



HAL
open science

Codebook-Based Solutions for Reconfigurable Intelligent Surfaces and Their Open Challenges

Jiancheng An, Chao Xu, Qingqing Wu, Derrick Wing Kwan Ng, Marco Di Renzo, Chau Yuen, Lajos Hanzo

► **To cite this version:**

Jiancheng An, Chao Xu, Qingqing Wu, Derrick Wing Kwan Ng, Marco Di Renzo, et al.. Codebook-Based Solutions for Reconfigurable Intelligent Surfaces and Their Open Challenges. *IEEE Wireless Communications*, 2024, 31 (2), pp.134 - 141. hal-04734213

HAL Id: hal-04734213

<https://hal.science/hal-04734213v1>

Submitted on 13 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Codebook-Based Solutions for Reconfigurable Intelligent Surfaces and Their Open Challenges

Jiancheng An, Chao Xu, Qingqing Wu, Derrick Wing Kwan Ng, Marco Di Renzo, Chau Yuen, and Lajos Hanzo

Abstract—Reconfigurable intelligent surfaces (RIS) is a revolutionary technology to cost-effectively improve the performance of wireless networks. We first review the existing framework of channel estimation and passive beamforming (CE & PBF) in RIS-assisted communication systems. To reduce the excessive pilot signaling overhead and implementation complexity of the CE & PBF framework, we conceive a codebook-based framework to strike flexible tradeoffs between communication performance and signaling overhead. Moreover, we provide useful insights into the codebook design and learning mechanisms of the RIS reflection pattern. Finally, we analyze the scalability of the proposed framework by flexibly adapting the training overhead to the specified quality-of-service requirements and then elaborate on its appealing advantages over the existing CE & PBF approaches. It is shown that our novel codebook-based framework can be beneficially applied to all RIS-assisted scenarios and avoids the curse of model dependency faced by its existing counterparts, thus constituting a competitive solution for practical RIS-assisted communication systems.

I. INTRODUCTION

Over the past decades, channel fading has been considered one of the most unfavorable factors limiting the performance of wireless communication systems. In particular, the signal attenuation caused by large-scale fading – including the distance-dependent path loss and shadow fading – constrains both the service coverage range and communication quality, especially for the emerging millimeter-wave (mmWave) and TeraHertz (THz) bands. Although both the fine-grained massive multiple-input multiple-output (MIMO) beamforming and the classic amplify- or decode-and-forward relaying schemes constitute effective solutions, they generally result in extra hardware costs and exceedingly high power consumption. Furthermore, the small-scale fading leads to rapid fluctuation of the received signal power. In general, adaptive modulation and coding as well as resource allocation are adopted for mitigating the impact of small-scale fading. However, these classic designs are typically implemented at the transmitters or receivers [1]. As a consequence, the wireless channel is still treated as a random “black box” and handled proactively.

Recently, *Reconfigurable Intelligent Surfaces (RIS)* have emerged as a promising technology for ameliorating wireless

propagation environments in real-time owing to the quantum leap in research advances on programmable metasurfaces [2]. Specifically, an RIS is a planar array comprising a large number of low-cost passive reflecting elements, each of which is capable of independently imposing a phase shift on the incident signals [3]. From the perspective of large-scale fading, an RIS is capable of mitigating the path loss by partially reconstructing the electromagnetic waves diffused in space without requiring additional energy, which is a substantial benefit over conventional relaying and active beamforming. As for small-scale fading, an RIS is capable of advantageously manipulating the electromagnetic waves to establish constructive signal superposition, thus creating a more reliable communication channel. Alternatively, an RIS may be optimized to increase the effective rank of a communication link for acquiring much-needed spatial multiplexing gain [4]. In stark contrast to existing active transceiver-based link adaptive technologies, an RIS offers a new design degrees-of-freedom (DoF) for enhancing the performance attained, ushering in a new era of smart programmable wireless environments.

In order to unlock the full potential of RISs, channel estimation (CE) and passive beamforming (PBF) constitute a pair of crucial challenges. On the one hand, a large number of pilots are required for probing the numerous reflected links [5]. Although advanced CE techniques have been conceived for reducing the pilot signaling overhead by exploiting the associated channel characteristics, the pilot overhead proportional to the number of RIS elements remains prohibitive [6]. On the other hand, most of the existing PBF designs pursue optimal performance by jointly optimizing the RIS and the active components in the network, which is nontrivial and requires bespoke solutions for different scenarios and applications [7]–[9]. Therefore, the scalability and complexity of the existing CE & PBF frameworks constitute a stumbling block for the large-scale deployment of RISs, not to mention their heavy dependence on channel estimation [10].

Against the above background, this article firstly offers a comprehensive review of the existing modularized CE & PBF designs. Then, we discuss the critical issues of excessive pilot overhead and high implementation complexity. In order to improve both energy efficiency and scalability, we then propose a novel codebook-based framework for RIS-assisted communications, which efficiently addresses these challenges. Specifically, by first generating an appropriate RIS reflection pattern (RP) codebook offline, one only has to concentrate on the end-to-end composite channel estimation and the transmitter design, which can be directly inherited from current communication systems. Based on the labeled codebook, an appropriate learning-based approach is applied for determining the optimized RP for a given set of quality-of-service (QoS)

J. An and C. Yuen are with the Engineering Product Development (EPD) Pillar, Singapore University of Technology and Design, Singapore 487372 (e-mail: jiancheng_an@163.com; yuenchau@sutd.edu.sg). C. Xu and L. Hanzo are with the School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: cx1g08@soton.ac.uk; lh@ecs.soton.ac.uk). Q. Wu is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: wu.qq1010@gmail.com). D. W. K. Ng is with the School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: w.k.ng@unsw.edu.au). M. Di Renzo is with Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des Signaux et Systèmes, 3 Rue Joliot-Curie, 91192 Gif-sur-Yvette, France (e-mail: marco.direnzo@universite-paris-saclay.fr).

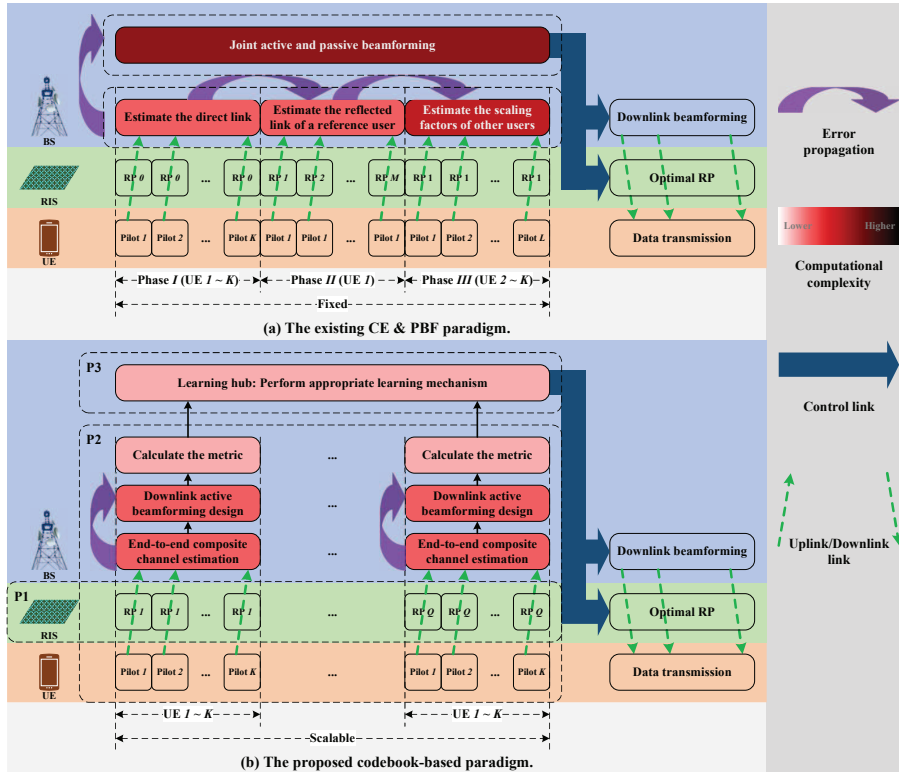


Fig. 1. Protocol comparison between the existing CE & PBF framework and the proposed codebook-based framework.

requirements. Moreover, we design both the codebook as well as the relevant learning mechanisms and demonstrate the compelling benefits of the proposed codebook-based framework over the state-of-the-art.

II. EXISTING CE & PBF FRAMEWORKS

In this section, we first survey the state-of-the-art CE & PBF designs conceived for RIS-assisted communication systems and then pose several critical questions concerning the existing RIS-oriented frameworks.

A. Channel Estimation (CE)

1) **Instantaneous CSI:** Channel state information (CSI) acquisition is extremely challenging for RIS-assisted systems because of the low-complexity passive nature of the RIS architecture. Specifically, due to the lack of sensing and advanced signal processing capability, it is unrealistic to separately estimate the base station (BS)-RIS and RIS-user equipment (UE) channels by harnessing traditional pilot-based techniques of active components. Instead, we opt for estimating the cascaded BS-RIS-UE channels [5]. Specifically, upon repeatedly transmitting the same pilot and evaluating the received signal under different RPs, all the reflected channels can be obtained after collecting sufficient observations. Unfortunately, the number of pilots required is much higher than that of RIS elements, which is prohibitive for practical applications.

For the sake of reducing the pilot overhead, advanced CE techniques may be conceived by exploiting the specific channel characteristics. In [6], a novel three-phase method was proposed for exploiting the commonality of the BS-RIS link among multiple UEs to reduce the pilot overhead. For

the sake of illustration, Fig. 1(a) portrays this three-phase CE approach. In *Phase I* consisting of K slots, all M RIS elements are switched off (i.e., RP 0 in Fig. 1(a)) for estimating the direct channels associated with K UEs. In *Phase II* of Fig. 1(a), all the reflecting elements are switched on and merely a reference UE (assumed to be UE 1) transmits pilot symbols to the BS. By leveraging the results of *Phase I*, the reflected channels associated with the reference UE can be estimated within M time slots. Finally, in *Phase III* of Fig. 1(a), all other $(K - 1)$ UEs transmit their pilot symbols to the BS. Since the reflected channels of the K UEs share the same BS-RIS link, the scaling factors between the different reflected channels are estimated instead of acquiring the full CSI, which requires at least $L = \max(K - 1, \lceil (K - 1)M/N \rceil)$ slots with N denoting the number of antennas at the BS. Interested readers may refer to [6] for the specific pilot design in *Phase III*. In [9], the different channel coherence time of the BS-RIS and RIS-UE links was exploited to further reduce the pilot overhead. Additionally, the exploitation of beamspace sparsity of mmWave communications and the spatial correlation between adjacent RIS elements can also mitigate the computational complexity as well as the pilot overhead at the cost of some performance erosion [3].

While pilot reduction has gradually become the prime objective of CE approaches conceived for RIS-assisted systems, their resultant performance erosion has been somewhat ignored [6]. Additionally, PBF has to be specifically tailored to accommodate the outputs of different CE approaches. In a nutshell, the existing modularized framework does not consider the intricate relationship between CE and PBF, it rather concentrates on the autonomous objective of each

stage, which generally fails to optimize the holistic system performance at a given pilot overhead. For example, if there are only limited feasible RPs for M reflecting elements due to practical hardware restrictions, it may become unnecessary to increase the pilot overhead proportionately with M . In contrast to the CE approaches that disregard the subsequent PBF requirements, it is a promising option to take into account the practical hardware constraints and ascertain what pivotal information is required for performing PBF.

2) **Statistical CSI:** In order to reduce the number of pilots required by RIS-assisted systems, statistical CSI-based methods have also attracted substantial interests. The authors of [11] maximized the sum rate of an RIS-assisted multi-user MIMO system by only exploiting the statistical location information of mobile users. Specifically, in the mmWave band, the channels of B5G/6G systems tend to largely rely on the line-of-sight (LoS) component, which allows aligning the RP with the pure LoS component [12]. As a benefit, the statistical CSI-based method significantly reduces the pilot overhead required for CSI acquisition; secondly, it also dispenses with frequent RIS reconfiguration as statistics vary slowly in time.

Although the statistical CSI-based method beneficially reduces the pilot overhead, it also has its disadvantages. Again, the statistical CSI-based method only performs well when strong LoS components are dominated. When the BS (UE) - RIS link is blocked, the statistical CSI-based method suffers from severe performance degradation. Furthermore, the statistical CSI-based scheme actually operates in a semi-static mode, thus reducing the DoF in optimizing the environment and failing to exploit the full potential of RIS. Therefore, striking a beneficial tradeoff between the reduced pilot overhead relying on statistical CSI and the excellent performance of instantaneous CSI requires future research.

B. Passive Beamforming (PBF)

PBF is another challenging task in RIS-assisted communications, which generally involves non-convex objective functions (OF) and constraints such as the discrete phase shifts [3]. On the other hand, as shown in Fig. 1(a), the PBF at the RIS usually has to be designed in conjunction with the active components in the wireless network for gaining optimal performance [8]. A practical approach is to first relax the non-convex modulus constraints by applying, for example, successive convex approximation (SCA) or semi-definite relaxation (SDR), and then map the solution attained into its nearest legitimate solution [7]. Alternatively, the heuristic alternating optimization (AO) technique is capable of obtaining a sub-optimal solution with polynomial-time complexity by iteratively optimizing the active and passive components assuming that the other is fixed [8]. Indeed, the PBF design can be further simplified by optimizing the phase shifts of one RIS element each time, while fixing that of the rest until convergence is achieved. Upon leveraging the aforementioned optimization philosophy, RISs have already been shown to be of great potential in enhancing the channel quality, suppressing co-channel interference, as well as preventing eavesdropping [13], [14].

To elaborate, the quadratic power scaling law shows that the received power increases 6 dB when doubling the number of reflecting elements in a far-field setting [8]. However, the conclusions of [8] are drawn based on the idealized assumption that all the reflected channels associated with different RIS elements are statistically independent, which requires adjacent RIS elements to be sufficiently far apart compared to the wavelength. In particular, the RIS's performance would be degraded by the channel's correlation and mutual coupling, which has to be quantitatively characterized by future research [10]. The calculus-based performance analysis is expected to be a powerful technique for characterizing the family of holographic RISs having infinitely densely deployed elements [2]. In summary, the performance characterization of RISs depends on the underlying channel model, which requires further investigations and should also be verified by experimental measurements [1].

III. THE CODEBOOK-BASED SOLUTION

A. The Codebook-Based Protocol

In this section, we develop a novel codebook-based framework for mitigating both the pilot overhead and the complexity. For the sake of illustration, Fig. 1(b) portrays the proposed codebook-based protocol. Instead of directly optimizing the RIS phase shifts, the codebook-based solution scans a limited set of RP candidates and learns the optimal RP based on the feedback from the controller. Thus, it becomes unnecessary to explicitly estimate the RIS-BS/UE channels. The detailed phases (\mathbf{P}) are summarized as follows:

- P1:** RIS appropriately extracts a codebook of RPs from the universal solution set taking into account the constraints imposed by the practical hardware structure, where the cardinality of the codebook is Q , as shown in Fig. 1(b).
- P2:** For each RP in the codebook, all the UEs send uplink pilots and the BS estimates the end-to-end composite channel without explicitly considering the existence of the RIS. The number of pilots within each subframe (i.e., K in Fig. 1(b)) depends on the total number of antennas at the UE side. Following this, the BS performs only the transmit beamformer optimization and evaluates the associated OF.
- P3:** After repeating **P2** for all RPs in the codebook, we now obtain Q labeled training samples, based on which the BS executes the appropriate learning algorithm. Finally, for a given input of QoS target, the learning system would return an optimal RP for downlink data transmission. Moreover, new observations can be employed during the data transmission phase to further improve the learning accuracy.

Before proceeding further, we first contrast the proposed codebook-based framework with existing CE & PBF designs. First of all, the codebook-based solution estimates the end-to-end composite channel at each time slot, which is exactly what both the transmitter and receiver need and thus avoids the potential pilot wastage of traditional approaches. The rest of the design problem is to adjust the size of the codebook according to the specific scenarios for striking beneficial tradeoffs

TABLE I
A SURVEY OF EXISTING CE & PBF FRAMEWORK AND CODEBOOK-BASED SOLUTIONS FOR RIS-ASSISTED COMMUNICATION SYSTEMS.

Scheme	Reference	Scenario	Metric	Overhead	Complexity	Key features	
CE & PBF	Statistical CSI-based PBF	[12]	MISO	Ergodic SE	$O(1)$	###	Configure RIS based on the statistical CSI
		[11]	MU-MIMO	Sum-Rate	$O(1)$	###	Configure RIS based on the position information of multiple users
	Instantaneous CSI acquisition	[5]	MU-MISO	CRLB	$O(MK)$	###	Exploit DFT-based RP to improve channel estimation accuracy
		[6]	MU-MISO	Overhead	$O(M + K + L)$	###	Exploit the common reflected channel of multiple users to reduce pilot overhead
	Passive beamforming	[14]	MISO	Secrecy rate	###	$O(N^3 + M^{3.5})$	Optimize RP to maximize the secrecy rate of the legitimate user
		[9]	MU-MISO	Weighted sum-rate	###	$O(KN^3 + KNM)$	Exploit two-timescale property to optimize RP and transmit beamforming
		[8]	MU-MISO	Transmit power	###	$O(KN^3 + KM^{3.5})$	Optimize RP to minimize the transmit power at the BS
		[7]	MU-MISO	Energy efficiency	###	$O(K^4 + M^4)$	Optimize RP to maximize the system energy efficiency
Codebook-based solution	[4]	MIMO	Capacity	$O(QN_s)$	$O(QN^3 + QNN_s^2)$	Select the optimal RP from a randomly-generated codebook having discrete phase shift	
	[13]	MU-MISO	Transmit power	$O(QK)$	$O(QNK^3)$	Select the optimal RP from a randomly-generated codebook having continuous phase shift	

SE: Spectral efficiency K: The number of UEs Q: The codebook size
CRLB: Cramer-Rao lower bound M: The number of RIS elements N_r : The number of receive antennas at the UE
L: $L = \max(K-1, \lceil (K-1)M/N \rceil)$ N: The number of transmit antennas at the BS

between the pilot overhead and the ultimate communication performance. As for the PBF design, the codebook is generated directly based on the legitimate RP set, which implicitly takes all the adverse practical factors into account during its learning process. By contrast, existing designs aim for the optimal solution under various idealized assumptions, e.g., the perfect CSI and flawless element tuning, which would result in severe mismatch errors, when considering the hardware imperfections. Although the CE & PBF framework can also take any adverse constraints into account at the beginning of the problem formulation, it has to be regularly updated to cater for various scenarios. For the sake of illustration, we boldly and explicitly contrast the codebook-based solution to its existing CE & PBF counterparts in Table I.

B. Benefits of the Codebook-Based Framework

Next, we elaborate on the advantages of the proposed codebook-based framework over the existing CE & PBF framework from the following six aspects.

1) **Backward Compatibility:** The proposed codebook-based framework can be directly applied to the conventional communication architecture for optimizing the RIS benefits [4]. The only additional configuration requirement is to appropriately increase the training overhead. During each training period, the BS repeats the end-to-end channel estimation and transmitter training in the same way as the existing framework. By contrast, the cascaded CE & PBF framework tends to replace the entire suite of existing protocols and thus has to be reformulated for different scenarios and setups [7].

2) **Scalable Pilot Overhead:** As shown in Fig. 1(b), the proposed scheme is capable of adjusting the training overhead according to the channel's coherence time so that the pilot overhead can be flexibly adapted to the specific scenario. This is in sharp contrast to the existing CE schemes [5],

[6], which require an excessive pilot signaling overhead for probing a large number of reflected channels. Although the progressive CE schemes of the conventional framework attempt to minimize the required pilot signaling overhead, they remain dependent on the number of reflecting elements, which essentially limits the practical application of RISs [15].

3) **Reduced Computational Complexity:** Compared to the computationally inefficient joint active and passive beamforming design, the codebook-based solution substantially reduces the system complexity by decoupling the optimization of PBF and that of the active elements, as shown in **P2** of Fig. 1(b). Moreover, when considering practical hardware imperfections, the reduced complexity may not necessarily lead to an inferior performance compared to the existing schemes relying on unrealistic idealized assumptions, especially in the case of large-scale deployments.

4) **Reduced Error Propagation:** In the existing CE & PBF frameworks, the successive estimation of different channels, the separate design of CE and PBF, as well as the mismatch between the practical RIS model and its relaxed counterpart will lead to error propagation (see Fig. 1(a)), thereby degrading the expected performance [6]. By contrast, in the codebook-based framework, error propagation only occurs at the active transmit beamformer design due to the imperfect CSI of the end-to-end composite channel, and its negative impact can be further mitigated by applying joint CE and transmit beamformer optimization. Hence, the proposed codebook-based framework substantially mitigates the error propagation compared to the existing CE & PBF design.

5) **Reduced Control Signaling:** The proposed codebook-based solution effectively reduces the capacity requirement of the control link. Assuming B legitimate phase shifts for each RIS element, the existing methods require $M \log_2 B$ signaling bits for configuring M RIS elements. By contrast, the control signaling overhead of our codebook-based solution is

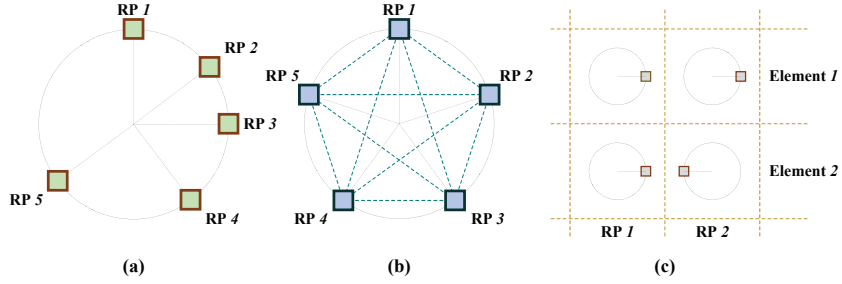


Fig. 2. The schematic of three RP codebooks: (a) Random ($M = 1$, $Q = 5$); (b) SDM ($M = 1$, $Q = 5$); (c) Orthogonal ($M = 2$, $Q = 2$).

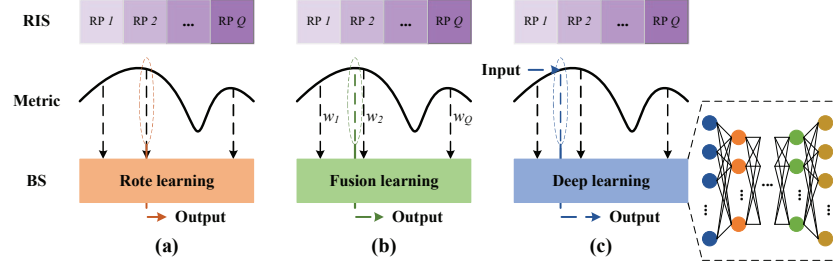


Fig. 3. The schematic of three learning mechanisms: (a) Rote learning; (b) Fusion learning; (c) Deep learning.

only $\log_2 Q$ bits. Therefore, the proposed scheme substantially reduces the backhaul overhead and delay.

6) **Stronger Robustness:** The proposed codebook-based scheme estimates the composite end-to-end channel without considering the individual channels reflected via the RIS. Specifically, the channel correlation, mutual coupling, component aging, and other adverse factors would result in severe performance degradation for the existing CE & PBF solution, which has not been fully characterized [13]. By contrast, all the aforementioned hardware imperfections have been implicitly taken into consideration during the on-site learning process of our codebook-based scheme, while the existing solutions would have to be redesigned or regularly updated for different scenarios. Therefore, the proposed scheme is more robust to hardware imperfections.

In summary, the existing CE & PBF framework can be characterized as “*sensing first and then configuring*”, whereas the proposed codebook-based solution is based on the “*learning from testing*” philosophy.

IV. RESEARCH OPPORTUNITIES AND CHALLENGES

Naturally, the specific generation of the RP codebook and the learning algorithm determine the performance attained, hence they are still under investigation, as discussed below.

A. Codebook Generation

Here we highlight three methods of generating the RP codebook, i.e., **PI** in Fig. 1(b).

1) **Random Codebook:** First of all, let us consider the codebook having Q identical RPs. Thus, we obtain Q observations of the same channel, which only attains a mean square error (MSE) performance gain of $10 \log_{10} Q$ dB for CE. However, by beneficially designing a codebook containing different RPs, one would obtain extra diversity gain. To this end, the most straightforward method is to randomly generate

the RP codebook, where each element in the codebook is randomly selected from the universal set of RPs [4]. For the sake of illustration, Fig. 2(a) shows a randomly generated codebook considering $M = 1$ and $Q = 5$.

2) **Sum Distance Maximized (SDM) Codebook:** It is evident that the randomly generated codebook of RPs does not exploit the limited training overhead effectively. To circumvent this issue, the authors of [13] conceived a heuristic method of generating Q different RPs by maximizing the sum of Euclidean distances between all pairs of RPs, as shown in Fig. 2(b). The SDM codebook has been shown to achieve beneficial received power gains. Nevertheless, since the SDM codebook cannot examine all the legitimate RPs, it only provides moderate gain compared to the random codebook [13]. Therefore, how to generate an appropriate codebook conducive to subsequent learning for adapting to a variety of application scenarios and optimization objectives remains an open question.

3) **Orthogonal Codebook:** In conventional beam control and tracking regimes, employing orthogonal beam codebooks (e.g., the DFT-based steering vector) has become a popular option. Hence in RIS-assisted communication systems, DFT-based RP design has also been employed for improving the estimation accuracy [6], but a broader family of orthogonal codebooks as well as the resultant performance analysis deserves further exploration. Specifically, an orthogonal codebook may be used for searching for the LoS component at a low overhead [4]. In this context Fig. 2(c) shows an orthogonal codebook considering $M = 2$ and $Q = 2$.

Note that in addition to the aforementioned environment-agnostic codebook designs, one could exploit the statistical CSI to design a CSI-aware codebook for further improving the performance at the expense of increasing the backhaul capacity. Furthermore, beam training could also be improved by dividing the potential service zone into several beams in a hierarchical manner [2].

B. Learning Mechanism

Then, a set of three learning mechanisms is introduced to perform PBF based on Q observations of the end-to-end composite channel [4], [13], i.e., **P3** in Fig. 1(b).

1) **Rote Learning**: Fig. 3(a) portrays the most straightforward learning method, namely rote learning [13], which is the most classic codebook-based scheme relying on memorizing multiple observations. Explicitly, the BS chooses the specific RP that performs the best in terms of the OF from the codebook, as usual. Thus, the control signaling overhead of $\log_2 Q$ bits is required to configure the RIS.

2) **Fusion Learning**: Another feasible technique is fusion learning, which synthesizes more reliable statistical inferences from multiple observations. As shown in Fig. 3(b), by weighting and superimposing different RPs in the pre-designed codebook using a set of weighting coefficients w_1, w_2, \dots, w_Q , one may attain a better performance by making each weighting coefficient proportional to the corresponding OF value. Naturally, the resultant performance is heavily dependent on the OF. Specifically, fusion learning will outperform rote learning for a convex optimization problem at the expense of increasing the backhaul capacity from $\log_2 Q$ bits to Q bits [4]. However, it is non-trivial to obtain a unified fusion learning framework suitable for all OFs. As such, the combination with some convex relaxation techniques, e.g., SCA, may be beneficially harnessed, but this requires further investigations.

3) **Machine Learning**: Machine learning (ML) techniques, especially deep learning, can be efficiently applied in our codebook-based framework instead of the aforementioned model-based methods. As shown in Fig. 3(c), the codebook is first labeled upon calculating the OF value for each RP. Following this, the performance metrics and corresponding reflection coefficients are used to train a deep neural network (DNN) as input and output, respectively. Then the desired QoS requirement is used as input and a well-trained DNN is utilized to return an appropriate RP to assist the data transmission. A series of new data can be employed for further training the DNN to improve its performance during the testing phase [1]. In contrast to the above two model-based learning mechanisms, which solve the associated maximization/minimization problems heuristically, the data-driven DNN introduced here is supervised with the aid of labeled samples, which aims for returning an optimized RP for satisfying the user's expected QoS target. Although advanced deep reinforcement learning (DRL) can be employed for updating the RP in real-time, more advanced unsupervised ML techniques capable of completely dispensing with labeled data for solving these challenging optimization problems require further exploration.

V. PERFORMANCE EVALUATION

A. Fundamental Tradeoff: Performance vs. Pilot Overhead

As expected, there is a fundamental performance vs. pilot overhead tradeoff in RIS-assisted systems. Specifically, the RIS performance relies on the CSI accuracy and hence on the pilot overhead. In Table II, we first illustrate the stylized performance versus overhead tradeoff for providing insights into the scalability of the codebook-based framework,

TABLE II
OVERHEAD AND PERFORMANCE COMPARISON BETWEEN CODEBOOK-BASED FRAMEWORK AND ITS EXISTING COUNTERPARTS.

	Reference	Overhead	Power scaling law
Random configuration	[8]	$O(1)$	$O(M)$
Optimal configuration	[8]	$O(M)$	$O(M^2)$
Random codebook	[4]	$O(Q)$	$O(M \ln Q)$
SDM codebook	[13]	$O(Q)$	$O(M^2 \text{sinc}^2(\pi/Q))$

where the rote learning mechanism is adopted. Explicitly, the performance characterized in Table II relies on theoretical evaluation and it is validated in terms of the received power scaling law (i.e., P_r) [4], [13]. Observe from Table II that the optimal RIS performance is characterized in terms of quadratic gain vs. the number of RIS elements [8], but this may only be attained at a proportionately increased training overhead. By contrast, the codebook-based solution is capable of striking a compelling performance vs. overhead trade-off by flexibly adjusting Q . When considering the maximum codebook size of $Q_{\max} = B^M$ with B denoting the number of phase shifts, we have $P_r \propto O(M^2 \log B)$ for the random codebook, which implies that the quadratic power scaling law could be achieved by exhaustively searching through the entire codebook space. Furthermore, the improved SDM codebook is capable of achieving a quadratic gain, but it suffers from some performance erosion determined by the codebook size. More advanced codebooks and learning mechanisms could achieve improved performance despite a limited pilot signaling overhead, but the associated performance evaluation is still an open research area.

B. Case Study

In this section, we provide some numerical results to show the benefits of the codebook-based solution. Given the strict page limit, we only employ the random codebook as well as the rote learning mechanism. In contrast to [13], we consider an RIS-assisted single-input single-output orthogonal frequency division multiplexing (SISO-OFDM) communication system, as shown in Fig. 4(a). All the parameters in our simulations are listed in Fig. 4(b).

First, we study the rate performance of the codebook-based solution under imperfect CSI in Fig. 4(c), where the AO method [8] and the random phase shift are adopted as benchmark schemes. The DFT-based RP is employed for improving the uplink channel estimation accuracy. Observe from Fig. 4(c) that the codebook-based scheme degrades and essentially it becomes the random phase shift based regime when $Q = 1$. Furthermore, compared to the AO method, the codebook-based scheme suffers from some performance penalty in the case of a high pilot power, but it outperforms the AO method in the prevalent practical scenario of having low pilot power at the UE side and an adequate number of training slots (e.g., $P_{UL} = -20$ dBm and $Q > 10$). In a nutshell, the estimation errors of the direct channels and all reflected channels severely degrade the existing CE &

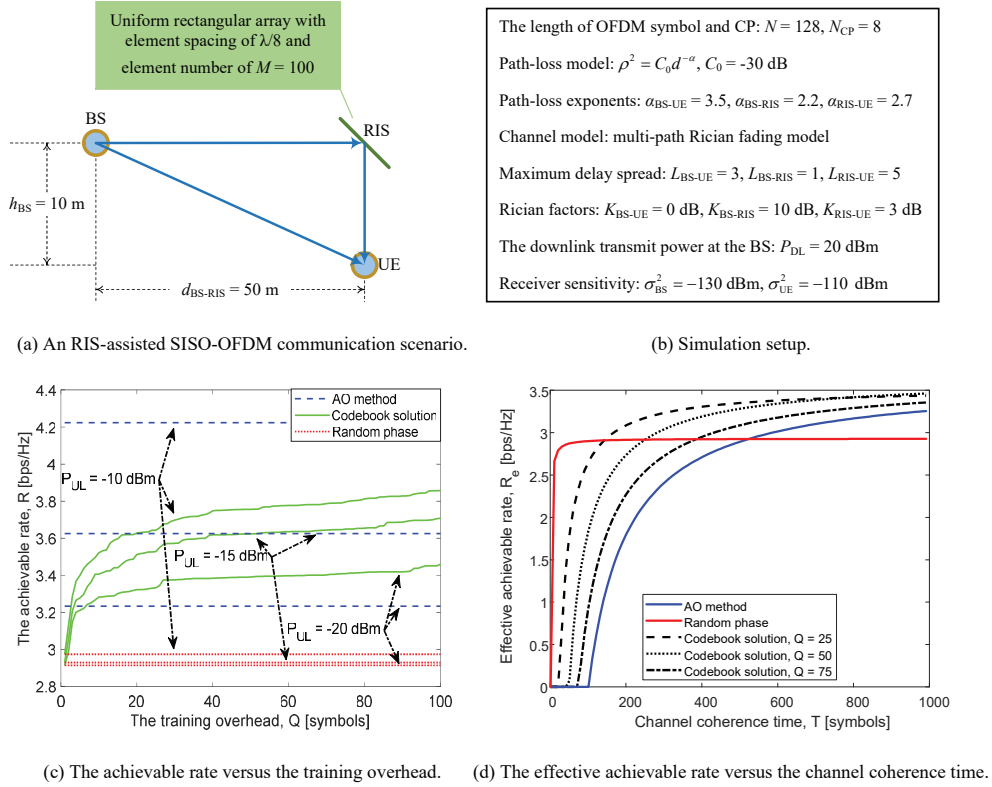


Fig. 4. Case study by considering an RIS-assisted SISO-OFDM system.

PBF performance. By contrast, the codebook-based scheme is shown to be more robust against imperfect CSI.

Then Fig. 4(d) shows the effective achievable rate versus the channel's coherence time with $P_{UL} = -15$ dBm. Specifically, the effective achievable rate is defined by $R_e = (1 - \tau/T)R$, where τ and T represent the training overhead and the channel coherence time, respectively. As seen from Fig. 4(d), when considering rapidly time-varying channels having short coherence time (e.g., $T < 500$), the AO method performs quite poorly in terms of its effective achievable rate, which is a consequence of its excessive overhead required for probing all the reflected channels. By contrast, the random phase shift having the lowest overhead results in the highest effective achievable rate. Fortunately, the scalable codebook-based scheme can dynamically adapt the training overhead Q , which may be optimized for maximizing the effective achievable rate vs the channel coherence time.

VI. CONCLUSIONS

A promising codebook-based framework was proposed for RIS-assisted systems to strike flexible performance and pilot overhead tradeoffs. Specifically, our codebook-based solution is demonstrated to be more suitable for ultra-dense wireless networks with large-scale RIS deployment. Moreover, we discussed the fundamental issues of the codebook-based framework, including the generation of the codebook as well as properly designed learning methods, and elaborated on its benefits over existing CE & PBF solutions. Since the codebook-based framework remains largely unexplored, this article will provide useful guidance for unleashing the full

potential of RIS in practical applications of future wireless networks.

ACKNOWLEDGMENT

This research is supported by the Ministry of Education, Singapore, under its MOE Tier 2 (Award number MOE-T2EP50220-0019). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not reflect the views of the Ministry of Education, Singapore. D. W. K. Ng is supported by the Australian Research Council's Discovery Project (DP210102169). The work of M. Di Renzo was supported in part by the European Commission through the H2020 ARIADNE project under grant agreement number 871464 and through the H2020 RISE-6G project under grant agreement number 101017011. L. Hanzo would like to acknowledge the financial support of the Engineering and Physical Sciences Research Council projects EP/W016605/1 and EP/P003990/1 (COALESCE) as well as of the European Research Council's Advanced Fellow Grant QuantCom (Grant No. 789028).

REFERENCES

- [1] M. Di Renzo, *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.*, vol. 38, pp. 2450–2525, Nov. 2020.
- [2] Y. Zhang, *et al.*, "Dual codebook design for intelligent omni-surface aided communications," *IEEE Trans. Wireless Commun.*, pp. 1–14, 2022, Early Access.
- [3] C. Pan, *et al.*, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, pp. 14–20, Jun. 2021.

- [4] J. An, *et al.*, “Low-complexity channel estimation and passive beamforming for RIS-assisted MIMO systems relying on discrete phase shifts,” *IEEE Trans. Commun.*, vol. 70, pp. 1245–1260, Feb. 2022.
- [5] T. L. Jensen and E. De Carvalho, “An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator,” in *Proc. IEEE ICASSP*, pp. 5000–5004, May 2020.
- [6] Z. Wang, L. Liu, and S. Cui, “Channel estimation for intelligent reflecting surface assisted multiuser communications: Framework, algorithms, and analysis,” *IEEE Trans. Wireless Commun.*, vol. 19, pp. 6607–6620, Oct. 2020.
- [7] C. Huang, *et al.*, “Reconfigurable intelligent surfaces for energy efficiency in wireless communication,” *IEEE Trans. Wireless Commun.*, vol. 18, pp. 4157–4170, Aug. 2019.
- [8] Q. Wu and R. Zhang, “Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming,” *IEEE Trans. Wireless Commun.*, vol. 18, pp. 5394–5409, Nov. 2019.
- [9] M.-M. Zhao, *et al.*, “Intelligent reflecting surface enhanced wireless networks: Two-timescale beamforming optimization,” *IEEE Trans. Wireless Commun.*, vol. 20, pp. 2–17, Jan. 2021.
- [10] G. Gradoni and M. Di Renzo, “End-to-end mutual coupling aware communication model for reconfigurable intelligent surfaces: An electromagnetic-compliant approach based on mutual impedances,” *IEEE Wireless Commun. Lett.*, vol. 10, pp. 938–942, May 2021.
- [11] A. Abrardo, D. Dardari, and M. Di Renzo, “Intelligent reflecting surfaces: Sum-rate optimization based on statistical position information,” *IEEE Trans. Commun.*, vol. 69, pp. 7121–7136, Oct. 2021.
- [12] Y. Han, *et al.*, “Large intelligent surface-assisted wireless communication exploiting statistical CSI,” *IEEE Trans. Veh. Technol.*, vol. 68, pp. 8238–8242, Aug. 2019.
- [13] J. An, *et al.*, “Joint training of the superimposed direct and reflected links in reconfigurable intelligent surface assisted multiuser communications,” *IEEE Trans. Green Commun. Netw.*, vol. 6, pp. 739–754, Jun. 2022.
- [14] M. Cui, G. Zhang, and R. Zhang, “Secure wireless communication via intelligent reflecting surface,” *IEEE Wireless Commun. Lett.*, vol. 8, pp. 1410–1414, Oct. 2019.
- [15] M. Najafi, *et al.*, “Physics-based modeling and scalable optimization of large intelligent reflecting surfaces,” *IEEE Trans. Commun.*, vol. 69, pp. 2673–2691, Apr. 2021.

Jiancheng An [M’22] received the B.S. degree in Electronics and Information Engineering and the Ph.D. degree in Information and Communication Engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2016 and 2021, respectively. From 2019 to 2020, he was a Visiting Scholar with the Next-Generation Wireless Group, University of Southampton, U.K. He is currently a research fellow with the Engineering Product Development (EPD) Pillar, Singapore University of Technology and Design (SUTD). His research interests include reconfigurable intelligent surfaces (RIS), and integrated sensing and communications (ISAC).

Chao Xu [SM’19] received a B.Eng. degree from Beijing University of Posts and Telecommunications, China, and a BSc(Eng) with First Class Honours from Queen Mary, University of London, UK, through a Sino-UK joint degree program in 2008, both in Telecommunications. He obtained a MSc degree with distinction in Radio Frequency Communication Systems and a Ph.D. degree in Wireless Communications from the University of Southampton, UK, in 2009 and 2015, respectively. He is currently a senior research fellow working at Next Generation Wireless Research Group, University of Southampton, UK. His research interests include index modulation, reconfigurable intelligent surfaces, noncoherent detection and turbo detection. He was awarded the Best M.Sc. Student in Broadband and Mobile Communication Networks by the IEEE Communications Society (United Kingdom and Republic of Ireland Chapter) in 2009. He also received 2012 Chinese Government Award for Outstanding Self-Financed Student Abroad and 2017 Dean’s Award, Faculty of Physical Sciences and Engineering, the University of Southampton.

Qingqing Wu [SM’21] was listed as the Clarivate ESI Highly Cited Researcher in 2022 and 2021, the Most Influential Scholar Award in AI-2000 by Aminer in 2021 and World’s Top 2% Scientist by Stanford University in 2020 and 2021. He was the recipient of the IEEE ComSoc Asia Pacific Best Young Researcher Award and Outstanding Paper Award in 2022, the IEEE ComSoc Young Author Best Paper Award in 2021.

Derrick Wing Kwan Ng [F’21] received his Ph.D. degree from the University of British Columbia in 2012. He is now working as an Associate Professor at the University of New South Wales, Sydney, Australia. His research interests include convex and non-convex optimization, physical layer security, wireless information and power transfer, and green (energy-efficient) wireless communications.

Marco Di Renzo [F’20] is a CNRS Research Director (Professor) and the Head of the Intelligent Physical Communications in the Laboratory of Signals and Systems at Paris-Saclay University - CNRS and CentraleSupélec. He serves as the Coordinator of the Communications and Networks Area of the Laboratory of Excellence DigiCosme, and as a Member of the Admission and Evaluation Committee of the Ph.D. School of Paris-Saclay University. He is a Fulbright Fellow at City University of New York, USA; an Fellow of IEEE, IET, AIAA, Vebleo; an Ordinary Member of EASA and the Academia Europaea; and a Highly Cited Researcher. He received the 2022 Michel Monpetit Prize from the French Academy of Sciences. He serves as the Editor-in-Chief of IEEE Communications Letters.

Chau Yuen [F’21] received the B.Eng. and Ph.D. degrees from Nanyang Technological University (NTU), Singapore, in 2000 and 2004, respectively. He was a Post-Doctoral Fellow with Lucent Technologies Bell Labs, Murray Hill, in 2005. From 2006 to 2010, he was with the Institute for Infocomm Research (I2R), Singapore. Since 2010, he has been with the Singapore University of Technology and Design. Dr. Yuen was a recipient of the Lee Kuan Yew Gold Medal, the Institution of Electrical Engineers Book Prize, the Institute of Engineering of Singapore Gold Medal, the Merck Sharp and Dohme Gold Medal, and twice a recipient of the Hewlett Packard Prize. He received the IEEE Asia Pacific Outstanding Young Researcher Award in 2012 and IEEE VTS Singapore Chapter Outstanding Service Award on 2019. Currently, he serves as an Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE System Journal, and IEEE Transactions on Network Science and Engineering. He served as the guest editor for several special issues, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS MAGAZINE, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING. He is a Distinguished Lecturer of IEEE Vehicular Technology Society.

Lajos Hanzo [F’04] (<http://www-mobile.ecs.soton.ac.uk>, https://en.wikipedia.org/wiki/Lajos_Hanzo) received his Master degree and Doctorate in 1976 and 1983, respectively from the Technical University (TU) of Budapest. He was also awarded the Doctor of Sciences (DSc) degree by the University of Southampton (2004) and Honorary Doctorates by the TU of Budapest (2009) and by the University of Edinburgh (2015). He is a Foreign Member of the Hungarian Academy of Sciences and a former Editor-in-Chief of the IEEE Press. He has served several terms as Governor of both IEEE ComSoc and of VTS. He has published widely at IEEE Xplore, 19 Wiley-IEEE Press books and has helped the fast-track career of 123 PhD students. Over 40 of them are Professors at various stages of their careers in academia and many of them are leading scientists in the wireless industry. He is also a Fellow of the Royal Academy of Engineering (FREng), of the IET and of EURASIP. He holds the Eric Sumner Field Award.