ML-Based Surrogate Modeling of Radiofrequency Quadrupole Accelerators

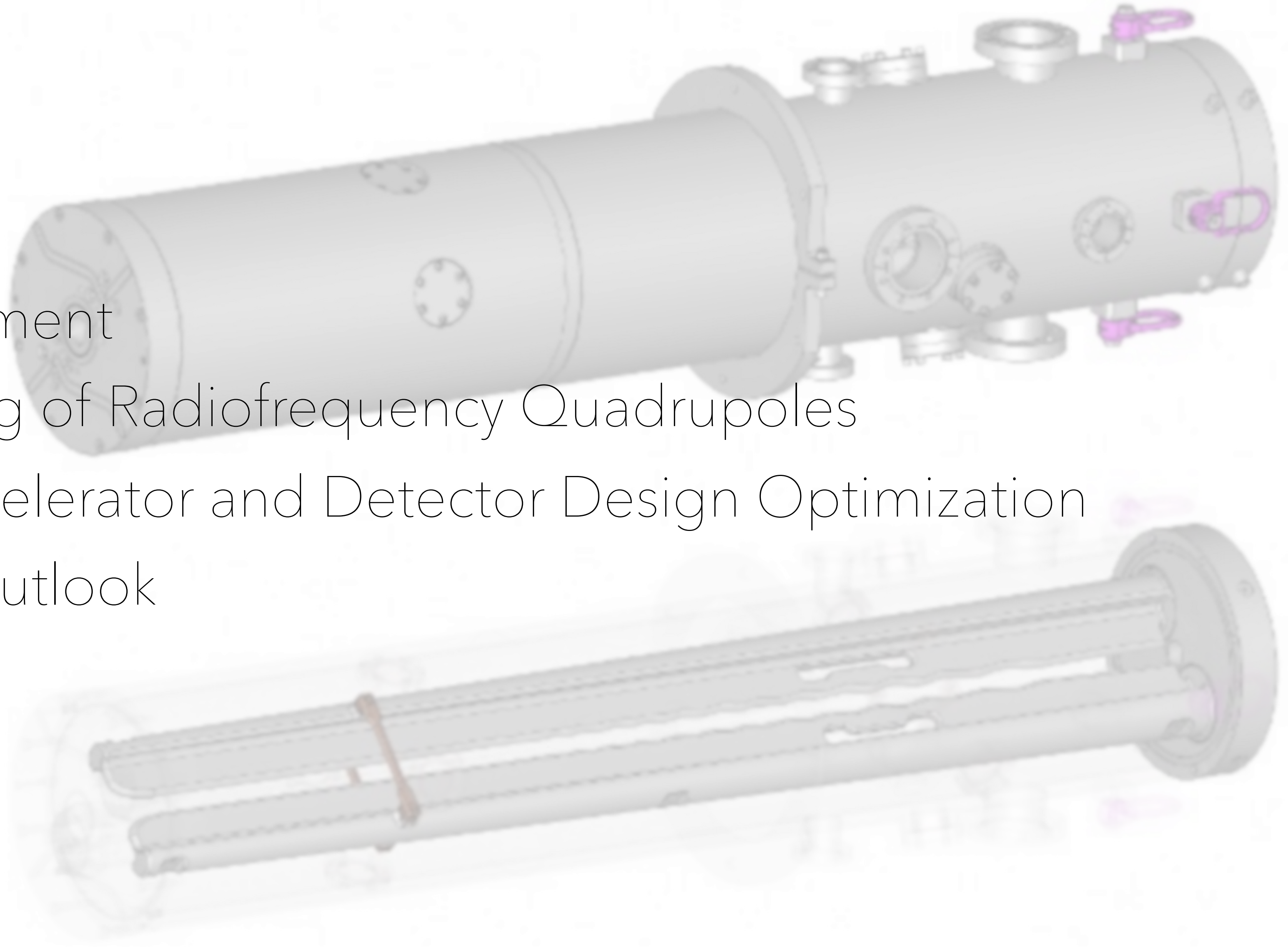
Joshua Villarreal

Neutrino Physics and Machine Learning

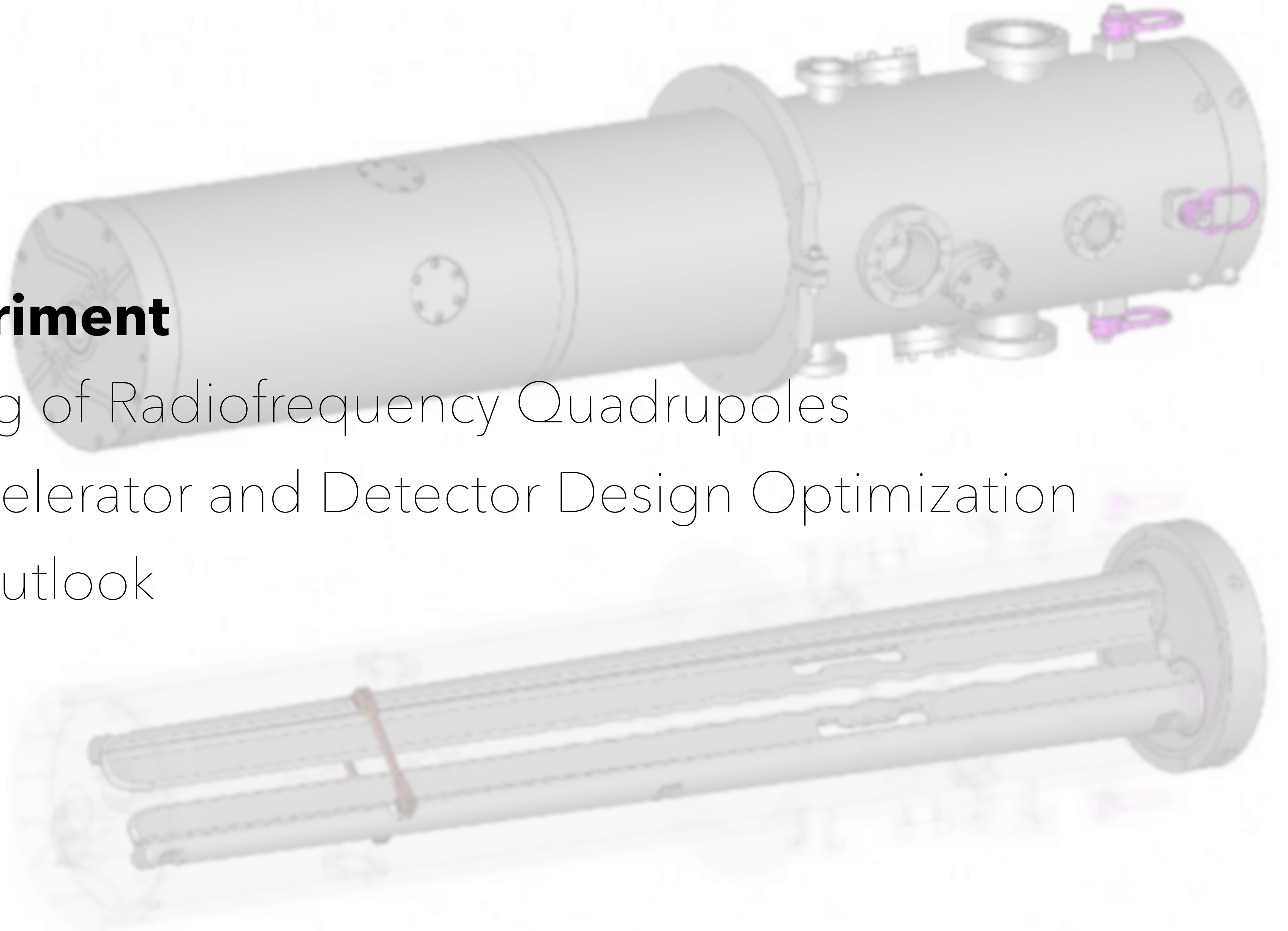
Wednesday, August 23, 2023

Outline

1. The IsoDAR Experiment
2. Surrogate Modeling of Radiofrequency Quadrupoles
3. Applications to Accelerator and Detector Design Optimization
4. Conclusions and Outlook

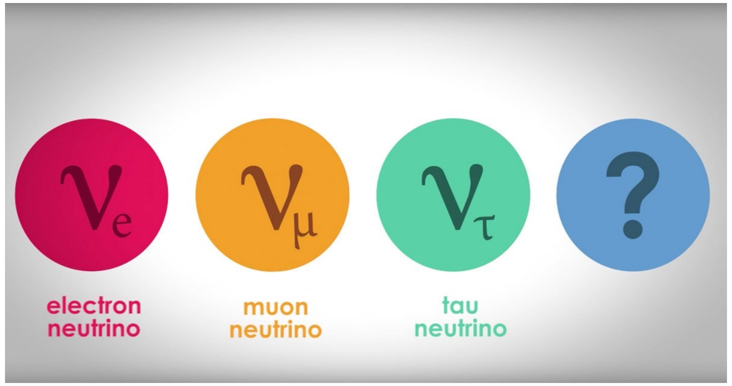


Outline



1. **The IsoDAR Experiment**
2. Surrogate Modeling of Radiofrequency Quadrupoles
3. Applications to Accelerator and Detector Design Optimization
4. Conclusions and Outlook

Introduction
An Ideal Sterile Neutrino Search



Consider the two-neutrino oscillation probability model

$$P = 1 - \sin^2 2\theta \sin^2[1.27\, \Delta m^2\, (L/E)]$$

- Disappearance/appearance probability is maximized when $1.27\, \Delta m^2\, (L/E) \approx \pi/2$, or when $\Delta m^2 \approx E/L$
- Electron antineutrino anomalies (both excesses and deficits) have been observed at neutrino experiments worldwide, and are thought to be signatures of oscillations of neutrinos into sterile neutrinos
- These anomalies tend to concentrate around the oscillation region of $\Delta m^2 \sim \mathcal{O}(1\ \mathrm{eV}^2)$
- In a perfect world…
  - Neutrino source has a well-understood production energy and flux
  - Large-volume detector with impressive sensitivity to neutrino interactions in the scintillation volume
  - $E/L \sim \mathcal{O}(1\ \mathrm{eV}^2)$ to give best chance of seeing neutrinos oscillating into a sterile state
  - High enough neutrino flux to achieve statistics at the 5σ level

Phys. Rev. Lett. 109 (2012) 141802

https://www.nevis.columbia.edu/daedalus/motiv/sterile.html

Introduction
An Ideal Sterile Neutrino Search

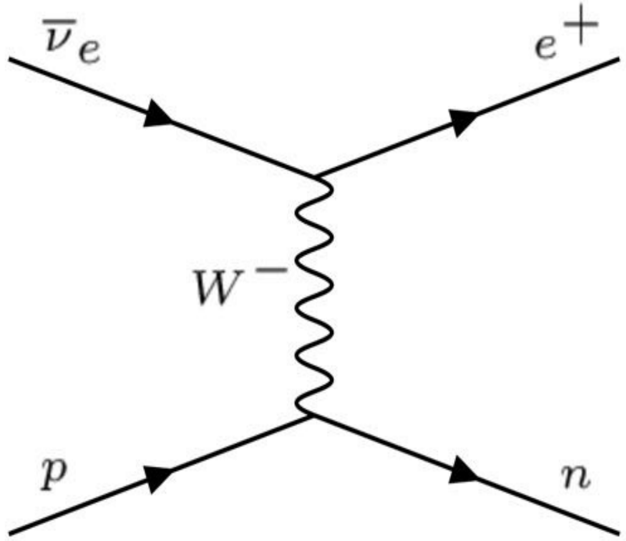
- In a perfect world… we have IsoDAR
  - Neutrino source has a well-understood production energy and flux

    The isotropic beta decay of ^{8}Li, where $\overline{\nu}_e$ have an average energy of 6.4 MeV.
  - Large-volume detector with impressive sensitivity to neutrino interactions in the scintillation volume

    An underground liquid scintillator that can detect IBD interactions peaking at an antineutrino energy of 9 MeV
  - $E/L \sim \mathcal{O}(1\ \mathrm{eV}^2)$ to give best chance of seeing neutrinos oscillating into a sterile state

    An antineutrino detector with a diameter $\sim\mathcal{O}(10\ \mathrm{m})$
  - High enough neutrino flux to achieve statistics at the 5σ level

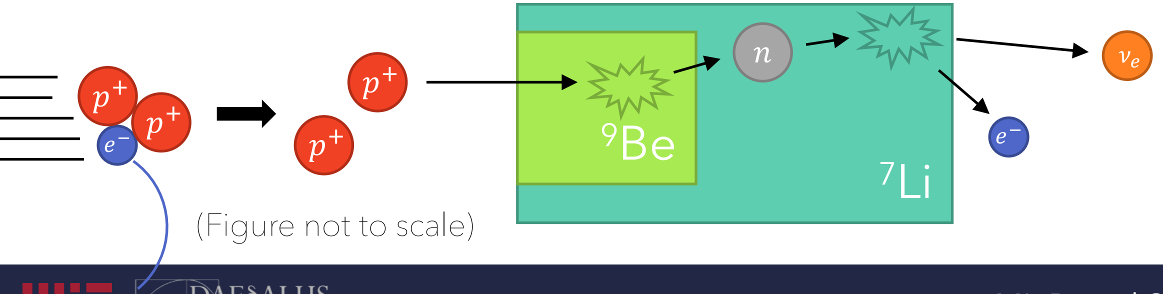
    Constantly produce ^{8}Li by bombarding ^{7}Li with neutrons, which we can produce by irradiating ^{9}Be with 10 mA of 60 MeV protons



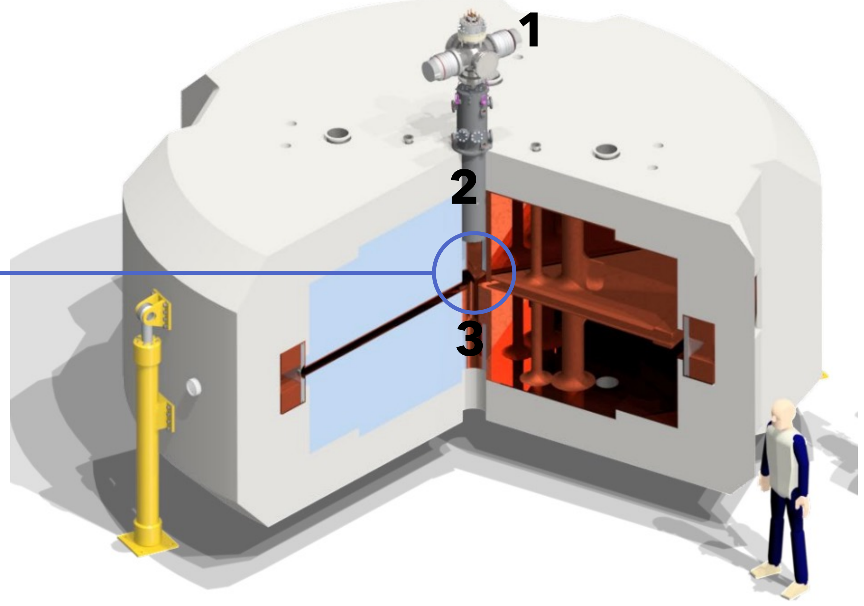
Phys. Rev. Lett. 109 (2012) 141802

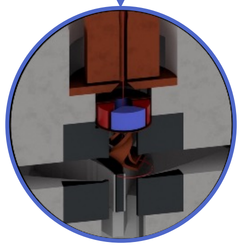
The IsoDAR Experiment
Overview

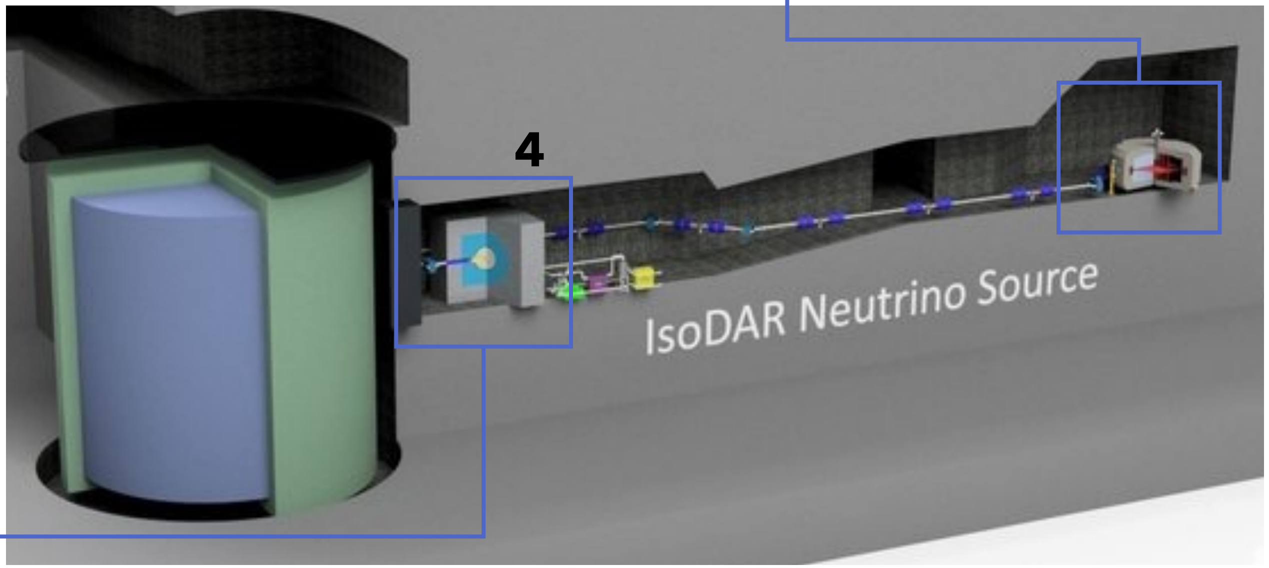
- Installation planned at Yemilab beside a kiloton-scale liquid scintillator
- IsoDAR is a proposed $\bar{\nu}_e$ source in which:
    1. H_2 plasma is created in the ion source
    2. Beam is pre-accelerated, focused, and bunched in the radiofrequency quadrupole (RFQ)
    3. 5 mA beam is injected into cyclotron & accel. to 60 MeV
    4. Electrons are dissociated from H_2^+, resulting 10 mA of 60 MeV protons irradiate target producing $\bar{\nu}_e$
- IsoDAR's high $\bar{\nu}_e$ production rate paves the way for state-of-the-art sensitivities to sterile neutrino searches and exotic neutrino property studies.



(Figure not to scale)

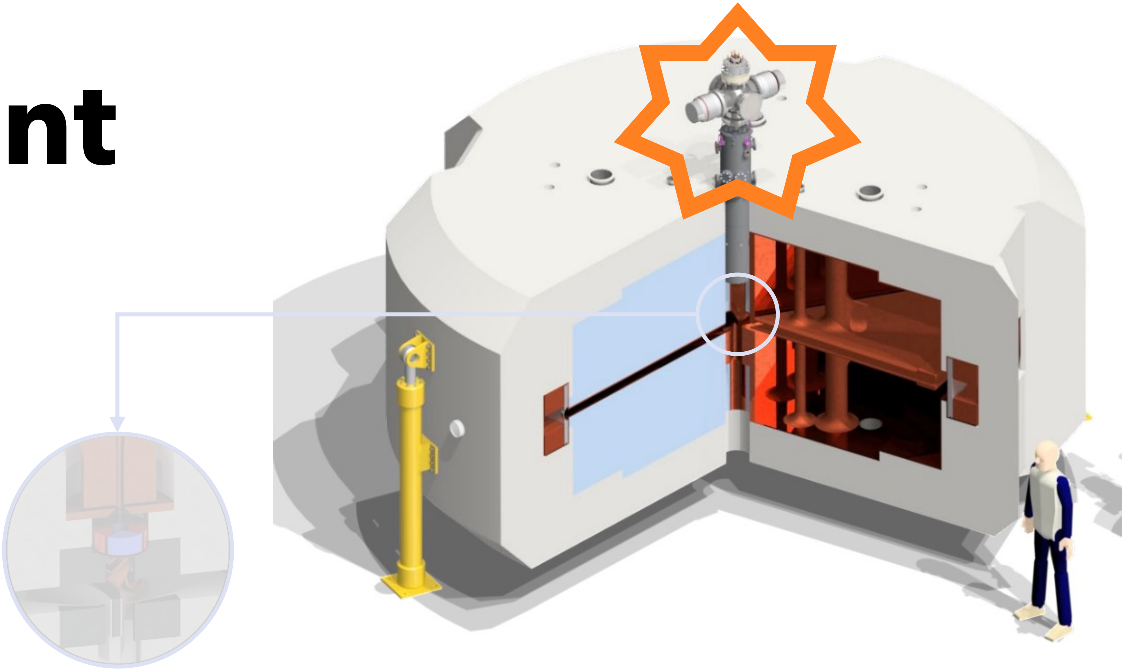


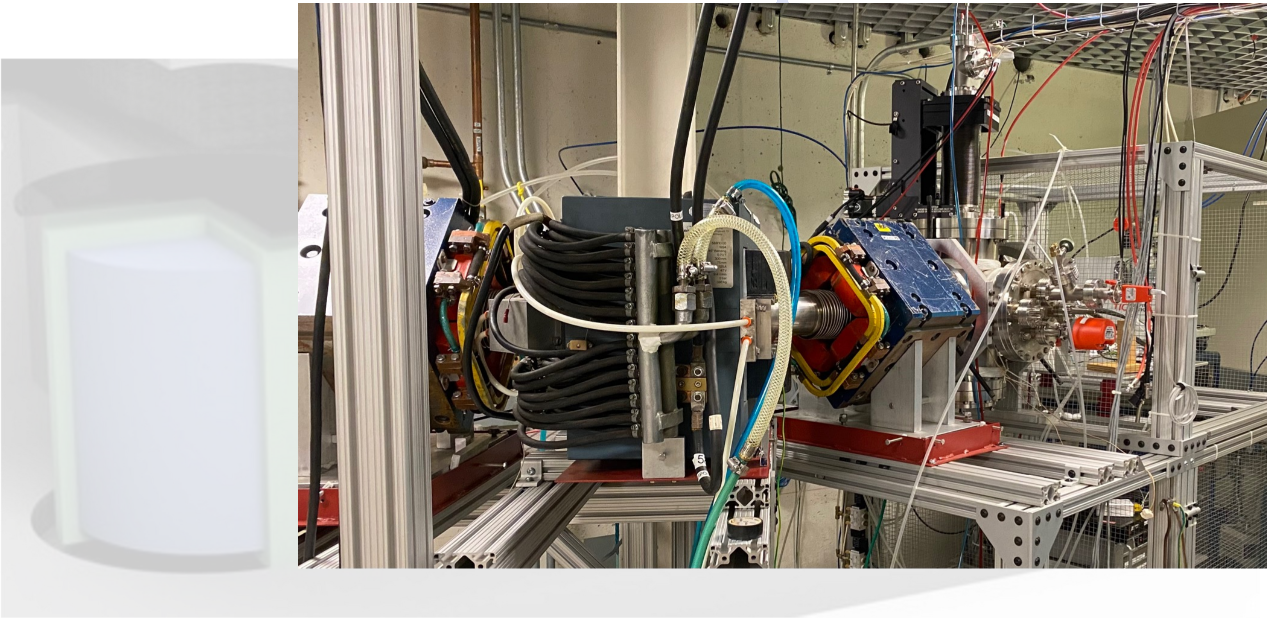




The IsoDAR Experiment
Ion Source

- Currently under commission at MIT's Plasma Science and Fusion Center
- Direct axial injection into the cyclotron require the ion source to have:
  - Low beam emittance
  - Minimally contaminated beam of H_2^+
  - High current (10 mA)
- Uses a multicusp ion source, with modifications, developed at LBNL
- Demonstrated this technology's feasibility by reaching an unprecedented 1 mA beam, low emittance (<0.05 π-mm-mrad, RMS, normalized), and high purity (80% H_2^+)

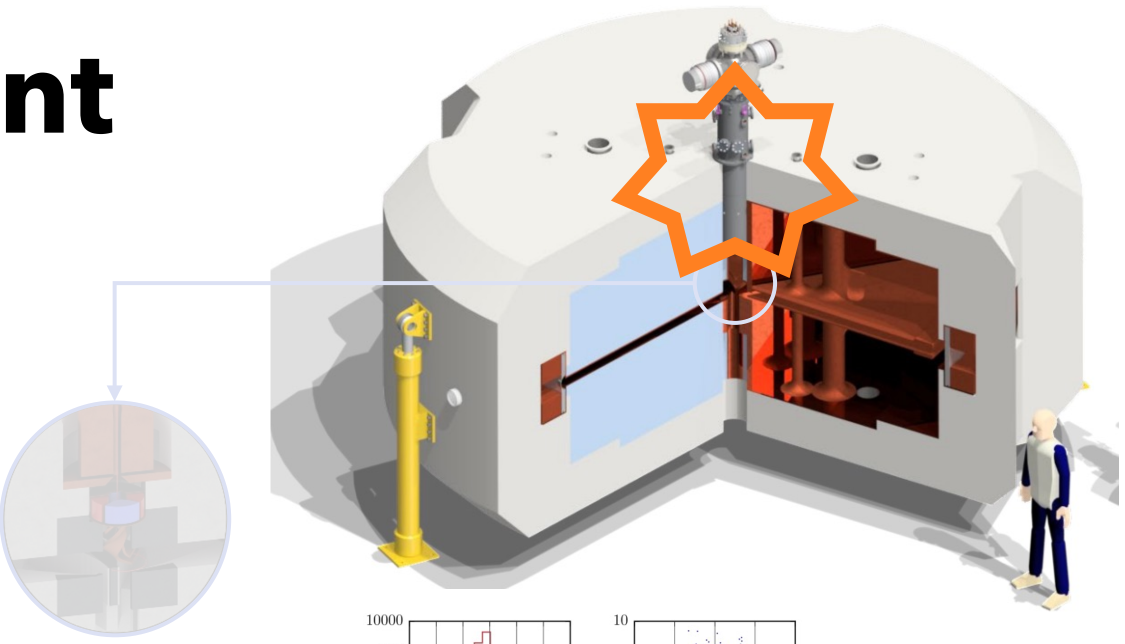




DOIs: 10.1063/1.4932395 & 10.1063/5.0063301

The IsoDAR Experiment
Pre-injection RFQ

- Currently being built at BEVATECH GmbH in Frankfurt, Germany
- Beam injection into cyclotrons generally difficult: has been shown that beam acceptances previously cap at 20%
- RFQs are especially useful for focusing, bunching, and even accelerating beams of low energy with high transmission while maintaining low beam emittances
- Necessary for clean injection into IsoDAR's spiral inflector and cyclotron
- Embedded in cyclotron yoke
- Subject of the remainder of this talk



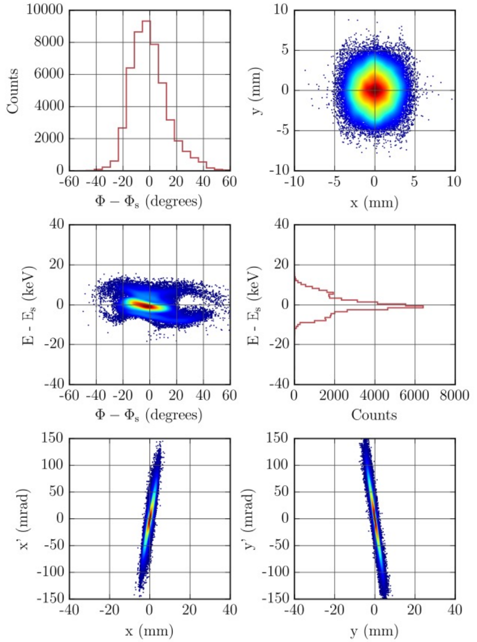
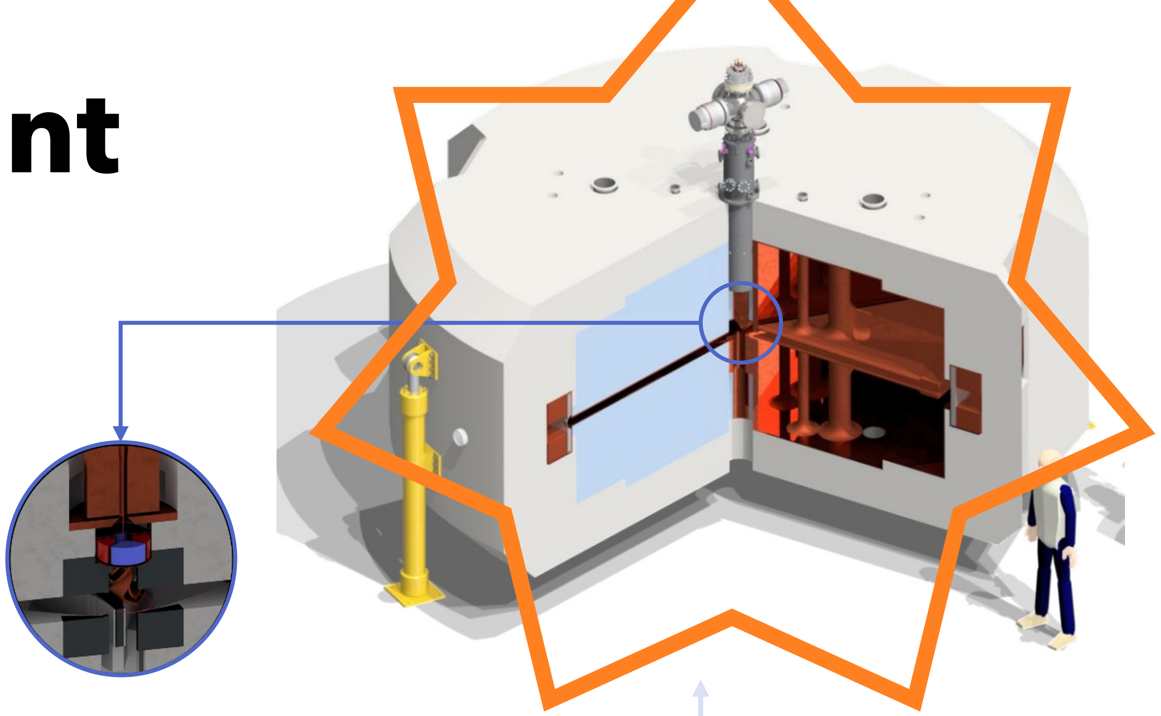


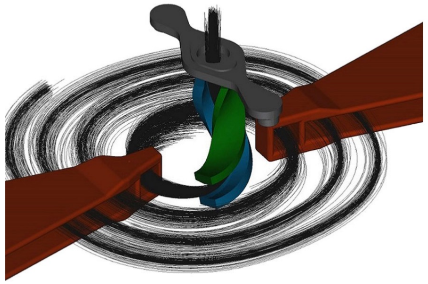
Figure 3: Phase Spaces of the RFQ output beam.

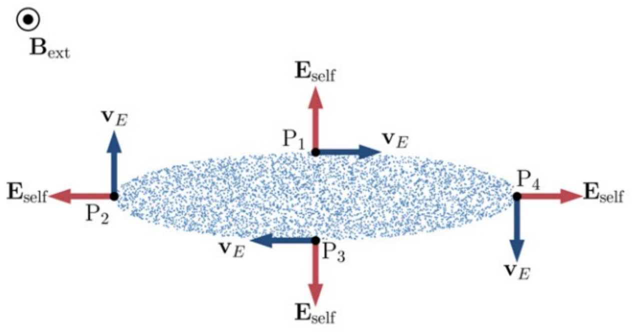
arXiv: 1710.00441, 1507.07258

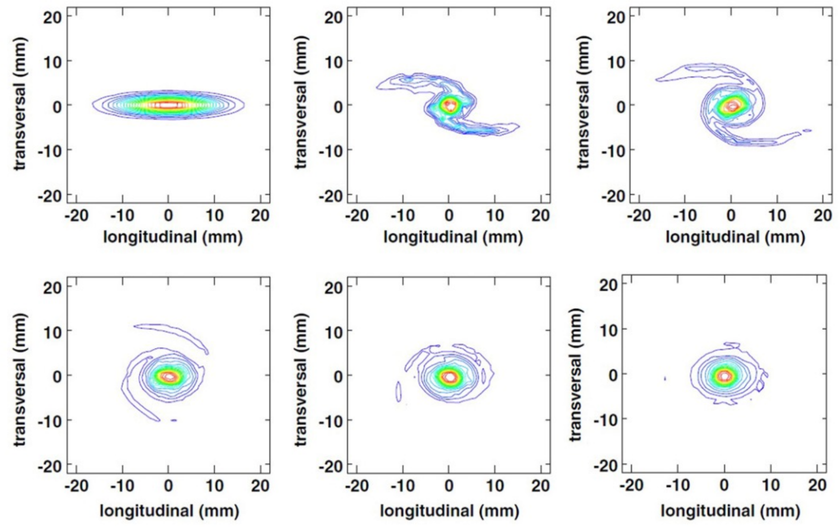
The IsoDAR Experiment
Injection and Cyclotron

- Beam is injected axially into cyclotron through a spiral inflector, where most of beam loss is expected to occur (~50%)
- IsoDAR cyclotron beam energy on-par to cyclotrons used for medical isotope production, but IsoDAR has much higher beam intensity
- High beam current poses space charge as a threat to clean extraction, but IsoDAR's cyclotron is designed to take advantage of *vortex motion* to allow for cleaner beam extraction
- Awarded ARDAP grant to develop 1.5 MeV/amu test cyclotron





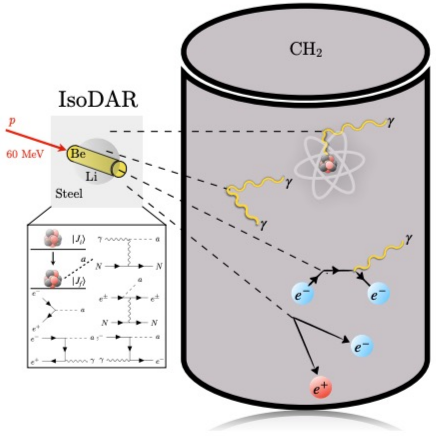




DOI 10.1088/1367-2630/ac5001

The IsoDAR Experiment
Relevance Beyond Sterile Neutrino Searches

- IsoDAR's nominal energy and neutrino intensity also optimal for exotic neutrino-decay searches
- Can provide newfound sensitivity to ALP searches when considering ALP couplings to nuclear de-excitation photons
- The development of cyclotrons in general can help to provide cost-effective particle physics experiments at universities and research centers worldwide
- Technology can be modified to significantly enhance medical isotope production



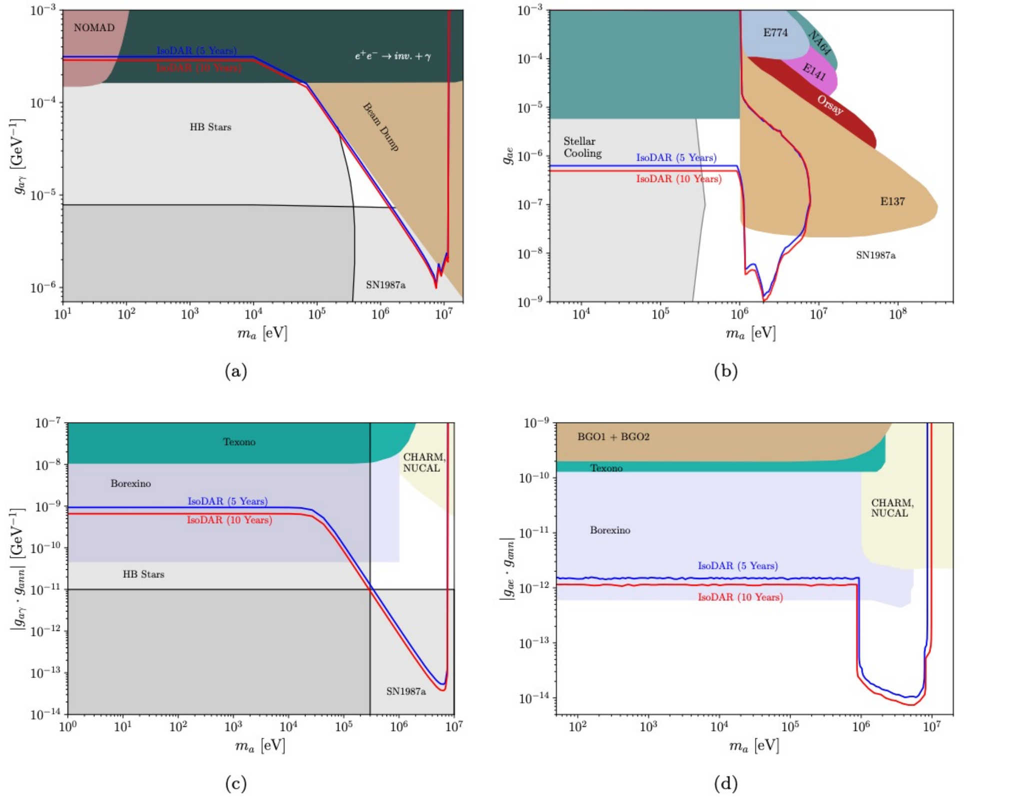


FIG. 4. Sensitivity contours at 90% CL, for 5 and 10 year exposures, using (a) couplings to photons, (b) couplings to electrons, (c) couplings to nucleons and photons, and (d) couplings to nucleons and electrons. In (c) and (d), ALPs are produced via nuclear transitions and propagate to the detector to subsequently scatter or decay via electron coupling (inverse Compton, $a \to e^+e^-$ decay) or photon coupling (inverse Primakoff, $a \to \gamma\gamma$ decay) channels.

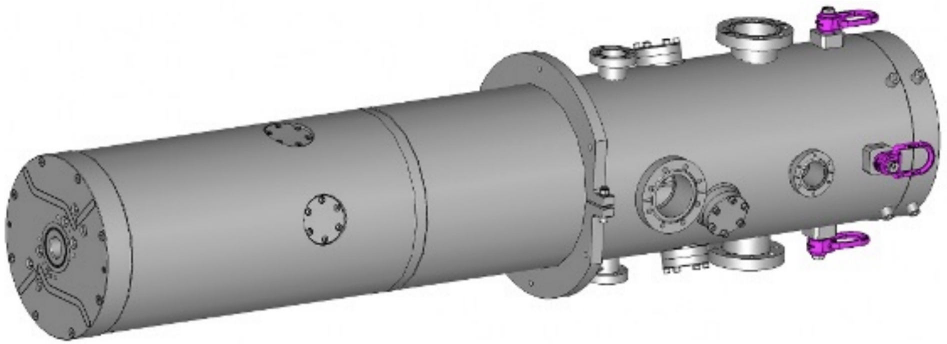
arXiv: 2207.13659, DOI: 10.1186/s41181-020-0090-3

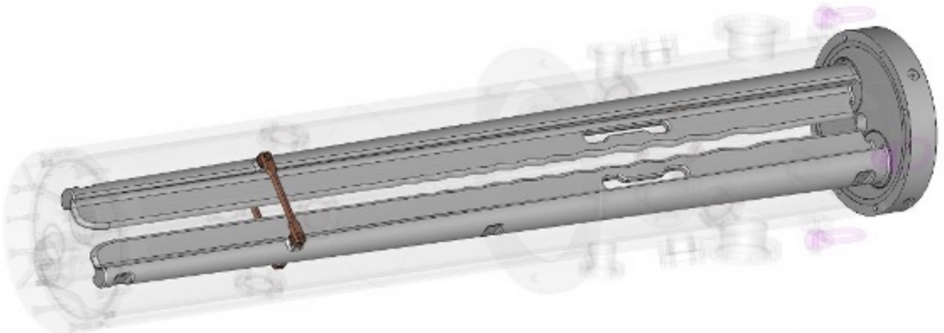
Outline

1. The IsoDAR Experiment
2. **Surrogate Modeling of Radiofrequency Quadrupoles**
3. Applications to Accelerator and Detector Design Optimization
4. Conclusions and Outlook

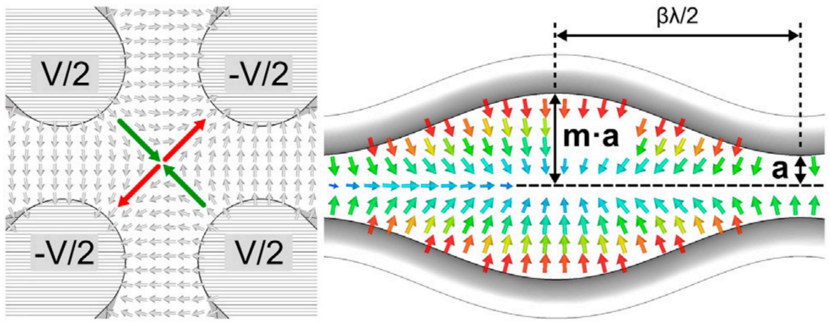
Radiofrequency Quadrupoles

- RFQs are linear accelerators in which four vanes (rods) apply oscillating EM fields, focusing throughgoing beam alternatingly in two transverse directions
- RFQs are necessitated by IsoDAR's physical space constraints and record-breaking beam current
  - Pre-acceleration before cyclotron injection
  - Early separation of p^+ and H_2^+
  - Focuses and bunches beam
- Simulation of RFQs (PARMTEQM, RFQGen) computationally costly
  - Exacerbated by IsoDAR's high beam current due to nonlinear space charge effects





Engineering model of IsoDAR's RFQ [1]



(Left) Instantaneous transverse electric field in an RFQ cell. (Right) Resultant longitudinal electric fields which create a net acceleration on one or many throughgoing charged particle. [2]

To an RFQ Surrogate Model

Uses of ML-enabled surrogate models

Uncertainty quantification

Predicted beam summary statistics can be used for global sensitivity analyses

Real-time feedback and commissioning

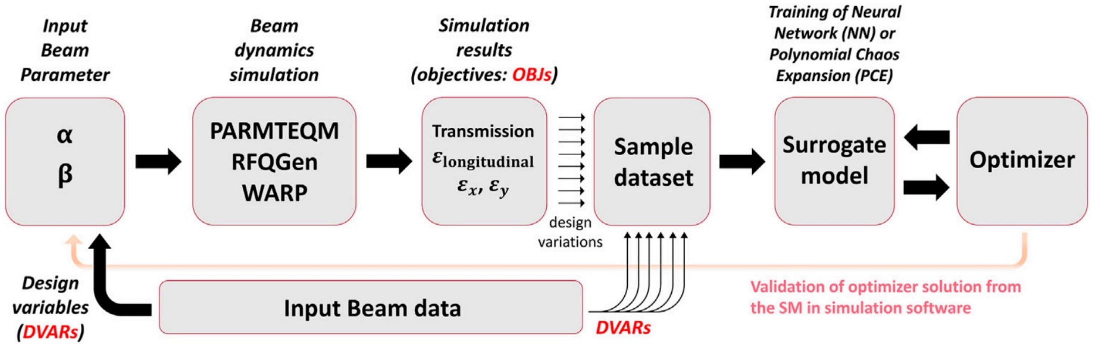
Rapid simulation of throughgoing beams is useful real-time information during tuning

Design optimization

Faster simulation reduces optimization convergence time

Can solve inverse problem for efficient initial guess of optimal design

- Surrogate models can serve as effective substitutes for computationally costly particle-in-cell simulations
- Fully connected deep neural networks (NNs) are well-suited to predicting beam output summary statistics (transmission, output energy, emittance) given RFQ design inputs

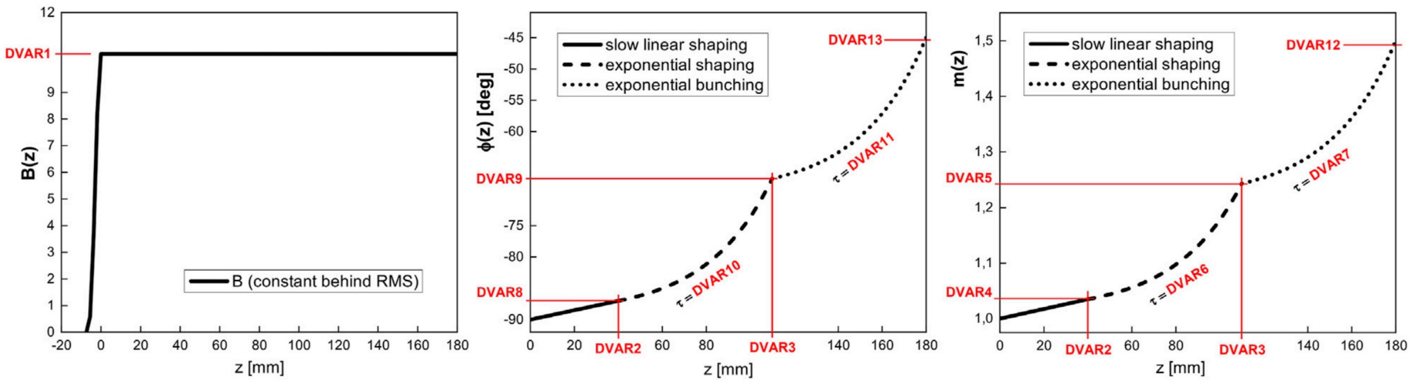


The use of a surrogate model for an RFQ design optimization scheme

https://doi.org/10.3389/fphy.2022.875889

Methods Overview

- >200,000 generated samples from high-fidelity simulation, same as previous work
    - 14 design variables (DVARs)
    - Extended with one feature corresponding to oddness/evenness of RFQ cells
    - 6 objective variables (OBJs)
- Code written in Julia
- Correlations between DVARs removed
- Initial NN hyperparameter scan expands limits explored previously
- Make appropriate cuts on data to eliminate nonphysical results (transmission)
- Design optimization leverages best-performing NN as surrogate model in Bayesian acquisition function

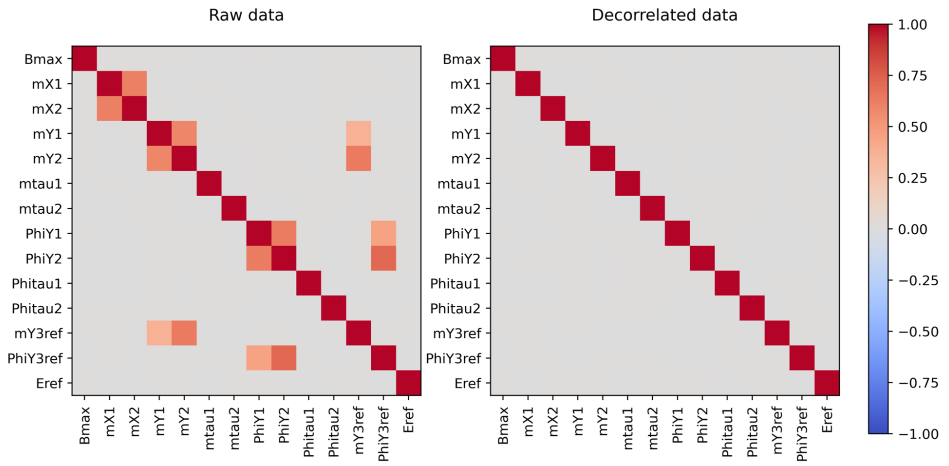


Parametrization of defining RFQ design characteristics used to build DVARs

```
DVAR definitions:
-----------------
DVAR1: Bmax   [ 8.5, 12.0 ]
DVAR2: mX1   [ 5, 140 ]
DVAR3: mX2   [ 15, 160 ]
DVAR4: mY1   [ 1.005, 1.7 ]
DVAR5: mY2   [ 1.055, 1.85 ]
DVAR6: mtau1   [ 1, 500 ]
DVAR7: mtau2   [ 1, 500 ]
DVAR8: PhiY1   [ -89.95, -30 ]
DVAR9: PhiY2   [ -87.45, -25 ]
DVAR10: Phitau1   [ 1, 500 ]
DVAR11: Phitau2   [ 1, 500 ]
DVAR12: mY3ref   [ 1.105, 2.0 ]
DVAR13: PhiY3ref   [ -84.95, -20 ]
DVAR14: Eref   [ 0.055, 0.075 ]
```

```
OBJ definitions:
----------------
OBJ1: transmission [%]
OBJ2: output energy [MeV]
OBJ3: RFQ length [cm]
OBJ4: longitudinal emittance [MeV*deg]
OBJ5: x-emittance [cm*mrad]
OBJ6: y-emittance [cm*mrad]
```

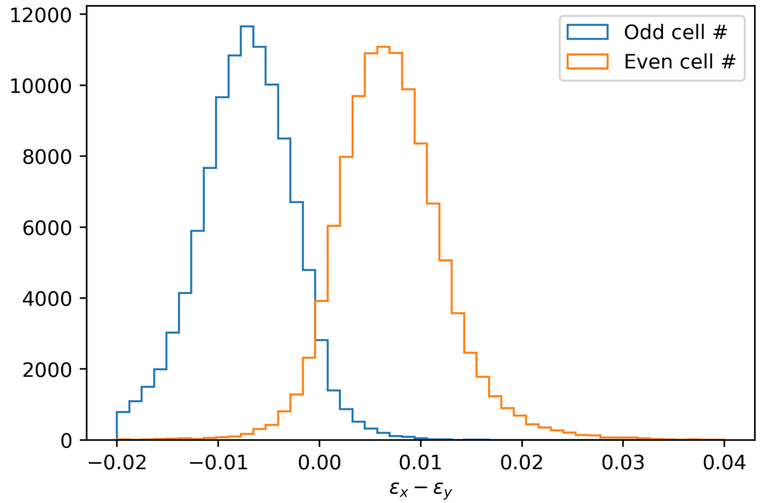
(Right) Correlation matrices before and after DVAR transformations



Methods
Predicting Emittances

- *Emittance* is the area that a beam occupies in phase space
- Beams with small emittances have particles confined to a small location having almost the same momentum

Difference between x- and y-emittances for RFQs in extended dataset having an odd number of cells and an even number of cells



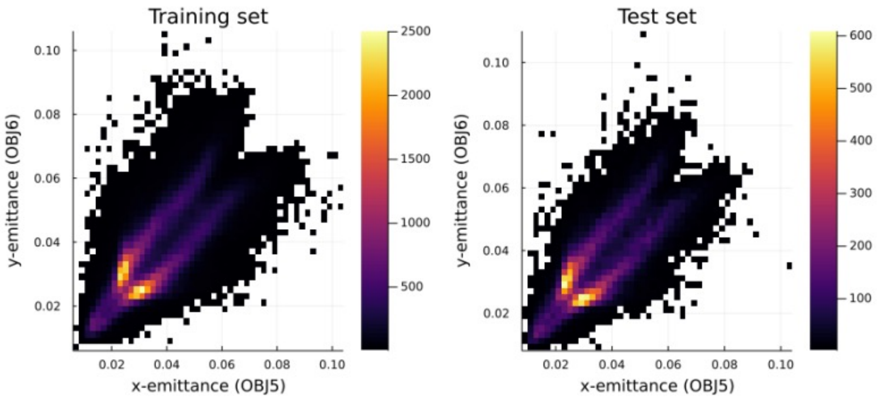


Figure 5. Joint distributions of *true* x and y emittances (OBJ5 and OBJ6, respectively) for sample training and test sets. From the inherent x, y symmetry of RFQs, we expect approximate symmetry about the 45° line.

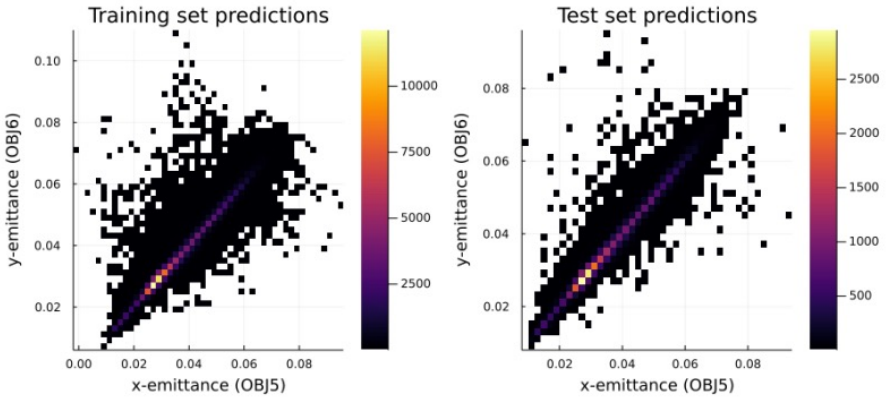
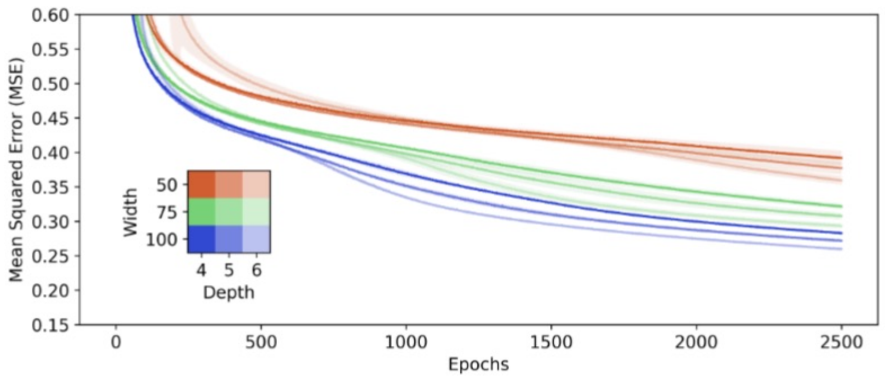


Figure 6. Joint distributions of *predicted* x and y emittances (OBJ5 and OBJ6, respectively) for sample training and test sets. Evident in these plots is the fact that the "double band" structure discussed in Fig. 5 is not recovered. We discuss in Sec. 2.8 that this effect is not due to training convergence to a local minimum in the output space, but an intrinsic characteristic of the dataset.

Results

Architecture Hyperparameter Grid Searches



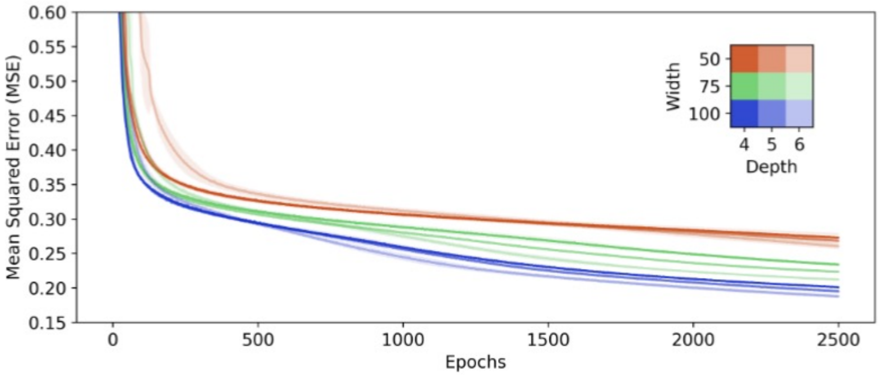


Figure 6. Training set loss curves (mean squared error) for scanned neural network architectures (Tab. 2), trained on the complete dataset (top) and samples having beam transmission $\geq 60\%$ (bottom). The solid line in the center of each curve represents the cross-fold loss mean, and one standard deviation is shaded above and below.

- All neural networks were trained to minimize MSE, had the same batch size of 1024, used sigmoid activations, and an ADAM optimizer with a constant learning rate of .1%

	Depth 4	Depth 5	Depth 6
Width 50	0.9884 ± 0.000311	0.9887 ± 0.000217	0.9893 ± 0.000127
Width 75	0.9900 ± 0.000169	0.9904 ± 0.0000864	0.9907 ± 0.000145
Width 100	0.9905 ± 0.0000763	0.9908 ± 0.0000698	0.9908 ± 0.00017

	Depth 4	Depth 5	Depth 6
Width 50	0.9901 ± 0.0002	0.9902 ± 0.000174	0.9905 ± 0.00026
Width 75	0.9912 ± 0.000129	0.9915 ± 0.000178	0.9918 ± 0.000217
Width 100	0.9918 ± 0.0000719	0.9917 ± 0.000152	0.9920 ± 0.0000623

Table 3. Aggregated validation-set R^2 scores for each set of hyperparameters (Tab. 2), for NNs trained on the complete dataset (top) and samples having transmission $\geq 60\%$ (bottom). Each R^2 score is computed across all objective variables, but is not representative of the prediction accuracy of individual objectives; MAPEs for each objective for each NN architecture are shown in Appendix C.

Results
NN Performance

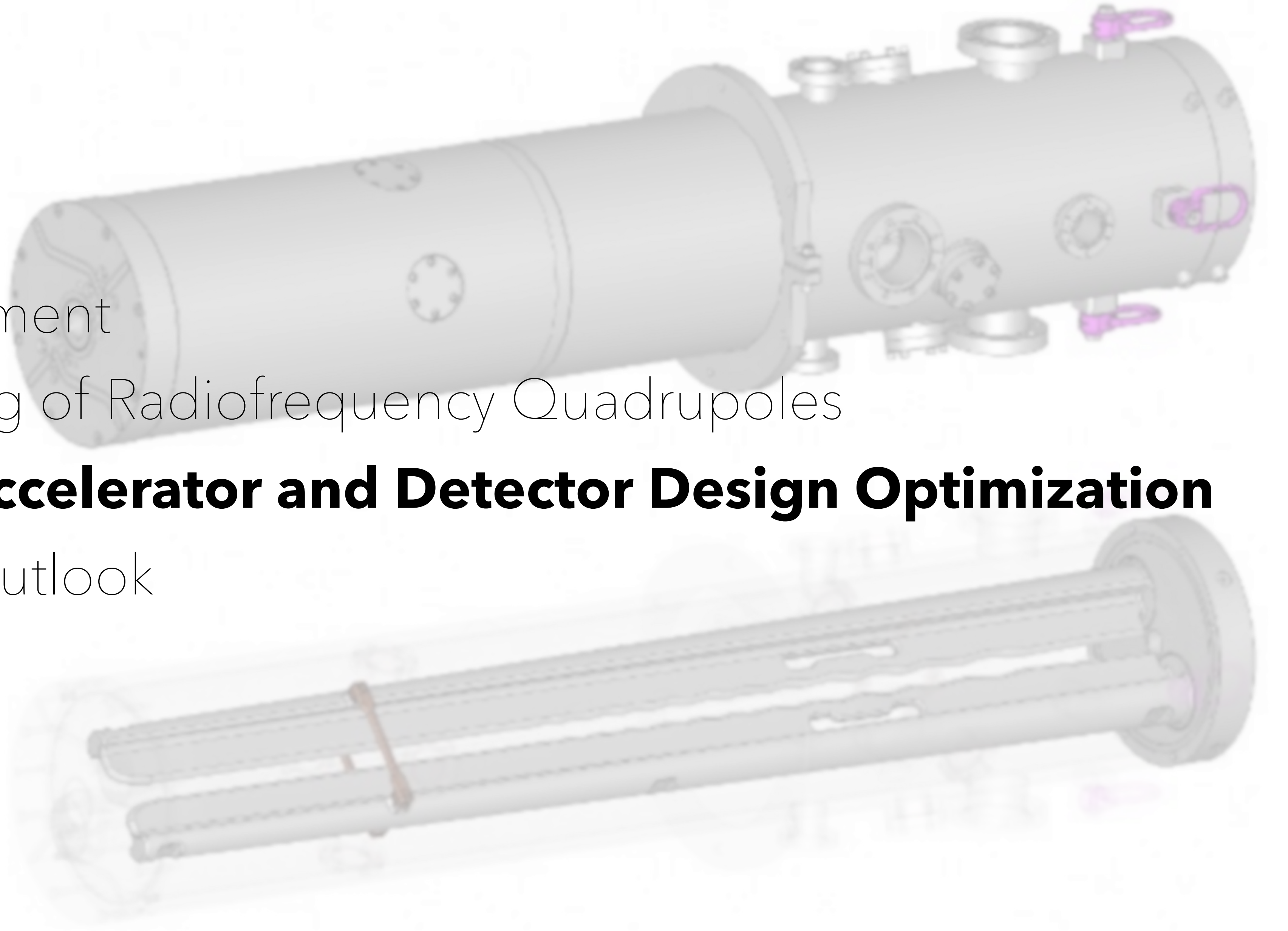
Label	Objective Variable	RFQNet1	RFQNet2	Ref. [17]
OBJ1	Transmission [%]	1.5%	0.97%	2.4%
OBJ2	E_{out} [MeV]	1.8%	1.8%	1.9%
OBJ3	RFQ length [cm]	1.3%	1.3%	2.0%
OBJ4	$\epsilon_{long.}$ [MeV-deg]	6.9%	5.8%	8.2%
OBJ5	ϵ_x [cm-mrad]	4.8%	4.1%	12.8%
OBJ6	ϵ_y [cm-mrad]	4.8%	4.0%	12.5%

Test-set mean absolute percent errors on top-performing NN predictions for each of the 6 objectives studied. Compared to previously best-performing surrogate model.

	RFQNet1	RFQNet2	Previous
Batch Size	1024	1024	256
Uses 15th feature variable: odd/even number of RFQ cells	✅	✅	❌
Data restricted to have transmission of at least 60%	❌	✅	❌

- Larger batch size alone was responsible for most of the reductions in transmission and longitudinal emittance predictions
- K-S tests between training-set and test-set errors indicate significant statistical differences in the distributions of residuals for certain objectives, hinting at overfitting
- Dropout regularization seemed to have inconsistent effects on the predictive performance of each of the 6 objectives studied

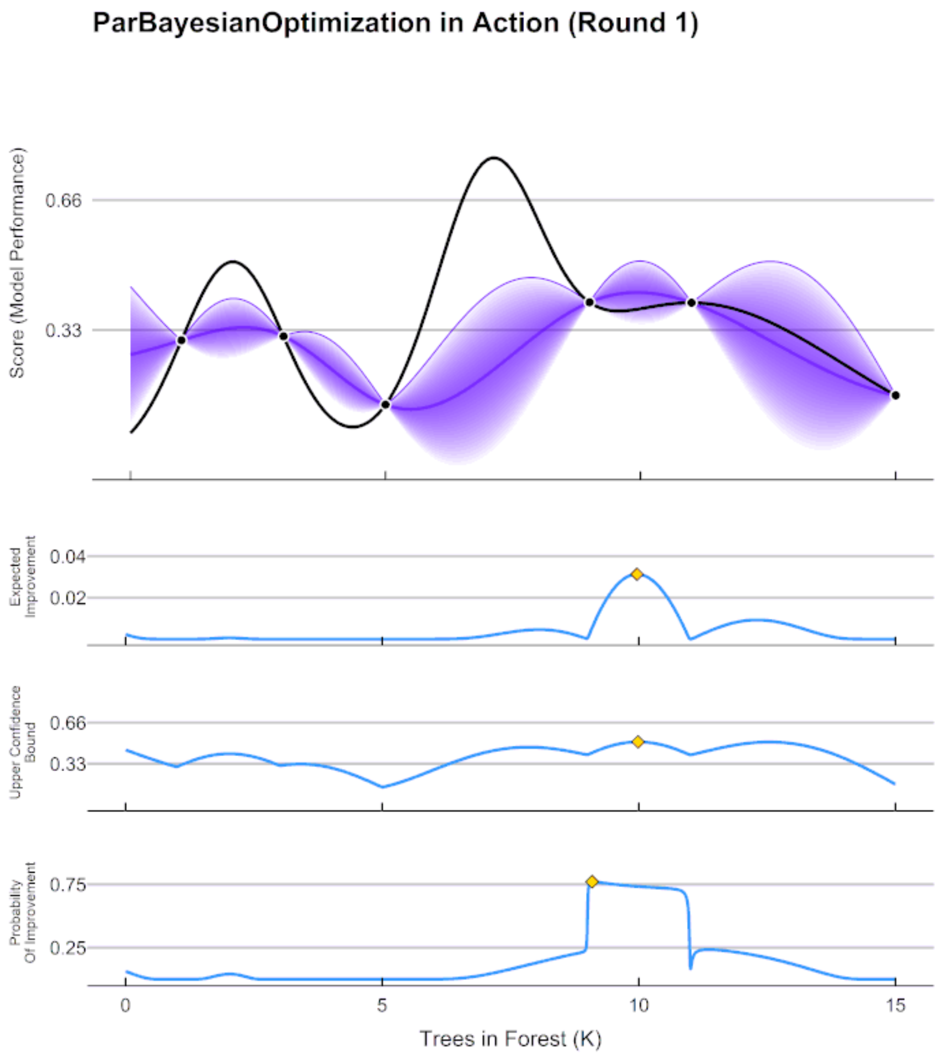
Outline



1. The IsoDAR Experiment
2. Surrogate Modeling of Radiofrequency Quadrupoles
3. **Applications to Accelerator and Detector Design Optimization**
4. Conclusions and Outlook

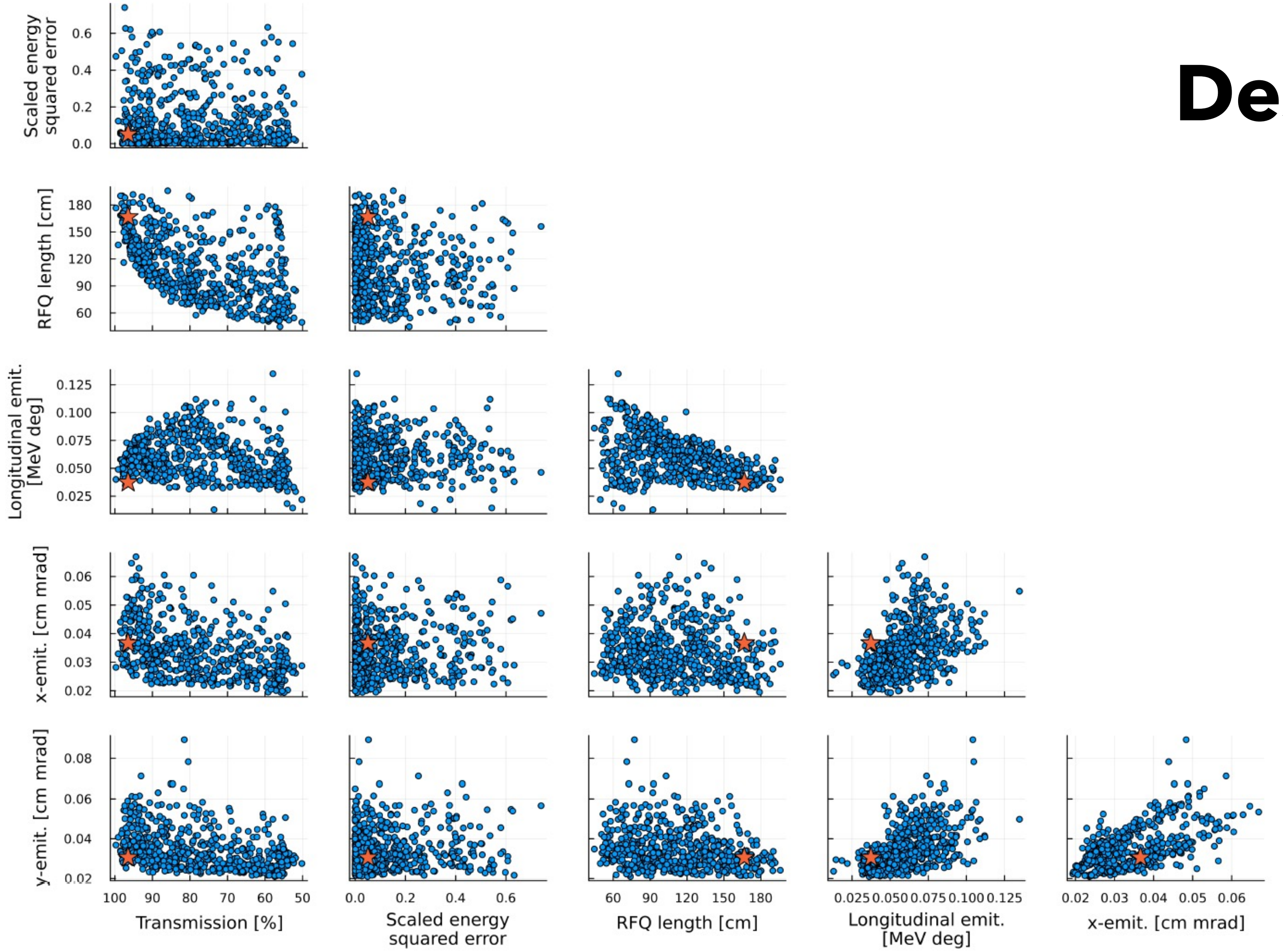
Bayesian Optimization

- Given some complex function f, we can algorithmically find an optimum by way of Bayesian Optimization
  - We can fit a Gaussian Process to f to build an *acquisition function* which helps us to determine which point in the input data to test for optimality next
  - Iteratively, we choose a potential optimum, update our Gaussian process with an evaluation of f, and choose the next optimum in the set
  - Finding optima if f is multidimensional means computing a *Pareto front* in which we can evaluate tradeoffs between the responses
- We employ Julia's `surrogates.jl` package to optimize the design of IsoDAR's RFQ using RFQNet2



GIF credit: https://commons.wikimedia.org/wiki/File:GpParBayesAnimationSmall.gif

Results
Design Optimization



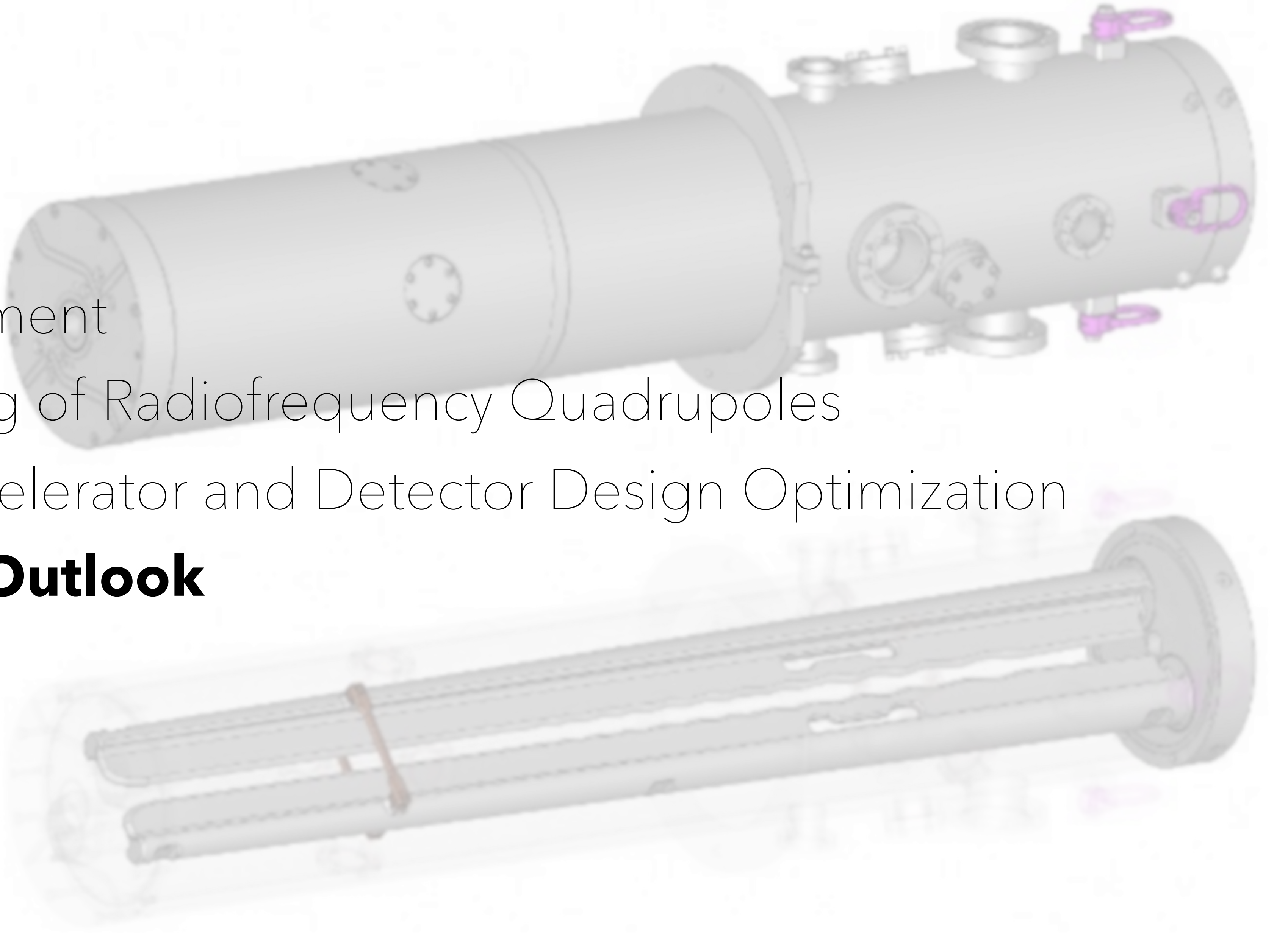
- RFQNet2 (incl. transmission cuts, 15th feature variable) as acquisition function in Bayesian optimization of RFQ design, relevant to IsoDAR's specifications:

```
max transmission
min (energy – 70 KeV)^2
min RFQ length
min longitudinal emittance
min transverse emittances
```

- Optimum identified as having $> 95\%$ transmission, $\epsilon_{long} < 0.04$ MeV °, and $\epsilon_x, \epsilon_y < 0.04$ cm mrad

Outline

1. The IsoDAR Experiment
2. Surrogate Modeling of Radiofrequency Quadrupoles
3. Applications to Accelerator and Detector Design Optimization
4. **Conclusions and Outlook**



Conclusions

- Julia is a programming language well-equipped to handle the computational rigors of surrogate model engineering
- Careful data selection criteria significantly improved NN predictive accuracy

Coming to a journal near you…
(2210.11451)

Future Work

- More sophisticated NNs may continue to improve predictive accuracy
  - Additional regularization strategies
  - Residual NNs, convolutional NNs
  - Step through RFQ one cell at a time to build final beam quality predictions