Generative diffusion-based channel estimation for pinching antenna-assisted indoor NFC

Chen, Y. ORCID: https://orcid.org/0009-0009-6278-0800 and Wang, Q. ORCID: https://orcid.org/0000-0002-1404-582X
(2026) Generative diffusion-based channel estimation for pinching antenna-assisted indoor NFC. Electronics, 15 (4). 730. ISSN 2079-9292 doi: 10.3390/electronics15040730
Available at https://centaur.reading.ac.uk/128518/

electronics

MDPI

Article

Generative Diffusion-Based Channel Estimation for Pinching Antenna-Assisted Indoor NFC

Yongji Chen [1,2] and Qi Wang [3,*]

[1] School of Data Science and Big Data Technology, Nanjing University of Information Science and Technology, Nanjing 210044, China; yongjichen@nuist.edu.cn or yongjichen@student.reading.ac.uk
[2] School of Computer Science, University of Reading, Reading RG6 6AY, UK
[3] School of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China
* Correspondence: 002086@nuist.edu.cn

Abstract

Pinching antenna systems (PASS) are waveguide-based antenna architectures featuring structural flexibility and high energy efficiency, making them attractive for indoor near-field communication (NFC). However, rich multipath propagation and spatial non-stationarity in practical indoor environments pose significant challenges to accurate channel estimation, especially under limited antenna activation and pilot resources. In this paper, the PASS channel estimation problem is reformulated from a generative inference perspective. A diffusion-model-driven channel estimation framework is proposed, where received signals are interpreted as noisy observations of latent near-field channel states, and channel estimation is performed via conditional reverse denoising diffusion. By exploiting waveguide-mediated near-field structures and sparse antenna activation, the proposed framework enables robust channel recovery in highly underdetermined settings. To better match indoor propagation characteristics, the diffusion-based inference emphasizes multipath-aware channel distributions, allowing joint modeling of deterministic waveguide effects and stochastic scattering, thereby alleviating model mismatch in conventional estimators. Simulation results show that the proposed method achieves stable channel estimation performance across different SNRs and antenna activation scales, while the computational complexity of the proposed framework is explicitly analyzed to assess its practical applicability.

Keywords: channel estimation; pinching antenna; indoor near-field communications; diffusion model

Academic Editor: Rocco Guerriero

Received: 10 January 2026
Revised: 2 February 2026
Accepted: 5 February 2026
Published: 9 February 2026

1. Introduction

As sixth-generation (6G) wireless communication systems continue to evolve toward higher carrier frequencies, wider bandwidths, and improved energy efficiency, the conventional communication paradigm dominated by far-field coverage is progressively extending to indoor near-field communication (NFC) scenarios [1,2]. In support of high-density device connectivity, immersive interactive applications, and complex indoor deployments, achieving precise controllability and reliable transmission of near-field electromagnetic radiation under constrained hardware complexity and energy budgets has emerged as a fundamental problem in the design of next-generation wireless system architectures [3].

Pinching antenna systems (PASS) have emerged as a reconfigurable antenna architecture based on dielectric waveguides, where low-cost pinching antenna elements can be

flexibly deployed to dynamically control radiation locations, the number of active radiating elements, and the spatial distribution of radiated energy [4–10]. Existing studies on PASS can be broadly categorized into three aspects: system architecture design, propagation and channel characteristics, and structural reconfiguration and resource optimization. From an architectural perspective, ref. [4] showed that PASS can establish strong line-of-sight (LoS) links with good structural scalability, while [5] revealed its advantages in forming LoS-dominant links, mitigating large-scale path loss, and enhancing multi-user performance. Building upon these foundations, subsequent works focused on resource optimization, interference management, and environmental awareness. In particular, ref. [6] studied a non-orthogonal multiple access (NOMA)-assisted downlink PASS and addressed sum-rate maximization via matching theory, whereas [7] extended PASS to multi-waveguide scenarios and developed game-theoretic and optimization-based solutions for joint waveguide assignment, antenna activation, decoding order, and power allocation. In addition, the impact of line-of-sight blockage was investigated in [8], which showed its performance benefits in multi-user scenarios through interference suppression, and was further exploited in [9] via blockage-aware waveguide assignment and antenna activation. Complementing these works, ref. [10] examined structural reconfiguration with tunable power radiation models and discrete antenna activation, systematically characterizing the relationship between structural configurations and communication performance. Overall, these studies establish PASS as a promising flexible antenna architecture for next-generation indoor communication systems.

However, whether the structural advantages of PASS can be translated into practical performance gains largely depends on the availability and accuracy of channel state information (CSI), which is particularly challenging to acquire in realistic indoor environments characterized by rich multipath propagation and strong spatial non-stationarity. Existing studies on PASS have mainly focused on system-level performance analysis and structural optimization under ideal or readily available CSI assumptions. For example, refs. [11,12] investigated uplink transmission in PASS by optimizing pinching antenna placement and deriving analytical or closed-form ergodic rate expressions, without explicitly addressing CSI acquisition or modeling in practical environments. Meanwhile, conventional pilot-based linear estimation and compressive sensing methods degrade significantly under complex multipath propagation and constrained antenna activation. Learning-driven approaches that model CSI as direct regression mappings often fail to capture the underlying distributional properties of wireless channels, which leads to limited generalization in PASS scenarios. Recent efforts have introduced generative deep learning [13] and PASS-oriented architectures such as PAMoE and PAformer [14] to improve estimation robustness. However, these methods largely remain within discriminative paradigms. Fundamentally, the PASS wireless channel should be viewed as a high-dimensional random object generated by structured waveguide-mediated propagation and stochastic scattering, with received signals being noisy observations of latent channel states. This observation motivates a re-examination of the modeling paradigm for PASS channel estimation. Instead of treating CSI recovery as a deterministic parameter reconstruction task, can channel estimation be formulated from a generative perspective as a distribution inference problem?

Diffusion models, as a rapidly emerging class of generative modeling approaches, have demonstrated distinctive advantages in learning high-dimensional random distributions and characterizing uncertainty. By constructing a bidirectional process consisting of forward perturbation and reverse denoising, diffusion models enable the progressive modeling and recovery of complex data distributions [15,16]. This mechanism exhibits a natural conceptual correspondence with wireless channel formation: multipath propagation and noise effects can be interpreted as a diffusion-like process, while channel

estimation corresponds to an inference task that gradually recovers latent propagation structures from random observations. In particular, for indoor near-field PASS scenarios, diffusion models provide a novel modeling framework capable of unifying deterministic waveguide-mediated propagation structures with stochastic scattering effects [17,18].

Motivated by the above considerations and the conceptual comparison summarized in Table 1, this work systematically investigates the application of diffusion-based generative modeling to channel estimation in pinching antenna-assisted indoor NFC systems. The main contributions of this paper are summarized as follows:

- We reformulate PASS channel estimation from a generative modeling perspective by introducing a diffusion-based framework that casts channel inference as a reverse denoising process from randomly perturbed latent channel states, enabling robust estimation under limited antenna activation and pilot overhead in indoor near-field PASS scenarios.
- We develop a multipath-aware diffusion inference method that jointly captures deterministic waveguide-mediated near-field propagation and stochastic scattering effects, thereby alleviating model mismatch in spatially non-stationary indoor environments.
- We demonstrate through simulations that the proposed approach consistently outperforms conventional and learning-based baselines across different signal-to-noise ratio (SNR) regimes and antenna activation scales. In addition, the computational complexity of the proposed framework is analyzed to evaluate its practical applicability.

Table 1. Importance comparison of channel estimation paradigms for PASS indoor NFC.

Aspect	Model-Based Methods	Learning-Based Regression	Proposed Generative Diffusion Framework
Modeling perspective	Deterministic inversion	Direct input–output mapping	Generative inference of latent channels
Multipath treatment	Explicit parametric modeling	Implicit data fitting	Distribution-aware modeling via diffusion
PASS scenario suitability	Limited	Moderate	High (near-field propagation and sparse activation)
Robustness under underdetermined observations	Performance degrades significantly	Sensitive to activation patterns	Robust in highly underdetermined regimes
Engineering usability	Interpretable but rigid	Limited generalization	Balanced robustness and interpretability

As shown in Figure 1, the remainder of this paper is organized as follows. Section 2 presents the system model of the considered pinching antenna-assisted near-field communication scenario. Section 3 introduces the proposed diffusion-based channel estimation framework and illustrates its core processing flow in detail. Section 4 provides extensive simulation results to evaluate the performance of the proposed method under different indoor propagation conditions. Finally, Section 5 concludes the paper and discusses potential directions for future work. For clarity, the key notations used throughout this paper are summarized in Table 2.

Table 2. List of notations.

Symbol	Meaning	Symbol	Meaning
$\mathcal{T}_i$	Pilot slot set for user i	$\mathbf{x}_0, \mathbf{x}_t$	Diffusion latent channel state
$\mathbf{y}_i, \mathbf{y}_i^{\mathrm{R}}$	Pilot observation vector	$\hat{\mathbf{x}}_0, \hat{\mathbf{h}}_i$	Estimated channel
$\mathbf{W}_i, \mathbf{W}_i^{\mathrm{R}}$	Sensing matrix	$\boldsymbol{\epsilon}$	Injected Gaussian noise
σ_n^2	Noise variance	β_t	Noise schedule at step t
$\mathbf{a}_t, \mathbf{A}_t$	Antenna activation vector/matrix	$\alpha_t, \bar{\alpha}_t$	Signal preservation factor
$\mathbf{g}$	In-waveguide channel vector	η_t	Consistency correction step size

Table 2. *Cont.*

Symbol	Meaning	Symbol	Meaning
$\mathbf{w}_t$	Effective combining vector	λ, λ_g	Free-space/guided wavelength
$\mathbf{h}_i$	Wireless channel vector	β_g	Guided-wave propagation constant
$\boldsymbol{v}_i$	LoS visibility indicator	$r_{i,n}, r_{l,n}$	User–antenna/waveguide distance
$d_{i,k}, d_{k,n}$	NLoS scattering distances	$\mathcal{CN}(\cdot)$	Complex Gaussian distribution

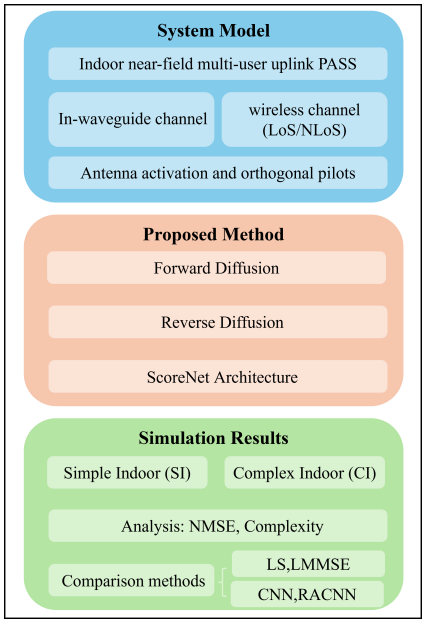


Figure 1. Overall research framework for diffusion-based channel estimation in PASS.

2. System Model

As illustrated in Figure 2, an indoor near-field multi-user uplink communication system assisted by pinching antennas is considered, where I single-antenna users concurrently transmit uplink signals to a base station (BS). At the BS, a dielectric waveguide is employed as a unified receiving and signal-combining structure. Along the waveguide, N pinching antenna (PA) units are deployed at candidate locations. Acting as passive electromagnetic coupling elements, these pinching antennas collect uplink signals radiated by the users in free space and couple them into the waveguide, where they are converted into guided electromagnetic waves. Each pinching antenna introduces a location-dependent coupling response that determines the amplitude and phase of the injected signal. The guided signals coupled by the selected (activated) pinching antennas are physically superposed along the waveguide and collected at a single feed point, which is connected to a single radio-frequency (RF) chain at the BS. To reflect practical implementation constraints, such as mechanical fixation and predefined coupling points in waveguide fabrication, we assume that pinching antennas can only be activated at a finite set of predefined discrete candidate locations along the waveguide, resulting in a discretized deployment grid.

2.1. Channel Model

An indoor near-field multi-user uplink communication scenario assisted by PASS is considered. The overall uplink channel consists of two components: the in-waveguide channeland the wireless propagation channel. The in-waveguide channel characterizes the guided electromagnetic propagation and coupling behavior between the pinching antennas and the feed point along the dielectric waveguide, while the wireless propagation channel models the near-field LoS and non-line-of-sight (NLoS) propagation between the users and the pinching antennas.

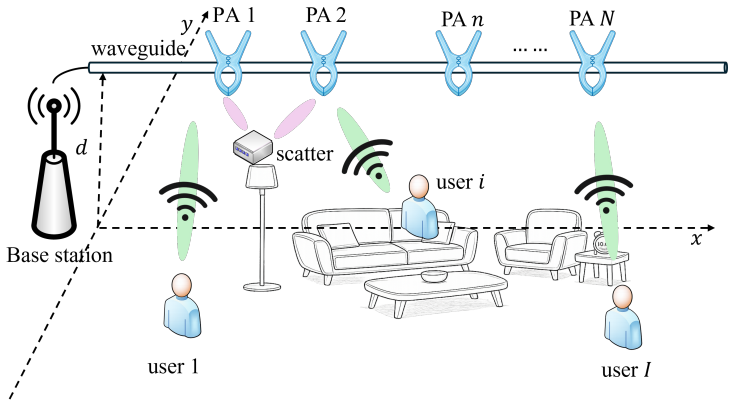


Figure 2. Pinching antennas assisted indoor NFC systems.

A three-dimensional Cartesian coordinate system is adopted to describe the system geometry. All users are assumed to be located on the ground plane $z = 0$. The position of the ith user is denoted by $\boldsymbol{\psi}_i = (x_i, y_i, 0)$. A dielectric waveguide is mounted at height $z = d$ and extends along the x-axis. Along the waveguide, N pinching antennas are deployed at predefined discrete locations. The position of the nth pinching antenna is given by $\boldsymbol{\psi}_n^{\mathrm{P}} = (x_n^{\mathrm{P}}, 0, d)$, $n \in \{1, \ldots, N\}$. The feed point of the waveguide is located at $\boldsymbol{\psi}_0^{\mathrm{P}} = (0, 0, d)$.

The propagation of guided electromagnetic waves inside the dielectric waveguide is characterized by the guided-wave propagation constant $\beta_g = \frac{2\pi}{\lambda_g}$, where $\lambda_g = \frac{\lambda}{n_e}$, λ denotes the free-space wavelength and n_e is the effective refractive index of the waveguide. Then, we let $r_{l,n}$ denote the axial propagation distance between the feed point and the nth pinching antenna along the waveguide, given by

$$r_{l,n} = \left\| \boldsymbol{\psi}_n^{\mathrm{P}} - \boldsymbol{\psi}_0^{\mathrm{P}} \right\| = \left| x_n^{\mathrm{P}} \right|. \tag{1}$$

Energy exchange between the pinching antennas and the waveguide is realized through electromagnetic coupling. The coupling strength of the nth pinching antenna is characterized by a coupling-amplitude coefficient $\alpha_n \in [0, 1]$, which represents the ratio of power exchanged between the waveguide and the pinching antenna and is determined by the antenna structure and coupling configuration [14]. Accordingly, the corresponding in-waveguide channel coefficient is expressed as

$$g_n = \alpha_n e^{-j\beta_g r_{l,n}}. \tag{2}$$

By stacking the coefficients associated with all pinching antennas, the in-waveguide channel vector is obtained as

$$\mathbf{g} = [g_1, \ldots, g_N]^{\mathrm{T}} \in \mathbb{C}^{N \times 1}. \tag{3}$$

2.1.1. Wireless Propagation Channel Model

The wireless propagation channel between the users and the pinching antenna array is now considered. The channel vector from the ith user to all pinching antennas is defined as

$$\mathbf{h}_i = [h_{i,1}, \ldots, h_{i,N}]^{\mathrm{T}} \in \mathbb{C}^{N \times 1}. \tag{4}$$

In indoor near-field environments, LoS propagation may not be uniformly available across all user–antenna links due to blockage and spatial geometry. To model this effect, a binary visibility vector is introduced as $\boldsymbol{v}_i = [v_{i,1}, \ldots, v_{i,N}]^{\mathrm{T}}$, $v_{i,n} \in \{0, 1\}$. Specifically,

$v_{i,n} = 1$ denotes the presence of a LoS path between the ith user and the nth pinching antenna, and $v_{i,n} = 0$ indicates its absence due to blockage.

Accordingly, the wireless channel is decomposed into a LoS component $\mathbf{h}_i^{\mathrm{LoS}}$ and NLoS component $\mathbf{h}_i^{\mathrm{NLoS}}$ as

$$\mathbf{h}_i = \boldsymbol{v}_i \odot \mathbf{h}_i^{\mathrm{LoS}} + \mathbf{h}_i^{\mathrm{NLoS}}. \tag{5}$$

2.1.2. LoS Channel Component

The distance between the ith user and the nth pinching antenna is given by

$$r_{i,n} = \left\| \boldsymbol{\psi}_i - \boldsymbol{\psi}_n^{\mathrm{P}} \right\|. \tag{6}$$

Under near-field conditions, spherical wavefront propagation is considered. By absorbing antenna directivity and constant scaling factors into an equivalent path-loss constant, the LoS channel coefficient is modeled as

$$h_{i,n}^{\mathrm{LoS}} = \eta^{\mathrm{LoS}} \frac{e^{-j\frac{2\pi}{\lambda} r_{i,n}}}{r_{i,n}}, \tag{7}$$

where η^{LoS} denotes an equivalent path-loss constant associated with a reference distance, which accounts for antenna-related gains and normalization factors.

2.1.3. NLoS Channel Component

Scattered propagation is modeled by assuming the existence of K dominant scatterers. We let $\boldsymbol{\phi}_k$ denote the position of the kth scatterer. The distances from the user to the scatterer and from the scatterer to the nth pinching antenna are given by

$$d_{i,k} = \| \boldsymbol{\psi}_i - \boldsymbol{\phi}_k \|, \qquad d_{k,n} = \left\| \boldsymbol{\phi}_k - \boldsymbol{\psi}_n^{\mathrm{P}} \right\|. \tag{8}$$

The complex gain of the kth scattering path is modeled as $\beta_{i,k} e^{j\Phi_{i,k}}$, where $\beta_{i,k}$ characterizes the scattering strength and $\Phi_{i,k}$ is uniformly distributed over $[0, 2\pi)$. The spherical-wave propagation response associated with the kth scatterer is modeled as a vector $\mathbf{b}_{i,k}$, whose nth element is given by

$$[\mathbf{b}_{i,k}]_n = \frac{e^{-j\frac{2\pi}{\lambda}(d_{i,k} + d_{k,n})}}{d_{i,k} d_{k,n}}. \tag{9}$$

The NLoS channel component is then expressed as

$$\mathbf{h}_i^{\mathrm{NLoS}} = \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \beta_{i,k} e^{j\Phi_{i,k}} \mathbf{b}_{i,k}. \tag{10}$$

2.2. Signal Model

Due to the shared dielectric waveguide and the single feed point connected to one RF chain, the BS obtains a scalar observation in each transmission slot. During the ιth transmission slot, the activation state of the N pinching antennas is described by a binary selection vector $\mathbf{a}_t = [a_{\iota,1}, \ldots, a_{\iota,N}]^{\mathrm{T}}, a_{\iota,n} \in \{0, 1\}$, where $a_{\iota,n} = 1$ indicates that the nth pinching antenna is activated in slot ι, and $a_{\iota,n} = 0$ otherwise. The corresponding diagonal selection matrix is given by $\mathbf{A}_\iota = \mathrm{diag}(\mathbf{a}_\iota)$.

We let $s_{i,\iota}$ denote the uplink transmit symbol of the ith user in slot ι, satisfying the average power constraint $\mathbb{E}\{|s_{i,\iota}|^2\} \leq P_i$. The additive noise at the BS is modeled as

circularly symmetric complex Gaussian noise $n_\iota \sim \mathcal{CN}(0, \sigma_n^2)$. The uplink signal received at the BS in the ιth slot is expressed as

$$y_\iota = \sum_{i=1}^{I} \mathbf{g}^{\mathrm{T}} \mathbf{A}_\iota \mathbf{h}_i \, s_{i,\iota} + n_\iota, \tag{11}$$

where $\mathbf{g}$ denotes the in-waveguide channel vector and $\mathbf{h}_i$ represents the wireless channel vector between the ith user and all pinching antennas.

For notational convenience, we define the effective combining vector in slot t as $\mathbf{w}_\iota \triangleq \mathbf{A}_\iota \mathbf{g}$. Accordingly, the received signal can be equivalently rewritten as

$$y_\iota = \sum_{i=1}^{I} \mathbf{w}_\iota^{\mathrm{T}} \mathbf{h}_i \, s_{i,\iota} + n_\iota. \tag{12}$$

During the channel estimation phase, an orthogonal pilot transmission strategy is employed. We let $\mathcal{T}_i$ denote the set of pilot slots assigned to the ith user. In each slot $\iota \in \mathcal{T}_i$, only the ith user transmits a known pilot symbol p_ι, while all other users remain silent. The pilot symbols are assumed to satisfy $\mathbb{E}\{|p_\iota|^2\} = 1$. Under this setting, the received signal in slot $\iota \in \mathcal{T}_i$ reduces to

$$y_\iota = \mathbf{w}_\iota^{\mathrm{T}} \mathbf{h}_i \, p_\iota + n_\iota. \tag{13}$$

3. Proposed Channel Estimation Method

A diffusion model–driven channel estimation framework tailored for PASS is proposed. As illustrated in Figure 3, the core idea is to reformulate the ill-conditioned and underdetermined uplink pilot inversion problem arising from limited pilot observations into a progressive reverse denoising diffusion inference process conditioned on the measured signals. In contrast to conventional methods that estimate channels via direct error minimization, the proposed approach learns an implicit prior distribution of PASS channels and enforces measurement consistency constraints during the reverse diffusion sampling stage. This design enables principled recovery of high-dimensional near-field channel vectors under limited antenna activation and constrained pilot resources.

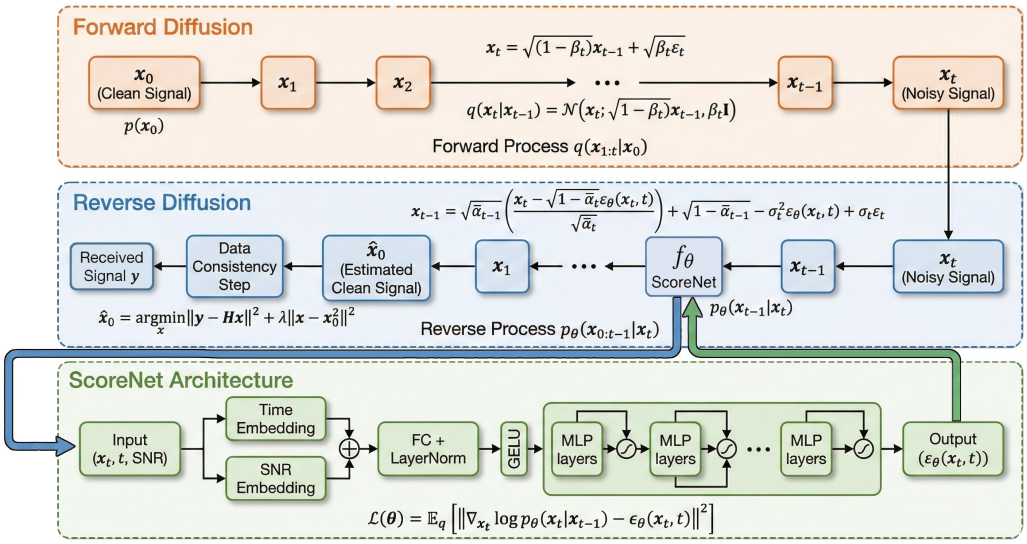


Figure 3. Proposed PADM model for channel estimation.

3.1. Observation Model and Problem Formulation

During the pilot transmission phase, for the ith user, the base station (BS) collects stacked pilot observations over the slots indexed by $\mathcal{T}_i$. The resulting linear observation model can be expressed as

$$\mathbf{y}_i = \mathbf{W}_i \mathbf{h}_i + \mathbf{n}_i, \tag{14}$$

where $\mathbf{y}_i \in \mathbb{C}^{|\mathcal{T}_i| \times 1}$ collects the received pilot observations, $\mathbf{h}_i \in \mathbb{C}^{N \times 1}$ denotes the wireless propagation channel vector from the ith user to the pinching antenna array, $\mathbf{W}_i \in \mathbb{C}^{|\mathcal{T}_i| \times N}$ is the effective observation matrix, and $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$ denotes the additive noise vector at the BS. The matrix $\mathbf{W}_i$ is formed by stacking the row vectors $\mathbf{w}_t^{\mathrm{T}}$ for all $t \in \mathcal{T}_i$, where $\mathbf{w}_t = \mathbf{A}_t \mathbf{g}$ is determined by the antenna activation pattern and the in-waveguide combining coefficients, and is assumed to be known at the BS.

To facilitate the application of diffusion-based generative models, which are typically defined over real-valued vector spaces, the complex-valued observation model is converted into an equivalent real-valued representation. Specifically, we define

$$\mathbf{h}_i^{\mathrm{R}} \triangleq \begin{bmatrix} \Re\{\mathbf{h}_i\} \\ \Im\{\mathbf{h}_i\} \end{bmatrix} \in \mathbb{R}^{2N}, \mathbf{y}_i^{\mathrm{R}} \triangleq \begin{bmatrix} \Re\{\mathbf{y}_i\} \\ \Im\{\mathbf{y}_i\} \end{bmatrix} \in \mathbb{R}^{2|\mathcal{T}_i|}, \tag{15}$$

and construct the corresponding real-valued sensing matrix as

$$\mathbf{W}_i^{\mathrm{R}} \triangleq \begin{bmatrix} \Re\{\mathbf{W}_i\} & -\Im\{\mathbf{W}_i\} \\ \Im\{\mathbf{W}_i\} & \Re\{\mathbf{W}_i\} \end{bmatrix} \in \mathbb{R}^{2|\mathcal{T}_i| \times 2N}. \tag{16}$$

With these definitions, the complex-valued observation model can be equivalently rewritten as the following real-valued linear system:

$$\mathbf{y}_i^{\mathrm{R}} = \mathbf{W}_i^{\mathrm{R}} \mathbf{h}_i^{\mathrm{R}} + \mathbf{n}_i^{\mathrm{R}}, \tag{17}$$

where $\mathbf{n}_i^{\mathrm{R}} \sim \mathcal{N}\left(\mathbf{0}, \frac{\sigma_n^2}{2} \mathbf{I}\right)$ denotes the real-valued noise vector obtained from the real and imaginary parts of $\mathbf{n}_i$.

The objective of this work is to infer the unknown channel vector $\mathbf{h}_i^{\mathrm{R}}$, and equivalently the original complex-valued channel $\mathbf{h}_i$, from the noisy and underdetermined observations $\mathbf{y}_i^{\mathrm{R}}$ given the known sensing matrix $\mathbf{W}_i^{\mathrm{R}}$. Rather than treating this as a deterministic inverse problem, this work seeks to formulate channel estimation from a generative and probabilistic inference perspective.

3.2. Forward Diffusion Modeling for PASS Channel Inference

Diffusion models aim to learn the prior distribution of high-dimensional random variables by progressively perturbing structured data into noise and subsequently performing inference through denoising. In this work, the near-field channel in PASS is treated as the latent variable of the diffusion process, such that the intrinsic statistical structure of PASS channels can be learned directly in the channel domain.

Specifically, the clean channel vector of the ith user is defined as the initial state of the diffusion process $\mathbf{x}_0 \triangleq \mathbf{h}_i^{\mathrm{R}} \in \mathbb{R}^{2N}$, where $\mathbf{h}_i^{\mathrm{R}}$ denotes the real-valued representation of the complex near-field channel. This formulation allows the diffusion model to capture key characteristics of PASS channels, including the coexistence of LoS and NLoS components, spatial non-stationarity, and long-range correlations induced by the waveguide geometry.

Due to sparse pinching antenna activation and limited pilot observations in practical PASS deployments, the underlying near-field channel state cannot be directly measured

or fully sampled and is therefore modeled as a latent variable x_0. Instead, the base station only acquires low-dimensional and noisy measurements constrained by the linear observation model

$$\mathbf{y}_i^{\mathrm{R}} = \mathbf{W}_i^{\mathrm{R}} \mathbf{x}_0 + \mathbf{n}_i^{\mathrm{R}}, \tag{18}$$

where $\mathbf{W}_i^{\mathrm{R}}$ denotes the known real-valued sensing matrix and $\mathbf{n}_i^{\mathrm{R}}$ represents additive noise. As a result, channel recovery constitutes a conditional inference problem subject to measurement constraints, rather than an unconditional data generation task.

To enable learning of the channel prior, a forward diffusion process is introduced to progressively inject Gaussian noise into $\mathbf{x}_0$. We let $\{\beta_t\}_{t=1}^{T}$ denote a predefined noise schedule, where β_t controls the noise variance injected at diffusion step t.

$$\alpha_t = 1 - \beta_t, \bar{\alpha}_t = \prod_{j=1}^{t} \alpha_j, \tag{19}$$

where α_t represents the signal preservation factor at step t and $\bar{\alpha}_t$ characterizes the cumulative attenuation of the original signal $\mathbf{x}_0$ up to step t.

The forward diffusion process follows a Markov chain with transition distribution

$$q(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)\mathbf{I}), \tag{20}$$

where $\mathbf{x}_t$ represents the noisy channel state at diffusion step t. By recursion, $\mathbf{x}_t$ admits the closed-form representation

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, \tag{21}$$

where $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ denotes standard Gaussian noise. This representation implies the marginal distribution

$$q(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}\big(\sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}\big). \tag{22}$$

As t increases, the structured PASS channel distribution is gradually transformed into an isotropic Gaussian distribution. The construction of a training tuple for the forward diffusion process is summarized in Algorithm 1.

Algorithm 1 Forward Diffusion Training Tuple Construction for PASS

1: **Input:** Channel dataset $\mathcal{D}$, diffusion steps T, noise schedule $\{\beta_t\}_{t=1}^{T}$, sensing matrix $\mathbf{W}_i^{\mathrm{R}}$, noise variance σ_n^2.
2: **Output:** A training tuple $(\mathbf{x}_t, t, \boldsymbol{\epsilon}, \mathbf{y}_i^{\mathrm{R}}, \mathbf{W}_i^{\mathrm{R}})$.
3: **Precompute:** $\alpha_t \leftarrow 1-\beta_t$, $\bar{\alpha}_t \leftarrow \prod_{j=1}^{t}\alpha_j$, for $t = \{1, \ldots, T\}$.
4: Sample a clean channel $\mathbf{x}_0 \sim \mathcal{D}$, where $\mathbf{x}_0 \triangleq \mathbf{h}_i^{\mathrm{R}}$.
5: Sample measurement noise $\mathbf{n}_i^{\mathrm{R}} \sim \mathcal{N}\Big(\mathbf{0}, \frac{\sigma_n^2}{2}\mathbf{I}\Big)$.
6: Form the pilot observation $\mathbf{y}_i^{\mathrm{R}} \leftarrow \mathbf{W}_i^{\mathrm{R}}\mathbf{x}_0 + \mathbf{n}_i^{\mathrm{R}}$.
7: Sample a diffusion step $t \sim \mathcal{U}\{1, \ldots, T\}$.
8: Sample injected noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
9: Construct the noisy latent $\mathbf{x}_t \leftarrow \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$.
10: **return** $(\mathbf{x}_t, t, \boldsymbol{\epsilon}, \mathbf{y}_i^{\mathrm{R}}, \mathbf{W}_i^{\mathrm{R}})$.

3.3. Conditional Reverse Diffusion with ScoreNet for PASS Channel Inference

The reverse diffusion process aims to progressively remove the injected noise and guide the samples from an isotropic Gaussian distribution back to the target channel

distribution. In this work, we adopt the noise-prediction parameterization and introduce a conditional ScoreNet to approximate the noise injected during the forward diffusion process. Specifically, a neural network is defined as

$$\boldsymbol{\epsilon}_{\boldsymbol{\theta}}\left(\mathbf{x}_t, t; \mathbf{y}_i^{\mathrm{R}}, \mathbf{W}_i^{\mathrm{R}}\right) \approx \boldsymbol{\epsilon}, \tag{23}$$

where $\boldsymbol{\theta}$ denotes the network parameters. Unlike unconditional diffusion models, PASS channel inference must explicitly account for the linear measurement constraint imposed by pilot observations. Therefore, the observation pair $(\mathbf{y}_i^{\mathrm{R}}, \mathbf{W}_i^{\mathrm{R}})$ is incorporated as conditioning information to steer the reverse denoising trajectory, such that the generated samples follow the learned channel prior while remaining consistent with the measurements. The overall conditional reverse diffusion procedure is summarized in Algorithm 2.

Algorithm 2 Conditional Reverse Diffusion for PASS Channel Inference

Input: Trained conditional ScoreNet $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\cdot)$; pilot observation $\mathbf{y}_i^{\mathrm{R}}$; sensing matrix $\mathbf{W}_i^{\mathrm{R}}$; diffusion steps T; schedule $\{\bar{\alpha}_t\}_{t=0}^{T}$ with $\bar{\alpha}_0 = 1$; step sizes $\{\eta_t\}_{t=1}^{T}$; noise variance σ_n^2.
Output: Estimated channel $\hat{\mathbf{h}}_i$.
Draw $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.
for $t = T, T-1, \ldots, 1$ **do**
  Predict noise $\hat{\boldsymbol{\epsilon}}_t \leftarrow \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{x}_t, t; \mathbf{y}_i^{\mathrm{R}}, \mathbf{W}_i^{\mathrm{R}})$.
  Estimate clean channel $\hat{\mathbf{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\hat{\boldsymbol{\epsilon}}_t\right)$.
  Deterministic reverse update $\tilde{\mathbf{x}}_{t-1} \leftarrow \sqrt{\bar{\alpha}_{t-1}}\,\hat{\mathbf{x}}_0 + \sqrt{1-\bar{\alpha}_{t-1}}\,\hat{\boldsymbol{\epsilon}}_t$.
  Compute data consistency gradient $\mathbf{g}_{t-1} \leftarrow \frac{1}{\sigma_n^2}(\mathbf{W}_i^{\mathrm{R}})^{\mathrm{T}}(\mathbf{W}_i^{\mathrm{R}}\tilde{\mathbf{x}}_{t-1} - \mathbf{y}_i^{\mathrm{R}})$.
  Measurement consistency correction $\mathbf{x}_{t-1} \leftarrow \tilde{\mathbf{x}}_{t-1} - \eta_t\,\mathbf{g}_{t-1}$.
end for
Recover complex channel $\hat{\mathbf{h}}_i \leftarrow \hat{\mathbf{x}}_0(1{:}N) + j\,\hat{\mathbf{x}}_0(N{+}1{:}2N)$.

3.3.1. ScoreNet Architecture

To accommodate the structural characteristics of PASS near-field channels, the ScoreNet is designed to jointly exploit spatial correlation and measurement conditioning. The input features are first mapped into a sequence of tokens via a linear projection layer. A Fourier embedding is then applied to enhance the representation of continuous spatial indices and long-range correlations. The diffusion step t is encoded through a time embedding and injected into the network. The backbone consists of a multi-layer Transformer encoder, whose self-attention mechanism captures global dependencies along the waveguide axis as well as non-local correlations induced by the coexistence of LoS and NLoS components. Finally, an output head produces the noise prediction $\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\cdot) \in \mathbb{R}^{2N}$. It is emphasized that the novelty of the proposed ScoreNet lies not in the use of a Transformer per se, but in the construction of a PASS-aware conditional denoiser that is jointly tailored to the near-field channel structure and the measurement constraints.

3.3.2. Training Objective

During training, clean channels $\mathbf{x}_0$ are sampled from a channel dataset. The diffusion step is uniformly sampled as $t \sim \mathcal{U}\{1, \ldots, T\}$, together with injected noise $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, to construct the noisy sample $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}$.

ScoreNet takes $(\mathbf{x}_t, t; \mathbf{y}_i^{\mathrm{R}}, \mathbf{W}_i^{\mathrm{R}})$ as input and is trained by minimizing the mean-squared error (MSE) loss

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{\mathbf{x}_0, t, \boldsymbol{\epsilon}}\left[\left\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{x}_t, t; \mathbf{y}_i^{\mathrm{R}}, \mathbf{W}_i^{\mathrm{R}})\right\|_2^2\right]. \tag{24}$$

This objective corresponds to the standard noise-prediction formulation derived from the variational lower bound, enabling the network to learn the multi-scale statistical structure of PASS channels across different noise levels.

3.3.3. Reverse Diffusion with Measurement Consistency

In the reverse diffusion stage, the noise prediction is first used to estimate the clean channel as

$$\hat{\mathbf{x}}_0(\mathbf{x}_t, t) = \frac{1}{\sqrt{\bar{\alpha}_t}}\Big(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{x}_t, t; \mathbf{y}_i^{\mathrm{R}}, \mathbf{W}_i^{\mathrm{R}})\Big). \tag{25}$$

Following a deterministic reverse diffusion update, an intermediate estimate of $\mathbf{x}_{t-1}$ is obtained as

$$\tilde{\mathbf{x}}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\,\hat{\mathbf{x}}_0(\mathbf{x}_t, t) + \sqrt{1-\bar{\alpha}_{t-1}}\,\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\mathbf{x}_t, t; \mathbf{y}_i^{\mathrm{R}}, \mathbf{W}_i^{\mathrm{R}}). \tag{26}$$

Since purely generative denoising may lead to samples that deviate from the actual measurements, a gradient-based measurement consistency correction is incorporated at each reverse step. Specifically, we define the data consistency cost function as

$$J(\mathbf{x}) = \frac{1}{2\sigma_n^2}\|\mathbf{y}_i^{\mathrm{R}} - \mathbf{W}_i^{\mathrm{R}}\mathbf{x}\|_2^2, \tag{27}$$

whose gradient is given by

$$\nabla_{\mathbf{x}} J(\mathbf{x}) = \frac{1}{\sigma_n^2}(\mathbf{W}_i^{\mathrm{R}})^{\mathrm{T}}\big(\mathbf{W}_i^{\mathrm{R}}\mathbf{x} - \mathbf{y}_i^{\mathrm{R}}\big). \tag{28}$$

At each reverse step, the final update is computed as

$$\mathbf{x}_{t-1} = \tilde{\mathbf{x}}_{t-1} - \eta_t \nabla_{\mathbf{x}} J(\tilde{\mathbf{x}}_{t-1}), \tag{29}$$

where $\eta_t > 0$ is a diffusion step and SNR-dependent step size that is adaptively scheduled during inference, rather than a fixed constant. Specifically, η_t is designed to balance the learned diffusion prior and the measurement consistency correction: a smaller step size is used in early reverse steps dominated by high noise to ensure stability, while a stronger correction is gradually allowed in later steps as the estimation approaches lower-noise regimes. The corresponding design rationale and derivation of the gradient-based correction are provided in Appendix A.

The reverse diffusion process is initialized from $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. After performing the reverse updates for $t = \{T, T-1, \ldots, 1\}$, the final estimate $\hat{\mathbf{x}}_0$ is obtained and mapped back to the complex-valued channel as

$$\hat{\mathbf{h}}_i = \hat{\mathbf{x}}_0(1{:}N) + j\,\hat{\mathbf{x}}_0(N{+}1{:}2N). \tag{30}$$

The resulting $\hat{\mathbf{h}}_i$ serves as the estimated uplink wireless propagation channel of the ith user over all pinching antenna locations.

4. Simulation Results

4.1. Simulation Setup and Benchmarking Methodology

To evaluate the performance of the proposed diffusion model-driven channel estimation framework for PASS, an indoor near-field multi-user uplink communication simulation platform is established. To reflect practical deployment constraints, the BS employs a dielectric waveguide as a unified receiving and combining structure, along which $Q = 100$

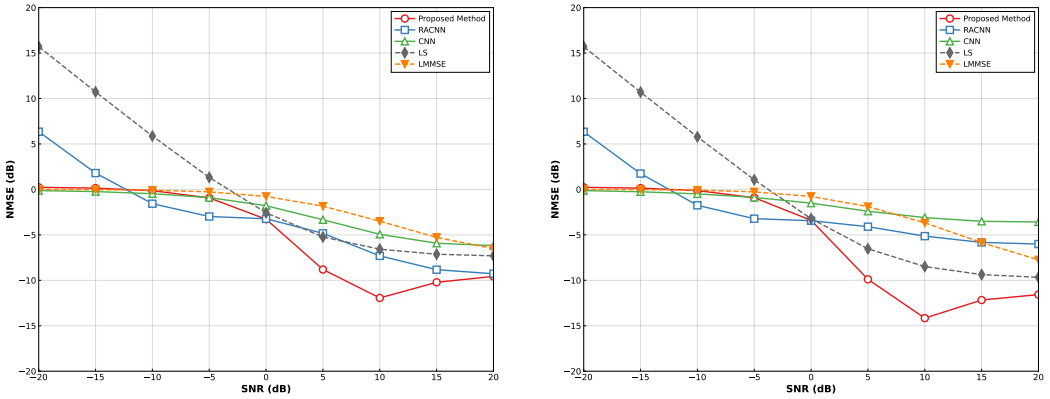
predefined discrete candidate locations are defined for selective pinching antenna activation, providing sufficiently fine spatial resolution while maintaining reasonable computational complexity. The system serves $I = 4$ single-antenna users randomly distributed in an indoor area of size 15×15 m^2. The waveguide is deployed at a height of $d = 3$ m, and the effective refractive index is set to $n_{\text{eff}} = 1$ as a system-level normalization to emphasize the relative propagation characteristics and structural effects of PASS, rather than detailed waveguide dispersion. The carrier frequencies are chosen as $f_c \in \{28, 73\}$ GHz. Two representative indoor propagation scenarios are considered. In the simple indoor (SI) scenario, each user–pinching antenna link contains $K = 4$ dominant scatterers, corresponding to relatively sparse reflections. In the complex indoor (CI) scenario, the number of scatterers is increased to $K = 8$ to model stronger NLoS effects induced by walls, furniture, and other indoor objects. In both scenarios, LoS paths are explicitly modeled whenever visibility conditions are satisfied.

Based on the above models, a PASS uplink channel dataset is generated. During training, the number of active pinching antennas is fixed to $N_{\text{tr}} = 16$, and 10^5 independent channel realizations are used for learning. During testing, the number of active pinching antennas is changed to $N_{\text{te}} = \{8, \ldots, 32\}$, which differs from the training configuration in order to evaluate the generalization and zero-shot capability of the proposed method under dynamic antenna activation. The diffusion length is set to $T = 1000$, with a linear noise schedule $\{\beta_t\}$ increasing from 10^{-4} to 0.02. Deterministic DDIM reverse diffusion is adopted during inference. The network input is organized as $[B, I, N_{\text{tr}}, 2]$, where B denotes the batch size, $I = 4$ is the number of users, $N_{\text{tr}} = 16$ is the number of pinching antennas during training, and the last dimension represents the real and imaginary parts of the channel coefficients. A six-layer MLP encoder with layer normalization and GELU activation is employed to capture long-range spatial correlations along the waveguide. The network is trained using the Adam optimizer with a cosine warm-up learning rate schedule, an initial learning rate of 3×10^{-4}, a batch size of $B = 256$, and 1000 training iterations.

For performance evaluation, the normalized mean square error (NMSE) is adopted, defined as $\text{NMSE} = \mathbb{E}\left\{\frac{\|\hat{\mathbf{h}}_i - \mathbf{h}_i\|_2^2}{\|\mathbf{h}_i\|_2^2}\right\}$. The proposed method is compared with classical linear estimators, including least-squares (LS) and linear minimum mean square error (LMMSE) estimators based on pinching antenna switching. In addition, two representative deep learning methods are included as learning-based baselines: (1) convolutional neural network (CNN), a 6-layer one-dimensional convolutional network with 64 filters per layer, batch normalization, and ReLU activation [19] and a (2) residual attention convolutional neural network (RACNN), consisting of 4 residual attention blocks that integrate both channel attention and spatial attention mechanisms, with global residual connections for learning channel corrections [20]. All baseline methods are evaluated under identical pilot transmission protocols, antenna activation patterns, and training data scales to ensure fair comparison.

4.2. Results and Analysis

As shown in Figure 4a for the simple indoor scenario at 28 GHz, the proposed method exhibits stable NMSE performance across the entire SNR range from -20 dB to 20 dB. In the low-SNR regime, the LS estimator suffers from severe noise amplification (NMSE of 15.71 dB at SNR $= -20$ dB), while RACNN degrades to 6.35 dB. In contrast, the proposed method maintains an NMSE of 0.23 dB at SNR $= -20$ dB, demonstrating strong robustness under noisy and underdetermined PASS observations. At medium-to-high SNRs, it achieves the best NMSE of -11.94 dB at SNR $= 10$ dB, outperforming CNN and RACNN by 7.0 dB and 4.6 dB, respectively, and consistently outperforms all baselines with smooth NMSE evolution.

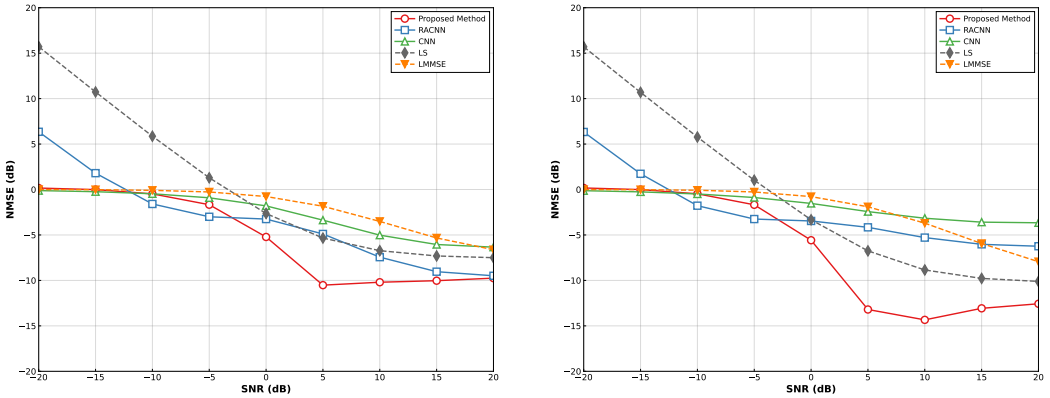


(**a**) 28 GHz (**b**) 73 GHz

Figure 4. NMSE performance versus SNR in simple indoor scenarios at different carrier frequencies.

A similar but more pronounced trend is observed in Figure 4b for the 73 GHz simple indoor scenario. Owing to stronger LoS dominance, the proposed method achieves improved performance of -14.16 dB at SNR $= 10$ dB. It preserves robust low-SNR performance (0.23 dB at SNR $= -20$ dB), outperforming LS by 15.5 dB. At SNR $= 10$ dB, the proposed method surpasses CNN and RACNN by 11.1 dB and 9.0 dB, respectively, while closely tracking and exceeding LMMSE in the medium-to-high SNR regime.

As shown in Figure 5a, in the complex indoor scenario at 28 GHz, the increased number of scatterers and the coexistence of LoS and NLoS components significantly increase the channel estimation difficulty, leading to performance degradation for most methods. In the low-SNR regime, the LS estimator suffers from severe noise amplification (NMSE of 15.71 dB at SNR $= -20$ dB), while learning-based baselines exhibit limited robustness. As SNR increases, LS gradually improves and outperforms CNN and LMMSE at high SNRs, reaching -7.50 dB at SNR $= 20$ dB, whereas RACNN achieves -9.49 dB and CNN remains around -6 dB. In contrast, the proposed method achieves the best NMSE of -10.52 dB at SNR $= 5$ dB and maintains stable performance around -10 dB across the medium-to-high SNR regime, consistently outperforming all baselines.



(**a**) 28 GHz (**b**) 73 GHz

Figure 5. NMSE performance versus SNR in complex indoor scenarios at different carrier frequencies.

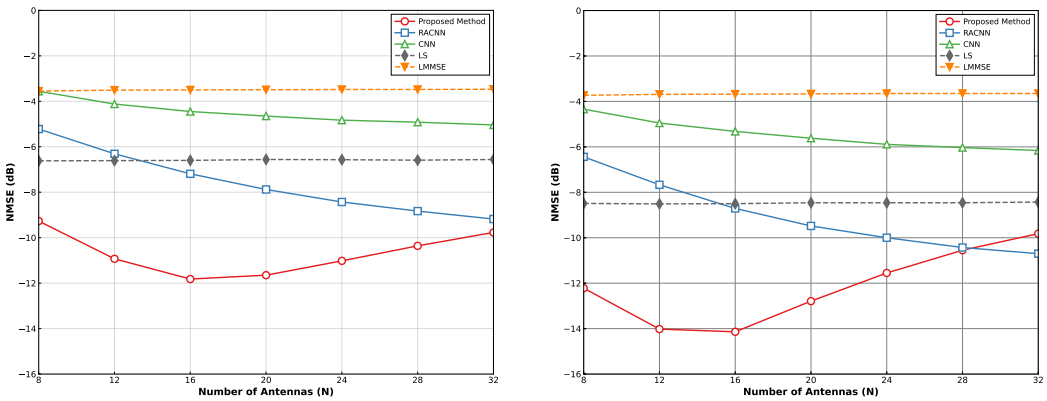
A similar trend is observed in Figure 5b for the 73 GHz complex indoor scenario, where higher carrier frequency further increases propagation loss and blockage sensitivity. The proposed method achieves -14.34 dB at SNR $= 10$ dB and maintains a stable NMSE floor between -12 dB and -14 dB. At high SNRs, LS improves to -10.11 dB at SNR $= 20$ dB, outperforming CNN and LMMSE, while CNN shows the weakest perfor-

mance (−3.66 dB). Overall, the proposed method provides gains exceeding 11 dB over CNN and 6 dB over LMMSE at SNR = 10 dB.

A comparison between Figures 4 and 5 reveals that increasing channel complexity affects different methods in fundamentally different ways. For the proposed method, increasing the number of scatterers from $K = 4$ to $K = 8$ does not degrade performance but instead leads to improved NMSE in certain SNR regimes. At 28 GHz, the NMSE improves from −8.81 dB to −10.52 dB at SNR = 5 dB, yielding a 1.7 dB gain, while at 73 GHz the improvement becomes more pronounced, increasing from −9.89 dB to −13.20 dB, corresponding to a 3.3 dB gain. This seemingly counter-intuitive behavior can be explained by a central limit theorem (CLT) effect: as the number of scatterers increases, the superposition of multiple random propagation paths causes the NLoS component to approach a Gaussian distribution, which is particularly well suited to the iterative denoising mechanism of diffusion models. The proposed generative framework naturally benefits from this statistical regularization, enabling more effective channel recovery in richer multipath environments.

In contrast, conventional estimators and discriminative learning-based baselines exhibit little sensitivity to the scatterer count. The NMSE of CNN changes by less than 0.1 dB between $K = 4$ and $K = 8$ across all SNRs, while RACNN and LMMSE show similarly negligible variation. These results indicate that such methods fail to exploit the evolving channel statistics induced by increased multipath richness, whereas the proposed diffusion-based estimator adaptively leverages CLT-induced regularity to achieve superior estimation accuracy.

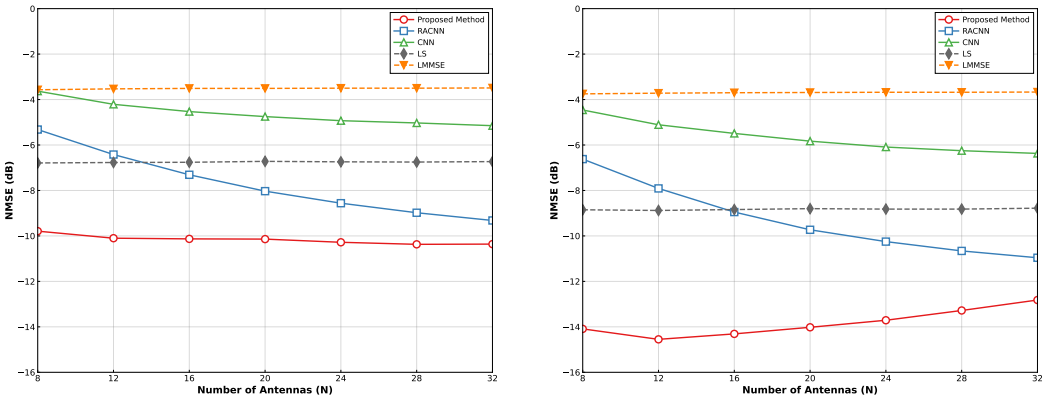
As shown in Figure 6, under SNR = 10 dB in simple indoor scenarios, the NMSE performance versus the number of pinching antennas N shows distinct adaptation behaviors across different estimation methods at both 28 GHz and 73 GHz. The linear baselines LS and LMMSE remain nearly flat with respect to N, indicating that increasing N provides no additional effective degrees of freedom under the single-waveguide PASS observation model. Learning-based baselines improve monotonically with N, with RACNN achieving an NMSE gain of about 4 dB as N increases from 8 to 32, and CNN showing a smaller improvement of 1.5–1.8 dB. The proposed method attains its best performance at moderate antenna scales (N = 12–16), reaching −11.8 dB at 28 GHz and −14.1 dB at 73 GHz. Although performance slightly degrades for larger N, it consistently outperforms all baselines across the entire range, with gains of 2–6 dB over RACNN and 6–10 dB over LMMSE, demonstrating robust and frequency-consistent behavior.



(**a**) 28 GHz (**b**) 73 GHz

Figure 6. NMSE performance versus the number of pinching antennas in simple indoor scenarios at SNR = 10 dB.

As shown in Figure 7, under the complex indoor scenario with SNR = 10 dB, the proposed method consistently achieves lower and more stable NMSE than all benchmark methods across different numbers of pinching antennas N at both 28 GHz and 73 GHz. At 28 GHz, the proposed method remains remarkably stable, ranging from −9.79 dB ($N = 8$) to −10.36 dB ($N = 32$), with a peak of −10.37 dB at $N = 28$ and a variation below 0.6 dB. In comparison, RACNN improves from −5.32 dB to −9.32 dB but still lags by 1–4.5 dB, while CNN remains 5–6 dB worse. LS shows limited sensitivity to N at about −6.75 dB, and LMMSE performs the worst at around −3.5 dB. At 73 GHz, the proposed method further improves, achieving NMSE values between −14.55 dB ($N = 12$) and −12.82 dB ($N = 32$), corresponding to a 2–4 dB gain over the 28 GHz case. RACNN improves to −10.96 dB at $N = 32$ but still trails by 2–7.5 dB, while CNN, LS, and LMMSE show only marginal gains with increasing N. Overall, Figure 7 demonstrates that the proposed method consistently outperforms conventional (LS, LMMSE) and learning-based (CNN, RACNN) estimators across antenna scales and carrier frequencies, delivering stable low NMSE rather than configuration-specific peak performance, which is well suited to practical PASS deployments with dynamic antenna activation and complex propagation conditions.



(**a**) 28 GHz (**b**) 73 GHz

Figure 7. NMSE performance versus the number of pinching antennas in complex indoor scenarios at SNR = 10 dB.

Figures 6a and 7a compare the channel estimation performance at 28 GHz under $K = 4$ and $K = 8$, respectively. When $K = 4$, the proposed method achieves NMSE values from −9.27 dB at $N = 8$ to a peak of −11.82 dB at $N = 16$, before increasing to −9.77 dB at $N = 32$, indicating an optimal operating region around $N = 16$ with a total variation of 2.55 dB. In contrast, when $K = 8$, the NMSE remains highly stable across all antenna configurations, confined to a narrow range of −9.79 dB to −10.37 dB (variation of 0.58 dB). Although the peak NMSE slightly degrades, the worst-case performance improves from −9.27 dB to −9.79 dB, demonstrating enhanced robustness. For learning-based baselines, RACNN and CNN exhibit nearly identical behaviors under both $K = 4$ and $K = 8$, while LS shows only a marginal improvement of about 0.15 dB and LMMSE remains unchanged at approximately −3.5 dB. Overall, at 28 GHz, increasing K mainly enhances the stability of the proposed method, whereas other approaches show limited sensitivity to user load variations.

Figures 6b and 7b illustrate the impact of user diversity on channel estimation at 73 GHz. When $K = 4$, the proposed method achieves its best NMSE of −14.14 dB at $N = 16$ but degrades to −9.82 dB at $N = 32$, resulting in a 4.32 dB variation and indicating increased sensitivity to antenna scaling under sparse user conditions. In contrast, when $K = 8$, the proposed method exhibits much improved stability, with NMSE values confined

to -14.55 dB to -12.82 dB (variation of 1.73 dB), and the worst-case NMSE improves by 3 dB, demonstrating enhanced robustness at large antenna scales. At $K = 8$, the proposed method delivers the best overall performance at 73 GHz, outperforming RACNN by 2–8 dB and CNN by 6–10 dB. LS achieves around -8.8 dB with only marginal improvement over the $K = 4$ case, while LMMSE remains near -3.7 dB with negligible sensitivity to K. Overall, increasing K from 4 to 8 simultaneously reduces NMSE and improves stability for the proposed method, which is particularly advantageous for dense millimeter-wave PASS deployments.

4.3. Discussion

4.3.1. Robustness Analysis

As shown in Figure 8, a sensitivity analysis is performed to assess the robustness of the proposed diffusion-based estimator with respect to key parameters. The number of DDIM inference steps S has negligible impact on performance: increasing S from 10 to 150 yields a nearly flat NMSE curve with a standard deviation of 0.034 dB and an overall variation below 0.1 dB, while $S = 50$ already achieves performance comparable to $S = 150$. The estimator is also highly robust to the initial reverse timestep t_{init}, with an NMSE variation below 0.05 dB and the lowest standard deviation (0.018 dB), owing to the LMMSE-based initialization. For the data consistency cap c_{dc}, increasing its value from 0.1 to around 1.5 results in a monotonic NMSE improvement of approximately 0.47 dB followed by saturation. In contrast, the DC step scale η_0 exhibits a stronger influence on performance, producing an almost linear NMSE reduction of about 1.65 dB as η_0 increases from 0.05 to 0.80. Finally, the NMSE shows a nonlinear dependence on SNR, being noise-limited at low SNRs, improving rapidly at moderate SNRs, and exhibiting mild degradation at high SNRs. Overall, the proposed estimator demonstrates strong robustness across inference and correction parameters, with the DC step scale serving as the dominant performance-controlling factor.

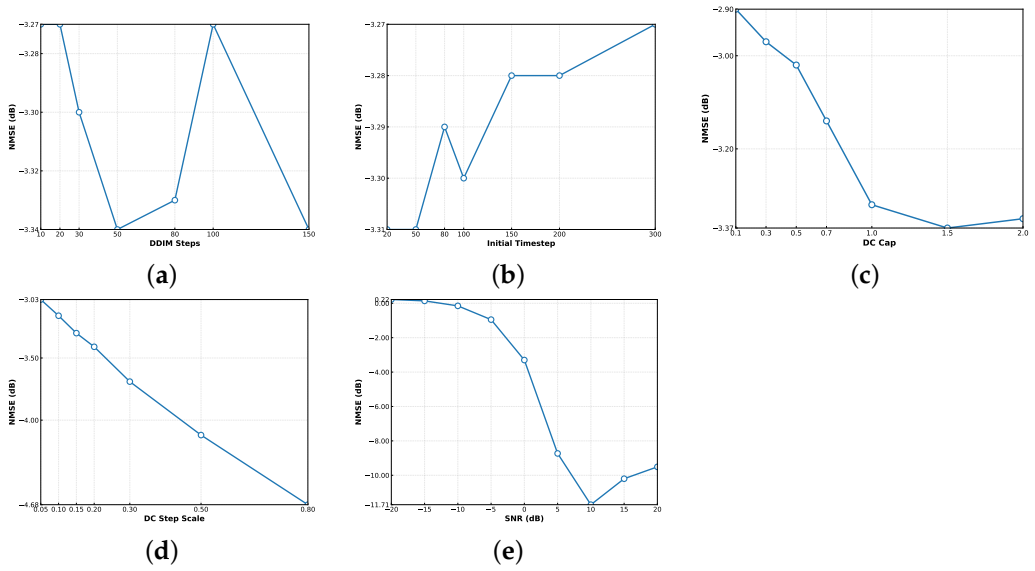


Figure 8. Robustness (**a**) sensitivity to the number of DDIM inference steps S; (**b**) sensitivity to the initialization timestep t_{init}; (**c**) sensitivity to the data-consistency cap c_{dc}; (**d**) sensitivity to the DC step scale η_0; (**e**) sensitivity to SNR.

4.3.2. Complexity Analysis

From the perspective of online inference, the computational complexity of different channel estimation methods is compared under identical system dimensions and observation settings. For conventional linear estimators, the dominant complexity is incurred by

matrix operations. Specifically, the LS estimator under underdetermined observations has a complexity of $\mathcal{O}(P^3 + P^2D)$, while the LMMSE estimator requires inversion of a $D \times D$ matrix and thus incurs a higher complexity of $\mathcal{O}(D^3 + D^2P)$. For deep learning-based methods such as CNN and RACNN, the online inference stage mainly consists of a finite number of one-dimensional convolution and attention operations, leading to a complexity that scales linearly with the channel dimension, i.e., $\mathcal{O}(D)$. The proposed diffusion model-driven approach introduces a multi-step reverse diffusion inference process, where each step involves one forward pass of the ScoreNet and one data consistency correction. As a result, the overall online inference complexity scales linearly with the number of diffusion steps S, i.e., $\mathcal{O}(S \cdot C_{\text{step}})$. Although additional iterative overhead is introduced, the complexity remains explicit and controllable, and avoids high-dimensional matrix inversion, making the proposed method suitable for underdetermined and sparsely activated scenarios. The computational complexity comparison is summarized in Table 3.

Table 3. Computational complexity comparison of channel estimation methods (online inference).

Method	Dominant Computational Complexity
LS	$\mathcal{O}(P^3 + P^2D)$
LMMSE	$\mathcal{O}(D^3 + D^2P)$
CNN	$\mathcal{O}(D)$
RACNN	$\mathcal{O}(D)$
Proposed Diffusion-Based Method	$\mathcal{O}(S \cdot C_{\text{step}})$

5. Conclusions

This paper studied the channel estimation problem for PASS in indoor near-field uplink communications and proposed a diffusion model-driven estimation framework tailored to the waveguide-fed and sparsely activated antenna architecture. By reformulating channel acquisition as a reverse denoising diffusion inference process, the proposed method exploits structural priors induced by near-field propagation, waveguide coupling, and multipath characteristics, enabling reliable channel recovery under highly underdetermined and noise-limited observations. Simulation results across simple and complex indoor scenarios at 28 GHz and 73 GHz demonstrate that the proposed approach achieves stable NMSE performance over a wide SNR range, outperforming conventional linear estimators and exhibiting improved robustness compared to existing learning-based baselines. These results indicate that the proposed framework is well suited for practical PASS deployments with uncertain channel and noise conditions. Several limitations of the proposed framework remain. As a data-driven generative inference approach, its effectiveness depends on the representativeness of the training data and may require retraining or fine-tuning when deployment scenarios deviate from the training distribution. Moreover, diffusion-related parameters introduce trade-offs between estimation accuracy and computational cost, and the current system model adopts simplified assumptions that do not fully capture practical hardware non-idealities, which merit further investigation in future work.

Author Contributions: Conceptualization, Y.C. and Q.W.; Methodology, Y.C.; Software, Y.C.; Validation, Y.C. and Q.W.; Formal analysis, Y.C.; Investigation, Y.C.; Resources, Q.W.; Data curation, Y.C.; Writing—original draft preparation, Y.C.; Writing—review and editing, Q.W.; Visualization, Y.C.; Supervision, Q.W.; Project administration, Q.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

We let T denote the total number of diffusion steps used during training and S denote the number of reverse diffusion steps used during inference. The gradient-based correction is applied during the inference stage. In the tth reverse diffusion inference step ($t = 0, 1, \ldots, S-1$), we define the normalized progress factor as

$$p_t \triangleq \frac{t}{\max(S-1, 1)} \in [0, 1], \tag{A1}$$

and set the step size for the gradient-based correction as

$$\eta_t = w_{\mathrm{dc}}(\mathrm{SNR}) \cdot \eta_0 \cdot \left(1 - \tfrac{1}{2} p_t\right) \cdot \left(1 + \tfrac{1}{2} p_t \, w_{\mathrm{dc}}(\mathrm{SNR})\right), \tag{A2}$$

where η_0 is a scale factor controlling the overall strength of the correction.

The SNR-adaptive weight is computed by first converting the SNR (in dB) to its linear scale

$$\mathrm{SNR}_{\mathrm{lin}} = 10^{\mathrm{SNR}/10}, \tag{A3}$$

and then defining an adaptive upper bound

$$c(\mathrm{SNR}) = 0.05 + 0.65 \cdot \sigma\left(\frac{\mathrm{SNR} + 5}{8}\right), \tag{A4}$$

where $\sigma(\cdot)$ denotes the sigmoid function. Finally, the data consistency (DC) weight is given by

$$w_{\mathrm{dc}}(\mathrm{SNR}) = \min\left(\frac{\mathrm{SNR}_{\mathrm{lin}}}{\mathrm{SNR}_{\mathrm{lin}} + 1}, \; c(\mathrm{SNR})\right). \tag{A5}$$

References

1. Cui, M.; Wu, Z.; Lu, Y.; Wei, X.; Dai, L. Near-field MIMO communications for 6G: Fundamentals, challenges, potentials, and future directions. *IEEE Commun. Mag.* **2023**, *61*, 40–46. [CrossRef]
2. Liu, Y.; Wang, Z.; Xu, J.; Ouyang, C.; Mu, X.; Schober, R. Near-field communications: A tutorial review. *IEEE Open J. Commun. Soc.* **2023**, *4*, 1999–2049. [CrossRef]
3. Chen, W.; Liu, Y.; Jafarkhani, H.; Eldar, Y.C.; Zhu, P.; Letaief, K.B. Signal processing and learning for next generation multiple access in 6G. *IEEE J. Sel. Top. Signal Process.* **2024**, *18*, 1146–1177. [CrossRef]
4. Yang, Z.; Wang, N.; Sun, Y.; Ding, Z.; Schober, R.; Karagiannidis, G.K. Pinching antennas: Principles, applications and challenges. *arXiv* **2025**, arXiv:2501.10753. [CrossRef]
5. Ding, Z.; Schober, R.; Poor, H.V. Flexible-antenna systems: A pinching-antenna perspective. *IEEE Trans. Commun.* **2025**, *73*, 9236–9253. [CrossRef]
6. Wang, K.; Ding, Z.; Schober, R. Antenna activation for NOMA assisted pinching-antenna systems. *IEEE Wireless Commun. Lett.* **2025**, *14*, 1526–1530. [CrossRef]
7. Wang, K.; Ding, Z.; Karagiannidis, G.K. Antenna activation and resource allocation in multi-waveguide pinching-antenna systems. *IEEE Trans. Wireless Commun.* **2025**, *25*, 4070–4082. [CrossRef]
8. Ding, Z.; Poor, H.V. LoS blockage in pinching-antenna systems: Curse or blessing? *IEEE Wireless Commun. Lett.* **2025**, *14*, 2798–2802. [CrossRef]
9. Wang, K.; Ouyang, C.; Liu, Y.; Ding, Z. Pinching-antenna systems with LoS blockages. *IEEE Wireless Commun. Lett.* **2025**, *14*, 4122–4126. [CrossRef]
10. Xu, X.; Mu, X.; Wang, Z.; Liu, Y.; Nallanathan, A. Pinching-antenna systems (PASS): Power radiation model and optimal beamforming design. *arXiv* **2025**, arXiv:2505.00218. [CrossRef]
11. Tegos, S.A.; Diamantoulakis, P.D.; Ding, Z.; Karagiannidis, G.K. Minimum data rate maximization for uplink pinching-antenna systems. *IEEE Wireless Commun. Lett.* **2025**, *14*, 1516–1520. [CrossRef]
12. Hou, T.; Liu, Y.; Nallanathan, A. On the performance of uplink pinching antenna systems (PASS). *IEEE Trans. Commun.* **2025**, *74*, 92–105. [CrossRef]

13. Rezaei, F.; Marvasti-Zadeh, S.M.; Tellambura, C.; Maaref, A. Adversarial score-based generative models for MMSE-achieving AmBC channel estimation. *IEEE Wireless Commun. Lett.* **2024**, *13*, 1053–1057. [CrossRef]
14. Xiao, J.; Wang, J.; Liu, Y. Channel estimation for pinching-antenna systems (PASS). *IEEE Commun. Lett.* **2025**, *29*, 1789–1793. [CrossRef]
15. Letafati, M.; Ali, S.; Latva-Aho, M. Diffusion model-aided data reconstruction in cell-free massive MIMO downlink: A computation-aware approach. *IEEE Wireless Commun. Lett.* **2024**, *13*, 3162–3166. [CrossRef]
16. Letafati, M.; Ali, S.; Latva-Aho, M. Conditional denoising diffusion probabilistic models for data reconstruction enhancement in wireless communications. *IEEE Trans. Mach. Learn. Commun. Netw.* **2025**, *3*, 133–146. [CrossRef]
17. Zhou, X.; Liang, L.; Zhang, J.; Jiang, P.; Li, Y.; Jin, S. Generative diffusion models for high dimensional channel estimation. *IEEE Trans. Wireless Commun.* **2025**, *24*, 5840–5854. [CrossRef]
18. Wang, Y.; Xu, Y.; Zhang, C.; Chen, Z.; Dai, M.; Wang, H. Channel estimation for RIS-assisted mmWave systems via diffusion models. *IEEE Commun. Lett.* **2025**, *30*, 597–601. [CrossRef]
19. Wen, C.; Jin, S.; Matthaiou, M.; Xiao, C.; Chen, S. Deep CNN-based channel estimation for mmWave massive MIMO systems. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 989–1000.
20. Lam, V.T.; Son, D.H.; Quynh, T.T.T.; Le, T.T. RACNN: Residual attention convolutional neural network for near-field channel estimation in 6G wireless communications. In *Proceedings of the 14th Conference on Information Technology and its Applications (CITA), Phnom Penh, Cambodia, 14–25 July 2025* ; Springer: Cham, Switzerland, 2025.