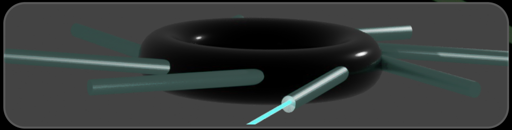
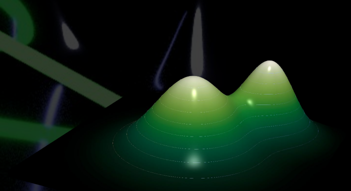
A Perspective on Autonomous Experimentation and Discovery

Marcus Michael Noack

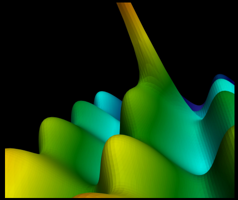


What is Autonomous Data Acquisition

CAMERA Workshop on AE and Main Takeaways

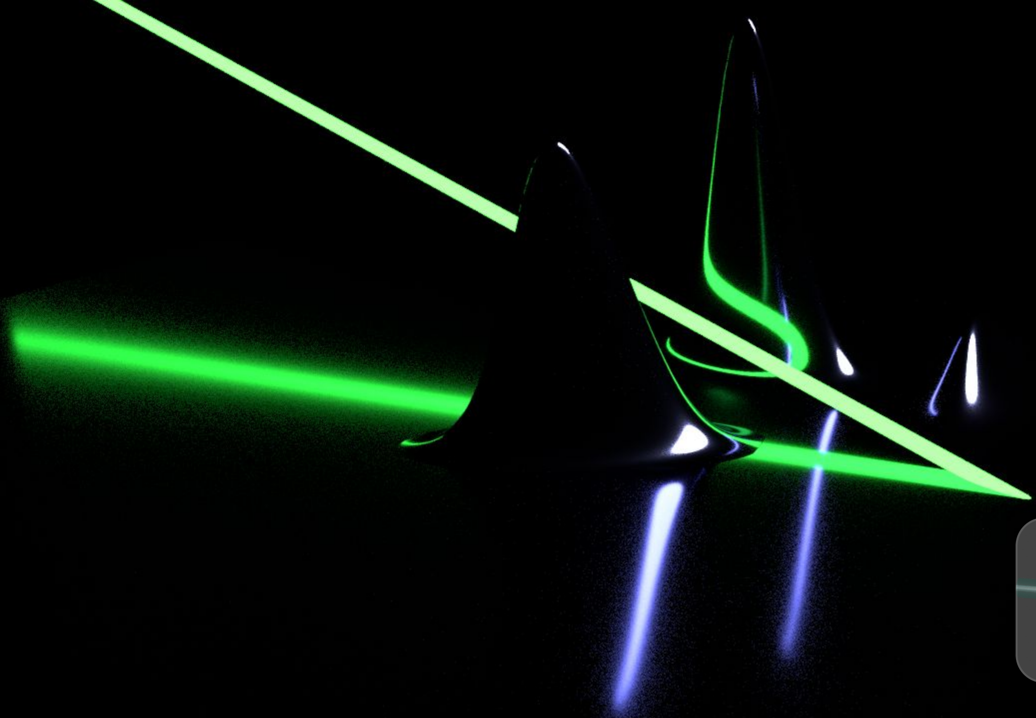


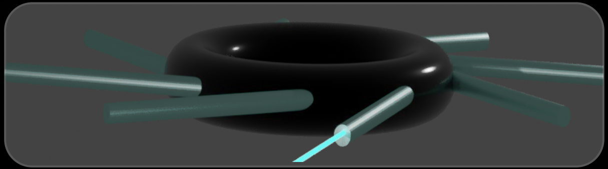
Basic Gaussian-Process-Driven Autonomous Data Acquisition



Mathematical Optimization for AE and ML

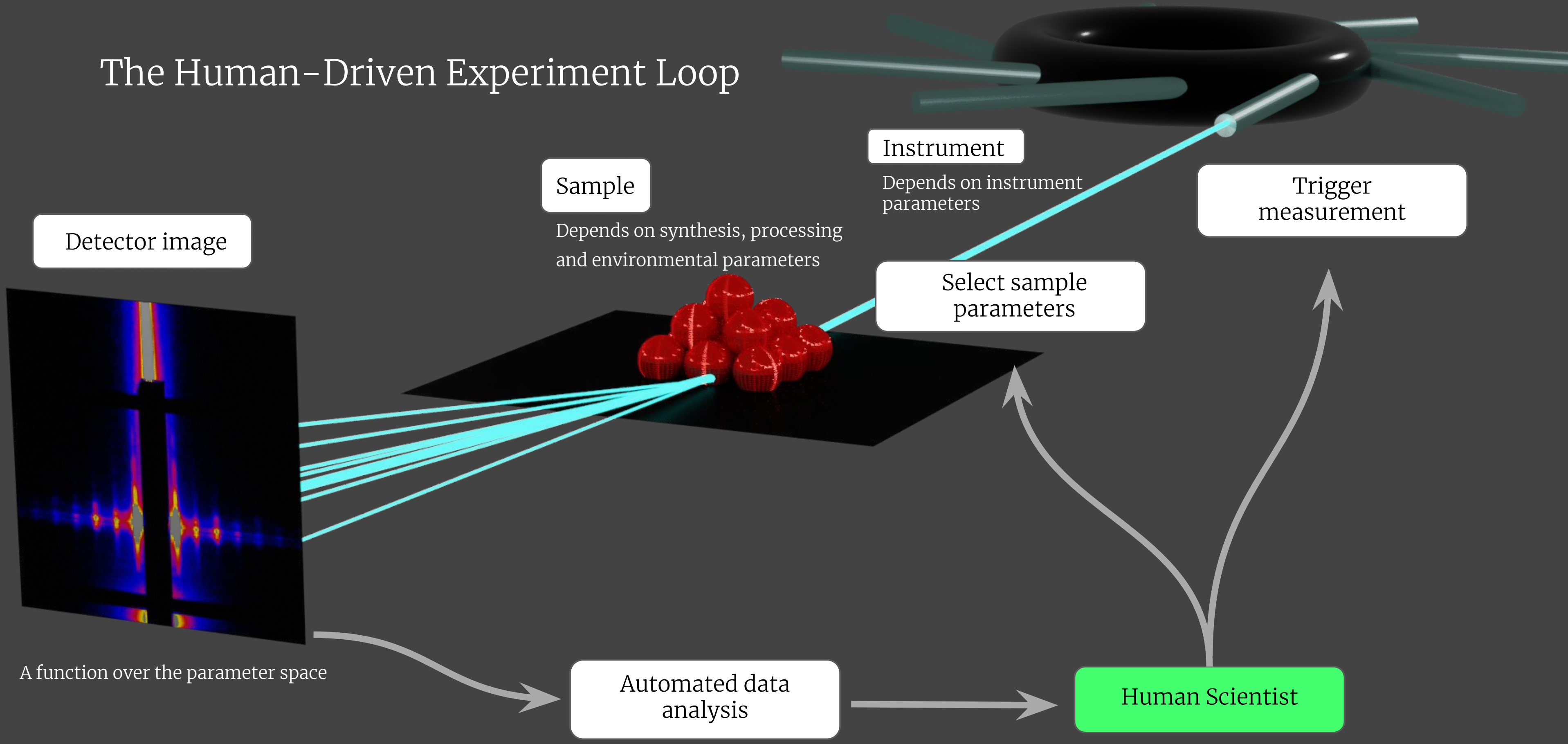
Bringing Autonomous Discovery to the Community





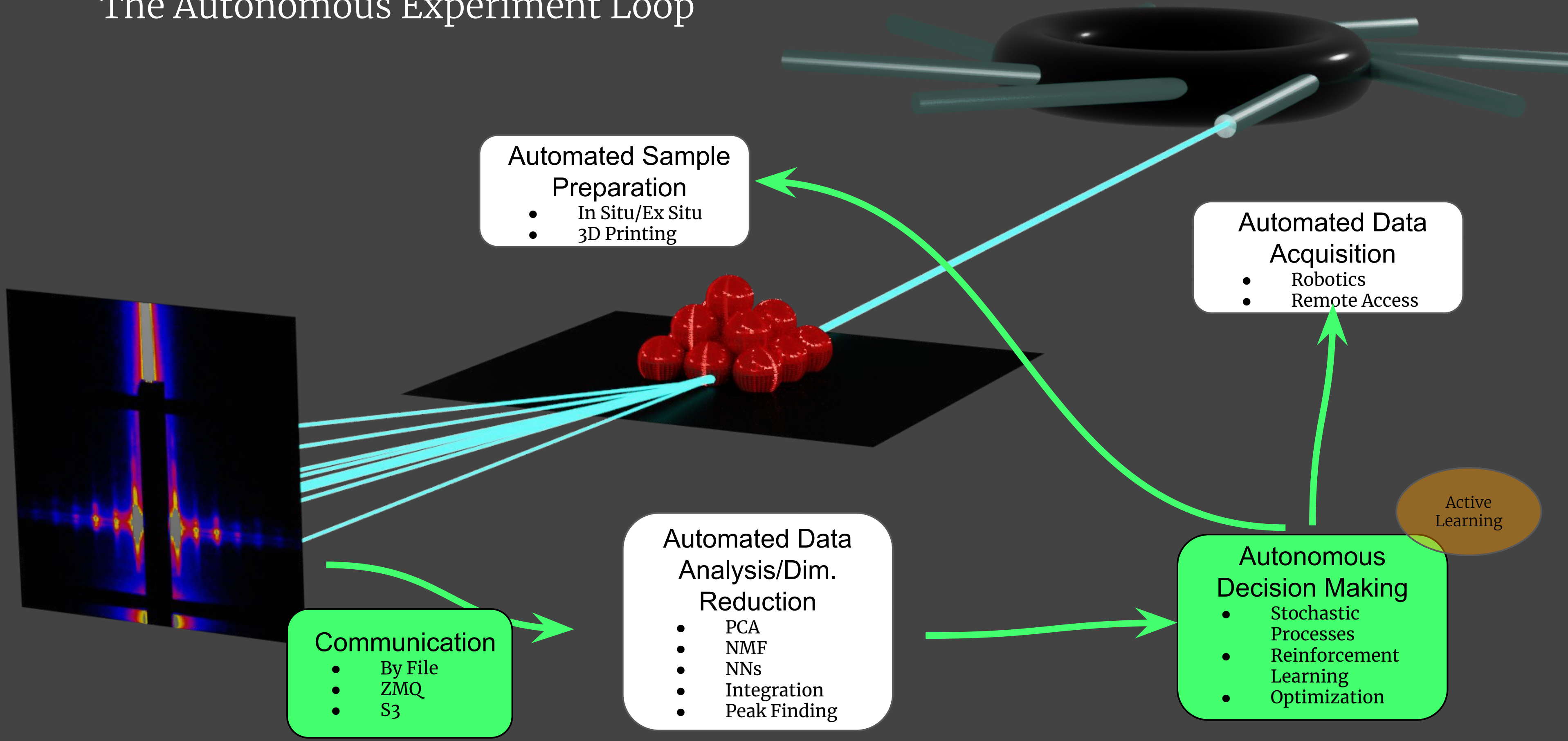
[What is Autonomous Data Acquisition]

The Human-Driven Experiment Loop

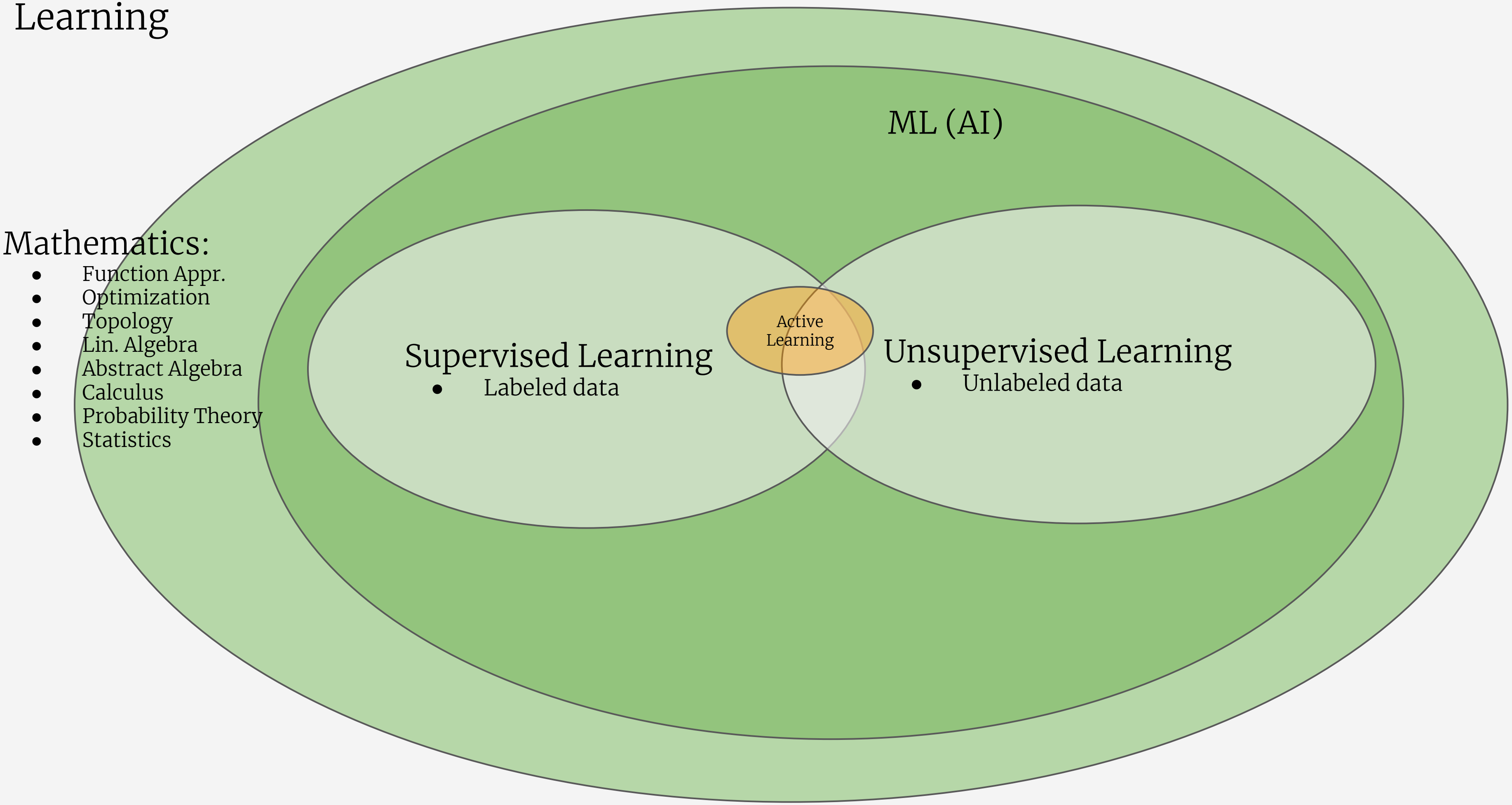


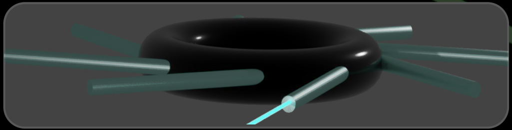
	Random	Scanning	Intuition
Inefficient	Prev. data not used/Non-optimal measurements	Prev. data not used/Non-optimal measurements	Not autonomous/ human attention required
Uninformative	No metric for quality	No metric for quality	No metric for quality
Biased			Relies on past experience Lack of reproducibility

The Autonomous Experiment Loop



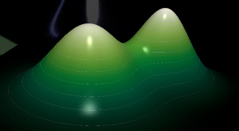
A Venn Diagram for ML and Active Learning



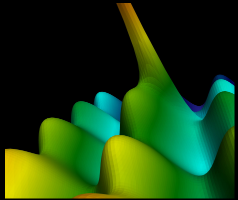


What is Autonomous Data Acquisition

CAMERA Workshop on AE and Main Takeaways

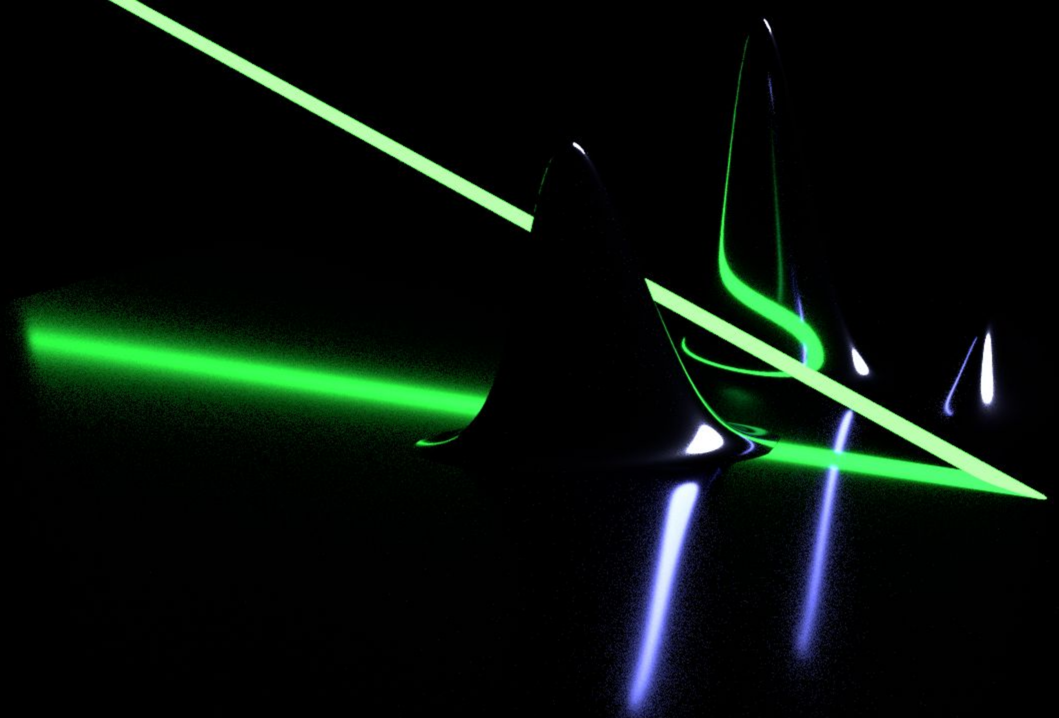


Basic Gaussian-Process-Driven Autonomous Data Acquisition



Mathematical Optimization for AE and ML

Bringing Autonomous Discovery to the Community



[CAMERA Workshop on AE and Main Takeaways]

Autonomous Discovery for Science and Engineering

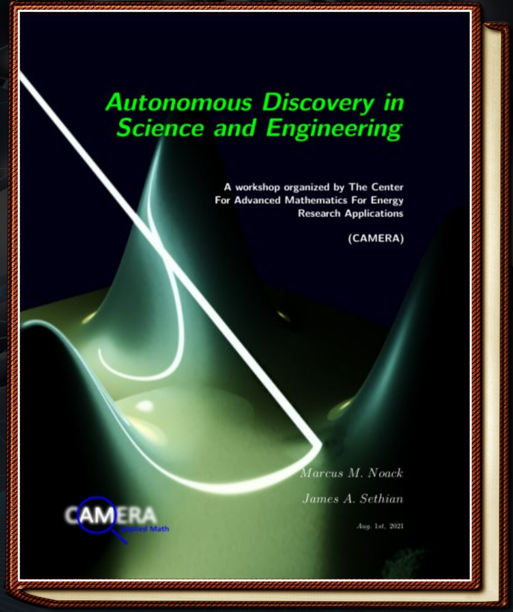
A three-day workshop for sharing recent developments in autonomous methods, sponsored by CAMERA —
The Center for Advanced Mathematics for Energy Research Applications

Dates: April 20th - 22nd, 2021

Methods and Algorithms: Gaussian Processes, Neural Networks, Reinforcement Learning, New Math, Optimization, Data Analytics and Infrastructure, UQ

Tutorials: gpCAM, Summit, CamLink, Escalate, Bluesky, Atinary SDLabs, DataFed, Cameo, AtomAI, ART

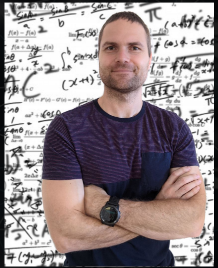
Applications: Microscopy, Spectroscopy, X-ray Scattering, Neutron Scattering, Autonomous Synthesis & Materials Discovery, Robotics & Remote Access



https://autonomous-discovery.lbl.gov/

DOI: https://doi.org/10.2172/1818491

The Organizing Committee



Marcus Noack, LBNL

Petrus Zwart, LBNL

Apurva Mehta, SLAC

Daniela Ushizima, LBNL

James Sethian, UCB

Kevin Yager, BNL

Simon Billinge, Columbia U.

Héctor García Martín, LBNL

Nicholas Schwarz, ANL

B. Reeja Jayan, CMU

Eva M. Herzig, Bayreuth U.

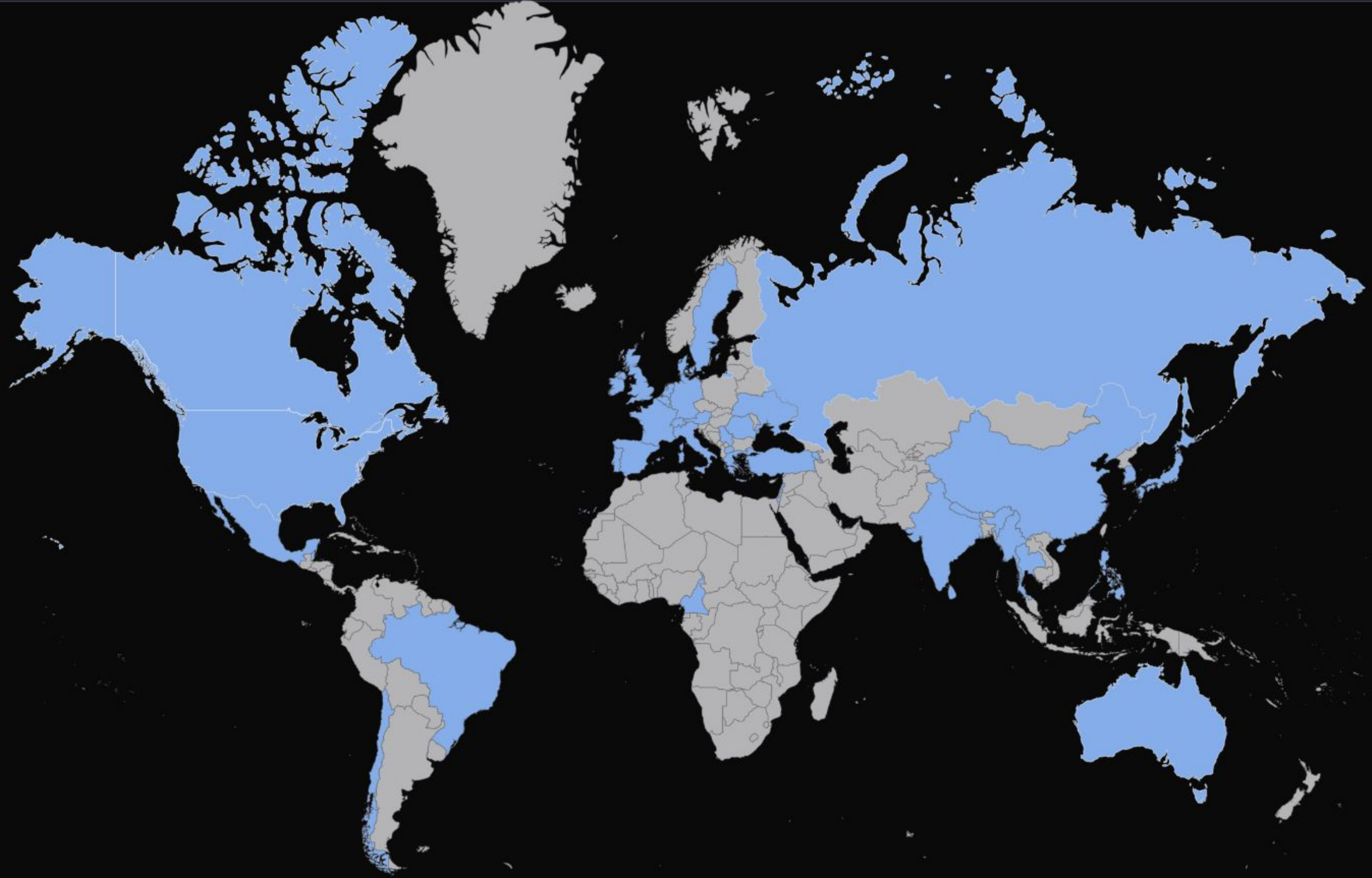
Alex Hexemer, LBNL

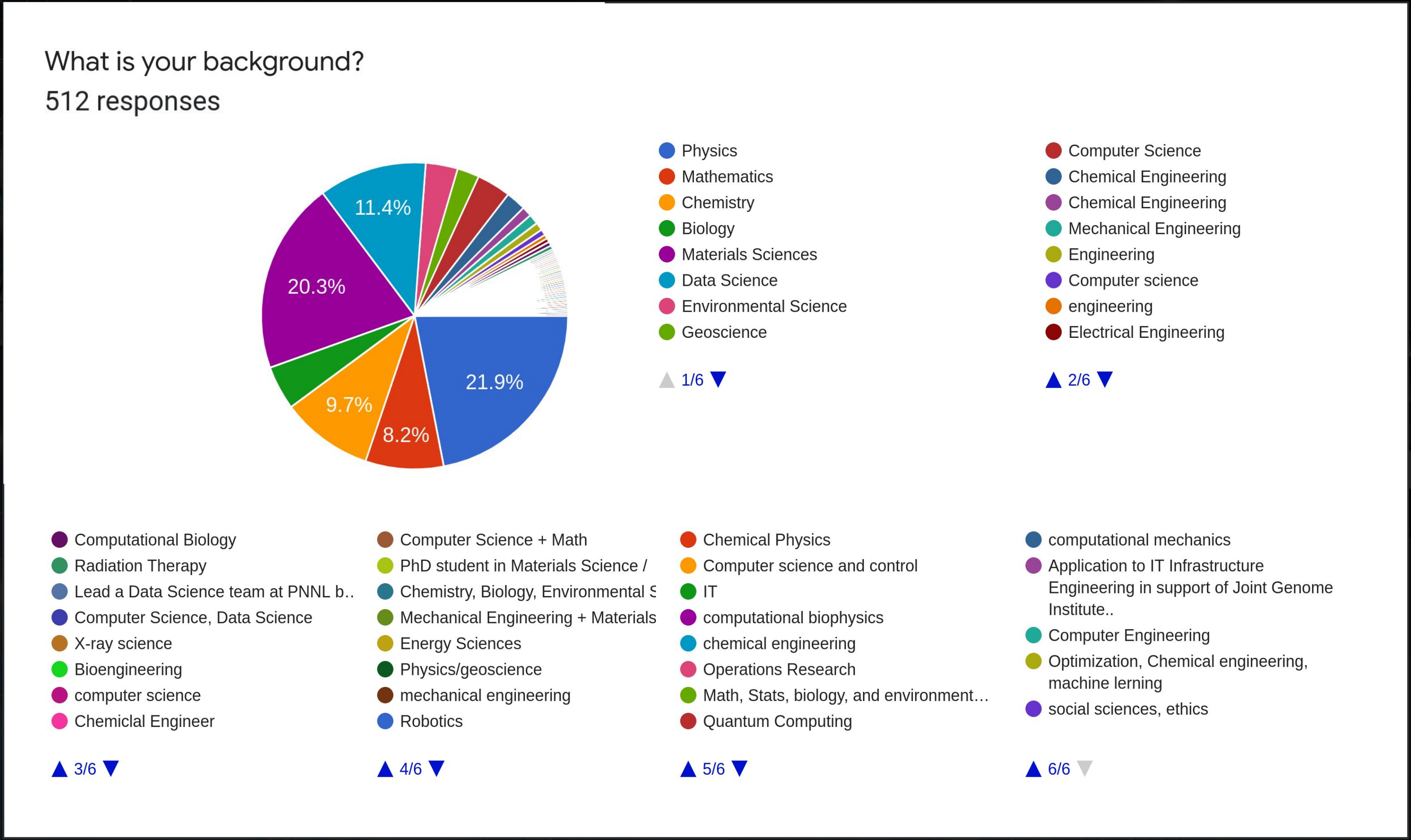
Bobby Sumpter, ORNL

Martin Boehm, ILL

Sergei Kalinin, ORNL

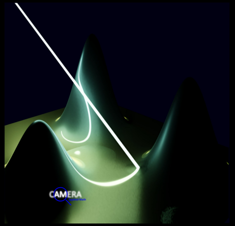
Aaron Gilad Kusne, NIST

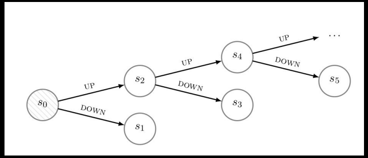



What is your background?
512 responses
11.4%
20.3%
21.9%
9.7%
8.2%
Physics
Mathematics
Chemistry
Biology
Materials Sciences
Data Science
Environmental Science
Geoscience
1/6
Computer Science
Chemical Engineering
Chemical Engineering
Mechanical Engineering
Engineering
Computer science
engineering
Electrical Engineering
2/6
Computational Biology
Radiation Therapy
Lead a Data Science team at PNNL b..
Computer Science, Data Science
X-ray science
Bioengineering
computer science
Chemiclal Engineer
3/6
Computer Science + Math
PhD student in Materials Science /
Chemistry, Biology, Environmental S
Mechanical Engineering + Materials
Energy Sciences
Physics/geoscience
mechanical engineering
Robotics
4/6
Chemical Physics
Computer science and control
IT
computational biophysics
chemical engineering
Operations Research
Math, Stats, biology, and environment…
Quantum Computing
5/6
computational mechanics
Application to IT Infrastructure Engineering in support of Joint Genome Institute..
Computer Engineering
Optimization, Chemical engineering, machine lerning
social sciences, ethics
6/6

Takeaways and Achievements: Methods/Algorithms

Modelling



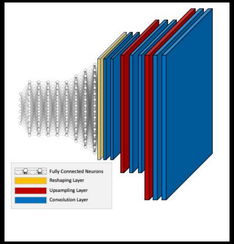


Gaussian and other stochastic processes

Reinforcement Learning

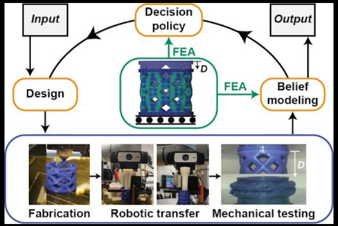
Take-Home Message 1: Gaussian processes and reinforcement learning are the most popular techniques for control, in different data regimes.

Analysis



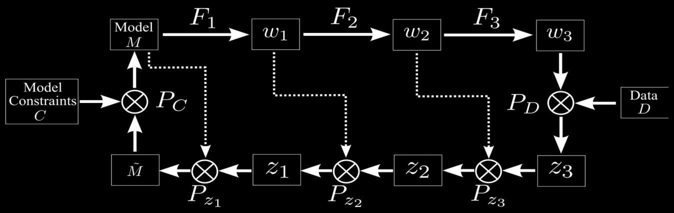
Take-Home Message 2: The data-analysis step is more and more often done automatically data-science tools (PCA, NN, Clustering,...).

Hardware/Robotics



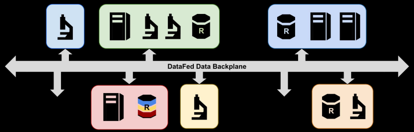
Take-Home Message 3: Instrument systems are increasingly built with automation in mind.

Training and Decision-Making



Take-Home Message 4: Efficient mathematical optimization under uncertainty and subject to constraints has come far but remains a challenge.

Communication Infrastructure



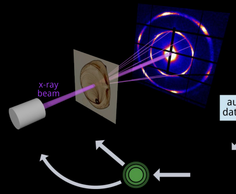
Take-Home Message 5: Data-management systems are emerging using lower-level tool (control, optimization, ...) to allow for standardization.

Takeaways and Achievements: Application

X-Ray Scattering

Autonomous Analytics and Control in X-ray Scattering

Kevin G. Yager
Center for Functional Nanomaterials, Brookhaven National Laboratory



Towards Understanding Solid-State Thin Film Dealloying Phenomena using Autonomous Synchrotron X-ray Characterization

Karen Chen-Wiegart[1,2,*]
Chonghang Zhao[1] Marcus Noack[4] Cheng-Chu Chung[1] Kedar Manandhare[2] Joshua Lynch[2] Hui Zhong[6] Ming Lu[3] Mingzhao Liu[3] Jianming Bai [2] Phillip Maffettone[2] Daniel Olds[2] Masafumi Fukuto[2] Ichiro Takeuchi[5] Sanjit Ghose[2] Thomas Caswell[2] Kevin Yager[3]
[1] Department of Materials Science and Chemical Engineering, Stony Brook University
[2] National Synchrotron Light Source – II, Brookhaven National Laboratory
[3] Center for Functional Nanomaterials, Brookhaven National Laboratory
[4] CAMERA, Computational Research Division, Lawrence Berkeley National Laboratory
[5] Department of Materials Science and Engineering, University of Maryland
[6] Department of Joint Photon Science Institute, Stony Brook University
* Karen.Chen-Wiegart@stonybrook.edu

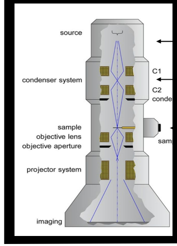
Microscopy

Progress Toward Rapid, Statistical Scanning Transmission Electron Microscopy

Steven R. Spurgeon[1,*], Matthew Olszta[1], Sarah Akers[2], Derek Hopkins[3], Kevin Fiedler[4]
[1] Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, [2] National Security Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, [3] Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352, [4] College of Arts and Sciences, Washington State University–Tri-Cities, Richland, WA 99352
* steven.spurgeon@pnnl.gov

SEM Image Denoising with Unsupervised Machine Learning for Better Defect Inspection and Metrology

Bappaditya Dey[1,2,*], Sandip Halder[1], Kasem Khalil[2], Gian Lorusso[1], Joren Severi[1], Philippe Leray[1], Magdy A. Bayoumi[2,3]
[1] imec, Kapeldreef 75, 3001 Leuven, Belgium, [2] The Center for Advanced Computer Studies, University of Louisiana at Lafayette, Louisiana, USA, [3] Department of Electrical and Computer Engineering, University of Louisiana at Lafayette, Louisiana, USA
* Bappaditya.Dey@imec.be



Machine-Driven Applications in Scanning Probe Microscopy at the Atomic Scale

John C. Thomas[1,*], Antonio Rossi[1], Darian Smalley[2], Zhuohang Yu[3,4], Tianyi Zhao[3,4], Mauricio Terrones[3,4,5], Marcus Noack[6], Masahiro Ishigami[2], Burno Schuler[1], Edward Barnard[1], Ed Wong[1], Alex Weber-Bargioni[1]
[1] Molecular Foundry, Lawrence Berkeley National Laboratory, California 94720, United States of America, [2] Department of Physics, University of Central Florida, Orlando, FL 32816, United States of America, [3] Department of Materials Science and Engineering, The Pennsylvania State University, University Park, PA 16082 United States of America, [4] Center for Two-Dimensional and Layered Materials, The Pennsylvania State University, University Park, PA, 16802 United States of America, [5] Department of Physics and Department of Chemistry, The Pennsylvania State University, University Park, PA, 16802 United States of America, [6] Center for Advanced Mathematics for Energy and Research Applications, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States of America
* jthomas@lbl.gov

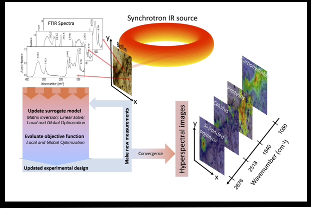
Spectroscopy

Autonomous Adaptive Data Acquisition for Scanning Hyperspectral Imaging

Elizabeth A. Holman
Chemistry and Chemical Engineering Division, California Institute of Technology, Pasadena, CA 91125
eholman@caltech.edu

Towards High-Throughput Autonomous Infrared Spectromicroscopy

Petrus H. Zwart[1,2,3,*], Liang Chen[1,3], Steven Lee[1,3,5], Patricia Valdespino Castillo[1,3], Marcus Noack[2,4], Hoi-Ying N. Holman[1,3]
[1] Molecular biophysics and Integrated Bioimaging Division, Lawrence Berkeley Laboratory, [2] Center for Advanced Mathematics for Energy Research Applications, Lawrence Berkeley Laboratory, [3] Berkeley Synchrotron Infrared Structural Biology Program, Lawrence Berkeley Laboratory, [4] Computational Research Division, Lawrence Berkeley Laboratory, [5] Department of Physics, University of California, Berkeley
* PHZwart@lbl.gov



Materials Science

Autonomous Research Systems for Materials Development

Benji Maruyama[1,*]
[1] Air Force Research Laboratory
* benji.maruyama@us.af.mil

Problem-Fluent Methods for Complex Decision Making in Autonomous Materials Research

Kristofer G. Reyes
Department of Materials Design and Innovation University at Buffalo, Computational Science Initiative Brookhaven National Laboratory
kreyes3@buffalo.edu

Accelerating Semiconductor Research through Robotic Automation and ML-Guided Experimentation

Aram Amassian
Materials Science and Engineering, Organic and Carbon Electronics Laboratories (ORaCEL), North Carolina State University, Raleigh, NC 27695
* aamassi@ncsu.edu

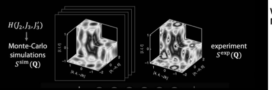
Autonomous Experimentation for Structural Design and Additive Manufacturing

Aldair E. Gongora[1], Kelsey L. Snapp[1], Emily Whiting[2], Patrick Riley[3], Kristofer G. Reyes[4,*], Elise F. Morgan[1,5,6,*], Keith A. Brown[1,6,7,8,*]
[1] Department of Mechanical Engineering, Boston University, Boston, MA 02215, USA, [2] Department of Computer Science, Boston University, Boston, MA 02215, USA, [3] Google Research, Mountain View, CA 94043, USA, [4] Department of Materials Design and Innovation, University at Buffalo, Buffalo, NY 14260, USA, [5] Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA, [6] Division of Materials Science & Engineering, Boston University, Boston, MA 02215, USA, [7] Physics Department, Boston University, Boston, MA 02215, USA
* kreyes3@buffalo.edu (K.G.R.), efmorgan@bu.edu (E.F.M.), brownka@bu.edu (K.A.B.)

Neutron Scattering

Improving the Measurement Strategy in Neutron Spectroscopy with Machine Learning

Martin Boehm[1,*], Tobias Weber[1], Yannick LeGoc[1], Marcus Noack[2], Paolo Mutti[1]
[1] Institut Laue-Langevin, 71 avenue des Martyrs, 38042 Grenoble, France, [2] Center for Advanced Mathematics for Energy Research Applications (CAMERA) and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, USA
* boehm@ill.fr



What is Involved in Using Machine Learning to Accelerate Understanding of Neutron Experiments?

Alan Tennant
[1] Neutron Science Division, Oak Ridge National Laboratory, Oak Ridge, TN 37830
[2] Shull Wollan Center, Oak Ridge National Laboratory, Oak Ridge, TN 37830
* tennantda@ornl.gov

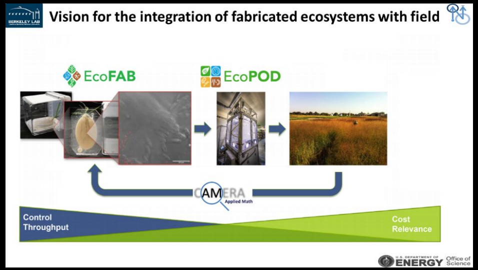
Interpretability and uncertainty quantification for machine learning analysis of inelastic neutron scattering

Keith T. Butler[1,3,*], Manh Duc Le[2], Jeyan Thiyagalingam[1,4], Toby G. Perring[2]
[1] SciML, Scientific Computing Department, STFC Rutherford Appleton Laboratory, Harwell Campus, Didcot, OX11 0QX, UK, [2] ISIS Neutron and Muon Source, STFC Rutherford Appleton Laboratory, Harwell Campus, Didcot, OX11 0QX, UK, [3] Department of Chemistry, University of Reading, Whiteknights, Reading RG6 6DX, UK, [4] Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ
* keith.butler@stfc.ac.uk

Biology

Developing Fabricated Ecosystems to Harness Plant-Microbe Interactions

Peter F. Andeer[1], Trent Northen[1,2,*]
[1] CEnvironmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 947202, [2] Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 947202
* TRNorthen@lbl.gov



Challenges

The Role of Co-Design: Much of the work is performed through co-design teams, bringing together needed expertise. The work has aspects of theory, modeling, algorithm design, data analysis, workflow, and software engineering.

Integrating Across Required Expertise: Teams (or in some cases, individuals) working in autonomous design often take on all the required roles, which requires a large breadth of expertise: it is challenging for a team to excel in all the necessary aspects.

Sharing Developments: There are significant opportunities to share advances across autonomous efforts. However, there is often inconsistent nomenclature and problem formulation.

Workflow and Infrastructure: For the most part, individual efforts center around homegrown workflows and infrastructures. There are opportunities to build work tools and infrastructures that can be shared.

Software: Understandably, many of the efforts described within are aimed at solving a particular set of scientific problems, and the emphasis is not on generalizable software. Openly-available software that is well-documented and properly maintained would be a step forward.

Co-Design

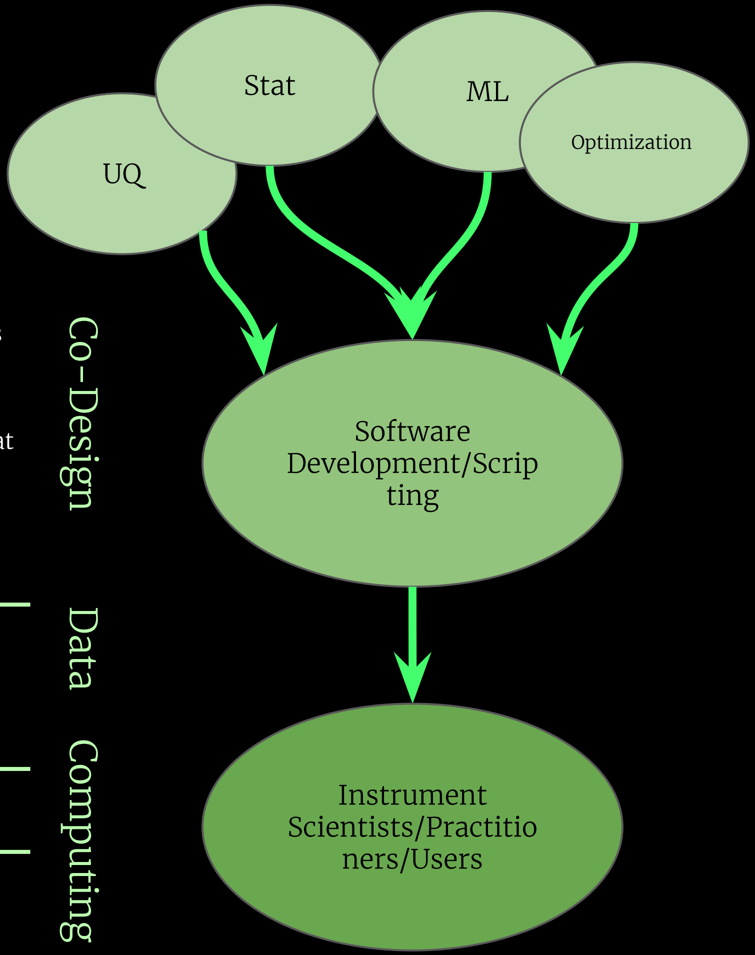
Shared Testbeds and Reproducible Research: It is challenging to cross-test different algorithms and methodologies with common accessible (FAIR) datasets with maintained standards.

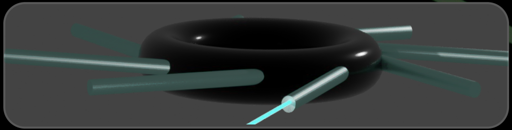
Data: Data != data. The structure of data has to be discussed before data taking and ML applications.

Data

Compute Resources: The field requires a new kind of compute-resource allocation, which keeps resources available throughout the experiment.

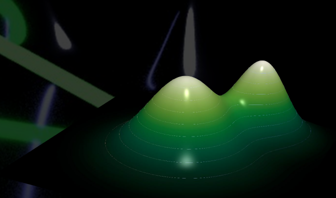
Computing



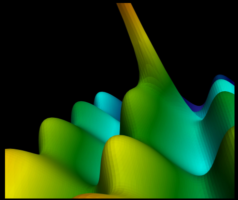


[What is Autonomous Data Acquisition]

[CAMERA Workshop on AE and Main Takeaways]

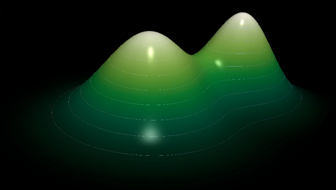


[Basic Gaussian-Process-Driven Autonomous Data Acquisition]



[Mathematical Optimization for AE and ML]

[Bringing Autonomous Discovery to the Community]

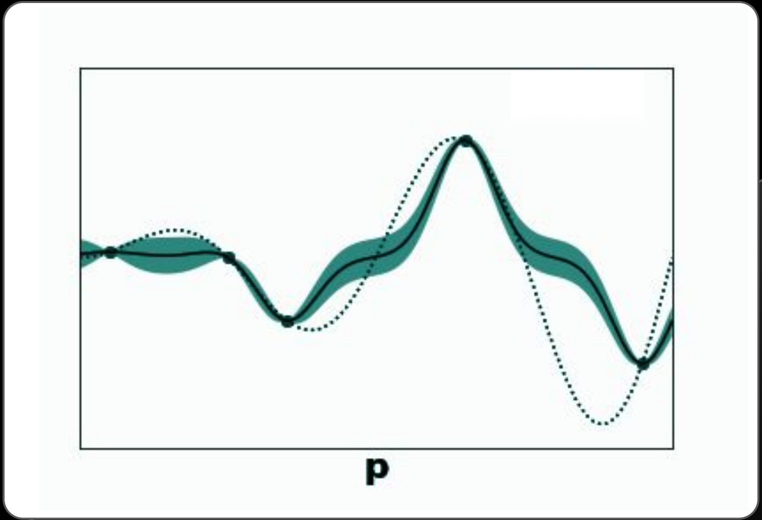


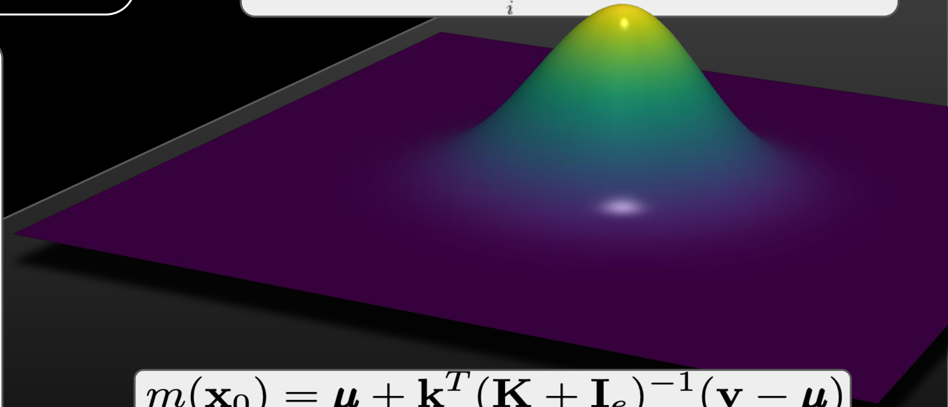
Basic Gaussian-Process-Driven Autonomous Data Acquisition

Gaussian Process Regression in a Nutshell

$$p(\mathbf{f}) \; = \; \frac{1}{2\pi^{d/2}}|\Sigma|^{-1/2} \; \exp[-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{f}-\boldsymbol{\mu})]$$

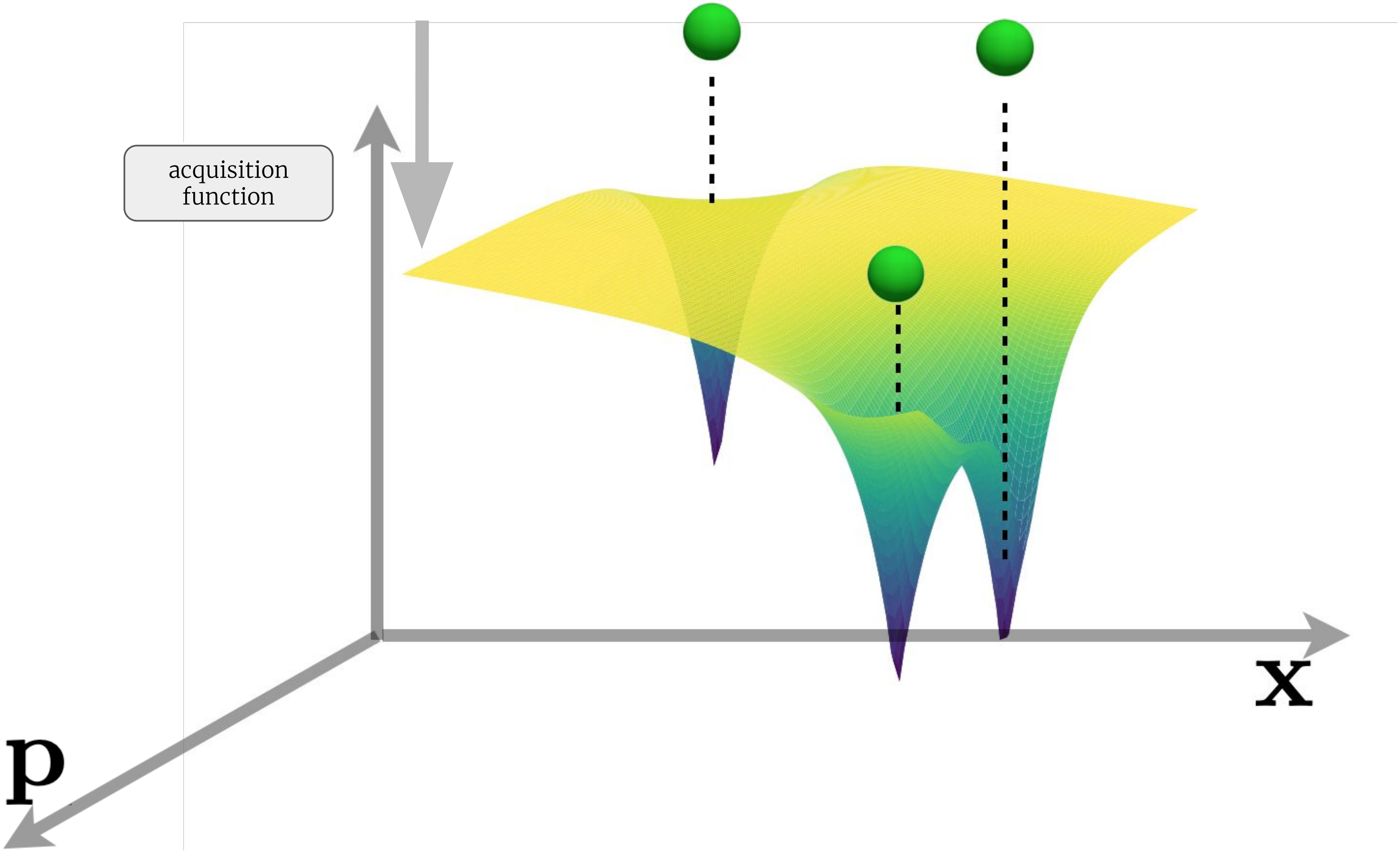
$$\mathcal{H} = \{f(\mathbf{x}) : f(\mathbf{x}) = \sum_{i}^{N} \alpha_i k(\mathbf{x}_i, \mathbf{x}), \forall \boldsymbol{\alpha} \in \mathcal{R}^N, \mathbf{x} \in \mathcal{R}^n\}$$



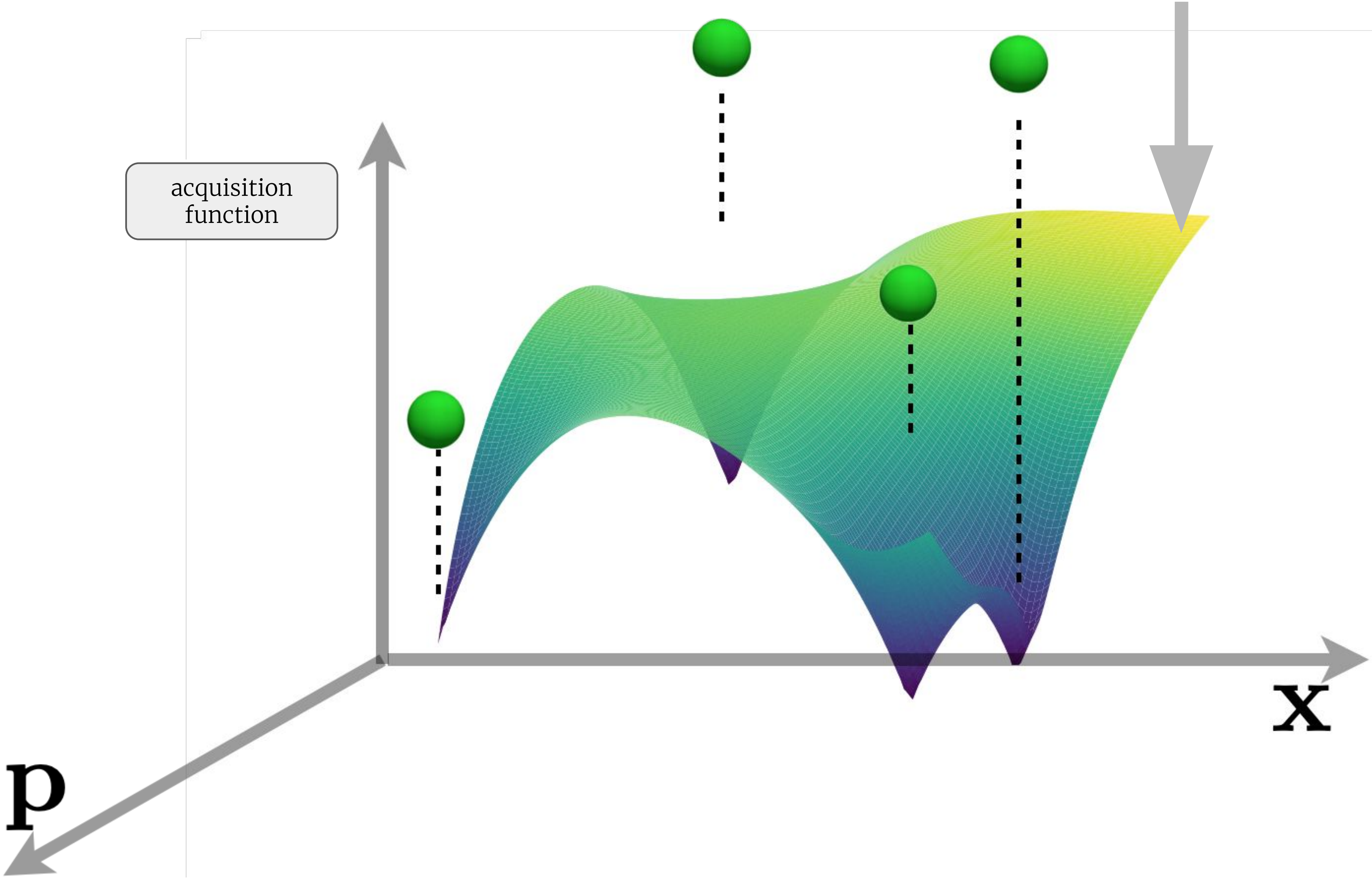


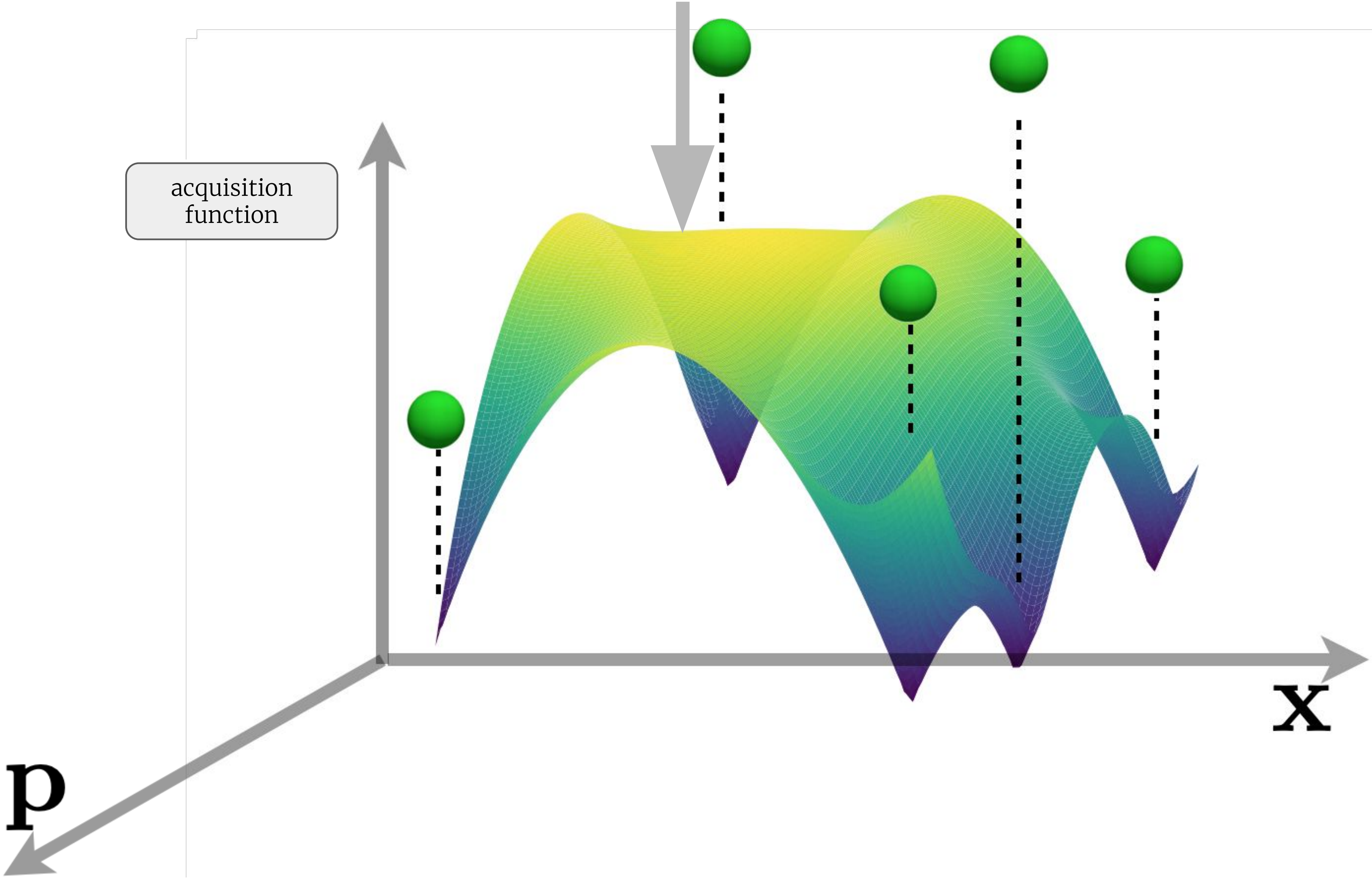
$$m(\mathbf{x}_0) = \boldsymbol{\mu} + \mathbf{k}^T(\mathbf{K} + \mathbf{I}_e)^{-1}(\mathbf{y} - \boldsymbol{\mu})$$

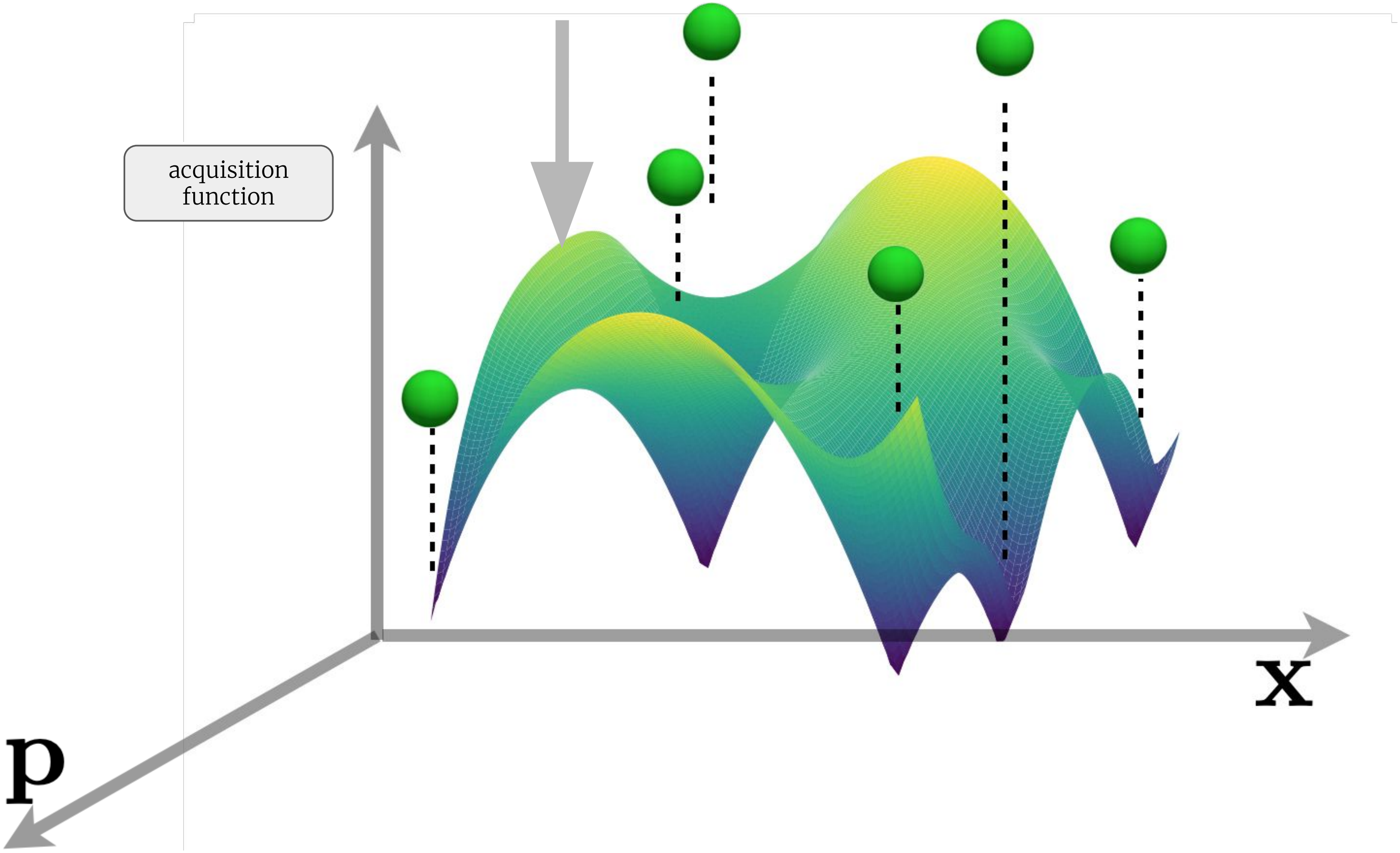
$$\sigma^2(\mathbf{x}_0) = k(\mathbf{x_0}, \mathbf{x_0}) - \mathbf{k}^T(\mathbf{K} + \mathbf{I}_e)^{-1}\mathbf{k}$$

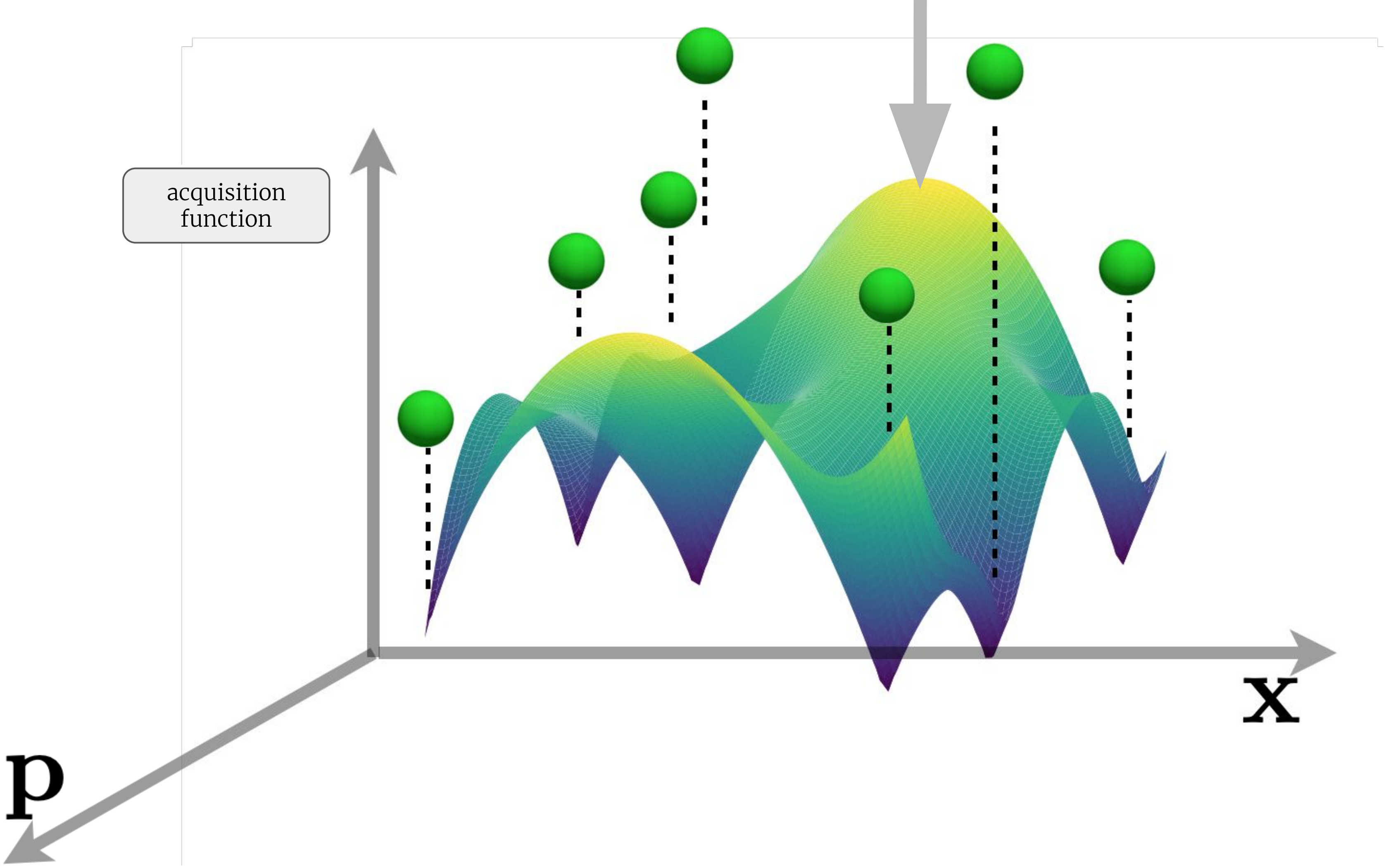


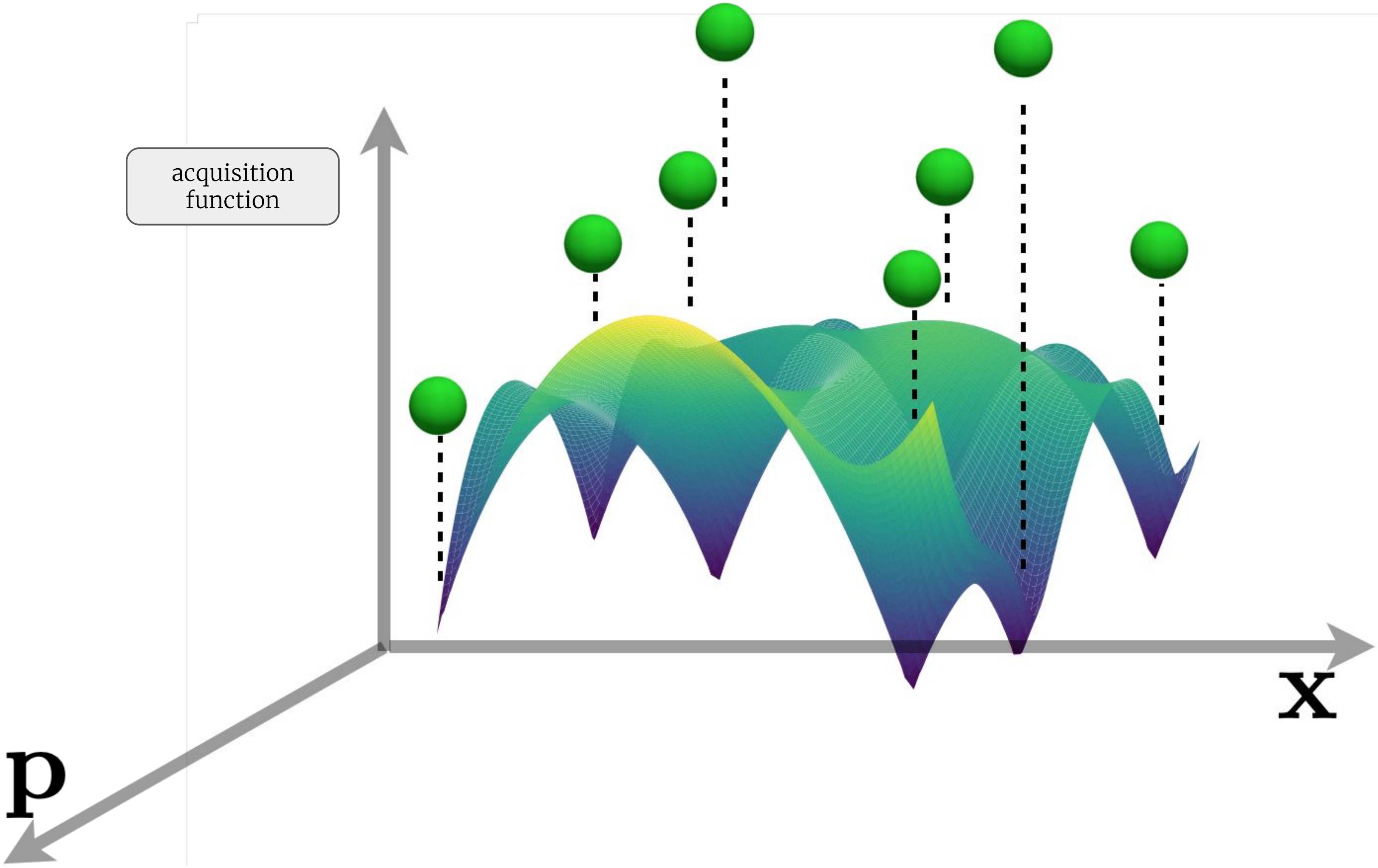
acquisition function
x
p


acquisition function
x
p


acquisition
function
x
p

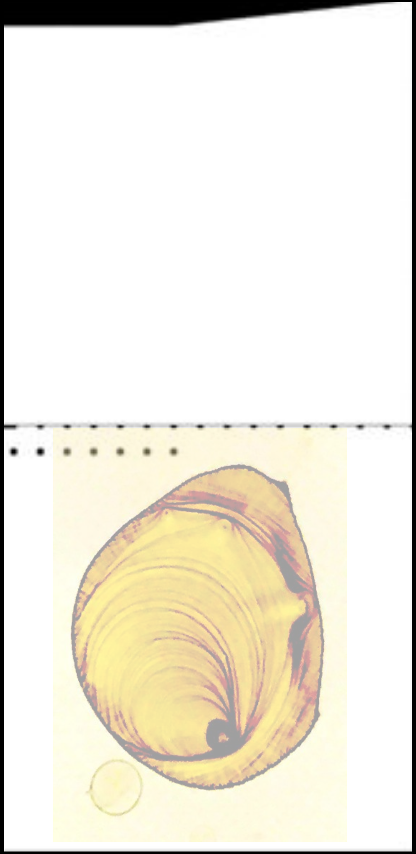

acquisition function
x
p


acquisition
function
x
p

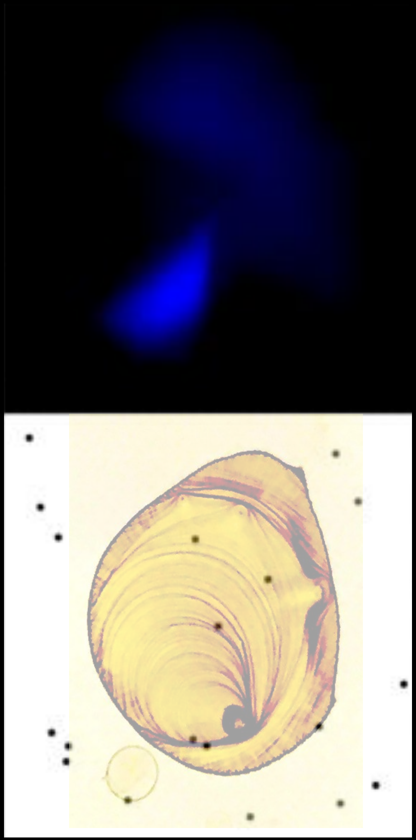

acquisition function
x
p

Our Very First Experiment: A Nanoparticle Stain Mapping Experiment

Facility: NSLS2, CFN @ BNL | Technique: SAXS | Achievement: Commissioning experiment



A Kriging-Based Approach to Autonomous Experimentation with Applications to X-Ray Scattering Marcus M. Noack, Kevin G. Yager, Masafumi Fukuto, Gregory S. Doerk, Ruipeng Li & James A. Sethian

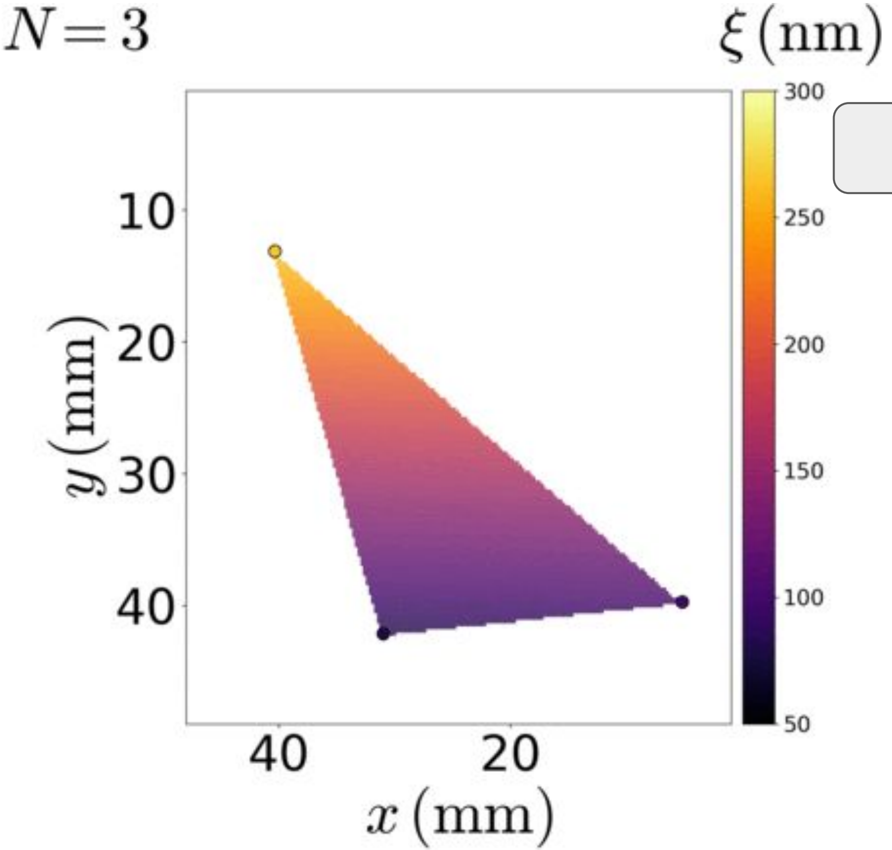


NSLS-II National Synchrotron Light Source II

Center for Functional Nanomaterials
Brookhaven National Laboratory

Autonomous SAXS Exploration of Nanoscale Ordering in a Blade-Coated Polymer-Grafted Nanorod Film

Facility: AFRL and NSLS II | Technique: SAXS | Achievement: 15% of data required, higher resolution in areas of interest

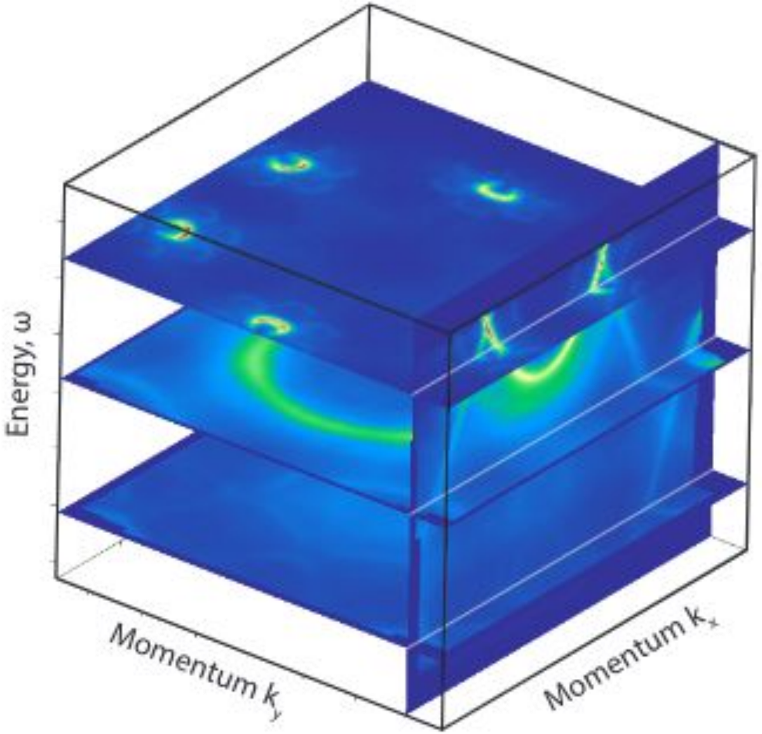


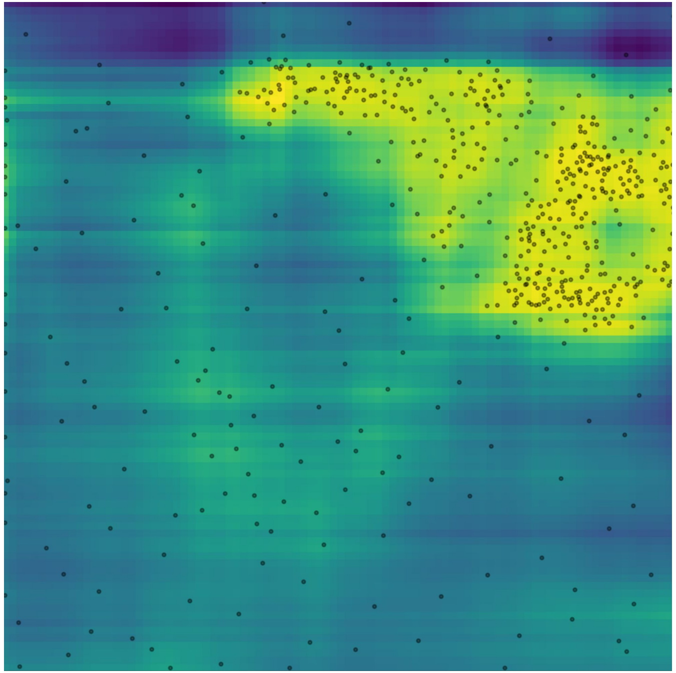
Grain Size

J. Streit, R. Vaia (AFRL), M. Fukuto, R. Li (BNL/NSLS-II), K. Yager (BNL/CFN), M. Noack (LBNL/CAMERA)

Autonomous Steering of ARPES Data Acquisition

Facilities: ALS @ LBNL | Technique: ARPES | Achievement: 12% of data required





K-Means-Driven Gaussian Process Data Collection for Angle-Resolved Photoemission Spectroscopy

Charles N. Melton, Marcus M. Noack, Taisuke Ohta, Thomas E. Beechem, Jeremy Robinson,Xiaotian Zhang, Aaron Bostwick, Chris Jozwiak, Roland J. Koch, Petrus H.
Zwart, Alexander Hexemer, and Eli Rotenberg

Autonomous Control of Synchrotron Infrared Spectroscopy

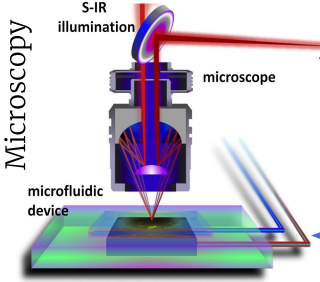
Facility: ALS @ LBNL | Technique: IR Spec. Micr. | Achievement: ~5% of data required, collected in ~10% of the time, materials targeted

ALS ADVANCED LIGHT SOURCE

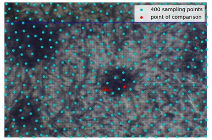
BSISB Berkeley Synchrotron Infrared Structural Biology Imaging Program

Hoi-Ying Holman, Petrus Zwart, Liang Chen, Steven Lee

Infrared Spectral Microscopy



GPR-based Surrogate Model



$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\mu\\ \mu_*\end{bmatrix}, \begin{bmatrix} K(X,X)+\sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*)\end{bmatrix}\right)$$

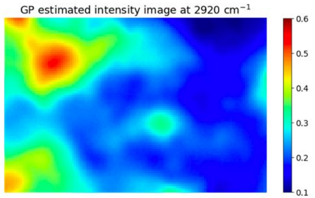
Adaptive Experimental Design

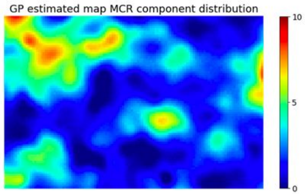
Spectral Map

MCR Component Map

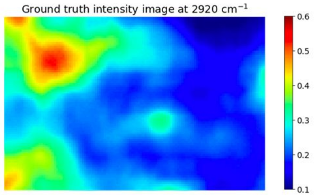
Estimated Spectral Accuracy

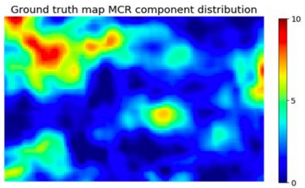
1 hour
400 sample points

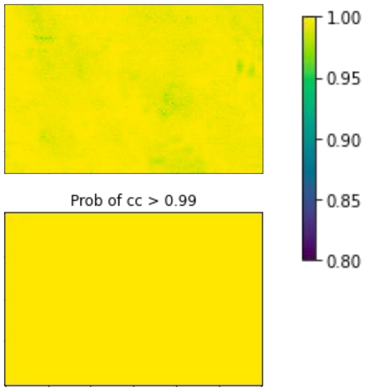


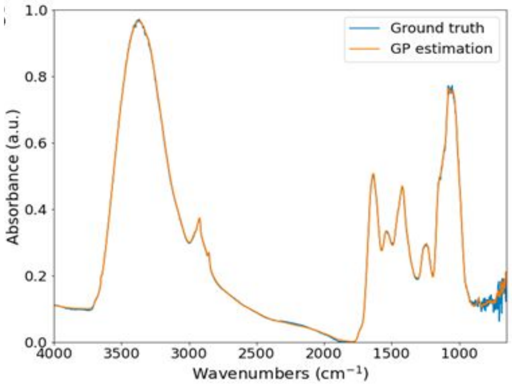


9 hours
9.8K sample points





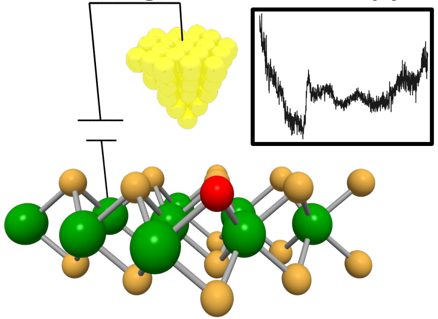




Autonomous Scanning Tunneling Spectroscopy

Facility: Molecular Foundry @ LBNL | Technique: STS Microscopy | Achievement: ~4% of data required, ~35 hrs vs ~1 mo acq. time

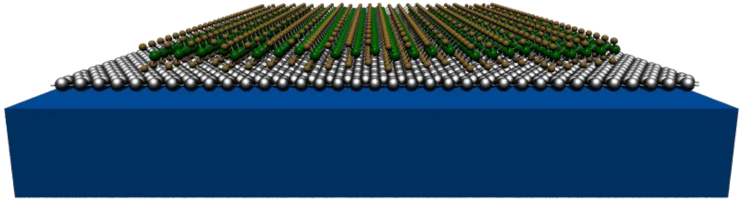
Scanning Probe Microscopy

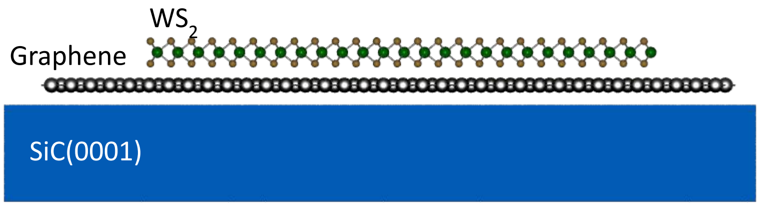


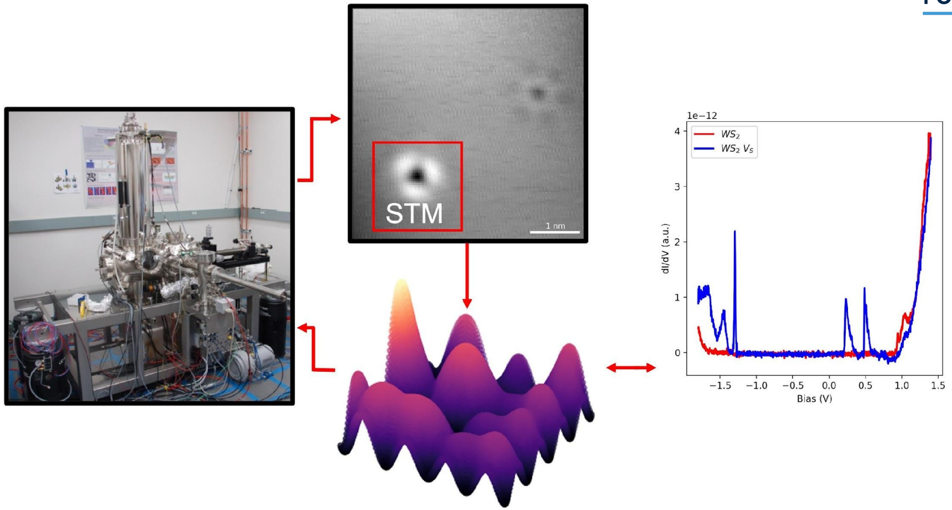
STM / STS

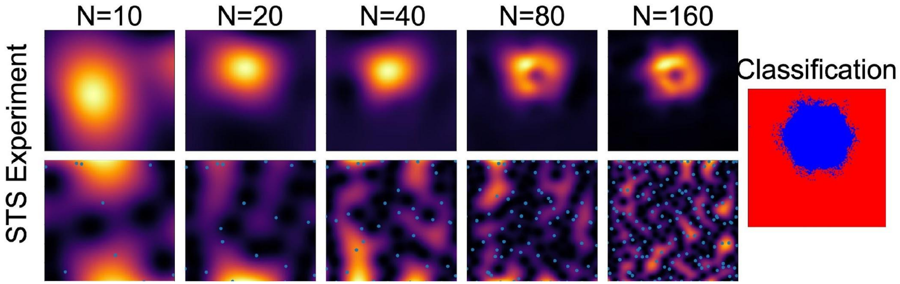
structure and electronic properties

Investigate Next Frontiers in 2D Quantum Materials





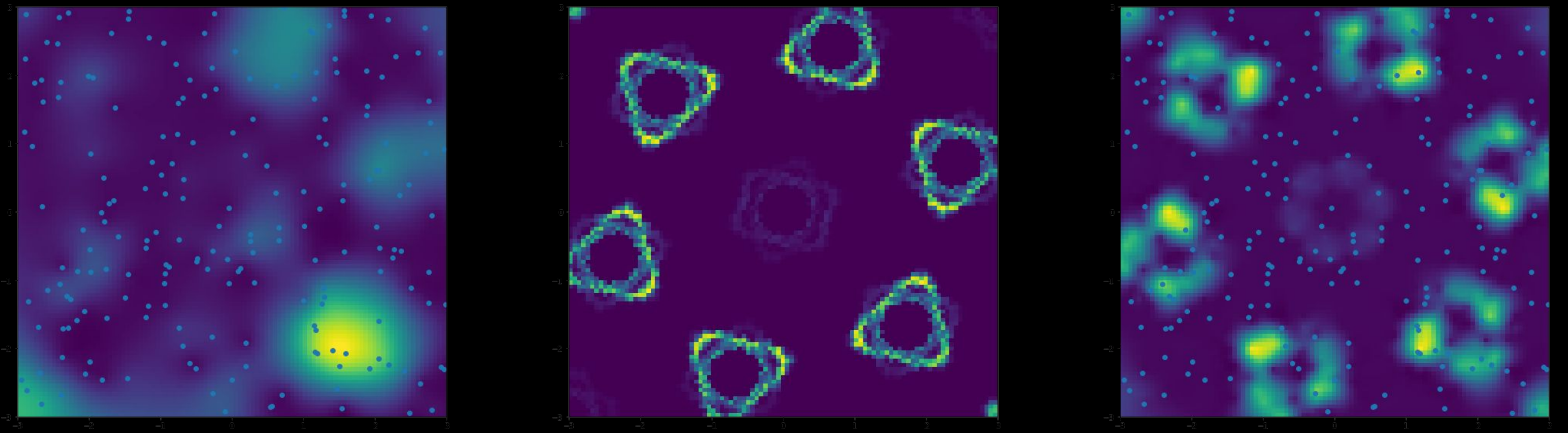




Thomas *et al.*, arxiv:2110.03351 (2021)

The Power of the RKHS: Domain-Informed Symmetry Constraints — Six-Fold Symmetry

$$k(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{36} \sum_{\phi} \sum_{\theta} \tilde{k}(\mathcal{R}_{\phi}\mathbf{x}_i, \mathcal{R}_{\theta}\mathbf{x}_j)$$

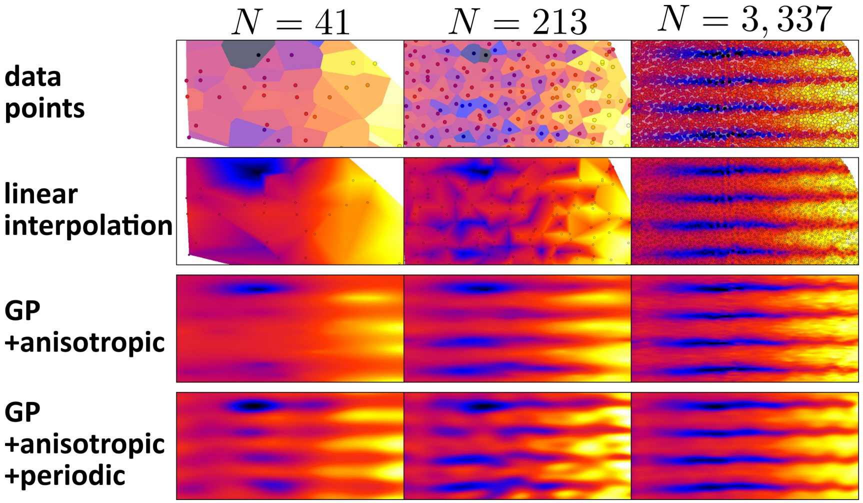


Martin Boehm
Paolo Mutti
Tobias Weber

NEUTRONS FOR SOCIETY

Physics Knowledge in the Form of Periodicity for X-Ray Scattering

Facility: NIST and NSLS II | Technique: SAXS/GISAXS | Achievement: Use of non-stationary kernels to learn and exploit local characteristics



Kevin Yager, Masafumi Fukuto, Jonathan Seppala @ CFN, BNL, NIST

Center for Functional Nanomaterials
Brookhaven National Laboratory

NIST National Institute of Standards and Technology U.S. Department of Commerce

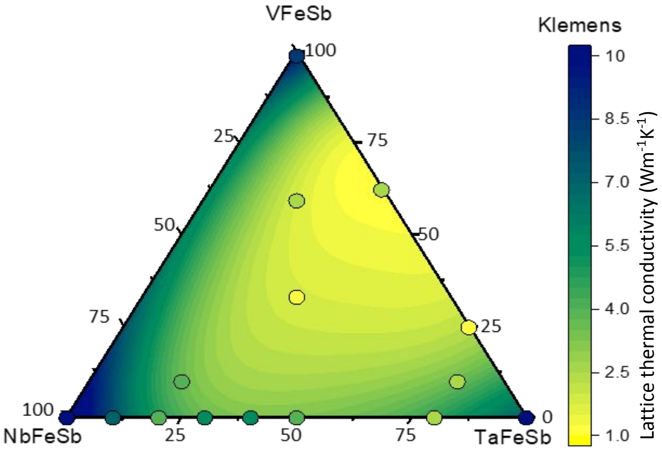
NSLS-II National Synchrotron Light Source II

BROOKHAVEN NATIONAL LABORATORY

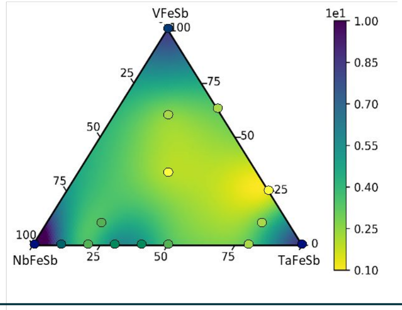
Physics-Aware Prediction of Lattice Thermal Conductivity of Alloys

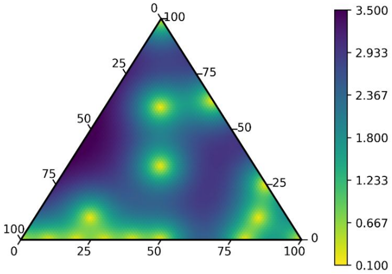
Facility: SLAC @ Stanford | Technique: Diffusivity, Heat Capacity and Density Measurements | Achievement: Physics-informed GP-driven steering

Physics-Based Model

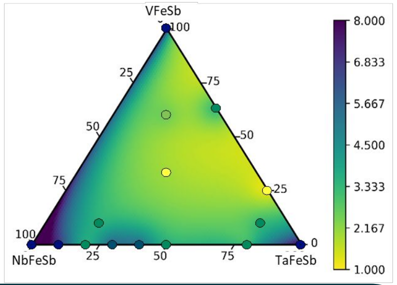


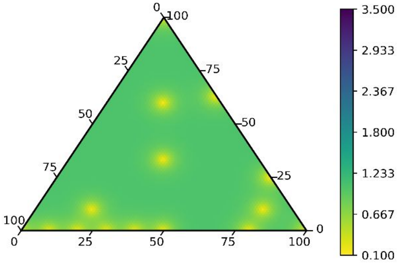
Data-Driven GPR





Physics-Aware GPR

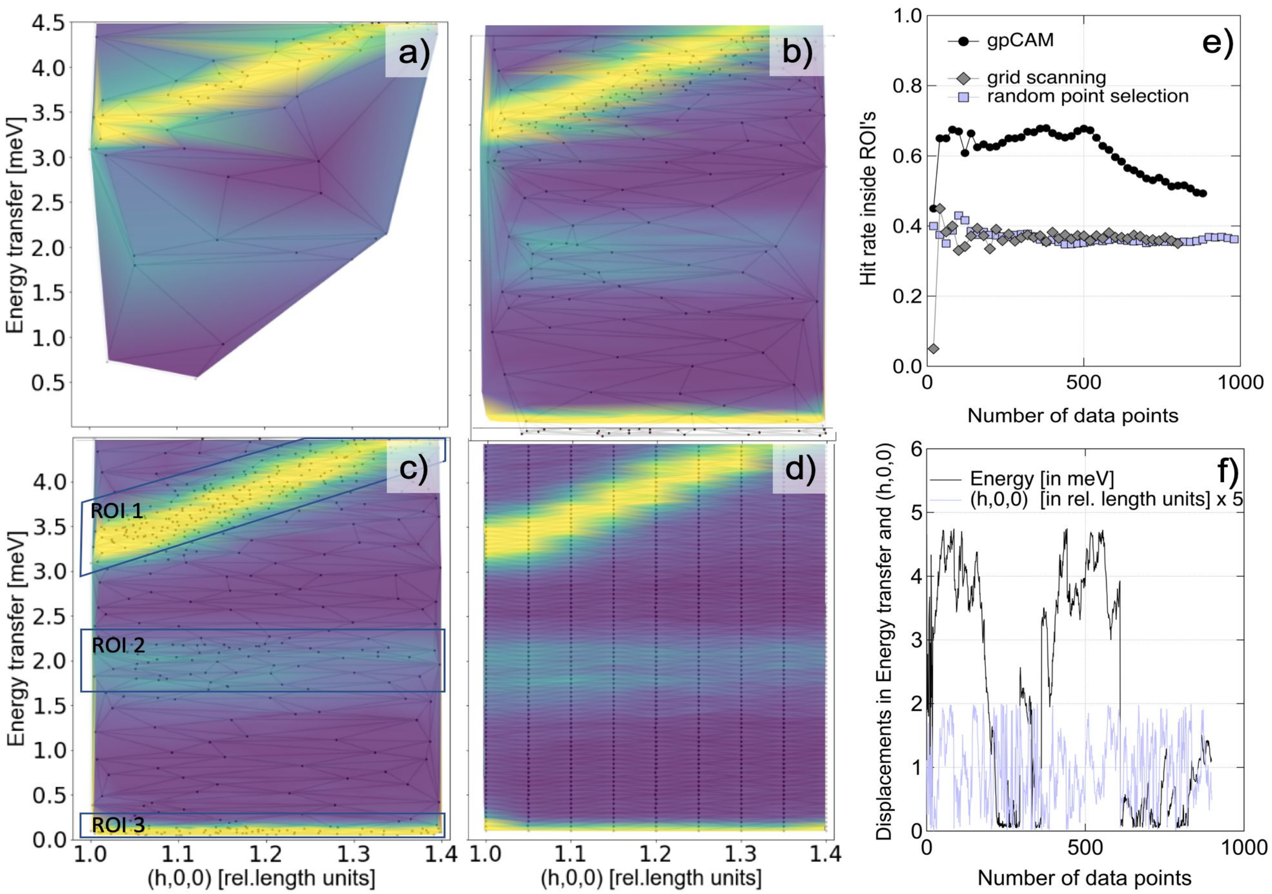




Suchismita Sarker, Apurva Mehta

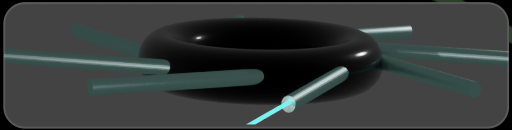
Targeted Autonomous Neutron Scattering

Facility: ILL, France | Technique: Inelastic Neutron Scattering | Achievement: More efficient exploration, experiment time decreased from several days to one night



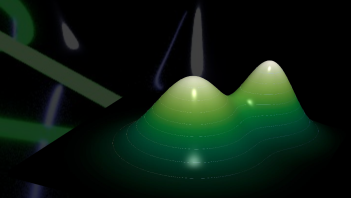
Martin Boehm
Paolo Mutti
Tobias Weber

NEUTRONS FOR SOCIETY

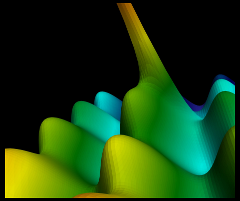


What is Autonomous Data Acquisition

CAMERA Workshop on AE and Main Takeaways

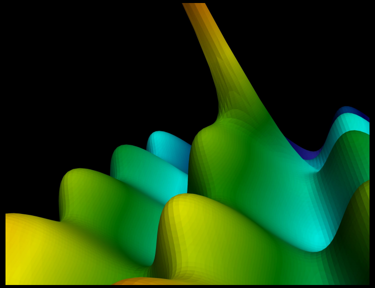


Basic Gaussian-Process-Driven Autonomous Data Acquisition



Mathematical Optimization for AE and ML

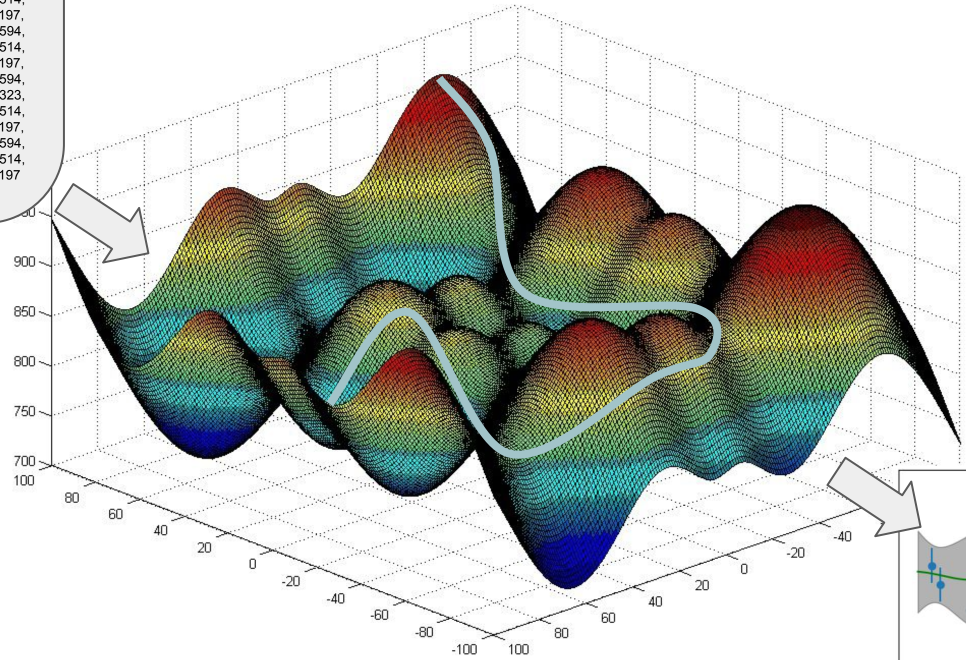
Bringing Autonomous Discovery to the Community

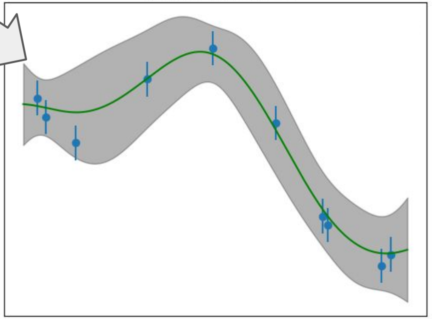


[Mathematical Optimization for AE and ML]

0.26436384, 0.21300795, 0.19834864, ..., 0.20183792, 0.82454492,0.81746336,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.09019633, 0.03045269, 0.55291218, ..., 0.66801905, 0.75265345,1.64352323,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.09019633, 0.03045269, 0.55291218, ..., 0.66801905, 0.75265345,1.64352323,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.09019633, 0.03045269, 0.55291218, ..., 0.66801905, 0.75265345,1.64352323,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197

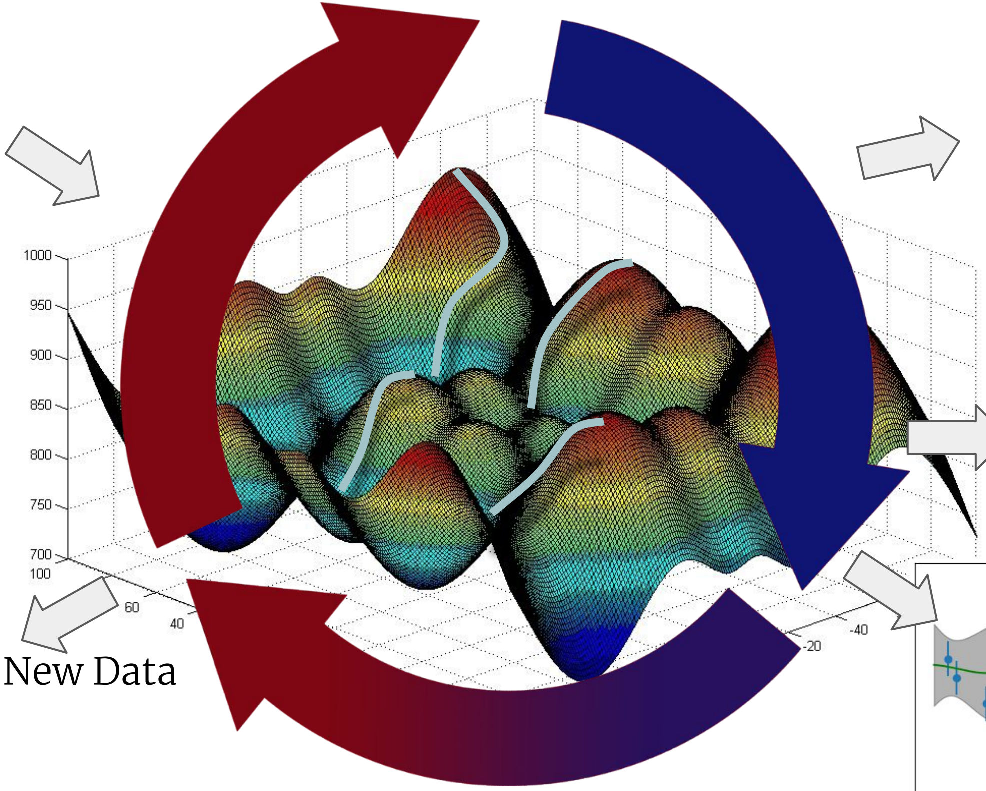
The Traditional Training/Optimization Workflow needs a Large Number of Function Evaluations and Blocks the Main Thread

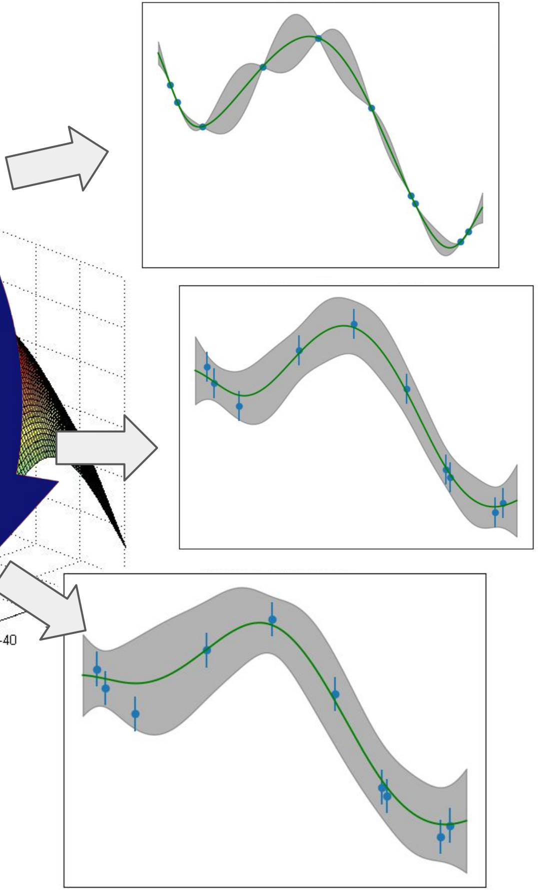




Minimizing Number of Function Evaluations: Asynchronous Distributed Training

0.26436384, 0.21300795, 0.19834864, ..., 0.20183792, 0.82454492,0.81746336,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.09019633, 0.03045269, 0.55291218, ..., 0.66801905, 0.75265345,1.64352323,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.09019633, 0.03045269, 0.55291218, ..., 0.66801905, 0.75265345,1.64352323,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.09019633, 0.03045269, 0.55291218, ..., 0.66801905, 0.75265345,1.64352323,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197,
0.43558082, 0.95638452, 0.99928695, ..., 0.63067478, 0.38601846,0.52014594,
0.68064155, 0.66793227, 0.02274104, ..., 0.37098925, 0.66477699,0.71282514,
0.13523289, 0.85643443, 0.43357488, ..., 0.71829634, 0.98986933,0.60671197





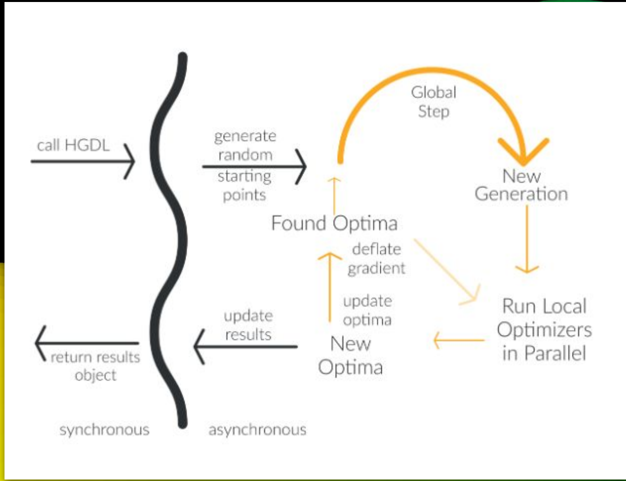
Kill/Restart/Ingest New Data

Optimization of the Log-Likelihood and Acquisition Functions with HGDL

Using DASK, pytorch and GPUs for High Performance Asynchronous Distributed Training

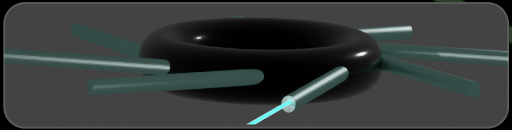
David Perryman

$$\log(L(\mathcal{D};\phi,\mu)) = -\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu}(\mathbf{x}))^T(\mathbf{K}(\phi)+\mathbf{I}_e)^{-1}(\mathbf{y}-\boldsymbol{\mu}(\mathbf{x})) - \frac{1}{2}\log(|\mathbf{K}(\phi)+\mathbf{I}_e|)$$



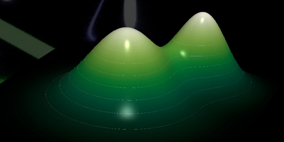
HGDL leads to:

1. a set of different interpretations of the data
2. a set of optimal measurements
3. HPC readiness of training and prediction
4. Asynchronous training

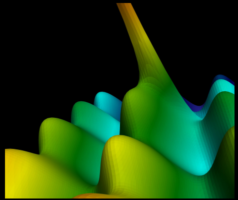


[What is Autonomous Data Acquisition]

[CAMERA Workshop on AE and Main Takeaways]

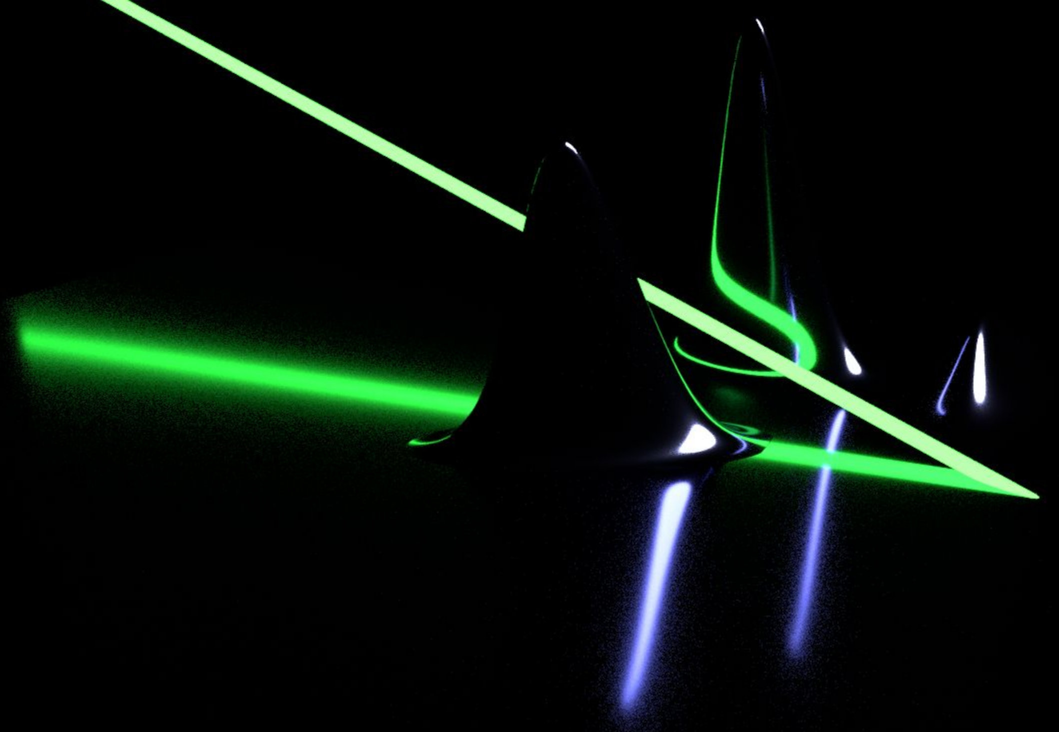


[Basic Gaussian-Process-Driven Autonomous Data Acquisition]



[Mathematical Optimization for AE and ML]

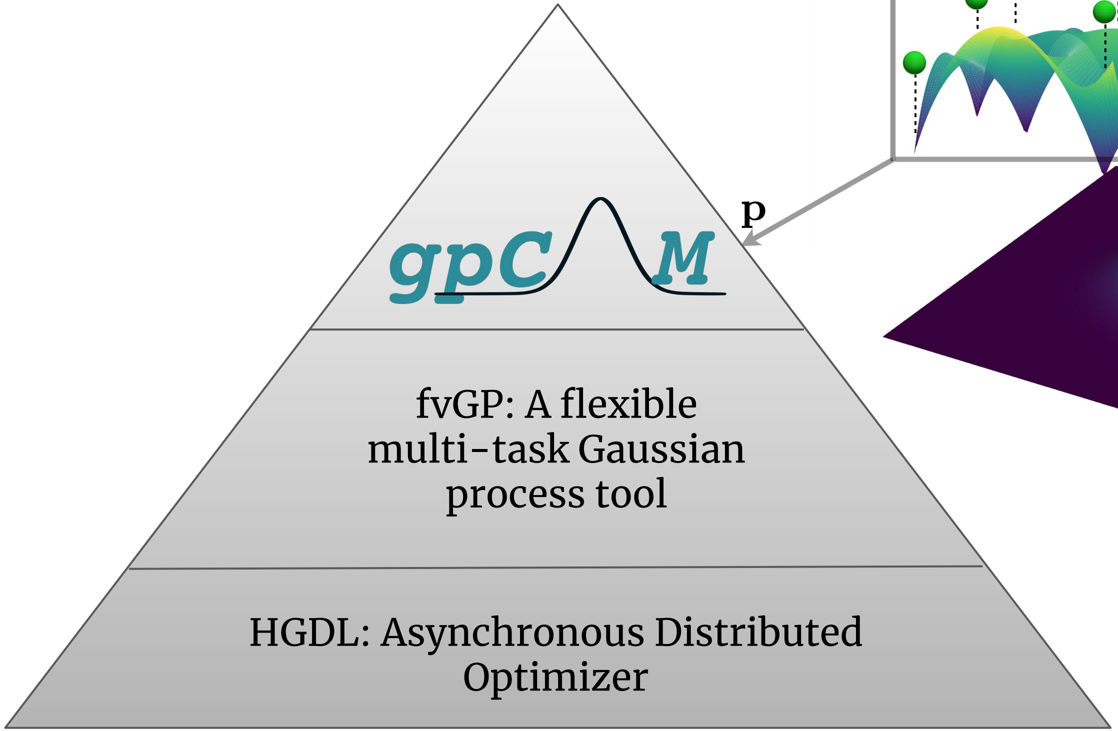
[Bringing Autonomous Discovery to the Community]

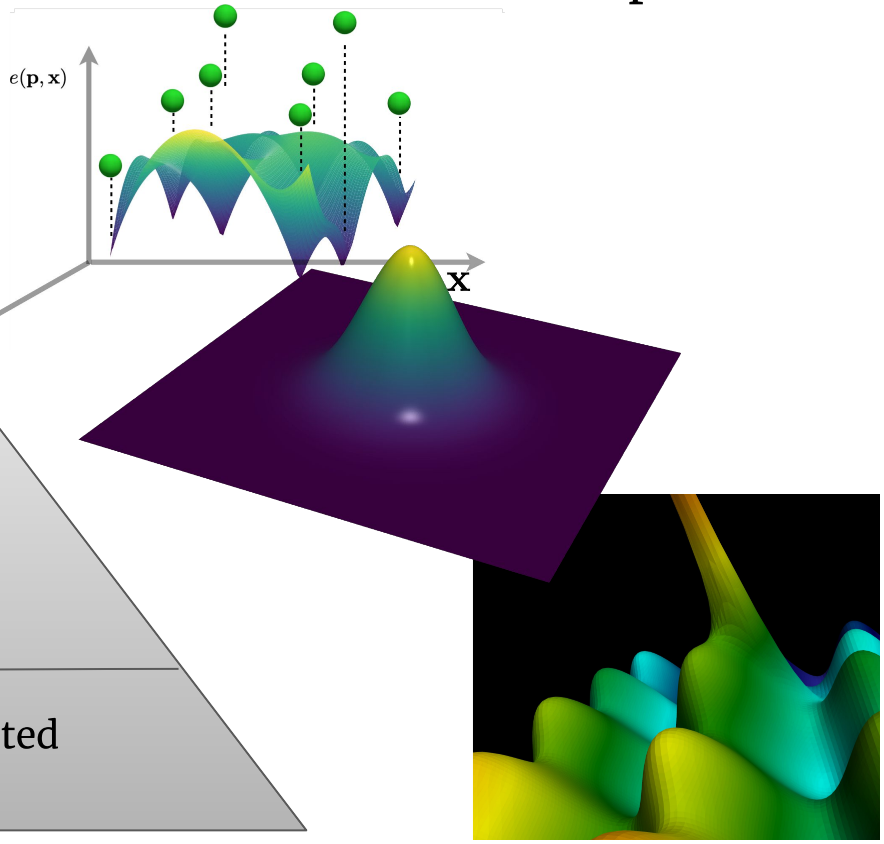


[Bringing Autonomous Discovery to the Community]


HGDL: Asynch.
Distributed Optimizer
fvGP: A flexible multi-task
Gaussian process tool
gpCAM

The products are three separate APIs that are built on top of each other:





```
pip install gpcam
```

```
from gpcam.autonomous_experimenter import AutonomousExperimenterGP
from instrument import instrument
import numpy as np

parameters = np.array([[3.0,45.8],
                       [4.0,47.0]])
init_hyperparameters = np.array([1,1,1])

hyperparameter_bounds =  np.array([[0.01,100],[0.01,100.0],[0.01,100]])

my_ae = AutonomousExperimenterGP(parameters, instrument, init_hyperparameters,
                                 hyperparameter_bounds,  init_dataset_size=10)

my_ae.train()

my_ae.go()
```

```
def instrument(data):
    for entry in data:
        entry["value"] = np.sin(np.linalg.norm(entry["position"]))
    return data
```

More information: gpcam.lbl.gov

BERKELEY LAB
Bringing Science Solutions to the World

MOLECULAR FOUNDRY

BSISB
Berkeley Synchrotron Infrared Structural Biology Imaging Program

ALS
ADVANCED LIGHT SOURCE

CRD
COMPUTATIONAL RESEARCH DIVISION

- James Sethian
- Alexander Hexemer
- Eli Rotenberg
- Charles Melton
- David Perryman
- Peter Zwart
- Hoi-Ying Holman
- Liang Chen
- Pablo Enfedaque
- Carolin Sutter-Fella
- Ron Ronald Pandolfi
- Dinesh Kumar
- Harinarayan Krishnan
- Daniela Ushizima
- Luca Francaviglia
- Maged Abdelsamie
- Edward Bernard
- Antonio Rossi
- Esther Singer
- John Thomas
- Steven Lee
- John Thomas

BROOKHAVEN NATIONAL LABORATORY

NSLS-II
NATIONAL SYNCHROTRON LIGHT SOURCE II

Center for Functional Nanomaterials
Brookhaven National Laboratory

- Masafumi Fukuto
- Kevin Yager
- Gregory Doerk
- Ruipeng Li
- Thomas Caswell
- Stuart Campbell
- Esther Tsai
- Aaron Stein
- Suwon Bae
- Sebastian Russel
- Yugang Zhang
- Guillaume Freychet
- Karen Chen-Wiegart
- Sanjit Ghose
- Cheng-Chu Chun
- Chonghang Zhao
- Dan Olds
- Phil Maffettone
- Joshua Lynch
- Aaron Michelson

COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

Berkeley
UNIVERSITY OF CALIFORNIA

UNIVERSITÄT BAYREUTH

- Michael Buchhorn
- Eva Herzig

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

- Lee Richter
- Gilad Kusne
- Jon Seppala
- Tyler Martin

Carnegie Mellon University

NEUTRONS FOR SOCIETY

- Martin Boehm
- Tobias Weber
- Yannick Le Goc
- Paul Steffens
- Paolo Mutti

SLAC
NATIONAL ACCELERATOR LABORATORY

- Apurva Mehta
- Suchismita Sarker
- Aditya Venkatraman
- Adi Hanuka
- Auralee Edelen

UNIVERSITY OF BIRMINGHAM

- Daniil Mironov
- Andrei Leonard Nicusan
- Kit Windows-Yule

TUM

- Astrid Schneidewind
- Marina Ganeva
- Mario Parente

AFRL
THE AIR FORCE RESEARCH LABORATORY
LEAD | DISCOVER | DEVELOP | DELIVER

- Jason Streit
- Richard Vaia

CAMERA

Applied Math

U.S. DEPARTMENT OF ENERGY

Office of Science

A Perspective on Autonomous Experimentation and Discovery

Marcus Michael Noack
MarcusNoack@lbl.gov
gpcam.lbl.gov

Questions?

gpCAM

CAMERA Applied Math