



HAL
open science

Ambiguity-Aware Segmented Estimation of Mutual Coupling in Large RIS: Algorithm and Experimental Validation

Philipp del Hougne

► **To cite this version:**

Philipp del Hougne. Ambiguity-Aware Segmented Estimation of Mutual Coupling in Large RIS: Algorithm and Experimental Validation. IEEE Transactions on Communications, 2026, 74, pp.4615-4628. <10.1109/tcomm.2026.3653083>. <hal-05491103>

HAL Id: hal-05491103

<https://hal.science/hal-05491103v1>

Submitted on 30 Apr 2026

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.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

Ambiguity-Aware Segmented Estimation of Mutual Coupling in Large RIS: Algorithm and Experimental Validation

Philipp del Hougne, *Member, IEEE*

Abstract—Optimizing a real-life RIS-parametrized wireless channel with a physics-consistent multiport-network model necessitates prior remote estimation of the mutual coupling (MC) between RIS elements. The number of MC parameters grows quadratically with the number of RIS elements, posing scalability challenges. Because of inevitable ambiguities, independently estimated segments of the MC matrix cannot be easily stitched together. Here, by carefully handling the ambiguities, we achieve a separation of the full estimation problem into three sequentially treated sets of smaller problems. We partition the RIS elements into groups. *First*, we estimate the MC for one group as well as the characteristics of the available loads. *Second*, we separately estimate the MC for each of the remaining groups, in each case with partial overlap with an already characterized group. *Third*, we separately estimate the MC between each distinct pair of groups. Full parallelization is feasible within the second and third sets of problems, and the third set of problems can furthermore benefit from efficient initialization. We experimentally validate our algorithm for a 4×4 MIMO channel parametrized by a 100-element RIS inside a rich-scattering environment. Our experimentally calibrated 5867-parameter multiport-network model achieves an accuracy of 40.5 dB, whereas benchmark models with limited or no MC awareness only reach 16.9 dB and 13.8 dB, respectively. Based on the experimentally calibrated models, we optimize the RIS for five wireless performance indicators. Experimental measurements with the optimized RIS configurations demonstrate only moderate benefits of MC awareness in RIS optimization in terms of the achieved performance. However, we observe that limited or no MC awareness markedly erodes the reliability of model-based predictions of the expected performance.

Index Terms—Reconfigurable intelligent surface, multiport-network theory, mutual coupling, parameter estimation, Virtual VNA, ambiguity, reverberation chamber, MIMO, separability, ambiguity-aware segmentation, gradient descent, optimization.

I. INTRODUCTION

Reconfigurable intelligent surfaces (RISs) enable wireless practitioners to tailor wireless channels to the needs of diverse wireless functionalities. Multiport-network theory (MNT) provides a rigorous framework to physics-consistently model such RIS-parametrized channels [1], [2]. Specifically, within MNT, each tunable lumped element that parametrizes an RIS element's response is treated as a lumped port terminated by

a tunable load. Importantly, MNT accurately captures mutual coupling (MC) between RIS elements. To date, the main focus of theoretical MNT-based studies has been on MC-aware optimizations of the RIS configuration [1]–[12]. However, this presupposes that the MNT model parameters are known.

Theoretical studies consider strongly simplified scenarios (e.g., minimally scattering thin-wire RIS elements in free space) for which the parameters can be determined analytically [1]. There are also some numerical efforts to extract the parameters for realistic RIS designs from full-wave simulations [13], [14]. However, practitioners need tools that do not require strong a priori knowledge. On the one hand, the RIS design may be completely unknown because it is proprietary, or partially unknown due to fabrication inaccuracies and component tolerances. On the other hand, the MC between RIS elements depends not only on the RIS design but also on the radio environment's geometry and material composition which cannot be known accurately in practice unless it is free space.¹ Techniques requiring minimal a priori knowledge to estimate the MNT model parameters given an experimental RIS-parametrized channel are thus essential.

Experimentally, the MNT model parameters can usually *not* be identified unambiguously. This limitation originates from the infeasibility of direct measurements and inevitable ambiguities in most remote measurements, as we now explain.

Directly measuring the MNT parameters would require connecting the lumped ports involved in the conceptual description of the RIS elements' tunable components to a vector network analyzer (VNA) or similar instrumentation. However, these conceptual ports associated with the RIS elements are not physically accessible in integrated RIS designs, i.e., one cannot connect them to a VNA instead of connecting them to the tunable loads.² Moreover, VNAs typically have less than ten ports, whereas RISs are envisioned to have hundreds of elements. Furthermore, for most real-life applications, VNAs are prohibitively costly. The integration of sensors into RIS elements does not appear to provide a viable alternative for directly measuring the MNT parameters. Indeed, sensor-equipped RIS elements were proposed for

This work was supported in part by the ANR France 2030 program (project ANR-22-PEFT-0005), the ANR PRCI program (project ANR-22-CE93-0010), the Rennes Métropole AES program (project "SRP"), the European Union's European Regional Development Fund, and the French Region of Brittany and Rennes Métropole through the contrats de plan État-Région program (projects "SOPHIE/STIC & Ondes" and "CyMoCoD").

P. del Hougne is with Univ Rennes, CNRS, IETR - UMR 6164, F-35000, Rennes, France (e-mail: philipp.del-hougne@univ-rennes.fr).

¹Experimental evidence for the dependence of the MC between RIS elements on the radio environment can be found in [15], [16], where a strong dependence of the MC strength on the radio environment was observed.

²Recent RIS prototypes with modular designs [17]–[19] offer direct access to the ports associated with the RIS elements. Such modular designs are therefore valuable in early-stage academic research but integrated designs will prevail in large-scale real-life deployment because they are more compact and cheaper to manufacture at scale.

channel estimation assuming MC-unaware cascaded channel models [20], [21]. Directly measuring the MC between the RIS elements would require, first, synchronization of the sensors, and, second, sensors capable of emitting waves. Ultimately, the sensor network integrated into the RIS would have to offer capabilities similar to a VNA with hundreds of ports, which appears unfeasible. Altogether, the MNT model parameters thus cannot be measured directly in experiments.

Hence, experimentally, the MNT model parameters can only be estimated remotely. However, remote estimates of MNT parameters are typically ambiguous. Indeed, an unambiguous MNT parameter estimation is only possible if each port associated with a RIS element can be terminated with three distinct and known individual loads, as well as known coupled loads [22], [23].³ These conditions, under which tunable loads can serve as “virtual” VNA ports [22], [23], are typically not satisfied. Most RIS prototypes only have two distinct individual loads (realized with a PIN diode) whose characteristics are only known approximately (due to component tolerances) or not at all. Fortunately, however, the ambiguities in the MNT model parameters are *operationally irrelevant* for practitioners who use them for MC-aware RIS optimization [19], [25], [26].

Few works have tackled the (inevitably ambiguous) remote MNT parameter estimation for RIS-parametrized channels without assuming strong a priori knowledge about the RIS design and/or the radio environment. The first technique, introduced and experimentally validated in [25], was a global gradient descent, seeking to simultaneously estimate all MNT parameters.⁴ The same technique was applied to beyond-diagonal RIS (BD-RIS) with tunable inter-element coupling in [26] based on a physics-compliant diagonal representation of BD-RIS-parametrized channels. However, a fundamental difficulty associated with a global gradient descent is that the number of unknowns scales quadratically with the number of RIS elements. Since RISs are envisioned to consist of hundreds of RIS elements, the computational requirements for a global gradient descent can become prohibitively costly.

Recently, [19] combined closed-form and gradient-descent steps to reduce the number of parameters estimated by gradient descent. Specifically, [19] limited the application of gradient descent essentially to estimating the MC between the RIS elements while using singular-value decompositions (SVDs) to estimate the transmission coefficients between antennas and RIS elements and fixing the MC between the antennas based on a single reference measurement. However, the fundamental computational challenges associated with the quadratic scaling persist.

Ideally, the estimation of the MC between the RIS elements could be separated into smaller chunks that can be independently tackled in separate (independent) gradient descents (not to be confused with alternating block gradient descent). Thereby, limitations on computational memory

could be accommodated and massive parallelization could be achieved. In experiments where measurement speed is the bottleneck, one could finalize the estimation of segment parameters for which the measurements are already completed while other measurements are still ongoing. Moreover, one may hypothesize that a gradient descent on smaller problems converges faster because the associated parameter landscape is smoother. However, the inevitable ambiguities fundamentally thwart such a separation into independent smaller problems because independently estimated blocks of the MC matrix cannot be straightforwardly stitched together. We overcome this challenge in the present work in which we propose an ambiguity-aware segmentation of the MC estimation, as detailed in Sec. I-A. Thereby, we remove the constraint (currently imposed by available computational resources) on the maximum number of RIS elements for which the MNT parameters can be estimated.

Given the current lack of experimental parameter estimation techniques for MNT models of RIS-parametrized channels, there is consequently also a lack of experimental MNT-based performance evaluations. While [25] considered a relatively large RIS with 68 1-bit-programmable elements, the performance evaluation therein was limited to a few antenna-level metrics optimized with a dictionary search. Meanwhile, [18], [19] conducted more extensive experimentally grounded performance evaluations of communications metrics for beyond-diagonal RIS and RIS with multi-bit-programmable elements, respectively, but the considered RISs were only composed of 8 RIS elements. Moreover, only [19] conducted experimental channel measurements for the optimized RIS configurations. Altogether, the literature does not contain any reports of extensive experimental MNT-based performance evaluations for *large* RIS. As detailed in Sec. I-A, we fill this research gap in the present work in which we benchmark the experimentally achieved performance with the MNT model against lower-fidelity models across five key performance indicators; we also analyze the accuracy with which each model predicts the achieved performance.

A. Contributions

In this article, we develop an ambiguity-aware segmentation of the MC parameter estimation problem for large RIS and leverage it for experimental MNT-based performance evaluations with a large-scale RIS. Our contributions are summarized as follows.

- 1) We present a separation of the estimation of the RIS's MC parameters into three sequentially tackled sets of subproblems based on a partition of the RIS elements into small groups. The first set only comprises one subproblem. The number of subproblems in the second set scales linearly with the number of groups; the problems can be formulated completely independently from each other and are hence fully parallelizable. The number of subproblems in the third set scales quadratically with the number of groups; the subproblems are completely independent from each other and hence fully parallelizable.

³Very recent work [24] suggests that two distinct known individual loads and known coupled loads are sufficient because the latter provide effectively a third termination distinct from the two individual loads.

⁴[25] used a coupled-dipole formulation that is operationally equivalent to the scattering-parameter representation of MNT [Sec. II.C.2, [26]].

- 2) We identify a low-cost initialization for the third set of subproblems.
- 3) We experimentally validate our technique for a 100-element RIS. Moreover, we benchmark the achieved accuracy against three simplified models. Furthermore, we systematically examine the influence of the number of antennas and the measurement noise on the achieved accuracy.
- 4) We conduct MC-aware optimizations with our calibrated MNT model for five communications performance metrics. Moreover, we benchmark the achieved performances against those achieved with three simplified models.

To the best of our knowledge, the present work is the first effort to segment the estimation of the MC between RIS elements. Related prior works [19], [25], [26] differ regarding the extent to which they separate the estimation of the MC between RIS elements from the estimation of other model parameters, but none of [19], [25], [26] segments the estimation of the MC between RIS elements.

B. Organization

The remainder of this article is organized as follows. In Sec. II, we describe our MNT system model, as well as three low-fidelity simplified models used for benchmarking. In Sec. III, we introduce our ambiguity-aware segmented parameter estimation technique for the MNT model step-by-step. In Sec. IV, we describe our experimental setup and detail the experimental validation of our technique. In Sec. V, we report extensive experimental performance evaluations based on our calibrated MNT-based model, as well as relevant benchmarking against simplified models. In Sec. VI, we discuss use cases of our technique for wireless practitioners. In Sec. VII, we conclude.

C. Notation

$\mathbf{A} = \text{diag}(\mathbf{a})$ denotes the diagonal matrix \mathbf{A} whose diagonal entries are \mathbf{a} . \mathbf{A}_{BC} denotes the block of the matrix \mathbf{A} whose row [column] indices are in the set \mathcal{B} [\mathcal{C}]. \mathcal{B}_i is the singleton containing the i th entry of \mathcal{B} . a_i denotes the i th entry of the vector \mathbf{a} . a_{ij} denotes the (i, j) th entry of the matrix \mathbf{A} . \top and \dagger denote the transpose and transpose-conjugate operations, respectively. $\mathbf{1}_q$ denotes a $q \times 1$ vector whose entries are unity.

II. SYSTEM MODEL

In this section, we first introduce the high-fidelity physics-consistent MNT system model in Sec. II-A. Then, in Sec. II-C, we relate it to three simplified system models that we consider for benchmarking in this paper. Finally, in Sec. II-B, we discuss the mapping from the control sequence chosen by the wireless practitioner to the physical load characteristics of the RIS elements.

A. MNT System Model

Any RIS-parametrized radio environment can be split into three entities: (i) a set of N_A antenna ports (via which waves can be injected and/or received), (ii) a set of N_S tunable lumped elements (which can be modeled as lumped ports terminated by tunable individual loads), and (iii) static scattering objects.⁵ The latter comprise both environmental scattering objects as well as the structural scattering of the antennas and the RIS elements. We assume that the items captured by (iii) are linear, passive, and reciprocal, which is reasonable because typically considered radio environments do not contain non-linear, active or non-reciprocal elements; beyond these general assumptions, we do *not* assume any specifics related to the radio environment or RIS element design. We further assume that all ports are sufficiently small to be described as lumped ports, which is reasonable because typical antenna ports as well as typical tunable components (e.g., PIN diodes or varactor diodes) are much smaller than the operating wavelength (see also Sec. IV-A). Finally, we assume for simplicity that the generators and detectors that inject and receive signals via the antenna ports are matched to a common reference impedance Z_0 , which is reasonable because such devices are commonly matched to 50Ω .

Irrespective of the complexity of the radio environment and the detailed RIS element design, entity (iii) is thus an N -port system, where $N = N_A + N_S$, characterized by its scattering matrix $\mathbf{S} \in \mathbb{C}^{N \times N}$ (defined using the same reference impedance Z_0 at all ports). Moreover, the set of N_S tunable loads constitutes an N_S -port scattering system characterized by its diagonal scattering matrix $\Phi \in \mathbb{C}^{N_S \times N_S}$. The i th diagonal entry of Φ is the reflection coefficient c_i of the i th individual load associated with the i th RIS element. We further define the vector $\mathbf{c} = [c_1, c_2, \dots, c_{N_S}] \in \mathbb{C}^{N_S}$ that is related to Φ as follows: $\Phi = \text{diag}(\mathbf{c})$. We denote by \mathcal{T} , \mathcal{R} and \mathcal{S} the sets of port indices associated with the N_T transmitting antennas, the N_R receiving antennas, and the N_S RIS elements, respectively, where $N_A = N_T + N_R$. According to MNT [2], [26], the end-to-end wireless channel matrix \mathbf{H} is defined as

$$\mathbf{H} = \mathbf{S}_{\mathcal{R}\mathcal{T}} + \mathbf{S}_{\mathcal{R}\mathcal{S}} (\Phi^{-1} - \mathbf{S}_{\mathcal{S}\mathcal{S}})^{-1} \mathbf{S}_{\mathcal{S}\mathcal{T}}. \quad (1)$$

In (1), $\mathbf{S}_{\mathcal{R}\mathcal{T}}$ represents the paths from the transmitters to the receivers that did not encounter the RIS; $\mathbf{S}_{\mathcal{R}\mathcal{S}}$ represents the paths from the RIS elements to the receivers; $\mathbf{S}_{\mathcal{S}\mathcal{T}}$ represents the paths from the transmitters to the RIS elements; $\mathbf{S}_{\mathcal{S}\mathcal{S}}$ captures the MC between the RIS elements; Φ captures the scattering characteristics of the loads associated with a chosen RIS configuration.

Remark 1. We omit the noise term for notation convenience in (1) and the following equations, but, of course, it is naturally accounted for in our experimental validation where measurement noise is inevitable; we carefully quantify the experimental noise strength in Sec. IV-A and, moreover, we explicitly examine the influence of the noise strength by adding

⁵This statement also applies to channels parametrized by BD-RIS [26].

further synthetic noise to the measurements in Sec. IV-B (in particular, Fig. 5 therein).

The matrix inversion in (1) can be rewritten as infinite sum whose convergence is guaranteed by passivity [9], [14], [26]:

$$\mathbf{H} = \mathbf{S}_{\mathcal{RT}} + \mathbf{S}_{\mathcal{RS}} \left[\sum_{k=0}^{\infty} (\Phi \mathbf{S}_{\mathcal{SS}})^k \right] \Phi \mathbf{S}_{\mathcal{ST}}. \quad (2)$$

The k th term in the infinite series corresponds to the family of paths that interact with the RIS k times [26].

B. Mapping from Control Sequence to Load Characteristics

The wireless practitioner typically does *not* know Φ . Moreover, the diagonal entries c_k of Φ can typically *not* take arbitrary values only constrained by passivity. Instead, given a RIS with m -bit-programmable elements, the practitioner defines a RIS control sequence $\mathbf{w} \in \{1, 2, \dots, 2^m\}^{N_S}$ and the resulting configurable physical RIS properties are

$$\Phi = f(\mathbf{w}). \quad (3)$$

We assume that the available individual loads are the same for all RIS elements, and that their biasing is mutually independent. Thus, f simply defines the i th diagonal entry of Φ as the w_i th entry of the 2^m -element vector $\mathbf{s} \in \mathbb{C}^{2^m}$ containing the realizable reflection coefficients.

Remark 2. In general, the realized loads at distinct RIS elements could differ by design and/or due to component tolerances, and/or they could be coupled, e.g., due to a limited power supply of the control circuitry.

C. Simplified System Models

Besides the high-fidelity MNT model described in Sec. II-A, we consider three lower-fidelity simplified models for benchmarking.

1) *LFMNT*: Our low-fidelity MNT model (LFMNT) uses the same forward model as MNT for inference, but the entries of the off-diagonal blocks of $\mathbf{S}_{\mathcal{SS}}$ are estimated (as detailed in Sec. III-E1) based on a modified system model in which the infinite sum in (2) is truncated after $k = 1$ so that MC effects are only captured to first order:

$$\mathbf{H}^{(1)} \approx \mathbf{S}_{\mathcal{RT}} + \mathbf{S}_{\mathcal{RS}} \Phi \mathbf{S}_{\mathcal{ST}} + \mathbf{S}_{\mathcal{RS}} \Phi \mathbf{S}_{\mathcal{SS}} \Phi \mathbf{S}_{\mathcal{ST}}. \quad (4)$$

2) *CASC*: Truncating the infinite sum after $k = 0$ is equivalent to assuming that all entries of $\mathbf{S}_{\mathcal{SS}}$ vanish (which implies zero MC between RIS elements and perfect matching at each RIS element port). One recovers the widespread simplified cascaded model:

$$\mathbf{H}^{\text{CASC}} = \mathbf{S}_{\mathcal{RT}} + \mathbf{S}_{\mathcal{RS}} \Phi \mathbf{S}_{\mathcal{ST}}. \quad (5)$$

This affine mapping in (5) still imposes some structure inherited from (1). Specifically, it still imposes that changing the configuration of a single RIS element results in a rank-one update of the channel matrix. Interestingly, a phenomenological argument to the same effect has been described based on a “keyhole” intuition in [27], inspired by older seminal work on multiple-input multiple-output (MIMO) systems without RIS [28].

3) *LR*: A generic, unstructured, affine model is

$$\mathbf{H}^{\text{LR}} = \mathbf{A} + \left[\sum_{k=1}^{N_S} w_k \mathbf{B}_k \right], \quad (6)$$

parametrized by $\mathbf{A} \in \mathbb{C}^{N_R \times N_T}$ and $\mathbf{B}_k \in \mathbb{C}^{N_R \times N_T}$. This LR model neglects not only MC-induced non-linearities in the mapping from \mathbf{c} to \mathbf{H} but also potential non-linearities in the mapping from \mathbf{w} to \mathbf{c} . In the special case of $m = 1$, the mapping from \mathbf{w} to \mathbf{c} is always affine⁶; for $m > 1$, the mapping from \mathbf{w} to \mathbf{c} is generally *not* affine⁷.

III. AMBIGUITY-AWARE SEGMENTED PARAMETER ESTIMATION

In this section, we begin by precisely formulating our problem statement in Sec. III-A. Then, we provide a high-level overview of our algorithmic approach in Sec. III-B. We go on to detail our algorithmic steps in Sec. III-C. Next, we provide technical details about our gradient-descent implementation in Sec. III-D. Finally, we explain how we choose the parameters for our three simplified benchmark models in Sec. III-E.

A. Problem Statement

Given the ability to measure \mathbf{H} for any desired \mathbf{w} , we seek to estimate a set of parameters $\hat{\mathbf{S}}_{\mathcal{RT}}, \hat{\mathbf{S}}_{\mathcal{RS}}, \hat{\mathbf{S}}_{\mathcal{SS}}, \hat{\mathbf{S}}_{\mathcal{ST}}$, and $\hat{\mathbf{s}}$ such that we can use (3) and then (1) to map any \mathbf{w} to the corresponding \mathbf{H} . We use the tilde to emphasize that our estimated parameters are generally different from $\mathbf{S}_{\mathcal{RT}}, \mathbf{S}_{\mathcal{RS}}, \mathbf{S}_{\mathcal{SS}}, \mathbf{S}_{\mathcal{ST}}$, and \mathbf{s} due to the inevitable ambiguities.

For clarity, we reiterate that according to our system model and problem statement, *first*, the reflection coefficients of the tunable loads associated with the RIS elements can only take 2^m distinct values but not any arbitrary value within or on the unit circle, *second*, the available states of the tunable loads are the same for all RIS elements, *third*, the tunable loads can be tuned independently from each other, *fourth*, we do not know the characteristics of the available states of the tunable loads, *fifth*, we do not have any prior knowledge about \mathbf{S} other than that it is symmetric due to reciprocity and sub-unitarity due to passivity and inevitable absorption. These five points reflect the reality of our experiment described in Sec. IV.

B. Algorithmic Overview

The crux consists in estimating the RIS’s MC matrix $\tilde{\mathbf{S}}_{\mathcal{SS}}$. However, we also require matching estimates of $\tilde{\mathbf{S}}_{\mathcal{RT}}, \tilde{\mathbf{S}}_{\mathcal{RS}}, \tilde{\mathbf{S}}_{\mathcal{ST}}$, and $\tilde{\mathbf{s}}$. Our technique consists of one preliminary phase, followed by three main phases that tackle the estimation of the RIS’s MC matrix. An optional additional final phase can be used for fine-tuning. A high-level overview of our technique is provided in Algorithm 1.

Preliminary Phase: We fix $\tilde{\mathbf{S}}_{\mathcal{RT}}$ based on a single reference measurement; moreover, based on N_S measurements, we fix in closed-form $\tilde{\mathbf{S}}_{\mathcal{RS}}^T$ and $\tilde{\mathbf{S}}_{\mathcal{ST}}$ up to row-wise scaling factors.

⁶For $m = 1$, there always exist $\alpha, \beta \in \mathbb{C}$ such that $c_i = s_{w_i} = \alpha w_i + \beta$, e.g., $\alpha = s_2 - s_1$ and $\beta = 2s_1 - s_2$.

⁷For $m > 1$, $c_i = s_{w_i} = \alpha w_i + \beta$ is exact only if all s_{w_i} lie equally spaced on a straight line in the complex plane.

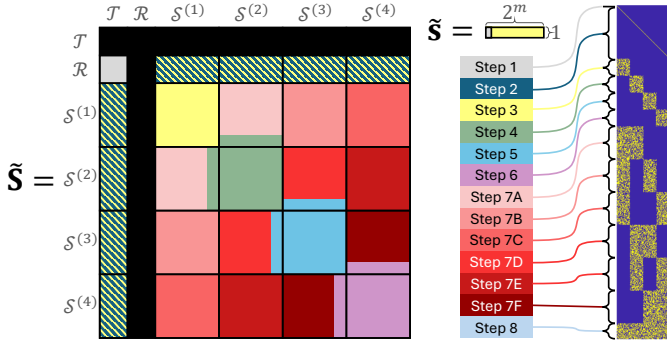


Fig. 1. Ambiguity-aware segmented MNT parameter estimation scheme. Colors indicate in what step a given segment is estimated. Black segments are not estimated. The estimation of hatched segments involves two steps. The right map illustrates the RIS configurations used in each step for a 1-bit programmable RIS whose reference state is colored blue. Step 8 is optional.

For the three main phases, we partition the RIS elements into groups.

Phase 1: We jointly estimate (i) the diagonal block of the MC matrix associated with the first group of RIS elements, (ii) the associated scaling factors for the rows of $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}}^\top$ and $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{T}}$ associated with the first group of RIS elements, as well as (iii) all unknown entries of $\tilde{\mathbf{s}}$.

Phase 2: For each remaining group of RIS elements in turn, we jointly estimate (i) the diagonal block of the MC matrix associated with that group of RIS elements, and (ii) the associated scaling factors for the rows of $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}}^\top$ and $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{T}}$ associated with that group of RIS elements.

Phase 3: For each pair of groups of RIS elements in turn, we estimate the corresponding off-diagonal block of the MC matrix.

Optional Final Phase: All estimated parameters are jointly fine-tuned.

The four phases must be conducted in this order, but the subproblems within Phase 2 and Phase 3 can be fully parallelized.

C. Detailed Algorithmic Steps

Preliminary Phase

Step 1: We measure $\mathbf{H}(\mathbf{w}_0)$, where $\mathbf{w}_0 = \mathbf{1}_{N_S}$, and define $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{T}} \triangleq \mathbf{H}(\mathbf{w}_0)$. This also implies that we fix $\tilde{s}_1 = 0$.

Step 2: We measure $\mathbf{H}(\mathbf{w}_i)$, where \mathbf{w}_i equals \mathbf{w}_0 except for the i th entry, and evaluate $\Delta_i = \mathbf{H}(\mathbf{w}_i) - \mathbf{H}(\mathbf{w}_0)$. We compute the SVD $\Delta_i = \mathbf{U}_i \Sigma_i \mathbf{V}_i^\dagger$ and identify the first left singular vector \mathbf{u}_i and the first right singular vector \mathbf{v}_i . We define $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}_i} \triangleq x_i \mathbf{u}_i$ and $\tilde{\mathbf{S}}_{\mathcal{S}_i\mathcal{T}} \triangleq y_i \mathbf{v}_i^\dagger$, where x_i and y_i are scaling factors that remain to be determined. We repeat this procedure for each RIS element in turn.

These first two steps are identical to [19]. MC does not appear to play an explicit role thus far because the rank-one property of updating a single RIS element's configuration is also found in the MC-unaware CASC model, as mentioned earlier in Sec. II-C2. However, it is the presence of MC that prevents us from using multi-flip updates with respect to \mathbf{w}_0 as done in the MC-unaware method in [27]. Indeed, the method in [27] cannot replace our Step 2 because in our problem

Algorithm 1: Ambiguity-Aware Segmented MNT Parameter Estimation

Preliminary phase:

1. Measure $\mathbf{H}(\mathbf{w}_0)$ with $\mathbf{w}_0 = \mathbf{1}_{N_S}$.
2. $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{T}} \leftarrow \mathbf{H}(\mathbf{w}_0)$ and $\tilde{s}_1 \leftarrow 0$.
3. **for** $i = 1, \dots, N_S$ **do**
 - (i) Measure $\mathbf{H}(\mathbf{w}_i)$ (where \mathbf{w}_0 and \mathbf{w}_i only differ in i th entry).
 - (ii) Compute the SVD $\mathbf{H}(\mathbf{w}_i) - \mathbf{H}(\mathbf{w}_0) = \mathbf{U}_i \Sigma_i \mathbf{V}_i^\dagger$.
 - (iii) \mathbf{u}_i [\mathbf{v}_i] \leftarrow dominant left [right] singular vector.
 - (iv) $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}_i} \leftarrow x_i \mathbf{u}_i$ and $\tilde{\mathbf{S}}_{\mathcal{S}_i\mathcal{T}} \leftarrow y_i \mathbf{v}_i^\dagger$;
 $x_i, y_i \in \mathbb{C}$ are still unknown.

end

Output of preliminary phase:

- $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{T}}$ fixed.
 - \tilde{s}_1 fixed.
 - $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}}$ and $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{T}}$ fixed up to scaling factors \mathbf{x} and \mathbf{y} .
-

Phase 1:

4. Measure $\mathbf{H}(\mathbf{w})$ for n_1 random configurations of RIS elements in $\mathcal{S}^{(1)}$ (all other RIS elements are in the same state as in \mathbf{w}_0).
5. Jointly estimate $\tilde{\mathbf{S}}_{\mathcal{S}^{(1)}\mathcal{S}^{(1)}}$, $\tilde{\mathbf{s}}$ (except \tilde{s}_1), $\mathbf{x}_{\mathcal{S}^{(1)}}$, and $\mathbf{y}_{\mathcal{S}^{(1)}}$ via gradient descent.

Output of Phase 1:

- $\tilde{\mathbf{S}}_{\mathcal{S}^{(1)}\mathcal{S}^{(1)}}$ fixed.
 - \tilde{s}_i fixed for $1 < i \leq 2^m$.
 - $\mathbf{x}_{\mathcal{S}^{(1)}}$ and $\mathbf{y}_{\mathcal{S}^{(1)}}$ fixed.
-

Phase 2:

6. **for** $i = 2, \dots, g$ **do**
 - (i) Measure $\mathbf{H}(\mathbf{w})$ for n_2 random configurations of RIS elements in $\mathcal{X}^{(i)} = \mathcal{S}^{(i)} \cup \mathcal{V}^{(i-1)}$ (all other RIS elements are in the same state as in \mathbf{w}_0).
 - (ii) Jointly estimate $\tilde{\mathbf{S}}_{\mathcal{X}^{(i)}\mathcal{X}^{(i)}}$ (except for already known entries), $\mathbf{x}_{\mathcal{S}^{(i)}}$, and $\mathbf{y}_{\mathcal{S}^{(i)}}$ via gradient descent.

end

Output of Phase 2:

- $\tilde{\mathbf{S}}_{\mathcal{S}^{(i)}\mathcal{S}^{(i)}}$ fixed for $1 < i \leq g$.
 - $\tilde{\mathbf{S}}_{\mathcal{S}^{(i)}\mathcal{V}^{(i-1)}}$ fixed for $1 < i \leq g$.
 - $\mathbf{x}_{\mathcal{S}^{(i)}}$ and $\mathbf{y}_{\mathcal{S}^{(i)}}$ fixed for $1 < i \leq g$.
-

Phase 3:

7. **foreach** unordered pair (k, l) with $1 \leq k < l \leq g$ **do**
 - (i) Measure $\mathbf{H}(\mathbf{w})$ for n_3 random configurations of RIS elements in $\mathcal{U}^{(kl)} = \mathcal{S}^{(k)} \cup \mathcal{S}^{(l)}$ (all other RIS elements are in the same state as in \mathbf{w}_0).
 - (ii) Get an initial estimate of $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}}$ (except already fixed entries) via multi-variable linear regression using (7).
 - (iii) Refine the estimate using gradient descent.

end

Output of Phase 3:

- $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}}$ (except already known entries) fixed for each unordered pair (k, l) with $1 \leq k < l \leq g$.
-

Output: $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{T}}$, $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}}$, $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{S}}$, $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{T}}$, $\tilde{\mathbf{s}}$.

(Optional) Global fine-tuning phase:

8. Measure $\mathbf{H}(\mathbf{w})$ for n_4 random RIS configurations.
9. Jointly refine $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{T}}$, $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}}$, $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{S}}$, $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{T}}$, and $\tilde{\mathbf{s}}$ using gradient descent.

Output of optional fine-tuning:

- Globally refined versions of $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{T}}$, $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}}$, $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{S}}$, $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{T}}$, and $\tilde{\mathbf{s}}$.
-

statement MC is not negligible, c_k cannot take any arbitrary value within or on the unit circle, and \mathbf{s} is not known.

Phase 1

Now, we deviate from [19] because we segment the estimation of $\tilde{\mathbf{S}}_{SS}$. We partition the N_S RIS elements into g non-overlapping groups of e RIS elements (except for the last group which contains $\lfloor \frac{N_S}{g} \rfloor$ elements); the set of port indices associated with the i th group is $\mathcal{S}^{(i)}$, such that $\bigcup_{i=1}^g \mathcal{S}^{(i)} = \mathcal{S}$. (See also Remark 6.) Without any prior knowledge about the RIS design or the radio environment (as assumed in this work), there is no preferential way in which RIS elements should be assigned into groups.

We introduce $\mathbf{x} = [x_1, x_2, \dots, x_{N_S}] \in \mathbb{C}^{N_S}$ and $\mathbf{y} = [y_1, y_2, \dots, y_{N_S}] \in \mathbb{C}^{N_S}$ for notational ease.

Step 3: We measure $\mathbf{H}(\mathbf{w})$ for n_1 known realizations of \mathbf{w} which equal \mathbf{w}_0 except for the entries associated with $\mathcal{S}^{(1)}$ which are chosen randomly. Using gradient descent, we jointly estimate $\tilde{\mathbf{S}}_{\mathcal{S}^{(1)}\mathcal{S}^{(1)}}$, $\tilde{\mathbf{s}}$ (except for its first entry that we fixed to 0 in Step 1), $\mathbf{x}_{\mathcal{S}^{(1)}}$, and $\mathbf{y}_{\mathcal{S}^{(1)}}$.

As illustrated in Fig. 1, this Step 3 identifies the top left diagonal block of $\tilde{\mathbf{S}}_{SS}$, and it simultaneously fixes the scaling factors on the associated rows and columns of $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{T}}$ and $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}}$, respectively, as well as the entries of $\tilde{\mathbf{s}}$ (except for the first one). It thus resembles Step 3 in [19] except that it is limited to $\mathcal{S}^{(1)}$ here instead of being applied globally to \mathcal{S} as in [19].

Phase 2

Step 4: We now seek to estimate $\tilde{\mathbf{S}}_{\mathcal{S}^{(2)}\mathcal{S}^{(2)}}$, and to simultaneously fix $\mathbf{x}_{\mathcal{S}^{(2)}}$, and $\mathbf{y}_{\mathcal{S}^{(2)}}$. We proceed similarly to Step 3 except for two modifications that are required to ensure compatibility with the ambiguous choices made in Step 3. On the one hand, we maintain $\tilde{\mathbf{s}}$ fixed after Step 3. On the other hand, instead of using random configurations of the RIS elements in $\mathcal{S}^{(2)}$ alone, we seek some overlap with the RIS elements in $\mathcal{S}^{(1)}$. Without any prior knowledge about the RIS design or the radio environment (as assumed in this work), there is no preferential way in which the overlapping RIS elements should be chosen. To that end, we define $\mathcal{X}^{(2)} = \mathcal{S}^{(2)} \cup \mathcal{V}^{(1)}$, where $\mathcal{V}^{(1)}$ is a set containing $e \gg v \geq 1$ distinct entries of $\mathcal{S}^{(1)}$ (see also Remark 7). We measure $\mathbf{H}(\mathbf{w})$ for n_2 known realizations of \mathbf{w} which equal \mathbf{w}_0 except for the entries associated with $\mathcal{X}^{(2)}$ which are chosen randomly. Using gradient descent, we jointly estimate $\tilde{\mathbf{S}}_{\mathcal{S}^{(2)}\mathcal{S}^{(2)}}$, $\tilde{\mathbf{S}}_{\mathcal{S}^{(1)}\mathcal{S}^{(2)}}$, $\tilde{\mathbf{S}}_{\mathcal{S}^{(2)}\mathcal{S}^{(1)}}$, $\mathbf{x}_{\mathcal{S}^{(2)}}$, and $\mathbf{y}_{\mathcal{S}^{(2)}}$.

Step (4 + i), where $1 \leq i < g$: We repeat Step 4 for the $(i+1)$ th diagonal block of $\tilde{\mathbf{S}}_{SS}$ by using an overlap with the v last RIS elements from $\mathcal{S}^{(i)}$. We jointly estimate $\tilde{\mathbf{S}}_{\mathcal{S}^{(i+1)}\mathcal{S}^{(i+1)}}$, $\tilde{\mathbf{S}}_{\mathcal{S}^{(i)}\mathcal{S}^{(i+1)}}$, $\tilde{\mathbf{S}}_{\mathcal{S}^{(i+1)}\mathcal{S}^{(i)}}$, $\mathbf{x}_{\mathcal{S}^{(i+1)}}$, and $\mathbf{y}_{\mathcal{S}^{(i+1)}}$. Alternatively, Step (4 + i) can be conducted using an overlap with the v last RIS elements from $\mathcal{S}^{(1)}$ instead of the v last RIS elements from $\mathcal{S}^{(i)}$. This alternative is appealing to maximize computational parallelization because it allows one to independently conduct the computational efforts of the $g-1$ steps that determine all but the first diagonal blocks of $\tilde{\mathbf{S}}_{SS}$.

Phase 3

Step (4 + g + i), where $1 \leq i \leq h$ and $h = \frac{g(g-1)}{2}$: We now seek to estimate the i th off-diagonal block of $\tilde{\mathbf{S}}_{SS}$ associated with the groups of RIS elements identified by $\mathcal{S}^{(k)}$ and $\mathcal{S}^{(l)}$, where $k \neq l$. We measure $\mathbf{H}(\mathbf{w})$ for n_3 known realizations

of \mathbf{w} which equal \mathbf{w}_0 except for the entries associated with $\mathcal{U}^{(kl)} = \mathcal{S}^{(k)} \cup \mathcal{S}^{(l)}$ which are chosen randomly. Using gradient descent, we estimate $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}} = \tilde{\mathbf{S}}_{\mathcal{S}^{(l)}\mathcal{S}^{(k)}}^\top$. If some entries of $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}}$ were already fixed in a previous step as part of the overlapping, we do not re-estimate them. Instead of initializing $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}}$ randomly for the gradient descent, we use a low-fidelity estimate obtained via multi-variable linear regression by assuming a truncation of (2) after $k=1$. The resulting expression

$$\mathbf{H}^{(1)} \approx \tilde{\mathbf{S}}_{\mathcal{R}\mathcal{T}} + \tilde{\mathbf{S}}_{\mathcal{R}\mathcal{U}^{(kl)}} \tilde{\Phi}_{\mathcal{U}^{(kl)}\mathcal{U}^{(kl)}} \tilde{\mathbf{S}}_{\mathcal{U}^{(kl)}\mathcal{T}} + \tilde{\mathbf{S}}_{\mathcal{R}\mathcal{U}^{(kl)}} \tilde{\Phi}_{\mathcal{U}^{(kl)}\mathcal{U}^{(kl)}} \tilde{\mathbf{S}}_{\mathcal{U}^{(kl)}\mathcal{U}^{(kl)}} \tilde{\Phi}_{\mathcal{U}^{(kl)}\mathcal{U}^{(kl)}} \tilde{\mathbf{S}}_{\mathcal{U}^{(kl)}\mathcal{T}} \quad (7)$$

is analogous to (4), and clearly affine in the sought-after $\tilde{\mathbf{S}}_{\mathcal{U}^{(kl)}\mathcal{U}^{(kl)}}$ of which $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}}$ is a block. Because \mathbf{H} is known from measurements and all other variables in (7) except for $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}}$ were fixed in previous steps, (7) can indeed be solved for $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}}$ via multi-variable linear regression given the n_3 matching pairs of \mathbf{H} and $\tilde{\Phi}$. Due to the approximate nature of (7), this yields only a low-fidelity estimate of $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}}$. However, this is nonetheless useful to initialize the gradient descent that seeks the high-fidelity estimate of $\tilde{\mathbf{S}}_{\mathcal{S}^{(k)}\mathcal{S}^{(l)}}$. Because these h steps aimed at retrieving off-diagonal blocks of $\tilde{\mathbf{S}}_{SS}$ are mutually independent, they can be fully parallelized.

Optional Final Phase

Final Step (optional): We measure $\mathbf{H}(\mathbf{w})$ for n_4 known, random realizations of \mathbf{w} . Using gradient descent, we jointly refine all previous estimates of $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{T}}$, $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}}$, $\tilde{\mathbf{S}}_{SS}$, $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{T}}$, and $\tilde{\mathbf{s}}$, which serve as initialization. If this optional final step for global fine-tuning is used, we denote the model with ‘‘MNTGFT’’; otherwise, we denote the model with ‘‘MNT’’.

The total number of measurements is

$$n_{\text{MNT}} = 1 + N_S + n_1 + (g-1)n_2 + \frac{g(g-1)}{2}n_3. \quad (8a)$$

$$n_{\text{MNTGFT}} = 1 + N_S + n_1 + (g-1)n_2 + \frac{g(g-1)}{2}n_3 + n_4. \quad (8b)$$

Remark 3. An affine initialization akin to the one used in Phase 3 to estimate the off-diagonal blocks of $\tilde{\mathbf{S}}_{SS}$ is not possible for any of the steps in Phase 1 or Phase 2 that estimate the diagonal blocks of $\tilde{\mathbf{S}}_{SS}$. The reason is that the steps in Phase 2 simultaneously fix the scaling factors associated with the corresponding blocks of $\tilde{\mathbf{S}}_{\mathcal{S}\mathcal{T}}$ and $\tilde{\mathbf{S}}_{\mathcal{R}\mathcal{S}}$. The step in Phase 1 identifying the first diagonal block moreover fixes all but the first of the entries of $\tilde{\mathbf{s}}$ that map into $\tilde{\Phi}$ via (3). Thus, the diagonal blocks are not the only unknowns in the steps that determine them. This prevents us from obtaining a low-fidelity estimate of a diagonal block by solving an approximate model originating from a truncation of (2) at $k=0$, akin to (5), via multi-variable linear regression.

Remark 4. For the sufficiently constrained, unambiguous parameter estimation underlying the related but distinct problem statement associated with the Virtual VNA technique, a purely closed-form version exists [22], [23]. Whether a similar purely closed-form version can be identified for our inevitably ambiguous problem statement in Sec. III-A is beyond the scope of this paper. We deliberately opt for gradient

descent here because it has proven to be more noise-robust in the Virtual VNA context [29]. However, it seems clear that without any a priori knowledge of \mathbf{s} , one requires gradient descent at least in Step 4 to identify $\tilde{\mathbf{s}}$; the case of $m = 1$ (1-bit-programmable RIS elements) may be an exception.

D. Gradient Descent Implementation

All gradient descents are implemented in TensorFlow. In each case, the cost \mathcal{C} to be minimized is defined as

$$\mathcal{C} = \left\langle \frac{\sum_{i,j} |h_{ij}^{\text{PRED}} - h_{ij}^{\text{MEAS}}|}{\sum_{i,j} |h_{ij}^{\text{MEAS}}|} \right\rangle, \quad (9)$$

where the superscripts PRED and MEAS indicate whether the channel gains are measured or predicted using (1), and $\langle \cdot \rangle$ indicates averaging over a batch of RIS configurations. Randomly initialized variables are drawn from a truncated normal distribution (mean: 0, standard deviation: 0.1). The batch size is 300 unless fewer examples are available. We use the Adam method for stochastic optimization with an initial step size of 10^{-3} that is gradually reduced by up to two orders of magnitude as a function of \mathcal{C} . We stop training after 4000 iterations and restore the parameter set that yielded the lowest seen cost.

E. Benchmark Models

We determine the parameters required by the three benchmark models introduced in Sec. II-C as follows.

1) *LFMNT*: We proceed as for the full MNT model except that we skip the gradient descents in Phase 3 that estimate off-diagonal blocks of the MC matrix. Instead, we use the initialization obtained via multi-variable linear regression as the final value.

2) *CASC*: We use the parameters estimated for the full MNT model except that we set all entries of $\tilde{\mathbf{S}}_{SS}$ to zero.

3) *LR*: We estimate the entries of \mathbf{A} and \mathbf{B}_k via multi-variable linear regression using the n_4 matching pairs of \mathbf{w} and \mathbf{H} measured in Step 8.

F. Computational Complexity

We now briefly characterize the *total* computational complexity and the *peak* computational complexity of the proposed ambiguity-aware segmented estimation scheme. Because we consider large RISs, $N_S \gg N_R$ and $N_S \gg N_T$. We assume that the inversion of the matrix $(\Phi^{-1} - \mathbf{S}_{SS}) \in \mathbb{C}^{N_S \times N_S}$ in (1) dominates the computational complexity with $\mathcal{O}(N_S^3)$. Each gradient descent iteration requires a forward evaluation and back-propagation of this computational complexity. Thus the total computational complexity of a gradient descent with I iterations that jointly estimates the entire MC matrix (as in [19], [25], [26]) has a *total* computational complexity of $\mathcal{O}(IN_S^3)$ and a *peak* computational complexity of $\mathcal{O}(N_S^3)$.

To keep our analysis tractable, and in line with our implementation for the present work, we assume that the

number of iterations I is fixed for all gradient descents. In our segmented approach, Phase 1 conducts a similar gradient descent but limited to a group of e RIS elements, implying a *total* computational complexity of $\mathcal{O}(Ie^3)$ and a *peak* computational complexity of $\mathcal{O}(e^3)$. Then, Phase 2 conducts $g - 1$ similar gradient descents limited to groups of $e + v$ RIS elements, each with a *total* computational complexity of $\mathcal{O}(I(e + v)^3)$ and a *peak* computational complexity of $\mathcal{O}((e + v)^3)$. Altogether, Phase 2 thus has a *total* computational complexity of $\mathcal{O}((g - 1)I(e + v)^3)$, while its *peak* computational complexity is $\mathcal{O}((e + v)^3)$. Finally, in Phase 3 we conduct $\frac{g(g-1)}{2}$ gradient descents for pairs of groups of RIS elements, each with a *total* computational complexity of $\mathcal{O}(I(2e)^3)$ and a *peak* computational complexity of $\mathcal{O}((2e)^3)$. Altogether, Phase 3 thus has a *total* computational complexity of $\mathcal{O}(\frac{g(g-1)}{2}I(2e)^3)$ and a *peak* computational complexity of $\mathcal{O}((2e)^3)$.

To summarize, the overall *peak* computational complexity of our segmented approach is thus $\mathcal{O}((2e)^3)$. With $e \approx N_S/g$, the *peak* computational complexity hence decreases cubically with g . One can thus choose g to ensure that the *peak* computational complexity does not exceed the limits of a given computational resource. This is the main impact of our segmented approach; indeed, existing approaches that jointly estimate all entries of the MC matrix [19], [25], [26] would simply fail if their *peak* computational complexity exceeded the limits of a given computational resource, implying that they could not be applied to large RIS.

Meanwhile, the overall *total* computational complexity of our segmented approach can be approximated as $\mathcal{O}(I(e^3 + (g - 1)(e + v)^3 + \frac{g(g-1)}{2}(2e)^3))$ which we can simplify assuming $e \approx N_S/g$ and $v \ll e$ to $\mathcal{O}(IN_S^3 \frac{4g-3}{g^2})$. This simple formula suggests that with $g \geq 4$ the *total* computational complexity of our segmented approach can actually drop below that of the non-segmented approach, but note that g cannot be chosen arbitrarily high (g cannot exceed N_S , and if g approaches N_S , the assumption $v \ll e$ cannot be satisfied); moreover, the number of required iterations I may vary with g but an investigation thereof is deferred to future work.

IV. EXPERIMENTAL VALIDATION

In this section, we describe in Sec. IV-A our experimental setup; then, we report in Sec. IV-B the experimental validation and analysis of our parameter estimation technique described in Sec. III.

A. Experimental Setup

Our RIS prototype is displayed in Fig. 2 and comprises 225 1-bit programmable RIS elements of which we use 100 in this paper; the remaining ones are kept in their reference state throughout all experiments. Specifically, each RIS element is a half-wavelength patch equipped with a PIN diode whose bias voltage we control individually. The inter-element spacing is roughly half a wavelength. The RIS elements' reflection coefficients under normal incidence associated with the two possible states roughly have a phase difference of π at 2.45 GHz. RIS design details are discussed in [30].

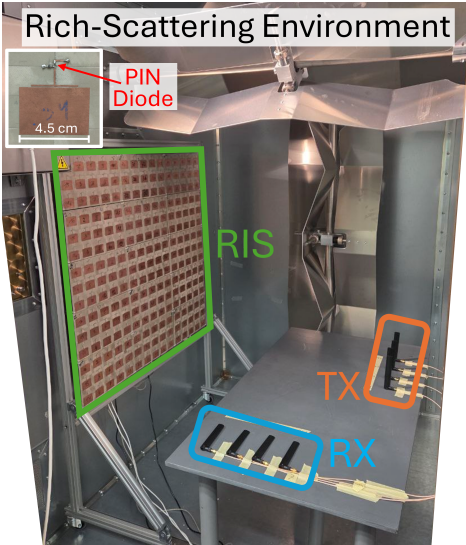


Fig. 2. Experimental setup (inset shows RIS element).

The pivotal technical detail of our RIS design that is relevant to the applicability of the MNT model described in Sec. II is that the size of the PIN diode is tiny compared to the wavelength (see inset in Fig. 2). Indeed, it is the PIN diode that we describe as lumped port terminated by a tunable load. Meanwhile, the size of the RIS element being on the order of half a wavelength and thus not electrically small is irrelevant for the applicability of the MNT model. Indeed, the RIS element void of the PIN diode constitutes a static scatterer that is treated as part of the radio environment. A direct consequence of our system model is that Δ_i from Step 2 is a rank-one matrix [22], [31], [32]. Experimentally, we have verified that this is approximately true by confirming that the ratio between the first and second singular value of Δ_i is large (greater than ten).

Our transmit and receive arrays each comprise four parallel, half-wavelength-spaced antennas (ANT-W63WS2-SMA). As seen in Fig. 2, they are placed alongside our RIS inside a reverberation chamber ($1.75 \times 1.5 \times 2 \text{ m}^3$) which constitutes an unknown, static, rich-scattering environment. To minimize the direct line-of-sight channel, the transmitting and receiving antennas are oriented perpendicular to each other. We use an eight-port VNA (two cascaded Keysight P5024B 4-port VNAs) to measure \mathbf{H} at our operating frequency of 2.45 GHz.

We conduct our experiment under well-controlled conditions: The signal-to-noise ratio (SNR) is 55.3 dB (estimated based on repeated measurements in quick succession of \mathbf{H} for the same \mathbf{w}) and the stability is 45.8 dB (estimated like the SNR, but based on measurements taken intermittently over the course of our experiment).

We emphasize that we only have minimal a priori knowledge about our experimental system, namely that all components are reciprocal and passive, and that all tunable loads are identical and have two possible states. We do not assume to know the two states' characteristics, nor the RIS element design, nor any specifics about our rich-scattering radio environment.

Remark 5. *The consideration of an RIS prototype with 1-bit-programmable RIS elements should not be mistakenly interpreted as a limitation of our algorithm to the special case of $m = 1$. Neither our problem statement in Sec. III-A nor the description of our algorithm in Sec. III-C assumes $m = 1$. In Step 3, all entries of $\tilde{\mathbf{s}}$ except the first one (that is always fixed to 0 in Step 1) are estimated. As mentioned in Sec. III-C, this step resembles Step 3 in [19], where it was verified experimentally for an RIS prototype with 6-bit-programmable elements (i.e., $m = 6$).*

B. Experimental MNT Model Parameter Estimation

We now experimentally validate the algorithm described in Sec. III for our experimental 4×4 MIMO channel parametrized by our RIS prototype with 100 1-bit-programmable elements, i.e., with $N_R = N_T = 4$, $N_S = 100$, and $m = 1$. The number of unknown, complex-valued, scalar MNT model parameters to be estimated is

$$n_u = N_R N_T + (N_R + N_T) N_S + \frac{N_S(N_S + 1)}{2} + (2^m - 1), \quad (10)$$

which evaluates to $n_u = 5867$ for our problem.

We partition the $N_S = 100$ RIS elements into $g = 4$ groups of $e = 25$ elements. We conduct the required measurements following the description in Sec. III with $n_1 = n_2 = \frac{e(e+1)}{2} = 325$, $n_3 = e^2 = 625$, $n_4 = 300$, and $v = 4$ (see Remark 6 and Remark 7 below). Because the measurement speed is the bottleneck in our experiments, we cannot exploit the full parallelizability within Phase 2 and Phase 3. Instead, we conduct the computations for any given subproblem as soon as the required data is available, while measurements for other subproblems may still be ongoing. For Phase 2, we choose a sequential overlap as illustrated in Fig. 1. The values of n_1 , n_2 , and n_3 are chosen generously so that we can test our algorithm under ideal conditions without being limited by the available data. We explore the influence of limited data by only using a fraction p of the available measurements for each segment after the preliminary phase; when p is below unity, we thus have $n_{\text{MNT}} = 1 + N_S + p(n_1 + (g-1)n_2 + \frac{g(g-1)}{2}n_3)$. The larger value n_3 (compared to $n_1 = n_2$) reflects the larger number of unknowns in off-diagonal blocks due to reciprocity. We furthermore measure \mathbf{H} for $n_5 = 30$ previously unseen random RIS configurations to assess the accuracy of our models.

Remark 6. *As discussed in Sec. III-F, the available computational resources impose a strict lower limit on the possible values of g , because for values of g below that limit the peak computational complexity would be too high. This is the primary consideration in choosing g . Secondary considerations can account for the realizable amount of parallelization given the available computational resources, as well as for the strength of the measurement noise. If the computational resources enable arbitrary parallelization, choosing a high value of g will reduce the computation time because the per-subproblem total computational complexity in Phase 2 and Phase 3 decreases as g increases. Another influential factor is the strength of the measurement noise. Lower values of g can improve noise robustness because they*

result in stronger variations of the measured realizations of \mathbf{H} . Moreover, higher values of v improve the robustness of the ambiguity alignment across groups against noise; however, higher values of v also slightly increase the computational complexity (see Sec. III-F). Thus, v should be chosen as low as possible while ensuring that robust ambiguity alignment is still achieved in the presence of the given measurement noise. Finally, it is conceivable that the value of g influences the smoothness of the parameter landscapes when the subproblems are smaller. Systematic investigations of these secondary considerations are beyond the scope of this paper. In our experiment with $N_S = 100$, $g = 4$ is a reasonable choice to validate our technique.

Remark 7. v should be significantly smaller than e or else the segmentation is effectively coarser than suggested by the value of g . At the same time, v should be sufficiently large to allow for a robust alignment of the ambiguities. While $v = 1$ is sufficient under ideal conditions, it may be insufficient if the relevant RIS element happens to only weakly influence the channel or be defective. In our experiment with $e = 25$, $v = 4$ is a reasonable choice.

To quantify the model accuracy, we use the following metric that is defined similar to an SNR, treating the model error as “noise”:

$$\zeta = \frac{\text{SD} [h_{ij}^{\text{MEAS}}(\mathbf{w})]}{\text{SD} [h_{ij}^{\text{MEAS}}(\mathbf{w}) - h_{ij}^{\text{PRED}}(\mathbf{w})]}, \quad (11)$$

where SD denotes the standard deviation across all entries of \mathbf{H} and the $n_5 = 30$ random, unseen realizations of \mathbf{w} ; the superscripts denote the experimentally measured ground truth (MEAS) and the model’s prediction (PRED).

With sufficient measurements, our technique from Sec. III achieves an accuracy of 40.5 dB (see MNT curve for $p = 1$ in Fig. 3). The amount of available data can be reduced to $p = 0.4$ without any notable effect on the accuracy, and even to $p = 0.2$ without any strong effect on the accuracy. Below $p = 0.2$, the accuracy drops rapidly as p is reduced further. It hence appears that the parameter estimation problem is ill-posed for $p < 0.2$. At $p = 0.2$, the total number of measured complex-valued scalars $N_R N_T n_{\text{MNT}}$ is roughly three times the number of unknown complex-valued scalar MNT parameters. We hypothesize that this factor of three would drop close to unity only in the absence of all experimental and numerical noise. Due to the non-linear nature of our system model in (1) the relatively low measurement noise and instabilities in our experiment are seemingly not negligible.

We see in Fig. 3 that the MNT accuracy drops monotonically as p is decreased (except for the regime of $p < 0.1$ in which the problem is severely ill-posed). The same is not true for MNTGFT which does *not* offer consistent gains compared to MNT. Meanwhile, the LFMNT model, aware of MC only to first order, only achieves accuracy up to 16.9 dB. The MC-unaware CASC model only achieves accuracies up to 13.8 dB. These results highlight the importance of capturing MC effects with high fidelity for accurate modelling. The

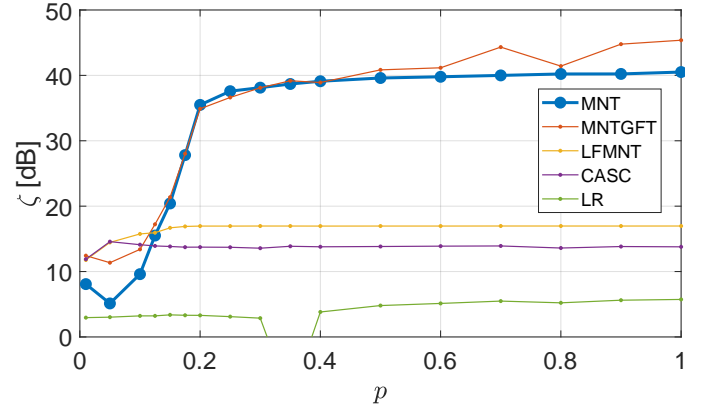


Fig. 3. Model accuracy metric ζ defined in (11) as a function of the fraction p of the available data used to estimate the model parameters, for the five considered cases.

completely unstructured LR model’s accuracy does not exceed 5.7 dB.

Next, we investigate the influence of the dimensions of the MIMO system on the achieved accuracy. By ignoring parts of the measured 4×4 MIMO channel, we can consider scenarios with a 3×3 channel, a 2×2 channel, and a 1×1 channel (which is a single-input single-output (SISO) scenario). For smaller MIMO systems, n_u is slightly lower, as seen in (10); however, n_u is dominated by the fixed size of the $\hat{\mathbf{S}}_{SS}$ block for large RIS like ours. At the same time, because each measurement yields $N_R N_T$ independent complex-valued scalars, the number of independent pieces of information acquired per measurement drops quadratically with the MIMO system dimension. The latter effect dominates the former, which is why we observe in Fig. 4 that more measurements are required for smaller MIMO systems to estimate the MNT parameters with comparable accuracy. Specifically, the 3×3 MIMO system reaches a comparable accuracy of 37.0 dB only at $p = 0.4$ (with marginal benefits of even larger p , reaching an accuracy of 39.7 dB at $p = 1$), and the 2×2 MIMO system reaches a comparable accuracy of 39.2 dB only at $p = 1$. For the SISO system, the parameter estimation appears to be always ill-posed for the range of considered values of p , and the accuracy never even reaches 10 dB. The ratio of the number of required measurements to n_u increases as the MIMO system dimension decreases, from 2.8 for 4×4 , via 3.3 for 3×3 , to 3.8 for 2×2 . We hypothesize that there is no easy way of predicting the minimal number of required measurements due to the complicated interplay between noise and the system model’s non-linearity.

Finally, we now examine the influence of measurement noise on the accuracy. Our segmented approach implies a complicated propagation of inaccuracies across its multiple phases. While a theoretical analysis of this error propagation is beyond the scope of this paper, we empirically evaluate the cumulative effect of this error propagation in Fig. 5 to gain qualitative insights. We consider for simplicity $N_R = N_T = 4$ and $p = 1$. To increase the noise level, we synthetically add complex-valued Gaussian noise to the measured channel matrices before applying our technique from Sec. III. The

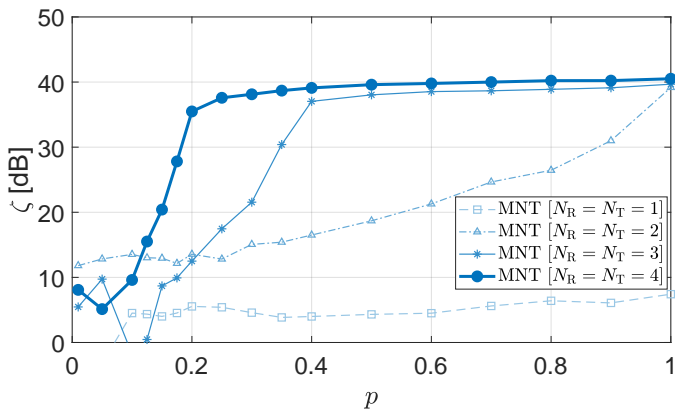


Fig. 4. MNT model accuracy as a function of the fraction p of the available data used to estimate the model parameters, for four differently dimensioned MIMO scenarios (MIMO collapses to SISO for $N_R = N_T = 1$).

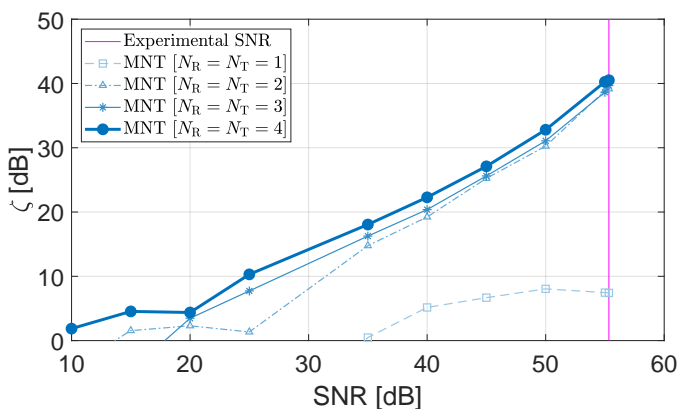


Fig. 5. MNT model accuracy as a function of the measurement SNR, for four differently dimensioned MIMO scenarios (MIMO collapses to SISO for $N_R = N_T = 1$) with $p = 1$.

largest considered value of SNR is the one imposed by our experimental measurement noise (i.e., without adding any synthetic noise). We expect our algorithm to be generally quite sensitive to noise because it essentially extracts the MNT model parameters from (minute) changes of the MIMO channel as a function of the RIS configuration. We observe indeed in Fig. 5 that our accuracy metric ζ [dB] scales roughly linearly with the SNR [dB]. The trend is very similar for all three considered MIMO systems, which reached similarly high accuracies at $p = 1$ in Fig. 4. In principle, the use of more measurements (i.e., larger values of p) can flexibly counteract the detrimental effect of measurement noise; to what extent such an approach is practical will depend on use-case specifics. Moreover, increasing the number v of overlapping RIS elements in Phase 2 can reduce the vulnerability to noise of the ambiguity alignment achieved between the groups.

V. EXPERIMENTAL MNT-BASED PERFORMANCE EVALUATION

We have observed substantial accuracy differences between the considered, experimentally calibrated models in Sec. IV-B. However, it remains unclear to what extent wireless

practitioners require the high MNT model accuracy. To answer this question, we perform extensive experimental performance evaluations in this section. We optimize the RIS configuration for five distinct end-to-end communications performance indicators based on the considered, experimentally calibrated models. Then, we experimentally measure the wireless channel matrix for each optimized RIS configuration. We examine to what extent the gap in model accuracy translates into a gap in end-to-end performance. We also examine to what extent the models faithfully predict the achieved performances.

This section constitutes the first experimental MNT-based performance evaluation for a large RIS. In Sec. V-A, we define the five considered key performance indicators (KPIs). In Sec. V-B, we describe our MC-aware discrete RIS optimization algorithm. In Sec. V-C, we present and discuss our experimental end-to-end performance evaluation.

A. Key Performance Indicators (KPIs)

We consider two SISO KPIs and three MIMO KPIs. These five KPIs are defined as follows.

SISO KPI 1: We aim to maximize the channel gain $|h_{ij}|^2$. This is a commonly considered KPI in theoretical papers on RIS [1], [3], [7]. Albeit on the antenna level, this KPI directly correlates with numerous end-to-end SISO communications metrics (e.g., capacity, bit-error rate, throughput).

SISO KPI 2: We assume that the i th TX communicates with the i th RX, such that the signals from the j th TX to the i th RX, where $j \neq i$, constitute undesired interferences. The rate achieved by the i th TX is hence $R_i = \log_2 \left(1 + P_T |h_{ii}|^2 / (\sum_{j \neq i} P_T |h_{ij}|^2 + \sigma^2) \right)$, where we assume that all TX transmit with power P_T ; σ quantifies the noise strength. This scenario represents a SISO interference channel. Our KPI is R_i , assuming $P_T/\sigma^2 = 100$ dB.

MIMO KPI 1: We aim to maximize the capacity C at low SNR by dominant eigenmode transmission. In this regime, $C = \log_2 (1 + P_T \|\mathbf{H}\|^2 / \sigma^2)$ so that maximizing C is equivalent to maximizing the spectral norm of the channel matrix: $\|\mathbf{H}\|^2$. Hence, our KPI is $\|\mathbf{H}\|^2$.

MIMO KPI 2: We assume the same scenario as in SISO KPI 2 but seek to maximize the sum rate as opposed to an individual rate. Our KPI is $R = \sum_i R_i$, assuming $P_T/\sigma^2 = 100$ dB.

MIMO KPI 3: We aim to maximize the capacity C at high SNR by multi-eigenmode transmission with uniform power allocation. In this regime, $C = \log_2 (\det(\mathbf{I}_{N_R} + P_T \mathbf{H} \mathbf{H}^\dagger / \sigma^2))$. Hence, our KPI is C , assuming $P_T/\sigma^2 = 100$ dB.

Given the 4×4 MIMO channel in our experiment described in Sec. IV (see Fig. 2), we can define 16 different versions of SISO KPI 1 and four different versions of SISO KPI 2. Thus, we separately optimize the RIS configuration for each version in turn, and we report the average of the optimized KPI over all versions in these two cases.

B. MC-Aware Discrete RIS Optimization

Optimizing the RIS configuration based on a system model to maximize a given KPI is typically complicated. The

problem is *high-dimensional* for large-scale RIS; the problem is *non-linear* due to MC governing the mapping from RIS configuration to channel, and the mapping from channel to KPI is usually also non-linear; the problem is *discrete* because only a few loads are available at each RIS element (two for the RIS elements in our prototype); the problem is also almost always *non-convex*. A closed-form solution only exists for a specific KPI under the assumption of an idealized fully-connected BD-RIS without any hardware constraints [10]. Existing theoretical works on MC-aware RIS optimization assume that the reflection coefficients of the tunable lumped elements can take any value on or within the unit circle [1]–[12]. Hence, these algorithms cannot be applied to our RIS prototype with 1-bit programmable elements. MNT-based discrete optimization has received little theoretical attention. Numerical and experimental works like [13], [33] facing discrete constraints in RIS optimization opted for coordinate-descent algorithms. Our recent theoretical study in [34] suggests that coordinate descent indeed yields a good trade-off between performance and computational cost.

In light of these considerations, we use the same coordinate-descent algorithm to optimize our RIS based on the five considered models (MNT, MNTGFT, LFMNT, CASC, LR) for the five considered KPIs defined in Sec. V-A. In each case, we define the objective function \mathcal{O} to be minimized as the negative of the considered KPI. As initialization, we choose the configuration out of the n_{MNT} experimentally measured ones that yields the smallest objective function. Then, we loop over the RIS elements until the objective function has not decreased for N_S iterations. Algorithm 2 summarizes our optimization procedure.

Algorithm 2: MC-Aware Binary RIS Optimization.

```

Evaluate the objective functions  $\{\mathcal{O}\}_{\text{init}}$  for the  $n_{\text{MNT}}$ 
experimentally measured RIS configurations  $\{\mathbf{c}\}_{\text{init}}$ .
Select  $\mathcal{O}_{\text{curr}}$  as the minimum from  $\{\mathcal{O}\}_{\text{init}}$  and  $\mathbf{c}_{\text{curr}}$  from  $\{\mathbf{c}\}_{\text{init}}$ 
at the same index such that  $\mathcal{O}_{\text{curr}} = \mathcal{O}(\mathbf{c}_{\text{curr}}) = \min(\{\mathcal{O}\}_{\text{init}})$ .
 $k \leftarrow 0, i \leftarrow 0$ 
while  $k < N_S$  do
     $i \leftarrow i + 1$ 
     $j \leftarrow ((i - 1) \bmod N_S) + 1$ 
     $\mathbf{c}_{\text{temp}} \leftarrow \mathbf{c}_{\text{curr}}$  with  $j$ th entry flipped.
     $\mathcal{O}_{\text{temp}} \leftarrow \mathcal{O}(\mathbf{c}_{\text{temp}})$ .
    if  $\mathcal{O}_{\text{temp}} < \mathcal{O}_{\text{curr}}$  then
         $\mathbf{c}_{\text{curr}} \leftarrow \mathbf{c}_{\text{temp}}$ 
         $\mathcal{O}_{\text{curr}} \leftarrow \mathcal{O}_{\text{temp}}$ 
         $k \leftarrow 0$ 
    else
         $k \leftarrow k + 1$ 
    end
end
Output:  $\mathbf{c}_{\text{curr}}$  and  $\mathcal{O}_{\text{curr}}$ .

```

C. Experimental Performance Evaluation Results

The KPIs achieved with the optimized RIS configurations are summarized in Table I and visualized in Fig. 6. Specifically, for each KPI we provide the value obtained on average over 300 random configurations (“RAND”), as well as with a configuration optimized based on each of the five considered models. We use the model parameters obtained in the calibrations with $p = 1$ and $N_R = N_T = 4$. Both

the model-based prediction of the KPI (“PRED”) and the corresponding experimentally measured KPI (“MEAS”) are provided.

For MNT and MNTGFT, we observe a very close agreement between prediction and measurement for all five KPIs. This observation not only reflects the high model accuracies already seen in Sec. IV-B but also testifies to the models’ generalizability. Indeed, the models were calibrated based on measurements corresponding to random RIS configurations while the optimized RIS configurations that maximize a KPI may be distributional outliers. Accurate predictions of KPIs are very important for wireless practitioners to inform decisions on resource allocation; a simple example would be which wireless link(s) to enhance with a RIS in a multi-user scenario. In contrast, we see notable discrepancies between prediction and measurement for the three low-fidelity models which tend to significantly overestimate the achieved KPIs.

The achieved performance gains relative to the RAND benchmark vary significantly across the five considered KPIs. Based on our experimentally calibrated MNT model, we achieve a 6.8-fold and 9.6-fold improvement over RAND for SISO KPI 1 and SISO KPI 2, respectively. For MIMO KPI 2, the MNT-based improvement over RAND is 4.5-fold. For MIMO KPI 1, the MNT-based improvement over RAND is 87.0 %. For MIMO KPI 3, the MNT-based improvement over RAND is only 5.2 %. The small improvement for MIMO KPI 3 makes sense because the rich-scattering conditions combined with the assumed SNR of 100 dB are already highly ideal conditions on which it is hard to improve substantially by tuning the channel matrix with the RIS. Generally, it appears that the SISO KPIs depend more strongly on the RIS configuration and are thus more amenable to benefiting from RIS optimization.

The results with the MNTGFT model reveal no substantial performance benefits originating from the final optional global fine-tuning of the model parameters. In fact, surprisingly, we find that the achieved MIMO KPI 1 and MIMO KPI 2 are actually slightly inferior. However, the MNTGFT model does predict the RIS configurations optimized with the MNT model to be better, revealing that the simple coordinate-descent algorithm simply got stuck in local minima in these two cases. These are thus artefacts originating from inefficient model-based optimization as opposed to model inaccuracies. Meanwhile, the results of LFMNT perfectly coincide with those of MNT because the optimized RIS configuration was the same for all five KPIs. We observe thus no difference between the performances achieved with MNT, MNTGFT and LFMNT across five diverse KPIs. Nonetheless, only MNT and MNTGFT accurately predict the KPIs achieved based on them.

The performances achieved with the MC-unaware CASC and LR models are still substantial but clearly inferior to those achieved with MNT, MNTGFT or LFMNT. Compared to RAND, CASC and LR still achieve roughly a 5.7-fold performance improvement for SISO KPI 1. CASC resp. LR achieve a 4.7-fold resp. 3.9-fold improvement for SISO KPI 2, a 73.2 % resp. 65.7 % improvement for MIMO KPI 1, a 2.2-fold resp. 1.6-fold improvement for MIMO KPI 2, and a 4.5 % resp. 4.9 % improvement for MIMO KPI 3.

TABLE I

EXPERIMENTALLY MEASURED KPIS WITH RIS CONFIGURATIONS OPTIMIZED TO MAXIMIZE THE KPIS BASED ON THE CONSIDERED, EXPERIMENTALLY CALIBRATED MODELS. THE MODELS' PREDICTIONS FOR THE MAXIMIZED KPI VALUES ARE INDICATED IN GRAY FONT. RAND CORRESPONDS TO THE AVERAGE OF THE MEASURED KPI VALUE ACROSS 300 RANDOM RIS CONFIGURATIONS. THE DATA IS VISUALIZED IN FIG. 6.

KPI \ Method	RAND	MNT		MNTGFT		LFMNT		CASC		LR	
	MEAS	MEAS	PRED	MEAS	PRED	MEAS	PRED	MEAS	PRED	MEAS	PRED
$ h_{ij} ^2 \times 10^3$	0.76	5.15	5.16	5.16	5.18	5.15	5.61	4.38	6.15	4.38	5.51
R_i [bits/s/Hz]	0.59	5.67	5.71	5.94	6.09	5.67	3.64	2.78	7.22	2.28	2.83
$\ \mathbf{H}\ ^2 \times 10^3$	8.40	15.70	15.70	15.28	15.26	15.70	16.35	14.54	17.52	13.91	21.33
$\sum_i R_i$ [bits/s/Hz]	2.37	10.75	10.71	8.91	8.85	10.75	6.80	5.31	12.07	3.86	5.81
C [bits/s/Hz]	97.64	102.72	102.72	102.76	102.76	102.72	102.72	102.05	103.18	102.44	104.39

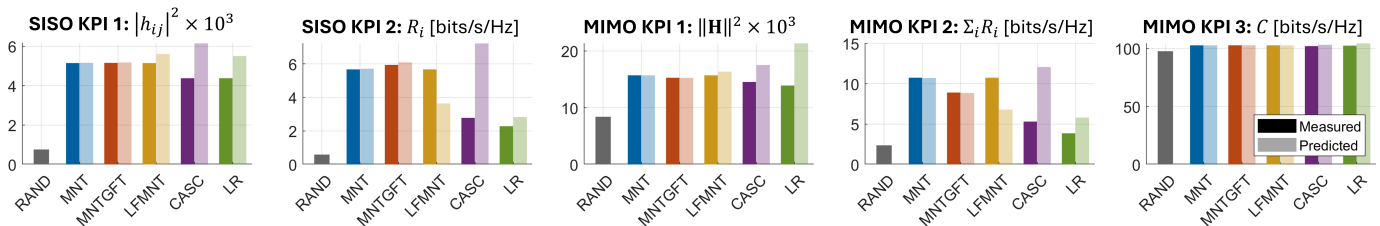


Fig. 6. Visualization of Table I.

These substantial improvements are surprising given the low accuracies observed in Sec. IV-B for the CASC and LR models. We hypothesize that the apparent robustness to model inaccuracies originates from the 1-bit constraint, in line with similar recent observations for small-scale RIS in [18], [19].

Altogether, models with limited or no MC awareness perform surprisingly well compared to MC-aware models in terms of the achieved KPIS. This observation questions whether the additional effort required to estimate the MC parameters is worthwhile. However, these lower-fidelity models predict the achieved KPIS rather inaccurately, which can be problematic for wireless practitioners facing resource allocation trade-offs.

VI. DISCUSSION

Prior to this article, there were three important open questions from the experimental perspective on the role of MC awareness in RIS optimization:

- 1) Do the benefits of MC awareness in RIS optimization found in theoretical studies manifest in experiments?
- 2) Is it possible at all to experimentally acquire MC awareness for large RIS?
- 3) Can the benefits outweigh the cost of experimentally acquired MC awareness?

Our key findings are summarized in the following answers:

- 1) The benefits in terms of achieved KPIS appear to be moderate, at least for few-bit-programmable RIS with half-wavelength inter-element spacing. However, the model-based predictions of the achieved KPIS are substantially more accurate with MC awareness.
- 2) Yes.
- 3) Possibly, e.g., in scenarios requiring accurate KPI predictions for resource allocation.

Given recent enthusiasm for MC awareness in RIS optimization, our rather nuanced conclusions in 1) and 3) are possibly surprising. Yet, or maybe precisely for

that reason, they constitute valuable takeaways for wireless practitioners, potentially saving them substantial calibration efforts dedicated to estimating the MC matrix.

The quadratic scaling of the number of unknowns, and thus the number of required measurements, with the number of RIS elements is a fundamental property of the MNT model. As such, it is not possible nor our intention to fundamentally modify this scaling. Instead, we accommodate the consequences of this scaling: our segmentation ensures the feasibility of the required computations for each segment. Without segmentation, the computational requirements would grow quadratically with the number of RIS elements and inevitably overwhelm any given computational resource for a sufficiently large RIS. Meanwhile, we do not expect our segmented approach to fundamentally affect the order of magnitude of the number of required measurements because the number of unknowns remains unchanged; quantitative evaluations of the measurement overhead for different hyperparameter combinations are deferred to future work.

Looking forward, we do not anticipate our method to be used in its current form once per coherence time in RIS-parametrized radio environments. Instead, we envision the following three use cases:

- 1) *Academic research*: Rigorous benchmarking of simplified lower-fidelity models requires well-controlled experiments with access to a high-fidelity ground-truth MC-aware model that our technique provides.
- 2) *One-off initialization*: In dynamic RIS-parametrized radio environments, many scattering objects are not dynamic. Thus, our technique can be applied once and the resulting parameters can subsequently serve as initialization for fine-tuning during each coherence time. It may furthermore be possible to only refine the most relevant parameters during each coherence time.
- 3) *Non-dynamic use cases*: Certain instances of massively parametrized wave systems are distinguished by the absence of uncontrolled perturbations. Examples include

RIS-parametrized wireless networks-on-chip [13], dynamic metasurface antennas (DMAs), and wave-domain physical neural networks (PNNs). In these cases, it is sufficient to apply our technique once during a one-off system model calibration.

Dynamic use cases require further research to limit the measurement overhead during each coherence time. Fundamentally, such reductions will likely require the use of prior knowledge. This can include prior knowledge about the RIS design and/or radio environment, prior knowledge of approximately suitable MNT parameters from a previous coherence time (as mentioned above), or assumptions of sparse representations (e.g., reduced-rank representations of the link between RIS and transceivers [35], [36] and/or of the RIS' MC matrix [16]).

VII. CONCLUSION

To summarize, we reported the first experimental characterization of MC for a large-scale RIS prototype, and its subsequent use for MNT-model-based RIS optimization in diverse performance evaluations. To facilitate the required estimation of the 5867 MNT model parameters, we introduced and experimentally validated an ambiguity-aware segmentation of the estimation problem. Our systematic analysis revealed the importance of high-fidelity MC characterization to achieve high model accuracies. We further evidenced that larger MIMO systems enable higher model accuracies with fewer measurements, and that the model accuracy [dB] scales roughly linearly with the SNR [dB]. The substantial accuracy differences between models with full, partial or no MC awareness did *not* translate into proportionate performance gain differences. Even MC-unaware models achieved strong performance improvements. Nonetheless, the important ability to predict the achieved performances based on the model did strongly correlate with the model accuracy.

Looking forward, *on the algorithmic side*, we encourage further investigations beyond the scope of this study to identify systematic guidelines for choosing the hyperparameters (g , v , n_1 , n_2 , n_3) of our proposed technique. Alternative solvers for the individual subproblems can also be explored. Furthermore, identifying data-driven or prior-knowledge-based routes to determining the most and/or least influential entries of the MC matrix are of interest. Moreover, efficient updates of outdated model parameters will be relevant for dynamic use cases. Finally, it appears feasible to modify our technique to cope with purely non-coherent detection [22], [23], [25], in order to alleviate the hardware requirements. Meanwhile, *on the hardware side*, it will be interesting to determine whether the observed discrepancies between model accuracy and wireless performance generalize to RIS with continuously tunable and/or more closely spaced elements. Beyond RIS, we expect our technique to find applications for DMAs and PNNs.

ACKNOWLEDGMENT

The author acknowledges I. Ahmed, F. Boutet, and C. Guitton, who, under the author's supervision, previously built the RIS prototype for the work presented in [30].

The author further acknowledges stimulating discussions with L. Le Magoarou.

REFERENCES

- [1] 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 Wirel. Commun. Lett.*, vol. 10, no. 5, pp. 938–942, 2021.
- [2] M. Nerini *et al.*, "A universal framework for multiport network analysis of reconfigurable intelligent surfaces," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 10, pp. 14 575–14 590, 2024.
- [3] X. Qian and M. Di Renzo, "Mutual coupling and unit cell aware optimization for reconfigurable intelligent surfaces," *IEEE Wirel. Commun. Lett.*, vol. 10, no. 6, pp. 1183–1187, 2021.
- [4] A. Abrardo *et al.*, "MIMO interference channels assisted by reconfigurable intelligent surfaces: Mutual coupling aware sum-rate optimization based on a mutual impedance channel model," *IEEE Wirel. Commun. Lett.*, vol. 10, no. 12, pp. 2624–2628, 2021.
- [5] R. Ma *et al.*, "RIS-assisted SWIPT network for internet of everything under the electromagnetics-based communication model," *IEEE Internet Things J.*, vol. 11, no. 9, pp. 15 402–15 415, 2024.
- [6] H. El Hassani *et al.*, "Optimization of RIS-aided MIMO—a mutually coupled loaded wire dipole model," *IEEE Wirel. Commun. Lett.*, vol. 13, no. 3, pp. 726–730, 2023.
- [7] H. Li *et al.*, "Beyond diagonal reconfigurable intelligent surfaces with mutual coupling: Modeling and optimization," *IEEE Commun. Lett.*, vol. 28, no. 4, pp. 937–941, 2024.
- [8] A. Abrardo *et al.*, "Design of reconfigurable intelligent surfaces by using S-parameter multiport network theory—optimization and full-wave validation," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 11, pp. 17 084–17 102, 2024.
- [9] D. Wijekoon *et al.*, "Phase shifter optimization in RIS-aided MIMO systems under multiple reflections," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 8, pp. 8969–8983, 2024.
- [10] M. Nerini *et al.*, "Global optimal closed-form solutions for intelligent surfaces with mutual coupling: Is mutual coupling detrimental or beneficial?" *IEEE Trans. Wirel. Commun.*, 2025.
- [11] D. Semmler *et al.*, "Performance analysis of systems with coupled and decoupled RISs," *Proc. ISWCS*, pp. 1–6, 2024.
- [12] B. Peng *et al.*, "RISnet: A domain-knowledge driven neural network architecture for RIS optimization with mutual coupling and partial CSI," *Trans. Wireless. Comm.*, vol. 24, no. 5, p. 4469–4482, Feb. 2025.
- [13] J. Tapie *et al.*, "Systematic physics-compliant analysis of over-the-air channel equalization in RIS-parametrized wireless networks-on-chip," *IEEE J. Sel. Areas Commun.*, vol. 42, no. 8, pp. 2026–2038, 2024.
- [14] P. Zheng *et al.*, "Mutual coupling in RIS-aided communication: Model training and experimental validation," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 11, pp. 17 174–17 188, 2024.
- [15] A. Rabault *et al.*, "On the tacit linearity assumption in common cascaded models of RIS-parametrized wireless channels," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 8, pp. 10 001–10 014, 2024.
- [16] P. del Hougne, "Experimental reduced-rank mutual coupling representation and estimation for large RIS," *arXiv:2511.11133*, 2025.
- [17] Z. Ming *et al.*, "A hybrid transmitting and reflecting beyond-diagonal reconfigurable intelligent surface with independent beam control and power splitting," *IEEE Trans. Microw. Theory Techn.*, 2025.
- [18] J. Tapie *et al.*, "Beyond-diagonal RIS prototype and performance evaluation," *IEEE Wirel. Commun. Lett.*, 2025.
- [19] P. del Hougne, "Experimental multiport-network parameter estimation and optimization for multi-bit RIS," *IEEE Wirel. Commun. Lett.*, 2025.
- [20] Q. Ma *et al.*, "Smart sensing metasurface with self-defined functions in dual polarizations," *Nanophotonics*, vol. 9, no. 10, pp. 3271–3278, 2020.
- [21] G. C. Alexandropoulos *et al.*, "Hybrid reconfigurable intelligent metasurfaces: Enabling simultaneous tunable reflections and sensing for 6g wireless communications," *IEEE Veh. Technol. Mag.*, vol. 19, no. 1, pp. 75–84, 2023.
- [22] P. del Hougne, "Virtual VNA: Minimal-ambiguity scattering matrix estimation with a fixed set of "virtual" load-tunable ports," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–19, 2025.
- [23] —, "Virtual VNA 2.0: Ambiguity-free scattering matrix estimation by terminating not-directly-accessible ports with tunable and coupled loads," *IEEE Trans. Antennas Propag.*, vol. 73, no. 7, pp. 4903–4908, 2025.

- [24] —, “Wireless multi-port sensing: Virtual-VNA-enabled de-embedding of an over-the-air fixture,” *IEEE Trans. Antennas Propag.*, 2025.
- [25] J. Sol *et al.*, “Experimentally realized physical-model-based frugal wave control in metasurface-programmable complex media,” *Nat. Commun.*, vol. 15, no. 1, p. 2841, 2024.
- [26] P. del Hougne, “A physics-compliant diagonal representation for wireless channels parametrized by beyond-diagonal reconfigurable intelligent surfaces,” *IEEE Trans. Wirel. Commun.*, vol. 24, no. 7, pp. 5871–5884, 2025.
- [27] S. E. Zegrar *et al.*, “Reconfigurable intelligent surface (RIS): Eigenvalue decomposition-based separate channel estimation,” *Proc. PIMRC*, pp. 1–6, 2021.
- [28] D. Chizhik *et al.*, “Keyholes, correlations, and capacities of multielement transmit and receive antennas,” *IEEE Trans. Wirel. Commun.*, vol. 1, no. 2, pp. 361–368, 2002.
- [29] J. Tapie and P. del Hougne, “Scalable multiport antenna array characterization with PCB-realized tunable load network providing additional “virtual” VNA ports,” *IEEE Antennas Wirel. Propag. Lett.*, 2025.
- [30] I. Ahmed *et al.*, “Over-the-air emulation of electronically adjustable Rician MIMO channels in a programmable-metasurface-stirred reverberation chamber,” *IEEE Trans. Antennas Propag.*, vol. 73, no. 4, pp. 2104–2119, 2025.
- [31] H. Prod’homme and P. del Hougne, “Efficient computation of physics-compliant channel realizations for (rich-scattering) RIS-parametrized radio environments,” *IEEE Commun. Lett.*, vol. 27, no. 12, pp. 3375–3379, 2023.
- [32] J. Sol *et al.*, “Optimal blind focusing on perturbation-inducing targets in sub-unitary complex media,” *Laser Photonics Rev.*, p. 2400619, 2024.
- [33] P. del Hougne *et al.*, “Optimally diverse communication channels in disordered environments with tuned randomness,” *Nat. Electron.*, vol. 2, no. 1, pp. 36–41, 2019.
- [34] C. Hammami *et al.*, “Statistical multiport-network modeling and efficient discrete optimization of RIS,” *arXiv:2508.01776*, 2025.
- [35] Z. Chen *et al.*, “Hybrid evolutionary-based sparse channel estimation for IRS-assisted mmWave MIMO systems,” *IEEE Trans. Wirel. Commun.*, vol. 21, no. 3, pp. 1586–1601, 2021.
- [36] P. Zheng *et al.*, “Mutual coupling-aware channel estimation and beamforming for RIS-assisted communications,” *IEEE Trans. Wirel. Commun.*, 2025.



Philipp del Hougne is a tenured CNRS Associate Professor (*Chargé de recherche, HDR*) affiliated with the Université de Rennes, France. He graduated in Physics from Imperial College London, United Kingdom, and received the Ph.D. degree in Physics from the Université Sorbonne Paris Cité, France. He subsequently held postdoctoral positions in Nice and Rennes, France, and Lausanne, Switzerland. In 2023, he received the *Habilitation à Diriger des Recherches* (HDR) degree in Electronics from the Université de Rennes, France.

He currently leads the Intelligent Wave Systems group at CNRS – IETR (Université de Rennes), France, which studies theory and applications of tunable complex wave systems to mold the flow of information through tailored wave-matter interactions, spanning applications from information extraction (imaging, sensing, localization), via information processing (analog wave-based computing), to information transfer (wireless communications).