Large-Scale GW Calculations on Pre-Exascale HPC Systems

BerkeleyGW: Method Developments and Code Optimization

M. Del Ben[1], F.H. da Jornada[1,2], A. Canning[1], N. Wichmann[3], K. Raman[4], R. Sasanka[4], C. Yang[1], S.G. Louie[1,2] and J. Deslippe[1]

[1]Lawrence Berkeley National Laboratory, [2]University of California at Berkeley, [3]CRAY, [4]Intel Corporation

The *GW* Method

Solve Dyson's equation:

$$\left[-\frac{1}{2}\nabla^2 + V_{\rm loc} + \Sigma(E_n)\right]\phi_n = E_n\phi_n, \quad (1)$$

$\Sigma(E_n) \rightarrow$ self-energy (non-Hermitian, non-local, energy-dependent operator)

- Perturbative expansion on the screened Coulomb interaction W
- First approximation $\Sigma = iGW$
- W obtained from Inverse Dielectric Matrix ϵ^{-1} of the system

In `BerkeleyGW` [1]:

- `Epsilon` code $\rightarrow$ Compute ϵ^{-1}
- `Sigma` code $\rightarrow$ Compute W from ϵ^{-1} and solve eq.1

`Epsilon` code: Inverse Dielectric Matrix ϵ^{-1}

Input: $\psi_{m\mathbf{k}}$, $\epsilon_{m\mathbf{k}}$, $\{\mathbf{q}\text{-points}\}$, $\{\omega_i\}$

1. Calculate plane-waves matrix elements (FFT's): $O(N_v N_c N_G \log N_G)$

$$M^G_{ja\mathbf{k}}(\mathbf{q}) = \langle\psi_{j\mathbf{k}+\mathbf{q}}|\, e^{i(\mathbf{G}+\mathbf{q})\cdot\mathbf{r}}\,|\psi_{a\mathbf{k}}\rangle$$

2. Calculate RPA polarizability (Matrix-Multiplication): $O(N_\omega N_v N_c N_G^2)$

$$\chi(\mathbf{q},\omega_i) = \mathbf{M}(\mathbf{q})^\dagger \mathbf{\Delta}_{ja\mathbf{k}}(\epsilon_{j\mathbf{k}}, \epsilon_{a\mathbf{k}}, \mathbf{q}, \omega_i)\mathbf{M}(\mathbf{q})$$

$\mathbf{\Delta}$ diagonal matrix containing the frequency dependence

3. Dielectric matrix ϵ and inversion: $O(N_\omega N_G^3)$

$$\epsilon_{GG'}(\mathbf{q},\omega_i) = \delta_{GG'} - v(\mathbf{q}+\mathbf{G})\chi_{GG'}(\mathbf{q},\omega_i)$$
$$\epsilon^{-1}(\mathbf{q},\omega_i) = (I - v\chi(\mathbf{q},\omega_i))^{-1}$$

Parameters: N_v number of valence bands, N_c number of conduction (empty) bands, N_G PW basis set size, N_ω number of frequencies.

Static Subspace Approximation [2]: Speed-Up the Calculation of $\epsilon^{-1}(\omega_i)$ for $\omega_i \neq 0$

- For $\omega_i = 0$: Standard calculation of $\bar{\chi}(0) = v^{\frac{1}{2}}\chi(0)v^{\frac{1}{2}}$

$$\chi(0) = \mathbf{M}^\dagger \mathbf{\Delta}_{ja\mathbf{k}}(0)\mathbf{M}$$

  Eigendecomposition $\bar{\chi}(0) = \mathbf{C}^{0\dagger}\mathbf{x}\mathbf{C}^0$, define $\mathbf{C}^0_s$ according to threshold $t_{\rm eigen}$
- Projection of $\mathbf{M}$ into the subspace spanned by $\mathbf{C}^0_s$

$$\overline{\mathbf{M}}^0_s = \mathbf{M}v^{\frac{1}{2}}\mathbf{C}^0_s$$

- For $\omega_i \neq 0$: direct computation of $\bar{\chi}_s(\omega_i)$

$$\bar{\chi}_s(\omega_i) = \overline{\mathbf{M}}^{0\dagger}_s \mathbf{\Delta}_{ja\mathbf{k}}(\omega_i)\overline{\mathbf{M}}^0_s$$

- Final evaluation of $\epsilon^{-1}(\omega_i)$ from $\bar{\chi}_s(\omega_i)$

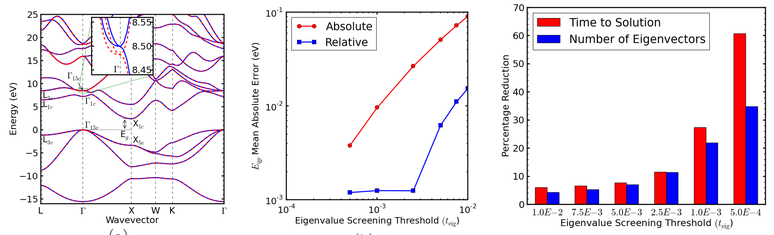
	Execution	Memory
Matrix Element	$O(N_v N_c N_G \log N_G)$	$O(N_v N_c N_G)$
Polarizability $\omega = 0$	$O(N_v N_c N_G^2)$	$O(N_G^2)$
Eigendecomposition: $\mathbf{C}^0_s$	$O(N_G^3)$	$O(N_G^2)$
$\overline{\mathbf{M}}^0_s$	$O(N_b N_v N_c N_G)$	$O(N_v N_c N_b)$
Polarizability $\omega \neq 0$	$O(N_\omega N_v N_c N_b^2)$	$O(N_\omega N_b^2)$
Inversion	$O(N_\omega N_b^3)$	$O(N_\omega N_b^2)$
I/O	$O(N_G N_b + N_\omega N_b^2)$	$O(N_G N_b + N_\omega N_b^2)$
Evaluation of $\epsilon^{-1}(\omega_i)$	$O(N_\omega N_b N_G^2)$	$O(N_\omega N_G^2)$

References

1. J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, Comput. Phys. Commun. 183, 1269 (2012).
2. M. Govoni and G. Galli, J. Chem. Theory Comput. 11, 2680 (2015) ; T.A. Pham, H.V. Nguyen, D. Rocca, and G. Galli, Phys. Rev. B 87, 155148 (2013) ; H.-V. Nguyen, T. A. Pham, D. Rocca, and G. Galli, Phys. Rev. B 85, 081101 (2012) ; H. F. Wilson, D. Lu, F. Gygi, and G. Galli, Phys. Rev. B 79, 245106 (2009) ; H. F. Wilson, F. Gygi, and G. Galli, Phys. Rev. B 78, 113303 (2008).
3. P. Umari, Geoffrey Stenuit, and Stefano Baroni, Phys. Rev. B 79, 201104-R (2009)
4. M. Del Ben, F.H. da Jornada, A. Canning, N. Wichmann, K. Raman, R. Sasanka, C. Yang, S.G. Louie and J. Deslippe, In Preparation

Benchmark Calculations

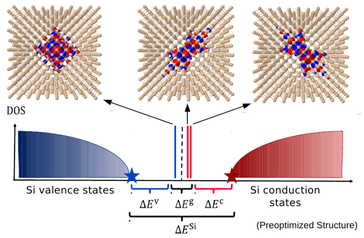
Silicon Carbide (β−SiC):



(a) (b) (c)

Figure: (a) band structure as obtained without (solid blue) and with the static subspace approximation (red dotted $t_{\rm eig} = 0.01$); (b) mean absolute error between the reference and approximate results (196 $E_{\rm QP}$ calculated); (c) percentage reduction in time to solution (red) and number of eigenvectors (blue) for the evaluation of ϵ^{-1}.

Single Vacancy V^0_1 (1000-Si Atoms Benchmark)



	ΔE^v	ΔE^g	ΔE^c	$\Delta E^{\rm Si}$	#Nodes	Time (min)
LDA	0.29	0.07	0.23	0.60	-	-
G_0W_0 (6Ry)	0.41	0.27	0.46	1.15	480	73
G_0W_0 (12Ry)	0.42	0.26	0.49	1.18	2048	157

Energies in eV
G_0W_0 = Full-Frequency Contour Deformation
Calculations Performed on Edison@NERSC (CRAY-XC30)

- Approximation tested for Insulators, Semiconductors, Metals, Clusters, Slabs
- Direct correlation between accuracy and $t_{\rm eigen}$ (5-25% of the eigenstates $\rightarrow\sim$ 1 meV accuracy)

`Sigma` code: Calculate $\Sigma_{lm}(\omega)$ and Solve Dyson's Equation

1. Calculate plane-waves matrix elements (FFT's): $O(N_\Sigma N_n N_G \log N_G)$

$$M^{-G}_{nm} = \langle\phi_n|\, e^{-i\mathbf{G}\cdot\mathbf{r}}\,|\phi_m\rangle$$

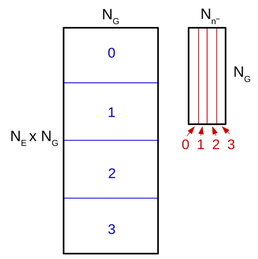
2. For any given pair of orbital functions $\{\phi_l, \phi_m\}$ calculate:

$$\Sigma_{lm}(E) = \frac{i}{2\pi}\int_0^\infty d\omega \sum_n \sum_{GG'} M^{-G}_{nl} \frac{\epsilon^{-1}_{GG'}(\omega)\cdot v(G')}{E - E_n - \omega} M^{-G'}_{nm}$$

Matrix-Multiplication + 2D Dot-Product $O(N_\Sigma N_\omega N_n N_G^2)$

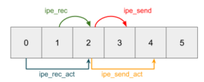
Parallelization Strategy

$\Sigma_{lm}(\omega)$ matrix elements distributed over Pools of processes. For each Pool:



Data distribution layout
$N_G \simeq 10^5$; $N_n \simeq 10^4$; $N_E \simeq 10^2$

- Left Matrix: $\epsilon^{-1}(\omega)$ distributed over rows ($N_G \times N_\omega$ combined index)
- Right Matrix: M^G_{nl} distributed over columns (FFT performed locally)
- At each cycle the matrix contraction is performed locally (ZGEMM)
- Ring communication restricted to contiguous MPI tasks



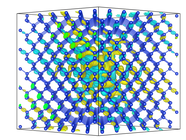
Non-blocking cyclic communication layout

- Communication can be performed over blocks of processes:
  - $N_{n''}$ size roughly constant independent on the ration OMP×MPI
  - Good ZGEMM performance independent on the number of MPI tasks employed

Same algorithm is used in the low-rank approximation case: matrix size $N_G \rightarrow N_b$

- Speed-up proportional to $(N_G/N_b)^2$
- Good performance achieved by using smaller pool sizes

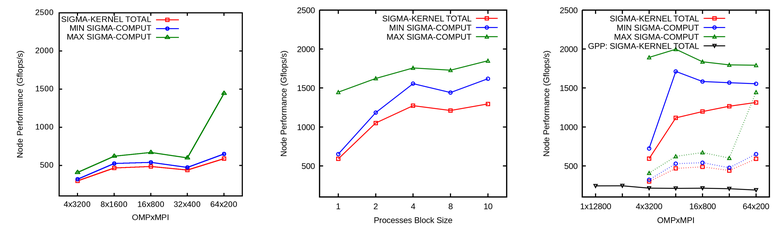
Performance Measurement on Cori-KNL@NERSC



Systems: Divacancy states in Silicon supercells containing 998 and 1726 atoms.

- FLOPs per node
- Best/worst performing node
- Strong Scaling
- Time to solution
- Comparison to peak performance

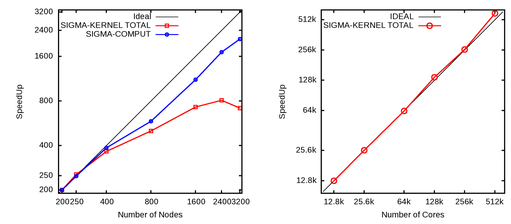
Single Pool Performance: 200 KNL Nodes



(a) (b) (c)

(a) Poor performance due to small Comput./Commun. ratio (different OMPxMPI ratio, no message blocking) (b) Improved performance for 64-OMPxMPI by using different processes block size (c) Different OMPxMPI ratio and process block size adjusted to give roughly constant n''.

Strong Scaling: Individual Pool and Full `Sigma`



(d) (e)

(a) Individual Pool scaling, total execution (red squares) and computationally intense part (blue circles), (b) Full `Sigma`, 200-NKL nodes per Pool.

Full `Sigma`: Best Performance

	998 Si	1726 Si
Number KNL Nodes	9600	9500
Number of Cores	633,600	627,000
Number $E_{\rm qp}$ Evaluated	48	38
Time to solution (s)	160	201
PetaFLOP/s	11.8	11.3
% Peak Performance	47	46

- `Sigma` kernel capable to scale to full-Cori
- `Sigma` achieves high fraction of peak performance
- Excellent time to solution (∼100 seconds) for systems made of thousands of atoms

Acknowledgments

This work was supported by the Center for Computational Study of Excited-State Phenomena in Energy Materials at the Lawrence Berkeley National Laboratory, which is funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division under Contract No. DE-AC02-05CH11231, as part of the Computational Materials Sciences Program.

mdelben@lbl.gov