

Energy-Efficiency Optimization with Model Convexification for Wireless Ad Hoc Networks with Multi-Packet Reception Capability

Jianshan Zhou, Daxin Tian, *Senior Member, IEEE*, Guixian Qu, Zhengguo Sheng, *Senior Member, IEEE*, Xuting Duan, and Victor C. M. Leung, *Life Fellow, IEEE*

Abstract—Energy efficiency is a significant requirement of resource management and design optimization in information networks. In this work, we propose an iterative fractional programming framework embedded with a distributed primal-dual extra-gradient projection algorithm, which addresses a wide class of the energy-efficiency optimization problems in wireless ad hoc networks with full-duplex radios and multi-packet reception capability. Specifically, we propose a model convexification mechanism by joining an affine transformation and an exponential transformation into the nonlinear fractional programming, which enables us to deal with the challenge arising from the complexity and non-convex structure of the original problem. With the model convexification, we can map the non-convex power control space into a convex space and equivalently derive a sequence of convex subproblems, which relaxes the convexity assumption widely adopted in the existing literature. We further propose a distributed primal-dual algorithm based on extra-gradient projection to solve the convex subproblem at each iteration of the fractional programming. The convergence of the proposed iterative fractional programming and the distributed optimization method is theoretically proven. Numerical results also verify the proposed method and demonstrate its superior performance over other representative distributed and centralized schemes in terms of achieving global energy efficiency.

Index Terms—Wireless ad hoc networks, energy efficiency, convex optimization, fractional programming, distributed computation.



1 INTRODUCTION

NOWADAYS, many existing and envisioned cyber-physical systems, such as connected and autonomous vehicle platoons [1], cooperative unmanned aerial vehicles (UAVs) [2], [3], energy-harvesting Internet of Things (IoT) [4], and many other multi-hop networks [5], rely on well-established wireless ad hoc networks to enable their applications and services. Besides, to enhance the performance of decoding or demodulating signals from multiple sources simultaneously, multi-packet reception (MPR) protocols have been combined with full-duplex radios for network nodes [3], [6]. These protocols can be realized by using some physical-layer self-interference cancellation technologies such as radio frequency (RF) interference cancellation, digital cancellation, and antenna cancellation. From many successful application domains, such as unmanned aerial vehicle (UAV) networks [3], wireless ad hoc or autonomous networks [7], [8], and wireless multimedia networks [9], [10], it is witnessed that the full-duplex radios and

multi-packet reception capability can bring great benefit to communication and networking systems in terms of system-level performance improvement. Hence, this technical paradigm has received much research attention. However, the introduction of full-duplex radios and multi-packet reception capability into a general wireless ad hoc network inevitably increases the system complexity due to the coupling power interferences from multiple concurrent signal transmitters. This integration poses an important challenge for system design and optimization. In particular, the impact of concurrent interferences on the quality of received signal cannot be fully ignored, even when the network nodes are equipped with advanced self-interference technologies [3], [7], [8].

Energy efficiency is a fundamental and significant requirement of wireless ad hoc networks with full-duplex radios and multi-packet reception capability. Such a system metric can play a critical role in affecting a broad range of communication and computing applications since network nodes are usually energy resource-limited [4], [11], [12]. Specifically, the energy-efficiency optimization objective is formulated as a ratio of the network-wide throughput or data rate over the network power consumption. It is concerned with the performance of the information transmission capacity when a unit of energy is consumed. Hence, a fractional form is used to characterize the energy-efficiency metric [13]–[15]. Its numerator and denominator are inherently two contradictory quantities. In addition, the constraints involved in the energy-efficiency optimization of a wireless ad hoc network usually impose restrictions on maximum power consumption or concurrent physical-layer

- Jianshan Zhou, Daxin Tian, and Xuting Duan are with Beijing Key Laboratory for Cooperative Vehicle Infrastructure Systems & Safety Control, School of Transportation Science and Engineering, Beihang University, Beijing 100191, China (e-mail: jianshanzhou@foxmail.com, dtian@buaa.edu.cn, duanxuting@buaa.edu.cn).
- Guixian Qu is with the Aero-engine System Collaborative Design Center, Research Institute of Aero-Engine, Beihang University, Beijing 100191, China (e-mail: guixianqu@foxmail.com).
- Zhengguo Sheng is with Department of Engineering and Design, the University of Sussex, Richmond 3A09, UK (e-mail: z.sheng@sussex.ac.uk).
- Victor C. M. Leung is with Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, B.C., V6T 1Z4 Canada (e-mail: vleung@ieee.org).

(Corresponding author: Daxin Tian)

interferences.

In mathematics, a general energy-efficiency optimization problem falls into a class of nonlinear constrained optimization (NCO) that can be appropriately solved by using many existing nonlinear optimization algorithms and software in centralized computation, such as interior-point methods, augmented Lagrangian methods, and gradient projection methods [16], [17]. However, in many practical situations, neither the convexity¹ nor the structural separability² can be well satisfied by the energy-efficiency optimization problem. In particular, when we take into account full-duplex radios and multi-packet reception capability, the coupling power interferences from multiple transmitters cannot be neglected. The coupling interference effect is usually captured by the well-known metric, i.e., the signal-to-interference-plus-noise ratio (SINR) on the concurrent channel. As shown in [18], SINR-based utility functions or their linear combinations are neither convex nor concave in the power feasible region due to the impact of the coupling interferences. The presence of non-convexity can make current convex optimization methods (like interior-point methods, sequential quadratic programming, and semidefinite cone programming [19], [20]) less effective or even infeasible in solving a global optimal and feasible point [21]. It remains an open issue to effectively tackle the nonlinear constrained energy-efficiency optimization problem in the presence of non-convex SINR-based utility functions.

In this work, we are motivated to establish a model convexification mechanism for mapping the non-convex system model into a convex form. We join an affine transformation and an exponential transformation into a fractional programming paradigm. Leveraging the model transformation and convexification, we propose a novel distributed optimization-enabled fractional programming framework for the global energy-efficiency maximization of a wireless ad hoc network with multi-packet reception capability. The proposed method provides a deep insight into how to remove the barrier arising from the non-convexity of energy-efficiency optimization in the form of fractional programming. In this way, the proposed framework enables legacy convex optimization techniques to come into play. The main contributions of this work are summarized as follows:

i) We propose an affine transformation for calculating the weighted sum-rate of a wireless ad hoc network with multi-packet reception capability. Using the affine transformation, we show that the network throughput utility function can be equivalently transformed into a simpler weighed-sum structure. This facilitates model convexification.

ii) Based on the transformed utility function, we extend Dinkelbach's method to establish a nonlinear constrained fractional programming framework to address the energy-

efficiency optimization problem. Even though this model does not fall into the category of convex-concave fractional programming, we prove that the framework can still solve the non-convex problem under the condition that the feasibility of solving a sequence of non-convex nonlinear constrained subproblems and the global optimality of solutions for those subproblems are guaranteed.

iii) To address a sequence of non-convex subproblems in the iterative framework, we propose an exponential transformation to convert the non-convex optimization structure into the convex structure. Based on the convexification, we propose a distributed primal-dual extra-gradient projection method that can handle the structural inseparability for practical distributed implementation, and finally prove its global convergence to an optimal solution.

The rest of this paper is organized as follows. Section 2 details a review on related works. Section 3 develops the system model and proposes an affine transformation. Section 4 extends the Dinkelbach's method and proposes a fractional programming framework. In Section 5, we propose the convexification mechanism and a distributed algorithm. Section 6 provides numerical results to validate the performance of the proposed method. Finally, Section 7 concludes this work and remarks future research directions.

2 RELATED WORKS

Admittedly, there exists a voluminous body of literature that aims at the optimal energy efficiency of communication systems ranging from various wireless ad hoc networks to cellular networks. Specifically, since the optimization objective is usually expressed as a fractional utility function, the theory of fractional programming has been widely utilized for inspiring the development of energy-efficiency optimization algorithms. Dinkelbach's method is one of the most famous fractional programming schemes. It had been originally proposed for solving a class of convex fractional programming problems (sometimes termed convex-concave problems), in which the numerator function is required to meet the convexity and the denominator should be concave concerning decision variables [22], [23]. Inspired by Dinkelbach's method or its variants, many researchers have been currently engaged in developing innovative design paradigms to address the energy-efficiency optimization challenges in different communication systems [15]. For example, [24] formulates the objective function as the ratio of the overall spectral efficiency over the total power consumption in a multi-user-multi-relay orthogonal frequency division multiple access (OFDMA) cellular network. Since their model does not incorporate the coupling interference structure, the problem can be simplified as quasi-concave fractional programming in the maximization form [24], and can be solved based on Dinkelbach's method. In [13], [14], Shen et al. have proposed a quadratic transformation similar to the classical Dinkelbach's transform. Their transformation is promising to handle multi-ratio fractional programming problems but it requires additional constraints that the objective function value must stay the same [13], [14]. Besides, the fractional programming method with the quadratic transformation depends on the concave-convex assumption to solve sum-of-functions-of-ratio problems (see Theorems

1. Convexity is a mathematical concept defined from the perspective of geometry. A convex optimization model requires that both its optimization objective function and constraint functions are convex.

2. The structural separability of a constrained optimization model means that its objective function and constraint functions are separable across multiple individuals' decision variables. When the objective and constraint functions of the model are separable, they can be naturally decomposed into a sequence of subproblems, each of which can be solved by a single individual. Thus, the overall model can be addressed in a distributed manner, and the solving algorithm can be well suited for distributed implementation.

3 and 4 in [13]). Another work [25] has proposed a matrix fractional programming method based on the quadratic transformation and applied it to solve the weighted sum of log utilities in a device-to-device (D2D) communication system. Due to the non-convexity of the model, an approximate model of the original problem needs to be properly chosen and solved, and only local convergence can be guaranteed (see Theorem 6 in [25]). In addition, the work [15] provides a comprehensive review of various classes of representative fractional problems including linear fractional problems, concave-convex fractional problems, and max-min fractional problems. Conventional convex fractional programming algorithms are also presented for energy-efficiency designs of different wireless networks [15]. It is observed from the above existing works and those therein that the convexity assumption plays a key role in enabling a wide range of fractional programming algorithms.

Another research focus related to the power control of wireless ad hoc networks is on distributed implementation. From the perspective of practical deployment, distributed computation is more appealing and promising than centralized computation which fully relies on a centralized entity. In particular, when the computing resource of each node in a network is limited, a distributed paradigm can reduce the computational burden undertaken by each node and a power optimization problem can be solved in a multi-node cooperative manner. Besides, in a wireless ad hoc network, every networking node, i.e., a computation agent, may not be able to access other nodes' individual utilities or decision constraint conditions, but they can only communicate the values of their decision variables locally [26]–[28]. In such situation, we need to design a distributed computation framework. Indeed, extensive efforts have been dedicated to investigating distributed optimization in various application contexts including communication, optimization, and control. Many well-known distributed algorithms, as well as their variants, have already been developed such as sub-gradient projection-based methods [29]–[33], primal-dual-based or dual-based methods [28], [34]–[41], diffusion adaptation-based methods [42]–[44], population game dynamics-based methods [45], [46]. However, most of those aforementioned methods, for instance, the prior works [28], [30]–[36], [38]–[41] and the references therein, originally aimed at solving distributed convex or concave optimization. These existing studies heavily depend on the convexity or concavity assumption on their problem formulation (i.e., on both the objective function and the constraints) [47]. Additionally, it can also be observed that some distributed algorithms only deal with unconstrained optimization problems like the works [29], [42]–[44] or focus on the problems only with linear inequalities [38], [46]. Differently, the work [37] has proposed an approximate dual sub-gradient algorithm that is effective to solve a class of non-convex distributed constrained optimization problems. Nonetheless, to guarantee their dual approach to converge, [37] adds a new assumption, termed the singleton dual optimal solution set, that the dual limit must have a single optimal solution in its feasible region.

It is also worth pointing out that the Lagrangian dual relaxation-based approach is a popular way to solve non-convex constrained optimization problems [16], which has

widely spawned different classes of convex optimization algorithms as shown in the aforementioned literature and therein. Unfortunately, when an energy-efficiency optimization problem of a wireless ad hoc network, subject to both individual power constraints and coupling interference constraints, cannot meet the mathematical convexity, convex optimization techniques proposed in the prior works cannot be applied directly. Due to the existence of a positive duality gap, the Lagrangian dual formulation is in general not equivalent to its primal non-convex problem [21], and only local convergence is guaranteed by primal-dual approaches.

To efficiently solve a non-convex optimization problem, a key idea is to equivalently transform the non-convex problem to a convex formulation by a proper model convexification. Much literature such as [5], [48]–[53] has shown that the natural exponential function can be used for development of such a transformation. For instance, early works [48], [49] have utilized the exponential transformation to uncover the hidden convexity or hidden concavity of a class of utility functions in feasible signal-to-interference ratio (SIR) regions. This idea motivates the convex optimization-based power control method for CDMA systems [50]. In existing works [5], [51]–[53], the exponential transformation has also been used for converting a non-convex power control problem into a convex problem in the form of geometric programming that aims at maximizing an objective function of network-wide SIR or minimizing an objective function of total power consumption in cellular networks. In fact, in an earlier work [54], the researchers studied the application of log-convexity in convex geometric programming. In mathematics, the exponential function has a powerful property, i.e., the log-convexity, which can be leveraged for model convexification. This mathematical technique is also called the log-convex transformation. However, the non-convex problem considered in this work is quite different from the above geometric programming problems, since its objective is in the fractional form where the numerator and denominator are two contradictory metrics. The existing exponential transformation technique cannot be directly applied in the non-convex energy-efficiency optimization.

Based on the above observations, we differentiate ourselves from the previous literature by relaxing the convexity assumption in system modeling in this work. Specifically, we propose an affine transformation and combine it with an exponential transformation to convert a constrained non-convex energy-efficiency optimization problem into a sequence of constrained convex optimization sub-problems. We then extend the idea of the original Dinkelbach methodology and propose a distributed primal-dual method based on an extra-gradient iteration mechanism for solving these convex sub-problems. This results in an iterative optimization framework, which guarantees globally-optimal convergence despite the non-convexity in the original problem. In summary, our transformation and optimization method provides a better understanding of how general fractional programming can benefit from the proposed model convexification. It can enrich current non-convex energy-efficiency optimization methodological frameworks and motivate the development of novel algorithms in other related fields.

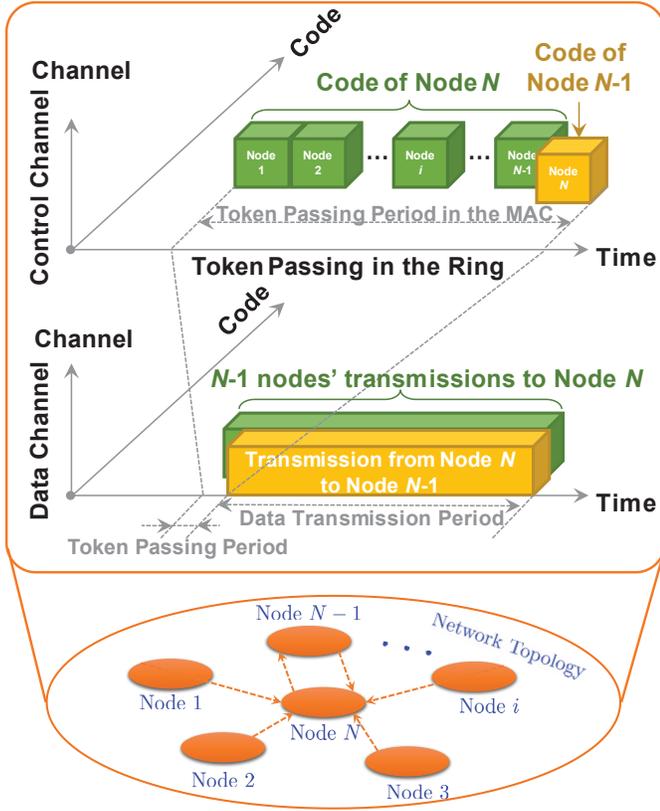


Fig. 1. An exemplary application scenario in which a wireless ad hoc network with full-duplex radios and MPR capability can be realized based on some existing token-based medium access control (MAC) schemes such as [3], [10]. In the MPR scenario, the coupling wireless interferences from transmitters and their self-interferences cannot be neglected.

3 SYSTEM MODEL AND PROBLEM FORMULATION

We consider a wireless ad hoc network where each node can adopt physical-layer self-interference cancellation technologies to realize full-duplex radio and has a code-division multiple-access (CDMA) transceiver capable of MPR capability [3], [7]–[10]. Similar to the existing study like [3], we focus on the scenario consisting of multiple one-hop communication nodes in close proximity, in which one node needs to receive data from several other nodes and this node also has a potential receiver in its neighborhood. It should be remarked that such a paradigm is quite general to cover many real application scenarios such as cluster-based wireless sensor networks. In wireless sensor networks composed of multiple one-hop clusters, a single head of each cluster needs to collect data from multiple sensors within the same cluster via one-hop communication and then a global master node also needs to receive information from different cluster heads. Besides, it is noted that, even though our system model presented here deals with a one-hop communication network, the specific application scenario does not alter our methodology and our proposed scheme can also be used in multi-hop routing scenarios. That is, our scheme can be integrated with many existing routing protocols and media access control (MAC) schemes like the token-based MAC [3] to support multi-hop routing applications. The fact is that the multi-hop routing procedure can usually be divided into

TABLE 1
Main Symbols and Definitions

Symbol	Value
\mathcal{N}	set of network nodes, where $\mathcal{N} \triangleq \{1, 2, \dots, N\}$
\mathbf{p}	column power vector, where $\mathbf{p} \triangleq [p_1, p_2, \dots, p_N]^T$
\mathcal{P}	global feasible power region
ψ_i	self-interference cancellation factor for node i
σ_i^2	average background noise power for node i
$I_i(\mathbf{p})$	physical-layer interference function with i
g_i	channel gain for node i
\mathbf{G}	diagonal matrix composed of channel gains
\mathbf{v}_i	auxiliary column vector for affine transformation
\mathbf{e}_i	unit column vector whose i -th element is 1
α	SINR threshold for successful packet capture
ω_i	normalized pre-allocated bandwidth
$f_i(\mathbf{p})$	data transmission rate for node i 's link
$F(\mathbf{p})$	network-wide weighted-sum rate
$F_{\text{affine}}(\mathbf{x})$	affine transformation function of column vector \mathbf{x}
$\tilde{F}(\mathbf{x})$	\mathbf{x} network throughput based on affine transformation
$\theta(\mathbf{p})$	cost-type energy-efficiency metric
$h(\theta)$	objective function with Dinkelbach's transformation
$\mathcal{M}_2(\theta)$	subproblem model with Dinkelbach's transformation
$\mathcal{M}_3(\theta)$	transformed subproblem model
$F_{\mathcal{M}_2}(\mathbf{x}, \mathbf{y})$	objective function of $\mathcal{M}_2(\theta)$
$\phi(x)$	exponential transformation on x
\mathbf{s}	counterpart of transformed \mathbf{x}
\mathcal{S}	global feasible region of column vector \mathbf{s}
$\phi(\mathbf{s})$	transformation on column vector \mathbf{s}
$\varphi_i(\phi(\mathbf{s}))$	SINR constraint after transformation for node i
$\boldsymbol{\varphi}(\mathbf{s})$	column vector of transformed SINR constraints
$J(\mathbf{s})$	objective function of $\mathcal{M}_3(\theta)$
$\boldsymbol{\lambda}$	column vector of Lagrangian multipliers
\mathcal{U}	box-type feasible set of Lagrangian multipliers
$\mathcal{L}(\mathbf{s}, \boldsymbol{\lambda})$	Lagrangian function of $\mathcal{M}_3(\theta)$
$G(\mathbf{s}, \boldsymbol{\lambda})$	primal-dual gradient function
L	Lipschitz constant for gradient of $J(\mathbf{s})$
L_G	Lipschitz constant for $G(\mathbf{s}, \boldsymbol{\lambda})$

a series of one-hop communications and thus our scheme can come into play in each one-hop communication process. To be specific, we define the number of the nodes in the wireless ad hoc network by N , and the set of the nodes as $\mathcal{N} \triangleq \{1, 2, \dots, N\}$. For any node $i \in \mathcal{N}$, its transmission power is denoted by p_i which is assumed to be ranged from $p_{i,\min}$ to $p_{i,\max}$, i.e., $p_i \in [p_{i,\min}, p_{i,\max}]$. The channel gain for any node $i \in \mathcal{N}$ is denoted by g_i and the average background noise power of the node i is assumed to be σ_i^2 . Accordingly, we let the collection of transmission power levels as a column vector \mathbf{p} , i.e., $\mathbf{p} = [p_1, p_2, \dots, p_N]^T \in \mathcal{P}$ where $\mathcal{P} = [p_{1,\min}, p_{1,\max}] \times [p_{2,\min}, p_{2,\max}] \times \dots \times [p_{N,\min}, p_{N,\max}]$. The main mathematical symbols and their physical meanings are also summarized in Table 1.

As shown in Fig. 1, we consider an exemplary application scenario where a specific node, for instance, the node N , would like to receive data packets transmitted from the other $N-1$ nodes, i.e., from $\{1, 2, \dots, N-1\}$, in a time slot, and simultaneously the node N can also transmit its data to one of the remaining $N-1$ nodes as its receiver, for instance,

the node $N - 1$. The self-interference cancellation effect for the node N is characterized by a coefficient ψ_N . According to the current literature [3], each node's MPR capability can be realized using the CDMA technique, and the MAC implementation is based on a token passing-based scheme. Specifically, a MAC token containing a code list is successively passed from one node to another in the ring. The code list carries a group of codes, i.e., the codes of each node in the ring to receive data packets. In general, the codes, usually called spreading codes, are orthogonal to each other. A transmitter will take the code of its corresponding receiver from the code list in the token when it holds the token and then immediately pass it to the next node in the ring. Multiple nodes with data to send can modulate their base-band signals with different orthogonal spreading codes on the same carrier, meanwhile avoiding serious interferences between their signals. Conversely, when a node employing a CDMA transceiver receives signals from multiple transmitters, it can use a matched filter to extract its targeted signal, as these concurrent signals have different orthogonal codes. In Fig. 1, each node $i, i = 1, 2, \dots, N - 1$, will select the same spreading code, i.e., the code of the node N , from the code list in the MAC token successively passed to themselves, since they have data to send to the same destination, i.e., the node N . At this point, the nodes ranging from 1 to $N - 1$ will cause interferences on the reception of the node N . Besides, for the node N with data to send to the node $N - 1$, it will select the code of the node $N - 1$ from the code list in the token. Due to the orthogonality of the codes corresponding to the nodes $N - 1$ and N , the nodes $i = 1, 2, \dots, N - 1$ concurrently sending data packets to the node N will not incur severe interferences on the reception of the node $N - 1$. Therefore, we present different interference functions on the reception of the nodes $N - 1$ and N , respectively. The interference functions are detailed in two different cases as follows³:

(1) For the transmission link from the node N to the node $N - 1$, the total interference at the receiver $N - 1$ is

$$I_N(\mathbf{p}) = \psi_{N-1} p_{N-1} + \sigma_{N-1}^2. \quad (1)$$

(2) For the transmission from any node $i \in \mathcal{N} \setminus \{N\}$ to N , the interference function associated with i is

$$I_i(\mathbf{p}) = \sigma_N^2 + \psi_N p_N + \frac{1}{N} \sum_{k=1, k \neq i}^{N-1} g_k p_k, \quad (2)$$

where N is also the spreading factor.

For simplicity, based on (1) and (2), we can re-arrange the interference function associated with any node $i \in \mathcal{N}$ as a unified linear form that couples the power controls of all the nodes in the network:

$$I_i(\mathbf{p}) = v_{0,i} + \mathbf{v}_i^T \mathbf{G} \mathbf{p} \quad (3)$$

where $\mathbf{v}_i = [v_{1,i}, v_{2,i}, \dots, v_{N,i}]^T$ and $v_{0,i} = \sigma_{N-1}^2$ for $i = N$ while $v_{0,i} = \sigma_N^2$ for all $i \in \mathcal{N} \setminus \{N\}$. When $i = N$, in

3. For notation simplicity, we use an interference function for each receiver to lump the total interference effects on its reception. The interference function incorporates the interference incurred by the noise power, the receiver's self-interference, and the potential interferences of other nodes with concurrent transmissions. The term "interference" indicates not only the self-interference and concurrent interferences but also the noise impact.

the vector $\mathbf{v}_N, v_{N-1,N} = \frac{\psi_{N-1}}{g_{N-1}}$ while $v_{l,N} = 0$ for all $l \in \mathcal{N} \setminus \{N-1\}$. When $i \in \mathcal{N} \setminus \{N\}$, in the vector $\mathbf{v}_i, v_{N,i} = \frac{\psi_N}{g_N}, v_{l,i} = \frac{1}{N}$ for all $l \in \mathcal{N} \setminus \{i, N\}$, while $v_{i,i} = 0$. \mathbf{G} is a diagonal matrix, i.e., the gain matrix $\mathbf{G} = \text{diag}\{g_1, g_2, \dots, g_N\}$. $I_i(\mathbf{p})$ is also called a coupling linear interference function.

Given a signal to interference plus noise ratio (SINR) threshold $\alpha > 0$, the signal can be correctly received over a transmission link from a transmitter $i \in \mathcal{N}$ when the SINR associated with this link is greater than the given threshold α [3], [7], [55], [56], i.e., satisfying the following inequality

$$\text{SINR}_i = \frac{g_i p_i}{I_i(\mathbf{p})} \geq \alpha. \quad (4)$$

Combining (3) and (4) then yields a group of inequalities

$$\alpha v_{0,i} + \alpha \mathbf{v}_i^T \mathbf{G} \mathbf{p} - g_i p_i \leq 0, \quad \forall i \in \mathcal{N}. \quad (5)$$

Now, we denote a unit column vector by \mathbf{e}_i , where the i -th component is equal to 1 while the other components are zero, i.e., $[\mathbf{e}_i]_i = 1$ while $[\mathbf{e}_i]_l = 0$ for all $l \in \mathcal{N} \setminus \{i\}$. We rearrange (5) as follows

$$\alpha v_{0,i} + (\alpha \mathbf{v}_i - \mathbf{e}_i)^T \mathbf{G} \mathbf{p} \leq 0, \quad \forall i \in \mathcal{N}. \quad (6)$$

Referring to the channel capacity of information transmission in the Shannon's sense, we derive the rate of the transmission link associated with node i and given its pre-allocated bandwidth $\omega_i > 0$ as follows

$$f_i(\mathbf{p}) = \omega_i \log_2 \left(1 + \frac{g_i p_i}{I_i(\mathbf{p})} \right), \quad \forall i \in \mathcal{N}. \quad (7)$$

Following (7), we can formulate the overall throughput utility of the wireless ad hoc network with full-duplex radios and MPR capability as the following weighted sum-rate

$$\begin{aligned} F(\mathbf{p}) &= \sum_{i=1}^N f_i(\mathbf{p}) = \sum_{i=1}^N \omega_i \log_2 \left(1 + \frac{g_i p_i}{I_i(\mathbf{p})} \right) \\ &= \sum_{i=1}^N \omega_i \log_2 \left(1 + \frac{g_i p_i}{v_{0,i} + \mathbf{v}_i^T \mathbf{G} \mathbf{p}} \right). \end{aligned} \quad (8)$$

Remark: We have formulated the system model that describes the weighted sum-rate of a wireless ad hoc network with a specific transmission-pair and interference topology as shown in Fig. 1. The network system and its variants have been widely investigated in the current literature such as [3], [7]–[10], and the system model can be applied to a wide variety of application scenarios, such as UAV-based wireless networks [3]. Even though our system modeling is based on a specific application scenario as illustrated in Fig. 1, the specific application scenario does not alter our methodology. The system modeling approach and the proposed optimization method can be adapted to other specific scenarios such as wireless sensor networks and vehicular ad hoc networks. It is also noted that we have modeled the fundamental communication requirement of the network nodes from the physical-layer perspective, using a general SINR threshold reception model (4) that is widely used in the current literature [3], [7], [55], [56]. It is because we consider full-duplex radios and MPR capability in the network that the physical-layer interferences caused by the concurrent data transmissions of multiple nodes cannot be neglected. Thus, we focus on the SINR-based channel capacity model.

This physical-layer model enables us to characterize the channel capacity concerning concurrent nodes' interferences and thus is widely adopted in the recent literature.

Remark: In Fig. 1, the network nodes with the CDMA implementation and the token passing-based MAC scheme are able to get the knowledge of the model parameters, e.g., the channel gains, and update the parameters for transmission optimization. Specifically, the token can also carry a channel gain list consisting of the channel gains or other channel state information (CSI) [3]. When a node receives the token and takes out its targeted code from the code list of the token, it will also update the channel gains in the channel gain list of the token according to its own channel measurements (i.e., the signal samples-based statistics). After that, the node will immediately pass the token to the next node in the ring, and the next token holder will repeat the similar procedure. In this way, the channel gains or other CSI parameters in the passed token can also be updated periodically and will be used for the nodes' transmission optimization.

Remark: We further highlight that the token passing in the ring is operated in the common control channel rather than in the data channel as in Fig. 1. The functionality of the common channel and the data channel is different. Concurrent data transmissions occur in the data channels. The common channel is usually used to broadcast some MAC messages such as tokens and beacons [3], [57], so as to enable nodes to find the network and establish the connection. For example, in [57], a vehicular node uses the common channel to request to join a cluster-based vehicular network and the data channel to transmit data. A cluster head will decide whether the requesting node is admitted to join the network or not. Once the requesting node is allowed to access the channel, it is added to the service list or the token list. In [3], [57], a transmission scheduling scheme is determined based on the information about the set of nodes accessing the same channel. Similar to the literature, we can construct the set of nodes that would like to send data according to the information on channel access before the nodes send data in the data channel.

3.1 Affine Transformation of Throughput Utility

When looking into (8) carefully, we are motivated to propose an affine transformation to convert the throughput utility into a simpler structure so as to facilitate the optimization design. Specifically, given a positive parameter a_i for each $i \in \mathcal{N}$, there must exist a real number x_i such that

$$a_i x_i = g_i p_i + (v_{0,i} + \mathbf{v}_i^T \mathbf{G} \mathbf{p}), \quad \forall i \in \mathcal{N}. \quad (9)$$

Let $\mathbf{A} = \text{diag}\{a_1, a_2, \dots, a_N\}$ and $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N]$, $\mathbf{v}_0 = [v_{0,1}, v_{0,2}, \dots, v_{0,N}]^T$, and $\mathbf{B} = (\mathbf{I} + \mathbf{V})^T$ where \mathbf{I} is a $N \times N$ unit matrix. We derive from (9) the following relationship

$$\mathbf{A} \mathbf{x} - \mathbf{v}_0 = \mathbf{B} \mathbf{G} \mathbf{p}. \quad (10)$$

Therefore, we can establish an affine transformation from \mathbf{x} to \mathbf{p} by $F_{\text{affine}} : \mathbf{x} \mapsto \mathbf{p}$, i.e.,

$$\mathbf{p} = F_{\text{affine}}(\mathbf{x}) = \mathbf{G}^{-1} \mathbf{B}^{-1} \mathbf{A} \mathbf{x} - \mathbf{G}^{-1} \mathbf{B}^{-1} \mathbf{v}_0. \quad (11)$$

This affine transformation F_{affine} establishes a one-to-one mapping relationship between the decision variables for power control, \mathbf{p} , and another set of decision variables, \mathbf{x} . This one-to-one mapping function provided by the affine transformation is bijective, indicating that we can uniquely determine the set of power decision variables \mathbf{p} by determining the new decision variables \mathbf{x} . At this point, we can substitute the power decision variables, \mathbf{p} , with the new decision variables, \mathbf{x} , to transform an original EE optimization problem with respect to \mathbf{p} into another equivalent one with respect to \mathbf{x} . Once we obtain optimal \mathbf{x} from the transformed problem, we can also determine optimal \mathbf{p} immediately.

Moreover, letting $c_{0,i} = v_{0,i} - \mathbf{v}_i^T \mathbf{B}^{-1} \mathbf{v}_0$, $\mathbf{b}_i^T = \mathbf{v}_i^T \mathbf{B}^{-1}$, and $\tilde{I}_i(\mathbf{x}) = c_{0,i} + \mathbf{b}_i^T \mathbf{A} \mathbf{x}$ for all $i \in \mathcal{N}$, we have

$$\tilde{I}_i(\mathbf{x}) = c_{0,i} + \mathbf{b}_i^T \mathbf{A} \mathbf{x} = v_{0,i} + \mathbf{v}_i^T \mathbf{G} \mathbf{p} = I_i(\mathbf{p}). \quad (12)$$

Combining (12) and (9), we obtain

$$\frac{a_i x_i}{\tilde{I}_i(\mathbf{x})} = \frac{g_i p_i}{I_i(\mathbf{p})} + 1, \quad \forall i \in \mathcal{N}. \quad (13)$$

Thus, the network throughput (8) can be rewritten as

$$\tilde{F}(\mathbf{x}) = \sum_{i=1}^N \tilde{f}_i(\mathbf{x}) = \sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i x_i}{\tilde{I}_i(\mathbf{x})} \right). \quad (14)$$

In addition, we let $\mathbf{p}_{\min} = [p_{1,\min}, p_{2,\min}, \dots, p_{N,\min}]^T$ and $\mathbf{p}_{\max} = [p_{1,\max}, p_{2,\max}, \dots, p_{N,\max}]^T$. By using (10), the bound constraint $\mathbf{p}_{\min} \leq \mathbf{p} \leq \mathbf{p}_{\max}$ is equivalent to the following bound constraint on \mathbf{x}

$$\mathbf{A}^{-1}(\mathbf{B} \mathbf{G} \mathbf{p}_{\min} + \mathbf{v}_0) \leq \mathbf{x} \leq \mathbf{A}^{-1}(\mathbf{B} \mathbf{G} \mathbf{p}_{\max} + \mathbf{v}_0), \quad (15)$$

and the linear inequality (6) is also equivalent to

$$(\alpha + 1)c_{0,i} + ((\alpha + 1)\mathbf{b}_i - \mathbf{e}_i)^T \mathbf{A} \mathbf{x} \leq 0, \quad \forall i \in \mathcal{N}. \quad (16)$$

Remark: With the proposed affine transformation F_{affine} , we can rearrange the transmission data rate in the form of $\log_2 \left(\frac{a_i x_i}{\tilde{I}_i(\mathbf{x})} \right)$ rather than $\log_2 \left(1 + \frac{g_i p_i}{I_i(\mathbf{p})} \right)$. The mathematical structure and property of the bound constraints and the linear inequalities are not changed under the affine transformation. The proposed affine transformation is beneficial to reduce the modeling complexity and, more importantly, makes the convexification transformation of a wide class of energy-efficiency optimization model possible, which will be shown in the following sections.

3.2 Non-convex Energy-Efficiency Optimization Model

In general, the energy-efficiency optimization problem of a wireless ad hoc network in the presence of concurrent wireless interferences can be modeled as

$$\begin{aligned} \mathcal{M}_0 : \quad & \min_{\mathbf{p}} : \theta(\mathbf{p}) = \frac{\sum_{i=1}^N p_i}{\sum_{i=1}^N \omega_i \log_2 \left(1 + \frac{g_i p_i}{I_i(\mathbf{p})} \right)} \\ & \text{s.t.} \quad \begin{cases} \alpha v_{0,i} + (\alpha \mathbf{v}_i - \mathbf{e}_i)^T \mathbf{G} \mathbf{p} \leq 0, \quad \forall i \in \mathcal{N}; \\ \mathbf{p} \in \mathcal{P}. \end{cases} \end{aligned} \quad (17)$$

In \mathcal{M}_0 , $\theta(\mathbf{p})$ denotes the energy consumption of the network to achieve per unit information transmission, which is expected to be minimized as much as possible to achieve

a high energy efficiency. The linear inequality constraint indicates the QoS-related requirement on the individual link SINR, and the bound constraint represents the actual physical restriction on the transmission power of each node.

With the proposed affine transformation in Subsection 3.1, we can equivalently convert \mathcal{M}_0 into another form as

$$\mathcal{M}_1 : \min_{\mathbf{x}} : \theta(\mathbf{x}) = \frac{\sum_{i=1}^N (C_i x_i + D_i)}{\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i x_i}{\tilde{I}_i(\mathbf{x})} \right)}$$

$$s.t. \begin{cases} (\alpha + 1)c_{0,i} + ((\alpha + 1)\mathbf{b}_i - \mathbf{e}_i)^T \mathbf{A}\mathbf{x} \leq 0, \\ \forall i \in \mathcal{N}; \\ \mathbf{x} \in \mathcal{X}. \end{cases} \quad (18)$$

where \mathcal{X} represents the feasible solution set of the bound constraints on \mathbf{x} as shown in (15). The coefficients $\mathbf{C} = \text{col}\{C_i, i \in \mathcal{N}\}$ and $\mathbf{D} = \text{col}\{D_i, i \in \mathcal{N}\}$ are determined by the affine transformation (11), i.e.,

$$\begin{cases} \mathbf{C} = \mathbf{A}^T (\mathbf{G}^{-1}\mathbf{B}^{-1})^T \mathbf{1} \\ \mathbf{D} = -\mathbf{v}_0^T (\mathbf{G}^{-1}\mathbf{B}^{-1})^T \mathbf{1}. \end{cases} \quad (19)$$

where $\mathbf{1}$ is the $N \times 1$ column vector all of whose elements are identical to 1.

From (17) and (19), it can be seen that both the original model and the resulting model after the affine transformation have the similar mathematical structure. To simplify notations, we use the set \mathcal{Q} to denote the feasible region of the solution \mathbf{x} for the model \mathcal{M}_1 in the following sections.

Remark: \mathcal{M}_1 is a nonlinear and nonconvex fractional programming problem and does not fall into the typical class of convex-concave fractional programming problems due to the non-concavity of its objective function. In general, it is difficult to directly solve a globally optimal solution to this problem by using conventional convex optimization methods. It is also noted that traditional Dinkelbach's methods depending on a strong convexity assumption may fail in solving such a non-convex model.

Remark: In the above model, we mainly focus on the power consumption of network communications. In reality, the power consumption of each node's circuit also contributes to the total energy consumption. Thus, the circuit power consumption can also be incorporated to develop a more complicated energy-efficiency optimization model. Nevertheless, this will require a precise model that can characterize the functional relationship between the circuit power consumption and some complicated factors such as the transmission power, the communication rate, and the hardware-specific conditions. It is meaningful to model the circuit power consumption and integrate it with the energy efficiency optimization, which is left as our future work.

4 ITERATIVE FRACTIONAL PROGRAMMING

Dinkelbach's method had been originally proposed for solving a class of convex-concave fractional programming problems, which is widely applied in many other fields due to its low complexity. The advantage of Dinkelbach's method is that it does not need to introduce additional constraints. Therefore, we extend Dinkelbach's method to develop an iterative algorithmic framework in our targeted context. More importantly, we will show that once the

feasibility and global optimality of solutions to all the non-convex optimization sub-problems can be guaranteed, the iterative fractional programming is applicable in terms of global convergence.

To tackle \mathcal{M}_1 , we introduce an additional auxiliary variable $\theta \in \mathbb{R}$ that is used to iteratively approach the optimal function value of \mathcal{M}_1 , i.e., the optimal objective value denoted by θ^* . Specifically, based on Dinkelbach's transformation, we can solve \mathcal{M}_1 by equivalently solving a sequence of the following optimization sub-problems, denoted by $\mathcal{M}_2(\theta)$, i.e.,

$$\mathcal{M}_2(\theta) : \min_{\mathbf{x}} : \sum_{i=1}^N (C_i x_i + D_i) - \theta \left[\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i x_i}{\tilde{I}_i(\mathbf{x})} \right) \right]$$

$$s.t. \begin{cases} (\alpha + 1)c_{0,i} + ((\alpha + 1)\mathbf{b}_i - \mathbf{e}_i)^T \mathbf{A}\mathbf{x} \leq 0, \\ \forall i \in \mathcal{N}; \\ \mathbf{x} \in \mathcal{X}. \end{cases} \quad (20)$$

in which the auxiliary variable θ is treated as a fixed parameter. Then, θ can be iteratively updated by

$$\theta[k+1] = \frac{\sum_{i=1}^N (C_i x_i^*[k] + D_i)}{\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i x_i^*[k]}{\tilde{I}_i(\mathbf{x}^*[k])} \right)}, \quad (21)$$

where k is used to denote the iteration index of updating θ , $k \in \mathbb{Z}_+$, and $x_i^*[k]$ denotes the optimal decision of the node $i \in \mathcal{N}$ determined at the k -th iteration, which is obtained by solving the sub-optimization problem $\mathcal{M}_2(\theta[k])$. $\mathbf{x}^*[k]$ is the column vectors of all the optimal decision variables at the k -th iteration, i.e., $\mathbf{x}^*[k] = [x_1^*[k], x_2^*[k], \dots, x_N^*[k]]^T$.

Define the parameterized optimal objective of $\mathcal{M}_2(\theta)$ as

$$h(\theta) = \min_{\mathbf{x} \in \mathcal{Q}} \left\{ \sum_{i=1}^N (C_i x_i + D_i) - \theta \left[\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i x_i}{\tilde{I}_i(\mathbf{x})} \right) \right] \right\}. \quad (22)$$

According to Jagannathan's theorem, we can introduce the following conclusion:

Theorem 1 (Jagannathan's theorem [58]). Let $\mathbf{x}^* \in \mathcal{Q}$ and $\theta^* = \theta(\mathbf{x}^*)$. \mathbf{x}^* is an optimal solution for the original problem \mathcal{M}_1 if and only if \mathbf{x}^* is optimal for the problem $\mathcal{M}_2(\theta^*)$.

Define $\theta[k+1]$ and $\theta[k+2]$ by $\theta[k+1] = \theta(\mathbf{x}[k])$ and $\theta[k+2] = \theta(\mathbf{x}[k+1])$, respectively. Using Theorem 1 above and resorting to Dinkelbach's theorem [59], we further have the corollaries as follows.

Corollary 1. At the optimal function value of the original problem \mathcal{M}_1 , θ^* , $h(\theta^*) = 0$ always holds true.

Corollary 2. Given that $\mathbf{x}[k] \in \mathcal{Q}$ and that $\mathbf{x}[k+1]$ solves $\mathcal{M}_2(\theta[k+1])$. If $\mathbf{x}[k]$ also solves $\mathcal{M}_2(\theta[k+1])$, $\mathbf{x}[k]$ is an feasible optimal solution for \mathcal{M}_1 , and $h(\theta[k+1]) = 0$ holds true. Otherwise, it always holds true that $\theta[k+2] < \theta[k+1]$.

Proof: Corollary 1 immediately follows Theorem 1 and can be proven in the similar logic in [59]. In the following, we mainly prove Corollary 2. According to Theorem 1 and Corollary 1, $\mathbf{x}[k]$ is an optimal solution for \mathcal{M}_1 and $h(\theta[k+1]) = 0$ under the condition that $\mathbf{x}[k]$ can solve the problem $\mathcal{M}_2(\theta[k+1])$. When $\mathbf{x}[k]$ does not satisfy the

solution optimality of $\mathcal{M}_2(\theta[k+1])$, there must exist some points $\tilde{\mathbf{x}} \in \mathcal{Q}$ such that the objective function value of $\mathcal{M}_2(\theta[k+1])$ at $\tilde{\mathbf{x}}$ satisfies

$$\sum_{i=1}^N (C_i \tilde{x}_i + D_i) - \theta[k+1] \left[\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i \tilde{x}_i}{\tilde{I}_i(\tilde{\mathbf{x}})} \right) \right] < \\ \sum_{i=1}^N (C_i x_i[k] + D_i) - \theta[k+1] \left[\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i x_i[k]}{\tilde{I}_i(\mathbf{x}[k])} \right) \right] = 0. \quad (23)$$

Since $\mathbf{x}[k+1]$ solves $\mathcal{M}_2(\theta[k+1])$, we can see

$$h(\theta[k+1]) = \sum_{i=1}^N (C_i x_i[k+1] + D_i) \\ - \theta[k+1] \left[\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i x_i[k+1]}{\tilde{I}_i(\mathbf{x}[k+1])} \right) \right] \\ \leq \sum_{i=1}^N (C_i \tilde{x}_i + D_i) - \theta[k+1] \left[\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i \tilde{x}_i}{\tilde{I}_i(\tilde{\mathbf{x}})} \right) \right] < 0, \quad (24)$$

which leads to

$$\theta(\mathbf{x}[k+1]) = \frac{\sum_{i=1}^N (C_i x_i[k+1] + D_i)}{\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i x_i[k+1]}{\tilde{I}_i(\mathbf{x}[k+1])} \right)} < \theta[k+1]. \quad (25)$$

This indeed is $\theta[k+2] < \theta[k+1]$. At this point, the corollary is proven. \square

Moreover, let $F_{\mathcal{M}_2}(\mathbf{x}, \mathbf{y})$ denote the objective function of $\mathcal{M}_2(\theta(\mathbf{x}))$, i.e.,

$$F_{\mathcal{M}_2}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^N (C_i y_i + D_i) - \theta(\mathbf{x}) \left[\sum_{i=1}^N \omega_i \log_2 \left(\frac{a_i y_i}{\tilde{I}_i(\mathbf{y})} \right) \right] \quad (26)$$

where $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ and $\mathbf{y} \in \mathcal{Q}$. Obviously, $F_{\mathcal{M}_2}(\mathbf{x}, \mathbf{y}) : \mathcal{Q} \times \mathcal{Q} \mapsto \mathbb{R}$, and this mapping is continuous on $\mathcal{Q} \times \mathcal{Q}$. Notice that $\mathcal{Q} \times \mathcal{Q}$ is a closed set, such that the value domain of $F_{\mathcal{M}_2}(\mathbf{x}, \mathbf{y})$ is also closed. Based on the continuity and boundedness of the function $F_{\mathcal{M}_2}(\mathbf{x}, \mathbf{y})$ with respect to \mathbf{x} and \mathbf{y} , we derive the following result:

Theorem 2. Let $\{\mathbf{x}[k]\}$ be a sequence of feasible points generated by solving a sequence of corresponding optimization sub-problems $\{\mathcal{M}_2(\theta[k]), \forall k \in \mathbb{Z}_+\}$, and $\lim_{k \rightarrow +\infty} \mathbf{x}[k] = \mathbf{p}^* \in \mathcal{Q}$. Such \mathbf{x}^* must be an optimal solution for the original problem \mathcal{M}_1 .

Proof: According to Corollary 2, it can be seen that

$$h(\theta(\mathbf{x}[k])) = F_{\mathcal{M}_2}(\mathbf{x}[k], \mathbf{x}[k+1]) \leq F_{\mathcal{M}_2}(\mathbf{x}[k], \mathbf{y}) \leq 0 \quad (27)$$

for all $\mathbf{x}[k] \in \mathcal{Q}$ and all $\mathbf{y} \in \mathcal{Q}$. This result indicates that $F_{\mathcal{M}_2}(\mathbf{x}, \mathbf{y})$ is bounded on $\mathcal{Q} \times \mathcal{Q}$. The equality of (27) is attained when $\mathbf{x}[k+1] = \mathbf{x}[k]$, i.e., $\mathbf{x}[k]$ solves $\mathcal{M}_2(\theta[k+1])$.

Furthermore, since $\theta(\mathbf{x}[k+1]) = \theta[k+2] < \theta[k+1] = \theta(\mathbf{x}[k])$ and $F(\mathbf{x}) > 0$ for $\mathbf{x} \in \mathcal{Q}$, we have $h(\theta(\mathbf{x}[k])) \leq h(\theta(\mathbf{x}[k+1]))$, i.e.,

$$F_{\mathcal{M}_2}(\mathbf{x}[k], \mathbf{x}[k+1]) \leq F_{\mathcal{M}_2}(\mathbf{x}[k+1], \mathbf{x}[k+2]) \leq 0, \quad (28)$$

which implies that $\{F_{\mathcal{M}_2}(\mathbf{x}[k], \mathbf{x}[k+1])\}$ is a monotonically decreasing sequence with respect to k . Therefore, the limit of $F_{\mathcal{M}_2}(\mathbf{x}[k], \mathbf{x}[k+1])$ must exist. Specifically, taking the limit on both sides of (28) can get $\lim_{k \rightarrow +\infty} F_{\mathcal{M}_2}(\mathbf{x}[k], \mathbf{x}[k+1]) =$

Algorithm 1: Iterative Fractional Programming

Input: The tolerable maximum number of iterations, $K > 0$, and the tolerable error, $\epsilon \geq 0$.
Output: An optimal solution and an optimal objective function value of \mathcal{M}_0

```

/* Initialization */
1 Select  $\mathbf{x}[0] \in \mathcal{Q}$  and  $\theta[1] = \theta(\mathbf{x}[0])$ . Set  $k = 0$ .
/* Now this is a While loop for iterations */
2 while  $k \leq K$  do
    /* Solve sub-problem at an iteration */
    3 Solve  $\mathcal{M}_2(\theta[k+1])$  to get an optimal feasible solution  $\mathbf{x}[k+1]$ .
    /* Update auxiliary parameter */
    4 Calculate  $\theta[k+2] = \theta(\mathbf{x}[k+1])$ .
    5 if  $\|h(\theta[k+1])\| \leq \epsilon$  then
    6     Break and return  $\mathbf{x}[k+1]$  and  $\theta[k+2]$ .
    7 else
    8     Set  $k = k + 1$ .
9 Return  $\mathbf{p}[k] = F_{\text{affine}}(\mathbf{x}[k])$  and  $\theta[k+1]$ .
```

$F_{\mathcal{M}_2}(\mathbf{x}^*, \mathbf{x}^*) = 0$. Hence, according to Theorem 1, \mathbf{x}^* is an optimal solution to \mathcal{M}_1 . \square

Now, based on the above theorems, we can propose an iterative fractional programming framework based on solving a series of $\mathcal{M}_2(\theta[k])$ as summarized in Algorithm 1.

Remark: Different from the original Dinkelbach's theoretical results, Theorem 2 establishes the global convergence of the iterative fractional programming, which does not require the fractional programming problem to be convex. However, Theorem 2 indicates that the convergence to an optimal solution to the original problem \mathcal{M}_1 relies on the feasibility and solution optimality of solving a sequence of non-convex sub-problems $\mathcal{M}_2(\theta[k])$ generated at all iterations k . In the following section, we dive into the subproblem and propose a distributed optimization method with a convexification transformation to deal with the challenge.

5 CONVEXIFICATION AND DISTRIBUTED OPTIMIZATION FOR SUCCESSIVE SUB-PROBLEMS

It has been shown in Section 4 that the applicability of the iterative fractional programming scheme relies on solving a globally optimal point of the constrained optimization sub-problems at all iterations. Thus, how to address $\mathcal{M}_2(\theta)$ plays a key role in the iterative fractional programming. Unfortunately, the objective function of $\mathcal{M}_2(\theta)$ is neither jointly convex nor jointly concave with respect to the decisions \mathbf{x} . It is difficult or even impossible to obtain a globally optimal solution to $\mathcal{M}_2(\theta)$ by directly applying existing convex optimization techniques. Hence, we first propose an algebraic transformation to realize the convexification of $\mathcal{M}_2(\theta)$ and pave the way for the development of a global distributed optimization algorithm.

5.1 Exponential Transformation-based Convexification

We introduce an exponential transformation denoted by $\phi(x) = \exp(x)$. Obviously, $\phi(x)$ is strictly monotonically increasing and strictly convex with respect to x , and $\phi(x) \in$

\mathbb{R}_{++} . There must exist a real number $s_i \in \mathbb{R}$ such that $\phi(s_i) = a_i x_i$, and such an s_i is unique for an $x_i, \forall i \in \mathcal{N}$. Let a column vector \mathbf{s} be $\mathbf{s} = [s_1, s_2, \dots, s_N]^T$, and define $\phi(\mathbf{s}) = [\phi(s_1), \phi(s_2), \dots, \phi(s_N)]^T$. It can be seen that $\phi(\mathbf{s}) = \mathbf{A}\mathbf{x}$. Since each x_i is bounded for all $i \in \mathcal{N}$, i.e., $x_i \in [x_{i,\min}, x_{i,\max}]$ where $x_{i,\min}$ and $x_{i,\max}$ are given in (15), we also have $s_i \in \mathcal{S}_i \triangleq [\ln(a_i x_{i,\min}), \ln(a_i x_{i,\max})]$ for all $i \in \mathcal{N}$. At this point, the feasible region of the bound constraints on \mathbf{s} is defined by $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \dots \times \mathcal{S}_N$. Recalling (4), (12) and (13), we rewrite the SINR-based constraints after the exponential transformation as

$$\varphi_i(\phi(\mathbf{s})) = \log_2(\alpha + 1) + \log_2 \left(\frac{\tilde{I}_i(\phi(\mathbf{s}))}{\phi(s_i)} \right) \leq 0, \quad (29)$$

where $\tilde{I}_i(\phi(\mathbf{s})) = c_{0,i} + \mathbf{b}_i^T \phi(\mathbf{s}), \forall i \in \mathcal{N}$.

Now, treating \mathbf{s} as new decision variables, solving the non-convex model $\mathcal{M}_2(\theta)$ boils down to solving the following model with respect to \mathbf{s} .

$$\begin{aligned} \mathcal{M}_3(\theta) : \min_{\mathbf{s}} : J(\mathbf{s}) &= \sum_{i=1}^N \frac{C_i \phi(s_i)}{a_i} \\ &+ \theta \left[\sum_{i=1}^N \omega_i \log_2 \left(\frac{\tilde{I}_i(\phi(\mathbf{s}))}{\phi(s_i)} \right) \right] \quad (30) \\ \text{s.t.} : \varphi_i(\phi(\mathbf{s})) &\leq 0, \forall i \in \mathcal{N}; \\ \mathbf{s} &\in \mathcal{S}. \end{aligned}$$

Let $\mathcal{R} \triangleq \{\mathbf{s} : \varphi_i(\phi(\mathbf{s})) \leq 0, \forall i \in \mathcal{N}\}$ for notation simplicity. From the above model $\mathcal{M}_3(\theta)$, we can derive the following result, the proof of which is detailed in Appendix 1 of the online supplemental material.

Theorem 3. Suppose $\mathcal{S} \cap \mathcal{R} \neq \emptyset$. $\mathcal{M}_3(\theta)$ is a convex constrained optimization problem that has a unique globally optimal solution in $\mathcal{S} \cap \mathcal{R}$.

Remark: Theorem 3 shows that the non-convex subproblem $\mathcal{M}_2(\theta)$ can be efficiently converted into a convex model by using the proposed affine transformation and exponential transformation. This brings into play the convex optimization methodology for energy-efficiency design.

5.2 Distributed Constrained Optimization

The constrained optimization model $\mathcal{M}_3(\theta)$ is derived based on the exponential transformation-based convexification of the model $\mathcal{M}_2(\theta)$. Ideally, $\mathcal{M}_3(\theta)$ can be efficiently solved by using some well-known convex optimization methods like interior point methods from the centralized computation point of view. However, considering the practical deployment of a network system, such centralized computation-based methods may not be effective if each node cannot fully access the global information on the whole system. Therefore, we focus on the development of a distributed optimization algorithm to address $\mathcal{M}_3(\theta)$ despite the presence of its variable-coupled objective and constraints. The key idea is to solve a corresponding fixed-point problem by combining the duality theory and the variational inequality theory.

First, we introduce a sequence of non-negative Lagrangian multipliers (dual decision variables) $\lambda_i \in \mathbb{R}_+, \forall i \in \mathcal{N}$, and let the set of these multipliers be a column

vector $\boldsymbol{\lambda}$, i.e., $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_N]^T \in \mathbb{R}_+^N$. For simplicity, we also re-write $\varphi_i(\mathbf{s}) = \varphi_i(\phi(\mathbf{s}))$ for all $i \in \mathcal{N}$, and introduce $\boldsymbol{\varphi}(\mathbf{s}) = [\varphi_1(\mathbf{s}), \varphi_2(\mathbf{s}), \dots, \varphi_N(\mathbf{s})]^T$ such that the nonlinear coupling constraints can be represented as a compact form $\boldsymbol{\varphi}(\mathbf{s}) \leq \mathbf{0}$. Using these notations, we formulate the Lagrangian function of $\mathcal{M}_3(\theta)$ as follows

$$\begin{aligned} \mathcal{L}(\mathbf{s}, \boldsymbol{\lambda}) &\triangleq J(\mathbf{s}) + \boldsymbol{\lambda}^T \boldsymbol{\varphi}(\mathbf{s}) \\ &= \sum_{i=1}^N \frac{C_i \phi(s_i)}{a_i} + \sum_{i=1}^N (\theta \omega_i + \lambda_i) \log_2 \left(\frac{\tilde{I}_i(\phi(\mathbf{s}))}{\phi(s_i)} \right) \quad (31) \\ &+ \log_2(\alpha + 1) \sum_{i=1}^N \lambda_i, \end{aligned}$$

and the dual problem of $\mathcal{M}_3(\theta)$ is proposed as follows

$$\begin{aligned} \mathcal{D}_3(\theta) : \max_{\boldsymbol{\lambda}} : V(\boldsymbol{\lambda}) &= \min_{\mathbf{s} \in \mathcal{S}} \mathcal{L}(\mathbf{s}, \boldsymbol{\lambda}) \quad (32) \\ \text{s.t.} : \boldsymbol{\lambda} &\in \mathbb{R}_+^N. \end{aligned}$$

It is noticed that the Lagrangian function $\mathcal{L}(\mathbf{s}, \boldsymbol{\lambda})$ is always linear with respect to the non-negative dual decision variables $\boldsymbol{\lambda}$. Thus, the dual problem $\mathcal{D}_3(\theta)$ is always convex.

To proceed with analysis, we introduce a weak assumption, i.e., the Slater's condition on $\mathcal{M}_3(\theta)$, that there exists a feasible point denoted by $\bar{\mathbf{s}} \in \mathcal{S}$ such that for all $i \in \mathcal{N}$ $\varphi_i(\bar{\mathbf{s}}) < 0$. $\bar{\mathbf{s}}$ is also termed an interior point of the feasible region $\mathcal{I} \triangleq \mathcal{S} \cap \mathcal{R}$, and \mathcal{I} satisfies the Slater's constraint qualification. Here, we need to point out that the introduced assumption does not impair the generality of our methodological framework since the model can always be reduced to the case meeting the Slater's constraint qualification. Specifically, if there exist some nodes, denoted by a set $\mathcal{N}_{\text{active}}$, such that their coupling interference constraints are active, i.e., $\varphi_j(\mathbf{s}) = 0$ for $j \in \mathcal{N}_{\text{active}}$, and the remaining constraints are inactive, i.e. $\varphi_i(\mathbf{s}) < 0$ for $i \in \mathcal{N} \setminus \mathcal{N}_{\text{active}} \neq \emptyset$, we can remove the decision variables of those nodes from \mathbf{s} and their active constraints from \mathcal{I} , which does not change the optimal solution of $\mathcal{M}_3(\theta)$. This is indeed the basic idea following the active-set method. That is, for each $j \in \mathcal{N}_{\text{active}}$, according to $\varphi_j(\mathbf{s}) = 0$ for $j \in \mathcal{N}_{\text{active}}$, we can see that s_j can be expressed as a function of the combination of the other decision variables $\{s_i : i \in \mathcal{N} \setminus \mathcal{N}_{\text{active}}\}$ as follows

$$s_j = \ln \left(\frac{\alpha \left(c_{0,j} + \sum_{i=1, i \neq j}^N b_{i,j} \phi(s_i) \right)}{1 - (\alpha + 1) b_{j,j}} \right), \quad (33)$$

where $b_{i,j}$ is the i -th element of the column vector \mathbf{b}_j . (33) implies that we only need to determine the partial decision variables associated with $\mathcal{N} \setminus \mathcal{N}_{\text{active}}$ and then can obtain the others associated with $\mathcal{N}_{\text{active}}$ by combining the partial decision variables whose interference constraints are inactive. In this way, we reduce the previous decision variable \mathbf{s} to a new one $\tilde{\mathbf{s}} = \text{col}\{s_i, i \in \mathcal{N} \setminus \mathcal{N}_{\text{active}}\}$. The previous model $\mathcal{M}_3(\theta)$ boils down to a lower-dimensional optimization problem with respect to $\tilde{\mathbf{s}}$, which meets the Slater's condition.

Based on the above Slater's condition, we have the following strong duality property for $\mathcal{M}_3(\theta)$ as follows.

Corollary 3 (Strong duality). Suppose that $\mathcal{M}_3(\theta)$ meets the Slater's condition. A dual optimal solution, denoted by $\boldsymbol{\lambda}^* \in \mathbb{R}_+^N$, exists for the dual problem $\mathcal{D}_3(\theta)$, and the

optimal objective function value of $\mathcal{D}_3(\theta)$ at $\boldsymbol{\lambda}^*$ is the same as that of the primal problem $\mathcal{M}_3(\theta)$ at its optimal solution $\mathbf{s}^* \in \mathcal{I}$.

Proof: Since $\mathcal{M}_3(\theta)$ is convex as stated in Theorem 3, its Slater's condition is a sufficient condition to make the strong duality hold. \square

Let the optimal objective function value of the dual problem $\mathcal{D}_3(\theta)$ be d^* , and thus we can see $J(\mathbf{s}^*) = V(\boldsymbol{\lambda}^*) = d^*$. Besides, since the optimal solutions of the primal and dual problems, $(\mathbf{s}^*, \boldsymbol{\lambda}^*)$, is always a saddle point of the Lagrangian function $\mathcal{L}(\mathbf{s}, \boldsymbol{\lambda})$, the following inequalities always hold true for $(\mathbf{s}, \boldsymbol{\lambda}) \in \mathcal{S} \times \mathbb{R}_+^N$,

$$\mathcal{L}(\mathbf{s}, \boldsymbol{\lambda}^*) \geq \mathcal{L}(\mathbf{s}^*, \boldsymbol{\lambda}^*) \geq \mathcal{L}(\mathbf{s}^*, \boldsymbol{\lambda}). \quad (34)$$

It is also recognized that for some given $\bar{\boldsymbol{\lambda}} \in \mathbb{R}_+^N$, $\nabla_{\mathbf{s}} \mathcal{L}(\mathbf{s}, \bar{\boldsymbol{\lambda}}) = \mathbf{0}$ can be treated as a group of linear equations with respect to $\phi(s_i)$, $\forall i \in \mathcal{N}$, where $\bar{\boldsymbol{\lambda}}$ are treated as fixed parameters. Hence, by solving these linear equations, we can obtain a zero-point solution denoted by $\tilde{\phi}_{\bar{\boldsymbol{\lambda}}} = [\tilde{\phi}_{\bar{\boldsymbol{\lambda}},1}, \tilde{\phi}_{\bar{\boldsymbol{\lambda}},2}, \dots, \tilde{\phi}_{\bar{\boldsymbol{\lambda}},N}]^T$ such that $\nabla_{\mathbf{s}} \mathcal{L}(\ln(\tilde{\phi}_{\bar{\boldsymbol{\lambda}}}), \bar{\boldsymbol{\lambda}}) = \mathbf{0}$, where $\ln(\tilde{\phi}_{\bar{\boldsymbol{\lambda}}})$ is a component-wise natural logarithmic function, i.e., $\ln(\tilde{\phi}_{\bar{\boldsymbol{\lambda}}}) = [\ln(\tilde{\phi}_{\bar{\boldsymbol{\lambda}},1}), \ln(\tilde{\phi}_{\bar{\boldsymbol{\lambda}},2}), \dots, \ln(\tilde{\phi}_{\bar{\boldsymbol{\lambda}},N})]^T$. We further denote $\mathbf{u} = \ln(\tilde{\phi}_{\bar{\boldsymbol{\lambda}}}) \in \mathbf{R}^N$, which is indeed the globally optimal point for the following unconstrained convex optimization problem

$$\mathbf{u} \in \operatorname{argmin}_{\mathbf{s} \in \mathbb{R}^N} \{J_{\bar{\boldsymbol{\lambda}}}(\mathbf{s}) = \mathcal{L}(\mathbf{s}, \bar{\boldsymbol{\lambda}})\}. \quad (35)$$

According to the saddle-point condition (34), we have

$$\mathcal{L}(\bar{\mathbf{s}}, \boldsymbol{\lambda}^*) = J(\bar{\mathbf{s}}) + \sum_{i=1}^N \lambda_i \varphi_i(\bar{\mathbf{s}}) \geq \mathcal{L}(\mathbf{s}^*, \bar{\boldsymbol{\lambda}}) \geq J_{\bar{\boldsymbol{\lambda}}}(\mathbf{u}), \quad (36)$$

which results in

$$\begin{aligned} J(\bar{\mathbf{s}}) - J_{\bar{\boldsymbol{\lambda}}}(\mathbf{u}) &\geq \sum_{i=1}^N \lambda_i (-\varphi_i(\bar{\mathbf{s}})) \geq \left(\sum_{i=1}^N \lambda_i \right) \min_{i \in \mathcal{N}} \{-\varphi_i(\bar{\mathbf{s}})\} \\ \Rightarrow \frac{J(\bar{\mathbf{s}}) - J_{\bar{\boldsymbol{\lambda}}}(\mathbf{u})}{\min_{i \in \mathcal{N}} \{-\varphi_i(\bar{\mathbf{s}})\}} &\geq \sum_{i=1}^N \lambda_i \end{aligned} \quad (37)$$

and

$$\begin{aligned} J(\bar{\mathbf{s}}) - J_{\bar{\boldsymbol{\lambda}}}(\mathbf{u}) &\geq \sum_{i=1}^N \lambda_i (-\varphi_i(\bar{\mathbf{s}})) \geq -\lambda_i \varphi_i(\bar{\mathbf{s}}) \\ \Rightarrow \frac{J(\bar{\mathbf{s}}) - J_{\bar{\boldsymbol{\lambda}}}(\mathbf{u})}{-\varphi_i(\bar{\mathbf{s}})} &\geq \lambda_i. \end{aligned} \quad (38)$$

Therefore, we derive a compact convex set \mathcal{U} for $\boldsymbol{\lambda}$, which contains the optimal dual solution $\boldsymbol{\lambda}^*$

$$\mathcal{U} \triangleq \left\{ \boldsymbol{\lambda} : 0 \leq \lambda_i \leq \frac{J(\bar{\mathbf{s}}) - J_{\bar{\boldsymbol{\lambda}}}(\mathbf{u})}{-\varphi_i(\bar{\mathbf{s}})}, \forall i \in \mathcal{N} \right\}. \quad (39)$$

Define the decision variable pair of \mathbf{s} and $\boldsymbol{\lambda}$ as $\mathbf{z} = \operatorname{col}\{\mathbf{s}, \boldsymbol{\lambda}\}$. The pair of the optimal primal solution and the optimal dual solution is represented by $\mathbf{z}^* = \operatorname{col}\{\mathbf{s}^*, \boldsymbol{\lambda}^*\}$. The primal-dual gradient-related function $G(\mathbf{s}, \boldsymbol{\lambda})$ is

$$G(\mathbf{s}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{s}} \mathcal{L}(\mathbf{s}, \boldsymbol{\lambda}) \\ -\nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{s}, \boldsymbol{\lambda}) \end{bmatrix} = \begin{bmatrix} \nabla_{\mathbf{s}} J(\mathbf{s}) + \sum_{i=1}^N \lambda_i \nabla_{\mathbf{s}} \varphi_i(\mathbf{s}) \\ -\boldsymbol{\varphi}(\mathbf{s}) \end{bmatrix}. \quad (40)$$

With (40), the equivalent formulation of the primal and the dual problems $\mathcal{M}_3(\theta)$ and $\mathcal{D}_3(\theta)$ can be represented in a variational formulation as follows:

$$(\mathbf{z} - \mathbf{z}^*)^T G(\mathbf{s}^*, \boldsymbol{\lambda}^*) \geq 0, \quad \mathbf{z} \in \mathcal{S} \times \mathcal{U}. \quad (41)$$

According to the theory of variational inequality (see [60]), the pair of $(\mathbf{s}^*, \boldsymbol{\lambda}^*)$ is an optimal solution for the primal and the dual problems $\mathcal{M}_3(\theta)$ and $\mathcal{D}_3(\theta)$ if and only if it is a solution for the variational inequality (41). More specifically, we can have the following corollary:

Corollary 4. If \mathbf{s}^* and $\boldsymbol{\lambda}^*$ solve the following fixed point problem $\mathcal{F}_3(\theta)$, then \mathbf{z}^* solves the variational inequality (41) and \mathbf{s}^* and $\boldsymbol{\lambda}^*$ are optimal solutions for $\mathcal{M}_3(\theta)$ and $\mathcal{D}_3(\theta)$, respectively.

$$\mathcal{F}_3(\theta) : \begin{cases} \mathbf{s}^* = \Pi_{\mathcal{S}}(\mathbf{s}^* - \nabla_{\mathbf{s}} \mathcal{L}(\mathbf{s}^*, \boldsymbol{\lambda}^*)); \\ \boldsymbol{\lambda}^* = \Pi_{\mathcal{U}}(\boldsymbol{\lambda}^* + \nabla_{\boldsymbol{\lambda}} \mathcal{L}(\mathbf{s}^*, \boldsymbol{\lambda}^*)). \end{cases} \quad (42)$$

Proof: This corollary can be proven by using the same logic in Proposition 1.5.8 of Volume I of [60]. \square

It is observed that besides \mathcal{S} , the constraint set of $\boldsymbol{\lambda}$, \mathcal{U} , is also a convex box-constraint set. Thus, we can decompose \mathcal{U} as $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \dots \times \mathcal{U}_N$ where $\mathcal{U}_i = \{\lambda : 0 \leq \lambda \leq (J(\bar{\mathbf{s}}) - J_{\bar{\boldsymbol{\lambda}}}(\mathbf{u})) / (-\varphi_i(\bar{\mathbf{s}}))\}$ for all $i \in \mathcal{N}$. This indicates that the fixed-point equations (42) can be further decomposed for motivating a distributed computation, in which the Euclidean projections are decoupled. To be specific, we resort to the extra-gradient iterative method to establish an iterative algorithm for solving the fixed-point equations (42) in $\mathcal{F}_3(\theta)$ in a distributed fashion as follows

$$\begin{cases} s_i [t + \frac{1}{2}] = \Pi_{\mathcal{S}_i}(s_i[t] - \gamma \nabla_{s_i} \mathcal{L}(s_i[t], \boldsymbol{\lambda}_i[t])) \\ \lambda_i [t + \frac{1}{2}] = \Pi_{\mathcal{U}_i}(\lambda_i[t] + \delta \varphi_i(s_i[t])) \\ s_i[t + 1] = \Pi_{\mathcal{S}_i}(s_i[t] - \gamma \nabla_{s_i} \mathcal{L}(s_i[t + \frac{1}{2}], \boldsymbol{\lambda}_i[t + \frac{1}{2}])) \\ \lambda_i[t + 1] = \Pi_{\mathcal{U}_i}(\lambda_i[t + \frac{1}{2}] + \delta \varphi_i(s_i[t + \frac{1}{2}])) \end{cases} \quad (43)$$

for all $i \in \mathcal{N}$, where $\gamma > 0$ and $\delta > 0$ are the primal and the dual step-lengths, $t \in \mathbb{Z}_+$ denotes the iteration index. The primal and the dual Euclidean projections can be reduced to

$$\Pi_{\mathcal{S}_i}(x) = \begin{cases} s_{i,\min}, & x < s_{i,\min} \\ x, & x \in \mathcal{S}_i \\ s_{i,\max}, & x > s_{i,\max} \end{cases}, \quad \forall i \in \mathcal{N} \quad (44)$$

and

$$\Pi_{\mathcal{U}_i}(x) = \begin{cases} 0, & x < 0 \\ x, & x \in \mathcal{U}_i \\ \frac{J(\bar{\mathbf{s}}) - J_{\bar{\boldsymbol{\lambda}}}(\mathbf{u})}{-\varphi_i(\bar{\mathbf{s}})}, & x > \frac{J(\bar{\mathbf{s}}) - J_{\bar{\boldsymbol{\lambda}}}(\mathbf{u})}{-\varphi_i(\bar{\mathbf{s}})} \end{cases}, \quad \forall i \in \mathcal{N}. \quad (45)$$

The distributed optimization algorithm is given in Algorithm 2. It is remarked that since the extra-gradient method is used, the extra projection operations are carried out at each iteration. Even though the extra-gradient iterations require twice the amount of computations when compared to a conventional gradient method, it can facilitate solving the pseudo monotone variational inequalities as in (41).

Algorithm 2: Extra-gradient-based primal-dual distributed algorithm for solving $\mathcal{M}_2(\theta)$

Input: The tolerable maximum number of iterations, $T_i > 0$, $i \in \mathcal{N}$, and the tolerable error, $\epsilon \geq 0$. The global energy-efficiency metric determined at the outer iteration $k + 1$, $\theta[k + 1]$, and the previous iteration $\mathbf{x}[k]$.

Output: An optimal solution of $\mathcal{M}_2(\theta[k + 1])$.

```

/* Initialization */
1 Logarithmically transform  $\mathbf{s}[t] = \ln(\mathbf{Ax}[k]) \in \mathcal{S}$ .
2 Select  $\lambda[t] \in \mathcal{U}$ .
3 Set  $\theta = \theta[k + 1]$  and  $t = 0$ .
/* While loop for iterations by each node */
4 while  $t \leq T_i$  for each  $i \in \mathcal{N}$  do
    /* Primal-dual extra-gradient updates */
5 Perform primal-dual extra-gradient updates with
    (43) to get new decisions  $s_i[t + 1]$  and  $\lambda_i[t + 1]$ .
    /* Distributed coordination */
6 Exchange primal-dual decisions with wireless
    communications over control channels.
7 if  $\max\{\|\mathbf{s}[t + 1] - \mathbf{s}[t]\|, \|\boldsymbol{\lambda}[t + 1] - \boldsymbol{\lambda}[t]\|\} \leq \epsilon$ 
    then
8 Exponentially transform
     $\mathbf{Ax}[k + 1] = \phi(\mathbf{s}[t + 1])$ .
9 Break and return  $\mathbf{x}[k + 1]$ .
10 else
11 Set  $t = t + 1$ .
12 Exponentially transform  $\mathbf{Ax}[k + 1] = \phi(\mathbf{s}[t])$ .
13 Return  $\mathbf{x}[k + 1]$ .

```

5.3 Convergence Analysis

It is noticed that the convergence of the iterative algorithm following (43) depends on the selection of both the non-negative primal and the non-negative dual step-lengths, γ and δ . At this point, γ and δ should be properly specified and sufficiently small such that the algorithmic convergence can be guaranteed. Therefore, we would like to establish the convergence condition in this subsection.

For each node i , let $F_i(\mathbf{s})$ and $h_i(\mathbf{s})$ denote

$$\begin{aligned}
 F_i(\mathbf{s}) &= \frac{C_i \phi(s_i)}{a_i} + \theta \left[\sum_{l=1}^N \omega_l \log_2 \left(\frac{\tilde{I}_l(\phi(\mathbf{s}))}{\phi(s_l)} \right) \right] \\
 h_i(\mathbf{s}) &= \nabla_{s_i} F_i(\mathbf{s}) \\
 &= \frac{C_i \phi(s_i)}{a_i} + \sum_{l=1, l \neq i}^N \frac{\theta \omega_l b_{i,l} \phi(s_i)}{(\ln 2) \tilde{I}_l(\phi(\mathbf{s}))} - \frac{\theta \omega_i}{\ln 2}
 \end{aligned} \tag{46}$$

for all $i \in \mathcal{N}$. It is seen that each $h_i(\mathbf{s})$ is a continuously differentiable function with respect to \mathbf{s} over \mathcal{S} . Besides, we also let $H(\mathbf{s}) = \nabla_{\mathbf{s}} J(\mathbf{s}) = [h_1(\mathbf{s}), h_2(\mathbf{s}), \dots, h_N(\mathbf{s})]^T$, $\boldsymbol{\varphi}(\mathbf{s}) = [\varphi_1(\mathbf{s}), \varphi_2(\mathbf{s}), \dots, \varphi_N(\mathbf{s})]^T$, and $\nabla_{\mathbf{s}} \boldsymbol{\varphi}(\mathbf{s}) = \text{col}\{\nabla_{s_i} \varphi_i(\mathbf{s}), i \in \mathcal{N}\}$. To proceed, we first show the Lipschitz continuity of $H(\mathbf{s})$ and $\nabla_{s_i} \varphi_i(\mathbf{s})$, $\forall i \in \mathcal{N}$, and derive their Lipschitz constants. Based on this, we further deduce the monotonicity and Lipschitz continuity of the mapping function $G(\mathbf{s}, \boldsymbol{\lambda})$ given in (40).

Lemma 1. Let L be $L = \sqrt{\sum_{j=1}^N \sum_{i=1}^N L_{j,i}^2}$ and $L_{j,i}$ be

$$L_{j,i} = \begin{cases} \frac{C_i \mu_i}{a_i} + \sum_{l=1, l \neq i}^N \frac{\theta \omega_l b_{i,l} \mu_i}{(\ln 2) |c_{0,l}|}, & j = i; \\ \sum_{l=1, l \neq i}^N \frac{\theta \omega_l b_{i,l} b_{j,l} \mu_i \mu_j}{(\ln 2) c_{0,l}^2}, & j \neq i \end{cases} \tag{47}$$

for $i, j = 1, 2, \dots, N$, where $\mu_i = \phi(s_{i, \max})$ for $i = 1, 2, \dots, N$. L is a Lipschitz constant for $H(\mathbf{s})$, i.e.,

$$\|H(\mathbf{s}^{(1)}) - H(\mathbf{s}^{(2)})\| \leq L \|\mathbf{s}^{(1)} - \mathbf{s}^{(2)}\| \tag{48}$$

holds true for any two $\mathbf{s}^{(1)}, \mathbf{s}^{(2)} \in \mathcal{S}$.

Lemma 2. Let L_i be $L_i = \sqrt{\sum_{l_1=1}^N \sum_{l_2=1}^N \kappa_{l_1, l_2}(i)}$ and $\kappa_{l_1, l_2}(i)$ be

$$\kappa_{l_1, l_2}(i) = \begin{cases} \frac{b_{l_1, i} \mu_{l_1}}{(\ln 2) c_{0,i}} + \sum_{k=1, k \neq l_1}^N \frac{b_{l_1, i} b_{k, i} \mu_{l_1} \mu_k}{(\ln 2) c_{0,i}^2}, & l_1 = l_2 \\ \frac{b_{l_1, i} b_{l_2, i} \mu_{l_1} \mu_{l_2}}{(\ln 2) c_{0,i}^2}, & l_1 \neq l_2 \end{cases} \tag{49}$$

for $l_1, l_2 = 1, 2, \dots, N$. L_i is a Lipschitz constant for $\varphi_i(\mathbf{s})$, i.e.,

$$\|\nabla_{\mathbf{s}} \varphi_i(\mathbf{s}^{(1)}) - \nabla_{\mathbf{s}} \varphi_i(\mathbf{s}^{(2)})\| \leq L_i \|\mathbf{s}^{(1)} - \mathbf{s}^{(2)}\|, \forall i \in \mathcal{N}, \tag{50}$$

holds true for any two $\mathbf{s}^{(1)}, \mathbf{s}^{(2)} \in \mathcal{S}$.

Lemma 3. $G(\mathbf{s}, \boldsymbol{\lambda})$ is monotone on $\mathcal{S} \times \mathbb{R}_+^N$ and is Lipschitz continuous on $\mathcal{S} \times \mathcal{U}$. One Lipschitz constant L_G can be formulated as

$$L_G = \sqrt{(L + \mu_{\varphi} + \mu_{\lambda} L_{\varphi})^2 + \mu_{\varphi}^2} \tag{51}$$

where $\mu_{\varphi} = \max_{\mathbf{s} \in \mathcal{S}} \{\|\nabla_{\mathbf{s}} \boldsymbol{\varphi}(\mathbf{s})\|\}$, $\mu_{\lambda} = \max_{\boldsymbol{\lambda} \in \mathcal{U}} \{\|\boldsymbol{\lambda}\|\}$, and $L_{\varphi} = \sqrt{\sum_{i=1}^N L_i}$.

The proofs of Lemmas 1, 2 and 3 are detailed in Appendices 2, 3 and 4, respectively, which are available in the online supplementary material. Based on these lemmas, we conclude the numerical convergence and the convergence rate for our distributed primal-dual algorithm using the extra-gradient mechanism in Algorithm 2 as follows.

Theorem 4. Let the pair of $(\mathbf{s}^*, \boldsymbol{\lambda}^*) \in \mathcal{S} \times \mathcal{U}$ be a solution of the variational inequality (41). If $0 < \gamma, \delta < 1/L_G$, the sequence generated by using (43), $\{\mathbf{s}[t], \boldsymbol{\lambda}[t]\}$, can converge to $(\mathbf{s}^*, \boldsymbol{\lambda}^*)$ at least R-linearly.

Proof: Let the pair of $(\mathbf{s}^*, \boldsymbol{\lambda}^*) \in \mathcal{S} \times \mathcal{U}$ be a solution of the variational inequality (41), $\{\mathbf{s}[t], \boldsymbol{\lambda}[t]\}$ and $\{\mathbf{s}[t + \frac{1}{2}], \boldsymbol{\lambda}[t + \frac{1}{2}]\}$ be the sequences generated by using (43). Combining Lemmas 1, 2 and 3, and following the same logic in Lemma 12.1.10 in Volume II of [60], we can see for all $t \in \mathbb{Z}_+$ that

$$\begin{cases} \|\mathbf{s}[t + 1] - \mathbf{s}^*\|^2 \leq \|\mathbf{s}[t] - \mathbf{s}^*\|^2 \\ \quad - \eta(\gamma) \left\| \mathbf{s} \left[t + \frac{1}{2} \right] - \mathbf{s}[t] \right\|^2; \\ \|\boldsymbol{\lambda}[t + 1] - \boldsymbol{\lambda}^*\|^2 \leq \|\boldsymbol{\lambda}[t] - \boldsymbol{\lambda}^*\|^2 \\ \quad - \eta(\delta) \left\| \boldsymbol{\lambda} \left[t + \frac{1}{2} \right] - \boldsymbol{\lambda}[t] \right\|^2, \end{cases} \tag{52}$$

where $\eta(\gamma) = 1 - \gamma^2 L_G^2$ and $\eta(\delta) = 1 - \delta^2 L_G^2$.

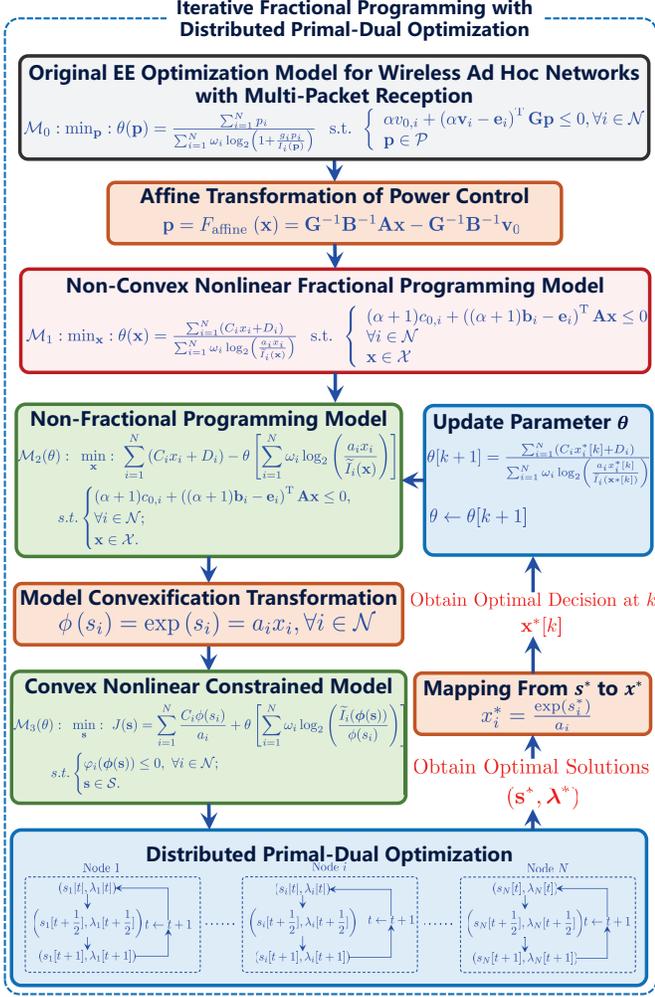


Fig. 2. The proposed methodological framework.

Since $0 < \gamma, \delta < 1/L_G$, we have $0 < \eta(\gamma), \eta(\delta) < 1$, and thus the generated sequence can guarantee the convergence. Also according to [60] (See Theorem 12.6.4 in Volume II), the convergence rate is at least R-linear. \square

By combining the distributed primal-dual optimization algorithm, we propose a novel iterative fractional programming framework as illustrated in Fig. 2. In this framework, we first transform a general non-convex energy-efficiency optimization problem into another one by using a proposed affine transformation scheme. Then, the fractional programming model is converted into a sequence of non-convex subproblems, which are further transformed into a sequence of convex subproblems by using the exponential transformation. The global optimum of the convex subproblems generated at each iteration, $(\mathbf{s}^*, \boldsymbol{\lambda}^*)$, are solved by using a proposed distributed primal-dual algorithm based on an extra-gradient projection mechanism. The optimal decision at an iteration k can be obtained by the mapping operator from $(\mathbf{s}^*, \boldsymbol{\lambda}^*)$ to $\mathbf{x}^*[k]$ and the optimal power control is also yielded by the affine transformation, i.e., $\mathbf{p}^*[k] = F_{\text{affine}}(\mathbf{x}^*[k])$. Thus, a new parameter $\theta[k+1]$ is constructed with $\mathbf{x}^*[k]$ for the next iteration.

Remark: From Fig. 2, we remark that the computing burden of each node is low since the whole optimization

problem is solved in a multi-node cooperative manner. Node N receiving and transmitting data is treated as a master node or a cluster head in the network. It mainly operates the steps of the model convexification transformation, decision variable mapping, and updating parameter θ . The other transmitting nodes (also including node N itself) only need to update their local primal and dual decision variables. The local decision variables are collected by node N via the common channel and used for updating the parameter θ . We highlight that these nodes do not need to solve the original complicated optimization problem. They only need to execute some parameter mapping or updating steps. Besides, these nodes 1 to $N-1$ do not need to exchange their local decision variables between themselves. Therefore, the proposed distributed paradigm also avoids the congestion in the common channel.

6 PERFORMANCE EVALUATION

In this section, we would like to conduct a series of numerical experiments to validate the effectiveness and the advantage of the proposed method. The experimental evaluation is twofold: (i) demonstrate the convergence of the proposed method and (ii) compare the global performance of the proposed method with some other representative centralized and distributed methods.

6.1 Global Convergence

First, we would like to examine the convergence of the distributed primal-dual optimization algorithm under different numbers of nodes. The primal and the dual step sizes are set to $\gamma = 5 \times 10^{-3}$ and $\delta = 1 \times 10^{-4}$, respectively, and the parameter θ of $\mathcal{M}_3(\theta)$ is fixed as $\theta = 0.01$ for the sake of demonstration. The error tolerance ϵ is set to $\epsilon = 1 \times 10^{-3}$. For the sake of methodological demonstration, we take into consideration a specific physical scenario similar to an unmanned aerial vehicle (UAV)-oriented wireless ad hoc network with full-duplex radios and MPR capability considered in the current literature [3] and adapt the physical-layer communication settings from the literature to our simulations here. The lower and the upper bounds of the power control of each node $i \in \mathcal{N}$ are set to $p_{i,\min} = 0$ dBm and $p_{i,\max} = 40$ dBm as used in [3], respectively. According to the existing works [3], [7], the background noise power and the self-interference power in a distributed network are usually much lower and the SINR threshold for each receiver can range within $[1, 10]$. Hence, we set the background noise power as $\sigma_i^2 = -p_{i,\max} = -40$ dBm, the SINR lower bound as $\alpha = 5$, and the self-interference cancellation coefficient as $\psi_i = 1 \times 10^{-3}$. In addition, the weights of all the nodes, ω_i , are identically set to 1, while the channel gains g_i are uniformly generated within $[0.01, 0.1]$. Here we remark that the above simulation settings are adopted for the intercomparison case study and our methodology is not limited to the application scenario specified in the simulations. Recalling that the proposed methodology explicitly considers a general SINR formulation, it can be adapted to distributed EE optimization over a wide range of wireless networks with full-duplex radios and MPR capability whose physical layer is characterized by the SINR utility.

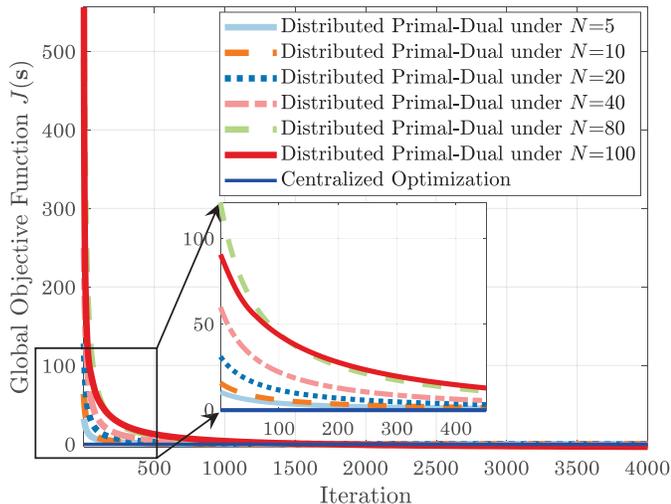


Fig. 3. The global convergence of the proposed distributed primal-dual algorithm with the extra-gradient mechanism under different node numbers.

Fig. 3 shows the convergence performance of the proposed distributed algorithm. For comparison, we also obtain the global optimal performance by using a centralized optimization based on the interior-point algorithm, which can be treated as the performance benchmark. From Fig. 3, it can be seen that the objective function $J(s)$ monotonously decreases along with numerical iterations under different N , and the proposed distributed algorithm can converge to the global optimal performance obtained by the centralized algorithm. The average numerical gap between the steady performance of the proposed algorithm and that of the centralized optimization algorithm is only about 3.6561% per node. Another fact can also be observed that the distributed algorithm needs more iterations to converge under a larger-scale network. Nevertheless, due to the distributed implementation, the convergence of the proposed algorithm is well guaranteed even when the node number N is increased from $N = 5$ to $N = 100$.

Next, we fix the node number N at $N = 100$ and further examine the proposed iterative fractional programming (IFP) combined with the distributed primal-dual optimization algorithm. In Fig. 4, we also compare the convergence performance of the proposed method with the centralized optimization algorithm. It is illustrated in this figure that the global energy cost metric $\theta(\mathbf{x})$ of the proposed method can converge consistently to the global optimal level obtained by the centralized algorithm. By comparison, Fig. 4 shows that our proposed method converges faster and it can arrive at the global optimum by fewer iterations, while the convergence rate of the centralized algorithm needs more iterations for convergence. Combining the results from Figs. 3 and 4, it is confirmed that the global convergence of the proposed method can be well guaranteed and the network can better benefit from the distributed computation in terms of convergence efficiency.

6.2 Performance Comparison

In this subsection, we evaluate and discuss the performance of the proposed IFP algorithm with the distributed primal-

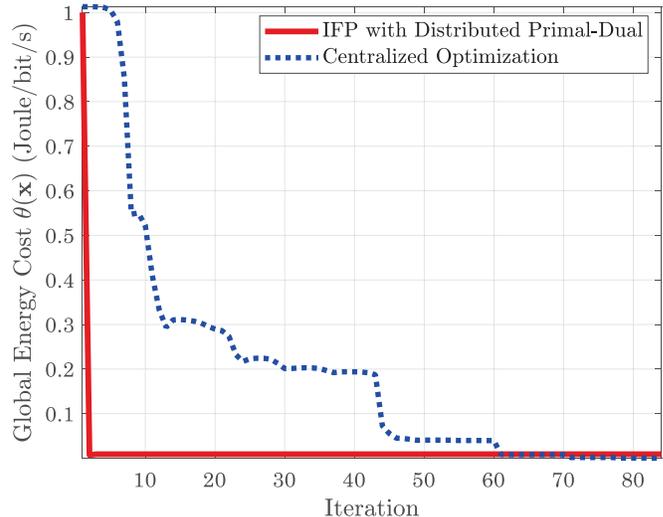


Fig. 4. The global convergence of the proposed iterative fractional programming algorithm under $N = 100$.

dual optimization (**Distributed IFP**) in terms of the global energy efficiency. Here, for comparison, we also implement four typical conventional methods for the power control of a general wireless ad hoc network, including a distributed random power control (**Distributed RPC**) method, a distributed maximum power-based control (**Distributed MPC**) method, a centralized network throughput optimization-based (**Centralized NTO**) method, and a centralized energy-efficiency optimization (**Centralized EEO**) method without model convexification. The distributed RPC method drives the nodes of the network to update their transmission powers in a random manner, i.e., randomly sampling the individual powers from the bounded interval $[p_{i,\min}, p_{i,\max}]$, which is similar to the random power control proposed in [61]. The distributed MPC method uses the maximal transmission powers of the network nodes for data transmissions, i.e., fixing the individual node's power at $p_{i,\max} = 40$ dBm, which is as the same as that in [3]. It is remarked that the distributed RPC method and the distributed MPC method are two typical baselines, which have been widely used for performance comparison in many other studies such as [2], [62]. In addition, we use the centralized NTO method and the centralized EEO method as two strong baselines that exploit the advanced nonlinear constrained programming algorithm, i.e., the sequential quadratic programming (SQP) based on the successive convex approximation technique [16], [19], to solve the original power optimization model in this paper. The successive convex approximation-based method is considered as one of the most efficient approaches to dealing with nonconvex large-scale optimization problems [63] and thus used as the strong benchmark for performance comparison here. Furthermore, we also remark that, to investigate the significance of our proposed model convexification mechanism as well as highlight the gained benefit from the proposed distributed optimization, these two competing methods are implemented in a centralized manner without integrating our model convexification in the simulation experiments. In summary, these compared methods follow either the distributed computation or the

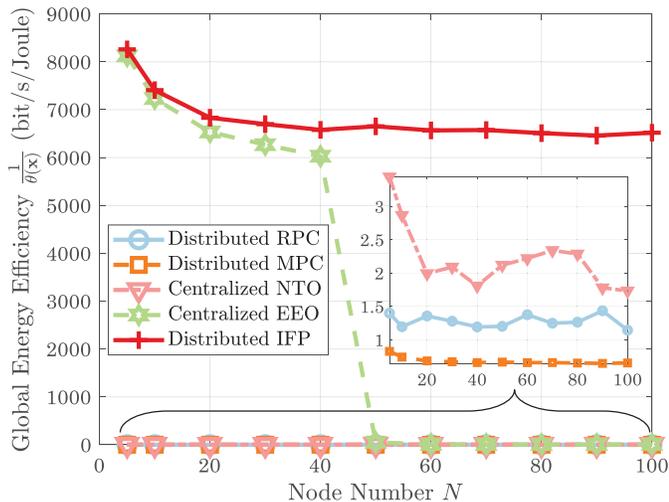


Fig. 5. The global energy efficiency of different methods under different node numbers.

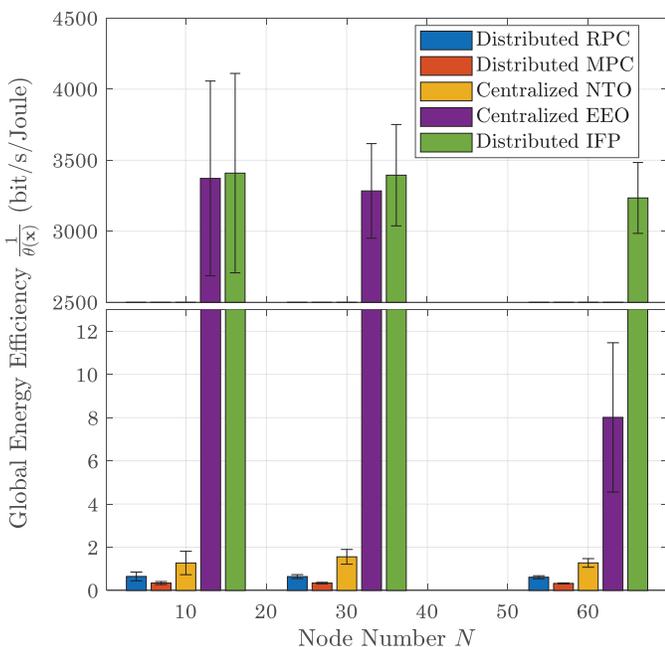


Fig. 6. The global energy efficiency of different methods under different node numbers and random weight settings.

centralized computation approaches and are representative for intercomparison simulation study.

We vary the node number N while adopting the same parameter settings as in Subsection 6.1. To facilitate the performance comparison between different methods, we focus the benefit-type energy-efficiency metric, $\frac{1}{\theta(\mathbf{x})}$, rather than the cost-type one, $\theta(\mathbf{x})$. Fig. 5 compares the global energy efficiency of different methods under different N . It is observed that the proposed method and the centralized EEO method can achieve similar energy efficiency when the node number is not large, e.g., under $N = 5, 10, 20$. Both these two methods can outperform the other three methods (Distributed RPC, Distributed MPC, and Centralized NTO). Nevertheless, when the node number increases, e.g., $N \geq 50$, the global performance of the centralized

EEO rapidly decreases and this centralized optimization method cannot guarantee the global optimality any longer, since it cannot find a global optimum and prematurely converges to a local point by directly solving the original non-convex model $\mathcal{M}_1(\theta)$. By contrast, our proposed method can achieve the highest performance among these methods and guarantee global optimality even when the node number becomes large. Specifically, the EE performance of our method is about 6.0438×10^3 bit/s/Joule more than that of the other methods on average. The underlying reason is that our method transforms the non-convex original model to a convex one by using the proposed convexification and thus enables the distributed convex optimization to come into play in solving the global optimization problem.

Moreover, we conduct Monte Carlo simulations to verify the effectiveness of our proposed method. The weight ω_i assigned for each $i \in \mathcal{N}$ is randomly generated from the interval $(0, 1)$. All the Monte Carlo simulations have been performed with 100 replications per condition point, and the results are demonstrated with the average levels and the corresponding standard deviation intervals in Fig. 6. From this figure, we observe that our method can still achieve the best energy efficiency among these compared methods. When compared to the two distributed methods, Distributed RPC and Distributed MPC, and the centralized optimization method, Centralized NTO, the global energy efficiency achieved by our method is about 3.3455×10^3 bit/s/Joule higher on average. This result indicates that our method can provide the global EE improvement of about three orders of magnitudes over the baselines. Besides, our proposed distributed method can also achieve the global energy-efficiency gain of about 3.2267×10^3 bit/s/Joule, i.e., providing two orders of magnitudes improvement in the global EE metric, over that of the centralized EEO method in the situation with a relatively large node number, i.e., $N = 60$. This result shows that the distributed computation with integrating the proposed model convexification better benefits the global EE optimization over a larger network. The centralized EEO method, even though it is implemented in a centralized optimization manner, suffers from the increased non-convexity and complexity with increasing the network size and its lack of the convexification transformation makes it fail in finding a globally optimum, yielding poor performance in a ad hoc network situation.

In Fig. 7, we compare the global performance of different methods under different background noise powers σ_i . In this experiment, the node number is fixed at $N = 30$. From Fig. 7, the global energy efficiency decreases along with increasing the noise power. The main reason is that a larger noise power can reduce the SINR of the communication links, i.e., making the quality of wireless channels degrade. Nevertheless, our proposed method still achieves the best global performance among all the comparative methods. Specifically, the global energy efficiency of our method is about 5.042×10^3 bit/s/Joule higher than that of the distributed RPC, the distributed MPC, and the centralized NTO methods on average. When compared to the centralized EEO method, our method can achieve the average EE performance gain of about 5.17%.

Fig. 8 shows the global performance of different methods under different self-interference coefficients ψ_i . It can

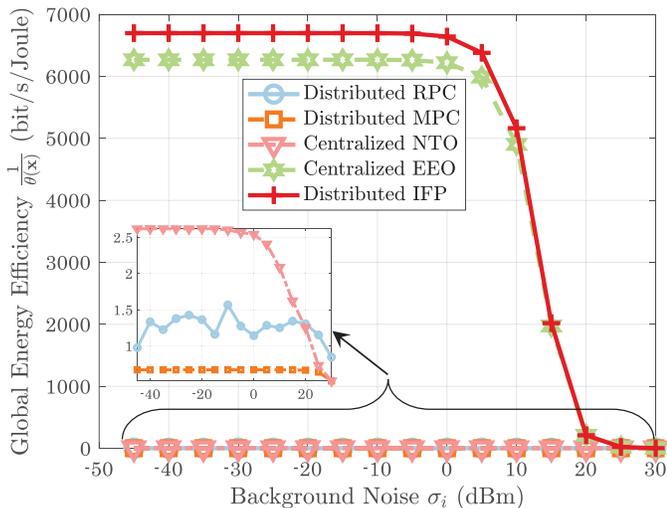


Fig. 7. The global energy efficiency of different methods under different levels of background noise power.

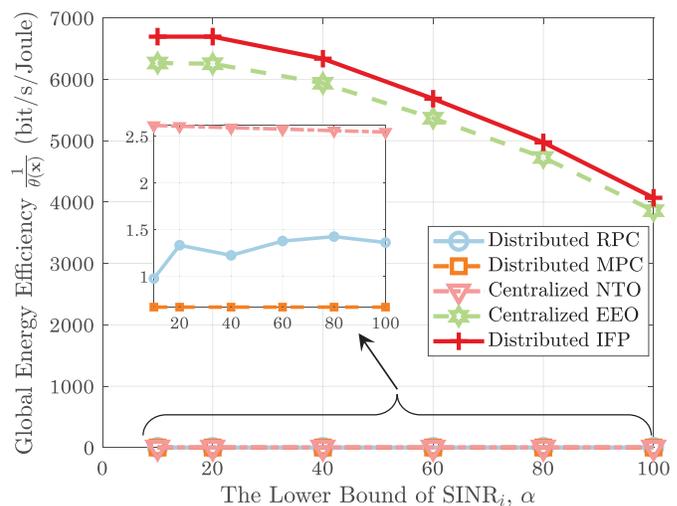


Fig. 9. The global energy efficiency of different methods under different SINR lower bounds.

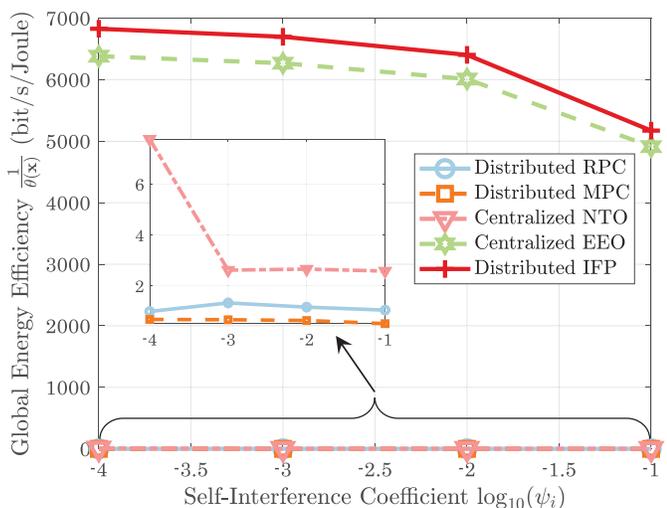


Fig. 8. The global energy efficiency of different methods under different self-interference coefficients.

be seen that increasing the self-interference coefficient can reduce the global energy efficiency of the network. The global performance of our proposed method is much greater than that of the distributed RPC, the distributed MPC, and the centralized NTO methods, respectively. In particular, the optimality achieved by our method is also better than that of the centralized EEO method. The global energy efficiency of our method increases by about 6.42% on average when compared to that of the centralized EEO method.

Finally, we compare the performance of different methods under different SINR lower bounds, i.e., α . From Fig. 9, we can find that a higher SINR bound can lead to a slight decrease in the performance of our proposed method and the centralized EEO method. Nonetheless, the global performance of our method is much better than those of the distributed RPC, the distributed MPC, and the centralized NTO methods, and is about 6.20% higher than that of the centralized EEO method on average. Combining all the results obtained under different scenarios, i.e., from Figs.

5 to 9, it can be confirmed that our proposed method, by using the convexification transformation, is able to guarantee better global optimality when compared to the centralized optimization method without convexification (i.e., the centralized EEO), and can outperform the conventional methods in terms of achieving much higher energy efficiency.

7 CONCLUSION

In this paper, we have investigated the global energy-efficiency optimization problem in general wireless ad hoc networks with consideration of multi-packet reception (MPR) capability. We have proposed a convexification transformation to map the complex non-convex problem into a convex problem and thus developed a novel iterative fractional programming framework. Based on duality and variational inequality theories, we have further proposed a distributed primal-dual optimization algorithm for solving a series of convex subproblems, which is embedded with iterative fractional programming. The convergence performance of the proposed iterative fractional programming with the distributed primal-dual optimization has been theoretically proven and numerically validated. Besides, experimental results have also shown the effectiveness and great advantage of the proposed method over the other conventional power control or optimization methods in terms of achieving global energy efficiency.

We remark that the proposed model convexification mechanism joining the affine transformation and the exponential transformation into Dinkelbach's method may facilitate addressing a broad class of non-convex and constrained energy-efficiency optimization problems characterized by the fractional programming. The methodological framework presented in this work enables the convex optimization theory to come into play and thus may motivate novel algorithm designs for non-convex and constrained resource optimization, such as the SINR-related network utility maximization, of a distributed information communication system in some other related fields. In future work, we will focus on the joint optimization of communication and

computation resources in mobile edge computing-enabled networks by integrating of the model convexification-based energy-efficiency optimization and computation offloading decision-making. We would also like to investigate the cross-layer optimization, guarantee the max-min energy-efficiency fairness, and allow for the integration of the proposed distributed optimization model and the spatial link diversity and channel coding. It is also expected to develop a prototype of the targeted system and facilitate practical applications based on the proposed distributed optimization over wireless networks.

ACKNOWLEDGMENTS

This research was supported in part by the National Key Research and Development Program of China under Grant No. 2022YFC3803700, the National Natural Science Foundation of China under Grant No. 52202391 and No. U20A20155, and the Opening Project of Ministry of Transport Key Laboratory of Technology on Intelligent Transportation Systems under Grant No. F20211746.

REFERENCES

- [1] J. Zhou, D. Tian, Z. Sheng, X. Duan, G. Qu, D. Zhao, D. Cao, and X. Shen, "Robust min-max model predictive vehicle platooning with causal disturbance feedback," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 15 878–15 897, Sep. 2022.
- [2] S. Zhang, J. Zhou, D. Tian, Z. Sheng, X. Duan, and V. C. M. Leung, "Robust cooperative communication optimization for multi-UAV-aided vehicular networks," *IEEE Wireless Communications Letters*, vol. 10, no. 4, pp. 780–784, 2021.
- [3] Y. Cai, F. R. Yu, J. Li, Y. Zhou, and L. Lamont, "Medium access control for unmanned aerial vehicle (UAV) ad-hoc networks with full-duplex radios and multipacket reception capability," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 390–394, Jan 2013.
- [4] J. Hu, W. Heng, X. Li, J. Wu, K. Wang, and G. Wang, "Energy-efficient power control with max-min fairness for energy harvesting ad hoc networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2019, pp. 1–5.
- [5] M. Chiang, C. W. Tan, D. P. Palomar, D. O'Neill, and D. Julian, "Power control by geometric programming," *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2640–2651, July 2007.
- [6] N. Pappas, M. Kountouris, A. Ephremides, and A. Traganitis, "Relay-assisted multiple access with full-duplex multi-packet reception," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3544–3558, July 2015.
- [7] G. D. Celik, G. Zussman, W. F. Khan, and E. Modiano, "MAC for networks with multipacket reception capability and spatially distributed nodes," *IEEE Transactions on Mobile Computing*, vol. 9, no. 2, pp. 226–240, Feb 2010.
- [8] J. H. Sarker and H. T. Mouftah, "A random access protocol with multi-packet reception for infrastructure-less wireless autonomic networks," *Wireless Personal Communications*, vol. 56, no. 3, pp. 447–456, Feb 2011.
- [9] Fei Yu, V. Krishnamurthy, and V. C. M. Leung, "Cross-layer optimal connection admission control for variable bit rate multimedia traffic in packet wireless cdma networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 542–555, Feb 2006.
- [10] H. Chen, F. Yu, H. C. B. Chan, and V. C. M. Leung, "A novel multiple access scheme over multi-packet reception channels for wireless multimedia networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1501–1511, April 2007.
- [11] Q. Wu, M. Tao, D. W. Kwan Ng, W. Chen, and R. Schober, "Energy-efficient resource allocation for wireless powered communication networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2312–2327, March 2016.
- [12] A. Zappone, L. Sanguinetti, G. Bacci, E. Jorswieck, and M. Debbah, "Energy-efficient power control: A look at 5G wireless technologies," *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1668–1683, April 2016.
- [13] K. Shen and W. Yu, "Fractional programming for communication systems—part i: Power control and beamforming," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2616–2630, May 2018.
- [14] —, "Fractional programming for communication systems—part ii: Uplink scheduling via matching," *IEEE Transactions on Signal Processing*, vol. 66, no. 10, pp. 2631–2644, May 2018.
- [15] A. Zappone and E. Jorswieck, *Energy Efficiency in Wireless Networks via Fractional Programming Theory*. now, 2015. [Online]. Available: <https://ieeexplore.ieee.org/document/8187084>
- [16] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, New York, NY, 2006. [Online]. Available: <https://doi.org/10.1007/978-0-387-40065-5>
- [17] S. Leyffer and A. Mahajan, *Nonlinear Constrained Optimization method and software*. ARGONNE NATIONAL LABORATORY, 2010. [Online]. Available: <https://wiki.mcs.anl.gov/leyffer/images/7/75/NLP-Solvers.pdf>
- [18] H. Boche and S. Naik, "Impact of interference coupling - loss of convexity," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, Nov 2009, pp. 1–6.
- [19] Zhi-Quan Luo and Wei Yu, "An introduction to convex optimization for communications and signal processing," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1426–1438, Aug 2006.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*. USA: Cambridge University Press, 2004.
- [21] Z. Luo and S. Zhang, "Duality gap estimation and polynomial time approximation for optimal spectrum management," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2675–2689, July 2009.
- [22] S. Schaible, "Fractional programming. i. duality," *Management Science*, vol. 22, no. 8, pp. 858–867, 1976. [Online]. Available: <http://www.jstor.org/stable/2630017>
- [23] —, "Fractional programming. ii, on Dinkelbach's algorithm," *Management Science*, vol. 22, no. 8, pp. 868–873, 1976. [Online]. Available: <http://www.jstor.org/stable/2630018>
- [24] K. T. K. Cheung, S. Yang, and L. Hanzo, "Achieving maximum energy-efficiency in multi-relay ofdma cellular networks: A fractional programming approach," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2746–2757, July 2013.
- [25] K. Shen, W. Yu, L. Zhao, and D. P. Palomar, "Optimization of MIMO device-to-device networks via matrix fractional programming: A minorization–maximization approach," *IEEE/ACM Transactions on Networking*, vol. 27, no. 5, pp. 2164–2177, Oct 2019.
- [26] A. Nedić and A. Ozdaglar, *Cooperative distributed multi-agent optimization*. Cambridge University Press, 2009, p. 340–386.
- [27] A. Nedić, *Distributed Optimization Over Networks*. Cham: Springer International Publishing, 2018, pp. 1–84. [Online]. Available: https://doi.org/10.1007/978-3-319-97142-1_1
- [28] J. Koshal, A. Nedić, and U. V. Shanbhag, "Multiuser optimization: Distributed algorithms and error analysis," *SIAM Journal on Optimization*, vol. 21, no. 3, pp. 1046–1081, 2011. [Online]. Available: <https://doi.org/10.1137/090770102>
- [29] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multi-agent optimization," *IEEE Transactions on Automatic Control*, vol. 54, no. 1, pp. 48–61, Jan 2009.
- [30] S. Sundhar Ram, A. Nedić, and V. V. Veeravalli, "Distributed stochastic subgradient projection algorithms for convex optimization," *Journal of Optimization Theory and Applications*, vol. 147, no. 3, pp. 516–545, December 2010.
- [31] P. Lin and W. Ren, "Distributed subgradient projection algorithm for multi-agent optimization with nonidentical constraints and switching topologies," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, Dec 2012, pp. 6813–6818.
- [32] H. Li, Q. Lü, and T. Huang, "Distributed projection subgradient algorithm over time-varying general unbalanced directed graphs," *IEEE Transactions on Automatic Control*, vol. 64, no. 3, pp. 1309–1316, March 2019.
- [33] Y. Zhu, W. Yu, G. Wen, G. Chen, and W. Ren, "Continuous-time distributed subgradient algorithm for convex optimization with general constraints," *IEEE Transactions on Automatic Control*, vol. 64, no. 4, pp. 1694–1701, April 2019.
- [34] T. Chang, A. Nedić, and A. Scaglione, "Distributed constrained optimization by consensus-based primal-dual perturbation method," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1524–1538, June 2014.
- [35] A. Falsone, K. Margellos, S. Garatti, and M. Prandini, "Dual decomposition for multi-agent distributed optimization

- with coupling constraints," *Automatica*, vol. 84, pp. 149 – 158, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0005109817303291>
- [36] Q. Yang and G. Chen, "Primal-dual subgradient algorithm for distributed constraint optimization over unbalanced digraphs," *IEEE Access*, vol. 7, pp. 85 190–85 202, 2019.
- [37] M. Zhu and S. Martínez, "An approximate dual subgradient algorithm for multi-agent non-convex optimization," *IEEE Transactions on Automatic Control*, vol. 58, no. 6, pp. 1534–1539, June 2013.
- [38] T. Chang, "A proximal dual consensus admm method for multi-agent constrained optimization," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3719–3734, July 2016.
- [39] D. Yuan, D. W. C. Ho, and G. Jiang, "An adaptive primal-dual subgradient algorithm for online distributed constrained optimization," *IEEE Transactions on Cybernetics*, vol. 48, no. 11, pp. 3045–3055, Nov 2018.
- [40] D. Yuan, D. W. C. Ho, and S. Xu, "Regularized primal-dual subgradient method for distributed constrained optimization," *IEEE Transactions on Cybernetics*, vol. 46, no. 9, pp. 2109–2118, Sep. 2016.
- [41] D. Yuan, S. Xu, and H. Zhao, "Distributed primal-dual subgradient method for multiagent optimization via consensus algorithms," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 6, pp. 1715–1724, Dec 2011.
- [42] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4289–4305, Aug 2012.
- [43] K. Yuan, B. Ying, X. Zhao, and A. H. Sayed, "Exact diffusion for distributed optimization and learning—part i: Algorithm development," *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 708–723, Feb 2019.
- [44] —, "Exact diffusion for distributed optimization and learning—part ii: Convergence analysis," *IEEE Transactions on Signal Processing*, vol. 67, no. 3, pp. 724–739, Feb 2019.
- [45] B. Yang and M. Johansson, *Distributed Optimization and Games: A Tutorial Overview*. London: Springer London, 2010, pp. 109–148. [Online]. Available: https://doi.org/10.1007/978-0-85729-033-5_4
- [46] J. Barreiro-Gomez, N. Quijano, and C. Ocampo-Martinez, "Constrained distributed optimization: A population dynamics approach," *Automatica*, vol. 69, pp. 101 – 116, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S000510981630036X>
- [47] M. Zhu and S. Martinez, "On distributed convex optimization under inequality and equality constraints," *IEEE Transactions on Automatic Control*, vol. 57, no. 1, pp. 151–164, Jan 2012.
- [48] Chi Wan Sung, "Log-convexity property of the feasible sir region in power-controlled cellular systems," *IEEE Communications Letters*, vol. 6, no. 6, pp. 248–249, June 2002.
- [49] S. Stanczak and H. Boche, "On the convexity of feasible QoS regions," *IEEE Transactions on Information Theory*, vol. 53, no. 2, pp. 779–783, Feb 2007.
- [50] H. Boche and S. Stanczak, "Convexity of some feasible QoS regions and asymptotic behavior of the minimum total power in cdma systems," *IEEE Transactions on Communications*, vol. 52, no. 12, pp. 2190–2197, Dec 2004.
- [51] D. Julian, Mung Chiang, D. O'Neill, and S. Boyd, "QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks," in *Proceedings. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2, June 2002, pp. 477–486.
- [52] C. W. Tan, D. P. Palomar, and M. Chiang, "Exploiting hidden convexity for flexible and robust resource allocation in cellular networks," in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, May 2007, pp. 964–972.
- [53] M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," *Found. Trends Netw.*, vol. 2, no. 4, p. 381–533, Apr. 2008. [Online]. Available: <https://doi.org/10.1561/1300000009>
- [54] A. Klinger and O. Mangasarian, "Logarithmic convexity and geometric programming," *Journal of Mathematical Analysis and Applications*, vol. 24, no. 2, pp. 388 – 408, 1968. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0022247X68900395>
- [55] D. Tse and S. Hanly, "Linear multiuser receivers: effective interference, effective bandwidth and user capacity," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 641–657, March 1999.
- [56] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, March 2000.
- [57] D. Niyato, E. Hossain, and P. Wang, "Optimal channel access management with qos support for cognitive vehicular networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 4, pp. 573–591, April 2011.
- [58] R. Jagannathan, "On some properties of programming problems in parametric form pertaining to fractional programming," *Management Science*, vol. 12, no. 7, pp. 609–615, 1966. [Online]. Available: <https://doi.org/10.1287/mnsc.12.7.609>
- [59] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967. [Online]. Available: <https://doi.org/10.1287/mnsc.13.7.492>
- [60] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag New York, 2003. [Online]. Available: <https://doi.org/10.1007/b97543>
- [61] X. Zhang and M. Haenggi, "Random power control in Poisson networks," *IEEE Transactions on Communications*, vol. 60, no. 9, pp. 2602–2611, 2012.
- [62] Y. Liu, J. Zhou, D. Tian, Z. Sheng, X. Duan, G. Qu, and D. Zhao, "Joint optimization of resource scheduling and mobility for UAV-assisted vehicle platoons," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 1–5.
- [63] G. Scutari and Y. Sun, *Parallel and Distributed Successive Convex Approximation Methods for Big-Data Optimization*. Cham: Springer International Publishing, 2018, pp. 141–308. [Online]. Available: https://doi.org/10.1007/978-3-319-97142-1_3



Jianshan Zhou received the B.Sc., M.Sc., and Ph.D. degrees in traffic information engineering and control from Beihang University, Beijing, China, in 2013, 2016, and 2020, respectively. He is an associate professor with the school of transportation science and engineering at Beihang University. From 2017 to 2018, he was a Visiting Research Fellow with the School of Informatics and Engineering, University of Sussex, Brighton, U.K. He was a Postdoctoral Research Fellow supported by the Zhuoyue Program of Beihang University and the National Postdoctoral Program for Innovative Talents from 2020 to 2022. He is or was the Technical Program Session Chair with the IEEE EDGE 2020, the IEEE ICUS 2022, the ICAUS 2022, the TPC member with the IEEE VTC2021-Fall track, and the Youth Editorial Board Member of the Unmanned Systems Technology. He is the author or co-author of more than 30 international scientific publications. His research interests include the modeling and optimization of vehicular communication networks and air-ground cooperative networks, the analysis and control of connected autonomous vehicles, and intelligent transportation systems. He was the recipient of the First Prize in the Science and Technology Award from the China Intelligent Transportation Systems Association in 2017, the First Prize in the Innovation and Development Award from the China Association of Productivity Promotion Centers in 2020, the Second Prize in the Beijing Science and Technology Progress Award in 2022, the National Scholarships in 2017 and 2019, the Outstanding Top-Ten Ph.D. Candidate Prize from Beihang University in 2018, the Outstanding China-SAE Doctoral Dissertation Award in 2020, and the Excellent Doctoral Dissertation Award from Beihang University in 2021.



Daxin Tian [M'13-SM'16] received the Ph.D. degree in Technology of Computer Application from Jilin University, China, in 2007. He is currently a Professor with the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests include mobile computing, intelligent transportation systems, vehicular ad hoc networks, and swarm intelligence. Prof. Tian leads about 11 research projects such as the projects funded by the National Natural Science Foundation and the

National Key Research and Development Program. He has authored/co-authored about 213 journal/conference papers, published 7 monographs and 2 translations, and authorized 34 invention patents. He was the recipient of the Second Prize of the National Science and Technology Award in 2015 and 2018, the First Prize of the Technical Invention Award of the Ministry of Education in 2017, the First Prize of the Science and Technology Award from the China Intelligent Transportation Association in 2017, the First Prize of the Innovation and Development Award from the China Association of Productivity Promotion Centers in 2020, and seven other ministerial and provincial science and technology awards. He also received the Changjiang Youth Scholars Program of China in 2018 and the Outstanding Youth Fund from the National Natural Science Foundation of China in 2019, the Forum Keynote Award from the 2019 Cyberspace Congress, the Outstanding Invited Speaker from the 2020 International Conference on Blockchain and Trustworthy Systems, and the Distinguished Young Investigator of China Frontiers of Engineering from Chinese Academy of Engineering in 2018. He was also awarded the Exemplary Reviewer for IEEE Wireless Communications Letters. He is a senior member of IEEE, CCF, and ITSC, and was or is the Editor-in-Chief of International Journal of Vehicular Telematics and Infotainment Systems, the Associate Editor of IEEE Transactions on Intelligent Vehicles, IEEE Internet of Things Journal, Complex System Modeling and Simulation, and Journal of Intelligent and Connected Vehicles.



Guixian Qu received the B.Sc. degree in transportation engineering from Shandong University of Technology, Shandong, China, in 2012, the M.Sc. and Ph.D. degrees from Beijing University of Technology, Beijing, China, in 2014 and 2019, respectively. She was a Postdoctoral Research Fellow of Beihang University. She is currently an Research Fellow of Research Institute of Aero-Engine, Beihang University. Her research interests include intelligent transportation systems, dynamics modeling and control, and distributed

optimization. She was the recipient of the First Prize in the Science and Technology Award from the China Communications and Transportation Association in 2019.



Zhengguo Sheng [SM'18] received the B.Sc. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2006, and the M.S. and Ph.D. degrees from Imperial College London, London, U.K., in 2007 and 2011, respectively. He is currently a Senior Lecturer with the University of Sussex, Brighton, U.K. Previously, he was with UBC, Vancouver, BC, Canada, as a Research Associate and with Orange Labs, Santa Monica, CA, USA, as a Senior Researcher. He has more than 100 publi-

cations. His research interests cover IoT, vehicular communications, and cloud/edge computing.



Xuting Duan received the Ph.D. degree in traffic information engineering and control from Beihang University, Beijing, China, in 2017. He is currently an assistant professor with the School of Transportation Science and Engineering, Beihang University, Beijing, China. His current research interests are focused on vehicular ad hoc networks.



Victor C. M. Leung [S'75-M'89-SM'97-F'03] is a Distinguished Professor of Computer Science and Software Engineering at Shenzhen University, Shenzhen, China, and a Professor Emeritus at the University of British Columbia (UBC), Vancouver, Canada. Before he retired from UBC at the end of 2018, he was a Professor of Electrical and Computer Engineering and holder of the TELUS Mobility Research Chair there. His research is in the broad areas of wireless networks and mobile systems. He has coauthored more

than 1300 journal/conference papers and book chapters. Dr. Leung is serving on the editorial boards of the IEEE Transactions on Green Communications and Networking, IEEE Transactions on Cloud Computing, IEEE Access, IEEE Network, and several other journals. He received the IEEE Vancouver Section Centennial Award, 2011 UBC Killam Research Prize, 2017 Canadian Award for Telecommunications Research, and 2018 IEEE TCGCC Distinguished Technical Achievement Recognition Award. He co-authored papers that won the 2017 IEEE ComSoc Fred W. Ellersick Prize, 2017 IEEE Systems Journal Best Paper Award, 2018 IEEE CSIM Best Journal Paper Award, and 2019 IEEE TCGCC Best Journal Paper Award. He is a Fellow of IEEE, the Royal Society of Canada, Canadian Academy of Engineering, and Engineering Institute of Canada. He is named in the current Clarivate Analytics list of "Highly Cited Researchers".

Supplementary Material for Energy-Efficiency Optimization with Model Convexification for Wireless Ad Hoc Networks with Multi-Packet Reception Capability

Jianshan Zhou, Daxin Tian, *Senior Member, IEEE*, Guixian Qu, Zhengguo Sheng, *Senior Member, IEEE*,
Xuting Duan, and Victor C. M. Leung, *Life Fellow, IEEE*



1 PROOF OF THEOREM 3

It can be seen that $\sum_{i=1}^N C_i \phi(s_i)/a_i$ is jointly convex with respect to \mathbf{s} since $a_i > 0$ and $\phi(s_i)$ is a strictly convex function of s_i for all $i \in \mathcal{N}$. Recall \mathcal{S} is a convex box-constraint set. To prove the convexity of $\mathcal{M}_3(\theta)$, we only need to prove that both the second term of its objective function and the inequality function $\varphi_i(\phi(\mathbf{s}))$ are jointly convex with respect to \mathbf{s} . In fact, for all $i \in \mathcal{N}$, it can be observed that $\tilde{I}_i(\phi(\mathbf{s}))$ is logarithmically convex, i.e., a log-convex function, with respect to \mathbf{s} . This further indicates the joint convexity of $\log_2(\tilde{I}_i(\phi(\mathbf{s}))/\phi(s_i))$ with respect to \mathbf{s} . Since the second term of its objective function and the inequality function $\varphi_i(\phi(\mathbf{s}))$ are a linear combination of $\log_2(\tilde{I}_i(\phi(\mathbf{s}))/\phi(s_i))$, respectively, they also meet the joint convexity.

2 PROOF OF LEMMA 1

For $j = 1, 2, \dots, N$ and $j = i$, we can have

$$\begin{aligned} [\nabla_{\mathbf{s}} H(\mathbf{s})]_{j,i} &= \frac{C_i \phi(s_i)}{a_i} \\ &+ \sum_{l=1, l \neq i}^N \frac{\theta \omega_l b_{i,l} \phi(s_i) \tilde{I}_l(\phi(\mathbf{s})) - \theta \omega_l b_{i,l}^2 \phi^2(s_i)}{(\ln 2) \tilde{I}_l^2(\phi(\mathbf{s}))}. \end{aligned} \quad (\text{S.1})$$

Notice $|c_{0,i}| \leq \tilde{I}_i(\phi(\mathbf{s}))$ for all $i \in \mathcal{N}$. We can see $\|[\nabla_{\mathbf{s}} H(\mathbf{s})]_{j,i}\| \leq L_{j,i}$ in this case.

For $j = 1, 2, \dots, N$ and $j \neq i$, we can also see

$$[\nabla_{\mathbf{s}} H(\mathbf{s})]_{j,i} = - \sum_{l=1, l \neq i}^N \frac{\theta \omega_l b_{i,l} b_{j,l} \phi(s_i) \phi(s_j)}{(\ln 2) \tilde{I}_l^2(\phi(\mathbf{s}))} \quad (\text{S.2})$$

and $\|[\nabla_{\mathbf{s}} H(\mathbf{s})]_{j,i}\| \leq L_{j,i}$.

In matrix theory, the spectral radius of a matrix is always not larger than any natural matrix norm. Since $L = \sqrt{\sum_{j=1}^N \sum_{i=1}^N L_{j,i}^2}$, the spectral radius of $\nabla_{\mathbf{s}} H(\mathbf{s})$, denoted by $\rho(\nabla_{\mathbf{s}} H(\mathbf{s}))$, always satisfies $\rho(\nabla_{\mathbf{s}} H(\mathbf{s})) \leq \|\nabla_{\mathbf{s}} H(\mathbf{s})\| \leq L$. At this point, the lemma is proven.

3 PROOF OF LEMMA 2

Notice that $\nabla_{\mathbf{s}}^2 \varphi_i(\mathbf{s}) = \nabla_{\mathbf{s}}^2 \log_2 \circ \tilde{I}_i$ for all $i \in \mathcal{N}$ and that for each $i = 1, 2, \dots, N$

$$\begin{aligned} & \left[\nabla_{\mathbf{s}}^2 \log_2 \circ \tilde{I}_i \right]_{l_1, l_2} = \\ & \begin{cases} \frac{b_{l_1, i} \exp(s_{l_1}) \tilde{I}_i(\phi(\mathbf{s})) - b_{l_1, i}^2 \exp(2s_{l_1})}{\ln 2 (\tilde{I}_i^2(\phi(\mathbf{s})))}, & l_1 = l_2; \\ \frac{-b_{l_1, i} b_{l_2, i} \exp(s_{l_1} + s_{l_2})}{\ln 2 (\tilde{I}_i^2(\phi(\mathbf{s})))}, & l_1 \neq l_2, \end{cases} \end{aligned} \quad (\text{S.3})$$

where $l_1, l_2 = 1, 2, \dots, N$.

By using the same logic in Lemma 1, we can also prove $\|\nabla_{\mathbf{s}}^2 \varphi_i(\mathbf{s})\| \leq \kappa_{l_1, l_2}(i)$ for all l_1, l_2 and i .

4 PROOF OF LEMMA 3

For any two pairs $\mathbf{z}_1 = \text{col}\{\mathbf{s}_1, \boldsymbol{\lambda}_1\}$, $\mathbf{z}_2 = \text{col}\{\mathbf{s}_2, \boldsymbol{\lambda}_2\}$, and $\mathbf{z}_1, \mathbf{z}_2 \in \mathcal{S} \times \mathcal{U}$, we can get based on (40)

$$\begin{aligned} & (G(\mathbf{z}_1) - G(\mathbf{z}_2))^{\text{T}} (\mathbf{z}_1 - \mathbf{z}_2) \\ &= (\nabla_{\mathbf{s}} J(\mathbf{s}_1) - \nabla_{\mathbf{s}} J(\mathbf{s}_2))^{\text{T}} (\mathbf{s}_1 - \mathbf{s}_2) \\ &+ \sum_{i=1}^N (\lambda_{1,i} \nabla_{\mathbf{s}} \varphi_i(\mathbf{s}_1) - \lambda_{2,i} \nabla_{\mathbf{s}} \varphi_i(\mathbf{s}_2))^{\text{T}} (\mathbf{s}_1 - \mathbf{s}_2) \quad (\text{S.4}) \\ &- \sum_{i=1}^N (\varphi_i(\mathbf{s}_1) - \varphi_i(\mathbf{s}_2)) (\lambda_{1,i} - \lambda_{2,i}) \end{aligned}$$

where we let $\lambda_{1,i}$ and $\lambda_{2,i}$ denote the i -th components in $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$, respectively. Recalling the convexity of $J(\mathbf{s})$, the Hessian matrix of $J(\mathbf{s})$, $\nabla_{\mathbf{s}}^2 J(\mathbf{s})$, must be positive definite. Hence, $\nabla_{\mathbf{s}} J(\mathbf{s})$ is monotone. This means that

$(\nabla_{\mathbf{s}}J(\mathbf{s}_1) - \nabla_{\mathbf{s}}J(\mathbf{s}_2))^T(\mathbf{s}_1 - \mathbf{s}_2) \geq 0$. Thus, (S.4) can be rearranged as

$$\begin{aligned}
& (G(\mathbf{z}_1) - G(\mathbf{z}_2))^T(\mathbf{z}_1 - \mathbf{z}_2) \\
& \geq \sum_{i=1}^N \lambda_{1,i} \left[\varphi_i(\mathbf{s}_2) - \varphi_i(\mathbf{s}_1) + (\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_1))^T(\mathbf{s}_1 - \mathbf{s}_2) \right] \\
& \quad + \sum_{i=1}^N \lambda_{2,i} \left[\varphi_i(\mathbf{s}_1) - \varphi_i(\mathbf{s}_2) - (\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2))^T(\mathbf{s}_1 - \mathbf{s}_2) \right] \\
& = \sum_{i=1}^N \frac{\lambda_{1,i}}{2} (\mathbf{s}_2 - \mathbf{s}_1)^T (\nabla_{\mathbf{s}}^2\varphi_i(\boldsymbol{\xi}_1)) (\mathbf{s}_2 - \mathbf{s}_1) \\
& \quad + \sum_{i=1}^N \frac{\lambda_{2,i}}{2} (\mathbf{s}_1 - \mathbf{s}_2)^T (\nabla_{\mathbf{s}}^2\varphi_i(\boldsymbol{\xi}_2)) (\mathbf{s}_1 - \mathbf{s}_2)
\end{aligned} \tag{S.5}$$

where the far right term of (S.5) follows the theory of Taylor's expansion. $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are two points that can be expressed as $\boldsymbol{\xi}_1 = a_1\mathbf{s}_1 + (1 - a_1)\mathbf{s}_2$ and $\boldsymbol{\xi}_2 = a_2\mathbf{s}_1 + (1 - a_2)\mathbf{s}_2$ where a_1 and a_2 are two some points within $[0, 1]$. According to the non-negativity of $\lambda_{1,i}$ and $\lambda_{2,i}$, and the positive definition of $\varphi_i(\mathbf{s})$, $\forall i \in \mathcal{N}$, we can see $(G(\mathbf{z}_1) - G(\mathbf{z}_2))^T(\mathbf{z}_1 - \mathbf{z}_2) \geq 0$. At this point, the monotonicity of $G(\mathbf{z})$ can be proven.

Next step, we would like to show the Lipschitz continuity of $G(\mathbf{z})$. According to (S.4) and Cauchy-Schwarz inequality, we can see

$$\begin{aligned}
\|G(\mathbf{z}_1) - G(\mathbf{z}_2)\| & \leq \|\nabla_{\mathbf{s}}J(\mathbf{s}_1) - \nabla_{\mathbf{s}}J(\mathbf{s}_2)\| \\
& \quad + \left\| \sum_{i=1}^N (\lambda_{1,i}\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_1) - \lambda_{2,i}\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2)) \right\| \\
& \quad + \|\varphi(\mathbf{s}_1) - \varphi(\mathbf{s}_2)\|
\end{aligned} \tag{S.6}$$

We can first get

$$\begin{aligned}
& \left\| \sum_{i=1}^N (\lambda_{1,i}\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_1) - \lambda_{2,i}\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2)) \right\| \\
& = \left\| \sum_{i=1}^N \lambda_{1,i} (\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_1) - \nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2)) \right. \\
& \quad \left. + \sum_{i=1}^N (\lambda_{1,i} - \lambda_{2,i}) \nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2) \right\|,
\end{aligned} \tag{S.7}$$

which further leads to

$$\begin{aligned}
& \left\| \sum_{i=1}^N (\lambda_{1,i}\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_1) - \lambda_{2,i}\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2)) \right\| \\
& \leq \sum_{i=1}^N \lambda_{1,i} \|\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_1) - \nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2)\| \\
& \quad + \sum_{i=1}^N |\lambda_{1,i} - \lambda_{2,i}| \|\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2)\| \\
& \leq \sum_{i=1}^N \lambda_{1,i} \|\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_1) - \nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2)\| + \mu_{\varphi} \|\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2\|.
\end{aligned} \tag{S.8}$$

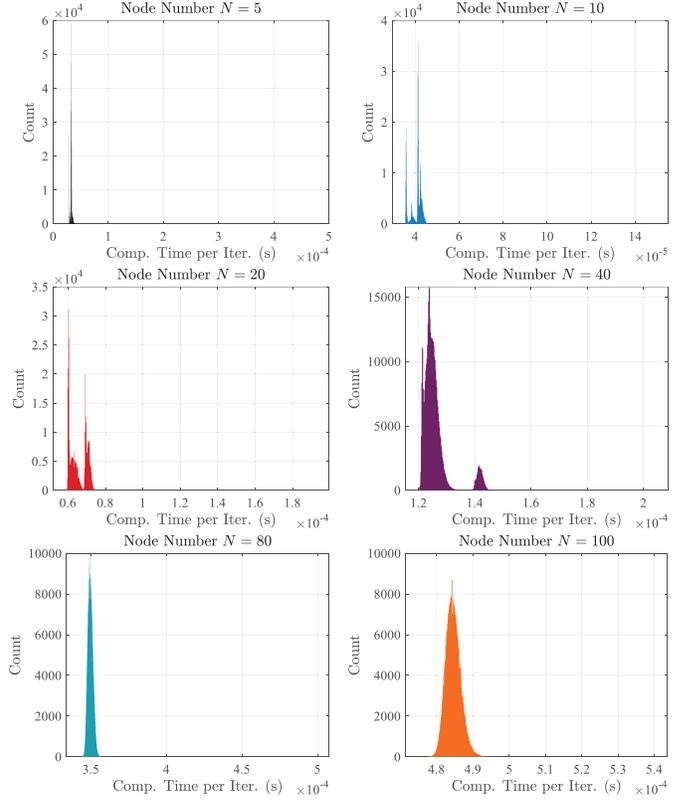


Fig. 1. The distribution of the measures on the computation time per iteration of each node using our distributed algorithm.

Using Cauchy-Schwarz inequality and the Lipschitz continuity of $\nabla_{\mathbf{s}}\varphi(\mathbf{s})$, $i \in \mathcal{N}$, we get

$$\begin{aligned}
& \sum_{i=1}^N \lambda_{1,i} \|\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_1) - \nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2)\| \\
& \leq \|\boldsymbol{\lambda}_1\| \sqrt{\sum_{i=1}^N \|\nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_1) - \nabla_{\mathbf{s}}\varphi_i(\mathbf{s}_2)\|^2} \\
& \leq \mu_{\lambda} L_G \|\mathbf{s}_1 - \mathbf{s}_2\|.
\end{aligned} \tag{S.9}$$

Using the mean-value theorem, we also get

$$\|\varphi(\mathbf{s}_1) - \varphi(\mathbf{s}_2)\| = \|(\nabla_{\mathbf{s}}\varphi(\boldsymbol{\xi}_3))^T(\mathbf{s}_1 - \mathbf{s}_2)\| \leq \mu_{\varphi} \|\mathbf{s}_1 - \mathbf{s}_2\| \tag{S.10}$$

Now, combining (50), (S.8), (S.9), (S.10) with (S.6) can get

$$\begin{aligned}
\|G(\mathbf{z}_1) - G(\mathbf{z}_2)\| & \leq (L + \mu_{\lambda} L_G + \mu_{\varphi}) \|\mathbf{s}_1 - \mathbf{s}_2\| + \mu_{\varphi} \|\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2\| \\
& \leq \sqrt{(L + \mu_{\lambda} L_G + \mu_{\varphi})^2 + \mu_{\varphi}^2} \|\mathbf{z}_1 - \mathbf{z}_2\|,
\end{aligned} \tag{S.11}$$

which concludes the lemma.

5 COMPUTATION TIME OF ALGORITHM

Our distributed EE optimization algorithm is implemented with MATLAB and run on a single computer with the specific hardware conditions: Intel(R) Core(TM) i7-8750H

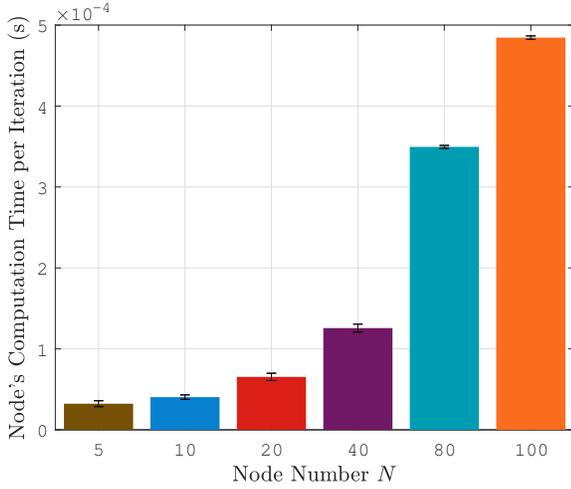


Fig. 2. The computation time per iteration of each node using our distributed algorithm.

CPU 2.20 GHz-2.21 GHz and RAM 8.00 GB.¹ To illustrate the computation time of our algorithm, we have conducted Monte Carlo simulations. The simulation settings are specified as the same as those in Subsection 6.1 of the main text. The Monte Carlo simulations are performed with 100 replications per network size. Note that the network nodes perform 4000 algorithmic iterations in each Monte Carlo simulation. We measure the computation time each node takes for each iteration execution and thus totally obtain about $4 \times 10^5 = 100 \times 4000$ samples per node. In Fig. 1, we show the distribution of the measures on the computation time per iteration of each node using our distributed algorithm. From this figure, it is seen that the computation time per iteration under different network sizes, $N = 5, 10, 20, 40, 80, 100$, ranges in the order of tens of microseconds to hundreds of microseconds. Due to the existence of random measurement errors in the wall-clock time of the computer hardware, the collected measures on the computation time per iteration generally do not follow a uniform distribution. Thus, we further evaluate the average results and the corresponding standard deviations based on the measures. Fig. 2 shows the computation time per iteration of each node using our proposed algorithm on average and the error bars indicate the corresponding

1. Here the computing platform is used for the network simulation purpose. This platform is a multi-task system that undertakes various software applications besides the simulation experiment. Thus, only partial computing resources are allocated to the simulation process. Our proposed optimization algorithm is not limited to the specific platform. In actual engineering implementation, the computing capacity of the hardware used to deploy the algorithm program is allowed to be lower than that used in the simulation experiment. From the perspective of practical engineering development, we need to choose suitable computing hardware for the algorithm. This choice will heavily depend on the deployment scenario of interest. For example, we can use a computing platform with a 64-bit Cortex-A53 processor for processing a wide range of complex tasks in power-constrained application scenarios, such as vehicle infotainment systems and mobile smart devices undertaking artificial intelligence and machine learning algorithms. Besides, the integration design of software and hardware and the code optimization are significant to improve the execution efficiency of the optimization algorithm. This issue is out of the scope of this work, and we consider it an important future direction to extend our study.

standard deviation intervals. From the figure, the average computation time per algorithm iteration of each node is about 3.2281×10^{-5} s, 4.0401×10^{-5} s, and 6.5399×10^{-5} s under $N = 5, 10, 20$, respectively. By comparison, the metric is about 1.2568×10^{-4} s, 3.4956×10^{-4} s, and 4.8454×10^{-4} s under $N = 40, 80, 100$, respectively.

Moreover, recall that, from Figure 3 in the main text of the paper, it takes about 500 to 800 iterations to reach convergence with a good numerical accuracy under the network size of $N = 5, 10, 20$. This means that the total computation time for algorithm convergence per node is 0.0161 s = $500 \times 3.2281 \times 10^{-5}$ s to 0.0523 s = $800 \times 6.5399 \times 10^{-5}$ s under $N = 5, 10, 20, 40$. Even when the number of node is larger, for instance, $N = 40, 80, 100$, our algorithm can also converge by about 1000 to 1500 iterations, which takes 0.1257 s = $1000 \times 1.2568 \times 10^{-4}$ s to 0.7268 s = $1500 \times 4.8454 \times 10^{-4}$ s. Hence we conclude that the convergence efficiency of the proposed algorithm is satisfactory to adapt to the potential variation of the model parameters, and the proposed algorithm is suitable for real-time computation in reality.²

REFERENCES

- [1] Y. Cai, F. R. Yu, J. Li, Y. Zhou, and L. Lamont, "Medium access control for unmanned aerial vehicle (UAV) ad-hoc networks with full-duplex radios and multipacket reception capability," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 390–394, Jan 2013.
- [2] D. Niyato, E. Hossain, and P. Wang, "Optimal channel access management with qos support for cognitive vehicular networks," *IEEE Transactions on Mobile Computing*, vol. 10, no. 4, pp. 573–591, April 2011.

2. It is also worth pointing out that in actual one-hop ad hoc networks, e.g., cluster-based vehicular networks and unmanned aerial vehicle (UAV) networks, the network size or the cluster size is usually limited because the transmission distance and the channel spectrum are restricted. For instance, the authors consider a UAV network consisting of only 10 UAVs in [1]. In [2], the average size of a vehicle cluster is controlled below 5 in a vehicular network. From Fig. 2 above and Fig. 3 in the main text, the average convergence time of our algorithm is about 30 ms which is indeed equal to the duration of only three 5G frames when the node number is 10. At this point, our algorithm can be highly efficiently executed, considering an actual networking cluster in which only several or ten nodes coexist.