



MAX CoE/ENCCS Workshop on efficient materials modelling on HPC with Quantum ESPRESSO, Siesta, and Yambo

### Hands-on session – Day 2

# PHONONS FOR HPC AND GPUs

Laura Bellentani and Iurii Timrov







iurii.timrov@psi.ch















### INTERATOMIC FORCE CONSTANTS

Let us consider a unit cell with Nat atoms:

$$s = 1...N_{at}$$

index of an atom in the unit cell

$$\alpha = \mathbf{X}, \mathbf{Y}, \mathbf{Z}$$

is the cartesian index

R

is the point in the Bravais lattice, identifying the position of a given cell

 $N_{\mathbf{R}}$ 

is the number of unit cells in the crystal



is the  $\alpha$  component of the displacement of the s-th atom

Matrix of Interatomic Force Constants:

$$C_{slpha,s'eta}(\mathsf{R},\mathsf{R}') = C_{slpha,s'eta}(\mathsf{R}-\mathsf{R}') = rac{\partial^2 E_{tot}}{\partial \mathsf{u}_{slpha}(\mathsf{R})\partial \mathsf{u}_{s'eta}(\mathsf{R}')}$$

# **SECULAR EQUATION**

Normal mode frequencies,  $\,\omega\,$  , and eigenvectors,  $\, ilde{{f u}}_{slpha}\,$  are determined by the secular equation:

$$\sum_{\mathbf{s}',\beta} \tilde{D}_{\mathbf{s}\alpha,\mathbf{s}'\beta}(\mathbf{q}) \, \tilde{\mathbf{u}}_{\mathbf{s}'\beta}(\mathbf{q}) = \omega_{\mathbf{q}}^2 \, \tilde{\mathbf{u}}_{\mathbf{s}\alpha}(\mathbf{q})$$

Interatomic Force Constants (IFC)

where

$$\tilde{D}_{s\alpha,s'\beta}(\mathbf{q}) = \frac{1}{\sqrt{M_s M_{s'}}} \sum_{\mathbf{R},\mathbf{R}'} \frac{\partial^2 E_{tot}}{\partial \mathbf{u}_{s\alpha}(\mathbf{R}) \partial \mathbf{u}_{s'\beta}(\mathbf{R}')} e^{i\mathbf{q}(\mathbf{R}'-\mathbf{R})}$$

is the **dynamical matrix**.

Diagonalization of the dynamical matrix gives phonon modes at q.

### DENSITY FUNCTIONAL PERTURBATION THEORY

Sternheimer equation (solve\_linter):

$$(H_{SCF}^{\mathbf{k}+\mathbf{q}} + \alpha P_v^{\mathbf{k}+\mathbf{q}} - \boldsymbol{\epsilon}_v^{\mathbf{k}}) |\Delta \psi_v^{\mathbf{k}+\mathbf{q}}\rangle = -P_c^{\mathbf{k}+\mathbf{q}} \Delta v_{SCF}^{\mathbf{q}}(\mathbf{r}) |\psi_v^{\mathbf{k}}\rangle$$

$$\uparrow \qquad \qquad \uparrow \qquad \qquad \uparrow$$

$$\mathsf{h\_psi} \qquad \qquad \mathsf{orthogonalize} \qquad \mathsf{apply\_dpot}$$

$$\Delta v_{SCF}^{\mathbf{q}}(\mathbf{r}) = \Delta v^{\mathbf{q}}(\mathbf{r}) + e^{2} \int \frac{\Delta n^{\mathbf{q}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} e^{-i\mathbf{q} \cdot (\mathbf{r} - \mathbf{r}')} d\mathbf{r}'$$

$$+ \frac{dv_{xc}(n)}{dn} \Big|_{n=n(\mathbf{r})} \Delta n^{\mathbf{q}}(\mathbf{r}).$$

$$\Delta n_v^{\mathbf{q}}(\mathbf{r}) = 4 \sum_{\mathbf{k}v} u_v^{\mathbf{k}*}(\mathbf{r}) \Delta u_v^{\mathbf{k}+\mathbf{q}}(\mathbf{r})$$
incdrhoscf

### DENSITY FUNCTIONAL PERTURBATION THEORY

Sternheimer equation:

$$(\boldsymbol{H}_{SCF}^{\mathbf{k}+\mathbf{q}} + \alpha \boldsymbol{P}_{v}^{\mathbf{k}+\mathbf{q}} - \boldsymbol{\epsilon}_{v}^{\mathbf{k}}) \boldsymbol{\Delta} \boldsymbol{\psi}_{v}^{\mathbf{k}+\mathbf{q}} = -\boldsymbol{P}_{c}^{\mathbf{k}+\mathbf{q}} \boldsymbol{\Delta} \boldsymbol{v}_{SCF}^{\mathbf{q}}(\mathbf{r}) \boldsymbol{\psi}_{v}^{\mathbf{k}}$$

 $C_{\mathcal{S}lpha,\mathcal{S}'eta}(\mathsf{R},\mathsf{R}')$ 

Rev. Mod. Phys. **73**, 515 (2001).



 $ilde{ extstyle D}_{m{s}lpha,m{s}'eta}(m{q})$ 



PHONONS

Phys. Rev. B **43**, 7231 (1991).

### **HPC** implementation

## CALCULATIONS AVAILABLE FROM ph.x

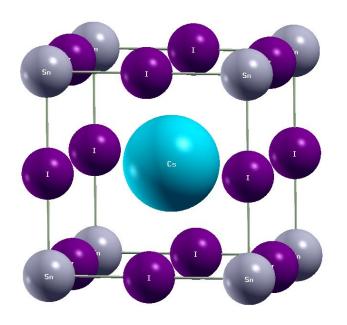
#### The phonon code works for a rather wide variety of systems and methods:

- ✓ Insulators (also polar insulators, with LO-TO splitting)
- ✓ Metals
- ✓ Magnetic systems at the scalar relativistic collinear level
- ✓ Spin-orbit coupling (fully relativistic approach)
- ✓ Electric field calculations: Born effective charges, dielectric tensor

#### Recent developments:

- ! Phonons for magnetic systems in the fully relativistic non-collinear approach
- ! Phonons within the DFT+U approach

### **EXERCISE ON LEONARDO: SIMULATION OF CnSnl3**



- \* Experimentally metallic due to self-doping
- \* In DFT it is a semiconductor (polar material)
- \* 5 atoms in the primitive unit cell
- \* 3 x 5 = 15 phonon modes

### **HPC** implementation

## EXERCISE ON LEONARDO: PHONON MODES AT Γ

cd /leonardo\_work/EUHPC\_TD02\_030/\$USER/max-coe-workshop/day2/example\_ph/

```
* step*/
    * README.md
    * submit.job

* inputs/
    * pw.CnSnI3.in:
    * ph.CnSnI3.in:
    * dyn.CnSnI3.in:
```

\*

solution/

folder for exercise

exercise description job file for Leonardo

folder containing the input files

vc-relax/SCF calculation input phonon calculation input acoustic sum rule input

directory with the reference results

# The phonon workflow PWSCF SIMULATION, STEP 1

The Phonon workflow for modes at a single q point



### The phonon workflow

## PWSCF SIMULATION, STEP 1

1. cd day2/exercise\_ph/step1/

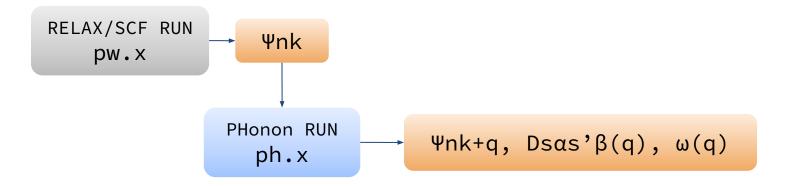
Perform a vc-relax calculation for CnSnI3 using the pw.x program.

- Copy .../inputs/pw.CnSnI3.in in the current folder and modify &CONTROL namelist to do a vc-relax
- Open <u>submit.job</u> and modify npw to use R&G on 4 MPIs: GPUs
- Submit the job file
- Copy the output directory out/ in the folder of the next step

```
$ cat pw.CnSnI3.in
        calculation = 'vc-relax'
        prefix = 'pwscf'
        outdir = './out'
&SYSTEM
   ecutwfc
                   = 80
   ecutrho
                    = 320
   occupations
                   = 'fixed'
   ntyp
                   = 3
   nat
                   = 5
   ibrav
                   = 0
&ELECTRONS
   conv thr
                   = 1e-14
&IONS
&CELL
        press = 0
        press_conv_thr = 0.05
ATOMIC SPECIES
Cs 132.90545196 Cs-nc-pbesol.upf
Sn 118.71 Sn-nc-pbesol.upf
I 126.90447 I-nc-pbesol.upf
K POINTS automatic
8 8 8 1 1 1
CELL_PARAMETERS angstrom
6.1821206415142775 0.00000000000000000
0.0000000000000000 6.1821206415142775
0.000000000000000 0.0000000000000 6.1821206415142775
ATOMIC_POSITIONS angstrom
     3.0910603207571383 3.0910603207571383 3.0910603207571383
      0.000000000000000 3.0910603207571383 0.0000000000000000
```

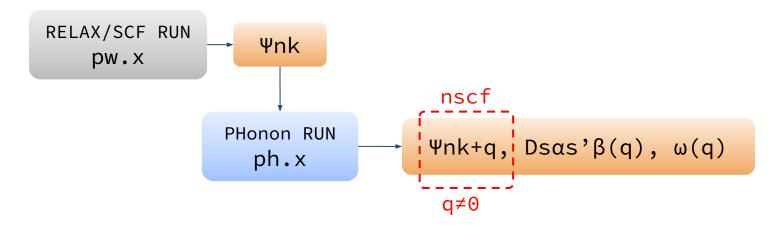
## PHONON CALCULATION, STEP 2

The Phonon workflow for modes at a single q point



## PHONON CALCULATION, STEP 2

The Phonon workflow for modes at a single q point



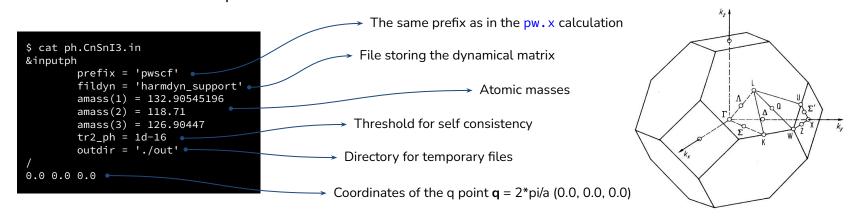
## PHONON CALCULATION, STEP 2

2. cd day2/exercise\_ph/step2/

**QE/INPUT\_PH** 

Perform a phonon calculation at  $\Gamma$  using the ph.x program.

Copy .../inputs/ph.CnSnI3.in in the current folder and modify the &inputph namelist; add coordinates of the Gamma point



- Submit the job file submit.job to run on 1 MPI: GPU

# The phonon workflow DYNMAT MATRIX, STEP 2

- Check the number of k points
   awk '/number of k/' ph.CnSnI3.out
- Check the number of irreducible representations
  - awk '/irreducible/' ph.CnSnI3.out
- Check the dynamical matrix in harmdyn\_ tail -n 97 harmdyn\_support





Optical modes

```
$ tail -n 97 harmdyn support
                          -0.154619 [THz] =
```

## ACOUSTIC SUM RULE (ASR) RULE, STEP 3

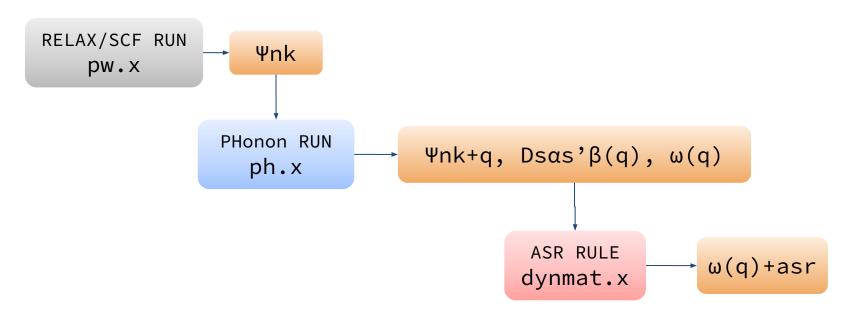
Because of the numerical inaccuracies the interatomic force constants do not strictly satisfy the acoustic sum rule (ASR). ASR comes directly from the continuous translational invariance of the crystal. If we translate the whole solid by a uniform displacement, the forces acting on the atoms must be zero.

For each 
$$lpha,eta$$
 and  $i:\sum_{\mathbf{L,j}}C_{lpha i,eta j}(\mathbf{R_L})=0$ 

As a consequence, the frequencies of the acoustic modes must be zero. ASR can be imposed with dynmat.x

## ACOUSTIC SUM RULE (ASR) RULE, STEP 3

The Phonon workflow for modes at a single q point



### The phonon workflow

# ACOUSTIC SUM RULE (ASR) RULE, STEP 3

3. cd day2/exercise\_ph/step3/

Apply the Acoustic Sum Rule (ASR) with dynmat.x

Because of the numerical inaccuracies the interatomic force constants do not strictly satisfy the acoustic sum rule (ASR). ASR comes directly from the continuous translational invariance of the crystal. If we translate the whole solid by a uniform displacement, the forces acting on the atoms must be zero.

For each 
$$lpha,eta$$
 and  $i:\sum_{\mathbf{L},\mathbf{j}}C_{lpha i,eta j}(\mathbf{R}_{\mathbf{L}})=0$ 

As a consequence, the frequencies of the acoustic modes must be zero. ASR can be imposed with dynmat.x

# ACOUSTIC SUM RULE (ASR) RULE, STEP 3

3. cd day2/exercise\_ph/step3/

**QE/INPUT\_DYNMAT** 

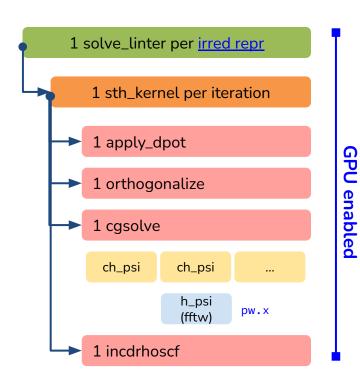
Apply the Acoustic Sum Rule (ASR) with dynmat.x

- Copy .../inputs/dyn.CnSnI3.in and add the 'crystal' ASR rule
- Copy ../step2/harmdyn\_support in the current folder
- Submit the job
- Check phonon modes with ASR rule applied in dyn.CnSnI3.out

```
$ cat dyn.CnSnI3.in
&input
  fildyn = 'harmdyn_support',
  asr = 'crystal'
/
The ASR rule to impose
```

File storing the dynamical matrix

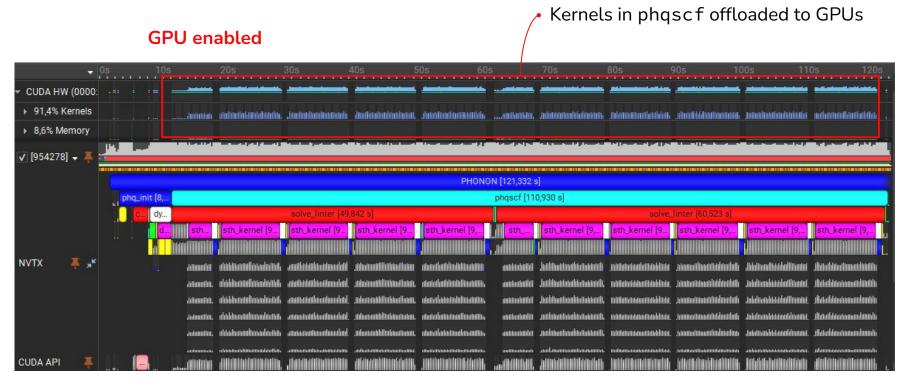
# The phonon workflow PHONON RUNNING ON GPUs



- GPU offloaded version currently available since 7.2
- Most of the routines in the call path for NC pseudopotentials are GPU-enabled
- GPU offload based on **OpenACC + CUDAFortran**
- h\_psi offload inherited from PWscf
- Offload of routines from LR\_Modules/ exploited also in TDDFPT
- Check with tracing tools ;-)

### **GPU** implementation

### **NSIGHT SYSTEM TRACE OF PHONON**



phqscf is the main driver for phonon mode calculation (trans=true)

### PARALLELIZATION HIERARCHIES FOR MULTI-GPU

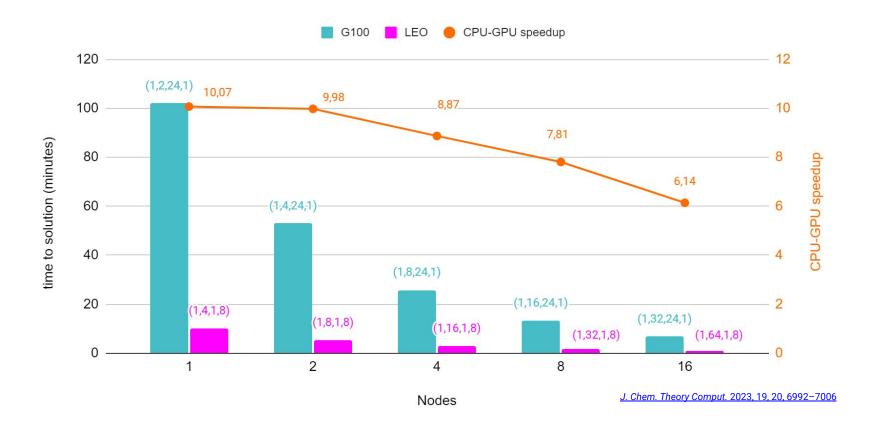
- \* R&G to distribute memory
  - ! distributed FFTs entail communications
- \* pools to distribute calculations on different k-points

```
IF ( lgamma ) THEN
   kunit = 1
   IF (noncolin.AND.domag) kunit = 2
ELSE
   kunit = 2
   IF (noncolin.AND.domag) kunit = 4
ENDIF
```

- \* images to distribute
  - \* <u>irreducible representations</u> (trans=true)
  - \* q-points (ldisp=true)

# GPU implementation PHONON AT LARGE SCALE

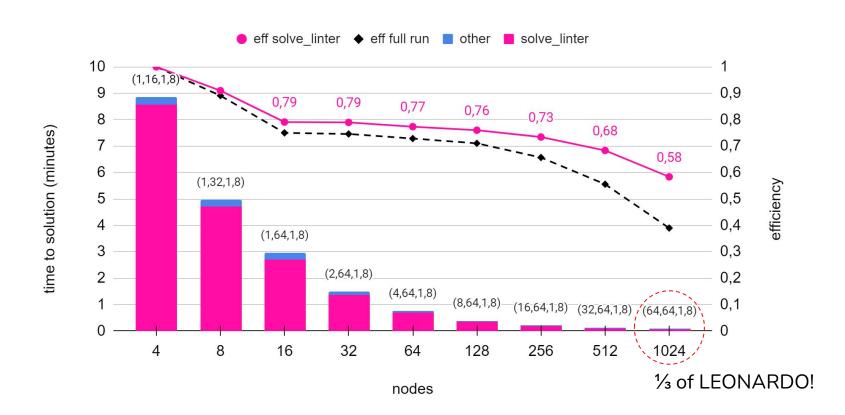
#### Excellent speed up with pool parallelism



## GPU implementation

### PHONON AT LARGE SCALE

Large scale with pools + image distribution



### The phonon workflow

### MULTI-GPU EXECUTION WITH POOLS, STEP 4

4. cd day2/exercise\_ph/step4/ GPU 0 GPU<sub>1</sub> With pool parallelism we distribute k-points among MPI ranks: GPU devices. rank 1 rank 0 mpirun -np N ph.x -nk npools ph.CnSnI3.in umber of k points= 20 cart, coord, in units 2pi/alat 0.0630215 0.0630215), wk = 0.0312500 0.0630215 0.0630215 0.0630215 0.1890645), wk = 0.0937500 0.0630215 0.0630215 0.3151076), wk = 0.0937590 0.0630215 0.0630215 0.4411506), wk = 0.0937500 0.0937500 0.0630215 0.1890645 0.1890645), wk 0.0630215 0.1890645 0.3151076), wk = 0.1875000 0.0630215 0.1890645 0.4411506), wk = 0.1875000 0.0630215 0.3151076 0.3151076), wk = 0.0937500 0.0630215 0.3151076 0.4411506), wk = 0.1875000 0.0937500 10) = 0.0630215 0.4411506 0.4411506), wk 11) = 0.1890645 0.1890645 0.1890645), wk = 0.0312500 k ( 12) = 0.1890645 0.1890645 0.3151076), wk = 0.0937500 13) = 0.1890645 0.1890645 0.4411506), wk = 0.0937500 14) = 0.1890645 0.3151076 0.0937500 0.3151076), wk = 15) = 0.1890645 0.3151076 0.4411506), wk = 0.1875000 16) = 0.1890645 0.4411506 0.0937500 0.4411506), wk = 17) = 0.3151076 0.3151076 0.0312500 0.3151076), wk =

18) =

19) = 1

20) = 1

0.3151076

0.3151076

0.4411506

0.3151076

0.4411506

0.4411506

0.4411506), wk =

0.4411506), wk =

0.4411506), wk =

0.0937500

0.0937500

0.0312500

GPU 1

GPU 0

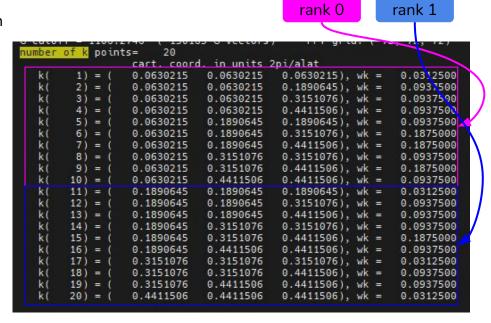
## MULTI-GPU EXECUTION WITH POOLS, STEP 4

4. cd day2/exercise\_ph/step4/

With pool parallelism we distribute k-points among MPI ranks: GPU devices.

mpirun -np N ph.x -nk npools ph.CnSnI3.in

- Copy the input file../step2/ph.CnSnI3.in
- Copy the folder ../step1/out
- Modify npools in submit.slurm to use
   2 pools: GPUs
- Submit the jobfile
- Check PHONON wall time tail ph.CnSnI3.out



### The phonon workflow

### MULTI-GPU EXECUTION WITH IMAGES, STEP 5

5. cd day2/exercise\_ph/step5/

With image parallelism at q point we distribute irreducible representations among MPI ranks: GPU devices.

mpirun -np N ph.x -ni nimages ph.CnSnI3.in

```
$awk '/There are / {x=NR+10} (NR<=x) {print $0} ' ph.CnSnI3.out
There are 5 irreducible representations

Representation 1 3 modes - To be done

Representation 2 3 modes - To be done

Representation 3 3 modes - To be done

Representation 4 3 modes - To be done

Representation 5 3 modes - To be done</pre>
```

## MULTI-GPU EXECUTION WITH IMAGES, STEP 5

5. cd day2/exercise\_ph/step5/

With image parallelism at q point we distribute irreducible representations among MPI ranks: GPU devices.

mpirun -np N ph.x -ni nimages ph.CnSnI3.in

```
 \text{ sawk '/There are / } \{x=NR+10\} (NR<=x) \{print $0\} ' ph.CnSnI3.out 
             5 irreducible representations
There are
                                                                    GPU 0
Representation
                           3 modes - To be done
                                                      rank 0
                   1
Representation
                   2
                           3 modes - To be done
                                                      rank 1
                                                                    GPU 1
Representation
                           3 modes - To be done
                                                                    GPU 2
                   3
                                                      rank 2
Representation
                           3 modes - To be done
                   4
                                                                    GPU 3
                                                      rank 3
Representation
                           3 modes - To be done
                    5
```

### The phonon workflow

## MULTI-GPU EXECUTION WITH IMAGES, STEP 5

5. cd day2/exercise\_ph/step5/

With image parallelism at q point we distribute irreducible representations among MPI ranks : GPU devices.

```
mpirun -np N ph.x -ni nimages ph.CnSnI3.in
mpirun -np 1 ph.x -ni 1 ph.CnSnI3.recover.in
```

- Copy the input file ../step2/ph.CnSnI3.in
- Copy ph.CnSnI3.in as ph.CnSnI3.recover.in and add recover=.true.in &inputph of the latter
- Copy the ../step1/out directory in the current folder
- Modify *nimages* in submit.slurm to distribute on 4 MPIs : GPUs
- Submit the jobfile
  - ! With image parallelism there is 1 output file for each image

! A recover run is needed to collect the IFCs and diagonalize the dynamical matrix

## MULTI-GPU EXECUTION WITH IMAGES, STEP 5

5. cd day2/exercise\_ph/step5/

With image parallelism at q point we distribute irreducible representations among MPI ranks : GPU devices.

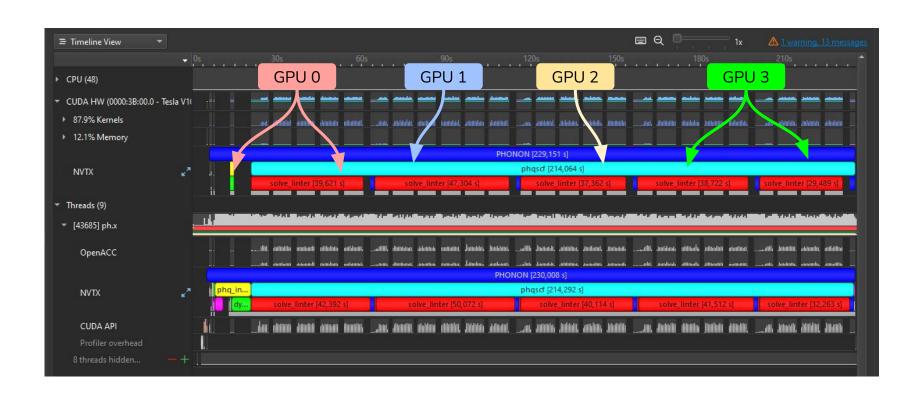
```
mpirun -np N ph.x -ni nimages ph.CnSnI3.in
mpirun -np 1 ph.x -ni 1 ph.CnSnI3.recover.in
```

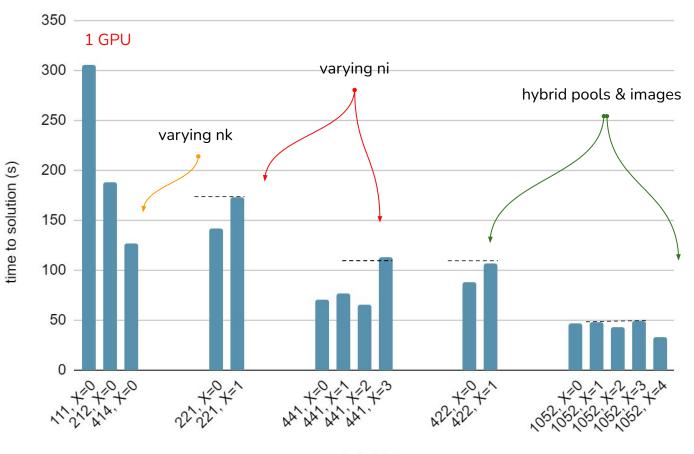
- Check the workfload for each image
  - ! image 0 has an extra scf run to compute the part of the dyn matrix not depending upon the change of Bloch functions
- Compare the wall times. Which image takes longer? Why?

```
$ awk '/I am image/ {x=NR+3} (NR<=x) {print $0} ' out.* 0</pre>
                      0 and my work is about
                                                 4 scf runs. I calculate:
I am image number
q point number
                   1, representations:
0 1
I am image number
                      1 and my work is about
                                                 3 scf runs. I calculate:
 point number
                   1, representations:
                                                 3 scf runs. I calculate:
I am image number
                      2 and my work is about
 point number
                   1, representations:
I am image number
                      3 and my work is about
                                                 6 scf runs. I calculate:
q point number
                   1, representations:
4 5
```

### The phonon workflow

# MULTI-GPU EXECUTION WITH IMAGES, STEP 5





np ni nk, X=image

# The phonon workflow HOW TO EXPLOIT GPUs

- \* Prioritize image and pool distribution if available
- \* Ensure the workload is balanced among MPI processes
- \* R&G is available, but good efficiency is limited to intra-node communications
- Use a GPU-aware version of the code to minimize the cost of H2D-D2H data movements

# CALCULATIONS AVAILABLE FROM ph.x

### Find the input option for your calculation at <a href="QE/INPUT\_PH">QE/INPUT\_PH</a>

```
    Single q calculation (trans = .true.) + ASR (TODAY)

# Dielectric constant (epsil = .true.), effective charges (zeu = .true.)

# LO-TO splitting in insulators and IR cross sections (dynmat.x)

# Raman cross sections (lraman=.true.)

# Phonon mode dispersion (ldisp = .true., ph.x + q2r.x + matdyn.x)

# Electron-phonon interaction coefficients
(electron_phonon='simple', 'interpolated',...)
```