The Journal of Open Source Software

DeepSZSim: A Python code for fast, tunable Sunyaev–Zeldovich effect submap simulations

Eve M. Vavagiakis [1,2*], **Samuel D. McDermott** [3*], **Humna Awan** [4,5*], **Elaine Y. Ran** [2*], **Kush Banker** [6*], **Samantha Usman** [3*], **Camille Avestruz** [4,5*], **and Brian D. Nord** [3,6,8*]

1 Department of Physics, Duke University, Durham, NC 27708, USA **2** Department of Physics, Cornell University, Ithaca, NY 14853, USA **3** Department of Astronomy and Astrophysics University of Chicago, Chicago, IL **4** Department of Physics University of Michigan, Ann Arbor, MI 48109 **5** Leinweber Center for Theoretical Physics, University of Michigan, Ann Arbor, MI 48109 **6** Department of Physics, The College, University of Chicago, 5801 S Ellis Ave, Chicago, IL 60637 **7** Kavli Institute for Cosmological Physics, University of Chicago, 5801 S Ellis Ave, Chicago, IL 60637 **8** Fermi National Accelerator Laboratory, Batavia, IL 60510, USA ***** These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Josh Borrow

Reviewers:

- @eunseongleee
- @avapolzin

Submitted: 29 October 2024
Published: unpublished

Summary

Current and upcoming measurements of the cosmic microwave background (CMB), the oldest observable light in the universe, elucidate the fundamental physics of our universe, including the development of large-scale structure. Galaxy clusters are the largest gravitationally bound structures in our universe and serve as powerful probes of this large-scale structure J. Richard Bond et al. (1996). Through measurements of galaxy clusters, we can derive insights into the growth of structure and place significant constraints on cosmology. Simulations of galaxy clusters that are well-matched to upcoming data sets are a key tool for addressing systematics (e.g., cluster mass inference) that limit current and future cluster-based cosmology constraints. However, most state-of-the-art simulations are too computationally intensive to produce multiple versions of significant systematic effects: from underlying gas physics to observational modeling uncertainties.

We present DeepSZSim, a novel user-friendly Python framework for generating simulations of the CMB and the thermal Sunyaev–Zel'dovich (tSZ) effect in galaxy clusters, which is based on average galaxy cluster thermal pressure profile models. DeepSZSim includes CMB power spectra generation using CAMB and simulated CMB temperature maps using `namaster` (Alonso et al., 2019), as well as tSZ signal modeling, instrument beam convolution, and noise. By tuning the input parameters based on a cosmology, distributions of halo mass and redshift, and experiment properties (e.g., map depth and observation frequency), users are able to generate a variety of simulated primary and secondary CMB anisotropy images. These simulations offer a fast and flexible method for generating large datasets to test mass inference methods like machine learning and simulation-based inference.

Statement of Need

DeepSZSim fits a unique niche within the plethora of existing CMB primary and secondary anisotropy simulations and software. These simulators and data sets range in size, detail, and accuracy, speed, and ease of use. Most simulators are computationally intensive, and most simulated datasets are not optimized for machine learning training sets. For example, N-body simulations provide the major setting for high-fidelity forward models of the universe Nelson et al. (2019). These simulations are uniquely capable of capturing both large-scale

The Journal of Open Source Software

and small-scale spatial modes of the cosmic web and the CMB and at multiple time steps. To achieve this, N-body simulations have high computational costs and are inflexible with respect to the specific physics models used in the simulations. Mechanistic forward modeling can provide much faster and at least somewhat lower fidelity (less-detailed) simulations Bolliet et al. (2023). Other methods deploy a combination of N-body and mechanistic models (Yamada & others, 2012). Machine learning has been tested for producing simulations of the CMB with generative adversarial networks (Han et al., 2021) and with autoencoders (Rothschild et al., 2022): unfortunately, machine learning methods for generative modeling lack interpretability and uncertainty quantification.

Overall, most software and simulations are difficult to access, especially for researchers new to these subjects, including students: they are not publicly available and are difficult to install. There is a need for codebases that are accessible, inexpensive, and multi-fidelity.

While lacking the fidelity of N-body simulations, DeepSZSim meets a need for a user-friendly, fast, and realistic simulation of CMB primary anisotropies via DeepCMBSim, alongside tSZ signal modeling with highly customizable observational inputs. This is valuable for building and testing a wide range of models for the classification and detection of CMB-related objects like SZ clusters. This is particularly useful for machine learning settings, which typically require a large amount of data, which is often not available from N-body simulations. DeepSZSim is currently being used by Deep Skies researchers to generate large catalogs of CMB+tSZ signal submaps to study galaxy cluster parameter inference via simulation-based inference.

Features

The DeepSZSim package uses the DeepCMBSim package to simulate large-scale modes of the CMB. These modes are combined with the SZ effect to generate realistic simulations of cluster images.

The full package workflow is demonstrated in Figure Figure 1.

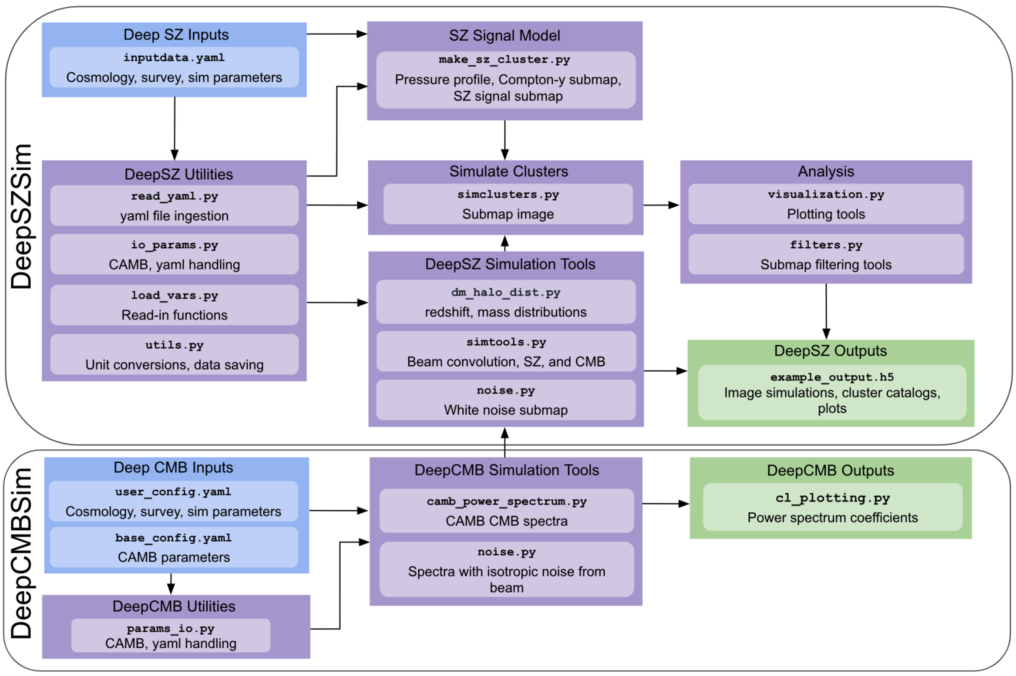

Figure 1: Software workflow for the `DeepSZSim` package, including the elements of `DeepCMBsim`.

The Journal of Open Source Software

DeepCMBSim Module

The DeepCMBsim package combines physical processes and sources of noise in a software framework that enables fast and realistic simulation of the CMB in which key cosmological parameters can be varied. DeepCMBSim simulates correlations of temperatures and polarization signals from the CMB, including large-scale gravitational lensing and BB polarization caused by non-zero tensor-to-scalar ratios.

DeepCMBSim's primary physics module is `camb_power_spectrum`, which defines the `CAMBPowerSpectrum` class. This calls CAMB `[@Lewis:1999bs; @Howlett:2012mh]`. The power spectrum of the noise follows the form in `[@Hu:2001kj]`, assuming statistical independence in the Stokes parameters `[@Knox:1995dq; @Zaldarriaga:1996xe]`. This software allows the user to specify cosmological parameters (e.g., omega matter, omega baryon, the lensing scale, the tensor-to-scalar ratio, which are inputs to CAMB) and experiment parameters (e.g., white noise level, beam size) in a `yaml` configuration file to permit a user-friendly interface to permit reproducible simulations. The default parameters reproduce the Planck 2018 cosmology `[@Planck:2018vyg]`.

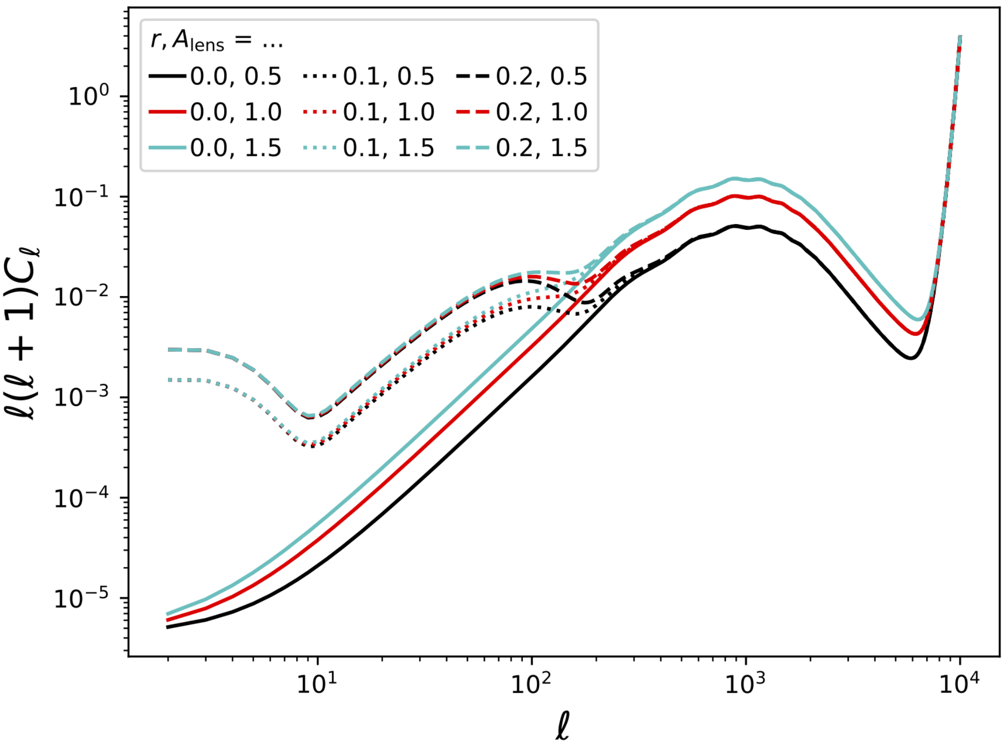

Figure 2: Example output angular spectra for the DeepCMBsim package for a set of tensor-to-scalar ratios r and lens scaling factors A_lens.

We present examples of the primary outputs from DeepCMBSim in Figure Figure 2.

SZ Cluster Simulation

DeepSZsim includes code for producing fast simulations of the thermal SZ effect for galaxy halos of varying mass and redshift, based on average thermal pressure profile fits from Battaglia et al. 2012 `[@Battaglia:2012]`. The output is an array of simulated submaps of the tSZ effect associated with galaxy halos, which can include simulated CMB, instrument beam convolution, and/or white noise.

The user provides inputs to generate an array of redshift and mass (M_200) for dark matter halos, the desired pixel and submap size for the output submaps, and inputs such as experiment properties (observation frequency, noise level, beam size) and a cosmological model. These inputs are easily customizable, or the user can run defaults based on the Atacama Cosmology Telescope [@ACT:2021] and Planck cosmology [@Planck:2019]. Cosmology computations depend on `colossus` [@Colossus:2018] and `astropy` [@Astropy:2013].

From these inputs, pressure profiles, Compton-y profiles, and tSZ signal maps are generated for the dark matter halo array Battaglia et al. (2012). Simulated CMB primary anisotropy maps can be generated through a dependency on `DeepCMBSim`. Final simulated submaps can include instrument beam convolution and white noise [@actnotebooks:2015]. Plotting functions for the simulations and an aperture photometry filter are included as tools. The submap handling functions rely on `pixell` [@pixell:2024].

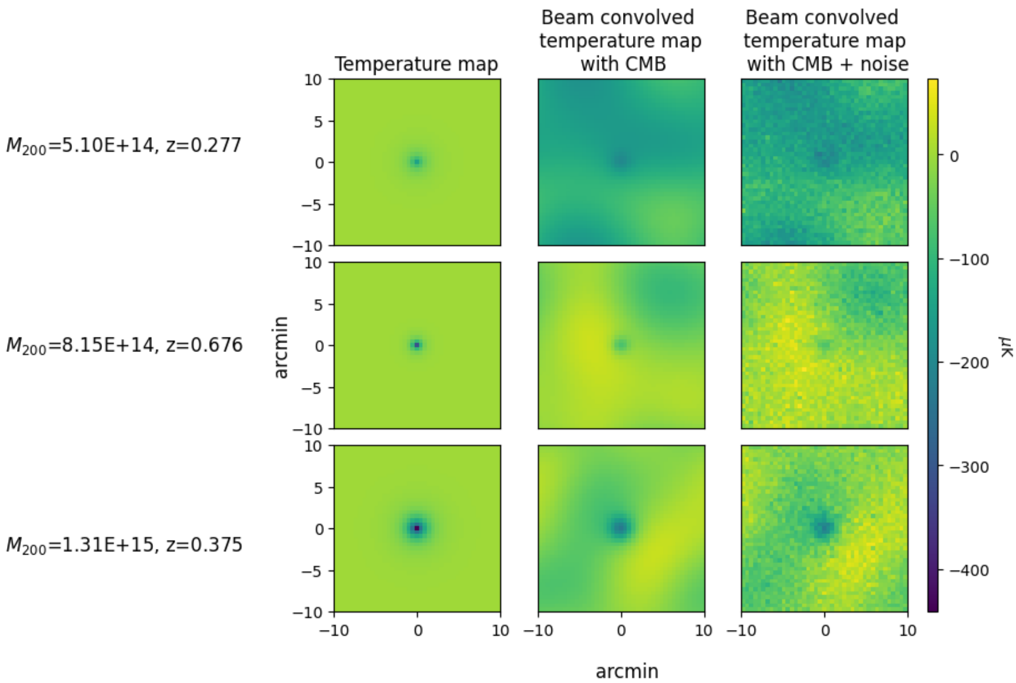

Figure 3: Example outputs for the `DeepSZsim` package for a set of masses, redshifts, and noise configurations.

We present examples of the primary outputs from `DeepSZSim` in Figure Figure 3.

Acknowledgements

E. M. Vavagiakis: methodology, software, writing, supervision. *S. D. McDermott*: methodology, software, writing, supervision. *H. Awan*: methodology, software, writing. *E. Y. Ran*: methodology, software *K. Banker*: methodology, software *S. Usman*: methodology, software *C. Avestruz*: conceptualization, methodology, project administration, funding acquisition, supervision, writing *Brian Nord*: conceptualization, methodology, project administration, funding acquisition, supervision, writing.

We acknowledge contributions from Maggie Voetberg, Junhe Zhang, Helen Tan, Ioana Corescu, and Antwine Willis.

We acknowledge the Deep Skies Lab as a community of multi-domain experts and collaborators who've facilitated an environment of open discussion, idea generation, and collaboration. This community was important for the development of this project.

Work supported by the Fermi National Accelerator Laboratory, managed and operated by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes.

This material is based upon work supported by the National Science Foundation under Grant No. AST-2202237.

This material is based upon work supported by the National Science Foundation under Grant No. 2009944.

References

Alonso, D., Sanchez, J., & Slosar, A. (2019). A unified pseudo-C_ℓ framework. *Monthly Notices of the Royal Astronomical Society*, *484*(3), 4127–4151. https://doi.org/10.1093/mnras/stz093

Arnaud, M., Pratt, G. W., Piffaretti, R., Böhringer, H., Croston, J. H., & Pointecouteau, E. (2010). The universal galaxy cluster pressure profile from a representative sample of nearby systems (REXCESS) and theYSZ–M500relation. *Astronomy and Astrophysics*, *517*, A92. https://doi.org/10.1051/0004-6361/200913416

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. (1986). The Statistics of Peaks of Gaussian Random Fields. *304*, 15. https://doi.org/10.1086/164143

Battaglia, N., Bond, J. R., Pfrommer, C., & Sievers, J. L. (2012). On the Cluster Physics of Sunyaev-Zel'dovich and X-Ray Surveys. II. Deconstructing the Thermal SZ Power Spectrum. *758*(2), 75. https://doi.org/10.1088/0004-637X/758/2/75

Bolliet, B., Kusiak, A., McCarthy, F., Sabyr, A., Surrao, K., Hill, J. C., Chluba, J., Ferraro, S., Hadzhiyska, B., Han, D., Macías-Pérez, J. F., Madhavacheril, M., Maniyar, A., Mehta, Y., Pandey, S., Schaan, E., Sherwin, B., Mancini, A. S., & Zubeldia, Í. (2023). *Class_sz I: Overview* (No. arXiv:2310.18482). arXiv. https://doi.org/10.48550/arXiv.2310.18482

Bond, J. Richard, Kofman, L., & Pogosyan, D. (1996). How filaments of galaxies are woven into the cosmic web. *Nature*, *380*(6575), 603–606. https://doi.org/10.1038/380603a0

Bond, J. R., & Myers, S. T. (1996). The Peak-Patch Picture of Cosmic Catalogs. I. Algorithms. *103*, 1. https://doi.org/10.1086/192267

Han, D., Sehgal, N., & Villaescusa-Navarro, F. (2021). Deep learning simulations of the microwave sky. *Phys. Rev. D*, *104*(12), 123521. https://doi.org/10.1103/PhysRevD.104.123521

Kaiser, N. (1986). Evolution and clustering of rich clusters. *222*, 323–345. https://doi.org/10.1093/mnras/222.2.323

Lesgourgues, J. (2011). *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview* (No. arXiv:1104.2932). arXiv. https://doi.org/10.48550/arXiv.1104.2932

Li, Z., Puglisi, G., Madhavacheril, M. S., & Alvarez, M. A. (2022). Simulated catalogs and maps of radio galaxies at millimeter wavelengths in Websky. *J. Cosmol. Astropart. Phys.*, *2022*(08), 029. https://doi.org/10.1088/1475-7516/2022/08/029

Nelson, D., Springel, V., Pillepich, A., Rodriguez-Gomez, V., Torrey, P., Genel, S., Vogelsberger, M., Pakmor, R., Marinacci, F., Weinberger, R., Kelley, L., Lovell, M., Diemer, B., &

The Journal of Open Source Software

Hernquist, L. (2019). The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, *6*(1), 2. https://doi.org/10.1186/s40668-019-0028-x

Rothschild, T., Nagai, D., Aung, H., Green, S. B., Ntampaka, M., & ZuHone, J. (2022). Emulating Sunyaev-Zeldovich Images of Galaxy Clusters using Auto-Encoders. *Monthly Notices of the Royal Astronomical Society*, *513*(1), 333–344. https://doi.org/10.1093/mnras/stac438

Sehgal, N., Bode, P., Das, S., Hernandez-Monteagudo, C., Huffenberger, K., Lin, Y.-T., Ostriker, J. P., & Trac, H. (2010). Simulations of the Microwave Sky. *ApJ*, *709*(2), 920–936. https://doi.org/10.1088/0004-637X/709/2/920

Stein, G., Alvarez, M. A., Bond, J. R., Engelen, A. van, & Battaglia, N. (2020). The Websky extragalactic CMB simulations. *J. Cosmol. Astropart. Phys.*, *2020*(10), 012. https://doi.org/2020100515130300

Yamada, K., & others. (2012). Imaging Simulations of the Sunyaev-Zel'dovich Effect for ALMA. *Publ. Astron. Soc. Jap.*, *64*, 102. https://doi.org/10.1093/pasj/64.5.102