Reconstructing Patchy Reionization with Deep Learning

Eric Guzman[1] and Joel Meyers[1]

[1]*Department of Physics, Southern Methodist University, 3215 Daniel Ave, Dallas, TX 75275, USA*
(Dated: August 23, 2021)

The precision anticipated from next-generation cosmic microwave background (CMB) surveys will create opportunities for characteristically new insights into cosmology. Secondary anisotropies of the CMB will have an increased importance in forthcoming surveys, due both to the cosmological information they encode and the role they play in obscuring our view of the primary fluctuations. Quadratic estimators have become the standard tools for reconstructing the fields that distort the primary CMB and produce secondary anisotropies. While successful for lensing reconstruction with current data, quadratic estimators will be sub-optimal for the reconstruction of lensing and other effects at the expected sensitivity of the upcoming CMB surveys. In this paper we describe a convolutional neural network, ResUNet-CMB, that is capable of the simultaneous reconstruction of two sources of secondary CMB anisotropies, gravitational lensing and patchy reionization. We show that the ResUNet-CMB network significantly outperforms the quadratic estimator at low noise levels and is not subject to the lensing-induced bias on the patchy reionization reconstruction that would be present with a straightforward application of the quadratic estimator.

I. INTRODUCTION

Observations of the cosmic microwave background (CMB) have provided a wealth of cosmological information. The coming generation of CMB surveys, including those from Simons Observatory [1], CCAT-prime [2], CMB-S4 [3], PICO [4], and CMB-HD [5], will map the CMB with unprecedented precision. In addition to providing greater sensitivity to primary CMB intensity and polarization fluctuations, next generation observations will allow for much more precise measurements of secondary CMB anisotropies [6].

Secondary anisotropies of the CMB are generated by interaction of CMB photons with structure which intervenes between our telescopes and the surface of last scattering. Prominent sources of secondary anisotropies include weak gravitational lensing by large-scale structure (see [7] for a review); the various manifestations of the Sunyaev-Zel'dovich effect [8–12], which describe Compton scattering of CMB photons with free electrons in galaxy clusters and the intergalactic medium; the integrated Sachs-Wolfe [13] and Rees-Sciama effects [14], characterizing how time-dependent gravitational potentials alter the energy of CMB photons; and the moving lens effect [15], where the transverse motion of massive objects imprints temperature fluctuations on the CMB. The sensitivity of upcoming experiments should allow some of these effects to be detected for the first time in the coming years [16–24]. These secondary anisotropies act both as a source of cosmological information and as a source of confusion for the primary anisotropies.

Sources of secondary anisotropies can be reconstructed due to the changes they imprint on the statistics of the primary CMB fluctuations. In particular, the primary CMB anisotropies are expected to be well described by nearly Gaussian fluctuations that can be fully characterized by their power spectra C_ℓ. Sources of secondary anisotropies break rotational invariance, induce off-diagonal correlations between the spherical harmonic coefficients, and cause the observed statistics to be non-stationary. This change to the statistics can be observed in the data, and thereby can be used to reconstruct the field causing the distortion [25, 26].

The standard technique for reconstructing distortions of the CMB fluctuations utilizes a quadratic estimator, constructed from a weighted product of two factors of CMB fluctuations [25]. The quadratic estimator has been used to great success to detect the effects of gravitational lensing at high significance with existing CMB surveys [27–29]. While the quadratic estimator works well with current data, it will be sub-optimal at the sensitivity anticipated in the next generation of CMB surveys. One reason for this is that higher order effects of the sources of secondary anisotropies become too important to ignore in very high fidelity CMB maps, and use of the quadratic estimator results in an estimate of the distortion field whose variance is limited by secondary anistropies rather than by instrumental noise. This challenge can be overcome by using maximum likelihood estimators [30, 31] or by iteratively removing the effects of the secondaries [32]. Another limiting factor in the use of the quadratic estimator is that the presence of more than one distortion field can lead to biases in the reconstructed fields. This bias can be avoided at the cost of increased variance with the use of 'bias-hardened' estimators [33].

One of the main motivations for an accurate reconstruction of the CMB lensing field is related to the search for primordial gravitational waves. CMB polarization can be split into curl-free E modes and divergence-free B modes, the latter of which are generated by primordial gravitational waves, but not by density fluctuations at linear order [34, 35]. Gravitational lensing of the CMB can convert E-mode polarization into B-mode polarization, and these lensing-induced B modes act as a source of confusion in searches for primordial gravitational waves [36, 37]. With an estimate of the lensing field, it is possible to delens the CMB, thereby removing some of the lensing B modes enabling greater sensitivity

to primordial gravitational waves [38–40].

Delensing the CMB is useful beyond the benefits it provides for primordial gravitational wave searches [41]. For example, delensing the temperature and E-mode polarization leads to sharper peaks in the power spectra, allowing for a better determination of parameters which impact peak positions [41, 42]. Delensing can also reduce the off-diagonal covariance induced by lensing [41] and thereby improve the sensitivity of the CMB to other effects, such as primordial non-Gaussianity [43].

In order to achieve the goals of next generation CMB surveys, delensing will be necessary (see e.g. [3, 44, 45]). However, achieving the required delensing performance is not without challenges. Delensing algorithms that go beyond the quadratic estimator are under development, but the implementations can be computationally intensive [31, 46–48]. In the near-term, delensing the CMB using external tracers of the lensing field holds promise [49], though at improved sensitivity, internal delensing, whereby the CMB is delensed using a map of the lensing field reconstructed from the observed CMB map, will be necessary to achieve the best delensing performance [32]. Internal delensing can lead to biases in the delensed CMB map, though these biases can be mitigated without much loss in sensitivity by splitting the data appropriately [50].

Reionization, the process by which the neutral gas filling the Universe became ionized by the first stars and galaxies between redshifts of about 20 and 6, was a complex and non-uniform process [51–59]. The spatial variations or patchiness in the process of reionization leads to effects on the observed CMB temperature and polarization [60–71]. Patchy reionization can produce B-mode polarization, either by the modulation of E-mode polarization or by scattering of remote temperature quadrupoles, and these secondary B modes can potentially act as a source of confusion in searches for primordial gravitational waves [72].

Quadratic estimators have been designed to reconstruct the spatial variations in optical depth that result from patchy reionization [65, 66]. However, when applied to maps of the lensed CMB, the patchy reionization quadratic estimator will be biased by the effects of lensing, since both effects lead to mode-coupling and the estimators are not independent [73]. In principle, the lensing estimator is also biased by the presence of patchy reionization, though the bias to the lensing estimate is expected to be small because the effects of patchy reionization are much smaller than those of lensing. These biases can be mitigated with delensing, and they can be essentially avoided with bias-hardening at the cost of a small increase to the variance of the estimator [73]. So far, only upper limits on the effects of patchy reionization exist [74, 75], though the effects may be detectable in future data [76].

In this paper, we apply the techniques of machine learning to reconstruct the effects of patchy reionization in simulated CMB data. It has been shown recently that machine learning is capable of producing promising results for lensing reconstruction and delensing at noise levels where the quadratic estimator is sub-optimal [77]. The focus of this work is to extend the techniques of Ref. [77] to the simultaneous reconstruction of CMB lensing and patchy reionization. As described above, reconstructing patchy reionization in the presence of lensing is challenging due to the extra variance coming from lensing-induced B-mode polarization and by the bias that can result from the sensitivity of the patchy reionization estimator to the effects of lensing. A main goal of this work is to determine the extent to which machine learning is capable of surmounting these challenges.

II. QUADRATIC ESTIMATOR

In this section we will review the construction of the quadratic estimator for CMB lensing [25] and for patchy reionization [65]. The quadratic estimator has become the standard technique for lensing reconstruction with current data, though it will be sub-optimal with the high precision data expected from future CMB surveys [31]. We will also discuss how the estimator for the spatial variation of the optical depth is biased in the presence of lensing [73].

A. Weak Gravitational Lensing

Cosmological structure which intervenes between our telescopes and the CMB surface of last scattering generates gravitational potentials which deflect CMB photons [7]. In real space this effect can be understood as a re-mapping of the unlensed CMB by a direction-dependent deflection angle,

$$T^{\text{len}}(\hat{\mathbf{n}}) = T^{\text{prim}}(\hat{\mathbf{n}} + \nabla\phi(\hat{\mathbf{n}})),$$
$$(Q^{\text{len}} \pm iU^{\text{len}})(\hat{\mathbf{n}}) = (Q^{\text{prim}} \pm iU^{\text{prim}})(\hat{\mathbf{n}} + \nabla\phi(\hat{\mathbf{n}})), \quad (1)$$

where $(T^{\text{prim}}, Q^{\text{prim}}, U^{\text{prim}})$ are the primordial CMB fields, ϕ is the lensing potential, $\hat{\mathbf{n}}$ is the line-of-sight direction, and the superscript 'len' refers to the lensed field [25]. In harmonic space, lensing of the CMB induces mode-coupling, or correlation, between fluctuations of different wavenumbers $\ell_1 \neq \ell_2$ that is proportional to the lensing potential ϕ. In the flat-sky approximation [25],

$$\left\langle X^{\text{len}}(\boldsymbol{\ell}_1)Y^{\text{len}}(\boldsymbol{\ell}_2)\right\rangle_{\text{CMB}} \propto f_{XY}^{\phi}(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2)\phi(\boldsymbol{\ell}), \quad (2)$$

where $\boldsymbol{\ell} = \boldsymbol{\ell}_1 + \boldsymbol{\ell}_2$ and $(X,Y) = (T,E,B)$. In this paper we focus on the EB mode-coupling since the corresponding estimator has the lowest variance for the noise levels we consider. For this case, the factor $f_{XY}^{\phi}(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2)$ in Eq. (2) is defined as

$$f_{EB}^{\phi}(\boldsymbol{\ell}_1,\boldsymbol{\ell}_2) = \left[(\boldsymbol{\ell}\cdot\boldsymbol{\ell}_1)C_{\ell_1}^{EE} - (\boldsymbol{\ell}\cdot\boldsymbol{\ell}_2)C_{\ell_2}^{BB}\right]$$
$$\times \sin 2(\varphi(\boldsymbol{\ell}_1) - \varphi(\boldsymbol{\ell}_2)). \quad (3)$$

The quadratic estimator provides a way of estimating any field leading to mode coupling by using a weighted product of pairs of observed CMB maps. The unbiased minimum-variance EB quadratic estimator for the CMB gravitational lensing potential is given by [25]

$$\hat{\phi}_{EB}(\boldsymbol{\ell}) = N_{EB}^{\phi}(\boldsymbol{\ell}) \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} E^{\mathrm{obs}}(\boldsymbol{\ell}_1) B^{\mathrm{obs}}(\boldsymbol{\ell}_2) F_{EB}^{\phi}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2), \tag{4}$$

with

$$F_{EB}^{\phi}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) = \frac{f_{EB}^{\phi}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)}{C_{\ell_1}^{EE,\mathrm{obs}} C_{\ell_2}^{BB,\mathrm{obs}}} . \tag{5}$$

The factor in front of Eq. (4) is an overall normalization factor defined as

$$N_{EB}^{\phi}(\boldsymbol{\ell}) = \left[\int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} f_{EB}^{\phi}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) F_{EB}^{\phi}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) \right]^{-1} , \tag{6}$$

which also gives the variance of the estimator. We define each observed field as the lensed field plus noise,

$$X^{\mathrm{obs}}(\boldsymbol{\ell}) = X^{\mathrm{len}}(\boldsymbol{\ell}) + N^X(\boldsymbol{\ell}), \tag{7}$$

where $X = (T, E, B)$ and $N^X(\boldsymbol{\ell})$ is the noise corresponding to field X, which we take to be a Gaussian random field whose power spectrum is given by

$$C_{\ell}^{TT,\mathrm{noise}} = \Delta_T^2 e^{\ell^2 \theta_{\mathrm{FWHM}}^2/(8\ln 2)} ,$$
$$C_{\ell}^{EE,\mathrm{noise}} = C_{\ell}^{BB,\mathrm{noise}} = \Delta_P^2 e^{\ell^2 \theta_{\mathrm{FWHM}}^2/(8\ln 2)} . \tag{8}$$

At the noise levels of future experiments, the lensing quadratic estimator is sub-optimal, and maximum likelihood techniques or an iterative delensing procedure is necessary to produce the best estimate of the lensing field [31, 32]. The basic idea of the iterative technique is to first apply the quadratic estimator on the observed maps, then use the resulting estimate of the lensing field to delens, next apply the quadratic estimator on the delensed maps, delens again with the resulting estimate of the residual lensing field, and iterate this procedure until convergence is achieved [32]. Simulations have shown that the iterative technique closely matches the performance of the maximum likelihood estimator [31, 32].

B. Patchy Reionization

The inhomogeneous nature of the reionization epoch causes several observable impacts on the CMB [60–69]. The patchiness of reionization leads to an anisotropic optical depth whose effects on the CMB can be separated into three categories: screening, scattering, and the kinetic Sunyaev-Zel'dovich effect.

First, the variation of the optical depth on the sky, $\tau(\hat{\mathbf{n}})$, leads to a spatially dependent screening of CMB fluctuations due to the scattering of CMB photons into and out of our line of sight. Next, Thomson scattering of remote temperature quadrupoles on the free electrons in ionized bubbles generates new polarization fluctuations. Finally, the radial velocity of ionized bubbles generates CMB temperature fluctuations through the kinetic Sunyaev-Zel'dovich effect.

For the purpose of this paper, we focus solely on the screening effect of patchy reionization, since we are interested in the effect that can be reconstructed from observations of CMB polarization alone. The scattering effect results in new polarization that is not correlated with the polarization generated at the surface of last scattering, and the kinetic Sunyaev-Zel'dovich effect only produces temperature fluctuations.

The screening effect modulates the amplitude of the temperature and polarization anisotropies, scaling them by a factor $e^{-\tau(\hat{\mathbf{n}})}$. For the purpose of the simulations described below, we first apply modulation followed by lensing giving maps in real space according to

$$T^{\mathrm{mod}}(\hat{\mathbf{n}}) = (T^{\mathrm{prim}}(\hat{\mathbf{n}}) e^{-\tau(\hat{\mathbf{n}})})^{\mathrm{len}}$$
$$(Q^{\mathrm{mod}} \pm i U^{\mathrm{mod}})(\hat{\mathbf{n}}) = ((Q^{\mathrm{prim}} \pm i U^{\mathrm{prim}})(\hat{\mathbf{n}}) e^{-\tau(\hat{\mathbf{n}})})^{\mathrm{len}}, \tag{9}$$

such that the superscript 'mod' refers to maps that have been first modulated by patchy reionization then lensed. This is an idealized treatment of the two effects, which in reality cannot be so cleanly separated in time. We further assume for our simulations that the lensing and patchy reionization fields are uncorrelated, though in reality they should exhibit some correlation, since they develop from the same underlying density fluctuations.

Following the formalism for the lensing quadratic estimator, an unbiased minimum variance EB estimator can be derived for $\tau(\hat{\mathbf{n}})$ (see [65, 73] for the full derivation) which takes the form

$$\hat{\tau}_{EB}(\boldsymbol{\ell}) = N_{EB}^{\tau}(\boldsymbol{\ell}) \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} E^{\mathrm{obs}}(\boldsymbol{\ell}_1) B^{\mathrm{obs}}(\boldsymbol{\ell}_2) F_{EB}^{\tau}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2), \tag{10}$$

where like Eq. (6), the factor

$$N_{EB}^{\tau}(\boldsymbol{\ell}) = \left[\int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} f_{EB}^{\tau}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) F_{EB}^{\tau}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) \right]^{-1} , \tag{11}$$

is a normalization that is equal to the variance of the estimator, with the mode coupling caused by patchy screening given by

$$f_{EB}^{\tau}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) = \left[C_{\ell_1}^{EE} - C_{\ell_2}^{BB} \right] \sin 2(\varphi(\boldsymbol{\ell}_1) - \varphi(\boldsymbol{\ell}_2)), \tag{12}$$

and the filter is

$$F_{EB}^{\tau}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) = \frac{f_{EB}^{\tau}(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2)}{C_{\ell_1}^{EE,\mathrm{obs}} C_{\ell_2}^{BB,\mathrm{obs}}} . \tag{13}$$

In Eq. (10) and throughout the rest of the paper, the $E^{\mathrm{obs}}, B^{\mathrm{obs}}$ are defined according to

$$X^{\mathrm{obs}}(\boldsymbol{\ell}) = X^{\mathrm{mod}}(\boldsymbol{\ell}) + N^X(\boldsymbol{\ell}), \tag{14}$$

with X and $N^X(\boldsymbol{\ell})$ as described below Eq. (7).

In the presence of lensing, the τ estimator is biased by a spurious signal that results from the mode-coupling induced by lensing. The amplitude of the bias is more than an order of magnitude greater than the true patchy reionization signal, and the bias remains significant even after 98% of the lensing signal has been removed [73]. The bias for the EB estimator can be calculated by how much of the lensing mode-coupling is picked up by the patchy reionization estimator [73]

$$\beta_{EB}(\boldsymbol{\ell}) = N_{EB}^\tau(\boldsymbol{\ell}) \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} f_{EB}^\phi(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) F_{EB}^\tau(\boldsymbol{\ell}_1, \boldsymbol{\ell}_2) \phi(\boldsymbol{\ell}) \,. \tag{15}$$

It is possible to construct a bias-hardened estimator that is insensitive to the effects of lensing and which does not have significantly higher variance than the estimator described above [73].

In addition to the bias, lensing also adds variance to the patchy reionization estimator by increasing the observed B-mode polarization power. A strategy to reconstruct the screening effect of patchy reionization using the tools above is therefore to first reconstruct the lensing field, use that estimated lensing field to delens the CMB, and finally to reconstruct the patchy reionization using a bias-hardened estimator. This procedure is difficult to implement in practice and would lead to complicated noise properties in the resulting estimates. One of the main goals of this paper is to compare the performance of an idealized version of the above procedure to the results that can be obtained through machine learning. We will show how our network is capable of making simultaneous estimates of a delensed polarization map, a lensing map, and a patchy reionization map in a single step.

III. DEEP LEARNING NETWORK AND METHODS

In this section, we provide a brief introduction to machine learning and its applications to image processing. We then discuss the design of the ResUNet-CMB architecture that we apply to the task of patchy reionization reconstruction with simulated CMB maps.

A. Deep Neural Networks

Machine learning is a process whereby computers learn the rules to complete tasks when provided a set of examples. In the past decade, artificial neural networks (ANNs) have become ubiquitous in machine learning, being applied to a wide range of problems. The method is inspired by biological systems in that ANNs use neurons as the building blocks. These neurons are grouped into layers and carry out tensor operations.

Deep learning makes use of these neural networks by using multiple successive layers of neurons, offering the ability to learn complex nonlinear relationships hidden in sets of data. There is a significant computational cost in training deep learning networks, but once trained, the networks are capable of very quickly making predictions with very low resource requirements. When dealing with image data in computer vision tasks, the convolutional neural network (CNN) has become the dominant deep learning network design.

It is natural to ask whether deep learning can be usefully applied to cosmology. Maps of the content of the Universe (such as the CMB) are 2D or 3D images, from which we can extract cosmological information using CNNs. As a result, deep learning applications have a growing presence in cosmology. Some recent applications of deep learning to the CMB involving CNNs are full-sky foreground removal [78], fast Wiener filtering [79], mass estimation of galaxy clusters [80, 81], and cosmic string detection [82].

Cosmological structure along our line of sight leaves characteristic distortions imprinted in the observed CMB polarization maps resulting in secondary anisotropies which can be identified and extracted by a CNN. These distortions are correlated over a range of scales, and so the hierarchical feature extraction of the CNN is well-suited to detect the impact of these distortions. In this paper, we employ a particular type of CNN called a Res-UNet to analyze secondary anisotropies of the CMB. This choice is motivated by the success of a ResUNet for flat-sky CMB lensing reconstruction [77]. In the sections that follow we discuss the ingredients that define a Res-UNet, we describe the specific ResUNet architecture we designed for simultaneous lensing and patchy reionization reconstruction, and we specify how our network is implemented and trained.

B. Residual U-Network (ResUNet)

Convolutional layers are the core of a CNN. Each neuron in a convolutional layer operates on a region of pixels of its input called the receptive field. In the first convolutional layer of a network, the receptive field is completely determined by the kernel size. The kernel size specifies the dimensions of the $k \times k$ filters which are slid across the input to extract features. Each filter is sensitive to a unique feature. Common choices for the kernel size use $k = 3, 5, 7, ...$, with odd numbers selected so symmetry about the central pixel can be maintained to reduce edge effects. For more details about convolutional layers see Ref. [83].

By chaining together convolutional layers, neurons in each successive layer become connected to a larger region of the initial input layer thereby increasing the size of each consecutive receptive field. The progressive increase in receptive field size of each convolutional layer makes the network sensitive to high-level features of
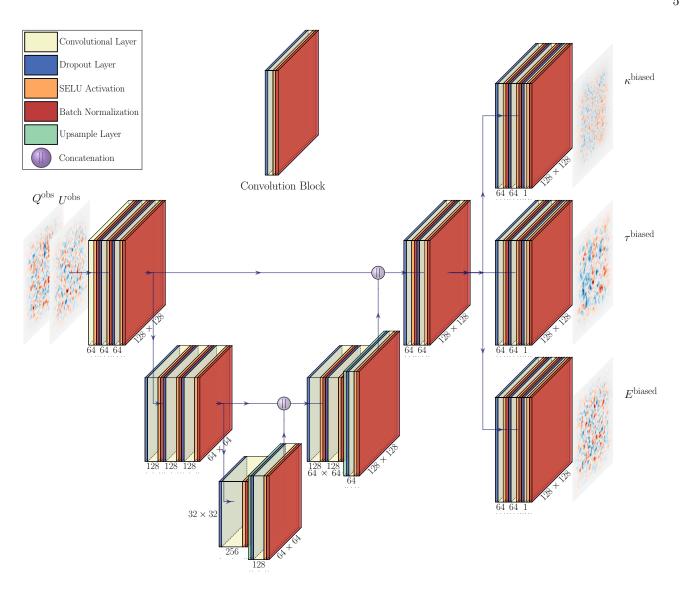
FIG. 1. ResUNet-CMB architecture with residual connections excluded for clarity[1]. The input images, $(Q^{\mathrm{obs}}, U^{\mathrm{obs}})$ are concatenated along the channel dimension. Convolutional layers are labeled with the number of filters they contain. The final images, $(\kappa^{\mathrm{biased}}, \tau^{\mathrm{biased}}, E^{\mathrm{biased}})$, are the output of the final batch normalization layer of each branch. The network has 5,292,367 total parameters and 5,287,113 trainable parameters. The receptive field size when ignoring residual and skip connections is 101×101 pixels for each output. Widths of layers are chosen for visual clarity and are not to scale.

larger scales. For tasks involving images, using a CNN is beneficial since mapping a neuron directly to every pixel would be computationally intractable.

Convolutional autoencoders are a type of CNN that take a tensor for input and produce an identically sized one for output. When used with image data, the tensor takes a shape of (image height, image width, number of channels). For example, if we have a 128×128 pixel RGB image, then the input tensor shape would be (128, 128, 3) where the red, green, and blue pixel values each represent a different channel. Employed with supervised learning, convolutional autoencoders are able to learn a nonlinear mapping between the input and output space.

There are two primary components in a convolutional autoencoder, the encoder and decoder. In the encoding phase, relevant features for the reconstruction of the output are learned by each convolutional layer and increase in complexity, becoming more abstract in deeper layers of the network. The output of each convolution is called a feature map since it is a tensor made of encoded features. The feature maps are continuously down-sampled at regular intervals lowering their dimensionality until a set minimum size is reached. This down-sampling helps increase the receptive field with respect to the input image while also reducing the propagation of unnecessary information.

———————

[1] Graphic made with publicly available code from `https://github.com/HarisIqbal88/PlotNeuralNet`.

The decoder takes the features learned in the last layer of the encoder, the minimum representation, and learns how to combine them to achieve the desired output. Starting from the minimum representation which is an abstract base, it is up-sampled while detail is added through consecutive convolutional layers. The final result is an output image with the same height and width as the initial input to the network.

Some localization information is lost through down-sampling. To combat this, the outputs of some layers in the encoder phase can be concatenated with those of the same size in the decoder. The concatenations are with respect to the channel dimension and are called skip connections. As the output is reconstructed from the minimum representation, the skip connections provide high resolution information just after up-sampling. These skip connections in combination with a convolutional autoencoder were introduced for the purpose of biomedical segmentation, and the resulting architecture is called a U-Net [84]. The encoding phase and decoding phase in the U-Net often mirror each other in both depth and the locations at which up- or down-sampling occur. As a result, this type of network is often presented in a U-shape, hence its name.

As the network grows deeper by stacking convolutional layers, it has been shown that performance and accuracy of the network decreases [85–87]. Residual connections, which take the input of a convolutional layer and add it element-wise to the output of a subsequent convolutional layer, allow more layers to be added to the network without degrading performance by helping to propagate parameter corrections to earlier layers and information to latter ones, through direct identity paths [86, 88, 89]. Their use can help improve network training and results. Adding residual and skip connections to a convolutional autoencoder yields an architecture called a ResUNet [90–92].

C. Network Architecture (ResUNet-CMB)

We focus here on the application of a ResUNet designed to reconstruct the effects of patchy reionization on the CMB polarization in the presence of lensing. After training, the ResUNet takes as input maps of the observed Q and U polarization ($Q^{\mathrm{obs}}, U^{\mathrm{obs}}$) and produces map-level estimates of the lensing convergence, the patchy reionization, and the primordial E-mode polarization ($\kappa, \tau, E^{\mathrm{prim}}$).

We use a modified version of the architecture defined in Ref. [77] which was shown to work in reconstructing the lensing convergence map and therefore provided an ideal starting point from which to build upon. Implementation of our network, which we call ResUNet-CMB, was done with the Keras package of TensorFlow 2.0[2].

A schematic representation of the ResUNet-CMB architecture is shown in Fig. 1 where the residual connections were left out for clarity. The network is made up of a series of convolution blocks. We define a convolution block as consisting of four components in the following order: dropout layer, convolutional layer, activation layer, and batch normalization layer. The convolutional layer is set with a kernel size of 5×5 and 'same' padding. For the activation we use the Scaled Exponential Linear Unit (SELU) function [93]. The dropout layer, with a dropout rate of 0.3 which we found to be the optimal value, is necessary to prevent the network from over-fitting. The dropout layer is omitted from the first convolution block in the network in order to prevent the irretrievable loss of initial information.

The final block of each branch in the decoder phase differs from the others in that the activation function is switched to a linear one. Down-sampling in the encoder phase is achieved by setting the stride to 2 in the convolutional layer. The stride refers to the number of pixels between applications of the filter to the input image. A dedicated up-sampling layer with nearest-neighbor interpolation was used for the decoding phase and was placed before the dropout layer to reduce the production of visual artifacts [94].

Skip connections occur every three convolution blocks in the encoder and concatenate along the channel dimension. Residual connections connect every two convolution blocks. The first residual connection takes the ($Q^{\mathrm{obs}}, U^{\mathrm{obs}}$) input layer and adds it element-wise to the batch normalization output of the second block. The next connection is from the output of the second block to that of the fourth block. The connections continue this pattern until the final two blocks of the decoder where no residual connections are used.

In both the encoding and decoding phases, some residual connections occur between convolution blocks with different input and output tensor dimensions. For such residual connections, a convolutional layer with linear activation and appropriate stride and number of filters is inserted into the connection in order to reconcile the tensor dimensions. If there is a difference in the image size between blocks in the decoder phase, a dedicated upsampling layer is used in addition to the convolutional layer. We found adding a batch normalization layer in every residual connection improved our results by lowering validation loss for all three outputs. In residual connections containing a convolutional layer, the batch normalization is placed after the convolutional layer.

D. Data Pipeline

The publicly available `CAMB`[3] software is used to produce the primordial CMB and lensing power spectra [95]. A cosmology with parameters of $H_0 = 67.9$ km s^{-1} Mpc^{-1}, $\Omega_b h^2 = 0.0222$, $\Omega_c h^2 = 0.118$, $n_s = 0.962$, $\tau = 0.0943$, and $A_s = 2.21 \times 10^{-9}$ was chosen since it allows a comparison with results from Ref. [77]. We take the patchy reionization spectrum $C_\ell^{\tau\tau}$ to match the $\bar{\tau} = 0.058$ model of Ref. [76].

Using the theory power spectra from `CAMB` and a modified version of `Orphics`[4], seven different maps are produced as Gaussian random fields: lensing convergence $\kappa \equiv \frac{1}{2}\nabla^2\phi$, patchy reionization τ, primordial E-mode polarization E^{prim}, primordial Q and U polarization Q^{prim} and U^{prim}, and observed Q and U polarization Q^{obs} and U^{obs}. To obtain the observed polarization maps we use Eq. (14), where the primordial maps (Q^{prim}, U^{prim}) are first modulated by τ, then lensed with κ, and finally summed with a noise map as in Eq. (9).

Each map is 128×128 pixels and covers a $5° \times 5°$ patch of sky. A cosine taper of $1.5°$ was applied to all maps to reduce edge effects. Instrument noise was implemented at the levels of $\Delta_T = 0.0$ μK-arcmin, 0.2 μK-arcmin, 1 μK-arcmin, and 2 μK-arcmin assuming $\Delta_P = \sqrt{2}\Delta_T$. We used $\theta_{\mathrm{FWHM}} = 1.4'$ for the cases of 1 μK-arcmin and 2 μK-arcmin, but $\theta_{\mathrm{FWHM}} = 1.0'$ for the 0.2 μK-arcmin case, the latter of which was chosen to make direct comparison with the results of Ref. [73].

For each noise level, 70000 sets of seven maps were produced with the random seed set to 1225 with the python `NumPy` package. In approximately 20% of the maps selected at random, we set both κ and τ to zero. We found that including these 'null maps' (meaning no lensing or modulation applied to ($Q^{\mathrm{prim}}, U^{\mathrm{prim}}$)) in training helps the network generalize and prevents it from predicting the presence of a signal when none is present in test data. In addition to the set of 70000, we also produce a separate collection of 7000 sets of maps, which we call the prediction set, on which most analysis of ResUNet-CMB predictions was conducted. No null maps were included in the prediction set.

The 70000 maps were split in a 80:10:10 ratio for training, validation, and test sets. Similar to what was done in Ref. [96], all maps in the training, validation and test sets are standardized with respect to values from the training set according to

$$X_{\mathrm{Processed}} = \frac{\left(X - \overline{X}_{\mathrm{Training}}\right)}{\sigma_{\mathrm{Training}}}. \qquad (16)$$

Here X represents a single unprocessed map from any data set. The mean and standard deviation are calculated for each map type over the full training set.

―――――――

[3] `https://camb.info`
[4] `https://github.com/msyriac/orphics`

ResUNet-CMB predictions are made by feeding in ($Q^{\mathrm{obs}}, U^{\mathrm{obs}}$) from the prediction set as input to the fully trained network. The input maps are also standardized according to Eq. (16) with the mean and standard deviation from the prediction set. Making a single prediction of ($\kappa, \tau, E^{\mathrm{prim}}$) regardless of noise level, takes approximately 0.021 seconds on average.

The ResUNet-CMB predictions are post-processed to rescale the outputs to physically meaningful values. This must be done because the network is trained on standardized maps, so the predictions made on the fully trained network will also be standardized. To perform the rescaling, we rearrange Eq. (16) to solve for X using the mean and standard deviation from the training set (since with real data we would not have a priori knowledge of the true ($\kappa, \tau, E^{\mathrm{prim}}$) maps).

E. Training

For each instrument noise level a separate network was trained. We chose to use a batch size of 32 to train the networks. The initial learning rate was set to 0.25 and decayed by a factor of 0.5 if after three epochs there was no improvement in validation loss. Individual learning rate decay of each trainable parameter was achieved by using the Adam optimizer with the TensorFlow default parameters [97]. If there was no improvement in the validation loss after 10 consecutive epochs then training was stopped and the best network saved. Training the noiseless network using a training split of 56000 sets of maps took an average of 10 hours on a single Nvidia Tesla V100 32GB.

IV. RESULTS

In this section we examine the performance of the ResUNet-CMB network. We analyze maps, power spectra, and reconstruction noise curves as evaluation metrics. As described in Section III D, we feed in the inputs ($Q^{\mathrm{obs}}, U^{\mathrm{obs}}$) from the prediction data set to the fully trained network which produces output maps ($\kappa, \tau, E^{\mathrm{prim}}$). The output maps are biased, and so we refer to the direct outputs of the network as ($\kappa^{\mathrm{biased}}, \tau^{\mathrm{biased}}, E^{\mathrm{biased}}$); we address how to treat the bias later in this section.

A. Maps

Figure 2 shows a single prediction for each of ($\kappa^{\mathrm{biased}}, \tau^{\mathrm{biased}}, E^{\mathrm{biased}}$) from the fully trained ResUNet-CMB network for the noiseless and 1 μK-arcmin case. We highlight these results since they allow for partial comparison to results from Ref. [77]. Our ResUNet-CMB network was able to reconstruct some of the largest scale features of $\tau(\hat{\mathbf{n}})$ despite the fact that the impact of patchy
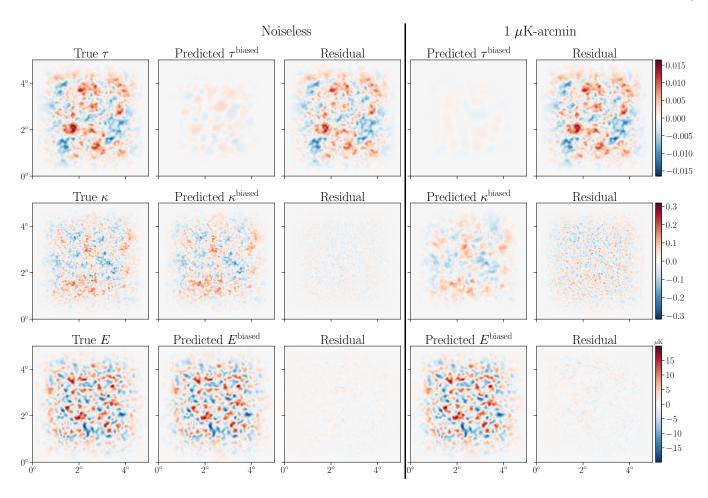
FIG. 2. Sample ResUNet-CMB predictions from fully trained networks for two noise levels compared to the true maps. The predicted maps shown here have not been rescaled as in Eq. (18) and thus have a multiplicative bias. Residual maps are calculated as the true map minus the predicted map.

reionization on the CMB polarization is orders of magnitude smaller than lensing for the chosen cosmological parameters. The noiseless κ^{biased} and E^{biased} are lacking features on only the smallest scales as can be seen in the residual maps which show the difference between the predicted and truth maps.

With an increase in noise, all predictions degrade with κ^{biased} and τ^{biased} seeing the largest change. The primary E-mode polarization is still faithfully reconstructed on most scales despite increasing instrument noise. The 1 μK-arcmin reconstruction of τ^{biased} has residual power that is about 18% higher for $\ell < 840$ than that of the noiseless reconstruction of τ^{biased}, where the residual power is defined as the power spectrum of the true map minus the predicted map. There is a larger fractional increase in the residual power for κ^{biased} with the addition of noise, especially in the range $56 < \ell < 1500$. The predictions of E^{biased} are most affected by noise on small scales, and the increase in the residual power peaks at about $\ell = 2900$ when comparing the noiseless and 1 μK-arcmin predictions.

B. Power Spectra

Next, we compare the recovered pseudo-power spectra from the ResUNet-CMB prediction at each noise level to the truth power spectrum for each map. In Fig. 3 we show a naively corrected $\langle C_\ell^{\tau^{\mathrm{biased}}\tau^{\mathrm{biased}}} \rangle$ pseudo-power spectrum for each noise level. To get the spectra we calculate the auto-power spectrum of each τ^{biased} map from the prediction set and average the results to find $\langle C_\ell^{\tau^{\mathrm{biased}}\tau^{\mathrm{biased}}} \rangle$. We then make a simple correction for power loss due to the window applied to each map by dividing the spectra by the mean of the squared cosine taper. The true input spectrum is the window-corrected average pseudo-power spectra of the noiseless truth maps $\tau(\hat{\mathbf{n}})$ in the prediction set.

As expected, reconstruction of τ by the ResUNet-CMB network is best at low ℓ (large angular scales) and degrades on small scales. As we increase instrumental noise, the fidelity of the reconstruction worsens on all scales. There is a larger decrease in the reconstructed power when going from the noiseless case to 1 μK-arcmin than there is between 1 μK-arcmin and 2 μK-arcmin.
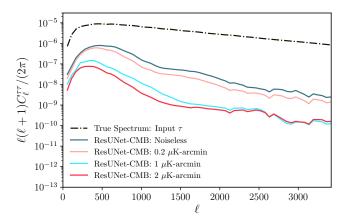
FIG. 3. Power spectra of the optical depth fluctuation maps τ^{biased} predicted by ResUNet-CMB for each of four noise levels averaged over the 7000 realizations included in the prediction set, compared to the averaged power spectrum of the true patchy reionization maps (black dashed-dot).
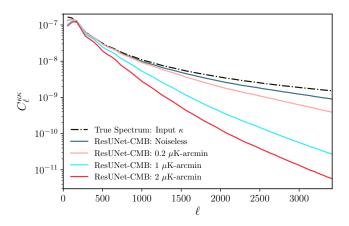

FIG. 4. Power spectra of the lensing convergence maps κ^{biased} predicted by ResUNet-CMB for each of four noise levels averaged over the 7000 realizations included in the prediction set, compared to the averaged power spectrum of the true lensing convergence maps (black dashed-dot).

Figure 4 and Fig. 5 show the power spectra for the ResUNet-CMB reconstructions, κ^{biased} and E^{biased}, respectively. For the κ^{biased} spectra, the reconstruction quickly worsens with increasing instrument noise, especially on small angular scales. Of the three ResUNet-CMB outputs, κ sees the strongest scale dependence in its deviation from the input spectrum.

The reconstructed primordial E-mode map gives results closer to the input than both the τ and κ ResUNet-CMB estimates across the entire range of scales we consider. This is not surprising, given that on the scales we consider the primordial E-mode polarization does not differ a great deal from the modulated and lensed E-mode polarization, the latter of which is just a linear combination of the input Q and U maps in the absence of
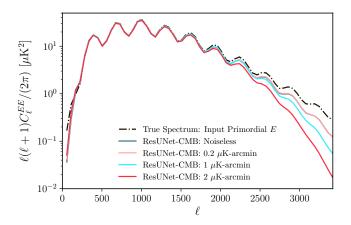

FIG. 5. Power spectra of the primordial E-mode polarization maps E^{biased} predicted by ResUNet-CMB for each of four noise levels averaged over the 7000 realizations included in the prediction set, compared to the averaged power spectrum of the true primordial E modes (black dashed-dot).

noise. We find the noiseless E-mode power spectrum recovery to be greater than about 98% for the ℓ range of $170 < \ell < 1400$. At about $\ell = 3000$ we find E-mode recovery of around 86%. At higher noise levels, the predicted primordial E-mode spectrum falls significantly below the input spectrum on small angular scales. The impact of noise on the reconstruction of E modes is smaller than that for κ and τ on all scales.

We note that the results for our ResUNet-CMB lensing reconstruction match well with what is found in Ref. [77]. However, we find that our reconstructed E-mode power is closer to the input spectrum on all scales than what was found in Ref. [77], though we have not been able to identify the cause of this difference.

C. Noise

We can also compare the results of the ResUNet-CMB network to the standard quadratic estimator. In order to do so, we need to define the reconstruction noise of the deep learning results. The ResUNet-CMB prediction of the modulation field, after post-processing, is a biased result τ^{biased}. Before calculating the reconstruction noise we first need to address this bias. Following treatment of the bias to the reconstructed lensing field from Ref. [77], we define a quantity

$$A_\ell = \left[\frac{\langle C_\ell^{\tau \tau^{\text{biased}}} \rangle}{\langle C_\ell^{\tau \tau} \rangle} \right]^{-1}, \qquad (17)$$

where $\langle \cdots \rangle$ represents the average over the entire prediction set. The quantity A_ℓ is used to rescale the biased output to recover an unbiased estimate of the optical depth fluctuations

$$\hat{\tau}(\boldsymbol{\ell}) = A_\ell \tau^{\text{biased}}(\boldsymbol{\ell}). \qquad (18)$$
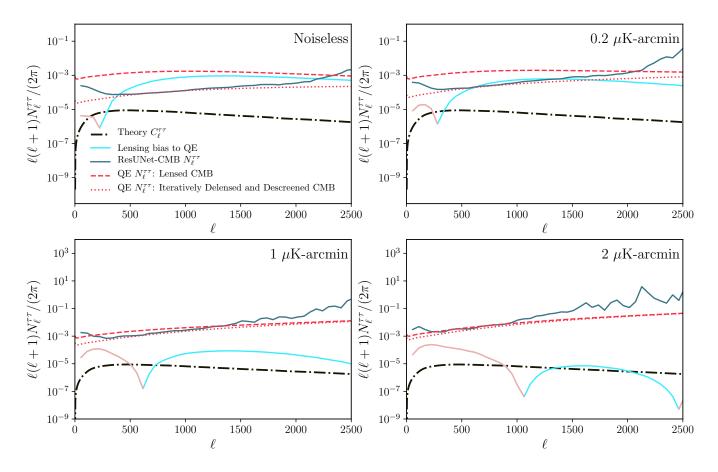
FIG. 6. Reconstruction noise power spectra of ResUNet-CMB normalized prediction $\hat{\tau}$ as defined in Eq. (19) (blue), standard quadratic estimator using lensed CMB spectra (QE; dashed red), and a quadratic estimator using iteratively delensed and descreened CMB spectra (QE; dotted red). The expected lensing bias to the EB quadratic estimator for τ (Eq. (15)) is plotted in cyan where it is positive and coral where negative.

Estimates on real data using ResUNet-CMB would also require rescaling by the same quantity A_ℓ. The unbiased estimate $\hat{\tau}$ can be directly compared to the unbiased estimate from the quadratic estimator discussed in Section II. The reconstruction noise spectrum can then be defined in a way analogous to the definition in Ref. [77],

$$N_\ell^{\tau\tau} = \langle C_\ell^{\hat{\tau}\hat{\tau}} \rangle - \langle C_\ell^{\tau\tau} \rangle, \qquad (19)$$

where we have used

$$\langle C_\ell^{\hat{\tau}\hat{\tau}} \rangle = A_\ell^2 \langle C_\ell^{\tau^{\text{biased}} \tau^{\text{biased}}} \rangle. \qquad (20)$$

It is this definition of the reconstruction noise that we will compare to the quadratic estimator.

We calculate the reconstruction noise of the EB quadratic estimator using Eq. (11) (with $\ell_{\text{max}} = 6100$). As discussed in Section II, the variance of the EB estimator is increased in the presence of lensing, due to the B-mode polarization induced by lensing. We therefore consider two versions of the quadratic estimator: one for which we used lensed CMB spectra, and another for which we employ an iterative delensing and descreening procedure to the CMB spectra.

The iteratively delensed and descreened estimate of the patchy reionization reconstruction noise was calculated as follows. First, we calculate the effects of patchy reionization on the CMB polarization spectra. We use these screened spectra to calculate an estimate of the lensing reconstruction noise using the iterative EB estimator [32]. The B-mode polarization from the patchy reionization acts as an effective source of noise for lensing reconstruction, thereby increasing the lensing reconstruction noise as compared to a case without patchy reionization. We use the resulting lensing reconstruction noise to compute delensed CMB spectra according to Ref. [41]. The resulting delensed CMB spectra were used in the estimate of the patchy reionization reconstruction noise using the EB estimator, with the residual lensing B-mode polarization acting as a source of noise for the patchy reionization reconstruction. If we had stopped the procedure at this point, we would find that ResUNet-CMB outperformed this delensed quadratic estimator in the low noise cases we considered. However, a lower patchy reionization reconstruction noise can be achieved by iteratively 'descreening' the CMB to remove the best es-
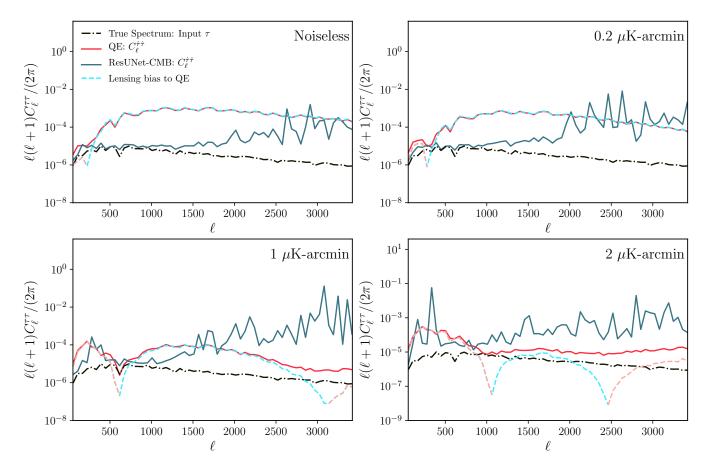
FIG. 7. Power spectra of the patchy τ field reconstructed by ResUNet-CMB (blue) and the quadratic estimator (QE; red) computed from reconstructed maps that are averaged over a set of 7000 simulations with a fixed lensing convergence and patchy reionization field, compared with the true input τ signal power spectrum (black dash-dot) and the expected lensing bias to the quadratic estimator (cyan dashed where the bias is positive, coral dashed where the bias is negative). The sharp features in the power spectra are dependent on the random seeds chosen for the simulations. The quadratic estimator spectra were calculated from maps with no cosine taper applied, and we perform a mean-field subtraction on the estimate, $\hat{\tau}$.

timate of the effects of patchy reionization on the polarization, in a procedure analogous to delensing [98]. We performed this iterative EB descreening estimate, then used the residual effects of patchy reionization on the polarization spectra to compute an improved estimate of the lensing reconstruction noise using the iterative EB estimator. This resulted in a lower lensing reconstruction noise which was then used to calculate an improved estimate of delensed spectra. We iterated this procedure of iterative EB lensing reconstruction and delensing followed by iterative EB patchy reionization reconstruction and descreening followed by iterative EB lensing reconstruction and delensing, etc., to convergence. The output after convergence of this procedure was used as our estimate of 'Iteratively Delensed and Descreened' patchy reionization reconstruction noise. Since the results of iterative delensing closely match the performance of the maximum likelihood lensing estimator [31, 32], it should be expected that the iterative delensing and descreening procedure provides a good approximation of the noise re-

sulting from the maximum likelihood estimate of simultaneous lensing and patchy reionization reconstruction, though the latter has not been demonstrated explicitly.

In addition to the extra variance due to lensing, the quadratic estimator also exhibits a bias to the reconstruction of the patchy reionization in the presence of lensing. We calculate this bias according to Eq. (15). The bias-hardened estimator from Ref. [73] has only a percent-level increase in variance compared to the standard estimator, and so we present the reconstruction noise computed from the standard estimator.

Figure 6 shows the patchy reionization reconstruction noise power spectra for the quadratic estimator and the normalized ResUNet-CMB predictions for each noise level. We also include in each panel the theoretical patchy reionization power spectrum and the bias on the standard quadratic estimator due to lensing. For all four noise levels shown, the ResUNet-CMB predictions have a lower reconstruction noise than the quadratic estimator using lensed spectra in the range $200 < \ell < 1000$. The range
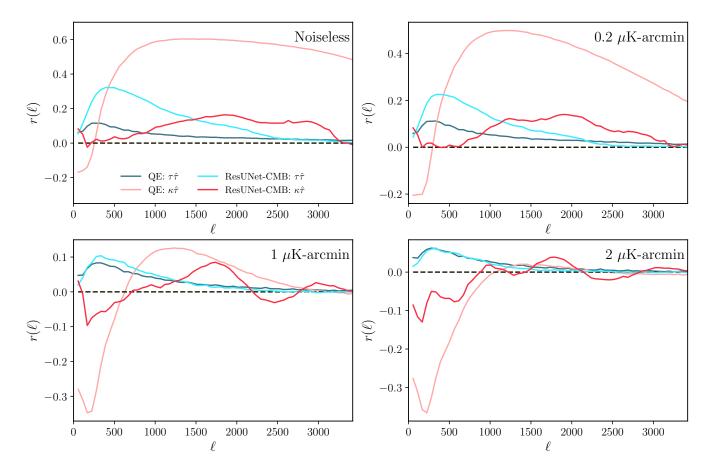
FIG. 8. Correlation coefficients of estimates of the patchy reionization field $\hat{\tau}$ from the standard quadratic estimator (QE) with the true patchy reionization field τ (blue) and the lensing convergence field κ (coral), and those from the ResUNet-CMB network with τ (cyan) and κ (red), all averaged over 7000 simulations. The simulations for the quadratic estimator did not use a cosine taper nor mean-field subtraction.

of angular scales where ResUNet-CMB outperforms the standard quadratic estimator grows as the noise level decreases. The ResUNet-CMB $N_\ell^{\tau\tau}$ matches well with the reconstruction noise of the quadratic estimator using iteratively delensed and descreened CMB spectra for a large ℓ range for all four noise levels. The top two panels demonstrate that the patchy reconstruction noise from ResUNet-CMB is also smaller than the bias due to lensing that is expected in the standard quadratic estimator over a wide range of scales, and this range is wider at lower noise. This behavior indicates that the ResUNet-CMB network learns to reconstruct the patchy τ field in a way that mitigates both the extra variance and the bias due to lensing. The ResUNet-CMB network is therefore able to successfully reconstruct both the lensing convergence and patchy reionization fields simultaneously in a way that is nearly optimal.

D. Bias

In Fig. 7 we show the reconstructed patchy reionization power spectrum from ResUNet-CMB and from the standard quadratic estimator for each noise level. We implement the EB quadratic estimator for τ as in Eq. (10), by using Symlens[5]. The results are computed by averaging over a set of 7000 simulated CMB maps with fixed κ and τ maps; the random seed for κ was set to 133 and for τ to 180. As expected, the quadratic estimator $C_\ell^{\hat{\tau}\hat{\tau}}$ matches well with the lensing bias computed from Eq. (15). The ResUNet-CMB estimator on the other hand follows more closely the input signal power spectrum, which is most easily seen for the low noise cases. At low noise the ResUNet-CMB predictions deviate from the input spectrum at small angular scales ($\ell \gtrsim 1500$), and for higher noise deviations appear even at larger scales.

To confirm that the ResUNet-CMB architecture is truly sensitive to the input τ and is not producing a

---

[5] https://github.com/simonsobs/symlens

spurious signal due to lensing or an artifact of the procedure, we calculate the correlation coefficients of each reconstruction according to

$$r_{Z\hat{\tau}}(\ell) = \frac{\langle C_\ell^{Z\hat{\tau}} \rangle}{\sqrt{\langle C_\ell^{ZZ} \rangle \langle C_\ell^{\hat{\tau}\hat{\tau}} \rangle}} \,, \tag{21}$$

where $Z = \kappa, \tau$. In Fig. 8 we compare the correlation coefficients of the ResUNet-CMB predictions to those of the standard quadratic estimator. Note that our definition of the correlation coefficient in Eq. (21) includes in the denominator the reconstruction noise on the patchy reionization field, and therefore these coefficients are always less than unity. It can clearly be seen that for all noise levels and on almost all scales, the standard quadratic estimator produces an estimate for the optical depth perturbation $\hat{\tau}$ which is more strongly correlated with the lensing convergence κ than it is with the true patchy reionization signal τ. The ResUNet-CMB estimate $\hat{\tau}$ has a larger correlation with the true patchy τ signal and smaller correlation with κ than does the quadratic estimator for nearly all scales at each noise level.

Furthermore, the ResUNet-CMB patchy reionization estimate is more strongly correlated with the true signal than with the lensing convergence on large scales, considerably so at low noise levels. While the quadratic estimator is dominated by the κ signal on almost all scales, the ResUNet-CMB estimator appears to exhibit a lensing bias only on small scales for the two lowest noise levels, and a reduced bias for the largest noise cases shown.

For $\ell > 2000$ in the 1 μK-arcmin and 2 μK-arcmin cases, the near zero positive correlation between $\hat{\tau}$ and the true patchy τ signal indicate ResUNet-CMB is unable to provide an accurate patchy reionization reconstruction on small scales at these noise levels. This might also explain similarity of the reconstructed patchy reionization spectra on small scales for these noise levels seen in Fig. 3.

E. Null Test

As a final check on the results of the ResUNet-CMB network, we perform a null test to ensure that no patchy reionization signal is reconstructed when none is present in the maps. We implement this test by making predictions of $(\kappa, \tau, E^{\mathrm{prim}})$ with the fully trained networks when the unlensed and unmodulated polarization maps $(Q^{\mathrm{prim}}, U^{\mathrm{prim}})$ (plus noise) are passed as the input. In Fig. 9, we plot the power spectra $\langle C_\ell^{\tau^{\mathrm{biased}}\tau^{\mathrm{biased}}} \rangle$ for each noise level that results from this null test. The spectra in Fig. 9 should be compared to the analogous spectra shown in Fig. 3 where the patchy reionization signal was present. We find that ResUNet-CMB reconstructs spectra $\langle C_\ell^{\tau^{\mathrm{biased}}\tau^{\mathrm{biased}}} \rangle$ that are much smaller for the null test than when a signal is present, especially on large angular scales where the reconstruction is most successful. In the noiseless case, the reconstructed spectrum $\langle C_\ell^{\tau^{\mathrm{biased}}\tau^{\mathrm{biased}}} \rangle$
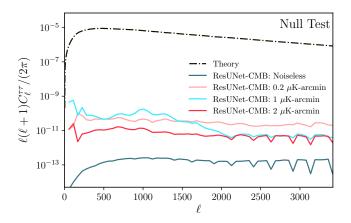

FIG. 9. Null test showing the average power spectra of the patchy reionization field τ^{biased} predicted from fully trained ResUNet-CMB networks on a data set of 7000 simulations with κ and τ set to zero. The results of this figure should be compared to those in Fig. 3. The theoretical patchy reionization spectrum that is used elsewhere in this work has been included in this plot (black dash-dot) to ease comparison, but for the null test shown in this figure, the true $C_\ell^{\tau\tau}$ vanishes.

in the null test is five to seven orders of magnitude smaller than the spectrum with patchy reionization included.

We found that the ResUNet-CMB network was much less successful on this null test when trained without including unlensed and unmodulated maps in the training set. In that case, the model tended to predict the presence of a patchy reionization signal with power spectra not much smaller than those shown in Fig. 3, even when no modulation was present in the maps. The inclusion of some unlensed and unmodulated polarization maps in the data set during training prevents a spurious signal from being reconstructed when lensing and patchy reionization are absent.

V. CONCLUSION

In this paper we described the construction of a deep learning network designed to reconstruct effects which alter the statistics of CMB fluctuations, the ResUNet-CMB. We showed that this network is capable of the simultaneous reconstruction of a lensing map, a patchy reionization map, and a delensed and descreened polarization map. The reconstruction of patchy reionization provided by ResUNet-CMB drastically reduced the lensing-induced bias present in the standard quadratic estimator, and it nearly matched the variance of an optimal iteratively delensed and descreened reconstruction scheme over a wide range of angular scales.

While the results from ResUNet-CMB are quite promising, we confirm the conclusion of other studies that with CMB polarization data alone, we are likely to be able to achieve only a statistical detection of the effects of patchy reionization [73]. The success of ResUNet-

CMB on the simultaneous reconstruction of two sources of secondary CMB fluctuations serves as a demonstration that machine learning tools can be usefully applied to the data expected from the next generation of CMB surveys.

The ResUNet-CMB network shows promising results, though improvements can still be made. The results of the network slightly under-performed idealized predictions for (κ, τ, E) at very low noise levels, especially on small scales. In Ref. [77] it is suggested that the reconstruction of high-ℓ modes is limited by the signal-to-noise ratio at those small scales. For the case of patchy reionization this issue is of particular importance since its effects are much smaller than lensing. This problem can potentially be partially mitigated by using higher resolution maps to include smaller physical scales, which should improve the reconstruction on all scales. Accommodating maps of a different resolution requires extensive restructuring of the network to ensure that features on all scales are being properly analyzed by the network while ensuring VRAM constraints are not exceeded. Another avenue to consider would be to first train the network on patchy reionization maps with a magnified signal, then slowly decrease the signal magnitude until the fiducial amplitude is reached. This method, used in Ref. [99] for a different purpose, could help with the reconstruction on scales dominated by noise. On the other hand, the noise levels where significant gains seem possible are much lower than those of currently planned CMB surveys, meaning that optimization in that very low noise regime may not lead to much practical improvement with forthcoming data.

One extension to our work that could improve patchy reionization reconstruction is to consider a broader set of input maps. One simple addition is to include CMB temperature maps in addition to polarization. More realistic simulations would also include correlations between the patchy reionization field and the matter density, leading to correlations with the lensing field, galaxy counts, or other tracers of structure. If such correlations were included, one could also consider adding maps of the cosmic infrared background, surveys of galaxies, or 21-cm intensity maps, the addition of which has been shown to improve patchy reionization reconstruction by other means [100–102].

Another extension of our work would be treat more realistic simulations, for example by adding astrophysical foregrounds or anisotropic noise. Applications of machine learning to simulated foregrounds on the full sky are already showing promising results [78]. Addressing the complications presented by more realistic simulations is a necessary step if machine learning is to be usefully applied to the data collected by upcoming CMB surveys.

We found a few modifications of standard machine learning techniques to be beneficial for our purposes. We included a batch normalization layer in all of our residual connections, contrary to the standard implementation of residual connections [86, 88]. We found that the inclusion of the batch normalization layer lowered validation loss across all outputs, thereby improving all reconstructions. The success of this change in our network prompts further questions about whether similar changes would be beneficial to other machine learning applications. When training the network, we included in the training set some maps which had both lensing and patchy reionization set to zero. We found this to be necessary for the trained network to successfully pass the null test for patchy reionization reconstruction. Without the null maps in the training set, we found that the network returned non-zero estimates for the patchy reionization signal, even when no patchy reionization was present in the maps of the test set. This procedure of including a subset of maps with varying conditions into the training set can be explored for further possible improvements to the network. The potential benefits of this modified subset training method should not be understated. Not only were the impacts to the quality of the reconstructions minimal, the change to the training also fixed a problem in the predictions of the trained network without requiring any modification to the architecture.

We did not explore the generalization of our predictions to maps generated with different cosmological parameters. Further tests are required to determine how well the network is able to reconstruct patchy reionization and lensing in CMB maps generated assuming different cosmologies. The success of the modified subset training provides a positive indication that a similar technique could be applied to improve the generalization capability of the network without much negative impact to the results.

In order to be useful for constraining cosmological parameters, the predictions made by any estimator need to be incorporated into a framework that is capable of rigorously quantifying uncertainties. Once trained, making predictions with ResUNet-CMB is very computationally efficient, and so the network provides good candidate to be included in a Monte Carlo exploration of posteriors. Another possibility would be to incorporate ResUNet-CMB into a Bayesian framework like that described in Ref. [47].

Our results act as a proof of concept that machine learning is capable of the simultaneous reconstruction of two fields that lead to the distortion of the statistics of CMB fluctuations. It is natural to ask whether machine learning can also be usefully applied to the reconstruction of other sources of secondary CMB fluctuations, perhaps by including additional maps as input. The ResUNet-CMB network described here is a useful model for such exploration, since new architecture aimed at similar tasks can easily be constructed by simply modifying the inputs and outputs of the publicly available network.

The work presented here adds to the list of examples that show machine learning is likely to play a positive role in the future of cosmology, especially when applied to the wealth of new data expected from forthcoming cosmological surveys.

ACKNOWLEDGMENTS

[1] **Simons Observatory** Collaboration, P. Ade *et al.*, "The Simons Observatory: Science goals and forecasts," *JCAP* **02** (2019) 056, `arXiv:1808.07445 [astro-ph.CO]`.

[2] M. Aravena *et al.*, "The CCAT-Prime Submillimeter Observatory," `arXiv:1909.02587 [astro-ph.IM]`.

[3] **CMB-S4** Collaboration, K. N. Abazajian *et al.*, "CMB-S4 Science Book, First Edition," `arXiv:1610.02743 [astro-ph.CO]`.

[4] **NASA PICO** Collaboration, S. Hanany *et al.*, "PICO: Probe of Inflation and Cosmic Origins," `arXiv:1902.10541 [astro-ph.IM]`.

[5] N. Sehgal *et al.*, "CMB-HD: An Ultra-Deep, High-Resolution Millimeter-Wave Survey Over Half the Sky," `arXiv:1906.10134 [astro-ph.CO]`.

[6] N. Aghanim, S. Majumdar, and J. Silk, "Secondary anisotropies of the CMB," *Rept. Prog. Phys.* **71** (2008) 066902, `arXiv:0711.0518 [astro-ph]`.

[7] A. Lewis and A. Challinor, "Weak gravitational lensing of the CMB," *Phys. Rept.* **429** (2006) 1–65, `arXiv:astro-ph/0601594`.

[8] Ya. B. Zeldovich and R. A. Sunyaev, "The Interaction of Matter and Radiation in a Hot-Model Universe," *Astrophys. Space Sci.* **4** (1969) 301–316.

[9] R. A. Sunyaev and Ya. B. Zeldovich, "The Interaction of matter and radiation in the hot model of the universe," *Astrophys. Space Sci.* **7** (1970) 20–30.

[10] R. A. Sunyaev and Ya. B. Zeldovich, "The Observations of relic radiation as a test of the nature of X-Ray radiation from the clusters of galaxies," *Comments Astrophys. Space Phys.* **4** (1972) 173–178.

[11] R. A. Sunyaev and Ya. B. Zeldovich, "Microwave background radiation as a probe of the contemporary structure and history of the universe," *Ann. Rev. Astron. Astrophys.* **18** (1980) 537–560.

[12] S. Y. Sazonov and R. A. Sunyaev, "Microwave polarization in the direction of galaxy clusters induced by the CMB quadrupole anisotropy," *Mon. Not. Roy. Astron. Soc.* **310** (1999) 765–772, `arXiv:astro-ph/9903287 [astro-ph]`.

[13] R. K. Sachs and A. M. Wolfe, "Perturbations of a Cosmological Model and Angular Variations of the Microwave Background," ApJ **147** (Jan., 1967) 73.

[14] M. J. Rees and D. W. Sciama, "Large-scale Density Inhomogeneities in the Universe," Nature **217** (Feb., 1968) 511–516.

[15] M. Birkinshaw and S. F. Gull, "A test for transverse motions of clusters of galaxies," Nature **302** (Mar., 1983) 315–317.

[16] K. M. Smith and S. Ferraro, "Detecting Patchy Reionization in the Cosmic Microwave Background," *Phys. Rev. Lett.* **119** no. 2, (2017) 021301, `arXiv:1607.01769 [astro-ph.CO]`.

[17] T. Louis, E. F. Bunn, B. Wandelt, and J. Silk, "Measuring Polarized Emission in Clusters in the CMB S4 Era," *Phys. Rev. D* **96** no. 12, (2017) 123509, `arXiv:1707.04102 [astro-ph.CO]`.

[18] A.-S. Deutsch, M. C. Johnson, M. Münchmeyer, and A. Terrana, "Polarized Sunyaev Zel'dovich tomography," *JCAP* **04** (2018) 034, `arXiv:1705.08907 [astro-ph.CO]`.

[19] A.-S. Deutsch, E. Dimastrogiovanni, M. C. Johnson, M. Münchmeyer, and A. Terrana, "Reconstruction of the remote dipole and quadrupole fields from the kinetic Sunyaev Zel'dovich and polarized Sunyaev Zel'dovich effects," *Phys. Rev. D* **98** no. 12, (2018) 123501, `arXiv:1707.08129 [astro-ph.CO]`.

[20] J. Meyers, P. D. Meerburg, A. van Engelen, and N. Battaglia, "Beyond CMB cosmic variance limits on reionization with the polarized Sunyaev-Zel'dovich effect," *Phys. Rev. D* **97** no. 10, (2018) 103505, `arXiv:1710.01708 [astro-ph.CO]`.

[21] K. M. Smith, M. S. Madhavacheril, M. Münchmeyer, S. Ferraro, U. Giri, and M. C. Johnson, "KSZ tomography and the bispectrum," `arXiv:1810.13423 [astro-ph.CO]`.

[22] S. C. Hotinli, J. Meyers, N. Dalal, A. H. Jaffe, M. C. Johnson, J. B. Mertens, M. Münchmeyer, K. M. Smith, and A. van Engelen, "Transverse Velocities with the Moving Lens Effect," *Phys. Rev. Lett.* **123** no. 6, (2019) 061301, `arXiv:1812.03167 [astro-ph.CO]`.

[23] S. Yasini, N. Mirzatuny, and E. Pierpaoli, "Pairwise Transverse Velocity Measurement with the Rees–Sciama Effect," *Astrophys. J. Lett.* **873** no. 2, (2019) L23, `arXiv:1812.04241 [astro-ph.CO]`.

[24] S. C. Hotinli, M. C. Johnson, and J. Meyers, "Optimal filters for the moving lens effect," `arXiv:2006.03060 [astro-ph.CO]`.

[25] W. Hu and T. Okamoto, "Mass reconstruction with cmb polarization," *Astrophys. J.* **574** (2002) 566–574, `arXiv:astro-ph/0111606`.

[26] T. Okamoto and W. Hu, "CMB lensing reconstruction on the full sky," *Phys. Rev.* **D67** (2003) 083002, `arXiv:astro-ph/0301031 [astro-ph]`.

[27] **Planck** Collaboration, N. Aghanim *et al.*, "Planck 2018 results. VIII. Gravitational lensing," *Astron. Astrophys.* **641** (2020) A8, `arXiv:1807.06210 [astro-ph.CO]`.

[28] W. Wu *et al.*, "A Measurement of the Cosmic Microwave Background Lensing Potential and Power Spectrum from 500 deg^2 of SPTpol Temperature and Polarization Data," *Astrophys. J.* **884** (2019) 70, `arXiv:1905.05777 [astro-ph.CO]`.

[29] O. Darwish *et al.*, "The Atacama Cosmology Telescope: A CMB lensing mass map over 2100 square degrees of sky and its cross-correlation with BOSS-CMASS galaxies," `arXiv:2004.01139 [astro-ph.CO]`.

[30] C. M. Hirata and U. Seljak, "Analyzing weak lensing of the cosmic microwave background using the

likelihood function," *Phys. Rev.* **D67** (2003) 043001, `arXiv:astro-ph/0209489 [astro-ph]`.

[31] C. M. Hirata and U. Seljak, "Reconstruction of lensing from the cosmic microwave background polarization," *Phys. Rev. D* **68** (2003) 083002, `arXiv:astro-ph/0306354`.

[32] K. M. Smith, D. Hanson, M. LoVerde, C. M. Hirata, and O. Zahn, "Delensing CMB Polarization with External Datasets," *JCAP* **06** (2012) 014, `arXiv:1010.0048 [astro-ph.CO]`.

[33] T. Namikawa, D. Hanson, and R. Takahashi, "Bias-Hardened CMB Lensing," *Mon. Not. Roy. Astron. Soc.* **431** (2013) 609–620, `arXiv:1209.0091 [astro-ph.CO]`.

[34] M. Kamionkowski, A. Kosowsky, and A. Stebbins, "Statistics of cosmic microwave background polarization," *Phys. Rev. D* **55** (1997) 7368–7388, `arXiv:astro-ph/9611125`.

[35] U. Seljak and M. Zaldarriaga, "Signature of gravity waves in polarization of the microwave background," *Phys. Rev. Lett.* **78** (1997) 2054–2057, `arXiv:astro-ph/9609169`.

[36] M. Zaldarriaga and U. Seljak, "Gravitational lensing effect on cosmic microwave background polarization," *Phys. Rev. D* **58** (1998) 023003, `arXiv:astro-ph/9803150`.

[37] A. Lewis, A. Challinor, and N. Turok, "Analysis of CMB polarization on an incomplete sky," *Phys. Rev. D* **65** (2002) 023505, `arXiv:astro-ph/0106536`.

[38] L. Knox and Y.-S. Song, "A Limit on the detectability of the energy scale of inflation," *Phys. Rev. Lett.* **89** (2002) 011303, `arXiv:astro-ph/0202286`.

[39] M. Kesden, A. Cooray, and M. Kamionkowski, "Separation of gravitational wave and cosmic shear contributions to cosmic microwave background polarization," *Phys. Rev. Lett.* **89** (2002) 011304, `arXiv:astro-ph/0202434`.

[40] U. Seljak and C. M. Hirata, "Gravitational lensing as a contaminant of the gravity wave signal in CMB," *Phys. Rev. D* **69** (2004) 043005, `arXiv:astro-ph/0310163`.

[41] D. Green, J. Meyers, and A. van Engelen, "CMB Delensing Beyond the B Modes," *JCAP* **1712** no. 12, (2017) 005, `arXiv:1609.08143 [astro-ph.CO]`.

[42] D. Baumann, D. Green, J. Meyers, and B. Wallisch, "Phases of New Physics in the CMB," *JCAP* **01** (2016) 007, `arXiv:1508.06342 [astro-ph.CO]`.

[43] W. R. Coulton, P. D. Meerburg, D. G. Baker, S. Hotinli, A. J. Duivenvoorden, and A. van Engelen, "Minimizing gravitational lensing contributions to the primordial bispectrum covariance," *Phys. Rev. D* **101** no. 12, (2020) 123504, `arXiv:1912.07619 [astro-ph.CO]`.

[44] K. Abazajian *et al.*, "CMB-S4 Science Case, Reference Design, and Project Plan," `arXiv:1907.04473 [astro-ph.IM]`.

[45] **CMB-S4** Collaboration, K. Abazajian *et al.*, "CMB-S4: Forecasting Constraints on Primordial Gravitational Waves," `arXiv:2008.12619 [astro-ph.CO]`.

[46] M. Millea, E. Anderes, and B. D. Wandelt, "Bayesian delensing of CMB temperature and polarization," *Phys. Rev. D* **100** no. 2, (2019) 023509, `arXiv:1708.06753 [astro-ph.CO]`.

[47] M. Millea, E. Anderes, and B. D. Wandelt, "Bayesian delensing delight: sampling-based inference of the primordial CMB and gravitational lensing," `arXiv:2002.00965 [astro-ph.CO]`.

[48] P. Diego-Palazuelos, P. Vielva, E. Martínez-González, and R. Barreiro, "Comparison of delensing methodologies and assessment of the delensing capabilities of future experiments," `arXiv:2006.12935 [astro-ph.CO]`.

[49] B. D. Sherwin and M. Schmittfull, "Delensing the CMB with the Cosmic Infrared Background," *Phys. Rev. D* **92** no. 4, (2015) 043005, `arXiv:1502.05356 [astro-ph.CO]`.

[50] N. Sehgal, M. S. Madhavacheril, B. Sherwin, and A. van Engelen, "Internal Delensing of Cosmic Microwave Background Acoustic Peaks," *Phys. Rev. D* **95** no. 10, (2017) 103512, `arXiv:1612.03898 [astro-ph.CO]`.

[51] R. Barkana and A. Loeb, "In the beginning: The First sources of light and the reionization of the Universe," *Phys. Rept.* **349** (2001) 125–238, `arXiv:astro-ph/0010468`.

[52] J. B. Wyithe and A. Loeb, "Reionization of hydrogen and helium by early stars and quasars," *Astrophys. J.* **586** (2003) 693–708, `arXiv:astro-ph/0209056`.

[53] R. Barkana and A. Loeb, "Unusually large fluctuations in the statistics of galaxy formation at high redshift," *Astrophys. J.* **609** (2004) 474–481, `arXiv:astro-ph/0310338`.

[54] S. Furlanetto, M. Zaldarriaga, and L. Hernquist, "The Growth of HII regions during reionization," *Astrophys. J.* **613** (2004) 1–15, `arXiv:astro-ph/0403697`.

[55] M. McQuinn, A. Lidz, O. Zahn, S. Dutta, L. Hernquist, and M. Zaldarriaga, "The Morphology of HII Regions during Reionization," *Mon. Not. Roy. Astron. Soc.* **377** (2007) 1043–1063, `arXiv:astro-ph/0610094`.

[56] A. Mesinger and S. Furlanetto, "Efficient Simulations of Early Structure Formation and Reionization," *Astrophys. J.* **669** (2007) 663, `arXiv:0704.0946 [astro-ph]`.

[57] N. Battaglia, H. Trac, R. Cen, and A. Loeb, "Reionization on Large Scales I: A Parametric Model Constructed from Radiation-Hydrodynamic Simulations," *Astrophys. J.* **776** (2013) 81, `arXiv:1211.2821 [astro-ph.CO]`.

[58] S. Mitra, T. R. Choudhury, and A. Ferrara, "Cosmic reionization after Planck II: contribution from quasars," *Mon. Not. Roy. Astron. Soc.* **473** no. 1, (2018) 1416–1425, `arXiv:1606.02719 [astro-ph.CO]`.

[59] P. Dayal and A. Ferrara, "Early galaxy formation and its large-scale effects," *Phys. Rept.* **780-782** (2018) 1–64, `arXiv:1809.09136 [astro-ph.GA]`.

[60] W. Hu, "Reionization revisited: secondary cmb anisotropies and polarization," *Astrophys. J.* **529** (2000) 12, `arXiv:astro-ph/9907103`.

[61] M. G. Santos, A. Cooray, Z. Haiman, L. Knox, and C.-P. Ma, "Small - scale CMB temperature and polarization anisotropies due to patchy reionization," *Astrophys. J.* **598** (2003) 756–766, `arXiv:astro-ph/0305471`.

[62] O. Zahn, M. Zaldarriaga, L. Hernquist, and M. McQuinn, "The Influence of non-uniform reionization on the CMB," *Astrophys. J.* **630** (2005)

657–666, arXiv:astro-ph/0503166.

[63] M. McQuinn, S. R. Furlanetto, L. Hernquist, O. Zahn, and M. Zaldarriaga, "The Kinetic Sunyaev-Zel'dovich effect from reionization," *Astrophys. J.* **630** (2005) 643–656, arXiv:astro-ph/0504189.

[64] O. Dore, G. Holder, M. Alvarez, I. T. Iliev, G. Mellema, U.-L. Pen, and P. R. Shapiro, "The Signature of Patchy Reionization in the Polarization Anisotropy of the CMB," *Phys. Rev. D* **76** (2007) 043002, arXiv:astro-ph/0701784.

[65] C. Dvorkin and K. M. Smith, "Reconstructing Patchy Reionization from the Cosmic Microwave Background," *Phys. Rev. D* **79** (2009) 043003, arXiv:0812.1566 [astro-ph].

[66] C. Dvorkin, W. Hu, and K. M. Smith, "B-mode CMB Polarization from Patchy Screening during Reionization," *Phys. Rev. D* **79** (2009) 107302, arXiv:0902.4413 [astro-ph.CO].

[67] N. Battaglia, A. Natarajan, H. Trac, R. Cen, and A. Loeb, "Reionization on Large Scales III: Predictions for Low-ℓ Cosmic Microwave Background Polarization and High-ℓ Kinetic Sunyaev-Zel'dovich Observables," *Astrophys. J.* **776** (2013) 83, arXiv:1211.2832 [astro-ph.CO].

[68] H. Park, P. R. Shapiro, E. Komatsu, I. T. Iliev, K. Ahn, and G. Mellema, "The Kinetic Sunyaev-Zel'dovich effect as a probe of the physics of cosmic reionization: the effect of self-regulated reionization," *Astrophys. J.* **769** (2013) 93, arXiv:1301.3607 [astro-ph.CO].

[69] M. A. Alvarez, "The Kinetic Sunyaev–Zel'dovich Effect From Reionization: Simulated Full-sky Maps at Arcminute Resolution," *Astrophys. J.* **824** no. 2, (2016) 118, arXiv:1511.02846 [astro-ph.CO].

[70] S. Paul, S. Mukherjee, and T. R. Choudhury, "Inevitable imprints of patchy reionization on the cosmic microwave background anisotropy," *Mon. Not. Roy. Astron. Soc.* **500** no. 1, (2020) 232–246, arXiv:2005.05327 [astro-ph.CO].

[71] T. R. Choudhury, S. Mukherjee, and S. Paul, "CMB constraints on a physical model of reionization," arXiv:2007.03705 [astro-ph.CO].

[72] S. Mukherjee, S. Paul, and T. R. Choudhury, "Is patchy reionization an obstacle in detecting the primordial gravitational wave signal?," *Mon. Not. Roy. Astron. Soc.* **486** no. 2, (2019) 2042–2049, arXiv:1903.01994 [astro-ph.CO].

[73] M. Su, A. P. Yadav, M. McQuinn, J. Yoo, and M. Zaldarriaga, "An Improved Forecast of Patchy Reionization Reconstruction with CMB," arXiv:1106.4313 [astro-ph.CO].

[74] V. Gluscevic, M. Kamionkowski, and D. Hanson, "Patchy Screening of the Cosmic Microwave Background by Inhomogeneous Reionization," *Phys. Rev. D* **87** no. 4, (2013) 047303, arXiv:1210.5507 [astro-ph.CO].

[75] T. Namikawa, "Constraints on Patchy Reionization from Planck CMB Temperature Trispectrum," *Phys. Rev. D* **97** no. 6, (2018) 063505, arXiv:1711.00058 [astro-ph.CO].

[76] A. Roy, A. Lapi, D. Spergel, and C. Baccigalupi, "Observing patchy reionization with future CMB polarization experiments," *JCAP* **05** (2018) 014, arXiv:1801.02393 [astro-ph.CO].

[77] J. Caldeira, W. Wu, B. Nord, C. Avestruz, S. Trivedi, and K. Story, "DeepCMB: Lensing Reconstruction of the Cosmic Microwave Background with Deep Neural Networks," *Astron. Comput.* **28** (2019) 100307, arXiv:1810.01483 [astro-ph.CO].

[78] M. A. Petroff, G. E. Addison, C. L. Bennett, and J. L. Weiland, "Full-sky Cosmic Microwave Background Foreground Cleaning Using Machine Learning," arXiv:2004.11507 [astro-ph.CO].

[79] M. Münchmeyer and K. M. Smith, "Fast Wiener filtering of CMB maps with Neural Networks," arXiv:1905.05846 [astro-ph.CO].

[80] N. Gupta and C. Reichardt, "Mass Estimation of Galaxy Clusters with Deep Learning II: CMB Cluster Lensing," arXiv:2005.13985 [astro-ph.CO].

[81] N. Gupta and C. L. Reichardt, "Mass Estimation of Galaxy Clusters with Deep Learning I: Sunyaev-Zel'dovich Effect," *Astrophys. J.* **900** no. 2, (2020) 110, arXiv:2003.06135 [astro-ph.CO].

[82] R. Ciuca, O. F. Hernández, and M. Wolman, "A Convolutional Neural Network For Cosmic String Detection in CMB Temperature Maps," *Mon. Not. Roy. Astron. Soc.* **485** (2019) 1377, arXiv:1708.08878 [astro-ph.CO].

[83] V. Dumoulin and F. Visin, "A guide to convolution arithmetic for deep learning," arXiv:1603.07285 [stat.ML].

[84] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," arXiv:1505.04597 [cs.CV].

[85] K. He and J. Sun, "Convolutional neural networks at constrained time cost," arXiv:1412.1710 [cs.CV].

[86] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," arXiv:1512.03385 [cs.CV].

[87] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," arXiv:1505.00387 [cs.LG].

[88] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," arXiv:1603.05027 [cs.CV].

[89] D. Balduzzi, M. Frean, L. Leary, J. Lewis, K. W.-D. Ma, and B. McWilliams, "The shattered gradients problem: If resnets are the answer, then what is the question?," arXiv:1702.08591 [cs.NE].

[90] B. Kayalibay, G. Jensen, and P. van der Smagt, "Cnn-based segmentation of medical imaging data," *CoRR* **abs/1701.03056** (2017) , arXiv:1701.03056. http://arxiv.org/abs/1701.03056.

[91] F. Milletari, N. Navab, and S. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," *CoRR* **abs/1606.04797** (2016) , arXiv:1606.04797. http://arxiv.org/abs/1606.04797.

[92] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters* **15** no. 5, (May, 2018) 749–753. http://dx.doi.org/10.1109/LGRS.2018.2802944.

[93] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," arXiv:1706.02515 [cs.LG].

[94] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill* (2016) . http://distill.pub/2016/deconv-checkerboard.

[95] A. Lewis, A. Challinor, and A. Lasenby, "Efficient computation of CMB anisotropies in closed FRW models," *Astrophys. J.* **538** (2000) 473–476, `arXiv:astro-ph/9911177`.

[96] J. Chardin, G. Uhlrich, D. Aubert, N. Deparis, N. Gillet, P. Ocvirk, and J. Lewis, "A deep learning model to emulate simulations of cosmic reionization," *Mon. Not. Roy. Astron. Soc.* **490** no. 1, (2019) 1055–1065, `arXiv:1905.06958 [astro-ph.CO]`.

[97] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," `arXiv:1412.6980 [cs.LG]`.

[98] P. D. Meerburg, J. Meyers, K. M. Smith, and A. van Engelen, "Reconstructing CMB fluctuations and the mean reionization optical depth," *Phys. Rev. D* **95** no. 12, (2017) 123538, `arXiv:1701.06992 [astro-ph.CO]`.

[99] R. Ciuca and O. F. Hernández, "Information Theoretic Bounds on Cosmic String Detection in CMB Maps with Noise," *Mon. Not. Roy. Astron. Soc.* **492** no. 1, (2020) 1329–1334, `arXiv:1911.06378 [astro-ph.CO]`.

[100] C. Feng and G. Holder, "Detecting Electron Density Fluctuations from Cosmic Microwave Background Polarization using a Bispectrum Approach," *Phys. Rev. D* **97** no. 12, (2018) 123523, `arXiv:1801.05396 [astro-ph.CO]`.

[101] C. Feng and G. Holder, "Searching for patchy reionization from cosmic microwave background with hybrid quadratic estimators," *Phys. Rev. D* **99** no. 12, (2019) 123502, `arXiv:1808.01592 [astro-ph.CO]`.

[102] S. C. Hotinli and M. C. Johnson, "Reconstructing large scales at cosmic dawn," `arXiv:2012.09851 [astro-ph.CO]`.