The HipGISAXS Software Suite: Recovering Nanostructures at Scale

Abhinav Sarje

Computational Research Division
Lawrence Berkeley National Laboratory

06.24.15
OLCF User Meeting
Oak Ridge, TN

Nanoparticle Systems

- Materials (natural or artificial) made up of nanoparticles.
- Sizes ranging from 1 nanometer to 1000s nanometers.
- Wide variety of applications in optical, electronic and biomedical fields. E.g.:
  - Inorganic nanomaterials in optoelectronics.
  - Organic material based nano-devices such as Organic Photovoltaics (OPVs), OLEDs.
  - Chemical catalysts, drug design and discovery, biological process dynamics.

Importance of structural information:

- Nanomaterials exhibit shape and size-dependent properties, unlike bulk materials which have constant physical properties regardless of size.
- Nanoparticle characterization is necessary to establish understanding and control of material synthesis and applications.

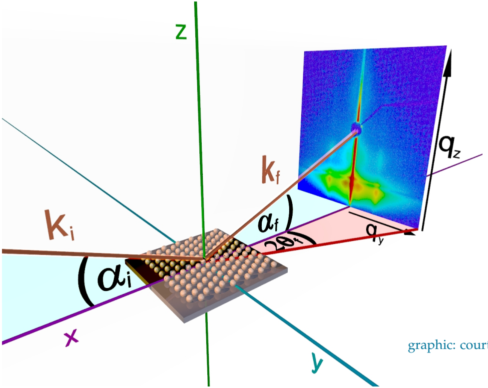
Nanoparticle Systems

- Materials (natural or artificial) made up of nanoparticles.
- Sizes ranging from 1 nanometer to 1000s nanometers.
- Wide variety of applications in optical, electronic and biomedical fields. E.g.:
  - Inorganic nanomaterials in optoelectronics.
  - Organic material based nano-devices such as Organic Photovoltaics (OPVs), OLEDs.
  - Chemical catalysts, drug design and discovery, biological process dynamics.

Importance of structural information:

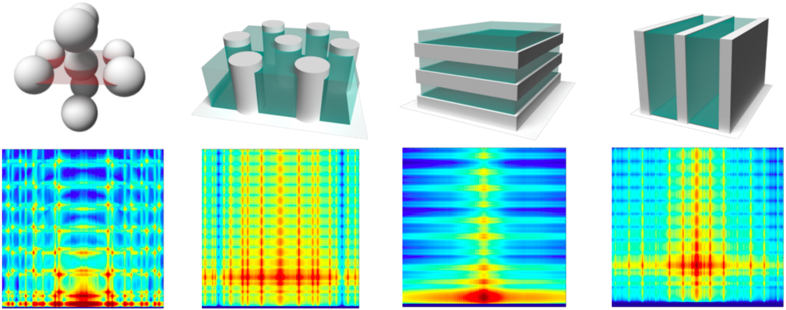
- Nanomaterials exhibit shape and size-dependent properties, unlike bulk materials which have constant physical properties regardless of size.
- Nanoparticle characterization is necessary to establish understanding and control of material synthesis and applications.

X-ray Scattering at Synchrotrons



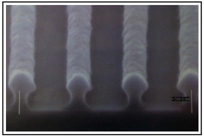
graphic: courtesy of A. Meyer, *www.gisaxs.de*

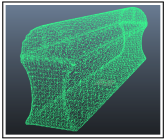
X-Ray Scattering: Examples

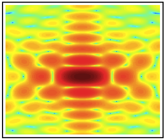


X-Ray Scattering: Complex Examples

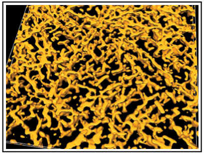
Gratings

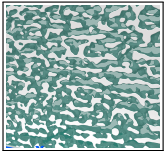


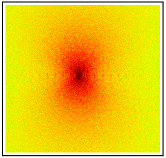




OPVs





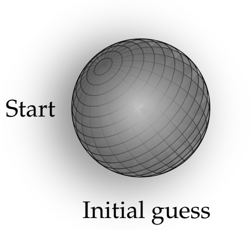


Real Sample

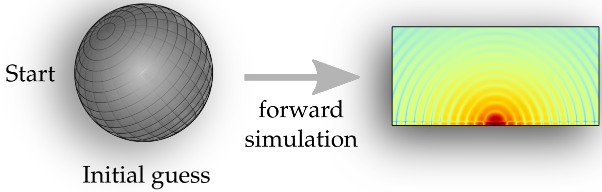
Model

Scattering Pattern

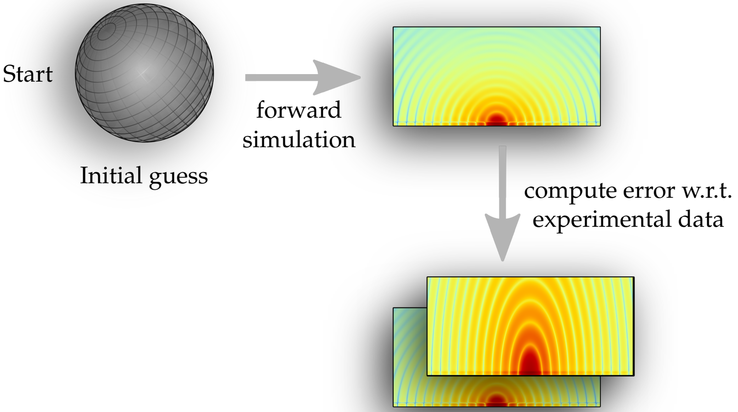
Computational Problems in Structure Recovery: Inverse Modeling



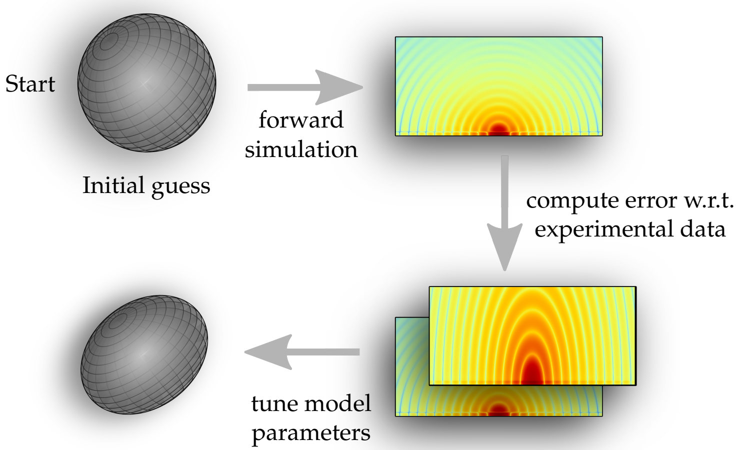
Computational Problems in Structure Recovery: Inverse Modeling



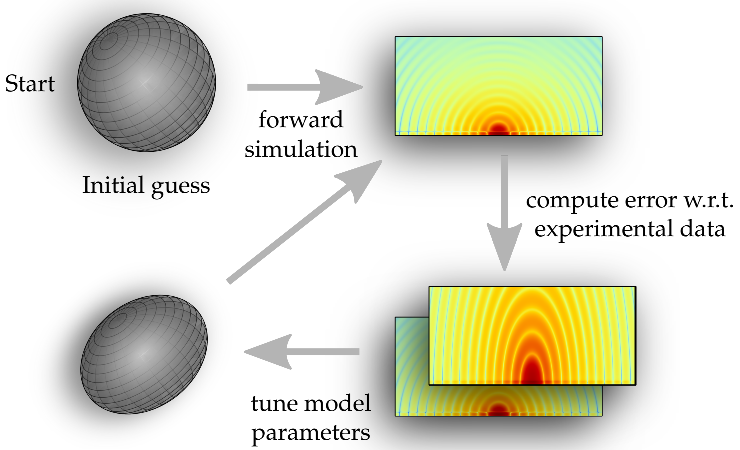
Computational Problems in Structure Recovery: Inverse Modeling



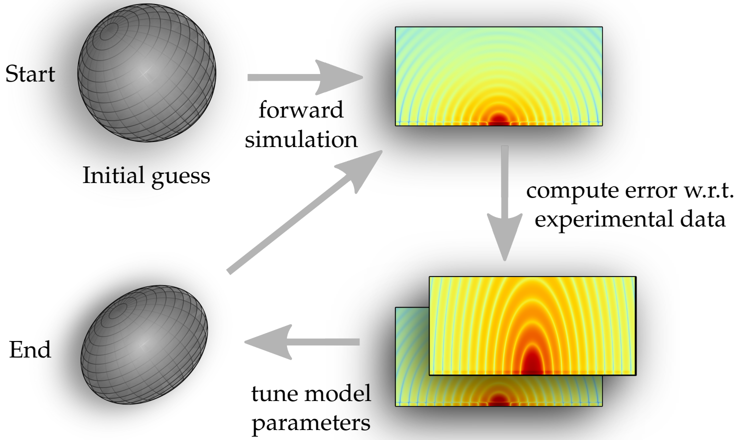
Computational Problems in Structure Recovery: Inverse Modeling



Computational Problems in Structure Recovery: Inverse Modeling



Computational Problems in Structure Recovery: Inverse Modeling



Need for High-Performance Computing

Data generation and analysis gap:

- High measurement rates of current state-of-the-art light beam detectors.
- Wait for days for analyzing data with previous softwares.
- Extremely inefficient utilization of facilities due to mismatch.
- *Example*: 100 MB raw data per second. Up to 12 TB per week.

Need for High-Performance Computing

High computational and accuracy requirements:

- Errors are proportional to the resolutions of various computational discretization.
- Higher resolutions require higher computational power.
- Example:
  - $O(10^7)$ to $O(10^{15})$ kernel computations for one simulation.
  - $O(10^2)$ experiments per material sample.
  - $O(10)$ to $O(10^3)$ forward simulations for inverse modeling per scattering pattern.

Need for High-Performance Computing

Science Gap:

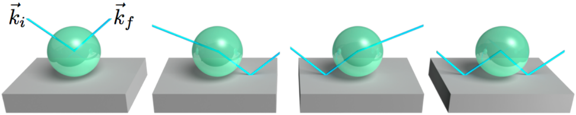
- Beam-line scientists lack access to high-performance algorithms and codes.
- In-house developed codes limited in compute capabilities and performance.
- Also, they are extremely slow – wait for days and weeks to obtain basic results.

Forward Simulations: Computing Scattered Light Intensities

Given:
1. a sample structure model, and
2. experimental configuration,

simulate scattering patterns.

Based on *Distorted Wave Born Approximation* (DWBA) theory.



Inverse Modeling

Forward simulation kernel: computing the scattered light intensities. E.g.

- FFT computations (SAXS)
- Complex form factor and structure factor computations (GISAXS)

Various inverse modeling algorithms:

- Reverse Monte-Carlo simulations for SAXS.
- Sophisticated optimization algorithms for GISAXS.
  - Gradient based: LMVM (Limited-Memory Variable-Metric.)
  - Derivative-free trust region-based: POUNDerS.
  - Stochastic: Particle Swarm Optmization.

Inverse Modeling

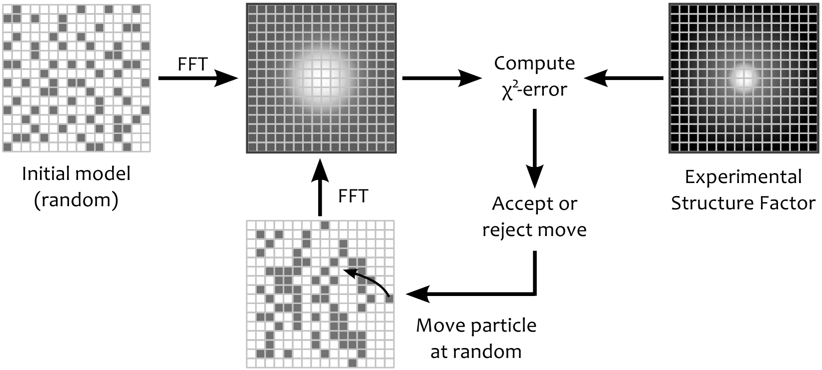
Forward simulation kernel: computing the scattered light intensities. E.g.

- FFT computations (SAXS)
- Complex form factor and structure factor computations (GISAXS)

Various inverse modeling algorithms:

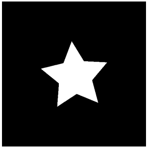
- Reverse Monte-Carlo simulations for SAXS.
- Sophisticated optimization algorithms for GISAXS.
  - Gradient based: LMVM (Limited-Memory Variable-Metric.)
  - Derivative-free trust region-based: POUNDerS.
  - Stochastic: Particle Swarm Optmization.

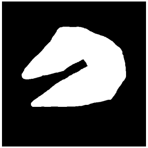
Reverse Monte Carlo Simulations

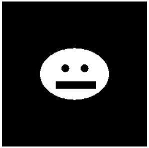


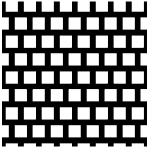
Reverse Monte Carlo Simulations: Validation

Actual Models



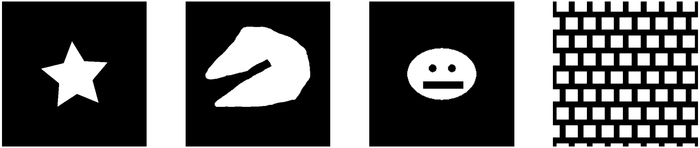




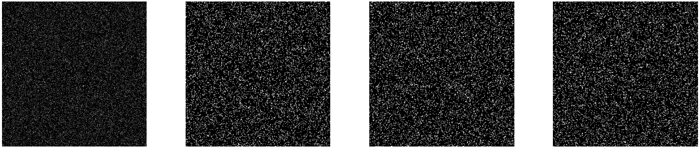


Reverse Monte Carlo Simulations: Validation

Actual Models

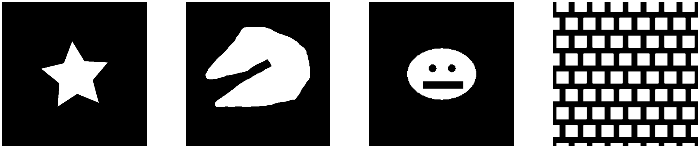


Initial Models

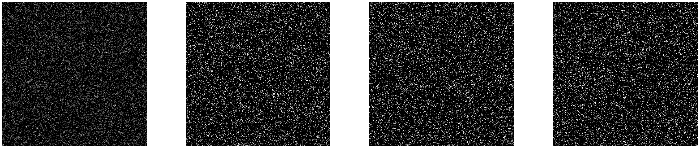


Reverse Monte Carlo Simulations: Validation

Actual Models



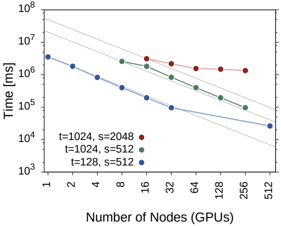
Initial Models

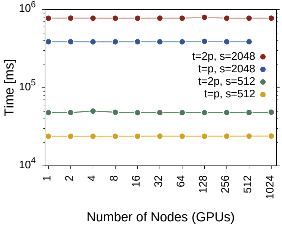


Recovered Models

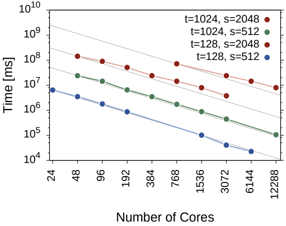
Reverse Monte Carlo Simulations: Strong and Weak Scaling

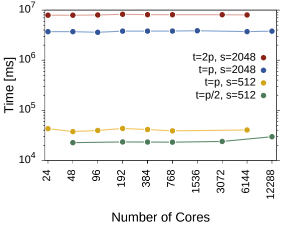
Titan





Hopper

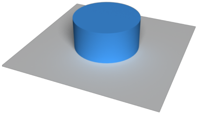




Limited-Memory Variable-Metric and POUNDerS

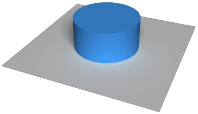
- Methods from the optimization package TAO.
- LMVM is a gradient-based method.
- POUNDerS is a derivative-free trust-region-based method.

Limited-Memory Variable-Metric and POUNDerS: Two Parameter Case

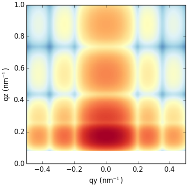


A Single Cylindrical Nanoparticle

Limited-Memory Variable-Metric and POUNDerS: Two Parameter Case

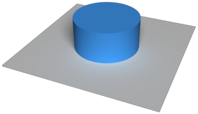


A Single Cylindrical Nanoparticle

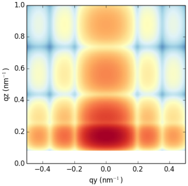


X-Ray Scattering Pattern

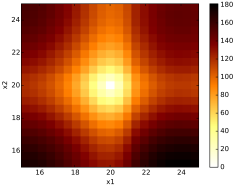
Limited-Memory Variable-Metric and POUNDerS: Two Parameter Case



A Single Cylindrical Nanoparticle

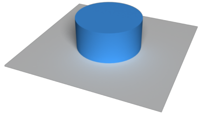


X-Ray Scattering Pattern

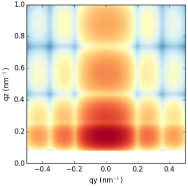


Objective Function Map

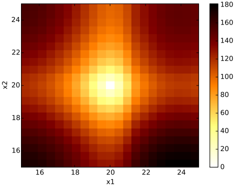
Limited-Memory Variable-Metric and POUNDerS: Two Parameter Case



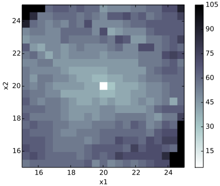
A Single Cylindrical Nanoparticle



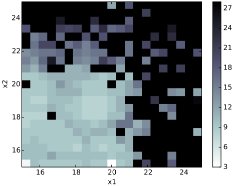
X-Ray Scattering Pattern



Objective Function Map

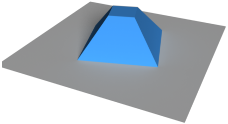


LMVM Convergence Map



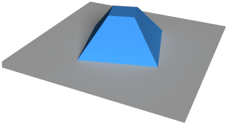
POUNDerS Convergence Map

Limited-Memory Variable-Metric and POUNDerS: Six Parameter Case

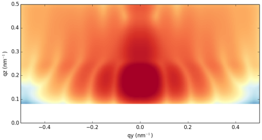


Pyramidal Nanoparticles forming a Lattice

Limited-Memory Variable-Metric and POUNDerS: Six Parameter Case

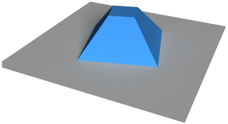


Pyramidal Nanoparticles forming a Lattice

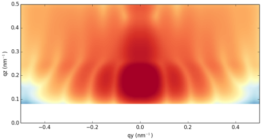


X-Ray Scattering Pattern

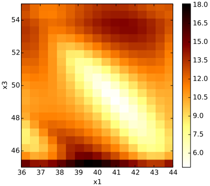
Limited-Memory Variable-Metric and POUNDerS: Six Parameter Case



Pyramidal Nanoparticles forming a Lattice

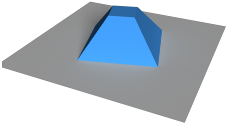


X-Ray Scattering Pattern

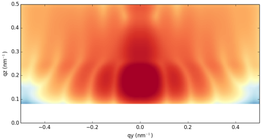


Objective Function Map

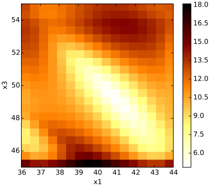
Limited-Memory Variable-Metric and POUNDerS: Six Parameter Case



Pyramidal Nanoparticles forming a Lattice

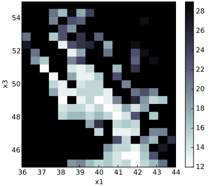


X-Ray Scattering Pattern



Objective Function Map

LMVM does not converge



POUNDerS Convergence Map

Particle Swarm Optimization

- Stochastic method.
- Multiple agents, *"particle swarm"*, search for optimal points in the parameter space.
- Agent velocities influenced by history of traveled paths.

Particle Swarm Optimization

- Stochastic method.
- Multiple agents, *"particle swarm"*, search for optimal points in the parameter space.
- Agent velocities influenced by history of traveled paths.

inertia coefficient

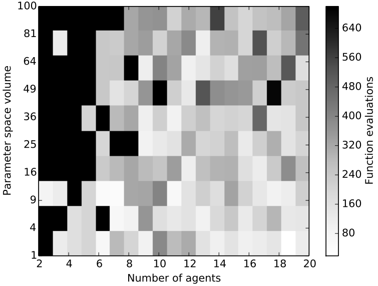
force coefficients

$$\boxed{\vec{v_i} \leftarrow \omega \vec{v_i} + (\vec{b_i} - \vec{x_i})\vec{r_1}\phi_1 + (\vec{b_g} - \vec{x_i})\vec{r_2}\phi_2}$$

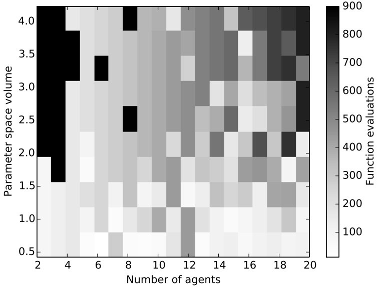
local best position

global best position

Particle Swarm Optimization: Fitting X-Ray Scattering Data

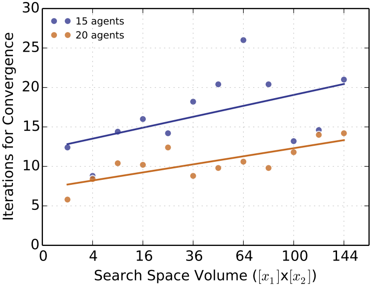


Fitting 2 Parameters

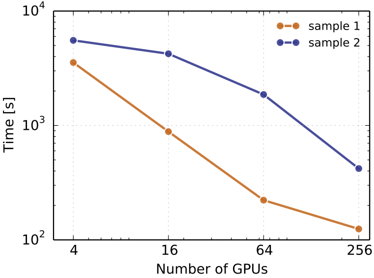


Fitting 6 Parameters

Particle Swarm Optimization: Performance

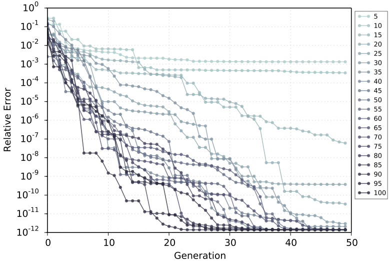


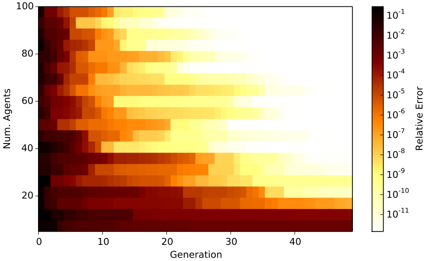
Convergence w.r.t. Search Space Volume



Strong Scaling on Titan

Particle Swarm Optimization: Agents vs. Generations





An Ongoing Work

We saw that:

- GPUs have brought data analysis time from days and weeks to just minutes and seconds.
- Derivative-based methods converge only for simple cases.
- Trust-region-based methods are very sensitive to initial guess.
- PSO does not require an initial guess.
- PSO is robust, nearly always converging, but expensive.
- Need better optimization algorithms.

An Ongoing Work

We saw that:

- GPUs have brought data analysis time from days and weeks to just minutes and seconds.
- Derivative-based methods converge only for simple cases.
- Trust-region-based methods are very sensitive to initial guess.
- PSO does not require an initial guess.
- PSO is robust, nearly always converging, but expensive.
- Need better optimization algorithms.

Near future:

- Opening gates to much more sophisticated analyses.
- We are applying machine learning for feature and structural classification to generate initial models to fit.

Our Current Core Team

- **Alexander Hexemer**, *Advanced Light Source, Berkeley Lab.*
- **Dinesh Kumar**, *Advanced Light Source, Berkeley Lab.*
- **Xiaoye S. Li**, *Computational Research Division, Berkeley Lab.*
- **Abhinav Sarje**, *Computational Research Division, Berkeley Lab.*
- **Singanallur Venkatakrishnan**, *Advanced Light Source, Berkeley Lab.*

Our Current Core Team

- **Alexander Hexemer**, *Advanced Light Source, Berkeley Lab.*
- **Dinesh Kumar**, *Advanced Light Source, Berkeley Lab.*
- **Xiaoye S. Li**, *Computational Research Division, Berkeley Lab.*
- **Abhinav Sarje**, *Computational Research Division, Berkeley Lab.*
- **Singanallur Venkatakrishnan**, *Advanced Light Source, Berkeley Lab.*

And we are open to collaborations ...

Further Information

1. **HipGISAXS: a high-performance computing code for simulating grazing-incidence X-ray scattering data**. Journal of Applied Crystallogrpahy, vol. 46, pp. 1781–1795, 2013.
2. **Massively parallel X-ray scattering simulations**. Supercomputing (SC), 2012.
3. **Tuning HipGISAXS on multi- and many-core supercomputers**. Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems, 2013.
4. **High performance inverse modeling with Reverse Monte Carlo simulations**. International Conference on Parallel Processing, 2014.

Acknowledgments

- Supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.
- Also supported by DoE Early Career Research grant awarded to Alexander Hexemer.
- Used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.
- Used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Thank you!

Web: **http://portal.nersc.gov/project/als/hipgisaxs**

http://www.github.com/hipgisaxs