New approaches to Bayesian uncertainty quantification for Nuclear Science

Peter Jacobs

Lawrence Berkeley National Laboratory

University of California, Berkeley

2024 Artificial Intelligence/Machine Learning Principal Investigator Exchange Meeting

Gaithersburg MD
December 4-5, 2024

Bayesian UQ project

Lead PI: Peter Jacobs, LBNL

Co-PIs:
- LBNL: Alan Poon, Jayson Vavrek
- Duke University: Simon Mak
- UC Berkeley: Yury Kolomensky, Uros Seljak
- Wayne State University: Chun Shen

BUQ project outline

Study of complex phenomena often requires the rigorous comparison of theory/model calculations and experimental measurements

→ Inverse Problem

Solution: Bayes's Theorem

$$P(\vec{\theta}|\mathrm{data}) = \frac{P(\mathrm{data}|\vec{\theta})P(\vec{\theta})}{P(\mathrm{data})}$$

$$\vec{\theta} : \text{Model parameters}$$

Widespread application in NP, HEP, Cosmology, Materials Science,…

- constrain model parameters
- validate the physical picture underlying the models
- discover new effects

Bayesian inference can be computationally challenging in practice because of

- large number of nuisance parameters
- high computational cost of the model calculations

BUQ project

Develop and deploy ML-based analysis tools for efficient Bayesian Inference in a broad range of NP experiments

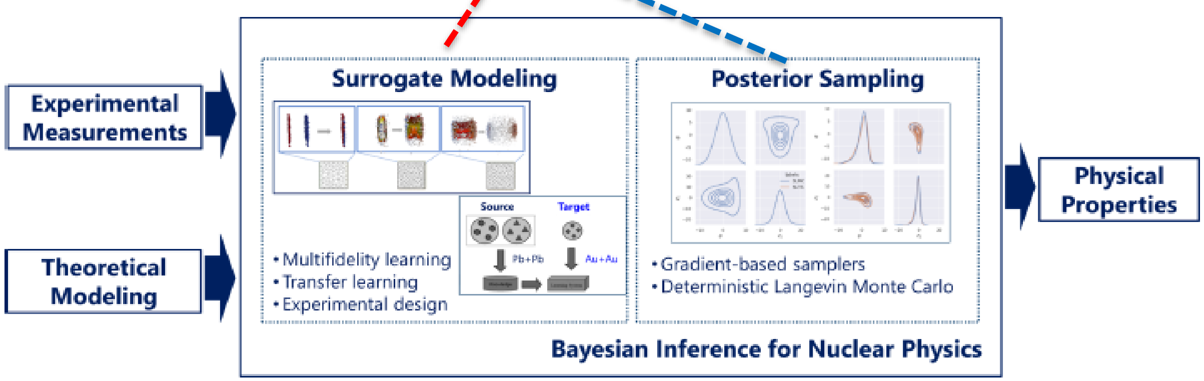
- mass and fundamental nature of the neutrino
- study of the Quark-Gluon Plasma
- mapping of natural and anthropogenic radiation environments

All require Bayesian inference, very different in character

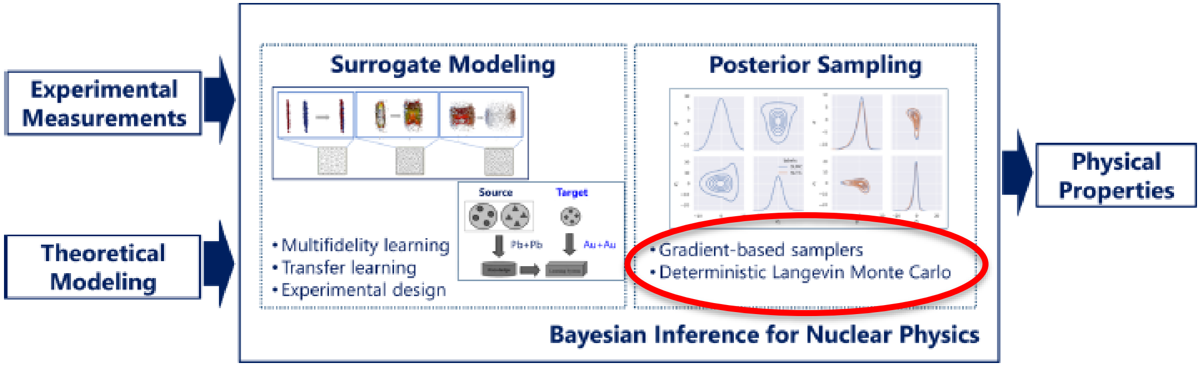
may be high-dimensional or multi-modal

may be computationally expensive

$$P(\vec{\theta}|\text{data}) = \frac{P(\text{data}|\vec{\theta})P(\vec{\theta})}{P(\text{data})}$$



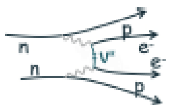
Uncertainty quantification



Meaningful measurement or parameter constraint requires the specification of uncertainty

- Standard discrete MC methods are not differentiable: uncertainty not well-defined
- UQ requires differentiable samplers: novel ML-based approaches

Neutrino-less double beta decay (0νββ)



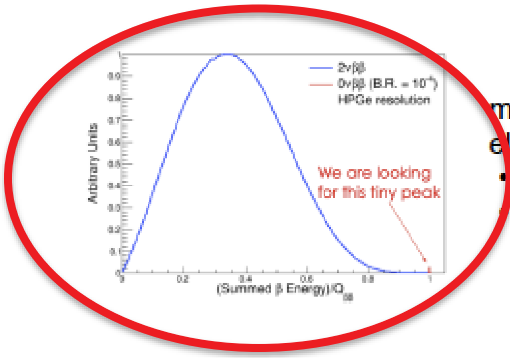
$\Delta L \neq 0$

Bayesian Inference and UQ for 0νββ decay

Experimental goal is to measure mono-energetic peak at Q_{bb}

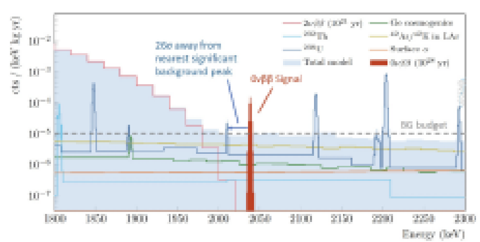
$$0\nu\beta\beta : (\mathbf{A}, \mathbf{Z}) \rightarrow (\mathbf{A}, \mathbf{Z}+\mathbf{2}) + \mathbf{2e^-}$$

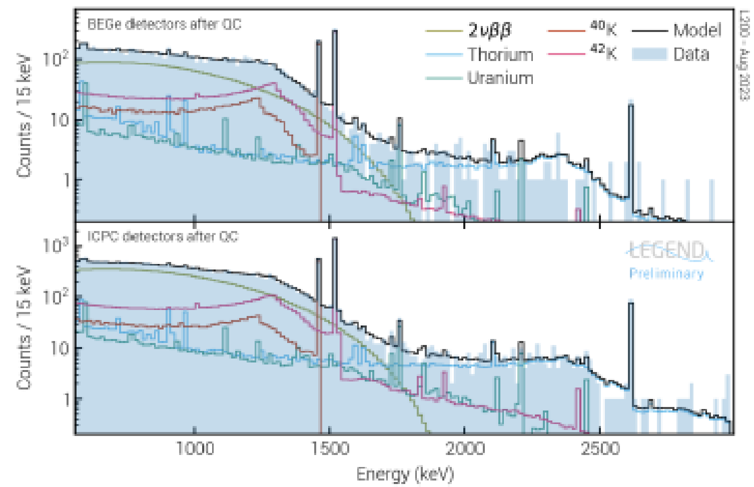
Experimental sensitivity: background (BI) > 1 $T^{0\nu}_{1/2} \propto \varepsilon \cdot a \cdot \sqrt{\frac{M \cdot t}{BI \cdot \Delta E}}$



measure sum energy spectrum of electrons
- 2νββ ➞ continuum
- 0νββ ➞ mono-energetic peak @ $Q_{\beta\beta}$

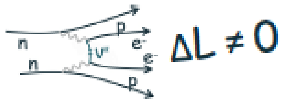
But this signal is buried under other backgrounds…



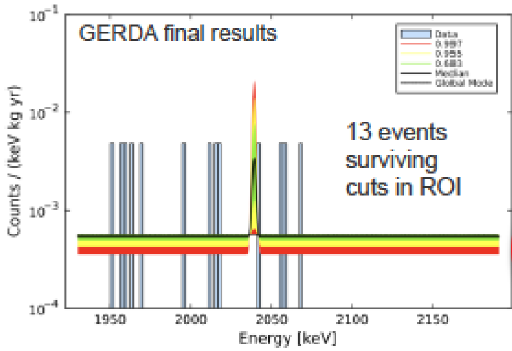


→ increase sensitivity by **background reduction (BI)** at $Q_{\beta\beta}$ and simultaneous increase of mass (M) and improvement of the energy resolution (ΔE)

11 | Ann-Kathrin Schütz Bayesian uncertainty quantification for 0νββ decay 04/10/2024



Bayesian Inference and UQ for 0νββ decay



Example GERDA/ L200 commissioning:
- < 10 detector datasets (< 20 cts in all datasets)
- $Q_{\beta\beta}$, σ_i, μ_i^B, μ_i^S
- → < 50 nuisance parameters

LEGEND-1000: 10x detector channels
- ~ 100 measurement campaigns
- ~ 150 detector datasets
- $Q_{\beta\beta}$, σ_i, μ_i^B, μ_i^S
- → 10^5-10^6 nuisance parameters

The total likelihood is constructed as the product of all $\mathcal{L}_i$ weighted with the Poisson term:

Poisson weight

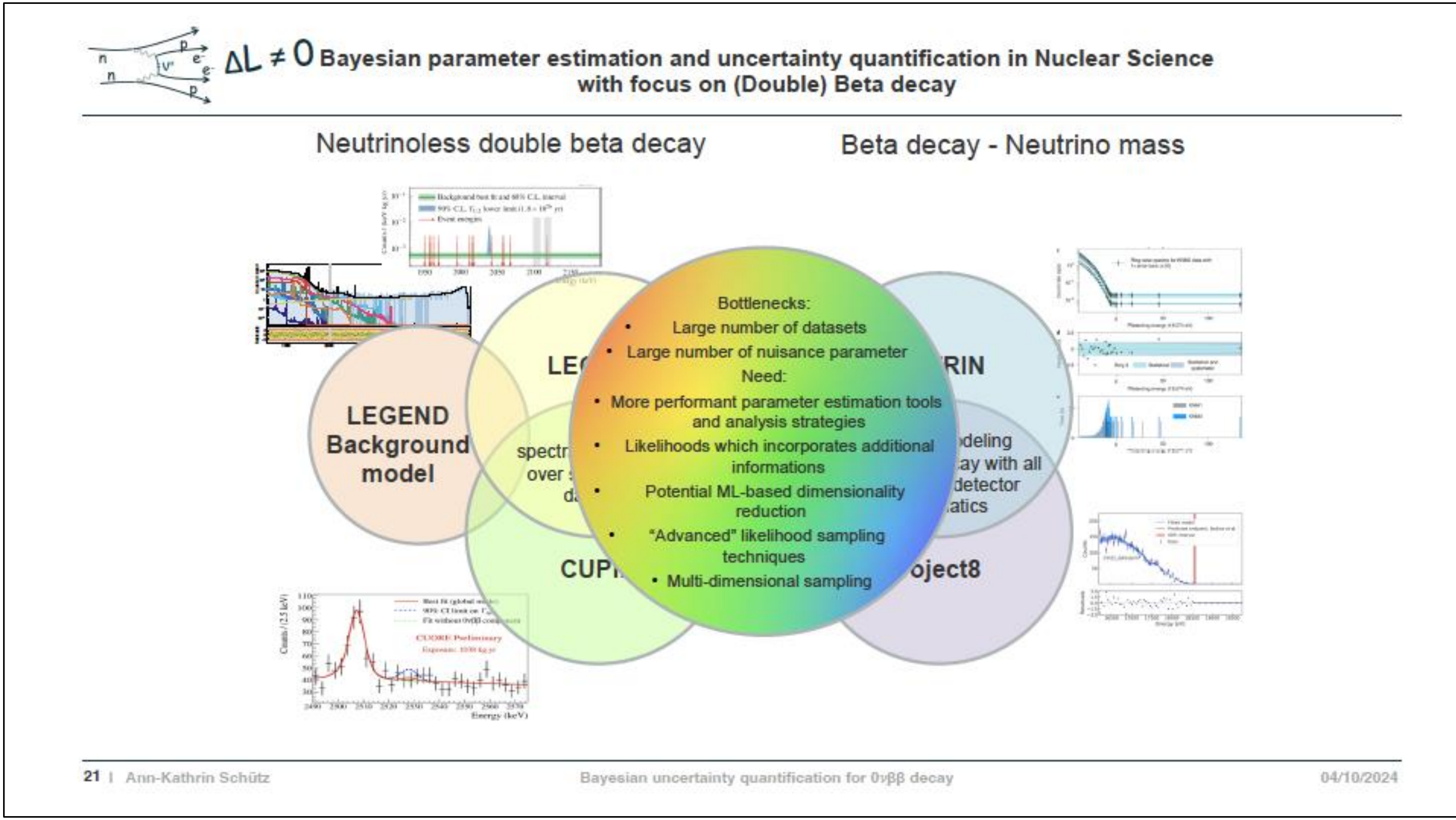
Flat background

Gaussian signal

Expected counts

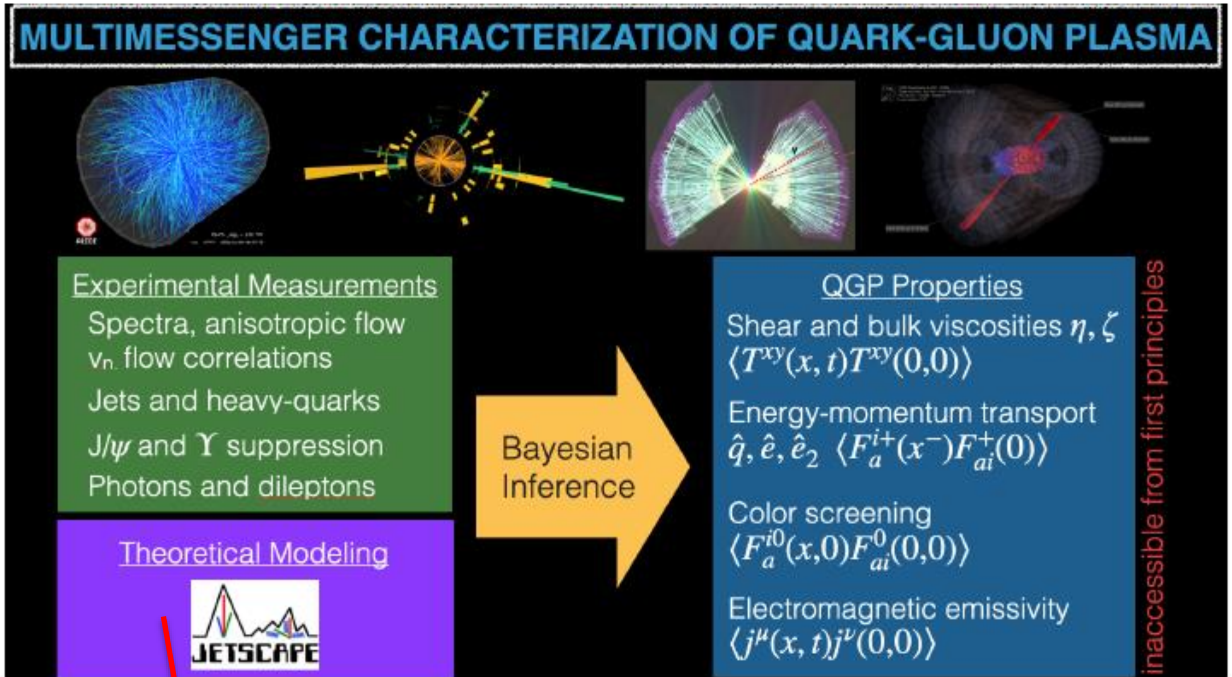
$$\mathcal{L} = \prod_k \left[\boxed{\frac{(\mu_{s,k}+\mu_{b,k})^{N_k}\, e^{-(\mu_{s,k}+\mu_{b,k})}}{N_k!}} \times \prod_{i=1}^{N_k} \frac{1}{\boxed{\mu_{s,k}+\mu_{b,k}}} \times \left(\boxed{\frac{\mu_{b,k}}{\Delta E}} + \boxed{\frac{\mu_{s,k}}{\sqrt{2\pi}\sigma_k} e^{-\frac{(E_i-Q_{\beta\beta})^2}{2\sigma_k^2}}} \right) \right]$$

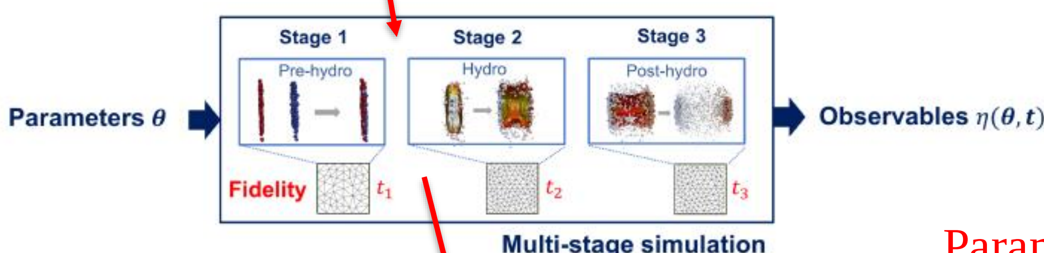
where:
- N_k total number of events observed in the ith partition
- E_i individual event energies in the ith partition
- $\sigma_k = FWHM_i/(2\sqrt{2ln2})$: energy resolution in ROI
  - the average FWHM across partitions is 3.29 keV



ΔL ≠ 0 Bayesian parameter estimation and uncertainty quantification in Nuclear Science with focus on (Double) Beta decay
Neutrinoless double beta decay
Beta decay - Neutrino mass
LEGEND Background model
LEG
spectr
over
CUP
RIN
deling
ay with all
detector
atics
oject8
Bottlenecks:
• Large number of datasets
• Large number of nuisance parameter
Need:
• More performant parameter estimation tools and analysis strategies
• Likelihoods which incorporates additional informations
• Potential ML-based dimensionality reduction
• "Advanced" likelihood sampling techniques
• Multi-dimensional sampling
CUORE Preliminary
Energy (keV)
21 | Ann-Kathrin Schütz
Bayesian uncertainty quantification for 0νββ decay
04/10/2024

Quark-gluon plasma





Parameter space is modest: 5~30

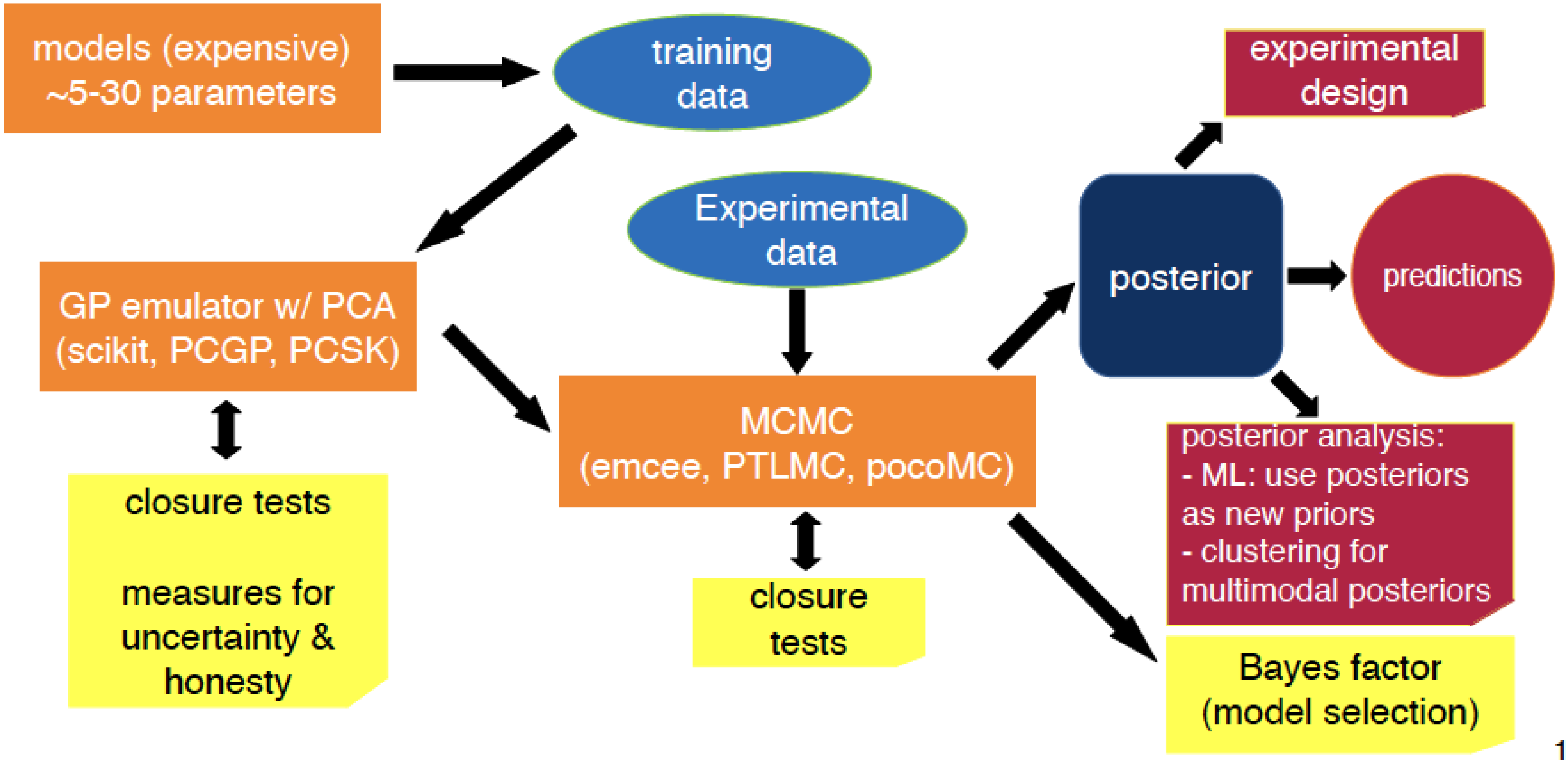
$$P(\vec{\theta}|\text{data}) = \frac{P(\text{data}|\vec{\theta})P(\vec{\theta})}{P(\text{data})}$$

Bottleneck: computationally expensive forward model

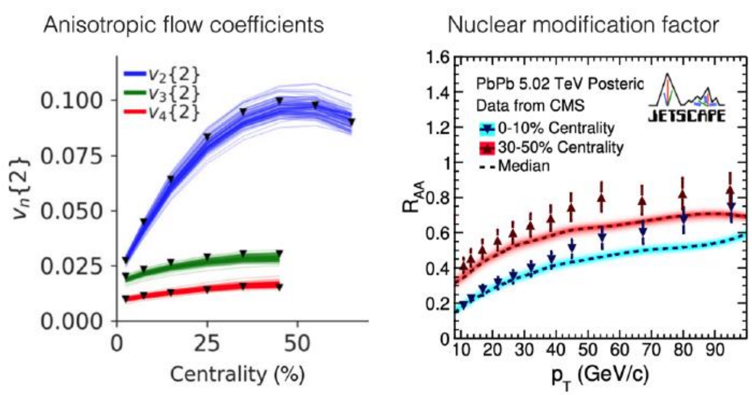
Bayesian inference: QGP

Hendrik, Lipei, Raymond

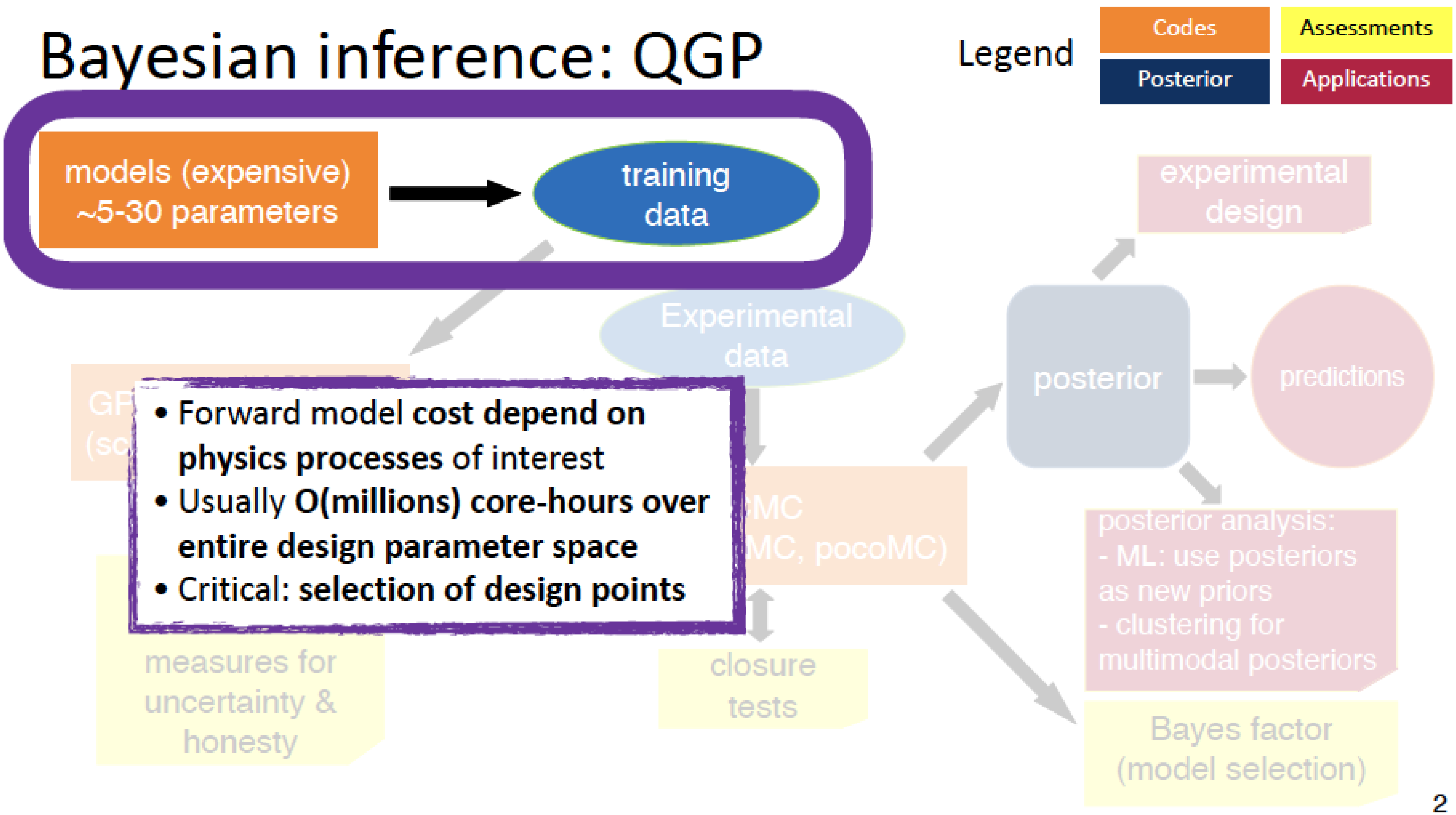
Legend: Codes | Assessments | Posterior | Applications



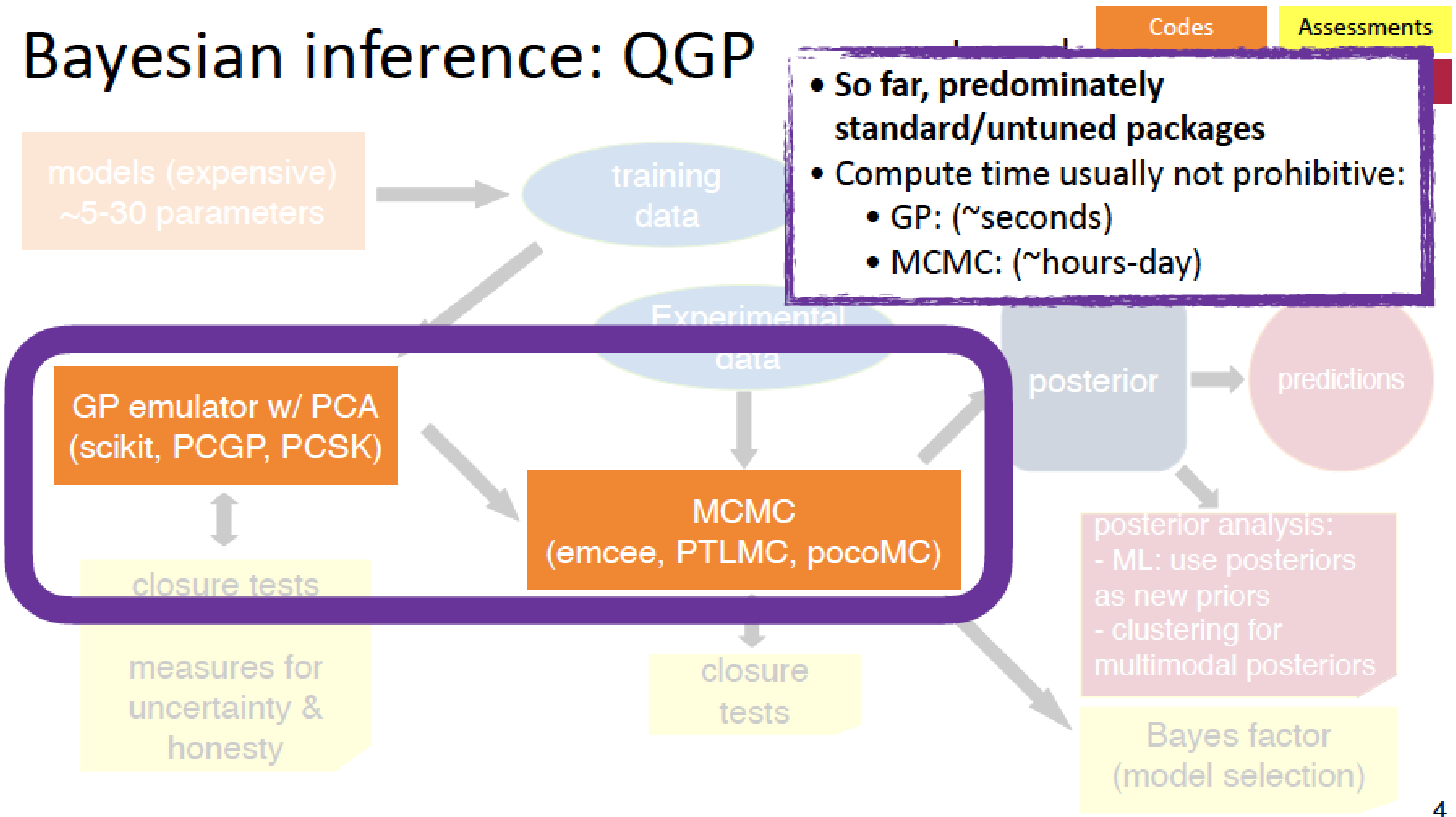
Posterior distributions for collective flow and jet quenching observables

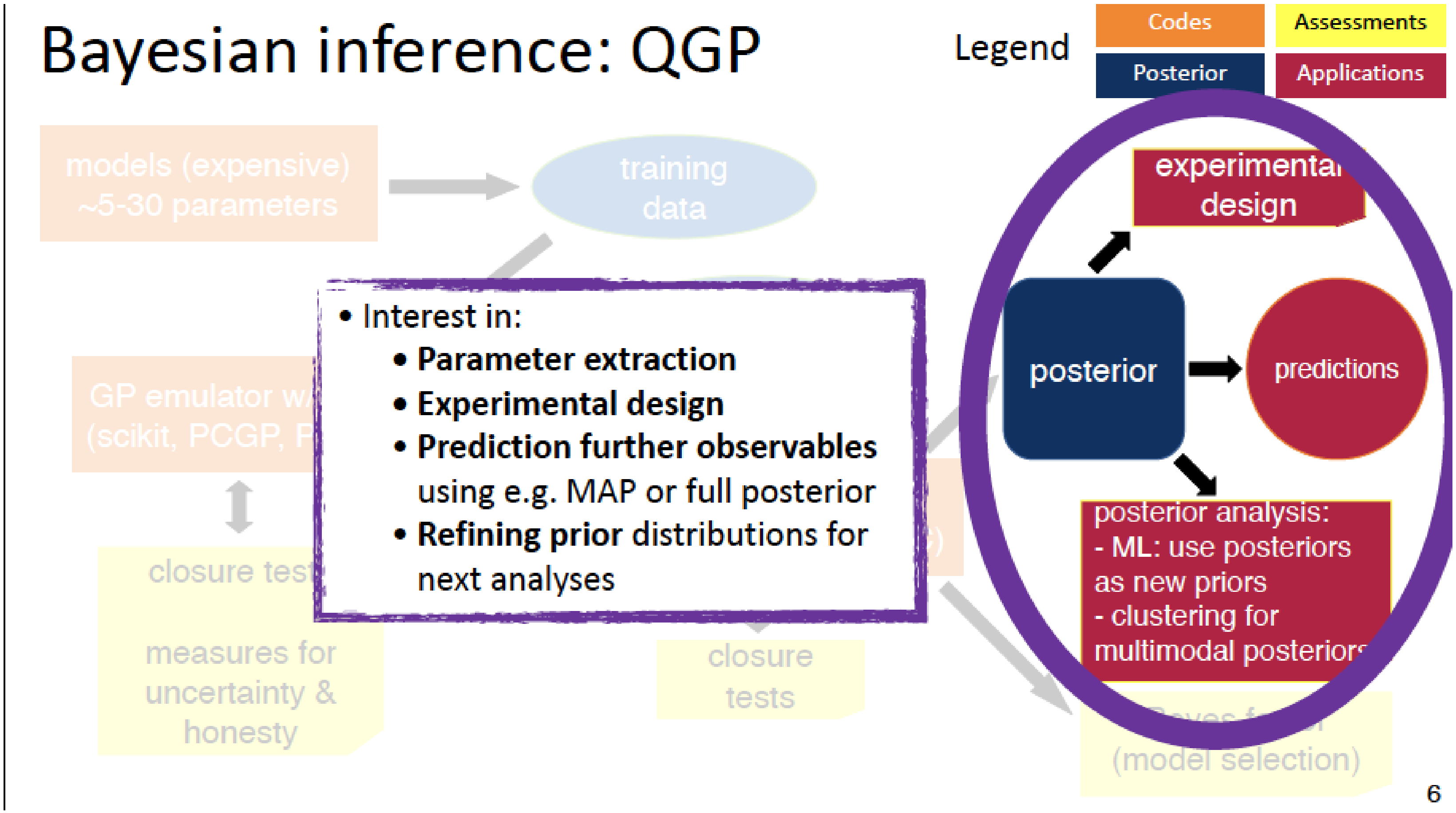


Bayesian inference: QGP



- Forward model **cost depend on physics processes** of interest
- Usually **O(millions) core-hours over entire design parameter space**
- Critical: **selection of design points**


Bayesian inference: QGP
Codes
Assessments
So far, predominately standard/untuned packages
Compute time usually not prohibitive:
GP: (~seconds)
MCMC: (~hours-day)
models (expensive)
~5-30 parameters
training data
Experimental data
GP emulator w/ PCA
(scikit, PCGP, PCSK)
MCMC
(emcee, PTLMC, pocoMC)
posterior
predictions
closure tests
measures for uncertainty & honesty
closure tests
posterior analysis:
- ML: use posteriors as new priors
- clustering for multimodal posteriors
Bayes factor
(model selection)
4



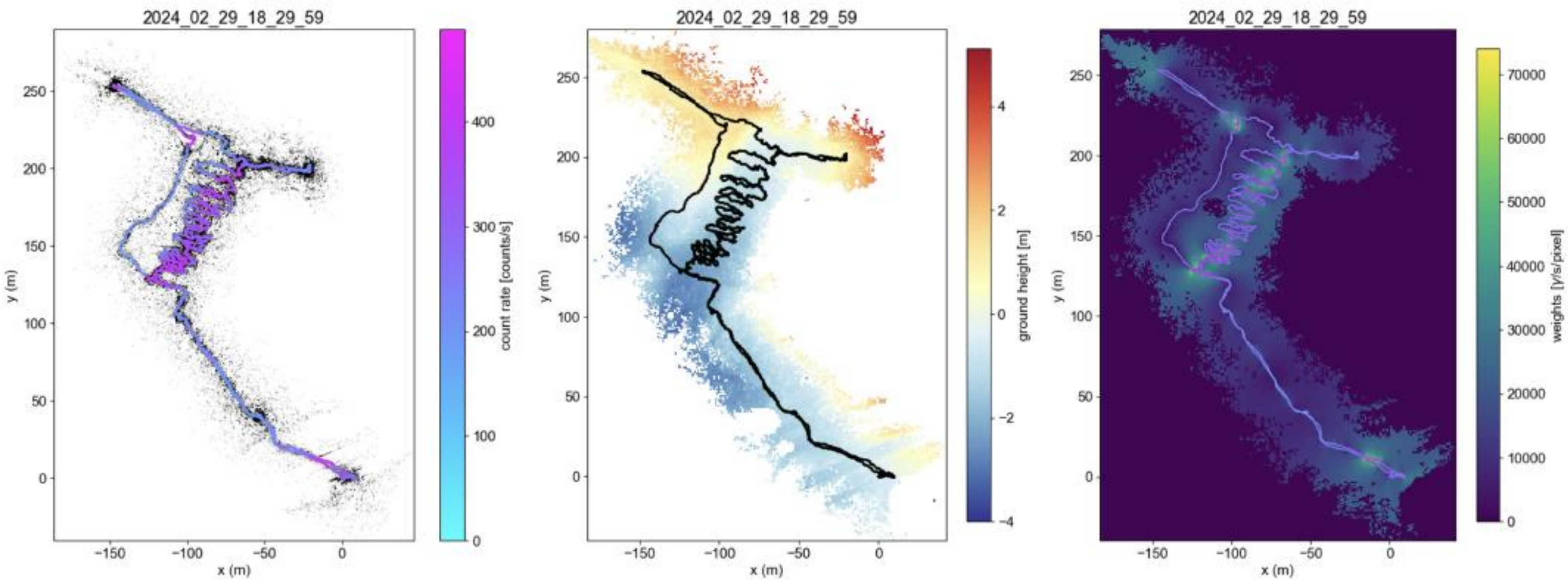
Differentiable samplers: UQ

Radiation imaging of the environment

Radiation mapping overall goals

Where is the radiation and how bad is it? I.e., what is the radiation *distribution* in the scene?

LiDAR + rad measurements → Scene representation → Rad image reconstruction

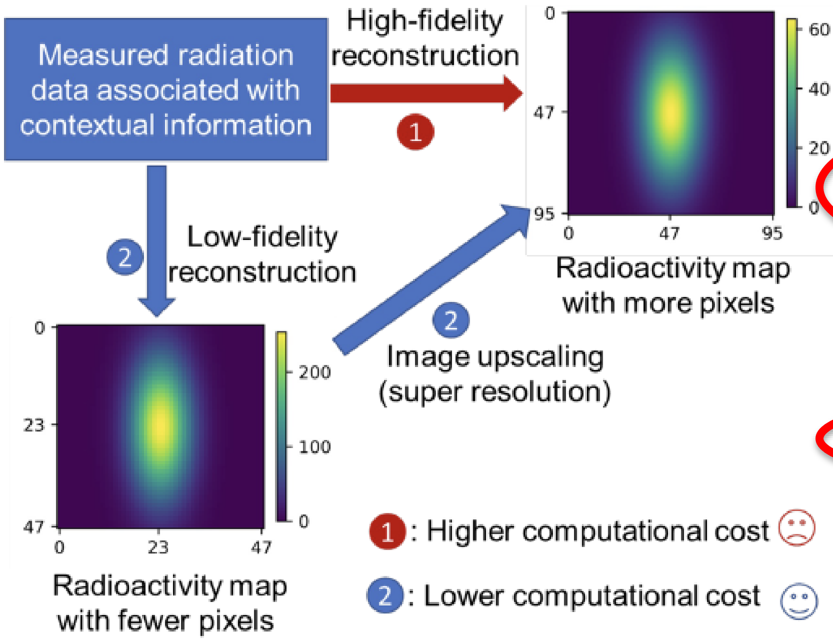


Inverse problem → Bayesian Inference

Unique computing challenge: accurate real-time image reconstruction

Multi-fidelity ML for reconstruction

Rad image reconstruction



Note: Only 2D map is considered

After low-fidelity reconstruction, perform image upscaling to generate the high-fidelity map

Image upscaling by machine learning algorithms with significantly less computation

Benefits of lower computation cost:

- acceleration of data processing to enable real-time or near real-time reconstruction of high-fidelity results
- reduction of computer memory usage so compact computer can be used on smaller, portable edge systems
    - (can't put an NVIDIA 4090 GPU on every system!)
- reduction of battery usage which allows more system operation time

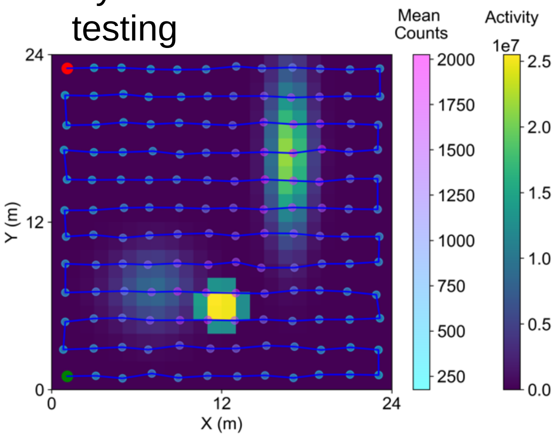
MCLMC for image reconstruction

Micro-canonical Langevin Monte Carlo

Fast MCMC sampler developed by Jakob Robnik and Uroš Seljak

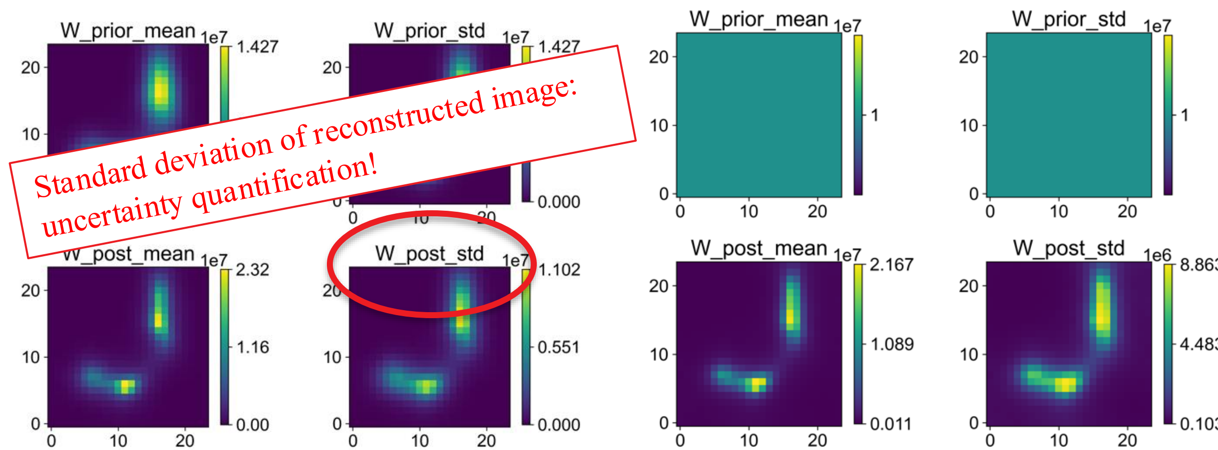
Robnik, Jakob, and Uroš Seljak. "Fluctuation without dissipation: Microcanonical langevin monte carlo." *arXiv preprint arXiv:2303.18221* (2023).

Synthetic data used for testing

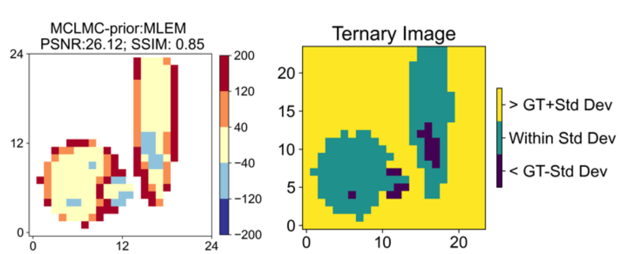


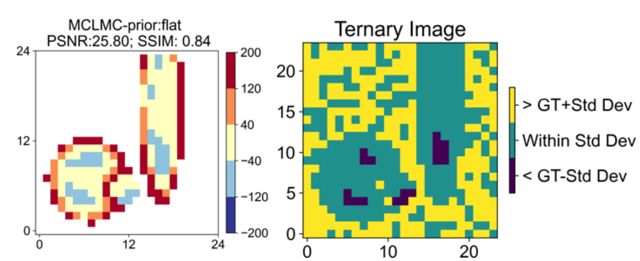
Measured counts plotted on top of ground truth radiation activity

Consistent sampling performance by MCLMC: different prior distribution of the radiation activity leads to similar posterior distribution



"Closure" test: high similarity achieved between the MCLMC reconstructed image with ground truth





Comparison of sampling times for MCMC samplers

For same reconstruction scenario, MCLMC is significantly faster than other MCMC samplers, which enables real-time (or near real-time) image reconstruction with uncertainty quantification.

sampler	Number of samples; computation time	convergence
MCLMC	1e5; ~ 13 s	Fully converged
HMC	1e4; ~ 1 hour 6 mins	Fully converged
	1000; ~ 7 mins	Some convergence, but not fully converged
	100; ~ 30 s	Far from convergence
NUTS	100; ~ 2 mins	Far from convergence
RMHMC	100; ~ 1 min 20s	Far from convergence
Metropolis-Adjusted Langevin Algorithm (MALA)	1000; ~ 40s	No convergence at all
	1e5; ~ 1 hour 9 mins	Far from convergence
random walk	100; 36 s	Far from convergence
	1e4; 6 mins	Far from convergence
ellip_slice		No convergence at all

BUQ project status

Neutrinos:

- evaluate the bottleneck and requirements of the Bayesian methodologies
- validate ML-based sampling approaches; successfully developed a Bayesian Optimization model combining Conditional Neutral Process with a Multi-Fidelity Gaussian Process.

Quark-Gluon Plasma:

- implement multi-fidelity approach to Bayesian Inference (Config model/Duke)
- systematic comparison of GP Emulators for (3+1)D bulk evolution (publication)
- build numerical framework to cross-compare MCMC algorithms (affine invariant, Parallel tempering, pocoMC).
- in progress: comparison of Bayesian evidence in various model setups.

Roch, Jahan and Shen, PRC 110 (2024) 044904

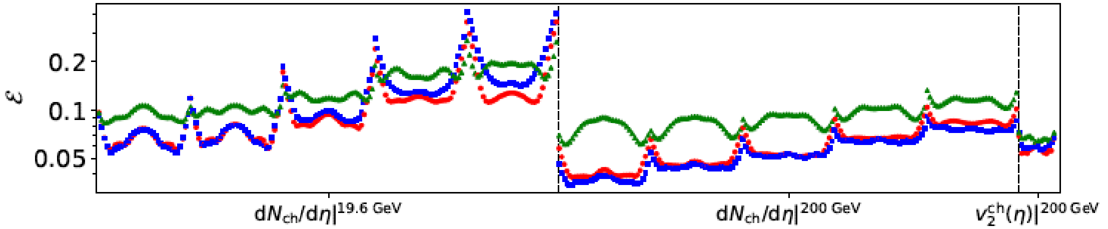
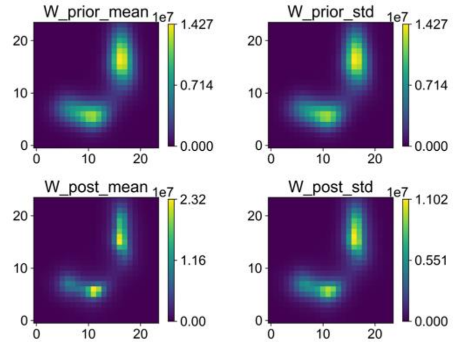


FIG. 2. (Color Online) The averaged RMS errors $\mathcal{E}$ for the three different types of GP emulators. Different training sets are separated by black lines. All emulators are trained with the same 970 LHD points.

Project status (cont'd)

Radiological imaging:

- development of image upscaling and image inpainting methods for both radiological image reconstruction and ground surface estimation (both multi-fidelity-based and non-multi-fidelity for comparison).
- Initial upscaling results show good performance. Initial inpainting results also show very good performance (Conference presentsation at IEEE NSS/MIC 2024)



Algorithms:

- working closely with all three NP groups (Neutrinos, Quark-Gluon Plasma, Radiological Imaging)
- Neutrinos: new multi-fidelity Bayesian optimization framework for neutrino shield simulations.
- QGP: new multi-fidelity Bayesian model for cost-efficient emulation of experimental observables (publication in preparation)
- Radiological Imaging: new Bayesian model for image inpainting with promising results (publication in preparation)

Budget

	FY23 ($k)	FY24 ($k)	Total ($k)
Allocated			
LBNL	338K	354K	692K
Duke	64K	66K	130K
UC Berkeley	182K	180K	362K
Wayne State	77K	39K	116K
total	**661K**	**639K**	**1300K**
Expenditures			
LBNL	306K	38K	344K
Duke	48K	4K	52K
UC Berkeley	49K	11K	60K
Wayne State	70K	6K	76K
Total	**473K**	**59K**	**532K**

FY23 expenditures were significantly below initial projection
Reason: postdoc hiring took several (many) months after start of project

- Positions require special skills, not common in the community
- In one case we waited 12 months for a student to graduate

Project is now fully staffed
→ we project the request of an NCE in FY26 for ~12 PD-months

Deliverables

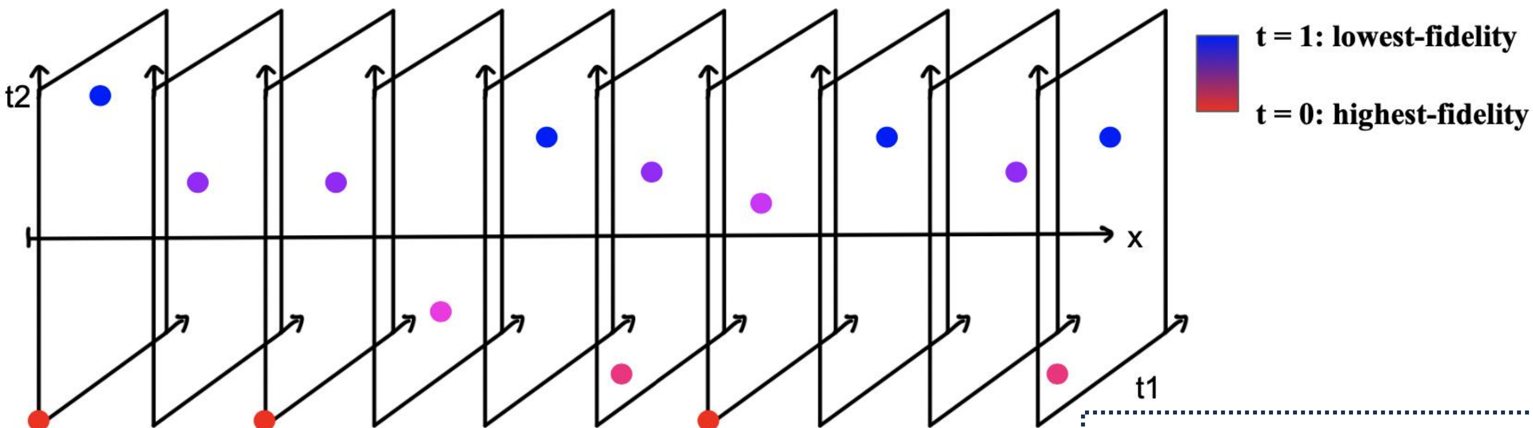
FY23
- Neutrinos: Implement fast gradient-based sampling; implement surrogate modeling.
- QGP: Implement Multi-fidelity learning and transfer-learning methods; initial performance studies. Integrate gradient-based posterior sampling.
- Radiological Mapping: Implement Multi-fidelity learning, transfer-learning, and gradient-based posterior sampling methods. Carry out initial performance studies.

FY24
- Neutrinos: explore new gradient-based sampling and surrogate modeling methods; implement new methods that integrate more detector information, explore performance.
- QGP: Explore performance of Multi-fidelity learning, transfer-learning, and gradient-based sampling, and utilize for novel, large-scale multi-messenger analyses of QGP data from RHIC and LHC.
- Radiological Mapping: Full assessment of new algorithms and first application in ongoing projects in the field; implement and assess new methods for dimensional reduction.

Extra slides

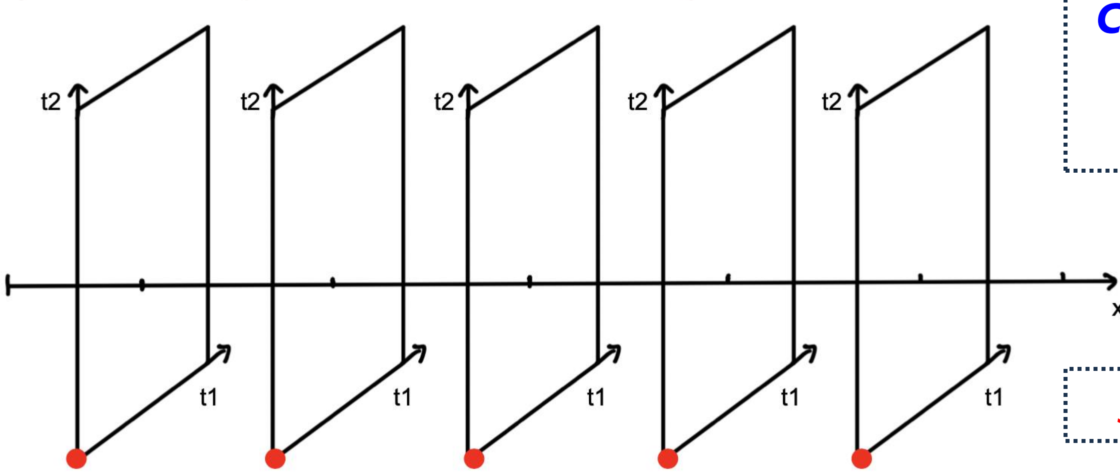
Multi-fidelity surrogates

Idea: Use **multi-fidelity** data $\{f(\boldsymbol{\theta}_i, \boldsymbol{t}_i)\}_{i=1}^{n}$ to train a GP **surrogate model** for predicting the **highest-fidelity** simulator $f(\boldsymbol{\theta}, \mathbf{0})$



CONglomerate multi-FIdelity *Gaussian process modeling*

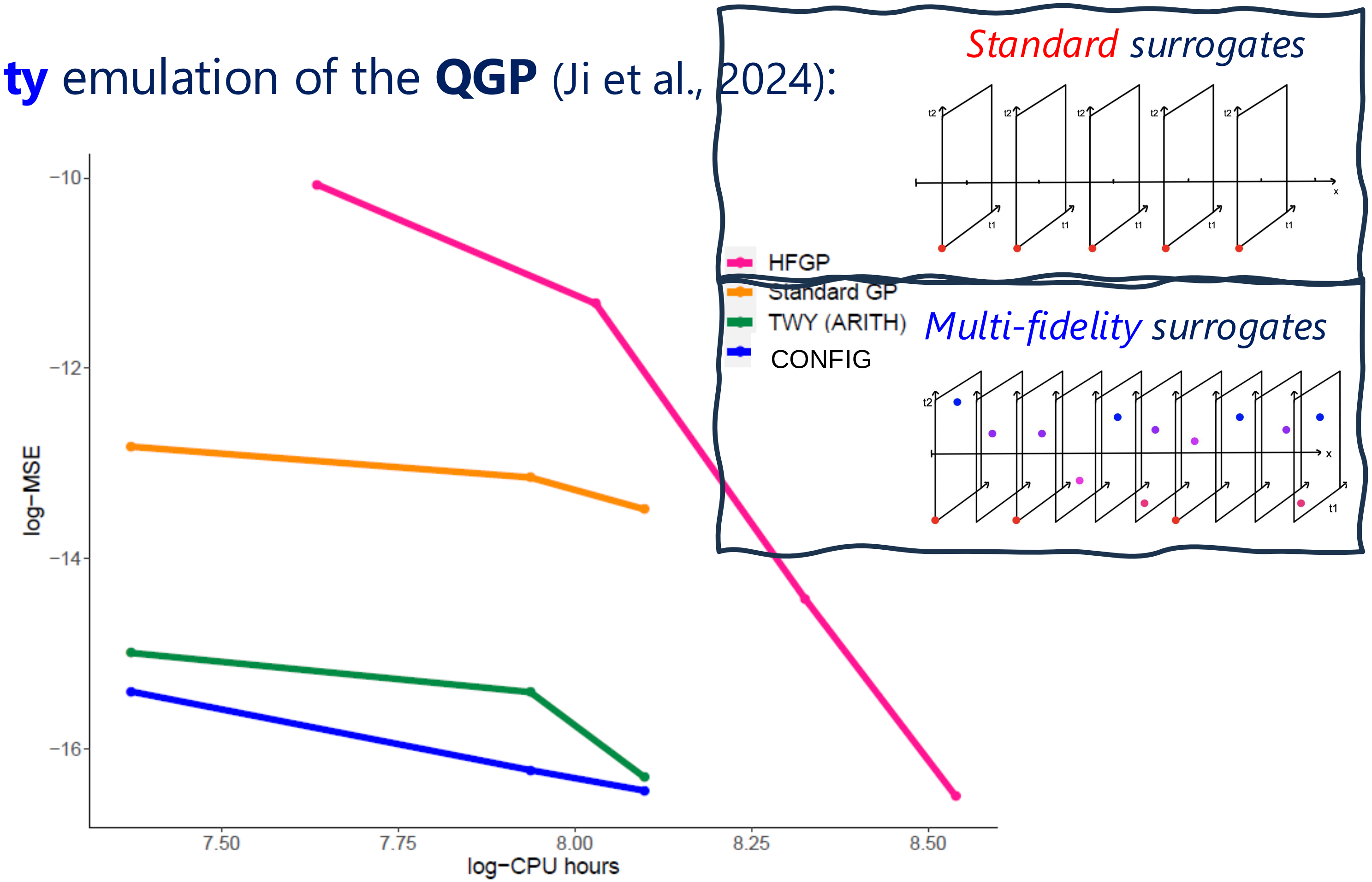
(**CONFIG**; Ji et al., 2024 *JUQ*)



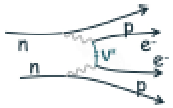
Standard surrogate modeling

Multi-fidelity surrogates

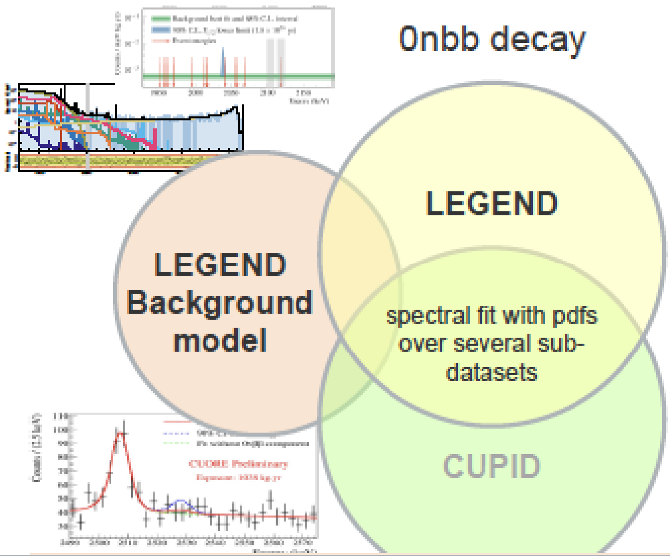
Multi-fidelity emulation of the **QGP** (Ji et al., 2024):

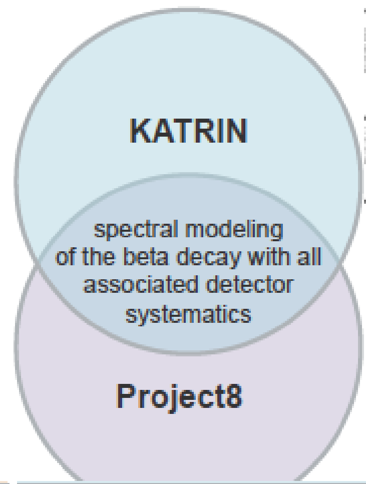


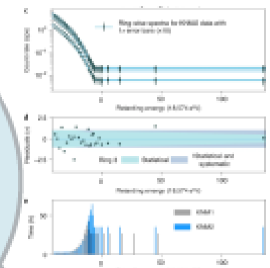
- $r = 2$ **fidelity** parameters (**spatial** mesh size, simulation **timestep**)

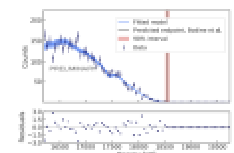


Bayesian parameter estimation and uncertainty quantification in Nuclear Science with focus on (Double) Beta decay









Bottlenecks:
- Increase of sub-datasets
- Marginalization over O(10^5) nuisance parameters describing detector systematics and bkg

Need:
- More performant parameter estimation tools and analysis strategies that will handle large datasets
- fit which incorporates additional informations into the likelihood (psd, veto signals,…

Bottlenecks:
- increasingly high dimensional data due to large number of detector systematics
- "Long" calculation time of detector response time

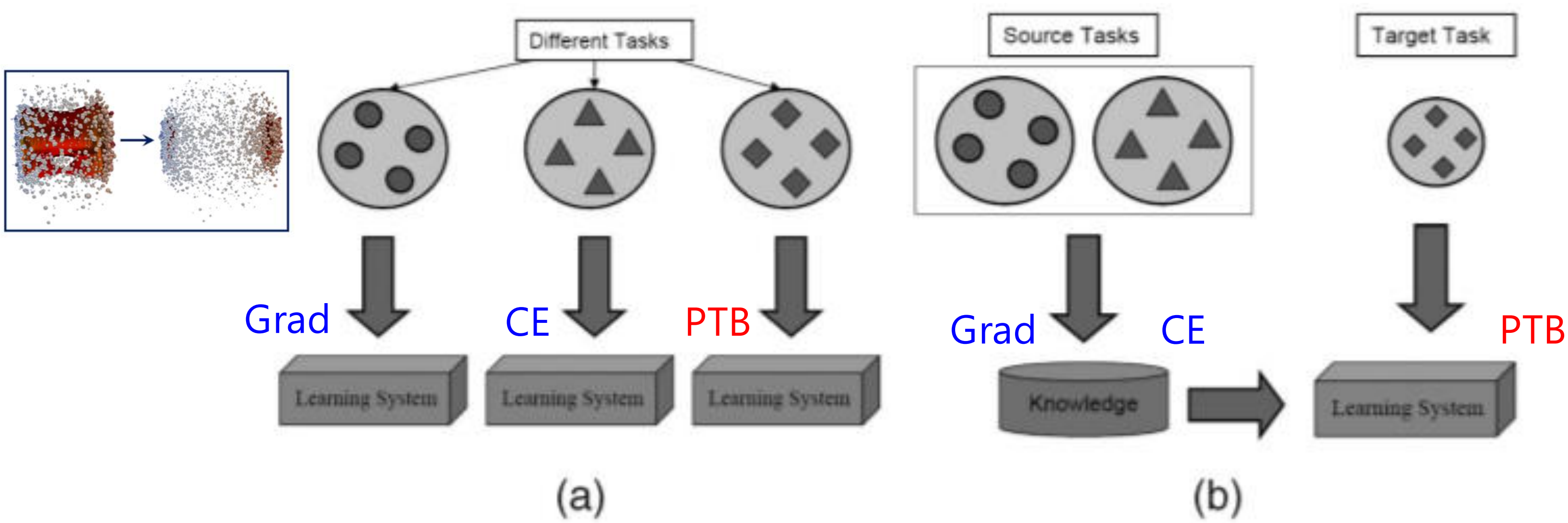
Need:
- differentiable model
- improved sampling tools e.j. gradient-based sampling
- multi-fidelity ML techniques

20 | Ann-Kathrin Schütz Bayesian uncertainty quantification for $0\nu\beta\beta$ decay 04/10/2024

Transfer learning surrogates

Transfer learning GPs (Liyanage et al., 2022 *PRC*; Wang et al., 2024+ *JUQ*):

- **Idea**: Using simulations on a **related** system (e.g., from previous analyses), apply **transfer learning** for **cost-efficient** surrogates on **target** system



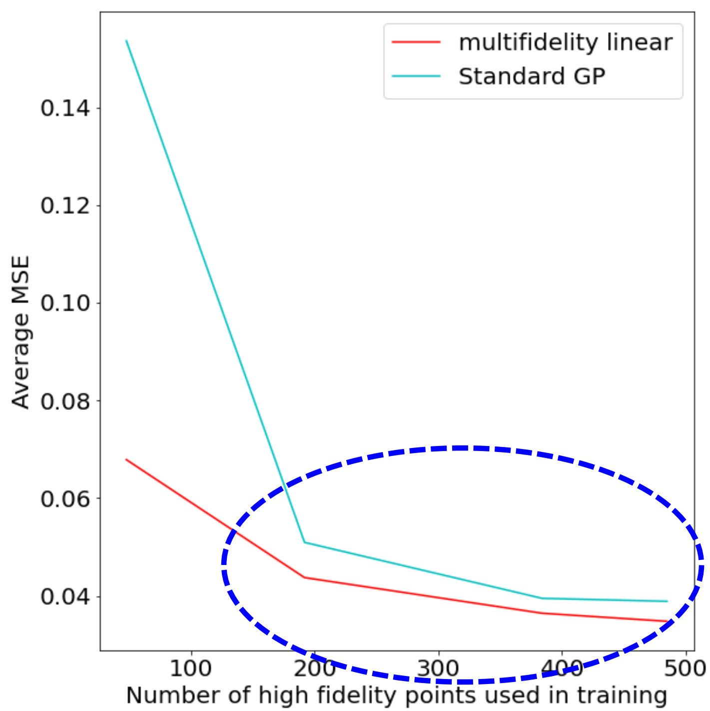
- **Comparably** accurate surrogates with **reduced** runs on **target** system
- … or more **accurate** surrogates with **comparable** runs on **target** system

Transfer learning surrogates

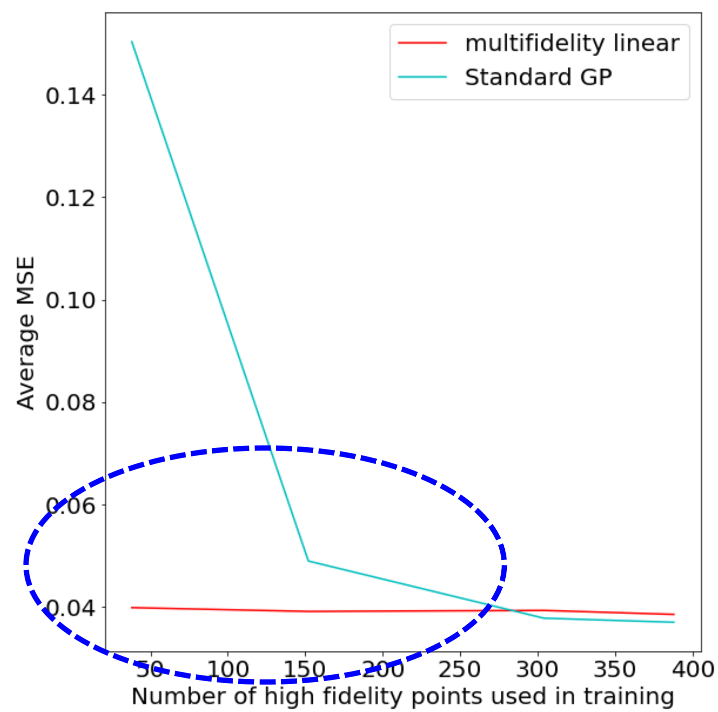
Liyanage et al. (2022 *PRC*):

- **Source**: Pb-Pb collisions at 2.76 TeV with Grad viscous correction

Target: Au+Au (Grad)

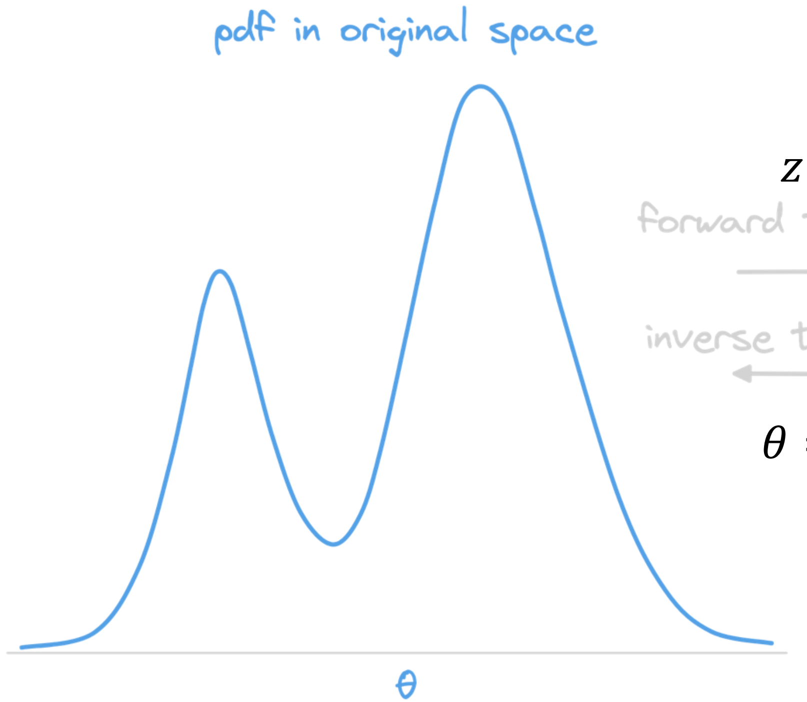


Target: Pb+Pb (CE)



More **accurate** surrogates at
reduced computational cost!

Normalizing flows

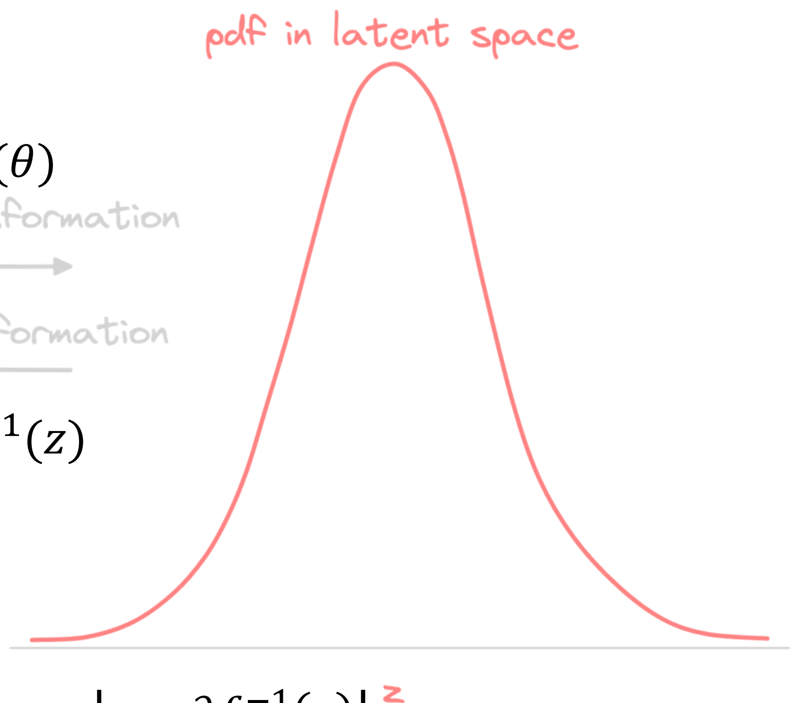


$$z = f(\theta)$$

forward transformation

inverse transformation

$$\theta = f^{-1}(z)$$



$$p_z(z) = p_\theta(f^{-1}(z)) \left| det \frac{\partial f^{-1}(z)}{\partial z} \right|$$

Normalizing flow preconditioning