Abstract—Reconfigurable intelligent surfaces (RISs) are considered as potential technologies for the upcoming sixth-generation (6G) wireless communication system. Various benefits brought by deploying one or multiple RISs include increased spectrum and energy efficiency, enhanced connectivity, extended communication coverage, reduced complexity at transceivers, and even improved localization accuracy. However, to unleash their full potential, fundamentals related to RISs, ranging from physical-layer (PHY) modelling to RIS phase control, need to be addressed thoroughly. In this paper, we provide an overview of some timely research problems related to the RIS technology, i.e., PHY modelling (including also physics), channel estimation, potential RIS architectures, and RIS phase control (via both model-based and data-driven approaches), along with recent numerical results. We envision that more efforts will be devoted towards intelligent wireless environments, enabled by RISs.

Index Terms—Reconfigurable intelligent surface, channel estimation, PHY modelling, channel sparsity, deep learning.

I. INTRODUCTION

The upcoming sixth-generation (6G) communication system is expected to significantly enhance the quality of services (QoSs). To this end, the key performance indicators in the current fifth generation (5G) need to be further improved. This can be reaped by Terahertz (THz) communications, reconfigurable intelligent surfaces (RISs), and their integration with artificial intelligence (AI) [1]–[3]. Among these potential candidates, a RIS, deployed with a large number of low-cost passive elements, is envisioned to provide the communications with a higher spectral efficiency/energy efficiency (SE/EE) [4]. A RIS can have various functionalities, e.g., reflection, polarization, refraction, and absorption, making intelligent control of the wireless propagation environment feasible [1], [5], [6]. With the introduction of the RIS to wireless communications, a new era, in which a controllable propagation environment becomes a reality, is deemed to come. Besides for the communications, the RIS can also be considered for localization together with millimeter wave (mmWave) multiple-input multiple-output (MIMO) networks [7]–[9].

In terms of hardware implementation, the RIS can be made of varactor diodes, crystal liquid, etc [10], [11]. As for physical-layer (PHY) modelling, the recent papers [12]–[14] focused on the analyses of path loss for different RIS behaviors, i.e., reflection vs. scattering, and the RIS phase profile is designed by addressing the corresponding integral equations for the electric and magnetic vector fields. Alternative physically feasible fading models were proposed under isotropic/rich scattering in [15], [16]. It should also be noted that a ray-tracing-based RIS channel model can be found in [17] for both indoor and outdoor environments.

In a RIS-assisted network, the channel state information (CSI) acquisition is essential for the RIS phase control, beamforming, resource allocation, and interference management [18]. Thus, efficient channel estimation (CE) algorithms are necessary. The RIS is often integrated into mmWave or (sub-)THz communications systems to enable line-of-sight (LOS)-like connectivity. In these frequency bands, the wireless channels are usually very directive and sparse with limited numbers of resolvable paths. Thus, the channel coefficient matrices are rank-deficient, resulting in inherent channel sparsity. A variety of compressive sensing (CS) techniques, e.g., atomic norm minimization, basis pursuit, approximate message passing (AMP), and mixed norm minimization, can be well-tailored for CE [19]–[24] in the RIS-aided system. In order to collect measurements, pilot sequences or beam training matrices are required. In the parametric channel models, heavily used in mmWave and THz communications, the channel parameters and locations of the terminals are closely coupled. Knowing one can infer the other. Thus the prior information on the mobile station (MS) or environment objects will facilitate the design of pilot sequences, which in turn improves the CE performance [25]. In addition to conventional model-based approaches, data-driven approaches, for instance, deep learning (DL) frameworks, can also be employed for CE, phase control, and symbol detection in RIS-aided communication systems. They are capable of mapping the received signals to the individual channels, RIS phase profile, and constellation points depending on the purpose of data-driven models.

RIS-assisted communication has gained much interest recently. Various researches have been conducted to realize the RIS, including but not limited to CE, performance analysis, design and optimization of the RIS. In this paper, we offer an overview of the recent research progress and results on RIS technology. In particular, we at first provide discussion on the physical modelling for the RIS. Then, we introduce efficient CE algorithms, novel architectures, and data-driven methods for the RIS-assisted communication system.

II. PHYSICAL LAYER MODELLING

A. Physical RIS Modelling

In the system models used in algorithm development, the RIS is typically modeled as an array of independent reflecting elements separated by \( \lambda/2 \) (\( \lambda \) = wavelength). This assumption
is valid if the RIS is implemented as a reflecting antenna array. Indeed, electrically tunable reflecting antenna arrays can be used to dynamically adjust their radiation patterns [10]. The control of the antenna elements can be implemented by loading the antenna elements with tunable impedances. The most often used tunable impedance in RIS prototypes is a varactor implemented with a p-i-n diode. By varying the load impedance, the resonance frequency of an antenna element is varied. This leads to the change in the phase response of the element at the frequency of the impinging wave. The desired reflecting pattern of the RIS can then be realized by joint control of the reflection phases of individual elements in a similar manner as in antenna arrays used in wireless receivers and transmitters.

However, in many of the publications discussing the potentials of the RIS in wireless systems, the RIS is assumed to be implemented with a metasurface. Those are electrically thin and dense two-dimensional arrays of structural elements possessing desired properties granted by their constitutive elements. Elements are called meta-cells, unit-cells or meta-atoms. The size of a meta-cell is typically much smaller than the signal’s wavelength, between $\lambda/10$ and $\lambda/5$ [26], [27]. Metasurfaces have the potential to offer better control on the propagation of the electromagnetic (EM) waves than the reflecting antenna arrays. In addition to steering the reflection or refraction angles of the impinging EM wave, metasurfaces can be designed to, e.g., absorb the waves to reduce the interference, change the polarization of the waves, and realize data modulation [5], [28].

B. RIS Channel Modelling

Understanding the behavior of frequency bands is crucial to optimize the frequency resources among different applications. The trend is to use higher frequencies in order to provide large capacities for data-hungry applications. The mmWave and THz band communications offer very high bandwidths with the cost of large propagation loss. Those are handled with highly directional high gain antenna systems. The large losses, especially in the non-line-of-sight (NLOS) paths, ultimately mean that the high-frequency channels tend to be sparse. That is, there are only few utilizable high gain paths and the common assumption of a rich scattering channel is no longer valid. Channel sparsity leads to challenges in aligning the transmit and receive beams and also to weak connections. A RIS may help by modifying the propagation channels by increasing the gain or creating a reflected LOS connection.

The channel modelling with RISs consists of the channel between the base station (BS) and the RIS, the RIS response, and the RIS–MS channels. As such, the BS–RIS and RIS–MS channels require novel channel models for mmWave and THz bands. There are many recent works aiming modelling mmWave and THz band propagation channels, such as [29], [30]. The specific RIS channel models depend not only on the frequency, but also on the placement scheme of the RISs. As RISs and BSs are stationary in the general, their positions can be optimized to minimize the channel path losses [31].

Fig. 1: The RIS-assisted communication system in the presence of LOS blockage. Different types of RIS architectures can be considered. MSs are in general randomly located and mobile. This causes a need to optimize the RIS beamformers in order to maximize the end-to-end link quality. From the channel point of view, the CE depends on the specific frequency-dependent path loss models. However, the mmWave and THz band channels share the sparsity due to weak penetration through the objects as well as lower dynamic range caused by large bandwidths. That is, the weak paths tend to disappear due to the high noise floor. On the other hand, to overcome the large path loss, directional antennas are required also in the CE phase. This is a challenge that we will be looked into below. In summary, the high-frequency channels are very promising due to large available frequency resources, but they also impose challenges in coping with low dynamic range and large path loss. The RISs are seen as very promising technologies to maximize the NLOS channel gains and to make it possible to increase the service area of the BSs and increases the QoSs at the same time.

III. RIS Channel Estimation

In the sequel, the signal processing techniques are based on the system model, depicted in Fig. 1, where the BS communicates with the MS via the RIS. The number of (antenna) elements at the BS, MS, and RIS are denoted by $N_t$, $N_r$, and $N$, respectively. The RIS can be purely passive, hybrid (a combination of passive and active RIS elements), or in the form of hybrid relay-reflecting (HR-RIS), i.e., a mixture functionality of a RIS and a relay [32]–[34]. The antenna configuration at the BS and MS depends on the practical applications, e.g., carrier frequency, physical size constraint, and power consumption limitation, etc. In this paper, a multi-antenna BS is assumed while both single- and multi-antenna MSs are considered, i.e., $N_t > 1$, $N_r \geq 1$.

For the purely passive RIS architectures without any baseband processing units, CE can only be done at BS or MS due to the unavailability of observations at the RIS. According to the literature, the CE methods are classified into two major categories: individual CE (also known as cascaded CE) [19], [35]–[37], and channel parameter estimation [20]–[23]. For the former, parallel factor tensor decomposition, matrix factorization & matrix completion and their variants can be applied. For the latter, inherent channel sparsity is adopted.
for the estimation algorithm design along with the assumption of parametric channel models. Compared to the individual CE, channel parameter estimation has the following advantages. (i) Instead of estimating the individual channel matrices, the number of parameters to be estimated is reduced significantly. (ii) The inherent channel sparsity can be leveraged in the CE algorithm development, and the availability of various CS techniques can be applied or modified for extracting the channel parameters with a small amount of training overhead.

On the contrary, for the ease of CE, active sensors can be added to the RIS, resulting in so-called hybrid RIS architecture [38]. With the introduction of active sensors (elements), the CE can be performed at the RIS, and the CE problem degrades to that for P2P channels. However, this requires more computational power, hardware complexity, and energy consumption for the RIS.

### A. Individual Channel Estimation

In [19], Wang et al. studied CE for the individual channels by using compressive sensing techniques, e.g., OMP and generalized AMP (GAMP). However, all the angular parameters were assumed to lie in a predefined grid, which leads to grid mismatch problem and limit its practical applications. In [36], [37], parallel factor tensor decomposition are applied to estimate the individual channel matrices, e.g., BS-RIS and RIS-MS channels with an iterative refinement of the individual CE conducted by using bilinear alternating least squares (BALS), while in [35] the individual channels are estimated in a sequential way by an alternation of matrix factorization and matrix completion. In these works, heavy training overhead is demanded as to achieve satisfactory estimation performance.

### B. Channel Parameter Estimation

Channel parameter estimation was considered in [20]–[23]. In [23], we made an intensive comparison on the CE performance for both passive and hybrid RIS architectures in terms of mean squared error (MSE) of the channel parameters, RIS gain, and SE. The CE for the passive and hybrid RIS architectures is addressed via atomic norm minimization along with multiple signal classification (MUSIC) algorithm or its variants. As for the passive RIS architecture, we proposed a two-stage sounding and CE procedure, detailed in [21]. By assuming downlink training (also applied for uplink training), the BS sends pilots to the MS, followed by the first stage of CE for the extraction of the angle of arrivals (AoAs) at the MS and angle of departure (AoD) at BS. The estimates are used to guide the design of the beam training matrix in the second stage sounding, aiming at the estimation of the angle differences associated with the RIS and products of propagation path gains. The hybrid RIS architecture considers alternating uplink and downlink training. Namely, the BS-RIS and RIS-MS channels are estimated at the RIS with pilots sent from the BS and MS, respectively.

Fig. 2 shows the simulation results for the estimation of the channel parameters, where $\phi_{R,M}$, $\delta$, and $\rho$ denotes the AoAs at MS, angle differences, and products of path gains, respectively. The performance is evaluated by considering the purely passive RIS and the hybrid RIS ($K = N_{RF} = 30$ and $K = N_{RF} = 20$), where $K$ is the number of active elements, and $N_{RF}$ is the number of radio frequency (RF) chains at the RIS. The MSE of the AoA shows that the purely passive RIS achieves better performance than the hybrid RIS under the same training overhead. A similar conclusion can be drawn for the estimation of products of path gains. However, the MSEs of the angle difference are approximate for the setup with $K = N_{RF} = 30$ and passive RIS architectures.

### C. Position-Aided Channel Estimation

In the parametric channel models, the channel parameters are calculated based on the locations/coordinates of the terminals and vice versa. Thus, the position information can offer prior information, albeit rough, for the channel parameters, e.g., AoAs and AoDs. This rough information can enhance the design of training beams used for CE, which in turn brings better estimation performance [25], [39]. In [25], we assumed a bounded position estimation error, which is equivalent to a reduced angular interval (compared to $[-\pi, \pi]$) after mapping the position to angular information. It was verified that better parameter estimation performance can be attained. Relying on the parameter estimates, a better RIS phase control matrix can be designed for communication purpose. The effect of prior location information on CE performance is illustrated in Fig. 3 for both single-path and multi-path scenarios. The training overhead in symbols are $T_i = 64$ and $T_i = 144$, respectively. For both scenarios, obvious improvement can be seen for the MSE performance with the aid of prior location information.

### IV. HYBRID RELAY-REFLECTING INTELLIGENT SURFACE

Intensive studies on the performance and design aspects of the RIS-assisted communication systems have shown the capability of RIS in improving the system capacity, sum-rate, and channel quality [40]–[45]. However, a main limitation of the RIS is that its passive beamforming limits the beamforming gains. It has been shown in [44], [46] that a very large number of reflecting elements need to be deployed in the RIS to perform better than decode-and-forward (DF) relaying; otherwise, it can be easily outperformed even by a half-duplex (HD) relay with few elements. In addition, a practical RIS with
limited-resolution phase shifts has considerable performance degradation [47], [48]. To overcome the aforementioned limitations of RIS, HR-RIS architectures and the concept of semi-active/passive beamforming have been introduced in [33]. The basic idea of the HR-RIS is to connect a few elements of the RIS to RF chains and power amplifiers (PAs) (RF-PA chains). As a result, HR-RIS can leverage the advantages while mitigating the disadvantages of RIS and relay.

A. Design Aspects of the HR-RIS

With an HR-RIS equipped with \( N \) elements, only \( K \ll N \) active relay elements are deployed, while the remaining elements are passive and only reflect the incident signals. Similar to the conventional RIS, the passive element can only shift the phase; in contrast, the active one can tune both the phase and amplitude of the incident signal. By optimizing the number of active elements, the HR-RIS can achieve high active beamforming gains while still almost preserving the passive beamforming gain. The analysis and numerical results in [33] have shown that only a few active elements, i.e., small \( K \), are sufficient for the HR-RIS to attain significant performance improvement with respect to the conventional RIS. In particular, increasing \( K \) does not always guarantee the improvement in the SE, while causing a linear increase in the total power consumption and hardware cost. Considering these aspects, the HR-RIS is suggested to deploy with a single or few active elements to attain remarkable improvement in the SE and EE.

Two architectures for the HR-RIS, namely, the fixed and dynamic HR-RIS have been proposed in [33]. In the former, the number and the positions of the active elements are predetermined and designed in manufacture. By contrast, those

B. Optimization of HR-RIS

Unlike the conventional RIS, a subset of elements in the HR-RIS are active ones that are supposed to modify not only the phases but also the amplitudes of the incident signals. Therefore, their amplitudes are not unity and required to be optimized. In [33], [34], the optimization of the coefficients of HR-RIS is formulated in a SE maximization problem, which is solved with an alternating optimization (AO) method [45]. The closed-form solution to the reflecting/relaying coefficient of the fixed HR-RIS is derived. For the dynamic HR-RIS, the problem becomes more challenging because the set of active elements is unavailable. To overcome this, a power allocation problem is formulated, enabling an efficient selection of active elements and obtaining their amplitudes based on the well-known water filling algorithm [45].

The SE and EE of the proposed HR-RIS are shown Fig. 5 in comparison with those of the conventional RIS, whose phases are either randomly generated or optimized using the AO scheme [45]. The results for the relay/RIS with only \( K, K \in [1, N] \), active/passive elements are also included for comparison. It is observed that increasing \( K \) provides the RIS (with only \( K \) elements) higher passive beamforming gains, and thus, its SE and EE monotonically increases with \( K \). By contrast, with a larger \( K \), the relay and HR-RIS can significantly amplify the signal and attain higher SEs, which, however, comes at the cost of degraded EEs due to the increase in power consumption. Furthermore, it is observed that the SEs of the HR-RIS schemes quickly grow with small \( K \), and almost converge with moderate or large \( K \). In particular, the best SE-EE tradeoff is attained with a small \( K \).

V. DATA-DRIVEN APPROACHES

Different from model-based approaches, data-driven schemes can be implemented with very low complexity (with only multiplication and addition operations, and element-wise
activation) [51]–[54]. The computation-heavy model training can be done offline. Data-driven approaches have been applied for CE and beamforming, RIS phase control, and symbol detection, etc. For example, RIS CE with the application of DL framework was first studied in [55]. To be specific, a twin convolutional neural network (CNN) is designed for mapping the received pilot signals to both direct and cascaded channels. The latest work on the application of deep neural network (DNN) to RIS phase control was included in [56].

A DL model can also be trained for passive beamforming with imperfect CSI. It is noted that phase shifts can be computed using closed-form solution obtained by traditionally hand-engineered methods, such as the semi-definite relaxation (SDR) [44] or AO [45]. However, these approaches generally require high complexity and feedback time, and thus, not amenable for real-time application [57]. In particular, when the CSI errors exist, the conventional closed-form solutions derived for the perfect CSI may become unreliable. In this scenario, an unsupervised learning-based DNN-aided passive beamforming can be employed [58]. More specifically, the DNN is trained to learn the phase shifts with the imperfect CSI, as illustrated in Fig. 6. In Fig. 7, we show the SEs of the $2 \times 1$ and $4 \times 1$ MISO systems with $N = \{30, 50\}$, respectively.

VI. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

In this paper, we have covered a wide range of closely-related RIS topics with certain solutions given. Nevertheless, more efforts are required in order to make the commercialization of RIS feasible in the foreseeable future. For instance, mature yet valid RIS channel models for different frequency bands are still unavailable. For the moving users, efficient yet effective channel tracking algorithms need to be further investigated. Also, a full understanding of the pros and cons for different RIS architectures is not yet reached. The coexistence of multi-RIS with different architectures will make the communication systems even more complicated, joint RIS activation/deactivation control, phase profile design, and user scheduling also needs to be further studied.

ACKNOWLEDGMENT

This work is supported by Horizon 2020, European Union’s Framework Programme for Research and Innovation, under grant agreement no. 871464 (ARIADNE) and partially supported by the Academy of Finland 6Genesis Flagship (grant 318927) and by the Academy of Finland (ROHM project, grant 319485). We would like to thank Dr. Henk Wymeersch, the co-author of [21], [25], and Dr. Quang-Doanh Vu, Dr. Kyungchun Lee, the co-authors of [33], [34], for their contributions to the development of the CE schemes and the HR-RIS architectures introduced in this paper.

REFERENCES
