Toward 1% calibration of CMB Lensing Cluster Mass Estimate with Deep Learning

S. Nakoneczny[1], D. Nagai[2], M. Ntampaka[3], S. Hassan[4], and F. Lanusse[5]

[1]National Centre for Nuclear Research, Astrophysics Division, Pasteura 7, 02-093 Warsaw, Poland

[2]Yale University, Physics Department, P.O. Box 208120, New Haven, CT 06520, USA

[3]Harvard University, Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS-51 Cambridge, MA 02138, USA

[4]New Mexico State University, Astronomy Department, 1780 E University Ave, Las Cruces, NM 88003, USA

[5]University of California in Berkeley, Center for Cosmological Physics, 366 LeConte Hall Berkeley, CA 94720, USA

September 2019

1 Introduction

Galaxy clusters are important probes of cosmology and structure growth in the Universe. Measuring the number density of galaxy clusters as a function of mass and redshift has the potential to yield some of the tightest constraints on the dark energy equation of state (Allen et al. 2011). However, accurately determining their masses remains challenging. Weak gravitational lensing directly probes the total matter content of a galaxy cluster and can provide an unbiased measurement of the total mass (Applegate et al. 2014). The lensed source can be either other galaxies or the cosmic microwave background (CMB). The CMB is particularly useful for high redshift clusters due to the difficulties in observing galaxies with sufficient signal-to-noise (SNR) in other frequencies at very high redshifts. The CMB also has simpler systematics uncertainties, as we have extremely good measurements of the statistical properties of the CMB and we know the precise redshift where the CMB photons originate. However, the SNR of the CMB-cluster lensing signal is comparatively weak for an individual galaxy cluster (Hu et al. 2007).

CMB-cluster lensing refers to the gravitational lensing of the CMB by massive galaxy clusters. The scales of interest are typically a few arc-minutes, corresponding to the angular size of galaxy clusters. On these scales, the CMB is well approximated as a gradient field across the position of the galaxy cluster. The direction of the CMB gradient is nearly independent for temperature and polarization. Gravitational lensing of this gradient by a galaxy cluster produces a dipole-like pattern oriented with the gradient, but with the hot and cold directions swapped (Seljak and Zaldarriaga 2000). We may also refer to this phenomenon as a butterfly pattern. For a given cluster mass and redshift, the magnitude of this dipole scales linearly with the magnitude of the CMB gradient.

The proposed non-machine learning methods of detecting CMB cluster lensing include Wiener filter approach to estimate the lensing distortion (Holder and Kosowsky 2004), a quadratic estimator (QE) which uses the correlation between the unlensed CMB gradient and the lensing signal (Melin and Bartlett 2015), maximum likelihood estimator (MLE) by fitting lensing templates to the lensed CMB maps (Lewis and King 2006; Baxter et al. 2015), and ML lensing estimator using a pixel-space likelihood (Raghunathan et al. 2017). While all these non-machine learning methods show different level of success, machine learning techniques, particularly convolutional neutral networks (CNN), are emerging as a powerful tool to perform classification and regression in Astronomy (Ntampaka et al. 2019b). Cluster masses were successfully predicted with CNN based on X-ray simulations (Ntampaka et al. 2019a). The CMB background was reconstructed in order to remove lensing effects on scales of several degrees (Caldeira et al. 2019). However, there is still no work which would precisely estimate cluster masses for large enough sample of clusters.

The ultimate goal of this project is to perform precise cluster mass estimation based on CMB cluster lensing for sample of clusters large enough for analyses of large cosmological surveys, such as Simons Observatory (Ade et al. 2019) and CMB-S4 (Abazajian et al. 2019). These surveys will significantly improve the sensitivity and resolutions of the CMB measurements, promising to advance our understanding of the cosmic structure formation from early universe to non-linear structure formation of the universe. At the same time, advances in computational cosmology has recently enabled creation of all-sky CMB lensing maps based on the light cone simulations of LSS (Takahashi et al. 2017). Such simulations allow to build a precise estimator of cluster mass, and perform tests before the final data is gathered by ongoing and forthcoming CMB experiments. The fact that MLE, which can be also considered as a machine learning method, achieves better results than QE, which fits lensing templates, justifies usage of machine learning and makes it highly complementary to the work done so far. It is in fact the most promising approach in extraction of small lensing signals from large and complex datasets. Therefore, we train deep learning method, CNN, to estimate cluster masses based on full-sky lensing simulation (Takahashi et al. 2017). In section 2, we present the methodology, including data description, patch extraction, testing methods and CNN training. Section 3 gives results for experiments with different source plane redshifts. In the last section 4 we conclude and provide outline of the future steps.

2 Methodology

2.1 Ray tracing simulation

We use a full-sky gravitational lensing simulation data (Takahashi et al. 2017). It provides 108 data sets generated by performing multiple-lens plane ray-tracing through high-resolution cosmological N-body simulations. The data sets include full-sky convergence and shear maps from redshifts $z = 0.05$ to 5.3 at intervals of $150h^{-1}$Mpc comoving radial distance (corresponding to a redshift interval of $\Delta z \approx 0.05$ at the nearby universe), enabling the construction of a mock shear catalog for an arbitrary source distribution up to z=5.3. The simulation also yields maps of gravitational lensing deflections for a source redshift at the last scattering surface. The CMB maps are available in two resolutions: $N_{side} = 4096, 8192$ corresponding to pixel sizes of 0.86 and 0.43 arcmin, respectively.

2.2 Data processing

The strength of CMB lensing pattern depends not only on cluster mass but also on cluster redshift. To start with a problem which gives consistent behaviour of cluster and lensing signal, we limit the halo dataset to minimum mass M_{200b} (the radius at a mean halo density 200 times larger than the background density) to $10^{14} M_{\odot}$ and redshift range to $0.1 - 0.3$. The data cuts are performed on M_{200b} to keep consistency between our approach and a dataset of extracted CMB images which we received from the authors of simulation in the first week of the project. When training machine learning (ML) model, we choose to use M_{500c} (the radii at halo density 500 times larger than the cosmological critical density) in order to stay consistent with cluster mass predictions from X-ray images (Ntampaka et al. 2019a). This inconsistency between masses within this work should be simplified in the later stages of the project. From the mass function analysis, we know that the massive clusters are rare and their abundance is possibly several orders of magnitudes less than the abundance of low mass clusters. Hence, as a starting simpler test, we flatten the mass function to create a homogeneous training dataset with equal samples for each mass bin. The distribution is flatten by binning it into 100 bins and limiting size of each bin to 100 halos, as shown in figure 1. This gives about 3000 halos for one realization of simulated data.

In order to extract CMB patches around halo locations, we use the HEALPix[1] library (Górski et al. 2005). We use simple cartesian projection, and every time a patch is extracted, the map is rotated in such a way which makes halo location appear in a projection's center, which minimizes any projection effects.

[1]http://healpix.sourceforge.net

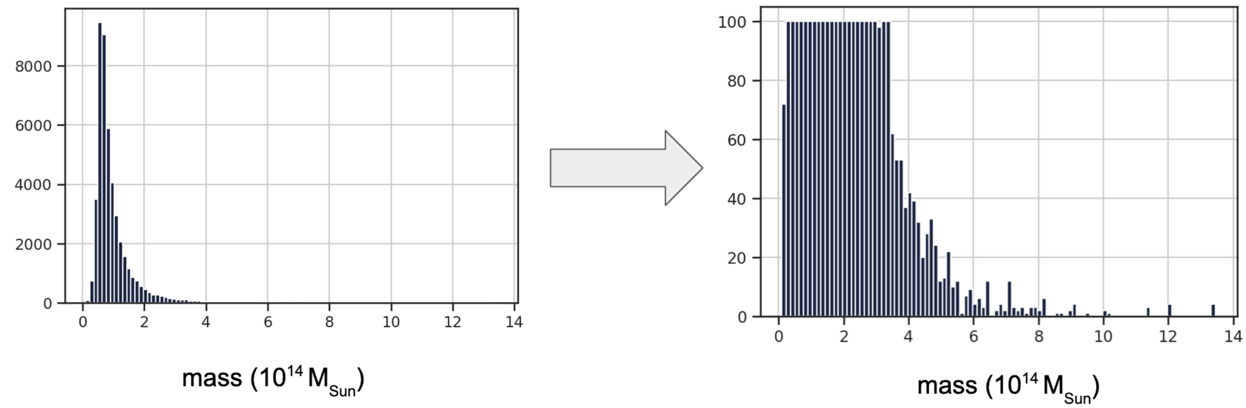

Figure 1: Histogram of M_{500c} for one realization of simulated data. *Left:* before flattening, *right:* after flattening. The completely unbalanced original set should be well balanced in mass range (0, 4) after the flattening.

The pixel size in projection is 0.5 arcmin. The patches extracted for CNN processing are of size 32x32 pixels which corresponds to 16x16 arcmin patches. This covers the most important lensing area which size is that of a cluster, about 2-4 arcmin in diameter. Figure 2 shows convergence and CMB maps for the most massive cluster with $M_{500c} = 13.45 \cdot 10^{14} M_{\odot}$. From left, the first image shows convergence from ray tracing which originates at redshift 1. It shows a clean convergence peek. The second image shows convergence for source redshift of 1100, related to CMB. Here, we can see that the images are getting more noisy due to projection effects from redshift 1100 to 0. We smooth those images with Gaussian kernel which sigma equals 0.5 arcmin (1 pixel), as shown in the 3rd image. The last image shows CMB temperature in μK, which is expected to contain a lensing "butterfly" pattern of scale around $10\mu K$. Unfortunately, we did not manage to find those patterns in the simulation. We speculate that the effect might be smaller than expect due to, for instance, projection effects. To clarify on that issue, it wold be necessary to look into a difference of lensed and unlensed images and look for the pattern in the image difference. The focus of the project goes to learning the cluster mass from convergence maps in order to see how projection effects between source redshift of 1 and 1100 affect the mass estimation.

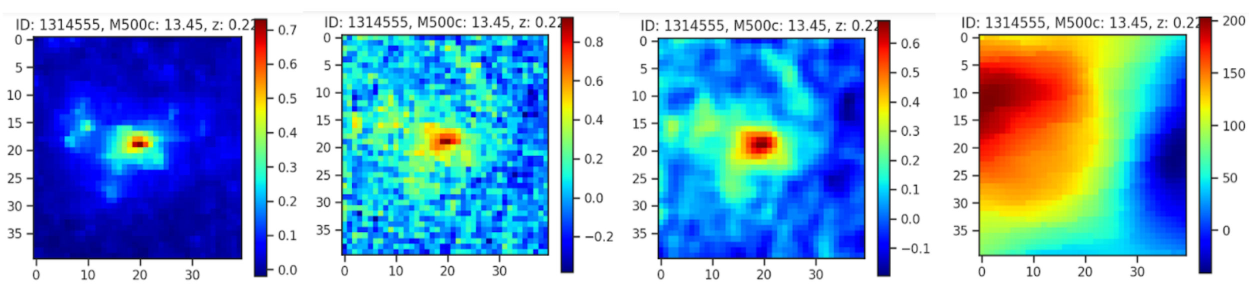

Figure 2: Extracted patches around the most massive halo with $M_{500c} = 13.45 M_{\odot}$. From left, *first:* convergence at source redshift 1, *second:* convergence at source redshift 1100 (CMB redshift), *third:* the same as second, smoothed with Gaussian kernel of size 0.5 arcmin (1 pixel), *fourth:* CMB temperature in μK. The lensing patterns is expected to be visible in the last image.

Figure 3 shows convergence maps for 4 clusters at $M_{500c} \approx 10^{14} M_{\odot}$. We can see that apart from central convergence peeks, there is also large scale structure imprinted around the clusters. This effect increase when we move from redshift 1 to 1100. We can see that already at source redshift 1, values in the central convergence peeks vary from 0.1 to 0.3, which makes it hard to precisely learn cluster mass from projected convergence maps.

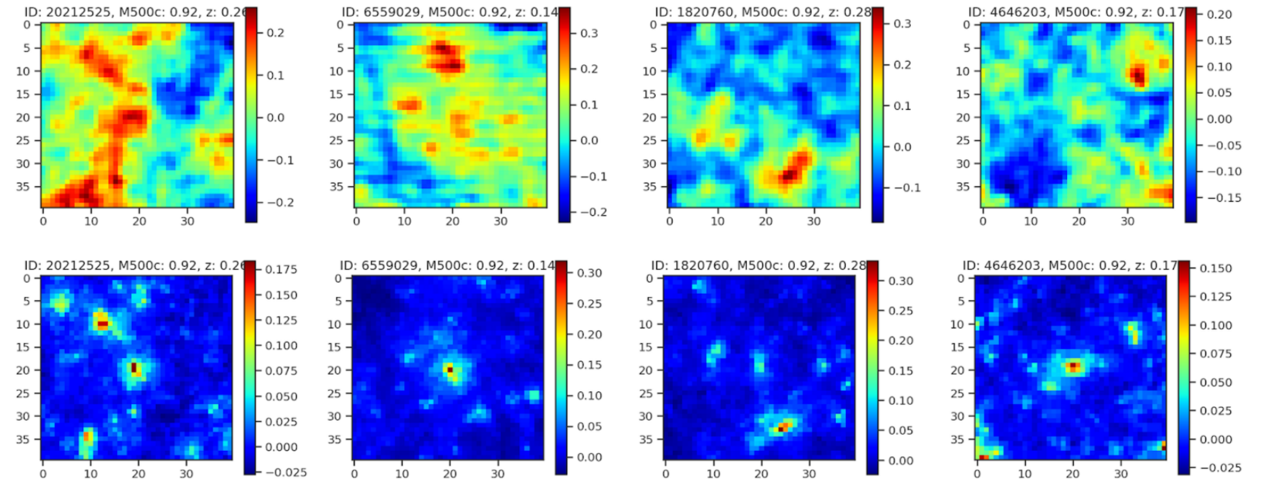

Figure 3: Comparison of redshift plane 1100 (top row) and 1 (bottom row) for 4 clusters (columns) at $M_{500c} \approx 10^{14} M_{\odot}$. LSS projection effects are already visible at source redshift 1 and become severe at source redshift 1100.

2.3 Testing methods

In order to validate the machine learning model, we extract 3 different validation sets from the training one. First, we choose 5% of both the most and the least massive clusters to create 2 validation sets. This leaves 90% of the training set from which we randomly pick 10% (9% in terms of the original dataset size) as the third and last validation set. The remaining data is used as training. As a given halo can be present at many locations, due to several ray paths affected by the halo gravitational potential, we use a group split about halo IDs in the creation of random validation set. It ensures that training halos are not used in validation. Such strategy of choosing validation sets evaluates the machine learning model on the hardest possible problems which include extrapolation tests on highest and lowest mass clusters. This is beneficial due to several reasons: it properly tests model overfitting, allows to trace such effects already during a training and shows how a network performs on different halo masses. The same random score may be achieved with completely different performance on the high and low mass end, and usually the non random validation sets lead to stopping the network training earlier than random validation sets. On those subsets, we report standard mean square error (MSE) and coefficient of determination (R2) which ranges from -1 to 1, with 0 corresponding to prediction of mean target values.

2.4 CNN architecture

The CNN architecture consists of the following layers:

1. input: 20x20 (0.5 arcmin pixel size)
2. convolution: 3x3x16
3. convolution: 3x3x32
4. convolution: 3x3x64
5. flatten (0.1 dropout): 14x14x64 = 12544
6. dense (0.1 dropout): 200
7. dense: 100
8. dense: 20

9. output: 1

The network was build and trained using Keras (Chollet et al. 2015) with Tensorflow backend (Abadi et al. 2015). All the layers, except for the output, use ReLU activation. The images were normalized using Keras feature wise center and normalization. Horizontal and vertical flips are used for data augmentation. Choosing a small image size of 20x20 pixels, with pixel size 0.5 arcmin, which crops the most around halo center proves to give the best results. For such small images, max pooling is not applied. We note that in order for the network to learn LSS projection effect it might be beneficial to include more area around the cluster center. More experiments would have to be performed to see whether such effects can be learned with enough precision to improve mass estimations. We also test different smoothing scales for the redshift source 1100, which shows that a small smoothing scale of 0.5 arcmin gives some improvement, while using larger smoothing tends to loose some information.

3 Results

In two main experiments, we estimate cluster mass from convergence maps originating at source redshift 1100 and around 1. Figure 4 shows training histories for those two experiments. The left plot visualizes the two most common errors: training and random test. Mass estimation based on cleaner maps from source redshift around 1 gives much better results, which shows how projections effects introduce "noise" into convergence maps. The closer plane experiment gives MSE=0.37, R2=0.68, and the CMB plane experiment gives MSE=0.64, R2=0.45. The low and high mass end tests give us interesting results, with much better results for the low mass test. This is contrary to the intuition, as we would expect high mass clusters to imprint stronger on the convergence maps allowing for easier mass estimation. The explanation is that high mass clusters are found in more dense areas of LSS which makes projection effects more visible. Figure 5 inspects relation between predicted and true mass. In the left plot, we can see results for experiment with the CMB plane, and the right one shows closer plane around redshift 1. The closer plane gives much better results at low mass end, but the problems due to LSS projection are significant from $M_{500c} > 10^{14} M_{\odot}$.

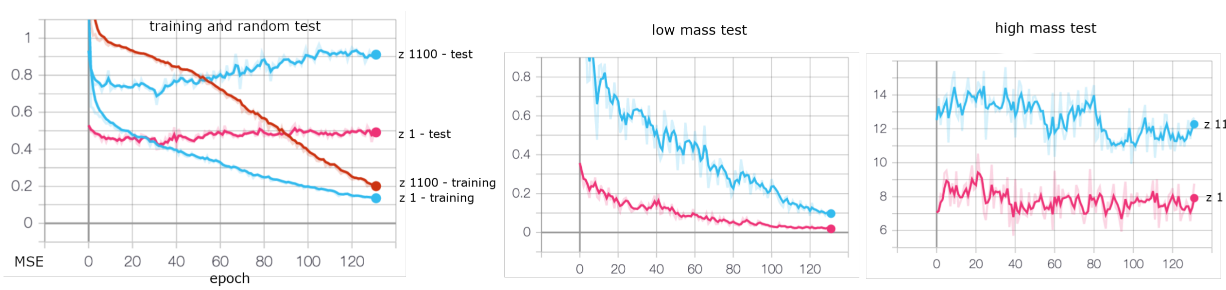

Figure 4: Training histories for experiments at plane redshift 1100 and around 1. *Left*: training and random test errors, *middle*: low mass end errors, *right*: high mass end errors. Source redshift around 1 and loow mass clusters give significantly better results, due to less projection effects from LSS.

4 Conclusion

We tried to estimate cluster mass based on CMB cluster lensing signal. This signal was already detected in surveys such as Planck, however, there is no pipeline which would perform cluster mass estimation for cluster sample large enough for analyses of forthcoming cosmological surveys, such as Simons Observatory and CMB-S4. On the theoretical side, recent advances in computational cosmology and data science provide access to all-sky, high-resolution CMB lensing simulations as well as a powerful and new machine learning tool, such as the convolution neural network (CNN). It is the most promising approach in detecting faint lensing signals in large and complex CMB datasets.

In the chosen lensing simulation (Takahashi et al. 2017), we were unable to visually confirm lensing patterns on scales of cluster size, 2-4 arcmin, and $10\mu K$ as estimated theoretically. Instead, we choose to

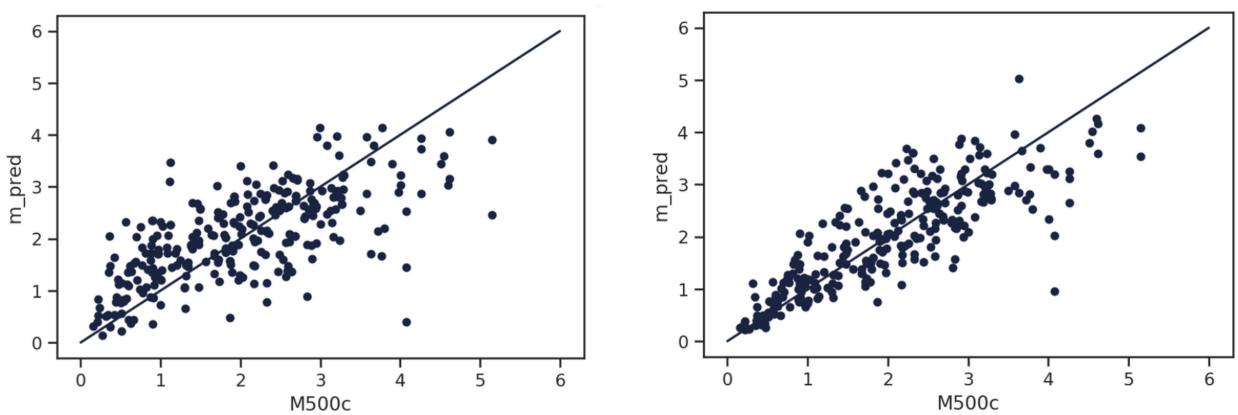

Figure 5: Scatter plots of predicted against true mass. *Left*: plane redshift 1100, *right*: plane redshift 1.

work with convergence maps, which are source of any lensing phenomenons. Findings of this analysis should be also important and helpful in future processing of temperature maps. We perform cluster mass estimation based on a training set of 3000 galaxy clusters. At CMB plane, we achieve MSE=0.64, R2=0.45. The results are not adequate for cosmological analysis yet, but we manage to analyse the projection effects originating at different source redshifts with respect to different mass clusters.

In order to create a working pipeline for cluster mass estimation based on CMB cluster lensing signal in temperature maps, we consider the following steps:

- make sure that CMB cluster lensing is detectable in training data
  - use the ray tracing code to get the non-lensed CMB temperature maps and look for lensing patterns in difference between original and lensed images
  - use a different training dataset on which standard CMB cluster lensing detection methods already work
- train a machine learning model for cluster lensing pattern detection
  - process lensed CMB temperature maps with CNN
  - variational auto-encoder for non-Gaussianities detection
- add cluster redshift information as additional input to CNN
- comparison with standard methods of CMB cluster lensing detection: QE, MLE
- CNN model interpretability

Our key results show how LSS projection effects highly affect the mass estimation, being significantly harder to process at higher source redshifts, especially for high mass clusters which are found in more dense regions of LSS. The best way to further look for the CMB cluster lensing signal would be to compare lensed and unlensed CMB images. Since the CMB unlensed maps were not readily available publicly, we plan to perform the ray-tracing simulations of CMB photons ourselves using the RAYTRIX code (Hamana and Mellier 2001; Sato et al. 2009; Schäfer et al. 2012). A promising machine learning method to test next is variational auto-encoder. Benefiting from the fact that CMB consists of Gaussian patterns, we will train a model to encode non-lensed CMB images into several Gaussian parameters. Using a trained network on lensed images would yield images with lensing effects removed and easily accessible in a difference of original and recreated images. Such network would be beneficial not only for lensing detection, but for any other problem resulting in non-Gaussian noise in CMB maps, such as thermal and kinematic Sunyaev-Zeldovich effects.

Acknowledgements

Many thanks to Kavli Foundation and organizers of the Kavli Summer Program in Astrophysics 2019 at University of California in Santa Cruz, where the project has been performed. Some of the results in this paper have been derived using the healpy and HEALPix package.

References

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

K. Abazajian, G. Addison, P. Adshead, Z. Ahmed, S. W. Allen, D. Alonso, M. Alvarez, A. Anderson, K. S. Arnold, C. Baccigalupi, and et al. CMB-S4 Science Case, Reference Design, and Project Plan. *arXiv e-prints*, July 2019.

Peter Ade, James Aguirre, Zeeshan Ahmed, Simone Aiola, Aamir Ali, David Alonso, Marcelo A. Alvarez, Kam Arnold, Peter Ashton, Jason Austermann, Humna Awan, Carlo Baccigalupi, Taylor Baildon, Darcy Barron, Nick Battaglia, Richard Battye, Eric Baxter, Andrew Bazarko, James A. Beall, Rachel Bean, Dominic Beck, Shawn Beckman, Benjamin Beringue, Federico Bianchini, Steven Boada, David Boettger, J. Richard Bond, Julian Borrill, Michael L. Brown, Sarah Marie Bruno, Sean Bryan, Erminia Calabrese, Victoria Calafut, Paolo Calisse, Julien Carron, Anthony Challinor, Grace Chesmore, Yuji Chinone, Jens Chluba, Hsiao-Mei Sherry Cho, Steve Choi, Gabriele Coppi, Nicholas F. Cothard, Kevin Coughlin, Devin Crichton, Kevin D. Crowley, Kevin T. Crowley, Ari Cukierman, John M. D'Ewart, Rolando Dünner, Tijmen de Haan, Mark Devlin, Simon Dicker, Joy Didier, Matt Dobbs, Bradley Dober, Cody J. Duell, Shannon Duff, Adri Duivenvoorden, Jo Dunkley, John Dusatko, Josquin Errard, Giulio Fabbian, Stephen Feeney, Simone Ferraro, Pedro Fluxà, Katherine Freese, Josef C. Frisch, Andrei Frolov, George Fuller, Brittany Fuzia, Nicholas Galitzki, Patricio A. Gallardo, Jose Tomas Galvez Ghersi, Jiansong Gao, Eric Gawiser, Martina Gerbino, Vera Gluscevic, Neil Goeckner-Wald, Joseph Golec, Sam Gordon, Megan Gralla, Daniel Green, Arpi Grigorian, John Groh, Chris Groppi, Yilun Guan, Jon E. Gudmundsson, Dongwon Han, Peter Hargrave, Masaya Hasegawa, Matthew Hasselfield, Makoto Hattori, Victor Haynes, Masashi Hazumi, Yizhou He, Erin Healy, Shawn W. Henderson, Carlos Hervias-Caimapo, Charles A. Hill, J. Colin Hill, Gene Hilton, Matt Hilton, Adam D. Hincks, Gary Hinshaw, Renée Hložek, Shirley Ho, Shuay-Pwu Patty Ho, Logan Howe, Zhiqi Huang, Johannes Hubmayr, Kevin Huffenberger, John P. Hughes, Anna Ijjas, Margaret Ikape, Kent Irwin, Andrew H. Jaffe, Bhuvnesh Jain, Oliver Jeong, Daisuke Kaneko, Ethan D. Karpel, Nobuhiko Katayama, Brian Keating, Sarah S. Kernasovskiy, Reijo Keskitalo, Theodore Kisner, Kenji Kiuchi, Jeff Klein, Kenda Knowles, Brian Koopman, Arthur Kosowsky, Nicoletta Krachmalnicoff, Stephen E. Kuenstner, Chao-Lin Kuo, Akito Kusaka, Jacob Lashner, Adrian Lee, Eunseong Lee, David Leon, Jason S. Y. Leung, Antony Lewis, Yaqiong Li, Zack Li, Michele Limon, Eric Linder, Carlos Lopez-Caraballo, Thibaut Louis, Lindsay Lowry, Marius Lungu, Mathew Madhavacheril, Daisy Mak, Felipe Maldonado, Hamdi Mani, Ben Mates, Frederick Matsuda, Loïc Maurin, Phil Mauskopf, Andrew May, Nialh McCallum, Chris McKenney, Jeff McMahon, P. Daniel Meerburg, Joel Meyers, Amber Miller, Mark Mirmelstein, Kavilan Moodley, Moritz Munchmeyer, Charles Munson, Sigurd Naess, Federico Nati, Martin Navaroli, Laura Newburgh, Ho Nam Nguyen, Michael Niemack, Haruki Nishino, John Orlowski-Scherer, Lyman Page, Bruce Partridge, Julien Peloton, Francesca Perrotta, Lucio Piccirillo, Giampaolo Pisano, Davide Poletti, Roberto Puddu, Giuseppe Puglisi, Chris Raum, Christian L. Reichardt, Mathieu Remazeilles, Yoel Rephaeli, Dominik Riechers, Felipe Rojas, Anirban Roy, Sharon Sadeh, Yuki Sakurai, Maria Salatino, Mayuri Sathyanarayana Rao, Emmanuel Schaan, Marcel Schmittfull, Neelima Sehgal, Joseph Seibert, Uros Seljak, Blake Sherwin, Meir Shimon, Carlos Sierra, Jonathan Sievers, Precious Sikhosana, Maximiliano Silva-Feaver, Sara M. Simon, Adrian Sinclair, Praween Siritanasak, Kendrick Smith, Stephen R. Smith, David Spergel, Suzanne T. Staggs, George Stein, Jason R. Stevens, Radek Stompor, Aritoki Suzuki, Osamu Tajima, Satoru Takakura, Grant Teply, Daniel B. Thomas, Ben Thorne, Robert Thornton, Hy Trac, Calvin Tsai, Carole Tucker, Joel Ullom, Sunny Vagnozzi, Alexander van Engelen, Jeff Van Lanen, Daniel D. Van Winkle, Eve M. Vavagiakis, Clara Vergès, Michael Vissers, Kasey Wagoner, Samantha Walker, Jon Ward, Ben Westbrook, Nathan Whitehorn, Jason Williams, Joel Williams, Edward J. Wollack, Zhilei Xu, Byeonghee Yu, Cyndia Yu, Fernando Zago, Hezi Zhang, Ningfeng Zhu, and The Simons Observatory collaboration. The Simons Observatory: science goals and forecasts. , 2019(2):056, Feb 2019. doi: 10.1088/1475-7516/2019/02/056.

S. W. Allen, A. E. Evrard, and A. B. Mantz. Cosmological Parameters from Observations of Galaxy Clusters. , 49:409–470, September 2011. doi: 10.1146/annurev-astro-081710-102514.

Douglas E. Applegate, Anja von der Linden, Patrick L. Kelly, Mark T. Allen, Steven W. Allen, Patricia R. Burchat, David L. Burke, Harald Ebeling, Adam Mantz, and R. Glenn Morris. Weighing the Giants - III. Methods and measurements of accurate galaxy cluster weak-lensing masses. , 439(1):48–72, Mar 2014. doi: 10.1093/mnras/stt2129.

E. J. Baxter, R. Keisler, S. Dodelson, K. A. Aird, S. W. Allen, M. L. N. Ashby, M. Bautz, M. Bayliss, B. A. Benson, L. E. Bleem, S. Bocquet, M. Brodwin, J. E. Carlstrom, C. L. Chang, I. Chiu, H. M. Cho, A. Clocchiatti, T. M. Crawford, A. T. Crites, S. Desai, J. P. Dietrich, T. de Haan, M. A. Dobbs, R. J. Foley, W. R. Forman, E. M. George, M. D. Gladders, A. H. Gonzalez, N. W. Halverson, N. L. Harrington, C. Hennig, H. Hoekstra, G. P. Holder, W. L. Holzapfel, Z. Hou, J. D. Hrubes, C. Jones, L. Knox, A. T. Lee, E. M. Leitch, J. Liu, M. Lueker, D. Luong-Van, A. Mantz, D. P. Marrone, M. McDonald, J. J. McMahon, S. S. Meyer, M. Millea, L. M. Mocanu, S. S. Murray, S. Padin, C. Pryke, C. L. Reichardt, A. Rest, J. E. Ruhl, B. R. Saliwanchik, A. Saro, J. T. Sayre, K. K. Schaffer, E. Shirokoff, J. Song, H. G. Spieler, B. Stalder, S. A. Stanford, Z. Staniszewski, A. A. Stark, K. T. Story, A. van Engelen, K. Vanderlinde, J. D. Vieira, A. Vikhlinin, R. Williamson, O. Zahn, and A. Zenteno. A Measurement of Gravitational Lensing of the Cosmic Microwave Background by Galaxy Clusters Using Data from the South Pole Telescope. , 806(2):247, Jun 2015. doi: 10.1088/0004-637X/806/2/247.

J. Caldeira, W. L. K. Wu, B. Nord, C. Avestruz, S. Trivedi, and K. T. Story. DeepCMB: Lensing reconstruction of the cosmic microwave background with deep neural networks. *Astronomy and Computing*, 28: 100307, July 2019. doi: 10.1016/j.ascom.2019.100307.

François Chollet et al. Keras. https://keras.io, 2015.

K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. , 622:759–771, April 2005. doi: 10.1086/427976.

T. Hamana and Y. Mellier. Numerical study of the statistical properties of the lensing excursion angles. , 327:169–176, October 2001. doi: 10.1046/j.1365-8711.2001.04685.x.

Gilbert Holder and Arthur Kosowsky. Gravitational Lensing of the Microwave Background by Galaxy Clusters. , 616(1):8–15, Nov 2004. doi: 10.1086/424808.

Wayne Hu, Simon DeDeo, and Chris Vale. Cluster mass estimators from CMB temperature and polarization lensing. *New Journal of Physics*, 9(12):441, Dec 2007. doi: 10.1088/1367-2630/9/12/441.

Antony Lewis and Lindsay King. Cluster masses from CMB and galaxy weak lensing. , 73(6):063006, Mar 2006. doi: 10.1103/PhysRevD.73.063006.

Jean-Baptiste Melin and James G. Bartlett. Measuring cluster masses with CMB lensing: a statistical approach. , 578:A21, Jun 2015. doi: 10.1051/0004-6361/201424720.

M. Ntampaka, J. ZuHone, D. Eisenstein, D. Nagai, A. Vikhlinin, L. Hernquist, F. Marinacci, D. Nelson, R. Pakmor, A. Pillepich, P. Torrey, and M. Vogelsberger. A Deep Learning Approach to Galaxy Cluster X-Ray Masses. , 876:82, May 2019a. doi: 10.3847/1538-4357/ab14eb.

Michelle Ntampaka, Camille Avestruz, Steven Boada, Joao Caldeira, Jessi Cisewski-Kehe, Rosanne Di Stefano, Cora Dvorkin, August E. Evrard, Arya Farahi, Doug Finkbeiner, Shy Genel, Alyssa Goodman, Andy Goulding, Shirley Ho, Arthur Kosowsky, Paul La Plante, Francois Lanusse, Michelle Lochner, Rachel Mandelbaum, Daisuke Nagai, Jeffrey A. Newman, Brian Nord, J. E. G. Peek, Austin Peel, Barnabas Poczos, Markus Michael Rau, Aneta Siemiginowska, Dougal J. Sutherland, Hy Trac, and Benjamin Wandelt. The Role of Machine Learning in the Next Decade of Cosmology. , 51(3):14, May 2019b.

S. Raghunathan, S. Patil, E. J. Baxter, F. Bianchini, L. E. Bleem, T. M. Crawford, G. P. Holder, A. Manzotti, and C. L. Reichardt. Measuring galaxy cluster masses with CMB lensing using a Maximum Likelihood estimator: statistical and systematic error budgets for future experiments. , 8:030, August 2017. doi: 10.1088/1475-7516/2017/08/030.

M. Sato, T. Hamana, R. Takahashi, M. Takada, N. Yoshida, T. Matsubara, and N. Sugiyama. Simulations of Wide-Field Weak Lensing Surveys. I. Basic Statistics and Non-Gaussian Effects. , 701:945–954, August 2009. doi: 10.1088/0004-637X/701/2/945.

B. M. Schäfer, L. Heisenberg, A. F. Kalovidouris, and D. J. Bacon. On the validity of the Born approximation for weak cosmic flexions. , 420:455–467, February 2012. doi: 10.1111/j.1365-2966.2011.20051.x.

Uroš Seljak and Matias Zaldarriaga. Lensing-induced Cluster Signatures in the Cosmic Microwave Background. , 538(1):57–64, Jul 2000. doi: 10.1086/309098.

R. Takahashi, T. Hamana, M. Shirasaki, T. Namikawa, T. Nishimichi, K. Osato, and K. Shiroyama. Full-sky Gravitational Lensing Simulation for Large-area Galaxy Surveys and Cosmic Microwave Background Experiments. , 850:24, November 2017. doi: 10.3847/1538-4357/aa943d.