Microlensing Detection and Inference via Learned Bayes Factors

Nolan Smyth[1,2,3] **Laurence Perreault-Levasseur**[1,2,3,4,5,6] **Yashar Hezaveh**[1,2,3,4,5,6]

[1]Ciela Institute, Montreal Institute for Astrophysics and Machine Learning, Montréal QC, Canada
[2]Mila – Quebec AI Institute, Montréal QC, Canada
[3]Department of Physics, Université de Montréal, Montréal QC, Canada
[4]Center for Computational Astrophysics, Flatiron Institute, New York, USA
[5]Perimeter Institute for Theoretical Physics, Waterloo, Canada
[6]Trottier Space Institute, McGill University, Montréal, Canada
nolan.smyth@umontreal.ca

Abstract

We present a unified framework for gravitational microlensing event detection and parameter inference. Traditional pipelines use deterministic hard cuts on photometric statistics, systematically missing low-magnification events in the finite-source regime. We instead frame detection as Bayesian model comparison using Evidence Networks, which learn calibrated Bayes factors from binary-labeled simulations, and combine this with Neural Posterior Estimation (NPE) for amortized parameter inference. Both share a transformer encoder that handles irregularly-sampled time series without imputation. On simulated Roman Space Telescope data, our Evidence Network achieves 99.9% detection efficiency with no false positives among $5{,}000$ Gaussian-noise realizations (a 95% upper bound on the false-positive rate of 6×10^{-4}). Gains are most dramatic in the extreme finite-source regime ($\rho \gtrsim 5$), where detection rates reach $\sim$95% versus $\sim$50% for hard cuts, precisely the short-duration free-floating planet events most constraining for formation scenarios. Our NPE provides calibrated posteriors, working towards real-time analysis at survey scale.

1 Introduction

The Nancy Grace Roman Space Telescope will survey over 200 million stars toward the Galactic bulge at 15-minute cadence, detecting $\sim$1,400 bound exoplanets [Penny et al., 2019] and $\sim$250 free-floating planets (FFPs) [Johnson et al., 2020] via gravitational microlensing, with lens mass and distance measurements achievable for a significant fraction of hosts [Terry et al., 2025]. FFPs are expected to outnumber bound terrestrial planets by a factor of 20 or more [Sumi et al., 2023], and the recent first direct mass measurement of an FFP via microlensing parallax [Dong et al., 2026] underscores the maturity of this technique. Characterizing the low-mass end of the FFP mass function requires sensitivity to short-duration, low-magnification events in the finite-source regime.

Realizing this science demands solving two coupled challenges: *detection* of microlensing signals amid billions of light curves, and *inference* of physical parameters for detected events. Current pipelines address these sequentially. Detection relies on deterministic *hard cuts*—threshold-based criteria such as $\Delta\chi^2 \geq 300$ with $N_{\text{consec}} \geq 6$ consecutive 3σ points [Johnson et al., 2020], or summed-significance variants [Sumi et al., 2011, Mroz et al., 2017]. Hard cuts are ad hoc, provide no probabilistic interpretation, and systematically miss low-magnification events. Inference then employs Markov Chain Monte Carlo (MCMC), which is computationally expensive and does not amortize across observations.

Preprint.

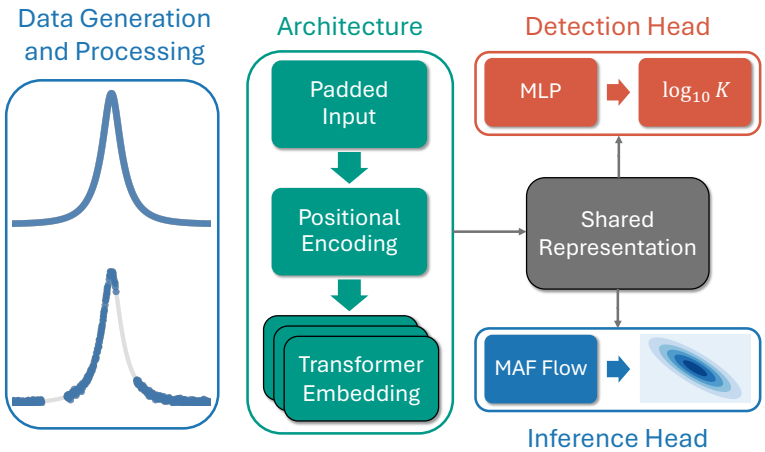

Figure 1: Joint architecture for microlensing detection and inference. A shared transformer encoder processes irregularly-sampled light curves into a fixed-dimensional summary $\mathbf{s}$. The detection head outputs calibrated Bayes factors $\log_{10} K$ via Evidence Networks, while the inference head uses a masked-autoregressive normalizing flow to estimate the full posterior $p(\boldsymbol{\theta}|\mathbf{x})$.

We present a unified neural framework combining *Evidence Networks* [Jeffrey and Wandelt, 2024] for Bayes factor estimation with *Neural Posterior Estimation* (NPE) [Papamakarios and Murray, 2018, Greenberg et al., 2019] for amortized posterior inference. Both share a transformer encoder for irregularly-sampled time series. Prior work has applied NPE to binary microlensing [Zhang et al., 2021], transformer embeddings for single-lens inference on irregular time-series data [Smyth et al., 2025], and normalizing flows with reversible-jump MCMC for microlensing model selection [Keehan et al., 2022]. Our framework extends these by unifying detection and characterization without requiring MCMC at inference time.

Our key contributions are: **(i)** a unified probabilistic framework performing both detection and inference within a Bayesian formalism; **(ii)** principled model comparison via Evidence Networks that recover calibrated Bayes factors from binary labels alone, replacing ad hoc $\Delta\chi^2$ thresholds; and **(iii)** improved detection completeness in the finite-source regime ($\sim$95% vs. $\sim$50% for hard cuts), where short-duration FFP events are most prevalent.

2 Method

Data and augmentation. We model finite-source point-lens (FSPL) microlensing events with five parameters $\boldsymbol{\theta} = (t_0, u_0, t_E, \rho, f_s)$: time of closest approach, impact parameter (in units of the Einstein radius θ_E), Einstein crossing time, normalized source radius $\rho \equiv \theta_s/\theta_E$ (where θ_s is the angular source size), and source flux fraction f_s such that the observed flux is $F(t) = f_s A(t, \boldsymbol{\theta}) + (1 - f_s)$ with A the magnification and $1 - f_s$ the blended light from unresolved stars. We adopt box-uniform priors: $t_0 \in [0, 20]$ days, $u_0 \in [0, 1.5]$, $\log_{10}(t_E/\text{day}) \in [-1, 1.3]$, $\log_{10}\rho \in [-2, 1]$, $f_s \in [0.1, 1.0]$. Light curves are simulated using `VBMicrolensing` [Bozza et al., 2018] at Roman's 15-minute cadence over a 20-day window. To build robustness, we apply on-the-fly augmentation: 0–3 seasonal gaps (1–10 days each), random dropout (0–60%), and Gaussian noise ($\sigma \in [0.001, 0.02]$). Each light curve is represented as a padded sequence of length $L = 1000$ with three channels per timestep: $(t_i^{\text{norm}}, F_i, \sigma_i)$. We apply recoverability filters to exclude unrecoverable events: at least 5 points within $1.5\,t_E$ of t_0, at least 5 points beyond $3.0\,t_E$ from t_0 (for baseline establishment), and peak signal-to-noise ratio > 5.

Shared transformer encoder. Both the Evidence Network and NPE share a transformer encoder [Vaswani et al., 2023] that maps variable-length input $\mathbf{X} \in \mathbb{R}^{L\times 3}$ to a fixed summary $\mathbf{s} \in \mathbb{R}^{256}$ (Figure 1). The architecture consists of an input projection, sinusoidal positional encoding, and 6 transformer layers with 8-head self-attention ($d_{\text{model}} = 256$, $d_{\text{ff}} = 512$). A learnable CLS token aggregates information via self-attention; padded positions are masked. This design naturally handles irregular sampling without interpolation.

Evidence Networks for detection. We frame detection as Bayesian model comparison. Given data $\mathbf{x}$, the Bayes factor $K = p(\mathbf{x}|\mathcal{M}_1)/p(\mathbf{x}|\mathcal{M}_0)$ quantifies evidence for the signal hypothesis $\mathcal{M}_1$ relative to noise $\mathcal{M}_0$. Unlike $\Delta\chi^2$ statistics—which are often used as a proxy for model comparison without rigorous Bayesian justification and fail to penalize model complexity—Bayes factors incorporate Occam's razor via marginalization over parameters.

Evidence Networks [Jeffrey and Wandelt, 2024] learn $\log_{10} K$ from binary-labeled data using the leaky parity-odd power (l-POP) exponential loss:

$$\mathcal{L}(f(\mathbf{x}), m) = \exp\left[\left(\tfrac{1}{2} - m\right) J_\alpha(f(\mathbf{x}))\right], \quad J_\alpha(z) = z + z|z|^{\alpha-1} \tag{1}$$

where $m \in \{0, 1\}$ is the binary label and $f(\mathbf{x})$ is the network output. For the optimal f^*, we have $\log_{10} K = J_\alpha(f^*(\mathbf{x}))$. The network consists of the shared transformer followed by a 3-layer MLP ($[256, 128, 1]$). It is trained on 100,000 light curves with equal signal/noise fractions.

Neural posterior estimation. For detected events, we estimate $p(\boldsymbol{\theta}|\mathbf{x})$ using NPE with a Masked Autoregressive Flow [Papamakarios et al., 2018] conditioned on the transformer summary $\mathbf{s}$, implemented via the `sbi` library [Boelts et al., 2025]. The flow is trained on 80,000 simulated pairs by maximizing $\mathbb{E}[\log p_\psi(\boldsymbol{\theta}|g_\phi(\mathbf{x}))]$. At test time, drawing 10,000 posterior samples requires $\sim$100 milliseconds—approximately $10^3\times$ faster than MCMC on the same hardware.

3 Results

Detection Performance We evaluate on a test set of 5,000 simulated events with realistic augmentation. Using a threshold of $\log_{10} K > 0.8$ (chosen for zero false positives on the validation set), the Evidence Network detects **4,996 of 5,000 signals (99.9%)** with no false positives on the 5,000 Gaussian-noise realizations—a one-sided Clopper–Pearson 95% upper bound on the false-positive rate of 5.99×10^{-4}, valid within the simulator's support (Section 4). This operating point lies on a broad plateau, not a knife edge: signal and noise $\log_{10} K$ distributions are cleanly separated (AUROC $= 0.9995$). For comparison, we compute hard-cut statistics using the *true* injected parameters—an oracle upper bound, since in practice one must fit the model, incurring optimization failures and local minima that would further degrade detection rates. Even with this advantage, hard cuts achieve only 95.9% [Johnson et al., 2020], 93.9% [Sumi et al., 2011], and 95.4% [Mroz et al., 2017].

Detection Efficiency vs. Physical Parameters Figure 2 shows detection efficiency as a function of each FSPL parameter. The advantage is most pronounced in the extreme finite-source regime ($\rho \gtrsim 5$), where efficiency remains $\sim$95% compared to $\sim$50% for hard cuts. This regime dominates short-duration events from low-mass FFPs—precisely the population most constraining for formation and ejection scenarios. Hard cuts fail because finite-source events have lower peak magnification despite distinctive morphological signatures (Figure 3); the Evidence Network extracts evidence from the full temporal structure.

Posterior Calibration We assess NPE calibration on the 4,996 detected events using the TARP diagnostic [Lemos et al., 2023]. Expected coverage closely tracks the credibility level across all parameters, confirming well-calibrated, unbiased posteriors. We additionally validate the Evidence Network on a toy model—detecting Gaussian bumps in white noise—where the Bayes factor admits a closed-form solution. The network is trained on binary signal/noise labels only, with no access to the analytic $\log K$ values. Despite this, it achieves near-oracle discriminative performance (AUC = 0.725 vs. oracle 0.726), demonstrating that Evidence Networks recover calibrated Bayes factors from binary labels alone without requiring explicit likelihood computation.

4 Discussion and Conclusion

Our unified framework replaces ad hoc hard cuts with calibrated Bayes factors via Evidence Networks, combined with amortized posterior inference via NPE—all sharing a single transformer encoder that handles irregular sampling without imputation.

Limitations. Our training assumes Gaussian photometric noise and does not model astrophysical false positives (stellar flares, eclipsing binaries, pulsating variables); the false-positive bound above is therefore valid only within this idealized noise model, and injecting or learning realistic contaminants during training is the most important next step [Legin et al., 2023]. The 20-day observation window

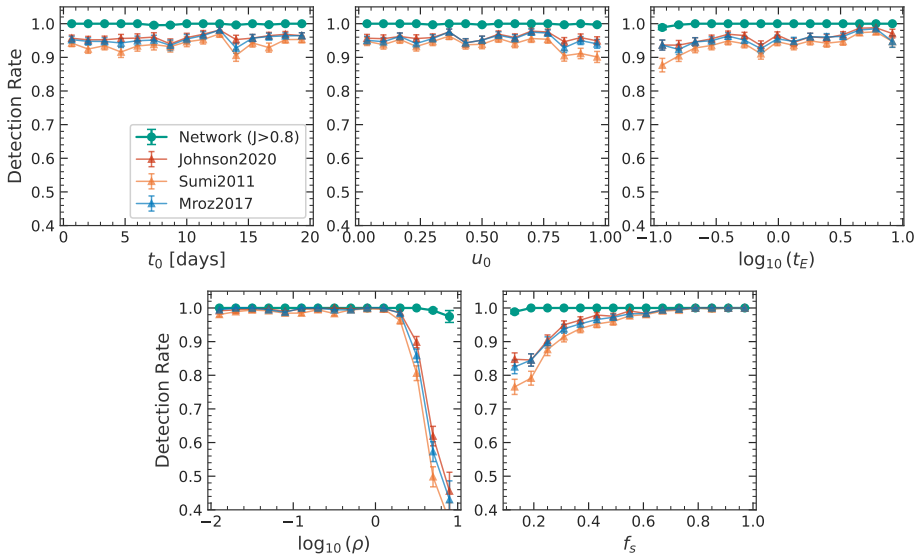

Figure 2: Detection efficiency as a function of FSPL parameters. The Evidence Network (blue) dramatically outperforms literature hard cuts in the extreme finite-source regime ($\rho \gtrsim 5$), recovering nearly twice as many events. Hard-cut statistics use oracle (true) parameters, so the comparison is conservative.

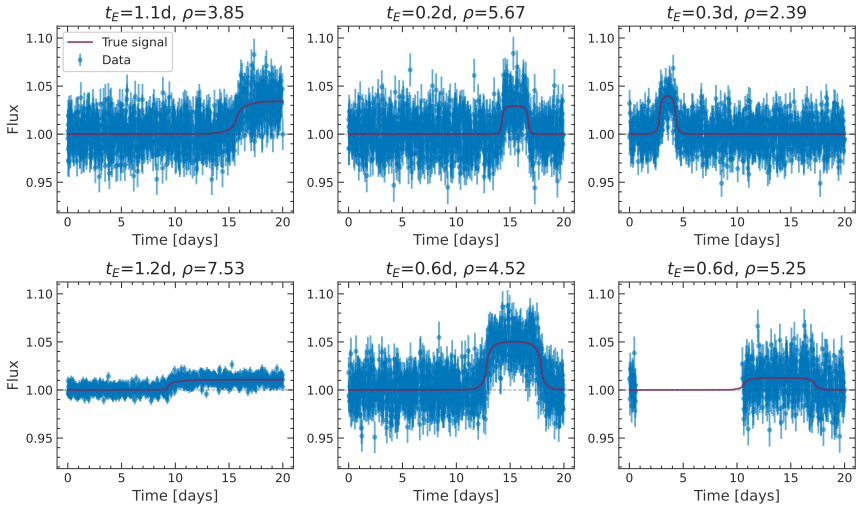

Figure 3: Six representative events detected by the Evidence Network ($\log_{10} K > 0.8$) but missed by all three literature hard cuts. These finite-source events ($\rho \gtrsim 2$) have low peak magnification despite distinctive morphological signatures (red curves), making them invisible to threshold-based statistics. The network extracts evidence from the full temporal structure rather than relying on peak flux alone.

limits applicability to longer-duration events from more massive lenses. Additionally, SBI methods can be sensitive to distributional shifts [Filipp et al., 2024]; the box-uniform priors above should ultimately be replaced by galactic population models.

Related approaches and outlook. Differentiable simulators (`microJAX` [Miyazaki and Kawahara, 2025], `microlux` [Ren and Zhu, 2025]) enable gradient-based inference via HMC [Ren et al., 2026]; these are complementary to SBI, which works with any black-box simulator and amortizes across observations. Natural next steps include binary lens models, hierarchical population inference, multi-band data, and other irregularly-sampled transients. Code: `https://github.com/NolanSmyth/evidence_networks_sbi_microlensing`

References

Matthew T. Penny, B. Scott Gaudi, Eamonn Kerins, Nicholas J. Rattenbury, Shude Mao, Annie C. Robin, and Sebastiano Calchi Novati. Predictions of the WFIRST Microlensing Survey I: Bound Planet Detection Rates. *The Astrophysical Journal Supplement Series*, 241(1):3, February 2019. ISSN 1538-4365. doi: 10.3847/1538-4365/aafb69.

Samson A. Johnson, Matthew T. Penny, B. Scott Gaudi, Eamonn Kerins, Nicholas J. Rattenbury, Annie C. Robin, Sebastiano Calchi Novati, and Calen B. Henderson. Predictions of the Nancy Grace Roman Space Telescope Galactic Exoplanet Survey II: Free-Floating Planet Detection Rates. *The Astronomical Journal*, 160(3):123, August 2020. ISSN 1538-3881. doi: 10.3847/1538-3881/aba75b.

Sean K. Terry, Etienne Bachelet, Farzaneh Zohrabi, Himanshu Verma, Alison Crisp, Macy Huston, Carissma McGee, Matthew Penny, Natasha S. Abrams, Michael D. Albrow, Jay Anderson, Fatemeh Bagheri, Jean-Phillipe Beaulieu, Andrea Bellini, David P. Bennett, Galen Bergsten, T. Dex Bhadra, Aparna Bhattacharya, Ian A. Bond, Valerio Bozza, Christopher Brandon, Sebastiano Calchi Novati, Sean Carey, Jessie Christiansen, William DeRocco, B. Scott Gaudi, Jon Hulberg, Stela Ishitani Silva, Sinclaire E. Jones, Eamonn Kerins, Somayeh Khakpash, Katarzyna Kruszynska, Casey Lam, Jessica R. Lu, Amber Malpas, Shota Miyazaki, Przemek Mroz, Arjun Murlidhar, David Nataf, Marz Newman, Greg Olmschenk, Rakek Poleski, Clement Ranc, Nicholas J. Rattenbury, Krzysztof Rybicki, Vito Saggese, Jennifer Sobeck, Keivan G. Stassun, Alexander P. Stephan, Rachel A. Street, Takahiro Sumi, Daisuke Suzuki, Aikaterini Vandorou, Meet Vyas, Jennifer C. Yee, Weicheng Zang, and Keming Zhang. Predictions of the Nancy Grace Roman Space Telescope Galactic Exoplanet Survey. IV. Lens Mass and Distance Measurements, October 2025.

Takahiro Sumi, Naoki koshimoto, David P. Bennett, Nicholas J. Rattenbury, Fumio Abe, Richard Barry, Aparna Bhattacharya, Ian A. Bond, Hirosane Fujii, Akihiko Fukui, Ryusei Hamada, Yuki Hirao, Stela Ishitani Silva, Yoshitaka Itow, Rintaro Kirikawa, Iona Kondo, Yutaka Matsubara, Shota Miyazaki, Yasushi Muraki, Greg Olmschenk, Clement Ranc, Yuki Satoh, Daisuke Suzuki, Mio Tomoyoshi, Paul J. Tristram, Aikaterini Vandorou, Hibiki Yama, and Kansuke Yamashita. Free-Floating planet Mass Function from MOA-II 9-year survey towards the Galactic Bulge, July 2023.

Subo Dong, Zexuan Wu, Yoon-Hyun Ryu, Andrzej Udalski, Przemek Mroz, Krzysztof A. Rybicki, Simon T. Hodgkin, Lukasz Wyrzykowski, Laurent Eyer, Thomas Bensby, Ping Chen, Sharon X. Wang, Andrew Gould, Hongjing Yang, Michael D. Albrow, Sun-Ju Chung, Cheongho Han, Kyu-Ha Hwang, Youn Kil Jung, In-Gu Shin, Yossi Shvartzvald, Jennifer C. Yee, Weicheng Zang, Dong-Jin Kim, Chung-Uk Lee, Byeong-Gon Park, Radoslaw Poleski, Jan Skowron, Michal K. Szymanski, Igor Soszynski, Pawel Pietrukowicz, Szymon Kozlowski, Dorota M. Skowron, Krzysztof Ulaczyk, Mariusz Gromadzki, Milena Ratajczak, Patryk Iwanek, Marcin Wrona, Mateusz J. Mroz, Guy Rixon, Diana L. Harrison, and Elme Breedt. A free-floating-planet microlensing event caused by a Saturn-mass object. *Science*, 391(6780):96–99, January 2026. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.adv9266.

T. Sumi, K. Kamiya, A. Udalski, D. P. Bennett, I. A. Bond, F. Abe, C. S. Botzler, A. Fukui, K. Furusawa, J. B. Hearnshaw, Y. Itow, P. M. Kilmartin, A. Korpela, W. Lin, C. H. Ling, K. Masuda, Y. Matsubara, N. Miyake, M. Motomura, Y. Muraki, M. Nagaya, S. Nakamura, K. Ohnishi, T. Okumura, Y. C. Perrott, N. Rattenbury, To Saito, T. Sako, D. J. Sullivan, W. L. Sweatman, P. J. Tristram, P. C. M. Yock, M. K. Szymanski, M. Kubiak, G. Pietrzynski, R. Poleski, I. Soszynski, L. Wyrzykowski, and K. Ulaczyk. Unbound or Distant Planetary Mass Population Detected by Gravitational Microlensing. *Nature*, 473(7347):349–352, May 2011. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature10092.

Przemek Mroz, Andrzej Udalski, Jan Skowron, Radoslaw Poleski, Szymon Kozlowski, Michal K. Szymanski, Igor Soszynski, Lukasz Wyrzykowski, Pawel Pietrukowicz, Krzysztof Ulaczyk, Dorota Skowron, and Michal Pawlak. No large population of unbound or wide-orbit Jupiter-mass planets. *Nature*, 548(7666):183–186, August 2017. ISSN 0028-0836, 1476-4687. doi: 10.1038/nature23276.

Niall Jeffrey and Benjamin D. Wandelt. Evidence Networks: Simple losses for fast, amortized, neural Bayesian model comparison, January 2024.

George Papamakarios and Iain Murray. Fast ε-free Inference of Simulation Models with Bayesian Conditional Density Estimation, April 2018.

David S. Greenberg, Marcel Nonnenmacher, and Jakob H. Macke. Automatic Posterior Transformation for Likelihood-Free Inference, May 2019.

Keming Zhang, Joshua S. Bloom, B. Scott Gaudi, Francois Lanusse, Casey Lam, and Jessica R. Lu. Real-Time Likelihood-Free Inference of Roman Binary Microlensing Events with Amortized Neural Posterior Estimation. *The Astronomical Journal*, 161(6):262, June 2021. ISSN 0004-6256, 1538-3881. doi: 10.3847/1538-3881/abf42e.

Nolan Smyth, Laurence Perreault-Levasseur, and Yashar Hezaveh. Transformer Embeddings for Fast Microlensing Inference, December 2025.

D. Keehan, J. Yarndley, and N. Rattenbury. Microlensing model inference with normalising flows and reversible jump MCMC. *Astronomy and Computing*, 41:100657, October 2022. ISSN 22131337. doi: 10.1016/j.ascom.2022.100657.

V. Bozza, E. Bachelet, F. Bartolić, T. Heintz, A. Hoag, and M. Hundertmark. VBBinaryLensing: A public package for microlensing light curve computation. *Monthly Notices of the Royal Astronomical Society*, 479(4):5157–5167, October 2018. ISSN 0035-8711, 1365-2966. doi: 10.1093/mnras/sty1791.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, August 2023.

George Papamakarios, Theo Pavlakou, and Iain Murray. Masked Autoregressive Flow for Density Estimation, June 2018.

Jan Boelts, Michael Deistler, Manuel Gloeckler, Álvaro Tejero-Cantero, Jan-Matthis Lueckmann, Guy Moss, Peter Steinbach, Thomas Moreau, Fabio Muratore, Julia Linhart, Conor Durkan, Julius Vetter, Benjamin Kurt Miller, Maternus Herold, Abolfazl Ziaeemehr, Matthijs Pals, Theo Gruner, Sebastian Bischoff, Nastya Krouglova, Richard Gao, Janne K. Lappalainen, Bálint Mucsányi, Felix Pei, Auguste Schulz, Zinovia Stefanidi, Pedro Rodrigues, Cornelius Schröder, Faried Abu Zaid, Jonas Beck, Jaivardhan Kapoor, David S. Greenberg, Pedro J. Gonçalves, and Jakob H. Macke. Sbi reloaded: A toolkit for simulation-based inference workflows, August 2025.

Pablo Lemos, Adam Coogan, Yashar Hezaveh, and Laurence Perreault-Levasseur. Sampling-Based Accuracy Testing of Posterior Estimators for General Inference, June 2023.

Ronan Legin, Alexandre Adam, Yashar Hezaveh, and Laurence Perreault Levasseur. Beyond Gaussian Noise: A Generalized Approach to Likelihood Analysis with non-Gaussian Noise. *The Astrophysical Journal Letters*, 949(2):L41, June 2023. ISSN 2041-8205, 2041-8213. doi: 10.3847/2041-8213/acd645.

Andreas Filipp, Yashar Hezaveh, and Laurence Perreault-Levasseur. Robustness of Neural Ratio and Posterior Estimators to Distributional Shifts for Population-Level Dark Matter Analysis in Strong Gravitational Lensing, November 2024.

Shota Miyazaki and Hajime Kawahara. microJAX: A Differentiable Framework for Microlensing Modeling with GPU-Accelerated Image-Centered Ray Shooting, October 2025.

Haibin Ren and Wei Zhu. A differentiable binary microlensing model using adaptive contour integration method. *The Astronomical Journal*, 169(3):170, March 2025. ISSN 0004-6256, 1538-3881. doi: 10.3847/1538-3881/adb1b2.

Haibin Ren, Weicheng Zang, Wei Zhu, Yoon-Hyun Ryu, Yuchen Tang, Jiyuan Zhang, Michael D. Albrow, Sun-Ju Chung, Andrew Gould, Cheongho Han, Kyu-Ha Hwang, Youn Kil Jung, In-Gu Shin, Yossi Shvartzvald, Hongjing Yang, Jennifer C. Yee, Dong-Jin Kim, Chung-Uk Lee, Byeong-Gon Park, Yunyi Tang, Dan Maoz, Shude Mao, and Qiyue Qian. KMT-2025-BLG-1314 and KMT-2025-BLG-1392: Two microlensing planetary/brown-dwarf candidates analyzed with differentiable code, March 2026.