HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HIP

# Case study: Data needs in computational materials physics – atom level simulations

Kai Nordlund

***Department of Physics and Helsinki Institute of Physics***

***University of Helsinki***

***Finland***

# Group and specific research topics

Prof. Kai Nordlund
Principal investigator

Doc. Antti Kuronen
Principal investigator

Doc. Flyura Djurabekova*
Principal investigator

Doc. Arkady Krasheninnikov
Principal investigator
(also Aalto univ., Helsinki)

Doc. Krister Henriksson
Fusion reactor mat's

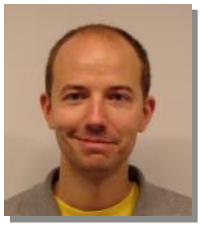
Doc. Jani Kotakoski
Nanostructures
(TU Wien, Austria)

Dr Carolina Björkas
Fusion reactor mat'ls

Dr Lotta Mether*
Plasma physics
(CERN, Switzerland)

M Sc Ane Lasa
Fusion reactor mat'ls

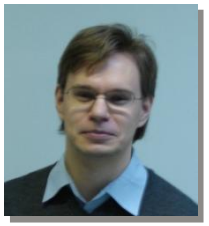
Dr Olli Pakarinen*
Ion tracks



Dr. Hannu-Pekka Komsa
$MoS_2$ nanostructures

Dr. Pi-Heng Chen (陈丕恒)
Quasicrystals

M Sc Andrea Meinander
Fusion reactor mat'ls



M Sc Ville Jansson
Fusion reactor mat'ls
(SCK-CEN, Belgium)

M Sc Jussi Polvi
Organic materials

M Sc Aarne Pohjonen*
Particle physics mat'ls

M Sc Stefan Parviainen*
Particle physics mat'ls

M Sc Marie Backman*
Nanostructures in silica

M Sc Avaz Ruzibaev*
Particle physics mat'ls

M Sc. Mohammad Ullah
GaN and ZnO

M Sc Laura Bukonte
Fusion reactor mat'ls

M Sc Andrey Ilinov
Nanomechanics

M Sc Kostya Avchachov*
Ion tracks

M Sc Aleksi Leino*
Nanostructures in silica

M Sc Wei Ren (任唯)
Carbon nanostructures

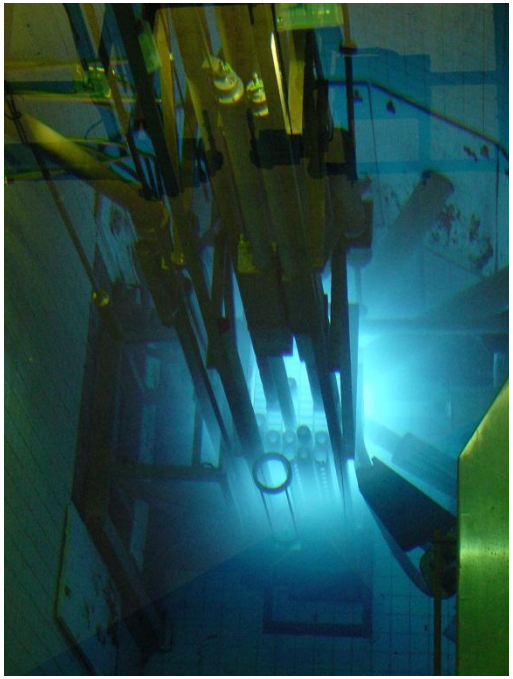
MSc Fredric Granberg
Nanowires

MSc Harriet Åhlgren
Graphene

MSc Morten Nagel
Fusion reactor mat'ls

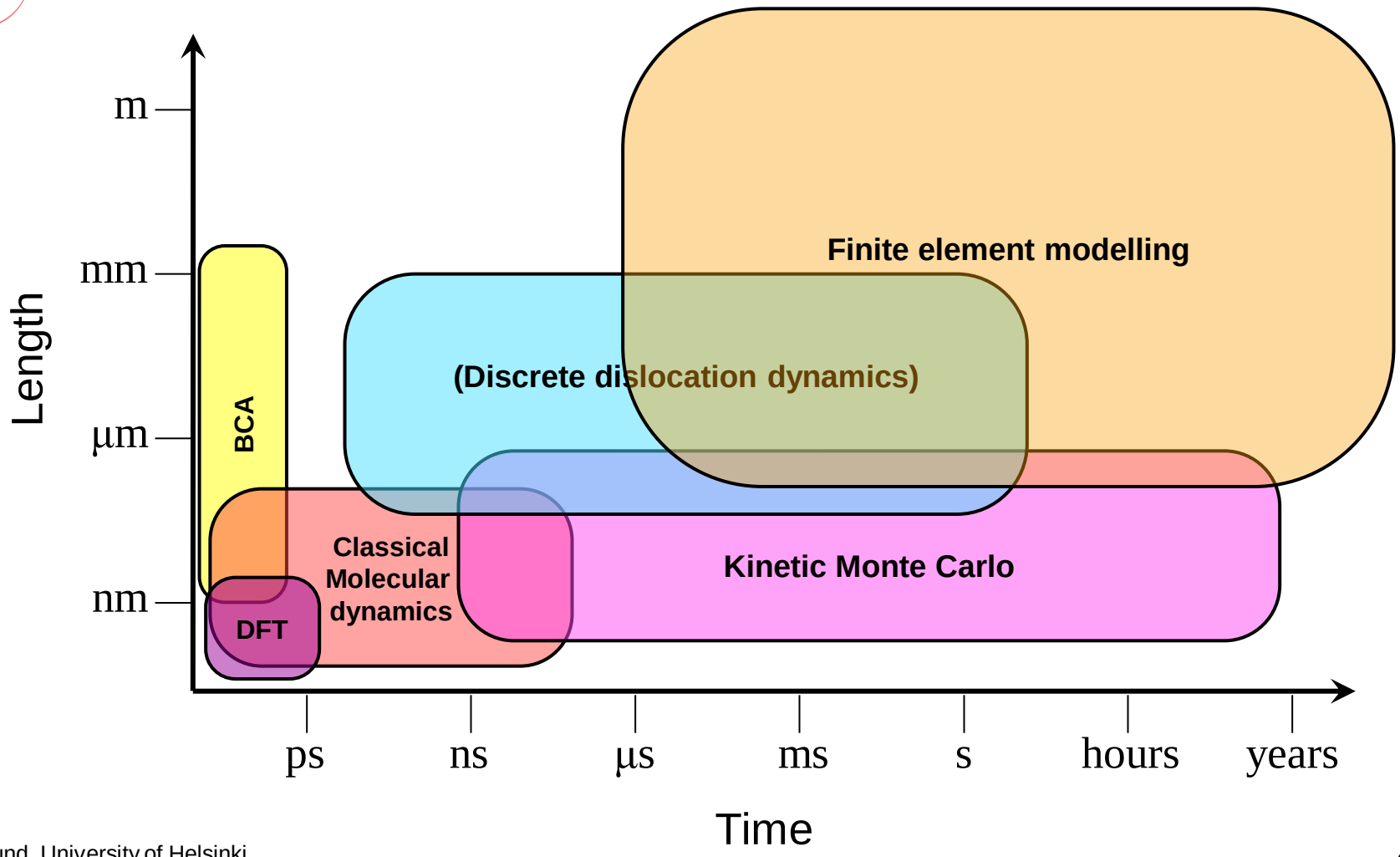
# Research theme

- Radiation effects on a material
  = what happens when an energetic particle from an accelerator or fission or fusion reactor hits a material??
- Can be harmful OR beneficial!



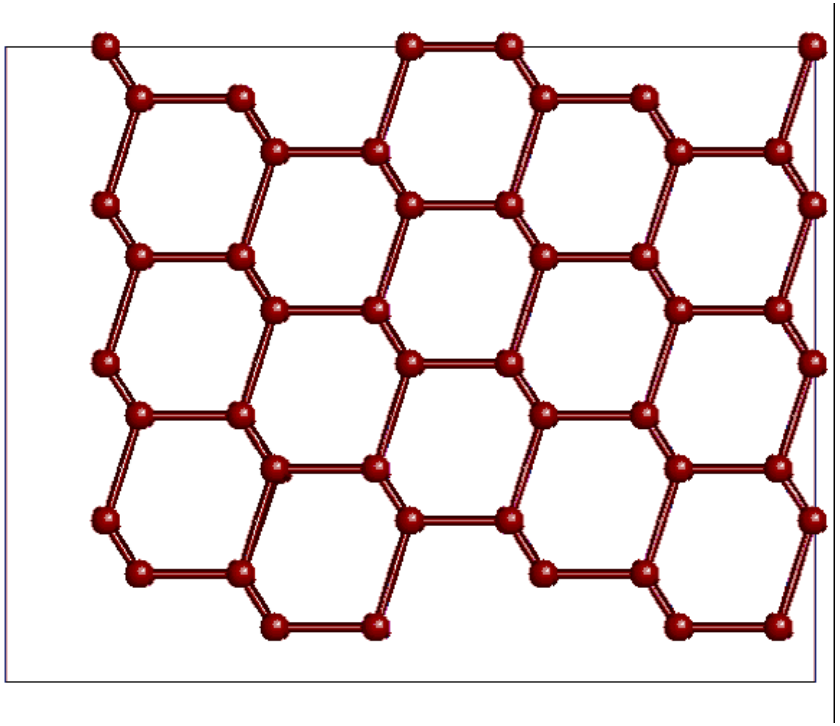
# Multiscale modelling in materials science

- The multiscale modelling framework



# Main simulation type

- The main simulation type in our group is molecular dynamics – MD
- Done on many different scales and materials
  - Quantum mechanical
  - Classical reactive bond-order potentials for organic materials
  - Classical many-body potentials for metals, semiconductors

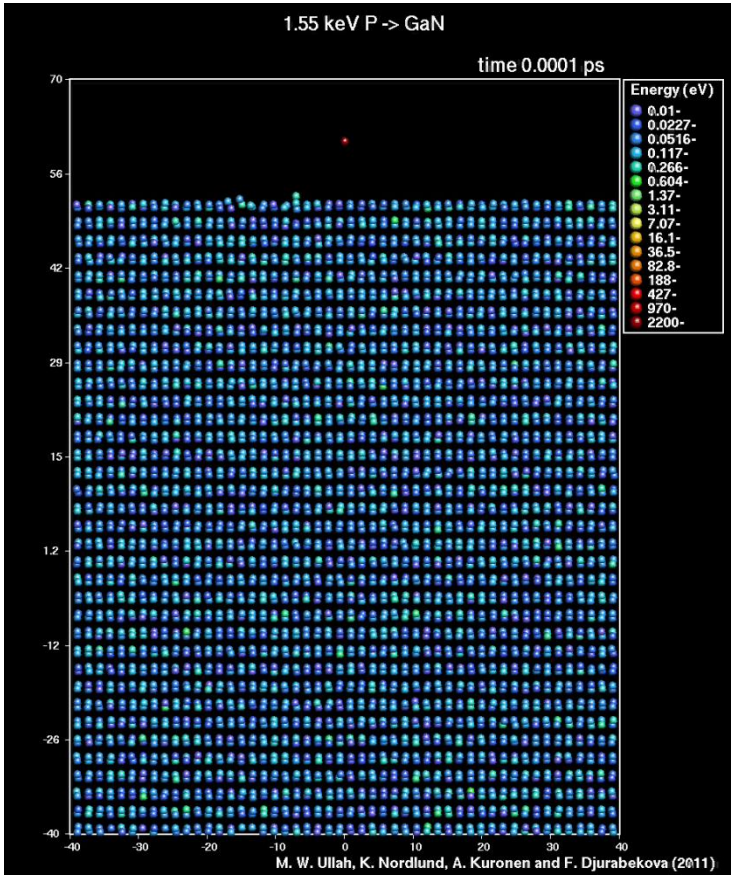


# Range of simulation sizes and data needs

- Initial input data usually trivial, 1 – 10 Mb
  - Intermediate inputs can be largish, 1 – 100 Gb
- Quantum mechanical:
  - 100 – 200 atoms, a few hundred events, < 1000 time steps
    - Output data trivial (< 1 Gb)
- Nanocluster  formation and deposition
  - 1000 – 100000 atoms, thousands of events @ 10000 time steps:
    - Output data 10 – 1000 Gb
- Damage buildup in fusion reactors:
  - 1000 – 10000 atoms, thousands of events @ 10000 time steps:
    - Output data 10  Gb – 1000 Gb
- keV cascades in solids
  - 1 – 1000 million atoms, hundreds of events:
    - Output data 1 – 100 Tb

# Example of user case

- Ion and nanocluster ion irradiation of GaN (Finnish-Russian collaboration ENIGAZ)
  - 5.5 million atoms
  - Hundreds of cascades
- Output structure: physical data + movie file
- Data need for project:
- Without movie file:
  - 1 Gb/event, 2 Tb total
- If movie files were stored:
  - 90 Gb/event, 180 Tb total

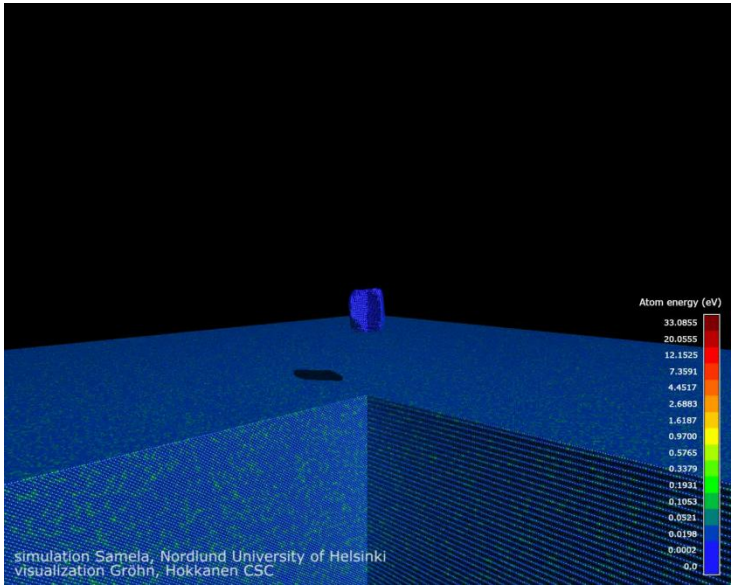


# A high-end user case

- Cratering in Au
- 50 million – 4 billion atoms
  - Single atom output at

    4 billion atoms = 250 Gb
  - Minimal movie file with

    10 frames 2.5 Tb




- If we would do 100 events > 250 Tb…

  (If all movie frames would be output > 2.5 Eb …)
- Minimal long-term storage only physics outputs and final atom coordinates: ~ 25 Tb

# Comments: pros and cons

- Big file size handling is really tedious:
  - Slow analysis, slow moving, even slow file copy within single machine
- Good news: file sizes are easily tunable:
  - Small # atoms: often like 10000 frames output
  - Huge # atoms: ~10 frames of output
  - Physics resolution is lost, but usually we can live with that
- Bad news: we would almost always want to store all final atom coordinates
- Intermediate-term storage space often also a problem:
  - Want to store many movie files in full for 1-2 years
  - Local disk suitable, but keeps filling up

# File format

- All input, most outputs are simple ascii files
  - Human (physicist) readable

```
Neighbours per atom =  26.801300 drmin      2.4481
 elstop allocating
*** first time in elstop in proc   0,  21 data points read in
*** Elstop atom type and file 1 Cuin/elstop.Cu.Cu.in                       , v=
  24644.
Total and last FDe =       0.247      0.247 eV for       1 atoms.
rec    -9.9889   -9.9950  -20.9007       0.0    4001     200.0000  0
t       0.0 ECM pos.      -7.96     -7.95    -16.59 Eksum      252.05
Nliq     0 at time       0.0
 Saving positions to md.movie at t=       0.41
  ... Done !
bpc P   -5.1780   -5.4358   14.2262 sz   36.150   36.150   36.150    47241.63
     Step     Time     Atoms     Temp          Epot_ave   Total En      Change
ec      1       0.41    4001   487.3960      -3.4842      -3.4212     0.0000
ec      2       0.81    4001   487.2544      -3.4842      -3.4212     0.0000
ec      3       1.22    4001   486.2748      -3.4842      -3.4213    -0.0002
```

```
        4001
 Frame number       1    0.20839     fs boxsize  36.150000   36.150000   36.150000
Cu   -17.171275   -17.171398   -17.170782  1         1
Cu   -17.171154   -15.363929   -15.363820  1         2
Cu   -15.363557   -17.171259   -15.363895  1         3
Cu   -15.363913   -15.363804   -17.171200  1         4
Cu   -17.171249   -17.171156   -13.556300  1         5
Cu   -17.171004   -15.363441   -11.748477  1         6
Cu   -15.363760   -17.170994   -11.749114  1         7
Cu   -15.363263   -15.363691   -13.555793  1         8
Cu   -17.171185   -17.171452    -9.941247  1         9
```

- Movie file can also be binary

# Sharing?

- Before publication, we do not want to share anything with outsiders

- After publication:
  - Some of the raw data is not really of interest to anybody else, only the final analyzed data which fits in a publication
  - Some data (atom coordinates in movie file and defect coordinates extracted from them) is of interest to ~ 3-30 other research groups in the world and a sharing system could be useful after publication
    - On the other hand just request by email works fine as well and likely is less effort

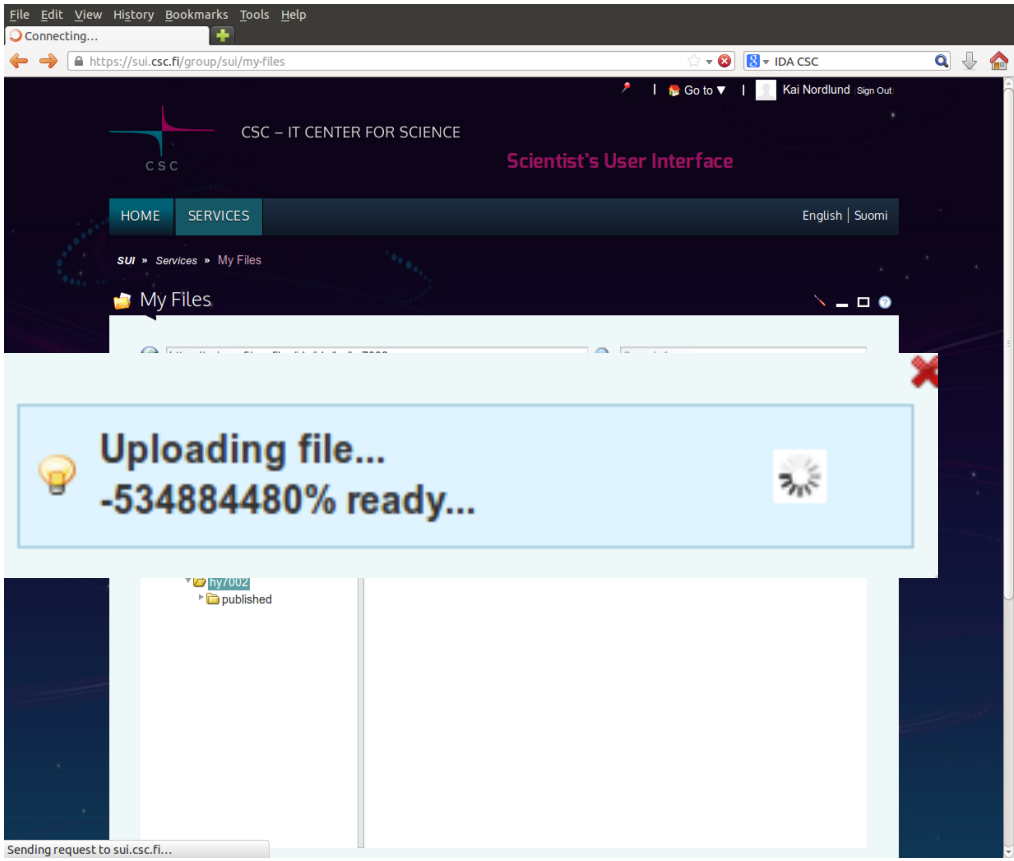
# Time scales for storage?

- Running files and all outputs:
  - ~1 PhD student lifetime (~5 years)
- Most important inputs and outputs:
  - ~ 1 Professor lifetime (~50 years)
  - For my own data storage has worked by data migration from one Unix disk to another for ~ 20 years now
  - Right now 70 Gb long-term storage (including all emails)

```
-rw-r--r-- 1 knordlun Domain Users   628 Jun 21  1995 Vofsize.ltersoff
-rw-r--r-- 1 knordlun Domain Users 16209 Jun 21  1995 coords.in
-rw-r--r-- 1 knordlun Domain Users   628 Jun 21  1995 Vofsize
-rw-r--r-- 1 knordlun Domain Users   116 Jun 21  1995 out.energyofsize
-rw-r--r-- 1 knordlun Domain Users 16209 Jun 21  1995 coords.in.bkp
-rw-r--r-- 1 knordlun Domain Users   374 Jun 21  1995 Vofsize.bkp
-rw-r--r-- 1 knordlun Domain Users  1914 Jun 21  1995 Vofsize.a
```

# First experience of IDA usage

- I tested the TTA/IDA storage recently for a set of nanocluster simulations
- Experience:
  - Graphical user interface non-intuitive, difficult to find place where to store
  - Worked fine except:



# Comments, questions?