

Article

# An Efficient Block Successive Upper-Bound Minimization Algorithm for Caching a Reconfigurable Intelligent Surface-Assisted Downlink Non-Orthogonal Multiple Access System

Xuan Zhou 

School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China; ee202130242019@mail.scut.edu.cn

**Abstract:** With the booming rollout of 5G communication, abundant new technologies have been proposed for quality of service requirements. In terms of the betterment in transmission coverage, mobile edge caching (MEC) has shown potential in reducing the transmission outage. The performance of MEC, meanwhile, can be promisingly enhanced by reconfigurable intelligent surfaces (RIS). Under this context, we explore a system comprising a small base-station (SBS) with limited cache capacity, two users, and one RIS. The SBS transmits the contents from the cache or fetches them from the remote backhaul hub to communicate with users through directional and possibly reflective channels. In this point-to-multipoint connection, non-orthogonal multiple access (NOMA) is applied, improving the capacity of the system. To minimize the outage probability, we first propose a caching policy from entropy perspective, based on which we investigate the beamforming and power allocation problem. The issue, however, is non-convex and involves multi-dimensional optimization. To address this, we introduce an efficient block successive upper-bound minimization algorithm, grounded in Gershgorin's circle theorem. This algorithm aims to find the globally optimal solution for power allocation and RIS beamformer, considering both the channel condition and content popularity. Numerical studies are performed to verify the effectiveness of the proposed algorithm.

**Keywords:** reconfigurable intelligent surface (RIS); mobile edge caching (MEC); non-orthogonal multiple access (NOMA); transmission outage probability



**Citation:** Zhou, X. An Efficient Block Successive Upper-Bound Minimization Algorithm for Caching a Reconfigurable Intelligent Surface-Assisted Downlink Non-Orthogonal Multiple Access System. *Electronics* **2024**, *13*, 791. <https://doi.org/10.3390/electronics13040791>

Academic Editor: Xiongwen Zhao

Received: 29 December 2023

Revised: 6 February 2024

Accepted: 8 February 2024

Published: 18 February 2024



**Copyright:** © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Existing Surveys and Contributions

With the commercial rollout of 5G communication, a significant surge in data traffic, diversified service demands, and the challenge of ubiquitous connectivity in an increasingly interconnected world have highlighted the necessity for advanced communication technology research. Under the context of betterment in network coverage, caching the requested contents at the transmitters near the requesting user, also known as mobile edge caching (MEC), is a promising solution to enhance the spectral efficiency and reduce the response latency. To strengthen the performance of MEC, a reconfigurable intelligent surface (RIS) can be applied to assist MEC with more spectral efficiency. An RIS is composed of up to hundreds of low-cost elements that can purposely modify the phases of incident signals so that the wireless channels are equivalently reconfigured without requiring complex radio-frequency chains, and thus can strengthen the synchronized signals at receiver [1–5]. Meanwhile, non-orthogonal multiple access (NOMA) can be applied in mobile communications. NOMA potentially outperforms orthogonal multiple access [6–9] by allocating asymmetric transmission power to users and adopting successive decoding. In [10], a comprehensive study for current advanced NOMA was performed, where applications such

as RIS-assisted NOMA were investigated. High efficient transmission schemes that combined RIS and NOMA, as well as deep machine learning, were also analyzed. Integrating MEC, RIS and NOMA into wireless communication network research represents a strategic response to the challenges of limited spectrum resources and the growing emphasis on network coverage and energy efficiency.

Initially, MEC provides more various services with higher efficiency. It shifts the communication paradigms from connection-centric to content-centric, which essentially provides a potential insight into the possible explorations of more efficient and flexible advanced techniques [11–13]. For recent progressive studies of MEC, the coordinated caching in unmanned aerial vehicle (UAV) and device-to-device networks was considered, where video contents were coordinately cached at both UAV and user terminals (UT) to maximize the cache utility [14]. A problem to maximize the cache utility by jointly optimizing UAV trajectory and the cache placement in both UAV and UT was formulated, which was then decomposed into three sub-problems and solved by iterative manner to address the integer non-convex optimization issues. In [15], in order to reduce the average perceived latency, joint cache placement and resource allocation for rate splitting multiple access were optimized and sub-optimally solved by dynamic programming and alternative searching method. A three-tier architecture of coordinated caching was introduced and optimization problem to minimize the content delivery latency was formulated, where graph feature embedding and searching were designed to approximate the solutions [16]. In [17], with the aim of improving the spectrum efficiency for profit maximization, nonlinear spectrum pricing schemes were adopted to schedule the caching of proactive pushing contents where an algorithm of modified-value iteration was applied to solve the formulated problem of Markov decision process. A federated learning method was proposed to address the problem of content popularity prediction in caching networks [18].

Meanwhile, RIS can assist MEC and other various configurations of communication systems with better overall performance, for it can strengthen synchronized signals by modifying phases. For example, RIS was combined with the simultaneous wireless information and power transfer (SWIPT) technique to facilitate energy efficiency (EE) [3], where an efficient algorithm was designed to obtain an apparent EE enhancement for RIS-aided multiple-input single-output system with SWIPT under constraints of practical settings. Moreover, mutual-coupling effects were investigated when RIS was employed to further improve the EE for a SWIPT system [19], where optimization problem was formulated to maximize EE by jointly optimizing impedance parameters of the RIS elements and beamforming vectors for the base-station (BS). The problem was then solved by elaborately invoked search methods. Additionally, Hybrid Beamforming was also examined and designed to accommodate the needs of sum rate maximization of multi-user for RIS-assisted SWIPT system [20]. To handle the non-ideal condition in practical condition, the imperfect channel state information was considered for RIS assisted communications, and robust hybrid beamforming vectors were designed to maximize the spectral efficiency for a typical system setting of a multi-antenna BS, multi-RIS, and multi-user [21]. In [22], a tractable analytical framework for RIS-assisted small-cell networks was introduced, based on which a cell-association method was proposed whose mathematical expressions of probabilities for user association and coverage, respectively, in general cases were also derived.

With regard to the assistance of RIS to MEC, there are recently a few works that consider the RIS deployment to enhance caching communications [23–25]. For instance, in order to optimally allocate the cache resource for BS and design the beamforming vectors for RIS, authors have formulated corresponding problems and obtained the sub-optimal solutions by individually searching algorithms for each optimization parameter [23]. RIS-assisted communication was considered in a wireless-powered caching system, where the optimization problem was formulated as a Stackelberg game process and solved by the sub-optimal approach of alternating optimization [24]. Moreover, in order to maximize the ratio of locally cached data over the computation requirement, RIS was utilized to assist the edge caching and computing for a wireless powered mobile network, in which

an exhaustive search was performed to find the optimal energy allocation after designing the caching placement and Lagrange dual method was applied to solve the optimization problem [25].

Though the above works have made significant progress in corresponding research issues, the searching algorithms or approaches adopted therein are generally not globally optimal, i.e., the existing methods usually apply the sub-optimal searching manners, e.g., block coordinated descent, alternating optimization, etc., to individually obtain each optimization parameter. Hence, it is still imperative to investigate more effective algorithms for global optimization of the RIS phase-shifts and power allocation in RIS-assisted caching network, particularly when we consider the promising technique of NOMA transmission mode in the system settings. The main novelty and contributions of this work consist of three-fold:

(1) A caching model based on entropy is proposed for RIS-assisted caching with downlink NOMA whose caching list is theoretically optimal from the view of information theory perspective.

(2) An analytical outage expression is applied and its function associated with the key parameters of transmit power and RIS phase-shifts is mathematically expanded to become more clearly interpreted.

(3) An efficient algorithm of block successive upper-bound minimization (BSUM) based on Gershgorin's circle theorem is proposed and elaborately designed to simultaneously search the globally optimal solutions for transmit power allocation and RIS phase-shifts to minimize the average transmission outage of the system.

## 1.2. Organization

Section 2 describes the wireless communication model, the effective channel condition concerning RIS, and the expressions of the spectral efficiency for users in NOMA scenario. Then, it illustrates the caching model for MEC and Zipf distribution for modeling the popularity of the requested content. Section 3 firstly proposes the caching strategy, based on which the section formulates the non-convex problem of minimizing the transmission outage concerning the beamformer and the power allocation. The effective BSUM algorithm is proposed in Section 4 to address the formulated non-convex problem. Section 5 presents the numerical simulation results. A list of abbreviations and symbols in this paper is presented by Table 1.

**Table 1.** Abbreviations and symbols.

MEC	mobile edge caching	$y_i$	the received signal
RIS	reconfigurable intelligent surface	$h_{eff}^{(i)}$	the effective channel
SBS	small base-station	$r$	the spectral efficiency
NOMA	non-orthogonal multiple access	$o_i$	the outage probability
UAV	unmanned aerial vehicle	$c_m^{(i)}$	the popularity profile
RHS	right-hand-side	$\mathcal{H}$	the caching entropy
w.r.t.	with respect to	$\bar{o}$	the average outage probability
$\lambda_i$	exponential distribution parameter	$\mathcal{L}$	the caching list
$\omega$	passive beamforming vector	$\mathbf{p}$	power allocation vector

## 2. Communication and Caching Models

### 2.1. Wireless Communication Model

**Notation:**  $|x|$  denotes the modulus of the complex number  $x$ ,  $\|\mathbf{x}\|$  denotes the 1-norm of the vector  $\mathbf{x}$ , while  $|\mathcal{K}|$  denotes the cardinality of the set  $\mathcal{K}$ .  $\Pr[\cdot]$  represents the probability of the event in the bracket.  $\angle[\omega]_n$  returns the angle of  $n$ -th complex element of the vector  $\omega$  in the bracket. The superscript  $T$  represents a transpose.

We consider one small base-station (SBS), two users, one RIS, and a backhaul-hub at remote site as shown in Figure 1, where the RIS is deployed to enhance the channel quality between the SBS and users, thereby improving communication performance. Assume that the RIS has  $M$  adjustable reflecting sub-surfaces with each comprising  $L$  elements. Let  $\mathbf{h}_{s,r} = [h_{s,r}^{(1)} h_{s,r}^{(2)} \dots h_{s,r}^{(M)}]^H$ ,  $\mathbf{h}_{r,u}^{(i)} = [h_{r,u}^{(i,1)} h_{r,u}^{(i,2)} \dots h_{r,u}^{(i,M)}]^H$  and  $h_{s,u}^{(i)}$ , where  $i \in \mathcal{I} \triangleq \{1, 2\}$ , denote the small-scale fading coefficients about the channels from SBS to RIS, from RIS to user  $i$ , and from SBS to user  $i$ , respectively, all yielding Rayleigh distribution. Moreover, let  $\mathbf{\Xi} \triangleq \text{diag}(a_1 e^{j\omega_1}, \dots, a_M e^{j\omega_M}) \in \mathbb{C}^{M \times M}$  denote the RIS reflecting matrix, where  $a_n$  and  $\omega_n$ ,  $n \in \mathcal{M} \triangleq \{1, \dots, M\}$ , are the amplitude and phase-shift for the  $n$ -th sub-surface at RIS, accordingly. Without loss of generality, the amplitude of each sub-surface is normalized to a unit. Denote by  $\alpha_{s,u}$  and  $\boldsymbol{\alpha}_{s-u}^{(i)} = [\alpha_{s-u}^{(i,1)} \alpha_{s-u}^{(i,2)} \dots \alpha_{s-u}^{(i,M)}]^H$ , respectively, the large-scale fading factors for the directional and the reflecting links from the SBS to user  $i$ , where each large-scale factor consists of the path-loss and shadowing fading coefficients. Denote by  $x_i \in \mathbb{C}$  for  $|x_i| = 1$  the transmission signal from the SBS to user  $i$ ,  $p_t$  the transmission power of the SBS at the reference distance of one meter (m), and  $y_i$  the received signal. Then, the received signal can be expressed by

$$y_i = (\alpha_{s,u} h_{s,u}^{(i)} + \mathbf{h}_{s,r}^H \mathbf{\Xi} (\boldsymbol{\alpha}_{s-u}^{(i)} \odot \mathbf{h}_{r,u}^{(i)})) p_t x_i + n_i, \tag{1}$$

where  $\odot$  stands for the Hadamard product, and  $n_i \sim \mathcal{CN}(0, \sigma_i^2)$  is the additive white Gaussian noise of zero-mean and variance  $\sigma_i^2$  at the receiver after band-pass filtering. Based on the communication model given by (1), the effective channel from SBS to user  $i$  can be expressed as

$$\begin{aligned} h_{eff}^{(i)} &\triangleq \alpha_{s,u} h_{s,u}^{(i)} + (\boldsymbol{\alpha}_{s-u}^{(i)} \odot \mathbf{h}_{r,u}^{(i)})^H \boldsymbol{\Theta}_i \boldsymbol{\omega} \\ &= \alpha_{s,u} h_{s,u}^{(i)} + (\boldsymbol{\alpha}_{s-u}^{(i)} \odot \mathbf{h}_{s-u}^{(i)})^H \boldsymbol{\omega}, \end{aligned} \tag{2}$$

where  $\boldsymbol{\Theta}_i \triangleq \text{diag}(h_{s,r}^{(1)}, h_{s,r}^{(2)}, \dots, h_{s,r}^{(M)})$ , while  $\boldsymbol{\omega} \triangleq [e^{j\omega_1} \dots e^{j\omega_M}]^H$  which performs as a phase-adjustable passive beamforming vector;  $\mathbf{h}_{s-u}^{(i)} = [h_{s-u}^{(i,1)} h_{s-u}^{(i,2)} \dots h_{s-u}^{(i,M)}]^H$  is referred to as the cascaded channel about the link of SBS-RIS-user  $i$ .

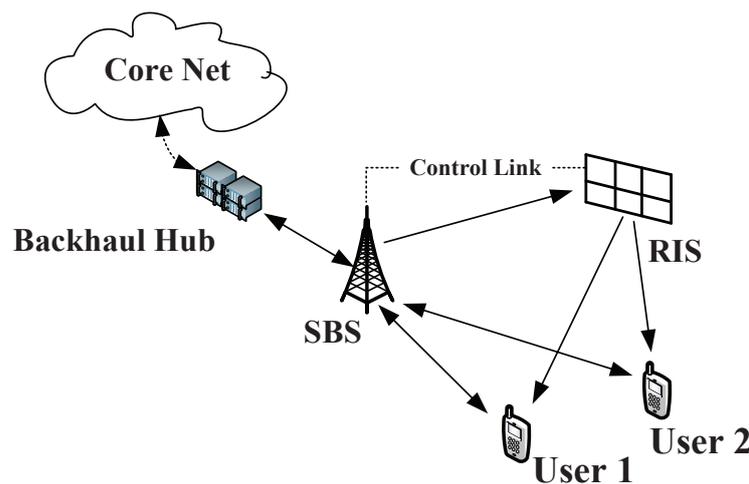


Figure 1. Illustration of the RIS-aided mobile edge caching.

A down-link NOMA transmission scheme is adopted for conveying the contents from SBS to user. Assuming that user 1 locates at the near side while user 2 at the far side, we decode the signals for user 2 first by considering the interference from user 1 as noise. Then, the signals for user 1 are decoded by subtracting the known signals of user 2 (though two users are considered to simplify the conceptual interpretation of NOMA decoding, the investigated outcomes can be extended to more users by pairing the users as commonly

assumed in NOMA transmission). Normally, the power allocation for user 1 and 2, denoted by  $p_1$  and  $p_2$ , is asymmetric. Then, the spectral efficiency for two users can be expressed as

$$r_1 = \log_2\left(1 + g_1(\omega, p_1)/\sigma_1^2\right) \tag{3}$$

$$r_2 = \log_2\left(1 + g_2(\omega, p_2)/(g_1(\omega) + \sigma_2^2)\right), \tag{4}$$

where  $g_i(\omega, p_i) \triangleq p_i|h_{eff}^{(i)}(\omega)|^2$  for  $i = 1, 2$ , in which  $p_1 > 0$ ,  $p_2 > 0$  and  $p_1 + p_2 = p_t$ . Moreover, the outage probability for user 1 and 2 is given by

$$o_i = \Pr[r_i < r_o], i \in \mathcal{I}, \tag{5}$$

where  $r_o$  is a minimum threshold of the quality of service.

### 2.2. Caching Model

Let  $\mathcal{K} \triangleq \{1, 2, \dots, K\}$ , where  $|\mathcal{K}| = K$ , be the set that comprises all content indexes requested by both users. The SBS has a limited cache space of  $C_x$ , where  $C_x < K$ . Consequently, it can only cache partial of the contents requested by both users. It is assumed that the contents are requested by the users with different interests, which is reasonable since different users have varied tastes in content. Then, sorted in descending popularity order, the content set for user  $i$  is given by  $\mathcal{X}_i = \{x_1^{(i)}, x_2^{(i)}, \dots, x_K^{(i)}\}$ , where  $x_m^{(i)}$  for  $m \in \mathcal{K}$  is not necessary to equal  $x_m^{(j)}$  if  $i \neq j$  since the users have their corresponding interests in contents as mentioned earlier. In addition, Zipf distribution is adopted to model the requested frequency for each content. Specifically, the popularity profile of the  $m$ -th content in  $\mathcal{X}_i$  can be written as

$$c_m^{(i)} = \frac{m^{-\xi_i}}{\sum_{n=1}^K n^{-\xi_i}}, \tag{6}$$

where  $m \in \mathcal{K}$  and  $\xi_i > 0$  denotes the Zipf parameter for user  $i$ . Note that  $\xi_i \neq \xi_j$  for  $i \neq j$  since we assume that the users have different interests for the contents. It can be observed by (6) that the Zipf parameter has major impact on and has formulated the contents' popularity profiles, which will weight more for the more popular contents with a larger Zipf parameter, i.e., the more popular contents will be requested more if  $\xi_i, \forall i \in \mathcal{I}$ , grows, and vice versa. Moreover, after sorted in a descent order of popularities, the content sets for users 1 and 2 are accordingly given by  $\mathcal{X}_1 = \{x_1^{(1)}, x_2^{(1)}, \dots, x_K^{(1)}\}$  and  $\mathcal{X}_2 = \{x_1^{(2)}, x_2^{(2)}, \dots, x_K^{(2)}\}$  where  $x_m^{(i)}$  is not necessary to equal  $x_m^{(j)}$  if  $i \neq j$  as aforementioned that the users have their corresponding interests for contents.

### 3. Problem Formulation

We aim to deliberately design a passive beamforming vector for RIS associated with the cached contents to minimize the average outage probability of the system. First, we investigate the strategy for allocating cache resources. It is assumed that the most popular  $L$  contents for user 1 and  $(C_x - L)$  contents for user 2 have been cached. To derive the best passive beamforming vector for RIS during the NOMA transmission, we first discuss the caching strategy from information theory perspectives for the problem formulation. First, we consider the question of how the contents shall be cached within the limited cache space at the SBS. Consider the caching solution, whose entropy can be expressed as

$$\mathcal{H} = - \sum_{m=1}^L c_m^{(1)} \log c_m^{(1)} - \sum_{n=1}^{C_x-L} c_n^{(2)} \log c_n^{(2)}. \tag{7}$$

From information perspectives, the better solution is to cache in a way such that the maximum  $\mathcal{H}$  is obtained. Then, we have the caching list in the space as

$$\mathcal{L} = \left\{ m, i \mid m = \arg \max_{m \in \mathcal{X}_i} (\mathcal{H}) \forall i \in \mathcal{I} \right\}. \tag{8}$$

Such a solution is somewhat counter-intuitive. One might argue that larger information corresponds to lower popularity, potentially leading to the caching of less popular content. However, this is not the truth based on the fundamental fact that entropy is the average information. Though less popularity does result in higher information, the overall average is proportionally smaller. Hence, the proposed solution is always optimal, based on information theory. Note that re-caching a content can be avoided by placement-schedule and, therefore, we neglect the discussion of the case that contents may be doubly cached in the caching list.

Since the channel power is an exponential random variable, the outage probability for the corresponding communication link can be written as

$$o_i = \Pr \left[ \frac{g_i(\omega, p_i)}{(g_{i-1}(\omega, p_{i-1}) + \sigma_i^2)} < \tau \right], \forall i \in \mathcal{I}, \tag{9}$$

where  $\tau = 2^{r_0} - 1$  and  $g_0(\omega, p_0) = 0$ . If requested by the user, the average outage probability is expressed by

$$\bar{o} = \frac{1}{2} \left( \sum_{m, i \in \mathcal{L}} c_m^{(i)} o_i + \sum_{m, i \notin \mathcal{L}} c_m^{(i)} o' \right), \tag{10}$$

where  $o_i$  denotes the outage probability from SBS to the  $i$ -th user, and  $o'$  denotes the two-hop cascading outage probability from the backhaul-hub to SBS and then to the requesting user.

Based on (9), by referring to [26] (Appendix I) and after a small mathematical manipulations, the closed-form expression of statistic outage probability for the links from SBS to user 1 and 2 can be expressed by

$$o_i = 1 - a_i e^{-\frac{\lambda_i(\omega) \sigma_i^2}{p_i} \tau}, \tag{11}$$

where  $a_1 = 1$  and  $a_2 = \frac{\lambda_2(\omega) p_2^{-1} \tau}{\lambda_2(\omega) p_2^{-1} + \lambda_1(\omega) p_1^{-1}}$ . Next, we discuss the impact of  $\omega$  on  $\lambda_i(\omega)$ , so that (11) can be expanded to an explicit form associated with the investigated parameter  $\omega$ . Since the small-fading channels are assumed to yield Rayleigh distribution, the power gains about the channels from SBS to users are exponential random variables, i.e.,  $|h_{eff}^{(i)}(\omega)|^2 \sim \exp(\lambda_i(\omega))$ , where  $\lambda_i$  for  $\forall i \in \mathcal{I}$  are exponential distribution parameters represented by the inversed average power gains of the channels impacted  $\omega$ , as described by the following proposition.

**Proposition 1.** Under the afore-assumption of quasi-static channels, for commonly applied Rayleigh channels, the exponential distribution parameter for the effective channels from SBS to the  $i$ -th user in (11) can be expressed as

$$\begin{aligned} \lambda_i(\omega) = & \left( \alpha_{s,u}^2 / \lambda_{s,u}^{(i)} + \sum_{j \in \mathcal{M}} \alpha_{s,u} \alpha_{s-u}^{(i,j)} J_0 \left( \omega_j + \angle h_{s-u}^{(i,j)} - \angle h_{s,u}^{(i)} \right) \right. \\ & \left. + \sum_{j,k \in \mathcal{M}} \alpha_{s-u}^{(i,j)} \alpha_{s-u}^{(i,k)} J_0 \left( \omega_j + \angle h_{s-u}^{(i,j)} - \omega_k - \angle h_{s-u}^{(i,k)} \right) \right)^{-1}, \end{aligned} \tag{12}$$

where  $\lambda_{s,u}^{(i)}$  denotes the inversed average power gain of  $h_{s,u}^{(i)}$ , and  $J_0$  represents the Bessel function of zeroth order.

**Proof.** (*Sketch of Proof*): The first term at right-hand-side (RHS) of (12) is straightforward via the definition of the exponential distribution parameter associated with the directional link, i.e., the average channel power of the link. Moreover, based on the Jakes's uniform scattering model [27], which can represent common Rayleigh channels, and without considering Doppler phase-shifts due to the assumption of quasi-static channels, the second term at RHS corresponds to the sum of cross-correlations between the signals of each reflecting-link and the directional-link, while the third term at RHS corresponds to the sum of cross-correlations among the signals of the reflecting-links (all phases in the equation refer to the synchronized phases at the receiver).  $\square$

Observing (12), it becomes apparent that after adjusting the phases with the associated RIS phase-shifts  $\omega_i$ , where  $i \in \mathcal{M}$ , the smaller the phase differences among the signals of reflecting and directional links, the larger  $J_0$  will be in both terms on the RHS. This, in turn, leads to a smaller exponential distribution parameter  $\lambda_i(\omega)$ . Our ultimate goal is to minimize  $\bar{\delta}$  by optimizing the passive-beamforming vector  $\mathbf{w}^T$  and the transmission power allocation  $\mathbf{p}^T$ . Since the two-hop cascading outage  $o'_i$  comprises the probability that the outages of two-hop transmissions simultaneously occur, and the outage for each hop individually occurs, based on the probability relations, it follows that  $o'_i = o_i + o_0 - o_i o_0$ , where  $o_0$  is the outage probability of the transmission from the backhaul-hub to SBS and is independent to  $o_i, \forall i \in \mathcal{I}$ . Then, (10) can be rewritten by

$$\bar{\delta} = \frac{1}{2} \left( \sum_{i \in \mathcal{I}} o_i + o_0 \sum_{i, j \notin \mathcal{L}} (1 - o_i) c_j^{(i)} \right), \quad (13)$$

where  $o_i$  is given by (11). Then, to achieve the design goal that minimizes  $\bar{\delta}$ , the problem is formulated as

$$(\mathbf{P1}) \quad \min_{\mathbf{w}} \bar{\delta}(\mathbf{w}) \quad (14)$$

$$\text{s.t.1} : 0 \leq \angle[\omega]_n \leq 2\pi, \quad \forall n \in \mathcal{M} \quad (15)$$

$$\text{s.t.2} : \|\mathbf{p}\| = p_t, \quad (16)$$

in which  $\mathbf{w} \triangleq [\omega^T \mathbf{p}^T]^T$ . (P1) is a non-convex problem since both optimized variables  $\mathbf{p} = [p_1 \ p_2]^T$  and  $\omega$  are coupled in the object function. Though the optimal solutions  $\omega^*$  and  $\mathbf{p}^*$  can be obtained by exhaustive searching, it may lead to exponential computational complexity.

## 4. Proposed Methods and Algorithms

### 4.1. Proposed Solution by an Efficient BSUM Algorithm

A sub-optimal solution of the formulated problem can be solved by stochastic approximation of the beamforming, and then by one-dimensional searching for power allocation. However, such sub-optimal solution may exhibit a noticeable gap from the optimal one, because the optimization of beamforming vectors is coupling with the optimization of the power allocation. In order to simultaneously optimize the parameters, we propose an efficient BSUM algorithm. First, two dual parameters  $\boldsymbol{\mu} = [\mu_1 \ \mu_2]^T \in \mathbb{R}^{2 \times 1}$  and  $\kappa$  are introduced. Then, a dual problem based on (P1) is expressed as

$$\mathcal{L}_{\bar{\delta}}(\mathbf{w}, \boldsymbol{\mu}) = \bar{\delta}(\mathbf{w}) + \boldsymbol{\mu}^T [-\omega \ \omega - 2\pi] + \kappa(\mathbf{1}^T \mathbf{p} - p_t). \quad (17)$$

Then, an efficient BSUM algorithm is proposed to solve the dual problem (17). Commonly, BSUM adopts an approximation obtained by Taylor's expansion [28]. However, not all of the Taylor expansion beyond the first-order can render the objective function convex. As such, traditional BSUM usually adopts the first-order Taylor expansion as the upper-bound for the object function. Though such an approximation ensures convexity in the

transformed problem, it may lead to a solution significantly distant from the optimal one due to the substantial gap between the approximation and the original problem.

Likewise, many existing optimization techniques, e.g., BCD algorithm and Newton’s method, are not applicable for the considered system. The lack of constructing an upper-bound function in BCD algorithm results in higher complexity when meeting non-convex problems. As for Newton’s method, it is particularly for solving unconstrained optimization problems, rather than the formulated one. To better overcome the non-convex issue, we propose an efficient BSUM approach that utilizes the second-order expansion of the objective function while approximating the upper-bound of the expansion using Gershgorin’s circle theorem, which contributes to the simultaneous search of the optimal beamforming vectors  $\omega$  and power allocation  $\mathbf{p}$ , namely  $\mathbf{w}$ . The efficient BSUM approach tends to search a global optimal solution rather than a local optimal one in general cases, on account of its simultaneous search of multi variables and sufficient initial points for searching. The comparison in terms of the applicable scenarios between the existing optimization algorithms and the proposed BSUM algorithm is summarized by Table 2.

**Table 2.** Comparison of applicable scenarios of algorithms.

Algorithm Name	Applicable Scenario
Newton’s Descent Method	Convex, Unconstrained optimization problems
BCD Algorithm	Non convex, Only suboptimal solution
The Traditional BSUM	Non convex, Only suboptimal solution
The Proposed BSUM	Non convex, Generally optimal with sufficient initial points

Let  $\mathcal{H}$  denote the Hessian matrix of the object function in (P1) valued at an instance of  $\mathbf{w}$  (denoted by  $\mathbf{w}_0$ ) and  $\Gamma_i = \sum_{j=1, j \neq i}^N |\{\mathcal{H}\}_{ij}|$ , where  $\{\mathcal{H}\}_{ij}$  is the  $i$ -th row and  $j$ -th column element of  $\mathcal{H}$ . We propose the following proposition before introducing the efficient BSUM.

**Proposition 2.** Suppose a diagonal matrix  $\Lambda_m \in \mathbb{R}^{M \times M}$ , whose diagonal element  $m_{ii} = 0$  if  $(\{\mathcal{H}\}_{ii} - \Gamma_i) \geq 0$ ; while  $m_{ii} = (\nu(\Gamma_i - \{\mathcal{H}\}_{ii}))$  if  $(\{\mathcal{H}\}_{ii} - \Gamma_i) < 0, \forall i$ , in which  $\nu \geq 1$ . Then, for properly chosen  $\nu$ ,  $\hat{\delta} = \bar{\delta}(\mathbf{w}_0) + \nabla \bar{\delta}(\mathbf{w} - \mathbf{w}_0) + (\mathbf{w} - \mathbf{w}_0)^T (\mathcal{H} + \Lambda_m)(\mathbf{w} - \mathbf{w}_0)$  is an upper-bound approximation for the second-order Taylor’s expansion of the function  $\bar{\delta}(\mathbf{w})$  at  $\mathbf{w}_0$  and it is convex.

**Proof.** (Sketch of Proof): Let  $\{\lambda_{\mathcal{H}}\}_i$  and  $\{\lambda_{(\mathcal{H} + \Lambda_m)}\}_i$  denote the  $i$ -th eigen-values of  $\mathcal{H}$  and  $(\mathcal{H} + \Lambda_m)$ , respectively. Then, according to Gershgorin’s circle theorem, we have  $(\{\mathcal{H}\}_{ii} - \Gamma_i) \leq \{\lambda_{\mathcal{H}}\}_i \leq (\{\mathcal{H}\}_{ii} + \Gamma_i)$  while  $0 \leq \{\lambda_{(\mathcal{H} + \Lambda_m)}\}_i \leq (2\Gamma_i)$ . Hence,  $\{\lambda_{(\mathcal{H} + \Lambda_m)}\}_i$  is positive semi-definite and is greater than or equal to  $\{\lambda_{\mathcal{H}}\}_i$ , indicating that  $\hat{\delta}$  is convex and serves as an upper-bound approximation for the second-order Taylor’s expansion of  $\bar{\delta}$ . On the other hand, it can be easily proved that the step-size equals  $(\mathcal{H} + \Lambda_m)^{-1} \nabla \mathbf{w}$  for searching the minimum of  $\hat{\delta}$ . This indicates a larger  $\nu$  corresponds to a smaller searching step-size. By selecting  $\nu$  appropriately, we can make the step-size sufficiently small, rendering higher-order terms in Taylor’s expansion of  $\bar{\delta}$  negligible. It follows that  $\hat{\delta}$  is an upper-bound of  $\bar{\delta}$ . This completes the proof.  $\square$

In Proposition 1,  $\nu$  is inversely proportional to the searching step-size towards the gradient (descent) direction. Though theoretically,  $\nu \geq 0$  ensures the searching moves towards the slope down-hill direction, it does not promise  $\hat{\delta}$  is the upper-bound of the object function  $\bar{\delta}$ . Consequently, the search may lead to an increasing point of  $\bar{\delta}$ . However, by choosing  $\nu \geq 1$  appropriately, the conditions required to perform the proposed BSUM are satisfied.

#### 4.2. Pseudo Codes of the Proposed Algorithm

We then propose the efficient BSUM algorithm to solve the problem (P1) as described in Algorithm 1.

**Algorithm 1:** The proposed efficient BSUM algorithm

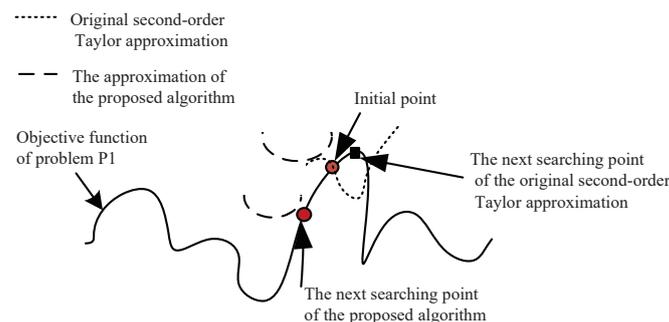
---

**Input:** Random times  $K$ ; search accuracy  $\epsilon$ ;  
step-size parameter  $\nu$ .  
**Output:**  $\mathbf{w}^*$

- 1 **for**  $i \leftarrow 1$  **to**  $K$  **do**
- 2     Initiate a random value  $\mathbf{w}_0$ ; **repeat**
- 3         1: Approximating  $\bar{\delta}$  in (17) by  $\hat{\delta}$  in Proposition 2;
- 4         2:  $\tilde{\mathbf{w}}(i)^* = \arg \min_{\mu, \mathbf{w}} \mathcal{L}_{\hat{\delta}}(\mathbf{w}, \mu)$ ;
- 5         3:  $\mathbf{w}_0 = \tilde{\mathbf{w}}(i)^*$ ;
- 6     **until**  $\|\tilde{\mathbf{w}}(i)^* - \mathbf{w}_0\| \leq \epsilon$ ;
- 7 **end**
- 8  $\mathbf{w}^* = \min(\tilde{\mathbf{w}}(i))$  for  $i = 1, \dots, K$ .

---

As the presented pseudo codes, the proposed BSUM algorithm constructs an upper-bound for the second-order expansion of the object function in (P1). Subsequently, it solves the dual problem through a series of iterative gradient descent steps. A small number of random times  $K$  are set to sample several initial points and choose the best one to better avoid the local minimum. Also, a graphic demonstration is provided in Figure 2 to furthermore interpret the proposed BSUM. As shown, the proposed algorithm always searches in a downhill direction. While in comparison, the searching point of the original second-order approximation may lie in a uphill direction.



**Figure 2.** A graphic demonstration of the proposed algorithm.

## 5. Numerical Simulation Results

In this section, we perform the numerical simulations to verify the effectiveness of the proposed search algorithm. The simulation settings are as follows. The number of RIS sub-surfaces and RIS element is set to 2 and 100, respectively. This setting is considered for a small-scale application with two users and one SBS. The spacing distance between adjacent RIS sub-surfaces is set as a practical distance of half wave-length, which can mitigate the interference among different sub-surfaces. The cache space of the SBS is set as 100, while each user has 80 content items, resulting in a total of 160 items, unless otherwise specified. The Zipf parameters for user 1 and user 2 are set as 1.2 and 0.6, respectively, to see the numerical performance under asymmetric content popularity settings for the two users. The transmission power is 25 dBm. The directional distances from the SBS to user 1 and 2 are 10 m and 20 m, respectively, while the distance from SBS to RIS is 12 m, from RIS to user 1 is 2 m, and from RIS to user 2 is 8 m. The topology of the distance setting is illustrated by Figure 3.  $r_o$  is set as 1 bps/Hz,  $o_0$  is assumed to be 0.05, the noise variances for both users are normalized to one, while the searching parameters  $\nu$  and  $K$  equal 1.5 and 5, correspondingly. In the simulations, the exhaustive searching refers to a brute-force computation to find the optimal among the quantitative beamformer, where the continuous phase in  $[0, 2\pi]$  is quantized to 1024 discrete phase-shift levels. Each user is individually allocated half of transmit power.

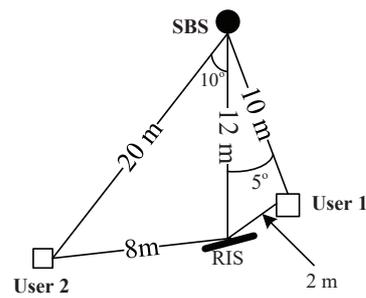


Figure 3. Illustration of the topology setting.

We first perform the simulation to find the optimal RIS beamformer using the proposed algorithm, which relies on the caching solution and takes into account the channel conditions for both users. As shown in Figure 4, we allocate half of the total transmit power to each user and aim to discover the optimal beamformer that minimizes the average outage probability. As illustrated, the average outage probability (i.e., the value of objective function of problem P1) corresponds to different phase-shift IDs, each corresponding to a quantized value. This variability highlights its inherent non-convex nature. However, by our proposed algorithm, the optimal value can be obtained, as evidenced by the comparison to exhaustive searching, where the performance of the proposed algorithm and that of exhaustive searching almost overlap. This also verifies the effectiveness of the proposed method in finding the optimal solution for an irregular and non-convex function (see the outline of the outage function versus different phase-shifts of the RIS beamformer).

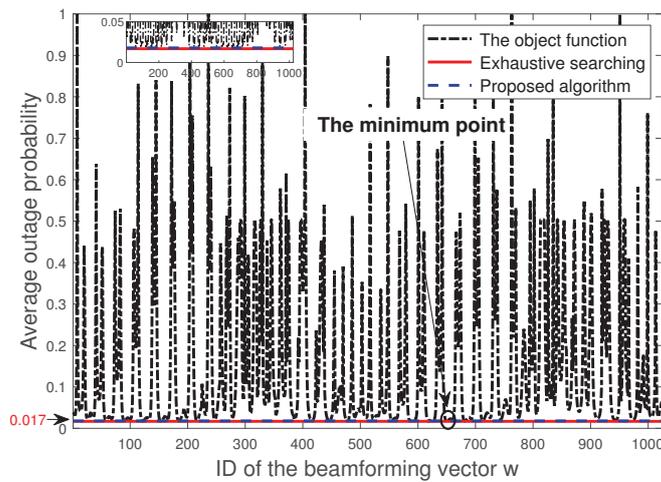
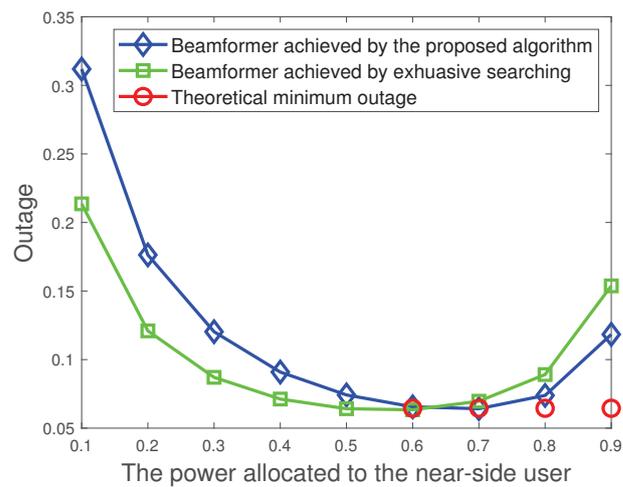


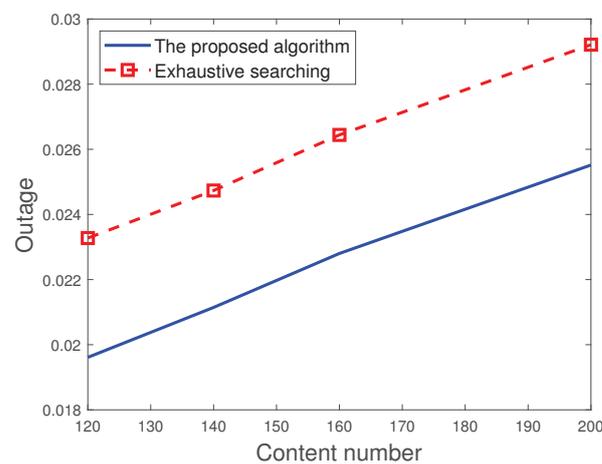
Figure 4. Transmission outage probability versus beamformer ID.

Then, we proceed with the simulation to evaluate the performance of the proposed algorithm in discovering globally optimal solutions for both beamformer and power allocation. Figure 5 indicates the outage probability comparison between the exhaustive searching and the proposed algorithm searching. As can be seen, the outage probabilities for both methods decay as the power ratio allocated to user 1 increases. When the power ratio approaches 0.7, the proposed algorithm achieves its minimal value. In contrast, the outage probability associated with the exhaustive searching slightly grows at the power ratio of 0.7, indicating a worse performance compared to our proposed algorithm. This is because our proposed algorithm can obtain globally optimal solutions for both beamformer and power allocation, whereas exhaustive searching focuses solely on finding the optimal beamformer. Hence, at the power ratio of 0.7, our method adopts its optimally peering beamformer, while the exhaustive searching remains using the sub-optimal beamformer achieved by one-dimensional searching.



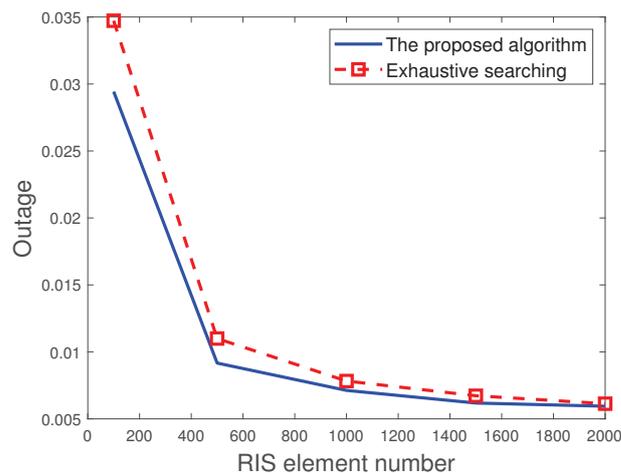
**Figure 5.** Outage versus power ratio allocated to the near-side user.

Next, we compare the performance with different content number. As shown in Figure 6, the outage probabilities of both compared methods increase as the content number grows. This is reasonable since, as the content-set becomes larger, more contents have to be conveyed by fetching them from remote backhaul-hub due to the limited cache space. Consequently, this leads to a worsening of the outage probability, primarily due to the more pronounced two-hop cascading outage probability. However, it is worth noting that the outage performance associated with the proposed method consistently outperforms that of the exhaustive searching. The reason is that our proposed method not only optimizes the beamformer as the exhaustive searching does, but also optimizes the power allocation. This is because the proposed method generally converges the search to globally optimal beamformer and power allocation, while the exhaustive method only searches for the optimal beamformer.



**Figure 6.** Performance comparison with different content number.

Lastly, we change the RIS element number to observe its impact on performance. As depicted in Figure 7, the outage probabilities obtained by both schemes decay significantly as the RIS element number grows. In the comparison, the proposed method generally outperforms the exhaustive searching one, the reason of which is analogous to the explanation provided in Figure 6. The proposed method tends to search and obtain the globally optimal solution for both beamformer and power allocation, while the exhaustive searching, in contrast, is searching for a sub-optimal solution solely for the beamformer vector.



**Figure 7.** Performance comparison with different RIS element number.

## 6. Conclusions

In this work, we investigated RIS-assisted caching communication with the adoption of NOMA for transmission. After proposing the caching strategy grounded in information theory, we designed an efficient searching algorithm based on Gershgorin's circle theorem to deal with the multi-variable non-convex problem. By setting the exhaustive searching as benchmark, we conducted numerical experiments, demonstrating that the proposed algorithm can effectively determine the optimal beamformer and power allocation. For beamformer optimization only, the outage performance of the proposed method almost overlaps that of the exhaustive searching. However, for optimization of both beamformer and power allocation, the proposed method exhibits an apparent performance gain over the exhaustive searching which typically yields a sub-optimal solution. Additional numerical analyses were also carried out to assess the impact of the number of requested contents and RIS elements on the outage probability. Other more complicated wireless channel model with more users in NOMA will be considered in the future work.

**Funding:** This research received no external funding.

**Data Availability Statement:** Data is contained within the article.

**Conflicts of Interest:** The author declares no conflicts of interest.

## References

1. Wu, Q.; Zhang, R. Intelligent Reflecting Surface Enhanced Wireless Network via Joint Active and Passive Beamforming. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 5394–5409. [[CrossRef](#)]
2. Huang, C.; Zappone, A.; Alexandropoulos, G.C.; Debbah, M.; Yuen, C. Reconfigurable Intelligent Surfaces for Energy Efficiency in Wireless Communication. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 4157–4170. [[CrossRef](#)]
3. Tang, J.; Peng, Z.; So, D.K.C.; Zhang, X.; Wong, K.K.; Chambers, J.A. Energy Efficiency Optimization for a Multiuser IRS-Aided MISO System With SWIPT. *IEEE Trans. Commun.* **2023**, *71*, 5950–5962. [[CrossRef](#)]
4. Li, D. How Many Reflecting Elements Are Needed for Energy- and Spectral-Efficient Intelligent Reflecting Surface-Assisted Communication. *IEEE Trans. Commun.* **2022**, *70*, 1320–1331. [[CrossRef](#)]
5. Wang, Z.; Zhao, X.; Tang, J.; Zhao, N.; So, D.K.C.; Zhang, X.Y.; Wong, K.K. Zero-Forcing Beamforming for RIS-Enhanced Secure Transmission. *IEEE Trans. Veh. Technol.* **2023**, *72*, 13666–13670. [[CrossRef](#)]
6. Tang, J.; Luo, J.; Liu, M.; So, D.K.C.; Alsusa, E.; Chen, G.; Wong, K.K.; Chambers, J.A. Energy Efficiency Optimization for NOMA With SWIPT. *IEEE J. Sel. Top. Signal Process.* **2019**, *13*, 452–466. [[CrossRef](#)]
7. Tang, J.; Yu, Y.; Liu, M.; So, D.K.C.; Zhang, X.; Li, Z.; Wong, K.K. Joint Power Allocation and Splitting Control for SWIPT-Enabled NOMA Systems. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 120–133. [[CrossRef](#)]
8. Qi, T.; Feng, W.; Chen, Y.; Nallanathan, A. Double QoS Guarantee for NOMA-Enabled Massive MTC Networks. *IEEE Internet Things J.* **2022**, *9*, 22657–22668. [[CrossRef](#)]
9. Chen, Z.; Tang, J.; Wen, M.; Li, Z.; Yang, J.; Zhang, X.Y.; Wong, K.K. Reconfigurable Intelligent Surface Assisted MEC Offloading in NOMA-Enabled IoT Networks. *IEEE Trans. Commun.* **2023**, *71*, 4896–4908. [[CrossRef](#)]

10. Dai, G.; Huang, R.; Yuan, J.; Hu, Z.; Chen, L.; Lu, J.; Fan, T.; Wan, D.; Wen, M.; Hou, T.; et al. Towards Flawless Designs: Recent Progresses in Non-Orthogonal Multiple Access Technology. *Electronics* **2023**, *12*, 4577. [[CrossRef](#)]
11. Li, C.; Chen, W. Content Pushing Over Idle Timeslots: Performance Analysis and Caching Gains. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 5586–5598. [[CrossRef](#)]
12. Li, C.; Chen, W.; Letaief, K.B. Joint Scheduling of Proactive Caching and On-Demand Transmission Traffics Over Shared Spectrum. *IEEE Trans. Commun.* **2021**, *69*, 8319–8334. [[CrossRef](#)]
13. Xu, X.; Tao, M. Decentralized Multi-Agent Multi-Armed Bandit Learning With Calibration for Multi-Cell Caching. *IEEE Trans. Commun.* **2021**, *69*, 2457–2472. [[CrossRef](#)]
14. Zhang, T.; Wang, Y.; Yi, W.; Liu, Y.; Nallanathan, A. Joint Optimization of Caching Placement and Trajectory for UAV-D2D Networks. *IEEE Trans. Commun.* **2022**, *70*, 5514–5527. [[CrossRef](#)]
15. Hua, Y.; Fu, Y.; Zhu, Q. Latency Minimization for Advanced RSMA-Enabled Wireless Caching Networks. *IEEE Wirel. Commun. Lett.* **2023**, *12*, 1329–1333. [[CrossRef](#)]
16. Wu, D.; Li, J.; He, P.; Cui, Y.; Wang, R. Social-Aware Graph-Based Collaborative Caching in Edge-User Networks. *IEEE Trans. Veh. Technol.* **2023**, *72*, 7926–7941. [[CrossRef](#)]
17. Hui, H.; Chen, W. Joint Scheduling of Proactive Pushing and On-Demand Transmission Over Shared Spectrum for Profit Maximization. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 107–121. [[CrossRef](#)]
18. Tao, Y.; Jiang, Y.; Zheng, F.C.; Wang, Z.; Zhu, P.; Tao, M.; Niyato, D.; You, X. Content Popularity Prediction Based on Quantized Federated Bayesian Learning in Fog Radio Access Networks. *IEEE Trans. Commun.* **2023**, *71*, 893–907. [[CrossRef](#)]
19. Ma, R.; Tang, J.; Zhang, X.; Wong, K.K.; Chambers, J.A. Energy-Efficiency Optimization for Mutual-Coupling-Aware Wireless Communication System Based on RIS-Enhanced SWIPT. *IEEE Internet Things J.* **2023**, *10*, 19399–19414. [[CrossRef](#)]
20. Chen, Z.; Tang, J.; Zhao, N.; Liu, M.; So, D.K.C. Hybrid Beamforming With Discrete Phase Shifts for RIS-Assisted Multiuser SWIPT System. *IEEE Wirel. Commun. Lett.* **2023**, *12*, 104–108. [[CrossRef](#)]
21. Chen, Z.; Tang, J.; Zhang, X.Y.; Wu, Q.; Chen, G.; Wong, K.K. Robust Hybrid Beamforming Design for Multi-RIS Assisted MIMO System With Imperfect CSI. *IEEE Trans. Wirel. Commun.* **2023**, *22*, 3913–3926. [[CrossRef](#)]
22. Zhao, D.; Wang, G.; Wang, J.; Zhou, Z. Reconfigurable Intelligent Surface-Assisted Millimeter Wave Networks: Cell Association and Coverage Analysis. *Electronics* **2023**, *12*, 4270. [[CrossRef](#)]
23. Chen, Y.; Wen, M.; Basar, E.; Wu, Y.C.; Wang, L.; Liu, W. Exploiting Reconfigurable Intelligent Surfaces in Edge Caching: Joint Hybrid Beamforming and Content Placement Optimization. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 7799–7812. [[CrossRef](#)]
24. Chu, Z.; Xiao, P.; Shojafar, M.; Mi, D.; Hao, W.; Shi, J.; Zhong, J. Intelligent-Reflecting-Surface-Empowered Wireless-Powered Caching Networks. *IEEE Internet Things J.* **2022**, *9*, 13153–13167. [[CrossRef](#)]
25. Chu, Z.; Xiao, P.; Shojafar, M.; Mi, D.; Hao, W.; Shi, J.; Zhou, F. Utility Maximization for IRS Assisted Wireless Powered Mobile Edge Computing and Caching (WP-MECC) Networks. *IEEE Trans. Commun.* **2023**, *71*, 457–472. [[CrossRef](#)]
26. Kandukuri, S.; Boyd, S. Optimal power control in interference-limited fading wireless channels with outage-probability specifications. *IEEE Trans. Wirel. Commun.* **2002**, *1*, 46–55. [[CrossRef](#)]
27. Goldsmith, A. *Wireless Communications*, 1st ed.; Cambridge University Press: New York, NY, USA, 2005.
28. Chi, C.Y.; Li, W.C.; Lin, C.H. *Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications*; CRC Press: Boca Raton, FL, USA, 2017.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.