Anomaly aware machine learning for dark matter direct detection at DARWIN

Andre Scaffidi and Roberto Trotta for the DARWIN collaboration.

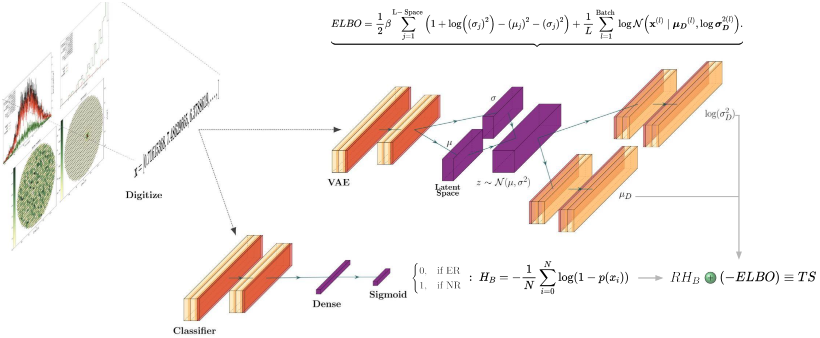
DARWIN

SISSA DATASCIENCE
Machine Learning for the Natural Sciences

SISSA

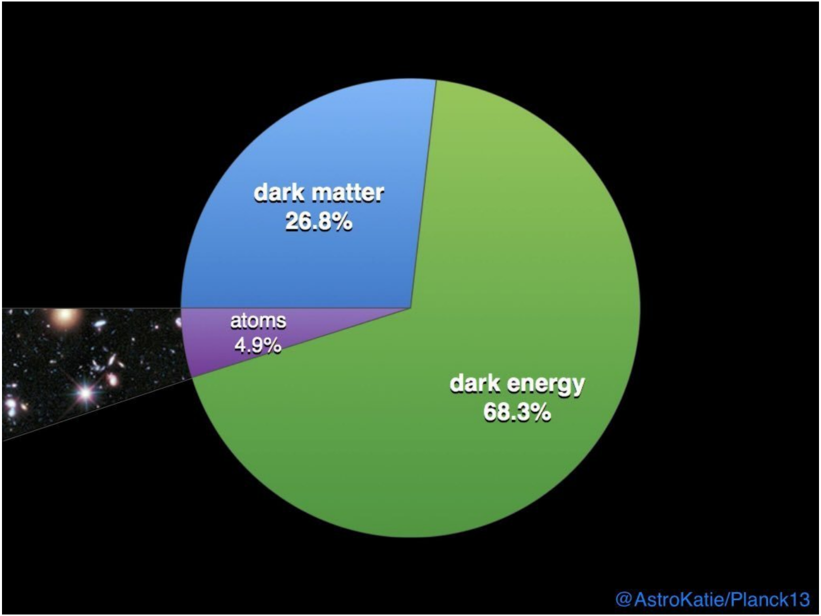
Overview

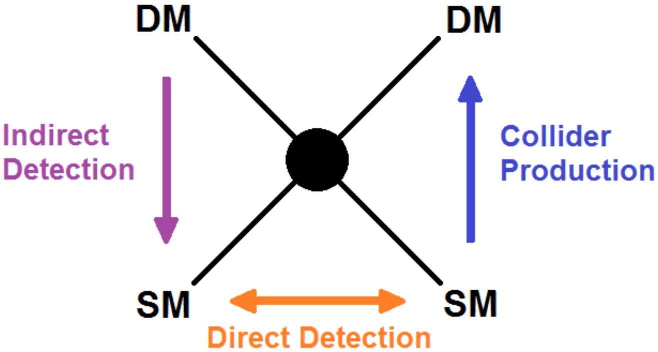
- Take simulated detector observables and construct anomaly detection task.
- $\Rightarrow$ New deep learning pipeline to improve upon traditional likelihood approaches.
- Can improve sensitivity over standard approach.



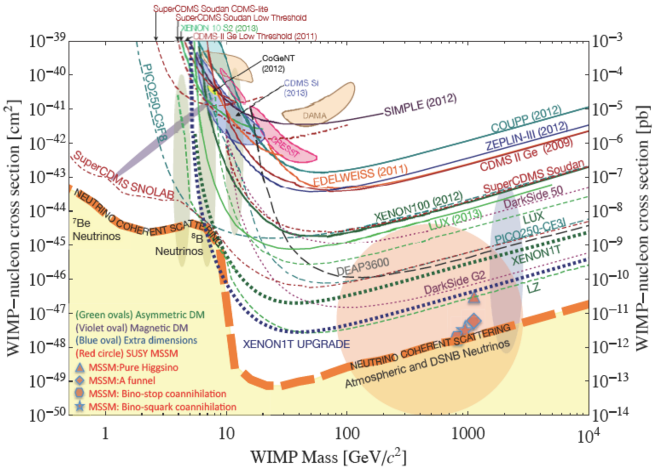
Dark Matter Direct Detection

The dark matter issue

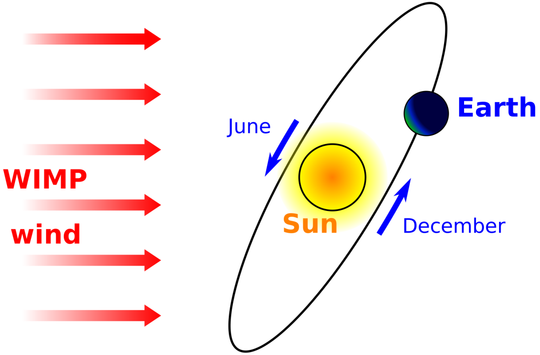



DM
DM
Indirect
Detection
Collider
Production
SM
SM
Direct Detection

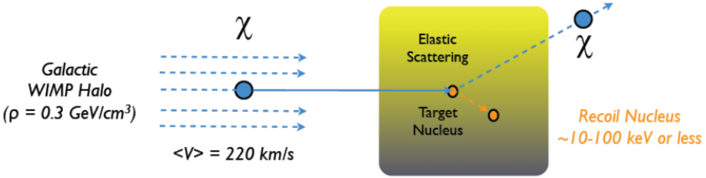
- Current and planned next-generation DD experiments are probing/will probe a very large portion of the parameter space of the WIMP (Weakly Interacting Massive Particle) model.



Direct detection: Schematic



Direct detection: The differential event rate



Scattering rate [Events/(keV kg day)].

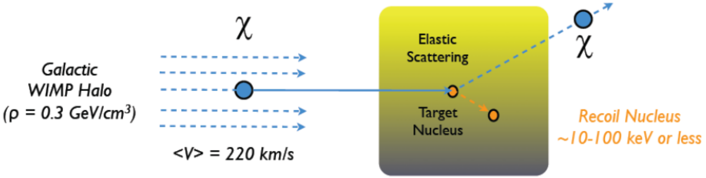
$$R(E,t) = \underbrace{\frac{\rho\sigma}{2m_\chi \mu_p^2}(A^{\text{eff}})^2 F^2(E)}_{\text{Particle physics.}} \underbrace{\eta(E,t)}_{\text{Astrophysics}} \quad , \quad \theta = \{m_\chi, \sigma\}$$

Expected number events after exposure MT:

$$\mu(E_i) = MT \int_0^\infty dE\, \epsilon(E)\, \phi(E, E_i)\, \mathbf{R(E)}$$

Don't need to concern ourselves with $\epsilon(E)\, \phi(E, E_i)$ when using SBI

Direct detection: The differential event rate



Scattering rate [Events/(keV kg day)].

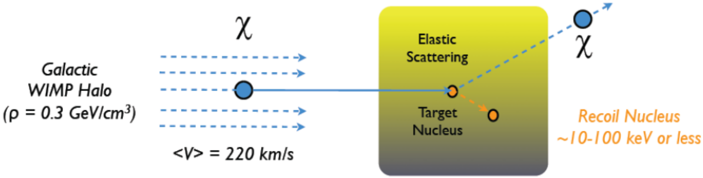
$$R(E,t) = \underbrace{\frac{\rho\sigma}{2m_\chi\mu_p^2}(A^{\text{eff}})^2F^2(E)}_{\text{Particle physics.}}\ \underbrace{\eta(E,t)}_{\text{Astrophysics}} \quad , \quad \theta = \{m_\chi, \sigma\}$$

Expected number events after exposure MT:

$$\mu(E_i) = MT\int_0^\infty dE\,\epsilon(E)\,\phi(E,E_i)\,\mathbf{R(E)}$$

Don't need to concern ourselves with $\epsilon(E)\,\phi(E,E_i)$ when using SBI

Direct detection: The differential event rate



Scattering rate [Events/(keV kg day)].

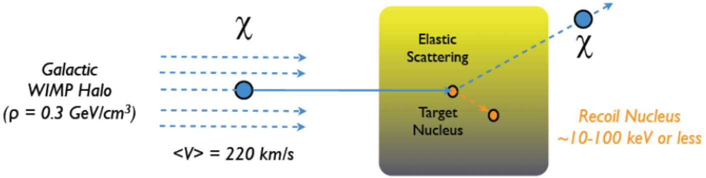
$$R(E,t) = \underbrace{\frac{\rho\sigma}{2m_\chi \mu_p^2} (A^{\text{eff}})^2 F^2(E)}_{\text{Particle physics.}} \quad \underbrace{\eta(E,t)}_{\text{Astrophysics}} \quad , \qquad \theta = \{m_\chi, \sigma\}$$

Expected number events after exposure MT:

$$\mu(E_i) = MT \int_0^\infty dE\, \epsilon(E)\, \phi(E, E_i)\, \mathbf{R(E)}$$

Don't need to concern ourselves with $\epsilon(E)\, \phi(E, E_i)$ when using SBI

Direct detection: The differential event rate



Scattering rate [Events/(keV kg day)].

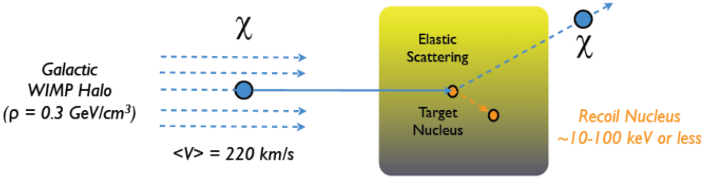
$$R(E,t) = \underbrace{\frac{\rho\sigma}{2m_\chi \mu_p^2} (A^{\mathrm{eff}})^2 F^2(E)}_{\text{Particle physics.}} \quad \underbrace{\eta(E,t)}_{\text{Astrophysics}} \quad , \qquad \theta = \{m_\chi, \sigma\}$$

Expected number events after exposure MT:

$$\mu(E_i) = MT \int_0^\infty dE \, \epsilon(E) \, \phi(E, E_i) \, \mathbf{R(E)}$$

Don't need to concern ourselves with $\epsilon(E)\,\phi(E, E_i)$ when using SBI

Direct detection: The differential event rate



Scattering rate [Events/(keV kg day)].

$$R(E,t) = \underbrace{\frac{\rho\sigma}{2m_\chi\mu_p^2}(A^{\text{eff}})^2F^2(E)}_{\text{Particle physics.}} \quad \underbrace{\eta(E,t)}_{\text{Astrophysics}} \quad , \quad \theta = \{m_\chi, \sigma\}$$

Expected number events after exposure MT:

$$\mu(E_i) = MT\int_0^\infty dE\,\epsilon(E)\,\phi(E,E_i)\,\mathbf{R(E)}$$

Don't need to concern ourselves with $\epsilon(E)\,\phi(E,E_i)$ when using SBI

Direct detection: Experiments

- Search for WIMP dark matter.
- Variety of detector targets/strategies.
- Leading are underground* “low-background” experiments. This work: DARWIN.
    - Methods tractable by any low background new physics search.

Direct detection: Experiments

- Search for WIMP dark matter.
- Variety of detector targets/strategies.
- Leading are underground* “low-background” experiments. This work: DARWIN.
    - Methods tractable by any low background new physics search.

Direct detection: Experiments

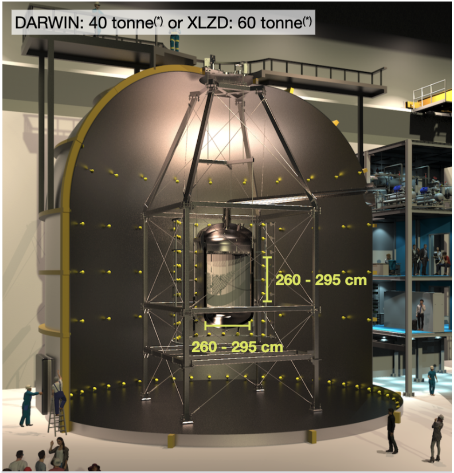
- Search for WIMP dark matter.
- Variety of detector targets/strategies.
- Leading are underground* "low-background" experiments. This work: DARWIN.
    - Methods tractable by any low background new physics search.

Direct detection: Experiments

- Search for WIMP dark matter.
- Variety of detector targets/strategies.
- Leading are underground* “low-background” experiments. This work: DARWIN.
    - Methods tractable by any low background new physics search.

DARWIN collaboration: Proposal

∼ 200 members



DARWIN collaboration

Direct detection: Traditional likelihood-based analysis

After some exposure → collect events:

$$\mathcal{L}(s+b) \sim \frac{e^{-\mu_s(\theta)-\mu_b(\theta)}}{n!} \prod_{i=1}^{n} \frac{d\left(N_s+N_b\right)}{dE}\left(E_i \mid \theta\right)$$

- Expected number of signal events: $\mu_s = MT \cdot \int dN_s/dE$
- Expected number of background events: $\mu_b = MT \cdot \int dN_b/dE$
- Spectral information

Model parameters $\theta = \{m_\chi, \sigma ...\}$ phenomenalogically determine two things:

- Number of expected events
- Signal spectrum 'shape'

} Important for ML analysis

Direct detection: Traditional likelihood-based analysis

After some exposure → collect events:

$$\mathcal{L}(s+b) \sim \frac{e^{-\mu_s(\theta)-\mu_b(\theta)}}{n!} \prod_{i=1}^{n} \frac{d\left(N_s+N_b\right)}{dE}\left(E_i \mid \theta\right)$$

- Expected number of signal events: $\mu_s = MT \cdot \int dN_s/dE$
- Expected number of background events: $\mu_b = MT \cdot \int dN_b/dE$
- Spectral information

Model parameters $\theta = \{m_\chi, \sigma ...\}$ phenomenalogically determine two things:

- Number of expected events
- Signal spectrum 'shape'

} Important for ML analysis

Direct detection: Traditional likelihood-based analysis

After some exposure $\rightarrow$ collect events:

$$\mathcal{L}(s+b) \sim \frac{e^{-\mu_s(\theta)-\mu_b(\theta)}}{n!} \prod_{i=1}^{n} \frac{d\left(N_s+N_b\right)}{dE}\left(E_i \mid \theta\right)$$

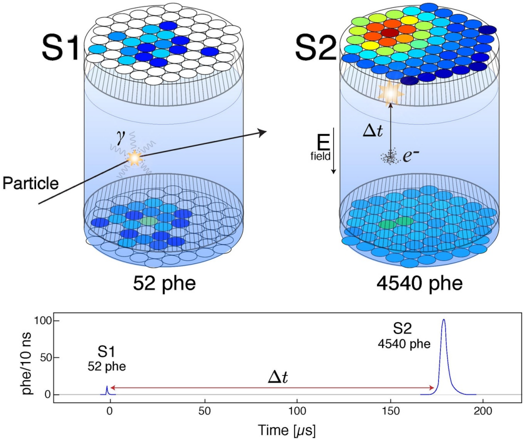
- Expected number of signal events: $\mu_s = MT \cdot \int dN_s/dE$
- Expected number of background events: $\mu_b = MT \cdot \int dN_b/dE$
- Spectral information

Model parameters $\theta = \{m_\chi, \sigma ...\}$ phenomenalogically determine two things:

- Number of expected events
- Signal spectrum 'shape'

} Important for ML analysis

Direct detection in TPCs: Events

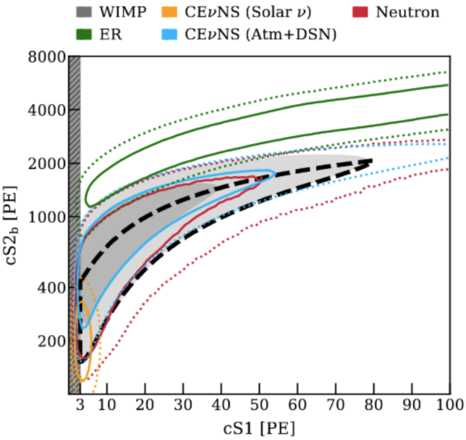


- S1: Prompt scintillation signal from recoil event.
- S2: Electron charges produced during ionization drift upwards → extracted into gaseous phase creating larger scintillation.

Direct detection: Traditional likelihood-based analysis

$$\prod_{i=1}^{n} \frac{d\left(N_s+N_b\right)}{dE}\left(E_i \mid \theta\right) \rightarrow \text{ 2D pdf derived from 'templates'}$$

Relies heavily on high-level summary statistics cS1,cS2:
$\Rightarrow \mathbf{E} = \mathbf{g}(\mathbf{cS1}, \mathbf{cS2})$

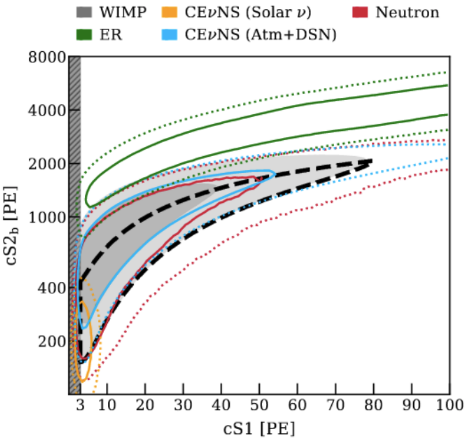


Direct detection: Traditional likelihood-based analysis

$$\prod_{i=1}^{n} \frac{d\left(N_s + N_b\right)}{dE}\left(E_i \mid \theta\right) \rightarrow$$ 2D pdf derived from 'templates'

Relies heavily on high-level summary statistics cS1,cS2:
$\Rightarrow \mathbf{E} = \mathbf{g}(\mathbf{cS1}, \mathbf{cS2})$

↳ Fitted analytically



M. Doerenkamp

Simulation based inference (SBI)

Simulation-Based Inference (in a nutshell)

Simulation-based inference is a statistical technique that allows us to make inferences about a population or process based on simulated data. It involves the following steps:

1. Generate simulated data.
2. Use deep neural nets to learn underlying features of simulated data.
3. Use trained models to inference.

Simulation-Based Inference (in a nutshell)

Simulation-based inference is a statistical technique that allows us to make inferences about a population or process based on simulated data. It involves the following steps:

1. Generate simulated data.
2. Use deep neural nets to learn underlying features of simulated data.
3. Use trained models to inference.

Simulation-Based Inference (in a nutshell)

Simulation-based inference is a statistical technique that allows us to make inferences about a population or process based on simulated data. It involves the following steps:

1. Generate simulated data.
2. Use deep neural nets to learn underlying features of simulated data.
3. Use trained models to inference.

Simulation-Based Inference (in a nutshell)

Simulation-based inference is a statistical technique that allows us to make inferences about a population or process based on simulated data. It involves the following steps:

1. Generate simulated data.
2. Use deep neural nets to learn underlying features of simulated data.
3. Use trained models to inference.

Benefits of SBI

- Can handle complex models with intractable likelihoods.
- Use deep neural nets to learn underlying features of simulated data/summary stats.
- Once a simulator has been established, possible to include arbitrarily complicated simulations into analysis: prompt readouts $\rightarrow$ high level summary stats.
- Makes no assumptions regarding the analytical form of the likelihood.
- Need no special treatment of nuisance parameters.
    - Can in principle simulate/calibrate any detector effects and learn them directly.

Benefits of SBI

- Can handle complex models with intractable likelihoods.
- Use deep neural nets to learn underlying features of simulated data/summary stats.
- Once a simulator has been established, possible to include arbitrarily complicated simulations into analysis: prompt readouts $\rightarrow$ high level summary stats.
- Makes no assumptions regarding the analytical form of the likelihood.
- Need no special treatment of nuisance parameters.
  - Can in principle simulate/calibrate any detector effects and learn them directly.

Benefits of SBI

- Can handle complex models with intractable likelihoods.
- Use deep neural nets to learn underlying features of simulated data/summary stats.
- Once a simulator has been established, possible to include arbitrarily complicated simulations into analysis: prompt readouts $\rightarrow$ high level summary stats.
- Makes no assumptions regarding the analytical form of the likelihood.
- Need no special treatment of nuisance parameters.
    - Can in principle simulate/calibrate any detector effects and learn them directly.

Benefits of SBI

- Can handle complex models with intractable likelihoods.
- Use deep neural nets to learn underlying features of simulated data/summary stats.
- Once a simulator has been established, possible to include arbitrarily complicated simulations into analysis: prompt readouts $\rightarrow$ high level summary stats.
- Makes no assumptions regarding the analytical form of the likelihood.
- Need no special treatment of nuisance parameters.
    - Can in principle simulate/calibrate any detector effects and learn them directly.

Benefits of SBI

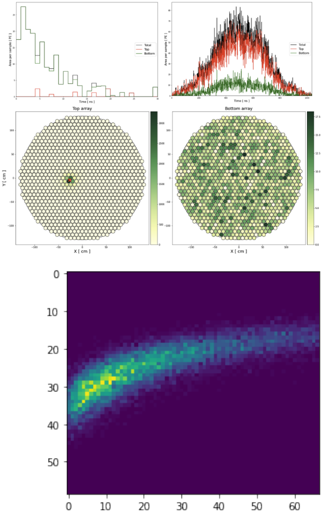
- Can handle complex models with intractable likelihoods.
- Use deep neural nets to learn underlying features of simulated data/summary stats.
- Once a simulator has been established, possible to include arbitrarily complicated simulations into analysis: prompt readouts $\rightarrow$ high level summary stats.
- Makes no assumptions regarding the analytical form of the likelihood.
- Need no special treatment of nuisance parameters.
    - Can in principle simulate/calibrate any detector effects and learn them directly.

Benefits of SBI

- Can handle complex models with intractable likelihoods.
- Use deep neural nets to learn underlying features of simulated data/summary stats.
- Once a simulator has been established, possible to include arbitrarily complicated simulations into analysis: prompt readouts $\rightarrow$ high level summary stats.
- Makes no assumptions regarding the analytical form of the likelihood.
- Need no special treatment of nuisance parameters.
  - Can in principle simulate/calibrate any detector effects and learn them directly.

Simulation-Based Inference with Neural Nets

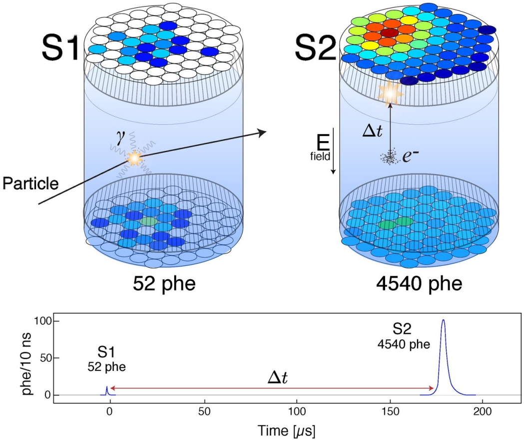
We have a variety of simulated data/summary stats available to us



Train NN to extract relevant features from simulated data. Effectively ‘learning’ the likelihood function directly from data.

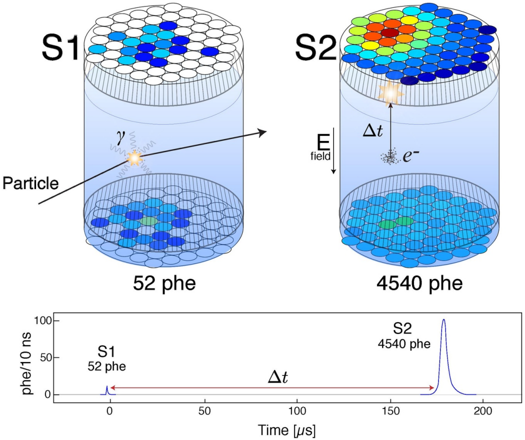
Simulation

Underground TPCs: Two types of events



- Nuclear Recoil (NR) → WIMPs
- (Dominant) Background → Electron Recoil (ER).
- Distance and ratio between S1/S2 peaks → NR vs. ER.
- NN can learn this instead!

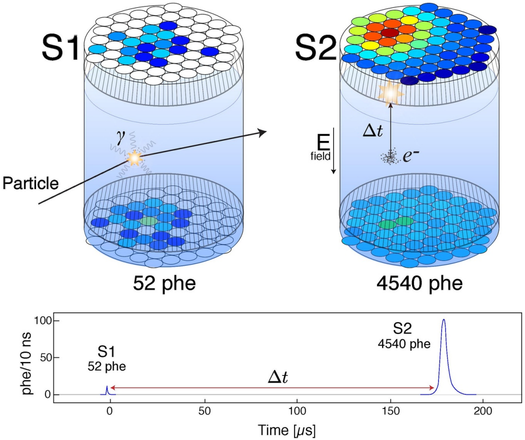
Underground TPCs: Two types of events



- Nuclear Recoil (NR) → WIMPs
- (Dominant) Background → Electron Recoil (ER).

- Distance and ratio between S1/S2 peaks → NR vs. ER.
- NN can learn this instead!

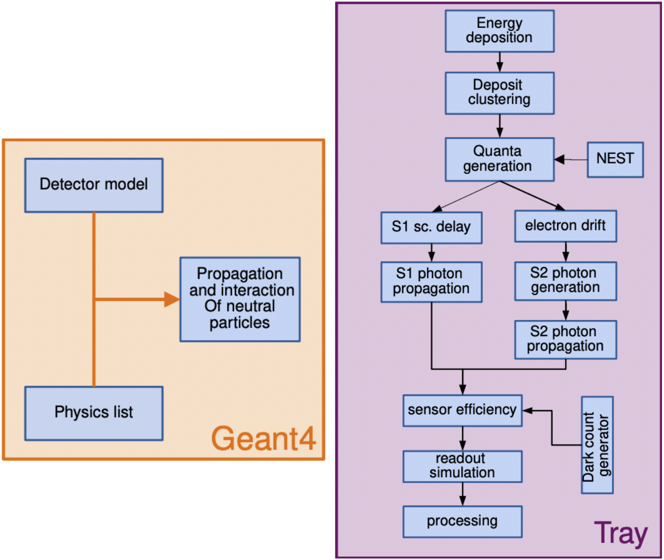
Underground TPCs: Two types of events



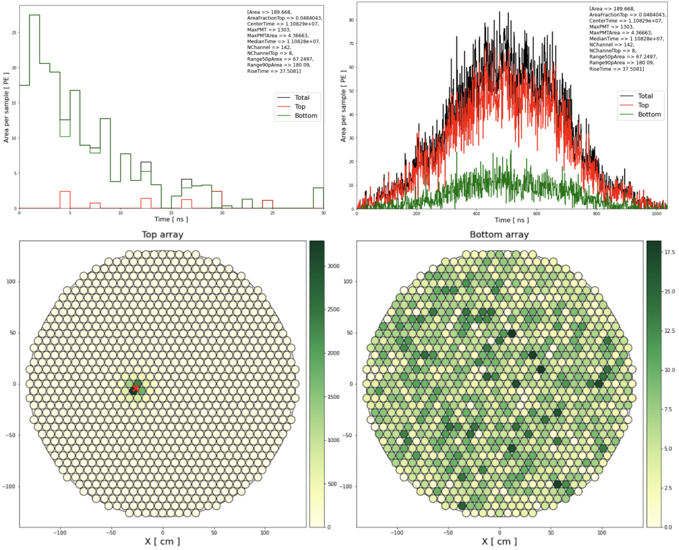
- Nuclear Recoil (NR) → WIMPs
- (Dominant) Background → Electron Recoil (ER).

- Distance and ratio between S1/S2 peaks → NR vs. ER.
- NN can learn this instead!

DARWIN: Simulation pipeline



DARWIN: NR event realisation



Nuclear recoil (NR) event example.

- SBI generally useful i.e parameter estimation DM mass/σ (Won't talk about it now.).
- Focus on 'Anomaly detection'.
  - Can we significantly detect excess NR (WIMP)?
- Increase power of technique with extensive simulation capabilities for DARWIN.

- SBI generally useful i.e parameter estimation DM mass/σ (Won't talk about it now.).
- Focus on 'Anomaly detection'.
    - Can we significantly detect excess NR (WIMP)?
- Increase power of technique with extensive simulation capabilities for DARWIN.

- SBI generally useful i.e parameter estimation DM mass/σ (Won't talk about it now.).
- Focus on 'Anomaly detection'.
    - Can we significantly detect excess NR (WIMP)?
- Increase power of technique with extensive simulation capabilities for DARWIN.

- SBI generally useful i.e parameter estimation DM mass/σ (Won't talk about it now.).
- Focus on 'Anomaly detection'.
    - Can we significantly detect excess NR (WIMP)?
- Increase power of technique with extensive simulation capabilities for DARWIN.

Analysis pipeline 1: Classification of recoil events

Classification: Signal vs. Background

- First primary objective in an analysis is to veto the dominant ER background.
- Binary classification: ER background vs. NR signal
- Traditional analyses → Must sacrifice NR acceptance due to ER events leaking into a low energy WIMP search region.
- Previous studies Sanz et. al, Herrero-Garcia et. al `arXiv:1911.09210, 2110.12248` for XENONnT.

Classification: Signal vs. Background

- First primary objective in an analysis is to veto the dominant ER background.
- Binary classification: ER background vs. NR signal
- Traditional analyses → Must sacrifice NR acceptance due to ER events leaking into a low energy WIMP search region.
- Previous studies Sanz et. al, Herrero-Garcia et. al `arXiv:1911.09210, 2110.12248` for XENONnT.

Classification: Signal vs. Background

- First primary objective in an analysis is to veto the dominant ER background.
- Binary classification: ER background vs. NR signal
- Traditional analyses → Must sacrifice NR acceptance due to ER events leaking into a low energy WIMP search region.
- Previous studies Sanz et. al, Herrero-Garcia et. al `arXiv:1911.09210, 2110.12248` for XENONnT.

Classification: Signal vs. Background

- First primary objective in an analysis is to veto the dominant ER background.
- Binary classification: ER background vs. NR signal
- Traditional analyses $\rightarrow$ Must sacrifice NR acceptance due to ER events leaking into a low energy WIMP search region.
- Previous studies Sanz et. al, Herrero-Garcia et. al `arXiv:1911.09210, 2110.12248` for XENONnT.

Classification: Signal vs. Background

Sanz et al. method:

1. Assume fixed WIMP mass 500 GeV and cross-section $\sigma = 10^{-45}$ cm^2 (34.2 live-days)
2. Generate image output of the simulated WIMP and "background" events
3. Train a **Convolutional Neural Network** (CNN) to classify WIMP vs. background

Classification: Signal vs. Background

Sanz et al. method:

1. Assume fixed WIMP mass 500 GeV and cross-section $\sigma = 10^{-45}$ cm^2 (34.2 live-days)
2. Generate image output of the simulated WIMP and "background" events
3. Train a **Convolutional Neural Network** (CNN) to classify WIMP vs. background

Classification: Signal vs. Background

Sanz et al. method:

1. Assume fixed WIMP mass 500 GeV and cross-section $\sigma = 10^{-45}$ cm^2 (34.2 live-days)
2. Generate image output of the simulated WIMP and "background" events
3. Train a **Convolutional Neural Network** (CNN) to classify WIMP vs. background

Classification: ER vs. NR

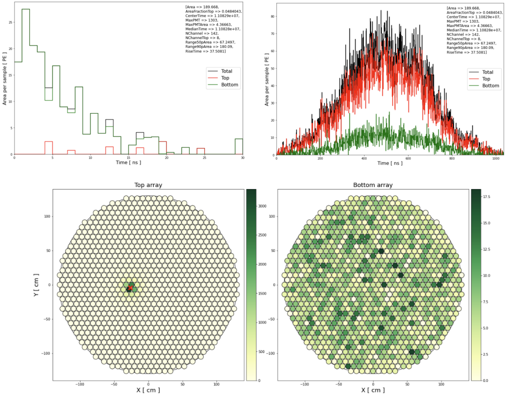
Our method:

1. Assume fixed WIMP mass 500 GeV and cross-section $\sigma = 10^{-45}$ cm^2 (34.2 live-days)
2. Generate image output of the simulated WIMP and "background" events
3. Train a **Convolutional Neural Network** (CNN) to classify WIMP vs. background

Discovered this can work with ensemble of WIMP masses.
Cross-section irrelevant for event-by-event bkg/signal classification.
I.e only learning if ER or NR.

Training data: Simulations

RAW event output S1, S2 PMT deposits (4-fold coincidence, 200 ns):

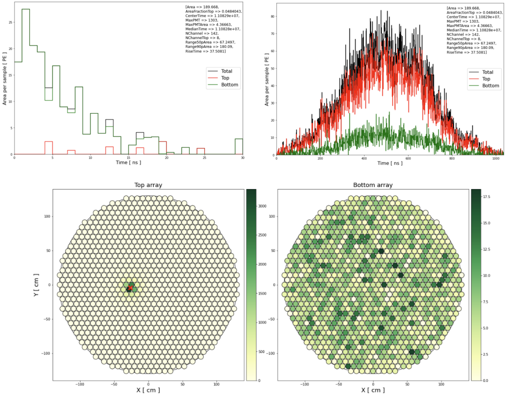


$\Longrightarrow$ $\mathbf{x} = [$S1WaveformTotal, S2WaveformTotal, S2Pattern $]$

Two distinct quanta: **Electron Recoil** (ER) and **Nuclear Recoil** (NR)

Training data: Simulations

RAW event output S1, S2 PMT deposits (4-fold coincidence, 200 ns):

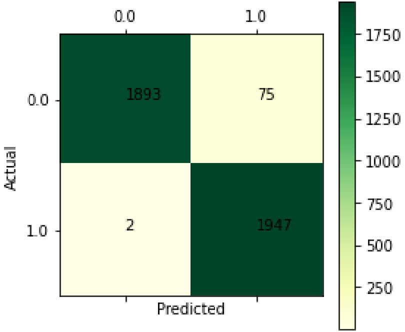


$\Longrightarrow \mathbf{x} = [\texttt{S1WaveformTotal}, \texttt{S2WaveformTotal}, \texttt{S2Pattern}]$

Two distinct quanta: **Electron Recoil** (ER) and **Nuclear Recoil** (NR)

Classification: ER vs. NR Results

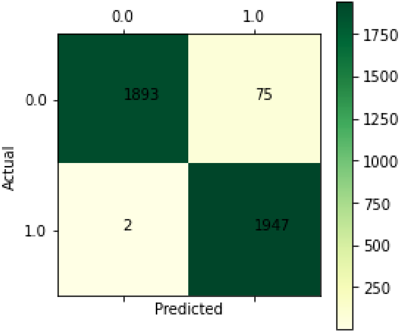
- Train on $\sim$ 40000 images. Take testing sub-sample of $\sim$ 40%
- Check performance $\rightarrow$ confusion matrix:



- Takeaway $\Rightarrow$ **98.03% accuracy**. (Recall = 98.07%, Precision = 96.39%)
- **This works regardless mass and cross-section: NR/ER are what matter.**

Classification: ER vs. NR Results

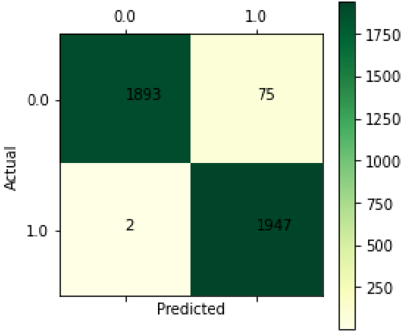
- Train on $\sim$ 40000 images. Take testing sub-sample of $\sim$ 40%
- Check performance $\rightarrow$ confusion matrix:



- Takeaway $\Rightarrow$ **98.03% accuracy**. (Recall = 98.07%, Precision = 96.39%)
- **This works regardless mass and cross-section: NR/ER are what matter.**

Classification: ER vs. NR Results

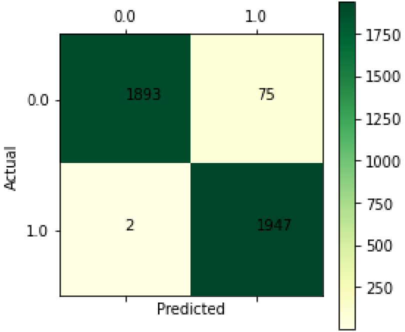
- Train on $\sim$ 40000 images. Take testing sub-sample of $\sim$ 40%
- Check performance $\rightarrow$ confusion matrix:



- Takeaway $\Rightarrow$ **98.03% accuracy**. (Recall = 98.07%, Precision = 96.39%)
- **This works regardless mass and cross-section: NR/ER are what matter.**

Classification: ER vs. NR Results

- Train on $\sim$ 40000 images. Take testing sub-sample of $\sim$ 40%
- Check performance $\rightarrow$ confusion matrix:



- Takeaway $\Rightarrow$ **98.03% accuracy**. (Recall = 98.07%, Precision = 96.39%)
- **This works regardless mass and cross-section: NR/ER are what matter.**

Great! But...

- Trained an effective event-by-event ER veto: In standard DARWIN pipeline, need to sacrifice NR acceptance!
  - Current DARWIN estimate 99.98% ER rejection. 30% NR acceptance.
  - **No need with SBI NN classifier.**
- However, no information regarding the energy of events: WIMPs manifest through number of events + spectral distribution!
- Can we learn the spectral information?

Great! But...

- Trained an effective event-by-event ER veto: In standard DARWIN pipeline, need to sacrifice NR acceptance!
    - Current DARWIN estimate 99.98% ER rejection. 30% NR acceptance.
    - **No need with SBI NN classifier.**
- However, no information regarding the energy of events: WIMPs manifest through number of events + spectral distribution!
- Can we learn the spectral information?

Great! But...

- Trained an effective event-by-event ER veto: In standard DARWIN pipeline, need to sacrifice NR acceptance!
  - Current DARWIN estimate 99.98% ER rejection. 30% NR acceptance.
  - **No need with SBI NN classifier.**
- However, no information regarding the energy of events: WIMPs manifest through number of events + spectral distribution!
- Can we learn the spectral information?

Great! But...

- Trained an effective event-by-event ER veto: In standard DARWIN pipeline, need to sacrifice NR acceptance!
  - Current DARWIN estimate 99.98% ER rejection. 30% NR acceptance.
  - **No need with SBI NN classifier.**
- However, no information regarding the energy of events: WIMPs manifest through number of events + spectral distribution!
- Can we learn the spectral information?

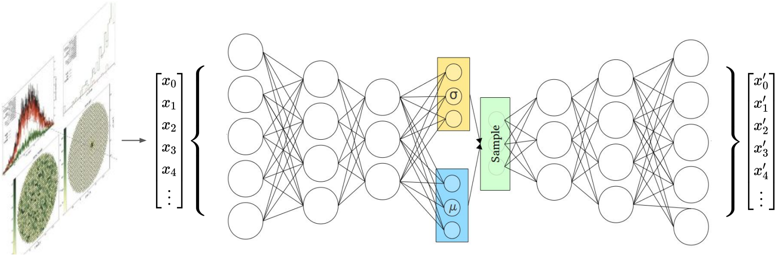
Great! But...

- Trained an effective event-by-event ER veto: In standard DARWIN pipeline, need to sacrifice NR acceptance!
    - Current DARWIN estimate 99.98% ER rejection. 30% NR acceptance.
    - **No need with SBI NN classifier.**
- However, no information regarding the energy of events: WIMPs manifest through number of events + spectral distribution!
- Can we learn the spectral information?

Analysis pipeline 2: Unsupervised approach

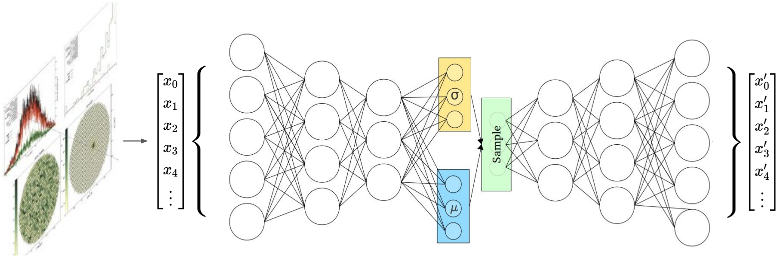
Generative Deep Learning: The Variational Auto-Encoder

- Variety of studies in HEP use these for anomaly detection tasks.
- Goal: Learn low dimensional representation (encoding) of data via dimensional reduction.
- Latent space (bottleneck) layer is a bunch of normal distributions parameterized by some $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.
- Our goal: Learn the latent representation of the background (ER) events. $\Rightarrow$ Spectral information (E).



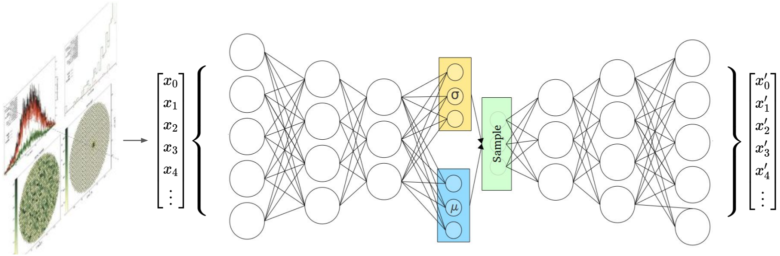
Generative Deep Learning: The Variational Auto-Encoder

- Variety of studies in HEP use these for anomaly detection tasks.
- Goal: Learn low dimensional representation (encoding) of data via dimensional reduction.
- Latent space (bottleneck) layer is a bunch of normal distributions parameterized by some $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.
- Our goal: Learn the latent representation of the background (ER) events. $\Rightarrow$ Spectral information (E).



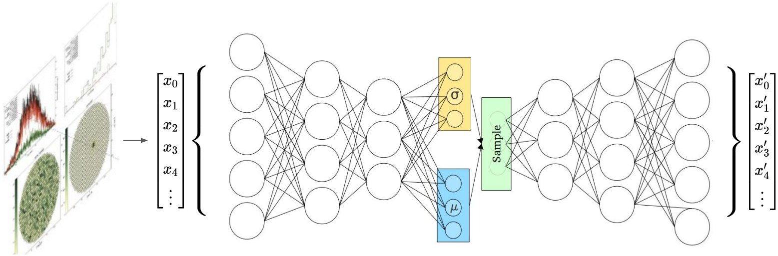
Generative Deep Learning: The Variational Auto-Encoder

- Variety of studies in HEP use these for anomaly detection tasks.
- Goal: Learn low dimensional representation (encoding) of data via dimensional reduction.
- Latent space (bottleneck) layer is a bunch of normal distributions parameterized by some $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.
- Our goal: Learn the latent representation of the background (ER) events. $\Rightarrow$ Spectral information (E).



Generative Deep Learning: The Variational Auto-Encoder

- Variety of studies in HEP use these for anomaly detection tasks.
- Goal: Learn low dimensional representation (encoding) of data via dimensional reduction.
- Latent space (bottleneck) layer is a bunch of normal distributions parameterized by some $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.
- Our goal: Learn the latent representation of the background (ER) events. $\Rightarrow$ Spectral information (E).



Variational-Auto-Encoder: Training

- Use same data as with supervised classification.
- Train VAE on just* ER data.
- Train by maximising evidence lower bound (ELBO):

$$\log p(x) \geq \text{ELBO} = \mathbb{E}_{q(z|x)} \left[\log \frac{p(x,z)}{q(z \mid x)} \right]$$

$$= E[\log p(x|z)] - \beta D_{KL}(q(z|x)||p(z))$$

$$x = \text{Input}$$

$$z = \text{Latent vector}$$

$$\beta = \text{Regularization parameter}$$

- **Loss** $\equiv -\text{ELBO}$

Variational-Auto-Encoder: Training

- Use same data as with supervised classification.
- Train VAE on just* ER data.
- Train by maximising evidence lower bound (ELBO):

$$\log p(x) \geq \text{ELBO} = \mathbb{E}_{q(z|x)}\left[\log \frac{p(x,z)}{q(z \mid x)}\right]$$

$$= E[\log p(x|z)] - \beta D_{KL}(q(z|x)||p(z))$$

$$x = \text{Input}$$

$$z = \text{Latent vector}$$

$$\beta = \text{Regularization parameter}$$

- **Loss** $\equiv -\text{ELBO}$

Variational-Auto-Encoder: Training

- Use same data as with supervised classification.
- Train VAE on just* ER data.
- Train by maximising evidence lower bound (ELBO):

$$\log p(x) \geq \text{ELBO} = \mathbb{E}_{q(z|x)} \left[\log \frac{p(x,z)}{q(z \mid x)} \right]$$

$$= E[\log p(x|z)] - \beta D_{KL}(q(z|x)||p(z))$$

$$x = \text{Input}$$

$$z = \text{Latent vector}$$

$$\beta = \text{Regularization parameter}$$

- **Loss** $\equiv -\text{ELBO}$

Variational-Auto-Encoder: Training

- Use same data as with supervised classification.
- Train VAE on just* ER data.
- Train by maximising evidence lower bound (ELBO):

$$\log p(x) \geq \text{ELBO} = \mathbb{E}_{q(z|x)}\left[\log \frac{p(x,z)}{q(z \mid x)}\right]$$

$$= E[\log p(x|z)] - \beta D_{KL}(q(z|x)||p(z))$$

$$x = \text{Input}$$

$$z = \text{Latent vector}$$

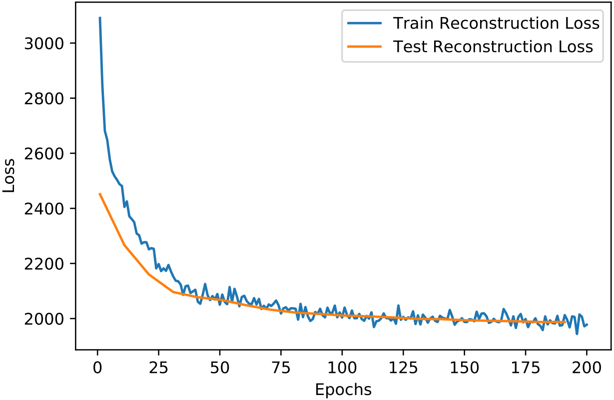
$$\beta = \text{Regularization parameter}$$

- **Loss** $\equiv -\text{ELBO}$

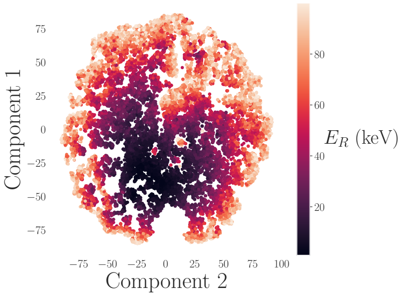
VAE: Training

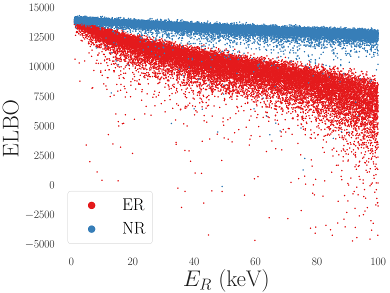
- Train the network for 200 epochs.

$$Loss \equiv -\text{ELBO}$$



Spectral information

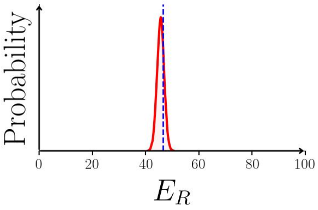


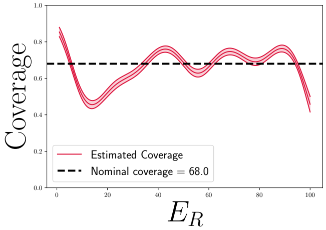


- Auto-encoder can learn underlying spectral information of events $\Rightarrow$ Sensitivity to WIMP mass.
- Can we also just fully reconstruct the energy of an event straight from the data? Yes!

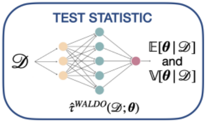
Follow up work: E reconstruction

Neural posterior density estimation (Masked auto-regressive flows)

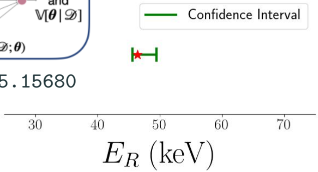


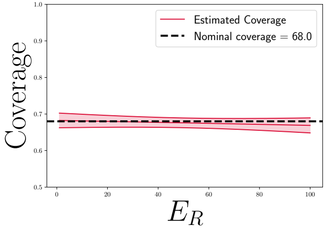


Neural posterior density estimation + WALDO



arXiv:2205.15680





Looking for non background-like events

Anomaly detection

- Traditionally unsupervised methods have been used.
- **Anomaly Detection**: Once trained, run data the network has never seen before through trained network.
- If VAE has learned the underlying properties of ER bkg events, any **non-background** events will in general have **higher loss** (smaller ELBO).
- Loss distribution of anomalous data (new physics) will show as an **excess** over background only loss distribution...

Anomaly detection

- Traditionally unsupervised methods have been used.
- **Anomaly Detection**: Once trained, run data the network has never seen before through trained network.
- If VAE has learned the underlying properties of ER bkg events, any **non-background** events will in general have **higher loss** (smaller ELBO).
- Loss distribution of anomalous data (new physics) will show as an **excess** over background only loss distribution...

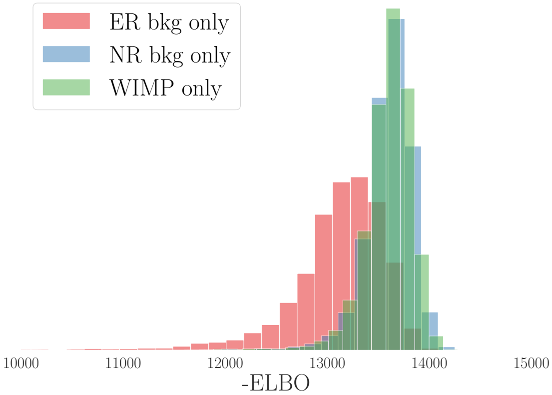
Anomaly detection

- Traditionally unsupervised methods have been used.
- **Anomaly Detection**: Once trained, run data the network has never seen before through trained network.
- If VAE has learned the underlying properties of ER bkg events, any **non-background** events will in general have **higher loss** (smaller ELBO).
- Loss distribution of anomalous data (new physics) will show as an **excess** over background only loss distribution...

Anomaly detection

- Traditionally unsupervised methods have been used.
- **Anomaly Detection**: Once trained, run data the network has never seen before through trained network.
- If VAE has learned the underlying properties of ER bkg events, any **non-background** events will in general have **higher loss** (smaller ELBO).
- Loss distribution of anomalous data (new physics) will show as an **excess** over background only loss distribution...

Anomaly detection



- Background loss distributions + WIMP loss distribution.
- Any* anomalous signal will show up as statistical deviation in (pseudo)data loss vs. (known) background loss.

Semi-unsupervised anomaly detection: New distance metric

Cool. But...

- A bit rubbish: Can we get greater separation (anomaly awareness) between these distributions?
- New 'anomaly function' that utilizes pre-trained supervised NN classifier:

$$TS = -ELBO + R\,H_B\ ,$$

where

- $H_B = -\frac{1}{N}\sum_{i=0}^{N}\log\left(1 - p\left(x_i\right)\right)$ (Binary cross-sentropy.)
- R scales the contribution of the cross-entropy term $\rightarrow$ makes it more/less supervised.

Semi-unsupervised anomaly detection: New distance metric

Cool. But...

- A bit rubbish: Can we get greater separation (anomaly awareness) between these distributions?
- New 'anomaly function' that utilizes pre-trained supervised NN classifier:

$$TS = -ELBO + R\,H_B\ ,$$

where

- $H_B = -\frac{1}{N}\sum_{i=0}^{N}\log\left(1 - p\left(x_i\right)\right)$ (Binary cross-sentropy.)
- R scales the contribution of the cross-entropy term $\rightarrow$ makes it more/less supervised.

Semi-unsupervised anomaly detection: New distance metric

Cool. But...

- A bit rubbish: Can we get greater separation (anomaly awareness) between these distributions?
- New 'anomaly function' that utilizes pre-trained supervised NN classifier:

$$TS = -ELBO + R\,H_B\ ,$$

where

- $H_B = -\frac{1}{N}\sum_{i=0}^{N}\log\left(1 - p\left(x_i\right)\right)$ (Binary cross-sentropy.)
- R scales the contribution of the cross-entropy term $\rightarrow$ makes it more/less supervised.

Semi-unsupervised anomaly detection: New distance metric

Cool. But...

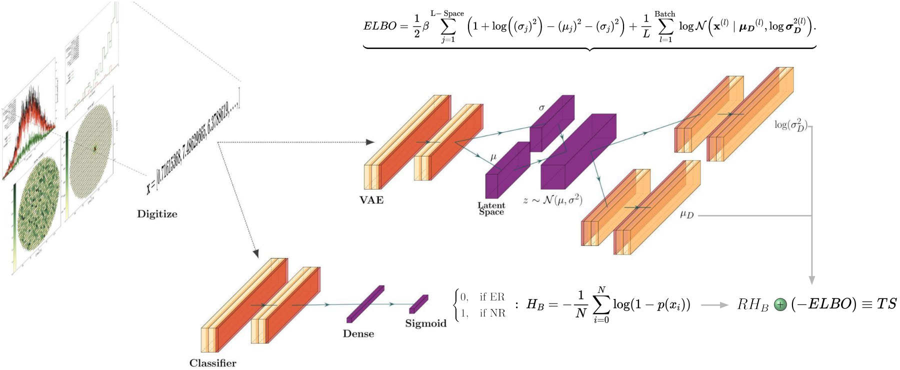
- A bit rubbish: Can we get greater separation (anomaly awareness) between these distributions?
- New 'anomaly function' that utilizes pre-trained supervised NN classifier:

$$TS = -ELBO + R\,H_B\ ,$$

where

- $H_B = -\frac{1}{N}\sum_{i=0}^{N}\log\left(1 - p\left(x_i\right)\right)$ (Binary cross-sentropy.)
- R scales the contribution of the cross-entropy term $\rightarrow$ makes it more/less supervised.

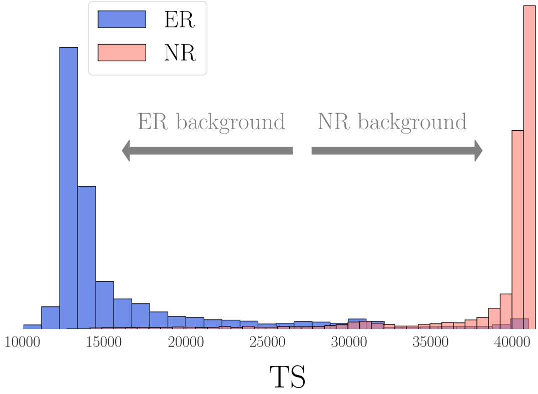
Semi-unsupervised anomaly detection: Full pipeline



Semi-unsupervised anomaly detection: New distance metric

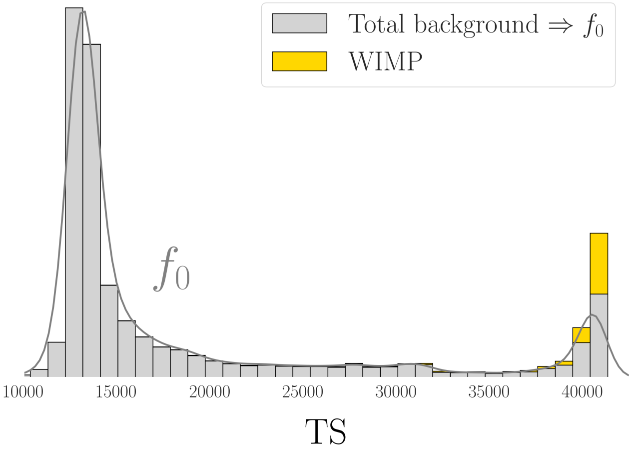
$$TS = (-ELBO) + R\,H_B\ ,$$

⇒ Semi-unsupervised. Much greater anomaly awareness!



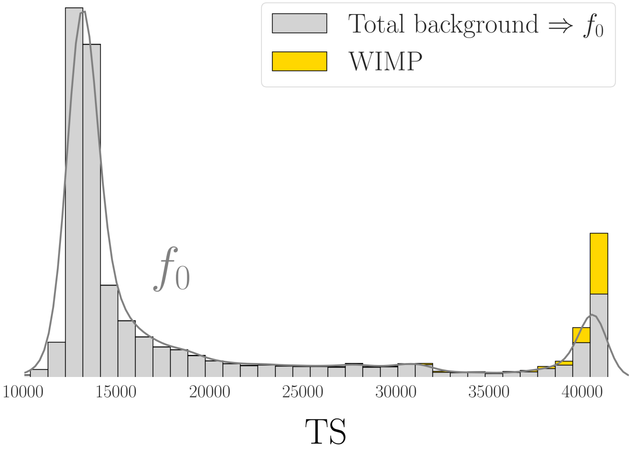
Pseudo-data sets

Re-weight anomaly score distributions TS according to expected ER+NR backgrounds and inject some WIMP signal: ER [2-10] keVee, NR [5-35] keVnr



Pseudo-data sets

Re-weight anomaly score distributions TS according to expected ER+NR backgrounds and inject some WIMP signal: ER [2-10] keVee, NR [5-35] keVnr



Dimensionally reduced analysis

- $\Rightarrow$ 1D analysis in TS space: Accept/reject
  $\mathcal{H}_0 : X \sim \mathcal{P}(x \mid \text{No signal})$.

$$\mathcal{L}(\mathbf{TS}|\mathcal{H}_0) \propto e^{-B} \prod_{i=1}^{N} \left(B f_0\left(TS_i\right)\right)$$

- Unbinned.
- Parametrically independent on WIMP model.
- No auxiliary terms required assuming simulations have suitably descriptive coverage.
- Capability to conduct ER only searches with same machinery.
- Backgrounds currently taken from templates: In principal can propagate uncertainties on the bkg from simulation (or even better, calibration).

Dimensionally reduced analysis

- $\Rightarrow$ 1D analysis in TS space: Accept/reject $\mathcal{H}_0 : X \sim \mathcal{P}(x \mid \text{No signal})$.

$$\mathcal{L}(\mathbf{TS}|\mathcal{H}_0) \propto e^{-B} \prod_{i=1}^{N} (B f_0 (TS_i))$$

- Unbinned.
- Parametrically independent on WIMP model.
- No auxiliary terms required assuming simulations have suitably descriptive coverage.
- Capability to conduct ER only searches with same machinery.
- Backgrounds currently taken from templates: In principal can propagate uncertainties on the bkg from simulation (or even better, calibration).

Dimensionally reduced analysis

- $\Rightarrow$ 1D analysis in TS space: Accept/reject $\mathcal{H}_0 : X \sim \mathcal{P}\left(x \mid \text{No signal}\right)$.

$$\mathcal{L}(\mathbf{TS}|\mathcal{H}_0) \propto e^{-B} \prod_{i=1}^{N} \left(B f_0\left(TS_i\right)\right)$$

- Unbinned.
- Parametrically independent on WIMP model.
- No auxiliary terms required assuming simulations have suitably descriptive coverage.
- Capability to conduct ER only searches with same machinery.
- Backgrounds currently taken from templates: In principal can propagate uncertainties on the bkg from simulation (or even better, calibration).

Dimensionally reduced analysis

- $\Rightarrow$ 1D analysis in TS space: Accept/reject $\mathcal{H}_0 : X \sim \mathcal{P}(x \mid \text{No signal})$.

$$\mathcal{L}(\mathbf{TS}|\mathcal{H}_0) \propto e^{-B} \prod_{i=1}^{N} (B f_0 (TS_i))$$

- Unbinned.
- Parametrically independent on WIMP model.
- No auxiliary terms required assuming simulations have suitably descriptive coverage.
- Capability to conduct ER only searches with same machinery.
- Backgrounds currently taken from templates: In principal can propagate uncertainties on the bkg from simulation (or even better, calibration).

Dimensionally reduced analysis

- $\Rightarrow$ 1D analysis in TS space: Accept/reject $\mathcal{H}_0 : X \sim \mathcal{P}(x \mid \text{No signal})$.

$$\mathcal{L}(\mathbf{TS}|\mathcal{H}_0) \propto e^{-B} \prod_{i=1}^{N} \left(B f_0 \left(TS_i\right)\right)$$

- Unbinned.
- Parametrically independent on WIMP model.
- No auxiliary terms required assuming simulations have suitably descriptive coverage.
- Capability to conduct ER only searches with same machinery.
- Backgrounds currently taken from templates: In principal can propagate uncertainties on the bkg from simulation (or even better, calibration).

Dimensionally reduced analysis

- $\Rightarrow$ 1D analysis in TS space: Accept/reject $\mathcal{H}_0 : X \sim \mathcal{P}\left(x \mid \text{No signal}\right)$.

$$\mathcal{L}(\mathbf{TS}|\mathcal{H}_0) \propto e^{-B} \prod_{i=1}^{N} \left(B f_0\left(TS_i\right)\right)$$

- Unbinned.
- Parametrically independent on WIMP model.
- No auxiliary terms required assuming simulations have suitably descriptive coverage.
- Capability to conduct ER only searches with same machinery.
- Backgrounds currently taken from templates: In principal can propagate uncertainties on the bkg from simulation (or even better, calibration).

Forecasting sensitivity

Median sensitivity

- Probability to accept/reject $\mathcal{H}_0$ after some exposure.
- Model independent.
- Simulate $\sim 10^4$ realisations of $-2\ln\mathcal{L}(\mathbf{TS}\,|\,\mathcal{H}_0)$ to ascertain the asymptotic form of $\mathcal{H}_0$.

$$p = \int_{q_{\rm med}}^{\infty} dq\;\mathcal{H}_0(q)\;.$$

Median sensitivity

- Probability to accept/reject $\mathcal{H}_0$ after some exposure.
- Model independent.
- Simulate $\sim 10^4$ realisations of $-2\ln\mathcal{L}(\mathbf{TS}\,|\,\mathcal{H}_0)$ to ascertain the asymptotic form of $\mathcal{H}_0$.

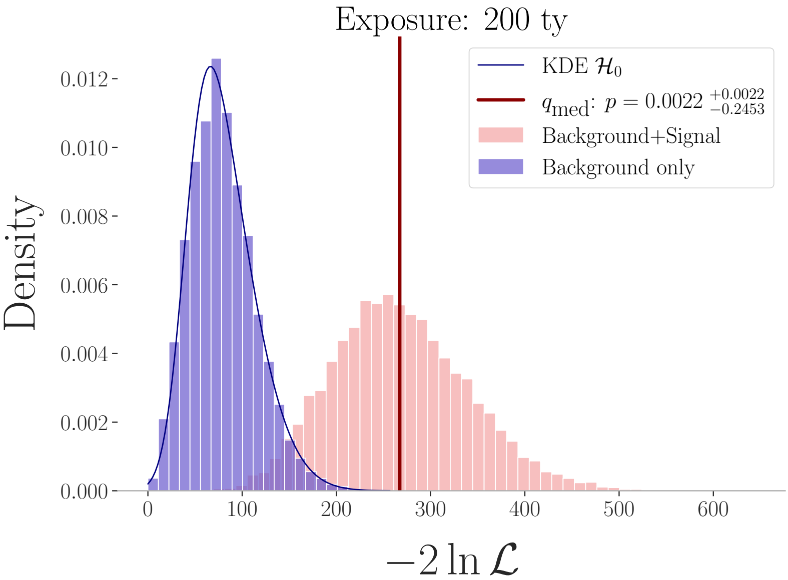
$$p = \int_{q_{\mathrm{med}}}^{\infty} dq\, \mathcal{H}_0(q)\ .$$

Median sensitivity

- Probability to accept/reject $\mathcal{H}_0$ after some exposure.
- Model independent.
- Simulate $\sim 10^4$ realisations of $-2\ln\mathcal{L}(\mathbf{TS}\,|\,\mathcal{H}_0)$ to ascertain the asymptotic form of $\mathcal{H}_0$.

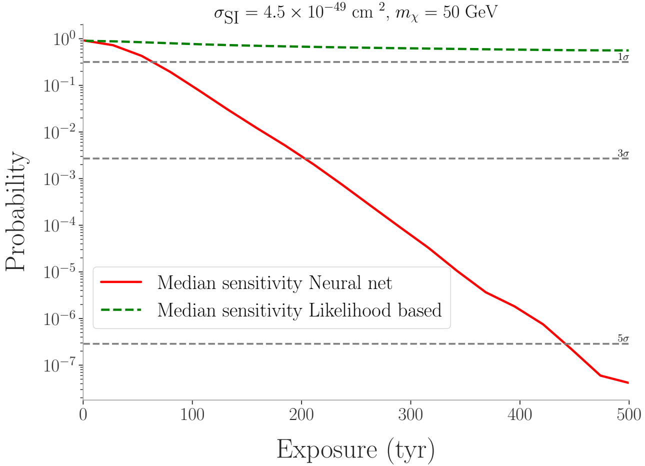
$$p = \int_{q_{\rm med}}^{\infty} dq\; \mathcal{H}_0(q)\;.$$

Median Sensitivity

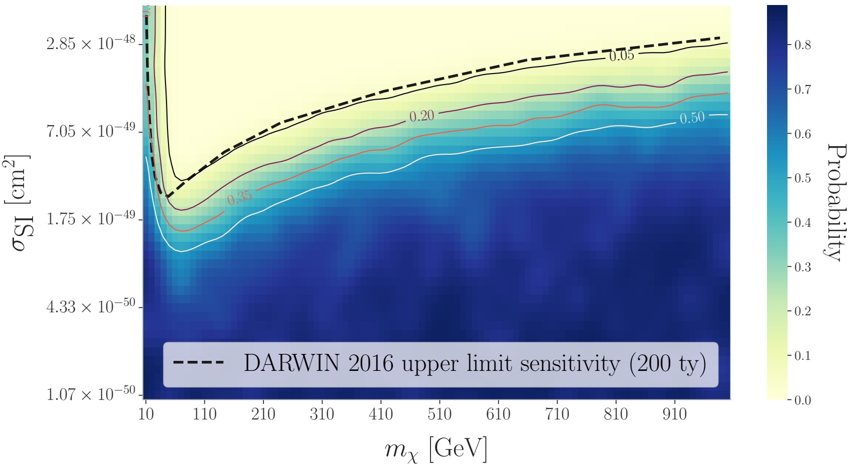


As a function of exposure

- Neural net
- Binned likelihood based: Median sensitivity [30% NR acceptance, 99% ER rejection]



Full sensitivity

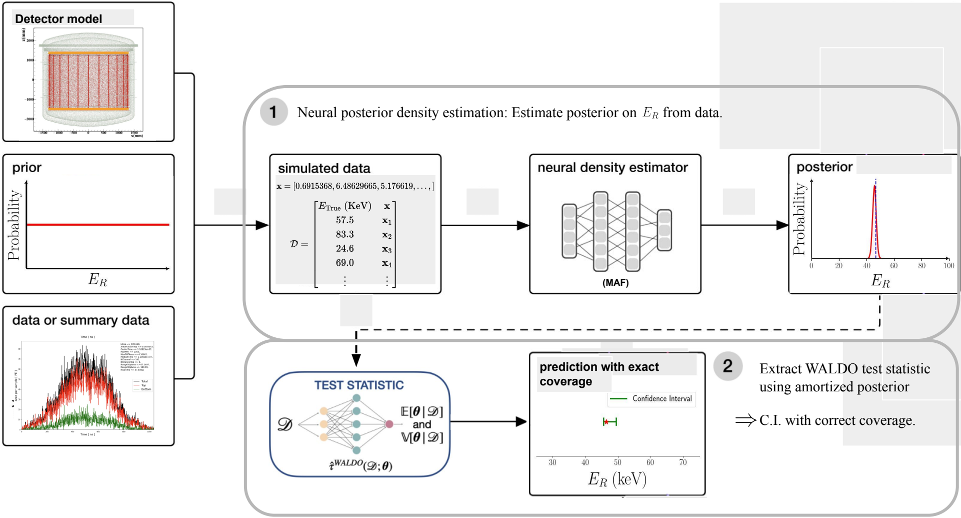


Caution: 90% C.L upper limit is model dependent → 'weaker' test.

Thank you!

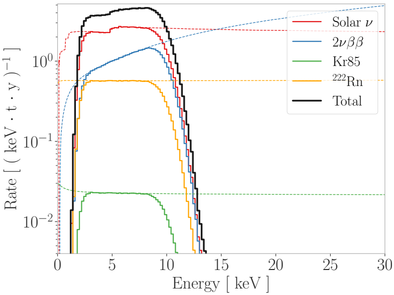
Backup Slides

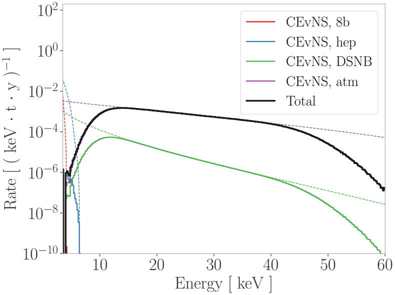
Energy reconstruction SBI with masked autoregressive flows



$$\tau^{\text{WALDO}}(\mathcal{D};\boldsymbol{\theta}_0) = \left(\mathbb{E}[\boldsymbol{\theta} \mid \mathcal{D}] - \boldsymbol{\theta}_0\right)^T \mathbb{V}[\boldsymbol{\theta} \mid \mathcal{D}]^{-1} \left(\mathbb{E}[\boldsymbol{\theta} \mid \mathcal{D}] - \boldsymbol{\theta}_0\right)$$

Backgrounds

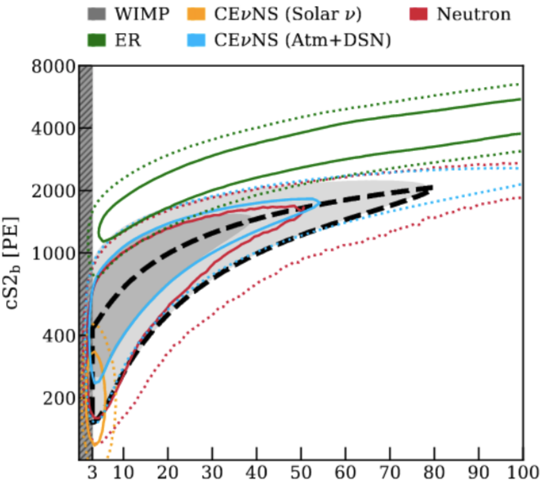




- Intrinsic and extrinsic.
- Coherent neutrino scattering provides dominant background for WIMP searches.

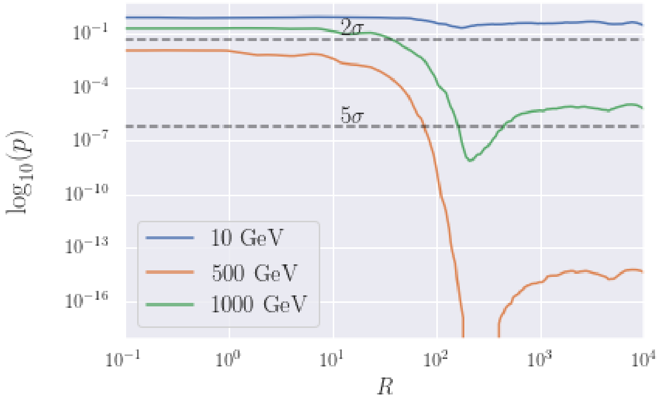
Binned likelihood based approach

$$\mathcal{L}(\mathbf{x}) = \prod_{i=1}^{bins} \frac{{\lambda_i}^{n_i}}{n_i!} e^{-\lambda_i} \quad : \text{ ER veto (99.98\%), fidiucilization etc.}$$



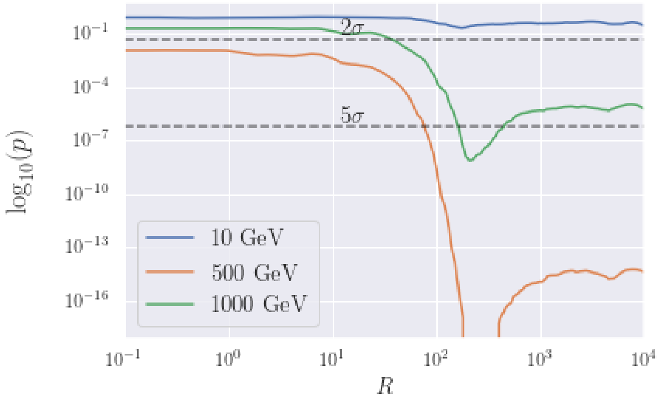
Effect of R

- Explore effect of the R parameter.
- Three mock data sets corresponding to 10, 500 and 1000 GeV at fixed $\sigma = 10^{-45}\text{cm}^2$, 5 $t{\cdot}$yr exposure.
- Best result for $R \sim 170$, but generally free to choose!



Effect of R

- Explore effect of the R parameter.
- Three mock data sets corresponding to 10, 500 and 1000 GeV at fixed $\sigma = 10^{-45}\text{cm}^2$, 5 $t{\cdot}$yr exposure.
- Best result for $R \sim 170$, but generally free to choose!



Effect of R

- Explore effect of the R parameter.
- Three mock data sets corresponding to 10, 500 and 1000 GeV at fixed $\sigma = 10^{-45}\text{cm}^2$, 5 $t{\cdot}$yr exposure.
- Best result for $R \sim 170$, but generally free to choose!