The BAND Software Framework

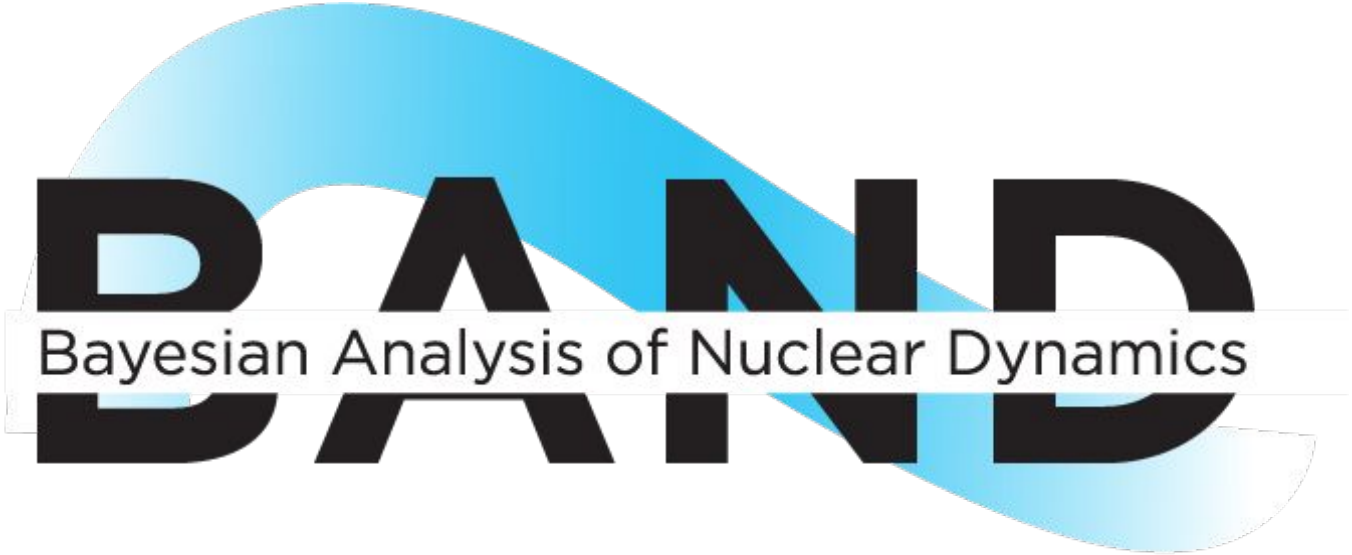
Kyle Godbey

Slides:
https://docs.google.com/presentation/d/17dKJUZav0h2AB87tdL6sU2SH1ijzjl13nczsyae9oi8/edit?usp=sharing

SANPC | 2024 SANPC Workshop
Newport News, VA

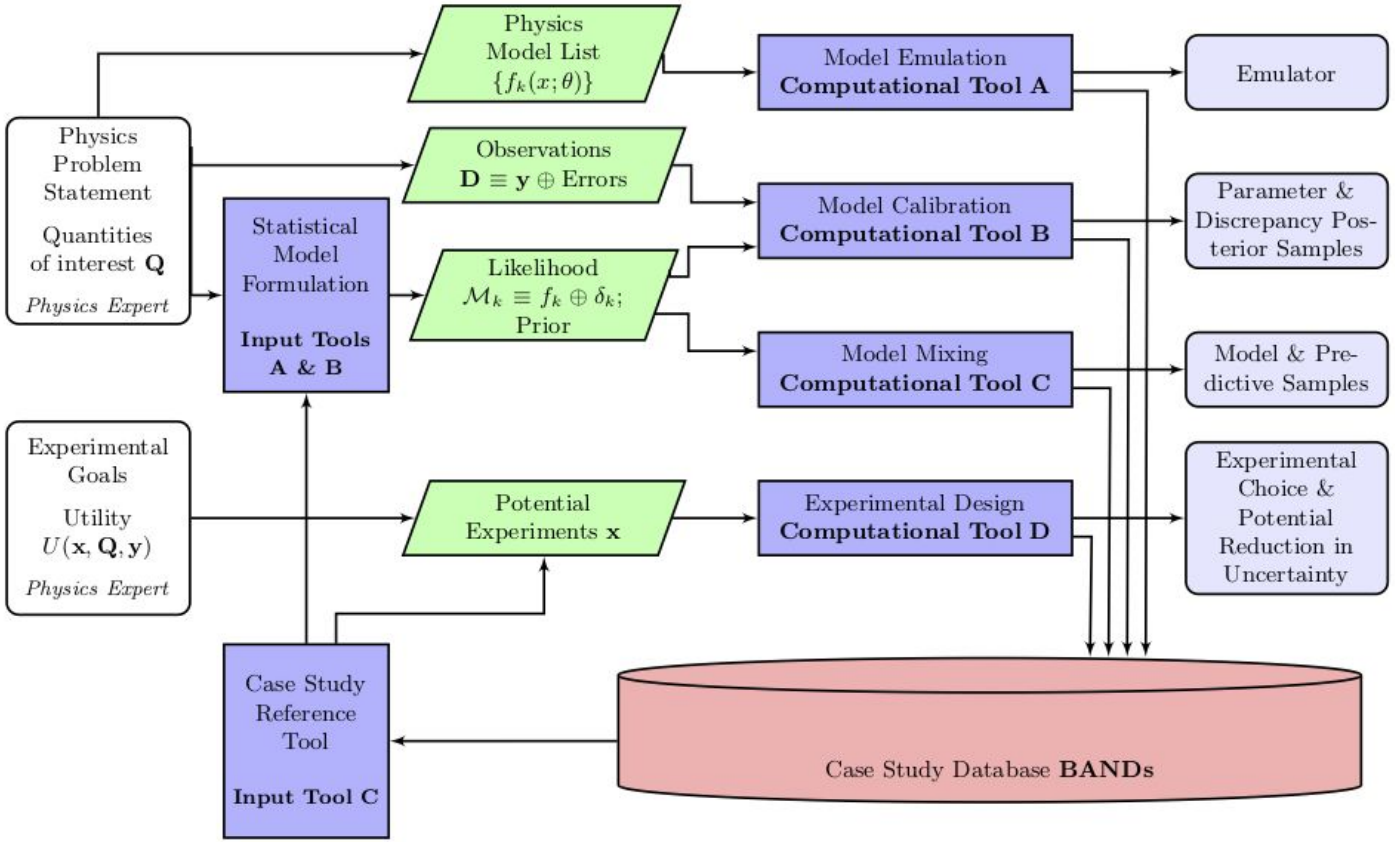
BAND
Bayesian Analysis of Nuclear Dynamics

Introducing: BAND

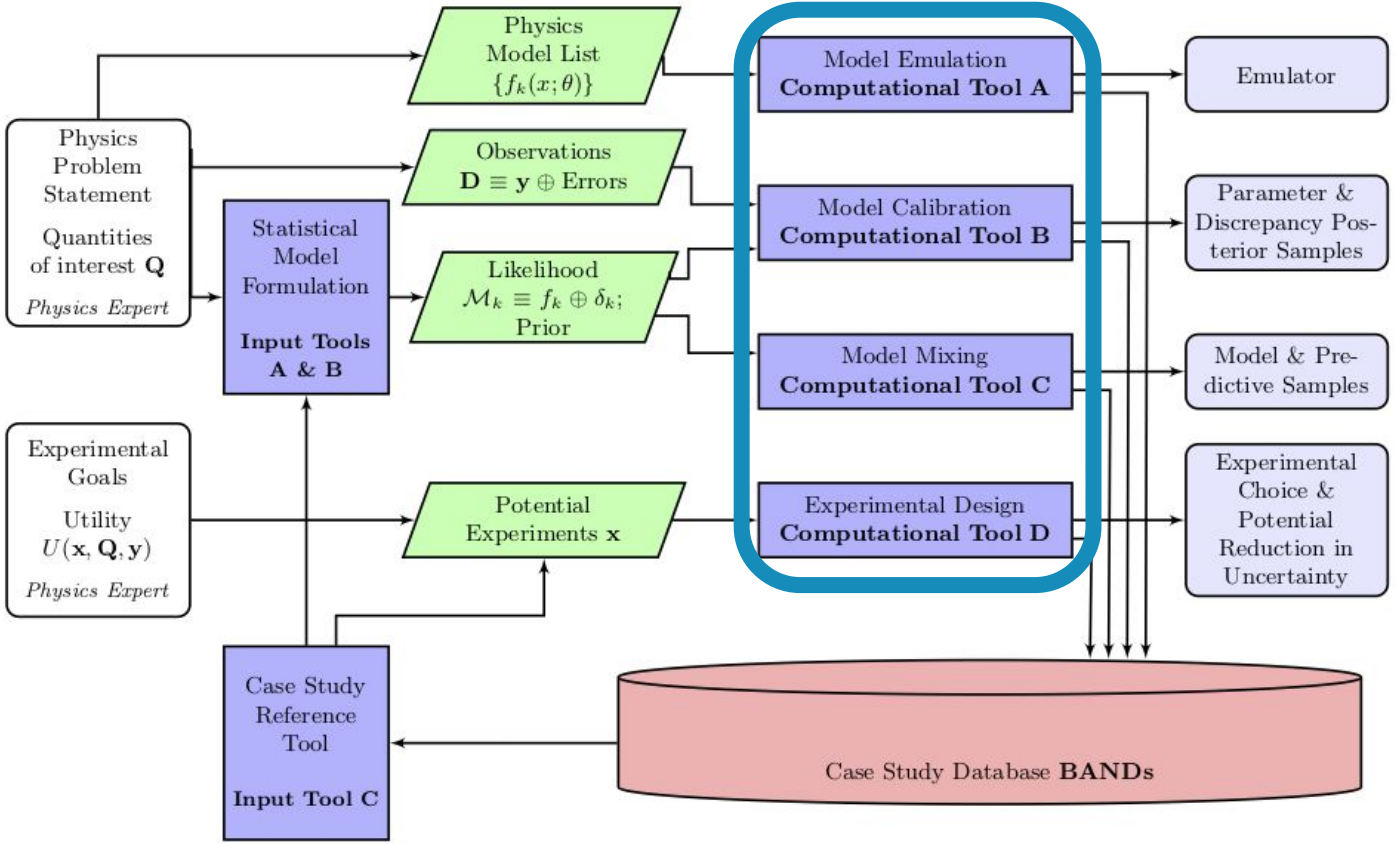


Supported by the NSF CSSI program under grant OAC-2004601

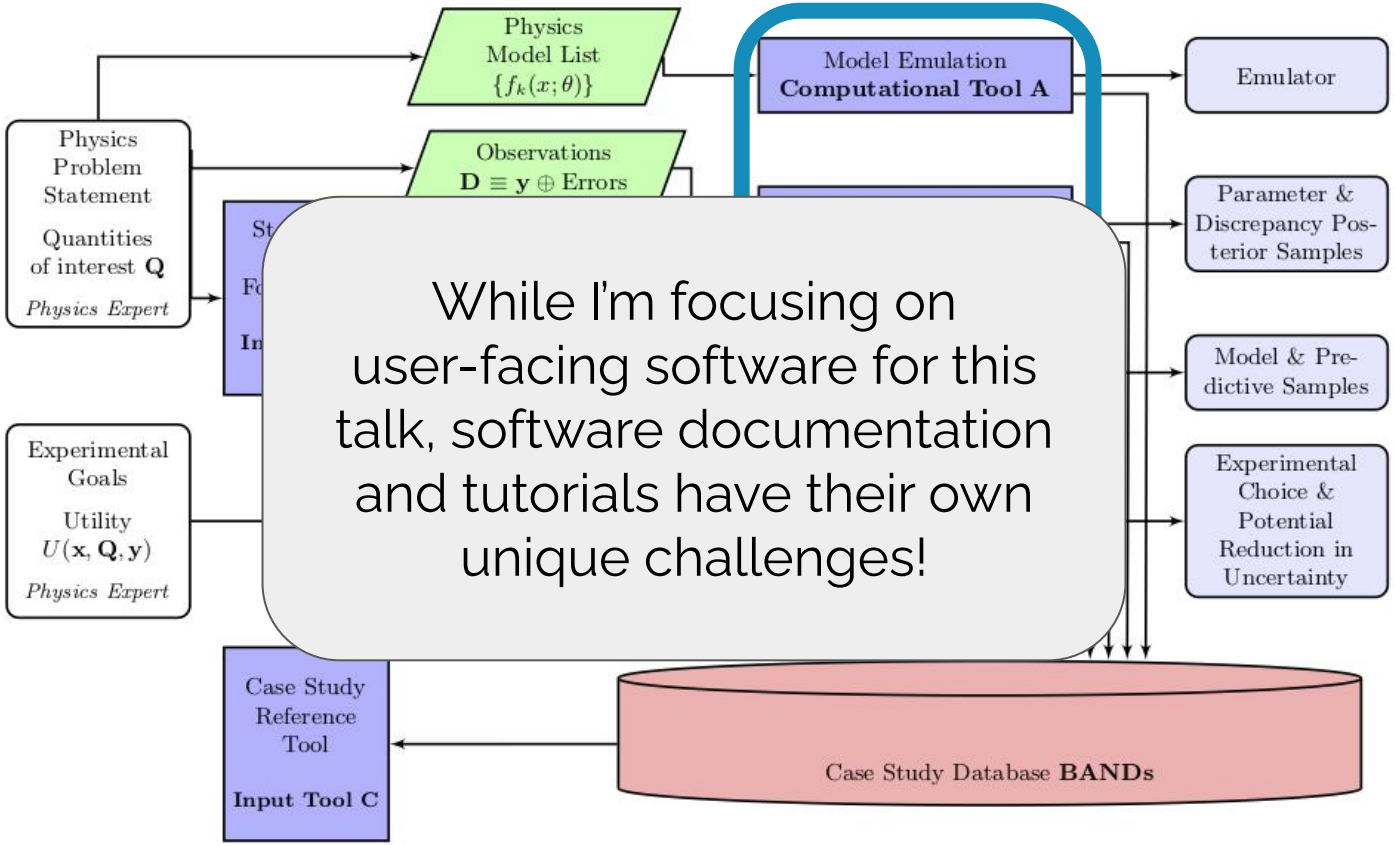
The Framework



The Framework



The Framework



While I'm focusing on user-facing software for this talk, software documentation and tutorials have their own unique challenges!

The Framework

Online:
https://bandframework.github.io/

On GitHub:
https://github.com/bandframework/bandframework

Software

External code delivery will be from the bandframework github repository

`surmise`

A Python package designed to provide a surrogate model interface for calibration, uncertainty quantification, and other tools.

O. Surer, M. Plumlee, S.M. Wild, M. Y-H. Chan
surmise Read the Docs

`Taweret`

A versatile Python package containing multiple model mixing techniques for a variety of use cases.

K. Ingles, D. Liyanage, A. C. Semposki, J. C. Yannotty
Taweret documentation

`SAMBA`

The SAndbox for Mixing using Bayesian Analysis, developed as a testing ground for multivariate model mixing on a toy model setup.

A. C. Semposki, R. J. Furnstahl, D. R. Phillips
SAMBA repository

`ParMOO`

ParMOO is a parallel multiobjective optimization solver that seeks to exploit simulation-based structure in objective and constraint functions.

T.H. Chang, S.M. Wild, H. Dickinson
parmoo Read the Docs

`BMEX`

The Bayesian Mass Explorer (BMEX) is a user-focused web application that provides a one-stop-shop for quantified theoretical model predictions of nuclear masses and related quantities.

K. Godbey, L. Buskirk, P. Giuliani
BMEX Web Application

`ROSE`

The Reduced Order Scattering Emulator (ROSE) is a Python package for building emulators using reduced basis methods for calculating nuclear scattering observables for user-defined interactions, including optical potentials.

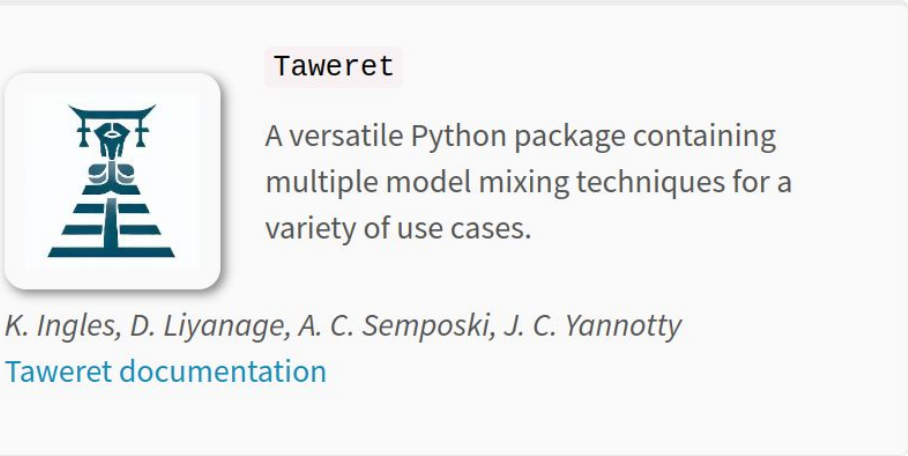
D. Odell, P. Giuliani, K. Godbey, K. Beyer, M. Y. Chan
ROSE Github

The Focus

Taweret - Generic Bayesian model mixing software with a variety of techniques and flexible data interface

The Focus

Taweret - Generic Bayesian model mixing software with a variety of techniques and flexible data interface



Point 1

A question to ask early and often: **is new software actually needed?**

JOSS
The Journal of Open Source Software

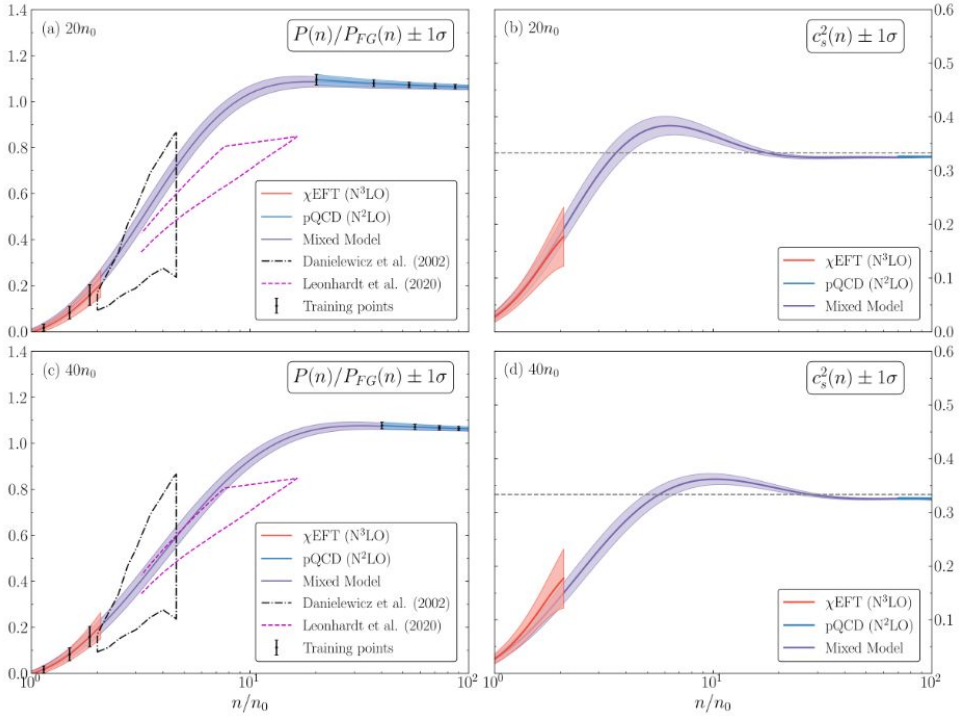
Taweret: a Python package for Bayesian model mixing

K. Ingles [1*¶], **D. Liyanage** [2*], **A. C. Semposki** [3*], **and J. C. Yannotty** [4*]

1 Illinois Center for Advanced Study of the Universe & Department of Physics, University of Illinois Urbana-Champaign, USA **2** Department of Physics, The Ohio State University, USA **3** Department of Physics and Astronomy & Institute of Nuclear and Particle Physics, Ohio University, USA **4** Department of Statistics, The Ohio State Univeristy, USA ¶ Corresponding author * These authors contributed equally.

From chiral EFT to perturbative QCD: a Bayesian model mixing approach to symmetric nuclear matter

A. C. Semposki [1,*] C. Drischler [1,2,†] R. J. Furnstahl [3,‡] J. A. Melendez [3,§] and D. R. Phillips [1,4,¶]



From chiral EFT to perturbative QCD: a Bayesian model mixing approach to symmetric nuclear matter

A. C. Semposki [1,*] C. Drischler [1,2,†] R. J. Furnstahl [3,‡] J. A. Melendez [3,§] and D. R. Phillips [1,4,¶]

Point 2

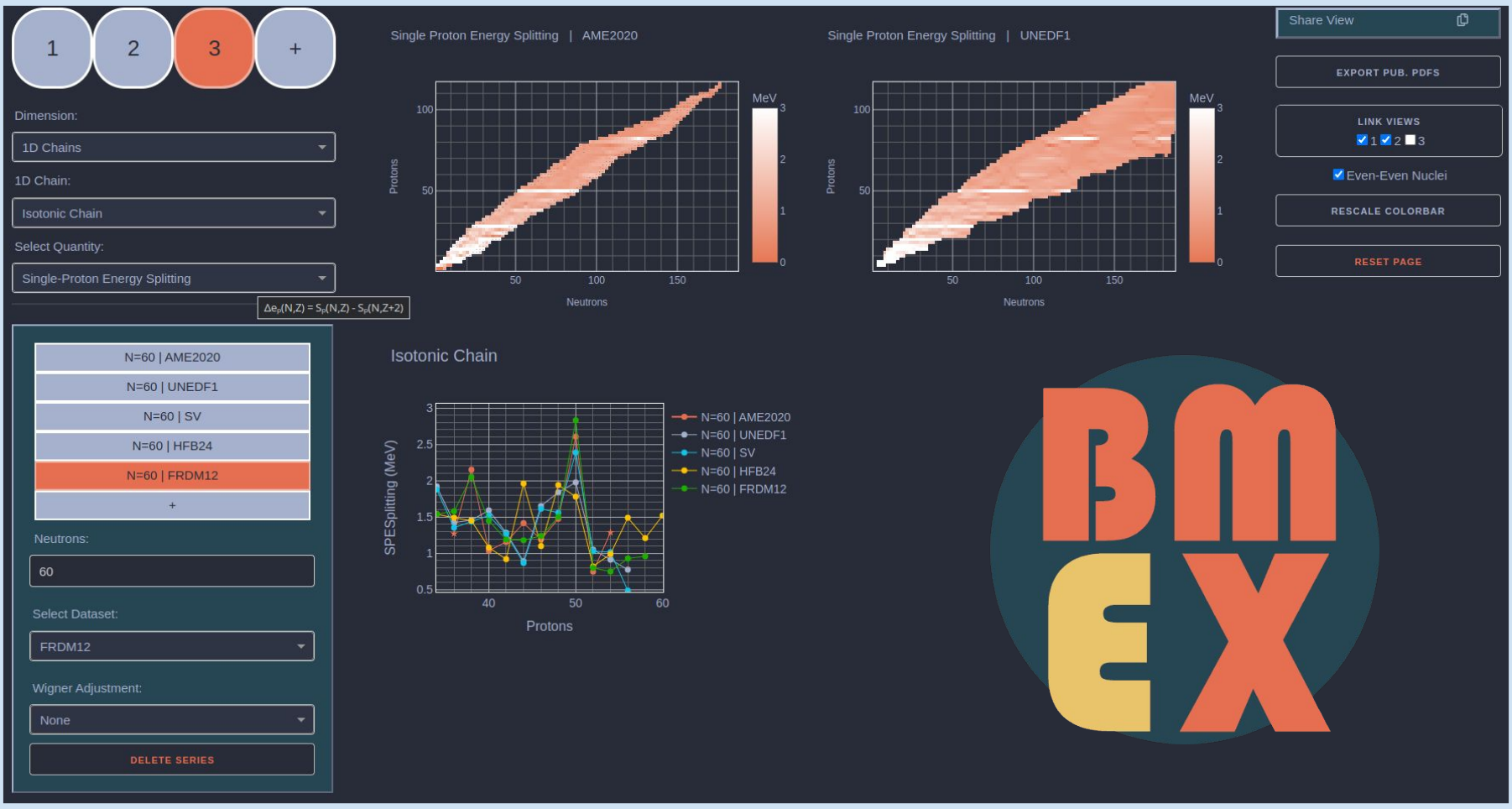
Our science goals drive development, but the needs for science applications and broad community engagement do not always align

From chiral EFT to perturbative QCD: a Bayesian model mixing approach to symmetric nuclear matter

A. C. Semposki [1,*] C. Drischler [1,2,†] R. J. Furnstahl [3,‡] J. A. Melendez [3,§] and D. R. Phillips [1,4,¶]

Point 3

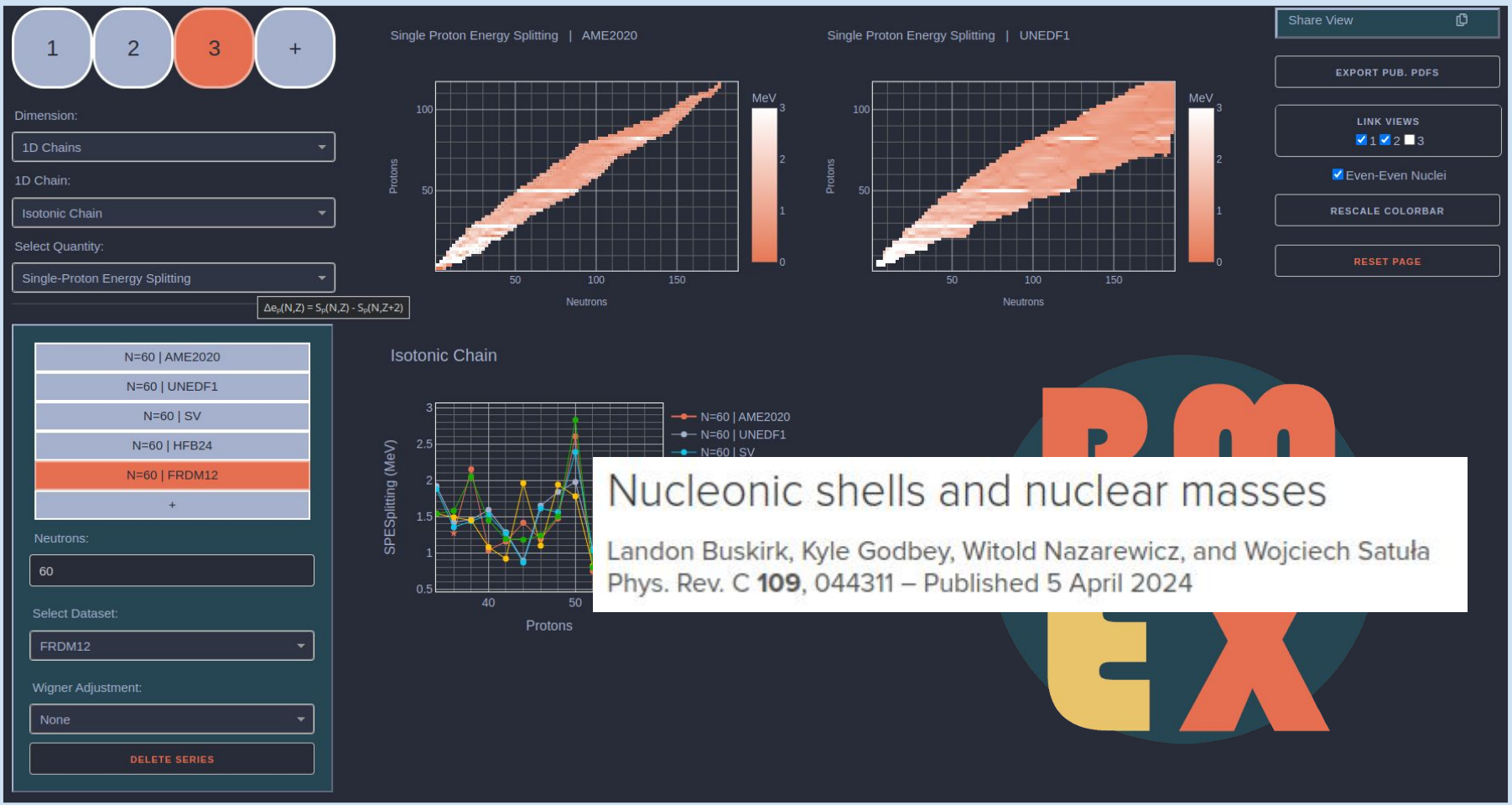
We should think of ways to ensure that developing useful research software is given the respect it warrants - particularly when we consider most development is done by junior colleagues



SANPC | 2024 SANPC Workshop
Newport News, VA

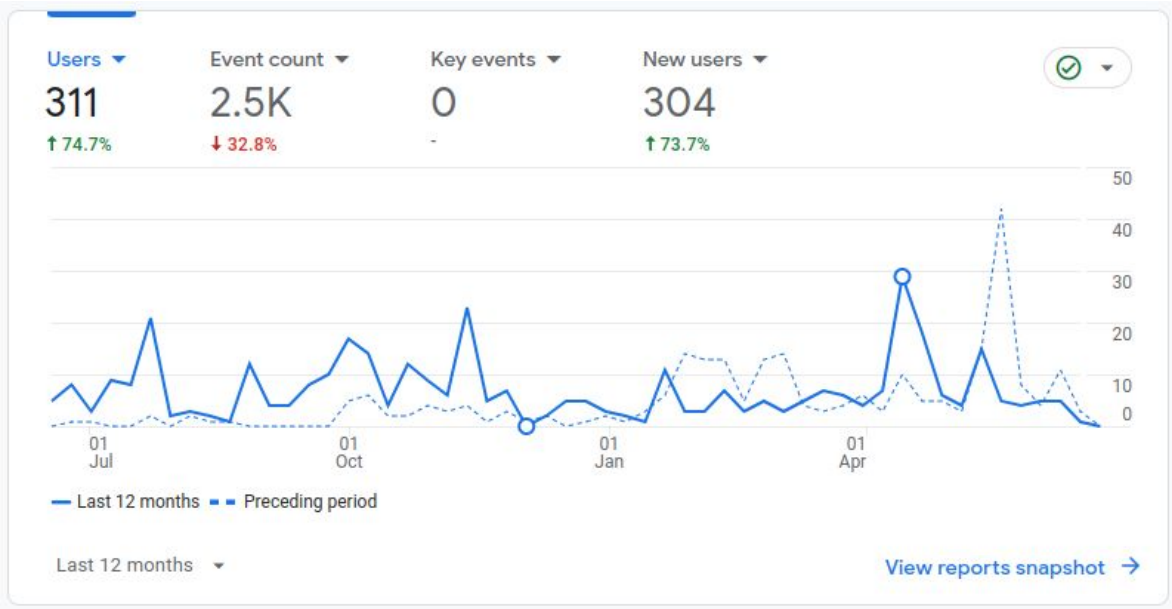
https://bmex.dev

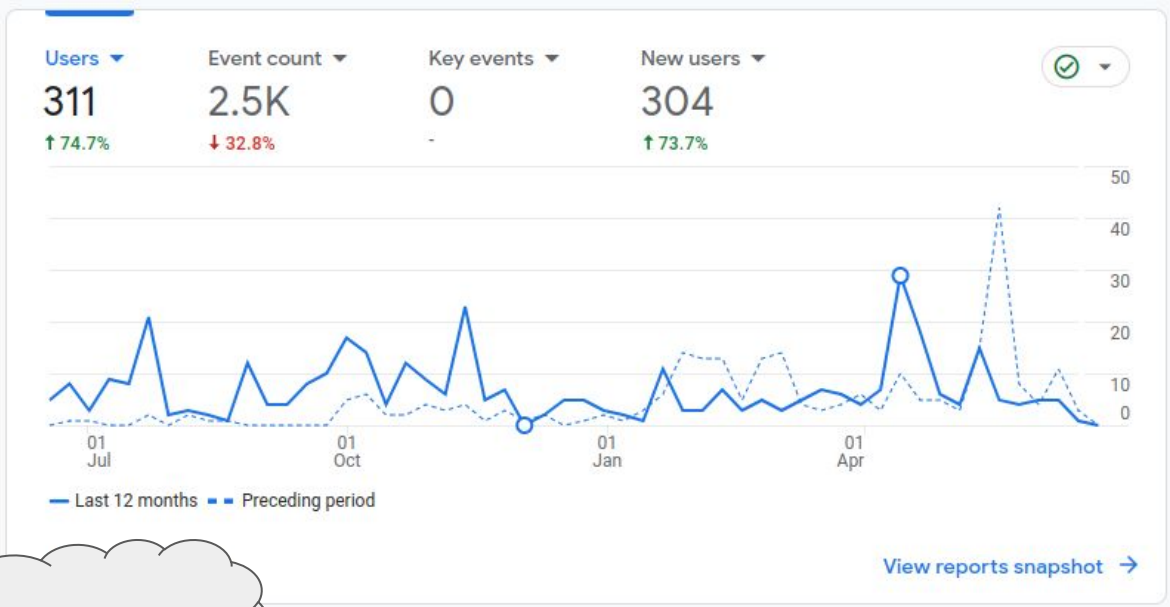
BAND
Bayesian Analysis of Nuclear Dynamics



Nucleonic shells and nuclear masses

Landon Buskirk, Kyle Godbey, Witold Nazarewicz, and Wojciech Satuła
Phys. Rev. C **109**, 044311 – Published 5 April 2024


Users
311
↑ 74.7%
Event count
2.5K
↓ 32.8%
Key events
0
-
New users
304
↑ 73.7%
50
40
30
20
10
0
01 Jul
01 Oct
01 Jan
01 Apr
Last 12 months
Preceding period
Last 12 months
View reports snapshot

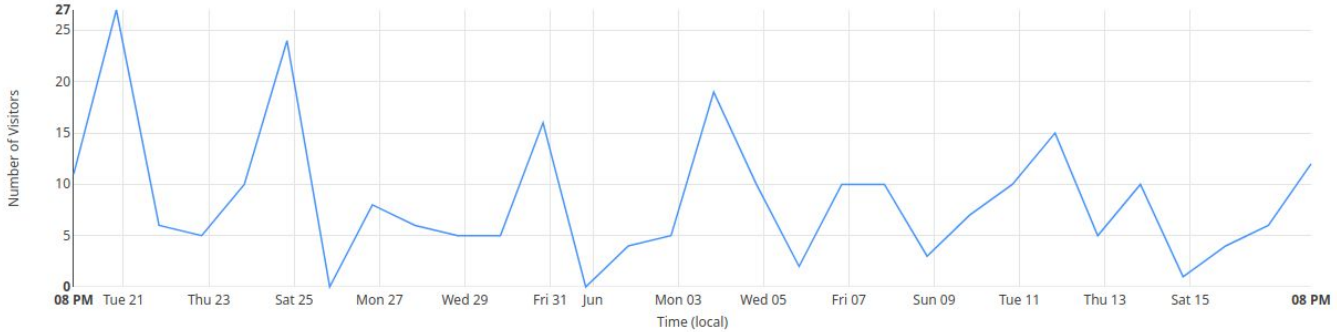


Point 4

How do we make this sustainable and enable future growth?

Unique Visitors

Total Unique Visitors	Maximum Unique Visitors	Minimum Unique Visitors
Previous 30 days	Per day	Per day
234	**27**	**0**

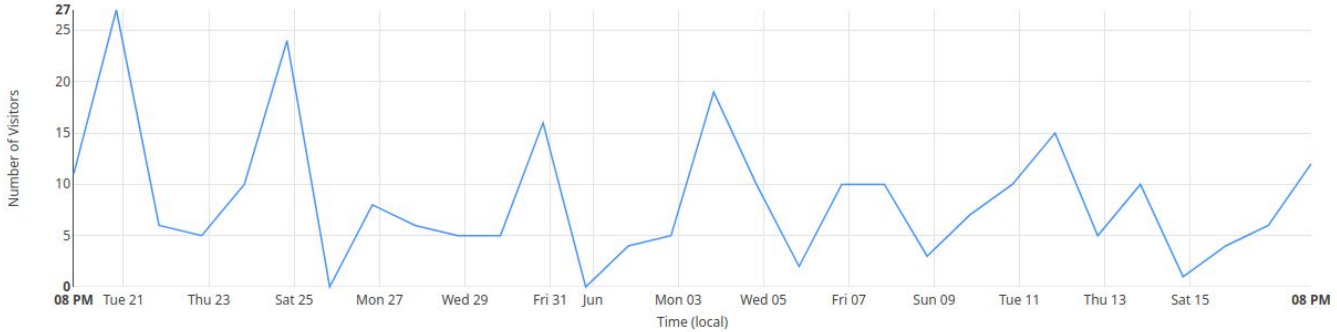


Point 4

How do we make this sustainable and enable future growth?

Unique Visitors

Total Unique Visitors	Maximum Unique Visitors	Minimum Unique Visitors
Previous 30 days	Per day	Per day
234	**27**	**0**



Point 5

This issue includes feature developments! Needs evolve and software should too.

Conclusions and Outlook

In short, research software faces challenges at all stages of development from conception to execution and from release to support and maintenance

The utopian solution would be a team of software engineers devoted to upkeep and development in perpetuity, but that's not realistic or warranted in most cases

Conclusions and Outlook

We want to build a broadly useful, open-source framework that meets the needs of the community

Conclusions and Outlook

We want to build a broadly useful, open-source framework that meets the needs of the community

The first challenge is even defining “community”, “needs”, and “useful”

Conclusions and Outlook

We want to build a broadly useful, open-source framework that meets the needs of the community

The first challenge is even defining “community”, “needs”, and “useful”

A sustainable framework should have a well defined scope from the beginning and a plan for growth beyond that scope

Conclusions and Outlook

Once the science goals are identified and software is starting to be developed, we need ways to ensure best practices are being adhered to

Conclusions and Outlook

Once the science goals are identified and software is starting to be developed, we need ways to ensure best practices are being adhered to

Maintenance is never trivial, but it's much easier if you're starting from a good base with well defined structures and policies

Conclusions and Outlook

At a **minimum** this should mean developers have access to training materials that are more specific for our needs in research software

Conclusions and Outlook

At a **minimum** this should mean developers have access to training materials that are more specific for our needs in research software

We all have a shared responsibility in trying to meet this need!

Conclusions and Outlook

Finally, some serious thought needs to be put into software sustainability and stewardship, if only for the reproducibility aspect

Conclusions and Outlook

Finally, some serious thought needs to be put into software sustainability and stewardship, if only for the reproducibility aspect

Reliance on free tiers of commercial offerings for source management and build systems is a **risk**, even if it's a small one right now

Conclusions and Outlook

If the software in question is a **service**, then things get even more complicated (and potentially expensive)

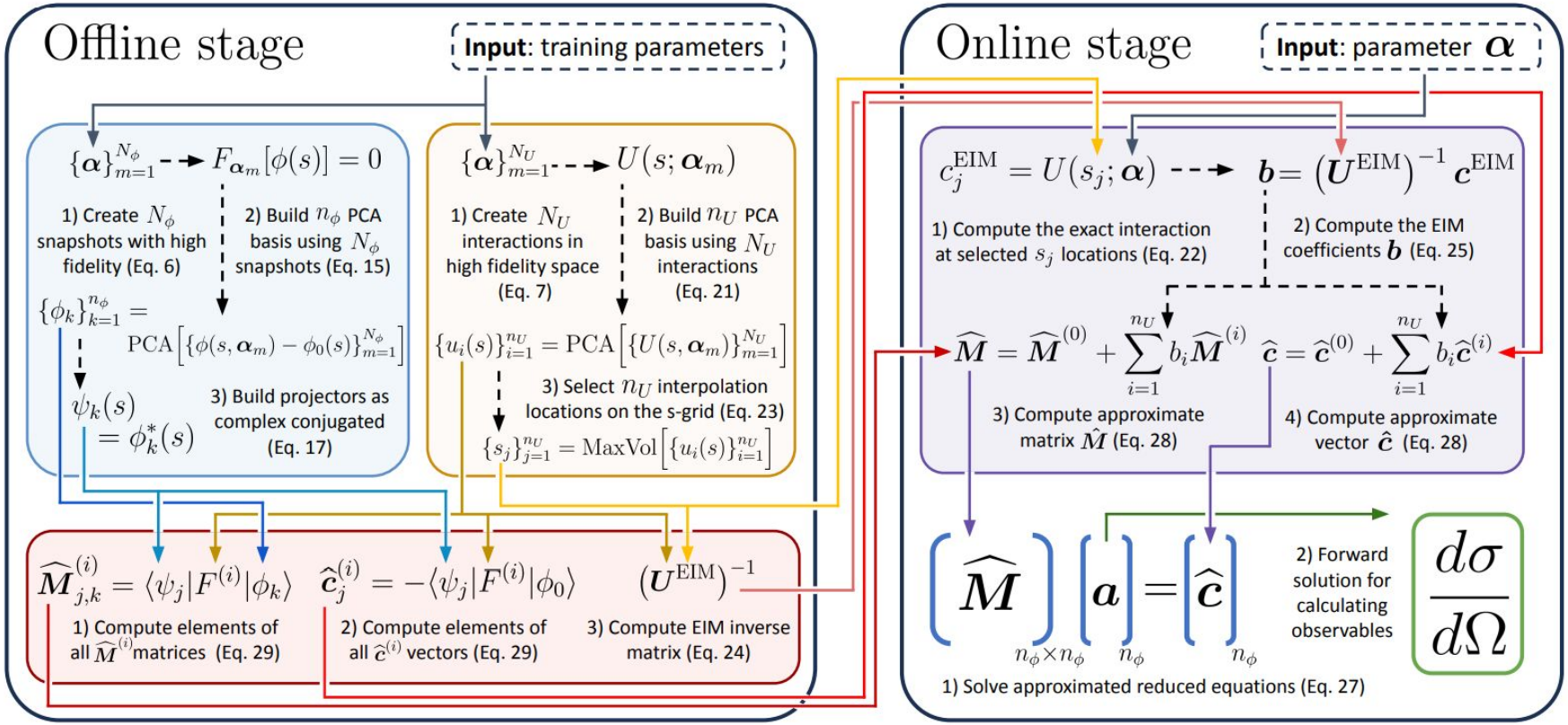
Conclusions and Outlook

If the software in question is a **service**, then things get even more complicated (and potentially expensive)

Systems like Jetstream2 are exactly what's needed, but a longer term allocation scheme would make it a firmer foundation for research infrastructure

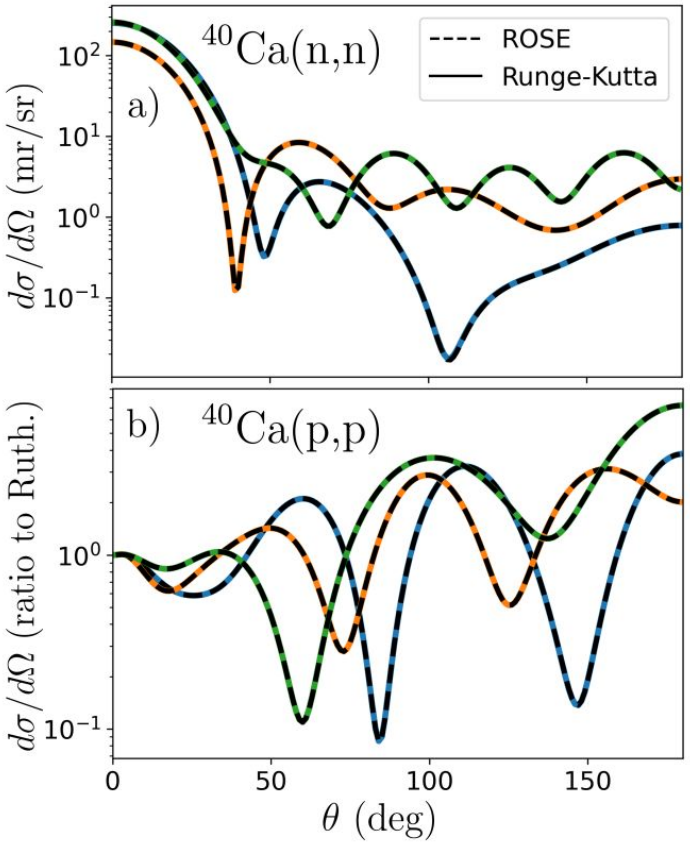
Thanks!

The Focus

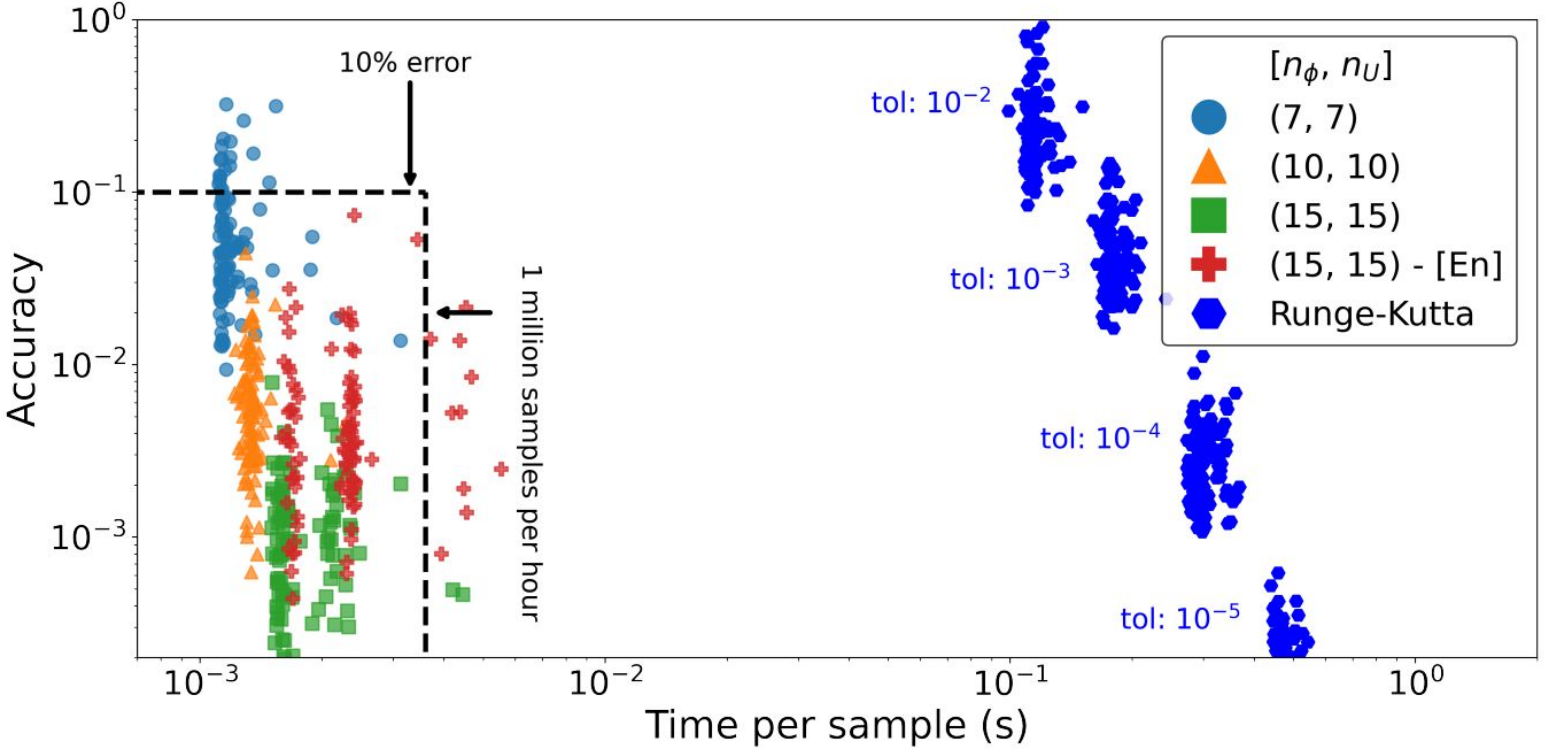

Offline stage
Input: training parameters
$\{\boldsymbol{\alpha}\}_{m=1}^{N_\phi} \dashrightarrow F_{\boldsymbol{\alpha}_m}[\phi(s)] = 0$
1) Create N_ϕ snapshots with high fidelity (Eq. 6)
2) Build n_ϕ PCA basis using N_ϕ snapshots (Eq. 15)
$\{\phi_k\}_{k=1}^{n_\phi} = \mathrm{PCA}\left[\{\phi(s,\boldsymbol{\alpha}_m) - \phi_0(s)\}_{m=1}^{N_\phi}\right]$
$\psi_k(s) = \phi_k^*(s)$
3) Build projectors as complex conjugated (Eq. 17)
$\{\boldsymbol{\alpha}\}_{m=1}^{N_U} \dashrightarrow U(s;\boldsymbol{\alpha}_m)$
1) Create N_U interactions in high fidelity space (Eq. 7)
2) Build n_U PCA basis using N_U interactions (Eq. 21)
$\{u_i(s)\}_{i=1}^{n_U} = \mathrm{PCA}\left[\{U(s,\boldsymbol{\alpha}_m)\}_{m=1}^{N_U}\right]$
3) Select n_U interpolation locations on the s-grid (Eq. 23)
$\{s_j\}_{j=1}^{n_U} = \mathrm{MaxVol}\left[\{u_i(s)\}_{i=1}^{n_U}\right]$
$\widehat{\boldsymbol{M}}_{j,k}^{(i)} = \langle\psi_j|F^{(i)}|\phi_k\rangle$
1) Compute elements of all $\widehat{M}^{(i)}$ matrices (Eq. 29)
$\widehat{\boldsymbol{c}}_j^{(i)} = -\langle\psi_j|F^{(i)}|\phi_0\rangle$
2) Compute elements of all $\widehat{c}^{(i)}$ vectors (Eq. 29)
$\left(\boldsymbol{U}^{\mathrm{EIM}}\right)^{-1}$
3) Compute EIM inverse matrix (Eq. 24)
Online stage
Input: parameter $\boldsymbol{\alpha}$
$c_j^{\mathrm{EIM}} = U(s_j;\boldsymbol{\alpha}) \dashrightarrow \boldsymbol{b} = \left(\boldsymbol{U}^{\mathrm{EIM}}\right)^{-1}\boldsymbol{c}^{\mathrm{EIM}}$
1) Compute the exact interaction at selected s_j locations (Eq. 22)
2) Compute the EIM coefficients $\boldsymbol{b}$ (Eq. 25)
$\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{M}}^{(0)} + \sum_{i=1}^{n_U} b_i \widehat{\boldsymbol{M}}^{(i)}$
3) Compute approximate matrix $\hat{\boldsymbol{M}}$ (Eq. 28)
$\widehat{\boldsymbol{c}} = \widehat{\boldsymbol{c}}^{(0)} + \sum_{i=1}^{n_U} b_i \widehat{\boldsymbol{c}}^{(i)}$
4) Compute approximate vector $\hat{\boldsymbol{c}}$ (Eq. 28)
$\widehat{\boldsymbol{M}}\ \boldsymbol{a} = \widehat{\boldsymbol{c}}$
$n_\phi \times n_\phi$ n_ϕ n_ϕ
1) Solve approximated reduced equations (Eq. 27)
2) Forward solution for calculating observables
$\frac{d\sigma}{d\Omega}$

The Goal

Supplementing high-fidelity simulations with ‘almost as good’ replacements

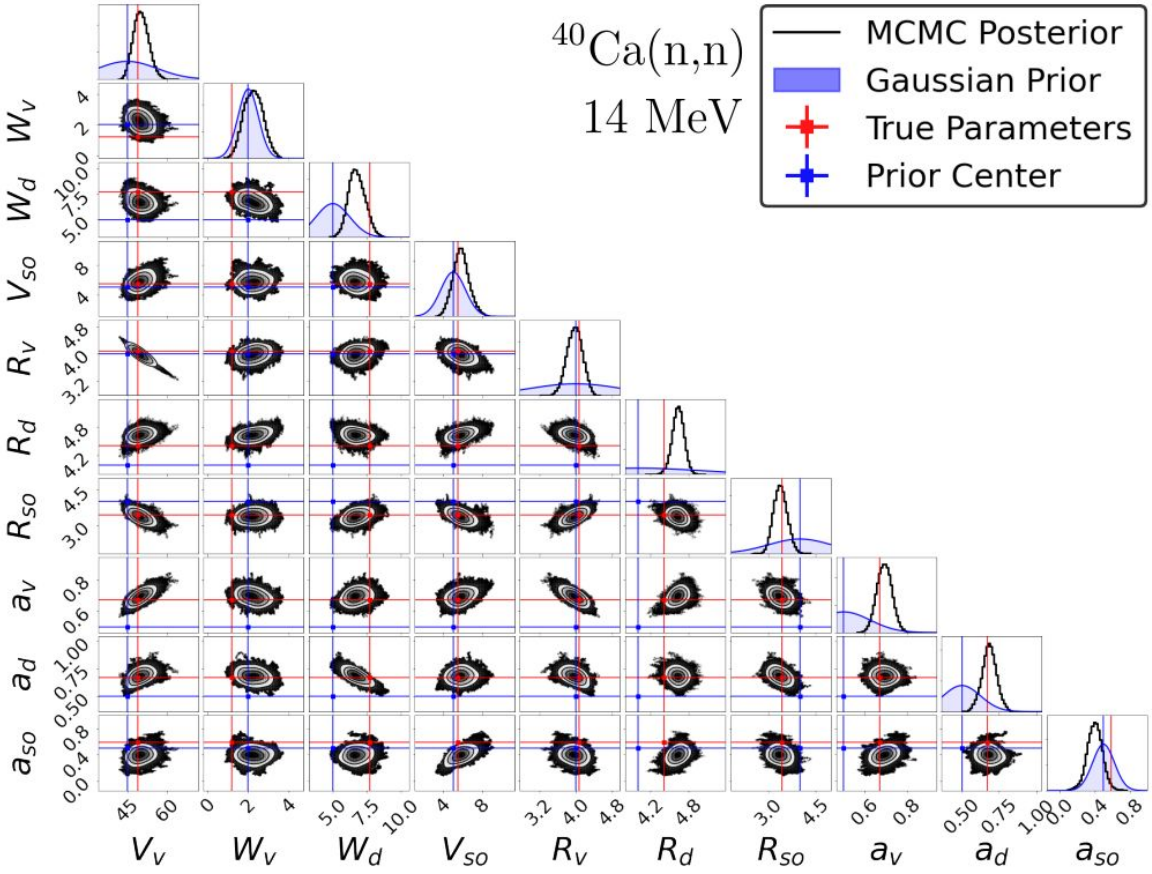


The Goal


10% error
1 million samples per hour
Accuracy
Time per sample (s)
tol: 10^{-2}
tol: 10^{-3}
tol: 10^{-4}
tol: 10^{-5}
$[n_\phi, n_U]$
(7, 7)
(10, 10)
(15, 15)
(15, 15) - [En]
Runge-Kutta

The Outcome

Principled Bayesian calibration is more accessible than ever before

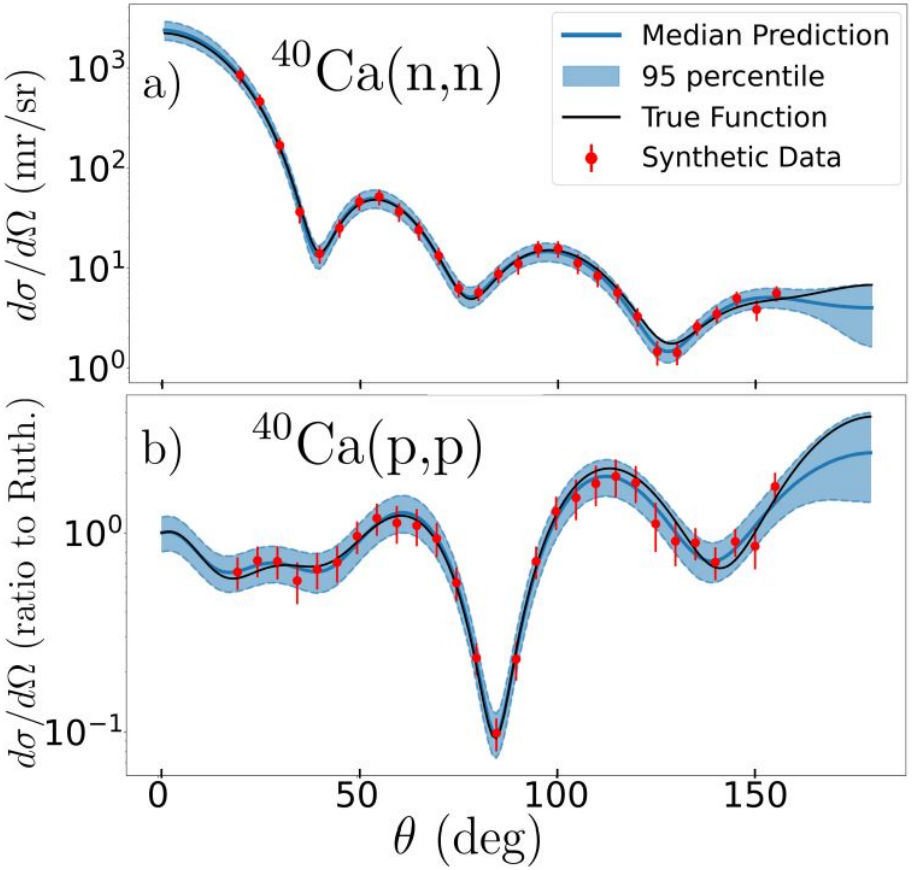


BAND
Bayesian Analysis of Nuclear Dynamics

The Outcome

The point is to get predictions, uncertainties, and covariances quickly

Integration with model mixing tools will help consider the wisdom of many models



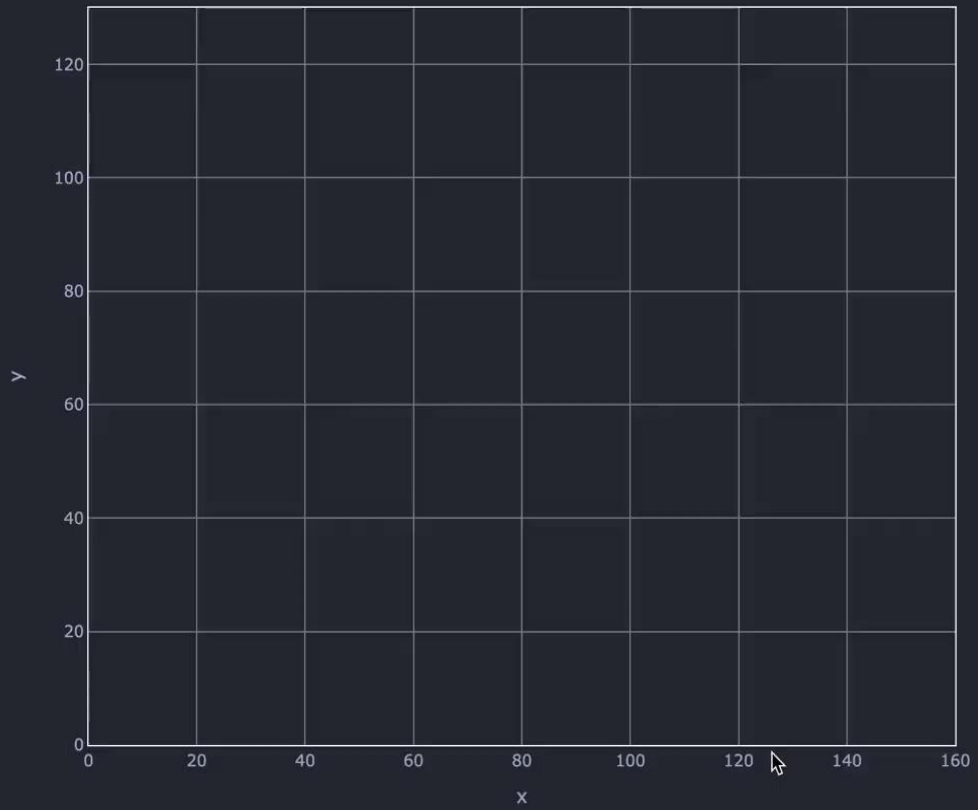
Outlook

- Continued feature development in both physics and emulation
  - Coupled-channels emulation
  - Black-box dimensionality reduction
- Expanded integration within and without the BAND framework
  - Bayesian model mixing
- Explore novel applications and user interaction modality
  - Cloud-enabled backends
  - Web applications

FIT

CLEAR

	x	y	δy
x			



Our Request

Continued feature development in both physics and emulation

What features are needed? What models are highest impact?

Expanded integration within and without the BAND framework

What does harmonious integration look like?

Explore novel applications and user interaction modality

What **tools** have the highest potential for impact across the pipeline?