AERIAL-IRS-ASSISTED LOAD BALANCING IN DOWNLINK NETWORKS

Shuyi Ren[†♯], *Beichen Huang*[†], *Xiaoyang Li*[♯], *and Kaiming Shen*[†]

[†]School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), China
[♯]Shenzhen Research Institute of Big Data, China

ABSTRACT

This work suggests a joint optimization of the *aerial intelligent reflecting surface* (AIRS) placement, passive beamforming, and *base station* (BS) association to improve the overall data throughput and fairness across downlink heterogeneous cellular networks. Differing from the related works in the literature that just seek to maximize *signal-to-interference-plus-noise ratio* (SINR), the paper takes into account the load balancing between macrocells and small cells. The resulting joint optimization problem is mixed continuous-discrete and has a highly bumpy landscape, so the traditional (sub)gradient-based tools are not suited. We propose a model-free approach based on *adaptive particle swarm optimization* (APSO) and blind beamforming, which recovers the solution from random explorations of the solution space. Simulations show that the proposed algorithm enables balanced traffic for the coexisting macro and small cells, and thereby achieves a higher network utility than the benchmark methods.

1. INTRODUCTION

Intelligent reflecting surface (IRS) is an emerging device that uses a large array of low-cost reflective elements to orient the radio waves toward the intended receiver [1]. While offering advantages like low cost and flexibility, IRS have been demonstrated to have the potential to enhance network efficiency in many aspects [2–5]. The method for improving the overall data rate is studied in [2]. Additionally, an IRS can be positioned near a base station (BS) to broaden signal coverage and simultaneously reduce the cost of the IRS-communication channel, as discussed in [3]. The optimization of BS-user association through changing IRS beamforming has been further investigated for heterogeneous IRS networks in [4] and for multi-access edge computing system in [5].

In the latest development of IRS, Aerial-IRS (AIRS) [6] attracts increasing attention. AIRS uses the unmanned aerial vehicle (UAV) to deploy the IRS in low-amplitude space, aligning the IRS panels parallel to the ground. Compared with the other practice of installing IRS on the exterior of building, which positions the IRS panels perpendicular to the ground, AIRS notably improves the mobility and the scope of coverage, as shown in Fig. 1. While AIRS inherits the advantages of IRS, it also possesses unique properties. AIRS helps maximize the worst-case signal-to-noise ratio [7], worst-case sum secrecy rate [8], and accumulated throughput [9]. [10] reveals a method for AIRS to improve the average achievable rate of the relaying network.

For both IRS and AIRS, the placement and beamforming are the most important properties. Existing studies have shown an optimal placement of IRS leads to a larger coverage of signal [11], and a stochastic geometry distribution of placement leads to a better performance [12]. And as mentioned previously, beamforming could improve the BS-user association [4, 5].

Despite the abundant studies, the effects of placement and beamforming on load balancing have not yet been explored in the literature. In this paper, we focus on optimizing the association between BSs and users for AIRS through iteratively optimizing the placement and beamforming in the scenario of ultra-dense heterogeneous downlink cellular networks where a large number of BSs are deployed [13]. As analyzed in [14], BS-user association is a vital factor for load balancing. Thus our method leads to an optimized load-balancing network.

The main contributions of this paper are three-fold. First, we analyze the effect of AIRS placement and beamforming on association and take both aspects into the design to maximize load balancing. Secondly, we construct and analyze an AIRS model capable of reflecting all incoming signals. It is more practically significant compared to previous studies in which AIRS only communicates with specific BSs or users. Third, we propose an algorithm without requiring any channel state information. In practice, the AIRS controller only requires signal quality information (i.e., data rate) from the user to improve the communication quality.

2. PROBLEM STATEMENT

Our cellular communication system consists of T BSs and U users, where an AIRS is deployed to assist the association, as shown in Fig. 1. We use (x, y) to denote the coordinates of the AIRS location. And the height of the AIRS is fixed in our setting. The AIRS comprises N reflective elements, with the beamforming vector denoted by $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_N)$, where each θ_n, $n = 1, \ldots, N$, represents the phase shift of

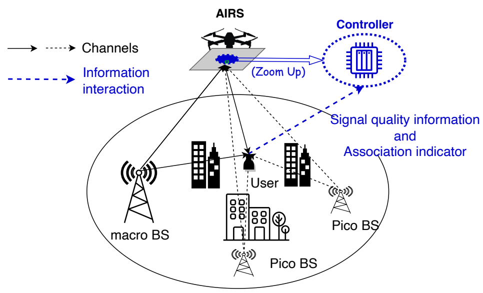


Fig. 1: AIRS-assisted downlink cellular network.

the nth reflective element. Furthermore, we assume that the choice of each θ_n is restricted to a uniform discrete set $\Phi_K = \{0, \omega, \cdots, (K-1)\omega\}$, where $\omega = \frac{2\pi}{K}$ for some given integer $K \geq 2$. We use $\tilde{H}_{t,u} \in \mathbb{C}$ to denote the channel of the direct path between tth BS and the uth user. And the sum of all N reflected path can be expressed with respect to AIRS location (x, y) and beamforming vector $\boldsymbol{\theta}$, i.e.,

$$G_{t,u}(x, y, \boldsymbol{\theta}) = \sum_{n=1}^{N} \tilde{H}_{t,n}(x, y)\tilde{H}_{n,u}(x, y)e^{j\theta_n}, \tag{1}$$

where $\tilde{H}_{t,n}$, $\tilde{H}_{n,u}$ represent the channel of incident signal from the tth BS to the nth AIRS reflective element and the channel of the reflected signal from the nth AIRS element to the uth user, respectively. Then the overall channel between the tth BS and the uth user can be expressed as,

$$H_{t,u}(x, y, \boldsymbol{\theta}) = \tilde{H}_{t,u} + G_{t,u}(x, y, \boldsymbol{\theta}). \tag{2}$$

The mean power of the transmit signal of the tth BS is denoted as P_t, and the mean power of the received signal at uth user can be computed as

$$p_{t,u}(x, y, \boldsymbol{\theta}) = P_t \left|H_{t,u}(x, y, \boldsymbol{\theta})\right|^2. \tag{3}$$

The corresponding signal-to-interference-plus-noise ratio is

$$\mathsf{SINR}_{t,u}(x, y, \boldsymbol{\theta}) = \frac{p_{t,u}(x, y, \boldsymbol{\theta})}{\sum_{b \neq t} p_{b,u}(x, y, \boldsymbol{\theta}) + \delta^2}, \tag{4}$$

where δ^2 is the background noise power.

We use a $T \times U$ binary matrix $\boldsymbol{Q} = [q_{t,u}]$ to denote the association indicator matrix for BS and user association, where $q_{t,u} = 1$ indicates that the uth user is associated with the tth BS, and $q_{t,u} = 0$ otherwise. Each user only associates with one BS at a time, i.e. $\sum_{t=1}^{T} q_{t,u} = 1$. This paper adopts a proportionally fair network utility optimization framework of maximizing the sum log-utility across all the users in the entire network. A key step in the problem formulation is an observation made in [13]. To be more specific, for a given set of users associated with a BS, the round-robin schedule is proportionally fair to balance the resources for the users. In our case, a total of $U_t(\boldsymbol{Q}) = \sum_{u=1}^{U} q_{t,u}$ users are associated with BS t. Applying round-robin schedule to the users, the long-term average rate of the uth user associated with the tth BS is given by

$$R_{t,u}(x, y, \boldsymbol{\theta}, \boldsymbol{Q}) = \frac{1}{U_t(\boldsymbol{Q})} \log\left(1 + \mathsf{SINR}_{t,u}(x, y, \boldsymbol{\theta})\right). \tag{5}$$

We seek the optimal x, y, $\boldsymbol{\theta}$ and $\boldsymbol{Q}$ that maximizing the overall network utility, i.e.,

$$\begin{aligned}
\max_{x,y,\boldsymbol{\theta},\boldsymbol{Q}} \quad & \sum_{u=1}^{U}\sum_{t=1}^{T} q_{t,u} \log\left(R_{t,u}(x, y, \boldsymbol{\theta}, \boldsymbol{Q})\right), && \text{(6a)}\\
\text{s.t.} \quad & (x, y) \in \mathbb{R}^2, && \text{(6b)}\\
& \theta_n \in \Phi_K, && \text{(6c)}\\
& \sum_{t=1}^{T} q_{t,u} = 1, \forall\, u. && \text{(6d)}
\end{aligned}$$

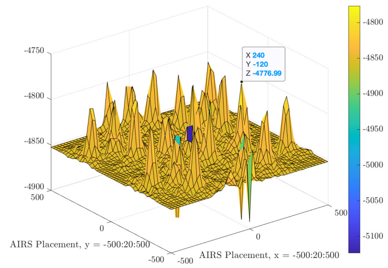
Constraint (6b) states that the potential location of AIRS is continuous in $\mathbb{R}^2$. Constraint (6c) restricts discrete phase shift of AIRS reflective element. Constraint (6d) ensures that each user can only associate with one BS.

Optimizing the AIRS beamformer is numerically difficult because the solution space is of huge size K^N. Our previous work [15] proposes the CSM method, and gives an acceptable solution. However, the real difficulty is optimizing the AIRS location. Even with a fixed beamformer and association indicator, the optimization is challenging due to the complex landscape of the objective function (as shown in Fig. 2). This complexity, characterized by an extremely large Lipschitz constant for its gradient, makes it difficult to apply traditional gradient methods. Furthermore, from a practical standpoint, the complete channel state information required in gradient calculations is too complex to obtain, given the limited computational resources in the network. This inherent difficulty, coupled with the current communication protocols, makes traditional gradient methods infeasible here, necessitating the exploration of alternative methods that can bypass the need for complete channel information.

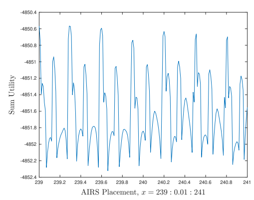
3. PROPOSED APPROACH

To deal with the above challenges, the optimization problem is solved with a black box approach in an iterative manner. One iteration begins with optimizing the placement of the AIRS, assuming a fixed beamformer. Then based on the optimized placement, the beamformer of the AIRS is optimized and go to the next iteration. Whenever the AIRS placement or beamformer is updated, the BS-user association is reassessed using the Max-SINR scheme, i.e., for each user, until no significant improvement of the overall network utility

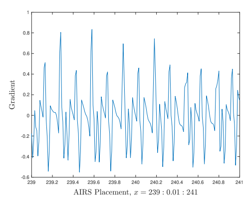
$$t^\star = \arg\max_{t=1,\ldots,T} \mathsf{SINR}_{t,u}, \text{ and } q_{t^\star,u} = 1. \tag{7}$$



(a) $x \in [-500, 500], y \in [-500, 500]$



(b)



(c)

Fig. 2: Landscape of the problem (6). Fig. 2b and 2c show the function value and gradient around the optimal point in Fig. 2a.

A detailed description of iteratively optimizing AIRS placement and beamformer as well as optimizing the association are presented in the following subsections.

3.1. Optimization of AIRS Placement and association

Given fixed association indicator matrix $\boldsymbol{Q}$ and AIRS beamformer $\boldsymbol{\theta}$, the function value of (6) is denoted as $F(x, y)$. We use a stochastic global search method, known as APSO, to optimize the AIRS location. This method helps avoid getting trapped in local optima by probabilistically adopting alternative sub-optimal solutions.

The solving approach begins with the random initialization of potential points. Each potential point has its own velocity which contains the update direction and step size. In the ith iteration, the set of L potential points is denoted as $\mathcal{R}_i = \{r_i^{(\ell)} \in \mathbb{R}^2 | \ell = 1, \ldots, L\}$ and the set of the velocities of these potential points is denoted as $\mathcal{V}_i = \{v_i^{(\ell)} \in \mathbb{R}^2 | \ell = 1, \ldots, L\}$. Further, two types of best-performing points are defined during the search process. The first one, $\mathsf{pBest}^{(\ell)}$, represents the best point up to the latest iteration of the ℓth potential point. The second one, nBest_i, represents the best point among all L potential points in the ith iteration, i.e.,

$$\mathsf{pBest}^{(\ell)} = \arg \max_{j=1,\ldots,i} F(r_j^{(\ell)}), \forall \ell = 1, \ldots, L, \tag{8}$$

$$\mathsf{nBest}_i = \arg \max_{\ell=1,\ldots,L} F(r_i^{(\ell)}), \text{at } i\text{th iteration}. \tag{9}$$

In each iteration, the velocity is updated based on the latest iteration velocity and information from the historical path of potential points. Then the ℓth potential point and its velocity at the ith iteration are updated as

$$\begin{aligned} \boldsymbol{v}_i^{(\ell)} &= w \boldsymbol{v}_{i-1}^{(\ell)} + c_1 \mathsf{rand}_1^{(\ell)} \left(\mathsf{pBest}^{(\ell)} - \boldsymbol{r}_{i-1}^{(\ell)} \right) \\ &\quad + c_2 \mathsf{rand}_2^{(\ell)} \left(\mathsf{nBest}_i - \boldsymbol{r}_{i-1}^{(\ell)} \right), \end{aligned} \tag{10}$$

$$\boldsymbol{r}_i^{(\ell)} = \boldsymbol{r}_{i-1}^{(\ell)} + \boldsymbol{v}_i^{(\ell)}, \tag{11}$$

where w is the inertia weight, c_1 and c_2 are the acceleration coefficients, and $\mathsf{rand}_1^{(\ell)}, \mathsf{rand}_2^{(\ell)} \overset{i.i.d.}{\sim} Uniform(0, 1)$ are two i.i.d. random variables uniformly distributed in $[0, 1]$. The update of inertia weight w and acceleration coefficients c_1, c_2 in each iteration follows the evolutionary state of searching, and thus the searching evolutionary state S needs to be further characterized.

The characterization of the searching evolutionary state at ith iteration is defined based on the geometric relationship between L potential points $\mathcal{R}_i$ and current most optimal point gBest_i. The geometric relationship is characterized by four evolutionary states: "exploration", "exploitation", "convergence" and "jumping-out". Specifically, the evolutionary factor, denoted as τ, is an important parameter in the searching evolutionary state. It is defined based on the mean distance $d^{(\ell)}$ of each potential point $\boldsymbol{r}_i^{(\ell)}$, where $d^{(\ell)} = \frac{1}{L-1} \sum_{l=1, l \neq \ell}^{L} \| r_i^{(\ell)} - r_i^{(l)} \|$. The mean distance of gBest is denoted as $d^{(globe)}$. $d^{(\mathsf{max})}$ and $d^{(\mathsf{min})}$ represent the maximum and minimum mean distances among the $L + 1$ points, which includes all L potential points $\mathcal{R}_i$ and the current most optimal point gBest_i. The evolutionary factor τ is then defined as, $\tau = \frac{d^{(globe)} - d^{(\mathsf{min})}}{d^{(\mathsf{max})} - d^{(\mathsf{min})}} \in [0, 1]$. The detailed mapping from the evolutionary factor τ to evolutionary state follows the membership function presented in [16].

The inertia weight w is updated using a sigmoid mapping $w(\tau) = \frac{1}{1 + 1.5 e^{-2.6\tau}}$ based on the evolutionary factor and state. The acceleration coefficients c_1 and c_2 are adaptively updated in each iteration, and the update processes are interpreted as 'self-cognition' and 'social influence', respectively. The former encourages exploration of local niches, while the latter drives convergence to the current globally optimal region [16]. The sum of c_1 and c_2 is necessitating normalized when their sum exceeds limit.

3.2. Optimization of AIRS Beamforming

After determining the placement of AIRS, the remaining problem is how to design the AIRS beamforming coefficient. The CSM method proposed in our previous work [15] is adopted. We generate M random samples, and use a scalar-valued utility $R^{(m)} = \sum_{u=1}^{U} R_u^{(m)}$, calculated by the sum of all users' overall rate (5), to quantify the performance of each random sample $\boldsymbol{\theta}_m$, m = 1,. . . ,M.

Let $\mathcal{Q}_{nk} \subseteq \{1, \ldots, T\}$ denote the subset of all those random samples with $\theta_{nm} = k\omega$, i.e.,

$$\mathcal{Q}_{nk} = \{t : \theta_{nm} = k\omega\}. \tag{12}$$

The conditional sample mean of $R^{(m)}$ within each subset $\mathcal{Q}_{nk}$ is computed via

$$\widehat{\mathbb{E}}[R|\theta_n = k\omega] = \frac{1}{|\mathcal{Q}_{nk}|} \sum_{m \in \mathcal{Q}_{nk}} R^{(m)}. \qquad (13)$$

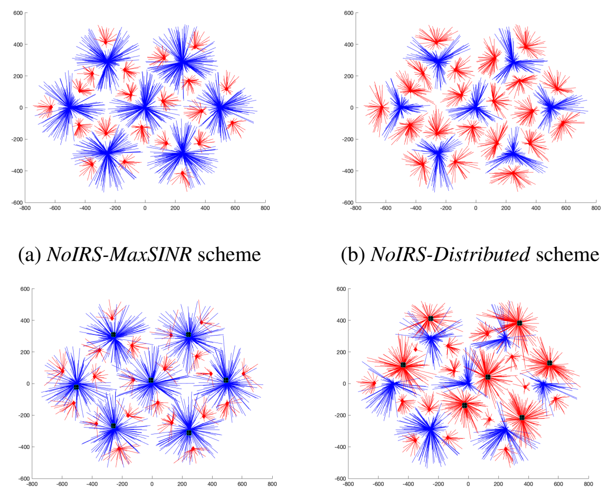
Each θ_n is set as some $\varphi \in \Phi_K$ with the highest conditional sample mean $\widehat{\mathbb{E}}[R|\theta_n = k\omega]$.

4. SIMULATION RESULTS

To validate the proposed optimization method, we employ the multi-cell wireless communication network model for simulations and algorithm testing. This model considers the network's geometry, which is a key factor influencing the optimization of AIRS placement, beamforming, and BS-user association. Specifically, the transmitted powers of macro and pico BSs in the model are $P_{\text{macro}} = 45$dBm and $P_{\text{pico}} = 36$dBm, respectively, and the background noise level is $\sigma^2 = -90$dBm. The AIRS consists of N reflecting elements, each with a size of 0.3m x 0.3m. We control the size of AIRS by changing the number of the reflecting elements N. Additionally, we choose a relatively large number of AIRS to simulate the application in a large area.

In the simulation, we use four models to compare, including models proposed in [17] and [18]. Apart from our proposed scheme, named *AIRS-Proposed*, the other three schemes are considered baseline models. The *NoIRS-MaxSINR* scheme solves the association problem based on the max-SINR scheme without the aid of AIRS. The *NoIRS-Distributed* scheme solves the association problem based on the PF scheme using a distributed method [13]. Lastly, the *AIRS-Random* scheme optimizes the AIRS beamforming based on random placements and determines the association based on the max-SINR scheme. The four schemes are evaluated in the scenario with 7 cells. Each cell contains one macro BS and three pico BSs with fixed locations and 400 uniformly distributed users. The size of each cell is with radius 500m and each cell has one AIRS with $40 * 40$ reflecting elements.

Fig. 3 displays the AIRS-aided association between BSs and users across multiple cells, with the AIRS, macro BSs, and pico BSs represented by a black square, red diamonds, and blue diamonds respectively. The results of four association schemes are sequentially presented in Fig. 3a-3d. It's observable that optimized placement and beamforming of AIRS effectively balance the load between pico and macro BSs, even without adopting the PF scheme. However, random AIRS placement, as shown in Fig. 3c, deteriorates load balancing performance, underscoring the significant impact of AIRS placement on association performance. Fig. 4 depicts load balancing across different schemes. The NoIRS-MaxSINR scheme results in highly unbalanced loads, overloading macro BSs and underutilizing pico BSs. And the issue is mitigated by the NoAIRS-Distributed scheme because the macro BSs and pico BSs almost share the users



(a) *NoIRS-MaxSINR* scheme (b) *NoIRS-Distributed* scheme

(c) *AIRS-Random* scheme (d) *AIRS-Proposed* scheme

Fig. 3: Association between BSs and users under different schemes

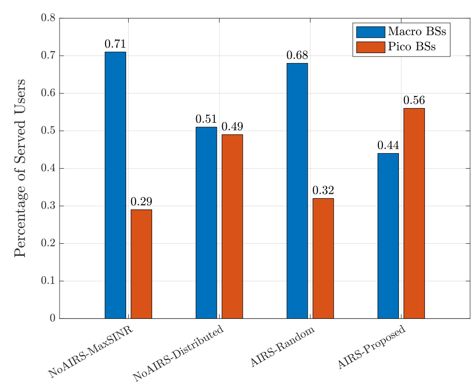


Fig. 4: Number of macro/pico users for various methods

equally. The AIRS-aided association provides a load balancing result similar to the NoAIRS-Distributed scheme, without requiring any intermediary pricing computations between the base station and the user [13, 14].

5. CONCLUSION

We propose a joint approach to optimize AIRS placement, beamforming, and association, aiming at a balanced network load. Our iterative solution to the fluctuating combinatorial optimization problem utilizes APSO and CSM and gives a good result. Compared to other approaches, we employ the max-SINR scheme for BS-user association, where the association decision is made solely based on the received signal quality information. Simulation results demonstrate the effectiveness of our method in enhancing communication efficiency by improving load balance.

6. REFERENCES

[1] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[2] D. Jin, Y. Xiao, Y. Li, G. Shi, and D. Niyato, "Optimizing intelligent reflecting surface-base station association for mobile networks," in *IEEE Int. Conf. Commun. (ICC)*, June 2021.

[3] Y. Huang, W. Mei, and R. Zhang, "Empowering base stations with co-site intelligent reflecting surfaces: User association, channel estimation and reflection optimization," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4940–4955, July 2022.

[4] P. Han, Z. Zhou, and Z. Wang, "Joint user association and passive beamforming in heterogeneous networks with reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3041–3045, Sept. 2021.

[5] W. He, D. He, X. Ma, X. Chen, Y. Fang, and W. Zhang, "Joint user association, resource allocation, and beamforming in RIS-assisted multi-server MEC systems," *IEEE Trans. Wireless Commun.*, pp. 1–1, Aug. 2023.

[6] C. You, Z. Kang, Y. Zeng, and R. Zhang, "Enabling smart reflection in integrated air-ground wireless network: IRS meets UAV," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 138–144, Dec. 2021.

[7] H. Lu, Y. Zeng, S. Jin, and R. Zhang, "Aerial intelligent reflecting surface: Joint placement and passive beamforming design with 3D beam flattening," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4128–4143, Feb. 2021.

[8] W. Wei, X. Pang, J. Tang, N. Zhao, X. Wang, and A. Nallanathan, "Secure transmission design for aerial IRS assisted wireless networks," *IEEE Trans. Commun.*, vol. 71, no. 6, pp. 3528–3540, June 2023.

[9] J. Xu, X. Kang, R. Zhang, Y.-C. Liang, and S. Sun, "Optimization for master-UAV-powered auxiliary-aerial-irs-assisted IoT networks: An option-based multi-agent hierarchical deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 9, no. 22, pp. 22887–22902, Nov. 2022.

[10] X. Pang, M. Sheng, N. Zhao, J. Tang, D. Niyato, and K.-K. Wong, "When UAV meets IRS: Expanding air-ground networks via passive reflection," *IEEE Wireless Commun.*, vol. 28, no. 5, pp. 164–170, Oct. 2021.

[11] S. Zeng, H. Zhang, B. Di, Z. Han, and L. Song, "Reconfigurable intelligent surface (RIS) assisted wireless coverage extension: RIS orientation and location optimization," *IEEE Commun. Lett.*, vol. 25, no. 1, pp. 269–273, Jan. 2021.

[12] X. Shi, N. Deng, N. Zhao, and D. Niyato, "Coverage enhancement in millimeter-wave cellular networks via distributed IRSs," *IEEE Trans. Commun.*, vol. 71, no. 2, pp. 1153–1167, Feb. 2023.

[13] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, June 2013.

[14] K. Shen and W. Yu, "Distributed pricing-based user association for downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1100–1113, June 2014.

[15] S. Ren, K. Shen, Y. Zhang, X. Li, X. Chen, and Z.-Q. Luo, "Configuring intelligent reflecting surface with performance guarantees: Blind beamforming," *IEEE Trans. Wireless Commun.*, vol. 22, no. 5, pp. 3355–3370, Nov. 2023.

[16] Z.-H. Zhan, J. Zhang, Y. Li, and H. S.-H. Chung, "Adaptive particle swarm optimization," *IEEE Trans. Syst., Man, Cybern., Part B (Cybern.)*, vol. 39, no. 6, pp. 1362–1381, Apr. 2009.

[17] W. Tang, M. Chen, X. Chen, J. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421–439, Jan. 2021.

[18] Y. Cheng, W. Peng, C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "RIS-aided wireless communications: Extra degrees of freedom via rotation and location optimization," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6656–6671, Aug. 2022.