Machine Learning, Markov Chain Monte Carlo, and Optimal Algorithms to Characterize the AdvACT Kilopixel Transition-Edge Sensor Arrays

Maria Salatino, Jason Austermann, James A. Beall, Steve Choi, Kevin T. Crowley, Shannon Duff, Shawn W. Henderson, Gene Hilton, S.-P. P. Ho, Johannes Hubmayr, Yaqiong Li, Michael D. Niemack, Sara M. Simon, Suzanne T. Staggs, and Edward J. Wollack

***Abstract*—Next-generation focal planes comprising dozens of kilopixel transition-edge sensor (TES) arrays require new methods to rapidly screen candidate arrays, evaluate array non-idealities in the field, identify outlier devices for removal, and optimize the array performance in the field. We demonstrate robust methods to estimate TES parameters (critical temperatures and thermal conductivity parameters) and their uncertainties using a custom Markov Chain Monte Carlo (MCMC) algorithm. We also constrain systematic effects in estimating the TES parameters from non-isothermal current-voltage curves (IVs) at approximately a $\sim$3% level. Additionally, for the first time, we have applied Machine Learning (ML) algorithms to tune detector arrays and optimize their performance.**

***Index Terms*—Kilopixel focal planes, Machine Learning, Markov Chain Monte Carlo, Transition-Edge Sensor.**

I. Introduction

FUTURE Cosmic Microwave Background (CMB) experiments will field hundreds of thousands of detectors. To prepare for these vast numbers of detectors, we present both studies and methods of operation for kilopixel arrays on Advanced ACTPol (AdvACT) [1], the new, bolometric camera of the Atacama Cosmology Telescope [2]. AdvACT is currently observing with about 6,000 polarization-sensitive AlMn TESes [3], [4] divided into three arrays: one high frequency (HF, 150 and 230 GHz) array [5] and two medium frequency (MF1 and MF2, 90 and 150 GHz) arrays [6]. The TESes are voltage-biased and read out with Time Domain Multiplexing (TDM) with a multiplexing factor (MUX) of 64 [7]–[9]. Each array is biased with up to 24 bias lines.

We present algorithms and methods developed using laboratory data taken with the AdvACT HF and MF2 arrays to measure detector parameters, quantify the level of detector systematics, and optimize the array operation. These techniques can be applied both to future CMB experiments with kilopixel focal planes (e.g., AliCPT-1, Simons Observatory [10], [11] and CMB-S4 [12], [13]) and to systems with different readout architectures (Frequency Domain Multiplexing [14], [15], microwave-SQUID multiplexing [16], [17]).

A common method used to estimate the TES parameters in a detector array is to measure the IV characteristics of the TESes at a series of different bath temperatures T_{bath} [5], [6]. We call a set of such measurements a temperature ramp acquisition. In Section II, we explore the sources and effects of non-isothermal IV acquisitions on the derived TES parameters. We then explore the intrinsic degeneracy in the estimated TES parameters using an MCMC algorithm (Section III).

An additional challenge in large detector arrays is tuning the voltage biases to reach optimal sensitivity. As the array size increases, the complexity and scale of this challenge requires automating this process [18]. For this reason, the next generation of CMB cameras will require innovative approaches to tune the detector arrays: for the first time, to our knowledge, we have applied ML algorithms to CMB instrumentation to address this concern. We present a practical example of this tuning procedure to rapidly optimize the detector array performance (Section IV). We offer conclusions in Section V.

II. Non-Isothermal IVs

During the IV acquisition, a decreasing current (from a few mAs to a μA) is sent through each bias line to drive all the detectors from normal to superconducting state. Subsequent analysis of these IVs enables estimates of each TES's critical temperature

Manuscript received October 31, 2018; accepted March 26, 2019. Date of publication April 11, 2019; date of current version June 20, 2019. This work was supported by the U.S. National Science Foundation under Grant 1440226. The development of multichroic detectors and lenses was supported by the NASA under Grant NNX13AE56G and Grant NNX14AB58G. *(Corresponding author: Maria Salatino.)*

M. Salatino is with the Physics Department, Princeton University, Princeton, NJ 08540 USA. She is now with the Stanford University and KIPAC, Stanford, CA 94304 USA (e-mail: marias5@stanford.edu).

J. Austermann, J. A. Beall, S. Duff, G. Hilton, and J. Hubmayr are with NIST, Boulder, CO 80305 USA (e-mail: jason.austermann@nist.gov; james.beall@nist.gov; shannon.duff@nist.gov; gene.hilton@nist.gov; johannes.hubmayr@nist.gov).

S. Choi, K. T. Crowley, S.-P. P. Ho, Y. Li, and S. T. Staggs are with the Physics Department, Princeton University, Princeton, NJ 08540 USA (e-mail: ktc35@berkeley.edu; spho@princeton.edu; yaqiongl@princeton.edu; staggs@princeton.edu).

S. W. Henderson and M. D. Niemack are with Cornell University, Ithaca, NY 14853 USA (e-mail: shawn@slac.stanford.edu; niemack@cornell.edu).

S. M. Simon is with Michigan University, Ann Arbor, MI 48109-1040 USA (e-mail: smsimon@umich.edu).

E. J. Wollack is with NASA/Goddard Space Flight Center, Greenbelt, MD 20771 USA (e-mail: edward.j.wollack@nasa.gov).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASC.2019.2910542

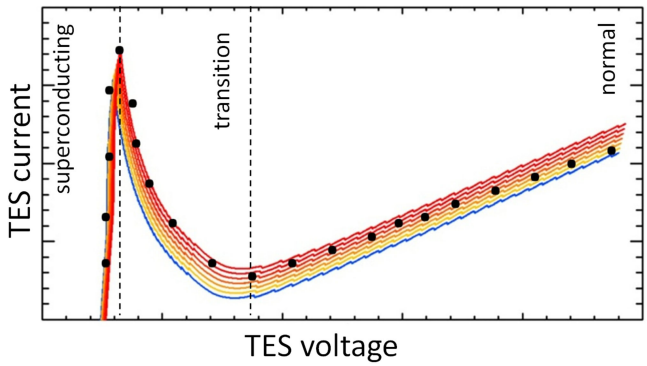

Fig. 1. IV cartoon. A typical IV consists of a set of experimental data (black dots); each data point belongs to a single isothermal IV; the entire IV covers a beam of IVs corresponding to a given temperature range.

T_c, differential thermal conductance $G = dP_{bias}/dT_{bath}$ [19], and saturation power P_{sat} (*i.e.*, the amount of electrical power required to drive the TES normal in the absence of optical loading). The IV analysis typically neglects the changes in the bath temperature due to the changing Joule power dissipated in the array of TESes as the current varies. In kilopixel focal planes, the simultaneous biasing of thousands of TESes with mAs of current heats the focal plane, resulting in IVs being acquired with varying effective bath temperatures T_{bath}, which we refer to as non-isothermal IVs. In this case, each data point of an IV can be thought of as belonging to a given isothermal IV, and the entire IV can then be considered to span a set of isothermal IVs within some temperature range (Fig. 1). Correcting for this effect is difficult because one only measures a single current and voltage value for each IV. Because each acquired IV follows the expected TES behavior (*i.e.,* superconducting, transition and normal state), the set of IVs are anticipated to span a small temperature range.

We have investigated the impact of non-isothermal IVs on the estimate of TES parameters with a toy model. Three possible heating effects are considered, where *dT* is the variation from its average value of the bath temperature estimated from a nearby thermometer:

$$\begin{aligned} &1)\ dT = 0\text{mK} \\ &2)\ dT = <T_{TESi}> \\ &3)\ dT = <T_{TESi}> + \sigma_{cal}. \end{aligned} \tag{1}$$

In the first (ideal) case no Joule heating is present from other TESes. In the second, many TESes are biased simultaneously and an increase of the bath temperature recorded by the nearby thermometer of a few milliKelvin is observed (which is the case for the AdvACT array). For this case a linear trend between *dT* and the average effective bath temperature at the colder ends of the TES thermal links ($<T_{TESi}>$) is modeled at each IV acquisition step. In the third scenario, a constant error, $\sigma_{cal} = 2$ mK, is adopted to account for offset errors in the thermometer calibration.

We estimate the TES parameters, (using eq. (2) in Section III), for two HF detectors using heating effect 1) for the T_{bath} values. The dominant systematic error arises from heating the sensor. In practice, this systematic error can be an order of magnitude larger than the statistical error from the fitting procedure. For a subset of HF TESes, we estimate the impact on the P_{sat} distribution at 100 mK. We then assume a heating effect of 5 mK, and neglect the σ_{cal} term to highlight the heating effect. The estimation of P_{sat} is affected by the temperature drift at a $\sim$3% level. Additionally, array heating correlates the sensors' data, artificially suggesting a more uniform P_{sat} distribution.

III. Estimating TES Parameters

Here we describe methods to estimate TES parameters from temperature ramp acquisitions. At each of a dozen T_{bath} values, IV curves for many hundreds of detectors (indexed by *i*) are fitted to estimate thousands of values of $P_{bias}(T_{bath}, i)$. The resulting distribution includes outliers that must be removed before estimating the TES parameters. With outlier we mean a detector property which deviates from the TES specs defined by, e.g., the TES design, such as an outlying P_{bias} value.

Rather than use the blunt method of excising all the data from one TES when it exhibits an outlying P_{bias} we build an estimator to identify outliers from the median of the TES P_{bias} values to recover TESes, which are otherwise operational.

For each MF sensor and T_{bath} value, we estimate the median P_{bias}, to identify and exclude outlier devices. We remove $P_{bias} < 0$, $P_{bias} > 20$ pW and P_{bias} values which do not follow the expected decreasing trend when T_{bath} increases. For each frequency channel, the resulting median P_{bias} versus T_{bath} trend represents the target behavior. To remove the outliers, for each TES, we compare each P_{bias} value at a given T_{bath} to $(1 \pm \gamma P_{bias})$. γ is tuned empirically to not discard operational TESes that have P_{bias} values away from the median. For $T_{bath} <$ 160 mK $\gamma = 0.4$, while for $T_{bath} >$ 160 mK $\gamma = 0.5$. For a P_{bias} value many γs away from the acceptance P_{bias} range, this method efficiently removes the outlier and does not discard the corresponding TES. Without an outlier estimator, these operational TESes would have been removed from the data analysis. Next a fit to the empirical equation,

$$P_{bias}\,(R = 0.9R_n) = K\,(T_c^n - T_{bath}^n) \tag{2}$$

estimates T_c, the constant *K* and the power index *n*. *K* and *n* depend on the thermal weak link to the bath [19]. T_{bath} is read out by a calibrated thermometer mounted on a copper support in contact to the array via copper beryllium springs. From eq. (2) we estimate *G* [19] and P_{sat}, which is defined as P_{bias} at 90% the normal resistance, R_n.

We fit eq. (2) with a custom IDL [20] script that minimizes the χ^2 function built from the measured P_{bias} and T_{bath} values. Temperature ramp acquisition is time consuming and lowers data acquisition efficiency. Repeated IV measurements show the P_{bias} estimates fluctuating at <1%, while T_{bath} is very stable. Based on these observations we adopt a 1% uncertainty estimate for P_{bias}, and no uncertainty in T_{bath}. For each TES, a matrix of $50 \times 50 \times 50 \times 90$ elements is created. The first three dimensions are populated by the *K*, *n* and T_c values generated within suitable ranges. The fourth dimension is populated by temperature values from the lowest T_{bath} to the possible highest T_c.

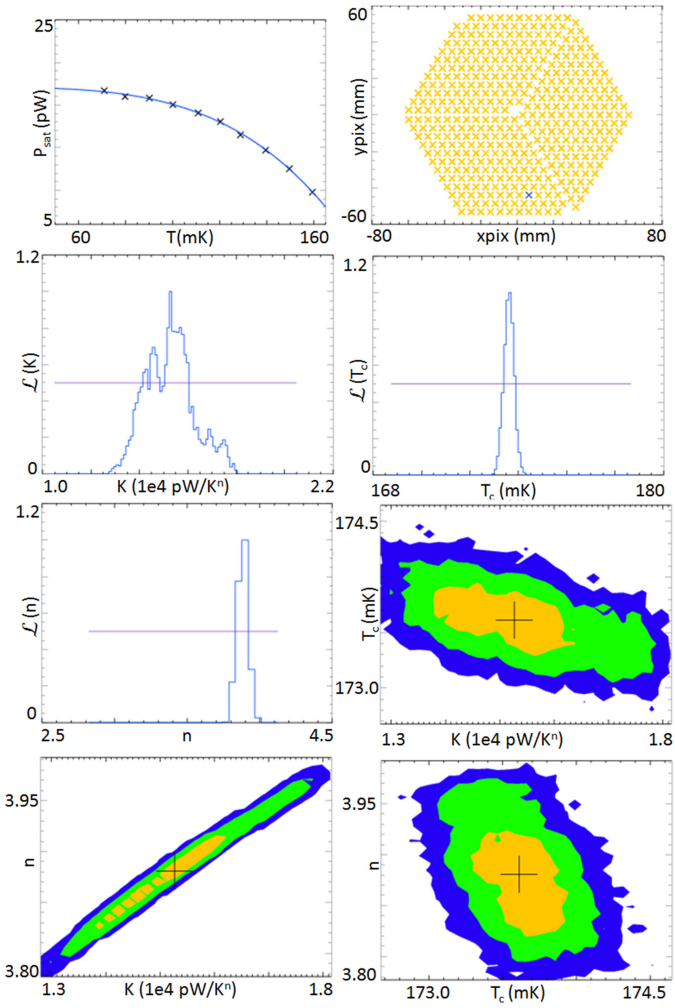

Fig. 2. T_c, n and K estimates for a typical AdvACT TES (MF2 c0r27, 90 GHz). Top left: experimental data and the MCMC fit (solid line). Top right: position on the focal plane of the detector under study. Center and bottom left: parameter likelihoods. Center right and bottom: contour plots, from inside to outside: 65, 90 and 95% confidence level regions.

We then determine the minimum of the χ^2 function from all the possible parameter combinations generated within the matrix.

The results from the χ^2 script provide the initial guess for a Metropolis-Hastings MCMC algorithm [21], [22] coded in another custom IDL script. For each TES, the script generates 10^7 points and discards the first 10^6 (*burn in*) [22]. Each point is drawn from a Gaussian distribution which is the product of Gaussian distributions of single parameters. We do not assume any prior. An advantage of this MCMC method is that it enables direct estimation of the uncertainties and correlations among the fitted parameters. A typical result from AdvACT MF2 is shown in Fig. 2. T_c is uncorrelated with n and K, while n and K are degenerate with each other. We estimate the TES parameters and errors from the mean and the standard deviation of the MCMC accepted points, respectively. Since T_c is uncorrelated with n and K, its value is well constrained by the MCMC runs. Since n and K are degenerate with each other, they have larger errors than T_c since we do not fix the n value a priori. As a result, the relative error in the estimate of T_c is $<1\%$ compared to 2–20%

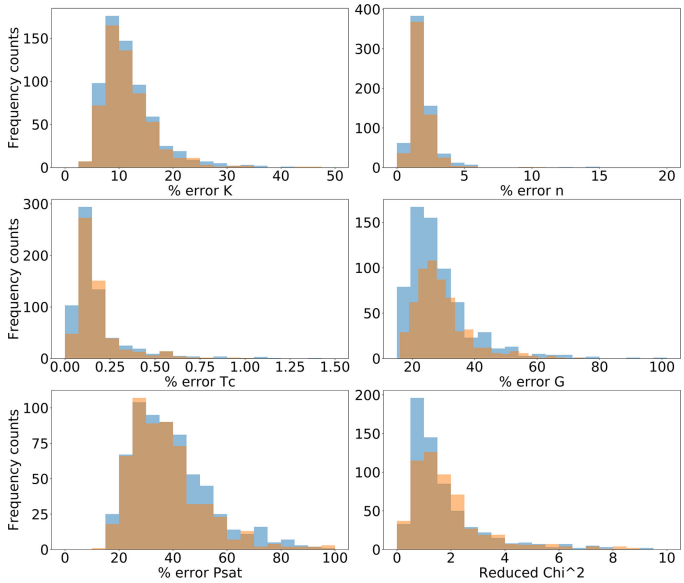

Fig. 3. Frequency counts of the estimated relative errors on the MF2 TES parameters from the MCMC algorithm [light gray (blue in online version) 90 GHz, dark gray (orange in online version) 150 GHz].

for K and n (Fig. 3). The errors are propagated in the estimates of TES properties, including P_{sat} at 100 mK, G, and the phonon NEP. These relative errors can be as large as 20–30%, being dominated by the K-n degeneracy.

IV. Machine Learning

To provide optimal bias voltages for each kilopixel detector array, it is important to first generate and then exclude a list of non-operational detectors, which we define as the "*deadlist*".

Readout tuning and detector deadlists for kilopixel arrays typically involve visually inspecting thousands of plots because an analytical approach is not practical. This operation requires a significant amount of time and is subjective. Traditionally, the detector deadlists are built at the beginning of the array qualification and are not updated unless major events occur. However, detectors flagged as non-operational in the deadlists, can actually perform well in certain conditions that vary between observations (e.g., changes in bias, bath temperature, weather, etc.). Deadlists can also slightly change between different readout tunings. Therefore, the possibility of building the deadlists and tuning the detector readout in an automated and fast (few minutes) way is appealing. It would enable dynamic deadlists that can change with weather conditions and readout configuration, which would allow for the recovery of more data.

ML algorithms have risen in popularity over the last few years [23], thanks to the diffusion of open source algorithms [24], [25], and have a variety of applications [26]–[31]. We have tested the possibility of automatically building a detector deadlist using ML algorithms to replace the need for visually inspecting thousands of plots.

With the TESes biased to be on a given point of their superconducting transition, we inject a small square wave signal (e.g., 10 μA wide with a period of 1s) or *bias step* [32] onto the TESes belonging to a given bias line. As an example, we have

built a database of bias step acquisitions for about 2,000 TESes divided into three classes: TESes showing bias steps, noise, and no bias steps.

We note that a ML algorithm is not strictly necessary in this case as an analytical approach would be more efficient, but we use it as an example of how ML algorithms could be used in two future applications that are in preparation: characterizing TES arrays and testing TES readout electronics, e.g., SQUIDs response. Each plot contains 2,500 data points, however, to keep the database dimensions manageable only a subset (e.g., 784-elements) of the array, $\boldsymbol{x}$, is employed. We choose 784 as practical number that works both for 1-dimensional and 2-dimensional ($784 = 28 \times 28$) ML problems.

The signals of the three classes are randomly sorted in the database. The signals in the dataset are randomly divided between 1500 training and 250 test signals. Using the knowledge of the correspondence of each signal to its label in the training set (*supervised learning*), we implement a ML algorithm with tensorflow [33]. We label each TES signal with a three-bit array (*one-hot encoding* [33]). For example, the true label $\boldsymbol{y}_-$ of a signal belonging to the class 0 is (1, 0, 0). In practice, an example of a correct predicted label $\boldsymbol{y}$ of a signal belonging to the class 0 could look like (0.85, 0.05, 0.1), while a wrongly predicted label could be (0.1, 0.7, 0.2). Each $\boldsymbol{y}$ can effectively be read as an array of probabilities, and is proportional to $\boldsymbol{x}$: $\boldsymbol{y} = \boldsymbol{w}\boldsymbol{x}$, where $\boldsymbol{w}$ is a weights matrix with dimensions (784, 3).

During the training phase, for each data point x_i of each signal, the ML algorithm estimates if the probability that the actual data point value increases, or decreases, the probability of belonging to a given class. The loss function, or cross-entropy, is defined as:

$$loss \propto -\sum\nolimits_{classes} y_{-}\log\left(y\right). \tag{3}$$

An optimizer estimator changes the weights w_i minimizing the difference between the true label and the predicted one, *i.e.*, estimating the gradient of the loss (*gradient descent of the loss* [33]). During the test phase, the ML algorithm works on the test signals using the knowledge on the weights learned during the training phase to predict to which class each test signal belongs to.

We find an accuracy in recognizing the test signals of 94.5% and a loss of 55%. The possible reason for the high loss value is still under investigation, but could be a result of overfitting, where the algorithm recognizes the features of the training set but is less capable of making predictions on the test sets. During the test phase, the ML algorithm estimates 93% of the signal labels correctly, which is in good agreement with what was found during the training phase. The ML algorithm runs in less than 1 minute compared to the several tens of minutes required for visually inspecting the ∼2,000 TESes bias step plots.

This positive result opens the prospect of using ML algorithms in CMB instrumentation.

V. Conclusion

In this work, we have presented methods to investigate the properties of kilopixel focal planes array. Biasing the focal plane arrays to drive them normal to measure the TES parameters can heat the focal plane, which introduces a systematic error of ∼3% in the estimate of the TES P_{sat}. This heating effect could be further reduced by tuning the current driving the TES normal. Large focal plane arrays are also affected by outliers in estimating the TES parameters from temperature ramp acquisitions. Defining an acceptance P_{bias} range built from the median P_{bias} values within an optimally defined range reliably excludes outliers, improving the estimate of TES parameters and enabling the recovery of TESes that would have otherwise been cut from the data analysis.

Additionally, estimating the TES parameters with an MCMC algorithm enables their errors and correlations/degeneracies to be evaluated. The error in estimating T_c values is few percent since T_c is uncorrelated with K and n. Given the degeneracy between K and n, their uncertainties are larger than that of T_c. This degeneracy is propagated to the estimate of P_{sat} and G, for which the relative error is ∼10–20%. The error on G affects the estimate of the phonon NEP.

Finally, we have demonstrated the possibility of tuning TESes with supervised ML with an accuracy in the recognition of the different TES behaviors of 94.5%. ML offers compelling opportunities for detector tuning and readout, enabling dynamic deadlists that can quickly change with weather conditions and readout configurations. The ML approach not only drastically reduces the time required to build a deadlist, but enables the recovery of data from detectors that are only operational in certain conditions and would otherwise be cut from all observations.

Acknowledgment

M.S. would like to thank R. Bellaveglia, G. Dujany, and G. Puglisi for useful discussions on ML, and C.G. Pappas on the paper.

References

[1] S. W. Henderson *et al.*, "Advanced ACTPol cryogenic detector arrays and readout," *J. Low Temp. Phys.*, vol. 184, pp. 772–779, 2016.
[2] R. J. Thorton *et al.*, "The Atacama Cosmology Telescope: The polarization-sensitive ACTPol instrument," *Astrophys. J. Suppl. Ser.*, vol. 227, no. 2, 2016, Art. no. 21.
[3] D. Li *et al.*, "AlMn transition edge sensors for Advanced ACTPol," *J. Low Temp. Phys.*, vol. 184, pp. 66–73, 2016.
[4] S. M. Duff *et al.*, "Advanced ACTPol multichroic polarimeter array fabrication process for 150 mm wafers," *J. Low Temp. Phys.*, vol. 184, pp. 634–641, 2016.
[5] S. P. Ho *et al.*, "Highly uniform 150 mm diameter multichroic polarimeter array deployed for CMB detection," in *Proc. SPIE*, vol. 9914, 2016, Art. no. 991418.
[6] S. K. Choi *et al.*, "Characterization of the mid-frequency arrays for Advanced ACTPol," *J. Low Temp. Phys.*, vol. 193, pp. 267–275, Jun. 2018.
[7] S. W. Henderson *et al.*, "Readout of two-kilopixel transition-edge sensor arrays for Advanced ACTPol," *Proc. SPIE*, vol. 9914, 2016, Art. no. 99141G.
[8] J. Beyer and D. Drung, "A SQUID multiplexer with superconducting-to-normalconducting switches," *Supercond. Sci. Technol.*, vol. 21, 2008, Art. no. 105022.
[9] W. B. Doriese *et al.*, "Developments in time-division multiplexing of X-ray transition-edge sensors," *J. Low Temp. Phys.*, vol. 184, no. 1–2, pp. 389–395, 2016.
[10] N. Galitzki *et al.*, "The Simons Observatory: Instrument overview," *Proc. SPIE*, vol. 10708, Jul. 2018, Art. no. 1070804, doi:10.1117/12.2312985.
[11] Simons Observatory website. 2016. [Online]. Available: https://simonsobservatory.org/
[12] The CMB-S4 collaboration, *CMB-S4 Technology Book*, 1st ed., 2017.

[13] CMB-S4 website. 2016. [Online]. Available: https://cmb-s4.org/
[14] M. A. Dobbs *et al.*, "Frequency multiplexed superconducting quantum interference device readout of large bolometer arrays for Cosmic Microwave Background measurements," *Rev. Sci. Instrum.*, vol. 83, no. 7, 2012, Art. no. 073113.
[15] A. Bender *et al.*, "Digital frequency domain multiplexing readout electronics for the next generation of millimeter telescopes," *Proc. SPIE*, vol. 9153, 2014, Art. no. 91531A.
[16] B. Dober *et al.*, "Microwave SQUID multiplexer demonstration for Cosmic Microwave Background imagers," *Appl. Phys. Lett.*, vol. 111, 2017, Art. no. 243510.
[17] S. W. Henderson *et al.*, "Highly-multiplexed microwave SQUID readout using the SLAC microresonator radio frequency (SMuRF) electronics for future CMB and sub-millimeter surveys," *Proc. SPIE*, vol. 10708, 2018, Art. no. 1070819.
[18] E. S. Battistelli *et al.*, "Automated SQUID tuning procedure for kilopixel arrays of TES bolometers on ACT," *Proc. SPIE*, vol. 7020, 2008, Art. no. 702028.
[19] K. D. Irwin and G. C. Hilton, "Transition-edge sensors," in *Cryogenic Particle Detection* (Topics in Applied Physics series 99). New York, NY, USA: Springer, 2005, pp. 63–150.
[20] Interactive Data Language (IDL) Software. 1977. [Online]. Available: https://www.harrisgeospatial.com/Software-Technology/IDL
[21] C. Andrieu *et al.*, *An Introduction to MCMC for Machine Learning*, vol. 50. Norwell, MA, USA: Kluwer, 2003, pp. 5–43.
[22] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[23] E. Mjolsness and D. De Coste, "Machine learning for science: State of the art and future prospects," *Science*, vol. 293, no. 5537, pp. 2051–2055, 2001.
[24] Keras website. 2015. [Online]. Available: https://keras.io/
[25] Pytorch website. 2016. [Online]. Available: https://pytorch.org/
[26] W. Fendt and B. D. Wandelt, "Pico: Parameters for the impatient cosmologist," *Astrophys. J.*, vol. 654, no. 1, pp. 2–11, 2007.
[27] C. P. Novaes *et al.*, "A neural-network based estimator to search for primordial non-gaussianity in Planck CMB maps," *J. Cosmol. Astropart. Phys.*, vol. 27, no. 09, 2015, Art. no. 064.
[28] S. Whiteson and D. Whiteson, "Machine learning for event selection in high energy physics," *Eng. Appl. Artif. Intell.*, vol. 8, pp. 1203–1217, 2009.
[29] P. Baldi, P. Sadowski, and D. Whiteson, "Searching for exotic particles in high-energy physics with deep learning," *Nature Commun.*,vol. 5, 2014, Art. no. 4308.
[30] A. Radovic *et al.*, "Machine learning at the energy and intensity frontiers of particle physics," *Nature*, vol. 560, pp. 41–48, 2018.
[31] R. Dodkins *et al.*, "MKID digital readout tuning with deep learning," *Astron. Comput.*, vol. 23, pp. 60–71, 2018.
[32] E. Grace, "Detector characterization, optimization, and operation for ACT-Pol," Ph.D. dissertation, Phys. Dept. Princeton Univ., Princeton, NJ, USA, 2016.
[33] Tensorflow website. 2015. [Online]. Available: https://www.tensorflow.org/