

Configuration requirements for 21-cm forest background quasar searches with Moon-based interferometer*

Siyuan Zhang (张思源)¹ Qi Niu (牛琦)¹ Yichao Li (李毅超)^{1†} Xin Zhang (张鑫)^{1,2,3‡}

¹Key Laboratory of Cosmology and Astrophysics (Liaoning), College of Sciences, Northeastern University, Shenyang 110819, China

²National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang 110819, China

³Key Laboratory of Data Analytics and Optimization for Smart Industry (Ministry of Education), Northeastern University, Shenyang 110819, China

Abstract: The 21-cm forest offers a powerful cosmological probe of the thermal history and small-scale structure of the intergalactic medium during the Epoch of Reionization (EoR). However, its success critically depends on the availability of high-redshift radio-loud quasars (HzRLQs) as background sources. In this work, we investigate the configuration requirements for a Moon-based low-frequency radio interferometer aimed at maximizing the detection of HzRLQs for future 21-cm forest studies. Building upon a previously developed quasar luminosity function (QLF), we forecast HzRLQ abundances under various array configurations. Assuming a total survey area of 10^4 deg^2 and 1 year of observation, we compare continuum surveys with 10 MHz bandwidth and 21-cm forest surveys with 5 kHz resolution. Our results show that a minimum collecting area of $\sim 6500 \text{ m}^2$ enables detection at $z \sim 6$, while SKA-like arrays ($N_{\text{st}} = 512$) extend the detection limit to $z \sim 10$ for 21-cm forest surveys and $z \sim 16$ for continuum surveys. Larger arrays with $N_{\text{st}} = 2048$ can reach $z \sim 11$ in 21-cm forest mode. We also explore configurations that maintain fixed collecting areas while increasing the number to enhance survey efficiency. This boosts source detection but significantly increases the data volume and computational demands. These results underscore the importance of optimizing array design for different survey goals and balancing sensitivity, spectral resolution, and data management. A well-designed Moon-based array could open a new observational window to reionization and early cosmic structure formation.

Keywords: quasars, radio loud quasars, reionization, HI line emission, surveys

DOI: 10.1088/1674-1137/ade957 **CSTR:** 32044.14.ChinesePhysicsC.49115105

I. INTRODUCTION

The 21-cm signal from the hyperfine transition of neutral hydrogen (HI) has become a powerful cosmological probe for the study of the early Universe, particularly, the Dark Ages, Cosmic Dawn, and Epoch of Reionization (EoR) [1]. The 21-cm absorption features imprinted by HI in the early Universe on the spectra of high-redshift radio-loud background sources have emerged as a promising probe of the intergalactic medium (IGM) during the Cosmic Dawn and EoR [2–4]. Recently, Shao *et al.* [5] proposed a novel cosmological probe known as the 21-cm forest, which provides a unique probe of small-scale structures during the EoR. The 21-cm forest relies on the absorption characteristics of HI in the intervening structures along the line of sight with respect to the spectra of the bright background radio sources, such as radio-

loud quasars. This technique provides valuable insights into the nature of dark matter (DM) and its role in shaping small-scale structure formation [6–10].

Breakthroughs in the 21-cm forest have been difficult to achieve because of ongoing challenges in both analytical modeling and observations. Because of the absence of analytical models, the parameter inference of the 21-cm forest relies on complex small-scale simulations [11, 12]. The substantial computational cost of simulations poses a significant challenge for constraining the astrophysical parameters. Recently, several new approaches have been proposed to address this challenge, including the deep-learning-driven likelihood-free parameter inference methods [13] and a halo-model-based analytical model for the one-dimensional power spectrum of the 21-cm forest [14].

Moreover, the detection of the 21-cm forest power

Received 22 April 2025; Accepted 23 June 2025; Published online 24 June 2025

* Supported by multiple funding sources: the National SKA Program of China (2022SKA0110200, 2022SKA0110203), the National Natural Science Foundation of China (12473091, 12473001) and the National 111 Project (B16009). Yichao Li acknowledges the support of the Fundamental Research Funds for the Central Universities (N2405008)

[†] E-mail: liyichao@mail.neu.edu.cn

[‡] E-mail: zhangxin@mail.neu.edu.cn

©2025 Chinese Physical Society and the Institute of High Energy Physics of the Chinese Academy of Sciences and the Institute of Modern Physics of the Chinese Academy of Sciences and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

spectrum remains challenging due to the limited sensitivity of existing low-frequency observational facilities. This constraint is especially critical for the identification of high-redshift radio-loud quasars (HzRLQs), which serve as essential background sources. While optical and near-infrared surveys have extended quasar observations to redshifts beyond 6 [15–17], the number of detected radio-loud quasars at such high redshifts remains small. A recent study [18] examined the abundance of HzRLQs and confirmed that observations are still constrained by flux limitations, even with the next-generation low-frequency radio telescope arrays, such as the low-frequency Square Kilometre Array (SKA-Low). This scarcity presents a significant obstacle to advancing 21-cm forest studies.

Advances in space technology now make it possible to utilize the far side of the Moon—a radio-quiet environment that is uniquely shielded from Earth's ionospheric aberrations and man-made radio frequency interference (RFI) [19–21]. This pristine electromagnetic environment provides an extraordinary platform for groundbreaking low-frequency radio astronomy observations [22]. The RFI-free conditions on the far side of the Moon allow the detection of weak cosmological signals, in particular, the 21-cm HI line from the Dark Ages, Cosmic Dawn, and EoR [23, 24].

In this work, we thoroughly explore the survey strategies for HzRLQs with the future Moon-based low-frequency array. Although the technology for building a large radio interferometric array like SKA-Low on the far side of the Moon remains underdeveloped, this study explores its immense potential and the fundamental configuration requirements for detecting HzRLQs. Specifically, we analyze two deployment scenarios: one involving variations in the effective receiving area and the other involving changes in the station diameter. These approaches are designed to address existing observational limitations. Additionally, we provide a detailed comparison of the HzRLQs under both the continuum and 21-cm forest surveys. However, these strategies also introduce new engineering challenges related to array optimization and data processing. Such a Moon-based observatory will provide transformative insights into the formation of the cosmic structure at the dawn of the universe and during its first billion years [25–27].

The remainder of this paper is organized as follows. In Section II, we describe the construction of the luminosity function for predicting the abundance of HzRLQs and the improvement of the model by including obscuration effects. In Section III, we present various scenarios for the construction of the low-frequency radio array on the Moon. In Section IV, we present our results, including the influence of obscuration effects on the prediction of high-redshift quasars, how to build future Moon-based arrays with the most significant detection of quasars and cost savings, and the baseline requirements to minimize

confusion effects. Finally, in Section V, we summarize our conclusions and outline future research directions.

II. LUMINOSITY FUNCTION MODEL

A. Physical-driven model

This study extends the physical-driven model established in previous work by incorporating the impact of optical observation biases on the detection of obscured quasars. A brief summary of the model is provided here, with detailed descriptions available in [18]. The key steps in constructing the model are as follows:

1. Calculating the black hole (BH) mass function.

Quasars are driven by the active behavior of BHs at the centers of galaxies. Therefore, the first step is to determine the abundance of BHs in galaxy centers. We use a dark matter halo mass function in Sheth–Tormen form [28] to describe the abundance of halos of different masses at different redshifts. We assume that each halo hosts a galaxy and that each galaxy contains a central BH. The BH mass function is then derived from the halo mass function using the mass relationship between BHs (M_{BH}) and their corresponding halos (M_{h}) [18, 29]:

$$M_{\text{BH}}(M_{\text{h}}, z) = A \left(\frac{M_{\text{h}}}{10^{12} M_{\odot}} \right)^{\frac{5}{3}} \left(\frac{\xi(z)(1+z)^3}{\Omega_{\text{m}}} \right)^{\frac{5}{6}} M_{\odot}, \quad (1)$$

where A is the amplitude parameter and $\xi(z)$ is a dimensionless parameter related to redshift z . In this work, we adopt the best-fit value $A \approx 3.2 \times 10^6$ from [18].

2. Determining the quasar duty cycle with quasar luminosity function (QLF). The quasar duty cycle $D_{\text{q}} = t_{\text{q}}/t_{\text{H}(z)}$ quantifies the fraction of black holes in the quasar phase, where t_{q} is the quasar lifetime and $t_{\text{H}(z)}$ is the cosmic age at redshift z . Assuming Eddington-limited accretion, we constrain t_{q} by fitting observational data of the QLF at high redshifts [30, 31].

3. Determining the fraction of radio-loud quasars whose radio flux exceeds the required threshold F_{th} .

We assume that 10% of quasars are radio-loud (according to observational experience, e.g., [32–34]). To account for the radio flux distribution of the quasar population at a given optical luminosity, we adopt the observed distribution of radio-loudness $R \equiv \log_{10}(L_{5\text{GHz}}/L_{4400})$, where $L_{5\text{GHz}}$ and L_{4400} are the luminosity values measured in the radio band (5 GHz) and optical band (4400 Å), respectively. We use a radio-loudness distribution function $N(R)$ constrained using the observation data from Bariuan *et al.* [35]. The fraction of radio-loud quasars over a flux threshold F_{th} is determined by integrating the radio-loud-

ness distribution function $N(R)$:

$$\epsilon(M_h, F_{th}) = \int_{R_0}^{\infty} dR N(R), \quad (2)$$

where the lower-bound $R_0 = R_0(M_{BH}(M_h, z), F_{th})$ is a function of black hole mass and flux threshold.

4. Estimate the abundance of HzRLQs. The abundance of HzRLQs is then estimated by integrating the weighted halo mass function:

$$N(\Delta z, \Delta \Omega, F_{th}) = \int_{\Delta z, \Delta \Omega} dV \int dM_h 10\% D_q \epsilon(M_h, F_{th}) \Phi_h(M_h), \quad (3)$$

where Δz and $\Delta \Omega$ are the redshift bin and survey sky area, respectively, dV is the cosmological volume element, and $\Phi_h(M_h) = dn/dM_h$ represents the halo mass function.

B. Obscuration-corrected QLF

The obscuration in the optical band may significantly affect predictions of the abundance of the HzRLQs. The distribution and nature of dust play a critical role in the observational properties of active galactic nuclei (AGNs). In particular, the molecular clouds and high-extinction regions can obscure AGNs, rendering them undetectable in optical surveys [36]. Quasars, especially at high redshifts, can be heavily obscured by dust in the optical band, which complicates their identification. This obscuration effect has been extensively studied, with findings indicating that a considerable fraction of AGNs are missed in optical surveys due to dust extinction [37]. Because the construction of the HzRLQ abundance model depends on the QLF in the optical band, dust obscuration in the optical band may lead to an underestimation of the abundance of HzRLQs. This underestimation has direct implications for studies such as the 21-cm forest, which rely on a complete census of background radio sources.

The obscuration renders the quasar luminosity in certain wavelength bands. To ensure the accuracy of the observed high-redshift quasar population, its luminosity function can be corrected by introducing an obscuration fraction, which is estimated based on the observations in the X-ray band [38]:

$$f_{obs} = \min[\psi_{max}, \max[\psi_0 - \beta \log(L_X/L_{X,0}), \psi_{min}]], \quad (4)$$

where $\phi_0 = 0.73$, $\psi_{max} = 0.84$, $\psi_{min} = 0.2$, $\beta = 0.24$, and $L_{X,0} = 10^{43.75} \text{ erg s}^{-1}$ [39, 40]. The X-band luminosity is estimated via the bolometric correction factor of the hard X-ray band [41]:

$$K_X(L_{bol}) = \frac{L_{bol}}{L_X} = a \left(1 + \left(\frac{\log(L_{bol}/L_{\odot})}{b} \right)^c \right), \quad (5)$$

where $a = 10.96$, $b = 11.93$, and $c = 17.79$. It is worth noting that the obscuration fraction function is primarily constrained using X-ray selected AGN samples at redshifts $z \lesssim 5$. Due to the limited observational data available for quasars at higher redshifts, the redshift evolution of the obscuration fraction remains uncertain. Therefore, we adopt a simplified approach and assume a redshift-independent obscuration fraction, acknowledging that this may introduce some uncertainties in our predictions.

We correct the luminosity function by multiplying the luminosity function of total quasars with the fraction of unobscured ones to obtain the luminosity function of the optically observed quasars:

$$\Phi = \Phi_{total}(M_{1450}, z) \times (1 - f_{obs}). \quad (6)$$

We compare the best-fit model with and without considering the obscuration effect in Fig. 1. The QLF model with the obscuration effect considered is shown by the solid curve. The dashed curve represents the best-fit model from Niu *et al.* [18]. The QLF model is fitted to the data by adjusting the mean quasar lifetime t_q until the likelihood function is maximized. Two main datasets are used in this work: one at $z \sim 6$ [30] and the other at

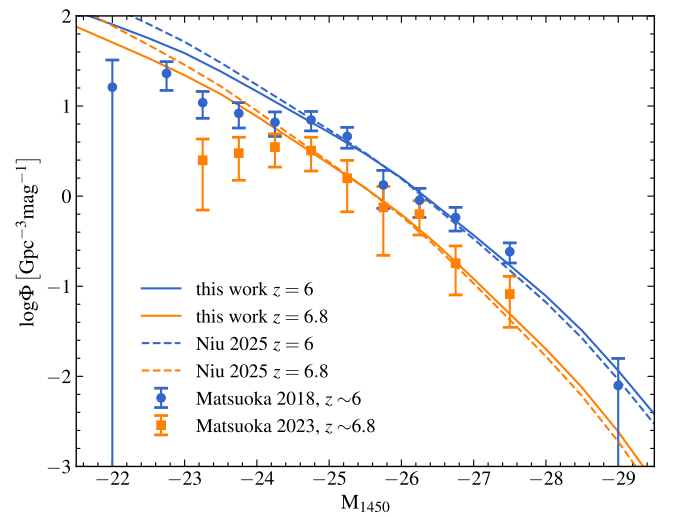


Fig. 1. (color online) Comparison of the best-fit QLF with and without considering the optical-band obscuration effect. Blue and orange curves denote models at redshifts $z = 6$ and $z = 6.8$, respectively. Solid lines illustrate models with obscuration effects in our work, whereas the dotted lines represent models without considering the obscuration effect. The blue and orange error bars show the measurements at $z \sim 6$ from Matsuoka *et al.* [30] and $z \sim 6.8$ from Matsuoka *et al.* [31]. Owing to observational incompleteness, the first two data points on the faint end are omitted.

$z \sim 6.8$ [31]. For the fits at $z = 6$ and $z = 6.8$, the least-square values are approximately 7.79 and 3.78, respectively. Considering the obscuration effect, the luminosity function is $t_q = 10^{5.79}$ at $z = 6$ and $t_q = 10^{5.76}$ at $z = 6.8$.

By comparing the results with and without obscuration effects, we find that incorporating obscuration leads to a slightly flatter QLF, a trend observed at both redshifts considered in our analysis. This is mainly because low-luminosity quasars are more likely to be obscured by dust, making them less likely to be detected in optical surveys. Including the obscuration effect can thus partially alleviate the discrepancy between theoretical models and observational QLF data at the faint end. However, the improvement is limited. As mentioned above, the obscuration fraction is based on low-redshift data, and its possible redshift evolution is not considered in this work. Additionally, the incompleteness of the observed sample at the faint end may still be the dominant factor contributing to the observed deviation.

III. MOON-BASED INTERFEROMETER CONFIGURATION

Existing research indicates that a significant detection of HzRLQs requires a radio interferometer with a sensitivity comparable to that of SKA-Low [18]. SKA-Low is designed to probe the early Universe through radio observations. It consists of 131,072 log-periodic dipole antennas, grouped into 512 stations, each containing 256 antennas [52]. Approximately 50% of these stations are concentrated within a 1 km diameter central core, while the remainder are distributed along three spiral arms extending up to 74 km. The total effective collecting area is approximately 419000 m², providing exceptional sensitivity to faint radio signals. Operating within a frequency range of 50 to 350 MHz, SKA-Low can observe the 21-cm signal at redshifts approximately between $z \sim 3$ and $z \sim 27$, covering key cosmic epochs, such as the EoR and Cosmic Dawn.

The observational noise variance of the radio interferometer is estimated by

$$\sigma = \sqrt{2} \frac{k_B T_{\text{sys}}}{A_{\text{eff}}} \frac{1}{\sqrt{N_{\text{st}}(N_{\text{st}} - 1) \Delta t \Delta \nu}}, \quad (7)$$

where k_B is the Boltzmann constant, A_{eff} refers to the effective receiving area of a station, N_{st} is the number of stations, $\Delta \nu$ is the frequency channel width, and Δt is the integration time. T_{sys} is the system temperature:

$$T_{\text{sys}} = T_{\text{sky}} + T_{\text{rx}}, \quad (8)$$

where T_{sky} is the mean brightness temperature from the sky and takes the form

$$T_{\text{sky}} = (2.73 + 25.2 \times (408/\nu_{\text{MHz}}(z))^{2.75}) \text{ K}, \quad (9)$$

where $\nu(z)$ represents the frequency in MHz corresponding to redshift z .

In this work, we consider that the surveys are performed with a continuum survey, which has the flux measurements integrated over the frequency band of 10 MHz to enhance the signal-to-noise ratio for continuum source detection. An integration bandwidth of 10 MHz was selected to achieve an optimal balance between resolving the continuum's spectral shape and maintaining sufficient detection sensitivity. This bandwidth is sufficiently narrow to capture spectral features while remaining broad enough to minimize noise. Due to the limited frequency resolution, the following 21-cm forest analysis requires additional deep-field observations. We also consider a 21-cm forest survey with a spectrum resolution of 5 kHz. Such high-frequency resolution significantly reduces the detection ability for HzRLQs, but it is fine enough for the 21-cm forest analysis without additional follow-up observations.

We assume $T_{\text{rx}} = 0.1T_{\text{sky}} + 40 \text{ K}$ [42], a constant total survey area of $S_{\text{tot}} = 10000 \text{ deg}^2$, and a total observation time of $t_{\text{tot}} = 1 \text{ yr}$. The observational noise variance may vary according to different effective collecting areas and survey efficiency. It is worth noting that the sky temperature T_{sky} dominates the noise term in the low frequency bands, as shown in Eq. (8). We systematically varied the receiver temperature to $T_{\text{rx}} = 0.1T_{\text{sky}} + 70 \text{ K}$ and found that it is only 6.37% more than the original. This analysis demonstrates that variations in receiver noise temperature have a negligible influence on the total system noise under these conditions.

a. Varying the effective collecting area

We consider the future construction of low-frequency interferometric arrays of varying sizes on the far side of the Moon. Using SKA-Low as a reference, we maintain a constant station diameter of $D = 40 \text{ m}$. The array size ranges between $N_{\text{st}} = 8, 32, 128$, and 512 up to 2048 stations. Notably, an array with 512 stations is comparable in size to the current SKA-Low, while the 2048-station configuration represents a more ambitious scenario, with an effective collecting area exceeding one square kilometer.

b. Varying the station diameter

The solid angle of the field of view (FoV) for a single pointing is estimated using the primary beam size of the station given by $\Omega \sim (1.2\lambda/D_{\text{st}})^2$, where D_{st} is the diameter of a station. For the subsequent analysis, we use the wavelength λ corresponding to 150 MHz to estimate the FoV. By varying the station diameter, the FoV for a single pointing changes, thereby affecting the survey effi-

ciency, while keeping the total survey area and observation time constant. To maintain consistency with the current SKA-Low, we ensure that the total effective collecting area remains at $\sim 419000\text{m}^2$, varying the number of stations and diameter of the station simultaneously. In particular, we consider the following station configurations: $\{(N_{\text{st}}, D_{\text{st}})\} = \{(512, 40\text{m}), (2048, 20\text{m}), (8192, 10\text{m})\}$.

IV. RESULTS AND DISCUSSIONS

Using the configuration parameters introduced in Section III, we estimate the system noise variance level for each case. We use 10 times the noise variance level as the lower bound of the integration of the predicted luminosity function to forecast the abundance of HzRLQs.

A. Configuration requirements with continuum survey

The number of HzRLQs detected (with their flux over 10 times the corresponding noise level) in continuum surveys is shown in Fig. 2, where different colors represent results for various interferometer size configurations. Because the predicted abundance depends on the width of the redshift bins, we fix the redshift bin $\Delta z = 0.5$ for the analysis of continuum surveys. Generally, the number of detected HzRLQs significantly reduced as redshift increased. The gray histogram corresponds to an array of eight 40 m-diameter stations, with an effective collecting area of 6546.875m^2 . This configuration enables the detection of approximately 10^3 radio-loud quasars at $z \sim 5$ and provides statistically significant detections ($N > 10$) up to $z \sim 10$.

Increasing the number of stations significantly enhances detection capabilities. With $N_{\text{st}} = 128$, the number of detected HzRLQs increases by an order of magnitude at each redshift bin, and the redshift detection limit extends to $z \sim 16$.

As discussed in Section III, an interferometer with 512 40 m-diameter stations achieves an effective collecting area comparable to that of the current SKA-Low. An even more ambitious configuration, with $N_{\text{st}} = 2048$, would push the total collecting area beyond one square kilometer. Such an interferometer would enable the robust detection of HzRLQs up to the beginning of the EoR or even Cosmic Dawn.

Increasing the effective collecting area of an interferometric array can significantly enhance its sensitivity to HzRLQs. Beyond simply scaling up the array, we also explore an alternative strategy that maintains the same total collecting area – comparable to that of the SKA – while boosting survey efficiency by employing smaller station diameters to achieve a larger FoV. The corresponding results are shown in Fig. 3, where different colors indicate configurations with varying station numbers and sizes.

Our findings demonstrate that even without increas-

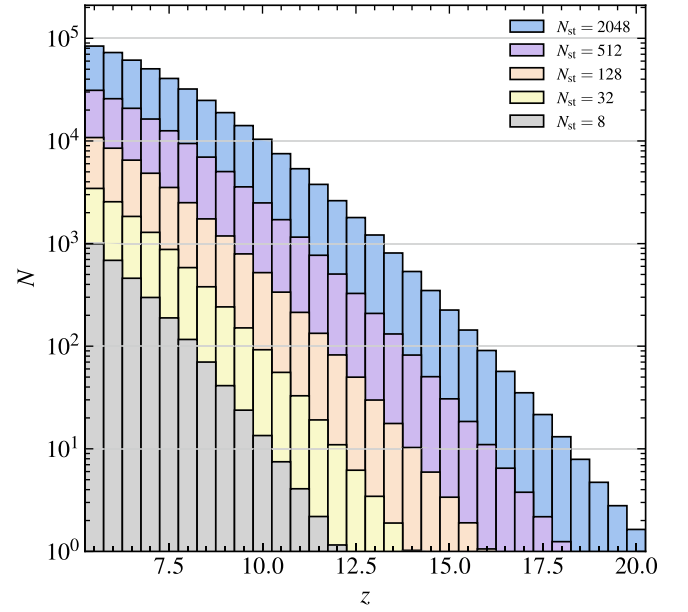


Fig. 2. (color online) Number of HzRLQs detected with continuum survey as a function of redshift with different sizes of interferometer. The results with $N_{\text{st}} = 8, 32, 128$, and 512 up to 2048 stations are shown in different colors.

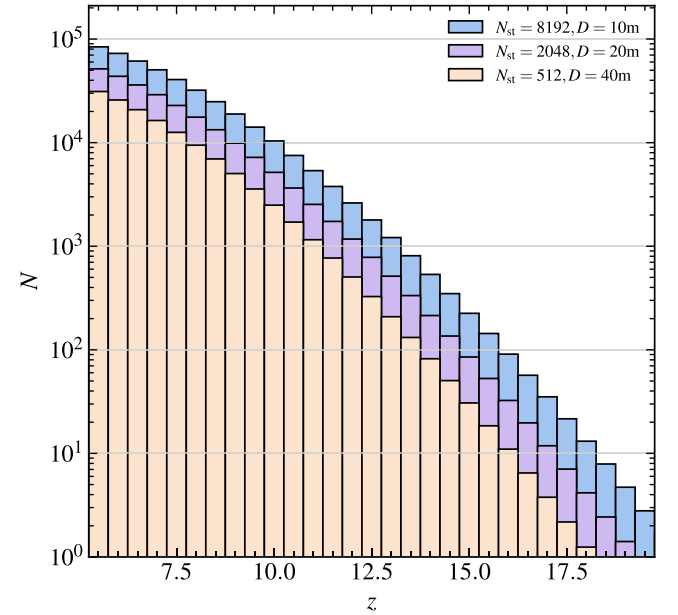


Fig. 3. (color online) Number of HzRLQs detected with continuum survey as a function of redshift with different diameters of station and number of station configurations. The results with $\{(N_{\text{st}}, D_{\text{st}})\} = \{(512, 40\text{m}), (2048, 20\text{m}), (8192, 10\text{m})\}$ are shown in different colors.

ing the total collecting area, the detection capability for HzRLQs can be improved by increasing the number of stations and reducing the diameter of each station. This configuration effectively enhances sky coverage and sur-

vey speed. However, it is important to note that such a design imposes substantial demands on the system, particularly in terms of data transmission, communication bandwidth, and computational resources required for data processing. These factors pose significant challenges for the data analysis pipelines of interferometric arrays and must be carefully considered in future instrument design.

Our analysis shows that the detectability of HzRLQs in continuum surveys is strongly dependent on both the size and configuration of the interferometric array. Increasing the number of stations significantly boosts sensitivity, enabling detection at higher redshifts, with large arrays ($N_{\text{st}} \geq 512$) capable of probing the EoR and potentially even Cosmic Dawn. Moreover, we demonstrate that optimizing array layout – by increasing station number while reducing individual station size – can further enhance survey efficiency without increasing the total collecting area. This approach increases the survey FoV and improves detection rates; however, it introduces substantial challenges in data transmission and processing. These trade-offs highlight the need for carefully balanced array designs that optimize both scientific return and technical feasibility in future low-frequency radio surveys targeting HzRLQs.

B. Configuration requirements with 21-cm forest survey

Through wide-area continuum surveys, large low-frequency radio interferometric arrays have the potential to detect a substantial number of high-redshift radio-loud quasars (HzRLQs). By integrating over frequency channels, continuum observations significantly enhance the signal-to-noise ratio. However, this integration comes at the expense of frequency resolution, which is critical for 21-cm forest power spectrum analysis. Hence, continuum detections often require dedicated follow-up observations to probe fine spectral structures. In this work, we assume a frequency resolution of 5 kHz for 21-cm forest surveys and investigate how different array sizes impact the detectability of HzRLQs. The corresponding results are shown in Fig. 2, where different colors represent low-frequency arrays of various sizes. For this analysis, we adopt a redshift bin width of $\Delta z = 1$.

Compared to continuum observations, the high spectral resolution required for 21-cm forest studies leads to a significant increase in observational noise, thereby reducing the detectability of HzRLQs. For a mini-sized array with only $n = 8$ stations, detecting HzRLQs beyond $z > 5$ becomes nearly impossible. However, for an array comparable to the ground-based SKA, with $n = 512$ stations, a moderate number of HzRLQs can still be detected, even though sensitivity remains lower than in continuum mode. In this configuration, the detection limit extends to approximately $z \sim 10$, which is sufficient for meaningful

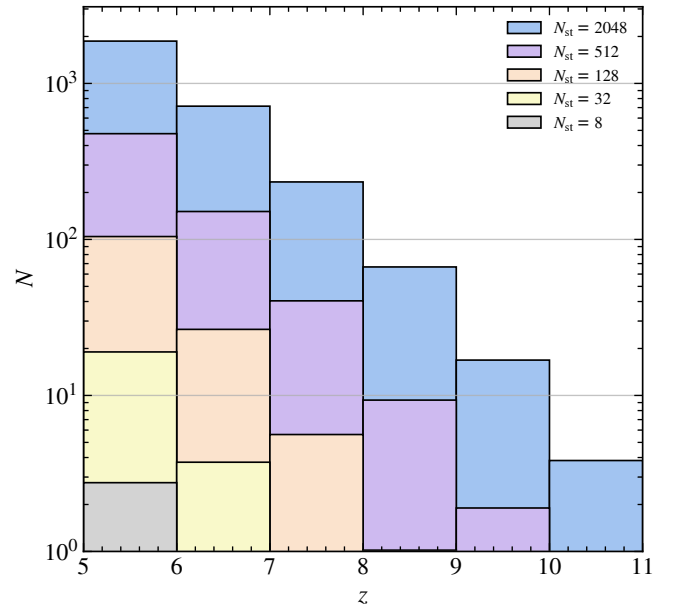


Fig. 4. (color online) Similar to Fig. 2 but for 21-cm forest survey with 5 kHz frequency resolution.

cosmological studies using the 21-cm forest. For extremely large arrays with $n = 2048$ stations, the detection redshift limit can be further extended to $z \sim 11$.

We also explore a scenario in which the total effective collecting area is constant but the number of stations is increased while reducing the diameter of each station. This approach enhances the FoV and survey efficiency, potentially improving the HzRLQ detection rate. The results are presented in Fig. 5, where different colors indicate different configurations of station number and size. Our findings show that reducing the station diameter by a factor of four and simultaneously increasing the number of stations by a factor of sixteen can substantially increase both the number of detectable HzRLQs and maximum redshift reached. However, this gain in survey efficiency comes with a trade-off: significantly greater data transmission, storage, and processing demands, posing major challenges for interferometric data analysis.

The detectability of HzRLQs in 21-cm forest surveys is governed by a complex interplay between array configuration and spectral resolution. While continuum surveys benefit from lower noise levels due to frequency integration, 21-cm forest observations require high spectral resolution and thus face higher noise. Larger and denser interferometric arrays can extend detection capabilities to higher redshifts, but they also introduce considerable computational and infrastructural challenges. Achieving an optimal balance between sensitivity, resolution, and data processing efficiency will be essential for the successful design and operation of future low-frequency radio arrays targeting the 21-cm forest.

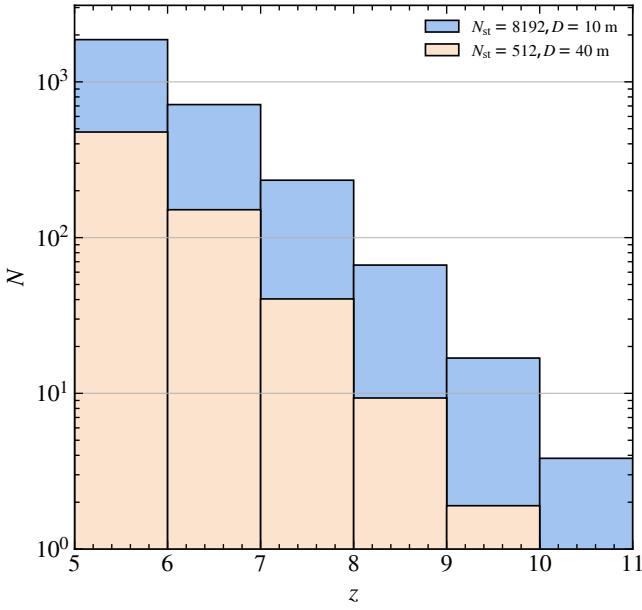


Fig. 5. (color online) Similar to Fig. 3 but for 21-cm forest survey with 5 kHz frequency resolution.

C. Confusion limit

The confusion limit is reached when the density of sources brighter than the observational noise becomes sufficiently high within the area of the synthesized beam. At this point, the ability of the radio array to resolve individual sources is limited not by sensitivity but by resolution. Consequently, the baseline length of the interferometric array plays a critical role in determining the array's effectiveness in mitigating source confusion. With the luminosity function constructed in Section II, we simulate the radio source samples with flux density ranging from $10^{-4} \sim 10^{-1}$ Jy, which is large enough to cover the faint end of the radio sources, within the redshift of $5.5 < z < 20.5$, and estimate the column number density as a function of flux density. We fit the flux-density cumulative column number density with a power law function:

$$N(> S) = CS^\beta, \quad (10)$$

where $N(> S)$ represents the column number density with source flux densities larger than S , and a power law function is characterized by the free parameters C and β . The best-fit values are $C = 3.94$, $\beta = -1.07$. The simulated flux-density cumulative column number density, as well as the best-fit curve, are shown in Fig. 6.

The confusion limit is then estimated via [43, 44]

$$\sigma_c = (\theta^2 m C)^{-1/\beta}, \quad (11)$$

where $C = 3.94$, $\beta = -1.07$ are the parameters of the best-fit power law function, $m = 30$ is the threshold parameter

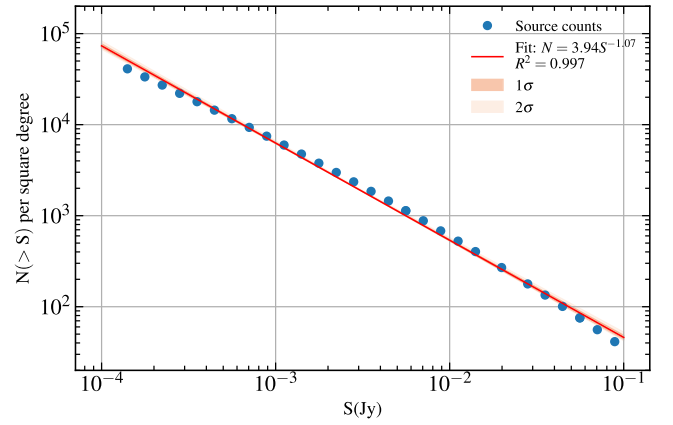


Fig. 6. (color online) Power-law fitting of source counts at $5.5 \lesssim z \lesssim 20.5$ and flux density at 200 MHz based on SKA-Low.

[45], and θ is the angular resolution. The angular resolution of the interferometer array is related to the maximum length of the baseline, e.g., $\theta = c/(\nu B_{\max})$, where c is the speed of light and ν is the corresponding frequency. To simplify the estimation, we ignore the frequency dependency and use $\nu = 200$ MHz in the following analysis.

We show the confusion limit as a function of the maximum baseline lengths in Fig. 7. The noise level with different numbers of 40 m stations is shown by horizontal dashed lines of different colors. It is evident that increasing the maximum baseline length significantly reduces the confusion limit. To ensure that the confusion limit remains below the system's thermal noise level, a minimum threshold for the maximum baseline should be established for configurations with varying numbers of stations. In general, a maximum baseline length of at least approximately 25 km is sufficient to render the confusion limit negligible.

D. Challenges in construction

As discussed in the previous sections, the substantial increase in the number of base stations brings a series of complex challenges in hardware design, data processing, and data storage [46].

The construction of large-scale radio arrays faces significant technical challenges as the number of antenna elements increases. Each station comprises multiple critical components, including antennas, low-noise amplifiers, filters, and analog-to-digital converters. With the dramatic rise in real-time data throughput, network congestion becomes a serious concern, particularly in distributed or remote configurations [47].

These challenges are further amplified in space-based applications such as lunar interferometer arrays, where observational data must be downlinked to ground-based data centers for further analysis. The resulting data volumes present substantial hurdles for storage, transport, and long-term data management.

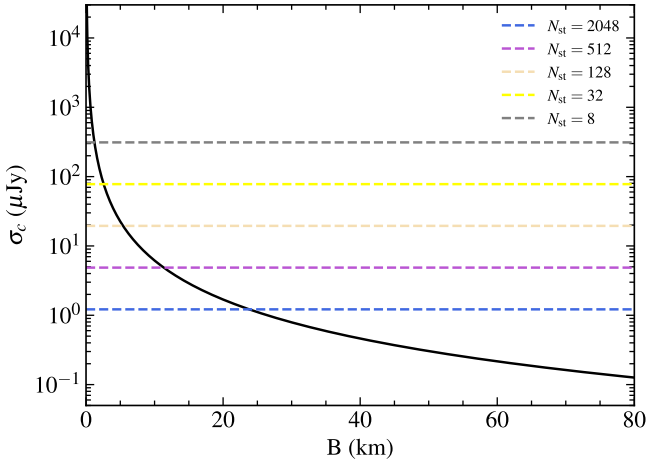


Fig. 7. (color online) The black line is the curve of the confusion limit σ_c as a function of the baseline length B . Different colored lines correspond to the observational noise under different configuration schemes.

From a computational perspective, system scaling results in the combinatorial growth of processing complexity. The calibration and imaging pipelines generate intermediate data products that often exceed the volume of raw observational data by orders of magnitude [48, 49].

Given that the data transfer rate of SKA-Low is as high as 7.15 Tbps [50] and the size of the data volume grows exponentially with the number of baselines ($N_{\text{st}}(N_{\text{st}} - 1)/2$), such a large amount of data makes it technically challenging to transmit raw observation data from the Moon-based interferometer directly back to Earth. To address the challenges of data transmission and real-time data processing in Moon exploration, we propose establishing a dedicated Moon-based data processing center. This facility will enable in situ data analysis, significantly reducing reliance on Earth-based processing while optimizing the scientific return from lunar missions.

Furthermore, dense station layouts heighten the risk of self-generated radio frequency interference (RFI), which can significantly degrade data quality. Mitigating these effects requires meticulous array design, including optimized station geometries, careful component isolation, and advanced electromagnetic shielding techniques [51].

Overall, these technical and infrastructural challenges must be carefully considered in the design of next-generation radio arrays, particularly those targeting high-resolution, wide-field observations such as the 21-cm forest.

V. CONCLUSION

This work investigates the scientific prospects and technical design requirements for detecting high-redshift radio-loud quasars (HzRLQs) using a Moon-based low-frequency interferometric array, with the goal of enabling 21-cm forest studies during the Epoch of Reioniza-

tion (EoR) and Cosmic Dawn.

We begin by improving the quasar luminosity function (QLF) model through the inclusion of dust obscuration effects. Because optical surveys tend to miss heavily obscured sources, particularly at the faint end, we introduce an X-ray-based correction to account for this bias. The obscuration-corrected QLF yields more accurate predictions for the abundance of HzRLQs and helps partially mitigate discrepancies between theoretical models and observations.

Assuming a total survey area of 10^4 deg^2 and integration time of 1 year, we analyze the detectability of HzRLQs in two complementary observing modes: continuum and 21-cm forest surveys. For continuum surveys, we adopt a bandwidth of 10 MHz, enabling significant signal-to-noise enhancement through frequency integration. In contrast, 21-cm forest observations require fine frequency resolution – assumed to be 5 kHz in this work – to resolve small-scale absorption features, which increases thermal noise and reduces source detectability.

We evaluate detection performance across a range of array sizes. For the smallest configuration considered ($N_{\text{st}} = 8$ stations with 40 m diameter), continuum surveys can detect HzRLQs up to $z \sim 10$, but the 21-cm forest survey becomes ineffective beyond $z \sim 5$. For SKA-scale arrays ($N_{\text{st}} = 512$), continuum surveys can detect quasars out to $z \sim 16$, while 21-cm forest observations remain feasible up to $z \sim 10$. For an extremely large array with $N_{\text{st}} = 2048$, the detection limit of the 21-cm forest survey extends to $z \sim 11$, supporting deeper exploration of the EoR. To minimize the confusion effects, the maximum baseline length must be $\geq 25 \text{ km}$.

We further explore configurations that maintain a fixed total collecting area ($\sim 419000 \text{ m}^2$) but increase the number of stations by reducing the station diameter. This strategy improves FoV and survey efficiency, enabling higher detection rates without additional collecting areas. For example, replacing the baseline configuration ($N_{\text{st}} = 512$, $D_{\text{st}} = 40 \text{ m}$) with ($N_{\text{st}} = 8192$, $D_{\text{st}} = 10 \text{ m}$) substantially increases the number of detectable HzRLQs and raises the detection redshift limit. However, this improvement comes with greater demands on real-time data transmission, onboard storage, calibration, and computing infrastructure, which is particularly challenging for a Moon-based observatory.

In summary, the detectability of HzRLQs depends sensitively on array configuration, frequency resolution, and survey design. Continuum surveys provide broad coverage with high signal-to-noise, while 21-cm forest observations offer detailed spectral information at the cost of sensitivity. Both modes benefit from careful optimization of array layout and observation parameters. Future efforts should focus on improving models of high-redshift quasar obscuration, refining array deployment strategies, and developing scalable data processing pipe-

lines. With these advancements, a lunar low-frequency interferometer could open a powerful new observational window into the reionization era, enabling high-precision studies of structure formation, dark matter, and the thermal history of the early Universe.

ACKNOWLEDGMENTS

We express our gratitude to ChatGPT for its assistance in refining the manuscript and enhancing the code to improve execution efficiency.

References

- [1] J. R. Pritchard and A. Loeb, *Phys. Rev. D* **78**, 103511 (2008)
- [2] C. L. Carilli, N. Y. Gnedin, and F. Owen, *Astrophys. J.* **577**, 22 (2002)
- [3] H. Shimabukuro, K. Ichiki, S. Inoue *et al.*, *Phys. Rev. D* **90**, 083003 (2014)
- [4] N. Thyagarajan, *Astrophys. J.* **899**, 16 (2020), arXiv: 2006.10070
- [5] Y. Shao, Y. Xu, Y. Wang *et al.*, *Nature Astron.* **7**, 1116 (2023)
- [6] L. Lopez-Honorez, O. Mena, Á. Moliné *et al.*, *JCAP* **08**, 004 (2016)
- [7] S. R. Furlanetto, S. P. Oh, and F. H. Briggs, *Phys.Rept.* **433**, 181 (2006)
- [8] J. R. Pritchard and A. Loeb, *Rept. Prog. Phys.* **75**, 086901 (2012)
- [9] Y. Shao, G. H. Du, T. N. Li *et al.*, *Phys. Lett. B* **862**, 139342 (2025)
- [10] H. Shimabukuro, Y. Xu, and Y. Shao, arXiv: 2504.14656
- [11] T. Šoltinský, J. S. Bolton, N. Hatch *et al.*, *Mon. Not. Roy. Astron. Soc.* **506**, 5818 (2021)
- [12] T. Šoltinský, J. S. Bolton, M. Molaro *et al.*, *Mon. Not. Roy. Astron. Soc.* **519**, 3027 (2023)
- [13] T. Y. Sun, Y. Shao, Y. Li *et al.*, *Commun. Phys.* **8**, 220 (2025)
- [14] Y. Shao, T. Y. Sun, M. L. Zhao *et al.*, arXiv: 2411.17094
- [15] X. B. Wu, F. Wang, X. Fan *et al.*, *Nature* **518**, 512 (2015)
- [16] X. Ding, M. Onoue, J. D. Silverman *et al.*, *Nature* **621**, 51 (2023)
- [17] Y. Zhang, T. An, A. Wang *et al.*, *A&A* **662**, L2 (2022), arXiv: 2205.05859
- [18] Q. Niu, Y. Li, Y. Xu *et al.*, *Astrophys. J.* **978**, 145 (2025)
- [19] J. Lazio, C. Carilli, J. Hewitt *et al.*, in *UV/Optical/IR Space Telescopes: Innovative Technologies and Concepts IV*, edited by H. A. MacEwen and J. B. Breckinridge (2009), p. 74360I.
- [20] C. W. James, R. D. Ekers, J. Álvarez-Muñiz *et al.*, *Phys. Rev. D* **81**, 042003 (2010)
- [21] B. T. Indermuehle, L. Harvey-Smith, M. Marquarding *et al.*, in *Observatory Operations: Strategies, Processes, and Systems VII* (2018), p. 107042S.
- [22] M. Klein Wolt, A. Aminaei, P. Zarka *et al.*, *Planet. Space Sci.* **74**, 167 (2012)
- [23] J. Burns, G. Hallinan, T. C. Chang *et al.*, arXiv: 2103.08623
- [24] R. S. Polidan, J. O. Burns, A. Ignatiev *et al.*, *Adv. Space Res.* **74**, 528 (2024)
- [25] T. B. H. Kuiper, D. L. Jones, M. J. Mahoney *et al.* in *Astrophysics from the Moon*, edited by M. J. Mumma and H. J. Smith (1990), pp. 522–527.
- [26] X. Chen, F. Gao, F. Wu *et al.*, *Phil. Trans. Roy. Soc. Lond. A* **382**, 20230094 (2024)
- [27] S. Iguchi, T. Yamada, Y. Yamasaki *et al.*, in *Space Telescopes and Instrumentation 2024: Optical, Infrared, and Millimeter Wave*, edited by L. E. Coyle, S. Matsuura, and M. D. Perrin (2024), p. 130922L.
- [28] R. K. Sheth and G. Tormen, *Mon. Not. Roy. Astron. Soc.* **329**, 61 (2002)
- [29] J. S. B. Wyithe and A. Loeb, *Astrophys. J.* **595**, 614 (2003)
- [30] Y. Matsuoka, M. A. Strauss, N. Kashikawa *et al.*, *Astrophys. J.* **869**, 150 (2018)
- [31] Y. Matsuoka, M. Onoue, K. Iwasawa *et al.*, *Astrophys. J. Lett.* **949**, L42 (2023)
- [32] D. Stern, S. G. Djorgovski, R. A. Perley *et al.*, *Astron. J.* **119**, 1526 (2000)
- [33] E. Bañados, B. P. Venemans, E. Morganson *et al.*, *Astrophys. J.* **804**, 118 (2015)
- [34] Y. Liu, R. Wang, E. Momjian *et al.*, *Astrophys. J.* **908**, 124 (2021)
- [35] L. G. C. Bariuan, B. Snios, M. Sobolewska *et al.*, *Mon. Not. Roy. Astron. Soc.* **513**, 4673 (2022)
- [36] L. Fanciullo, V. Guillet, F. Boulanger *et al.*, *A&A* **602**, A7 (2017)
- [37] H. L. Guo, B. Q. Chen, H. B. Yuan *et al.*, *Astrophys. J.* **906**, 47 (2021)
- [38] W. Li, K. Inayoshi, M. Onoue *et al.*, *Astrophys. J.* **950**, 85 (2023)
- [39] Y. Ueda *et al.* (MAXI Collaboration), *Publ. Astron. Soc. Jap.* **63**, S937 (2011)
- [40] Y. Ueda, M. Akiyama, G. Hasinger *et al.*, *Astrophys. J.* **786**, 104 (2014)
- [41] F. Duras, A. Bongiorno, F. Ricci *et al.*, *A&A* **636**, A73 (2020)
- [42] D. J. Bacon *et al.* (SKA Collaboration), *Publ. Astron. Soc. Austral.* **37**, e007 (2020)
- [43] W. S. Jeong, C. P. Pearson, H. M. Lee *et al.*, *Mon. Not. Roy. Astron. Soc.* **369**, 281 (2006)
- [44] A. Cohen, Tech. Rep. Memo #17, Long Wavelength Array (2004).
- [45] D. W. Hogg, *Astron. J.* **121**, 1207 (2001)
- [46] A. J. Beasley, R. Murowinski, and M. Tarenghi, *strophys. Space Sci.* **313**, 1 (2008)
- [47] B. Wang, X. Zhu, C. Gao *et al.*, *Sci. Rep.* **5**, 13851 (2015)
- [48] T. An, *Sci. China Phys. Mech. Astron.* **62**, 989531 (2019)
- [49] S. Wang, S. Prunet, S. Mignot *et al.*, *A&A* **692**, A61 (2024)
- [50] S. Guo, Y. LU, T. An *et al.*, *Sci. Sin. Phys. Mech. Astron.* **53**, 229504 (2023)
- [51] Y. Huang, X. P. Wu, Q. Zheng *et al.*, *A&A* **16**, 36 (2016)
- [52] <https://www.skao.int/en/explore/telescopes/SKA-Low>