Exploring the BSM parameter space with Simulation-Based Inference method

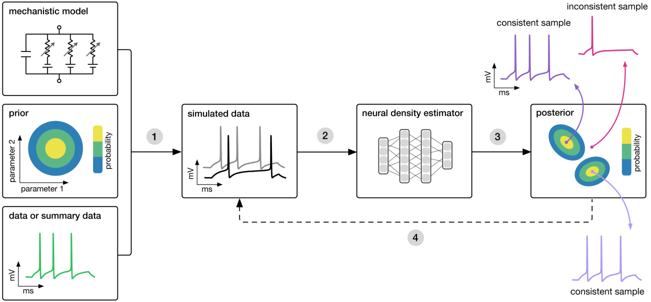
Arpita Mondal

Based on arXiv:2502.11928 (hep-ph), JHEP 12 (2025) 138

21th April, 2026

Motivation to Simulation-Based Inference

- ✔ Understanding BSM physics theories in the context of experimental results is the cornerstone of much research in HEP community
- ✔ Mapping out regions of parameter space that produce given observables is especially difficult for theories with many free parameters - ex. MSSM
- ✔ Exploration $\rightarrow$ define space of interest, uniformly sample, accept the points satisfy theoretical/experimental constraints
- ✔ Number of required samples grows exponentially with dimension, becomes computational expensive, complex likelihood to compute



Source: 10.21105/joss.02505

Amortized SBI

- Amortized → once trained, they can provide posterior for any new observation without additional simulations, making them particularly efficient for repeated inference tasks
- Sequential → effectively concentrates computational effort on the most relevant regions, reducing wasted evaluations
- Based on Bayes theorem, $p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{\theta})p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})}$
- Neural Posterior Estimation (NPE): one directly trains the density estimator to learn the posterior, $p(\boldsymbol{\theta}|\boldsymbol{x}) \propto q_\phi(\boldsymbol{\theta}|\boldsymbol{x})$
- Neural Likelihood Estimation (NLE): trains to get likelihood and then we get posterior, $p(\boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{\theta})q_\phi(\boldsymbol{x}|\boldsymbol{\theta})$
- Neural Ratio Estimation (NRE): trains to learn the likelihood-to-evidence ratio, $r(\boldsymbol{x}, \boldsymbol{\theta}) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})}$ and then we get posterior, $p(\boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{\theta})r(\boldsymbol{x}|\boldsymbol{\theta})$
- ϕ contains the hyperparameters of Neural Network

Exploring parameter space of pMSSM using SBI method

- ✔ Model $\rightarrow$ phenomenological MSSM (pMSSM), 19 parameter space
- ✔ Simulator $\rightarrow$ FeynHiggs & micrOMEGA
- ✔ Focus on Higgs sector (mass & coupling strength), B-hadron decay branching ratios
- ✔ The tree-level Higgs mass depends on $\tan\beta$ and M_A.
- ✔ Radiative correction to the tree level mass and this correction depends on many other pMSSM parameters like A_t, μ and soft SUSY breaking parameters.

Exploring parameter space of pMSSM using SBI method

- Model $\rightarrow$ phenomenological MSSM (pMSSM), 19 parameter space
- Simulator $\rightarrow$ FeynHiggs & micrOMEGA
- Focus on Higgs sector (mass & coupling strength), B-hadron decay branching ratios
- The tree-level Higgs mass depends on $\tan\beta$ and M_A.
- Radiative correction to the tree level mass and this correction depends on many other pMSSM parameters like A_t, μ and soft SUSY breaking parameters.

- M_1 = 500 GeV
- $M_3 = m_{\tilde{l}} = m_{\tilde{q}}$ = 4000 GeV
- $A_l = A_d = A_u = A_c$ = 0
- Most relevant five parameter

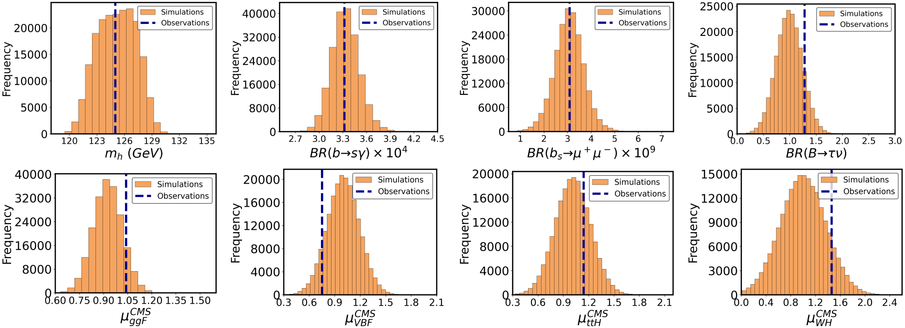
- M_2 = (550-5000) GeV
- μ = (550-5000) GeV
- $\tan\beta$ = 1-60
- M_A = (100-5000) GeV
- $|A_t|$ = (0-8000) GeV

Sample preparation

- We have simulated $\sim$ 200,000 samples
- To make this similar with practical data, we add some noise to each observable
- New sample $f'(x) = f(x) + \mathcal{N}(0, \sigma^2)$, $f(x)$ is actual sample
- The best-fit point of each observables must be within the distribution

Sample preparation

- We have simulated $\sim$ 200,000 samples
- To make this similar with practical data, we add some noise to each observable
- New sample $f'(x) = f(x) + \mathcal{N}(0, \sigma^2), f(x)$ is actual sample
- The best-fit point of each observables must be within the distribution



Ablation Study

Fixed Part		Efficiency with varying hyperparameter							
MAF 1 layer NT=20, TBS=512 LR=0.0001	HF value	2	10	50	100	400	600	800	1000
	Efficiency (%)	96.67	97.29	98.70	99.34	99.51	98.88	97.68	96.88
MAF 1 layer HF=400, TBS=512 LR=0.0001	NT value	1	2	5	10	20	40	70	100
	Efficiency (%)	96.4	97.10	97.80	98.88	99.51	99.56	99.60	99.62
MAF 1 layer HF=400, NT=20 LR=0.0001	TBS value	16	32	64	128	256	512	1024	2048
	Efficiency (%)	98.09	98.87	99.16	99.27	99.45	99.51	98.96	97.60
MAF 1 layer HF=400, NT=20 TBS=512	LR value	0.00001	0.0001	0.001	0.01	0.1	0.3	0.5	1.0
	Efficiency (%)	98.68	99.51	99.56	99.27	99.05	98.78	98.06	97.60
HF=400, NT=20 TBS=512, LR=0.0001	MAF layer	1		2		3		4	
	Efficiency (%)	99.51		99.32		99.02		99.60	

- HF : Hidden Feature, NT: number of transforms or blocks, LR: Learning Rate, TBS: Training Batch size, MAF: density estimator

Ablation Study

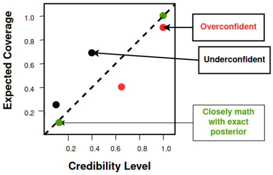
Fixed Part		Efficiency with varying hyperparameter							
MAF 1 layer NT=20, TBS=512 LR=0.0001	HF value	2	10	50	100	400	600	800	1000
	Efficiency (%)	96.67	97.29	98.70	99.34	99.51	98.88	97.68	96.88
MAF 1 layer HF=400, TBS=512 LR=0.0001	NT value	1	2	5	10	20	40	70	100
	Efficiency (%)	96.4	97.10	97.80	98.88	99.51	99.56	99.60	99.62
MAF 1 layer HF=400, NT=20 LR=0.0001	TBS value	16	32	64	128	256	512	1024	2048
	Efficiency (%)	98.09	98.87	99.16	99.27	99.45	99.51	98.96	97.60
MAF 1 layer HF=400, NT=20 TBS=512	LR value	0.00001	0.0001	0.001	0.01	0.1	0.3	0.5	1.0
	Efficiency (%)	98.68	99.51	99.56	99.27	99.05	98.78	98.06	97.60
HF=400, NT=20 TBS=512, LR=0.0001	MAF layer	1		2		3		4	
	Efficiency (%)	99.51		99.32		99.02		99.60	

- HF : Hidden Feature, NT: number of transforms or blocks, LR: Learning Rate, TBS: Training Batch size, MAF: density estimator

Features	NPE	NLE	NRE
Density function used (1 layer)	MAF	MAF	ResNET
Density function features	``hidden_features''=400 ``num_transforms''=20	``hidden_features''=200 ``num_transforms''=20	``hidden_features''=200 ``num_blocks''=20

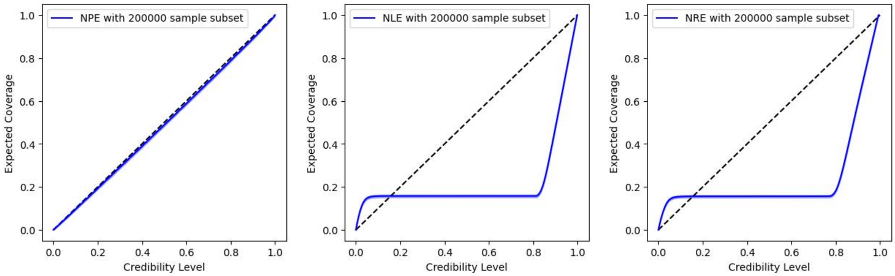
Common hyperparameters
Batch size = 512, Learning rate = 10^{-4}

Validation Process - TARP test

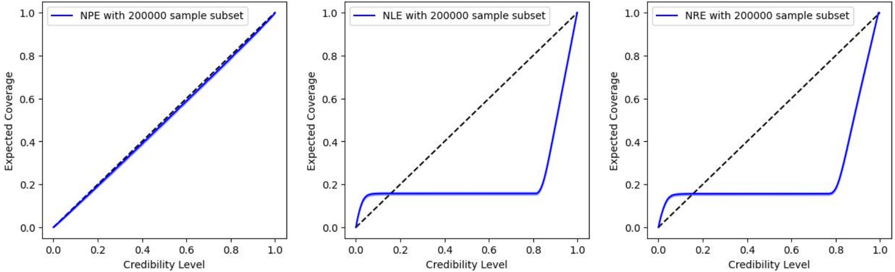


- Take a test point $\boldsymbol{x}^i_{\text{test}} \rightarrow$ true parameter $\boldsymbol{\theta}^i_{\text{test}}$
- Choose a random reference point $(\boldsymbol{\theta}^i_r)$ from a Uniform distribution in the range $[0, 1]$.
- Draw N_p points from estimated posterior $\{\boldsymbol{\theta}^{i,p}_{\text{predicted}}\}_{p=1}^{p=N_\text{p}}$.
- Next we calculate the euclidean distance between the reference point $\boldsymbol{\theta}^i_r$ and $\boldsymbol{\theta}^{i,p}_{\text{predicted}}$: $\left\{d(\boldsymbol{\theta}^{i,p}_{\text{predicted}}, \boldsymbol{\theta}^i_r)\right\}_{p=1}^{p=N_p}$.
- We also calculate distance $d(\boldsymbol{\theta}^i_{\text{test}}, \boldsymbol{\theta}^i_r)$ between true parameter ($\boldsymbol{\theta}^i_{\text{test}}$) and the reference point.
- Compute the fraction of the points $\left\{d(\boldsymbol{\theta}^{i,p}_{\text{predicted}}, \boldsymbol{\theta}^i_r)\right\}_{p=1}^{p=N_\text{p}}$ with distance smaller than $d(\boldsymbol{\theta}^i_{\text{test}}, \boldsymbol{\theta}^i_r)$. This fraction is called the credibility level
- Steps are repeated for all the N_{test} points in the test dataset.
- Draw the histogram with c_i. The cumulative distribution of c multiplied with the bin width of the histogram is called the Expected Coverage Probability (ECP)

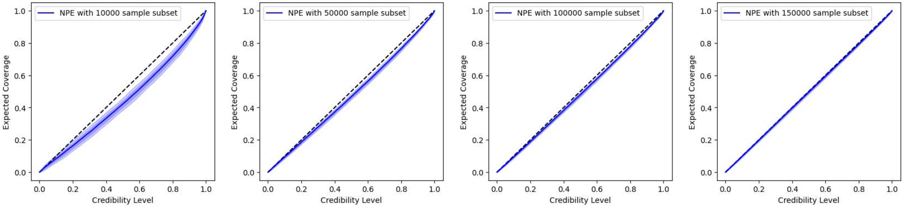
Results of TARP test



Results of TARP test

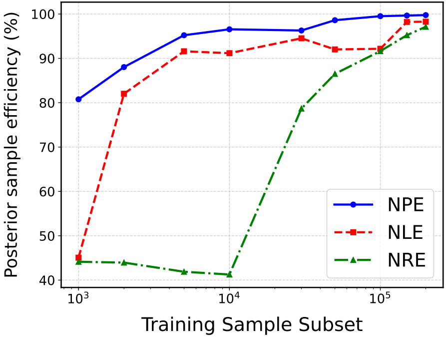


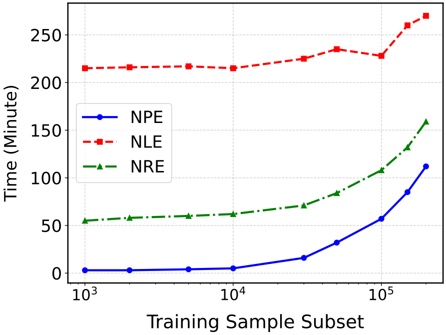
- NPE works better
- Find minimum no of samples → train with 1000, 2000, . . . , 200000
- 100,000 sample is enough for NPE method



Posterior sample efficiency

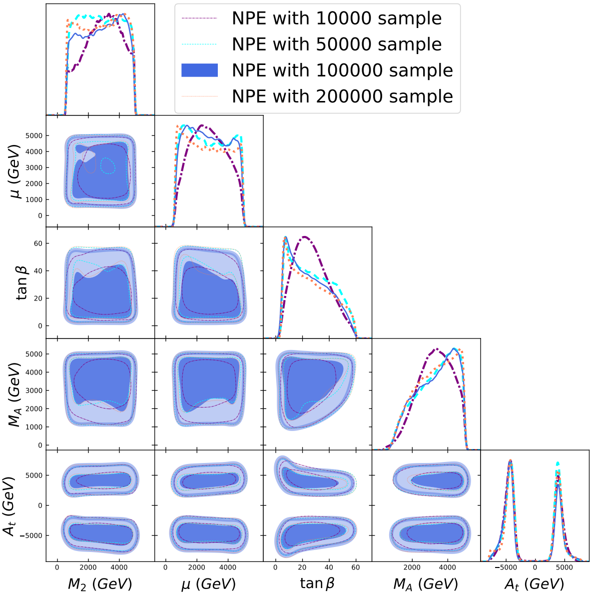
$$\text{Posterior efficiency } (\%) = \frac{\text{No of posterior samples within } 3\sigma \text{ range of observable}}{\text{Total number of posterior samples drawn}} \times 100$$



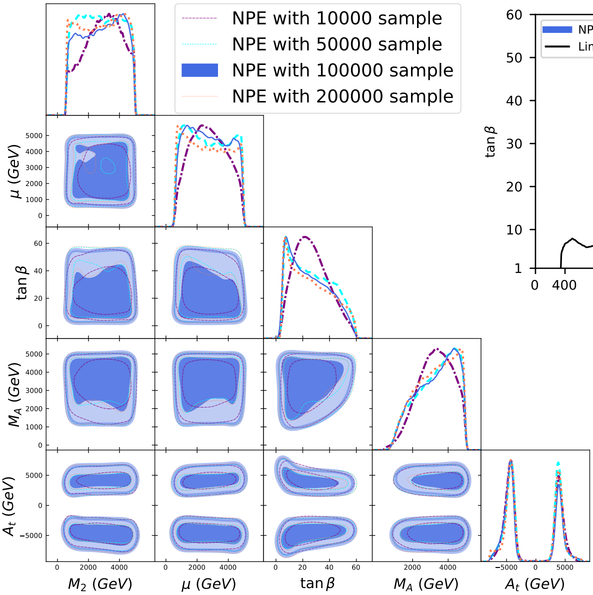


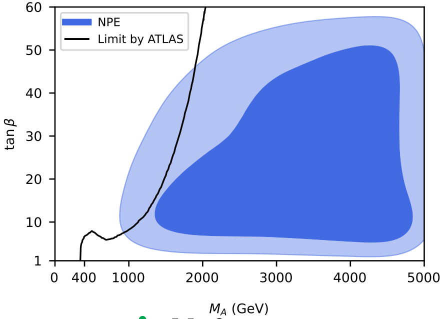
- Posterior sample efficiency saturates at 100,000 samples for NPE
- NPE requires only 50% of total sample
- NLE & NRE require larger sample set and still cannot achieve similar efficiency
- NLE (NRE) takes $\sim$ 4 (2) times longer than the NPE

Posterior distribution



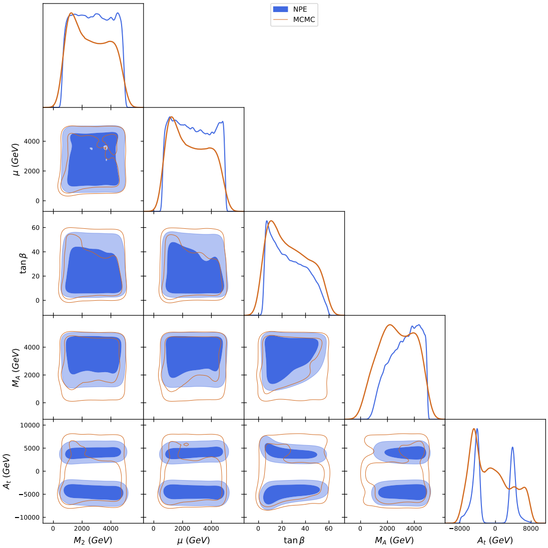
Posterior distribution





- M_2 & μ indirectly affects the Higgs mass $\rightarrow$ mostly flat
- $H \rightarrow \tau^+\tau^-$ channel provides most stringent limit on $M_A - \tan\beta$ plane
- 1σ region is still allowed

Performance Comparison: NPE Algorithm vs. MCMC Method

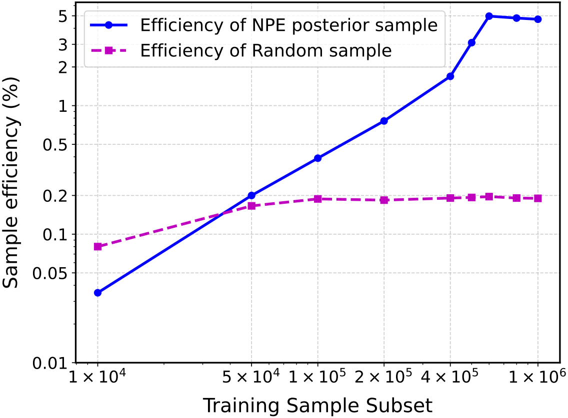


- MCMC run takes roughly 72 hours to complete, whereas the NPE method, using a subset of 1×10^5 sample subset, completes in only 24 hours.

NPE with Dark Matter constraints

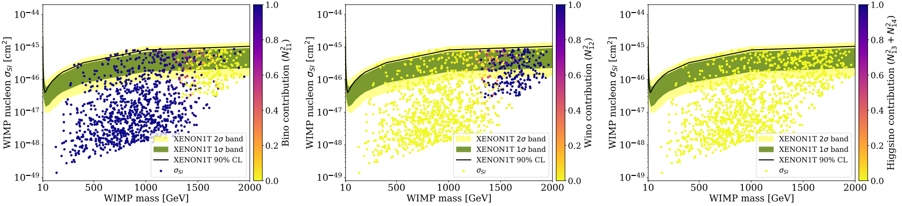
- Relic density of DM from Planck 2018, Spin-independent and spin dependent nucleon cross-sections from XENON1T, PICO-60 & PANDAX-4T

- M_1 = 100-2000
- M_2 = 100-5000
- M_3 = 2500-5000
- μ = 100-5000
- M_A = 100-5000
- $\tan\beta$ = 1-60
- $\tilde{l}$ = 1000-5000
- $\tilde{q}$ = 2000-5000
- $|A_t|$ = 0-8000



- Efficiency of random sample $\sim$ 0.2%
- Efficiency of NPE sample $\sim$ 5%, 25 times of the efficiency of random sample

Impact of DM constraints on NPE posterior sample



- $\widetilde{\chi}_1^0$ is the WIMP DM here
- For all posterior sample we compute the observables and put all the constraints
- Evidently, the higgsino-dominated scenario is completely ruled out by the XENON1T data
- the wino-dominated scenario is allowed only in the higher mass range, i.e., for DM masses in excess of 1400 GeV.
- The bino-dominated DM scenario is still allowed in the low mass region

Summary

- We have considered three amortized SBI method to show the efficiency of parameter space scan
- We considered pMSSM focusing on the Higgs sector only first
- We showed through validation test (TARP), sample efficiency and computational time NPE only works much better than NLE & NRE methods
- We also have compared this NPE method with MCMC and showed that NPE works with lesser sample set and much faster than MCMC
- We also employed NPE for larger parameter space with DM constraints
- Showed that with NPE the posterior sample efficiency by increases 25 times for 10 million sample set
- From SBI-predicted posterior, we get wino-like DM at higher mass region $\sim$1400 GeV, but in the lower we can still get bino-like DM