SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2026, Vol. 69, Iss. 4, 142302:1–142302:16
https://doi.org/10.1007/s11432-025-4666-y

Beam-structured precoding for network massive MIMO systems via Hamiltonian-based optimization

Wen-Jie ZHU[1,2], Yuxuan ZHANG[1,2], Ziyu XIANG[1,2], Ding SHI[2], Li YOU[1,2] & Xiqi GAO[1,2*]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China*
[2]*Purple Mountain Laboratories, Nanjing 211100, China*

Received 8 May 2025/Revised 19 August 2025/Accepted 4 November 2025/Published online 2 March 2026

Abstract Massive multiple-input multiple-output (MIMO) has received widespread recognition for its substantially improved spectral efficiency, and it remains a fundamental technology in future wireless communication networks. Its extension to network massive MIMO enables joint transmission across base stations (BSs), but also introduces significant challenges due to the high dimension of the channel matrices and the associated optimization variables. Our work investigates the precoder design by leveraging beam-structured precoding under a Hamiltonian-based framework. We begin by introducing a beam-based channel model and formulating the precoder design problem in the beam domain. Then we show that the optimal beam-domain precoder for each user terminal (UT) only occupies beams corresponding to its non-zero beam-domain channel elements, a design referred to as beam-structured precoding, which results in a lower-dimensional optimization problem. The corresponding problem is handled using a Hamiltonian system, where the objective function is interpreted as potential energy, transforming the optimization problem into a physical system's energy minimization task. The system's dynamical equations are then solved numerically with a RATTLE integrator, providing a principled approach to explore the solution space, with reduced computational complexity and favorable performance. Through simulation results, we verify the effectiveness of our method by demonstrating notable complexity savings while maintaining high performance.

Keywords network massive MIMO transmission, beam-structured precoding, sum-rate maximization, Hamiltonian-based optimization

Citation Zhu W-J, Zhang Y X, Xiang Z Y, et al. Beam-structured precoding for network massive MIMO systems via Hamiltonian-based optimization. Sci China Inf Sci, 2026, 69(4): 142302, https://doi.org/10.1007/s11432-025-4666-y

1 Introduction

The significant impact of massive multiple-input multiple-output (MIMO) technology has reshaped modern wireless networks. By deploying large-scale antennas at base stations (BSs), massive MIMO enables dense user connectivity through spatial multiplexing, offering substantial gains in spectral and energy efficiency [1,2]. Massive MIMO has become fundamental in addressing the growing demand for ultra-high data rates [3]. To further improve overall system performance, network massive MIMO has attracted increasing attention, where the BSs jointly transmit data through coordinated signal processing. Unlike traditional independent BS operation, which results in severe inter-cell interference, network massive MIMO coordinates antennas from all the BSs as a unified virtual array [4]. Such coordination exploits spatial multiplexing gains to constructively combine desired signals while suppressing interference. As a result, inter-cell interference is transformed into a controllable resource [5,6], unlocking additional degrees of freedom and enhancing spectral efficiency and link reliability, particularly for cell-edge users [7,8].

As the demand for higher data rates and better efficiency increases, achieving high spectral efficiency in network massive MIMO systems has become a primary design challenge. This requires solving large-scale optimization problems that balance the enhancement of desired signals with the suppression of inter-cell interference, for which sum-rate maximization serves as a key objective [9]. The weighted sum mean-squared-error (WMMSE) framework has become one of the most widely used approaches for tackling this optimization problem, and has demonstrated effectiveness in various scenarios [10–13]. However, its extension to massive MIMO systems is often limited by computational complexity and slow convergence, especially as network size increases [14–16]. Future wireless networks are expected to operate at higher frequencies while also supporting a growing number of BS antennas and served users. In such scenarios, leveraging the inherent sparsity of the channel becomes a promising direction for designing efficient precoding strategies, as it can reduce the effective dimension of the optimization problem.

* Corresponding author (email: xqgao@seu.edu.cn)

Despite this reduction in the dimension, the resulting precoding optimization problem remains high-dimensional and non-convex, posing significant challenges for efficient solution. Hamiltonian-based optimization provides a powerful approach to tackling such problems effectively. By interpreting the objective function as the system's potential energy, it transforms the optimization task into a minimum energy problem within a dynamical framework. The optimization process is guided by the system's dynamical equations, which describe how the system evolves, ultimately converging to an optimal configuration [17,18]. Unlike traditional methods, Hamiltonian-based optimization does not rely on initial values for the solution [19]. This robustness stems from the introduction of a kinetic term, which provides inertia to the system, enabling it to move smoothly through the search space. As a result, the system can traverse regions with small gradients, reducing the likelihood of premature convergence to local minima and improving its ability to escape suboptimal solutions, even when starting from less favorable initial points. To solve the system's dynamical equations and obtain the optimal solution, numerical methods are employed. These methods provide discrete updates, effectively guiding the optimization process toward the optimum [20–22]. The authors of [23] demonstrated that in the iterative precoding process based on minorize-maximization, the Hamiltonian framework can effectively eliminate the computational burden of high-dimensional matrix inversion. However, the full potential of Hamiltonian-based optimization for solving precoding problems, especially in large-scale and complex network settings, remains to be further investigated.

In this paper, we propose a beam-structured precoder design approach for network massive MIMO systems using Hamiltonian-based optimization. The main contributions of this paper are summarized as follows.

- A beam-based channel model is formulated to capture the spatial characteristics of network massive MIMO systems. Based on this model, the precoder design problem is reformulated in the beam domain. It is further demonstrated that, for each UT, the optimal beam-domain precoder can focus on the beams corresponding to the non-zero elements of its own beam-domain channel, resulting in a beam-structured precoding (BSP) that exploits the sparsity to reduce the optimization dimension.
- To solve the resulting BSP problem, we adopt a Hamiltonian-based optimization approach, modeling the optimization process as a dynamical system whose motion equations guide the solution trajectory. The evolution of the system is computed with a numerical integrator, which ensures stable convergence behavior and effective progression toward a well-optimized precoder.
- Simulation results validate that the proposed BSP approach achieves comparable sum-rate performance to spatial-domain precoding, while significantly reducing computational complexity and maintaining its effectiveness under finite antenna settings.

The rest of this paper is organized as follows. Section 2 introduces the beam-based channel model for network massive MIMO systems. In Section 3, we demonstrate how the optimal sum-rate maximizing precoder can be derived in the beam domain, which leads to the BSP design. Building on this formulation, Section 4 presents a Hamiltonian-based approach to solving the BSP problem, where the system's dynamical equations guide the optimization process. Section 5 presents simulation results, followed by the conclusion in Section 6.

We use the following notations throughout the paper. Vectors and matrices are, respectively, defined as bold lowercase and uppercase letters. Superscripts $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ stand for conjugate, transpose and conjugate transpose, respectively. $\mathbf{I}_M$ represents the size-M identity matrix; $\mathbf{0}_{M\times N}$ denotes the $M \times N$ all-zero matrix. $\bar{\jmath} = \sqrt{-1}$. $\mathbb{C}^{M\times N}$ means the $M \times N$-dimensional complex vector space. The notation $\triangleq$ is used for definitions. Let $\mathcal{CN}(\mu, \sigma^2)$ denote the circular symmetric complex Gaussian distribution with mean μ and standard deviation σ. Let $[\mathbf{A}]_{m,n}$ and $[\mathbf{A}]_{m,:}$ denote the (m, n)-th entry and the m-th row of matrix $\mathbf{A}$, where the element indices start from 1. $[\mathbf{a}]_m$ and $[\mathbf{a}]_{\mathcal{B}}$ denote m-th element and elements with index set $\mathcal{B}$ of vector $\mathbf{a}$, respectively. $|\mathcal{B}|$ denotes the cardinality of set $\mathcal{B}$. $[\mathcal{B}]_n$ represents the n-th number of $\mathcal{B}$ in ascending order, counted from 1. $\mathrm{bdiag}(\mathbf{A}_1, \ldots, \mathbf{A}_n)$ means a block diagonal matrix with $\mathbf{A}_1, \ldots, \mathbf{A}_n$ on the diagonal. $\mathbb{E}\{\cdot\}$ evaluates the ensemble expectation of the input parameter. The derivative with respect to time is described via the dot notation as $\frac{d}{dt}\mathbf{a} = \dot{\mathbf{a}}$. $\odot$ and $\otimes$ denote the Hadamard and Kronecker product operators, respectively. Let $\|\mathbf{a}\|$ denote the 2-norm of vector $\mathbf{a}$. The cardinality of set $\mathcal{B}$ is defined as $|\mathcal{B}|$.

2 System model

Consider the problem of downlink transmission within a network massive MIMO system. The system comprises K single-antenna user terminals (UTs) and L BSs, each equipped with an M-antenna uniform linear array (ULA). Throughout the paper, we use k, u to indicate the user index and use ℓ, r for the BS index. The BSs jointly transmit the signals to users, where user data and instantaneous channel state information (CSI), assumed to be known at the BSs, are shared globally via high-speed and error-free backhauls. The overall system configuration is illustrated

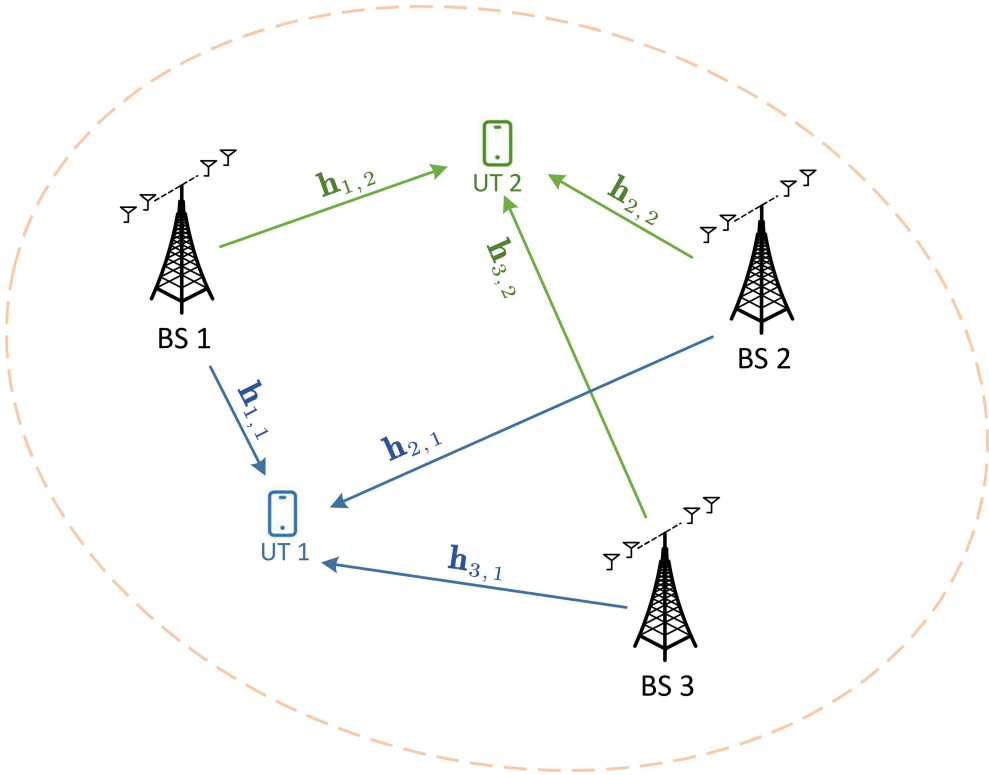

Figure 1 (Color online) Illustration of a downlink network massive MIMO system.

in Figure 1.

For a given direction with the azimuth and elevation angles of departure (AoDs) defined as ϕ^{az} and ϕ^{el}, the directional cosine is $\Omega \triangleq \sin\phi^{\mathrm{az}}\cos\phi^{\mathrm{el}}$. With a half-wavelength antenna spacing, the steering vector pointing towards Ω is [24]

$$\mathbf{v}(\Omega) = \frac{1}{\sqrt{M}}\left[1, e^{-\bar{\jmath}\pi\Omega}, \ldots, e^{-\bar{\jmath}\pi(M-1)\Omega}\right]^{\mathrm{T}} \in \mathbb{C}^{M\times 1}. \tag{1}$$

Suppose that there are $P_{k,\ell}$ physical paths from the ℓ-th BS to the k-th user, and the p-th path has a complex-valued coefficient of $\alpha_{k,\ell,p}$. The channel frequency response between the ℓ-th BS and the k-th user is

$$\mathbf{h}_{k,\ell} = \sqrt{M}\sum_{p=1}^{P_{k,\ell}} \alpha_{k,\ell,p}\mathbf{v}(\Omega_{k,\ell,p}). \tag{2}$$

In the physical channel model (2), the AoD corresponding to each path is arbitrary, resulting in infinitely many possible directional cosines $\Omega_{k,\ell,p}$. Thus, it is difficult to capture the precise coefficient and directional cosine [25]. To counter this, we divide the domain $\mathcal{D} = \{\Omega|\Omega \in (-1,1]\}$ of the directional cosine into disjoint subjects as $\mathcal{D} = \bigcup_{n=1}^{N}\Lambda_n$. With uniform division, each subset is written as $\Lambda_n = \left[\frac{2(n-1)-N}{N}, \frac{2n-N}{N}\right)$, where $N = FM$, and $F \geqslant 1$ is defined as the fine factor. The directional cosines in set Λ_n are approximated by $\Omega_n = \frac{2(n-1)-N}{N}$, which tends to be asymptotically accurate with an infinitely large number of divided subsets. Let $\mathcal{P}_{k,\ell} = \{\Omega_{k,\ell,1}, \ldots, \Omega_{k,\ell,P_{k,\ell}}\}$ denote the path parameter set from the ℓ-th BS to the k-th user. With the sampled steering vectors pointing towards $\Omega_n, \forall n$, the channel in (2) can be approximated by a beam-based channel model [26, 27] as

$$\mathbf{h}_{k,\ell} = \sum_{n=1}^{N}\tilde{h}_{k,\ell,n}\mathbf{v}(\Omega_n), \tag{3}$$

where $\tilde{h}_{k,\ell,n} = \sqrt{M}\sum_{\Omega_{k,\ell,p}\in\mathcal{P}_{k,\ell}\cap\Lambda_n}\alpha_{k,\ell,p}$ is the beam-domain channel coefficient. Define the beam matrix that consists of sampled steering vectors as

$$\mathbf{V} = [\mathbf{v}(\Omega_1), \ldots, \mathbf{v}(\Omega_N)] \in \mathbb{C}^{M\times N}. \tag{4}$$

Then, the beam-based channel model in (3) can be rewritten in a compact form as

$$\mathbf{h}_{k,\ell} = \mathbf{V}\tilde{\mathbf{h}}_{k,\ell}, \tag{5}$$

where $\tilde{\mathbf{h}}_{k,\ell} \in \mathbb{C}^{N\times 1}$ is the beam-domain channel vector, expressed as

$$\tilde{\mathbf{h}}_{k,\ell} = \left[\tilde{h}_{k,\ell,1}, \ldots, \tilde{h}_{k,\ell,N}\right]^{\mathrm{T}}. \tag{6}$$

The beam matrix $\mathbf{V}$ in (4) incorporates a fine factor F to control the number of sampled steering vectors. The conventional discrete Fourier transform (DFT)-based channel model corresponds to the case $F = 1$. When $F > 1$, the angular domain is sampled more densely, resulting in improved spatial resolution [25, 28]. Increasing the fine factor mitigates the dispersive effect presented in the DFT-based channel model, thereby better preserving the sparsity in the beam domain [29]. Therefore, an appropriate choice of fine factor provides a favorable trade-off between complexity and accuracy.

The received signal y_k of the k-th UT is expressed as

$$y_k = \sum_{\ell} \mathbf{h}_{k,\ell}^{\mathrm{H}}\mathbf{p}_{k,\ell}d_k + \sum_{\ell}\sum_{u\neq k} \mathbf{h}_{k,\ell}^{\mathrm{H}}\mathbf{p}_{u,\ell}d_u + n_k, \tag{7}$$

where $\mathbf{p}_{k,\ell} \in \mathbb{C}^{M\times 1}$ is the precoding vector for the k-th user from the ℓ-th BS, satisfying power budget constraint $\sum_{\ell} \mathbf{p}_{k,\ell}^{\mathrm{H}}\mathbf{p}_{k,\ell} \leqslant P_{\ell}$, d_k with $\mathbb{E}_d\{d_k d_k^*\} = 1$ is the data stream intended for the k-th user, jointly transmitted by the BSs, and $n_k \sim \mathcal{CN}(0, \sigma^2)$ is the circularly symmetric complex Gaussian noise. Define the collective precoding vector for the k-th user as the concatenation of precoding vectors from all the BSs, given by $\mathbf{p}_k = \left[\mathbf{p}_{k,1}^{\mathrm{T}}, \ldots, \mathbf{p}_{k,L}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{ML}$. Similarly, stacking the downlink channel vectors from all the BSs to the k-th user yields the collective channel vector $\mathbf{h}_k = [\mathbf{h}_{k,1}^{\mathrm{T}}, \ldots, \mathbf{h}_{k,L}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{ML}$. Then, the received signal in (7) can be rewritten as

$$y_k = \mathbf{h}_k^{\mathrm{H}}\mathbf{p}_k d_k + \sum_{u\neq k} \mathbf{h}_k^{\mathrm{H}}\mathbf{p}_u d_u + n_k. \tag{8}$$

Considering the sparsity of the massive MIMO channel, we pick out the non-zero elements, whose indices constitute the set $\mathcal{N}_{k,\ell} = \{n|[\tilde{\mathbf{h}}_{k,\ell}]_n \neq 0\}$ in ascending order, with cardinality $N_{k,\ell} = |\mathcal{N}_{k,\ell}|$. For each BS, define the set comprising indices of non-zero elements of all the users in ascending order as $\mathcal{N}_{\ell} = \bigcup_k \mathcal{N}_{k,\ell}$, whose cardinality is $\hat{N}_{\ell} = |\mathcal{N}_{\ell}|$. Define the number of non-zero beams from all the BSs as $\hat{N} = \sum_{\ell} \hat{N}_{\ell}$. Thus, the collective channel vectors $\{\mathbf{h}_k\}$ can be expressed in the form of the beam-based channel model as

$$\mathbf{h}_k = \hat{\mathbf{V}}\hat{\mathbf{h}}_k, \forall k, \tag{9}$$

where

$$\hat{\mathbf{V}} = \mathrm{bdiag}(\hat{\mathbf{V}}_1, \ldots, \hat{\mathbf{V}}_L) \in \mathbb{C}^{ML\times\hat{N}}, \tag{10}$$

$$\hat{\mathbf{V}}_{\ell} = [\mathbf{V}]_{:,\mathcal{N}_{\ell}} \in \mathbb{C}^{M\times\hat{N}_{\ell}}, \forall \ell, \tag{11}$$

$$\hat{\mathbf{h}}_k = [\hat{\mathbf{h}}_{k,1}^{\mathrm{T}}, \ldots, \hat{\mathbf{h}}_{k,L}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{C}^{\hat{N}\times 1}, \forall k, \tag{12}$$

$$\hat{\mathbf{h}}_{k,\ell} = [\tilde{\mathbf{h}}_{k,\ell}]_{\mathcal{N}_{\ell}} \in \mathbb{C}^{\hat{N}_{\ell}\times 1}, \forall k, \ell. \tag{13}$$

Focusing on non-zero beams, these definitions enable the design of a more efficient beam-domain precoding strategy, which will be further developed in Section 3.

3 Precoder design in the beam domain

This section begins with the formulation of the precoder design problem in network massive MIMO transmission. Then we show that the sum-rate maximizing precoder can be obtained by designing it in the beam domain under the beam-based channel model. Furthermore, we show that the optimal beam-domain precoder for each user only needs to focus on the beams corresponding to the non-zero elements of its own beam-domain channel. This leads to a structured precoding approach, referred to as BSP, which effectively exploits the channel sparsity to reduce the dimension of the precoding problem.

3.1 Problem formulation

We assume that the instantaneous CSI of all associated users is available at the BSs, while each UT has access to its own instantaneous CSI and the covariance r_k of the aggregate interference-plus-noise $n'_k = \sum_{u\neq k} \mathbf{h}_k^{\mathrm{H}}\mathbf{p}_u d_u + n_k$. Under such assumption, when the UT treats n'_k as Gaussian noise, the achievable rate of the k-th user is given by [30]

$$\mathcal{R}_k = \log\left(1 + \frac{\mathbf{p}_k^{\mathrm{H}}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\mathbf{p}_k}{\sigma^2 + \sum_{u\neq k}\mathbf{p}_u^{\mathrm{H}}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\mathbf{p}_u}\right). \tag{14}$$

Our task is to find a precoding strategy to maximize the sum-rate with per-BS power constraints, formulated as the following optimization problem:

$$\begin{aligned}\mathbf{P1}: \max_{\{\mathbf{p}_k\}_{k=1}^K} &\sum_k \mathcal{R}_k,\\ \text{s.t.} &\sum_k \mathbf{p}_k^{\mathrm{H}}\mathbf{D}_\ell\mathbf{p}_k \leqslant P_\ell, \forall \ell,\end{aligned} \tag{15}$$

where $\mathbf{D}_\ell \in \mathbb{R}^{ML\times ML}$ is a block-diagonal selection matrix with the ℓ-th diagonal block being $\mathbf{I}_M$ and others being $\mathbf{0}_{M\times M}$.

Solving **P1** is inherently difficult because of the high-dimensional nature of the system, the non-convexity of the objective function, and the coupled per-BS power constraints. These factors complicate the search for an optimal solution and increase the computational complexity, making it essential to develop efficient precoding strategies to address these issues.

3.2 Beam-structured precoding

We first establish that the spatial-domain precoders $\mathbf{p}_k, \forall k$ lie in the range space of $\bar{\mathbf{V}} = \mathbf{I}_L \otimes \mathbf{V} \in \mathbb{C}^{ML\times NL}$, as stated in the following theorem.

Theorem 1. For any $\mathbf{p}_k \in \mathbb{R}^{ML\times 1}$, there exists a $\mathbf{q}_k \in \mathbb{C}^{NL\times 1}$ such that

$$\mathbf{p}_k = \bar{\mathbf{V}}\mathbf{q}_k. \tag{16}$$

Proof. As defined in (4), the $M \times N$ ($M \leqslant N$) matrix $\mathbf{V}$ is of Vandermonde type with distinct generating points, and therefore is full row rank, i.e., $\mathrm{rank}(\mathbf{V}) = M$. Thus, $\mathrm{rank}(\bar{\mathbf{V}}) = \mathrm{rank}(\mathbf{I}_L)\,\mathrm{rank}(\mathbf{V}) = LM$, and $\bar{\mathbf{V}}$ is also full row rank. By the result that any full row rank matrix $\mathbf{A}$ admits a solution to $\mathbf{p} = \mathbf{A}\mathbf{q}$ for every $\mathbf{p}$ [31], the claim follows.

Substituting (16) into (14), the sum-rate maximization problem **P1** can be reformulated as

$$\begin{aligned}\mathbf{P2}: \max_{\{\mathbf{q}_k\}_{k=1}^K} &\sum_k \log\left(1 + \frac{\mathbf{q}_k^{\mathrm{H}}\bar{\mathbf{V}}^{\mathrm{H}}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\bar{\mathbf{V}}\mathbf{q}_k}{\sigma^2 + \sum_{u\neq k}\mathbf{q}_u^{\mathrm{H}}\bar{\mathbf{V}}^{\mathrm{H}}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\bar{\mathbf{V}}\mathbf{q}_u}\right),\\ \text{s.t.} &\sum_k \mathbf{q}_k^{\mathrm{H}}\bar{\mathbf{V}}^{\mathrm{H}}\mathbf{D}_\ell\bar{\mathbf{V}}\mathbf{q}_k \leqslant P_\ell, \forall \ell.\end{aligned} \tag{17}$$

For **P2**, which optimizes over $\{\mathbf{q}_k\}$, we derive the following theorem, stating that each user needs to focus on only the beams corresponding to the non-zero elements of its own beam-domain channel vector.

Theorem 2. When $M \to \infty$ and $\mathcal{N}_{k,\ell} \cap \mathcal{N}_{u,\ell} = \emptyset, \forall \ell, k \neq u$, problem **P2** admits an optimal solution $\{\mathbf{q}_k^\star\}$ satisfying

$$[\mathbf{q}_k^\star]_{(\ell-1)N+n} = 0, \text{ for } n \notin \mathcal{N}_{k,\ell}, \forall k, \ell. \tag{18}$$

Proof. For the ℓ-th BS, removing from each stationary point $\mathbf{p}_{k,\ell}^\star, \forall k$, the components lying in the null space of the channel matrix $\mathbf{H}_\ell^{\mathrm{H}} = [\mathbf{h}_{1,\ell}, \ldots, \mathbf{h}_{K,\ell}]^{\mathrm{H}}$ leads to the notion of a positive stationary point $\hat{\mathbf{p}}_{k,\ell}$, which resides in the column space of $\mathbf{H}_\ell$. Since these null-space components contribute no useful signal power but only increase the total transmit power, it is sufficient to consider positive stationary points without loss of generality. Consequently, any positive stationary point that lies in the column space of the channel matrix can be represented by new variables $\{\mathbf{x}_{k,\ell}^\star \in \mathbb{C}^{K\times 1}\}$ as

$$\hat{\mathbf{p}}_{k,\ell} = \mathbf{H}_\ell \mathbf{x}_{k,\ell}^\star, \forall k, \ell, \tag{19}$$

and the original problem can be reformulated with stationary points equivalent to those in the original formulation [32, Theorem 2]. Furthermore, since the channel matrix can be decomposed as $\mathbf{H}_\ell = \hat{\mathbf{V}}_\ell\hat{\mathbf{H}}_\ell$, where

$\hat{\mathbf{H}}_\ell = [\hat{\mathbf{h}}_{1,\ell}, \ldots, \hat{\mathbf{h}}_{K,\ell}]$, we define the intermediate variable $\hat{\mathbf{q}}^\star_{k,\ell} = \hat{\mathbf{H}}_\ell \mathbf{x}^\star_{k,\ell} \in \mathbb{C}^{\hat{N}_\ell \times 1}$. Substituting this definition into (19) yields $\hat{\mathbf{p}}_{k,\ell} = \hat{\mathbf{V}}_\ell \hat{\mathbf{H}}_\ell \mathbf{x}^\star_{k,\ell} = \hat{\mathbf{V}}_\ell \hat{\mathbf{q}}^\star_{k,\ell}, \forall k, \ell$, or equivalently, in the aggregated form

$$\hat{\mathbf{p}}_k = \hat{\mathbf{V}} \hat{\mathbf{q}}^\star_k, \forall k, \tag{20}$$

where $\hat{\mathbf{q}}^\star_k = \left[(\hat{\mathbf{q}}^\star_{k,1})^{\mathrm{T}}, \ldots, (\hat{\mathbf{q}}^\star_{k,L})^{\mathrm{T}}\right]^{\mathrm{T}}$. Eq. (20) has the same structure as (16), except that here $\hat{\mathbf{q}}^\star_k$ lies in the subspace corresponding to the union of non-zero beam index sets of all users. With (9) and (20), solving **P2** is equivalent to solving the following problem with optimization variables $\{\hat{\mathbf{q}}_k\}$:

$$\begin{aligned} \mathbf{P3}: \max_{\{\hat{\mathbf{q}}_k\}_{k=1}^K} & \sum_k \log\left(1 + \frac{\hat{\mathbf{q}}_k^{\mathrm{H}} \hat{\mathbf{V}}^{\mathrm{H}} \hat{\mathbf{V}} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathrm{H}} \hat{\mathbf{V}}^{\mathrm{H}} \hat{\mathbf{V}} \hat{\mathbf{q}}_k}{\sigma^2 + \sum_{u \neq k} \hat{\mathbf{q}}_u^{\mathrm{H}} \hat{\mathbf{V}}^{\mathrm{H}} \hat{\mathbf{V}} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathrm{H}} \hat{\mathbf{V}}^{\mathrm{H}} \hat{\mathbf{V}} \hat{\mathbf{q}}_u}\right), \\ \text{s.t.} & \sum_k \hat{\mathbf{q}}_k^{\mathrm{H}} \hat{\mathbf{V}}^{\mathrm{H}} \mathbf{D}_\ell \hat{\mathbf{V}} \hat{\mathbf{q}}_k \leqslant P_\ell, \forall \ell. \end{aligned} \tag{21}$$

The element of $\hat{\mathbf{V}}^{\mathrm{H}} \hat{\mathbf{V}}$ can be obtained from the corresponding entries of $\mathbf{V}^{\mathrm{H}} \mathbf{V}$:

$$\left[\mathbf{V}^{\mathrm{H}} \mathbf{V}\right]_{i,j} = \mathbf{v}(\Omega_i)^{\mathrm{H}} \mathbf{v}(\Omega_j) = \frac{1}{M} e^{-\bar{\jmath}(M-1)\pi(i-j)/N} \cdot \frac{\sin(M\pi(i-j)/N)}{\sin(\pi(i-j)/N)}, \tag{22}$$

which is a Dirichlet kernel that decays as $|i-j|$ increases. As $M \to \infty$, this kernel asymptotically approaches a Dirac delta function. Since $\hat{\mathbf{V}}$ consists of the columns of $\mathbf{V}$ corresponding to the union of the non-zero supports in all users' beam-domain channel vectors, and the number of users is finite, the selected columns are typically well-separated in terms of their indices. As a result, when the number of BS antennas grows sufficiently large, the inner product between different columns of $\hat{\mathbf{V}}$ vanishes asymptotically, and we have $\lim_{M\to\infty} \hat{\mathbf{V}}^{\mathrm{H}} \hat{\mathbf{V}} = \mathbf{I}_{\hat{N}}$. Thus, when $M \to \infty$, **P3** is equivalent to

$$\begin{aligned} \mathbf{P4}: \max_{\{\hat{\mathbf{q}}_k\}_{k=1}^K} & \sum_k \log\left(1 + \frac{\hat{\mathbf{q}}_k^{\mathrm{H}} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathrm{H}} \hat{\mathbf{q}}_k}{\sigma^2 + \sum_{u \neq k} \hat{\mathbf{q}}_u^{\mathrm{H}} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathrm{H}} \hat{\mathbf{q}}_u}\right), \\ \text{s.t.} & \sum_k \hat{\mathbf{q}}_k^{\mathrm{H}} \hat{\mathbf{D}}_\ell \hat{\mathbf{q}}_k \leqslant P_\ell, \forall \ell, \end{aligned} \tag{23}$$

where $\hat{\mathbf{D}}_\ell \in \mathbb{R}^{\hat{N} \times \hat{N}}$ is a block-diagonal selection matrix with the ℓ-th diagonal block being $\mathbf{I}_{\hat{N}_\ell}$ and others being $\mathbf{0}_{\hat{N}_r \times \hat{N}_r}, \forall r \neq \ell$.

Define a mask vector $\mathbf{m}_u \in \mathbb{C}^{\hat{N} \times 1}$ as

$$[\mathbf{m}_u]_{\sum_{r=0}^{\ell-1} \hat{N}_r + n} = \begin{cases} 1, & [\mathcal{N}_\ell]_n \in \mathcal{N}_{k,\ell}, \\ 0, & \text{else}. \end{cases} \tag{24}$$

For a specific user u, let $\hat{\mathbf{q}}'_u = \hat{\mathbf{q}}^\star_u \odot \mathbf{m}_u$, and $\hat{\mathbf{q}}'_k = \hat{\mathbf{q}}^\star_k, \forall k \neq u$, where $\hat{\mathbf{q}}^\star_k, \forall k$ are the optimal solutions to **P4**. Since $(\hat{\mathbf{q}}'_u)^{\mathrm{H}} \hat{\mathbf{D}}_\ell \hat{\mathbf{q}}'_u \leqslant (\hat{\mathbf{q}}^\star_u)^{\mathrm{H}} \hat{\mathbf{D}}_\ell \hat{\mathbf{q}}^\star_u$, the constructed precoding vectors $\{\hat{\mathbf{q}}'_k\}$ satisfy the power constraint $\sum_k (\hat{\mathbf{q}}'_k)^{\mathrm{H}} \hat{\mathbf{D}}_\ell \hat{\mathbf{q}}'_k \leqslant P_\ell$.

According to the definition of the optimal solution, we have

$$\sum_k \mathcal{R}_k(\hat{\mathbf{q}}'_1, \ldots, \hat{\mathbf{q}}'_K) \leqslant \sum_k \mathcal{R}_k(\hat{\mathbf{q}}^\star_1, \ldots, \hat{\mathbf{q}}^\star_K). \tag{25}$$

On the other hand, under the assumption that $\mathcal{N}_{k,\ell} \cap \mathcal{N}_{u,\ell} = \emptyset, \forall \ell, k \neq u$, we have $(\hat{\mathbf{q}}'_u)^{\mathrm{H}} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathrm{H}} \hat{\mathbf{q}}'_u \leqslant (\hat{\mathbf{q}}^\star_u)^{\mathrm{H}} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathrm{H}} \hat{\mathbf{q}}^\star_u$, $\forall k \neq u$ and $(\hat{\mathbf{q}}'_k)^{\mathrm{H}} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathrm{H}} \hat{\mathbf{q}}'_k = (\hat{\mathbf{q}}^\star_k)^{\mathrm{H}} \hat{\mathbf{h}}_k \hat{\mathbf{h}}_k^{\mathrm{H}} \hat{\mathbf{q}}^\star_k, \forall k \neq u$. Thus,

$$\mathcal{R}_k(\hat{\mathbf{q}}'_1, \ldots, \hat{\mathbf{q}}'_K) \geqslant \mathcal{R}_k(\hat{\mathbf{q}}^\star_1, \ldots, \hat{\mathbf{q}}^\star_K), \forall k \neq u. \tag{26}$$

For user u, we have $(\hat{\mathbf{q}}'_k)^{\mathrm{H}} \hat{\mathbf{h}}_u \hat{\mathbf{h}}_u^{\mathrm{H}} \hat{\mathbf{q}}'_k = (\hat{\mathbf{q}}^\star_k)^{\mathrm{H}} \hat{\mathbf{h}}_u \hat{\mathbf{h}}_u^{\mathrm{H}} \hat{\mathbf{q}}^\star_k, \forall k \neq u$ and $(\hat{\mathbf{q}}'_u)^{\mathrm{H}} \hat{\mathbf{h}}_u \hat{\mathbf{h}}_u^{\mathrm{H}} \hat{\mathbf{q}}'_u = (\hat{\mathbf{q}}^\star_u)^{\mathrm{H}} \hat{\mathbf{h}}_u \hat{\mathbf{h}}_u^{\mathrm{H}} \hat{\mathbf{q}}^\star_u$. Thus,

$$\mathcal{R}_u(\hat{\mathbf{q}}'_1, \ldots, \hat{\mathbf{q}}'_K) = \mathcal{R}_u(\hat{\mathbf{q}}^\star_1, \ldots, \hat{\mathbf{q}}^\star_K). \tag{27}$$

Eqs. (26) and (27) together yield

$$\sum_k \mathcal{R}_k(\hat{\mathbf{q}}'_1, \ldots, \hat{\mathbf{q}}'_K) \geqslant \sum_k \mathcal{R}_k(\hat{\mathbf{q}}^\star_1, \ldots, \hat{\mathbf{q}}^\star_K). \tag{28}$$

By combining (25) with (28), we obtain

$$\sum_k \mathcal{R}_k(\hat{\mathbf{q}}'_1,\ldots,\hat{\mathbf{q}}'_K)=\sum_k \mathcal{R}_k(\hat{\mathbf{q}}^\star_1,\ldots,\hat{\mathbf{q}}^\star_K). \tag{29}$$

Thus, it suffices to focus on the elements in $\{\hat{\mathbf{q}}^\star_u\}$ corresponding to the non-zero beams of each user's channel vector, which means $[\hat{\mathbf{q}}^\star_k]_{\sum_{r=1}^{\ell-1}\check{N}_r+n}=0$, for $[\mathcal{N}_\ell]_n\notin\mathcal{N}_{k,\ell},\forall k,\ell$ and is in accordance with (18). This completes the proof of Theorem 2.

Under the asymptotic condition $M\to\infty$, the disjoint-support condition $\mathcal{N}_{k,\ell}\cap\mathcal{N}_{u,\ell}=\emptyset,\forall\ell,k\neq u$ in Theorem 2 serves as a sufficient condition that enables a lower-dimensional characterization of the solution space. In large-scale systems, the angular resolution in the beam domain becomes increasingly fine as the number of antennas grows and the fine factor $F>1$ is applied, thereby narrowing each beam's spatial coverage and reducing the overlap of support sets. Moreover, user scheduling strategies in wireless systems often prioritize spatially separable users, which naturally promotes disjoint or near-disjoint supports in the beam domain.

For each user, the number of non-zero beams from all the BSs is $\check{N}_k=\sum_\ell N_{k,\ell},\forall k$. According to Theorem 2, the optimal precoder for each user only needs to involve its own non-zero beams, thereby reducing the dimension of interest to $\sum_k\check{N}_k$. Thus, the precoding structure in (16) becomes

$$\mathbf{p}_k=\bar{\mathbf{V}}\boldsymbol{\Phi}_k\mathbf{u}_k,\forall k, \tag{30}$$

where $\mathbf{u}_k\in\mathbb{C}^{\check{N}_k\times1}$ is constructed by extracting from $\mathbf{q}_k$ the components corresponding to the non-zero positions of its beam-domain channel vector, and $\{\boldsymbol{\Phi}_k\in\mathbb{R}^{NL\times\check{N}_k}\}$ are the non-zero beam indicating matrices, mapping the lower-dimensional non-zero elements to the NL-dimensional beam spaces as

$$\boldsymbol{\Phi}_k=\mathrm{bdiag}(\boldsymbol{\Phi}_{k,1},\ldots,\boldsymbol{\Phi}_{k,L}),\forall k, \tag{31}$$

$$[\boldsymbol{\Phi}_{k,\ell}]_{m,n}=\begin{cases}1, & m=[\mathcal{N}_{k,\ell}]_n, n=1,\ldots,N_{k,\ell},\\ 0, & \text{else.}\end{cases} \tag{32}$$

We refer to this design as beam-structured precoding, since the spatial-domain precoder is structured as the product of a beam matrix $\bar{\mathbf{V}}$ and the compact beam-domain precoding vector $\mathbf{u}_k$ that is restricted to non-zero beams via the beam indicating matrices $\{\boldsymbol{\Phi}_k\}$. Similarly, with $\bar{\mathbf{V}}$ and $\{\boldsymbol{\Phi}_k\}$, Eq. (9) can be rewritten as

$$\mathbf{h}_k=\bar{\mathbf{V}}\boldsymbol{\Phi}_k\mathbf{g}_k,\forall k, \tag{33}$$

where $\{\mathbf{g}_k\in\mathbb{C}^{\check{N}_k\times1}\}$ are the compact beam-domain channel vectors derived by extracting all the non-zero beams as

$$\mathbf{g}_k=[[\tilde{\mathbf{h}}_{k,1}]^{\mathrm{T}}_{\mathcal{N}_{k,1}},\ldots,[\tilde{\mathbf{h}}_{k,L}]^{\mathrm{T}}_{\mathcal{N}_{k,L}}]^{\mathrm{T}},\forall k. \tag{34}$$

The rate of the k-th user in (14) can be rewritten as

$$\mathcal{R}_k=\log\left(1+\frac{\mathbf{u}_k^{\mathrm{H}}\mathbf{B}_{k,k}\mathbf{g}_k\mathbf{g}_k^{\mathrm{H}}\mathbf{B}_{k,k}\mathbf{u}_k}{\sigma^2+\sum_{u\neq k}\mathbf{u}_u^{\mathrm{H}}\mathbf{B}_{u,k}\mathbf{g}_k\mathbf{g}_k^{\mathrm{H}}\mathbf{B}_{k,u}\mathbf{u}_u}\right), \tag{35}$$

where $\mathbf{B}_{k,u}=\boldsymbol{\Phi}_k^{\mathrm{T}}\bar{\mathbf{V}}^{\mathrm{H}}\bar{\mathbf{V}}\boldsymbol{\Phi}_u\in\mathbb{C}^{\check{N}_k\times\check{N}_u}$. Thus, the beam-domain precoding problem can be formulated as

$$\begin{aligned}\mathbf{P5}:\ &\max_{\{\mathbf{u}_k\}_{k=1}^K}\ \sum_k\log\left(1+\frac{\mathbf{u}_k^{\mathrm{H}}\mathbf{B}_{k,k}\mathbf{g}_k\mathbf{g}_k^{\mathrm{H}}\mathbf{B}_{k,k}\mathbf{u}_k}{\sigma^2+\sum_{u\neq k}\mathbf{u}_u^{\mathrm{H}}\mathbf{B}_{u,k}\mathbf{g}_k\mathbf{g}_k^{\mathrm{H}}\mathbf{B}_{k,u}\mathbf{u}_u}\right),\\ &\text{s.t.}\ \sum_k\mathbf{u}_k^{\mathrm{H}}\mathbf{B}_{k,k}\check{\mathbf{D}}_{k,\ell}\mathbf{u}_k\leqslant P_\ell,\forall\ell,\end{aligned} \tag{36}$$

where $\check{\mathbf{D}}_{k,\ell}\in\mathbb{R}^{\check{N}_k\times\check{N}_k}$ is a block diagonal matrix with the ℓ-th diagonal block being $\mathbf{I}_{N_{k,\ell}}$ and others being $\mathbf{0}_{N_{k,r}\times N_{k,r}}$.

As a result of the reformulation in **P5**, the dimension of the optimization space becomes $\check{N}=\sum_k\sum_\ell N_{k,\ell}$. In our simulation setting, the average of $N_{k,\ell}$ is approximately $0.2M$. Compared to the original problem, the computational complexity can be significantly reduced.

4 Precoder design with Hamiltonian-based optimization

In this section, we tackle problem **P5** through a Hamiltonian system, which offers a physics-inspired framework for optimization. In this setting, the objective function $f(\mathbf{u}) = -\sum_k \mathcal{R}_k$ is taken as the potential energy, while the collective compact beam-domain precoding vector for all UTs $\mathbf{u} = \left[\mathbf{u}_1^{\mathrm{T}}, \ldots, \mathbf{u}_K^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{\check{N}\times 1}$ is referred to as the generalized coordinate [18,33]. This formulation models the optimization process through an energy-based dynamical system, where the system evolves according to a set of governing equations derived from physical principles. The underlying energy structure helps the system explore the solution space, naturally guiding $\mathbf{u}$ toward a state that minimizes the objective function [34]. While the original Hamiltonian system conserves energy, we incorporate a dissipative mechanism to allow energy decay, thereby promoting convergence to a steady-state solution [35,36]. These dynamics are numerically solved using an integrator that provides discrete updates, thereby driving the optimization process forward [21,22].

4.1 Hamiltonian-based optimization

The optimization problem can be expressed in the form of

$$\begin{aligned} \mathbf{P6}: \min_{\mathbf{u}} \ & f(\mathbf{u}), \\ \text{s.t. } & \boldsymbol{\psi}(\mathbf{u}) \leqslant \mathbf{0}, \end{aligned} \tag{37}$$

where $\boldsymbol{\psi}(\mathbf{u}) : \mathbb{C}^{\check{N}} \to \mathbb{R}^L$ is the vector of constraint functions, expressed as

$$\boldsymbol{\psi}(\mathbf{u}) = \Big[\sum_k \mathbf{u}_k^{\mathrm{H}}\mathbf{B}_{k,k}\check{\mathbf{D}}_{k,1}\mathbf{u}_k - P_1, \ldots, \sum_k \mathbf{u}_k^{\mathrm{H}}\mathbf{B}_{k,k}\check{\mathbf{D}}_{k,L}\mathbf{u}_k - P_L\Big]^{\mathrm{T}}. \tag{38}$$

The unconstrained Hamiltonian, which corresponds to the system's total energy, is calculated as the sum of kinetic and potential energies [37,38],

$$\mathcal{H}(\mathbf{u}, \mathbf{n}) = T(\mathbf{n}) + f(\mathbf{u}), \tag{39}$$

where $\mathbf{n} = \mathbf{M}\dot{\mathbf{u}}$ is the momentum characterized by a mass matrix $\mathbf{M}$ that is positive definite[1)], and $T(\mathbf{n}) = \frac{1}{2}\dot{\mathbf{u}}^{\mathrm{H}}\mathbf{M}\dot{\mathbf{u}} = \frac{1}{2}\mathbf{n}^{\mathrm{H}}\mathbf{M}^{-1}\mathbf{n}$ is the kinetic energy. The pair $(\mathbf{u}, \mathbf{n})$ is called phase space coordinates. In the Hamiltonian system, the phase space represents the state of the system, with the kinetic term capturing its dynamic behavior and governing how the system transitions from one state to another [34]. When applied to optimization problems, the system's state evolves analogously in the phase space, where the kinetic term helps guide the search process effectively toward the optimal solution within the optimization space.

To incorporate constraints, define a total Hamiltonian by adding constraint terms with Lagrange multipliers to the original Hamiltonian as follows [39]:

$$\mathcal{H}_{\text{total}} = \mathcal{H}(\mathbf{u}, \mathbf{n}) + \boldsymbol{\psi}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda}, \tag{40}$$

where $\boldsymbol{\lambda} = [\lambda_1, \ldots, \lambda_L]^{\mathrm{T}}$ are the Lagrangian multipliers. According to complementary slackness $\boldsymbol{\psi}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{0}$, a constraint is only active on the boundary of the feasible region, while being typically inactive within the region [40]. Specifically, when an inequality constraint is inactive (e.g., $\psi_\ell(\mathbf{u}) < 0$), the corresponding Lagrangian multiplier λ_ℓ is zero, and the system is effectively unconstrained with respect to this inequality. When this inequality constraint becomes active ($\lambda_\ell \neq 0$), it acts as an equality constraint, meaning $\psi_\ell(\mathbf{u}) = 0$. Thus, these inequality constraints behave as equality constraints, with an "active/inactive" switch indicating whether they participate in the system dynamics or not [41].

The dynamics of the system under these constraints are described by the Hamilton's equations $\dot{\mathbf{u}} = \partial\mathcal{H}_{\text{total}}/\partial\mathbf{n}, \dot{\mathbf{n}} = -\partial\mathcal{H}_{\text{total}}/\partial\mathbf{u}$ [38,41], yielding the following differential-algebraic system of equations (DAE) [42]:

$$\dot{\mathbf{u}} = \frac{\partial\mathcal{H}}{\partial\mathbf{n}}, \tag{41a}$$

$$\dot{\mathbf{n}} = -\frac{\partial\mathcal{H}}{\partial\mathbf{u}} - \mathbf{J}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda}, \tag{41b}$$

$$\boldsymbol{\psi}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{0}, \quad \boldsymbol{\lambda} \geqslant \mathbf{0}, \tag{41c}$$

1) For any positive-definite mass matrix $\mathbf{M}$, we can apply a linear transformation $\mathbf{u} = \mathbf{M}^{-1/2}\hat{\mathbf{u}}, \mathbf{n} = \mathbf{M}^{1/2}\hat{\mathbf{n}}$, and re-define the objective and constraint functions as $f(\mathbf{u}) = \hat{f}(\mathbf{M}^{1/2}\mathbf{u}), \boldsymbol{\psi} = \hat{\boldsymbol{\psi}}(\mathbf{M}^{1/2}\mathbf{u})$. Under this change of coordinates, the system is equivalent to a Hamiltonian system with the identity mass matrix $\mathbf{I}$.

where $\mathbf{J}(\mathbf{u}) = [\nabla\psi_1(\mathbf{u}), \ldots, \nabla\psi_L(\mathbf{u})]^{\mathrm{T}}$ is the $L \times \check{N}$ Jacobian matrix of the constraints at point $\mathbf{u}$. These equations drive the system's dynamics, guiding its evolution while maintaining feasibility under the given constraints. Differentiating the constraint twice yields

$$\begin{aligned} \mathbf{0} &= \mathbf{J}(\mathbf{u})\frac{\partial \mathcal{H}}{\partial \mathbf{n}}, \\ \mathbf{0} &= \frac{\partial}{\partial \mathbf{u}}\Big(\mathbf{J}(\mathbf{u})\frac{\partial \mathcal{H}}{\partial \mathbf{n}}\Big)\frac{\partial \mathcal{H}}{\partial \mathbf{n}} - \mathbf{J}(\mathbf{u})\frac{\partial^2 \mathcal{H}}{\partial \mathbf{n}^2}\Big(\frac{\partial \mathcal{H}}{\partial \mathbf{u}} + \mathbf{J}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda}\Big), \end{aligned} \tag{42}$$

from which we can calculate $\boldsymbol{\lambda}$.

The differentiation of $\mathcal{H}$ with respect to time can be written as

$$\dot{\mathcal{H}} = \frac{\partial \mathcal{H}}{\partial \mathbf{n}}\dot{\mathbf{n}} + \frac{\partial \mathcal{H}}{\partial \mathbf{u}}\dot{\mathbf{u}} + \{\mathcal{H}, \mathcal{H}\} = \frac{\partial \mathcal{H}}{\partial \mathbf{n}}\Big(- \frac{\partial \mathcal{H}}{\partial \mathbf{u}} - \mathbf{J}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda}\Big) + \frac{\partial \mathcal{H}}{\partial \mathbf{u}}\frac{\partial \mathcal{H}}{\partial \mathbf{n}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{n}}\mathbf{J}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda}, \tag{43}$$

where $\{A, B\}$ is the Poisson bracket of the two dynamical variables, defined as $\sum_i \Big(\frac{\partial A}{\partial u_i}\frac{\partial B}{\partial n_i} - \frac{\partial A}{\partial n_i}\frac{\partial B}{\partial u_i}\Big)$ [43]. According to the complementary slackness, the differentiation in (43) vanishes, which means the Hamiltonian $\mathcal{H}$ is constant along solutions. Trajectories of such a conservative system oscillate on the level sets in phase space, rather than converging to a single solution [35, 36].

To reach the minimum of f, we incorporate a dissipation term that gradually reduces the system's total energy, driving the system towards a level set corresponding to a low energy and aligned with the system's optimum. The dissipative force can usually be expressed by adding a damping term $-\gamma\mathbf{n}$ associated with the momentum $\mathbf{n}$, where $\gamma > 0$ is the damping coefficient, reflecting the rate of energy loss. The modified Hamiltonian equations become [36]

$$\dot{\mathbf{u}} = \frac{\partial \mathcal{H}}{\partial \mathbf{n}}, \tag{44a}$$

$$\dot{\mathbf{n}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{u}} - \mathbf{J}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda} - \gamma\mathbf{n}, \tag{44b}$$

$$\boldsymbol{\psi}(\mathbf{u})^{\mathrm{T}}\boldsymbol{\lambda} = \mathbf{0}, \quad \boldsymbol{\lambda} \geqslant \mathbf{0}. \tag{44c}$$

The dissipative Hamiltonian system is designed so that the total energy decreases over time, causing the system state to evolve toward a lower energy configuration. Since the objective function $f(\mathbf{u}) = -\sum_k \mathcal{R}_k$ is embedded as the potential energy, this energy dissipation naturally drives the system toward a minimizer of the objective. Solving the dissipative Hamiltonian DAE (44) yields a trajectory $(\mathbf{u}(t), \mathbf{n}(t))$ that asymptotically converges to a stationary point $(\mathbf{u}^\star, \mathbf{0})$, where $\mathbf{u}^\star$ minimizes the objective function.

4.2 Numerical solution with the RATTLE scheme

In a Hamiltonian system, the gradient of the potential energy, which is the same as the gradient of the objective function, reflects the direction and magnitude of the system's state changes. Together with the gradients of the constraints, they provide the evolution information under the constraint conditions, ensuring that the system converges correctly towards the optimal solution. Therefore, before using the numerical integrator to solve (44), we first define the gradients of the optimization objective and constraint functions.

Theorem 3. The gradient of the objective function $f(\mathbf{u})$ is given as $\nabla f(\mathbf{u}) = (\nabla f(\mathbf{u}_1), \ldots, \nabla f(\mathbf{u}_K))$, with

$$\nabla f(\mathbf{u}_k) = -\hat{r}_k^{-1}\mathbf{C}_{k,k}\mathbf{u}_k - \sum_{u\neq k}(\hat{r}_u^{-1} - r_u^{-1})\mathbf{C}_{u,k}\mathbf{u}_k, \tag{45}$$

where $r_k = \sigma^2 + \sum_{u\neq k}\mathbf{u}_u^{\mathrm{H}}\mathbf{C}_{k,u}\mathbf{u}_u$, $\hat{r}_k = r_k + \mathbf{u}_k^{\mathrm{H}}\mathbf{C}_{k,k}\mathbf{u}_k$, and $\mathbf{C}_{k,u} = \mathbf{B}_{u,k}\mathbf{g}_k\mathbf{g}_k^{\mathrm{H}}\mathbf{B}_{k,u} \in \mathbb{C}^{\check{N}_u\times\check{N}_u}$. The gradient of the ℓ-th constraint function $\psi_\ell(\mathbf{u}) = \sum_k \mathbf{u}_k^{\mathrm{H}}\mathbf{B}_{k,k}\check{\mathbf{D}}_{k,\ell}\mathbf{u}_k - P_\ell$ is calculated as

$$\nabla\psi_\ell(\mathbf{u}) = \check{\mathbf{B}}_\ell\mathbf{u}, \tag{46}$$

where $\check{\mathbf{B}}_\ell = \mathrm{bdiag}(\mathbf{B}_{1,1}\check{\mathbf{D}}_{k,1}, \ldots, \mathbf{B}_{K,K}\check{\mathbf{D}}_{k,L}) \in \mathbb{C}^{\check{N}\times\check{N}}$.

Proof. The gradient of $f(\mathbf{u})$ with respect to $\mathbf{u}_k$ can be calculated as

$$\nabla f(\mathbf{u}_k) = -\frac{\partial \mathcal{R}_k}{\partial \mathbf{u}_k^*} - \sum_{u\neq k}\frac{\partial \mathcal{R}_u}{\partial \mathbf{u}_k^*}. \tag{47}$$

Rewrite (35) as

$$\mathcal{R}_k = \log(1 + r_k^{-1}\mathbf{u}_k^{\mathrm{H}}\mathbf{C}_{k,k}\mathbf{u}_k) = \log\left(\sigma^2 + \sum_u \mathbf{u}_u^{\mathrm{H}}\mathbf{C}_{k,u}\mathbf{u}_u\right) - \log\left(\sigma^2 + \sum_{u\neq k} \mathbf{u}_u^{\mathrm{H}}\mathbf{C}_{k,u}\mathbf{u}_u\right). \tag{48}$$

Thus, $\frac{\partial \mathcal{R}_k}{\partial \mathbf{u}_k^*}$ and $\frac{\partial \mathcal{R}_u}{\partial \mathbf{u}_k^*}(u \neq k)$ can be calculated, respectively, as

$$\frac{\partial \mathcal{R}_k}{\partial \mathbf{u}_k^*} = \hat{r}_k^{-1}\mathbf{C}_{k,k}\mathbf{u}_k, \tag{49}$$

$$\frac{\partial \mathcal{R}_u}{\partial \mathbf{u}_k^*} = (\hat{r}_u^{-1} - r_u^{-1})\mathbf{C}_{u,k}\mathbf{u}_k, \forall u \neq k. \tag{50}$$

Thus, we have

$$\nabla f(\mathbf{u}_k) = -\hat{r}_k^{-1}\mathbf{C}_{k,k}\mathbf{u}_k - \sum_{u\neq k}(\hat{r}_u^{-1} - r_u^{-1})\mathbf{C}_{u,k}\mathbf{u}_k. \tag{51}$$

For the constraint function $\psi_\ell(\mathbf{u}) = \sum_k \mathbf{u}_k^{\mathrm{H}}\mathbf{B}_{k,k}\check{\mathbf{D}}_{k,\ell}\mathbf{u}_k - P_\ell$, we have

$$\nabla\psi_\ell(\mathbf{u}_k) = \frac{\partial \psi_\ell(\mathbf{u})}{\partial \mathbf{u}_k^*} = \mathbf{B}_{k,k}\check{\mathbf{D}}_{k,\ell}\mathbf{u}_k. \tag{52}$$

Thus, the gradient of $\psi_\ell(\mathbf{u})$ can be constituted as

$$\nabla\psi_\ell(\mathbf{u}) = \begin{bmatrix} \mathbf{B}_{1,1}\check{\mathbf{D}}_{1,\ell}\mathbf{u}_1 \\ \vdots \\ \mathbf{B}_{K,K}\check{\mathbf{D}}_{K,\ell}\mathbf{u}_K \end{bmatrix} = \check{\mathbf{B}}_\ell\mathbf{u}. \tag{53}$$

This completes the proof of Theorem 3.

With the gradient of the constraint functions, the Jacobian matrix can be derived as

$$[\mathbf{J}]_{\ell,:}(\mathbf{u}) = \nabla^{\mathrm{T}}\psi_\ell(\mathbf{u}) = (\check{\mathbf{B}}_\ell\mathbf{u})^{\mathrm{T}}. \tag{54}$$

Substituting the expression of $\mathbf{J}(\mathbf{u})$ into (42), we derive the expression of $\boldsymbol{\lambda}$. Specifically, the partial derivatives of $\mathcal{H}$ are $\frac{\partial \mathcal{H}}{\partial \mathbf{n}} = \frac{1}{2}\mathbf{M}^{-1}\mathbf{n}$, and $\frac{\partial \mathcal{H}}{\partial \mathbf{u}} = \nabla f(\mathbf{u})$. Thus, Eq. (42) can be rewritten as

$$\frac{1}{2}(\mathbf{J}(\mathbf{u})\mathbf{M}^{-1}\mathbf{J}(\mathbf{u})^{\mathrm{T}})\boldsymbol{\lambda} = \frac{1}{4}\mathbf{b} - \frac{1}{2}\mathbf{J}(\mathbf{u})\mathbf{M}^{-1}\nabla f(\mathbf{u}), \tag{55}$$

where $\mathbf{b} = [\mathbf{n}^{\mathrm{T}}\mathbf{M}^{-1}\check{\mathbf{B}}_1\mathbf{M}^{-1}\mathbf{n}, \ldots, \mathbf{n}^{\mathrm{T}}\mathbf{M}^{-1}\check{\mathbf{B}}_L\mathbf{M}^{-1}\mathbf{n}]^{\mathrm{T}}$. The multipliers can be calculated as

$$\boldsymbol{\lambda} = (2\mathbf{J}(\mathbf{u})\mathbf{M}^{-1}\mathbf{J}(\mathbf{u})^{\mathrm{T}})^{-1}(\mathbf{b} - 2\mathbf{J}(\mathbf{u})\mathbf{M}^{-1}\nabla f(\mathbf{u})). \tag{56}$$

Note that the Lagrangian multipliers corresponding to inactive constraints should be zero. For each BS, if its power constraint is inactive, i.e., $\psi_\ell(\mathbf{u}) < 0$, we set the associated multiplier λ_ℓ to zero to strictly adhere to the complementary slackness condition.

To solve the dissipative Hamiltonian equations (44), we turn to the RATTLE integrator, a well-known method for solving this DAE by integrating the system's dynamics while enforcing rigid constraints at each time step [21, 22]. The RATTLE integrator preserves the structure of the continuous-time dissipative dynamics, enabling the discrete-time solution to faithfully follow the energy decay trajectory of the system [44]. As the integration proceeds, the total energy of the system gradually decreases due to dissipation, and the trajectory asymptotically converges to a steady state, which corresponds to a local minimizer of the objective function $f(\mathbf{u})$. We employ the dissipative leapfrog symmetric second-order RATTLE method, whose update equations are given by [36]

$$\mathbf{n}^{(d+1/2)} = e^{-\gamma h/2}\mathbf{n}^{(d)} - \frac{h}{2}\Big(\frac{\partial \mathcal{H}(\mathbf{u}^{(d)})}{\partial \mathbf{u}} + \mathbf{J}(\mathbf{u}^{(d)})^{\mathrm{T}}\boldsymbol{\lambda}^{(d)}\Big), \tag{57a}$$

$$\mathbf{u}^{(d+1)} = \mathbf{u}^{(d)} + h\frac{\partial \mathcal{H}(\mathbf{n}^{(d+1/2)})}{\partial \mathbf{n}}, \tag{57b}$$

$$\mathbf{n}^{(d+1)} = e^{-\gamma h/2}\Big(\mathbf{n}^{(d+1/2)} - \frac{h}{2}\Big(\frac{\partial \mathcal{H}(\mathbf{u}^{(d+1)})}{\partial \mathbf{u}} + \mathbf{J}(\mathbf{u}^{(d+1)})^{\mathrm{T}}\boldsymbol{\mu}^{(d)}\Big)\Big), \tag{57c}$$

where h is the step size for the time discretization, $\boldsymbol{\lambda}^{(d)}$ and $\boldsymbol{\mu}^{(d)}$ are calculated according to the constraints.

Algorithm 1 summarizes the solution to the linear BSP problem under the Hamiltonian framework, where the RATTLE integrator in (57) is employed to compute the system dynamics. Owing to the intrinsic dissipative structure of the Hamiltonian system, the total energy monotonically decreases over time, driving the system toward a steady state that corresponds to a minimizer of the optimization problem **P6**. The RATTLE integrator preserves this energy-decaying behavior at the discrete level, ensuring that the numerical trajectory faithfully follows the continuous-time convergence path [38].

Algorithm 1 BSP-hamiltonian design with a dissipative second order RATTLE integrator.

Input: Initialized precoder $\mathbf{u}^{(0)}$ and momentum $\mathbf{n}^{(0)}$, and accuracy ϵ.

Set $d=0$;
repeat
  $\mathbf{n}^{(d+1/2)} \leftarrow e^{-\gamma h/2}\mathbf{n}^{(d)} - \frac{h}{2}(\nabla f(\mathbf{u}^{(d)}) + \mathbf{J}(\mathbf{u}^{(d)})^{\mathrm{T}}\boldsymbol{\lambda}^{(d)})$;
  $\mathbf{u}^{(d+1)} \leftarrow \mathbf{u}^{(d)} + \frac{h}{2}\mathbf{M}^{-1}\mathbf{n}^{(d+1/2)}$;
  Compute $\nabla f(\mathbf{u}^{(d+1)})$, $\mathbf{J}(\mathbf{u}^{(d+1)})$ and $\boldsymbol{\lambda}^{(d+1)}$ according to (45), (54) and (56), respectively;
  $\boldsymbol{\mu}^{(d+1)} \leftarrow (\mathbf{J}(\mathbf{u}^{(d+1)})\mathbf{M}^{-1}\mathbf{J}(\mathbf{u}^{(d+1)})^{\mathrm{T}})^{-1}(\frac{2}{h}\mathbf{J}(\mathbf{u}^{(d+1)})\mathbf{M}^{-1}\mathbf{n}^{(d+1/2)} - \mathbf{J}(\mathbf{u}^{(d+1)})\mathbf{M}^{-1}\nabla f(\mathbf{u}^{(d+1)}))$;
  $\mathbf{n}^{(d+1)} \leftarrow e^{-\gamma h/2}(\mathbf{n}^{(d+1/2)} - \frac{h}{2}(\nabla f(\mathbf{u}^{(d+1)}) + \mathbf{J}(\mathbf{u}^{(d+1)})^{\mathrm{T}}\boldsymbol{\mu}^{(d)}))$;
  Set $d \leftarrow d+1$, and calculate $\mathcal{R}_k^{(d)}, \forall k$ according to (35);
until $|\sum_k \mathcal{R}_k^{(d)} - \sum_k \mathcal{R}_k^{(d-1)}| \leqslant \epsilon$.

4.3 Computational complexity

We analyze the computational complexity in two aspects: design complexity, which refers to the cost of computing the precoding vectors $\mathbf{u}_k, \forall k$, and implementation complexity, which corresponds to the cost of generating the transmitted signal based on the designed precoders.

(1) Design complexity: The BSP via Hamiltonian-based optimization Algorithm 1 utilizes a single-layer numerical scheme to iteratively advance the solution in discrete time. At each iteration, the computational complexity of Algorithm 1 comes mainly from computing $\nabla f(\mathbf{u})$ as (45), which involves calculating $\hat{r}_k = \sigma^2 + \sum_u \mathbf{u}_u^{\mathrm{H}}\mathbf{C}_{k,u}\mathbf{u}_u, \forall k$, as well as evaluating the gradient $\nabla f(\mathbf{u})$ itself. Notably, the value of $\mathbf{C}_{k,u}$ depends only on the propagation environment, and therefore should be computed once and reused in each iteration. Furthermore, $\mathbf{C}_{k,u}$ is treated symbolically as the outer product of two vectors, $(\mathbf{B}_{u,k}\mathbf{g}_k)(\mathbf{B}_{u,k}\mathbf{g}_k)^{\mathrm{H}}$, throughout the calculation of $\nabla f(\mathbf{u})$, rather than explicitly formed as a matrix. For notational convenience, let $N_{\mathrm{avg}} = \frac{1}{KL}\sum_k\sum_\ell N_{k,\ell}$ denote the average number of non-zero beams across all BS-UT pairs. Thus, the computational complexity of the algorithm is $\mathcal{O}(N_{\mathrm{iter}} \times 3K^2LN_{\mathrm{avg}})$, where N_{iter} denotes the total iterations.

We compare our algorithm with the widely used WMMSE method, which has been extended for joint processing/transmission (JP/JT) in [45, Algorithm 1] and [11, Algorithms 1–3]. In typical network massive MIMO systems where $ML > K$, the complexity of the WMMSE algorithm can be expressed as $\mathcal{O}(N_{\mathrm{out}}^{\mathrm{W}}(K^3 + 3K^2ML))$, where $N_{\mathrm{out}}^{\mathrm{W}}$ represents the total outer iterations required by the WMMSE procedure. Owing to the sparse nature of the channel (i.e., $N_{\mathrm{avg}} < M$), and the fact that the BSP via Hamiltonian-based optimization does not involve the cubic term K^3, its per-iteration complexity is strictly lower than that of WMMSE. Regarding the iteration count, the WMMSE method requires a greater number of iterations as the system scales up [15, 46], whereas Algorithm 1 achieves convergence with fewer iterations, which will be validated by the simulation results in Section 5.

(2) Implementation complexity: To obtain the transmitted signal $\mathbf{x} = \bar{\mathbf{V}}\sum_k \boldsymbol{\Phi}_k\mathbf{u}_k d_k$, we decompose the processing across the BSs and generate each BS's signal as $\mathbf{x}_\ell = \mathbf{V}\sum_k \boldsymbol{\Phi}_k\mathbf{u}_{k,\ell}d_k, \forall \ell$, where $\mathbf{u}_{k,\ell} \in \mathbb{C}^{N_{k,\ell}}$ forms part of the compact beam-domain precoding vector $\mathbf{u}_k = [\mathbf{u}_{k,1}^{\mathrm{T}}, \ldots, \mathbf{u}_{k,L}^{\mathrm{T}}]^{\mathrm{T}}, \forall k$. We first calculate the beam-domain transmitted signal as $\mathbf{s}_\ell = \sum_k \boldsymbol{\Phi}_k\mathbf{u}_{k,\ell}d_k, \forall \ell$, which has complexity $\mathcal{O}(KLN_{\mathrm{avg}})$. Next, each $\mathbf{x}_\ell = \mathbf{V}\mathbf{s}_\ell$ is obtained efficiently by applying an N-point fast Fourier transform (FFT) to $\mathbf{s}_\ell$, with a complexity of $\mathcal{O}(N\log_2 N)$ per BS. Thus, the total implementation complexity of BSP is $\mathcal{O}(KLN_{\mathrm{avg}} + LN\log_2 N)$.

For comparison, generating $\mathbf{x} = \sum_k \mathbf{p}_k d_k$ directly from spatial-domain precoders $\mathbf{p}_k, \forall k$ incurs a complexity of $\mathcal{O}(KLM)$. While BSP introduces an additional $\mathcal{O}(LN\log_2 N)$ due to FFT, this is relatively negligible. In sparse scenarios of massive MIMO systems, since $N_{\mathrm{avg}} < M$, the dominant implementation complexity term $\mathcal{O}(KLN_{\mathrm{avg}})$ of BSP is smaller than that of the spatial-domain precoding method, which is $\mathcal{O}(KLM)$.

(3) Overall complexity: Combining the design and implementation phases, the overall computational complexity of the BSP via Hamiltonian-based optimization is given by $\mathcal{O}(3N_{\mathrm{iter}}K^2LN_{\mathrm{avg}} + KLN_{\mathrm{avg}} + LN\log_2 N)$. This total complexity is significantly lower than that of conventional spatial-domain approaches.

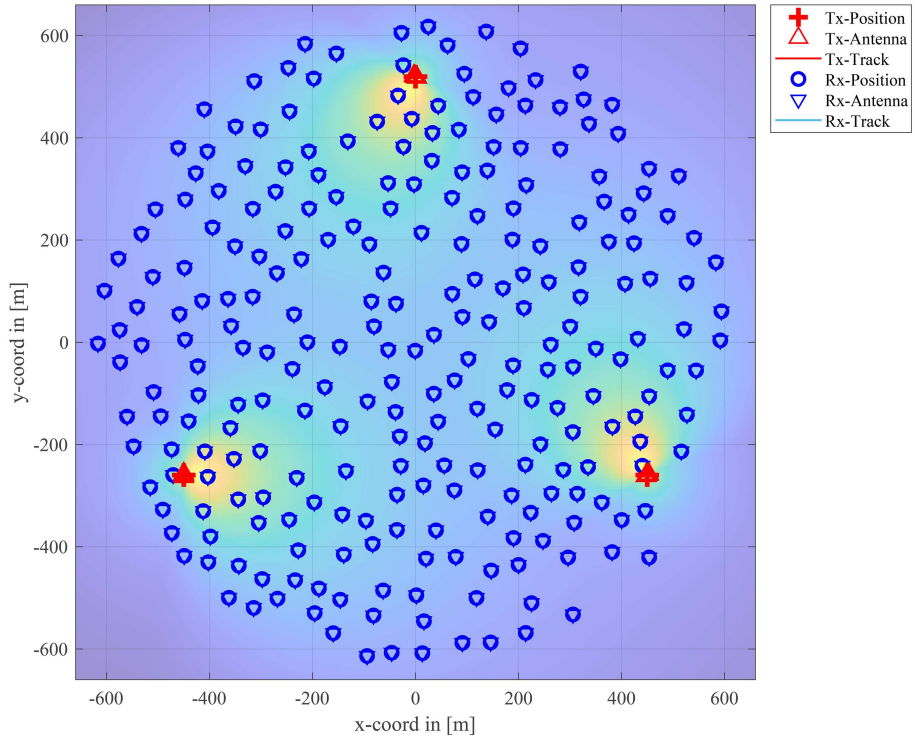

Figure 2 (Color online) The layout of a network massive MIMO system.

5 Simulation results

In this section, we evaluate the performance of the BSP via Hamiltonian-based optimization in the network massive MIMO system. Using the widely-recognized QuaDRiGa model [47], the channel coefficients are obtained under the "3GPP 38.901 UMa NLOS" scenario. The network consists of $L = 3$ BSs, each with $M = 512$ "3gpp-3d" antennas spaced at half-wavelength intervals, and $K = 256$ users each with a single antenna. The center frequency is set at 10 GHz. As depicted in Figure 2, the BSs are located at coordinates $(0, 300\sqrt{3})$, $(-450, -150\sqrt{3})$, and $(450, -150\sqrt{3})$, and the users are distributed uniformly in a circle of radius $360\sqrt{3}$ m. The BSs are elevated at 25 m, while the users are at 1.5 m height. The noise power is set to $\sigma^2 = 1$, while different signal-to-noise ratios (SNRs) are obtained with the varying power budgets $P_\ell = \sigma^2 \text{SNR}$. The fine factor is set as $F = 2$ (unless otherwise stated), which is sufficient to provide good performance [28]. Both of the iterative algorithms are initialized with the maximum ratio transmission (MRT) approach (unless otherwise stated). In this section, Algorithm 1 employs a damping coefficient $\gamma = 0.4$, identity mass matrix $\mathbf{M} = \mathbf{I}$ (unless otherwise stated), and the step size follows an adaptive scheme as $h^{(d+1)} = (\frac{r}{\delta^{(d)}})^{\theta/2} h^{(d)}$ [48, 49], where $\delta^{(d)} = \|\frac{h^{(d)}}{2}(\nabla f(\mathbf{u}^{(d)}) - \nabla f(\mathbf{u}^{(d+1)}))\|$, $h^{(0)} = 0.07, r = 2$ and $\theta = 0.1$.

We first conduct statistical analysis over 100 runs to examine the ratio of the number of non-zero beams per user $\check{N}_k$, to the total number of base station antennas ML under different fine factors F. As shown in Figure 3, the results reveal the sparsity of the beam domain. Moreover, compared to the case of $F = 1$, where the average ratio is approximately 0.3, the sparsity is further enhanced when $F \geqslant 2$, with the mean value dropping to 0.2. However, the additional gain from increasing F beyond 2 is marginal, making $F = 2$ a balanced choice.

Figure 4 illustrates the impact of the fine factor F on the sum-rate performance across different SNR levels. When comparing $F = 1$ with higher values, a significant performance improvement is observed, particularly in the high-SNR regime. This demonstrates the effectiveness of a large F in enhancing system performance. However, when comparing between $F = 2$ and 3, the sum-rate improvements become marginal, as highlighted in the zoomed-in inset. This indicates that while a larger value of F yields better performance, further increasing the fine factor yields diminishing returns. Therefore, selecting $F = 2$ in the numerical simulations of this study is a reasonable choice.

Figure 5 depicts the sum-rate performance of the BSP via Hamiltonian-based optimization in Algorithm 1, labeled as "BSP-Hamiltonian (Converged)", compared with several representative precoding schemes: the spatial-domain "WMMSE" precoder based on Algorithms 1–3 in [11], evaluated both with full convergence and under a complexity-

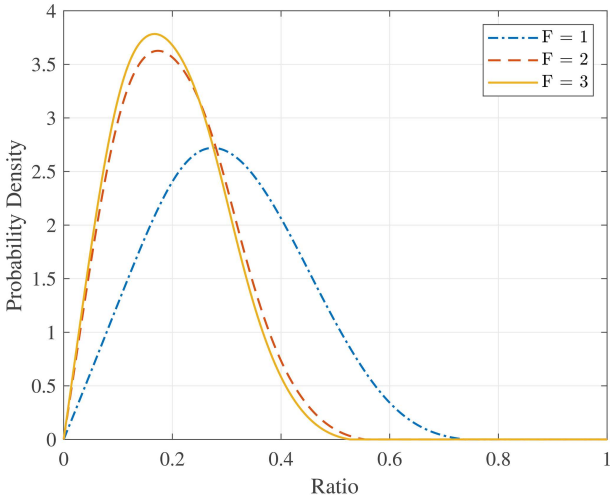

Figure 3 (Color online) Probability density function of the ratio between the number of non-zero beams $\check{N}_k$ and total BS antennas ML under different fine factors.

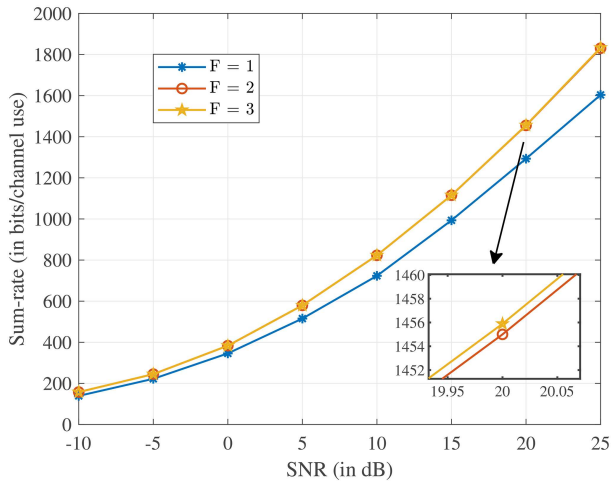

Figure 4 (Color online) Sum-rate with different fine factors.

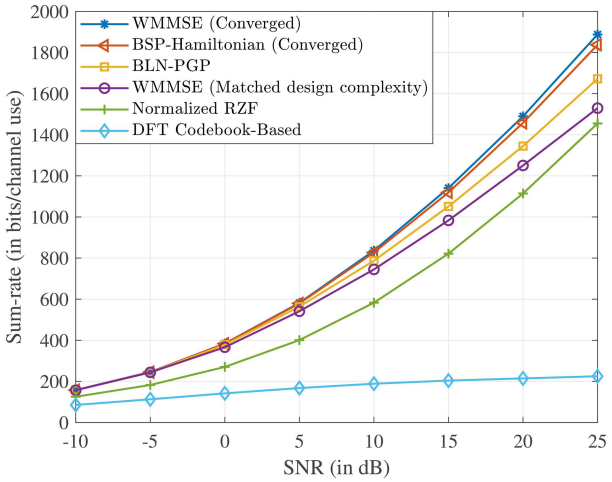

Figure 5 (Color online) Sum-rate of different precoding methods.

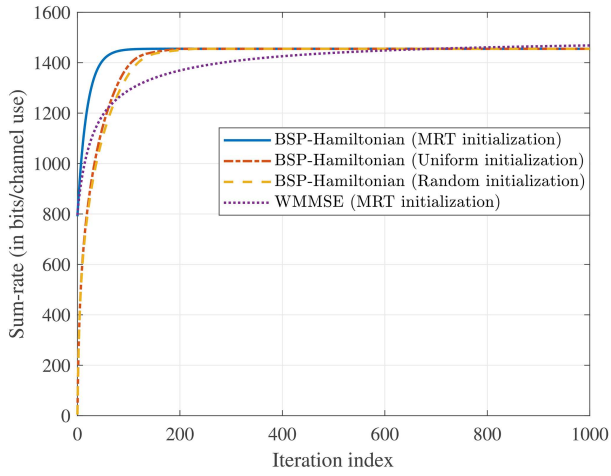

Figure 6 (Color online) Convergence trajectories of BSP-Hamiltonian with different initializations, compared with WMMSE, when SNR = 20 dB.

matched setting where the total design complexity is aligned with BSP-Hamiltonian for fair comparison; the "BLN-PGP" scheme from [50], a recent learning-based precoder derived from unfolding the parallel gradient projection (PGP) algorithm; the "Normalized RZF" method built upon the regularized zero-forcing (RZF) precoding approach proposed in [51], where the centralized precoding directions are computed from the aggregate channel across all BSs, and each BS's precoding vectors are individually normalized to satisfy the local per-BS power constraint; and the "DFT Codebook-Based" method from [52], where each BS selects its precoders from a DFT codebook. As expected, the sum-rate increases monotonically with SNR for all of the methods. Among them, the performance obtained by BSP via Hamiltonian-based optimization is closely matched to that of the WMMSE precoder when both are iterated to convergence, validating the effectiveness of BSP. Under matched computational complexity, the BSP method also outperforms WMMSE by a noticeable margin at high SNR. The BLN-PGP method exhibits competitive performance in the low-to-medium SNR range, but degrades at high SNR due to the less effective interference suppression under dense user scenarios. The normalized RZF method provides moderate performance by mitigating interference, yet its effectiveness is limited by the lack of direct sum-rate optimization. The DFT codebook-based method suffers from coarse beam quantization and rigid structure, resulting in the lowest sum-rate among all methods.

To evaluate how the BSP-Hamiltonian method performs under different initializations, Figure 6 presents its

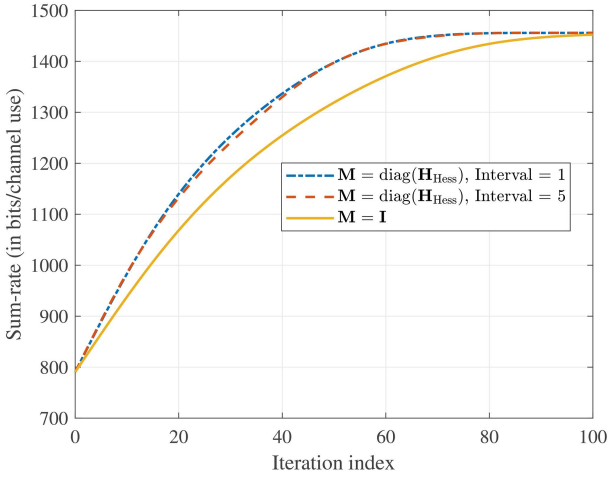

Figure 7 (Color online) Convergence trajectories of BSP-Hamiltonian with different mass matrices $\mathbf{M}$ when SNR = 20 dB.

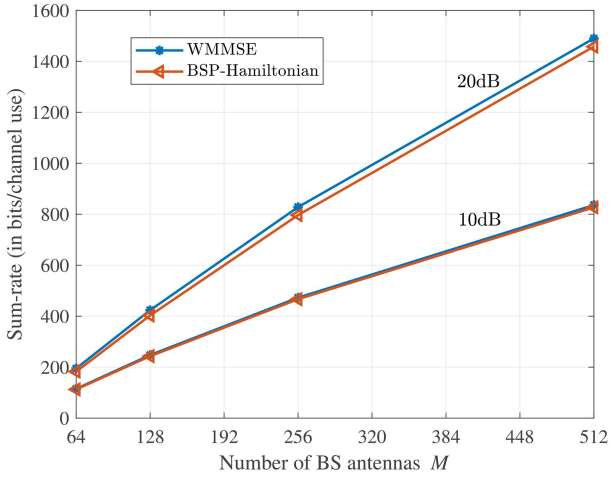

Figure 8 (Color online) Sum-rate performance comparison under different numbers of antennas M.

convergence behavior under MRT initialization, uniform initialization, and random initialization, evaluated at SNR = 20 dB. Among the three settings, the MRT initialization requires fewer iterations to reach convergence. Nevertheless, all settings converge within approximately 200 iterations to achieve comparable sum-rate performance, indicating low sensitivity to initialization in terms of final performance [19]. In contrast, the WMMSE algorithm requires significantly more iterations to stabilize (more than 800). These results demonstrate the convergence efficiency and initialization resilience of the Hamiltonian-based method.

Figure 7 evaluates the impact of the mass matrix $\mathbf{M}$ on the convergence behavior. In order to avoid the prohibitive computational cost for large-dimensional matrix inversions, the approach proposed in [53] is adopted, where $\mathbf{M}$ is set as the diagonal of the Hessian $\mathbf{H}_{\text{Hess}}$. To further alleviate the computational burden, in addition to per-iteration updates of the diagonal entries, a less frequent update interval is adopted, where the diagonal entries of the Hessian are updated every 5 iterations. As shown in the figure, both diagonal Hessian-based cases can accelerate the convergence compared to the identity matrix scheme, while maintaining comparable final sum-rate performance. Moreover, updating the diagonal less frequently still achieves similar convergence behavior, making it a favorable trade-off in practice.

Figure 8 illustrates the sum-rate performance of the BSP compared with the WMMSE method under different numbers of BS antennas M. As observed, the beam-structured approach achieves performance close to the WMMSE solution across a wide range of antenna array sizes, confirming its effectiveness even when M is finite. Notably, while the absolute gap between the two schemes grows mildly due to the overall increase in sum-rate, the relative gap monotonically decreases. These results demonstrate that the proposed beam-structured design remains highly competitive in practical systems with finite antennas.

Figure 9 compares the design and implementation complexity of BSP-Hamiltonian and WMMSE under different numbers of users. As shown in Figure 9(a), BSP-Hamiltonian consistently exhibits lower design complexity than WMMSE. This is due to the exploitation of the beam-domain channel sparsity, which reduces BSP-Hamiltonian's per-iteration complexity, while Hamiltonian-based optimization further reduces the iteration count, making BSP-Hamiltonian more efficient overall. In terms of implementation complexity, as shown in Figure 9(b), BSP-Hamiltonian also maintains a clear advantage over WMMSE. These results highlight the scalability of BSP-Hamiltonian, especially in large-scale systems.

6 Conclusion

In this paper, we proposed a beam-structured precoder design using Hamiltonian-based optimization for network massive MIMO systems. By introducing a beam-based channel model and reformulating the precoder design problem in the beam domain, we demonstrated that the optimal beam-domain precoder of each user can focus on beams corresponding to its own non-zero beam-domain channel elements, thereby reducing the optimization dimension. The resulting problem was handled using a Hamiltonian system, whose dynamical equations efficiently guide the optimization process with stable convergence. Simulation results showed that the proposed BSP achieves

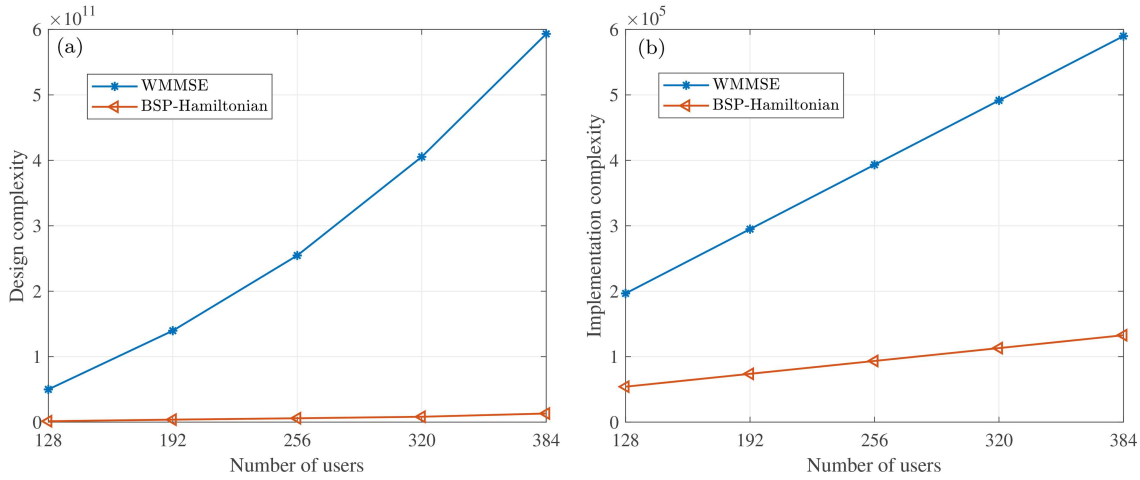

Figure 9 (Color online) Complexity comparison between BSP-Hamiltonian and WMMSE algorithms under different numbers of users. (a) Design complexity; (b) implementation complexity.

a comparable sum-rate performance to the spatial-domain WMMSE precoding, while significantly reducing the computational and implementation complexity. For instance, under the simulation setup with $M = 512$, $K = 256$, and $L = 3$, at 20 dB SNR, our approach attains 97.86% of WMMSE's sum-rate with only 2.31% of its complexity. These results validate the practicality and efficiency of the proposed method in finite-dimensional settings.

Acknowledgements This work was supported by Jiangsu Province Major Science and Technology Project (Grant No. BG2024005), Jiangsu Province Basic Research Project (Grant No. BK20192002), Fundamental Research Funds for the Central Universities (Grant No. 2242022k60007), Key R&D Plan of Jiangsu Province (Grant No. BE2022067), National Natural Science Foundation of China (Grant No. 62322104), Jiangsu Funding Program for Excellent Postdoctoral Talent (Grant No. 2025ZB025), and Basic Research Program of Jiangsu (Grant No. BK20250290).

References

1 Marzetta T L, Larsson E G, Yang H, et al. Fundamentals of Massive MIMO. Cambridge: Cambridge University Press, 2016
2 Dreifuerst R M, Heath R W. Massive MIMO in 5G: how beamforming, codebooks, and feedback enable larger arrays. IEEE Commun Mag, 2023, 61: 18–23
3 You L, Huang Y F, Zhong W, et al. Robust online energy efficiency optimization for distributed multi-cell massive MIMO networks. Sci China Inf Sci, 2023, 66: 132302
4 Ngo H Q, Interdonato G, Larsson E G, et al. Ultradense cell-free massive MIMO for 6G: technical overview and open questions. Proc IEEE, 2024, 112: 805–831
5 Gesbert D, Hanly S, Huang H, et al. Multi-cell MIMO cooperative networks: a new look at interference. IEEE J Sel Areas Commun, 2010, 28: 1380–1408
6 Wang K, Wang X, Xu W, et al. Coordinated linear precoding in downlink multicell MIMO-OFDMA networks. IEEE Trans Signal Process, 2012, 60: 4264–4277
7 Wei Z, Masouros C, Wong K K, et al. Multi-cell interference exploitation: enhancing the power efficiency in cell coordination. IEEE Trans Wireless Commun, 2020, 19: 547–562
8 Zhang Z, Dai L. A joint precoding framework for wideband reconfigurable intelligent surface-aided cell-free network. IEEE Trans Signal Process, 2021, 69: 4085–4101
9 Xiao X, You L, Wang K Z, et al. Distortion-aware beamforming design for multi-beam satellite communications with nonlinear power amplifiers. Sci China Inf Sci, 2024, 67: 162302
10 Atzeni I, Gouda B, Tölli A. Distributed precoding design via over-the-air signaling for cell-free massive MIMO. IEEE Trans Wireless Commun, 2021, 20: 1201–1216
11 Wu Z K, Fei Z S. Precoder design in downlink CoMP-JT MIMO network via WMMSE and asynchronous ADMM. Sci China Inf Sci, 2018, 61: 082306
12 Xiang Z, Gao X, Li K X, et al. Massive MIMO downlink transmission for multiple LEO satellite communication. IEEE Trans Commun, 2024, 72: 3352–3364
13 Xiang Z, Sun R, Gong X, et al. Massive MIMO uplink transmission for multiple LEO satellite communication. IEEE Trans Aerosp Electron Syst, 2025, 61: 4852–4865
14 Vu Q D, Tran L N, Juntti M. Noncoherent joint transmission beamforming for dense small cell networks: global optimality, efficient solution and distributed implementation. IEEE Trans Wireless Commun, 2020, 19: 5891–5907
15 Sun R, Wang C, Lu A A, et al. Precoder design for massive MIMO downlink with matrix manifold optimization. IEEE Trans Signal Process, 2024, 72: 1065–1080
16 Zhang Y, Mitran P, Rosenberg C. Joint resource allocation for linear precoding in downlink massive MIMO systems. IEEE Trans Commun, 2021, 69: 3039–3053
17 França G, Sulam J, Robinson D P, et al. Conformal symplectic and relativistic optimization. In: Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, 2020. 16916–16926
18 Diakonikolas J, Jordan M I. Generalized momentum-based methods: a Hamiltonian perspective. SIAM J Optim, 2021, 31: 915–944
19 Anand A, Brown K R. Hamiltonian-based graph-state ansatz for variational quantum algorithms. Phys Rev A, 2025, 111: 012437
20 Andersen H C. Rattle: a "velocity" version of the shake algorithm for molecular dynamics calculations. J Comput Phys, 1983, 52: 24–34
21 Reich S. Symplectic integration of constrained Hamiltonian systems by Runge-Kutta methods. Technical Report 93-20, Department of Computer Science, University of British Columbia, Vancouver, 1993
22 Leimkuhler B J, Skeel R D. Symplectic numerical integrators in constrained hamiltonian systems. J Comput Phys, 1994, 112: 117–125
23 Zhang Y X, Lu A-A, Xia X-G, et al. Cross-subcarrier precoder design for massive MIMO-OFDM downlink with symplectic optimization. Sci China Inf Sci, 2025, 68: 122301
24 Tse D, Viswanath P. Fundamentals of Wireless Communication. Cambridge: Cambridge University Press, 2005

25 Lu A A, Gao X, Xiao C. Robust linear precoder design for 3D massive MIMO downlink with a posteriori channel model. IEEE Trans Veh Technol, 2022, 71: 7274–7286
26 Shi D, Song L, Zhou W, et al. Channel acquisition for HF skywave massive MIMO-OFDM communications. IEEE Trans Wireless Commun, 2023, 22: 4074–4089
27 Shi D, Song L, Gao X, et al. Beam structured signal detector for HF skywave massive MIMO-OFDM communications. IEEE Trans Wireless Commun, 2024, 23: 19459–19474
28 Yang J, Lu A A, Chen Y, et al. Channel estimation for massive MIMO: an information geometry approach. IEEE Trans Signal Process, 2022, 70: 4820–4834
29 Xie H, Gao F, Zhang S, et al. A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model. IEEE Trans Veh Technol, 2017, 66: 3170–3184
30 Sun R, You L, Lu A A, et al. Precoder design for user-centric network massive MIMO with matrix manifold optimization. IEEE J Sel Areas Commun, 2025, 43: 705–719
31 Strang G. Introduction to Linear Algebra. Wellesley: Wellesley-Cambridge Press, 2016
32 Zhao X, Shi Q. A universal low-dimensional subspace structure in beamforming design: theory and applications. IEEE Trans Signal Process, 2025, 73: 1775–1791
33 Girolami M, Calderhead B. Riemann manifold Langevin and Hamiltonian Monte Carlo methods. J R Statistical Soc Ser B-Statistical Methodology, 2011, 73: 123–214
34 Zhong Y D, Dey B, Chakraborty A. Symplectic ODE-Net: learning Hamiltonian dynamics with control. In: Proceedings of 8th International Conference on Learning Representations (ICLR 2020), Addis Ababa, 2020. 1–17
35 França G, Jordan M I, Vidal R. On dissipative symplectic integration with applications to gradient-based optimization. J Stat Mech: Theory Exp, 2021, 2021: 043402
36 Ghirardelli M. Optimization via conformal Hamiltonian systems on manifolds. J Comput Dyn, 2024, 11: 354–375
37 Goldstein H, Poole C, Safko J, et al. Classical Mechanics. 3rd ed. San Francisco: Addison-Wesley, 2002
38 Hairer E, Wanner G, Lubich C. Geometric Numerical Integration, Structure-Preserving Algorithms for Ordinary Differential Equations. 2nd ed. Berlin/Heidelberg: Springer-Verlag, 2006
39 Dirac P A. Lectures on Quantum Mechanics. New York: Dover Publications, 2001
40 Boyd S, Vandenberghe L. Convex Optimization. Cambridge: Cambridge University Press, 2004
41 França G, Barp A, Girolami M, et al. Optimization on manifolds: a symplectic approach. 2021. ArXiv:2107.11231v2
42 Taylor J R. Classical Mechanics. Sausalito: University Science Books, 2005
43 Plummer H C. An Introductory Treatise on Dynamical Astronomy. Cambridge: Cambridge University Press, 1918
44 Leimkuhler B, Reich S. Simulating Hamiltonian Dynamics. Cambridge: Cambridge University Press, 2004
45 Kaleva J, Tölli A, Juntti M J, et al. Decentralized joint precoding with pilot-aided beamformer estimation. IEEE Trans Signal Process, 2018, 66: 2330–2341
46 Schmidt D A, Shi C, Berry R A, et al. Comparison of distributed beamforming algorithms for MIMO interference networks. IEEE Trans Signal Process, 2013, 61: 3476–3489
47 Jaeckel S, Raschkowski L, Borner K, et al. QuaDRiGa: a 3-D multi-cell channel model with time evolution for enabling virtual field trials. IEEE Trans Antennas Propagat, 2014, 62: 3242–3256
48 Gaines J G, Lyons T J. Variable step size control in the numerical solution of stochastic differential equations. SIAM J Appl Math, 1997, 57: 1455–1484
49 Hairer E, Norsett S P, Wanner G. Solving Ordinary Differential Equations I: Nonstiff Problems. 2nd ed. Berlin: Springer, 1993
50 Zhu M, Chang T H, Hong M. Learning to beamform in heterogeneous massive MIMO networks. IEEE Trans Wireless Commun, 2023, 22: 4901–4915
51 Peel C B, Hochwald B M, Swindlehurst A L. A vector-perturbation technique for near-capacity multiantenna multiuser communication part I: channel inversion and regularization. IEEE Trans Commun, 2005, 53: 195–202
52 Yang D, Yang L L, Hanzo L. DFT-based beamforming weight-vector codebook design for spatially correlated channels in the unitary precoding aided multiuser downlink. In: Proceedings of IEEE International Conference on Communications (ICC 2010), Cape Town, 2010. 1–5
53 Ji X, Zhang Z, Holbrook A, et al. Gradients do grow on trees: a linear-time $O(N)$-dimensional gradient for statistical phylogenetics. Mol Biol Evol, 2020, 37: 3047–3060