5th international Abinit develeoper workshop
11 October – 14 April, 2011
Han-sur-Lesse, Belgium

Implementation of the Bethe-Salpeter formalism in Abinit.

M. Giantomassi

Université Catholique de Louvain
Louvain-la-Neuve, Belgium

ETSF
European Theoretical Spectroscopy Facility

abinit.org

UCL
Université catholique de Louvain
UNIVERSITAS CATHOLICA LOVANIENSIS SEDES SAPIENTIAE 1425

Overview

- A brief introduction to the Bethe-Salpeter formalism
- BSE in the electron-hole representation
- The GW+BSE code of Abinit
- Implementation details
- Future developments

Overview

- A brief introduction to the Bethe-Salpeter formalism
- BSE in the electron-hole representation
- The GW+BSE code of Abinit
- Implementation details
- Future developments

MBPT and excitations

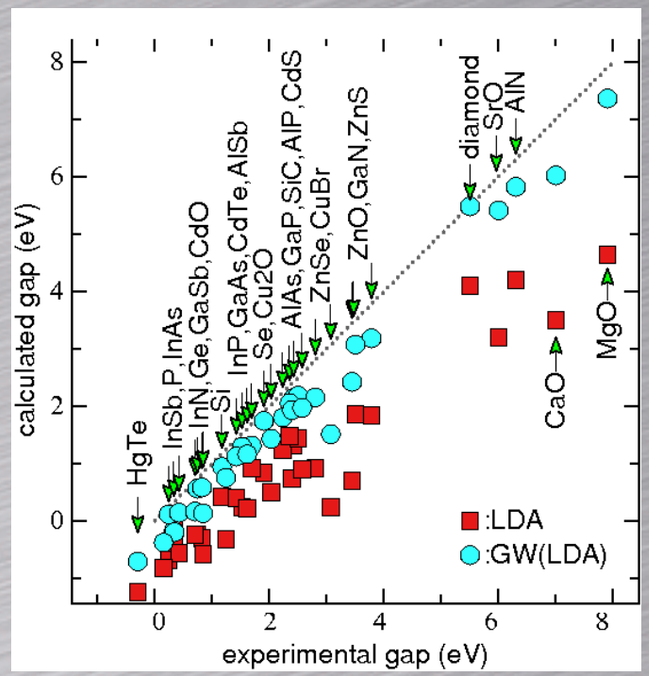
$$\left(-\frac{1}{2}\Delta + v_{\rm ext}(\mathbf{r}) + v_{Hxc}[n](\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r})$$

Lagrangian multiplier

$$\hat{h}_0(\mathbf{r}_1)\Psi_i(\mathbf{r}_1) + \int \Sigma(\mathbf{r}_1,\mathbf{r}_2;\epsilon_i)\Psi_i(\mathbf{r}_2)\,\mathrm{d}\mathbf{r}_2 = \epsilon_i\Psi_i(\mathbf{r}_1).$$

Pole of the Green function

Phys. Rev. Lett. 96, 226402 (2006)



GW gaps are in much better agreement with experiments

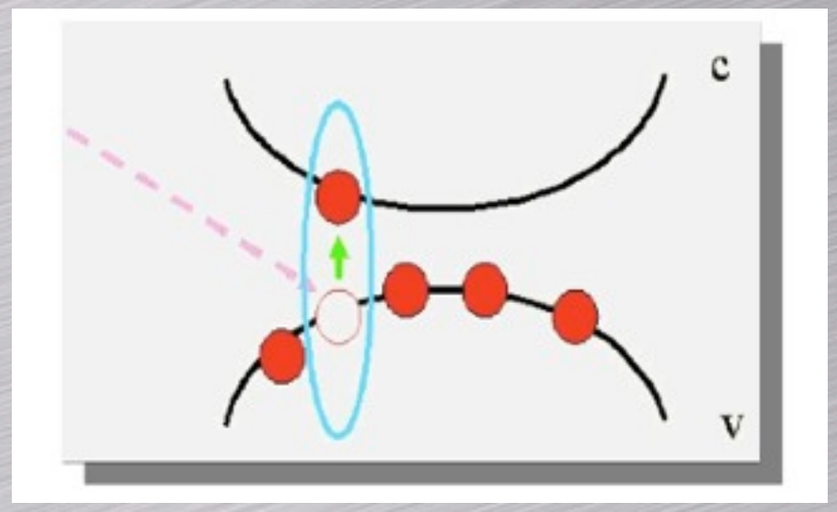
- Charged excitation energies are well described within the GW approximation for the self-energy

- Absorption spectra are directly connected to the many-body irreducible polarizability $\tilde{\chi}$

$$\epsilon_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q};\omega) = \delta_{\mathbf{G}_1\mathbf{G}_2} - v(\mathbf{q}+\mathbf{G}_1)\,\tilde{\chi}_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{q};\omega),$$

$$\epsilon_M^{\rm LF}(\omega) = \lim_{\mathbf{q}\to 0}\frac{1}{\epsilon^{-1}_{00}(\mathbf{q},\omega)}$$

Local field effects included



Neutral excitation energies are the poles of the irreducible polarizability

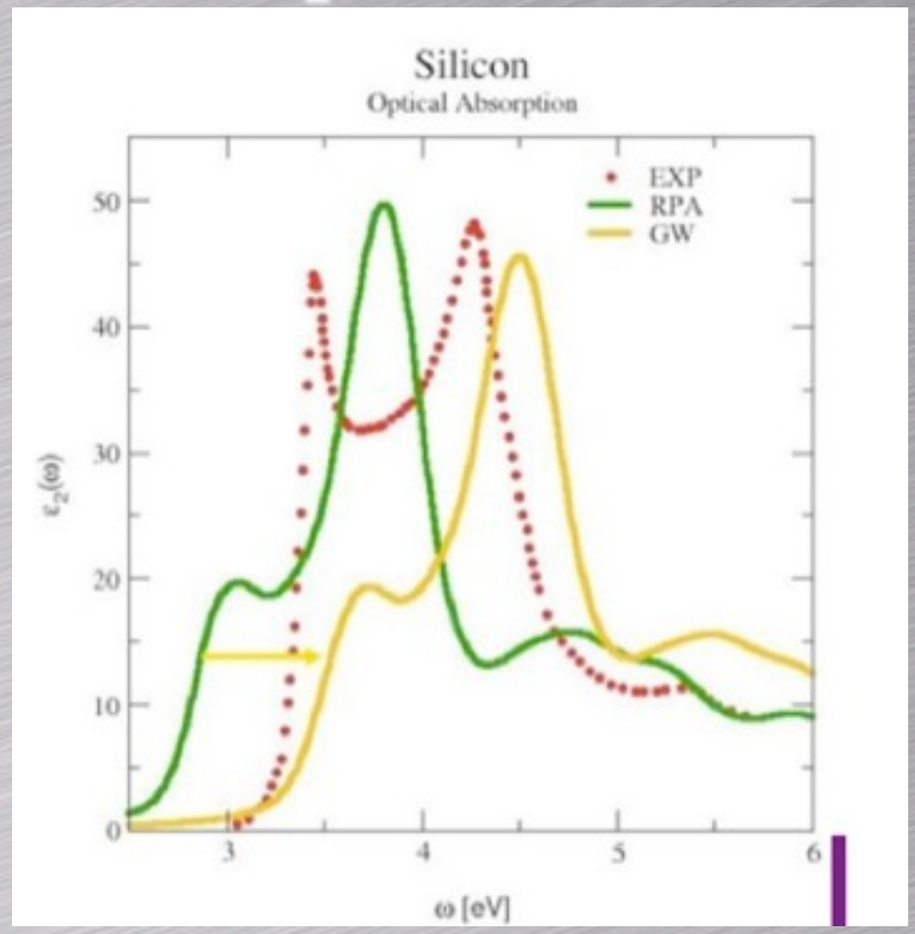
Ab-initio absorption spectra

RPA polarizabilites:

$$\chi^{\mathrm{KS}}(12) = -i\, G^{\mathrm{KS}}(12) G^{\mathrm{KS}}(21)$$

$$\chi^{\mathrm{GW}}(12) = -i\, G^{\mathrm{GW}}(12) G^{\mathrm{GW}}(21)$$

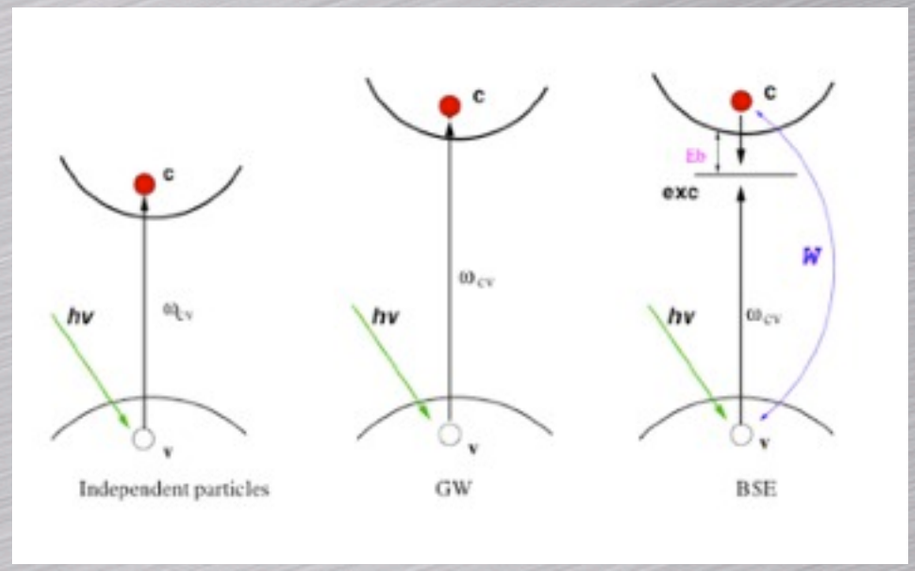
- RPA with GW corrections leads to a blue-shifted spectrum
- The first peak is missing. Important phenomena are not captured by the RPA!



The exact many-body polarizability

$$\tilde{\chi}(12) = -iG(13)\,\Gamma(34;2)\,G(41)$$

Vertex function

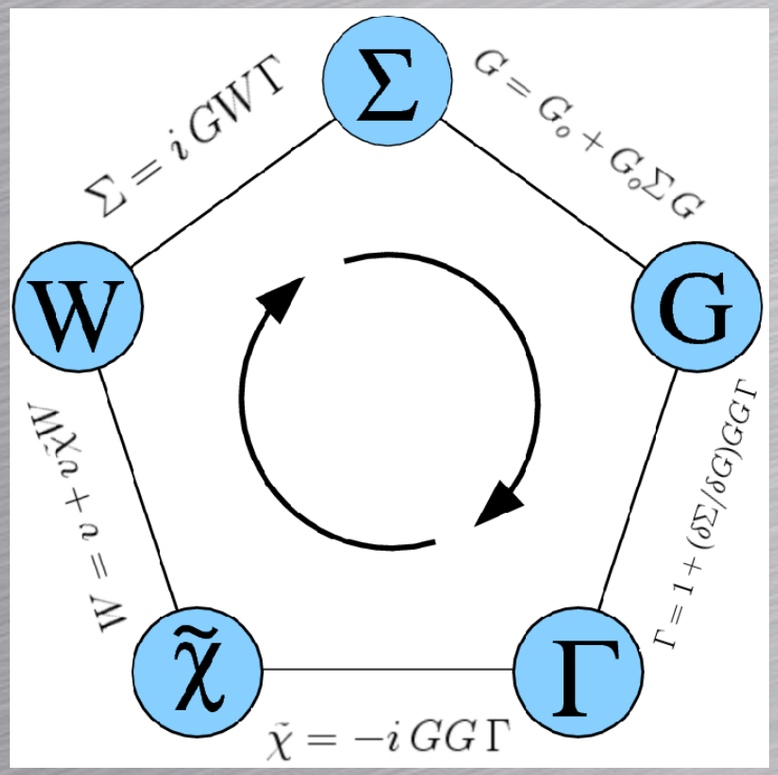


Vertex corrections are needed to describe the phenomena involved in neutral excitations

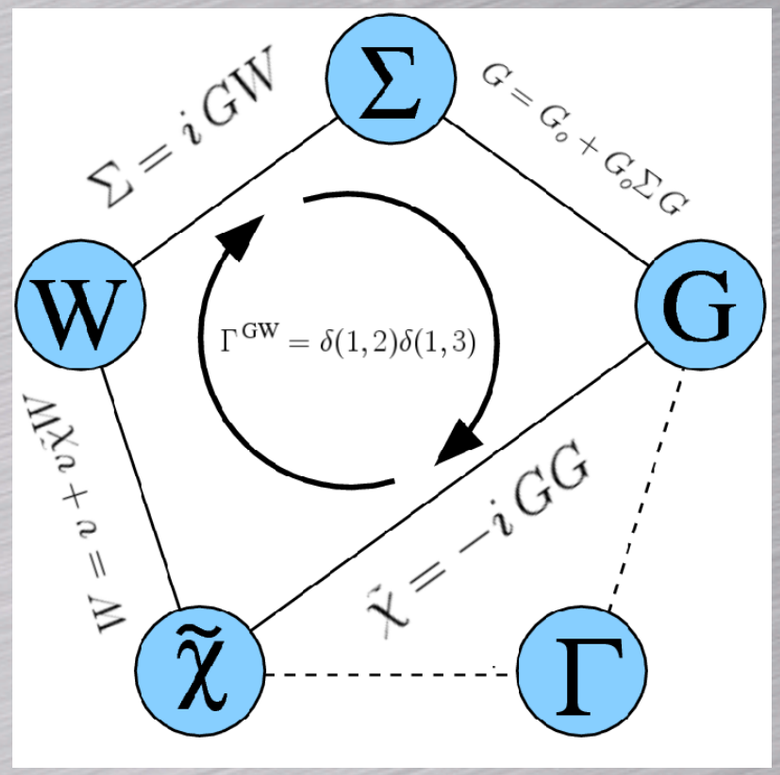
Hedin's pentagon and BSE

Phys. Rev. 139, A796–A823 (1965)

Full set of equations



The GW approximation



Beyond GW: the second iteration of the pentagon

$$\Sigma_{\mathrm{GW}}(12) = iG(12)\,W(12) \quad \Rightarrow \quad \frac{\delta\Sigma_{GW}(12)}{\delta G(34)} = i\delta(13)\delta(24)W(12) + iG\frac{\delta W}{\delta G}$$

Using $\Gamma = 1 + \frac{\delta\Sigma}{\delta G} G\,G\,\Gamma$ one obtains an equation for Γ and a new approximation for $\tilde{\chi}$

BSE in a nutshell

Rev. Mod. Phys. 74, 601–659 (2002)

The BSE must be formulated in terms of the four-point functions $L(11',22')$ and $L^0(11',22')$

$$\tilde{\chi}(12) = L(11,22) \qquad \chi^0(12) = L^0(11,22)$$

Contracting gives the many-body polarizability

Integral equation for L:

$$L = L^0 + L^0\,K\,L \implies L = [1 - L^0\,K]^{-1}\,L^0$$

Local field effects are included by using the modified kernel

Screened interaction between electron and hole

$$K(1234) = \delta(12)\delta(34)\bar{v}(13) - \delta(13)\delta(24)W(12)$$

with the modified Coulomb interaction

$$\begin{cases} \bar{v}(\mathbf{q}) = v(\mathbf{q}) \quad \text{if } \mathbf{q} \neq 0 \\ \\ \bar{v}(\mathbf{q}=0) = 0 \end{cases}$$

The inversion of $\epsilon_{\mathbf{G}_1\mathbf{G}_2}$ is thus avoided!

Overview

- A brief introduction to the Bethe-Salpeter formalism
- BSE in the electron-hole representation
- The GW+BSE code of Abinit
- Details on the implementation
- Future developments

BSE in the e-h basis set

Phys. Rev. B 62, 8, 4927 (2000)

$$L = \left[1 - L^0 K\right]^{-1} L^0$$

1. Select a finite basis set thus discretizing the equation
2. Solve the problem with matrix algebra

1) Kohn-Sham states are used to expand the four-point functions

$$F(1234) = \sum_{\substack{(n_1 n_2)\\(n_3 n_4)}} F_{(n_1 n_2)(n_3 n_4)}\, \psi^\dagger_{n_1}(1)\psi_{n_2}(2)\, \psi_{n_3}(3)\psi^\dagger_{n_4}(4) \qquad n = (b, \mathbf{k}, \sigma)$$

L^0 is **diagonal** in the KS basis set $\qquad L^0_{(n_1 n_2)(n_3 n_4)}(\omega) = \dfrac{(f_{n_2} - f_{n_1})}{(\epsilon_{n_2} - \epsilon_{n_1} - \omega)} \delta_{n_1 n_3} \delta_{n_2 n_4}$

2) After some algebra one obtains:

Two-particle Hamiltonian

$$L_{(n_1 n_2)(n_3 n_4)}(\omega) = \left[H - \omega\right]^{-1}_{(n_1 n_2)(n_3 n_4)} (f_{n_4} - f_{n_3})$$

We have assumes a static W and an energy gap $\qquad (n_1, n_2) \Rightarrow \begin{cases} (c, v, \mathbf{k}, \sigma) \\ (v, c, \mathbf{k}, \sigma) \end{cases}$

Spectra from the BSE

$$\epsilon_M(\omega) = 1 - \lim_{\mathbf{q}\to 0} v(\mathbf{q})\,\tilde{\chi}_{00}(\mathbf{q},\omega)$$

Local field effects are included via the modified kernel

Using the matrix notation in the e-h basis set

$$L = [H - \omega]^{-1} F$$

$$F = \left(\begin{array}{c|cc} & |v'c'\rangle & |c'v'\rangle \\ \hline \langle vc| & 1 & 0 \\ \langle cv| & 0 & -1 \end{array}\right)$$

and using $\tilde{\chi}(12) = L(11,22)$

the macroscopic dielectric function can be expressed as:

$$\epsilon_M(\omega) = 1 - \lim_{\mathbf{q}\to 0} v(\mathbf{q})\,\langle P(\mathbf{q})|[H - \omega]^{-1} F|P(\mathbf{q})\rangle$$

Dipole operator in the e-h representation

$$P(\mathbf{q})_{n_1 n_2} = \langle n_2|e^{i\mathbf{q}\cdot\mathbf{r}}|n_1\rangle = \delta_{n_1 n_2} + i\mathbf{q}\cdot\langle n_2|\mathbf{r}|n_1\rangle + O(q^2)$$

Selection rules for the dipole:

1. spin
2. irreducible representations

The BS Hamiltonian

Phys. Rev. B 62, 4927–4944 (2000)

In spin-unpolarized systems only singlet states contribute to the optical properties

`nsppol=1`

$$\bar{v} - W \rightarrow 2\bar{v} - W$$

$$H = \left(\begin{array}{c|cc} & |v'c'\rangle & |c'v'\rangle \\ \hline \langle vc| & R & C \\ \langle cv| & -C^* & -R^* \end{array}\right)$$

$R = R^\dagger$ Resonant block

$C = C^t$ Coupling block

$H \neq H^\dagger$ due to C

R is diagonal dominant:

Transition energies on the diagonal

$$R_{(vc)(v'c')} = (\epsilon_c - \epsilon_v)\delta_{vv'}\delta_{cc'} + K_{(vc)(v'c')}$$

In extended systems, C is smaller than R

$$C_{(vc)(c'v')} = K_{(vc)(c'v')}$$

Tamm-Dancoff approximation (TDA) neglects

$$H^{\mathrm{TDA}} = \left(\begin{array}{c|cc} & |v'c'\rangle & |c'v'\rangle \\ \hline \langle vc| & R & 0 \\ \langle cv| & 0 & -R^* \end{array}\right)$$

Spin structure of the BSE

Phys. Rev. B 77, 184408 (2008)

$$P(\mathbf{q})_{n_1 n_2} \approx_{\mathbf{q}\to 0} \delta_{n_1 n_2} + i\mathbf{q}\cdot\langle n_2|\mathbf{r}|n_1\rangle$$

Only spin-preserving transitions (violet region) contribute to $\epsilon_M(\omega)$

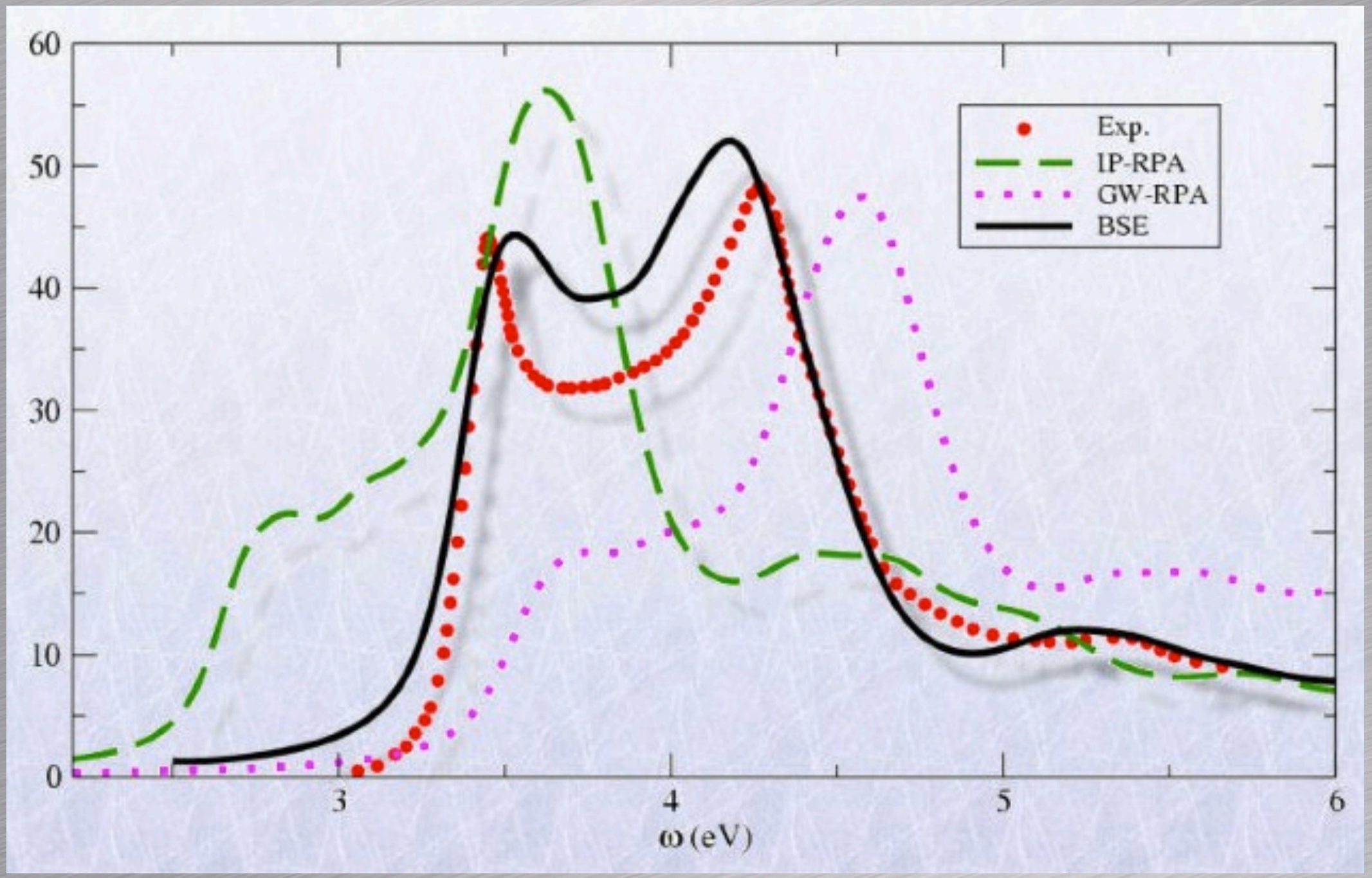
$$H = \left(\begin{array}{c|cc|cc} & |\uparrow\uparrow\rangle & |\downarrow\downarrow\rangle & |\uparrow\downarrow\rangle & |\downarrow\uparrow\rangle \\ \hline \langle\uparrow\uparrow| & T-W+\bar{v} & \bar{v} & 0 & 0 \\ \langle\downarrow\downarrow| & \bar{v} & T-W+\bar{v} & 0 & 0 \\ \hline \langle\uparrow\downarrow| & 0 & 0 & T-W & \\ \langle\downarrow\uparrow| & 0 & 0 & 0 & T-W \end{array}\right)$$

The resonant block now consists of four spin-dependent blocks:

$$R = \left(\begin{array}{c|cc} & |v'c'\uparrow\rangle & |v'c'\downarrow\rangle \\ \hline \langle vc\uparrow| & (T-W+\bar{v})^{\uparrow\uparrow} & \bar{v}^{\uparrow\downarrow} \\ \langle vc\downarrow| & \bar{v}^{\downarrow\uparrow} & (T-W+\bar{v})^{\downarrow\downarrow} \end{array}\right) \qquad \bar{v}^{\downarrow\uparrow} = (\bar{v}^{\uparrow\downarrow})^H$$

Does it work?

Rev. Mod. Phys. 74, 601–659 (2002)



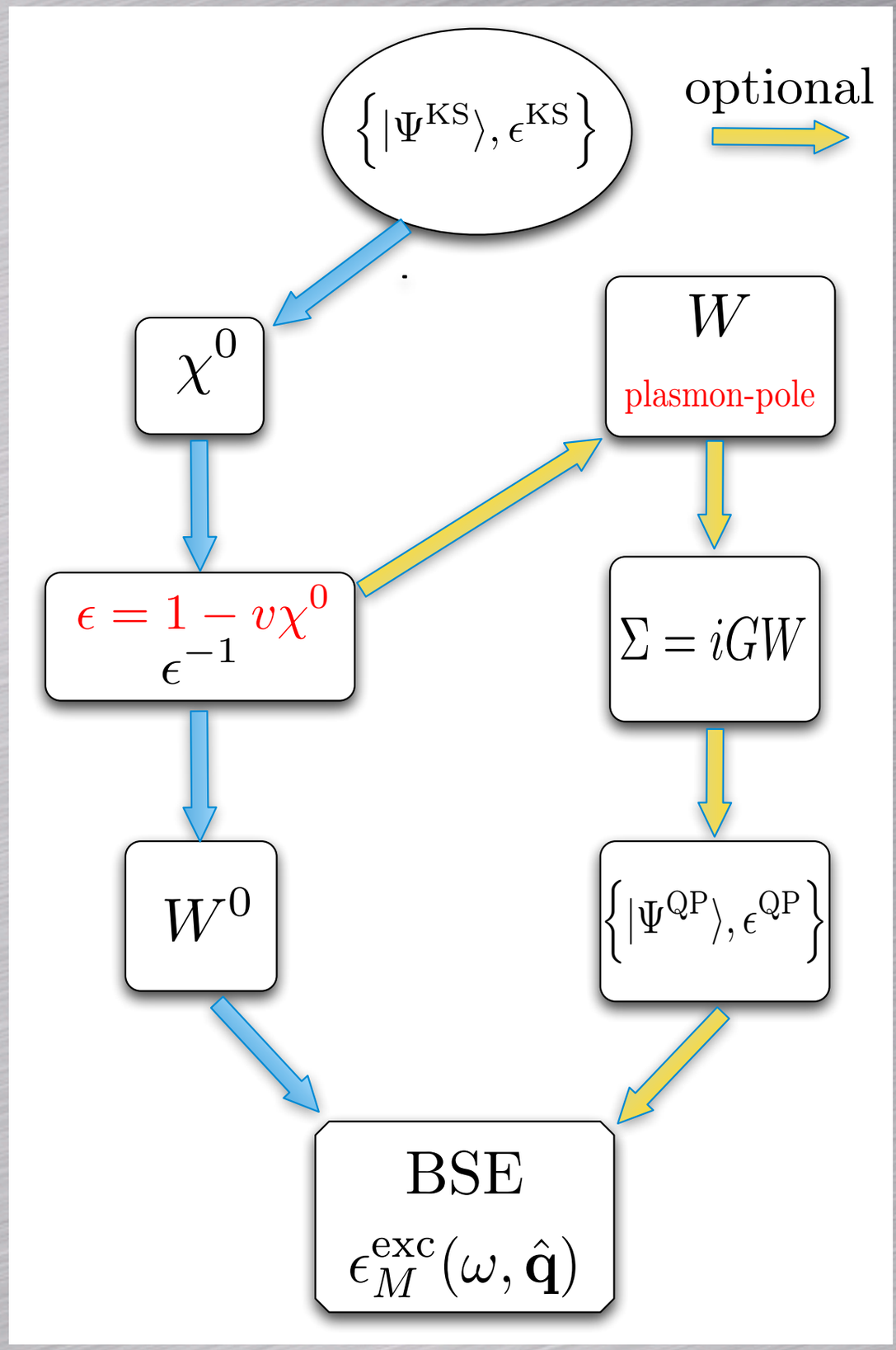
Silicon absorption spectrum

Overview

- A brief introduction to the Bethe-Salpeter formalism
- BSE in the electron-hole representation
- The GW+BSE code of Abinit
- Details of the implementation
- Future developments

GW+BSE flowchart

- Files are used to connect the different steps
- Different MPI algorithms optimized for the different run-level
- Fortran-IO or MPI-IO for reading and writing
- Big arrays are MPI-distributed with the exception of W
- Matrix-vector operations are done in parallel using a column-wise decomposition
- In-core and out-of-core solution for W
- The first NC implementation of the BSE code was based on routines from the EXC code. Many thanks to the EXC developers for sharing their code



BSE with plane waves

Exchange term

$$\bar{v}_{(vc\mathbf{k})(v'c'\mathbf{k}')} = \frac{1}{V} \sum_{\mathbf{G}\neq 0} \bar{v}(\mathbf{G}) \, \langle c\mathbf{k}|e^{i\mathbf{G}\cdot\mathbf{r}}|v\mathbf{k}\rangle\langle v'\mathbf{k}'|e^{-i\mathbf{G}\cdot\mathbf{r}}|c'\mathbf{k}'\rangle$$

`bs_exchange_term = 0` to exclude this term (no local field effects)

Coulomb term

$$W_{(vc\mathbf{k})(v'c'\mathbf{k}')} = \frac{1}{V} \sum_{\mathbf{G}_1\mathbf{G}_2} W^0_{\mathbf{G}_1\mathbf{G}_2}(\mathbf{k}'-\mathbf{k})\langle v'\mathbf{k}'|e^{i(\mathbf{q}+\mathbf{G}_1)\cdot\mathbf{r}}|v\mathbf{k}\rangle\langle c\mathbf{k}|e^{-i(\mathbf{q}+\mathbf{G}_2)\cdot\mathbf{r}}|c'\mathbf{k}'\rangle$$

The set of k-points
defines the q-mesh for W

The most CPU demanding term

`bs_coulomb_term =`
- 0 --> Diagonal approximation for W
- 1 --> Full W
- 3 --> Model dielectric function

Oscillator matrix elements

$$\langle \mathbf{k}-\mathbf{q}, b_1 | e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} | \mathbf{k}, b_2 \rangle = \hat{\mathcal{F}}\left[u_{\overline{\mathbf{k}-\mathbf{q}} b_1} \, u^{\dagger}_{\mathbf{k} b_2} \right] (\mathbf{G}-\mathbf{G}_0)$$

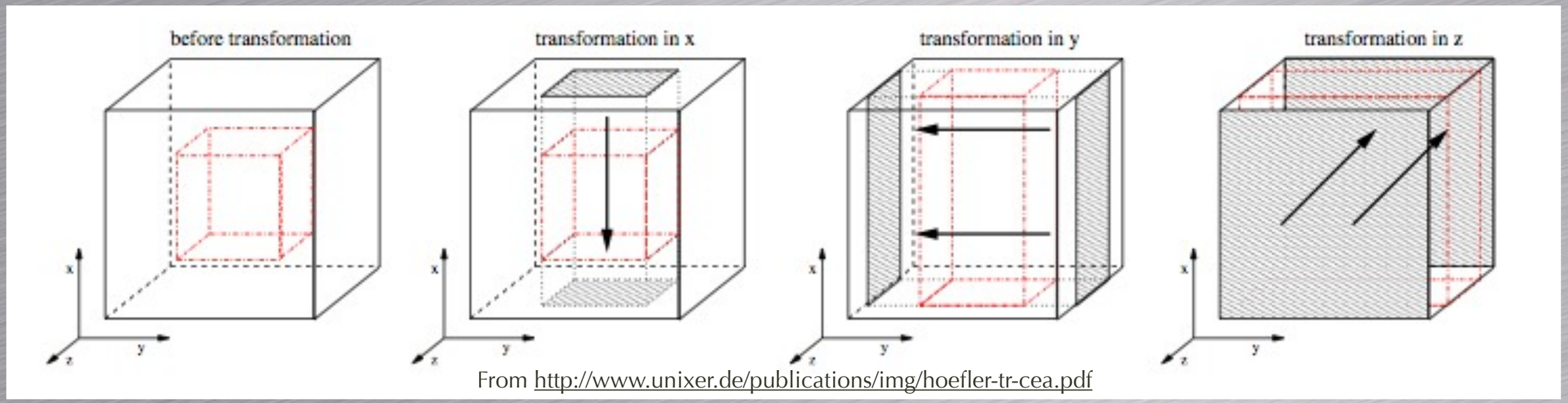
$$\mathbf{k}-\mathbf{q} = \overline{\mathbf{k}-\mathbf{q}} + \mathbf{G}_0, \quad \overline{\mathbf{k}-\mathbf{q}} \in BZ$$

`fftgw` to control the aliasing due to the convolution

Zero padded FFT leads to a significant speed-up

Support for → Goedecker FFT library (`fftalg` 112)

Support for → FFTW3, INTEL-MKL, IBM-EESL (`fftalg` 312)



From http://www.unixer.de/publications/img/hoefler-tr-cea.pdf

PAW oscillator matrix elements

$$\langle \Psi_{b_1\mathbf{k}-\mathbf{q}} | e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}} | \Psi_{b_2\mathbf{k}} \rangle = \mathrm{PW} + \sum_{ij} \langle \tilde{\Psi}_{b_1\mathbf{k}-\mathbf{q}} | \tilde{p}_i \rangle \langle \tilde{p}_j | \tilde{\Psi}_{b_2\mathbf{k}} \rangle \times$$

Precomputed using a spline fit and stored in paw_pwij_t

$$e^{-i(\mathbf{q}+\mathbf{G})\cdot\mathbf{R}_i} \left[\langle \phi_i | e^{-i(\mathbf{q}+\mathbf{G})\cdot(\mathbf{r}-\mathbf{R}_i)} | \phi_j \rangle - \langle \tilde{\phi}_i | e^{-i(\mathbf{q}+\mathbf{G})\cdot(\mathbf{r}-\mathbf{R}_i)} | \tilde{\phi}_j \rangle \right]$$

The PAW form factors needed for the spline are tabulated on a 1D-mesh (`m_paw_pwij.F90`)

$$\int_0^{r_c^a} j_l(|\mathbf{q}+\mathbf{G}|r) \left(\phi_{n_i l_i} \phi_{n_j l_j} - \tilde{\phi}_{n_i l_i} \tilde{\phi}_{n_j l_j} \right) \mathrm{d}r$$

PAW projections are symmetrized on-the-fly in `paw_symcprj.F90`

$$\langle \tilde{p}_i^a | \tilde{\Psi}_{n\mathcal{R}\mathbf{k}} \rangle = e^{i\mathbf{k}\cdot\mathbf{L}} \sum_\alpha D^{l_i}_{\alpha m_i}(\mathcal{R}^{-1}) \langle \tilde{p}^{a'}_{n_i l_i \alpha} | \tilde{\Psi}_{n\mathbf{k}} \rangle$$

$$\mathcal{R}^{-1}(\mathbf{R}^a - \mathbf{t}) = \mathbf{R}^{a'} + \mathbf{L}$$

BSE solvers

Three different solvers can be selected using `bs_algorithm`:

1. **Direct diagonalization**:
   - Lapack or ScaLapack+MPI-IO (complete or partial diago)
   - Eigenvectors, energies, DOS, oscillator strengths and excitonic amplitudes
   - Bad scaling with the size of the matrix
2. **Haydock iterative method**
   - Very efficient, excellent MPI scalability
   - Only optical spectra and an approximated DOS
3. **Iterative diagonalization with the preconditioned CG method**
   - Direct access to binding energies, DOS, wavefunctions ...
   - Efficient provided that the number of eigenvectors << N_{eh}
   - Coupling is not supported yet

BSE spectra with diagonalization

Phys. Rev. Lett. 80, 4510–4513 (1998)

$$L = [H - \omega]^{-1} F$$

The inversion for each frequency is avoided by using the spectral decomposition of H

For a non-singular operator

$$H|\lambda\rangle = \epsilon_\lambda |\lambda\rangle$$

$$O_{\lambda\lambda'} = \langle\lambda|\lambda'\rangle$$

$$H = \sum_{\lambda\lambda'} \epsilon_\lambda |\lambda\rangle O_{\lambda\lambda'} \langle\lambda'|$$

$$[H - \omega]^{-1} = \sum_{\lambda\lambda'} |\lambda\rangle \frac{O^{-1}_{\lambda\lambda'}}{(\epsilon_\lambda - \omega)} \langle\lambda'|$$

The inverse for all frequencies at the price of a single diagonalization!

- TDA allows one to use standard methods (CG or direct diago)
- Only the resonant block is needed for TDA calculations
- The inclusion of the coupling block requires a more involved treatment...

Lanczos-Haydock algorithm

Comput. Phys. Commun. 20, 11 (1980)

$\langle P|(\omega - R)^{-1}|P\rangle$ can be calculated bypassing completely the diagonalization!

Dense Hermitian matrix

$$R = R^\dagger \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{pmatrix}$$

Lanczos basis

Real symmetric tridiagonal form

$$\begin{pmatrix} a_1 & b_2 & & & \\ b_2 & a_2 & b_3 & & \\ & b_3 & * & * & \\ & & * & * & * \\ & & & * & * \end{pmatrix}$$

- Only simple matrix-vector multiplications are required
- Only three vectors are needed to construct the Lanczos basis

First step

$$b_1 = 0$$

$$|1\rangle = \frac{|P\rangle}{\||P\rangle\|}$$

First vector of the Lanczos basis

Lanczos chain

$i = 1$

$$a_i = \langle i|R|i\rangle$$

$$|i+1\rangle = \frac{R|i\rangle - a_i|i\rangle - b_{i-1}|i-1\rangle}{b_i + 1}$$

$$b_{i+1} = \|R|i\rangle - a_i|i\rangle - b_i|i-1\rangle\|$$

$i = i + 1$

Iterative solution of the BSE

Phys. Rev. B 59, 5441–5451 (1999)

$$R^k = \begin{pmatrix} a_1 & b_2 & 0 & \cdots & 0 \\ b_2 & a_2 & b_3 & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & b_{k-1} & a_{k-1} & b_k \\ 0 & \cdots & 0 & b_k & a_k \end{pmatrix}$$

Continued fraction

$$\langle P|(\omega - R)^{-1}|P\rangle = \cfrac{\|P\|^2}{\omega - a_1 - \cfrac{b_2^2}{\omega - a_2 - \cfrac{b_3^2}{\cdots}}}$$

Terminator

- The number of iterations required to converge is almost independent on the size of the matrix (~100-200)
- Easy to MPI parallelize
- Terminator helps to converge the spectrum. Assuming $\alpha_n = \alpha_\infty,\ \beta_n = \beta_\infty \ \text{for}\, n > n^0$

$$t(\omega) = \frac{1}{2\beta_\infty^2}\left\{(\omega - \alpha_\infty) - \sqrt{(\omega - \alpha_\infty)^2 - 4\beta_\infty^2}\right\}$$

- Formalism can be generalized to non-Hermitian matrices. See NanoLetters, **6**, 257, (2010)
- Eigenvalues and eigenvectors are not accessible

Overview

- A brief introduction to the Bethe-Salpeter formalism
- BSE in the electron-hole representation
- The GW+BSE code of Abinit
- Implementation details
- Future developments

Wavefunction descriptor

- A single wavefunction is represented by the Fortran datatype `wave_t`
- `wave_t` contains three buffers for $u(\mathbf{G}), u(\mathbf{r}), \langle p_i|\tilde{\Psi}\rangle$
- The wavefunction descriptor, `Wfd`, is a container storing:

    i) The array of `wave_t`: `Wfd%Wave(b,k,s)`

    ii) Internal tables for performing zero-padded FFT

    iii) G-vectors and form factors for basic operations in G-space

    iv) Tables with the MPI distribution of the states

- The internal status can be changed at run-time (e.g. the FFT mesh)
- Loops are MPI-parallelized depending on the availability of the states

Basic Methods

(A) $u(\mathbf{G}) \Rightarrow u(\mathbf{r})$

```
call wfd_get_ur(Wfd, ib, ik, spin, ur)
```

(B) $u(\mathbf{G}) \Rightarrow \langle p_i | \tilde{\Psi} \rangle$

```
call wfd_get_cprj(Wfd, ib, ik, spin, Crystal, Cp)
```

(C) FFT1 => FFT2

```
call wfd_change_ngfft(Wfd, Crystal, Psps, new_ngfft)
```

(D) $|\Psi\rangle \Rightarrow V_{\rm nl}|\Psi\rangle$
&

```
call wfd_vnlpsi(Wfd, band, ik, spin, npw, Crystal,&
                                          Psps, Hamk, vnl_psi, opaw_psi)
```

- Bands, k-points and spins are accessed using their global index
- FFT is skipped if $u(\mathbf{r})$ is already in memory
- Execution stops and dump an error file if the wave function is not available

```
subroutine wfd_mkrho(Wfd,Crystal,Psps,Kmesh,Bstr,ngfftf,nfftf,rhor)

  ! Recalculate the internal FFT tables if needed.
  call wfd_change_ngfft(Wfd,Cryst,Psps,ngfftf)

  ! Distribute the states according to their availability.
  Iter_bks = wfd_iterator_bks(Wfd, bks_mask=ABS(occ)>=tol8)

  ! Summing over (b,k,s).
  do spin=1,nsppol
    do ik=1,nkibz
      do ib_iter=1,iter_len(Iter_bks,ik,spin)
        ib = yield(Iter_bks,ib_iter,ik,spin) ! Retrieve my band index.

        call wfd_get_ur(Wfd,ib,ik,spin,ur)    u(G) ⇒ u(r)

        do ir=1,nfftf  ! Accumulate my density.
          rhor(ir,spin) = rhor(ir,spin) + &
&           occ(ib,ik,spin)*CONJG(ur(ir))*ur(ir)*wt(ik)
        end do

      end do
    end do
  end do

  ! Gather the total rhor.
  call xsum_mpi(rhor,Wfd%comm,ierr)
```

MPI parallelized!

$$n(\mathbf{r}) = \sum_{n\mathbf{k}\sigma}^{\mathrm{IBZ}} f_{n\mathbf{k}\sigma} |\Psi_{n\mathbf{k}\sigma}|^2$$

Pros and Cons

- Flexible, easy to use and to extend
- Support different levels of memory distribution
- Loops are MPI-parallelized automatically
- States can be replicated among the nodes
- Useless states can be deallocated during the run if needed
- Different instances of the same object

- Too flexible!
- Bands are not contiguous in memory, workspace arrays might be needed for particular algorithms
- The internal buffers must be declared as pointers (F90 limitation)

Overview

- A brief introduction to the Bethe-Salpeter formalism
- BSE in the electron-hole representation
- The GW+BSE code of Abinit
- Implementation details
- Future developments

Future Developments

- Inhomogeneous k-meshes
- Better control of memory
- Interpolation schemes in k-space

- Temperature effects due to e-ph coupling
- Non-collinear magnetism and spin orbit
- Beyond static W: dynamical BSE
- Generalization to finite momentum transfer