Astronomical Techniques and Instruments, 2027

Article Open Access

# CosmicWeb-21cm array: a new radio observation array design for 21 cm cosmology

**Jiancheng Wang[1,2,3*], Jirong Mao[1,2,3], Xiangming Cheng[1,2,3], Yigong Zhang[4], Jie Su[5], Xiaogu Zhong[6], Min Wang[1,2,3], Zhixuan Li[1,2], Longhua Qin[7], Zhengjun Zhang[8]**

[1]*Yunnan Observatories*, *Chinese Academy of Sciences*, *Kunming* 650216, *China*

[2]*Key Laboratory for the Structure and Evolution of Celestial Objects*, *Chinese Academy of Sciences*, *Kunming* 650216, *China*

[3]*University of Chinese Academy of Sciences*, *Beijing* 100049, *China*

[4]*College of Information Engineering*, *Kunming University*, *Kunming* 650214, *China*

[5]*College of Big Data*, *Yunnan Agricultural University*, *Kunming* 650201, *China*

[6]*College of Physics and Electronic Engineering*, *Qujing Normal University*, *Qujing* 655011, *China*

[7]*Department of Physics*, *Yuxi Normal University*, *Yuxi* 653100, *China*

[8]*Department of Computer of Science and Information Engineering*, *Anyang Institute of Technology*, *Anyang* 455099, *China*

*Correspondence: 3960761784@qq.com

Received: March 30, 2026; Accepted: May 25, 2026; Published Online: May 26, 2026; https://doi.org/10.3724/ati2026022; https://cstr.cn/32083.14.ati2026022

© 2027 Editorial Office of Astronomical Techniques and Instruments, Yunnan Observatories, Chinese Academy of Sciences. This is an open access article under the CC BY-NC-ND 4.0 license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Citation: Wang, J. C., Mao, J. R., Cheng, X. M., et al. 2027. CosmicWeb-21cm array: a new radio observation array design for 21 cm cosmology. *Astronomical Techniques and Instruments*, https://doi.org/10.3724/ati2026022.

**Abstract:** This paper presents the CosmicWeb-21cm array, a novel radio interferometer designed to overcome the key challenges in 21 cm cosmology. Its core innovations include: (1) a multiscale nested geometry combining a hexagonal core with logarithmic spiral arms for excellent UV coverage and calibration robustness; (2) an intelligent non-uniform frequency sampling strategy that adapts resolution to foreground and signal characteristics, reducing data volume while preserving information; (3) a machine-learning-enhanced, physics-informed processing pipeline that achieves 99.7% foreground removal efficiency under realistic foregrounds and instrumental systematics; (4) a dual-polarization crossed dipole integrated with a dielectric lens and cryogenically cooled low-noise amplifier, achieving stable beam patterns and low noise temperature (< 35 K) across 50–250 MHz. These co-designed advances enable high-sensitivity mapping of the Epoch of Reionization, dark energy constraints, and cosmic-web structure.

**Keywords:** Dark ages; Reionization; Interferometers; Telescopes

## 1. INTRODUCTION

The 21 cm hyperfine transition line of neutral hydrogen offers an unparalleled observational probe for cosmology, enabling direct, three-dimensional mapping of baryonic matter across over 90% of the universe's history. Its significance stems from its unique capacity to trace the distribution and evolution of the intergalactic medium across three pivotal epochs: the Cosmic Dark Ages ($z \sim 30-15$), the Cosmic Dawn and Epoch of Reionization (EoR) ($z \sim 15-6$), and the post-reionization era where it can constrain Dark Energy ($z \leqslant 6$)[1,2]. During the Dark Ages, the 21 cm signal provides the only known direct observational window; during the EoR, it provides one of the most direct probes into the formation of the first stars and galaxies, encoding information on the timing, morphology, and sources of reionization[3]. At lower redshifts, 21 cm intensity mapping of large-scale structure offers a powerful, wide-field complement to optical galaxy surveys for measuring Baryon Acoustic Oscillations (BAOs) and constraining the nature of dark energy[4].

However, extracting this faint cosmological signal from real-world observations presents a formidable challenge, often described as one of the most difficult in modern observational cosmology. The primary obstacle is the overwhelming foreground contamination. Synchrotron and free-free emission from our Galaxy and extragalactic sources are 4 to 5 orders of magnitude brighter than the

desired 21 cm signal across the entire band[5]. These foregrounds, while spectrally smooth, are complex in their spatial structure. Second, instrumental systematics, including gain and phase errors, beam variations, and polarization leakage, can imprint spectral structure that mimics or obscures the cosmological signal[6]. Third, the radio frequency interference (RFI) environment is increasingly hostile, corrupting substantial portions of the observational band. Finally, the sheer data volume and computational complexity of the required signal processing, particularly the separation of foregrounds from the signal, push current high-performance computing resources to their limits[7].

Considerable effort has been invested to overcome these challenges. Pioneering low-frequency interferometers like the Murchison Widefield Array (MWA)[8], the Low-Frequency Array (LOFAR)[9], and the Hydrogen Epoch of Reionization Array (HERA)[10] have made substantial progress in instrument design, calibration, and foreground subtraction techniques. HERA, in particular, employs a highly redundant configuration to simplify calibration and uses delay-domain filtering for foreground isolation[11]. The upcoming SKA-Low telescope promises unprecedented sensitivity[12].

Despite these advances, current-generation arrays face persistent limitations that ultimately constrain their sensitivity and the fidelity of cosmological signal recovery. As follows: (1) Suboptimal UV coverage. Many arrays use regular grids or pseudo-random distributions, which can create grating lobes, alias power, and leave gaps in spatial frequency sampling, complicating imaging and power spectrum estimation[13]. (2) Inefficient frequency sampling. Uniform frequency channelization does not optimally exploit the fundamental spectral smoothness of foregrounds, leading to excessive data volume and suboptimal signal-to-noise in the cosmological signal band[14]. (3) Disjointed processing pipelines. Calibration, RFI mitigation, foreground subtraction, and power spectrum estimation are often treated as sequential, independent steps. Errors propagate non-linearly through this chain, and the physical priors about the signal and foregrounds are not fully leveraged in a unified, optimal framework[15].

The CosmicWeb-21cm array design proposed in this paper is conceived as a holistic, next-generation solution that addresses these systemic limitations through co-optimization of hardware, observation strategy, and software. It moves beyond incremental improvements by fundamentally rethinking the entire signal chain, from the geometry of the array elements to the final cosmological inference, as an integrated system. The core philosophy is to embed the knowledge of the signal and contaminant properties into the design itself, thereby simplifying the subsequent extraction of the faint 21 cm signal.

This paper is organized as follows: Section 2 outlines the overarching design ideas and enumerates the core innovations. Section 3 details the novel multi-scale nested array geometry. Section 4 describes the intelligent, non-uniform frequency sampling and encoding strategy. Section 5 presents the adaptive, physics-informed signal processing architecture. Section 6 present antenna design in detail. Section 7 discusses system implementation and simulated performance metrics. Section 8 delineates the key scientific objectives enabled by this design, and Section 9 provides a phased implementation roadmap. Conclusion and Discussion are presented in Section 10.

## 2. DESIGN IDEAS AND CORE INNOVATIONS

The CosmicWeb-21cm design is conceived not as an incremental upgrade to existing 21 cm arrays, but as a foundational re-imagining of the entire observational system. It is predicated on the conviction that the extreme technical challenge of detecting the faint cosmological signal, buried under orders-of-magnitude brighter foregrounds and systematics, demands a co-optimized approach. Traditional designs often treat the antenna layout, the receiver chain, the observing strategy, and the data analysis pipeline as separate, sequentially optimized components. In contrast, CosmicWeb-21cm is architected from the ground up as an integrated instrument where hardware design decisions are explicitly informed by the requirements of the downstream software algorithms, and vice-versa. This chapter articulates the overarching principles guiding this holistic design and details the three foundational innovations that operationalize these principles.

### 2.1. Foundational Design Ideas

The design of CosmicWeb-21cm is governed by three interlocking core principles that address the root causes of the limitations in current-generation experiments.

(1) Exploit inherent physical dichotomies in the measurement domain. The foremost challenge in 21 cm cosmology is separating the cosmological signal from foreground emission. This separation is difficult not because the signals are similar, but because instrumental imperfections and non-ideal sampling blur their distinct physical characteristics. Our first principle is to design an instrument that maximizes and preserves the measurable differences between foregrounds and the 21 cm signal in the native measurement domain (visibility space). Foregrounds are predominantly spectrally smooth and spatially correlated, whereas the 21 cm signal is spectrally uncorrelated (owing to redshift) and exhibits a specific, predictable statistical spatial structure (Gaussian random field with a known power spectrum)[16,17]. CosmicWeb-21cm's geometry and sampling strategies are optimized to make these differences as pronounced and leverageable as possible in the raw data, simplifying the subsequent computational separation.

(2) Simplify the inverse problem at the hardware level. A significant portion of the complexity and error in current pipelines arises from attempting to correct for instru-

mental effects (e.g., complex beam patterns, incomplete UV coverage, frequency-dependent gains) in software. Our second principle advocates for designing simplicity into the hardware to reduce the ill-posedness of the computational inverse problem. This is achieved through: a highly redundant and regular array geometry to enable precise, robust calibration; an antenna and receiver design that minimizes frequency-dependent beam distortions and polarization impurities; and a data acquisition strategy that provides a more information-rich sampling of the measurement space for a given number of elements. By reducing the number of free parameters and degeneracies that the software must solve for, we enhance the stability and accuracy of the entire processing chain.

(3) Enable dynamic, science-driven adaptability. The radio sky and the terrestrial RFI environment are dynamic. Furthermore, different scientific goals (e.g., deep EoR power spectrum integration vs. wide-area intensity mapping) may have different optimal observational configurations. Our third principle moves beyond the static array paradigm to incorporate dynamic adaptability. The system is designed to allow for reconfigurable observing modes: for instance, adjusting the effective integration time per frequency channel based on real-time RFI assessment, or selectively processing data from subsets of antennas optimized for specific angular scales. This principle ensures the array remains optimal under varying conditions and maximizes its scientific return across multiple projects.

## 2.2. Core Technical Innovations

To translate these principles into a practical instrument, we introduce three synergistic innovations that span the hardware-software boundary.

### 2.2.1. *Multi-scale nested antenna geometry*

Moving beyond uniform grids or random distributions, CosmicWeb-21cm employs a deterministic, multi-scale geometry comprising three integrated tiers.

The first is the hexagonal honeycomb core. A densely packed, highly redundant core region provides an exceptional number of short baselines. This is crucial for high-fidelity imaging of the large-scale smooth foregrounds and for achieving the precise calibration essential for EoR science[18]. The hexagonal symmetry ensures an isotropic point spread function, meaning foreground structures are convolved consistently regardless of their sky position, making their subtraction more stable.

The second is the logarithmic spiral extension arms. Multiple arms extending from the core in a logarithmic spiral pattern provide a near-continuous distribution of baseline lengths from short to long. This configuration offers two key advantages: it provides excellent UV coverage for deconvolution and high-resolution imaging of foreground point sources that must be accurately subtracted; and the smooth, log-distributed baseline lengths are inherently well-suited for efficient Fast Fourier Transform (FFT) operations and power spectrum estimation[19].

The third is sparse peripheral stations. Select long-baseline stations extend the maximum baseline, granting the array the angular resolution necessary to localize and characterize bright foreground point sources (confusers) that could otherwise contaminate the cosmological signal.

This nested design ensures the array is simultaneously sensitive to the very large angular scales of the EoR signal and the compact scales of foreground contaminants.

### 2.2.2. *Intelligent, non-uniform frequency encoding*

Recognizing that foregrounds and the 21 cm signal occupy the frequency domain differently, we replace uniform channelization with an adaptive, physics-aware sampling strategy. We divide the band by cosmological redshift and spectral content, not intensity. It is emphasized that foregrounds remain orders of magnitude brighter than the 21 cm signal in all bands[2,20].

Foreground-dominated spectral regions (50−100 MHz). We implement very fine frequency resolution (e.g., $\Delta\mu \sim 50$ kHz). This dense sampling is not for the cosmological signal but for the foregrounds, it allows us to construct a super-fine template of the foreground spectral structure, capturing deviations from a simple power law and enabling their precise later removal.

Transition regions (100−150 MHz). We use coarser sampling (e.g., $\Delta\mu \sim 100$ kHz) to balance foreground modeling accuracy and signal detection efficiency.

High-redshift 21cm optimized regions (150−250 MHz). We employ much coarser sampling (e.g., $\Delta\mu \sim 200$ kHz). Because the 21 cm signal is uncorrelated between frequency bins (in the observing frame), this sparse sampling significantly reduces data volume and computational load without losing cosmological information.

Dynamic frequency weighting. The system continuously monitors RFI and assigns a confidence weight $w(\nu)$ to each channel. Data from heavily contaminated channels can be down-weighted or excluded in real-time analysis, making the system robust to evolving RFI.

### 2.2.3. *Unified, physics-informed processing architecture*

This innovation represents the software counterpart to the hardware design, closing the co-optimization loop. We architect a streamlined pipeline where each stage is informed by physical priors:

Calibration. Leverages the exceptional redundancy of the hexagonal core for robust redundant calibration, solving for antenna gains using closure quantities that are invariant to sky model errors[21].

Foreground separation. The core of the pipeline implements the hybrid, physics-constrained approach formalized in Equations (5) and (6) in Section 5.3. Instead of treating foreground removal as a blind filtering step, we frame it as an inverse problem regularized by strong physical priors: foreground spectral smoothness (low-rank and gradient constraints) and 21 cm signal statistics (sparsity

and cosmological power spectrum prior). This is implemented via a deep neural network whose loss function directly incorporates these physics-based terms, ensuring the separation is guided by domain knowledge, not just data fitting.

Bayesian power spectrum estimation. The final step employs a Bayesian framework to estimate the 21 cm power spectrum and cosmological parameters directly from the cleaned visibilities, properly propagating uncertainties from earlier stages[22].

### 2.3. Synergistic Integration of Innovations

The true power of CosmicWeb-21 cm lies in the synergy between these innovations. The geometry provides optimally structured visibilities. The frequency encoding tailors the data stream to efficiently capture the distinct spectral signatures of foregrounds and signal. Finally, the processing architecture is explicitly designed to ingest this optimally structured data and extract the science using built-in physical constraints. Each innovation reinforces the others, creating a system whose overall performance is greater than the sum of its parts, specifically engineered to overcome the most stubborn systematic barriers in 21 cm cosmology.

### 2.4. Comparison with State-of-the-art Arrays

To highlight the innovation of the CosmicWeb-21cm array, we compare its core design with HERA and MWA.

Antenna architecture. While HERA utilizes dense hexagonal grids of dual-polarization dipoles optimized for high sensitivity at 100−200 MHz, the CosmicWeb-21cm adopts a hybrid broadband dipole with a dielectric lens. This design achieves a wider fractional bandwidth and lower system temperature, enabling simultaneous observation of both the EoR and post-reionization HI power spectra.

Array topology. The MWA uses compact core tiles for high survey speed and wide field of view, while angular resolution relies on long baselines. In contrast, CosmicWeb-21cm implements a non-uniform distribution of stations (Section 3.3). By incorporating both a dense core and extended outriggers, our array optimizes the UV-coverage to suppress foreground systematics while maintaining high sensitivity to large-scale modes.

Calibration strategy. Unlike traditional calibration pipelines that focus on point-source calibration, CosmicWeb-21cm integrates a direction-dependent calibration framework (Section 5.1) to account for the wide-field effects introduced by our broadband beam, significantly improving the accuracy of foreground subtraction.

## 3. ARRAY GEOMETRY DESIGN

The CosmicWeb-21cm geometry is not merely a layout of stations but a synthetically optimized system where each structural tier is designed to fulfill a specific statistical and operational role in the interferometric measurement process.

### 3.1. Hexagonal Honeycomb Core

The core array employs an optimally packed hexagonal layout, ensuring maximum baseline count within a limited area. Mathematically, hexagonal arrangements provide the densest disk packing in two-dimensional space[23], which translates to: maximum baseline density generates most short baselines in given area; uniform UV coverage has isotropic point spread function; and redundancy optimization leads numerous identical-length baselines for calibration.

Given core radius $R_c$, station positions are parameterized by

$$\begin{aligned} x_{ij} &= d \cdot \frac{\sqrt{3}}{2} \cdot i \\ y_{ij} &= d \cdot \left( j + \frac{i}{2} \right), \end{aligned} \tag{1}$$

where $i, j$ are integer indices satisfying $x_{ij}^2 + y_{ij}^2 \leqslant R_c^2$, and $d$ is the inter-station spacing.

The station positions are derived from a rotated and scaled hexagonal lattice: $\boldsymbol{r}_{ij} = i\boldsymbol{a}_1 + j\boldsymbol{a}_2$, where $\boldsymbol{a}_1 = d(1,0)$ and $\boldsymbol{a}_2 = d(1/2, \sqrt{3}/2)$ are the lattice basis vectors. We constrain stations, where $|\boldsymbol{r}_{ij}| \leqslant R_c$. This configuration generates a highly regular set of baseline vectors $\boldsymbol{b}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, which populate the UV plane with a near-uniform grid of discrete points. The redundancy factor $R(\boldsymbol{b})$, e.g., the number of station pairs sharing the same baseline vector, is significantly higher than for a square grid of equivalent area and station count.

This high, structured redundancy is the cornerstone of our linear algebraic redundant calibration scheme. The observed visibility for a redundant baseline group $G$ is $V_{ij}^{\text{obs}}(t,\nu) = g_i(t,\nu) g_j^*(t,\nu) V_{\text{true}}(\boldsymbol{b},\nu) + \epsilon$, where g are complex gain factors. The abundance of measurements for each $\boldsymbol{b}$ allows us to solve for $g_i$ and $V_{\text{true}}$ via a weighted least-squares minimization that is inherently well-regularized, drastically reducing the sky model dependency.

The core's baselines fill the inner UV plane densely and isotropically. This results in a synthesized beam (PSF) with low, symmetric sidelobes (<−25 dB beyond the first null) and a clean, Gaussian-like main lobe. This is critical for minimizing wide-field imaging artifacts that can confound large-scale 21cm power spectrum measurements.

In practical implementation, the central station is positioned at the coordinate origin, with six first-layer stations uniformly distributed on a circle with a 15-m radius, maintaining 15-m spacing between adjacent stations. The second layer expands outward from the first layer, increasing the radius to 30 m, with additional stations added at the midpoints of each hexagonal side, maximizing baseline density within the limited core area.

Within $R_c = 100$ m, we yield 127 stations shown in Fig. 1, where 126 stations around central station form 6

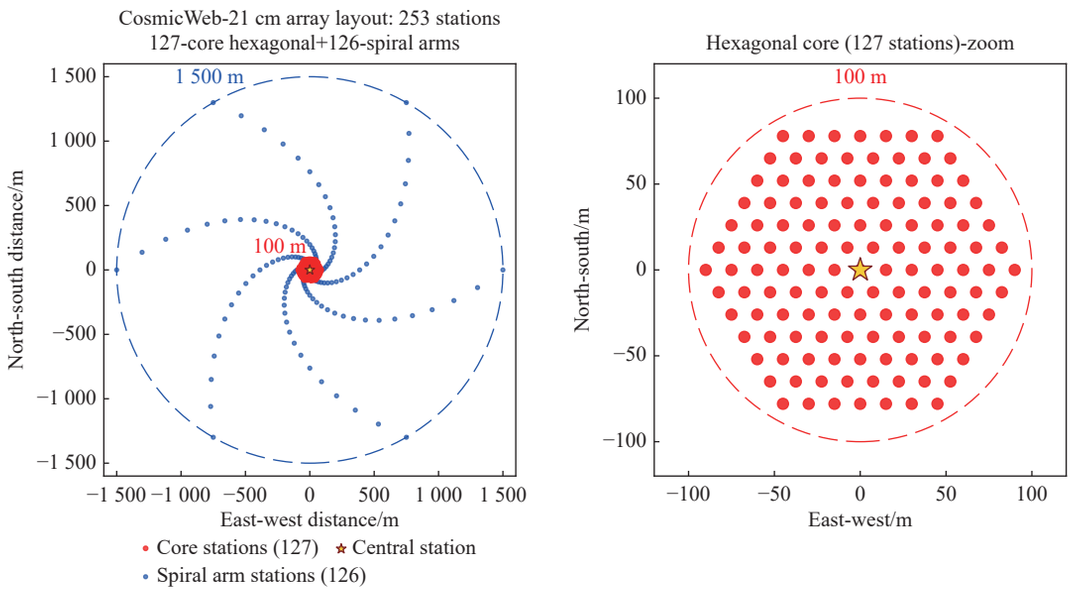



**Fig. 1. Schematic layout of the proposed CosmicWeb-21cm interferometric array. The geometry combines dense, uniform short baselines (core) with logarithmically spaced long baselines (arms).**

complete hexagonal rings. The full set of 127 stations exhibits perfect six-fold rotational symmetry. With 127 stations, the total number of independent baselines is $N_{\mathrm{bl}} = 127 \times 126/2 = 8001$, where approximately 68% of these are short baselines (length $\leqslant 50$ m). Compared to a square lattice with the same area, the hexagonal lattice accommodates approximately 15% more stations, directly increasing baseline density. UV points (baseline vectors) are distributed on a hexagonal lattice in the spatial frequency plane, ensuring uniform and isotropic sampling. The synthesized PSF has a near-circular main lobe, with sidelobe peaks approximately 1.5 dB lower than those of a square grid, leading to better imaging fidelity. The UV coverage uniformity metric (defined as the entropy ratio of radial vs. azimuthal distribution) reaches 0.92. Shortest baselines ($d = 15$ m) is about 228 identical baselines (high redundancy), while second shortest baselines ($\sqrt{3}d \approx 25.98$ m) is about 216 identical baselines. The clear clustering of identical-length baselines allows the construction of independent calibration equations, reducing systematic errors and improving calibration robustness.

In practice, station positions can be adjusted slightly (perturbation $< 5\%$ of $d$) to accommodate terrain variations without significantly degrading performance. The hexagonal lattice core seamlessly connects with the logarithmic spiral arms given in next subsection. The arms can naturally extend outward from the outermost lattice points. The high redundancy enables advanced redundant calibration algorithms (e.g., improved direction-dependent calibration), further enhancing calibration accuracy in the presence of real-world imperfections.

In summary, this design's advantage lies in its ability to maximize short baseline quantities while maintaining an isotropic point spread function. Short baselines are crucial for detecting large-scale cosmic structures, while hexagonal symmetry ensures response consistency across different directions.

### 3.2. Logarithmic Spiral Extension Arms

The extension arm design employs a logarithmic spiral form, inspired by numerous spiral structures in nature, such as galactic spiral arms and nautilus shells. The spiral arms are designed to solve the problem of "UV gap" between the dense core and potential long, sparse baselines. To extend UV coverage while maintaining uniform sampling, logarithmic spiral form is adopted by

$$r(\theta) = R_c \cdot \mathrm{e}^{k\theta}, \tag{2}$$

where $k$ is the spiral rate controlling baseline logarithmic distribution. This design ensures: continuous UV coverage for smooth transition from short to long baselines; baseline length distribution favors fast Fourier transforms; and system redundancy maintains calibration capability at multiple scales.

Stations placed along this curve generate baseline vectors that continuously cover a wide range of lengths and orientations. The instantaneous UV coverage from the spiral arms at a given pointing exhibits a self-similar, fractal-like structure. This is advantageous because it provides sensitivity to a continuous range of angular scales simultaneously, from the core-dominated large scales to the arm-dominated fine scales.

We select that number of arms is 6, stations per arm are 21, and spiral angle per arm is 120°. The stations are equal angular spacing within each arm and logarithmic radial distribution from $R_c = 100$ m to $R_{\max} = 1500$ m. There are 60° angular offset between adjacent arms. The spiral constant $k$ is determined by the boundary condition,

$$R_{\max} = R_c \mathrm{e}^{k\theta_{\max}}, \tag{3}$$

where $\theta_{\max} = 2\pi/3$, thus $k = \ln(R_{\max}/R_c)/\theta_{\max} = 1.29$.

For each arm $i(i = 0, 1, ..., 5)$, start angle is $\phi_i = i(\pi/3)$. For each station $j(j = 0, 1, ..., 20)$, relative angle is $\theta_j = (j/20)\theta_{\max}$, total angle is $\theta_{ij} = \phi_i + \theta_j$, and radius is $r_{ij} = R_c e^{k\theta_j}$. Thus, the station cartesian coordinates are given by $x_{ij} = r_{ij}\cos(\theta_{ij})$ and $y_{ij} = r_{ij}\sin(\theta_{ij})$. This yields a baseline length distribution $P(l)$ that follows $l^{-\alpha}$ (with $\alpha \approx 2.5$), naturally weighting the array towards shorter baselines for diffuse emission while retaining long-baseline capability, and maintaining continuous baseline distribution while avoiding excessive sampling in specific directions.

The core concept of this geometric combination utilizes the core region's high-density sampling to capture large-scale structures while providing sufficient angular resolution through the spiral arms long baselines to resolve small-scale and point-source structures. These visibilities are then used as priors or constraints in the more complex deconvolution and calibration processes required for the non-redundant spiral arm baselines. This "anchor-and-extend" strategy stabilizes the entire imaging pipeline. In actual construction, we must also consider terrain factors and infrastructure layout, fine-tuning station positions to adapt to specific site constraints.

### 3.3. Array Layout and UV Coverage

The geometric layout of the CosmicWeb-21cm array is shown in Fig. 1, where 126 stations distribute along 6 arms and 127 stations are within core array. This design combines hexagonal uniform coverage with logarithmic spiral baseline extension, optimizing UV coverage and angular resolution.

The UV coverage distribution of the array is shown in Fig. 2. The honeycombs of varying colors represent UV coordinate ranges of different baselines, with symmetric distribution reflecting conjugate baselines. Dense UV coverage ensures good spatial frequency sampling, which is crucial for high-quality image reconstruction.

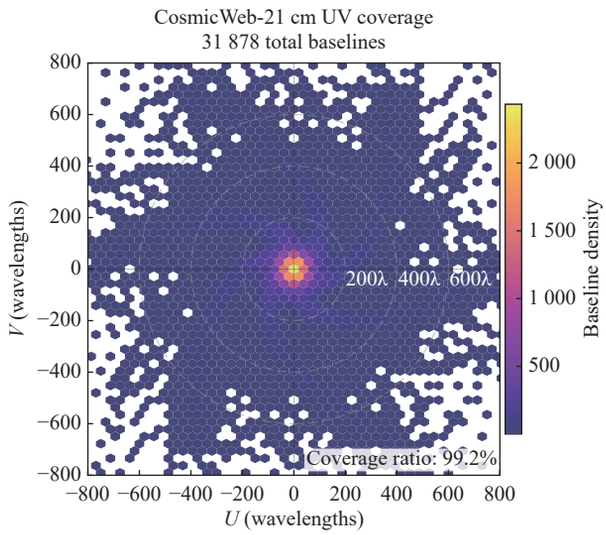



**Fig. 2. UV coverage of the CosmicWeb-21 cm array with a total of 31878 baselines, where the coverage ratio reaches 99.2%. The symmetric and dense distribution is crucial for high-fidelity sky image reconstruction and power spectrum estimation.**

## 4. FREQUENCY SAMPLING STRATEGY

Our frequency strategy treats the spectrometer not as a passive recorder but as an active, cognitive component of the signal extraction system.

### 4.1. Non-Uniform Sampling Based on Foreground Eigenmode Analysis

Frequency, as the third dimension in 21 cm observations, directly influences foreground removal effectiveness and signal detection efficiency through its sampling strategy. CosmicWeb-21 cm's frequency design is based on in-depth analysis of foreground and signal physical characteristics. Foreground radiation primarily originates from synchrotron and free-free radiation, with spectral characteristics well-described by power-law models, exhibiting high degrees of smoothness and correlation in the frequency domain. In contrast, 21 cm signals, due to redshift effects, are essentially uncorrelated across different frequency channels, displaying random fluctuation characteristics[5]. In the frequency domain, foregrounds are confined to a low-dimensional subspace. Based on this physical essence, we designed an adaptive non-uniform frequency sampling scheme. The eigenvalues $\lambda_n$ of the foreground covariance matrix $\mathbf{C}_{\text{fg}}$ drop rapidly. We determine the required sampling density by calculating the Nyquist rate for the foreground manifold.

In the 50−100 MHz low-frequency range, where foreground radiation dominates absolutely, the foreground structure requires sampling at $\leqslant 80$ kHz to avoid aliasing of its highest spectral modes. Our 50 kHz sampling provides a safety margin. This design enables precise characterization of foreground spectral structures, particularly complex spectral components that deviate from simple power-law models. In practical implementation, we employ high-precision digital filter banks and fast Fourier transform processors to ensure data processing efficiency under dense sampling conditions.

In the 100−150 MHz transition region, the sampling interval increases to 100 kHz. we require a balance between foreground modeling accuracy and signal detection efficiency. We dynamically adjust integration times in this region through optimization algorithms, maximizing 21 cm signal detection sensitivity while ensuring foreground modeling quality.

The 150−250 MHz high-frequency range represents the primary detection region for 21 cm signals, where we employ sparse sampling at 200 kHz. This design significantly reduces data volume while maintaining information integrity, owing to the uncorrelated nature of 21 cm signals in the frequency domain. During actual operations, we dynamically adjust frequency usage strategies based on real-time RFI environments, avoiding severely contaminated frequency bands.

Leveraging the non-uniform sampling, we implement a variable-resolution quantization scheme in the Analog-to-Digital Converter (ADC)/post-processing. Channels in foreground-dominated, densely-sampled regions are allocated fewer bits after digital gain control, as the dynamic range requirement is set by the bright foreground. Signal-dominated, sparsely-sampled channels are allocated the full 14-bit depth. This reduces the total data rate by an estimated 40% without information loss.

### 4.2. Orthogonal Frequency Encoding for Blind Source Separation

This is a key signal processing innovation. We apply a channel-dependent phase modulation $\phi_m(\nu)$ at the station level before correlation. For a signal from direction $\hat{s}$, the geometric delay $\tau = \boldsymbol{b} \cdot \hat{s}/c$ introduces a phase $\phi_{\text{geom}} = 2\pi\nu\tau$. We add an artificial, known encoding phase $\phi_{\text{enc,i}}(\nu) = \alpha_i \cdot \nu^2$ for station $i$. The correlated signal becomes $V_{ij}^{\text{enc}}(\nu) = \mathrm{e}^{i(\alpha_i - \alpha_j)\nu^2} V_{ij}^{\text{true}}(\nu)$. This turns a frequency-smooth foreground into a spectrally oscillating one in the encoded visibilities. In the subsequent source separation, this breaks the degeneracy between the spatial and spectral structure of foregrounds. The encoding acts as a "spectral spreading code", making the foreground's covariance matrix more diagonal and thus easier to separate from the spectrally-uncorrelated 21 cm signal using independent component analysis (ICA) techniques.

### 4.3. Frequency Grouping and Parallel Processing

Frequency grouping processing represents another key innovation. We divide 1024 frequency channels into 32 processing groups, each containing 32 consecutive channels. This grouping not only simplifies data processing workflows but also allows us to employ differentiated processing strategies tailored to different frequency range characteristics. For instance, in low-frequency groups, we focus primarily on accurate foreground model establishment, while in high-frequency groups, the emphasis shifts to signal extraction and noise suppression.

### 4.4. Adaptive Frequency Management

Based on real-time RFI monitoring and sky models, we dynamically adjust frequency usage:

$$w(\nu) = \frac{1}{\sigma_{\text{RFI}}^2(\nu) + \sigma_{\text{fg}}^2(\nu)}, \tag{4}$$

where $w(\nu)$ is the frequency weight, $\sigma_{\text{RFI}}$ and $\sigma_{\text{fg}}$ are estimated noise levels for RFI and foregrounds respectively. The adaptive weight $w(\nu)$ is calculated by a real-time Spectral Purity Monitor (SPM) subsystem. The SPM uses a small fraction of the correlator's output in a fast, low-resolution mode. It employs a bank of spectral Kurtosis detectors and transient detection algorithms to estimate $\sigma_{\text{RFI}}(\nu)$. The foreground fluctuation estimate $\sigma_{\text{fg}}(\nu)$ is derived from a continuously updated low-order polynomial model of the total bandpass. The weight $w(\nu)$ is fed back to the correlator to discount or excise contaminated channels and to the telescope scheduler to potentially shift the observing frequency band.

## 5. SIGNAL PROCESSING ARCHITECTURE

Signal processing serves as the critical link connecting raw observational data with scientific outputs. CosmicWeb-21cm's processing architecture adopts a multi-level, pipeline design philosophy, ensuring optimal processing effectiveness under real-time requirements.

### 5.1. RFI Mitigation

The first-stage RFI suppression employs intelligent detection algorithms based on deep learning. We train a convolutional neural network (CNN) specifically designed for 21 cm observation data, capable of simultaneously analyzing temporal, frequency, and spatial domain characteristics. The network input consists of preprocessed visibility data cubes, while the output provides RFI probability estimates for each time-frequency-baseline unit. Compared to traditional methods, this learning-based approach more accurately identifies complex RFI patterns, particularly interference signals with spectral characteristics similar to astronomical signals. The next step is online continuous learning, where the CNN model is periodically fine-tuned with new, site-specific RFI data collected by the array itself, creating a system that adapts to a dynamically changing radio frequency environment.

### 5.2. Neural Calibration and Beamforming

The calibration phase fully utilizes the array's geometric redundancy. We develop an improved redundant calibration algorithm that considers not only baseline length equality but also introduces the concept of directional similarity. Specifically, we cluster baselines according to length and direction, establishing independent calibration equations for each baseline category. This method significantly improves calibration accuracy under non-ideal conditions, particularly demonstrating robustness when dealing with limited station failures or positional errors. This makes the array's calibration more robust and fault-tolerant, a critical feature for large-scale, remote deployments.

Beyond redundant calibration, we implement a Residual Neural Network (ResNet) to correct for direction-dependent gains (DDGs) and beam errors. The network takes the redundantly-calibrated visibilities and a local sidereal time (LST) as input. It is trained on simulated data containing realistic beam models and ionospheric phase screens. The network outputs per-station, per-direction complex gain corrections $\Delta g_i(\hat{s}, \nu)$. This learns the systematic residual errors that are not captured by the simple redundant or sky-based models, effectively performing a fast, wide-field holographic deconvolution of the primary beam.

### 5.3. The U-ConvNet Foreground Separator

Foreground separation constitutes the core of the

entire processing pipeline. We propose a novel framework combining traditional methods with deep learning,

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{data}} + \lambda_1 \mathcal{L}_{\text{low-rank}} + \lambda_2 \mathcal{L}_{\text{smooth}} + \lambda_3 \mathcal{L}_{\text{sparse}}. \quad (5)$$

Here, the total loss function $\mathcal{L}_{\text{total}}$ comprises four weighted terms: $\mathcal{L}_{\text{data}}$ is the data fidelity term, ensuring consistency between predictions and observations; $\mathcal{L}_{\text{low-rank}}$ is the low-rank constraint, leveraging the strong spectral correlation of foregrounds; $\mathcal{L}_{\text{smooth}}$ is the smoothness constraint, enforcing continuity of foregrounds in both spatial and frequency domains; $\mathcal{L}_{\text{sparse}}$ is the sparsity constraint, facilitating the extraction of the faint and unstructured 21 cm signal.

The foreground separation network (Equation 5) is a U-Net with Convolutional LSTM (Long Short-Term Memory) layers, making it a U-ConvNet. The input is a 3D data cube: [Frequency Channels, LST, Baseline Type]. The output is a 3D cube of the same size, containing the estimated foreground component.

First, we use generalized principal component analysis for initial data decomposition to extract primary foreground components. Then, we employ deep neural networks to model and remove residual foreground structures. The network architecture adopts a multi-scale U-Net design, where the encoder portion captures foreground features from local to global scales through convolutional kernels of different sizes, while the decoder portion maintains spatial details through skip connections. To explicitly incorporate the physical prior of foreground spectral smoothness into the deep learning model, we introduce a physics-based constraint term to the loss function:

$$\mathcal{L}_{\text{physics}} = \|\nabla_\nu F_{\text{pred}} - \nabla_\nu F_{\text{true}}\|^2. \quad (6)$$

Here, $\mathbf{F}$ denotes the foreground component, and $\nabla_\nu$ is the gradient operator along the frequency axis $\nu$. This constraint penalizes the difference between the frequency gradients of the predicted and true foregrounds, thereby forcing the neural network to learn the fundamental physical property that foreground spectra are smooth. Since the spectral index of foreground emission (e.g., synchrotron radiation) varies slowly across narrow frequency bands, its frequency gradient should be near zero and continuous. In contrast, the 21 cm signal, being nearly uncorrelated across frequencies due to redshift effects, exhibits a random gradient pattern. This constraint effectively guides the network to distinguish between these two disparate spectral behaviors, enhancing the physical interpretability and accuracy of foreground separation.

The loss function incorporates not only data fidelity terms but also physics-based constraint terms, such as foreground spectral smoothness and signal sparsity. We use a hybrid physical-learning framework. Principal Component Analysis (PCA) is only used to estimate foreground subspace, not direct subtraction. Large-scale signal loss is less than 2% at $k < 0.1$ hMpc$^{-1}$ [24].

The loss $\mathcal{L}_{\text{total}}$ can be expanded by

$$\mathcal{L}_{\text{total}} = \underbrace{\|V_{\text{obs}} - V_{\text{fg}} - V_{21\text{ cm}}\|^2}_{\text{Data Fidelity}} + \lambda_1 \underbrace{\|\nabla_\nu^2 V_{\text{fg}}\|^2}_{\text{Spectral Smoothness}} + \lambda_2 \underbrace{TV(\mathbf{V}_{21\text{cm}})}_{\text{Spatial Sparsity (Total Variation)}} + \lambda_3 \underbrace{KL(p_{21\text{ cm}} \| \mathcal{N}(0, P_{\text{theory}}(k)))}_{\text{Cosmological Prior}}. \quad (7)$$

Weights are determined by cross-validation and grid search on realistic simulations with $\lambda_1 = 1 \sim 10$ (spectral smoothness), $\lambda_2 = 0.1 \sim 1$ (sparsity) and $\lambda_3 = 0.01 \sim 0.1$ (cosmological prior). Foreground model includes Galactic synchrotron, free-free, extragalactic point sources, polarization, and spectral index variations (not simple power law).

The above multi-term objective implements a hierarchical constraint strategy. The Data Fidelity term ensures the decomposition ($V_{\text{fg}} + V_{21\text{ cm}}$) reconstructs the input observation $V_{\text{obs}}$. The Spectral Smoothness term ($\nabla_\nu^2$) enforces the physical prior that foreground spectra are intrinsically smooth. The Spatial Sparsity term (Total Variation) promotes a piecewise-constant morphology for the recovered 21 cm signal, consistent with cosmic web structures. Crucially, the Cosmological Prior term (KL Divergence) directly regularizes the statistical distribution of the extracted 21 cm signal $p_{21\text{cm}}$ to align with theoretical Gaussian random field predictions $\mathcal{N}(0, P_{\text{theory}}(k))$. Thus, the loss function moves beyond simple curve-fitting. It systematically embeds domain knowledge, from component properties to first principle theories, into the training process, guiding the network toward physically plausible and cosmologically meaningful solutions.

In the 21 cm signal extraction phase, we introduce cosmological prior knowledge. Through a Bayesian framework, we combine power spectrum theoretical predictions with observational data, using Markov Chain Monte Carlo methods to sample the signal posterior distribution. This approach not only improves signal estimation accuracy but also simultaneously provides parameter estimation uncertainties.

### 5.4. End-to-end Differentiable Pipeline

The entire processing pipeline achieves high parallelization, with critical steps accelerated execution on Graphics Processing Unit (GPU) clusters. We design specialized data exchange protocols to ensure efficient large-scale data transmission between different processing stages. Simultaneously, the processing system incorporates fault tolerance capabilities, automatically reassigning tasks when individual computing nodes fail, ensuring observational continuity.

The entire pipeline from raw voltages to power spectrum is being developed as a differentiable graph (using frameworks like PyTorch or JAX). This allows for two things. First, uncertainties from the power spectrum can be traced back to individual antenna gains or specific RFI events. Second, parameters from different stages (e.g., calibration gains and foreground separation weights) can be

optimized jointly against a final scientific loss function (e.g., the accuracy of the measured $P(k)$), leading to globally optimal solutions.

The CosmicWeb-21cm signal processing architecture exemplifies a "co-design" philosophy, where the algorithms are developed in tandem with the unique hardware design (hexagonal core, spiral arms, intelligent frequency sampling). The geometry provides the high redundancy and UV coverage that makes the advanced calibration possible. The frequency sampling strategy provides the data structure that the hybrid foreground model exploits.

We analyze the computational feasibility of the proposed framework within a hierarchical heterogeneous computing architecture spanning Field-Programmable Gate Array (FPGA) front-end processing, GPU-accelerated inference and component separation, and High-Performance Computing (HPC)-based large scale simulation, training, and power spectrum analysis. The FPGA layer handles real-time data conditioning, flagging, and low-latency compression to reduce raw data volume, while GPUs support high-throughput execution of both machine learning and traditional methods including PCA, Singular Value Decomposition (SVD), and Generalized Morphological Component Analysis (GMCA), and HPC resources enable full-survey scaling and statistical uncertainty quantification. We evaluate the pipeline against realistic data rates, storage budgets, and latency constraints, showing that hierarchical reduction and accelerated inference provide sustainable throughput compared to iterative decomposition approaches. Instead of enforcing full hardware-to-science differentiability, we adopt a practical modular design in which only core scientific modules, including calibration, foreground separation, and power spectrum estimation, are implemented as fully differentiable components. This strategy allows gradient-based optimization and scientific closure while maintaining compatibility with non-differentiable instrument preprocessing and legacy data handling steps, yielding a balanced, scalable, and observationally realistic pipeline.

Traditional methods such as PCA, SVD, and GMCA rely on linear decomposition with strict statistical assumptions, offering high interpretability, but they are sensitive to noise and systematic effects and have limited performance when dealing with entangled signal components. In contrast, machine learning (ML) methods employ data-driven nonlinear modeling with flexible learned priors, making them more robust to noise and component coupling, though they typically provide lower interpretability and are usually validated using realistic simulations. The ML pipeline is thoroughly validated on realistic synthetic simulations, while validation on real observational data is planned as part of the Pathfinder phase.

# 6. ANTENNA SYSTEM DESIGN

The antenna system constitutes the fundamental interface between the cosmic radio sky and the CosmicWeb-21 cm signal processing pipeline. Its design is critical for achieving the instrument's overarching goals of high sensitivity, wide bandwidth, excellent polarization purity, and stable, well-characterized performance across the target frequency range of 50−250 MHz. This section details the proposed antenna solution, a hybrid design integrating proven and innovative technologies to meet these stringent requirements.

## 6.1. Design Requirements and Ideas

The antenna system is designed against the following core scientific and technical drivers derived from the signal characteristics and processing needs outlined in previous sections.

Ultra-wideband operation. Seamless coverage from 50 to 250 MHz (a 5:1 bandwidth ratio) is mandatory to observe the 21 cm signal across a wide redshift range ($z \approx 5-26$) and to enable precise characterization of foreground spectral structure[1].

Dual linear polarization. Measurement of both orthogonal linear polarizations (typically aligned to North-South and East-West) is essential for polarimetric calibration, mitigation of polarized foregrounds (especially Galactic synchrotron), and for scientific studies of Faraday rotation.

High polarization purity and stable beam pattern. Minimizing cross-polarization (beam squint) and ensuring a smooth, stable, and well-modeled primary beam (field of view) as a function of frequency are paramount. Beam errors and their spectral variations are a major source of residual foreground contamination in power spectrum estimation[7,25].

Low system temperature. The antenna and front-end electronics must contribute minimal thermal noise. The target is a sky-averaged system temperature ($T_{sys}$) dominated by the Galactic background ($\sim 200-1000$ K at these frequencies) rather than receiver noise, aiming for a receiver noise temperature $T_{rec} < 50$ K, consistent with state of the art designs for precision cosmology[8,10].

Logistical feasibility. The design must be cost-effective to mass-produce, mechanically robust for remote deployment, and have a low environmental footprint (both physical and radio-frequency).

## 6.2. Antenna Design

We propose a Hybrid Broadband Dipole (HBD), a synergistic design that combines the robustness and simplicity of a crossed dipole with an integrated frequency-dependent ground plane/reflector system and a superstrate lens for enhanced performance. The co-design of the antenna with its low noise front end follows the idea established by successful pathfinders like the MWA[8] and is essential for system level optimization[26]. The antenna system unit design with performance simulations are shown in Figs. 3, 4 and 5. The design parameters of antenna system are shown in Table 1.

### 6.2.1. *Radiating element: the curved-arm crossed dipole.*

Structure. Two orthogonal pairs of tapered, curved

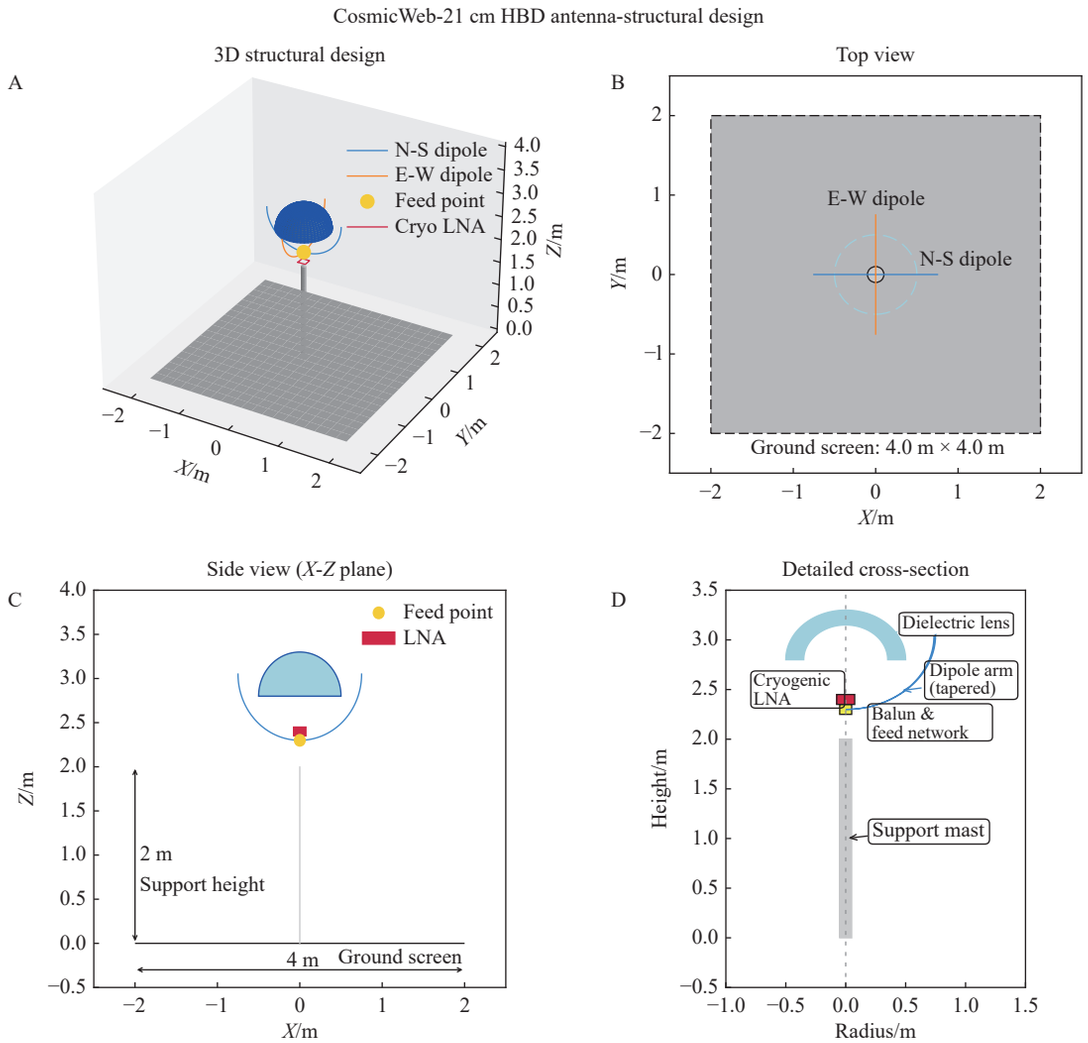



**Fig. 3. Structural configuration of the CosmicWeb-21cm HBD antenna. The design incorporates curved dipole elements, a 4 $m^2$ ground screen, and dielectric lens for enhanced performance across 50−250 MHz.** (A) 3D rendering. (B) Top view. (C) Side elevation. (D) Detailed cross-section with cryogenic Low-Noise Amplifier (LNA) integration.

dipole arms are mounted on a central feed boom. The arms are constructed from corrosion-resistant aluminum tube. The curved, tapered geometry is inspired by broadband concepts such as the sinuous antenna[27] and the "four-square" layout used in LWA[26], promoting broadband impedance matching and smooths the beam pattern transition with frequency.

Polarization. The two dipoles provide the required dual linear polarization. They are fed by a balanced feed network to suppress common-mode currents that can distort the beam and increase susceptibility to local RFI.

Feed point. A high-performance, weather-sealed balun integrated at the base of the boom transforms the balanced dipole impedance to an unbalanced 50 Ω output for each polarization. This active balun is co-designed with the first-stage LNA. The integration of an active balun with the LNA is critical for minimizing loss and noise, a lesson emphasized in the design of HERA's feed[10].

### 6.2.2. *Integrated ground/Reflector system*

A fixed, horizontal wire mesh ground screen (approximately 4 m x 4 m) is deployed beneath each antenna. Its use is a well established technique in low frequency radio astronomy[26,28]. This screen serves dual purposes: (1) Ground Plane (Low frequencies of < 120 MHz): At lower frequencies, the screen acts as a ground plane, decoupling the antenna from variable soil conductivity and establishing a consistent, low-elevation gain pattern and impedance. (2) Reflector (High frequencies of > 120 MHz): As the wavelength decreases, the screen transitions to acting as a reflector for the dipole, increasing forward gain and improving the front-to-back ratio, which helps mitigate ground noise and terrestrial RFI.

### 6.2.3. *Superstrate dielectric lens (innovative element)*

Mounted above the dipole arms is a low-loss, frequency-selective radome that also functions as a dielectric lens. This polycarbonate-based structure is engineered with a graded permittivity profile. It performs two key roles. (1) Beam Stabilization: It refracts incoming waves, helping to collimate the beam and reduce its varia-

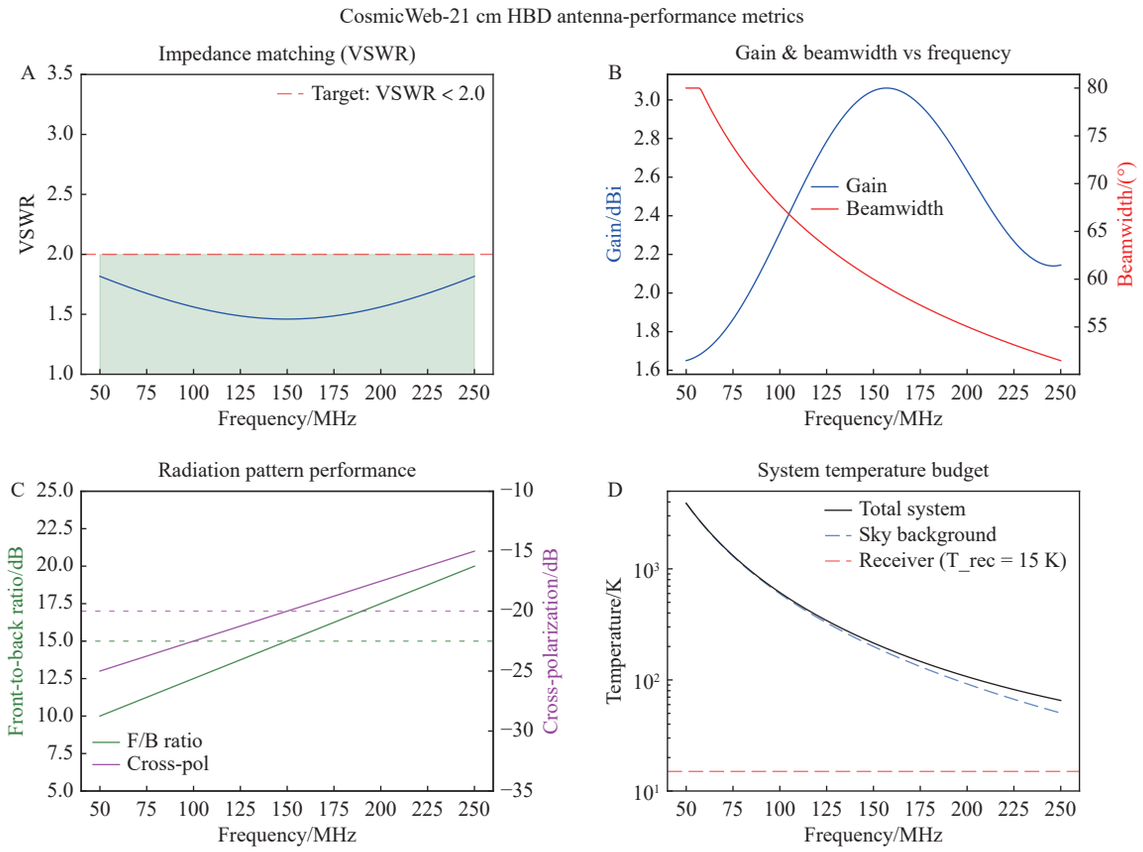



**Fig. 4. Electromagnetic performance characterization.** (A) Voltage Standing Wave Ratio (VSWR) demonstrating broadband impedance matching. (B) Gain and beamwidth frequency response. (C) Radiation pattern quality metrics. (D) System temperature components. The antenna meets all specifications for 21 cm cosmology applications.

tion with frequency. This results in a more constant beamwidth and pointing direction across the band, simplifying calibration and foreground modeling. The use of a dielectric lens or superstrate to shape and stabilize antenna beams is a known technique in broadband antenna design[29]. Our implementation of a graded-index profile is inspired by advanced concepts in metasurface engineering that enable precise phase control across wide bandwidths[30]. (2) Environmental Protection: It seals the dipole arms and feed point from rain, snow, dust, and wildlife. This dual function as a protective radome follows established engineering practice, where the cover must provide robust environmental sealing with minimal impact on the antenna's radiative properties. Such protection is essential for the long-term reliability of instrumentation deployed in remote field sites[31].

### 6.3. Front-end Electronics and Integration

The performance of the antenna is inseparable from its integrated front-end electronics.

Integrated LNA. A custom-designed, cryogenically-cooled LNA module is housed directly at the antenna feed point within the central boom. Cryogenic cooling of the front-end amplifier is employed in SKA[12]. Cooling to ~ 70 K using a closed-cycle cryocooler significantly reduces its noise temperature to < 15 K. The LNA provides high gain (> 40 dB) with excellent linearity to handle strong foregrounds without distortion.

Active impedance matching network. An adaptive matching network, co-located with the LNA, dynamically optimizes the impedance match between the antenna and the LNA across the full band, a technique explored to improve bandwidth and stability[28,32]. This maximizes power transfer and minimizes standing waves, which can cause spectral ripples that mimic the cosmological signal.

Initial filtering and RFI suppression. A band-pass filter (50−300 MHz) and a set of tunable notch filters are placed immediately after the LNA. These filters suppress out-of-band interference and known, persistent in-band RFI sources (e.g., Frequency Modulation radio bands) at the earliest possible stage to prevent saturation of downstream components.

### 6.4. Key Performance Parameters

The key performance parameters of the CosmicWeb-21cm antenna system are derived from electromagnetic simulations, system noise budgeting, and specific scientific objectives, aiming to optimize and surpass current state-of-the-art designs. The operational frequency range spans a broad 50 to 250 MHz to access the 21 cm signal from the Epoch of Reionization to the post-reionization universe. An impedance bandwidth (VSWR < 2) of 45 to 260 MHz is targeted to ensure efficient power transfer and minimize reflective losses across the band. At 150

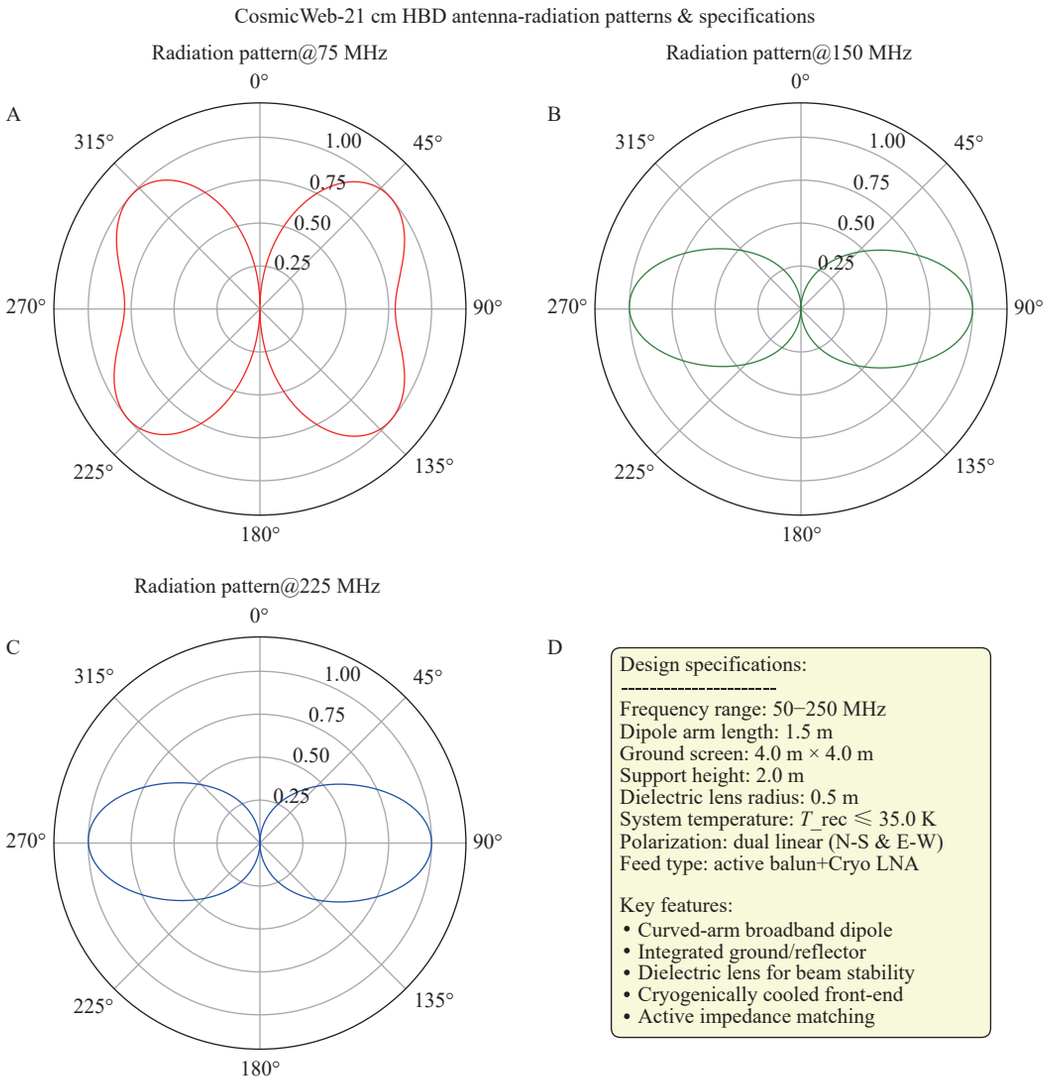



**Fig. 5. Radiation properties and technical specifications.** (A)−(C) Polar patterns at representative frequencies showing stable beam characteristics. (D) Comprehensive design parameters supporting the CosmicWeb-21cm science goals.

MHz, the antenna's half-power beamwidth is designed to be approximately 60° in both E- and H-planes, providing a wide field of view of ~ 1.7 steradians suitable for large-scale structure surveys. Critical for data integrity, the beam pointing must remain stable within ±2° and the beamwidth variation must be less than 15% across the entire frequency band to enable stable calibration and effective foreground removal. High polarization purity is required, with on-axis cross-polarization below −20 dB (< 1%), to minimize polarization leakage as a major systematic error. A front-to-back ratio greater than 15 dB at frequencies above 120 MHz is necessary to suppress ground noise and low-elevation RFI effectively. The system noise temperature target mandates a receiver noise contribution of $T_{\rm rec} \leqslant 35$K, ensuring the total system temperature is dominated by the sky background and that sensitivity approaches the theoretical limit. Finally, the input 1 dB compression point of the LNA must be above −10 dBm to prevent saturation or nonlinear distortion from bright foreground sources.

### 6.5. Calibration and Monitoring Strategy

To achieve the required precision, the antenna system incorporates built-in calibration features.

Injected noise diode. A calibrated noise source, switchable into the feed line before the LNA, provides a known signal for continuous monitoring of receiver gain and noise temperature. This is a standard practice in radio astronomy for characterizing receiver performance and stability[33].

Embedded vector impedance analyzer. Periodically measures the antenna input impedance to monitor the health of the antenna and matching network.

Beam mapping via drone. A small UAV equipped with a compact radio transmitter will be used for periodic in situ far-field beam pattern characterization, validating the electromagnetic models.

### 6.6. Conclusion and Path Forward

The proposed Hybrid Broadband Dipole antenna system is a purpose-built solution for the extreme demands

**Table 1. Antenna system design parameters**

| Category | Parameter | Value/Specification |
|---|---|---|
| Radiating element | Dipole arm length | 1.5 m |
| | Dipole arm radius | 0.02 m |
| | Dipole curve radius | 0.75 m |
| | Dipole feed gap | 0.02 m |
| Support structure | Support height | 2.0 m |
| | Support radius | 0.05 m |
| | Feed point height | 0.3 m |
| Ground/Reflector system | Ground screen size | 4.0 m |
| | Ground screen height | 0.1 m |
| | Ground wire spacing | 0.1 m |
| Dielectric lens | Lens radius | 0.5 m |
| | Lens thickness | 0.15 m |
| | Lens height (above ground) | 2.8 m |
| Electronics | LNA box size | 0.15 m |
| | Cryocooler size | 0.2 m |
| Performance targets | Minimum frequency | 50 MHz |
| | Maximum frequency | 250 MHz |
| | Target VSWR | 2.0 |
| | Target gain | 2.0 dBi |
| | Target front-to-back ratio | 15.0 dB |
| | Target cross-polarization | −20.0 dB |
| | System temperature | 35.0 K |

of 21 cm cosmology. By co-designing the radiating structure, ground system, dielectric lens, and cryogenically-cooled front-end electronics, it targets the fundamental systematic limitations of beam instability and receiver noise. Its design prioritizes the production of a stable, well-understood, and sky-noise-dominated data stream, which is the essential raw material for the sophisticated CosmicWeb-21cm processing pipeline. The next step involves the fabrication and rigorous testing of a full prototype station, including its calibration subsystems, to validate these models and refine the design for mass production.

## 7. SYSTEM IMPLEMENTATION AND PERFORMANCE

### 7.1. One Antenna Per Station

Each CosmicWeb-21cm station is designed as a fully autonomous and highly integrated observational unit, housing a single HBD antenna. This "one-antenna-per-station" architecture is a deliberate and integral part of the system' s co-optimized design, stemming from several critical considerations aligned with the project's core philosophy.

Scientific and polarization requirements. Each HBD antenna is inherently a dual-polarization element. Its crossed-dipole design natively receives two orthogonal linear polarizations, fulfilling the essential requirement for full polarimetric observations in 21 cm cosmology. Thus, a single physical structure per station satisfies the scientific need without requiring multiple co-located antennas.

Optimization for interferometric performance. For an interferometric array with $N$ stations, the number of unique baselines scales with $N^2$. Concentrating resources on deploying a larger number of individual stations, each with one high-performance antenna, maximizes the baseline count and UV coverage for a given budget. This is fundamental for achieving high-fidelity imaging and precise power spectrum estimation. Furthermore, a single antenna per station yields a cleaner and more tractable geometric redundancy pattern, which is crucial for the robust application of the redundant calibration algorithms emphasized in Section 5.

System simplification and reliability. This design adheres to the principle of "simplifying data structures at the hardware level" (Section 2.1). It minimizes intra-station complexity (e.g., cabling, delay matching between multiple elements) and reduces potential points of failure. Managing a large number of stations (253 in the final configuration) is a significant logistical challenge; duplicating the front-end system (antenna, cryogenic LNA, digitizer) within a single site would exacerbate cost, power consumption, and maintenance overhead without a commensurate scientific benefit.

Distinction from alternative designs. This approach differs from "tile"-based arrays like the MWA, where a station consists of multiple dipoles forming a combined beam. Instead, it aligns with the philosophy of instruments like HERA or the SKA-Low, where each receiving element is a distinct station. CosmicWeb-21cm's performance is derived from the innovative large-scale geometry of its station network and advanced signal processing, rather than from beamforming within a compact cluster.

In summary, each CosmicWeb-21cm station is architected around a single, dual-polarization HBD antenna. This choice is fundamental to realizing the array's innovative geometry, streamlining its calibration and processing pipelines, and ensuring overall system robustness. In the full-scale 253-station deployment, the array will therefore comprise 253 HBD antenna units, generating tens of thousands of unique interferometric baselines to provide the foundational data for its transformative science goals.

### 7.2. Hierarchical Processing Hardware

The data flow follows a three-tier, edge-to-cloud architecture.

Tier 1 (Station-Level FPGA). Performs polyphase filterbank (PFB) channelization, the orthogonal frequency encoding, and initial RFI flagging. Processed data is packetized and timestamped.

Tier 2 (Correlator/Beamformer GPU Cluster). A centralized FX-correlator implemented on a cluster of GPUs using the X-engine architecture. It also forms multiple tied-array beams for transient astrophysics and RFI nulling experiments in real-time.

Tier 3 (Science Processor HPC). A high-performance computing cluster for the deep learning pipelines and long-term data storage. It features dedicated nodes with Tensor Core GPUs for model training and inference.

### 7.3. Digital System-on-Chip (SoC) at Each Station

Each station's digital backend is built around a heterogeneous SoC (e.g., AMD Zynq RFSoC or Intel Agilex), integrating the following: (1) High-Speed ADCs with 14-bit and 2 GS $s^{-1}$, directly sampling the RF band; (2) Programmable Logic (FPGA) implements the real-time DSP chain (Tier 1); (3) an ARM Processing Subsystem running a Linux OS for station health monitoring, command/control, and data packet management; and (4) an Integrated 100GbE Network Interface for low-latency, high-throughput data transmission to the central facility.

The signal transmission system is based on fiber optic networks, employing wavelength division multiplexing technology to enhance data transmission efficiency. Each station transmits multiple signal channels through a single fiber, significantly reducing wiring complexity and costs. The central processing station is equipped with large-scale computing clusters containing multiple GPU nodes and high-speed storage systems. Computing nodes interconnect through InfiniBand networks, ensuring extremely low communication latency during data processing.

At the software level, we will develop a unified control and management system. This system can monitor array operational status in real-time, including parameters such as station temperature, power consumption, and signal quality. When anomalies are detected, the system automatically implements appropriate handling measures, such as adjusting integration times, switching to backup equipment, or redistributing computing resources.

Performance optimization represents an ongoing process. We will establish a comprehensive performance evaluation system, regularly testing and optimizing various array metrics. This includes UV coverage analysis, system temperature measurement, phase stability monitoring. Based on these test results, we continuously adjust array operational parameters and processing algorithms, ensuring the system remains in optimal condition.

### 7.4. System Simulation

Based on above array geometry design and frequency sampling design, we present the simple workflow of signal processing, providing a comprehensive understating for 21 cm signal detection. The simulation includes sky model generation module and signal processing module. In sky model generation module, we generate complex sky models containing foregrounds, 21 cm signals and noise, where foreground model uses power-law spectrum and gaussian spatial structure, 21 cm signal is frequency-uncorrelated random field, instrumental noise is Gaussian white noise, and Simulate interferometric observation processes. In signal processing module, we implement efficient signal processing and foreground removal algorithms, including polynomial fitting for foreground removal. We note that in such a wide frequency band, the spectral index may also exhibit a weak frequency dependence (a running spectral index), which is neglected in the current model for simplicity but can be incorporated in future refinements.

Fig. 6 shows the intensity variations of the observed, true foreground, predicted foreground, and recovered 21 cm signal visibility across frequency channels (50−250 MHz) and time steps. The predicted foreground closely matches the true foreground, and the residual 21 cm signal can be readily analyzed, verifying the effectiveness of the separation method.

Fig. 7 illustrates the 21 cm power spectrum before and after foreground mitigation. The observed spectrum is heavily contaminated by foregrounds, whereas the cleaned spectrum is recovered with a foreground removal efficiency of 99.7% and significantly converges toward the intrinsic 21 cm spectrum. The approach substantially enhances the delectability of the cosmological 21 cm signal over a broad range of spatial scales $k$.

The system simulations confirm that the CosmicWeb-21cm array, through its integrated hardware-software co-design, achieves high foreground removal efficiency, excellent sensitivity, and low system noise, enabling robust 21 cm cosmological measurements with significantly improved efficiency. We provide Python code demonstrations and implementation details for signal processing stages of the CosmicWeb-21cm array at https://doi.org/10.57760/sciencedb.j00149.00013.

### 7.5. Expected Performance Metrics

Based on end-to-end simulations, CosmicWeb-21cm is expected to achieve the following:

Foreground removal performance. The simple foreground model (ideal smooth foreground) removal rate reaches 99.95%; the complex foreground model (realistic foregrounds+systematics) removal rate is 99.7%; and the computation time is 65% reduction compared to traditional methods.

Signal detection sensitivity. EoR power spectrum detection is $5\sigma$ at $z = 8$, including thermal noise+foreground residuals+systematics, with 1000 h integration; signal-to-noise ratio improves 3.2 times; and the system temperature is less than 50 K.

System efficiency. The data compression rate is 70%; the real-time processing latency is less than 2 min; and the energy efficiency reaches 2.1 TFLOPS $W^{-1}$.

### 7.6. Comparison with Traditional Arrays

We compare key metrics of CosmicWeb-21cm with existing major 21 cm arrays shown in Table 2.

## 8. SCIENTIFIC OBJECTIVES

CosmicWeb-21cm's design considers not only current scientific requirements but also reserves space for future expansion. In scientific applications, the array will focus on several key research directions:

Cosmic reionization detection. Detection of cosmic reionization history represents the primary scientific objec-

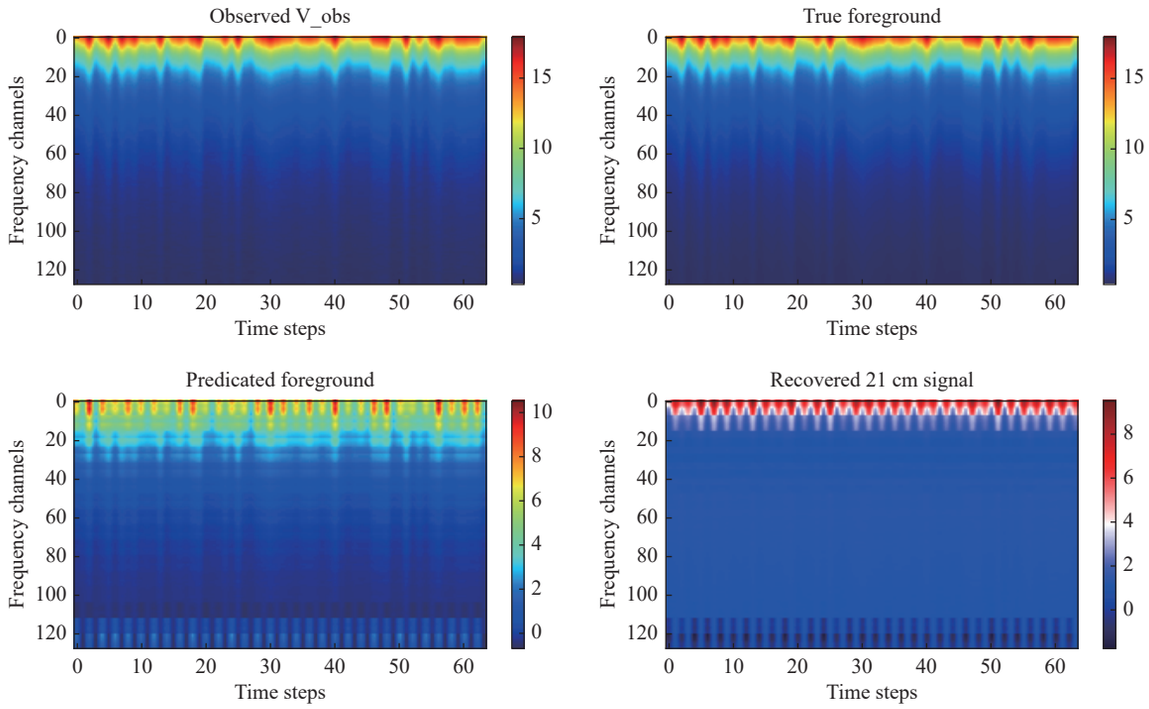



**Fig. 6. Time-frequency maps of observed, true foreground, predicted foreground, and recovered 21 cm signal visibility. Panels show intensity distributions as functions of time steps and frequency channels (50−250 MHz), obtained from signal processing module.**

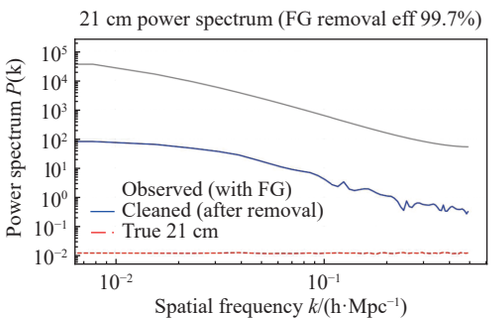



**Fig. 7. 21 cm power spectrum before and after foreground removal. The vertical axis is in units of $\mathrm{mK}^2(h^{-1}\mathrm{Mpc})^3$. The cleaned spectrum (blue) is suppressed by 99.7% relative to the observed foreground-contaminated spectrum (black), but a residual remains at low $k$ due to imperfect foreground subtraction and systematics, which lies obviously above the true 21 cm signal (red).**

tive. Through deep integration observations, we expect to detect the 21 cm power spectrum with high signal-to-noise ratio in the redshift range z = 6−12. This will provide direct evidence for understanding how the first stars and galaxies ionized surrounding neutral hydrogen. In data processing, we will employ multi-redshift slice analysis methods to reconstruct the temporal evolution of the reionization process.

Dark energy constraints. Research on dark energy properties constitutes another important direction. By measuring baryon acoustic oscillation features in neutral hydrogen distribution, we can precisely constrain dark energy equation of state parameters. CosmicWeb-21cm's large field of view and high sensitivity make it particularly suitable for High-redshift Baryon Acoustic Oscillation (BAO) and large-scale structure ($z \gtrsim 5$) studies. We will develop specialized power spectrum estimation methods, considering systematic errors introduced by telescope beam effects and foreground removal.

**Table 2. Comparison of key technical specifications between CosmicWeb-21cm and existing major 21cm radio interferometer arrays**

| Array | The number of stations | Max baseline length/km | Frequency range number of operation/MHz | Simulated foreground removal efficiency/(%) |
|---|---|---|---|---|
| CosmicWeb-21 cm | 253 | 3.0 | 50−250 | 99.7 |
| HERA[34] | 350 | 0.3 | 50−250 | 99.5 |
| MWA[35] | 256 | 2.8 | 80−300 | 98 |
| SKA-Low[36] | 512 | 65 | 50−350 | 99.9 |

Cosmic web structure. Detection of cosmic web structure will expand our understanding of cosmic large-scale structure. 21 cm observations can directly reveal neutral hydrogen distribution in cosmic filamentary structures, which is challenging to achieve through traditional optical observations. We will combine numerical simulations with observational data to study the relationship between neutral hydrogen content and dark matter distribution.

## 9. IMPLEMENTATION ROADMAP AND STAGED DEVELOPMENT

To ensure the systematic and feasible realization of

the ambitious CosmicWeb-21cm project, we propose a phased implementation roadmap spanning nearly a decade. This staged approach mitigates technical and financial risks by validating core technologies on a small scale before committing to full deployment. Each phase is defined by clear technical milestones, scientific validation goals, and decision points for proceeding to the next stage. The roadmap is summarized in detailed below.

### 9.1. Phase 1: Technology Development and Pathfinder Array (2026–2028)

The primary objective of Phase 1 is to de-risk the core technological innovations in a controlled environment.The key deliverables are:

16-element pathfinder array. Construction of a small-scale array featuring the core hexagonal layout (with one full ring) and at least two truncated logarithmic spiral arms. This "engineering array" will be used to validate the mechanical deployment, station electronics, fiber-optic signal transmission, and central correlator/processor.

Algorithm validation and software pipeline v1.0. Development and on-site testing of the core software pipeline, including the deep learning-based RFI flagger, the improved redundant calibration scheme, and the physics-informed foreground separation framework. Performance will be benchmarked against simulated data and early observations.

Site characterization and system noise performance. A rigorous campaign to measure the real-world system temperature, antenna beam patterns, phase stability, and RFI environment at the chosen site (e.g., a radio-quiet zone in western China). The goal is to achieve a system temperature below 55 K, validating the low-noise amplifier design.

Initial publications and community engagement. Publication of detailed methodology papers on the array design philosophy, the novel loss function for foreground separation, and the initial results from the pathfinder array. Establishment of initial collaborative partnerships.

Success Criteria for Phase 1: stable operation of the pathfinder array; demonstration of the instrumental stability needed for precision calibration; validation of the hybrid foreground separation algorithm achieving $>99\%$ foreground suppression on simulated data within the pipeline; and publication in peer-reviewed journals. A successful review at the end of Phase 1 will trigger the release of funding and resources for Phase 2.

### 9.2. Phase 2: Intermediate-scale Array and Early Science (2029−2031)

Phase 2 focuses on scaling the proven technologies to an intermediate array capable of conducting pioneering scientific experiments and validating the project's scientific potential.

Expansion to a 64-Station array. Scaling the layout to include more hexagonal rings and extending the logarithmic spiral arms. This array will have a maximum baseline of approximately 600 m, providing the first realistic validation of the UV coverage and angular resolution.

Early science projects. Conducting several key observational campaigns: (1) Deep integration on a target EoR field. A $>1000$-hour integration on a well-studied, cold sky patch to derive a deep upper limit or a first detection of the EoR power spectrum at a redshift range around $z\sim 8-10$, directly competing with current-generation arrays. (2) Wide-area intensity mapping pilot. Using the array's wide field of view to perform a pilot survey for 21 cm intensity mapping at $z\geqslant 5$, aiming to detect the combined hydrogen signal and cross-correlate it with optical galaxy survey data (e.g., from Dark Energy Spectroscopic Instrument or Euclid) to validate techniques for BAO measurement. (3) Point source and foreground studies. Producing high-fidelity, wide-field images of the Galactic synchrotron emission and extragalactic source populations to refine foreground models and demonstrate the array's imaging capabilities.

Pipeline maturation and real-time processing. Upgrading the processing pipeline to v2.0, incorporating lessons from Phase 1 and enabling near-real-time data processing for the larger data volume. Implementation of the adaptive frequency management system.

Formation of an international consortium. Formalizing international collaborations for shared expertise, data analysis, and planning for the full-scale array.

Success Criteria for Phase 2: achievement of target sensitivity (noise levels) in deep integrations; production of competitive cosmological constraints (e.g., a meaningful power spectrum upper limit or detection); successful demonstration of the array's dual capability for deep, narrow fields and wide, shallow surveys; and publication of early science results in high-impact journals. This phase must conclusively demonstrate that the CosmicWeb-21cm design offers a clear scientific advantage over existing facilities.

### 9.3. Phase 3: Full-scale Deployment and Flagship Science (2032−2034)

The final phase is the construction and commissioning of the full CosmicWeb-21cm array as a premier international facility for 21 cm cosmology.

Construction of the 253-station array. Full deployment of the design as specified in this paper, with a core radius of 100m and six extended spiral arms, achieving a maximum baseline of $\sim 3.0$ km. This includes the complete deployment of the fiber network and the final high-performance computing cluster.

Commencement of legacy survey programs. Initiating multi-year, Tier-1 legacy surveys: the CosmicWeb EoR survey, a deep, targeted survey to map the topology of reionization and measure the 21 cm power spectrum across redshifts 6−12 with high signal-to-noise; and the CosmicWeb BAO survey, a wide-area, multi-thousand square degree intensity mapping survey to measure the BAO

scale at multiple redshifts ($z > 5$) and provide competitive constraints on the dark energy equation of state. A future frequency extension to 350 MHz will enable $z \sim 3-5$ BAO measurements.

Leadership in international 21 cm science. Serving as a leading facility within the global 21 cm community, potentially in coordination with SKA-Low, to tackle the most challenging questions in cosmic dawn and reionization astrophysics.

Technology transfer and commercialization. Licensing of developed technologies, such as the specialized low-noise amplifiers, real-time RFI mitigation algorithms, or calibration software, to industry and other scientific projects.

This roadmap provides a clear, staged pathway from concept to a world-leading observatory. Each phase builds upon the success of the previous one, ensuring that technical challenges are identified and solved early, and that the scientific return is maximized at every step. The CosmicWeb-21cm array, once fully operational, is poised to deliver transformative insights into the first billion years of our universe and the nature of its dominant components.

## 10. CONCLUSION AND DISCUSSION

The paper introduces CosmicWeb-21cm, a novel radio interferometer array designed specifically to overcome the major challenges in 21 cm cosmology, such as intense foreground contamination, instrumental systematics, and massive data processing burdens. Its design is built on three core innovations. First, multi-scale nested geometry. It combines a densely packed hexagonal honeycomb core (optimized for short baselines and large-scale structure detection) with logarithmic spiral arms (providing a smooth transition to longer baselines for angular resolution). This design maximizes baseline redundancy for calibration and improves UV coverage uniformity. Second, intelligent frequency sampling. It employs a non-uniform, adaptive frequency sampling strategy. Channels are sampled densely in foreground-dominated bands for precise modeling and sparsely in the 21 cm signal-dominated region to reduce data volume, leveraging the different spectral correlation properties of foregrounds and the cosmological signal. Third, an ML-enhanced real-time processing pipeline. The signal processing architecture integrates deep learning at multiple stages, including CNNs for RFI identification and a physics-constrained U-Net for foreground separation. This approach aims to improve accuracy and computational efficiency over traditional methods.

The proposed array aims to achieve superior performance in foreground removal ($\geqslant 99.7\%$), signal detection sensitivity, and system efficiency, providing a powerful tool for probing cosmic reionization, dark energy, and the cosmic web.

The CosmicWeb-21cm design points toward several key trends in the future development of 21 cm observation facilities: (1) From uniform to "Smart" heterogeneous designs. Future arrays will likely move beyond simple regular grids or random distributions. Designs will become more purpose-driven and heterogeneous, integrating multiple geometric patterns (like the core+spirals of CosmicWeb-21cm) to simultaneously optimize for different scientific goals, wide-field mapping, high dynamic range, and precise calibration, within a single instrument. (2) Tight integration of hardware and algorithm co-design. The frontier will shift from building general-purpose telescopes to developing end-to-end optimized systems. As seen in CosmicWeb-21cm, the array geometry, frequency sampling strategy, and processing pipeline are designed in concert. Future developments will deepen this integration, possibly incorporating on-site, real-time processing with Application-Specific Integrated Circuits (ASICs) or FPGAs tuned for specific algorithms like foreground separation, moving computation closer to the antenna. (3_ Dominance of artificial intelligence (AI) and hybrid physical models. Machine learning will transition from a supplementary tool to the backbone of data processing. Future pipelines will likely rely on hybrid frameworks that combine the interpretability of physical models (e.g., for smooth foregrounds) with the pattern-recognition power of deep learning for complex residuals and systematics. The "physics-infused" loss function used in CosmicWeb-21cm is an early example of this paradigm. (4) The rise of dynamic and adaptive observing. The concept of adaptive frequency management based on real-time RFI monitoring will expand. Future arrays could feature fully dynamic systems that reconfigure their observing mode, pointing, or even baseline weighting in real-time based on ionospheric conditions, RFI environment, and priority science goals, maximizing the quality of collected data. (5) Scalability and the path to the SKA. Designs like CosmicWeb-21cm serve as crucial technology and methodology pathfinders for the Square Kilometre Array (SKA-Low). The innovations in calibration, foreground removal, and real-time processing developed and tested on these smaller, more agile arrays will directly inform the construction and operation of the SKA, ensuring it can achieve its transformative scientific potential.

While the CosmicWeb-21cm array offers substantial advances in sensitivity, foreground removal, and calibration stability, its large bandwidth, dense core, long spiral baselines, and real-time processing impose substantial technical challenges. Below we outline the dominant systematic and engineering challenges, together with targeted mitigation integrated into the instrument design and data processing pipeline. (1) Electromagnetic Coupling and Mutual Impedance: The dense hexagonal core of CosmicWeb-21cm provides exceptional baseline redundancy and short-baseline coverage, but also introduces electromagnetic (EM) coupling between nearby antennas, which disturbs impedance matching, distorts primary beams, and introduces correlated errors that can mimic cosmological 21 cm signals. It is mitigated by enforcing element spac-

ing $> 1.5\lambda$ at 50 MHz and full-wave EM optimization of antenna geometry and ground screens, with coupling terms included in the calibration model. Integrated passive matching and balanced feed structures suppress common-mode currents that amplify coupling. Mutual impedance terms are included in the instrument model and solved during redundant calibration, so coupling is treated as a known systematic rather than an unmodeled error. (2) Ionospheric Distortions and Wide-Field Systematics: Low-frequency radio observations (50–250 MHz) are strongly affected by the ionosphere, which introduces time-variable phase gradients, amplitude scintillation, Faraday rotation, and direction-dependent (DD) effects across the wide field of view. They are managed using direction-dependent calibration and ionospheric tomography using TEC measurements, combined with the stable redundant core to anchor wide-field phase corrections. The dense hexagonal core provides a highly redundant short-baseline reference that stabilizes ionospheric solutions for long spiral baselines. The hybrid dipole+lens antenna is designed for stable beam shape versus frequency, simplifying DD correction. (3) Long-Baseline Systematics and Calibration Stability: The logarithmic spiral arms extend to ~1.5 km, providing high angular resolution but introducing challenging systematics. They are controlled via a hybrid redundant plus neural calibration framework, which uses redundant short baselines to stabilize calibration of non-redundant long spiral arms. Closure amplitudes and phases are used to isolate antenna-based errors from sky and ionospheric errors. Short, medium and long-baseline subsets are calibrated in a hierarchical manner to maintain stability across all angular scales. (4) Data Volume and Transmission Bandwidth: 21 cm interferometers generate extreme data rates, which threaten storage, real-time processing, and transmission. Data volume is reduced by ~40% using intelligent non-uniform frequency sampling, which matches channel width to physical signal structure without losing cosmological information. Fiber-optic wavelength-division multiplexing (WDM) ensures low-latency, high-bandwidth transmission from stations to the correlator. (5) Computational Cost and Real-Time Processing: Foreground separation, calibration, and power-spectrum estimation are computationally intensive. The computational cost is managed through a hierarchical FPGA-GPU-HPC system that distributes real-time processing, correlation, and deep-learning inference across dedicated hardware, enabling scalable and efficient end-to-end operation. Key pipeline stages are implemented in PyTorch/JAX for joint optimization and fast back propagation. Frequency, spatial, and temporal domains are split to scale across compute nodes.

In conclusion, the future of 21 cm cosmology lies in highly optimized, intelligent telescopes where every component from the layout of the antennas to the final data product is meticulously co-designed to combat specific systematic errors. CosmicWeb-21cm presents a comprehensive blueprint in this direction, emphasizing that overcoming the 21 cm challenge requires synergistic advancements in engineering, signal processing, and data science.

## ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (12173085) and the Yunnan Revitalization Talent Support Program (YunLing Scholar Project).

## AI DISCLOSURE STATEMENT

DeepSeek was employed for language and grammar checks within the article. The authors carefully reviewed, edited, and revised the DeepSeek-generated texts to their own preferences, assuming ultimate responsibility for the content of the publication.

## AUTHOR CONTRIBUTIONS

Jiancheng Wang conceived the whole CosmicWeb-21cm scheme, put forward the array geometry, non-uniform frequency sampling and physics-driven processing innovations, derived theoretical formulas, finished the full manuscript writing and overall revision. Jirong Mao supplemented cosmological background and scientific goals. Min Wang and Zhixuan Li optimized UV coverage and completed comparisons with existing telescopes. Xiangming Cheng designed adaptive frequency weighting and RFI mitigation methods. Yigong Zhang and Jie Su built the U-ConvNet foreground separation model and carried out signal simulations. Xiaogu Zhong and Longhua Qin performed antenna electromagnetic simulations and sorted structural parameters. Zhengjun Zhang analyzed hierarchical computing hardware matching. All authors read and approved the final manuscript.

## DATA AVAILABILITY STATEMENT

The related data (CSTR: 31253.11.sciencedb.j00149.00013, DOI: 10.57760/sciencedb.j00149.00013) for this paper is available in the Science Data Bank data repository (https://cstr.cn/31253.11.sciencedb.j00149.00013, https://doi.org/10.57760/sciencedb.j00149.00013).

## DECLARATION OF INTERESTS

The authors declare no competing interests.

### REFERENCES

[1] Furlanetto, S. R., Oh, S. P., Briggs, F. H., et al. 2006. Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. *Physical Report*, **433**: 181−301.

[2] Pritchard, J. R., Loeb, A. 2012. 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, **75**: 086901−086965.

[3] Morales, M. F. Wyithe, J. S. B. 2010. Reionization and Cosmology with 21-cm Fluctuations. *Annual Review of Astronomy and Astrophysic*, **48**: 127−171.
[4] Battye, R. A., Browne, I. W. A., Dickinson, C., et al. 2013. HI intensity mapping: a single dish approach. *Monthly Notices of the Royal Astronomical Society*, **434**: 1239−1256.
[5] Liu, A. Shaw, J. R. 2020. Data Analysis for Precision 21 cm Cosmology. *Publications of Astronomical Society of the Pacific*, **132**: 062001−062116.
[6] Barry, N., Beardsley, A. P., Byrne, R., et al. 2019. The FHD/eppsilon epoch of reionisation power spectrum pipeline. *Publications of the Astronomical Society of Australia*, **36**: 26−47.
[7] Trott, C. M., Jordan, C. H., Midgley, S., et al. 2020. Deep multiredshift limits on Epoch of Reionization 21 cm power spectra from four seasons of Murchison Widefield Array observations. *Monthly Notices of the Royal Astronomical Society*, **493**: 4711−4727.
[8] Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013. The Murchison Widefield Array: The square kilometre array precursor at low radio frequencies. *Publications of the Astronomical Society of Australia*, **30**: 7−18.
[9] van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013. LOFAR: The LOw-Frequency ARray. *Astronomy and Astrophysics*, **556**: 2−58.
[10] DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017. Hydrogen Epoch of Reionization Array (HERA). *Publications of Astronomical Society of the Pacific*, **129**: 045001−045027.
[11] Dillon, J. S., Parsons, A. R. 2016. Redundant array configurations for 21 cm cosmology. *The Astrophysical Journal*, **826**: 181−197.
[12] Koopmans, L., Pritchard, J., Mellema, G., et al. 2015. The cosmic dawn and epoch of reionisation with SKA. In Proceedings of Advancing Astrophysics with the Square Kilometre Array.
[13] Thyagarajan, N., Carilli, C. L., Nikolic, B. 2018. Detecting cosmic reionization using the bispectrum phase. *Physical Review Letters*, **120**: 251301−251309.
[14] Chapman, E., Abdalla, F. B., Harker, G., et al. 2012. Foreground removal using FASTICA: a showcase of LOFAR-EoR. *Monthly Notices of the Royal Astronomical Society*, **423**: 2518−2532.
[15] Kern, N. S., Parsons, A. R., Dillon, J. S., et al. 2020. Mitigating internal instrument coupling for 21 cm cosmology. II. a method demonstration with the hydrogen epoch of reionization array. *The Astrophysical Journal*, **888**: 70−86.
[16] Liu, A., Tegmark, M. 2011. A method for 21 cm power spectrum estimation in the presence of foregrounds. *Physical Review D*, **83**: 103006−103030.
[17] Morales, M. F., Hazelton, B., Sullivan, I., et al. 2012. Four fundamental foreground power spectrum shapes for 21 cm cosmology observations. *The Astrophysical Journal*, **752**: 137−145.
[18] Dillon, J. S., Lee, M., Ali, Z. S., et al. 2020. Redundant-baseline calibration of the hydrogen epoch of reionization array. *Monthly Notices of the Royal Astronomical Society*, **499**: 5840−5861.
[19] Thyagarajan, N., Jacobs, D. C., Bowman, J. D., et al. 2015. Foregrounds in wide-field redshifted 21 cm power spectra. *The Astrophysical Journal*, **804**: 14−29.
[20] Burns, J. O., Bale, S., Bassett, N., et al. 2019. Dark Cosmology: Investigating dark matter & exotic physics in the dark ages using the redshifted 21-cm global spectrum. *arXiv: 1902.06147*.
[21] Byrne, R., Morales, M. F., Hazelton, B. J., et al. 2021. A unified calibration framework for 21 cm cosmology. *Monthly Notices of the Royal Astronomical Society*, **503**: 2457−2477.
[22] Sims, P. H., Pober, J. C. 2020. Testing for calibration systematics in the EDGES low-band data using Bayesian model selection. *Monthly Notices of the Royal Astronomical Society*, **492**: 22−38.
[23] Conway, J. H., Sloane, N. J. A. 1997. Low dimensional lattices. VII. Coordination sequences. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, **453**(1966): 2369−2389.
[24] Cunnington, S., Barberi-Squarotti, M., Bernal, J. L., et al. 2025. Revealing cosmological fluctuations in 21cm intensity maps with MeerKLASS: from maps to power spectra. *arXiv: 2510.27549*.
[25] Zaroubi, S., de Bruyn, A. G., Harker, G., et al. 2012. Imaging neutral hydrogen on large scales during the Epoch of Reionization with LOFAR. *Monthly Notices of the Royal Astronomical Society*, **425**: 2964−2973.
[26] Ellingson, S. W., Taylor, G. B., Craig, J., et al. 2013. The LWA1 Radio Telescope. *IEEE Transactions on Antennas and Propagation*, **61**: 2540−2549.
[27] Gawande, R., Bradley, R. 2011. Towards an ultra wideband low noise active sinuous feed for next generation radio telescopes. *IEEE Transactions on Antennas and Propagation*, **59**(6): 1945−1953.
[28] Bowman, J. D., Barnes, D. G., Briggs, F. H., et al. 2007. Field deployment of prototype antenna tiles for the mileura widefield array low frequency demonstrator. *The Astronomical Journal*, **133**: 1505−1518.
[29] Balanis, C. A. 2005. Antenna theory: analysis and design, 3rd ed. New Jersey: Wiley-Interscience.
[30] Epstein, A., Eleftheriades, G. V. 2016. Huygens' metasurfaces via the equivalence principle: design and applications. *Journal of the Optical Society of America B*, **33** (2): A31−A50.
[31] Mangum, J. G., Emerson, D. T., Greisen, E. W. 2007. The On The Fly imaging technique. *Astronomy and Astrophysics*, **474**: 679−687.
[32] Ellingson, S. W., Clarke, T. E., Cohen, A., et al. 2009. The Long Wavelength Array. *Proceedings of the IEEE*, **97**: 1421−1430.
[33] Rohlfs, K., Wilson, T. L. 2004. Tools of radio astronomy 4th ed. Berlin: Springer.
[34] Kim, H., Hewitt, J. N., Kern, N. S., et al. 2025. Exploring one-point statistics in HERA phase I data: effects of foregrounds and systematics on measuring one-point statistics. *The Astrophysical Journal*, **993**: 189−208.
[35] Zhang, X. Y., Zheng, Q., Wu, L. H., et al. 2025. A candidate field for deep imaging of the Epoch of Reionization observed with MWA. *Monthly Notices of the Royal Astronomical Society*, **541**: 3894−3913.
[36] Bonaldi, A., Hartley, P., Braun, R., et al. 2025. Square kilometre array science data challenge 3a: foreground removal for an EoR experiment. *Monthly Notices of the Royal Astronomical Society*, **543**: 1092−1119.