

Predicting 21cm-line map from Lyman α emitter distribution with Generative Adversarial Networks

Shintaro Yoshiura,^{1*} Hayato Shimabukuro,² Kenji Hasegawa³ and Keitaro Takahashi^{4,5}

¹*The University of Melbourne, School of Physics, Parkville, VIC 3010, Australia*

²*South-Western Institute for Astronomy Research (SWIFAR), Yunnan University (YNU), Kunming 650500, People's Republic of China*

³*Department of Physics and Astrophysics, Nagoya University Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan*

⁴*Faculty of Science, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan*

⁵*International Research Organization for Advanced Science and Technology, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The radio observation of 21cm-line signal from the Epoch of Reionization (EoR) enables us to explore the evolution of galaxies and intergalactic medium in the early universe. However, the detection and imaging of 21cm-line signal are tough due to the foreground and instrumental systematics. In order to overcome these obstacles, as a new approach, we propose to take a cross correlation between observed 21cm-line data and 21cm-line images generated from the distribution of the Lyman- α emitters (LAEs) through machine learning. In order to create 21cm-line maps from LAE distribution, we apply conditional Generative Adversarial Network (cGAN) trained with the results of our numerical simulations. We find that 21cm-line brightness temperature maps and the neutral fraction maps can be well reproduced at large scales. Furthermore, we show that the cross correlation is detectable at $k < 0.2 \text{ Mpc}^{-1}$ by combining 400 hours of MWA Phase II observation and LAE deep survey of the Subaru Hyper Suprime Camera. Our new approach of cross correlation with image construction using the cGAN can not only boost the detectability of EoR 21cm-line signal but also allow us to estimate the 21cm-line auto-power spectrum.

Key words: cosmology: dark ages, reionization, first stars

1 INTRODUCTION

One of the milestones in the history of the Universe is the epoch of reionization (EoR). The ultraviolet (UV) photons escaped from ionizing sources such as early galaxies ionize the hydrogen atoms in the intergalactic medium (IGM) and generate ionized bubbles around ionizing sources. The morphology and topological property of ionized region depends on the nature of the ionizing sources. However, the dominant source of ionizing photons still remains unknown. The redshifted cosmological 21cm line from the neutral hydrogen atom is one of promising tools to probe the morphology of ionized bubbles in the IGM during the EoR.

Currently, low-frequency radio telescopes such as the the Giant Metrewave Radio Telescope EoR Experiment (GMRT, [Paciga et al. 2013](#)) the Donald C. Backer Precision Array for Probing the Epoch of Reionization (PAPER, [Parsons et al. 2010](#)), the Murchison Widefield Array (MWA, [Tingay et al. 2013](#)) and the LOw Frequency ARray (LOFAR, [van Haarlem et al. 2013](#)) are operating to detect the 21cm

line signal from the EoR. These ongoing radio telescopes have provided upper limits on the 21cm-line power spectrum, and these upper limits have been gradually updated thanks to sophisticated analysis (e.g [Barry et al. 2019](#); [Ali et al. 2015](#); [Paciga et al. 2011](#); [Patil et al. 2017](#))

However, we are facing many difficulties in detecting the 21cm line. Especially, observations of the 21cm line from the EoR is obstructed by the bright foreground contamination and observational systematic errors. The foreground contamination is mainly due to the synchrotron emission from our Galaxy and extragalactic radio sources and is about 4 orders of magnitude larger than the 21cm-line signal. Further, the observational data are polluted by various systematics such as Earth's ionosphere, the beam-shape error and the radio frequency interference. Thus, it is required to perform a foreground removal/avoidance with a high accuracy and overcome these systematics adequately in order to detect the 21cm-line signal.

The cross correlation between the 21cm line and other emission lines is one of powerful methods to reduce the effects of contamination from the foreground and systematics because the foreground and systematics should not corre-

* E-mail: syoshiura@unimelb.edu.au

late with the partner lines of the cross correlation. High- z galaxies have been studied as the partner of the cross correlation in previous works (e.g. Lidz et al. 2009; Park et al. 2014; Sobacchi et al. 2016; Hutter et al. 2017; Moriwaki et al. 2019; Feng et al. 2017; Heneka et al. 2017; Wiersma et al. 2013; Feng et al. 2017). Since the ionizing photon emitted from high- z massive galaxies generate large ionized regions around them, the 21cm-line signal negatively correlates with galaxies at large scales. In addition to the negative correlation at large scales, the 21cm-line signal can positively correlate with galaxies at small scales due to matter density fluctuation.

In Kubota et al. (2018) and Yoshiura et al. (2018), we have studied the Lyman- α emitters (LAEs) as a partner of the cross correlation. The LAE is a kind of high- z galaxies which emit strong Lyman- α line. So far, Subaru Hyper Supreme Cam (HSC) has detected a large number of LAEs in large survey fields at $z > 5.7$ (Ouchi et al. 2018; Shibuya et al. 2018; Konno et al. 2018) and further surveys at higher redshifts are ongoing. We have found the cross power spectrum between the 21cm line and Lyman- α emitters can be detected by combining the MWA, LAE surveys by the Subaru HSC and a follow-up by the Prime Focus Spectrograph (PFS, Takada et al. 2014; Tamura et al. 2016) to determine precise redshifts of the LAEs.

In this work, we propose a new cross correlation approach to detect the 21cm-line signal. First, we predict 21cm-line maps from the observed spatial distribution of LAEs with the conditional Generative Adversarial Networks (cGAN), which is one of the ANN techniques¹. Specifically, to predict 21cm-line maps from LAE distribution, we use an image-to-image translator developed in Isola et al. (2016) based on the cGAN technique. As mentioned above, the 21cm-line signal correlates with the LAE distribution at scales of a wide range, and therefore it would be possible to predict the 21cm-line maps from the observed LAE distribution if the network has learned the 21cm-LAE cross correlation properly in advance. The predicted 21cm-line maps can be used as the partner of the cross correlation to not only boost the detectability of 21cm-line signal but also estimate its auto-correlation which cannot be obtained by the conventional cross-correlation method with galaxies.

We show the feasibility of our new method using simulations. First, we develop a network with training data sets, which consist of input data (LAE spatial distribution) and correct output data (21cm-line map). We employ the result of two types of numerical simulations of galaxy formation and reionization to prepare the training data sets. Then we evaluate the performance of the network with test data sets which are also obtained from the numerical simulations. Furthermore, we investigate the detectability of the cross correlation between the predicted 21cm-line maps and the observed 21cm-line data assuming 21cm-line observations by the MWA and Subaru HSC and PFS observations. Finally,

we argue the estimation of the 21cm auto-power spectrum using the cross correlation.

This paper is structured as follows. In Sec. 2, we describe the 21cm line and our reionization simulation model. In Sec. 3, we introduce the cGAN method and a training data set used in this work. In Sec. 4, we outline our method for evaluation of the detectability of the 21cm line. In Secs. 5 and 6, we show our results and give a discussion. Finally, we summarise our work in Sec. 7.

2 21CM LINE AND SIMULATION

In this work, we attempt to construct 21cm-line maps from spatial distribution of LAEs. The redshifted 21cm line, emitted from neutral hydrogen, is measured as the brightness temperature which is described as (Furlanetto et al. 2006, e.g.),

$$\delta T_b = 27 x_{\text{HI}} (1 + \delta_m) \left(1 - \frac{T_{\text{CMB}}}{T_S} \right) \times \left(\frac{\Omega_b h^2}{0.023} \right) \left(\frac{0.15}{\Omega_m h^2} \frac{1+z}{10} \right)^{1/2} \text{ mK}, \quad (1)$$

where x_{HI} is the neutral fraction, δ_m is the matter density fluctuation, T_{CMB} is the CMB temperature, T_S is the spin temperature, h is the hubble parameter, Ω_b and Ω_m are the density parameter of baryon and matter, respectively. We assume the IGM is sufficiently heated and $T_S \gg T_{\text{CMB}}$. This assumption is reasonable since we focus on lower redshift $z = 6.6$.

For modeling the 21cm line and the LAE distribution, we employ a radiative transfer (RT) numerical simulation and a semi-numerical simulation for solving the ionization. By comparing the networks developed from two types of simulation, we discuss the effect of model of ionization to the accuracy of signal prediction. The LAE distribution is evaluated by solving a Lyman- α 1D RT from the data. We describe the models below.

2.1 Radiative Transfer

We briefly describe our RT reionization simulation which is identical to that studied in Kubota et al. (2018).

Our simulation employs the matter density fields obtained from a massive cosmological N-body simulation (Ishiyama et al. 2012, 2009). The RT equations are solved in a box of 160 Mpc³ gridded in 256³ cells. The area of the box, 160² deg², approximately corresponds to 1 deg² at $z = 6.6$. As the model of ionizing sources, we employ the result of a radiative hydrodynamic (RHD) simulation (Hasegawa & Semelin 2013; Hasegawa et al. 2016). The RHD simulation regulates the star formation by taking into account the UV feedback and the supernovae feedback. The clumping factor depends on the matter density and the neutral fraction. It is worth noting that the IGM ionization history of the model is consistent with the constraints by quasar spectra (Fan et al. 2006) and Thomson scattering optical depth to the CMB (Planck Collaboration et al. 2016).

We assume the HSC deep survey at $z=6.6$. However, in order to prepare a large number of training data set, we use the IGM data at 20 different redshifts in a range of $6.3 < z <$

¹ The applications of the machine learning technique, in particular artificial neural network (ANN), have been suggested for the analysis of the 21cm line at the EoR (e.g. Shimabukuro & Semelin 2017; Doussot et al. 2019; Gillet et al. 2018; Hassan et al. 2019a; La Plante & Ntampaka 2019; Kern et al. 2017; Jennings et al. 2018; Schmit & Pritchard 2018; Hassan et al. 2019b,a; Li et al. 2019; Shimabukuro et al. 2020)

7.4. For the calculation of the LAE distribution, we solve the Lyman- α radiative transfer using the IGM data and use the halo distribution at $z = 6.6$ to avoid the evolution of the halo mass function.

2.2 Semi-Numerical Model

We use a simple model calculating the reionization process with a semi-numerical scheme to compare the performance of the networks. The halo distribution and the matter density field are identical to those used in the RT model. The semi-numerical method used in this work is similar to that of 21cmFAST (Zahn et al. 2011; Mesinger et al. 2011), while the latest version of 21cmFAST (Park et al. 2019) adopts a more realistic model.

First, the number of ionizing photon emitted from a halo into IGM is modeled as

$$\dot{N}_{\text{ion}} = f_{\text{esc},c} \left(\frac{M_{\text{halo}}}{10^{10}} \right)^{\alpha_{\text{esc}}} \dot{N}_{\text{ion,int}}, \quad (2)$$

where $f_{\text{esc},c}$ is the escape fraction of ionizing photons, M_{halo} is the halo mass, α_{esc} controls a mass dependence of the $f_{\text{esc},c}$. The number of ionizing photon produced in a halo is given as

$$\dot{N}_{\text{ion,int}} \propto (10^{-0.58M_{\text{UV}}}), \quad (3)$$

where we convert M_{halo} to the UV magnitude M_{UV} based on Shimizu et al. (2014).

The ionization state is determined by comparing the number of ionizing photons and the number of neutral hydrogen atoms within a sphere of radius R centred at a cell. We start the comparison from $R = R_{\text{max}}$ and eventually decrease the R to the size of a cell. Here R_{max} is the maximum mean free path of ionizing photon (Mesinger et al. 2011). If the number of ionizing photons is larger than the number of neutral hydrogen atoms, the central cell of the sphere is regarded as ionized. In fact, we assume the neutral fraction is 10^{-4} , rather than exactly zero, inside the ionized bubble. We normalize $\dot{N}_{\text{ion,int}}$ so that the neutral fraction is 0.2 for the model with $f_{\text{esc},c} = 0.001$, $\alpha_{\text{esc}} = -1.0$ and $R_{\text{max}} = 30\text{Mpc}$ at $z=6.6$.

For preparing the training and test data sets, we vary the parameters in ranges of $f_{\text{esc},c} = [0.001, 0.5]$, $\alpha_{\text{esc}} = [-1.0, 0.0]$ and $R_{\text{max}} = [5, 50]$. We label the model with $-1.0 < \alpha_{\text{esc}} < -0.5$ as S_l and the model with $-0.5 < \alpha_{\text{esc}} < 0$ as S_h .

2.3 LAE model

The distribution of observable LAEs is obtained by solving 1D RT of Lyman- α photons using the IGM data described above. The model of LAEs is identical to the model used in Kubota et al. (2018); Yoshiura et al. (2018). The intrinsic luminosity of Lyman- α is based on the model of the RHD simulation and given as $L_{\alpha} = 10^{42} (M_{\text{halo}}/10^{10})^{1.1} [\text{erg/s}]$. We use the line profile model of Yajima et al. (2018), and derive the transmission rate by solving the optical depth through the IGM along the line of sight (LoS). We identify halo as LAEs if the luminosity is larger than $4.2 \times 10^{42} [\text{erg/s}]$ and $2.5 \times 10^{42} [\text{erg/s}]$ for the HSC deep and ultra-deep surveys, respectively.

The LAE model has free parameters such as the Lyman- α escape fraction f_{α} , the galactic wind velocity V_{out} , and the HI column density within the galaxy N_{HI} . In practice, the LAE parameters can be calibrated so that the LAE distribution is consistent with latest observations. Although the calibration of the LAE parameters depends on the model of reionization, we fix the LAE parameters which are consistent with that of Kubota et al. (2018) for simplicity: $f_{\alpha} = 0.3$, $V_{\text{out}} = 150 [\text{km/s}]$ and $N_{\text{HI}} = 10^{19} [\text{cm}^{-2}]$. We leave the calibration of the LAE model considering the reionization model for future works. It should be worth to note that Inoue et al. (2018) has proposed stochastic models of LAE, which can explain various statistical properties of LAEs measured by the HSC survey.

3 IMAGE TO IMAGE TRANSLATOR

In order to construct maps of the 21cm line and neutral fraction from the LAE distribution, we use a technique of GAN. In this section, we describe the methodology and training data set.

3.1 Generative Adversarial Networks

The GAN, originally developed in Goodfellow et al. (2014), consists of two networks, a generator and a discriminator. The generator aims to produce fake images which are not distinguishable from true images. The discriminator attempts to distinguish true images and fake images created by the generator. Updated GAN techniques have been applied not only in the context of reionization but also in cosmology. For example, previous works such as Mustafa et al. (2019); Zamudio-Fernandez et al. (2019) have succeeded in making independent images of weak lensing convergence maps and the cosmic neutral hydrogen distribution.

In Isola et al. (2016), they have modified the conditional GAN (cGAN) to the image-to-image translation problem, and the code has been made public². In this work, we use a port of the code³. Below we describe the methodology briefly. See Isola et al. (2016) for more details.

The generator G and discriminator D are chosen so that the objective \mathcal{L}_{pix} is minimized and maximized with respect to them, respectively,

$$(G, D) = \arg \min_G \max_D \mathcal{L}_{\text{pix}}. \quad (4)$$

The objective function is given as

$$\mathcal{L}_{\text{pix}} = \mathcal{L}_{\text{cGAN}}(G, D) + \lambda \mathcal{L}_{\text{L1}}(G) \quad (5)$$

where a hyper parameter λ works as the weight of the term. The first term is a general objective for the cGAN, and it is given by

$$\begin{aligned} \mathcal{L}_{\text{cGAN}}(G, D) &= \mathbb{E}_{x,y} [\log D(x, y)] \\ &+ \mathbb{E}_x [\log (1 - D(x, G(x)))], \end{aligned} \quad (6)$$

where $\mathbb{E}[a]$ represents the expectation value of a , x is an input image and y is a true image which is the genuine pair image of the x . The generator, G , makes a fake image, $G(x)$,

² <https://phillipi.github.io/pix2pix/>

³ <https://github.com/affinelayer/pix2pix-tensorflow>

from an input x . The discriminator, $D(x, z)$, judges if a given image z is y or a fake image $G(x)$ and returns a value from 0 (fake) to 1 (true).

In Isola et al. (2016), they have introduced L_1 as the second term in Eq. (5). The L_1 is given as

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y}(\|y - G(x)\|_1). \quad (7)$$

The term represents the difference of the generated fake image $G(x)$ and the true image y . By introducing the L_1 term, the generator is trained to generate images close to true images and to deceive the discriminator. In Isola et al. (2016), they have argued that the L_1 can reduce the artificial behavior of output images.

In our context, the input image x is an image of LAE distribution and y is the true image of the 21cm-line signal corresponding to the x . Both are calculated from our numerical simulation in this paper. The $G(x)$ is a fake image of 21cm-line signal generated from the LAE distribution x . The discriminator D is given a pair of images of the LAE distribution and 21cm-line map and judges if the latter is a true or fake image.

Next, we describe the architecture briefly. See a document of this implementation⁴ for details. The architecture of the G is so called encoder-to-decoder⁵. In the network, input LAE image is processed through 8 down sampling convolution layers and 8 up sampling de-convolution layers. The size of filter is 4×4 and the stride is 2. The discriminator consists of 5 down sampling convolution layers. The pair of images is down sampled through 4 layers and output of fifth layer is a 30×30 image which is used to judge the image to be true or fake.

3.2 Data Sets and Training

Here we describe the data sets used to train and test the network. They are obtained from the numerical simulations described in previous section.

First, because the pixel value of input and output images needs to be in a range of [0-255], we normalize the 21cm brightness temperature images by 20 [mK] and the LAE images by $L_{\max} = 2.1 \times 10^{43}$ [erg/s]. In order to improve the efficiency of training, we classify the images into three groups based on the average neutral fraction such as $x_{\text{HI}} < 0.1$, $0.1 < x_{\text{HI}} < 0.3$, and $0.3 < x_{\text{HI}} < 0.5$.

We label the data set as follows. We refer the RT model and semi-numerical model as H and S , respectively. For S models, we label the model with $-1.0 < \alpha_{\text{esc}} < -0.5$ and $-0.5 < \alpha_{\text{esc}} < 0$ as S_l and S_h , respectively. The 3 bins of the average neutral fraction of $x_{\text{HI}} < 0.1$, $0.1 < x_{\text{HI}} < 0.3$, and $0.3 < x_{\text{HI}} < 0.5$ are referred as 1, 2 and 3, respectively. Thus, for example, a network trained using the RT simulation data of the neutral fraction bin of $0.1 < x_{\text{HI}} < 0.3$ is referred as

model	H_1	H_2	H_3
N_{train}	188	100	76
N_{test}	92	52	40

model	$S_{l,2}$	$S_{l,3}$	$S_{h,2}$	$S_{h,3}$
N_{train}	140	203	267	157
N_{test}	68	116	133	72

Table 1. The number of the training data sets N_{train} and the test data sets N_{test} . The RT simulation and the semi-numerical models are labeled as H and S , respectively. The subscript i ($i = 1, 2$ and 3) indicates the neutral fraction ($x_{\text{HI}} < 0.1$, $0.1 < x_{\text{HI}} < 0.3$, and $0.3 < x_{\text{HI}} < 0.5$, respectively). In the semi-numerical model, l and h indicate the power-law index of the escape fraction is lower and higher than -0.5 , respectively.

“ H_2 ”. Note that the HSC deep survey is assumed in most cases and we add UD when the HSC ultra-deep survey is assumed such as “ $S_{l,2}\text{UD}$ ”.

We need a large number of training data sets to train the network effectively. However, we have only one realization of RT simulation due to its high computational. Thus, we increase the number of quasi-independent data sets from 1 simulation box as follows. Because the redshift uncertainty of the Lyman- α survey of the Subaru HSC is $\Delta z = 0.1$ which roughly corresponds 40 Mpc, we divide the simulation box into 4 slices of $(160 \text{ Mpc})^2 \times 40 \text{ Mpc}$. By doing this along the 3 axes of the box, we obtain 12 slices which correspond to the sky area of 1 deg^2 and the redshift width of 0.1. Further, we randomly shift the center of the maps and create quasi-independent images to obtain 500 and 100 images as training and test data, respectively.

If the training data is identical to the test data, the network is overtrained and loses the versatility. To avoid this, we use slices divided along x and y directions as training data and those divided along z direction as test data. Thus, the test data and the training data are independent. We list the number of original training and test images in table. 1.

Hyper parameters such as λ , the number of training iteration and the batch size should be chosen by the convergence of the training so that the network produces images correctly from the input LAE distribution. We performed training with combinations of λ of 1, 10 and 100 and the batch size of 32 and 64 and chose $\lambda = 100$ and batch size of 64 as our parameters since the evolution of the objective function is stable for all training data sets. Furthermore, in order to avoid the over fitting, we stop training at 300 iterations.

4 STATISTICAL ANALYSIS

Here, we describe the data analysis method used in this work. For the evaluation of an accuracy of our network, we calculate the correlation coefficient in real space, which is given as

$$r = \frac{\sum_i^N (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i^N (a_i - \bar{a})^2} \sqrt{\sum_i^N (b_i - \bar{b})^2}}, \quad (8)$$

where N is the number of pixels, a_i is i th pixel value of output images, b_i is i th pixel value of true images and \bar{a} represents the averaged value.

⁴ <https://affinelayer.com/pix2pix/>

⁵ In original literature (Isola et al. 2016), a U-Net structure (Ronneberger et al. 2015) is employed to transfer the shape of the input image to the output image efficiently. The U-net structure propagates the feature map from i th layer to $(n-i)$ th layer where the feature map is an output tensor filtered at a certain layer and n is the total number of layer. We note that our results in later section show acceptable quality without the U-Net structure. The introducing the U-Net might improve the result.

A major tool to study the 21cm line is the auto power spectrum, which is given by

$$P(|\mathbf{k}|) = (2\pi)^2 \delta_D(\mathbf{k} - \mathbf{k}') \langle \delta(\mathbf{k}) \delta(\mathbf{k}') \rangle, \quad (9)$$

where \mathbf{k} is wavenumber in 2D Fourier space, δ_D is Dirac's delta function, δ is the fluctuation in 2D Fourier space.

We also use the 21cm-LAE correlation coefficient which is given as,

$$r_k^{21, \text{LAE}}(k) = \frac{P_{21, \text{LAE}}(k)}{\sqrt{P_{21}(k) P_{\text{LAE}}(k)}}, \quad (10)$$

where P_{21} is the 21cm power spectrum and P_{LAE} is the power spectrum of LAE distribution. The 21cm-LAE cross power spectrum is given as,

$$P_{21, \text{LAE}}(|\mathbf{k}|) = (2\pi)^2 \delta_D(\mathbf{k} - \mathbf{k}') \langle \delta(\mathbf{k}) \delta_{\text{LAE}}(\mathbf{k}') \rangle, \quad (11)$$

where δ_{LAE} is the fluctuation of the LAE distribution.

In this work, we attempt to calculate the cross correlation between the true 21cm-line image and the output 21cm-line image, which is given as $P_X^{T, O}(|\mathbf{k}|) = (2\pi)^2 \delta_D(\mathbf{k} - \mathbf{k}') \langle \delta_{\text{true}}(\mathbf{k}) \delta_{\text{out}}(\mathbf{k}') \rangle$, where δ_{true} and δ_{out} are the fluctuation in the true image and the output image, respectively. The correlation coefficient is given as

$$r_k^{T, O}(k) = \frac{P_X^{T, O}(k)}{\sqrt{P_{\text{true}}(k)} \sqrt{P_{\text{out}}(k)}}, \quad (12)$$

where P_{true} and P_{out} indicates the auto power spectrum of the true image and the output image, respectively.

Ultimately, in the future work, the output image is used for the cross correlation with the observed 21cm line. In this work, we assume true images as the observed 21cm line and estimate the detectability of the cross power spectrum. We evaluate the error which is written as,

$$2 \quad \sigma_X^2(|\mathbf{k}|) = \frac{1}{\epsilon \Delta k k S_{\text{area}} / 2\pi} \times (P_X^2(|\mathbf{k}|) + (P_{\text{obs}}(|\mathbf{k}|) + N_{\text{obs}}(|\mathbf{k}|))(P_{\text{out}}(|\mathbf{k}|) + N_{\text{pix}}(|\mathbf{k}|))), \quad (13)$$

where P_X is the cross power spectrum between the observed 21cm line image and the output image and P_{obs} is the power spectrum of observed 21cm line image. The number of sample in k bin is $\epsilon \Delta k k S_{\text{area}} / 2\pi$ where Δk is the width of k bin and S_{area} is the area of image. In practice, we replace the term by the number of \mathbf{k} samples used in power spectrum calculation. N_{obs} is thermal noise, and N_{pix} is the error due to the network.

For the 21cm line observation, we assume an observation by the MWA Phase II with compact array, and the thermal noise can be estimated by

$$N_{\text{obs}}(\mathbf{k}) = \left(\frac{\lambda_o^2}{A_e} \right) \left(\frac{T_{\text{sys}}}{\sqrt{Bt}} \right)^2 \frac{D_M^2(z)}{n(\mathbf{k})}, \quad (14)$$

where $n(\mathbf{k})$ is the number of baseline in a k bin, λ_o is the observed wavelength, $A_e = 20[\text{m}^2]$ is the effective area of antenna, $T_{\text{sys}} = 255 [\text{K}]$ is the system temperature, $B = 2.5\text{MHz}$ is the band width roughly corresponding to the HSC redshift error, t is the observation time and $D_M(z)$ is the comoving distance to z . For estimation of baseline distribution $n(\mathbf{k})$, we use the array distribution of MWA⁶.

⁶ <http://www.mwatelescope.org/telescope/configurations/phase-ii>

The error on the cross power spectrum can be reduced by adding independent sample images. Thus, the error is evaluated as σ_X / \sqrt{N} , where N is the number of samples. For example, since the total field of view of the HSC deep field survey is 27deg^2 and the size of image is approximately 1deg^2 in this work, we can use 27 samples to calculate the cross correlation.

5 RESULTS

We here show results of the 21cm-line image prediction by our cGAN network, the power spectrum and the 21cm-LAE cross power spectrum. We have 7 training data sets (i.e. H_1 , H_2 , H_3 , $S_{l,2}$, $S_{l,3}$, $S_{h,2}$ and $S_{h,3}$) for deep and ultra deep surveys. Thus, 14 different networks are developed.

The structure of this section is as follows. We first check the property of data used for training in section 5.1. Next, we compare the predicted images, their auto-power spectra and cross correlation coefficients with those of true images in sections 5.2 and 5.3. In section 5.4, in order to see the robustness of our method, we attempt to predict 21cm-line images of H_2 model using networks trained with data sets of other models. Finally, the detectability of cross-power spectrum between the observed 21cm-line image and predicted image is discussed in section 5.5, and we attempt to estimate 21cm-line auto-power spectrum using the cross-power spectrum in 5.6. Note that the true model is identical to the model used for training network in this section except in 5.4.

5.1 Training Data Property

Before showing results, it would be useful to show the 21cm-line auto-power spectrum and the 21cm-LAE correlation coefficient of the test data sets. Here it should be noted that the statistical property of training and test data is the same.

In upper panels of Fig. 1, 21cm-line auto-power spectra, $\Delta_{21} \propto k^2 P_{21}$, of 7 training data sets are shown with the standard deviation. As can be seen, they have a peak at $k = 0.2 \text{ Mpc}^{-1}$ in all models. Although the power spectra of many models are consistent with each other within sample variance, they show a weak dependence on the model parameters. Since the amplitude of power spectrum is proportional to x_{HI}^2 , those of high- x_{HI} models (H_3 , $S_{l,3}$ and $S_{h,3}$) are systematically larger than those of low- x_{HI} models. Besides, although S_l models is consistent with S_h models at all scales within sample variance, S_l models have slightly lower power than S_h models at large scale. This indicates that the lower α_{esc} value reduces the ionization contribution from heavier galaxies and mitigates fluctuations due to large ionized bubbles.

The lower panels of Fig. 1 show the 21cm-LAE cross correlation coefficient, $r_k^{T, O}$. As we indicated before, 21cm-line signal and the LAE distribution have a negative correlation at large scales because the inside-out reionization is driven by galaxies in all models. Furthermore, at small scales, the 21cm line correlates positively with LAEs in the H model. This positive correlation is caused by partially neutral regions within the ionized bubbles as shown in Kubota et al. (2018). On the other hand, in the S models, the neutral fraction at ionized bubbles is completely zero. Thus, the 21cm line does not correlate with LAEs at $k > 1 \text{ Mpc}^{-1}$. Although

the sample variance is rather large, we can see a tendency that the cross correlation is weaker for lower- x_{HI} models.

5.2 Image Reconstruction

Fig. 2 shows examples of the input LAE distribution, the true 21cm-line image and the predicted 21cm-line image for $S_{h,3}$ model. Here, the predicted image is generated by a network trained with data of $S_{h,3}$ model, that is, the correct model. The neutral regions tend to lie far from LAEs in the predicted image, as they should be, and the large-scale feature is very similar between the true and predicted images. However, the small-scale structure is not reproduced well. This will be due to the weak correlation between LAE distribution and the small-scale fluctuations in 21cm line, as we saw in bottom panels of Fig. 1.

In Table 2, we summarize the average correlation coefficient between true and predicted images, r , and the average number of LAEs for all 14 models. The value of r of H models is generally less than 0.3, while it is $0.4 \sim 0.5$ for S models. This would be due to the larger number of training data sets (see Table 1), the larger number of LAEs and the simple distribution of the neutral fraction. It should be noted that the r is a measure of overall similarity between the two images. In fact, as we saw in Fig. 2, the large-scale feature is reproduced better than the small-scale feature.

To see this quantitatively, in Fig. 3, we show the scale-dependent correlation coefficient between the two images, $r_k^{T,O}$, defined in Eq. (12). The value of $r_k^{T,O}$ is about 0.4 for H_1 and H_3 models, and 0.6 for H_2 model at the largest scale (left panel), while it is larger than 0.7 for S_h models (right panel) and S_l models (not shown). On the other hand, the correlation is close to zero at small scales ($k > 0.2 \text{ Mpc}^{-1}$) for all models. Nevertheless, looking at the small scales more carefully, $r_k^{T,O}$ for H_1 and H_2 models is non-zero (~ 0.1) at $k \sim 1 \text{ Mpc}^{-1}$, while it is consistent with zero for S_h models. As we saw in Fig. 1, in H models, $r_k^{21,\text{LAE}}$ is also positive at small scales and this is caused from the correlation between the LAE distribution and residual HI. Thus, our network learns the correlation between the 21cm line and LAEs at both large and small scales correctly.

Next, we consider the relation between the quality of reproduction and the number of LAEs. As indicated before, the neutral fraction is not zero even in ionized bubbles for H models due to the recombination and ionization equilibrium so that neutral hydrogen can survive even around large halos. Therefore, although the averaged neutral fraction of H_i and S_i models ($i = 2$ and 3) are almost the same and they are based on the same halo distribution, there is a large difference in the number of LAEs.

The number of LAEs increases when we assume the ultra deep survey, and the r for the ultra deep survey is larger than that of deep survey as listed in Table 2. In Fig. 4, we show the $r_k^{T,O}$ between true and output images for H_2 and H_2, UD models. It improves from 0.6 to 0.8 with an assumption of the ultra deep survey. These facts indicate that, as the number of LAEs increases, the relation between the LAE and 21cm-line signal is learned more accurately.

However the situation is different for H_1 model. Although a large number of LAEs are detectable in H_1 model, the similarity between the true and predicted images is weaker than other models as shown in table 2 and Fig. 3.

In fact, in H_1 model, the average neutral fraction is very low (< 0.1) so that only a few neutral islands exist and the 21cm-LAE correlation is weak. This is the reason why the reconstruction of the 21cm-line image of H_1 model is difficult.

5.3 Statistical Property

We now investigate statistical metrics such as the 21cm power spectrum P_k and the 21cm-LAE cross correlation coefficient $r_k^{21,\text{LAE}}$ of predicted images and compare them with those of true images.

In the upper panels of Fig. 5, we show the 21cm-line auto-power spectrum of predicted and true images of H_2 and $S_{h,2}$ models. Interestingly, their power spectra are consistent within sample variance even at small scales, although, as we saw in the previous subsection, the correlation between true and predicted images is poor at small scales. This result indicates that the network succeeds to learn the amplitude of 21cm-line fluctuations at all scales but fails to learn the phase of fluctuations at small scales.

The bottom panels of Fig. 5 show the correlation coefficient between the 21cm-line map and the LAE distribution, $r_k^{21,\text{LAE}}$. As with the case of the 21cm power spectrum, predicted images are consistent with the true images within sample variance. The P_k and $r_k^{21,\text{LAE}}$ of other models, which are not shown here, have the same accuracy as Fig. 5. Therefore, our network can predict 21cm-line images which are statistically consistent with true images.

5.4 Cross Test

So far, we checked the quality of the networks which were trained with the correct models. In fact, we cannot know the correct model a priori. Thus, in practice, we propose to take a cross correlation between the observed data and predicted 21cm-line images of various models, and find the model which has the strongest correlation. To see how this works, we here calculate the cross correlation between 21cm-line images of H_2 model and predicted images obtained by networks trained by correct and different models.

In Fig. 6, we show the cross correlation between true images of H_2 model and predicted images of other models, $r_k^{H_2,O}$, with a shade representing the sample variance calculated as σ/\sqrt{N} where σ is the standard deviation, N is the number of samples and we assume $N = 25$ considering the HSC deep survey.

As we see in the left panel, even for H models with different neutral fraction (H_1 and H_3 models), the correlation coefficients between the predicted images and the true image of H_2 model is relatively high (~ 0.4 for H_1 model and ~ 0.6 for H_3) at large scales. The correlation coefficient is the highest for H_2 model and it is just slightly larger than that for H_3 model. Interestingly, the predicted images by S models shown in the right panel correlate even stronger (~ 0.8) with H_2 -model image at large scales.

Relatively strong correlation for all models at large scales is due to the fact that they are all based on the inside-out reionization scenario. Thus, on the one hand it is not easy to measure neutral fraction and/or distinguish models, but on the other hand the detectability of 21cm-line signal

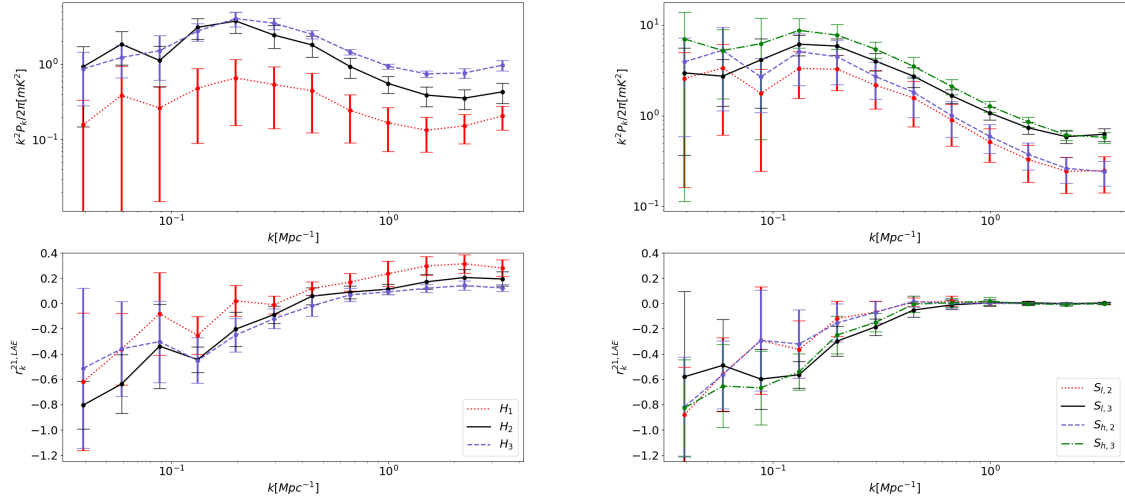


Figure 1. Comparison of the properties of the test data sets. Top row shows the 21cm-line auto-power spectrum, and bottom row shows the correlation coefficient between the 21cm-line signal and LAE distribution. In the left column, solid line and dashed line represent H_1 , H_2 and H_3 models, respectively. In the right column, dotted line, solid line, dashed line and dot-dashed line are $S_{l,2}$, $S_{l,3}$, $S_{h,2}$ and $S_{h,3}$ models, respectively. The error bars represent the sample variance.

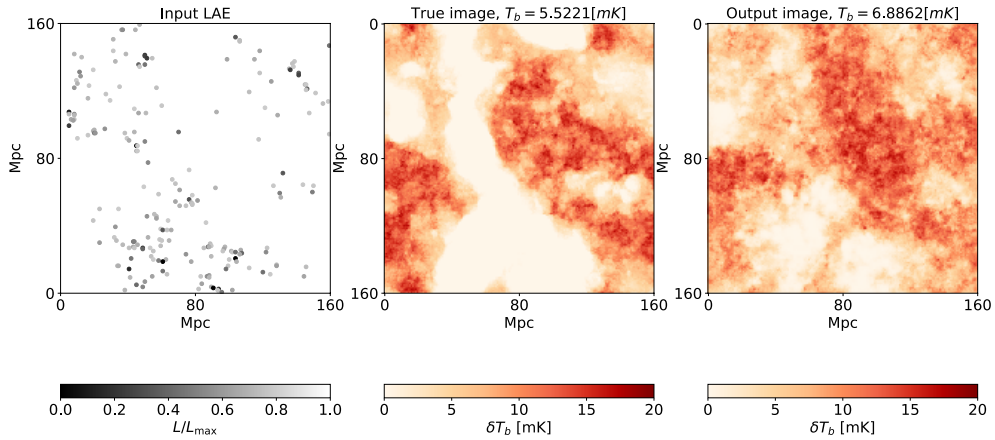


Figure 2. Example of the 21cm-line image reconstruction in $S_{h,3}$ model. The panels show the input LAE distribution, the true 21cm-line image, and the output (predicted) image from left to right. In the true image, neutral hydrogen is located far from LAEs and this feature is well learned in the output image. The output image resembles the true image at large scales, the small-scale structure is not well reconstructed.

by this method does not depend on the detail of the model used to create training data.

Here it should be noted that the correlation, $r_k^{H_2, O}$, for H models is positive at small scales ($k > 1 \text{ Mpc}^{-1}$) while it vanishes for S models. Thus, observations at small scale may be useful to identify the true model.

5.5 Detectability

We evaluate the detectability of the cross power spectrum between the observed 21cm line and output images assuming the MWA Phase II compact array observation and the HSC deep survey. The error is calculated using Eq. (14). The total survey area is 25 deg^2 , and we assume the integration time of the MWA is 100 hours per pointing. We note that the field-of-view of MWA at $z = 6.6$ is much larger than the

model	H_1	H_2	H_3	H_1, UD	H_2, UD	H_3, UD
r	0.14 ± 0.094	0.27 ± 0.079	0.14 ± 0.066	0.22 ± 0.092	0.37 ± 0.061	0.32 ± 0.060
N_{LAE}	209	167	72	586	483	231
model	$S_{l,2}$		$S_{l,3}$	$S_{l,2}, \text{UD}$		$S_{l,3}, \text{UD}$
r	0.43 ± 0.13		0.40 ± 0.095	0.48 ± 0.14		0.54 ± 0.073
N_{LAE}	221		163	558		436
model	$S_{h,2}$		$S_{h,3}$	$S_{h,2}, \text{UD}$		$S_{h,3}, \text{UD}$
r	0.50 ± 0.16		0.50 ± 0.10	0.62 ± 0.14		0.60 ± 0.11
N_{LAE}	218		181	558		468

Table 2. Correlation coefficient, r , between the true 21cm-line image and the output (predicted) 21cm-line image. The average number of input LAEs per image is also listed. For the model named with UD, we assume the ultra deep survey of the HSC, and then the number of LAEs drastically increases and the cross correlation becomes strong.

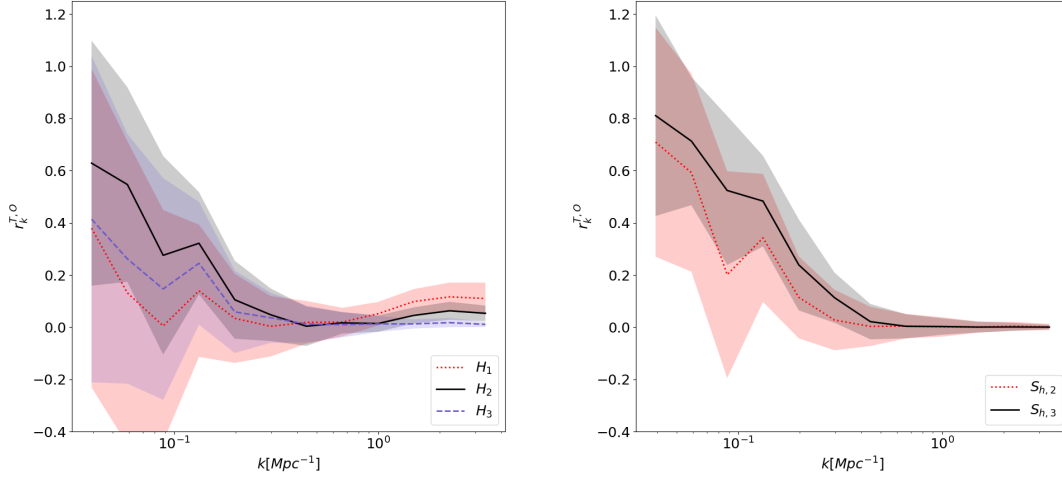


Figure 3. Correlation coefficient between the true and output 21cm-line images. Left and right panels are the result of (H_1, H_2, H_3) and $(S_{2,h}, S_{3,h})$ models. The shade shows the sample variance. The cross correlation is relatively strong at large scales ($k < 0.1 \text{ Mpc}^{-1}$).

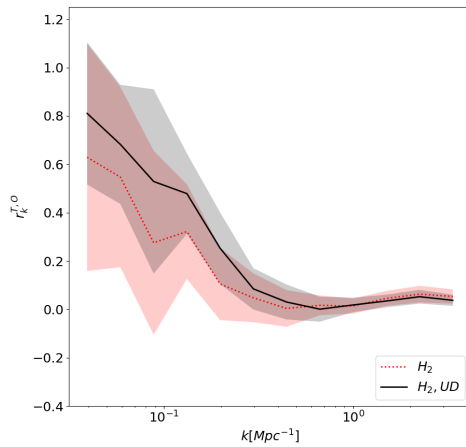


Figure 4. Same as Fig. 3 but for H_2 (dotted) and H_2, UD (solid) models. The ultra deep observation increases the number of LAEs and improves the correlation between the true and output 21cm-line images.

HSC deep field. However, in practice, the HSC deep field is separated into 4 different patches of sky, and therefore at least 4 pointings (400 hours) of the MWA observation is required in total. Note that the model used for training is identical to the true model in this analysis.

Fig. 7 shows the cross power spectrum between observed and predicted 21cm-line images for H_2 and $S_{h,2}$ cases. The shaded region represents the observational error on the cross power spectrum. The cross power spectrum is detectable at $k < 0.2 \text{ Mpc}^{-1}$ for both models. The cumulative signal-to-noise ratio (SNR), which is calculated as $\text{SNR}^2 = \sum_k (P_X(k)/\sigma_X(k))^2$, exceeds 4.

The cross power spectrum cannot be detected at small scales since the MWA phase II with compact array does not have long baselines sufficient for resolving small-scale fluctuations. The detection at small scales might be possible by the SKA1-Low, although the cross power spectrum is weak at small scales.

In practice, the foreground would be the largest source of error. The foreground power spectrum is more than 8 orders of magnitude larger than the 21cm auto-power spectrum at $k \sim 0.1 \text{ Mpc}^{-1}$. The foreground term enhances the error at all scales, and the detection is not possible with-

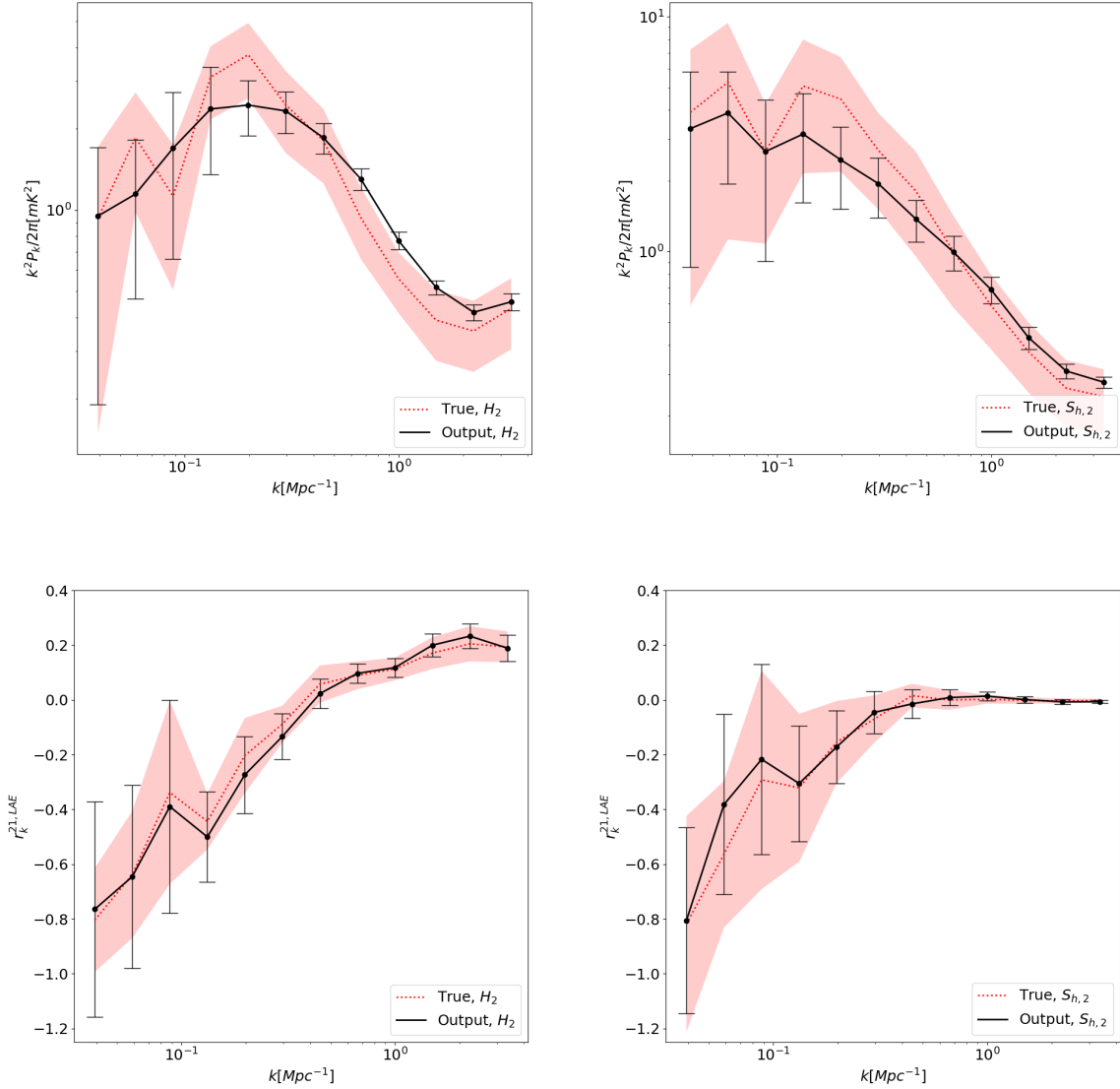


Figure 5. Top and bottom panels represent the 21cm-line auto-power spectra and the cross correlation coefficient between the 21cm-line signal and LAEs. Left and right panels show the comparison of true (dotted) and output (solid) images for H_2 and $S_{h,2}$ models, respectively. Shade and error bars represent the sample variance evaluated from the true and output samples, respectively. The network successes to generate output images with the 21cm-line auto-power spectrum and 21cm-LAE cross power spectrum which are consistent with those of the true images. In other models not shown here, the quality of the networks is similar.

out foreground removal. The foreground avoidance method, based on spectral smoothness of synchrotron emission, is not applicable in this work since we only use spacial information. If we extend our method to 3 dimensional map, the foreground avoidance method is applicable and the cGAN construction can be more useful. To achieve this, we need to identify the redshift of LAEs with a high precision, and a spectroscopic observation is required.

We also mention that the foreground contamination should be reduced by integrating independent samples since the foregrounds contribute only to the statistical variance but not to the average. Note that the error due to foregrounds cannot be reduced in the same way in the case of the 21cm-line auto-power spectrum. Currently, the most serious

obstacle of 21cm-line auto-power spectrum measurement is the contamination from the combination of foregrounds and various systematics due to instruments and analysis. The cross correlation technique has a potential to mitigate such systematic errors and should be essential to validate the detection of the 21cm line in future observation.

5.6 Estimation of Error in 21cm-line Auto-Power Spectrum

As discussed before, because we don't know the correct reionization model a priori, we need to seek a model which has the highest correlation with observed 21cm-line images. Once the best model is selected, the 21cm-line auto-power

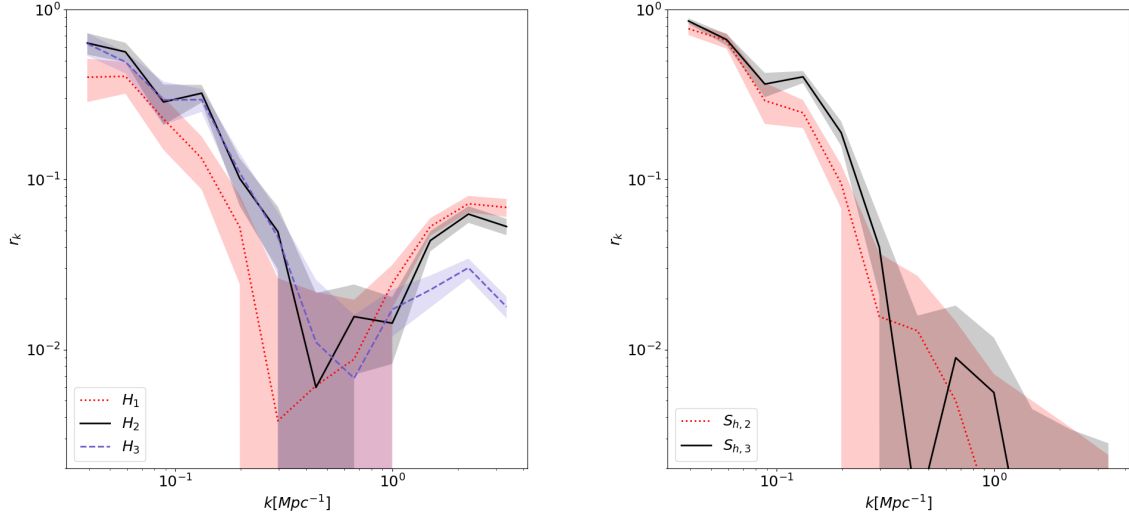


Figure 6. Correlation coefficient between the predicted images of H_2 and other models. The output images are generated from the input LAE distribution of H_2 model by the networks trained by 21cm-line images of other models. Shades show the sample variance, and the survey area of 25 deg^2 is assumed. It is seen that the cross correlation is measurable at large scales even if the model of network is not correct.

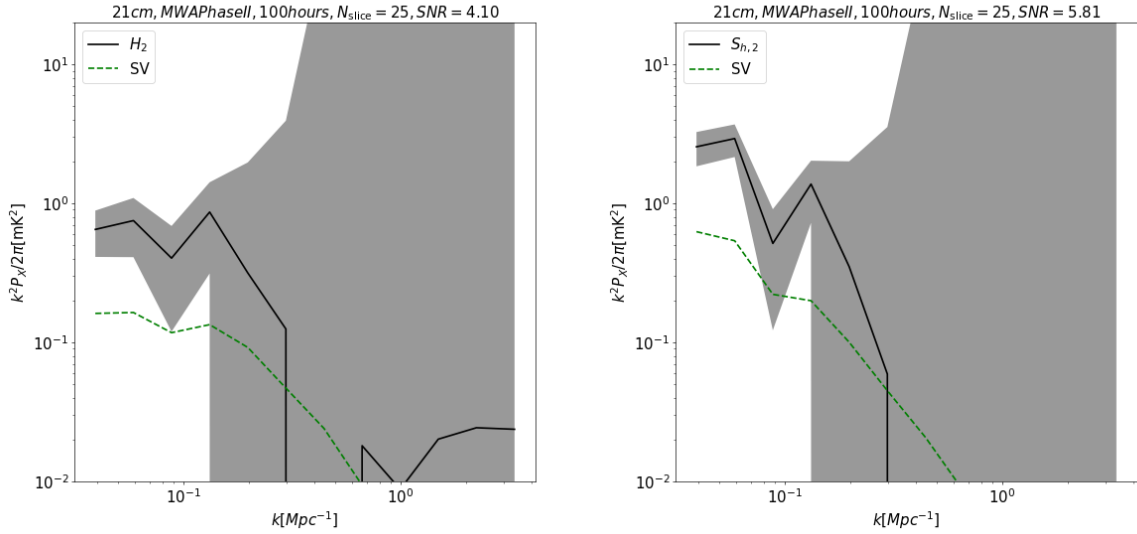


Figure 7. Detectability of the cross power spectrum between observed 21cm line maps and output images in the H_2 and the $S_{h,2}$ cases. Observation of HSC deep survey with the FoV of 25 deg^2 and the MWA Phase II observation of 100 hours per slice are assumed. Solid line is the cross power spectrum, the dashed line is expected sample variance error (SV) and the shade is the error calculated from Eq. (14) including the thermal noise of the MWA. The cross correlation is observable at large scales ($k < 0.2 \text{ Mpc}^{-1}$). The cumulative signal to noise ratio is larger than 4 for both models.

spectrum can be calculated with the predicted images of the selected model as demonstrated in Fig. 5. However, since the cross power spectrum could be measured only at large scales, the uncertainty at small scales should be properly propagated to the estimation of auto-power spectrum.

It is possible to estimate the error in 21cm-line auto-power spectrum as follows, using the measured cross power spectrum between the observed 21cm-line and predicted im-

ages. First, 21cm-line auto-power spectrum can be written as,

$$P_{21}^{1/2}(k) = \frac{P_X(k)}{r_k(k) P_{21,\text{pre}}^{1/2}(k)}. \quad (15)$$

Therefore, by using a model of r_k denoted as \tilde{r}_k , we can estimate P_{21} from the observed cross power spectrum and auto-power spectrum of predicted image. We assume that

the \tilde{r}_k is empirically obtained from simulations with a statistical error of Δr_k which is replaced by the sample variance of $r_k^{T,O}$. Then the error on the estimated power spectrum can be described as

$$\sigma^2 \approx \left(\frac{\sigma_X^2}{P_X^2} + \frac{\Delta^2 r_k}{\tilde{r}_k^2} + \frac{\Delta^2 P_{\text{out}}}{4P_{\text{out}}^2} \right) \frac{P_X}{\tilde{r}_k P_{\text{out}}^{1/2}}, \quad (16)$$

where σ_X is evaluated in Eq. (14). The error on the 21cm-line auto-power spectrum of the predicted image $\Delta^2 P_{\text{out}}$ is given as the sample variance of P_{out} which is shown in Fig. 5. We note that the total error is also reduced by $1/\sqrt{N}$, and $N = 25$ represents the assumed survey area in deg^2 .

Fig. 8 shows the 21cm power spectrum estimated by this method. The left and right panels are the results of H_2 and $S_{h,2}$ models. The error is large at small scales due to lack of sensitivity of MWA Phase II.

The HSC ultra deep survey can improve the correlation coefficient, and the error on the power spectrum prediction is smaller than that of the deep survey as shown in Fig. 8. However, the improvement is not drastic, and the assumption of field-of-view ($\sim 25 \text{ deg}^2$) for ultra deep survey may be too optimistic.

6 DISCUSSION

6.1 Improving Network

There are some strategies to improve the efficiency and the accuracy of the network, which we summarise here:

(i) *Calibration of training model* - Since we do not know the true model a priori, we need to prepare a large number of networks trained by various models. We are able to reduce the number of the network by model calibration using other observed quantities, and then the learning can be easier than that of this work. For example, we can remove models which are not consistent with observations such as statistical property of LAEs, constraints on the neutral fraction as a function of redshift and Thomson optical depth of cosmic microwave background photons.

(ii) *Increase the number of independent training data sets* - In this article, we used only one realization of the N-body simulation for the training and test data sets. Furthermore, to avoid training data becomes identical to test data, we used images integrated along x and y directions as training data, and the image integrated along z direction is used as the test data. As a result, the number of training sets is quite smaller than other works applying CNN to 21cm line maps of 1000 training data (e.g. Gillet et al. (2018); Hassan et al. (2019a); La Plante & Ntampaka (2019)). Thus, the accuracy of our estimation would be improved if we could increase the number of original training data sets as many as previous studies.

(iii) *Optimizing hyper parameters* - The cGAN method has some hyper parameters such as the weight λ and the number of training epochs which can affect the quality of image reconstruction. Although the hyper parameters in this work have been chosen from some combinations of values so that the objective function is stable and the output images do not show artificial structures, optimization by comparing a larger number of hyper parameter sets can improve the accuracy of the network.

(iv) *Using alternative architecture* - Other network architecture might improve the construction such as the U-net used in Isola et al. (2016). Additionally, the accuracy of the construction can be improved using additional input information such as the UV luminosity of the LAE.

We leave these improvements for future works.

6.2 Comparison With 21cm-LAE Cross Correlation

Here, we compare the detectability of 21cm-line signal between the conventional 21cm-LAE cross power spectrum and our new cross correlation introduced in this work. As discussed in our previous works (Kubota et al. 2018; Yoshiura et al. 2018), 1,000 hours of MWA Phase I observation is required for detecting the 21cm-LAE cross power spectrum with $\text{SNR} \sim 4$. Although only 100 hours of observation is required to achieve the $\text{SNR} \sim 4$ for the new cross correlation method, we assumed the MWA Phase II with the redundant array which improves the sensitivity roughly 4 times at $k \sim 0.1 \text{ Mpc}^{-1}$ (Wayth et al. 2018) and we employed a model of thermal noise lower than that used in our previous works. Therefore, if we assume the identical observation specification, the detectability of is almost comparable.

However, it should be noted that, as discussed above, the accuracy of prediction can be boosted by improving the neural network architecture. If a future improved network can enhance the correlation between the true and predicted images at all scales, the detectability of our new cross correlation would become better than the conventional 21cm-LAE cross correlation.

Another advantage of our new cross-correlation method is that the 21cm-line map and its auto-power spectrum can be estimated without a model uncertainty. In case of the conventional 21cm-LAE cross correlation, we can construct a model of reionization and LAEs which produces a cross-power spectrum consistent with the observation data. Then, the model can predict a 21cm-line map and its auto-power spectrum. However, the observed cross-power spectrum cannot narrow down a model and the predicted auto-power spectrum has a model uncertainty.

On the other hand, in the current case, the network which generates the 21cm-line map is trained by simulation data sets based on a specific model of reionization and LAEs. However, once the generated 21cm-line map has a correlation with the observed data, the generated map reflects the true 21cm-line map independently of the model used to train the network. Thus, the estimation of the 21cm-line map and statistics are more direct for the current method.

6.3 Neutral Fraction Map

We have discussed the image prediction of the 21cm-line brightness temperature. However, the cGAN can predict the neutral fraction map, which cannot be measured directly via the 21cm line. The cross correlation between the observed 21cm line and the reconstructed neutral fraction images can be useful to extract the information of ionized region from the observation of 21cm line, and it will effectively reveal the property of ionizing sources.

Fig. 9 is an example of the reconstruction of a neutral

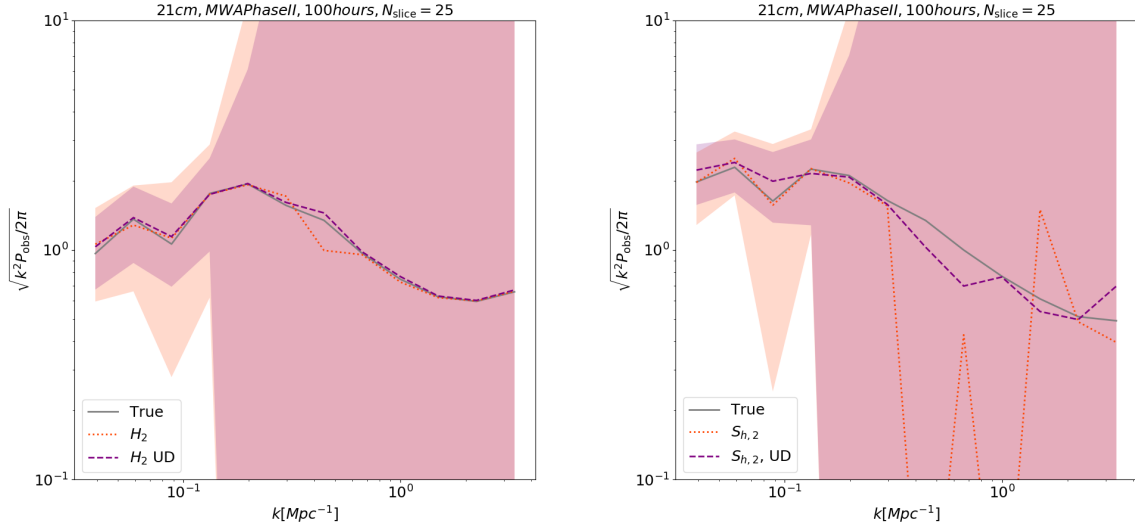


Figure 8. Prediction of the 21cm power spectrum from the cross power spectrum in the H_2 and the $S_{h,2}$ model is shown in left and right panels. Same as Fig. 16, we assume 100 hours of MWA observation per slice and $N = 25$. The solid line is the true 21cm power spectrum, the dotted line is the predicted power spectrum and the dashed line is the predicted power spectrum with the assumption of the HSC ultra deep survey. Light colored shade region indicates the error in the case with the HSC deep survey, and the darker shade is that of the ultra deep survey. Here, we use averaged $r_k^{T,O}$ as an empirical correlation coefficient, and therefore the true power spectrum is well consistent with the predicted spectrum. Note that The survey area of 25deg^2 might be over assumption for the ultra deep survey.

fraction map. As in the case of 21cm line, from the distribution of LAEs, the network can predict a map of the neutral fraction map which resembles to the true one.

Table. 3 shows the correlation coefficient between the true and predicted neutral hydrogen maps, r , its standard deviation and the average number of LAEs, N_{LAE} . The result is similar to that of 21cm line which is summarized in table. 2.

Here it should be noted that the prediction of averaged x_{HI} was not accurate in this work. The averaged x_{HI} of output images tends to be the mean value of training data set. For example, the output neutral fraction of the $S_{h,3}$ model is approximately 0.4 regardless of input LAEs. The prediction of the averaged value might require to add the averaged x_{HI} in the optimizing function instead of L_1 term.

6.4 Spherical Bubble Model

In the model of the inside-out reionization, ionized bubbles are generated around galaxies. Although the size and the shape of bubbles depend on the source property and the density of the neutral hydrogen distribution, a spherical bubble (SB) model, in which we assume ionized bubbles can be replaced with spheres around ionizing sources, is reasonable model at large scales. Thus, the SB model is useful for estimation of the map of neutral fraction from distribution of ionizing sources. Note that the SB model might be available for the 21cm line, but the fluctuation of the matter density cannot be taken into account. Thus, here, we use the SB model only for the neutral fraction map.

Instead of the cGAN network, we can use the SB model to predict the distribution of the neutral fraction. For the H_2 , UD model, we make ionized regions with the radius R

around input LAEs and assume the IGM outside of bubbles is neutral. An example image of LAE distribution, true image, output image and image of the SB model is shown in Fig. 10.

The radius R is a free parameter in the SB model, and we choose $R = 16$ Mpc so that the r_k between the true image and cGAN output image is consistent with that of the true image and the SB model as shown in left panel of Fig. 11. However, if we calibrate the SB model based on the auto power spectrum, R of 22 Mpc is chosen. Since the R is highly ambiguous, the SB model is too simple to predict the neutral fraction map, and the cGAN method can be more useful than the SB model.

7 SUMMARY

In this work, we proposed a new approach to detect EoR 21cm-line signal, which measures cross correlation between observed 21cm-line map and a 21cm-line map predicted from the LAE distribution through machine learning. Specifically, we applied the conditional GAN method to the translation of an LAE distribution map to a 21cm-line map. The network is trained using LAE distribution maps, 21cm-line brightness temperature maps and neutral fraction maps obtained by the RT numerical simulation (H models) and the semi-numerical simulation (S models). For preparing the LAE distribution, we solved the Lyman- α radiative transfer using the IGM data and models of the line profile.

The network can reproduce the maps of the 21cm-line brightness temperature from the LAE distribution. The accuracy depends on the training model and the average neutral fraction. As a metric to assess the quality of network, we have calculated the cross correlation coefficient between the

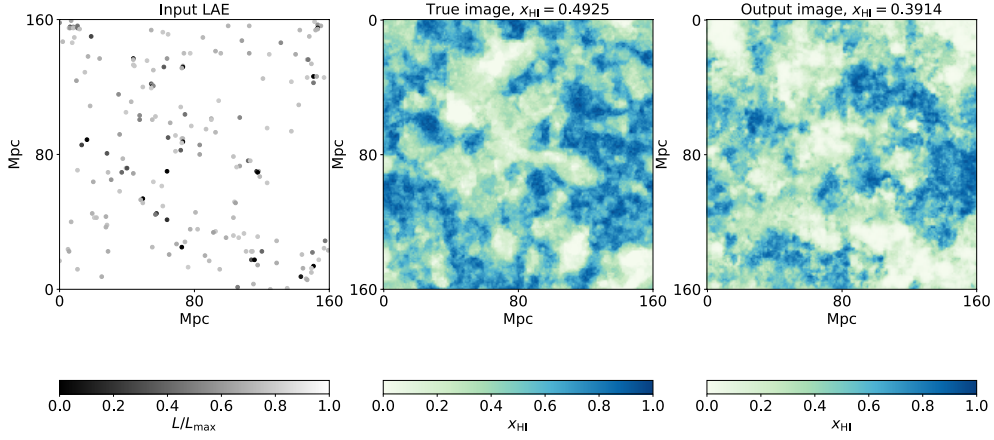


Figure 9. Example of the reconstruction for the neutral fraction in the $S_{h,3}$ model. The panels show the input LAE distribution, the true image, and the output image from left to right. Same as 2, the large scale structure of the output image is comparable to the true image.

model	H_1	H_2	H_3	H_1, UD	H_2, UD	H_3, UD
r	0.14 ± 0.11	0.28 ± 0.077	0.18 ± 0.071	0.18 ± 0.11	0.40 ± 0.071	0.33 ± 0.061
N_{LAE}	214	169	73	603	482	233
model		$S_{l,2}$	$S_{l,3}$		$S_{l,2}, \text{UD}$	$S_{l,3}, \text{UD}$
r		0.44 ± 0.14	0.40 ± 0.084		0.50 ± 0.12	0.53 ± 0.079
N_{LAE}		221	161		558	434
model		$S_{h,2}$	$S_{h,3}$		$S_{h,2}, \text{UD}$	$S_{h,3}, \text{UD}$
r		0.50 ± 0.16	0.48 ± 0.11		0.58 ± 0.14	0.61 ± 0.11
N_{LAE}		217	179		560	465

Table 3. Same as Table. 2, but for neutral fraction images. Since we randomly chosen the training sample, the average number of LAEs is not identical to the Table. 3.

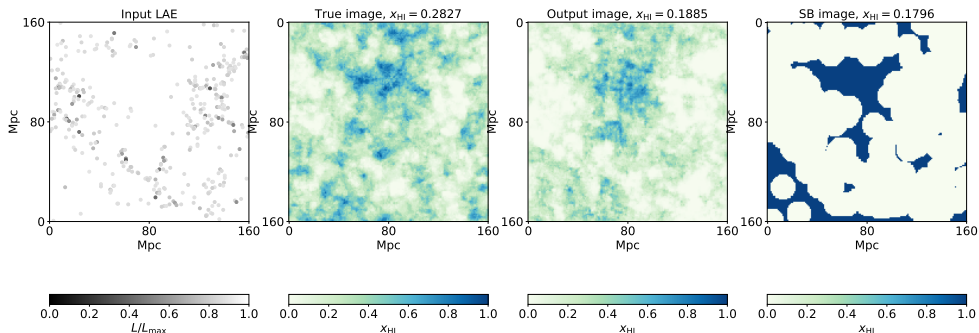


Figure 10. The panels show examples of the input LAE distribution, the true neutral fraction image, the cGAN output and the SB model images from left to right. This is the H_2, UD model. In the SB model, we use $R = 16\text{Mpc}$.

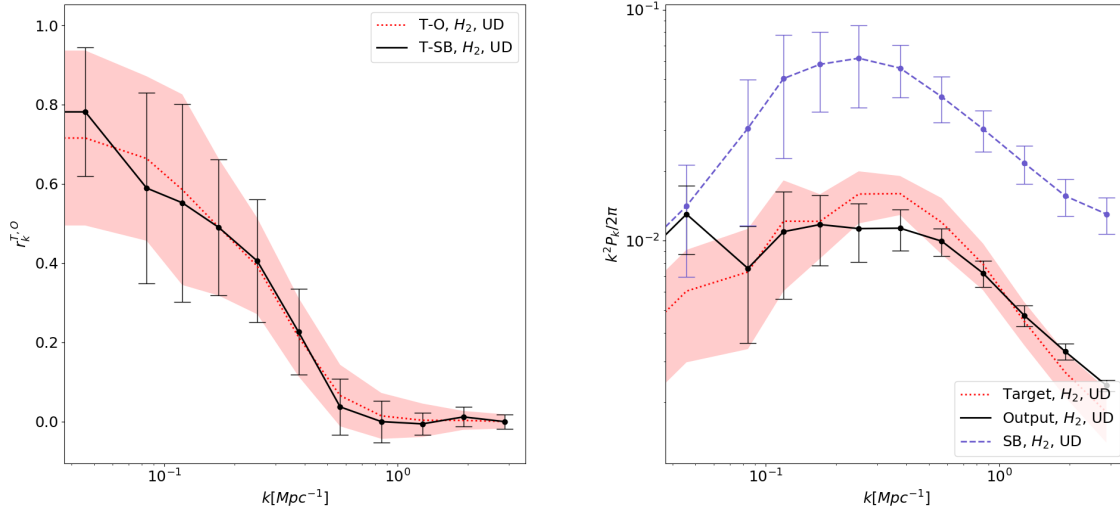


Figure 11. Left : Dotted line is the correlation coefficient between the true image of the neutral fraction and the cGAN output image, and the solid line is that between the true image and the image in the SB model. We compare the H_2 , UD model with the SB model with R of 16 Mpc. The shade and error bar show the sample variance. Right : The dotted, solid and dashed lines are the power spectrum of the neutral fraction for the true image, the cGAN output image and the SB model.

true and predicted images. Although the overall correlation is rather weak ($0.14 \sim 0.27$ for H models and $0.4 \sim 0.5$ for S models) assuming the HSC deep survey, the correlation at large scales ($k \sim 0.04 \text{ Mpc}^{-1}$) is much better ($0.4 \sim 0.5$ for H models and $0.7 \sim 0.8$ for S models). We also found that the correlation can be stronger by increasing the number of LAEs with the HSC ultra deep survey.

We also used the 21cm-line auto-power spectrum and the 21cm-LAE correlation coefficient to evaluate the accuracy of the reconstruction. The result of cGAN output images is consistent with that of true images at all scales within sample variance. This indicates that the network learns the fluctuation in 21cm-line signal and the correlation between 21cm-line signal and LAE distribution correctly.

We have evaluated the detectability of the cross correlation between observed 21cm-line maps and predicted images. We found that the cross power spectrum can be measured at large scales ($k < 0.2 \text{ Mpc}^{-1}$) with 400 hours of MWA Phase II observation and the HSC deep survey.

As a final remark, the cross correlation is a crucial observable to distinguish the 21cm-line signal from residuals of foregrounds. Furthermore, if the cross correlation were detected, we have to compare the signal with a number of numerical simulations in order to interpret the result. Our method can confirm the detection and allow us to interpret the reionization scenario simultaneously. Nevertheless, there is room for improvement on our method in many ways: optimizing the network hyper-parameters, increasing the number of training data sets, and adding more information. By boosting the detectability, our method can be an essential tool for future data analysis of the EoR 21cm-line observation.

ACKNOWLEDGEMENTS

We would like to thank Hidenobu Yajima for providing us with the Lyman- α transmission code and Tomoaki Ishiyama for conducting the N-body simulation used in this work. We also thank to Taisuke Nakashima and Kenji Kubota for helpful discussion for the work. SY is supported by JSPS Overseas Research Fellowships. HS was supported by the NSFC (Grant No.11850410429), the China Postdoctoral Science Foundation, the Tsinghua International Postdoctoral Fellowship Support Program, and the International Postdoctoral Fellowship from the Ministry of Education and the State Administration of Foreign Experts Affairs of China. KT is partially supported by Grand-in-Aid from the Ministry of Education, Culture, Sports, and Science and Technology (MEXT) of Japan No. 16H05999, Bilateral Joint Research Projects of JSPS, and the ISM Cooperative Research Program 2020-ISMCRP-2017.

REFERENCES

- Ali Z. S., et al., 2015, *ApJ*, **809**, 61
- Barry N., et al., 2019, arXiv e-prints, [p. arXiv:1909.00561](https://arxiv.org/abs/1909.00561)
- Doussot A., Eames E., Semelin B., 2019, arXiv e-prints, [p. arXiv:1904.04106](https://arxiv.org/abs/1904.04106)
- Fan X., et al., 2006, *AJ*, **132**, 117
- Feng C., Cooray A., Keating B., 2017, *ApJ*, **846**, 21
- Furlanetto S., Oh S. P., Briggs F., 2006, *Phys. Rept.*, **433**, 181
- Gillet N., Mesinger A., Greig B., Liu A., Ucci G., 2018, preprint, ([arXiv:1805.02699](https://arxiv.org/abs/1805.02699))
- Goodfellow I. J., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y., 2014, arXiv e-prints, [p. arXiv:1406.2661](https://arxiv.org/abs/1406.2661)
- Hasegawa K., Semelin B., 2013, *MNRAS*, **428**, 154
- Hasegawa K., et al., 2016, arXiv e-prints, [p. arXiv:1603.01961](https://arxiv.org/abs/1603.01961)

- Hassan S., Andrianomena S., Doughty C., 2019a, arXiv e-prints, p. [arXiv:1907.07787](#)
- Hassan S., Liu A., Kohn S., La Plante P., 2019b, *MNRAS*, **483**, 2524
- Heneka C., Cooray A., Feng C., 2017, *ApJ*, **848**, 52
- Hutter A., Dayal P., Müller V., Trott C. M., 2017, *ApJ*, **836**, 176
- Inoue A. K., et al., 2018, *PASJ*, **70**, 55
- Ishiyama T., Fukushige T., Makino J., 2009, *PASJ*, **61**, 1319
- Ishiyama T., Nitadori K., Makino J., 2012, arXiv e-prints, p. [arXiv:1211.4406](#)
- Isola P., Zhu J., Zhou T., Efros A. A., 2016, CoRR, abs/1611.07004
- Jennings W. D., Watkinson C. A., Abdalla F. B., McEwen J. D., 2018, preprint, ([arXiv:1811.09141](#))
- Kern N. S., Liu A., Parsons A. R., Mesinger A., Greig B., 2017, *ApJ*, **848**, 23
- Konno A., et al., 2018, *PASJ*, **70**, S16
- Kubota K., Yoshiura S., Takahashi K., Hasegawa K., Yajima H., Ouchi M., Pindor B., Webster R. L., 2018, *MNRAS*, **479**, 2754
- La Plante P., Ntampaka M., 2019, *ApJ*, **880**, 110
- Li W., et al., 2019, *MNRAS*, **485**, 2628
- Lidz A., Zahn O., Furlanetto S. R., McQuinn M., Hernquist L., Zaldarriaga M., 2009, *ApJ*, **690**, 252
- Mesinger A., Furlanetto S., Cen R., 2011, *MNRAS*, **411**, 955
- Moriwaki K., Yoshida N., Eide M. B., Ciardi B., 2019, *MNRAS*, p. 2236
- Mustafa M., Bard D., Bhimji W., Lukić Z., Al-Rfou R., Kratochvil J. M., 2019, *Computational Astrophysics and Cosmology*, **6**, 1
- Ouchi M., et al., 2018, *PASJ*, **70**, S13
- Paciga G., et al., 2011, *MNRAS*, **413**, 1174
- Paciga G., et al., 2013, *MNRAS*, **433**, 639
- Park J., Kim H.-S., Wyithe J. S. B., Lacey C. G., 2014, *MNRAS*, **438**, 2474
- Park J., Mesinger A., Greig B., Gillet N., 2019, *MNRAS*, **484**, 933
- Parsons A. R., et al., 2010, *AJ*, **139**, 1468
- Patil A. H., et al., 2017, *ApJ*, **838**, 65
- Planck Collaboration et al., 2016, *A&A*, **596**, A108
- Ronneberger O., Fischer P., Brox T., 2015, arXiv e-prints, p. [arXiv:1505.04597](#)
- Schmit C. J., Pritchard J. R., 2018, *MNRAS*, **475**, 1213
- Shibuya T., et al., 2018, *PASJ*, **70**, S14
- Shimabukuro H., Semelin B., 2017, *MNRAS*, **468**, 3869
- Shimabukuro H., Mao Y., Tan J., 2020, arXiv e-prints, p. [arXiv:2002.08238](#)
- Shimizu I., Inoue A. K., Okamoto T., Yoshida N., 2014, *MNRAS*, **440**, 731
- Sobacchi E., Mesinger A., Greig B., 2016, *MNRAS*, **459**, 2741
- Takada M., et al., 2014, *PASJ*, **66**, R1
- Tamura N., et al., 2016, in Proc. SPIE. p. 99081M ([arXiv:1608.01075](#)), doi:10.1117/12.2232103
- Tingay S. J., et al., 2013, *PASA*, **30**, e007
- Wayth R. B., et al., 2018, *Publ. Astron. Soc. Australia*, **35**, 33
- Wiersma R. P. C., et al., 2013, *MNRAS*, **432**, 2615
- Yajima H., Sugimura K., Hasegawa K., 2018, *MNRAS*, **477**, 5406
- Yoshiura S., Line J. L. B., Kubota K., Hasegawa K., Takahashi K., 2018, *MNRAS*, **479**, 2767
- Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, *MNRAS*, **414**, 727
- Zamudio-Fernandez J., Okan A., Villaescusa-Navarro F., Bilaloglu S., Derin Cengiz A., He S., Perreault Levasseur L., Ho S., 2019, arXiv e-prints, p. [arXiv:1904.12846](#)
- van Haarlem M. P., et al., 2013, *A&A*, **556**, A2

This paper has been typeset from a \LaTeX file prepared by the author.