New algorithms and improvements to the BerkeleyGW excited state code

James R. Chelikowsky, Andrew Canning, Jack R. Deslippe, Felipe H. da Jornada, Steven G. Louie, Jeffrey B. Neaton, Johannes Lischner, Fang Liu, Diana Y. Qiu, Jaime S. de Sousa, Sahar Sharifzadeh, Derek Vigil-Fowler, Chao Yang

BERKELEY LAB

U.S. DEPARTMENT OF ENERGY | Office of Science

Dimension reduction techniques for the GW self energy approximation

Motivation

- The GW approximation to the self energy requires the frequency dependent screening dielectric matrix $\epsilon(\omega) = I - V\chi_0(\omega)$ to be computed efficiently.
- The standard approach requires computing the irreducible polarizability operator $\chi_0(\omega)$, which in turn requires all eigenpairs of a single particle Hamiltonian to be computed.
- **Goal**: Eliminate the need to compute the unoccupied eigenpairs.

Compute $v\chi_0$ without forming χ_0 explicitly

- It follows from eigenvector perturbation analysis that $V\chi_0(\omega)$ can be obtained by solving the Sternheimer equation
  $(H - \varepsilon_v + \omega + i\eta)\Delta\Psi_{v,\omega} = -P^{\perp}_{occ}\,(\mathrm{Diag}(\psi_v)V) \qquad (P^{\perp}_{occ} = I - \sum_v \psi_v\psi_v^*)$
  $(V\chi_0)(\omega) = 2\left[\sum_v \mathrm{Diag}(\psi_v^*)\Delta\Psi_{v,\omega}\right]$
- For each frequency, a brute force approach would require solving $n \times n \times n_v$ equations to obtain an $n \times n$ dielectric matrix.
- Use dimension reduction techniques to reduce the number of equations to be solved.

Low-rank structure of $v\chi_0$

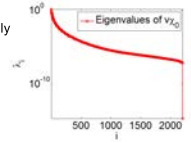
Wison, Lu, Gygi, Galli (2009)

- Eigenvalues of $V\chi_0$ decrease to 0 rapidly
- Low-rank approximation: $V\chi_0 \approx U\Lambda U^*$

$V\chi_0$ low rank

⇒ $\Delta\Psi$ low rank?

⇒ $-P^{\perp}_{occ}\,(\mathrm{Diag}(\psi_v)V)$ low rank?



Constructing low-rank approximation to ϵ and Σ

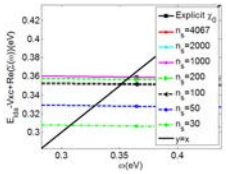
- $-P^{\perp}_{occ}\,(\mathrm{Diag}(\psi_v)v) \approx U_v\Sigma_v V_v^*$ (Frequency independent!)
- Solve $(H - \varepsilon_v + \omega + i\eta)\Delta\Psi_v = U_v$ ($n \times n_s$ matrix, $n_s \ll n$)
- $\epsilon(\omega) \approx I - \sum_v \mathrm{Diag}(\psi_v^*)\,\Delta\Psi_v\Sigma_v V_v^*$
- Further rank reduction $\epsilon(\omega) \approx I - X(\omega)CY^*$
- $\epsilon^{-1}(\omega) = I + X(\omega)C^{-1}Y^*$ (Inverting an $n_s \times n_s$ matrix C!)
- $\langle\psi_n|\Sigma_c(\omega)|\psi_n\rangle = tr(Z_Y^*H^{-1}(\omega)Z_X(\omega)C^{-1})$
  $Z_X(\omega) = \mathrm{Diag}(\psi_n)X(\omega), \quad Z_Y = \mathrm{Diag}(\psi_n)Y$

Numerical results for a methane molecule

Accuracy in $\epsilon^{-1}(0)$

Truncation level	Our approach	Galli's approach
2205	0.16E-15	0.80E-14
2000	0.44E-07	0.48E-07
1000	0.18E-05	0.10E-05
200	0.11E-03	0.22E-03
100	0.44E-03	0.11E-02
50	0.15E-02	0.38E-02
20	0.52E-02	0.14E-01

Accuracy in $\Sigma(\omega)$



Observation

For small molecules, a 90% truncation of the singular values/vectors of $V\chi_0$ yields sufficiently accurate approximation of the self energy and quasiparticle self energy. Both computational time and storage requirement can be reduced significantly by the low-rank approximation.

Improved Interpolation Schemes

Motivation

- Excitons are correlated electron-hole pairs which can be predicted from the solutions of the Bethe-Salpeter equation (BSE).
- Because of their correlated nature, it's necessary to use interpolation schemes to solve the BSE.

Interpolation of the Bethe-Salpeter Equation

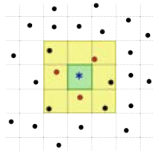
- In BerkeleyGW, we calculate the Kernel of the BSE on a coarse grid and interpolate them using the projection between wave functions the **coarse** and **fine** grids.

$$\langle vc\mathbf{k}_{\mathrm{fi}}|K|v'c'\mathbf{k}'_{\mathrm{fi}}\rangle = \sum_{i_1,i_2,i_3,i_4} C^{\mathbf{k}_{\mathrm{co}}}_{c,n_1} C^{*\mathbf{k}_{\mathrm{co}}}_{v,n_2} C^{*\mathbf{k}'_{\mathrm{co}}}_{c',n_3} C^{\mathbf{k}'_{\mathrm{co}}}_{v',n_4} \langle n_2 n_1 \mathbf{k}_{\mathrm{co}}|K|n_4 n_3 \mathbf{k}'_{\mathrm{co}}\rangle$$

- We implemented space decomposition algorithms and caching techniques to speedup the interpolation: over **10^4 speedup** for graphene.

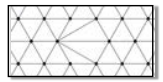
Before: $O(N_{\mathrm{co}}^2 N_{\mathrm{fi}}^3)$

After: $O(N_{\mathrm{co}} N_{\mathrm{fi}})$

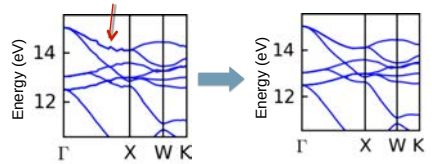


Delaunay Tessellation

- In order to obtain smooth interpolated band structures, we implemented an algorithm that first tessellates the k-points based on Delaunay triangulation. This removes interpolation discontinuities.



Delaunay tessellation of an inhomogeneous k-mesh.



Old interpolation algorithm | New algorithm

Ongoing and Future Works

- Improve the interpolation of the dielectric matrix, which is ill-behaved for systems with reduced dimensionality.
- Support interpolation of kernels generated without the Tamm-Dancoff approximation (partially implemented)

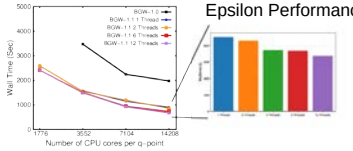
Collaborators **Developers:** Felipe H. da Jornada, Jack R. Deslippe and Steven G. Louie

Node Level Parallelism in BerkeleyGW 1.1

Motivation

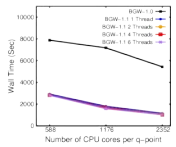
- Improve the scalability of GW implementations to massively parallel DOE machines – utilizing both inter- and intra-node parallelism

On Node Parallelism



Epsilon Performance

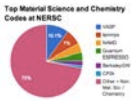
- **2X – 10X performance improvement throughout Package**
- **Hybrid OpenMP/MPI model added to support current and next-generation DOE machines.**
- **New efficient algorithms to reduce complexity. DFT orbital requirements reduced 5x.**



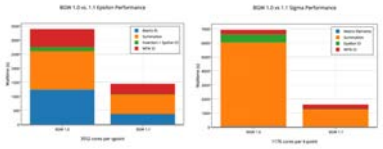
Sigma Performance

Impact

- **BerkeleyGW was 2.2% of the entire NERSC workload in the first 8 months of 2013. Compared to 1.2% in 2012.**
- **Users Study: photovoltaics, Interpretation of DOE Light-Source photoemission spectra, LED, Electronic, transport, and optical properties of novel**



BerkeleyGW 1.1 vs 1.0



- Substantial speedups throughout package from IO (next panel) OpenMP and vectorization.

Collaborators **Developers:** Jack R. Deslippe, Felipe H. da Jornada, Derek Vigil-Fowler, Fang Liu, Chao Yang and Steven G Louie

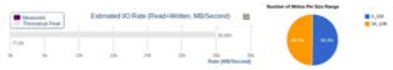
Improvements in BerkeleyGW

Motivation

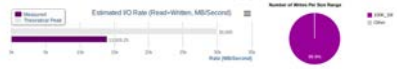
- For large systems run at scale on DOE supercomputers, IO has become a major bottleneck in BerkeleyGW Performance.
- We overcome this bottleneck with parallel-IO using DOE supported HDF5 libraries.

Parallel IO

BGW 1.0 IO Signature

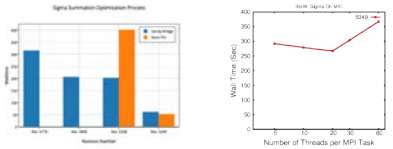


BGW 1.1 IO Signature



- **IO improved from 90 MB/sec to > 2 GB/sec**
- **Average write size increased by orders of magnitude**
- **Utilize DOE supported parallel HDF5 libraries and Lustre filesystems features: striping across multiple disks**

Support for Many-Core Architectures



Rev 4770: Initial Code

Rev 4896: Refactor code to have loops targeting MPI, OpenMP, SIMD

Rev 5338: OpenMP Pragmas added

Rev 5349: Vectorization Ensured

Ongoing and Future Works

- Parallel IO for more files formats (wavefunctions)
- Support of GPUs
- Collaboration with Intel engineers to ensure vectorization, optimization

Collaborators **Developers:** Jack R. Deslippe, Felipe H. da Jornada, Derek Vigil-Fowler, Fang Liu, David A. Strubbe, Chao Yang and Steven G. Louie

This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract No. DE-AC02-05CH11231.

MSD Materials Sciences Division