
Probing Dark Matter Substructure with Stellar Streams and Neural Simulation-Based Inference

Joeri Hermans
University of Liège
joeri.hermans@doct.uliege.be

Nilanjan Banik
Texas A&M University
banik@tamu.edu

Christoph Weniger
University of Amsterdam
c.weniger@uva.nl

Gianfranco Bertone
University of Amsterdam
g.bertone@uva.nl

Gilles Louppe
University of Liège
g.louppe@uliege.be

Abstract

This work applies recent advances in simulation-based inference to probe dark matter substructures in the Milky Way halo. By simulating the stellar density variations caused by subhalo impacts on stellar streams, and tying those to the WDM thermal relic mass, the technique allows us to efficiently compute preliminary, and statistically diagnosed constraints on the dark matter particle mass. Using existing GD-1 data, we find that the proposed approach has the potential to constrain the WDM mass above 15 keV, provided that simulator systematics can be controlled.

1 Introduction and cosmological context

Cold Dark Matter (CDM) models [1; 2] predict a hierarchical collapse in which large haloes form through the merging of smaller dark matter clumps [3; 4; 5]. This process is driven by CDM’s scale-free halo mass function [6; 7] and depends on the initial conditions of the matter power spectrum, which in turn anticipates the existence of dark matter haloes down to $10^{-4} M_{\odot}$ (solar masses) [8]. Warm Dark Matter (WDM) models [9; 10; 11] on the other hand, in which the dark matter particle is much lighter, influence structure formation down to the scale of dwarf galaxies. While at large scales the collapse in WDM is hierarchical as well, it becomes strongly suppressed below $10^9 M_{\odot}$, where the non-negligible velocity dispersion of dark matter particles prevents haloes to form and smooths the density field instead [12]. The discrepancies between CDM and WDM at these smaller scales provide a powerful test bed to examine the nature of the dark matter particle. Perturbations in the stellar density of cold globular cluster streams [13; 14; 15; 16; 17; 18] caused by gravitational interactions with (small-scale) dark matter substructure [19; 20; 21; 22] in the Milky Way halo make streams an ideal probe to characterize the dark matter particle. However, such perturbations are also attributable to interactions with baryonic structures in our Galaxy, such as the bar, spiral arms and the galactic population of the Giant Molecular Clouds. It is therefore crucial to study stream systems sufficiently distant from these baryonic structures, such as GD-1, to confidently detect subhalo impacts [23].

While forward modeling these complex interactions is relatively straightforward, the simulation model does not easily lend itself to statistical inference; the computation of the likelihood involves solving an intractable inverse problem which rests on the integration of all stochastic execution paths implicitly defined by computer code. Despite the intractability, it remains however possible to infer the underlying probabilities by relying on likelihood-free approximations. This approach is generally referred to as *likelihood-free* or *simulation-based* inference [24]. Previous analyses [13; 14] relied on Approximate Bayesian Computation (ABC) [25] to constrain the thermal relic mass of the dark matter particle by *manually* crafting a summary statistic based on the (partial) power spectrum of a stream’s stellar density variations. However, ABC is *only* exact whenever the summary statistic is *sufficient*

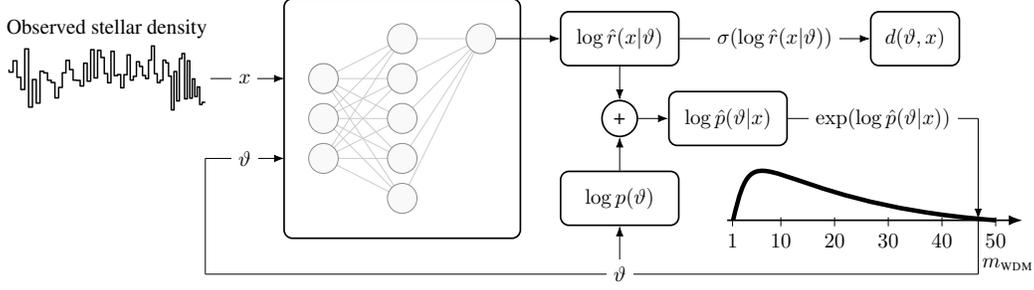


Figure 1: Graphical representation of the inference procedure after training the ratio estimator (neural network). The ratio estimator accepts a target parameter ϑ and an observable x as inputs, which are subsequently used to approximate the likelihood-to-evidence ratio $\hat{r}(x|\vartheta)$. The discriminator output $d(\vartheta, x)$ — the sigmoidal projection $\sigma(\cdot)$ of $\log \hat{r}(x|\vartheta)$ — is only used during training.

and the distance function chosen to express the similarity between observed and simulated data tends to 0, which in practice is never achievable. This work proposes to solve the *intractability* of the likelihood by learning an amortized mapping from target or model parameters ϑ and observables x to posterior densities by solving a *tractable* minimization problem. The learned mapping has the potential to increase the statistical power of an analysis since the procedure *automatically* learns an internal representation of the data. To support the reproducibility of this work, we provide all code¹. Every result is annotated with `</>`, which links to the code used to generate it.

2 Posterior inference through density ratio estimation

In this work we jointly infer **target parameters** ϑ which include the WDM thermal relic mass m_{WDM} and the stream age t_{age} . Given the Bayesian perspective of this analysis, we define the priors over the WDM mass m_{wdm} and stream age t_{age} as `uniform(1, 50)` keV and `uniform(3, 7)` billion years (Gyr) respectively. The upper bound corresponds to a half-mode mass of $\sim 4 \times 10^4 M_{\odot}$, whereas the prior over the stream age is based on the stream length and width. The length of the stream and the associated velocity dispersion indicate a stream age of at least 3.4 Gyr [26]. However, a possible explanation for this is that the stream is in fact older, but that the thicker regions are formed by encounters with dark matter or baryonic structures. Following this argument, and using the stream thickness of the thinnest regions, the upper limit is estimated to be about 7 Gyr [14]. **Observables** x encapsulate the (normalized) stellar density of mock streams — produced by the simulation or forward model — and the *observed* GD-1 normalized stellar density.

The Bayesian paradigm finds model parameters compatible with observation by computing the *posterior* $p(\vartheta|x)$. Evaluating Bayes' rule in our setting is not possible as the likelihood $p(x|\vartheta)$ is per definition intractable. To enable the tractable evaluation of the posterior, we have to rely on likelihood-free surrogates for key components in Bayes' rule. Note that Bayes' rule can be factorized into the product of the tractable prior probability and the likelihood-to-evidence ratio $r(x|\vartheta)$:

$$p(\vartheta|x) = p(\vartheta) \frac{p(x|\vartheta)}{p(x)} = p(\vartheta) \frac{p(\vartheta, x)}{p(\vartheta)p(x)} = p(\vartheta)r(x|\vartheta). \quad (1)$$

Hermans et al. [27] show that an amortized estimator $\hat{r}(x|\vartheta)$ of the intractable likelihood-to-evidence ratio can be obtained by training a discriminator $d(\vartheta, x)$ with inputs ϑ and x , to distinguish between samples from the joint $p(\vartheta, x)$ with class label 1 and samples from the product of marginals $p(\vartheta)p(x)$ with class label 0 using a discriminative criterion such as the binary cross entropy. Whenever the training criterion is minimized, the authors theoretically demonstrate that the optimal discriminator $d(\vartheta, x)$ models the Bayes-optimal decision function

$$d(\vartheta, x) = \frac{p(\vartheta, x)}{p(\vartheta, x) + p(\vartheta)p(x)}. \quad (2)$$

¹github.com/JoeriHermans/constraining-dark-matter-with-stellar-streams-and-ml

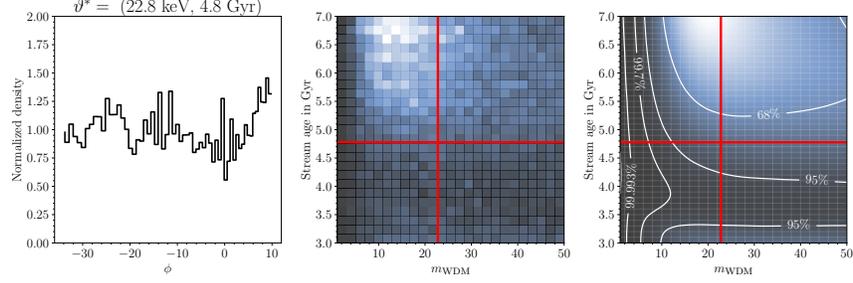


Figure 2: The columns show, from left to right, the observable, the approximate ABC posterior and our method respectively. The red cross indicates the groundtruth. \langle/\mathbb{D}

Subsequently, given a model parameter ϑ and an observable x , we can use the discriminator as a density *ratio estimator* to compute the likelihood-to-evidence ratio

$$r(x|\vartheta) = \frac{1 - d(\vartheta, x)}{d(\vartheta, x)} = \frac{p(\vartheta, x)}{p(\vartheta)p(x)} = \frac{p(x|\vartheta)}{p(x)}. \quad (3)$$

The posterior probability density can be approximated for arbitrary ϑ and x by computing $\log p(\vartheta|x) \approx \log p(\vartheta) + \log \hat{r}(x|\vartheta)$, provided that ϑ and x are supported by the product of marginals $p(\vartheta)p(x)$. The ratio estimator can likewise be used to compute a credible region (CR) at a desired level of uncertainty α by constructing a region Θ which satisfies $\int_{\Theta} p(\vartheta)r(x|\vartheta) d\vartheta = 1 - \alpha$. Since many such regions Θ exist, we select the highest posterior density region, which is the smallest CR.

3 Experiments and results

Simulations 10 million pairs $(\vartheta, x) \sim p(\vartheta, x)$ are drawn from the simulation model for training, and 100,000 for testing. The simulations in the training dataset are reused in our ABC analyses.

Ratio estimator training The ratio estimators use SELU [28] activations and were trained using the ADAMW [29] optimizer for 50 epochs with a batch-size of 4096. We found that larger batch-sizes, for our setting, generalized better. This work considers a RESNET-50 [30] architecture with 1 dimensional convolutions without dilations. Because our methodology treats ϑ as an input feature, we cannot easily condition the convolutional layers of the RESNET-based architectures on ϑ . This would require conditional convolutions [31] or hypernetworks [32] to generate specialized kernels for a given ϑ . To retain the simplicity of our architecture, we inject the dependency on ϑ in the fully connected trunk.

Approximate Bayesian Computation Our experimental trials with ABC consider $s(x) = \mathbb{V}[x/x_o]$ as a summary statistic, where x_o is the observable of interest. Given that every observable has a fixed number of bins, we can divide the synthetic observable x by the observable of interest x_o . Ideally, if the observables match perfectly, then the variance of the stellar density ratios is 0.

3.1 Statistical quality

Before making any scientific conclusion, it is crucial to verify the result of the involved statistical computation. This is especially challenging in the likelihood-free setting because the likelihood model is per definition intractable. The amortization of the ratio estimators enables us to quickly approximate the posterior of any observable x supported by the marginal model $p(x)$. It is therefore computationally viable to assess the quality of the approximation by determining whether the empirical *coverage* probability corresponds to the desired confidence level. In every trial, we evaluate the interval construction on 10,000 unseen observables. This is repeated 10 times to build up a robust statistic. The empirical coverage probability of a ratio estimator is therefore based on approximately 100,000 unseen observables in total. Although credible regions do not necessarily have a frequentist interpretation, they closely approximate the nominal coverage probability. All architectures exhibit similar performance, the RESNET-50 architecture yields 0.675 ± 0.006 , 0.944 ± 0.002 , 0.996 ± 0.001 for the confidence levels 0.68, 0.95 and 0.997 respectively \langle/\mathbb{D} . The bias of the ratio estimator is probed by assessing the convergence of the mode towards the nominal target parameters for 1000 independent and identically distributed observables. Figure 3 demonstrates a single trial.

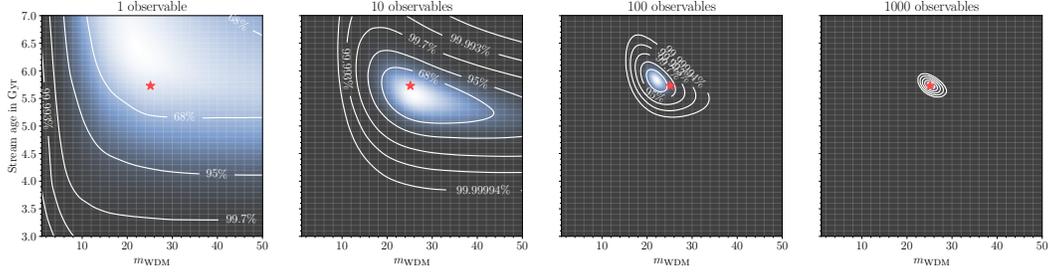


Figure 3: Demonstration of the bias probe. The figures show, from left to right, the posteriors for 1, 10, 100 and 1000 independent and identically distributed mock GD-1 observables. As the number of observables increases, the posteriors are becoming increasingly more tight around the nominal value. This indicates that the posteriors do not, in expectation, introduce significant bias. $\langle/\mathit>$

3.2 Direct comparison against ABC

We find that the structure of the posteriors are, for most mock simulations, in strong agreement. However, we observe that in some cases our ABC results depend strongly on the adopted summary statistics. This highlights the difficulty in manually constructing an effective summary statistic for high-dimensional data, while this aspect is completely automated in our approach. Diagnostics based on (approximate) posterior samples exist [33], but are not computationally viable because the posteriors have to be numerically approximated for every test-observable. Our method does not suffer from this issue, because the estimation of the posterior density is amortized.

3.3 Preliminary constraints on m_{WDM} using data on GD-1

We compute *preliminary* constraints on m_{WDM} based on the observed stellar density along the GD-1 stream which was obtained using *Gaia* proper motions [34; 35] and *Pan-STARRS* photometry [36]. We stress that **our simulation model does not account for baryonic effects, disturbances caused by massive ($> 10^9 M_{\odot}$) subhalos, and effects induced by variations in the Milky Way potential.** The posteriors and corresponding credible regions are shown in Figure 4. In this context, all independent analyses have a strong preference for CDM over WDM. After marginalizing the stream age, the proposed methodology yields $m_{\text{WDM}} \geq 17.5$ keV (95% CR) and $m_{\text{WDM}} \geq 10.5$ keV (99.7% CR). Assuming the posterior approximated by ABC is exact, we find $m_{\text{WDM}} \geq 10.2$ keV (95% CR) and $m_{\text{WDM}} \geq 5.0$ keV (99.7% CR). Our results are promising and indicate that with the proposed

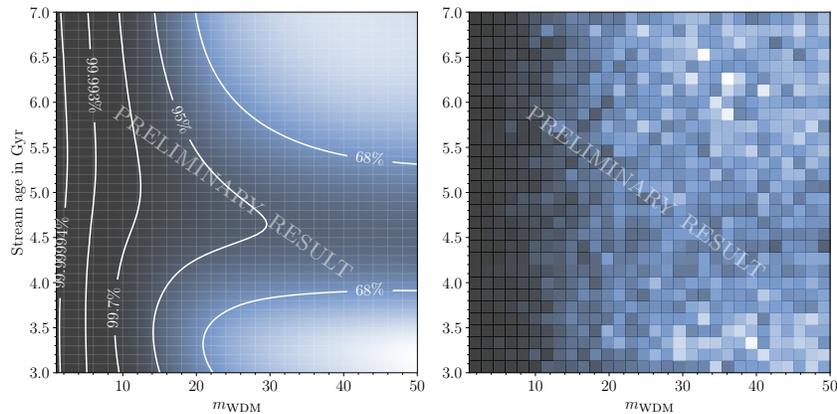


Figure 4: Posteriors based on the observed stellar density variations of GD-1. The figure shows, from left to right, our method and ABC respectively. The constraints are based on credible regions. All posteriors indicate a preference for CDM over WDM within the assumed simulation model. $\langle/\mathit>$

method we could effectively exclude the sterile neutrino [10; 37; 38; 39; 40] as a possible candidate for the dark matter particle, provided simulator systematics can be brought under control.

Broader impact

The approach could provoke a shift in the way likelihood-free statistical inference is done in the physical sciences. Previous approaches, such as Approximate Bayesian Computation, require a hand-crafted sufficient summary statistic and distance metric. The specific definition of both components can significantly influence the approximated posteriors. Instead, our approach automatically *learns* an efficient internal representation of the relation between the model parameters and the observables, and therefore permits domain-scientists to pivot from attempting to solve the intractable inverse problem — by defining *assumed* sufficient summary statistics — to the more natural forward modelling of the phenomena of interest. The machine learning component would thus handle the inference aspect. Finally, the amortization of the posterior estimates enables thorough statistical validation of the neural networks to ensure qualitative posteriors. Something which is not feasible in ABC analyses.

Acknowledgments and Disclosure of Funding

Joeri Hermans would like to thank the National Fund for Scientific Research for his FRIA scholarship. Gilles Louppe is recipient of the ULiège - NRB Chair on Big data and is thankful for the support of NRB. All authors would like to thank the developers of the *Jupyter* [41] project, *Matplotlib* [42], *Numpy* [43], *PyTorch* [44], and *Scikit-Learn* [45] for enabling this work.

References

- [1] P. J. E. Peebles. Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations. *Astrophysical Journal*, 263:L1–L5, December 1982. doi: 10.1086/183911.
- [2] G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees. Formation of galaxies and large-scale structure with cold dark matter. *Nature*, 311:517–525, October 1984. doi: 10.1038/311517a0.
- [3] Ben Moore, Sebastiano Ghigna, Fabio Governato, George Lake, Thomas Quinn, Joachim Stadel, and Paolo Tozzi. Dark matter substructure within galactic halos. *The Astrophysical Journal Letters*, 524(1):L19, 1999.
- [4] Vladimir Avila-Reese, Claudio Firmani, and Xavier Hernández. On the formation and evolution of disk galaxies: Cosmological initial conditions and the gravitational collapse. *The Astrophysical Journal*, 505(1):37, 1998.
- [5] DH Zhao, HJ Mo, YP Jing, and Gerhard Boerner. The growth and structure of dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 339(1):12–24, 2003.
- [6] Stefan Hofmann, Dominik J Schwarz, and Horst Stoecker. Damping scales of neutralino cold dark matter. *Physical Review D*, 64(8):083507, 2001.
- [7] Aurel Schneider, Robert E Smith, and Darren Reed. Halo mass function and the free streaming scale. *Monthly Notices of the Royal Astronomical Society*, 433(2):1573–1587, 2013.
- [8] Edmund Bertschinger. Effects of cold dark matter decoupling and pair annihilation on cosmological perturbations. *Physical Review D*, 74(6):063509, 2006.
- [9] JR Bond and AS Szalay. The collisionless damping of density fluctuations in an expanding universe, 1983.
- [10] Scott Dodelson and Lawrence M Widrow. Sterile neutrinos as dark matter. *Physical Review Letters*, 72(1):17, 1994.
- [11] Paul Bode, Jeremiah P. Ostriker, and Neil Turok. Halo Formation in Warm Dark Matter Models. *Astrophysical Journal*, 556(1):93–107, July 2001. doi: 10.1086/321541.

- [12] Robert E Smith and Katarina Markovic. Testing the warm dark matter paradigm with large-scale structures. *Physical Review D*, 84(6):063507, 2011.
- [13] Nilanjan Banik, Gianfranco Bertone, Jo Bovy, and Nassim Bozorgnia. Probing the nature of dark matter particles with stellar streams. *Journal of Cosmology and Astroparticle Physics*, 2018(07):061, 2018.
- [14] Nilanjan Banik, Jo Bovy, Gianfranco Bertone, Denis Erkal, and TJJ de Boer. Novel constraints on the particle nature of dark matter from stellar streams. *arXiv preprint arXiv:1911.02663*, 2019.
- [15] Ana Bonaca, Charlie Conroy, David W. Hogg, Phillip A. Cargile, Nelson Caldwell, Rohan P. Naidu, Adrian M. Price-Whelan, Joshua S. Speagle, and Benjamin D. Johnson. High-resolution spectroscopy of the GD-1 stellar stream localizes the perturber near the orbital plane of Sagittarius. *Astrophys. J. Lett.*, 892(2):L37, 2020. doi: 10.3847/2041-8213/ab800c.
- [16] Ana Bonaca, David W Hogg, Adrian M Price-Whelan, and Charlie Conroy. The spur and the gap in gd-1: Dynamical evidence for a dark substructure in the milky way halo. *The Astrophysical Journal*, 880(1):38, 2019.
- [17] Jo Bovy, Anita Bahmanyar, Tobias K Fritz, and Nitya Kallivayalil. The shape of the inner milky way halo from observations of the pal 5 and gd-1 stellar streams. *The Astrophysical Journal*, 833(1):31, 2016.
- [18] Nilanjan Banik and Jo Bovy. Effects of baryonic and dark matter substructure on the Pal 5 stream. *Mon. Not. Roy. Astron. Soc.*, 484(2):2009–2020, 2019. doi: 10.1093/mnras/stz142.
- [19] RA Ibata, GF Lewis, MJ Irwin, and T Quinn. Uncovering cold dark matter halo substructure with tidal streams. *Monthly Notices of the Royal Astronomical Society*, 332(4):915–920, 2002.
- [20] Kathryn V Johnston, David N Spergel, and Christian Haydn. How lumpy is the milky way’s dark matter halo? *The Astrophysical Journal*, 570(2):656, 2002.
- [21] Joo Heon Yoon, Kathryn V Johnston, and David W Hogg. Clumpy streams from clumpy halos: detecting missing satellites with cold stellar structures. *The Astrophysical Journal*, 731(1):58, 2011.
- [22] Raymond G Carlberg. Dark matter sub-halo counts via star stream crossings. *The Astrophysical Journal*, 748(1):20, 2012.
- [23] Nilanjan Banik and Jo Bovy. Effects of baryonic and dark matter substructure on the pal 5 stream. *Monthly Notices of the Royal Astronomical Society*, 484(2):2009–2020, 2019.
- [24] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 2020.
- [25] Donald B Rubin. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, pages 1151–1172, 1984.
- [26] Jeremy J Webb and Jo Bovy. Searching for the gd-1 stream progenitor in gaia dr2 with direct n-body simulations. *Monthly Notices of the Royal Astronomical Society*, 485(4):5929–5938, 2019.
- [27] Joeri Hermans, Volodimir Begy, and Gilles Louppe. Likelihood-free MCMC with Amortized Approximate Ratio Estimators. *arXiv e-prints*, art. arXiv:1903.04057, Mar 2019.
- [28] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. In *Advances in neural information processing systems*, pages 971–980, 2017.
- [29] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [30] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

- [31] Brandon Yang, Gabriel Bender, Quoc V Le, and Jiquan Ngiam. Condconv: Conditionally parameterized convolutions for efficient inference. In *Advances in Neural Information Processing Systems*, pages 1307–1318, 2019.
- [32] David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016.
- [33] Sean Talts, Michael Betancourt, Daniel Simpson, Aki Vehtari, and Andrew Gelman. Validating bayesian inference algorithms with simulation-based calibration. *arXiv preprint arXiv:1804.06788*, 2018.
- [34] Gaia Collaboration and Brown. Gaia Data Release 2. Summary of the contents and survey properties. *Astronomy and Astrophysics*, 616:A1, August 2018. doi: 10.1051/0004-6361/201833051.
- [35] Gaia Collaboration and Prusti. The Gaia mission. *Astronomy and Astrophysics*, 595:A1, November 2016. doi: 10.1051/0004-6361/201629272.
- [36] Kenneth C. Chambers and Pan-STARRS Team. The Pan-STARRS1 Survey Data Release. In *American Astronomical Society Meeting Abstracts 229*, volume 229 of *American Astronomical Society Meeting Abstracts*, page 223.03, January 2017.
- [37] Xiangdong Shi and George M Fuller. New dark matter candidate: nonthermal sterile neutrinos. *Physical Review Letters*, 82(14):2832, 1999.
- [38] Kevork Abazajian, George M Fuller, and Mitesh Patel. Sterile neutrino hot, warm, and cold dark matter. *Physical Review D*, 64(2):023501, 2001.
- [39] Takehiko Asaka and Mikhail Shaposhnikov. The ν msm, dark matter and baryon asymmetry of the universe. *Physics Letters B*, 620(1-2):17–26, 2005.
- [40] Alexey Boyarsky, Oleg Ruchayskiy, and Mikhail Shaposhnikov. The role of sterile neutrinos in cosmology and astrophysics. *Annual Review of Nuclear and Particle Science*, 59:191–214, 2009.
- [41] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, and Carol Willing. Jupyter notebooks – a publishing format for reproducible computational workflows. In F. Loizides and B. Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87 – 90. IOS Press, 2016.
- [42] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.
- [43] Stefan Van Der Walt, S Chris Colbert, and Gael Varoquaux. The numpy array: a structure for efficient numerical computation. *Computing in Science & Engineering*, 13(2):22, 2011.
- [44] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.
- [45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.