Minimizing Excitation Quantization Bits in Pattern Synthesis via Mixed-Integer Programming

Xuejing Zhang, Xue Shi, Cheng Liu, and Xuepan Zhang

***Abstract*—This article addresses the problem of minimizing excitation quantization bits in pattern synthesis to reduce costs while meeting performance requirements on the beampattern. To achieve this task, we propose a framework leveraging mixed-integer programming (MIP), which models problems by introducing both continuous and discrete variables. The proposed method is applicable to various pattern synthesis scenarios and general array geometries, with the capability to minimize the quantization bits for both excitation phase and excitation amplitude. In addition, the method supports joint optimization, allowing for optimal beampattern performance while minimizing quantization bits. The optimality of the proposed MIP model is theoretically guaranteed, and it can be efficiently solved using off-the-shelf solvers. Our approach differs from existing methods that predetermine quantization bits, as it obtains the minimal excitation quantization bits under the given beampattern specifications. Extensive simulations are presented to demonstrate the effectiveness and superiority of the proposed method.**

***Index Terms*—Array pattern synthesis, excitation quantization, mixed-integer programming (MIP), phased array, quantization sidelobe.**

I. Introduction

PATTERN synthesis, as one of the fundamental problems in array signal processing, plays a pivotal role particularly in radar and wireless communication. It involves the deliberate design of specific radiation patterns to optimize performance criteria like directivity, resolution, and interference mitigation [1], [2], [3]. In radar applications, pattern synthesis enables the precise control of beam shapes for enhanced target detection and tracking capabilities. In wireless communication, a well-designed beampattern facilitates efficient spectrum utilization and minimizes interference among users.

Due to its significant applications, pattern synthesis has attracted considerable research interest, leading to the development of numerous techniques over the past few decades [4], [5], [6], [7]. For example, several classical algorithms [8], [9], [10] can offer closed-form expressions for specific radiation patterns. However, their applicability is limited to certain array geometries, such as uniform linear arrays (ULAs). To achieve pattern synthesis for general arrays, evolutionary algorithms were proposed to seek optimal excitations through stochastic approaches [11], [12], [13]. However, these methods are generally time-consuming, which limits their practical applications.

With the development of convex optimization theory [14], it has been successfully applied to pattern synthesis in recent years [15], [16], [17], [18], [19]. For instance, Lebret and Boyd [20] have illustrated how convex programming (CP) can be employed to design optimal patterns for arbitrary antenna arrays. Fuchs and Rondineau [21] presented simple and efficient procedures for synthesizing array patterns while controlling the excitations, by utilizing CP and norm minimization. Wang et al. [22] addressed magnitude control and robust design in array pattern synthesis and formulated the synthesis task as semidefinite programming problems. Fuchs [23] presented a general procedure for efficiently solving nonconvex array synthesis problems based on the semidefinite relaxation (SDR) technique [24]. Different from convex optimization methods, another type of pattern synthesis algorithm designs beampatterns with the aid of adaptive array theory [25]. Although having the ability to precisely control the array response as presented in [26], [27], and [28], they exhibit limited flexibility in controlling mainlobe width and mainlobe level. In addition to the aforementioned methods, there are also several works that achieve pattern synthesis using different approaches, such as utilizing the beam superposition [29], Fourier–Bessel series expansion [30], and spectral factorization (SF) [31].

It is worth noting that the aforementioned pattern synthesis algorithms do not consider the quantization effects on excitation phase and amplitude. In practice, phased array systems often employ digital phase shifters and digital attenuators to control the excitation phase and excitation amplitude, respectively. Since the digital devices have finite resolution, it results in quantization on the excitations and deteriorates the overall performance on the radiation pattern [32], [33], [34]. To achieve pattern synthesis under the constraint of excitation phase quantization, Ismail and Hamici [35] proposed a quantized particle swarm optimization (PSO) algorithm. It searches for the optimal phase solutions within quantized values to minimize the sidelobe level while preserving the main beam. A binary differential evolution algorithm is introduced in

Received 4 April 2025; revised 11 June 2025; accepted 3 July 2025. Date of publication 15 July 2025; date of current version 30 October 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62101101, in part by Sichuan Science and Technology Program under Grant 2024NSFSC1433, and in part by the Peng Cheng Shang Xue Education Fund under Grant XY2021602. *(Corresponding author: Xuejing Zhang.)*

Xuejing Zhang and Cheng Liu are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zhangxuejing@uestc.edu.cn; lc0740@std.uestc.edu.cn).

Xue Shi is with the College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu 610059, China (e-mail: waterns@126.com).

Xuepan Zhang is with Hangzhou Institute of Technology, Xidian University, Hangzhou 311200, China (e-mail: xpzhang7@163.com).

Digital Object Identifier 10.1109/TAP.2025.3587538

[36] to design a low-sidelobe beampattern with quantized excitation phase. Considering phase quantization or amplitude quantization, Keizer [37] proposed a low-sidelobe pattern synthesis algorithm based on quantized iterative Fourier transform (QIFT). Although the sidelobe level can be improved, this method is only applicable to arrays with specific geometries. Apart from the aforementioned work, there are also studies that discuss phase round-off methods to reduce beam pointing errors caused by quantized phases, as detailed in [38], [39], and [40].

It should be pointed out that the above pattern synthesis algorithms with quantized excitations require a predetermined number of excitation quantization bits, which then guides the design of the radiation pattern for the given quantization level. In practice, increasing the number of quantization bits leads to higher costs and increased complexity in the hardware system. Therefore, a balance needs to be achieved between performance and practicality. Provided that the performance of the radiation pattern (e.g., sidelobe level and accuracy of mainlobe axis) can be satisfied, it is usually desirable to minimize the number of quantization bits as much as possible. To our knowledge, the problem of minimizing the number of excitation quantization bits in pattern synthesis has received limited attention. Thompson et al. [41] proposed a quantized phase bit allocation algorithm, which employs simulated annealing (SA) to minimize the phase quantization bits. However, only phase quantization is considered in [41], while the excitation amplitude must be preassigned. Moreover, the employed SA algorithm may suffer from drawbacks such as slow convergence and sensitivity to parameters. The report in [42] presented design rules for determining the number of quantization bits required for excitation phase and amplitude. Nevertheless, the conclusions drawn from [42] are only applicable to the Taylor taper.

In this article, we discuss how to minimize the number of excitation quantization bits in pattern synthesis, in order to reduce the cost while meeting the performance requirements on the beampattern. We found that the problem of minimizing quantization bits can be effectively modeled via mixed-integer programming (MIP) [43], [44], [45]. The MIP problem typically includes both continuous and discrete variables, and it can be efficiently solved using general-purpose off-the-shelf solvers, such as GUROBI [46] and CPLEX [47]. In this article, we study the problem of minimizing quantization bits in different pattern synthesis scenarios. Apart from purely minimizing the quantization bits, the proposed scheme can also be extended to joint optimization scenarios, aiming to achieve optimal beampattern performance while minimizing the quantization bits. The main contributions of this article can be summarized as follows.

1) We propose a framework to achieve the minimization of excitation quantization bits in pattern synthesis, utilizing MIP. Our work is applicable to different scenarios of pattern synthesis and is suitable for general array geometries (e.g., linear arrays, planar arrays, and conformal arrays), with the capability to minimize the quantization bits for both excitation phase and excitation amplitude.
2) Different from existing methods that often predetermine the excitation quantization bits, the proposed method can obtain the minimal excitation quantization bits under the given beampattern specifications. Furthermore, the proposed method can be extended to joint optimization scenarios, enabling the achievement of optimal beampattern performance while minimizing the excitation quantization bits.
3) The optimality of the proposed MIP model is theoretically guaranteed, and it can be efficiently solved using off-the-shelf solvers. We conduct performance evaluations of the proposed method in different scenarios through simulations. The results confirm the effectiveness and superiority of the proposed method.

The rest of the article is organized as follows. In Section II, we introduce some preliminaries on excitation quantization in pattern synthesis. In Section III, we present the proposed method by using the minimization of phase quantization bits as an example. Extensions to other scenarios and further explorations are presented in Section IV. Simulation results are shown in Section V and conclusion is drawn in Section VI.

A. Notations

We use bold upper case and lower case letters to represent matrices and vectors, respectively. In particular, we use $\mathbf{1}$ and $\mathbf{0}$ to represent an all-one vector and all-zero vector, respectively, where the vector dimension is specified by the subscript or inferred from the context. $(\cdot)^T$ denotes the transpose operation, $|\cdot|$ calculates absolute value and $\angle(\cdot)$ returns the argument of the input. The real part of a complex number is denoted by $\Re(\cdot)$. We use $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}_+$, and $\{0, 1\}$ to denote the set of real numbers, the set of complex numbers, the set of positive integers, and the set of binary elements (where each element can only take either 0 or 1), respectively. $\otimes$ denotes the Kronecker product. $\Psi_1 \subsetneq \Psi_2$ indicates that Ψ_1 is a proper subset of Ψ_2. We use the notation $\preceq$ to represent elementwise comparison between matrices. For example, $\boldsymbol{A} \preceq \boldsymbol{B}$ implies that each element of $\boldsymbol{A}$ is not greater than the corresponding element of $\boldsymbol{B}$ in the same position.

II. Preliminaries

We first introduce some preliminaries on pattern synthesis and excitation quantization. Take a 1-D linear array as an example, although our method can be readily generalized to arbitrary array geometries (e.g., planar and conformal arrays). Let the number of array elements be N. Under the assumption of far field and narrowband, the array radiation pattern can be expressed as

$$f(\theta) = \left|\boldsymbol{w}^T \boldsymbol{a}(\theta)\right| \tag{1}$$

where $\boldsymbol{w} = [w_1, \ldots, w_N]^T$ represents the complex excitation vector, $\boldsymbol{a}(\theta)$ is the steering vector for the spatial angle θ and is defined as

$$\boldsymbol{a}(\theta) = \left[g_1(\theta)\, e^{-j\omega\tau_1(\theta)}, \ldots, g_N(\theta)\, e^{-j\omega\tau_N(\theta)}\right]^T \tag{2}$$

where $g_n(\theta)$ represents the element pattern of the nth antenna, ω is the operating frequency, and $\tau_n(\theta)$ is the time delay

between the nth antenna element and the reference one, $n = 1, \ldots, N$. For a given antenna array, the purpose of pattern synthesis is to design the excitation vector $\boldsymbol{w}$ to meet specific requirements, such as achieving desired directivity, reducing peak sidelobe level (PSL), or synthesizing the beam with a specific shape.

For practical phased arrays, the excitation phase (i.e., $\angle w_n$, $n = 1, \ldots, N$) is typically controlled by the digital phase shifter. Considering the finite resolution of the digital phase shifter, for Q-bit quantization, the excitation phase is required to take values from the following finite set:

$$\Psi_Q \triangleq \left\{0, \delta_Q, 2\delta_Q, \ldots, \left(2^Q - 1\right)\delta_Q\right\} \tag{3}$$

where $\delta_Q = 2\pi/2^Q$ represents the phase resolution corresponding to the Q-bit phase quantization.

The excitation amplitude (i.e., $|w_n|$, $n = 1, \ldots, N$) of an antenna array is typically controlled by digital attenuators. For U-bit quantization, the excitation amplitude is required to take values from the following finite set [42]:

$$\Upsilon_U \triangleq \left\{\alpha_1^{(U)}, \alpha_2^{(U)}, \ldots, \alpha_{2^U}^{(U)}\right\} \tag{4}$$

where $\alpha_{2^U}^{(U)}$ (in decibels) represents the given dynamic range of the digital attenuator and the elements of Υ_U are calculated as follows:

$$\alpha_i^{(U)} = \frac{(i-1)\cdot \alpha_{2^U}^{(U)}}{2^U - 1},\ i = 1, \ldots, 2^U. \tag{5}$$

It should be noted that the elements of Υ_U are all described in decibels. For example, assuming the dynamic range is 35 dB, for a 3-bit amplitude quantization, we have $\Upsilon_3 = \{0, 5, 10, 15, 20, 25, 30, 35\}$, where the unit is in decibels.

III. Minimizing Phase Quantization Bits via MIP

In this section, we investigate the minimization of phase quantization bits in pattern synthesis. We formulate the problem using MIP and discuss dimensionality reduction techniques for accelerated solving.

A. Minimizing Phase Quantization Bits for Single Focused-Beam Synthesis

For simplicity, we consider focused-beam synthesis with a single beam. The problem of pattern synthesis under Q-bit excitation phase quantization can be formulated as follows:

$$\text{find}\ \ \boldsymbol{w} \tag{6a}$$
$$\text{s.t.}\ \ \Re\left[\boldsymbol{w}^T \boldsymbol{a}\left(\bar{\theta}_0\right)\right] = 1 \tag{6b}$$
$$\left|\boldsymbol{w}^T \boldsymbol{a}(\theta)\right| \le \rho(\theta),\ \ \theta \in \Theta \tag{6c}$$
$$\angle w_n \in \Psi_Q,\ \ n = 1, \ldots, N \tag{6d}$$

where $\bar{\theta}_0$ is the mainlobe axis, Θ represents the sidelobe region, and $\rho(\theta)$ sets an upper bound for the sidelobe level. In problem (6), the constraint (6d) indicates that the excitation phase must be selected from a given set Ψ_Q, as defined in (3). It is precisely due to the discrete constraint on the excitation that we have taken the real part of the mainlobe beam response, as shown in constraint (6b), to prevent the problem from being infeasible.

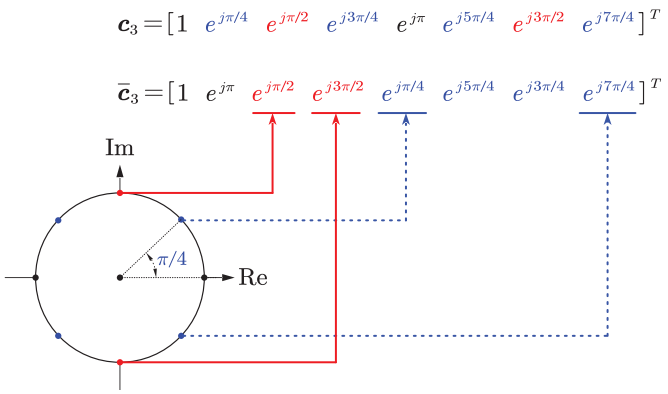


Fig. 1. Illustration of $\boldsymbol{c}_M$ and $\bar{\boldsymbol{c}}_M$ when $M = 3$.

Considering that lower quantization bits can reduce system costs, our focus is on minimizing the phase quantization bit Q, given the parameters $\bar{\theta}_0$, Θ, and $\rho(\theta)$. This corresponds to the following problem:

$$\min_{\boldsymbol{w}, Q}\ \ Q \tag{7a}$$
$$\text{s.t.}\ \ \Re\left[\boldsymbol{w}^T \boldsymbol{a}\left(\bar{\theta}_0\right)\right] = 1 \tag{7b}$$
$$\left|\boldsymbol{w}^T \boldsymbol{a}(\theta)\right| \le \rho(\theta),\ \ \theta \in \Theta \tag{7c}$$
$$\angle w_n \in \Psi_Q,\ \ n = 1, \ldots, N \tag{7d}$$
$$Q \in \mathbb{Z}_+. \tag{7e}$$

Unlike problem (6) where we have preassigned the quantization bit, in the above problem (7), the quantization bit Q is an optimization variable. This increases the difficulty of problem (7), because the candidate set of excitation phase is discrete and undetermined. It should also be noted that the excitation amplitude can take continuous values. Therefore, problem (7) involves both continuous and discrete variables.

Before addressing problem (7), we first notice that the set of a higher bit quantization includes all sets of lower bit quantization. Specifically, assuming M_1 and M_2 are two quantization bit levels, and $M_1 < M_2$, then the phase candidate set Ψ_{M_1} must be a proper subset of the set Ψ_{M_2}, i.e.,

$$\Psi_{M_1} \subsetneq \Psi_{M_2}. \tag{8}$$

In other words, a lower bit quantization phase candidate set can be formed by selecting some elements from a higher bit one.

To facilitate the description of the set relationship in (8) and to model the problem of minimizing phase quantization bits described in (7), we consider a general phase quantization bit M. We can then define the following 2^M-dimensional vector $\boldsymbol{c}_M$ by enumerating the complex exponentials of the candidate phases corresponding to the M-bit quantization

$$\boldsymbol{c}_M \triangleq \left[1, e^{j2\pi/2^M}, \ldots, e^{j\left(2^M-1\right)\cdot 2\pi/2^M}\right]^T. \tag{9}$$

By appropriately reordering the elements of vector $\boldsymbol{c}_M$ (denoted as $\bar{\boldsymbol{c}}_M$ after reordering), we can ensure that the first 2^m elements of $\bar{\boldsymbol{c}}_M$ correspond exactly to the candidate phases

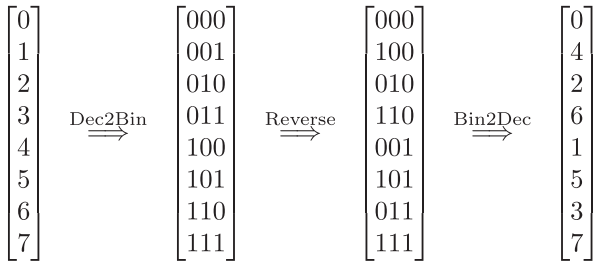


Fig. 2. Schematic of calculating r_i using reverse binary operation when $M = 3$.

for m-bit quantization, $m = 1, \ldots, M$. Taking $M = 3$ as an example, we reorder vector $\boldsymbol{c}_3$ into the following vector $\bar{\boldsymbol{c}}_3$:

$$\bar{\boldsymbol{c}}_3 = \left[1, e^{j\pi}, e^{j\pi/2}, e^{j3\pi/2}, e^{j\pi/4}, e^{j5\pi/4}, e^{j3\pi/4}, e^{j7\pi/4}\right]^T.$$

It can be seen that the first two and the first four elements of vector $\bar{\boldsymbol{c}}_3$ correspond to the candidate phases for 1- and 2-bit quantization, respectively, as illustrated in Fig. 1.

For a general phase quantization bit M, we can see from (9) that the phase of the ith element in $\boldsymbol{c}_M$ is $2\pi(i-1)/2^M$, $i = 1, \ldots, 2^M$. Due to the exchange of element order, we can always represent the ith element of $\bar{\boldsymbol{c}}_M$ as $2\pi(r_i - 1)/2^M$, where r_i is the new index of the phase after reordering, $i = 1, \ldots, 2^M$. It can be verified that by performing a reversed binary operation, we can obtain r_i from i, thereby deriving the vector $\bar{\boldsymbol{c}}_M$. The process can be divided into the following steps.

1) For a given $i \in \{1, \ldots, 2^M\}$, obtain the M-bit binary code of $(i-1)$.
2) Reverse the resulting binary code from left to right to get the reversed binary code.
3) Obtain the decimal value x_i corresponding to the reversed binary code.
4) Let $r_i - 1 = x_i$, thus obtaining $r_i = x_i + 1$.

To provide an intuitive explanation, Fig. 2 illustrates the schematic of the above process using $M = 3$ as an example, which is related to Fig. 1. In Fig. 2, "Dec2Bin" and "Bin2Dec" represent decimal to binary conversion and binary to decimal conversion, respectively, while "Reverse" stands for the M-bit binary reversal operation.

Based on the above analysis, we can predetermine a sufficiently large M, requiring only that it satisfies

$$Q_* \le M \tag{10}$$

where Q_* represents the optimal value for problem (7). By introducing an amplitude matrix $\boldsymbol{D}$, the excitation vector $\boldsymbol{w}$ in the problem (7) can be expressed in the following form:

$$\boldsymbol{w} = \boldsymbol{D}\bar{\boldsymbol{c}}_M \tag{11}$$

where $\boldsymbol{D} \in \mathbb{R}^{N \times 2^M}$ is the amplitude matrix to be determined, $\bar{\boldsymbol{c}}_M$ represents the reordered vector corresponding to M-bit phase quantization.

With (11), we can achieve the selection of quantized excitation phase for each antenna by designing the matrix $\boldsymbol{D}$. Note that each row of $\boldsymbol{D}$ must contain precisely one nonzero element, representing the amplitude of the corresponding antenna.

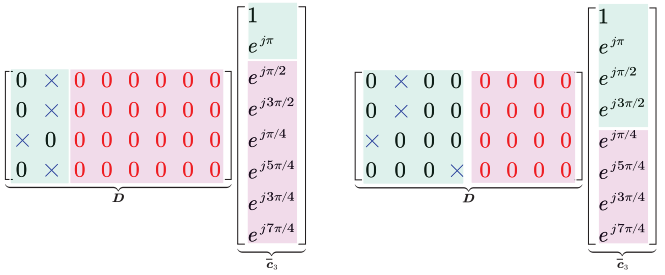


Fig. 3. Illustration of matrix $\boldsymbol{D}$ under different quantization bit requirements when $N = 4$ and $M = 3$. (a) 1-bit phase quantization. (b) 2-bit phase quantization.

To facilitate this, we introduce a binary matrix $\boldsymbol{B} \in \{0,1\}^{N \times 2^M}$ and impose the following two constraints:

$$\boldsymbol{B}\mathbf{1} = \mathbf{1} \tag{12}$$

$$\mathbf{0} \le \boldsymbol{D} \le \eta \cdot \boldsymbol{B} \tag{13}$$

where η represents a positive constant greater than the maximum of excitation amplitudes. Constraint (12) ensures that each row of matrix $\boldsymbol{B}$ has exactly one element equal to 1. In the presence of constraint (13), it is observed that an element in $\boldsymbol{D}$ can only take a nonzero value if the corresponding element in $\boldsymbol{B}$ is 1. The above two constraints guarantee that each row of $\boldsymbol{D}$ can only have one nonzero element, which corresponds to the excitation amplitude of the antenna. The column number where the nonzero element is located corresponds to the index of the selected excitation phase in $\bar{\boldsymbol{c}}_M$.

To minimize the number of phase quantization bits, it is necessary to select phases from the front of the vector $\bar{\boldsymbol{c}}_M$ as far as possible. This requires that the nonzero columns of the amplitude matrix $\boldsymbol{D}$ are located as far forward as possible, while the zero columns should be arranged toward the end. For an intuitive explanation, Fig. 3 presents a schematic representation of matrix $\boldsymbol{D}$ obtained under different quantization bits, when $N = 4$ and $M = 3$. The cross symbol in Fig. 3 denotes some nonzero elements and the same applies throughout the subsequent figures. Specifically, Fig. 3(a) corresponds to 1-bit phase quantization, while Fig. 3(b) corresponds to 2-bit phase quantization. It can be observed that the fewer phase quantization bits there are, the less the number of nonzero columns in $\boldsymbol{D}$, and their positions are more toward the front.

In fact, by considering the effect of constraint (13) and applying appropriate constraints to the binary matrix $\boldsymbol{B}$, we can effectively control the priority of the nonzero columns in $\boldsymbol{D}$, thus preferentially selecting phases that are ranked at the beginning of $\bar{\boldsymbol{c}}_M$. Specifically, we introduce an M-dimensional binary vector $\boldsymbol{q}$ and constrain that

$$\boldsymbol{B}^T\mathbf{1} \le N \cdot \boldsymbol{F}\boldsymbol{q} \tag{14}$$

where the matrix $\boldsymbol{F}$ is defined as

$$\boldsymbol{F} \triangleq \begin{bmatrix} \mathbf{1}_2 & & & & \\ & \mathbf{1}_2 & & & \\ & & \mathbf{1}_4 & & \\ & & & \ddots & \\ & & & & \mathbf{1}_{2^{M-1}} \end{bmatrix} \in \mathbb{R}^{2^M \times M} \tag{15}$$

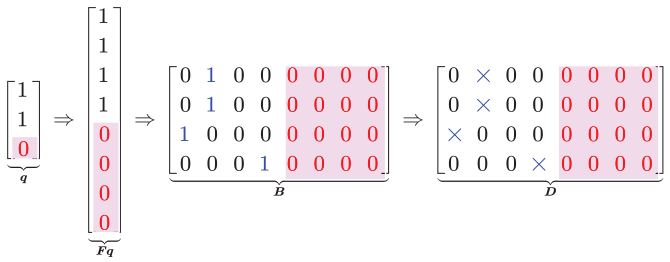


Fig. 4. Illustration of the constraint relationship between $\boldsymbol{q}$, $\boldsymbol{B}$, and $\boldsymbol{D}$ when $N = 4$, $M = 3$, and $Q = 2$.

with $\mathbf{1}_k$ standing for the all-one vector of dimension k. For $\forall i \in \{1, \ldots, M\}$, if the first i elements of $\boldsymbol{q}$ are ones and the remaining elements are zero, then all the ones in $\boldsymbol{B}$ are located within its first 2^i columns, under the above constraint (14). This further implies that the required number of bits for phase quantization does not exceed i. Therefore, if the binary vector $\boldsymbol{q}$ is monotonically nonincreasing, the number of ones in $\boldsymbol{q}$ can be used to represent the required number of bits for phase quantization.

To enforce the monotonically nonincreasing property of vector $\boldsymbol{q}$, we impose the following constraint:

$$\boldsymbol{G}_M \boldsymbol{q} \preceq \mathbf{0}. \tag{16}$$

The matrix $\boldsymbol{G}_M$ is defined as

$$\boldsymbol{G}_M \triangleq \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(M-1)\times M} \tag{17}$$

where the subscript M is used to indicate the column number of $\boldsymbol{G}_M$.

Under the effect of the above constraints (14) and (16), the resulting vector $\boldsymbol{q}$ has leading elements being ones, while the remaining elements are zeros. Moreover, the number of ones in $\boldsymbol{q}$ is equal to the required number of phase quantization bits, i.e.,

$$Q = \mathbf{1}^T \boldsymbol{q}. \tag{18}$$

Therefore, we can achieve the minimization of phase quantization bit Q by minimizing the sum of the elements in the binary vector $\boldsymbol{q}$. To further provide an intuitive explanation, Fig. 4 presents a schematic illustrating the constraint relationship between $\boldsymbol{q}$, $\boldsymbol{B}$, and $\boldsymbol{D}$, when $N = 4$, $M = 3$, and $Q = 2$.

Based on the above analysis, by substituting (11) into the problem (7), we can ultimately formulate the problem of minimizing phase quantization bits as follows:

$$\min_{\boldsymbol{D},\boldsymbol{B},\boldsymbol{q}} \ \mathbf{1}^T \boldsymbol{q} \tag{19a}$$

$$\text{s.t. } \Re\left[\bar{\boldsymbol{c}}_M^T \boldsymbol{D}^T \boldsymbol{a}\left(\bar{\theta}_0\right)\right] = 1 \tag{19b}$$

$$\left|\bar{\boldsymbol{c}}_M^T \boldsymbol{D}^T \boldsymbol{a}\left(\theta\right)\right| \le \rho\left(\theta\right), \ \theta \in \Theta \tag{19c}$$

$$\mathbf{0} \preceq \boldsymbol{D} \preceq \eta \cdot \boldsymbol{B} \tag{19d}$$

$$\boldsymbol{B}\mathbf{1} = \mathbf{1} \tag{19e}$$

$$\boldsymbol{B}^T \mathbf{1} \preceq N \cdot \boldsymbol{F}\boldsymbol{q} \tag{19f}$$

$$\boldsymbol{G}_M \boldsymbol{q} \preceq \mathbf{0} \tag{19g}$$

$$\boldsymbol{B} \in \{0, 1\}^{N \times 2^M} \tag{19h}$$

$$\boldsymbol{q} \in \{0, 1\}^M. \tag{19i}$$

The above formulation (19) is an MIP problem, where some variables (see matrix $\boldsymbol{B}$ and vector $\boldsymbol{q}$) can only take on binary values, while others (see matrix $\boldsymbol{D}$) can take on continuous values. Note that if we omit the discrete constraints (19h) and (19i), we obtain a convex optimization problem. In fact, for the above problem (19), we can solve it using MIP solvers, such as GUROBI [46] and CPLEX [47].

Remark 1: In models (6), (7), and (19), we impose the constraint $\Re[\boldsymbol{w}^T \boldsymbol{a}(\bar{\theta}_0)] = 1$ to fix the gain at mainlobe axis. This formulation is adopted because the mainlobe response $\boldsymbol{w}^T \boldsymbol{a}(\bar{\theta}_0)$ in these models may take complex values when excitation phase quantization is considered, making it impossible to predetermine its value. The linear constraint $\Re[\boldsymbol{w}^T \boldsymbol{a}(\bar{\theta}_0)] = 1$ can simplify the optimization model. Moreover, since $|\boldsymbol{w}^T \boldsymbol{a}(\bar{\theta}_0)| \ge \Re[\boldsymbol{w}^T \boldsymbol{a}(\bar{\theta}_0)]$, the constraint $\Re[\boldsymbol{w}^T \boldsymbol{a}(\bar{\theta}_0)] = 1$ guarantees that the beam gain at the mainlobe axis satisfies $|\boldsymbol{w}^T \boldsymbol{a}(\bar{\theta}_0)| \ge 1$. Theoretically, our method could potentially achieve a normalized PSL lower than the preset one.

Remark 2: From a theoretical perspective, the model (19) is inherently a nonconvex optimization problem due to the presence of binary variables. This nonconvex characteristic leads to the possible existence of multiple global optimal solutions that typically achieve identical quantization bit numbers while exhibiting different weight vector solutions. Although computationally challenging, the model remains tractable, and the MIP solvers can guarantee finding at least one optimal solution under proper configurations. In practical implementations, particularly for large-scale arrays, we must balance solution optimality against computation time. Our simulation results demonstrate that for moderate-scale arrays, the proposed method reliably obtains optimal solutions (as verified by solver termination status) within a reasonable time. For more complex scenarios, it efficiently produces high-quality suboptimal solutions.

B. Dimensionality Reduction for Accelerated Solving

For a MIP problem, a common method to accelerate its solving is to reduce the dimensionality of the optimization variables, including both integer and continuous variables. In fact, we can equivalently halve the variable dimensionality of problem (19), thus obtaining a more simplified MIP formulation.

To reduce the dimensionality of variables, we first observe that for any quantization bit M, the phase candidates can be divided into two groups, where the phases of one group can be obtained by shifting the phases of the other group by π. Equivalently, this implies that half of the elements in vector $\boldsymbol{c}_M$ or $\bar{\boldsymbol{c}}_M$ can be obtained through multiplying the other half of the elements by negative one, due to the fact that $e^{j\pi} = -1$. Take $M = 3$ as an example, we can rewrite $\boldsymbol{c}_3$ as

$$\boldsymbol{c}_3 = \left[1, e^{j\pi/4}, e^{j\pi/2}, e^{j3\pi/4}, -1, -e^{j\pi/4}, -e^{j\pi/2}, -e^{j3\pi/4}\right]^T.$$

It can be observed that by introducing negative signs, we can equivalently represent vector $\boldsymbol{c}_3$ using only half of its elements.

TABLE I
COMPARISON OF MATRIX DIMENSIONS

Matrix	Dimension	Matrix	Dimension
$\bar{\boldsymbol{c}}_M$	2^M	$\check{\boldsymbol{c}}_M$	2^{M-1}
$\boldsymbol{D}$	$N \times 2^M$	$\check{\boldsymbol{D}}$	$N \times 2^{M-1}$
$\boldsymbol{B}$	$N \times 2^M$	$\check{\boldsymbol{B}}$	$N \times 2^{M-1}$
$\boldsymbol{F}$	$2^M \times M$	$\check{\boldsymbol{F}}$	$2^{M-1} \times (M-1)$

In fact, for any phase quantization bit M, it is sufficient to retain only the first half of the elements in $\boldsymbol{c}_M$, as all the remaining elements of $\boldsymbol{c}_M$ can be represented by introducing negative signs to this subset. This further allows us to reduce the dimensionality of variables $\boldsymbol{D}$, $\boldsymbol{B}$ and $\boldsymbol{q}$, by utilizing a shortened $\boldsymbol{c}_M$ or $\bar{\boldsymbol{c}}_M$.

Specifically, by eliminating elements with argument angles belonging to $[\pi, 2\pi)$ from $\bar{\boldsymbol{c}}_M$, we obtain a 2^{M-1}-dimensional vector $\check{\boldsymbol{c}}_M$. Similar to (11), we can express the excitation vector $\boldsymbol{w}$ as

$$\boldsymbol{w} = \check{\boldsymbol{D}}\check{\boldsymbol{c}}_M \tag{20}$$

where $\check{\boldsymbol{D}} \in \mathbb{R}^{N\times 2^{M-1}}$ is the generalized amplitude matrix, and its elements can take negative values. It should be clarified that when an element of $\check{\boldsymbol{D}}$ takes a negative value, the actual excitation amplitude corresponds to the absolute value of that element, while the excitation phase is the selected phase plus π.

To further minimize the number of phase quantization bits, we introduce a binary matrix $\check{\boldsymbol{B}} \in \{0,1\}^{N\times 2^{M-1}}$. Similar to constraints (12)–(14), we impose the following constraints:

$$\check{\boldsymbol{B}}\mathbf{1} = \mathbf{1} \tag{21}$$

$$-\eta \cdot \check{\boldsymbol{B}} \preceq \check{\boldsymbol{D}} \preceq \eta \cdot \check{\boldsymbol{B}} \tag{22}$$

$$\check{\boldsymbol{B}}^T\mathbf{1} \preceq N \cdot \check{\boldsymbol{F}}\boldsymbol{q} \tag{23}$$

where

$$\check{\boldsymbol{F}} \triangleq \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & \mathbf{1}_2 & & \\ & & & \ddots & \\ & & & & \mathbf{1}_{2^{M-2}} \end{bmatrix} \in \mathbb{R}^{2^{M-1}\times M}. \tag{24}$$

It should be noted that, unlike constraint (13), in the aforementioned constraint (22), we allow the elements of matrix $\boldsymbol{D}$ to take negative values. With the constraint (16), the formulation in (18) can be retained.

Clearly, compared with the model in Section III-A, the dimensions of matrices and optimization variables have been reduced. For clarity, Table I compares the dimensions of various matrices before and after dimensionality reduction, for a generalized M. Ultimately, we obtain the following formulation for minimizing phase quantization bits:

$$\min_{\check{\boldsymbol{D}},\check{\boldsymbol{B}},\boldsymbol{q}} \ \mathbf{1}^T\boldsymbol{q} \tag{25a}$$

$$\text{s.t. } \Re\left[\check{\boldsymbol{c}}_M^T\check{\boldsymbol{D}}^T\boldsymbol{a}\left(\bar{\theta}_0\right)\right] = 1 \tag{25b}$$

$$\left|\check{\boldsymbol{c}}_M^T\check{\boldsymbol{D}}^T\boldsymbol{a}\left(\theta\right)\right| \le \rho\left(\theta\right),\ \theta \in \Theta \tag{25c}$$

$$-\eta \cdot \check{\boldsymbol{B}} \preceq \check{\boldsymbol{D}} \preceq \eta \cdot \check{\boldsymbol{B}} \tag{25d}$$

$$\check{\boldsymbol{B}}\mathbf{1} = \mathbf{1} \tag{25e}$$

$$\check{\boldsymbol{B}}^T\mathbf{1} \preceq N \cdot \check{\boldsymbol{F}}\boldsymbol{q} \tag{25f}$$

$$\boldsymbol{G}_M\boldsymbol{q} \preceq \mathbf{0} \tag{25g}$$

$$\check{\boldsymbol{B}} \in \{0,1\}^{N\times 2^{M-1}} \tag{25h}$$

$$\boldsymbol{q} \in \{0,1\}^M. \tag{25i}$$

The above formulation (25) is an MIP problem. Compared with (19), due to the approximately halved variable dimension, the solving of formulation (25) using an MIP solver is expected to be faster.

IV. EXTENSIONS AND FURTHER EXPLORATIONS

In Section III, we considered how to minimize the phase quantization bits for single focused-beam synthesis and discussed how to reduce the model dimension for accelerated solution finding. Notably, our proposed method can be extended to address a wider range of scenarios. Next, we present several extensions for our approach.

A. Minimizing Phase Quantization Bits for Phase-Only Reconfigurable Beampattern Synthesis

Let us first explore the minimization of phase quantization bits for phase-only reconfigurable beampattern synthesis. In phase-only reconfigurable arrays, the beam direction or beam shape can be changed solely by adjusting the excitation phase. In this case, we need to jointly design the common excitation amplitude and the distinct excitation phases corresponding to each beam under phase quantization, with the goal of minimizing the required number of phase quantization bits.

Taking the focused-beam scanning as an example and denoting the number of beams as K, we can formulate the problem as follows:

$$\min_{\{\boldsymbol{w}_k\}_{k=1}^K,\boldsymbol{\alpha},Q} \ Q \tag{26a}$$

$$\text{s.t. } \Re\left[\boldsymbol{w}_k^T\boldsymbol{a}\left(\bar{\theta}_k\right)\right] = 1\ \forall k \in \mathbb{K} \tag{26b}$$

$$\left|\boldsymbol{w}_k^T\boldsymbol{a}\left(\theta\right)\right| \le \rho_k\left(\theta\right),\ \theta \in \Theta_k\ \forall k \in \mathbb{K} \tag{26c}$$

$$\angle w_{k,n} \in \Psi_Q,\ n = 1,\ldots,N\ \forall k \in \mathbb{K} \tag{26d}$$

$$|\boldsymbol{w}_k| = \boldsymbol{\alpha}\ \forall k \in \mathbb{K} \tag{26e}$$

$$Q \in \mathbb{Z}_+ \tag{26f}$$

where the subscript k corresponds to the kth beam, and the set $\mathbb{K}$ is defined as $\mathbb{K} \triangleq \{1,2,\ldots,K\}$. We use $\bar{\theta}_k$, Θ_k, $\rho_k(\theta)$, and $\boldsymbol{w}_k$ to represent the mainlobe axis, the sidelobe region, the sidelobe level, and the excitation vector corresponding to the kth beam, respectively. Additionaly, $w_{k,n}$ in (26d) represents the nth entry of $\boldsymbol{w}_k$, $\boldsymbol{\alpha}$ in (26e) denotes the common excitation amplitude vector, and the set Ψ_Q is the candidate phase set for Q-bit phase quantization as defined in (3).

In the above problem (26), the optimization variables include the common excitation amplitude vector $\boldsymbol{\alpha}$, the excitation vectors $\{\boldsymbol{w}_k\}_{k=1}^K$ corresponding to each beam, and the phase quantization bit Q. It should be noted that the excitation amplitude vector $\boldsymbol{\alpha}$ can take continuous values, while the phase of the excitation vector and the phase quantization bit

Q can only take a finite number of discrete values. To address problem (26), we follow (11) and define the excitation vector of each beam as

$$\boldsymbol{w}_k = \boldsymbol{D}_k \bar{\boldsymbol{c}}_M,\ k = 1, \ldots, K \tag{27}$$

where $\boldsymbol{D}_k \in \mathbb{R}^{N \times 2^M}$ is the amplitude matrix corresponding to the kth beam and M is a predetermined number of quantization bits, similar to what we have done in (10).

Following the modeling approach in Section III-A, we can formulate the problem of minimizing phase quantization bits for phase-only reconfigurable beampattern synthesis as follows:

$$\min_{\{\boldsymbol{D}_k, \boldsymbol{B}_k\}_{k=1}^K, \boldsymbol{\alpha}, \boldsymbol{q}} \quad \mathbf{1}^T \boldsymbol{q} \tag{28a}$$

$$\text{s.t. } \Re\left[\bar{\boldsymbol{c}}_M^T \boldsymbol{D}_k^T \boldsymbol{a}\left(\bar{\theta}_k\right)\right] = 1\ \forall k \in \mathbb{K} \tag{28b}$$

$$\left|\bar{\boldsymbol{c}}_M^T \boldsymbol{D}_k^T \boldsymbol{a}\left(\theta\right)\right| \le \rho_k\left(\theta\right),\ \theta \in \Theta_k\ \forall k \in \mathbb{K} \tag{28c}$$

$$\mathbf{0} \le \boldsymbol{D}_k \le \eta \cdot \boldsymbol{B}_k\ \forall k \in \mathbb{K} \tag{28d}$$

$$\boldsymbol{D}_k \mathbf{1} = \boldsymbol{\alpha}\ \forall k \in \mathbb{K} \tag{28e}$$

$$\boldsymbol{B}_k \mathbf{1} = \mathbf{1}\ \forall k \in \mathbb{K} \tag{28f}$$

$$\boldsymbol{B}_k^T \mathbf{1} \le N \cdot \boldsymbol{F}\boldsymbol{q}\ \forall k \in \mathbb{K} \tag{28g}$$

$$\boldsymbol{G}_M \boldsymbol{q} \le \mathbf{0} \tag{28h}$$

$$\boldsymbol{B}_k \in \{0,1\}^{N \times 2^M}\ \ \forall k \in \mathbb{K} \tag{28i}$$

$$\boldsymbol{q} \in \{0,1\}^M. \tag{28j}$$

Clearly, the above formulation is an MIP problem. Note that in constraint (28e), the common excitation amplitude vector $\boldsymbol{\alpha}$ is modeled as a dummy variable to ensure that

$$\boldsymbol{D}_1 \mathbf{1} = \boldsymbol{D}_2 \mathbf{1} = \cdots = \boldsymbol{D}_K \mathbf{1}. \tag{29}$$

This guarantees that each beam has the same excitation amplitude, thereby satisfying phase-only control. Furthermore, it is evident that when considering a single beam (i.e., $K = 1$), the formulation (28) degrades to (19). In addition, for problem (28), we can similarly employ the method presented in Section III-B to reduce the dimension of optimization variables, thereby accelerating its solving using an MIP solver.

Remark 3: In model (28), we treat $\boldsymbol{\alpha}$ as an optimization variable primarily because we have explicitly limited the mainlobe gain of each beam [see constraint (28b)]. By setting $\boldsymbol{\alpha}$ as an optimization variable, we allow for the optimized design of antenna power, while permitting unequal power allocation across antennas. The vector $\boldsymbol{\alpha}$ obtained in model (28) represents the normalized amplitude coefficients. The actual power of each antenna can be obtained by scaling these coefficients accordingly. On the other hand, the proposed model can also be applied to scenarios with given antenna power. In this case, we use $\bar{\boldsymbol{\alpha}}$ to denote the predetermined power of each antenna. The beam gains should now be treated as variables. The constraints (28b), (28c), and (28e) should be modified as

$$\Re\left[\bar{\boldsymbol{c}}_M^T \boldsymbol{D}_k^T \boldsymbol{a}\left(\bar{\theta}_k\right)\right] = \beta_k\ \forall k \in \mathbb{K} \tag{30a}$$

$$\left|\bar{\boldsymbol{c}}_M^T \boldsymbol{D}_k^T \boldsymbol{a}\left(\theta\right)\right| \le \beta_k \cdot \rho_k\left(\theta\right),\ \theta \in \Theta_k\ \forall k \in \mathbb{K} \tag{30b}$$

$$\boldsymbol{D}_k \mathbf{1} = \bar{\boldsymbol{\alpha}}\ \forall k \in \mathbb{K} \tag{30c}$$

where β_k denotes the newly introduced optimization variable, with $k \in \mathbb{K}$.

B. Minimizing Amplitude Quantization Bits for Amplitude-Only Low-Sidelobe Synthesis

As the second extension, we now investigate the problem of minimizing the quantization bits for amplitude-only low-sidelobe synthesis. Unlike the phase-only scenario discussed in Section IV-A, here we design only the excitation amplitude to achieve the desired low-sidelobe beampattern with the minimal number of amplitude quantization bits.

Following (7), which is used to optimize the quantization bits of excitation phase, we can similarly formulate the problem for minimizing the quantization bits in the amplitude-only low-sidelobe synthesis scenario as follows:

$$\min_{\boldsymbol{w}, U} \quad U \tag{31a}$$

$$\text{s.t. } \Re\left[\boldsymbol{w}^T \boldsymbol{a}\left(\bar{\theta}_0\right)\right] = \beta \tag{31b}$$

$$\left|\boldsymbol{w}^T \boldsymbol{a}\left(\theta\right)\right| \le \beta \cdot \rho\left(\theta\right),\ \theta \in \Theta \tag{31c}$$

$$w_n \in \Upsilon_U,\ n = 1, \ldots, N \tag{31d}$$

$$U \in \mathbb{Z}_+. \tag{31e}$$

In (31), β stands for the desired beam gain at the mainlobe axis $\bar{\theta}_0$, and Υ_U represents the set of excitation amplitudes for U-bit amplitude quantization, as defined in (4).

To solve the problem (31), we predetermine a sufficiently large J, requiring only that it satisfies

$$U_* \le J \tag{32}$$

where U_* represents the optimal value for problem (31). Essentially, the role of J is similar to the role of M in (10). With the preassigned J, we can further enumerate all candidate excitation amplitudes that do not exceed J bits, thereby obtaining the following vector $\boldsymbol{z}$ for excitation amplitude candidates:

$$z = \left[\alpha_1^{(1)}, \alpha_2^{(1)}, \alpha_1^{(2)}, \alpha_2^{(2)}, \alpha_3^{(2)}, \alpha_4^{(2)}, \ldots, \alpha_1^{(J)}, \ldots, \alpha_{2^J}^{(J)}\right]^T.$$

It is easy to verify that $\boldsymbol{z}$ is a P-dimensional vector, where

$$P = 2^{J+1} - 2. \tag{33}$$

In addition, it should be pointed out that, unlike excitation phase, the excitation amplitude candidates for low-bit quantization may not be fully covered by those for high-bit quantization.

In the amplitude-only and amplitude quantization scenario, the weight vector coefficients can only be selected from the elements of $\boldsymbol{z}$. Naturally, we can express the weight vector $\boldsymbol{w}$ as follows:

$$\boldsymbol{w} = \boldsymbol{B}z \tag{34}$$

where $\boldsymbol{B}$ is an $N \times P$ binary matrix to be determined.

To select a unique excitation amplitude for each antenna, we constrain $\boldsymbol{B}$ to satisfy

$$\boldsymbol{B}\mathbf{1} = \mathbf{1}. \tag{35}$$

Furthermore, with the binary matrix $\boldsymbol{B}$, we can control the priority of excitation amplitude selection, thereby minimizing the number of quantization bits for excitation amplitude. Specifically, we first note that if l-bit amplitude quantization is used, all the one elements in $\boldsymbol{B}$ should appear between the

(2^l-1)th column and the $(2^{l+1}-2)$th column, $l=1,\ldots,J$. Then, similar to constraint (14) in the phase quantization scenario, we introduce a J-dimensional binary vector $\boldsymbol{h}$ and impose the following constraint:

$$\boldsymbol{B}^T\mathbf{1} \preceq N\cdot \boldsymbol{Y}\boldsymbol{h} \tag{36}$$

where the matrix $\boldsymbol{Y}$ is defined as

$$\boldsymbol{Y} \triangleq \begin{bmatrix} \mathbf{1}_2 & -\mathbf{1}_2 & & & & & \\ & \mathbf{1}_4 & -\mathbf{1}_4 & & & & \\ & & \mathbf{1}_8 & -\mathbf{1}_8 & & & \\ & & & \ddots & \ddots & & \\ & & & & \mathbf{1}_{2^{J-1}} & -\mathbf{1}_{2^{J-1}} & \\ & & & & & \mathbf{1}_{2^J} \end{bmatrix} \in \mathbb{R}^{P\times J}.$$

Similar to the vector $\boldsymbol{q}$ in (16), we impose the following monotone nonincreasing constraint on $\boldsymbol{h}$ in (36):

$$\boldsymbol{G}_J\boldsymbol{h} \preceq \mathbf{0}. \tag{37}$$

Under the above constraints (35)–(37), we can use the number of ones in $\boldsymbol{h}$ to represent the required number of amplitude quantization bits. In other words, we have

$$U = \mathbf{1}^T\boldsymbol{h}. \tag{38}$$

Taking the case where 3-bit amplitude quantization is required as an example, we have $\boldsymbol{h}=[1,1,1,0,\ldots,0]^T$ and $\boldsymbol{Y}\boldsymbol{h}=[0,0,0,0,0,0,1,1,1,1,1,1,1,1,0,\ldots,0]^T$. Recalling constraint (36), all the one elements of matrix $\boldsymbol{B}$ are located between the 7th and 14th columns, which exactly corresponds to 3-bit amplitude quantization.

Finally, the problem of minimizing amplitude quantization bits for amplitude-only low-sidelobe synthesis can be formulated as

$$\min_{\boldsymbol{B},\boldsymbol{h}} \ \mathbf{1}^T\boldsymbol{h} \tag{39a}$$
$$\text{s.t.} \ \left|z^T\boldsymbol{B}^T\boldsymbol{a}(\theta)\right| \le \rho\cdot\Re\left(z^T\boldsymbol{B}^T\boldsymbol{a}\left(\bar{\theta}_0\right)\right),\ \theta\in\Theta \tag{39b}$$
$$\boldsymbol{B}\mathbf{1}=\mathbf{1} \tag{39c}$$
$$\boldsymbol{B}^T\mathbf{1} \preceq N\cdot\boldsymbol{Y}\boldsymbol{h} \tag{39d}$$
$$\boldsymbol{G}_J\boldsymbol{h}\preceq\mathbf{0} \tag{39e}$$
$$\boldsymbol{B}\in\{0,1\}^{N\times P} \tag{39f}$$
$$\boldsymbol{h}\in\{0,1\}^J. \tag{39g}$$

It should be noted that, since all the candidates for excitation amplitude have already been given, we use constraint (39b) to automatically determine the beam gain at mainlobe axis $\bar{\theta}_0$. This provides more flexibility for minimizing the amplitude quantization bits. The problem (39) is an integer programming problem, which can also be solved using an MIP solver.

C. Minimizing Total Quantization Bits for Excitation Phase and Excitation Amplitude

In the aforementioned scenarios, we assumed the presence of quantization in either the excitation phase or excitation amplitude. In the third extension, we discuss how to minimize the total number of quantization bits when both excitation phase and excitation amplitude are quantized simultaneously.

Considering that the various excitation phase candidates and excitation amplitude candidates can be combined with each other, we define the following vector $\boldsymbol{v}$ to represent all possible candidate states for excitation:

$$\boldsymbol{v} \triangleq \bar{\boldsymbol{c}}_M \otimes z \in \mathbb{C}^V \tag{40}$$

where $V \triangleq P\cdot 2^M$, $\bar{\boldsymbol{c}}_M\in\mathbb{C}^{2^M}$ and $\boldsymbol{z}\in\mathbb{R}^P$ have been defined in Sections III-A and IV-B, respectively. By introducing an $N\times V$ binary matrix $\boldsymbol{B}$, we can express the weight vector as

$$\boldsymbol{w} = \boldsymbol{B}\boldsymbol{v}. \tag{41}$$

Similar to the previous discussion, to select an excitation state for each antenna, we add the following constraint to $\boldsymbol{B}$:

$$\boldsymbol{B}\mathbf{1}=\mathbf{1}. \tag{42}$$

Following the approach from Section III-A, we introduce an M-dimensional binary vector $\boldsymbol{q}$ to represent the number of phase quantization bits. The following constraints can thus be imposed:

$$\boldsymbol{B}^T\mathbf{1} \preceq N\cdot\boldsymbol{\Xi}\boldsymbol{q} \tag{43}$$
$$\boldsymbol{G}_M\boldsymbol{q}\preceq\mathbf{0} \tag{44}$$

where $\boldsymbol{\Xi}\triangleq\boldsymbol{F}\otimes\mathbf{1}_P$. It is easy to observe that the above constraint (43) is an extension of the previous constraint (14). The required number of phase quantization bits (denoted as Q) can be finally expressed as

$$Q = \mathbf{1}^T\boldsymbol{q}. \tag{45}$$

Before representing the number of amplitude quantization bits, we first re-express the vector $\boldsymbol{v}$ as

$$\boldsymbol{v} = \boldsymbol{E}(z\otimes\bar{\boldsymbol{c}}_M) \tag{46}$$

and rewrite the weight vector $\boldsymbol{w}$ as

$$\boldsymbol{w} = \boldsymbol{B}\boldsymbol{E}(z\otimes\bar{\boldsymbol{c}}_M). \tag{47}$$

In (46) and (47), $\boldsymbol{E}$ is a permutation matrix satisfying

$$\boldsymbol{E} = \sum_{j=1}^{2^M}\left(\boldsymbol{e}_j^T\otimes\boldsymbol{I}_P\otimes\boldsymbol{e}_j\right)\in\mathbb{R}^{V\times V} \tag{48}$$

where $\boldsymbol{e}_j$ is a 2^M-dimensional vector with its jth element being one and all other elements being zero, $j=1,\ldots,2^M$. Compared with (40), the vector $\boldsymbol{z}$ is placed before $\bar{\boldsymbol{c}}_M$ in (46). This facilitates our model for the required number of amplitude quantization bits. Specifically, we introduce a J-dimensional binary vector $\boldsymbol{h}$, and impose the following constraints:

$$\boldsymbol{E}^T\boldsymbol{B}^T\mathbf{1} \preceq N\cdot\boldsymbol{T}\boldsymbol{h} \tag{49}$$
$$\boldsymbol{G}_J\boldsymbol{h}\preceq\mathbf{0} \tag{50}$$

where $\boldsymbol{T}\triangleq\boldsymbol{Y}\otimes\mathbf{1}_{2^M}$. Note that the above constraint (49) can be seen as an extension to constraint (36). Ultimately, we can express the required number of amplitude quantization bits (denoted as U) as

$$U = \mathbf{1}^T\boldsymbol{h}. \tag{51}$$

Combining (45) and (51), the problem of minimizing the total number of quantized bits can be formulated as

$$\min_{\boldsymbol{B},\boldsymbol{q},\boldsymbol{h}} \ \mathbf{1}^T\boldsymbol{q}+\mathbf{1}^T\boldsymbol{h} \tag{52a}$$
$$\text{s.t.} \ \left|\boldsymbol{v}^T\boldsymbol{B}^T\boldsymbol{a}(\theta)\right| \le \rho\cdot\Re\left(\boldsymbol{v}^T\boldsymbol{B}^T\boldsymbol{a}\left(\bar{\theta}_0\right)\right),\ \theta\in\Theta \tag{52b}$$
$$\boldsymbol{B}\mathbf{1}=\mathbf{1} \tag{52c}$$
$$\boldsymbol{B}^T\mathbf{1}\le N\cdot\boldsymbol{\Xi}\boldsymbol{q} \tag{52d}$$
$$\boldsymbol{E}^T\boldsymbol{B}^T\mathbf{1}\le N\cdot\boldsymbol{T}\boldsymbol{h} \tag{52e}$$
$$\boldsymbol{G}_M\boldsymbol{q}\le\mathbf{0} \tag{52f}$$
$$\boldsymbol{G}_J\boldsymbol{h}\le\mathbf{0} \tag{52g}$$
$$\boldsymbol{B}\in\{0,1\}^{N\times V} \tag{52h}$$
$$\boldsymbol{q}\in\{0,1\}^{M} \tag{52i}$$
$$\boldsymbol{h}\in\{0,1\}^{J}. \tag{52j}$$

Similar to problem (39), the above problem (52) is an integer programming problem that can be solved using an MIP solver.

D. Joint Optimization of Quantization Bits and Sidelobe Level

In the previous discussions, our sole objective is to minimize the number of quantization bits for excitation amplitude or phase, under the given upper bound for sidelobe level. This results in the possibility that, although the number of quantization bits has been minimized, the resulting PSL may not be the optimal one. In this section, we take the focused-beam synthesis as an example and explore how to achieve the minimization of PSL while also minimizing the phase quantization bits. This presents a joint optimization problem, where we prioritize the minimization of phase quantization bits and subsequently consider the minimization of PSL.

For simplicity, we assume that the upper bound sidelobe level is uniform along the angle, and denote it as $\bar{\rho}$. Following the formulation in (19), we introduce a new variable s and re-express the constraint (19c) as

$$\left|\bar{\boldsymbol{c}}_M^T\boldsymbol{D}^T\boldsymbol{a}(\theta)\right|\le s,\ \theta\in\Theta \tag{53}$$
$$s\le\bar{\rho} \tag{54}$$

where s is used as an optimization variable to represent the obtained PSL. Compared with constraint (19c), although there seems to be no fundamental change, by introducing the variable s, we can achieve joint optimization of both phase quantization bits and PSL.

Specifically, we first note that within the framework of formulation (19), the variable s should be positive and typically less than the normalized mainlobe gain, which is taken as 1 as indicated by the constraint (19b). Concisely, we have

$$0<s<1. \tag{55}$$

In addition, the number of quantization bits $\mathbf{1}^T\boldsymbol{q}$ varies discretely with a minimal step size of one. Therefore, if we minimize the combined objective function defined as

$$\chi(\boldsymbol{q},s)\triangleq\mathbf{1}^T\boldsymbol{q}+s \tag{56}$$

we can achieve the lowest PSL while also minimizing the phase quantization bits. This is because the reduction in $\chi(\boldsymbol{q},s)$ due to changes in $\boldsymbol{q}$ is always greater than the variation caused by s. Thus, it allows us to prioritize minimizing the phase quantization bits. Moreover, to achieve a lower value on the objective function, s needs to be adjusted to its minimum value, which corresponds to the optimal PSL.

Ultimately, the problem of joint optimization for minimizing quantization bits and sidelobe level can be formulated as

$$\min_{\boldsymbol{D},\boldsymbol{B},\boldsymbol{q},s} \ \mathbf{1}^T\boldsymbol{q}+s \tag{57a}$$
$$\text{s.t.} \ \Re\left[\bar{\boldsymbol{c}}_M^T\boldsymbol{D}^T\boldsymbol{a}\left(\bar{\theta}_0\right)\right]=1 \tag{57b}$$
$$\left|\bar{\boldsymbol{c}}_M^T\boldsymbol{D}^T\boldsymbol{a}(\theta)\right|\le s,\ \theta\in\Theta \tag{57c}$$
$$s\le\bar{\rho} \tag{57d}$$
$$\mathbf{0}\le\boldsymbol{D}\le\eta\cdot\boldsymbol{B} \tag{57e}$$
$$\boldsymbol{B}\mathbf{1}=\mathbf{1} \tag{57f}$$
$$\boldsymbol{B}^T\mathbf{1}\le N\cdot\boldsymbol{F}\boldsymbol{q} \tag{57g}$$
$$\boldsymbol{G}_M\boldsymbol{q}\le\mathbf{0} \tag{57h}$$
$$\boldsymbol{B}\in\{0,1\}^{N\times 2^M} \tag{57i}$$
$$\boldsymbol{q}\in\{0,1\}^{M}. \tag{57j}$$

It is easy to see that the above formulation is an MIP problem, which can be solved using MIP solvers. It should be pointed out that the above approach is also applicable to other joint optimization scenarios, such as simultaneously minimizing amplitude quantization bits and sidelobe level.

E. Minimizing Quantization Bits for Shaped-Beam Synthesis

In the above studies, we solely considered the focused-beam scenario. In fact, the proposed approach can also achieve minimization of the number of quantized excitation bits for the shaped-beam synthesis scenario. Taking the phase quantization scenario as an example, it only requires modifying the mainlobe level constraint to

$$l(\theta)\le\left|\bar{\boldsymbol{c}}_M^T\boldsymbol{D}^T\boldsymbol{a}(\theta)\right|\le u(\theta),\ \theta\in\bar{\Theta} \tag{58}$$

where $\bar{\Theta}$ denotes the constrained interval for the shaped beam, and $l(\theta)$ and $u(\theta)$ denote the lower and upper bound level, respectively. Unlike the focused-beam scenario, the additional lower bound constraint on the beampattern amplitude (i.e., $l(\theta)\le|\bar{\boldsymbol{c}}_M^T\boldsymbol{D}^T\boldsymbol{a}(\theta)|$) is nonconvex. To simplify the analysis, we consider a discretized angle $\theta_i\in\bar{\Theta}$ and show how to reformulate the following constraint into a tractable form:

$$\left|\bar{\boldsymbol{c}}_M^T\boldsymbol{D}^T\boldsymbol{a}(\theta_i)\right|\ge l(\theta_i). \tag{59}$$

Geometrically, the above constraint means that the point $\boldsymbol{\nu}_i\triangleq[\Re(\bar{\boldsymbol{c}}_M^T\boldsymbol{D}^T\boldsymbol{a}(\theta_i))\ \Im(\bar{\boldsymbol{c}}_M^T\boldsymbol{D}^T\boldsymbol{a}(\theta_i))]^T$ must lie on or outside the circle of radius $l(\theta_i)$ in the complex plane. We consider a regular P-sided circumscribed polygon of the circle with radius $l(\theta_i)$. Approximately, we allow $\boldsymbol{\nu}_i$ to lie outside the circumscribed polygon, which leads to the following constraints:

$$\Re\left[\bar{\boldsymbol{c}}_M^T\boldsymbol{D}^T\boldsymbol{a}(\theta_i)\,e^{j2\pi\frac{p}{P}}\right]+\varsigma\cdot\boldsymbol{f}_i(p)\ge l(\theta_i),\ p=1,\ldots,P \tag{60}$$
$$\mathbf{1}^T\boldsymbol{f}_i\le P-1 \tag{61}$$

where ς is a sufficiently large positive constant, and $\boldsymbol{f}_i$ is a P-dimensional binary optimization variable, with $\boldsymbol{f}_i(p)$ denoting

its pth element. It can be observed that when $\boldsymbol{f}_i(p) = 0$, the inequality $\Re[\bar{\boldsymbol{c}}_M^T \boldsymbol{D}^T \boldsymbol{a}(\theta_i) e^{j2\pi(p/P)}] \geq l(\theta_i)$ is satisfied, demonstrating that $\boldsymbol{\nu}_i$ is located exterior to a polygon edge and then (59) is satisfied. Therefore, we constrain $\boldsymbol{f}_i$ to contain at least one zero element [i.e., constraint (61)], thereby guaranteeing the validity of constraint (59).

It can be observed that after the aforementioned approximate reformulation, we can obtain a tractable MIP model for the shaped-beam synthesis scenario. Moreover, by combining with the models established in Sections III and IV, we can achieve quantization bit minimization for different excitation quantization conditions.

Remark 4: In fact, the problem of minimizing the number of quantization bits for different pattern synthesis scenarios discussed in this article can also be solved using the exhaustive method. Specifically, the number of phase/amplitude quantization bits is initially specified based on the particular scenario, and then a generic pattern synthesis method is utilized to design the beampattern with the given quantization bit number. By validating whether the design outcomes satisfy the requirements and iteratively adjusting the number of quantization bits, an optimal solution with the minimal number of quantization bits can ultimately be identified. Evidently, this exhaustive search method necessitates repeated pattern synthesis designs, resulting in a longer design time. Different from the exhaustive method, the proposed method can find the minimal number of quantization bits in a single solution process.

V. Numerical Results

In this section, simulations are presented to demonstrate the effectiveness and superiority of the proposed method.[1] Unless otherwise specified, we set $\eta = 20$, $M = 6$ and $J = 5$ for the proposed method, and use a linear array comprising 20 uniformly spaced isotropic elements with a spacing of 0.5λ. Our comparative methods include the Chebyshev taper, Taylor taper, CP method in [20], SA method in [41], QIFT method in [37], accurate array response control (A^2RC) method in [26], sequential convex optimization (SCO) method in [18], iterative approximation (IA) method in [19], and the SF method in [31]. The proposed MIP models are solved by GUROBI [46]. The simulations are conducted using a computing platform, the processor is Intel[2] Core[3] i7-10750H CPU at 2.60 GHz.

A. Minimizing Phase Quantization Bits

In the first example, we consider the problem of minimizing phase quantization bits in pattern synthesis. For fair comparison, all benchmark methods incorporate excitation phase quantization with the same number of quantization bits as the proposed method.

1) Single Focused-Beam Pattern Synthesis: In the first case, we consider the pattern synthesis problem with a single focused beam, as discussed in Section III-A. The mainlobe axis is steered to $\bar{\theta}_0 = 20°$, and the mainlobe region is $[14°, 26°]$. We set the desired PSL to -20 dB. To minimize the phase quantization bits, we solve the proposed model (19) as well as the model (25) that considers dimensionality reduction. Simulation shows that the minimal phase quantization bits obtained from both models are $Q_* = 4$, although their excitation vectors are different. This is because there may be multiple sets of optimal solutions corresponding to the same phase quantization bits, as we have mentioned in Remark 2.

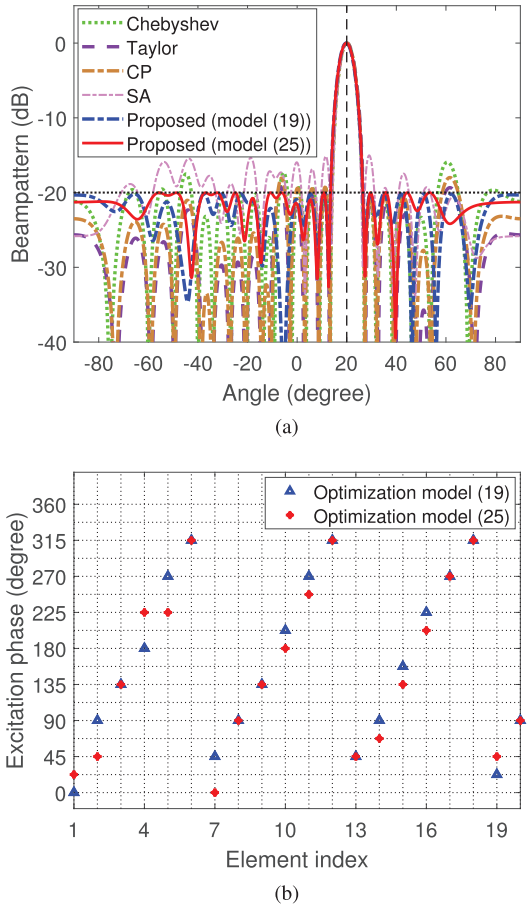


Fig. 5. Beampattern and excitation phase results in the scenario of single focused-beam pattern synthesis. (a) Beampattern comparison of different methods. (b) Excitation phase distribution of the proposed two models.

TABLE II
Resulting Excitation Amplitudes of the Proposed Two Models for Single Focused-Beam Pattern Synthesis

n	$\lvert w_n^{(1)}\rvert$	$\lvert w_n^{(2)}\rvert$	n	$\lvert w_n^{(1)}\rvert$	$\lvert w_n^{(2)}\rvert$
1	0.4890	0.3898	11	0.5969	0.7220
2	0.3207	0.3366	12	0.6656	0.6163
3	0.4174	0.4547	13	0.6571	0.5763
4	0.4825	0.3472	14	0.6029	0.5863
5	0.5676	0.5465	15	0.5799	0.5566
6	0.4757	0.5500	16	0.4709	0.4878
7	0.6095	0.4987	17	0.4159	0.4631
8	0.8061	0.7127	18	0.2795	0.4651
9	0.5231	0.7940	19	0.3804	0.3950
10	0.6667	0.5253	20	0.4188	0.3336

[1]The MATLAB codes for the proposed method are available online at https://zhangxuejing7.github.io/HomePage/

[2]Registered trademark.

[3]Trademarked.

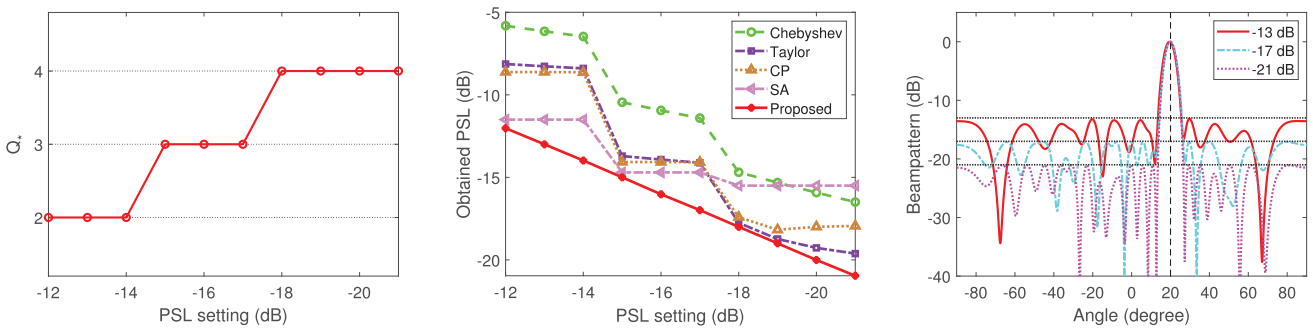


Fig. 6. Illustration of results under different PSL settings for single focused-beam pattern synthesis. (a) Curve of the obtained phase quantization bit versus PSL setting for the proposed method. (b) Curves of the obtained PSL versus PSL setting for different methods. (c) Beampattern results of the proposed method with different PSL settings.

TABLE III
COMPARISON OF PSL, DIRECTIVITY, AND HPBW FOR DIFFERENT ALGORITHMS IN FOCUSED-BEAM SYNTHESIS

	PSL (dB)	Directivity (dBi)	HPBW (deg)
Chebyshev	-15.92	12.797	5.8
Taylor	-19.27	12.783	6.0
CP	-17.99	12.818	5.9
SA	-15.03	12.487	5.5
Proposed (model (19))	-20.08	12.578	6.1
Proposed (model (25))	-20.01	12.470	6.1

TABLE IV
PSL AND RUNNING TIME OF THE EXHAUSTIVE SEARCH METHOD UNDER DIFFERENT PHASE QUANTIZATION BITS

	1-bit	2-bit	3-bit	4-bit
PSL (dB)	0	-14.49	-19.58	-21.11
Running Time (s)	8.25	49.94	138.91	677.74

Fig. 5(a) presents the radiation patterns of the two models. We can see that both models can achieve satisfactory radiation pattern results. Fig. 5(b) and Table II provide the resulting excitation phase and excitation amplitude of the proposed method, respectively. For the weight vectors obtained from the two models (denoted as $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$, respectively), we have $|\boldsymbol{w}_1^T\boldsymbol{a}(\bar{\theta}_0)| = 1.0201$ and $|\boldsymbol{w}_2^T\boldsymbol{a}(\bar{\theta}_0)| = 1.0008$, which confirms the prediction in Remark 1 It should be noted that solving model (25) (with a running time of 112 s) is faster than solving model (19) (with a running time of 568 s), which is consistent with theoretical prediction. The model (25) will be adopted in the subsequent simulations. Fig. 5(a) also presents the beampattern results of different methods. We have considered 4-bit phase quantization for all the presented results. Compared with the proposed method, the other four beampatterns exhibit quantized sidelobes, failing to meet the desired requirement on sidelobe level. To provide a comprehensive evaluation of the algorithm performance, Table III compares the PSL, directivity, and half-power beamwidth (HPBW) obtained by different algorithms. The results demonstrate the fundamental performance tradeoffs between these parameters.

In addition, to demonstrate the advantages of the proposed algorithm over the exhaustive search method mentioned in Remark 4, we minimize the PSL with different phase quantization bits. Table IV presents the achieved PSL values across various quantization levels, along with the running time required by the exhaustive approach. Notably, while the exhaustive method can obtain the same minimum quantization bits ($Q_* = 4$) as our algorithm, its total computation time (approximately 875 s) significantly exceeds that of the proposed method.

2) Result Comparison Under Different PSL Settings: To further assess the performance of the proposed method, we conduct simulations under different PSL settings, with other parameters remaining unchanged. Fig. 6(a) shows the obtained phase quantization bits Q_* of the proposed method with different PSL settings. It is evident that as the PSL decreases, the required number of phase quantization bits increases gradually. Accordingly, Fig. 6(b) compares the resulting PSLs obtained by different methods. We have considered phase quantization for the four comparison methods and have used the same number of quantization bits as the proposed one [shown in Fig. 6(a)]. As indicated by Fig. 6(b), all four comparison methods produce quantization sidelobes. In contrast, the proposed method can achieve satisfactory sidelobe performance when considering phase quantization. As an illustration, Fig. 6(c) depicts the beampattern results obtained by the proposed method when the PSL is set to -13, -17, and -21 dB, respectively. The effectiveness of the proposed method can thus be verified.

B. Minimizing Amplitude Quantization Bits for Amplitude-Only Low-Sidelobe Synthesis

In the second example, we consider the problem of minimizing amplitude quantization bits for amplitude-only low-sidelobe synthesis, as discussed in Section IV-B. We assume that the dynamic range of the discrete attenuator is 30 dB and set $J = 8$. The desired focused-beam is steered to 0° with a target PSL lower than -26 dB. The first null width of the mainlobe is 20°. For the benchmark methods (whose original excitation amplitudes may be continuous, such as Chebyshev and Taylor tapers), we apply additional amplitude

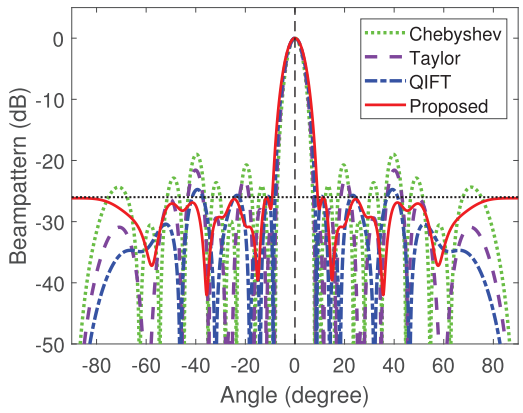


Fig. 7. Beampattern comparison of different methods for amplitude-only low-sidelobe synthesis.

TABLE V
RESULTING EXCITATION AMPLITUDES OF THE PROPOSED METHOD FOR AMPLITUDE-ONLY LOW-SIDELOBE SYNTHESIS

n	$\lvert w_n\rvert$ (dB)	n	$\lvert w_n\rvert$ (dB)	n	$\lvert w_n\rvert$ (dB)
1	-12.8571	8	0	15	-4.2857
2	-8.5714	9	0	16	-8.5714
3	-8.5714	10	0	17	-8.5714
4	-4.2857	11	0	18	-25.7143
5	-4.2857	12	0	19	-25.7143
6	-4.2857	13	0	20	-30
7	0	14	-4.2857		

TABLE VI
COMPARISON OF PSL, DIRECTIVITY, AND HPBW FOR DIFFERENT ALGORITHMS IN AMPLITUDE-ONLY LOW-SIDELOBE SYNTHESIS

	PSL (dB)	Directivity (dBi)	HPBW (deg)
Chebyshev	-18.99	12.611	6.0
Taylor	-21.58	12.419	6.4
QIFT	-24.74	12.097	6.8
Proposed	-26.02	11.749	7.4

quantization to their weight vectors, using the same number of quantization bits as in the proposed method.

With the given parameter settings, the proposed method yields an amplitude quantization of $U_* = 3$ bits. Fig. 7 compares the beampatterns obtained by different methods. It can be seen that the Chebyshev pattern, the Taylor pattern, and the QIFT method all produce varying degrees of quantized sidelobes, while the proposed method can effectively meet the given requirement on sidelobe level. Table V presents the excitation amplitudes obtained by the proposed method. Table VI provides a comparative analysis of PSL, directivity, and HPBW among different algorithms.

To further evaluate the performance of the proposed method under different PSL specifications, Fig. 8 shows the simulation results across multiple PSL conditions. Specifically, Fig. 8(a) illustrates the resulting amplitude quantization bits U_* obtained by the proposed method under different PSL settings. It is evident that a lower PSL requires more amplitude quantization bits. Fig. 8(b) compares the actual PSLs obtained by different

TABLE VII
ELEMENT LOCATIONS OF NONUNIFORM LINEAR ARRAY

n	$p_n(\lambda)$	n	$p_n(\lambda)$	n	$p_n(\lambda)$	n	$p_n(\lambda)$
1	0.00	4	1.13	7	2.35	10	3.61
2	0.49	5	1.59	8	2.86	11	4.03
3	0.77	6	2.03	9	3.27	12	4.53

methods when amplitude quantization is considered [with the same quantization bits as in Fig. 8(a)]. It can be observed that the obtained PSL matches the set PSL for the proposed method. As an illustration, Fig. 8(c) presents the beampattern results obtained by the proposed method under several different PSL settings.

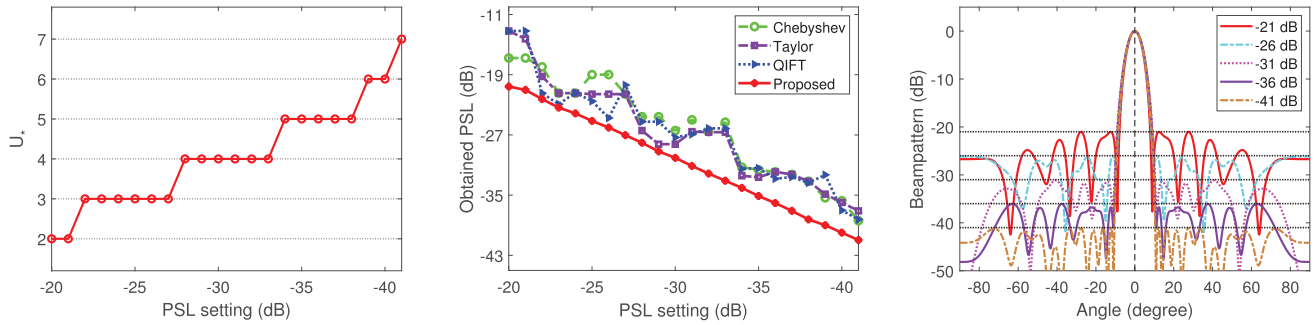
To demonstrate the applicability of the proposed algorithm to 2-D planar arrays, we consider a 5×5 uniform rectangular array with half-wavelength element spacing. Let $u = \sin(\theta_e)\cos(\theta_a)$ and $v = \sin(\theta_e)\sin(\theta_a)$, where θ_e and θ_a stand for elevation and azimuth angles, respectively. The beam steers to $(u_0, v_0) = (0, 0)$, and the desired PSL is set to -16 dB. The proposed algorithm achieves a minimum amplitude quantization of $U_* = 2$ bits. The resulting PSL is -16.14 dB, which satisfies the prescribed sidelobe requirement. The corresponding radiation beampattern and element-wise excitation amplitudes are presented in Fig. 9(a) and (b), respectively.

C. Minimizing Total Quantization Bits for Excitation Phase and Excitation Amplitude

In the third case, we consider minimizing the total number of quantization bits for excitation phase and excitation amplitude, as discussed in Section IV-C. In this scenario, we use a 12-element nonuniform linear array with element positions $\{p_n\}_{n=1}^{12}$ shown in Table VII. We assume that the dynamic range of the discrete attenuator is 50 dB. The beam is steered to 25°, with the desired sidelobes having nonuniform levels, as indicated by the dashed line in Fig. 10(a). By minimizing the total number of quantization bits, the proposed method results in $Q_* = 2$ and $U_* = 3$. In other words, 2-bit phase quantization and 3-bit amplitude quantization are sufficient to meet the given radiation requirements. Fig. 10(a) presents the radiation patterns obtained by different methods. Under the same number of quantization bits, CP method and A^{2}RC method raise the sidelobe levels due to excitation quantization. In addition, we can see that the A^{2}RC method lacks flexibility in controlling the mainlobe width. Finally, Fig. 10(b) shows the resulting excitation phase and excitation amplitude for the proposed method.

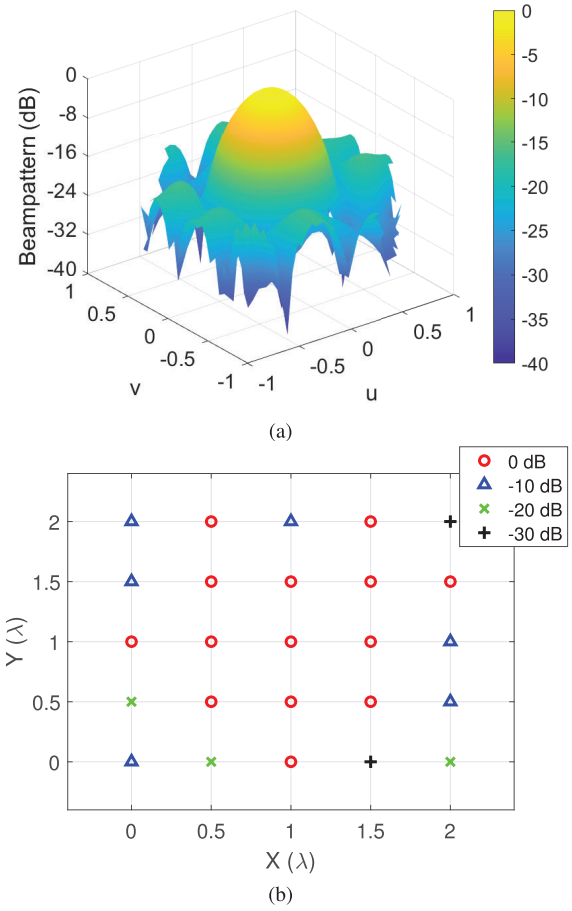
D. Joint Optimization of Phase Quantization Bits and Sidelobe Level

As an extension, in the fourth example we consider the joint optimization of phase quantization bits and sidelobe level, as elaborated in Section IV-D. Specifically, we sequentially consider pattern synthesis for a phase-only reconfigurable array, as discussed in Section IV-A, and the shaped-beam synthesis scenario mentioned in Section IV-E.



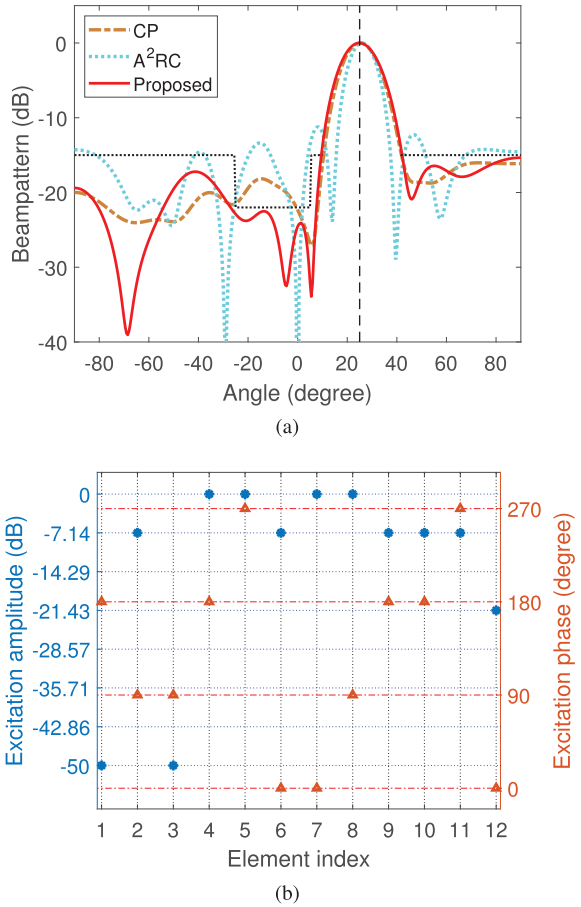
(a) (b) (c)

Fig. 8. Illustration of results under different PSL settings for amplitude-only low-sidelobe synthesis. (a) Curve of the obtained amplitude quantization bit versus PSL setting for the proposed method. (b) Curves of the obtained PSL versus PSL setting for different methods. (c) Beampattern results of the proposed method with different PSL settings.



(a) (b)

Fig. 9. Beampattern and excitation results of the proposed method for planar array in amplitude-only low-sidelobe synthesis. (a) Beampattern result. (b) Excitation amplitudes.



(a) (b)

Fig. 10. Beampattern and excitation results using quantized excitation phase and amplitude. (a) Beampattern comparison of different methods. (b) Excitation distribution of the proposed method.

1) Pattern Synthesis for Phase-Only Reconfigurable Array: In this case, we consider three beams steered to 0°, 10°, and 20°, with the beamwidth between the first nulls set to 16°. Setting the PSL for each scanning beam to −13 dB, our objective is to minimize the required number of phase quantization bits with the lowest PSL for the scanning beams. Simulation result reveals that we can achieve a PSL of −14.42 dB for the scanning beams with only 2-bit phase quantization. Fig. 11(a) illustrates the obtained phase-only reconfigurable patterns. The corresponding excitation phases for each beam and the common excitation amplitudes are presented in Fig. 11(b) and Table VIII, respectively.

2) Shaped-Beam Synthesis: In the scenario of shaped-beam synthesis, we assume that the mainlobe region is $[-15°, 15°]$,

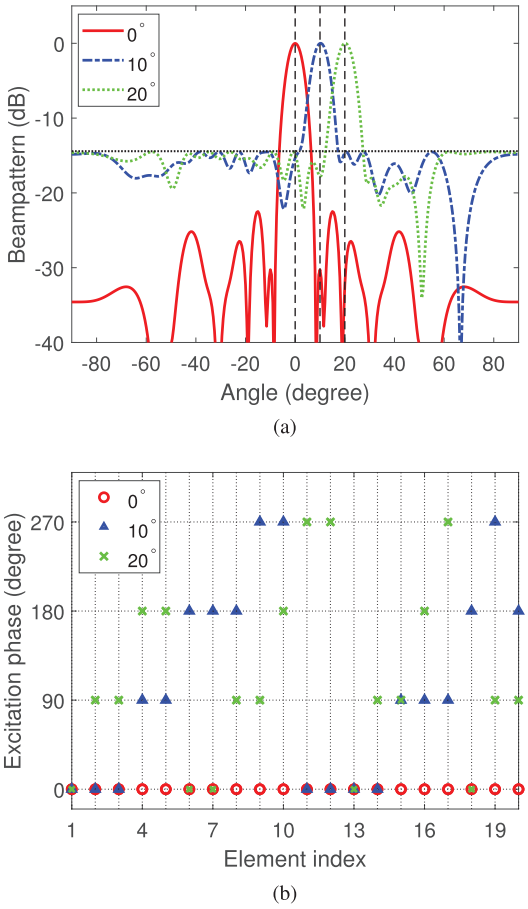


Fig. 11. Beampattern and excitation phase results in the scenario of phase-only beampattern synthesis. (a) Beampattern result of the proposed method. (b) Excitation phase distribution of the proposed method.

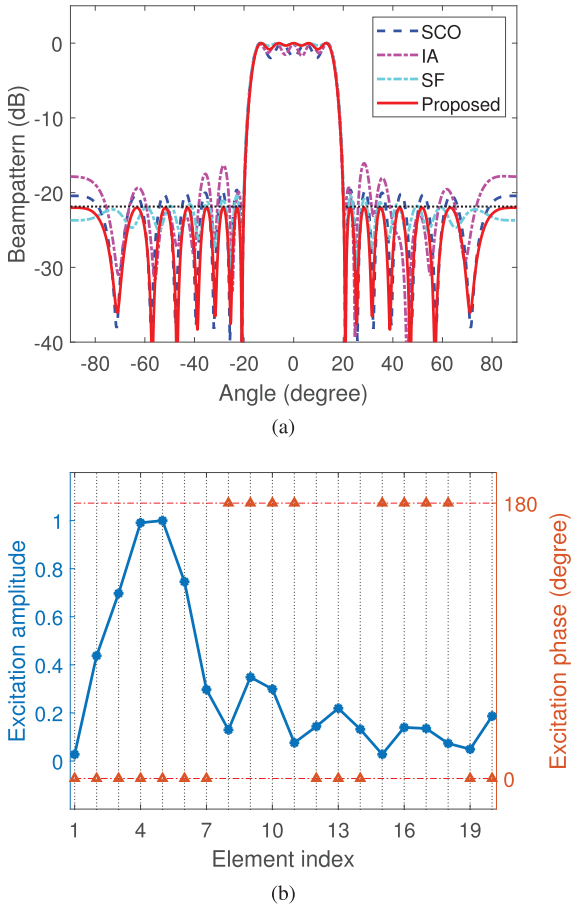


Fig. 12. Beampattern and excitation results in the scenario of shaped-beam synthesis. (a) Beampattern comparison of different methods. (b) Excitation distribution of the proposed method.

TABLE VIII

RESULTING COMMON EXCITATION AMPLITUDES OF THE PROPOSED METHOD FOR PHASE-ONLY BEAMPATTERN SYNTHESIS

n	$\lvert w_n\rvert$	n	$\lvert w_n\rvert$	n	$\lvert w_n\rvert$	n	$\lvert w_n\rvert$
1	0.4431	6	0.7582	11	0.8623	16	0.5042
2	0.5703	7	0.9298	12	0.5774	17	0.1694
3	0.7209	8	0.9525	13	0.7215	18	0.3494
4	0.3429	9	1.1384	14	0.5436	19	0.2648
5	0.7874	10	0.5651	15	0.5453	20	0.1414

TABLE IX

COMPARISON OF PSL, DIRECTIVITY, AND HPBW FOR DIFFERENT ALGORITHMS IN SHAPED-BEAM SYNTHESIS

	PSL (dB)	Directivity (dBi)	HPBW (deg)
SCO	-19.64	6.075	33.2
IA	-16.08	5.955	33.0
SF	-20.41	5.657	33.2
Proposed	-21.86	5.673	32.8

and the allowable ripple level in the mainlobe is 0.3 dB. Then, we constrain the PSL to be below −20 dB and aim to minimize both phase quantization bits and actual PSL simultaneously. For the proposed approach as presented in Section IV-E, we take $P = 8$. Simulation result shows that we can achieve the above requirements with 1-bit phase quantization ($Q_* = 1$), resulting in a PSL of −21.86 dB.

With the same phase quantization bit, Fig. 12(a) presents the shaped patterns obtained by different methods. We can observe that the SCO method, IA method, and SF method all produce quantized sidelobes of varying degrees. In addition, the SCO method and IA method exhibit poor ripple performance. For the proposed method, the resulting shaped beampattern is satisfactory. The excitation distribution of our method can be found in Fig. 12(b). To provide a comprehensive evaluation of the algorithm performance, Table IX compares the PSL, directivity, and HPBW obtained by different algorithms.

VI. CONCLUSION

In this article, we have presented a framework for minimizing the number of excitation quantization bits in pattern synthesis using MIP. The proposed method is versatile, applicable to different pattern synthesis scenarios and general array geometries. Moreover, our method can minimize the quantization bits for both excitation phase and excitation amplitude. Compared with existing methods that often predetermine quantization bits, the proposed method can determine the minimal excitation quantization bits required to meet specific beampattern performance. In addition, we have shown that the proposed method can be extended to joint optimization scenarios, balancing the need for optimal beampattern performance

with the minimization of excitation quantization bits. The optimality of our MIP model can be theoretically guaranteed, and its effectiveness has been demonstrated through simulations across various scenarios.

REFERENCES

[1] J.-J. Xiao and A. Nehorai, "Optimal polarized beampattern synthesis using a vector antenna array," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 576–587, Feb. 2009.

[2] G. Oliveri, L. Poli, and A. Massa, "Maximum efficiency beam synthesis of radiating planar arrays for wireless power transmission," *IEEE Trans. Antennas Propag.*, vol. 61, no. 5, pp. 2490–2499, May 2013.

[3] Y. Zhang, T. Osman, and A. Alkhateeb, "Online beam learning with interference nulling for millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 5109–5124, May 2024.

[4] C.-Y. Tseng and L. J. Griffiths, "A simple algorithm to achieve desired patterns for arbitrary arrays," *IEEE Trans. Signal Process.*, vol. 40, no. 11, pp. 2737–2746, Nov. 1992.

[5] W. Wei, G. Cui, X. Yu, R. Liu, and X. Wang, "Asymmetric beampattern synthesis for rectangular planar array via window function design," *IEEE Trans. Signal Process.*, vol. 72, pp. 400–414, 2024.

[6] R. Verdú-Monedero and J.-L. Gómez-Tornero, "On the use of the Z transform of LTI systems for the synthesis of steered beams and nulls in the radiation pattern of leaky-wave antenna arrays," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2275–2290, May 2019.

[7] P. Rocca, G. Oliveri, R. J. Mailloux, and A. Massa, "Unconventional phased array architectures and design methodologies—A review," *Proc. IEEE*, vol. 104, no. 3, pp. 544–560, Mar. 2016.

[8] C. L. Dolph, "A current distribution for broadside arrays which optimizes the relationship between beam width and side-lobe level," *Proc. IRE*, vol. 34, no. 6, pp. 335–348, Jun. 1946.

[9] H. Unz, "Linear arrays with arbitrarily distributed elements," *IRE Trans. Antennas Propag.*, vol. 8, no. 2, pp. 222–223, Mar. 1960.

[10] A. Koretz and B. Rafaely, "Dolph–Chebyshev beampattern design for spherical arrays," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2417–2420, Jun. 2009.

[11] K.-K. Yan and Y. Lu, "Sidelobe reduction in array-pattern synthesis using genetic algorithm," *IEEE Trans. Antennas Propag.*, vol. 45, no. 7, pp. 1117–1122, Jul. 1997.

[12] D. W. Boeringer and D. H. Werner, "Particle swarm optimization versus genetic algorithms for phased array synthesis," *IEEE Trans. Antennas Propag.*, vol. 52, no. 3, pp. 771–779, Mar. 2004.

[13] V. Murino, A. Trucco, and C. S. Regazzoni, "Synthesis of unequally spaced arrays by simulated annealing," *IEEE Trans. Signal Process.*, vol. 44, no. 1, pp. 119–122, Jan. 1996.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[15] J. Yang, J. Lin, Q. Shi, and Q. Li, "An ADMM-based approach to robust array pattern synthesis," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 898–902, Jun. 2019.

[16] K. Tang and X. Zhang, "Auto-determining mainlobe width for beampattern synthesis via relaxation optimization," *IEEE Signal Process. Lett.*, vol. 29, pp. 314–318, 2022.

[17] S. E. Nai, W. Ser, Z. L. Yu, and H. Chen, "Beampattern synthesis for linear and planar arrays with antenna selection by convex optimization," *IEEE Trans. Antennas Propag.*, vol. 58, no. 12, pp. 3923–3930, Dec. 2010.

[18] L. Li, X. Zhang, and Z. He, "Beampattern synthesis with dynamic range ratio constraint via sequential convex optimization," *IEEE Antennas Wireless Propag. Lett.*, vol. 23, pp. 3227–3231, 2024.

[19] S. Lei et al., "Power-gain pattern synthesis of array antenna with dynamic range ratio restriction," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, pp. 2691–2695, 2019.

[20] H. Lebret and S. Boyd, "Antenna array pattern synthesis via convex optimization," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 526–532, Mar. 1997.

[21] B. Fuchs and S. Rondineau, "Array pattern synthesis with excitation control via norm minimization," *IEEE Trans. Antennas Propag.*, vol. 64, no. 10, pp. 4228–4234, Oct. 2016.

[22] F. Wang, V. Balakrishnan, P. Y. Zhou, J. J. Chen, R. Yang, and C. Frank, "Optimal array pattern synthesis using semidefinite programming," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1172–1183, May 2003.

[23] B. Fuchs, "Application of convex relaxation to array synthesis problems," *IEEE Trans. Antennas Propag.*, vol. 62, no. 2, pp. 634–640, Feb. 2014.

[24] Z.-Q. Luo, W.-K. Ma, A. M. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 20–34, May 2010.

[25] P. Y. Zhou and M. A. Ingram, "Pattern synthesis for arbitrary arrays using an adaptive array method," *IEEE Trans. Antennas Propag.*, vol. 47, no. 5, pp. 862–869, May 1999.

[26] X. Zhang, Z. He, B. Liao, X. Zhang, Z. Cheng, and Y. Lu, "A^{2}RC: An accurate array response control algorithm for pattern synthesis," *IEEE Trans. Signal Process.*, vol. 65, pp. 1810–1824, 2017.

[27] X. Zhang, Z. He, B. Liao, X. Zhang, and W. Peng, "Pattern synthesis for arbitrary arrays via weight vector orthogonal decomposition," *IEEE Trans. Signal Process.*, vol. 66, pp. 1286–1299, 2018.

[28] X. Zhang, Z. He, B. Liao, Y. Yang, J. Zhang, and X. Zhang, "Flexible array response control via oblique projection," *IEEE Trans. Signal Process.*, vol. 67, no. 12, pp. 3126–3139, Jun. 2019.

[29] J.-Y. Li, Y.-X. Qi, and S.-G. Zhou, "Shaped beam synthesis based on superposition principle and Taylor method," *IEEE Trans. Antennas Propag.*, vol. 65, no. 11, pp. 6157–6160, Nov. 2017.

[30] Y. Li, K. C. Ho, and C. Kwan, "3-D array pattern synthesis with frequency invariant property for concentric ring array," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 780–784, Feb. 2006.

[31] S.-P. Wu, S. Boyd, and L. Vandenberghe, "FIR filter design via semidefinite programming and spectral factorization," in *Proc. 35th IEEE Conf. Decis. Control*, vol. 1, Dec. 1996, pp. 271–276.

[32] M. Clenet and G. A. Morin, "Visualization CF radiation-pattern characteristics of phased arrays using digital phase shifters," *IEEE Antennas Propag. Mag.*, vol. 45, no. 2, pp. 20–35, Apr. 2003.

[33] D. Xia and X. Zhang, "Synthesis of sparse arrays with discrete phase constraints via mixed-integer programming," *IEEE Antennas Wireless Propag. Lett.*, vol. 23, pp. 1030–1034, 2024.

[34] S. Lee and H.-J. Song, "Accurate statistical model of radiation patterns in analog beamforming including random error, quantization error, and mutual coupling," *IEEE Trans. Antennas Propag.*, vol. 69, no. 7, pp. 3886–3898, Jul. 2021.

[35] T. H. Ismail and Z. M. Hamici, "Array pattern synthesis using digital phase control by quantized particle swarm optimization," *IEEE Trans. Antennas Propag.*, vol. 58, no. 6, pp. 2142–2145, Jun. 2010.

[36] S. Goudos, "Antenna design using binary differential evolution: Application to discrete-valued design problems," *IEEE Antennas Propag. Mag.*, vol. 59, no. 1, pp. 74–93, Feb. 2017.

[37] W. P. M. N. Keizer, "Low sidelobe phased array pattern synthesis with compensation for errors due to quantized tapering," *IEEE Trans. Antennas Propag.*, vol. 59, no. 12, pp. 4520–4524, Dec. 2011.

[38] M. Smith and Y. Guo, "A comparison of methods for randomizing phase quantization errors in phased arrays," *IEEE Trans. Antennas Propag.*, vol. AP-31, no. 6, pp. 821–828, Nov. 1983.

[39] W. Jiang, Y. Guo, T. Liu, W. Shen, and W. Cao, "Comparison of random phasing methods for reducing beam pointing errors in phased array," *IEEE Trans. Antennas Propag.*, vol. 51, no. 4, pp. 782–787, Apr. 2003.

[40] Y. Zhang, H. Sun, and Q. H. Liu, "Integer linear programming roundoff method for arbitrary planar phased-array scanning," *IEEE Trans. Antennas Propag.*, vol. 67, no. 5, pp. 3040–3047, May 2019.

[41] D. Thompson, M. Yeary, C. Fulton, and B. McGuire, "Optimized beam steering approach for improved sidelobes in phased array radars using a minimal number of control bits," *IEEE Trans. Antennas Propag.*, vol. 63, no. 1, pp. 106–112, Jan. 2015.

[42] B. C. Brock, "The application of Taylor weighting, digital phase shifters, and digital attenuators to phased-array antennas," Sandia National Laboratories, Albuquerque, NM, USA, and Livermore, CA, USA, Sandia Rep. SAND2008-1687, 2008.

[43] S. Bourguignon, J. Ninin, H. Carfantan, and M. Mongeau, "Exact sparse approximation problems via mixed-integer programming: Formulations and computational performance," *IEEE Trans. Signal Process.*, vol. 64, no. 6, pp. 1405–1419, Mar. 2016.

[44] J. I. Echeveste, M. Á. G. de Aza, J. Rubio, and J. Zapata, "Synthesis of coupled antenna arrays using digital phase control via integer programming," *IET Microw., Antennas Propag.*, vol. 12, no. 6, pp. 999–1003, May 2018.

[45] K. Miao, Y. Zhang, C. Yao, and H. Sun, "Sidelobe suppression by phase-only tapering with integer linear programming," in *Proc. CIE Int. Conf. Radar (Radar)*, Dec. 2021, pp. 1989–1992.

[46] *Gurobi Optimizer Reference Manual*, Gurobi Optimization, Briarwood, TX, USA, 2023.

[47] *IBM ILOG CPLEX Optimization Studio CPLEX Users Manual*, IBM, Raleigh, NC, USA, 2018.