# 11th i-CoMSE Workshop:
# Mesoscale Particle-Based Modeling

## Mississippi State University
## July 21–25, 2025

## Session 15: Reproducibility and open-source software
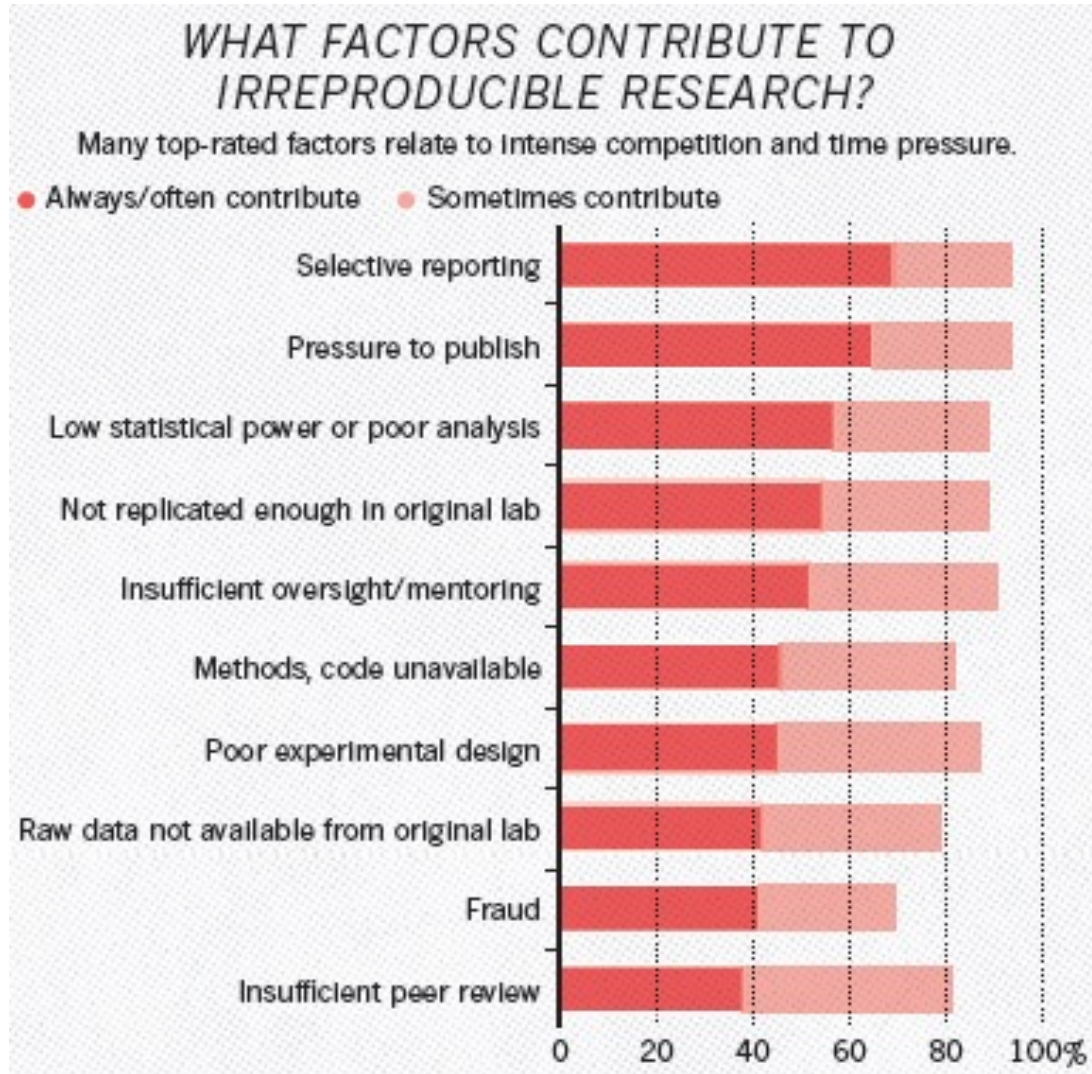
i-Co SE

# Dissemination

- There is no point in doing research if you keep it to yourself!

- Publicly funded research aims to:
  - Promote universal scientific literacy, fundamental understanding of our world
  - Inform policy (energy, health, education, etc.)
  - Advance technologies for addressing societal challenges

- What are some ways you share your research?

# Reproducibility

- Science is not science if it can't be reproduced.

- Science must be reproducible to have *impact.*

- Reproduction is the key step in translating an observation of a single researcher into *knowledge.*

# Reproducibility



WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?
Many top-rated factors relate to intense competition and time pressure.
● Always/often contribute   ● Sometimes contribute

For simulations, what are some concerns you might have for reproducible science?

Survey results from 1500 scientists and researchers [Nature **533**, 452 (2016)]

# Best practices for reproducibility

- Prefer use of scripts (not GUIs or hand-typed commands) to execute your workflow. They create a self-documenting record of your procedure!

- Verify your procedures by reproducing an established result or benchmark.

- Document your code with descriptive comments.

- Keep your data organized and backed up. Does your lab have a data management plan? Do you follow it?

- Share your data and codes. Use public repositories if appropriate.

# Best practices for reproducibility

- Clearly describe all details needed to replicate your procedure.
- Activity
  - How many details can you find are missing in this methods section?

# Open-source software

- "… software that features freely available source code, which users may view, modify, adopt, and share for both commercial and noncommercial purposes."
  - https://github.com/resources/articles/software-development/what-is-open-source-software
- Software will be distributed under a specific license that governs how it can be used.
- Source code can usually be found on a public repository like GitHub.

# Asking for help

- We all need help when we are using simulation software, even if we are experts in it.

- What resources have you used to answer questions about your research / simulations?

- Think about a time someone asked you for help with their research / simulations. What strategies did you use to answer their question?

# Best practices for asking for help

- Search the documentation: does it mention what you want to do?

- Search the discussion board: did someone else have a similar question?

- Describe what you are trying to achieve clearly and concisely.

  - We don't need the full backstory of your work, so it's often better to ask about a specific feature you're trying to use.

  - But, some broader context is still helpful because we might know a way to achieve what you're trying to do that isn't the specific command you asked about.

- Try not to ask questions about skills you can learn elsewhere or other software (ex: "How do I slice a NumPy array?")

- You can ask about whether the software can do something, but scientific questions about how to create a model, what type of measurement to use, etc. are better directed to your research advisor.

# Best practices for asking for help

**Ramon Crehuet**                                                                Oct 14, 2020, 9:01:02 AM   ☆   ⬅   ⋮
to hoomd-users

Dear HOOMD users,
I am quite new to HOOMD. I have two (apparently) very simple questions, which I am not able to solve by reading the documentation.

1. Is it possible to store the potential energy of the configuration that is saved in a GSD trajectory? `dynamic` does not recognize the quantity 'potential_energy', even though `analyze.log` does. I know I could save the energy in a text file with the logger, I was just wondering if it could be stored in the GSD file.
2. If I load a trajectory with `gsd.hoomd.open('trajectory.gsd', 'rb')`, can I calculate the energy of the each frame? I guess I also need to define a system (with the same number of particles and the interactions), but once this is done, how can I assign the coordinates in the trajectory to the system?

Context explanation:
I plan to run long trajectories keeping only uncorrelated frames in the GSD. I would like to calculate the energy of these frames when I slightly modify the parameters of the interaction potential. That is why I wanted to store the initial energy of the frames and why I need to re-calculate the energy of the frames with the new parameters.

Thanks for your time.
Best regards,
Ramon

**Joshua A. Anderson**                                                           Oct 16, 2020, 8:01:36 AM   ☆   ⬅   ⋮
to hoomd...@googlegroups.com

Ramon,

Thanks for asking. The just released HOOMD-blue v3.0.0-beta.1 has a completely new API that we designed specifically with these types of use-cases in mind. There aren't specific tutorials for these yet in the documentation. Here are complete scripts to demonstrate using a LJ liquid.

You can save the potential energy of the system in a GSD file using hoomd.md.compute.ThermodynamicQuantities and hoomd.logging.Logger:
```
```

# Reporting a bug

- We all wish our code was perfect… but there are always bugs to be found.

- Think about a time you have found a bug in your own code. How did you approach fixing it?

- How do you explain to someone else what you think is wrong so they can verify and fix it?

# Best practices for reporting a bug

- Try the discussion board first: has someone had a similar problem? Maybe it isn't a bug!

- Check the issue list: has someone already reported the problem? If yes, is what you're seeing the same or different?

- Create a *minimal reproducer*: the smallest, shortest running script that can reproduce your problem.
  - The scientific context is less important than the specific commands and error.

- Report your bug with all requested information
  - What version of the software? How was it installed or built (configuration, dependencies)? Run on what hardware, OS, etc.?
  - Attach the minimal reproducer script.
  - Typically, paste *text* not screenshots. Developers have limited time, and we can copy and run text more easily!

# Best practices for reporting a bug



https://github.com/lammps/lammps/issues/2970

# Community

- We are all part of one scientific community.

## Lammps fix nvt doesn't work

■ LAMMPS   ■ LAMMPS Mailing List Mirror

**1**        ✉ Apr 2013

**C**        ✉ Apr 2013

I recommend that unless you have the data/arguments to prove that a given feature in Lammps doesn't work, next time try for email subject: "Problems with setting the temp with fix nvt in Lammps", or "My brain doesn't work and thus I cannot understand why the temp from fix nvt is not the one I expect…" , etc. In your case chances are the fix nvt is working just fine and it is you are the one who doesn't know how to use it.

- How could a reply like this negatively impact the poster? The software developers? The community?

# Community

- We are all part of one scientific community.

- Many open-source software developers and people answering questions are volunteering their time to do so because they care about their community.

- Treat each other right!
  - Disagreements are OK and healthy, but be polite.
  - You may not get a reply right away, and that's OK.
  - No flaming, abusive language, harassment, etc.

- Pay it forward: ask questions when you start, share your expertise when you're ready.

- Be empathetic and give the benefit of the doubt if someone isn't following best practices. We were all new at some point!

- An open-source project may have a code of conduct that you should read and adhere to. The code of conduct should also outline procedures for how violations of it will be handled.