Relaying with Deadline Constraint: Energy Minimization with Full Channel State Information

Chin Keong Ho, Peng Hui Tan, and Sumei Sun
Institute for Infocomm Research, A*STAR, Singapore 138632.
e-mail: {hock,phtan,sunsm}@i2r.a-star.edu.sg

***Abstract*—We consider a time-slotted source-relay-destination network where data is to be delivered by a deadline. Our goal is to minimize the sum transmission energy by power allocation for each node and time slot, with knowledge of *full channel state information* (CSI) in the form of the SNRs of all slots before the deadline. We assume that both receivers, namely the relay and the destination, can accumulate mutual information across previous and present slots by using earlier received packets for joint decoding. We obtain the structure of the optimal solution in a semi-analytical closed-form. The structural result shows that the source and relay jointly perform a generalized form of water-filling over slots. Finally, we use the structural result to develop a heuristic scheme that uses only *causal CSI*, in the form of the SNRs of only past and present slots.**

I. Introduction

We consider a time-slotted communication system subject to a deadline constraint, i.e., a fixed number of bits is to be delivered from a source S to a destination D within a fixed number of slots. This scenario is relevant for applications such as multimedia streaming, where information is considered useful only if it is delivered within the deadline. We assume a relay R helps via a two-phase protocol: the relay tries to decode the source's message in Phase 1, then participates in relaying in Phase 2. Each phase consists of one or more slots.

Our objective is to minimize the sum transmission energy used by all nodes over all slots, by power allocation for each node and slot, subject to the deadline constraint. Implicitly via the power allocation, we also determine the optimal slot where Phase 1 transits to Phase 2. We assume availability of *full* channel state information (CSI), in terms of the link SNRs of all slots. Our contributions are:

(i) For static channels given full CSI, we obtain a closed-form solution. In contrast to a continuous-time system, in (discrete-time) slotted systems the optimal time instant that the relay decodes the message may not always coincide with the time that the relay starts transmitting (at slot boundaries).

(ii) For general fading channels given full CSI, we obtain a semi-analytical closed-form solution. The structural solution shows that the source and relay jointly perform a generalized form of water-filling over slots, and allows an efficient numerical computation of the optimal solution.

(iii) Based on the observations from (ii), we propose a heuristic solution for the intractable (but practical) problem assuming that only *causal* CSI is available, i.e., the SNRs in future slots are not a priori known. Numerical results show the significant power saving is possible.

	slot 1	slot 2	slot 3
source S	$\boxed{\mathbf{x}_1(w)}$	$\boxed{\mathbf{x}_2(w)}$	$\boxed{\mathbf{x}_3(w)}$
Power at S	p_1^{S}	p_2^{S}	p_3^{S}
relay R		$\boxed{\widetilde{\mathbf{x}}_2(\widehat{w})}$	$\boxed{\widetilde{\mathbf{x}}_3(\widehat{w})}$
Power at R	$p_1^{\mathsf{R}} = 0$	p_2^{R}	p_3^{R}

Fig. 1. Protocol for $K = 3$, assuming R decodes w as $\widehat{w}$ in slot 1.

Literature Survey: This energy minimization problem given causal CSI, but without relaying, was studied in [1]–[3]. In [4], different deadlines were considered for separately arrived packets. In [5], a continuous-time framework was considered. In [6]–[11], relaying is considered. Different from these works, we consider a slotted system with deadline and we employ (i) *inter-slot relaying* such that the transition from Phase 1 to 2 occurs only at slot boundaries, and (ii) *mutual-information accumulation* (MIA) such that both S and R use all received packets for joint decoding. The MIA technique is similar to the use of an ARQ scheme that employs incremental redundancy via retransmissions [12], except that in our context, retransmissions are deliberate and both S and R perform decoding. The motivation of using inter-slot relaying is that it is readily implemented, while that of using MIA is that it increases the achievable rate and conversely saves transmission power for a given transmission rate.

II. System Model

A source S delivers message W of nB nats over $K \in \mathbb{Z}^+$ slots to a destination D with the help of a relay R. Each slot consists of n channel uses over time. Thus, the effective transmission rate is $R = B/K$ nats per channel use. Each node has one antenna and is subject to the half-duplex constraint.

A. Coding Scheme

The source uses independent Gaussian codebooks for each slot k. For every message $W \in \{1, \cdots, e^{nB}\}$, a codeword $\mathbf{x}_k^{\mathsf{S}}(W)$ of length n is generated according to $\mathcal{CN}(\mathbf{0}_n, \mathbf{I}_n)$, the i.i.d. complex-valued zero-mean unit-variance n-variate Gaussian distribution. The relay independently generates its codebook, with codewords $\{\mathbf{x}_k^{\mathsf{R}}(W)\}$, based on the same steps.

B. Relaying Protocol

Since the nodes are subject to the half-duplex constraint, we divide the relaying protocol into two phases. In Phase 1, R has

978-1-4673-0990-5/12/$31.00 ©2012 IEEE

not yet decoded the message. In Phase 2, R has decoded the message and so can help to relay. We employ a decode-and-forward relaying scheme where each receiver can accumulate mutual information over slots. An example of the power allocation for $K = 3$ slots is shown in Fig. 1.

For any Y-to-X link in slot k, let the scalar $\gamma_k^{\mathsf{YX}} \geq 0$ be the channel SNR, while the n channel phases are denoted by $\boldsymbol{\theta}_k^{\mathsf{YX}}$. Let $\boldsymbol{\Theta}_k^{\mathsf{YX}}$ be the diagonal matrix with $\boldsymbol{\theta}_k^{\mathsf{YX}}$ as its diagonal elements. We assume that the transmitter Y and the receiver X know γ_k^{YX}, but only the receiver knows $\boldsymbol{\theta}_k^{\mathsf{YX}}$.

1) Phase 1: Consider Phase 1, before the relay has decoded the message W. The source transmits the codeword with power p_k^{S}. The codeword is received at node $\mathsf{X} \in \{\mathsf{R}, \mathsf{D}\}$, which denotes the relay or the destination respectively, as

$$\mathbf{y}_k^{\mathsf{X}} = \sqrt{p_k^{\mathsf{S}} \gamma_k^{\mathsf{SX}}} \cdot \boldsymbol{\Theta}^{\mathsf{SX}} \cdot \mathbf{x}_k^{\mathsf{S}} + \mathbf{v}_k^{\mathsf{X}} \tag{1}$$

where $\mathbf{v}_k^{\mathsf{X}} \sim \mathcal{CN}(\mathbf{0}_n, \mathbf{I}_n)$ is additive white Gaussian noise (AWGN). Using all past received packets $\{\mathbf{y}_i^{\mathsf{R}}, i = 1, \cdots, k\}$ for joint typical decoding, the relay reliably decodes W if [13]

$$\sum_{i=1}^{k} I(p_i^{\mathsf{S}} \gamma_i^{\mathsf{SR}}) \geq B \tag{2}$$

where $I(x) := \log(1 + x)$ is the mutual information for Gaussian channels with SNR $x \geq 0$ and log is the natural logarithm. Thus, the mutual information accumulates over slots until it exceeds B, upon which the relay decodes W.

2) Phase 2: We denote $\widetilde{K}$ as the earliest possible slot index such that (2) holds. We assume $1 \leq \widetilde{K} \leq K - 1$; otherwise, relaying is not necessary and Phase 2 is not needed. Consider Phase 2 that runs from slot $k = \widetilde{K} + 1$ to $k = K$, in which R can perform relaying. Both S and R transmit concurrently using their independent codebooks. The destination thus receives

$$\mathbf{y}_k^{\mathsf{D}} = \sqrt{\gamma_k^{\mathsf{SD}} p_k^{\mathsf{S}}} \cdot \boldsymbol{\Theta}^{\mathsf{SD}} \cdot \mathbf{x}_k^{\mathsf{S}} + \sqrt{\gamma_k^{\mathsf{RD}} p_k^{\mathsf{R}}} \cdot \boldsymbol{\Theta}^{\mathsf{RD}} \cdot \mathbf{x}_k^{\mathsf{R}} + \mathbf{v}_k^{\mathsf{D}} \tag{3}$$

for $\widetilde{K} + 1 \leq k \leq K$. We assume a phase fading channel, i.e., each element in $\boldsymbol{\theta}^{\mathsf{SD}}$ and $\boldsymbol{\theta}^{\mathsf{RD}}$ is i.i.d. with a uniform distribution over $[0, 2\pi)$ [14]. Using *all* received packets $\{\mathbf{y}_k^{\mathsf{D}}, k = 1, \cdots, K\}$ for joint typical decoding, the destination reliably decodes message W if [14]

$$\sum_{k=1}^{\widetilde{K}} I(p_k^{\mathsf{S}} \gamma_k^{\mathsf{SD}}) + \sum_{k=\widetilde{K}+1}^{K} I(p_k^{\mathsf{S}} \gamma_k^{\mathsf{SD}} + p_k^{\mathsf{R}} \gamma_k^{\mathsf{RD}}) \geq B. \tag{4}$$

III. Optimal Solution with Full CSI

We formulate the problem with full CSI in Section III-A. Then we obtain the solution for the static channel in Section III-B and for the general fading channels in Section III-C.

A. Optimization Problem

We wish to minimize the total transmission power (or interchangeably energy) by varying the transmission power $p_k^{\mathsf{S}}, p_k^{\mathsf{R}} \geq 0$ for all k (and hence also $\widetilde{K}$ implicitly), such that the destination can reliably decode nB nats after K slots of transmission.

We assume *full CSI* is available, i.e., the SNRs of all links of all slots are known before transmission begins. Thus the minimum total transmission power serves as the lower bound when causal CSI is available, wherein the problem often becomes intractable, see e.g. [1]–[3], [11]. Moreover, the results developed for full CSI may be translated to multi-carrier systems where resources are allocated in the frequency domain.

We can separate the analysis into two cases. In the *no-relaying* case, S transmits directly to D. In the *active-relay* case, R transmits with positive power for some slots in Phase 2. By obtaining the minimum sum power for both cases separately, we can then implement the case that incurs a smaller minimum sum power. The optimal power allocation for the no-relaying case is straightforwardly obtained by the water-filling algorithm, see e.g. [15]. The optimal power allocation the active-relay case is, however, challenging.

Henceforth, we focus on the active-relay case. Recall that $1 \leq \widetilde{K} \leq K-1$ is the earliest possible slot index such that (2) holds, after which the relay starts to perform relaying. Taking $\widetilde{K}$ as an auxillary variable to be optimized, the sum-energy optimization problem is formulated as Problem P0, given by

$$J^{\star} \triangleq \min_{1 \leq \widetilde{K} \leq K-1} \min_{\{p_k^{\mathsf{S}} \geq 0, p_k^{\mathsf{R}} \geq 0\}} \sum_{k=1}^{K} p_k^{\mathsf{S}} + p_k^{\mathsf{R}} \text{ s.t. } (2), (4) \text{ hold.} \tag{5}$$

Without loss of optimality, only the *stronger* node, i.e., the node with the larger SNR to the destination, transmits in Phase 2; otherwise it can be shown that the sum power can be reduced without violating the constraints (2),(4). Henceforth, we assume only the stronger node transmits in Phase 2 with power $p_k \geq 0, \widetilde{K} + 1 \leq k \leq K$. Then the equivalent channel in Phase 2 is given by $\widetilde{\gamma}_k \triangleq \max(\gamma_k^{\mathsf{SD}}, \gamma_k^{\mathsf{RD}})$. For notational consistency, we re-write p_k^{S} in Phase I as $p_k, 1 \leq k \leq \widetilde{K}$.

Given $\widetilde{K}$, the *inner minimization problem* in Problem P0 can then be stated equivalently as Problem P1:

$$J(\widetilde{K}) \triangleq \min_{\mathbf{p} \geq \mathbf{0}} \sum_{k=1}^{K} p_k \tag{6a}$$

$$\text{s.t.} \sum_{k=1}^{\widetilde{K}} I\left(\gamma_k^{\mathsf{SR}} p_k\right) \geq B \tag{6b}$$

$$\sum_{k=1}^{\widetilde{K}} I\left(\gamma_k^{\mathsf{SD}} p_k\right) + \sum_{k=\widetilde{K}+1}^{K} I\left(\widetilde{\gamma}_k p_k\right) \geq B \tag{6c}$$

where $\mathbf{p}$ denotes a length-K vector with elements $\{p_k\}$, and $\mathbf{0}$ is an all-zero vector. After solving for P1 for all $\widetilde{K}$, we can then obtain the optimal solution for P0 as

$$J^{\star} = \min_{1 \leq \widetilde{K} \leq K-1} J(\widetilde{K}) \tag{7}$$

which only involves a comparison of $K - 1$ terms. Subsequently we take $\widetilde{K}$ as fixed and focus on Problem P1.

Problem P1 is a convex optimization problem, as the objective function is linear in $\mathbf{p}$ and the feasible set is convex. Hence, the optimal solution can be solved numerically via convex optimization techniques [15]. However, it is insightful to obtain the optimal solution, or at least some of its structural properties, in closed form, so that we can obtain heuristic but close-to-to-optimal solutions that can be efficiently implemented.

B. Static Channel

We first consider the *static channel* where $\gamma_k^{\mathsf{SD}} = \gamma^{\mathsf{SD}}, \gamma_k^{\mathsf{SR}} = \gamma^{\mathsf{SR}}, \gamma_k^{\mathsf{RD}} = \gamma^{\mathsf{RD}}$ for all k, i.e., the SNR for each link is the same for all slots. It turns out that $J(\widetilde{K})$, denoted specifically as $J^{\text{static}}(\widetilde{K})$ for the static case, can be obtained in closed-form.

Without loss of generality, we introduce an auxiliary variable[1] $0 \le \widetilde{B} \le B$ such that $\widetilde{B} = B - \sum_{k=\widetilde{K}+1}^{K} I(\widetilde{\gamma} p_k)$ where $\widetilde{\gamma} = \widetilde{\gamma}_k = \max(\gamma_k^{\mathsf{SD}}, \gamma_k^{\mathsf{RD}})$ which is independent of k for the static channel. We can interpret $\widetilde{B}$ as the mutual information needed by the destination in Phase 1 for decoding the message. We shall use the observation that when $\widetilde{B}$ is fixed, P1 is decoupled into two simpler optimization subproblems. Specifically, by varying over $\widetilde{B}$, it can be easily verified that the minimum transmission energy can be expressed as

$$J^{\text{static}}(\widetilde{K}) = \min_{0 \le \widetilde{B} \le B} J_1^{\text{static}}(\widetilde{B}, \widetilde{K}) + J_2^{\text{static}}(\widetilde{B}, \widetilde{K}) \quad (8)$$

where $J_1^{\text{static}}, J_2^{\text{static}}$ are the respective solutions to these problems for fixed $\widetilde{B}$ and $\widetilde{K}$:

$$J_1^{\text{static}} \triangleq \min_{\mathbf{p}_1 \ge \mathbf{0}} \sum_{k=1}^{\widetilde{K}} p_k \quad (9a)$$

$$\text{s.t.} \ \sum_{k=1}^{\widetilde{K}} I\left(\gamma^{\mathsf{SR}} p_k\right) \ge B, \ \sum_{k=1}^{\widetilde{K}} I\left(\gamma^{\mathsf{SD}} p_k\right) \ge \widetilde{B} \quad (9b)$$

$$J_2^{\text{static}} \triangleq \min_{\mathbf{p}_2 \ge \mathbf{0}} \sum_{k=\widetilde{K}+1}^{K} p_k \ \text{ s.t.} \ \sum_{k=\widetilde{K}+1}^{K} I\left(\widetilde{\gamma} p_k\right) = B - \widetilde{B}. \quad (10)$$

where $\mathbf{p}_1, \mathbf{p}_2$ denote vectors with the first $\widetilde{K}$ elements and the last $K - \widetilde{K}$ elements of $\mathbf{p}$, respectively.

The optimal solution to (8) is expressed in closed-form in Theorem 1. First, we need Lemma 1. We assume that $\gamma^{\mathsf{SD}} < \gamma^{\mathsf{SR}}$, otherwise no relaying is needed as the relay cannot accumulate more mutual information than the destination. All proofs are given in the Appendix, unless otherwise specified.

Lemma 1: Let $\gamma^{\mathsf{SD}} < \gamma^{\mathsf{SR}}$. The optimal power allocation for (9) and (10) are independent of k, given respectively by

$$q_1^\star(\widetilde{B}) = \max\{q_1'(\widetilde{B}), q_1''\}, \quad (11)$$

$$q_2^\star(\widetilde{B}) = \left(\exp\left((B - \widetilde{B})/(K - \widetilde{K})\right) - 1\right)/\widetilde{\gamma} \quad (12)$$

where $q_1'(\widetilde{B}) \triangleq (e^{\widetilde{B}/\widetilde{K}} - 1)/\gamma^{\mathsf{SD}}, q_1'' \triangleq (e^{B/\widetilde{K}} - 1)/\gamma^{\mathsf{SR}}$. Thus, $J_1^{\text{static}}(\widetilde{B}) = \widetilde{K} q_1^\star(\widetilde{B}), J_2^{\text{static}}(\widetilde{B}) = (K - \widetilde{K}) q_2^\star(\widetilde{B})$.

Remark 1: From Lemma 1, the optimal power allocation $q_1^\star, q_2^\star$ to subproblems (9) and (10), respectively, are independent of the slot index. This implies that the optimal power allocation is quasi-constant which may change at most once at the transition from Phase 1 to Phase 2. Moreover, as seen from Appendix A, this optimal solution is unique[2].

Theorem 1: Let $f(x) := \widetilde{K} p_1'(x) + (K - \widetilde{K}) q_2^\star(x), 0 \le x \le B$. Given $1 \le \widetilde{K} \le K - 1$, the optimal solution in (8) is

$$J^{\text{static}}(\widetilde{K}) = f(\max\{\widetilde{B}_1^\star, \widetilde{B}_2^\star\}) \quad (13)$$

where $\widetilde{B}_1^\star$ solves $q_1'(\widetilde{B}_1^\star) = q_1''$ and $\widetilde{B}_2^\star$ is the unique stationary point of $f(\widetilde{B})$, i.e., $\widetilde{B}_1^\star = \widetilde{K} \log\left(1 - \frac{\gamma^{\mathsf{SD}}}{\gamma^{\mathsf{SR}}} + \frac{\gamma^{\mathsf{SD}}}{\gamma^{\mathsf{SR}}} e^{B/\widetilde{K}}\right)$, $\widetilde{B}_2^\star = \frac{\widetilde{K}}{K}\widetilde{B} - \frac{\widetilde{K}(K-\widetilde{K})}{K} \log\left(\frac{\widetilde{\gamma}}{\gamma^{\mathsf{SD}}}\right)$.

Remark 2: In the information-theoretic literature [13], [14], the continuous-time relaying system was considered. The optimal durations for Phase 1 and Phase 2 are determined by a time-sharing variable α; e.g., see [11] where $0 < \alpha < 1$ is optimized. A key feature of the continuous-time system is that the optimal time instant $t_{\text{decode}}^\star$ when the relay correctly decodes the message, equals the optimal time instant $t_{\text{relay}}^\star$ when the relay starts relaying. In the discrete-time system, however, it can be shown that $t_{\text{decode}}^\star < t_{\text{relay}}^\star$ hold in some situations. Thus, the intuition for the continuous-time system does not immediately carry over to the slotted system.

C. Fading Channel

Next, we consider the general case where the SNRs can change over slots. We call this the fading channel since the time variation is typically due to fading. The optimal solution for Problem P1 can be expressed in a semi-analytical closed-form in Theorem 2. First, we need Lemma 2 and Lemma 3.

In subsequent derivations, we use the fact that Problem P1 is a convex optimization problem. Since Slater's condition is satisfied and so strong duality holds, we can apply the Karush-Kuhn-Tucker (KKT) optimality conditions [15]. The Lagrangian function of the dual problem is $\mathcal{L}(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \triangleq \sum_{k=1}^{K} p_k - \sum_{k=1}^{K} \mu_k p_k - \lambda_1 \left(\sum_{k=1}^{\widetilde{K}} I\left(\gamma_k^{\mathsf{SR}} p_k\right) - B\right) - \lambda_2 \left(\sum_{k=1}^{\widetilde{K}} I\left(\gamma_k^{\mathsf{SD}} p_k\right) + \sum_{k=\widetilde{K}+1}^{K} I\left(\widetilde{\gamma}_k p_k\right) - B\right)$, where $\boldsymbol{\lambda} \triangleq [\lambda_1, \lambda_2], \boldsymbol{\mu} \triangleq [\mu_1, \cdots, \mu_K]$ are non-negative Lagrangian multipliers. We denote the optimal power allocation and Lagrange multipliers that satisfy the KKT conditions with a superscript $\star$.

Lemma 2: Assume $1 \le \widetilde{K} \le K - 1$ and the relay is active, i.e., R transmits with positive power for some slot in Phase 2. Given the optimal $\boldsymbol{\lambda} \ge \mathbf{0}$, the optimal power allocation $\mathbf{p}^\star$ satisfies

$$p_k^\star(\boldsymbol{\lambda}) = \begin{cases} [p_k^\circ(\lambda_1, \lambda_2)]^+, & 1 \le k \le \widetilde{K} \text{ (Phase 1)} \quad (14a) \\ [\lambda_2 - 1/\widetilde{\gamma}_k]^+, & \widetilde{K} + 1 \le k \le K \text{ (Phase 2)} \quad (14b) \end{cases}$$

such that the primal constraints (6b) and (6c), and the complementary slackness condition (CSC) that

$$\lambda_1 \left(\sum_{k=1}^{\widetilde{K}} I\left(\gamma_k^{\mathsf{SR}} p_k^\star\right) - B\right) = 0 \quad (15a)$$

[1]Clearly $\widetilde{B} \le B$ as $I(\cdot) \ge 0$. It is not necessary to consider $\widetilde{B} < 0$, otherwise p_k can be further reduced in Phase 2 and yet satisfy the constraints.

[2]From the proof, it can be deduced that the quasi-constant and unique properties hold for any function $I(\cdot)$ that is strictly concave.

$$\lambda_2\left(\sum_{k=1}^{\widetilde{K}} I\left(\gamma_k^{\mathsf{SD}} p_k^\star\right)+\sum_{k=\widetilde{K}+1}^{K} I\left(\widetilde{\gamma}_k p_k^\star\right)-B\right)=0 \tag{15b}$$

are satisfied. In (14a), $p_k^\circ(\lambda_1,\lambda_2)$ is defined as the non-negative solution of the equation

$$1=\frac{\lambda_1\gamma_k^{\mathsf{SR}}}{1+\gamma_k^{\mathsf{SR}}p_k^\circ}+\frac{\lambda_2\gamma_k^{\mathsf{SD}}}{1+\gamma_k^{\mathsf{SD}}p_k^\circ}, \tag{16}$$

i.e., $p_k^\circ=\phi_{1k}+\sqrt{\phi_{1k}^2+\phi_{2k}}$ where

$$\phi_{1k}\triangleq\frac{1}{2}\left(\lambda_1+\lambda_2-\frac{1}{\gamma_k^{\mathsf{SR}}}-\frac{1}{\gamma_k^{\mathsf{SD}}}\right) \tag{17}$$

$$\phi_{2k}\triangleq\frac{\lambda_1}{\gamma_k^{\mathsf{SD}}}+\frac{\lambda_2}{\gamma_k^{\mathsf{SR}}}-\frac{1}{\gamma_k^{\mathsf{SR}}\gamma_k^{\mathsf{SD}}}. \tag{18}$$

The necessary and sufficient condition for $p_k^\circ\geq 0$ is

$$\lambda_1\gamma_k^{\mathsf{SR}}+\lambda_2\gamma_k^{\mathsf{SD}}\geq 1. \tag{19}$$

Remark 3: To solve Problem P1, we need to only search $\boldsymbol{\lambda}$ over a smaller optimization domain, specifically given by $\boldsymbol{\Lambda}\triangleq\{\boldsymbol{\lambda}\geq\mathbf{0}: p_k^\star(\boldsymbol{\lambda})$ satisfies (6b), (6c), $k\in\mathcal{K}$, and (15)$\}$, as

$$J(\widetilde{K})\overset{(a)}{=}\max_{\boldsymbol{\lambda}\geq\mathbf{0},\boldsymbol{\mu}\geq\mathbf{0}}\min_{\mathbf{p}\geq\mathbf{0}}\mathcal{L}(\mathbf{p},\boldsymbol{\lambda},\boldsymbol{\mu})\overset{(b)}{=}\max_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\sum_{k=1}^{K}p_k^\star(\lambda_1,\lambda_2) \tag{20}$$

where (a) follows from strong duality and (b) follows from Lemma 2. To simplify further, we need Lemma 3.

Lemma 3: Without loss of optimality, the following properties hold for the optimization in (20).

1) The constraint (6c) in the set $\boldsymbol{\Lambda}$ holds with equality. Hence (15b) also holds. Thus, (20) becomes

$$\max_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\sum_{k=1}^{K}p_k^\star(\lambda_1,\lambda_2)=\max_{\lambda_2\in\Lambda_2}\sum_{k=1}^{K}p_k^\star(\widetilde{\lambda}_1(\lambda_2),\lambda_2) \tag{21}$$

where $\widetilde{\lambda}_1(\lambda_2)$ is the solution of λ_1 such that (6c) holds with equality given λ_2 and $\Lambda_2\triangleq\{\lambda_2\geq 0:\widetilde{\lambda}_1(\lambda_2)\geq 0,(\widetilde{\lambda}_1(\lambda_2),\lambda_2)$ satisfies (6b), (15a)$\}$.
2) The functions $\widetilde{\lambda}_1(\lambda_2)$ and $p_k^\star(\widetilde{\lambda}_1(\lambda_2),\lambda_2)$ in Phase 1 in (21) are both decreasing[3] in $\lambda_2\geq 0$.
3) $\widetilde{\lambda}_1(\lambda_2)\geq 0$ if and only if (iff) $\lambda_2\leq\lambda_2'$, where λ_2' is the solution of λ_2 in $\widetilde{\lambda}_1(\lambda_2)=0$.
4) The constraint (6b) for the optimization in (21) is satisfied iff $\lambda_2\leq\lambda_2''$, where λ_2'' is the value of λ_2 such that (6b) and (6c) holds with equality.

Theorem 2: The optimal solution to Problem P1 is

$$J(\widetilde{K})=\sum_{k=1}^{K}p_k^\star(\widetilde{\lambda}_1(\lambda_2^\star),\lambda_2^\star), \tag{22}$$

where the optimal $\lambda_2^\star$ is given by $\lambda_2^\dagger\triangleq\min\{\lambda_2',\lambda_2''\}$.

Proof: From 1) in Lemma 3, Problem P1 reduces to the problem in (21). Furthermore, from 3) and 4) in Lemma 3, a necessary condition for the optimal $\lambda_2^\star$ is that it must be smaller than both λ_2' and λ_2'', i.e., $0\leq\lambda_2^\star\leq\lambda_2^\dagger\triangleq\min\{\lambda_2',\lambda_2''\}$. Without loss of generality, Problem P1 becomes

$$J(\widetilde{K})=\max_{\lambda_2\in\Lambda_2'}\sum_{k=1}^{K}p_k^\star(\widetilde{\lambda}_1(\lambda_2),\lambda_2) \tag{23}$$

where $\Lambda_2'\triangleq\{0\leq\lambda_2\leq\lambda_2^\dagger:(\widetilde{\lambda}_1(\lambda_2),\lambda_2)$ satisfies (15a)$\}$. From 2) in Lemma 3, we choose λ_2 as large as possible to minimize $p_k^\star(\widetilde{\lambda}_1(\lambda_2),\lambda_2)$ for all k. Thus, the optimal solution is $\lambda_2^\star=\lambda_2^\dagger$. To complete the proof, note that the constraint (15a) is not violated. If $\lambda_2^\star=\lambda_2'$, we get (6); if $\lambda_2^\star=\lambda_2''$, we get $\lambda_1=0$; both cases imply (15a). ■

Remark 4 (Water-filling is optimal in Phase 2): For the slots in Phase 2, the optimal power allocation in (14b) is equivalent to performing water-filling with the water level λ_2 over the slots given the channel $\widetilde{\gamma}_k$. Although water-filling is optimal in Phase 2, λ_2 also appears in the power allocation in Phase 1. Hence λ_1,λ_2 cannot be optimized independently.

Interestingly, the optimal power allocation displays a similar water-filling behavior as in the optimal power allocation for a point-to-point communication system, in that (i) strictly positive power is allocated iff the channel SNRs are sufficiently large according to some *power allocation threshold*, and (ii) the power allocated is *monotonic* with respect to the channel SNR, i.e, more power is allocated for channels with higher SNR. Remarks 5 and 6 elaborates both remarks, respectively.

Remark 5 (Power Allocation Threshold in Phase 1): The optimal power allocation (14a) in Phase 1 is no longer a straightforward water-filling algorithm, unlike in Phase 2. This is because the source needs to transmit common information or *multicast* to two nodes, and the link SNRs of both links jointly determine the optimal power allocation. Nevertheless, we see from (14a) and (19) that no power is allocated if $\lambda_1\gamma_k^{\mathsf{SR}}+\lambda_2\gamma_k^{\mathsf{SD}}<1$. That is, a linearly weighted SNR, with the Lagrange multipliers as weights, serves as a threshold to determine if the source should transmit at all. This differs from the conventional water-filling where there is only one single SNR threshold (known typically as the water level).

Remark 6 (Monotonicity): Given $\lambda_1,\lambda_2\geq 0$, the optimal power allocated $p_k^\star, k=1,\cdots,K$, is non-decreasing as any of the channel SNRs $\gamma_k^{\mathsf{RD}},\gamma_k^{\mathsf{SR}},\widetilde{\gamma}_k$ increases, as verified via (14a) and (16) for $k=1,\cdots,\widetilde{K}$, and also via (14b) for $k=\widetilde{K}+1,\cdots,K$.

IV. A Heuristic Scheme with Causal CSI

So far we assume that full CSI is available. In practice, only causal CSI, in the form of present and past SNRs, is available. We shall employ the optimal structural result in Lemma 2 to develop a heuristic scheme that uses only causal CSI. Observe that for a given $\widetilde{K}$:

- The optimal power allocation depends only on two variables λ_1 and λ_2, which can be determined optimally if full CSI is available.
- Given λ_1 and λ_2, only causal CSI, namely the SNRs of the current slot, is needed to obtain the optimal power allocation.

[3] By decreasing (increasing), we mean non-increasing (decreasing).

Hence, the problem of requiring the CSI of future slots is resolved given λ_1 and λ_2. These observations motivate the following key ideas in the heuristic scheme:

- We fix λ_1 and λ_2 over different SNR conditions, denoted as λ_1^{fix} and λ_2^{fix} respectively.
- Given λ_1^{fix} and λ_2^{fix}, we then compute the power allocation according to $p_k^\star$ in Lemma 2. The relay will then decode the message at some slot index denoted as $\widetilde{K}^{\text{realized}}$, which may be greater than the maximum slot length K.
- We introduce another parameter $\widetilde{K}^{\text{fix}} \leq K$, as the maximum slot index where the relay has to decode the message. In the proposed algorithm, we will force the relay to decode the message at slot index $\widetilde{K}^{\text{realized}} \leq \widetilde{K}^{\text{fix}}$.

Thus, the set of parameters $\mathcal{P} \triangleq \{\lambda_1^{\text{fix}}, \lambda_2^{\text{fix}}, \widetilde{K}^{\text{fix}}\}$ is fixed for a given network scenario, but the instantaneous SNRs can still change over different links and slots.

Typically, the average SNR conditions is known, say by long-term measurements, for a given network scenario. Given the average SNRs, we can then perform an offline optimization to determine the best set of parameters $\mathcal{P}$. Specifically, we can obtain the average sum power by Monte Carlo simulations for a reasonable set of $\mathcal{P}$, then choose the best $\mathcal{P}$ that gives the smallest average sum power.

We now describe in detail our proposed heuristic algorithm. In the algorithm, we track the mutual information accumulated at R and D, denoted as I_k^{R} and I_k^{D}, respectively, for slot $k \in \mathcal{K}$. If $I_k^{\mathsf{R}} \geq B$, we proceed to Phase 2; if $I_k^{\mathsf{D}} \geq B$, we exit the algorithm. First, consider the slots in the Phase 1.

- If the slot index $k = \widetilde{K}^{\text{fix}}$, we allocate the power to ensure the relay decodes the message; if less power is needed for the destination to decode the message, we then instead allocate this required power. Thus, the allocated power is $p_k = \min\left\{\frac{2^{B-I_k^{\mathsf{D}}}-1}{\gamma_k^{\mathsf{SD}}}, \frac{2^{B-I_k^{\mathsf{R}}}-1}{\gamma_k^{\mathsf{SR}}}\right\}$.
- If the slot index $k < \widetilde{K}^{\text{fix}}$, we allocate the power according to Lemma 2; if less power is needed for the destination to decode the message, we then instead allocate this required power. Thus, the allocated power is $p_k = \min\left\{\frac{2^{B-I_k^{\mathsf{D}}}-1}{\gamma_k^{\mathsf{SD}}}, [p_k^\circ(\lambda_1, \lambda_2)]^+\right\}$.

Next, consider the slots in the Phase 2.

- If the slot index $k = K$, we allocate power such that the destination decodes the message, i.e., $p_k = \frac{2^{B-I_k^{\mathsf{D}}}-1}{\gamma_k^{\mathsf{SD}}}$.
- If the slot index $k < K$, we perform the same power allocation as in Phase 1 for $k < \widetilde{K}^{\text{fix}}$.

To obtain numerical results, we assume we need to deliver one bit per slot per channel use over $K = 10$ slots. We assume independent Rayleigh-fading channels for every link and slot. The distances from S to R, R to D, and S to D are d, $1-d$ and 1, respectively, where d is varied. The average SNR of each link of distance $\tilde{d}$ is given by $\bar{\gamma}\tilde{d}^{-l}$, where $\bar{\gamma} = 10$ is the normalized average SNR and $l = 4$ is the path loss exponent.

The average sum power is obtained over 50000 independent Monte Carlo runs for varying distance d, see Fig. 2. For our heuristic algorithm, each cross in Fig. 2 corresponds to a set of parameters $\mathcal{P}$. We recall that the heuristic algorithm uses only causal CSI for power allocation. Note that in practice, once the relay placement is fixed, the best $\mathcal{P}$ is chosen for implementation. As benchmark, we consider the case where full CSI is available where the optimal power allocation is obtained by Theorem 2. We also consider the case where no relay is present and the optimal power allocation can be obtained by the conventional water filling algorithm; in this case, the sum power does not change with the relay placement.

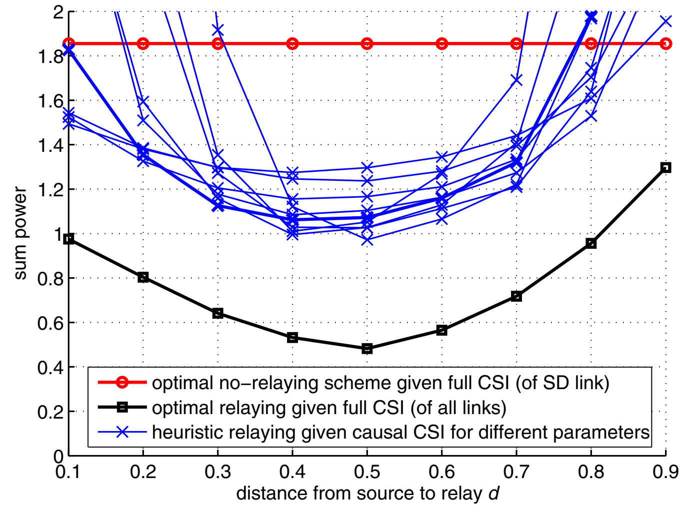

Fig. 2. Sum power incurred to deliver $B = 1$ bit per slot per channel use over $K = 10$ slots. The sum power incurred by the proposed heuristic scheme is plotted with $\times$ for different parameters $\lambda_1^{\text{fix}}, \lambda_2^{\text{fix}}$ and $\widetilde{K}^{\text{fix}}$.

From Fig. 2, we see that with relaying, and even with causal CSI, the reduction of sum power is significant, especially if the relay is placed midway between S and D. Hence, the availability of CSI is effective in reducing the energy consumption.

V. Conclusion

We have obtained analytical solutions to the problem of energy minimization of a slotted relay system that guarantee data delivery within a deadline, assuming full CSI is available. The results can be of interest to many communication scenarios, such as in cellular systems where the energy cost is substantial, yet the data needs to be delivered within some strict deadline. We also used the insights to develop heuristic schemes assuming causal CSI is available.

Appendix

A. Proof of Lemma 1

We sketch the proof to obtain J_1^{static}. It can be shown that the optimal power to obtain J_1^{static} is unique and is the same for all k in Phase 2. This can be proved for example by contradiction, using the fact that $I(\cdot)$ is strictly convex. Without loss of optimality, we thus let $p_k = p$ for all k in (9). Since $I(\cdot)$ is a strictly increasing function, we can then obtain $p^\star = q_1^\star$ and thus $J_1^{\text{static}} = \widetilde{K} q_1^\star$. The proof for J_2^{static} is similar and is omitted.

B. Proof of Theorem 1

Fix $1 \le \widetilde{K} \le K-1$. Using Lemma 1, (8) becomes

$$J^{\text{static}} = \min_{0 \le \widetilde{B} \le B} \widetilde{K} q_1^\star(\widetilde{B}) + (K - \widetilde{K}) q_2^\star(\widetilde{B}) \quad (24a)$$

$$= \min_{\widetilde{B}_1^\star \le \widetilde{B} \le B} \widetilde{K} q_1^\star(\widetilde{B}) + (K - \widetilde{K}) q_2^\star(\widetilde{B}) \quad (24b)$$

$$= \min_{\widetilde{B}_1^\star \le \widetilde{B} \le B} f(\widetilde{B}) \quad (24c)$$

$$= f\left([\widetilde{B}_2^\star]_{\widetilde{B}_1^\star}^{B} \right) \quad (24d)$$

where $[x]_y^z, y \le z$, denotes a function of x that takes the value of x if $y \le x \le z$, of y if $x \le y$ and of z is $x \ge z$. To see that (24b) holds, suppose we restrict the solution space instead to $0 \le \widetilde{B} \le \widetilde{B}_1^\star$, which implies $q_1^\star(\widetilde{B}) = \max\{q_1'(\widetilde{B}), q_1''\} = q_1''$ independent of $\widetilde{B}$. Since $q_2^\star(\widetilde{B})$ is decreasing in $\widetilde{B}$, the optimal solution must be $\widetilde{B}_1^\star$. Thus, it is sufficient to consider the solution space $\widetilde{B}_1^\star \le \widetilde{B} \le B$. Next, (24c) holds because in the refined solution space $q_1^\star(\widetilde{B}) = q_1'(\widetilde{B})$ and so the objective function is given by $f(\widetilde{B})$; (24d) holds because $f(\cdot)$ is a convex function, and so to determine the global optimal solution it is sufficient to compare the extreme boundary solutions $\widetilde{B}_1^\star$ and B, as well as the stationary solution which can be obtained as $\widetilde{B}_2^\star$. Finally, we obtain (13) as $\widetilde{K} < K$ implies $\widetilde{B}_2^\star < \widetilde{K}/KB < B$, so $[\widetilde{B}_2^\star]_{\widetilde{B}_1^\star}^{B} = \max\{\widetilde{B}_1^\star, \widetilde{B}_2^\star\}$.

C. Proof of Lemma 2

Consider Phase 2. We omit the standard derivations that use the KKT conditions to obtain the water-filling solution in (14b); e.g. see [15, Example 5.2]. For subsequent derivations, we use the fact that $\lambda_2^\star > 0$; this follows from (14b) and the assumption that R is active, i.e., $p_k^\star > 0$ for some k in Phase 2.

Consider Phase 1. From the KKT optimality conditions,

$$\partial \mathcal{L}/\partial p_k = 0 \Rightarrow 1 = \mu_k + \frac{\lambda_1 \gamma_k^{\mathsf{SR}}}{1 + \gamma_k^{\mathsf{SR}} p_k} + \frac{\lambda_2 \gamma_k^{\mathsf{SD}}}{1 + \gamma_k^{\mathsf{SD}} p_k}. \quad (25)$$

The two solutions of p_k in (25) are $p_k^+ = \Phi_{1k} + \sqrt{\Phi_{1k}^2 + \Phi_{2k}}$ and $p_k^- = \Phi_{1k} - \sqrt{\Phi_{1k}^2 + \Phi_{2k}}$, where

$$\Phi_{1k} \triangleq \frac{1}{2} \left(\frac{\lambda_1 + \lambda_2}{1 - \mu_k} - \frac{1}{\gamma_k^{\mathsf{SR}}} - \frac{1}{\gamma_k^{\mathsf{SD}}} \right), \quad (26)$$

$$\Phi_{2k} \triangleq \frac{1}{1 - \mu_k} \left(\frac{\lambda_1}{\gamma_k^{\mathsf{SD}}} + \frac{\lambda_2}{\gamma_k^{\mathsf{SR}}} \right) - \frac{1}{\gamma_k^{\mathsf{SR}} \gamma_k^{\mathsf{SD}}}. \quad (27)$$

Given $p_k \ge 0$, from (25) we get

$$\lambda_1 \gamma_k^{\mathsf{SR}} + \lambda_2 \gamma_k^{\mathsf{SD}} \ge 1 - \mu_k \quad (28)$$

implying $\Phi_{2k} \ge 0$. Thus, $p_k^- \le 0$ is infeasible except at $p_k^- = 0$. Without loss of optimality, we take p_k^+ as the only feasible solution for (25). We will use the CSC [15] that $\mu_k^\star p_k^\star = 0$.
(a) Suppose $\lambda_1^\star \gamma_k^{\mathsf{SR}} + \lambda_2^\star \gamma_k^{\mathsf{SD}} \le 1$. Let $\eta \triangleq {\lambda_1^\star}^2 \gamma_k^{\mathsf{SR}}/(1 + \lambda_1^\star p_k^\star) + {\lambda_2^\star}^2 \gamma_k^{\mathsf{SD}}/(1 + \lambda_2^\star p_k^\star)$, where $\eta > 0$ since $\lambda_2^\star > 0$ as noted earlier. From (25), after some algebraic manipulations we get $\mu_k^\star \ge p_k^\star \eta$ and so $\mu_k^\star > 0$. From the CSC $\mu_k^\star p_k^\star = 0$, $p_k^\star = 0$.
(b) Suppose $\lambda_1^\star \gamma_k^{\mathsf{SR}} + \lambda_2^\star \gamma_k^{\mathsf{SD}} > 1$. From (25), $\mu_k^\star < p_k^\star \eta$ where $\eta > 0$. Thus, $p_k^\star > 0$; otherwise if $p_k^\star \le 0$ then $\mu_k^\star < 0$. Then from the CSC $\mu_k^\star p_k^\star = 0$, we get $\mu^\star = 0$. Thus, $p_k^\star = p_k^+$ with $\mu_k = 0$, i.e., $p_k^\star = p_k^\circ$ after some algebraic manipulations.
From (a) and (b) above, we get $p_k^\star$ in (14a) where λ_1, λ_2 are the optimal Lagrangian multipliers. By inspecting (16), $p_k^\circ \ge 0$ holds iff (19) holds. Finally, (6b) and (6c) (primal constraints), and (15) (CSC) follow from the KKT optimality conditions.

D. Proof of Lemma 3

1) If the inequality in (6c) is strict, we can reduce $p_k^\star$ in Phase 1 to reduce the sum power, which implies that the power allocation cannot be optimal. By contradiction, for the power allocation to be optimal, (6c) must hold with equality.

2) Suppose that (6c) holds with equality, so we can write $\widetilde{\lambda}_1 = \lambda_1$. As λ_2 increases, from (14b) $p_k^\star$ in Phase 2 increases. From (6c) with equality, $p_k^\star$ in Phase 1 decreases. Hence from (14a), p_k° also decreases. Given that λ_2 increases and p_k° decreases, from (16) we deduce that $\widetilde{\lambda}_1$ decreases.

3) From 2) above, $\widetilde{\lambda}_1(\lambda_2)$ is decreasing in λ_2. This monotonic relationship implies that $\widetilde{\lambda}_1(\lambda_2) \ge 0$ iff $\lambda_2 \le \lambda_2'$.

4) From 2) above, $p_k^\star(\widetilde{\lambda}_1(\lambda_2), \lambda_2)$ in Phase 2 is decreasing in λ_2. Thus the LHS of (6b) is also decreasing in λ_2. This monotonic relationship implies that (6b) holds iff $\lambda_2 \le \lambda_2''$.

References

[1] A. Fu, E. Modiano, and J. Tsitsiklis, "Optimal transmission scheduling over a fading channel with energy and deadline constraints," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 630–641, Mar. 2006.

[2] J. Lee and N. Jindal, "Energy-efficient scheduling of delay constrained traffic over fading channels," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1855–1875, Apr. 2009.

[3] ——, "Asymptotically optimal policies for hard-deadline scheduling over fading channels," Jun 2009. [Online]. Available: http://arxiv.org/abs/0906.5397

[4] W. Chen, M. J. Neely, and U. Mitra, "Energy-efficient scheduling with individual packet delay constraints over a fading channel," *Wireless Networks*, vol. 15, no. 5, pp. 601–618, Jul. 2009.

[5] M. Zafer and E. Modiano, "A calculus approach to energy-efficient data transmission with quality-of-service constraints," *IEEE/ACM Trans. on Networking*, vol. 17, no. 3, pp. 898 –911, Jun. 2009.

[6] M. Chen, S. Serbetli, and A. Yener, "Distributed power allocation strategies for parallel relay networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 552–561, Feb. 2008.

[7] Z. Zhou, S. Zhou, J.-H. Cui, and S. Cui, "Energy-efficient cooperative communication based on power control and selective single-relay in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3066–3078, Aug. 2008.

[8] D. Rajan, "Power efficient delay allocation in multihop wireless networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 4, pp. 1813 –1825, Jul. 2007.

[9] L. Lin, X. Lin, and N. Shroff, "Low-complexity and distributed energy minimization in multihop wireless networks," *IEEE/ACM Trans. Netw.*, vol. 18, no. 2, pp. 501 –514, Apr. 2010.

[10] W. Chen, M. J. Neely, and U. Mitra, "Delay-constrained energy-efficient scheduling over a multihop link," Jun. 2007, pp. 1406 –1410.

[11] C. K. Ho, P. H. Tan, and S. Sun, "Relaying for energy-efficient scheduling with deadline," in *Proc. IEEE Int. Conf. on Commun.*, Cape Town, South Africa, May 2010.

[12] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inf. Theory*, vol. 47, no. 5, pp. 1971–1988, Jul. 2001.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, Inc., 2006.

[14] G. Kramer, I. Maric, and R. D. Yates, *Cooperative Communications*, ser. Foundations and Trends in Networking. Hanover, MA: Now Publishers Inc., 2006, vol. 1, no. 3-4.

[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.