Friedrich-Alexander-Universität
Erlangen-Nürnberg

Using End-to-End Optimization to Improve the Measurement of Neutrinos from the Galactic Plane with IceCube

Oliver Janik for the IceCube Collaboration

28.10.25

Friedrich-Alexander-Universität
Erlangen-Nürnberg

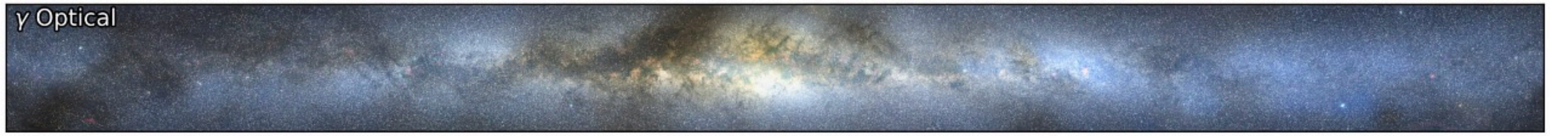
ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

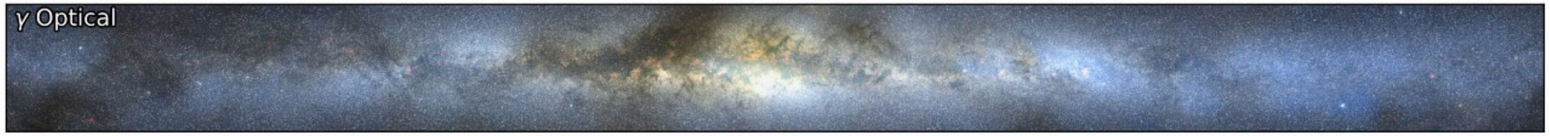
Using **End-to-End Optimization** to Improve the Measurement of Neutrinos from the Galactic Plane with IceCube

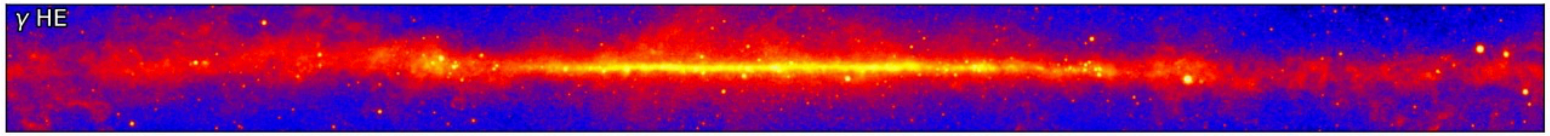
Oliver Janik for the IceCube Collaboration
28.10.25

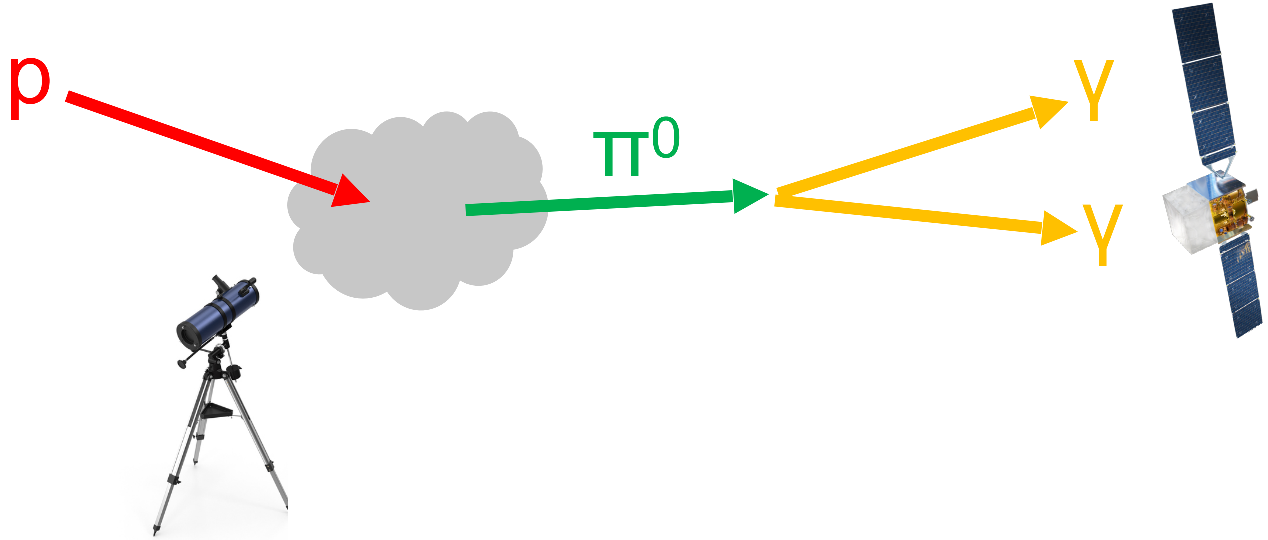
ICECUBE NEUTRINO OBSERVATORY

Measuring the Galactic Plane Neutrino Flux with IceCube

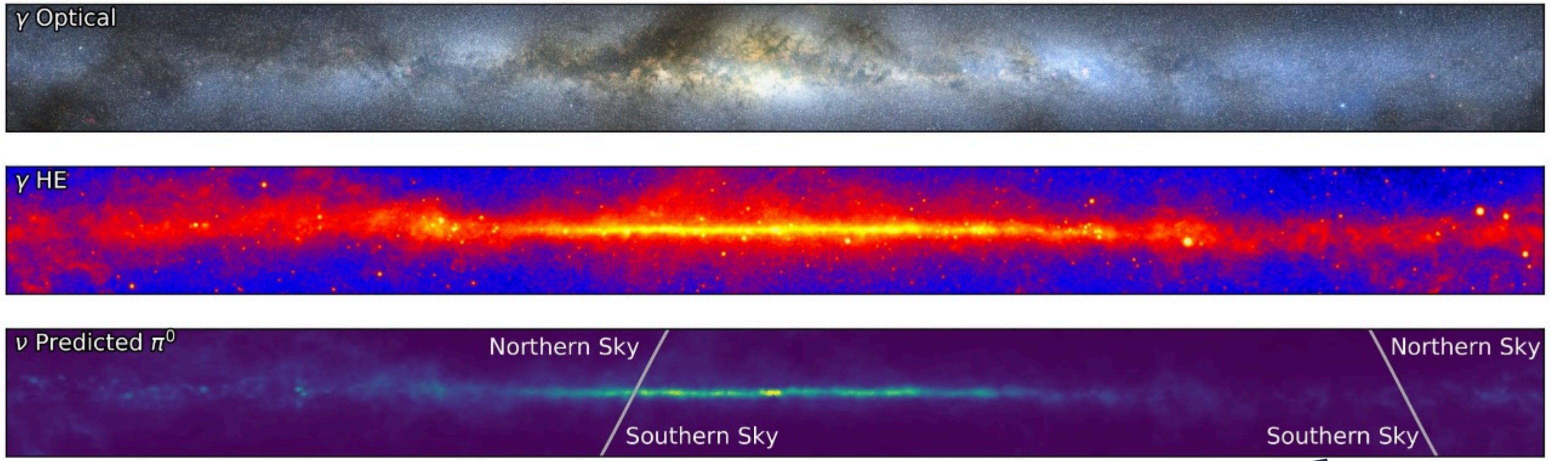



γ Optical

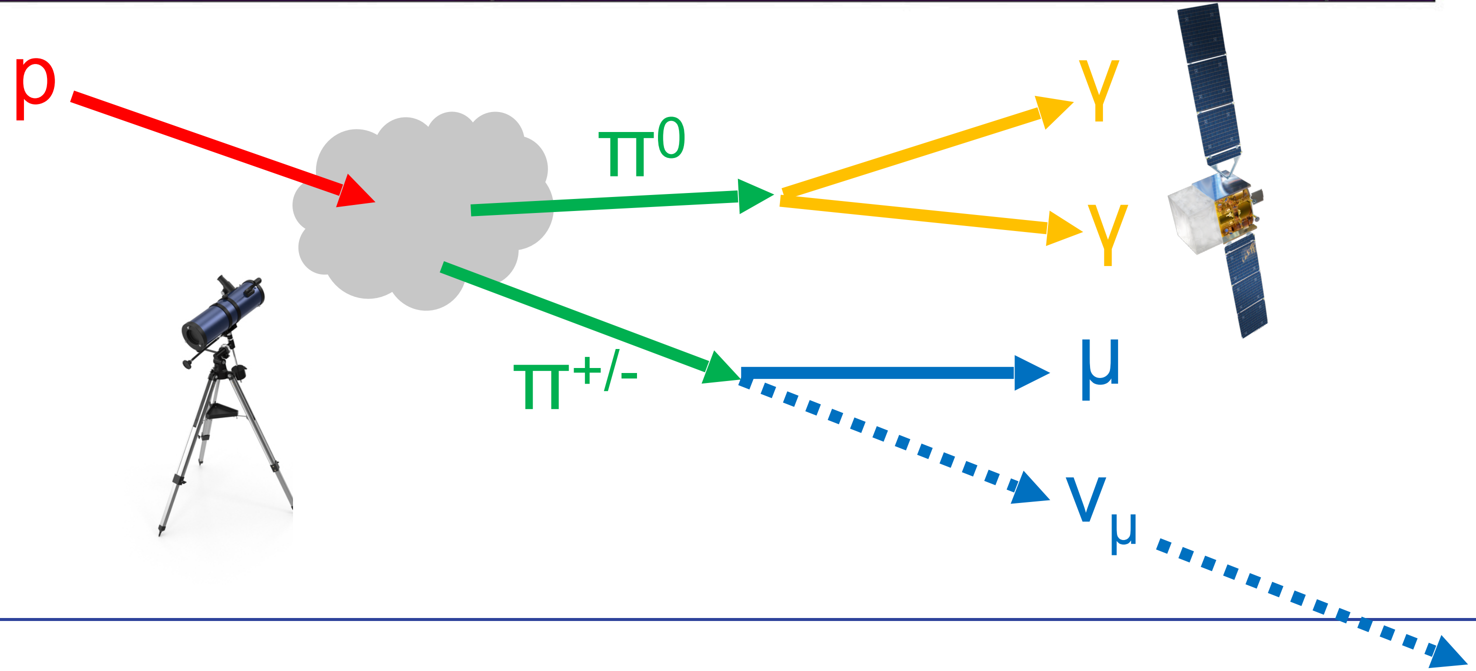

γ HE


p
π0
γ
γ

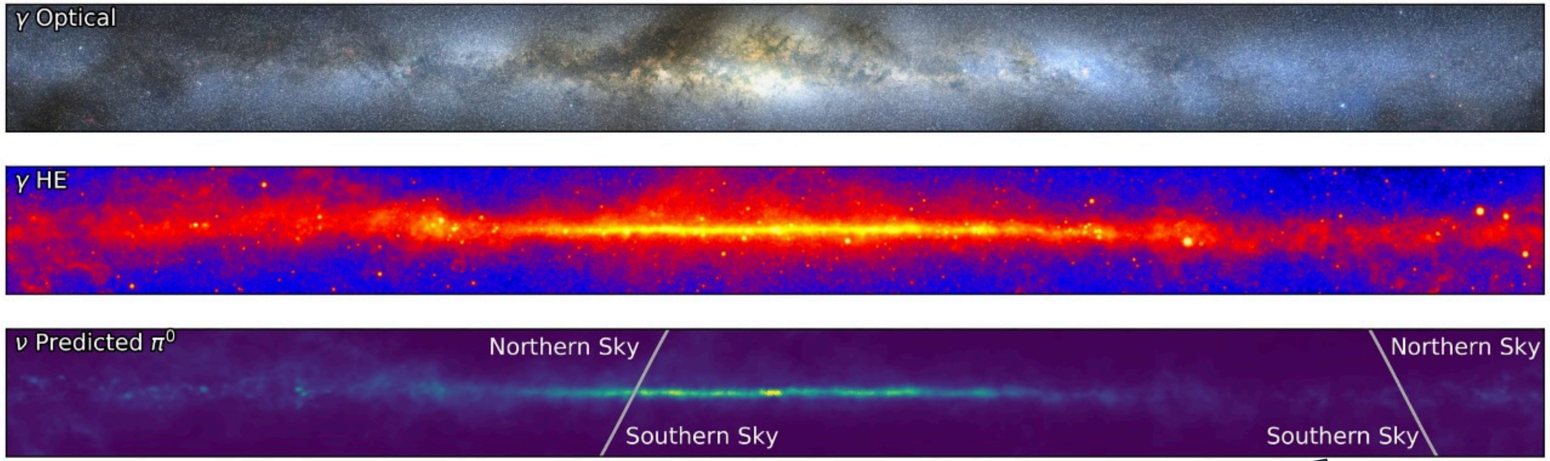
Measuring the Galactic Plane Neutrino Flux with IceCube



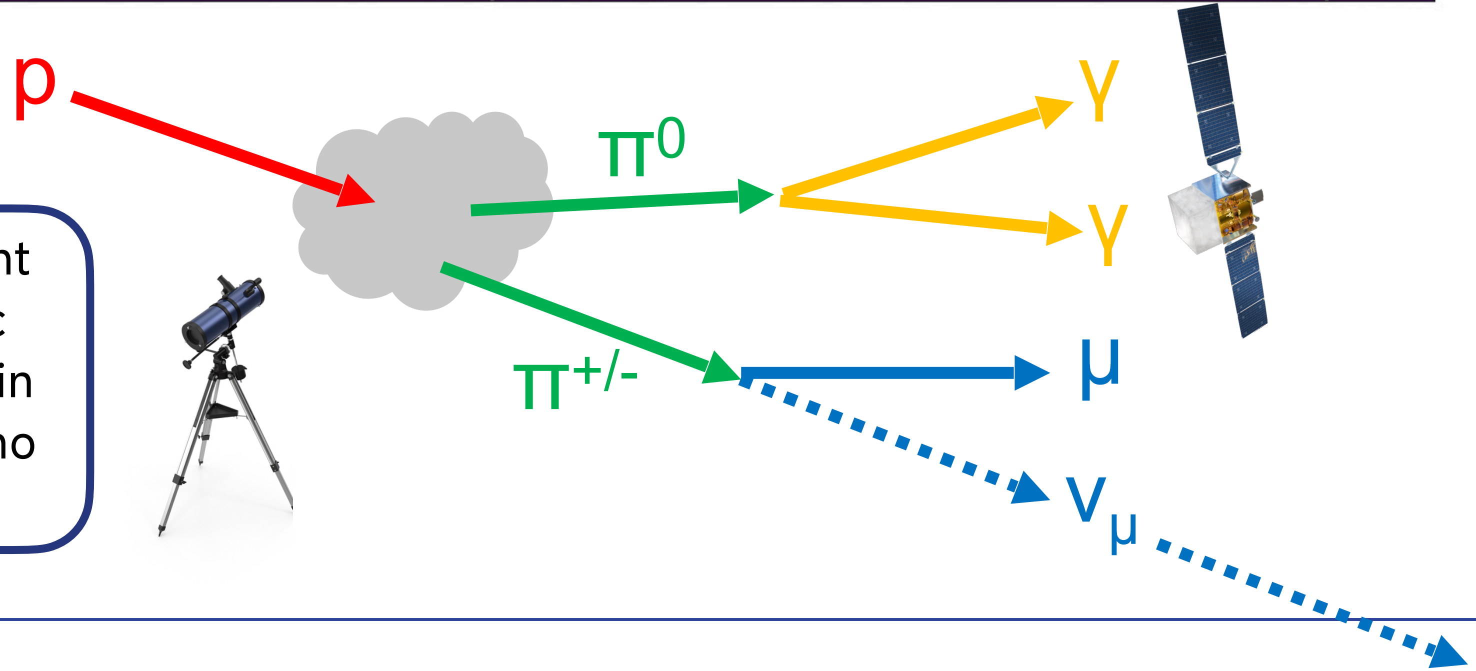
IceCube Collaboration*† Observation of high-energy neutrinos from the Galactic plane. *Science* **380**, 1338-1343 (2023). DOI: 10.1126/science.adc9818



Measuring the Galactic Plane Neutrino Flux with IceCube



IceCube Collaboration*† Observation of high-energy neutrinos from the Galactic plane. *Science* **380**, 1338-1343 (2023). DOI: 10.1126/science.adc9818



A precise measurement of the diffuse galactic neutrino flux can help in finding galactic neutrino point sources

Measuring the Galactic Plane Neutrino Flux with IceCube

1km³ of instrumented South Pole ice

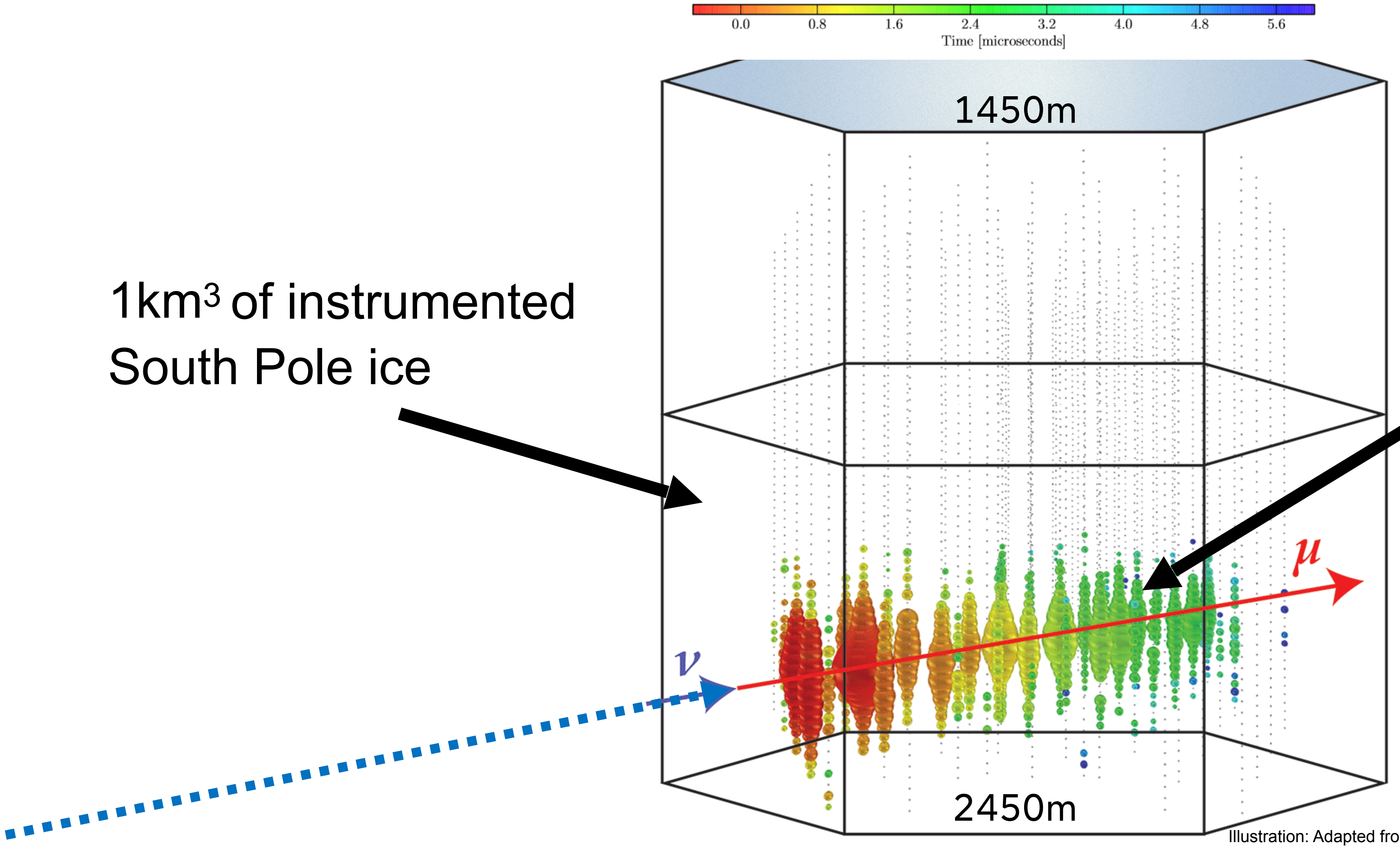
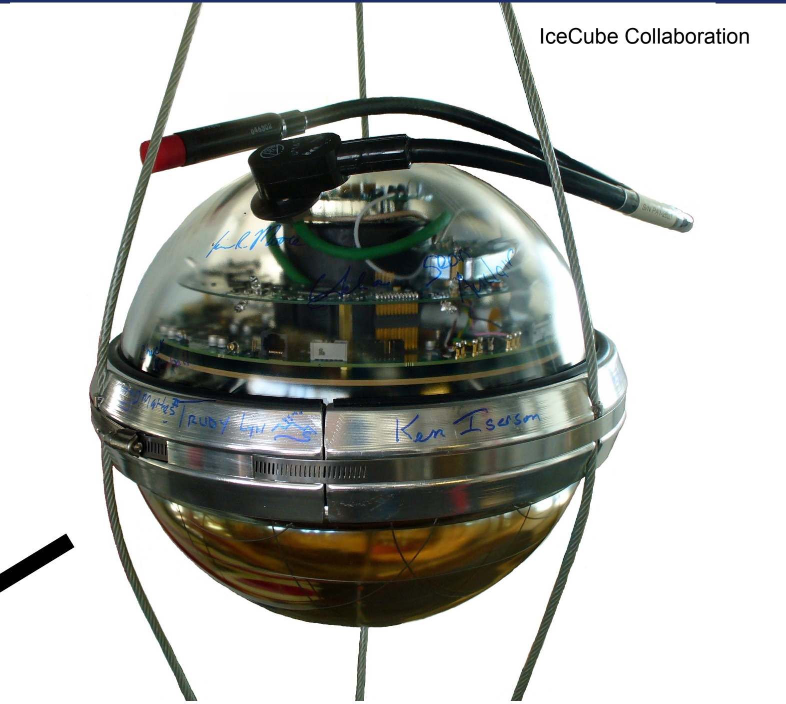


Illustration: Adapted from APS/Joan Tycko; Neutrino event: IceCube



IceCube Collaboration

5160 Light Sensors

Measuring the Galactic Plane Neutrino Flux with IceCube

IceCube Collaboration

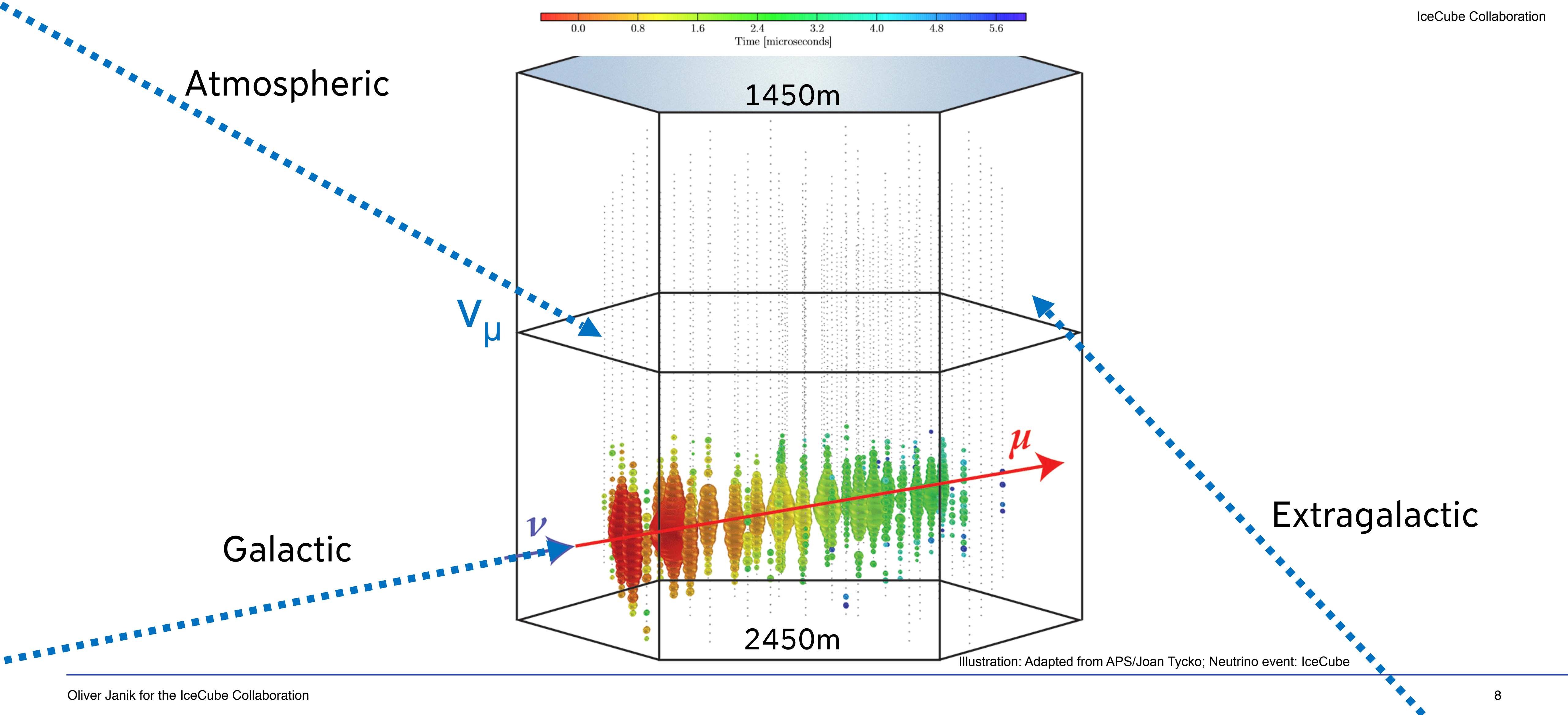
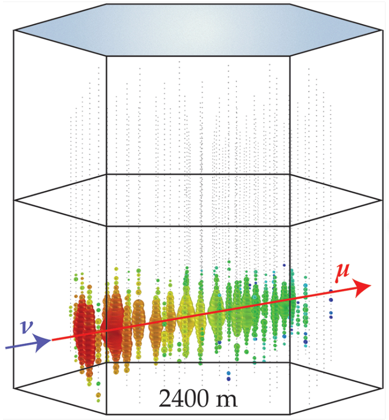


Illustration: Adapted from APS/Joan Tycko; Neutrino event: IceCube

Forward Folding Likelihood Fits

Monte Carlo Simulation



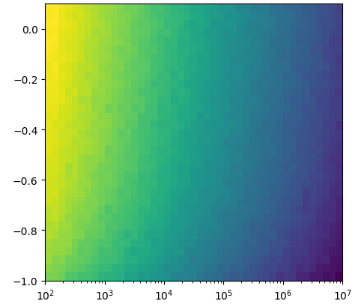
Reconstruction

N-Dimensional Input Variables

- -Energy
- -Zenith
- -Right Ascension
- - ...

MC events weighted according to flux model

N-Dimensional Histogram

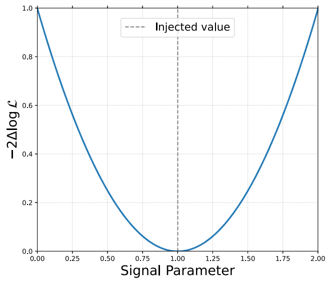


Data

Likelihood

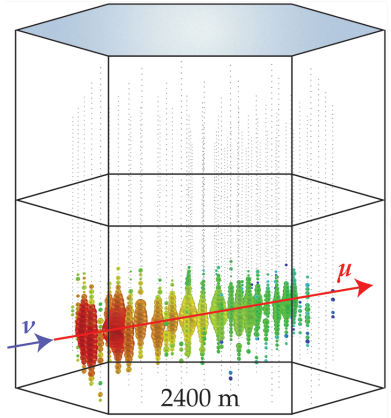
$$\mathcal{L} = \frac{\lambda(\boldsymbol{\theta})^k e^{-\lambda(\boldsymbol{\theta})}}{k!}$$

Best fit & Sensitivities



Forward Folding Likelihood Fits

Monte Carlo Simulation



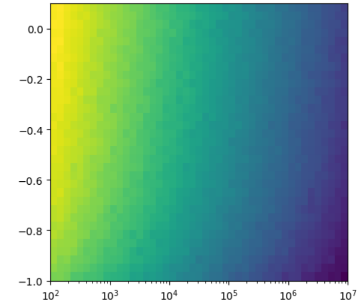
Reconstruction

N-Dimensional Input Variables

- -Energy
- -Zenith
- -Right Ascension
- - ...

MC events weighted according to flux model

N-Dimensional Histogram

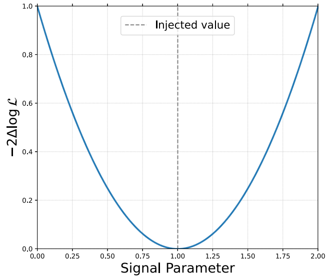


Data

Likelihood

$$\mathcal{L} = \frac{\lambda(\boldsymbol{\theta})^k e^{-\lambda(\boldsymbol{\theta})}}{k!}$$

Best fit & Sensitivities

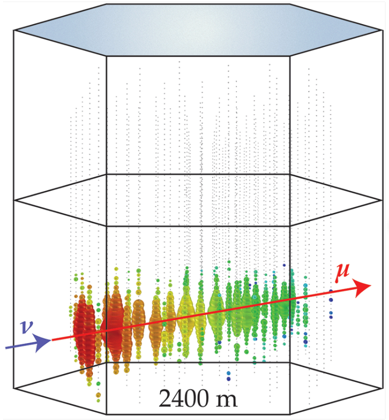


Higher dimensionality of histogram leads to exponentially more bins

Will quickly run out of MC statistics!

Forward Folding Likelihood Fits

Monte Carlo Simulation

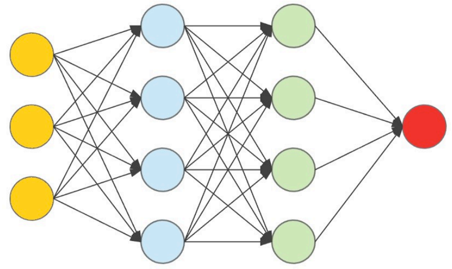


Reconstruction

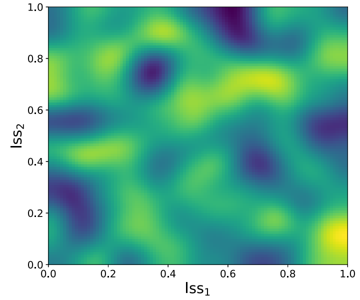
N-Dimensional Input Variables

- -Energy
- -Zenith
- -Right Ascension
- - ...

Neural Network



K-Dimensional Histogram

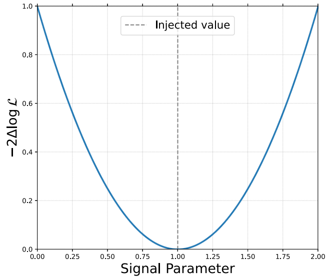


Data

Likelihood

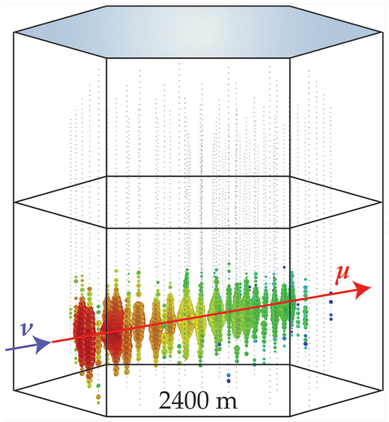
$$\mathcal{L} = \frac{\lambda(\boldsymbol{\theta})^k e^{-\lambda(\boldsymbol{\theta})}}{k!}$$

Best fit & Sensitivities



Forward Folding Likelihood Fits

Monte Carlo Simulation

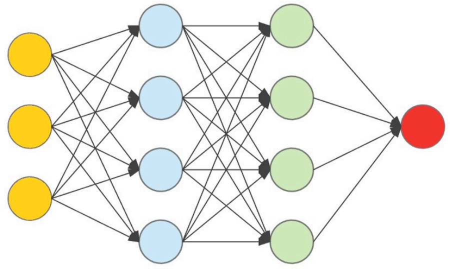


Reconstruction

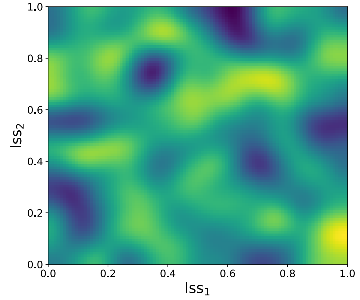
N-Dimensional Input Variables

- -Energy
- -Zenith
- -Right Ascension
- - ...

Neural Network



K-Dimensional Histogram

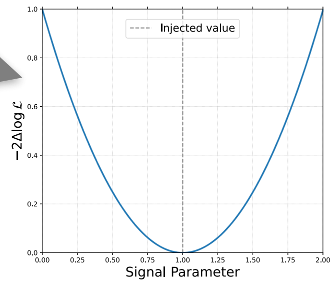


Data

Likelihood

$$\mathcal{L} = \frac{\lambda(\boldsymbol{\theta})^k e^{-\lambda(\boldsymbol{\theta})}}{k!}$$

Best fit & Sensitivities

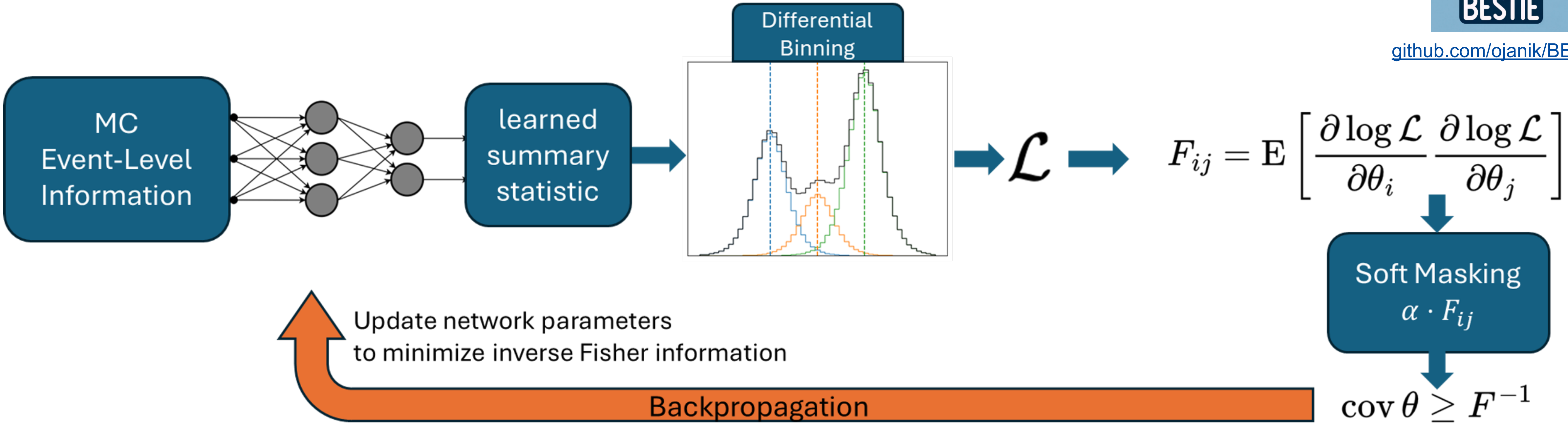


Increase sensitivity on the measurement of the normalization of the diffuse galactic neutrino flux

BESTIE Workflow

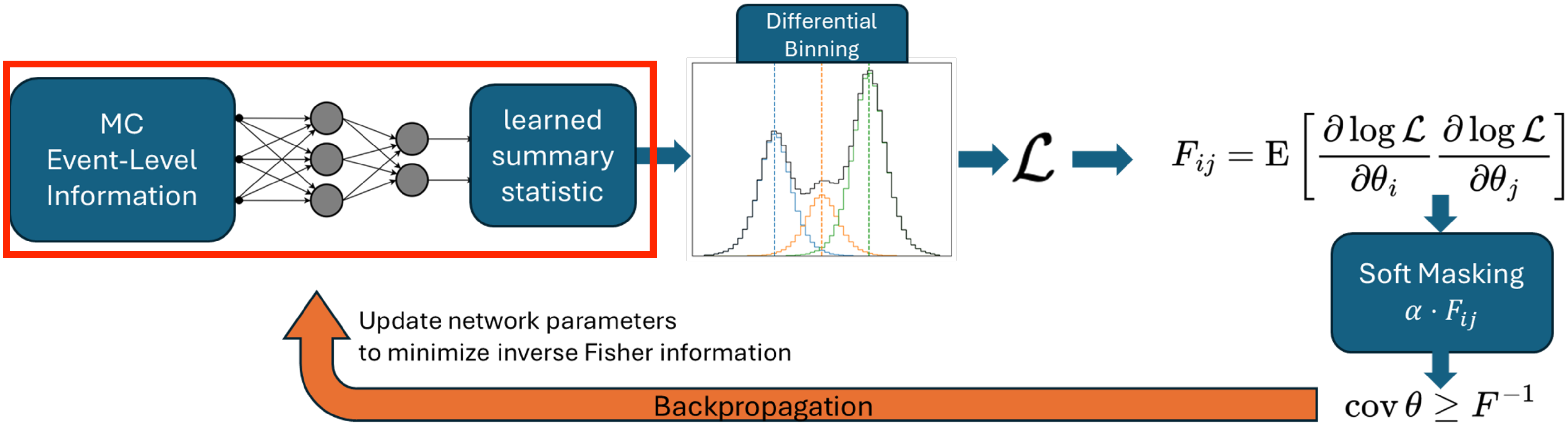
Binned End-to-end optimized Summary sTatistics for the Icecube Experiment

github.com/ojanik/BESTIE



BESTIE Workflow

Binned End-to-end optimized Summary sTatistics for the Icecube Experiment



Data Sampling and Neural Network

Total MC dataset with O(10^7) events

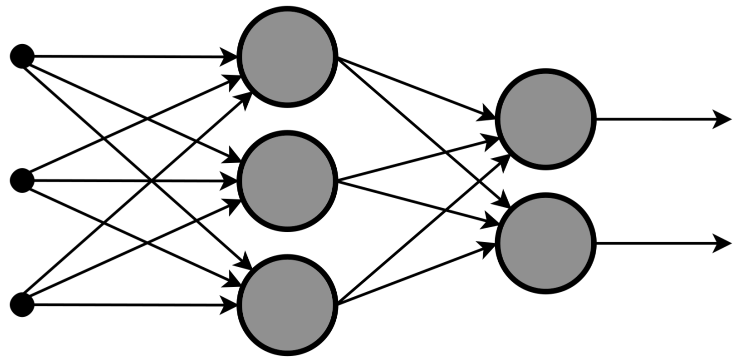
- Reconstructed Energy
- Reconstructed Zenith
- Reconstructed Right Ascension

Used in the standard analysis of the galactic plane

- Angular Uncertainty

Additional input

Sample ~100k events



Dense Neural Network

ξ **learned summary statistic**

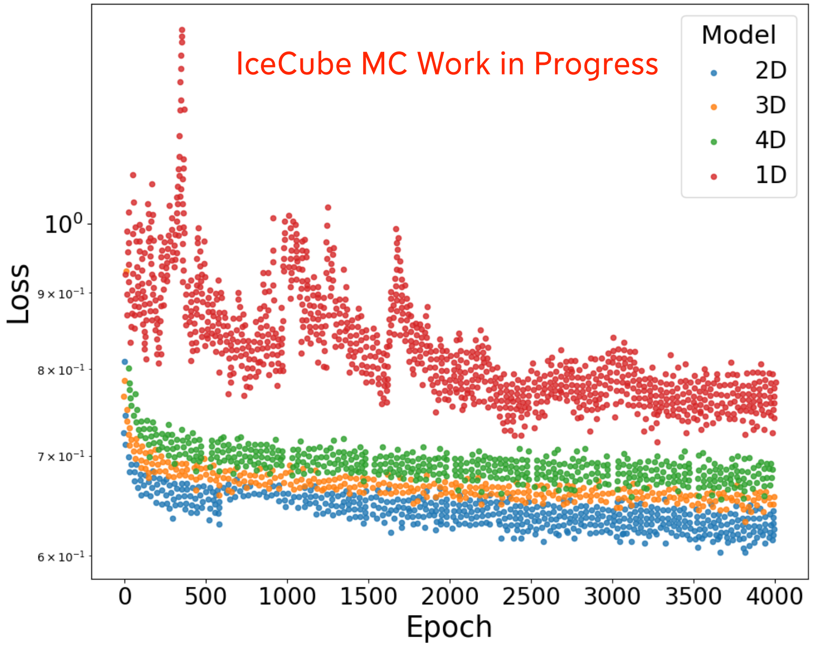
Data Sampling and Neural Network

Total MC dataset with O(10^7) events

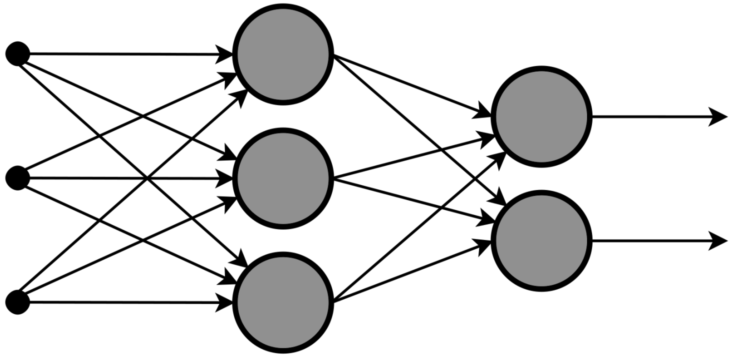
Reconstructed Energy
Reconstructed Zenith
Reconstructed Right Ascension
Angular Uncertainty

Used in the standard analysis of the galactic plane

Additional input



Sample ~100k events

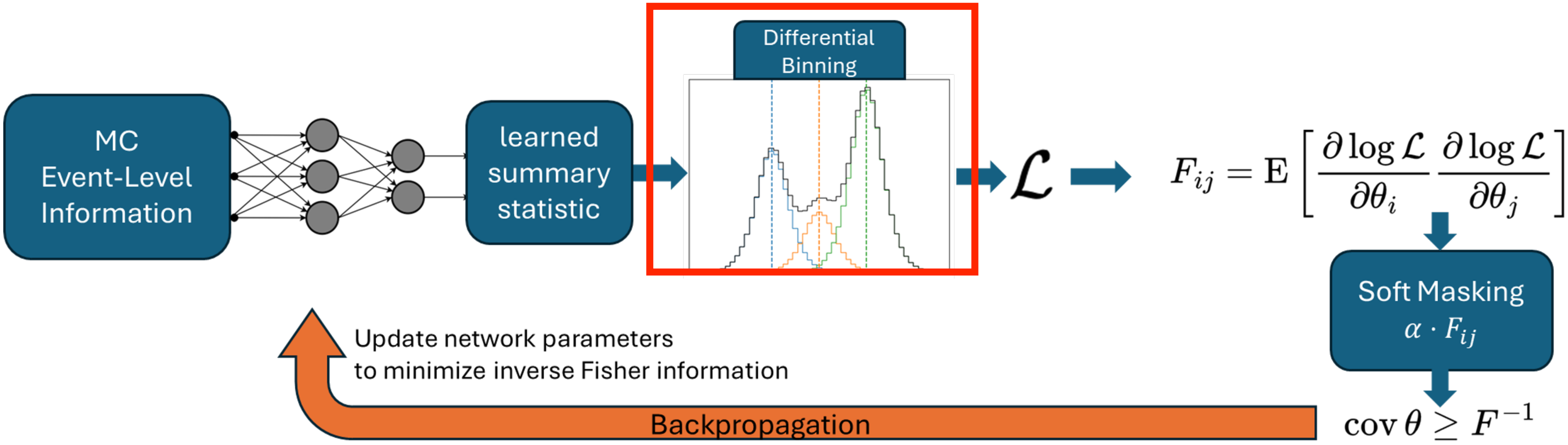


Dense Neural Network

ξ learned summary statistic

BESTIE Workflow

Binned End-to-end optimized Summary sTatistics for the Icecube Experiment



Differential Binning

Signal and background parameters

Bin count

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

Event weights

Indicator function

Differential Binning

Signal and background parameters

Bin count

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

Event weights

Indicator function

Standard indicator function:

$$I_k(\xi_n) = \begin{cases} 1, if\ b_k \leq \xi_n < b_{k+1} \\ 0, else \end{cases}$$

Differential Binning

Signal and background parameters

Bin count

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

Event weights

Indicator function

Standard indicator function:

$$I_k(\xi_n) = \begin{cases} 1, if\ b_k \leq \xi_n < b_{k+1} \\ 0, else \end{cases}$$

Differential indicator function:

$$I_k(\xi_n) = \tanh\left(\frac{\xi_n - b_k}{s}\right)\tanh\left(\frac{b_{k+1} - \xi_n}{s}\right) + 1$$

Differential Binning

Signal and background parameters

Bin count

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

Event weights

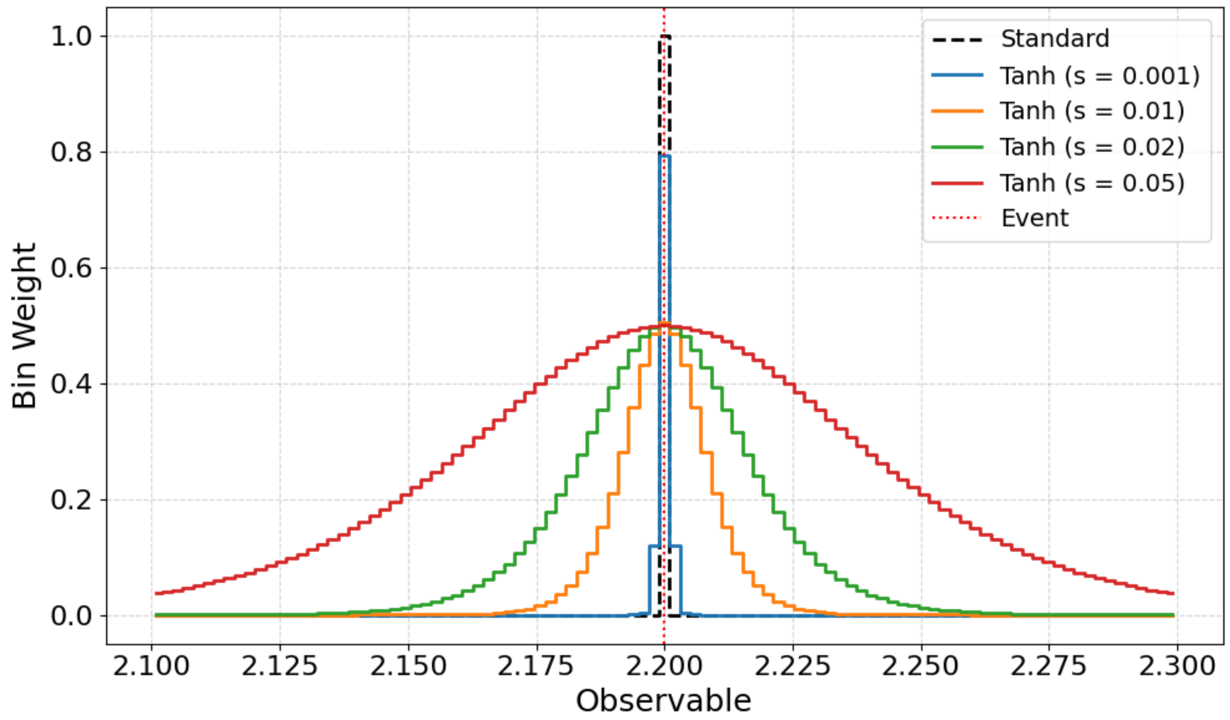
Indicator function

Standard indicator function:

$$I_k(\xi_n) = \begin{cases} 1, if\ b_k \leq \xi_n < b_{k+1} \\ 0, else \end{cases}$$

Differential indicator function:

$$I_k(\xi_n) = \tanh\left(\frac{\xi_n - b_k}{s}\right) \tanh\left(\frac{b_{k+1} - \xi_n}{s}\right) + 1$$



Differential Binning

Signal and background parameters

Bin count

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

Event weights

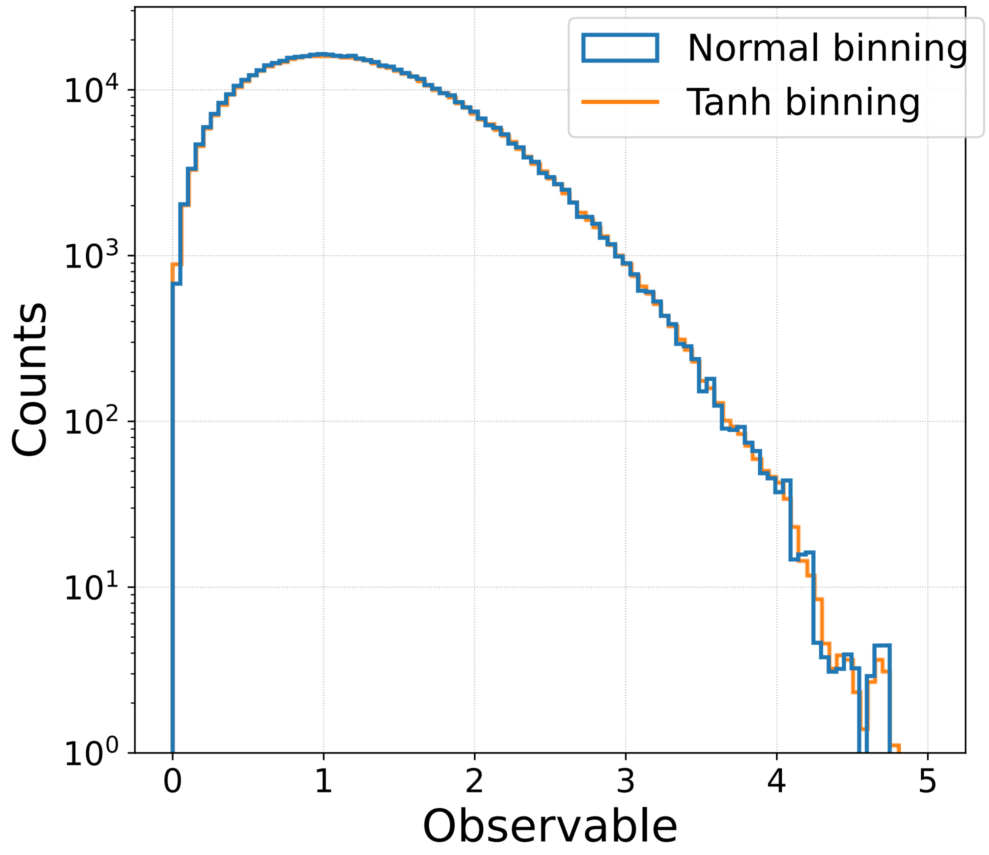
Indicator function

Standard indicator function:

$$I_k(\xi_n) = \begin{cases} 1, if\ b_k \leq \xi_n < b_{k+1} \\ 0, else \end{cases}$$

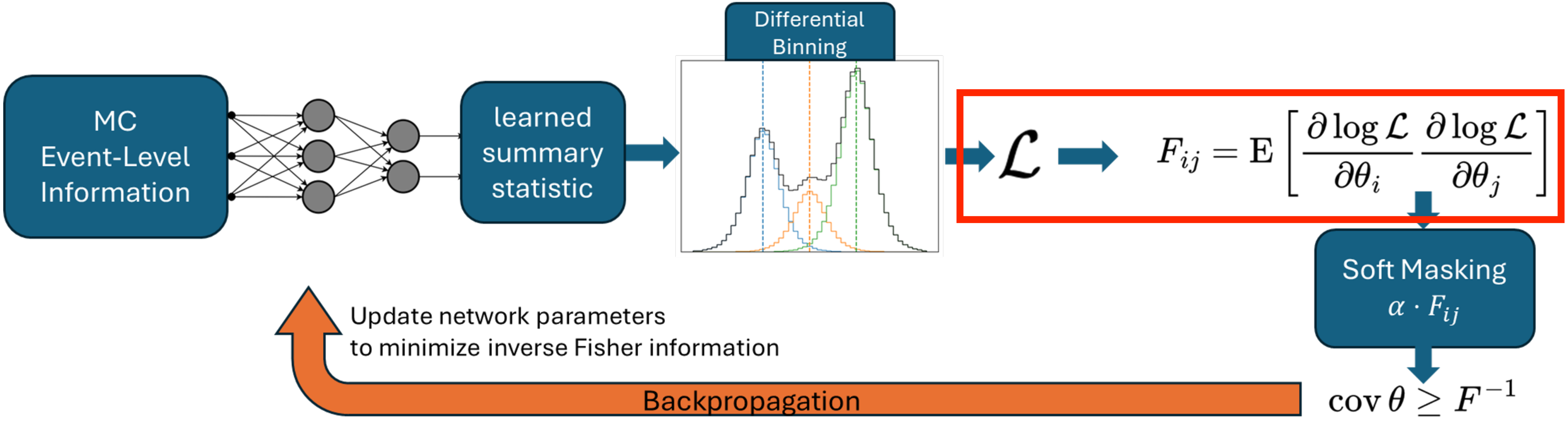
Differential indicator function:

$$I_k(\xi_n) = \tanh\left(\frac{\xi_n - b_k}{s}\right) \tanh\left(\frac{b_{k+1} - \xi_n}{s}\right) + 1$$



BESTIE Workflow

Binned End-to-end optimized Summary sTatistics for the Icecube Experiment



Fisher Information

The per-bin Fisher Information is given by:

$$F_{ij,k} = \mathrm{E}\left[\frac{\partial \log \mathcal{L}_k}{\partial \theta_i} \frac{\partial \log \mathcal{L}_k}{\partial \theta_j}\right]$$

Fisher Information

The per-bin Fisher Information is given by:

$$F_{ij,k} = \mathrm{E}\left[\frac{\partial \log \mathcal{L}_k}{\partial \theta_i} \frac{\partial \log \mathcal{L}_k}{\partial \theta_j}\right]$$

Insert Poisson Likelihood

$$\mathcal{L}(\lambda \mid \mu) = \frac{\lambda^\mu e^{-\lambda}}{\mu!}$$

$$F_{ij,k} = \frac{1}{\lambda} \frac{\partial \lambda_k}{\partial \theta_i} \frac{\partial \lambda_k}{\partial \theta_j}$$

Fisher Information

The per-bin Fisher Information is given by:

$$F_{ij,k} = \mathrm{E}\left[\frac{\partial \log \mathcal{L}_k}{\partial \theta_i}\frac{\partial \log \mathcal{L}_k}{\partial \theta_j}\right]$$

Insert Poisson Likelihood

$$\mathcal{L}(\lambda \mid \mu) = \frac{\lambda^{\mu} e^{-\lambda}}{\mu!}$$

$$F_{ij,k} = \frac{1}{\lambda}\frac{\partial \lambda_k}{\partial \theta_i}\frac{\partial \lambda_k}{\partial \theta_j}$$

Insert Bincount

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

$$F_{ij,k} = \frac{\sum_n \frac{\partial w_n(\theta)}{\partial \theta_i} I_k(\xi_n) \cdot \sum_n \frac{\partial w_n(\theta)}{\partial \theta_j} I_k(\xi_n)}{\sum_n w_n(\theta) I_k(\xi_n)}$$

Fisher Information

The per-bin Fisher Information is given by:

$$F_{ij,k} = \mathrm{E}\left[\frac{\partial \log \mathcal{L}_k}{\partial \theta_i} \frac{\partial \log \mathcal{L}_k}{\partial \theta_j}\right]$$

Insert Poisson Likelihood

$$\mathcal{L}(\lambda \mid \mu) = \frac{\lambda^\mu e^{-\lambda}}{\mu!}$$

$$F_{ij,k} = \frac{1}{\lambda} \frac{\partial \lambda_k}{\partial \theta_i} \frac{\partial \lambda_k}{\partial \theta_j}$$

Insert Bincount

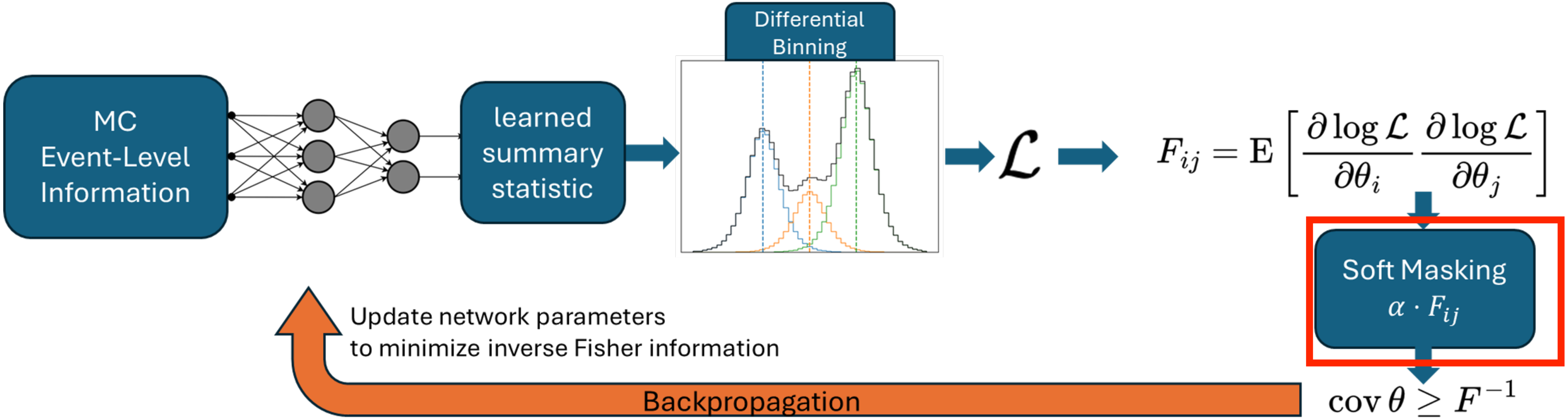
$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

Can precalulate the weights and derivatives to save computation time during training!

$$F_{ij,k} = \frac{\sum_n \boxed{\frac{\partial w_n(\theta)}{\partial \theta_i}} I_k(\xi_n) \cdot \sum_n \boxed{\frac{\partial w_n(\theta)}{\partial \theta_j}} I_k(\xi_n)}{\sum_n \boxed{w_n(\theta)} I_k(\xi_n)}$$

BESTIE Workflow

Binned End-to-end optimized Summary sTatistics for the Icecube Experiment



The Matter of Insufficient MC Statistics

Poisson likelihood does not account for insufficient MC statistics

$$\mathcal{L}(\lambda \mid \mu) = \frac{\lambda^{\mu} e^{-\lambda}}{\mu!}$$

Expectation:

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k\left(\xi_n\right)$$

The Matter of Insufficient MC Statistics

To account for limited MC statistics, use effective likelihood:

$$\mathcal{L}(\lambda \mid \mu) = \frac{\lambda^{\mu} e^{-\lambda}}{\mu!} \longrightarrow \mathcal{L}_{\mathrm{Eff}}(\vec{\theta} \mid \mu) = \left(\frac{\lambda}{\sigma^2}\right)^{\frac{\lambda^2}{\sigma^2}+1} \Gamma\left(\mu + \frac{\lambda^2}{\sigma^2} + 1\right) \left[\mu! \left(1 + \frac{\lambda}{\sigma^2}\right)^{\mu + \frac{\lambda^2}{\sigma^2}+1} \Gamma\left(\frac{\lambda^2}{\sigma^2} + 1\right)\right]$$

Expectation:

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k\left(\xi_n\right)$$

Absolute uncertainty:

$$\sigma_k^2 = \sum_n w_n^2(\boldsymbol{\theta}) I_k\left(\xi_n\right)$$

The Matter of Insufficient MC Statistics

To account for limited MC statistics, use effective likelihood:

$$\mathcal{L}(\lambda \mid \mu) = \frac{\lambda^{\mu} e^{-\lambda}}{\mu!} \longrightarrow \mathcal{L}_{\text{Eff}}(\vec{\theta} \mid \mu) = \left(\frac{\lambda}{\sigma^2}\right)^{\frac{\lambda^2}{\sigma^2}+1} \Gamma\left(\mu + \frac{\lambda^2}{\sigma^2} + 1\right)\left[\mu!\left(1 + \frac{\lambda}{\sigma^2}\right)^{\mu + \frac{\lambda^2}{\sigma^2}+1} \Gamma\left(\frac{\lambda^2}{\sigma^2} + 1\right)\right]$$

Expectation:

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k\left(\xi_n\right)$$

Absolute uncertainty:

$$\sigma_k^2 = \sum_n w_n^2(\boldsymbol{\theta}) I_k\left(\xi_n\right)$$

During training: Want to use simple solution for Fisher information found with the Poisson likelihood. Not given for the effective likelihood.

Soft Masking

Expectation:

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

Absolute uncertainty:

$$\sigma_k^2 = \sum_n w_n^2(\boldsymbol{\theta}) I_k(\xi_n)$$

Relative uncertainty:

$$\sigma_{rel,k} = \frac{\sigma_k}{\lambda_k}$$

Soft Masking

Expectation:
$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

Absolute uncertainty:
$$\sigma_k^2 = \sum_n w_n^2(\boldsymbol{\theta}) I_k(\xi_n)$$

Relative uncertainty:
$$\sigma_{rel,k} = \frac{\sigma_k}{\lambda_k}$$

During training mask the per-bin Fisher Information:

$$F'_{ij,k} = F_{ij,k} \cdot \left(1 + \frac{1}{1 + \exp((t - \sigma_{rel,k})/s)}\right)$$

Threshold, above which rel. uncertainty masking occurs, e.g., 5%

Sharpness, how abrupt the masking is

Soft Masking

Expectation:

$$\lambda_k = \sum_n w_n(\boldsymbol{\theta}) I_k(\xi_n)$$

Absolute uncertainty:

$$\sigma_k^2 = \sum_n w_n^2(\boldsymbol{\theta}) I_k(\xi_n)$$

Relative uncertainty:

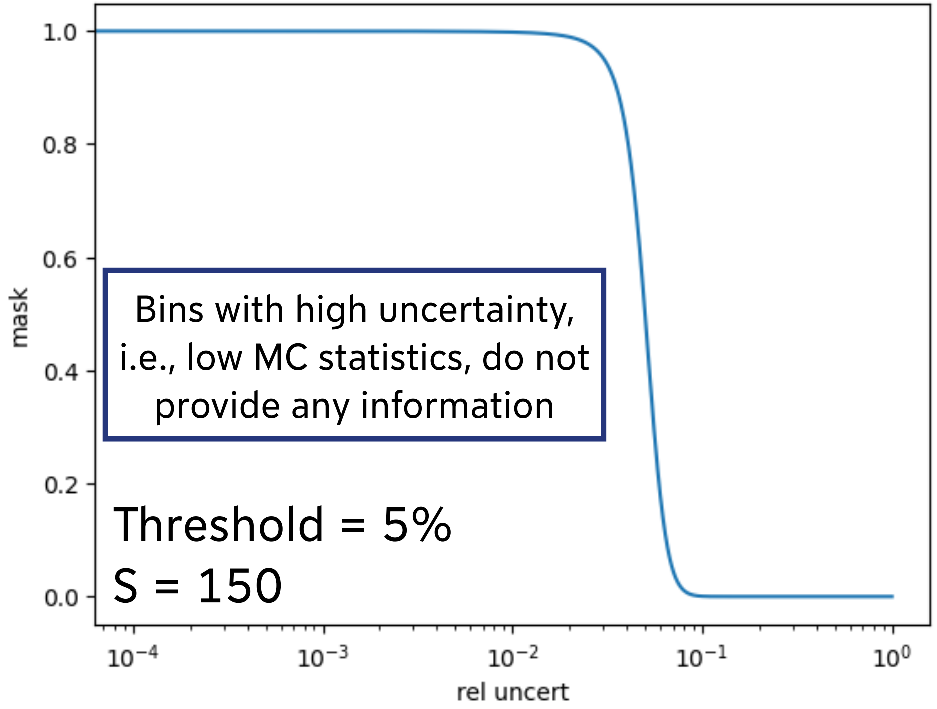
$$\sigma_{rel,k} = \frac{\sigma_k}{\lambda_k}$$

During training mask the per-bin Fisher Information:

$$F'_{ij,k} = F_{ij,k} \cdot \left(1 + \frac{1}{1 + \exp((t - \sigma_{rel,k})/s)}\right)$$

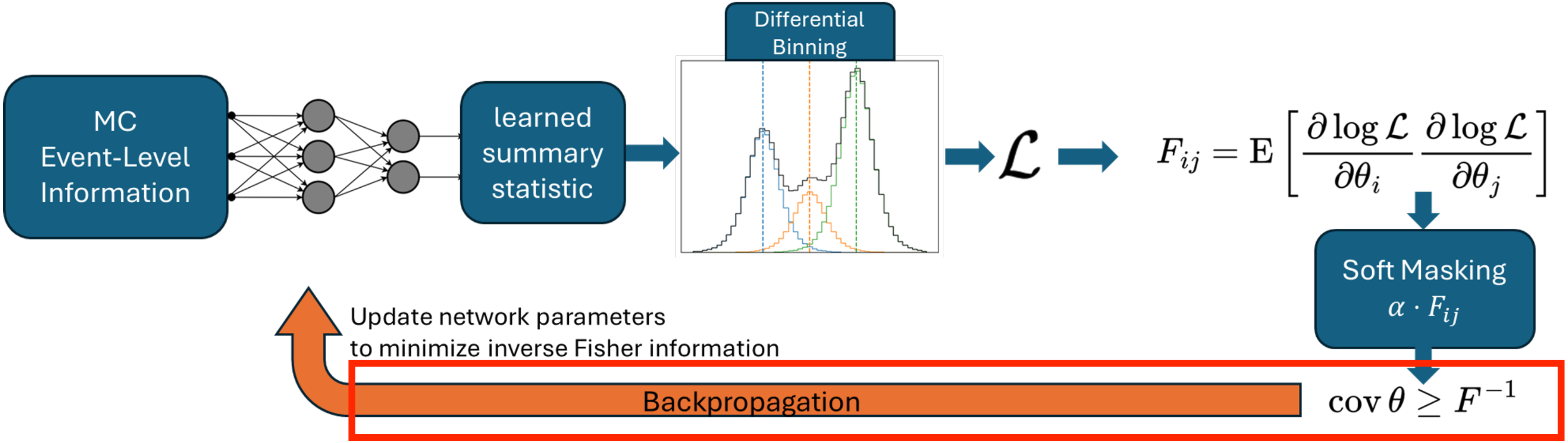
Threshold, above which rel. uncertainty masking occurs, e.g., 5%

Sharpness, how abrupt the masking is



BESTIE Workflow

Binned End-to-end optimized Summary sTatistics for the Icecube Experiment



Loss

Get estimate of the covariance from Cramér-Rao-Bound

$$\operatorname{cov} \boldsymbol{\theta} \geq \left(\sum_{k} F'_{ij,k} \right)^{-1}$$

Get estimate of the covariance from Cramér-Rao-Bound

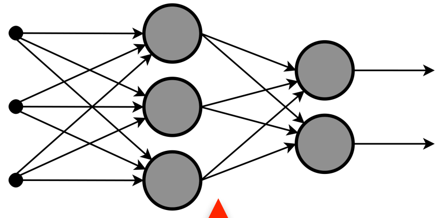
$$\operatorname{cov}\boldsymbol{\theta} \geq \left(\sum_k F'_{ij,k}\right)^{-1}$$

Take relevant part of the
covariance as loss, i.e.,
the estimated uncertainty
of the galactic flux normalization

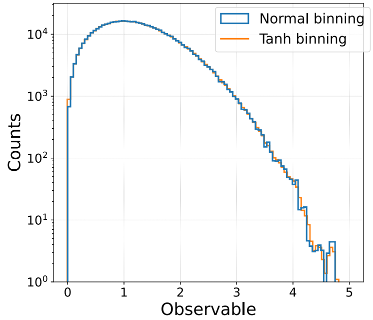
$$\mathrm{Var}(\theta_{gal})$$

Loss

Get estimate of the covariance from Cramér-Rao-Bound



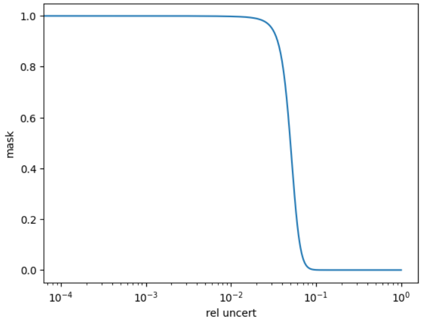
$$\operatorname{cov}\boldsymbol{\theta} \geq \left(\sum_k F'_{ij,k}\right)^{-1}$$



Take relevant part of the covariance as loss, i.e., the estimated uncertainty of the galactic flux normalization

$$\mathrm{Var}(\theta_{gal})$$

Backpropagation

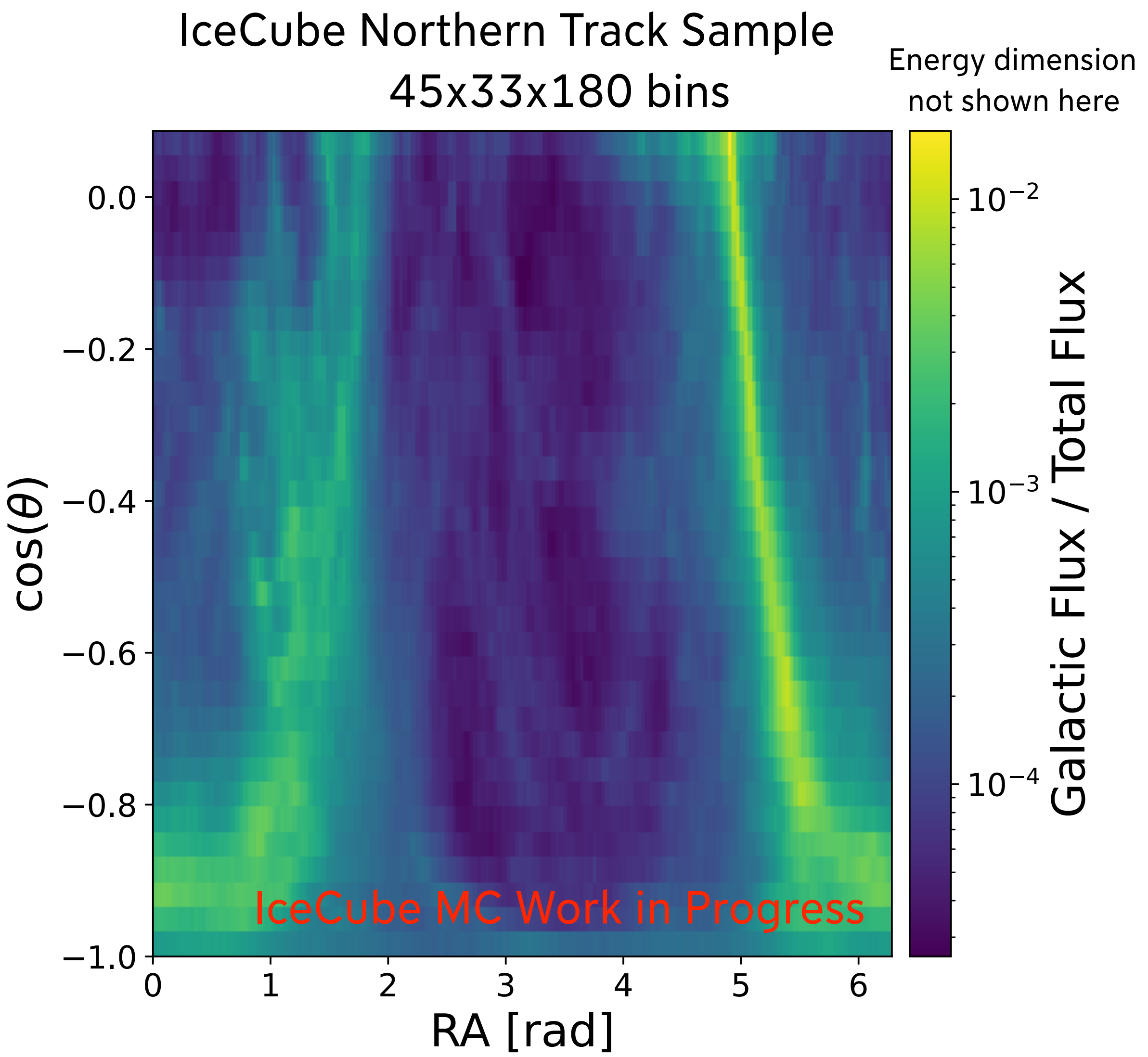


$$F_{ij,k} = \frac{\sum_n \frac{\partial w_n(\theta)}{\partial \theta_i} I_k(\xi_n) \cdot \sum_n \frac{\partial w_n(\theta)}{\partial \theta_j} I_k(\xi_n)}{\sum_n w_n(\theta) I_k(\xi_n)}$$

Results

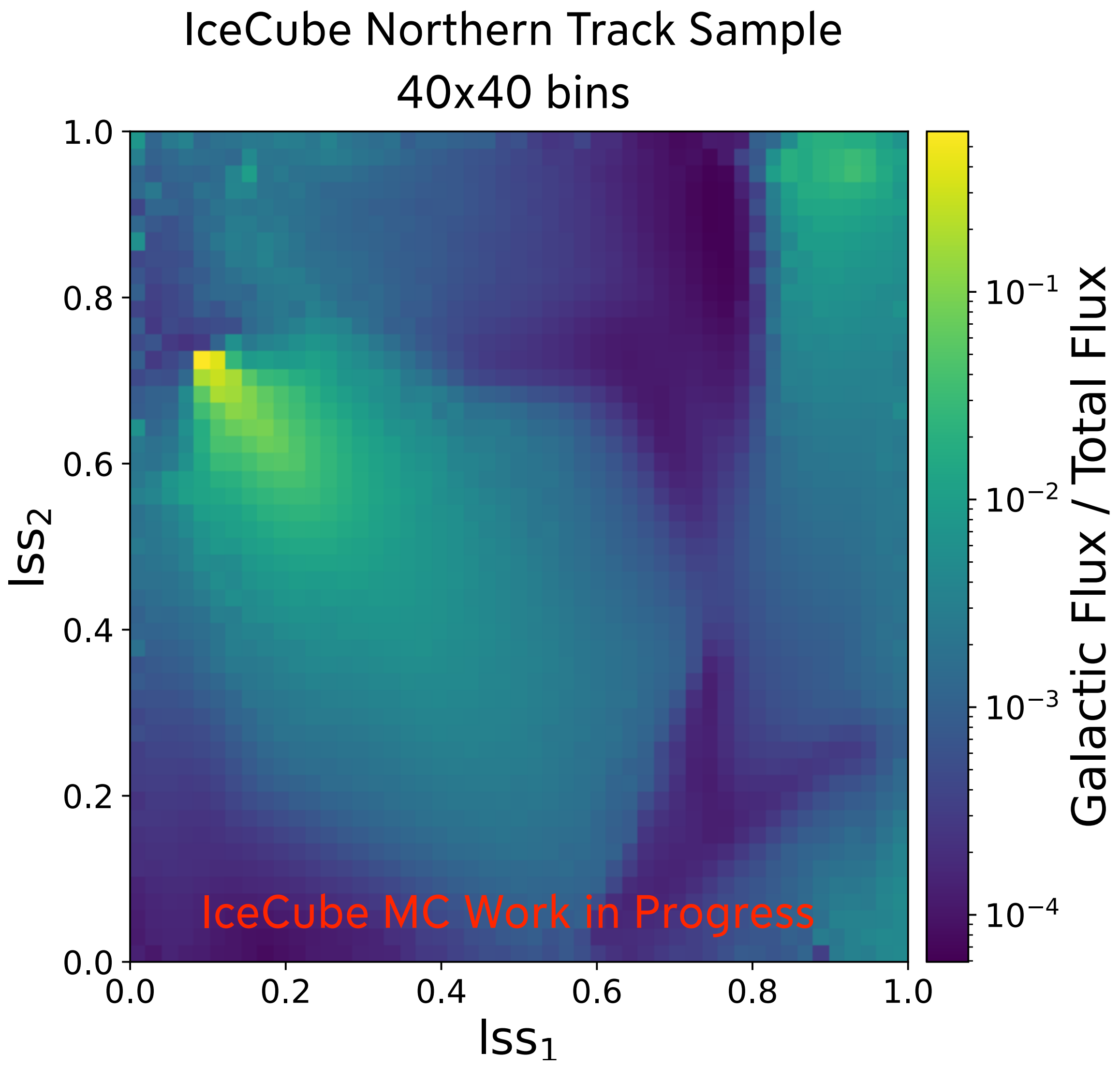
Results Histograms

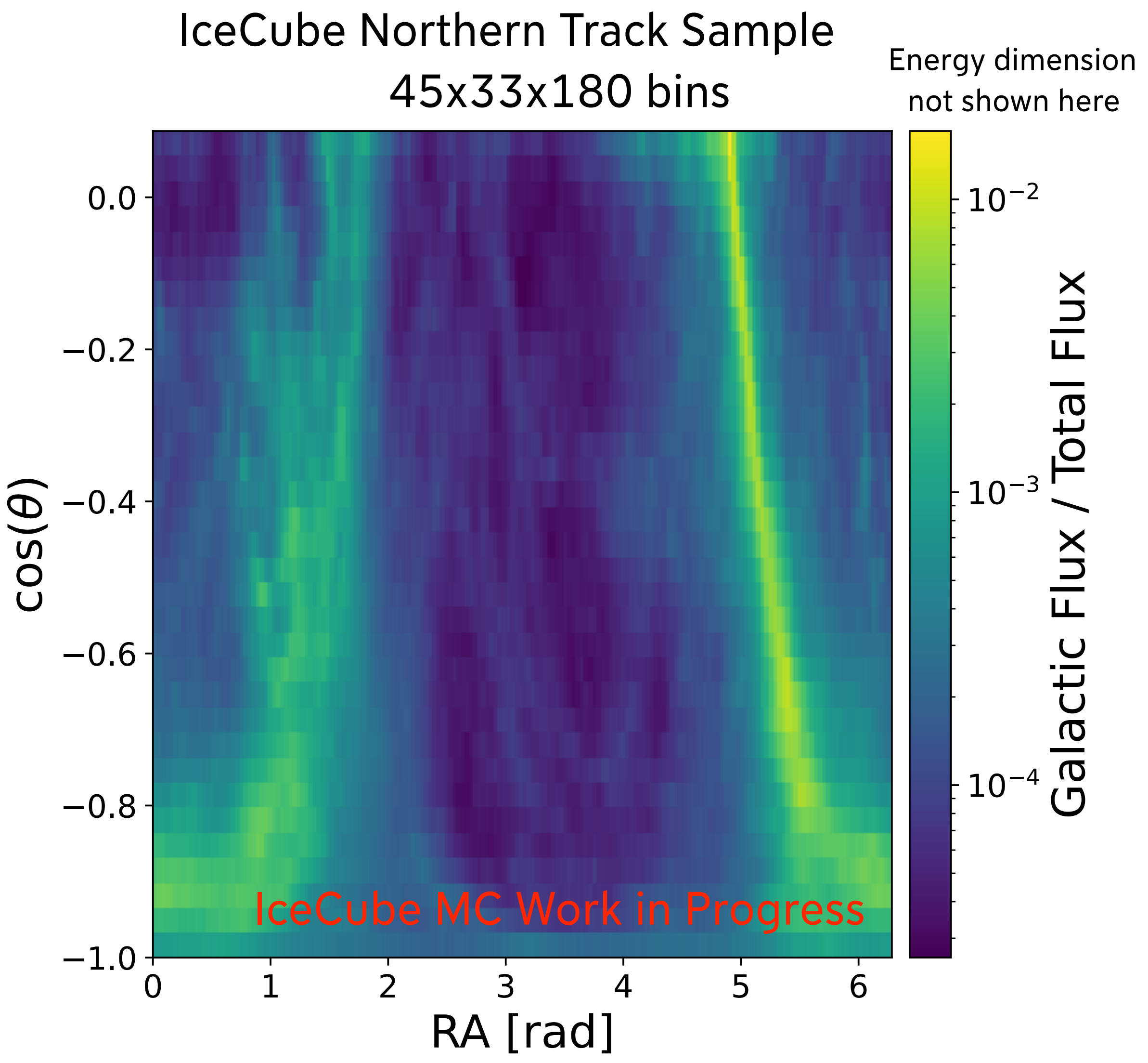
Using IceCube Northern Tracks



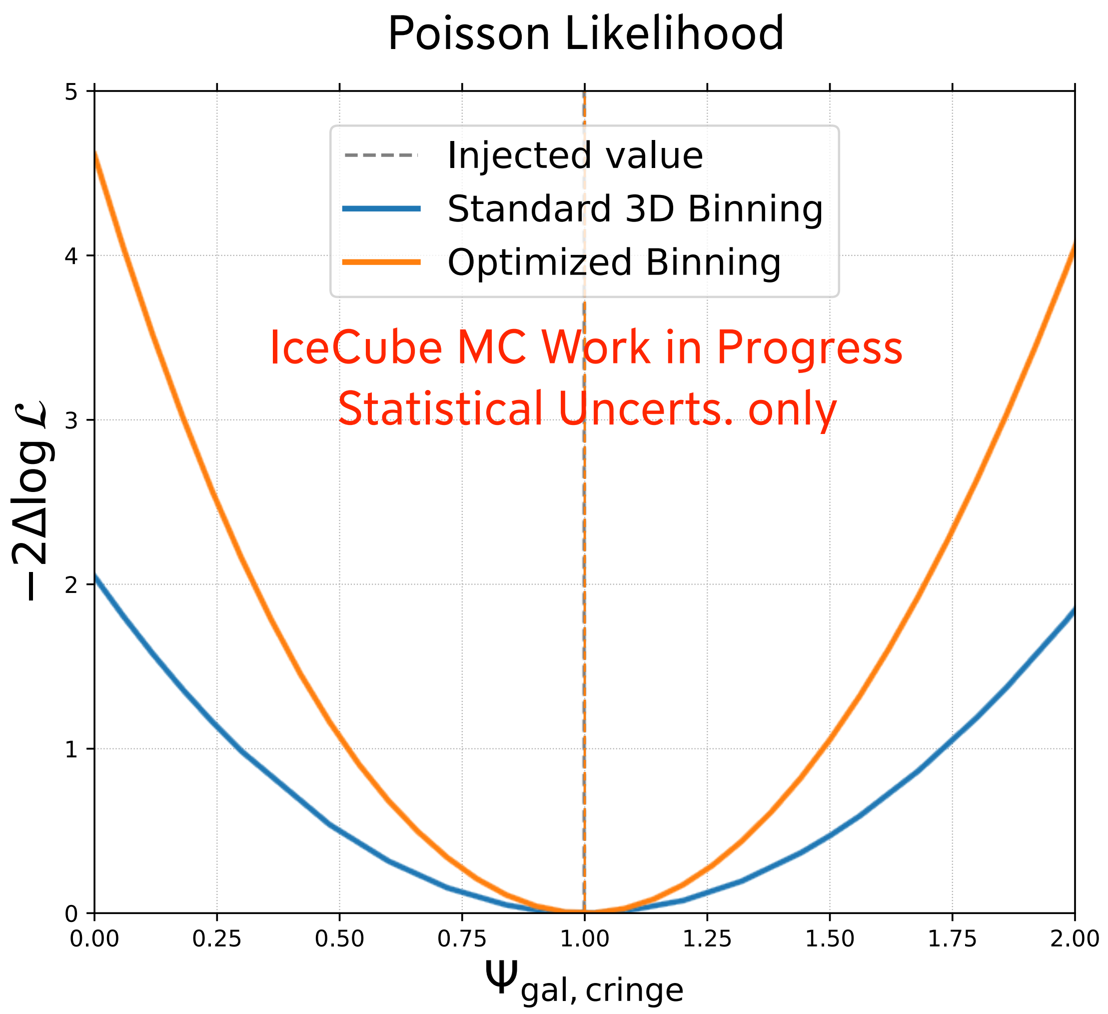
Results Histograms

Using IceCube Northern Tracks

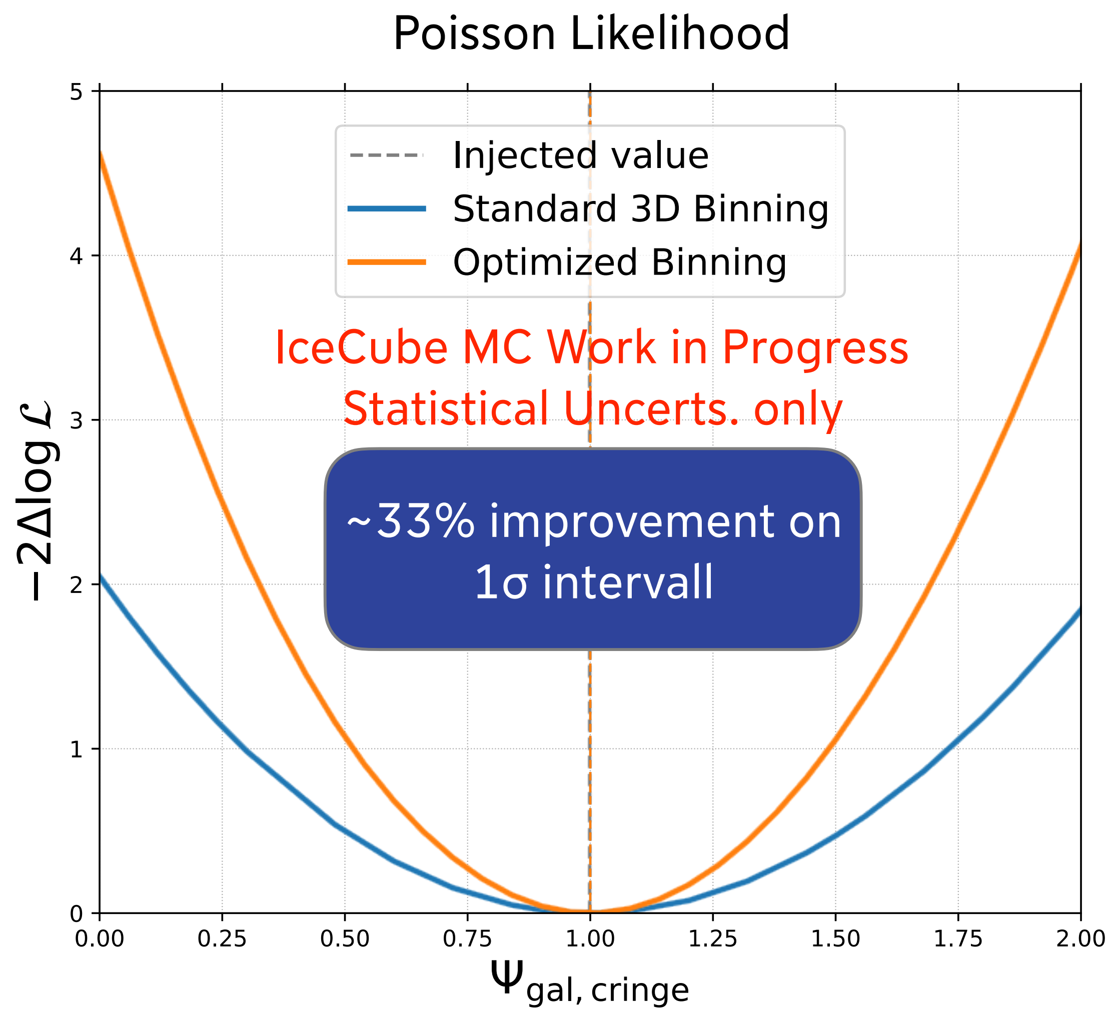




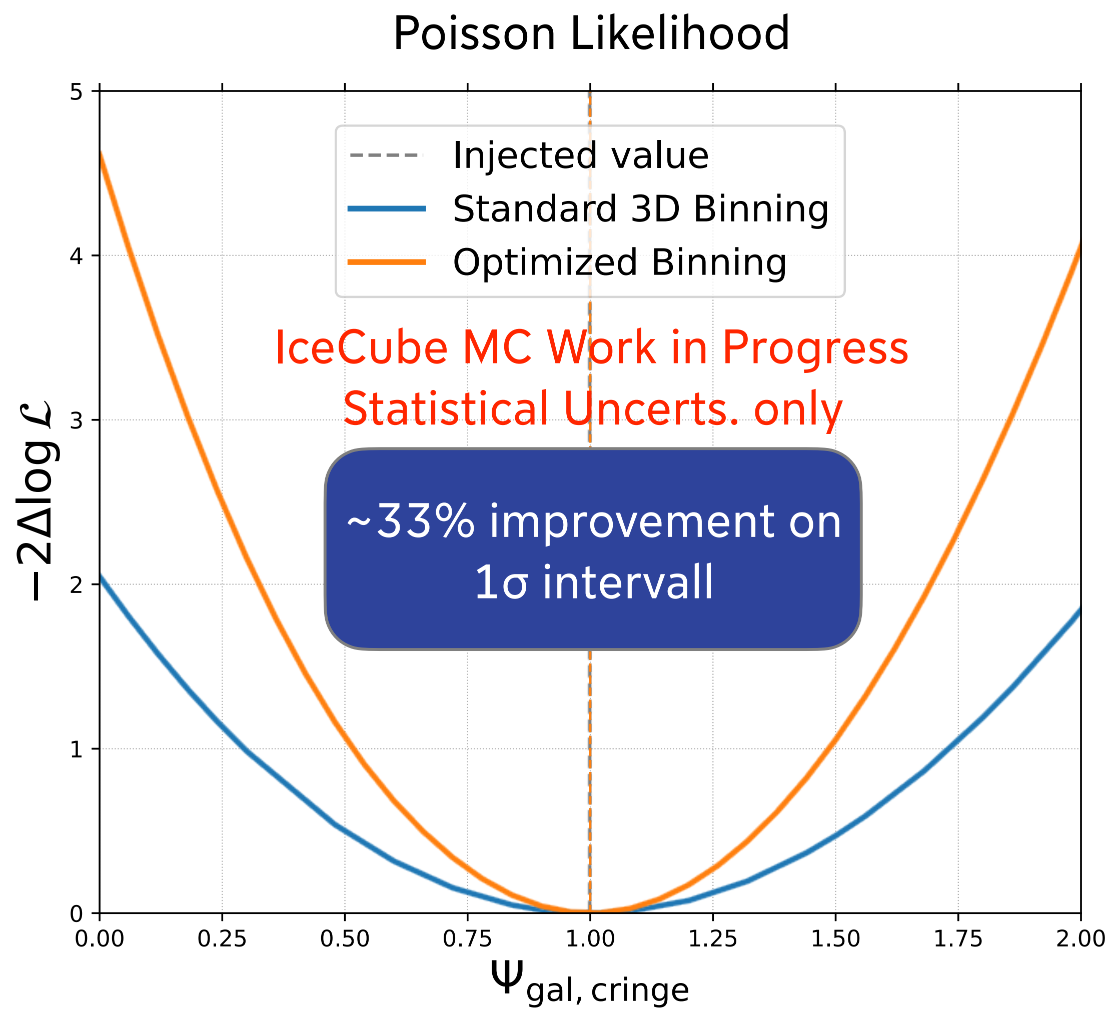
Results Likelihood Profile

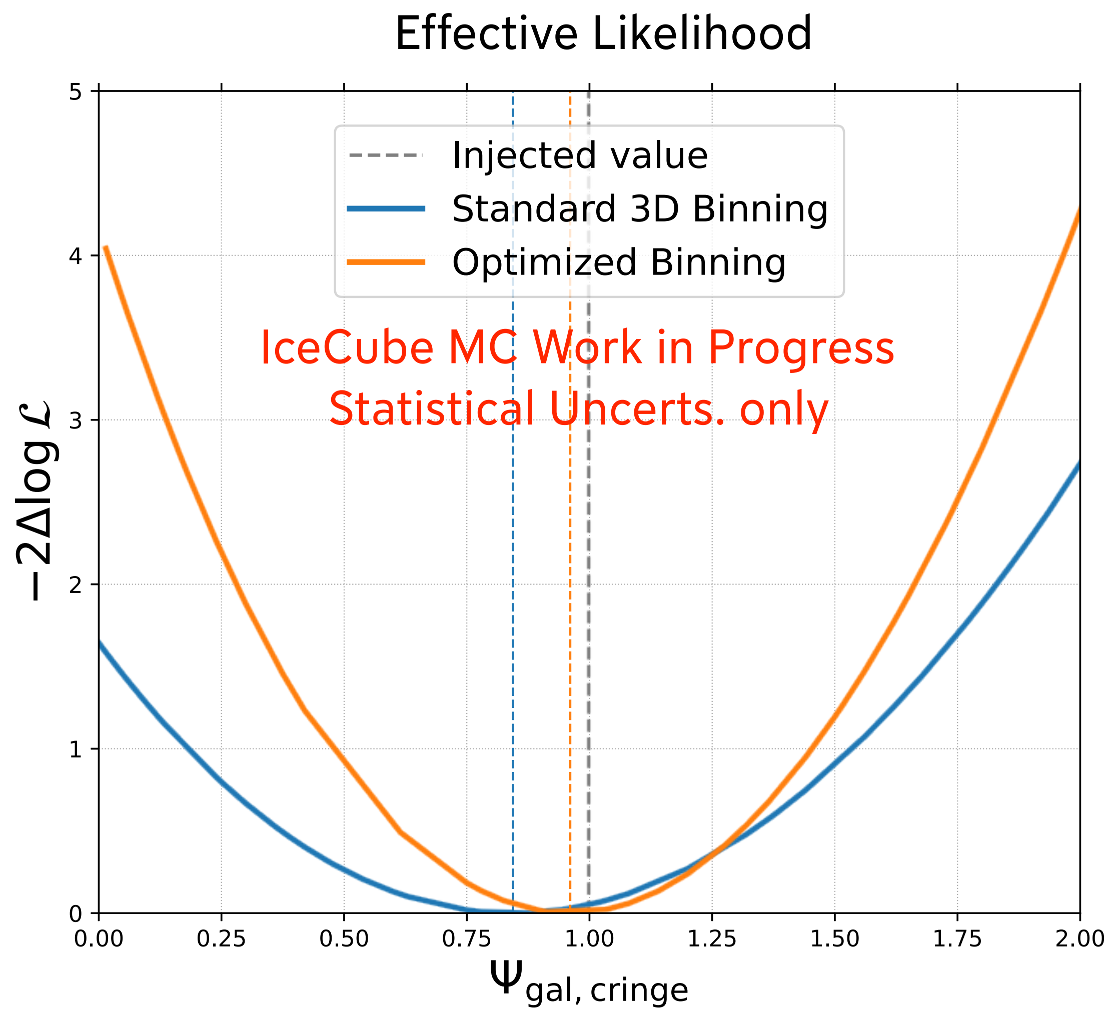


Results Likelihood Profile



Results Likelihood Profile

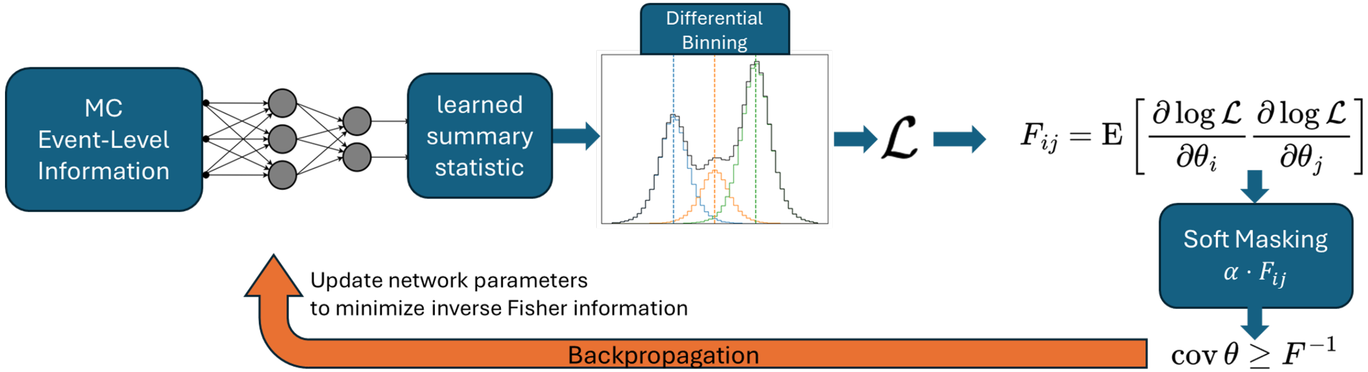


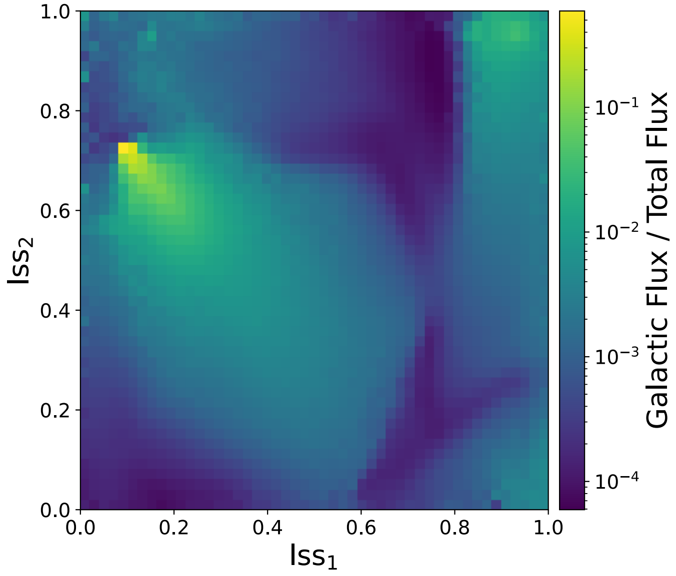


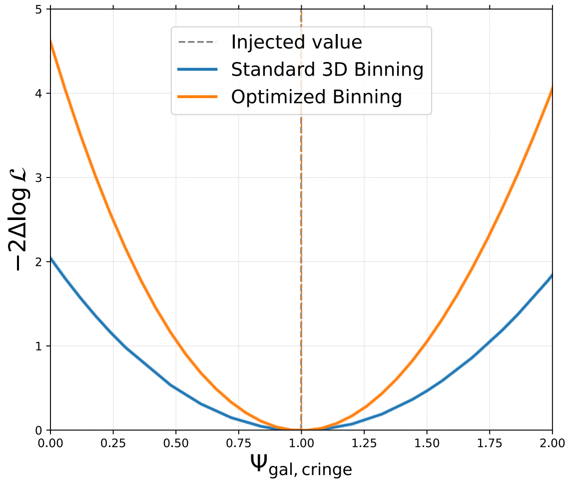
Summary

- Build fully differentiable analysis pipeline
- Improvement to standard binning by ~33%

- Use increased sensitivity to do a precise measurement of the galactic neutrino flux
- Constraining this flux can help in identifying galactic neutrino point sources







ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

Backup

Data preprocessing

Event features
Reco Energy
Reco Zenith
Reco RA
Angular uncertainty

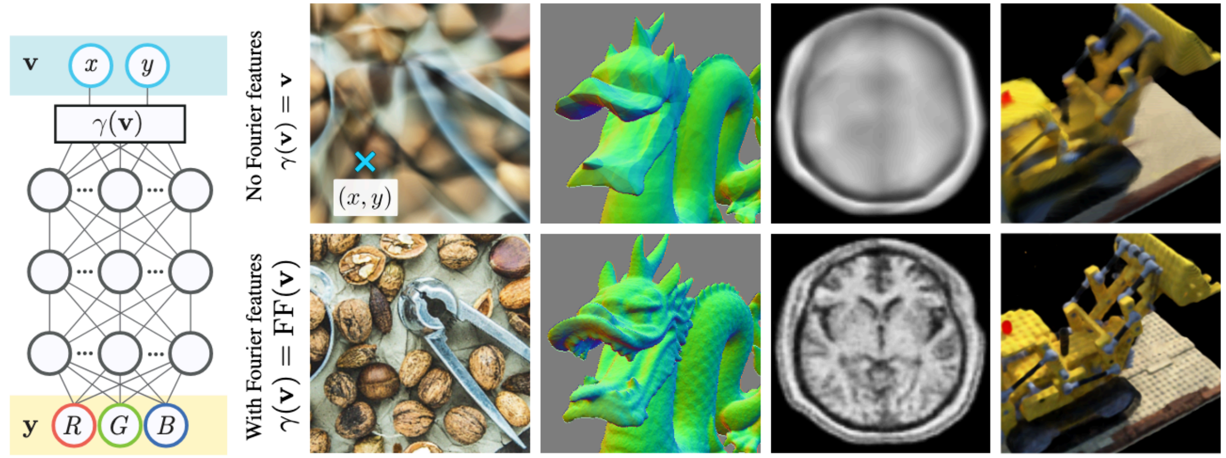
Normalize data

Select data from analysis region

Fourier Feature Mapping

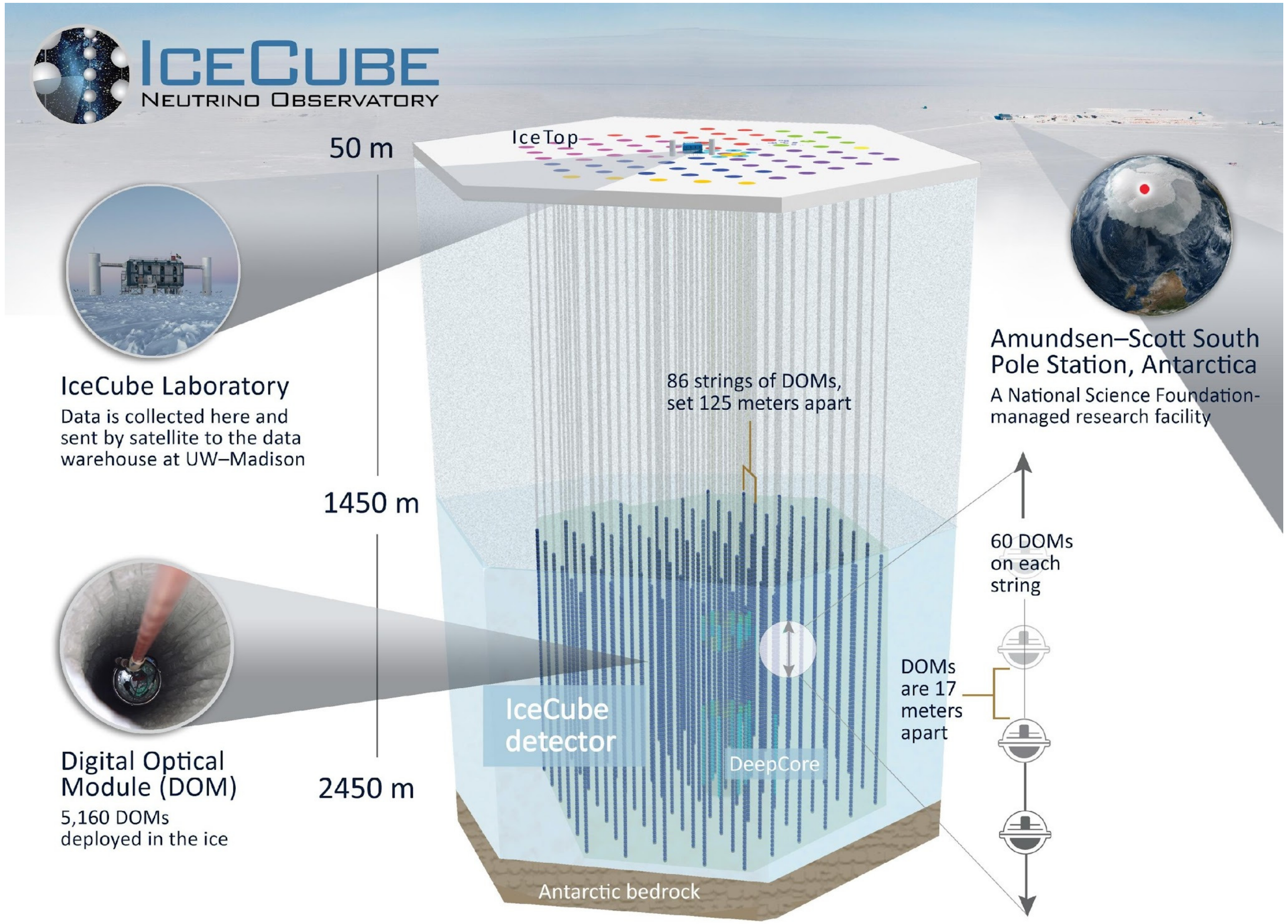
$$\gamma(\boldsymbol{v}) = \begin{bmatrix} \cos(\boldsymbol{B}\boldsymbol{v}) \\ \sin(\boldsymbol{B}\boldsymbol{v}) \end{bmatrix}$$

B sampled from $\mathcal{N}(0, \sigma^2)$



(a) Coordinate-based MLP (b) Image regression $(x,y) \rightarrow$ RGB (c) 3D shape regression $(x,y,z) \rightarrow$ occupancy (d) MRI reconstruction $(x,y,z) \rightarrow$ density (e) Inverse rendering $(x,y,z) \rightarrow$ RGB, density

IceCube Neutrino Observatory



Differential Binning

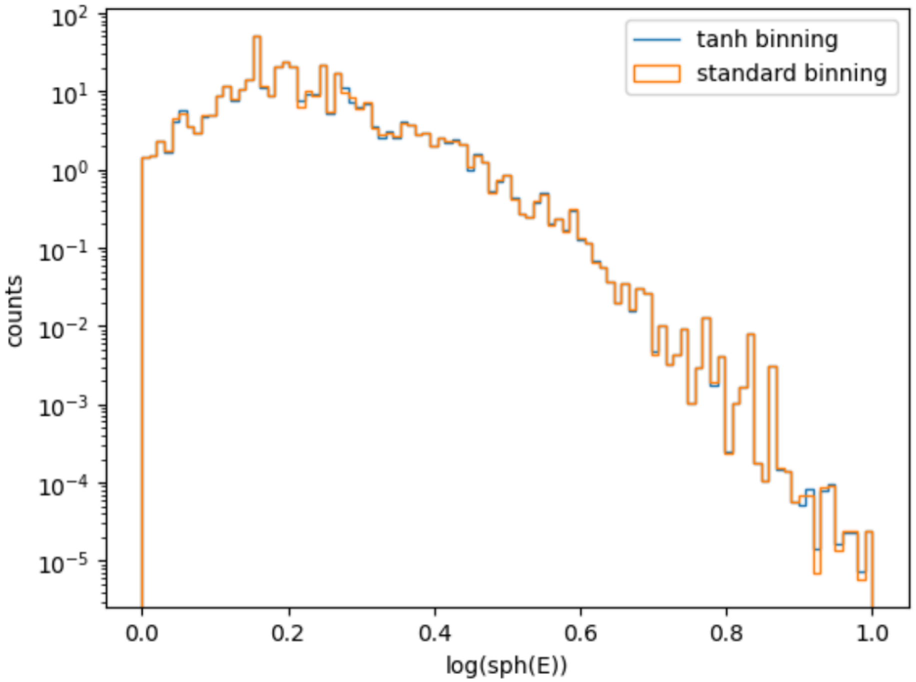
summary statistic

Bin edges

$$\text{Norm} \cdot 0.5 \left(\tanh\left(\frac{\text{lss} - b_i}{s} \right) \cdot \tanh\left(\frac{b_{i+1} - \text{lss}}{s} \right) + 1 \right)$$

Bandwidth

~twice as fast to evaluate as a binned kde

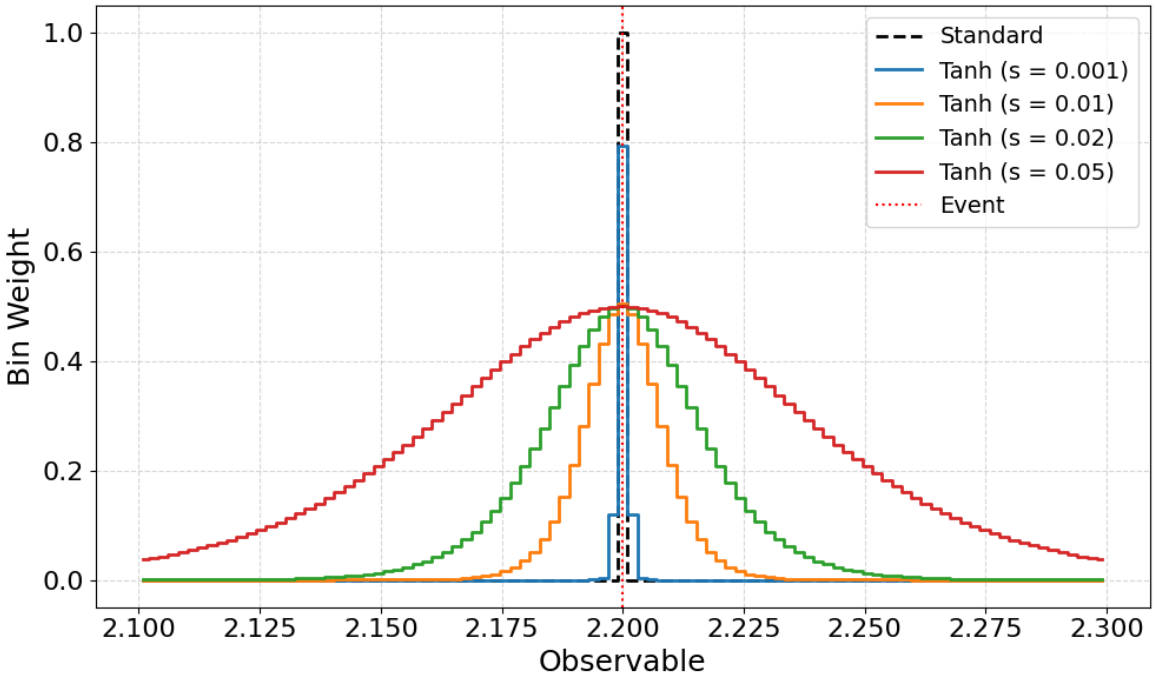


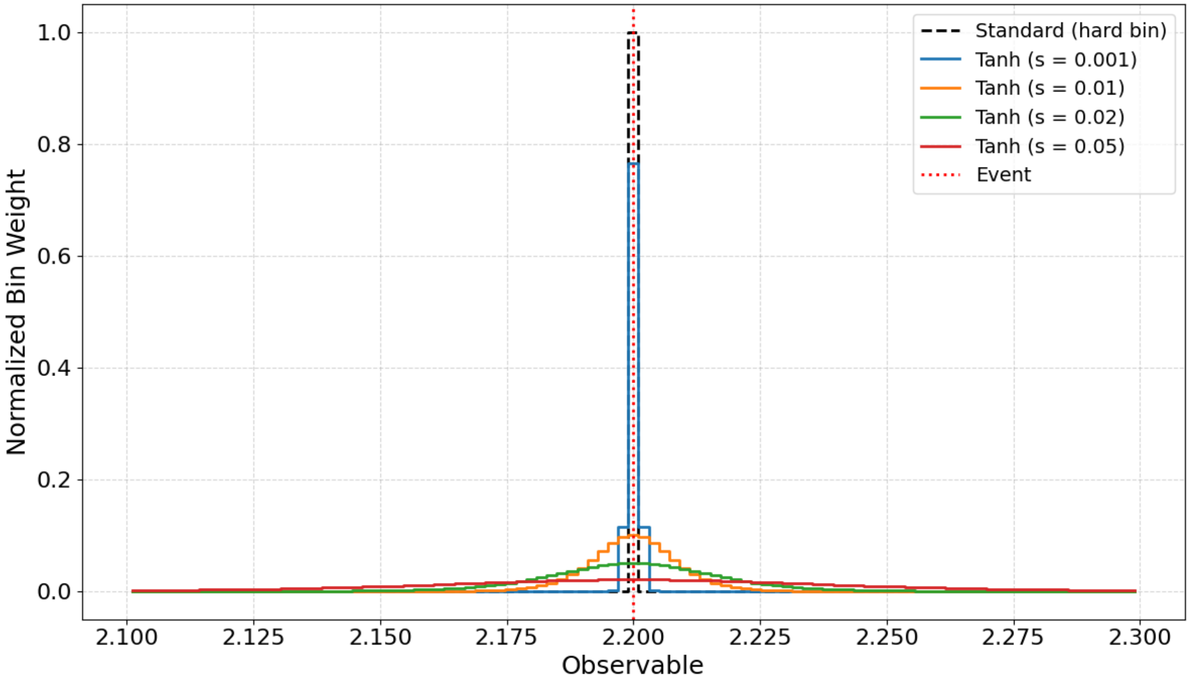
For N-dimensional:

$$\mathcal{B}_{i_1 i_2 ... i_D} = \bigotimes_{d=1}^{D} \frac{\text{Norm}}{2} \left(\tanh\left(\frac{\text{lss}_d - b_{i_d}^{(d)}}{s_d} \right) \cdot \tanh\left(\frac{b_{i_d+1}^{(d)} - \text{lss}_d}{s_d} \right) + 1 \right)$$

Outer product over all binning dimensions

Normalization of Bins





Optimization with Multiple Datasets