# A Fast Machine Learning Based Algorithm for MKID Readout Power Tuning

Rupert H. Dodkins[1,*], Kieran O'Brien[1], Niranjan Thatte[1], Sumedh Mahashabde[1], Neelay Fruitwala[2], Seth Meeker[2], Alex Walter[2], Paul Szypryt[2] and Ben Mazin[2]
[1]*Department of Astronomy, University of Oxford, Oxford, OX1 3RH, UK*
[2]*Department of Physics, University California, Santa Barbara, California 93106, USA*
*Contact: rupert.dodkins@physics.ox.ac.uk

***Abstract***— **As high pixel count Microwave Kinetic Inductance Detector (MKID) arrays become widely adopted, there is a growing demand for automated device readout calibration. These calibrations include ascertaining the optimal driving power for best pixel sensitivity, which, because of large variations in MKID behavior, is typically performed by manual inspection. This process takes roughly 1 hour per 1000 MKIDs, making the manual characterization of ten-kilopixel scale arrays unfeasible. We propose the concept of using a machine-learning algorithm, based on a convolution neural network (CNN) architecture, which should reliably tune ten-kilopixel scale MKID arrays on the order of several minutes.**

## INTRODUCTION

Microwave Kinetic Inductance Detectors, or MKIDs, are superconducting detectors that sense photons by measuring the change in ac surface impedance produced by the separation of Cooper pairs [1]. Each MKID device is a superconducting thin film lithographically patterned into an array of high quality factor microwave resonators. Through passive frequency domain multiplexing, arbitrarily large arrays can be fabricated and thousands of pixels can be read out per feed-line using room temperature electronics. The current generation of MKID devices are kilopixel arrays [2] and in the UVOIR (UV, optical and IR) regime, the current generation are many tens of thousands of pixels [3].

While fabrication of large arrays is comparatively straightforward, maintaining consistency across each array is challenging. Ideally a wide frequency sweep through the device should yield transmission dips at equal spacing with uniform, high quality factors. In reality there is a distribution in the separation between resonances and the quality factors, and therefore the power handling ability. Due to this variety in behavior, the probe signal power allocations have been, until now, mostly decided by manual human inspection.

## MKID DIGITAL READOUT

For readouts in both the THz [4] and UVOIR regime [5], a software defined radio transceiver is used to generate and monitor a comb of microwave frequencies to drive each of the resonators. Since the quality factors of the resonators are high,

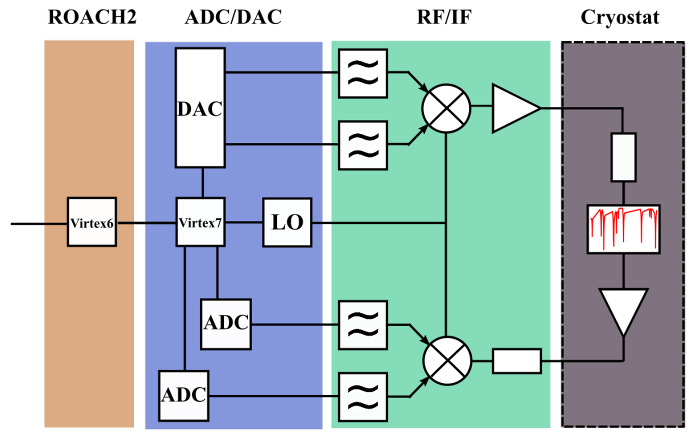


Fig. 1: A block diagram showing the general functions of the hardware that make up a single feed-line of the second generation UVOIR Software Defined Radio digital readout [5]. The CASPER ROACH2 boards [6] contain firmware, that produces the baseband frequencies to drive the resonators; as well as channelizes the incoming signal, performs optimal filtering, triggering, and creates photon packets to send to the data acquisition machine.

the carrier frequency for a particular pixel does not interact with other pixels while propagating through the array. The complexity of device readout is transferred to the digital backend, which demultiplexes the sum of probe tones using custom firmware. The number of resonators a feed-line can probe is limited by bandwidth of the analogue-to-digital converter (ADC); therefore, many feed-line readout systems operate in parallel, each probing a fraction of the array. Fig. 1 shows the readout architecture of a single feed-line for the UVOIR second-generation readout [5].

In order to detect the phase (or amplitude) modulations of each resonator during observations, the readout uses I/Q carrier signals, where $I$ and $Q$ represent the magnitude of the real and imaginary signal components respectively. $I$ is a sine wave at the resonator frequency and $Q$ is the same waveform offset in phase by one-quarter cycle. The phase between the readout tone and the resonant frequency is calculated using the equation $\Phi = \arctan(Q/I)$. During observations, the readout electronics constantly monitor the phase of each pixel and if that phase crosses a threshold then a photon packet is created and sent to the data acquisition machine.

### DRIVE SIGNAL POWER TUNING

Readout tone powers need to be evaluated and programmed into the readout in order to probe the resonators under optical illumination. The cryogenic amplifier (usually a HEMT amplifier) noise and Two-Level System Noise [7] place limits on the sensitivity of the device. One technique to reduce these contributions is to drive each resonator with as much power as possible. However, in the high-power regime resonators exhibit a nonlinear response. If the power is sufficiently high, the resonator can occupy two stable states, and the resonator undergoes bifurcation. This has been attributed to an inherent nonlinearity in the kinetic inductance [8] and readout power heating [9]. The transmission profile distorts away from the Lorentzian profile (Fig. 2), which can be detrimental to the phase measurements and therefore sensitivity. The goal in readout tuning is therefore to identify the power for the onset of bifurcation and step down a few dB – which is essentially a classification problem.

De Visser et al. [10] have shown that, in the THz regime, even though the readout photons are below the Cooper pair binding energy, the readout photons will heat the quasiparticle population, which leads to an increase in generation-recombination noise. This study assumes that optimum sensitivity is achieved by driving the resonators at the highest power prior to bifurcation, and describes an optimization to that method. In theory, the algorithm could be optimized to any criteria so long as the training data exists.

Presently, the optimal power for each resonator is evaluated through manual visual inspection. The user will study each resonator, observing the resonance loops, the derivative of the transmission spectra with respect to frequency, termed $v_{IQ}$, and the relative change of these parameters at increasing power. With this information at hand they can ascertain if, and why, a resonator is showing non-ideal behavior, and make an informed estimate of the optimal power.

For the second generation UVOIR readout [5], the input data is created by sweeping the digital readout frequency across the resonators and then stepping the power of the digital readout in 1 dB intervals. The power limits are chosen such that each resonator is sampled in the low power regime and after it has transitioned into the bifurcation regime. This creates a datacube for each resonator consisting of $I$ and $Q$ magnitudes at each frequency and power. Hierarchical Data Format (HDF5) [11] is used to store each of these datacubes for every pixel on a feed-line.

Fig. 2 shows two example resonator datacubes. The first resonator in Fig. 2 (a) shows a resonator with near-ideal transmission spectra and power handling ability. The resonator clearly bifurcates at the power proceeding lime-green (power index 12 from the lowest power).

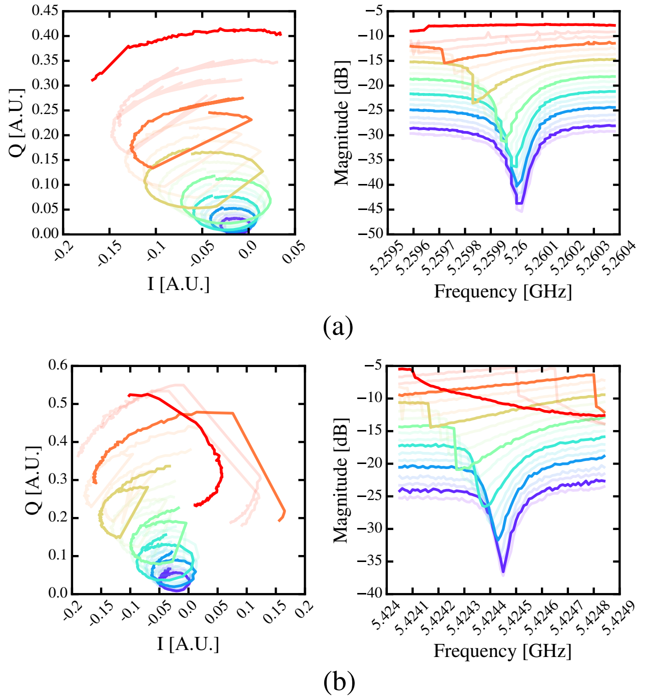


Fig. 2: (a) The resonance loop and transmission spectrum of an ideal resonator sampled at a range of powers. (b) The resonance loop and transmission spectrum of two near-colliding resonators sampled at a range of powers. The red and blue trends are sampled at the highest and lowest powers respectively; high saturation colors are separated by 3dB. The units of $I$ and $Q$ are uncalibrated raw data produced by the Analog to Digital converter in the digital readout.

The second resonator, Fig. 2b, shows a resonator datacube where the sampling window becomes contaminated at high powers with an adjacent resonator from higher frequencies and the initial resonator simultaneously translates out of the sampling window at low frequencies. Depending on the sampling window, it will sometimes appear as though an unbifurcated resonator disappears and a bifurcated resonator reappears, or vice versa. Sometimes, two resonator profiles will merge, and it is non-trivial to analytically fit the merger. Other times a well-separated resonator will show non-ideal power-handling behavior, for example the resonator in Fig. 2a shows multiple discontinuities at the penultimate power sample. Or simply, a resonator could have low signal to noise ratio.

When preparing an MKID device for observations it is often preferable to include as many resonators as possible, to ensure sufficient pixel count. However, there are many ways resonators can show non-ideal behavior and it is challenging to tune analytical algorithms to accurately account for each type of behavior.

### DEEP LEARNING CLASSIFICATION

Deep learning is a machine learning method used to discover and model high level abstractions in data [12], and the

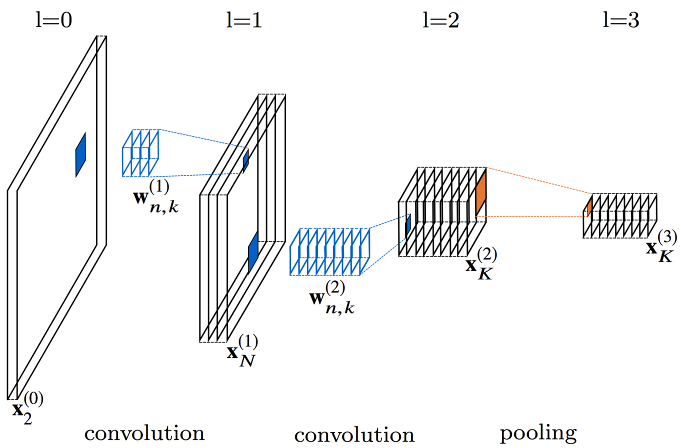


Fig. 3: The first four layers of a generic convolution neural network. The $n$ and $k$ labels for the feature maps are displayed in the reference frame of layer 2. For clarity, the $n$ dimension of the weight tensors are not displayed, nor are any of the bias vectors. The input image has two spatial dimensions (frequency and power) and an effective depth of two (amplitudes of $I$ and $Q$). The filled squares in each layer represent a projection to all depths. For a given weight, the filled square of the previous feature map displays the receptive field and, on the current feature map, the destination of the weighted sums along the $n$ dimension. Between layers 2 & 3 is a 1/4 subsampling or pooling

objective in classification is simply to create some model which can take an input vector and assign the correct label to it. In deep learning this is accomplished with an interconnected network of multiple layers each consisting of many transfer functions termed neurons. Each neuron takes the input from the previous layer, transforms it according to a learnable parameter and then passes the output through a nonlinear activation function. During training, these parameters are tuned in successive iterations such that the outputs of the final layer converge on the labels of the input images for that batch. For resonator power tuning, the target label would be the optimal power index, or in this study, the highest power prior to bifurcation.

The more layers in the neural network, the more complex the features the model can discern. However, a excessive number of parameters in the neural network compared to the amount of training data, increases the chances of overfitting, meaning the model does not generalize well to unseen input data. Image recognition algorithms typically use millions of training images. For resonator power tuning, using manual inspection data as the reference training set, this magnitude of training data will not be available. To maintain the model sophistication in this instance, more training data could be created through label preserving transformations, or regularization techniques such as dropout [13], early stopping and batch normalization [14] could be utilized.

In a fully connected neural network (FCNN), each neuron computes a weighted sum of all the neurons in the layer that precedes it. These networks suffer from the 'curse of dimensionality', making them computationally expensive and, in some cases, easy to overfit. The goal of this study is to create a fast and rigorous algorithm. Given the limited scope of the training data, an alternative neural network architecture should be more effective.

Convolution neural networks (CNN) conversely are locally connected networks. Smaller weight matrices are convolved across the preceding layer, probing smaller receptive fields and exploiting spatially local correlations. To increase the number of free parameters, a four-dimensional weight tensor is convolved across the previous layer producing a tensor, $\boldsymbol{x}$, with depth $K$ for an input depth $N$.

$$\boldsymbol{x}_k^{(l)} = a\left(\sum_n^N \boldsymbol{w}_{n,k}^{(l)} * \boldsymbol{x}_n^{(l-1)} + \boldsymbol{b}_k^{(l)}\right)$$

where $n = 0,1 \dots N$ and $k = 0,1 \dots K$. The weights, $\boldsymbol{w}^{(l)}$, act as filters extracting two-dimensional features from each layer in the $n$ dimension. The biases $\boldsymbol{b}^{(l)}$ are also learnable parameters. The area of each successive layer in a CNN architecture decreases (assuming no padding) which means the learnable parameters of the successive layers extract successively larger scale features.

For resonator power tuning, the power-sweep datacubes are three-dimensional, and the spatial information between all the dimensions could be best exploited with a CNN architecture. Fig. 3 shows an arbitrary implementation of the first few layers of a CNN-based algorithm for resonator power tuning. The benefits of both FCNN and CNN architectures could be exploited by using several convolution and pooling layers, followed by fully connected layers.

## Conclusions

It has been shown that a simple machine learning algorithm is an interesting new method for characterizing resonator arrays for readout. The advantages are that it can potentially characterize an array of many thousands of resonators in just a few seconds – compared to manual inspection, which takes many human hours. This advancement is vitally important as more kilopixel resonator arrays come online and as $10^4$ pixel scale array are developed. Dodkins et al. (2017) [15] details the implementation of a CNN algorithm for resonator classification and provides a comparison with conventional analytical algorithms in order to quantify the accuracy.

## Acknowledgment

The authors would like to thank the Science Technology Facilities Council for supporting this research through their studentship and Long Term Attachment programs.

## References

[1] Day, P. K., LeDuc, H. G., Mazin, B. A., Vayonakis, A., & Zmuidzinas, J. "A broadband superconducting detector suitable for use in large arrays," *Nature*, vol. *425*, pp. 817-821, 2003.

[2] Catalano, A., Adam, R., Ade, P., André, P., Aussel, H., Beelen, A., ... & Calvo, M. "The NIKA2 commissioning campaign: performance and first results". arXiv preprint arXiv:1605.08628. 2016.

[3] Meeker, S., Mazin, B., Jensen-Clem, R., Walter, A., Szypryt, P., Strader, M., & Bockstiegel, C. "Design and development status of mkid integral

field spectrographs for high contrast imaging," *in Proc. Adaptive Optics for Extremely Large Telescopes 4–Conference Proceedings* 2015 paper 1.1

[4] van Rantwijk, J., Grim, M., van Loon, D., Yates, S., Baryshev, A., & Baselmans, J. "Multiplexed readout for 1000-pixel arrays of microwave kinetic inductance detectors," *IEEE Transactions on Microwave Theory and Techniques*, vol. 64, pp. 1876-1883, 2016.

[5] Strader, M. J. "*Digitial readout for microwave kinetic inductance detectors and applications in high time resolution astronomy,"* (Doctoral dissertation, University of California, Santa Barbara). 2016.

[6] CASPER Group Collaboration of Astrophysical Signal Processing and Electronics Research. [online] Available: https://casper.berkeley.edu/ 2009.

[7] Gao, J. "*The physics of superconducting microwave resonators"*. California Institute of Technology. 2008.

[8] Swenson, L. J., Day, P. K., Eom, B. H., Leduc, H. G., Llombart, N., McKenney, C. M., ... & Zmuidzinas, J. "Operation of a titanium nitride superconducting microresonator detector in the nonlinear regime," *Journal of Applied Physics*, vol. 113, pp. 104501, 2013

[9] Thomas, C. N., Withington, S., & Goldie, D. J. "Electrothermal model of kinetic inductance detectors," *Superconductor Science and Technology*, vol. *28*, pp. 045012. 2015.

[10] De Visser, P. J., Withington, S., & Goldie, D. J. "Readout-power heating and hysteretic switching between thermal quasiparticle states in kinetic inductance detectors," *Journal of Applied Physics*, vol. *108*, 114504. 2010.

[11] The HDF Group Hierarchical data format version 5. [online] Available: http://www.hdfgroup.org/HDF5 2000.

[12] Bengio, Y., & LeCun, Y. "Scaling learning algorithms towards AI," *Large-scale kernel machines*, vol. *34*, pp. 1-41 2007.

[13] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929-1958, 2014.

[14] Ioffe, S., & Szegedy, C. "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *International Conference on Machine Learning* pp. 448-456 Jun 2015

[15] Dodkins, R., Fruitwala N., O'Brien K., Thatte N., Mahashabde S., Meeker S., Walter A., Szypryt P., Mazin B., in prep. "MKID Digital Readout Tuning with Deep Learning". Monthly Notices of the Royal Astronomical Society 2017.