Indirect searches for dark matter with photons and neutrinos

iDMEu Kick-off meeting

Christoph Weniger, University of Amsterdam

11 May 2021

Overview

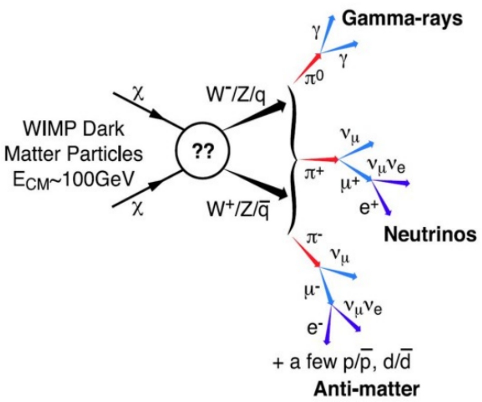
- Dark matter searches with photons and neutrinos
- Challenges in and opportunities for indirect searches
- Conclusions

Disclaimer: This is not a comprehensive overview of the state of the field. Instead, you get my perspective on some of the root causes of pain in our field, which might be related and useful for your field as well.

1) Photons from dark matter

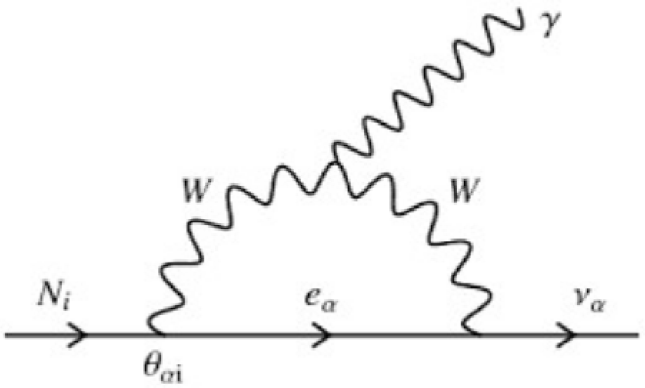
Dark matter self-annihilation

example: WIMPs



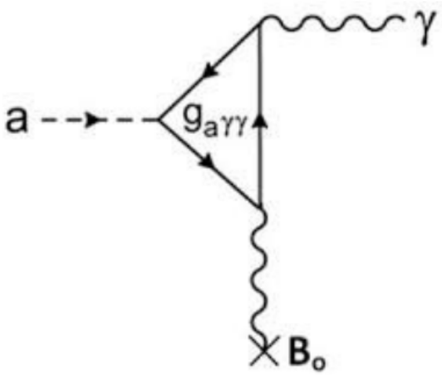
Dark matter decay

example: sterile neutrinos



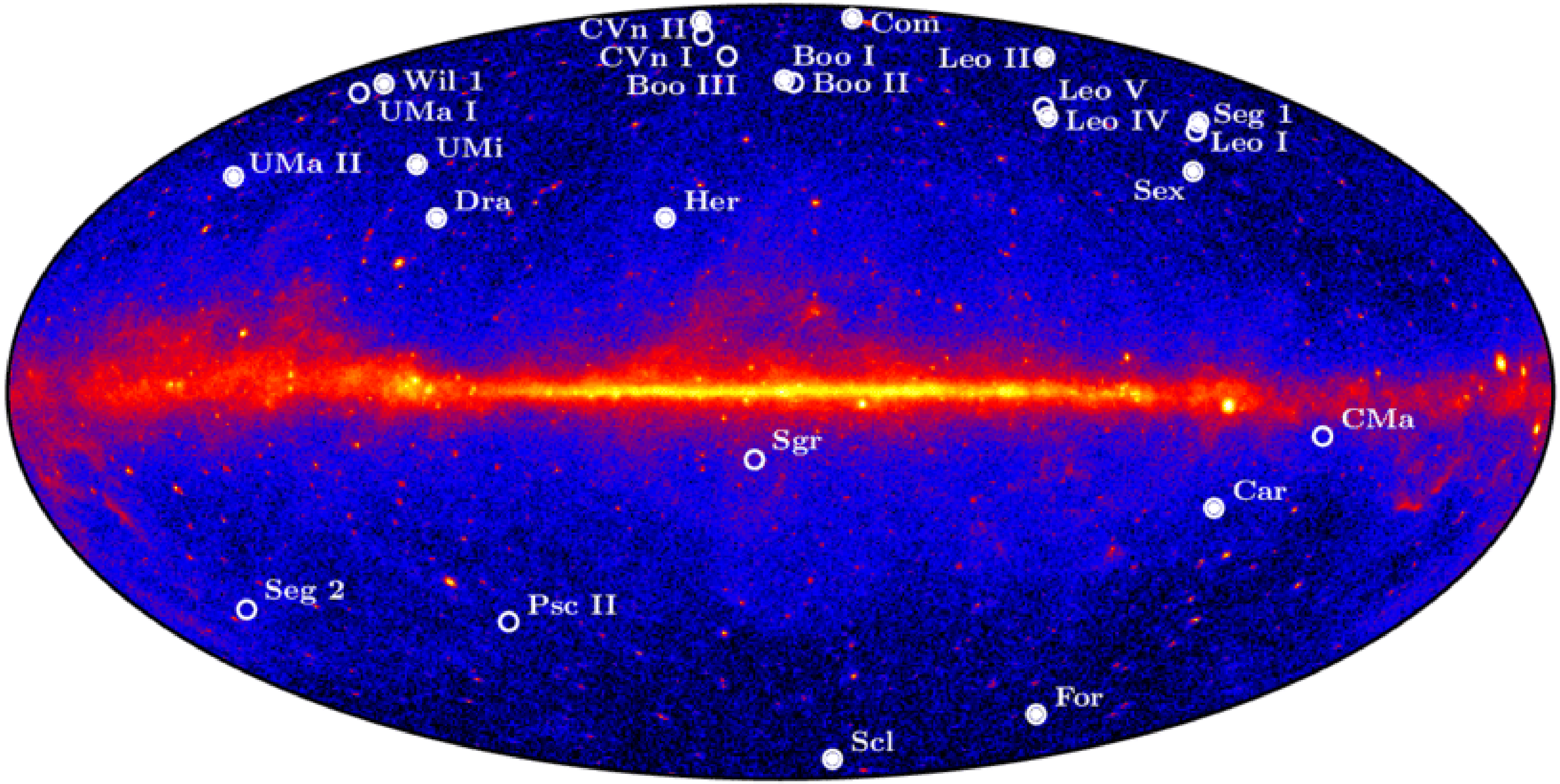
Dark matter oscillation

example: axions



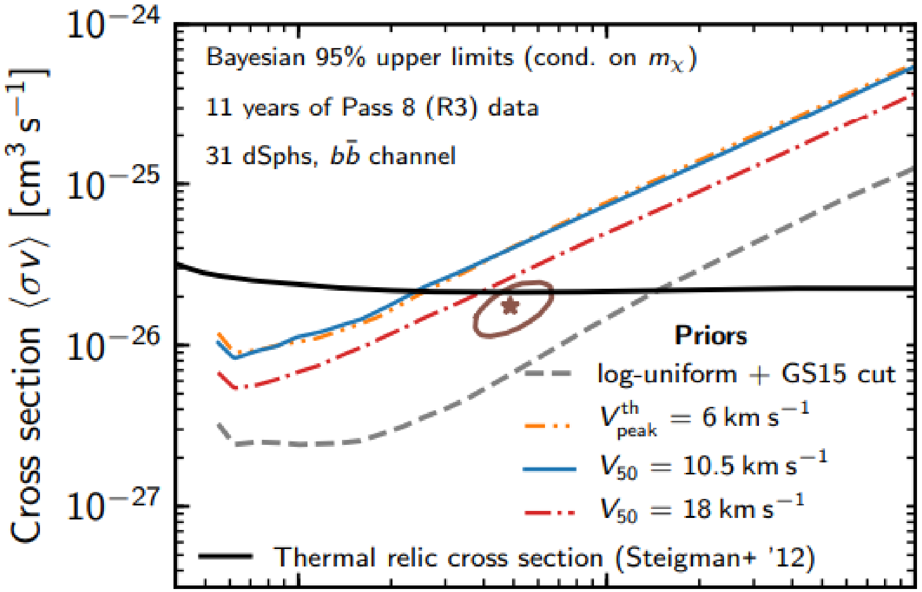
Also: secondary radiation of $e^{\pm}$ upscattering the interstellar radiation field, synchrotron losses, etc

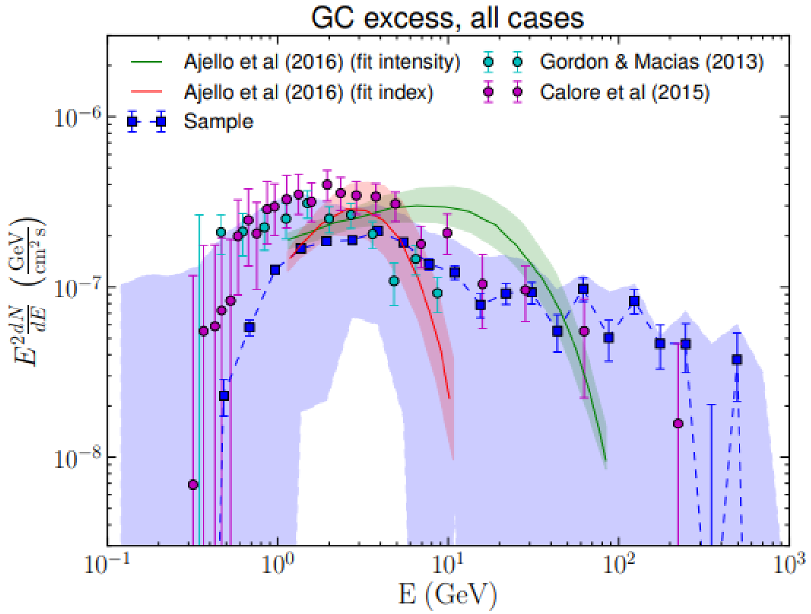
Gamma rays and WIMP searches



Credit: Fermi LAT collaoration

Gamma rays and WIMPs - Recent highlights

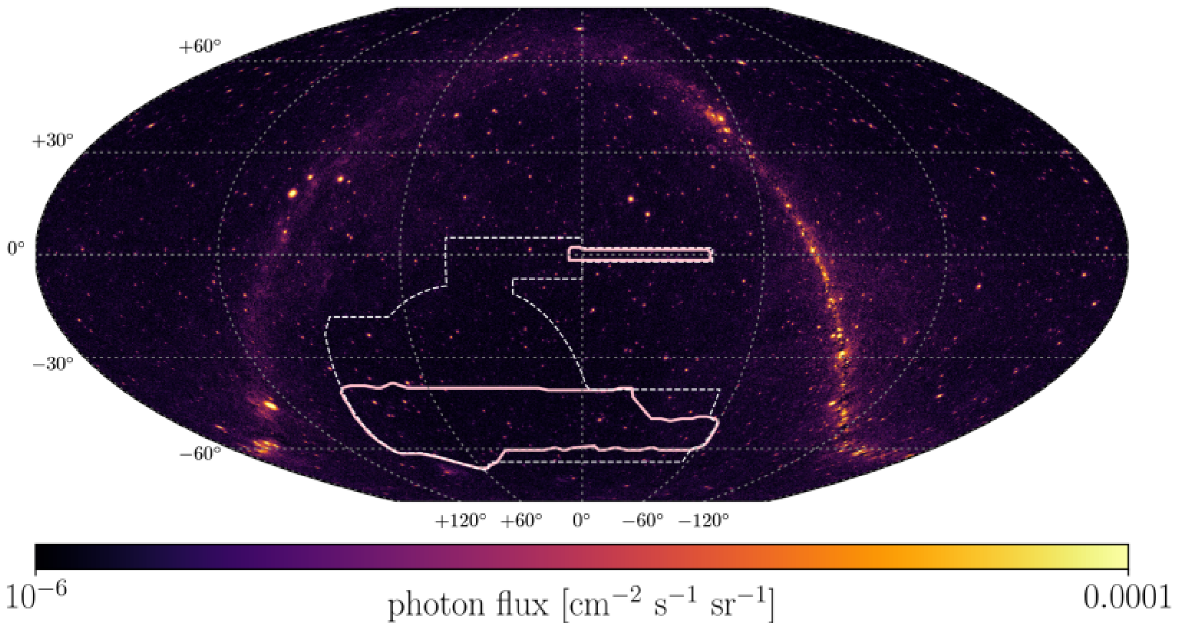




Structure Formation Models Weaken Limits on WIMP Dark Matter from Dwarf Spheroidal Galaxies, Ando+ 2021.

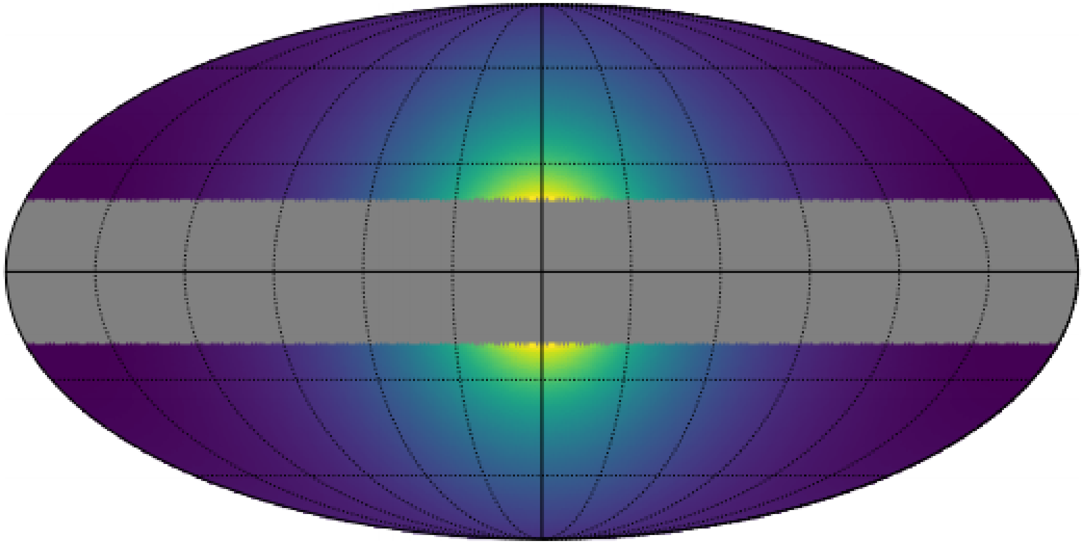
The Fermi Galactic center GeV excess is a candidate for a dark matter signal, since Goodenough & Hooper 2008, here from Fermi coll. 2017.

Future: AMEGO (?), CTA (~2025?), LHAASO (~2021?)



Detection of cross-correlation between gravitational lensing and gamma rays, 1907.13484

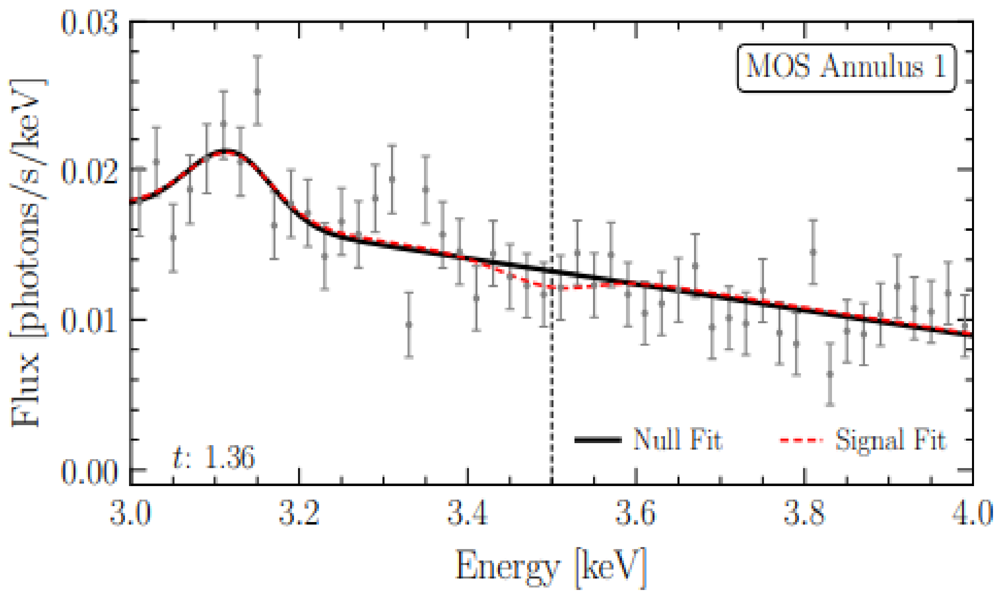
X-ray searches for Sterile neutrino Dark Matter

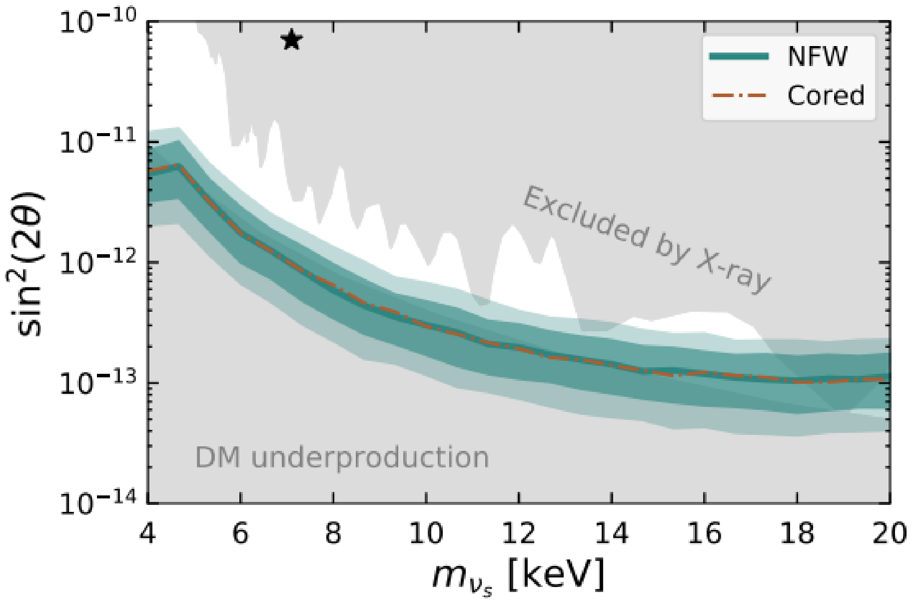


Sterile neutrino signal from the Galactic DM halo (with Galactic disk masked).

Credit: Dekker+ 2021

Sterile neutrinos searches - Recent developments



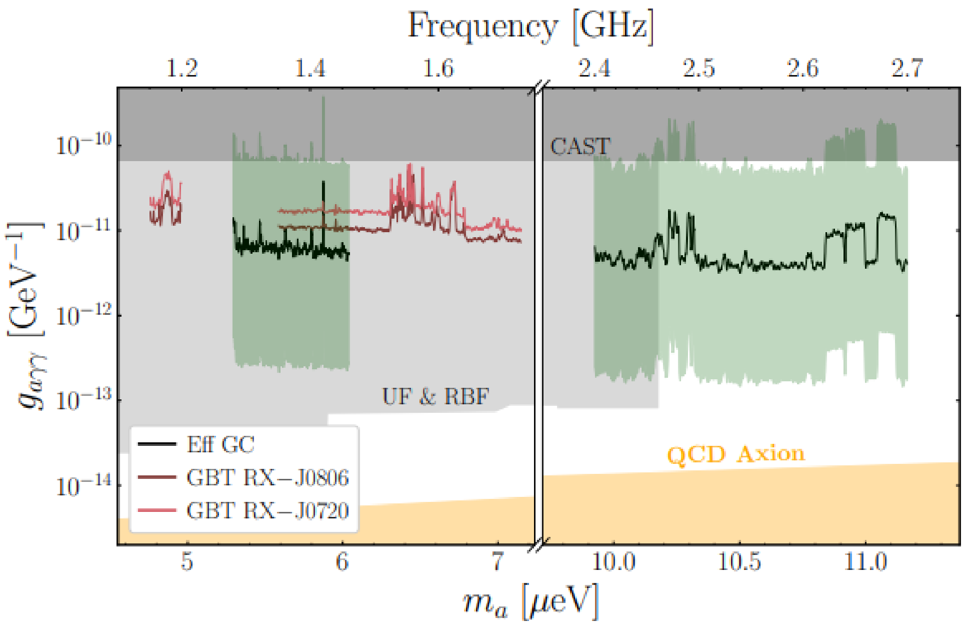


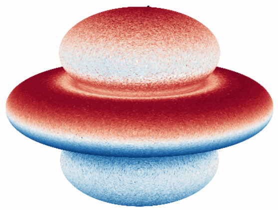
A deep search for decaying dark matter with XMM-Newton **blank-sky observations**, using Gaussian processes, does not find any indications for a 3.5 keV line, Foster+ 2021, 2102.02207.

Prospects, searches for sterile neutrinos and axionlike particles from the Galactic halo with eROSITA, Dekker+ 2021, 2103.13241.

Future: eROSITA, XRISM (2023?), Athena (2031?)

Radio searches for axions - Recent developments



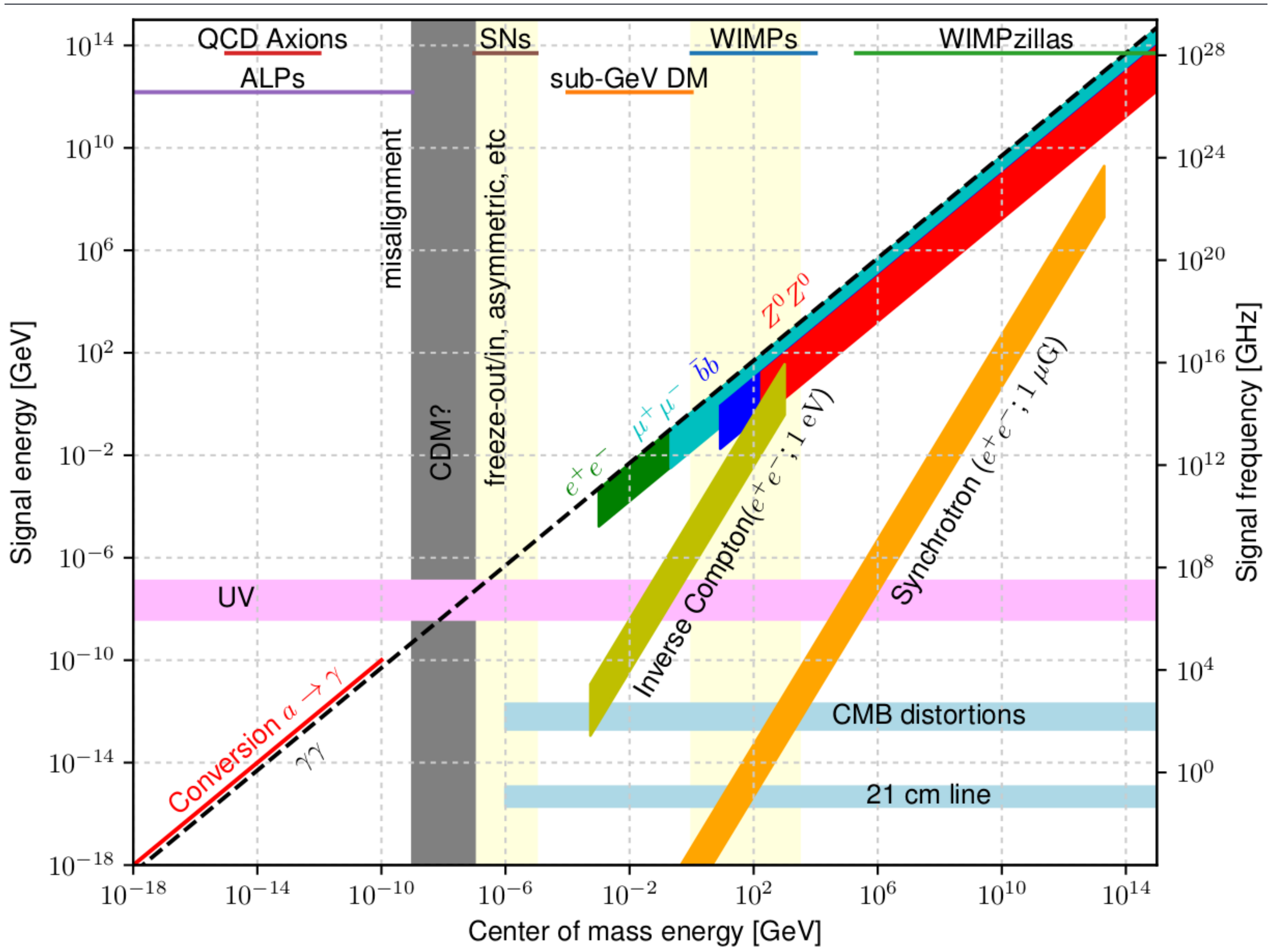


Green Bank and Effelsberg Radio Telescope Searches for Axion Dark Matter Conversion in Neutron Star Magnetospheres, Foster+ 2020

Axion-Photon Conversion in Neutron Star Magnetospheres: The Role of the Plasma in the Goldreich-Julian Model, Witte+ 2021

Future: GBT, Effelsberg, Sardinia, MWA, SKA (~2025?)

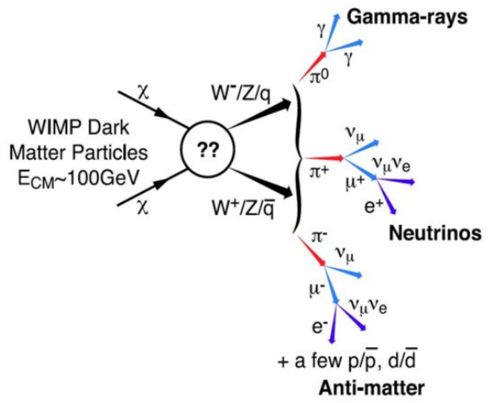
Dark matter mass vs frequency: Outlook

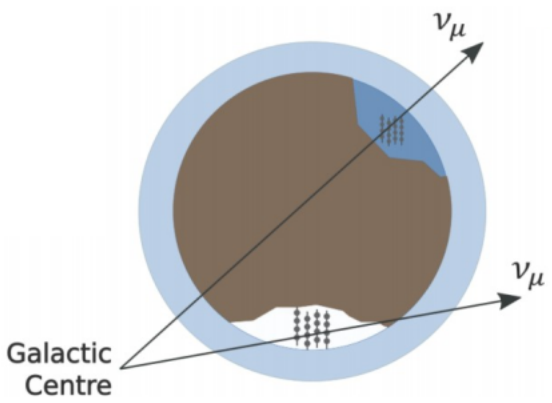


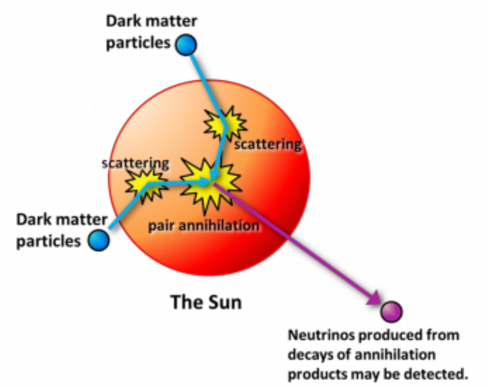
List of upcoming experiments

- LHAASO ~2021
- CTA ~2025
- AMEGO
- XRISM ~2022
- SKA ~2025

2) Neutrinos from dark matter: Mechanisms

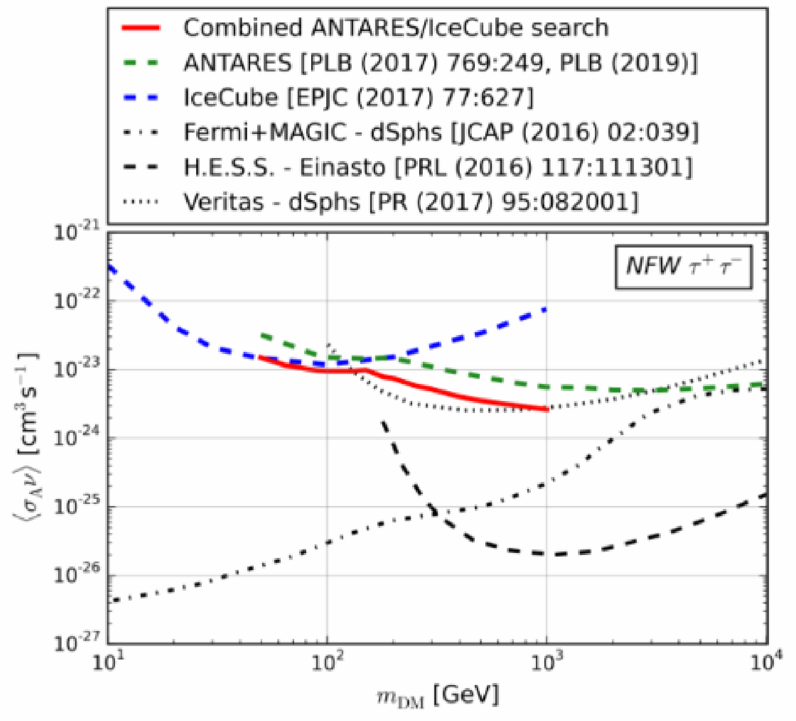


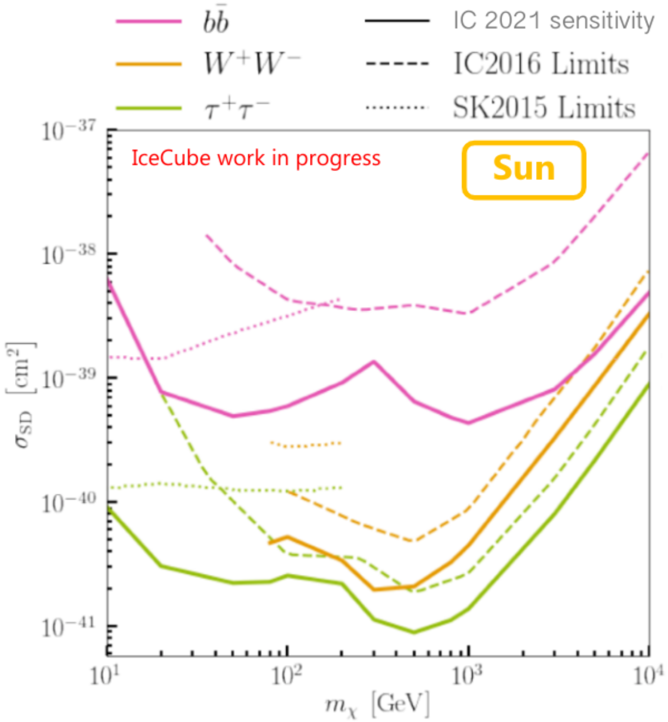




- Neutrinos are produced in the annihilation of WIMPs (mostly through pion decay)
- Using the earth as shield for atmospheric neutrinos
- DM annihilation in the Sun leads to HE neutrino flux

Neutrino searches status



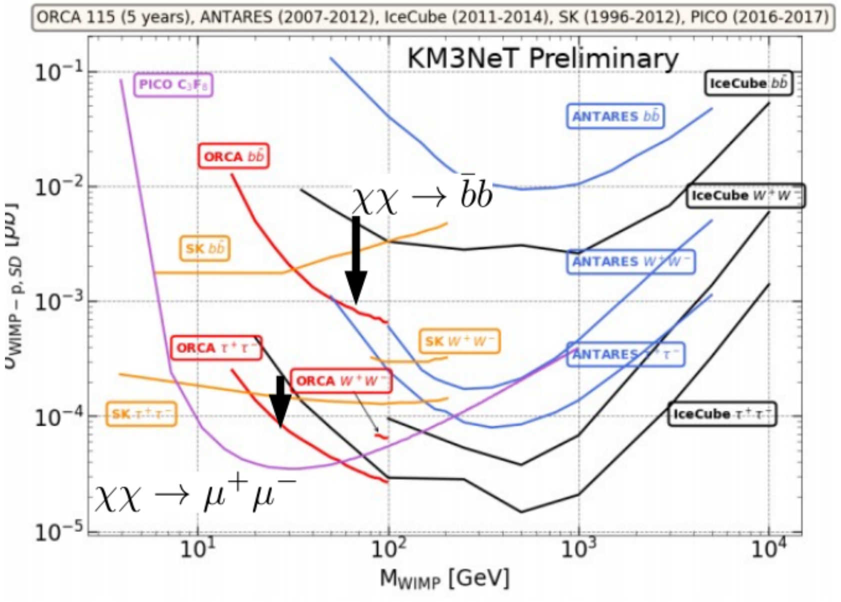


Combined IceCube/ANTARES search for dark matter in the Galactic center, ANTARES, 2003.06614.

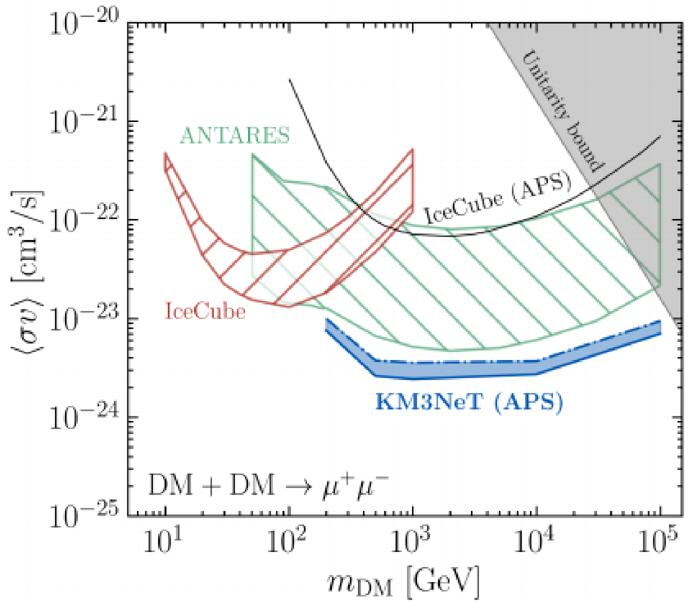
IceCube limits on SD WIMP-nucleon cross-section from observations of the Sun, Stuttard+, INDEES 2021.

Future: IceCube Upgrade, KM3NeT ORCA, IceCube-Gen2, KM3NeT ARCA, IceCube 8yr+ results

Neutrinos: Experimental Outlook



Forcasts for Neutrinos from the Sun limits (ICRC 2019, Gozzini)

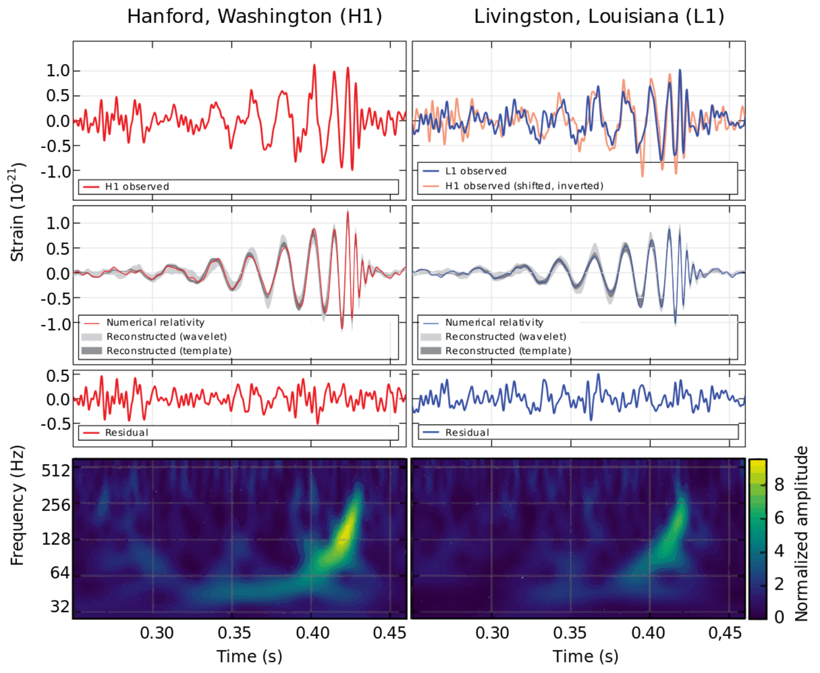


Robust Limits from Upcoming Neutrino Telescopes and Implications on Minimal Dark Matter Models, du Pree 2021

Future: IceCube 8yr+ results, IceCube-Gen2, IceCube 7 string update, IceCube Upgrade, KM3NeT ORCA/ARCA

3) How are discoveries made?

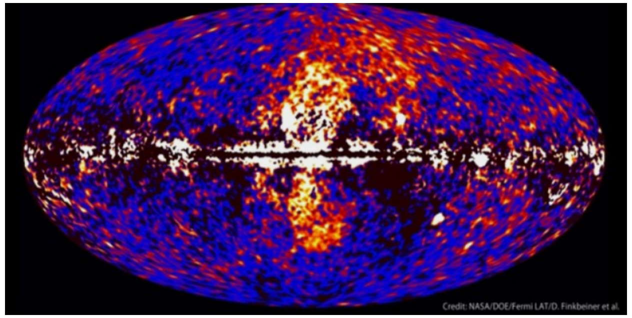
- (D1) **Evidence** against the null hypotehsis
  - The false positive rate should be so low that we almost certainly do not look at statistical noise.
  - The signal should reproduce with more data.
- (D2) **No alternative hypothesis**
  - Exclude known unknowns
  - No instrumental artefact, astrophyiscal backgrounds, etc
- (D3) **Plausibility**
  - Avoid base rate fallacy
  - Account for other null detections
- (D4) **Simplicity**
  - Exclude unknown unknowns
  - Signals "looks like how is should look like"
  - Corroborating evidence from multiple sources.



LIGO measurement of GW150914

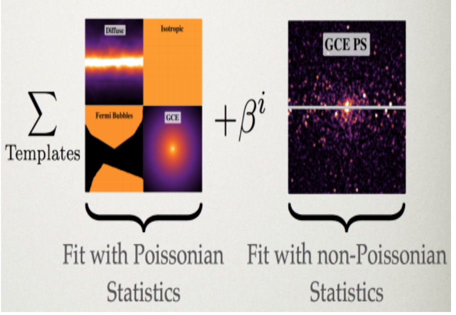
Discussion inspired by: Michele Vallisneri

Many indirect dark matter search strategies



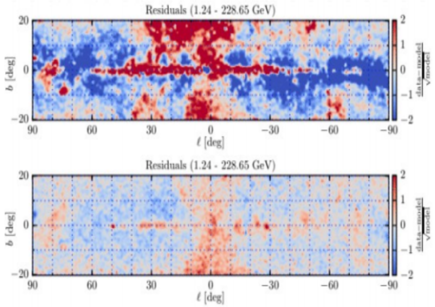
Template fits

Credit: NASA/DOE/FermiLAT/Finkbeiner+, 2010



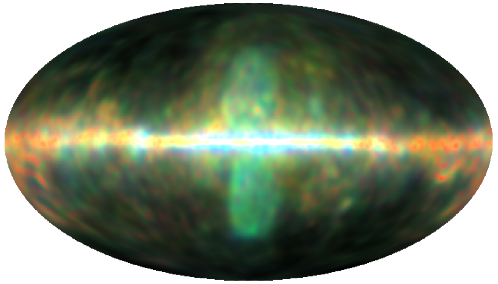
Non-Poissonian templates

Credit: Collin, Rodd, Safdi, 2016



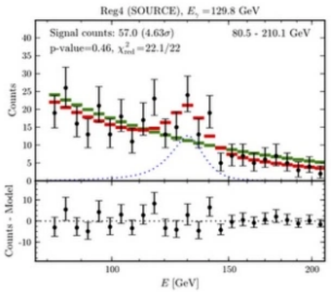
Adaptive templates

Credit: Storm+, 2017



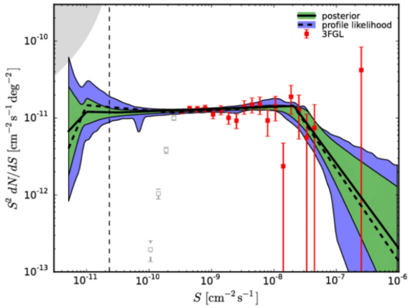
Spectral decomposition

Credit: Ensslin+, 2015



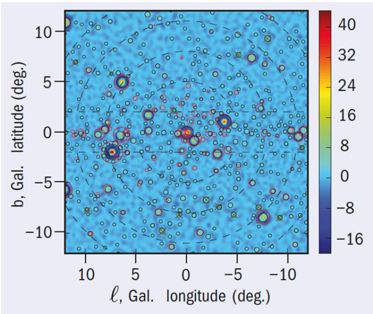
Line searches

Credit: CW, 2012



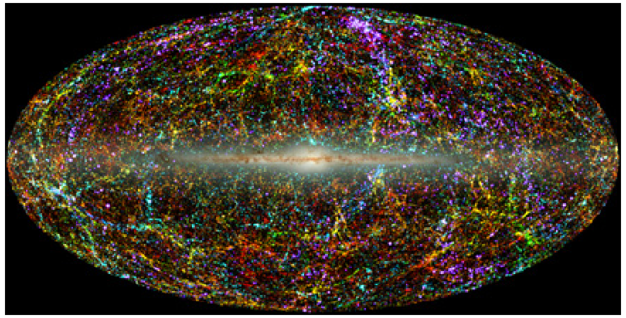
One-point statistics

Credit: Zecchlin+, 2015



Wavelet filtering

Credit: Bartels+, 2016

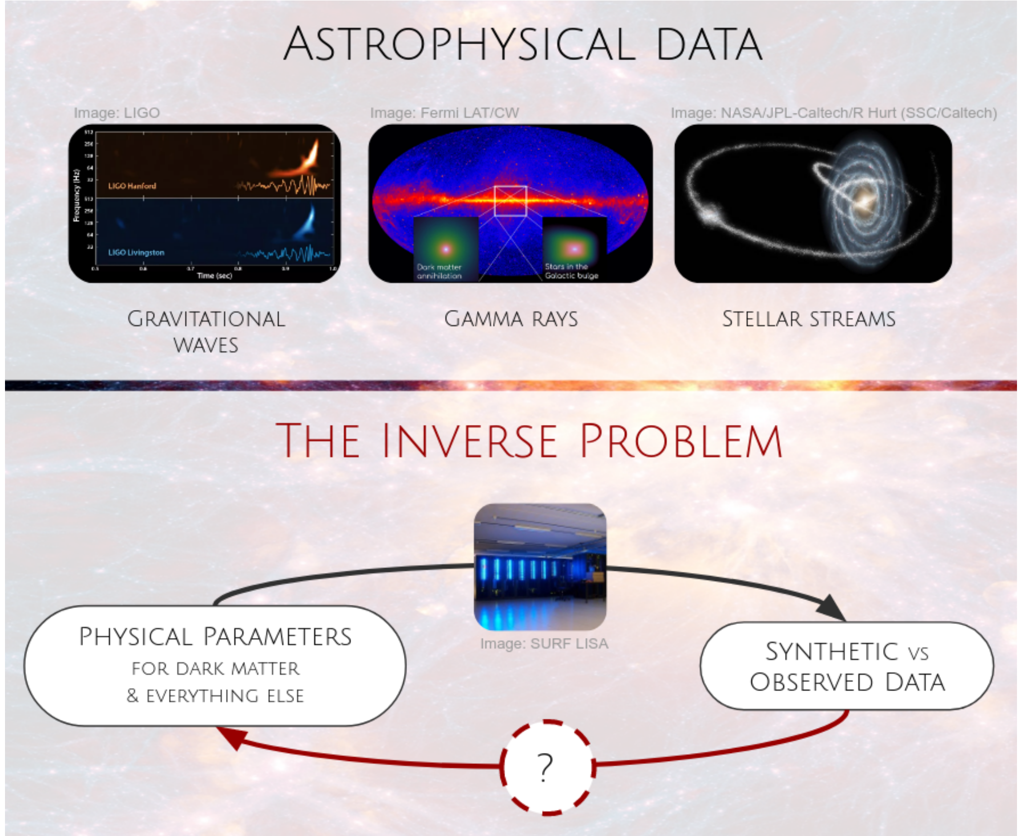


Cross-correlation studies

Credit: 2MASS, 2006

What is actually the problem that these methods try to solve?

The inverse problem



$$p(\mathrm{DM} = \mathcal{H}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathrm{DM} = \mathcal{H})p(\mathrm{DM} = \mathcal{H})}{p(\mathbf{x})}$$

- $\mathbf{x}$: Observational data
- $\mathcal{H}$: DM hypothesis
  $\mathcal{H} = \text{WIMP, axion, sterile neutrino}\ldots$
- $p(\mathbf{x}|\mathrm{DM} = \mathcal{H})$: Likelihood of data $\mathbf{x}$ given $\mathcal{H}$
- $p(\mathrm{DM} = \mathcal{H}|\mathbf{x})$: Posterior probability of $\mathcal{H}$
- $p(\mathbf{x})$: Bayesian evidence (normalizing factor)

$$\text{Odds ratio} = \frac{p(\mathrm{DM} = \mathcal{H}|\mathbf{x})}{p(\mathrm{DM} \neq \mathcal{H}|\mathbf{x})}$$

$$\mathrm{TS} = -2\ln\frac{p(\mathbf{x}|\boldsymbol{\nu}_{\mathrm{DM}} = 0)}{\max_{\boldsymbol{\nu}_{\mathrm{DM}}}\ p(\mathbf{x}|\boldsymbol{\nu}_{\mathrm{DM}})}$$

Marginal likelihoods require integration

The marginal likelihood requires integrating over DM parameters $\boldsymbol{\nu}$ and all other parameters $\boldsymbol{\eta}$.

$$p(\mathbf{x}|\mathrm{DM} = \mathcal{H}) = \int_V d^N\nu\, d^M\eta\ p(\mathbf{x}|\boldsymbol{\nu}, \boldsymbol{\eta})p(\boldsymbol{\nu}|\mathcal{H})p(\boldsymbol{\eta})$$

- $p(\mathbf{x}|\boldsymbol{\nu}, \boldsymbol{\eta})$: Likelihood of the data given all parameters.
- $p(\boldsymbol{\nu}|\mathcal{H})$: Prior knowledge about your DM parameters
- $p(\boldsymbol{\eta})$: Prior knowledge about everything else

Bayesian Net

Our knowledge about the data as well as the underlying physics can be encoded in a Bayesian net.

$$p(\mathbf{x}|\boldsymbol{\nu},\boldsymbol{\eta})p(\boldsymbol{\eta})p(\boldsymbol{\nu}|\mathcal{H}) \equiv p(\boldsymbol{x},\boldsymbol{\nu},\boldsymbol{\eta}|\mathcal{H}) =$$

Likelihood-based inference is hard

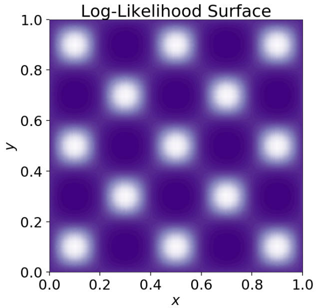
In likelihood-based approaches, we have to formally solve the integral

$$p(\mathbf{x}|\mathrm{DM} = \mathcal{H}) = \int_V d^N\nu\, d^M\eta\;\; p(\mathbf{x}|\boldsymbol{\nu}, \boldsymbol{\eta})p(\boldsymbol{\nu}|\mathcal{H})p(\boldsymbol{\eta})$$

Techniques

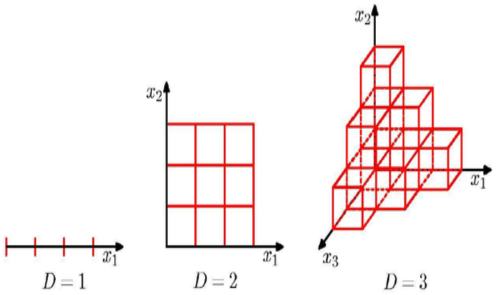
- Bayesian: MCMC, Nested Sampling
- Frequentist (optimization): Minuit, Gradient based techniques (BFGS).

Multimodal posteriors



(Dynesty 1.1)

Curse of dimensionality

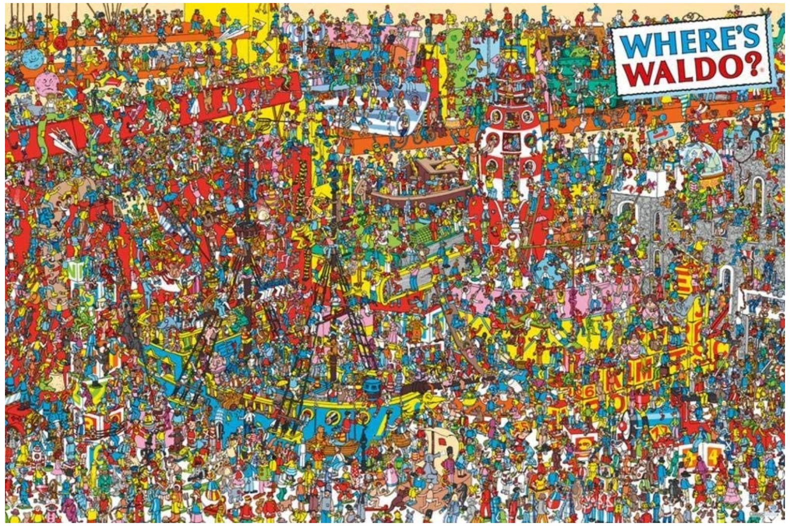


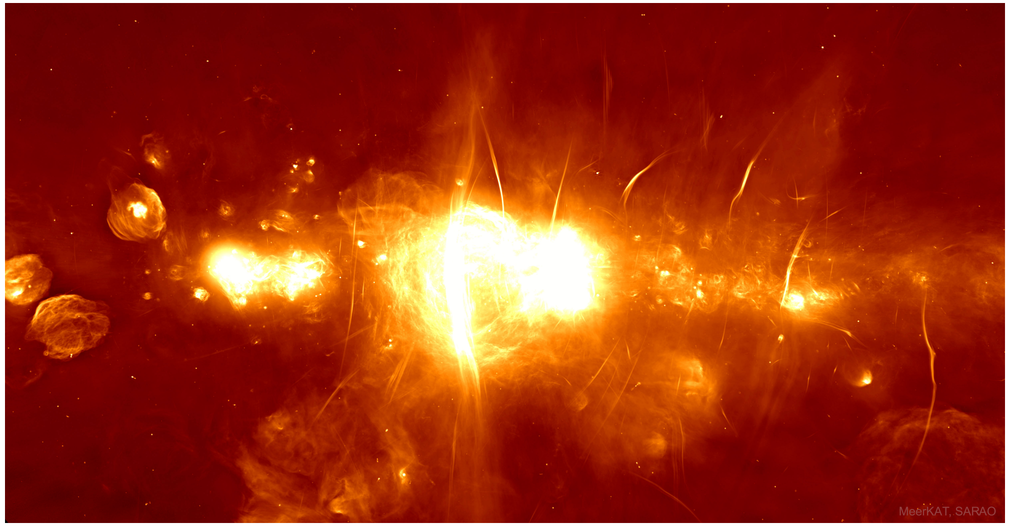
(Bishop 2007)

No simulation reuse

Why is likelihood-based inference hard?

Two extreme examples to make my point.





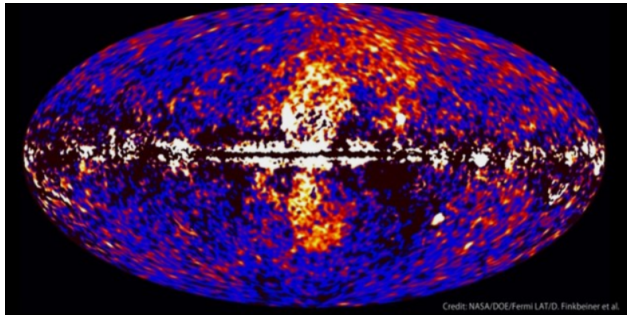
In order to detect Waldo ($\boldsymbol{\nu}$) robustly against modeling systematics, we have to model accurately each and every person in that image ($\boldsymbol{\eta}$).

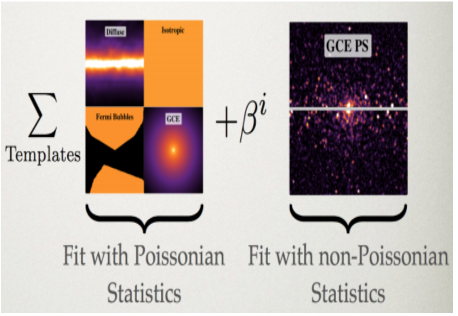
In order to identify radio lines from axion dark matter in the Galactic center ($\boldsymbol{\nu}$) robustly against modeling systematics, we have to model accurately all CR sources, propagation and emission in that image ($\boldsymbol{\eta}$).

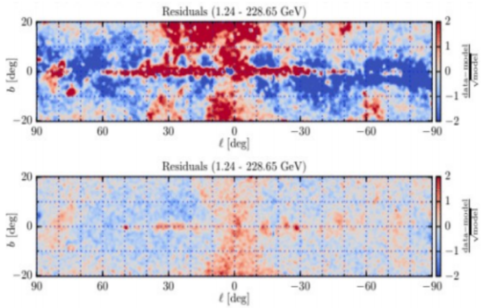
Back to indirect DM searches

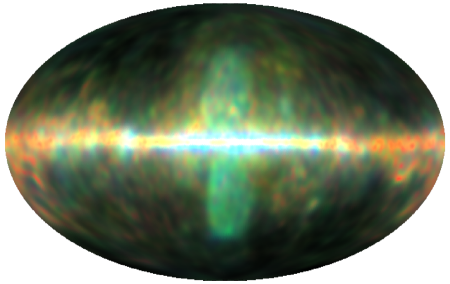
Analyses are defined through their compromises:

- usage of summary statistics
- masking of difficult to model data
- model approximations / simplistic models
- inference approximations









Template regression

- $p(\mathbf{x}_{\text{pi0}}, \boldsymbol{\eta}_{\text{diff}}) \to \delta(\mathbf{x}_{\text{pi0}} - \bar{\mathbf{x}}_{\text{pi0}})$
- $p(\mathbf{A}) \to \prod_i \text{Unif}(A_i)$
- Poisson → Gaussian

Non-Poissonian templates

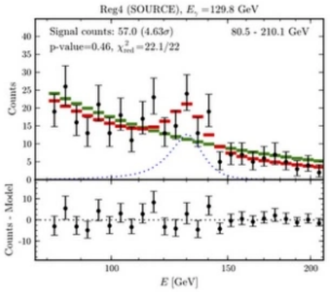
- Absorb stochastic point sources in effective non-Poissonian likelihood
- Neglect energy information

Adaptive templates

- Impose ad-hoc correlation structure on templates
- Nearest neighbor regularization rather than proper Bayesian priors

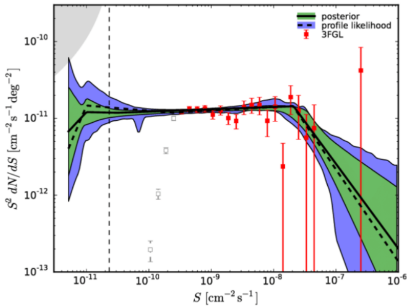
Spectral decomposition

- Remove spatial information & templates
- Assume spectrum of each physical component is the same everywhere



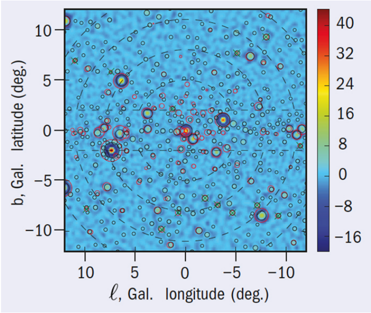
Line searches

- No spatial information, only summary energy spectra
- ROI is small energy range
- Approximate backgrounds with power-law
- “Trade systematics for statistical uncertainties”



One-point statistics

- Model likelihood of count statistics histogram rather than likelihood of skymaps
- Mask difficult sky regions
- Focus on single energy band



Wavelet filtering

- Put data with complex diffuse backgrounds through a matched (wavelet) filter
- Otherwise similar to One-point statistics analysis

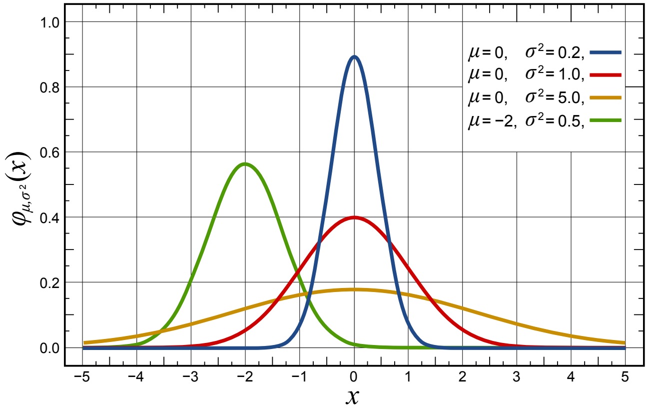
Cross-correlation studies

- Project data onto tracers of DM mass
- Neglect effects of local Universe
- Sometimes neglect correlation between overlapping tracers

Can we deal with marginalization (integration of nuisance parameters $\boldsymbol{\eta}$) in a different way?

Evaluating vs sampling marginal likelihoods

EVALUATING the marginal likelihood is HARD



Evaluating the **probability density** $p(\mathbf{x}|\boldsymbol{\nu})$ for a given value of $\mathbf{x}$ **is hard**, since we have to perform the integral

$$p(\mathbf{x}|\boldsymbol{\nu}) = \int d\boldsymbol{\eta}\, p(\mathbf{x}|\boldsymbol{\nu},\boldsymbol{\eta})p(\boldsymbol{\eta})$$

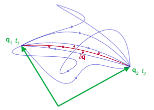
SAMPLING from the marginal likelihood is EASY

Sampling the probability density $p(\mathbf{x}|\boldsymbol{\nu})$ **is simple:**

1. Draw $\boldsymbol{\eta} \sim p(\boldsymbol{\eta})$ from the prior.
2. Draw $\mathbf{x}$ from the simulator $\mathbf{x} \sim p(\mathbf{x}|\boldsymbol{\nu},\boldsymbol{\eta})$.
3. Forget about $\boldsymbol{\eta}$.
4. You now have a sample from $\mathbf{x} \sim p(\mathbf{x}|\boldsymbol{\nu})$.

Neural simulation-based inference

Consider the following minimization problem (following Hermans+2019)



$$L[d(\mathbf{x},\boldsymbol{\nu})] = -\int d\mathbf{x}\, d\boldsymbol{\nu}\, d\boldsymbol{\eta}\ \left[p(\mathbf{x}|\boldsymbol{\nu},\boldsymbol{\eta})p(\boldsymbol{\nu})p(\boldsymbol{\eta}) \ln d(\mathbf{x},\boldsymbol{\nu}) + p(\mathbf{x})p(\boldsymbol{\nu})p(\boldsymbol{\eta}) \ln(1 - d(\mathbf{x},\boldsymbol{\nu}))\right]$$

Minimizing the functional w.r.t. the function d yields formally the optimal Bayesian classifier

$$d(\mathbf{x},\boldsymbol{\nu}) = \frac{p(\boldsymbol{\nu}|\mathbf{x})}{p(\boldsymbol{\nu}|\mathbf{x}) + p(\boldsymbol{\nu})} = \frac{p(\mathbf{x}|\boldsymbol{\nu})/p(\mathbf{x})}{p(\mathbf{x}|\boldsymbol{\nu})/p(\mathbf{x}) + 1}\ .$$

Since we know the prior $p(\boldsymbol{\nu})$, we can easily obtain the posterior $p(\boldsymbol{\nu}|\mathbf{x})$.

There are quite a few algorithms that formally approximate the (marginal) posterior: neural likelihood estimation, neural posterior estimation, neural ratio estimation. See Cranmer+2019 for a recent discussion (but the field is developing fast!).

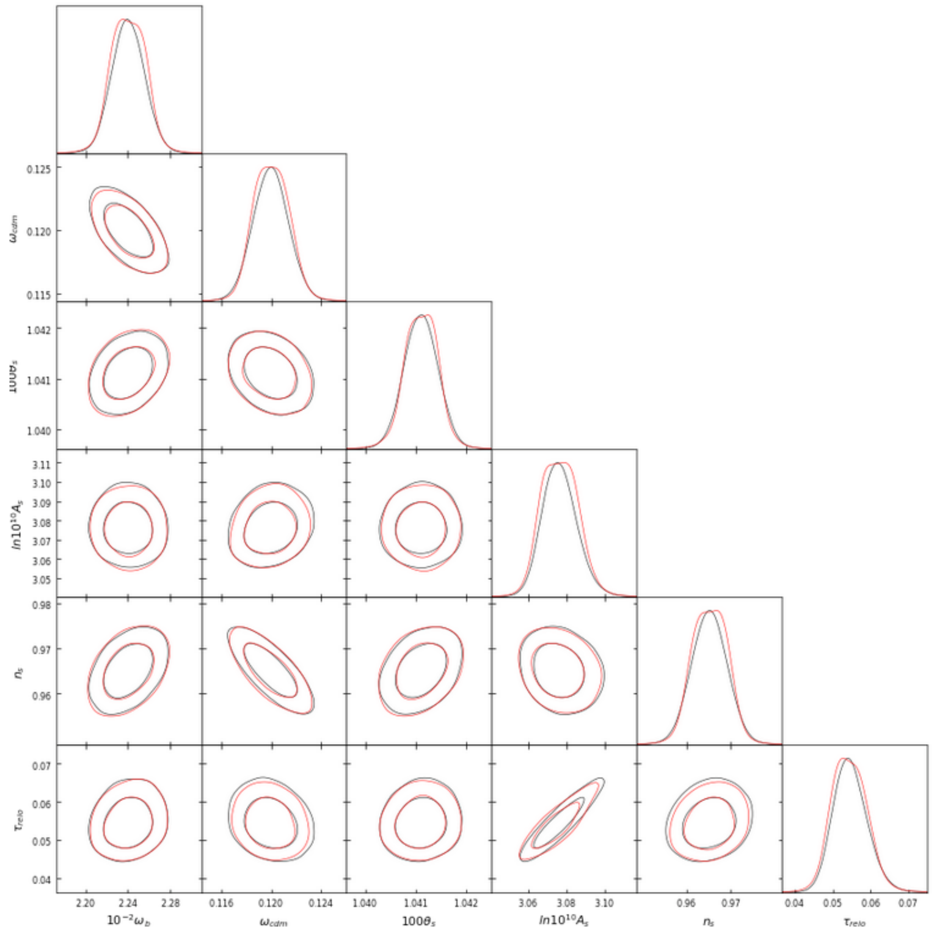
How to do this in practice?

1. Generate samples $\mathbf{x}, \boldsymbol{\nu} \sim p(\mathbf{x}|\boldsymbol{\nu})p(\boldsymbol{\nu})$ as training data.
2. Come up with some parametrized $d_\phi(\mathbf{x}, \boldsymbol{\nu})$ as a flexible function of parameters ϕ (e.g. a neural network).
3. Optimize ϕ using stochastic gradient descent.

The trained network now represents **Bayesian marginal posteriors** for the dark matter parameters, $p(\boldsymbol{\nu}|\mathbf{x})$, **for any possible observation** $\mathbf{x}$.

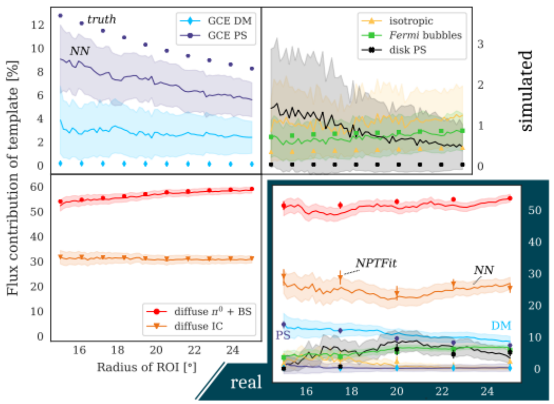
This is just one example. There are many neural-network based methods with similar goals and results. This is an open extremely promising research field!

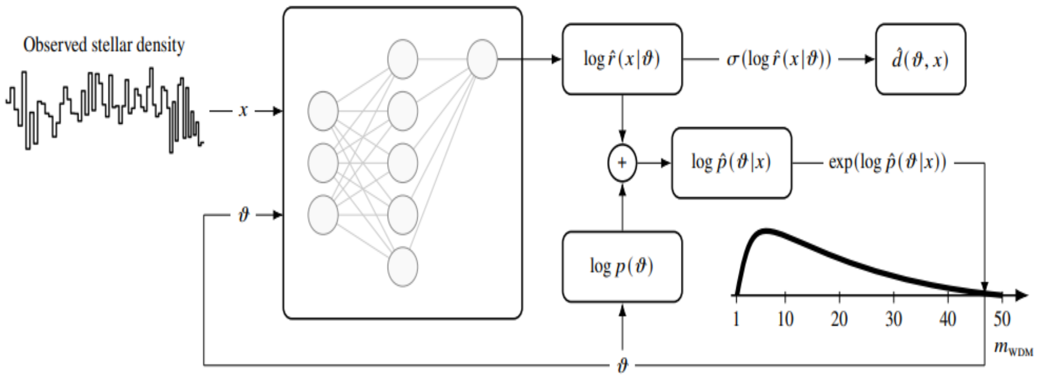
Simulation-based vs sampling-based posteriors



Some ongoing work: Cosmological parameter inference with neural ratio estimation (black, 5000 simulator runs) and MCMC (red, 60000 runs). Credit: Alex Cole, using swyft.

Examples and applications



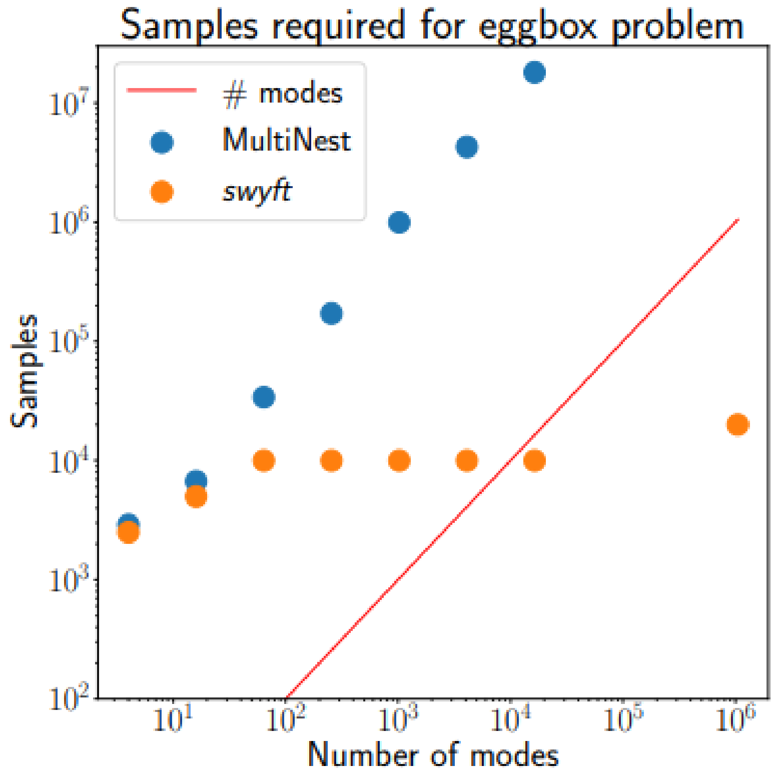


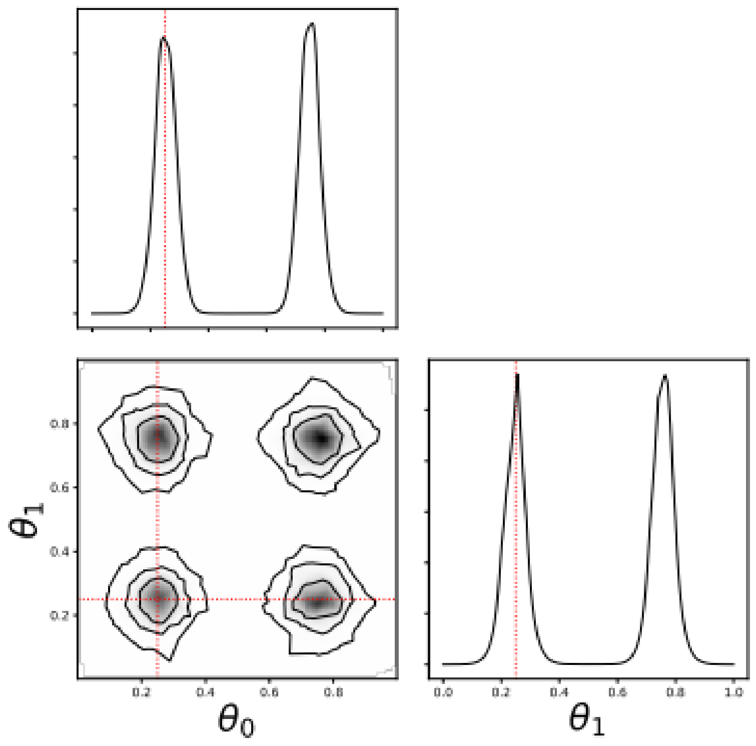
- Re-analysis of the Galactic center excess, comparing point sources with DM hypothesis List+ 2020
- The analysis is actually based on Bayesian neural networks
- Different method, but the goal is the same.

- An analysis of stellar streams, Hermans+ 2019.
- Marginalization over dozens of parameters that describe the encounter history.

“Inference super powers”

- Consider a high-dimensional eggbox posterior, with two modes in each direction. Assuming 20 parameters, this give $2^{20} \sim 10^6$ modes.
- We can effectively marginalize over likelihoods with 1 Mio modes, using only 10 thousand samples.





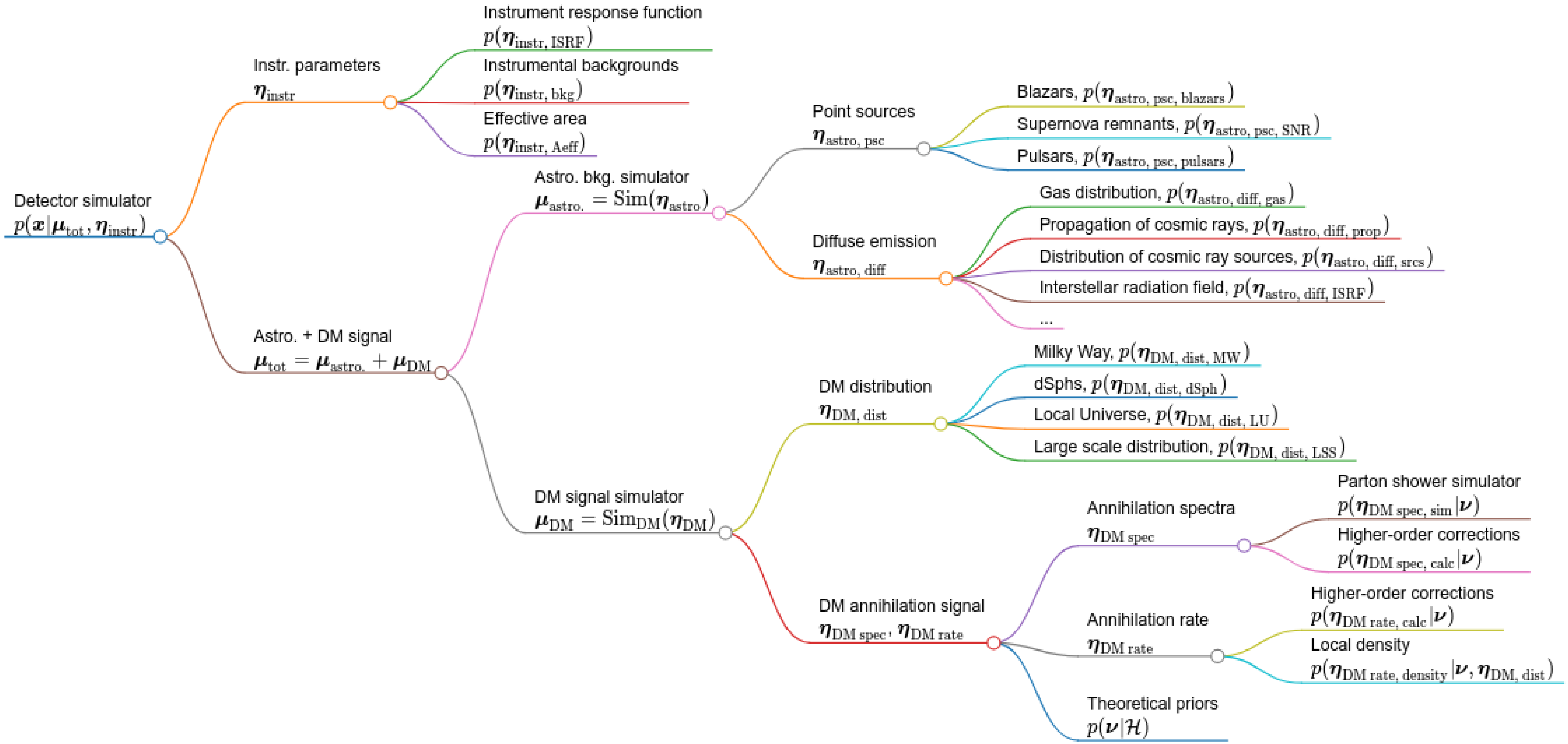
From Miller+2020.

Potential benefits for our community

- **Better models**:
  - Extra parameters do not cause extra inference costs.
  - Typically simulator-based inference requires far fewer simulator runs than likelihood-based methods.
  - Simulations runs can be re-used, so even detailed and slow simulators are an option.
  - **Focus on improving physical models rather than statistical methods.**
- **Better inference**:
  - Statistical inference is done “automatically” through an NN.
  - Thanks to amortized posteriors, one can do consistency checks like coverage-tests, etc.
  - Less worries about biases and information loss due to imperfect summary statistics, cuts into data, etc.
- **Better science**:
  - Discussions can focus on the best physics assumptions, rather than on statistical techniques.
  - We actually might agree on the detection of a dark matter signal with indirect searches.
- But, there are also challenges:
  - Simulation-based inference does not provide goodness-of-fit.
  - There are other methods (e.g. variational inference, ELBO maximization) that can help to maximize the model evidence and ensure good agreement between data and model.

Back to indirect searches

Instead of integrating, we have to sample from $p(\mathbf{x}|\boldsymbol{\nu},\boldsymbol{\eta})p(\boldsymbol{\eta})p(\boldsymbol{\nu}|\mathcal{H}) \equiv p(\boldsymbol{x},\boldsymbol{\nu},\boldsymbol{\eta}|\mathcal{H})$



What do we need?

- Intermediate products (gas maps, interstellar radiation field, source population models, …)
  - Current situation:
    - Usually one or very few versions are published, sometimes $\bar{\mathbf{m}} \pm \Delta\mathbf{m}$.
  - Ideally we need:
    - Programs that randomly generate, e.g., gas maps from the data posterior, $p(\mathbf{x}_{\text{gas}}|21\text{cm})$
    - Alterantively a catalog of random samples from that program.
- Experimental likelihoods (cosmic rays, results from the Cherenkov Telescopes, …)
  - Current situation:
    - Often flux with statistical (+ sometimes systematic) errors, with unclear correlation structure
  - Ideally we need:
    - directly measured data in some representation
    - a sampler to generate possible data realizations given the physical flux, $\mathbf{x} \sim p(\mathbf{x}|\boldsymbol{\mu})$
      - fast convolution with instrument response functions
    - but, even just covariance information about systematics errors is a good start
- Education
  - Ideally, students should be trained not only in single statistical methods, but have a robust knowledge of Bayesian inference and networks.
  - Students need to be familiar with training simple neural networks.
- Tools
  - Few people would use nested sampling if everybody had to write their own nested sampler.
  - We need standard tools that work in >90% of the use cases. Examples: SBI, SWYFT

Conclusions

"The focus of iDMEu, more than on showcasing achievements or the current status of the field, is on the challenges within the communities and in terms of cross-talk with others, and on current and future needs."

- Our current way of modeling and learning from data is not sustainable
  - Robust results require realistic and detailed models, in particular as data becomes better.
    - We are less and less "statistics limited" and more and more "background limited"
  - Simulator side
    - More detailed models –> Slower simulators –> Can afford **less simulator runs**
  - Inference side
    - More detailed models –> More parameters –> **Need more simulators runs** to converge
- Relevant for present (e.g. Fermi LAT, XMM-Newton) and upcoming data (CTA, XRISM, eROSITA, SKA).
- Simulation based inference can alleviate many of the pain points
  - More simulation efficient
  - No extra costs for extra uncertainties
  - Simulation re-use is possible
  - There are other powerful methods that I didn't had time to cover
- Realizing the potential of simulator based inference should happen at the community level, since it requires adjustments in how we publish results, and how we construct physical and statistical models.