Optimization-Based Phase-Shift Codebook Design for Large IRSs

Walid R. Ghanem, Member, IEEE, Vahid Jamali, Member, IEEE, Malte Schellmann, Hanwen Cao, Joseph Eichinger, and Robert Schober, Fellow, IEEE

Abstract—In this letter, we focus on large intelligent reflecting surfaces (IRSs) and propose a new codebook construction method to obtain a set of pre-designed phase-shift configurations for the IRS unit cells. Since the complexity of online optimization and the overhead for channel estimation scale with the size of the phase-shift codebook, the design of small codebooks is of high importance. We consider both continuous and discrete phase-shift designs and formulate the codebook construction as optimization problems. To solve the optimization problems, we propose an optimal algorithm for the discrete phase-shift design and a locally optimal solution for the continuous design. Simulation results show that the proposed algorithms facilitate the construction of codebooks of different sizes and with different beamwidths. Moreover, the performance of the discrete phase-shift design with 2-bit quantization is shown to approach that of the continuous phase-shift design. Finally, our simulation results show that the proposed designs enable large transmit power savings compared to the existing linear and quadratic codebook designs.

Index Terms—Intelligent reflecting surfaces (IRSs), phase-shift codebook, optimization problem.

I. INTRODUCTION

RECENTLY, intelligent reflecting surfaces (IRSs) have been proposed to shape wireless communication channels [1], [2], [3], [4], [5]. IRSs typically comprise large numbers of reconfigurable elements and are deployed between the base station (BS) and the user to establish a virtual line-of-sight (LoS) link. By properly configuring the elements, the IRS can provide a high passive beamforming gain. However, due to the required large number of IRS elements, online optimization and channel estimation in IRS-assisted systems is challenging.

Most existing works [3], [4], [5], [6], [7], [8] aim to optimize all IRS elements individually for real-time communication, which leads to computationally complex resource allocation problems for large numbers of elements [1]. Unfortunately, most algorithms in the literature, e.g., [3], [4], [5], [6], [7], [8], may not be applicable for such large IRSs since their complexity will render the optimization problem intractable. To overcome this issue, the authors of [1] proposed to design a set of phase-shift configurations based on a codebook framework in an offline stage, and then select the best codeword from the set in an online optimization stage for a given channel realization. Hence, the complexity of online optimization and the corresponding channel estimation overhead scale with the IRS codebook size and not explicitly with the number of reflecting elements. In [9] and [10], the authors employed predefined phase-shift configurations based on the discrete Fourier transform (DFT) matrix for channel estimation in IRS-assisted systems. However, the codebooks in [9] and [10] contain the same number of phase-shift configurations as there are IRS elements, while in [1] the codebook has also to be large to attain good performance [2]. To overcome this problem, the authors of [2] proposed a quadratic codebook design, which features a parameter to control the size of the codebook. However, the shape of the generated beams cannot be explicitly controlled. Furthermore, the authors of [1], [9], [10], and [2] focus on continuous IRS phase shifts, whereas in practice, discrete IRS phase shifts may be preferred to reduce the implementation cost [6], [11].

Notations: $\mathbb{C}^N$ represents the set of all $N \times 1$ vectors with complex valued entries. The circularly symmetric complex Gaussian distribution with mean $\mu$ and variance $\sigma^2$ is denoted by $\mathcal{CN}(\mu, \sigma^2)$, $\sim$ stands for “distributed as”, and $\mathbb{E}\{\cdot\}$ denotes statistical expectation. $\nabla_x f(x)$ denotes the gradient vector of function $f(x)$ and its elements are the partial derivatives of $f(x)$. $a \otimes b$ denotes the Kronecker product of two vectors. $A^T$, $A^H$, $\text{Rank}(A)$, and $\text{Tr}(A)$ denote transpose, Hermitian transpose, rank, and trace of matrix $A$, respectively. $\|A\|_2$ denotes the nuclear norm and spectral norm of matrix $A$, respectively. $\lambda_{\text{max}}(A)$ is the eigenvector associated with the maximum eigenvalue of matrix $A$. $x^*$ is the optimal value of optimization variable $x$. Finally, $\text{Diag}(a)$ denotes a diagonal matrix with the elements of vector $a$ on its main diagonal.

II. SYSTEM MODEL

In this section, we present the system and signal models and the codebook concept for IRS-assisted wireless systems.

A. System Model

We consider a downlink system comprising a single-antenna BS that serves a single-antenna user. The downlink communication is assisted by a large IRS comprising $Q$ reflecting unit cells. We consider the case where the IRS is in the far field of the BS and the user. In this letter, we focus on the offline design and assume the direct link between the BS and the user is blocked. However, later, in the online optimization, the non-negligible direct link can be inserted into the end-to-end design. The near field design has to take into consideration the BS and user equipment (UE) locations with respect to the IRS. The optimization-based codebook design for the near field scenario is an interesting case for future studies.

The proposed codebook is directly applicable for the phase shift of each tile. The near field design has to take into consideration the BS and user equipment (UE) locations with respect to the IRS. The optimization-based codebook design for the near field scenario is an interesting case for future studies.

1The near field design has to take into consideration the BS and user equipment (UE) locations with respect to the IRS. The optimization-based codebook design for the near field scenario is an interesting case for future studies.

2To facilitate the support of multiple users, the IRS can be divided into tiles [1]. Then, using the proposed codebook, each tile may serve one user (single beam) and multiple users are served by different tiles (e.g., multiple beams are created by the IRS via different tiles). In this case, the proposed codebook is directly applicable for the phase shift of each tile.
channel as done in our previous studies, see [1], [12] for more details.

Throughout this letter, we assume that the IRS is a planar uniform array comprising \( Q_y \) \((Q_z)\) unit cells spaced \( d_y \) \((d_z)\) meter apart in \( y\)-direction \((z\)-direction\), indexed by \( q_y = 0, \ldots, Q_y - 1 \) \((q_z = 0, \ldots, Q_z - 1)\), where \( Q = Q_y Q_z \) and each unit cell has an area of \( A_{\text{uc}} = d_y d_z \). Moreover, for future reference, we define \( A_y(q_y, q_z) = A_y(q_y) + A_y(q_z) \) and \( A_z(q_y, q_z) = A_z(q_y) + A_z(q_z) \) with \( A_y(q_y) = \sin(\theta) \sin(\phi) \) and \( A_z(q_z) = \cos(\theta) \), where \( q_y = (\theta, \phi) \) and \( q_z = (\theta, \phi) \) are the angle of arrival (AoA) of the incident wave and the angle of departure (AoD) of the reflected wave, respectively, and \( \theta \) and \( \phi \) denote the elevation and azimuth angles, respectively, see [1] for more details. For convenience of presentation, we use either \( q_y \) or \( q_z \) to refer to the phase-shift applied by the \((q_y, q_z)\)-th or equivalently the \(q\)-th IRS element, respectively. Moreover, \( w \in \mathbb{C}^Q \) is a vector collecting variables \( w_{q_y, q_z} = e^{j\nu_{q_y, q_z}}, \forall q_y, q_z \), or equivalently \( w_q = e^{j\nu_q}, \forall q \).

### B. Signal Model

Assuming the signal arrives from AoA \( \Psi_t \), and phase shifts \( w_{q_y, q_z}, \forall q_y, q_z \), are applied by the elements of the IRS, the IRS response function towards AoD \( \Psi_r \) is given as follows [1] and [2]:

\[
g(\Psi_t, \Psi_r) = \sum_{q_y=0}^{Q_y-1} \sum_{q_z=0}^{Q_z-1} e^{jkd_y \beta_y} e^{jkd_z \beta_z} A_y(q_y) A_z(q_z) q_w e^{j\nu_{q_y, q_z}},
\]

where \( k = \frac{2\pi}{\lambda} \) is the wave number, \( \lambda \) is the wavelength, and \( \hat{g} = \frac{4\pi}{kd_y d_z} \) [2]. The IRS response function determines the end-to-end path loss of the IRS-assisted virtual link [1, Lemma 1], \( P_L_{\text{IRS}} = |g(\Psi_t, \Psi_r)|^2 P_{L_{\text{t}}} P_{L_{\text{r}}} \), where \( P_{L_{\text{t}}} \) and \( P_{L_{\text{r}}} \) denote the free-space path loss from the BS to the IRS and from the IRS to the user, respectively.

### C. Codebook Design

In order to maximize the power reflected in the direction where the user is located, each unit cell of the IRS has to apply an appropriate phase-shift such that the waves propagating from all unit cells in the direction of interest add up coherently. Moreover, a number of beams are needed to cover the entire area to be illuminated by the IRS. The set of phase-shift configurations needed to generate these beams is referred to as the codebook and a given phase-shift configuration is a codeword. The goal of this letter is to design suitable codewords, i.e., a codebook, such that a signal arriving from any arbitrary AoA can be reflected towards any desired AoD. A straightforward way to do this is to discretize \( \Psi_t \), into intervals and to design for each interval of \( \Psi_t \), a complete codebook for reflecting the signal towards a desired discretized \( \Psi_r \). However, as can be observed from (1), the response function depends on \( \Psi_t \) and \( \Psi_r \) only via \( A_y(\Psi_t, \Psi_r), \forall \Psi_r \). Thus, it is preferable to discretize \( A_y(\Psi_t, \Psi_r), \forall \Psi_r \), because different \( \Psi_t \), and \( \Psi_r \) can yield the same value \( A_y(\Psi_t, \Psi_r), \forall \Psi_r \).

In other words, the same phase-shift configuration can be used to reflect different AoAs to different AoDs. To facilitate the presentation, the IRS response function is rewritten as follows:

\[
g(\beta_y, \beta_z) = \hat{g} w^H (y(\beta_y) \otimes z(\beta_z)), \tag{2}
\]

where \( y(\beta_y) = [1, \ldots, e^{jkd_y q_y (Q_y - 1)}], z(\beta_z) = [1, \ldots, e^{jkd_z q_z (Q_z - 1)}], \beta_y \triangleq A_y(\Psi_t, \Psi_r), \) and \( \beta_z \triangleq A_z(\Psi_t, \Psi_r) \).

To design a codebook, we discretize \( \beta_i \in [-\beta_i/2, \beta_i/2], \forall \Psi_r \), into \( M_y, \forall \Psi_r \), intervals, respectively. Note that, the values of \( \beta_i, \forall \Psi_r \), are between -2 and 2. Thus, the entire range of \( \beta_i \) is \( \beta_i = 4 \). However, due to the periodicity of the exponential function \( e^{jkd_y q_z} \), it may not be necessary to consider the entire range of 4, but we need to consider only \( k d_y \beta_i \leq 2\pi \), and thus \( \beta_i \leq \lambda / d_y \). In summary, \( \beta_i \leq \min\{4, \lambda / d_y\} \).

The codebook contains \( M_y \) \( M_z \) codewords and each codeword is characterized by given continuous intervals in \( y\)- and \( z\)-direction, i.e., \( B_{y,m_y}, \forall m_y = \{1, \ldots, M_y\} \), and \( B_{z,m_z}, \forall m_z = \{1, \ldots, M_z\} \). In this letter, the collection of intervals \( B_{y,m_y} \) and \( B_{z,m_z} \), are obtained by uniform quantization of \( [-\beta_i/2, \beta_i/2], \forall \Psi_r \), into \( \Psi_r \), respectively.

### III. CONTINUOUS PHASE-SHIFT DESIGN

In this section, we design the codebook assuming the IRS phase shifts are not discretized, i.e., \( \nu_{q_y, q_z} \in [-\pi, \pi] \).

#### A. Problem Formulation

Mathematically, for continuous phase shifts, the optimization problem for designing a codeword \( w_{m_y, m_z} \) for intervals \( B_{y,m_y}, \forall m_y \in \{1, \ldots, M_y\} \) and \( B_{z,m_z}, \forall m_z \in \{1, \ldots, M_z\} \), is formulated as follows:

maximize \( \alpha \)

s.t. C1 : \( |g_{m_y, m_z}(\beta_y, \beta_z)|^2 \geq \alpha, \forall \beta_y \in B_{y,m_y}, \beta_z \in B_{z,m_z}, \forall \Psi_r \)

C2 : \( |w_{q_y, q_z}| = 1, \forall q_y = \{0, \ldots, Q_y - 1\}, q_z = \{0, \ldots, Q_z - 1\} \),

where \( g_{m_y, m_z}(\beta_y, \beta_z) = w^H_{m_y, m_z}(y(\beta_y) \otimes z(\beta_z)) \)

\( w_{m_y, m_z} \in \mathbb{C}^Q \) is the phase-shift vector containing \( w_{q_y, q_z}, \forall q_y, q_z \). Constraint C2 is the unit-modulus constraint for the IRS elements to ensure that each element has unit magnitude. Moreover, to facilitate the optimization, we discretize the continuous intervals \( B_{y,m_y} \) and \( B_{z,m_z} \) into \( P_y \) and \( P_z \) discrete points, respectively, and collect these points in the two discrete sets \( B_{y,m_y} \) and \( B_{z,m_z} \), respectively, where \( P_y \) and \( P_z \) are the cardinality of discrete sets \( B_{y,m_y} \) and \( B_{z,m_z} \). We maximize the reflected power for the \( \beta_i, \forall \Psi_r \), belonging to discrete sets \( B_{y,m_y} \) and \( B_{z,m_z} \), by maximizing \( \alpha \). To obtain the complete codebook of \( M \) codewords, optimization problem (3) has to be solved for all sets \( B_{y,m_y}, \forall m_y = \{1, \ldots, M_y\} \), and \( B_{z,m_z}, \forall m_z = \{1, \ldots, M_z\} \).

There is no systematic approach for solving general non-convex optimization problems. However, in the following, we show that problem (3) can be reformulated as a difference of convex (D.C.) functions programming problem. This reformulation allows the application of a Taylor series approximation to obtain a locally optimal solution of (3).
B. Proposed Solution

In the following, we transform (3) into an equivalent semi-definite programming (SDP) problem. Let us define $W_{m_y,m_z} = w_{m_y,m_z} W_{m_y,m_z}^H$ and exploit the following identity [8]:

$$|w_{m_y,m_z}^H(y(\beta_y) \otimes z(\beta_z))|^2 = \operatorname{Tr}(W_{m_y,m_z}^H(y(\beta_y) \otimes z(\beta_z))(y(\beta_y) \otimes z(\beta_z))^H).$$

(4)

Based on (4), optimization problem (3) can be rewritten in the following equivalent SDP form:

$$\begin{align*}
\text{maximize} & \quad \alpha \\
\text{s.t.} & \quad C1: G_{m_y,m_z}(\beta_y, \beta_z) \geq \alpha, \forall \beta_y \in \mathcal{B}_{y,m_y}, \beta_z \in \mathcal{B}_{z,m_z}, \\
& \quad C2: \operatorname{Rank}(W_{m_y,m_z}) = 1, C3: W_{m_y,m_z} \succeq 0, \\
& \quad C4: \operatorname{Diag}(W_{m_y,m_z}) = 1, \\
\end{align*}$$

(5)

where $G_{m_y,m_z}(\beta_y, \beta_z) = \operatorname{Tr}(W_{m_y,m_z}^H(y(\beta_y) \otimes z(\beta_z))(y(\beta_y) \otimes z(\beta_z))^H)$ and constraints C2, C3, and C4 are imposed to guarantee that $W_{m_y,m_z} = w_{m_y,m_z} W_{m_y,m_z}^H$ and $|w_{y,m_z}| = 1$ hold after optimization, respectively.

Optimization problem (5) is still non-convex due to the rank constraint. To handle the rank constraint, we rewrite C2 equivalently as $C2 : \|W_{m_y,m_z}\|_* - \|W_{m_y,m_z}\|_2 \leq 0$, see [8]. Constraint C2 is in the form of a differences of two convex functions. However, C2 is still non-convex. To overcome this issue, we adopt the penalty method [13, 14] and rewrite (5) as follows:

$$\begin{align*}
\text{maximize} & \quad \alpha - \eta(i) \left(\|W_{m_y,m_z}\|_* - \|W_{m_y,m_z}\|_2\right) \\
\text{s.t.} & \quad C1, C3, C4, \\
\end{align*}$$

(6)

where $\eta(i)$ is the penalty factor employed in the $i$-th iteration of the proposed Algorithm 1, see below. The objective function in (6) is still non-convex. To tackle this issue, we determine the first-order Taylor approximation of $\|W_{m_y,m_z}\|_2$ at initial point $W_{m_y,m_z}^{(i)}$ as follows:

$$\begin{align*}
\|W_{m_y,m_z}\|_2 & \geq \|W_{m_y,m_z}^{(i)}\|_2 \\
& + \operatorname{Tr}\left[\lambda_{\max}(W_{m_y,m_z}^{(i)}) \times \lambda_{\max}(W_{m_y,m_z}^{(i)}) (W_{m_y,m_z} - W_{m_y,m_z}^{(i)})\right]. \\
\end{align*}$$

(7)

By substituting (7) into (6), we obtain the following approximated optimization problem:

$$\begin{align*}
\text{maximize} & \quad \alpha - \eta(i) \left(\|W_{m_y,m_z}\|_* - \|W_{m_y,m_z}\|_2\right) \\
& - \operatorname{Tr}\left[\lambda_{\max}(W_{m_y,m_z}^{(i)}) \times \lambda_{\max}(W_{m_y,m_z}^{(i)}) (W_{m_y,m_z} - W_{m_y,m_z}^{(i)})\right] \\
\text{s.t.} & \quad C1, C3, C4. \\
\end{align*}$$

(8)

By substituting (7) into (6), we obtain the following approximated optimization problem:

$$\begin{align*}
\text{maximize} & \quad \alpha - \eta(i) \left(\|W_{m_y,m_z}\|_* - \|W_{m_y,m_z}\|_2\right) \\
& - \operatorname{Tr}\left[\lambda_{\max}(W_{m_y,m_z}^{(i)}) \times \lambda_{\max}(W_{m_y,m_z}^{(i)}) (W_{m_y,m_z} - W_{m_y,m_z}^{(i)})\right] \\
\text{s.t.} & \quad C1, C3, C4. \\
\end{align*}$$

(8)

Optimization problem (8) is convex because the objective function is concave and the constraints span a convex set. Therefore, it can be efficiently solved by standard convex optimization solvers such as CVX [15]. Algorithm 1 summarizes the main steps for solving (5) in an iterative manner, where the solution of (8) in iteration $i$ is used as the initial point for the next iteration $i + 1$. Assuming the maximum number of iterations $I_{\max}$ is chosen sufficiently large, and since problem (3) is reformulated as a difference-of-convex problem in (6) and TSA is used to convexify the problem, according to [13], Algorithm 1 produces a sequence of improved feasible points until convergence to a locally optimum point of (3). Codeword $w_{m_y,m_z}$ can be recovered as $w_{m_y,m_z} = \sqrt{\delta_{\max}} v$, where $v$ is the normalized eigenvector corresponding to the maximum and only non-zero eigenvalue $\delta_{\max} = \operatorname{Tr}(W_{m_y,m_z})$.

Algorithm 2 summarizes the main steps for obtaining the complete reflection codebook containing $M$ codewords.

Remark 1: Optimization problem (8) is an SDP problem. Thus, the complexity order of Algorithm 1 per iteration is $O\left(\left|\mathcal{B}_{y,m_y}\right| + \left|\mathcal{B}_{z,m_z}\right|\right)^3 + \left|\mathcal{B}_{y,m_y}\right| + \left|\mathcal{B}_{z,m_z}\right|\right)^2 Q^2 + \left|\mathcal{B}_{y,m_y}\right| + \left|\mathcal{B}_{z,m_z}\right|\right)^3 \approx O\left(\left|\mathcal{B}_{y,m_y}\right| + \left|\mathcal{B}_{z,m_z}\right|\right)^3\right)$, where $O$ is the big-O notation.

IV. DISCRETE PHASE-SHIFT CODEBOOK DESIGN

For ease of practical implementation, we also consider the case where the phase-shift of each IRS element can take only $2^b$ levels.

A. Optimization Problem

Similar to problem (3) for continuous phase shifts, the discrete phase-shift optimization problem can be formulated as follows:

$$\begin{align*}
\text{maximize} & \quad \alpha \\
\text{s.t.} & \quad C1 : |g_{m_y,m_z}(\beta_y, \beta_z)|^2 = \|w_{m_y,m_z}^H(y(\beta_y) \otimes z(\beta_z))|^2 \\
& \geq \alpha, \\
& \forall \beta_y \in \mathcal{B}_{y,m_y}, \beta_z \in \mathcal{B}_{z,m_z}, \\
& \forall \nu_q \in \{S\}, \forall q \in \{0, \ldots, Q - 1\}. \\
\end{align*}$$

(9)
Optimization problem (9) is a non-convex mixed-integer problem which is difficult to solve. However, we show in the following that problem (9) can be transformed into the canonical form of a binary linear problem which allows the application of CVX to solve it optimally.

**B. Proposed Solution**

We start by writing the relation:

\[
|w_{m_y,m_z}^H (y(\beta_y) \otimes z(\beta_z))^T|^2 = w_{m_y,m_z}^H A_{\beta_y,\beta_z} w_{m_y,m_z},
\]

where \( A_{\beta_y,\beta_z} = (y(\beta_y) \otimes z(\beta_z)) (y(\beta_y) \otimes z(\beta_z))^H \) is a positive semi-definite matrix. We can further rewrite (10) as follows:

\[
w_{m_y,m_z}^H A_{\beta_y,\beta_z} w_{m_y,m_z} = \sum_{q=0}^{Q-1} \sum_{i=0}^{Q-1} |A_{\beta_y,\beta_z}(q,i)| \left( \cos(\text{Ang}(A_{\beta_y,\beta_z}(q,i))) + j \sin(\text{Ang}(A_{\beta_y,\beta_z}(q,i))) \right) \left( \cos(\nu_q - \nu_i) + j \sin(\nu_q - \nu_i) \right),
\]

where \( A_{\beta_y,\beta_z}(q,i) \) is the \((q,i)\)-th entry of matrix \( A_{\beta_y,\beta_z} \), and \( \text{Ang}(A_{\beta_y,\beta_z}(q,i)) \) is the phase of \( A_{\beta_y,\beta_z}(q,i) \). Optimization problem (11) is non-linear with respect to the phase-shift due to the non-linear functions \( \cos(\cdot) \) and \( \sin(\cdot) \). To transform these non-linear functions, we apply a linear transformation as in [6] and [16]. Let us define \( \nu_q = \bar{\alpha}^T x_q \), where \( x_q \) is a binary vector of dimension \( 2^h \) satisfying \( \|x_q\| = 1 \), and \( \bar{\alpha} = [0, \Delta \nu, \ldots, (S - 1) \Delta \nu] \).

Moreover, we define \( \Delta \nu_i = \nu_q - \nu_i \). Similarly, we define binary vector \( y_{q,i} \) of dimension \( 2^{h+1} - 1 \). Thus, the following relation holds: \( \bar{\alpha}^T (x_q - x_i) = \bar{\alpha}^T y_{q,i} \), where \( \bar{\alpha} = [- (S - 1) \Delta \nu, \ldots, - \Delta \nu, 0, \Delta \nu, \ldots, (S - 1) \Delta \nu] \).

Optimization problem (9) can be rewritten as the following equivalent mixed binary linear optimization problem:

\[
\begin{align*}
\text{maximize } & \alpha \\
\text{s.t. } & C1: \sum_{q=0}^{Q-1} \sum_{i=0}^{Q-1} |A_{\beta_y,\beta_z}(q,i)| \left( \cos(\text{Ang}(A_{\beta_y,\beta_z}(q,i))) e^T \right) y_{q,i} \geq \alpha, \\
& \forall \beta_y \in \mathcal{G}_{y,m_y}, \beta_z \in \mathcal{G}_{z,m_z}, \\
& C2: x_q(j) \in \{0, 1\}, \forall j, q, \quad C3: \sum_{j=1}^{2^h} x_q(j) = 1, \forall q, \\
& C4: y_{q,i}(u) \in \{0, 1\}, \forall u, q, i, \quad 2^{h+1} - 1, \\
& C5: \sum_{u=1}^{2^{h+1} - 1} y_{q,i}(u) = 1, \forall q, i, \\
& C6: \bar{\alpha}^T (x_q - x_i) = \bar{\alpha}^T y_{q,i}, \forall q, i.
\end{align*}
\]

where \( y \) denotes the collection of optimization variables \( y_{q,i}, \forall q, i, x_q(j) \) is the \( j \)-th entry of vector \( x_q \), \( y_{q,i}(u) \) is the \( u \)-th entry of vector \( y_{q,i} \), \( e^T = \cos(\alpha^T) \), and \( s^T = \sin(\alpha^T) \). Optimization problem (12) is in the form of a binary linear optimization problem which can be solved using CVX [15]. The phase of the \( q \)-th element of \( w_{m_y,m_z} \) is given by \( \bar{\alpha}^T x_q \).

**Remark 2:** The complexity of problem (12) is exponential in \( Q \) and \( |S| \). However, (12) can be solved optimally using branch-and-bound implemented in modern solvers such as MOSEK. The convergence of the branch-and-bound can be found in [17] for interested readers.

**Remark 3:** The discrete phase-shift reflection codebook can be generated using Algorithm 2 by solving problem (12) to design the codewords \( w_{m_y,m_z} \).

**V. SIMULATION RESULTS**

Figs. 1a, 1b, 1c, and 1d show 2D heat-map plots for \( \bar{g}_{\max} g_{y,m_y} (\beta_y, \beta_z)^2 \) in dB, where \( y_{\max} = \bar{g} Q \), see (1), for different codebook designs. We consider \( M = 9 \) and design intervals \( B_{y,m_y} = [-0.3333, 0.3333] \) and \( B_{z,m_z} = [-0.3333, 0.3333] \). As can be seen, the proposed continuous design attains a more uniform coverage in the area of interest compared to the linear and quadratic designs. Moreover, Fig. 1 reveals that for 2-bit quantization, the proposed algorithm also provides uniform coverage in the area of the main lobe.

In Fig. 2, we show the trade-off between the average transmit power of the BS required to achieve an SNR of \( \gamma_{\text{req}} = 10 \) dB at the user and the codebook size \( M \). We assume the direct link between the BS and the user is blocked. The corresponding link distances are \( d_1 = d_2 = 150 \) m. The free-space path loss of the links is \( P_{\text{link}} = (c / (4 \pi f d_1))^2 \) and \( P_{\text{Lr}} = (c / (4 \pi f d_2))^2 \), where \( c \) is the speed of light and \( f = 3.4 \) GHz is the carrier frequency. The bandwidth is 20 MHz and the noise power spectral density is \(-174 \) dBm/Hz. In the following, we show the performance of the considered codebooks in LoS and non-LoS channels.

**a) LoS channel:** We consider LoS channels between the BS and the IRS and between the IRS and the user. The results in Fig. 2 were obtained by averaging over \( 10^5 \) uniformly random realizations of the AoAs and AoDs of the IRS. The transmit power required to ensure \( \gamma_{\text{req}} \) is given by \( P_{\text{req}} = \bar{g}^2 \max_{w_{m_y,m_z}} \|h^T \text{Diag}(w_{m_y,m_z}) g\|_2^2 / \sigma^2 \), where \( \sigma^2 \) is the receiver noise power, \( h \in \mathbb{C}^Q \) is the channel vector between IRS and user, and \( g \in \mathbb{C}^Q \) is the channel vector between BS and IRS. \( h \) and \( g \) are modelled using the path losses and the steering vectors at the IRS. We also show two performance upper bounds, one achieved with fully-controllable IRS phase shifts (asymptotically achievable as \( M \to \infty \)) and an unattainable bound for \( M \leq Q \) given in [2] leading to required transmits powers of \( P_{\text{req,F}} = \bar{g} \gamma_{\text{req}} \) and \( P_{\text{req,l}} = \bar{g} \gamma_{\text{req}} / 2 \), respectively. The first bound is not function of the codebook as this bound is obtained by optimizing the phase-shifts of each element while the second bound is when we have an optimal codebook where the IRS reflections pattern have rectangular shapes [2].

As can be seen from Fig. 2, the proposed codebook designs require a lower transmit power than the linear and the quadratic
of paths. For simplicity, we assume the channel from the BS to the IRS does not exist.

Fig. 1. Heat maps for different codebook designs, $Q = 225$, $M = 9$, $I_{\text{max}} = 25$, $d_y = d_z = \frac{\lambda}{2}$.

Fig. 2. Average total transmitted power versus codebook size, $Q = 121$, $\gamma_{\text{req}} = 10$ dB.

Fig. 3. Average total transmitted power versus codebook size for different values of $K$ in dB, $Q = 121$, $\gamma_{\text{req}} = 10$ dB, number of paths $L = 6$.

To model the non-LoS channel, we assume the relative power of non-LoS paths according to the value $K$.

As can be seen in Fig. 3, for all considered schemes for all considered $K$ values, the schemes converges to the same transmit power for the LoS channel, as for large codebooks, there exists a codeword which can be aligned the LoS path.

**Design, especially for small codebook sizes, $M \ll Q$, which is the regime of interest in practice. This is due to the fact that the proposed designs cover the design interval of $B_{y,m}$ for all codewords more uniformly than the baseline schemes. For $M \approx Q$, the linear design slightly outperforms the proposed and the quadratic designs. This is because although the linear design yields the most narrow beam, with $M = Q$, it can still cover the entire angular space. Hence, further widening of the beam, as is done for the proposed and the quadratic designs, is not required and degrades the performance.

Fig. 2 also shows the impact of phase-shift quantization on the average total transmit power. We observe that a 1-bit phase-shift quantization requires a higher average transmitted power compared to continuous phase shifts. This is expected since due to coarse discrete phase shifts, the signals reflected by the IRS cannot be perfectly aligned in phase at the receiver, thus resulting in a power loss.

In contrast, for 2-bit quantization, the performance is close to that for continuous phase shifts.

b) **Non-LoS channel:** To model the non-LoS channel, we define the relative powers of the LoS and non-LoS paths as $K = \frac{P_{L_0}}{\sum_{l=1}^{L-1} P_{L_l}}$, where $P_{L_0}$ is the power of the LoS path, $P_{L_l}$ is the power of each non-LoS path, and $L$ is the number of paths. For simplicity, we assume the channel from the BS to the IRS and from the IRS to the user have the same value of $K$.

We assume free-space path loss for the LoS paths and change the relative power of non-LoS paths according to the value of $K$.

References

[1] M. Najafi et al., “Physics-based modeling and scalable optimization of large intelligent reflecting surfaces,” IEEE Trans. Commun., vol. 69, no. 4, pp. 2673–2691, Apr. 2021.

[2] V. Jamali et al., “Power efficiency, overhead, and complexity tradeoff of IRS codebook design—Quadratic phase-shift profile,” IEEE Commun. Lett., vol. 25, no. 6, pp. 2048–2052, Jun. 2021.

[3] Q. Wu and K. Zhang, “Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming,” IEEE Trans. Wireless Commun., vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[4] C. Huang et al., “Reconfigurable intelligent surfaces for energy efficiency in wireless communication,” IEEE Trans. Wireless Commun., vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[5] M. D. Renzo et al., “Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come,” EURASIP J. Wireless Commun. Netw., vol. 2019, no. 1, pp. 4157–4170, May 2019.

[6] Q. Q. Wu and R. Zhang, “Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts,” IEEE Trans. Commun., vol. 68, no. 3, pp. 1838–1851, May 2020.

[7] Y. Tang et al., “Joint transmit and reflective beamforming design for IRS-assisted multiser MISO SWIPT systems,” in Proc. IEEE Intern. Conf. Commun. (ICC), Jun. 2020, pp. 1–6.

[8] X. Yu et al., “Power-efficient resource allocation for multiuser MISO systems via intelligent reflecting surfaces,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2020, pp. 1–6.

[9] B. Zheng and R. Zhang, “Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization,” IEEE Wireless Commun. Lett., vol. 9, no. 4, pp. 518–522, Apr. 2020.

[10] B. Zheng et al., “Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design,” IEEE Trans. Wireless Commun., vol. 19, no. 12, pp. 8315–8329, Dec. 2020.

[11] M. Rahal et al., “Arbitrary beam pattern approximation via RISs with measured element responses,” 2022, arXiv:2203.07225.

[12] D. Xu et al., “Optimal resource allocation design for large IRS-assisted swipt systems: A scalable optimization framework,” IEEE Trans. Commun., vol. 70, no. 2, pp. 1423–1441, Feb. 2022.

[13] H. A. Le Thi et al., “Exact penalty and error bounds in DC programming,” J. Global Optim., vol. 52, no. 3, pp. 509–535, Sep. 2011.

[14] X. Yu et al., “Robust and secure wireless communications via intelligent reflecting surfaces,” IEEE J. Sel. Areas Commun., vol. 38, no. 11, pp. 2637–2652, Nov. 2020.

[15] M. Grant and S. Boyd. (Mar. 2014). Disciplined Convex Programming, Version 2.1. [Online]. Available: http://cvxr.com/cvx

[16] B. Di et al., “Hybrid beamforming for reconfigurable intelligent surface based multi-user communication systems: Achievable rates with limited discrete phase shifts,” IEEE J. Sel. Areas Commun., vol. 38, no. 8, pp. 1809–1822, Aug. 2020.

[17] H. T. R. Horst, Global Optimization Deterministic Approaches. Berlin, Germany: Springer-Verlag, 1990.