

Acquisition of channel state information for mmWave massive MIMO: traditional and machine learning-based approaches[†]

Chenhao QI^{1*}, Peihao DONG², Wenyan MA¹, Hua ZHANG¹,
Zaichen ZHANG¹ & Geoffrey Ye LI^{3*}

¹*School of Information Science and Engineering, Southeast University, Nanjing 210096, China;*

²*College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;*

³*Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK*

Received 24 February 2021/Revised 29 March 2021/Accepted 13 April 2021/Published online 7 July 2021

Abstract The accuracy of channel state information (CSI) acquisition directly affects the performance of millimeter wave (mmWave) communications. In this article, we provide an overview on CSI acquisition, including beam training and channel estimation for mmWave massive multiple-input multiple-output systems. The beam training can avoid the estimation of a high-dimension channel matrix, while the channel estimation can flexibly exploit advanced signal processing techniques. In addition to introducing the traditional and machine learning-based approaches in this article, we also compare different approaches in terms of spectral efficiency, computational complexity, and overhead.

Keywords beam training, channel estimation, machine learning, massive MIMO, millimeter wave (mmWave) communications

Citation Qi C H, Dong P H, Ma W Y, et al. Acquisition of channel state information for mmWave massive MIMO: traditional and machine learning-based approaches. *Sci China Inf Sci*, 2021, 64(8): 181301, <https://doi.org/10.1007/s11432-021-3247-2>

1 Introduction

Millimeter wave (mmWave) communications have attracted extensive interest from academia, industry, and government as they can make full use of abundant frequency resources at high frequency band to achieve ultra-high-speed data transmission [1–4]. The mmWave communication systems are usually equipped with large antenna arrays, known as mmWave massive multiple-input multiple-output (MIMO), to generate highly directional beams and compensate for the severe path loss in high frequency band. However, the performance of directional beamforming largely relies on the accuracy of channel state information (CSI) acquisition. Compared to the traditional MIMO systems, the CSI acquisition in mmWave massive MIMO systems is challenging. On one hand, the large antenna arrays form a high-dimension channel matrix, whose estimation consumes more resources, e.g., pilot sequence overhead, sounding beam overhead, and computational complexity. On the other hand, the mmWave massive MIMO typically employs a hybrid beamforming architecture, where the radio frequency (RF) chains are much fewer than the antennas. Therefore, we can only obtain a low-dimension signal from the RF chains instead of directly getting a high-dimension signal from the frontend antennas, which makes CSI acquisition much more challenging than usual.

CSI acquisition includes beam training and channel estimation. The beam training sounds the mmWave massive MIMO channel with analog transmit and receive beams to find the beam pairs best fitting for the transmission, which can avoid the estimation of a high-dimension channel matrix. The channel estimation

* Corresponding author (email: qch@seu.edu.cn, geoffrey.li@imperial.ac.uk)

† Invited article

focuses on estimating a high-dimension channel matrix, which flexibly exploits advanced signal processing techniques, such as compressed sensing (CS). Both beam training and channel estimation can exploit machine learning (ML) techniques in addition to the traditional approaches.

In this article, we provide an overview on CSI acquisition for mmWave massive MIMO. We first discuss beam training approaches, including beam sweeping, hierarchical beam training, and ML-based beam training. Then we present channel estimation approaches, including the CS-based sparse channel estimation, array signal processing based channel estimation, and ML-based channel estimation. Finally, we compare different approaches in terms of spectral efficiency (SE), computational complexity, and incurred overhead, and identify some future research topics in this area.

We use the following notations. Symbols for vectors (lower case) and matrices (upper case) are in boldface. $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and conjugate transpose (Hermitian), respectively. Let $\mathbb{C}^{P \times Q}$ denote the set of complex-valued matrices with P rows and Q columns. According to the convention, $\mathcal{O}(\cdot)$ denotes the order of complexity.

2 Channel modeling and system overview

In this section, we first briefly review the state-of-the-art channel models for mmWave massive MIMO systems, followed by the elaboration on the dedicated transceiver architecture. At the end of this section, an enhanced-mmWave massive MIMO system with reconfigurable intelligent surface (RIS) is briefly discussed.

2.1 mmWave massive MIMO channel modeling

It has been shown that mmWave massive MIMO channels follow the Saleh-Valenzuela model [5, 6], which can be further divided into the narrowband model [7] and the wideband model [8]. The narrowband model is a summation of the product of several transmitting channel steering vectors and receiving channel steering vectors, where each element involved in the summation corresponds to a channel multipath component (MPC). We consider a multiuser mmWave massive MIMO system comprising a base station (BS) and U users. Let N_T and N_R denote the number of antennas at the BS and each user, respectively. Define the channel matrix between the BS as $\mathbf{H}_u \in \mathbb{C}^{N_T \times N_R}$ and the u th user as

$$\mathbf{H}_u = \sqrt{\frac{N_T N_R}{L_u}} \sum_{i=1}^{L_u} g_{u,i} \boldsymbol{\alpha}(N_T, \theta_{u,i}) \boldsymbol{\alpha}^H(N_R, \varphi_{u,i}), \quad (1)$$

where L_u and $g_{u,i}$ denote the total number of resolvable paths and the channel fading coefficient of the i th path for the u th user, respectively. The steering vector $\boldsymbol{\alpha}(N, \theta)$ is defined as

$$\boldsymbol{\alpha}(N, \theta) = \frac{1}{\sqrt{N}} \left[1, e^{j\pi\theta}, \dots, e^{j\pi\theta(N-1)} \right]^T. \quad (2)$$

Define the angle of arrival (AoA) and angle of departure (AoD) of the i th path of the u th user as $\vartheta_{u,i}$ and $\varphi_{u,i}$, respectively, which are uniformly distributed over $[-\pi, \pi)$. Then $\theta_{u,i} \triangleq \sin \vartheta_{u,i}$ and $\phi_{u,i} \triangleq \sin \varphi_{u,i}$ if the distances between adjacent antennas at the BS and the users are with half-wavelength. Since the non-line-of-sight (NLOS) MPCs are usually much weaker than the line-of-sight (LOS) MPCs, the mmWave MIMO channel is sparse in the angle domain, which only has a small number of significant entries. However, due to the channel power leakage caused by the limited resolution of phase shifters, this sparse property is not ideal [9], which brings the challenge for CSI acquisition.

While the narrowband model assumes the same delay for different MPCs, the wideband model further considers the different delays of MPCs [10]. To tackle the frequency-selective fading caused by the multipath delay spread in the wideband channel, orthogonal frequency division multiplexing (OFDM) with K subcarriers is typically used to convert a wideband channel into multiple narrowband channels. Define $\mathbf{H}_u^k \in \mathbb{C}^{N_T \times N_R}$ as the channel matrix on the k th subcarrier between the BS and the u th user can be expressed as

$$\mathbf{H}_u^k = \sum_{d=0}^{D-1} \mathbf{H}_{u,d} e^{-j \frac{2\pi k d}{K}}, \quad (3)$$

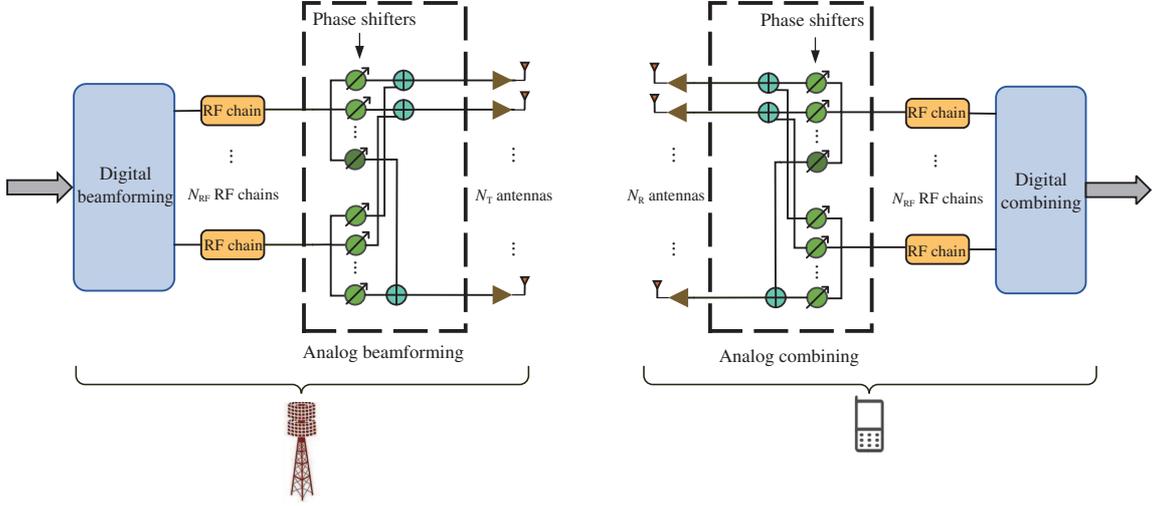


Figure 1 (Color online) Hybrid architecture for mmWave MIMO transceiver.

where D denotes the number of delay taps of the channel. The channel matrix at the d th delay tap can be expressed as

$$\mathbf{H}_{u,d} = \sqrt{\frac{N_T N_R}{L_u}} \sum_{i=1}^{L_u} g_{u,i} p_u(dT_s - \tau_{u,i}) \boldsymbol{\alpha}(N_T, \theta_{u,i}) \boldsymbol{\alpha}^H(N_R, \phi_{u,i}), \quad (4)$$

where $p_u(t)$, T_s , and $\tau_{u,i}$ denote the pulse shaping, the sampling interval, and the delay of the i th path for the u th user, respectively.

As an extension of the wideband channel model, a practical model in 3GPP TR 38.901 replaces each channel MPC with a channel cluster, which further includes a number of channel rays with some minor angle offsets. The 3GPP TR 38.901 reference implementation, known as the quasi deterministic radio channel generator (QuaDRiGa) [11], has provided a system-level software to simulate realistic radio channels including mmWave channels.

Based on the wideband model, the beam squint effect, which is caused by different delays of the received signals from different receiving antennas, has been investigated [12, 13]. However, most studies neglect the beam squint effect by assuming that the received signal contributed by the same MPC has the same delay for different antennas. Sometimes it is necessary to further consider the detailed aspects of the mmWave MIMO channel model so that the designed mmWave system can better match the practical scenarios.

2.2 mmWave MIMO transceiver architecture

To balance complexity and performance, hybrid architecture, as shown in Figure 1, is widely used in mmWave MIMO transceiver, where several antennas share one RF chain and signal processing is partially in the digital and partially in the analog domains. Properly designed hybrid beamforming and combining, with much lower complexity, can approach the achievable rate performance of fully-digital beamforming [14]. Different from the fully-digital beamforming that needs the same number of RF chains as that of the antennas, hybrid beamforming and combining uses much fewer RF chains than the antennas and substantially reduces the hardware costs. From Figure 1, hybrid beamforming at the transmitter side typically includes analog beamforming and digital beamforming, where the former generates highly directional beams based on large antenna arrays and the latter mitigates the interference among parallel data streams supported by multiple RF chains. Analog beamforming connects RF chains to antennas by the phase shifters, switches, or electromagnetic lens. Hybrid combining works similarly as hybrid beamforming, but at the receiver side.

There are two different connection modes for the phase shifters: full-connection and partial-connection modes. In the full-connection mode, each RF chain connects to all antennas while in the partial-connection mode, each RF chain only connects to a subset of antennas and therefore can further reduce the hardware complexity but with some performance degradation. The phase shifters can only change the phases of

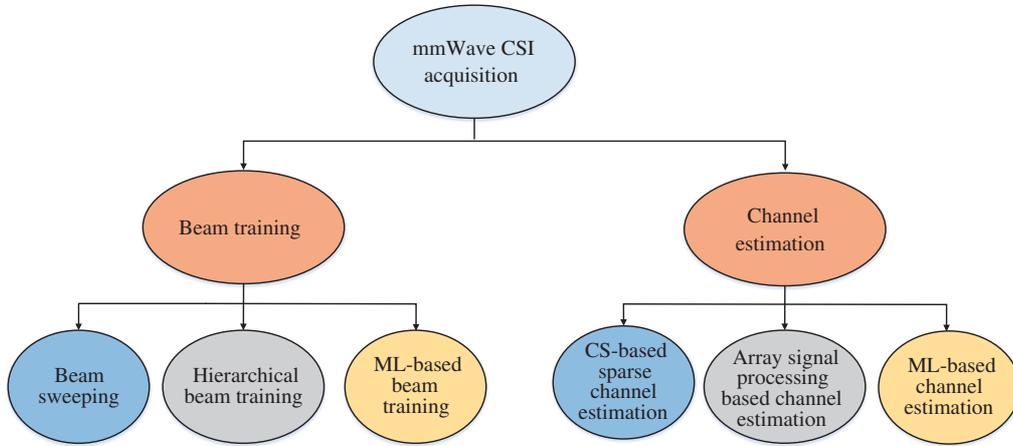


Figure 2 (Color online) CSI acquisition for mmWave massive MIMO.

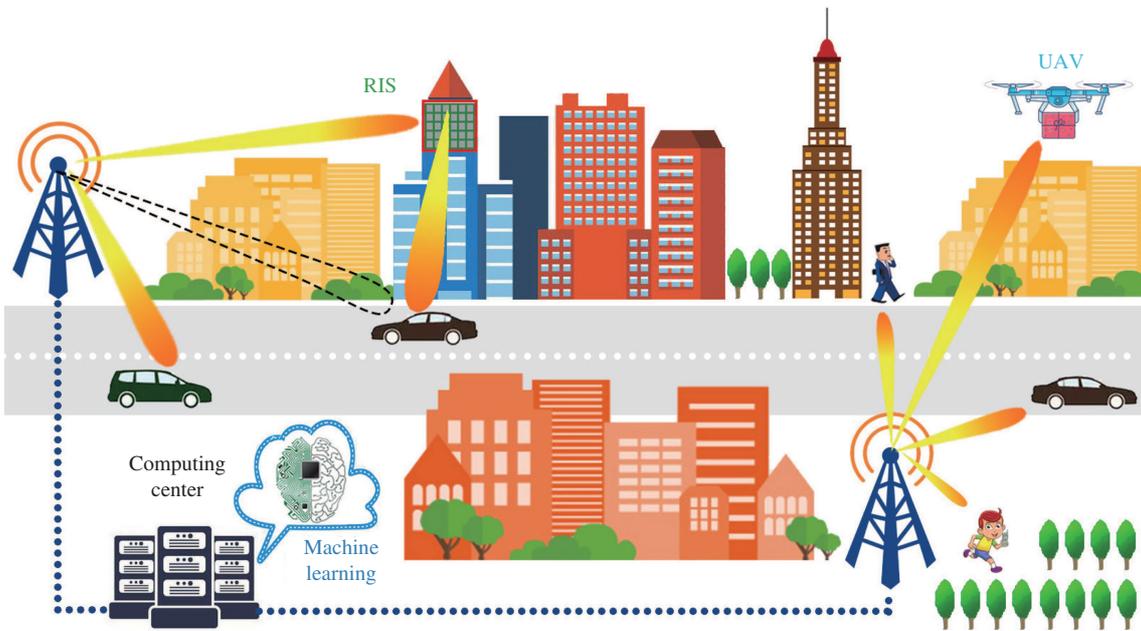


Figure 3 (Color online) Illustration of mmWave massive MIMO communications system.

signals usually with limited resolution. To further reduce the hardware complexity, low-cost switches with only on-and-off binary states can be used to partially replace the phase shifters [15,16]. But the time efficiency as well as the signal jitter when switching between two states should also be considered. These hardware constraints bring new challenges to CSI acquisition comparing with other communication systems. We will focus on CSI acquisition for mmWave communications in this article, as shown in Figure 2.

2.3 RIS to enhance signal coverage

Recently, RIS has been proposed to improve signal coverage [17–20]. The RIS, sometimes also known as the intelligent reflecting surface (IRS), can reflect the incident wireless signal by changing its amplitude and phase, which functions similarly as a mirror to reflect the incident light. Therefore, it can help cover the area without the LOS channel path, as illustrated in Figure 3, which well compensates for the deficiency of the mmWave MIMO signal that mostly relies on the directional transmission. In mobile mmWave wireless communications with ground vehicles [21] or unmanned aerial vehicles (UAVs) [22],

the effective channel links between the users and the BS may be blocked by the buildings or other users. In this context, the RIS can effectively reflect the signal and recover the link. Therefore, the RIS can improve the reliability of the mmWave communications. But different from wireless relays, RIS is generally implemented by low-cost hardware, such as meta-material antennas, positive intrinsic-negative diodes, field effect transistors, and can passively reflect the signal in the RF frontend without any baseband signal processing capabilities. The RIS introduces additional channel links between the transmitter and the RIS as well as those between the RIS and the receiver [23–25]. As a result, the CSI acquisition in the RIS-assisted mmWave massive MIMO system becomes much more complicated and is different from that in mmWave systems. The traditional and ML-based CSI acquisition approaches can be performed with the assistance of computing center that provides powerful computing capacity, as shown in Figure 3. Since the RIS is a recent emerging technique for mmWave communications, there is no much mature work in CSI acquisition for RIS-assisted mmWave massive MIMO system and therefore we will skip the detailed discussion in this article.

3 Beam training

Beam training is a process to find a pair of transmit and receive beams that best align with the strongest MPCs of the mmWave MIMO channel. Sometimes beam training is also called beam alignment [26]. For the mmWave massive MIMO using electromagnetic lens that generally function as a DFT transformation from the angle space to the beamspace, beam training is also called beam selection. Codebook-based beam training is a popular and general method [27–30]. For this method, a codebook is first established at the transmitter and the receiver, and each codeword in the codebook, similar to the channel steering vector, generates a beam. During the beam training, we use a pair of codewords selected from the codebooks at the transmitter and receiver, respectively, to generate a pair of beams and then measure the received signal power. The pair of codewords corresponding to the largest received signal power is identified as the result of beam training. As shown in Figure 2, the beam training approaches can be categorized into three branches, including beam sweeping, hierarchical beam training, and ML-based beam training. In brief, beam training identifying the best transmit and receive beams can avoid the challenge coming from estimating a high-dimension mmWave MIMO channel matrix in the scale of the number of antennas. With beam training, we only need to estimate a low-dimension equivalent channel matrix in the scale of the number of the RF chains. Once the beam training is finished, we use the classical methods such as least square or minimum mean squared error estimation, to estimate the equivalent channel matrix with a small number of pilot symbols.

3.1 Beam sweeping

The most straightforward approach on beam training is testing all pairs of codewords to find the best one, which exhaustively selects each pair of codewords one by one and is thus very similar to sweeping different angles in the space. As a result, this approach is called beam sweeping [31]. After beam sweeping, we can find the best pair of codewords and use them for transmit and receive beamforming. However, we may generate even better fine-grained beams by further performing angle estimation based on the best pair of codewords [32,33], since the codewords used in the beam sweeping usually point at coarse-grained discrete angles.

To reduce the overhead of beam sweeping, one scheme only uses a subset of codewords for the beam sweeping, which is called partial beam sweeping [34]. As shown in Figure 4, the transmit and receive codewords generating transmit and receive beams, respectively, can form two-dimensional grids, where each grid represents a pair of transmit and receive codewords. The partial beam sweeping divides the beam training into two stages for the initial test (INTS) and the additional test (ADTS). While the original beam sweeping exhaustively tests all grids in Figure 4, the partial beam sweeping method only tests half or fewer of all the grids in the stage of INTS. By exploiting the coherence of different grids caused by the mmWave channel power leakage, the neighboring two columns and two rows with the largest average energy are identified, where the two tested grids on the cross of the identified two columns and two rows are predicted as the best codeword pair or best grid. Then the untested grids in the neighbor of the predicted best grid are tested in the ADTS, which only needs a very small training overhead [34]. Furthermore, to break down the constraint that the grid selection in the INTS should be equally spaced, the probabilistic selection is introduced for the grid selection so that the number of the selected grids

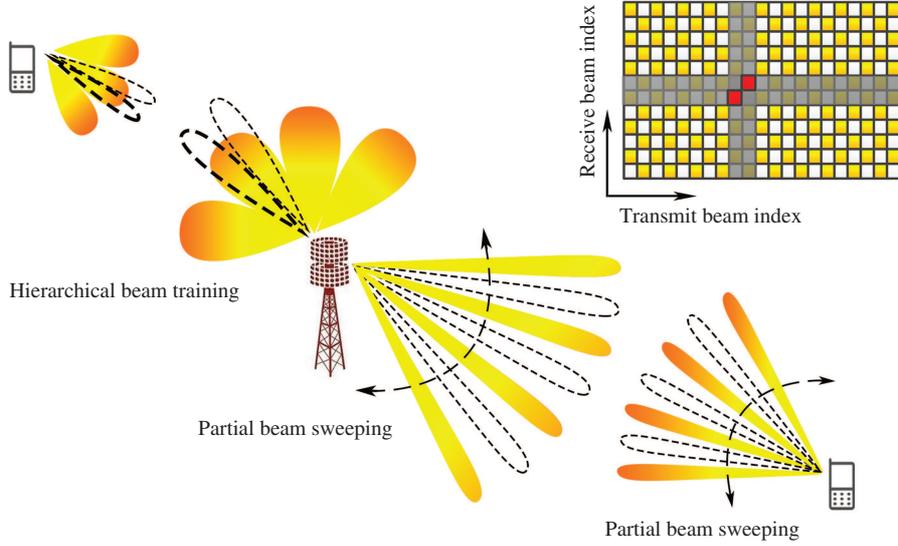


Figure 4 (Color online) Illustrations of partial beam sweeping and hierarchical beam training.

in the INTS can be set randomly. Compared with the beam sweeping, the partial beam sweeping can substantially reduce the beam training overhead with small sacrifice of SE performance.

Aside from directly reducing the training overhead, another option is to consider more efficient use of the total training energy budget. In [35], the beam sweeping is divided into two stages. In the first stage, all beam pairs are tested in the same way as the beam sweeping, but a set of less favorable pairs determined from the received signal profile is then eliminated. In the second stage, an extra measurement is taken for each of the survived pairs and the best beam pair is determined by comprehensively considering these two stages.

Auxiliary beams can also be used to assist the beam sweeping as well as the beam tracking. Different from using only one beam for beam sweeping, multiple beams with one main beam and some auxiliary beams are used together for beam sweeping. Since we can compare the power between the main beam and the auxiliary beams, estimation of channel AoA and AoD can be improved [36]. Auxiliary beams are also capable of tracking the angle variations in mobile mmWave communications scenarios, where high-resolution angle tracking strategies can be designed [37].

3.2 Hierarchical beam training

Another approach to reducing the beam training overhead is hierarchical beam training, which uses a fast codeword selection strategy based on a hierarchical codebook [38–44]. The hierarchical codebook has multiple layers with increasing numbers of codewords from the upper to the lower layers. That is, the upper layer consists of a small number of low-resolution codewords that generate wide beams and the lower layer has a large number of high-resolution codewords that generate narrow beams with highly directional beam gain. As shown in Figure 4, the hierarchical beam training usually first tests the mmWave channel with some wide beams generated by low-resolution codewords at the upper layer and then narrows down the beam width layer by layer until a best codeword at the bottom layer is obtained.

Many studies focus on the codeword design under different hardware constraints, including the limited resolution, constant envelope of phase shifters, and the limited number of RF chains. To design a codeword with a wide beam coverage, the simplest scheme is to power off some antennas; but will consequently reduce the total radiation power and affect the signal coverage. A better scheme is to divide a large antenna array into several subarrays [39], where each subarray generates a beam and we optimize the weighted summation of the subarray beams to approximate the objective beam pattern. To further improve the performance of beam pattern for the generated beams under the hardware constraints, e.g., reducing the mainlobe fluctuations or enlarging the sidelobe attenuation of the generated beam, we may design ideal codewords in the first step by ignoring the hardware constraints and then generate practical codewords considering the hardware constraints to approximate the ideal codeword by alternative minimization in the second step [30]. For the non-convex codeword design problem under the constraints

of the transition band and the ripple on mainlobe or sidelobes, an efficient algorithm is proposed and closed-form expressions of the beam training error rate are derived [40]. To support simultaneous beam training for multiple users and therefore improve the efficiency of multiuser beam training, adaptive hierarchical codebook based on multi-mainlobe codewords can be further considered, where each mainlobe can be pointed at a spatial region that a user may be probably located in [45, 46]. As such, the total training overhead can be substantially reduced.

3.3 ML-based approaches

Recently, ML has been introduced for the mmWave beam training by exploiting the temporal, frequency or spatial correlation of the mmWave communications. Two main branches of ML for beam training include supervised learning and reinforcement learning (RL), where the labeled training dataset is required by the former but not by the latter. The RL enables an agent to learn how to take a good action on the environment aiming at getting the largest rewards from the environment, by continuous exploration and exploitation.

3.3.1 Supervised learning-based approaches

The commonly used methods for supervised learning include the methods based on neural network (NN), such as deep NN (DNN), convolutional NN (CNN), recurrent NN (RNN). To reduce the overhead of beam sweeping, a beam alignment method only testing partial beams can be used to predict the beam distribution and the best beam pair based on the NN, where the NN is trained offline to exploit the channel spatial correlation using simulated mmWave channel data [47]. To improve the efficiency of beam training, beams with different widths, e.g., wide beams and narrow beams, can be used. Given the wide beam measurements, a super-resolution-inspired deep learning method based on the CNN can be used to estimate the beam qualities of narrow beams and therefore select the best narrow beam for data transmission. By capturing the spatial correlation as well as the temporal correlation, a beam qualities prediction model based on a convolutional long short term memory (LSTM) network is used to estimate the current beam qualities from the historical data [48]. In fact, we may treat a pair of transmit and receive beams as a pixel of an image, where the beam training can be treated as an image reconstruction problem. Based on the off-line eigen-beam extraction from the original beam-domain receive power map (BDRPM) and partial beam measurements, we can reconstruct a full BDRPM by the DNN [49].

To assist the beam training, additional information including the positions and orientation of users [50, 51], situational awareness [52], and those coming from sub-6GHz channel [53] and light detection and ranging (LIDAR) sensors [54] can be employed. In [50], a pair of transmit and receive beams can also be considered as a fingerprint at a particular position. Based on an NN trained by a fingerprint database, we can predict the candidate beam pairs given the position of the users, which will only incur small beam training overhead. In [51], the position and orientation are input to a DNN that essentially functions as a multi-labeled classifier to obtain the probability of each beam pair being the best one under indoor scenarios. By collecting and exploiting the situational awareness information, e.g., the positions and types of the receiver and its neighbors, the mmWave beam training can be formulated as a classification problem, where the ensemble learning methods can be leveraged to design efficient and robust schemes [52]. It is well known that the sub-6 GHz wireless channels have better propagation condition than the mmWave ones. If sub-6 GHz channels are available, the mapping functions to predict the best mmWave beam pair from the sub-6 GHz channels can be better modeled by the DNN than the analytical methods [53]. In autonomous driving, LIDAR is a sensor widely equipped for high-resolution mapping and position, where the LIDAR sensor data can be used to train the CNN to find a set of beams for mmWave communications [54].

3.3.2 RL-based approaches

Different from the supervised learning that relies on the labeled training dataset, the RL does not need such kind of information and therefore is more flexible. Aiming at obtaining the largest rewards from the environment, an agent learns how to take a good action towards the environment through continuous exploration and exploitation. As a straightforward approach of RL, multi-armed bandit (MAB) can be applied for mmWave beam training. MAB addresses a sequential decision problem that an agent selects an arm with a trade-off between exploitation and exploration to maximize the expected reward.

Each pair of transmit and receive beams can be regarded as an arm of the MAB model. Then the beam training can be formulated as the problem of finding a best arm, which can be typically solved by upper confidence bound (UCB) algorithms. In [55], the correlation and unimodality properties across beam training are exploited, which inspires the proposed unimodal beam alignment algorithm with the asymptotically optimal performance. In [56], the beam training is divided into two stages, including the beam pair selection stage and the beam refinement stage. In the first stage, coarse beam directions are identified by the proposed greedy UCB and risk-aware greedy UCB algorithms. In the second stage, a modified optimistic optimization algorithm further refines the results from the first stage. Considering the multipath propagation of mmWave channels, M arms with M largest UCBs instead of only one arm are selected at each time slot, where the transmission rate is iteratively updated according to the feedback [57].

In addition to the UCB algorithm, the Bayesian methods can also be used to solve the beam alignment problem based on the MAB model. The linear Thompson sampling can be integrated with sparse Bayesian learning and Kalman filters for smart exploration of feasible actions [58]. In [59], a second-best preference policy is proposed to restrict the value function and select the beam pair with the current sub-optimal priority to achieve the best balance between exploitation and exploration.

Besides the UCB and Bayesian algorithms, a fast ML method based on contextual information is proposed to use rough user location and aggregate the received data to learn and continuously adapt to the actual channel environment [60], while a hierarchical beam alignment method exploiting the inherent correlation among beams determines the optimal beam with accelerated BA process and reduced beam training overhead [61].

Compared to the MAB model, it is more sophisticated to model the beam training as a Markov decision process, where Q-learning methods are typically employed. In [62], the estimation of beam steering angle is divided into two phases. In the first phase, Q-learning based beam tracking finds the optimal beam. In the second phase, several beams neighboring the optimal beam found in the first phase are tested, where the tested beam pair with the largest received signal strength is then selected for the modified auxiliary-beam-pair-based angle estimation. In fact, the Q table in the Q-learning methods can be well approximated by the DNN, which is named as deep Q-network (DQN). In [63], the DQN is used to learn the change of the mmWave communication channels and intelligently train the beams with low overhead.

4 Channel estimation

Channel estimation, as the other category of CSI acquisition, aims at accurately estimating the mmWave massive MIMO channels. As shown in Figure 2, the mainstream channel estimation methods can be divided into three branches, including CS-based sparse channel estimation, array signal processing-based channel estimation, and ML-based channel estimation.

4.1 CS-based sparse channel estimation

Extensive studies on channel measurement and modeling have demonstrated that the scattering environment in mmWave frequencies is sparser than its lower frequency counterpart, such as in the sub-6 GHz channel. Therefore, we can exploit the channel sparsity by formulating channel estimation as a sparse recovery problem and solve it using the CS algorithms. In general, we first obtain the AoDs/AoAs and then estimate the channel gains. Let \mathbf{z} denote a sparse vector, where its locations and values of the non-zero entries indicate the quantized AoDs and AoAs and path gains, respectively. The performance loss caused by the quantization error proves to be minor with a fine-grain enough candidate set for AoDs and AoAs [64]. Owing to the channel sparsity, the number of non-zero entries in \mathbf{z} is much smaller than its dimension.

The orthogonal matching pursuit (OMP) algorithm is popular for sparse channel estimation for mmWave wireless system [23, 64–66]. During the downlink channel training, random beamforming and projection can be used to estimate the mmWave channel, where the OMP algorithm is adopted to estimate the quantized AoDs and AoAs corresponding to the largest entries of \mathbf{z} and a lower bound on the achievable rate as a function of the CS measurements is provided [65]. However, the proposed method in [65] only estimates the LOS channel path of the mmWave channel, which does not make full use of mmWave channel resources. To solve this problem, OMP-based estimation methods for multiple channel paths are proposed [64, 66]. Different from the random beamforming adopted in [65], a hierarchical codebook

with adaptive CS algorithm to recover \mathbf{z} is designed in [64], which can invoke hybrid processing during the pilot transmission and thus yield more efficient beam patterns to capture the non-zero entries in \mathbf{z} . The AoD, AoA, and the gain of each channel path are iteratively distilled from the residual signal, where the residual signal is iteratively updated by subtracting the components contributed by the previously estimated paths. In [66], the OMP-based channel estimation is evaluated with a redundant dictionary, which consists of array response vectors with finely quantized angle grids non-uniformly distributed in $[0, \pi]$. The lower and upper bounds of the sum-of-squared errors of the proposed method are analytically derived, while the pilot vectors are designed to minimize the coherence of the sensing matrix. Note that the receiving noise in the mmWave system is not white since it is multiplied by the hybrid combiner at the receiver. In [67], a simultaneous-weighted-OMP algorithm exploits the common support among different subcarriers, where the spatial noise components are simultaneously whitened by Cholesky factorization of the hybrid combiner and the most likely support index in the sparse beamspace channels can be estimated more accurately before whitening. Since the OMP algorithm only saves the best candidate atom having the largest projection with the residue at each iteration, it estimates the dominant channel entries sequentially and greedily, which cannot guarantee the globally optimal performance. In [68], a regularized multipath matching pursuit algorithm better than OMP is proposed for sparse channel estimation, where multiple candidate atoms instead of only one atom are kept at each iteration and a regularization step is introduced to reduce the computational complexity.

Sparse channel estimation for mmWave massive MIMO with electromagnetic lens structure draws the research interest, since the electromagnetic lens with switches have lower hardware complexity than the phase shifter networks. In [69], mmWave system with 2D lens is considered, where the structural characteristics of beamspace channel show that the support of the sparse beamspace channel is uniquely determined by its largest entry. Therefore, only one projection is needed to determine the beamspace channel support of each path, which can substantially reduce the computational complexity. Better performance can be achieved by the proposed method in [69], especially in the low signal-to-noise ratio (SNR) region, compared to the classical CS algorithms. The mmWave system with 3D lens is considered in [70, 71]. It is shown that the most power of the beamspace channel matrix is concentrated on a much smaller submatrix, which inspires the adaptive support detection method to model the beamspace channel support as a rectangle, with the adaptively modified width and length of the rectangle [70]. It is further demonstrated in [71] that the main power of the beamspace channel is more concentrated on an area of dual crossing than the rectangle, which brings forward a dual crossing channel estimation method to iteratively refine the selection of dominant entries until the stop condition is met. At each iteration, two adjacent columns and rows with the largest channel power are identified to form the area of dual crossing, where the selection of the dominant entries is adjusted by deleting the weakest outer entry at dual crossing corners and then adding a new entry outside of the strongest outer entry.

To mitigate the performance loss caused by the quantization error, we may increase the resolution of the AoD/AoA candidate set, which however incurs high hardware complexity. To tackle this problem, off-grid super-resolution channel estimation based on gradient descent [72], atomic norm minimization [73], Dirichlet kernel [74], and distributed grid matching pursuit [75] has been investigated. Inheriting the formulated sparse recovery problem, the intractable l_0 -norm minimization is equivalently transformed into the minimization of a surrogate function [72]. By resorting to the iterative reweighting method, the above equivalent problem is then solved with the AoDs and AoAs updated based on gradient descent. If the initialized channel sparse level is set larger than the genuine one, we can find the genuine number of paths by iteratively pruning those false paths. To reduce the computational complexity, singular value decomposition-based preconditioning can be further used by properly initializing the AoDs and AoAs. Channel estimation can also be formulated as an atomic norm minimization problem, which avoids discretizing the angle spaces of the AoA/AoD into grids and therefore achieves high-accuracy channel estimation [73]. Further reformulating the atomic norm minimization problem as a semi-definite program, alternating direction method of multipliers can be employed where each iteration only needs closed-form computation. By assuming that the beamspace parameters can be any value in the beamspace domain and do not restrict to predefined discrete angles, implicit Dirichlet kernel structure in the Fourier domain is used, where the OMP algorithm is employed to find the maximum of the Dirichlet kernel peaks instead of recovering all the non-zero entries in the virtual DFT domain [74]. In [75], a channel estimation method based on distributed grid matching pursuit is developed to iteratively detect and adjust the channel support, where the adaptive measurement matrix is determined by the most possible channel paths in the outer loop iteration and the AoAs and AoDs estimation associated with the LOS path is

improved with the grid matching strategy in the inner loop iteration.

4.2 Array signal processing based channel estimation

Array signal processing technique, originally proposed for the AoA estimation in radar system with a large antenna array, can also be applied for mmWave massive MIMO channel estimation. The estimating signal parameters via rotational invariance techniques (ESPRIT) algorithm and multiple signal classification (MUSIC) algorithm are the classical AoA estimation algorithms used in radar systems. But different from the radar system, the mmWave massive MIMO typically employs a hybrid beamforming architecture where the RF chains are much fewer than the antennas, meaning that we can only obtain a low-dimension signal from the RF chains instead of directly getting a high-dimension signal from the frontend antennas. Compared to the high-dimension mmWave MIMO channel matrix to be estimated, the low-dimension signal from the RF chains is not enough.

Channel estimation methods based on the ESPRIT algorithm are proposed in [76–80]. The training signals at both the BS and each user are designed to obtain the low-dimensional effective channel [76, 77], while the spatial smoothing and the forward backward averaging techniques are adopted to alleviate the impact of coherent signals caused by multiple AoAs or AoDs close to each other. Moreover, the channel matrix is converted to a real matrix so that the computational complexity can be reduced. To avoid the information loss from the high-dimension signal to the low-dimension signal, a straightforward method is to power off some antennas so that the time slots for channel estimation equal that of the antennas, by which we estimate the AoAs and AoDs through a submatrix of the high-dimension mmWave MIMO channel matrix and eventually combine the submatrices together to obtain an estimate of the full channel matrix. Two methods based on the beamspace ESPRIT algorithm are proposed in [78, 79], where the hybrid precoding matrix and combining matrix are designed as the DFT matrix so that the beamspace ESPRIT algorithm can be directly applied. However, the hybrid precoding matrix and combining matrix can only be partial DFT matrix, since the time slots for channel estimation are less than the total antennas. Consequently, the AoAs and AoDs to be estimated are restricted in a small range, indicating that the aforementioned methods are not generally applicable for the estimation of any AoAs and AoDs. Note that powering off antennas will reduce the total radiation power and therefore reduce the signal coverage. A better method is to turn off fewer antennas, e.g., only one antenna [80]. To achieve high-resolution estimation of AoAs and AoDs using the ESPRIT method, multiple stages of pilot transmission can be considered. AoAs and AoDs can be first estimated by a two-dimensional ESPRIT scheme and then paired, based on which the channel gain can be estimated by the least squares (LS) method. To further reduce the pilot overhead, a one-dimensional ESPRIT and minimum searching-based scheme only needing two stages are proposed, where the AoAs are first estimated and then the AoDs are obtained by searching the minimum value within the identified mainlobe. Besides, hybrid beamforming and combining matrices for the pilot transmission can be designed to yield almost equal received powers for any AoA and AoD so that robust super-resolution channel estimation can be achieved.

Channel estimation methods based on the MUSIC algorithm are proposed in [81, 82]. The beamspace 2D MUSIC algorithm in [81] estimates the AoAs and AoDs of the incoherent signals and then the LS method is adopted to estimate the channel gain for each path. Similar to [78], the hybrid precoding and combining matrices are partial DFT matrices, indicating that the beamspace 2D MUSIC algorithm can only estimate AoAs and AoDs in a small range. For the coherent signals, the beamspace 2D MUSIC algorithm is also used [82]. Since the signals are coherent, different signals can merge into one signal, which decreases the number of independent sources. In this context, the received signals are multiplied by an exchange matrix so that the number of antennas in subarrays equals the total number of array entries. However, the hybrid precoder and combiner can only be partial DFT matrices, which also suffers from the drawbacks of a small estimation range.

4.3 ML-based solutions

The fast development of ML motivates its application in channel estimation for mmWave massive MIMO. In particular, deep learning (DL) based methods have attracted many interests since they can extract features via multiple neural layers and accelerate the computation dramatically resorting to parallel computing. The DNN can be trained to estimate the channel parameters including AoAs, AoDs, and channel gains.

ML-based channel estimation using the denoising methods is proposed in [83–88]. In [83], the beamspace mmWave massive MIMO channel matrix is treated as a 2D image, where a learned denoising-based approximate message passing (LDAMP) NN integrating denoising CNN estimates the mmWave channel by subtracting the estimated residual noise from the noisy channel. To further improve the estimation performance of [83], the denoising process is divided into several stages, mainly including the preprocessing stage and the fine processing stage [84, 85]. In [84], a learned approximate message passing (LAMP)-based network with deep residual learning for channel estimation includes two stages. In the first stage, the LAMP network exploits the beamspace channel sparsity to get a preliminary channel estimation, while in the second stage the deep residual learning further refines the coarse estimation obtained in the first stage and reduces the impact of channel noise. In [85], a fully convolutional denoising approximate message passing algorithm has three stages. In the first stage the LAMP network processes the received signal and the mmWave channels; in the second stage, a fully convolutional denoising network extracts noise characteristics and obtains the noise level; and in the final stage, the feature channel is increased or decreased, respectively, by down-sampling or up-sampling. To accurately estimate the beamspace channel, the original soft-threshold shrinkage function in the LAMP network can be replaced by a derived Gaussian mixture shrinkage function, based on the prior information that the beamspace channel elements can be modeled by the Gaussian mixture distribution [86]. Note that the above denoising-based channel estimation methods are tailored to specific noise levels. To improve the flexibility and efficiency of denoising-based channel estimation, a fast and flexible denoising CNN is proposed, where the sparse mmWave channel matrix is also treated as a natural image [87]. By using a flexible noise level map as the input, the proposed denoising network can be used for a wide range of SNR levels. Moreover, to reduce the training and testing latency, each image in the proposed denoising network is divided into several sub-images.

In fact, the ML can also be directly used to predict the mmWave channel besides functioning as denoisers [89–92]. In [89], the DNN learns the mapping function between the received omni-beam patterns and the mmWave channel, while the LSTM is adopted to track the channel. However, in practical mmWave wireless system equipped with a large antenna array, it is difficult to obtain the omni-directional signal, which requires powering off all the antennas except one antenna and consequently reduces the radiation power and signal coverage. Since the mmWave channel exhibits sparse property in beamspace, the beamspace channel amplitude can be introduced for the training of the DNN [90]. Then based on it, the significant entries in the beamspace channel vector can be predicted. This method is more flexible than directly predicting the positions of nonzero entries, especially when the beamspace channel is not ideally sparse. The non-zero entries of the sparse channel can be obtained simultaneously by a well trained DNN with significantly reduced running time, instead of the sequential greedy search by heuristic sparse recovery algorithms, such as OMP. Moreover, the DNN can be trained under a noisy channel environment with interference, making it robust to various channel conditions. To provide accurate channel estimation close to the performance of the ideal minimum mean-squared error estimator, a robust framework on DL-based channel estimation has been proposed for mmWave massive MIMO-OFDM systems [91]. The main idea is using the deep CNN to capture the spatial, frequency, and temporal channel correlation simultaneously, which is difficult for the traditional methods. Figure 5 illustrates the framework of deep CNN-based channel estimation, where N_T and N_R are the numbers of antennas at the transmitter and the receiver, respectively, and $M_T[n]$ and $M_R[n]$ are the numbers of beamforming and combining vectors, respectively, for pilot transmission in the n th coherence interval. As shown in Figure 5, the received pilots at Q subcarriers in the current coherence interval are first processed by the tentative estimation module. Then the tentatively estimated channels together with those estimated in the previous $S - 1$ coherence intervals are input into the developed deep CNN to obtain the estimated channels at Q subcarriers of the current coherence interval. Specifically, the input of the deep CNN consists of $2QS$ real-valued matrices, which are first processed by sixty-four $3 \times 3 \times 4$ convolutional kernels. The obtained feature maps are then filtered by eight convolutional layers with sixty-four $3 \times 3 \times 64$ kernels. The output layer applies four $3 \times 3 \times 4$ convolutional kernels to process the extracted feature maps to approximate the real channel matrices. Through proper design, the CNN is able to capture the channel temporal correlation efficiently with low complexity even without incorporating the long short-term memory architecture. Besides improving the estimation accuracy and saving the running time, the proposed framework also reduces the pilot overhead. Specifically, several successive coherence intervals are grouped as a channel estimation unit, within which the pilot overhead can be reduced gradually at the cost of limited performance loss. The developed CNN-based framework is robust to different propagation scenarios even unseen in the offline

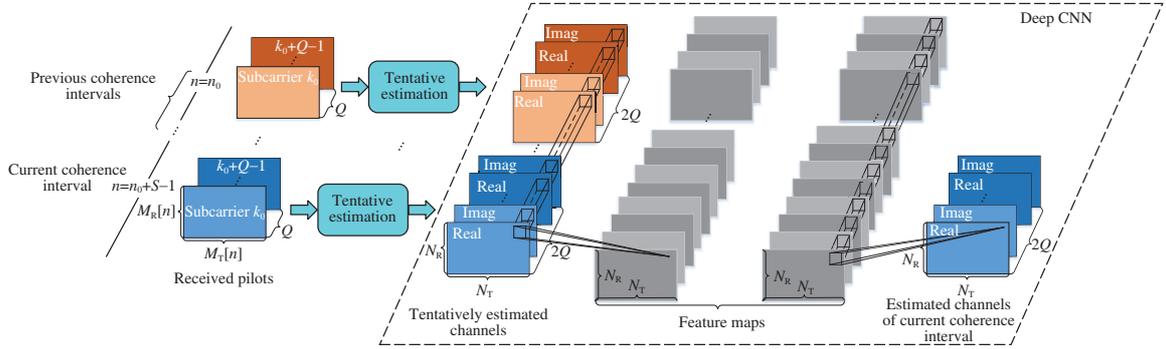

Figure 5 (Color online) Framework on deep CNN-based channel estimation.

Table 1 Comparisons of different channel estimation or beam training schemes

	Scheme name	Computational complexity	Overhead
Beam training	Partial beam sweeping [34]	$\mathcal{O}(U(\frac{N_T}{2N_{RF}} + 2) + 4U^3)$	$2(\frac{N_T}{2N_{RF}} + 3)$
	Hierarchical beam training [39]	$\mathcal{O}(U \log_2 N_T + 4U^3)$	$2U \log_2 N_T + U$
	MAB-based scheme [55]	$\mathcal{O}(N_T^3)$	Depend on convergence speed
Channel estimation	Adaptive CS-based scheme [64]	$\mathcal{O}(U \log_2 N_T + 4U^3)$	$2U \log_2 N_T + U$
	ESPRIT-based scheme [80]	$\mathcal{O}(8UT^3 + 2UL^2(10L + 5T))$	T
	Deep CNN-based scheme [91]	$\mathcal{O}(UN_T^2 + UN_T \sum_{l=1}^{L_c} F_l^2 N_{l-1} N_l)$	T

training stage and without any prior knowledge of the channel conditions.

5 Performance evaluation

In this section, we compare several typical approaches introduced in Sections 3 and 4 in terms of SE, computational complexity, and overhead.

We consider a multiuser mmWave massive MIMO system, where a BS equipped with $N_T = 64$ antennas and $N_{RF} = 4$ RF chains serves $U = 3$ single-antenna users. The mmWave massive MIMO channels are generated based on the Saleh-Valenzuela model. In Table 1, we compare three beam training schemes and three channel estimation schemes, including partial beam sweeping [34], hierarchical beam training [39], the MAB-based scheme [55], the adaptive CS-based scheme [64], the ESPRIT-based scheme [80], and the deep CNN-based scheme [91].

5.1 SE performance

Figure 6 compares the SE at different SNRs for the six schemes. From the figure, at the low SNR region, the partial beam sweeping, the MAB-based scheme, and the deep CNN-based scheme outperform the other three schemes. At SNR = -5 dB, the deep CNN-based scheme has 56.8%, 143.1% and 73.2% performance improvement over the adaptive CS-based scheme, the ESPRIT-based scheme, and the hierarchical beam training scheme, respectively, while the partial beam sweeping-based scheme has 104.7%, 217.4% and 126.1% performance improvements over the adaptive CS-based scheme, the ESPRIT-based scheme, and the hierarchical beam training scheme, respectively. Since the narrow beams with large beam gain are used by both the partial beam sweeping and the MAB-based scheme, better anti-noise performance can be achieved. On the contrary, for the hierarchical beam training and adaptive CS-based schemes relying on hierarchical codebooks, wide beams with relatively small beam gain are used at the upper layer, which leads to worse anti-noise performance. The deep CNN-based scheme can train the neural network under the predefined noisy condition and can guarantee better anti-noise performance than the other two schemes. At SNR = 15 dB, the ESPRIT-based scheme has 8.3%, 15.7%, 18.0%, 23.6%, and 36.3% performance improvement over the deep CNN-based scheme, the partial beam sweeping scheme, the adaptive CS-based scheme, the MAB-based scheme, and the hierarchical beam training scheme, respectively. Since the ESPRIT-based scheme employs singular value decomposition to achieve super-resolution estimation of the AoA and AoD, it performs the best in the high SNR region and the worst in the low SNR region among all the schemes.

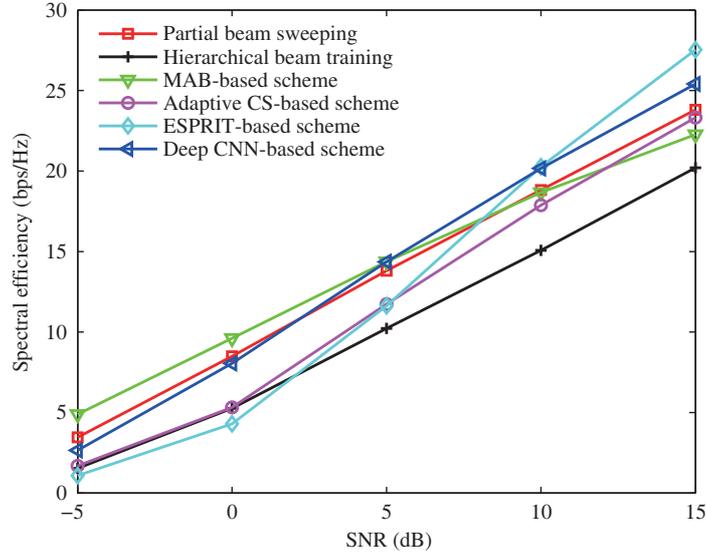


Figure 6 (Color online) Comparisons of spectral efficiency for different schemes in terms of SNR.

5.2 Computational complexity

The computational complexity for the six schemes is compared in Table 1. To compare computational complexity for different schemes, we set the parameters of the deep CNN-based scheme as [91], where $Q = 1$, the number of convolution layers $L_c = 10$, the side length of the filters used by the l th layer $F_l = 3$, the number of input feature maps of the l th layer $N_{l-1} = 2$, for $l = 1$ and $N_{l-1} = 64$, for $l = 2, \dots, 10$, and the number of output feature maps of the l th layer $N_l = 64$, for $l = 1, \dots, 9$ and $N_l = 2$, for $l = 10$. By setting the number of channel paths and the length of pilot symbols to be $L = 2$ and $T = 8$, respectively, we can figure out that the computational complexities of partial beam sweeping, hierarchical beam training, MAB-based scheme, adaptive CS-based scheme, ESPRIT-based scheme, and deep CNN-based scheme, are in the orders of the magnitudes of 10^2 , 10^2 , 10^5 , 10^2 , 10^4 and 10^7 , respectively. In brief, the deep CNN-based scheme has the largest computational complexity among all the schemes but can be accelerated dramatically via parallel computing.

5.3 Overhead comparison

To compare the overhead for different schemes in the unit of time slots, we assume either a test of a beam pair or transmission of a pilot symbol occupies one time slot in Table 1. Note that both the ESPRIT-based scheme and deep CNN-based scheme need T time slots, which is independent of the numbers of the antennas, served users, and RF chains. Since both the adaptive CS-based scheme and the hierarchical beam training scheme are based on hierarchical codebooks, they have the same overhead, which is larger than that of partial beam sweeping for a small N_T , and vice versa.

6 Conclusion and open issues

In this article, we have provided an overview on CSI acquisition for mmWave massive MIMO, including beam training and channel estimation, with focus on the traditional and the ML-based approaches. The in-depth review along with a comprehensive performance comparison demonstrates the promising prospect of mmWave massive MIMO communications from theory to practice. However, there are still many open issues for future research. Here are some of them.

6.1 CSI acquisition for mobile mmWave massive MIMO

Since the mmWave massive MIMO uses highly directional beams to compensate the path loss, the CSI acquisition and tracking are critical for reliable communications, especially in the mobile scenario, e.g., communications with ground vehicles or UAVs. How to efficiently manage the beams and design the channel tracking algorithms in case of channel blockage is an open issue.

6.2 CSI acquisition for RIS-assisted mmWave massive MIMO

The RIS introduced in Section 2 can improve the signal coverage of mmWave massive MIMO by creating additional channel links. The beam training and channel estimation for RIS-assisted mmWave massive MIMO worth further investigation, especially in the dynamic wireless environment with multiple RISs. The low-cost and passive characteristics of RIS and hardware constraints should be taken into account.

6.3 CSI acquisition based on RL

RL can interactively learn how to take a good action based on the reward and does not rely on the labeled training dataset, which well matches with the mmWave channel with short coherence time. How to effectively integrate RL in the CSI acquisition, i.e., conceive smart beam training strategies or intelligent channel estimation schemes, deserves more studies.

6.4 Advanced CSI acquisition techniques

Although the beam training and channel estimation are presented in two separate categories in this article, they are not independent to each other. Sometimes we may combine them together, e.g., first using beam training benefit from directional transmission gain and then using advanced estimation methods coming from channel estimation category for flexible parameter estimation. More efficient and intelligent CSI acquisition techniques are interesting and worth future study.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 62071116, 61960206005).

References

- 1 Heath R W, Gonzalez-Prelcic N, Rangan S, et al. An overview of signal processing techniques for millimeter wave MIMO systems. *IEEE J Sel Top Signal Process*, 2016, 10: 436–453
- 2 Xiao M, Mumtaz S, Huang Y M, et al. Millimeter wave communications for future mobile networks. *IEEE J Sel Areas Commun*, 2017, 35: 1909–1935
- 3 Yang X, Matthaiou M, Yang J, et al. Hardware-constrained millimeter-wave systems for 5G: challenges, opportunities, and solutions. *IEEE Commun Mag*, 2019, 57: 44–50
- 4 Zhang J A, Huang X J, Dyadyuk V, et al. Massive hybrid antenna array for millimeter-wave cellular communications. *IEEE Wireless Commun*, 2015, 22: 79–87
- 5 Tse D, Viswanath P. *Fundamentals of Wireless Communication*. Cambridge: Cambridge University Press, 2007
- 6 Rappaport T S, Sun S, Mayzus R, et al. Millimeter wave mobile communications for 5G cellular: it will work! *IEEE Access*, 2013, 1: 335–349
- 7 Sayeed A, Behdad N. Continuous aperture phased MIMO: basic theory and applications. In: *Proceedings of Allerton Conference on Communication, Control, and Computing*, Illinois, 2010. 1196–1203
- 8 Schniter P, Sayeed A. Channel estimation and precoder design for millimeter-wave communications: the sparse way. In: *Proceedings of Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, 2014. 273–277
- 9 Ma W Y, Qi C H. Beam-space channel estimation for millimeter wave massive MIMO system with hybrid precoding and combining. *IEEE Trans Signal Process*, 2018, 66: 4839–4853
- 10 Rappaport T S, MacCartney G R, Samimi M K, et al. Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design. *IEEE Trans Commun*, 2015, 63: 3029–3056
- 11 Jaeckel S, Raschkowski L, Borner K, et al. QuaDRiGa: a 3-D multi-cell channel model with time evolution for enabling virtual field trials. *IEEE Trans Antenna Propagat*, 2014, 62: 3242–3256
- 12 Wang B L, Gao F F, Jin S, et al. Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems. *IEEE Trans Signal Process*, 2018, 66: 3393–3406
- 13 Wang B L, Jian M N, Gao F F, et al. Beam squint and channel estimation for wideband mmWave massive MIMO-OFDM systems. *IEEE Trans Signal Process*, 2019, 67: 5893–5908
- 14 El Ayach O, Heath R W, Abu-Surra S, et al. Low complexity precoding for large millimeter wave MIMO systems. In: *Proceedings of IEEE International Conference on Communications*, Ottawa, 2012. 3724–3729
- 15 Mendez-Rial R, Rusu C, Gonzalez-Prelcic N, et al. Hybrid MIMO architectures for millimeter wave communications: phase shifters or switches? *IEEE Access*, 2016, 4: 247–267
- 16 Yu X H, Zhang J, Letaief K B. A hardware-efficient analog network structure for hybrid precoding in millimeter wave systems. *IEEE J Sel Top Signal Process*, 2018, 12: 282–297
- 17 Hu S, Rusek F, Edfors O. The potential of using large antenna arrays on intelligent surfaces. In: *Proceedings of IEEE Vehicular Technology Conference*, Sydney, 2017. 1–6
- 18 Basar E, Di Renzo M, de Rosny J, et al. Wireless communications through reconfigurable intelligent surfaces. *IEEE Access*, 2019, 7: 116753
- 19 Di Renzo M, Zappone A, Debbah M, et al. Smart radio environments empowered by reconfigurable intelligent surfaces: how it works, state of research, and the road ahead. *IEEE J Sel Areas Commun*, 2020, 38: 2450–2525
- 20 Dai L L, Wang B C, Wang M, et al. Reconfigurable intelligent surface-based wireless communications: antenna design, prototyping, and experimental results. *IEEE Access*, 2020, 8: 45913–45923
- 21 Choi J, Va V, Gonzalez-Prelcic N, et al. Millimeter-wave vehicular communication to support massive automotive sensing. *IEEE Commun Mag*, 2016, 54: 160–167
- 22 Xiao Z Y, Xia P F, Xia X G. Enabling UAV cellular with millimeter-wave communication: potentials and approaches. *IEEE Commun Mag*, 2016, 54: 66–73

- 23 Wang P L, Fang J, Duan H P, et al. Compressed channel estimation for intelligent reflecting surface-assisted millimeter wave systems. *IEEE Signal Process Lett*, 2020, 27: 905–909
- 24 Zhang J M, Qi C H, Li P, et al. Channel estimation for reconfigurable intelligent surface aided massive MIMO system. In: *Proceedings of IEEE International Workshop on Signal Processing Advances in Wireless Communications*, Atlanta, 2020. 1–5
- 25 Wan Z W, Gao Z, Gao F F, et al. Terahertz massive MIMO with holographic reconfigurable intelligent surfaces. *IEEE Trans Commun*, 2021. doi: 10.1109/TCOMM.2021.3064949
- 26 Liu C S, Li M, Hanly S V, et al. Millimeter wave beam alignment: large deviations analysis and design insights. *IEEE J Sel Areas Commun*, 2017. doi: 10.1109/JSAC.2017.2699360
- 27 Wang J Y, Lan Z, Pyo C W, et al. Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems. *IEEE J Sel Areas Commun*, 2009, 27: 1390–1399
- 28 Song J, Choi J, Love D J. Common codebook millimeter wave beam design: designing beams for both sounding and communication with uniform planar arrays. *IEEE Trans Commun*, 2017, 65: 1859–1872
- 29 Xiao Z Y, Xia P F, Xia X G. Codebook design for millimeter-wave channel estimation with hybrid precoding structure. *IEEE Trans Wireless Commun*, 2017, 16: 141–153
- 30 Chen K J, Qi C H, Li G Y. Two-step codeword design for millimeter wave massive MIMO systems with quantized phase shifters. *IEEE Trans Signal Process*, 2020, 68: 170–180
- 31 Giordani M, Polese M, Roy A, et al. A tutorial on beam management for 3GPP NR at mmWave frequencies. *IEEE Commun Surv Tut*, 2019, 21: 173–196
- 32 Hofstetter E, DeLong D. Detection and parameter estimation in an amplitude-comparison monopulse radar. *IEEE Trans Inform Theor*, 1969, 15: 22–30
- 33 Li Z H, Qi C H, Li G Y. Low-complexity multicast beamforming for millimeter wave communications. *IEEE Trans Veh Technol*, 2020, 69: 12317–12320
- 34 Sun X Y, Qi C H, Li G Y. Beam training and allocation for multiuser millimeter wave massive MIMO systems. *IEEE Trans Wireless Commun*, 2019, 18: 1041–1053
- 35 Li M, Liu C S, Hanly S V, et al. Explore and eliminate: optimized two-stage search for millimeter-wave beam alignment. *IEEE Trans Wireless Commun*, 2019, 18: 4379–4393
- 36 Zhu D L, Choi J, Heath R W. Auxiliary beam pair enabled AoD and AoA estimation in closed-loop large-scale millimeter-wave MIMO systems. *IEEE Trans Wireless Commun*, 2017, 16: 4770–4785
- 37 Zhu D L, Choi J, Cheng Q, et al. High-resolution angle tracking for mobile wideband millimeter-wave systems with antenna array calibration. *IEEE Trans Wireless Commun*, 2018, 17: 7173–7189
- 38 Noh S, Zoltowski M D, Love D J. Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment. *IEEE Trans Wireless Commun*, 2017, 16: 5689–5701
- 39 Xiao Z Y, He T, Xia P F, et al. Hierarchical codebook design for beamforming training in millimeter-wave communication. *IEEE Trans Wireless Commun*, 2016, 15: 3380–3392
- 40 Zhang J J, Huang Y M, Shi Q J, et al. Codebook design for beam alignment in millimeter wave communication systems. *IEEE Trans Commun*, 2017, 65: 4980–4995
- 41 Lin C, Li G Y, Wang L. Subarray-based coordinated beamforming training for mmWave and sub-THz communications. *IEEE J Sel Areas Commun*, 2017, 35: 2115–2126
- 42 Sun X Y, Qi C H. Codeword selection and hybrid precoding for multiuser millimeter-wave massive MIMO systems. *IEEE Commun Lett*, 2019, 23: 386–389
- 43 Sun Y H, Qi C H. Weighted sum-rate maximization for analog beamforming and combining in millimeter wave massive MIMO communications. *IEEE Commun Lett*, 2017, 21: 1883–1886
- 44 Chen K J, Qi C H. Beam training based on dynamic hierarchical codebook for millimeter wave massive MIMO. *IEEE Commun Lett*, 2019, 23: 132–135
- 45 Chen K J, Qi C H, Dobre O A, et al. Simultaneous multiuser beam training using adaptive hierarchical codebook for mmwave massive MIMO. In: *Proceedings of IEEE Global Communications Conference*, Hawaii, 2019. 1–6
- 46 Qi C H, Chen K J, Dobre O A, et al. Hierarchical codebook-based multiuser beam training for millimeter wave massive MIMO. *IEEE Trans Wireless Commun*, 2020, 19: 8142–8152
- 47 Ma W Y, Qi C H, Li G Y. Machine learning for beam alignment in millimeter wave massive MIMO. *IEEE Wireless Commun Lett*, 2020, 9: 875–878
- 48 Echigo H, Cao Y W, Bouazizi M, et al. A deep learning-based low overhead beam selection in mmWave communications. *IEEE Trans Veh Technol*, 2021, 70: 682–691
- 49 Lin C H, Kao W C, Zhan S Q, et al. BsNet: a deep learning-based beam selection method for mmwave communications. In: *Proceedings of IEEE Vehicular Technology Conference*, Hawaii, 2019. 1–6
- 50 Satyanarayana K, El-Hajjar M, Mourad A A M, et al. Deep learning aided fingerprint-based beam alignment for mmWave vehicular communication. *IEEE Trans Veh Technol*, 2019, 68: 10858–10871
- 51 Rezaie S, Manchón C N, de Carvalho E. Location-and orientation-aided millimeter wave beam selection using deep learning. In: *Proceedings of IEEE International Conference on Communications*, Dublin, 2020. 1–6
- 52 Wang Y Y, Klautau A, Ribero M, et al. mmWave vehicular beam selection with situational awareness using machine learning. *IEEE Access*, 2019, 7: 87479–87493
- 53 Alrabeiah M, Alkhateeb A. Deep learning for mmWave beam and blockage prediction using sub-6 GHz channels. *IEEE Trans Commun*, 2020, 68: 5504–5518
- 54 Klautau A, Gonzalez-Prelcic N, Heath R W. LIDAR data for deep learning-based mmWave beam-selection. *IEEE Wireless Commun Lett*, 2019, 8: 909–912
- 55 Hashemi M, Sabharwal A, Koksals C E, et al. Efficient beam alignment in millimeter wave systems using contextual bandits. In: *Proceedings of IEEE International Conference on Computer Communications*, Hawaii, 2018. 2393–2401
- 56 Va V, Shimizu T, Bansal G, et al. Online learning for position-aided millimeter wave beam training. *IEEE Access*, 2019, 7: 30507–30526
- 57 Cheng M, Wang J B, Wang J Y, et al. A fast beam searching scheme in mmWave communications for high-speed trains. In: *Proceedings of IEEE International Conference on Communications*, Shanghai, 2019. 1–6
- 58 Booth M B, Suresh V, Michelusi N, et al. Multi-armed bandit beam alignment and tracking for mobile millimeter wave communications. *IEEE Commun Lett*, 2019, 23: 1244–1248
- 59 Hussain M, Michelusi N. Second-best beam-alignment via Bayesian multi-armed bandits. In: *Proceedings of IEEE Global*

- Communications Conference, Hawaii, 2019. 1–6
- 60 Sim G H, Klos S, Asadi A, et al. An online context-aware machine learning algorithm for 5G mmWave vehicular communications. *IEEE/ACM Trans Netw*, 2018, 26: 2487–2500
 - 61 Wu W, Cheng N, Zhang N, et al. Fast mmWave beam alignment via correlated bandit learning. *IEEE Trans Wireless Commun*, 2019, 18: 5894–5908
 - 62 Kim S, Kwon G, Park H. Q-learning-based low complexity beam tracking for mmWave beamforming system. In: *Proceedings of International Conference on Information and Communication Technology Convergence*, Jeju Island, 2020. 1451–1455
 - 63 Zhang J J, Huang Y M, Wang J H, et al. Intelligent beam training for millimeter-wave communications via deep reinforcement learning. In: *Proceedings of IEEE Global Communications Conference*, Hawaii, 2019. 1–7
 - 64 Alkhateeb A, El Ayach O, Leus G, et al. Channel estimation and hybrid precoding for millimeter wave cellular systems. *IEEE J Sel Top Signal Process*, 2014, 8: 831–846
 - 65 Alkhateeb A, Leus G, Heath R W. Compressed sensing based multi-user millimeter wave systems: how many measurements are needed? In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Dubai, 2015. 2909–2913
 - 66 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
 - 67 Rodriguez-Fernandez J, Gonzalez-Prelicic N, Venugopal K, et al. Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems. *IEEE Trans Wireless Commun*, 2018, 17: 2946–2960
 - 68 Tao J, Qi C H, Huang Y M. Regularized multipath matching pursuit for sparse channel estimation in millimeter wave massive MIMO system. *IEEE Wireless Commun Lett*, 2019, 8: 169–172
 - 69 Gao X Y, Dai L L, Han S F, et al. Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array. *IEEE Trans Wireless Commun*, 2017, 16: 6010–6021
 - 70 Gao X Y, Dai L L, Han S F, et al. Beamspace channel estimation for 3D lens-based millimeter-wave massive MIMO systems. In: *Proceedings of International Conference on Wireless Communications and Signal Processing*, Yangzhou, 2016. 1–5
 - 71 Ma W Y, Qi C H. Channel estimation for 3-D lens millimeter wave massive MIMO system. *IEEE Commun Lett*, 2017, 21: 2045–2048
 - 72 Hu C, Dai L L, Mir T, et al. Super-resolution channel estimation for MmWave massive MIMO with hybrid precoding. *IEEE Trans Veh Technol*, 2018, 67: 8954–8958
 - 73 Tsai Y, Zheng L, Wang X D. Millimeter-wave beamformed full-dimensional MIMO channel estimation based on atomic norm minimization. *IEEE Trans Commun*, 2018, 66: 6150–6163
 - 74 Anjinappa C K, Zhou Y, Yapici Y, et al. Channel estimation in mmWave hybrid MIMO system via off-grid Dirichlet kernels. In: *Proceedings of IEEE Global Communications Conference*, Hawaii, 2019. 1–6
 - 75 Gao Z, Hu C, Dai L L, et al. Channel estimation for millimeter-wave massive MIMO with hybrid precoding over frequency-selective fading channels. *IEEE Commun Lett*, 2016, 20: 1259–1262
 - 76 Liao A W, Gao Z, Wu Y P, et al. 2D unitary ESPRIT based super-resolution channel estimation for millimeter-wave massive MIMO with hybrid precoding. *IEEE Access*, 2017, 5: 24747–24757
 - 77 Liao A W, Gao Z, Wang D M, et al. Terahertz ultra-massive MIMO-based aeronautical communications in space-air-ground integrated networks. *IEEE J Sel Area Commun*, 2021. doi: 10.1109/JSAC.2021.3071834
 - 78 Zhang J S, Haardt M. Channel estimation and training design for hybrid multi-carrier mmWave massive MIMO systems: the beamspace ESPRIT approach. In: *Proceedings of European Signal Processing Conference*, Kos, 2017. 385–389
 - 79 Zhang J S, Haardt M. Channel estimation for hybrid multi-carrier mmwave MIMO systems using three-dimensional unitary esprit in DFT beamspace. In: *Proceedings of International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, Curacao, 2017. 1–5
 - 80 Ma W Y, Qi C H, Li G Y. High-resolution channel estimation for frequency-selective mmWave massive MIMO systems. *IEEE Trans Wireless Commun*, 2020, 19: 3517–3529
 - 81 Guo Z Y, Wang X D, Heng W. Millimeter-wave channel estimation based on 2-D beamspace MUSIC method. *IEEE Trans Wireless Commun*, 2017, 16: 5384–5394
 - 82 Wu J, Heng W, Hu J M, et al. Long-term channel estimation for mm-Wave hybrid beamforming systems of coherent signals based on 2-D MUSIC algorithm. In: *Proceedings of IEEE Vehicular Technology Conference*, Illinois, 2018. 27–30
 - 83 He H T, Wen C K, Jin S, et al. Deep learning-based channel estimation for beamspace mmWave massive MIMO systems. *IEEE Wireless Commun Lett*, 2018, 7: 852–855
 - 84 Wei Y, Zhao M M, Zhao M J, et al. An AMP-based network with deep residual learning for mmWave beamspace channel estimation. *IEEE Wireless Commun Lett*, 2019, 8: 1289–1292
 - 85 Zhang Y H, Mu Y F, Liu Y, et al. Deep learning-based beamspace channel estimation in mmWave massive MIMO systems. *IEEE Wireless Commun Lett*, 2020, 9: 2212–2215
 - 86 Wei X H, Hu C, Dai L L. Deep learning for beamspace channel estimation in millimeter-Wave massive MIMO systems. *IEEE Trans Commun*, 2021, 69: 182–193
 - 87 Jin Y, Zhang J Y, Jin S, et al. Channel estimation for cell-free mmWave massive MIMO through deep learning. *IEEE Trans Veh Technol*, 2019, 68: 10325–10329
 - 88 Liang P Z, Fan J C, Shen W H, et al. Deep learning and compressive sensing-based CSI feedback in FDD massive MIMO systems. *IEEE Trans Veh Technol*, 2020, 69: 9217–9222
 - 89 Moon S, Kim H, Hwang I. Deep learning-based channel estimation and tracking for millimeter-wave vehicular communications. *J Commun Netw*, 2020, 22: 177–184
 - 90 Ma W Y, Qi C H, Zhang Z C, et al. Sparse channel estimation and hybrid precoding using deep learning for millimeter wave massive MIMO. *IEEE Trans Commun*, 2020, 68: 2838–2849
 - 91 Dong P H, Zhang H, Li G Y, et al. Deep CNN-based channel estimation for mmWave massive MIMO systems. *IEEE J Sel Top Signal Process*, 2019, 13: 989–1000
 - 92 Ye H Y, Gao F F, Qian J, et al. Deep learning-based denoise network for CSI feedback in FDD massive MIMO systems. *IEEE Commun Lett*, 2020, 24: 1742–1746