Speeding Up the GW code: Parallelism, PPM's ..

R. Shaltaf, G.-M. Rignanese and X. Gonze

Université Catholique de Louvain - Louvain-la-Neuve, Belgium

3rd International ABINIT developer workshop
22/09/2006

The code contains two main subroutines

- screening.f90 contains all the needed operations to calculate χ_0 and ϵ and hence ϵ^{-1}
- sigma.f90 we calculate the self energy

supporting routines

most of the routines that are needed by the main subroutines are included in the /src/05gw

starting from the DFT wavefunction and eigenvalues, the aim is to calculate χ_0

$$\chi^0_{\mathbf{G},\mathbf{G}'} = 2 \sum_{n,n',k} (f_{n,\mathbf{k}} - f_{n',\mathbf{k}+\mathbf{q}}) \frac{< \phi_{n',\mathbf{k}+\mathbf{q}} | e^{-i(\mathbf{q}+\mathbf{G}).\mathbf{r}} | \phi_{n,\mathbf{k}} >< \phi_{n,\mathbf{k}+\mathbf{q}} | e^{i(\mathbf{q}+\mathbf{G}).\mathbf{r}} | \phi_{n,\mathbf{k}} >}{\epsilon_{n,k} - \epsilon_{n',\mathbf{k}+\mathbf{q}} - \omega - i\delta}$$

note that the calculations on the (q=0) involve more operations than other q points

mrgscr

it is a good idea to partition the calculation using nqptdm and qptdm and then using 'mrgscr' utility to merge the screening files

starting from the DFT wavefunction and eigenvalues, the aim is to be able to calculate χ_0

$$\chi^0_{\mathbf{G},\mathbf{G}'} = 2 \sum_{n,n',k} (f_{n,\mathbf{k}} - f_{n',\mathbf{k}+\mathbf{q}}) \frac{< \phi_{n',\mathbf{k}+\mathbf{q}} | e^{-i(\mathbf{q}+\mathbf{G}).\mathbf{r}} | \phi_{n,\mathbf{k}} >< \phi_{n,\mathbf{k}+\mathbf{q}} | e^{i(\mathbf{q}+\mathbf{G}).\mathbf{r}} | \phi_{n,\mathbf{k}} >}{\epsilon_{n,k} - \epsilon_{n',\mathbf{k}+\mathbf{q}} - \omega - i\delta}$$

Paralellization possibilities

- k points
- bands
- q points
- ω

starting from the DFT wavefunction and eigenvalues, and a screening file that contains the ε^{-1} calculate the self energy Σ

$$< \phi_j|\Sigma_x|\phi_j >= -\frac{4\pi}{V}\sum_i\sum_{\mathbf{q},\mathbf{G}}\frac{\rho^2_{ij}(\mathbf{q}+\mathbf{G})}{|\mathbf{q}+\mathbf{G}|^2}$$

$$\begin{aligned} < \phi_j|\Sigma_c|\phi_j > &= -\frac{2\pi}{V}\sum_i\sum_{q,\mathbf{G}}\frac{\rho^*_{ij}(\mathbf{q}+\mathbf{G})\rho_{ij}(\mathbf{q}+\mathbf{G})}{|\mathbf{q}+\mathbf{G}||\mathbf{q}+\mathbf{G}'|} \\ &\times \frac{\Omega^2_{\mathbf{G},\mathbf{G}^{prime}}(q)}{\tilde{\omega}(\mathbf{q})-\epsilon_i+\tilde{\omega}(\mathbf{q})(2f_i-1)} \end{aligned}$$

k points parallelization

- available starting v4.6.5
- strongly localized in the code
- the number of MPI statements is very minimal

$\longrightarrow$ most of the load is done by master proc
$\longrightarrow$ the parallelism take place inside

  - cchi0,cchiq0 in case of screening
  - csigme in case of self energy

no memory saving

bands parallelization

- available starting v5.3.0
- more delocalized
- number of MPI communications increased

$\longrightarrow$ most of the load is shared by all proc,
the inversion of DM still done sequentially by master process

memory saving option is available

Crucial input variables for the parallization

- gwpara
  - case 1 k points
  - case 2 bands
  - case 3 mixed ..not yet
- mkmem
  - 0 or
  - nkpt
- localrdwf
  - 0
  - 1

```
wf(nr,min_band_per_proc:max_band_per_proc,nkpt
oc(nkpt,min_band_per_proc:max_band_per_proc)
en(nkpt,min_band_per_proc:max_band_per_proc)

if(gwpara=1.or.mkmem==0)
min_band_per_proc=1
max_band_per_proc=nbands
end if

if(localrdwf)then
read data
else
if(me==0)then
read data
xcast_mpi(data)
end if
```

- master reads KS and casts different blocks of wavefunctions, bands, occ among proc's
- each proc performs FFT on its set of wavefunctions
- each proc takes part of calculations of different quantities (some communications can take place here)
- master casts valence bands to each proc in case of screening, or the bands at which we calculate corrections in case of self energy
- master collects the results from different proc's

MPI communications

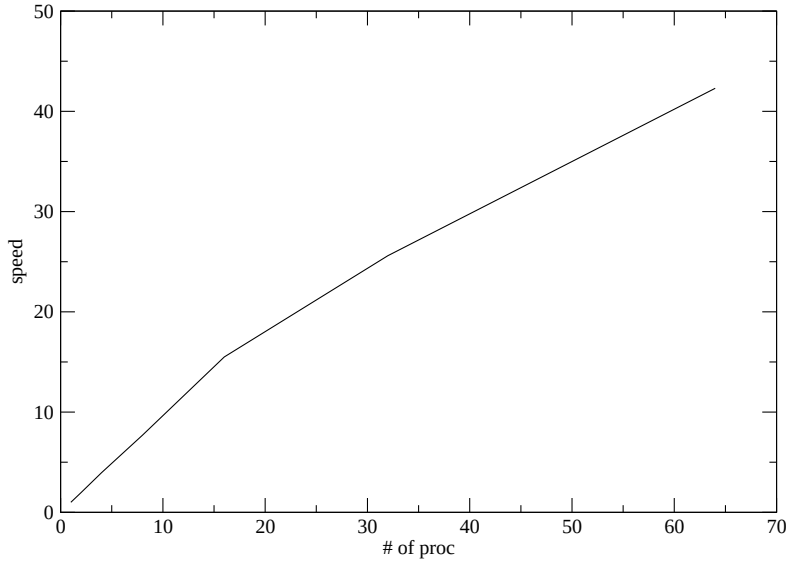
bands

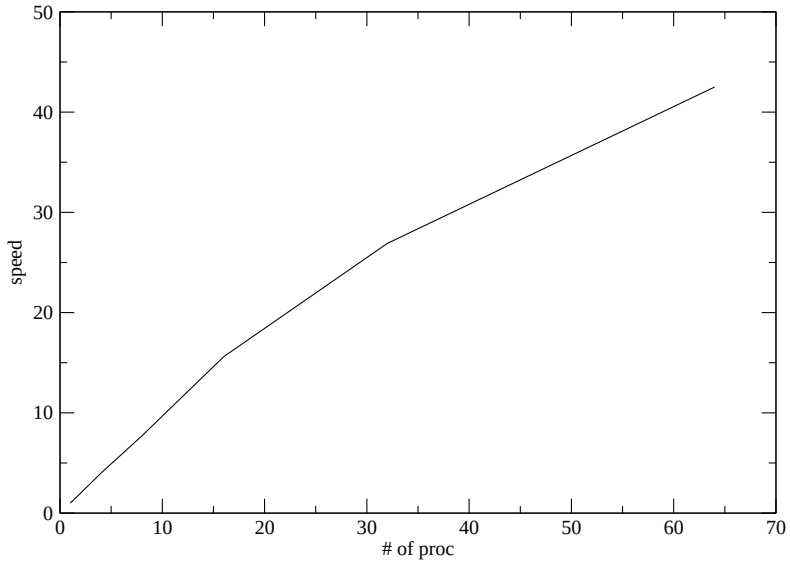
- testlda and rdkss(localrdwf ==0)
- fftwfn (mkmem==0)
- density and fermi
- in chi0 and chi0q0

k points

- testlda and rdkss(localrdwf ==0)
- non
- non
- in chi0 and chi0q0

obviously the number of communications for band parallelization is much larger


speed
# of proc
0
10
20
30
40
50
60
70


speed
# of proc
0
10
20
30
40
50
60
70

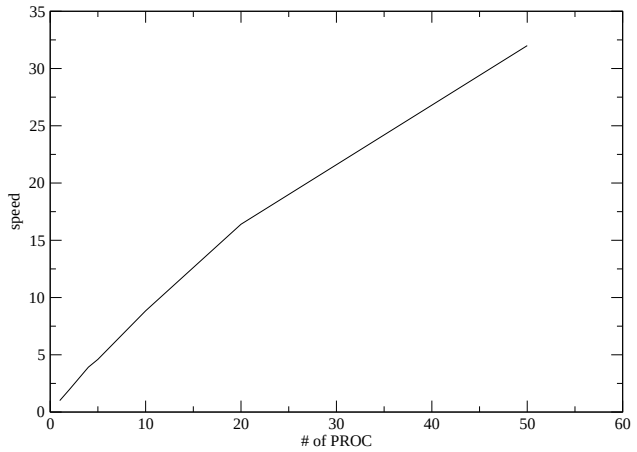
!

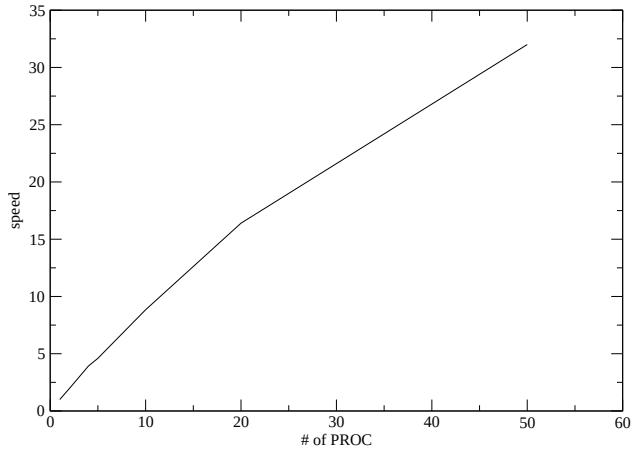
Figure: vise

Speeding Up the GW code: Parallelism, PPM's ..

Shaltaf

Technical details





Plasmon pole models

- the dynamic dependence of the screening is extracted from inverse DM calculated on a very limited number of frequences 1 or 2 the imaginary part of the ϵ^{-1} is approximated by a single plasmon pole at each G, G' vectors

$$\epsilon^{-1} = \delta_{G,G'} - \frac{\Omega}{\omega - \epsilon} \tag{1}$$

- in PPM of Godby and Needs, the dielectric matrix is fiited along the imaginary axis

$$< \phi_j | \Sigma_x | \phi_j >= -\frac{4\pi}{V} \sum_i \sum_{\mathbf{q},\mathbf{G}} \frac{\rho^2_{ij}(\mathbf{q}+\mathbf{G})}{|\mathbf{q}+\mathbf{G}|^2}$$

$$\begin{aligned} < \phi_j | \Sigma_c | \phi_j > &= -\frac{2\pi}{V} \sum_i \sum_{q,\mathbf{G}} \frac{\rho^*_{ij}(\mathbf{q}+\mathbf{G})\rho_{ij}(\mathbf{q}+\mathbf{G})}{|\mathbf{q}+\mathbf{G}||\mathbf{q}+\mathbf{G}'|} \\ &\times \frac{\Omega^2_{\mathbf{G},\mathbf{G}^{prime}}(q)}{\tilde{\omega}(\mathbf{q}) - \epsilon_i + \tilde{\omega}(\mathbf{q})(2f_i - 1)} \end{aligned}$$

Farid Engel

starting from an approximate full polarizability

$$\tilde{\chi}(\omega, q) = [\omega^2 L(q)^{-1} + \chi(0, q)^{-1}]^{-1}$$

$$L = (\mathbf{q} + \mathbf{G})(\mathbf{q} + \mathbf{G}')\rho_0(\mathbf{G} - \mathbf{G}') = -\frac{2}{\pi}\chi^1_{\mathbf{G},\mathbf{G}'}(\mathbf{q})$$

to obtain the PPM we solve the generalized eigenvalue problem

$$\chi(0, q)\mathbf{x}_{pq} = -\frac{1}{\omega_p^2(q)} L(q)\mathbf{x}_{pq}$$

with plasmon pole frequencies ω_p, and defining scaled plasmon pole vectors

$$\Theta(\mathbf{G}) = \frac{1}{\sqrt{2\omega_p}} \frac{4\pi}{|\mathbf{q} + \mathbf{G}|^2} y_{p,\mathbf{q}}(\mathbf{G}); y_{p,\mathbf{q}} = L(\mathbf{q})\mathbf{x}_{p\mathbf{q}}$$

the self energy can be written as

$$< i|\Sigma|j >= \frac{1}{V} \sum_{i,p,q} \frac{1}{\omega - \epsilon_i + (2f_i - 1)\omega_p} |\sum_{\mathbf{G}} \rho^*_{ij}\Theta_{p,\mathbf{q}}|^2$$

advantages of E-F ppmodel

- the number of plasmon pole frequencies reduced from N^2 to N and thus the memory required by them
- the memory required by ε^{-1} reduced to half as we need only ε^{-1} evaluated at one frequency
- they enable faster evaluation of self energy

- Nw GN EF
- 27 15.6 12.1
- 51 35.2 16.1
- 65 54.6 19.5
- 113 141.5 38.6
- 281 282 67