# Near-Field Beamforming for Stacked Intelligent Metasurfaces-assisted MIMO Networks

Anastasios Papazafeiropoulos, Pandelis Kourtessis, Symeon Chatzinotas, Dimitra I. Kaklamani, Iakovos S. Venieris

*Abstract*—Stacked intelligent metasurfaces (SIMs) have recently gained significant interest since they enable precoding in the wave domain that comes with increased processing capability and reduced energy consumption. The study of SIMs and high frequency propagation make the study of the performance in the near field of crucial importance. Hence, in this work, we focus on SIM-assisted multiuser multiple-input multiple-output (MIMO) systems operating in the near field region. To this end, we formulate the weighted sum rate maximisation problem in terms of the transmit power and the phase shifts of the SIM. By applying a block coordinate descent (BCD)-relied algorithm, numerical results show the enhanced performance of the SIM in the near field with respect to the far field.

*Index Terms*—Reconfigurable intelligent surface (RIS), stacked intelligent metasurfaces (SIM), near-field communications, 6G networks.

## I. Introduction

Reconfigurable intelligent surface (RIS) has recently emerged as a fundamental technology that increases network capacity while accounting for energy sustainability [1]. Generally, an RIS consists of an artificial surface with a large number of nearly passive elements that can shape the propagation environment [2], [3].

Most existing works on RIS have relied on the assumption of a single-layer surface, which limits the degrees of freedom concerning the adjustment of the beam patterns [2]–[4]. Moreover, it has been shown that conventional RISs do not have the capability of inter-user interference suppression [4].

These observations motivated the proposition of stacked intelligent metasurface (SIM), which includes a stack of an array of intelligent surfaces similar to the structure of artificial neural networks [5]. Note that a SIM is not a mathematical abstraction. Among its remarkable properties, we draw attention to its processing capability, where the forward propagation eventuates at the speed of light. Specifically, a SIM-based transceiver of point-to-point multiple-input multiple-output (MIMO) communication systems has been proposed in [5], where the combining and the precoding are implemented as the electromagnetic (EM) waves propagate through the corresponding SIMs. In [6] and [7], a SIM is integrated at the base station (BS) to enable beamforming in the wave domain based on instantaneous and statistical CSI, respectively. In [8] and [9], contrary to [5] and [6], we have proposed more general hybrid wave-digital architectures and more efficient

algorithms that enable the simultaneous optimization of all parameters.

In parallel, as we move on to higher frequencies and to larger services, the region of a near field may include distances of several hundred metres [10]. Notably, in the near-field region, EM waves exhibit distinct propagation characteristics compared to the far field, e.g., from planar-wave propagation we result in spherical-wave propagation [11]. The spherical wave propagation in the near-field introduces a new distance dimension, which facilitates interference mitigation and increases the performance. Although several works have studied the performance of RIS in the near-field region [10], [11], no work has studied studied the performance of SIM in this region, which motivates this work.

In this work, we consider a SIM-assisted multiuser MIMO operating the near field.[1] In other words, contrary to previous works [5]–[9], we focus on the performance in the near field and we assume multi-antenna users. We develop the weighted sum rate maximisation problem to optimize both the transmit power and the SIM by applying a block coordinate descent (BCD)-based algorithm. Among the results, we observe that the use of a SIM in the near field results in increased weighted sum rate compared to far-field beamforming.

*Notation*: Matrices and vectors are represented by boldface upper and lower case symbols, respectively. The notations $(\cdot)^\mathsf{T}$, $(\cdot)^\mathsf{H}$, and $\mathrm{tr}(\cdot)$ denote the transpose, Hermitian transpose, and trace operators, respectively. Also, the symbol $\mathbb{E}[\cdot]$ denotes the expectation operator. The notation $\mathrm{diag}(\mathbf{A})$ represents a vector with elements equal to the diagonal elements of $\mathbf{A}$. The notation $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$ represents a circularly symmetric complex Gaussian vector with zero mean and a covariance matrix $\boldsymbol{\Sigma}$.

## II. System and Channel Models

In this section, we present the system and channel models for the SIM-assisted near-field multi-user MIMO communication system.

### A. System Model

We consider the downlink of a multiuser MIMO wireless system with a BS equipped with a uniform linear array (ULA) of $M_{\mathrm{BS}}$ antennas that serve $K$ multi-antenna users, each having $M$ antennas. A SIM is incorporated into the BS to provide precoding in the EM wave domain, as shown in Fig. 1. The SIM is implemented by an array of $L$ metasurfaces that each one of them is made by a uniform planar array (UPA) of $N = N_y \times N_z$ meta-atoms with $N_y$ and $N_z$ denoting the

[1]As mentioned before, contrary to an RIS, a SIM has an enhanced capability of adjusting the beam patterns, can suppress the multi-user interference, the forward propagation eventuates at the speed of light, and the precoding takes place as the EM waves propagate through it. The latter two mean that there is no need for digital beamforming with the accompanied radio frequency (RF) chains, which significantly reduces the hardware cost and energy consumption, while the precoding delay is reduced since the processing performed in the wave domain.
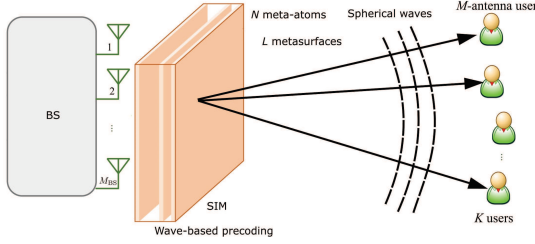
Fig. 1: A SIM-aided MIMO system in the near field.

number of meta-atoms in the directions of $y$ and $z$ axes. We denote the sets $\mathcal{N} = \{1, \ldots, N\}$, $\mathcal{L} = \{1, \ldots, L\}$, $\mathcal{M} = \{1, \ldots, M\}$, and $\mathcal{K} = \{1, \ldots, K\}$. All metasurfaces consist of an identical number of meta-atoms. Also, the SIM is connected to a smart controller to optimize the phase shifts of the EM waves propagated through each of its meta-atoms. We denote $\mathbf{\Phi}^l = \mathrm{diag}(\boldsymbol{\phi}_l) \in \mathbb{C}^{N \times N}$, where $\boldsymbol{\phi}_l = [\phi_1^l, \ldots, \phi_N^l]^\mathsf{T} \in \mathbb{C}^{N \times 1}$ is the phases-shift matrix of the SIM. In particular, we have $\phi_n^l = e^{j\theta_n^l}$, where $\theta_n^l \in [0, 2\pi)$, $n \in \mathcal{N}, l \in \mathcal{L}$, the phase shift by the $n$-th meta-atom on the $l$-th transmit metasurface layer. We assume that the phase shifts are continuously-adjustable. We have assumed that the modulus of the elements of the coefficient matrix equals 1 to assess the maximum achievable rate [12].

### B. Channel Model

The coefficient of transmission from the $\tilde{n}$th meta-atom on the $(l-1)$th layer to the $n$th meta-atom on the $l$th layer of the SIM is provided by the Rayleigh-Sommerfeld diffraction theory [13] as

$$w_{n,\tilde{n}}^l = \frac{A_t \cos x_{n,\tilde{n}}^l}{r_{n,\tilde{n}}^l} \left( \frac{1}{2\pi r_{n,\tilde{n}}^l} - j\frac{1}{\lambda} \right) e^{j2\pi r_{n,\tilde{n}}^l/\lambda}, l \in \mathcal{L}, \quad (1)$$

where $A_t$ denotes the area of each meta-atom, $r_{n,\tilde{n}}^l$, is the transmission distance, and $x_{n,\tilde{n}}^l$ is the angle between the propagation direction and the normal direction of the $(l-1)$th metasurface layer,. Overall, the SIM can be modeled as

$$\mathbf{G} = \mathbf{\Phi}^L \mathbf{W}^L \cdots \mathbf{\Phi}^2 \mathbf{W}^2 \mathbf{\Phi}^1 \in \mathbb{C}^{N \times N}, \quad (2)$$

where $\mathbf{W}^l \in \mathbb{C}^{N \times N}, l \in \mathcal{L}/\{1\}$ is the coefficient matrix of transmission between the $(l-1)st$ layer and the $l$th layer. In the case of $\mathbf{W}_k^1 \in \mathbb{C}^{N \times M_{\mathrm{BS}}}$, it is the transmission coefficient matrix between the $M_{\mathrm{BS}}$ BS antennas and the first layer of the SIM.

Regarding the description of the near-field LoS channel and without any loss of generality, we assume that the SIM is implemented in the $YZ$-plane while all the users are located in the $XY$-plane. The reference element of the SIM and the reference antenna of user $k$ are given by $(0, y_s, z_s)$ and $(x_k, y_k, 0)$, respectively. The coordinates of the element $n$ of the outer surface of the SIM are given by

$$\mathbf{s}_n = [0, c_y(n)d_S + y_s, c_z(n)d_S + z_f]^T, \quad (3)$$

where $c_y(n) = \mod(n-1, N_y)$ and $c_z(n) = \lfloor n-1/N_y \rfloor$ and $d_S = \lambda/2$ is the SIM element spacing. In the case of the ULAs of users, we assume that they are parallel to the $y$-axis. Hence, the coordinates of antenna $m$ of user $k$ are given by

$$\mathbf{u}_m^k = [x_k, (m-1)d_m + y_z, 0]^T, \quad (4)$$

where $d_m = \lambda/2$ is the user antenna spacing.

The near-field LoS channel between the $n$-th STAR element and the $m$-th antenna of user $k$, where the EM signals undergo free space path-loss while propagating in spherical wavefronts, is given by [14]

$$[\mathbf{H}_k]_{mn} = \alpha_{mn}^k \exp(-j2\pi r_{mn}^k/\lambda), \quad (5)$$

where $r_{mn}^k = \|\mathbf{u}_m^k - \mathbf{s}_n\|_2$ and $\alpha_{mn}^k = \frac{\lambda}{4\pi r_{mn}^k}$ denote the distance between the $m$-th antenna of user $k$ and the $n$-th element of the outer surface of the SIM and the free space path-loss coefficient, respectively. Notably, $r_{mn}^k$ can be written as

$$r_{mn}^k = \|[\hat{d}_k \cos \hat{\vartheta}_k, \hat{d}_k \sin \hat{\vartheta}_k + (m-1)\bar{d}_m, 0]^T - \mathbf{s}_n\|_2, \quad (6)$$

where $\hat{d}_k$ and $\hat{\vartheta}_k = \arctan(y_k/x_k)$ are the distance and angle of user $k$, respectively.

*Remark 1:* The term $r_{mn}^k$ includes both the distance and angle information of user $k$. This property results in performance and coverage improvements in multi-user wireless systems since the SIM can transmit the signal not only at a certain angle but also at a certain distance.

*Remark 2:* If we assumed parallel waves, i.e., a far-field LoS channel between the SIM and user $k$, it would be written as

$$\mathbf{H}_k^{\mathrm{far}} = \sqrt{\beta_k MN} \mathbf{e}_k(\alpha_k) \mathbf{e}_{\mathrm{SIM}}^{\mathsf{H}}(\varphi_k, \vartheta_k), \quad (7)$$

where $\beta_k$ is the path-loss coefficient between the SIM and user $k$, which is the same as in the case of near field. Also, $\varphi_k$, $\vartheta_k$ denote the azimuth AoD and the elevation AoD with respect to the SIM, and $\alpha_k$ is the AoA concerning user $k$. The AoA and AoD are defined based on the location of user $k$ with respect to the SIM. Moreover, by denoting $\kappa = 2\pi/\lambda$, we have

$$\mathbf{e}_k(\alpha_k) = [1, e^{j\kappa d_m \cos \alpha_k}, \cdots, e^{j\kappa(M-1)d_m \cos \alpha_k}]^\mathsf{T} \quad (8)$$

$$\mathbf{e}_{\mathrm{SIM}}(\varphi_k, \vartheta_k) = [1, e^{j\kappa d_S \sin \varphi_k \sin \vartheta_k}, \ldots, e^{j\kappa(N_x-1)d_S \sin \varphi_k \sin \vartheta_k}]^\mathsf{T}$$
$$\otimes [1, e^{j\kappa d_S \cos \vartheta_k}, \cdots, e^{j\kappa(N_y-1)d_S \cos \vartheta_k}]^\mathsf{T}. \quad (9)$$

*Remark 3:* The near-field LoS channel allows an increase in DoFs by providing multiple information streams since the rank of the LoS channel in the near-field case can be approximated with the DoFs of the spheroidal waves, which can also be enhanced with distance [15]. On the contrary, the far-field LoS MIMO channel in (7) is just of rank one.

## III. DOWNLINK DATA TRANSMISSION AND ACHIEVABLE RATE

In this section, we present the downlink transmission model, which will lead to the corresponding achievable rate of a SIM-assisted near-field MIMO system.

### A. Downlink Data Transmission

During the downlink transmission, wave-based beamforming takes place thanks to the SIM, which is in contrast to conventional digital precoding, where each symbol is assigned to an individual beamforming vector. Under this setting, the BS selects a set of $K$ appropriate antennas from the total of $M_{\mathrm{BS}}$ antennas since each data stream has to be transmitted directly from the corresponding antenna at the BS. For the sake of simplicity, we assume that $M_{\mathrm{BS}} = KM$ [5].

The received signal by user $k$ is given by

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{G} \mathbf{W}_k^1 \mathbf{P}_k \mathbf{x}_k + \mathbf{H}_k \mathbf{G} \sum_{i \neq k}^K \mathbf{W}_i^1 \mathbf{P}_i \mathbf{x}_i + \mathbf{n}_k, \quad (10)$$

where $\mathbf{n}_k \in \mathbb{C}^{M \times 1}$ is AWGN with distribution $\mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$ and $\mathbf{x}_k \in \mathbb{C}^{M \times 1}$ is the symbol vector with $\mathbb{E}\{\mathbf{x}_k \mathbf{x}_k^{\mathsf{H}}\} = \mathbf{I}_M$ and $\mathbb{E}\{\mathbf{x}_k \mathbf{x}_{k'}^{\mathsf{H}}\} = \mathbf{0}$, for $k \neq k'$. Also, $\mathbf{P}_k \in \mathbb{C}^{M \times M}$ is a diagonal matrix, where its $m$-th diagonal entry denotes the square root of the power allocated to the $m$-th data stream of the $k$-th user. The total transmit power constraint at the BS is given by

$$\sum_{k=1}^{K} \|\mathbf{P}_k\|_F^2 \leq P, \tag{11}$$

where $P$ denotes the transmit power budget at the BS.

### B. Achievable Rate

The achievable rate of $k$-th user is given by

$$R_k(\mathbf{P}, \boldsymbol{\Phi}) = \log |\mathbf{I} + \mathbf{H}_k \mathbf{W}_k^1 \mathbf{P}_k^2 (\mathbf{W}_k^1)^{\mathsf{H}} \mathbf{H}_k^{\mathsf{H}} \mathbf{Q}_k^{-1}|, \tag{12}$$

where $\mathbf{P} = \{\mathbf{P}_i, \forall i \in \mathcal{K}\}$, $\boldsymbol{\Phi} = \{\boldsymbol{\Phi}^l, \forall l \in \mathcal{L}\}$, and $\mathbf{Q}_k$ is the interference-plus-noise covariance matrix given by

$$\mathbf{Q}_k = \sum_{i \neq k}^{K} \mathbf{H}_k \mathbf{G} \mathbf{W}_i^1 \mathbf{P}_i^2 (\mathbf{W}_i^1)^{\mathsf{H}} \mathbf{G}^{\mathsf{H}} \mathbf{H}_k^{\mathsf{H}} + \sigma^2 \mathbf{I}_M. \tag{13}$$

## IV. PROBLEM FORMULATION AND OPTIMIZATION

Herein, we first formulate the optimization problem, and then we solve it by using the BCD method.

### A. Problem Formulation

Herein, we consider the maximization of the weighted sum rate by optimizing the transmit power and the phase-shifts of the SIM. In particular, the optimization problem can be formulated as

$$(\mathcal{P}1) \quad \max_{\mathbf{P}, \boldsymbol{\Phi}} \sum_{i=1}^{K} \eta_i R_i(\mathbf{P}, \boldsymbol{\Phi}) \tag{14a}$$

$$\text{s.t} \quad \mathbf{G} = \boldsymbol{\Phi}^L \mathbf{W}^L \cdots \boldsymbol{\Phi}^2 \mathbf{W}^2 \boldsymbol{\Phi}^1 \mathbf{W}^1, \tag{14b}$$

$$\boldsymbol{\Phi}^l = \text{diag}(\phi_1^l, \dots, \phi_N^l), l \in \mathcal{L}, \tag{14c}$$

$$|\phi_n^l| = 1, n \in \mathcal{N}, l \in \mathcal{L}, \tag{14d}$$

$$\sum_{k=1}^{K} \|\mathbf{P}_k\|_F^2 \leq P, \tag{14e}$$

where the constraints (14b)-(14d) concern the SIM elements, while the constraint (14e) corresponds to the power constraint of the BS. Note that $\eta_k$ denotes the access priority for user $k$.

The coupling between the optimization variables of the transmit power and phase shifts of each surface, which also have constant modulus constraints, and the non-convexity of problem $(\mathcal{P}1)$, make the solution challenging. To this end, we reformulate $(\mathcal{P}1)$ in terms of the mean-square error (MSE).

Let $\mathbf{U}_k \in \mathbb{C}^{M \times M}$ be the linear combining matrix for user $k$, the estimated signal vector of user $k$ can be written as

$$\tilde{\mathbf{x}}_k = \mathbf{U}_k^{\mathsf{H}} \mathbf{r}_k. \tag{15}$$

Application of the weighted minimum mean square error (WMMSE) technique allows to transform (14a) into a more tractable expression. Thus, $(\mathcal{P}1)$ can be written as [16]

$$(\mathcal{P}2) \quad \min_{\mathbf{Z}, \mathbf{U}, \mathbf{P}, \boldsymbol{\Phi}} \sum_{i=1}^{K} \eta_i g_i(\mathbf{Z}, \mathbf{UP}, \boldsymbol{\Phi}) \tag{16a}$$

$$\text{s.t} \quad (14b) - (14e), \tag{16b}$$

where the function $g_i(\mathbf{Z}, \mathbf{U}, \mathbf{P}, \boldsymbol{\Phi})$ is given by

$$g_i(\mathbf{Z}, \mathbf{U}, \mathbf{P}, \boldsymbol{\Phi}) = \log |\mathbf{Z}_i| - \text{tr}(\mathbf{Z}_i \mathbf{E}_i) + M \tag{17}$$

with $\mathbf{U} = \{\mathbf{U}_k, \forall i \in \mathcal{K}\}$, $\mathbf{Z} = \{\mathbf{Z}_i \succeq \mathbf{0}, \forall i \in \mathcal{K}\}$, and $\mathbf{E}_k$ corresponding to the set of combining matrices, the set of auxiliary matrices, and the MSE of user $k$.

The MSE of user $k$ is obtained as

$$\begin{aligned} \mathbf{E}_k &= \mathbb{E}\{(\tilde{\mathbf{x}}_k - \mathbf{x}_k)(\tilde{\mathbf{x}}_k - \mathbf{x}_k)^{\mathsf{H}}\} \\ &= (\mathbf{U}_k^{\mathsf{H}} \mathbf{H}_k \mathbf{G} \mathbf{W}_k^1 \mathbf{P}_k - \mathbf{I}_M)(\mathbf{U}_k^{\mathsf{H}} \mathbf{H}_k \mathbf{G} \mathbf{W}_k^1 \mathbf{P}_k - \mathbf{I}_M)^{\mathsf{H}} \\ &\quad + \mathbf{U}_k^{\mathsf{H}} \mathbf{H}_k \mathbf{G} \mathbf{W}_{\bar{k}}^1 \mathbf{P}_{\bar{k}}^2 (\mathbf{W}_{\bar{k}}^1)^{\mathsf{H}} \mathbf{G}^{\mathsf{H}} \mathbf{H}_k^{\mathsf{H}} \mathbf{U}_k + \sigma^2 \mathbf{U}_k^{\mathsf{H}} \mathbf{U}_k, \end{aligned} \tag{18}$$

where $\mathbf{P}_{\bar{k}} = \text{diag}(\mathbf{P}_1, \dots, \mathbf{P}_{k-1}, \mathbf{P}_{k+1}, \dots, \mathbf{P}_K)$ and $\mathbf{W}_{\bar{k}}^1 = [\mathbf{W}_1^1, \dots, \mathbf{W}_{k-1}^1, \mathbf{W}_{k+1}^1, \dots, \mathbf{W}_K^1] \in \mathbb{C}^{N \times (K-1)M}$.

The tractability of the reformulated optimization problem $(\mathcal{P}2)$ because of the concavity of the objective function with respect to $\mathbf{Z}, \mathbf{U}, \mathbf{P}, \boldsymbol{\Phi}$ allows its solution by using the BCD method in terms of four blocks, where in each iteration, one variable is optimized while keeping the other variables fixed.

*1) Optimization with respect to $\mathbf{U}$:* The optimal $\mathbf{U}$ is obtained from (16a) by solving $\partial g_i / \partial \mathbf{U}_i = 0$, $\forall i \in \mathcal{K}$ while keeping $\mathbf{Z}, \mathbf{P}, \boldsymbol{\Phi}$ fixed. Specifically, we obtain

$$\mathbf{U}_k^{\text{opt}} = (\mathbf{Q}_k + \mathbf{H}_k \mathbf{W}_k^1 \mathbf{P}_k^2 (\mathbf{W}_k^1)^{\mathsf{H}} \mathbf{H}_k^{\mathsf{H}})^{-1} \mathbf{H}_k \mathbf{W}_k^1 \mathbf{P}_k. \tag{19}$$

*2) Optimization with respect to $\mathbf{Z}$:* In a similar way, the optimal $\mathbf{Z}$ is obtained from (16a) by solving $\partial g_i / \partial \mathbf{Z}_i = 0$, $\forall i \in \mathcal{K}$ while keeping $\mathbf{U}, \mathbf{P}, \boldsymbol{\Phi}$ fixed. Thus, we obtain

$$\mathbf{Z}_k^{\text{opt}} = (\mathbf{E}_k^{\text{opt}})^{-1}, \tag{20}$$

where $\mathbf{E}_k^{\text{opt}}$ results by substituting $\mathbf{U}_k^{\text{opt}}$ into (18).

Note that by using the Woodbury matrix identity, the objective function $(\mathcal{P}2)$ can be rewritten as

$$\begin{aligned} g_i(\mathbf{Z}, \mathbf{U}, \mathbf{P}, \boldsymbol{\Phi}) &= \log \det((\mathbf{E}_k^{\text{opt}})^{-1}) \\ &= \log \det(\mathbf{I}_M + \mathbf{H}_k \mathbf{W}_k^1 \mathbf{P}_k^2 (\mathbf{W}_k^1)^{\mathsf{H}} \mathbf{H}_k^{\mathsf{H}} \mathbf{Q}_k^{-1}), \end{aligned} \tag{21}$$

where we have substituted (19) and (20). We observe that (21) equals $R_i(\mathbf{P}, \boldsymbol{\Phi})$, which means that $(\mathcal{P}1)$ and $(\mathcal{P}2)$ are equivalent regarding the solutions of $\mathbf{P}$ and $\boldsymbol{\Phi}$.

*3) Optimization with respect to $\mathbf{P}$:* In this case, with $\mathbf{Z}, \mathbf{U}, \boldsymbol{\Phi}$ assumed fixed, the weighted MSE minimization problem $(\mathcal{P}2)$ can be written as

$$(\mathcal{P}2.1) \quad \min_{\mathbf{P}} \sum_{i=1}^{K} g(\mathbf{P}) \tag{22a}$$

$$\text{s.t} \quad (14e), \tag{22b}$$

where $g(\mathbf{P}) = \sum_{i=1}^{K}(\text{tr}(\mathbf{P}_i^{\mathsf{H}} \mathbf{A} \mathbf{P}_i) - 2\text{tr}(\text{Re}(\mathbf{B}_i \mathbf{P}_i)))$, in which, we have denoted $\mathbf{B}_i = \eta_i \mathbf{Z}_i \mathbf{U}_i^{\mathsf{H}} \mathbf{H}_i \mathbf{G} \mathbf{W}_i^1$ and $\mathbf{A} = \sum_{k=1}^{K} \eta_k (\mathbf{W}_k^1)^{\mathsf{H}} \mathbf{G}^{\mathsf{H}} \mathbf{H}_k^{\mathsf{H}} \mathbf{U}_k \mathbf{Z}_k \mathbf{U}_k^{\mathsf{H}} \mathbf{H}_k \mathbf{G} \mathbf{W}_k^1$.

The problem $(\mathcal{P}2.1)$ is a a quadratically constrained quadratic program, which can be solved by using CVX [17]. Also, $(\mathcal{P}2.1)$ could be solved in closed form with reduced complexity by applying the Lagrangian dual decomposition method [18].

*4) Optimization with respect to $\boldsymbol{\Phi}$:* Given $\mathbf{Z}, \mathbf{U}, \mathbf{P}$, after substituting (18) into (17), the optimization problem with respect to the phase shifts matrices is formulated as

$$(\mathcal{P}2.2) \quad \min_{\boldsymbol{\phi}_l} \sum_{i=1}^{K} g_i(\phi_l) \tag{23a}$$

$$\text{s.t} \quad (14b) - (14d), \tag{23b}$$

where

$$g_i(\boldsymbol{\phi}_l) = \text{tr}(\mathbf{G}^\mathsf{H}\mathbf{C}_i\mathbf{G}\mathbf{D} - 2\,\text{Re}(\mathbf{E}_i\mathbf{G})). \tag{24}$$

with $\mathbf{C}_i = \eta_i\mathbf{H}_k^\mathsf{H}\mathbf{U}_k\mathbf{Z}_k\mathbf{U}_k^\mathsf{H}\mathbf{H}_k$, $\mathbf{D} = \mathbf{W}_i^1(\sum_{k=1}^K \mathbf{P}_k^2)(\mathbf{W}_i^1)^\mathsf{H}$, and $\mathbf{E}_i = \eta_i\mathbf{W}_i^1\mathbf{P}_i\mathbf{Z}_k\mathbf{U}_k^\mathsf{H}\mathbf{H}_k$ The problem $(\mathcal{P}2.2)$ is non-convex due to the unit-modulus constraint and the non-convex constraints regarding the phase shifts.

Herein, we are going to apply the projected gradient ascent algorithm until convergence to a locally optimal solution to $(\mathcal{P}2.2)$. Specifically, by starting from $\boldsymbol{\phi}_l^0$, we shift along the gradient of $g_i(\boldsymbol{\phi}_l)$. Next, we project the new point $\boldsymbol{\phi}_l$ onto $\Phi_l$ to hold the new points in the feasible set. For ease of exposition, we have defined the set $\Phi_l = \{|\phi_n^l| = 1, n \in \mathcal{N}\}$. Note that $\phi_n^l$ has to be found inside the unit circle because of the unit-modulus constraint. Also, we denote $P_{\Phi_l}(\cdot)$ the projection onto $\Phi_l$.

The following iteration describes the algorithm. In particular, we have

$$\boldsymbol{\phi}_l^{i+1} = P_{\Phi_l}(\boldsymbol{\phi}_l^i + \mu_i\nabla_{\boldsymbol{\phi}_l}g_j(\boldsymbol{\phi}_j^i)). \tag{25}$$

The step size is obtained by the Armijo-Goldstein backtracking line search method. We have $\mu_i = T_i\bar{\kappa}^{m_i}$, where $\bar{\kappa} \in (0,1)$ and $T_i > 0$ with $m_i$ being the smallest positive integer that satisfies

$$g_j(\boldsymbol{\phi}_l^{i+1}) \geq B_{T_i\bar{\kappa}^{m_i}}(\boldsymbol{\phi}_l^i; \boldsymbol{\phi}_l^{i+1}), \tag{26}$$

where

$$B_\mu(\boldsymbol{\phi}_l; \mathbf{x}) = g_j(\boldsymbol{\phi}_l) + \langle \nabla_{\boldsymbol{\phi}_l}g_j(\boldsymbol{\phi}_l), \mathbf{x} - \boldsymbol{\phi}_l \rangle - \frac{1}{\mu}\|\mathbf{x} - \boldsymbol{\phi}_l\|_2^2 \tag{27}$$

is the quadratic approximation of $f(\boldsymbol{\phi}_l)$.

The following lemma provides the complex-valued gradient.

*Lemma 1:* The complex gradient $\nabla_{\boldsymbol{\phi}_l}g_i(\boldsymbol{\phi}_l)$ is given in closed-form by

$$\nabla_{\boldsymbol{\phi}_l}g_i(\boldsymbol{\phi}_l) = \text{diag}(\mathbf{J}_l^*\mathbf{D}^\mathsf{T}\mathbf{G}^\mathsf{T}\mathbf{C}_i^\mathsf{T}\mathbf{R}_l^*). \tag{28}$$

*Proof:* To obtain $\nabla_{\boldsymbol{\phi}_l}g_i(\boldsymbol{\phi}_l)$, we first derive the following differential

$$\begin{aligned}
d(g_j(\boldsymbol{\phi}_l^i)) &= \text{tr}\,(d(\mathbf{G}^\mathsf{H})\mathbf{C}_i\mathbf{G}\mathbf{D} + \mathbf{G}^\mathsf{H}\mathbf{C}_id(\mathbf{G})\mathbf{D} \\
&\quad - 2\,\text{Re}(\mathbf{E}_id(\mathbf{G}))) \tag{29} \\
&= \text{tr}\,(\mathbf{J}_l^\mathsf{H}d(\boldsymbol{\Phi}_l^\mathsf{H})\mathbf{R}_l^\mathsf{H}\mathbf{C}_i\mathbf{G}\mathbf{D} + \mathbf{G}^\mathsf{H}\mathbf{C}_i\mathbf{R}_ld(\boldsymbol{\Phi}_l)\mathbf{J}_l\mathbf{D} \\
&\quad - 2\,\text{Re}(\mathbf{E}_i\mathbf{R}_ld(\boldsymbol{\Phi}_l)\mathbf{J}_l)), \tag{30}
\end{aligned}$$

where, in (30), we have substituted $d(\mathbf{G}) = \mathbf{R}_ld(\boldsymbol{\Phi}_l)\mathbf{J}_l$ with $\mathbf{R}_l = \boldsymbol{\Phi}_L\mathbf{W}^L\cdots\boldsymbol{\Phi}_{l+1}\mathbf{W}^{l+1}$, and $\mathbf{J}_l = \mathbf{W}^l\boldsymbol{\Phi}_{l-1}\mathbf{W}^{l-1}\cdots\boldsymbol{\Phi}_1$.

Having derived the differential, we obtain

$$\nabla_{\boldsymbol{\phi}_l}g_i(\boldsymbol{\phi}_l) = \frac{\partial}{\partial\boldsymbol{\phi}_l^*}g_i(\boldsymbol{\phi}_l) = \text{diag}(\mathbf{J}_l^*\mathbf{D}^\mathsf{T}\mathbf{G}^\mathsf{T}\mathbf{C}_i^\mathsf{T}\mathbf{R}_l^*). \tag{31}$$

∎

*5) Overall Algorithm:* A summary of the proposed algorithm for solving (16) is provided in Algorithm 1. Given that (16a) is non-decreasing in each iteration and since it is upper bounded because the transmit power is limited, Algorithm 1 is guaranteed to converge to a stationary point.

The complexity of Algorithm 1 is mainly yielded from Problems $(\mathcal{P}2.1)$ and $(\mathcal{P}2.2)$. In particular, the complexity of $(\mathcal{P}2.1)$ is $O(KM_{\text{BS}}^3)$ and the complexity $(\mathcal{P}2.2)$ is $O(KL(N^2+M^2))$. Hence, the overall computational complexity of Algorithm 1 is provided by $O(I_{\text{BCD}}K(M_{\text{BS}}^3 + L(N^2 + M^2)))$, where $I_{\text{BCD}}$ is the number of BCD iterations.

---

**Algorithm 1** Overall Algorithm

1: Initialize feasible $\mathbf{P}$ and $\boldsymbol{\Phi}$ that satisfy (14b)-(14e),
2: **repeat**
3:     Given $\mathbf{P}$ and $\boldsymbol{\Phi}$, update the combining matrices $\mathbf{U}$ using (19).
4:     Given $\mathbf{P}$, $\boldsymbol{\Phi}$, and $\mathbf{U}$ update the auxiliary matrices $\mathbf{Z}$ using (20).
5:     Given $\boldsymbol{\Phi}$, $\mathbf{U}$, and $\mathbf{Z}$ update the transmit power $\mathbf{P}$ by solving (22).
6:     Given $\mathbf{P}$, $\mathbf{U}$, and $\mathbf{Z}$ update the phase shifts matrices $\boldsymbol{\Phi}$ by solving (23).
7: **until** The increase of (16) is less than a threshold $\epsilon$.

---

## V. NUMERICAL RESULTS

Herein, we assess the performance of the weighted sum rate of a SIM-assisted near-field multi-user MIMO communication system by depicting analytical results and Monte Carlo simulations. The layout assumes that the SIM is parallel to the $y - z$ plane and centered along the $x-$axis. The users are located in a circular region with a radius between 2 m and 4 m. The spacing between adjacent meta-atoms is assumed to be $\lambda/2$ while the size of each meta-atom is $\lambda/2 \times \lambda/2$. Also, we have $d_{\text{SIM}} = T_{\text{SIM}}/L$, where $T_{\text{SIM}} = 5\lambda$ is the thickness of the SIM. Regarding the distance between the $\tilde{n}-$th meta-atom of the $(l-1)-$st metasurface and the $n-$th meta-atom of the $l-$st metasurface, it is given by $d_{n,\tilde{n}}^l = \sqrt{d_{\text{SIM}}^2 + d_{n,\tilde{n}}^2}$, where

$$d_{n,\tilde{n}} = \frac{\lambda}{2}\sqrt{\lfloor|n - \tilde{n}|/N_x\rfloor^2 + [\text{mod}(|n - \tilde{n}|, N_x)]^2}. \tag{32}$$

In the case of the transmission distance between the $m-$th antenna of the BS and the $\tilde{n}$-th meta-atom on the first metasurface layer, it is given by (33). Moreover, we have $\cos x_{n,\tilde{n}}^l = d_{\text{SIM}}/d_{n,\tilde{n}}^l, \forall l \in \mathcal{L}$. The path loss is given by $\tilde{\beta}_k = C_0(d_k/\hat{d})^{-\alpha}$, where $C_0 = (\lambda_2/4\pi\hat{d})$ is the free space path loss at the reference distance of $\hat{d} = 1$ m, and $\alpha = 2.5$ is the path-loss exponent. The carrier frequency and the system bandwidth are 10 GHz and 20 MHz, respectively. Furthermore, we assume $M_{\text{BS}} = 4$, $K = 4$, $M = 2$, $N = 40$, and $L = 4$. The Rayleigh distance of a SIM with $N = 40$ elements at the above frequency of 5 GHz is approximately 5 m, which means that all users are inside the region of the near field.

To show better the improvement in the near field, we assume two settings: i) a random user setup, where users lie in different angles, and ii) an inline user setup, where users lie in the same line, i.e., they are located in the same angle with respect to the SIM.

In Fig. 2, we depict the achievable weighted sum rate versus the nu6mber of elements $N$ of each surface while varying the number of surfaces $L$. For the sake of comparison, we have also depicted the far-field scenario obtained based on (7). In this case, the angles of the channels are obtained based on the Uniform distribution. We observe that the rate increases with $N$ because a larger $N$ results in a higher beamforming gain but in the case of the far-field, especially in the inline user setting, the increase is insignificant because both the gain and the interference become higher with $N$. Moreover, the near-field beamforming achieves a higher rate than the far-field one because the near-field LoS channels transmit more data streams to the users since they have a higher rank. In other words, a degrees of freedom enhancement appears. Also, under the same type of beamforming, i.e., in the near-field or far-field beamforming, the random user scenario appears better

$$d_{\tilde{n},m}^1 = \sqrt{d_{\text{SIM}}^2 + \left[\left(\text{mod}(\tilde{n}-1, N_x) - \frac{N_x-1}{2}\right)\frac{\lambda}{2} - \left(m - \frac{N_t+1}{2}\right)\frac{\lambda}{2}\right]^2 + \left(\lceil \tilde{n}/N_x \rceil - \frac{N_y+1}{2}\right)^2 \frac{\lambda_2}{4}}. \tag{33}$$
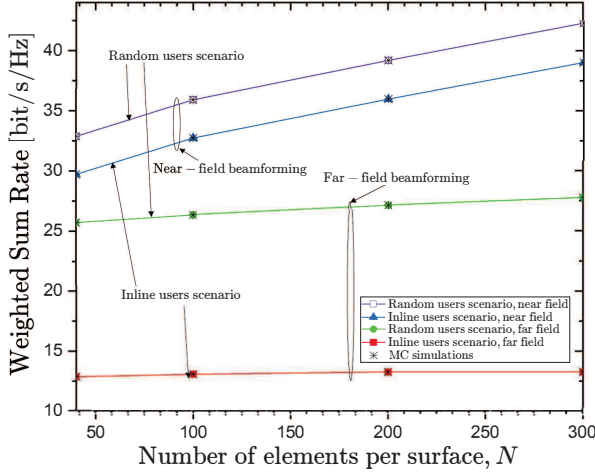


Fig. 2: Weighted sum rate of the SIM-aided MIMO architecture with respect to the number of meta-atoms $N$.



Fig. 3: Weighted sum rate of the SIM-aided MIMO architecture with respect to the transmit power $P$.

performance compared to the inline user scenario because the latter comes with higher inter-user interference. The reason is that in the inline user scenario, only the user distance information is leveraged to mitigate the inter-user interference, while in the random user scenario both the distance and angles information are leveraged. Notably, Monte Carlo (MC) simulations verify the analytical results.

In Fig. 3, we show the achievable weighed sum rate versus the transmit power while varying the number of surfaces $L$ in the scenario of random users setup. As can be seen, the weighted sum rate increases in all cases with power because a higher power budget enables the reception of stronger signals from the users. Also, the SIM outperforms significantly the case of a single surface, i.e., $L = 1$. Generally, by increasing the number of surfaces of the SIM, the rate increases because the SIM can manage to mitigate the inter-user interference in the wave domain [5]. Furthermore, we show the comparison between near and far-field beamforming, and we observe that the former performs better because the latter includes higher inter-user interference.

## VI. CONCLUSION

This paper presented the study of the weighted sum rate of SIM-assisted multiuser MIMO communication systems in the near field. A BCD-based algorithm was used to solve the non-convex optimization. Numerical results showed that near-field beamforming can improve the performance compared to far field and the degrees of freedom are facilitated by the near-field channels.

## REFERENCES

[1] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, 2020.

[2] C. Huang *et al.*, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, 2019.
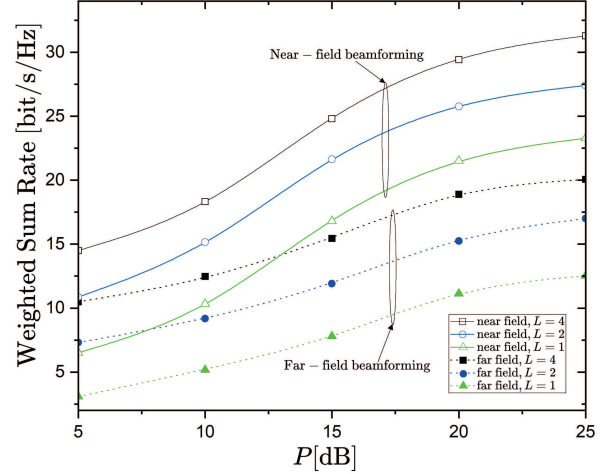
[3] A. Papazafeiropoulos *et al.*, "Achievable rate of a STAR-RIS assisted massive MIMO system under spatially-correlated channels," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 1550–1564, 2024.

[4] H. Guo *et al.*, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064–3076.

[5] J. An *et al.*, "Stacked intelligent metasurfaces for multiuser downlink beamforming in the wave domain," *arXiv preprint arXiv:2309.02687*, 2023.

[6] ——, "Stacked intelligent metasurfaces for efficient holographic MIMO communications in 6G," *IEEE J. Sel. Areas Commun.*, 2023.

[7] A. Papazafeiropoulos *et al.*, "Achievable rate optimization for large stacked intelligent metasurfaces based on statistical CSI," *accepted with Minor Changes in IEEE Wireless Commun. Let.*, pp. 1–1, 2024.

[8] ——, "Achievable rate optimization for stacked intelligent metasurface-assisted holographic MIMO communications," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2024.

[9] A. Papazafeiropoulos, P. Kourtessis, and S. Chatzinotas, "Performance of double-stacked intelligent metasurface-assisted multiuser massive MIMO communications in the wave domain," *arXiv preprint arXiv:2402.16405*, 2024.

[10] E. Björnson and L. Sanguinetti, "Power scaling laws and near-field behaviors of massive MIMO and intelligent reflecting surfaces," *IEEE O. J. Commun. Soc.*, vol. 1, pp. 1306–1324, 2020.

[11] X. Mu *et al.*, "Reconfigurable intelligent surface-aided near-field communications for 6G: Opportunities and challenges," *IEEE Veh. Tech. Mag.*, vol. 19, no. 1, pp. 65–74, 2024.

[12] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, 2019.

[13] X. Lin *et al.*, "All-optical machine learning using diffractive deep neural networks," *Science*, vol. 361, no. 6406, pp. 1004–1008, 2018.

[14] H. Zhang *et al.*, "Beam focusing for near-field multiuser MIMO communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7476–7490, 2022.

[15] D. A. Miller, "Communicating with waves between volumes: Evaluating orthogonal spatial channels and limits on coupling strengths," *Applied optics*, vol. 39, no. 11, pp. 1681–1699, 2000.

[16] Q. Shi *et al.*, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331–4340, 2011.

[17] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab software for disciplined convex programming (2008)," *Web page and software available at https://web.stanford.edu/ boyd/cvxbook*, 2015.

[18] C. Pan *et al.*, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Trans. Wireless Commun.*, vol. 19, no. 8, pp. 5218–5233, 2020.