

Article

A Model-Driven Channel Estimation Method for Millimeter-Wave Massive MIMO Systems

Qingli Liu *, Yangyang Li and Jiaxu Sun

Communication and Network Laboratory, Dalian University, Dalian 116622, China

* Correspondence: liuqingli@dlu.edu.cn

Abstract: Aiming at the problem of low estimation accuracy under a low signal-to-noise ratio due to the failure to consider the “beam squint” effect in millimeter-wave broadband systems, this paper proposes a model-driven channel estimation method for millimeter-wave massive MIMO broadband systems. This method considers the “beam squint” effect and applies the iterative shrinkage threshold algorithm to the deep iterative network. First, the millimeter-wave channel matrix is transformed into a transform domain with sparse features through training data learning to obtain a sparse matrix. Secondly, a contraction threshold network based on an attention mechanism is proposed in the phase of beam domain denoising. The network selects a set of optimal thresholds according to feature adaptation, which can be applied to different signal-to-noise ratios to achieve a better denoising effect. Finally, the residual network and the shrinkage threshold network are jointly optimized to accelerate the convergence speed of the network. The simulation results show that the convergence speed is increased by 10% and the channel estimation accuracy is increased by 17.28% on average under different signal-to-noise ratios.

Keywords: millimeter wave; massive MIMO; channel estimation; beam squint; model driven



Citation: Liu, Q.; Li, Y.; Sun, J. A Model-Driven Channel Estimation Method for Millimeter-Wave Massive MIMO Systems. *Sensors* **2023**, *23*, 2638. <https://doi.org/10.3390/s23052638>

Academic Editor: Davy P. Gaillot

Received: 16 December 2022

Revised: 10 February 2023

Accepted: 14 February 2023

Published: 27 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

At present, the fifth-generation (5G) wireless communication network technology has entered a critical period of large-scale application, and the ultimate goal is to realize the mobile Internet and the Internet of Everything [1]. However, the 5G communication system cannot meet all the needs of future wireless communication systems, especially the explosive growth of mobile data traffic and the need for business diversification. Many universities and some foreign scholars have started the research and development of the sixth-generation (6G) wireless communication network technology [2]. To meet the application requirements of the ultra-high transmission rate (huge traffic) and ultra-high connection density (giant connection), the spectrum needs all spectrums below 6 GHz, i.e., millimeter wave, terahertz, and optical band. While the spectrum resources are abundant, its path loss is also high. To overcome the huge attenuation in the millimeter-wave transmission path, the millimeter-wave technology is usually used in combination with massive multiple-input multiple-output (MIMO) technology. Higher energy efficiency and spectrum efficiency can be provided by using large-scale antennas [3]. In recent years, a new digital modulation technique, spatial modulation, has also been extensively studied, which strikes a good balance between spectral efficiency and energy efficiency [4,5]. At the same time, the channel estimation of this single input stream multiple input multiple output technology (SM-MIMO) has also been studied [6]. Therefore, accurate channel state information is indispensable for any communication system, especially for millimeter-wave MIMO-OFDM broadband systems with “beam squint” effect; accurate CSI is the main factor affecting the performance of the system, so accurate channel estimation becomes a key problem to be solved.

Aiming at the problem of channel estimation in the millimeter-wave massive MIMO system, relevant scholars have used the sparse characteristic of millimeter-wave propagation [7] to adopt the restoration method of compressed sensing, implementing algorithms such as Orthogonal Matching Pursuit (OMP) [8–10], Basis Pursuit De-Noising (BPDN) [11], and some other similar algorithms [12]. However, the performance of the compressed sensing algorithm is greatly affected by the prior sparsity and the iteration step size, and the anti-noise ability of the algorithm is poor in the environment of low signal-to-noise ratio, so it is not suitable for the channel estimation of the millimeter-wave massive MIMO broadband system. The authors of [13] proposed a Support Detection (SD) method by using the structural characteristics of the beam space channel of a millimeter wave. This method decomposes the channel estimation problem into a series of sub-problems, reducing pilot overhead and ensuring reliable performance. In [14], a Cospase Analysis Approximate Message Passing (SCAMPI) algorithm based on isomorphism analysis was proposed. This method treats the channel matrix as a two-dimensional image, transforms the channel estimation problem into an image reconstruction problem, and improves channel estimation accuracy. However, the methods proposed in the above documents are all applicable to narrowband systems, but in practice, to achieve higher data rates, millimeter-wave systems are more likely to be broadband. For broadband systems, [15] proposed a method called Synchronous Orthogonal Matching Pursuit (SOMP), which first estimates the support of the wideband beam spatial channel for some subcarriers, then creates public support, and finally restores the entire beam space channel. Some other similar algorithms [16,17] also assume common support, ignoring the “beam squint” effect [18] in wideband millimeter-wave systems; i.e., larger bandwidths cause subcarrier directions to change less at different frequencies. Large bandwidths in turn lead to differences in spatial channels, so the assumed common support is wrong. The authors of [19] proved that each path component in a broadband system has a unique frequency-dependent sparse structure, and proposed a Continuous Support Detection (SSD) method based on applying continuous interference cancellation. This method estimates all sparse path components one by one, but because only the least squares algorithm is used to recover the channel vector, the estimation accuracy in the low SNR area is low. Due to the huge path loss and limited transmit power, the uplink Signal–Noise Ratio (SNR) is usually low [20]. In [21], the authors proposed a model-driven unsupervised generalized expectation consensus (Learned Denoising-based GEC, LDGEC) signal recovery network based on learning denoising for broadband beam space channel estimation, which combines the denoising convolutional neural network DnCNN into the generalized expectation consistent signal recovery algorithm, although it has achieved better performance than the algorithm based on compressed sensing, it can only obtain -7 dB NMSE performance when $\text{SNR} = 0$, so there is still room for improvement in the estimation accuracy under low SNR.

In view of the above method, the channel estimation accuracy is low under a low signal-to-noise ratio for wideband systems with “beam squint” effect, and this paper proposes a new channel estimation method based on Adaptive Threshold Learning Iterative Shrinkage Threshold Network (ALISTA-Net). The method uses model-driven deep learning, based on the iterative shrinkage threshold algorithm, and the algorithm is expanded into the deep iterative network, and the millimeter-wave channel matrix is transformed into the sparse transform domain through training data learning. In the beam domain denoising stage, the Shrinkage Threshold Network (ST-Net) based on the attention mechanism is proposed. The network learns from the idea of the attention mechanism to provide a set of thresholds adaptively through feature learning so that the model is suitable for situations with different noise contents. The denoising effect under different signal-to-noise ratios is used to improve the accuracy of channel estimation; at the same time, to improve the network convergence speed, residual learning is introduced into the network structure. Simulation experiments show that compared with the existing channel estimation algorithms for broadband systems, this method has achieved higher estimation accuracy under different SNRs.

The structure of this paper is divided into the following parts: Section 2 introduces the system model and the description of channel estimation. Section 3 introduces the structure of the ALISTA-Net network proposed in this paper. Section 4 analyzes and discusses the experimental results. Section 5 introduces the conclusion of this paper.

2. System Model and Problem Description

2.1. System Model

In this paper, in Time Division Duplexing (TDD) mode, a wide-band beam space massive MIMO-OFDM system based on lens array is considered, as shown in Figure 1. The signals are received by the antenna at the receiving end. These signals are processed by radio frequency to obtain baseband analog signals, and then the analog signals are converted into digital signals and then processed synchronously. After the cyclic prefix is removed, they are demodulated by FFT and then processed by baseband signals.

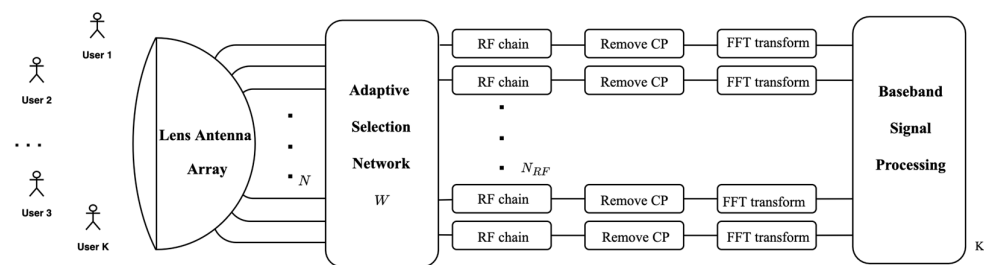


Figure 1. Broadband millimeter-wave MIMO-OFDM system structure based on lens antenna array.

The system is equipped with an N element-lens antenna array and a base station with RF chains to simultaneously serve single-antenna users. This paper considers the widely used Saleh–Valenzuela channel model, then the $N \times 1$ spatial channel of the user at the subcarrier m ($m = 1, 2, \dots, M$) can be expressed as

$$\mathbf{h}_m = \sqrt{\frac{N}{L}} \sum_{l=1}^L \beta_l e^{-j2\pi\tau_l f_m} \boldsymbol{\alpha}(\phi_{l,m}) \quad (1)$$

L is the number of paths, and β_l and τ_l are the complex gain and time delay of the l -th path, respectively. Furthermore, $\boldsymbol{\alpha}(\phi_{l,m})$ is the array response vector, and $\phi_{l,m}$ is the spatial direction at the subcarrier m , defined as

$$\phi_{l,m} = \frac{f_m}{c} d \sin \theta_l \quad (2)$$

Among them, $f_m = f_c + \frac{f_b}{M} + \left(m - 1 - \frac{M-1}{2}\right)$ is the frequency of the subcarrier m , f_c and f_b are the carrier frequency and bandwidth, respectively, c is the speed of light, θ_l is the physical direction, and $d = 0.5 \cdot c / f_c$ is the antenna spacing. In millimeter-wave broadband systems, a larger f_b will lead to $\phi_{l,m}$ a larger change in subcarriers of different frequencies, which in turn will lead to a difference in the spatial channel \mathbf{h}_m , which is called the “beam squint” effect. When the base station is equipped with a uniform linear lens antenna array, the array response vector $\boldsymbol{\alpha}(\phi_{l,m})$ can be expressed as

$$\boldsymbol{\alpha}(\phi_{l,m}) = e^{-j2\pi\phi_{l,m}\mathbf{P}_a} \quad (3)$$

$\mathbf{P}_a = \left[-\frac{N-1}{2}, -\frac{N+1}{2}, \dots, \frac{N-1}{2}\right]$ denotes the indices of different antennas. The lens antenna array shown in Figure 1 can transform the space domain in (1) into the beam space domain. The lens antenna array is similar to the $N \times N$ spatial discrete Fourier transform

(DFT) matrix \mathbf{F} . Therefore, the wideband beam spatial channel $\tilde{\mathbf{h}}_m$ at the subcarrier m is expressed as

$$\tilde{\mathbf{h}}_m = \mathbf{F}^H \mathbf{h}_m = \sqrt{\frac{N}{L}} \sum_{l=1}^L \beta_l e^{-j2\pi\tau_l f_m} \tilde{\mathbf{c}}_{l,m} \quad (4)$$

where $\tilde{\mathbf{c}}_{l,m}$ denotes the l -th path component at the subcarrier m in the beam space, and $\tilde{\mathbf{c}}_{l,m}$ is determined $\phi_{l,m}$ by

$$\tilde{\mathbf{c}}_{l,m} = \mathbf{F}^H \boldsymbol{\alpha}(\phi_{l,m}) = [\Xi(\phi_{l,m} - \bar{\phi}_1), \Xi(\phi_{l,m} - \bar{\phi}_2), \dots, \Xi(\phi_{l,m} - \bar{\phi}_N)]^T \quad (5)$$

$\Xi(x) = \frac{\sin N\pi x}{\sin \pi x}$ is the function of Dirichlet, and $\bar{\phi}_n = \frac{1}{N} \left(n - \frac{N+1}{2} \right)$ ($n = 1, 2, \dots, N$) is the predefined spatial orientation of the lens antenna array.

2.2. Problem Description

The received signal vector at the base station can be expressed as

$$\mathbf{y}_{m,q} = \mathbf{W}_q (\tilde{\mathbf{h}}_m s_{m,q} + \mathbf{n}_{m,q}) \quad (6)$$

Among them, $\mathbf{y}_{m,q} \in \mathbb{C}^{N_{RF} \times 1}$ and $s_{m,q}$ is defined as the transmission pilot of the subcarrier m at the moment q ($q = 1, 2, \dots, Q$) (each user only sends one pilot at each moment). $\mathbf{n}_{m,q} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2 \mathbf{I})$ denotes a Gaussian noise vector, σ^2 denoting the noise power. $\mathbf{W}_q \in \mathbb{C}^{N_{RF} \times N}$ is a fixed adaptive selection network for different subcarriers. Since the pilot signal is known at the receiver side, it is assumed that $s_{m,q} = 1$. After Q instants of pilot transmission, the total measurement signal can be obtained as

$$\bar{\mathbf{y}}_m = [\mathbf{y}_{m,1}^T, \dots, \mathbf{y}_{m,Q}^T]^T = \bar{\mathbf{W}} \tilde{\mathbf{h}}_m + \mathbf{n}_m^{eq} \quad (7)$$

$\bar{\mathbf{W}} = [\mathbf{W}_1^T, \mathbf{W}_2^T, \dots, \mathbf{W}_Q^T]^T \in \mathbb{C}^{QN_{RF} \times N}$ is expressed as the overall combination matrix, and $\mathbf{n}_m^{eq} = [(\mathbf{W}_1 \mathbf{n}_{m,1})^T, \dots, (\mathbf{W}_Q \mathbf{n}_{m,Q})^T]^T$ is expressed as the effective noise vector. In this paper, a low-cost 1-bit phase shifter is used in the adaptive selection network, where elements of $\bar{\mathbf{W}}$ are randomly selected in the set $\frac{1}{\sqrt{QN_{RF}}} \{-1, +1\}$. By stacking the M beam space channel vectors into a matrix and converting the complex values to real values, the following signal recovery problem description can be obtained

$$\mathbf{Y} = \bar{\mathbf{W}} \mathbf{H} + \mathbf{N} \quad (8)$$

\mathbf{Y} represents the received signal, and $\mathbf{H} = [\text{Re}(\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \dots, \tilde{\mathbf{h}}_M), \text{Im}(\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \dots, \tilde{\mathbf{h}}_M)] \in \mathbb{R}^{N \times 2M}$ is the beam frequency matrix. Since $\tilde{\mathbf{h}}_m$ is sparse, if the beam frequency matrix is regarded as a natural image, the compressed image restoration method can be used for beam space channel estimation. Therefore, a channel estimation network based on model-driven deep learning is developed.

3. Channel Estimation Network Based on Model-Driven Deep Learning

This article applies the Iterative Shrinkage Thresholding Algorithm (ISTA) [22] to the deep iterative network. The main update steps of the ISTA algorithm to restore the signal are as follows

$$\mathbf{R}^{(k)} = \hat{\mathbf{H}}^{(k-1)} - \rho \bar{\mathbf{W}}^T (\bar{\mathbf{W}} \hat{\mathbf{H}}^{(k-1)} - \mathbf{Y}) \quad (9)$$

$$\hat{\mathbf{H}}^{(k)} = \underset{\mathbf{H}}{\text{argmin}} \frac{1}{2} \|\mathbf{R}^{(k)} - \mathbf{H}\|_2^2 + \lambda \|\boldsymbol{\Psi} \mathbf{H}\|_1 \quad (10)$$

wherein k represents the iteration index of ISTA, and ρ represents the iteration step size. $\boldsymbol{\Psi}$ is the transformation matrix that makes $\boldsymbol{\Psi} \mathbf{H}$ sparse in the transform domain. The two update

steps (9) and (10) correspond to gradient descent and optimization procedures, respectively. Therefore, $\mathbf{R}^{(k)}$ can be regarded as a noisy signal. In the broadband millimeter-wave system, to reduce the recovery error caused by manually setting parameters (such as λ , ρ and Ψ), this paper adopts the idea of learnable parameters, expands the iterative steps of the ISTA algorithm to ALISTA-Net, using the powerful learning power of neural networks to obtain an optimal set of parameters, and an adaptive selection threshold method based on ST-Net network is proposed in the beam domain denoising stage to reduce the loss-of-sparse recovery. The network consists of cascaded K layers with the same structure, which strictly corresponds to the K sub-iterative processes in the ISTA algorithm. The network structure is shown in Figure 2.

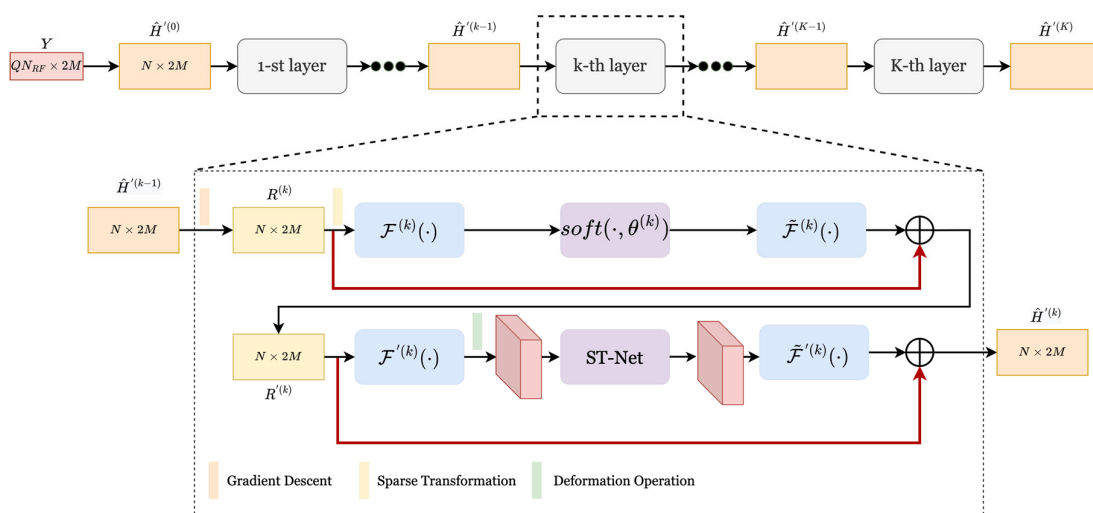


Figure 2. ALISTA-Net network structure diagram.

Taking the received signal \mathbf{Y} and $\hat{\mathbf{H}}^{(0)}$ as the input of the ALISTA-Net network, the estimation of the channel matrix $\hat{\mathbf{H}}^{(K)}$ is used as the final output of the network. Using learnable parameters, learnable sparse matrices, and introducing an attention-based shrinkage threshold network (ST-Net) to replace manually set hyperparameters (ρ , λ , Ψ and thresholds), this work aims to select the best parameters during network training to compensate for recovery loss caused by improper manual settings. In this stage, the operation to expand to ALISTA-Net mainly includes the following processes:

1. Gradient descent process: $\mathbf{R}^{(k)}$ can be rewritten as

$$\mathbf{R}^{(k)} = \hat{\mathbf{H}}^{(k-1)} - \rho^{(k)} \overline{\mathbf{W}}^\top (\overline{\mathbf{W}} \hat{\mathbf{H}}^{(k-1)} - \mathbf{Y}) \tag{11}$$

Among them, $\rho^{(k)}$ are the parameters that can be learned from the data, rather than manually setting each layer to share.

2. (Inverse) sparse transformation and residual learning process: The specific method of using sparse transformation is as follows

$$\mathcal{F}^{(k)}(\mathbf{R}^{(k)}) = \mathbf{B}^{(k)} \cdot \text{ReLU}(\mathbf{A}^{(k)} \cdot (\mathbf{R}^{(k)})^\top) \tag{12}$$

Among them, $\mathbf{A}^{(k)} \in \mathbb{R}^{\omega_1 \times 2M}$ and $\mathbf{B}^{(k)} \in \mathbb{R}^{\omega_2 \times \omega_1}$ are two learnable matrices, $\|\mathcal{F}^{(k)}(\mathbf{R}^{(k)}) - \mathcal{F}^{(k)}(\mathbf{H})\|_2^2 \approx \alpha(k) \|\mathbf{R}^{(k)} - \mathbf{H}\|_2^2$ as proved in [22], so the optimization process in Formula (10) can be rewritten as

$$\hat{\mathbf{H}}^{(k)} = \underset{\mathbf{H}}{\operatorname{argmin}} \left(\frac{1}{2} \left\| \mathcal{F}^{(k)}(\mathbf{R}^{(k)}) - \mathcal{F}^{(k)}(\mathbf{H}) \right\|_2^2 + \beta^{(k)} \left\| \mathcal{F}^{(k)}(\mathbf{H}) \right\|_1 \right) \quad (13)$$

$\beta^{(k)} = \lambda^{(k)} \times \alpha^{(k)}$, the optimization problem in Equation (13). can be viewed as a noise reduction problem, which is solved by the soft threshold operation in the frequency domain, i.e., $\operatorname{soft}(x, \theta) = \operatorname{sign}(x)(\max(0, |x| - \theta))$, and $\operatorname{sign}(\cdot)$ is a sign function. $\tilde{\mathcal{F}}^{(k)}$ represents the inverse transformation that converts the signal from the sparse transform domain to the beam domain, $\mathcal{F}^{(k)}$ and $\tilde{\mathcal{F}}^{(k)}$ the network structure of and is symmetrical, consisting of two fully connected layers, but with different weights. So, the solution of Equation (13) is as follows

$$\hat{\mathbf{H}}^{(k)} = \tilde{\mathcal{F}}^{(k)} \left(\operatorname{soft} \left(\mathcal{F}^{(k)}(\mathbf{R}^{(k)}), \theta^{(k)} \right) \right) \quad (14)$$

$\theta^{(k)}$ represents the k -th threshold of the contraction operation, which is also a learnable parameter. To speed up the convergence speed of the network, residual learning is introduced into ALISTA-Net, which is the part connected by the red line in Figure 2, so Equation (14) is rewritten as

$$\hat{\mathbf{H}}^{(k)} = \mathbf{R}^{(k)} + \tilde{\mathcal{F}}^{(k)} \left(\operatorname{soft} \left(\mathcal{F}^{(k)}(\mathbf{R}^{(k)}), \theta^{(k)} \right) \right) \quad (15)$$

Modules $\mathbf{R}'^{(k)}$ and $\mathbf{R}^{(k)}$ modules are similar, but the input is changed, as follows

$$\mathbf{R}'^{(k)} = \hat{\mathbf{H}}^{(k)} - \rho'^{(k)} \overline{\mathbf{W}}^\top \left(\overline{\mathbf{W}} \hat{\mathbf{H}}^{(k)} - \mathbf{Y} \right) \quad (16)$$

To transform and denoise in the beam domain, it is first necessary for $\mathbf{R}'^{(k)}$ to perform a transpose operation. In the denoising stage, inspired by the deep residual shrinkage network, a shrinkage threshold network (ST-Net) based on the attention mechanism is proposed, which will be introduced later. Finally, the inverse transposition operation is performed, so the output can be expressed by the following formula

$$\hat{\mathbf{H}}'^{(k)} = \mathbf{R}'^{(k)} + \mathbf{T}' \left(\tilde{\mathcal{F}}^{(k)} \left(\operatorname{ST-Net} \left(\mathcal{F}'^{(k)} \left(\mathbf{T} \left(\mathbf{R}'^{(k)} \right) \right), \theta'^{(k)} \right) \right) \right) \quad (17)$$

3. Shrink Threshold Network (ST-Net) Based on Attention Mechanism: The choice of threshold in Equation (17) has a great influence on the estimation accuracy of the ALISTA-Net model. Since the noise amount of each input is different, ST-Net is introduced in the beam domain denoising stage, inspired by the deep residual shrinkage network DRSN-CW structure in the literature [23]. The basic module is shown in Figure 3. The structure of the proposed shrinkage threshold network based on the attention mechanism is shown in Figure 4. First, the threshold in the soft threshold function is automatically learned and set by the network, which reduces the loss of accuracy caused by the inaccurate manual setting of the threshold. Secondly, the threshold value in the soft threshold function of the network is a positive number within the appropriate value range, to avoid the output situation of all zeros. At the same time, each sample has its own unique set of thresholds, making the model more applicable to situations with different noise contents.

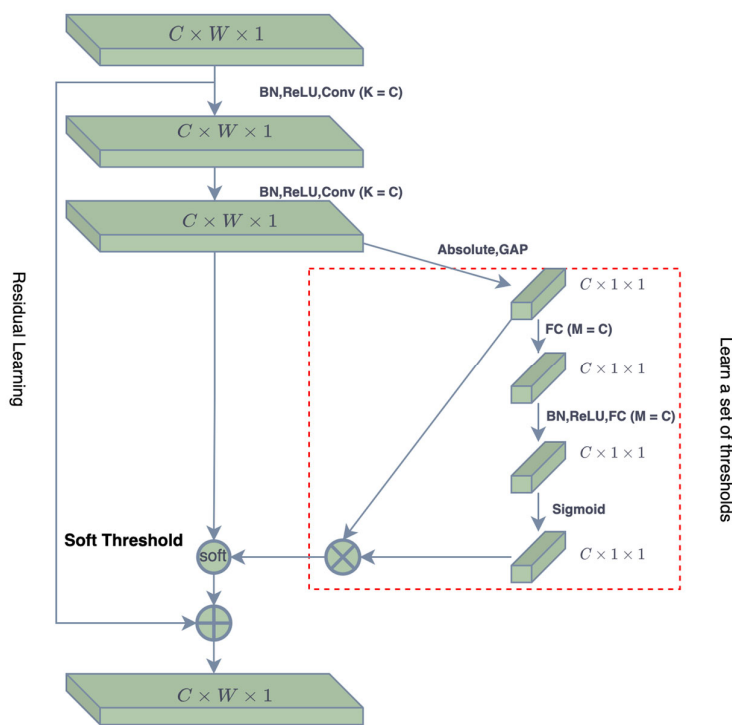


Figure 3. The basic module of deep residual shrinkage network.

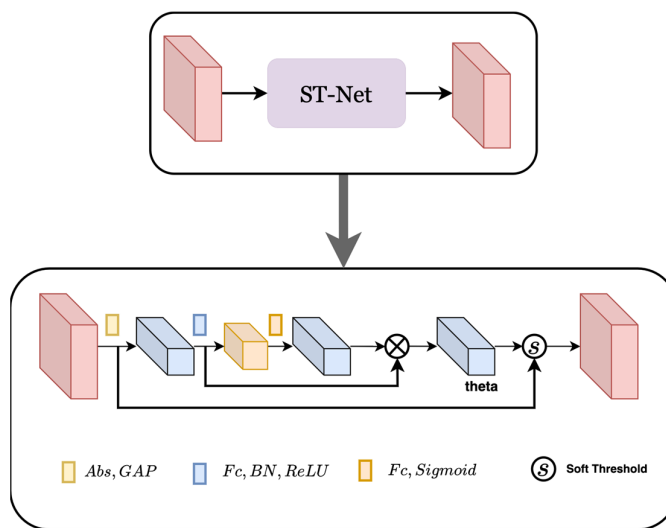


Figure 4. ST-Net network structure diagram.

The network structure draws on the idea of the attention mechanism. First, by taking the absolute value of the input, the absolute value of the sparse transformation feature map is simplified into a 16-dimensional vector by using Global Average Pooling (GAP). To make the intermediate output value of each layer the value more stable, and reduce the risk of model overfitting, the obtained features are input into two fully connected networks \mathcal{N}_1 and \mathcal{N}_2 with batch normalization (BN), and these two networks have 4 and 16 neurons, respectively, using Sigmoid and ReLU as activation functions to generate scaling vectors. The iteration threshold can be obtained by multiplying the dimensionality reduction vector and the scaling vector. The obtained threshold is a 16-dimensional vector, which changes in different iteration stages, and a set of thresholds for each channel can be adaptively obtained according to the characteristics of the input feature map. Finally, the contraction operation is performed according to the contraction function in the formula.

4. Training parameters: The trainable parameter set in ALISTA-Net is used Θ . According to the above description, Θ can be expressed as follows

$$\Theta = \left\{ \rho^{(k)}, \rho'^{(k)}, \theta^{(k)}, \mathcal{F}^{(k)}(\cdot), \tilde{\mathcal{F}}^{(k)}(\cdot), \mathcal{F}'^{(k)}(\cdot), \tilde{\mathcal{F}}'^{(k)}(\cdot), \mathcal{N}_{1,k}, \mathcal{N}_{2,k} \right\}_{k=1}^K$$

Among them, K is the total number of iterations of ALISTA-Net, and all parameters will be learned as parameters of the neural network.

5. Loss function: Since ALISTA-Net contains K stages, the loss function of the training process designed in this paper is as follows

$$\begin{aligned} \mathcal{L}(\Theta) &= \mathcal{L}_{mse} + \mu \mathcal{L}_{iteration} + \zeta \mathcal{L}_{symmetry} \\ &= \|\hat{\mathbf{H}}_K - \mathbf{H}\|_2^2 + \mu \sum_{k=1}^K \|\hat{\mathbf{H}}_K - \mathbf{H}\|_2^2 + \zeta \sum_{k=1}^K \left\| \tilde{\mathcal{F}}(\mathcal{F}(\hat{\mathbf{H}}_K) - \mathbf{H}) \right\|_2^2 \end{aligned} \quad (18)$$

\mathcal{L}_{mse} is a measure of the loss between the output and the true value, and is a key component of the loss function. $\mathcal{L}_{iteration}$ is the sum of the errors between the true value and the iterated value of ALISTA-Net in each stage. This helps to produce estimates that are close to the true value at each stage. $\mathcal{L}_{symmetry}$ is to establish $\mathcal{F} \cdot \tilde{\mathcal{F}} = I$ as much as possible.

4. Simulation Verification and Result Analysis

4.1. Simulation Data and Parameter Settings

To complete the training and testing of the ALISTA-Net network model, the specific parameter settings of the system model channel in this paper are shown in Table 1. The parameters of the training plan are set as shown in Table 2.

Table 1. Channel Parameter Setting of Millimeter-Wave Massive MIMO System.

Parameter	Value
Number of lens elements	32
Number of RF chains	8
Carrier frequency	28 GHz
Bandwidth	4 GHz
Number of subcarriers	32
Maximum delay	20ns
Physical direction of the path	$\theta_l \sim \mu(-\pi/2, \pi/2)$
Delay of the path	$\tau_l \sim \mu(0, 20ns)$
Number of resolvable paths	$L = 3$

Table 2. Training plan parameter settings.

Parameter	Value
Training set	10,000
Validation set	1280
Testing set	2560
Batch Size	64
Optimizer	Adam
Learning rate	0.0001
SNR	$[-10, 20]$
Maximum training iterations	5000

4.2. Simulation Result Analysis

In this paper, the normalized mean squared error (NMSE) is used as the difference between the channel matrix generated by the estimation estimate and the real channel matrix, which is defined as

$$NMSE = E \left\{ \frac{\|\mathbf{H}^{(K)} - \mathbf{H}\|_2^2}{\|\mathbf{H}\|_2^2} \right\} \quad (19)$$

wherein the smaller the value of NMSE, the higher the accuracy of channel estimation by the model.

- Ablation experiment

To compare the NMSE performance of the expanded iterative network (LISTA-Net) and the expanded iterative network (ALISTA-Net) using ST-Net, this paper designs an ablation simulation experiment for analysis. The simulation results are shown in Figures 5 and 6.

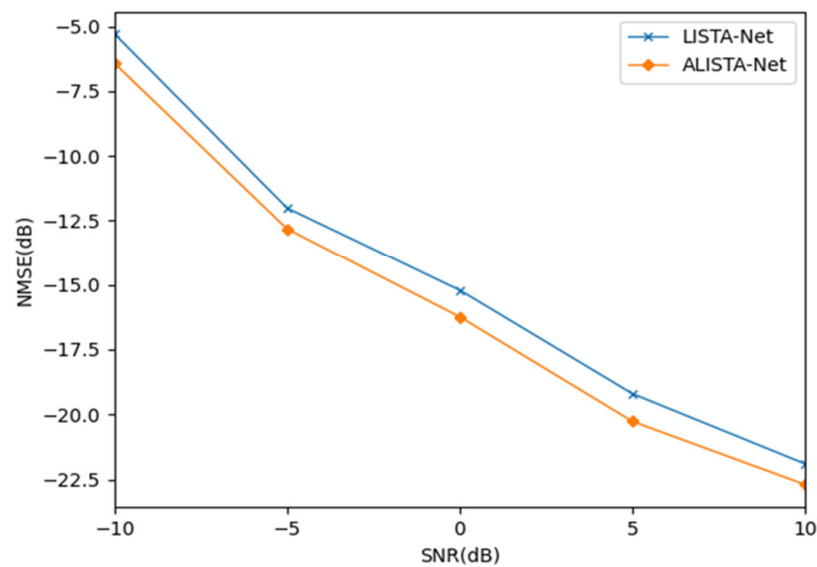


Figure 5. Performance Comparison of ALISTA Net and LISTA Net under Different SNRs.

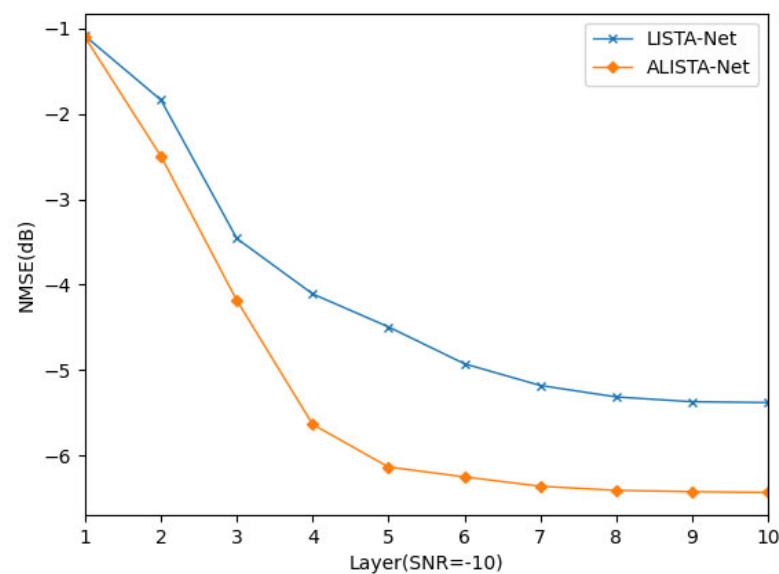


Figure 6. Convergence comparison between ALISTA Net and LISTA Net when SNR = -10.

As shown in Figure 5, the abscissa is the signal-to-noise ratio, and the ordinate is the normalized mean square error. As the SNR increases, the NMSE decreases gradually and the estimation accuracy increases. This is because as the signal-to-noise ratio increases, the amount of noise decreases gradually, and the impact on channel estimation gradually decreases, so the estimation accuracy gradually increases. Under five different signal-to-noise ratios, the NMSE of ALISTA-Net is lower than that of LISTA-Net, and the estimation accuracy is better than that of LISTA-Net. When SNR = -10 , compared with LISTA-Net, ALISTA-Net has achieved 1.06 dB accuracy improvement. This is due to the proposed ST-Net adopting a strategy of providing different thresholds for each channel according to different noise characteristics instead of sharing thresholds for all channels.

In Figure 6, the abscissa is the number of layers deployed in the network, and the ordinate is the normalized mean square error. As the number of layers increases, the NMSE gradually decreases, and the convergence speed of ALISTA-Net is fast. Specifically, the ALISTA-Net method proposed in this paper can converge within six layers.

Table 3 lists the time it takes for the two methods to complete channel estimation in the online phase. The result obtained is the online test time per 10 epochs. The running speed can reflect the practicability of the algorithm in the time environment.

Table 3. Running time.

Algorithm	Time/s
LISTA-Net	9.17
ALISTA-Net	9.25

Since ALISTA-Net introduces the ST-Net module, the running time of ALISTA-Net is slightly higher than that of LISTA-Net, but the final time consumption is acceptable in the transmission of real-time communication systems.

- Performance analysis

As shown in Figure 7, the ALISTA-Net algorithm proposed in this paper can not only obtain higher accuracy but also have a faster convergence speed than the current classic broadband system channel estimation method. Specifically, the proposed algorithm can converge within six layers; thus, the online stage can be deployed with a small number of layers, speeding up the training rate.

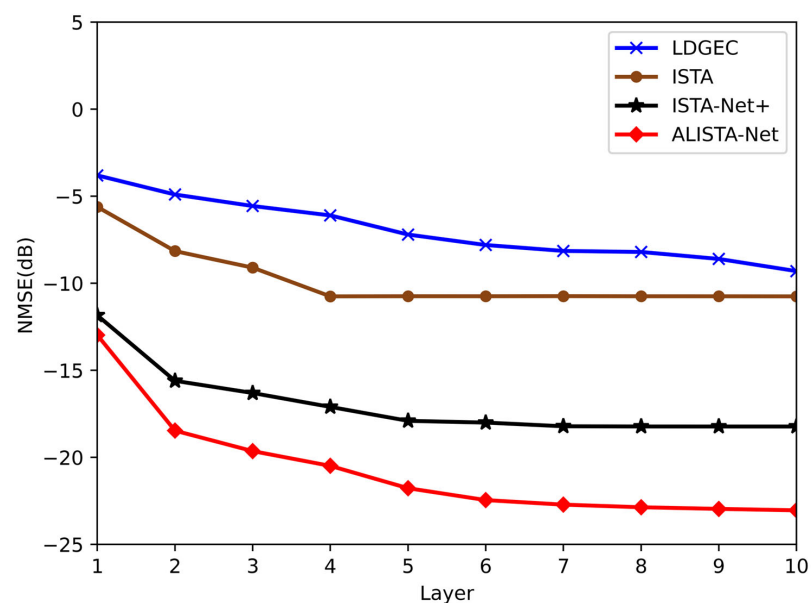


Figure 7. Convergence comparison between ALISTA Net and other channel estimation algorithms.

As shown in Figure 8, the ALISTA-Net method proposed in this paper achieves NMSE lower than that of the existing classic wideband system channel estimation algorithm in the range of 0–20 dB SNR. When SNR = 0, compared with the algorithm based on compressed sensing, ALISTA-Net obtains the largest gain and achieves NMSE of -16.24 dB, which verifies the effectiveness of the proposed scheme. Compared with deep learning-based ISTA-Net⁺ [24], a lower NMSE value is also achieved, which indicates that the proposed algorithm based on model-driven deep learning has been optimized for millimeter-wave massive MIMO broadband system channel estimation.

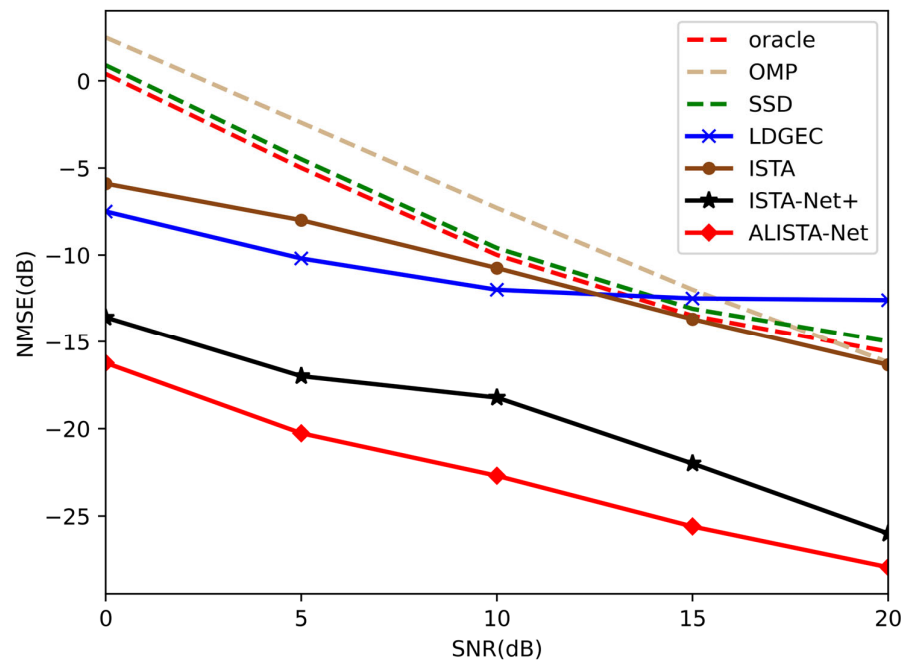


Figure 8. Comparison of NMSE performance between ALISTA Net and other channel estimation algorithms.

- Complexity analysis

As shown in Table 4, the complexity of the proposed ALISTA-Net is compared with other channel estimation methods, and ALISTA-Net achieves a lower computational complexity than LDGEC and ISTA-Net⁺. This is because the computational complexity of ALISTA-Net is mainly determined by matrix multiplication, and the computational complexity of LDGEC is mainly determined by matrix inversion. The computational complexity of ISTA-Net⁺ is mainly determined by the convolution operation, where K is the filter size, C_{in} and C_{out} are the number of input channels and output channels of the convolution, respectively (in [21], $C_{in} = C_{out} = 32$). Although the number of trainable parameters of ALISTA-Net is higher than that of LISTA-Net and ISTA-Net⁺, it is known that both ALISTA-Net achieves better estimation accuracy than LISTA-Net and ISTA-Net⁺ based on the results of ablation experiments and performance tests.

Table 4. Complexity Analysis.

Method	Parameters	Computational Complexity
LISTA-Net	1.65×10^5	$O(QN_{RF}NM)$
SSD	0	$O(MN_{RF}QL^2\Omega^2)$
OMP	0	$O(MN_{RF}QL^3\Omega^3)$
LDGEC	5.19×10^5	$O(MN^3)$
ISTA	2	$O(QN_{RF}NM)$
ISTA-Net ⁺	3.76×10^4	$O(MNK^2C_{in}C_{out})$
ALISTA-Net	1.68×10^5	$O(QN_{RF}NM)$

5. Conclusions

Aiming at the problem of low estimation accuracy of existing channel estimation algorithms in millimeter-wave massive MIMO broadband systems with a low signal-to-noise ratio, this paper proposes a channel estimation algorithm based on model-driven deep learning considering the “beam squint” effect. Ablation experiments show that, compared with LISTA-Net without ST-Net, ALISTA-Net achieves a higher estimation accuracy and convergence speed in low SNR areas. Performance experiments show that compared with existing wideband system algorithms, ALISTA-Net can still achieve higher estimation accuracy in areas with low SNR. Therefore, the ALISTA-Net proposed in this paper is suitable for channel estimation in millimeter-wave massive MIMO broadband systems and has faster convergence speed and higher channel estimation accuracy.

Author Contributions: Conceptualization, Q.L. and Y.L.; methodology, Q.L.; software, Y.L.; validation, Y.L. and J.S.; formal analysis, Y.L.; investigation, Y.L.; resources, Q.L.; data curation, Q.L. and Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L. and Q.L.; visualization, Y.L.; supervision, Q.L. and Y.L.; project administration, Q.L. and Y.L.; funding acquisition, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 61931004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The processed data required to reproduce these findings cannot be shared as the data also form part of an ongoing study.

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

References

- Bockelmann, C.; Pratas, N. Massive machine-type communications in 5G: Physical and MAC-layer solutions. *IEEE Commun. Mag.* **2016**, *54*, 59–65. [\[CrossRef\]](#)
- Zhao, Y.J.; Yu, G.H.; Xu, H.Q. 6G mobile communication networks: Vision, challenge, and key technologies. *Sci. China Inf. Sci.* **2019**, *49*, 963–987. [\[CrossRef\]](#)
- Lee, J.; Choi, K.J.; Kim, K.S. Massive MIMO full-duplex for high-efficiency next generation WLAN systems. In Proceedings of the 2016 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Republic of Korea, 19–21 October 2016; pp. 1152–1154.
- Wen, M. A Survey on Spatial Modulation in Emerging Wireless Systems: Research Progresses and Applications. *IEEE J. Sel. Areas Commun.* **2019**, *37*, 1949–1972. [\[CrossRef\]](#)
- Li, J.; Dang, S.; Huang, Y.; Chen, P.; Qi, X.; Wen, M.; Arslan, H. Composite Multiple-Mode Orthogonal Frequency Division Multiplexing with Index Modulation. *IEEE Trans. Wirel. Commun.* **2022**. [\[CrossRef\]](#)
- Wu, X.; Claussen, H.; Di Renzo, M.; Haas, H. Channel Estimation for Spatial Modulation. *IEEE Trans. Commun.* **2014**, *62*, 4362–4372. [\[CrossRef\]](#)
- Pi, Z.; Khan, F. An introduction to millimeter-wave mobile broadband systems. *IEEE Commun. Mag.* **2011**, *49*, 101–107. [\[CrossRef\]](#)
- Tropp, J.A.; Gilbert, A.C. Signal Recovery from Random Measurements Via Orthogonal Matching Pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666. [\[CrossRef\]](#)

9. Lee, J.; Gil, G.-T.; Lee, Y.H. Channel Estimation via Orthogonal Matching Pursuit for Hybrid MIMO Systems in Millimeter Wave Communications. *IEEE Trans. Commun.* **2016**, *64*, 2370–2386. [[CrossRef](#)]
10. Manoj, A.; Kannu, A.P. Multi-user millimeter wave channel estimation using generalized block OMP algorithm. In Proceedings of the 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Sapporo, Japan, 4–6 July 2017; pp. 1–5.
11. Malla, S.; Abreu, G. Channel estimation in millimeter wave MIMO Systems: Sparsity enhancement via reweighting. In Proceedings of the 2016 International Symposium on Wireless Communication Systems (ISWCS), Poznań, Poland, 20–23 September 2016; pp. 230–234.
12. Cheng, X.; Li, L.; Du, L. ADMM-based channel estimation for mmWave massive MIMO systems. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 152–157.
13. Gao, X.; Dai, L.; Han, S. Reliable BeamSpace Channel Estimation for Millimeter-Wave Massive MIMO Systems with Lens Antenna Array. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 6010–6021. [[CrossRef](#)]
14. Yang, J.; Wen, C.K.; Jin, S. BeamSpace Channel Estimation in mmWave Systems Via Cosparsity Image Reconstruction Technique. *IEEE Trans. Commun.* **2018**, *66*, 4767–4782. [[CrossRef](#)]
15. Gao, Z.; Hu, C.; Dai, L.; Wang, Z. Channel Estimation for Millimeter-Wave Massive MIMO With Hybrid Precoding Over Frequency-Selective Fading Channels. *IEEE Commun. Lett.* **2016**, *20*, 1259–1262. [[CrossRef](#)]
16. Dong, Y.; Chen, C.; Yi, N.; Lu, G.; Jin, Y. Channel Estimation Using Low-Resolution PSs for Wideband mmWave Systems. In Proceedings of the 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, NSW, Australia, 4–7 June 2017; pp. 1–5.
17. Wang, Y.; Xu, W.; Zhang, H.; You, X. Wideband mmWave Channel Estimation for Hybrid Massive MIMO With Low-Precision ADCs. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 285–288. [[CrossRef](#)]
18. Brady, J.H.; Sayeed, A.M. Wideband communication with high-dimensional arrays: New results and transceiver architectures. In Proceedings of the 2015 IEEE International Conference on Communication Workshop (ICCW), London, UK, 8–12 June 2015; pp. 1042–1047.
19. Gao, X.; Dai, L.; Zhou, S.; Sayeed, A.M.; Hanzo, L. Wideband BeamSpace Channel Estimation for Millimeter-Wave MIMO Systems Relying on Lens Antenna Arrays. *IEEE Trans. Signal Process.* **2019**, *67*, 4809–4824. [[CrossRef](#)]
20. Heath, R.W.; González-Prelcic, N.; Rangan, S.; Roh, W.; Sayeed, A.M. An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems. *IEEE J. Sel. Top. Signal Process.* **2016**, *10*, 436–453. [[CrossRef](#)]
21. He, H.; Wang, R.; Jin, S.; Wen, C.-K.; Li, G.Y. BeamSpace channel estimation in terahertz communications: A model-driven unsupervised learning approach. *arXiv* **2020**, arXiv:2006.16628.
22. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Society. Ser. B Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
23. Zhao, M.; Zhong, S.; Fu, X.; Tang, B.; Pecht, M. Deep Residual Shrinkage Networks for Fault Diagnosis. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4681–4690. [[CrossRef](#)]
24. Zhang, J.; Ghanem, B. ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–21 June 2018; pp. 1828–1837.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.