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
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

Pile-ups - min active zone

\[ f_{n,t} = \frac{1}{2\sqrt{\pi}Dt} \exp\left[-\frac{n-a^2}{4Dt}\right] \exp\left[-\frac{n+a^2}{4Dt}\right] g_{\xi} d\xi \]

\[ \langle N \rangle = \frac{\int_{\min} f_{n,t} \ dn}{\int_{\max} f_{n,t} \ dn} \]

constant for \( a^2 >> 4Dt \)

\[ \sqrt{t} \text{ for } a^2 << 4Dt \]

Wall - max active zone

Diffusive kinetics in heterogeneous media

\[ \frac{\partial f_{n,t}}{\partial t} = D \frac{\partial^2 f_{n,t}}{\partial n^2}, \text{ where } D = \frac{1}{2} s (k_d + k_u) \]

\[ f_{n,t} = \frac{\sqrt{a}}{\sqrt{n} \sqrt{Dt}} \exp\left[-\frac{n+a}{2\sqrt{Dt}}\right] t_i\left(\frac{2\sqrt{Dt}}{\sqrt{n}}\right) \]

\[ \langle N \rangle = \frac{\int_{\max} n f_{n,t} \ dn}{\int_{\min} f_{n,t} \ dn} \]

constant for \( a^2 >> 4Dt \)

\[ t \text{ for } a^2 << 4Dt \]

Gatsenko, Baskova, Gordienko, Proc. of Cracow Grid Workshop (CGW’09), Cracow, Poland, pp. 264-273 (2010)
Molecular Dynamics by LAMMPS
(cluster, DCI on DG)

By porting MD to DG-SG DCI!

~2000 tasks simultaneously

Page in Wikipedia

Power (~1.2 TFLOPs)

Please, see 3D images with anaglyph glasses

By porting MD to DG-SG DCI!
• Design/code the physical process (actors, interactions)
  – atoms, potentials, forces, ambience, etc. (small in LAMMPPS 4GL script)
• Design/code the initial configuration of atoms (positions and velocities of atoms)
  – input datafile (BIG in LAMMPPS text format)
  – input file (small in LAMMPPS 4GL script)
• Schedule/code the output (snapshots of positions and velocities - BIG, physical properties - small)
What is the Main Aim of a 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

<table>
<thead>
<tr>
<th>Task</th>
<th>Software</th>
<th>Infrastructure</th>
<th>Runtime</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Molecular Dynamics (MD) simulation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large samples (10⁵-10⁶ atoms)</td>
<td>LAMMPS (MPI-binary)</td>
<td>Cluster</td>
<td>&gt;1-10 ... ∞ days</td>
</tr>
<tr>
<td>Many (~10³) small (10²-10⁴ atoms) samples</td>
<td>LAMMPS (sequential binary)</td>
<td>DCI (BOINC Desktop Grid + Service Grid)</td>
<td>&gt;1-100 hours</td>
</tr>
<tr>
<td><strong>Post-processing</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Derivative physical values</td>
<td>debyer, XRD, ND, ...</td>
<td>Desktop, cluster</td>
<td>&gt;1-100 hours</td>
</tr>
<tr>
<td>Statistics on results</td>
<td>R (no binary)</td>
<td>Desktop, cluster</td>
<td>&gt;1-10 hours</td>
</tr>
<tr>
<td><strong>Visualization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3D cross-sections for many (10²) snapshots</td>
<td>Ovito (GUI-only), AtomEye</td>
<td>Desktop, cluster</td>
<td>&gt;1-100 hours</td>
</tr>
<tr>
<td>3D video of evolution</td>
<td>ffmpeg</td>
<td>Desktop, cluster, DCI</td>
<td>&gt;1-10 min</td>
</tr>
</tbody>
</table>
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
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.
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
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](http://scigate.imp.kiev.ua/liferay/web/guest/lammps-wf)
WF-components: LAMMPS+Pizza+AtomEye+XRD+ND+R+FFMPEG
Invariant \textit{(execs & envir)} and variable \textit{(input & scripts)} parts

1. LAMMPS
   - LAMMPS binary + command line args

2. Pizza -> XYZ-data
   - Python-envir + Python-script (Pizza) + input Python-script

3. ND debyer binary + command line args

4. R -> ND-plot
   - R-envir + command line args + R-script

5. XRD debyer binary + command line args

6. R -> RDF-plot:
   - R-envir + command line args + R-script

7. R -> AtomEye (debyer binary) + command line args + input script

8. R -> SS-plot:
   - R-envir + command line args + R-script

9. AtomEye -> 3D Visualization Images
   - AtomEye binary + command line args + input script

10. FFMPEG -> 3D Video:
    - FFMPEG binary + command line args
Job Runtime (Resources): Short (Server)+Med (DCI)+Long (Cluster)

1. LAMMPS (<1-10-∞ days)
2. Pizza -> XYZ-data (<1-10 min)
3. ND (<1-10 days)
4. R -> ND-plot (<1-10 min)
5. XRD (<1-10 days)
6. R -> XRD-plot (<1-10 min)
7. R -> RDF plot (<1-10 min)
8. R -> SS-plot (<1-10 min)
9. AtomEye -> 3D Visualization Images (<1 hour)
10. FFMPEG -> 3D Visualization Video (<1-10 min)
Output Data: HUGE text + SMALL text + PLOTs + IMAGEs + VIDEO

1. LAMMPS (> 1-10...GB)
2. Pizza -> XYZ-data (> 1-10...GB)
3. ND (< 10-100 MB)
4. R -> ND-plot (< 1 MB)
5. XRD (< 10-100 MB)
6. R -> XRD-plot (< 1 MB)
7. R -> RDF plot (< 1 MB)
8. R -> SS-plot (< 1 MB)
9. AtomEye-> 3D Visualization Images (< 1 MB)
10. FFmpeg-> 3D Visualization Video (< 10 MB)
Results: Rough + Processed + PLOTS + IMAGES + VIDEO

1. LAMMPS
   - 3D Visualization

2. Pizza -> XYZ-data
   - atoms: x, y, z, element

3. ND-data (scattering)
   - atoms: x, vx, fx, CFG, element, RDF, SS, T, σ...

4. R -> ND-plot
   - RDF-plot

5. XRD-data (scattering)

6. R -> XRD-plot
   - XRD-plot

7. R -> RDF-plot

8. R -> SS-plot

9. AtomEye -> 3D Visualization Image

10. FFMPEG -> 3D Visualization Video

Additional Tools:
- 23
Workflow as a Hub for Virtual Experimental Labs in Physics

1. LAMMPS
2. Pizza -> XYZ-data
3. Neutron Diffraction (ND)
4. R -> ND plot
5. X-Ray Diffraction (XRD)
6. R -> XRD plot
7. R -> Radial Distribution Function (RDF) plot
8. R -> Stress-Strain (SS) plot
9. AtomEye -> 3D Visualization Images
10. FFMPEG -> 3D Visualization Video

SLS (Swiss Light Source) & LHC

SOLEIL synchrotron (France)

JEOL R005 (Japan) – world’s champion – 0.5A

Testing Machine (H&P)

X-Ray Diffractometer

Testing Machine (H&P)
Use Case 2: Set of nanocrystals - different statistical realizations

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

Massive MD-simulations of 400-1000 samples (<010×010) (example: 100 m/s, ...

Fitting PDF and CDF to Weibull distribution

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: 2x4 nm

Size: 2x16 nm
Use Case 4: Manipulations with carbon nanotubes

Detachment of m-CNTs after application of driving force per atom $F=0.17 \text{ eV/A}$ 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
   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!