Energy-Efficient Training-Assisted Transmission Strategies for Closed-Loop MISO Systems

Xin Liu, Shengqian Han, *Member, IEEE*, and Chenyang Yang, *Senior Member, IEEE*

***Abstract*—This paper studies energy-efficient transmission strategies for a closed-loop downlink multiple-input–single-output (MISO) system, where a communication period consists of three phases for uplink training, downlink data sending, and base station (BS) idling. For both delay-tolerant and delay-sensitive services, the durations of the three phases are optimized, aimed at maximizing the energy efficiency (EE) of the system. To this end, we derive the approximate average net spectrum efficiency (SE) and outage probability with imperfect uplink channel estimation, which are used to characterize the quality-of-service (QoS) requirements for the two kinds of services, respectively. The impact of QoS requirement, signal-to-noise ratio (SNR), and circuit power consumption on the optimal transmission durations is analyzed. For delay-tolerant services, analytical results show that the EE-oriented design leads to a longer training duration than the SE-oriented design in general. For delay-sensitive services, it is shown that introducing BS idling is crucial in improving the EE. The challenges and opportunities of applying the proposed transmission strategies in current and future cellular systems are discussed, and the transmission strategies are extended from single-user single-service to multiuser mixed-service scenarios. Simulation results demonstrate the significant EE gain of the EE-oriented design over the SE-oriented design in both single-user and multiuser scenarios.**

***Index Terms*—Base station (BS) idling, energy efficiency (EE), quality of service (QoS), training design.**

I. Introduction

ENERGY-efficient communications are becoming more and more desirable for future wireless networks to reduce the associated energy consumption, carbon emission, and operational cost [1]. To improve the energy efficiency (EE) while ensuring the quality-of-service (QoS) requirements of users, traffic and QoS-aware design is an important principle, based on which some energy-efficient transmission strategies have been proposed. For example, traffic-aware power-sharing policies considering spatial traffic load difference were studied in [2], base station (BS) sleeping control and power matching schemes were studied in [3] to achieve a good tradeoff between energy saving and traffic delay, and opportunistic transmission scheduling schemes and advanced relay transmission schemes exploiting the delay tolerance of services were investigated in [4] and [5], respectively. In [6] and [7], a typical QoS-aware strategy, i.e., BS idling, was investigated in the time domain and the frequency–time domain, respectively, which provides on-demand services to users according to their QoS requirements and is commonly recognized as a promising approach to reduce energy consumption. In [8], a QoS-based antenna switching-off technique was proposed, where BS idling is in the spatial domain.

The application of BS idling requires knowledge of channel state information (CSI) at the BS, based on which the BS can transmit all data during a part of the time slots with an adaptive rate, while remaining in idle mode during other time slots to save energy. In [6] and [7], perfect CSI is assumed at the BS for the optimization of BS idling, which leads to optimistic EE performance since the impact of channel estimation errors and the resource usage, as well as energy consumption for channel acquisition are not taken into account.

In time-division duplex (TDD) systems, the BS can obtain the downlink CSI by estimating the uplink CSI from the received uplink training signals based on the channel reciprocity. In the literature, training design for the maximization of system spectral efficiency (SE) has been well studied. In [9], the length of training signals for a multiple-input multiple-output (MIMO) system was optimized to maximize the lower bound of the channel capacity. It is shown that the optimal training length is equal to the number of transmit antennas when the transmit power of training and data is jointly optimized. The optimal training length that maximizes the net uplink SE and downlink SE in multiuser MIMO systems was investigated in [10] and [11], respectively.

However, the training designed for high SE does not necessarily lead to high EE when taking into account the circuit power consumption and signaling overhead. The training design aimed at EE maximization was first studied in [12] for a single-input single-output (SISO) system, where the length and transmit power of training signals are optimized. In [13], energy-efficient power allocation between training and data symbols was investigated for a training-based downlink MIMO system. Both the works in [12] and [13] considered the optimization of downlink training, either without or with the circuit power consumptions. The EE-oriented design of uplink training for closed-loop systems was studied in our preliminary work [14], where only delay-tolerant services are considered, and BS idling is not taken into account.

Manuscript received January 1, 2014; revised April 29, 2014; accepted June 8, 2014. Date of publication August 27, 2014; date of current version July 14, 2015. This work was supported in part by the National High Technology Research and Development Program of China under Grant 2014AA01A705, by the National Natural Science Foundation of China under Grant 61120106002, by the National Basic Research Program of China under Grant 2012CB316003, and by the Fundamental Research Funds for the Central Universities. This paper was presented in part at the IEEE Wireless Communications and Networking Conference, Shanghai, China, April 7–10, 2013. The review of this paper was coordinated by Dr. N.-D. Dao. *(Corresponding author: Shengqian Han.)*

The authors are with the School of Electronics and Information Engineering, Beihang University, Beijing 100191, China (e-mail: liuxinjlu@ee.buaa.edu.cn; sqhan@ee.buaa.edu.cn; cyyang@buaa.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2014.2352646

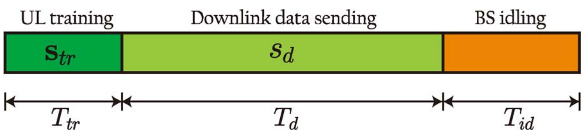


Fig. 1. Communication period of the considered TDD systems consisting of uplink training, downlink data sending, and idling, where $T_{\rm tr}+T_{\rm d}+T_{\rm id}=T$.

In this paper, we study the energy-efficient transmission strategies for a training-assisted downlink closed-loop system, where a communication period is divided into three phases for uplink training, downlink data sending, and BS idling. We first consider a single-user multiple-input single-output (MISO) system, which can act as a good start for continuing the design of energy-efficient closed-loop strategies for general multiuser MIMO systems. Noting that maximizing EE should not sacrifice the user experience and different traffics impose different QoS provisions, we consider both delay-tolerant and delay-sensitive services in the analysis. For each kind of services, the durations of the three phases are jointly optimized, aimed at maximizing EE, and the impact of QoS requirement, signal-to-noise ratio (SNR), and circuit power on the optimal durations and system EE is analyzed. We then discuss the challenges and opportunities of applying the proposed transmission strategies in current and future cellular systems and provide one way to extend the transmission strategy designs from single-user single-service to multiuser mixed-service scenarios. Simulation results validate our analytical analysis and demonstrate the EE gain of the EE-oriented design over the SE-oriented design.

Notations: $\mathbf{I}$ denotes the identity matrix, $|\cdot|$ and $\|\cdot\|$ represent the magnitude and Euclidian norm, respectively, $(\cdot)^H$ denotes conjugate transpose, and $\otimes$ is the Kronecker product. $\mathbb{N}^+$ and $\mathbb{N}$ denote the set of positive integers and nonnegative integers, respectively.

II. System, Power Consumption, and Service Models

A. System Model

Consider a TDD MISO system where the BS is equipped with M antennas and the user has a single antenna. We assume block-fading channels, which remain constant during each communication period but are independent between different periods. As shown in Fig. 1, the communication period has the duration of T symbols, which is divided into three phases including uplink training, downlink data sending, and BS idling with the durations of $T_{\rm tr}$, $T_{\rm d}$, and $T_{\rm id}$ symbols, respectively. In the uplink training phase, the user sends uplink training signals to the BS. Then, in the downlink data sending phase, the BS estimates the uplink CSI and obtains the downlink CSI based on channel reciprocity, with which the downlink precoder is computed for data transmission. In the BS idling phase, the BS switches into idle mode to save energy if the QoS requirement of the user has been satisfied. In the paper, we consider the BS idling that can operate in the symbol level. To realize the fast deactivation and reactivation of the hardware, some components of the BS may not be completely turned off, but they can operate with low power consumption. For example, the millisecond-level BS idling has been considered in [15]. Compared with the long-term BS idling, which can be only used during the traffic off-peak hours, the employed BS idling is a whole-day strategy and is recognized as an important approach for energy saving [7].

During the uplink training phase, the training signals $\mathbf{s}_{\rm tr}$ are sending from the user over the $T_{\rm tr}$ symbols, where $\mathbf{s}_{\rm tr}\in\mathbb{C}^{T_{\rm tr}\times 1}$, and $\mathbf{s}_{\rm tr}^H\mathbf{s}_{\rm tr}=P_uT_{\rm tr}$ with P_u denoting the transmit power of the user. Then, the received signal at the BS can be expressed as

$$\mathbf{y}_u=\mathbf{S}_{\rm tr}\mathbf{h}+\mathbf{n}_u \tag{1}$$

where $\mathbf{S}_{\rm tr}=\mathbf{I}\otimes\mathbf{s}_{\rm tr}$ is an $MT_{\rm tr}\times M$ training matrix, $\mathbf{h}\sim\mathcal{CN}(\mathbf{0},\sigma_h^2\mathbf{I})$ is an $M\times 1$ channel vector whose elements are independent and identically distributed complex Gaussian random variables with zero mean and variance σ_h^2, and $\mathbf{n}_u\sim\mathcal{CN}(\mathbf{0},\sigma_u^2\mathbf{I})$ is additive white Gaussian noise (AWGN) at the BS.

We consider a minimum mean square error (MMSE) channel estimator. Then, the relationship between the true value of channel vector $\mathbf{h}$ and its MMSE estimate $\hat{\mathbf{h}}$ can be modeled as

$$\mathbf{h}=\hat{\mathbf{h}}+\mathbf{e} \tag{2}$$

where $\hat{\mathbf{h}}\sim\mathcal{CN}(\mathbf{0},\sigma_{\hat{h}}^2\mathbf{I})$, $\mathbf{e}\sim\mathcal{CN}(\mathbf{0},\sigma_e^2\mathbf{I})$ is the estimation error vector, and $\hat{\mathbf{h}}$ and $\mathbf{e}$ are uncorrelated [16]. It is not hard to show that $\sigma_{\hat{h}}^2=\sigma_h^2\gamma_uT_{\rm tr}/(1+\gamma_uT_{\rm tr})$ and $\sigma_e^2=\sigma_h^2/(1+\gamma_uT_{\rm tr})$, where $\gamma_u\triangleq P_u\sigma_h^2/\sigma_u^2$ is the average uplink SNR.

During the downlink data sending phase, a maximum ratio transmission precoder is employed [17], which can be expressed as $\mathbf{w}=\sqrt{P_{\rm d}}(\hat{\mathbf{h}}/\|\hat{\mathbf{h}}\|)$, where $P_{\rm d}$ is the transmit power of the BS. The received signal at the user is

$$y_{\rm d}=\mathbf{h}^H\mathbf{w}s_{\rm d}+n_{\rm d} \tag{3}$$

where $s_{\rm d}$ is the downlink data symbol, and $n_{\rm d}\sim\mathcal{CN}(0,\sigma_{\rm d}^2)$ is the AWGN at the user.

By assuming perfect estimation of the downlink equivalent channel $\mathbf{h}^H\mathbf{w}$, the downlink instantaneous SNR can be obtained from (2) and (3) as

$$\text{SNR}=\frac{P_{\rm d}}{\sigma_{\rm d}^2}\left|\frac{\mathbf{h}^H\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|}\right|^2=\frac{P_{\rm d}}{\sigma_{\rm d}^2}\left|\|\hat{\mathbf{h}}\|+\frac{\mathbf{e}^H\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|}\right|^2 \tag{4}$$

and the instantaneous net SE of the system during the whole period is

$$\begin{aligned}SE^{\rm in}&=\frac{T_{\rm d}}{T}\log_2(1+\text{SNR})\\&=\frac{T_{\rm d}}{T}\log_2\left(1+\frac{P_{\rm d}}{\sigma_{\rm d}^2}\left|\|\hat{\mathbf{h}}\|+\frac{\mathbf{e}^H\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|}\right|^2\right).\end{aligned} \tag{5}$$

B. Power Consumption Model

Considering the fact that the overall power consumption at the user is far less than that at the BS, we only consider the power consumed by the BS.

During the uplink training phase, the BS operates in receiving mode. The power consumption can be modeled as [1]

$$P_{\mathrm{bu}} = \frac{M \cdot \left(P_{\mathrm{RF}}^{\mathrm{RX}} + P_{\mathrm{BB}}^{\mathrm{RX}}\right)}{(1-\sigma_{\mathrm{dc}})(1-\sigma_{\mathrm{MS}})(1-\sigma_{\mathrm{cool}})} \triangleq M \cdot P_c^{\mathrm{RX}} \quad (6)$$

where $P_{\mathrm{RF}}^{\mathrm{RX}}$ and $P_{\mathrm{BB}}^{\mathrm{RX}}$ are the power consumed by radio frequency (RF) links and baseband (BB) signal processing such as synchronization and channel estimation; σ_{dc}, σ_{MS}, and σ_{cool} are the loss factors used to reflect the power incurred by dc–dc power supply, mains supply, and active cooling, respectively; and $P_c^{\mathrm{RX}} \triangleq (P_{\mathrm{RF}}^{\mathrm{RX}} + P_{\mathrm{BB}}^{\mathrm{RX}})/((1-\sigma_{\mathrm{dc}})(1-\sigma_{\mathrm{MS}})(1-\sigma_{\mathrm{cool}}))$ is the circuit power per antenna when the BS operates in receiving mode.

During the downlink data sending phase, the BS consumes both transmit and circuit power. The power consumption can be modeled as

$$P_{\mathrm{bd}} = \frac{\frac{P_{\mathrm{d}}}{\eta} + M \cdot \left(P_{\mathrm{RF}}^{\mathrm{TX}} + P_{\mathrm{BB}}^{\mathrm{TX}}\right)}{(1-\sigma_{\mathrm{dc}})(1-\sigma_{\mathrm{MS}})(1-\sigma_{\mathrm{cool}})} \triangleq \mu \cdot P_{\mathrm{d}} + M \cdot P_c^{\mathrm{TX}} \quad (7)$$

where $P_{\mathrm{RF}}^{\mathrm{TX}}$ and $P_{\mathrm{BB}}^{\mathrm{TX}}$ are the power consumed by RF links and BB signal processing such as precoding, modulation, and coding, η is the efficiency of the power amplifier (PA), $\mu \triangleq 1/(\eta(1-\sigma_{\mathrm{dc}})(1-\sigma_{\mathrm{MS}})(1-\sigma_{\mathrm{cool}}))$ is a parameter that reflects the loss of transmit power, and $P_c^{\mathrm{TX}} \triangleq (P_{\mathrm{RF}}^{\mathrm{TX}} + P_{\mathrm{BB}}^{\mathrm{TX}})/((1-\sigma_{\mathrm{dc}})(1-\sigma_{\mathrm{MS}})(1-\sigma_{\mathrm{cool}}))$ is the circuit power per antenna when the BS operates in transmitting mode. Obviously, the power consumption of the BS during the downlink data sending phase is larger than that during the uplink training phase, i.e., $P_{\mathrm{bd}} > P_{\mathrm{bu}}$.

When the BS operates in idling mode, its power consumption P_{id} is far less than P_{bu} and P_{bd}. Therefore, we assume $P_{\mathrm{id}} = 0$ in the analytical analysis, which does not change the conclusions we obtained.

Then, the total power consumption during a communication period can be expressed as

$$P_{\mathrm{tot}} = \frac{P_{\mathrm{bu}} T_{\mathrm{tr}} + P_{\mathrm{bd}} T_{\mathrm{d}}}{T}. \quad (8)$$

C. Service Model

As previously mentioned, the basic principle of energy-efficient design is to maximize the EE of the system and, at the same time, guarantee the QoS for each user. Different services impose different QoS constraints and have different traffic models, which lead to different optimization problems and EE gains. In this paper, we consider two classes of services with the following models to illustrate the EE gains of the systems with "on-demand" transmission strategies, which have very different QoS requirements.

- For *delay-tolerant services*, the traffic model provided in [18] is used, which reflects the services such as best efforts. It is assumed that there is always a sufficiently large backlog of data in the buffer for transmission. Since the data arrival process may be discontinuous in time and the amount of data arrived may be variable, the transmit strategy needs to ensure that the average transmission rate is no smaller than the average arrival rate.
- For *delay-sensitive services*, the traffic model provided in [19] is used. It models the delay-sensitive services such as voice over Internet protocol, where packets regularly arrive, and each must be transmitted within a given delay bound with a given outage probability.

III. Delay-Tolerant Services

To achieve high EE for the system whereas guarantee QoS for the user with delay-tolerant services, we optimize the durations for uplink training, downlink data sending, and BS idling, where the average net SE is used to characterize the user's rate requirement and to compute the EE of the system.

A. Average Net SE and Problem Formulation

From (5), the average net SE is

$$\mathrm{SE} = \mathbb{E}\left[\frac{T_{\mathrm{d}}}{T}\log_2\left(1 + \frac{P_{\mathrm{d}}}{\sigma_{\mathrm{d}}^2}\left|\|\hat{\mathbf{h}}\| + \frac{\mathbf{e}^H\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|}\right|^2\right)\right] \quad (9)$$

where the expectation is taken over both channel estimate $\hat{\mathbf{h}}$ and estimation error $\mathbf{e}$, whose closed-form expression is difficult to obtain.

Considering that an explicit expression of the average net SE is crucial for formulating and solving the EE optimization problem, in the sequel, we seek its approximate expression. It is easy to show that $\|\hat{\mathbf{h}}\|$ is the square root of the sum of squares of M independent complex Gaussian random variables with zero mean and variance $\sigma_{\hat{h}}^2$, and $\mathbf{e}^H\hat{\mathbf{h}}/\|\hat{\mathbf{h}}\|$ is a complex Gaussian variable with zero mean and σ_e^2. Moreover, it is obvious that $\sigma_{\hat{h}}^2$ is a monotonically increasing function of the average uplink SNR γ_u, whereas σ_e^2 is a monotonically decreasing function of γ_u. This suggests that when M is large and γ_u is high, the term $\mathbf{e}^H\hat{\mathbf{h}}/\|\hat{\mathbf{h}}\|$ in (2) is far less than the term $\|\hat{\mathbf{h}}\|$ in a high probability, and the instantaneous downlink SNR can be approximated as

$$\mathrm{SNR} = \frac{P_{\mathrm{d}}}{\sigma_{\mathrm{d}}^2}\left|\|\hat{\mathbf{h}}\| + \frac{\mathbf{e}^H\hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|}\right|^2 \approx \frac{P_{\mathrm{d}}}{\sigma_{\mathrm{d}}^2}\|\hat{\mathbf{h}}\|^2 \triangleq \widetilde{\mathrm{SNR}} \quad (10)$$

where $\widetilde{\mathrm{SNR}} \sim \Gamma(k,\theta)$ is subject to Gamma distribution with shape k and scale θ, $k = M$, $\theta = \gamma_{\mathrm{d}}\gamma_u T_{\mathrm{tr}}/(1+\gamma_u T_{\mathrm{tr}})$, and $\gamma_{\mathrm{d}} \triangleq P_{\mathrm{d}}\sigma_{\hat{h}}^2/\sigma_{\mathrm{d}}^2$ is the average downlink SNR.

Based on the moment-matching method in [20], if a random variable $X \sim \Gamma(k,\theta)$, then $Y = X + c$ can be approximated as a Gamma distributed variable satisfying $\Gamma(k_y,\theta_y)$, where $c \geq 0$ is a constant, $k_y = (k\theta+c)^2/k\theta^2$, and $\theta_y = k\theta^2/(k\theta+c)$. The approximation is accurate when k is large enough. Then, from (10), $1+\widetilde{\mathrm{SNR}}$ is an approximate Gamma distributed variable satisfying $\Gamma(k_1,\theta_1)$, where

$$k_1 = \frac{(k\theta+1)^2}{k\theta^2} = M + \frac{1}{M\theta^2} + \frac{2}{\theta}, \; \theta_1 = \frac{k\theta^2}{k\theta+1} = \frac{M\theta^2}{M\theta+1}. \quad (11)$$

For the variable $Y \sim \Gamma(k_y, \theta_y)$, $\mathbb{E}[\ln(Y)] = \psi(k_y) + \ln(\theta_y)$, where $\psi(k_y)$ is the Digamma function [20]. With this property, the average net SE can be approximated as

$$\mathrm{SE} \approx \frac{T_\mathrm{d}}{T}\frac{1}{\ln 2}\mathbb{E}\Big[\ln(1+\widetilde{\mathrm{SNR}})\Big] \approx \frac{T_\mathrm{d}}{T}\frac{1}{\ln 2}\left(\psi(k_1)+\ln(\theta_1)\right). \quad (12)$$

When the value of M is large, k_1 will be large, and $\psi(k_1) = \ln(k_1) + \mathcal{O}(1/k_1) \approx \ln(k_1)$ [20]. Then, the average net SE can be further approximated as

$$\mathrm{SE} \approx \frac{T_\mathrm{d}}{T}\log_2(k_1\theta_1) = \frac{T_\mathrm{d}}{T}\log_2\left(1+\frac{M\gamma_\mathrm{d}\gamma_u T_\mathrm{tr}}{1+\gamma_u T_\mathrm{tr}}\right) \triangleq \widetilde{\mathrm{SE}}(T_\mathrm{tr}, T_\mathrm{d}, T_\mathrm{id}) \quad (13)$$

which is accurate when the number of antennas M is large, and the average uplink SNR is high. We will show the accuracy of the approximation via simulations later.

Then, the SE-oriented transmission strategy optimization problem, aimed at maximizing the average net SE, can be formulated as

$$\max_{T_\mathrm{tr}, T_\mathrm{d}, T_\mathrm{id}} \quad \widetilde{\mathrm{SE}}(T_\mathrm{tr}, T_\mathrm{d}, T_\mathrm{id}) \quad (14a)$$
$$\text{s.t.} \quad T_\mathrm{tr} + T_\mathrm{d} + T_\mathrm{id} = T \quad (14b)$$
$$T_\mathrm{tr} \in \mathbb{N}^+, \quad T_\mathrm{d}, T_\mathrm{id} \in \mathbb{N} \quad (14c)$$

where $T_\mathrm{tr} \geq 1$ is considered for the closed-loop system.

We use SE_max to denote the optimal value of the objective function of problem (14), which is the maximum achievable average net SE of the system.

To maximize the EE of the considered system supporting delay-tolerant services, the objective function is defined as the ratio of the average net SE to the total power consumed at the BS during a communication period, which is

$$\widetilde{\mathrm{EE}}(T_\mathrm{tr}, T_\mathrm{d}, T_\mathrm{id}) = \frac{\widetilde{\mathrm{SE}}(T_\mathrm{tr}, T_\mathrm{d}, T_\mathrm{id})}{P_\mathrm{tot}} = \frac{T_\mathrm{d}}{P_\mathrm{bu}T_\mathrm{tr}+P_\mathrm{bd}T_\mathrm{d}}\log_2\left(1+\frac{M\gamma_\mathrm{d}\gamma_u T_\mathrm{tr}}{1+\gamma_u T_\mathrm{tr}}\right). \quad (15)$$

The problem to optimize the durations for uplink training, downlink data sending, and BS idling, aimed at maximizing the EE of the system, can be formulated as

$$\max_{T_\mathrm{tr}, T_\mathrm{d}, T_\mathrm{id}} \quad \widetilde{\mathrm{EE}}(T_\mathrm{tr}, T_\mathrm{d}, T_\mathrm{id}) \quad (16a)$$
$$\text{s.t.} \quad \widetilde{\mathrm{SE}}(T_\mathrm{tr}, T_\mathrm{d}, T_\mathrm{id}) \geq SE_\mathrm{min} \quad (16b)$$
$$T_\mathrm{tr} + T_\mathrm{d} + T_\mathrm{id} = T \quad (16c)$$
$$T_\mathrm{tr} \in \mathbb{N}^+, \quad T_\mathrm{d}, T_\mathrm{id} \in \mathbb{N} \quad (16d)$$

where SE_min is the minimum average net SE requirement for the user to support delay-tolerant services.

B. Solution of Optimal Durations

Proposition 1: The optimal solutions of problems (14) and (16) are achieved when $T_\mathrm{tr} + T_\mathrm{d} = T$.

Proof: See Appendix A. ■

It suggests that the BS never operates in idling mode when supporting delay-tolerant services. This is because with a minimal average rate constraint, the EE can be always improved by transmitting more data with a higher rate in a longer duration. Based on Proposition 1, we have $T_\mathrm{d} = T - T_\mathrm{tr}$, then (13) and (15) can be expressed as

$$\widetilde{\mathrm{SE}}(T_\mathrm{tr}) = \frac{T-T_\mathrm{tr}}{T}\log_2\left(1+\frac{M\gamma_\mathrm{d}\gamma_u T_\mathrm{tr}}{1+\gamma_u T_\mathrm{tr}}\right) \quad (17)$$

$$\widetilde{\mathrm{EE}}(T_\mathrm{tr}) = \frac{T-T_\mathrm{tr}}{P_\mathrm{bu}T_\mathrm{tr}+P_\mathrm{bd}(T-T_\mathrm{tr})}\log_2\left(1+\frac{M\gamma_\mathrm{d}\gamma_u T_\mathrm{tr}}{1+\gamma_u T_\mathrm{tr}}\right). \quad (18)$$

Then, problem (16) can be simplified as

$$\max_{T_\mathrm{tr}} \quad \widetilde{\mathrm{EE}}(T_\mathrm{tr}) \quad (19a)$$
$$\text{s.t.} \quad \widetilde{\mathrm{SE}}(T_\mathrm{tr}) \geq SE_\mathrm{min} \quad (19b)$$
$$1 \leq T_\mathrm{tr} \leq T \quad (19c)$$
$$T_\mathrm{tr} \in \mathbb{N}^+. \quad (19d)$$

To solve problem (19), we first remove constraint (19d) by regarding T_tr as a continuous variable within $[1, T]$ and then round up the optimal solution to the nearest integer. The same approach can be used to solve the SE-oriented problem (14). The following proposition shows that such a relaxed version of the problem is convex.

Proposition 2: Both $\widetilde{\mathrm{EE}}(T_\mathrm{tr})$ in (19a) and $\widetilde{\mathrm{SE}}(T_\mathrm{tr})$ in (19b) are concave functions of T_tr, and their first-order derivatives, i.e., $\widetilde{\mathrm{EE}}'(T_\mathrm{tr})$ and $\widetilde{\mathrm{SE}}'(T_\mathrm{tr})$, are monotonically decreasing functions of T_tr.

Proof: See Appendix B. ■

Therefore, the solutions to the relaxed versions of the SE-oriented and EE-oriented optimization problems can be numerically found with efficient algorithms [21]. Although we cannot find their closed-form solutions, we are able to compare the difference in the optimal training durations toward maximizing the EE and toward maximizing the SE.

C. Difference in Training Duration Optimized Toward EE and SE

Define T^*_trSE and T^*_trEE as the optimal uplink training durations of the relaxed versions of the SE-oriented optimization and the EE-oriented optimization, respectively. Denote $\Delta T^*_\mathrm{tr} = T^*_\mathrm{trEE} - T^*_\mathrm{trSE}$ as the difference in the optimal training durations under the two criteria.

1) Impact of QoS Requirement: The following proposition reflects the impact of the QoS requirement on the difference.

Proposition 3: $\Delta T^*_\mathrm{tr} \geq 0$ always holds; the value of ΔT^*_tr remains constant first and then decreases with the increase of the minimum SE requirement, i.e., SE_min; and the equality holds when $SE_\mathrm{min} = SE_\mathrm{max}$.

Proof: See Appendix C. ■

Note that the required minimum average SE by the user does not necessarily equal to $SE_{\max}$. It implies that, in general, the EE-oriented optimization leads to a longer training duration for delay-tolerant services. This is because the power consumption of the BS in receiving mode is smaller than that in transmitting mode so that a longer uplink training duration will reduce the total power consumption of the BS.

2) Impact of Uplink SNR: To show the impact of uplink SNR on the difference in optimal training duration, we examine (17) and (18) in the extreme case where $\gamma_u \to \infty$.

For $T_{\rm tr} \geq 1$, we have $\widetilde{\rm SE}(T_{\rm tr}) \to (1-(T_{\rm tr}/T))\log_2(1+M\gamma_{\rm d})$ from (17) and $\widetilde{\rm EE}(T_{\rm tr}) \to (T-T_{\rm tr})/(P_{\rm bu}T_{\rm tr}+P_{\rm bd}(T-T_{\rm tr}))\log_2(1+M\gamma_{\rm d})$ from (18). The asymptotic expressions of both $\widetilde{\rm SE}(T_{\rm tr})$ and $\widetilde{\rm EE}(T_{\rm tr})$ are monotonically decreasing functions of $T_{\rm tr}$. Therefore, $T^*_{\rm trEE} = T^*_{\rm trSE} = 1$ when $\gamma_u \to \infty$.

This implies that at high uplink SNR, both criteria lead to the same short training duration.

3) Impact of Circuit Power Consumption: To simplify the notations, we set $P_c^{\rm TX} = \lambda P_c^{\rm RX}$, where $\lambda > 1$ means that the circuit power consumed at the BS in transmitting mode is larger than that in receiving mode. Then, the ratio of $P_{\rm bd}$ and $P_{\rm bu}$ can be expressed as

$$\frac{P_{\rm bd}}{P_{\rm bu}} = \frac{\lambda M P_c^{\rm RX} + \mu P_{\rm d}}{M P_c^{\rm RX}} = \lambda + \frac{\mu P_{\rm d}}{M P_c^{\rm RX}}. \tag{20}$$

Substituting (20) into (C.6) in Appendix C, we have

$$\widetilde{\rm SE}'\left(T^o_{\rm trEE}\right) = -\frac{\widetilde{\rm SE}\left(T^o_{\rm trEE}\right)}{\frac{T}{\frac{\mu P_{\rm d}}{M P_c^{\rm RX}}+\lambda-1} + T - T^o_{\rm trEE}} \tag{21}$$

where $T^o_{\rm trEE}$ is the training duration that maximizes $\widetilde{\rm EE}$ without the minimum average SE constraint, as defined in Appendix C.

It is shown from (21) that $\widetilde{\rm SE}'(T^o_{\rm trEE})$ will increase with $P_c^{\rm RX}$ but is always less than zero. Since $\widetilde{\rm SE}'(T_{\rm tr})$ is monotonically decreasing and $\widetilde{\rm SE}'(T^*_{\rm trSE}) = 0$ where $T^*_{\rm trSE}$ is the training duration that maximizes the average SE, $T^o_{\rm trEE}$ will decrease with the increase of $P_c^{\rm RX}$ but is always larger than $T^*_{\rm trSE}$. According to the analysis of the relationship among $T^o_{\rm trEE}$, $T^*_{\rm trEE}$, and $T^*_{\rm trSE}$ in Appendix C, $T^*_{\rm trEE}$ will decrease with the increase of $P_c^{\rm RX}$ but always exceed $T^*_{\rm trSE}$. Therefore, the difference in optimal training duration under the two criteria, $\Delta T^*_{\rm tr}$ will decrease with the increase of the circuit power.

If $\lambda = 1$, i.e., $P_c^{\rm TX} = P_c^{\rm RX}$, $\widetilde{\rm SE}'(T^o_{\rm trEE})$ will approach zero when $P_c^{\rm RX}$ approaches infinity, i.e., the optimal training duration $T^o_{\rm trEE}$ maximizing EE also maximizes $\widetilde{\rm SE}$. Therefore, $\Delta T^*_{\rm tr}$ will approach zero in this case.

In summary, $T^*_{\rm trEE} = T^*_{\rm trSE} = 1$ in the high-uplink-SNR regime, and $\Delta T^*_{\rm tr}$ approaches zero when the minimum SE requirement approaches the maximum achievable SE or when the circuit power is very large and $P_c^{\rm TX} = P_c^{\rm RX}$. In general scenarios, $T^*_{\rm trEE} > T^*_{\rm trSE}$, and $\Delta T^*_{\rm tr}$ decreases with the increase of circuit power.

IV. Delay-Sensitive Services

To achieve high EE and guarantee the QoS of delay-sensitive services, again, we optimize the durations for uplink training, downlink data sending, and BS idling, where the QoS requirement can be characterized by the outage probability for transmitting a given number of bits over the communication period, as shown in Fig. 1.

A. Outage Probability and Problem Formulation

To model the QoS requirement, where a given number of bits (e.g., B bits) should be transmitted within the duration T over bandwidth W with the assistance of $T_{\rm tr}$ training symbols, we define a required net SE, i.e., $SE_r \triangleq B/WT$. Due to channel fluctuation and channel estimate errors, the required number of bits may not be reliably transmitted in the communication period even when the BS and the user employ their maximal transmit power. The outage probability is defined as the probability that the instantaneous net SE, i.e., ${\rm SE}^{\rm in}$, is not larger than the required value of SE_r, i.e.,

$$\begin{aligned} &P_{\rm out}(T_{\rm tr}, T_{\rm d}, T_{\rm id}|SE_r) \\ &\quad = \Pr({\rm SE}^{\rm in} \leq SE_r) \\ &\quad = \Pr\left(\frac{T_{\rm d}}{T}\log_2\left(1+\frac{P_{\rm d}}{\sigma_{\rm d}^2}\left|\|\hat{\mathbf{h}}\|+\frac{\mathbf{e}\hat{\mathbf{h}}^H}{\|\hat{\mathbf{h}}\|}\right|^2\right) \leq SE_r\right). \end{aligned} \tag{22}$$

For notational simplicity, we define $\Psi \triangleq |\|\hat{\mathbf{h}}\| + (\mathbf{e}\hat{\mathbf{h}}^H/\|\hat{\mathbf{h}}\|)|^2$. Noting that $(\mathbf{e}\hat{\mathbf{h}}^H/\|\hat{\mathbf{h}}\|) \sim \mathcal{CN}(0, \sigma_e^2)$, then for a given $\|\hat{\mathbf{h}}\|$, Ψ is a noncentral chi-squared distributed random variable with two degrees of freedom and noncentrality parameter $\|\hat{\mathbf{h}}\|^2/\sigma_e^2$, whose cumulative distribution function (cdf) is denoted by $F_{(nc-\chi^2,2)}(\cdot)$. Therefore, the outage probability for a given $\|\hat{\mathbf{h}}\|$ can be expressed as

$$\begin{aligned} P_{\rm out}\left(T_{\rm tr}, T_{\rm d}, T_{\rm id}|SE_r, \|\hat{\mathbf{h}}\|\right) &= \Pr\left(\Psi \leq \frac{2^{\frac{SE_rT}{T_{\rm d}}}-1}{\frac{P_{\rm d}}{\sigma_{\rm d}^2}}\right) \\ &= F_{(nc-\chi^2,2)}\left(\frac{2^{\frac{SE_rT}{T_{\rm d}}}-1}{\frac{P_{\rm d}}{\sigma_{\rm d}^2}}\right). \end{aligned} \tag{23}$$

Further considering that $\|\hat{\mathbf{h}}\|$ is a central chi-distributed random variable with $2M$ degrees of freedom, from the proof in [22, App.], we can obtain the outage probability as

$$\begin{aligned} &P_{\rm out}(T_{\rm tr}, T_{\rm d}, T_{\rm id}|SE_r) \\ &\quad = \frac{\sum_{i=0}^{M-1}\binom{M-1}{i}(\gamma_uT_{\rm tr})^i\Gamma_{i+1}\left(\frac{2^{\frac{SE_rT}{T_{\rm d}}}-1}{\gamma_{\rm d}}\right)}{(1+\gamma_uT_{\rm tr})^{M-1}} \end{aligned} \tag{24}$$

where $\binom{M-1}{i}$ is the number of combinations, and $\Gamma_M(x) = 1/(M-1)!\int_0^x t^{M-1}e^{-t}dt$ is the incomplete Gamma function.

To maximize the EE of the considered system supporting delay-sensitive services, the objective function is defined as $SE_r/P_{\rm tot}$. Denoting ϵ as the maximum acceptable outage probability, then the optimization problem can be formulated as

$$\max_{T_{\rm tr},T_{\rm d},T_{\rm id}} \quad \frac{SE_r}{P_{\rm bu}T_{\rm tr}+P_{\rm bd}T_{\rm d}} \tag{25a}$$

$$\text{s.t.} \quad P_{\rm out}(T_{\rm tr},T_{\rm d},T_{\rm id}|SE_r)\le\epsilon \tag{25b}$$

$$T_{\rm tr}+T_{\rm d}+T_{\rm id}=T \tag{25c}$$

$$T_{\rm tr}\in\mathbb{N}^+,\ T_{\rm d},\ T_{\rm id}\in\mathbb{N}. \tag{25d}$$

Since SE_r is a predetermined value, problem (25) is equivalent to the following problem:

$$\min_{T_{\rm tr},T_{\rm d},T_{\rm id}} \quad P_{\rm bu}T_{\rm tr}+P_{\rm bd}T_{\rm d} \tag{26a}$$

$$\text{s.t.} \quad P_{\rm out}(T_{\rm tr},T_{\rm d},T_{\rm id}|SE_r)\le\epsilon \tag{26b}$$

$$T_{\rm tr}+T_{\rm d}+T_{\rm id}=T \tag{26c}$$

$$T_{\rm tr}\in\mathbb{N}^+,\ T_{\rm d},\ T_{\rm id}\in\mathbb{N}. \tag{26d}$$

B. Solution of Optimal Durations

The optimization problem (26) is hard to solve, because the expression of the outage probability in (24) is rather involved. To circumvent this problem, we approximate the outage probability by using the approximate SNR in (10), which can be rewritten as

$$\widetilde{\rm SNR}=\frac{P_{\rm d}}{\sigma_{\rm d}^2}\|\hat{\mathbf{h}}\|^2=\frac{\gamma_{\rm d}\gamma_u T_{\rm tr}}{2(1+\gamma_u T_{\rm tr})}\cdot\nu \tag{27}$$

where $\nu\triangleq 2\|\hat{\mathbf{h}}\|^2/\sigma_{\hat{h}}^2$ is a chi-squared distributed variable with $2M$ degrees of freedom. Then, from (22), the outage probability can be approximated as

$$\begin{aligned}&\widetilde{P_{\rm out}}(T_{\rm tr},T_{\rm d},T_{\rm id}|SE_r)\\&=\Pr\left(\frac{T_{\rm d}}{T}\log_2(1+\widetilde{\rm SNR})\le SE_r\right)\\&=\Pr\left(\nu\le\frac{2}{\gamma_{\rm d}}\left(1+\frac{1}{\gamma_u T_{\rm tr}}\right)\left(2^{\frac{SE_rT}{T_{\rm d}}}-1\right)\right)\\&=F_{(\chi^2,2M)}\left(\frac{2}{\gamma_{\rm d}}\left(1+\frac{1}{\gamma_u T_{\rm tr}}\right)\left(2^{\frac{SE_rT}{T_{\rm d}}}-1\right)\right)\end{aligned} \tag{28}$$

where $F_{(\chi^2,2M)}(\cdot)$ is the cdf of a chi-square random variable with $2M$ degrees of freedom.

Further considering the monotonicity of $F_{(\chi^2,2M)}(\cdot)$, problem (26) can be reformulated as

$$\min_{T_{\rm tr},T_{\rm d},T_{\rm id}} \quad P_{\rm bu}T_{\rm tr}+P_{\rm bd}T_{\rm d} \tag{29a}$$

$$\text{s.t.} \quad \frac{2}{\gamma_{\rm d}}\left(1+\frac{1}{\gamma_u T_{\rm tr}}\right)\left(2^{\frac{SE_rT}{T_{\rm d}}}-1\right)\le\nu_c \tag{29b}$$

$$T_{\rm tr}+T_{\rm d}+T_{\rm id}=T \tag{29c}$$

$$T_{\rm tr}\in\mathbb{N}^+,\ T_{\rm d},\ T_{\rm id}\in\mathbb{N} \tag{29d}$$

where $\nu_c\triangleq F_{(\chi^2,2M)}^{-1}(\epsilon)$ is a constant determined by ϵ, and $F_{(\chi^2,2M)}^{-1}(\cdot)$ represents the inverse function of $F_{(\chi^2,2M)}(\cdot)$.

To solve problem (29), we first replace $T_{\rm id}$ by $T_{\rm id}=T-T_{\rm tr}-T_{\rm d}$ from (29c) and then relax $T_{\rm d}$ as a continuous variable within $[0,T]$ and $T_{\rm tr}$ as a continuous variable within $[\varsigma,T]$ with ς denoting a very small positive number near zero, because we consider $T_{\rm tr}\ge 1$ for the closed-form system. After solving the relaxed problem, we round up the optimal solution of $T_{\rm tr}$ to the nearest positive integer and find the minimal integer around the optimal solution of $T_{\rm d}$ to ensure the QoS constraint (29b) is satisfied. In the sequel, we show that the relaxed version of problem (29) is convex.

Proposition 4: The optimal solution of the relaxed version of problem (29) makes constraint (29b) hold with equality.

Proof: See Appendix D. ■

According to Proposition 4, the duration for downlink data sending can be expressed as

$$T_{\rm d}=\frac{SE_rT}{\log_2\left(1+\frac{\gamma_{\rm d}\gamma_u\nu_c}{2}\frac{T_{\rm tr}}{1+\gamma_uT_{\rm tr}}\right)} \tag{30}$$

then problem (29) can be relaxed by omitting the constraints in (29d) as

$$\min_{T_{\rm tr}} \quad P_{\rm bu}T_{\rm tr}+P_{\rm bd}\frac{SE_rT}{\log_2\left(1+\frac{\gamma_{\rm d}\gamma_u\nu_c}{2}\frac{T_{\rm tr}}{1+\gamma_uT_{\rm tr}}\right)} \tag{31a}$$

$$\text{s.t.} \quad T_{\rm tr}+\frac{SE_rT}{\log_2\left(1+\frac{\gamma_{\rm d}\gamma_u\nu_c}{2}\frac{T_{\rm tr}}{1+\gamma_uT_{\rm tr}}\right)}\le T \tag{31b}$$

$$\varsigma\le T_{\rm tr}\le T \tag{31c}$$

where (30) is used to transform the equality constraint (29c) into the inequality constraint (31b).

For the sake of notational simplicity, we denote (31a) and the left-hand side of the inequality constraint (31b) as

$$\begin{aligned}h(T_{\rm tr})&\triangleq P_{\rm bu}T_{\rm tr}+P_{\rm bd}\frac{SE_rT}{\log_2\left(1+\frac{\gamma_{\rm d}\gamma_u\nu_c}{2}\frac{T_{\rm tr}}{1+\gamma_uT_{\rm tr}}\right)}\\&=a_1T_{\rm tr}+\frac{a_2}{\log_2\left(1+\frac{b_1T_{\rm tr}}{1+b_2T_{\rm tr}}\right)}\end{aligned} \tag{32}$$

$$c(T_{\rm tr})\triangleq T_{\rm tr}+\frac{SE_rT}{\log_2\left(1+\frac{\gamma_{\rm d}\nu_c}{2}\frac{\gamma_uT_{\rm tr}}{1+\gamma_uT_{\rm tr}}\right)} \tag{33}$$

where a_1, a_2, b_1, and b_2 denote $P_{\rm bu}$, $P_{\rm bd}SE_rT$, $\gamma_{\rm d}\gamma_u\nu_c/2$, and γ_u, respectively. In the following proposition, we show that problem (31) is convex.

Proposition 5: Both $h(T_{\rm tr})$ in (32) and $c(T_{\rm tr})$ in (33) are convex functions of $T_{\rm tr}$.

Proof: See Appendix E. ■

It follows that the relaxed problem (31) can be numerically solved with efficient convex optimization algorithms [21].

C. Impact of QoS Requirement, SNR, and Circuit Power on the Optimal Solution

We first examine the properties of the objective function, i.e., $h(T_{\rm tr})$, and the left-hand side of the inequality constraint (31b), i.e., $c(T_{\rm tr})$, as follows.

When $T_{\rm tr} \to 0$, the term $\log_2(1+(b_1T_{\rm tr}/1+b_2T_{\rm tr}))$ in (E.1) given in Appendix E approaches zero, hence the first-order derivative $h'(T_{\rm tr}) \to -\infty$ as shown in (E.1). When $T_{\rm tr} \to T$, $T_{\rm d}$ will approach zero considering constraint (29c). Further noting from (30) that $T_{\rm d} = SE_rT/\log_2(1 + (\gamma_{\rm d}\gamma_u\nu_c/2)T_{\rm tr}/(1 + \gamma_uT_{\rm tr})) = SE_rT/\log_2(1 + b_1T_{\rm tr}/(1 + b_2T_{\rm tr}))$, we know that $\log_2(1 + b_1T_{\rm tr}/(1 + b_2T_{\rm tr}))$ will approach infinity when $T_{\rm tr} \to T$. Substituting it into (E.1), we know that $h'(T_{\rm tr})$ will approach a_1, which is a positive number. Since $h''(T_{\rm tr}) > 0$ always holds as shown in (E.2), it is not hard to find that $h(T_{\rm tr})$ first decreases and then increases within the duration $(0, T)$. Similarly, we can prove that $c(T_{\rm tr})$ also decreases at first and then increases within the duration $(0, T)$.

Denote $[T_{\rm tr}^{\rm low}, T_{\rm tr}^{\rm up}]$ as the feasible set of $T_{\rm tr}$ for problem (31), which is the intersection of the sets defined by its two constraints (30b) and (30c). Let $[z^{\rm low}, z^{\rm up}]$ denote the set defined by constraint (30c), i.e., $\{T_{\rm tr}|c(T_{\rm tr}) \le T\}$. Since we have shown that $c(T_{\rm tr})$ is convex and decreases at first and then increases within the duration $(0, T)$, further noting that $c(\varsigma) > T$ and $c(T) > T$, we can obtain that $z^{\rm low}$ and $z^{\rm up}$ are the two solutions to the equation $c(T_{\rm tr}) = T$, and $\varsigma < z^{\rm low} \le z^{\rm up} < T$, i.e., $[z^{\rm low}, z^{\rm up}]$ is a subset of $[\varsigma, T]$ defined by (30b). Therefore, we obtain that $T_{\rm tr}^{\rm low} = z^{\rm low}$ and $T_{\rm tr}^{\rm up} = z^{\rm up}$.

Denote $T_{\rm tr}^v$ as the duration for uplink training that minimizes $h(T_{\rm tr})$ without constraint (31b) and $T_{\rm tr}^c$ as the duration for uplink training that minimizes $c(T_{\rm tr})$, respectively. Since both $h(T_{\rm tr})$ and $c(T_{\rm tr})$ are convex functions and they first decrease and then increase within the duration $(0, T)$, we have $h'(T_{\rm tr}^v) = 0$ and $c'(T_{\rm tr}^c) = 0$. It is easy to see that $T_{\rm tr}^c \in [T_{\rm tr}^{\rm low}, T_{\rm tr}^{\rm up}]$ since $T_{\rm tr}^c$ minimizes $c(T_{\rm tr})$ so that the inequality constraint in (31b) is satisfied.

Let $T^*_{\rm trEEout}$, $T^*_{\rm dEEout}$, and $T^*_{\rm idEEout}$ denote the optimal solutions of the relaxed problem (31). We have the following proposition.

Proposition 6:

- $T_{\rm tr}^v \ge T_{\rm tr}^c$ always holds, and the equality holds if and only if $SE_r = 0$.
- Denote SE_r^o as the SE_r with which $T_{\rm tr}^v = T_{\rm tr}^{\rm up}$ holds. Then, when $SE_r \le SE_r^o$, $T^*_{\rm trEEout} = T_{\rm tr}^v$ that increases with SE_r; when $SE_r > SE_r^o$, $T^*_{\rm trEEout} = T_{\rm tr}^{\rm up}$ that decreases with SE_r.
- When $P_c^{\rm TX} = P_c^{\rm RX}$ and both of them are sufficiently large, we have $T_{\rm tr}^v \le T_{\rm tr}^{\rm up}$, and $T^*_{\rm trEEout} = T_{\rm tr}^v$ that increases with SE_r and decreases with circuit power.

Proof: See Appendix F. ■

Due to the difficulty in finding a closed-form solution of $T^*_{\rm trEEout}$, it is not easy to analyze the impact of general average uplink SNR on $T^*_{\rm trEEout}$. However, we can gain some insight from the results at high average uplink SNR. When $\gamma_u \to \infty$, from (30), the optimal duration for downlink data sending becomes $T^*_{\rm dEEout} = SE_rT/\log_2(1 + (\gamma_{\rm d}\nu_c/2))$, which increases with the SE requirement linearly and decreases with the increase of average downlink SNR. Since $T^*_{\rm dEEout}$ is independent of $T^*_{\rm trEEout}$ in this case, by substituting it into the optimization problem (31), we can obtain that $T^*_{\rm trEEout} = \varsigma$. It suggests that the energy-efficient strategy for delay-sensitive services reduces the resources for uplink training to increase the opportunity of BS idling, which is different from the strategy for delay-tolerant services that keeps the BS active all the time.

Remark 1: The different behaviors for BS idling under the two kinds of services can be explained as follows. For delay-tolerant services, the user has only a minimal average rate constraint. Then, for any given duration for uplink training, the BS can always improve the system EE by transmitting more data in a longer duration [i.e., the EE is an increasing function of $T_{\rm d}$ for any given $T_{\rm tr}$, as shown by (15)], which can reduce the impact of the circuit power consumed for receiving uplink training signals at the BS. Therefore, BS idling is not needed in this case. For delay-sensitive services, the amount of data to be transmitted is fixed in the given duration; therefore, the BS can switch into the idle mode to reduce the power consumption when all the data have been transmitted.

Remark 2: The proposed transmission strategy for delay-sensitive services cannot be directly applied to the real-time video service with a time-varying data rate. A simple way to handle this problem is to consider the worst-case design. For example, we can optimize the duration for uplink training before the packets arrive based on the highest possible data rate of real-time video services. Then, given the uplink training, one can determine the durations for downlink data sending and idling for each arrived packet.

V. Discussions and Extensions

In the previous sections, we have designed transmission strategies for two kinds of services in a single-user scenario. Here, we analyze the challenges and opportunities of applying the proposed strategies to current and future cellular systems and provide one way to extend the single-user single-service designs to the multiuser scenario with mixed services.

A. Challenges and Opportunities

The designed transmission strategies in previous sections cannot be applied to existing cellular systems, e.g., the long-term evolution (LTE) system, due to the following reasons. First, the proposed method optimizes the duration for uplink training, which may exceed the maximal uplink training duration of the LTE system. Second, a macro or micro BS in the LTE system often serves multiple users with mixed traffics. With the proposed single-user designs, the optimized durations for uplink training, downlink data sending, and BS idling may be nonidentical for different users, which will lead to the echo interference at the BS caused by transmitting to and receiving from different users simultaneously. Third, the proposed method considers the symbol-level BS idling, and different users may require different BS idling durations. This requires the BS to operate in idle mode for one user and in active mode for another user simultaneously.

Nevertheless, these challenges can be overcome with the emerging advanced technologies for future cellular systems. First, the ultradense networks lead to very small coverage of each BS, where the number of users in every small cell will be very limited, and the user-specific system parameter configuration will become possible [23]. Second, effective echo interference cancelation techniques enable the BS to transmit and receive simultaneously in a full-duplex mode [24]. Third, the fast hardware deactivating/reactivating technology supports the symbol-level BS idling, and the bandwidth adaptation technology enables BS idling in the frequency domain, which allows multiple users to have different BS idling durations in the time domain [15].

B. Extension to Multiple Users With Mixed Services

In the following, we provide one way to extend the proposed strategies to the multiuser scenario with mixed services, which takes into account the constraint on the maximal uplink training duration and does not need the echo interference cancelation techniques.

Assume that the BS can be idle in both frequency and time domains. BS idling in the frequency domain is also known as bandwidth adaptation, which can reduce the maximum RF output power of the BS by adaptively adjusting the occupied bandwidth according to the QoS requirement when given the power spectral density of the signals. Then, the operating point of the PA can be adapted to the required output power for saving energy [25]. BS idling in the time domain saves energy by deactivating power-consuming hardware components when there are no data to transmit. Although BS idling schemes in frequency and time domains are based on different principles, they perform similarly in energy saving in practice as analyzed in [25].

We consider the orthogonal-frequency-division-multiple-access-based multiuser transmission, which is a widely used technique for broadband cellular systems. Suppose that the BS serves K users with mixed services, where the users occupy orthogonal subcarriers. Let $T_{\rm UL}$ and $T_{\rm DL}$ denote the durations of uplink and downlink subframes of the system, and $T_{\rm UL}^k < T_{\rm UL}$ denote the duration that can be used to transmit uplink training for user k. The remaining uplink duration, i.e., $T_{\rm UL} - T_{\rm UL}^k$, is used by user k to transmit its uplink data. Define $T^k = T_{\rm UL}^k + T_{\rm DL}$ as the total communication period of user k and $T_{\rm tr}^k$, $T_{\rm d}^k$, and $T_{\rm id}^k$ as the durations for uplink training, downlink data sending, and BS idling of user k. It should be noted that BS idling can occur not only in the downlink subframe but also in the uplink subframe when $T_{\rm tr}^k < T_{\rm UL}^k$. Then, the transmission strategy optimization problem for user k can be formulated as follows:

$$\max_{T_{\rm tr}^k, T_{\rm d}^k, T_{\rm id}^k} \quad EE_k \tag{34a}$$

$$\text{s.t.} \quad \text{QoS requirement of user } k \tag{34b}$$

$$T_{\rm tr}^k + T_{\rm d}^k + T_{\rm id}^k = T^k \tag{34c}$$

$$T_{\rm tr}^k \in \mathbb{N}^+,\ T_{\rm d}^k,\ T_{\rm id}^k \in \mathbb{N} \tag{34d}$$

$$T_{\rm tr}^k \le T_{\rm UL}^k \tag{34e}$$

$$T_{\rm d}^k \le T_{\rm DL} \tag{34f}$$

where objective function (34a) and QoS constraint (34b) are determined by the service type of user k, which are given by (16) for delay-tolerant services and by (25) for delay-sensitive services.

Compared with optimization problems (16) and (25) for the single-user scenario, we add two new constraints (34e) and (34f). Since both constraints are convex, we can use the same approaches for problems (16) and (25) to solve problem (34).

In the transmission mechanism, we limit the uplink training duration of all users not to exceed the common duration of uplink subframe $T_{\rm UL}$ and the downlink data sending duration of all users not to exceed the common downlink subframe $T_{\rm DL}$. In this manner, echo interference at the BS can be avoided, and the transmission strategies of multiple users can be separately optimized.

Remark 3: In the paper, we suppose that the transmit power of the BS and the user is constant. The power consumption of the user is not taken into account in the EE of the downlink system. When the transmit power of a user can be adjusted, it is not hard to find that the EE is maximized when the user transmits with its maximal power. This implies that the transmit power of the user will be constant regardless of the QoS requirement or SNR. Therefore, transmit power control at the user side will not affect the results we obtained. On the other hand, transmit power control at the BS side provides more freedom to balance the transmit and circuit power and, hence, will improve the EE. However, the joint optimization of the transmit power at each time slot and the durations for uplink training, downlink data sending, and BS idling is nonconvex and difficult to solve. This makes the analysis of the impact of transmit power control at the BS on the performance of the proposed transmission strategies very complicated. We will leave this interesting problem for future work.

VI. Simulation Results

Here, we use simulations to validate the analytical analysis and evaluate the maximum EE for different services. Unless otherwise specified, in the simulations, we assume that the BS has $M = 4$ antennas with the transmit power of 46 dBm, the users transmit with 23 dBm, the communication period T is set to 140 symbols corresponding to the duration of one LTE frame, and the maximum outage probability for delay-sensitive services is 5%. The power consumption parameters are configured as in [1] for a macro BS, where the per-antenna circuit power of the BS in transmitting and receiving modes is set to 47.1 W, and the parameter reflecting transmit power loss, i.e., μ, is set as 4.24. The path-loss model is set as $35.3 + 37.6 \log_{10}(d)$, where d is the distance between the BS and the user in meters [26]. Since the average downlink SNR is usually higher than the average uplink SNR, we set $\gamma_{\rm d} = \gamma_u + 10$ dB.

In the following, we first evaluate the accuracy of the approximations, validate the analysis for the optimal durations, and show the EE gain by simulating the single-user scenario under the two kinds of services, where each BS serves only one user. After that, we show the EE gain for the multiuser scenario by simulating a case where each BS serves two users with mixed services. Following the previous definitions, we use $SE_{\min}$ and

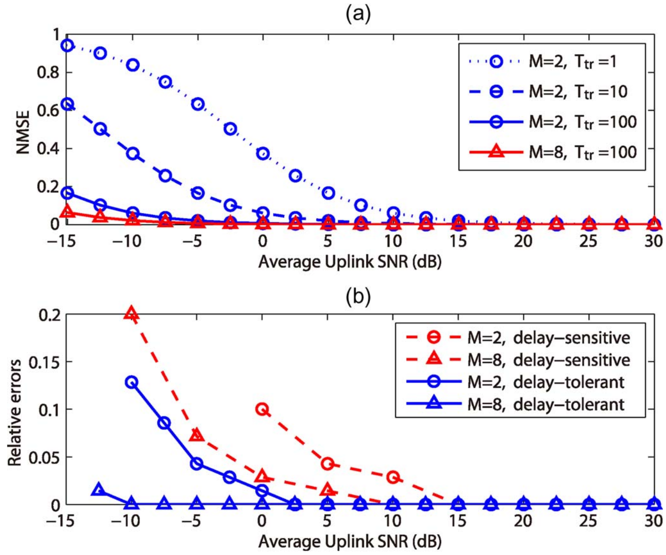


Fig. 2. NMSE of the approximate SNR and relative errors of the designed durations, where $SE_{\min} = SE_r = 1$ b/s/Hz is considered in Fig. 2(b).

SE_r to denote the data rate requirements of a user with delay-tolerant and delay-sensitive services, respectively. Herein, both $SE_{\min}$ and SE_r are normalized by the bandwidth occupied by the user, or in other words, unit bandwidth per user is supposed in the simulations. For the two kinds of services, different packet schedulers are assumed based on the QoS requirements as described in Section II-C, with which the BS periodically removes some of the data from the buffer of a user and transmits to the user according to the optimized transmission strategies.

A. Accuracy of the Approximations

We first examine the accuracy of the SNR approximation given in (10), which are used in the optimizations for both delay-tolerant and delay-sensitive services. In Fig. 2(a), the normalized mean square error (NMSE) of the SNR approximation, which is defined by $\mathbb{E}\{|\mathrm{SNR} - \widetilde{\mathrm{SNR}}|^2\}/\mathbb{E}\{|\mathrm{SNR}|^2\}$, is plotted, where different uplink SNRs γ_u, number of antennas M, and uplink training durations T_{tr} are considered. It is shown that the accuracy of the SNR approximation is poor when γ_u and M are small and improves with the growth of γ_u and M. Moreover, we find that the NMSE decreases with the increase of T_{tr}, because increasing uplink training duration improves the quality of channel estimation. Note that in our work, the value of T_{tr} is not fixed, which adapts to the SNR and QoS requirements. A low SNR generally leads to a large T_{tr}, which can compensate the accuracy degradation of the SNR approximation to a certain extent in the low-SNR regime. To see this effect, we then plot the relative errors between the designed durations and the optimal durations, defined by $(|T_{\mathrm{tr}}^{\mathrm{opt}} - T_{\mathrm{tr}}| + |T_{\mathrm{d}}^{\mathrm{opt}} - T_{\mathrm{d}}| + |T_{\mathrm{id}}^{\mathrm{opt}} - T_{\mathrm{id}}|)/T$, in Fig. 2(b), where T_{tr}, T_{d}, and T_{id} denote the designed durations based on both the SNR approximation and the moment matching for Gamma distribution, and $T_{\mathrm{tr}}^{\mathrm{opt}}$, $T_{\mathrm{d}}^{\mathrm{opt}}$, and $T_{\mathrm{id}}^{\mathrm{opt}}$ are the optimal durations that are obtained by exhaustive searching over all possible combinations of the durations for uplink training, downlink data sending, and BS idling. During the exhaustive search, for each combination, we obtain the average net SE by Monte Carlo methods and compute the outage probability based on the exact expression (24); therefore, the optimal durations are not based on any approximations. It is shown that the errors caused by the approximations are small (less than 0.2) even for small γ_u and M. We can also see that the relative errors for delay-sensitive services are larger than those for delay-tolerant services, because the delay-sensitive services usually have a shorter uplink training duration to increase BS idling duration, whereas BS idling never occurs for delay-tolerant services. The impact of the duration errors on the system EE is evaluated in Fig. 3, where the EEs corresponding to the designed durations and the optimal durations are plotted. We can see that the small duration errors have a negligible impact on the EE for delay-tolerant services, and only a very small gap can be observed for delay-sensitive services under small γ_u and M.

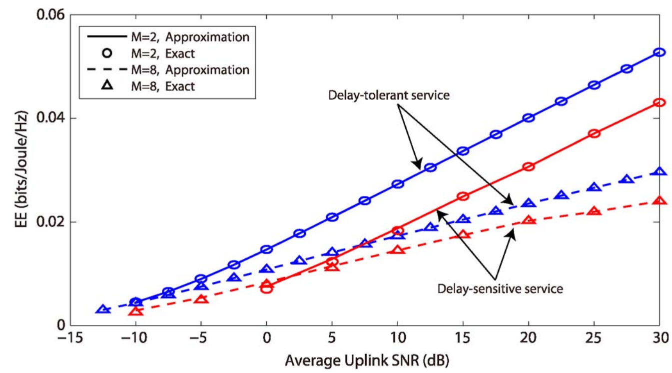


Fig. 3. Exact and approximate maximum EE versus the average uplink SNR with $SE_{\min} = SE_r = 1$ b/s/Hz.

B. Validation on the Analysis of Optimal Transmission Durations

Here, we validate the analytical analysis of the impact of QoS requirement, SNR, and circuit power on the optimal transmission durations for delay-tolerant and delay-sensitive services, respectively.

1) Delay-Tolerant Services: Fig. 4(a) shows the impact of SNR on the uplink training durations optimized toward SE and EE. We can see that when the SNR exceeds 15 dB, $T_{\mathrm{trEE}}^* = T_{\mathrm{trSE}}^* = 1$, and $\Delta T_{\mathrm{tr}}^* = 0$. In general cases, the EE-oriented optimization requires more training symbols than the SE-oriented optimization. In Fig. 4(b), we show the impact of circuit power on ΔT_{tr}^*. It is shown that when the circuit power is small, there is an obvious difference in training duration between the EE-oriented and the SE-oriented optimization, and the difference decreases with the increase of the circuit power. When $P_c^{\mathrm{TX}} = P_c^{\mathrm{RX}}$ and the circuit power is very large, $\Delta T_{\mathrm{tr}}^* \to 0$. These results agree well with our analytical analysis.

2) Delay-Sensitive Services: Fig. 5(a) shows the optimal duration for uplink training versus the QoS requirement with different values of circuit power. We can see that when circuit power is very large (e.g., $P_c^{\mathrm{RX}} = 50$ W), the optimal training duration increases with SE_r. In general, when SE_r is small, the optimal training duration increases with the increase of

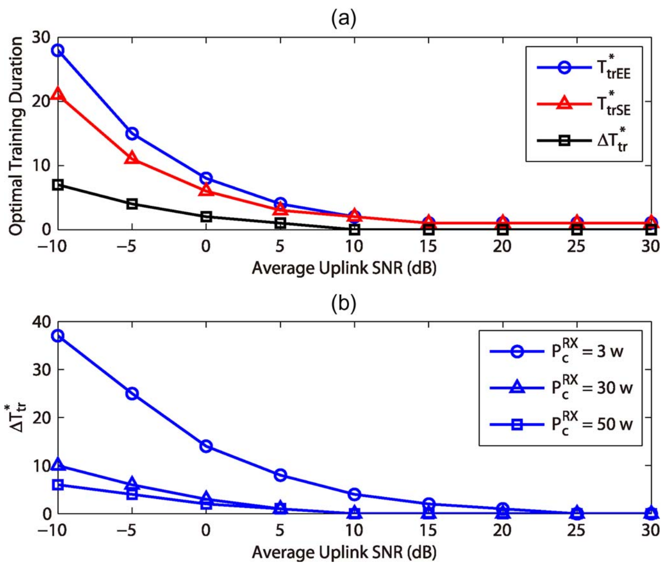


Fig. 4. Optimal training duration and optimal training duration difference for delay-tolerant services with $SE_{\min} = 1$ b/s/Hz.

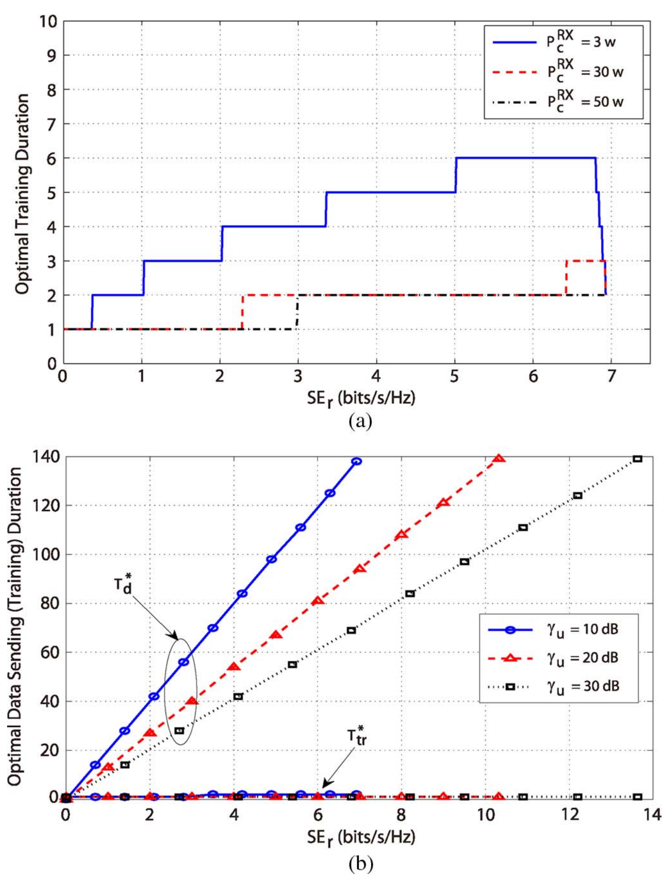


Fig. 5. Optimal durations for uplink training and downlink data sending versus net SE requirement for delay-sensitive services. (a) Optimal training duration with $\gamma_u = 10$ dB. (b) Optimal data sending duration.

SE_r and the decrease of $P_c^{\rm RX}$. When SE_r is large, the optimal training duration decreases with the increase of SE requirement. Fig. 5(b) shows the optimal duration for downlink data sending and uplink training versus the QoS requirement with different values of SNR. It shows that when the SNR is very high, e.g., 20 and 30 dB, one training symbol is optimal. The optimal

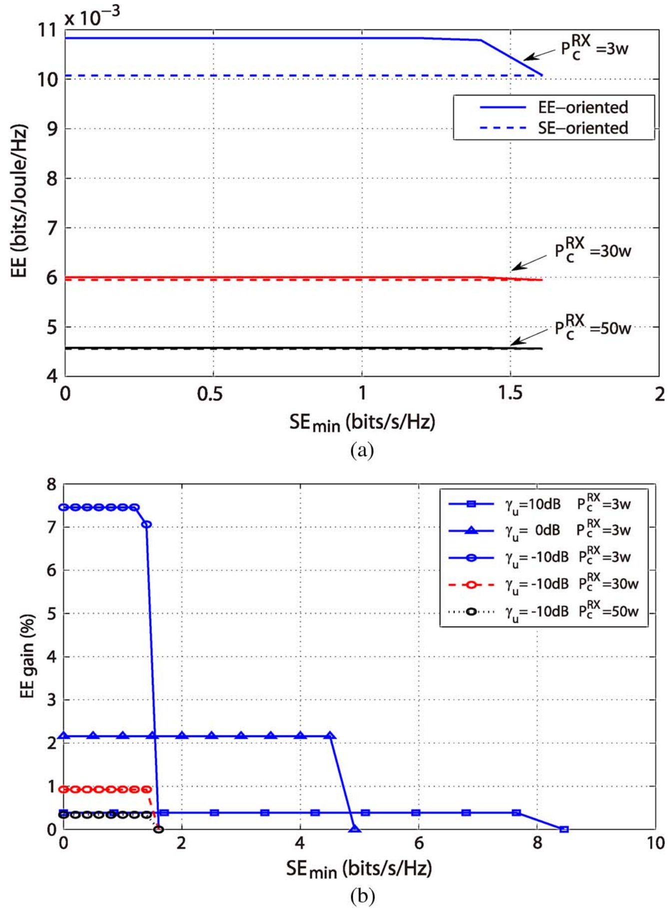
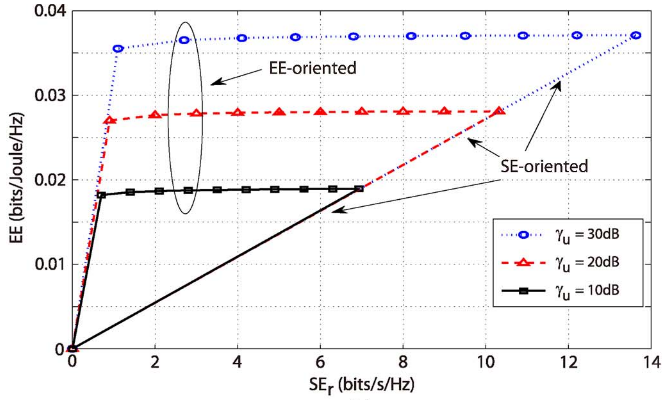


Fig. 6. Maximum EE achieved by the EE-oriented and SE-oriented optimization and EE gain of the EE-oriented optimization over the SE-oriented optimization for delay-tolerant services. (a) Maximum EE with $\gamma_u = -10$ dB. (b) EE gain.

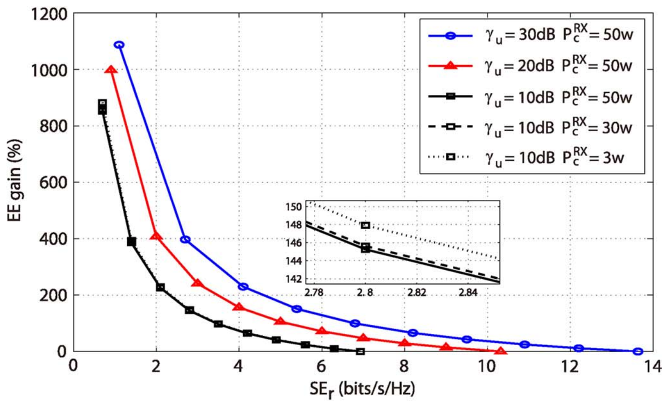
transmit duration increases with SE_r linearly and decreases with the increase of SNR. These results agree well with the analysis in Section IV.

C. Maximum EE and EE Gain Versus QoS Requirement

1) Delay-Tolerant Services: In Fig. 6(a), we compare the EE achieved by the EE-oriented optimization (16) and the SE-oriented optimization that maximizes the SE in (14). Since the optimal training duration maximizing the SE is independent of $SE_{\min}$, the corresponding EE is a constant. When the circuit power is small, we can see that the EE achieved by the EE-oriented optimization first remains constant and then decreases to the EE achieved by SE-oriented optimization when $SE_{\min} = SE_{\max}$, as indicated by Proposition 3. Fig. 6(b) shows the EE gain of the EE-oriented optimization over the SE-oriented optimization, which is defined as $(\text{EE}(T^*_{\rm trEE}) - \text{EE}(T^*_{\rm trSE}))/\text{EE}(T^*_{\rm trSE})$, where $\text{EE}(T^*_{\rm trEE})$ and $\text{EE}(T^*_{\rm trSE})$ are, respectively, the EE achieved by the EE-oriented and the SE-oriented optimization following the definitions in Section III. It is shown that the EE gain first remains constant and then decreases to zero with the increase of the minimum net SE requirement. Moreover, the EE gain decreases with the growth of circuit power, because the optimal training duration



(a)



(b)

Fig. 7. Maximum EE achieved by the EE-oriented and SE-oriented optimization and EE gain of the EE-oriented optimization over the SE-oriented optimization for delay-sensitive services. (a) Maximum EE with $P_c^{\mathrm{RX}} = 50$ W. (b) EE gain.

difference under the two criteria decreases with the increase of circuit power, as shown in Fig. 4(b). The EE gain decreases with the increase of SNR and approaches zero when the SNR is sufficiently high, i.e., the EE-oriented optimization reduces to the SE-oriented optimization in this case, which agrees with our previous analysis.

2) Delay-Sensitive Services: In Fig. 7(a), we compare the EE achieved by the EE-oriented optimization (25) and by the SE-oriented optimization. Again, because the optimal training and data sending durations maximizing the SE do not depend on the QoS requirement, the power consumption is fixed, and the achieved EE increases with the QoS requirement linearly. We can see that the EE achieved by the EE-oriented optimization is always higher than that by the SE-oriented optimization, unless the SE requirement is equal to zero or equal to the maximum achievable SE. In Fig. 7(b), we evaluate the EE gain of the EE-oriented optimization over the SE-oriented optimization. It is shown that the EE gain is remarkable at low SE requirements. The EE gain improves with the increase of SNR. This is because the system can satisfy the SE requirement with fewer downlink data sending symbols for high SNR and low QoS requirement. It is also shown that higher EE gain can be achieved with lower circuit power, but the impact is relatively little. This is because the impact of circuit power mainly comes from uplink training phase; however, the optimal training duration is short when SNR is high.

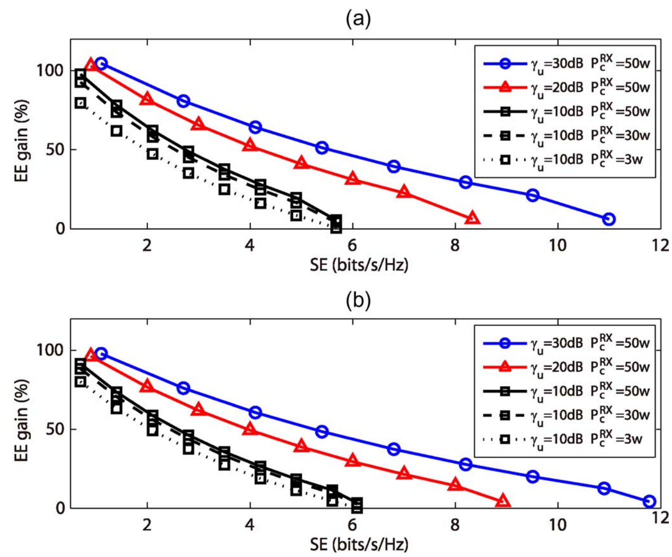


Fig. 8. EE gain of the EE-oriented optimization over the SE-oriented optimization for $K = 2$ users and mixed services, where $SE_{\min} = SE_r$ are given in the x-axis, and $T_{\mathrm{UL}}^k : T_{\mathrm{DL}} = 1 : 4$ and $1 : 6$ are considered in Fig. 8(a) and (b).

D. EE Gain for Multiple Users and Mixed Services

Here, we evaluate the EE gain of the proposed EE-oriented optimization over the SE-oriented optimization in the scenarios with multiple users and mixed services. Specifically, we consider that the BS serves $K = 2$ users that occupy orthogonal subcarriers, where user 1 is with the delay-tolerant service, user 2 is with the delay-sensitive service, and $SE_{\min} = SE_r$ is assumed for the two kinds of services. The total communication periods of the two users are set as $T^1 = T^2 = 140$, within which the maximal uplink training durations for the two users are set to the same, i.e., $T_{\mathrm{UL}}^1 = T_{\mathrm{UL}}^2$. To reflect different configurations of uplink–downlink subframe durations, we consider $T_{\mathrm{UL}}^k / T_{\mathrm{DL}} = 1/4$ as 1/6 in Fig. 8(a) and (b).[1] We can see that under both configurations, the EE gain is evident particularly for small SE requirements. Moreover, given the SE requirement, the EE gain increases with the SNR. The result coincides with the analysis in Fig. 7(b) for the pure delay-sensitive services but contradicts with the result in Fig. 6(b) for the pure delay-tolerant services, where the EE gain decreases with the SNR. The difference comes from the new constraint introduced for the multiuser case that downlink data sending cannot use the time slots in the uplink subframe. This leads to the result that even for the delay-tolerant services, BS idling may occur in some uplink time slots to reduce the receiving power consumption, while we have proved that this will not happen in the single-user case. A higher SNR will reduce the uplink training duration, which leads to the increase of the BS idling duration in the uplink subframe and, hence, achieves a larger EE gain. Similarly, the BS idling duration in the uplink subframe will increase with the growth of circuit power to save more power, which leads to the improvement of the EE gain, as

[1]The values of $T_{\mathrm{UL}}^k / T_{\mathrm{DL}}$ are smaller than the configurations of uplink–downlink subframe durations because $T_{\mathrm{UL}}^k < T_{\mathrm{UL}}$ is only the uplink training duration.

shown in the figures. In addition, by comparing Fig. 8(a) and (b), we can find that a longer downlink subframe can support a higher SE requirement.

VII. Conclusion

In this paper, we have investigated the energy-efficient transmission strategies for downlink TDD closed-loop MISO systems, where the three-phase communication is considered consisting of uplink training, downlink data sending, and BS idling. To optimize the durations of the three phases aimed at maximizing the downlink EE while guaranteeing the user's QoS, we first derived the closed-form expressions of the approximate average net SE and outage probability, which are employed to characterize the QoS of delay-tolerant and delay-sensitive services, respectively. Then, we proposed methods to find the optimal durations of the three phases and analyzed the impact of QoS requirement, SNR, and circuit power on the optimal durations for each kind of services. For delay-tolerant services, we showed that the EE-oriented design requires longer uplink training than the SE-oriented design in general, and they will coincide for high uplink SNR and large circuit power, or when the minimal SE requirement is equal to the maximum achievable SE. Moreover, we found that the BS will not operate in idle mode with the optimal strategy for supporting delay-tolerant services. For delay-sensitive services, we showed that the optimal duration for uplink training first increases and then decreases with the increase of QoS requirement in general. In the high-average-uplink-SNR regime, one training duration is optimal, and the optimal duration for downlink data sending increases with the QoS requirement linearly and decreases with SNR. Moreover, we found that BS idling plays an important role in reducing circuit power consumption when supporting delay-sensitive services. We also provided a simple way to apply the obtained transmission strategies for the scenarios with multiple users and mixed services. Simulation results validated our analytical analysis and demonstrated a significant EE gain of the EE-oriented design over the SE-oriented design in both single-user single-service and multiuser mixed-service scenarios.

Appendix A
Proof of Proposition 1

For the SE-oriented problem (14), it is easy to see that $\widetilde{\mathrm{SE}}(T_{\mathrm{tr}}, T_{\mathrm{d}}, T_{\mathrm{id}})$ is maximized when $T_{\mathrm{id}} = 0$, since $\widetilde{\mathrm{SE}}(T_{\mathrm{tr}}, T_{\mathrm{d}}, T_{\mathrm{id}})$ is an increasing function of T_{d} for any given T_{tr} according to (13).

For the EE-oriented problem (16), we can prove by contradiction that the maximum EE of (16) is obtained when $T_{\mathrm{tr}} + T_{\mathrm{d}} = T$ as follows.

First, we assume that the maximum EE is obtained when $T_{\mathrm{tr}} + T_{\mathrm{d}} < T$, which satisfies constraint (16c). Denote T_{tr1} and T_{d1} as the optimal uplink training duration and optimal downlink data sending duration, respectively. Then, the maximum EE can be denoted as $\widetilde{\mathrm{EE}}(T_{tr1}, T_{d1})$, and the corresponding SE, $\widetilde{\mathrm{SE}}(T_{tr1}, T_{d1}) \geq SE_{\min}$ satisfying constraint (16b).

We can find another downlink data sending duration T_{d2}, which is larger than T_{d1} but satisfies (16c). From (13) and (15), we can find that both $\widetilde{\mathrm{EE}}(T_{\mathrm{tr}}, T_{\mathrm{d}})$ and $\widetilde{\mathrm{SE}}(T_{\mathrm{tr}}, T_{\mathrm{d}})$ are the monotonically increasing functions of T_{d}. Therefore, there will be $\widetilde{\mathrm{SE}}(T_{tr1}, T_{d2}) > \widetilde{\mathrm{SE}}(T_{tr1}, T_{d1}) \geq SE_{\min}$, which means that constraint (16b) is satisfied, and $\widetilde{\mathrm{EE}}(T_{tr1}, T_{d2}) > \widetilde{\mathrm{EE}}(T_{tr1}, T_{d1})$. Obviously, it is contradictory with the assumption that $\widetilde{\mathrm{EE}}(T_{tr1}, T_{d1})$ is the maximum EE. It follows that the initial assumption that the maximum EE is obtained when $T_{\mathrm{tr}} + T_{\mathrm{d}} < T$ must be false. Further considering constraint (16c), the maximum EE is obtained when $T_{\mathrm{tr}} + T_{\mathrm{d}} = T$, proving the proposition.

Appendix B
Proof of Proposition 2

To simplify the notation, we rewrite (18) as

$$\widetilde{\mathrm{EE}}(T_{\mathrm{tr}}) = \frac{T - T_{\mathrm{tr}}}{P_{\mathrm{bu}}T_{\mathrm{tr}} + P_{\mathrm{bd}}(T - T_{\mathrm{tr}})} \log_2\left(1 + \frac{M\gamma_{\mathrm{d}}\gamma_u T_{\mathrm{tr}}}{1 + \gamma_u T_{\mathrm{tr}}}\right) \triangleq g_1(T_{\mathrm{tr}}) \cdot g_2(T_{\mathrm{tr}}) \tag{B.1}$$

where $g_1(T_{\mathrm{tr}}) = (T - T_{\mathrm{tr}})/(P_{\mathrm{bu}}T_{\mathrm{tr}} + P_{\mathrm{bd}}(T - T_{\mathrm{tr}})) \geq 0$ and $g_2(T_{\mathrm{tr}}) = \log_2(1 + M\gamma_{\mathrm{d}}\gamma_u T_{\mathrm{tr}}/(1 + \gamma_u T_{\mathrm{tr}})) \geq 0$.

Considering that $P_{\mathrm{bd}} > P_{\mathrm{bu}}$, it is easy to show that

$$g_1'(T_{\mathrm{tr}}) = -\frac{P_{\mathrm{bu}}T}{[P_{\mathrm{bu}}T_{\mathrm{tr}} + P_{\mathrm{bd}}(T - T_{\mathrm{tr}})]^2} < 0$$

$$g_1''(T_{\mathrm{tr}}) = \frac{-2P_{\mathrm{bu}}T(P_{\mathrm{bd}} - P_{\mathrm{bu}})}{[P_{\mathrm{bd}}T - (P_{\mathrm{bd}} - P_{\mathrm{bu}})T_{\mathrm{tr}}]^3} < 0$$

$$g_2'(T_{\mathrm{tr}}) = \frac{1}{\ln 2}\frac{M\gamma_{\mathrm{d}}\gamma_u}{(1 + \gamma_u T_{\mathrm{tr}} + M\gamma_{\mathrm{d}}\gamma_u T_{\mathrm{tr}})(1 + \gamma_u T_{\mathrm{tr}})} > 0$$

$$g_2''(T_{\mathrm{tr}}) = -\frac{M\gamma_{\mathrm{d}}\gamma_u}{\ln 2}\frac{2\left(\gamma_u + \gamma_u^2 T_{\mathrm{tr}} + M\gamma_{\mathrm{d}}\gamma_u^2 T_{\mathrm{tr}}\right) + M\gamma_{\mathrm{d}}\gamma_u}{[(1 + \gamma_u T_{\mathrm{tr}} + M\gamma_{\mathrm{d}}\gamma_u T_{\mathrm{tr}})(1 + \gamma_u T_{\mathrm{tr}})]^2} < 0.$$

Then, we can obtain the derivatives of $\widetilde{\mathrm{EE}}(T_{\mathrm{tr}})$ as

$$\begin{aligned}\widetilde{EE}'(T_{\mathrm{tr}}) &= g_1'(T_{\mathrm{tr}})g_2(T_{\mathrm{tr}}) + g_1(T_{\mathrm{tr}})g_2'(T_{\mathrm{tr}}),\\ \widetilde{EE}''(T_{\mathrm{tr}}) &= g_1''(T_{\mathrm{tr}})g_2(T_{\mathrm{tr}}) + 2g_1'(T_{\mathrm{tr}})g_2'(T_{\mathrm{tr}})\\ &\quad + g_1(T_{\mathrm{tr}})g_2''(T_{\mathrm{tr}}).\end{aligned} \tag{B.2}$$

We can see that $\widetilde{EE}''(T_{\mathrm{tr}}) < 0$ always holds. Therefore, $\widetilde{EE}(T_{\mathrm{tr}})$ is a concave function of T_{tr}, and $\widetilde{EE}'(T_{\mathrm{tr}})$ is a monotonically decreasing function of T_{tr}.

Similarly, we rewrite (17) as

$$\widetilde{\mathrm{SE}}(T_{\mathrm{tr}}) = \frac{T - T_{\mathrm{tr}}}{T}\log_2\left(1 + \frac{M\gamma_{\mathrm{d}}\gamma_u T_{\mathrm{tr}}}{1 + \gamma_u T_{\mathrm{tr}}}\right) = f_1(T_{\mathrm{tr}}) \cdot f_2(T_{\mathrm{tr}})$$

where $f_1(T_{\mathrm{tr}}) = (T - T_{\mathrm{tr}}/T) \geq 0$ and $f_2(T_{\mathrm{tr}}) = \log_2(1 + M\gamma_{\mathrm{d}}\gamma_u T_{\mathrm{tr}}/(1 + \gamma_u T_{\mathrm{tr}})) \geq 0$, whose derivatives are $f_1'(T_{\mathrm{tr}}) = -(1/T) < 0$, $f_1''(T_{\mathrm{tr}}) = 0$, $f_2'(T_{\mathrm{tr}}) = (1/\ln 2)M\gamma_{\mathrm{d}}\gamma_u/((1 +$

$\gamma_u T_{\rm tr}+M\gamma_{\rm d}\gamma_u T_{\rm tr})(1+\gamma_u T_{\rm tr}))>0$, and $f_2''(T_{\rm tr})=g_2''(T_{\rm tr})<0$. Then, we have

$$\widetilde{\rm SE}'(T_{\rm tr}) = f_1'(T_{\rm tr})f_2(T_{\rm tr}) + f_1(T_{\rm tr})f_2'(T_{\rm tr}),$$
$$\widetilde{\rm SE}''(T_{\rm tr}) = f_1''(T_{\rm tr})f_2(T_{\rm tr}) + 2f_1'(T_{\rm tr})f_2'(T_{\rm tr}) + f_1(T_{\rm tr})f_2''(T_{\rm tr}). \quad \text{(B.3)}$$

Since $\widetilde{\rm SE}''(T_{\rm tr})<0$ always holds, $\widetilde{\rm SE}(T_{\rm tr})$ is a concave function of $T_{\rm tr}$, and $\widetilde{\rm SE}'(T_{\rm tr})$ is monotonically decreasing.

APPENDIX C
PROOF OF PROPOSITION 3

We begin with providing the derivatives of EE and SE, i.e., $\widetilde{EE}'(T_{\rm tr})$ and $\widetilde{\rm SE}'(T_{\rm tr})$, at two extreme cases of the training duration with $T_{\rm tr}=\xi$ and $T_{\rm tr}=T$, where ξ is a very small positive value that approaches zero.

From (B.2) and (B.3) in Appendix B, we have

$$\widetilde{EE}'(\xi) \approx \frac{1}{\ln 2}\frac{M\gamma_{\rm d}\gamma_u}{P_{\rm bd}} > 0 \quad \text{(C.1)}$$

$$\widetilde{EE}'(T) = -\frac{1}{P_{\rm bu}T}\log_2\left(1+\frac{M\gamma_{\rm d}\gamma_u T}{1+\gamma_u T}\right) < 0 \quad \text{(C.2)}$$

$$\widetilde{\rm SE}'(\xi) \approx \frac{M\gamma_{\rm d}\gamma_u}{\ln 2} > 0 \quad \text{(C.3)}$$

$$\widetilde{\rm SE}'(T) = -\frac{1}{T}\log_2\left(1+\frac{M\gamma_{\rm d}\gamma_u T}{1+\gamma_u T}\right) < 0. \quad \text{(C.4)}$$

Considering $\widetilde{EE}'(\xi)>0$ and $\widetilde{EE}'T)<0$ from (C.1) and (C.2), as well as the fact that $\widetilde{EE}'(T_{\rm tr})$ is monotonically decreasing from Proposition 2, it follows that $\widetilde{EE}(T_{\rm tr})$ first increases and then decreases. Let $T_{\rm trEE}^o$ denote the training duration that maximizes $\widetilde{EE}$ without constraint (19b) on the minimum SE. Since $\widetilde{EE}(T_{\rm tr})$ is a concave function, we know that $\widetilde{EE}'(T_{\rm trEE}^o)=0$.

Since $\widetilde{\rm SE}'(T_{\rm tr})$ is monotonically decreasing, and $\widetilde{\rm SE}'(0)>0$, $\widetilde{\rm SE}'(T)<0$ from (C.3) and (C.4), $\widetilde{\rm SE}(T_{\rm tr})$ first increases and then decreases with $T_{\rm tr}$. Considering that $\widetilde{\rm SE}(T_{\rm tr})$ is concave, $T_{\rm trSE}^*$ can be found from $\widetilde{\rm SE}'(T_{\rm trSE}^*)=0$.

From (17) and (18), we can obtain the relationship between the approximations of the EE and the SE as

$$\widetilde{EE}(T_{\rm tr}) = \frac{\widetilde{\rm SE}(T_{\rm tr})T}{P_{\rm bu}\left(T+\left(\frac{P_{\rm bd}}{P_{\rm bu}}-1\right)(T-T_{\rm tr})\right)}. \quad \text{(C.5)}$$

By taking the derivative with respect to $T_{\rm tr}$, respectively, to the left- and right-hand sides of (C.5) and considering that $\widetilde{EE}'(T_{\rm trEE}^o)=0$, we can obtain

$$\widetilde{\rm SE}'(T_{\rm trEE}^o) = -\frac{\widetilde{\rm SE}(T_{\rm trEE}^o)}{\frac{T}{\frac{P_{\rm bd}}{P_{\rm bu}}-1}+T-T_{\rm trEE}^o}. \quad \text{(C.6)}$$

Obviously, $\widetilde{\rm SE}'(T_{\rm trEE}^o)<0$. Further considering that $\widetilde{\rm SE}'(T_{\rm trSE}^*)=0$ and $\widetilde{\rm SE}'(T_{\rm tr})$ is monotonically decreasing, we have $T_{\rm trEE}^o > T_{\rm trSE}^*$.

If $T_{\rm trEE}^o$ satisfies constraint (19b), we know that $T_{\rm trEE}^* = T_{\rm trEE}^o > T_{\rm trSE}^*$.

Now, we consider the case when $T_{\rm trEE}^o$ does not satisfy (19b). Denote $[T_{\rm tr}^{\min}, T_{\rm tr}^{\max}]$ as the feasible set of $T_{\rm tr}$ defined by the constraint in (19b). It is easy to see that $T_{\rm trSE}^* \in [T_{\rm tr}^{\min}, T_{\rm tr}^{\max}]$ since $T_{\rm trSE}^*$ maximizes the SE so that the constraint in (19b) is satisfied. Further considering the fact $T_{\rm trEE}^o > T_{\rm trSE}^*$, if $T_{\rm trEE}^o$ is outside of the feasible set, there must be $T_{\rm trEE}^o > T_{\rm tr}^{\max}$. As we have analyzed, $\widetilde{EE}(T_{\rm tr})$ first increases and then decreases with $T_{\rm tr}$, and the transition point is at $T_{\rm trEE}^o$. Therefore, we have $\widetilde{EE}(T_{\rm tr}^{\min}) \le \widetilde{EE}(T_{\rm trSE}^*) \le \widetilde{EE}(T_{\rm tr}^{\max}) < \widetilde{EE}(T_{\rm trEE}^o)$. Consequently, $T_{\rm trEE}^* = T_{\rm tr}^{\max} \ge T_{\rm trSE}^*$ in this case. The equality $T_{\rm trEE}^* = T_{\rm trSE}^*$ holds if $T_{\rm tr}^{\max} = T_{\rm trSE}^*$, i.e., if $SE_{\min}$ is equal to the maximum achievable net SE, i.e., $SE_{\max}$.

Based on the given results, the impact of $SE_{\min}$ on $\Delta T_{\rm tr}^*$ can be observed as follows. For small $SE_{\min}$ that satisfies $SE_{\min} \le \widetilde{\rm SE}(T_{\rm trEE}^o)$, we have $T_{\rm trEE}^* = T_{\rm trEE}^o > T_{\rm trSE}^*$, and thus, $\Delta T_{\rm tr}^*$ remains constant. With the increase of $SE_{\min}$ so that $SE_{\min} > \widetilde{\rm SE}(T_{\rm trEE}^o)$, we have $T_{\rm trEE}^* = T_{\rm tr}^{\max} \ge T_{\rm trSE}^*$. It is easy to see that $T_{\rm tr}^{\max}$ is a monotonically decreasing function of $SE_{\min}$, which means that $\Delta T_{\rm tr}^*$ decreases with the increase of $SE_{\min}$ in this case, and $\Delta T_{\rm tr}^*=0$ when $SE_{\min}=SE_{\max}$.

APPENDIX D
PROOF OF PROPOSITION 4

First, we assume that the minimum energy consumption is obtained when $2/\gamma_{\rm d}(1+1/(\gamma_u T_{\rm tr}))(2^{SE_r T/T_{\rm d}}-1)<\nu_c$, which satisfies constraint (29b). Denote T_{tr3} and T_{d3} as the optimal durations for uplink training and downlink data sending, respectively. Then, the minimum energy consumption can be denoted as $P_{\rm bu}T_{tr3}+P_{\rm bd}T_{d3}$, and $T_{tr3}+T_{d3}\le T$ satisfying the constraint in (29c).

From (29b), we can find that $2/\gamma_{\rm d}(1+1/(\gamma_u T_{\rm tr}))(2^{SE_r T/T_{\rm d}}-1)$ is a monotonically decreasing function of $T_{\rm d}$. Therefore, we can find another downlink data sending duration T_{d4}, which is smaller than T_{d3}, but still satisfies (29b) and (29c). Then there will be $P_{\rm bu}T_{tr3}+P_{\rm bd}T_{d4} < P_{\rm bu}T_{tr3}+P_{\rm bd}T_{d3}$, which is contradictory with the assumption that $P_{\rm bu}T_{tr3}+P_{\rm bd}T_{d3}$ is the minimum energy consumption. It follows that the initial assumption that the minimum energy consumption is obtained when $2/\gamma_{\rm d}(1+1/(\gamma_u T_{\rm tr}))(2^{SE_r T/T_{\rm d}}-1)<\nu_c$ must be false, i.e., constraint (29b) should hold with equality.

APPENDIX E
PROOF OF PROPOSITION 5

From (32), we can obtain the derivatives of $h(T_{\rm tr})$ as (E.1) and (E.2), shown at the bottom of the next page.

We can see that $h''(T_{\rm tr})>0$ always holds. Therefore, $h(T_{\rm tr})$ is a convex function of $T_{\rm tr}$.

Similarly, we can prove that $c''(T_{\rm tr})>0$ always holds by replacing $P_{\rm bu}$ and $P_{\rm bd}$ in (32) with 1; hence, $c(T_{\rm tr})$ in (33) is also a convex function of $T_{\rm tr}$.

APPENDIX F
PROOF OF PROPOSITION 6

Considering $h'(T_{\rm tr}^v)=0$ and $c'(T_{\rm tr}^c)=0$, we have from (E.1) that

$$j\left(T_{\rm tr}^v\right)=\left(1+b_1T_{\rm tr}^v+b_2T_{\rm tr}^v\right)\left(1+b_2T_{\rm tr}^v\right)\times\left[\log_2\left(1+\frac{b_1T_{\rm tr}^v}{1+b_2T_{\rm tr}^v}\right)\right]^2=\frac{b_1}{\ln 2}SE_rT\frac{P_{\rm bd}}{P_{\rm bu}} \quad \text{(F.1)}$$

$$j\left(T_{\rm tr}^c\right)=\left(1+b_1T_{\rm tr}^c+b_2T_{\rm tr}^c\right)\left(1+b_2T_{\rm tr}^c\right)\times\left[\log_2\left(1+\frac{b_1T_{\rm tr}^c}{1+b_2T_{\rm tr}^c}\right)\right]^2=\frac{b_1}{\ln 2}SE_rT \quad \text{(F.2)}$$

where $j(T_{\rm tr})=(1+b_1T_{\rm tr}+b_2T_{\rm tr})(1+b_2T_{\rm tr})[\log_2(1+b_1T_{\rm tr}/(1+b_2T_{\rm tr}))]^2$ is a monotonically increasing function of $T_{\rm tr}$.

Since $b_1>0$ and $P_{\rm bd}/P_{\rm bu}>1$ always hold, we have $j(T_{\rm tr}^v)\geq j(T_{\rm tr}^c)$, and the equality holds if and only if $SE_r=0$. Since $j(T_{\rm tr})$ is monotonically increasing, we have that $T_{\rm tr}^v\geq T_{\rm tr}^c$ always holds, and the equality holds if and only if $SE_r=0$.

In the following, we consider two cases.

1) $T_{\rm tr}^v$ Satisfies the Inequality Constraint in (31b): In this case, $T_{\rm tr}^v\leq T_{\rm tr}^{\rm up}$ and thus $T_{\rm trEEout}^*=T_{\rm tr}^v$. From (F.1), we have that $j(T_{\rm tr}^v)$ increases with the increase of SE requirement, i.e., SE_r. Since $j(T_{\rm tr}^v)$ is a monotonically increasing function of $T_{\rm tr}^v$, hence the optimal training duration $T_{\rm trEEout}^*=T_{\rm tr}^v$ increases with the increase of the SE requirement in this case.

2) $T_{\rm tr}^v$ Does Not Satisfy the Inequality Constraint in (31b): Noting that $T_{\rm tr}^v\geq T_{\rm tr}^c$, if $T_{\rm tr}^v$ is out of the feasible set $[T_{\rm tr}^{\rm low}, T_{\rm tr}^{\rm up}]$, there must be $T_{\rm tr}^v>T_{\rm tr}^{\rm up}$. As we have analyzed, $h(T_{\rm tr})$ first decreases and then increases with $T_{\rm tr}$, and the transition point is at $T_{\rm tr}^v$. Therefore, we have $h(T_{\rm tr}^{\rm low})\geq h(T_{\rm tr}^c)\geq h(T_{\rm tr}^{\rm up})>h(T_{\rm tr}^v)$. Consequently, $T_{\rm trEEout}^*=T_{\rm tr}^{\rm up}$ when $T_{\rm tr}^v>T_{\rm tr}^{\rm up}$.

As we have analyzed, $c(T_{\rm tr})$ first decreases and then increases with $T_{\rm tr}$, and the transition point is at $T_{\rm tr}^c$. Hence, the upper bound of the feasible set of training duration, i.e., $T_{\rm tr}^{\rm up}$, must belong to the increasing interval of $c(T_{\rm tr})$, i.e., $c(T_{\rm tr}^{\rm up})$ is an increasing function of $T_{\rm tr}^{\rm up}$. The value of $T_{\rm tr}^{\rm up}$ can be obtained when the constraint in (31b) holds with the equality $c(T_{\rm tr}^{\rm up})=T$, i.e.,

$$c\left(T_{\rm tr}^{\rm up}\right)=T_{\rm tr}^{\rm up}+\frac{SE_rT}{\log_2\left(1+\frac{\gamma_{\rm d}\nu_0}{2}\frac{\gamma_uT_{\rm tr}^{\rm up}}{1+\gamma_uT_{\rm tr}^{\rm up}}\right)}=T. \quad \text{(F.3)}$$

Obviously, $c(T_{\rm tr}^{\rm up})$ linearly increases with the increase of SE_r. Therefore, to ensure that $c(T_{\rm tr}^{\rm up})=T$ holds, the value of $T_{\rm tr}^{\rm up}$ must decrease when SE_r increases. It follows that $T_{\rm trEEout}^*=T_{\rm tr}^{\rm up}$ decreases with the increase of the SE requirement in this case.

The given two cases will be, respectively, active for different values of SE_r. For small SE_r, we have from (F.1) and (F.2) that $j(T_{\rm tr}^v)\approx j(T_{\rm tr}^c)$ and then $T_{\rm tr}^v\approx T_{\rm tr}^c$. Since $T_{\rm tr}^c\leq T_{\rm tr}^{\rm up}$ always holds, we have $T_{\rm tr}^v\leq T_{\rm tr}^{\rm up}$, which belongs to Case One. With the increase of SE_r, there will be a value of SE_r, under which both $T_{\rm tr}^{\rm up}$ and $T_{\rm tr}^{\rm low}$ will approach $T_{\rm tr}^c$, i.e., the feasible set is a point. We have shown that $T_{\rm tr}^v>T_{\rm tr}^c$ when $SE_r\neq 0$; therefore, we have $T_{\rm tr}^v>T_{\rm tr}^{\rm up}$ when SE_r is sufficiently large, which belongs to Case Two.

Considering that $T_{\rm tr}^v$ increases with SE_r while $T_{\rm tr}^{\rm up}$ decreases with SE_r, we can always find a SE_r (denoted by SE_r^o) with which $T_{\rm tr}^v=T_{\rm tr}^{\rm up}$ holds. When $SE_r\leq SE_r^o$, we have $T_{\rm tr}^v\leq T_{\rm tr}^{\rm up}$, and $T_{\rm trEEout}^*=T_{\rm tr}^v$, which increases with SE_r; when $SE_r>SE_r^o$, we have $T_{\rm tr}^v>T_{\rm tr}^{\rm up}$, and $T_{\rm trEEout}^*=T_{\rm tr}^{\rm up}$, which decreases with SE_r.

To analyze the impact of circuit power, we substitute (20) into (F.1) and have that

$$j\left(T_{\rm tr}^v\right)=\frac{b_1}{\ln 2}SE_rT\left(\lambda+\frac{\mu P_{\rm d}}{MP_c^{\rm RX}}\right). \quad \text{(F.4)}$$

When $P_c^{\rm RX}$ is large and $\lambda=1$, i.e., $P_c^{\rm TX}=P_c^{\rm RX}$, we have that $j(T_{\rm tr}^v)\approx j(T_{\rm tr}^c)$, and hence, $T_{\rm tr}^v\approx T_{\rm tr}^c$. Since $T_{\rm tr}^c\leq T_{\rm tr}^{\rm up}$ always holds, we have $T_{\rm tr}^v\leq T_{\rm tr}^{\rm up}$ in this scenario, which belongs to Case One, regardless of the value of SE_r. Therefore, we have $T_{\rm trEEout}^*=T_{\rm tr}^v$, which increases with SE_r. Moreover, it can be observed from (F.4) that $j(T_{\rm tr}^v)$ decreases with the increase of circuit power, i.e., $P_c^{\rm RX}$. Since $j(T_{\rm tr}^v)$ is a monotonically increasing function of $T_{\rm tr}^v$, we have $T_{\rm trEEout}^*=T_{\rm tr}^v$, which decreases with circuit power.

REFERENCES

[1] G. Auer *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[2] J. Kwak, K. Son, Y. Yi, and S. Chong, "Greening effect of spatio-temporal power sharing policies in cellular networks with energy constraints," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4405–4415, Dec. 2012.

$$h'(T_{\rm tr})=a_1-\frac{a_2b_1}{\ln 2}\frac{1}{(1+b_1T_{\rm tr}+b_2T_{\rm tr})(1+b_2T_{\rm tr})\left[\log_2\left(1+\frac{b_1T_{\rm tr}}{1+b_2T_{\rm tr}}\right)\right]^2} \quad \text{(E.1)}$$

$$h''(T_{\rm tr})=\frac{a_2b_1}{\ln 2}\frac{[b_1+2b_2(1+b_1T_{\rm tr}+b_2T_{\rm tr})]\log_2\left(1+\frac{b_1T_{\rm tr}}{1+b_2T_{\rm tr}}\right)+\frac{2b_1}{\ln 2}}{[(1+b_1T_{\rm tr}+b_2T_{\rm tr})(1+b_2T_{\rm tr})]^2\left[\log_2\left(1+\frac{b_1T_{\rm tr}}{1+b_2T_{\rm tr}}\right)\right]^3} \quad \text{(E.2)}$$

[3] J. Wu, S. Zhou, and Z. Niu, “Traffic-aware base station sleeping control and power matching for energy-delay tradeoffs in green cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4196–4209, Aug. 2013.

[4] M. I. Poulakis, A. D. Panagopoulos, and P. Constantinou, “Channel-aware opportunistic transmission scheduling for energy-efficient wireless links,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 192–204, Jan. 2013.

[5] C. Sun and C. Yang, “Energy-efficient hybrid one- and two-way relay transmission,” *IEEE Trans. Veh. Technol.*, vol. 62, no. 8, pp. 3737–3751, Oct. 2013.

[6] J. Xu, L. Qiu, and C. Yu, “Improving network energy efficiency through cooperative idling in the multi-cell systems,” *EURASIP J. Wireless Commun. Netw.*, vol. 2011, no. 165, pp. 1–18, 2011.

[7] S. Han, C. Yang, and A. Molisch, “Spectrum and energy efficient cooperative base station doze,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 285–296, Feb. 2014.

[8] J. Xu and L. Qiu, “Energy efficiency optimization for MIMO broadcast channels,” *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, pp. 690–701, Feb. 2013.

[9] B. Hassibi and B. Hochwald, “How much training is needed in multiple-antenna wireless links?” *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[10] J. Hoydis, M. Kobayashi, and M. Debbah, “Optimal channel training in uplink network MIMO systems,” *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824–2833, Jun. 2011.

[11] M. Kobayashi, N. Jindal, and G. Caire, “Training and feedback optimization for multiuser MIMO downlink,” *IEEE Trans. Commun.*, vol. 59, no. 8, pp. 2228–2240, Aug. 2011.

[12] M. Gursoy, “On the capacity and energy efficiency of training-based transmissions over fading channels,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4543–4567, Oct. 2009.

[13] Z. Xu *et al.*, “Energy-efficient power allocation between pilots and data symbols in downlink OFDMA systems,” in *Proc. IEEE GLOBECOM*, Dec. 2011, pp. 1–6.

[14] X. Liu, S. Han, C. Yang, and C. Sun, “Energy-efficient uplink training design for closed-loop MISO systems,” in *Proc. IEEE WCNC*, Apr. 2013, pp. 2548–2553.

[15] T. Bohn *et al.*, “D4.1: Most promising tracks of green radio technologies,” INFSO-ICT-247733, Tech. Rep., Dec. 2010. [Online]. Available: https://www.ict-earth.eu/

[16] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, Time Series Analysis*. Reading, MA, USA: Addison-Wesley, 1991.

[17] T. Lo, “Maximum ratio transmission,” *IEEE Trans. Commun.*, vol. 47, no. 10, pp. 1458–1461, Oct. 1999.

[18] R. Zhang, “Optimal dynamic resource allocation for multi-antenna broadcasting with heterogeneous delay-constrained traffic,” *IEEE J. Sel. Areas Signal Process.*, vol. 2, no. 2, pp. 243–255, Apr. 2008.

[19] J. Lee and N. Jindal, “Energy-efficient scheduling of delay constrained traffic over fading channels,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1866–1875, Apr. 2009.

[20] R. Heath, M. Kountouris, and T. Bai, “Modeling heterogeneous network interference with using Poisson point processes,” *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4114–4126, Aug. 2013.

[21] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, Dec. 2004.

[22] V. S. Annapureddy, D. V. Marathe, T. R. Ramya, and S. Bhashyam, “Outage probability of multiple-input single-output (MISO) systems with delayed feedback,” *IEEE Trans. Commun.*, vol. 57, no. 2, pp. 319–326, Feb. 2009.

[23] I. Hwang, B. Song, and S. S. Soliman, “A holistic view on hyper-dense heterogeneous and small cell networks,” *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, Jun. 2013.

[24] S. Hong *et al.*, “Applications of self-interference cancellation in 5G and beyond,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, Feb. 2014.

[25] A. Ambrosy, O. Blume, H. Klessig, and W. Wajda, “Energy saving potential of integrated hardware and resource management solutions for wireless base stations,” in *Proc. IEEE PIMRC*, 2011, pp. 2418–2423.

[26] “Further advancements for E-UTRA physical layer aspects (Release 9),” Sophia-Antipolis Cedex, France, Tech. Rep. TR 36.814, Mar. 2010.

Xin Liu received the B.S.E. degree in electronics and information engineering from Jilin University, Changchun, China, in 2011 and the M.S.E. degree in signal and information processing from Beihang University, Beijing, China, in 2014.

His research interests include energy-efficient training and transmission design in the areas of wireless communications. After graduation, he joined Microsoft China as an Account Technology Strategist.

Shengqian Han (S'05–M'12) received the B.S. degree in communication engineering and the Ph.D. degree in signal and information processing from Beihang University (formerly Beijing University of Aeronautics and Astronautics), Beijing, China, in 2004 and 2010, respectively.

From 2010 to 2012, he held a postdoctoral research position with the School of Electronics and Information Engineering, Beihang University, where he is currently a Lecturer. His research interests include multiple-antenna techniques, cooperative communication, and energy-efficient transmission in the areas of wireless communications and signal processing.

Chenyang Yang (SM'08) received the M.S.E. and Ph.D. degrees in electrical engineering from Beihang University (formerly Beijing University of Aeronautics and Astronautics), Beijing, China, in 1989 and 1997, respectively.

She is currently a Full Professor with the School of Electronics and Information Engineering, Beihang University. She has published various papers and filed many patents in the fields of signal processing and wireless communications. Her recent research interests include signal processing in network multiple-input multiple-output, cooperative communication, energy-efficient transmission, and interference management.

Dr. Yang was the Chair of the IEEE Communications Society Beijing Chapter from 2008 to 2012. She has served as a Technical Program Committee Member for many IEEE conferences, such as the IEEE International Conference on Communications and the IEEE Global Telecommunications Conference. She currently serves as an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, an Associate Editor-in-Chief for the *Chinese Journal of Communications*, and an Associate Editor-in-Chief for the *Chinese Journal of Signal Processing*. She was nominated as an Outstanding Young Professor of Beijing in 1995 and was supported by the First Teaching and Research Award Program for Outstanding Young Teachers of Higher Education Institutions by the Ministry of Education (P.R.C. “TRAPOYT”) during 1999–2004.