SEARCHES FOR SOURCES OF UHECRS AND NEUTRINOS

Francesca Capel

Max Planck Institute for Physics, Munich, Germany

Theory meeting experiments: Particle astrophysics & cosmology

19th Rencontres du Vietnam, 5th-11th January 2023

MAX-PLANCK-GESELLSCHAFT

Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

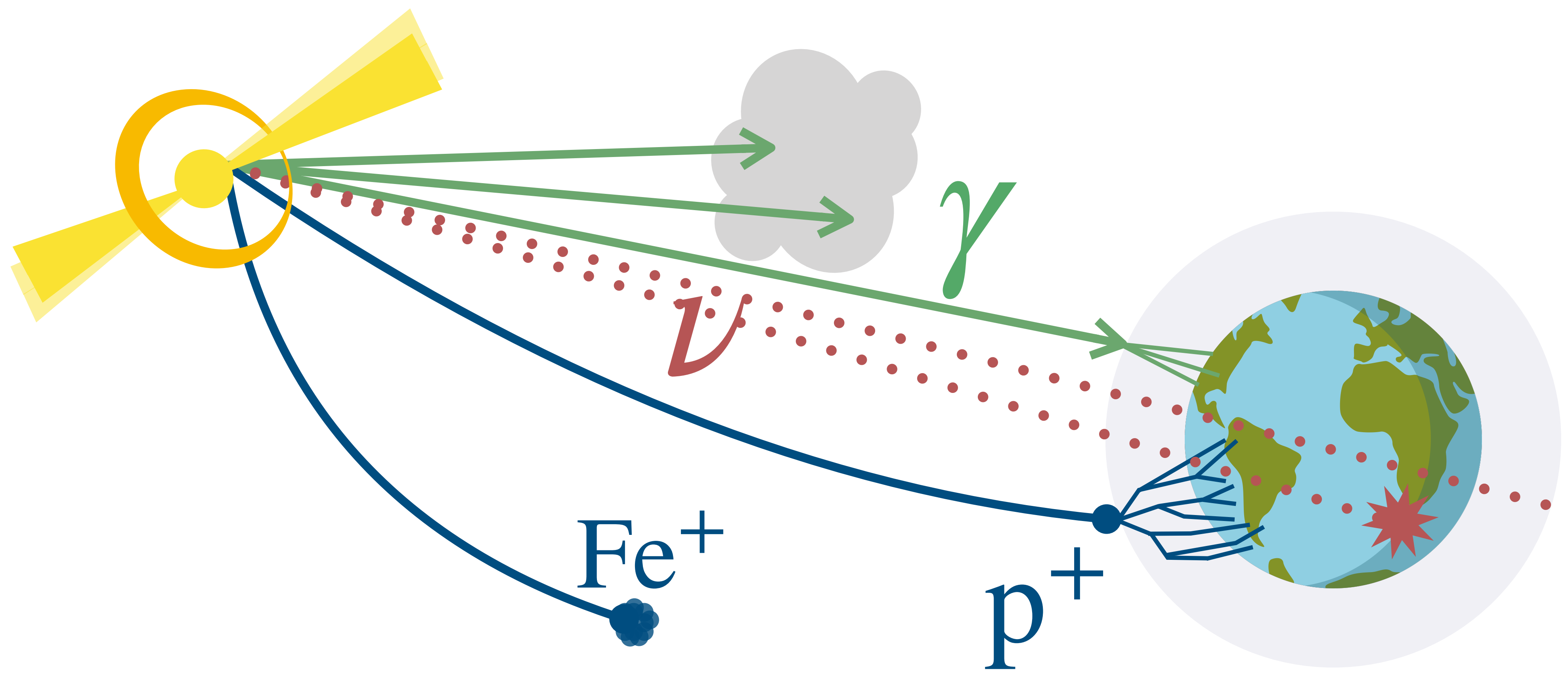
ODSL
ORIGINS Data Science Lab

SFB 1258
Neutrinos
Dark Matter
Messengers

INTRODUCTION

Goal: Identification of energetic particle accelerators

Photons, cosmic rays and neutrinos connected by hadronuclear and photohadronic interactions



$$
\begin{array}{c}
pp \text{ or } p\gamma \\
\swarrow \quad \searrow \\
\pi^0 \qquad\qquad \pi^\pm \\
\downarrow \qquad \swarrow \quad \searrow \\
\gamma\gamma \quad \nu_\mu \quad \mu \to \nu_\mu \\
\qquad\qquad\quad \swarrow \quad \searrow \\
\qquad\qquad\quad e \qquad \nu_e
\end{array}
$$

Most direct probes right now

UHECRs: > EeV

Auger, TA

Less deflected, smaller horizon

HE nu: 100 TeV – PeV

IceCube (Antares, Baikal, KM3Net)

Primary CRs ~ 1 – 10 PeV

Signature of hadronic interactions

This talk: Statistical methods for source searches with UHECR and HE nu

OUTLINE

1

Motivation

Challenges & solutions

2

Applications:

- UHECRs
- HE nu

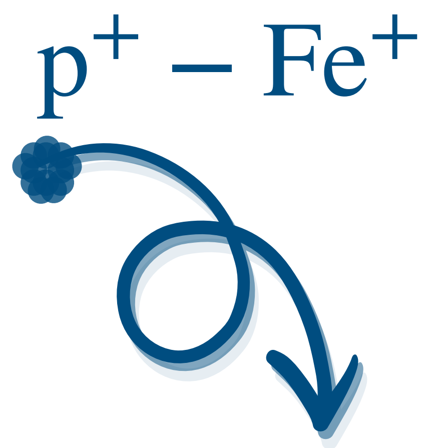
3

Next steps

Conclusions

MOTIVATION

Large-scale experiments have been collecting data for ~10 – 20 years
Many exciting results, but still working towards a coherent bigger picture



UHECRs

AGN, Starburst galaxies
Aab et al. 2018

Large scale structure
Aab et al. 2017, Ding et al. 2021

HE nu

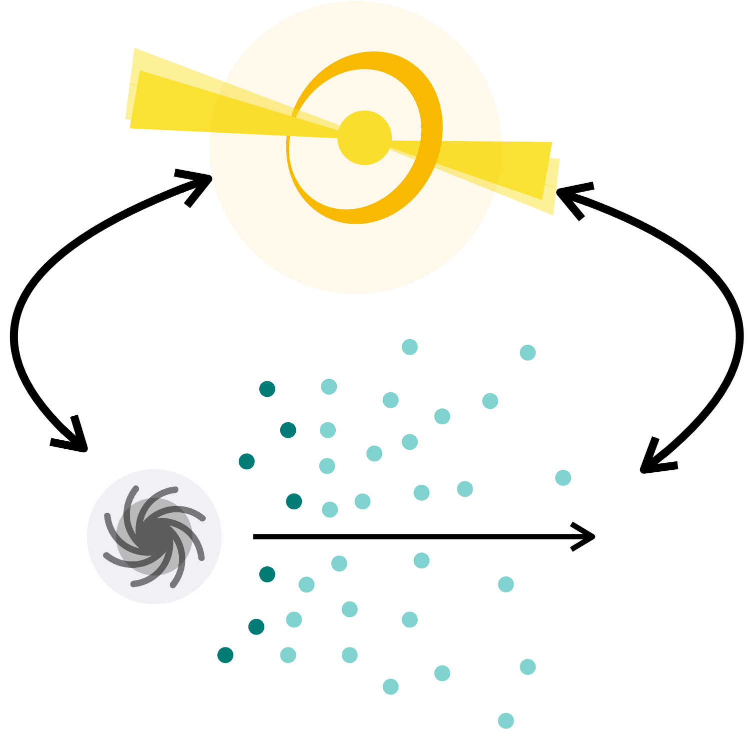
AGN, blazars
IceCube et al. 2018a, 2018b, Plavin et al . 2020, Buson et al. 2022…

NGC 1068/Seyfert II galaxies
Abbasi et al. 2022

Lots of relevant multi-messenger data from different observatories

Individual sources

Multi-wavelength data
Variability
Morphology
Redshift



Populations

Source densities
Luminosity functions
Cosmological evolution

Increasing complexity of both data and models

MOTIVATION

What to do?

- Bigger and better experiments
- Advanced statistical analysis

MOTIVATION

What to do?

See talks in the future experiments session
Karle, Suomijärvi, Battisti

Bigger and better experiments

Advanced statistical analysis

Prepare for next-generation experiments

Make the most of existing data

Improve model–data connection

Many recent developments in computational statistics

CHALLENGES

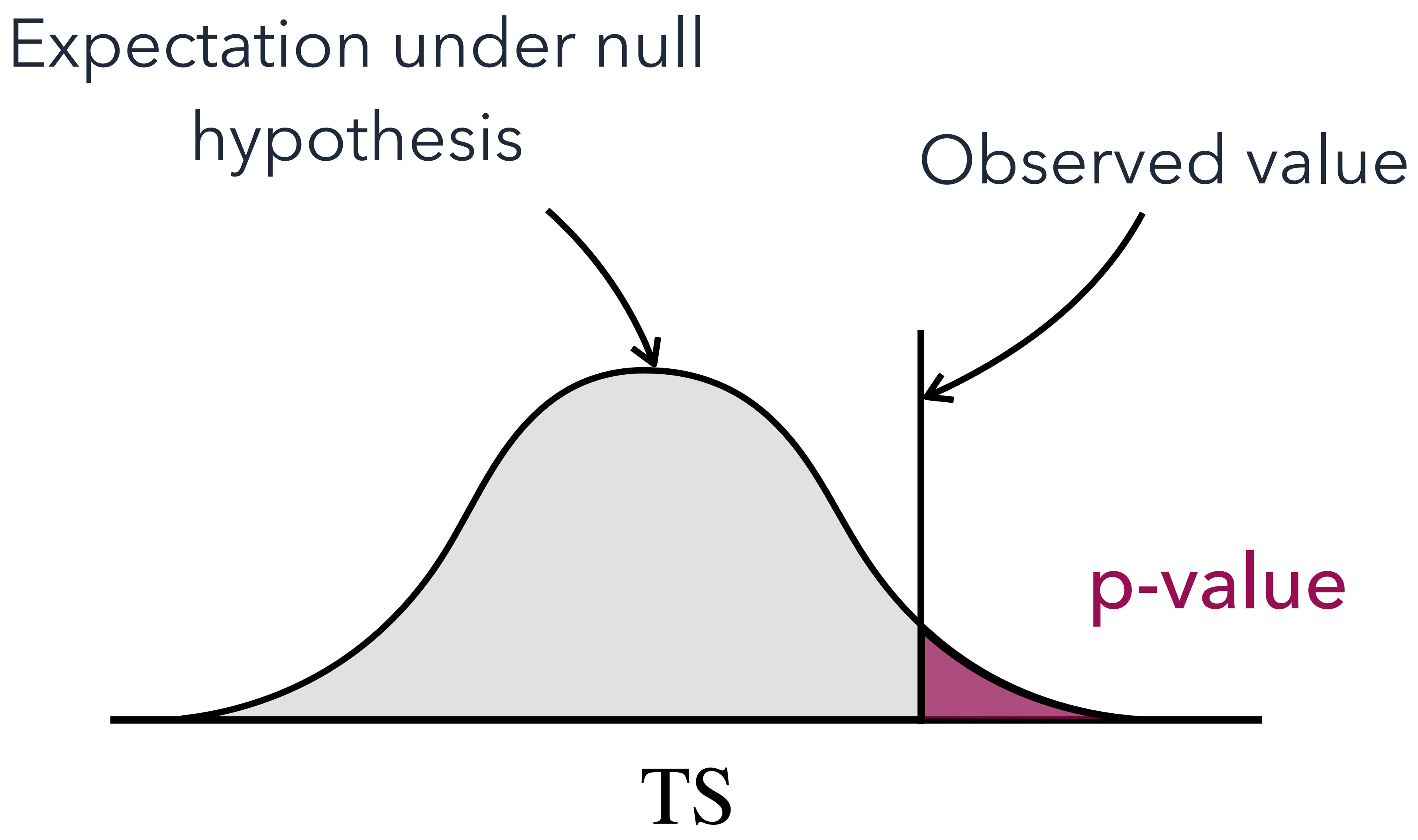
Typical approach to source searches: **Null hypothesis significance testing**
Significance is computed via a **test statistic** and calibrated via simulations

E.g. Likelihood-ratio test

Alternative hypothesis + best-fit parameters

$$\mathrm{TS} \propto \log \frac{\mathscr{L}(\hat{\theta}_{\mathrm{source}})}{\mathscr{L}(\theta_{\mathrm{BG}})}$$

Null hypothesis



Best fit parameters usually found via e.g. grid scans, Minuit

Braun et al. 2008, 2010, Aab et al. 2018…

CHALLENGES

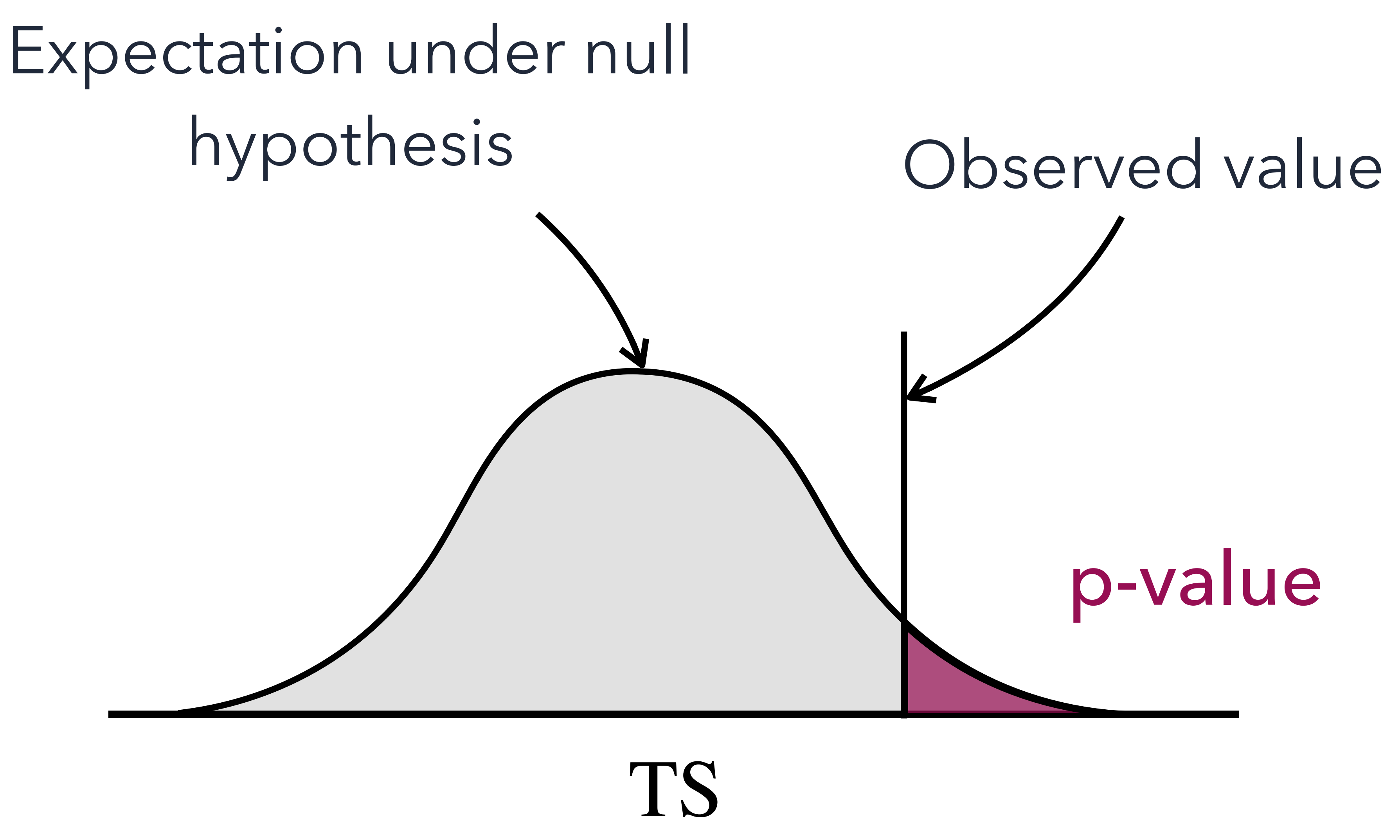
Typical approach to source searches: **Null hypothesis significance testing**
Significance is computed via a **test statistic** and calibrated via simulations

E.g. Likelihood-ratio test

Alternative hypothesis + best-fit parameters

$$\mathrm{TS} \propto \log \frac{\mathscr{L}(\hat{\theta}_{\mathrm{source}})}{\mathscr{L}(\theta_{\mathrm{BG}})}$$

Null hypothesis



Best fit parameters usually found via e.g. grid scans, Minuit

Braun et al. 2008, 2010, Aab et al. 2018…

Some limitations:

- Number of **free model parameters**
- **Complexity** in theoretical and data models
- **Determination of uncertainties** in best-fit parameters
- **Interpretation** always in relation to null hypothesis
- Analysis choices and sensitivity driven by **trial-factor corrections**

See also: Gigerenzer et al. 2004, Wasserstein & Lazar 2016

SOLUTIONS

What alternatives do we have to improve or complement these analyses?

- Number of **free model parameters**
- **Complexity** in theoretical and data models
- **Determination of uncertainties** in best-fit parameters
- **Interpretation** always in relation to null hypothesis
- Analysis choices and sensitivity driven by **trial-factor corrections**

SOLUTIONS

What alternatives do we have to improve or complement these analyses?

- Number of **free model parameters**
- **Complexity** in theoretical and data models
- **Determination of uncertainties** in best-fit parameters

- **Interpretation** always in relation to null hypothesis
- Analysis choices and sensitivity driven by **trial-factor corrections**

SOLUTIONS

What alternatives do we have to improve or complement these analyses?

- Number of **free model parameters**
- **Complexity** in theoretical and data models
- **Determination of uncertainties** in best-fit parameters

Markov chain Monte Carlo

Hierarchical modelling

- **Interpretation** always in relation to null hypothesis
- Analysis choices and sensitivity driven by **trial-factor corrections**

SOLUTIONS

What alternatives do we have to improve or complement these analyses?

- Number of **free model parameters**
- **Complexity** in theoretical and data models
- **Determination of uncertainties** in best-fit parameters

Markov chain Monte Carlo

Hierarchical modelling

- **Interpretation** always in relation to null hypothesis
- Analysis choices and sensitivity driven by **trial-factor corrections**

Bayesian inference

SOLUTIONS

Markov chain Monte Carlo

Algorithm to numerically approximate **high-dimensional integrals** (e.g. expectation values, variances of parameters)

Exact convergence in the limit of infinite samples

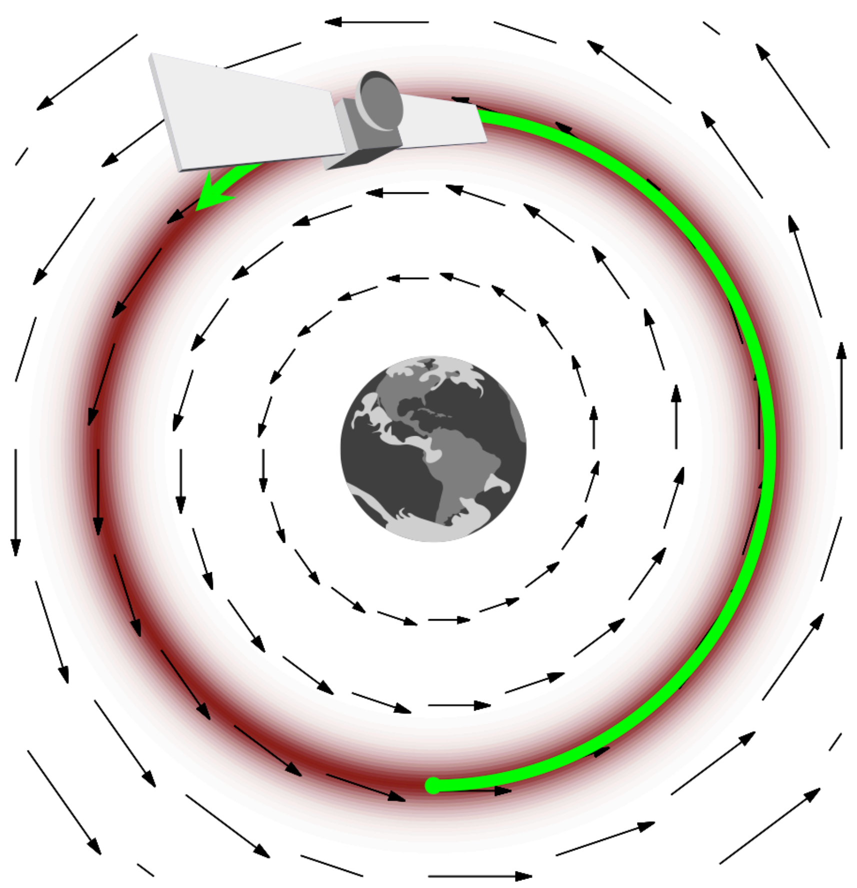
$$\int_\Theta d\theta\, f(\theta)p(\theta) = \lim_{N\to\infty} \frac{1}{N}\sum_{n=0}^{N} f(\theta_n)$$

Hamiltonian Monte Carlo

A type of Markov chain Monte Carlo that uses **Hamiltonian dynamics** to move **efficiently** through high-dimensional parameter spaces

$$\theta \longrightarrow (\theta, p) \qquad \frac{d\theta}{dt} = \frac{\partial H}{\partial p}$$

$$H(\theta, p) \equiv \log P(\theta, p) \qquad \frac{dp}{dt} = -\frac{\partial H}{\partial \theta}$$



Betancourt (2014)

SOLUTIONS

Markov chain Monte Carlo

Algorithm to numerically approximate **high-dimensional integrals** (e.g. expectation values, variances of parameters)

Exact convergence in the limit of infinite samples

$$\int_\Theta d\theta\, f(\theta)p(\theta) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N} f(\theta_n)$$

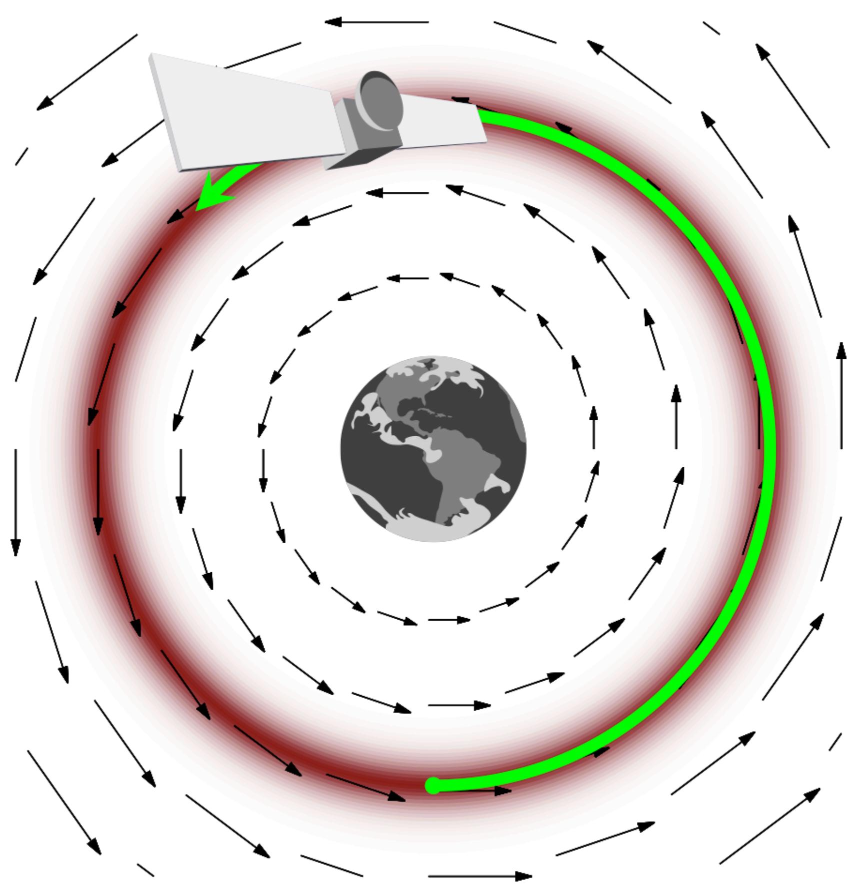
Large numbers of free parameters possible

Hamiltonian Monte Carlo

A type of Markov chain Monte Carlo that uses **Hamiltonian dynamics** to move **efficiently** through high-dimensional parameter spaces

$$\theta \longrightarrow (\theta, p) \qquad \frac{d\theta}{dt} = \frac{\partial H}{\partial p}$$

$$H(\theta, p) \equiv \log P(\theta, p) \qquad \frac{dp}{dt} = -\frac{\partial H}{\partial \theta}$$



Betancourt (2014)

Uncertainty quantification for free

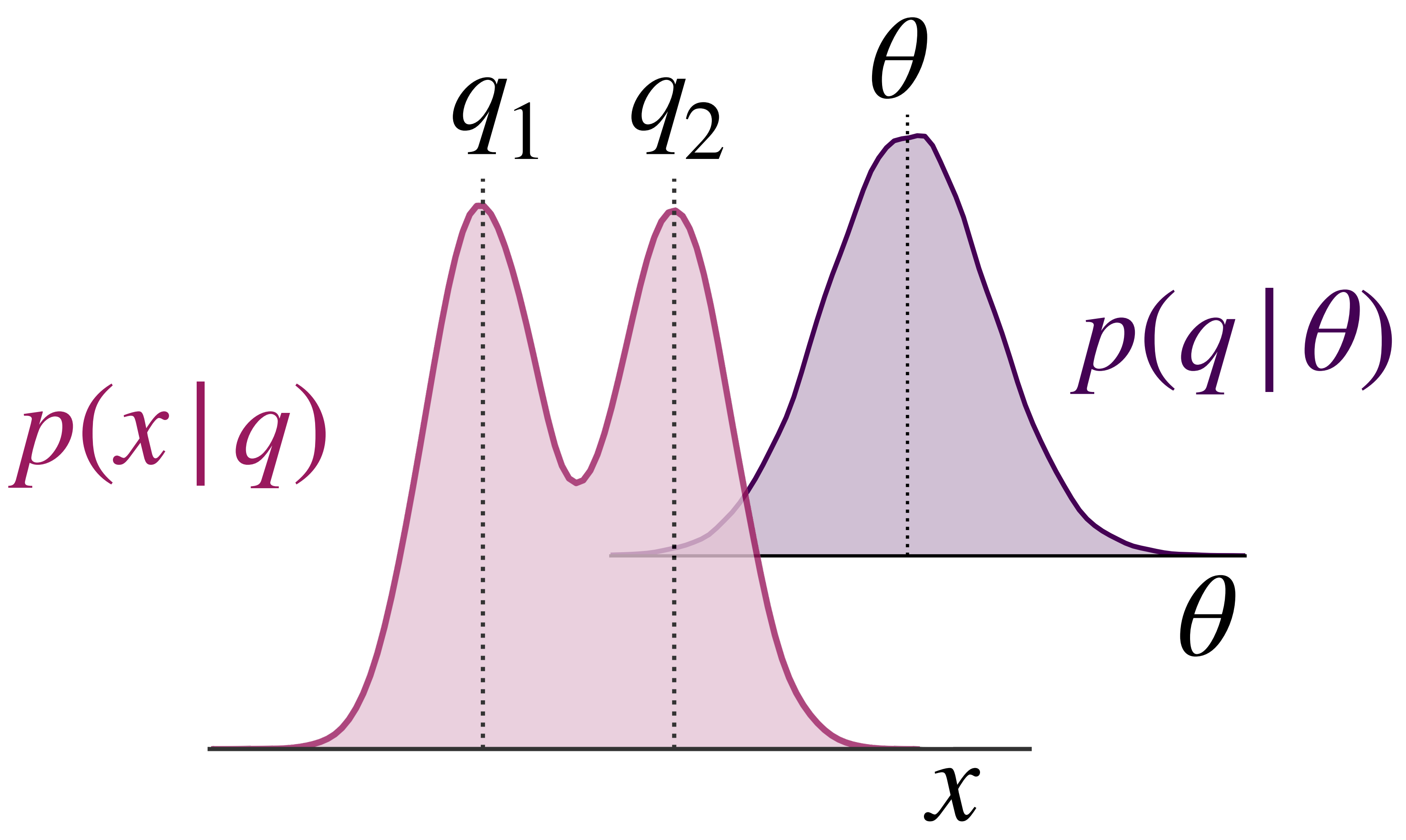
SOLUTIONS

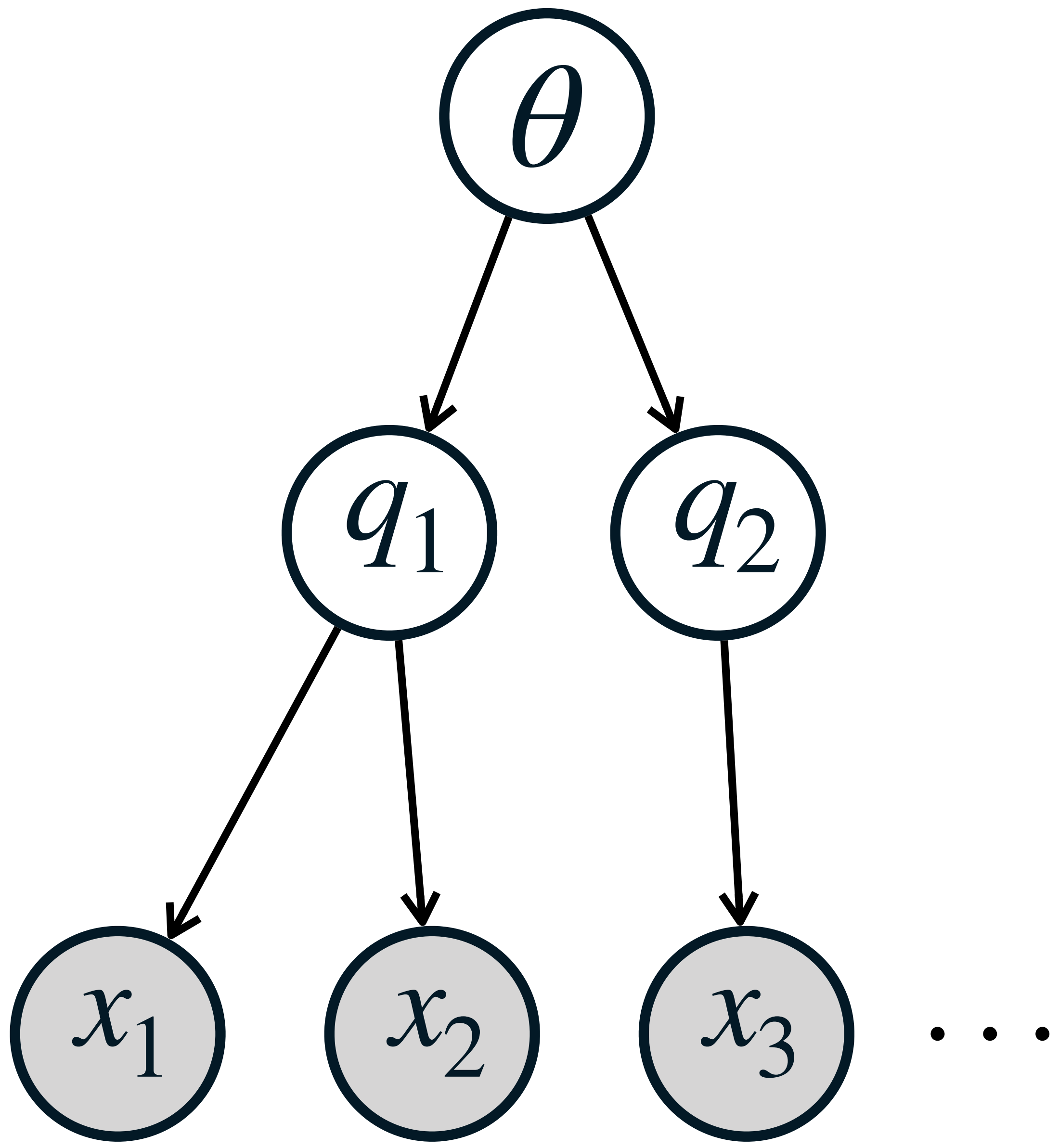
Include more complexity into analysis

Hierarchical modelling

Organise the free parameters into a hierarchy that describes the **data generating process**

$$\mathscr{L}(x, \theta) = p(x \,|\, q)\; p(q \,|\, \theta) \ldots$$





High-level parameters

Latent parameters

Observations

SOLUTIONS

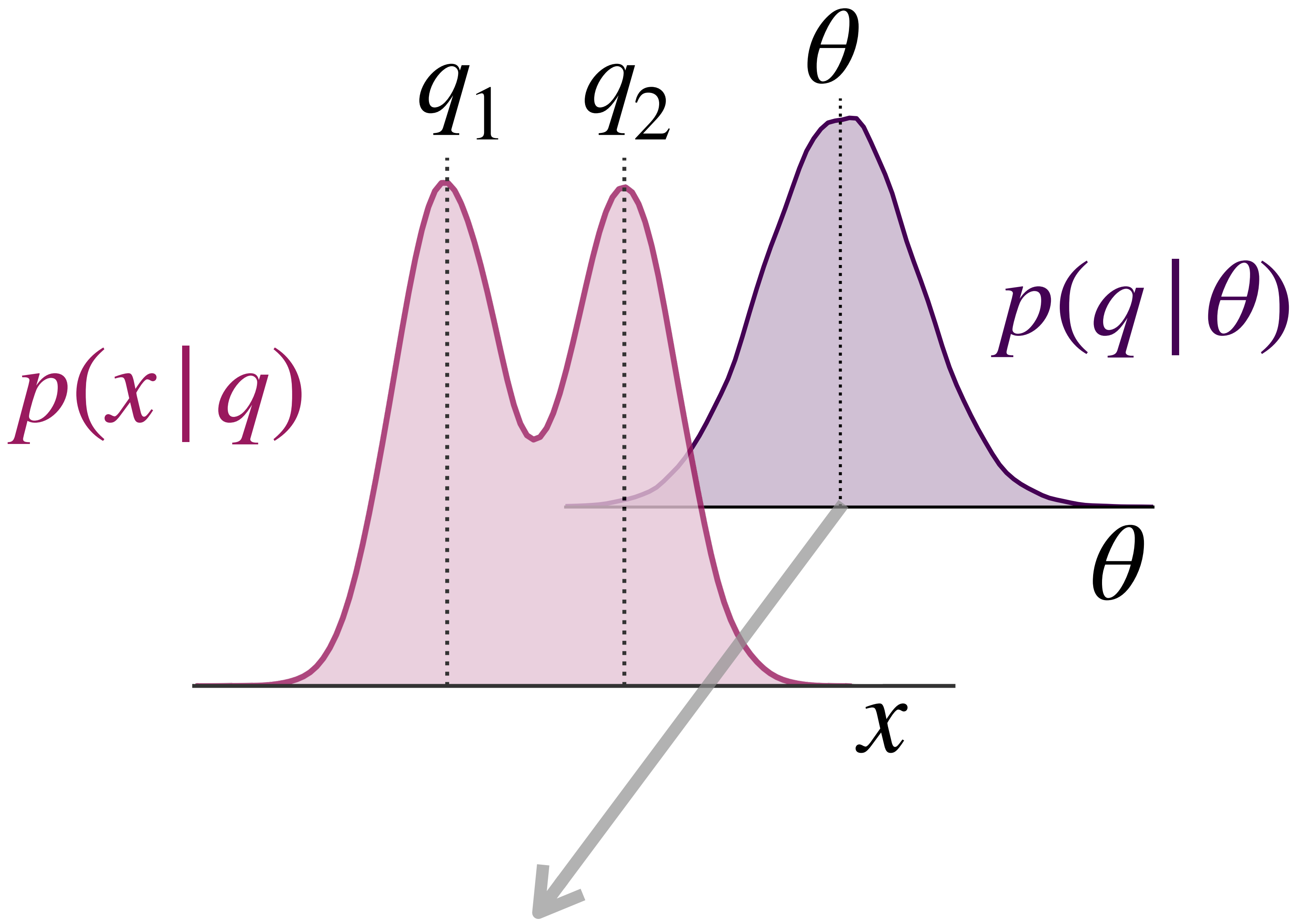
Include more complexity into analysis

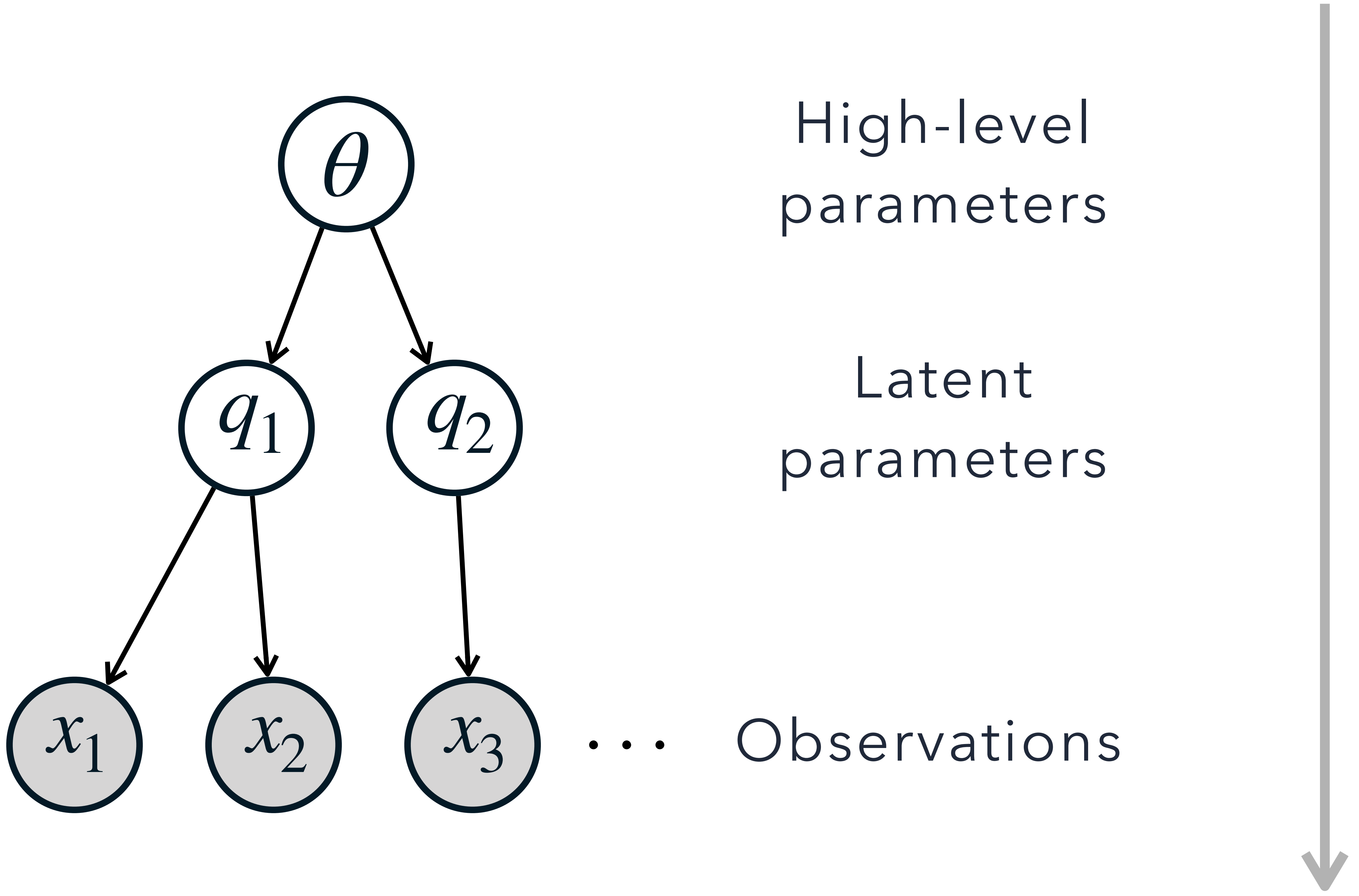
Simulation

Hierarchical modelling

Organise the free parameters into a hierarchy that describes the **data generating process**

$$\mathscr{L}(x, \theta) = p(x \mid q)\ p(q \mid \theta) \dots$$





SOLUTIONS

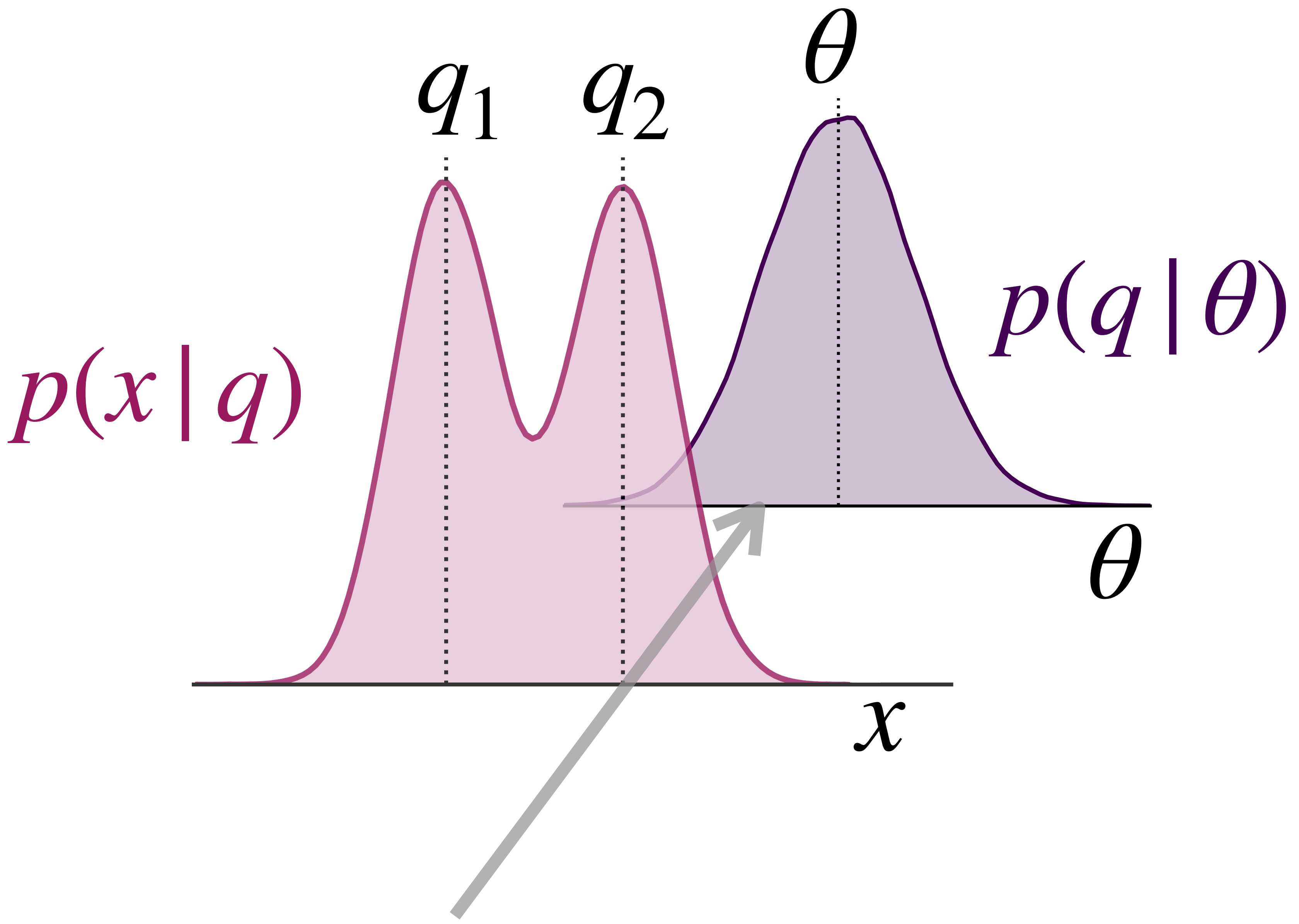
Include more complexity into analysis

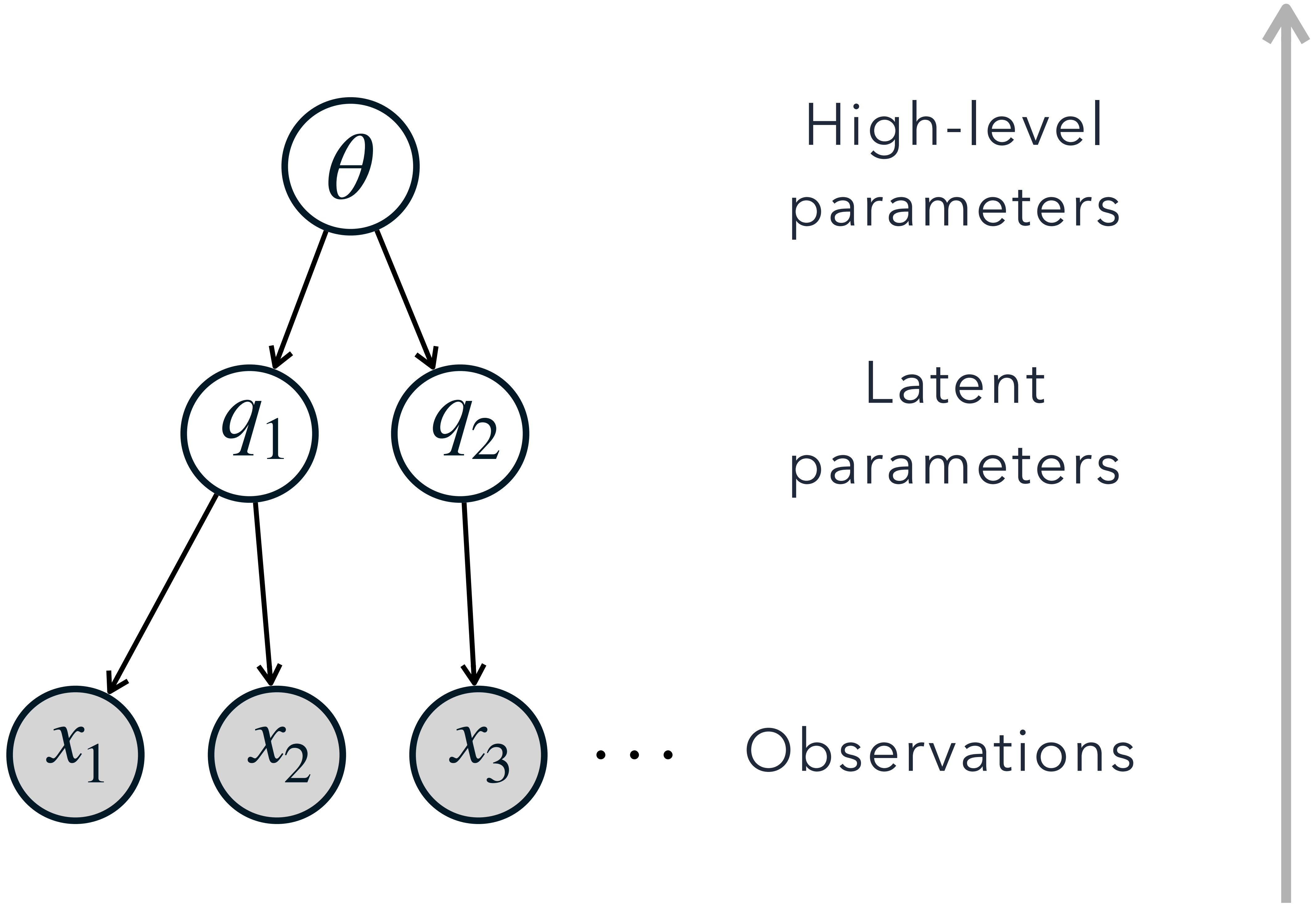
Inference

Hierarchical modelling

Organise the free parameters into a hierarchy that describes the **data generating process**

$$\mathscr{L}(x, \theta) = p(x \,|\, q)\; p(q \,|\, \theta) \ldots$$





SOLUTIONS

Bayesian inference

Frequentist: Construct procedures with frequency guarantees

"If repeating the experiment with only background, how unlikely is this result?"

If testing multiple hypothesis, more likely to find extreme results

Bayesian: Quantify degree of belief

"What is the probability this particle originates from this source?"

Impact of extreme results is mitigated by priors and model structure

Wasserman 2004, Gelman et al. 2012 & 2014, Kruschke et al. 2017

Complementary analyses and interpretations

SOLUTIONS

Bayesian inference

Frequentist: Construct procedures with frequency guarantees
"If repeating the experiment with only background, how unlikely is this result?"
If testing multiple hypothesis, more likely to find extreme results

Bayesian: Quantify degree of belief
"What is the probability this particle originates from this source?"
Impact of extreme results is mitigated by priors and model structure

Wasserman 2004, Gelman et al. 2012 & 2014, Kruschke et al. 2017

Complementary analyses and interpretations

Summary
- Markov chain Monte Carlo
- Hierarchical modelling
- Bayesian inference

Application to source searches

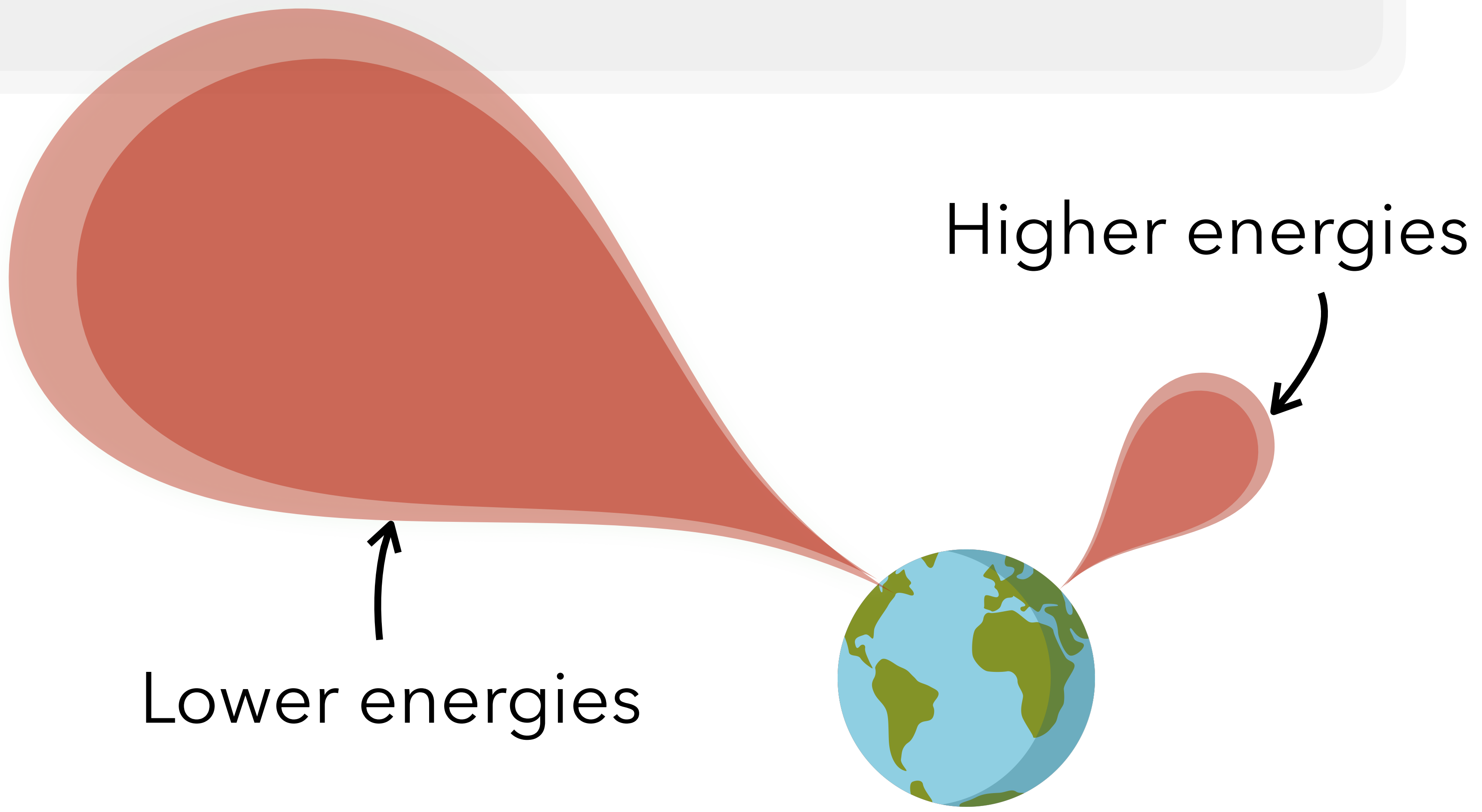
1 UHECRs   2 HE nu

UHECR SOURCE SEARCHES

Focus: Including complexity of UHECR phenomenology into statistical analysis and allowing for more direct interpretation of the results

UHECR propagation

- **Energy losses** through interactions
- **Deflections** in magnetic fields
- **Photodisintegration** of nuclei



Possible origins of a detected UHECR depend on measured **arrival direction**, **energy**, **composition** and **propagation history**

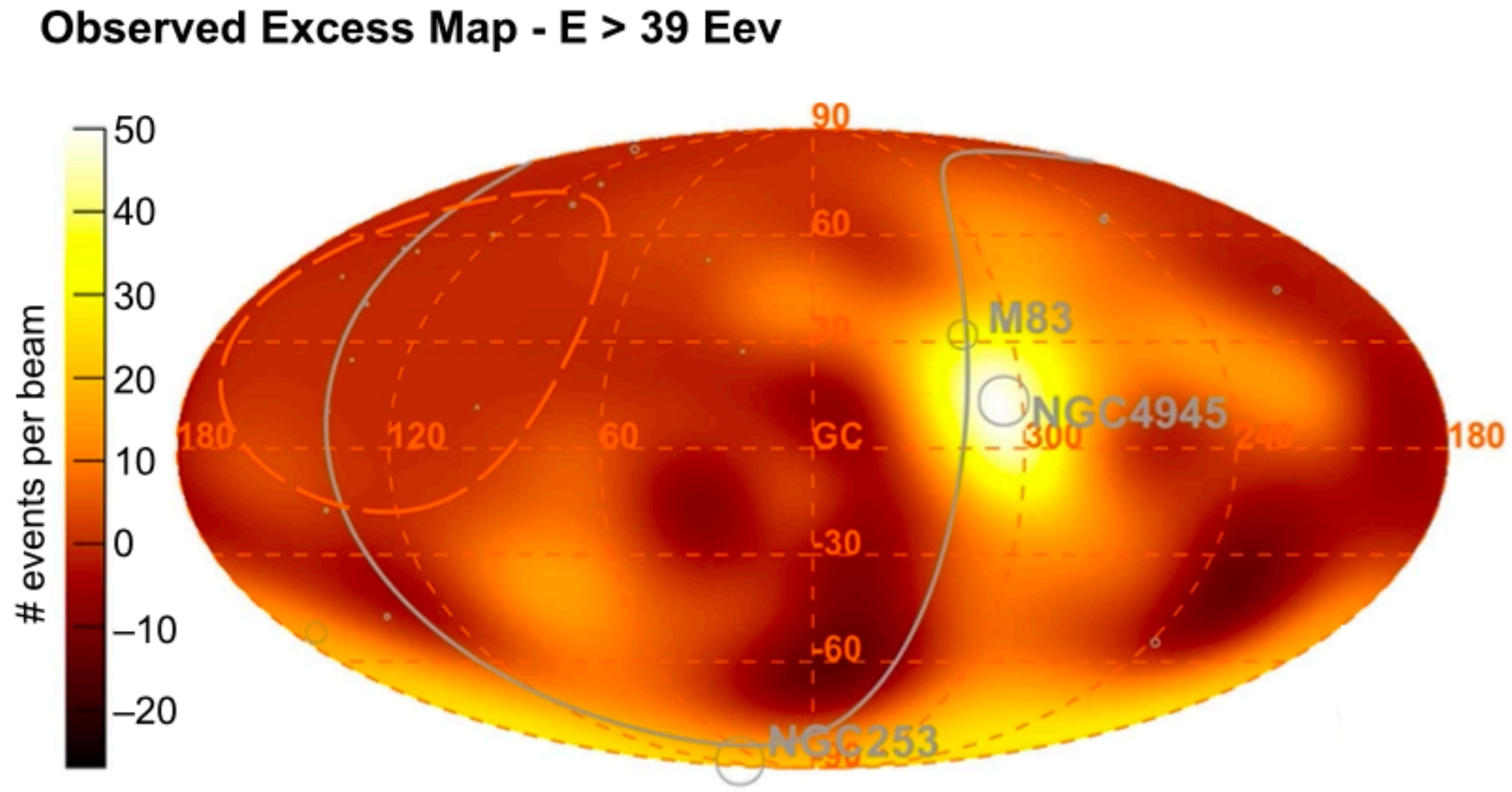
Baseline approach

2 free parameters; **search radius** and **associated fraction**

Evidence presented as **rejection of isotropy**

Penalty for scan over threshold energies

Aab et al. 2018, see also earlier talk by Andringer for latest work



UHECR SOURCE SEARCHES

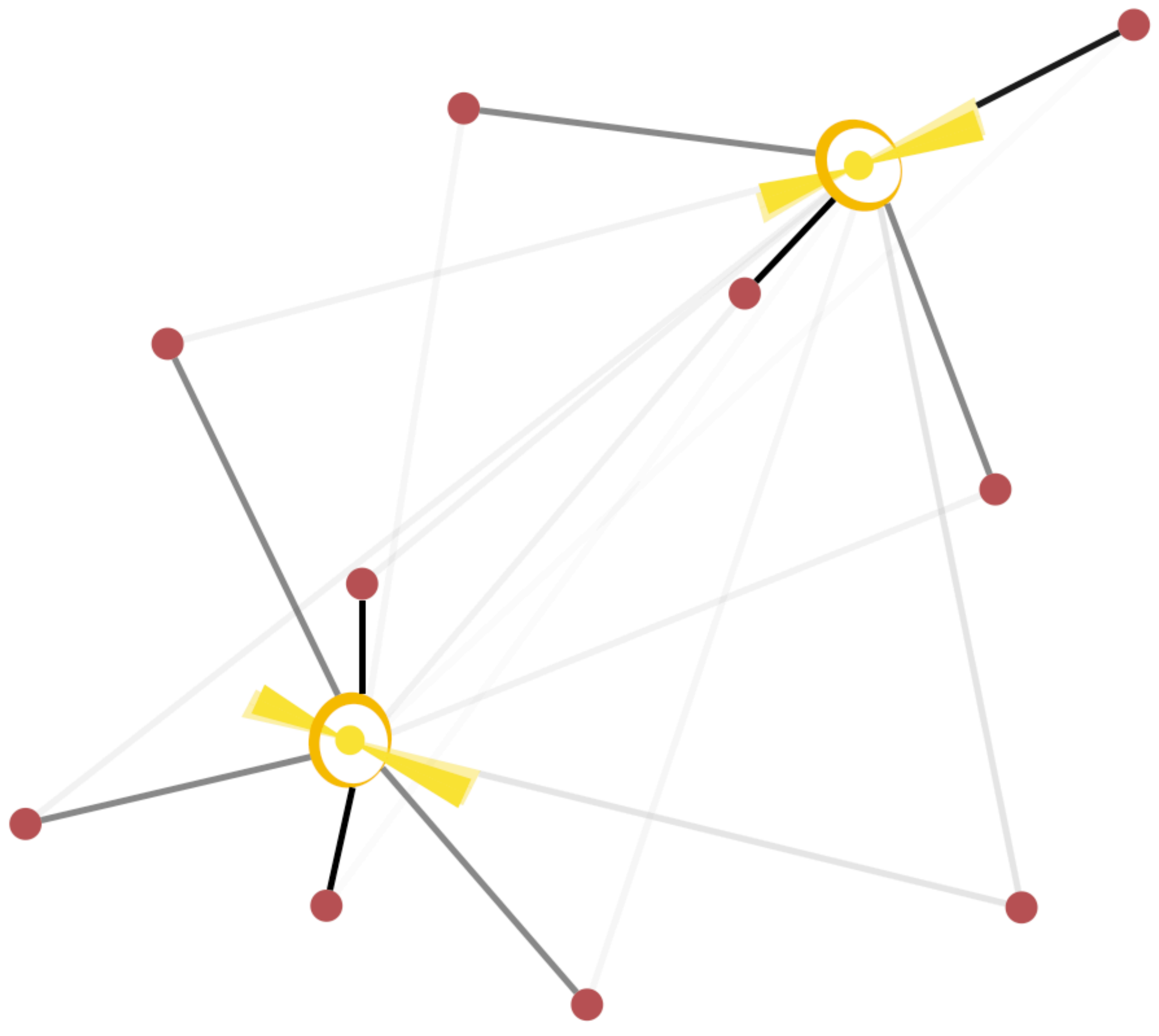
Bayesian hierarchical model

Joint fit of spectrum and arrival directions

~100s free parameters, including **physical hyperparameters**

Results in terms of **individual source-UHECR associations**

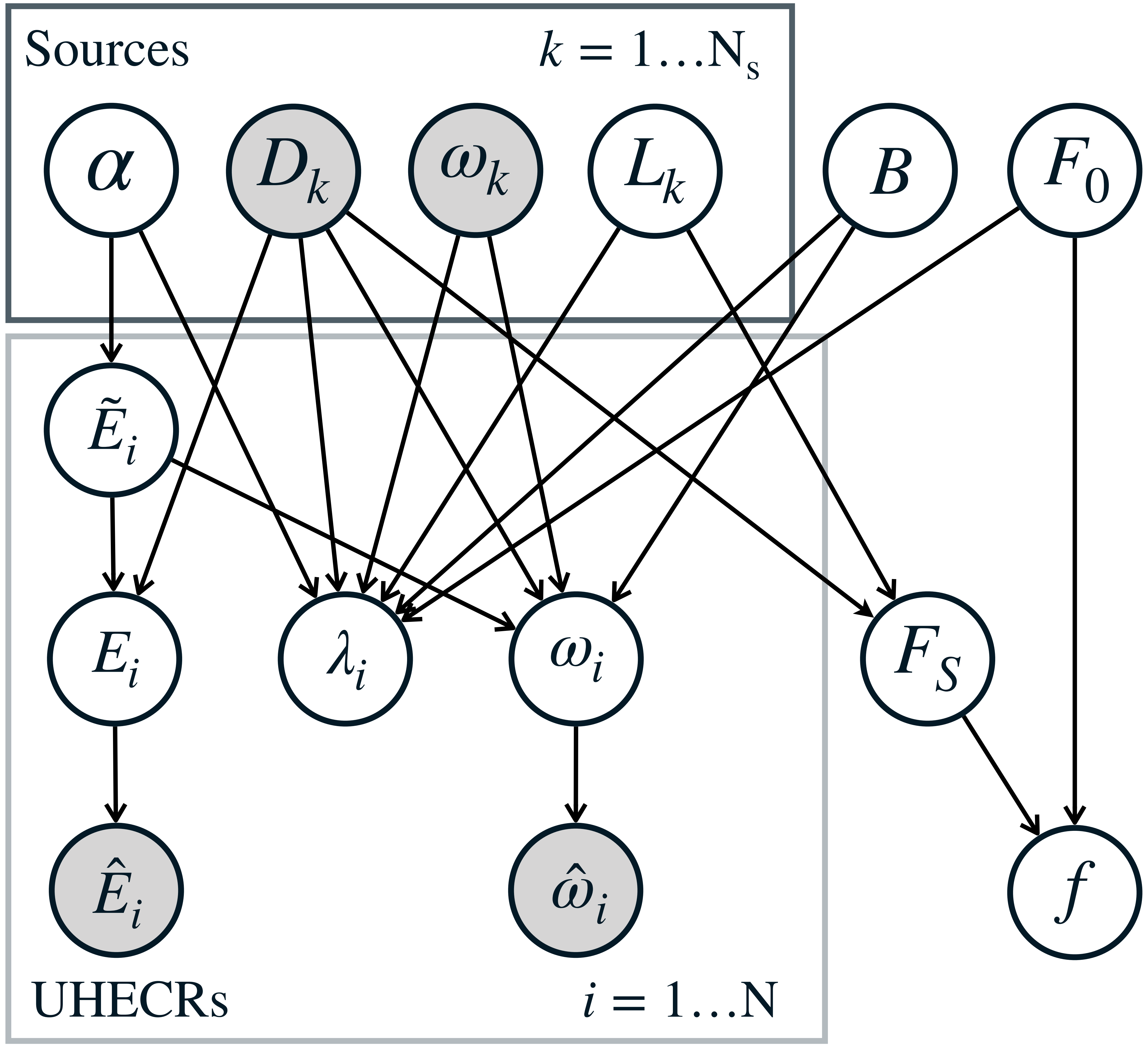
Unphysical associations removed



$P(\text{associated} \,|\, \text{data})$

Implemented via **Hamiltonian Monte Carlo** in **Stan**

Betancourt et al. 2017
Carpenter et al. 2017

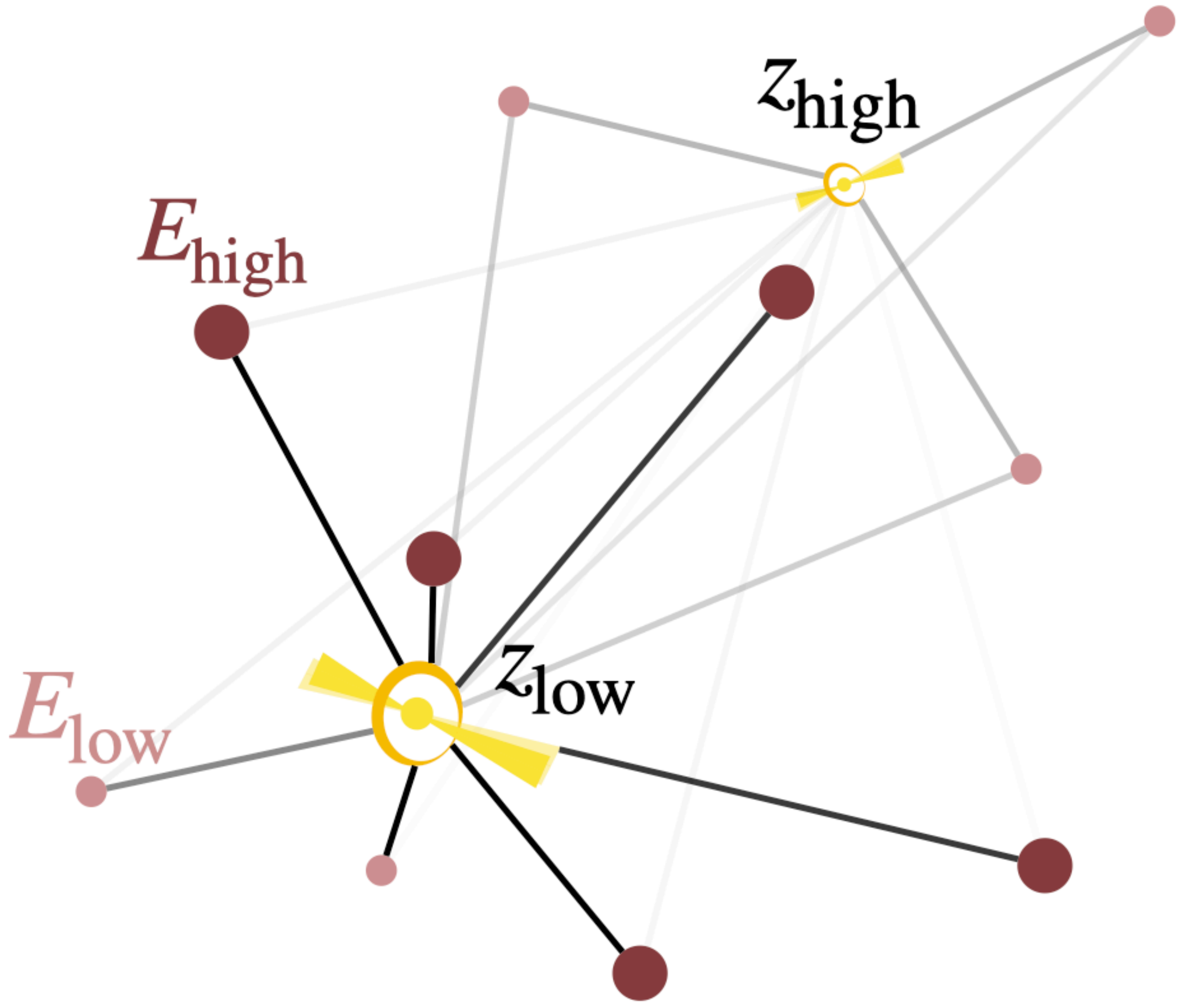


Assuming only protons

UHECR SOURCE SEARCHES

Bayesian hierarchical model

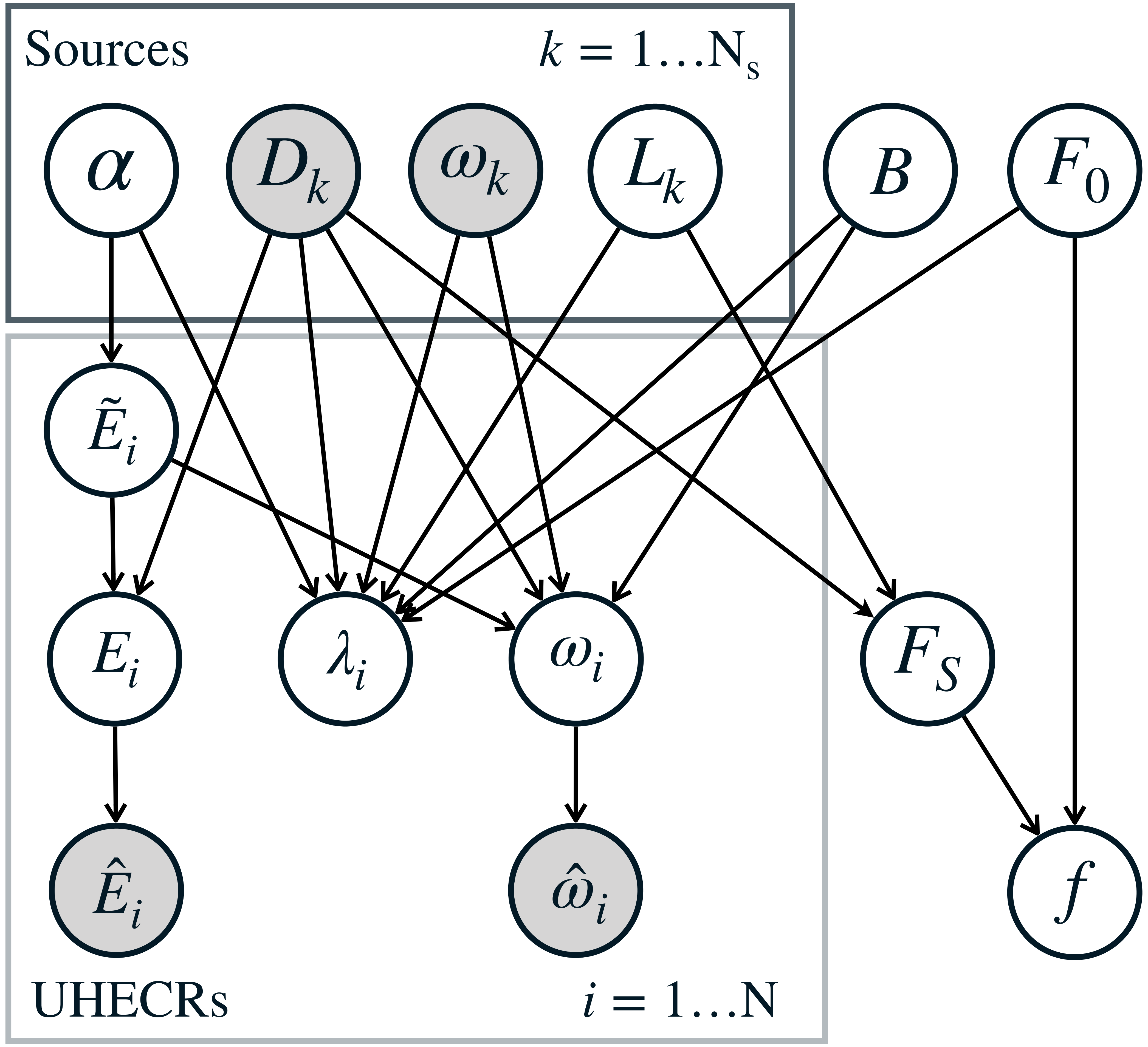
- Joint fit of spectrum and arrival directions
- **~100s free parameters**, including **physical hyperparameters**
- Results in terms of **individual source-UHECR** associations
- **Unphysical associations removed**



$P(\text{associated} \mid \text{data})$

Implemented via **Hamiltonian Monte Carlo** in **Stan**

Betancourt et al. 2017
Carpenter et al. 2017



Assuming only protons

UHECR SOURCE SEARCHES

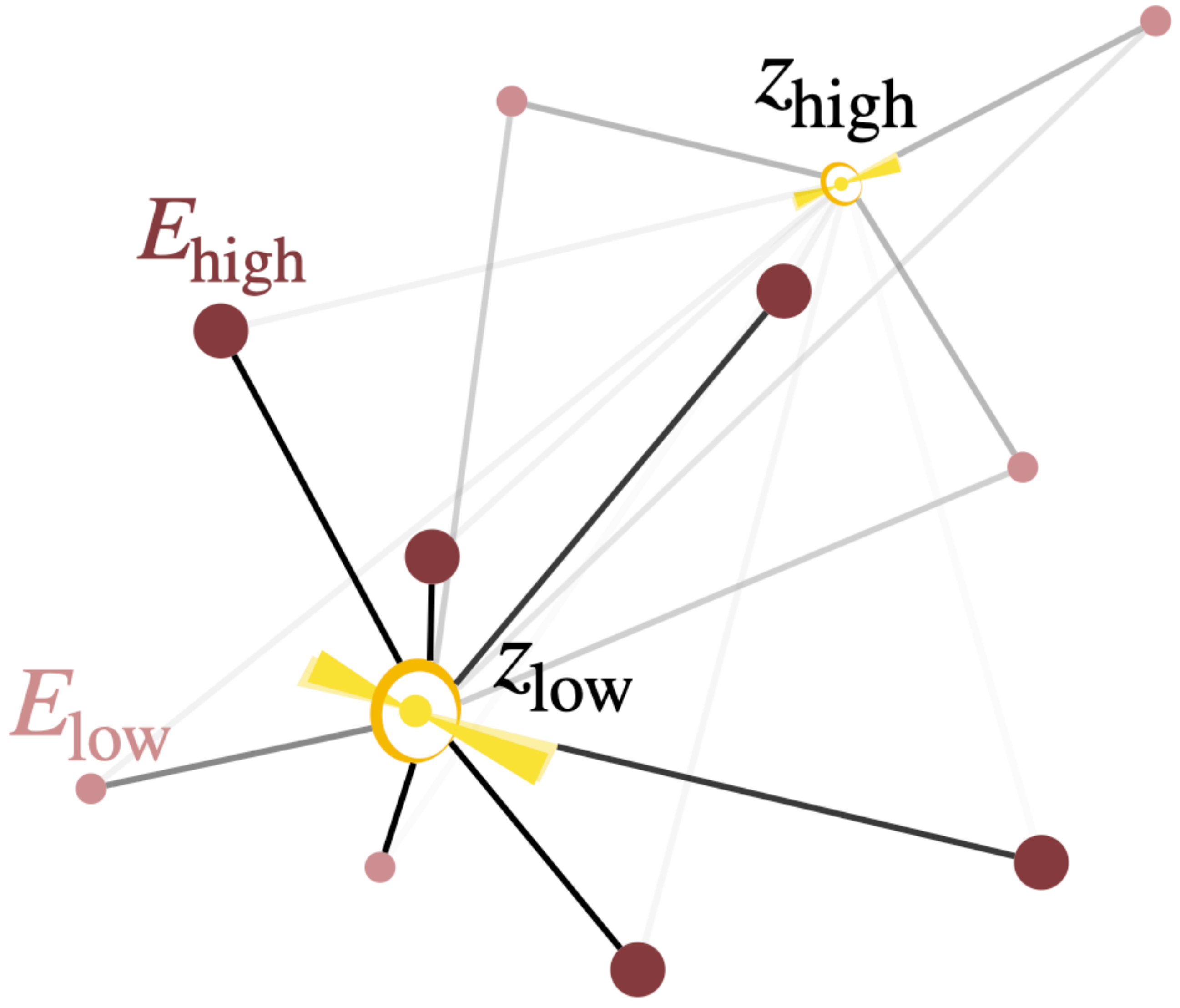
Bayesian hierarchical model

Joint fit of spectrum and arrival directions

~100s free parameters, including **physical hyperparameters**

Results in terms of **individual source-UHECR associations**

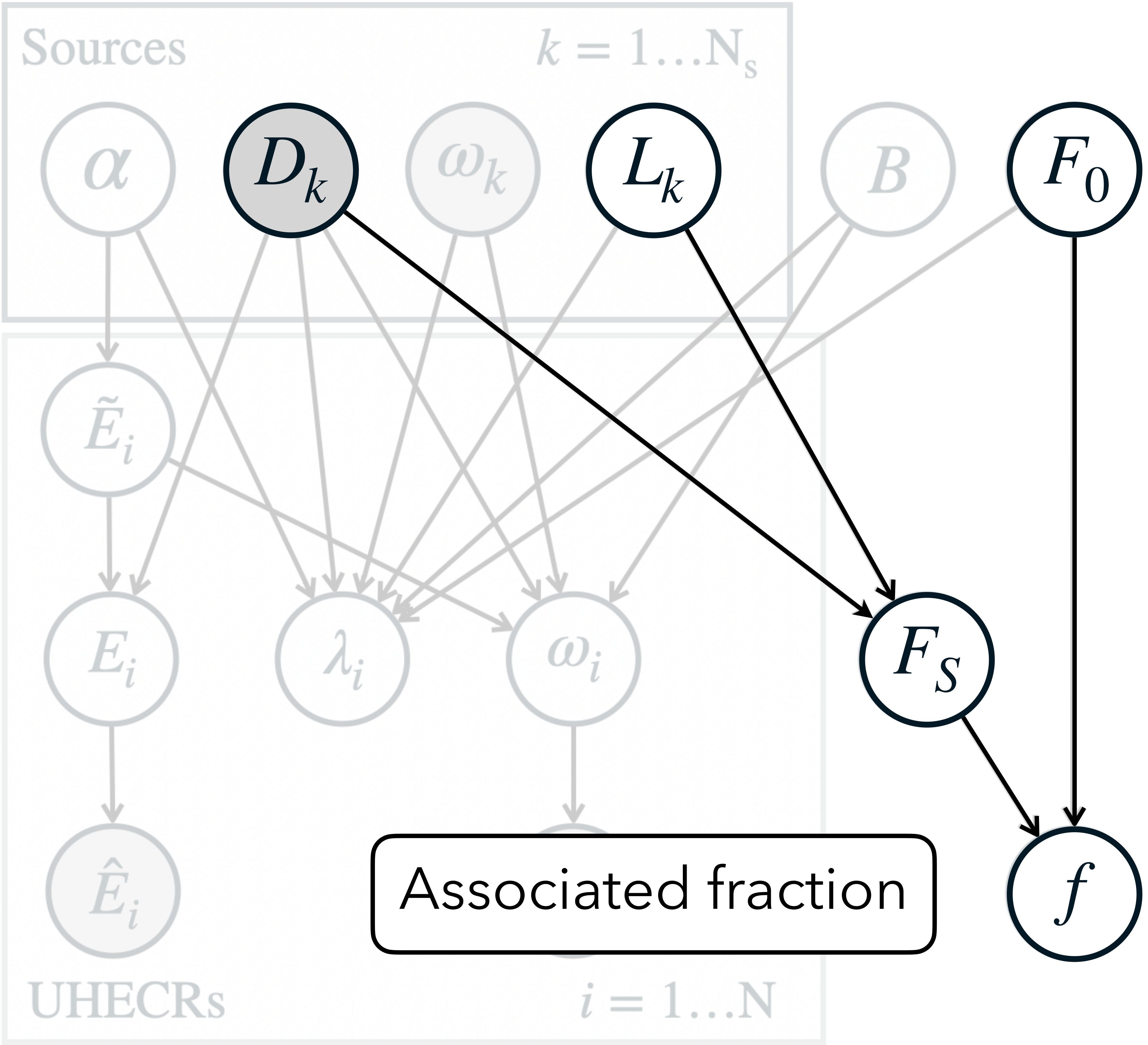
Unphysical associations removed



$P(\text{associated} \,|\, \text{data})$

Implemented via **Hamiltonian Monte Carlo** in **Stan**

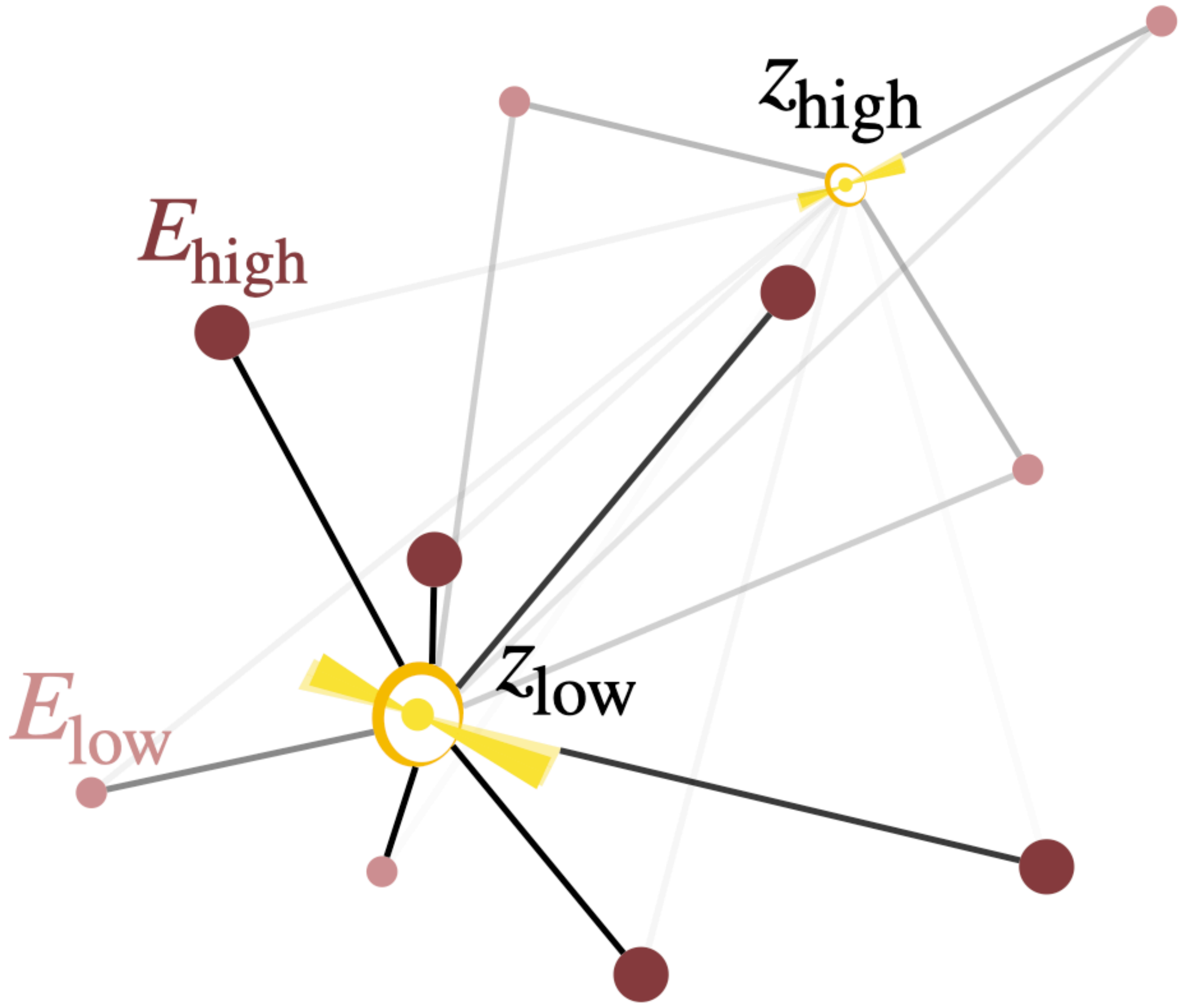
Betancourt et al. 2017
Carpenter et al. 2017



UHECR SOURCE SEARCHES

Bayesian hierarchical model

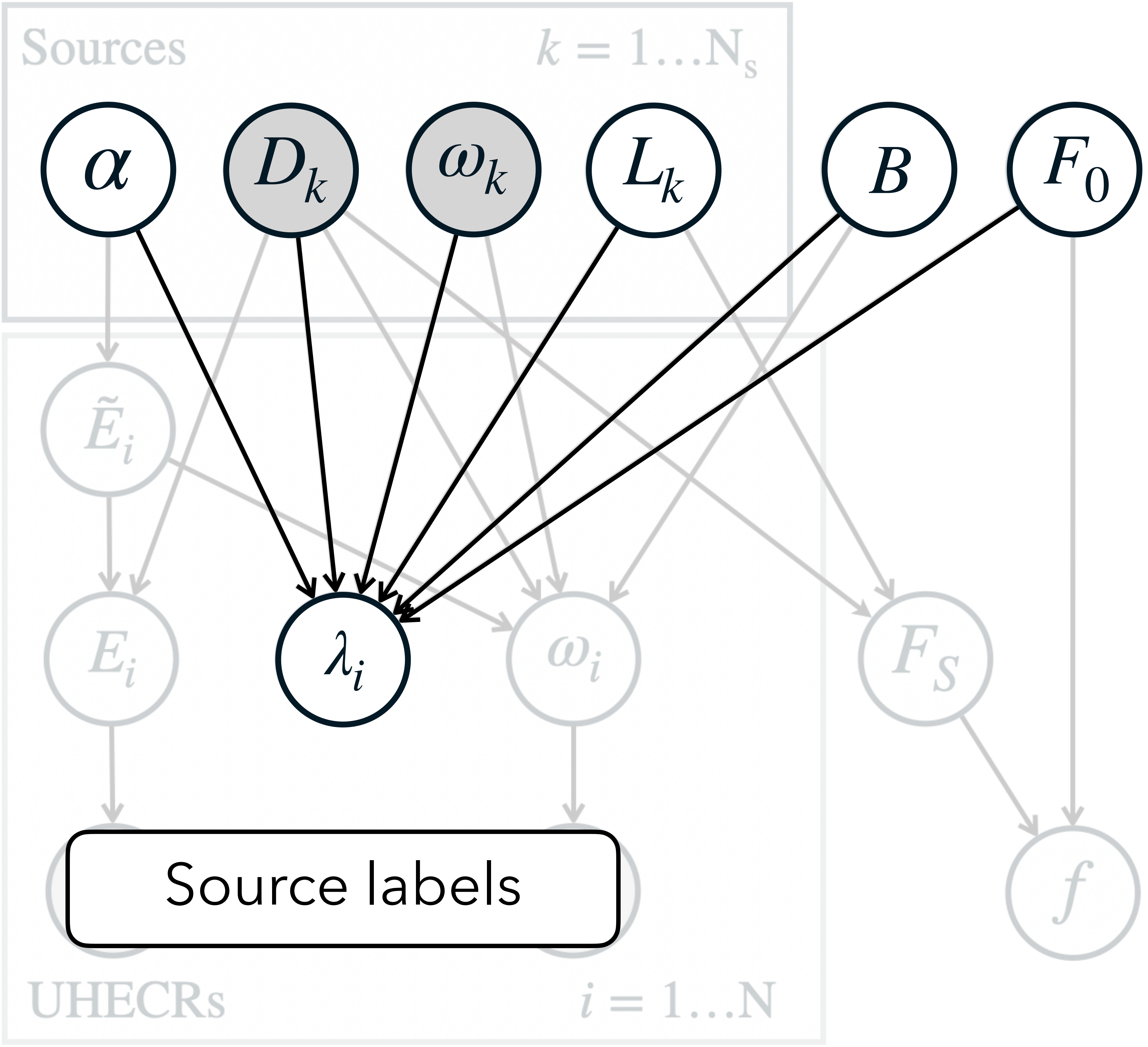
- Joint fit of spectrum and arrival directions
- ~100s **free parameters**, including **physical hyperparameters**
- Results in terms of **individual source-UHECR** associations
- **Unphysical associations removed**



$P(\text{associated} \mid \text{data})$

Implemented via **Hamiltonian Monte Carlo** in **Stan**

Betancourt et al. 2017
Carpenter et al. 2017



UHECR SOURCE SEARCHES

Application to starburst galaxy catalog & Auger 2014 public data

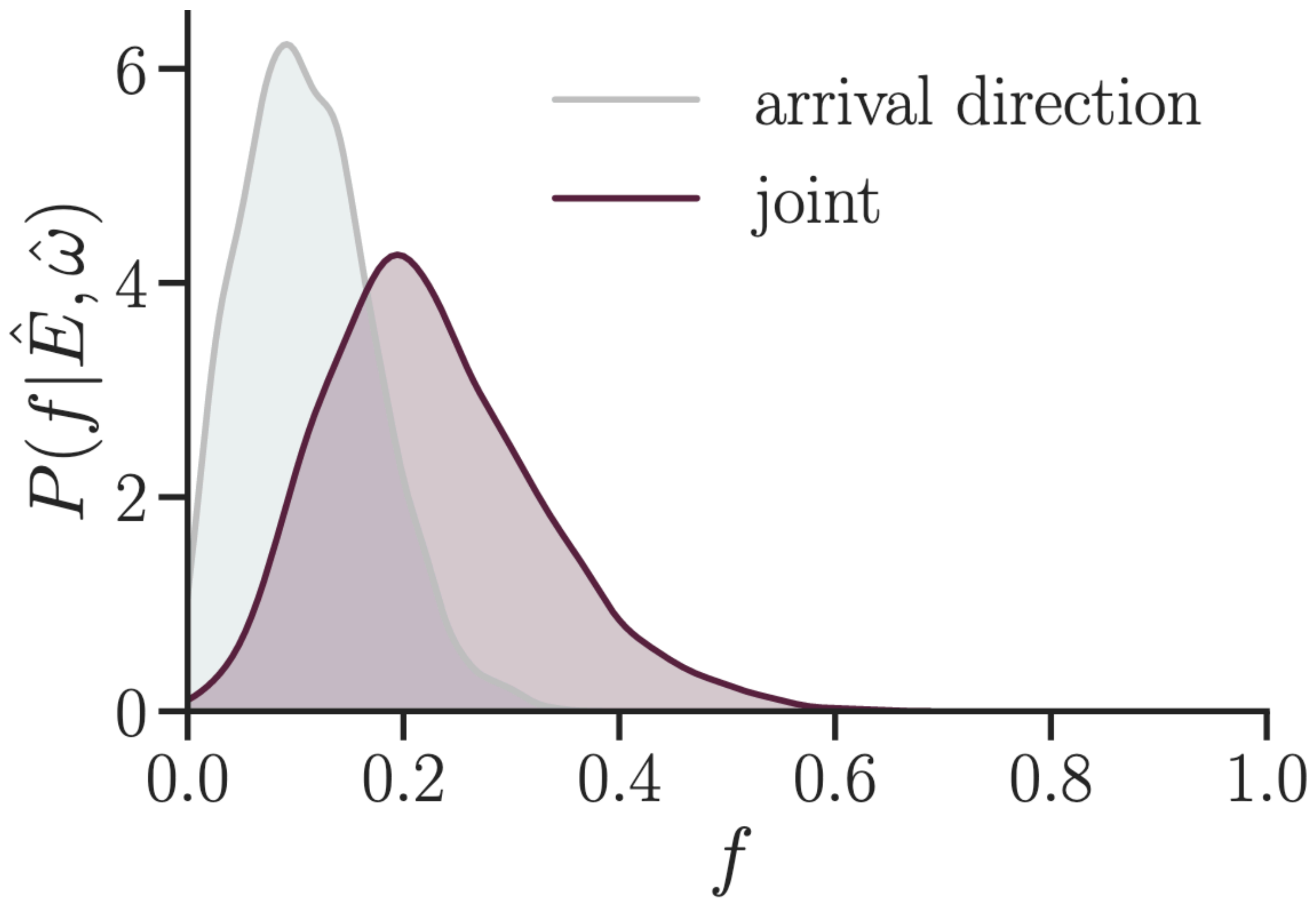
Including **energy-dependent deflections** allows for more possible associations

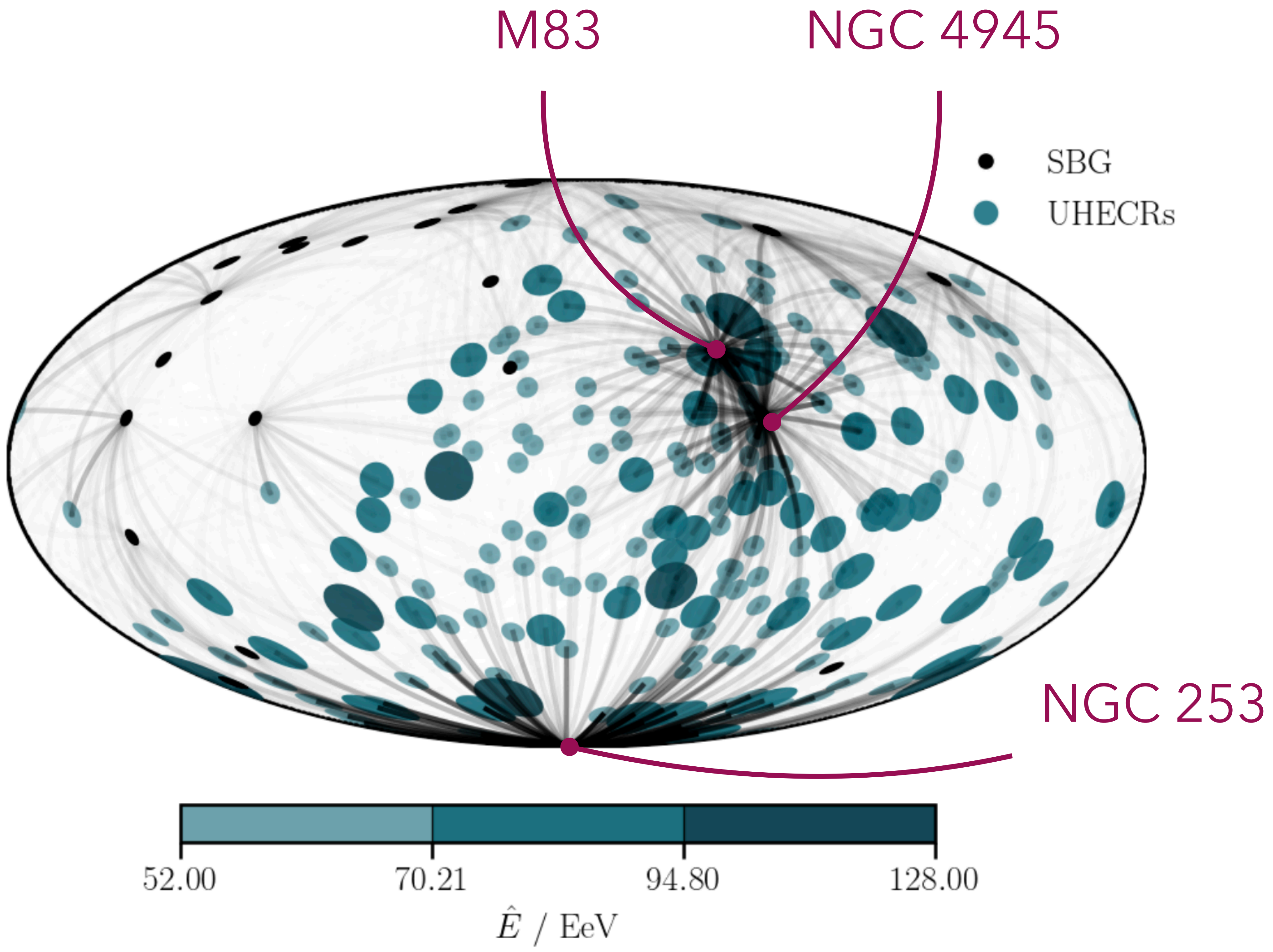
Uncertainties on the energies has a large impact on the horizon

Highest energy UHECRs lack obvious associations

Assumptions

- Only protons
- Continuous energy losses
- Extra-Galactic B field is a random Gaussian field
- Galactic B field not considered





Capel & Mortlock 2019, https://github.com/cescalara/uhecr_model

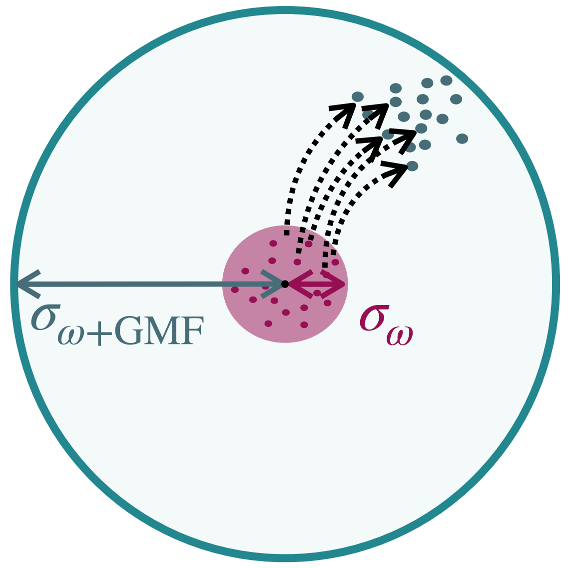
Auger data release: Aab et al. 2015

UHECR SOURCE SEARCHES

Extension to including UHECR composition, Galactic B field and TA 2014 public data

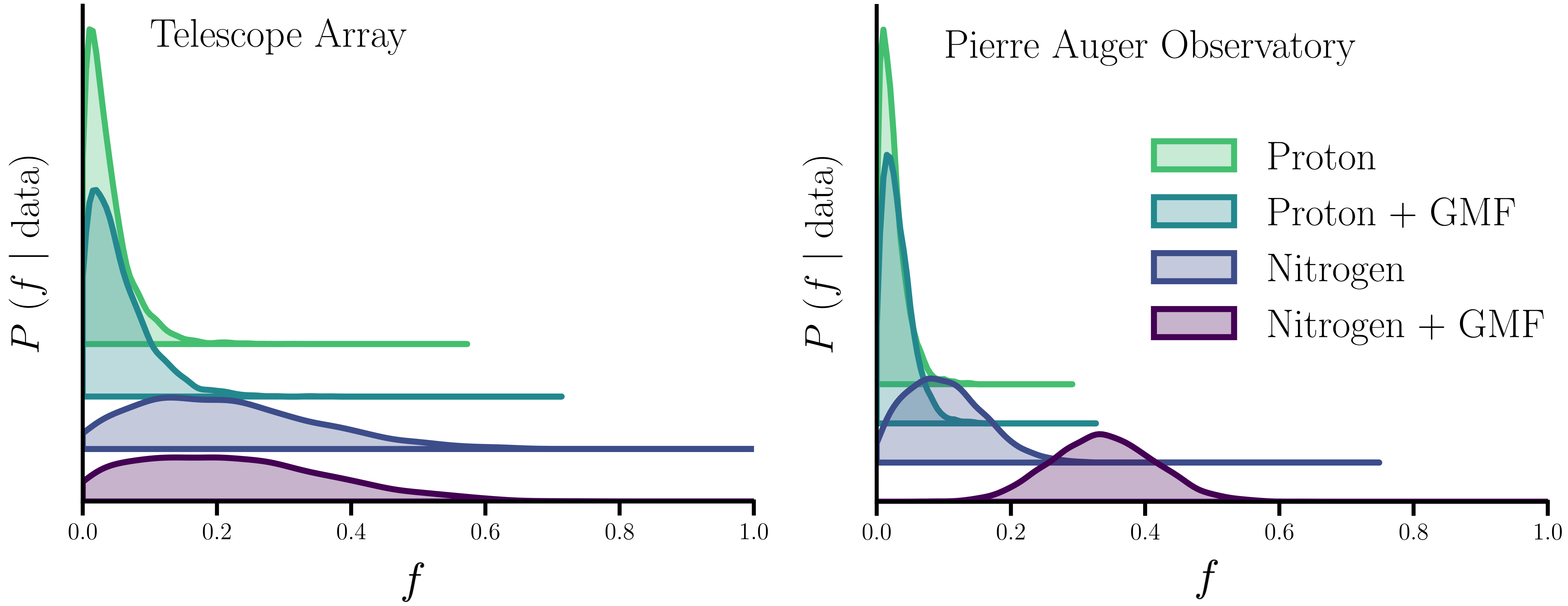
Assume fixed composition at sources, e.g. Nitrogen

Galactic deflections are approximated by backtracking with CRPropa 3 (JF12 model)



Narrow prior on Extragalactic B

Energy losses approximated via loss lengths as implemented in CRPropa3



Work in progress: Watanabe, Fedynitch, Capel & Sagawa, UHECR 2022
CRPropa3: Batista et al. 2022, TA data release: Abbasi et al. 2014

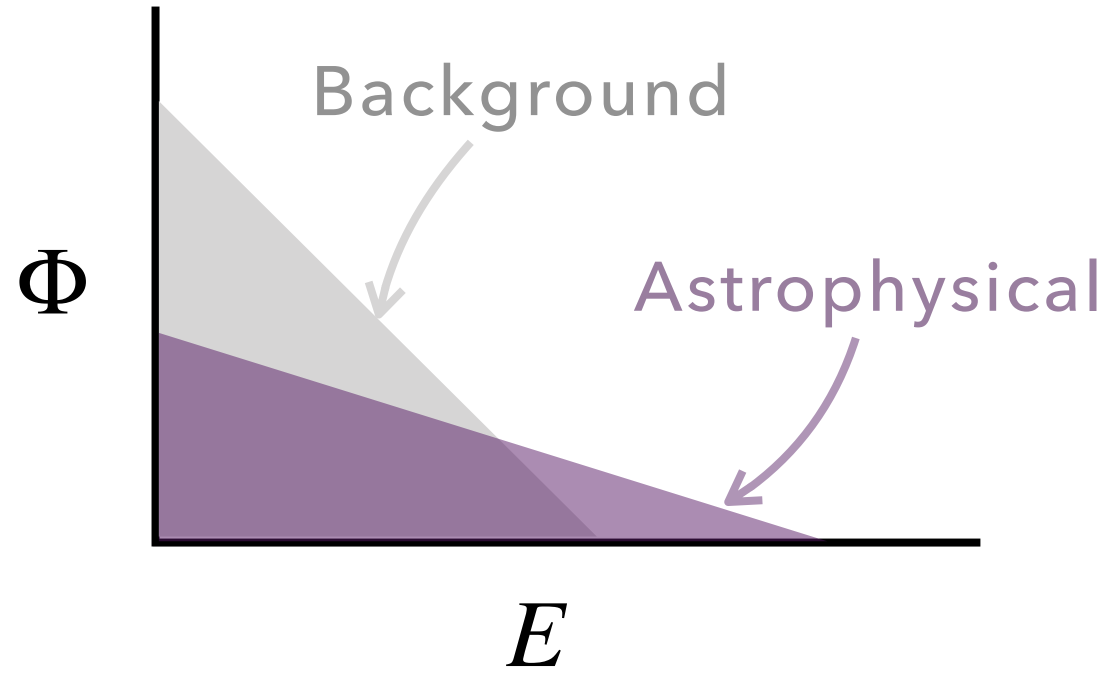
See also combined fit of directions, energy, composition: Bister et al. (Auger), UHECR 2022

To include more realistic propagation: Likelihood-free or emulator-based methods necessary

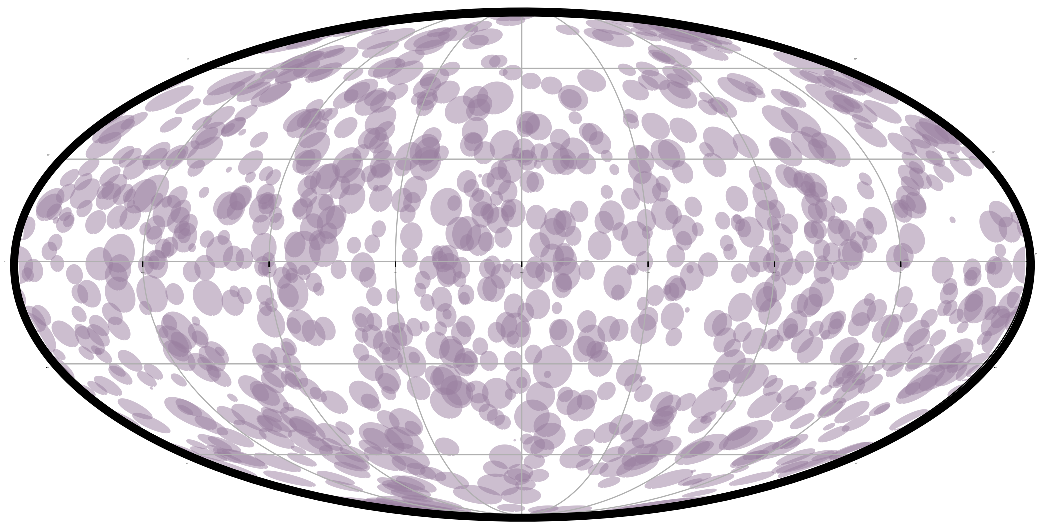
NEUTRINO SOURCE SEARCHES

Focus: Complementary information from individual sources and population, leveraging multi-messenger information

IceCube neutrino observations

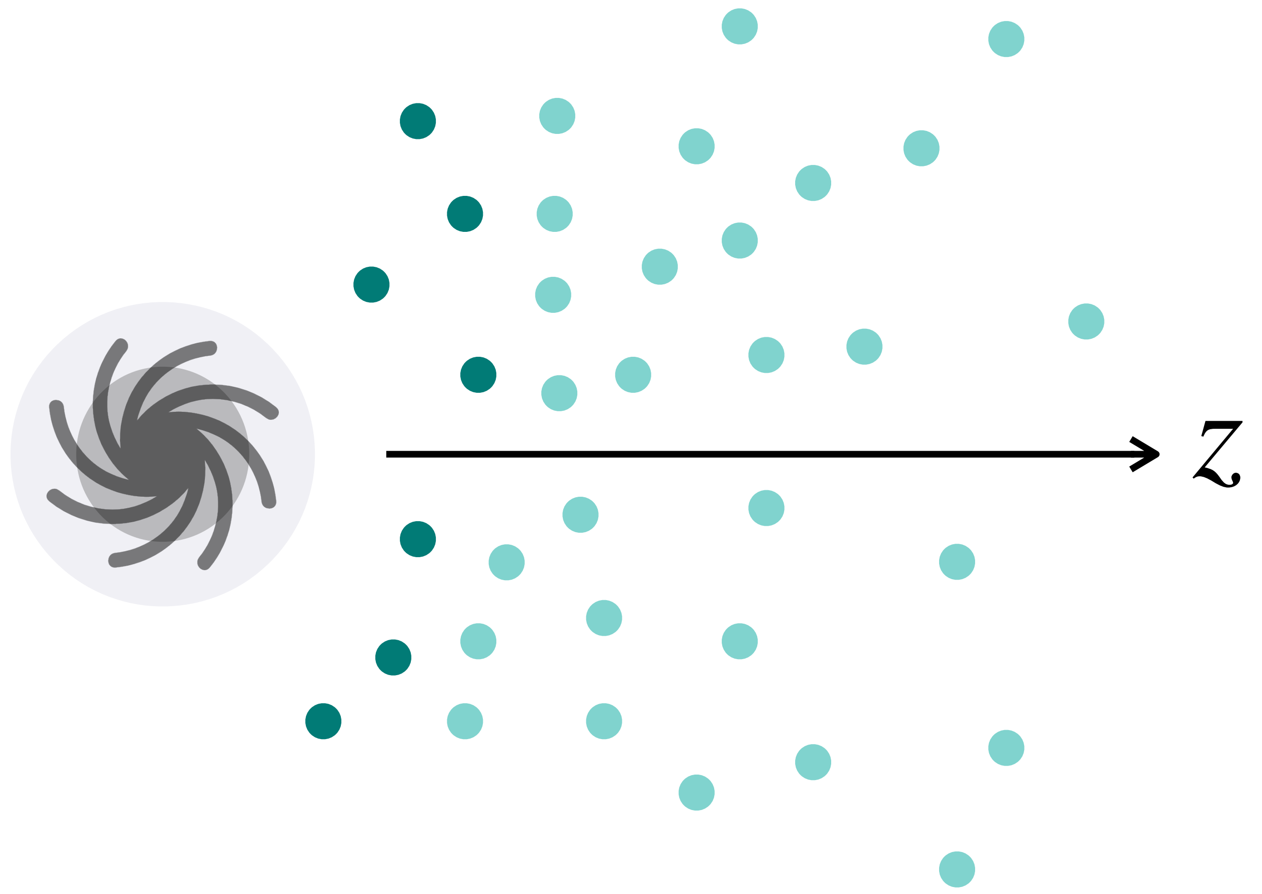


An astrophysical flux...



...but no/few obvious point sources

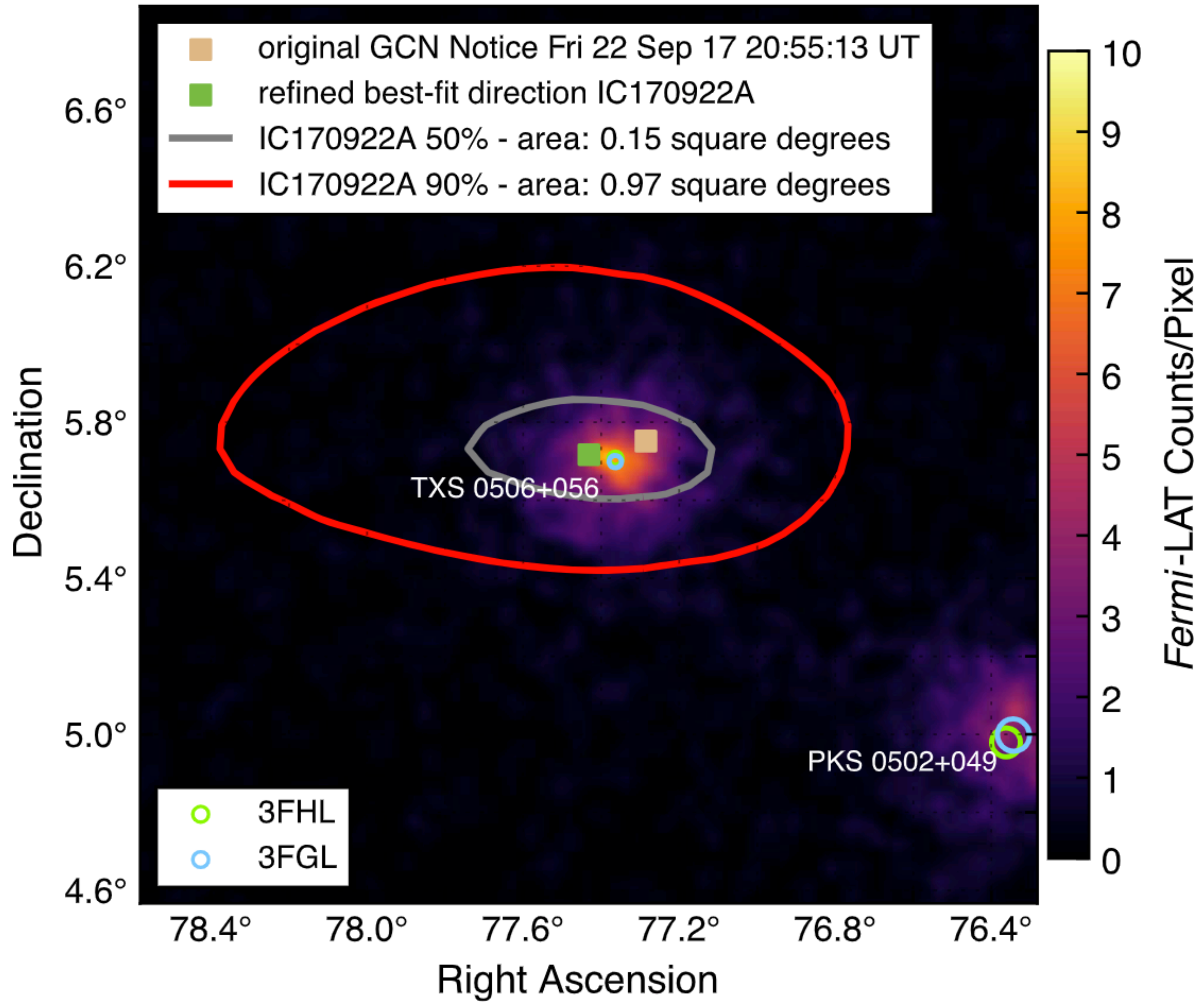
Physical picture



Extragalactic sources characterised by a **density, luminosity and cosmological evolution**

NEUTRINO SOURCE SEARCHES

Example: Blazar–neutrino connection



IceCube Collaboration et al. (2018)

TXS 0506+056 and IceCube 170922A

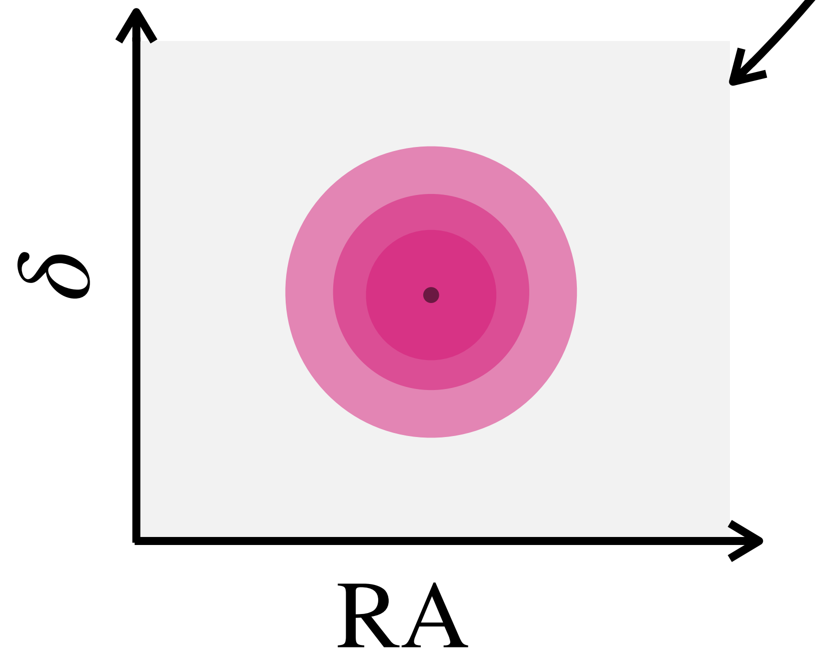
- "Signalness" of ν is ~ 0.6
- Blazars are relatively common
- Blazar flare duration of ~ 6 months
- **3σ statistical significance**
- ν-γ connection is still unclear

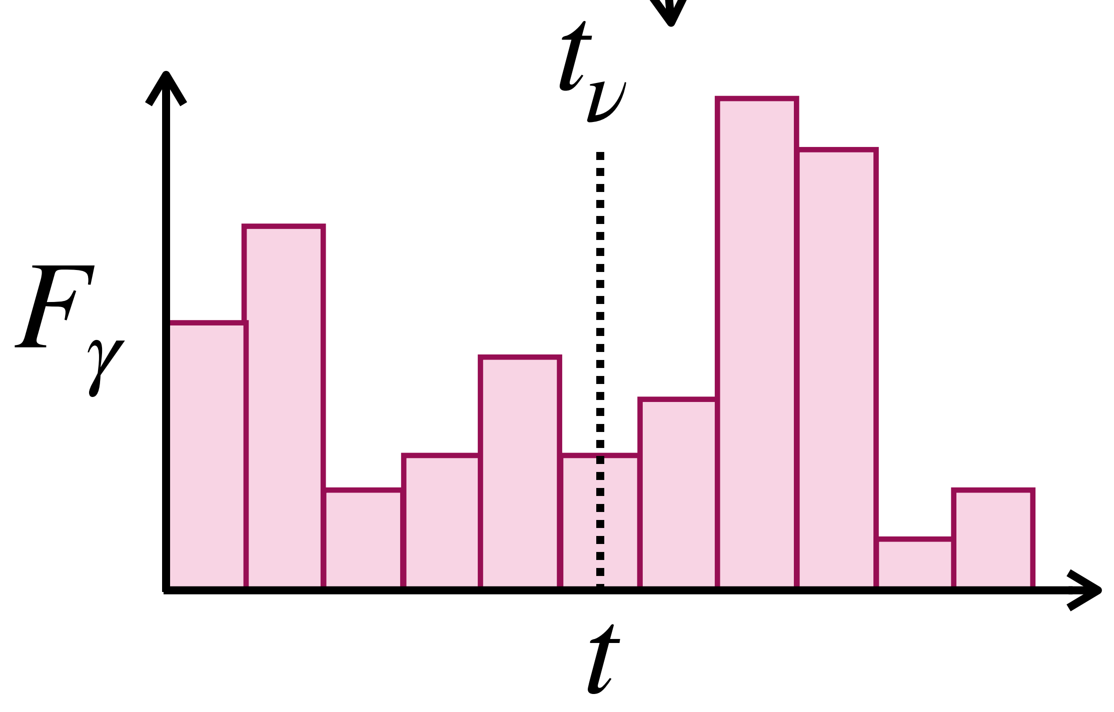
NEUTRINO SOURCE SEARCHES

One neutrino event tested with **~2300 extragalactic γ-ray sources** seen by Fermi-LAT

$$\mathscr{L} = n_s \, p(\nu \,|\, \text{source}) \; + (1 - n_s) \; p(\nu \,|\, \text{background})$$

$$p(\nu \,|\, \text{source}) = \sum_{i=1}^{N_{\text{src}}} p(\vec{x}_\nu \,|\, \vec{x}_i) w_i(t_\nu) w_{\text{acc}}(\vec{x}_i)$$





Key assumptions for ~3σ result

1. The neutrino flux is proportional to the γ-ray flux

$$F^\nu \propto F^\gamma \qquad\qquad F^\nu = Y_{\nu\gamma} F^\gamma$$

2. All sources have the same neutrino spectrum

IceCube Collaboration et al. (2018)

NEUTRINO SOURCE SEARCHES

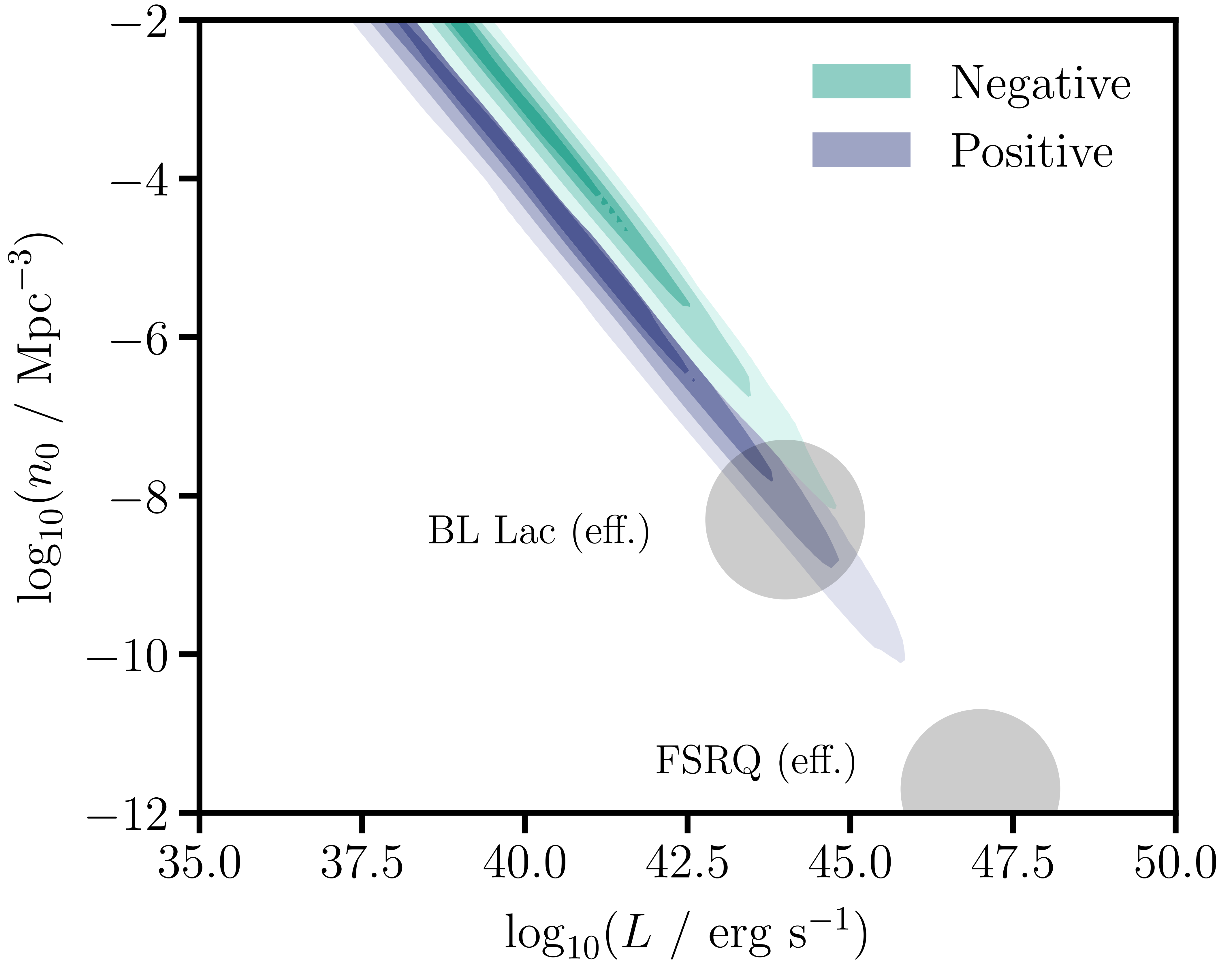
If blazars are the main neutrino sources:

- They must be numerous and powerful enough to produce the observed astrophysical flux
- They cannot be too rare or bright, as then point sources would be detected

We used a Bayesian hierarchical model to find the constraints on the **density** and **luminosity** of a **general population** neutrino sources

- TXS 0506+056 is either a BL Lac or FSRQ blazar (e.g. Padovani et al. 2019)
- In both cases sources are **strongly constrained**

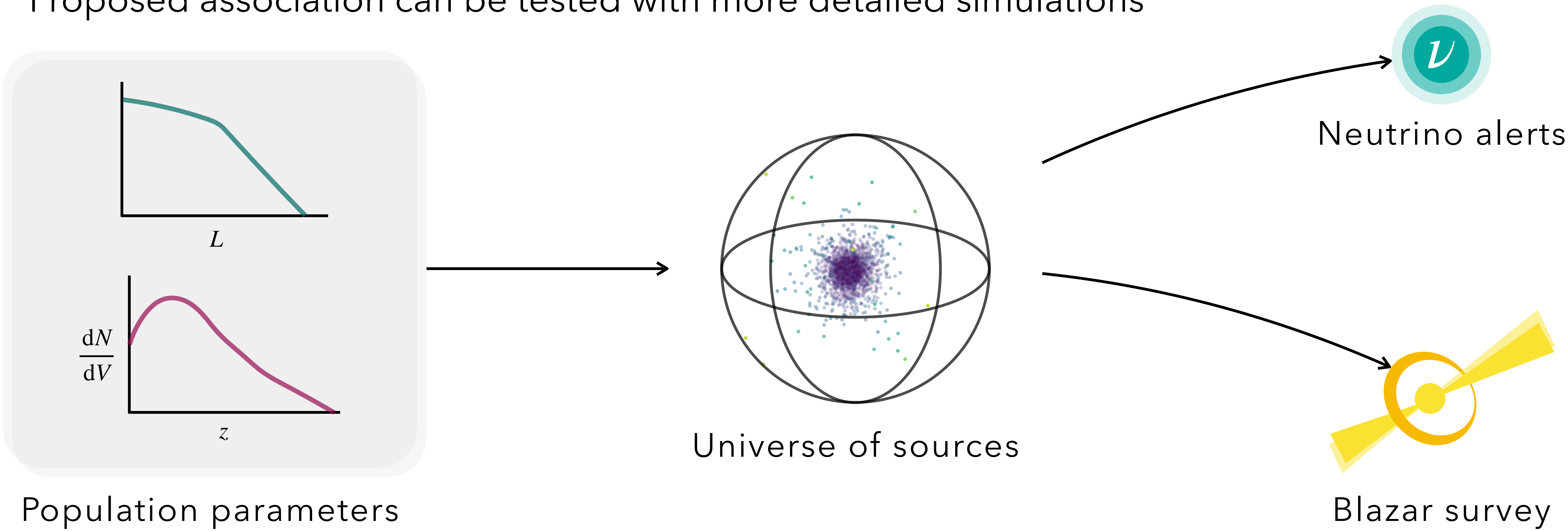
Capel, Mortlock & Finley 2020



See also: Lipari et al. 2008, Silvestri & Barwick 2010, Ahlers & Halzen 2014, Kowalski 2015, Murase & Waxman 2016, Palladino et al. 2020

NEUTRINO SOURCE SEARCHES

Proposed association can be tested with more detailed simulations



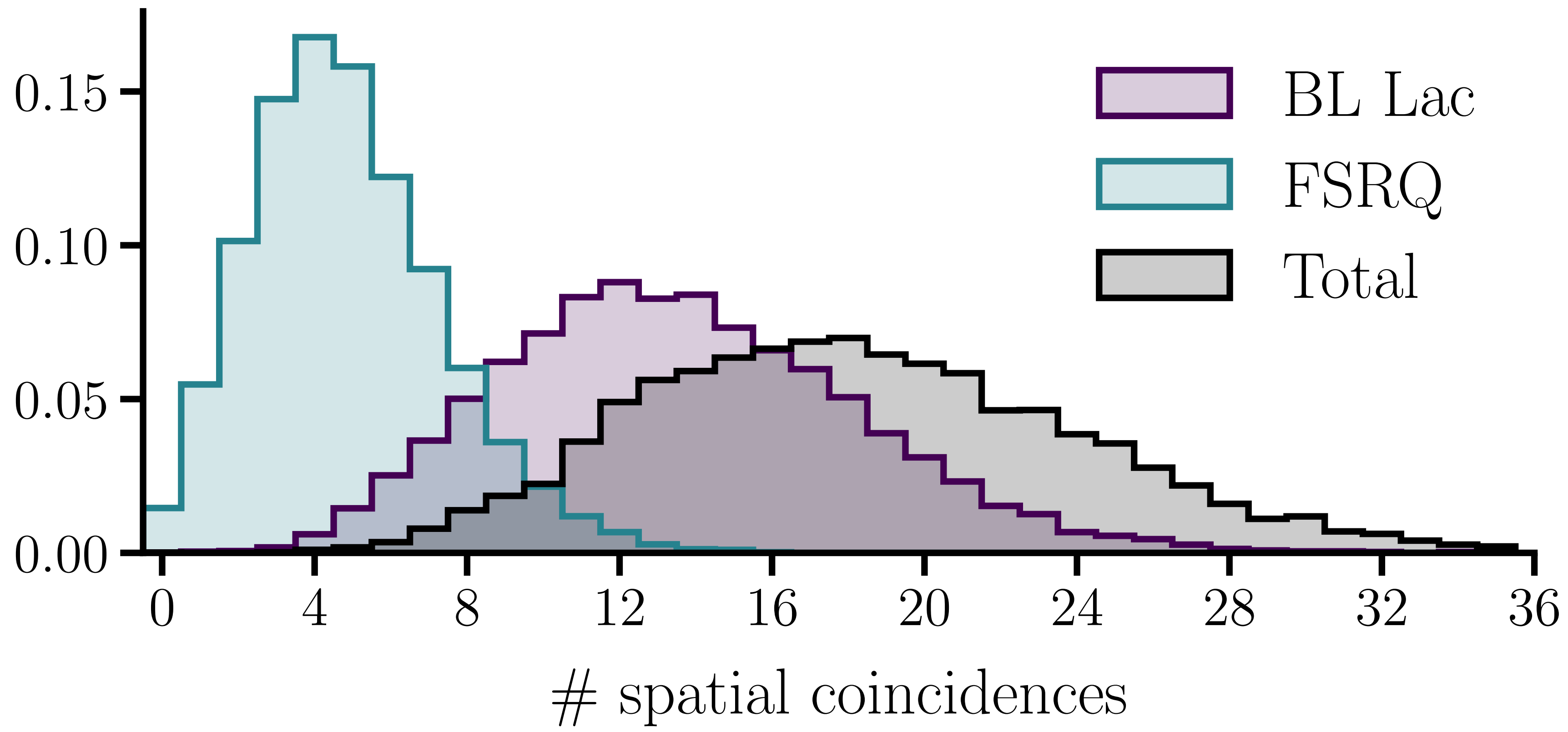
Our "reference model" can reproduce the Fermi 4FGL, FAVA and IceCube alert catalog

Burgess & Capel 2021, https://github.com/grburgess/popsynth

Abdollahi et al. 2020, Abdollahi et al. 2017, Aartsen et al. 2018.

NEUTRINO SOURCE SEARCHES

Assuming **no blazar–neutrino connection**, how often do chance coincidences occur?



We expect ~20 spatial coincidences between neutrino alerts and blazars

This expectation is consistent with observations

7.6% $^{+5.1}_{-3.8}$

Of 10 year surveys have chance coincidences between neutrino alerts and flaring blazars

To find a low chance coincidence of **~0.1% (3σ)** it is necessary to **weight based on the γ-ray flux or other rare source properties**

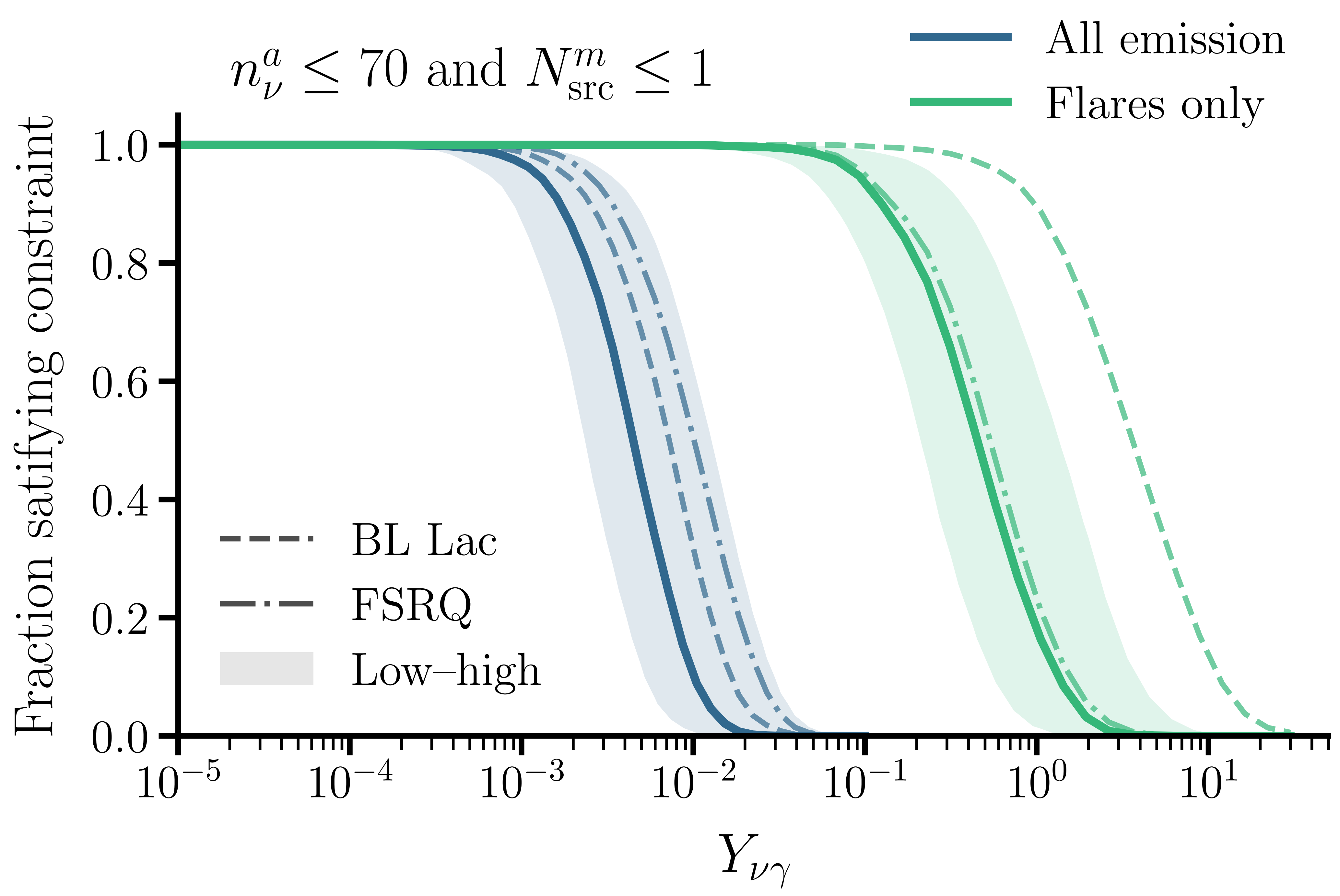
Blazar modelling implies γ-ray–nu connection unlikely

See e.g. earlier talk by Karwin & in this session Rodrigues

Capel, Burgess, Mortlock & Padovani 2022

NEUTRINO SOURCE SEARCHES

Following the assumptions of association likelihood for the blazar–neutrino connection, how many neutrinos should we expect to see from the blazar population?



n_ν^a Number of neutrino alerts

$N_{\rm src}^m$ Number of sources producing more than one neutrino alert

Neutrinos would be overproduced compared to observations, unless:

1. The γ-ray connection is small $\quad Y_{\nu\gamma} \ll 1$
2. Only a small subset of sources contribute

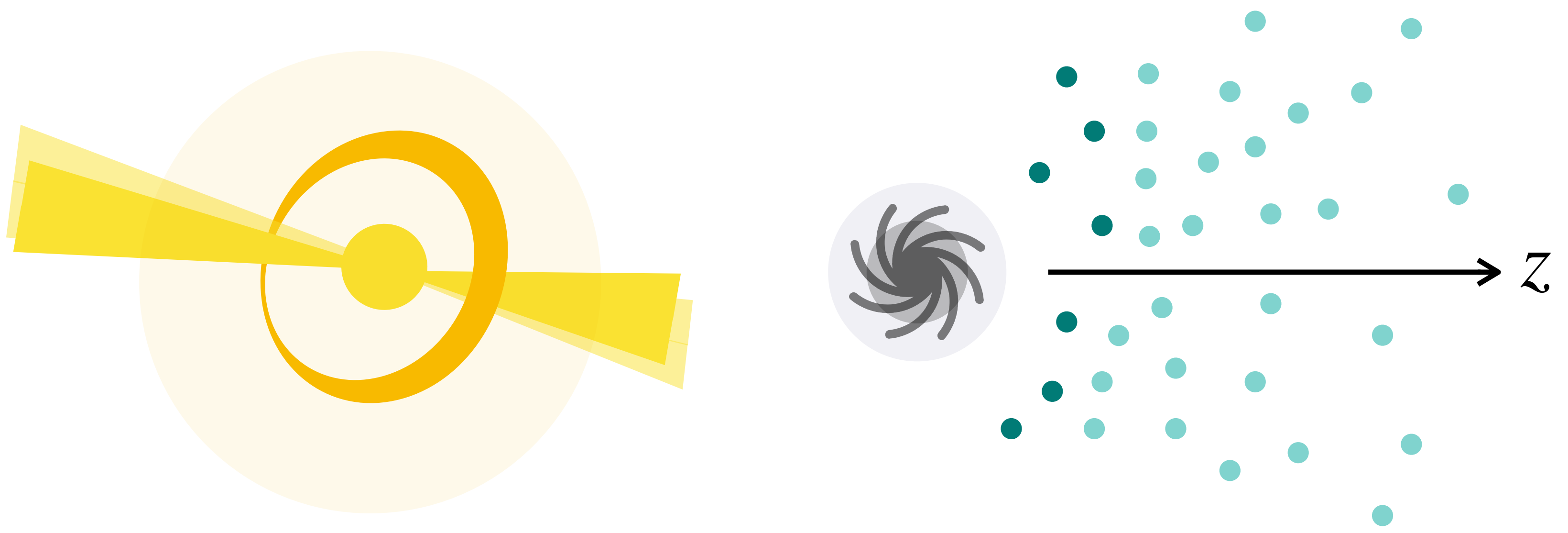
Both of these factors motivate changes to the likelihood used

Capel, Burgess, Mortlock & Padovani 2022

CONCLUSIONS

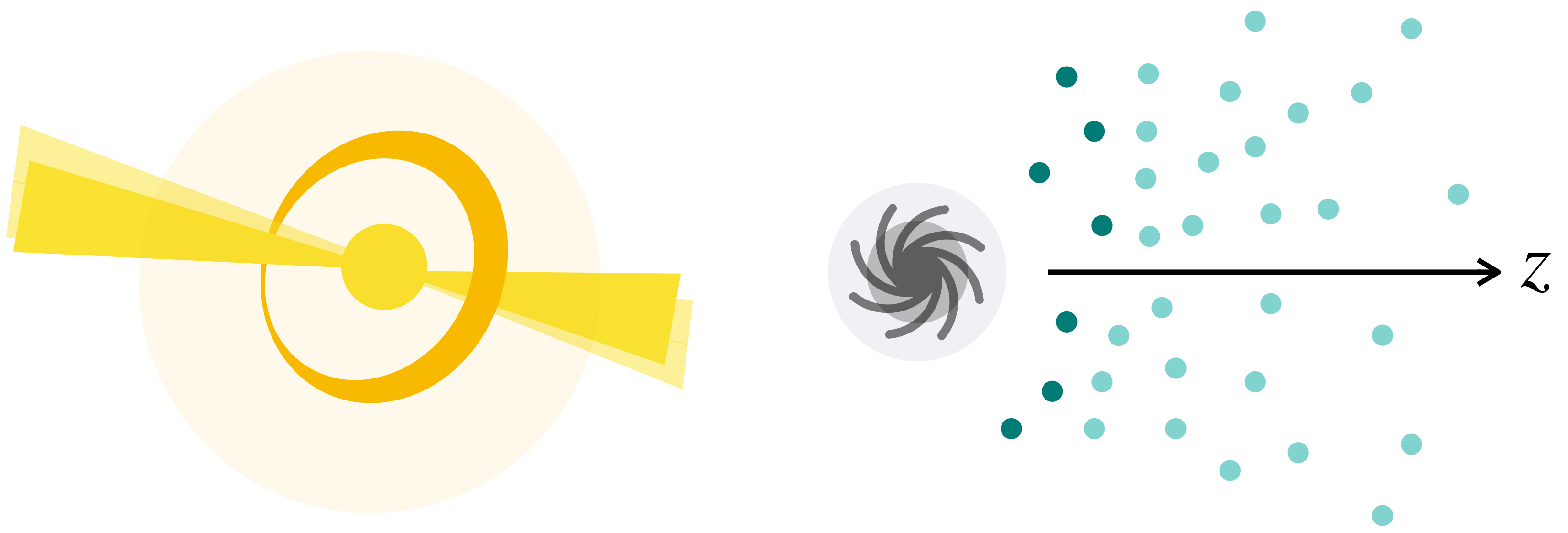
CONCLUSIONS

The **underlying physics**, **source populations**, and **multi-messenger information** can tell us if possible associations make sense



CONCLUSIONS

The **underlying physics**, **source populations**, and **multi-messenger information** can tell us if possible associations make sense



We should bring more of this information into **statistical analyses** that we use to search for sources