RESEARCH   **Open Access**

Energy efficiency analysis of one-way and two-way relay systems

Can Sun[*] and Chenyang Yang

Abstract

Relaying is supposed to be a low energy consumption technique since the long distance transmission is divided into several short distance transmissions. When the power consumptions (PCs) other than that consumed by transmitting information bits is taken into account, however, relaying may not be energy efficient. In this article, we study the energy efficiencies (EEs) of one-way relay transmission (OWRT) and two-way relay transmission (TWRT) by comparing with direct transmission (DT). We consider a system where two source nodes transmit to each other with the assistance of a half-duplex amplify-and-forward relay node. We first find the maximum EEs of DT, OWRT, and TWRT by optimizing the transmission time and the transmit powers at each node. Then we compare the maximum EEs of the three strategies, and analyze the impact of circuit PCs and data amount. Analytical and simulation results show that relaying is not always more energy efficient than DT. Moreover, TWRT is not always more energy efficient than OWRT, despite that it is more spectral efficient. The EE of TWRT is higher than those of DT and OWRT in symmetric systems where the circuit PCs at each node are identical and the numbers of bits to be transmitted in two directions are equal. In asymmetric systems, however, OWRT may provide higher EE than TWRT when the numbers of bits in two directions differ significantly.

1 Introduction

Since the explosive growth of wireless services is sharply increasing their contributions to the carbon footprint and the operating costs, energy efficiency (EE) has drawn more and more attention recently as a new design goal for various wireless communication systems [1-3], compared with spectral efficiency (SE) that has been the design focus for decades.

A widely used performance metric for EE is the number of transmitted bits per unit of energy. When only transmit power is taken into account, the EE monotonically decreases with the increase of the SE [4] at least for point-to-point transmission in additive white Gaussian noise (AWGN) channel. In that case, when we minimize the transmit power, the EE will be maximized [5]. In practical systems, however, not only the power for transmitting information bits but also various signaling and circuits contribute to the system energy consumption (EC), which fundamentally change the relationship between the SE and EE. Specifically, when the circuit power consumption (PC) is considered, the

optimization problem that minimizes the overall transmit power does not necessarily lead to an energy efficient design [2].

Relaying is viewed as an energy saving technique because it can reduce the transmit power by breaking one long range transmission into several short range transmissions [3]. In fact, relaying has been extensively studied from another viewpoint, i.e., it is able to extend the coverage, enhance the reliability as well as the capacity of wireless systems [6]. One-way relay transmission (OWRT) can reduce the one-hop communication distance and provide spatial diversity, but its SE will also reduce to 1/2 of that of direct transmission (DT) when practical half-duplex relay is applied [7]. Fortunately, two-way relay transmission (TWRT) can recover the SE loss when properly designed [8-10]. However, it is not well-understood whether these relay strategies are energy efficient, when various energy costs in addition to transmit power are considered.

Considering both the transmit power and the receiver processing power, the EE of decode-and-forward (DF) OWRT systems was studied with single-antenna and multi-antenna nodes in [11,12], respectively. In [13], after accounting for the energy cost of acquiring channel

* Correspondence: saga@ee.buaa.edu.cn
School of Electronics and Information Engineering, Beihang University, Beijing 100191, China

information, relay selection for an OWRT system with multiple DF relays was optimized to maximize the EE. In [14], the EE of DF OWRT was compared with that of DT, where the result shows that OWRT is more energy efficient when the distance between source and destination is large, otherwise DT is better. In [15,16], the EEs of OWRT and base station cooperation transmission were compared, where the overall energy costs including those from manufacture and deployment were considered. In [17], TWRT was shown to be more energy efficient than OWRT via simulations, where only transmit power was considered in the EC model. In [5], the EE of TWRT was compared with those of OWRT and DT, with optimized relay position and transmit power at each node. It shows that when the relay is placed at the midpoint of two source nodes, TWRT consumes less energy than OWRT and DT. Again, only transmit power was considered in the EC model. When we take into account the energy costs other than that contributed by the transmit power, what is the results of comparison between relaying and DT? Will TWRT still be more energy efficient than OWRT?

In this article, we analyze the EEs of TWRT, OWRT, and DT by studying a simple amplify-and-forward (AF) relay system. In literature, there are other relay protocols such as DF and compress-and-forward (CF) that provide higher rate regions than AF. However, AF is also widely considered in practice [6], and is superior to DF in outage performance for TWRT when the channel gains from two source nodes to the relay node are symmetric [18]. Moreover, the system models differ a lot among the relay protocols. In order to analyze the maximal EE, we need to find the relationship between end-to-end data rate and transmit power. With AF protocol, we can obtain the data rate-transmit power relationship by deriving the signal-to-noise ratio (SNR) at the destination. With DF protocol, the end-to-end data rate is quite different, which is modeled as the lower one of the achievable data rates in two hops. When considering CF, the case is even more complicated since its transmission and processing procedure is usually very complex, which is rather involved for analysis. Here we focus on AF relay as a good start, while the EEs of other relay protocols will be considered in future studies.

We consider a delay-constrained system, where B bits of message should be transmitted as a block within a duration T. This model is widely used for applications with strict delay constraints on data delivery, e.g., Voice-over-IP and sensor networks, where the message is generated periodically and must be transmitted with a hard deadline [19-21]. Note that the energy consumed by transmitting information decreases as the transmission duration increases [4], but the energy consumed by circuits increases with the duration. Therefore, in such a

system we can adjust the transmission duration to reduce the overall EC as long as the transmission duration is shorter than the block length T. In other word, the system may transmit the B bits in a shorter duration than T and then switch to an idle status until the next block [21]. During the idle status, a part of the transceiver hardware can be shut down, which can be exploited to improve the EE.

Specifically, we first maximize the EEs of TWRT, OWRT, and DT by optimizing transmission time and transmit powers, respectively, for the three strategies. We then compare the optimized EEs of TWRT with those of OWRT and DT. We show that when all the three strategies operate with optimized transmission time and power, relaying is *not* always more energy efficient than DT. Moreover, TWRT is *not* always more energy efficient than OWRT if the numbers of bits to be transmitted in two directions are unequal, or the circuit PCs at each node are different.
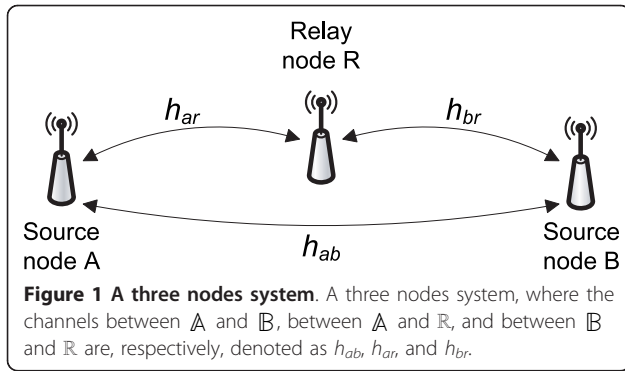
The rest of this article is organized as follows. System model and the ECs of the three transmit strategies are, respectively, described in Sections 2 and 3. Then the EEs of different strategies are optimized in Section 4. In Section 5, the optimized EEs are compared under varies circuit PCs and numbers of transmitted bits. Simulation results are given in Section 6. Section 7 concludes the article.

2 System model

Consider a system consisting of two source nodes $\mathbb{A}$ and $\mathbb{B}$, and an AF half-duplex relay node (RN) $\mathbb{R}$, each equipped with a single antenna. We consider a delay constrained system, where the information bits are generated periodically and must be transmitted in a block within a hard deadline T. In each block, nodes A and B, respectively, intends to transmit B_{ab} and B_{ba} bits to each other with bandwidth W. In practice, the information bits to be transmitted in each block compose a packet or a frame, depending on application scenarios. In the following, we use the term "packet size" to refer the amount of data in each block, i.e., B_{ab} and B_{ba}.

The channels among three nodes are assumed as frequency-flat fading channels, which are respectively, denoted as h_{ab}, h_{ar}, and h_{br}, as shown in Figure 1. We assume perfect channel knowledge at each node. The noise power N_0 is assumed to be identical at each node.

To reduce the EC, the system may not use the entire duration T for transmission in each block. After B_{ab} and B_{ba} bits have been transmitted, the nodes can operate at an idle status until next block. In other word, each node has three modes: transmission, reception, and idle. The PCs in these modes are, respectively, denoted as $P^t/\epsilon + P^{ct}$, P^{cr}, and P^{ci}, where P^t is the transmit power, $\epsilon \in (0, 1]$ denotes the power amplifier efficiency, P^{ct}, P^{cr}, and

Figure 1 A three nodes system. A three nodes system, where the channels between $\mathbb{A}$ and $\mathbb{B}$, between $\mathbb{A}$ and $\mathbb{R}$, and between $\mathbb{B}$ and $\mathbb{R}$ are, respectively, denoted as h_{ab}, h_{ar}, and h_{br}.

Figure 2 Transmission procedure in each block. Bidirectional transmission procedure in each block, where **(a)** is for direct transmission, **(b)** is for one-way relaying, and **(c)** is for two-way relaying.

P^{ci} are, respectively, the circuit PCs in transmission, reception, and idle modes.

The circuit PCs in P^{ct} and P^{cr} consist of two parts: the power consumed by baseband processing and radio frequency (RF) circuits. The PC of RF circuit is usually assumed independent of data rate [6,21], while there are different assumptions for the PC of baseband processing circuit. In systems with low complexity baseband processing, the baseband PC can be neglected compared with the RF PC [6,21]. Otherwise, the baseband PC is not negligible and increases with data rate [22]. In this article, we consider the first case, where P^{ct} and P^{cr} only consist of RF PC, which are modeled as constants independent of data rate. Modeling P^{ct} and P^{cr} as functions of data rate leads to a different optimization problem, which will be considered in our future study.

The PC in idle mode P^{ci} is modeled as a constant, and $P^{ci} \leq P^{ct}$, $P^{ci} \leq P^{cr}$. Subscripts $(\cdot)_a$, $(\cdot)_b$, and $(\cdot)_r$ will be used to denote the PCs at different nodes.

3 Energy consumptions of three transmit strategies

We consider three transmit strategies, DT, OWRT, and TWRT, to complete the bidirectional communication between the two source nodes. In the following, we respectively introduce their ECs.

3.1 Direct transmission

In DT, nodes $\mathbb{A}$ and $\mathbb{B}$ transmit to each other without the assistance of RN. The transmission procedure is shown in Figure 2a. During each block, the system first allocates a duration T_{ab} for the transmission from node $\mathbb{A}$ to $\mathbb{B}$, where node $\mathbb{A}$ is in transmit mode and node $\mathbb{B}$ is in receive mode. Then the system allocates a duration T_{ba} for the transmission from node $\mathbb{B}$ to $\mathbb{A}$, where node $\mathbb{A}$ is in receive mode and node $\mathbb{B}$ is in transmit mode. After the B_{ab} and B_{ba} bits are transmitted, the system turns into idle status during $T - T_{ab} - T_{ba}$, where both nodes $\mathbb{A}$ and $\mathbb{B}$ are in idle mode. The EC of DT can be obtained as
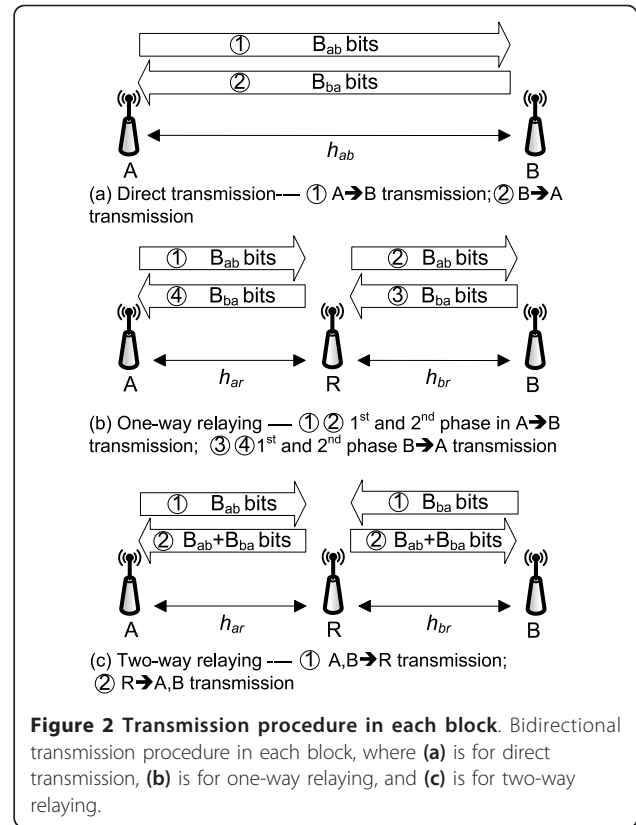
$$
\begin{aligned}
E_D &= T_{ab}\left(P_a^t/\varepsilon + P_a^{ct} + P_b^{cr}\right) + T_{ba}\left(P_b^t/\varepsilon + P_b^{ct} + P_a^{cr}\right) \\
&\quad + (T - T_{ab} - T_{ba})\left(P_a^{ci} + P_b^{ci}\right) \\
&= T_{ab}\left(P_a^t/\varepsilon + P_D^{c1} + P_D^{ci}\right) + T_{ba}\left(P_b^t/\varepsilon + P_D^{c2} - P_D^{ci}\right) + TP_D^{ci}
\end{aligned}
\tag{1}
$$

where $P_D^{c1} \triangleq P_a^{ct} + P_b^{cr}$ and $P_D^{c2} \triangleq P_b^{ct} + P_a^{cr}$ are, respectively, the total circuit PCs in $\mathbb{A} \rightarrow \mathbb{B}$ and $\mathbb{B} \rightarrow \mathbb{A}$ transmission, and $P_D^{ci} \triangleq P_a^{ci} + P_b^{ci}$ is the total circuit PC in idle duration.

Given T_{ab} and T_{ba}, nodes $\mathbb{A}$ and $\mathbb{B}$ should, respectively, transmit with data rates of B_{ab}/T_{ab} and B_{ba}/T_{ba} bits-per-second (bps) to exchange the B_{ab} and B_{ba} bits messages, which are given by Shannon capacity formula as

$$
\frac{B_{ab}}{T_{ab}} = W\log_2\left(1 + \frac{P_a^t|h_{ab}|^2}{N_0}\right), \quad \frac{B_{ba}}{T_{ba}} = W\log_2\left(1 + \frac{P_b^t|h_{ab}|^2}{N_0}\right).
\tag{2}
$$

Since Shannon capacity formula represents the maximum achievable data rates under given transmit powers, the transmit power derived via this formula is the minimum transmit power that can support the required data rates. As a result, we can analyze the maximal EE for a given SE. We will also use the Shannon capacity formula to represent the relationship between data rates and transmit powers in OWRT and TWRT cases later.

3.2 One-way relay transmission

In OWRT, each of the $\mathbb{A} \to \mathbb{B}$ and $\mathbb{B} \to \mathbb{A}$ transmission is divided into two hops, thus the bidirectional transmission needs four phases, as shown in Figure 2b. For example, in $\mathbb{A} \to \mathbb{B}$ transmission, node $\mathbb{A}$ transmits to RN in the first phase, and RN transmits to node $\mathbb{B}$ in the second phase. With the AF relay protocol, the two phases in each direction employ identical time duration. For simplifying the analysis, we do not consider the direct link in OWRT. Although this will degrade the performance of OWRT, we will show later that it does not affect our comparison results for the EE.

The system allocates a duration T_{ab} for $\mathbb{A} \to \mathbb{B}$ transmission. During the first half of T_{ab}, node $\mathbb{A}$ transmits to RN, and thus node $\mathbb{A}$ is in transmit mode, node $\mathbb{R}$ is in receive mode, and node $\mathbb{B}$ is idle. During the second half of T_{ab}, RN forwards the information to node $\mathbb{B}$, and thus node $\mathbb{R}$ is in transmit mode, node $\mathbb{B}$ is in receive mode, and node $\mathbb{A}$ is idle. Then, the system allocates a duration T_{ba} for $\mathbb{B} \to \mathbb{A}$ transmission. Finally, the system turns into idle status during $T - T_{ab} - T_{ba}$ after the bidirectional transmission. The EC of OWRT can be obtained as

$$
\begin{aligned}
E_O &= \frac{T_{ab}}{2}\left(P_a^t/\varepsilon + P_a^{ct} + P_r^{cr} + P_b^{ci} + P_{r1}^t/\varepsilon + P_r^{ct} + P_b^{cr} + P_a^{ci}\right) \\
&\quad + \frac{T_{ba}}{2}\left(P_b^t/\varepsilon + P_b^{ct} + P_r^{cr} + P_a^{ci} + P_{r2}^t/\varepsilon + P_r^{ct} + P_a^{cr} + P_b^{ci}\right) \\
&\quad + (T - T_{ab} - T_{ba})\left(P_a^{ci} + P_b^{ci} + P_r^{ci}\right) \\
&= T_{ab}\left(\frac{P_a^t + P_{r1}^t}{2\varepsilon} + P_O^{c1} - P_O^{ci}\right) + T_{ba}\left(\frac{P_b^t + P_{r2}^t}{2\varepsilon} + P_O^{c2} - P_O^{ci}\right) + TP_O^{ci},
\end{aligned}
\tag{3}
$$

where P_{r1}^t and P_{r2}^T are, respectively, the relay transmit powers in $\mathbb{A} \to \mathbb{B}$ and $\mathbb{B} \to \mathbb{A}$ links, $P_O^{c1} \triangleq \left(P_a^{ct} + P_r^{cr} + P_b^{ci} + P_r^{ct} + P_b^{cr} + P_a^{ci}\right)/2$ and $P_O^{c2} \triangleq \left(P_b^{ct} + P_r^{cr} + P_a^{ci} + P_r^{ct} + P_a^{cr} + P_b^{ci}\right)/2$ are, respectively, the overall circuit PCs in $\mathbb{A} \to \mathbb{B}$ and $\mathbb{B} \to \mathbb{A}$ transmission, and $P_O^{ci} \triangleq P_a^{ci} + P_b^{ci} + P_r^{ci}$ is the overall circuit PC in idle duration where all three nodes operate in idle mode.

The required bidirectional data rates can be obtained from the capacity formula and the expression of SNR for OWRT derived in [23], which are respectively,

$$
\frac{B_{ab}}{T_{ab}} = \frac{W}{2}\log_2\left(1 + \frac{P_a^t P_{r1}^t |h_{ar}|^2 |h_{br}|^2}{|h_{ar}|^2 P_a^t N_0 + |h_{br}|^2 P_{r1}^t N_0 + N_0^2}\right)
\tag{4}
$$

$$
\frac{B_{ba}}{T_{ba}} = \frac{W}{2}\log_2\left(1 + \frac{P_b^t P_{r2}^t |h_{br}|^2 |h_{ar}|^2}{|h_{br}|^2 P_b^t N_0 + |h_{ar}|^2 P_{r2}^t N_0 + N_0^2}\right)
\tag{5}
$$

where the factor 1/2 is due to the two-phase transmission in each direction.

3.3 Two-way relay transmission

In TWRT, the bidirectional transmission is completed in two phases, as shown in Figure 2c. In the first phase, both nodes $\mathbb{A}$ and $\mathbb{B}$ transmit to RN, where the nodes $\mathbb{A}$ and $\mathbb{B}$ are in transmit mode and the node $\mathbb{R}$ is in receive mode. In the second phase, RN broadcasts its received signal to the nodes $\mathbb{A}$ and $\mathbb{B}$, where the node $\mathbb{R}$ is in transmit mode, and the nodes $\mathbb{A}$ and $\mathbb{B}$ are in receive mode. After receiving the superimposed signal, each of the source nodes $\mathbb{A}$ and $\mathbb{B}$ removes its own transmitted signal via self-interference cancelation [8], and obtains its desired signal sent from the other source node. The two phases employ identical durations as in OWRT.

The system allocates duration T_{TWR} to the bidirectional transmission, and then turns into idle status during $T - T_{\mathrm{TWR}}$. The EC of TWRT is obtained as

$$
\begin{aligned}
E_T &= \frac{T_{\mathrm{TWR}}}{2}\left(P_a^t/\varepsilon + P_b^t/\varepsilon + P_a^{ct} + P_b^{ct} + P_r^{cr}\right) + \frac{T_{\mathrm{TWR}}}{2}\left(P_r^t/\varepsilon + P_r^{ct} + P_a^{cr} + P_b^{cr}\right) \\
&\quad + (T - T_{\mathrm{TWR}})\left(P_a^{ci} + P_b^{ci} + P_r^{ci}\right) \\
&= T_{\mathrm{TWR}}\left(\frac{P_a^t + P_b^t + P_r^t}{2\varepsilon} + P_T^c - P_T^{ci}\right) + TP_T^{ci},
\end{aligned}
\tag{6}
$$

where $P_T^c \triangleq \left(P_a^{ct} + P_b^{ct} + P_r^{ct} + P_r^{cr} + P_a^{cr} + P_b^{cr}\right)/2$ and $P_T^{ci} \triangleq P_a^{ci} + P_b^{ci} + P_r^{ci}$ are the overall circuit PCs in the bidirectional transmission duration and the idle duration, respectively.

The required bidirectional data rates can be obtained from the capacity formula and the SNR expression of TWRT derived in [23], which are respectively,

$$
\frac{B_{ab}}{T_{\mathrm{TWR}}} = \frac{W}{2}\log_2\left(1 + \frac{P_a^t P_r^t |h_{ar}|^2 |h_{br}|^2}{|h_{ar}|^2 P_a^t N_0 + |h_{br}|^2 P_b^t N_0 + |h_{br}|^2 P_r^t N_0 + N_0^2}\right),
\tag{7}
$$

$$
\frac{B_{ba}}{T_{\mathrm{TWR}}} = \frac{W}{2}\log_2\left(1 + \frac{P_b^t P_r^t |h_{br}|^2 |h_{ar}|^2}{|h_{ar}|^2 P_a^t N_0 + |h_{br}|^2 P_b^t N_0 + |h_{ar}|^2 P_r^t N_0 + N_0^2}\right),
\tag{8}
$$

where the factor 1/2 is due to the two-phase transmission.

4 Energy efficiency optimization for three transmit strategies

In this section, we optimize the EEs for DT, OWRT, and TWRT. The EE is defined as the number of bits transmitted in two directions per unit of energy, i.e.,

$$
\eta_{\mathrm{EE}} = \frac{B_{ab} + B_{ba}}{E},
\tag{9}
$$

where E is the EC per block, which respectively equals to E_D, E_O or E_T in DT, OWRT, or TWRT.

To guarantee a fair comparison, we maximize the EEs of DT, OWRT, and TWRT with the same packet

sizes B_{ab} and B_{ba}. From the definition of η_{EE}, we see that EE maximization is equivalent to EC minimization for a given pair of B_{ab} and B_{ba}. Consequently, we will minimize the EC per block for different strategies by optimizing transmission time and power of each node.

We consider that the transmission time should not exceed the duration of a block T, and the transmit power of each node should be less than the maximum transmit power P_{max}^t. Note that the system may not be able to transmit B_{ab} and B_{ba} bits within the duration T even if the maximum transmit power is used. In this case an outage occurs. Since we assume perfect channel knowledge at each node, the nodes can estimate the transmit power and the transmission time required for each block, which depend on the channel distribution and packet sizes B_{ab} and B_{ba}. Once the channel statistics and the packet sizes are given, the outage probability is fixed. In practice, the packet sizes B_{ab} and B_{ba} can be pre-determined according to the quality of service (QoS) requirements, channel environment, and the acceptable outage probability. We will use Monte-Carlo simulation to find the maximal B_{ab} and B_{ba} that ensure the outage probability to be lower than a threshold, e.g., 10%. Then, we only need to consider the EE optimization when the packet sizes are smaller than the maximum B_{ab} and B_{ba}.

4.1 Direct transmission
As shown in (3), the EC of DT is a function of the transmit powers P_a^t and P_b^t as well as the transmission time T_{ab} and T_{ba}. The EC can be minimized by jointly optimizing the transmit powers and transmission time as follows,

$$\min_{T_{ab}, T_{ba}, P_a^t, P_b^t} T_{ab}\left(\frac{P_a^t}{\varepsilon} + P_D^{c1} - P_D^{ci}\right) + T_{ba}\left(\frac{P_b^t}{\varepsilon} + P_D^{c2} - P_D^{ci}\right) + TP_D^{ci} \quad (10)$$
$$\text{s.t.} \quad T_{ab} + T_{ba} \leq T, \ P_a^t \leq P_{max}^t, P_b^t \leq P_{max}^t.$$

To solve this joint optimization problem, we first express the transmit powers P_a^t and P_b^t as functions of the transmission time T_{ab} and T_{ba} by using (2), which are respectively,

$$P_a^t = \frac{N_0}{|h_{ab}|^2}\left(2^{\frac{B_{ab}}{WT_{ab}}} - 1\right), P_b^t = \frac{N_0}{|h_{ab}|^2}\left(2^{\frac{B_{ba}}{WT_{ba}}} - 1\right). \quad (11)$$

By substituting (11) into both the objective function and the constraints of (10), the problem (10) can be reformulated as follows,

$$\min_{T_{ab}, T_{ba}} T_{ab}\left[\frac{N_0\left(2^{\frac{B_{ab}}{WT_{ab}}} - 1\right)}{\varepsilon|h_{ab}|^2} + P_D^{c1} - P_D^{ci}\right] + T_{ba}\left[\frac{N_0\left(2^{\frac{B_{ab}}{WT_{ab}}} - 1\right)}{\varepsilon|h_{ab}|^2} + P_D^{c2} - P_D^{ci}\right] + TP_D^{ci} \quad (12)$$
$$\text{s.t.} \ T_{ab} + T_{ba} \leq T, T_{ab} \geq T_{min1}, T_{ab} \geq T_{min2}.$$

where

$$T_{min1} = \frac{B_{ab}}{W\log_2\left(1 + \frac{P_{max}^t|h_{ab}|^2}{N_0}\right)}, T_{min2} = \frac{B_{ba}}{W\log_2\left(1 + \frac{P_{max}^t|h_{ab}|^2}{N_0}\right)}. \quad (13)$$

The minimum value constraints on T_{ab} and T_{ba} are due to the transmit power constraints, without which the data rates B_{ab}/T_{ab} and B_{ba}/T_{ba} will be too high to be supported even with the maximal transmit powers.

Note that the problem in (12) is equivalent to the joint optimization problem in (10), where now only the transmission time needs to be optimized. In the objective function of the problem in (12), the first term is a function of T_{ab} and not related to T_{ba}. It is easy to show that its second order derivative with respect to T_{ab} is positive. Thus it is a convex function of T_{ab}. Similarly, the second term in the objective function is a convex function of T_{ba}. The last term is independent of the transmission time. Therefore, the objective function is convex with respect to T_{ab} and T_{ba}. All the constraints in (12) are also convex.[a] Then the problem can be solved by using efficient convex optimization techniques, such as gradient descent algorithm [24].

4.2 One-way relay transmission
Similar to the DT case, we first express the transmit powers as functions of the transmission time using (4) and (5). Then the joint optimization of transmit power and transmission time can be solved with two steps: first find the optimal transmit powers as functions of the transmission time, then optimize the transmission time to minimize the EC.

For a given T_{ab}, both P_a^t and P_b^t can be obtained from (4), where multiple feasible solutions exist. In order to minimize the EC, we find the transmit powers that minimize the sum power as follows,

$$\min_{P_a^t, P_{r1}^t} P_a^t + P_{r1}^t$$
$$\text{s.t.} \ P_a^t \leq P_{max}^t, P_{r1}^t \leq P_{max}^t, \ \text{and (4)}. \quad (14)$$

To ensure that all the constraints in (14) can be satisfied, the data rate B_{ab}/T_{ab} should be less than the maximum data rate supported by the maximum transmit power. This turns into a minimum value constraint for the transmit time, which is

$$T_{ab} \geq B_{ab} \bigg/ \left[\frac{W}{2} \log_2 \left(1 + \frac{(P^t_{\max})^2 |h_{ar}|^2 |h_{br}|^2}{|h_{ar}|^2 P^t_{\max} N_0 + |h_{br}|^2 P^t_{\max} N_0 + N_0^2} \right) \right] \triangleq T_{\min 1}. \quad (15)$$

Denote the minimum value of $P^t_a + P^t_{r1}$ as $P_{\min 1}(T_{ab})$, where $T_{ab} \geq T_{\min 1}$. It can be derived as a piecewise function as follows (see Appendix 1),

$$P_{\min 1}(T_{ab}) = \begin{cases} \frac{C_1 |h_{br}|^2 P^t_{\max} N_0 + C_1 N_0^2}{(|h_{ar}|^2 |h_{br}|^2 P^t_{\max} - C_1 |h_{ar}|^2 N_0)} + P^t_{\max}, & T_{\min 1} \leq T_{ab} < T_{d1} \\ C_1 N_0 \left(\frac{1}{|h_{br}|^2} + \frac{1}{|h_{ar}|^2} \right) + \frac{2\sqrt{C_1^2 + C_1 N_0}}{|h_{ar} h_{br}|}, & T_{ab} \geq T_{d1} \end{cases} \quad (16)$$

or,

$$P_{\min 1}(T_{ab}) = \begin{cases} P^t_{\max} + \frac{C_1 |h_{ar}|^2 P^t_{\max} N_0 + C_1 N_0^2}{(|h_{ar}|^2 |h_{br}|^2 P^t_{\max} - C_1 |h_{br}|^2 N_0)}, & T_{\min 1} \leq T_{ab} < T_{d2} \\ C_1 N_0 \left(\frac{1}{|h_{br}|^2} + \frac{1}{|h_{ar}|^2} \right) + \frac{2\sqrt{C_1^2 + C_1 N_0}}{|h_{ar} h_{br}|}, & T_{ab} \geq T_{d2} \end{cases} \quad (17)$$

where $C_1 \triangleq 2^{2B_{ab}/(T_{ab}W)} - 1$, the demarcation points T_{d1} and T_{d2} are defined in Appendix 1. If $T_{d1} \geq T_{d2}$, $P_{\min 1}(T_{ab})$ follows (16), otherwise, it follows (17).

The piecewise function can be explained as follows. When T_{ab} is large, the data rate is low and both P^t_a and P^t_{r1} are below their maximum value, then $P_{\min 1}(T_{ab})$ follows the second part in (16) or (17). As T_{ab} decreases, one of P^t_a and P^t_{r1} will achieve its maximum value. When $T_{ab} = T_{d1}$, we have $P^t_{r1} = P^t_{\max}$, and when $T_{ab} = T_{d2}$, $P^t_a = P^t_{\max}$. If $T_{d1} \geq T_{d2}$, P^t_{r1} achieves its maximum value first, $P_{\min 1}(T_{ab})$ follows the first part in (16). Otherwise, P^t_a achieves its maximum value first, $P_{\min 1}(T_{ab})$ follows the first part in (17). When T_{ab} decreases to $T_{\min 1}$, both P^t_a and P^t_{r1} achieve the maximum value. For simplicity, we refer the first part in (16) or (17) as "one-max" interval, because one of the nodes uses its maximum transmit power. We refer the second part in (16) or (17) as "non-max" interval, since neither of the nodes uses its maximum transmit power.

For a given T_{ba}, we can also find the values of P^t_b and P^t_{r2} that minimize their summation. Following an analogous procedure, the minimum value of $P^t_b + P^t_{r2}$ denoted as $P_{\min 2}(T_{ba})$ can be derived as a piecewise function of transmission time T_{ba}, which are respectively,

$$P_{\min 2}(T_{ba}) = \begin{cases} \frac{C_2 |h_{ar}|^2 P^t_{\max} N_0 + C_2 N_0^2}{(|h_{ar}|^2 |h_{br}|^2 P^t_{\max} - C_2 |h_{br}|^2 N_0)} + P^t_{\max}, & T_{\min 2} \leq T_{ba} < T_{d3} \\ C_2 N_0 \left(\frac{1}{|h_{br}|^2} + \frac{1}{|h_{ar}|^2} \right) + \frac{2\sqrt{C_2^2 + C_2 N_0}}{|h_{ar} h_{br}|}, & T_{ba} \geq T_{d3} \end{cases} \quad (18)$$

or,

$$P_{\min 2}(T_{ba}) = \begin{cases} P^t_{\max} + \frac{C_2 |h_{br}|^2 P^t_{\max} N_0 + C_2 N_0^2}{(|h_{ar}|^2 |h_{br}|^2 P^t_{\max} - C_2 |h_{ar}|^2 N_0)}, & T_{\min 2} \leq T_{ba} < T_{d4} \\ C_2 N_0 \left(\frac{1}{|h_{br}|^2} + \frac{1}{|h_{ar}|^2} \right) + \frac{2\sqrt{C_2^2 + C_2 N_0}}{|h_{ar} h_{br}|}, & T_{ba} \geq T_{d4} \end{cases} \quad (19)$$

where $C_2 \triangleq 2^{2B_{ba}/(T_{ba}W)} - 1$, the demarcation points T_{d3} and T_{d4} can be derived similarly as T_{d1} and T_{d2} in $P_{\min 1}(T_{ab})$. If $T_{d3} \geq T_{d4}$, $P_{\min 2}(T_{ba})$ follows (18), otherwise, it follows (19). The minimum value constraint for T_{ba}, i.e., $T_{ba} \geq T_{\min 2}$, is also due to the maximum transmit power constraint like that for T_{ab} in (15), and $T_{\min 2}$ can be derived similarly as $T_{\min 1}$.

Then the optimization problem that minimizes the EC can be formulated as follows,

$$\min_{T_{ab}, T_{ba}} \ T_{ab} \left(\frac{P_{\min 1}(T_{ab})}{2\varepsilon} + P^{c1}_O - P^{ci}_O \right) + T_{ba} \left(\frac{P_{\min 2}(T_{ba})}{2\varepsilon} + P^{c2}_O - P^{ci}_O \right) + T P^{ci}_O \quad (20)$$
$$\text{s.t.} \quad T_{ab} + T_{ba} \leq T, T_{ab} \geq T_{\min 1}, T_{ba} \geq T_{\min 2}.$$

We can show that the first term in the objective function is a quasi-convex function of T_{ab} (see Appendix 2). Similarly, the second term is a quasi-convex function of T_{ba}. The last term is a constant. However, the sum of two quasi-convex functions may not be quasi-convex. Therefore, we solve this problem using the following approach.

First, we assume that the optimal solution for (20) satisfies $T^{\text{opt}}_{ab} + T^{\text{opt}}_{ba} < T$. In this case, the first constraint in (20) can be omitted. Since the second constraint is only related to T_{ab}, and the last constraint is only related to T_{ba}, the joint optimization problem can be decoupled into two subproblems, i.e., optimizing T_{ab} to minimize the first term in objective function with the constraint $T_{ab} \geq T_{\min 1}$, and optimizing T_{ba} to minimize the second term in objective function with the constraint $T_{ba} \geq T_{\min 2}$. Because we have proved that the first two terms in the objective function are, respectively, quasi-convex functions with respect to T_{ab} and T_{ba}, both the two subproblems can be solved via quasi-convex optimization techniques such as bisection algorithm [24].

If the optimized T_{ab} and T_{ba} from the two subproblems satisfy $T^{\text{opt}}_{ab} + T^{\text{opt}}_{ba} < T$, then our assumption holds, and we obtain the optimal transmission time. Otherwise, the optimal solution for (20) must satisfy $T^{\text{opt}}_{ab} + T^{\text{opt}}_{ba} = T$. In this case, we only need to find the optimal T^{opt}_{ab}, where a scalar searching is applied, and the optimal T^{opt}_{ba} can be obtained as $T^{\text{opt}}_{ba} = T - T^{\text{opt}}_{ab}$.

4.3 Two-way relay transmission

Analogous to the previous sections, we first derive the transmit powers as functions of the transmission time.

For a given T_{TWR}, we can find P^t_a, P^t_b, and P^t_r from (7) and (8), where multiple feasible solutions exist. To minimize the EC, again we find P^t_a, P^t_b, and P^t_r that minimize their summation from the following problem,

$$\min_{P^t_a, P^t_b, P^t_r} \ P^t_a + P^t_b + P^t_r$$
$$\text{s.t.} \quad P^t_a \leq P^t_{\max}, P^t_b \leq P^t_{\max}, P^t_r \leq P^t_{\max}, (7) \text{ and } (8). \quad (21)$$

Following a similar derivation as in the case of OWRT, the minimum value of $P_a^t + P_b^t + P_r^t$ can be obtained as a piecewise function of the transmission time T_{TWR}, which is denoted as $P_{\min}(T_{\mathrm{TWR}})$.

When T_{TWR} is large, the data rates B_{ab}/T_{TWR} and B_{ba}/T_{TWR} are low, and all transmit powers are below their maximum values. The optimal transmit powers are derived with similar method in Appendix 1 as follows,

$$P_a^{t-\mathrm{opt}} = \frac{C_1 N_0}{|h_{ar}|^2} + \frac{N_0(C_1^2 + C_1 + C_1 C_2)}{|h_{ar}h_{br}|\sqrt{(C_1 + C_2)(C_1 + C_2 + 1)}}, \quad (22a)$$

$$P_b^{t-\mathrm{opt}} = \frac{C_2 N_0}{|h_{br}|^2} + \frac{N_0(C_2^2 + C_2 + C_1 C_2)}{|h_{ar}h_{br}|\sqrt{(C_1 + C_2)(C_1 + C_2 + 1)}}, \quad (22b)$$

$$P_r^{t-\mathrm{opt}} = \frac{C_1 N_0}{|h_{br}|^2} + \frac{C_2 N_0}{|h_{ar}|^2} + \frac{N_0\sqrt{(C_1 + C_2)(C_1 + C_2 + 1)}}{|h_{ar}h_{br}|}. \quad (22c)$$

where $C_1 \triangleq 2^{\frac{2B_{ab}}{WT_{\mathrm{TWR}}}} - 1$ and $C_2 \triangleq 2^{\frac{2B_{ba}}{WT_{\mathrm{TWR}}}} - 1$. The corresponding $P_{\min}(T_{\mathrm{TWR}})$ is the sum of (22a), (22b), and (22c).

When T_{TWR} decreases, the data rates increases, then $P_a^{t-\mathrm{opt}}, P_b^{t-\mathrm{opt}}$, and $P_r^{t-\mathrm{opt}}$ increase until one of them achieves the maximum value $P_{\max}^t$. By setting (22a), (22b), and (22c) to be $P_{\max}^t$, respectively, we can obtain $T_{\mathrm{TWR}} = T_{d1}$ when $P_a^{t-\mathrm{opt}} = P_{\max}^t$, $T_{\mathrm{TWR}} = T_{d2}$ when $P_b^{t-\mathrm{opt}} = P_{\max}^t$, and $T_{\mathrm{TWR}} = T_{d3}$ when $P_r^{t-\mathrm{opt}} = P_{\max}^t$. Without loss of generality, we assume that $T_{d1} \geq T_{d2}$ and $T_{d1} \geq T_{d3}$ (similar results can be obtained for other cases). In this case, $P_a^{t-\mathrm{opt}}$ achieves the maximum value first, i.e., node $\mathbb{A}$ transmits with the maximum transmit power. By substituting $P_a^t = P_{\max}^t$ into (7) and (8), we have

$$P_a^{t-\mathrm{opt}} = P_{\max}^t, \quad (23a)$$

$$P_b^{t-\mathrm{opt}} = \frac{C_1 C_2 N_0^2(|h_{ar}|^2 - |h_{br}|^2) + C_2|h_{ar}|^2|h_{br}|^2 P_{\max}^t N_0}{C_1|h_{ar}|^2|h_{br}|^2 P_{\max}^t N_0}, \quad (23b)$$

$$P_r^{t-\mathrm{opt}} = \frac{C_1 C_2 N_0^2(|h_{ar}|^2 - |h_{br}|^2) + P_{\max}^t N_0|h_{ar}|^2(C_1|h_{ar}|^2 + C_2|h_{br}|^2) + C_1|h_{ar}|^2 N_0^2}{|h_{ar}|^2|h_{br}|^2(|h_{ar}|^2 P_{\max}^t - C_1 N_0)}. \quad (23c)$$

The corresponding $P_{\min}(T_{\mathrm{TWR}})$ can be obtained by adding (23a), (23b), and (23c).

When T_{TWR} further decreases, the data rates further increases, $P_b^{t-\mathrm{opt}}$ and $P_r^{t-\mathrm{opt}}$ in (23) increase until one of them achieves its maximum value. Without loss of generality, assume that $P_b^{t-\mathrm{opt}}$ in (23b) achieves $P_{\max}^t$ first. The corresponding value of T_{TWR} is denoted as $T_{\min}$,

which can be obtained by setting (23b) to be $P_{\max}^t$. Then both nodes $\mathbb{A}$ and $\mathbb{B}$ transmit with the maximum power. Substituting $P_a^t = P_b^t = P_{\max}^t$ into (7) and (8), we need to find one P_r^t from two equations, which has no solution. Therefore, $T_{\min}$ is the minimum value of T_{TWR} due to the maximum transmit power constraint. Finally, the minimal sum transmit power is obtained as

$$P_{\min}(T_{\mathrm{TWR}}) = \begin{cases} (23a) + (23b) + (23c), T_{\min} \leq T_{\mathrm{TWR}} < T_{d1} \\ (22a) + (22b) + (22c), T_{\mathrm{TWR}} \geq T_{d1}, \end{cases} \quad (24)$$

where its first and second parts are, respectively, referred to as "one-max" and "non-max" interval for simplicity as that in the case of OWRT.

Then the optimization problem that minimizes the EC can be formulated as

$$\min_{T_{\mathrm{TWR}}} T_{\mathrm{TWR}} \left(\frac{P_{\min}(T_{\mathrm{TWR}})}{2\varepsilon} + P_T^c - P_T^{ci} \right) + TP_T^{ci} \quad (25)$$

$$\text{s.t. } T_{\min} \leq T_{\mathrm{TWR}} \leq T.$$

Using the similar method in Appendix 2, we can prove that the objective function is a quasi-convex function of T_{TWR}. Therefore, efficient quasi-convex optimization techniques [24] can be applied to solve the problem.

5 Energy efficiency analysis

In this section, we compare the EEs of different transmit strategies, and analyze the impact of various channels and system settings.

From the objective functions in (20) and (25), we can see that the expressions of the ECs of OWRT and TWRT are quite complex because the minimal sum transmit powers are piecewise functions with very complicated expressions, i.e., (16), (17), (18), (19), and (24). To gain useful insight into the EE comparison, we consider the following two approximations.

Approximation 1: In the piecewise functions of $P_{\min 1}(T_{ab}), P_{\min 2}(T_{ba})$, and $P_{\min}(T_{\mathrm{TWR}})$, we only consider the "non-max" interval, where none of the nodes achieves its maximum transmit power.

We take the function $P_{\min 1}(T_{ab})$ in (16) as an example to explain the approximation. In the "non-max" interval, as transmission time T_{ab} decreases, both transmit powers at nodes $\mathbb{A}$ and $\mathbb{B}$, i.e., P_a^t and P_{r1}^t, increase for supporting the increased data rate B_{ab}/T_{ab}. In the "one-max" interval, P_{r1}^t has achieved its maximum value. As T_{ab} decreases, only P_a^t can increase to support the increased data rate, thus P_a^t grows much faster than that in "non-max" interval and approaches its maximum value rapidly. Therefore, the range $(T_{\min 1}, T_{d1})$ of the "one-max" interval is very short, and in most cases the optimized $T_{ab}^{\mathrm{opt}} \notin (T_{\min 1}, T_{d1})$. Instead,

$T_{ab}^{\text{opt}} \in (T_{d1}, +\infty)$. Based on this observation, we only consider the "non-max" interval in range $(T_{d1}, +\infty)$.

Since we only consider the case where none of the nodes achieve its maximal transmit power, we do not need to consider the maximum transmit power constraints. Therefore it is not necessary to consider the corresponding minimum value constraints on the transmission time in this section.

Approximation 2: In the expressions of $P_{\text{min1}}(T_{ab})$, $P_{\text{min2}}(T_{ba})$, *and* $P_{\text{min}}(T_{\text{TWR}})$, *we respectively consider that*

$$\frac{2B_{ab}}{2^{\frac{2B_{ab}}{WT_{ab}}} - 1} \approx 2^{\frac{2B_{ab}}{WT_{ab}}}, \frac{2B_{ba}}{2^{\frac{2B_{ba}}{WT_{ba}}} - 1} \approx 2^{\frac{2B_{ba}}{WT_{ba}}}, \quad (26a)$$

$$2^{\frac{2B_{ab}}{WT_{\text{TWR}}}} + 2^{\frac{2B_{ba}}{WT_{\text{TWR}}}} - 2 \approx 2^{\frac{2B_{ab}}{WT_{\text{TWR}}}} + 2^{\frac{2B_{ab}}{WT_{\text{TWR}}}} - 1. \quad (26b)$$

We take (26a) as an example to explain the approximation, which affects the values of the transmit power $P_{\text{min1}}(T_{ab})$ and $P_{\text{min2}}(T_{ba})$ in OWRT. When the SEs in two directions, i.e., $B_{ab}/(WT_{ab})$ and $B_{ba}/(WT_{ba})$ are high, it is easy to see that the approximations in (26a) are accurate. On the other hand, when the SEs are low, the transmit powers $P_{\text{min1}}(T_{ab})$ and $P_{\text{min2}}(T_{ba})$ are much lower than the circuit PC. Then the approximations on transmit powers have little impact on the analysis of EC.

By applying these approximations, the ECs of OWRT and TWRT can be simplified as

$$E_O \approx T_{ab} \left[\frac{N_0}{2\varepsilon|h_{\text{eff}}|^2} \left(2^{\frac{2B_{ab}}{WT_{ab}}} - 1 \right) + P_O^{c1} - P_O^{ci} \right]$$
$$+ T_{ba} \left[\frac{N_0}{2\varepsilon|h_{\text{eff}}|^2} \left(2^{\frac{2B_{ba}}{WT_{ba}}} - 1 \right) + P_O^{c2} - P_O^{ci} \right] + TP_O^{ci}, \quad (27)$$

$$E_T \approx T_{\text{TWR}} \left[\frac{N_0}{2\varepsilon|h_{\text{eff}}|^2} \left(2^{\frac{2B_{ab}}{WT_{\text{TWR}}}} + 2^{\frac{2B_{ab}}{WT_{\text{TWR}}}} - 2 \right) + P_T^c - P_T^{ci} \right] + TP_T^{ci}, \quad (28)$$

where $|h_{\text{eff}}| \triangleq 1/\left(\frac{1}{|h_{ar}|} + \frac{1}{|h_{br}|} \right)$ can be viewed as an equivalent channel gain between two source nodes due to the usage of the relay.

For the convenience of comparison, we rewrite the EC of DT in the same form as follows,

$$E_D = T_{ab} \left[\frac{N_0}{\varepsilon|h_{ab}|^2} \left(2^{\frac{B_{ab}}{WT_{ab}}} - 1 \right) + P_D^{c1} - P_D^{ci} \right]$$
$$+ T_{ba} \left[\frac{N_0}{\varepsilon|h_{ab}|^2} \left(2^{\frac{B_{ba}}{WT_{ba}}} - 1 \right) + P_D^{c2} - P_D^{ci} \right] + TP_D^{ci}. \quad (29)$$

5.1 Baseline case

As a baseline for further analysis, we first consider the case where all the circuit PCs are zero and the packet sizes in two directions are symmetric, i.e., $P^{ct} = P^{cr} = P^{ci} = 0$ and $B_{ab} = B_{ba} \triangleq B$. Then the ECs of OWRT, TWRT, and DT shown in (27), (28), and (29) are decreasing functions of the transmission time. As a result, the system will use the entire duration T for transmission. Due to the symmetric packet sizes, the optimal values of T_{ab} and T_{ba} are identical in DT and OWRT. This means that the optimal transmission time in DT and OWRT are $T_{ab}^{\text{opt}} = T_{ba}^{\text{opt}} = T/2$, and that in TWRT is $T_{\text{TWR}}^{\text{opt}} = T$. After substituting the optimal transmission time into (27), (28), and (29), the minimum ECs can be obtained as

$$E_D^{\text{min}} = \frac{N_0 T}{\varepsilon} \frac{\left(2^{\frac{2B}{WT}} - 1 \right)}{|h_{ab}|^2}, E_O^{\text{min}} \approx \frac{N_0 T}{\varepsilon} \frac{\left(2^{\frac{4B}{WT}} - 1 \right)}{2|h_{\text{eff}}|^2}, E_T^{\text{min}} \approx \frac{N_0 T}{\varepsilon} \frac{\left(2^{\frac{2B}{WT}} - 1 \right)}{|h_{\text{eff}}|^2}, \quad (30)$$

from which we can see that the optimal EE, $\eta_{\text{EE}}^{\text{opt}} = \frac{2B}{E^{\text{min}}}$, is a decreasing function of the packet size B in the three strategies. This implies that the maximal EE is achieved when B approaches zero.

Now, we compare the EEs of the three strategies. First, it shows from (30) that $E_O^{\text{min}}/E_T^{\text{min}} \geq 1$, which means that TWRT is more energy efficient than OWRT.

Second, we see that $E_D^{\text{min}}/E_T^{\text{min}} = |h_{\text{eff}}|^2/|h_{ab}|^2$, i.e., the EE comparison between TWRT and DT depends on the effective channel gain $|h_{\text{eff}}|$ and the direct link channel gain $|h_{ab}|$. If $|h_{\text{eff}}| > |h_{ab}|$, TWRT is more energy efficient, otherwise, DT is more energy efficient. To gain further insight into this comparison, we consider an AWGN channel,[b] where $|h_{ab}|^2$ is normalized as 1, the distance from the RN to nodes $\mathbb{A}$ and $\mathbb{B}$ are, respectively, d and $1 - d$. Then $|h_{ar}|^2 = \left(\frac{1}{d} \right)^\alpha$ and $|h_{br}|^2 = \left(\frac{1}{1-d} \right)^\alpha$, where α is the path loss attenuation factor. Then the equivalent channel gain becomes

$$|h_{\text{eff}}| = 1/\left(\frac{1}{|h_{ar}|} + \frac{1}{|h_{br}|} \right) = \frac{1}{d^{\alpha/2} + (1-d)^{\alpha/2}}, \quad (31)$$

which is related to the RN position. To maximize $|h_{\text{eff}}|$, the optimal relay position is the midpoint of the two source nodes, i.e., $d = 0.5$. In this case, $|h_{\text{eff}}| = 2^{\alpha/2}/2$. When $\alpha > 2$, which is true in most practical channel environments, $|h_{\text{eff}}| = 2^{\alpha/2}/2 > |h_{ab}| = 1$, and TWRT is more energy efficient than DT.

Third, for DT and OWRT we have

$$E_D^{\min}/E_O^{\min} = \frac{|h_{\text{eff}}|^2}{|h_{ab}|^2} \frac{2(2^{\frac{2B}{WT}} - 1)}{2^{\frac{4B}{WT}} - 1} = \frac{|h_{\text{eff}}|^2}{|h_{ab}|^2} \frac{2}{2^{\frac{2B}{WT}} + 1}. \quad (32)$$

If $|h_{\text{eff}}| \leq |h_{ab}|$, since $\dfrac{2}{2^{\frac{2B}{WT}} + 1} \leq 1$ we have $E_D^{\min}/E_O^{\min} \leq 1$, i.e., DT is more energy efficient than OWRT.

If $|h_{\text{eff}}| > |h_{ab}|$, the comparison result depends on the packet size B. When $B \to 0, \dfrac{2}{2^{\frac{2B}{WT}} + 1} \to 1$, then $E_D^{\min}/E_O^{\min} \to |h_{\text{eff}}|^2/|h_{ab}|^2 \geq 1$. It means that in low traffic region, OWRT is more energy efficient. When $B \to \infty, \dfrac{2}{2^{\frac{2B}{WT}} + 1} \to 0$, then $E_D^{\min}/E_O^{\min} \to 0 < 1$. It means that in high traffic region, DT is more energy efficient. An intuitive explanation is as follows. On one hand, OWRT needs two-phase for transmission in each direction, thus the data rate in each phase should be twice of that in DT, which requires more transmit power. On the other hand, OWRT has higher equivalent channel gain, which reduces the required transmit power. In low traffic region, doubling the lower data rate has little impact on the transmit power, and thus OWRT is more energy efficient due to higher equivalent channel gain.

Here we argue that even if OWRT exploits the direct link between $\mathbb{A}$ and $\mathbb{B}$ for spatial diversity, the conclusion will still be the same. With the direct link, the equivalent channel gain can be improved. However, the improvement is rather limited in most cases, because the signal attenuation between the two source nodes is much larger than that between the source nodes and the RN. Furthermore, OWRT has 1/2 spectral efficiency loss with respect to DT and TWRT, which cannot be recovered from the SNR gain.

5.2 Impact of circuit power consumption

In this subsection we assume symmetric packet size, i.e., $B_{ab} = B_{ba} = B$, but consider the non-zero circuit PCs in practical systems. Then the ECs in (27), (28), and (29) are no longer monotonically decreasing functions of the transmission time. With the increase of the transmission time, the transmit energy decreases since the required data rate reduces, however, the circuit energy increases linearly. We take TWRT as an example to analyze the EE.

The optimal transmission time in TWRT can be obtained by taking the derivative of E_T in (28) with respect to T_{TWR} and setting it to be zero, which is

$$\frac{dE_T}{dT_{\text{TWR}}} \approx \frac{d}{dT_{\text{TWR}}} \left\{ T_{\text{TWR}} \left[\frac{N_0}{\varepsilon |h_{\text{eff}}|^2} \left(2^{\frac{2B}{WT_{\text{TWR}}}} - 1 \right) + P_T^c - P_T^{ci} \right] + T P_T^{ci} \right\} \quad (33a)$$

$$= \left[\frac{N_0 \left(2^{\frac{2B}{WT_{\text{TWR}}}} - 1 \right)}{\varepsilon |h_{\text{eff}}|^2} + P_T^c - P_T^{ci} \right] - \frac{N_0 \ln 2}{\varepsilon |h_{\text{eff}}|^2} 2^{\frac{2B}{WT_{\text{TWR}}}} \frac{2B}{WT_{\text{TWR}}} \quad (33b)$$

$$\triangleq \left[\frac{N_0(2^{\eta_{\text{SE-T}}} - 1)}{\varepsilon |h_{\text{eff}}|^2} + P_T^c - P_T^{ci} \right] - \frac{N_0 \ln 2}{\varepsilon |h_{\text{eff}}|^2} 2^{\eta_{\text{SE-T}}} \eta_{\text{SE-T}} = 0 \Big|_{\eta_{\text{SE-T}} = \eta_{\text{SE-T}}^{\text{opt}}}, \quad (33c)$$

where $\eta_{\text{SE-T}} \triangleq \dfrac{2B}{WT_{\text{TWR}}}$ is the bidirectional SE of TWRT.

Although it is difficult to obtain a closed form solution of the optimal T_{TWR}, some observations can be obtained from (33). The optimal SE that minimizes the EC should satisfy (33c), from which we can see that $\eta_{\text{EE-T}}^{\text{opt}}$ does not depend on the packet size B. Therefore, the optimal transmission time $T_{\text{TWR}}^{\text{opt}} = \dfrac{2B}{W \eta_{\text{SE-T}}^{\text{opt}}}$ increases linearly with B. Considering that T_{TWR} should not exceed the time duration of a block T, we obtain the following observation.

Observation 1: In high traffic region, $T_{\text{TWR}}^{\text{opt}} = T$. In low traffic region where $\dfrac{2B}{W \eta_{\text{SE-T}}^{\text{opt}}} \leq T$, the optimal transmission time $T_{\text{TWR}}^{\text{opt}} = \dfrac{2B}{W \eta_{\text{SE-T}}^{\text{opt}}}$ increases linearly with the packet size B.

In high traffic region, the transmission time $T_{\text{TWR}}^{\text{opt}} = T$, then the bidirectional SE $\dfrac{2B}{WT}$ increases linearly with the packet size B, thus the transmit energy increases exponentially with B according to the capacity formula. In this case, the transmit EC is much larger than the circuit EC, thus the EE will be almost the same as that in zero circuit PC scenario.

In low traffic region, when the system transmits with the optimal transmission time $T_{\text{TWR}}^{\text{opt}} = \dfrac{2B}{W \eta_{\text{SE-T}}^{\text{opt}}}$, the equality in (33b) equals to zero. Then we have

$$T_{\text{TWR}}^{\text{opt}}\left[\frac{N_0(2^{\frac{2B}{WT_{\text{TWR}}^{\text{opt}}}}-1)}{\varepsilon|h_{\text{eff}}|^2}+P_T^c-P_T^{ci}\right]=\frac{2BN_0}{\varepsilon|h_{\text{eff}}|^2W}(\ln 2)2^{\frac{2B}{WT_{\text{TWR}}^{\text{opt}}}}$$

$$=\frac{2BN_0}{\varepsilon|h_{\text{eff}}|^2W}(\ln 2)2^{\eta_{\text{SE-T}}^{\text{opt}}},$$

where the first equality comes from the fact that (33b) equals to zero, and the second equality comes from $T_{\text{TWR}}^{\text{opt}}=\frac{2B}{W\eta_{\text{SE-T}}^{\text{opt}}}$.

By substituting $B_{ab}=B_{ba}=B$ and $T_{\text{TWR}}=T_{\text{TWR}}^{\text{opt}}$ into the EC of TWRT in (28), and then substituting (34), the minimum EC of TWRT can be obtained as

$$E_T^{\min}=\frac{2BN_0}{\varepsilon|h_{\text{eff}}|^2W}(\ln 2)2^{\eta_{\text{SE-T}}^{\text{opt}}}+TP_T^{ci}, \quad (34)$$

and the optimal EE of TWRT is given by

$$\eta_{\text{EE-T}}^{\text{opt}}=\frac{2B}{\frac{2BN_0}{\varepsilon|h_{\text{eff}}|^2W}(\ln 2)2^{\eta_{\text{SE-T}}^{\text{opt}}}+TP_T^{ci}}, \quad (35)$$

from which we can obtain the following observation.

Observation 2: In low traffic region, if the circuit PC in idle mode $P_T^{ci}=0$, we have $\eta_{\text{EE-T}}^{\text{opt}}=\frac{\varepsilon|h_{\text{eff}}|^2W}{N_0(\ln 2)2^{\eta_{\text{SE-T}}^{\text{opt}}}}$. Since we have shown that $\eta_{\text{SE-T}}^{\text{opt}}$ does not depend on the packet size B, $\eta_{\text{EE-T}}^{\text{opt}}$ also does not change with B in this case. If $P_T^{ci}\neq 0$, $\lim_{B\to 0}\eta_{\text{EE-T}}^{\text{opt}}=0$ since a large portion of energy is consumed in the idle duration.

Note that although $\lim_{B\to 0}\eta_{\text{EE-T}}^{\text{opt}}=0$ due to the non-zero idle mode circuit PC, this observation does not mean that the idle duration is unnecessary. If the system transmits with the entire duration T, where $T>T_{\text{TWR}}^{\text{opt}}$, it can save the EC in idle mode, but it wastes more EC in transmission mode because it does not transmit with the optimal transmission time. Finally, more energy will be consumed and the EE will be reduced. We will show this impact later in simulations.

Observation 2 shows that if $P_T^{ci}=0$, $\eta_{\text{EE-T}}^{\text{opt}}$ does not change with B in low traffic region, where $\frac{2B}{W\eta_{\text{SE-T}}^{\text{opt}}}\leq T$, i.e., $B\leq TW\eta_{\text{SE-T}}^{\text{opt}}/2$. In other words, EE is insensitive to the packet size when $B\in(0,TW\eta_{\text{SE-T}}^{\text{opt}}/2)$. We can show that such a region becomes wider as the circuit power

P_T^c increases. By taking derivative with respect to P_T^c at both side of (33c), we obtain

$$1-\frac{N_0(\ln 2)^2}{\varepsilon|h_{\text{eff}}|^2}2^{\eta_{\text{SE-T}}^{\text{opt}}}\eta_{\text{SE-T}}^{\text{opt}}\frac{d\eta_{\text{SE-T}}^{\text{opt}}}{dP_T^c}=0, \quad (36)$$

from which we can see that $\frac{d\eta_{\text{SE-T}}^{\text{opt}}}{dP_T^c}=\frac{\varepsilon|h_{\text{eff}}|^2}{N_0(\ln 2)^2 2^{\eta_{\text{SE-T}}^{\text{opt}}}\eta_{\text{SE-T}}^{\text{opt}}}\geq 0$, i.e., as the circuit power P_T^c increases, $\eta_{\text{SE-T}}^{\text{opt}}$ increases, and then the region $(0,TW\eta_{\text{SE-T}}^{\text{opt}}/2)$ extends.

Following analogous procedure, we can obtain the same observations as in the Observations 1 and 2 for DT and OWRT. The optimal EEs of DT and OWRT in low traffic region can be obtained as

$$\eta_{\text{EE-D}}^{\text{opt}}=\frac{2B}{\frac{BN_0}{\varepsilon|h_{ab}|^2W}(\ln 2)(2^{\eta_{\text{SE-D1}}^{\text{opt}}}+2^{\eta_{\text{SE-D2}}^{\text{opt}}})+TP_D^{ci}}, \quad (37)$$

$$\eta_{\text{EE-O}}^{\text{opt}}=\frac{2B}{\frac{BN_0}{\varepsilon|h_{\text{eff}}|^2W}(\ln 2)(2^{\eta_{\text{SE-O1}}^{\text{opt}}}+2^{2\eta_{\text{SE-O2}}^{\text{opt}}})+TP_O^{ci}}, \quad (38)$$

where $\eta_{\text{SE-D1}}^{\text{opt}}$ and $\eta_{\text{SE-D2}}^{\text{opt}}$ are the optimal SEs in $\mathbb{A}\to\mathbb{B}$ and $\mathbb{B}\to\mathbb{A}$ directions in DT, $\eta_{\text{SE-O1}}^{\text{opt}}$ and $\eta_{\text{SE-O2}}^{\text{opt}}$ are those in OWRT, none of them depends on the packet size B. We omit the detailed derivations for concise.

Since it is difficult to derive closed form expressions for the optimal transmission time and the optimal SEs, there are also no closed form expressions for the optimal EEs. We will use simulations to compare the EEs of DT, OWRT, and TWRT under non-zero circuit PCs.

5.3 Impact of unequal data amounts in two directions

In this section, we assume that the circuit PCs are identical at each node, and consider that the packet sizes in two directions differ. Define $B_{ab}=\beta B_s$ and $B_{ba}=(1-\beta)B_s$, where B_s is the overall number of bits to be transmitted in two directions, and β is a factor to reflect the traffic asymmetry. We will show that once B_s is given, the minimum ECs of DT and OWRT are independent of β, but the EC of TWRT is minimized when $\beta=0.5$. In other words, the asymmetric packet sizes in two directions only reduces the EE of TWRT.

Proposition 1. *The minimum EC of OWRT does not depend on β.*

Proof. Since, we assume $P_O^{c1} = P_O^{c2} \triangleq P_O^c$, the EC of OWRT in (27) can be rewritten as

$$E_O = T_{ab} \left[\frac{N_0 \left(2^{\frac{2\beta B_s}{WT_{ab}}} - 1 \right)}{2\varepsilon |h_{\text{eff}}|^2} + P_O^c - P_O^{ci} \right]$$

$$+ T_{ba} \left[\frac{N_0 \left(2^{\frac{2(1-\beta)B_s}{WT_{ba}}} - 1 \right)}{2\varepsilon |h_{\text{eff}}|^2} + P_O^c - P_O^{ci} \right] + T P_O^{ci}. \tag{39}$$

To minimize the EC, the optimal transmit time should satisfy that (see Appendix 3),

$$\frac{\beta B_s}{T_{ab}^{\text{opt}}} = \frac{(1-\beta)B_s}{T_{ba}^{\text{opt}}} \triangleq R_O, \tag{40}$$

i.e., the data rates on the two directions are identical, where R_O is not a function of β. Then the minimum E_O can be obtained as follows by substituting (40) into (39),

$$E_O^{\min} = \frac{\beta B_s}{R_O} \left[\frac{N_0 \left(2^{\frac{2R_O}{W}} - 1 \right)}{2\varepsilon |h_{\text{eff}}|^2} + P_O^c - P_O^{ci} \right] + \frac{(1-\beta)B_s}{R_O} \left[\frac{N_0 \left(2^{\frac{2R_O}{W}} - 1 \right)}{2\varepsilon |h_{\text{eff}}|^2} + P_O^c - P_O^{ci} \right] + T P_O^{ci}$$

$$= \frac{B_s}{R_O} \left[\frac{N_0 \left(2^{\frac{2R_O}{W}} - 1 \right)}{2\varepsilon |h_{\text{eff}}|^2} + P_O^c - P_O^{ci} \right] + T P_O^{ci}, \tag{41}$$

which is not a function of β.

This proposition is easy to understand intuitively. Because with the optimized transmission time, the OWRT system transmits with the same data rate on each direction, and each bit is transmitted with identical data rate R_O and thus with identical time duration $1/R_O$. Therefore, the energy consumed by each bit is identical no matter in which direction it is transmitted. Then the minimum EC only depends on the overall number of transmitted bits B_s.

The minimum EC of DT, $E_D^{\min}$, can be obtained in a similar way, which also does not depend on β. We do not show the results for concise.

Proposition 2. *The minimum EC of TWRT is a function of β, and its minimum value is achieved when β = 0.5.*

Proof. The EC of TWRT in (28) can be rewritten as,

$$E_T = T_{\text{TWR}} \left[\frac{N_0 \left(2^{\frac{2\beta B_s}{WT_{\text{TWR}}}} - 1 \right)}{2\varepsilon |h_{\text{eff}}|^2} + \frac{P_T^c - P_T^{ci}}{2} \right]$$

$$+ T_{\text{TWR}} \left[\frac{N_0 \left(2^{\frac{2(1-\beta)B_s}{WT_{\text{TWR}}}} - 1 \right)}{2\varepsilon |h_{\text{eff}}|^2} + \frac{P_T^c - P_T^{ci}}{2} \right] + T P_T^{ci}. \tag{42}$$

If the transmission time in two directions could be different,[c] the EC becomes

$$E_{T1} = T_{\text{TWR1}} \left[\frac{N_0 \left(2^{\frac{2\beta B_s}{WT_{\text{TWR1}}}} - 1 \right)}{2\varepsilon |h_{\text{eff}}|^2} + \frac{P_T^c - P_T^{ci}}{2} \right]$$

$$+ T_{\text{TWR2}} \left[\frac{N_0 \left(2^{\frac{2(1-\beta)B_s}{WT_{\text{TWR2}}}} - 1 \right)}{2\varepsilon |h_{\text{eff}}|^2} + \frac{P_T^c - P_T^{ci}}{2} \right] + T P_T^{ci}. \tag{43}$$

Note that the only difference of E_T and E_{T1} is the transmission time in their first and second terms. With less constraints on the transmission time, the minimum value of E_{T1} achieved by optimizing T_{TWR1} and T_{TWR2} is a lower bound of the minimum value of E_T by optimizing T_{TWR}, i.e., $E_T^{\min} = \min_{T_{\text{TWR}}}(E_T) \geq \min_{T_{\text{TWR1}}, T_{\text{TWR2}}}(E_{T1}) = E_{T1}^{\min}$.

Following the analogous procedure as we analyze the OWRT system, we can show that $E_{T1}^{\min}$ is not a function of β. Moreover, using similar method as in Appendix 3, we can prove that the optimal T_{TWR1} and T_{TWR2} that minimize (43) satisfy $\frac{\beta B_s}{T_{\text{TWR1}}^{\text{opt}}} = \frac{(1-\beta)B_s}{T_{\text{TWR2}}^{\text{opt}}}$. It suggests that only when $\beta = 0.5$, $T_{\text{TWR1}}^{\text{opt}} = T_{\text{TWR2}}^{\text{opt}}$. In this case, by choosing $T_{\text{TWR}} = T_{\text{TWR1}}^{\text{opt}} = T_{\text{TWR2}}^{\text{opt}}$, E_T in (42) equals to $E_{T1}^{\min}$. Therefore, only when $\beta = 0.5$, $E_{T1}^{\min}$ equals to its lower bound $E_{T1}^{\min}$. Then proposition 2 is true.

6 Simulation results

In this section, we evaluate the EEs of the three transmission strategies, DT, OWRT, and TWRT, and validate previous analysis via simulations.

Simulation parameter settings are summarized in Table 1, where we consider that three nodes are located on a straight line, and the RN is at the midpoint of two source nodes. In this case, the equivalent channel gain in relaying achieves the maximal value. The small scale fading channels are independent and identically distributed (i.i.d.) Rayleigh block fading, which remain constant during one block but are independent from one block to another. All the results are averaged over 500 channel realizations.

The increase of distance D, noise power N_0, and attenuation factor α all result in higher required transmit power. Since their impacts are similar, we only show the impact of α. Because the increase of block duration T is equivalent to a reduction of the transmitted bits number per unit of time, we set T as a constant and change the values of B_{ab} and B_{ba}.

From [6,21], the circuit PCs in practical systems usually range from dozens to hundreds of mW. Therefore, we set the circuit PCs in this range in the simulations. The power amplifier efficiency e is set as 0.35 [21].

6.1 Baseline case

We first compare the EEs of different strategies in the baseline case where the circuit PCs are zero and the packet sizes $B_{ab} = B_{ba}$.

To show the EEs in different channel conditions, we set the attenuation factor α as 2 or 4. Since we are more interested in comparing the EEs rather than showing their absolute values, we normalize the EEs by the maximum EE of DT system for each α. The normalized EE is shown in Figure 3, and the corresponding outage
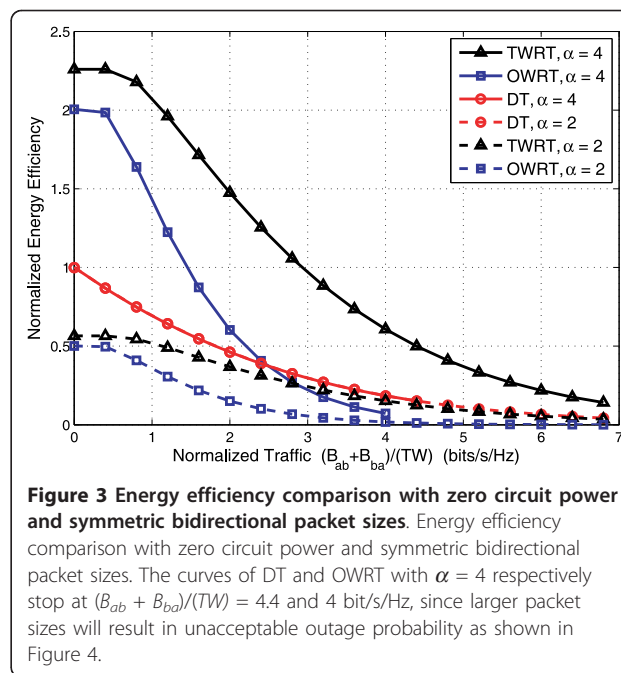
Figure 3 Energy efficiency comparison with zero circuit power and symmetric bidirectional packet sizes. Energy efficiency comparison with zero circuit power and symmetric bidirectional packet sizes. The curves of DT and OWRT with $\alpha = 4$ respectively stop at $(B_{ab} + B_{ba})/(TW) = 4.4$ and 4 bit/s/Hz, since larger packet sizes will result in unacceptable outage probability as shown in Figure 4.

probability is shown in Figure 4. The x-axis is the overall number of transmitted bits in two directions normalized by the block duration and bandwidth, i.e., $(B_{ab} + B_{ba})/(TW)$, which can be viewed as the average bidirectional SE per block.[d]

In Figure 3, because of the normalization, the EE curves of DT under different α overlap. It shows that the spectral efficient strategy TWRT is also energy efficient with respect to OWRT. When the attenuation factor is large, i.e., $\alpha = 4$, the EE of TWRT is higher than that of DT, while when $\alpha = 2$ the result is just the opposite. The comparison between DT and OWRT depends both on the packet size and the channel condition. When $\alpha = 2$, DT always outperforms OWRT.

Table 1 List of important parameters

Symbol	Definition	Simulation setting
D	Distance between source nodes $\mathbb{A}$ and $\mathbb{B}$	100 m
$d, D\text{-}d$	Distance between $\mathbb{A}$ and $\mathbb{B}$ to relay node $\mathbb{R}$	50 m
PL	Path loss attenuation	$30 + 10\log_{10}(\text{distance}^{\alpha})$ dB
α	Path loss attenuation factor	2, 4
N_0	Noise power at each node	-94 dBm
W	Bandwidth	10MHz
T	Block duration	5 ms
B_{ab}, B_{ba}	Packet sizes in two directions	≥ 0
ϵ	Power amplifier efficiency	0.35
$P_{\max}^t$	Maximum transmit power	40dBm
$P_a^{ct}, P_b^{ct}, P_r^{ct}$	Circuit power in transmit mode at each node	From 0 to hundreds of mW
$P_a^{ct}, P_b^{ct}, P_r^{ct}$	Circuit power in receive mode at each node	From 0 to hundreds of mW
$P_a^{ct}, P_b^{ct}, P_r^{ct}$	Circuit power in idle mode at each node	From 0 to hundreds of mW
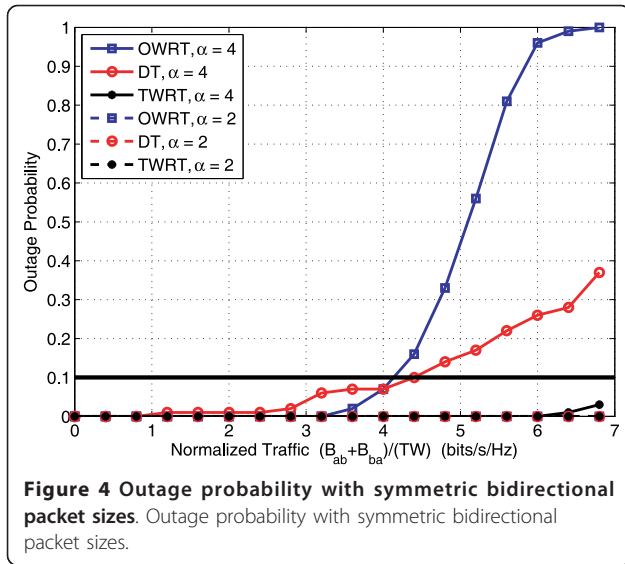
Figure 4 Outage probability with symmetric bidirectional packet sizes. Outage probability with symmetric bidirectional packet sizes.

Figure 5 Energy efficiency of TWRT with different circuit powers. Energy efficiency of TWRT with different circuit powers: the attenuation factor $\alpha = 4$, circuit power consumptions at each node are identical, i.e., $P_a^{ct} = P_b^{ct} = P_r^{ct} \triangleq P^{ct}, P_a^{cr} = P_b^{cr} = P_r^{cr} \triangleq P^{cr}$ and $P_a^{ci} = P_b^{ci} = P_r^{ci} \triangleq P^{ci}$. In the lowest curve, the system transmits with entire duration without optimizing the transmission time, and thus there is no idle duration. In all other curves, the system transmits with the optimized transmission time.

When $\alpha = 4$, OWRT is superior to DT in low traffic region, but is inferior to DT in high traffic region. All these results agree well with our analysis.

Figure 4 shows that when $\alpha = 2$ the outage probabilities of DT, OWRT, and TWRT are zero for the considered packet sizes. When $\alpha = 4$, the outage probabilities all increase. We see that TWRT offers lowest outage probability, and thus can support larger packet size given the same outage probability.

Since we only consider the case where the outage probability is lower than an acceptable threshold, say 10%, the EE curves of OWRT or DT when $\alpha = 4$ is only plotted for the scenarios where $(B_{ab} + B_{ba})/(TW)$ is lower than 4 or 4.4 bits/s/Hz in Figure 3. In the following sections, we use the same method to determine the maximal packet sizes for DT, OWRT and TWRT, which ensure the outage probability to be lower than 10%.

6.2 Non-zero circuit power consumption
In Figure 5, we take TWRT as an example to show the impact of different circuit powers. We present the maximal EEs, which are achieved by the optimized transmission time and transmit power, i.e., there may be idle duration in each block. For comparison, we provide the baseline case again where the circuit PCs are zero. To show the necessity of the transmission time optimization, we also show the EE for a system who transmits with the entire block duration (i.e., there is no idle duration).

As expected, the non-zero circuit PC reduces the EE. It shows that the circuit PC only affects the EE in low traffic region, i.e., in low SE region. While in high SE region, since the transmit PC is much higher than the circuit PC, the EEs are almost the same for different

circuit PCs. That is to say, the high and low SE regions are, respectively, "transmit power dominant" and "circuit power dominant".
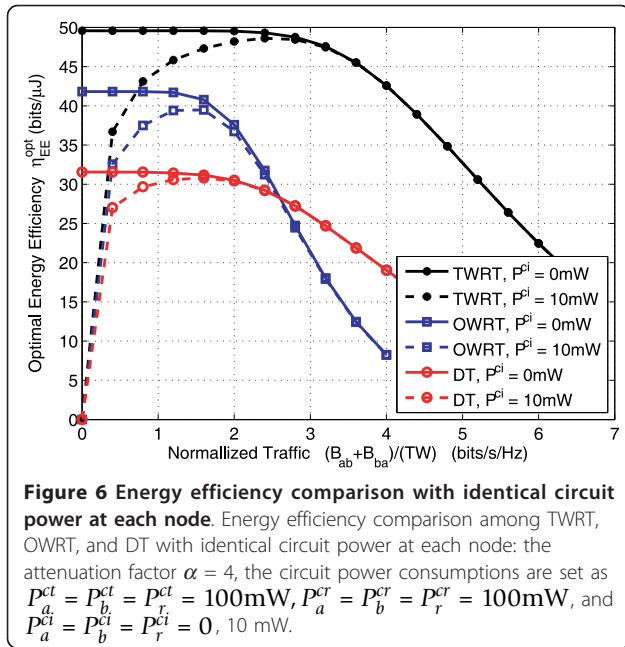
When we assume the circuit PC in idle mode $P^{ci} = 0$, i.e., there exists an idle duration but its PC can be ignored, the EE does not change with SE in the "circuit power dominant" region. As the circuit PCs in the transmit and receive modes P^{ct} and P^{cr} increase, this region becomes wider.

When $P^{ci} \neq 0$, the EE reduces to zero as the packet size decreases. Comparing the lowest two curves where $P^{ci} = 10$ mW, we can see that the EE will decrease if we do not consider the idle duration, i.e., do not optimize the transmission time. Moreover, it is shown that when the PC in idle mode is not negligible, there is a non-zero optimal packet size that maximizes the maximal EE.

All these results agree with our earlier analytical analysis. We do not show the results of OWRT and DT, which are similar as those of TWRT.

In Figure 6, we compare the EEs of different strategies with equal circuit PC at each node, where $\alpha = 4$. It shows that the EE of TWRT is always higher than that of OWRT. Since the path loss is severe, TWRT outperforms DT. OWRT is superior to DT in low traffic region, but becomes inferior in high traffic region. These results are the same as those in zero circuit PC scenario.

From Figure 6, we see that the idle mode circuit power P^{ci} only affects the energy efficiencies in low

Figure 6 Energy efficiency comparison with identical circuit power at each node. Energy efficiency comparison among TWRT, OWRT, and DT with identical circuit power at each node: the attenuation factor $\alpha = 4$, the circuit power consumptions are set as $P_a^{ct} = P_b^{ct} = P_r^{ct} = 100\text{mW}, P_a^{cr} = P_b^{cr} = P_r^{cr} = 100\text{mW}$, and $P_a^{ci} = P_b^{ci} = P_r^{ci} = 0$, 10 mW.

traffic region, and the comparison result among different strategies will not change no matter P^{ci} is zero or not. Since the different EE curves are more distinguishable when the circuit power in idle mode is zero, in the following we set the circuit power in idle status $P^{ci} = 0$.
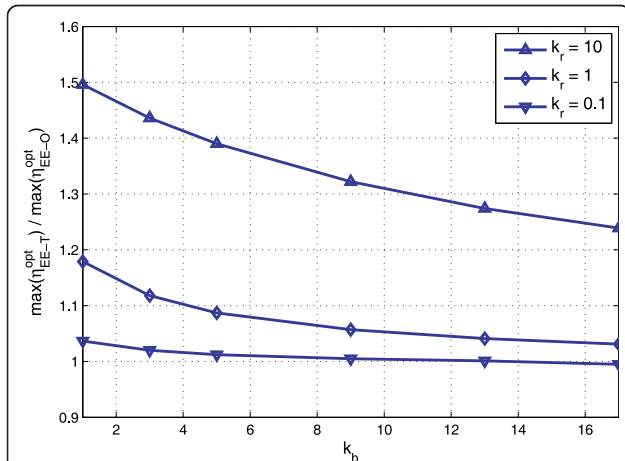
Figure 7 Performance gain of TWRT over OWRT in energy efficiency with unequal circuit powers at each node. The gain of the maximal energy efficiency of TWRT over that of OWRT, considering unequal circuit power consumptions at each node. The circuit powers of node $\mathbb{A}$ in transmit, receive, and idle modes are, respectively, $P_a^{ct} = 50\text{mW}$, $P_a^{cr} = 100\text{mW}$, and $P_a^{ci} = 0\text{mW}$. The circuit powers of nodes $\mathbb{B}$ and $\mathbb{R}$ are $\left(P_b^{ct}, P_b^{cr}, P_b^{ci}\right) = k_b\left(P_a^{ct}, P_a^{cr}, P_a^{ci}\right)$ and $\left(P_r^{ct}, P_r^{cr}, P_r^{ci}\right) = k_r\left(P_a^{ct}, P_a^{cr}, P_a^{ci}\right)$ where k_b and k_r reflect the unequal circuit powers at the three nodes. The attenuation factor $\alpha = 4$.

mW. Note that the circuit powers in transmit and receive modes P^{ct} and P^{cr} are still non-zero.

In Figure 7, we compare the EEs with unequal circuit PCs at each node. We set the circuit PCs as $p_b^{ct} = k_b p_a^{ct}, p_b^{cr} = k_b p_a^{cr}$, where $k_b \geq 1$, which means that node $\mathbb{B}$ consumes more circuit power than node $\mathbb{A}$. We also set $p_r^{ct} = k_r p_a^{ct}, p_r^{cr} = k_r p_a^{cr}$, where $k_r \geq 1$ or $k_r \leq 1$, which reflects the cases where the RN consumes more circuit power or less circuit power than node $\mathbb{A}$ depending on specific application scenarios.

It is easy to understand that if the circuit PC at the RN is high, the advantage of relay transmission over direct transmission shrinks and vice versa. Therefore, we focus on the comparison between OWRT and TWRT in Figure 7. We plot the performance gain of the maximal EE of TWRT over that of OWRT, i.e., $\dfrac{\max(\eta_{\text{EE-T}}^{\text{opt}})}{\max(\eta_{\text{EE-O}}^{\text{opt}})}$, in order to observe whether TWRT is more energy efficient than OWRT, and how much performance gain TWRT can achieve.

From the simulation results in Figure 7, we can see that as k_b increases, i.e., the difference of the circuit PCs at the two source nodes becomes larger, the benefit of TWRT over OWRT shrinks. The OWRT even become more energy efficient than TWRT when the relay circuit PC is low.

6.3 Unequal bidirectional packet sizes
Finally, we compare the maximal EEs with unequal bidirectional packet sizes, which are shown in Figure 8. It shows that the EEs of DT and OWRT do not depend on the ratio B_{ab}/B_{ba}, but the EE of TWRT reduces as
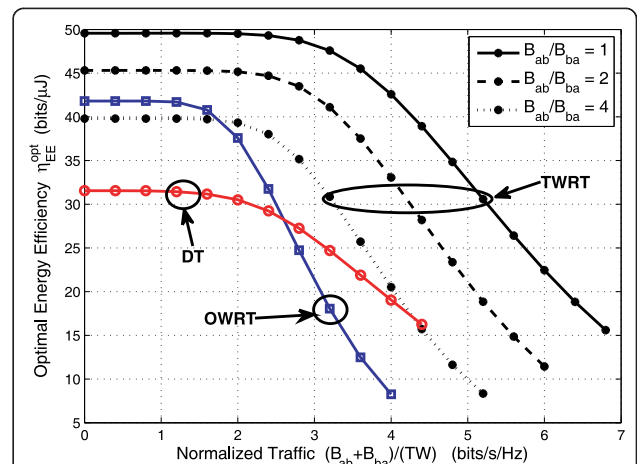
Figure 8 Impact of unequal bidirectional packet sizes. Impact of unequal bidirectional packet sizes: the attenuation factor $\alpha = 4$, circuit power consumptions $P_a^{ct} = P_b^{ct} = P_r^{ct} = 100\text{mW}, P_a^{cr} = P_b^{cr} = P_r^{cr} = 100\text{mW}$, and $P_a^{ci} = P_b^{ci} = P_r^{ci} = 0\text{mW}$.

the difference between B_{ab} and B_{ba} increases, and may even become lower than those of OWRT and DT.

Note that in all the simulations, we did not consider the Approximations 1 and 2 employed in the beginning of Section 5. We can see that the analytical results using those approximations agree well with the simulation results. This validates the previous theoretical analysis.

7 Conclusion

In this article, we studied the energy efficiencies of OWRT and TWRT, and compared with direct transmission. We first found the maximal energy efficiencies of three strategies by jointly optimizing the bidirectional transmission time and the transmit power. We then compared their maximal energy efficiencies with either zero or non-zero circuit power consumptions, and reveal the mechanisms to improve the energy efficiency of the three transmission strategies under different scenarios.

Analytical and simulation results showed that in symmetric systems with equal circuit power at each node and equal packet sizes in two directions, the spectral efficient two-way relaying is also more energy efficient than one-way relaying, but two-way relaying only provides higher energy efficiency than direct transmission when the path loss attenuation is large. In asymmetric systems where the circuit power consumptions at each node are different or the bidirectional packet sizes are unequal, the advantage of two-way relaying diminishes because it can not simultaneously minimize the energy consumed by the transmissions in two directions. One-way relaying may offer higher energy efficiency, depending on the difference between the amount of data in two directions. Compared with the joint transmit power and transmission time optimization, only optimizing the transmit power has a loss in EE when the packet size is small. All the comparison results reveal that relaying is not always more energy efficient than direct transmission, and the two-way relaying does not not always offer higher energy efficiency than one-way relaying. To save the energy consumption, a system should choose the most suitable transmission strategy considering its required amount of data to be transmitted, channel statistics, hardware circuit powers, and so on.

We also showed the relationship between the energy efficiency and the spectral efficiency, i.e., the required amount of data normalized by bandwidth and time duration, for all the three transmission strategy, which is largely dependent on the circuit power consumption. With zero circuit power, the energy efficiency achieves its maximum value as the spectral efficiency approaches zero. With non-zero circuit powers in transmit and receive duration but negligible circuit powers in idle duration, energy efficiency does not change with spectral efficiency in low traffic region but reduce sharply in high traffic region. With non-zero circuit powers in all the transmit, receive and idle modes, there exists a non-zero optimal spectral efficiency that maximizes the maximal energy efficiency.

Appendix 1: Solution of optimization problem (14)

From (4), the transmit power at node $\mathbb{A}$ can be expressed as a function of the transmit power at the RN in $\mathbb{A} \to \mathbb{B}$ link as

$$P_a^t = \frac{C_1 |h_{br}|^2 P_{r1}^t N_0 + C_1 N_0^2}{|h_{ar}|^2 |h_{br}|^2 P_{r1}^t - C_1 |h_{ar}|^2 N_0} \triangleq f(P_{r1}^t), \qquad (44)$$

where $C_1 \triangleq 2^{2B_{ab}/(T_{ab}W)} - 1$.

By substituting (44) into both the objective function and the constraints of (14), the optimization problem can be rewritten as

$$\begin{aligned} &\min_{P_{r1}^t} f(P_{r1}^t) + P_{r1}^t \\ &\text{s.t. } f(P_{r1}^t) \leq P_{\max}^t, P_{r1}^t \leq P_{\max}^t, \end{aligned} \qquad (45)$$

which only depends on P_{r1}^t.

It is easy to show that the objective function is convex by taking its second order derivative with respect to P_{r1}^t, which is positive. Without the two constraints in this problem, the optimal P_{r1}^t can be obtained as follows by setting the first order derivative of the objective function with respect to P_{r1}^t as zero,

$$P_{r1}^{t-opt} = \frac{C_1 N_0}{|h_{br}|^2} + \frac{\sqrt{C_1^2 + C_1} N_0}{|h_{ar} h_{br}|}. \qquad (46)$$

Then the corresponding optimal transmit power at node A can be obtained by substituting (46) into (44),

$$P_a^{t-\mathrm{opt}} = f(P_{r1}^{t-\mathrm{opt}}) = \frac{C_1 N_0}{|h_{ar}|^2} + \frac{\sqrt{C_1^2 + C_1} N_0}{|h_{ar} h_{br}|}. \qquad (47)$$

We can see that both $P_{r1}^{t-\mathrm{opt}}$ and $P_a^{t-\mathrm{opt}}$ are increasing functions of $C_1 = 2^{2B_{ab}/(T_{ab}W)} - 1$, thus are decreasing functions of T_{ab}. Therefore, when T_{ab} is high enough, both $P_{r1}^{t-\mathrm{opt}}$ and $P_a^{t-\mathrm{opt}} = f(P_{r1}^{t-\mathrm{opt}})$ will satisfy the two constraints in (45). Then (46) and (47) are the optimal solutions of the problem (14).

As T_{ab} decreases, both $P_{r1}^{t-\mathrm{opt}}$ and $P_a^{t-\mathrm{opt}}$ increase, until one of them achieve its maximum value. By substituting (46) and (47) into $P_{r1}^{t-\mathrm{opt}} = P_{\max}^t$ and $P_{r1}^{t-\mathrm{opt}}$, respectively, we can derive the corresponding

demarcation point $T_{ab} = T_{d1}$ where $P_{r1}^{t-\text{opt}}$ achieves its maximal value, and can also derive the corresponding $T_{ab} = T_{d2}$ where $P_a^{t-\text{opt}}$ achieves its maximal value. The derived T_{d1} and T_{d2} are given by

$$T_{d1} = \frac{2B_{ab}}{W\log_2\left(1 + \frac{|h_{br}|^2\left[N_0 + 2P_{\max}^t|h_{ar}|^2 - \sqrt{N_0^2 + 4P_{\max}^t N_0|h_{ar}|^2 + 4(P_{\max}^t)^2|h_{ar}h_{br}|^2}\right]}{2N_0(|h_{br}|^2 - |h_{ar}|^2)}\right)}, \quad (48)$$

$$T_{d2} = \frac{2B_{ab}}{W\log_2\left(1 + \frac{|h_{br}|^2\left[N_0 + 2P_{\max}^t|h_{br}|^2 - \sqrt{N_0^2 + 4P_{\max}^t N_0|h_{br}|^2 + 4(P_{\max}^t)^2|h_{ar}h_{br}|^2}\right]}{2N_0(|h_{br}|^2 - |h_{ar}|^2)}\right)}. \quad (49)$$

If $T_{d1} \geq T_{d2}$, as T_{ab} decreases, $P_{r1}^{t-\text{opt}}$ achieves its maximal value first, then we have

$$P_{r1}^{t-\text{opt}} = P_{\max}^t. \quad (50)$$

The corresponding $P_a^{t-\text{opt}}$ can be obtained by substituting (50) into (44), which is

$$P_a^{t-\text{opt}} = \frac{C_1|h_{br}|^2 P_{\max}^t N_0 + C_1 N_0^2}{|h_{ar}|^2|h_{br}|^2 P_{\max}^t - C_1|h_{ar}|^2 N_0}. \quad (51)$$

If $T_{d1} < T_{d2}$, as T_{ab} decreases, $P_a^{t-\text{opt}}$ achieves its maximal value first, then we have

$$P_a^{t-\text{opt}} = P_{\max}^t. \quad (52)$$

The corresponding $P_{r1}^{t-\text{opt}}$ can be derived using (44) by substituting (52),

$$P_{r1}^{t-\text{opt}} = \frac{C_1|h_{ar}|^2 P_{\max}^t N_0 + C_1 N_0^2}{|h_{ar}|^2|h_{br}|^2 P_{\max}^t - C_1|h_{br}|^2 N_0}. \quad (53)$$

By adding (46) and (47), (50) and (51), and (52) and (53), we can obtain the expressions of $P_{\min 1}(T_{ab}) = \min(P_a^t + P_{r1}^t)$ in (16) and (17).

Appendix 2: Proof of quasi-convexity of the objective function in (20)

We consider the case that $P_{\min 1}(T_{ab})$ follows (16), the conclusion is the same if it follows (17). Since $P_{\min 1}(T_{ab})$ is a piecewise function of T_{ab}, $T_{ab}\left(\frac{P_{\min 1}(T_{ab})}{2\varepsilon} + P_O^{c1} - P_O^{ci}\right)$ is also a piecewise function. For simplicity, we define

$$T_{ab}\left(\frac{P_{\min 1}(T_{ab})}{2\varepsilon} + P_O^{c1} - P_O^{ci}\right) = \begin{cases} f_l(T_{ab}), & T_{\min 1} \leq T < T_{d1} \\ f_r(T_{ab}), & T \geq T_{d1}. \end{cases} \quad (54)$$

By taking the second order derivative of $f_l(T_{ab})$, we have $f_l''(T_{ab}) \geq 0$ when $T_{\min 1} \leq T < T_{d1}$. Therefore, $f_l(T_{ab})$ is a convex function in the range $T_{\min 1} \leq T < T_{d1}$.

Then we will show that $f_r(T_{ab})$ is a quasi-convex function in the range $T > T_{d1}$, where we will use the following lemma.

Lemma 1. *Suppose that a function $f(x)$ is second order differentiable in (x_L, x_R), $\lim_{x \to x_L} f'(x) < 0$, $\lim_{x \to r_R} f'(x) > 0$, and $f''(x)$ only has one zero point in (x_L, x_R). Then $f(x)$ is a quasi-convex function on (x_L, x_R).*

Proof. Since $f(x)$ is second order differentiable, $f'(x)$ is continuous on (x_L, x_R). Considering that $\lim_{x \to x_L} f'(x) < 0$, $\lim_{x \to r_R} f'(x) > 0$ $f'(x)$ at least has one zero point in (x_L, x_R). We then show that $f'(x)$ can only has one zero point.

Assume that $f'(x)$ has three or more zero points such that $f'(a) = f'(b) = f'(c) = 0$. According to Rolle's theorem, there exists a point $x_1 \in (a, b)$ such that $f''(x_1) = 0$, and also a point $x_2 \in (b, c)$ such that $f''(x_2) = 0$. This conflicts with the assumption that $f''(x)$ only has one zero point.

Assume that $f'(x)$ has two zero points such that $f'(a) = f'(b) = 0$, $a, b \in (x_L, x_R)$. According to Rolle's theorem, there is a point $x_1 \in (a, b)$ which satisfies $f''(x_1) = 0$. Without loss of generality, we assume that $f'(x_1) > 0$. Considering that $\lim_{x \to x_R} f'(x) > 0$, and in (x_1, x_R), $f'(x)$ only has one zero point $f'(b) = 0$, therefore, $f'(b) = 0$ is the minimum value of $f'(x)$ in (x_1, x_R), and thus $f''(b) = 0$. Then we have two zero points for $f''(x)$, which conflicts with the assumption that $f''(x)$ only has one zero point.

Consequently $f'(x)$ can only has one zero point. Assume that $f'(x_M) = 0$. Then in (x_L, x_M), $f'(x) < 0$, $f(x)$ is non-increasing, while in (x_M, x_R), $f'(x) > 0$, $f(x)$ is non-decreasing, which means that $f(x)$ is a quasi-convex function in (x_L, x_R) [24].

By taking the derivative of $f_r(T_{ab})$, we find that $f_r'(0) \to -\infty$, and $\lim_{T_{ab} \to \infty} f_r'(T_{ab}) = P_O^{c1} - P_O^{ci} \geq 0$ since the circuit PC in the idle mode is lower than that in the transmit or receive mode. We also find that

$$f_r''(T_{ab}) = k_1[k_2 + k_3 g(T_{ab})], \quad (55)$$

where $k_1 = \frac{2(\ln 2)^2 B_{ab}^2 N_0}{(W^2 T_{ab}^3 \varepsilon)} 2^{\frac{2B_{ab}}{(WT_{ab})}} > 0, k_2 = \frac{1}{|h_{ar}|^2} + \frac{1}{|h_{br}|^2} > 0, k_3 = \frac{1}{2|h_{ar}h_{br}|} > 0,$, k_2 and k_3 do not depend on T_{ab}, and $g(T_{ab})$ is given by

$$g(T_{ab}) = \frac{4\left(2^{\frac{2B_{ab}}{WT_{ab}}} - 1\right)^2 + 2\left(2^{\frac{2B_{ab}}{WT_{ab}}} - 1\right) - 1}{\left(2^{\frac{2B_{ab}}{WT_{ab}}} - 1\right)\sqrt{\left(2^{\frac{2B_{ab}}{WT_{ab}}} - 1\right)2^{\frac{2B_{ab}}{WT_{ab}}}}}. \quad (56)$$

We can easily obtain that $\lim_{T_{ab} \to 0} g(T_{ab}) = 4$, $\lim_{T_{ab} \to +\infty} g(T_{ab}) = -\infty$ and $g'(T_{ab}) < 0$, for

$T_{ab} > 0$. Then $g(T_{ab})$ strictly monotonically decreases from 1 to -∞ when $T_{ab} > 0$. Therefore, $f''(T_{ab})$ in (55) only has one zero point. According to Lemma 1, $f_r(T_{ab})$ is a quasi-convex function on $(0, +\infty)$, and thus a quasi-convex function on $[T_{d1}, +\infty)$.

Based on the expression of T_{d1} derived in Appendix 1, we can obtain that $\lim_{T_{ab}\to T_{d1}} f'_l(T_{ab}) = \lim_{T_{ab}\to T_{d1}} f'_r(T_{ab}) \triangleq \delta$. If $\delta \leq 0$, $T_{ab}\left(\frac{P_{\min 1}(T_{ab})}{2\varepsilon} + P_O^{c1} - P_O^{ci}\right) = f_l(T_{ab})$ monotonically decreases in $[T_{\min 1}, T_{d1})$ due to the convexity of $f_l(T_{ab})$, while $T_{ab}\left(\frac{P_{\min 1}(T_{ab})}{2\varepsilon} + P_O^{c1} - P_O^{ci}\right) = f_r(T_{ab})$ first decreases and then increases in $[T_{d1},+\infty)$ due to the quasi-convexity of $f_r(T_{ab})$. Therefore, $T_{ab}\left(\frac{P_{\min 1}(T_{ab})}{2\varepsilon} + P_O^{c1} - P_O^{ci}\right)$ is quasi-convex in $[T_{\min 1}, +\infty)$. If $\delta > 0$, the same is true.

Appendix 3: Derivation of the optimal transmission time

Recall that in Approximation 1, we only consider the case where none of the nodes achieves its maximal transmit power and thus we can ignore the minimum value constraints on transmission time. Then the optimization problem of the transmission time is given by

$$
\min_{T_{ab},T_{ba}} T_{ab}\left[\frac{N_0\left(2^{\frac{2\beta B_s}{WT_{ab}}}-1\right)}{2\varepsilon|h_{\text{eff}}|^2}+P_O^c-P_O^{ci}\right]+T_{ba}\left[\frac{N_0\left(2^{\frac{2(1-\beta)B_s}{WT_{ba}}}-1\right)}{2\varepsilon|h_{\text{eff}}|^2}+P_O^c-P_O^{ci}\right]+TP_O^{ci} \quad (57)
$$
$$
\text{s.t. } T_{ab}+T_{ba}\leq T
$$

This is a convex problem, where the optimal T_{ab} and T_{ba} should satisfy the following Karush-Kuhn-Tucker (KKT) conditions,

$$\lambda(T_{ab}^{\text{opt}} + T_{ba}^{\text{opt}} - T) = 0, \quad (58a)$$

$$\frac{N_0\left(2^{\frac{2\beta B_s}{WT_{ab}^{\text{opt}}}}-1\right)}{2\varepsilon|h_{\text{eff}}|^2}+P_O^c-P_O^{ci}-\frac{N_0}{2\varepsilon|h_{\text{eff}}|^2}2^{\frac{2\beta B_s}{WT_{ab}^{\text{opt}}}}(\ln 2)\frac{2\beta B_s}{WT_{ab}^{\text{opt}}}=-\lambda, \quad (58b)$$

$$\frac{N_0\left(2^{\frac{2(1-\beta)B_s}{WT_{ba}^{\text{opt}}}}-1\right)}{2\varepsilon|h_{\text{eff}}|^2}+P_O^c-P_O^{ci}-\frac{N_0}{2\varepsilon|h_{\text{eff}}|^2}2^{\frac{2(1-\beta)B_{ba}}{WT_{ba}^{\text{opt}}}}(\ln 2)\frac{2(1-\beta)B_s}{WT_{ba}^{\text{opt}}}=-\lambda, \quad (58c)$$

where λ is the Lagrange multiplier.

We can see that the expressions in the left-hand side of (58b) and (58c) equal to each other. Therefore, the optimal transmission time satisfies

$$\frac{\beta B_s}{T_{ab}^{\text{opt}}} = \frac{(1-\beta)B_s}{T_{ba}^{\text{opt}}} \triangleq R_O. \quad (59)$$

Substituting (59) into the KKT conditions, it is easy to see that R_O is not a function of β.

Endnotes
[a]The feasible region of the EE optimization problem may be empty, which implies an outage of a block. Thereby we do not need to optimize for this block. Similar case also exists in the OWRT and TWRT optimization problems.
[b]It should be noted that AWGN channel is appropriate for modeling free space propagation where $\alpha = 2$. We consider different path loss attenuation factors here, which may be an abuse of the terminology of "AWGN channel".
[c]This can not happen in practice, which is considered only for the proof.
[d]The average bidirectional SE per block takes into account the entire duration of a block, which includes not only the transmission time but also the idle duration.

References
1. L Correia, D Zeller, O Blume, D Ferling, Y Jading, I Godor, G Auer, L Perre, Challenges and enabling technologies for energy aware mobile radio networks. IEEE Commun Mag. **48**(11), 66–72 (2010)
2. G Li, Z Xu, C Xiong, C Yang, S Zhang, Y Chen, S Xu, Energy-efficient wireless communications: tutorial, survey and open issues. IEEE Commun Mag. **18**(6), 28–35 (2011)
3. C Han, T Harrold, S Armour, I Krikidis, S Videv, PM Grant, H Haas, JS Thompson, I Ku, C Wang, T Le, MR Nakhai, J Zhang, L Hanzo, Green radio: radio techniques to enable energy-efficient wireless networks. IEEE Commun Mag. **49**(6), 46–54 (2011)
4. Y Chen, S Zhang, S Xu, G Li, Fundamental trade-offs on green wireless networks. IEEE Commun Mag. **49**(6), 30–37 (2011)
5. Y Li, X Zhang, M Peng, W Wang, Power provisioning and relay positioning for two-way relay channel with analog network coding. IEEE Signal Process Lett. **18**(9), 517–520 (2011)
6. M Dohler, Y Li, *Cooperative Communications, Hardware, Channel & Phy*, (Wiley, UK, 2010)
7. J Laneman, D Tse, G Wornell, Cooperative diversity in wireless networks: efficient protocols and outage behavior. IEEE Trans Inf Theory. **50**(12), 3062–3080 (2004). doi:10.1109/TIT.2004.838089
8. B Rankov, A Wittneben, Spectral efficient protocols for half-duplex fading relay channels. IEEE J Select Areas Commun. **25**(2), 379–389 (2007)

9. T Oechtering, E Jorswieck, R Wyrembelski, H Boche, On the optimal transmit strategy for the MIMO bidirectional broadcast channel. IEEE Trans Commun. **57**(12), 3817–3826 (2009)

10. C Sun, Y Li, B Vucetic, C Yang, Transceiver design for multi-user multi-antenna two-way relay channels, in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'10)*, Miami, Florida, USA, (December 2010), pp. 1–5

11. C Bae, W Stark, End-to-end energy-bandwidth tradeoff in multihop wireless networks. IEEE Trans Inf Theory. **55**(9), 4051–4066 (2009)

12. C Chen, W Stark, S Chen, Energy-bandwidth efficiency tradeoff in MIMO multi-hop wireless networks. IEEE J Select Areas Commun. **29**(8), 1537–1546 (2011)

13. R Madan, N Mehta, A Molisch, J Zhang, Energy-efficient cooperative relaying over fading channels with simple relay selection. IEEE Trans Wirel Commun. **7**(8), 3013–3025 (2008)

14. S Wang, J Nie, Energy efficiency optimization of cooperative communication in wireless sensor networks. EURASIP J Wirel Commun Netw. **2010**, 1–8 (2010)

15. D Cao, S Zhou, C Zhang, Z Niu, Energy saving performance comparison of coordinated multi-point transmission and wireless relaying, in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'10)*, Miami, Florida, USA, (December 2010), pp. 1–5

16. P Rost, Opportunities, benefits, and constraints of relaying in mobile communication systems, (Technische University Dresden, 2009) PhD thesis

17. H Xu, B Li, XOR-assisted cooperative diversity in OFDMA wireless networks: Optimization framework and approximation algorithms, in *Proceedings of IEEE International Conference on Computer Communications (INFOCOM'09)*, Rio de Janeiro, Brazil, (April 2009), pp. 2141–2149

18. Q Li, S Ting, A Pandharipande, Y Han, Adaptive two-way relaying and outage analysis. IEEE Trans Wirel Commun. **8**(6), 3288–3299 (2009)

19. M Zafer, E Modiano, Optimal rate control for delay-constrained data transmission over a wireless channel. IEEE Trans Inf Theory. **54**(9), 4020–4039 (2008)

20. J Lee, N Jindal, Energy-efficient scheduling of delay constrained traffic over fading channels. IEEE Trans Wirel Commun. **8**(4), 1866–1875 (2009)

21. S Cui, A Goldsmith, A Bahai, Energy-constrained modulation optimization. IEEE Trans Wirel Commun. **4**(5), 2349–2360 (2005)

22. S Howard, C Schlegel, K Iniewski, Error control coding in low-power wireless sensor networks: when is ECC energy-efficient?. EURASIP J Wirel Commun Netw. **2006**, 1–14 (2006)

23. R Louie, Y Li, B Vucetic, Practical physical layer network coding for two-way relay channels: performance analysis and comparison. IEEE Trans Wirel Commun. **9**(2), 764–777 (2010)

24. S Boyd, L Vandenberghe, Convex Optimization, (Cambridge University Press, New York, 2004)