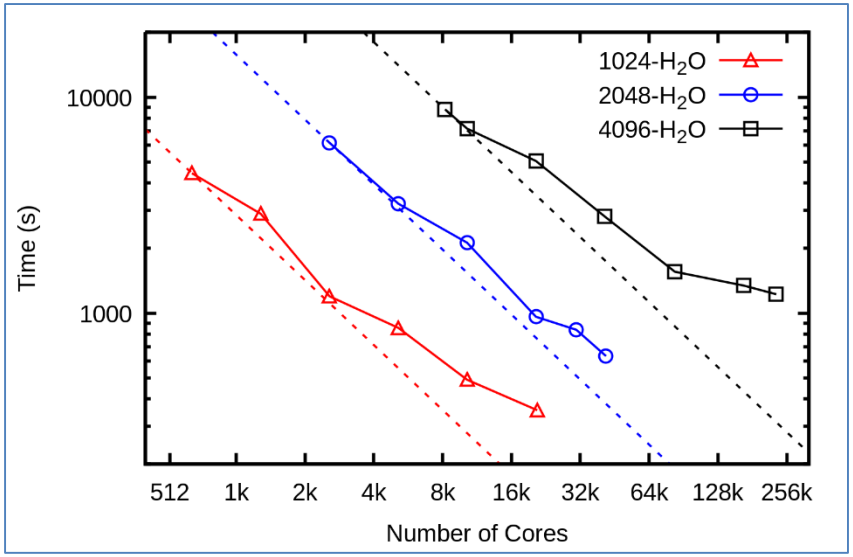


Time to solution for full SCF convergence compared to direct solvers (ScaLAPACK and ELPA). (a) Water-1024, (b) Complex, (c) BiLayer and (d) SiDivac. Actual times are given in parenthesis. For SiDivac, B^* and D^* are times obtained with a larger basis (about 1.7 times larger than in B and D, with 160 KNL nodes).

Strong Scaling







Left: OpenMP threads per MPI task for a fixed number of MPI tasks (2560), Water-1024 (method D in the figure above). Right: time to solution for bulk liquid water with 1024, 2048 and 4096 molecules (method D in the previous slide).

Summary

Main conclusions:

- Unconstrained CG offers good parallel scalability and outperforms standard diagonalization
- Implementation within a localized basis set allows for efficient sparse and dense linear algebra implementations

Ongoing and future work:

- Sub-group parallelization for small matrix multiplication
- Implementation in a plane wave basis framework (http://qboxcode.org)
- GPU implementation and comparisons with other iterative strategies
- Applications of unconstrained minimization in other areas

Thank you !