

Simulation based inference for efficient theory space sampling: An application to supersymmetric explanations of the anomalous muon $g - 2$

Logan Morrison¹, Stefano Profumo¹, Nolan Smyth, and John Tamanas^{1*}

*Department of Physics and Santa Cruz Institute for Particle Physics, University of California,
Santa Cruz, California 95064, USA*

 (Received 10 May 2022; accepted 29 November 2022; published 14 December 2022)

For the purpose of minimizing the number of sample model evaluations, we propose and study algorithms that utilize (sequential) versions of likelihood-to-evidence ratio neural estimation. We apply our algorithms to a supersymmetric interpretation of the anomalous muon magnetic dipole moment in the context of a phenomenological minimal supersymmetric extension of the standard model, and recover nontrivial models in an experimentally constrained theory space. Finally we summarize further potential possible uses of these algorithms in future studies.

DOI: [10.1103/PhysRevD.106.115016](https://doi.org/10.1103/PhysRevD.106.115016)

I. INTRODUCTION

The pursuit of understanding beyond the Standard Model (BSM) physics theories in the context of experimental results is the cornerstone of much research in high energy physics (HEP). However, despite the increasing accuracy of weak-scale observables, it remains nontrivial to perform the inverse calculation, i.e., to determine regions of parameter space yielding the observables.

Mapping out regions of parameter space that produce given observables is especially difficult for theories with many free parameters—a famous example being the minimal supersymmetric model (MSSM) (see, e.g., [1] for a review) and downsized versions thereof, in parameter space, such as the phenomenological MSSM (pMSSM) [2,3]. The exploration of the parameter space of these theories is typically carried out, in the HEP literature, by defining a search space in the space of pMSSM parameters, uniformly sampling it, and then rejecting all samples which do not conform with theoretical or experimental constraints (see, e.g., [4–6], albeit there is a very vast literature on the topic). However, the number of required samples grows exponentially with the dimension of the search space, so these methods are extremely computationally intensive, if not intractable, when calculations of the output predictions for a given set of input parameters are slow.

There has been a recent upsurge of developments in the broader scientific literature regarding the inverse problem

of restricting parameter spaces of a forward model purely through sampling (i.e., calculating observables from model parameters; see, e.g., [7–9] for a few recent examples). The introduction of neural networks to simulation-based inference (SBI) frameworks, in particular, has led to explosive improvements in accuracy and precision and ablation studies comparing them all [10–13].

In this work, we introduce the sequential neural ratio estimation (SNRE) algorithm to the problem of sampling from an experimentally constrained version of the pMSSM. Although not explicitly stated, recent work [14] has used methods from the simulation-based inference literature—namely neural likelihood estimation. We show how that method fits into the larger SBI framework and discuss an alternative, though related, approach that uses likelihood-to-evidence ratio neural estimation. This fits in the context of “likelihood-free” methods in which drawing samples from a forward model is possible, but the likelihood is expensive or intractable to evaluate. Additionally, we demonstrate sequential versions of two algorithms that can significantly reduce the number of model evaluations required.

Specifically, here we are interested in the following example case study: we perform sampling of the pMSSM parameter space, producing experimentally viable predictions for (i) the relic density of dark matter ($\Omega_\chi h^2$) in the form of the lightest neutralino, which has mass M_χ , (ii) the Higgs mass (m_h), (iii) the anomalous magnetic moment of the muon (a_μ), and (iv) WIMP-nucleon cross sections (σ^{SI}).

We organize the manuscript as follows: Sec. II details the general SBI framework and describes the SNRE algorithm in particular; Sec. III details the search space of the pMSSM to which we will be applying SNRE; in Sec. IV we demonstrate the improved sampling efficiency of sequential algorithms; we show an application of SNRE to an experimentally constrained pMSSM parameter space in

*jtamanas@ucsc.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

Sec. V; and finally, in Sec. VI we summarize our findings and further discuss potential future uses of SBI.

II. SIMULATION-BASED INFERENCE

This work aims to use the simulation-based inference (SBI) framework to sample from a large-dimensional parameter space while evaluating the model output as few times as possible for computational efficiency. There are essentially four overarching approaches to SBI: rejection sampling with approximate Bayesian computation (ABC), posterior estimation (NPE), likelihood-to-evidence ratio estimation (NRE), and likelihood estimation (NLE). For a full review of simulation-based inference and other methods therein, we refer the reader to Ref. [15].

We begin by stating Bayes' rule:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}, \quad (1)$$

where $P(X|\theta)$ is the likelihood of the data given the model parameters, $P(\theta)$ is the prior probability of the model parameters, and $P(X)$ is the evidence. In this work, θ refers to supersymmetric parameters, and X refers to various observables of our choosing, e.g., the Higgs mass, the DM relic density, and the muon anomalous magnetic moment.

The NRE approach begins by approximating the likelihood-to-evidence ratio, $r(X, \theta) = \frac{P(X|\theta)}{P(X)}$, and then using it to sample from the posterior distribution of the model parameters with Hamiltonian Monte Carlo (HMC). HMC can be thought of as a Markov Chain Monte Carlo (MCMC) [16,17] algorithm where a chain's position and momentum are sampled from a joint distribution and then evolved according to Hamilton's equations. HMC makes use of derivatives to calculate its transitions, which makes it not directly applicable using the `micrOMEGAS` [18] package, but it can still be used on the neural network side. For a full review of HMC, see [19].

A. Likelihood-to-evidence ratio estimation

Here, we will follow the approach of Ref. [12] to approximate the likelihood-to-evidence ratio: Assume we have two classes of simulations, Y_0 and Y_1 , each of which is composed of pairs of $\{X, \theta\}$. We define Y_0 and Y_1 such that $Y_0 = \{X, \theta\} \sim P(X)P(\theta)$ and $Y_1 = \{X, \theta\} \sim P(X|\theta)$. We use curly brackets to denote a set of values. The optimal classifier (one which minimizes the binary cross-entropy loss) between Y_0 and Y_1 is given by

$$d^*(X, \theta) = \frac{P(X|\theta)}{P(X|\theta) + P(X)P(\theta)} \quad (2)$$

which can, in turn, be used to express the likelihood-to-evidence ratio:

$$r^*(X, \theta) = \frac{P(X|\theta)}{P(X)} = \frac{P(X, \theta)}{P(X)P(\theta)} = \frac{d^*(X, \theta)}{1 - d^*(X, \theta)}, \quad (3)$$

where we used the fact that the joint distribution $P(X, \theta) = P(X|\theta)P(\theta)$. Thus, approximating $r^*(X, \theta)$ is equivalent to finding the optimal classifier $d^*(X, \theta)$, which we can accomplish by training a classifier to distinguish between Y_0 and Y_1 . For the rest of this work, we shall use $r(X, \theta)$ to denote the approximate likelihood-to-evidence ratio.

In practice, we produce samples of Y_1 by first drawing θ from the prior, $P(\theta)$, and then passing it through our simulator (the `micrOMEGA` package) to generate X . We begin sampling from Y_0 in the same way: we draw θ from the prior and plug it into our simulator to generate X . After this, however, we *resample* $\theta' \sim P(\theta)$ and use the new value in its place, i.e., $(X, \theta') \sim Y_1$.

We will use a neural network to model $r(X, \theta)$, which allows us to take advantage of the vast literature and resources available for deep learning applications. We will use an ensemble of multilayer perceptrons (MLPs). Ensembling models and taking their average was shown in [20] to produce more accurate, conservative posteriors than would be obtained by simply training a single model.

Although not applicable for this work, it is worth pointing out that the data can be of virtually any format. For example, the authors of [12] have shown the ability of this algorithm to ingest 128×128 pixel images of strongly lensed galaxies and produce accurate posteriors of its Einstein radius.

B. Sequential neural likelihood-to-evidence ratio estimation

If the prior is not carefully crafted to be a good approximation of the posterior to start with, we expect many of the samples to produce observables far from the experimental values of interest. These can be seen as wasted evaluations of the model, the very thing we are trying to limit. However, with a trained NRE, we have a better approximation of the posterior available to us. We therefore elect to iterate the training procedure outlined in the previous section, but replace the sampling of θ from the prior with a sampling from the intermediate posterior. The full algorithm is outlined in Algorithm 1.

Empirically, numerical experiments show that sequential versions of SBI tend to converge to the true posteriors with fewer simulator evaluations than their nonsequential counterparts [13]. Intuitively, this is because the approximate posterior will be most accurate in the densest regions of the parameter space *as specified by the training data*. However, when we are sampling from the posterior, we are most interested in its regions of highest probability. Therefore, we can reduce the number of model evaluations required by iteratively closing in on the true posterior. We demonstrate this in Sec. IV where we run trials to determine how

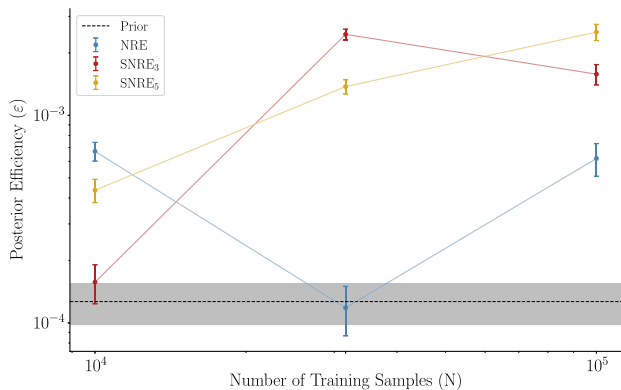


FIG. 1. The fraction of points sampled from the posterior that lie within the ranges specified in Eq. (8). The gray band indicates the efficiency of sampling directly from the prior for comparison with standard MCMC based approaches.

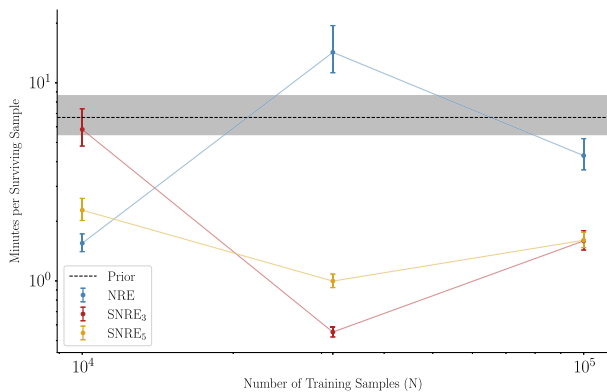


FIG. 2. The expected amount of time (minutes) needed to obtain one sample in the ranges specified in Eq. (8).

efficient (the fraction of viable simulations) naive, NRE, and SNRE sampling, respectively, are. Our results are summarized in Figs. 1 and 2.

Finally, it is important to note that either version of SBI is not guaranteed to converge to the true posterior and can often produce overconfident bounds on parameter space with exceedingly high computational budgets required to calibrate [20]. By using an ensemble of classifiers, which we do for all of our applications, we partially alleviate this problem.

III. PHENOMENOLOGICAL MINIMAL SUPERSYMMETRIC MODEL

The fully general, unconstrained minimal supersymmetric extension to the Standard Model (MSSM) contains 105 free parameters [21], principally connected with the soft supersymmetry breaking pattern. By enforcing no flavor changing neutral currents, no new sources of CP violation, and first and second generation universality, these parameters are restricted to a 19-dimensional parameter space,

Algorithm 1. Sequential Neural Likelihood-to-Evidence Ratio Estimation (SNRE).

```

Let  $p_r(\theta|X_0)$  be the posterior in round  $r$ .
Let  $p_{r=0}(\theta|X_0) = P(\theta)$ .
while  $r < R$  do
  set prior to  $p_r(\theta|X_0)$ 
  while  $n \leq N$  do
    Sample  $\theta_n \sim p_r(\theta|X_0)$ 
    Sample  $X_i \sim P(X|\theta_n)$ 
    Add  $\{\theta_n, X_n\}$  to  $\mathcal{D}$ 
  end while
  while  $d \leq D$  do
    Sample  $\{\theta_A, X_A\}_d \sim \mathcal{D}$ 
    Sample  $\{\theta_B, X_B\}_d \sim \mathcal{D}$ 
    Assign labels  $y = 1$  for  $\{\theta_A, X_A\}_d$  and  $\{\theta_B, X_B\}_d$ 
    Assign labels  $y = 0$  for  $\{\theta_B, X_A\}_d$  and  $\{\theta_A, X_B\}_d$ 
  end while
  (re-)train  $r(X, \theta)$  to classify  $\{\theta_A, X_A\}_D$ ,  $\{\theta_B, X_B\}_D$ ,
   $\{\theta_B, X_A\}_D$ , and  $\{\theta_A, X_B\}_D$  by minimizing the binary
  cross-entropy loss.
   $p_r(\theta|X_0) = r(X, \theta)P(\theta)$ 
end while

```

known as the “phenomenological” MSSM (pMSSM) [22]. Here, since we intend to illustrate the algorithms outlined above as applied to the pMSSM in connection with supersymmetric contributions to the anomalous magnetic moment of the muon, we entertain the possibility that the supersymmetric scalar particles associated with second generation leptons, smuons, be lighter than, and not as usually assumed degenerate with, their first generation counterparts, selectrons. Therefore, we relax the universality assumption in the slepton sector of the pMSSM, which adds two more parameters to the parameter space of interest. To simplify, and since they are largely irrelevant to quantities of interest, especially the muon anomalous magnetic moment, we assume all squark masses to be degenerate at a relatively large mass scale set to 4 TeV, large enough so that they have negligible effects in our calculations.

To utilize the SBI framework, one must have a forward model of the likelihood, i.e., a function which takes in a set of parameters, θ , and produces an observable, X , from some underlying distribution. The underlying distribution need not be known explicitly, for our goal is to learn it from the samples. Additionally, we must provide a prior distribution of the parameters. For our applications, we will use a uniform prior for the parameters, with bounds chosen to be consistent with the literature and listed in Table I.

In the following applications, we aim to produce posterior distributions conditioned on a combination of observables calculated by the micrOMEGAs package [18]. Although not arbitrarily accurate, micrOMEGAs produces a *deterministic* output, so we assume Gaussian error bars on these calculations. Thus, our likelihood is given by

TABLE I. Domains used for the pMSSM soft supersymmetry breaking parameters. All parameters (aside from $\tan\beta$) are in units of GeV. We set all squark mass parameters to 4 TeV.

Parameter	Domain	Description
$ \mu $	[100, 4000]	Higgs mixing parameter
$ M_1 $	[50, 1000]	Bino mass parameter
$ M_2 $	[100, 4000]	Wino mass parameter
M_3	[400, 4000]	Higgsino mass parameter
M_{L_1}	[100, 4000]	Left-handed selectron mass
M_{L_2}	[100, 1000]	Left-handed smuon mass
M_{L_3}	[100, 4000]	Left-handed stau mass
M_{r_1}	[100, 4000]	Right-handed selectron mass
M_{r_2}	[100, 1000]	Right-handed smuon mass
M_{r_3}	[100, 4000]	Right-handed stau mass
M_A	[400, 4000]	Pseudoscalar Higgs mass
$\tan\beta$	[1, 60]	Ratio of Higgs VEVs v_u/v_d
$ A_t $	[0, 4000]	Trilinear Higgs-stop coupling
$ A_b $	[0, 4000]	Trilinear Higgs-sbottom coupling
$ A_\tau $	[0, 4000]	Trilinear Higgs-stau coupling

$\mathcal{N}(X(\theta), \sigma_X)$, where \mathcal{N} is the normal distribution and $X(\theta) = (\Omega_{\text{DM}}(\theta), a_\mu(\theta), m_h(\theta), p_{\text{Xenon1T}})$ is the vector of calculated observables from micrOMEGAs. The variance of the distribution is $\sigma_X^2 = \sigma_{\text{exp}}^2 + \sigma_{\text{th}}^2$ with theoretical Gaussian uncertainties are:

$$\begin{aligned} \sigma_X &= (\sigma_\Omega, \sigma_{a_\mu^{\text{SUSY}}}, \sigma_{m_h}, \sigma_{p_{\text{Xenon1T}}}) \\ &= (0.01, 65 \times 10^{-11}, 1.0 \text{ GeV}, 10^{-5}). \end{aligned} \quad (4)$$

pMSSM parameters, θ are sampled from a uniform prior, with bounds shown in Table I. When sampling from our posterior, we set the observed values of X to their experimental results [23–27]

$$\begin{aligned} X_0 &= (\Omega_\chi h^2, m_h, a_\mu^{\text{SUSY}}, p_{\text{Xenon1T}}) \\ &= (0.12, 125 \text{ GeV}, 251 \times 10^{-11}, 0.0455). \end{aligned} \quad (5)$$

Here we are assuming that the discrepancy between the SM and experimentally observed anomalous magnetic moment of the muon is solely due to supersymmetric contributions $a_\mu^{\text{SUSY}} = a_\mu^{\text{exp}} - a_\mu^{\text{SM}}$. As Xenon1T has not yet observed WIMP dark matter, we set its p-value equal to $\text{CDF}_{\mathcal{N}}(2\sigma)$ and treat all p-values greater than this value identically via the pre-processing described in Eqs. (A1)–(A4).

IV. BENEFITS FROM SEQUENTIAL TRAINING

A key question in algorithm optimization is to seek the minimization of the time required to obtain N samples from a distribution. Here, we can approximate the amount of time spent computing, T , via

$$T = Ns + Rt + (R - 1)v + w, \quad (6)$$

where N is the total number of micrOMEGAs calculations performed, s is the time per micrOMEGAs evaluation, R is the number of sequential rounds of training, t is the time to train the network in each sequential round, v is the amount of time to sample from the intermediate posteriors during SNRE training, and w is the amount of time to sample from the final SNRE posterior. In this application, we limit R , t , v , w , so that the majority of the time is spent evaluating micrOMEGAs, i.e., $T \approx Ns$. N is often chosen to be large enough so that the number of samples surviving the constraints is larger than some number, M . We can write

$$N = M/\varepsilon + D, \quad (7)$$

where ε is the efficiency of our micrOMEGAs samples, and D is the number of samples used to train our models (for traditional methods, $D = 0$). If the efficiency is sufficiently small, the number of required samples becomes prohibitively large, which is the typical case when sampling θ from the prior. The goal of SBI is to minimize N by maximizing ε . We illustrate this effect by calculating the efficiency of over a range of training sample sizes and SNRE training rounds.

A. Setup

The benefits of the SBI framework become clear when looking at the efficiency of sampling, i.e., the fraction of samples that satisfy the given constraints of the problem. To illustrate this, we perform the pMSSM posterior approximation exercise while enforcing observables to be within 3-sigma of their central values:

$$\begin{aligned} 0.09 &< \Omega_\chi h^2 < 0.15 \\ 122 \text{ GeV} &< m_h < 128 \text{ GeV} \\ 56 \times 10^{-11} &< a_\mu^{\text{SUSY}} < 445 \times 10^{-11} \\ 0.0455 &< p_{\text{Xenon1T}}, \end{aligned} \quad (8)$$

where $a_\mu^{\text{SUSY}} = (g_\mu^{\text{SUSY}} - 2)/2$ is the supersymmetric contribution to the muon anomalous magnetic moment and p_{Xenon1T} is the p-value of the model as determined by Xenon1T.

We choose to run the SBI pipeline on three different choices of R . Specifically we set $R = 1$ for NRE, $R = 3$ for SNRE₃, and $R = 5$ for SNRE₅.

We compare each choice of R with the ‘‘prior’’ baseline, which calculates observables using samples from the prior, and then subjects them to the same filtering as shown in Eq. (8). We emphasize that samples from the prior do not correspond to samples from the posterior, and we only use the samples as a point of comparison.

We allot a budget of 150,000 micrOMEGAs calculations and split them into training sets and inference sets. We use the training sets to train the networks, and the inference sets are the evaluations from which our final samples come.

Ideally, the training sets will be as small as possible to maximize the number of posterior samples in our inference set. However, increasing the size of the training set could increase the efficiency of posterior samples, thus enabling equal performance with the weaker algorithm producing a larger inference set. For each method, we run a trial on a training set of sizes 10,000, 31,000, and 100,000, with the remaining calculations left for inference. For SNRE_R , we distribute the calculations equally among the R rounds.

B. Results

The results demonstrate several notable features. First, in Fig. 1 we observe the expected trend among the sequential algorithms; as the number of points used to train increases, the number of points sampled from the approximate posteriors which lie within the ranges specified in Eq. (8) also increases. However, this is not true for NRE, which performed worse for the medium-sized training set. The absence of this scaling for NRE is likely due to overfitting, which we did not prevent by early stopping. Finally, as the efficiency is highest for SNRE_5 , it is the preferred model in the event that additional computational resources become available after the training phase, e.g., in future follow-up analyses.

We determine the average amount of time from start to finish required to generate a valid sample from the final posterior and display the results in Fig. 2. Although slower overall, we determine that the efficiency of SNRE_R gives it a factor of up to 3 times faster sampling than NRE, even when the inference set is significantly smaller. Our results suggest sequential algorithms are consistently the most efficient when they use $\mathcal{O}(10\%)$ of the allotted calculations for training.

V. APPLICATION TO PMSSM

A. SBI setup

We now examine the posteriors and experimental observables produced by an application to the pMSSM. In this search, we are interested in finding the regions of parameter space most likely to contain the correct relic density, Higgs mass, provide a good fit to the observed anomalous magnetic moment of the muon, and are not excluded by Xenon1T.

Inspired by the results of Sec. IV, we run SNRE for a total of 10 rounds, each with 20,000 new samples calculated with micrOMEGAS. Next, we assemble the likelihood-to-evidence approximators with an ensemble of 5 binary classifiers whose inputs are a concatenation of X and θ and whose outputs are the logit of the classification probability. Each network consists of three hidden layers, each consisting of 256 units. Next, we train networks to minimize the binary cross-entropy loss between samples from the likelihood, $p(X|\theta)$, and the joint probability, $p(X)p(\theta)$.

Finally, we average the ensemble outputs during inference time to produce an approximate likelihood-to-evidence ratio of a given sample. When we sample with HMC during intermediate training steps, we use 20 chains with 2000 warmup steps and 32 chains with 2000 warmup steps during the final inference time.

B. Results

In Fig. 3, we show samples from the approximate posterior as one-dimensional histograms. The multi-modal and irregularly shaped distribution of points demonstrates the ability of flexible neural methods to hone in on these regions. The results exhibit several features previously found in the literature despite not being enforced *a priori* in our search (outside of choosing bounds on the prior.) Additionally, we plot two-dimensional histograms for a subset of the dimensions in Figs. 4 and 5.

These results demonstrate, for example, the learned dependence on the relative signs of μ and the gaugino parameters. In addition, we find that most models have a binolike LSP with light smuons whose masses are near the LSP mass. Physically, these correspond to the bino coannihilation with smuons driving thermal relic abundance. One expects these results as binolike LSPs have smaller WIMP-nucleon cross sections, and lighter sleptons contribute more to a_μ , while also suppressing the relic density. We note the generally larger values of M_{L_2} compared to M_{r_2} . This trend is mostly due to the sneutrino requiring a larger value of M_{L_2} because its mass is typically lighter than the left-handed smuon. The right-handed smuon mass, on the other hand, is generally equal to M_{r_2} .

Of the 96,000 samples generated from the approximated posterior, 179 survive the cuts laid out in Eq. (8), resulting in an efficiency of 0.00186, similar to the results shown in Fig. 1. The surviving points are shown on top of Xenon1T and ATLAS [28,29] constraints in Figs. 6 and 7, respectively. Interestingly, our results indicate that the experiments are complementary: regions of low direct detection constraining power exist in high ATLAS constraining power regions, and vice versa. We expect future analyses to probe the majority of the surviving parameter space.

VI. CONCLUSIONS

We introduced a simulation-based inference framework capable of analyzing high-dimensional parameter spaces and applied the framework to supersymmetric extensions to the Standard Model. We showed how equivalent amounts of computational time result in orders of magnitude more pMSSM models of experimental interest than naive sampling from the prior. Additionally, we demonstrated how sequential sampling methods are even more efficient than their nonsequential counterparts by performing a light scan over hyperparameters.

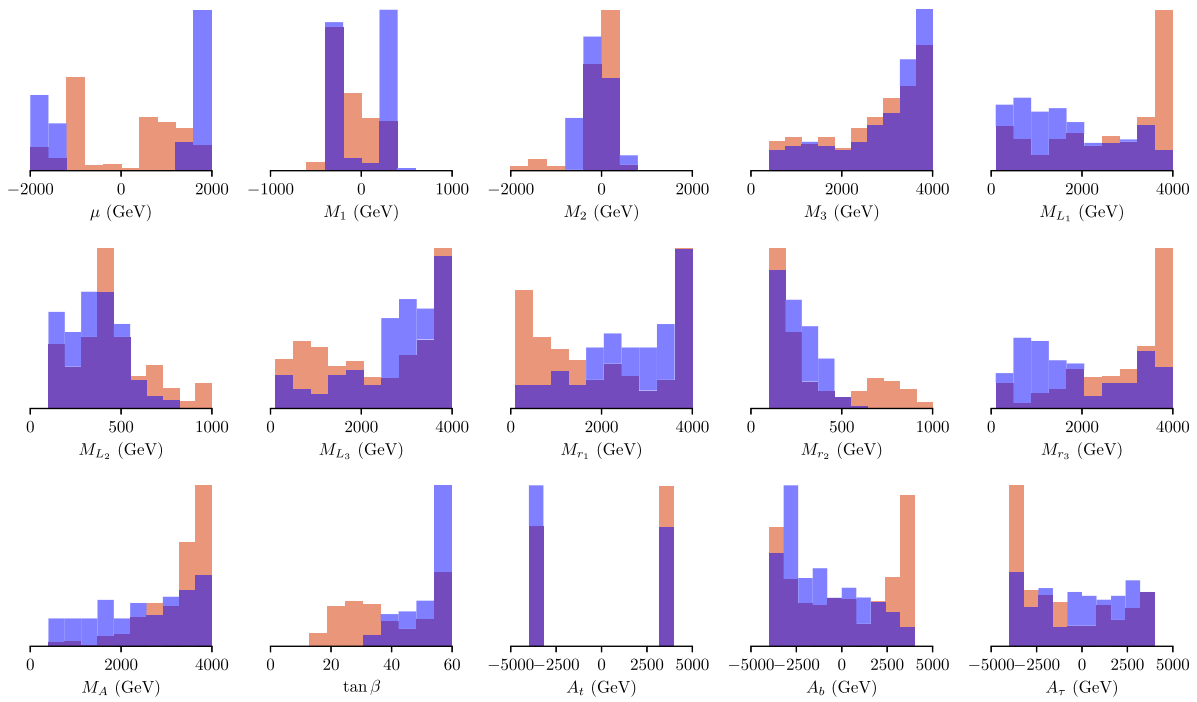


FIG. 3. Samples from the posterior which lie within the experimental constraints specified in Eq. (8) obtained by running SNRE with hyperparameters listed in Table II are shown in salmon. We note the overdensity of small smuon masses which affect a_μ and Ω_χ by coannihilations. See Figs. 5 and 4 for corner plots with subsets of these parameters. Shown in blue are the filtered samples from the posterior for the nonsequential NRE, again trained using the hyperparameters in Table II. We see the SNRE algorithm produces similar coverage for most parameters while providing greater sampling efficiency.

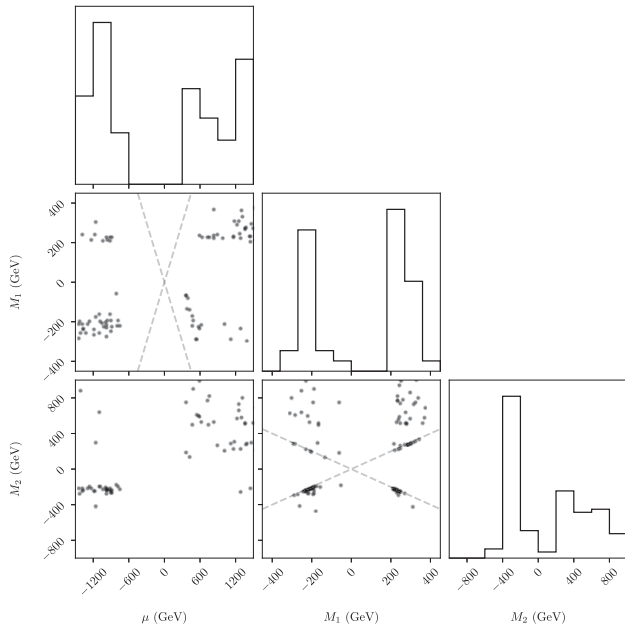


FIG. 4. Corner plots of μ and gaugino mass parameters created from the same samples as Fig. 3. Dotted lines correspond to $\mu = M_1$ and $M_2 - 2 = M_1$, respectively. We see a strong bias toward binolike and winolike LSPs. Direct detection constraints force μ toward larger values of its range. Additionally, we see a large dependence on the relative signs of μ and M_2 .

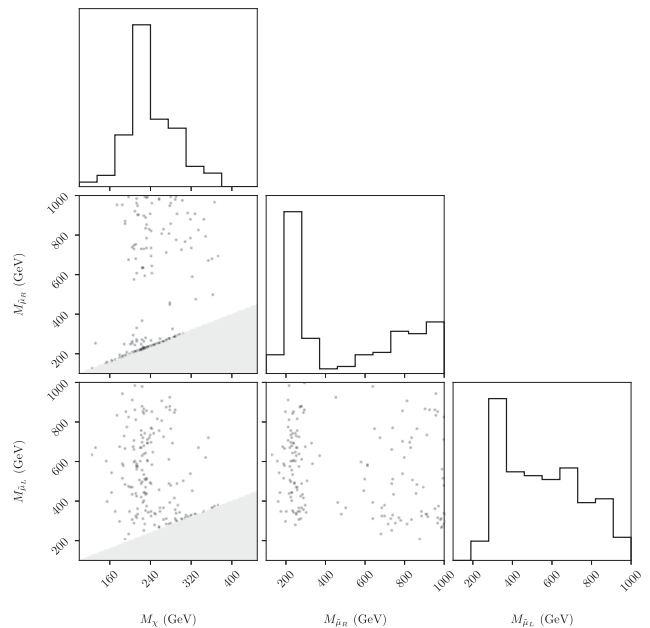


FIG. 5. Corner plots of smuon and dark matter masses calculated from the same samples as Fig. 3. Shaded regions do not have a neutralino LSP which, although not explicitly excluded, are not experimentally viable. We note the large concentration of light right-handed smuons.

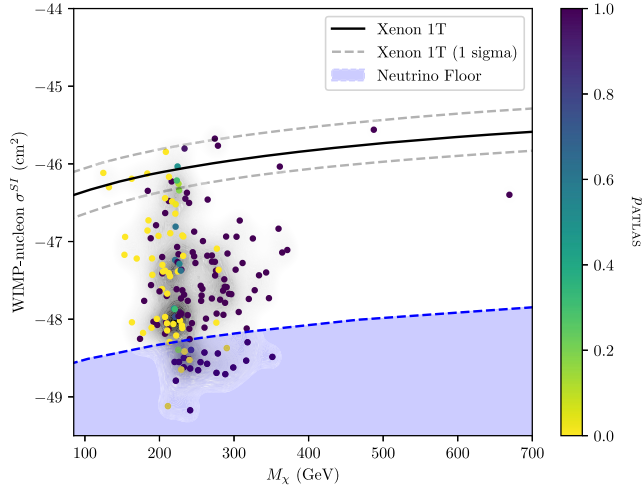


FIG. 6. Samples from the approximate posterior which lie within the experimental constraints specified in Eq. (8). The color of each point corresponds to the p-value from compressed spectra constraints reported by ATLAS. The background is shaded according to a gaussian kernel density estimate in order to visualize the concentration of points on these axes. We note the ability of ATLAS to probe models which lie in the neutrino floor.

Finally, we applied the framework to sample from regions of the pMSSM parameter space yielding the correct thermal dark matter relic density, Higgs mass, and the muon’s anomalous magnetic moment, which direct detection experiments have not yet excluded. We found that the most likely regions of this space are tantalizingly close to current ATLAS bounds and are likely to be covered in future direct dark matter detection experiments.

Future exercises similar to the present one could make use of calibration methods presented in Ref. [30] to avoid overly confident constraints on the approximated posterior and therefore avoid excluding potentially viable regions of parameter space. Additionally, one can consider higher-dimensional observables. Rather than summaries of observed weak-scale quantities, it is potentially of interest to operate directly on the raw data obtained from experiments.

ACKNOWLEDGMENTS

The authors would like to thank Michael Hance, Giordon Stark, Jeff Shahinian, and Michael Holzbock. This work is partly supported by the U.S. Department of Energy Grant No. de-sc0010107 and the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1842400 to N. S.

APPENDIX: TRAINING DETAILS

Prior to entering neural network, is data are preprocessed to restrict the dynamic range of the inputs. For the relic

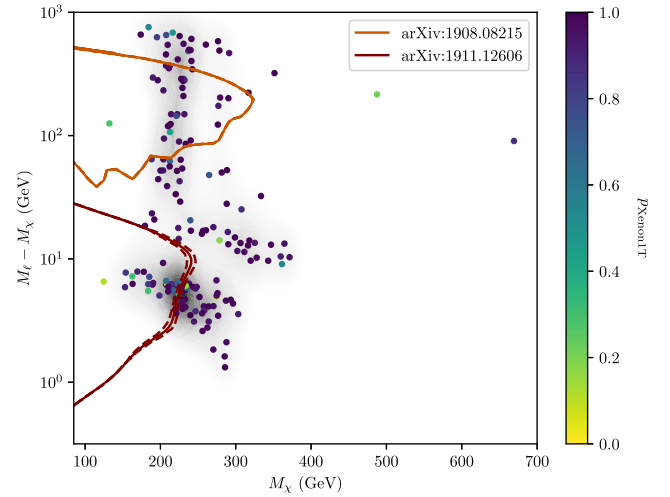


FIG. 7. Same as Fig. 6 but plotted on ATLAS constraints with Xenon1T p-values, as calculated by micromEGAS. The mass splitting is taken with respect to the lightest slepton mass, M_ℓ . Dotted lines correspond to $\pm 1\sigma$ constraints. We expect future analyses to constrain much of the viable parameter space with small mass splittings.

density, we enforce the value to be greater than 10^{-4} , convert to log-space and rescale:

$$\Omega_\chi h^2 \rightarrow 10 \times \log_{10}(\text{clip}(\Omega_\chi h^2, 10^{-4}, \infty)) \quad (\text{A1})$$

The Higgs mass is restricted to be larger than 118 GeV and made $\mathcal{O}(1)$ by subtracting off a value close to its known mass 123.864 GeV and rescaling to be $\mathcal{O}(1)$

$$m_h \rightarrow \frac{\text{clip}(m_h, 118, \infty) - 123.864}{2.2839}. \quad (\text{A2})$$

The SUSY contribution to the muon anomalous magnetic moment is restricted to be within 10^{-11} and 10^{-8} , converted to log-space and made $\mathcal{O}(1)$

$$a_\mu^{\text{SUSY}} \rightarrow \log_{10} \text{clip}(a_\mu, 10^{-11}, 10^{-8}) + 9.5. \quad (\text{A3})$$

TABLE II. Hyperparameters common to all (S)NRE algorithms tested in all applications.

Hyperparameter	Value
Hidden layers	3
Hidden layer size	256
Ensemble size	5
Learning rate	3×10^{-4}
Weight decay	0
Max gradient norm	10^{-3}
Validation interval	50
Patience	200
Batch size	256
Training split	0.95

Lastly, we restrict the p-value from Xenon1T to be positive and less than 0.0455. We then rescale to be $\mathcal{O}(1)$:

$$p_{\text{Xenon1T}} \rightarrow 10 \times \text{clip}(p_{\text{Xenon1T}}, 0, 0.0455). \quad (\text{A4})$$

In the above, $\text{clip}(x, x_{\min}, x_{\max})$ is given by:

$$\text{clip}(x, x_{\min}, x_{\max}) = \begin{cases} x_{\min} & \text{if } x < x_{\min} \\ x_{\max} & \text{if } x > x_{\max} \\ x & \text{otherwise} \end{cases} \quad (\text{A5})$$

The hyperparameters used in the networks are listed in Table II.

We used the SaxBI package [31] with JAX [32] and Flax [33] backends to initialize and fit the models to perform approximate Bayesian inference. We used the optax-repository's [34] implementation of the adaptive moment estimation (Adam) algorithm [35] to optimize our models. We used numpyro's implementation of Hamiltonian Monte Carlo (HMC) [36,37] to sample from the posterior distribution of the model parameters. We run our simulations using micrOMEGAS v5.2.13 [18] with SOFTSUSY v4.1.7 [38] backend. Our entire codebase is open-source and can be found here [39]. All computations were run on a machine with an AMD Ryzen 7 3700X processor, 16 GB of RAM, and no GPU.

-
- [1] D. J. H. Chung, L. L. Everett, G. L. Kane, S. F. King, J. D. Lykken, and L.-T. Wang, *Phys. Rep.* **407**, 1 (2005).
- [2] C. F. Berger, J. S. Gainer, J. L. Hewett, and T. G. Rizzo, *J. High Energy Phys.* **02** (2009) 023.
- [3] S. S. AbdusSalam, B. C. Allanach, F. Quevedo, F. Feroz, and M. Hobson, *Phys. Rev. D* **81**, 095012 (2010).
- [4] L. Roszkowski, E. M. Sessolo, and A. J. Williams, *J. High Energy Phys.* **02** (2015) 014.
- [5] G. Bertone, F. Calore, S. Caron, R. Ruiz, J. S. Kim, R. Trotta, and C. Weniger, *J. Cosmol. Astropart. Phys.* **04** (2016) 037.
- [6] M. Van Beekveld, W. Beenakker, M. Schutten, and J. De Wit, *SciPost Phys.* **11**, 049 (2021).
- [7] J. Alsing, T. Charnock, S. Feeney, and B. Wandelt, *Mon. Not. R. Astron. Soc.* **488**, 4440 (2019).
- [8] B. K. Miller, A. Cole, G. Louppe, and C. Weniger, *arXiv:2011.13951*.
- [9] A. Cole, B. K. Miller, S. J. Witte, M. X. Cai, M. W. Grootes, F. Nattino, and C. Weniger, Fast and credible likelihood-free cosmology with truncated marginal neural ratio estimation, *arXiv:2111.08030*.
- [10] G. Papamakarios, D. C. Sterratt, and I. Murray, *arXiv:1805.07226*.
- [11] G. Papamakarios and I. Murray, *arXiv:1605.06376*.
- [12] J. Hermans, V. Begy, and G. Louppe, *arXiv:1903.04057*.
- [13] J.-M. Lueckmann, J. Boelts, D. S. Greenberg, P. J. Gonçalves, and J. H. Macke, *arXiv:2101.04653*.
- [14] J. Hollingsworth, M. Ratz, P. Tanedo, and D. Whiteson, *Eur. Phys. J. C* **81**, 1138 (2021).
- [15] K. Cranmer, J. Brehmer, and G. Louppe, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 30055 (2020).
- [16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* **21**, 1087 (1953).
- [17] W. K. Hastings, *Biometrika* **57**, 97 (1970).
- [18] G. Bélanger, F. Boudjema, A. Pukhov, and A. Semenov, *Nuovo Cimento C* **33**, 111 (2010).
- [19] R. Neal, in *Handbook of Markov Chain Monte Carlo*, edited by S. Brooks, A. Gelman, G. Jones, and X.-L. Meng (Chapman and Hall/CRC, 2011), pp. 113–162, [10.1201/b10905](https://doi.org/10.1201/b10905).
- [20] J. Hermans, A. Delaunoy, F. Rozet, A. Wehenkel, and G. Louppe, *arXiv:2110.06581*.
- [21] S. P. MARTIN, in *Perspectives on Supersymmetry* (World Scientific, Singapore, 1998), pp. 1–98.
- [22] A. Djouadi *et al.*, *arXiv:hep-ph/9901246*.
- [23] N. Aghanim *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A6 (2020).
- [24] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins *et al.*, *Phys. Lett. B* **716**, 1 (2012).
- [25] S. Chatrchyan, V. Khachatryan, A. Sirunyan, A. Tumasyan, W. Adam, E. Aguilo, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan *et al.*, *Phys. Lett. B* **716**, 30 (2012).
- [26] B. Abi *et al.* (Muon g-2 Collaboration), *Phys. Rev. Lett.* **126**, 141801 (2021).
- [27] G. W. Bennett *et al.* (Muon g-2 Collaboration), *Phys. Rev. D* **73**, 072003 (2006).
- [28] ATLAS Collaboration, *Eur. Phys. J. C* **80**, 123 (2020).
- [29] ATLAS Collaboration, *Phys. Rev. D* **101**, 052005 (2020).
- [30] K. Cranmer, J. Pavez, and G. Louppe, *arXiv:1506.02169*.
- [31] J. Tamanas, SaxBI: Likelihood-based Inference with JAX (2022), <https://github.com/jtamanas/SaxBI>, version 1.0.0.
- [32] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: Composable transformations of Python + NumPy programs (2018), <http://github.com/google/jax>, version 0.2.5.
- [33] J. Heek, A. Levskaaya, A. Oliver, M. Ritter, B. Rondepierre, A. Steiner, and M. van Zee, Flax: A neural network library and ecosystem for JAX (2020), <http://github.com/google/flax>, version 0.4.0.
- [34] M. Hessel, D. Budden, F. Viola, M. Rosca, E. Sezener, and T. Hennigan, Optax: Composable gradient transformation and optimisation, in *JAX!* (2020), <http://github.com/deepmind/optax> version 0.0.1.
- [35] D. P. Kingma and J. Ba, *arXiv:1412.6980*.
- [36] D. Phan, N. Pradhan, and M. Jankowiak, *arXiv:1912.11554*.
- [37] E. Bingham, J. P. Chen, M. Jankowiak, F. Obermeyer, N. Pradhan, T. Karaletsos, R. Singh, P. A. Szerlip, P. Horsfall, and N. D. Goodman, *J. Mach. Learn. Res.* **20**, 1 (2019).
- [38] B. Allanach, *Comput. Phys. Commun.* **143**, 305 (2002).
- [39] <https://www.github.com/jtamanas/MSSM>.