Communication Schemes with Constrained Reordering of Resources

Petar Popovski, *Senior Member, IEEE*, Zoran Utkovski, *Member, IEEE*, and Kasper F. Trillingsgaard

***Abstract*—This paper introduces a communication model inspired by two practical scenarios. The *first scenario* is related to the concept of *protocol coding*, where information is encoded in the actions taken by an existing communication protocol. We investigate strategies for protocol coding via combinatorial reordering of the labelled user resources (packets, channels) in an existing, primary system. However, the degrees of freedom of the reordering are constrained by the operation of the primary system. The *second scenario* is related to communication systems with energy harvesting, where the transmitted signals are constrained by the energy that is available through the harvesting process. We have introduced a communication model that covers both scenarios and elicits their key feature, namely the constraints of the primary system or the harvesting process. We have shown how to compute the capacity of the channels pertaining to the communication model when the resources that can be reordered have binary values. The capacity result is valid under arbitrary error model in which errors in each resource (packet) occur independently. Inspired by the information–theoretic analysis, we have shown how to design practical error–correcting codes suited for the communication model. It turns out that the information–theoretic insights are instrumental for devising superior design of error–control codes.**

***Index Terms*—Protocol coding, capacity, secondary channel, energy harvesting.**

I. Introduction

A. Motivating Scenarios

THE communication models and schemes treated in this paper are motivated by two scenarios.

1) Secondary channel: Consider Fig. 1, where a cellular base station (BS) serves a group of *primary terminals* in its range. It is assumed that the cellular system is frame–based (WiMax [1], LTE [2], etc.). The metadata contained in the frame header informs the terminals how to receive/interpret the actual data that follows. The frame header is commonly encoded more robustly compared to the data, such that it can be reliably received in an area that is larger than the nominal coverage area, as depicted on Fig. 1. In such a context, while still using the same infrastructure, we can introduce new *secondary devices*, which are able to operate in the extended coverage area. These can be e. g. machine-type devices [3], such as sensors or actuators, that are controlled by the cellular BS. The secondary devices are simple and have a limited functionality, capable to decode only the frame header, but not the complex high–rate modulation used for data.

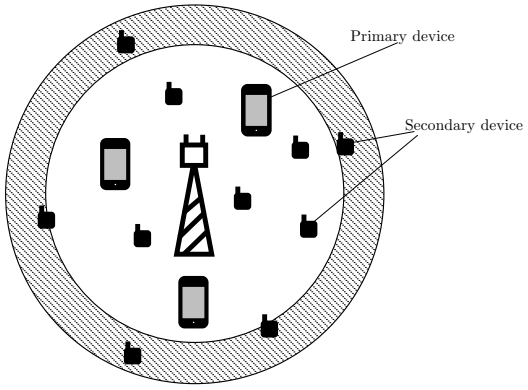


Fig. 1. Illustration of a secondary communication through protocol coding in cellular systems. A primary device can decode any information sent by the base station, while the secondary device has a limited functionality can only decode the information sent by protocol coding. The range of the primary communication system (white circle) is smaller than the range of the secondary information (shaded circle).

In such a setting, the BS can send reliable data to the secondary devices in the following way. Assume that there are F frequency channels and each of the channels can be allocated either to primary user 0 or user 1. The actual allocation is announced in the frame header and is received by all devices, primary and secondary. Then the BS can encode additional, *secondary data*, into the actual arrangement of the users on the channels. For example, if $F = 4$ and there are two channels allocated to each user, then the number of bits that can be sent by reordering the users across the channels is $\log_2 \binom{4}{2} = 2.58$ [bits/frame]. However, the challenge is that the resources available for reordering are not controlled by the transmitter of the secondary data. The primary system schedules the resources to the primary users based on criteria that are independent of the secondary communication. Therefore, the number of channels $s \leq F$ allocated to user 1 in a given frame is a constraint imposed by the primary system. The key question treated in this paper is how to encode secondary information by reordering the resources (users to channels), provided that the primary system provides a constraint through the random choice of s.

2) Energy harvesting: This is a class of systems in which the energy that is available for communication is supplied through a process of harvesting, such that the energy supply is

The associate editor coordinating the review of this paper and approving it for publication was Dr. A Khisti. Manuscript received November 26, 2010; revised November 19, 2012.

P. Popovski is with the Department of Electronic Systems, Aalborg University, Denmark (e-mail: petarp@es.aau.dk).

Z. Utkovski was with the Department of Information Technology at University of Ulm, Germany. He is now with the Faculty of Computer Science, University Goce Delčev Štip, R. Macedonia (e-mail: zoran.utkovski@uni-ulm.de).

K. F. Trillingsgaard is with Aalborg University, Denmark (e-mail: kasperft@gmail.com).

Digital Object Identifier 10.1109/TCOMM.2013.09.120027

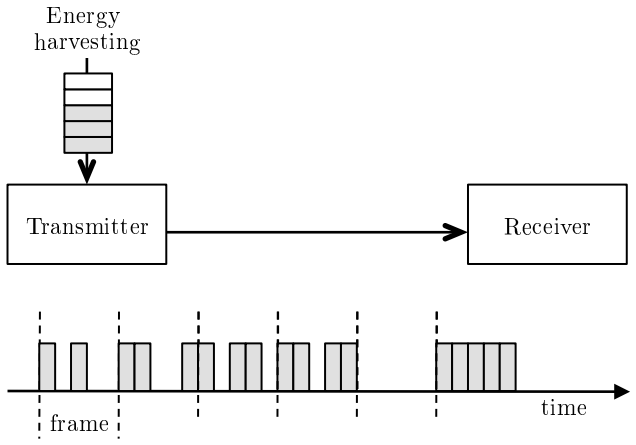


Fig. 2. Communication system with energy harvesting and ON/OFF signaling. The frame size if $F = 5$.

dependent on external factors. Consider the setting on Fig. 2, where communication is organized in frames of length F. The system uses ON/OFF modulation, such that the symbol 'ON' or '1' consumes a single quantum of energy, while the symbol 'OFF' or '0' does not consume energy. The value of F is chosen to correspond to the size of the energy buffer, such that in a given frame the transmitter can harvest between 0 and F energy quanta. In this paper we assume that *all* the energy quanta harvested in a given frame must be used in the next frame; see the discussion in Section V. The transmitter sends data by ordering the s ON symbols in the frame in one of the $\binom{F}{s}$ possible ways. Since s is externally constrained and can even be random, the challenge is similar to the secondary communication: to encode information through reordering, but under a constraint.

B. Related Work and Contributions

Models of communication with reordering of resources and ideas related to secondary communication have appeared before in various contexts and under different names. to describe the essence of those ides, we use the term *protocol coding*: encode information in the actions taken by a certain (existing) communication protocol. An early work that mentions the possibility to send data by modulating the random access protocol is [4], but in a rather "negative" context, since the model used *explicitly prohibits* to decide the protocol actions based on user data. The seminal work [5] uses a form of protocol coding, as information is modulated in the arrival times of data packets. More recent works on possible encoding of information in relaying scenarios through *protocol–level* choice of whether to transmit or receive is presented in [6], [7] and, [8]. At a conceptual level, protocol coding bridges information theory and networking [9]. The idea to send data by reordering packets is certainly not new and has been presented in several works [10], [11], [12]. However, a distinction for our work is the constraint put on the reordering, which gives rise to completely novel communication strategies. The practical coding strategies are related to the frequency permutation arrays for power line communications [13], [14].

Communication systems with energy harvesting is an emerging research area. In [15], [16], [17], [18] the authors describe continuous-time systems with energy harvesting and compute offline transmission schemes that are optimal in terms of throughput or minimization of the time for completing a transmission. Slot-based energy harvesting systems with an infinite-capacity battery are treated in [19], where the authors use an information-theoretic model and introduce the save-and-transmit scheme, which is proved to achieve the capacity of the AWGN channel.

Preliminary results of this work have appeared in [20] and [21]. In [20] we have introduced the secondary channel, assuming that the primary packets pass through an erasure channel, while in [21] we have used the Z–channel. In this paper we present a general communication model that corresponds to the two motivating scenarios. We relate the model to the channels with causal side information at the transmitter (CSIT) [22]. However, using the specific features of the communication model, we provide an explicit characterization of the capacity-achieving strategies for general binary-input memoryless channels. We then show how the insights from the information-theoretic analysis can be used to devise practical coding strategies, based on trellis codes. Besides the two described scenarios, the communication model gives rise to communication channels whose analysis goes beyond the purpose of current applications, as it is of more general information-theoretic interest.

The paper is organized as follows. Section II introduces the communication model that covers both scenarios. Section III contains information–theoretic results about the considered model and shows how to calculate the capacity. Those results are used in Section IV to show practical coding strategies, along with numerical illustration. Section V discusses practical features and applications of the secondary channels based on protocol coding. The last section concludes the paper and outlines directions for future work.

II. System Model

The communication model defined here encompasses the features of both the secondary data and the energy harvesting. A transmitter sends data to the receiver in frames, each frame consisting of F *slots*. Each slot can have the value 0 or 1. In the context of secondary communication, 0 and 1 can be interpreted as addresses of the primary terminals to which the packets are scheduled. In the context of energy harvesting, $0(1)$ means absence (presence) of transmission. A given frame has s slots with value 1 and $F - s$ slots with value 0. The number of $1-$slots in a frame, s, is termed *state* of the frame. We assume that the frame state is selected in a random and memoryless fashion, thus modeling the behavior of an external factor (primary scheduler in secondary communication or nature in energy harvesting). Specifically, the probability that a frame is in state s is binomial

$$P_S(s) = \binom{F}{s} a^s (1-a)^{F-s} \tag{1}$$

For the energy harvesting scenario, a can be understood as the probability that an energy quantum arrives in a given slot of the previous frame, which is now at disposal for the current frame.

The transmitter knows the frame state causally, at the start of the frame, such that it only send information by reordering the 1s and 0s in the frame. For example, if $F = 4$ and $S = 3$, then the possible transmit symbols are $1110, 1101, 1011, 0111$. But, if $S = F = 4$, the transmitter cannot send any data in the corresponding frame. We use the term *symbol* to denote a frame that consists of F slots. A symbol represents a single channel use in our system. An input symbol is an $F-$dimensional binary vector $\mathbf{x} = (x_1, x_2, \dots x_F) \in \mathcal{X} = \{0,1\}^F$. An output symbol is also $F-$dimensional vector $\mathbf{y} = (y_1, y_2, \dots y_F) \in \mathcal{Y} = \mathcal{J}^F$, where the cardinality is $|\mathcal{J}| \geq 2$.

Each slot is sent over a binary-input memoryless channel, defined by the distribution $p_{Y|X}(y|x)$, where $x \in \{0,1\}$, $y \in \mathcal{J}$ and $|\mathcal{J}| = J$. For example, for the error model with erasures, $\mathcal{J} = \{0, 1, \epsilon\}$, corresponds to the secondary communication in which the packet header is either received correctly or it is erased (wrong checksum). Another example is the Z-channel, where $\mathcal{J} = \{0,1\}$ and $p_{Y|X}(1|0) = 0$, i.e. errors can occur only when 1 is sent. The Z-channel can be used in the energy harvesting scenario, where the ON signal ($X = 1$) corresponds to transmission of, for example, a particular spread-spectrum sequence of length n. The motivation for this model can be explained as follows. In absence of energy transmission, the probability that the noise produces the spread-spectrum sequence decreases exponentially with n, such that we can approximate that situation by setting $p_{Y|X}(1|0) = 0$.

In general, the communication channel applicable to a single slot can be described by J transition probabilities, represented by a vector:

$$\mathbf{q}_i = (q_{i1}, q_{i2}, \dots q_{iJ}) \qquad i = 0, 1 \tag{2}$$

where $q_{ij} = P(y = j | x = i)$ and some q_{ij} can be equal to 0. Using the elementary channels applied to each slot, we can define the transition probabilities for the channel of interest $\mathbf{X} - \mathbf{Y}$:

$$P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{f=1}^{F} q_{x_f y_f} \tag{3}$$

The key constraint on the communication comes from the random channel state $s \in \mathcal{S} = \{0, 1, \dots F\}$. The set of input symbols $\mathcal{X}$ is partitioned into $F + 1$ subsets $\mathcal{X}_s$ defined as follows:

$$\mathbf{x} \in \mathcal{X}_s \Leftrightarrow \sum_{i=1}^{F} x_i = s \tag{4}$$

When the frame state is $S = s$, then the transmitter can only sent an input symbol $\mathbf{x} \in \mathcal{X}_s$.

III. Computing the Capacity for the Communication Channel with Constrained Reordering

A simple upper bound on the capacity of the considered channels is 1 bit per slot i. e. F bits per frame. The communication model considered here is related, but not identical, to the channels with causal state information at the transmitter (CSIT) [22]. In a channel with causal CSIT, the state $S = s$ is memoryless, while the channel is defined by specifying $P_{\mathbf{Y}|\mathbf{X},S}(\mathbf{y}|\mathbf{x}, s)$ for all $s \in \mathcal{S}, \mathbf{x} \in X, \mathbf{y} \in \mathcal{Y}$. Shannon showed that instead of considering the original channel with CSIT, one can consider an ordinary, discrete memoryless channel with equal capacity, but with a larger input alphabet. The input variable of the equivalent channel is T and each possible input letter t, termed *strategy* [23], represents a mapping from the state alphabet $\mathcal{S}$ to the input alphabet $\mathcal{X}$ of the original channel. A particular strategy $t \in \mathcal{T}$ is defined by the vector of size $|\mathcal{S}|$: $(t(1), \dots t(|\mathcal{S}|))$, where $t(s) \in \mathcal{X}$. Therefore, if each $s \in \mathcal{S}$ can be mapped to any $\mathbf{x} \in \mathcal{X}$, then the total number of possible strategies is $|\mathcal{X}|^{|\mathcal{S}|}$ and therefore $|\mathcal{T}| \leq |\mathcal{X}|^{|\mathcal{S}|}$. The key result is that, to achieve the capacity, it is sufficient that the channel input of the $n-$th channel use $\mathbf{x}_n$ depends only on the message and the current state $S = s_n$, but not the past states.

In our communication model, the state s defines which inputs are possible to use, i. e., $S = s$ implies $\mathbf{x} \in \mathcal{X}_s$. While causal CSIT channels with input-cost constraints have been studied in the literature, to our knowledge no prior work where the cost constraints also involve the state variables exists. Thus, the results for channels with causal CSIT cannot be directly applied, as $P_{\mathbf{Y}|\mathbf{X},S}(\mathbf{y}|\mathbf{x}, s)$ is not defined if $\mathbf{x} \notin \mathcal{X}_s$. On the other hand, it is intuitively clear that the result for the channels with causal CSIT can be used if we restrict the strategies only to the ones that have valid mappings, i. e. $t(s) \in \mathcal{X}_s$. This is proven in the following proposition.

Proposition 1: Let a channel and its state be memoryless. For given channel input $\mathbf{x}$, the channel is defined by $p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})$, where $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$. The channel is restricted by a state, such that if $S = s$, then the input symbol must be $\mathbf{x} \in \mathcal{X}_s$, where $\cup_{s\in\mathcal{S}} \mathcal{X}_s = \mathcal{X}$ and $\mathcal{X}_{s_1} \cap \mathcal{X}_{s_2} = \emptyset$ if $s_1 \neq s_2$. Then the capacity of the channel is:

$$C = \max_{P_U(\cdot), f:\mathcal{U}\times s\to\mathcal{X}_s, \forall s\in\mathcal{S}} I(U; \mathbf{Y}) \tag{5}$$

the joint distribution of the random variables $S, U, \mathbf{X}, \mathbf{Y}$ is given by

$$P_{S,U,\mathbf{X},\mathbf{Y}}(s, u, \mathbf{x}, \mathbf{y}) = P_S(s) P_U(u) \delta(\mathbf{x}, f(u, s)) P_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) \tag{6}$$

where U is auxiliary random variable with support set $\mathcal{U}$ and $|\mathcal{U}| \leq \min\left\{\prod_{s\in\mathcal{S}} |\mathcal{X}_s|, |\mathcal{Y}|\right\}$. The indicator function is defined as $\delta(\mathbf{x}, f(u, s)) = 1$, for $\mathbf{x} = f(u, s)$ and $\delta(\mathbf{x}, f(u, s)) = 0$ otherwise.

Proof: The proof is along the line of the proof [23] (pages 456-457). The key argument is that, to achieve the capacity, then at a certain channel use n it is sufficient that U depends only on the message M, but not the sequence of the previous states. The fact that for given $U = u$ and $S = s$ the function f must be constrained to be $f(u, s) \in \mathcal{X}_s$ does not alter this argument. Note that each fixed $U = u$ defines one bijective mapping $t_u : \mathcal{S} \to \mathcal{X}$, restricted such that $t_u(s) \in \mathcal{X}_s$. Then the number of possible mappings is $\prod_{s\in\mathcal{S}} |\mathcal{X}_s|$, which is an upper bound on the required cardinality $|\mathcal{U}|$. Following the properties of mutual information ([24], Section 8.3), it should also be $|\mathcal{U}| \leq |\mathcal{Y}|$. This proves the proposition. ■

Proposition 1 covers channels that are more general than our communication model, where the set of possible states $\mathcal{S}$ defines partitioning of $\mathcal{X}$ into $|\mathcal{S}|$ subsets.

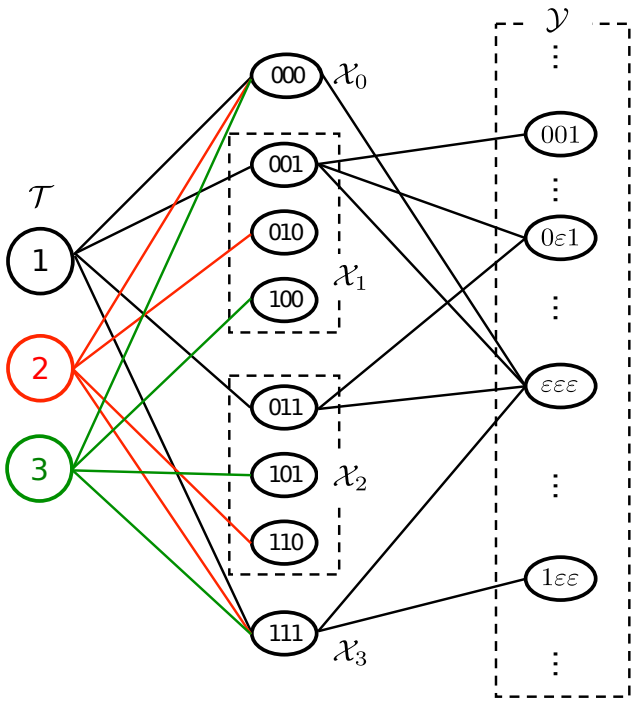


Fig. 3. Example choice of the probability distribution $P_{\mathbf{X}|T}$ with $F = 3$ and $\mathcal{T} = \{1, 2, 3\}$. The transition probabilities on the channel $\mathbf{X} - \mathbf{Y}$ are not marked, but it is assumed that each packet 0 or 1 can become erased ϵ independently with probability p.

Instead of using U as a random variable that indexes a set of functions, we can equivalently use an auxiliary variable T. Each $T = t$ corresponds to one function, as defined above, such that for given $T = t$ and each $s \in \mathcal{S}$, there is a single *representative of t in s*, which is $\mathbf{x} = t(s) \in \mathcal{X}_s$. We use the terms "strategies" and "input symbols" interchangeably. Hence, $\mathcal{T}$ consists of the input symbols $\{1, 2, \dots |\mathcal{T}|\}$. The set of $F+1$ representatives $\{\mathbf{x}_s(t)\}$ for given t will be called a *multisymbol* of t.

Example: We illustrate the communication strategies through a case with $F = 3$ and erasure channel. Fig. 3 illustrates the capacity-achieving strategies. It is sufficient to have $|\mathcal{T}| = 3$ strategies. The four edges going out of each strategy $T = 1, 2, 3$ define the four respective representatives of that strategy; e. g., the representatives of $T = 2$ are $\mathbf{x} \in \{000, 010, 110, 111\}$. Intuitively, the representatives of the same strategy should be as similar to each other as possible, while being as different as possible from the representatives of the other strategies. The similarity is measured in terms of Hamming distance among the representatives. Thus, the representatives of $T = 2$ in $\mathcal{X}_1$ and $\mathcal{X}_2$ are 010 and 110, respectively, and not 010 and 101. The reason is that in the former case the Hamming distance between the representatives is 1, which is minimal possible, while it is 3 in the latter case.

Before stating the main theorem, we will need several definitions.

Definition 1: A multisymbol associated with a given strategy $t \in \mathcal{T}$ is a set of $F + 1$ vectors

$$\mathcal{M}_t = \{\mathbf{x}_0, \mathbf{x}_1, \dots \mathbf{x}_{F-1}, \mathbf{x}_F\} \tag{7}$$

such that $\mathbf{x}_s \in \mathcal{X}_s$ for each $s \in \mathcal{S}$.

We use the following notation. $w_H(\mathbf{x})$ is the Hamming weight of the vector $\mathbf{x}$, while $d_H(\mathbf{x}_1, \mathbf{x}_2)$ is the Hamming distance between the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$.

Definition 2: A multisymbol $\mathcal{M}_t$ is termed *minimal multisymbol* if for each pair $\mathbf{x}_{s_1}, \mathbf{x}_{s_2} \in \mathcal{M}_t$ where $s_2 > s_1$ the following holds:

$$d_H(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = w_H(\mathbf{x}_{s_2}) - w_H(\mathbf{x}_{s_1}) \tag{8}$$

Definition 3: The *basic* multisymbol $\mathcal{M}^b$ has its representative $00\cdots011\cdots1$ in $\mathcal{X}_s$ that starts with $F - s$ consecutive zeros and ends with s consecutive ones.

For providing certain properties of multisymbols, it is useful to define a permutation of a multisymbol.

Definition 4: The permutation of a multisymbol $\mathcal{M}$ is

$$\mathcal{M}' = \gamma_\pi(\mathcal{M}) \tag{9}$$

where each $\mathbf{x}'_s \in \mathcal{M}'$ is obtained from the corresponding $\mathbf{x}_s \in \mathcal{M}$ by permuting the packets according to a given permutation π of length F.

For example, when $F = 3$, $\mathcal{M} = \mathcal{M}^b$ and $\pi = 321$, the permuted multisymbol is $\mathcal{M}' = \{000, 100, 110, 111\}$. Note that the permutation preserves the Hamming distance between any two representatives $d_H(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = d_H(\mathbf{x}'_{s_1}, \mathbf{x}'_{s_2}) = s_2 - s_1$. Since any minimal multisymbol can be obtained from the basic one via permutation, it follows that there are in total $F!$ different minimal multisymbols.

The secondary channel can be represented by a cascade of two channels $T - \mathbf{X} - \mathbf{Y}$. In order to express $I(T; \mathbf{Y})$, we write $I(T, \mathbf{X}; \mathbf{Y}) = I(T; \mathbf{Y}) + I(\mathbf{X}; \mathbf{Y}|T) = I(\mathbf{X}; \mathbf{Y}) + I(T; \mathbf{Y}|\mathbf{X})$. and using the Markov property for the cascade we get $I(T; \mathbf{Y}|\mathbf{X}) = 0$, which implies:

$$I(T; \mathbf{Y}) = I(\mathbf{X}; \mathbf{Y}) - I(\mathbf{X}; \mathbf{Y}|T) \tag{10}$$

such that we can write

$$\begin{aligned} C &= \max_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(T; \mathbf{Y}) \\ &\leq \max_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X}; \mathbf{Y}) - \min_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X}; \mathbf{Y}|T) \end{aligned} \tag{11}$$

where the equality is achieved if and only if there is a pair of distributions $\big(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\big)$ that simultaneously attains the max/min in the first/second term, respectively.

Let us consider the term $I(\mathbf{X}; \mathbf{Y})$ and see which distribution $P_{\mathbf{X}}(\cdot)$ can maximize it. Due to the constraints, $P_{\mathbf{X}}(\cdot)$ cannot be an arbitrary distribution on $\mathcal{X}$, but it has to belong to the following set of distributions:

$$\mathcal{P}_{\mathbf{X},S} = \left\{ P_{\mathbf{X}}(\cdot) | \sum_{\mathbf{x} \in \mathcal{X}_s} P_{\mathbf{X}}(\mathbf{x}) = P_S(s), \forall s = 0, 1, \cdots F \right\} \tag{12}$$

We then define:

$$C_{XY} = \max_{P_{\mathbf{X}}(\cdot) \in \mathcal{P}_{\mathbf{X},S}} I(\mathbf{X}; \mathbf{Y}) \tag{13}$$

The distribution in $P_{\mathbf{X}}(\cdot) \in \mathcal{P}_{\mathbf{X},S}$ that maximizes $I(\mathbf{X}; \mathbf{Y})$ is given by the following lemma, proved in in Appendix A.

Lemma 1: The distribution $P_{\mathbf{X}}(\cdot) \in \mathcal{P}_{\mathcal{X},\mathcal{S}}$ that achieves C_{XY} is, for all s and each $\mathbf{x} \in \mathcal{X}_s$:

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{P_S(s)}{\binom{F}{s}} \tag{14}$$

Proof: The proof is given in Appendix A. ∎

For particular $t \in \mathcal{T}$, the mutual information $I(\mathbf{X};\mathbf{Y}|T=t)$ is determined by $P_{\mathbf{X}|T}(\cdot|T=t)$, which is defined by the particular choice of the multisymbol $\mathcal{M}_t$. Therefore we will write:

$$I(\mathbf{X};\mathbf{Y}|T=t) = I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t) \tag{15}$$

Of special interest $I(\mathbf{X};\mathbf{Y}|T=t)$, when the multisymbol $\mathcal{M}_t$ is a minimal one, $\mathcal{M}_t = \mathcal{M}^m$:

$$I(\mathbf{X};\mathbf{Y}|T=t) = I(\mathbf{X};\mathbf{Y}|\mathcal{M}^m) = I_m \tag{16}$$

We can now state the main theorem:

Theorem 1: The capacity of the channel with constrained resource reordering is computed as

$$C = C_{XY} - I_m \tag{17}$$

where C_{XY} and I_m are given by (13) and (16), respectively. The capacity is achieved when for each strategy $t \in \mathcal{T}$:

- $P_{\mathbf{X}|T}(\mathbf{x}|T=t) = P_S(s)$ if $\mathbf{x} \in \mathcal{M}_t \cap \mathcal{X}_s$ and $P_{\mathbf{X}|T}(\mathbf{x}|T=t) = 0$ otherwise;
- $\mathcal{M}_t$ is a minimal multisymbol.

while T is a uniform random variable with cardinality $|\mathcal{T}| = \text{lcm}\left(\binom{F}{0}, \binom{F}{1}, \ldots, \binom{F}{F}\right)$, where lcm stands for "least common multiplier".

Proof: We first show that we can choose the distributions $\left(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\right)$ such that $I(\mathbf{X};\mathbf{Y}) = C_{XY}$. The distributions that can be induced by $\left(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\right)$ are a subset of $\mathcal{P}_{\mathbf{X},S}$, such that

$$\max_{P_T(\cdot), P_{\mathbf{X}|T}(\cdot)} I(\mathbf{X};\mathbf{Y}) \leq \max_{P_{\mathbf{X}}(\cdot) \in \mathcal{P}_{\mathbf{X},S}} I(\mathbf{X};\mathbf{Y}) \tag{18}$$

We now show that it is possible to select the pair $\left(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\right)$ that can result in the distribution $P_{\mathbf{X}}(\cdot)$ given by (14) in Lemma 1, thus achieving equality in (18). Let

$$|\mathcal{T}| = \text{lcm}\left(\binom{F}{0}, \binom{F}{1}, \ldots, \binom{F}{F}\right) = L \tag{19}$$

and let T be uniformly distributed over $\mathcal{T}$. Each $T=t$ has a single representative in each $\mathcal{X}_s$, such that $P_{\mathbf{X}|T}(\mathbf{x}|T=t) = P_S(s)$ if $\mathbf{x} \in \mathcal{M}_t \cap \mathcal{X}_s$ and $P_{\mathbf{X}|T}(\mathbf{x}|T=t) = 0$ otherwise. Since by the definition of L, the value

$$m_s = \frac{L}{|\mathcal{X}_s|} = \frac{L}{\binom{F}{s}} \tag{20}$$

is integer, we can choose the multisymbols in a way that each $\mathbf{x} \in \mathcal{X}_s$ is a representative of exactly m_s strategies $t \in \mathcal{T}$. Then for any s and any $\mathbf{x} \in \mathcal{X}_s$ it follows

$$P_{\mathbf{X}}(\mathbf{x}) = m_s \frac{1}{L} P_S(s) = \frac{P_S(s)}{\binom{F}{s}} \tag{21}$$

which is identical to the distribution given by Lemma 1. We have thus shown how to choose $\left(P_T(\cdot), P_{\mathbf{X}|T}(\cdot)\right)$ to maximize $I(\mathbf{X};\mathbf{Y})$ under the transmit constraints.

Regarding the minimization of $I(\mathbf{X};\mathbf{Y}|T)$, we write:

$$\begin{aligned} I(\mathbf{X};\mathbf{Y}|T) &= \sum_{t\in\mathcal{T}} P_T(t) I(\mathbf{X};\mathbf{Y}|T=t) \\ &= \frac{1}{|\mathcal{T}|} \sum_{t\in\mathcal{T}} I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t) \end{aligned} \tag{22}$$

For a fixed multisymbol $\mathcal{M}_t$, we decompose:

$$I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t) = H(\mathbf{Y}|\mathcal{M}_t) - H(\mathbf{Y}|\mathbf{X},\mathcal{M}_t) \tag{23}$$

We consider first:

$$H(\mathbf{Y}|\mathbf{X},\mathcal{M}_t) = \sum_{s=0}^{F} P_S(s) H(\mathbf{Y}|\mathbf{x}_s(t)) \tag{24}$$

Since each component of $\mathbf{x}_s$ uses identical memoryless channels, $H(\mathbf{Y}|\mathbf{x}_s(t))$ depends only on the Hamming weight s, but not on how the 0s and 1s are arranged in $\mathbf{x}_s$. This is proven in Lemma 2 in Appendix B, such that (24) can be rewritten as

$$H(\mathbf{Y}|\mathbf{X},\mathcal{M}_t) = \sum_{s=0}^{F} P_S(s) H_s \tag{25}$$

where H_s is given by (35). Considering the remaining member in (23), Lemma 3 in Appendix C shows that all minimal multisymbol result in an equal and minimal value of $H(\mathbf{Y}|\mathcal{M}_t)$. Therefore, if $\mathcal{M}^m$ is a minimal multisymbol, then

$$I(\mathbf{X};\mathbf{Y}|\mathcal{M}_t) = I_m \tag{26}$$

where I_m does not depend on the actual multisymbol as long as it is minimal.

It remains to show that it is possible to find $|\mathcal{T}| = \text{lcm}\left(\binom{F}{0}, \binom{F}{1}, \ldots, \binom{F}{F}\right) = L$ different minimal multisymbols that simultaneously maximize $I(\mathbf{X};\mathbf{Y})$ and minimize $I(\mathbf{X};\mathbf{Y}|T) = I_m$ in (11). We represent multisymbols by a directed graph. Fig. 4(a) shows a choice of a set of minimal multisymbols with $F=4$ and the corresponding directed graph is depicted on Fig. 4(b). Each node in the graph represents a particular $\mathbf{x} \in \mathcal{X}$. An edge exists between $\mathbf{x}_s \in \mathcal{X}_s$ and $\mathbf{x}_{s+1} \in \mathcal{X}_{s+1}$ if and only if the Hamming distance is $d_H(\mathbf{x}_s, \mathbf{x}_{s+1}) = 1$. The directed edge from $\mathbf{x}_s$ to $\mathbf{x}_{s+1}$ exists if they can both belong to a same minimal multisymbol $\mathcal{M}_t$. A multisymbol is represented by a path of length F that starts at $00\cdots0$ and ends at $11\cdots1$. To each edge we can assign a nonnegative integer, which denotes the number of multisymbols (paths) that contain that edge. On Fig. 4(b), each edge that starts from 0000 has a weight 3, each edge between an element of $\mathcal{X}_1$ and $\mathcal{X}_2$ has a weight 1, etc. The weight of each edge between $\mathbf{x}_s$ and $\mathbf{x}_{s+1}$ can be treated as an outgoing weight for $\mathbf{x}_s$ and incoming weight for $\mathbf{x}_{s+1}$.

Using the graph representation, we need to prove that, for each $s = 0 \ldots F-1$, it is possible to match all outgoing weights from $\mathcal{X}_s$ to all incoming weights from $\mathcal{X}_{s+1}$. Since L divides each $\binom{F}{s}$, the number of multisymbols that contain $\mathbf{x}_s \in \mathcal{X}_s$ is an integer $m_s = \frac{L}{\binom{F}{s}}$. The number of outgoing edges from $\mathbf{x}_s$ is $(F-s)$, while the number of incoming edges to $\mathbf{x}_s$ is s. The sum of incoming weights and the sum of outgoing weights for $\mathbf{x}_s$ is equal to m_s. Note that the average outgoing weight for $\mathbf{x}_s$ is $\frac{m_s}{F-s}$, while the average incoming weight for any $\mathbf{x}_{s+1} \in \mathcal{X}_{s+1}$ is $\frac{m_{s+1}}{s+1}$. However, the following holds $\frac{m_s}{F-s} = \frac{L}{\binom{F}{s}(F-s)} = \frac{L}{\binom{F}{s+1}(s+1)} = \frac{m_{s+1}}{s+1}$ i. e. the average outgoing weight from $\mathcal{X}_s$ is equal to the average incoming weight at $\mathcal{X}_{s+1}$, which is a necessary condition for the multisymbols that achieve the secondary capacity. We now prove that for each outgoing weight from $\mathcal{X}_s$ there is a matched

incoming weight at $\mathcal{X}_{s+1}$.We choose the weight of each edge to be either $w_1 = \lfloor \frac{m_s}{F-s} \rfloor$ or $w_2 = \lceil \frac{m_s}{F-s} \rceil$. Then b weights have to be chosen to be equal to $w_2 = \lceil \frac{m_s}{F-s} \rceil$, where b is given by

$$m_s = a(F-s) + b,\ a \in \{\mathbb{N} \cup 0\},\ 0 \leq b \leq F-s-1. \quad (27)$$

There are $s+1$ incoming edges at $\mathbf{x}_{s+1}$. The weight of each incoming edge is also either w_1 or w_2, since $\frac{m_s}{F-s} = \frac{m_{s+1}}{s+1}$. In order to satisfy the condition that the total incoming weight of $\mathbf{x}_{s+1}$ is m_{s+1}, d weights should be chosen to be equal to w_2, where d is given by

$$m_{s+1} = c(s+1) + d,\ c \in \{\mathbb{N} \cup 0\},\ 0 \leq d \leq s. \quad (28)$$

If (27) and (28) are satisfied, then $b\binom{F}{s} = d\binom{F}{s+1}$ needs to be fulfilled, which follows from $\binom{F}{s+1} = \binom{F}{s}\frac{F-s}{s+1}$ and the equality of average incoming/outgoing weights. For each outgoing weight from $\mathcal{X}_s$ there is a matched incoming weight at $\mathcal{X}_{s+1}$. Since $L \leq F!$, it will be always possible to select L different paths.

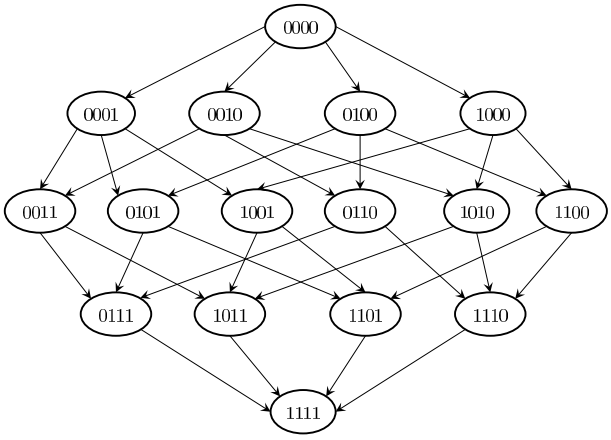
Therefore, it is always possible to select a set of L minimal multisymbols that achieve the upper bound in (11), which proves the theorem.

■

As it can be seen from Fig. 4, if $F = 4$ it turns out that $\frac{m_s}{F-s}$ is always an integer, such that all the outgoing/incoming weights to the same node are identical. This is not the case if, e. g., $F = 7$, then $L = 105$, $m_1 = 15$ and $\frac{m_1}{7-1} = \frac{15}{6}$, such that each node from $\mathcal{X}_1$ has 3 outgoing edges of weight 3 and 3 of weight 2.

t	$\{\mathbf{x}_s(t)\}$
1	(0000, 0001, 0011, 0111, 1111)
2	(0000, 0001, 0101, 1101, 1111)
3	(0000, 0001, 1001, 1011, 1111)
4	(0000, 0010, 0011, 1011, 1111)
5	(0000, 0010, 0110, 0111, 1111)
6	(0000, 0010, 1010, 1110, 1111)
7	(0000, 0100, 0101, 0111, 1111)
8	(0000, 0100, 0110, 1110, 1111)
9	(0000, 0100, 1100, 1101, 1111)
10	(0000, 1000, 1001, 1101, 1111)
11	(0000, 1000, 1010, 1011, 1111)
12	(0000, 1000, 1100, 1110, 1111)

(a)



(b)

Fig. 4. Selection of the representative sets for $F = 4$ that achieve the capacity. (a) Multisymbols for the 12 inputs (b) Graph representation of the process for selecting the multisymbols $\mathbf{x}_s(t)$.

IV. Practical Coding Strategies

In this part of the paper we consider practical coding strategies for the communication model introduced in Section II. The first thing to be noted is that, for finite packet (codeword) length, there will always be a nonzero probability of error, even if the channel itself does not introduce error. To see this, note that an unfortunate sequence of states can occur: for example, in all frames that constitute the packet, the state is $S = 0$. We will call these error *encoder errors*.

In order to emphasize the utility of the information-theoretic analysis presented so far, we take the following approach. As a reference, we first present a *"naïve" coding strategy*, which represents a design that can be undertaken without using the framework of multisymbols. We then present code design inspired by the information-theoretic analysis.

A. Naïve Coding Strategy

The naïve strategy works as follows. We take any usual error-correction code of rate R and interleave the output of this code, e. g. by using a pseudo-random interleaver. The motivation for using an interleaver is to break the burst bit errors that can occur within one secondary symbol (frame), both due to encoder or channel error. For example, for a frame length $F = 4$ we take four of the coded and interleaved bits and look at the current state of the channel, i. e., how many 1s we can transmit in the next frame. Then, we pick any (e. g. randomly) of the possible frames, obtained by permuting the packets, that has minimal possible Hamming distance. For example, let the coded bits be 0101 and let the state be $S = 3$. Then the Hamming distance of the "true information" 0101 from 0111, 1101 is 1 (minimal possible), while it is 3 from 1011 or 1110. Hence, when the system needs to transmit 0101 and the state is $S = 3$, it chooses randomly between 0111 and 1101.

B. Coding Strategy inspired from the Information–Theoretic Analysis

Here we propose a coding strategy which is inspired by the capacity results for the secondary communication channel. Recall that the result stated by Theorem 1 is quite general and holds for all classes of memoryless channels $\mathbf{X} - \mathbf{Y}$ with binary inputs. Among other channels, it holds for the erasure channel, the binary symmetric channel and the Z channel. As already argued, for a uniform distribution over $\mathcal{T}$, this capacity can be achieved by a set $\mathcal{T}$ of cardinality $L = \text{lcm}\left(\binom{F}{0}, \binom{F}{1}, \ldots, \binom{F}{F}\right)$. The L multisymbols should be minimal, meaning that the Hamming distance between two adjacent symbols is 1, $d_H(x_s, x_{s+1}) = 1$. From the viewpoint of capacity, the choice of the multisymbols is irrelevant, as long they are minimal and the distribution of X fulfills the required condition. However, the choice of the multisymbols does affect the performance of the error-correcting code constructed based on the multisymbol framework.

Our aim is to use the multisymbol framework in the construction of practical coding schemes which are better suited for the secondary communication channel than the

naïve approach. The question to ask is which criterion, e.g. distance metric we are going to use in the selection of the multisymbols. We adopt a heuristic approach and take the *expected Hamming distance* as the metric of interest. For two multisymbols t_1 and t_2, this distance is defined as follows

$$E_{d_H}(t_1,t_2)=\sum_{s=0}^{F}P_S(s)d_H(x_s(t_1),x_s(t_2)), \qquad (29)$$

where d_H is the Hamming distance between the two vectors. Clearly, considering the triviality of the states $S=0$ and $S=F$, we can simplify to:

$$E_{d_H}(t_1,t_2)=\sum_{s=1}^{F-1}P_S(s)d_H(x_s(t_1),x_s(t_2)) \qquad (30)$$

The motivation behind this is that this metric incorporates the state of the channel which can not be controlled by the secondary system.

With this in mind, we can construct a convolutional code by using the multisymbols framework and the expected Hamming distance as design criterion. We define a trellis for the convolutional code with a certain number of states. Each trellis state contains two outgoing paths, each of them corresponding to one possible input binary symbol. Also, each state has two incoming paths. Each branch in the trellis is associated with an input symbol and an output symbol, where the input symbol is binary and the output symbol is one of the L multisymbols. The trellis has L branches, such that each multisymbol is associated with only one single-step transition in the trellis diagram.

For designing the trellis transitions, we use the known rules from trellis coding: the output symbols on the branches exiting from the same state should be maximally separated in terms of the expected Hamming distance. The same is valid for the output symbols associated with the two branches that enter the same state. In order to illustrate the code construction, we take the example with $F=4$, where the minimal cardinality of the uniform auxiliary variable T is $L=\mathrm{lcm}\{\binom{4}{0},\binom{4}{1},\ldots,\binom{4}{4}\}=12$.

There are multiple ways in which the multisymbols can be chosen, and different sets have different features. We can get useful insights about the expected Hamming distance spectrum if we use the representation of the multisymbols as paths in the directed graph, as shown in Fig. 5 c). In order to maximize the expected Hamming distance between multisymbols, we have to choose the multisymbols such to avoid, if possible, to have multisymbols with common edges. Indeed, for two different multisymbols t_1 and t_2 which share a common edge $(x_j(t),x_{j+1}(t))$, the terms in the expected Hamming distance

$$E_{d_H}(t_1,t_2)=\sum_{s=0}^{F}P_S(s)d_H(x_s(t_1),x_s(t_2)), \qquad (31)$$

associated with that edge will be 0. The necessary condition to avoid a common edge between the nodes from $\mathcal{X}_s$ and $\mathcal{X}_{s+1}$, where $s\leq\lfloor F/2\rfloor-1$, is that $L/\binom{F}{s}\leq F-s$. In other words, the edge weight should be at most 1. In general, the error performance of a code depends on the whole distance spectrum, which may be very difficult to control. We therefore turn to the minimal expected Hamming distance as a heuristic, not optimal, indicator related to the the code performance.

A representative example of choice of 12 multisymbols for $F=4$ is given on Fig. 4(a). We observe that no two multisymbols are identical and the choice of the multisymbols is capacity achieving. The multisymbols are constructed by using each edge of the graph exactly once, except for the edges between $\mathcal{X}_0=\{0000\}$ and $\mathcal{X}_1=\{0001,0010,0100,1000\}$, where common edges can not be avoided. Additionally, common edges are avoided later in the graph, by an adequate choice of the paths associated with the multisymbols. For example, we choose $t_2=\{0000,0001,0101,1101,1111\}$ instead of $t_2=\{0000,0001,0101,0111,1111\}$ in order to avoid a common edge with $t_1=\{0000,0001,0011,0111,1111\}$ in the last section of the graph. The minimal expected Hamming distance for this choice of multisymbols is 1.

C. An Example of Trellis Code Design and Performance Results

The coding scheme we propose is designed as a concatenation of an outer error correcting code, an interleaver and an encoder, as given in Fig. 5 (a). The outer error correcting code is a convolutional code with rate $1/2$, thus $2n$ binary symbols are generated from n symbols. The inner code is trellis based, as discussed. We associate multisymbols with the transitions in the trellis such that the output symbols on the branches exiting from the same state are maximally separated in terms of expected Hamming distance. The same is valid for the output symbols associated with the two branches that enter the same state. The trellis encoder codes incoming binary symbols into multisymbols which are then impaired by the channel. For this illustrative code design and performance evaluation, we assume a binary erasure channel.

The symbol errors from a trellis code come in bursts since ending up in a wrong state implies more than one symbol error. To avoid bursts of errors, an interleaver is used. The interleaver is implemented as matrix with dimensions $\lambda\times\frac{2n}{\lambda}$ with $2n$ divisible by λ. We consider trellis-based coding scheme for $F=4$, which defines a trellis with 12 branches. One option is to consider a code with 4 states and 3 branches from each state or a code, which implies that the source information is originally encoded in ternary symbols. Another, more practical option, is to have a trellis with 6 states and 2 branches from each state. The code construction uses the trellis code with 6 states to avoid mapping from binary symbols to ternary symbols. This means that one binary symbol is transmitted for each multisymbol. The trellis design for the set of multisymbols introduced in Section IV is shown in Fig. 5 (b). The multisymbols which are associated with the transitions in the trellis are chosen such that the output symbols on the branches exiting from the same state are maximally separated in terms of the expected Hamming distance, which yields a distance of 2.5 between the output multisymbols for all states. This is the maximal distance in the distance spectrum. The same is valid for the output symbols associated with the two branches that enter the same state.

The simulations are performed with $\lambda=8$ and for packet lengths $N=6,14,30,62$. These packet lengths are chosen

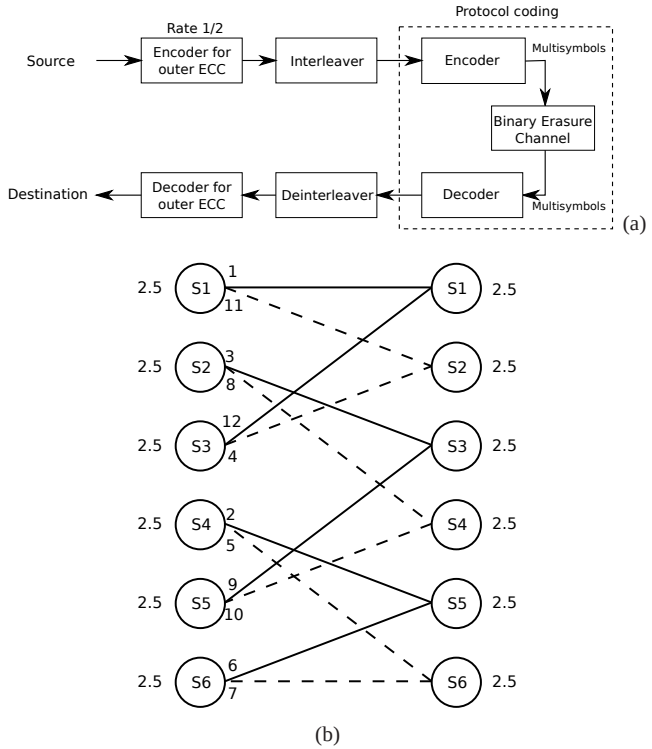


Fig. 5. Code Design. a) Block diagram of the code (b) Trellis construction for the set of multisymbols.

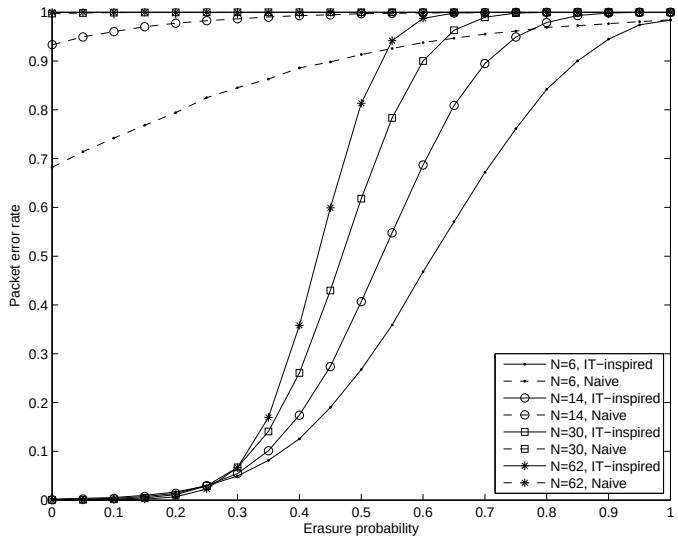


Fig. 6. Performance comparison between the naïve coding scheme and the scheme motivated from the multiuser framework

such that $N + 2$ (two tail bits are added by the outer convolutional code) is divisible by $\lambda = 8$. The results are averaged over 10000 iterations. We compare the performance of the coding scheme inspired by the multisymbol framework and the naïve coding scheme, which does not account for the specifics of the secondary channel. The simulation results for the packet error rate (PER) for different erasure probability are shown in Fig. 6. The results present a clear evidence that the information-theoretic analysis carries a practical significance for the secondary communications channels.

V. Discussion

In the model used for energy harvesting, we have forced the transmitter to send all the harvested energy quanta during the next frame. This creates a situation in which the transmitter sends e. g. zero-rate symbols (all ones). A way to amend this situation is to assume that the harvesting buffer has a size of $B = \frac{F}{2}$. In that case the states with $s > \frac{F}{2}$ are reached with probability 0 and the same approach presented in the paper can be used to devise communication strategies.

We have already described a generic application of secondary channels: communication with newly introduced devices, with limited functionality, in an area that is larger than the original coverage area. The secondary rate is low, even when protocol coding operates close to capacity, so it is hard to use it for rate advantage. Furthermore, the secondary rate depends on the current load (traffic, number of users) in the primary system. For example, if protocol coding is done by allocating the users to the channels in a cellular system, the best secondary rate is obtained when each channel can be allocated to a different user, since this maximizes the possible number of reorderings. Finally, the new secondary devices can have low-complexity, limited implementation of the primary protocol stack. In the extreme case, secondary data is encoded with presence/absence of packet, such that a secondary device needs to use only power detection.

Header compression [25] may appear as a competitor as it works in a somewhat opposite way: tries to compress the overhead whenever the actual communication scenario allows it. However, this is not necessarily canceling the secondary channel: e. g, the MAC–layer identifiers may be compressed, but still all the users have to be differentiated and the secondary channel arises from reordering their identifiers. Interestingly, the secondary capacity can be used to assess the performance margin of a certain primary protocol/system. Intuitively, if in a given scenario the secondary capacity is non–zero, then the operation of the primary system is not optimal.

Secondary channels can be used to send low-rate control data. For example, secondary data can be regarded as *expanded "future use" bits*: in many standards there are unspecified, free bits for future use and protocol coding practically unleashes "hidden" future use bits in the protocol, which may be indispensable as the system evolves. Another usage can be signaling in cognitive radio, where the cognitive (secondary) users may cause interference to the incumbent (primary) user. Protocol coding inherently introduces a possibility to provide in–band information about spectrum availability, e. g. through a Cognitive Pilot Channel (CPC) [26]. For example, if the primary system is a digital TV broadcaster, then secondary channel can be defined by reordering of the TV packets, which empowers the TV broadcast tower to dynamically inform about spectrum availability. Finally, in the emerging machine–to–machine (M2M) communication [3], cellular networks embrace a large number of low–cost, low–power devices, that have different traffic/behavior from the usual cellular users. Such a device device is mostly in a low–power "sleep" mode an it may be tuned receive on the secondary channel. Upon receiving a downlink trigger from the BS, it can wake up

another radio interface to send information. Thus, protocol coding offers an opportunity to introduce universal wake–up beacons.

A. Protocol Coding in WiMAX: A Brief Case Study

Although we have considered only reordering of binary resources, it is of interest to see how much secondary capacity can be offered in a practical system. In WiMAX [1], the downlink and uplink control information is transmitted at the beginning of each frame, which includes preamble, frame control header (FCH) and MAP message. The MAP message indicates the resource allocation for downlink and uplink data and control signal transmission. The Base Station (BS) translates the QoS requirements of the Subscriber Stations (SSs) into the appropriate number of allocated slots. The BS informs about the scheduling to all SSs by using the DL_MAP (Downlink Medium Access Protocol) and UL_MAP (Uplink Medium Access Protocol) messages in the beginning of each frame [27]. Protocol coding can be implemented by reordering the slots allocated in a frame. The secondary users for which this information is intended have only to read the broadcast DL_MAP and UL_MAP messages. For example, when the number of slots reserved for each of the SSs is 6,9,2,10,7,6,10,15,15,20 respectively, 289 secondary bits can be sent by reordering of the resources. Assuming a frame duration of $5ms$, this translates to we can have ≈ 58 [kbps] of additional information, which is in the frame headers that are robustly protected [28]. In order to get an idea about the the distance where the MAP message is "detectable", compared to the information data, we resort to the propagation model in [28], with the total path loss is given by $L = 126.2 + 36 \log d$ [dB], where d is in kilometers. The MAP is protected with $6-$times repetition coding, while and BPSK is used for both MAP message and data, which results in distance $d' \approx 1.65\, d$ where the header is detectable compared to the distance d for the user data.

VI. Conclusions and Future Work

We have introduced a class of communication channels with reordering of resources that are applicable in two different scenarios: (1) creation of a secondary channel over an existing primary system and (2) energy harvesting systems. The first scenario corresponds to the concept of protocol coding, where information is modulated in the actions taken by the communication protocol of an existing, primary system. Communication schemes with reordering of resources have been introduced before, but the key feature of the communication model is that it works under constraints that are put by the primary system or the energy harvesting process. We have shown how to compute the capacity of those channels when the resources that can be reordered have binary values. The capacity result is valid under arbitrary error model in which errors in each resource (packet) occur independently. The insights obtained from the capacity–achieving communication strategies have been used to demonstrate a design of practical error–correcting codes suited for the considered communication model.

It may be argued that the model with only two primary is limiting, but extension to K primary addresses entails complexity that is outside the scope of this initial paper on the topic. Yet, the results with binary secondary inputs provide novel insights for the communication strategies and set the basis for generalizations to $K > 2$. Another question for future work is how to compute the capacity and which coding strategies to use when the scheduling process in the primary system is generalized (buffering, retransmission, etc.).

Acknowledgment

The authors would like to thank Prof. Osvaldo Simeone (New Jersey Institute of Technology) for useful discussions on the channels with causal channel state information at the transmitter. The authors would also like to thank the Editor Ashish Khisti for providing very valuable suggestions that have improved the manuscript.

Appendix A
Proof of Lemma 1

Proof: We generalize the Theorem 4.5.1 from [29] to reflect the fact that the maximization is over $\mathcal{P}_{\mathbf{X},S}$ rather than $\mathcal{P}_{\mathbf{X}}$. Let us denote $P_{\mathbf{X}}(\mathbf{x}_{s,k}) = \alpha_{s,k}$ where $\mathbf{x}_{s,k}$ is the $k-$th element (e. g. in a lexicographic order) within the set $\mathcal{X}_s$. Let $\boldsymbol{\alpha} = (\alpha_{0,1}, \alpha_{1,1}, \alpha_{1,2}, \ldots, \alpha_{F,F})$ be the 2^F-dimensional probability vector. Then $I(\mathbf{X};\mathbf{Y}) = f(\alpha)$ and the maximization problem is:

$$\max f(\boldsymbol{\alpha}) \qquad \text{such that} \quad \sum_{k=1}^{K_s} \alpha_{s,k} = p_s, \quad \forall s \in \mathcal{S} \tag{32}$$

where $p_s = P_S(s)$ and $K_s = |\mathcal{X}_s| = \binom{F}{s}$. The constraint $\sum_{s,k} \alpha_{s,k} = 1$ is redundant, since $\sum_s p_s = 1$. We need to use $(F+1)$ Lagrangian multipliers and maximize $f(\boldsymbol{\alpha}) - \sum_s \lambda_s (\sum_k \alpha_{s,k} - p_s)$. The necessary and sufficient KKT conditions for each s,k are given as $\frac{\partial f}{\partial \alpha_{s,k}} = \lambda_s$ when $\alpha_{s,k} > 0$ and $\frac{\partial f}{\partial \alpha_{s,k}} \leq \lambda_s$ when $\alpha_{s,k} = 0$. We have:

$$\frac{\partial f}{\partial \alpha_{s,k}} = I(\mathbf{X}_{s,k} = \mathbf{x}_{s,k}; \mathbf{Y}) - \log e \tag{33}$$

where we have defined:

$$I(\mathbf{X} = \mathbf{x}_{s,k}; \mathbf{Y}) = \sum_{\mathbf{y} \in \mathcal{Y}} p(\mathbf{y}|\mathbf{x}_{s,k}) \log \frac{p(\mathbf{y}|\mathbf{x}_{s,k})}{\sum_{s,k} \alpha_{s,k} p(\mathbf{y}|\mathbf{x}_{s,k})} \tag{34}$$

The necessary and sufficient conditions for an input probability vector $\boldsymbol{\alpha} \in \mathcal{P}_{\mathcal{X},\mathcal{S}}$ to maximize this mutual information are then stated as follows. For a set of numbers $\{C_s\}$, where $C_s = \lambda_s + \log e$ and $s \in \mathcal{S}$: If $\alpha_{s,k} > 0$ then $I(\mathbf{X} = \mathbf{x}_{s,k}; \mathbf{Y}) = C_s$; otherwise, if $\alpha_{s,k} = 0$ then $I(\mathbf{X} = \mathbf{x}_{s,k}; \mathbf{Y}) \leq C_s$. Let $\mathcal{Y}_A$ be the set of all $\mathbf{y}$ whose elements are permutations of a certain $\mathbf{y}_A$. The $K_s \times |\mathcal{Y}_A|$ sub–matrix that contains $p(\mathbf{y}|\mathbf{x}_{s,k})$ which correspond to the inputs from the state $S = s$ and the outputs from the subset $\mathcal{Y}_A$ exhibits a symmetry: each row of this sub–matrix is a permutation of each other row and each column is a permutation of each other column. Using the definition of a symmetric channel from [29] and setting all the inputs $\mathbf{x} \in \mathcal{X}_s$ equiprobable with $\alpha_{s,k} = \frac{p_s}{K_s}$. Then the output probabilities $p(\mathbf{y}) = \sum_s \frac{p_s}{K_s} \sum_k p(\mathbf{y}|\mathbf{x}_{s,k})$ are the same for all $\mathbf{y} \in \mathcal{Y}_A$. Further it follows that the $K_s \times |\mathcal{Y}_A|$ sub-matrix containing the elements $p(\mathbf{y}|\mathbf{x}_{s,k}) \log \frac{p(\mathbf{y}|\mathbf{x}_{s,k})}{\sum_{s,k} \alpha_{s,k} p(\mathbf{y}|\mathbf{x}_{s,k})}$

has the same permutation properties as $p(\mathbf{y}|\mathbf{x}_{s,k})$, and hence the sum of these terms in (34) is the same for all $\mathbf{x} \in \mathcal{X}_s$. ■

APPENDIX B

Lemma 2: The conditional entropy for $\mathbf{x}_s \in \mathcal{X}_s$, having a Hamming weight of s, is given by:

$$H(\mathbf{Y}|\mathbf{X}=\mathbf{x}_s) = sH(\mathbf{q}_1) + (F-s)H(\mathbf{q}_0) = H_s \quad (35)$$

where $H(\mathbf{q}_i) = -\sum_{j=1}^{J} q_{ij}\log_2 q_{ij}$ for $i = 0, 1$ and $\mathbf{q}_i$ is given by (2).

Proof: In order to determine $H(\mathbf{Y}|\mathbf{X} = \mathbf{x}) = -\sum_{\mathbf{y}\in\mathcal{J}^F} P(\mathbf{y}|\mathbf{x})\log_2 P(\mathbf{y}|\mathbf{x})$, we use the fact that $P(\mathbf{y}|\mathbf{x}) = \prod_{f=1}^{F} q_{x_f y_f}$ is a product distribution, such that we can write $H(\mathbf{Y}|\mathbf{X}=\mathbf{x})$ as:

$$\begin{aligned} &-\sum_{\mathbf{y}\in\mathcal{J}^F}\prod_{i=1}^{F} q_{x_i y_i}\sum_{j=1}^{F}\log_2 q_{x_j y_j} \\ &\qquad = -\sum_{j=1}^{F}\sum_{y_1\in\mathcal{J}}\cdots\sum_{y_F\in\mathcal{J}}\log_2 q_{x_j y_j}\prod_{i=1}^{F} q_{x_i y_i} \end{aligned}$$

where (a) follows from changing the order of summation. If we consider the component $j = 1$:

$$\begin{aligned} &-\sum_{y_1\in\mathcal{J}}\cdots\sum_{y_F\in\mathcal{J}}\log_2 q_{x_1 y_1}\prod_{i=2}^{F} q_{x_i y_i} \\ &\qquad = -\sum_{y_1\in\mathcal{J}} q_{x_1 y_1}\log_2 q_{x_1 y_1}\sum_{y_2\in\mathcal{J}}\cdots\sum_{y_F\in\mathcal{J}}\prod_{i=2}^{F} q_{x_i y_i} \\ &\qquad \overset{(b)}{=} -\sum_{y_1\in\mathcal{J}}\log_2 q_{x_1 y_1}\cdot q_{x_1 y_1} = H(\mathbf{q}_{x_1}) \end{aligned} \quad (36)$$

where (b) follows from $\sum_{y_2\in\mathcal{J}}\cdots\sum_{y_F\in\mathcal{J}}\prod_{i=2}^{F} q_{x_i y_i} = 1$. Doing the same for $j = 2\ldots F$ shows that each $x_j = i$, $i = 0, 1$, contributes $H(\mathbf{q}_i)$ to $H(\mathbf{Y}|\mathbf{X}=\mathbf{x})$, which proves the lemma. ■

APPENDIX C

Lemma 3: The entropy $H(\mathbf{Y}|\mathcal{M}_t)$ is minimized when $\mathcal{M}_t$ is an arbitrary minimal multisymbol.

Proof: We first consider a special type of $P_S(\cdot)$, in which only two states $s_1, s_2 \in \mathcal{S}$ occur with non-zero probability $P_S(s_1) = \lambda$ and $P_S(s_2) = 1-\lambda$, such that $\mathcal{M}_t = \{\mathbf{x}_{s_1}, \mathbf{x}_{s_2}\}$. Due to the symmetry implied by Lemma 2, without losing generality, we first pick an arbitrary $\mathbf{x}_{s_1} \in \mathcal{X}_{s_1}$. The question is how to pick $\mathbf{x}_{s_2} \in \mathcal{X}_{s_2}$ in order to minimize $H(\mathbf{Y}|\mathcal{M}_t)$. Recall that $w_H(\mathbf{x}) = s$ for $\mathbf{x} \in \mathcal{X}_s$ and, without losing generality, assume $s_2 > s_1$. Let $g_{uv}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$, where $u, v \in \{0, 1\}$ denote the number of positions f at which $x_{s_1 f} = u$ and $x_{s_2 f} = v$. Using similar arithmetics as in Lemma 2:

$$\begin{aligned} H(\mathbf{Y}|\mathcal{M}_t) = {} & g_{00}H(\mathbf{q}_0) + g_{11}H(\mathbf{q}_1) + g_{01}H(\lambda\mathbf{q}_0 + (1\text{-}\lambda)\mathbf{q}_1) \\ & + g_{10}H((1\text{-}\lambda)\mathbf{q}_0 + \lambda\mathbf{q}_1) \end{aligned} \quad (37)$$

The Hamming distance is $d_H(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) = g_{01} + g_{10}$. Since $w_H(\mathbf{x}_{s_1}) < w_H(\mathbf{x}_{s_2})$, it follows that $g_{10}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) < g_{01}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$. Assume that $g_{10}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) > 0$ and let there be f_1, f_2 such that:

$$(x_{s_1,f_1}, x_{s_2,f_1}) = (1, 0) \qquad (x_{s_1,f_2}, x_{s_2,f_2}) = (0, 1) \quad (38)$$

Let $\mathbf{z}_{s_2}$ be another representative from $\mathcal{X}_{s_2}$, obtained by swapping the positions f_1, f_2 in $\mathbf{x}_{s_2}$, but keeping the other values of $\mathbf{x}_{s_2}$, such that $z_{s_2,f_1} = 1$ and $z_{s_2,f_2} = 0$. Then:

$$\begin{aligned} g_{00}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + 1 &= g_{00}(\mathbf{z}_{s_1}, \mathbf{z}_{s_2}) \\ g_{11}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) + 1 &= g_{11}(\mathbf{z}_{s_1}, \mathbf{z}_{s_2}) \\ g_{01}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) - 1 &= g_{01}(\mathbf{z}_{s_1}, \mathbf{z}_{s_2}) \\ g_{10}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2}) - 1 &= g_{10}(\mathbf{z}_{s_1}, \mathbf{z}_{s_2}) \end{aligned} \quad (39)$$

Using the concavity of the entropy function, we can write:

$$\begin{aligned} &H(\lambda\mathbf{q}_0 + (1\text{-}\lambda)\mathbf{q}_1) + H((1\text{-}\lambda)\mathbf{q}_0 + \lambda\mathbf{q}_1) \\ &\quad \geq \lambda H(\mathbf{q}_0) + (1\text{-}\lambda)H(\mathbf{q}_1) + (1\text{-}\lambda)H(\mathbf{q}_0) + \lambda H(\mathbf{q}_1) \\ &\quad = H(\mathbf{q}_0) + H(\mathbf{q}_1) \end{aligned} \quad (40)$$

Using (39) and (40) it follows:

$$\begin{aligned} H_{\mathbf{x}_{s_1},\mathbf{x}_{s_2}} &= g_{00}H(\mathbf{q}_0) + g_{11}H(\mathbf{q}_1) + g_{01}H(\lambda\mathbf{q}_0 + (1-\lambda)\mathbf{q}_1) \\ &\quad + g_{10}H((1-\lambda)\mathbf{q}_0 + \lambda\mathbf{q}_1) \\ &\geq g_{00}H(\mathbf{q}_0) + g_{11}H(\mathbf{q}_1) \\ &\quad + (g_{01}-1)H(\lambda\mathbf{q}_0 + (1-\lambda)\mathbf{q}_1) \\ &\quad + (g_{10}-1)H((1-\lambda)\mathbf{q}_0 + \lambda\mathbf{q}_1) = H_{\mathbf{x}_{s_1},\mathbf{z}_{s_2}} \end{aligned}$$

where $g_{uv} = g_{uv}(\mathbf{x}_{s_1}, \mathbf{x}_{s_2})$ and $H_{\mathbf{x}_{s_1},\mathbf{x}_{s_2}} = H(\mathbf{Y}|\mathcal{M}_t = \{\mathbf{x}_{s_1}, \mathbf{x}_{s_2}\})$. We can analogously continue the swap the positions in $\mathbf{x}_{s_2}$ until getting $g_{10} = 0$. Each swap does not increase $H(\mathbf{Y}|\mathcal{M}_t)$, which means that when $g_{10} = 0$, $H(\mathbf{Y}|\mathcal{M}_t)$ is minimal.

We now consider a general $P_S(\cdot)$. As indicated above, $H(\mathbf{Y}|\mathcal{M}_t)$ can be written as:

$$H(\mathbf{Y}|\mathcal{M}_t) = \sum_{f=1}^{F} H(\mathbf{u}_f) \quad (41)$$

where $\mathbf{u}_f$ is the probability distribution that corresponds to the $f-$th position, defined as:

$$\mathbf{u}_f = \sum_{s=0}^{F} P_s\left[(1-x_{s,f})\mathbf{q}_0 + x_{s,f}\mathbf{q}_1\right] \quad (42)$$

where $x_{s,f} \in \{0, 1\}$. Without losing generality, let us take the first value x_{s1} of each of the representatives $\mathbf{x}_s$ can create $(F+1)-$dimensional vector $\mathbf{z}_1$. In a similar way $\mathbf{z}_2$ is created, such that:

$$\mathbf{z}_1 = (x_{01}, x_{11}, \cdots x_{F1}) \qquad \mathbf{z}_2 = (x_{02}, x_{12}, \cdots x_{F2}) \quad (43)$$

The probability distribution vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ can be written as:

$$\begin{aligned} \mathbf{u}_1 &= (Q_{00} + Q_{01})\mathbf{q}_0 + (Q_{10} + Q_{11})\mathbf{q}_1 \\ \mathbf{u}_2 &= (Q_{00} + Q_{10})\mathbf{q}_0 + (Q_{01} + Q_{11})\mathbf{q}_1 \end{aligned} \quad (44)$$

where $Q_{uv} = \sum_{s\in\mathcal{G}_{uv}(z_1,z_2)} P_s$ and the sets $\mathcal{G}_{uv}(z_1, z_2) = \{s|x_{s,1} = u, x_{s,2} = v\}$ for $u, v \in \{0, 1\}$.

We now show that the contribution of the positions 1 and 2 to the entropy $H(\mathbf{Y}|\mathcal{M}_t)$ is minimized when one of the sets $\mathcal{G}_{01}, \mathcal{G}_{10}$ is empty. Let us start with a multisymbol $\{\mathbf{x}_s\}$ in which none of the sets $\mathcal{G}_{01}(z_1, z_2), \mathcal{G}_{10}(z_1, z_2)$ is empty. Without losing generality, we will "empty" the set $\mathcal{G}_{01}(z_1, z_2)$ as follows: If there is $s \in \mathcal{S}$ such that $x_{s,1} = 0, x_{s,2} = 1$, these two positions in the representative $\mathbf{x}_s$ are swapped. That is, if there is a representative $\mathbf{x} = 01\cdots$, it is changed to

$10\cdots$. Using the concavity of the entropy, we can show that these swapping operations can decrease the contribution of the positions $f = 1, 2$ to the entropy (41). Note that after swapping (44), the new distributions are:

$$\begin{aligned} \mathbf{u}_1' &= Q_{00}\mathbf{q}_0 + (Q_{01} + Q_{10} + Q_{11})\mathbf{q}_1 \\ \mathbf{u}_2' &= (Q_{00} + Q_{01} + Q_{10})\mathbf{q}_0 + Q_{11}\mathbf{q}_1 \end{aligned} \tag{45}$$

Using the concavity property, it can be shown that (see Lemma 4 in Appendix D)

$$H(\mathbf{u}_1) + H(\mathbf{u}_2) \geq H(\mathbf{u}_1') + H(\mathbf{u}_2') \tag{46}$$

where $\mathbf{u}_1, \mathbf{u}_2$ and $\mathbf{u}_1', \mathbf{u}_2'$ are given by (44) and (45), respectively. Analogously, the contribution from the two positions will decrease to the value (46) if the set $\mathcal{G}_{10}(z_1, z_2)$ is emptied. ■

APPENDIX D

Lemma 4: Let $\mathbf{q}_0$ and $\mathbf{q}_1$ be vectors of equal dimensions, each representing a probability distributions. Let $\mathbf{Q} = \{Q_1, Q_2, Q_3, Q_4\}$ be a probability distribution. Then the following holds:

$$\begin{aligned} &H\big((Q_1 + Q_2)\mathbf{q}_0 + (Q_3 + Q_4)\mathbf{q}_1\big) \\ &\quad + H\big((Q_1 + Q_3)\mathbf{q}_0 + (Q_2 + Q_4)\mathbf{q}_1\big) \\ &\geq H\big(Q_1\mathbf{q}_0 + (Q_2 + Q_3 + Q_4)\mathbf{q}_1\big) \\ &\quad + H\big((Q_1 + Q_2 + Q_3)\mathbf{q}_0 + Q_4\mathbf{q}_1\big) \end{aligned} \tag{47}$$

Proof: The members on the left-handed side of (47) can be written as:

$$\begin{aligned} H\big((Q_1 + Q_2)\mathbf{q}_0 + (Q_3 + Q_4)\mathbf{q}_1\big) &= H\big(\lambda\mathbf{v}_1 + (1\text{-}\lambda)\mathbf{v}_2\big) \\ H\big((Q_1 + Q_3)\mathbf{q}_0 + (Q_2 + Q_4)\mathbf{q}_1\big) &= H\big((1\text{-}\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2\big) \end{aligned}$$

where $\mathbf{v}_1 = Q_1\mathbf{q}_0 + (Q_2 + Q_3 + Q_4)\mathbf{q}_1$, $\mathbf{v}_2 = (Q_1 + Q_2 + Q_3)\mathbf{q}_0 + Q_4\mathbf{q}_1$, and $\lambda = \frac{Q_3}{Q_2 + Q_3}$. Since $H(\cdot)$ is concave, we finalize the proof by writing:

$$\begin{aligned} &H\big(\lambda\mathbf{v}_1 + (1\text{-}\lambda)\mathbf{v}_2\big) + H\big((1\text{-}\lambda)\mathbf{v}_1 + \lambda\mathbf{v}_2\big) \\ &\quad \geq \lambda H(\mathbf{v}_1) + (1\text{-}\lambda)H(\mathbf{v}_2) + (1\text{-}\lambda)H(\mathbf{v}_1) + \lambda H(\mathbf{v}_2) \\ &\quad = H(\mathbf{v}_1) + H(\mathbf{v}_2) \end{aligned}$$

■

REFERENCES

[1] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX*. Prentice-Hall, 2007.

[2] 3GPP, "LTE–Advanced," [Online]. Available: http://www.3gpp.org/article/lte-advanced.

[3] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 66–74, Apr. 2011.

[4] J. L. Massey, *Channel Models for Random–Access Systems*, ser. Performance Limits in Communication Theory and Practice, NATO Advances Studies Institutes Series E142. Kluwer Academic, 1988, pp. 391–402.

[5] V. Anantharam and S. Verdu, "Bits through Queues," *IEEE Trans. Inf. Theory*, vol. 44, pp. 4–18, Jan. 1996.

[6] G. Kramer, "Models and theory for relay channels with receive constraints," in *Proc. 42 Annual Allerton Conf. Commun., Control Comput.*, Urbana-Champaign, IL, USA, Sept. 2004.

[7] T. Lutz, C. Hausl, and R. Kötter, "Bits Through Relay Cascades with Half–Duplex Constraint," 2009, submitted, (arXiv:0906.1599).

[8] P. Popovski and O. Simeone, "Protocol coding for two-way communications with half-duplex constraints," in *IEEE GLOBECOM*, Miami, FL, USA, Dec. 2010.

[9] A. Ephremides and B. Hajek, "Information theory and communication networks: An unconsummated union," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2416–2434, Oct. 1998.

[10] K. Ashan, *Covert Channel Analysis and Data Hiding in TCP/IP*. M. Sc. thesis, Dept. of Electrical and Computer Engineering, University of Toronto, Aug. 2002.

[11] R. C. Chakinala, A. Kumarasubramanian, R. Manokaran, G. Noubir, C. P. Rangan, and R. Sundaram, "Steganographic communication in ordered channels," in *Inf. Hiding, Lecture Notes Comput. Science*, vol. 4437. Springer-Verlag, 2009.

[12] A. El-Atawy and E. Al-Shaer, "Building covert channels over the packet reordering phenomenon," in *Proc. IEEE INFOCOM*, Apr. 2009.

[13] A. J. H. Vinck, "Coded modulation for power line communications," *AEÜ J.*, pp. 45–49, Jan. 2000.

[14] W. Chu, C. J. Colbourn, and P. Dukes, "Constructions for permutation codes in powerline communications," *Designs, Codes Cryptography, Kluwer Academic Publishers*, vol. 32, pp. 51–64, 2004.

[15] J. Yang and S. Ulukus, "Transmission completion time minimization in an energy harvesting system," in *2010 44th Annual Conf. Inf. Sciences Syst. (CISS)*, Princeton, USA, Mar. 2010, pp. 1–6.

[16] K. Tutuncuoglu and A. Yener, "Short-term throughput maximization for battery limited energy harvesting nodes," in *2011 IEEE International Conf. Commun. (ICC)*, Xi'an, China, June 2011, pp. 1–5.

[17] V. Sharma, U. Mukherji, V. Joseph, and S. Gupta, "Optimal energy management policies for energy harvesting sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1326–1336, Apr. 2010.

[18] B. Devillers and D. Gunduz, "Energy harvesting communication system with battery constraint and leakage," in *2011 IEEE GLOBECOM Workshops*, Houston, USA, Dec. 2011, pp. 383–388.

[19] O. Ozel and S. Ulukus, "Information-theoretic analysis of an energy harvesting communication system," in *2010 IEEE 21st International Symp. Personal, Indoor Mobile Radio Commun. Workshops (PIMRC Workshops)*, Istanbul, Turkey, Sept. 2010, pp. 330–335.

[20] P. Popovski and Z. Utkovski, "On the secondary capacity of the communication protocols," in *IEEE GLOBECOM*, Honolulu, HI, USA, Dec. 2009.

[21] Z. Utkovski and P. Popovski, "Protocol coding with reordering of user resources: Capacity results for the z-channel," in *49th Annual Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sept. 2011.

[22] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Research Development*, vol. 2, pp. 289–293, Oct. 1958.

[23] G. Keshet, Y. Steinberg, and N. Merhav, *Channel coding in the presence of side information*, ser. Foundations Trends Commun. Inf. Theory, 2007, vol. 4, no. 6.

[24] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2nd Edition, 2006.

[25] H. Hannu, L.-E. Jonsson, R. Hakenberg, T. Koren, K. Le, Z. Liu, A. Martensson, A. Miyazaki, K. Svanbro, T. Wiebke, T. Yoshimura, and H. Zheng, "RObust header compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed," in *RFC 3095*, July 2001.

[26] J. Perez-Romero, O. Sallent, R. Agusti, and L. Giupponi, "A novel on-demand cognitive pilot channel enabling dynamic spectrum allocation," in *Proc. IEEE DySPAN*, Dublin, Ireland, Apr. 2007.

[27] A. Sayenko, O. Alanen, J. Karhula, and T. Hämäläinen, "Wimax overview and system performance," in *Proc. 9th ACM MSWiM*, Torremolinos, Spain, Oct. 2006.

[28] F. Wang, A. Ghosh, C. Sankaran, and P. Fleming, "Wimax overview and system performance," in *Proc. IEEE VTC Fall*, Sept. 2006.

[29] R. Gallager, *Information Theory and Reliable Communication*. John Wiley and Sons, 1968.

Petar Popovski (S'97, A'98, M'04, SM'10) received the Dipl.-Ing. in electrical engineering and Magister Ing. in communication engineering from Sts. Cyril and Methodius University, Skopje, Macedonia, in 1997 and 2000, respectively and Ph. D. from Aalborg University, Denmark, in 2004.

He is currently a Professor at Aalborg University, where he held faculty positions since 2004. From 2008 to 2009 he held part-time position as a wireless architect at Oticon A/S. He has more than 140 publications in journals, conference proceedings and books and has more than 25 patents and patent applications. He has received the Young Elite Researcher award and the SAPERE AUDE career grant from the Danish Council for Independent Research. He has received six best paper awards, including three from IEEE. His research interests are in the broad area of wireless communication and networking, information theory and protocol design.

Dr. Popovski serves on the editorial board of several journals, including IEEE Communications Letters (Senior Editor), IEEE JSAC Cognitive Radio Series, IEEE Transactions on Communications and IEEE Transactions on Wireless Communications.

Zoran Utkovski (M'10) received the Dipl.-Ing. in electrical engineering from Ss. Cyril and Methodius University in Skopje, R. Macedonia, in 2000, M.Sc. with distinction in communications engineering from Chalmers University of Technology, Sweden in 2004, and Dr.-Ing. with distinction in communications engineering from University of Ulm, Germany, in 2010.

He is currently an Assistant Professor at University Goce Delčev Štip, R. Macedonia, where he is head of the Signal Processing and Communication Systems group. From 2005 to 2010, he was with University of Ulm, Germany. Since 2011, he has been a member of the Laboratory for Complex Systems and Networks at the Macedonian Academy of Sciences and Arts. His main research interests are in the field of information and communication theory, with a focus on the geometric aspects of coding for non-coherent communication in wireless networks.

Dr. Utkovski is a co-author of the Best Poster Award paper at the IEEE Communication Theory Workshop in 2010.

Kasper F. Trillingsgaard received the B. Sc. in communication systems from Aalborg University, Denmark in 2011.

He is currently pursuing his M.Sc. degree in wireless communication at Aalborg University, Denmark. His research interests include wireless communications, information theory and coding theory.