Application of the Science Gateway Portal on the Basis of WS-PGRADE Technology for Simulation of Aggregation Kinetics and Molecular Dynamics Simulations of Metal-Organic Nanostructure

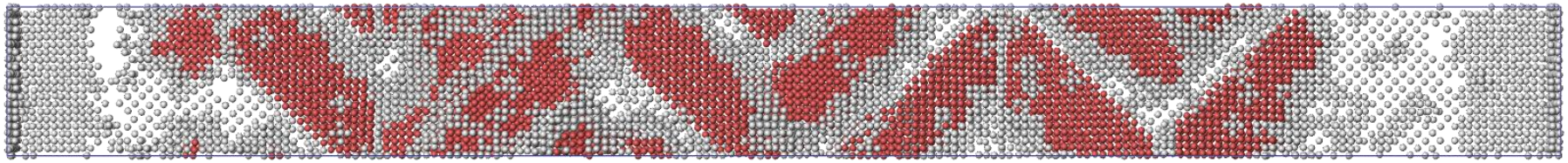
O.Baskova, O.Gatsenko, L.Bekenev, E.Zasimchuk and Yuri Gordienko

G.V.Kurdyumov Institute for Metal Physics (IMP),
National Academy of Sciences, Kiev, Ukraine

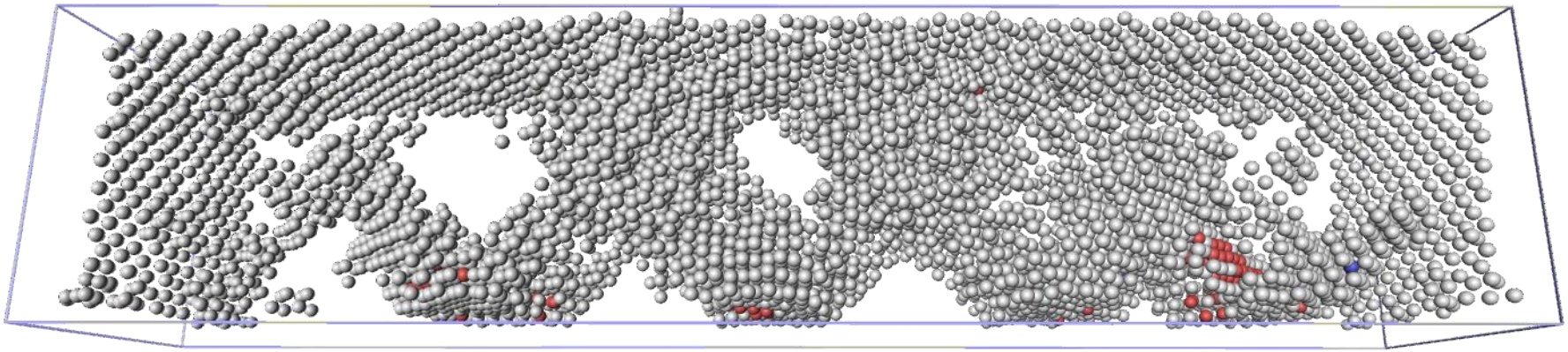
SCI-BUS is supported by the FP7 Capacities Programme under contract nr RI-283481

Scientific Problem: nanoscale research & manufacturing

Increase a range of simulated parameters and find their "magic" (critical) values for atomic self-organization and nanoscale manufacturing.



2D super-lattice on Al surface



3D hierarchic network of voids in Al bulk

Available Computing Infrastructure

- Local Cluster (MPI jobs)
- Service Grid (as a part of the National Grid Initiative)
- Desktop Grid “SLinCA@Home” connected to SG by EDGeS-bridge

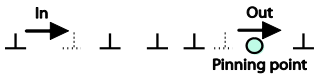
  (made during EDGeS and DEGISCO EU FP7 projects)

Monte Carlo app (cluster, DCI on Desktop Grid)

Theory

Simulations in Desktop Grid

Pile-ups - min active zone



Diffusive kinetics

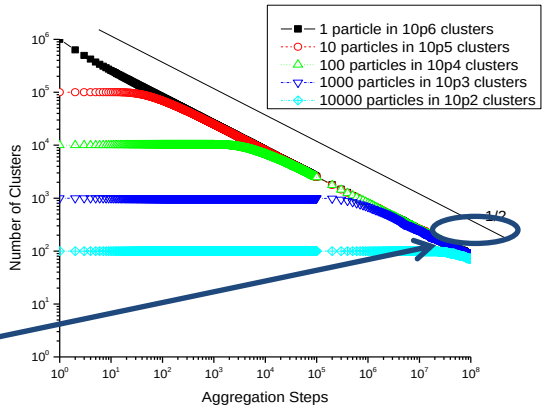
$$\frac{\partial f(n,t)}{\partial t} = D\frac{\partial^2 f(n,t)}{\partial n^2}, \text{ где } D = \tfrac{1}{2} s\ k_d + k_a$$

$$f(n,t) = \frac{1}{2\sqrt{\pi D t}} \int_0^\infty \left\{ \exp\left[-\frac{(n-\xi)^2}{4Dt} \right] - \exp\left[-\frac{(n+\xi)^2}{4Dt} \right] \right\} g(\xi)\, d\xi$$

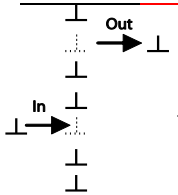
$$\langle N \rangle = \frac{\int_m^M n f(n,t)\, dn}{\int_m^M f(n,t)\, dn}$$

constant for $a^2 >> 4Dt$

$\sqrt{t}$ for $a^2 << 4Dt$



Wall - max active zone



Diffusive kinetics in heterogeneous media

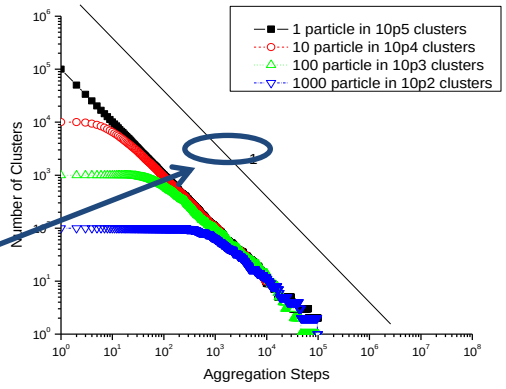
$$\frac{\partial f(n,t)}{\partial t} = D\frac{\partial^2 [n f(n,t)]}{\partial n^2}, \text{ где } D = \tfrac{1}{2} s\ k_d + k_a$$

$$f(n,t) = \frac{\sqrt{a}}{Dt\sqrt{n}} \exp\left[-\frac{n+a}{Dt} \right] I_1\left(\frac{2\sqrt{na}}{Dt} \right)$$

$$\langle N \rangle = \frac{\int_m^M n f(n,t)\, dn}{\int_m^M f(n,t)\, dn}$$

constant for $a^2 >> 4Dt$

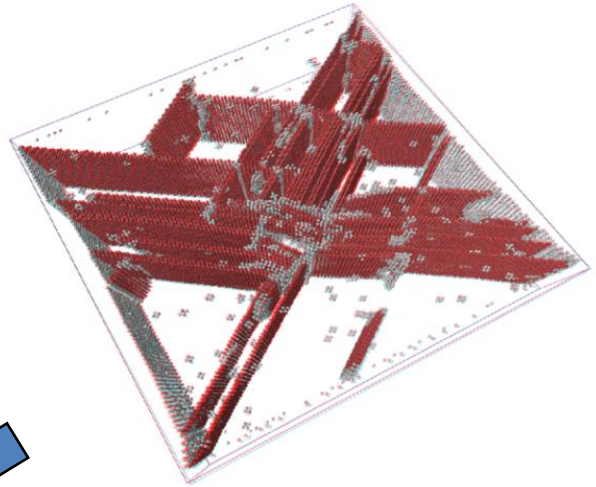
t for $a^2 << 4Dt$

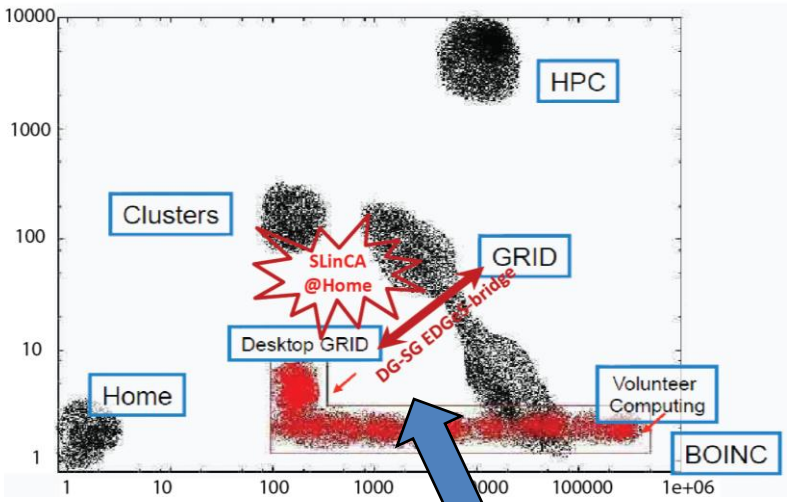


Gatsenko, Baskova, Gordienko, Proc. of Cracow Grid Workshop (CGW'09), Cracow, Poland, pp.264-273 (2010)
Gordienko, International Journal of Modern Physics B (2012), online: Arxiv preprint arXiv:1104.5381 (2011)

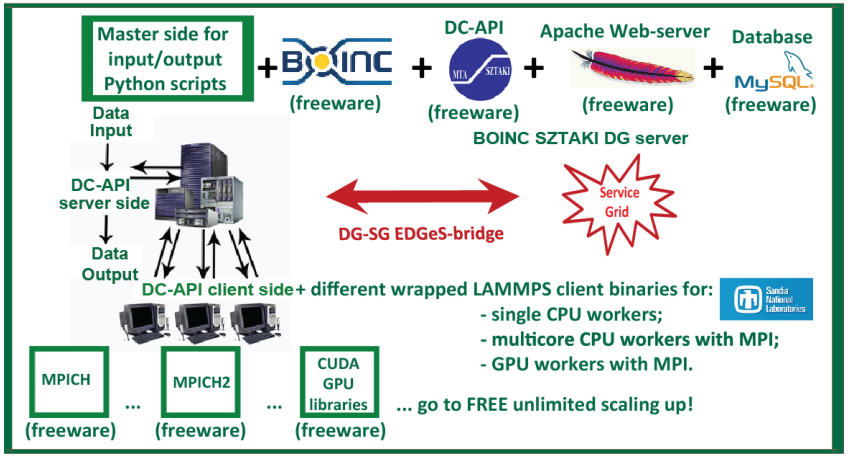
Molecular Dynamics by LAMMPS (cluster, DCI on DG)

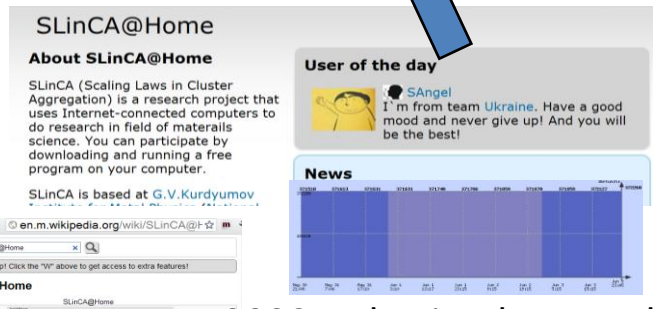
Please, see 3D images with anaglyph glasses





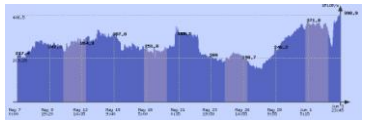
By porting MD to DG-SG DCI!





Page in Wikipedia

~2000 tasks simultaneously

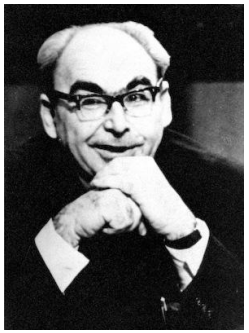


Power (~1.2 TFLOPs)

Typical User Scenario in Molecular Dynamics Simulations

- Design/code the physical process (actors, interactions)
  - atoms, potentials, forces, ambience, etc. (**small** in LAMMPS 4GL script)
- Design/code the initial configuration of atoms (positions and velocities of atoms)
  - input datafile (**BIG** in LAMMPS text format)
  - input file (**small** in LAMMPS 4GL script)
- Schedule/code the output (snapshots of positions and velocities - **BIG**, physical properties - **small**)

What is the Main Aim of scientist?



"A mathematician is a device for turning coffee into theorems."

Alfréd Rényi
prominent Hungarian mathematician

Brute-force generalization: "A scientist is a device for turning anything (coffee, time, money, …) into publications."

(C) YG :)

What is the essence of scientific publication (in materials science, at least)?

Many-page text is IMPORTANT, but essence of paper are: plots, figures, photos!

Well-structured information (post-processed data)!

Main Aim (in short): run simulation to get publication (by clever post-processing the rough data)!

Previously Used Workflow

Task	Software	Infrastructure	Runtime
Molecular Dynamics (MD) simulation			
Large samples (10^5-10^6 atoms)	LAMMPS (MPI-binary)	Cluster	>1-10 … ∞ days
Many (~10^3) small (10^2-10^4 atoms) samples	LAMMPS (sequential binary)	DCI (BOINC Desktop Grid + Service Grid)	>1-100 hours
Post-processing			
Derivative physical values	debyer, XRD, ND, …	Desktop, cluster	>1-100 hours
Statistics on results	R (no binary)	Desktop, cluster	>1-10 hours
Visualization			
3D cross-sections for many (10^2) snapshots	Ovito (GUI-only), AtomEye	Desktop, cluster	>1-100 hours
3D video of evolution	ffmpeg	Desktop, cluster, DCI	>1-10 min

Technical Problems and Ways to Solution

1. Heterogeneous software (binaries, scripts, data formats) of various kinds:
   de facto standard (R, LAMMPS, AtomEye, ffmpeg, …)
   newly born (Ovito, debyer, pizza, …)
   > WS-PGRADE: WF with closed jobs linked in LEGO-style
2. Heterogeneous hardware (local, cluster, DCI)
   > gUSE: resources customized for different jobs.
3. Complex manual operation for their reconciliation
   > WF with “provide input”/”get output” needs only
4. Ad hoc change of physical process after initial data output
   > multistage WF with intermediate output
5. Long learning curve for usual scientists as to DCI internals
   > user-friendly WF constructor and GUI for input/output

Main Milestones to Aim

1. **Smooth access** to heterogeneous software & hardware
2. **Division of roles**:
   a) **Admin** (expert in Computer Science?): portal activities,
   b) **Power User** (principal scientist): **science task formulation**,
   c) **User** (scientists, students): **science task operation** (run simulation, post-process data, visualization)
3. **More complex WF** (added modules, ad hoc changes, …) , **BUT(!)**
4. … **NO additional complexities** (Q: is it naive? A: NO!):
   1. **NO changes in executables** (they are already used!)
   2. **NO changes in input/output formats** (linked to executables)
   3. ALL changes **by scripts & command line arguments ONLY**
5. **Short learning curve** for "non-Computer-Science" scientists

Desirable User Scenarios

Basic idea: separate the "physics" and "computer science" activities.

Power User (scientific task -> definition only):

- Actually design/code a physical process

End User (scientific task -> operation only):

- Manage numerous jobs (submit, monitor, report) by user-friendly interface
- Monitor progress of calculations
- Get results for post-processing and interpretation.

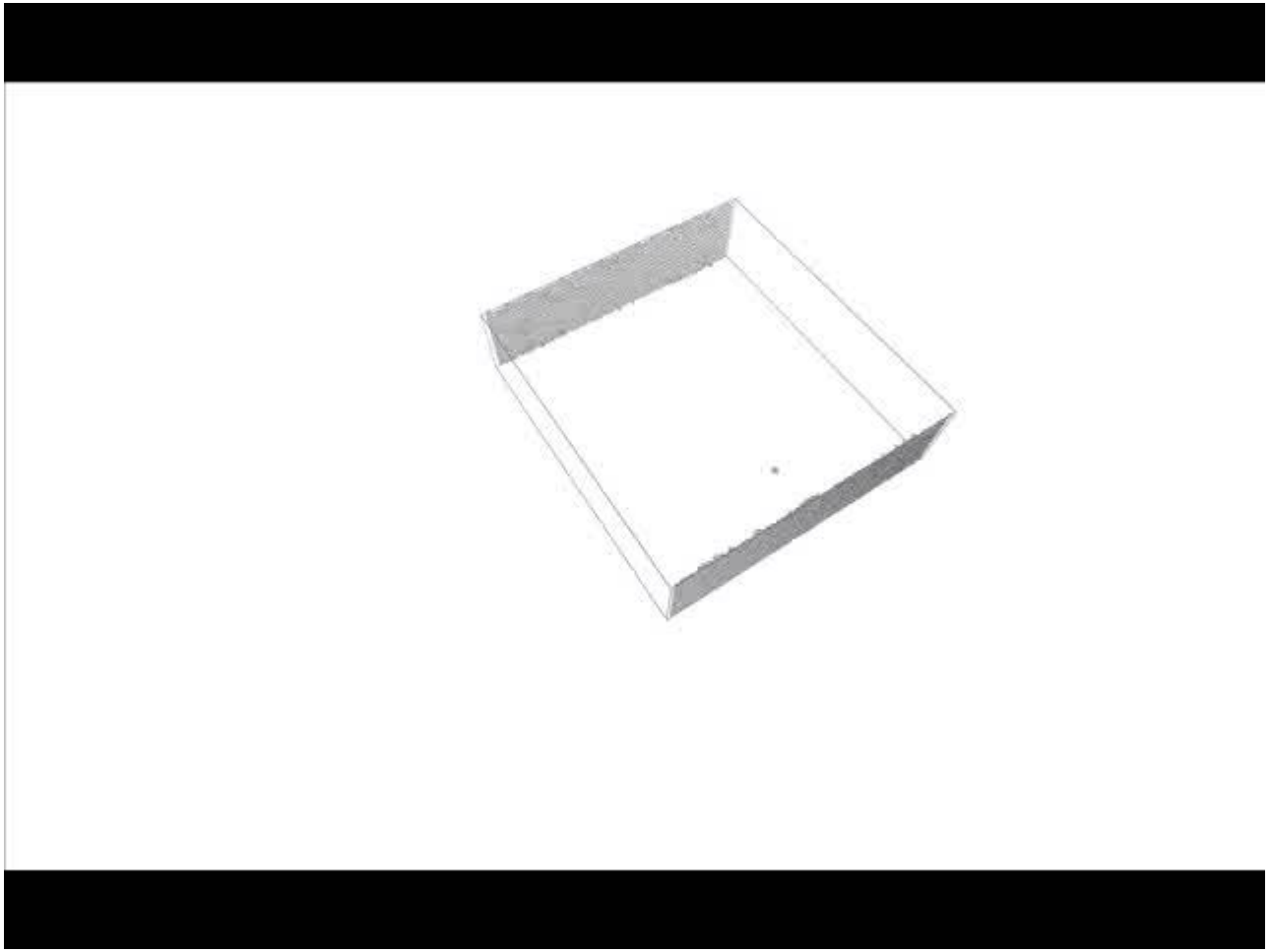
Use Cases

1. mechanical properties (strength, plasticity,…) of a nanocrystal under various conditions
2. … of an ensemble of nanocrystals under the same conditions
3. manipulations with graphene - tension, impact, etc.
4. … with carbon nanotubes (CNTs) – adsorption, conductance, strength, …
5. … with complex metal-organic compounds.

Use Case 1:

Tension of nanocrystal under different conditions

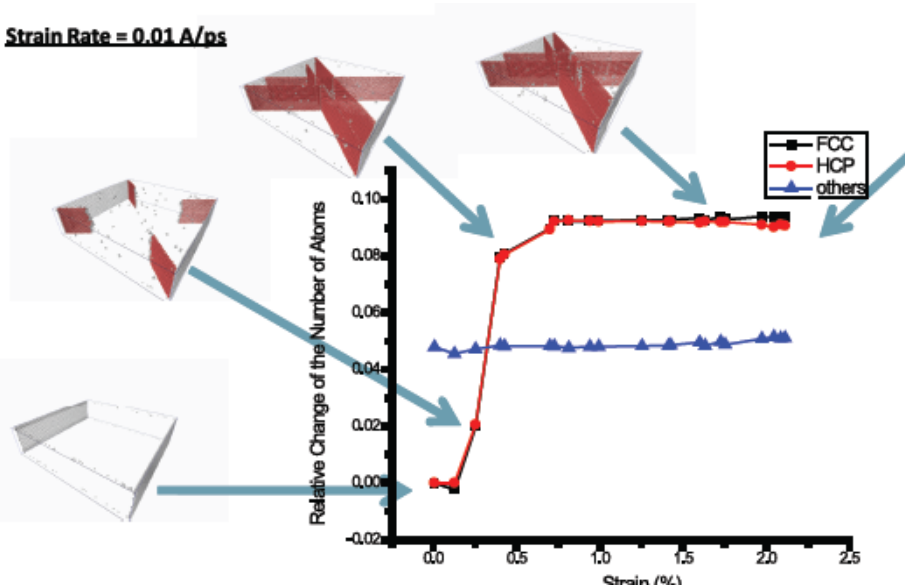
Typical Example: tension of Al nanocrystal

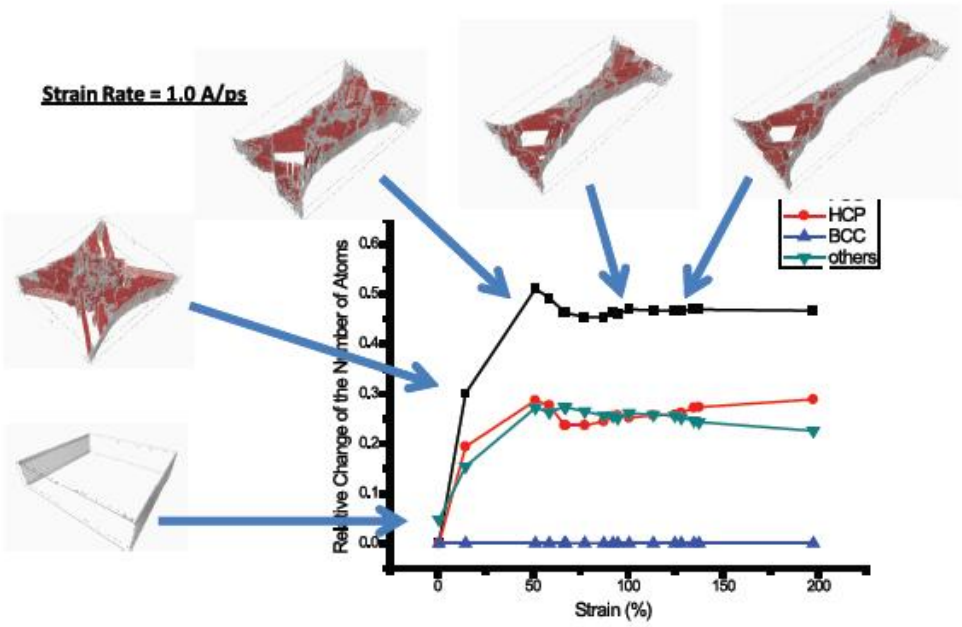


14

Post-processing tasks: strain-stress, defect evolution…

External mechanical influence with different values of strain rate…





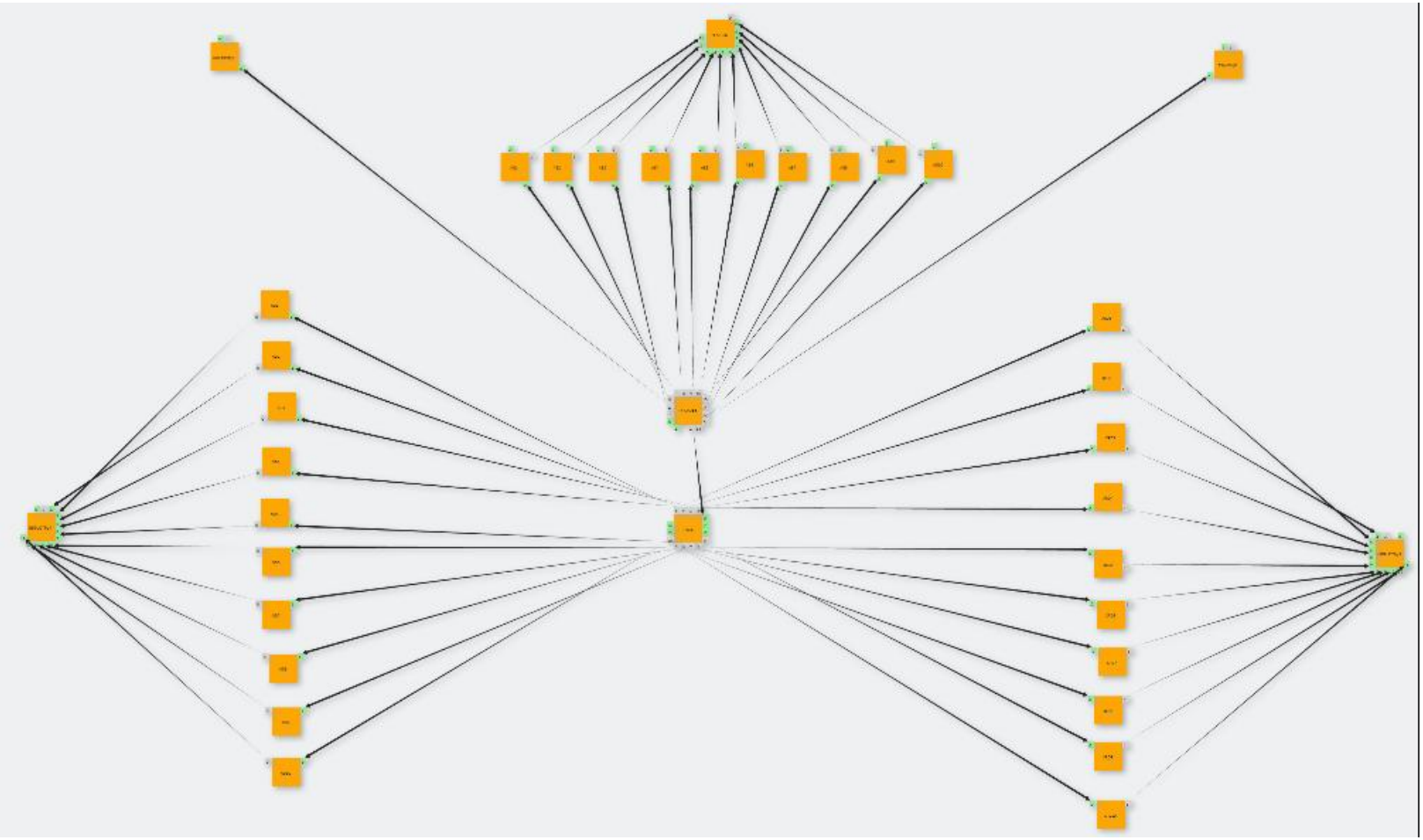
How it can be implemented?

Let’s see at the example of

WS-PGRADE-based workflow

for this Use Case 1

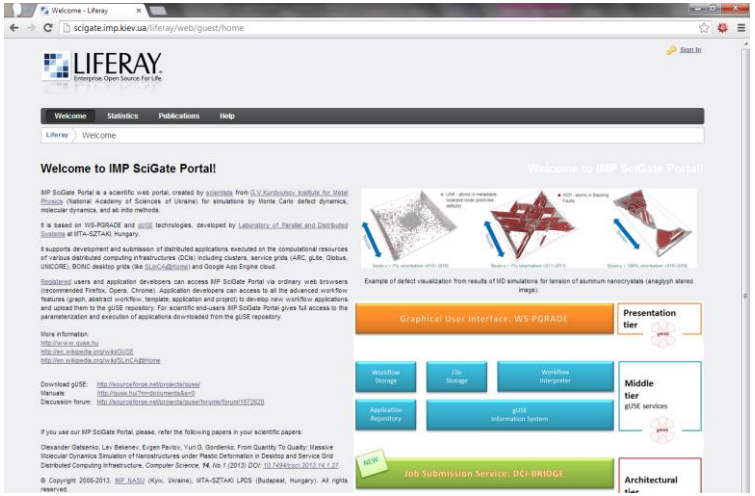
Typical definition of LAMMPS-workflow (Power User role)

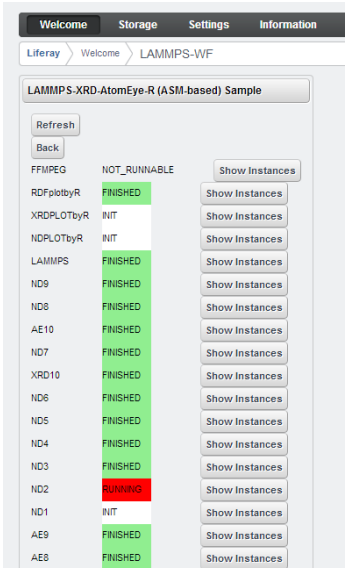


Simple scheme, BUT big work behind curtains for reconciliation of various modules: binaries, data input-output formats, etc.

Typical execution of LAMMPS-workflow (End User role)

IMP SciGate portal (WS-PGRADE+gUSE)

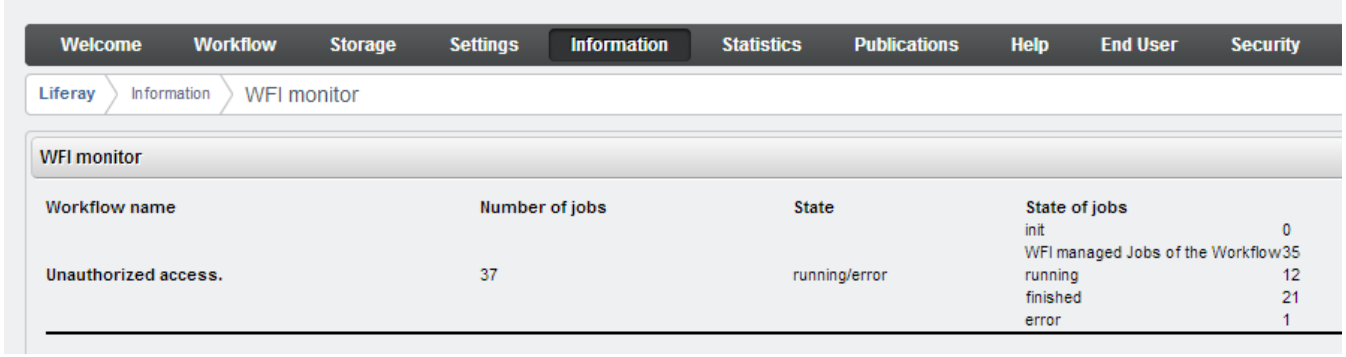




Monitoring the state of jobs in the workflow:

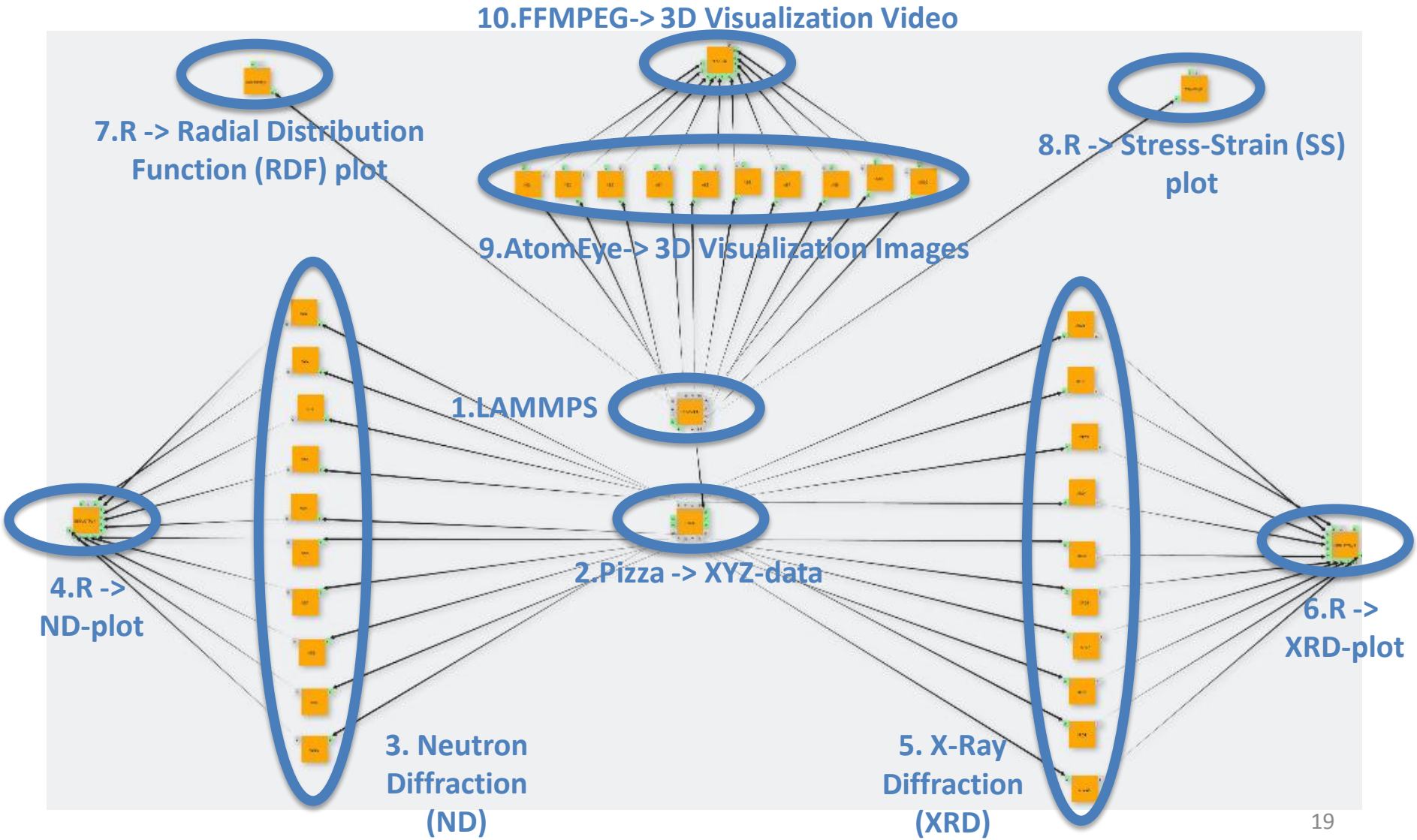
RUNNING
FINISHED
ERROR
INITIATED

Monitoring the workflows

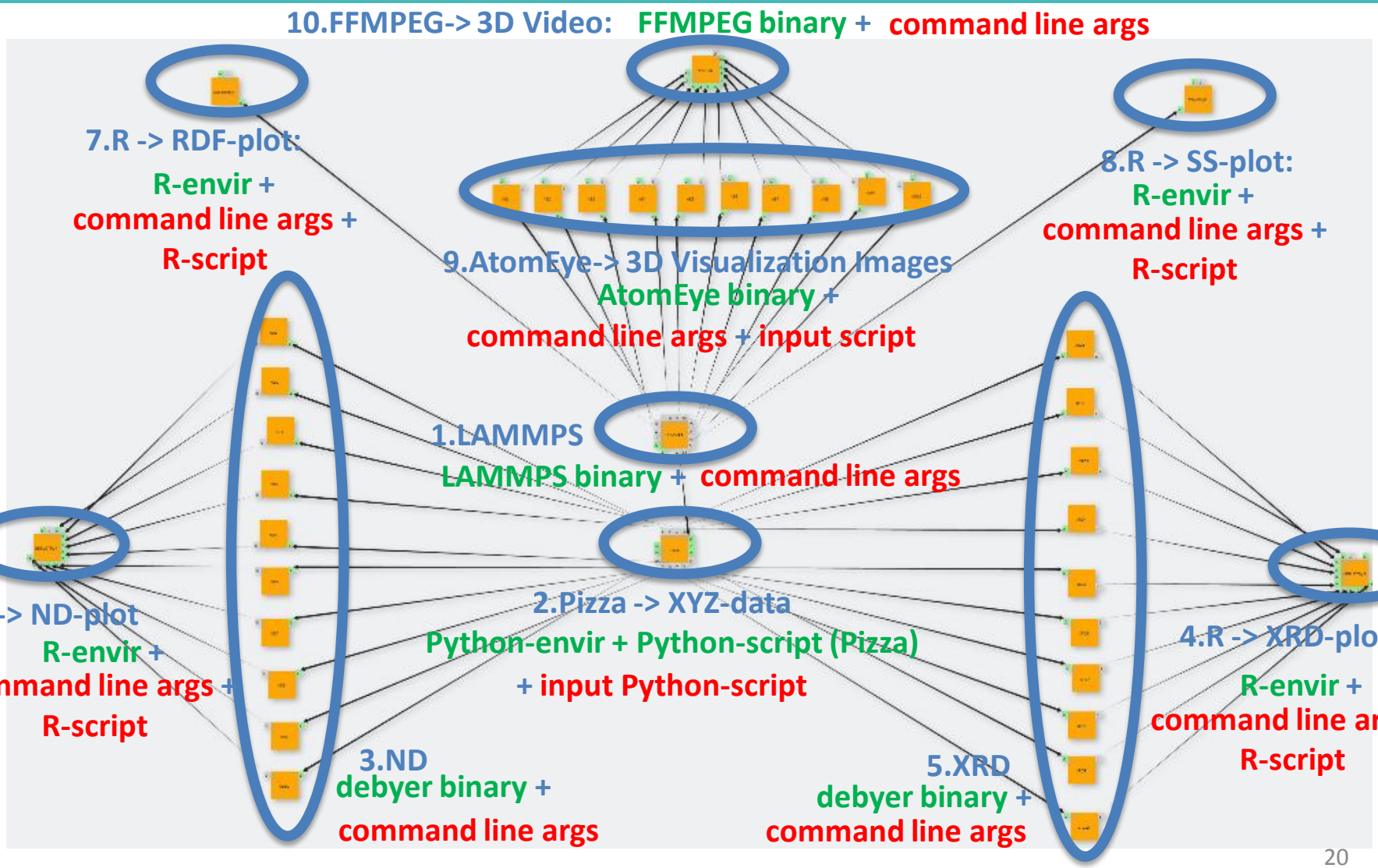


Demo for Use Case 1: http://scigate.imp.kiev.ua/liferay/web/guest/lammps-wf

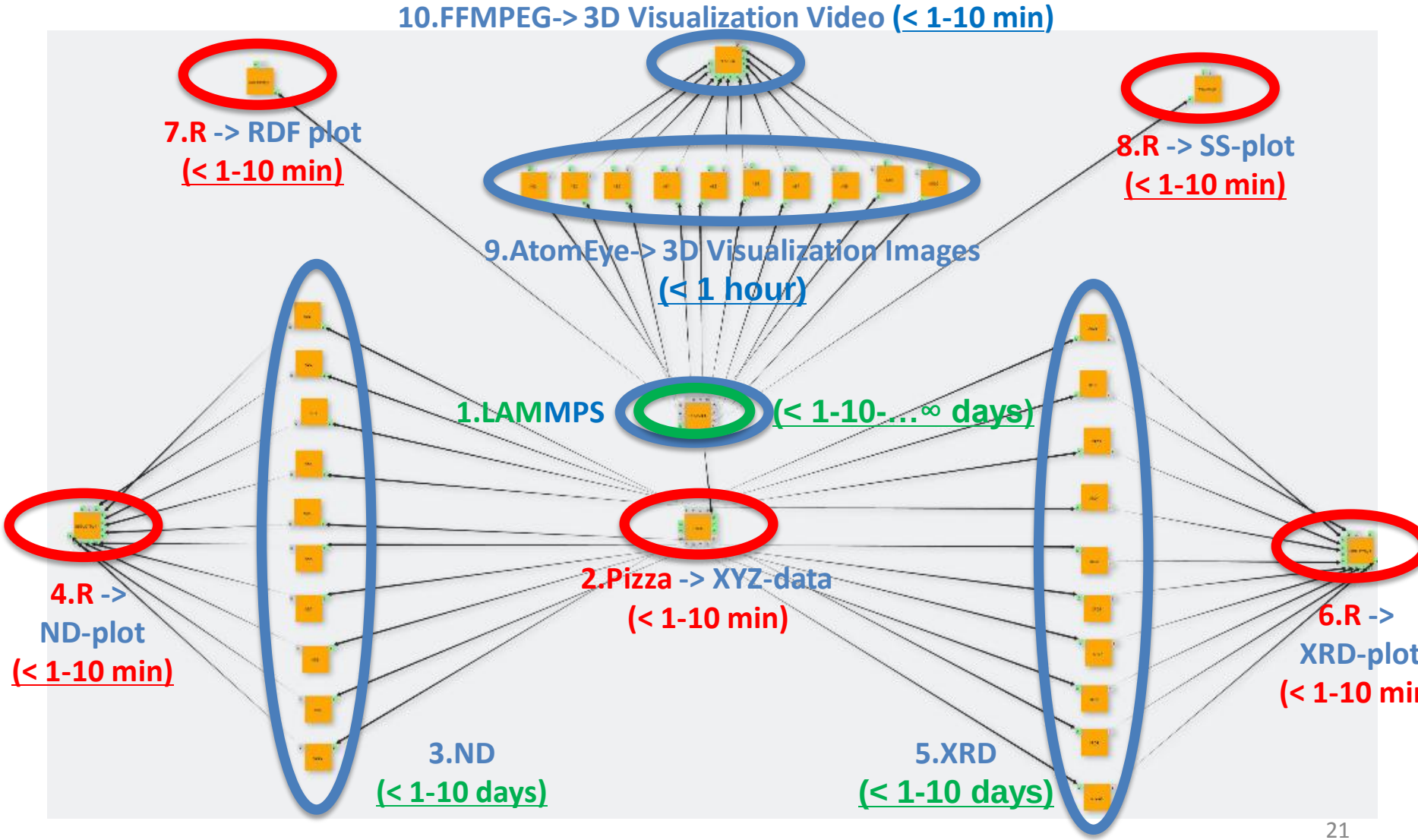
WF-components: LAMMPS+Pizza+ AtomEye+XRD+ND+R+FFMPEG



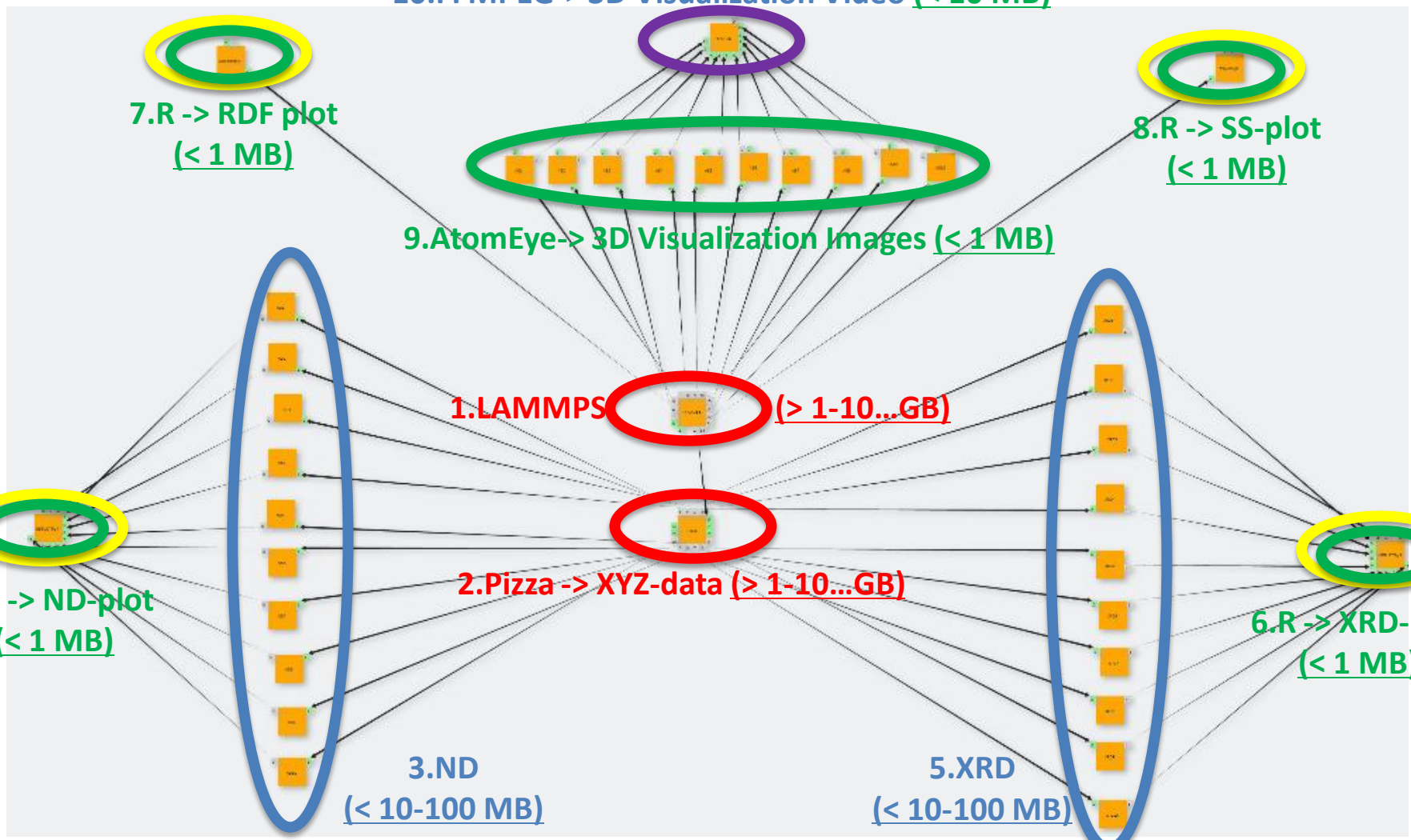
Invariant (execs & envir) and variable (input & scripts) parts



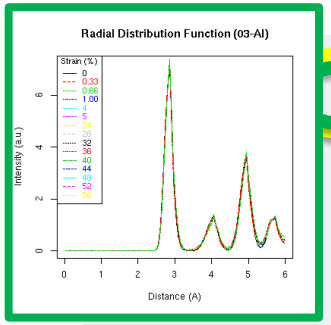
Job Runtime (Resources): Short (Server)+Med (DCI)+Long (Cluster)



Output Data: HUGE text + SMALL text + PLOTs + IMAGEs + VIDEO

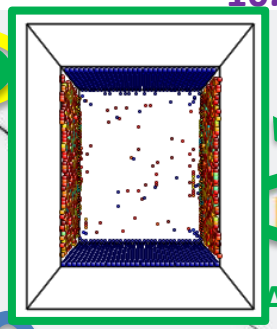


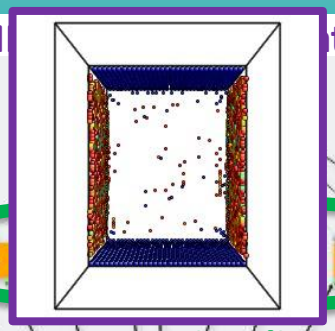
Results: Rough + Processed + PLOTs + IMAGEs + VIDEO



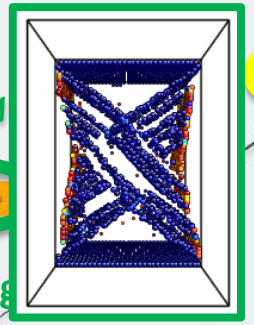
7.R -> RDF-plot

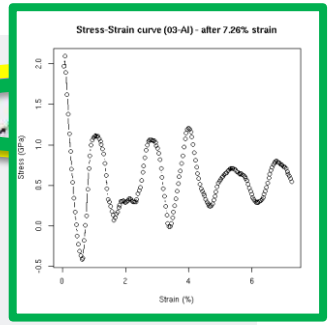
10.FFM ... tion Video



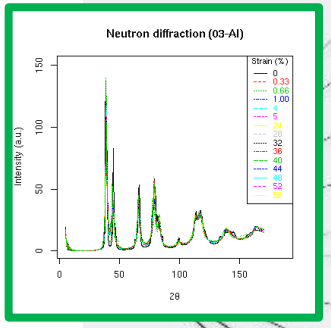


AtomEye-> 3D Visualization Imag





8.R -> SS-plot



4.R -> ND-plot

atoms: x,...,vx,...fx,...,CFG,element,RDF,SS,T,σ,...

1.LAMMPS

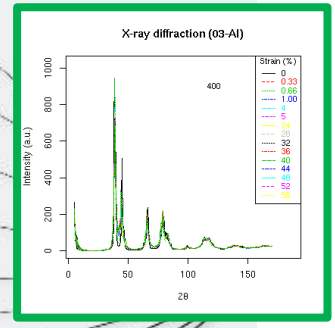
```
ITEM: TIMESTEP
0
ITEM: NUMBER OF ATOMS
```

```
# Time-averaged data for fix rdf1
# TimeStep Number-of-rows
# Row c_myRDF[1] c_myRDF[
1000 100
1 0.03 0 0
2 0.09 0 0
3 0.15 0 0
4 0.21 0 0
```

```
# Fix print output for fix def1
100 0.0033290374126790484542 3.1895106651385676244 3.1806225146104827317
200 0.0070518319386856168074 3.4672571571176282745 3.4973300610684812106
300 0.010452461553787689014 3.4832409281382759048 3.6462118184797946974 2
400 0.014175256079794478978 3.4900966503289105885 3.7870485968929648024 2.
500 0.017898050605801046897 3.4811879221455157563 3.9207326787306642224 2
600 0.021441864625749609 3.5289024104154171901 4.0664675482419125885 2.
700 0.024735105937216925881 3.5686066182493836633 4.1774318185941385906 2.
```

2.Pizza -> XYZ-data

atoms: x, y, z, element



6.R -> XRD-plot

```
#debyer-pattern x-ray lambda=1.5418 cut-off=181.708
#from 5 to 170 step 0.1
5.05 413.423
5.15 481.225
5.25 474.675
5.35 406.157
5.45 324.443
5.55 274.975
5.65 268.06
```

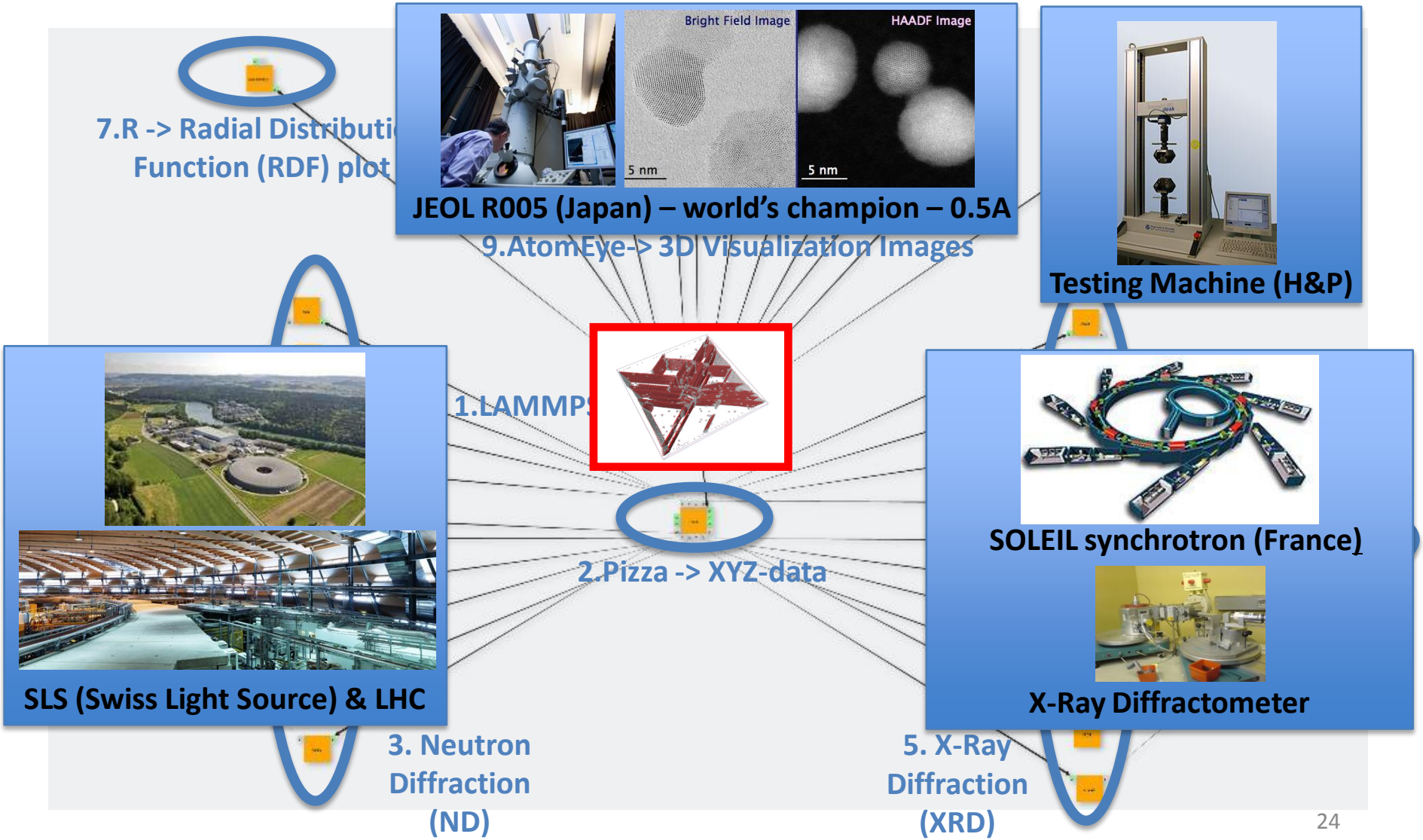
3.ND-data (scattering)

```
2826
Atoms
Al 2.02500070278 5.54475000029e-09 2.86378595843
Al 6.0749898711 5.54475000029e-09 2.86378595843
Al 10.1250212922 5.54475000029e-09 2.86378595843
Al 14.1750202112 5.54475000029e-09 2.86378595843
Al 18.2250191302 5.54475000029e-09 2.86378595843
Al 22.2750180491 5.54475000029e-09 2.86378595843
Al 26.3250169681 5.54475000029e-09 2.86378595843
Al 30.3750158871 5.54475000029e-09 2.86378595843
Al -5.67885679939e-08 1.4318924085 1.43189353917
Al 2.02500070278 5.54475000029e-09 1.71973027999e-07
Al 4.0500066627 1.4318924085 1.43189353917
```

```
#debyer-pattern neutron lambda=1.5418 cut-off=181.708
#from 5 to 170 step 0.1
5.05 29.9966
5.15 34.9563
5.25 34.5208
5.35 29.573
5.45 23.6519
5.55 20.0705
5.65 19.5902
```

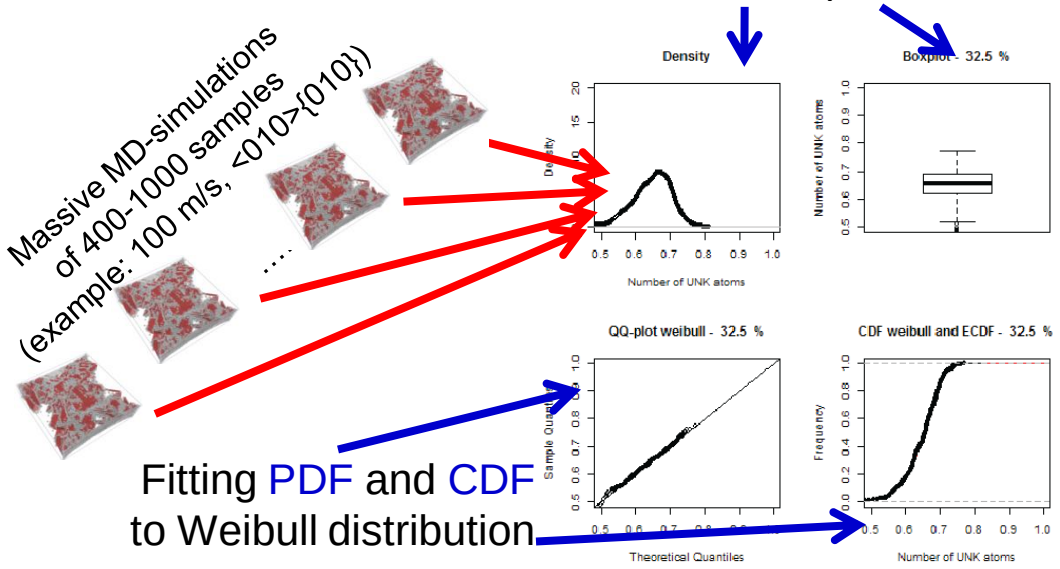
5.XRD-data (scattering)

Workflow as a Hub for Virtual Experimental Labs in Physics

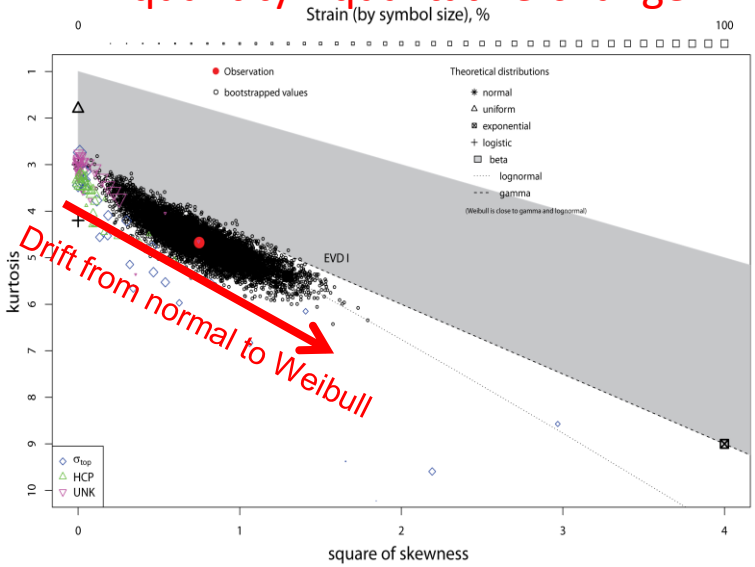


Use Case 2: Set of nanocrystals - different statistical realizations

Distribution (PDF) of concentrations of defects in the ensemble of ~1000 samples



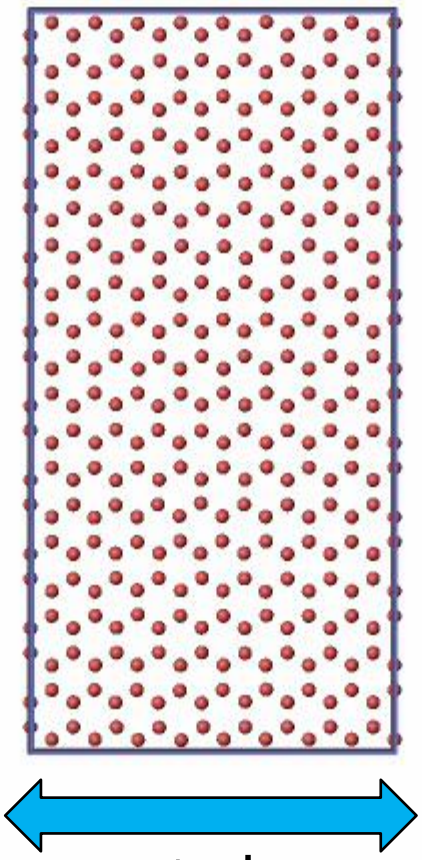
Drift of PDF (from normal to Weibull) in ensemble of ~1000 samples: quantity->qualitative change



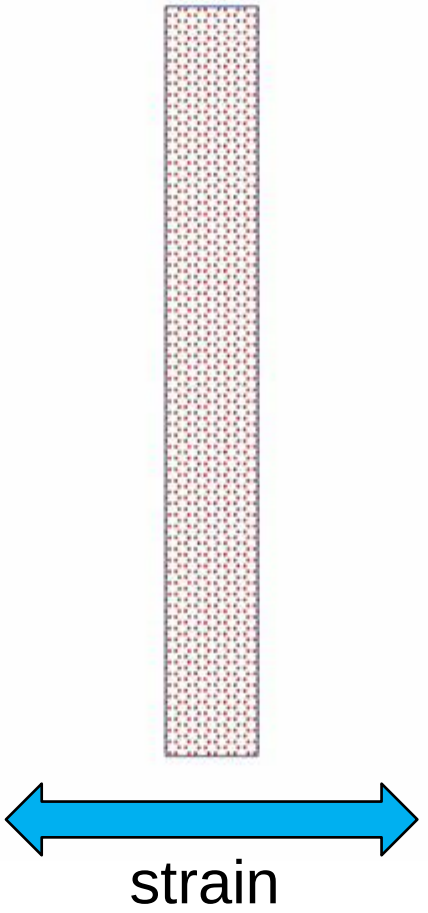
Parameter sweeping allow to find transition from quantity to new quality: observe change of defect distribution with strain, i.e. change of deformation mode!

Use Case 3: Graphene behavior for various parameters

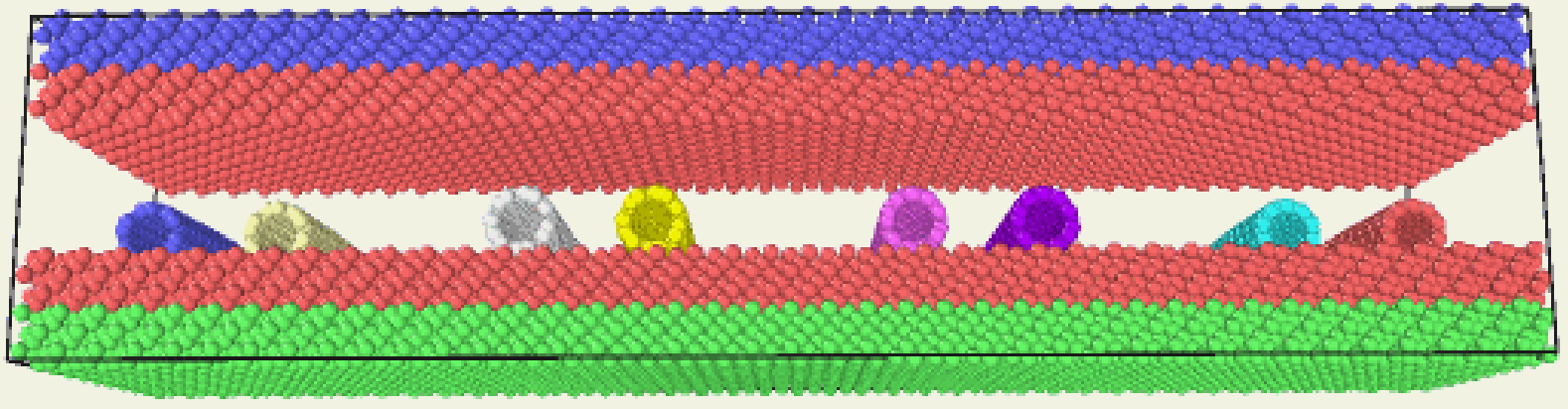
Size: 2x**4** nm



Size: 2x**16** nm



Use Case 4: Manipulations with carbon nanotubes



Detachment of m-CNTs after application of driving force per atom **F=0.17 eV/A** and usage of the second Si-substrate (“stamp”) in the presence of s-CNTs:
two m-CNT c(6,6); two s-CNT c(7,5), two s-CNT c(9,2), and two m-CNT c(10,0)
(from left to right).

From Milestones -> to Conclusions

1. Smooth access to heterogen. soft & hard? **YES (soft), MAYBE (hard)**
2. Division of roles? **YES (at least, 3 levels)**
   a) Admin: portal activities **Q:Expert in comp.sci? A.NO!**
   b) Power User (principal scientist): **science task formulation -> WF definition**
   c) User (scientists, students): **science task operation (simulate, post-process, visualize) -> WF usage (input, start, stop, output)**
3. More complex WF (added modules, ad hoc changes, …) **-> YES**
4. … LOW level of added complexities: **Q: is it true? A: YES!**
   1. NO changes in binaries **–> YES**
   2. NO changes in input/output formats **–> YES, but with intermediate conversion scripts**
   3. ALL changes by scripts & command line arguments **–> YES**
5. Short learning curve for usual scientists? **–> YES, shorter**

Hardships (non-critical)

- **Small number of ports (MAX=16 for gUSE 3.5.5 at the moment)**
  - **limit scale-up for additional modules (now job-replicator is used)**
- **Output file naming convention (alphanumeric only)**
  - **cause problems with legacy code with special symbols**
- **Info like “stdout” and “stderr” are not provided (“No information …” message only) for some errors in WS-PGRADE**
- **Sometimes “stdout” from binary goes to “stderr” of portal (why?)**

Questions (recommendations) to developers of…

WS-PGRADE

- More ports in jobs?
- High-level constructions (LOOP, SWITCH, …)?

gUSE

- More detailed step-by-step “Use-Case Guides” for
  - configuration of connection to various (ARC, Google) resources,
  - complex workflows with conditional branching,
  - best practices (from your experience) on users/resources management

Thank you for efforts

in making these things possible

and

for your attention!