Sum-Rate Maximization in Holographic MIMO Communications with Stacked Intelligent Metasurfaces

Sajjad Nassirpour, Toan-Van Nguyen, Tharmalingam Ratnarajah, and Duy H. N. Nguyen
Department of Electrical and Computer Engineering, San Diego State University, San Diego, USA
Email: {snassirpour, tnguyen58, duy.nguyen}@sdsu.edu and t.ratnarajah@ieee.org

ICC 2025 - IEEE International Conference on Communications | 979-8-3315-0521-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICC52391.2025.11161367

***Abstract*—In this paper, we investigate holographic multiple-input multiple-output communications in a multi-user network assisted by stacked intelligent metasurface (SIM) elements. We use discrete phase shifters as the SIM elements and explore the sum-rate maximization problem in a downlink scenario. Unlike past studies, we assume that the transmit powers at the base station follow a discrete set of power levels, addressing circuit design limitations in generating continuous-valued powers. We propose an alternative optimization technique to iteratively optimize the transmit powers and SIM elements. Specifically, we introduce the filled function (FF) optimization method to optimize the SIM elements and apply a modified version of the FF method to obtain the optimal transmit powers. We assess our optimization methods against state-of-the-art benchmarks, including successive refinement (SR) and mapped projected gradient ascent (MPGA) for SIM elements, as well as water-filling for optimal transmit powers. Simulation results show that our proposed method outperforms SR and MPGS in terms of sum-rate, even though it uses discrete transmit powers.**

***Index Terms*—Stacked intelligent metasurface, reconfigurable intelligent surface, holographic MIMO, discrete phase shifter.**

I. Introduction

The increasing demand for higher data rates and massive device connectivity are key drivers of the next generation of wireless networks. Massive multiple-input multiple-output (MIMO) is a well-known technology to support multiple users with high data rates, but it presents challenges in terms of high implementation costs and energy consumption. To mitigate these issues, reconfigurable intelligent surface (RIS) has been proposed as an intermediary node between the base station (BS) and users, enhancing data rates and spectral efficiency in a more cost- and energy-efficient way [1]–[3]. Despite the advantages, RIS-assisted networks experience substantial path loss, necessitating the use of large RIS [4], which in turn increases the feedback overhead for channel estimation.

On the other hand, over the past decade, massive MIMO technology has been well developed and understood. Now, a fair question arises: what will the next generation of MIMO systems be? In response to this, the concept of Holographic MIMO (HMIMO) communications has recently been introduced [5], [6]. HMIMO communications refer to the physical process of fully restoring three-dimensional (3D) target scenes in a realistic way, achieved through communication devices equipped with holographic-type radios and electromagnetic (EM)-domain signal processing [6]. However, HMIMO technology is still in its early stages and requires further investigation, particularly concerning practical implementation.

Recently, stacked intelligent metasurface (SIM) has been proposed as an initial step toward realizing HMIMO. SIM consists of multiple layers of RIS elements that can be positioned near the BS, effectively addressing the path loss issue seen in traditional RIS setups. Additionally, SIM offers three key advantages over conventional MIMO systems. First, SIM can replace traditional digital beamforming, eliminating the need for high-resolution digital-to-analog and analog-to-digital converters (DAC/ADC), thereby lowering hardware costs. Second, SIM reduces the number of required radio-frequency (RF) chains, leading to lower energy consumption. Third, SIM minimizes latency by performing precoding in the electromagnetic domain [7]–[9].

In [7], achievable rates in a single-user SIM-assisted MIMO network were studied using an alternative optimization (AO) technique with projected gradient ascent (PGA) for SIM element optimization. Then, [8] extended this to a multi-user setting, applying the water-filling (WF) method for transmit power allocation and PGA for SIM element optimization.

The above studies assumed continuous-valued phase shifters (PSs) as the SIM elements, whereas discrete PSs are more cost-effective. The study [9] considered a multi-user SIM-assisted multiple-input single-output (MISO) network with discrete PSs. This leads to a mixed-integer-non-linear programming (MINLP) problem, which is non-convex and NP-hard. The authors of [9] proposed two methods for optimizing the discrete PSs: (i) *Mapped PGA (MPGA) method:* This approach begins by relaxing the optimization problem under the assumption of continuous-valued PSs. The PGA method is then used to optimize the PSs. Since the resulting solution is not directly applicable to discrete PSs, it is mapped to the nearest discrete PSs. However, MPGA does not directly solve the MINLP problem, which potentially leads to performance degradation. (ii) *Successive refinement (SR) method:* It tackles the MINLP problem directly and optimizes one PS at a time through a one-dimensional search while holding the other PSs constant. It repeats this process until convergence.

Nevertheless, the SR method converges to local optimums, and its performance relies on the initial solution. To address this issue, in this paper, we propose an optimization method

979-8-3315-0521-9/25/$31.00 ©2025 IEEE

based on the filled function (FF) approach to optimize discrete-valued SIM elements in a multi-user MISO network. The idea of the FF optimization method was first proposed for the continuous domain in [10], and later adapted for the discrete domain in [11]. This method allows us to move from one local optimum solution to another better one. The FF method consists of two parts: local search and global search. The local search is similar to the SR method, while the global search defines an auxiliary optimization problem based on the FF to aid the optimization process in finding a new solution, superior to the one obtained from the local search. Moreover, previous works typically assumed continuous-valued transmit powers at the BS and optimized them via the water-filling (WF) method. However, generating such powers is challenging due to circuit design constraints. In this work, we overcome this limitation by considering a discrete set of transmit power levels and applying a modified version of the FF (mFF) method for optimization.

Using discrete transmit powers transforms the MINLP problem into a nonlinear integer programming (NILP) problem. We use the AO technique to iteratively optimize both the transmit powers and the SIM elements until convergence is achieved. We compare the performance of our proposed FF optimization method with five benchmarks, including the SR and MPGA methods, in terms of sum-rate and computational complexity. We show through simulation that the FF method offers a higher sum-rate gain than the SR and MPGA methods, even though the latter utilize continuous-valued transmit powers.

The remainder of this paper is structured as follows: Section II introduces the system model and formulates the problem. Sections III and IV detail our proposed optimization methods. Section V presents the numerical analysis, and Section VI provides the conclusion of the paper.

Notation: We denote vectors, matrices, and scalars using bold lowercase, bold uppercase, and italic letters, respectively. We use $\mathbb{C}^{a\times b}$ to represent the set of complex matrices with dimensions $a \times b$, and $\mathrm{diag}(\mathbf{d})$ to form a diagonal matrix from vector $\mathbf{d}$. The identity matrix of size $K \times K$ appears as $\mathbf{I}_K$, and we express the Euclidean norm and absolute value of $\mathbf{d}$ as $||\mathbf{d}||$ and $|\mathbf{d}|$, respectively. We use $\log_2(\cdot)$ for the logarithmic function with base 2. The conjugate transpose of matrix $\mathbf{D}$ is written as $\mathbf{D}^H$, and we use calligraphic letters to represent sets, with $|\mathcal{D}|$ indicating the cardinality of set $\mathcal{D}$.

II. System Model and Problem Formulation

A. System Model

We focus on a downlink scenario in a SIM-assisted MISO network, as depicted in Fig. 1, where the BS uses low-resolution DACs and a SIM to serve multiple users simultaneously. Here, the goal is to jointly optimize the transmit powers at the BS and the SIM elements to maximize sum-rate.

The network consists of a BS with M antennas, K single-antenna users, and a SIM with L metasurface layers, each containing N meta-atoms, as shown in Fig. 1. We define $\mathcal{K} = \{1, 2, \ldots, K\}$, $\mathcal{N} = \{1, 2, \ldots, N\}$, and $\mathcal{L} = \{1, 2, \ldots, L\}$ as the set of users, meta-atoms, and metasurface layers, respectively. We assume a separate low-error link for the SIM-controller, which is responsible for adjusting the SIM elements. It is important to note that if $M > K$, we can benefit from a digital beamformer at the BS in addition to power allocation and SIM optimization, but this requires high-resolution digital-to-analog converters (DACs). Therefore, in this work, we focus on $M = K$ and aim to maximize sum-rate solely through power allocation and SIM element optimization. In this scenario, the received signal at the k^{th} user, i.e., $\mathsf{Rx}_k, k \in \mathcal{K}$, at time t is given by:

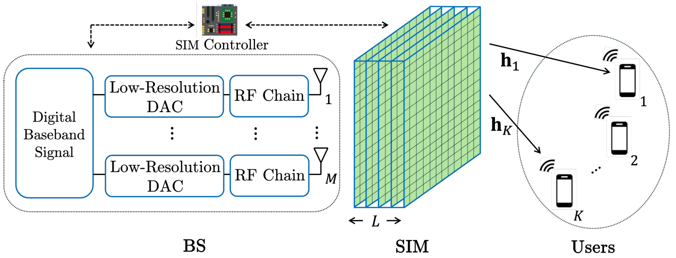

Fig. 1. A downlink multi-user SIM-assisted MISO network.

$$y_k(t) = \mathbf{h}_k^{\mathrm{H}}(t)\boldsymbol{\Psi}(t)\sum_{k'\in\mathcal{K}} \mathbf{w}_k^{[1]}\sqrt{p_{k'}(t)}s_{k'}(t) + n_k(t), \tag{1}$$

where $\mathbf{h}^{\mathrm{H}}(t) \in \mathbb{C}^{1\times N}$ is the channel between Rx_k and the elements in the last layer of the SIM at time t, and $\boldsymbol{\Psi}(t) \in \mathbb{C}^{N\times N}$ denotes the SIM configuration matrix at time t. Moreover, $\mathbf{w}_k^{[1]} \in \mathbb{C}^{N\times 1}$ represents the transmission vector from the k^{th} antenna element to the first layer of the SIM. The n^{th} element of $\mathbf{w}_k^{[1]}$ follows the Rayleigh-Sommerfeld diffraction theory and is equal to [12]:

$$\left[\mathbf{w}_k^{[1]}\right]_n = \frac{d_x d_y \cos\xi_{n,k}^{[1]}}{d_{n,k}^{[1]}}\left(\frac{1}{2\pi d_{n,k}^{[1]}} - j\frac{1}{\lambda}\right)e^{j\left(\frac{2\pi}{\lambda}\right)d_{n,k}^{[1]}}, \tag{2}$$

where d_x and d_y denote the meta atom's dimensions, $\xi_{n,k}^{[1]}$ is the angle between the propagation direction and the normal direction of the first SIM layer, $d_{n,k}^{[1]}$ is the distance between the k^{th} antenna element and the n^{th} meta atom in the first SIM layer, and λ represents the wavelength. Furthermore, in (1), $p_{k'}(t)$ is the transmit power allocated to $\mathsf{Rx}_{k'}, k' \in \mathcal{K}$, at time t. Past studies usually assumed that the transmit power $p_{k'}(t)$ takes continuous values, satisfying $p_{k'}(t) \geq 0$ and $\sum_{k'\in\mathcal{K}} p_{k'}(t) \leq P_T$, where P_T is the total power budget [8]. In contrast, in this paper, we consider the practical limitations of circuit design, which make it challenging to generate continuous values of $p_{k'}(t)$. We assume $p_{k'}(t) \in \mathcal{P}_U$, where $\mathcal{P}_U$ is a discrete set with U possible choices as below:

$$\mathcal{P}_U = \left\{\frac{P_T}{U}, \frac{2P_T}{U}, \ldots, P_T\right\}. \tag{3}$$

Moreover, $s_k(t)$ is the information symbol intended to Rx_k, which is an independent and identically distributed (i.i.d.) random variable with zero mean and unit variance, and $n_k(t)$ is a zero-mean additive white Gaussian (AWGN) noise with variance σ_k^2 at Rx_k at time t.

We assume that all operations take place within a single coherence time. Consequently, we will omit the time notation for the remainder of this paper.

SIM Configuration Matrix: Here, the SIM consists of multiple layers arranged sequentially. Thus, characterizing the SIM configuration matrix $\mathbf{\Psi}$ necessitates identifying two parts: the SIM configuration matrix within each layer and the transmission matrix from one layer to the next.

To achieve the first part of $\mathbf{\Psi}$, we use PSs as SIM elements and represent the n^{th} PS in the ℓ^{th} layer as $\varrho_n^{[\ell]} e^{j\theta_n^{[\ell]}}$, where $n \in \mathcal{N}$ and $\ell \in \mathcal{L}$, with $\varrho \in [0,1]$ and $\theta_n^{[\ell]} \in [0, 2\pi)$ denoting the corresponding amplitude and phase, respectively. For simplicity, we assume that $\varrho_n^{[\ell]} = 1$. Unlike prior works [8], [9], where $\theta_n^{[\ell]}$ could take any continuous value in $[0, 2\pi)$, resulting in higher cost due to infinite-bit resolution PSs, in this paper, we assume that each SIM element has a b-bit PS with 2^b quantized levels. These quantized levels are denoted by set $\mathcal{Q}_b$, which is given by:

$$\mathcal{Q}_b = \left\{0, \frac{2\pi}{2^b}, \frac{4\pi}{2^b}, \ldots, \frac{2\pi(2^b-1)}{2^b}\right\}. \tag{4}$$

Then, the SIM configuration matrix in the ℓ^{th} layer, $\mathbf{\Theta}^{[\ell]}$, is written as:

$$\mathbf{\Theta}^{[\ell]} = \operatorname{diag}\left(e^{j\boldsymbol{\theta}^{[\ell]}}\right), \quad \boldsymbol{\theta}^{[\ell]} = \left[\theta_1^{[\ell]}, \theta_2^{[\ell]}, \ldots, \theta_N^{[\ell]}\right]. \tag{5}$$

To characterize the second part of $\mathbf{\Psi}$, we shift our focus to the transmission matrix from the $(\ell-1)^{\text{th}}$ layer to the ℓ^{th} layer of $\mathbf{\Psi}$, where $2 \leq \ell \leq L$, denoted as $\mathbf{W}^{[\ell]}$. Similar to (2), $\mathbf{W}^{[\ell]}$ also adheres to Rayleigh-Sommerfeld diffraction theory, and the element in the n^{th} row and n'^{th} column of $\mathbf{W}^{[\ell]}$ is expressed as:

$$\left[\mathbf{W}^{[\ell]}\right]_{n,n'} = \frac{d_x d_y \cos \xi_{n,n'}^{[\ell]}}{d_{n,n'}^{[\ell]}} \Big(\frac{1}{2\pi d_{n,n'}^{[\ell]}} - j\frac{1}{\lambda}\Big) e^{j\frac{2\pi}{\lambda} d_{n,n'}^{[\ell]}}, \tag{6}$$

where $n, n' \in \mathcal{N}$, $\xi_{n,n'}^{[\ell]}$ is the angle between the propagation direction and the normal direction of the ℓ^{th} SIM layer, and $d_{n,n'}^{[\ell]}$ denotes the distance between the n^{th} meta-atom in the $(\ell-1)^{\text{th}}$ layer and the n'^{th} meta-atom in the ℓ^{th} layer.

As a result, the SIM configuration matrix $\mathbf{\Psi}$ is obtained as:

$$\mathbf{\Psi} = \mathbf{\Theta}^{[L]} \mathbf{W}^{[L]} \mathbf{\Theta}^{[L-1]} \ldots \mathbf{\Theta}^{[2]} \mathbf{W}^{[2]} \mathbf{\Theta}^{[1]}. \tag{7}$$

Channel Model: We use $\mathbf{h}_k = \sqrt{\beta_k}\mathbf{h}_0$ to denote the channel between the BS and Rx_k, where β_k represents the large-scale fading between the BS and Rx_k and and $\mathbf{h}_0$ is the small-scale fading. In particular, $\beta_k = C_0 d_k^{-\alpha}$, where C_0 represents the signal loss at a reference distance, d_k is the distance between Rx_k and the BS, and α is the path loss exponent. Further, we assume that $\mathbf{h}_0$ follows the i.i.d. Rayleigh fading model, with the channel state information (CSI) being constant within a given coherence time and globally available.

According to the above, we compute the signal-to-interference-plus-noise ratio (SINR) at Rx_k as below:

$$\text{SINR}_k = \frac{p_k \left|\mathbf{h}_k^{\text{H}} \mathbf{\Psi} \mathbf{w}_k^{[1]}\right|^2}{\sum\limits_{k' \in \mathcal{K}, k' \neq k} p_{k'} \left|\mathbf{h}_k^{\text{H}} \mathbf{\Psi} \mathbf{w}_{k'}^{[1]}\right|^2 + \sigma_k^2}. \tag{8}$$

B. Problem Formulation

As mentioned earlier, our goal is to jointly optimize transmit powers and SIM elements to maximize sum-rate. We define $\mathbf{p} = [p_1, p_2, \ldots, p_K]$ and $\boldsymbol{\theta} = [\boldsymbol{\theta}^{[1]}, \boldsymbol{\theta}^{[2]}, \ldots, \boldsymbol{\theta}^{[L]}]$ and then formulate our optimization problem, which is given by:

$$\begin{aligned} \text{P1}: \ \max_{\mathbf{p},\boldsymbol{\theta}} \ & \sum_{k \in \mathcal{K}} \log_2(1 + \text{SINR}_k) && (9)\\ \text{s.t.} \ & \sum_{k \in \mathcal{K}} p_k \leq P_T, && (9.\text{a})\\ & p_k \in \mathcal{P}_U, \ k \in \mathcal{K}, && (9.\text{b})\\ & \theta_n^{\ell} \in \mathcal{Q}_b, \ n \in \mathcal{N}, \ \ell \in \mathcal{L}. && (9.\text{c}) \end{aligned}$$

To tackle P1, we reformulate it as a minimization problem:

$$\begin{aligned} \text{P2}: \ \min_{\mathbf{p},\boldsymbol{\theta}} \ & f(\mathbf{p}, \boldsymbol{\theta}) && (10)\\ \text{s.t.} \ & (9.\text{a}), (9.\text{b}), (9.\text{c}), \end{aligned}$$

where $f(\mathbf{p}, \boldsymbol{\theta}) \stackrel{\triangle}{=} -\sum_{k \in \mathcal{K}} \log_2(1 + \text{SINR}_k)$. As demonstrated in (10), P2 is a non-convex NILP problem. The exhaustive search method is the only approach that guarantees finding the global optimal solution, albeit with prohibitive computational complexity. To address this, we propose using the AO technique. We start by treating $\mathbf{p}$ as constant and optimizing $\boldsymbol{\theta}$ to find $\boldsymbol{\theta}^*$. Next, we fix $\boldsymbol{\theta}$ at $\boldsymbol{\theta}^*$ and focus on optimizing the power vector $\mathbf{p}$. We alternate between these optimization steps until convergence is achieved, either through a predetermined number of iterations, i.e., I_{AO}, or when no further improvement is observed.

The next two sections provide a detailed explanation of how we achieve the optimal solutions for $\boldsymbol{\theta}$ and $\mathbf{p}$, respectively.

III. SIM Element Optimization

In this section, we aim to optimize $\boldsymbol{\theta}$, while $\mathbf{p}$ is set to be constant (i.e., $\mathbf{p} = \mathbf{p}_0$), which converts P2 to the following optimization problem:

$$\begin{aligned} \text{P3}: \ \min_{\boldsymbol{\theta}} \ & f(\mathbf{p}_0, \boldsymbol{\theta}) && (11)\\ \text{s.t.} \ & (9.\text{c}). \end{aligned}$$

As observed, P3 is also a non-convex NILP problem. In the literature, two main approaches are commonly employed to address P3: (i) relaxing the integer parameters and applying state-of-the-art methods, such as gradient ascent, to optimize $\boldsymbol{\theta}$, followed by mapping the resulting solution to a discrete set. This methodology is used in the MPGA method. However, this approach inherently compromises performance, as it does not directly target the NILP problem [9]. (ii) Employing heuristic techniques, e.g., the SR method, which directly tackles the NILP problem in P3 and provides a low-complexity search algorithm to find the optimal solution [13].

In this work, we adopt the second approach by introducing the FF method, which specifically leverages the Sigmoid FF to shift from a local optimum solution (i.e., local minimizer) to a better one. Our proposed FF method contains two stages: local search and global search, where the former starts from an initial solution, say $\boldsymbol{\theta}_0$, and finds the corresponding local minimizer, $\boldsymbol{\theta}_0^*$, while the latter aims to move from the local minimizer $\boldsymbol{\theta}_0^*$ to another local minimizer $\boldsymbol{\theta}_1^*$ such that $f(\mathbf{p}_0, \boldsymbol{\theta}_1^*) \leq f(\mathbf{p}_0, \boldsymbol{\theta}_0^*)$.

Stage I - Local Search. The goal of this stage is to obtain a local minimizer. To do so, we define $\bar{\mathcal{N}}(\boldsymbol{\theta}_0)$ as the set of the neighbors of $\boldsymbol{\theta}_0$ and then search among the neighbors to find the best solution. Here, $\bar{\mathcal{N}}(\boldsymbol{\theta}_0)$ is given by:

$$\bar{\mathcal{N}}(\boldsymbol{\theta}_0) = \boldsymbol{\theta}_0 \cup \{\boldsymbol{\theta}_0 + \boldsymbol{\lambda}_e, \boldsymbol{\lambda}_e \in \boldsymbol{\Lambda}\}, \tag{12}$$

where $\boldsymbol{\lambda}_e$ is a vector with length $E \overset{\triangle}{=} L \times N$, and the e^{th} element of $\boldsymbol{\lambda}_e, 1 \leq e \leq E$ is chosen from $\mathcal{Q}_b \backslash \{0\}$ and the others are set to zero. Moreover, $\boldsymbol{\Lambda}$ is the direction set equals to $\boldsymbol{\Lambda} = \{\boldsymbol{\lambda}_e, e = 1, 2, \ldots, E\}$. Next, we have:

$$\boldsymbol{\theta}_0^* = \underset{\boldsymbol{\theta} \in \bar{\mathcal{N}}(\boldsymbol{\theta}_0)}{\arg\min}\ f(\mathbf{p}_0, \boldsymbol{\theta}). \tag{13}$$

We denote the number of iterations needed to find the local minimizer as i^{loc}, and define $i^{\text{loc}}_{\max}$ as its maximum value. If $\boldsymbol{\theta}_0^* \neq \boldsymbol{\theta}_0$ and $i^{\text{loc}} \leq i^{\text{loc}}_{\max}$, we set $\boldsymbol{\theta}_0 \leftarrow \boldsymbol{\theta}_0^*$ and repeat the procedure by following (12) and (13). Otherwise, we call $\boldsymbol{\theta}_0^*$ as the local minimizer of $f(\mathbf{p}_0, \boldsymbol{\theta})$. Algorithm 1 explains how the local search works.

Algorithm 1: Local Search Algorithm

```
1  Input: $\boldsymbol{\theta}_0$, $i^{\text{loc}}_{\max}$;
2  Output: $\boldsymbol{\theta}_0^*$;
3  $i^{\text{loc}} = 0$; $\boldsymbol{\theta}_0^* = \boldsymbol{\theta}_0$;
4  for all $\boldsymbol{\lambda}_e \in \boldsymbol{\Lambda}$ do
5      $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \boldsymbol{\lambda}_e$;
6      if $f(\mathbf{p}_0, \hat{\boldsymbol{\theta}}) < f(\mathbf{p}_0, \boldsymbol{\theta}^*)$ then
7          $\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}$;
8  $i^{\text{loc}} \leftarrow i^{\text{loc}} + 1$;
9  if $\boldsymbol{\theta}_0^* \neq \boldsymbol{\theta}_0$ and $i^{\text{loc}} \leq i^{\text{loc}}_{\max}$ then
10     $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_0^*$;
11     Go to line 4;
12 else
13     $\boldsymbol{\theta}_0^*$ is the local minimizer.
```

Notice that local search is equivalent to the SR method [9], [13] when $i^{\text{loc}}_{\max} \rightarrow \infty$. Although it provides a low-complexity approach, it only converges to the local minimizer $\boldsymbol{\theta}_0^*$, which heavily relies on the initial choice of $\boldsymbol{\theta}_0$. This leads to a question: is it possible to move from the current local minimizer $\boldsymbol{\theta}_0^*$ to a better solution? The answer is yes. To achieve this, the FF method applies the global search as follows.

Stage II - Global Search. In this part, we mainly aim to shift from one local minimizer to another superior local minimizer. To this end, the global search begins from an initial random solution $\boldsymbol{\theta}_0$ and applies Algorithm 1 to locate $\boldsymbol{\theta}_0^*$. Next, the filled function $F(\boldsymbol{\theta}, \boldsymbol{\theta}_0^*)$ comes into play, facilitating the shift from the current local minimizer $\boldsymbol{\theta}_0^*$ to an improved solution. In this work, we define $F(\boldsymbol{\theta}, \boldsymbol{\theta}_0^*)$ using the Sigmoid function as follows:

$$F_r(\boldsymbol{\theta}, \boldsymbol{\theta}_0^*) = \left(\frac{2 + \gamma \|\boldsymbol{\theta} - \boldsymbol{\theta}_0^*\|^2}{1 + \gamma \|\boldsymbol{\theta} - \boldsymbol{\theta}_0^*\|^2}\right) q_r\left(f\left(\mathbf{p}_0, \boldsymbol{\theta}\right) - f\left(\mathbf{p}_0, \boldsymbol{\theta}_0^*\right)\right), \tag{14}$$

where r is the heuristic optimization parameter, and

$$q_r(t) = \begin{cases} t + r, & t \leq -r, \\ \frac{1}{1 + e^{\frac{-6}{r}(t + r/2)}}, & -r < t < 0, \\ 1, & t \geq 0, \end{cases} \tag{15}$$

with

$$\gamma = \begin{cases} 0, & f\left(\mathbf{p}_0, \boldsymbol{\theta}\right) - f\left(\mathbf{p}_0, \boldsymbol{\theta}_0^*\right) \leq -r, \\ 1, & \text{otherwise}. \end{cases} \tag{16}$$

In [11], the authors showed that a valid filled function must meet three necessary conditions to ensure the shift from one local minimizer to a better one, and [14] proved that $F_r(\boldsymbol{\theta}, \boldsymbol{\theta}_0^*)$ in (14) fulfills these conditions. Next, we consider $F(\boldsymbol{\theta}, \boldsymbol{\theta}_0^*)$ as the objective function of a new optimization problem, called auxiliary optimization problem, which is represented by:

$$\begin{aligned} \text{P4}: \ \min_{\boldsymbol{\theta}} \ & F(\boldsymbol{\theta}, \boldsymbol{\theta}_0^*) \\ \text{s.t.} \quad & \boldsymbol{\theta}, \boldsymbol{\theta}_0^* \in \mathcal{Q}_b. \end{aligned} \tag{17}$$

To optimize P4, we follow Algorithm 1 assuming $\boldsymbol{\theta}_0^*$ and $F(\boldsymbol{\theta}, \boldsymbol{\theta}_0^*)$ as the input and objective function, respectively. This results in a new solution as $\boldsymbol{\theta}_1 = \bar{\boldsymbol{\theta}}_0$. Consequently, we run the local search algorithm for $\boldsymbol{\theta}_1$ using $f(\mathbf{p}_0, \boldsymbol{\theta})$ to attain $\boldsymbol{\theta}_1^*$. The filled function guarantees that $f(\mathbf{p}_0, \boldsymbol{\theta}_1^*) \leq f(\mathbf{p}_0, \boldsymbol{\theta}_0^*)$.

We provide Algorithm 2 to describe the global search, where $\boldsymbol{\theta}_0, i^{\text{loc}}_{\max}$, $i^{\text{filled}}_{\max}$, r, and ϵ are the inputs and $\boldsymbol{\theta}^{**}$ is the output. We use $i^{\text{filled}}_e, e = 0, 1, \ldots, E$ to denote the number of times our method searches for a new solution by scanning the e^{th} neighbor of a local minimizer obtained from P4. Then, we define $i^{\text{filled}}_{\max}$ as the maximum number of times the global search uses the filled function to search for a new solution. We also use ϵ as the stopping condition. More precisely, we stop the procedure if there is no further improvement in the process (i.e., $r < \epsilon$) or $\sum_{e=0}^{E} i^{\text{filled}}_e > i^{\text{filled}}_{\max}$.

IV. Transmit Power Optimization

Following the AO technique, once P4 is optimized, we treat $\boldsymbol{\theta}$ as constant and set it to its optimal value, $\boldsymbol{\theta}^*$, and then focus on optimizing the transmit power $\mathbf{p}$ in the subsequent problem.

$$\begin{aligned} \text{P5}: \min_{\mathbf{p}} \ & f(\mathbf{p}, \boldsymbol{\theta}^*) \\ \text{s.t.} \quad & \text{(9.a), (9.b)}, \end{aligned} \tag{18}$$

which is non-convex because of discrete transmit powers. In [11], the author showed that the FF-based optimization methods can handle any general non-convex NILP problem when the only constraint is selecting from a discrete set. However, the problem P5 includes an additional constraint specified in (9.a) (i.e., $\sum_{k \in \mathcal{K}} p_k \leq P_T$), alongside the discrete transmit power constraint in (9.b). Therefore, in this section, we modify the FF method to obtain the mFF method by devising a strategy to ensure that for any possible solution of $p_k, k \in \mathcal{K}$, the (9.a) constraint holds. To achieve this, we divide $\mathcal{K}$ into two sets $\mathcal{I}_{\mathcal{K}}$ and $\mathcal{D}_{\mathcal{K}}$ where each set has length $\frac{K}{2}$. We call $\mathcal{I}_{\mathcal{K}}$ and $\mathcal{D}_{\mathcal{K}}$ independent set and dependent set, respectively. We assign each user in $\mathcal{I}_{\mathcal{K}}$ (say $i \in \mathcal{I}_{\mathcal{K}}$) to only

Algorithm 2: Global Search Algorithm

```
1  Input: $\boldsymbol{\theta}_0, i_{\max}^{\text{loc}}, i_{\max}^{\text{filled}}, r, \epsilon$;
2  Output: $\boldsymbol{\theta}^{**}$;
3  $\ell = 0$; $r_0 = r$, $i_e^{\text{filled}} = 0$, for $e = 0, 1, \ldots, E$;
4  $\boldsymbol{\theta}^{**} = \boldsymbol{\theta}_0$;
5  Run Algorithm 1 with $\boldsymbol{\theta}_\ell$ and (11) to find $\boldsymbol{\theta}_\ell^*$;
6  if $f(\mathbf{p}_0, \boldsymbol{\theta}_\ell^*) < f(\mathbf{p}_0, \boldsymbol{\theta}^{**})$ then
7      $\boldsymbol{\theta}^{**} = \boldsymbol{\theta}_\ell^*$; $r = r_0$;
8      $e = 1$;
9      Run Algorithm 1 with $\boldsymbol{\theta}_\ell^*$ and (17) to get $\bar{\boldsymbol{\theta}}_\ell$;
10     $i_{e-1}^{\text{filled}} \leftarrow i_{e-1}^{\text{filled}} + 1$;
11     if $e = 1$ then
12         $\ell \leftarrow \ell + 1$;
13         $\boldsymbol{\theta}_\ell = \bar{\boldsymbol{\theta}}_{\ell-1}$;
14     else
15         $\boldsymbol{\theta}_\ell = \bar{\boldsymbol{\theta}}_\ell$;
16     Go to line 5;
17 if $e \leq E$ then
18     $\boldsymbol{\theta}_\ell^* = \boldsymbol{\theta}_{\ell-1}^* + \boldsymbol{\lambda}_e$;
19     $e \leftarrow e + 1$; Go to line 9;
20 else
21     if $r < \epsilon$ or $\sum_{e=0}^{E} i_e^{\text{filled}} > i_{\max}^{\text{filled}}$ then
22         $\boldsymbol{\theta}^{**}$ is the global minimizer.
23     else
24         $r \leftarrow \frac{r}{10}$; $\ell \leftarrow \ell - 1$;
25         Go to line 8;
```

one user in $\mathcal{D}_\mathcal{K}$ (say $j \in \mathcal{D}_\mathcal{K}$) and define (i, j) as a user pair. Then, we define $\mathbf{p}_{\mathcal{I}_\mathcal{K}} = [p_i | i \in \mathcal{I}_\mathcal{K}]$ and $\mathbf{p}_{\mathcal{D}_\mathcal{K}} = [p_j | j \in \mathcal{D}_\mathcal{K}]$, and rewrite P5 as follows:

$$\text{P6} : \min_{\mathbf{p}_{\mathcal{I}_\mathcal{K}}} \quad f(\mathbf{p}_{\mathcal{I}_\mathcal{K}}, \boldsymbol{\theta}^*) \tag{19}$$

$$\text{s.t.} \quad p_i + p_j = \frac{P_T}{K/2}, \quad (i, j) \text{ is a user pair,} \tag{19.a}$$

$$p_i, p_j \in \bar{\mathcal{P}}_U, \qquad i \in \mathcal{I}_\mathcal{K}, \ j \in \mathcal{D}_\mathcal{K}, \tag{19.b}$$

where $\bar{\mathcal{P}}_U = \left\{ \frac{P_T}{U}, \frac{2P_T}{U}, \ldots, \frac{(\frac{U}{K/2} - 1)P_T}{U} \right\}$. Since there are $\frac{K}{2}$ user pairs in $\mathcal{K}$, (19.a) and (19.b) guarantee that any choice of $\mathbf{p}_{\mathcal{I}_\mathcal{K}}$ satisfies the (9.a) constraint. Thus, we apply the FF method to P6 to optimize $\mathbf{p}_{\mathcal{I}_\mathcal{K}}$ and compute $\mathbf{p}_{\mathcal{D}_\mathcal{K}}$ accordingly.

Here, there are $I_{\text{mFF}} = \frac{K!}{(\frac{K}{2})!(\frac{K}{2})!}$ possible distinct options for $\mathcal{I}_\mathcal{K}$. Therefore, the mFF method optimizes P6 for all choices of $\mathcal{I}_\mathcal{K}$ and picks the best one as the optimal solution. Similar to the FF method, the mFF method incorporates local and global search algorithms. We define $\bar{\mathcal{N}}(\mathbf{p}_{\mathcal{I}_{\mathcal{K},0}})$, the set of the neighbors of power vector $\mathbf{p}_{\mathcal{I}_{\mathcal{K},0}}$, as below:

$$\bar{\mathcal{N}}(\mathbf{p}_{\mathcal{I}_{\mathcal{K},0}}) = \mathbf{p}_{\mathcal{I}_{\mathcal{K},0}} \cup \left\{ \mathbf{p}_{\mathcal{I}_{\mathcal{K},0}} + \boldsymbol{\lambda}_m, \boldsymbol{\lambda}_m \in \boldsymbol{\Lambda}_m \right\}, \tag{20}$$

where $\boldsymbol{\lambda}_m$ is a vector with length $K/2$, such that the m^{th} element of $\boldsymbol{\lambda}_m, 1 \leq m \leq K/2$, is chosen from $\bar{\mathcal{P}}_U \backslash \{0\}$ and the others are set to zero. Also, $\boldsymbol{\Lambda}_m$ denotes the direction set equals to $\boldsymbol{\Lambda}_m = \{\boldsymbol{\lambda}_m, m = 1, 2, \ldots, K/2\}$. We use $\mathbf{p}_{\mathcal{I}_{\mathcal{K},\ell}}$ to represent the solution in the ℓ^{th} iteration.

The local and global search algorithms of the mFF method are analogous to Algorithms 1 and 2 in Section III. Therefore, due to space limitations, we omit the detailed descriptions of the mFF method's local and global search algorithms.

V. Numerical Analysis

In this section, we present the simulation results to evaluate the performance of our proposed SIM-assisted network in terms of sum-rate and computational complexity, which incorporates discrete transmit power levels and discrete PSs. For our simulations, the BS is positioned at $(0, 0)$, and users are randomly distributed within an area of 100m^2. We set $K = 4$, $\sigma_k^2 = -80\text{dBm}$, $b = 2\text{bit}$, $\alpha = 2$, $C_0 = -30\text{dB}$ at reference distance of 1m, $\lambda = 0.125\text{m}$, $r = 10$, $\epsilon = 0.01$, $i_{\max}^{\text{loc}} = NL$, $i_{\max}^{\text{filled}} = 100$, and $I_{\text{AO}} = 5$. We conduct 50 simulation trials and report the average results.

Sum-rate: In Fig. 2, Fig. 3, and Fig. 4, we compare the performance of our proposed optimization method with five benchmark schemes in terms of sum-rate. The figures present the sum-rate against varying values of N, L, and P_T, respectively. Each figure isolates the effect of one parameter by keeping the other two constant. Specifically, in Fig. 2, N varies while $L = 2$ and $P_T = 26\text{dBm}$. In Fig. 3, L changes and N is fixed at 25 and P_T at 26dBm. Further, Fig. 4 shows the sum-rate performance when P_T varies and $L = 2$ and $N = 25$. Here, the optimization methods are represented as a tuple $(\cdot, \cdot)$, where the first entry denotes the method used for transmit power optimization, and the second entry specifies the method for optimizing SIM elements. As expected, the figures indicate that (WF, PGA) delivers the best results due to access to continuous-valued powers and PSs. The (mFF, PGA) method ranks second, showing only a 6% performance drop compared to (WF, PGA) when $N = 49$, $L = 2$, and $P_T = 26\text{dBm}$. Under the same setting, the (WF, FF) approach follows with an 11% decrease in performance relative to (WF, PGA), but it surpasses both the (WF, MPGA) and (WF, SR) methods [9] by employing a global search to shift from one local optimum to a superior one. Lastly, the (mFF, FF) method achieves 77% of the performance of (WF, PGA), which is attributed to its reliance on discrete transmit powers and PSs. Although (mFF, FF) utilizes discrete transmit power, it outperforms the (WF, MPGA) and (WF, SR) methods due to the superior performance of the FF method compared to MPGA and SR.

Computational Complexity: Table I presents the computational complexity of our proposed method alongside other benchmarks, where I_{WF} and I_{GA} denote the number of iterations in the WF and PGA methods, respectively. Here, (WF, PGA) exhibits the lowest complexity as it focuses solely on continuous-valued sets. On the other hand, (WF, FF) involves greater complexity than (WF, SR) and (WF, MPGA), highlighting the trade-off between complexity and sum-rate performance in our FF optimization method. Moreover, Table I shows that (mFF, FF) incurs the highest complexity due to the involvement of discrete transmit powers and PSs.

VI. Conclusion

We studied HMIMO communications by examining a downlink scenario in a SIM-assisted MISO network, where the BS generates discrete transmit powers and discrete PSs are

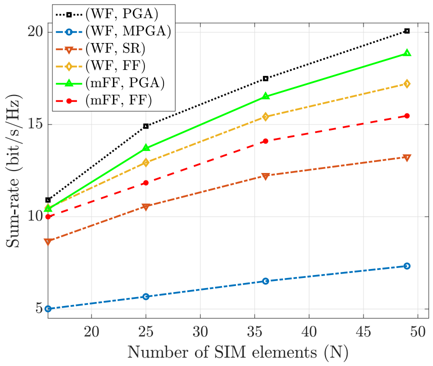

Fig. 2. $L = 2$, and $P_T = 26$ dBm.

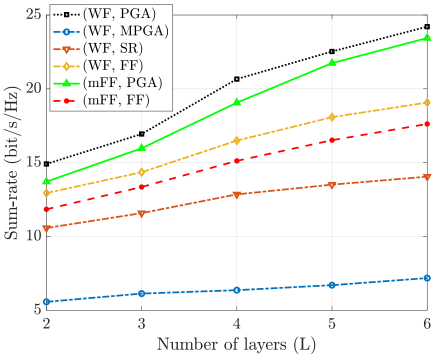

Fig. 3. $N = 25$ and $P_T = 26$ dBm.

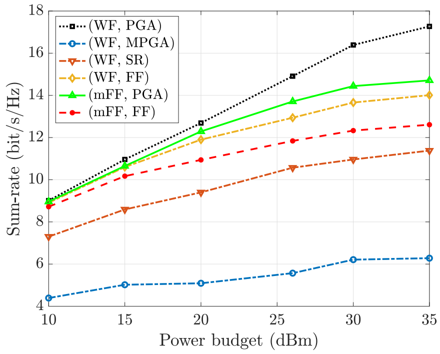

Fig. 4. $L = 2$, and $N = 25$.

TABLE I
COMPUTATIONAL COMPLEXITY

Method	**Complexity**
(WF, PGA)	$\mathcal{O}\left(I_{\mathrm{AO}} K^2 (4N+3) \left[I_{\mathrm{WF}} + 2 I_{\mathrm{GA}} LN\right]\right)$ [9]
(WF, MPGA)	$\mathcal{O}\left(I_{\mathrm{AO}} K^2 (4N+3) \left[I_{\mathrm{WF}} + 2 I_{\mathrm{GA}} (LN)^2\right]\right)$ [9]
(WF, SR)	$\mathcal{O}(I_{\mathrm{AO}} K^2 (4N+3) \left[I_{\mathrm{WF}} + (\lvert\mathcal{Q}_b\rvert - 1)(LN)^2\right])$ [9]
(WF, FF)	$\mathcal{O}\left(I_{\mathrm{AO}} K^2 (4N+3) \left[I_{\mathrm{WF}} + 2(\lvert\mathcal{Q}_b\rvert - 1)(LN)^2 i_{\max}^{\mathrm{Filled}}\right]\right)$ [2]
(mFF, PGA)	$\mathcal{O}\left(I_{\mathrm{AO}} K^2 (4N+3) \left[I_{\mathrm{mFF}} (U/K - 0.5) i_{\max}^{\mathrm{Filled}} + 2 I_{\mathrm{GA}} LN\right]\right)$
(mFF, FF)	$\mathcal{O}\left(I_{\mathrm{AO}} K^2 (4N+3) i_{\max}^{\mathrm{Filled}} \left[I_{\mathrm{mFF}} (U/K - 0.5) + 2(\lvert\mathcal{Q}_b\rvert - 1)(LN)^2\right]\right)$

used as the SIM elements. We introduced an AO technique to iteratively optimize both the transmit power and SIM elements. Specifically, we introduced the mFF method to optimize the transmit powers and the FF method to adjust the SIM elements. We evaluated the performance of our proposed approach against five baselines, demonstrating that it achieves a higher sum-rate compared to two state-of-the-art methods (i.e., the MPGA and SR methods), even when those methods have access to continuous-valued transmit powers. Since the mFF and FF methods are heuristic, a promising direction for future research would be to explore optimization techniques based on non-convex approximations. Additionally, extending the results by integrating SIM with beyond-diagonal RIS could offer another valuable avenue for further investigation.

VII. ACKNOWLEDGMENT

This work was supported by the U.S. National Science Foundation under grants ECCS-2146436, CCF-2225576, and CCF-2322190.

REFERENCES

[1] E. Björnson, Ö. Özdogan, and E. G. Larsson, “Reconfigurable intelligent surfaces: Three myths and two critical questions,” *IEEE Communications Magazine*, vol. 58, no. 12, pp. 90–96, 2020.

[2] S. Nassirpour, A. Vahid, D.-T. Do, and D. Bharadia, “Beamforming design in reconfigurable intelligent surface-assisted IoT networks based on discrete phase shifters and imperfect CSI,” *IEEE Internet of Things Journal*, 2023.

[3] S. Nassirpour, N. Kusashima, J. Flordelis, and A. Vahid, “Mix-and-conquer: Beamforming design with interconnected RIS for multi-user networks,” in *ICC 2024-IEEE International Conference on Communications*. IEEE, 2024, pp. 3725–3730.

[4] Z. Kang, C. You, and R. Zhang, “Active-passive IRS aided wireless communication: New hybrid architecture and elements allocation optimization,” *IEEE Transactions on Wireless Communications*, 2023.

[5] C. Huang, S. Hu, G. C. Alexandropoulos, A. Zappone, C. Yuen, R. Zhang, M. Di Renzo, and M. Debbah, “Holographic MIMO surfaces for 6g wireless networks: Opportunities, challenges, and trends,” *IEEE wireless communications*, vol. 27, no. 5, pp. 118–125, 2020.

[6] T. Gong, P. Gavriilidis, R. Ji, C. Huang, G. C. Alexandropoulos, L. Wei, Z. Zhang, M. Debbah, H. V. Poor, and C. Yuen, “Holographic MIMO communications: Theoretical foundations, enabling technologies, and future directions,” *IEEE Communications Surveys & Tutorials*, 2023.

[7] A. Papazafeiropoulos, J. An, P. Kourtessis, T. Ratnarajah, and S. Chatzinotas, “Achievable rate optimization for stacked intelligent metasurface-assisted holographic MIMO communications,” *IEEE Transactions on Wireless Communications*, 2024.

[8] J. An, M. Di Renzo, M. Debbah, and C. Yuen, “Stacked intelligent metasurfaces for multiuser beamforming in the wave domain,” in *ICC 2023-IEEE International Conference on Communications*. IEEE, 2023, pp. 2834–2839.

[9] J. An, M. Di Renzo, M. Debbah, H. V. Poor, and C. Yuen, “Stacked intelligent metasurfaces for multiuser downlink beamforming in the wave domain,” *arXiv preprint arXiv:2309.02687*, 2023.

[10] R. GE, “A filled function method for finding a global minimizer of a function of several variables,” *Mathematical programming*, vol. 46, no. 1, pp. 191–204, 1990.

[11] C.-K. Ng, L.-S. Zhang, D. Li, and W.-W. Tian, “Discrete filled function method for discrete global optimization,” *Computational Optimization and Applications*, vol. 31, no. 1, pp. 87–115, 2005.

[12] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, “All-optical machine learning using diffractive deep neural networks,” *Science*, vol. 361, no. 6406, pp. 1004–1008, 2018.

[13] Q. Wu and R. Zhang, “Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts,” *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1838–1851, 2019.

[14] S. Nassirpour, A. Gupta, A. Vahid, and D. Bharadia, “Power-efficient analog front-end interference suppression with binary antennas,” *IEEE Transactions on Wireless Communications*, vol. 22, no. 4, pp. 2592–2605, 2022.