nanoHUB: online simulation and data

1.6 M visitors,
16,000+ simulation users per year

Ale Strachan
strachan@purdue.edu
Purdue University
3 Technologies That Could Create Trillion-Dollar Markets Over the Next Decade

By Greg Satell  Updated April 21, 2019 / Original February 17, 2019

Predictive materials models & experiments

Sequential learning for materials discovery

Data science & machine learning

NanoHUB

Cyber-infrastructure
An engaged & growing community

- 620+ simulation tools
- 1,800+ contributors
- \sim16,000+ annual simulation users
- 1,300,000+ simulations per year
- 1.6+ million visitors per year

www.nanohub.org/usage
Apps connected to powerful research codes

*Designed for end users – instructors, students, domain experts*

[Image of Nanomaterial mechanics explorer](https://nanohub.org/tools/nanomatmech)

**Powered by LAMMPS**

[Image of DFT Material Properties Simulator](https://nanohub.org/tools/dftmatprop)

**Powered by Quantum Espresso**

APPifying research scientific codes

Science codes

Rappture

Apps

Rapid APPlication infrastructTURE

Xml description of INs and OUTs

Automatic GUI from xml

Schred

LAMMPS

Keras

TensorFlow

siesta

NEMO5

Polymer Modeler

Padre

Abinit

NAMD  

Science codes

Developers

Users
Jupyter: end to end scientific workflows

1. Use *Polymer Modeler* to create amorphous polymer system

2. Visualize structure

3. MD simulations using *LAMMPS* on HPC resources

4. Post process results and plot

Publish your workflow with a few clicks
We containerize it for reproducibility
Jupyter in App mode

Thermal CNT

This software computes thermal conductivities of single-wall carbon nanotubes. Two versions of the program are available: a tool version for fast set-up and simulations and a tutorial version for students.

Choose the version you want to use:
- Tool version
- Tutorial version

References & tool validation

These tools use the open-source software GROMACS (GROMOS Machine for Chemical Simulations) to estimate thermal conductivities of single-wall carbon nanotubes (SW-CNT). We use harmonic potentials (OPLS-AA force field) for the phononic thermal transfer. The procedure used here to compute the CNT thermal conductivities is explained in the tutorial version of this tool. More details can be found in: E. Chialazzo, T. Asen, "Enhancing surface heat transfer by carbon nanotubes: towards an alternative to nanofluids?", Nanoscale Res. Lett. 6 (2011) 349.

In the graphics below we report the thermal conductivities obtained with the tool version for carbon nanotube lengths between 5 and 50 nm. The conductivity scales with the CNT length as $\lambda \sim L^{0.45}$. The results are in line with those obtained by: R. A. Shelby, K. Torok, Y. Bayazitoglu, "Nanoscale heat transfer: 3D nano-hot spots, 1D heat transfer, and thermal transport of single wall carbon nanotubes", Int. J. Heat Mass Transf. 53 (2010) 5884-5887 and M. Alaghemandi, E. Albayrak, M.C. Börö, F. Möller-Pesth, "The thermal conductivity and thermal rectification of carbon nanotubes studied using reverse non-equilibrium molecular dynamics simulations", Nanotechnology 20 (2009) 115704.

Simulation powered book chapter?
Impact on research

2,300+ citations
5,600+ authors

2,000 use in research
1,600+ Outside NCN papers
Impact on education

- Simulations in ALL top 50 US NWR Engineering Schools
- 1,500+ courses
- 44% HBCUs, 37% of all HHE institutions, & 24% of all MSIs
- 170+ companies in nanoHUB courses
Publish for reproducibility and discoverability

- All tools have DOIs
- All tools are indexed by
  - Web of Science
  - Google Scholar
Cyberinfrastructure making apps more powerful

Automatic UQ
Making Simulations and Data More Powerful

1. Automatic UQ

2. Simulation results database

3. Interoperable cyberinfrastructure

Explore data

Machine learning
Data science, machine learning & materials

Learn from data, by example, without (or with little) underlying physics/chemistry information

Dimensionality reduction in multiscale modeling

Predictive models where we lack physics

Design of experiments / optimization

Interatomic potentials via machine learning
Introduction to Machine Learning for Materials Science

The tutorials here will give you an insight into the usage of Machine Learning to apply physics to materials science.

- **Get started** Click on the links below to begin each tutorial.
- **Important** To exit individual tutorials and return to this page, use File -> Close.

**Querying databases, Organizing and Plotting Data:**
- Query Pyrmatgen and Mendeleev for properties like Young's modulus and melting point.
- Organize data into Pandas dataframes and python dictionaries and plot using Pyplot.

**Linear Regression to predict material properties:**
- Perform linear regression using the scikit learn package and predict Young's modulus.
- Visualize trends in data and 'goodness of fit' of linear model.

**Neural Network Regression to predict material properties:**
- Use neural networks to perform non-linear, higher order regression.
- Visualize trends and compare non-linear model to linear regression.

https://nanohub.org/tools/mseml
Data science for design of experiments

Can ML help reduce the number of experiments to achieve a desired goal?

- Existing data
- Predictive model of state of knowledge
- Query information source
- Information acquisition function

Goal achieved?
YES: SUCCESS
NO: Add data & restart

Maximizing Li+ conductivity in solid oxides
Predictive models of current knowledge

Deep neural network

Random forests

https://nanohub.org/tools/citrinetools
Finding best conductor with the fewest experiments

- Run the notebook online
- Adapt it to your specific problem

https://nanohub.org/tools/citrinetools
Learn faster exploring community data

Use non-numerical inputs for clustering Sankey diagrams to explore clusters

Find patterns in the data correlation analysis

Machine learning models

Detect outliers
Visually explore the simulation database

DFT Material Properties Simulator

Sankey diagrams to explore simulation DB

https://nanohub.org/tools/dftmatprop
Explore correlations between inputs & outputs

Outputs (equilibrium volume and bulk modulus) correlate with KE cutoff

Outputs (equilibrium volume and bulk modulus) correlate with number of k-points
Learn from previous community runs

Silicon

Aluminum
nanoHUB community

- Work Plan
- Pilot projects
- Development
- Engagement through Content Stewardship
- Wish lists
- Support Tickets
- User Contributions
- Analytics

Focus / Effort

Self-Serve Infrastructure

Partner Projects

Custom Dev

nanoMFG

nanoBIO

GRANTA

KIM
**Vision:** The nanoMFG node aims to assemble, develop, and implement advanced cyberinfrastructure tools for research, education, and industrial deployment of integrated, nanoscale manufacturing processes and systems.

**About the Node:**
- Launched in 2017, Illinois and Berkeley
- Develops tools that integrate computational and experimental data around nanomanufacturing processes
- For more info: info-nanomfgnode@illinois.edu

**Get Involved with nanoMFG:**
- Annual Workshop on Data-Science Enabled Advanced in Nanomanufacturing, Spring 2020
- Summer Fellowships for Students
- External Call for Proposals released in Febr 2020

[Image of the NanoMFG node curators and additional information]
Get involved with nanoMFG

**Workshop on Data-Science Enabled Advances in Nanomanufacturing**
- Annual workshop held in Feb/Mar
- Bring together members of the nanoMFG community from industry, labs, and academia
- Presentations, hands-on activities, panel discussions, poster sessions, and networking

**Call For Tool Proposals**
- Focuses on a targeted need for the nanomanufacturing community identified in Needs Assessment Report
- Includes full support for a graduate research assistant or partial support for a post-doc
- Includes two weeks onsite to work with our software development team on site; continued support for remote development
- Call released annually in early spring for next academic year

**Opportunities for Students**
- Opportunities for graduate students
- Two-year graduate fellowship for URM students
- Summer fellowships at Illinois and Berkeley
- Research Experience for Undergraduates via NCSA FoDOMMaT, NCN SURF

The nanoMFG node is committed to promoting diversity and providing an inclusive environment for all participants. Students, post-docs, and faculty from backgrounds traditionally underrepresented in STEM are especially encouraged to apply.

Questions? Contact ertekin@illinois.edu or jayr@illinois.edu
The safe and successful application of nanotechnology in the biological realm is a challenge due to the inherent **multiscale nature of biology**. Engineering **nanoBIO devices** requires the knowledge of nanodevice - biological systems interactions at the bioenvironment, cell, and tissue levels.

**Outreach Activities & Community**

- **Frameworks** for developer support
  - To facilitate new tool contributions
- **Fellowships** for External Collaborators to Develop Apps
- **Code Nano** and Annual **Engineered nanoBIO Workshops**
  - Come check them out!
- **Open-Source** Code Development on GitHub
- **Visualization and Data Analytics Resources**: community for effective inspection of results

[https://nanohub.org/groups/nanobio/](https://nanohub.org/groups/nanobio/)

Contact: Geoffrey Fox (gcf@Indiana.edu)
New Node Highlights: Machine Learning for Nanoscale Simulations on nanoHUB

ML surrogates for MD simulations:
- learn pre-identified features associated with the simulation outputs
- generate accurate predictions for unseen parameters
- enable instantaneous predictions
- improve interactivity with anytime access to simulation results

Check out the ML prediction layer on nanoHUB Tool: Ions in Nanoconfinement
https://nanohub.org/tools/nanoconfinement

Ideas and Details in:
International Conference of Computational Science (LNCS, 2019, vol 11537, pp 116-130)
International Journal of High-Performance Computing (Accepted; arXiv:1910.14620)
nanoHUB

- APPifiying research codes
  - Use emerging/research codes in the classroom
  - Use of Apps by domain experts for discovery
  - Developers: recognition for sharing Apps and data
- Do more with simulations and data
  - Automatic UQ – more bang per click
  - Simulation data automatically stored and discoverable
  - Re-use data to learn and train ML models
  - Drive simulations from sequential learning workflows
- Data analytics
  - Provide recommendations to users
  - Speculative computing

1.6 M visitors every year, 16,000+ simulation users per year
Thanks

nanohub.org
nanohub.org/whypublsih

Lynn Zentner
Mike Zentner
Gerhard Klimeck

Martin Hunt  Steve Clark  Daniel Mejia  Saaketh Desai  Tanya Faltens