Effective Capacity Maximization of Two-Way Full-Duplex and Half-Duplex relays with Finite Block Length Packets Transmission

Mohammad Lari · **Zahra Keshavarz Gandomani**

Abstract In order to satisfy the delay requirements of telecommunication systems, in this paper, we present a cooperative network with the short packet transmission in the Rayleigh fading channel. The desired relay can be implemented as a two-way half-duplex (HD) or a two-way full-duplex (FD). Also, for more accurate satisfaction and reduction of communication delays, sending and receiving with short packets is considered. Effective capacity appropriately measures the transmission rate under the delay constraint. Therefore, it is considered as a performance evaluation criterion here. With a two-way relay, two nodes exchange data with each other using a relay simultaneously. The priorities and requirements of the two nodes are not necessarily the same. Therefore, to increase performance, the system is modeled and solved as a multi-objective problem. In this way, the available power in the network is divided between the relay and two nodes, and the effective capacity of the two nodes is maximized. Depending on the different conditions, the optimal amount of allocated power to relay and nodes is calculated. However, due to the complexity and time consuming calculations, an approximate method which speeds up the calculation is presented. The approximated solution has a very close performance to the optimal allocated power. Finally, various comparisons have been made in different conditions between the performance of two-way HD and two-way FD relays. The improvement of multi-objective power allocation has been shown, especially when the relay is not located in the middle of two nodes.

Keywords Effective Capacity · Finite Block Length Packet · Full-Duplex Relays · Half-Duplex Relays · Multi-Objective Optimization · Power Allocation · Two Way Relays

1 Introduction

Fifth generation of telecommunication systems brings a new communication era, which would transform several industry aspects and people's lives. The Internet of Things (IoT)

M. Lari
Electrical and Computer Engineering Faculty, Semnan University, Semnan, Iran
E-mail: m_lari@semnan.ac.ir

Zahra Keshavarz Gandomani
Electrical and Computer Engineering Faculty, Semnan University, Semnan, Iran
E-mail: zahra_keshavarz@semnan.ac.ir

[1, 2], self-driving and autonomous vehicles [3], online control and industrial automation [4], and virtual reality [5, 6] are instances of the fifth generation technology applications. For example, the IoT is an extensive concept, which denotes that all our surrounding objects are interconnected. This concept is applicable in all areas of energy, transportation, health, manufacturing, production, etc. to create an intelligent world with optimal use of resources [1, 2, 7]. To respond to the new requirements discussed and targeted in the fifth generation of communications, the network structure and its various parameters will undergo a fundamental transformation. Consequently, it will be possible to provide new services and meet different needs in advance.

The delay parameter and its association with delay-sensitive traffic are ignored or underestimated in previous generations [8, 9]. Human being has been pivotal in designing the previous generations of telecommunication systems (first to the fourth generation). As the human senses are less sensitive to low latencies, the delay parameter is usually not guaranteed in previous generations and tens of milliseconds delay were reasonably normal [10]. These circumstances are not tolerable for new requirements such as self-driving cars or remote surgery. In these applications and many other situations (the details of which are provided in various sources such as [7, 11, 12]), considering even a few milliseconds delay and guaranteeing its level of latency are compulsory [10]. In such cases, if the delay exceeds the threshold, the packet will be unusable and is usually discarded even if received completely and without any error. If a self-driving car transmits data to the approaching car or when overtaking a car with delay, the result will be catastrophic. Therefore, paying attention to delay at different network layers and transmission-reception techniques are focused recently. So, this paper will guarantee the statistical delay of packet by proposing a cooperative model for data transmission in the shortest possible time and appropriate allocated power.

Delay is an extensive concept created in a telecommunication systems or networks for various reasons [13, 14]. Therefore, it is necessary to pay attention to its origin to reduce delay and guarantee its maximum level. One reason for the delay is to transmit and decode relatively long packets in the receiver. Since it is possible to achieve a rate equal to the channel capacity for (infinitely) long packets in Shannon's theorem, the packets transmitted from the transmitter to the receiver have been long in recent years and conventional telecommunications. If these long packets are designed appropriately, (like using the Turbo codes or Low-Density Parity Check (LDPC) codes) data transmission rate will be very close to Shannon's rate. However, transmitting these packets is time-consuming, and their decoding delay is high. Therefore, they are not applicable in delay-sensitive traffics. On the other hand, in many applications, such as sensor networks or the Internet of Things, the data generated by the transmitters is very low and does not exceed a few bytes and it is not efficient to use long packets to transmit such data [15, 16]. Therefore, transferring the Finite Block Length (FBL) packets for data transmission and analyzing the performance of the telecommunication system by sending FBL packets are highly considered [17]. The conducted studies indicate that the communication rate differs from Shannon's rate when using the FBL packets and the error of the packets in the receiver is not zero. However, the transmission and decoding time of these packets is lower; therefore, this type of communication is being analyzed and implemented in the fifth generation of telecommunication systems [17]. Besides the transmission and decoding delays, the transmitter buffer may cause delay. In addition, the limited-time and frequency intervals in the telecommunication system could also cause a delay in packet transmission. Several other factors are also available in telecommunication networks, which might cause delay but they are not discussed here (for further information see [13]).

To provide high-quality service to delay-sensitive traffics, in this paper, a cooperative system including two nodes and a half-duplex (HD) two-way, or a full-duplex (FD) two-way relay between two nodes are considered. For ease of understanding of the content, the half-duplex and full-duplex performance is abbreviated as HD and FD, respectively. Using a simple relaying system, four time slots or frequencies are required to transfer data between two nodes. However, if a two-way relay is applied, the number of time or frequency slots is reduced. Two-way relaying can be applied for direct connection between devices without any interface (i.e. device-to-device connection or D2D) in a telecommunication network [18]. In a two-way relay with HD function, the relay receives data from both nodes within a time slot (or frequency) and retransmits the received data to both nodes in the next interval. Therefore, the number of required time intervals (or frequencies) is equal to 2. Similarly, two-way relay with FD function receives data from nodes and transmits data back to them in the same interval. Therefore, only 1 time (or frequency) interval is required. In this case, the nodes should also act as FD. Thus, the relays and nodes face the self-interference problem i.e. power transmission leakage to the self-receiver that should be managed and reduced properly. Accordingly, besides increasing the spectral gain, the use of two-way relays reduces delay [19, 20].

The transmitter buffer is another bottleneck that causes delay. If the packet input and output rate to the buffer is not adjusted correctly, the suppression of data in the transmitter buffer will increase and this will cause a delay. Since the packet output rate from the transmitter buffer is proportional to the random capacity and varies with the wireless channel, the proper adjustment of the packet input rate to the buffer is very important. If the input rate is too much, packet latency and delay in the buffer will be high. On the other hand, if the input rate is too low, it will not be possible to provide high-speed services. This input rate, which is adjusted to the requested service quality and guarantees the statistical delay of packets latency in the buffer, is called the effective capacity. To maximize the effective capacity of two nodes in this paper, the optimal power is calculated for allocation to relay and nodes. Since the two nodes are different and their requested rate and quality of service and constraints are not necessarily identical, the effective capacity of both nodes must be maximized. Therefore, a multi-objective optimization problem arises. This problem will be discussed below and resolved appropriately. Then, the effective capacity of the system will be discussed under different circumstances based on system parameters like the residual self-interference, the received small packet error, and service quality parameter. Compared to similar studies, the innovations of this research paper are as follows.

• Optimizing the effective capacity of the FD relay and two-way relay is the main innovation of this article. Using the FD and two-way relays can reduce the data transmission delay between nodes and they are suitable to transmit the delay-sensitive data.

• Another innovation of this paper is to provide a multi-objective model for the optimization problem and the detailed study and provision of the relationships associated with it. The nature of the problem is multi-objective. However, most previous articles and reviews have not optimized these systems on a multi-objective basis. Therefore, considering the problem as a multi-objective phenomenon provides interesting and different points.

• Here, it is proved that the optimization problem's solution is time-consuming and has high computational complexity. Therefore, an approximate but simpler solution, which is very close to the optimal performance, is proposed that could contribute to solving various situations quickly in a shorter time.

The rest of this paper is structured as follows. Section 2 provides an overview of the previous research related to the subject of the article. Section 3 describes the telecommunication of FBL packets, the two-way HD, and FD relays, effective capacity, and channel

model. Section 4 examines and solves the multi-objective optimization problem to maximize the effective capacity of two nodes. Section 5 provides the results of numerous simulations and comparisons and finally the paper is concluded in Section 6.

2 Review of the previous studies

Ultra-Reliable Low Latency Communications (URLLC) is one of the scenarios considered in the fifth generation of telecommunication systems. The amount of information transferred between transmitter and receiver is low in this type of communication. However, even this low information must be received with low delay and error to be applicable [13, 21]. Article [16] was the beginning of the research on this type of communication and references [22–24] were the continuation of this research trend. Since the quality of service in this type of communication is widely different from conventional communication (in conventional communication, the delay is not important, but transmission rate matters; however in URLLC, the transmission rate is not so important, but low delay and insignificant error matter), references [22–24] have addressed resource management to provide appropriate services to users in both categories. For example, [22] has considered a cellular network including different users with the various requested quality of services. One group is the conventional network users who require high rates and do not care about delay. The other category is the users who care about low delay and error rather than the transmission rate. The base station has to distribute time and frequency intervals and its power among users to maximize the capacity. In other cases, such as [25], increasing the throughput rate of URLLC users and reducing their error under delay constraint are considered. For further reviews and other issues see [11, 13, 18].

In recent years, as the telecommunication systems and user requirements became complex, the multi-objective optimization of telecommunication systems is focused [26, 27]. In this optimization method, several objective functions are optimized simultaneously rather than a single one. Thus, performance improvement compared to one-goal optimization is expected [28]. Therefore, in [22, 25] the problem is proposed and solved as multi-objective due to the presence of different objectives in URLLC.

FD communication and relays are emphasized due to their potential performance improvements [29]. Besides increasing the spectral efficiency, the use of FD relays reduces the transmission time between source and destination. Therefore, it is appropriate for delay-sensitive applications. For example, the cellular network users in [30] need low-latency service. FD relay is devised to provide the corresponding services to cell edge users. Self-interference reduction processes are Time-consuming and increases the delay. Therefore, the FD relay decreases the self-interference with the least possible processing and the residual interference is processed at the base station. Then, if necessary, this interference is used to reduce error. The authors in [31] calculated and compared the reliability when using the FBLs in HD and FD relays. Then, they specified the threshold value for selecting the type of relay between HD and FD. Moreover, the FD state of base station performance is considered to provide the URLLC service to users [32, 33]. Like the FD relays, the two-way relays can increase the spectral efficiency and reduce the delay by declining the required time intervals when transferring data between two nodes [34].

Effective capacity is a recent definition of capacity, which considers the delay of storing data in the buffer. Effective capacity is a proper criterion to analyze the delay-sensitive users' performance and guaranteeing the quality of their requested service. One of the uses of low-latency transmission is telecommunication between vehicles. Therefore, the effec-

tive capacity is maximized by allocating power and bandwidth and considering the delay constraint in [35]. In this paper, channel variations are rapid because of the high speed of vehicles; therefore, large-scale channel state information are applied to allocate power. Another application of low-latency transmission is the industrial application and Machine Type Communications (MTC). In [36], an industrial connection with some parallel fading channels is considered. The entry of transmitted packets is scattered and only the channel state information is statistically available in the transmitter. Accordingly, the effective capacity and quality of service are investigated and an upper bound is calculated for the possibility of packet error. In [37, 38], the resource allocation to maximize the effective capacity of the Cloud Radio Access Network (C-RAN) is investigated for URLLC transmission. The authors of [39] has addressed the performance of a telecommunication system with amplify and forward relay when transmitting the FBL packets. By allocating power in this paper, the effective capacity is maximized. Similar power allocation is addressed in [40] to maximize the effective capacity of a downlink multi-user network when transmitting the FBL packets. However, there is no reliable article addressing the effective capacity of low-latency transmission in FD relays. Therefore, one of the innovations of this paper is to consider this model for the delay-sensitive data packets.

3 System Model

Here, the cooperative system including two-way HD and FD relays with transmitting the FBL packets is addressed. Effective capacity is also explained as the data entry rate to the transmitter buffer, which also guarantees the statistical delay of packets.

3.1 Short Packet Transmission

Suppose that the transmitter is going to send s bits to the receiver. These bits are first encoded in the transmitter and converted into the m symbols. Then, they are transmitted to the receiver as a packet with m symbols. Since the length of the packet is finite, it is received in the receiver with the error probability of ε and the transmission rate is written as [16]:

$$\frac{s}{m} \tag{1}$$

In other words, in m times of using the channel, s bits are transmitted from the transmitter to the receiver. Dissimilar codes have different functions. For a certain amount of error ε in the signal to noise ratio (SNR) γ, the maximum value of rate is equal to r and defined as [16, 41, 42]:

$$r = \log_2(1+\gamma) - \sqrt{\frac{\gamma(\gamma+2)}{m(\gamma+1)^2}} Q^{-1}(\varepsilon)\log_2(e) + \frac{\log_2(m)}{m} \tag{2}$$

where $Q^{-1}(.)$ depicts the inverse of Q-function [16, 41, 42]. The throughput rate r in (2) shows the maximum number of bits that are transmitted from the transmitter to the receiver with the error rate less than or equal to ε in m times of using the normalized time and bandwidth channel. Therefore, the unit of r is expressed in bits per channel use (bits/c.u.). If m tends to the infinity and ε tends to zero, r in (2) tends to the famous Shannon's capacity [16, 43]. Therefore, sometimes r is considered as the channel capacity with a limited number of transmission [16]. It should be noted that according to (2), r is a function of γ, m and ε; however, it is abbreviated as r for brevity.

3.2 Two-Way Relaying

Suppose that nodes A and B are going to transmit packets to each other and this data trans-
mission is done through the relay R . There is also no direct path between A and B . Nor-
mally, in the first time slot, node A transmits its packet to relay R . In the second slot, the
relay re-sends the packet to node B . Similarly, in the next two time slots, i.e. the third
and fourth slots, node B sends a packet to node A . Therefore, four time slots are required
to share two packets between two nodes. This transmission method is not cost-effective in
terms of spectral efficiency and transmission delay. However, in the HD and FD two-way
relays (Figures 1a and 1b), just two and one time slots are required to transmit a packet from
node A to B and vice versa, respectively.

Suppose that node A is going to send the short packet x_A with the length m_A to node B
and node B is going to send the short packet x_B with the length m_B to node A . Both nodes
send their packets to the relay R simultaneously. Node A sends its packet with power P_A
and node B sends its packet with power P_B . For the two-way relay with HD performance,
after normalizing the power, the relay re-transmits the received packets for both nodes with
power P_R in the next slot. Both nodes know their transmitted packets. Therefore, if they
have access to the channel coefficients, they could remove the effect of their own packet
from the received packet from the relay completely. In this case

$$y_A = \sqrt{P_R}\, h_A\, \beta \left(\sqrt{P_B}\, h_B\, x_B + w_R \right) + w_A \tag{3}$$

specifies the received packet in node A . In (3), h_A and h_B represent the coefficient of the
flat fading channel between node A and relay R and node B and relay R . Moreover, β is
the variable gain of the relay to normalize the power of the received packet, w_A and w_R
are the white Gaussian noise with a unique power in the node A and the relay. The received
Packet is written in node B similarly. By simplifying the relationship and assuming that the
powers of x_A and x_B are equal to one, the Signal to Noise Ratio (SNR) in nodes A and B
is calculated as follows when $H_A = |h_A|^2$ is the power gain of the channel between node
A and the relay and $H_B = |h_B|^2$ is the power gain of the channel between node B and the
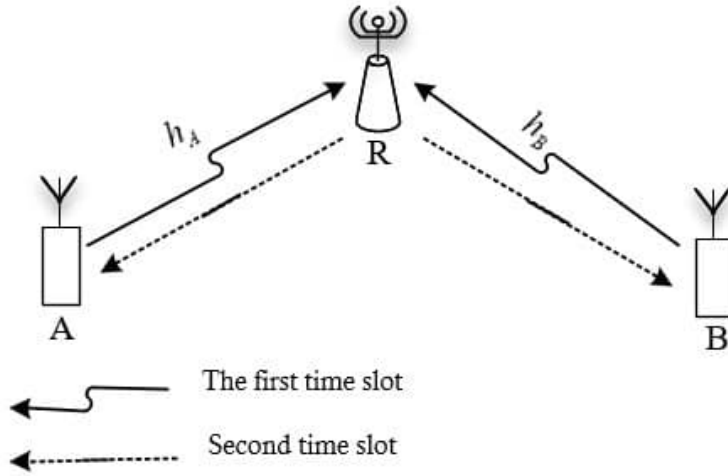relay [44].

$$\gamma_A = \frac{P_B\, P_R\, H_A\, H_B}{P_R\, H_A + P_A\, H_A + P_B\, H_B + 1} \tag{4}$$

$$\gamma_B = \frac{P_A\, P_R\, H_A\, H_B}{P_R\, H_B + P_B\, H_B + P_A\, H_A + 1} \tag{5}$$
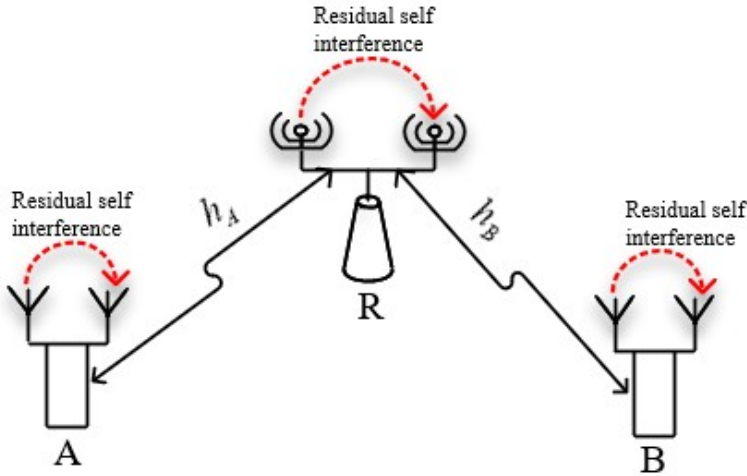
In the two-way relay with FD function, the relay normalizing the power and re-sends the
received packets to both nodes with power P_R in the same time. The two nodes receive the
transmitted packet at the same time slot too. Both nodes and relays have self-interference
in addition to noise because of simultaneous sending and receiving. Again, the nodes know
their transmitted packets and they can remove them. Therefore, the received packet in node
A is

$$y_A = \sqrt{P_R}\, h_A\, \beta \left(\sqrt{P_B}\, h_B\, x_B + \sqrt{P_R}\, \omega_R\, x_R + w_R \right) + \sqrt{P_A}\, \omega_A\, x_A + w_A \tag{6}$$

The received packet by node B is re-written in the same manner. The repeated parameters
are the same as the HD mode. x_R is the received packet by the relay after normalizing
the power or the same as the transmitted packet by the relay. However, $0 \le \omega_A \le 1$ and
$0 \le \omega_R \le 1$ indicate the coefficient of self-interference cancellation in node A and relay

(a) Half-Duplex (HD)

(b) Full-Duplex (FD)

Fig. 1 Two-way relaying.

R . In other words, $\sqrt{P_A}\,\omega_A\,x_A$ is the residual self-interference in node A and $\sqrt{P_R}\,\omega_R\,x_R$ represents the residual self-interference in relay. If $\omega_A = \omega_R = 0$, then the self-interference is removed completely. Otherwise, If $\omega_A = \omega_R = 1$, then there is no self-interference cancellation at node A and relay and the FD relay is likely to work improperly. Therefore, interference management in FD relays is very important. It is worth noting to explain why it is assumed that each node can remove its packet re-sent through the relay but it cannot remove its self-interference completely. The justification is that self-interference enters the

receiver with high power. Because of the indeterminate and nonlinear effects of different parts of the transmitter and receiver, it cannot be removed completely from the receiver. In contrast, the re-sent packet of each node to itself enters the receiver with low power. After the interference removal process, the residual part is insignificant and within the noise power limit. Therefore, it is not considered as a separate interference. By simplifying (6), the Signal to Interference plus Noise Ratio (SINR) in nodes A and B appears as [44]:

$$\gamma_A = \frac{P_B P_R H_A H_B}{P_R^2 H_A \Omega_R + P_R H_A + (P_A \Omega_A + 1)(P_A H_A + P_B H_B + P_R \Omega_R + 1)} \tag{7}$$

$$\gamma_B = \frac{P_A P_R H_A H_B}{P_R^2 H_B \Omega_R + P_R H_B + (P_B \Omega_B + 1)(P_A H_A + P_B H_B + P_R \Omega_R + 1)} \tag{8}$$

Since the residual interference cannot be estimated, (7) and (8) use the mean residual power interference coefficient as $\Omega_R = \mathbb{E}\{\omega_R^2\}$, $\Omega_A = \mathbb{E}\{\omega_A^2\}$, and $\Omega_B = \mathbb{E}\{\omega_B^2\}$. Then, the following assumptions are made to simplify the problem without losing its general concept.

$$m_A = m_B = m$$

$$P_A = P_B = P$$

$$\Omega_A = \Omega_B = \Omega_R = \Omega$$

3.3 Effective Capacity

The effective capacity is defined as the input rate of the packets to a system's buffer, which can pass through the channel and a certain quality of service could be guaranteed for it [45, 46]. Quality of service refers to the statistical delay in the transmitter buffer, which is defined as [45–47]:

$$\mathbb{P}\{D > D_0\} \approx \eta e^{-\theta D_0} \tag{9}$$

where $\mathbb{P}\{.\}$ is the probability of occurrence, D is the packet delay in the transmitting buffer, D_0 is a threshold for the delay, η is the probability that the transmitter buffer is not empty and θ denotes the quality of service exponent. As the value of θ is set to a higher level, the probability of the delay is reduced and the quality of service will be stricter. In contrast, for a lower value of θ, the probability of delay is higher and the quality of service is not too strict. After specifying the amount of requested quality by each user using the parameter θ, the effective capacity R_E is calculated as:

$$R_E = -\frac{1}{m\theta} \log \left[\mathbb{E}\{e^{-rm\theta}(1-\varepsilon) + \varepsilon\} \right]. \tag{10}$$

It should be noted that $\log(.)$ represents a natural logarithm function.

3.4 Communication Channel

The communication channel between the nodes and the relay is considered as Rayleigh flat fading. The channel coefficient between nodes A and B with relay R is equal to $h_A = h_{0,A} d_A^{-\alpha/2}$ and $h_B = h_{0,B} d_B^{-\alpha/2}$ respectively. $h_{0,A}$ and $h_{0,B}$ represent uncorrelated complex Gaussian coefficients with zero mean and unit variance. Moreover, d_A and d_B determine the distance between nodes A and B to the relay and $d_A + d_B = 1$. α denotes the path loss exponent of the channel.

Table 1 Main parameters of this paper.

Abbreviation	Description
m	Short packet length
ε	Error probability of received short packet
P	Transmitted power of node A and B
P_{R}	Transmitted power of relay R
P_{tot}	Total available power which is divided between nodes and relay
Ω	Mean residual power interference coefficient
r_i	Throughput rate of node i
H_i	Power gain of the channel between node i and relay R
γ_i	SNR of node i in HD mode or SINR of node i in FD mode
θ_i	Quality of service exponent in node i
$R_{E,i}$	Effective capacity of node i

3.5 Table of Parameters

For convenience, the main variables of the article are presented in Table 1. To summarize some of the relations and also in Table 1, the index i displays one of the nodes A or B ($i \in \{A, B\}$). Moreover, $\bar{i}$ depicts the complementary of i in the set. So, if $i = A$, then $\bar{i} = B$.

4 Maximizing the Effective Capacity of the Two-Way Relay

In this section, the appropriate power allocation between nodes A and B and relay R is calculated. The allocated power maximize the effective capacity when short packet is transmitted by node A and node B. It is assumed that the total available power i.e. P_{tot} is optimally distributed between the nodes and relay. This type of power allocation is commonly used in FD amplify and forward relay [48], in FD wireless power communication networks [49] and two-way fixed-gain amplify and forward relay [50]. Therefore, $P_{\mathrm{R}} + 2P = P_{tot}$. Here, the problem is provided and solved for the HD and FD two-way relays individually.

The condition of two nodes A and B are different and their requested rate and quality of service is not necessarily the same. Therefore, the optimization problem is considered and solved as a multi-objective problem. In the multi-objective optimization problem, there are two or more objective functions to be optimized instead of one single objective function. These functions are usually not aligned and do not optimize at the same point. Therefore, the multi-objective optimization problems have a set of optimal solutions called Pareto optimal solutions instead of a specific optimal solution [27]. In multi-objective minimization (multi-objective maximization), the optimal Pareto point is selected when there is no other point in the feasible design space in which the value of all objective functions is less than or equal to (greater than or equal to) the value of objective functions at the optimal Pareto point [27]. Finally, the expert user suggests a point as an answer according to the priority of the objective functions if it is necessary. Solving the multi-objective optimization problem has various techniques. For more details see [27].

Before designing and solving the problem, lemma 1 is addressed. This theorem will be necessary for solving the discussed problems.

Lemma 1 *If $\gamma > 1$, $m > 100$ and $\varepsilon > 10^{-23}$, the throughput rate in* (2) *is strictly increasing and concave in terms of γ.*

$$\forall \{\gamma > 1,\ m > 100,\ \varepsilon > 10^{-23}\} \implies \left\{ \frac{\partial r}{\partial \gamma} > 0,\ \frac{\partial^2 r}{\partial \gamma^2} \leq 0 \right\}$$

Proof Please refer to [39].

In most communication systems, a high SNR (about 10 and above) is desired under normal constraints. Therefore, in most cases we have $\gamma > 1$. The constraint $m > 100$ is not restrictive at all. That is because 100 bits equals 12.5 bytes, which is not a significant value in the transmission of the short packets. The error rate of $\varepsilon = 10^{-23}$ is insignificant and it is never required. Thus, the constraint of lemma 1 is not strict, and lemma 1 is always approved.

4.1 Multi-Objective Optimization Problem in the Two-Way HD Relay

The multi-objective optimization problem in a two-way HD relay with short packet transmission to maximize the effective capacity of two nodes is as follows:

$$\begin{cases} \max_{P_R} R_{E,A} \\ \max_{P_R} R_{E,B} \\ \text{s.t.} P_R < P_{tot} \end{cases} \tag{11}$$

In the optimization problem (11), the effective capacity of the nodes is written according to (10) as follows:

$$R_{E,i} = -\frac{1}{m\theta_i} \log \left[\mathbb{E}\{ e^{-r_i \frac{m}{2}\theta_i}(1 - \varepsilon_i) + \varepsilon_i \} \right], \quad i \in \{A, B\}. \tag{12}$$

We use $\frac{m}{2}$ in (12) because in the HD relay, the number of using channel m is twice the length of the transmitted packet. Now, it is proved that the optimization problem (11) is concave. Accordingly, first lemma 2 and 3 are proved, and finally by lemma 4, the concavity of the problem (11) is presented.

Lemma 2 *The SNR is concave in terms of the relay power P_R and only has a single maximum point. That is:*

$$\frac{\partial^2 \gamma_i}{\partial P_R^2} < 0, \quad i \in \{A, B\}$$

Proof For ease of writing, the lemma is proved for node A. The relations can also be simply written and proved for node B. Using (4), the second-order derivative of γ_A relative to P_R is:

$$\frac{\partial^2 \gamma_A}{\partial P_R^2} = -\frac{4H_A H_B (H_A P_{tot} + 1)(H_A P_{tot} + H_B P_{tot} + 2)}{(H_A (P_{tot} + P_R) + H_B (P_{tot} - P_R) + 2)^3} \tag{13}$$

Since $P_R < P_{tot}$, the denominator of (13) is always positive. Thus, the second derivative in (13) is constantly negative and γ_A is concave relative to P_R. Now, the first derivative is as follows:

$$\frac{\partial \gamma_A}{\partial P_R} = \frac{H_A H_B (P_{tot} - 2P_R)}{H_A (P_{tot} + P_R) + H_B (P_{tot} - P_R) + 2} - \frac{H_A H_B P_R (H_A - H_B)(P_{tot} - P_R)}{(H_A (P_{tot} + P_R) + H_B (P_{tot} - P_R) + 2)^2} \tag{14}$$

If (14) is set equal to zero, a quadratic equation is obtained with two roots. These two roots must be real. By calculating the roots and the second-order derivative in both roots, one of the roots will be acceptable because the second-order derivative becomes positive in the other root, which is contrary to the second-order derivative under relation (13). $P_R^{\star A}$ in (15) is the optimal power of the relay, which makes the first-order derivative zero and maximizes γ_A.

$$P_R^{\star A} = -\frac{H_A\,P_{tot} + H_B\,P_{tot} + 2 - \sqrt{2(H_A\,P_{tot} + 1)(H_A\,P_{tot} + H_B\,P_{tot} + 2)}}{H_A - H_B} \tag{15}$$

Lemma 3 $R_{E,A}$ *relative to* r_A *and* $R_{E,B}$ *relative to* r_B *is strictly increasing and concave. This means that:*

$$\frac{\partial R_{E,i}}{\partial r_i} > 0, \quad \frac{\partial^2 R_{E,i}}{\partial r_i^2} < 0, \quad i \in \{A, B\}.$$

Proof The first and second-order derivatives are calculated according to (16) and (17). In (16), since $\varepsilon_i \leq 0.5$ and the exponential function is always positive, the numerator and denominator of the relation are positive. Therefore, the first-order derivative is positive and the effective capacity is strictly increasing relative to the throughput rate. Therefore, the second-order derivative is also negative and the effective capacity is concave relative to the throughput rate.

$$\frac{\partial R_{E,i}}{\partial r_i} = \frac{1}{2}\frac{\mathbb{E}\left\{e^{-r_i\frac{m}{2}\theta_i}\left(1 - \varepsilon_i\right)\right\}}{\mathbb{E}\left\{e^{-r_i\frac{m}{2}\theta_i}\left(1 - \varepsilon_i\right) + \varepsilon_i\right\}}, i \in \{A, B\} \tag{16}$$

$$
\begin{aligned}
\frac{\partial^2 R_{E,i}}{\partial r_i^2} &= \\
\frac{1}{2}&\frac{\frac{-\theta_i m}{2}\mathbb{E}\left\{e^{-r_i\frac{m}{2}\theta_i}(1-\varepsilon_i)\right\}\left(\mathbb{E}\left\{e^{-r_i\frac{m}{2}\theta_i}(1-\varepsilon_i)\right\}+\varepsilon_i\right)}{\left(\mathbb{E}\left\{e^{-r_i\frac{m}{2}\theta_i}(1-\varepsilon_i)+\varepsilon_i\right\}\right)^2} + \frac{1}{2}\frac{\frac{\theta_i m}{2}\left(\mathbb{E}\left\{e^{-r_i\frac{m}{2}\theta_i}(1-\varepsilon_i)\right\}\right)^2}{\left(\mathbb{E}\left\{e^{-r_i\frac{m}{2}\theta_i}(1-\varepsilon_i)+\varepsilon_i\right\}\right)^2} = \\
\frac{-\theta_i m \varepsilon_i}{4}&\frac{\mathbb{E}\left\{e^{-r_i\frac{m}{2}\theta_i}(1-\varepsilon_i)\right\}}{\left(\mathbb{E}\left\{e^{-r_i\frac{m}{2}\theta_i}(1-\varepsilon_i)+\varepsilon_i\right\}\right)^2}, i \in \{A, B\}
\end{aligned}
\tag{17}
$$

Lemma 4 *If lemma 1 holds, the effective capacity is concave relative to the relay power* P_R *and only has one maximum point. This means that:*

$$\frac{\partial^2 R_{E,i}}{\partial P_R^2} < 0, \quad i \in \{A, B\}$$

.

Proof This lemma is again proved for node A. Proof for node B is done in the same manner. According to the chain rule, the second-order derivative of $R_{E,A}$ relative to P_R equals

$$\frac{\partial^2 R_{E,A}}{\partial P_R^2} = \frac{\partial^2 R_{E,A}}{\partial r_A^2}\left(\frac{\partial r_A}{\partial P_R}\right)^2 + \frac{\partial^2 r_A}{\partial P_R^2} + \frac{\partial R_{E,A}}{\partial r_A}. \tag{18}$$

On the right side of (18), according to lemma 3, the first fraction is negative. The second fraction is always positive. The fourth fraction is also positive according to lemma 3. If it is proved that the third fraction is constantly negative, the negation of (18) will be proved. Therefore, using the chain rule, the third fraction is rewritten as:

$$\frac{\partial^2 r_A}{\partial P_R^2} = \frac{\partial^2 r_A}{\partial \gamma_A^2}\left(\frac{\partial \gamma_A}{\partial P_R}\right)^2 + \frac{\partial^2 \gamma_A}{\partial P_R^2} + \frac{\partial r_A}{\partial \gamma_A}. \tag{19}$$

On the right side of (19), according to lemma 1, the first fraction is negative and the fourth fraction is positive. According to lemma 2, the second fraction is also positive and the third fraction is negative. Therefore, (19) is negative and it is possible to claim that the effective capacity of node A is concave relative to the power of the relay. Now, the first-order derivative is written as follows:

$$\frac{\partial R_{E,A}}{\partial P_R} = \frac{\partial R_{E,A}}{\partial r_A}\frac{\partial r_A}{\partial \gamma_A}\frac{\partial \gamma_A}{\partial P_R}. \tag{20}$$

On the right side of (20), according to lemma 3, the first fraction is non-zero. According to lemma 2, the second fraction is not zero. Based on lemma 2, the third fraction becomes zero only at the point $P_R^{\star A}$. Therefore, it can be concluded that the effective capacity of node A is concave relative to P_R and reaches its maximum value only at the point $P_R^{\star A}$.

Since the functions of the multi-objective problem (11) are concave, the weighted sum is a proper solution for them. The weighted sum method is very simple and it is the necessary and sufficient condition to extract all the optimal Pareto points in the case of multi-objective concave (or convex) problems [27]. In the weighted sum method, the first objective function of the optimization problem (11) with the coefficient w and the second objective function with the coefficient $1-w$ are summed up and they form a new objective function. The coefficient $0 \le w \le 1$ presents the priority of the first objective function and the coefficient $0 \le 1-w \le 1$ indicates the priority value of the second objective function. Accordingly, the optimization problem (11) is simplified to

$$\begin{cases} \min_{P_R} J \\ \text{s.t.} P_R < P_{tot} \\ J = \frac{w}{m\theta_A}\log\left[\mathbb{E}\left\{e^{-r_A\frac{m}{2}\theta_A}(1-\varepsilon_A)+\varepsilon_A\right\}\right] \\ \quad + \frac{1-w}{m\theta_B}\log\left[\mathbb{E}\left\{e^{-r_B\frac{m}{2}\theta_B}(1-\varepsilon_B)+\varepsilon_B\right\}\right]. \end{cases} \tag{21}$$

It should be noted that the first and second objective functions of the problem (11) are summed up with the negative sign. Thus, maximization of the problem (11) is minimized in problem (21).

Due to the statistical expectation in the definition of effective capacity, the numerical solution of problem (21) is time-consuming. Therefore, the problem is simplified to achieve a faster solution. To simplify J in (21), due to the small probability of error, ε_A and ε_B are ignored and J is simplified to

$$J \approx \frac{w}{m\theta_A}\log\left[\mathbb{E}\left\{e^{-r_A\frac{m}{2}\theta_A}(1-\varepsilon_A)\right\}\right] + \frac{1-w}{m\theta_B}\log\left[\mathbb{E}\left\{e^{-r_B\frac{m}{2}\theta_B}(1-\varepsilon_B)\right\}\right]. \tag{22}$$

Then, using the approximation of logarithm of the sum of the exponential functions according to (23) [51],

$$\log\mathbb{E}\{e^z\} \approx \max\{z\} + C \tag{23}$$

J is again simplified to

$$J \approx \frac{w}{m\theta_A}\left[\max\left\{-r_A\frac{m}{2}\theta_A + C + \log(1-\varepsilon_A)\right\}\right] + \frac{1-w}{m\theta_B}\left[\max\left\{-r_B\frac{m}{2}\theta_B + C + \log(1-\varepsilon_B)\right\}\right]$$
$$= \left[\max\left\{-\frac{w}{2}r_A\right\} + \frac{w}{m\theta_A}C + \frac{w}{m\theta_A}\log(1-\varepsilon_A)\right] + \left[\max\left\{-\frac{1-w}{2}r_B\right\} + \frac{1-w}{m\theta_B}C + \frac{1-w}{m\theta_B}\log(1-\varepsilon_B)\right] \tag{24}$$

In (23), $z \in \mathbb{Z}$ denotes an arbitrary random variable and $\max z$ specifies the largest value of a random variable in the set $\mathbb{Z}$. In addition, C is a constant value and does not depend

on a random variable . This constant value also appears in (24). Now, J is inserted in the optimization problem (21) and the simplified problem is written as:

$$\begin{cases} \min_{P_{\mathrm{R}}} \quad \{\max\{-\frac{w}{2}r_{\mathrm{A}}\} + \max\{-\frac{1-w}{2}r_{\mathrm{B}}\} + \prime C\} \\ \mathrm{s.t.} P_{\mathrm{R}} < P_{tot} \end{cases} \tag{25}$$

and $\prime C$ indicates fixed parameters that do not affect optimization.

The optimization problem (25) is a min-max optimization problem [27]. Therefore, the solution to this type of problem is used here as well. First, two auxiliary variables τ_{A} and τ_{B} are defined and the problem (25) is converted to

$$\begin{cases} \min_{P_{\mathrm{R}}} \tau_{\mathrm{A}} + \tau_{\mathrm{B}} \\ \mathrm{s.t.} \tau_{\mathrm{A}} \geq -\frac{w}{2}r_{\mathrm{A}} \\ \mathrm{s.t.} \tau_{\mathrm{B}} \geq -\frac{1-w}{2}r_{\mathrm{A}} \\ \mathrm{s.t.} P_{\mathrm{R}} < P_{tot} \end{cases} \tag{26}$$

For further simplification, using the auxiliary variable $\tau = \tau_{\mathrm{A}} + \tau_{\mathrm{B}}$, the optimization problem is rewritten again as:

$$\begin{cases} \min_{P_{\mathrm{R}}} \tau \\ \mathrm{s.t.} \tau \geq -\frac{w}{2}r_{\mathrm{A}} - \frac{1-w}{2}r_{\mathrm{B}} \\ \mathrm{s.t.} P_{\mathrm{R}} < P_{tot} \end{cases} \tag{27}$$

The optimization problem (27) has no statistical expectation. Therefore, is much easier to solve than the optimization problem (21). To speed up convergence in the numerical solution, the following relationship can be used as a starting point.

$$P_{\mathrm{R}} = wP_{\mathrm{R}}^{\star\mathrm{A}} + (1-w)P_{\mathrm{R}}^{\star\mathrm{B}}$$

The answer to the problem (27) is called $P_{\mathrm{R}}^{\star}$. The SNR of nodes A and B in this optimal point are $\gamma_{\mathrm{A}}^{\star}$ and $\gamma_{\mathrm{B}}^{\star}$.

To establish the assumptions of lemma 1, $\gamma_{\mathrm{A}}^{\star} > 1$ and $\gamma_{\mathrm{B}}^{\star} > 1$ are needed. Generally, we assume $\gamma_{\mathrm{A}}^{\star} > \gamma_{T,\mathrm{A}}$ and $\gamma_{\mathrm{B}}^{\star} > \gamma_{T,\mathrm{B}}$. This comparison may support the assumption of lemma 1, or satisfy the minimum throughput rate of node A and node B . Now, if the constraint $\gamma_{\mathrm{A}}^{\star} > \gamma_{T,\mathrm{A}}$ is not met in node A , node A is not transmitted and the relay uses its power $P_{\mathrm{R}}^{\star\mathrm{B}}$ to transmit the data of node B . Similarly, if the constraint $\gamma_{\mathrm{B}}^{\star} > \gamma_{T,\mathrm{B}}$ is not met, node B is not transmitted and the power $P_{\mathrm{R}}^{\star\mathrm{A}}$ is allocated to the relay.

Lemma 2 shows that the SNR is concave relative to the relay power and only has one maximum point. Therefore, the equation $\gamma_{\mathrm{A}} = \gamma_{T,\mathrm{A}}$ has exactly two roots $P_{T,\mathrm{A}}^1$ and $P_{T,\mathrm{A}}^2$; and the equation $\gamma_{\mathrm{B}} = \gamma_{T,\mathrm{B}}$ must have two roots equal to $P_{T,\mathrm{B}}^1$ and $P_{T,\mathrm{B}}^2$. Therefore, if $P_{\mathrm{R}}^{\star} < P_{T,\mathrm{A}}^1$ or $P_{\mathrm{R}}^{\star} > P_{T,\mathrm{A}}^2$, then node A is not transmitted. Similarly, if $P_{\mathrm{R}}^{\star} < P_{T,\mathrm{B}}^1$ or $P_{\mathrm{R}}^{\star} > P_{T,\mathrm{B}}^2$, then node B will have no transmission. In (28), $P_{T,i}^1$ and $P_{T,i}^2$ are calculated for $i \in \{\mathrm{A},\mathrm{B}\}$. Moreover, the value σ_1 is given in (29).

$$\begin{aligned} P_{T,i}^1 &= \frac{(H_{\bar{i}}-H_i)\gamma_{T,i}+H_{\mathrm{A}}H_{\mathrm{B}}\gamma_{T,i}-\sigma_1}{2H_{\mathrm{A}}H_{\mathrm{B}}} \\ P_{T,i}^2 &= \frac{(H_{\bar{i}}-H_i)\gamma_{T,i}+H_{\mathrm{A}}H_{\mathrm{B}}\gamma_{T,i}+\sigma_1}{2H_{\mathrm{A}}H_{\mathrm{B}}} \end{aligned} \tag{28}$$

$$\begin{aligned} \sigma_1 = \big[&H_{\mathrm{A}}^2 H_{\mathrm{B}}^2 P_{tot}^2 - 6H_i^2 H_{\bar{i}} P_{tot}\gamma_{T,i} - 2H_i H_{\bar{i}}^2 P_{tot}\gamma_{T,i} \\ &+ (H_{\mathrm{A}}\gamma_{T,i})^2 + (H_{\mathrm{B}}\gamma_{T,i})^2 \\ &- 2H_{\mathrm{A}}H_{\mathrm{B}}\gamma_{T,i}^2 - 8H_{\mathrm{A}}H_{\mathrm{B}}\gamma_{T,i} \big] \end{aligned} \tag{29}$$

4.2 Multi-Objective Optimization Problem in the Two-Way FD Relay

To maximize the effective capacity of nodes in a cooperative system with a two-way FD relay and short packet transmission, the following optimization problem is used.

$$
\begin{cases}
\max\limits_{P_R} R_{E,A} \\
\max\limits_{P_R} R_{E,B} \\
\text{s.t.} P_R < P_{tot}
\end{cases}
\tag{30}
$$

In this problem, the effective capacity of nodes A and B can be written as:

$$
R_{E,i} = -\frac{1}{m\theta_i} \log\left[\left\{ e^{-r_i m \theta_i} (1 - \varepsilon_i) + \varepsilon_i \right\}\right], i \in \{A, B\}.
\tag{31}
$$

In FD relay, the SINR is not concave relative to the relay power. This means that the second-order derivative of γ_i, $i \in \{A, B\}$ is not always negative relative to P_R. Moreover, the effective capacity is not concave relative to the relay power and the second-order derivative of $R_{E,i}$, $i \in \{A, B\}$ is not constantly negative relative to P_R. Therefore, lemma 2 and 4 are not generally true for the FD relay. However, it is possible to prove that the SINR and effective capacity only have one maximum point relative to P_R. Therefore, lemma 5 is presented as follows.

Lemma 5 *The SINR and the effective capacity have only one maximum point relative to the relay power.*

Proof For simplicity, the proof is provided for node A. Again, the relations are simply written and proved for node B. γ_A for FD relay is written in (7). Since the value of γ_A is constantly positive and becomes zero for $P_R = 0$ and $P_R = P_{tot}$, it has a definite maximum point. Then, the quadratic equation is obtained by calculating the first-order derivative relative to P_R and the two values of P_R are calculated as follows:

$$
\begin{aligned}
P_R^1, P_R^2 =& \\
& -\frac{(H_A P_{tot} + H_B P_{tot} + 2)(\Omega P_{tot} + 2)}{(H_A - H_B)(\Omega P_{tot} + 2) + 2\Omega(H_A P_{tot} + 1)} \\
& \pm \frac{2\sqrt{(H_A P_{tot} + 1)(\Omega P_{tot} + 1)(\Omega P_{tot} + 2)(H_A P_{tot} + H_B P_{tot} + 2)}}{(H_A - H_B)(\Omega P_{tot} + 2) + 2\Omega(H_A P_{tot} + 1)}
\end{aligned}
\tag{32}
$$

It should be noted that the second fraction sign is positive for P_R^1 and negative for P_R^2. First, it is assumed that P_R^1 is the answer and $P_R^1 \geq 0$. After multiplying both sides of the equation by the fraction's denominators and achieving an algebraic expression, its sign is determined. After simplification, it is proved that $H_A \leq 0$ which is not true. Therefore, there is only one maximum point as:

$$
\begin{aligned}
P_R^{\star A} =& \\
& -\frac{(H_A P_{tot} + H_B P_{tot} + 2)(\Omega P_{tot} + 2)}{(H_A - H_B)(\Omega P_{tot} + 2) + 2\Omega(H_A P_{tot} + 1)} \\
& -\frac{2\sqrt{(H_A P_{tot} + 1)(\Omega P_{tot} + 1)(\Omega P_{tot} + 2)(H_A P_{tot} + H_B P_{tot} + 2)}}{(H_A - H_B)(\Omega P_{tot} + 2) + 2\Omega(H_A P_{tot} + 1)}.
\end{aligned}
\tag{33}
$$

Then, similar to the relations of lemma 4 and considering the assumptions of lemma 1, it is proved that $R_{E,A}$ is maximized at the point $P_R^{\star A}$.

Although the effective capacity is not concave relative to the relay power, lemma 5 shows that the effective capacity only has one maximum point. Thus, during the numerical solution of the problem, if the solution algorithm converges, the obtained solution will be the global maximum answer. On the other hand, when the self-interference is eliminated properly and $\Omega \to 0$, lemmas 2 and 4 are established and the effective capacity is concave. Therefore, the optimal answer for nodes A and B will be close to $P_R^{\star A}$ and $P_R^{\star B}$ even if Ω is small but not necessarily close to zero. Therefore, the following point is applied as a starting point to increase the computing speed in multi-objective optimization.

$$P_R = wP_R^{\star A} + (1-w)P_R^{\star B}$$

For non-concave optimization (30), the weighted sum method is a sufficient condition to achieve the optimal Pareto point. This means that if this method is applied, the answers will be the optimal Pareto points. This method is not a necessary condition to calculate the optimum points. However, since the weighted sum method is simple, it is applied in FD problem too. If other methods that are the necessary and sufficient condition to achieve the optimal Pareto point are used, the results of the HD problem will no longer comparable with FD. Since the results obtained in HD and FD mode are going to be compared, the weighted sum method will be applied to solve the problem (30) (articles such as [22] have used this method to solve the non-convex problem). Therefore, by combining the weighted sum of two objective functions, the multi-objective optimization problem is simplified to

$$\begin{cases} \min_{P_R} J \\ \text{s.t.} P_R < P_{tot} \\ J = \frac{w}{m\theta_A} \log \left[\mathbb{E}\left\{ e^{-r_A\, m\theta_A} \left(1 - \varepsilon_A\right) + \varepsilon_A \right\} \right] \\ + \frac{1-w}{m\theta_B} \log \left[\mathbb{E}\left\{ e^{-r_B\, m\theta_B} \left(1 - \varepsilon_B\right) + \varepsilon_B \right\} \right]. \end{cases} \tag{34}$$

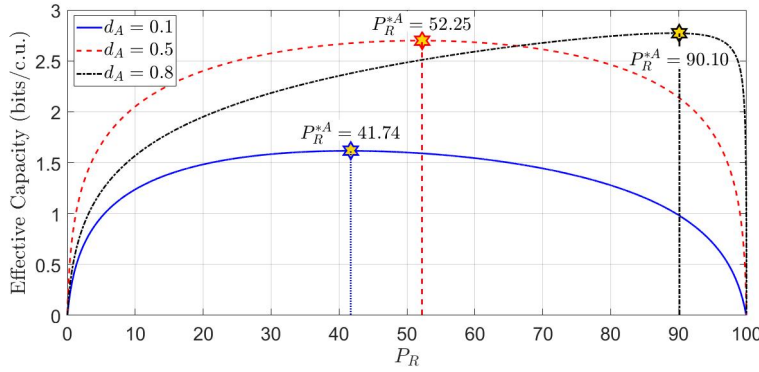Similar to equations (22) to (26), problem (34) is also simplified as a min-max problem (35).

$$\begin{cases} \min_{P_R} \tau \\ \text{s.t.} \tau \geq -\frac{w}{2}r_A - \frac{1-w}{2}r_B \\ \text{s.t.} P_R < P_{tot} \end{cases} \tag{35}$$

Once again w and $1-w$ indicate the importance of the effective capacity of nodes A and B , respectively, and r_i, $i \in \{A, B\}$ is the throughput rate of node i in the FD relay.

5 Simulation, Numerical Results, and Comparison

This section provides simulations, numerical results, and a comparison of performance in different modes. Here the main parameters include $m = 100$, $P_{tot} = 1000\ W$, $\alpha = 4$, $\varepsilon_A = \varepsilon_B = 10^{-4}$, $\theta_A = \theta_B = 10^{-3}$, $d_A = 0.5$, $d_A + d_B = 1$ and $\gamma_{R,A} = \gamma_{T,B} = 1$. If any of these parameters is altered, the new value will be mentioned.

Figure 2 shows the effective capacity of node A in terms of HD relay power for three values of $d_A = 0.1$, $d_A = 0.5$ and $d_A = 0.8$. This figure specifies $P_R^{\star A}$ for which $R_{E,A}$ is maximized. According to lemma 4, the effective capacity of each node is only maximized for a single value of relay power. Meanwhile, if d_A increases, the relay will approach node B and the packet of node B will reach the relay even with the small amount of power. On the other hand, since the relay is away from node A , it has to spend more power to transmit packets to node A . Therefore, as the relay R takes distance from node A and approaches

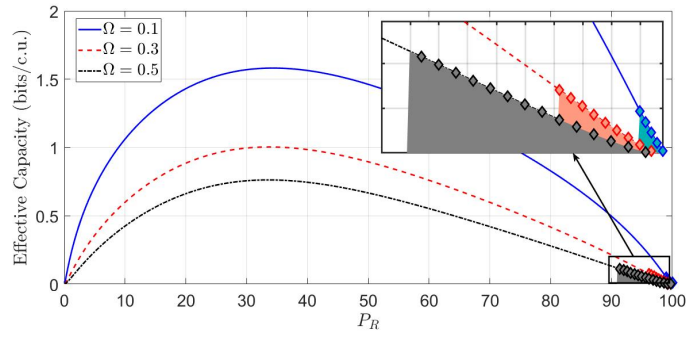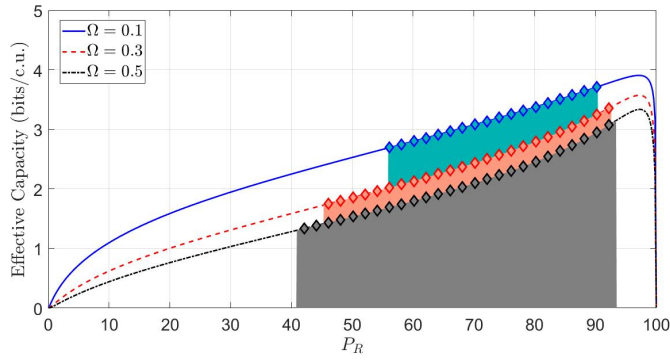**Fig. 2** Effective capacity of node A vs. HD relay power and different node-relay distance.

node B , more power should be allocated to the relay. This issue is also illustrated in Figure 2.

Lemma 5 showed that the effective capacity in the FD relay is not necessarily concave relative to the relay power. However, the effective capacity always maximizes at one point. Moreover, by decreasing the mean residual power interference coefficient Ω, the FD SINR tends to the SNR of the HD relay. Figure 3 indicates the effective capacity of node A in terms of FD relay power for three values of $\Omega = 0.1$, $\Omega = 0.3$ and $\Omega = 0.5$. In Figure 3a $d_A = 0.1$ and in Figure 3b we have $d_A = 0.9$. The effective capacity only has one maximum point in all three cases. However, the effective capacity is not constantly concave and parts of the curve in which the second-order derivative is positive are highlighted in Figure 3. Moreover, these parts are reduced by decreasing Ω. Since the power interference coefficient is less than 0.1 in real constraints, the effective capacity in FD relay and under practical constraints $\Omega \ll 1$ can be considered concave.
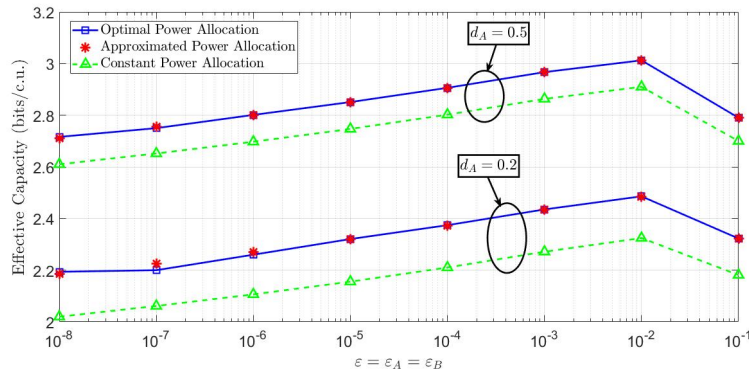
Figure 4 indicates the weighted sum of the effective capacities for two nodes A and B with an identical weight of $w = 0.5$ in terms of error probability in HD relay. The effective capacity with approximate power allocation (optimization problem (27)) matches with effective capacity with optimal power allocation (optimization problem (11)) properly. Since solving the simplified optimization problem (27) has higher execution speed, we use the approximated solution for further comparison. The execution speeds are compared in Table 2. Figure 4 presents the weighted sum of the effective capacities of two nodes with equal power allocation $P_{tot}/3$ between nodes A and B and relay R . Accordingly, effective capacity with optimal power allocation provides better performance than equal power allocation between nodes and relays. This superior performance is most noticeable under the constraints that the relay is not placed in the middle of the nodes.

Figure 4 illustrates the weighted sum of the effective capacity of two nodes with an identical weight of $w = 0.5$ for $d_A = 0.2$ and $d_A = 0.5$. The identical weight of $w = 0.5$ means that nodes A and B have equal importance and the multi-objective optimization problem attempts to maximize the effective capacity of both nodes with the same priority. When the relay is placed between two nodes, the weighted sum of the effective capacity is more than the case that the relay is distant from node B. If the relay is distant from node B , the attenuation of the relay-node B channel is more than that of the node A . Accordingly, part of the available power is consumed to compensate for this additional attenuation so that two
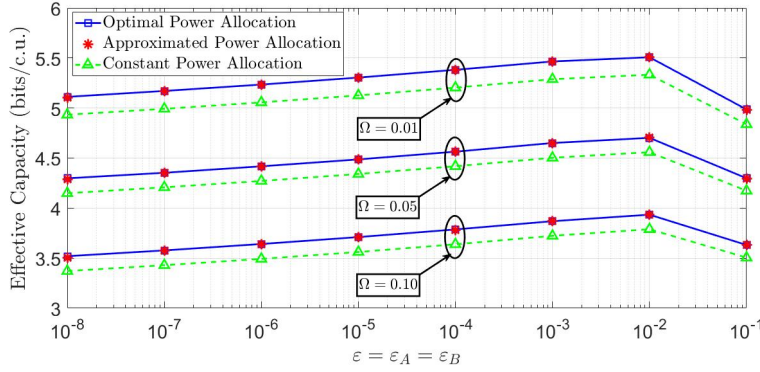
(a) $d_A = 0.1$

(b) $d_A = 0.9$

Fig. 3 Effective capacity of node A vs. FD relay power and different mean residual power interference coefficient (a) $d_A = 0.1$, (b) $d_A = 0.9$.

Fig. 4 Weighted sum of the effective capacity vs. error probability in HD relay with optimal and approximated power allocation and constant power allocation.
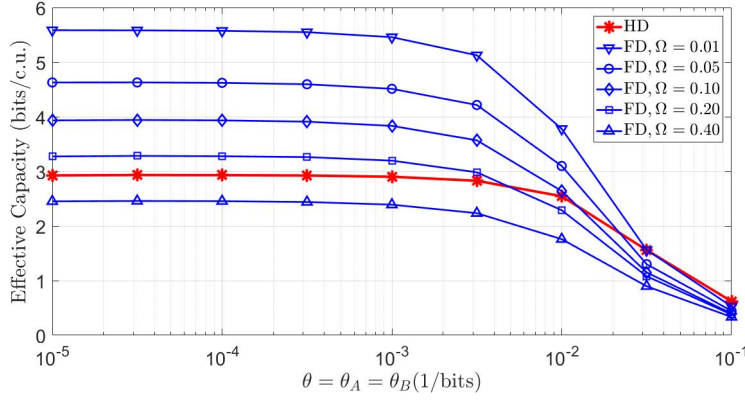
Fig. 5 Weighted sum of the effective capacity vs. error probability in FD relay with optimal and approximated power allocation and constant power allocation.

nodes with the same priority would become similar in terms of channel attenuation. Then, the residual power is divided to maximize the weighted sum of the effective capacities of nodes A and B . Therefore, the performance of the system with the same priority between nodes is greater for the case that the relay is located in the middle of two nodes. Moreover, the effective capacity is maximized at a certain value of the error probability. Therefore, the values of m, ε and θ should be proportional to each other. Similar to the simulations in Figure 4, other simulations are conducted for the weighted sum of the effective capacity in terms of different lengths of packet m that have similar results. To avoid redundancy, these figures are not provided here.
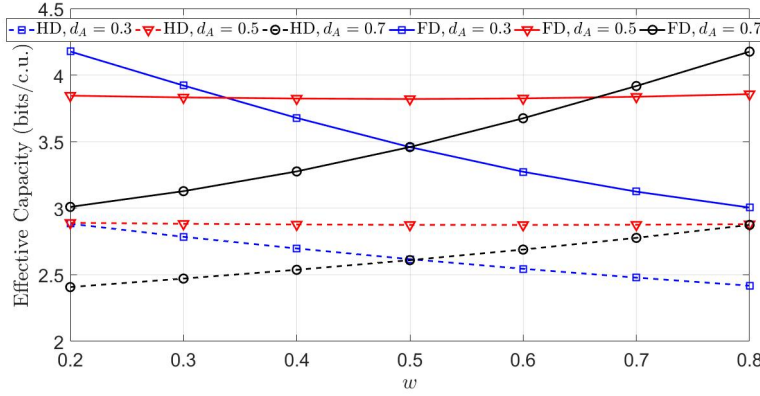
Figure 5 presents the weighted sum of the effective capacity with the identical weight $w = 0.5$ in terms of the error probability $\varepsilon = \varepsilon_A = \varepsilon_B$ in the FD relay between nodes A and B . This simulation is conducted for three diferent conditions with almost no self-interference and $\Omega = 0.01$, moderate self-interference and $\Omega = 0.05$, and high self-interference and $\Omega = 0.1$. The weighted sum of effective capacity is matched between the optimal solution of the problem (34) and the approximate solution of the problem (35), which indicates the high accuracy of the simplified optimization problem. Solving the simplified optimization problem has a much higher execution speed. Therefore, the approximate power allocation will be applied for further comparisons. The execution time to solve the optimal and simplified problem is compared in Table 2.

Figure 5 also shows the weighted sum of the effective capacity of the two nodes with equal power allocation $P_{tot}/3$. Effective capacity with optimal power allocation enjoys much better performance than the identical power allocation between the nodes and relays. Besides, according to our expect, by reducing the self-interference power coefficient Ω, the effective capacity will increase.

Fiure 6 provides a comparison between the weighted sum of the effective capacity in the HD and FD relays. The FD relay has self-interference, which affects the overall performance of the system. Accordingly, at low self-interference values (e.g. $\Omega = 0.01$), the effective capacity of an FD relay is almost twice greater than that of the HD relay. The reason for this phenomenon is that the FD relay uses only one-time or frequency interval to receive packet and transmit them to the nodes. However, the HD relay uses two-time or frequency

Fig. 6 Comparison of weighted sum of the effective capacity in HD relay and FD relay with different mean residual power interference coefficient Ω.

Fig. 7 Comparison of weighted sum of the effective capacity in HD relay and FD relay vs. weight w and different distance between relay and nodes.

intervals to transmit the same packet. When the self-interference is removed properly, the performance of the FD relay is twice that of the HD relay.

By increasing the self-interference coefficient Ω, the FD relay's performance decreases. Even at large quantities of θ, which indicates the strict quality of service and very low buffer delay, the HD relay shows a better performance. In other words, self-interference's effect is more highlighted at large quantities of θ and strict quality of service. Therefore, if strict service quality is required, it is suggested to remove self-interference in the FD relay completely or use the HD relay. With further increase in θ, the FD relay performance will eventually be worse than the HD relay at all values of θ. In such cases, the self-interference is so severe that it can even compensate the double use of HD relays in time or frequency intervals.
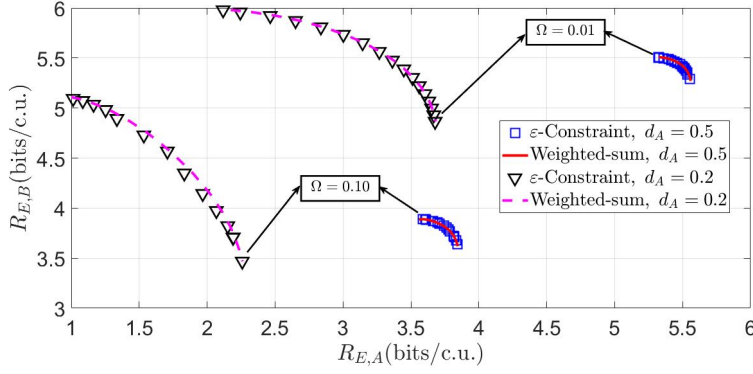
In the proposed model, the priority of nodes A and B are not necessarily the same. Therefore, in transforming the multi-objective problem and providing the weighted sum of

the effective capacity, the weight w might have different values. Figure 7 shows the weighted sum of the effective capacity in terms of w for the HD relay and FD relay with $\Omega = 0.1$. To provide a better comparison, the distance of the relay to node A is set to the three values of $d_A = 0.3$, $d_A = 0.5$ and $d_A = 0.7$. When $d_A = 0.5$ and the relay is in the middle of two nodes, the coefficient w does not have much effect on the weighted sum of the capacity. If the relay is in the middle of two nodes, the SNR (and SINR) of the two nodes are symmetric. On the average case, the weighted sum of the effective capacity of the system is unique under all values of w. However, when the relay is not in the middle of two nodes, the priority of nodes and the w value have a relatively significant effect on the weighted sum of the effective capacity. However, when the relay is close to node A and $d_A = 0.3$, the weighted sum of the effective capacity decreases with increasing w (increasing the performance priority of node A). In this case, the relay channel to node B has higher attenuation than the relay channel to node A . In lower values of w, the power is divided to compensate the poor condition of the relay channel to node B , and relatively good performance is results. However, as w increases and the priority of node A is enhances, the power is divided such that node A achieves an appropriate performance. Therefore, the performance of node B is ignored. In this case, the inappropriate condition of the relay channel to node B have an impact on the overall system performance and reduces the weighted sum of the effective capacity. When $d_A = 0.7$, the situation is completely reversed and by increasing w, the weighted sum of capacity is enhanced. Therefore, to achieve proper performance, when the priority of nodes is different, it is necessary to locate the relay between nodes in a right place and it has a significant effect on the overall system performance.

In the present paper, the problem is solved as multi-objective optimization instead of using a single-objective optimization approach. In the single-objective problem, the objective function is defined as the mean effective capacity of two nodes. Therefore $R_E = (R_{E,A} + R_{E,B})/2$ is maximized. The solution to this problem is equivalent to solving the multi-objective problem with $w = 0.5$. Therefore, modeling the problem as multi-objective and its solution is more comprehensive than the conventional single-objective model. Therefore, the results should be better. This improvement in performance is quite evident in Figure 7 when the relay is not set in the middle of nodes A and B . For example, when the relay is near node A and $d_A = 0.3$, the single-objective solution's answer is $R_E = 3.46$(bits/c.u.) in FD mode ($R_E = 2.61$(bits/c.u.) in the HD mode). However, in the multi-objective solution, if the weight w is chosen properly, the weighted sum of the effective capacity may be higher than the mean effective capacity by about 25 percent in the FD mode (15 percent in HD mode). Another point in multi-objective optimization is the possibility of prioritizing the performance of nodes, which improves the overall system performance.

It was discussed in Section 4.2 that the weighted sum method is the only sufficient condition to reach the optimal Pareto point in the non-concave multi-objective problems. To evaluate the performance of this method, the calculated Pareto point by the weighted sum method and ε-constraint method are plotted in Figure 8. The ε-constraint method is the necessary and sufficient condition to reach all the optimal Pareto points in non-concave problems [27]. This method is very simple when there are only two objective functions in the multi-objective optimization problem. Accordingly, one of the objective functions such as $R_{E,B}$ in (30) is transferred to the problem constraints and the optimization problem is rewritten as:

$$\begin{cases} \max_{P_R} R_{E,A} \\ \text{s.t.} R_{E,B} > \mu \\ \text{s.t.} P_R < P_{tot} \end{cases} \qquad (36)$$

Fig. 8 Pareto optimal frontier with the weighted sum and ε-constraint methods.

Table 2 Comparison of execution time in the optimal and approximated problem.

relay mode	HD		
$\varepsilon = \varepsilon_A = \varepsilon_B$	10^{-8}	10^{-5}	10^{-2}
optimal solution	1415	1356	1327
approximated solution	564	561	561

relay mode	FD		
Ω	0.01	0.05	0.10
optimal solution	1935	1657	1622
approximated solution	553	540	547

In this problem, the Pareto points is calculated by altering μ and solving the problem (in the ε-constraint method, when one of the objective functions is added to the problem constraints, this function is compared with ε and this is the reason of naming this method. However, in this article, ε is used elsewhere; therefore $R_{E,B}$ is compared with μ in the added constraint).

In Figure 8, the optimal Pareto frontier is plotted by the weighted sum and ε-constraint in FD relay. For a detailed analysis, the distance between the relay and node A is assumed as $d_A = 0.5$ and $d_A = 0.2$ individually. Moreover, the intensity of self-interference is considered as the low and relatively high values of $\Omega = 0.01$ and $\Omega = 0.10$ respectively. The calculated Pareto frontier by both methods are almost identical. Therefore, the solution to the non-concave optimization problem (30) with the weighted sum method is very close to the optimal performance.

If the relay is in the middle of two nodes and $d_A = 0.5$, the performance is symmetric and the range of variations for two nodes is the same. In addition, the range of variations is not extensive. However, when the relay is close to node A , symmetry is lost and the variation interval increases significantly. For example, when $\Omega = 0.01$, the effective capacity of node A varies within the range of $R_{E,A} = 2.1$(bits/c.u.) to $R_{E,A} = 3.7$(bits/c.u.) and that of node B varies between $R_{E,B} = 4.9$(bits/c.u.) and $R_{E,B} = 6.0$(bits/c.u.). If $\Omega = 0.10$, the effective capacity of node A varies within the range of $R_{E,A} = 1.0$(bits/c.u.) to $R_{E,A} = 2.2$(bits/c.u.) and that of node B varies between $R_{E,B} = 3.5$(bits/c.u.) and $R_{E,B} = 5.1$(bits/c.u.). Therefore, especially when the relay is not located between the two nodes A and B , power allocation is very efficient, which can significantly affect the overall system performance.

Table 2 provides a comparison between the execution time of optimization problem in both optimal and approximated problem for HD relay with 3 different error probability and FD relay with 3 different values of Ω. The simulations were implemented on a laptop with an 8-core Intel (R) Core (TM) i7-3632QM CPU @ 2.20GHz processor with 6GB of memory. For any level of error probability, 1000 random samples are generated as a channel and simulations are performed to achieve the gradient value of less than 0.1 at each point. Then the required time for optimal and approximated solution are recorded. This procedure is repeated 100 times and the results of the average runtime are presented in Table 2. The execution time of the approximated problem is much shorter than that of the optimal problem. This difference increases with the increasing number of channel samples.

6 Conclusions

Considering the significance of low-latency data transmission in the fifth generation of communications, this article addressed transmission of the short packets between two nodes. Short packet data transmission requires short transmission time and decoding. Therefore, it is appropriate to delay-sensitive traffics. Moreover, the two-way relay is applied for data transmission between two nodes. The relay type could be either HD or FD. If an FD relay is applied, the data transmission time between two nodes is reduced. Finally, the weighted sum of the effective capacity for two nodes is considered to evaluate the performance. The effective capacity indicates the transferable rate through the channel by guaranteeing the statistical delay of packet in the transmitter buffer.

In the proposed model, the optimal power for effective capacity maximization of two nodes is allocated when the packets are transferred between these two nodes with a two-way HD or FD relay. It is also proved that the resulting performance is better than identical power allocation among the nodes. Since the performance priority of the two nodes is not necessarily the same, the problem optimization is considered multi-objective. In the case of using the HD relay, it is proved that the optimization problem is concave and it was solved by the weighted sum method. However, in the case of using the FD relay, although the problem has one optimal point, it is not concave. Therefore, the weighted sum method was also applied to solve the optimization problem in these conditions. As the optimization problem involves statistical means, its solution is long and time-consuming. Thus, the approximate method was obtained to calculate the optimal power of the relays and nodes. The answer of solving the approximate problem has an appropriate consistency with the results of the optimal solution. It is also proved that multi-objective optimization provides better results than solving the problem with the single-objective approach. In the single-objective case, the effective capacity of two nodes with similar weights is summed up and maximized. However, in the multi-objective case, the priority and weight of the nodes are not necessarily equal. Therefore, the overall system performance is better than the performance of the system in single-mode especially when the relay is not in the middle of two nodes or their priority is not the same. Regarding the performance of HD and FD relays, the FD relay outperforms the HD relay when the self-interference is reduced to an appropriate level. With increasing θ and the strict service requirement of the nodes, the performance of both HD and FD relays decreases naturally. However, in this case, HD outperforms FD gradually. Therefore, it is concluded that if the strict quality of service is considered, it is advised to remove self-interference in the FD relay largely or use the HD relay.

References

1. Lin Zhang and Ying-Chang Liang. Average throughput analysis and optimization in cooperative iot networks with short packet communication. *IEEE Transactions on Vehicular Technology*, 67(12):11549–11562, 2018.
2. Aditya Tiwary, Manish Mahato, Abhitesh Chidar, Mayank Kumar Chandrol, Mayank Shrivastava, and Mohit Tripathi. Internet of things (iot): Research, architectures and applications. *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4(3):23–27, 2018.
3. Dennis Miller. Blockchain and the internet of things in the industrial sector. *IT professional*, 20(3):15–18, 2018.
4. Martin Wollschlaeger, Thilo Sauter, and Juergen Jasperneite. The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0. *IEEE industrial electronics magazine*, 11(1):17–27, 2017.
5. Zhihan Lv. Virtual reality in the context of internet of things. *Neural Computing and Applications*, 32(13):9593–9602, 2020.
6. Christopher G Coogan and Bin He. Brain-computer interface control in a virtual reality environment and applications for the internet of things. *IEEE Access*, 6:10840–10849, 2018.
7. Gordon J Sutton, Jie Zeng, Ren Ping Liu, Wei Ni, Diep N Nguyen, Beeshanga A Jayawickrama, Xiaojing Huang, Mehran Abolhasan, Zhang Zhang, Eryk Dutkiewicz, et al. Enabling technologies for ultra-reliable and low latency communications: From phy and mac layer perspectives. *IEEE Communications Surveys & Tutorials*, 21(3):2488–2524, 2019.
8. Haythem A Bany Salameh, Sufyan Almajali, Moussa Ayyash, and Hany Elgala. Spectrum assignment in cognitive radio networks for internet-of-things delay-sensitive applications under jamming attacks. *IEEE Internet of Things Journal*, 5(3):1904–1913, 2018.
9. Mithun Mukherjee, Suman Kumar, Mohammad Shojafar, Qi Zhang, and Constandinos X Mavromoustakis. Joint task offloading and resource allocation for delay-sensitive fog networks. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2019.
10. Ping Yang, Yue Xiao, Ming Xiao, and Shaoqian Li. 6g wireless communications: Vision and potential techniques. *IEEE Network*, 33(4):70–75, 2019.
11. Changyang She, Chenyang Yang, and Tony QS Quek. Radio resource management for ultra-reliable and low-latency communications. *IEEE Communications Magazine*, 55(6):72–78, 2017.
12. Mahyar Shirvanimoghaddam, Mohammad Sadegh Mohammadi, Rana Abbas, Aleksandar Minja, Chentao Yue, Balazs Matuz, Guojun Han, Zihuai Lin, Wanchun Liu, Yonghui Li, et al. Short block-length codes for ultra-reliable low latency communications. *IEEE Communications Magazine*, 57(2):130–137, 2018.
13. He Chen, Rana Abbas, Peng Cheng, Mahyar Shirvanimoghaddam, Wibowo Hardjawana, Wei Bao, Yonghui Li, and Branka Vucetic. Ultra-reliable low latency cellular networks: Use cases, challenges and approaches. *IEEE Communications Magazine*, 56(12):119–125, 2018.
14. Mohammad Lari. Transmission delay minimization in wireless powered communication systems. *Wireless Networks*, 25(3):1415–1430, 2019.
15. Yifan Gu, He Chen, Yonghui Li, Lingyang Song, and Branka Vucetic. Short-packet two-way amplify-and-forward relaying. *IEEE Signal Processing Letters*, 25(2):263–267, 2017.
16. Giuseppe Durisi, Tobias Koch, and Petar Popovski. Toward massive, ultrareliable, and low-latency wireless communication with short packets. *Proceedings of the IEEE*, 104(9):1711–1726, 2016.
17. Yury Polyanskiy, H Vincent Poor, and Sergio Verdú. Channel coding rate in the finite blocklength regime. *IEEE Transactions on Information Theory*, 56(5):2307–2359, 2010.
18. Daquan Feng, Changyang She, Kai Ying, Lifeng Lai, Zhanwei Hou, Tony QS Quek, Yonghui Li, and Branka Vucetic. Toward ultrareliable low-latency communications: Typical scenarios, possible solutions, and open issues. *IEEE Vehicular Technology Magazine*, 14(2):94–102, 2019.
19. Xilin Cheng, Bo Yu, Xiang Cheng, and Liuqing Yang. Two-way full-duplex amplify-and-forward relaying. In *MILCOM 2013-2013 IEEE Military Communications Conference*, pages 1–6. IEEE, 2013.
20. Zhengquan Zhang, Zheng Ma, Zhiguo Ding, Ming Xiao, and George K Karagiannidis. Full-duplex two-way and one-way relaying: average rate, outage probability, and tradeoffs. *IEEE Transactions on Wireless Communications*, 15(6):3920–3933, 2016.
21. Hyoungju Ji, Sunho Park, Jeongho Yeo, Younsun Kim, Juho Lee, and Byonghyo Shim. Ultra-reliable and low-latency communications in 5g downlink: Physical layer aspects. *IEEE Wireless Communications*, 25(3):124–130, 2018.
22. Mostafa Darabi and Lutz Lampe. Multi objective resource allocation for joint embb and urllc traffic with different qos requirements. In *2019 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2019.
23. Ruizhe Qi, Xuefen Chi, Linlin Zhao, and Wanting Yang. Martingales-based aloha-type grant-free access algorithms for multi-channel networks with mmtc/urllc terminals co-existence. *IEEE Access*, 8:37608–37620, 2020.

24. Ali A Esswie and Klaus I Pedersen. Opportunistic spatial preemptive scheduling for urllc and embb coexistence in multi-user 5g networks. *IEEE Access*, 6:38451–38463, 2018.

25. Mahdi Haghifam, Mohammad Robat Mili, Behrooz Makki, Masoumeh Nasiri-Kenari, and Tommy Svensson. Joint sum rate and error probability optimization: Finite blocklength analysis. *IEEE Wireless Communications Letters*, 6(6):726–729, 2017.

26. Emil Bjornson, Eduard Axel Jorswieck, Mérouane Debbah, and Bjorn Ottersten. Multiobjective signal processing optimization: The way to balance conflicting metrics in 5g systems. *IEEE Signal Processing Magazine*, 31(6):14–23, 2014.

27. Jin-Hee Cho, Yating Wang, Ray Chen, Kevin S Chan, and Ananthram Swami. A survey on modeling and optimizing multi-objective systems. *IEEE Communications Surveys & Tutorials*, 19(3):1867–1901, 2017.

28. Mohammad Lari and Sina Asaeian. Multi-objective antenna selection in a full duplex base station. *Wireless Personal Communications*, 110(2):781–793, 2020.

29. Caihong Kai, Yibo Yi, Min Peng, and Wei Huang. An amplify-and-forward full-duplex cooperative relay scheme for low-latency downlink transmission in cran. *IEEE Communications Letters*, 2020.

30. Yufei Jiang, Hanjun Duan, Xu Zhu, Zhongxiang Wei, Tong Wang, Fu-Chun Zheng, and Sumei Sun. Toward urllc: A full duplex relay system with self-interference utilization or cancellation. *IEEE Wireless Communications*, 28(1):74–81, 2021.

31. Yifan Gu, He Chen, Yonghui Li, and Branka Vucetic. Ultra-reliable short-packet communications: Half-duplex or full-duplex relaying? *IEEE Wireless Communications Letters*, 7(3):348–351, 2017.

32. Keshav Singh, Sudip Biswas, Meng-Lin Ku, and Mark F Flanagan. Transceiver design for full-duplex ultra-reliable low-latency communications with finite blocklength. In *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2020.

33. Jie Zeng, Tiejun Lv, Zhipeng Lin, Ren Ping Liu, Jiajia Mei, Wei Ni, and Y Jay Guo. Achieving ultrareliable and low-latency communications in iot by fd-scma. *IEEE Internet of Things Journal*, 7(1):363–378, 2019.

34. Kuo-Guan Wu, Feng-Tsun Chien, Yu-Feng Lin, and Min-Kuan Chang. Sinr and delay analyses in two-way full-duplex swipt-enabled relaying systems. *IEEE Transactions on Communications*, 2020.

35. Chongtao Guo, Le Liang, and Geoffrey Ye Li. Resource allocation for low-latency vehicular communications: An effective capacity perspective. *IEEE Journal on Selected Areas in Communications*, 37(4):905–917, 2019.

36. Jinho Choi. An effective capacity-based approach to multi-channel low-latency wireless communications. *IEEE Transactions on Communications*, 67(3):2476–2486, 2018.

37. Hong Ren, Nan Liu, Cunhua Pan, Maged Elkashlan, Arumugam Nallanathan, Xiaohu You, and Lajos Hanzo. Power-and rate-adaptation improves the effective capacity of c-ran for nakagami-*m* fading channels. *IEEE Transactions on Vehicular Technology*, 67(11):10841–10855, 2018.

38. Jihas Khan and Lillykutty Jacob. Resource allocation for comp enabled urllc in 5g c-ran architecture. *IEEE Systems Journal*, 2020.

39. Yulin Hu, M Cenk Gursoy, and Anke Schmeink. Optimal power allocation for amplify and forward relaying with finite blocklength codes and qos constraints. In *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, 2018.

40. Yulin Hu, Mustafa Ozmen, M Cenk Gursoy, and Anke Schmeink. Optimal power allocation for qos-constrained downlink networks with finite blocklength codes. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2018.

41. Yulin Hu, James Gross, and Anke Schmeink. On the performance advantage of relaying under the finite blocklength regime. *IEEE Communications Letters*, 19(5):779–782, 2015.

42. Yulin Hu, Yao Zhu, M Cenk Gursoy, and Anke Schmeink. Swipt-enabled relaying in iot networks operating with finite blocklength codes. *IEEE Journal on Selected Areas in Communications*, 37(1):74–88, 2018.

43. Yulin Hu, James Gross, and Anke Schmeink. On the capacity of relaying with finite blocklength. *IEEE Transactions on Vehicular Technology*, 65(3):1790–1794, 2015.

44. Xilin Cheng, Bo Yu, Xiang Cheng, and Liuqing Yang. Two-way full-duplex amplify-and-forward relaying. In *MILCOM 2013 - 2013 IEEE Military Communications Conference*, pages 1–6, 2013.

45. Mohammad Lari, Abbas Mohammadi, Abdolali Abdipour, and Inkyu Lee. Characterization of effective capacity in antenna selection mimo systems. *journal of communications and networks*, 15(5):476–485, 2013.

46. Mohammad Lari, Abbas Mohammadi, Abdolali Abdipour, and Inkyu Lee. Characterization of effective capacity in af relay systems. *IEICE Electronics Express*, 9(7):679–684, 2012.

47. Deli Qiao, M Cenk Gursoy, and Senem Velipasalar. Throughput-delay tradeoffs with finite blocklength coding over multiple coherence blocks. *IEEE Transactions on Communications*, 67(8):5892–5904, 2019.

48. Leonardo Jimenez Rodriguez, Nghi H. Tran, and Tho Le-Ngoc. Optimal power allocation and capacity of full-duplex af relaying under residual self-interference. *IEEE Wireless Communications Letters*, 3(2):233–236, 2014.

49. Hyungsik Ju and Rui Zhang. Optimal resource allocation in full-duplex wireless-powered communication network. *IEEE Transactions on Communications*, 62(10):3528–3540, 2014.

50. Wided Hadj Alouane. Power allocation for asymmetric two-way fixed-gain af relaying networks. *Telecommunication Systems*, 71(4):553–559, 2019.

51. Alan Jeffrey and Daniel Zwillinger. *Table of integrals, series, and products*. Elsevier, 2007.