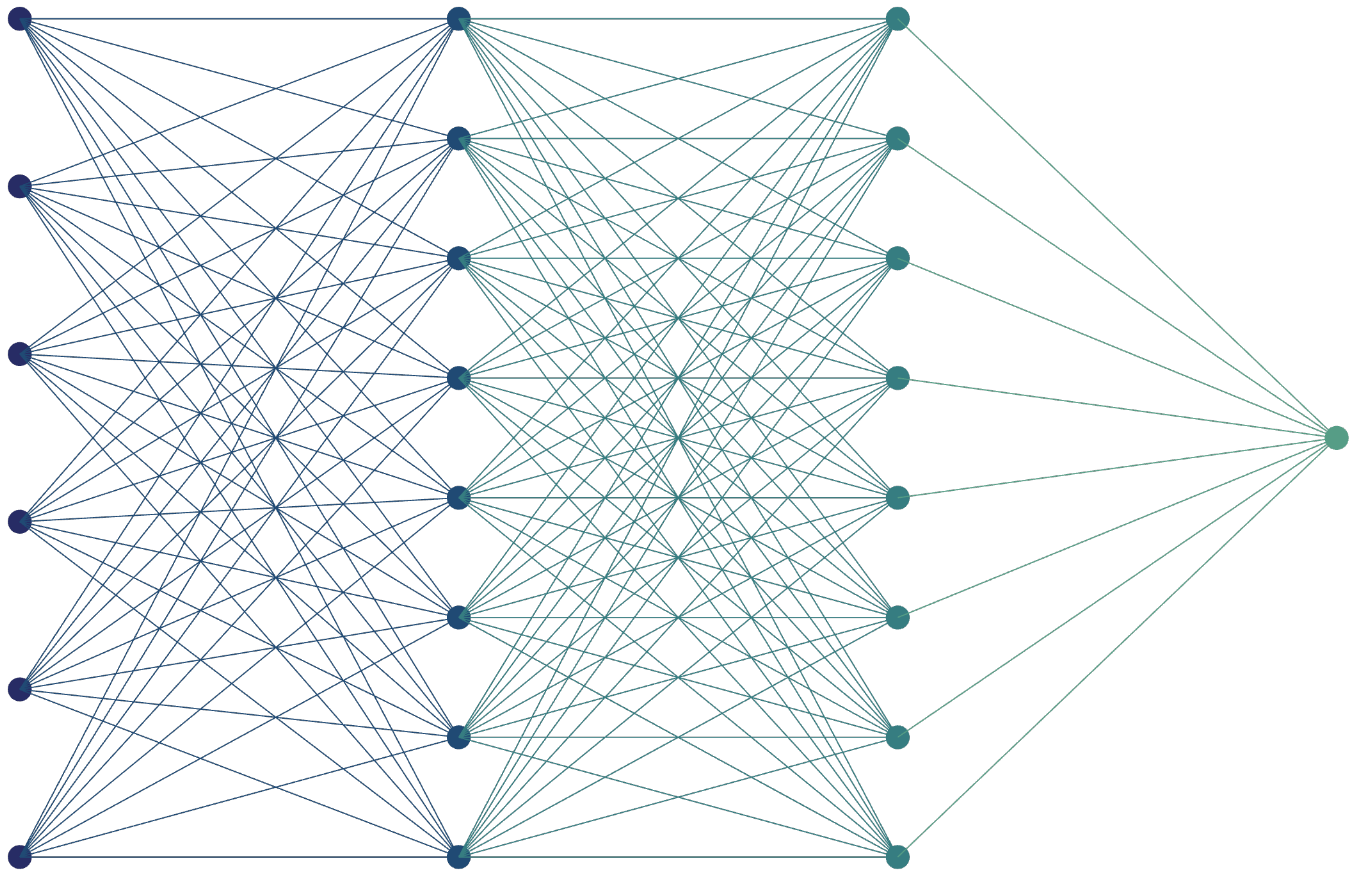
ML Methods for the keV sterile neutrino search with the TRISTAN-Phase of the KATRIN Experiment

Luca Fallböhmer –

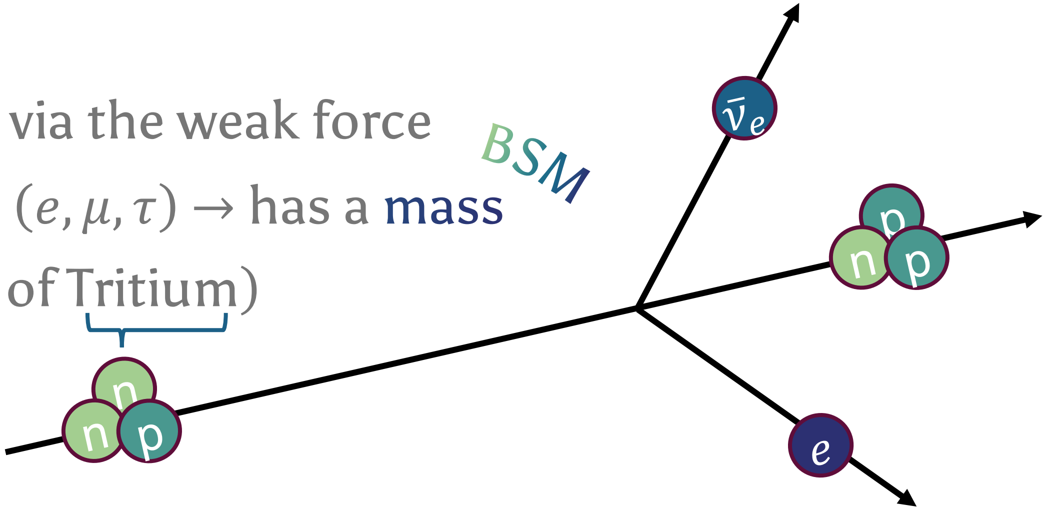
NPML 2025

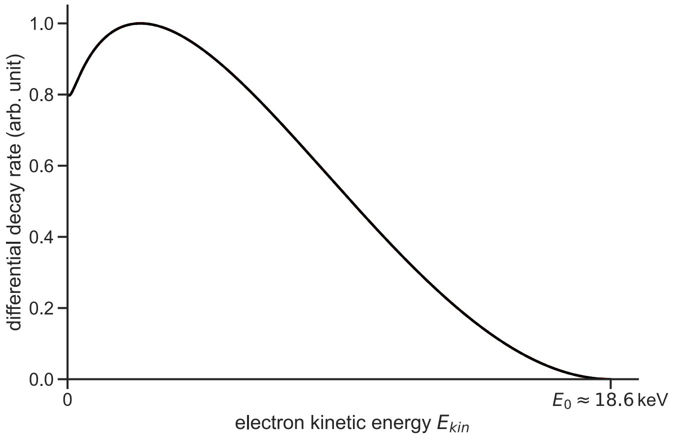


Neutrino Physics and the KATRIN Experiment

The Neutrino ν

- Lepton that only interacts via the weak force
- Oscillates between flavors $(e, \mu, \tau) \rightarrow$ has a **mass**
- Is created in β-decay (e.g. of Tritium)

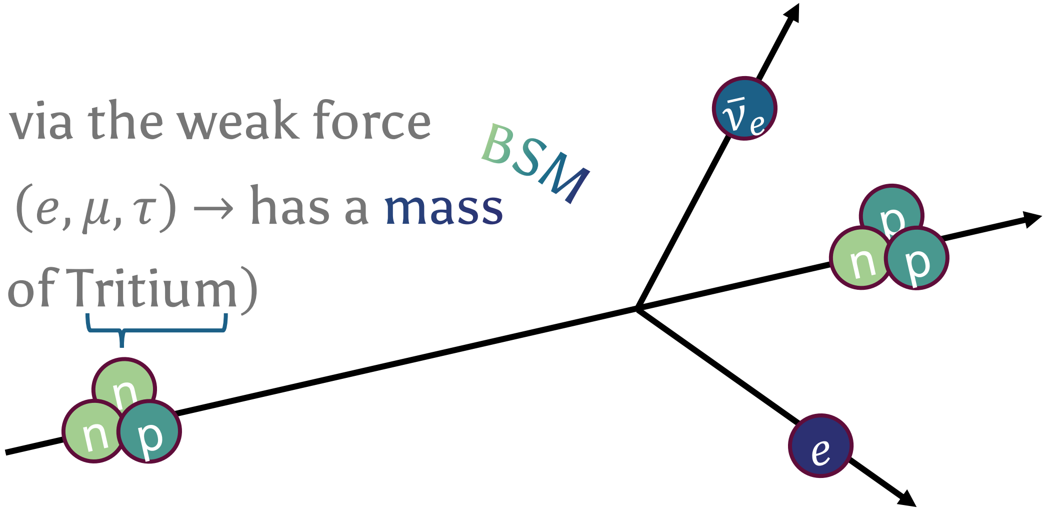


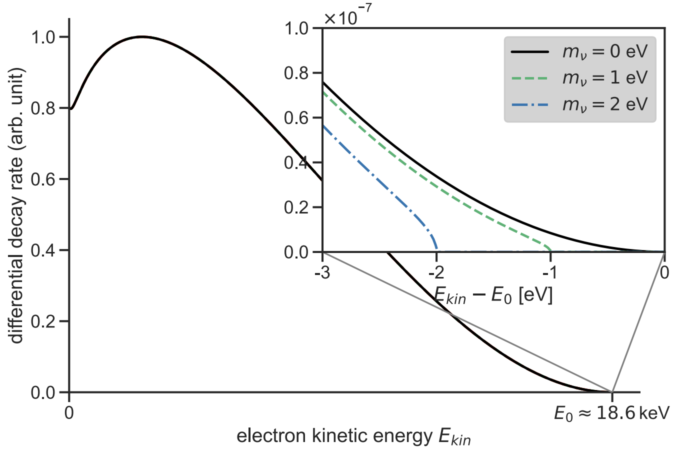


Neutrino Physics and the KATRIN Experiment

The Neutrino ν

- Lepton that only interacts via the weak force
- Oscillates between flavors $(e, \mu, \tau) \rightarrow$ has a **mass**
- Is created in β-decay (e.g. of Tritium)

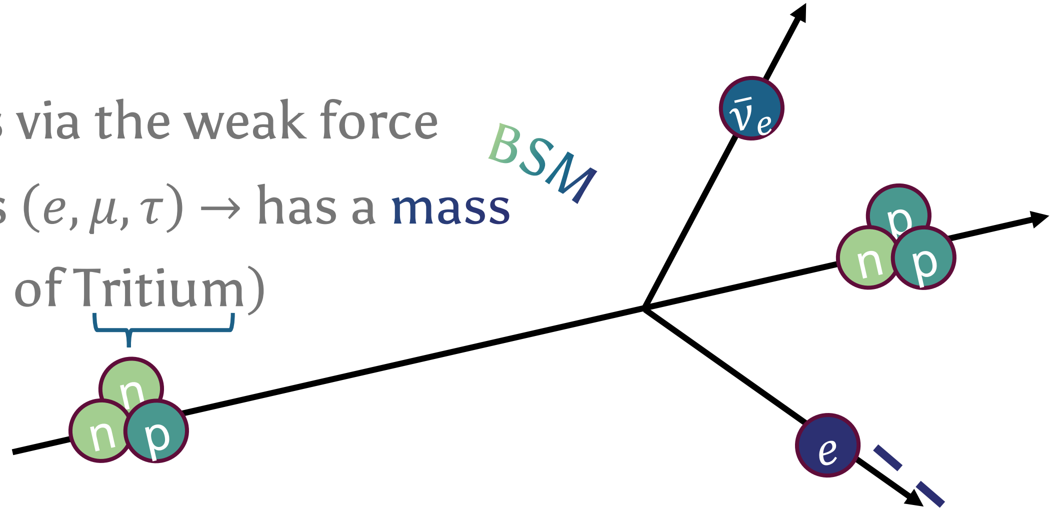


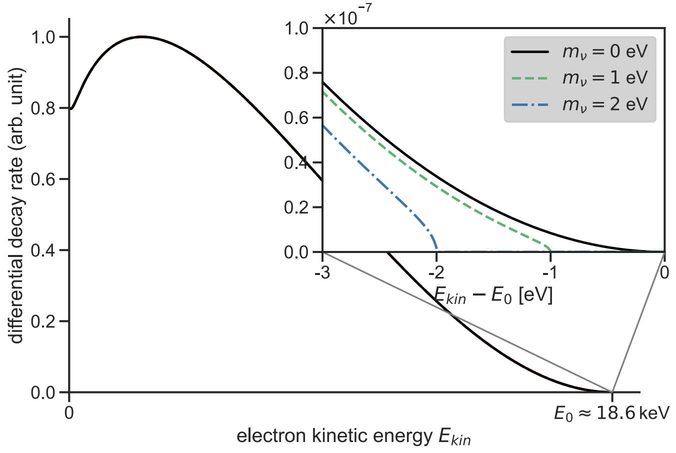


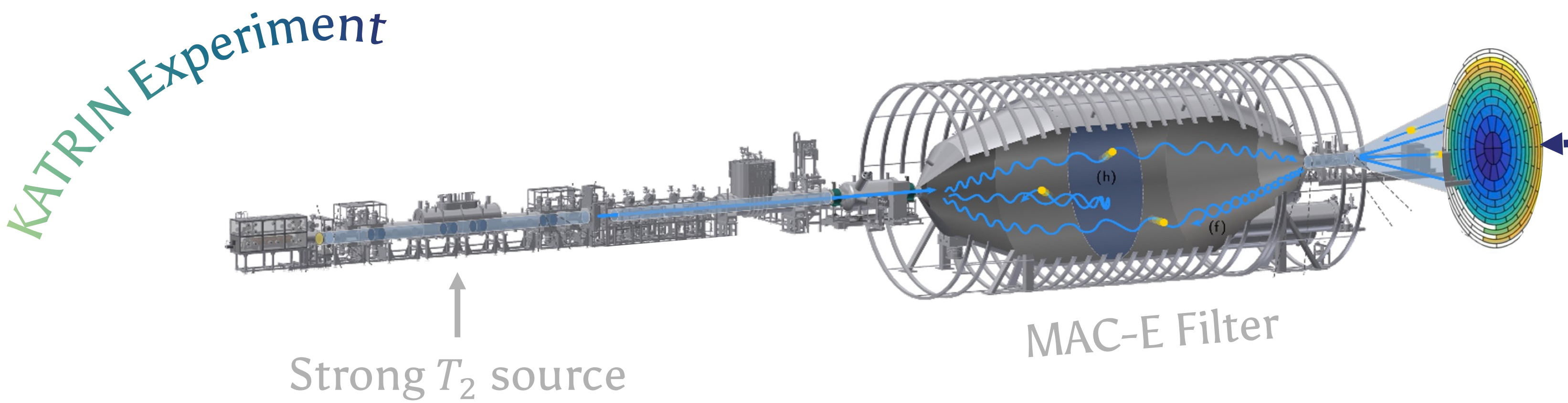
Neutrino Physics and the KATRIN Experiment

The Neutrino ν

- Lepton that only interacts via the weak force
- Oscillates between flavors $(e, \mu, \tau) \rightarrow$ has a **mass**
- Is created in β-decay (e.g. of Tritium)



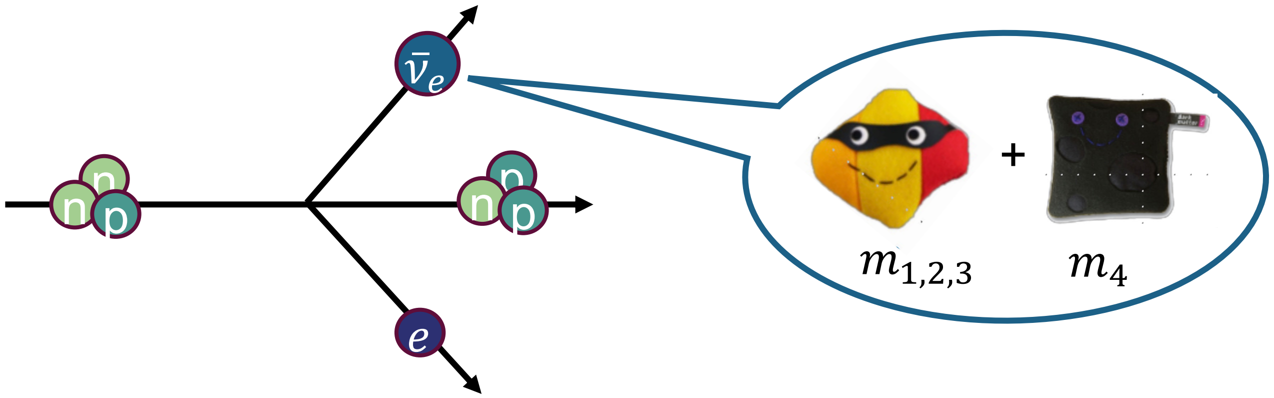




KATRIN measures:

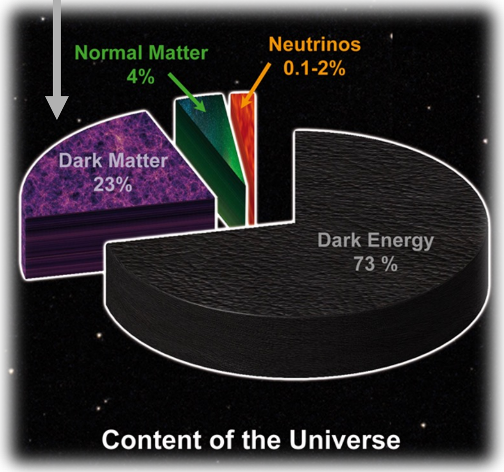
- The kinematic endpoint region
- A low number of counts (<1 CPS)
- In integral mode

… but there's more!



The Sterile Neutrino ν_4

- Hypoth. particle, no interaction via weak force
- Excellent dark matter candidate if m_4 on a keV-scale



… but there's more!

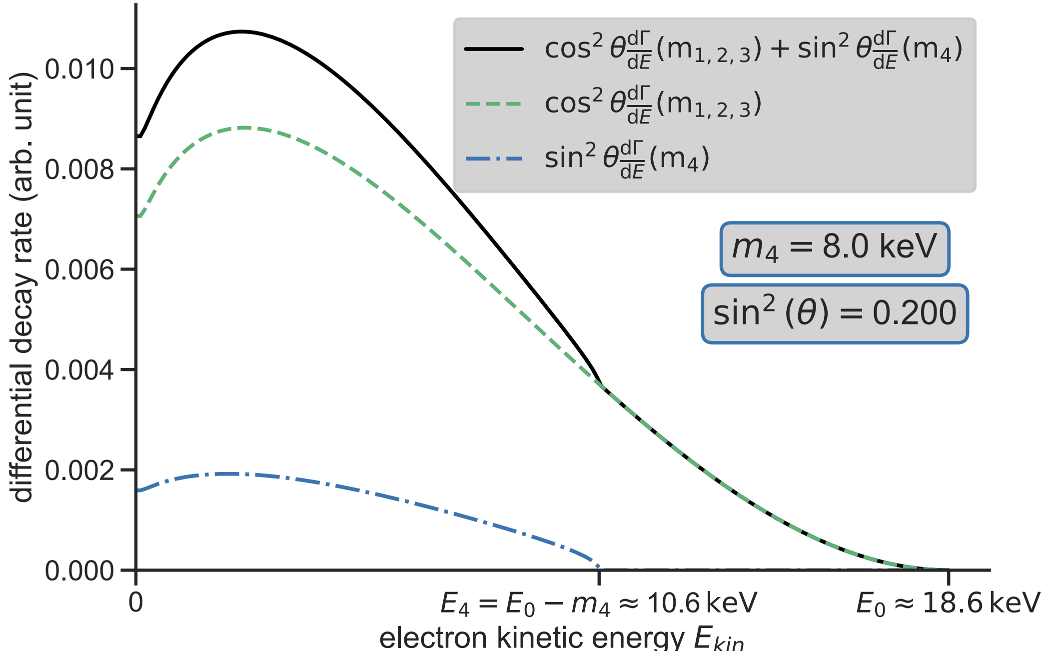
The Sterile Neutrino ν_4

- Hypoth. particle, no interaction via weak force
- Excellent dark matter candidate if m_4 on a keV-scale

… but there's more!

The Sterile Neutrino ν_4

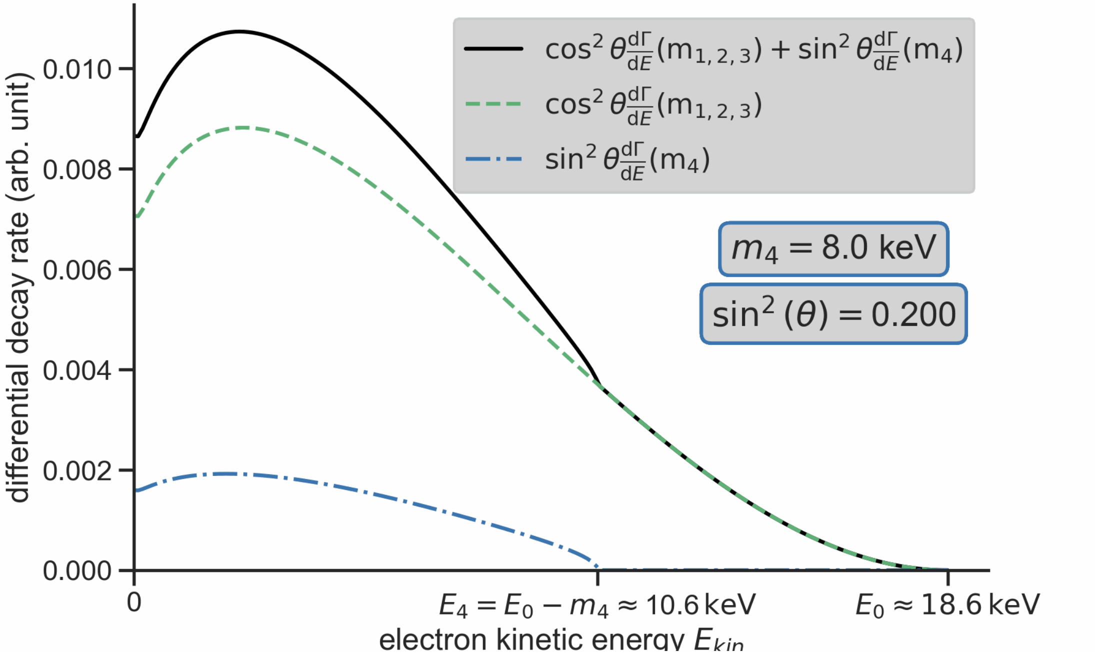
- Hypoth. particle, no interaction via weak force
- Excellent dark matter candidate if m_4 on a keV-scale
- Appears as a kink in the spectral shape



… but there's more!

The Sterile Neutrino ν_4

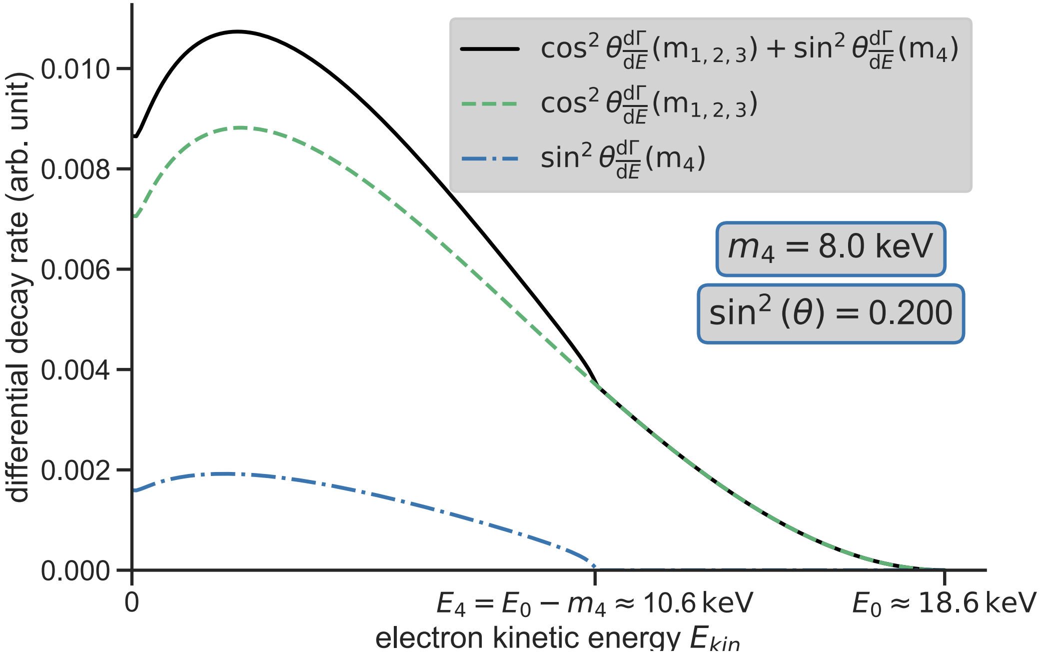
- Hypoth. particle, no interaction via weak force
- Excellent dark matter candidate if m_4 on a keV-scale
- Appears as a kink in the spectral shape

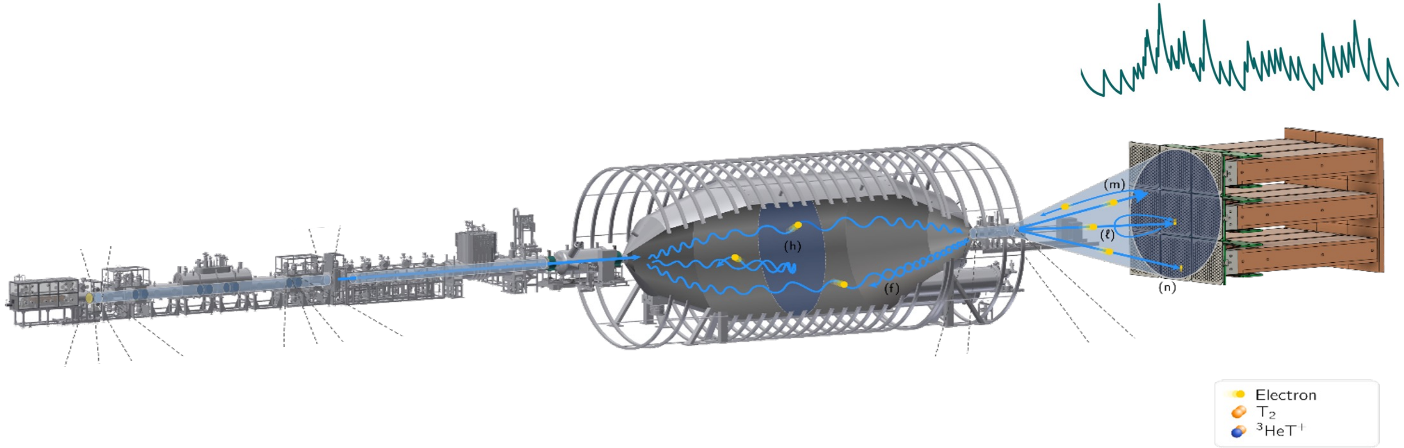


… but there's more!

The Sterile Neutrino ν_4

- Hypoth. particle, no interaction via weak force
- Excellent dark matter candidate if m_4 on a keV-scale
- Appears as a kink in the spectral shape

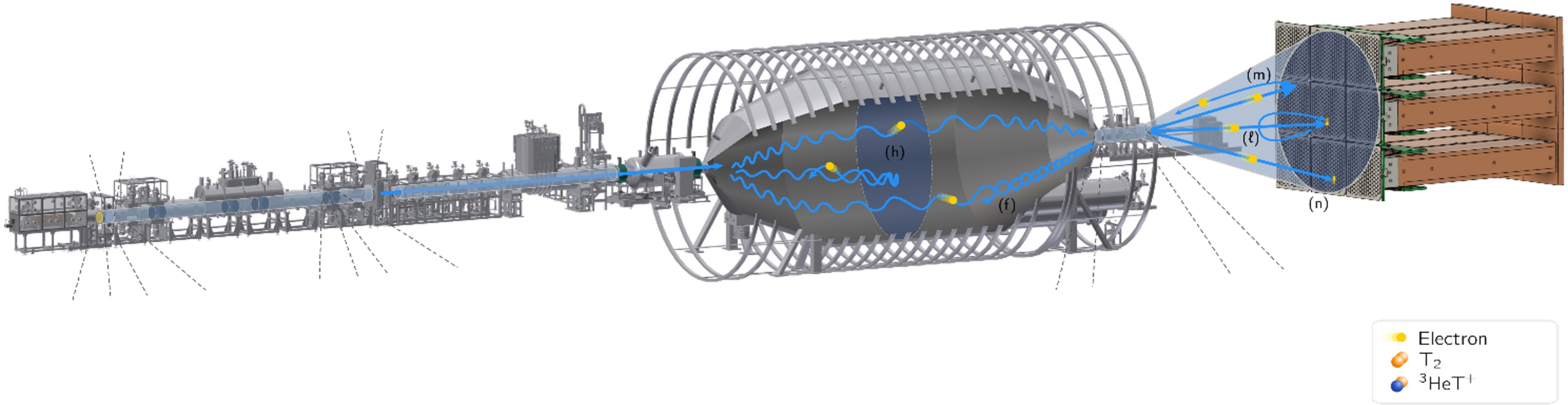




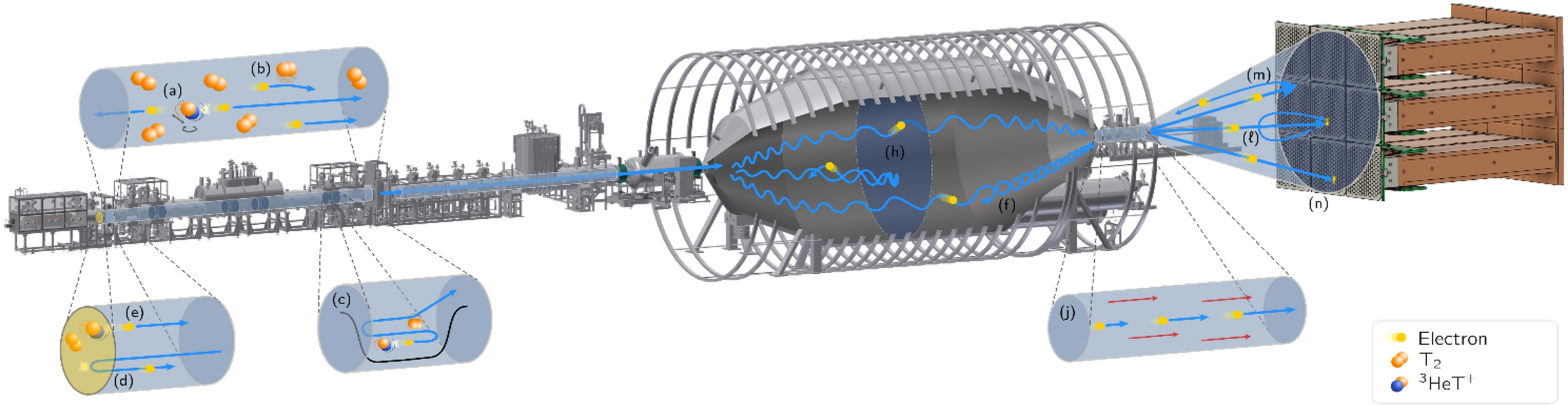
The TRISTAN Detector and DAQ System

- Silicon Drift Detector
- Consists of ~1000 pixels
- Can handle very high rates (~10^8 CPS total)
- Differential measurement

Challenge: Systematic Effects

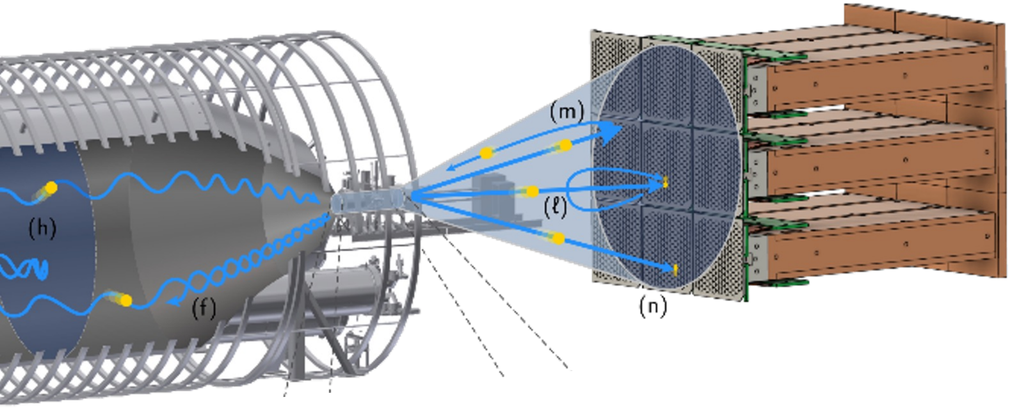


Challenge: Systematic Effects

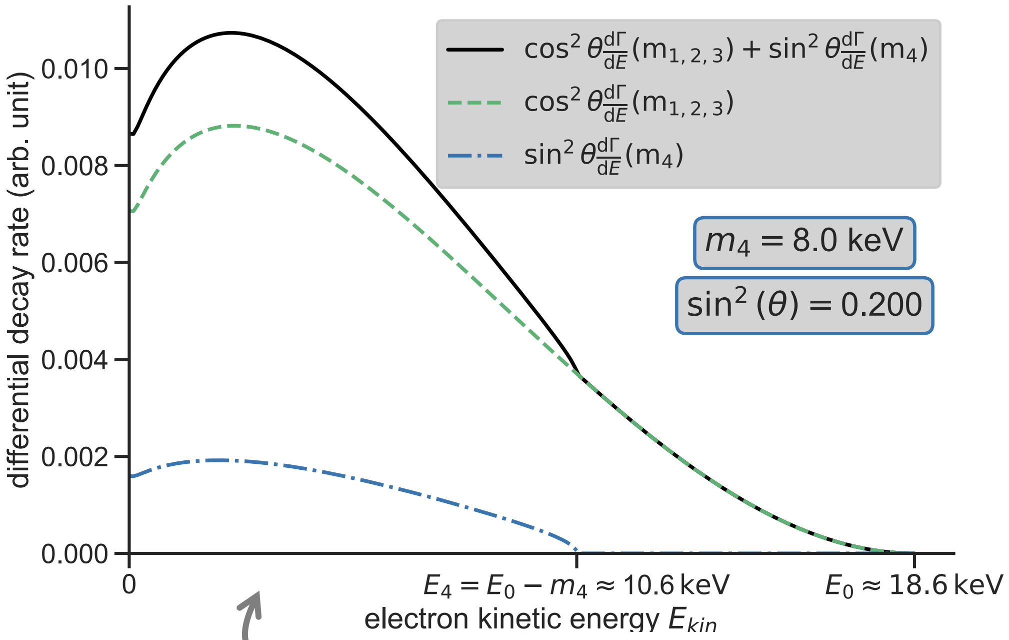


Challenge: Systematic Effects

Percent-level systematic effects
(e.g. Backscattering @ Detector)

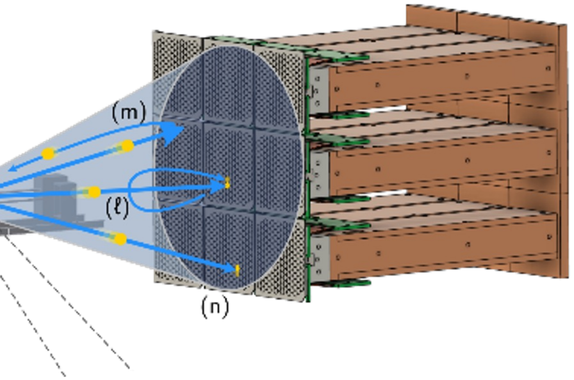


While…



Trying to be sensitive to ppm-level signal (collecting $\mathcal{O}(10^{14})$ events)

Challenge: Systematic Effects



Model systematic effects via Monte Carlo simulations

(need to be very precise)

Fit prediction to data

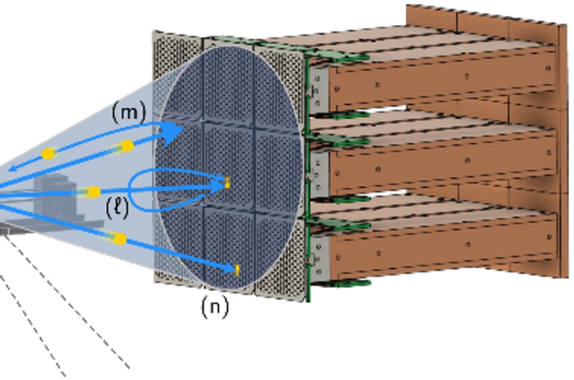
$$Model = \boldsymbol{A} \times \Gamma(E, \boldsymbol{m_4}, \mathbf{sin}^2\boldsymbol{\theta}, \boldsymbol{y_{nuis}})$$

free amplitude

Sterile parameters

Nuisance parameters

Challenge: Systematic Effects



Model systematic effects via Monte Carlo simulations

(need to be very precise)

Fit prediction to data

$$Model = A \times \Gamma\left(E, \boldsymbol{m_4}, \mathbf{sin}^2\boldsymbol{\theta}, \boldsymbol{y_{nuis}}\right)$$

free amplitude

Sterile parameters

Nuisance parameters

But…

- **Limited accuracy of MC**
- **Limited statistics of MC**

How to Deal with MC Limitations using ML?

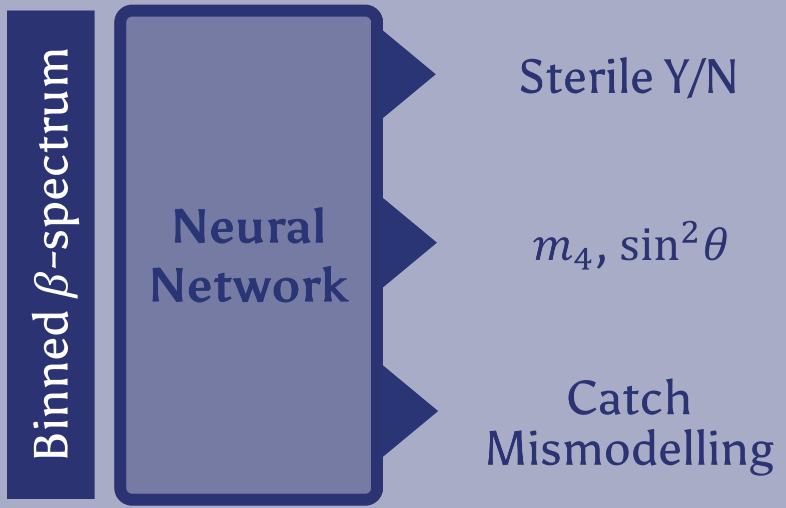
First Ideas (work in progress)

How to Deal with MC Limitations using ML?

First Ideas (work in progress)

Limited Accuracy of MC:

→„Directly“ Search for kink-like signature with Neural Networks

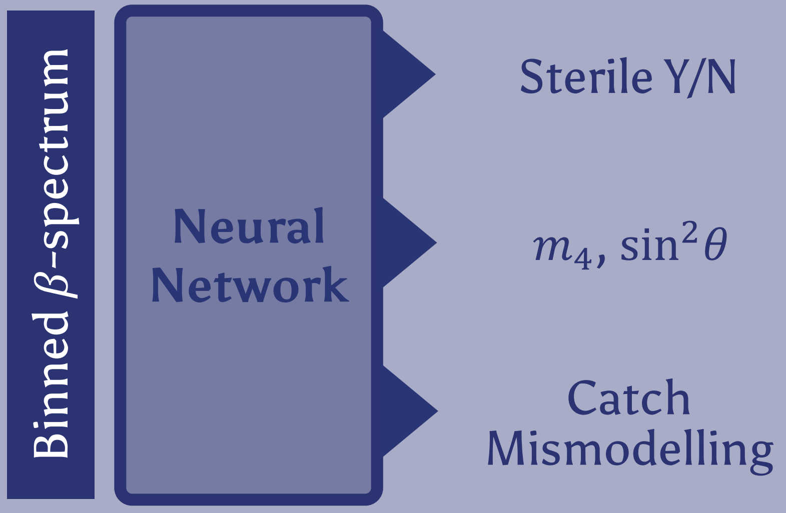


How to Deal with MC Limitations using ML?

First Ideas (work in progress)

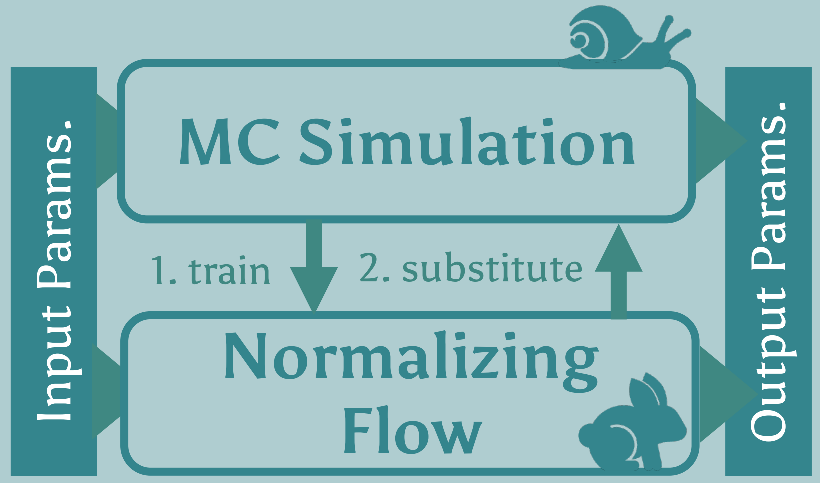
Limited Accuracy of MC:

→„Directly“ Search for kink-like signature with Neural Networks



Limited Statistics of MC:

→ Speed Up MC sampling using Normalizing Flows



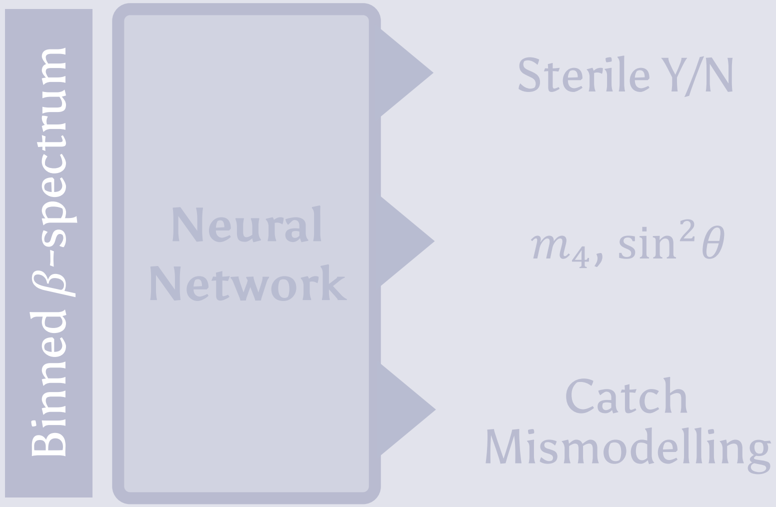
How to Deal with MC Limitations using ML?

First Ideas (work in progress)

Maybe today

Limited Accuracy of MC:

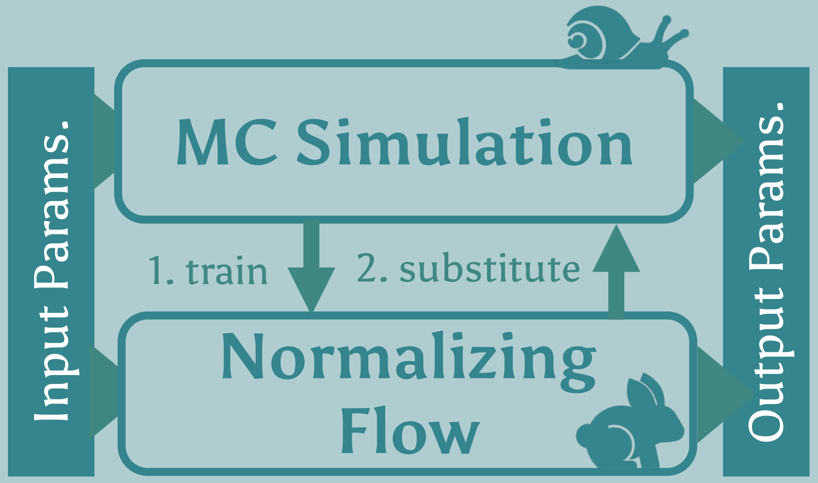
→„Directly“ Search for kink-like signature with Neural Networks



TODAY

Limited Statistics of MC:

→ Speed Up MC sampling using Normalizing Flows

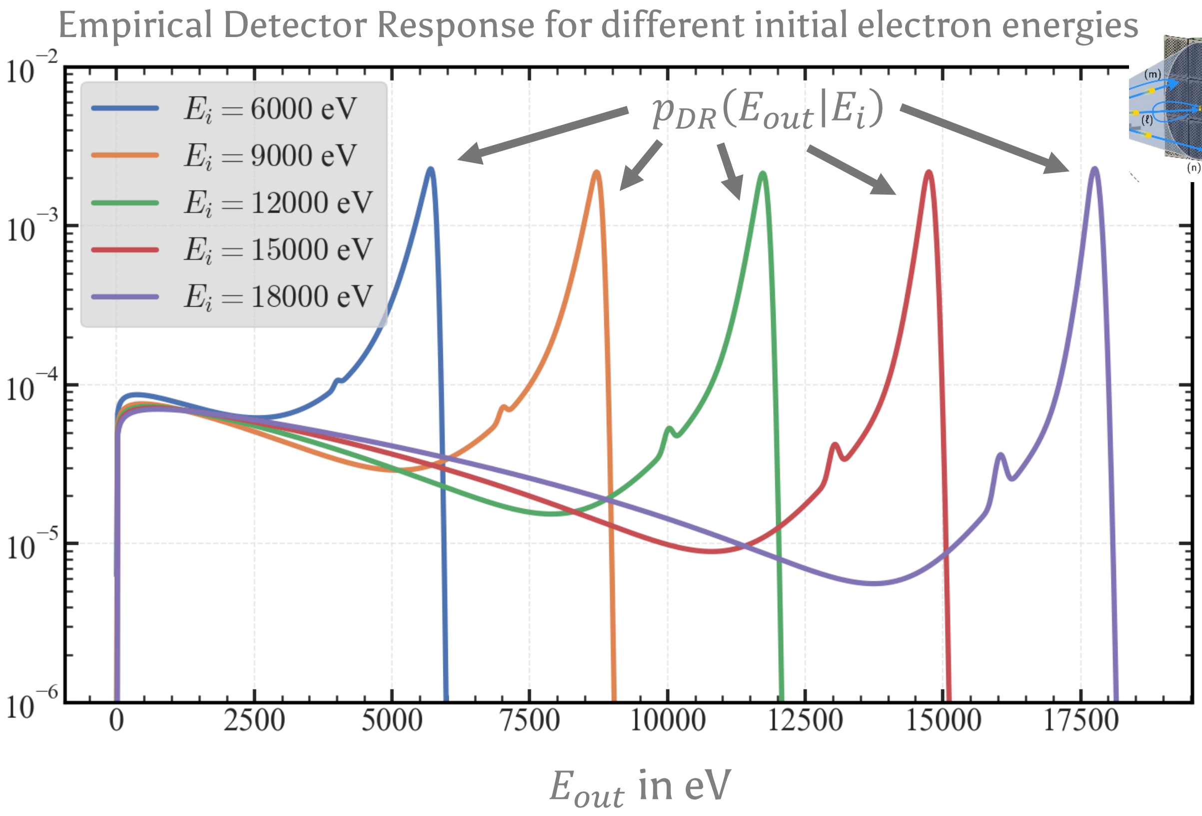


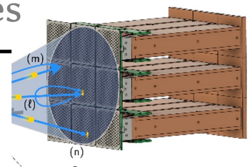
Flow Based Monte Carlo - Setting

Feasibility study on simple example:

Are the flows:

- Accurate Enough?
- Fast Enough?
- Can we draw more samples than we train with?





Flow Based Monte Carlo - Setting

Feasibility study on simple example:

Are the flows:

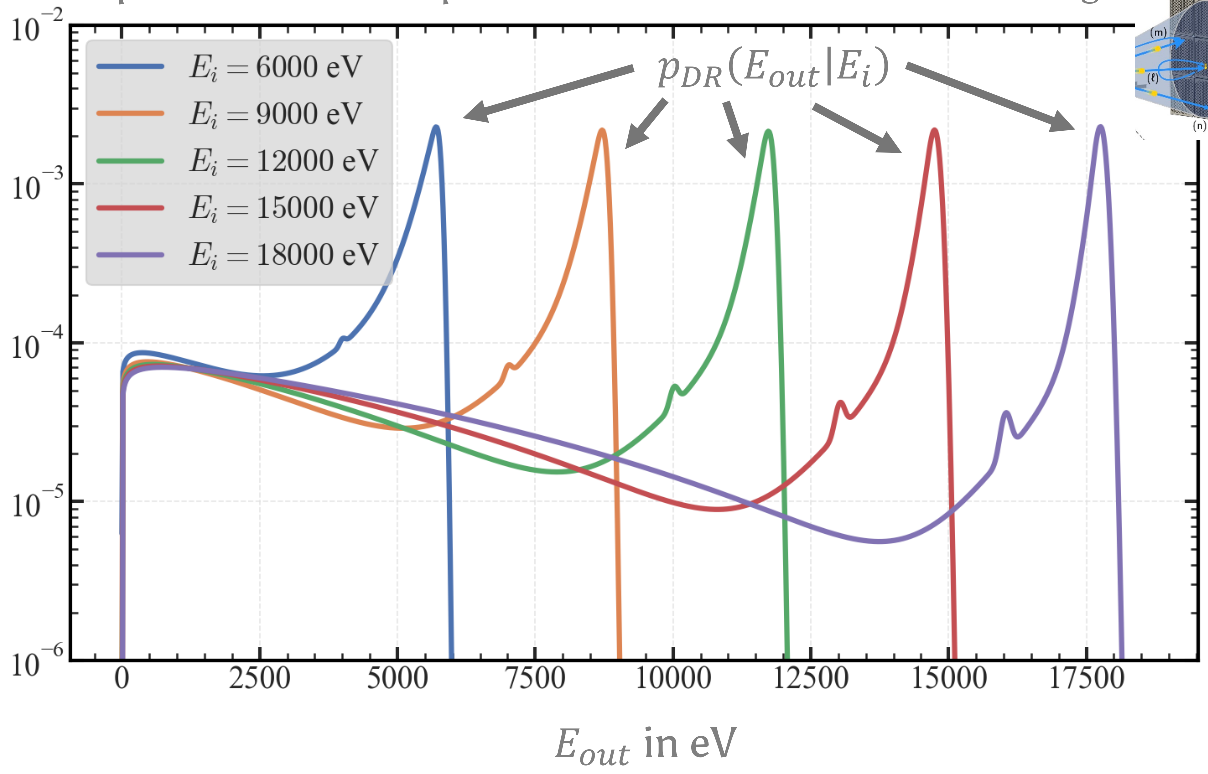
- Accurate Enough?
- Fast Enough?
- Can we draw more samples than we train with?

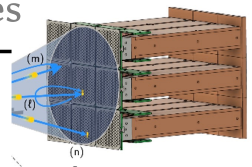
Training Data

1. $E_i \sim \mathcal{U}(0, 18.6)$ keV
2. $E_{out} \sim p_{DR}(E_{out}|E_i)$

repeat N-times for N samples

Empirical Detector Response for different initial electron energies





Flow Based Monte Carlo - Setting

Feasibility study on simple example:

Are the flows:

- Accurate Enough?
- Fast Enough?
- Can we draw more samples than we train with?

Training Data

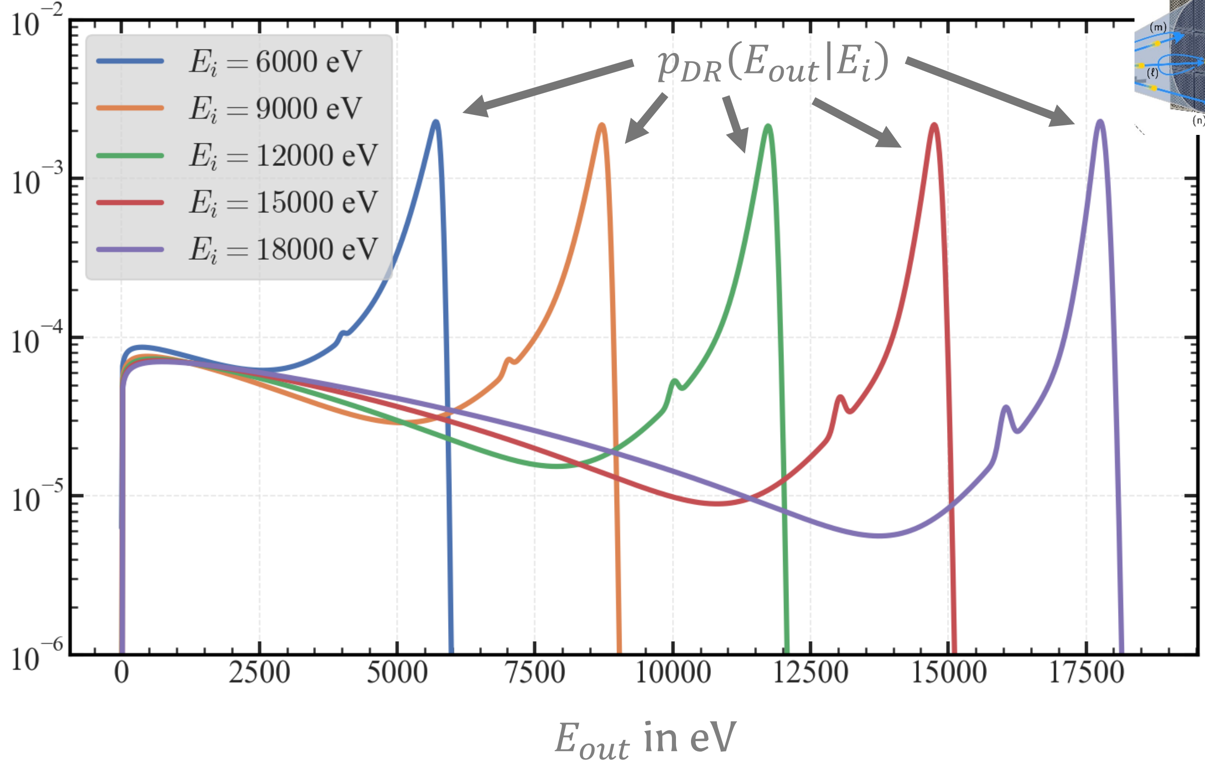
1. $E_i \sim \mathcal{U}(0, 18.6)$ keV
2. $E_{out} \sim p_{DR}(E_{out}|E_i)$

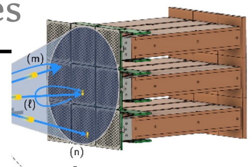
repeat N-times for N samples

Models / Techniques

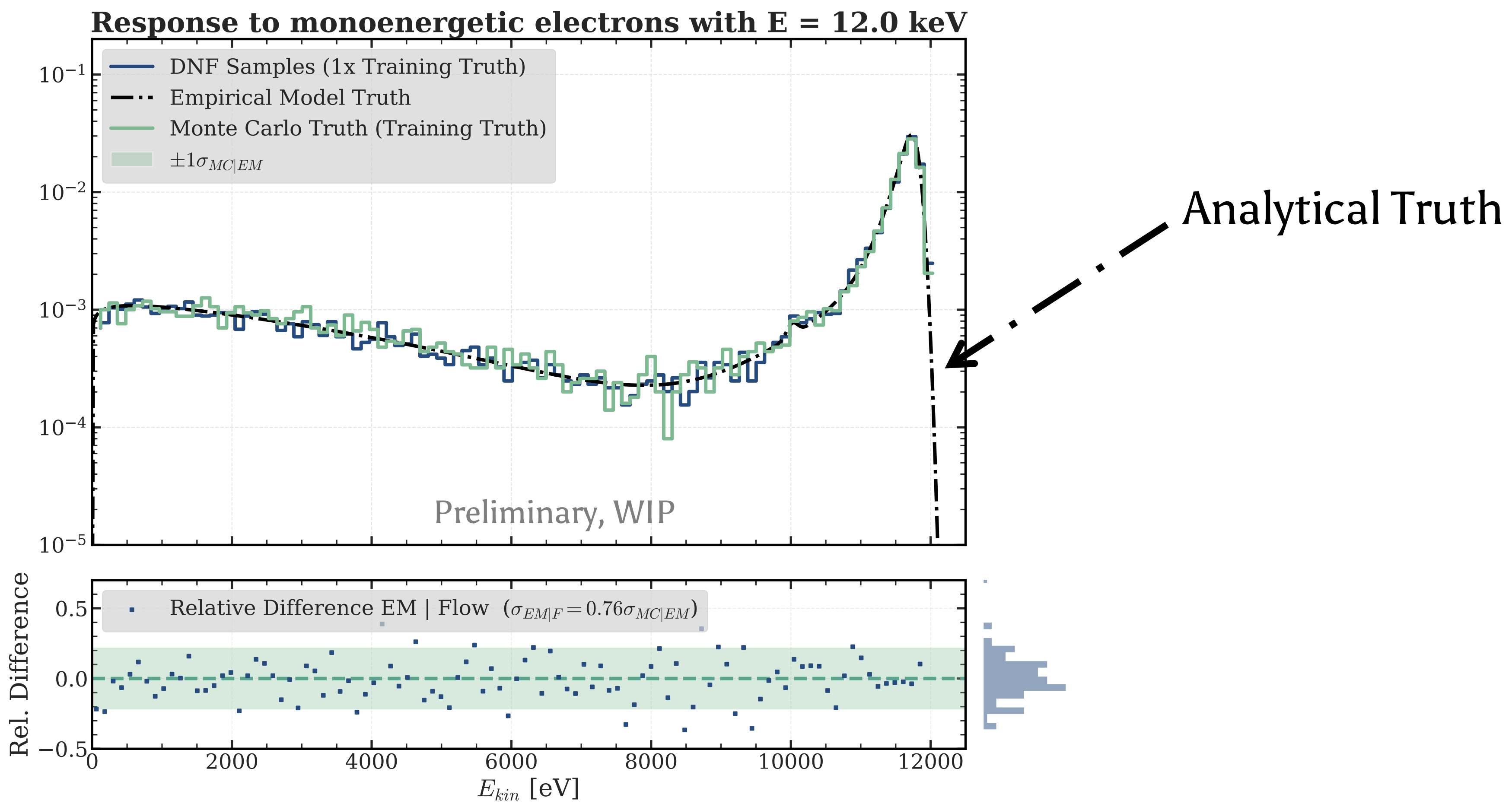
1. Discrete Normalizing Flow (RQ-Spline)
2. MLP trained in Flow-Matching scenario (OT CFM)

Empirical Detector Response for different initial electron energies

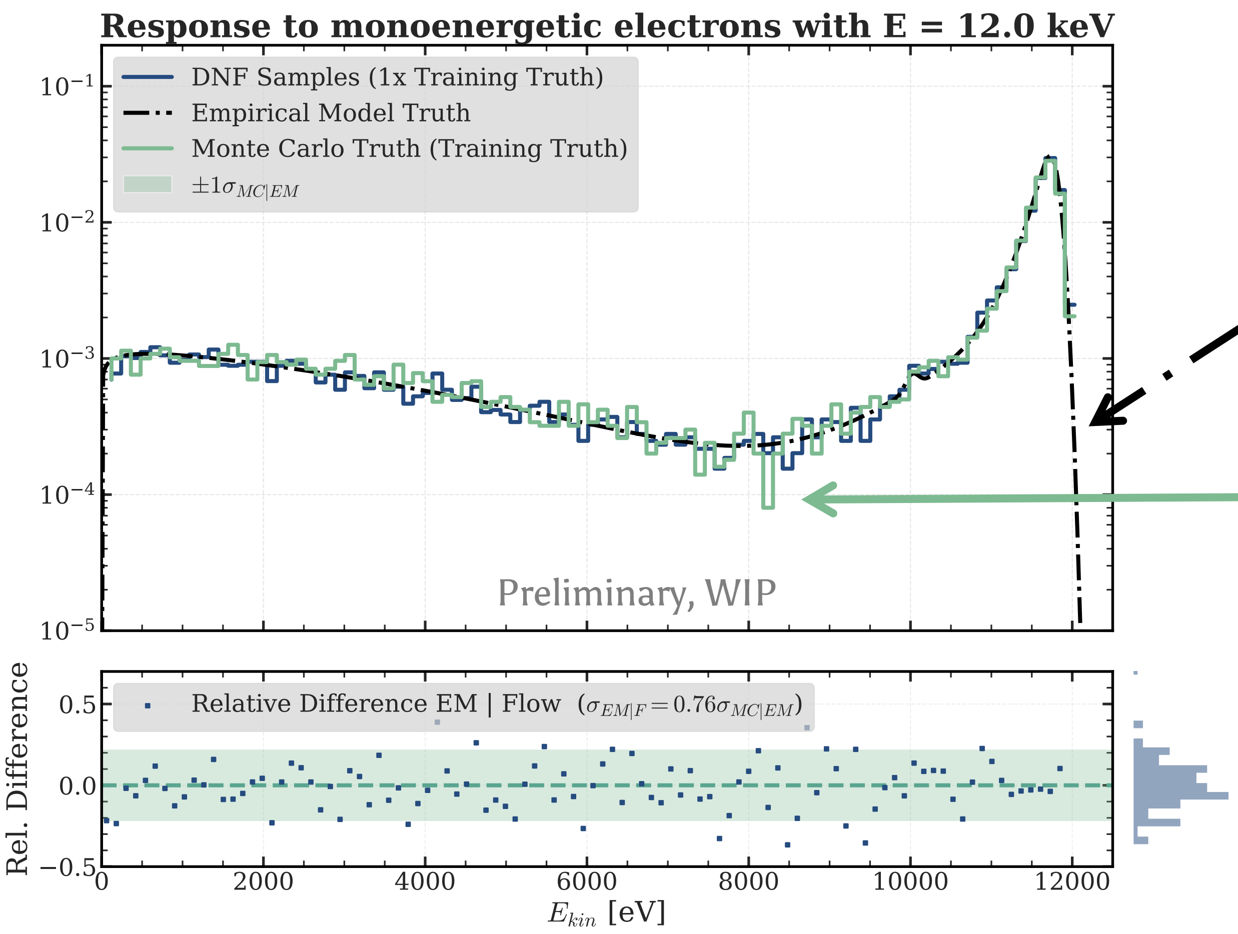




Flow Based Monte Carlo - Results



Flow Based Monte Carlo - Results



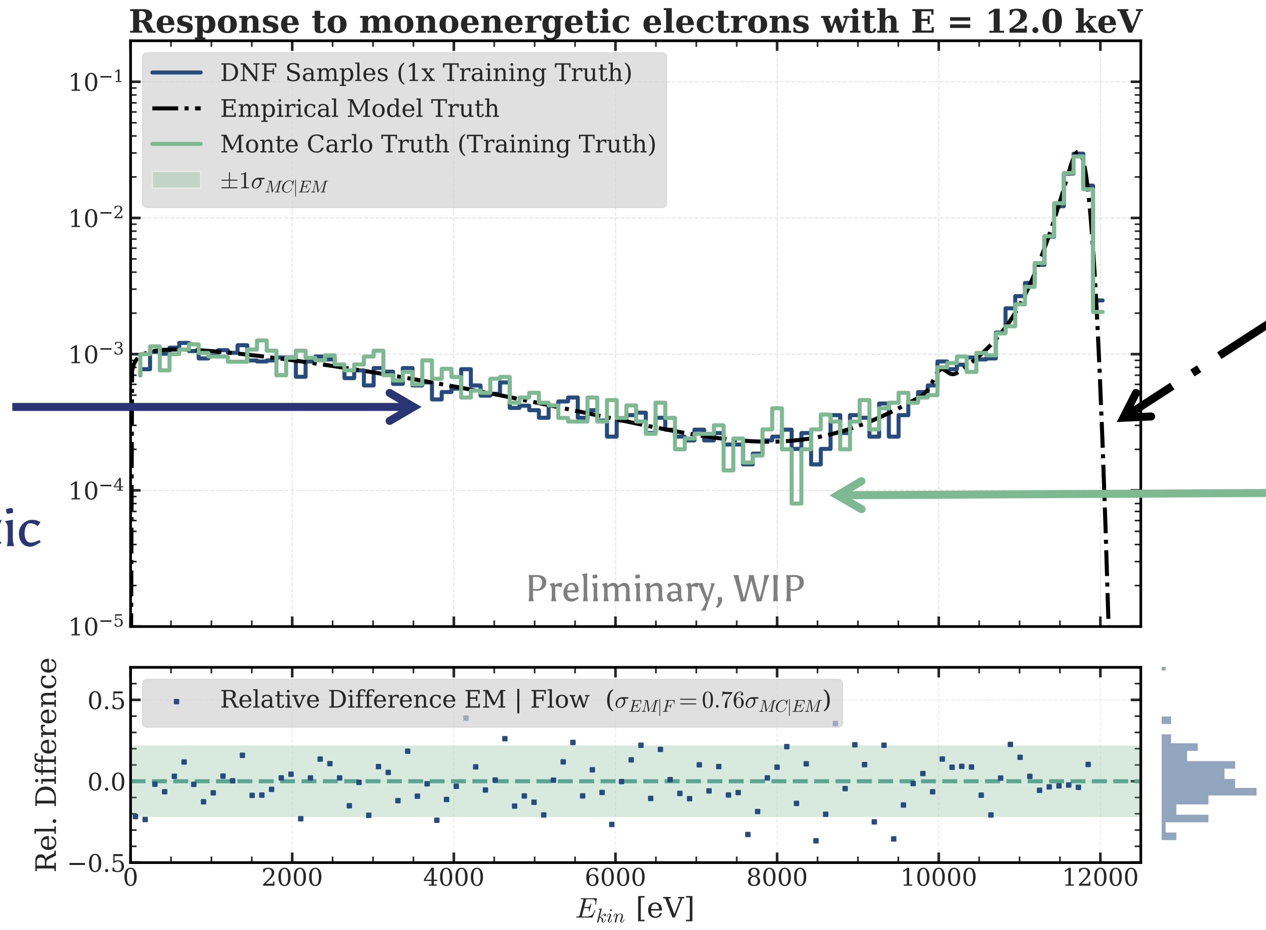
Analytical Truth

Samples from Analytical Truth used as training data („MC-truth“)

Flow Based Monte Carlo - Results

Samples from trained Flow

(@ same statistic of MC Truth)



Analytical Truth

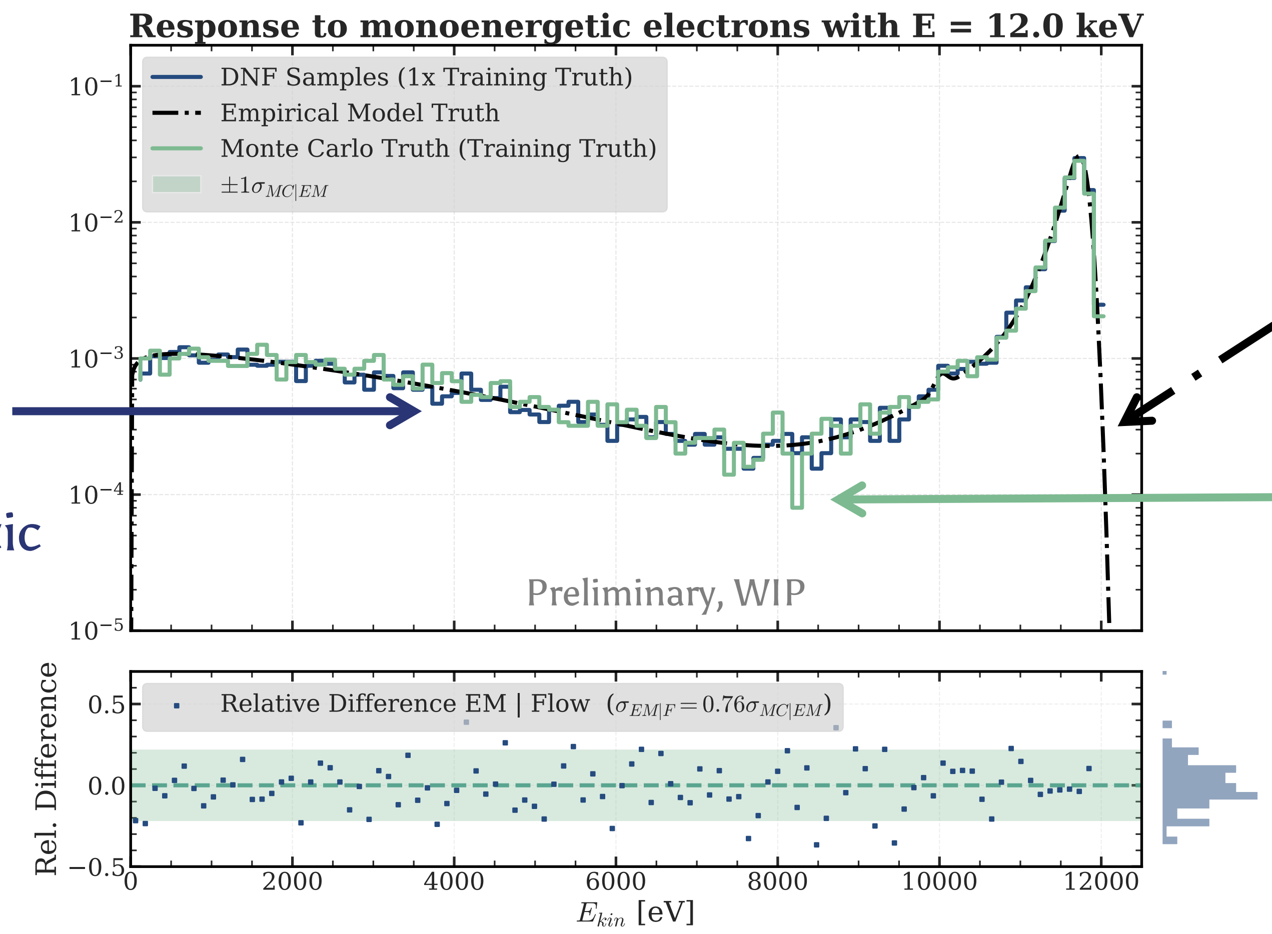
Samples from Analytical Truth used as training data („MC-truth")

Flow Based Monte Carlo - Results

@ same statistic of MC Truth: Inference speedup by 500x with virtually no quality loss

Samples from trained Flow

(@ same statistic of MC Truth)



Analytical Truth

Samples from Analytical Truth used as training data („MC-truth“)

Flow Based Monte Carlo - Limitations

What happens at large statistics N?

Dataset Sizes become prohibitive (e.g. 10^{13} events → PB-scale)

Training Times become prohibitive

Can we draw more samples than we train with ?

Flow Based Monte Carlo - Limitations

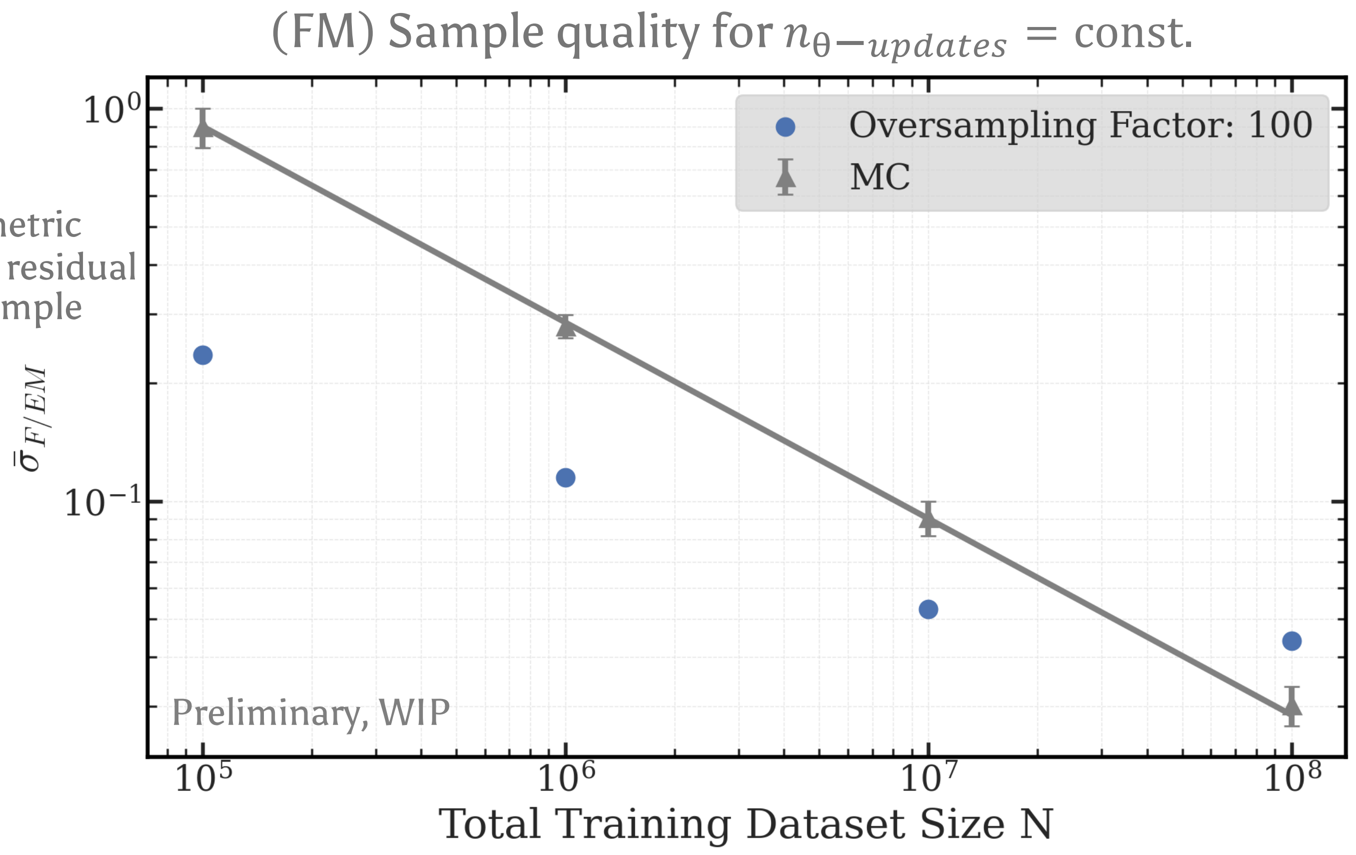
What happens at large statistics N?

Dataset Sizes become prohibitive (e.g. 10^{13} events → PB-scale)

Training Times become prohibitive

Can we draw more samples than we train with ?

Quality metric based on residual truth - sample



Flow Based Monte Carlo - Limitations

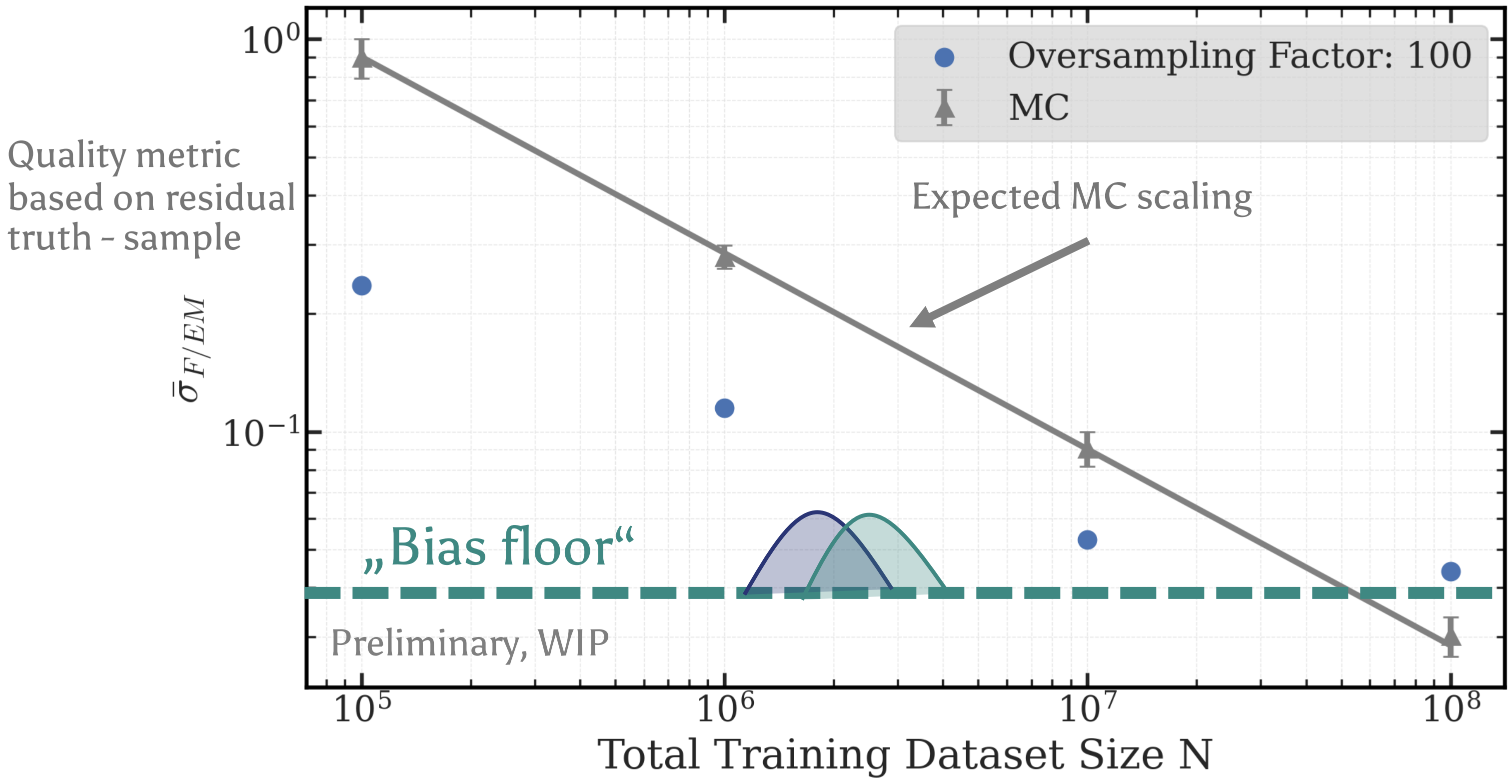
What happens at large statistics N?

Dataset Sizes become prohibitive (e.g. 10^{13} events → PB-scale)

Training Times become prohibitive

Can we draw more samples than we train with ?

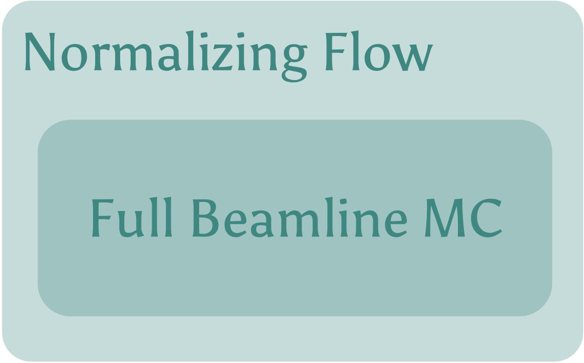
(FM) Sample quality for $n_{\theta-updates}$ = const.



∃ Limit when keeping $n_{\theta-updates}$ = const!

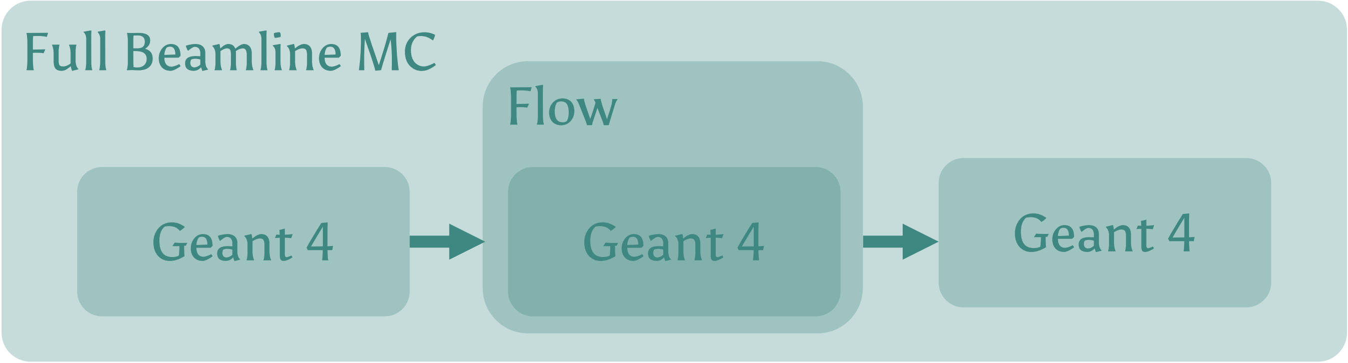
Diminishing returns for increased effort

Flow Based Monte Carlo – Where would this be used?



Highest speedup

not modular



Only speed up bottleneck MCs

Highly modular

Conclusion & Outlook

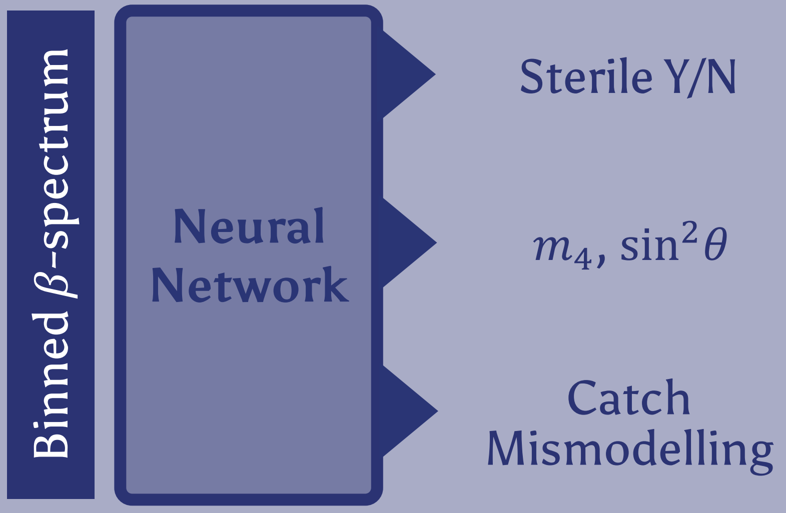
- Accuracy & speed of sample generation demonstrated
- Currently scalable to datasets with $\mathcal{O}(10^9)$ events
  → fast generator suitable for sensitivity studies & smaller calibration measurements
- still very early in development, lots of work in progress

How to Deal with MC Limitations using ML?

First Ideas (work in progress)

Limited Accuracy of MC:

→„Directly“ Search for kink-like signature with Neural Networks



Limited Statistics of MC:

→ Speed Up MC sampling using Normalizing Flows

Search for Sterile Signature with NN

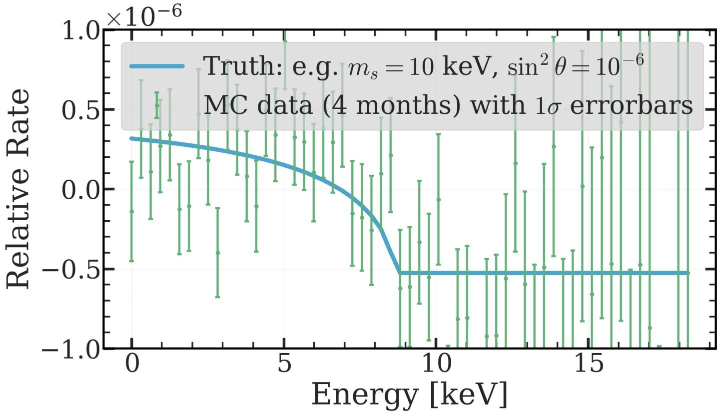
Motivation:

NN could learn higher-order correlations to "semantically" learn the kink-like signature

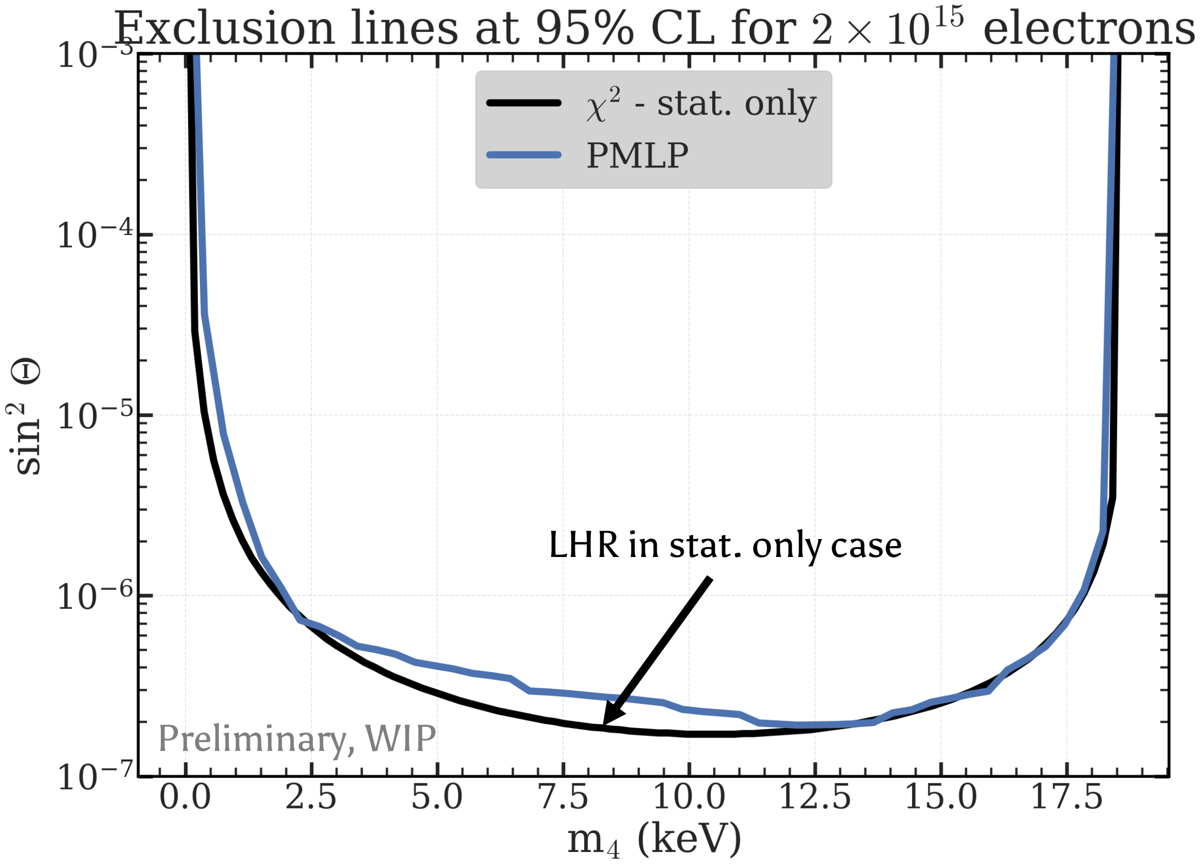
First Check:

Are Neural Networks even sensitive to the sterile neutrino signature?

Example MC Signature



~1 yr.



Search for Sterile Signature with NN

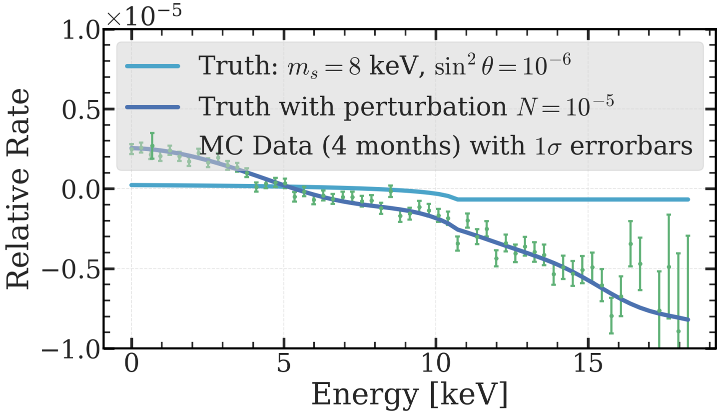
Motivation:

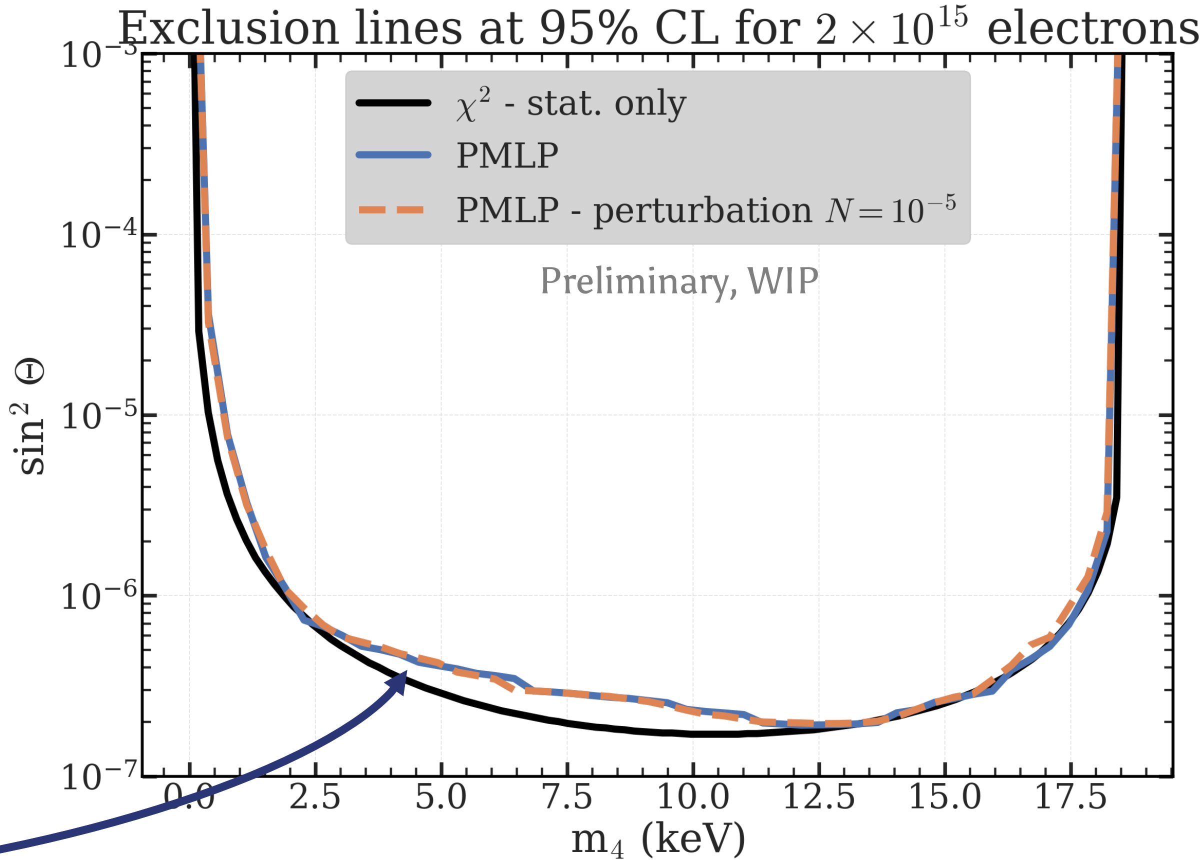
NN could learn higher-order correlations to "semantically" learn the kink-like signature

First Check:

Are Neural Networks even sensitive to the sterile neutrino signature?

MC Sig. w/ smooth bias





Add Bias ∉ {Training Data} → No sensitivity loss for NN

Search for Sterile Signature with NN

Another Example:

Can do bin-wise tests for the stat. significance of sterile signature or mismodelling, etc.

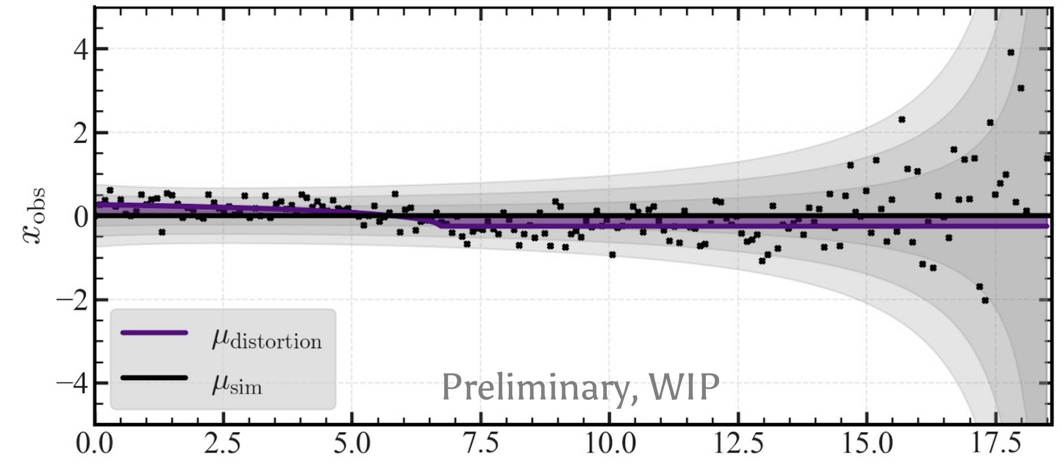
Idea:

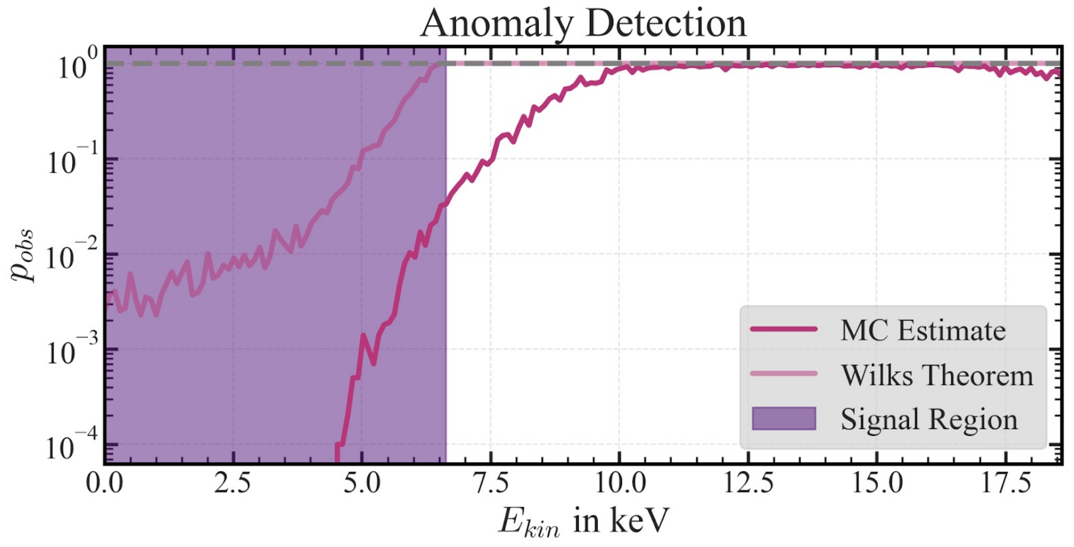
Train bin-wise classifier, NN will learn LHR for each bin when minimizing BCE („LHR-trick“)

How to get p-values:

Use a MC estimate, for a global p-value: take care to include “look-elsewhere-effect“!

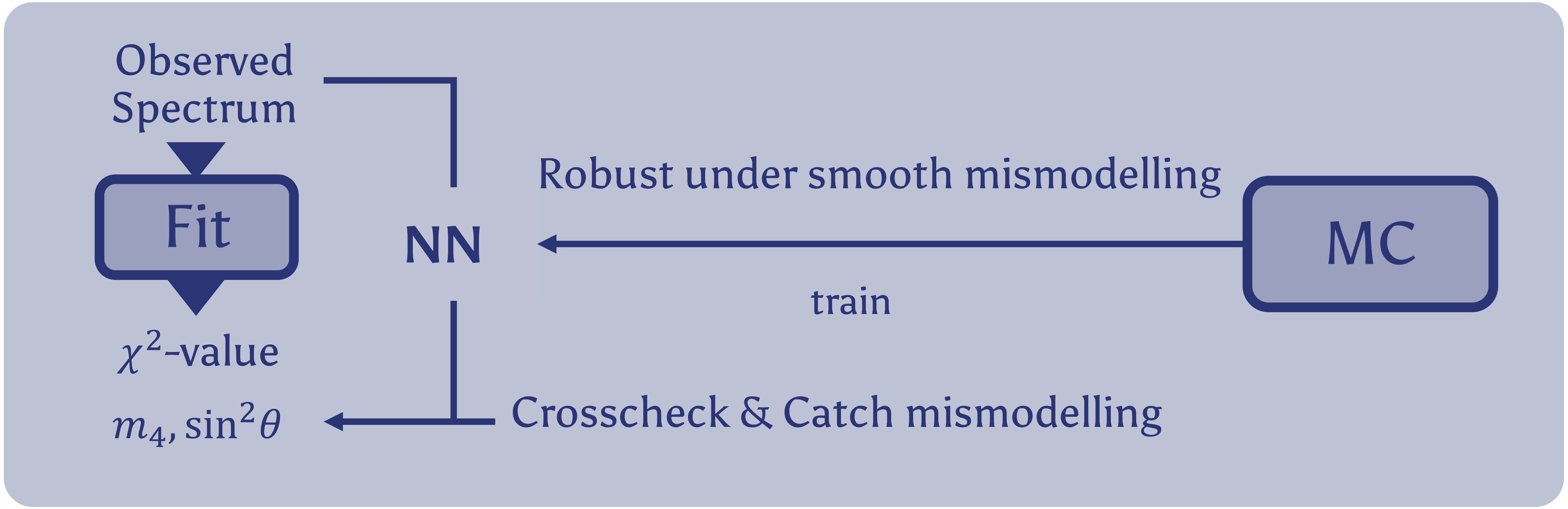
Bin-Wise stat. Significance of sterile signature





Search for Sterile Signature with NN - Outlook

Where would this be used?



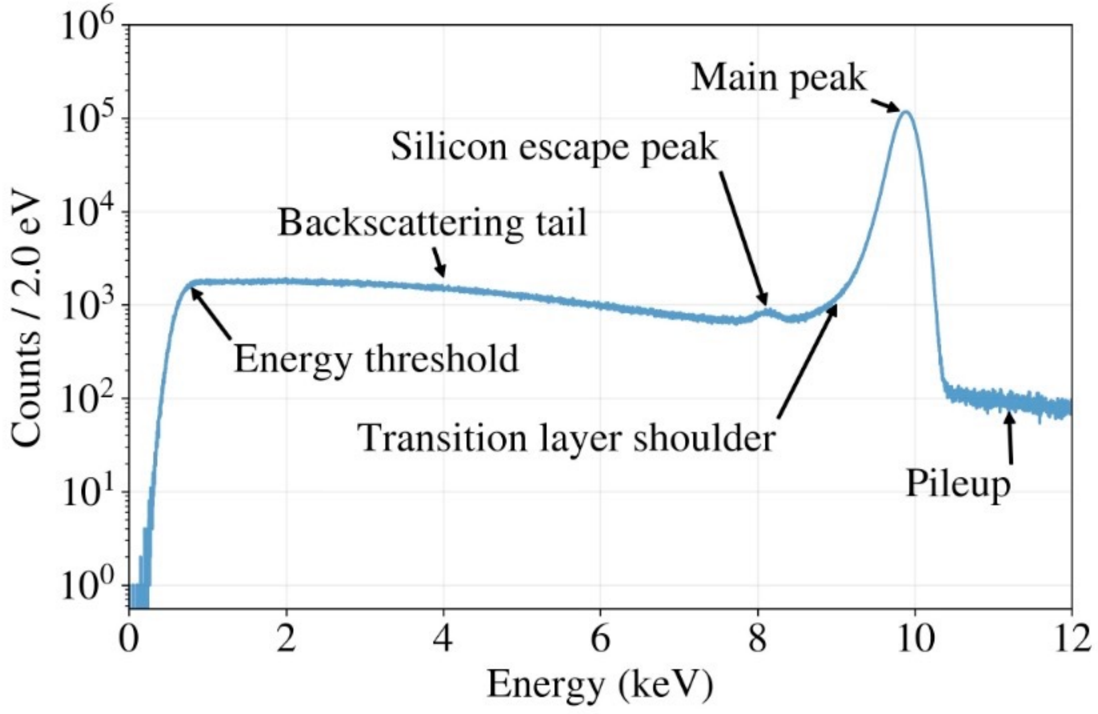
Limitations

- Still somewhat reliant on the MC (used as training data, cant be too far off)
- Still to be tested for non-smooth / larger group of perturbations

Backup 1: Flow Based MC

Real Response of TRISTAN Detector

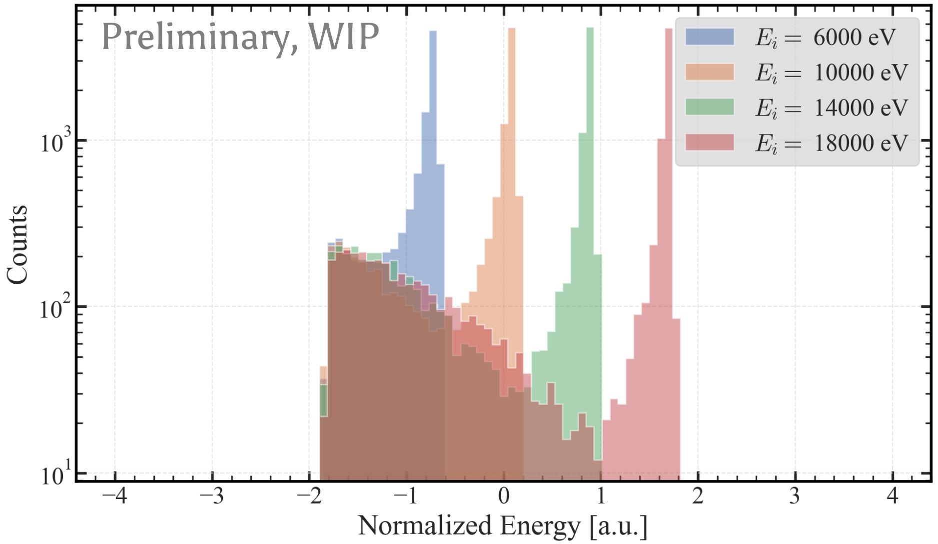
To a monoenergetic electron beam



Flow Based Monte Carlo - Data

Simple Standardization

„Choice of units“ $\vec{x} \rightarrow \frac{\vec{x} - mean(\vec{x})}{var(\vec{x})}$ with $\vec{x} = (E_{out}, E_i)$
(also called z-score normalization)



z-score norm

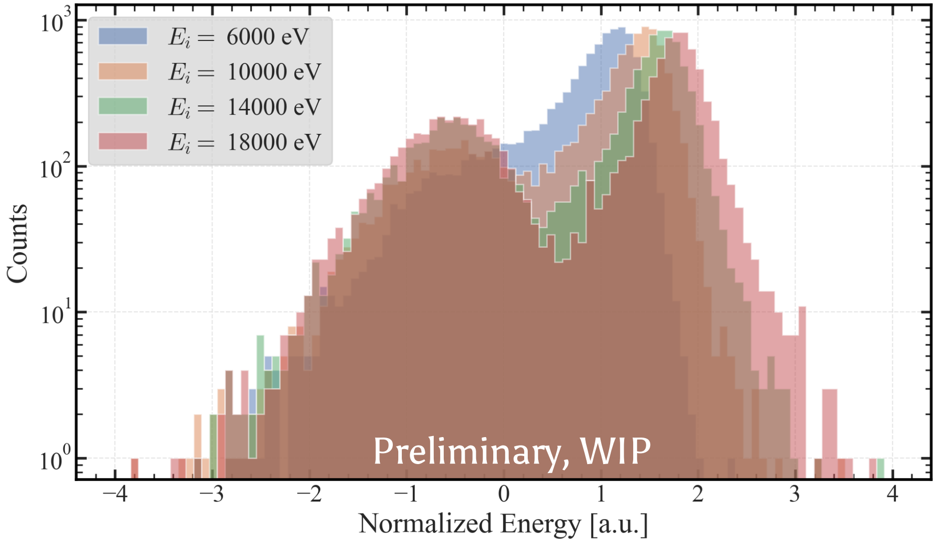
Energy-based Standardization

Want to mitigate sharp cutoffs in E_{out} close to E_i

$$u = \frac{E_{out}}{E_i} \in [0, 1]$$

$$\tilde{u} = \alpha + (1 - 2\alpha)u \text{ with } \alpha = 10^{-6}$$

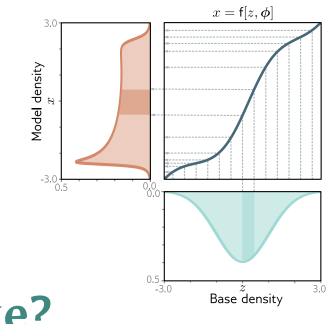
$$u_{final} = \log_{10}\left(\frac{\tilde{u}}{1 - \tilde{u}}\right)$$



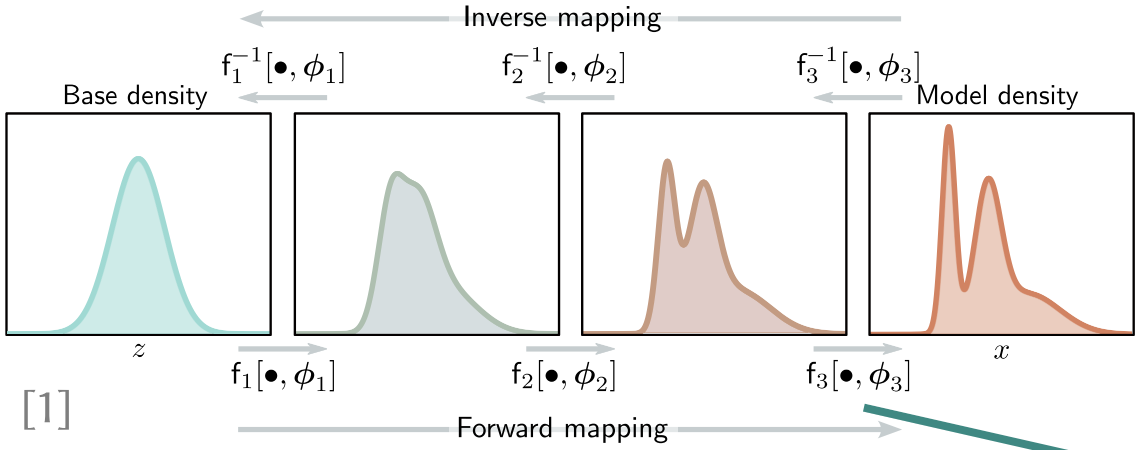
„energy“ norm

[1] https://udlbook.github.io/udlbook/

Flow Based Monte Carlo - Model



(Conditional) Discrete Normalizing Flows



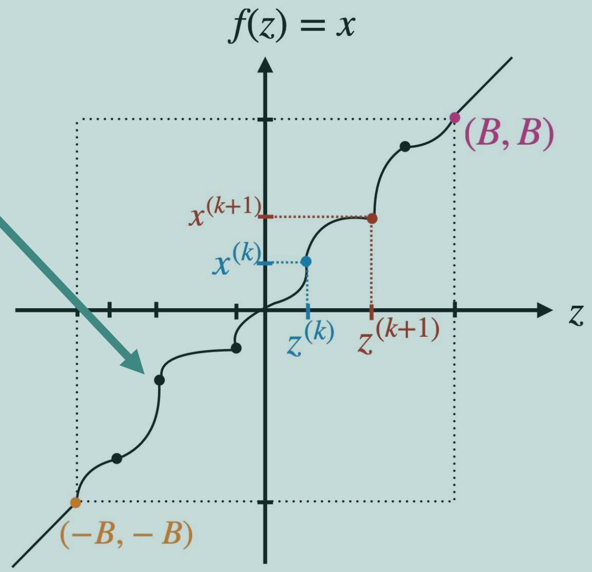
[1]

How does a RQS transform look like?

Additional Conditional Info concatenated to NN input

NN

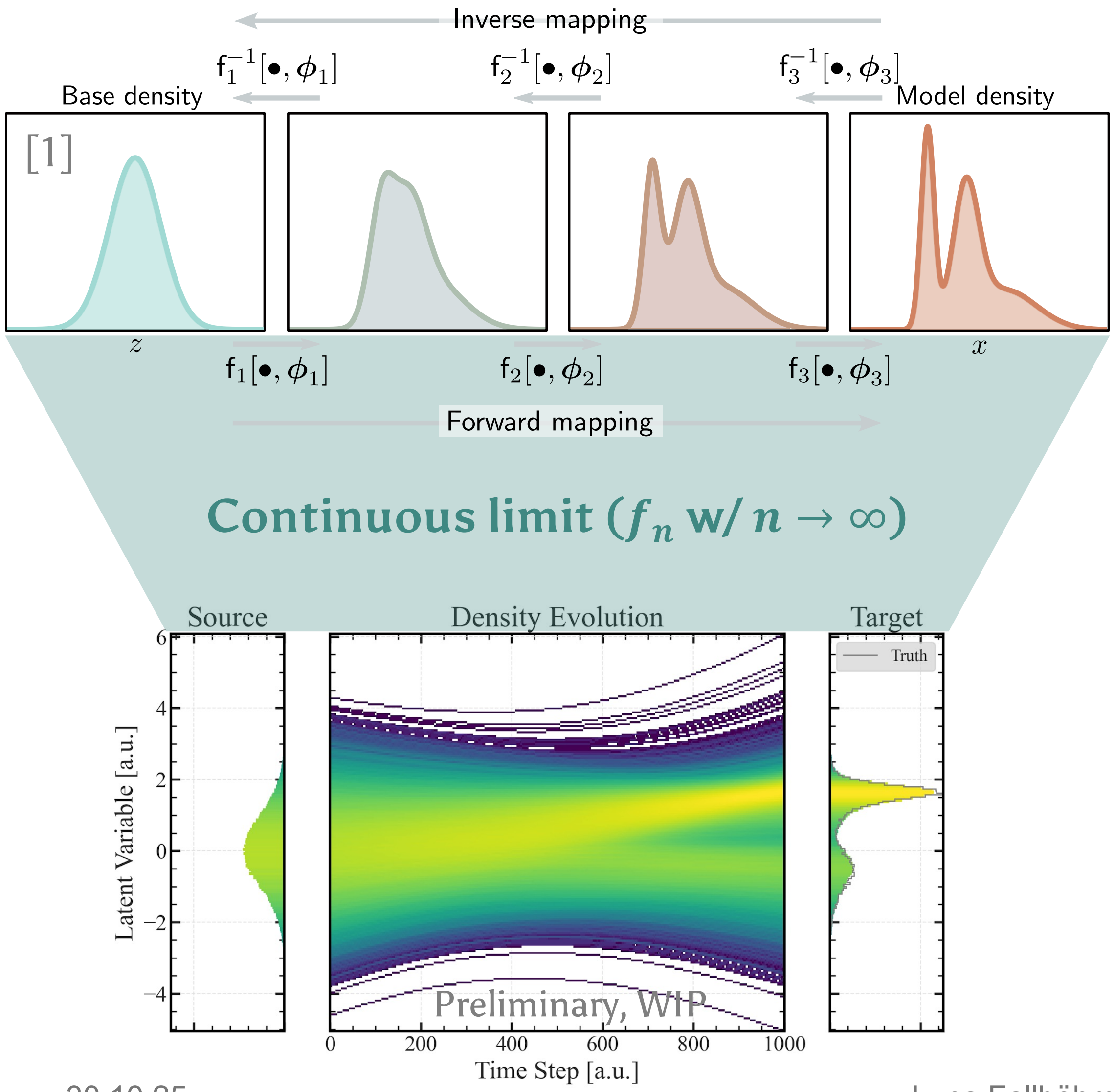
NN just learns the positions & derivatives @ these positions, constraining the RQ-spline



How do they work?

- Minimize KL-Div. between model and target density by minimizing NLL of samples!
- NNs learn series of invertible transforms while keeping Jacobians in check (Jacobian <-> „density info“)
- Allows sampling & density estimation!

Flow Based Monte Carlo - Model



Going from discrete transformations to continuous field

Updates to latent variable z become infinitesimally small

$$\frac{dz(t,\text{ctxt})}{dt} = u(z, t, \text{ctxt});\ t \in [0, 1]$$

w/ $z(t = 0) = z_0$ and $z(t = 1) = x$

Neural ODE: train NN to predict $u(z, t, \text{ctxt})$, call it $v_\theta(z, t, \text{ctxt})$

To do training / density est.:

$$z_0 = x - \int_0^1 v_\theta(z, t, \text{ctxt})dt$$

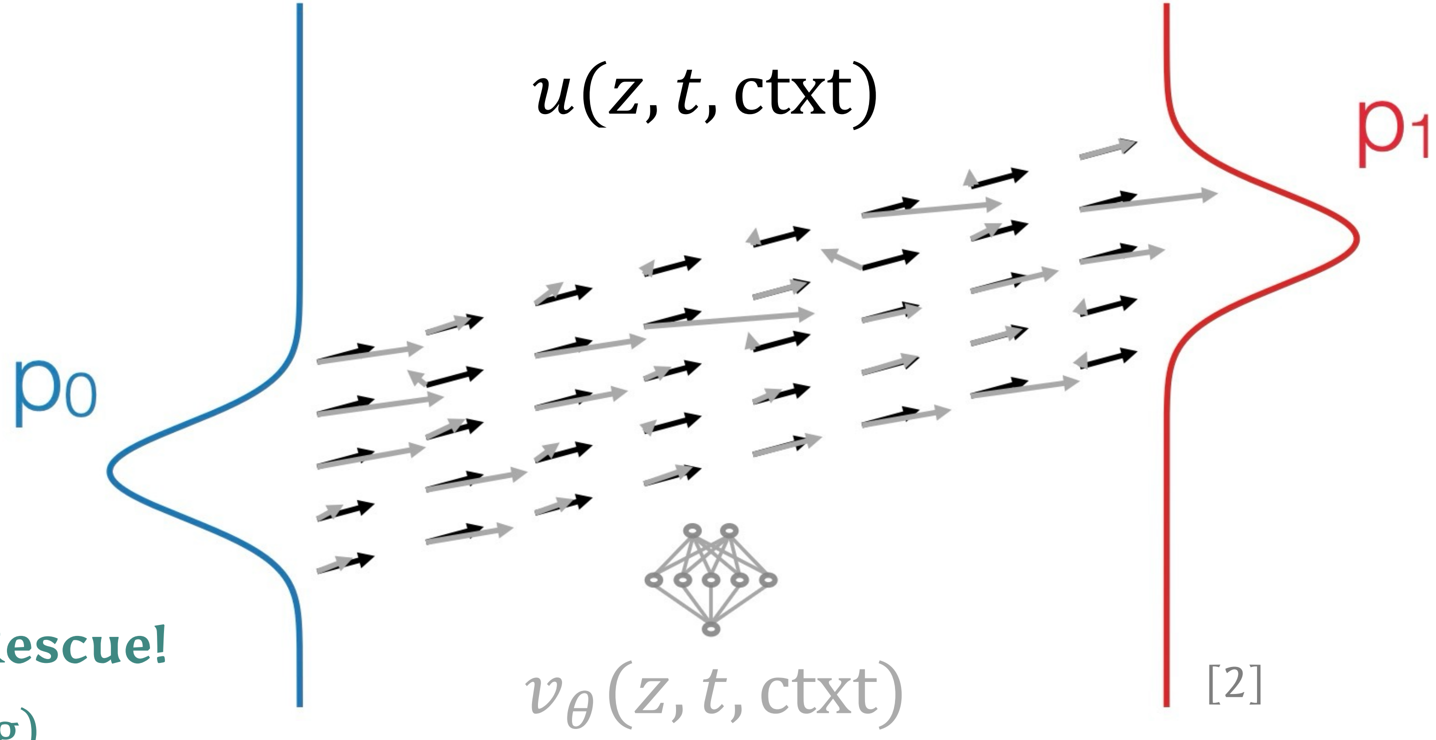
Integration @ Training Time!

Flow Based Monte Carlo - Model

Do regression on the ODE velocity field!

Integration @ training time:

- Very time consuming
- (doesn‘t scale nicely to higher D)



(Conditional) Flow Matching to the Rescue!

(no integration required for training)

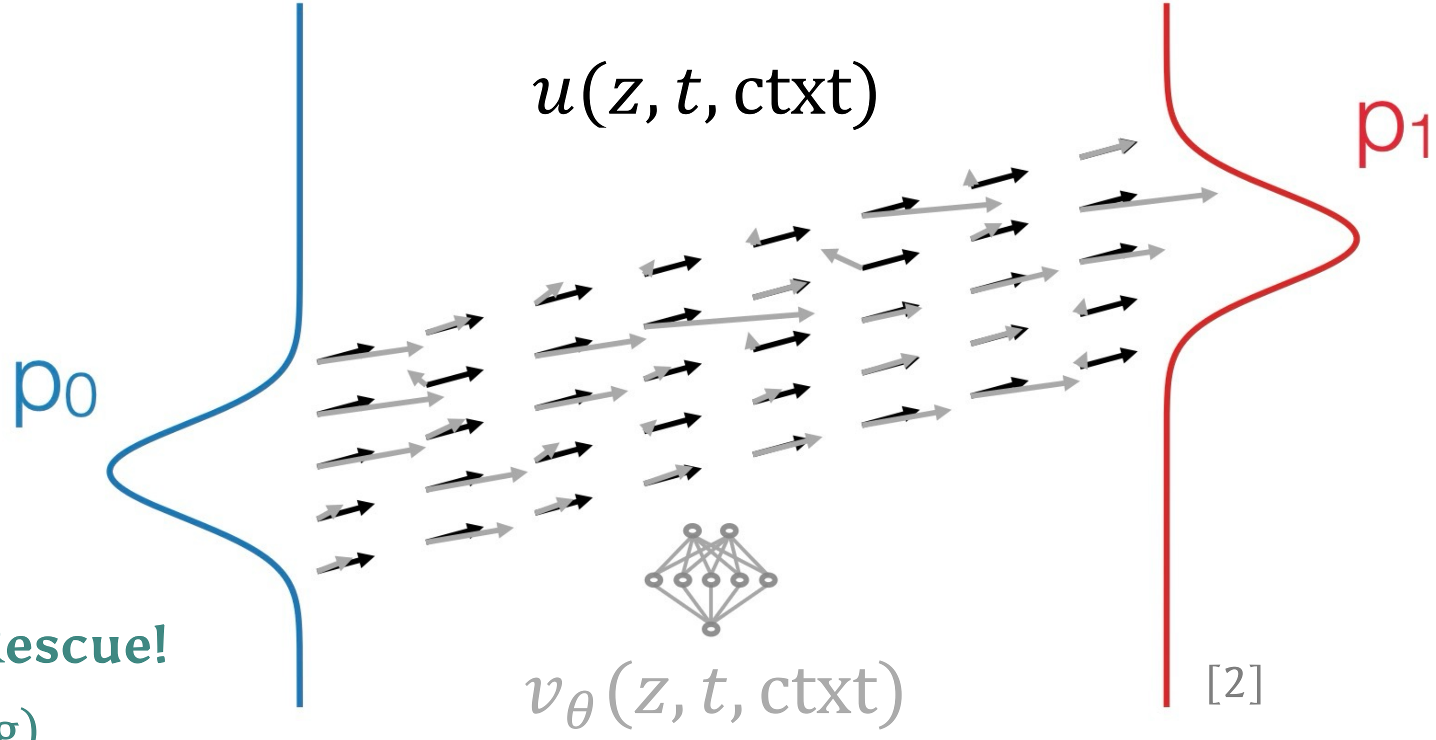
$$\mathcal{L}_{CFM} = \mathbb{E}_{t,z_0,x}[\|v_\theta(z,t,\mathrm{ctxt}) - u(z,t,\mathrm{ctxt} \mid z_0, x)\|^2]$$

Flow Based Monte Carlo - Model

Do regression on the ODE velocity field!

Integration @ training time:

- Very time consuming
- (doesn‘t scale nicely to higher D)



(Conditional) Flow Matching to the Rescue!

(no integration required for training)

$$\mathcal{L}_{CFM} = \mathbb{E}_{t,z_0,x}[||v_\theta(z, t, \mathrm{ctxt}) - x - z_0||^2]$$

& straight line approx.

Flow Based Monte Carlo - Model

What we want:

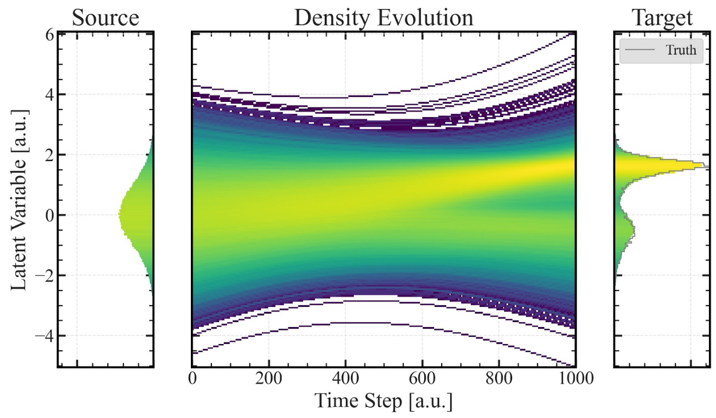
- Accurate sampling
- Fast sampling

Lightweight model

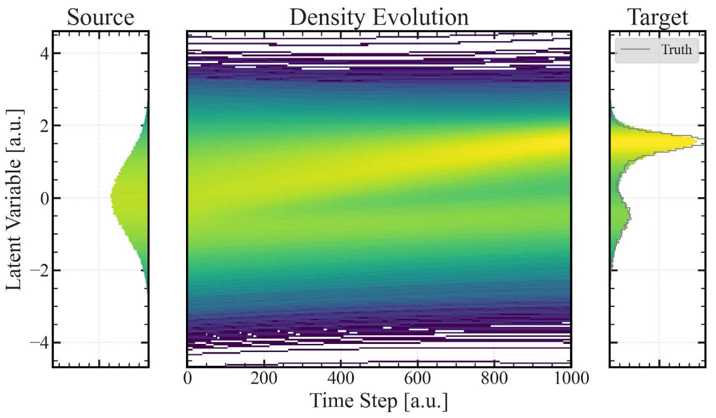
Straight prob. paths

Conditional Flow Matching with:

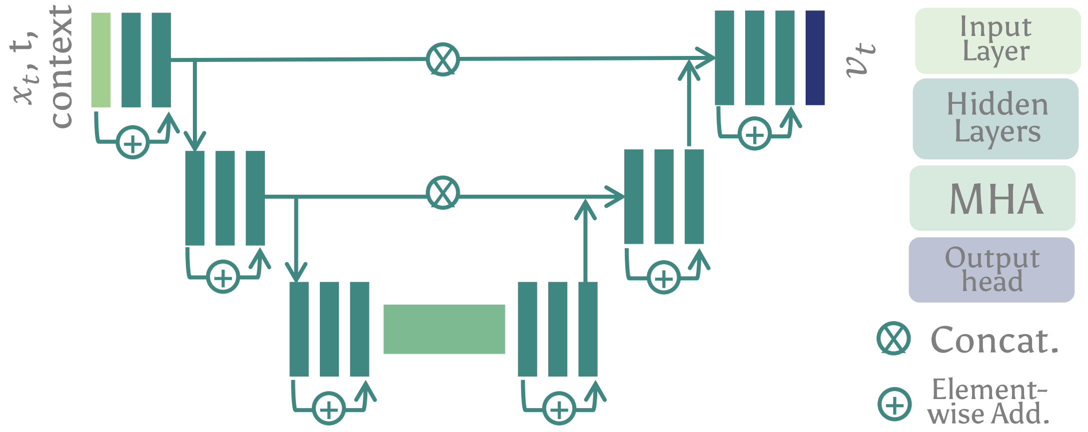
- Velocity Field Predictor: MLP with Unet-like Structure & self attn. at bottleneck
- Optimal Transport Coupling between Source and Target



CFM



CFM w/ (batch-wise) optimal transport coupling



Architecture

Unet-like to retain info from different scales

Backup 2: “Kink search” with Neural Networks

Exclusion Lines

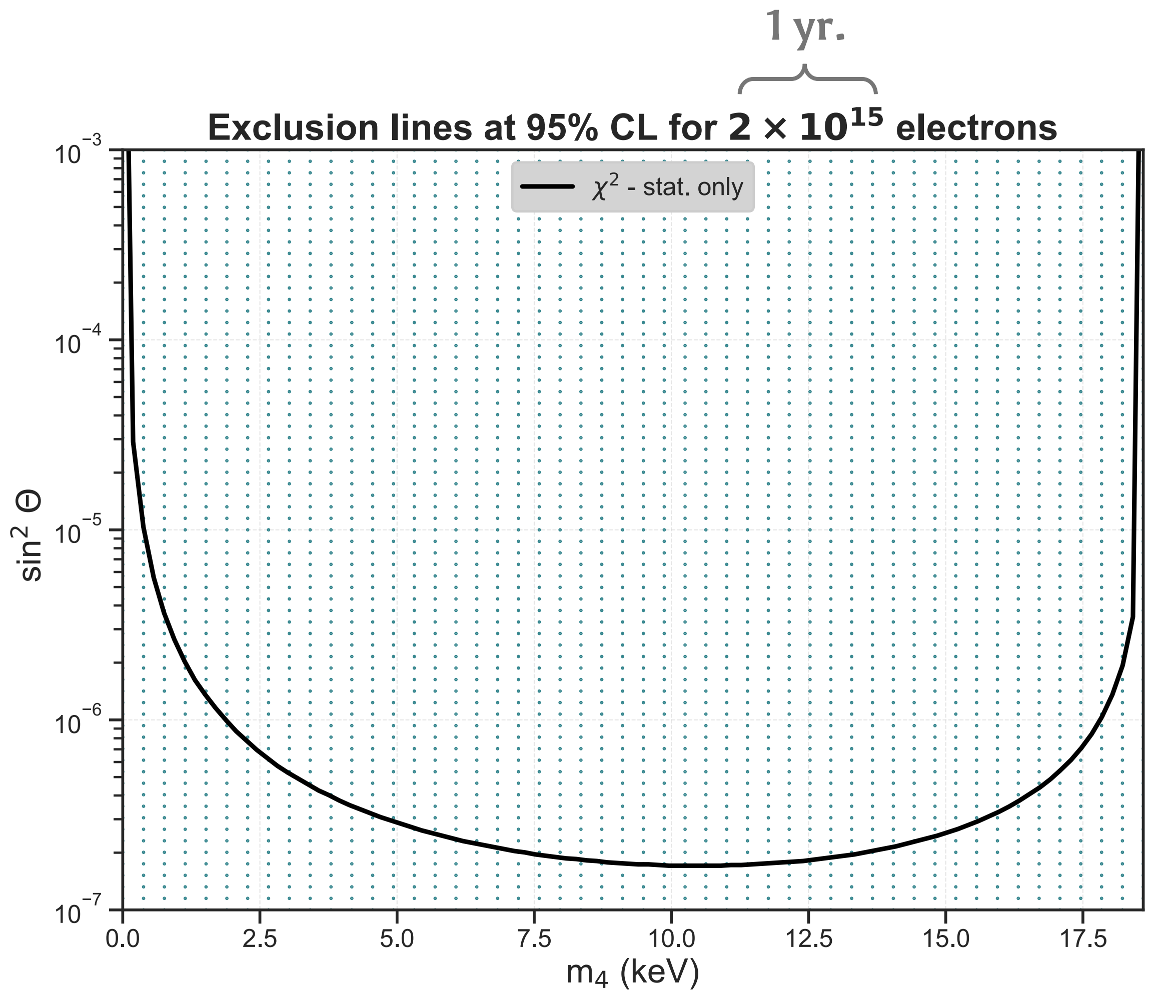
Goal: Compare NN to "traditional" method -> draw an exclusion line

χ^2-based approach

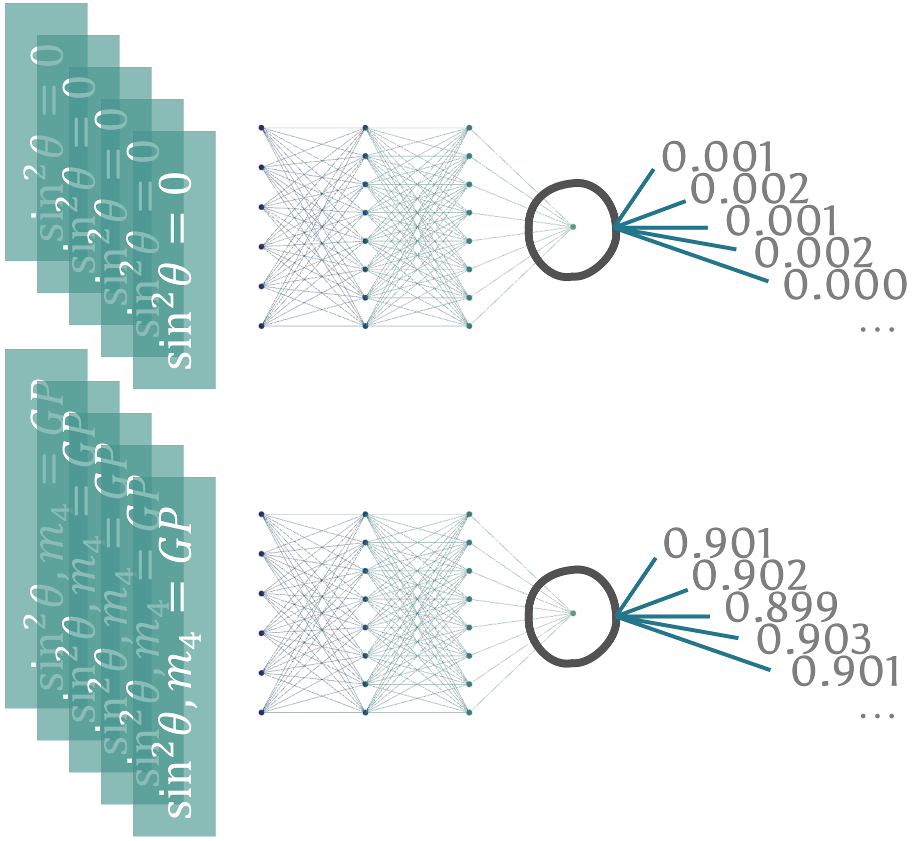
Gridscan over $\left(m_4, \sin^2(\theta)\right)$

- χ^2 computed at every gridpoint kl
- Contour drawn at 95% CL
  - ${\chi_{crit}}^2 = 5.99$ if Wilks' theorem holds
- Nuisance parameters: global signal amplitude

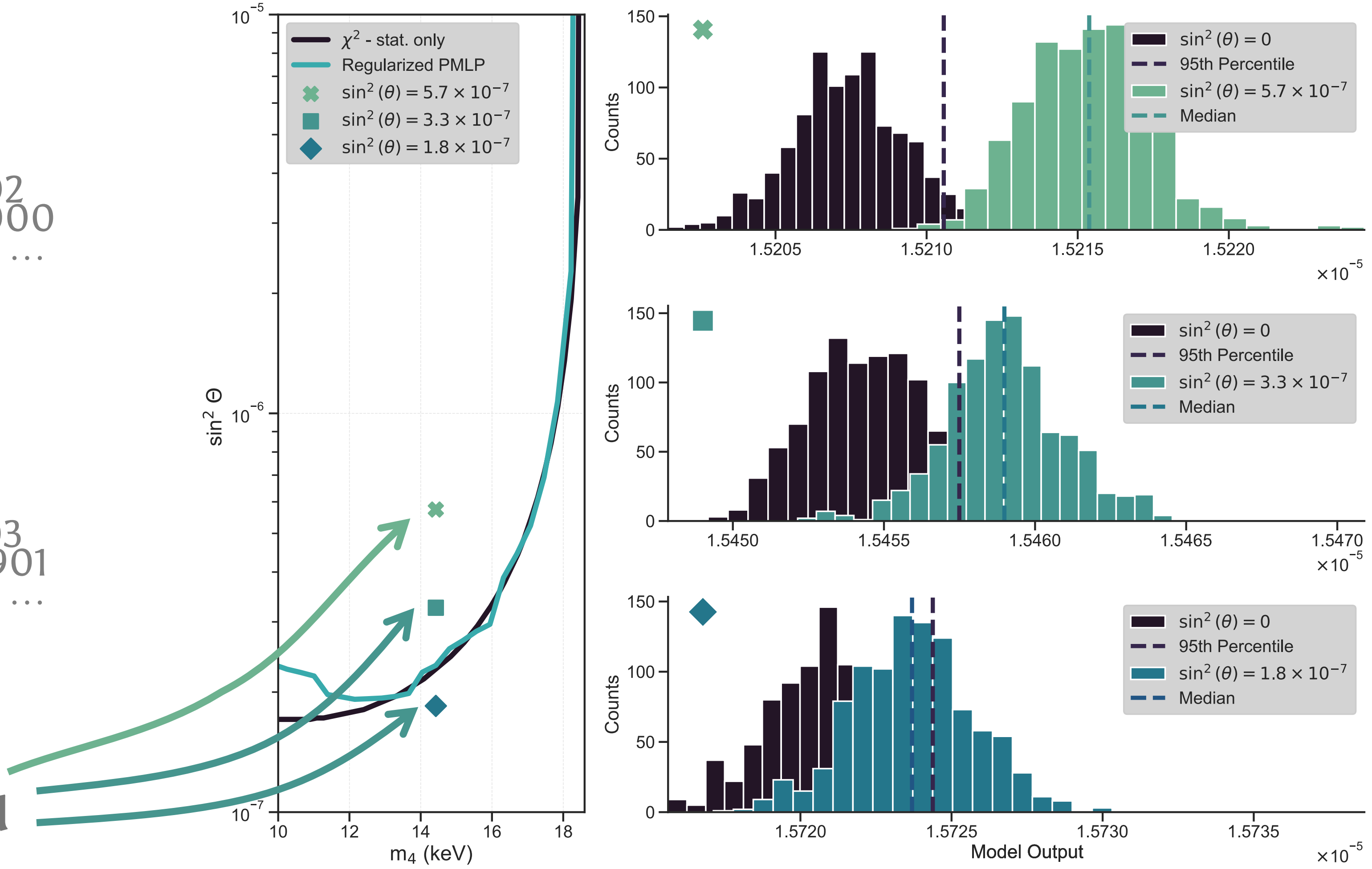
$$\chi^2_{kl}(\vec{p}) = \left(\vec{\Gamma}_{kl}(\vec{p}) - \vec{\Gamma}_{\mathrm{ref}}\right)^T V^{-1} \left(\vec{\Gamma}_{kl}(\vec{p}) - \vec{\Gamma}_{\mathrm{ref}}\right)$$



Exclusion Lines using Neural Networks

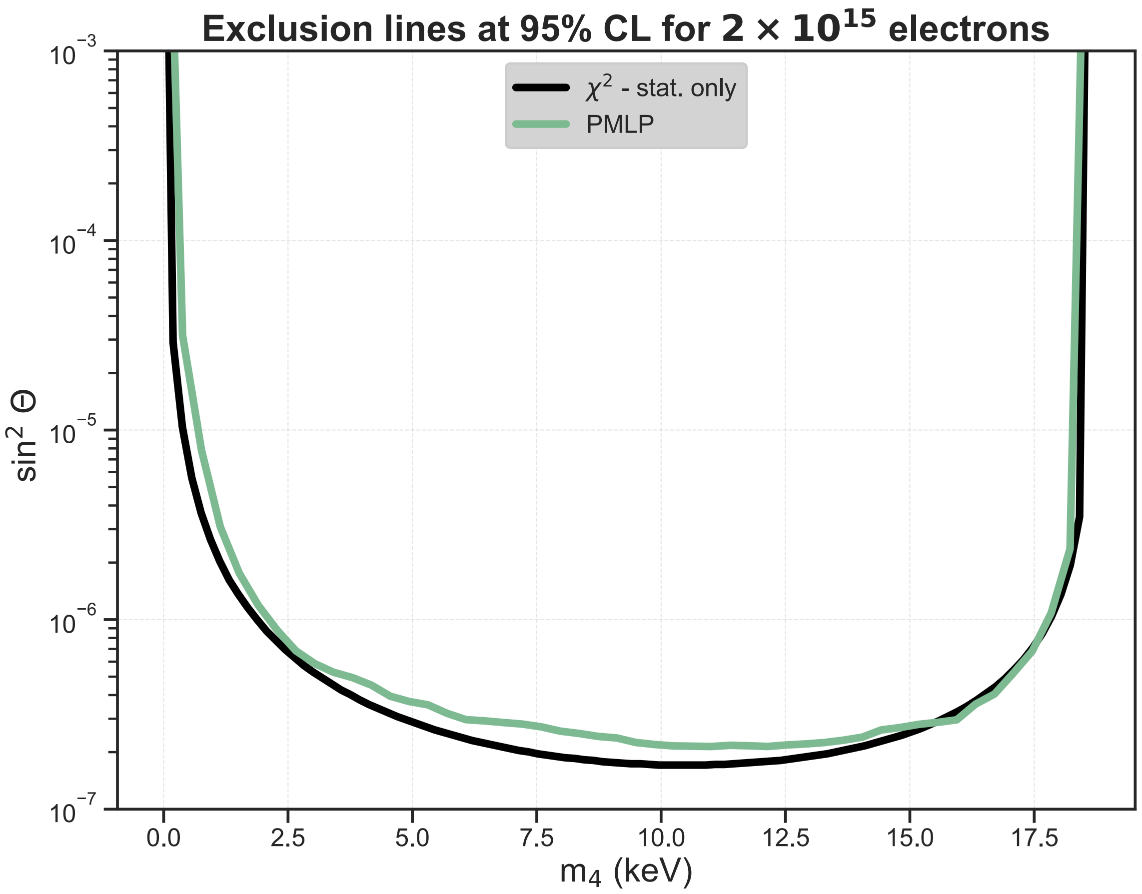


Scan through $\left(m_4, \sin^2(\theta)\right)$ **grid**



Three Takeaways

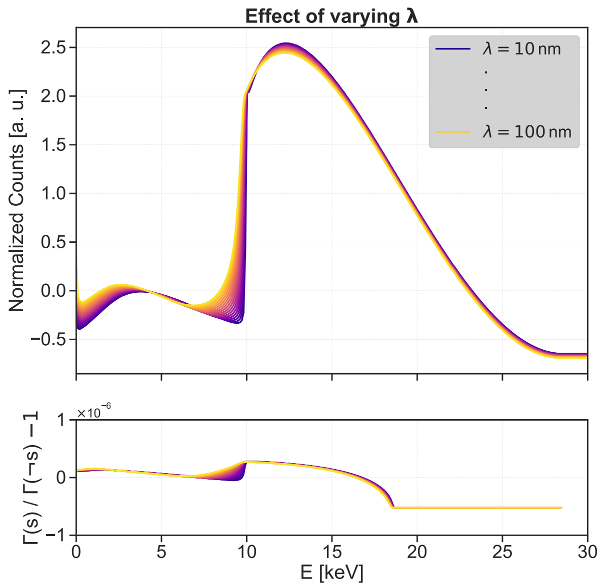
1. NNs are very sensitive to the sterile neutrino signature

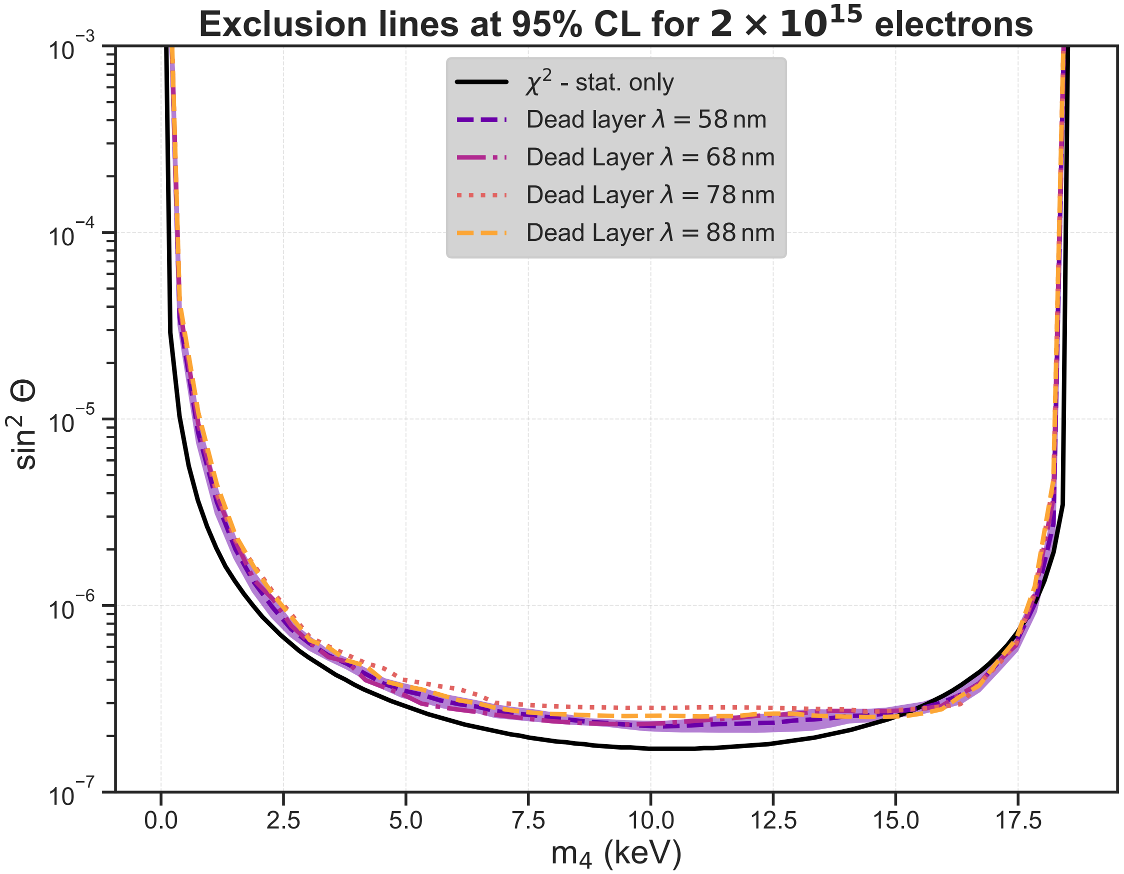


Three Takeaways

1. NNs are very sensitive to the sterile neutrino signature

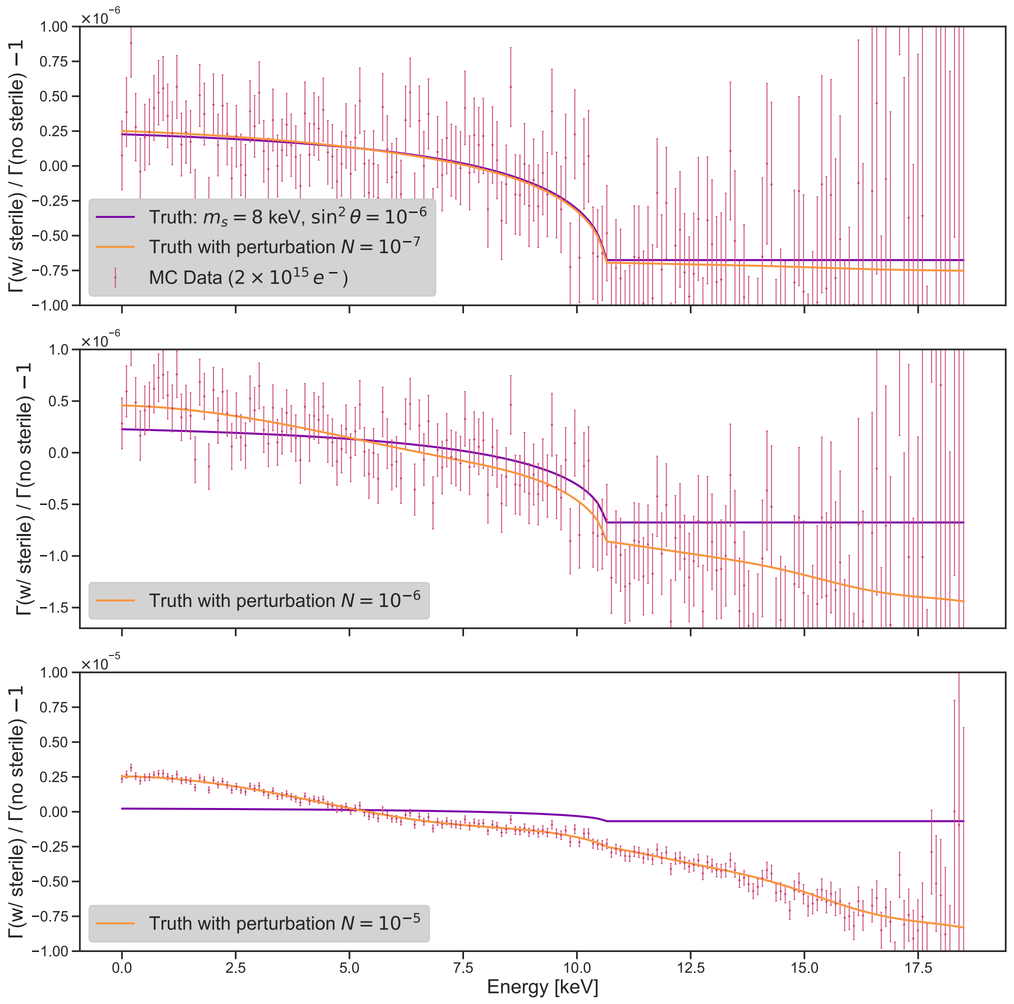
2. Can handle detector related syst. effects and uncertainties





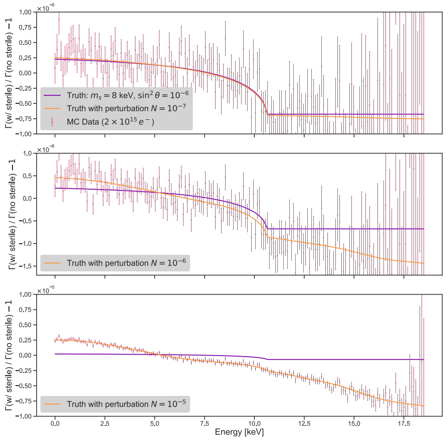
Three Takeaways

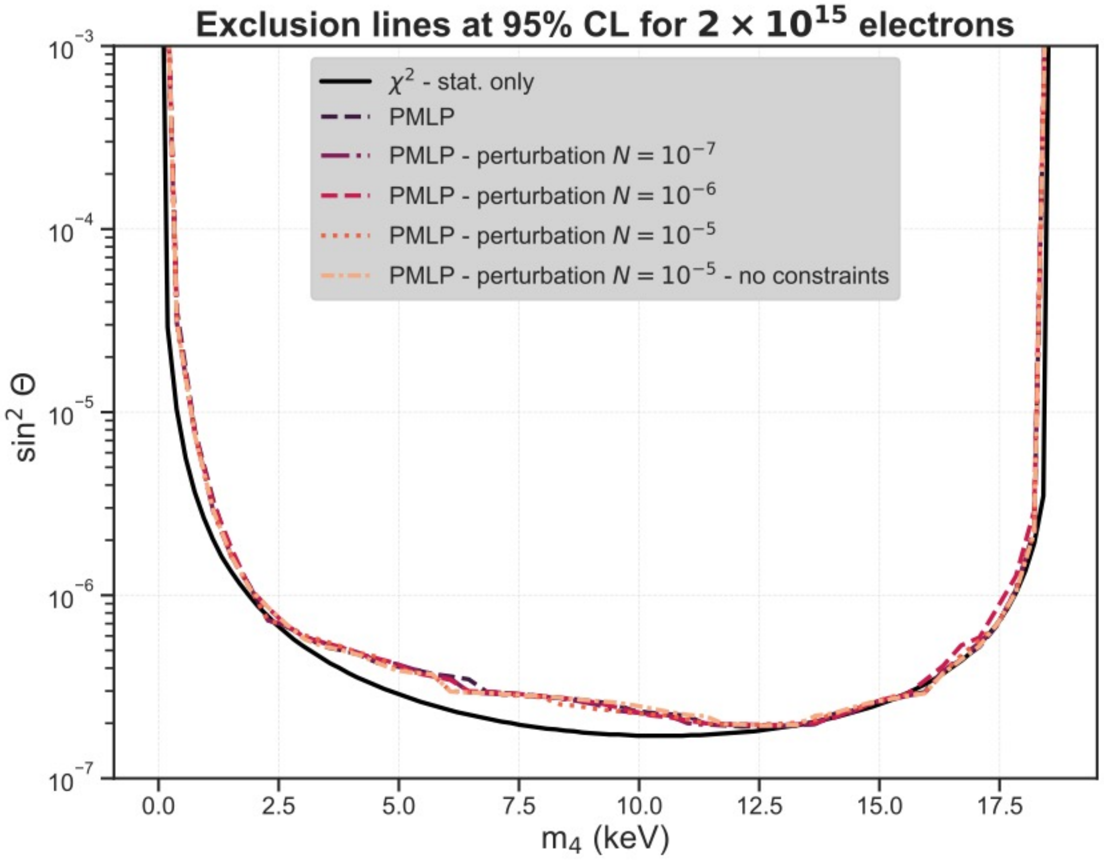
1. NNs are very sensitive to the sterile neutrino signature
2. Can handle detector related syst. effects and uncertainties
3. Can handle small smooth modelling inaccuracies



Three Takeaways

1. NNs are very sensitive to the sterile neutrino signature
2. Can handle detector related syst. effects and uncertainties
3. Can handle small smooth modelling inaccuracies



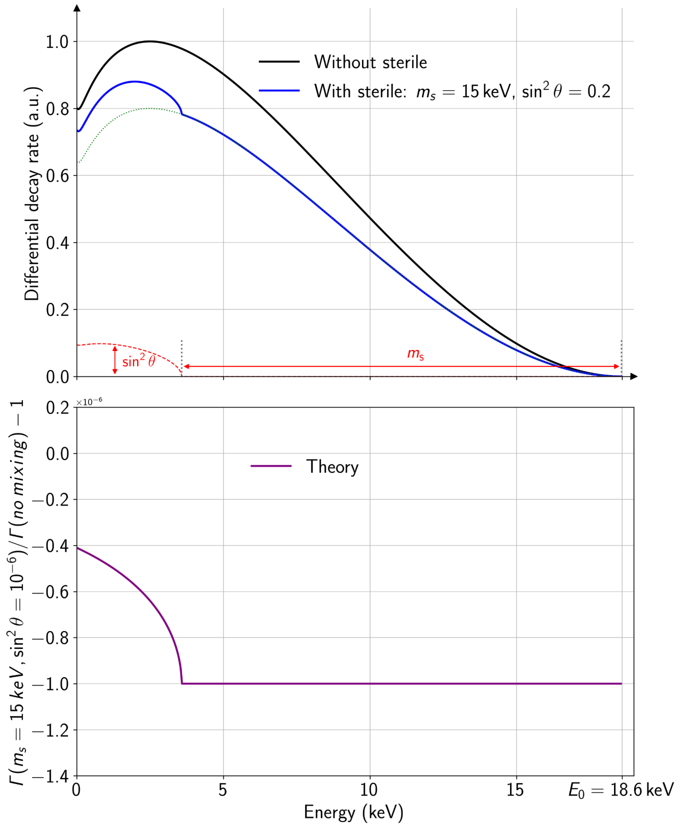


Caveats

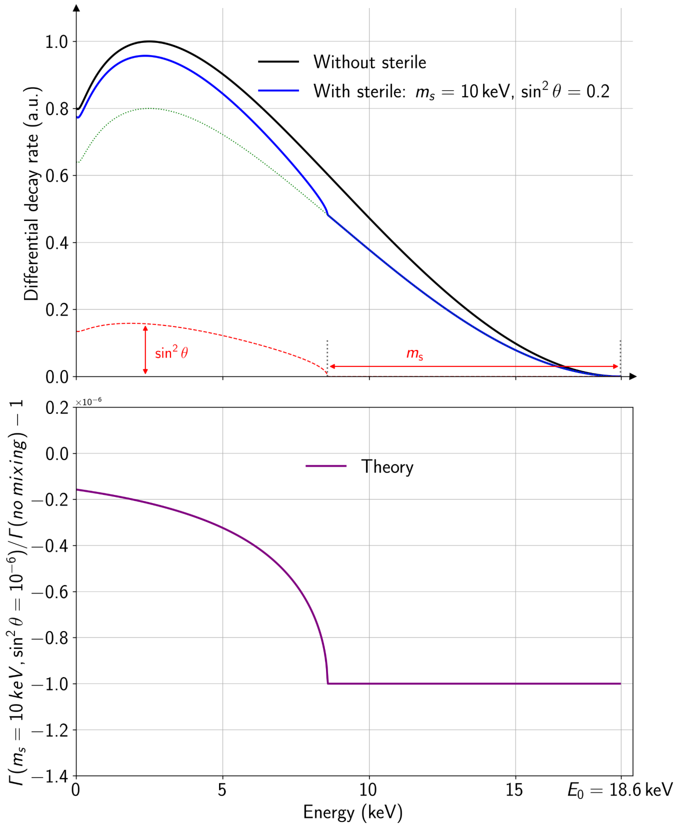
1. Not completely model independent (training data has to come from somewhere)
2. Not 100% transparent
(only likelihood-ratio estimator)

Training Data

$m_s = 5\ \mathrm{keV}$
$\sin^2\theta = 10^{-6}$



$m_s = 10\ \mathrm{keV}$
$\sin^2\theta = 10^{-6}$

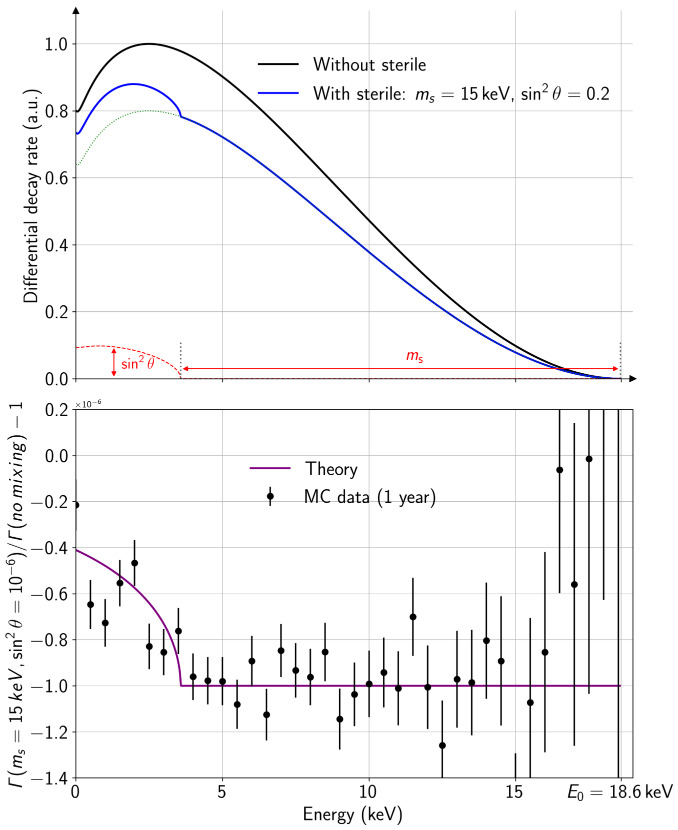


Training Data should include:

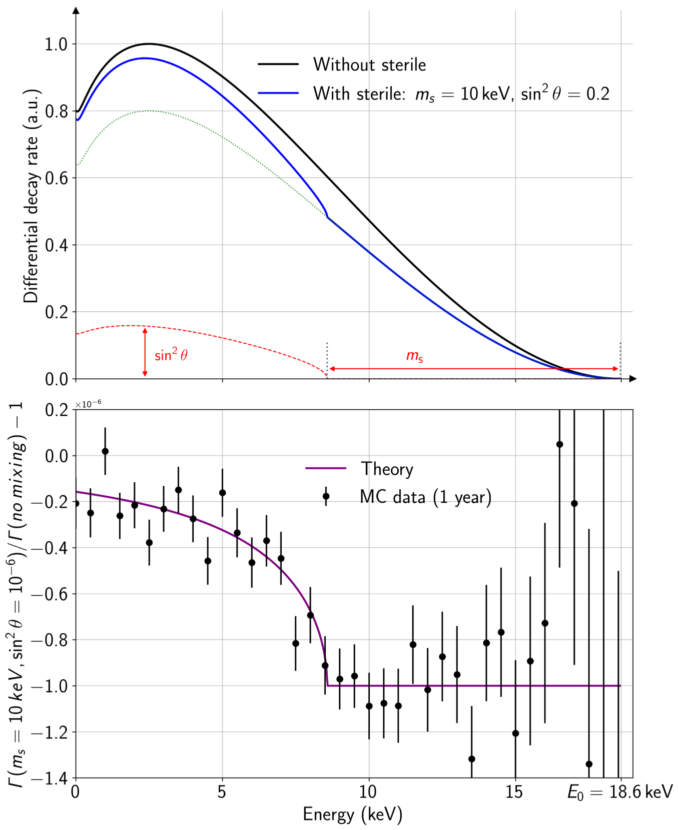
- Different sterile signatures

Training Data

$m_s = 5$ keV
$\sin^2\theta = 10^{-6}$



$m_s = 10$ keV
$\sin^2\theta = 10^{-6}$

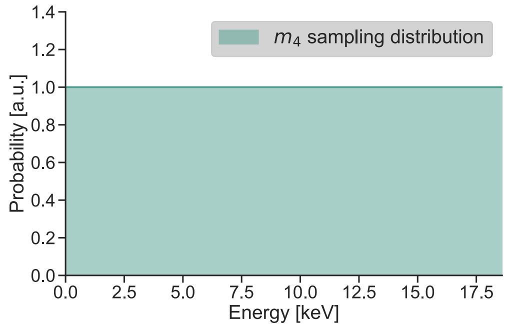


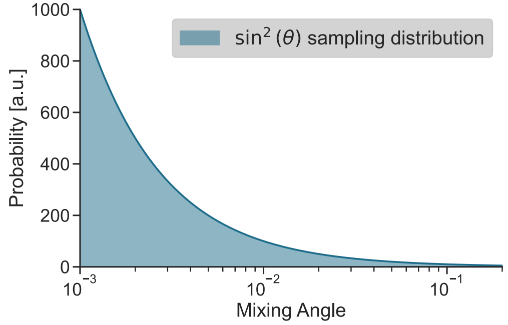
Training Data should include:

- Different sterile signatures
- Statistical fluctuations

Training Data

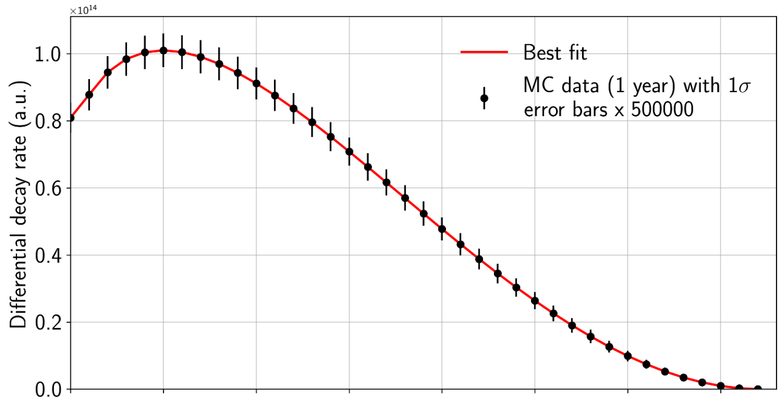
Theoretical Prediction



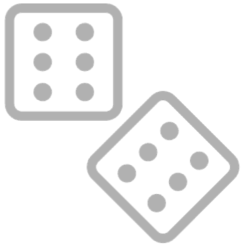


Sample $\mathrm{m}_4, \sin^2(\theta)$

50% probability to set $\sin^2(\theta) = 0$



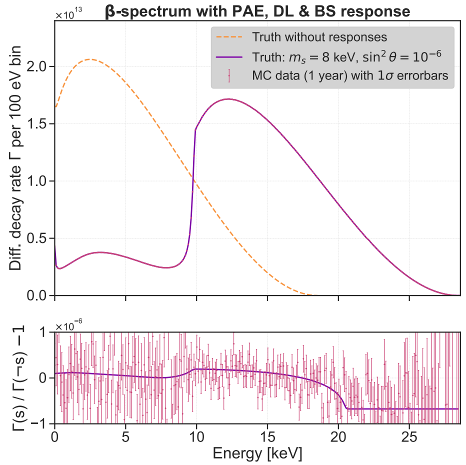
Statistically fluctuate sample



Repeat N times

Training Data

e.g.



Label 0 for $\sin^2\theta = 0$

Add labels for each sample

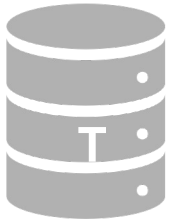
Label 1 for $\sin^2\theta > 0$

Optionally add syst. effects

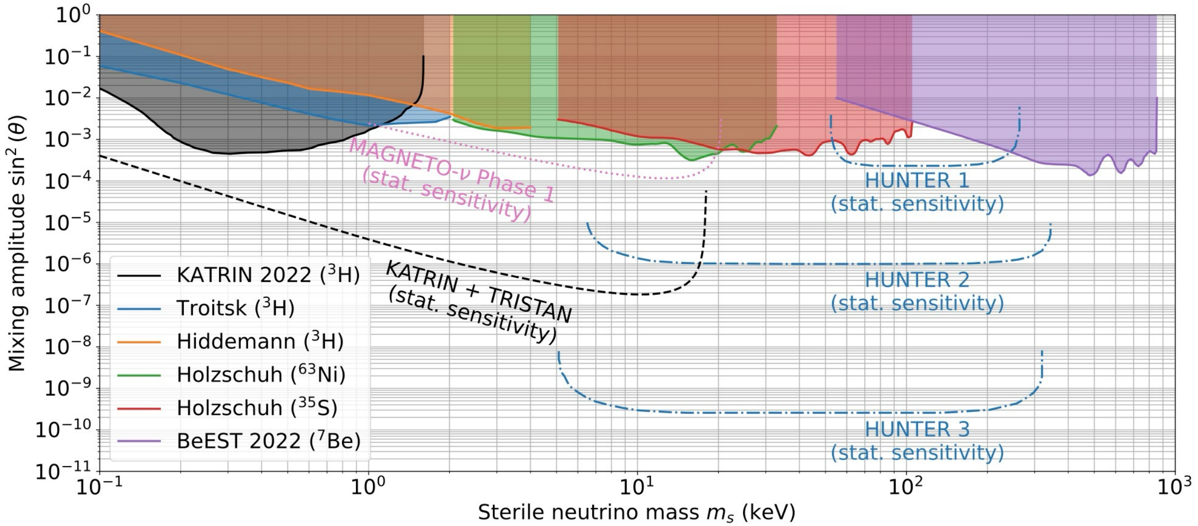
Done via convolution with Response Matrices

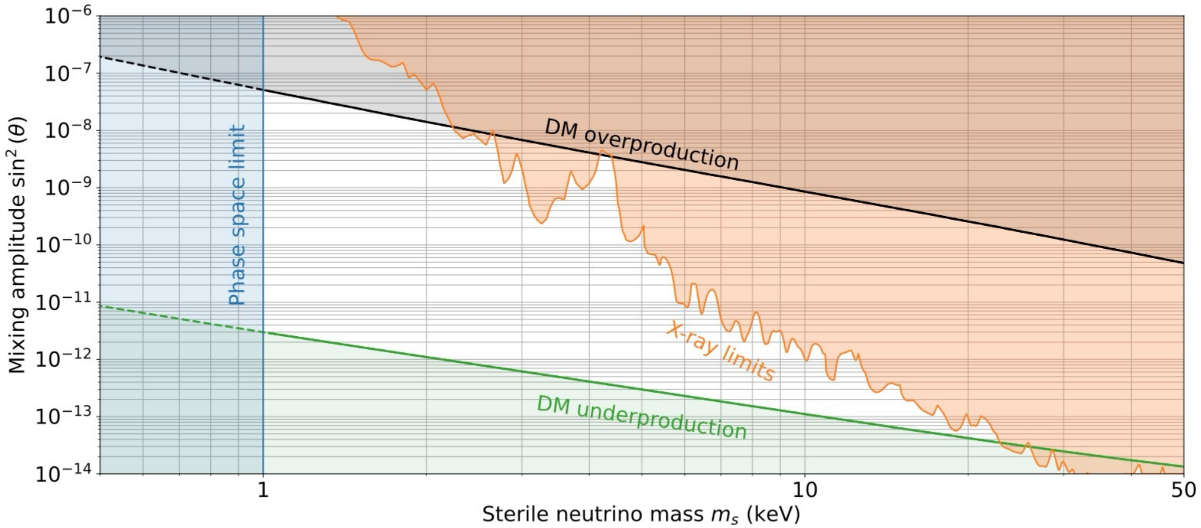
Feature standardization

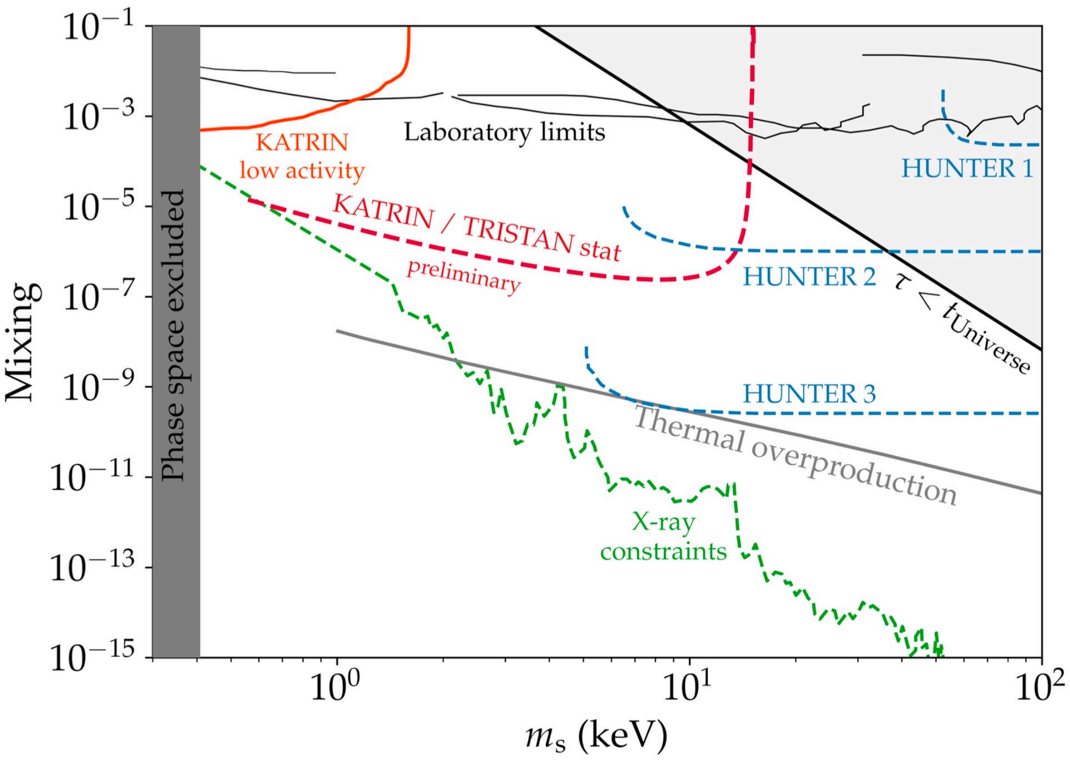
Training Dataset



Backup: Sterile Neutrino Parameter Space

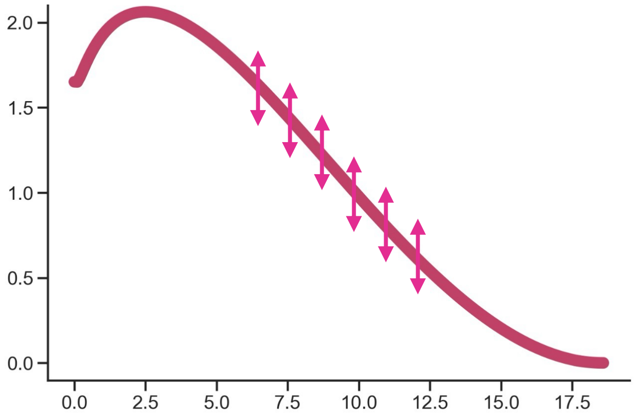






Backup: Sensitivity Studies – Contour Uncertainty

Statistical fluctuations in spectra



Statistical fluctuations in histograms