# Device Activity Detection and Channel Estimation Using Score-Based Generative Models in Massive MIMO

TANG Chenyue[1], LI Zeshen[1], CHEN Zihan[2], Howard H. YANG[1]
(1. ZJU-UIUC Institute, Zhejiang University, Haining 314400, China;
2. Singapore University of Technology and Design, Singapore 487372, Singapore)

DOI: 10.12142/ZTECOM.202501007
https://kns.cnki.net/kcms/detail/34.1294.TN.20250312.1558.002.html, published online March 13, 2025
Manuscript received: 2025-01-02

**Abstract:** The growing demand for wireless connectivity has made massive multiple-input multiple-output (MIMO) a cornerstone of modern communication systems. To optimize network performance and resource allocation, an efficient and robust approach is joint device activity detection and channel estimation. In this paper, we present an approach utilizing score-based generative models to address the underdetermined nature of channel estimation, which is data-driven and well-suited for the complex and dynamic environment of massive MIMO systems. Our experimental results, based on a comprehensive dataset generated through Monte-Carlo sampling, demonstrate the high precision of our channel estimation approach, with errors reduced to as low as −45 dB, and exceptional accuracy in detecting active devices.

**Keywords:** activity detection; channel estimation; inverse problem; score-based generative model; massive MIMO

**Citation** (Format 1): TANG C Y, LI Z S, CHEN Z H, et al. Device activity detection and channel estimation using score-based generative models in massive MIMO [J]. ZTE Communications, 2025, 23(1): 53 – 62. DOI: 10.12142/ZTECOM.202501007
**Citation** (Format 2): C. Y. Tang, Z. S. Li, Z. H. Chen, et al., "Device activity detection and channel estimation using score-based generative models in massive MIMO," *ZTE Communications*, vol. 23, no. 1, pp. 53 – 62, Mar. 2025. doi: 10.12142/ZTECOM.202501007.

## 1 Introduction

### 1.1 Motivation

The advent of the Internet of Things (IoT) era is marked by a significant increase in the number of connected devices, each capable of sensing and communicating, which has brought about a new set of challenges in network connectivity[1 – 2]. The IoT, with its expected massive device connectivity, is poised to revolutionize various aspects of daily life and socio-economic activities, from smart homes and cities to healthcare applications. These applications require ubiquitous connectivity, making massive machine-type communications (mMTC) a critical component of the upcoming 6G networks[3]. MMTC aims to provide wireless connectivity to a vast number of devices with low-complexity and low-power, which is essential for realizing IoT-based applications but also poses significant challenges in terms of network management and efficiency[4 – 6].

One of the key enablers for mMTC is the massive multiple-input multiple-output (MIMO) technology[7], which is expected to significantly improve spectral and energy efficiency at the base station (BS) level. However, a major challenge lies in acquiring accurate channel state information (CSI) for mMTC, as the pilot-aided training overhead for uplink channel estimation scales with the number of devices, which can be extremely large in a massive connection scenario[8]. A typical characteristic of mMTC traffic is its sporadic pattern, with most devices designed to remain in sleep mode for energy conservation and only a limited number active for data transmission at any given time interval[9]. This sporadic nature entails the design of joint device activity detection and channel estimation to reduce the training overhead for channel estimation.

Traditional methods for channel estimation often use dimension reduction techniques (e.g., the discrete Fourier transform) to reduce the pilot sequence length and computational complexity, which may lead to performance degradation due to the off-grid effect and energy leakage[10]. These methods also fail to capitalize on the common sparsity across different frequency bands. To address these limitations, a novel sparse Bayesian learning (SBL) framework for joint device activity detection and channel estimation has been proposed, exploiting additional sparsity structures to significantly enhance sparse

**TANG Chenyue and LI Zeshen are co-first authors.**

recovery performance[11].

Inspired by the potential of score-based generative models in specialized applications such as magnetic resonance imaging (MRI) reconstruction[12–13], we introduce a training and inference algorithm for wireless channel estimation using score-based generative models in massive MIMO communication scenarios. This approach models the log-distribution of channels by learning the high-dimensional gradient, known as the score, providing a distribution learning framework for modeling high-dimensional millimeter-wave (mmWave) MIMO channels in a stochastic environment. Unlike traditional methods, our approach uses score-based generative models to learn the score of the distribution in an unsupervised manner, independent of pilot symbols. Device activity detection and probabilistic channel estimation are achieved by sampling from the posterior distribution using annealed Langevin dynamics, tackling challenges in out-of-distribution settings, wide signal-to-noise ratio (SNR) ranges, and interference scenarios.

### 1.2 Related Work

Many problems in engineering applications, ranging from signal processing and computer vision to machine learning and statistics, can be formulated as linear inverse problems[14]. To solve these linear inverse problems, researchers have proposed various approaches, such as compressed sensing methods[15] and deep learning techniques[16]. Pre-trained generative priors have also been used in solving linear inverse problems, surpassing classical compressed sensing approaches[17–18]. With the emergence of deep generative models in density estimation[19–21], there has been a surge of interest in developing linear inverse algorithms with data-driven priors[22–23]. Owing to the powerful representational capabilities of deep generative models, they can effectively learn accurate prior knowledge given sufficient data samples[14]. Their potential in solving linear inverse problems is gaining increasing attention.

Massive connectivity is a key requirement for future wireless cellular networks to support mMTC[5]. In large-scale wireless cellular networks, user detection and channel estimation can be viewed as high-dimensional linear inverse problems, as scheduling a large number of occasionally active users on a separate control channel may incur significant overhead. Studies such as Refs. [24] and [25] investigate a random access protocol in which each active user picks one of the orthogonal signature sequences at random and sends it to the BS, and a connection is established if the selected preamble is not used by the other users. Refs. [5] and [26] propose the use of the approximate message passing (AMP)[27] algorithm for joint user activity detection and channel estimation, and further show that a state evolution analysis[28] of the AMP algorithm allows an analytic characterization of the missed detection and false alarm probabilities for device detection.

As a novel class of generative models, diffusion models (DM), also known as score-based generative models, have achieved remarkable performance in density estimation and image generation[20, 29]. Originally, DM was introduced for unconditional image generation; however, they have since been widely applied to conditional probability distributions, enabling tasks such as conditional image generation[30]. Supervised end-to-end training of deep learning-based methods has been successfully applied to wireless MIMO channel estimation[31–32], introducing a powerful and robust deep learning algorithm in the form of the learned denoising approximate message passing (L-DAMP) algorithm[33]. Furthermore, Ref. [34] employs annealed Langevin dynamics and score-based models to efficiently train generative models on simulated datasets, achieving performance superior to that of generative adversarial networks (GANs).

### 1.3 Contributions

The principal contributions of this paper are summarized as follows:

• We introduce an approach that leverages score-based generative models to achieve joint active device detection and channel estimation for massive MIMO communications. Our solution delivers accurate estimates without imposing any assumptions on the dimensionality or sparsity of the channels, thereby providing a flexible and robust method for real-world applications.

• We generate simulated data with varying sizes and complexity, closely capturing the diverse and dynamic nature of massive MIMO environments. This capability allows our model to be trained and tested under conditions that accurately reflect real-world wireless propagation scenarios.

• Through extensive numerical simulations, we validate the effectiveness of our method. The results indicate that the accuracy of active device detection exceeds 98% under high SNR conditions. Additionally, the normalized mean square error (NMSE) can be reduced to as low as −45 dB, highlighting the superior performance of our approach in channel state estimation and active user detection in massive MIMO systems.

### 1.4 Organization

The remainder of this paper is organized as follows. Section 2 presents the massive MIMO system model, the inverse problem, and the procedures involved in score-based generative models. Section 3 details the training phase of the proposed method, focusing on the generation of channel data and the training of the score function, as well as the inference (testing) stage. Section 4 provides numerical simulation results and discussions. Sections 5 and 6 conclude the paper and present the future work.

## 2 Preliminaries

### 2.1 Massive MIMO

Massive MIMO is a key technology for next-generation wireless communication systems, characterized by the deployment of a large number of antennas at the BS to serve multiple users simultaneously[35]. This configuration allows for significant improvements in spectral efficiency, energy efficiency, and overall system performance. By leveraging spatial multiplexing and beamforming techniques, massive MIMO can effectively mitigate interference, increase data rates, and improve the reliability of wireless links. The large number of antennas enables the BS to exploit the spatial diversity of the channel, leading to more precise CSI estimation and better resource allocation. As a result, massive MIMO serves as a key enabler for next-generation wireless networks, including 5G and beyond, addressing the growing demand for high-speed, low-latency, and high-capacity communication services[36].

### 2.2 Inverse Problem

Inverse problems are ubiquitous in various scientific and engineering fields, where the goal is to infer the parameters or states of a system from observed data. A common linear model used to describe inverse problems is expressed as:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{P} + \boldsymbol{Z} \quad (1),$$

where $\boldsymbol{Y}$ represents the observed data, $\boldsymbol{X}$ is the unknown system matrix or operator, $\boldsymbol{P}$ is the known parameter vector, and $\boldsymbol{Z}$ is the noise term. The objective is to estimate $\boldsymbol{X}$ from $\boldsymbol{Y}$ and $\boldsymbol{P}$, but reconstructing the underlying causes from their observed effects is inherently complex, particularly in real-world scenarios. This challenge is further compounded by the presence of noise and the potential for the problem to be ill-posed, meaning that solutions may not exist, may not be unique, or may be excessively sensitive to noise. To address these issues, regularization techniques such as Tikhonov or total variation regularization are commonly applied. These methods add constraints to the problem to stabilize the solution[37].

In situations where the number of unknowns surpasses the number of measurements, referred to as under-determined problems, the challenge intensifies. The disparity between the number of unknowns and the available data leads to a scenario with an infinite number of potential solutions to $\boldsymbol{X}$ that could align with the equation $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{P} + \boldsymbol{Z}$. This scenario is particularly problematic as it significantly increases the risk of inaccurate or unstable solutions[13]. To combat these difficulties, optimization methods and Bayesian approaches are employed. These strategies incorporate prior knowledge and provide a framework for managing the uncertainty associated with the estimates. Furthermore, recent progress in machine learning and generative modeling has introduced innovative approaches to address these challenges. These advancements offer new methods to handle the instability and uncertainty inherent in under-determined problems, thereby improving the reliability and accuracy of the solutions derived from noisy and incomplete data.

Specifically, score-based generative models have shown promise in addressing under-determined inverse problems by leveraging the underlying data distribution to generate plausible solutions. These models represent a powerful approach capable of capturing the complex structures of high-dimensional data distributions without explicit parametric forms, making them especially suitable for applications where data distribution is complex or not easily characterized by traditional models[17]. Such applications are particularly relevant for real-world wireless environments.

### 2.3 Score-Based Generative Model

A core hypothesis of this study is that the characteristics of wireless channels can be represented as samples drawn from a common probability distribution, which has been widely adopted in both theoretical and practical wireless communications research[38]. Score-based generative models, which have demonstrated their effectiveness on natural image benchmark datasets, are a class of generative models that generate data by estimating the gradient of the data distribution[13]. This approach diverges from traditional generative modeling techniques, which often rely on explicit parameterization of the data distribution. Instead, score-based generative models learn the gradient field of the data distribution in a non-parametric fashion, providing a flexible framework for capturing complex data distributions[20]. Fig. 1 fully displays the process of handling inverse problems using a score-based generative model.

#### 2.3.1 Learning Score Function

The score function for a point $\boldsymbol{X}$ is represented as:

$$\boldsymbol{\psi}_{\boldsymbol{X}}(\boldsymbol{X}) = \nabla \log p_{\boldsymbol{X}}(\boldsymbol{X}) \quad (2),$$

where $\boldsymbol{X}$ denotes the data point, $p_{\boldsymbol{X}}(\boldsymbol{X})$ is the probability density distribution of this data point, and $\boldsymbol{\psi}_{\boldsymbol{X}}(\boldsymbol{X})$ is a matrix of size $M \times N$. The score function encapsulates the local density

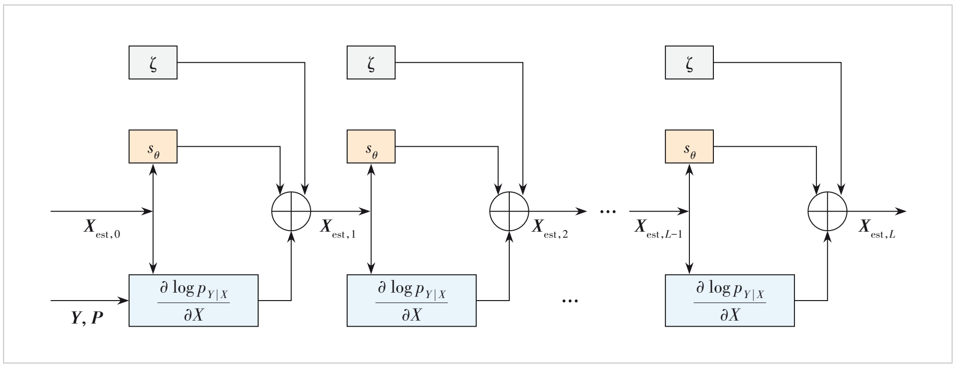



**Figure 1. A step-by-step process for estimating $X$ by employing a score-based model in conjunction with the known matrices $Y$ and $P$**

information of the data distribution, which is instrumental in the generative process. In practice, $\boldsymbol{\psi}_X(\boldsymbol{X})$ can be used to guide the optimization process for channel estimation by iteratively updating the channel estimate in the direction that maximizes the likelihood of the observed data. For example, if the score function indicates a high likelihood of a certain channel coefficient being non-zero, the algorithm can focus on refining the estimate of that coefficient, leading to more accurate channel estimation overall. The goal is to learn a model $s_\theta$ capable of generating $s_\theta(\boldsymbol{X})$ to approximate $\boldsymbol{\psi}_X(\boldsymbol{X})$.

### 2.3.2 Denoising Score Matching

While $\boldsymbol{\psi}_X$ and the explicit score matching $\mathbb{E}_{X\sim p_X}\left[\left\| s_\theta(\boldsymbol{X}) - \boldsymbol{\psi}_X(\boldsymbol{X})\right\|_2^2\right]$ are often intractable, denoising score matching is proposed to address this issue. Ref. [39] demonstrates that the loss function $\mathcal{L}(s_\theta)$ we used for training is equivalent to the loss function of the explicit score matching, as long as $\log p_{\tilde{X}|X}(\tilde{\boldsymbol{X}}|\boldsymbol{X})$ is differentiable with respect to $\tilde{\boldsymbol{X}}$. This approach transforms the task of learning the score function of the original data distribution (which is nearly impossible in the real world) into learning the score of the perturbed distribution by using $\mathcal{L}(s_\theta)$. By synthesizing corrupted data samples $\tilde{\boldsymbol{X}}$ and learning the score of the conditional distribution $p_{\tilde{X}|X}$, the following objective is used:

$$\mathcal{L}(s_\theta) = \mathbb{E}_{X\sim p_X, \tilde{X}\sim p_{\tilde{X}}}\left[\left\| s_\theta(\tilde{\boldsymbol{X}}) - \nabla \log p_{\tilde{X}|X}(\tilde{\boldsymbol{X}}|\boldsymbol{X})\right\|_2^2\right] \quad (3).$$

Since IoT allows the use of arbitrary noise distributions for training and learning the score at arbitrarily perturbed inputs, we set the perturbation $\boldsymbol{U}$ as i.i.d. Gaussian, with zero mean and covariance matrix $\sigma_U^2\boldsymbol{I}$, i.e.,

$$\nabla \log p_{\tilde{X}|X}(\tilde{\boldsymbol{X}}|\boldsymbol{X}) = -\boldsymbol{U}/\sigma_U^2 \quad (4).$$

A learnable model proposed by Ref. [20] is used to learn $s_\theta$. The model (in our work, such a deep neural network described in Section 3) uses a weighted version of $\mathcal{L}(s_\theta)$ at multiple noise levels to train a single score-based model for an individual datum within a batch, represented by:

$$\mathcal{L}_{\text{score}}(\theta) = \mathbb{E}_{j, X\sim p_X, U_j\sim p_{U_j}}\left[\sigma_{U_j}^2\left| s_\theta(\boldsymbol{X} + \boldsymbol{U}_j) + \frac{\boldsymbol{U}_j}{\sigma_{U_j}^2}\right|_2^2\right] \quad (5).$$

Weighing the predicted score at each noise level is to formulate denoising score matching as a variance-exploding (VE) diffusion process[40].

### 2.3.3 Posterior Sampling Using Score Functions

Once the score function is learned, it can be used to perform posterior sampling, which is a key step in the channel estimation process. Posterior sampling involves drawing samples from the posterior distribution of the CSI conditioned on the received pilot symbols. Given the known matrices $\boldsymbol{Y}$ and $\boldsymbol{P}$, the posterior distribution of matrix $\boldsymbol{X}$ can be expressed using the Bayes' rule:

$$p_{X|Y}(\boldsymbol{X}|\boldsymbol{Y}) = \frac{p_{Y|X}(\boldsymbol{Y}|\boldsymbol{X}) \cdot p_X(\boldsymbol{X})}{p_Y(\boldsymbol{Y})} \quad (6).$$

Expanding the logarithm of the posterior distribution, we get

$$\log p_{X|Y}(\boldsymbol{X}|\boldsymbol{Y}) = \log p_{Y|X}(\boldsymbol{Y}|\boldsymbol{X}) + \log p_X(\boldsymbol{X}) - \log p_Y(\boldsymbol{Y}) \quad (7).$$

Taking the gradient with respect to $\boldsymbol{X}$, we obtain

$$\nabla \log p_{X|Y}(\boldsymbol{X}|\boldsymbol{Y}) = \nabla \log p_{Y|X}(\boldsymbol{Y}|\boldsymbol{X}) + \nabla \log p_X(\boldsymbol{X}) \quad (8),$$

since $\nabla \log p_Y(\boldsymbol{Y}) = 0$. For all $\boldsymbol{Y}$, the gradient of the posterior distribution simplifies to:

$$\boldsymbol{\psi}_{X|Y}(\boldsymbol{X}|\boldsymbol{Y}) = \boldsymbol{\psi}_{Y|X}(\boldsymbol{Y}|\boldsymbol{X}) + \boldsymbol{\psi}_X(\boldsymbol{X}) \quad (9).$$

This result shows that the gradient of the posterior distribution is a combination of the gradient of the likelihood function and the gradient of the prior distribution. The likelihood function is derived from $\boldsymbol{Y}$, while the prior distribution is learned using the score-based generative model.

### 2.3.4 Annealed Langevin Dynamics for Posterior Sampling

To sample from the posterior distribution, we use annealed Langevin dynamics, which is an iterative process that updates the channel estimate $\boldsymbol{X}_{\text{est}}$ in a manner that maximizes the posterior probability. We introduce time-varying hyperparameters $\alpha_t$ and $\beta_t$ as an enhancement, based on the method proposed in Ref. [41]. The update rule for annealed Langevin dynamics is given by

$$\boldsymbol{X}_{\text{est}, i+1} = \boldsymbol{X}_{\text{est}, i} + \alpha_i \cdot \Big(\nabla \log p_{Y|X}(\boldsymbol{Y}|\boldsymbol{X}_{\text{est}, i}) + \nabla \log p_X(\boldsymbol{X}_{\text{est}, i})\Big) + \sqrt{2\beta \cdot \alpha_i} \cdot \sigma_{U_i} \cdot \zeta \quad (10),$$

where $\alpha_i$ is the step size that decays over time; $\beta$ is a hyperparameter that controls the amount of noise added to the update; $\sigma_{U_i}$ is the noise level at the $i$-th step; $\zeta\sim\mathcal{CN}(0, \boldsymbol{I})$ is Gaussian noise added to maintain diversity in the samples. The parameters $\alpha_i$, $\beta$, and $\sigma_{U_i}$ are critical for the performance of the proposed method. The learning rate $\alpha_i$ is chosen through a grid search to balance convergence speed and accuracy. The initial value of the regularization parameter $\beta$ is empirically set to 0.9 for robustness to noise. The noise variance $\sigma_{U_i}$ is estimated from the training data using a maximum likelihood approach.

The gradient of the likelihood function $\nabla \log p_{Y|X}(\boldsymbol{Y}|\boldsymbol{X}_{\text{est}, i})$ can be derived from $\boldsymbol{P}$. For Gaussian noise, this gradient is given by

$$\nabla \log p_{Y|X}(Y|X_{\text{est},i}) = \frac{(X_{\text{est},i}P - Y)P^H}{\sigma^2_{\text{pilot}}} \tag{11}$$

The term $\nabla \log p_X(X_{\text{est},i})$ represents the gradient of the prior distribution, which is learned using the score-based generative model. This gradient is approximated by the learned score function $s_\theta(X_{\text{est},i})$.

## 3 System Model

Consider a single-cell massive MIMO network, where a BS equipped with $M$ antennas serves $N$ potential users, denoted by the set $\mathcal{N} = \{1,\cdots,N\}$. Each user device is equipped with a single antenna. This setup is typical for an uplink massive access scenario, where the BS efficiently manages data transfer and communication from numerous users within its coverage area. Fig. 2 illustrates an example.

In our system model for device activity detection and channel estimation, the sporadic nature of user traffic can be characterized by a user activity indicator for each user. We denote this indicator by

$$\lambda_n = \begin{cases} 1, & \text{if user } n \text{ is active} \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

The probability of a user being active is $\epsilon$, and the probability of being inactive is $1-\epsilon$, such that $\Pr[\lambda_n = 1] = \epsilon$ and $\Pr[\lambda_n = 0] = 1-\epsilon$. The set of active users within a coherence block is defined as $\mathcal{K} = \{n : \lambda_n = 1\}$, and the number of active users is $K = |\mathcal{K}|$.

The transmitted signal for each user $n$ is given by

$$x_n = \lambda_n h_n \tag{13}$$

where $h_n$ represents the channel coefficient for user $n$. The matrix $X$ is formed by stacking the transmitted signals of all users, i.e., $X = [x_1,\cdots,x_N]^T$.

During the training phase, the BS receives a matrix $Y$, which is modeled as the product of the transmitted signal matrix $X$, the pilot matrix $P$, and the addition of additive white Gaussian noise $Z$. The model can be expressed as

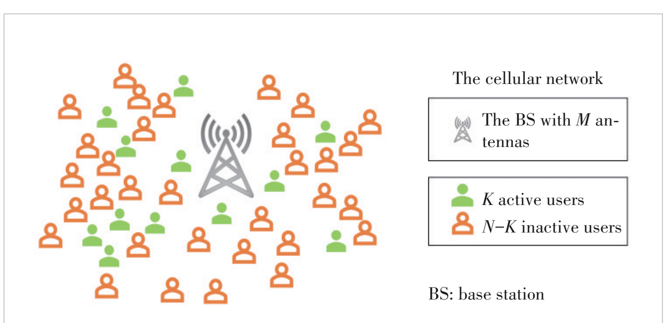



**Figure 2. System model of a massive device communication network**

$$Y = XP + Z \tag{14}$$

Here, the channel state information matrix $X \in \mathbb{C}^{M\times N}$, which is a complex matrix of size $M$ times $N$, where $M$ represents the number of receive antennas and $N$ denotes the total number of users. $Z$ is the Gaussian noise matrix with elements distributed as $\mathcal{CN}[0,\sigma^2 I]$. $P$ is the pilot matrix where each entry is a randomly chosen (fixed for all test samples) quadrature phase shift keying (QPSK) symbol with unit amplitude and low-resolution phase. Pilot symbols $L_{\text{pilot}}$ are selected from a pre-designed codebook, with each symbol $p_i$ belonging to $\mathbb{C}^N$. These pilot symbols are utilized to facilitate the estimation process. The transmitted pilot matrix $P$ is constructed from these symbols, and it is common practice in communication standards to pre-specify these pilot sequences.

The pilot density is defined by $\alpha = L_{\text{pilot}}/N$. When $\alpha < 1$, it implies that the number of received pilots is less than the total number of possible pilot transmissions, i.e., $ML_{\text{pilot}} < MN$. This situation leads to an under-determined inverse problem for channel estimation, where there are more unknowns (channel coefficients) than the number of equations provided by the received pilots.

Following the methodology outlined in Ref. [34], we employ a score-based generative model to accomplish two critical tasks: channel estimation and device activity detection. The pseudocode is shown in Algorithm 1. This model operates on a data-driven approach, which is particularly effective in addressing under-determined scenarios. In under-determined problems, the number of unknowns exceeds the number of equations, making the system's solution unstable and sensitive to noise. However, our approach can stabilize these solutions by learning the underlying data distribution, thus providing a robust framework for estimation.

**Algorithm 1:** Device activity detection and channel estimation via score-based generative models in massive MIMO systems

**Input:** Pilot matrix $P$, received pilots $Y$, pretrained score-based model $s_\theta$, received noise power $\sigma^2_{\text{pilot}}$, inference noise levels $\sigma^2_{U_i}$, hyperparameters $L, Q, \alpha_0, \beta$, and $r < 1$.
**Generate random initial estimate:** $X_{\text{est},0} \sim \mathcal{CN}(0, I)$
**For** $i = 1, 2, \cdots, L$
  Set annealed noise level $\sigma \leftarrow \sigma_{U_i}$.
  **For** $q = 1, 2, \cdots, Q$
    Generate annealing noise $\zeta \sim \mathcal{CN}(0, I)$.
    $$X_{\text{est},q} \leftarrow X_{\text{est},q-1} + \alpha_0 \cdot r^i \cdot \frac{(X_{\text{est}}P - Y)P^H}{\sigma^2_{\text{pilot}} + \sigma^2} + \alpha_0 \cdot r^i \cdot s_\theta(X_{\text{est}}) + \sqrt{2\beta \cdot \alpha_0 \cdot r^i} \cdot \sigma \cdot \zeta$$
    Count the number of zero rows in $X_{\text{est}}$ to find $N - K$.
**Output:** Estimate channel matrix $X_{\text{est}}$, and then get the NMSE and activity detection accuracy.

The objective of our model is to estimate the CSI using the received pilot matrix $\boldsymbol{Y}$ and the known pilot matrix $\boldsymbol{P}$, and to determine the number of inactive users indicated by $\lambda_n = 0$ in the channel matrix $\boldsymbol{X}$. The process is divided into two main phases: training and inference.

1) Training phase: This initial step involves training the score-based generative model by minimizing the loss function as detailed in Section 2.3.2. To compute the score function $s_\theta$, we train a deep neural network, as depicted in Fig. 3. The neural network is trained to learn the score function that approximates the gradient of the log-likelihood of the data distribution. Moreover, it is fully convolutional, enabling it to process matrices of varying sizes, which is crucial for the dynamic nature of massive MIMO systems. This is a one-time setup process for the wireless device, typically conducted offline using a high-performance computing server and a dataset comprising either precise channel measurements or simulated channel data. The loss function quantifies the discrepancy between the model's predictions and the actual data, guiding the model to learn the data distribution effectively.

Particularly, our approach employs a Monte-Carlo simulation to generate synthetic massive MIMO channel data. User positioning is modeled to simulate random distribution within a defined area, reflecting real-world spatial randomness. Path loss is calculated using $\text{Path Loss/dB} = 128.1 + 37.6 \log_{10} d$ [42], which is a standard model for signal attenuation in wireless communication. Firstly, it allows for the modeling of complex channel behaviors by simulating a large number of random variables, which is essential for accurately representing the multipath fading effects in wireless communication channels[43]. Secondly, this approach facilitates the assessment of system performance under various conditions, providing a robust framework for optimizing and understanding the behavior of massive MIMO systems[36]. Then, the neural network begins with 2D downsampling and convolutional layers designed to extract meaningful features from these input datasets. To enhance the model's ability to learn complex patterns, Rectified Linear Unit (ReLU) activation functions are adopted to introduce nonlinearity. The model then employs 2D upsampling with additional convolutional layers to reconstruct the data to its original dimensions. In the closing act of the methodology, a 2D average pooling layer serves to compress feature maps, enhance noise resilience, and streamline subsequent layers by reducing dimensionality and focusing on dominant features.

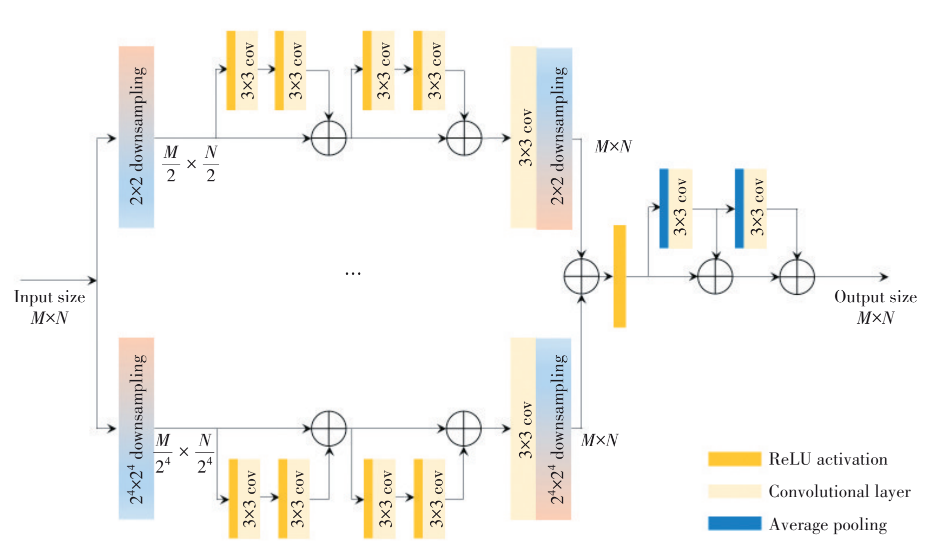
2) Inference (testing) phase: In this phase, channel estimation is treated as an optimization problem and solved using the iterative algorithm presented in Sections 2.3.3 and 2.3.4. The pre-trained model, combined with the received pilots, is utilized to recover the CSI. This phase is designed to operate independently of the training stage, enabling adaptability to various real-world conditions, including interference and quantization effects on the received pilots.

The complexity of each step is analyzed as follows.

1) Initialization: The initialization step involves generating a random initial estimate $\boldsymbol{X}_{\text{est},0} \sim \mathcal{CN}(0, \boldsymbol{I})$, which has a complexity of $O(MN)$, where $M$ and $N$ are the dimensions of the channel matrix.

2) Outer loop $i = 1, 2, \cdots, L$: Setting annealed noise level $\sigma \leftarrow \sigma_{U_i}$ involves negligible computational complexity.

3) Inner loop $q = 1, 2, \cdots, Q$: Generating annealing noise $\zeta \sim \mathcal{CN}(0, \boldsymbol{I})$ has a complexity of $O(MN)$. The update step for $\boldsymbol{X}_{\text{est},q}$ involves several matrix operations:




ReLU: Rectified Linear Unit

**Figure 3. An elaborate schematic representation of model $s_\theta$ utilizing the RefineNet architecture. This fundamental block is cascaded $D$ times in sequence**

- $\boldsymbol{X}_{\text{est},q} \leftarrow \boldsymbol{X}_{\text{est},q-1} + \alpha_0 \cdot r^i \cdot \dfrac{(\boldsymbol{X}_{\text{est}}\boldsymbol{P} - \boldsymbol{Y})\boldsymbol{P}^H}{\sigma^2_{\text{pilot}} + \sigma^2}$: This step involves matrix multiplication and division, with a complexity of $O(MNP)$, where $P$ is the number of pilots.
- $+\alpha_0 \cdot r^i \cdot s_\theta(\boldsymbol{X}_{\text{est}})$: The complexity of this step depends on the model $s_\theta$, assumed to be $O(f(MN))$, where $f$ is a function of the model complexity.
- $+\sqrt{2\beta \cdot \alpha_0 \cdot r^i} \cdot \sigma \cdot \zeta$: This step has negligible complexity.

4) Counting zero rows: Counting the number of zero rows in $\boldsymbol{X}_{\text{est}}$ to find $N - K$ has a complexity of $O(MN)$.

Thus, the total time complexity is dominated by the inner loop operations, particularly the matrix multiplications and the model $s_\theta$ evaluation. Therefore, the total complexity is approximately $O(LQ \cdot MNP + LQ \cdot f(MN))$. The space complexity is primarily determined by the storage requirements for $\boldsymbol{X}_{\text{est}}$ and other intermediate variables, which is $O(MN)$.

## 4 Experiments

We conduct simulations with different configurations of $M$ (the number of antennas), $N$ (the total number of users), and $K$ (the number of active users) to generate input channel matrices that vary in size and complexity via Monte-Carlo methodology. This approach allows us to assess the efficacy of our model under diverse conditions, verifying its robustness and capability to accurately mirror the dynamics of realistic scenarios. Additionally, we perform a comparative analysis of our proposed score-based generative model against the traditional linear minimum mean square error (LMMSE) method for channel estimation. This comparison is conducted across various SNR levels to evaluate their performance in terms of NMSE and activity detection accuracy. All experiments are conducted using PyTorch on an NVIDIA RTX 3090 GPU.

The channel matrix is initialized by assigning random positions to users and computing the path loss as a function of their distance from the BS. Subsequently, it constructs the channel coefficients using complex Gaussian random variables to simulate the multipath fading effects and compiles these into a data matrix for each simulation iteration. After generating datasets, we assess the performance of Algorithm 1 through simulations, spanning various SNR levels. Our evaluation criteria include the accuracy of channel estimation, the error rates throughout a simulated communication system that employs coding, and the computational overhead associated with both training and inference phases. To mimic real-world deployment scenarios, we examine situations where the algorithm is challenged with data distributions that differ from those encountered during training. This assessment is conducted without any foreknowledge of the test environment's characteristics, without modifying the model to adapt to the new distribution, and without conducting any additional training specifically for the test conditions.

For fine-tuning the hyperparameters in our channel estimation methods, we utilize a subset of 500 channel realizations sampled from the training distribution. In the testing phase, we create a fresh dataset consisting of 50 channel realizations for each target distribution, ensuring that the random seed used differs from those used in the training and validation phases. For the pilot signals $\boldsymbol{P}$, we construct matrices with dimensions $N \times L_{\text{pilot}}$, filled with QPSK elements that are randomly selected to represent unit-power, two-bit phase-quantized random beamforming vectors. To standardize the channel measurements, we apply normalization using the mean channel power calculated from the training dataset, which is derived from all training samples and their respective entries. The average SNR is then determined using the formula $N/\sigma^2_{\text{pilot}}$, where $N$ is the number of transmit antennas and $\sigma^2_{\text{pilot}}$ is the variance of the pilot signals.

Our proposed model demonstrates rapid convergence, as indicated by the swift reduction in training loss during the initial steps (Fig. 4), stabilizing at a low value by the completion of training.

In our comparative analysis of channel estimation techniques, the proposed score-based generative model outperforms the LMMSE method (Fig. 5). The traditional LMMSE method[44] exhibits higher noise levels and is notably less accurate in estimating user activity rates, especially in poor channel conditions. Only under sufficiently good channel conditions can the traditional method approach the performance of our generative learning approach.

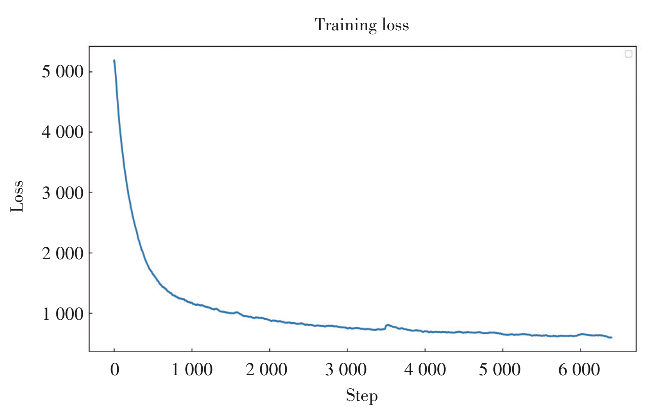
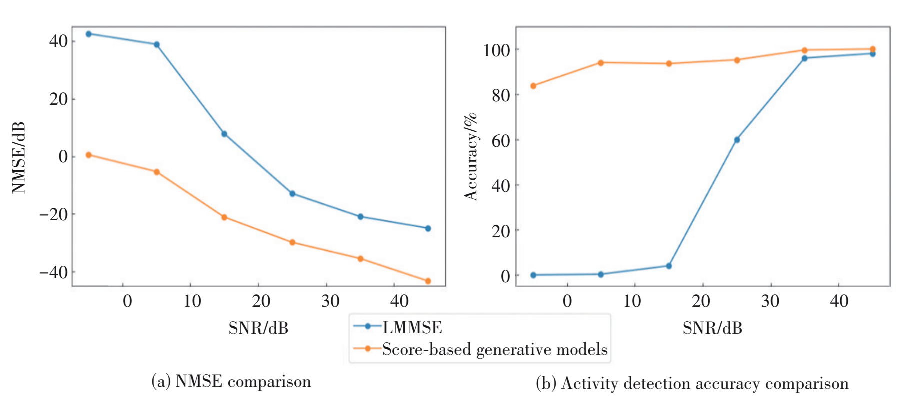



**Figure 4. Training loss of the score-based generative model over steps**




(a) NMSE comparison (b) Activity detection accuracy comparison

LMMSE: linear minimum mean square error NMSE: normalized mean square error SNR: signal-to-noise ratio

**Figure 5. Performance comparison of channel estimation methods**

We also input different datasets to evaluate the algorithm's performance. Three sets of comparative experiments were conducted to observe the impact of varing $M$, $N$, and $K$ on the algorithm's effectiveness (Fig. 6). The overall experimental results indicate the following:

1) Performance improvement: Our proposed method demonstrates a substantial enhancement in performance compared to traditional methods. This is particularly evident when evaluating the impact of varying $M$, $N$, and $K$ on the algorithm's effectiveness. Our method consistently shows lower NMSE and higher accuracy across different configurations, indicating a superior capability in channel estimation and device activity detection.

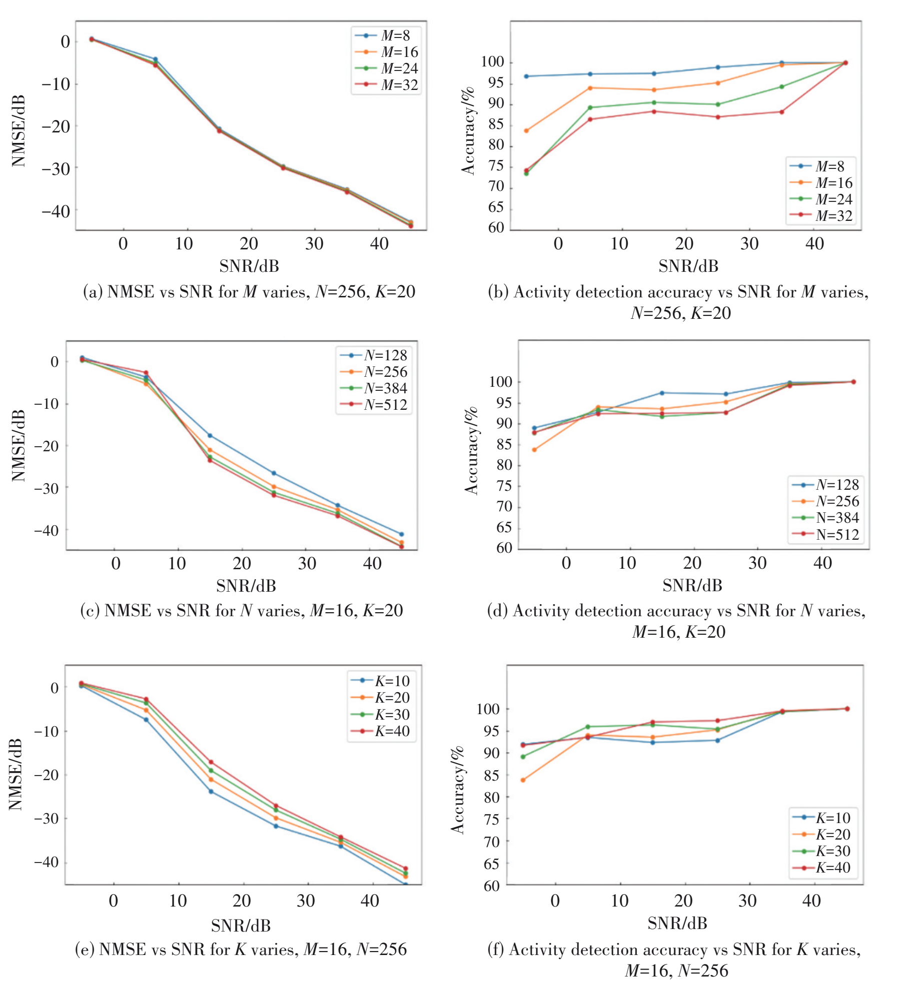
2) Robustness: The proposed score-based generative model demonstrates robustness under varying channel conditions, maintaining low estimation errors despite changes in channel conditions.

The numerical comparative analysis is as follows:

1) When only $M$ (number of receiving antennas) varies, the maximum absolute difference in NMSE across all SNR levels is only 1.15 dB (when SNR=10 dB), indicating a minimal impact on channel estimation. However, due to the increase in matrix size, the sensitivity to data increases, and the accuracy of active detection is poor under very poor channel conditions. The smaller the $M$, the faster the accuracy approaches 100% (e.g., $M$=8, 16). Nevertheless, a perfect accuracy rate of 100% can be achieved when SNR=45 dB.

2) When only $N$ (total number of devices) varies, the NMSE curve indicates a slight overall improvement in model performance. This suggests that our model is particularly suitable for massive MIMO scenarios. Meanwhile, the overall active detection accuracy tends towards 100% more rapidly as the SNR increases.

3) Changes in $K$ (number of active users) do not affect the shape of the channel. The active detection accuracy remains high even under the worst channel conditions. Moreover, the CSI estimation becomes closer to the ground truth as the number of active users decreases. This is applicable to IoT scenarios, where devices are typically designed to remain inactive most of the time to conserve energy, with only a few devices active transmitting data at any given interval. This indicates that using our model to assess device activity rates and perform more accurate channel estimation could optimize device activity patterns in the future, further reducing energy consumption and improving energy efficiency.




(a) NMSE vs SNR for $M$ varies, $N$=256, $K$=20

(b) Activity detection accuracy vs SNR for $M$ varies, $N$=256, $K$=20

(c) NMSE vs SNR for $N$ varies, $M$=16, $K$=20

(d) Activity detection accuracy vs SNR for $N$ varies, $M$=16, $K$=20

(e) NMSE vs SNR for $K$ varies, $M$=16, $N$=256

(f) Activity detection accuracy vs SNR for $K$ varies, $M$=16, $N$=256

NMSE: normalized mean square error SNR: signal-to-noise ratio

**Figure 6. Diagrams for simulaion results**

## 5 Conclusions

In this paper, we propose a novel method for joint device activity detection and channel estimation in massive MIMO networks, enabling accurate channel estimation to enhance energy efficiency and communication performance.

We employ score-based generative models, an innovative generative approach that integrates deep neural networks without making any assumptions about the received pilot matrix, the transmitted pilot matrix, and the pilot density. During our simulation experiments, we generated a comprehensive dataset using Monte-Carlo sampling. Since the deep neural network framework used to learn the scoring function is fully convolutional, the model can flexibly adapt to inputs of various sizes. We conducted a series of comparative experiments under varying conditions, including varying numbers of antennas, total users, and active users. The results demonstrate that as channel conditions improve, channel estimation is highly precise, with errors reduced to as low as −45 dB, and the detection of active devices is exceptionally accurate. As the number of users increases, the NMSE decreases, indicating that our approach is highly suitable for massive MIMO scenarios. Moreover, a smaller number of active users indicates a sparser channel matrix, yet changes in activity have a minimal impact on our model's performance, confirming that our method is entirely data-driven.

## 6 Future Work

The proposed score-based generative model for joint device activity detection and channel estimation demonstrates significant potential for application in next-generation wireless systems. Future work will explore the adaptation of this method to mmWave channels, which have unique characteristics, such as higher frequency bands and more severe path loss. Additionally, we plan to investigate the integration of this approach into 5G and 6G deployments, where massive connectivity and high spectral efficiency are critical requirements.

**References**

[1] CHEN X M, NG D W K, YU W, et al. Massive access for 5G and beyond [J]. IEEE journal on selected areas in communications, 2021, 39(3): 615-637. DOI: 10.1109/JSAC.2020.3019724
[2] MIUCCIO L, PANNO D, RIOLO S. An energy-efficient DL-aided massive multiple access scheme for IoT scenarios in beyond 5G networks [J]. IEEE Internet of Things journal, 2023, 10(9): 7936 – 7959. DOI: 10.1109/JIOT.2022.3231128
[3] VISWANATHAN H, MOGENSEN P E. Communications in the 6G era [J]. IEEE access, 2020, 8: 57063 – 57074. DOI: 10.1109/ACCESS.2020.2981745
[4] WEI F, CHEN W, WU Y P, et al. Toward 5G wireless interface technology: enabling nonorthogonal multiple access in the sparse code domain [J]. IEEE vehicular technology magazine, 2018, 13(4): 18 – 27. DOI: 10.1109/MVT.2018.2867280
[5] LIU L, YU W. Massive connectivity with massive MIMO: part I: device activity detection and channel estimation [J]. IEEE transactions on signal processing, 2018, 66(11): 2933 – 2946. DOI: 10.1109/TSP.2018.2818082
[6] DING J, QU D M, FENG M J, et al. Dynamic preamble-resource partitioning for critical MTC in massive MIMO systems [J]. IEEE Internet of Things journal, 2021, 8(20): 15361 – 15371. DOI: 10.1109/JIOT.2021.3064061
[7] RUSEK F, PERSSON D, LAU B K, et al. Scaling up MIMO: opportunities and challenges with very large arrays [J]. IEEE signal processing magazine, 2013, 30(1): 40 – 60. DOI: 10.1109/MSP.2011.2178495
[8] SHEN J C, ZHANG J, CHEN K C, et al. High-dimensional CSI acquisition in massive MIMO: Sparsity-inspired approaches [J]. IEEE systems journal, 2017, 11(1): 32 – 40. DOI: 10.1109/JSYST.2015.2448661
[9] LIU L, LARSSON E G, YU W, et al. Sparse signal processing for grant-free massive connectivity: a future paradigm for random access protocols in the Internet of Things [J]. IEEE signal processing magazine, 2018, 35 (5): 88 – 99. DOI: 10.1109/MSP.2018.2844952
[10] DAI J S, LIU A, LAU V K N. FDD massive MIMO channel estimation with arbitrary 2D-array geometry [J]. IEEE transactions on signal processing, 66(10): 2584 – 2599. DOI: 10.1109/tsp.2018.2807390
[11] RAJORIYA A, RUKHSANA S, BUDHIRAJA R. Centralized and decentralized active user detection and channel estimation in mMTC [J]. IEEE transactions on communications, 70(3): 1759 – 1776. DOI: 10.1109/tcomm.2022.3141401
[12] JALAL A, ARVINTE M, DARAS G, et al. Robust compressed sensing MRI with deep generative priors [C]//Proc. 35th International Conference on Neural Information Processing Systems. ACM, 2021: 14938 – 14954. DOI: 10.5555/3540261.3541406
[13] SONG Y, SHEN L Y, XING L, et al. Solving inverse problems in medical imaging with score-based generative models [EB/OL]. (2021-11-15) [2024-08-30]. https://arxiv.org/abs/2111.08005v2
[14] MENG X M, KABASHIMA Y. Diffusion model based posterior sampling for noisy linear inverse problems [EB/OL]. (2022-11-20) [2024-12-11]. https://arxiv.org/abs/2211.12343v4
[15] CANDÈS E J, ROMBERG J K, TAO T. Stable signal recovery from incomplete and inaccurate measurements [J]. Communications on pure and applied mathematics, 2006, 59(8): 1207 – 1223. DOI: 10.1002/cpa.20124
[16] WU Y, ROSCA M, LILLICRAP T. Deep compressed sensing [C]//36th International Conference on Machine Learning. PMLR, 2019: 6850 – 6860
[17] AALI A, ARVINTE M, KUMAR S, et al. Solving inverse problems with score-based generative priors learned from noisy data [C]//Proc. 57th Asilomar Conference on Signals, Systems, and Computers. IEEE, 2023: 837 – 843. DOI: 10.1109/IEEECONF59524.2023.10477042
[18] ALKHATEEB A, LEUS G, HEATH R W. Compressed sensing based multi-user millimeter wave systems: how many measurements are needed? [C]//Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2015: 2909 – 2913.. DOI: 10.1109/icassp.2015.7178503
[19] REZENDE D J, MOHAMED S, REZENDE D J, et al. Variational inference with normalizing flows [C]//Proc. 32nd International Conference on International Conference on Machine Learning - Volume 37. ACM, 2015: 1530 – 1538. DOI: 10.5555/3045118.3045281
[20] SONG Y, ERMON S. Generative modeling by estimating gradients of the data distribution [J]. 2019: 11918 – 11930. DOI: 10.5555/3454287.3455354
[21] SONG Y, ERMON S, SONG Y, et al. Improved techniques for training score-based generative models [C]//Proc. 34th International Conference on Neural Information Processing Systems. ACM, 2020: 12438 – 12448. DOI: 10.5555/3495724.3496767
[22] MENG X M, KABASHIMA Y. Quantized compressed sensing with score-based generative models [EB/OL]. (2022-11-02) [2024-12-11]. https://arxiv.org/abs/2211.13006v4
[23] BORA A, JALAL A, PRICE E, et al. Compressed sensing using generative models [C]//Proc. 34th International Conference on Machine Learning. ACM, 2017: 537 – 546. DOI: 10.5555/3305381.3305437
[24] HASAN M, HOSSAIN E, NIYATO D. Random access for machine-to-machine communication in LTE-advanced networks: issues and approaches [J]. IEEE communications magazine, 2013, 51(6): 86 – 93. DOI: 10.1109/MCOM.2013.6525600
[25] PRATAS N K, THOMSEN H, STEFANOVIĆ Č, et al. Code-expanded random access for machine-type communications [C]//Proc. IEEE Globecom Workshops. IEEE, 2012: 1681 – 1686. DOI: 10.1109/GLOCOMW.2012.6477838
[26] CHEN Z L, YU W. Massive device activity detection by approximate message passing [C]//Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2017: 3514 – 3518. DOI: 10.1109/ICASSP.2017.7952810
[27] DONOHO D L, MALEKI A, MONTANARI A. Message-passing algorithms for compressed sensing [J]. Proceedings of the national academy of sciences of the United States of America, 2009, 106(45): 18914 – 18919. DOI: 10.1073/pnas.0909892106
[28] BAYATI M, MONTANARI A. The dynamics of message passing on dense graphs, with applications to compressed sensing [J]. IEEE transac-

tions on information theory, 2011, 57(2): 764 – 785. DOI: 10.1109/TIT.2010.2094817

[29] HO J, JAIN A, ABBEEL P. Denoising diffusion probabilistic models [C]//34th International Conference on Neural Information Processing Systems. NIPS, 2020: 6840 – 6851

[30] BATZOLIS G, STANCZUK J, SCHÖNLIEB C B, et al. Conditional image generation with score-based diffusion models [EB/OL]. (2021-11-26) [2024-12-11]. https://arxiv.org/abs/2111.13606v1

[31] SOLTANI M, POURAHMADI V, MIRZAEI A, et al. Deep learning-based channel estimation [J]. IEEE communications letters, 2019, 23(4): 652 – 655. DOI: 10.1109/LCOMM.2019.2898944

[32] HE H T, WEN C K, JIN S, et al. Deep learning-based channel estimation for beamspace mmWave massive MIMO systems [J]. IEEE wireless communications letters, 2018, 7(5): 852 – 855. DOI: 10.1109/LWC.2018.2832128

[33] METZLER C A, MOUSAVI A, BARANIUK R G, et al. Learned D-amp [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. ACM, 2017: 1770 – 1781. DOI: 10.5555/3294771.3294940

[34] ARVINTE M, TAMIR J I. MIMO channel estimation using score-based generative models [J]. IEEE transactions on wireless communications, 2023, 22(6): 3698 – 3713. DOI: 10.1109/TWC.2022.3220784

[35] LARSSON E G, EDFORS O, TUFVESSON F, et al. Massive MIMO for next generation wireless systems [J]. IEEE communications magazine, 52 (2): 186 – 195. DOI: 10.1109/mcom.2014.6736761

[36] BORGES D, MONTEZUMA P, DINIS R, et al. Massive MIMO techniques for 5G and beyond: opportunities and challenges [J]. Electronics, 2021, 10(14): 1667. DOI: 10.3390/electronics10141667

[37] TARANTOLA A. Inverse problem theory and methods for model parameter estimation [M]. Philadelphia, USA: Society for Industrial and Applied Mathematics, 2005. DOI: 10.1137/1.9780898717921

[38] BOTTARELLI M, EPIPHANIOU G, BEN ISMAIL D K, et al. Physical characteristics of wireless communication channels for secret key establishment: a survey of the research [J]. Computers & security, 2018, 78: 454 – 476. DOI: 10.1016/j.cose.2018.08.001

[39] VINCENT P. A connection between score matching and denoising autoencoders [J]. Neural computation, 2011, 23(7): 1661 – 1674. DOI: 10.1162/neco_a_00142

[40] SONG Y, DURKAN C, MURRAY I, et al. Maximum likelihood training of score-based diffusion models [EB/OL]. (2021-01-22) [2024-08-30]. https://arxiv.org/abs/2101.09258v4

[41] ROBERTS G O, TWEEDIE R L. Exponential convergence of Langevin distributions and their discrete approximations [J]. Bernoulli, 1996, 2(4): 341. DOI: 10.2307/3318418

[42] DANUFANE F H, DI RENZO M, DE ROSNY J, et al. On the path-loss of reconfigurable intelligent surfaces: an approach based on Green's theorem applied to vector fields [J]. IEEE transactions on communications, 2021, 69(8): 5573 – 5592. DOI: 10.1109/TCOMM.2021.3081452

[43] DATTA T, KUMAR N A, CHOCKALINGAM A, et al. A novel Monte-Carlo-sampling-based receiver for large-scale uplink multiuser MIMO systems [J]. IEEE transactions on vehicular technology, 2013, 62(7): 3019 – 3038. DOI: 10.1109/TVT.2013.2260572

[44] AIT AOUDIA F, HOYDIS J. End-to-end learning for OFDM: from neural receivers to pilotless communication [J]. IEEE transactions on wireless communications, 2022, 21(2): 1049 – 1063. DOI: 10.1109/TWC.2021.3101364

## Biographies

**TANG Chenyue** (chenyue.23@intl.zju.edu.cn) received her BE degree from Central South University, China in 2023. She is currently working toward her master's degree in electrical engineering and information technology at Zhejiang University, China. Her current research interests include statistical signal processing, inverse problems, and machine learning.

**LI Zeshen** received his BE degree from Jilin University, China in 2022. He is working toward his PhD degree in information and communication engineering at Zhejiang University, China. His current research interests include federated learning, edge computing, and distributed machine learning.

**CHEN Zihan** received his BE degree in communication engineering from the Yingcai Honors College, University of Electronic Science and Technology of China (UESTC) in 2018 and PhD degree from the Singapore University of Technology and Design (SUTD)-National University of Singapore (NUS) Joint PhD Program in 2022. Currently, he is a postdoctoral research fellow with SUTD. His research mainly focuses on network intelligence, machine learning, and edge computing.

**Howard H. YANG** received his BE degree in communication engineering from Harbin Institute of Technology, China in 2012, and MSc degree in electronic engineering from the Hong Kong University of Science and Technology, China in 2013, and PhD degree in electrical engineering from the Singapore University of Technology and Design, Singapore in 2017. He is currently an assistant professor with Zhejiang University/University of Illinois at Urbana-Champaign Institute, Zhejiang University, China. His research interests currently focus on the modeling of modern wireless networks, high dimensional statistics, graph signal processing, and machine learning.