Speeding-up GW Calculations for Large-scale Quasiparticle Predictions

Peihong Zhang
Department of Physics, University at Buffalo, SUNY
The DFT band gap “problem”

- The DFT bandgap “problem”

<table>
<thead>
<tr>
<th>Material</th>
<th>LDA (eV)</th>
<th>Exp (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>diamond</td>
<td>3.9</td>
<td>5.48</td>
</tr>
<tr>
<td>Silicon</td>
<td>0.52</td>
<td>1.17</td>
</tr>
<tr>
<td>Germanium</td>
<td>~0</td>
<td>0.74</td>
</tr>
<tr>
<td>LiF</td>
<td>6.0</td>
<td>9.4</td>
</tr>
</tbody>
</table>

KS orbital energies cannot simply be interpreted as quasiparticle energies.
A successful approach for understanding the excited states of many-electron systems is based on the quasiparticle (QP) concept and the Green function method.

The QP equation:

\[ H^0(r)\psi_{nk}^{QP}(r) + \int \Sigma(r, r', E_{nk}^{QP})\psi_{nk}^{QP}(r')dr' = E_{nk}^{QP}\psi_{nk}^{QP}(r) \]

\[ H^0 = T + V^H + V^{ext} \]

- The self-energy contains all effects of e-e interactions.

The difficulty of quasiparticle calculations is that the self-energy is non-Hermitian, nonlocal and energy dependent.
The GW approximation to the electron self-energy

\[ E_{nk}^{QP} = \varepsilon_{nk}^{KS} + \langle n\vec{k} | \Sigma - V_{xc} | n\vec{k} \rangle \]

\[ \Sigma^{GW} = iGW \]

\( G \): electron Green function

\( W \): screened Coulomb interaction

<table>
<thead>
<tr>
<th>Material</th>
<th>LDA (eV)</th>
<th>GWA\textsuperscript{(a)} (eV)</th>
<th>Exp (eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>diamond</td>
<td>3.9</td>
<td>5.6</td>
<td>5.48</td>
</tr>
<tr>
<td>Silicon</td>
<td>0.52</td>
<td>1.29</td>
<td>1.17</td>
</tr>
<tr>
<td>Germanium</td>
<td>(~0)</td>
<td>0.75</td>
<td>0.74</td>
</tr>
<tr>
<td>LiF</td>
<td>6.0</td>
<td>9.1</td>
<td>9.4</td>
</tr>
</tbody>
</table>

Typical accuracy of ab initio G\textsuperscript{0}W\textsuperscript{0} methods: \( \sim 0.1 – 0.2 \) eV

Hedin, 1965; Hybertsen and Louie, 1986
First-principles GW method

KS mean-field calculation → Dielectric function

One particle Green function G → Self-energy

\[ \Sigma = i \int \frac{dE'}{2\pi} e^{-i\delta E'} G(r, r'; E - E') W(r, r'; E') \]

RPA

\[ \varepsilon = 1 - \nu \chi \]

\[ W = \varepsilon^{-1} \nu \]

Hybertsen and Louie, 1986
Not all materials are “GW friendly”

- The band gap of InN was found to be 0.7 ~ 0.8 eV in 2002! By experimentalists!!! Previously “accepted” value was 1.9 eV
  - What have theorists been doing???

- ZnO: LDA band gap: 0.7 eV
  - Earlier GW calculations by several groups: 1.5 ~ 2.5 eV
  - Experimental gap: ~ 3.6 eV

- CuCl: LDA band gap: 0.32 eV
  - “Straightforward” GW calculations give Eg < 2.0 eV
  - Experiment: 3.4 eV

- GWA does not work for these materials?

- The devil is in the detail: Convergence behavior of several truncation parameters has to be carefully examined!
The dielectric function

- The dielectric function: \( \varepsilon = 1 - \nu \chi^0 \)

\[
\chi_{G,G'}^{0}(q, \omega) = \frac{1}{2} \sum_{v,v,k} M_{vc}(k, q, G) M^*(k, q, G') \times \\
\left( \frac{1}{E_{vk+q} - E_{ck} - \omega + i\delta} + \frac{1}{E_{vk+q} - E_{ck} + \omega + i\delta} \right)
\]

- Two truncation parameters in practical calculations
  - Number of conduction bands (Nc) included in the summation
  - Kinetic energy cutoff (G_{cut}) for the dielectric matrix \( \varepsilon_{GG'}(q, \omega) \)

- In practical calculations, truncations are almost always applied
  - Nc number of conduction bands needed to converge the GW results scales with the system volume
  - Nc and anc G_{cut} are both highly materials dependent
The Coulomb-hole self-energy

- The self-energy is usually decomposed into two terms

\[ \Sigma(r, r'; E) = \Sigma_{\text{SEX}}(r, r'; E) + \Sigma_{\text{COH}}(r, r'; E) \]

where \( \Sigma_{\text{SEX}} \): screened exchange; \( \Sigma_{\text{COH}} \): Coulomb hole

\[ \langle \Sigma_{\text{COH}} \rangle_{nk} = \frac{1}{2} \sum_{m} \left[ \sum_{\bar{q}, \bar{G}, \bar{G'}} [M_{\bar{G}}^{n,m}(\bar{k}, \bar{q})]^* M_{\bar{G}}^{n,m}(\bar{k}, \bar{q}) \right] \Omega_{\bar{G}\bar{G}',(\bar{q})}^2 \times \frac{[\tilde{\omega}_{\bar{G}\bar{G}',(\bar{q})}[E - \epsilon_{m,\bar{k} - \bar{q}} - \tilde{\omega}_{\bar{G}\bar{G}',(\bar{q})}]}{[\tilde{\omega}_{\bar{G}\bar{G}',(\bar{q})}]^2} \nu(\bar{q} + \bar{G}') \]

\[ M_{\bar{G}}^{nm}(k, q) = \langle mk - q| e^{-i(q + G) \cdot r} | nk \rangle \]

- The summation in principle should include all conduction bands
Convergence issues in GW calculations

• Two important cutoff parameters:
  - Band summation
  - Kinetic energy cutoff of the dielectric matrix
    \[
    \begin{align*}
    \varepsilon_{GG'}(\vec{q}) \\
    E_{cut} = |G_{cut}|^2 / 2
    \end{align*}
    \]

• The convergence issue in GW calculations is well recognized but often “ignored” (sometimes intentionally)
  - Hard to perform convergence tests for all GW calculations
  - For some materials (e.g., Si, Ge), GW results may converge quickly
  - For others (e.g. ZnO, MgO, CuCl), unconverged results may lead to false predictions and/or wrong interpretations
Quasiparticle Gap of ZnO (LDA+U/GW)

A small kinetic energy cutoff for the dielectric matrix leads to a false convergence behavior.

Previous result

\[ E_{\text{cut}} = |G_{\text{max}}|^2 / 2 \]

\[ \epsilon_{GG'}(\vec{q}) \]

\[ W_{GG'}(q, \omega) = \epsilon_{GG'}^{-1}(q, \omega) v(q + G') \]
Quasiparticle gap of MgO

Previous results

Quasiparticle gap of CuCl

A few hundred bands

Dielectric matrix cutoff: 10 Ry

10,000 bands for a two-atom system!?
Why are GW calculations for some materials more difficult to converge than others

VBM: Very localized O 2p
CBM: Delocalized Zn 4s

ZnO (LDA+U)

E (eV)

VBM: Very localized O 2p
CBM: Delocalized Zn 4s

Coulomb Hole Energy (eV)
How about CuCl?

- VBM: Extremely localized Cu 3d; CBM: Very delocalized Cu 4s
- LDA severely underestimates the Cu 3d levels

LDA gap: 0.32 eV
LDA+U is a better starting point for subsequent GW calculation

- LDA+U gives a better starting point for GW many-body treatment

LDA+U gap: 1.48 eV
GW calculations for large systems? High-throughput GW calculations?

- (Forget about ZnO or CuCl) For a 2-atom MgO, we need about 1,000 conduction bands to converge the result.

- Suppose we are interested in a system containing 200 atoms (e.g., a supercell containing a defect), we will need

\[1000 \times (200/2) = 100,000 \text{ conduction bands}\]

- To achieve the same level of convergence.

- Not only calculating the wave functions is hard (if possible), but storing these wave functions is extremely problematic, not to mention the subsequent GW calculations.
Speed up GW calculation for large systems?

Do we really need an explicit band-by-band summation?

DOS of MgO

Speed up GW calculation for large systems

- Contributions from high energy states are calculated by an energy integration:

\[ \chi_{G,G'}^0(q, \omega = 0) \approx \sum_{c} \left[ \sum_{vk} \sum \frac{M_{vc}(k, q, G) M_{vc}^*(k, q, G')}{E_{vk+q} - E_{ck}} \right] \]

\[ + \int_{E_0} \sum_{vk} \frac{M_{vE}(k, q, G) M_{vE}^*(k, q, G')}{E_{vk+q} - E'} g(E') dE' \]

Low-energy states

High-energy states

\[ g(E) = \frac{\Omega}{\pi^2} \sqrt{2(E - V_{xc}(0))} : \text{free-electron-like DOS} \]
Accuracy/performance of the new method

QP band structure of MgO

QP band gap of ZnO

Dielectric matrix cutoff: 60 Ry
### Sensitivity to the integration energy grid

<table>
<thead>
<tr>
<th>$\Delta E$ (eV)</th>
<th>New method</th>
<th>Conventional method</th>
<th>Speed-up factor</th>
<th>$\Delta E_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_0 + N_E$</td>
<td>$E_g$</td>
<td>$N_c$</td>
<td>$E_g$</td>
</tr>
<tr>
<td>1.5</td>
<td>895</td>
<td>7.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>640</td>
<td>7.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>510</td>
<td>7.88</td>
<td>8,000</td>
<td>7.86</td>
</tr>
<tr>
<td>3.0</td>
<td>435</td>
<td>7.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.5</td>
<td>385</td>
<td>7.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>320</td>
<td>7.84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

QP band structure of MgO  
(A 16-atom supercell)
Quasiparticle gap of CuCl

Speedup factor for small systems: 10 times

Conventional method

new method
### Large scale GW calculations

- **Band gap of MgO supercells**

<table>
<thead>
<tr>
<th># of atoms</th>
<th>New method</th>
<th>Conventional method</th>
<th>Speed-up factor</th>
<th>( \Delta E_g )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( N_0 + N_E )</td>
<td>( E_g )</td>
<td>( N_c )</td>
<td>( E_g )</td>
</tr>
<tr>
<td>2</td>
<td>170</td>
<td>7.86</td>
<td>1,000</td>
<td>7.86</td>
</tr>
<tr>
<td>16</td>
<td>320</td>
<td>7.84</td>
<td>8,000</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>920</td>
<td>7.89</td>
<td>32,000</td>
<td>7.86</td>
</tr>
<tr>
<td>128</td>
<td>1060</td>
<td>7.83</td>
<td>64,000</td>
<td></td>
</tr>
<tr>
<td>256</td>
<td>1580</td>
<td>7.86</td>
<td>128,000</td>
<td></td>
</tr>
</tbody>
</table>

- A speed-up factor of nearly two orders of magnitude is achieved
- Numerical error: less than \( \pm 0.05 \) eV

Hybrid organic-inorganic halide perovskite solar cell

- 2009: 3.8% (CH$_3$NH$_3$PbI$_3$), Kojima et al, JACS 131, 6050 (2009)
- 2011: 6.5%, Park et al, Nanoscale 3, 4088 (2011)
- 2015: 21.0%, EPFL
- 2016: 22.1%, KRICT and UNIST

Fastest-advancing solar cell technology to date in terms of conversion efficiency!
Hybrid organic-inorganic perovskites

- Several phases
- Random orientation of organic molecules
## Quasiparticle band gap of MAPbI₃

<table>
<thead>
<tr>
<th>Phase</th>
<th>Structure</th>
<th>( E_{g}^{GW+SOC} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cubic (primitive)</td>
<td>PbI(_3^-)</td>
<td>ideal struct.</td>
</tr>
<tr>
<td></td>
<td>MAPbI(_3)</td>
<td>MA [001]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA [011]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA [111]</td>
</tr>
<tr>
<td>Tetragonal (primitive)</td>
<td>MAPbI(_3)</td>
<td>MA [001]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA [011]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MA [111]</td>
</tr>
<tr>
<td>C &amp; T (supercell)</td>
<td>MAPbI(_3)</td>
<td>with MA directions randomized</td>
</tr>
<tr>
<td>Experiment</td>
<td>MAPbI(_3)</td>
<td>1.51∼1.66</td>
</tr>
</tbody>
</table>

PRB 93, 085202 (2016)
2D materials

- Reported GW band gap: 2.41 ~ 2.84 eV

- Methods: G\textsuperscript{0}W\textsuperscript{0}, G\textsuperscript{1}W\textsuperscript{0}, self-consistent GW, etc

- Parameters used:
  Number of conduction bands: 96 to 9000
  k-point sampling: 6x6x1 to 24x24x1

- Our method can be applied to 2D materials as well

PRB 85, 205302 (2012); PRB 86, 115409 (2012);
PRB 87, 155304 (2013); PRB 88, 045412 (2013);
PRL 115, 119901 (2013)
2D materials

Fully converged GW calculations for 2D materials with a speed-up factor of about 20 ~ 30
Conclusions

• Despite the tremendous success, accurate and efficient predictions of excited-states properties of complex solid systems remain a major challenge

• We have developed a powerful and simple-to-implement method that can drastically accelerate fully converged GW calculations for large systems, enabling
  - Fast and accurate GW calculations for complex materials

• Speed-up factor:
  Small systems (< 10 atoms): 10 ~ 20 times
  Medium-sized systems (10 ~ 50 atoms): 20 ~ 40 times
  Large systems (50 ~ 500 atoms): 40 ~ 100 times
Acknowledgement

- Collaborators:
  - Weiwei Gao, Tesfaye Abtew, Weiyi Xia (Buffalo)
  - Yiyang Sun, Shengbai Zhang (RPI)
  - Xiang Gao (Beijing CSRC)

- Funding

- Computing:

ICQMS @ SHU