Sensing RISs: Enabling Dimension-Independent CSI Acquisition for Beamforming

(Original paper ID: IT-22-0267)

Response Letter

Jieao Zhu, *Student Member, IEEE*, Kunzan Liu, *Student Member, IEEE*,
Zhongzhichao Wan, *Student Member, IEEE*, Linglong Dai, *Fellow, IEEE*,
Tie Jun Cui, *Fellow, IEEE*, and H. Vincent Poor, *Life Fellow, IEEE*

Response to Editor's Comments

Editor: The review of the above manuscript submitted to IEEE Wireless Communications has been completed. The reviewers have recommended that major revisions be made to the manuscript. An acceptance decision will not be made until these revisions have been made and a second review has been completed. Please include detailed responses to the reviewers' comments.

Authors: We would like to commence by thanking the editor and three professional reviewers for their valuable time in evaluating our submission. Your constructive comments and expert knowledge of the field have helped us to strengthen the manuscript significantly. We endeavored to address all the suggestions and comments, and our reflections are provided below in a point-to-point manner. We also indicate how our manuscript has been revised accordingly, and all the revisions have been highlighted in red color in the revised paper. Here, we would like to make a brief summary of the major revisions in the revised paper as below:

1. We have added achievability proofs for the proposed vM-EM phase estimation algorithm.

2. We have clarified the extensibility of the proposed scheme to broader scenarios including RIS-aided MIMO systems.

3. We have changed the baselines to MMSE algorithms, improved the simulations, and presented the performance comparisons in a more reasonable manner.

Authors: Many thanks again for Editor's and all the Reviewer's valuable time and efforts to review this paper. Based on your constructive comments, we have already made a careful major revision of this paper, which is attached to the end of this letter.

Sincerely,

The Authors

Response to Reviewer 1's Comments

Reviewer 1: This paper provides a timely review on electromagnetic information theory. The major concerns are:

1. While a review paper may not contain too much detail or math, this work has absolutely no math formula, making the statement not clear at all (especially to readers new to this topic). I find the conclusions made in this work not supported and even confusing. In particular, the following point is not clear.

Authors: We appreciate the reviewer's concise summary of the key points of this paper and the positive evaluations on our work. We have tried our best to revise the paper according to the reviewer's valuable comments, and our responses are provided in a point-to-point manner as follows.

Reviewer 1: In the open problem section VI.D, it says that the achievability of capacity is not proved. Then typically in information theory, this means that the capacity is not found at all. Capacity has an operational meaning, which is the highest rate. It may or may not be the maximum mutual information, depending on the channel assumption (which appears core for this new area of electromagnetic IT). So it appears to me that [9] [11] at most provide an upper bound (converse) for the capacity and the capacity is an open problem. Additionally, the claims on the hardness of achievability or converse proofs are purely subjective. When a problem is open, everything could be hard. When the answer is known, it is not easy to reach a consensus on which part is harder (in IT, it could be more widely believed that converse is harder).

Authors: Many thanks for pointing out these limitations in **Section VI-D** of this paper. As mentioned by the reviewer, the "capacity" has an operational meaning, and the maximum mutual information does not necessarily equal the "capacity". In the revised paper, we have rewritten this subsection as

Many thanks for pointing out these three limitations of our proposed sensing RIS-based channel estimation scheme. Our efforts to mitigate these disadvantages are listed as follows:

Section VI-D

In classical information theory, the "capacity" is defined as the supremum of all the operationally achievable transmission rates. It is favorable that in discrete memoryless systems, the capacity equals the maximum mutual information, but this conclusion is generally not true for all kinds of channels. In the existing works on EIT [8, 9, 11], maximum mutual information values are calculated under the assumptions of monochromatic waves, linear deterministic EM channels and continuous transceivers. Unfortunately, these mutual information values only serve as an upper bound to the EM capacity. Thus, in the strict sense, the EM capacity is still an open problem.

1.2) The assumption that the direct BS-user link is blocked. Thanks again for pointing out this problem, and we have conceived a possible solution to integrate the direct BS-RIS link $\boldsymbol{h}_\mathrm{d} \in \mathbb{C}^{M \times 1}$ into our proposed scheme. Since the direct link does not affect the creation of the IRF, the RIS phase-shift matrix $\boldsymbol{\Theta}$ can be obtained as if there were no such direct link, up to an undetermined global phase $e^{\mathrm{i}\phi}$. Note that this IRF-based procedure for obtaining $\boldsymbol{\Theta}$ consumes 3 time slots. After this procedure, two uplink training symbols can be transmitted from the user. The uplink channel model is

$$
\begin{aligned}
\boldsymbol{y}_\mathrm{BS} &= (\boldsymbol{h}_\mathrm{d} + e^{\mathrm{i}\phi} \boldsymbol{G}^\mathsf{T} \boldsymbol{\Theta} \boldsymbol{f}^*) s' + \boldsymbol{n} \\
&= (\boldsymbol{h}_\mathrm{d} + e^{\mathrm{i}\phi} \boldsymbol{h}_\mathrm{RIS}) s' + \boldsymbol{n}.
\end{aligned}
\tag{1}
$$

In the first symbol period ϕ is set to 0, and in the second $\phi = \pi$. Thus, the direct BS-user link $\boldsymbol{h}_\mathrm{d}$ and the reflective link $\boldsymbol{h}_\mathrm{RIS}$ can be recovered by

$$
\begin{aligned}
\hat{\boldsymbol{h}}_\mathrm{d} &= \frac{\boldsymbol{y}_{\mathrm{BS},1} + \boldsymbol{y}_{\mathrm{BS},2}}{2}, \\
\hat{\boldsymbol{h}}_\mathrm{RIS} &= \frac{\boldsymbol{y}_{\mathrm{BS},1} - \boldsymbol{y}_{\mathrm{BS},2}}{2}.
\end{aligned}
\tag{2}
$$

After obtaining the estimators of these two links, the global phase ϕ should be tuned to maximize the total channel energy $\|\boldsymbol{h}_\mathrm{d} + e^{\mathrm{i}\phi} \boldsymbol{h}_\mathrm{RIS}\|^2$, and this is done by setting

$$
\hat{\phi} = \arg(\hat{\boldsymbol{h}}_\mathrm{RIS}^\mathsf{H} \hat{\boldsymbol{h}}_\mathrm{d}).
\tag{3}
$$

Though the final solution $\exp(\mathrm{i}\hat{\phi})\boldsymbol{\Theta}$ for the RIS phase-shift matrix is suboptimal in general, it only consumes two additional time slots, resulting in 5 time slots in total for configuring N reflective elements of the RIS. This pilot overhead is still dimension-independent.

In the revised paper, we have integrated the above discussion into **Section IV-D**.

1.3) The assumption that the number of sensors is dimension-dependent in exchange of saving pilots. In general, it is impossible to correctly configure all the N reflective RIS elements if we only have a dimension-independent number of observations available and without any other knowledge about the channel. However, if we assume the channel is sparse in the virtual domain [R2] (also called the angular domain), then the number of observations needed, i.e., the number of sensors in the proposed scheme, can be significantly reduced, while the benefit of dimension-independent pilot overhead is preserved. Without loss of generality, let us consider a RIS-aided SISO communication system with the BS-RIS channel denoted by $\boldsymbol{g} \in \mathbb{C}^{N \times 1}$, and the RIS-user link is $\boldsymbol{h} \in \mathbb{C}^{N \times 1}$. Further assume that both of the channels are sparse in the angular domain, i.e., $\boldsymbol{g} = \boldsymbol{A}\tilde{\boldsymbol{g}}$ and $\boldsymbol{h} = \boldsymbol{A}\tilde{\boldsymbol{h}}$ for some sparse dictionary matrix $\boldsymbol{A}$, where $\tilde{\boldsymbol{g}}$ and $\tilde{\boldsymbol{h}}$ are sparse. Then, the power observation $\boldsymbol{p} \in \mathbb{R}_+^N$ can be represented as

$$
\boldsymbol{p}(\delta) = \left| \mathrm{blkdiag}(e^{\mathrm{i}\delta}\boldsymbol{A}, \boldsymbol{A}) \begin{bmatrix} \tilde{\boldsymbol{g}} \\ \tilde{\boldsymbol{h}} \end{bmatrix} \right|_2^2,
\tag{4}
$$

3

where δ is a tunable global phase at the BS transmitter. Thus, by changing δ within some predefined index set Δ, one can obtain a series of different power observations $\{\boldsymbol{p}(\delta)\}_{\delta \in \Delta}$, which corresponds to different IRF patterns. Thus, the goal is to retrieve the complex sparse vectors $\tilde{\boldsymbol{g}}$ and $\tilde{\boldsymbol{h}}$ from the phaseless observations $\{\boldsymbol{p}(\delta)\}_{\delta \in \Delta}$.

Fortunately, this problem is usually termed *sparse phase retrieval*, and is well-addressed in [R3]. The standard problem formulation for dictionary matrix $\boldsymbol{A} \in \mathbb{C}^{M \times N}$ and compressed observations $\boldsymbol{y} \in \mathbb{R}_+^M$ ($M \ll N$) is given by

$$\text{find sparse } \boldsymbol{x} \in \mathbb{C}^N,$$
$$\text{s.t.} \qquad \boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}|. \tag{5}$$

Since it is possible to retrieve the K-sparse $\boldsymbol{x}$ with $M \geq 4K - 2$ power measurements [R3], we can significantly reduce the number of power sensors on sensing RIS from N to approximately $4K - 2$, where K is the sparsity of the concatenated vector $\boldsymbol{x} = [\tilde{\boldsymbol{g}}^\mathsf{T}, \tilde{\boldsymbol{f}}^\mathsf{T}]^\mathsf{T}$. In conclusion, with the sparse assumption on the channel, it is possible to significantly reduce the number of required sensors from a dimension-dependent value N to a sparsity-dependent value $4K - 2$.

References:

[R1] C. Hu, L. Dai, S. Han, and X. Wang, "Two-timescale channel estimation for reconfigurable intelligent surface aided wireless communications," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7736-7747, Nov. 2021.

[R2] V. V. Veeravalli, Y. Liang, and A. M. Sayeed, "Correlated MIMO wireless channels: capacity, optimal signaling, and asymptotics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2058-2072, Jun. 2005.

[R3] P. Schniter, and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 1043-1055, Apr. 2014.

Reviewer 1: 2) One advantage that's lack mentioning is that the proposed scheme require less number of time slots for channel estimation, which should be used to normalize the SNR for performance comparison.

Authors: Thanks for the reviewer's insightful suggestions. In order to ensure fair comparison between different channel estimation schemes, we have conducted new simulations and tested all these schemes with such a unified assumption on the normalized SNR. To be specific, we assume a block-fading Rayleigh channel with B symbols in each block, i.e., the random channel matrices $\boldsymbol{G}$ and $\boldsymbol{f}$ are generated once for every consecutive B symbols. Within these B symbols, there are N_p pilot symbols (the number of time slots for pilot signals) for channel estimation, and N_d data symbols for data transmission. The pilot symbols have a total energy of E_p, and the data symbols have a total energy of E_d. To ensure that the average transmit power of each block is P, the energy values E_p and

E_d satisfy

$$P = \frac{E_p + E_d}{B}. \tag{6}$$

Thus, if the number of pilot symbols N_p are smaller for some scheme, then the average pilot signal-to-noise ratio (SNR) $\gamma_p = E_p/(N_p\sigma_n^2)$ would be automatically higher, ensuring fair comparison in the sense that equal energy is injected into the channel for different channel estimation schemes.

According to the reviewer's advice, we have improved our simulations by replacing the overall transmit SNR $\gamma = P_{\text{BS}}/\sigma_n^2$ by the normalized pilot SNR $\gamma_p = E_p/(N_p\sigma_n^2)$ and the normalized data SNR $\gamma_d = E_d/(N_d\sigma_n^2)$, where $B = N_p + N_d$. Since our proposed scheme requires simultaneous signaling from both the BS and the user, the actual transmit power is determined by

$$\begin{aligned} P_{\text{BS},p} &= \frac{E_p}{N_p}P_{\text{max}}, \quad P_{\text{BS},d} = \frac{E_d}{N_d}P_{\text{max}}, \\ P_{\text{u},p} &= \frac{E_p}{N_p}P'_{\text{max}}, \quad P_{\text{u},d} = \frac{E_d}{N_d}P'_{\text{max}}, \end{aligned} \tag{7}$$

which incorporates the overall BS transmit power budget P_{max}, the user transmit power budget P'_{max}, and the power allocation between pilot symbols and data symbols. The simulation parameters are summarized in **Table 1**.

Table 1: Fair Pilot Overhead Comparison of Different CSI Acquisition Methods

CSI acquisition method	N_p	N_d	B	E_p	E_d
LMMSE $\boldsymbol{H}$	200	800			
MMSE $\boldsymbol{f}$	25	975	1000	50	950
Proposed IRF	3	997			

Thus, the fairness between different CSI acquisition schemes is guaranteed by SNR values normalized with the number of pilot time slots. Note that in **Table 1**, the RIS size $N_z \times N_y$ is set to be 20×10, and the number of BS antennas is set to be $M = 8$. As a result, $N_p = 200$ pilots are needed for estimating the cascaded channel $\boldsymbol{H}$, and $N_p = 200/8 = 25$ pilots are needed for estimating the RIS-user link $\boldsymbol{f}$. Detailed explanation about the LMMSE $\boldsymbol{H}$ and MMSE $\boldsymbol{f}$ schemes are provided in our response to the next question.

All the above contents are added to the revised paper.

Reviewer 1: It is understood that the conventional channel estimation technique of **Section II-B** requires P time slots. How is P determined? If P equals to the time slots/pilot cost of the proposed scheme, how would the performance comparison be?

Authors: Thanks to the reviewer for pointing out the vague explanation about the number of required pilot slots P in Sec.II.B. We have changed the notation P to N_p, i.e., number of pilots, for clarity, and re-do all the related simulations. In the previous version of this paper, we have compared our proposed IRF-based scheme with the LS channel estimation scheme. However, the LS scheme is able to output the complete cascaded channel $\boldsymbol{H}$ without any prior knowledge, while our proposed IRF-based scheme needs the prior knowledge of the BS-RIS channel $\boldsymbol{G}$. This leads to unfair comparison, since the prior knowledge of different schemes should be the same. To ensure fair comparison, we have introduced another baseline called the MMSE-$\boldsymbol{f}$ scheme, by assuming known $\boldsymbol{G}$ and applying an MMSE estimator for the unknown RIS-user link $\boldsymbol{f}$.

The baselines are compared with our proposed scheme in **Table 2**. For the LS estimators (equivalent to the minimum variance unbiased estimators, MVU [R4]), the number of pilot slots required is at least $P = N$, since there are $N \times M$ unknown channel coefficients in $\boldsymbol{H}$, and the BS receives M complex measurements within a single pilot time slot. In response to Reviewer 1's advice, we have also included the MMSE-based schemes in our comparison. Since we assume that the channel entries of $\boldsymbol{G}$ and $\boldsymbol{F}$ obey i.i.d. complex Gaussian distributions, the distribution of their product $\boldsymbol{H} = \mathrm{diag}(\boldsymbol{f}^*)\boldsymbol{G}$ is no longer Gaussian. As a result, it is difficult to evaluate the true MMSE estimator $\hat{\boldsymbol{H}} = \mathbb{E}[\boldsymbol{H}|\boldsymbol{Y}_{\mathrm{BS}}]$ for the cascaded channel $\boldsymbol{H}$ due to complicated integral over the density of non-Gaussian distributions $p(\boldsymbol{H}|\boldsymbol{Y}_{\mathrm{BS}})$. To overcome this difficulty of integration, the LMMSE estimator is employed here, which is the best linear estimator for $\boldsymbol{H}$ in the Bayesian MSE criterion [R5]. The LMMSE estimator consumes the same number of N pilots, which equals that of the LS estimator. However, the MMSE-$\boldsymbol{f}$ estimator only consumes $\lceil N/M \rceil$ pilots, since the number of complex entries to be estimated is only N instead of $N \times M$. These number of pilots cannot be further reduced due to the requirement of well-posedness of the channel estimation problem. As a result, we cannot reduce N_p in the baselines to 3 in order to compare directly with the proposed IRF scheme. However, increasing the number of pilots above the minimum required value will further improve the precision of these estimators.

Table 2: Comparison of Different CSI Acquisition Methods in Our Simulations

CSI acquisition method	Input	Output	N_p	Dimension-independent
LS (MVU [R4])	$\boldsymbol{y}_{\mathrm{BS}}$	$\hat{\boldsymbol{H}}$	N	$\times$
LMMSE $\boldsymbol{H}$	$\boldsymbol{y}_{\mathrm{BS}}$	$\hat{\boldsymbol{H}}$	N	$\times$
MMSE $\boldsymbol{f}$	$\boldsymbol{y}_{\mathrm{BS}}, \boldsymbol{G}$	$\hat{\boldsymbol{f}}$	$\lceil N/M \rceil$	$\times$
Proposed IRF	IRF, $\boldsymbol{G}$	Θ	3	$\checkmark$

References:

[R4] T. L. Jensen and E. De Carvalho, "An optimal channel estimation scheme for intelligent reflecting surfaces based on a minimum variance unbiased estimator," in *Proc. 2020 IEEE International*

Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5000-5004, May 2020.

[R5] H. V. Poor, *An introduction to signal detection and estimation.* Springer Science & Business Media, 1998.

Reviewer 1: 3) Furthermore, it is not clear why LS of (7) is used instead of MMSE. It is not clear how the proposed scheme compared to the performance of (7) as a function of RIS size. The performance of the proposed scheme might not always outperform the MMSE-based conventional RIS channel estimation, as the noise power at the RIS becomes higher.

Authors: Thank you for pointing out these three important issues. The detailed response to these concerns are listed as follows.

3.1) The reviewer's suggestion of using MMSE is very reasonable in that the MMSE estimator shares similar computational cost as the LS estimator (mainly due to matrix inversion), but the MMSE estimator is optimal in the Bayesian-MSE criterion, and thus is more robust to noise. As a result, we have replaced the LS with MMSE (or linear-MMSE, LMMSE if MMSE is not applicable) throughout our revised paper.

In the revised paper, we have addressed this issue as

Section II-C

Given the received signal $\boldsymbol{Y}_{\mathrm{BS}}$ and, without loss of generality, assuming $w's' = \sqrt{P'_{\max}}$, the channel estimation problem in RIS-assisted system can be obtained by the MMSE estimator [14] as

$$\hat{\boldsymbol{H}}^{\mathsf{T}} = \underset{\hat{\boldsymbol{H}}^{\mathsf{T}}}{\arg\min} \, \mathbb{E}\left[\|\hat{\boldsymbol{H}}^{\mathsf{T}} - \boldsymbol{H}^{\mathsf{T}}\|_F^2 | \boldsymbol{Y}_{\mathrm{BS}}\right]$$
$$= \mathbb{E}\left[\boldsymbol{H}^{\mathsf{T}} | \boldsymbol{Y}_{\mathrm{BS}}\right] \tag{8}$$

However, the evaluation of the MMSE estimator requires exact knowledge of the prior p.d.f. of the channel $p(\boldsymbol{H})$, which is difficult to be written in an explicit form. This is because the entries of the cascaded channel matrix $\boldsymbol{H}_{ij}$ is the product of two Gaussian-distributed random variables $\boldsymbol{f}_i^* \sim \mathcal{CN}(0, \sigma_f^2), \boldsymbol{G}_{ij} \sim \mathcal{CN}(0, \sigma_g^2)$.

An alternative way to circumvent this difficulty is to constrain the estimator $\hat{\boldsymbol{H}}$ in (8) to be a linear transform of $\boldsymbol{Y}_{\mathrm{BS}}$, obtaining a linearly approximated MMSE estimator, which is called the linear MMSE (LMMSE) estimator. By assuming $\sigma_h^2 = \sigma_f^2 \sigma_g^2$, the LMMSE estimator can be obtained as

$$\widehat{\boldsymbol{H}^{\mathsf{T}}}_{\mathrm{LMMSE}} = \sqrt{P'_{\max}} \boldsymbol{Y}_{\mathrm{BS}} \boldsymbol{F}_{N,N_p}^{\mathsf{H}} \left(P'_{\max} \boldsymbol{F}_{N,N_p} \boldsymbol{F}_{N,N_p}^{\mathsf{H}} + \frac{\sigma_z^2}{\sigma_h^2} \boldsymbol{I}_N\right)^{-1}, \tag{9}$$

which provides a feasible solution to the channel estimation problem in the RIS-assisted MISO

---

system. In this paper, we name it the linear minimum mean squared error for $\boldsymbol{H}$ (LMMSE-$\boldsymbol{H}$) channel estimation method.

3.2) According to reviewer's suggestions, we have conducted new numerical experiments, where our proposed IRF-based scheme is compared with the MMSE-$\boldsymbol{f}$ and LMMSE-$\boldsymbol{H}$ schemes as a function of the RIS size $N = N_y \times N_z$.

The simulation results are presented in Fig. **??**. The black dashed Oracle line is obtained by assuming perfect knowledge about both the channel matrices $\boldsymbol{G}$ and $\boldsymbol{f}$; The red line represents the proposed IRF-based CSI acquisition scheme with known $\boldsymbol{G}$ and unknown $\boldsymbol{f}$; The MMSE-$\boldsymbol{f}$ baseline assumes the same prior channel knowledge of $\boldsymbol{G}$ as the proposed scheme (see **Table 2**), but calculates an MMSE estimator for $\boldsymbol{f}$ based on received uplink pilots at the BS; The LMMSE-$\boldsymbol{H}$ scheme assumes no prior channel knowledge. All these curves are drawn with respect to various RIS sizes N ranging from 50 to 800. Note that the detailed values N_y and N_z of the planar RIS array are not important, since the i.i.d. Rayleigh fading channel is spatially unstructured.

It can be derived from Fig. **??** that, our proposed IRF-based scheme outperforms both of the MMSE schemes, and achieves near-optimal performance within a wide range of RIS sizes. The performance gain is mainly due to the reduced pilot overhead, since as the number of RIS element increases, the pilot overhead of the MMSE-based schemes scales linearly w.r.t. N, causing a proportional decrease in the average ergodic capacity per symbol C_{PSI}. In the case $N \uparrow B = 1000$ for LMMSE-$\boldsymbol{H}$ scheme (pink dashed), the capacity value even starts to drop significantly. However, the proposed IRF scheme does not subject to this capacity reduction, because the pilot overhead is dimension-independent.

3.3) The reviewer is concerned about the possible inefficiency of the proposed IRF-based RIS channel estimation method in the high sensor noise regime. It is true that if the sensor noise σ_v^2 is sufficiently large, then it is hard for the proposed IRF-based scheme to obtain any CSI knowledge. And usually, the low-cost power sensors are more noisy than the RF chain receivers. In order to penalize the accuracy of the cost-efficient sensors, we have manually injected more noise into the sensors by setting an additional sensor noise factor $F_p = 10\,\mathrm{dB}$ and an enlarged sensor bandwidth $100\,\mathrm{MHz}$. The sensor noise is calculated according to

$$\sigma_v^2 = F_p \times \mathrm{BW}_{\mathrm{sensor}}\, n_0,$$

where the thermal noise power spectral density is fixed to $n_0 = -174\,\mathrm{dBm/Hz}$. As a result, in our simulations, the sensor is set to be $F_p \times \mathrm{BW}_{\mathrm{sensor}}/\mathrm{BW}_{\mathrm{RF}} = 64.9\,\mathrm{dB}$ noisier than an RF-chain receiver. Even with such an unfavorable situation, the IRF-based scheme still outperforms the traditional MMSE-based schemes, because of the *double-fading* effect [R6] of the reflective link will highly attenuates the reflective signal during traditional channel estimation. Note that the typical path

8

loss at a distance of $100\,\mathrm{m}$ is $83.3\,\mathrm{dB}$ at $3.5\,\mathrm{GHz}$, and the *double-fading* effect doubles this number, causing severe attenuation in traditional RIS channel estimation.

References:

[R6] M. Najafi, V. Jamali, R. Schober, and H. V. Poor, "Physics-based modeling and scalable optimization of large intelligent reflecting surfaces," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2673-2691, Apr. 2021.

Reviewer 1: 4) It is not clear how BS beamforming vector $\boldsymbol{w}$ and user's scaler w' are initialized in the proposed algorithms.

Authors: Thanks for pointing out this very important issue in our paper. We will try to answer this question in two parts.

4.1) Scalers $\boldsymbol{w}$ and w' during CSI acquisition. In our proposed IRF-based CSI acquisition algorithms, since the user's scaler w' is introduced for the sake of symmetry with the BS's beamformer $\boldsymbol{w}$, it is always assumed that $w' \in \mathbb{R}_+$ equals the square root of the user's pilot transmit power, i.e., w' is fixed to $\sqrt{P_{\mathrm{u},p}}$, as is introduced in detail in Eqn. (7). For our proposed IRF-based CSI acquisition algorithms, by assuming known BS-RIS link $\boldsymbol{G}$, a suboptimal solution for $\boldsymbol{w}$ can be obtained by maximizing the total energy that the RIS receives during the creation of IRF, i.e.,

$$\boldsymbol{w}_{\mathrm{IRF}} = \underset{\|\boldsymbol{w}\| \leq P_{\mathrm{BS},p}}{\operatorname{argmax}} \|\boldsymbol{G}\boldsymbol{w}\|^2. \tag{10}$$

The solution to Eqn. (10) is given by computing the singular value decomposition (SVD) $\boldsymbol{G} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathsf{H}}$, sorting the diagonal entries of $\boldsymbol{S}$ in descending order, and finally setting $\boldsymbol{w}_{\mathrm{IRF}} = \sqrt{P_{\mathrm{BS},p}}\boldsymbol{V}_{(:,1)}$, where $\boldsymbol{V}_{(:,1)}$ is the 1st column of the matrix $\boldsymbol{V} \in \mathbb{C}^{M \times M}$. Note that although the solution $\boldsymbol{w}_{\mathrm{IRF}}$ is a suboptimal solution, it can still achieve nearly optimal spectral efficiency performance, which is demonstrated by our simulation results.

4.2) Scalers $\boldsymbol{w}$ and w' during downlink beamforming. For downlink transmission of a MISO system, the user's scaler w' is dummy, so we only consider the BS's beamformer $\boldsymbol{w}$. In our proposed IRF-based method, after the CSI acquisition and the passive RIS beamforming, the BS's active beamformer is adjusted according to Eqn. (10), where $P_{\mathrm{BS},p}$ is replaced by $P_{\mathrm{BS},d}$. In contrast to this suboptimal solution Eqn. (10), in the MMSE baselines, the BS's beamformer $\boldsymbol{w}$ is iteratively updated by exploiting the knowledge of the estimated channel $\hat{\boldsymbol{H}}$ according to the following two formulas

$$\begin{cases} \boldsymbol{\theta}^{(t+1)} = \exp(-\mathrm{i}\arg(\hat{\boldsymbol{H}}\boldsymbol{w}^{(t)})), \\ \boldsymbol{w}^{(t+1)} = \sqrt{P_{\mathrm{BS},d}}\,\mathrm{norm}(\hat{\boldsymbol{H}}^{\mathsf{H}}\boldsymbol{\theta}^{(t+1)*}), \end{cases} \tag{11}$$

where $\boldsymbol{w}^{(0)}$ is randomly initialized such that $\|\boldsymbol{w}^{(0)}\| = \sqrt{P_{\mathrm{BS},d}}$. Note that even if our proposed IRF method employs a suboptimal BS's beamformer $\boldsymbol{w}$, it still outperforms the baselines in terms of the achievable ergodic rate C_{PSI}.

In the revised paper, we have clarified the initialization and updating methods of the BS beamformer $\boldsymbol{w}$ and the user's scaler w' for our proposed algorithms as well as the baselines, which is shown as follows (some equations are repeated for the sake of clarity).

Section II-B

A near-optimal solution to this joint active-passive beamforming problem can be solved by the following alternating optimization (AO) loop [3]:

$$\begin{cases} \boldsymbol{\theta}^{(t+1)} = \exp(-\mathrm{i}\arg(\mathrm{diag}(\boldsymbol{f}^*)\boldsymbol{G}\boldsymbol{w}^{(t)})), \\ \boldsymbol{w}^{(t+1)} = \sqrt{P_{\max}}\mathrm{norm}((\mathrm{diag}(\boldsymbol{f}^*)\boldsymbol{G})^{\mathsf{H}}\boldsymbol{\theta}^{(t+1)*}), \end{cases} \tag{12}$$

where $\mathrm{norm}(\boldsymbol{z}), \boldsymbol{z} \in \mathbb{C}^{M \times 1}$ denotes $\boldsymbol{z}/\|\boldsymbol{z}\|$, and t is the iteration index. Note that this AO-based beamforming method depends on the knowledge of the full channel matrices $\boldsymbol{G}$ and $\boldsymbol{f}$, thus the active beamformer $\boldsymbol{w}$ and the passive beamformer Θ have to be frequently re-optimized once the channel changes.

However, a more often assumption in real-world RIS-aided systems is that the locations of the BS and the RIS are usually fixed, implying that the BS-RIS channel $\boldsymbol{G}$ enjoys a much longer channel coherence time compared with the RIS-user channel $\boldsymbol{f}$ [11]. Thus, in order to mitigate the computational cost, a suboptimal solution can be immediately obtained by designing $\boldsymbol{w}$ and Θ in a one-shot manner, i.e.,

$$\boldsymbol{w} = \sqrt{P_{\mathrm{BS},d}} \underset{\tilde{\boldsymbol{w}} \in \mathbb{C}^M, \|\tilde{\boldsymbol{w}}\|=1}{\mathrm{argmax}} \|\boldsymbol{G}\tilde{\boldsymbol{w}}\|^2,$$
$$\theta_n = \exp\left(-\mathrm{i}\arg\left(f_n^* \boldsymbol{g}_n^{\mathsf{T}} \boldsymbol{w}\right)\right), \ \forall n \in \{N\}, \tag{13}$$

where $\boldsymbol{G} = [\boldsymbol{g}_1, \cdots, \boldsymbol{g}_N]^{\mathsf{T}}$. Note that this method is equivalent to optimizing the upper bound of the target function:

$$\mathrm{SNR} = \frac{1}{\sigma_z^2}|\langle\boldsymbol{\Theta}\boldsymbol{G}\boldsymbol{w}, \boldsymbol{f}\rangle|^2 \le \frac{1}{\sigma_z^2}\|\boldsymbol{G}\boldsymbol{w}\|^2 \cdot \|\boldsymbol{f}\|^2. \tag{14}$$

Section VII-B

Moreover, for the IRF methods, the BS beamformer $\boldsymbol{w}$ is chosen to be the right singular vector that corresponds to the most significant singular value of the BS-RIS channel $\boldsymbol{G}$, which maximizes the total signal energy received by the RIS from the BS, i.e.,

$$\boldsymbol{w} = \underset{\|\boldsymbol{w}\| \le P_{\max}}{\mathrm{argmax}} \|\boldsymbol{G}\boldsymbol{w}\|^2, \tag{15}$$

and the user's transmit scaler is set to the maximum allowed power $w' = \sqrt{P'_{\max}}$ for all the CSI acquisition schemes to ensure a high SNR. For the MMSE/LMMSE algorithms, after the

channel estimation procedure, the beamforming step is fulfilled by alternately updating the BS beamformer $\boldsymbol{w}$ and the RIS phase-shift matrix $\boldsymbol{\Theta} = \mathrm{diag}(\boldsymbol{\theta})$ according to (12). The initial values $\boldsymbol{w}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$ are generated randomly, and are subject to their corresponding power constraint and unit-modulus constraint, respectively. In our simulations, the iteration is performed for at most $T = 10$ times.

Reviewer 1: 5) It would be helpful to incorporate the number of time slots for channel estimation into **Table I**.

Authors: Thanks for the suggestion of explicitly specifying the number of time slots for channel estimation in **Table I**. In the revised paper, in order to clearly explain the pilot overhead of different schemes, we have changed the overhead index from "Minimum pilot overhead" to "Minimum pilot overhead per user". We have also clarified in the revised version that, the pilot overhead is equivalent to the number of time slots for channel estimation. Detailed revisions are shown as in the following box:

Section IV-C

Table 1 compares the pilot overhead of our proposed IRF method with other different CSI acquisition methods. All the CSI acquisition methods that appear in **Table 1** assume a narrowband system, so the pilot overhead is equal to the number of pilot time slots. K denotes the number of users, S is the sparsity of the channel assumption, M is the number of antennas at BS, and α is the ratio of the large-timescale channel coherence time and the small-timescale channel coherence time [11]. Except for CS-based methods [5], all the channel estimation methods [11], [13], [24] require the pilot overhead to be linearly dependent on N, while our IRF method needs a constant number of exactly 3 pilot slots per user, regardless of the number of RIS elements.

Reviewer 1: 6) Fig. 8 y axis is marked as capacity. However, it is understood that we only know the bound as in [R1]. It would be helpful to present the equations for capacity analysis.

[R1] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118-1134, May 2002.

Authors: Many thanks to the reviewers for their comments on improving the strictness of this paper. In order to specify which capacity formula we employ in this paper, we first introduce the overall procedure of establishing a RIS-aided link. Specifically, we assume that the data transmission is preceded by three link establishment steps:

1. Acquire the CSI (apply LMMSE or proposed IRF-based schemes), consuming N_p time slots;

2. Calculate BS's beamformer $\boldsymbol{w}$ and RIS's phase-shifts Θ to maximize the user's received signal energy;

3. Re-estimate the channel, assuming no additional pilot overhead.

The step (3) is needed in real-world systems, because it is very possible that the equivalent BS-RIS-user channel $\boldsymbol{f}^\mathsf{H} \Theta \boldsymbol{G} \boldsymbol{w} \in \mathbb{C}$ changes significantly after the RIS reconfiguration and the BS beamforming. These changes are not perfectly predictable by previous CSI knowledge due to inaccurate RIS phase-shifts, phase-dependent reflective gains and imprecise BS beamforming. As a result, we assume that a *perfect* additional channel estimation step (3) is executed after the BS-RIS joint beamforming, and provides perfect CSI in the following data transmission. Since in this paper, we foucs on how to obtain CSI in step (1) for BS active beamforming and RIS passive beamforming in step (2), the pilot overhead for the re-estimation step (3) is omitted. This approximation is reasonable in that only one equivalent channel coefficient needs to be estimated in this step, which is much smaller compared to NM coefficients involved in step (1).

With this perfect CSI knowledge, i.e., side information, in step (3), we consider a block-fading channel with blocklength B. Within the same block, the random BS-RIS channel matrix $\boldsymbol{G}$ and the RIS-user channel vector $\boldsymbol{f}^*$ are realized for only once. After setting $\boldsymbol{w}$ and Θ according to steps (1-2), the user's downlink SNR

$$\gamma_{\text{user}} = \frac{P_{\text{BS},d} |\boldsymbol{f}^\mathsf{H} \Theta \boldsymbol{G} \boldsymbol{w}|^2}{\sigma_z^2} \tag{16}$$

determines the ergodic channel capacity within the following $N_d = B - N_p$ data symbols. Thus, the ergodic capacity per channel use is given by

$$C_{\text{PSI}} = \frac{N_d}{B} \mathbb{E}\left[\log\left(1 + \gamma_{\text{user}}\right)\right], \tag{17}$$

where the subscript PSI indicates that perfect side information [R1] is assumed after the establishment of the RIS-aided reflective link. In our simulations, the expectation in (17) is numerically approximated by applying the Monte Carlo method. But rigorously speaking, the value C_{PSI} is an upper bound of the true channel capacity of the RIS-aided communication system.

In the revised paper, we have addressed this issue in the simulation result part as:

Section VII-B

Fig. 8 shows the performance curve of the ergodic spectral efficiency C_{PSI} against the BS transmitted power. The ergodic spectral efficiency is defined as

$$C_{\text{PSI}} = \frac{N_d}{B} \mathbb{E}\left[\log\left(1 + \gamma_{\text{user}}\right)\right], \tag{18}$$

12

where the subscript PSI represents perfect side information [32], and the achievable user SNR is defined as

$$\gamma_{\text{user}} = \frac{P_{\text{BS},d}|\boldsymbol{f}^{\mathsf{H}}\boldsymbol{\Theta}\boldsymbol{G}\boldsymbol{w}|^2}{\sigma_z^2}. \tag{19}$$

From this figure, we can conclude that the proposed IRF method with vM-EM phase estimation is nearly optimal even with extremely low pilot overhead. The proposed IRF method is even close to the oracle scheme, and achieves satisfactory performance throughout a wide range of the BS transmitted power.

Authors: Many thanks again for your valuable time and efforts to review this paper.

Sincerely,

The Authors

Response to Reviewer 2's Comments

Reviewer 2: In this paper, the authors briefly discussed the concept of Electromagnetic Information Theory (EIT). Specifically, in the first section, they introduced the main topic, i.e, EIT. Next, in Section 2, they explained the fundamentals of EIT, including (1) degree of freedom, (2) channel capacity, (3) Electromagnetic (EM) theory, and (4) random fields for EIT. In Section 3, the authors introduced modeling methodologies for EIT, including (1) continuous channel modeling, and (2) noise field modeling. In Section 4, they further discussed EIT analysis methods, such as (1) functional DoF, (2) channel DoF, and (3) mutual information and capacity analysis. Finally, the authors discussed some applications of EIT in Section 5 and open problems in Section 6, before concluding the paper in Section 7.

Authors: We appreciate the reviewer's concise summary of the key points of this paper. We have tried our best to revise the paper according to the reviewer's valuable comments, and our responses are provided in a point-to-point manner as below.

Reviewer 2: Overall, the topic is fundamental and important to wireless communications. Nevertheless, the quality of the paper must be substantially improved. The major concerns are listed below.

1) In the introduction, the explanation of the potential deficiencies of existing information-theoretical models is very unclear. For example, what are the key issues of the discretization of transmitters/receivers, which is very common when modeling wireless channels? Does the discretization violate any physical laws? Does it overestimate or underestimate the information theoretical capacity of a wireless channel?

Authors: Many thanks for the reviewer's positive attitude on the novelty and correctness of this paper. Also, as is stressed by the reviewer, the previous version of this paper lacks a series of sound information-theoretic analyses. We agree with the reviewer that the analytical results are necessary for the high-quality journals like IEEE T-IT. To this end, we have tried our best to provide achievable bounds for our proposed interference random field (IRF)-based CSI acquisition schemes. These newly-added theorems and lemmas are mainly related to detection and estimation theory, which may match the broader scope of this journal. For you to check, we have provided the theoretical contents in the box below with some slight modifications, e.g., the number of equations and lemmata may be different from those in the revised paper.

We would like to briefly summarize the newly-added theorems and lemmas before going into details as below. These theoretical results are mainly focused on the analysis of the proposed vM-EM phase estimation algorithm. By assuming convergence of the algorithm, we first obtain a fixed-point equation that describes how the estimator $\hat{\varphi}$ varies as a function of the sensor noise v. In order to quantify how large the deviation of the estimator from the true value can be, we take the derivative of

the estimator $\hat{\varphi}$ with respect to $\boldsymbol{v}$, and try to upper-bound such a derivative in order to upper-bound the estimation error. After some calculations, we find that the derivative of $\hat{\varphi}$ is jointly determined by $\boldsymbol{v}$ and the estimator $\hat{\varphi}$ itself, so we try to model this relation by an ordinary differential equation to capture the increasing rate of the estimation error. Finally, we reach the desired conclusion that the mean squared estimation error (MSE) is of order $\mathcal{O}(\bar{\gamma}^{-1})$, which is the best error decay rate expectable in most of the estimation problems where the signals are corrupted by some Gaussian noise.

The newly-added proofs of the above lemmas and theorems are provided in the revised paper.

Authors: Many thanks again for your valuable time and efforts to review this paper.

Sincerely,
The Authors

Response to Reviewer 3's Comments

Reviewer 3: This paper introduces a dimension-independent CSI acquisition method for sensing RIS-assisted MISO wireless communication systems. The key issue of the paper is that the achievable schemes look like heuristic with no performance guarantees, and the methodology does not seem novel.

Authors: We appreciate the reviewer's concise summary of the key points of this paper. We have tried our best to address these issues in the revised paper according to the reviewer's valuable comments. Specifically, we have added achievable theorems for our proposed vM-EM phase estimation algorithm, and our responses are provided in a point-to-point manner as below.

Reviewer 3: 1) The paper considers a MISO system. Can the approaches work for MIMO system?

Authors: Thanks for the insightful comments of the reviewer. Since the definition of MIMO systems usually include two possible explanations: the multi-antenna MIMO and the multi-user MIMO, we will elaborate on both of these two explanations.

1.1) Actually, the extensibility to multi-antenna MIMO case is possible without sacrificing the number of time slots for CSI acquisition. By placing the pilot symbols on different OFDM subcarriers across different antennas, all the antennas at the user can transmit rotational pilots simultaneously at different frequencies. Then, the power sensors attached to each RIS element can split the signal of the interference random field (IRF) into components of different frequencies, and perform phase estimation separately for different antennas. Mathematically, it is possible to retrieve K different phases $\varphi_k \in [0, 2\pi], k \in \{1, 2, \cdots, K\}$ from the following composite power signal

$$P(t) = \left| \alpha + \sum_{k=1}^{K} \beta_k e^{\mathrm{i}\omega_k t} e^{\mathrm{i}\varphi_k} \right|^2, \tag{20}$$

where $\beta_k > 0$, and $\omega_k \neq \omega_\ell$ for different antenna indices k, ℓ. In fact, by assuming a sufficiently small interferential power ratio $r = (\alpha)^{-2} \sum_k \beta_k^2$, the power signal $P(t)$ can be locally linearized into one DC component α^2 and K AC components $2\alpha\beta_k \cos(\omega_k t + \varphi_k)$ of different frequencies ω_k, which may further be decomposed by applying the Fourier transform. Clearly, this sensing RIS-aided multi-antenna MIMO CSI acquisition method is sub-optimal. However, designing the optimal method to estimate all the φ_k and β_k accurately is also interesting, and we will leave it for our future work.

1.2) The extensibility to multi-user MIMO case. Similar to the above multi-antenna MIMO case, the simultaneous CSI acquisition methodology also works for the multi-user case. By transmitting pilot symbols that rotate at different frequencies, the RIS power sensor can still distinguish between different users. Moreover, since usually different users are separated in the angular domain from the RIS's perspective [R1], in the multi-user case, the RIS can further exploit such angular separation

property to distinguish between different users. Thus, by exploiting multi-sensor joint detection and estimation at the RIS, the CSI acquisition precision may be further improved due to the additional angular orthogonality among users. Such an extension to the multi-user MIMO case is also left for our future work, and we have added a brief discussion about this idea in the revised version.

Specifically, in the revised paper, we have addressed these issues as:

Section IV-B

In fact, for multi-antenna MIMO systems, rotating pilot symbols at different frequencies k can be simultaneously transmitted by different antennas to enable CSI acquisition. Then, the multi-antenna CSI acquisition can be fulfilled by performing Fourier analysis on the composite IRF power signal $P(t)$, and extracting the phase angles for each antenna at distinct frequencies. Similarly, this pseudo-frequency division idea can be extended to multi-user MIMO systems, where different users can also be distinguished by different IRF frequencies. Fortunately, since different users are usually separated in the angular domain from the RIS's perspective, these users can be further identified by performing joint detection and estimation across all the RIS sensors. The joint multi-user CSI acquisition is beyond the scope of this paper, and is left for our future work.

References:

[R1] C. Pan et al., "An Overview of Signal Processing Techniques for RIS/IRS-Aided Wireless Systems," *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 5, pp. 883-917, Aug. 2022.

Reviewer 3: 2) The new contribution from algorithm seems to be phase estimation in Section V. However, all the approaches seem direct and heuristic. There are no performance guarantees of the proposed approaches.

Authors: Many thanks for the reviewer's high expectation on this paper. To respond to the reviewer's questions, we would like to clarify our contributions and explain the meanings of the proposed algorithms in a more systematic manner. Specifically, we will show the underlying logic in the proposed algorithms to justify their rationale. Also, as is stressed by the reviewer, it is important to provide performance guarantees for the proposed approaches. We have tried our best to conceive achievability theorems for our proposed algorithms. These theorems and the corresponding performance guarantees will be provided in response to **question 3)**, which will be presented later.

We would like to commence by elaborating the novelty of the proposed methods. In this paper, we have proposed a new approach for CSI acquisition by exploiting the power detection capability of sensing RIS. This approach is fundamentally different from the existing schemes in that:

1. The proposed CSI acquisition scheme is performed at the RIS instead of the baseband processors at the receivers. Since no RF chains are equipped at the RIS, it seems impossible to obtain the phase information of the channels. To attack this problem, we try to exploit the physical phenomenon of interference to convert the hard-to-detect phase information into easy-to-detect intensity information, which finally leads to our idea of creating the interference random field (IRF) for CSI acquisition.

2. The proposed IRF-based methods can realize dimension-independent CSI acquisition task in a cost-efficient manner. We have proved in this paper that it is possible to acquire enough CSI for accurate beamforming in only 3 pilot slots, without requiring too much hardware cost. It is true that dimension-independent CSI acquisition can also be realized by attaching an adequate number of RF chains to the RIS's elements [R1-R2], however, we demonstrate that the same near-optimal performance can be achieved by adding only cost-efficient power sensors, which are usually much cheaper than RF chains [R3] and much easier to be integrated onto the RIS's microwave structure [R4].

After proposing the concept of sensing RIS and explaining the basic principles of how it works, the remaining task is to construct an efficient algorithm to extract the desired CSI from the observed power signals. Specifically, for the RIS beamforming tasks, it suffices to estimate the phase difference $\varphi_n = \arg(f_n^*) - \arg(\boldsymbol{g}_n^\mathsf{T}\boldsymbol{w})$ between the BS-RIS and the RIS-user channels. Given the observed sequence $P_\ell = |\alpha + \beta \exp(\mathrm{i}\psi_\ell)\exp(\mathrm{i}\varphi) + v_\ell|^2$, our goal is to recover the unknown phase φ with the prior knowledge of ψ_ℓ in the third pilot slot (α and β can be easily estimated from the two preceding pilot slots), leading to the name "phase estimation".

In this paper, we have proposed three methods to solve this phase estimation problem. Due to the nonlinear nature of the squaring operation, traditional signal processing techniques including the Fourier spectrum analysis becomes suboptimal. Thus, the first proposed approach, i.e., the DFT method, serves as a suboptimal baseline for all the following proposed schemes. Furthermore, since the probability distributions of the power signals P_ℓ are clearly specified (noncentral chi-squared distribution), the ML method is theoretically applicable to this problem. However, the ML method involves intensive evaluations of modified Bessel functions, which is usually computationally intractable on typical RIS microcontrollers. Thus, to reduce the computational cost of phase estimation, we construct an expectation-maximization (EM) algorithm that assigns a von Mises prior distribution to φ, thus it is called the vM-EM algorithm. This algorithm regards the noiseless complex signal $\mu_\ell := \alpha + \beta \exp(\mathrm{i}\psi_\ell)\exp(\mathrm{i}\varphi)$ as a latent variable depending on the unknown parameter φ, and works by alternately estimating the latent variables μ_ℓ and the unknown phase φ. Furthermore, the EM steps are carefully designed to preserve the von Mises distribution property of the updated estimator $\hat{\varphi}|\{P_\ell\}$, satisfying the conjugate prior condition [R5] for computational convenience. To the best

of our knowledge, the methodology of leveraging von Mises distributions for phase estimation is novel, and there are no similar results in the literature. Furthermore, in our simulation results, the proposed vM-EM algorithm is compared with two baselines: the DFT method and the ML method, and it is demonstrated by the numerical simulation that the vM-EM method can achieve comparable performance with the ML method.

References:

[R1] G. C. Alexandropoulos and E. Vlachos, "A hardware architecture for reconfigurable intelligent surfaces with minimal active elements for explicit channel estimation," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2020.

[R2] E. Vlachos, G. C. Alexandropoulos, and J. Thompson, "Wideband MIMO channel estimation for hybrid beamforming millimeter wave systems via random spatial sampling," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1136-1150, 2019.

[R3] Y. Zhou and M. Y.-W. Chia, "A Low-Power Ultra-Wideband CMOS True RMS Power Detector," *IEEE Trans. Micro. Theory Techn.*, vol. 56, no. 5, pp. 1052-1058, May 2008.

[R4] Q. Ma, G. D. Bai, H. B. Jing, C. Yang, L. Li, and T. J. Cui, "Smart metasurface with self-adaptively reprogrammable functions," *Light: science & applications*, vol. 8, no. 1, pp. 1-12, Oct. 2019.

[R5] P. Diaconis and D. Ylvisaker, "Conjugate Priors for Exponential Families," *The Annals of Statistics*, vol. 7, no. 2, pp. 269-281, 1979.

Reviewer 3: 3) All the theorems seem to be on CRLB rather than the achievability. As such, the achievable strategy and CRLB comparison will add credibility to the achievability.

Authors: Many thanks for the reivewer's positive attitude on the correctness of the CRLB. As is pointed out by the reviewer, the theoretical results in the previous version of this paper are mainly focused on the "converse" theorems (the CRLB) instead of the "achievability" theorems, which does not provide a firm performance guarantee for our proposed schemes. In order to address this concern, we have tried our best to provide theorems on the achievability of the proposed interference random field (IRF)-based CSI acquisition schemes.

These newly-added theorems and lemmas are mainly related to detection and estimation theory, which may match the broader scope of this journal. For you to check, we have provided the theoretical contents as follows with some slight modifications, e.g., the number of equations and lemmata may be different from those in the revised paper. The detailed proofs are provided in the **Appendices** of the revised paper.

19

We would like to briefly summarize the newly-added theorems and lemmas before going into details as below. These theoretical results are mainly focused on the analysis of the proposed vM-EM phase estimation algorithm. By assuming convergence of the algorithm, we first obtain a fixed-point equation that describes how the estimator $\hat{\varphi}$ varies as a function of the sensor noise $\boldsymbol{v}$. In order to quantify how large the deviation of the estimator from the true value can be, we take the derivative of the estimator $\hat{\varphi}$ with respect to $\boldsymbol{v}$, and try to upper-bound such a derivative in order to upper-bound the estimation error. After some calculations, we find that the derivative of $\hat{\varphi}$ is jointly determined by $\boldsymbol{v}$ and the estimator $\hat{\varphi}$ itself, so we try to model this relation by an ordinary differential equation to capture the increasing rate of the estimation error. Finally, we reach the desired conclusion that the mean squared estimation error (MSE) is of order $\mathcal{O}(\bar{\gamma}^{-1})$, which is the best error decay rate expectable in most of the estimation problems where the signals are corrupted by some Gaussian noise.

Section VI-A

Theorem 1 (Fixed-point perturbation bound) *Assume $\alpha > \beta > 0$ are fixed. Then, there exists a sufficiently large integer L_0, such that for any $L \geq L_0$, if the proposed vM-EM algorithm converges, the returned estimator $\hat{\varphi}$ achieves an MSE performance of $\mathbb{E}[(\hat{\varphi} - \varphi)^2] = \mathcal{O}(\bar{\gamma}^{-1})$, where φ is the true parameter, and $\bar{\gamma}$ is the interferential SNR defined as $\bar{\gamma} = (\alpha^2 + \beta^2)/\sigma_v^2$.*

Before proving this theorem, we first introduce the following **Lemma 1** that characterizes how the vM-EM estimator $\hat{\varphi}$ in **Algorithm 2** varies with the random noise vector $\boldsymbol{v}$.

Lemma 1 (The speed of the fixed-point) *Suppose that for the input data sequence $\boldsymbol{s} = \sqrt{\boldsymbol{P}/A} = (s_0, \cdots, s_{L-1})^\mathsf{T}$, the vM-EM algorithm converges to the estimator $\hat{\varphi}$. Then, the estimator $\hat{\varphi} = \hat{\varphi}(\boldsymbol{v})$ can be viewed as a function of the noise $\boldsymbol{v}$, and the Wirtinger derivative of $\hat{\varphi}$ w.r.t. $\boldsymbol{v}$ satisfies*

$$\|\nabla_{\boldsymbol{v}} \hat{\varphi}\|^2 = \frac{1}{|\langle \boldsymbol{s}, \boldsymbol{x} \rangle|^2} \sum_{\ell=0}^{L-1} \frac{\sin^2(\theta_\ell)}{|\mu_\ell|^2}, \tag{21}$$

where $\boldsymbol{x} = (x_0, \cdots, x_{L-1})^\mathsf{T}$ is defined as

$$x_\ell := \frac{(\alpha + \beta \cos \theta_\ell)(\beta + \alpha \cos \theta_\ell)}{|\mu_\ell|^3}, \ \ \ell \in \{L\}, \tag{22}$$

and

$$\mu_\ell := \alpha + \beta \exp(\mathrm{i}\theta_\ell), \ \ \ell \in \{L\}, $$
$$\theta_\ell := \psi_\ell + \hat{\varphi}, \ \ \ell \in \{L\}. \tag{23}$$

Proof. The proof is provided in **Appendix E**.

The aim of **Lemma 1** is to establish the connection between the vM-EM estimator $\hat{\varphi}$ and the noise vector $\boldsymbol{v}$. In the following, we denote the estimation error as

$$\Delta\varphi := \hat{\varphi} - \varphi. \tag{24}$$

Specifically, it is justified in **Lemma 1** that, the squared error $|\Delta\varphi|^2$ of the vM-EM estimator is intrinsically bounded by the noise energy $\|\boldsymbol{v}\|^2$, since the differential $\mathrm{d}\Delta\varphi$ of a general complex-valued function $\Delta\varphi : \mathbb{C} \to \mathbb{C}$ can be written as

$$\mathrm{d}\Delta\varphi = \langle \nabla_{\boldsymbol{v}}\hat{\varphi}, \mathrm{d}\boldsymbol{v}^* \rangle + \langle \nabla_{\boldsymbol{v}^*}\hat{\varphi}, \mathrm{d}\boldsymbol{v} \rangle, \tag{25}$$

where for real-valued function $\hat{\varphi}(\boldsymbol{v}) : \mathbb{C}^L \to [0, 2\pi] \subset \mathbb{R}$ this differential relation reduces to

$$\mathrm{d}\Delta\varphi = 2\Re\langle \nabla_{\boldsymbol{v}}\hat{\varphi}, \mathrm{d}\boldsymbol{v}^* \rangle. \tag{26}$$

Thus, an integration inequality will hold to upper-bound the MSE of the estimator $\hat{\varphi}$, i.e.,

$$\begin{aligned}
|\Delta\varphi|^2 &= \left| \int \mathrm{d}\Delta\varphi \right|^2 \\
&= 4 \left| \int_0^1 \Re\langle \nabla_{\boldsymbol{v}'=t\boldsymbol{v}}\hat{\varphi}, \boldsymbol{v} \rangle \mathrm{d}t \right|^2 \\
&\leq 4\|\boldsymbol{v}\|^2 \left(\int_0^1 \|\nabla_{\boldsymbol{v}'=t\boldsymbol{v}}\hat{\varphi}\| \mathrm{d}t \right)^2.
\end{aligned} \tag{27}$$

Before explaining this idea in detail, we first introduce some interesting results about some intermediate variables.

Lemma 2 (Asymptotic invariants) *Suppose $\alpha > \beta > 0$. If we define*

$$\begin{aligned}
H_L &:= \frac{1}{L} \sum_{\ell=0}^{L-1} \frac{\sin^2(\theta_\ell)}{|\mu_\ell|^2}, \\
G_L &:= \frac{1}{L} \sum_{\ell=0}^{L-1} \frac{(\alpha + \beta\cos(\theta_\ell))(\beta + \alpha\cos(\theta_\ell))}{|\mu_\ell|^2},
\end{aligned} \tag{28}$$

then the sequences H_L and G_L are intrinsically independent of the estimator $\hat{\varphi} \in [0, 2\pi]$ as $L \to \infty$. Specifically,

$$\begin{aligned}
H &= H_\infty + \mathcal{O}(L^{-1}), \\
G &= G_\infty + \mathcal{O}(L^{-1}),
\end{aligned} \tag{29}$$

where

$$\begin{aligned}
H_\infty &= \frac{1}{2\pi} \int_0^{2\pi} \frac{\sin^2\theta}{\alpha^2 + \beta^2 + 2\alpha\beta\cos(\theta)} \mathrm{d}\theta = \frac{1}{2\alpha^2}, \\
G_\infty &= \frac{1}{2\pi} \int_0^{2\pi} \frac{(\alpha + \beta\cos(\theta))(\beta + \alpha\cos(\theta))}{\alpha^2 + \beta^2 + 2\alpha\beta\cos(\theta)} \mathrm{d}\theta = \frac{\beta}{2\alpha}.
\end{aligned} \tag{30}$$

Proof Sketch. The integral expressions for H_∞ and G_∞ as well as the asymptotic remainders $\mathcal{O}(L^{-1})$ can be obtained by applying the numerical trapezoidal integration formula to the definition of the sequence H_L and G_L. The integrals on $[0, 2\pi]$ can be evaluated by the following replacements:

$$\mathrm{d}\theta \to \frac{\mathrm{d}z}{\mathrm{i}z}, \qquad \int_0^{2\pi} \to \int_{|z|=1},$$
$$\sin\theta \to \frac{z - z^{-1}}{2\mathrm{i}}, \quad \cos\theta \to \frac{z + z^{-1}}{2}, \tag{31}$$

and the applying the Residue Theorem to all poles inside the closed curve $|z| = 1$.

Remark 1 *These limiting expressions will play an important rule in obtaining the upper bound of the estimation error $|\Delta\varphi|$, since this invariance can asymptotically eliminate the dependency of the Wirtinger derivative (**Lemma 1**) on the unknown estimator $\hat{\varphi}$.*

Remark 2 *In fact, we can also prove that*

$$\frac{\|\boldsymbol{x}\|^2}{L} = \frac{2\alpha^2 - \beta^2/2}{4\alpha^2(\alpha^2 - \beta^2)} + \mathcal{O}(L^{-1}), \;\; \alpha > \beta > 0 \tag{32}$$

by the same numerical integration approximation technique and the Residue Theorem.

Lemma 3 (ODE bound) *Suppose for the input data sequence $\boldsymbol{s} = \sqrt{P/A} = (s_0, \cdots, s_{L-1})^{\mathsf{T}}$, the vM-EM algorithm converges, and the estimation error is denoted by $\Delta\varphi = \hat{\varphi} - \varphi$. Then, for any sufficiently large $L \geq L_0$, there exists some positive $\delta = \delta(L) > 0$ and $C = C(L) > 0$, such that for any noise vector $\boldsymbol{v}$ satisfying $\|\boldsymbol{v}\|/\sqrt{L} \leq \delta$, the estimation error is upper-bounded by*

$$|\Delta\varphi| \leq C \frac{\|\boldsymbol{v}\|}{\sqrt{L}}. \tag{33}$$

Proof. The proof is provided in **Appendix F**. From the above **Lemma 3**, we can directly prove **Theorem 1**. Choose $L \geq L_0$, and then let us compute the MMSE of the vM-EM estimator, i.e.,

$$\mathbb{E}[(\hat{\varphi} - \varphi)^2] = \mathbb{E}[|\Delta\varphi|^2]$$
$$= \mathbb{E}\left[|\Delta\varphi|^2 \middle| \frac{\|\boldsymbol{v}\|}{\sqrt{L}} \leq \delta\right] \mathbb{P}\left[\frac{\|\boldsymbol{v}\|}{\sqrt{L}} \leq \delta\right] + \mathbb{E}\left[|\Delta\varphi|^2 \middle| \frac{\|\boldsymbol{v}\|}{\sqrt{L}} > \delta\right] \mathbb{P}\left[\frac{\|\boldsymbol{v}\|}{\sqrt{L}} > \delta\right]$$
$$\leq C^2 \mathbb{E}\left[\frac{\|\boldsymbol{v}\|^2}{L}\right] + \pi^2 \mathbb{P}\left[\frac{\|\boldsymbol{v}\|}{\sqrt{L}} > \delta\right] \tag{34}$$
$$\overset{(a)}{\leq} C^2 \sigma_v^2 + \pi^2 \frac{\mathbb{E}\left[\left(\frac{\|\boldsymbol{v}\|}{\sqrt{L}}\right)^r\right]}{\delta^r},$$

where (a) comes from applying the Markov inequality, and $r > 0$ can be arbitrarily chosen. Particularly, by setting $r = 2$, we obtain

$$\mathbb{E}[(\hat{\varphi} - \varphi)] \leq \sigma_v^2(C^2 + \pi^2/\delta^2) = \mathcal{O}(\bar{\gamma}^{-1}), \tag{35}$$

which completes the proof of **Theorem 1**.

Remark 3 *The conclusion of* **Theorem 1** *guarantees that as* $\bar{\gamma} \to \infty$, *the MSE of the vM-EM estimator decays at a rate of at least* $(\bar{\gamma})^{-1}$. *This conclusion is verified in the following numerical simulation, where it is shown that the MSE curve of the proposed vM-EM algorithm has an asymptotic slope of* -1 *in the logarithmic coordinate.*

Authors: Many thanks again for your valuable time and efforts to review this paper.

Sincerely,
The Authors