On the Capacity of Generalized Ising Channels

Artyom Sharov Ron M. Roth

***Abstract*—Nearly tight lower and upper bounds on the capacity of generalized Ising channels are presented. For the case where feedback is allowed, a closed-form expression for the capacity is found for channel error probability $p \in [0, p_0]$, where $p_0 \approx 0.398324$. A near-capacity-achieving family of encoders is presented for the values $p \in [0, p_0]$. Two lower bounds on that capacity for larger values of p are presented, one of which is tight on the interval $[p_0, 0.5]$.**

***Index Terms*—granular media, dynamic programming, channel capacity, magnetic recording, feedback.**

I. Introduction

Let $\mathbb{Z}^+$ denote the set of positive integers. For a real parameter $p \in [0, 1]$, we define a *one-dimensional generalized Ising channel* $\mathtt{Is}(p)$ through the following relation between an input random process $\mathcal{X} = (X_t)_{t \in \mathbb{Z}^+}$ over $\Sigma = \{0, 1\}$ and an output random process $\mathcal{Y} = (Y_t)_{t \in \mathbb{Z}^+}$ over Σ, for any $t \in \mathbb{Z}^+$:

$$Y_t = X_{t-Z_t} , \tag{1}$$

where $\mathcal{Z} = (Z_t)_{t \in \mathbb{Z}^+}$ is an i.i.d. Bernoulli process characterized by p (in other words, $\mathrm{Prob}(Z_t = 1) = p$ independently of the other entries of $\mathcal{Z}$), and $X_0 \stackrel{\triangle}{=} 1$. Defining a state process $\mathcal{S} = (S_t)_{t \in \mathbb{Z}^+}$ by $S_t = X_{t-1}$ gives an alternative way of representing the channel $\mathtt{Is}(p)$, specifically,

$$Y_t = \begin{cases} X_t & \text{with probability } 1-p \\ S_t & \text{with probability } p \end{cases}, \text{ for any } t \in \mathbb{Z}^+.$$

Figure 1 demonstrates how this channel operates at any time t given the state S_t. To the best of our knowledge, this family of channels was first defined in [15, Sec. 4], where the first lower and upper bounds on its capacity were given.

One particular representative of this family of channels, $\mathtt{Is}(0.5)$, is known in the literature as the *one-dimensional Ising channel* [4], which bears its name due to its affinity to the Ising model in statistical mechanics [14][16]. Berger and Bonomi studied the information-theoretic properties of $\mathtt{Is}(0.5)$, specifically, they established the value 0.5 of the zero-error capacity [4, Th. 6] and gave the first numeric lower and upper bounds on the value of the (Shannon) channel capacity [4, Ths. 4 and 5], which was found to be strictly greater than 0.5. The zero-error capacity of $\mathtt{Is}(0.5)$, attained by a simple code construction, also appears in more recent papers in the context of *grain errors* under a combinatorial error model [19, Sec. 2], as well as in a probabilistic setting [15, Prop. 5]. Elishco and Permuter [10] introduced feedback to $\mathtt{Is}(0.5)$ and found its capacity to be approximately 0.575522 along with a simple zero-error coding scheme achieving that capacity [10, Ths. 1 and 2]. The setting of added feedback can be generalized to any channel $\mathtt{Is}(p)$ with $p \in [0, 1]$, thereby resulting in a new channel denoted hereafter by $\mathtt{Is}_{\mathtt{FB}}(p)$. Figure 2 shows schematically a block diagram of a transmission of a message M through $\mathtt{Is}_{\mathtt{FB}}(p)$. The message is decoded into $\widetilde{M}$ and one bit is fed back at a delay of one time slot. The idea of [10], as well as of several other earlier papers [2][20], is based on the observation that the capacity of certain channels with feedback can be recast as an *infinite-horizon dynamic program* [5, Ch. 8], whose optimal average reward equals the capacity of the channel. This technique was proposed by Tatikonda in his Ph.D. thesis [27], followed by his subsequent works [28][30][31]. The optimal average reward of the dynamic program in these works was found by solving the matching Bellman equation [1, Sec. 3][3].

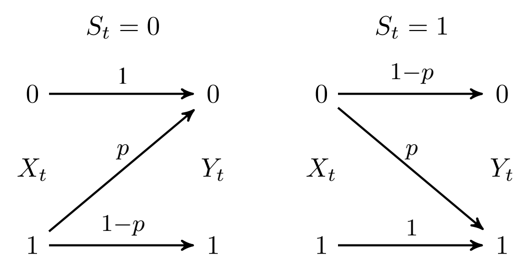

Fig. 1. The channel $\mathtt{Is}(p)$ as a function of the state S_t.

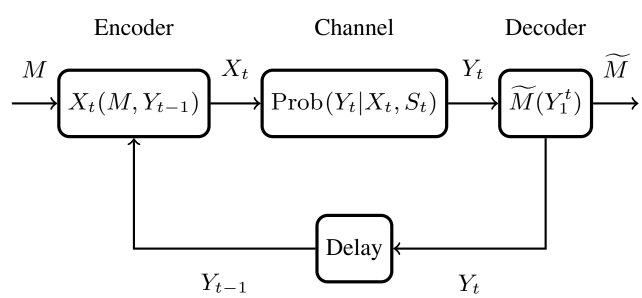

Fig. 2. The channel $\mathtt{Is}_{\mathtt{FB}}(p)$.

The study of $\mathtt{Is}(p)$ is motivated by the behavior of bit-patterned media during shingled writing [12][15, Sec. 3-B], where the magnetic field emanating from the write head may cause overlapping patterns of substitution errors, resembling those of $\mathtt{Is}(p)$. In our earlier work on granular media [25], we modeled these substitution errors as *overlapping grain errors*. The storage capacity of the medium can be increased

Artyom Sharov and Ron M. Roth are with the Computer Science Department, Technion, Haifa 3200003, Israel.

Emails: {sharov, ronny}@cs.technion.ac.il.

This work was supported in part by Grants Nos. 1092/12 and 1396/16 from the Israel Science Foundation. Part of this work was presented at the IEEE International Symposium on Information Theory, Hong Kong, June 2015

by supplementing the shingled magnetic recording with a hardware that detects during a write whether the previously-written bit along the track has been overrun by the current write (yet no backtracking is allowed to correct that bit); it can be shown that introducing such a capability is equivalent to adding feedback to $\mathtt{Is}(p)$, hence our interest in studying $\mathtt{Is}_{\mathsf{FB}}(p)$ as well.[1]

Let $\langle s, s' \rangle$ denote the set $\{s, s{+}1, s{+}2, \ldots, s'\}$ for any $s, s' \in \mathbb{Z}^+$ such that $s \le s'$. For an infinite vector $(A_t)_{t \in \mathbb{Z}^+}$ and $s, s' \in \mathbb{Z}^+$ such that $s \le s'$, let $A_s^{s'}$ denote the finite sub-vector $(A_t)_{t \in \langle s, s' \rangle}$. Denote by $\mathsf{cap}(p)$ the *capacity* of $\mathtt{Is}(p)$, namely,[2]

$$\mathsf{cap}(p) = \lim_{n \to \infty} \frac{1}{n} \sup_{\mathrm{Prob}(X_1^n)} I(X_1^n; Y_1^n) \tag{2}$$

[11, Secs. 4.6 and 5.9], where the processes $\mathcal{X} = (X_t)_{t \in \mathbb{Z}^+}$ and $\mathcal{Y} = (Y_t)_{t \in \mathbb{Z}^+}$ are related by (1), and where $I(X_1^n; Y_1^n)$ is the mutual information between X_1^n and Y_1^n. Massey [18], inspired by Marko [17], defined the (normalized) *directed information* of a channel given by the prefix of length n of the random input and output processes, $(X_t)_{t \in \mathbb{Z}^+}$ and $(Y_t)_{t \in \mathbb{Z}^+}$, respectively, as

$$\frac{1}{n} I(X_1^n \to Y_1^n) = \frac{1}{n} \sum_{i=1}^{n} I(X_1^i; Y_i \mid Y_1^{i-1}) ,$$

and then showed that taking the supremum over all distributions $\mathrm{Prob}(X_i \mid X_1^{i-1} Y_1^{i-1})$ and then letting n go to ∞ yields an expression which bounds from above the capacity of the underlying channel with feedback. In a series of subsequent works [21][22][28], this expression was also shown to bound the feedback capacity from below for a class of stationary channels, where the output is a function of the current and (a finite number of) past symbols of the input process $\mathcal{X}$ and of a predefined stationary ergodic error process $\mathcal{Z}$. Thus, Massey's normalized directed information was shown to be asymptotically equivalent to the capacity of the underlying channel with feedback, for that class of channels. As the channel $\mathtt{Is}_{\mathsf{FB}}(p)$ is of this ilk, one can define its capacity $\mathsf{cap}_{\mathsf{FB}}(p)$ in the aforementioned fashion, namely,[3]

$$\mathsf{cap}_{\mathsf{FB}}(p) = \lim_{n \to \infty} \frac{1}{n} \sup_{\mathrm{Prob}(X_i \mid X_1^{i-1} Y_1^{i-1})} \sum_{i=1}^{n} I(X_1^i; Y_i \mid Y_1^{i-1}) . \tag{3}$$

The rest of the paper is structured as follows. In Section II, we present improvements on the lower and upper bounds on the capacity $\mathsf{cap}(p)$ of $\mathtt{Is}(p)$. Next, in Section III, using the technique suggested by Tatikonda, we develop a closed-form expression for the capacity $\mathsf{cap}_{\mathsf{FB}}(p)$ of $\mathtt{Is}_{\mathsf{FB}}(p)$ for $p \in [0, p_0]$, where $p_0 \approx 0.398324$ is the unique solution of the equality $\frac{p}{1-p} = \frac{2^{\mathrm{H}(p)}}{2^{\mathrm{H}(p)}+1}$ on $[0,1]$, and $\mathrm{H}(p)$ is the binary entropy of p, for $p \in [0,1]$. Finally, in Section IV, we give three lower bounds on $\mathsf{cap}_{\mathsf{FB}}(p)$, for $p \in [0,1]$. One of these lower bounds is near-capacity-achieving on the interval $[0, p_0]$; another lower bound can be shown to be tight on the interval $[p_0, 0.5]$, employing the latest result by Sabag *et al.* [23].

[1] Specifically, we model the shingled write channel by $W_{t-1} = X_{t-1+Z_t}$, where Z_t is Bernoulli i.i.d. with $\mathrm{Prob}(Z_t = 1) = q$ and $(W_t)_t$ is the sequence of recorded bits; namely, the t-th bit X_t overwrites the previously recorded bit W_{t-1} with probability q. This channel is therefore equivalent to $\mathtt{Is}(1{-}q)$, where $Y_t = W_{t-1}$.

[2] In [11], the limit (2) was proven to exist.

[3] In [20, Th. 3], the limit (3) was proven to exist.

II. Bounds on $\mathsf{cap}(p)$

In the following two theorems, the proofs of which are deferred to Appendix A, we establish improved lower and upper bounds on $\mathsf{cap}(p)$.

Theorem 2.1: Let $p \in [0,1]$ and let $\mu, \zeta \in \mathbb{Z}^+$. Then

$$\mathsf{cap}(p) \ge \varrho(\mu, \zeta, p) \triangleq \sup_{P_\mu \in \mathrm{Mar}_\mu,\ \mathcal{X} \sim P_\mu} \Big\{ H(Y_{\zeta+1} \mid Y_2^\zeta X_1) - \mathrm{H}(p) \cdot \mathrm{Prob}(X_2 \ne X_1) \Big\} , \tag{4}$$

where Mar_μ stands for the set of all the stationary Markov chains of order μ over Σ, $\mathcal{X} \sim P_\mu$ denotes that $\mathcal{X}$ is distributed according to the Markov chain P_μ, and $H(\cdot \mid \cdot)$ stands for the conditional binary entropy function. □

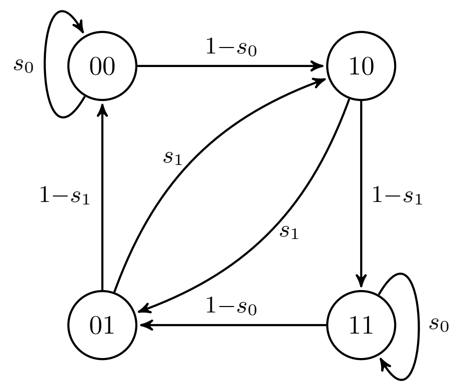

Fig. 3. Markov chain of order 2 from Example 2.3.

Theorem 2.2: Let $p \in [0,1]$ and let b_0, $\mathsf{b}_{1;1}$, b_1, $\mathsf{b}_{1;2}$, b_2, $\mathsf{b}_3 \in [0,1]$ be real numbers such that

$$\mathsf{b}_{1;1} \le \mathsf{b}_1, \mathsf{b}_{1;2} \le \mathsf{b}_2, \sum_{i \in \langle 1,3 \rangle} \mathsf{b}_i \le \mathsf{b}_0,\ 4\mathsf{b}_0 - \sum_{i \in \langle 1,3 \rangle} (4{-}i)\mathsf{b}_i \le 1 . \tag{5}$$

For each $i \in \langle 0, 2 \rangle$, let $\beta_i \in [0,1]$ be real numbers such that

$$\begin{aligned} \beta_0 &\triangleq \mathsf{b}_0 - 2p(1{-}p)\mathsf{b}_1, \\ \beta_1 &\triangleq p^2 \mathsf{b}_1 + p(1{-}p)\mathsf{b}_2 + (\mathsf{b}_1 {-} p \cdot \mathsf{b}_{1;1})(1{-}p)^2, \\ \beta_2 &\triangleq p^2 \mathsf{b}_2 {+} p(1{-}p)(\mathsf{b}_3 {+} \mathsf{b}_1 {-} p \cdot \mathsf{b}_{1;1}) {+} (1{-}p)^2(\mathsf{b}_2 {-} p \cdot \mathsf{b}_{1;2}). \end{aligned} \tag{6}$$

Also, let

$$\mathsf{t}(\beta_0, \beta_1, \beta_2) \triangleq \Big(1{-}2\beta_0{+}\beta_1\Big) \mathrm{H}\Big(\frac{\beta_0{-}\beta_1{-}\beta_2}{1{-}2\beta_0{+}\beta_1}\Big) + \beta_0 \mathrm{H}\left(\frac{\beta_1}{\beta_0}\right) + (\beta_0{-}\beta_1) \mathrm{H}\left(\frac{\beta_2}{\beta_0{-}\beta_1}\right) . \tag{7}$$

Then

$$\mathsf{cap}(p) \le \rho(p) \triangleq \max_{(\mathsf{b}_0, \mathsf{b}_{1;1}, \mathsf{b}_1, \mathsf{b}_{1;2}, \mathsf{b}_2, \mathsf{b}_3) \in U} \Big\{ \mathsf{t}(\beta_0, \beta_1, \beta_2) - \mathsf{b}_0 \cdot \mathrm{H}(p) \Big\} , \tag{8}$$

where $U \subseteq [0,1]^6$ is the region given by the inequalities (5).□

Example 2.3: Let $(\mu, \zeta) = (2, 3)$. For $\boldsymbol{x} \in \Sigma^2$, let $s_{\boldsymbol{x}}$ and $\varsigma_{\boldsymbol{x}}$ denote the transition probability $\mathrm{Prob}(X_2 = 0 \mid$

$X_1X_0 = \boldsymbol{x}$) and the stationary probability $\mathrm{Prob}(X_1X_0 = \boldsymbol{x})$, respectively, of a Markov chain P_2. Due to the symmetry of $\mathtt{Is}(p)$ to bit complementation, we also take $\varsigma_{\boldsymbol{x}} = \varsigma_{\overline{\boldsymbol{x}}}$ and $s_{\boldsymbol{x}} = 1-s_{\overline{\boldsymbol{x}}}$ where $\overline{\boldsymbol{x}}$ stands for the bit-wise complement of $\boldsymbol{x}$. Define for brevity $\varsigma_{x_1} = \varsigma_{x_10}$ and $s_{x_1} = s_{x_10}$ for any $x_1 \in \Sigma$. Now we can characterize P_2 using a graph on four states as in Figure 3, where the states 00 and 11 have stationary probability $\varsigma_0 = \varsigma_{00} = \varsigma_{11}$ and the states 01 and 10 have stationary probability $\varsigma_1 = \varsigma_{10} = \varsigma_{01}$.

One can readily check that the expression for $\varrho(\mu, \zeta, p)$ from (4) for $(\mu, \zeta) = (2, 3)$ simplifies to solving the following maximization problem:

$$\varrho(\mu{=}2, \zeta{=}3, p) = \max_{s_0, s_1 \in [0,1]} \left\{ 2 \sum_{i \in \langle 0,3 \rangle} \ell_i \cdot \mathrm{H}\left(\frac{k_i}{\ell_i}\right) - 2\varsigma_1 \mathrm{H}(p) \right\},$$

where

$$\begin{aligned}
k_0 &= \mathrm{Prob}(Y_4Y_3Y_2X_1 = 0000) \\
&= \varsigma_0(s_0^2 + ps_0(1-s_0) + p(1-p)s_1(1-s_0)) \\
&\quad + \varsigma_1 p(1-p)s_1(1-s_1+p \cdot s_1), \\
k_1 &= \mathrm{Prob}(Y_4Y_3Y_2X_1 = 0100) \\
&= \varsigma_0(1-p)^2 s_1(1-s_0) \\
&\quad + \varsigma_1 p(ps_1(1-s_1) + p^2 s_1^2 + (1-p)(1-s_0)(1-s_1)), \\
k_2 &= \mathrm{Prob}(Y_4Y_3Y_2X_1 = 0010) = \varsigma_1(1-p)^2 s_1(1-s_1+p \cdot s_1), \\
k_3 &= \mathrm{Prob}(Y_4Y_3Y_2X_1 = 0110) \\
&= \varsigma_1(1-p)(p \cdot s_1(1-s_1) + p^2 s_1^2 + (1-p)(1-s_0)(1-s_1)), \\
\ell_0 &= \mathrm{Prob}(Y_3Y_2X_1 = 000) \\
&= \varsigma_0(s_0 + p(1-s_0)) + \varsigma_1 p(1-p)s_1, \\
\ell_1 &= \mathrm{Prob}(Y_3Y_2X_1 = 100) \\
&= \varsigma_0(1-p)(1-s_0) + \varsigma_1 p(1-s_1+p \cdot s_1), \\
\ell_2 &= \mathrm{Prob}(Y_3Y_2X_1 = 010) = \varsigma_1(1-p)^2 s_1, \\
\ell_3 &= \mathrm{Prob}(Y_3Y_2X_1 = 110) = \varsigma_1(1-p)(1-s_1+p \cdot s_1),
\end{aligned}$$

and

$$(\varsigma_0, \varsigma_1) = \left(\frac{0.5(1-s_1)}{2-s_0-s_1}, \frac{0.5(1-s_0)}{2-s_0-s_1}\right).$$

□

The proofs of Theorems 2.1 and 2.2, to be found in Appendix A, follow a standard technique of bounding the mutual information from below and from above. The lower bound of Theorem 2.1 arises after assuming a certain Markovian property on the distribution of $\mathcal{X}$, whereas the upper bound of Theorem 2.2 emerges as a result of employing the method of types both on $\mathcal{X}$ and $\mathcal{Y}$.

The quantities $\mathsf{b}_0, \mathsf{b}_{1;1}, \mathsf{b}_1, \mathsf{b}_{1;2}, \mathsf{b}_2, \mathsf{b}_3$ in Theorem 2.2 stand for numbers of runs[4] of various lengths in the typical words $\boldsymbol{x}$ of a capacity-achieving probability distribution of the input process X_1^n, normalized by the word length n. As a result of the transmission of any such $\boldsymbol{x}$, any typical output word will have asymptotically the same number β_0 of runs and the same number β_i of runs of length i, for each $i \in \langle 1, 2 \rangle$, per symbol. The quantity $\mathsf{t}(\beta_0, \beta_1, \beta_2)$ stands for the growth rate (with respect to n) of a set of words of length n with β_0 runs and β_i runs of length i, for each $i \in \langle 1, 2 \rangle$, per symbol.

[4]By a run we mean a consecutive subword $x_i\ x_{i+1}\ \ldots\ x_{i'}$ of $\boldsymbol{x} = (x_i)_{i \in \langle 1,n \rangle}$ such that $x_i = x_{i+1} = \ldots = x_{i'}$, where $x_{i-1} \neq x_i$ (if $i \geq 2$) and $x_{i'} \neq x_{i'+1}$ (if $i' \leq n-1$).

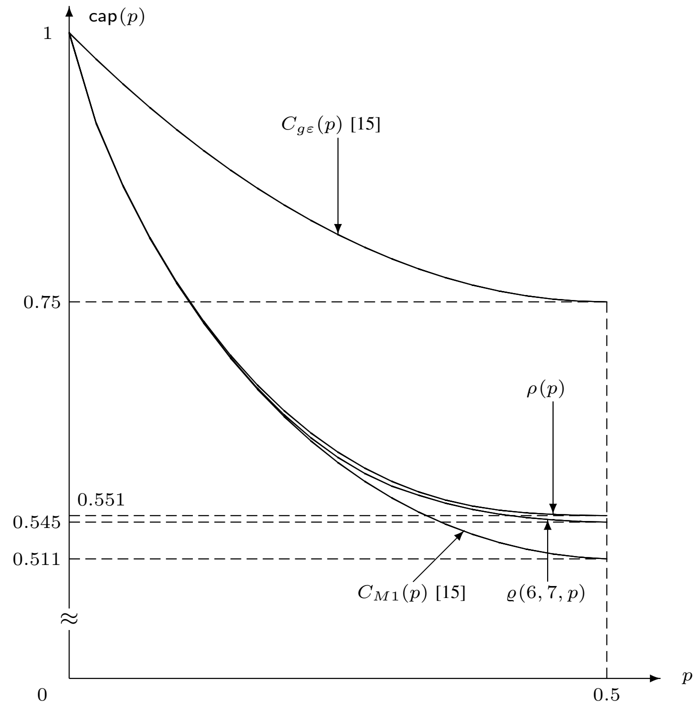

Fig. 4. Bounds on $\mathsf{cap}(p)$.

Figure 4 shows $\varrho(\mu{=}6, \zeta{=}7, p)$ and $\rho(p)$ alongside the best lower bound $C_{M1}(p)$ derived from [15, Eq. (12)] and the best upper bound $C_{g\varepsilon}(p) \triangleq 1-p(1-p)$ from [15, Eq. (3)] on $\mathsf{cap}(p)$ (the vertical axis in the figure is not to scale for values less than 0.511). Due to [15, Prop. 1], which shows that $\mathsf{cap}(p) = \mathsf{cap}(1-p)$, we can limit ourselves to the range $[0, 0.5]$ of p (as is done in Figure 4), since all the lower and upper bounds in Figure 4 on the range $[0.5, 1]$ are obtained by their reflection with respect to $p = 0.5$. One can observe the improvement of $\varrho(6, 7, p)$ and $\rho(p)$ over $C_{M1}(p)$ and $C_{g\varepsilon}(p)$, and the near-tightness of the new lower and upper bounds.

Remark 2.4: Using a generalization of the classical Blahut–Arimoto algorithm [29], one can obtain another set of lower bounds on $\mathsf{cap}(p)$; however, with comparable computational power and running time, the bounds that we obtained with this method were looser than $\varrho(\mu, \zeta, p)$. □

III. THE VALUE OF $\mathsf{cap}_{\mathsf{FB}}(p)$ FOR THE LOW RANGE OF p

The main result of this section is the following theorem whose proof will take up the rest of this section and Appendix B.

Theorem 3.1: Let $p_0 \in [0, 1]$ be the unique solution of the equality $\frac{p}{1-p} = \frac{2^{\mathrm{H}(p)}}{2^{\mathrm{H}(p)}+1}$. Then, for $p \in [0, p_0]$,

$$\mathsf{cap}_{\mathsf{FB}}(p) = \mathrm{H}\Big(\frac{1}{2^{\mathrm{H}(p)}+1}\Big) - \frac{\mathrm{H}(p)}{2^{\mathrm{H}(p)}+1}.$$

□

In Section III-A, we present a general formulation of an infinite-horizon average-reward dynamic program (DP), which is instantiated in Section III-B with quantities related to $\mathtt{Is}_{\mathsf{FB}}(p)$ to represent the problem of finding $\mathsf{cap}_{\mathsf{FB}}(p)$ as a DP. In Section III-C, we present the specialization of the

Bellman equation to our case by substituting the quantities from Section III-B into the general form of that equation. All three sections closely follow the presentation in [10, Sec. 4], albeit with several changes in notation.

A. DP formulation

A comprehensive survey on average-cost dynamic programming, *inter alia* containing the definitions of this section, is [1].

An *average-reward DP problem* describes a dynamic system evolving in (discrete) time $t \in \mathbb{Z}^+$ according to the septuple $\mathcal{T} = (\mathcal{B}, \mathcal{A}, \mathcal{D}, \iota, \upsilon, \mathsf{e}, \mathsf{r})$, where

- $\mathcal{B}$ is a Borel space containing all the system *states* b_t;
- $\mathcal{A}$ is a compact subset of a Borel space containing all the system *actions* a_t;
- $\mathcal{D}$ is a measurable space of all system *disturbances* d_t;
- ι is a probability measure of the initial state b_1;
- $\upsilon : \mathcal{D} \times \mathcal{B} \times \mathcal{A} \to [0, 1]$ is a conditional probability distribution of the disturbance d_t given b_t and a_t;
- $\mathsf{e} : \mathcal{B} \times \mathcal{A} \times \mathcal{D} \to \mathcal{B}$ is the function by which the system evolves from one state to another in the following fashion: $b_{t+1} = \mathsf{e}(b_t, a_t, d_t)$, for any $t \in \mathbb{Z}^+$;
- $\mathsf{r} : \mathcal{B} \times \mathcal{A} \to \mathbb{R}$ is a bounded real *reward* function.

The system transitions from one state b_t to another in discrete time $t \in \mathbb{Z}^+$ while undertaking actions a_t based on some prescribed policy and experiencing disturbances d_t, which cause random deviations from the course of undertaken action. The goal of the system is to maximize the average reward gained by it over time, namely, to maximize

$$\Psi = \Psi(\mathcal{T}) \triangleq \sup_{\boldsymbol{\pi}} \liminf_{n\to\infty} \tfrac{1}{n} \mathrm{E}_{\boldsymbol{\pi}} \left(\textstyle\sum_{t\in\langle 1,n\rangle} \mathsf{r}(b_t, a_t) \right) , \tag{9}$$

where $\boldsymbol{\pi}=(\pi_t)_{t\in\mathbb{Z}^+}$ is the system *policy*, given by an infinite vector of functions $\pi_t : \mathcal{B} \times \mathcal{D}^{t-1} \to \mathcal{A}$ mapping $(b_1, d_1, d_2, d_3, \ldots, d_{t-1})$ into a_t; in other words, π_t determines which action is undertaken by the system given the initial state and the entire history of disturbances prior to time t, for each $t \in \mathbb{Z}^+$.

B. Recasting the problem of finding $\mathsf{cap_{FB}}(p)$ as a DP

Given the channel $\mathtt{Is_{FB}}(p)$ with channel input process $\mathcal{X} = (X_t)_{t\in\mathbb{Z}^+}$, output process $\mathcal{Y} = (Y_t)_{t\in\mathbb{Z}^+}$, and state process $\mathcal{S} = (S_t)_{t\in\mathbb{Z}^+}$ taking values $(x_i)_i$, $(y_i)_i$, and $(s_i)_i$, respectively, we associate the following quantities with the components of the septuple $\mathcal{T}$ for all $t \in \mathbb{Z}^+$. Define the DP state at time t as the following vector of length 2

$$b_t = (b_t(s))_{s\in\Sigma} \triangleq (\mathrm{Prob}(S_t{=}s \mid Y_1^{t-1} = \boldsymbol{y}_1^{t-1}))_{s\in\Sigma} \ , \tag{10}$$

where $\boldsymbol{y}_1^{t-1} = (y_i)_{i=1}^{t-1} \in \Sigma^{t-1}$. The DP action at time t is given by the 2×2 stochastic matrix

$$\begin{aligned} a_t &= (a_t(s,x))_{s,x\in\Sigma} \triangleq \\ &\quad (\mathrm{Prob}(X_t = x \mid S_t = s, Y_1^{t-1} = \boldsymbol{y}_1^{t-1}))_{s,x\in\Sigma} \ ; \end{aligned} \tag{11}$$

the DP disturbance at time t is simply $d_t \triangleq y_t$. Set

$$\Xi(y, s, x) \triangleq \mathrm{Prob}(Y_t = y | S_t = s, X_t = x) \ , \tag{12}$$

and notice that this distribution is completely defined by the channel model, as it was presented in Figure 1. The initial distribution ι of b_1 is defined so that $\mathrm{Prob}(b_1 = 1) = 1$ and $\mathrm{Prob}(b_1 \in (0, 1)) = 0$; the disturbance distribution is given by

$$\upsilon(y \mid b_t, a_t) = \textstyle\sum_{s,s'\in\Sigma} b_t(s) a_t(s, s') \Xi(y, s, s') \ . \tag{13}$$

The evolution function e can be expressed recursively for $\widetilde{s} \in \Sigma$ and $t \in \mathbb{Z}^+$ as[5]

$$b_{t+1}(\widetilde{s}) = \frac{\sum_{s\in\Sigma} b_t(s) a_t(s, \widetilde{s}) \Xi(y_t, s, \widetilde{s})}{\sum_{s,s'\in\Sigma} b_t(s) a_t(s, s') \Xi(y_t, s, s')} \ ; \tag{14}$$

finally, the reward function is the conditional mutual information

$$\mathsf{r}(b_t, a_t) \triangleq I(S_t, X_t; Y_t \mid Y_1^{t-1} = \boldsymbol{y}_1^{t-1}) \ , \tag{15}$$

which depends only on the distribution $\mathrm{Prob}(S_t = s, X_t = x, Y_t = y \mid Y_1^{t-1} = \boldsymbol{y}_1^{t-1})$, which, in turn, is a product of $b_t(s)$, $a_t(s, x)$ and $\Xi(y, s, x)$, as they appeared in (10)–(12). Substituting (15) into (9) and noticing that looking for the supremum on the range of policies $\boldsymbol{\pi}$ in our case is equivalent to optimizing on the distribution $\mathrm{Prob}(X_t \mid S_t, Y_1^{t-1})$ yields

$$\Psi = \sup_{\boldsymbol{\pi}=\mathrm{Prob}(X_t|S_t,Y_1^{t-1})} \liminf_{n\to\infty} \frac{1}{n} \mathrm{E}_{\boldsymbol{\pi}} \Big(\sum_{t\in\langle 1,n\rangle} I(S_t, X_t; Y_t | Y_1^{t-1}{=}\boldsymbol{y}_1^{t-1}) \Big). \tag{16}$$

It was proved in [20, Th. 1] that the right-hand side of (16) equals the capacity of the underlying channel if its state process $\mathcal{S}$ evolves according to the equation $s_{t+1} = g(s_t, x_t, y_t)$ for some function $g(\cdot)$ and if there is a nonzero probability of reaching any channel state s from any other channel state s' in a finite number of steps. Clearly, both prerequisites hold in our case, therefore (16) transforms to merely $\Psi = \mathsf{cap_{FB}}(p)$.

C. Statement and solution of the Bellman equation

The Bellman equation in its general form ([1, Th. 2.1]), which comes next, gives an alternative characterization of the optimal average reward Ψ. We present a slightly modified discrete form of that integral equation below (see [1, Th. 5.1]).

Theorem 3.2 ([1]): The optimal average reward of a dynamic system characterized by $\mathcal{T}$ equals $\Psi \in \mathbb{R}$ if there exists a function $\mathsf{h} : \mathcal{B} \to \mathbb{R}$ such that for all $b \in \mathcal{B}$

$$\Psi + \mathsf{h}(b) = \sup_{a\in\mathcal{A}} \Big(\mathsf{r}(b, a) + \sum_{d\in\mathcal{D}} \upsilon(d \mid b, a) \mathsf{h}(\mathsf{e}(b, a, d)) \Big) \ . \tag{17}$$

□

To state a specialized form of the identity (17), we first represent the expressions (13)–(15) as functions of $b_t(0)$, $\gamma_t \triangleq (1{-}b_t(0)) \cdot a_t(1, 1)$ and $\delta_t \triangleq b_t(0) \cdot a_t(0, 0)$, in equalities (18)–(20), respectively:

$$\begin{aligned} \upsilon(0 \mid b_t, a_t) &= \upsilon = \upsilon(b_t(0), \delta_t, \gamma_t) = \\ &= (1{-}p)(1{+}\delta_t{-}\gamma_t) + (2p{-}1) \cdot b_t(0), \\ \upsilon(1 \mid b_t, a_t) &= 1 - \upsilon, \end{aligned} \tag{18}$$

$$\begin{aligned} b_{t+1}(0) &= \begin{cases} \frac{1}{\upsilon}(\delta_t + (1{-}p)(1{-}b_t(0){-}\gamma_t)) & y_t = 0 \\ \frac{p}{1-\upsilon}(1{-}b_t(0){-}\gamma_t) & y_t = 1 \end{cases}, \\ b_{t+1}(1) &= 1{-}b_{t+1}(0), \end{aligned} \tag{19}$$

$$\begin{aligned} \mathsf{r}(b_t, a_t) &= H(Y_t \mid Y_1^{t-1}{=}\boldsymbol{y}_1^{t-1}) {-} H(Y_t \mid X_{t-1}^t, Y_1^{t-1}{=}\boldsymbol{y}_1^{t-1}) \\ &= \mathrm{H}(\mathrm{Prob}(Y_t = 0 \mid Y_1^{t-1} = \boldsymbol{y}_1^{t-1})) \\ &\quad - \mathrm{H}(p) \cdot \mathrm{Prob}(X_t \neq X_{t-1} \mid Y_1^{t-1} = \boldsymbol{y}_1^{t-1}) \\ &= \mathrm{H}(\upsilon) {+} \mathrm{H}(p)(\delta_t{+}\gamma_t{-}1). \end{aligned} \tag{20}$$

[5]To see how the recurrence (14) is obtained, the reader is referred to [20, Eq. (35)].

Letting $b \triangleq b_t(0)$ and getting rid of the rest of the subindexes t allows us to rewrite (17) as follows. Let

$$\mathsf{z}(b,\gamma,\delta;\mathsf{h}) \triangleq \mathrm{H}(\upsilon)+\mathrm{H}(p)(\delta+\gamma-1)+(1-\upsilon)\mathsf{h}\Big(\frac{p}{1-\upsilon}(1-b-\gamma)\Big) + \upsilon\cdot\mathsf{h}\Big(\frac{1}{\upsilon}(\delta+(1-p)(1-b-\gamma))\Big), \tag{21}$$

then for $b \in [0,1]$,

$$\Psi+\mathsf{h}(b) = \sup_{(\gamma,\delta)\in[0,1-b]\times[0,b]} \mathsf{z}(b,\gamma,\delta;\mathsf{h}) \ . \tag{22}$$

Using the iterative method we describe later on in Section IV-C for finding a lower bound on $\mathsf{cap}_{\mathsf{FB}}(p)$ on the interval $[0,p_0]$, we can obtain a discretized estimate $\boldsymbol{\chi}^{(m)}$ on the values of the function h, as well as a lower bound on the number Ψ. Based on these two quantities, we can correctly guess the actual function h and the actual number Ψ solving (22). Let $p \in [0,p_0]$ (where p_0 is defined in the statement of Theorem 3.1), then $b^\star = b^\star(p) = \frac{1}{2^{\mathrm{H}(p)}+1}$,

$$\Psi^\star = \Psi^\star(p) = \mathrm{H}(b^\star) - b^\star\cdot\mathrm{H}(p), \tag{23}$$

$$\mathsf{h}^\star(b) = \mathsf{h}^\star(b,p) = \begin{cases} \mathrm{H}(p)b+\Psi^\star & b\in[0,b^\star] \\ \mathrm{H}(b) & b\in[b^\star,1-b^\star] \\ \mathrm{H}(p)(1-b)+\Psi^\star & b\in[1-b^\star,1] \end{cases} . \tag{24}$$

The proof that $\Psi^\star$ and $\mathsf{h}^\star$ indeed solve (22) appears in Proposition B.3 in Appendix B, wherefrom follows the upcoming corollary proving Theorem 3.1.

Corollary 3.3: Let $p \in [0,p_0]$, then $\mathsf{cap}_{\mathsf{FB}}(p) = \Psi^\star$. □

IV. Lower bounds on $\mathsf{cap}_{\mathsf{FB}}(p)$

In this section, we present three lower bounds on $\mathsf{cap}_{\mathsf{FB}}(p)$, for values of p in the range $[0,1]$. The first lower bound is near-capacity-achieving on the interval $[0,p_0]$ and is obtained by exhibiting an actual zero-error coding scheme, inspired by an idea from [26, Sec. 2-B]. The second lower bound is a generalization of the coding scheme of [10, Th. 2] (presented in Section IV-B), whereas the third one is an iterative technique based on the Bellman equation (presented in Section IV-C). Thanks to a recent work by Sabag *et al.* [23], the second lower bound is tight, as we show in Appendix C, thereby extending our knowledge of the feedback capacity to the interval $[0,0.5]$.

A. Near-capacity-achieving encoder for $p\in[0,p_0]$.

The encoder for the values $p\in[0,p_0]$ is inspired by the method of "lost ones retransmission" from [26, Sec. 2-B]. We next explain how to adapt this method to our setting.

Our feedback-equipped zero-error encoder will operate in iterations. In the first iteration, we will send to the receiver an encoding of the actual information message. In subsequent iterations, we will transmit an encoding of an index (i.e., concise representation) of the error combination that occurred in the previous iteration and detected through feedback. The information message and the indices to error combinations will be encoded using a Markovian source and will be transmitted to the receiver through the channel $\mathtt{Is}_{\mathsf{FB}}(p)$ bit-by-bit with repetitions of bits corrupted by the channel and detected through feedback. During some periods in the transmission, we will employ a simple code which repeats every input bit; clearly, such a code can be easily decoded by reading off the values of the odd-indexed coordinates (unaffected by the channel). With high probability, this code will be used on a logarithmic number of input bits in total, which will not affect the encoder rate asymptotically.

Let μ be a positive integer and let $P = P_\mu$ be a Markov chain with μ states indexed by integers from 1 through μ, whose transition probabilities are parametrized by real variables $q_1, q_1, \ldots, q_\mu \in [0,1]$, as shown in Figure 5. We

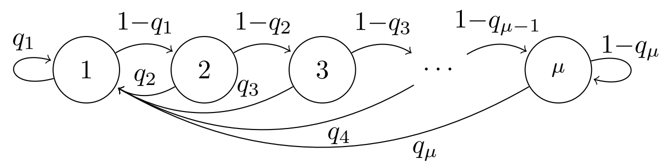

Fig. 5. Markov chain P_μ from Section IV-A.

interpret the states 1 through $\mu-1$ of P as lengths of the last run of the sequence Y_1^t generated so far at the output of the channel; the state μ will stand for lengths μ and above. The probability of switching the last output bit (and starting a new run of length 1) in a state $i \in \langle 1,\mu\rangle$ is q_i; the probability of continuing the current run of 0's or 1's from the state i is accordingly $1-q_i$. Denote the stationary probability of a state $i\in\langle 1,\mu\rangle$ of P by ς_i and the entropy rate $\sum_{i\in\langle 1,\mu\rangle}\varsigma_i\,\mathrm{H}(q_i)$ of P by $\mathsf{H}(P)$.

Let ϵ be an arbitrarily small positive real number, standing for a control parameter that determines the encoder rate, and let $N = N(\epsilon,p,\mu)$ be a sufficiently large positive integer, standing for the length of the information message to be sent. Let $O(\epsilon)$ denote a constant multiple of ϵ throughout the rest of this section. The Markov chain P can be viewed as an encoder[6] $\mathsf{E} = \mathsf{E}_{P,\ell,\epsilon}$ (with a corresponding decoder $\mathsf{D} = \mathsf{D}_{P,\ell,\epsilon}$), which takes an information message $M = (M_i)_{i\in\langle 1,\ell\rangle}$ of length ℓ and outputs a typical (for the Markov chain P) sequence of length $\frac{\ell}{\mathsf{H}(P)} + O(\epsilon)$ indexed by M, such that the per-symbol number of runs of length j is $\varsigma_j q_j + O(\epsilon)$ for all $j\in\langle 1,\mu-1\rangle$ and the per-symbol number of runs of length at least μ is $\varsigma_\mu q_\mu + O(\epsilon)$. Figure 6 demonstrates the `transmit_all` function which takes ϵ, the Markov chain P_μ, and a message word M of length N as arguments and transmits an encoding of M to the receiver along with recurring encodings of errors from the previous iterations using the auxiliary functions `transmit_segment` and `rep`. The function `transmit_segment` takes $\widehat{\epsilon}$ (derived from the original ϵ), the Markov chain P, and a word $\widehat{M}$ of length ℓ and transmits the encoding of $\widehat{M}$ using $\mathsf{E}_{P,\ell,\widehat{\epsilon}}$ to the receiver with retransmissions of corrupted previously-transmitted bits. It returns the set of coordinates $\mathcal{E}$ where transmission errors occurred, and sets $\mathcal{E}_j$, $j\in\langle 1,\mu\rangle$, such that $\mathcal{E}_j$, for $j\in\langle 1,\mu-1\rangle$, contains the positions of ends of runs of length j in the output and $\mathcal{E}_\mu$ contains the positions of the ends of the remaining runs. Note that in lines 8, 12, and 13 we are adding indexes of positions as they appear in the output;

[6]To see how to mimic a given Markov chain by a fixed-rate encoder, the reader is referred to [24, Sec. 3].

we could have represented the same positions more concisely by disregarding the retransmitted bits of the output (which is equivalent to not adding $|\mathcal{E}|$ in these three algorithm lines). We will employ this concise representation in the function `typical_index_message`, which will be explained later on. The function `rep` takes a message $\widehat{M} = (\widehat{M}_i)_i$ as its first argument and the length of the transmission output ℓ as its second argument and transmits $(\widehat{M}_i\widehat{M}_i)_i$ padded with 0's to the required output length ℓ. In the first invocation of `transmit_segment` in line 18, we send an encoding of the original message M, after which we proceed with sending the index message J corresponding to the error combination from the previous transmission, in a recurring manner (lines 20–35). To that end, in each of the iterations of the while loop (lines 21–24), we first send the sizes of the sets $\mathcal{E}_j \cap \mathcal{E}$ and $\mathcal{E}_j$ for $j \in \langle 1, \mu\rangle$ obtained in the previous invocation of `transmit_segment`. In line 25, we use the function `typical_index_message` (whose definition is omitted from Figure 6) to generate a message J, whose contents represent the index of the error combination that occurred in the previous invocation of `transmit_segment`, in the set of $\widehat{\epsilon}$-typical error combinations, which satisfy,

$$\left|\frac{|\mathcal{E}_j \cap \mathcal{E}|}{|\mathcal{E}_j \cap \mathcal{E}| + |\mathcal{E}_{j-1} \setminus \mathcal{E}|} - p\right| \leq \widehat{\epsilon}\,, \quad \text{for } j \in \langle 2, \mu{-}1\rangle\,, \tag{25}$$

as well as

$$\left|\frac{|\mathcal{E}_\mu \cap \mathcal{E}|}{|\mathcal{E}_\mu| + |\mathcal{E}_{\mu-1} \setminus \mathcal{E}|} - p\right| \leq \widehat{\epsilon}\,. \tag{26}$$

The numerator in (25) stands for the number of input runs of length $j{-}1$ whose right ends were corrupted, whereas the denominator counts the number of all input runs of length $j{-}1$, for $j \in \langle 2, \mu{-}1\rangle$. Similarly, the numerator and the denominator in (26) stand for the number of input runs of length at least μ whose right ends were corrupted and the number of all input runs of length at least μ, respectively.

For a length ℓ large enough, the probability that the function `transmit_segment` generates an index to an $\widehat{\epsilon}$-typical combination is at least $1{-}\widehat{\epsilon}$. The contents of J are essentially a concatenation of $\mu{-}1$ indexes enumerated by integers $j = 2, 3, \ldots, \mu$, such that the j-th index represents the position of the obtained set of coordinates $\mathcal{E}_j \cap \mathcal{E}$ within an ordered list of sets of coordinates standing for typical errors at the right ends of output runs of length j. The function also returns a Boolean b which equals true if and only if the error combination is $\widehat{\epsilon}$-typical; in the unlikely scenario that it is not $\widehat{\epsilon}$-typical, we send the entire message M by means of the `rep` function (preceded by a special synchronization symbol, omitted from the algorithm and used by the decoder to detect the special case) and then halt at line 28. The iteration of the while loop ends with the transmission of J by means of `transmit_segment` in line 34. We leave the while loop when the size $|J|$ of the index message falls below some prescribed integer (say, $\sqrt{N}$), in which case we transmit J using the simple encoding of the `rep` function (line 32). The decoding of the above transmission is straightforward, and we therefore omit the details.

We now proceed with the asymptotic rate analysis of the presented encoder. All the times when the `rep` function is used

---

FUNCTION: `rep`
INPUT: message $\widehat{M} = (\widehat{M}_i)_i$, ℓ
OUTPUT: message M of length ℓ
  // || is the concatenation operator
01: $M \leftarrow 0^{\ell-2\cdot\texttt{length}(\widehat{M})}||(\widehat{M}_i\widehat{M}_i)_{i\in\left\langle 1,\texttt{length}(\widehat{M})\right\rangle}$

FUNCTION: `transmit_segment`
INPUT: $\widehat{\epsilon}$, Markov chain P_μ, message $\widehat{M} = (\widehat{M}_i)_{i\in\langle 1,\ell\rangle}$
OUTPUT: set $\mathcal{E}$, sets $(\mathcal{E}_j)_{j\in\langle 1,\mu\rangle}$
02: $\mathcal{E} \leftarrow \emptyset$
03: $\boldsymbol{\chi} = (X_i)_{i\in\langle 1,\texttt{length}(\boldsymbol{\chi})\rangle} \leftarrow \mathsf{E}_{P_\mu,\ell,\widehat{\epsilon}}(\widehat{M})$
04: **for** $i \in \langle 1, \texttt{length}(\boldsymbol{\chi})\rangle$ **do** {
05:   transmit X_i through $\mathsf{Is}_{\mathsf{FB}}(p)$ and receive Y_i through feedback
06:   **if**$(Y_i \neq Y_{i-1})$ {
07:     $j \leftarrow \texttt{min}$(length of the last output run ending with Y_{i-1}, μ)
08:     $\mathcal{E}_j \leftarrow \mathcal{E}_j \cup \{i - 1 + |\mathcal{E}|\}$
09:   }
10:   **if**$(X_i \neq Y_i)$ {
11:     $j \leftarrow \texttt{min}$(length of the last output run ending with Y_i, μ)
12:     $\mathcal{E}_j \leftarrow \mathcal{E}_j \cup \{i + |\mathcal{E}|\}$
13:     $\mathcal{E} \leftarrow \mathcal{E} \cup \{i + |\mathcal{E}|\}$
14:     retransmit X_i
15:   }
16: }

FUNCTION: `transmit_all`
INPUT: ϵ, Markov chain P_μ, message $M = (M_i)_{i\in\langle 1,N\rangle}$
17: $\widehat{\epsilon} \leftarrow \epsilon/(N\left\lfloor\log_{\mathsf{H}(P)/\mathsf{i}}\sqrt{N}\right\rfloor)$
  // first invocation
18: $(\mathcal{E}, \mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_\mu) \leftarrow \texttt{transmit\_segment}(\widehat{\epsilon}, P_\mu, M)$
19: $J \leftarrow M$
20: **while**(**true**) { // subsequent invocations
21:   **for** $j \in \langle 1, \mu\rangle$ **do** {
22:     transmit $\texttt{rep}(|\mathcal{E}_j \cap \mathcal{E}|\,, 2\lceil\log_2 N\rceil)$
23:     transmit $\texttt{rep}(|\mathcal{E}_j|\,, 2\lceil\log_2 N\rceil)$
24:   }
25:   $(b, J) \leftarrow \texttt{typical\_index\_message}(p, \widehat{\epsilon}, \mathcal{E}, \mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_\mu)$
26:   **if**(b **is false**) {
    // we precede the transmission with a special symbol
27:     transmit $\texttt{rep}(M, 2N)$
28:     **exit**
29:   }
30:   **if**$(|J| < \sqrt{N})$ { // stopping condition
31:     transmit $\texttt{rep}(J, 2\sqrt{N})$
32:     **break**
33:   }
34:   $(\mathcal{E}, \mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_\mu) \leftarrow \texttt{transmit\_segment}(\widehat{\epsilon}, P_\mu, J)$
35: }

Fig. 6. Transmission procedure for the encoder from Section IV-A.

to transmit bits to the receiver, except at line 27, can be neglected for the purposes of our analysis, because altogether the expected number of bits sent in all such transmissions is sublinear in N. Only bits transmitted via `transmit_segment` and the $2N$ bits transmitted at line 27 will thus count towards our analysis. The `transmit_segment` function incurs two rate penalties: for the repetition of corrupted bits (inflating the message length by a factor of at most $1{+}p\varsigma_1 + O(\widehat{\epsilon})$) and for the use of the encoding $\mathsf{E}_{P,\ell,\widehat{\epsilon}}$ (inflating the message length by a factor of at most $\frac{1}{\mathsf{H}(P)} + O(\widehat{\epsilon})$). Overall, the inflation factor of `transmit_segment` is at most $\frac{1+p\varsigma_1}{\mathsf{H}(P)} + O(\widehat{\epsilon})$ bits. Denote the length of a message transmitted in the j-th invocation of `transmit_segment` by N_j (taking retransmissions into account). Denote by t_j the value of $|J|$ at the end of the j-th iteration of the while loop and by $\mathsf{i} = \mathsf{i}(p, P_\mu)$ the rate of `typical_index_message`. If the function `typical_index_message` receives an $\widehat{\epsilon}$-

typical error combination $\mathcal{E}, \mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_\mu$, it returns a message J of length inflated by at most $\mathsf{i} + O(\widehat{\epsilon})$ relative to the length of message $\boldsymbol{\chi}$ obtained in the preceding invocation of `transmit_segment`. Then, for any integer $j \geq 1$, one has

$$N_{j+1} \leq t_j \left(\frac{1+p\varsigma_1}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right)$$

and

$$t_j \leq \frac{N_j(\mathsf{i} + O(\widehat{\epsilon}))}{1+p\varsigma_1 - O(\widehat{\epsilon})} \leq \frac{N_j \cdot \mathsf{i}}{1+p\varsigma_1} + O(\widehat{\epsilon}) \;,$$

implying

$$\frac{N_{j+1}}{N_j} \leq \frac{\mathsf{i}}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \tag{27}$$

and

$$\frac{t_{j+1}}{t_j} \leq \frac{\mathsf{i}}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \;. \tag{28}$$

At the end of the first iteration, we have

$$N_1 \leq N \left(\frac{(1+p\varsigma_1)}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right) \tag{29}$$

and

$$t_1 \leq N \left(\frac{1}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right) (\mathsf{i} + O(\widehat{\epsilon})) \leq N \left(\frac{\mathsf{i}}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right) \;. \tag{30}$$

If the exit statement in line 28 is never reached, the while loop ends at iteration $m+1$, where m is the largest integer satisfying $t_m \geq \sqrt{N}$. Since

$$t_m = t_1 \prod_{j\in\langle 1, m-1\rangle} \frac{t_{j+1}}{t_j} \overset{(28)}{\leq} t_1 \left(\frac{\mathsf{i}}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right)^{m-1} \overset{(30)}{\leq} N \left(\frac{\mathsf{i}}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right)^{m} ,$$

one has

$$m+1 \leq \left\lfloor \log_{1/(\frac{\mathsf{i}}{\mathsf{H}(P)} + K_1\widehat{\epsilon})} \sqrt{N} \right\rfloor + 1$$

for some positive constant K_1. The probability of never reaching the exit statement in line 28 is therefore at least

$$(1-\widehat{\epsilon})^{\lfloor \log_{1/(\mathsf{i}/\mathsf{H}(P)+K_1\widehat{\epsilon})} \sqrt{N} \rfloor + 1} \geq (1-\widehat{\epsilon})^{\lfloor \log_{\mathsf{H}(P)/\mathsf{i}} \sqrt{N} \rfloor} = \left(1 - \frac{\epsilon}{N \left\lfloor \log_{\mathsf{H}(P)/\mathsf{i}} \sqrt{N} \right\rfloor} \right)^{\lfloor \log_{\mathsf{H}(P)/\mathsf{i}} \sqrt{N} \rfloor},$$

which for N large enough is at least $\mathrm{e}^{-2\epsilon/N} \geq 1 - \frac{2\epsilon}{N}$. The total length of messages transmitted by `transmit_segment` under the assumption that the exit statement in line 28 is never reached is at most

$$\begin{aligned}\sum_{j\in\langle 1, m+1\rangle} N_j &\overset{(27)}{\leq} N_1 \sum_{j\in\langle 0, m\rangle} \left(\frac{\mathsf{i}}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right)^j \\ &\leq N_1 \sum_{j\in\mathbb{Z}^+\cup\{0\}} \left(\frac{\mathsf{i}}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right)^j \\ &\overset{(29)}{\leq} N \left(\frac{(1+p\varsigma_1)}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right) \frac{1}{1 - \frac{\mathsf{i}}{\mathsf{H}(P)} - O(\widehat{\epsilon})} \\ &\leq N \left(\frac{(1+p\varsigma_1)}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right) \frac{1}{1 - \frac{\mathsf{i}}{\mathsf{H}(P)}} \;. \end{aligned} \tag{31}$$

When the exit statement is reached, the total length of transmitted messages is increased by $2N$; in this case we can bound their length from above by

$$\sum_{j\in\langle 1, m+1\rangle} N_j + 2N \overset{(31)}{\leq} N \left(\frac{(1+p\varsigma_1)}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right) \frac{1}{1 - \frac{\mathsf{i}}{\mathsf{H}(P)}} + 2N \leq K_2 N$$

for some positive constant K_2. The expected length of messages transmitted by the `transmit_segment` function, for N large enough, is therefore at most

$$N \left(\frac{(1+p\varsigma_1)}{\mathsf{H}(P)} + O(\widehat{\epsilon}) \right) \frac{1}{1 - \frac{\mathsf{i}}{\mathsf{H}(P)}} + \frac{2\epsilon}{N} \cdot K_2 N \leq N \left(\frac{(1+p\varsigma_1)}{\mathsf{H}(P) - \mathsf{i}} + \epsilon \right),$$

implying that the rate of `transmit_all` is at least

$$\frac{\mathsf{H}(P) - \mathsf{i}}{1 + p\varsigma_1 + (\mathsf{H}(P) - \mathsf{i})\epsilon} \geq \frac{\mathsf{H}(P) - \mathsf{i}}{1 + p\varsigma_1} - \epsilon \;.$$

We conclude that by taking ever-decreasing values of $\epsilon > 0$, we will be able to approach a rate of $\frac{\mathsf{H}(P)-\mathsf{i}}{1+p\varsigma_1}$.

To finish our analysis, we need to write out the expression for i; to that end, we first need to explain how `typical_index_message` works. We notice that in a coding scheme with feedback like ours, where bits get repeated immediately after being corrupted, from the standpoint of the receiver, all the errors in the output occur at the ends of runs. Moreover, the output runs of length 1 got transmitted to the receiver without disruptions, whereas an output run of any given length $j \geq 2$ originated in either an input run of length j that was transmitted error-free or in an input run of length $j-1$ whose end position was "smeared" into the beginning of the following input run. Thus, for unique decoding by the receiver, it suffices for the encoder to point at the runs of lengths 2 and above whose right ends were corrupted by the channel. A typical sequence $\boldsymbol{\chi}$ of length ℓ obtained at line 3 of Figure 6 will typically yield in the output

- $|\mathcal{E}_1| \sim \kappa_1 \ell$ error-free runs of length 1, where $\kappa_1 \triangleq \varsigma_1 q_1 (1-p)$;
- $|\mathcal{E}_j| \sim \kappa_j \ell$ runs of length j, for any given $j \in \langle 2, \mu-1\rangle$, with $|\mathcal{E}_j \cap \mathcal{E}| \sim \lambda_j \ell$ runs having an error at the rightmost bit, where
$$\lambda_j \triangleq \varsigma_{j-1} q_{j-1} p \quad \text{and} \quad \kappa_j \triangleq \varsigma_j q_j (1-p) + \lambda_j \;; \tag{32}$$
- $|\mathcal{E}_\mu| \sim \kappa_\mu \ell$ runs of length μ, with $|\mathcal{E}_\mu \cap \mathcal{E}| \sim \lambda_\mu \ell$ runs having an error at the rightmost bit, where
$$\lambda_\mu \triangleq \varsigma_{\mu-1} q_{\mu-1} p \quad \text{and} \quad \kappa_\mu \triangleq \varsigma_\mu q_\mu + \lambda_\mu \;. \tag{33}$$

The number of typical error combinations in the transmission through $\mathtt{Is}_{\mathsf{FB}}(p)$ (that are also typical with respect to the Markov chain P) equals, up to a factor that is sub-exponential in ℓ, the following multinomial coefficient:

$$\binom{\kappa_2 \ell}{\lambda_2 \ell} \binom{\kappa_3 \ell}{\lambda_3 \ell} \cdot \ldots \cdot \binom{\kappa_\mu \ell}{\lambda_\mu \ell},$$

whose asymptotic growth rate (with respect to the sequence length ℓ) equals the rate of `typical_index_message`, namely,

$$\mathsf{i} = \sum_{j\in\langle 2, \mu\rangle} \kappa_j \, \mathrm{H}\left(\frac{\lambda_j}{\kappa_j} \right).$$

The following theorem sums up the above discussion.

Theorem 4.1: Let $p \in [0,1]$ and let μ be a positive integer. Let P_μ be a Markov chain defined as in Figure 5. Then

$$\mathsf{cap}_{\mathsf{FB}}(p) \geq \varrho_1(p,\mu) = \max_{(q_j)_{j\in\langle 1,\mu\rangle}\in[0,1]^\mu} \frac{\mathrm{H}(P_\mu) - \sum_{j\in\langle 2,\mu\rangle} \kappa_j \,\mathrm{H}\left(\frac{\lambda_j}{\kappa_j}\right)}{1+p\varsigma_1},$$

where κ_j and λ_j for $j \in \langle 2,\mu\rangle$ are defined in (32),(33). □

The above scheme proves to be near-capacity-achieving on the interval $[0, p_0]$; for instance, the largest difference between $\mathsf{cap}_{\mathsf{FB}}(p)$ and $\varrho_1(p, 9)$ is at $p = p_0$, where $\mathsf{cap}_{\mathsf{FB}}(p_0) \approx 0.595041$ and $\varrho_1(p_0, 9) \approx 0.594737$ (within 0.051% from the capacity).

B. Generalization of the encoder from [10, Th. 2]

Next, we present a second coding scheme for $\mathtt{Is}_{\mathsf{FB}}(p)$, which is a generalization of the scheme in [10, Th. 2].

The finite-state machine appearing in Figure 7 schematically describes both an encoder E and a decoder D for $\mathtt{Is}_{\mathsf{FB}}(p)$. The input to the encoder is a message $M = (M_i)_{i\in\langle 1,n\rangle}$ drawn from the set of binary vectors of length n starting with a 1, whose number of runs[7] is within $(\xi \pm \epsilon)n$ for an arbitrarily small $\epsilon > 0$, where $\xi = \xi(p) \in [0,1]$ maximizes the expression $\frac{\mathrm{H}(\xi)}{2-\xi(1-p)}$. In other words, the number of alternations from a 0 to a 1 or from a 1 to a 0, as we move along the bits of M, is within $(\xi \pm \epsilon)n - 1$. After transmitting the first bit $X_1 = M_1 = 1$ (and receiving the first output bit $Y_1 = 1$ through the feedback), we enter state S_1 of E; this state will designate (both in E and in D) the event that the previous bit was transmitted safely (viz., $X_{t-1} = Y_{t-1}$). We remain in this state as long as the bit Y_t received through the feedback at time t is different from the previously received bit Y_{t-1} and keep transmitting the following bits of the message M in order. Next, we consider the case where the last two received bits Y_t and Y_{t-1} are the same. Since being in the state S_1 implies that $X_{t-1} = Y_{t-1}$, we have $Y_t = Y_{t-1}$ either when the last two transmitted bits are the same ($X_t = X_{t-1}$), or when there was an error in the last transmitted bit ($X_t \neq Y_t$); in both events we enter state S_0, and, on the next time slot $t+1$, retransmit the last bit X_t. From state S_0 we always transition back to state S_1. This way, $(1-p)((\xi \pm \epsilon)n-1)$ bits of M are transmitted once and the remaining $n-(1-p)((\xi \pm \epsilon)n-1)$ bits — twice. Since the exponential growth rate of words with $(\xi \pm \epsilon)n$ runs, for $\epsilon \to 0$, tends to $\mathrm{H}(\xi)$, the average rate of the encoder in Figure 7 is

$$\varrho_2(p) = \frac{\mathrm{H}(\xi)}{1\cdot\xi(1-p)+2\cdot(1-\xi(1-p))} = \frac{\mathrm{H}(\xi)}{2-\xi(1-p)} \ . \tag{34}$$

The decoder D is equivalent to the decoder given in [10, Th. 2]; for completeness, we briefly describe its operation next. It starts in state S_1 upon receiving the first bit $Y_1 = 1$ from the channel. In state S_1 at time t, we always decode the next message bit as Y_t, whereas in state S_0 we wait one time slot. As with the encoder E, we remain in S_1, as long as the last two received bits are different ($Y_t \neq Y_{t-1}$) and transition to S_0 otherwise; from S_0 we deterministically transition to S_1. This scheme guarantees error-free decoding, as the encoder at time $t+1$ always retransmits the t-th bit, once it detects that $Y_t = Y_{t-1}$, hence after waiting one time slot in state S_0, we are guaranteed to receive $Y_{t+1} = X_t$, which represents the next message bit sent by the encoder.

[7]Messages with a prescribed number of runs can be generated by applying enumerative coding [8][13, Ch. 6] to raw input bits that are i.i.d. Bernoulli with success probability 0.5.

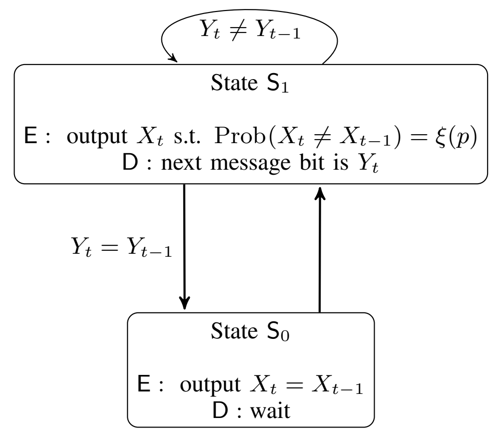

Fig. 7. Encoder E and decoder D for $\mathtt{Is}_{\mathsf{FB}}(p)$.

```
INPUT: p ∈ [0,1], vector χ = (x_j)_{j∈⟨0,Λ⟩}
OUTPUT: vector χ' = (x'_j)_{j∈⟨0,Λ⟩}
  χ' ← 0;
  for j ∈ ⟨0,Λ⟩ do {
    b ← j/Λ;
    for (δ ← 0; δ ≤ b; δ ← δ+1/Λ) do {
      for (γ ← 0; γ ≤ 1−b; γ ← γ+1/Λ) do {
        υ ← (1−p)(1+δ−γ) + (2p−1)b;
        (d1, d2) ← ( p/(1−υ) (1−b−γ), 1/υ (δ + (1−p)(1−b−γ)) );
        (m1, m2) ←
          ( (⌊Λd1⌋ +1−Λd1) x_{⌊Λd1⌋} + (Λd1− ⌊Λd1⌋) x_{⌊Λd1⌋+1},
            (⌊Λd2⌋ +1−Λd2) x_{⌊Λd2⌋} + (Λd2− ⌊Λd2⌋) x_{⌊Λd2⌋+1} );
        x'_j ← max { x'_j, H(υ) + H(p)(δ+γ−1) + (1−υ)·m1+υ·m2 };
      }
    }
  }
```

Fig. 8. One iteration of a numeric method for the lower bound $\varrho_3(p,\Lambda)$.

The aforementioned coding scheme was proved in [10] to be capacity-achieving at $p = 0.5$ with code rate ≈ 0.575522. It is also capacity-achieving at $p = p_0$, where $\mathsf{cap}_{\mathsf{FB}}(p)$, as it appears in Corollary 3.3, and the code rate $\varrho_2(p)$ from (34) can be shown to be equal (to approximately 0.595045). Recently, Sabag *et al.* [23, Th. 1] have shown a general technique, called the Q-contexts mapping, for obtaining upper bounds on the feedback capacity of a subset of finite-state channels, to which the generalized Ising channels belong. Using this technique, one can demonstrate that the lower bound $\varrho_2(p)$ is, in fact, tight on the entire interval $[p_0, 0.5]$, as shown in Appendix C.

C. Iterative method

Another technique for obtaining simple lower bounds on $\mathsf{cap}_{\mathsf{FB}}(p)$ is to start with a "guess" $\mathsf{h}_0 : [0,1] \to \mathbb{R}$ bounding the function $\mathsf{h}^\star$ from below, and perform the following iterations for $i \in \mathbb{Z}^+$:

$$\mathsf{h}_i(b) = \sup_{(\gamma,\delta)\in[0,1-b]\times[0,b]} \mathsf{z}(b,\gamma,\delta;\mathsf{h}_{i-1}), \quad b \in [0,1] \ . \quad (35)$$

Since for any solution h of the Bellman equation (22), the function $\mathsf{h}+c$ is also a solution for any constant $c \in \mathbb{R}$, for the purposes of this method we can choose $\mathsf{h}^\star(0) = 0$. One can prove by induction on i that, for any $i \in \mathbb{Z}^+ \cup \{0\}$ and $b \in [0,1]$,

$$\mathsf{h}_i(b) \le \mathsf{h}^\star(b) + i \cdot \Psi^\star \ . \quad (36)$$

Concretely, with $\mathsf{h}_0 \equiv 0$ as the initial guess, the induction basis $\mathsf{h}_0(b) \le \mathsf{h}^\star(b)$ readily holds. As for the induction step, assume that (36) holds for $i = m \in \mathbb{Z}^+ \cup \{0\}$. By definition,

$$\mathsf{h}_{m+1}(b) = \sup_{(\gamma,\delta)\in[0,1-b]\times[0,b]} \mathsf{z}(b,\gamma,\delta;\mathsf{h}_m),$$

and since $\mathsf{z}(b,\gamma,\delta;\mathsf{h})$ is monotonically increasing and additive in h, we have

$$\begin{aligned}\mathsf{h}_{m+1}(b) &\stackrel{(36)}{\le} \sup_{(\gamma,\delta)\in[0,1-b]\times[0,b]} \mathsf{z}(b,\gamma,\delta;\mathsf{h}^\star(b)+m\cdot\Psi^\star) \\ &= \sup_{(\gamma,\delta)\in[0,1-b]\times[0,b]} \mathsf{z}(b,\gamma,\delta;\mathsf{h}^\star(b)) + m\cdot\Psi^\star \\ &= (\mathsf{h}^\star(b)+\Psi^\star) + m\cdot\Psi^\star = \mathsf{h}^\star(b) + (m+1)\Psi^\star \ ,\end{aligned}$$

which completes the inductive proof. After substituting $b = 0$ into (36) and dividing both sides by i, we conclude that

$$\frac{\mathsf{h}_i(0)}{i} \le \Psi^\star \text{ for any } i \in \mathbb{Z}^+ \ . \quad (37)$$

One technical difficulty arising while calculating good lower bounds of this kind is the continuous nature of the functions h_i. While it is relatively easy to find h_1 analytically, the computation of the functions h_i for larger values of i becomes unwieldy. This problem, in turn, is remedied by discretizing the values of the parameter $b \in [0,1]$ at the expense of some slack in the bound. Figure 8 shows a procedure that gets a vector $\boldsymbol{\chi} = (x_j)_{j\in\langle 0,\Lambda\rangle}$ of length $\Lambda+1$, representing a discretized guess of the function $\mathsf{h}^\star$, bounding it from below, and returns a vector $\boldsymbol{\chi}' = (x'_j)_{j\in\langle 0,\Lambda\rangle}$ of the same length, which is an improvement over $\boldsymbol{\chi}$. Lemma 4.2, coming next, proves the concavity of all the functions h_i provided that the initial guess h_0 is concave, wherefrom it follows that $\boldsymbol{\chi}'$ is a discretized lower bound on h_i provided that $\boldsymbol{\chi}$ is a discretized lower bound on h_{i-1}.

Lemma 4.2: Let h_0 be a concave function on the interval $[0,1]$. Then the functions h_i, obtained as in (35), are concave for all $i \in \mathbb{Z}^+$. Hence, referring to vectors $\boldsymbol{\chi}$ and $\boldsymbol{\chi}'$ in Figure 8, if $\boldsymbol{\chi} \le (\mathsf{h}_{i-1}(j/\Lambda))_{j\in\langle 0,\Lambda\rangle}$ for some $i \in \mathbb{Z}^+$, then $\boldsymbol{\chi}' \le (\mathsf{h}_i(j/\Lambda))_{j\in\langle 0,\Lambda\rangle}$.

Proof: Denote by γ_j and δ_j the maximizing values of γ and δ in the right-hand side of (35) at points $b = b_j \in [0,1]$, for $j \in \{1,2\}$. Then, for any $\Upsilon \in [0,1]$, at the point $b = \Upsilon b_1 + (1-\Upsilon)b_2$ we have

$$\begin{aligned}\mathsf{h}_i(b) &= \sup_{(\gamma,\delta)\in[0,1-b]\times[0,b]} \mathsf{z}(b,\gamma,\delta;\mathsf{h}_{i-1}) \\ &\ge \mathsf{z}(b, \Upsilon\gamma_1 + (1-\Upsilon)\gamma_2, \Upsilon\delta_1 + (1-\Upsilon)\delta_2; \mathsf{h}_{i-1}) \ .\end{aligned}$$

Lemma B.1 (see Appendix B) proves that the function $\mathsf{z}(b,\gamma,\delta;\mathsf{h}_{i-1})$ is concave in the variables (b,γ,δ) provided h_{i-1} is concave, hence

$$\begin{aligned}\mathsf{h}_i(b) &\ge \Upsilon\cdot\mathsf{z}(b_1,\gamma_1,\delta_1;\mathsf{h}_{i-1}) + (1-\Upsilon)\mathsf{z}(b_2,\gamma_2,\delta_2;\mathsf{h}_{i-1}) \\ &= \Upsilon\cdot\mathsf{h}_i(b_1) + (1-\Upsilon)\mathsf{h}_i(b_2) \ ,\end{aligned}$$

thereby implying the concavity of the function h_i. By induction, this implies that if the initial guess h_0 is a concave function, so are h_i for all $i \in \mathbb{Z}^+$.

Next, assume that the entries of the input vector $\boldsymbol{\chi} = (x_j)_{j\in\langle 0,\Lambda\rangle}$ in the procedure in Figure 8 are lower bounds on the values of a concave function $\mathsf{h}_{i-1}(b)$ at the points $b = j/\Lambda$, for all $j \in \langle 0,\Lambda\rangle$ and for some $i \in \mathbb{Z}^+$. Due to the concavity of h_{i-1}, we have $(\mathsf{m}_1,\mathsf{m}_2) \le (\mathsf{h}_{i-1}(\mathsf{d}_1),\mathsf{h}_{i-1}(\mathsf{d}_2))$, hence $x'_j \le \mathsf{h}_i(j/\Lambda)$ for all $j \in \langle 0,\Lambda\rangle$. □

We can apply the procedure of Figure 8 till the difference between two consecutive lower bounds $\frac{\boldsymbol{\chi}^{(m)}}{m} - \frac{\boldsymbol{\chi}^{(m-1)}}{m-1}$ is sufficiently small. If we start, as suggested before, with the initial guess of $\boldsymbol{\chi}^{(0)} = (0)_{j\in\langle 0,\Lambda\rangle}$, then, by repeated invocation of Lemma 4.2 it follows that the quantity $\varrho_3(p,\Lambda) = \frac{\boldsymbol{\chi}^{(m)}(0)}{m}$, obtained while reaching the stopping condition after m iterations, is at most $\mathsf{h}_m(0)/m$, which, in turn, due to (37), is a lower bound on $\Psi^\star$. By the continuity of $\mathsf{z}(b,\gamma,\delta;\mathsf{h})$ in b,γ,δ wherever $b \in [0,1]$ and $(\gamma,\delta) \in [0,1-b]\times[0,b]$, the larger Λ is, the closer the vector $\boldsymbol{\chi}^{(m)}$ gets to h_m (in the points which are integer multiples of $1/\Lambda$), for any $m \in \mathbb{Z}^+$. This makes the lower bound obtained through the above discretization technique as close as needed to $\mathsf{h}_m(0)/m$, which, in turn, is a lower bound on $\Psi^\star$ by (37).

Figure 9 shows the curve $p \mapsto \mathsf{cap}_{\mathsf{FB}}(p)$ on the interval $[0,0.5]$, given by the following piecewise function

$$\begin{cases} \mathrm{H}\left(\frac{1}{2^{\mathrm{H}(p)}+1}\right) - \frac{\mathrm{H}(p)}{2^{\mathrm{H}(p)}+1} & p \in [0,p_0] \\ \varrho_2(p) & p \in [p_0,0.5] \end{cases} ,$$

alongside best lower bounds $p \mapsto \varrho_2(p)$, $p \mapsto \varrho_3(p,\Lambda{=}1000)$ on $\mathsf{cap}_{\mathsf{FB}}(p)$ and alongside a lower bound $p \mapsto \varrho(6,7,p)$ on the capacity $\mathsf{cap}(p)$ of the feedback-free channel $\mathtt{Is}(p)$ (and thus on $\mathsf{cap}_{\mathsf{FB}}(p)$) on the intervals $[0.5,0.525]$, $[0.525,0.7]$ and $[0.7,1]$, respectively (the vertical axis in the figure is not to scale for values less than 0.545). The lower bound $\varrho(6,7,p)$ on $\mathsf{cap}(p)$, plotted only on the interval $[0.5,1]$, is insignificantly better than $\varrho_3(p,\Lambda{=}1000)$ on the interval $[0.7,1]$. We do not plot the lower bound $\varrho_1(p,\mu)$ on the interval $[0,p_0]$, as even with $\mu = 9$, it is virtually indistinguishable from $\mathsf{cap}_{\mathsf{FB}}(p)$ therein. Marked by bold dots are the values of the capacity at $p = p_0$ and $p = 0.5$. Interestingly, at $p = p_0$, the obtained piecewise curve is C^1 continuous, as the left and the right derivatives of the function $\mathsf{cap}_{\mathsf{FB}}(p)$ at $p = p_0$ are both equal to approximately -0.201111. For comparison, on the interval $[0.5,1]$, we also plot the upper bounds $\rho_1(p)$ and $\rho_2(p)$ due to Theorem C.3 and Remarks C.6 and C.7 in Appendix C; these bounds are visibly distinguishable from one another only on the interval $[p_0,1-p_0]$. Based on numeric computations of $\rho_1(p)$ on the interval $[0.849946,0.899]$, we can also conclude that the curve $p \mapsto \mathsf{cap}_{\mathsf{FB}}(p)$ is not symmetric with respect to $p = 0.5$ on the interval $[0,1]$, unlike its counterpart without feedback [15, Prop. 1]. Marked by smaller dots is the extension of the curve $p \mapsto \varrho_2(p)$ to the interval $[0,p_0]$; one can observe

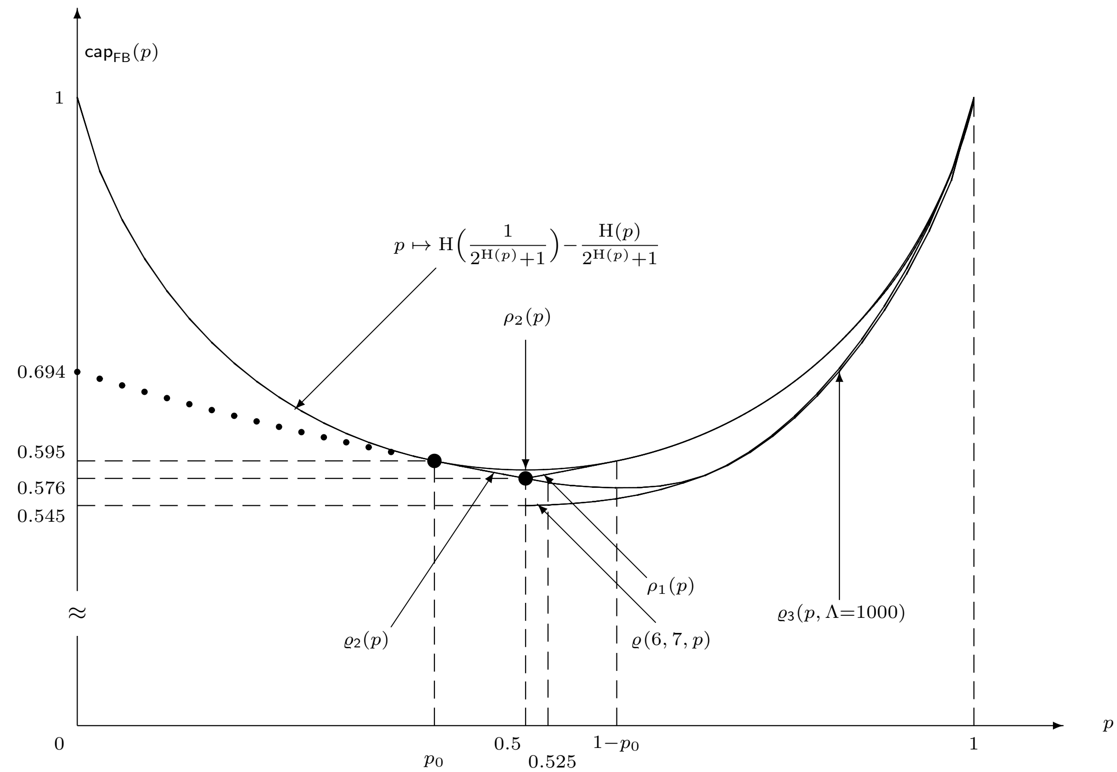

Fig. 9. Bounds on $\mathsf{cap}_{\mathsf{FB}}(p)$.

its suboptimality with respect to the value of the capacity therein.

Appendices

Appendix A.

We can assume in the proof of Theorems 2.1 and 2.2 that the input distribution $\mathcal{X}$ behaves as a stationary Markov chain due to the following observation.

Remark A.1: Following the definitions in [6], given a finite input alphabet $\Sigma_{\mathcal{X}}$ and a finite set of channel states $\Sigma_{\mathcal{S}}$, a *finite-state channel* is a collection of $|\Sigma_{\mathcal{X}}|$ stochastic matrices $C(x)$ of size $|\Sigma_{\mathcal{S}}| \times |\Sigma_{\mathcal{S}}|$ for each $x \in \Sigma_{\mathcal{X}}$, whose rows and columns are indexed by the states of $\Sigma_{\mathcal{S}}$. The entry (s, s') in matrix $C(x)$ stands for the conditional probability of moving from state s to state s' by the channel after feeding symbol $x \in \Sigma_{\mathcal{X}}$ to it. Additionally, the authors of [6] assume the following definition of an indecomposable matrix: a non-negative square matrix $A = (a_{i,i'})_{i,i'}$, whose rows and columns are indexed with integers in $\langle 1, w\rangle$, is said to be *indecomposable* if for any two indexes $i_0, i_1 \in \langle 1, w\rangle$, there exist non-negative integers t_0, t_1 and an index $i_2 \in \langle 1, w\rangle$ such that $A^{t_0}(i_0, i_2) > 0$ and $A^{t_1}(i_1, i_2) > 0$. Furthermore, [6, Th. 1] gives a necessary and sufficient condition for the indecomposability of a channel, specifically, a channel is indecomposable if and only if $\prod_{i\in\langle 1,t\rangle} C(x_i)$ is an indecomposable matrix for any $x_i \in \Sigma_{\mathcal{X}}$ and $i \in \langle 1, t\rangle$. Finally, [6, Th. 4] and the proof of [6, Th. 3] establish that the capacity of any indecomposable finite-state channel can be achieved by a family of stationary Markov sources.

An Ising channel per this definition is given by $\Sigma_{\mathcal{X}} = \Sigma_{\mathcal{S}} = \Sigma$ and the two following indecomposable matrices

$$C(0) = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \text{ and } C(1) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix},$$

obtained from the fact that the input process $\mathcal{X} = (X_t)_{t\in\mathbb{Z}^+}$ and the state process $\mathcal{S} = (S_t)_{t\in\mathbb{Z}^+}$ of an Ising channel satisfy $S_t = X_{t-1}$ for every $t \in \mathbb{Z}^+$. Additionally, since for any $x_i \in \Sigma$ and $i \in \langle 1, t\rangle$, one has $\prod_{i\in\langle 1,t\rangle} C(x_i) = C(x_t)$, an Ising channel is indecomposable, concluding the proof of the observation. □

Proof of Theorem 2.1: By the chain rule, we have

$$\begin{aligned}\frac{1}{n} I(X_1^n; Y_1^n) &= \frac{1}{n} H(Y_1^n) - \frac{1}{n} H(Y_1^n \mid X_1^n) \\ &\geq \frac{1}{n} H(Y_1^n) - \frac{1}{n} \sum_{i\in\langle 2,n\rangle} H(Y_i \mid Y_1^{i-1} X_1^n) - \frac{1}{n}.\end{aligned}$$

By definition, Y_i is independent of Y_1^{i-1} and $X_1^{i-2} X_{i+1}^n$ given X_i and X_{i-1}, for any $i \in \langle 2, n\rangle$; therefore

$$\begin{aligned}\frac{1}{n} I(X_1^n; Y_1^n) &\geq \frac{1}{n} H(Y_1^n) - \frac{1}{n} \sum_{i\in\langle 2,n\rangle} H(Y_i \mid X_i X_{i-1}) - \frac{1}{n} \\ &= \frac{1}{n} H(Y_1^n) - \frac{1}{n} \mathrm{H}(p) \sum_{i\in\langle 2,n\rangle} \mathrm{Prob}(X_i \neq X_{i-1}) - \frac{1}{n}. \quad (38)\end{aligned}$$

Now, due to Remark A.1, we can assume that $\mathcal{X}$ is distributed according to the distribution of a (time-invariant) stationary Markov chain, implying that

$$\frac{1}{n} I(X_1^n; Y_1^n) \geq \frac{1}{n} H(Y_1^n) - \frac{n-1}{n} \mathrm{H}(p) \cdot \mathrm{Prob}(X_2 \neq X_1) - \frac{1}{n}. \quad (39)$$

We proceed by bounding the quantity $\frac{1}{n}H(Y_1^n)$ from below. For $\zeta \in \mathbb{Z}^+$, by the chain rule, one has

$$\frac{1}{n}H(Y_1^n) \geq \frac{1}{n} \sum_{i\in\langle\zeta+1,n\rangle} H(Y_i \mid Y_1^{i-1}) \ .$$

Since conditioning does not increase entropy,

$$\frac{1}{n}H(Y_1^n) \geq \frac{1}{n} \sum_{i\in\langle\zeta+1,n\rangle} H(Y_i \mid Y_1^{i-1}X_{i-\zeta}),$$

but Y_i does not depend on $Y_1^{i-\zeta}$ given $X_{i-\zeta}$, hence

$$\begin{aligned}\frac{1}{n}H(Y_1^n) &\geq \frac{1}{n} \sum_{i\in\langle\zeta+1,n\rangle} H(Y_i \mid Y_{i-\zeta+1}^{i-1}X_{i-\zeta}) \\ &= \frac{n-\zeta}{n}H(Y_{\zeta+1} \mid Y_2^{\zeta}X_1) \ . \end{aligned} \tag{40}$$

From (2), (39), and (40) we obtain the lower bound $\varrho(\mu,\zeta,p)$ on $\mathsf{cap}(p)$, as in (4). Thanks to Remark A.1, the expressions $H(Y_{\zeta+1} \mid Y_2^{\zeta}X_1)$ and $\mathrm{Prob}(X_2 \neq X_1)$ in (4) can now be computed assuming $\mathcal{X}$ is distributed according to a stationary Markov chain P_μ of order μ over Σ. Note that the difference between the normalized mutual information $\frac{1}{n}I(X_1^n;Y_1^n)$ and the lower bound $\varrho(\mu,\zeta,p)$ is $O(\frac{\zeta}{n})$, that is why the lower bound $\varrho(\mu,\zeta,p)$ gets tight as $\mu,\zeta\to\infty$ and $\frac{\zeta}{n}\to 0$, namely, $\mathsf{cap}(p) = \lim_{\mu,\zeta\to\infty,\frac{\zeta}{n}\to 0}\varrho(\mu,\zeta,p)$. □

Proof of Theorem 2.2: We start by establishing an upper bound on $\frac{1}{n}I(X_1^n;Y_1^n)$. By the chain rule, we have

$$\begin{aligned}\frac{1}{n}I(X_1^n;Y_1^n) &= \frac{1}{n}H(Y_1^n) - \frac{1}{n}H(Y_1^n \mid X_1^n) \\ &\leq \frac{1}{n}H(Y_1^n) - \frac{1}{n}\sum_{i\in\langle 2,n\rangle} H(Y_i \mid Y_1^{i-1}X_1^n) \ .\end{aligned}$$

Recall that Y_i is independent of Y_1^{i-1} and $X_1^{i-2}X_{i+1}^n$ given X_i and X_{i-1}, for any $i \in \langle 2,n\rangle$; therefore

$$\begin{aligned}\frac{1}{n}I(X_1^n;Y_1^n) &\leq \frac{1}{n}H(Y_1^n) - \frac{1}{n}\sum_{i\in\langle 2,n\rangle} H(Y_i \mid X_iX_{i-1}) \\ &= \frac{1}{n}H(Y_1^n) - \frac{1}{n}\mathrm{H}(p)\sum_{i\in\langle 2,n\rangle} \mathrm{Prob}(X_i \neq X_{i-1}) \ .\end{aligned}$$

For each $n \in \mathbb{Z}^+$, let $\mathsf{pd}(p,n)$ be a probability distribution of X_1^n, such that $X_1^n \sim \mathsf{pd}(p,n)$ implies that $\frac{1}{n}I(X_1^n;Y_1^n) \to \mathsf{cap}(p)$ as $n\to\infty$. For a word $\boldsymbol{x} \in \Sigma^n$, let $\mathsf{n}_0(\boldsymbol{x})$ denote the number of runs in $\boldsymbol{x}$, let $\mathsf{n}_i(\boldsymbol{x})$ denote the number of runs of length i in $\boldsymbol{x}$, for $i \in \langle 1,3\rangle$, and let $\mathsf{n}_{1;i}(\boldsymbol{x})$ denote the number of runs of length i preceded by runs of length 1, for $i \in \{1,2\}$. Also, let $\mathsf{b}_0 \in [0,1]$ be the number of runs per symbol, let $\mathsf{b}_i \in [0,\frac{1}{i}]$ be the number of runs of length i per symbol, for each $i \in \langle 1,3\rangle$, and let $\mathsf{b}_{1;i}$ be the number of runs of length i preceded by runs of length 1 per symbol, for both $i \in \{1,2\}$, in a typical word $\boldsymbol{x} \in \Sigma^n$ of the probability distribution $\mathsf{pd}(p,n)$.[8] In other words, for each $i \in \langle 0,3\rangle$, $\mathsf{n}_i(\boldsymbol{x})$ is within $(\mathsf{b}_i{\pm}\epsilon)n$, and for both $i \in \{1,2\}$, $\mathsf{n}_{1;i}(\boldsymbol{x})$ is within $(\mathsf{b}_{1;i} \pm \epsilon)n$ with probability at least $1{-}\epsilon$, for any arbitrarily small $\epsilon > 0$ and n large enough. To be valid numbers of runs per symbol, the b_i's and the $\mathsf{b}_{1;i}$'s have to satisfy (5). Notice that when $X_1^n \sim \mathsf{pd}(p,n)$ and $n\to\infty$, one has

$$\frac{1}{n}\sum_{i\in\langle 2,n\rangle} \mathrm{Prob}(X_i \neq X_{i-1}) \longrightarrow \mathsf{b}_0 \ . \tag{41}$$

A run of length 1 at location e disappears when $Z_e^{e+1} = 10$, hence the expected number of disappearances of runs of length 1 is within

$$\mathrm{Prob}(Z_1^2{=}10)(\mathsf{b}_1{\pm}\epsilon) = p(1{-}p)(\mathsf{b}_1{\pm}\epsilon)n \ ,$$

and each such disappearance decreases $\mathsf{n}_0(\boldsymbol{x})$ by 2. Therefore, by a variant of the joint asymptotic equipartition property [9, Th. 8.6.1] for stationary and ergodic random processes (see Remark A.1), a word $\boldsymbol{x} \in \Sigma^n$, when transmitted through $\mathtt{Is}(p)$, will result in a word $\boldsymbol{y} \in \Sigma^n$ in which, with probability at least $1{-}\epsilon$, $\mathsf{n}_0(\boldsymbol{y})$ is within $(\mathsf{b}_0{\pm}\epsilon{-}2p(1{-}p)(\mathsf{b}_1{\mp}\epsilon))n$. A run of length $i \in \{1,2\}$ ending at location e is created in $\boldsymbol{y}$ by one of the following four disjoint events at $\boldsymbol{x}$:

(i) a run of length i ends at location $e{-}1$ and $Z_{e-i} = Z_e = 1$;

(ii) a run of length $i{+}1$ ends at location e, and $Z_{e-i} = 1$, $Z_{e+1} = 0$;

(iii) a run of length i ends at location e, $Z_{e-i+1} = Z_{e+1} = 0$, yet it is not preceded by a run of length 1 or $Z_{e-i} = 0$ (otherwise, this run is of length $i{+}2$ at least);

(iv) for $i = 2$, a run of length 1 ends at location $e{-}1$, and $Z_{e-1} = 0$, $Z_e = 1$, yet it is not preceded by a run of length 1 or $Z_{e-2} = 0$ (otherwise, this run is of length 3 at least).

These considerations imply that with probability $\geq 1{-}\epsilon$, $\mathsf{n}_1(\boldsymbol{y})/n$ is within

$$p^2(\mathsf{b}_1 \pm \epsilon) + p(1{-}p)(\mathsf{b}_2 \pm \epsilon) + (\mathsf{b}_1 \pm \epsilon{-}p(\mathsf{b}_{1;1} \mp \epsilon))(1{-}p)^2$$

and, with probability $\geq 1{-}\epsilon$, $\mathsf{n}_2(\boldsymbol{y})/n$ is within

$$\begin{aligned}&p^2(\mathsf{b}_2 \pm \epsilon) + p(1{-}p)(\mathsf{b}_3{+}\mathsf{b}_1 \pm 2\epsilon - p(\mathsf{b}_{1;1} \mp \epsilon)) + \\ &(1{-}p)^2(\mathsf{b}_2 \pm \epsilon{-}p(\mathsf{b}_{1;2} \mp \epsilon)) \ .\end{aligned}$$

As $n\to\infty$ and $\epsilon\to 0$, one has $\mathsf{n}_i(\boldsymbol{y})/n \to \beta_i$, where the β_i's for all $i \in \langle 0,2\rangle$ are given by (6). The exponential growth rate of a set of words with β_0 runs per symbol and β_i runs of length i per symbol for $i \in \{1,2\}$ is readily seen to be $\mathsf{t}(\beta_0,\beta_1,\beta_2)$, as defined in (7). Thus, we can claim that, when $X_1^n \sim \mathsf{pd}(p,n)$ as $n\to\infty$ and $\epsilon\to 0$, the quantity $\frac{1}{n}H(Y_1^n)$ is bounded from above by $\mathsf{t}(\beta_0,\beta_1,\beta_2)$. This observation, alongside (41), yields the following upper bound on $\frac{1}{n}I(X_1^n;Y_1^n)$ for large values of n:

$$\frac{1}{n}I(X_1^n;Y_1^n) \leq \mathsf{t}(\beta_0,\beta_1,\beta_2) - \mathsf{b}_0 \cdot \mathrm{H}(p) + o(1) \ ,$$

where $o(1)$ goes to 0 as $n\to\infty$, which leads to the upper bound $\rho(p)$ on $\mathsf{cap}(p)$, as in (8). □

Remark A.2: One natural way to improve the upper bound $\rho(p)$ in Theorem 2.2 is by additionally taking into consideration numbers of runs of lengths $\mathsf{n}_i(\boldsymbol{x})$ for $i > 3$. However as i increases, the formulas for β_i's quickly become unwieldy, since, for $i \geq 3$, we also need to take into account the

[8] As $n\to\infty$, there is no guarantee that $\mathsf{n}_0(\boldsymbol{x})/n$, $\mathsf{n}_i(\boldsymbol{x})/n$ for $i \in \langle 1,3\rangle$, and $\mathsf{n}_{1;i}(\boldsymbol{x})/n$ for $i \in \{1,2\}$, in a word $\boldsymbol{x} \in \Sigma^n$, typical of the distribution $\mathsf{pd}(p,n)$, converge. However, by the well-known Bolzano–Weierstrass theorem, since all six such sequences of per-symbol numbers of runs are bounded from above (by, say, 1), each such sequence has a convergent subsequence. Taking a convergent subsequence of the sequence $\mathsf{n}_0(\boldsymbol{x})/n$ and applying the same Bolzano–Weierstrass type of argument to this subsequence, we find a subsubsequence on which both $\mathsf{n}_0(\boldsymbol{x})/n$ and $\mathsf{n}_1(\boldsymbol{x})$ converge. Continuing in the same vein for all six sequences of per-symbol number of runs yields a subsequence, convergent for all six of them. The quantities b_i for $i \in \langle 0,3\rangle$ and $\mathsf{b}_{1;i}$ for $i \in \{1,2\}$, in fact, represent the limits of their respective convergent subsequences.

possibility that a run of length at most $i-2$ in the input word $\boldsymbol{x}$ merges with the two preceding runs, when a preceding run of length 1 is eliminated, forming a run of length i in the output word $\boldsymbol{y}$.

Another improvement on $\rho(p)$ comes from bounding the quantity $\frac{1}{n}H(X_1^n)$ (which is an upper bound on $\frac{1}{n}I(X_1^n;Y_1^n)$) from above using the statistics on the number of runs per symbol in typical input words $\boldsymbol{x}$, i.e. using b_i's and $\mathsf{b}_{1;i}$'s. As a consequence, the maximand of the right-hand side of (8) can be replaced with a minimum between that upper bound on $\frac{1}{n}H(X_1^n)$ and $\mathsf{t}(\beta_0,\beta_1,\beta_2)-\mathsf{b}_0\cdot\mathrm{H}(p)$; however, it turns out that the improvement on $\rho(p)$ that this modification renders is insignificant. □

APPENDIX B.

By analogy to [10, Lemma 6], we will use the following technical lemma on the concavity of the supremized expression $\mathsf{z}(b,\gamma,\delta;\mathsf{h})$ to prove Proposition B.3 which comes next.

Lemma B.1: The expression $\mathsf{z}(b,\gamma,\delta;\mathsf{h})$, defined in (21), is concave in (b,γ,δ) for any concave function $\mathsf{h}:[0,1]\to\mathbb{R}$.

Proof: The functions $(b,\gamma,\delta)\mapsto\mathrm{H}(\upsilon(b,\gamma,\delta))$ and $(b,\gamma,\delta)\mapsto\mathrm{H}(p)(\delta+\gamma-1)$ are readily concave in (b,γ,δ), when $b\in[0,1]$, $\gamma\in[0,1-b]$, and $\delta\in[0,b]$. As the sum of concave functions is a concave function, it suffices to prove that the third and the fourth terms of $\mathsf{z}(b,\gamma,\delta;\mathsf{h})$ are concave in (b,γ,δ). Since these two proofs are similar, for brevity we will show the proof for the third term only.

Let $b_1,b_2\in[0,1]$, and let $\Gamma_i\in[0,1-b_i]$ and $\Delta_i\in[0,b_i]$ for $i\in\{1,2\}$. Set $\Phi_i=1-\upsilon(b_i,\Gamma_i,\Delta_i)$ for $i\in\{1,2\}$, then, for any $\Upsilon\in[0,1]$, due to the concavity of h, one has

$$\begin{aligned}&\frac{\Upsilon\Phi_1}{\Upsilon\Phi_1+(1-\Upsilon)\Phi_2}\mathsf{h}\Big(\frac{p(1-b_1-\Gamma_1)}{\Phi_1}\Big)\\&+\frac{(1-\Upsilon)\Phi_2}{\Upsilon\Phi_1+(1-\Upsilon)\Phi_2}\mathsf{h}\Big(\frac{p(1-b_2-\Gamma_2)}{\Phi_2}\Big)\\&\le\mathsf{h}\Big(\frac{p(1-\Upsilon b_1-(1-\Upsilon)b_2-\Upsilon\Gamma_1-(1-\Upsilon)\Gamma_2)}{\Upsilon\Phi_1+(1-\Upsilon)\Phi_2}\Big).\end{aligned}$$

Multiplying both sides of the last inequality by

$$\Upsilon\Phi_1+(1-\Upsilon)\Phi_2$$

yields the concavity of $(1-\upsilon)\mathsf{h}\Big(\frac{p}{1-\upsilon}(1-b-\gamma)\Big)$. □

Remark B.2: The claim of Lemma B.1 is, in fact, a particular case of a more general phenomenon, specifically, the concavity of the *perspective* of a multivariate real concave function [7, Sec. 3.2.6]. □

Proposition B.3: Let $p\in[0,p_0]$. The constant $\Psi^\star$ and function $\mathsf{h}^\star$, defined in (23) and (24), respectively, solve (22); moreover, the expression $\mathsf{z}(b,\gamma,\delta;\mathsf{h}^\star)$ is supremized for

$$\gamma^\star(b)=\begin{cases}\frac{p}{1-p}(b-b^\star)+1-b^\star & b\in[0,b^\star]\\ 1-b & b\in[b^\star,1]\end{cases},\tag{42}$$

$$\delta^\star(b)=\begin{cases}b & b\in[0,1-b^\star]\\ \frac{p}{1-p}(1-b^\star-b)+1-b^\star & b\in[1-b^\star,1]\end{cases}.\tag{43}$$

Proof: Denote by z_γ and z_δ the partial derivatives $\frac{\partial\mathsf{z}}{\partial\gamma}$ and $\frac{\partial\mathsf{z}}{\partial\delta}$, respectively, of $\mathsf{z}(b,\gamma,\delta;\mathsf{h}^\star)$. Due to Lemma B.1 (applied with $\mathsf{h}=\mathsf{h}^\star$ and fixed b) and the symmetry of the problem with respect to $b=0.5$, it suffices to prove the following Karush–Kuhn–Tucker conditions [7, Sec. 5.5.3] to show that $\mathsf{z}(b,\gamma,\delta;\mathsf{h}^\star)$ is supremized at $\gamma^\star$ and $\delta^\star$, as defined in (42), (43):

- For $b\in[0,b^\star]$, $\mathsf{z}_\gamma(b,\gamma^\star,\delta^\star;\mathsf{h}^\star)=0$, $\mathsf{z}_\delta(b,\gamma^\star,\delta^\star;\mathsf{h}^\star)>0$,
- For $b\in(b^\star,0.5]$, $\mathsf{z}_\gamma(b,\gamma^\star,\delta^\star;\mathsf{h}^\star),\mathsf{z}_\delta(b,\gamma^\star,\delta^\star;\mathsf{h}^\star)>0$.

Set $\mathsf{h}_b^\star=\frac{\partial\mathsf{h}^\star}{\partial b}$, then

$$\begin{aligned}\mathsf{z}_\gamma&=(1-p)\log_2\frac{\upsilon}{1-\upsilon}+\mathrm{H}(p)\\&-(1-p)\mathsf{h}^\star\left[\frac{1}{\upsilon}(\delta+(1-p)(1-b-\gamma))\right]\\&+\frac{p(1-p)(\delta-b)}{\upsilon}\mathsf{h}_b^\star\left[\frac{1}{\upsilon}(\delta+(1-p)(1-b-\gamma))\right]\\&+(1-p)\mathsf{h}^\star\left[\frac{p}{1-\upsilon}(1-b-\gamma)\right]\\&+\frac{p(pb-1+(1-p)\delta)}{1-\upsilon}\mathsf{h}_b^\star\left[\frac{p}{1-\upsilon}(1-b-\gamma)\right],\end{aligned}$$

and

$$\begin{aligned}\mathsf{z}_\delta&=(1-p)\log_2\frac{1-\upsilon}{\upsilon}+\mathrm{H}(p)\\&+(1-p)\mathsf{h}^\star\left[\frac{1}{\upsilon}(\delta+(1-p)(1-b-\gamma))\right]\\&+\frac{p(pb+(1-p)(1-\gamma))}{\upsilon}\mathsf{h}_b^\star\left[\frac{1}{\upsilon}(\delta+(1-p)(1-b-\gamma))\right]\\&-(1-p)\mathsf{h}^\star\left[\frac{p}{1-\upsilon}(1-b-\gamma)\right]\\&+\frac{p(1-p)(1-b-\gamma)}{1-\upsilon}\mathsf{h}_b^\star\left[\frac{p}{1-\upsilon}(1-b-\gamma)\right].\end{aligned}$$

For $\gamma=\gamma^\star$ and $\delta=\delta^\star$, the quantity

$$\frac{p}{1-\upsilon}(1-b-\gamma)=\frac{p(b^\star-b)}{(1-p)(1-b^\star)}$$

is between 0 and $\frac{pb^\star}{(1-p)(1-b^\star)}=\frac{p}{(1-p)2^{\mathrm{H}(p)}}$ as b ranges over $[0,b^\star]$. Yet for $p\in[0,p_0]$, we have $\frac{p}{(1-p)2^{\mathrm{H}(p)}}\le b^\star$, therefore for any $b\in[0,b^\star]$,

$$\begin{aligned}&\mathsf{z}_\gamma(b,\gamma^\star,\delta^\star;\mathsf{h}^\star)\\&=(1-p)\log_2\frac{b^\star}{1-b^\star}+\mathrm{H}(p)-(1-p)\Psi^\star\\&+(1-p)\mathsf{h}^\star\left[\frac{p(b^\star-b)}{(1-p)(1-b^\star)}\right]-\frac{p(1-b)}{1-b^\star}\mathrm{H}(p)\\&=p\,\mathrm{H}(p)-(1-p)\Psi^\star\\&+(1-p)\Big(\mathrm{H}(p)\frac{p(b^\star-b)}{(1-p)(1-b^\star)}+\Psi^\star\Big)-\frac{p(1-b)\,\mathrm{H}(p)}{1-b^\star}\\&=0,\end{aligned}$$

and

$$\begin{aligned}&\mathsf{z}_\delta(b,\gamma^\star,\delta^\star;\mathsf{h}^\star)\\&=(1-p)\log_2\frac{1-b^\star}{b^\star}+\mathrm{H}(p)+(1-p)\Psi^\star-p\,\mathrm{H}(p)\\&-(1-p)\mathsf{h}^\star\left[\frac{p(b^\star-b)}{(1-p)(1-b^\star)}\right]+\frac{p(b^\star-b)}{1-b^\star}\mathsf{h}_b^\star\left[\frac{p(b^\star-b)}{(1-p)(1-b^\star)}\right]\\&=(1-p)(2\,\mathrm{H}(p)+\Psi^\star)-(1-p)\Big(\frac{p\,\mathrm{H}(p)(b^\star-b)}{(1-p)(1-b^\star)}+\Psi^\star\Big)\\&+\mathrm{H}(p)\frac{p(b^\star-b)}{1-b^\star}\\&=2(1-p)\,\mathrm{H}(p)>0.\end{aligned}$$

For $b \in (b^\star, 0.5]$, we have

$$\mathsf{z}_\gamma(b,\gamma^\star,\delta^\star;\mathsf{h}^\star) = (1-p)\log_2\frac{b}{1-b} + \mathrm{H}(p) - (1-p)\Psi^\star + (1-p)\Psi^\star - p\,\mathrm{H}(p) = (1-p)\Big(\mathrm{H}(p) - \log_2\frac{1-b}{b}\Big) > 0,$$

and

$$\mathsf{z}_\delta(b,\gamma^\star,\delta^\star;\mathsf{h}^\star) = (1-p)\log_2\frac{1-b}{b} + \mathrm{H}(p) + (1-p)\Psi^\star - p\,\mathrm{H}(p) - (1-p)\Psi^\star = (1-p)\Big(\mathrm{H}(p) + \log_2\frac{1-b}{b}\Big) > 0\ .$$

Finally, to conclude that $\Psi^\star$ and $\mathsf{h}^\star$ indeed solve (22), it is left to verify that $\Psi^\star + \mathsf{h}^\star(b) = \mathsf{z}(b,\gamma^\star,\delta^\star;\mathsf{h}^\star)$. Readily, for $b \in [0, b^\star]$,

$$\begin{aligned}\mathsf{z}(b,\gamma^\star,\delta^\star;\mathsf{h}^\star) &= \mathrm{H}(b^\star) - \mathrm{H}(p)\frac{b^\star-b}{1-p} + b^\star\Psi^\star + (1-b^\star)\mathsf{h}^\star\left[\frac{p(b^\star-b)}{(1-p)(1-b^\star)}\right] \\ &= \Psi^\star + \mathrm{H}(p)b + \mathrm{H}(b^\star) - \mathrm{H}(p)b^\star \\ &= \Psi^\star + \mathsf{h}^\star(b),\end{aligned}$$

and for $b \in [b^\star, 0.5]$,

$$\mathsf{z}(b,\gamma^\star,\delta^\star;\mathsf{h}^\star) = \mathrm{H}(b) + b\Psi^\star + (1-b)\Psi^\star = \mathrm{H}(b) + \Psi^\star = \Psi^\star + \mathsf{h}^\star(b).$$

□

Remark B.4: As follows from the proof of Proposition B.3, the largest value p_0 of p, for which the proposition applies, is the unique solution of the equality $\frac{p}{1-p} = \frac{2^{\mathrm{H}(p)}}{2^{\mathrm{H}(p)}+1}$ in the interval $[0,1]$. □

APPENDIX C.

We start off by proving an auxiliary lemma.

Lemma C.1: Let $p \in [0,1] \setminus \{0.5\}$ and let U_p be a domain of values of $\mathsf{x} \in [0,1]$ satisfying the following system of inequalities:

$$0 \le \frac{p - \mathsf{x} + \mathsf{x}^2 - p\mathsf{x}^2}{(p-\mathsf{x})(1-p) + p^2\mathsf{x}^2} \le 1 \tag{44}$$

and

$$0 \le \frac{1-p-\mathsf{x}+p\mathsf{x}^2}{(1-p)^2 - p\mathsf{x} + p(1-p)\mathsf{x}^2} \le 1\ . \tag{45}$$

Then

$$U_p = \begin{cases} [\frac{p}{1-p}, 1] & \text{if } p \in [0, 0.5) \\ [\frac{1-p}{p}, 1] & \text{if } p \in (0.5, p_1] \\ [\frac{p(1-\sqrt{1-4(1-p-p^2)})}{2(1-p-p^2)}, 1] & \text{if } p \in (p_1, 1] \end{cases}, \tag{46}$$

where $p_1 \approx 0.618034$ is the solution of the equation $p^2 = 1-p$ on the interval $[0,1]$.

Proof: The numerator of the fraction in (45) is non-negative for any $\mathsf{x} \in [0,1]$ and any $p \in [0, 0.5)$. Therefore, the double inequality in (45) holds if and only if

$$1-p-\mathsf{x}+p\mathsf{x}^2 \le (1-p)^2 - p\mathsf{x} + p(1-p)\mathsf{x}^2,$$

which is equivalent to

$$(p-\mathsf{x})(1-p) + p^2\mathsf{x}^2 \le 0. \tag{47}$$

In other words, the denominator of the fraction in (44) is non-positive. In light of that, in order for the double inequality in (44) to hold also, in addition to inequality (47) above, the numerator of (44) must be non-positive and at least the denominator of (44), i.e.,

$$p - \mathsf{x} + \mathsf{x}^2 - p\mathsf{x}^2 \le 0. \tag{48}$$

$$(p-\mathsf{x})(1-p) + p^2\mathsf{x}^2 \le p - \mathsf{x} + \mathsf{x}^2 - p\mathsf{x}^2. \tag{49}$$

For $p \in [0, 0.5)$, the inequality (47) holds for any $\mathsf{x} \in [\frac{1-p-\sqrt{(1-p)(1-p-4p^3)}}{2p^2}, 1]$ and the inequality (48) holds for any $\mathsf{x} \in [\frac{p}{1-p}, 1]$, whereas the inequality (49) holds for any $\mathsf{x} \in [0,1]$.

Altogether, U_p is the intersection of the interval

$$\left[\frac{1}{2p^2}(1 - p - \sqrt{(1-p)(1-p-4p^3)}), 1\right]$$

with $[\frac{p}{1-p}, 1]$ and $[0,1]$. One can verify that

$$\frac{1-p-\sqrt{(1-p)(1-p-4p^3)}}{2p^2} < \frac{p}{1-p} \Leftrightarrow 4p^4(2p-1) < 0,$$

which holds for any $p \in [0, 0.5)$, therefore $U_p = [\frac{p}{1-p}, 1]$ for any $p \in [0, 0.5)$, proving the first case of (46).

The denominator of (44) is positive for any $p \in (0.5, 1]$, hence in that interval (44) is equivalent to satisfying the following two inequalities:

$$p-\mathsf{x}+(1-p)\mathsf{x}^2 \ge 0 \text{ and } p-\mathsf{x}+(1-p)\mathsf{x}^2 \le (p-\mathsf{x})(1-p) + p^2\mathsf{x}^2,$$

which, in turn, is equivalent to satisfying both

$$\left(\mathsf{x} \le 1 \text{ or } \mathsf{x} \ge \frac{p}{1-p}\right) \text{ and }$$

$$\left|2(1-p-p^2)\mathsf{x} - p\right| \le \sqrt{1-4(1-p-p^2)}\ . \tag{50}$$

Since $\mathsf{x} \in [0,1]$, the inequalities (50) simplify to

$$\mathsf{x} \in \left[\frac{p}{2(1-p-p^2)}(1 - \sqrt{1-4(1-p-p^2)}), 1\right]\ . \tag{51}$$

Similarly, for $p \in (0.5, 1]$, the numerator of (45) is at least the denominator of (45), and, so, both must be non-positive, i.e.,

$$1-p-\mathsf{x}+p\mathsf{x}^2 \le 0 \text{ and } (1-p)^2 - p\mathsf{x} + p(1-p)\mathsf{x}^2 \le 0\ ,$$

that is,

$$\mathsf{x} \in \left[\frac{1-p}{p}, 1\right] \text{ and } |2p(1-p)\mathsf{x} - p| \le \sqrt{p^2 - 4p(1-p)^3}\ . \tag{52}$$

Since $\frac{1-p}{p} \ge \frac{1}{2p(1-p)}(p - \sqrt{p^2-4p(1-p)^3})$ for any $p \in [0,1]$, the inequalities in (52) together are equivalent to $\mathsf{x} \in [\frac{1-p}{p}, 1]$. Intersecting the interval $[\frac{1-p}{p}, 1]$ with (51) when $p \in (0.5, 1]$ results in the remaining two cases of (46). □

Next, we proceed with the main proof that $\mathsf{cap}_{\mathsf{FB}}(p) = \varrho_2(p)$ on the interval $(p_0, 0.5)$ by citing a specialized variant of the main result of [23], preceded by several related definitions.

Let Q be the directed labeled graph in Figure 10, whose set of states is $\mathcal{Q} = \{v_0, v_1, v_2, v_3\}$. The labels on its edges correspond to the binary output alphabet of $\mathtt{Is}_{\mathsf{FB}}(p)$, whereas the states v_0 through v_3, respectively, categorize the last run of an output sequence into one of the following four types: runs of 0's of even length; runs of 0's of odd length; runs of 1's of even length; runs of 1's of odd length. Construct the following directed labeled graph G:

- The set of states of G is $\Sigma \times \mathcal{Q}$ where each state $(s, v) \in \Sigma \times \mathcal{Q}$ corresponds to a realization s of $\mathtt{Is}_{\mathsf{FB}}(p)$ and a category v of the output Y of $\mathtt{Is}_{\mathsf{FB}}(p)$.
- For any pair of states $(s, v), (s', v') \in \Sigma \times \mathcal{Q}$, include a directed edge $(s, v) \to (s', v')$ with label y in G if and only if the edge $v \to v'$ with label y is in Q and $y \in \{s, s'\}$.

Remark C.2: Notice that the graph G is strongly connected.

Let $\alpha_i \in [0,1]$ be real numbers for all $i \in \langle 0, 3\rangle$. Construct the Markov chain $P = P(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$ with G as its support graph by assigning the transition probabilities from one state of P to another as shown in Figure 11, where the α_i's for $i \in \langle 0, 3\rangle$ stand for probabilities of producing a given bit value (0 or 1) as the next input bit, conditioned on being in a given state of $\Sigma \times \mathcal{Q}$. Namely, for a random pair (S, Q) with realizations in $\Sigma \times \mathcal{Q}$,

- $\alpha_0 = \mathrm{Prob}(X = 1 \mid (S, Q) = (0, v_0))$;
- $\alpha_1 = \mathrm{Prob}(X = 0 \mid (S, Q) = (1, v_0))$;
- $\alpha_2 = \mathrm{Prob}(X = 1 \mid (S, Q) = (0, v_1))$;
- $\alpha_3 = \mathrm{Prob}(X = 0 \mid (S, Q) = (1, v_1))$.

Also, we employ the symmetry of generalized Ising channels to bit complementation by assuming right away (see Remark C.5 for an outline of a justification of this assumption) that

- $\alpha_0 = \mathrm{Prob}(X = 0 \mid (S, Q) = (1, v_2))$;
- $\alpha_1 = \mathrm{Prob}(X = 1 \mid (S, Q) = (0, v_2))$;
- $\alpha_2 = \mathrm{Prob}(X = 0 \mid (S, Q) = (1, v_3))$;
- $\alpha_3 = \mathrm{Prob}(X = 1 \mid (S, Q) = (0, v_3))$.

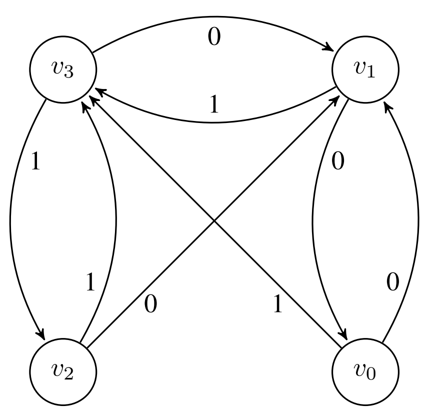

Fig. 10. Directed labeled graph Q.

Edges which are assigned zero transition probabilities are removed from G. Notice that for $(\alpha_i)_{i\in\langle 0,3\rangle} \in (0,1)^4$ and $p \in (0,1)$, the (modified) support graph of P is G which has only one strongly connected component, which, in turn, ensures the existence of a unique stationary distribution induced by P.

Define $\mathcal{M}$ to be the set of all Markov chains

$$P = P(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$$

for $(\alpha_i)_{i\in\langle 0,3\rangle}$ in the open hypercube $(0,1)^4$. For a state $(s, v) \in \Sigma \times \mathcal{Q}$ in a Markov chain $P \in \mathcal{M}$, let $\varsigma_{(s,v)}$ denote its (unique) stationary probability. Due to the symmetry of P to bit complementation, one has

$$\varsigma_{(0,v_0)} = \varsigma_{(1,v_2)},\ \varsigma_{(1,v_0)} = \varsigma_{(0,v_2)},$$
$$\varsigma_{(0,v_1)} = \varsigma_{(1,v_3)},\ \varsigma_{(1,v_1)} = \varsigma_{(0,v_3)},$$

which, in turn, facilitates the derivation of the stationary distribution. Specifically,

$$\begin{aligned}
\varsigma_{(0,v_0)} &= \frac{1}{2\ell}(1-\alpha_2 - (1-\alpha_2)\alpha_3 p + \alpha_1\alpha_2\alpha_3 p^2 - \alpha_1\alpha_2\alpha_3 p^3),\\
\varsigma_{(1,v_0)} &= \frac{1}{2\ell}(\alpha_2 p - (1+\alpha_0)\alpha_2\alpha_3 p^2 + \alpha_0\alpha_2\alpha_3 p^3),\\
\varsigma_{(0,v_1)} &= \frac{1}{2\ell}(1 - (1+\alpha_0)\alpha_3 p + \alpha_0\alpha_3 p^2),\\
\varsigma_{(1,v_1)} &= \frac{1}{2\ell}(\alpha_0(1-\alpha_2)p + \alpha_1\alpha_2 p^2),
\end{aligned}$$

where

$$\begin{aligned}
\ell &= \ell(\alpha_0, \alpha_1, \alpha_2, \alpha_3)\\
&\triangleq 2-\alpha_2 + (\alpha_0+\alpha_2-\alpha_0\alpha_2-2\alpha_3-\alpha_0\alpha_3+\alpha_2\alpha_3)p\\
&\quad + (\alpha_1\alpha_2+\alpha_0\alpha_3-\alpha_2\alpha_3-\alpha_0\alpha_2\alpha_3+\alpha_1\alpha_2\alpha_3)p^2\\
&\quad + (\alpha_0-\alpha_1)\alpha_2\alpha_3 p^3.
\end{aligned}$$

Denote by $\varsigma_{v_i} \triangleq \varsigma_{(0,v_i)}+\varsigma_{(1,v_i)}$ the stationary probability of having the last run of an output sequence from category v_i, for $i \in \langle 0, 3\rangle$. Let $\mathsf{H}_P(Y \mid Q) = \sum_{i\in\langle 0,3\rangle} \varsigma_{v_i} \mathsf{H}_P(Y \mid v_i)$ denote the conditional entropy of the output Y of the Markov chain P given the category of (the last run of) an output sequence. Again, due to the symmetry to bit complementation, we can write $\mathsf{H}_P(Y \mid Q) = 2\sum_{i\in\langle 0,1\rangle} \varsigma_{v_i} \mathsf{H}_P(Y \mid v_i)$. Define

$$\begin{aligned}
\mathsf{z} &= \mathsf{z}(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = \mathsf{z}_p(\alpha_0, \alpha_1, \alpha_2, \alpha_3)\\
&\triangleq 2 \cdot \Big[\sum_{i\in\langle 0,1\rangle} \varsigma_{v_i}\, \mathrm{H}\Big(\frac{\varsigma_{(0,v_i)}}{\varsigma_{v_i}}(1-\alpha_{2i}+\alpha_{2i}p)\\
&\qquad\qquad + \frac{\varsigma_{(1,v_i)}}{\varsigma_{v_i}}\alpha_{2i+1}(1-p)\Big)\\
&\qquad - (\alpha_0\varsigma_{(0,v_0)}+\alpha_1\varsigma_{(1,v_0)}+\alpha_2\varsigma_{(0,v_1)}+\alpha_3\varsigma_{(1,v_1)})\, \mathrm{H}(p)\Big],
\end{aligned}$$

and note that due to the continuity of $\mathsf{z}(\alpha_0, \alpha_1, \alpha_2, \alpha_3)$, we can extend the set $\mathcal{M}$ to its compact closure $\mathrm{cl}(\mathcal{M})$ while preserving the equality

$$\begin{aligned}
&\sup_{(\alpha_i)_{i\in\langle 0,3\rangle}\in(0,1)^4} \{\mathsf{z}(\alpha_0, \alpha_1, \alpha_2, \alpha_3)\}\\
&\qquad = \sup_{(\alpha_i)_{i\in\langle 0,3\rangle}\in[0,1]^4} \{\mathsf{z}(\alpha_0, \alpha_1, \alpha_2, \alpha_3)\}.
\end{aligned}$$

We obtain an upper bound $\rho_1(p)$ on $\mathsf{cap}_{\mathsf{FB}}(p)$ by specializing [23, Th. 1] to the case[9] of generalized Ising channels:

$$\begin{aligned}
\rho_1(p) &\triangleq \sup_{P\in\mathrm{cl}(\mathcal{M})} \{\mathsf{H}_P(Y \mid Q) - \mathrm{Prob}_P(X \neq S)\, \mathrm{H}(p)\}\\
&= \sup_{P\in\mathrm{cl}(\mathcal{M})} \Big\{ 2 \sum_{i\in\langle 0,1\rangle} \varsigma_{v_i} \mathsf{H}_P(Y \mid v_i)\\
&\qquad - 2(\alpha_0\varsigma_{(0,v_0)}+\alpha_1\varsigma_{(1,v_0)}+\alpha_2\varsigma_{(0,v_1)}+\alpha_3\varsigma_{(1,v_1)})\, \mathrm{H}(p) \Big\}\\
&= \sup_{(\alpha_i)_{i\in\langle 0,3\rangle}\in[0,1]^4} \{\mathsf{z}(\alpha_0, \alpha_1, \alpha_2, \alpha_3)\},
\end{aligned}$$

[9]To see that it is indeed a specialization, note that the set $\mathcal{P}_\pi$ and the expression $I(X, S; Y \mid Q)(= \mathsf{H}(Y \mid Q) - \mathsf{H}(Y \mid Q, X, S))$ in [23, Th. 1] are contained in $\mathrm{cl}(\mathcal{M})$ and equal to $\mathsf{H}_P(Y \mid Q) - \mathsf{H}_P(Y \mid Q, X, S)$, respectively, in our notation. Finally, $\mathsf{H}_P(Y \mid Q, X, S) = \mathrm{Prob}_P(X \neq S)\, \mathrm{H}(p)$.

	$(0,v_0)$	$(1,v_0)$	$(0,v_1)$	$(1,v_1)$	$(0,v_2)$	$(1,v_2)$	$(0,v_3)$	$(1,v_3)$
$(0,v_0)$	0	0	$1-\alpha_0$	$\alpha_0 p$	0	0	0	$\alpha_0(1-p)$
$(1,v_0)$	0	0	$\alpha_1(1-p)$	0	0	0	$\alpha_1 p$	$1-\alpha_1$
$(0,v_1)$	$1-\alpha_2$	$\alpha_2 p$	0	0	0	0	0	$\alpha_2(1-p)$
$(1,v_1)$	$\alpha_3(1-p)$	0	0	0	0	0	$\alpha_3 p$	$1-\alpha_3$
$(0,v_2)$	0	0	$1-\alpha_1$	$\alpha_1 p$	0	0	0	$\alpha_1(1-p)$
$(1,v_2)$	0	0	$\alpha_0(1-p)$	0	0	0	$\alpha_0 p$	$1-\alpha_0$
$(0,v_3)$	0	0	$1-\alpha_3$	$\alpha_3 p$	0	$\alpha_3(1-p)$	0	0
$(1,v_3)$	0	0	$\alpha_2(1-p)$	0	$\alpha_2 p$	$1-\alpha_2$	0	0

Fig. 11. Transition probabilities of the Markov chain P.

leading to the following theorem, which is our main tool for proving that $\mathsf{cap}_{\mathsf{FB}}(p) = \varrho_2(p)$ on the interval $(p_0, 0.5)$.

Theorem C.3: Let $p \in [0,1]$. Then

$$\mathsf{cap}_{\mathsf{FB}}(p) \le \sup_{(\alpha_i)_{i\in\langle 0,3\rangle}\in[0,1]^4} \{\mathsf{z}(\alpha_0,\alpha_1,\alpha_2,\alpha_3)\}\,. \quad (53)$$

□

Remark C.4: Notice that when $\alpha_0 = \alpha_1 = \alpha_3 = 0$, the right-hand side of (53) reduces to

$$\frac{1}{2-\alpha_2+\alpha_2 p}\Big((1-\alpha_2+\alpha_2 p)\,\mathrm{H}\left(\frac{\alpha_2 p}{1-\alpha_2+\alpha_2 p}\right) + \mathrm{H}(\alpha_2(1-p)) - \alpha_2\,\mathrm{H}(p)\Big),$$

which can be simplified further to $\frac{\mathrm{H}(\alpha_2)}{2-\alpha_2(1-p)}$, which, in turn, is identical to the expression for $\varrho_2(p)$ in (34) (with ξ replaced by α_2). □

Remark C.5: It can be verified that the function $\mathsf{H}_P(Y \mid Q)$ is concave if expressed in terms of the following 16 variables

$$\kappa_{x,s,v} = \mathrm{Prob}(X=x, S=s, Q=v), \quad x,s\in\Sigma, v\in\mathcal{Q},$$

which, in turn, satisfy certain linear equalities and inequalities that form a convex domain. By the symmetry of the channel to bit complementation it then follows that the function $\mathsf{H}_P(Y \mid Q)$ is maximized when $\kappa_{x,s,v_i} = \kappa_{\overline{x},\overline{s},v_{(i+2) \bmod 4}}$, which, in turn, implies the assumed symmetry to bit complementation of $\mathrm{Prob}(X=x \mid (S,Q)=(s,v))$ at the maximum, as well as of the implied stationary distribution, $\varsigma_{(s,v)}$, at that point. □

In view of Remark C.4, to prove that $\mathsf{cap}_{\mathsf{FB}}(p) = \varrho_2(p)$ on the interval $(p_0, 0.5)$, it suffices to show that on that interval the upper bound of (53) is attained for $\alpha_0 = \alpha_1 = \alpha_3 = 0$ and $\alpha = \xi$, where $\xi \in [0,1]$ maximizes the expression $\frac{\mathrm{H}(\xi)}{2-\xi(1-p)}$ (see Section IV-B). To demonstrate that this point is a global maximum of z in the hypercube $(\alpha_i)_{i\in\langle 0,3\rangle} \in [0,1]^4$, next we will show that for any p in the interval $(p_0, 0.5)$,

(M1) any global maximum of $\mathsf{z}(\alpha_0,\alpha_1,\alpha_2,\alpha_3)$ on the hypercube $(\alpha_i)_{i\in\langle 0,3\rangle} \in [0,1]^4$ occurs when $\alpha_1 = 0$;

(M2) any global maximum of $\mathsf{z}(\alpha_0,0,\alpha_2,\alpha_3)$ on the hypercube $(\alpha_i)_{i\in\{0,2,3\}} \in [0,1]^3$ occurs when $\alpha_3 = 0$;

(M3) by switching the variables $(\alpha_0,\alpha_2) \to (\mathsf{x},\mathsf{y})$, we can recast the problem of finding the global maximum of the function $\mathsf{z}(\alpha_0,0,\alpha_2,0)$ over $(\alpha_0,\alpha_2) \in [0,1]^2$ as a problem of finding the global maximum of a function $\widetilde{\mathsf{z}}(\mathsf{x},\mathsf{y}) = \widetilde{\mathsf{z}}_p(\mathsf{x},\mathsf{y})$ over a compact domain $U_p \subset [0,1]^2$ defined by certain inequalities in x and y; moreover, the preimage of the global maximum of $\widetilde{\mathsf{z}}(\mathsf{x},\mathsf{y})$ on $[0,1]^2$ will be outside of U_p and that maximum will be the only extremum of the function $\widetilde{\mathsf{z}}(\mathsf{x},\mathsf{y})$ on $[0,1]^2$.

Assertions (M1)–(M3) imply that a global maximum of $\widetilde{\mathsf{z}}(\mathsf{x},\mathsf{y})$ over U_p must occur on the boundary of U_p, which is equivalent to saying that $\max_{(\alpha_0,\alpha_2)\in[0,1]^2} \mathsf{z}(\alpha_0,0,\alpha_2,0)$ occurs at either

$$\big\{(\alpha_0,\alpha_2)\in[0,1]^2 : \alpha_0 = 0\big\}\,, \text{ or}$$
$$\big\{(\alpha_0,\alpha_2)\in[0,1]^2 : \alpha_0 = 1\big\}\,, \text{ or}$$
$$\big\{(\alpha_0,\alpha_2)\in[0,1]^2 : \alpha_2 = 0\big\}\,, \text{ or}$$
$$\big\{(\alpha_0,\alpha_2)\in[0,1]^2 : \alpha_2 = 1\big\}\ .$$

Finally, we will prove that

(M4) the following inequality holds

$$\varrho_2(p) = \max_{\alpha_2\in[0,1]} \mathsf{z}(0,0,\alpha_2,0) \ge \max\{\mathsf{f}_0(p), \mathsf{f}_1(p)\}\,,$$

where

$$\mathsf{f}_0(p) \triangleq \max_{\alpha_2\in[0,1]} \mathsf{z}(1,0,\alpha_2,0), \text{ and}$$

$$\mathsf{f}_1(p) \triangleq \max_{\alpha_0\in[0,1]} \mathsf{z}(\alpha_0,0,0,0)\ .$$

Since $\mathsf{z}(\alpha_0,0,1,0) \equiv 0$ and since the function $\mathsf{z}(0,0,\alpha_2,0)$ for $\alpha_2 \in [0,1]$ is maximized at $\alpha_2 = \xi$, assertions (M1)–(M4) will imply $\mathsf{cap}_{\mathsf{FB}}(p) = \varrho_2(p)$ on $p \in (p_0, 0.5)$.

Assertion (M1) can be shown in two steps. First, searching for a maximum numerically[10] one can establish that for any $p \in (p_0, 0.5)$,

$$\max_{(\alpha_0,\alpha_1,\alpha_2,\alpha_3,p)\in[0,1]\times[0.2,1]\times[0,1]^2\times(p_0,0.5)} \mathsf{z}(\alpha_0,\alpha_1,\alpha_2,\alpha_3) \le \varrho_2(0.5) \le \varrho_2(p)\,, \quad (54)$$

$$\max_{(\alpha_0,\alpha_1,\alpha_2,\alpha_3,p)\in[0,1]^2\times[0.7,1]\times[0,1]\times(p_0,0.5)} \mathsf{z}(\alpha_0,\alpha_1,\alpha_2,\alpha_3) < \varrho_2(0.5) - 0.017 \le \varrho_2(p)\,,$$

hence, for any $p \in (p_0, 0.5)$, a global maximum of z must be attained for some $\alpha_1 \in [0, 0.2]$ and for some $\alpha_2 \in [0, 0.7]$.

[10] The first inequality in (54) is satisfied with equality at $\alpha_0 = \alpha_0^* \approx 0.758052$, $\alpha_2 = \alpha_3 = 0$, $p = 0.5$ and any $\alpha_1 \in [0.2, 1]$. The two-dimensional Taylor series of z in the variables α_2 and α_3 at $\alpha_2 = \alpha_3 = 0$ with $\alpha_0 = \alpha_0^*$ and $p = 0.5$ is $\varrho_2(0.5) + (\kappa_1\alpha_1)\cdot\alpha_2 + \kappa_2\cdot\alpha_3 + (\kappa_3 - \kappa_4\alpha_1)\cdot\alpha_2\cdot\alpha_3 + O(\alpha_2^2) + O(\alpha_3^2)$, where $(\kappa_i)_{i\in\langle 1,4\rangle} \approx (-0.149691, -0.15932, 0.100453, 0.0599745)$. In other words, the series is at most $\varrho_2(0.5) + (\kappa_1\alpha_1)\cdot\alpha_2 + (\kappa_2+\kappa_3)\cdot\alpha_3 + O(\alpha_2^2) + O(\alpha_3^2)$.

Secondly, again using numerical[11] software,

$$\max_{(\alpha_0,\alpha_1,\alpha_2,\alpha_3,p)\in[0,1]\times[0,0.2]\times[0,0.7]\times[0,1]\times(p_0,0.5)} \frac{\partial \mathsf{z}}{\partial \alpha_1} \leq 0 \ , \tag{55}$$

which implies (M1).

Assertion (M2) is now proved by establishing numerically[12] that for any $p \in (p_0, 0.5)$

$$\max_{(\alpha_0,\alpha_2,\alpha_3,p)\in[0,1]\times[0,0.7]\times[0,1]\times(p_0,0.5)} \frac{\partial \mathsf{z}(\alpha_0, 0, \alpha_2, \alpha_3)}{\partial \alpha_3} \leq 0 \ . \tag{56}$$

To prove assertion (M3), write $\widetilde{\mathsf{z}}(\mathsf{x},\mathsf{y}) = \mathsf{z}(\alpha_0, 0, \alpha_2, 0)$, where

$$\mathsf{x} = \frac{(1-\alpha_2)(1-\alpha_0+\alpha_0 p)}{1-\alpha_2+\alpha_2 p} \text{ and } \mathsf{y} = \frac{1-\alpha_2+\alpha_2 p}{1+\alpha_0(1-\alpha_2)p} \ .$$

This results in recasting the maximization problem as

$$\begin{aligned} &\max_{(\mathsf{x},\mathsf{y})\in U_p} \{\widetilde{\mathsf{z}}(\mathsf{x},\mathsf{y})\} \\ &= \max_{(\mathsf{x},\mathsf{y})\in U_p} \left\{ \frac{1}{\mathsf{y}+1} \left(\mathsf{y}\,\mathrm{H}(\mathsf{x}) + \mathrm{H}(\mathsf{y}) - (1-\mathsf{x}\mathsf{y})\,\mathrm{H}(p) \right) \right\} \end{aligned}$$

where U_p consists of all pairs $(\mathsf{x},\mathsf{y}) \in [0,1]^2$ such that

$$0 \leq \frac{p - \mathsf{y} + \mathsf{x}\mathsf{y} - p\mathsf{x}\mathsf{y}}{p - p^2 - \mathsf{y} + p\mathsf{y} + p^2\mathsf{x}\mathsf{y}} \leq 1, \text{ and} \tag{57}$$

$$0 \leq \frac{p - 1 + \mathsf{y} - p\mathsf{x}\mathsf{y}}{-(1-p)^2 + p\mathsf{y} - p\mathsf{x}\mathsf{y} + p^2\mathsf{x}\mathsf{y}} \leq 1 \ . \tag{58}$$

One can check that the only extremum of $\widetilde{\mathsf{z}}(\mathsf{x},\mathsf{y})$ on $[0,1]^2$ occurs at $\mathsf{x} = \mathsf{y} = \frac{2^{\mathrm{H}(p)}}{2^{\mathrm{H}(p)}+1}$ and that it is a (global) maximum. Substituting $\mathsf{x} = \mathsf{y}$ into (57) and (58) yields the domain U_p given by inequalities (44) and (45). By Lemma C.1 we get that $U_p = [\frac{p}{1-p}, 1]$ for $p \in (p_0, 0.5)$. Since, for such values of p,

$$\frac{2^{\mathrm{H}(p)}}{2^{\mathrm{H}(p)}+1} < \frac{p}{1-p}$$

(equality holds for $p = p_0$, by the definition of p_0), we get that $\frac{2^{\mathrm{H}(p)}}{2^{\mathrm{H}(p)}+1} \notin U_p$, thereby concluding the proof of (M3).

Finally, to prove assertion (M4), we first demonstrate that

$$\varrho_2(p) = \max_{\alpha_2\in[0,1]} \mathsf{z}(0,0,\alpha_2,0) \geq \mathsf{f}_0(p) \ .$$

By Jensen's inequality,

$$\begin{aligned} \mathsf{z}(1,0,\alpha_2,0) &= \frac{(1-\alpha_2+\alpha_2 p)\,\mathrm{H}(\frac{(1-\alpha_2)p}{1-\alpha_2+\alpha_2 p})}{2-\alpha_2+p} \\ &\quad + \frac{(1+(1-\alpha_2)p)\,\mathrm{H}(\frac{1-\alpha_2+\alpha_2 p}{1+(1-\alpha_2)p})}{2-\alpha_2+p} - \frac{\mathrm{H}(p)}{2-\alpha_2+p} \\ &\leq \mathrm{H}\left(\frac{1-\alpha_2+p}{2-\alpha_2+p}\right) - \frac{\mathrm{H}(p)}{2-\alpha_2+p} \\ &= \mathrm{H}\left(\frac{1}{2-\alpha_2+p}\right) - \frac{\mathrm{H}(p)}{2-\alpha_2+p}. \end{aligned}$$

[11]The inequality (55) is satisfied with equality at $\alpha_2 = 0$ and any $(\alpha_0, \alpha_1, \alpha_3, p)$ in the domain $U = [0,1] \times [0,0.2] \times [0,1] \times (p_0, 0.5)$. The coefficient of α_2 of the Taylor series of $\partial\mathsf{z}/\partial\alpha_1$ in α_2 at $\alpha_2 = 0$ attains the maximum of $\kappa_2^* \approx -0.0150434$ over the domain U, i.e., the Taylor series is at most $\kappa_2^* \cdot \alpha_2 + O(\alpha_2^2)$.

[12]The inequality (56) is satisfied with equality at $\alpha_0 = 0$ and any (α_2, α_3, p) in the domain $U = [0,0.7] \times [0,1] \times (p_0, 0.5)$. The coefficient of α_0 of the Taylor series of $\partial\mathsf{z}(\alpha_0,0,\alpha_2,\alpha_3)/\partial\alpha_3$ in α_0 at $\alpha_0 = 0$ attains the maximum of $\kappa_0^* \approx -0.0557648$ over the domain U, i.e., the Taylor series is at most $\kappa_0^* \cdot \alpha_0 + O(\alpha_0^2)$.

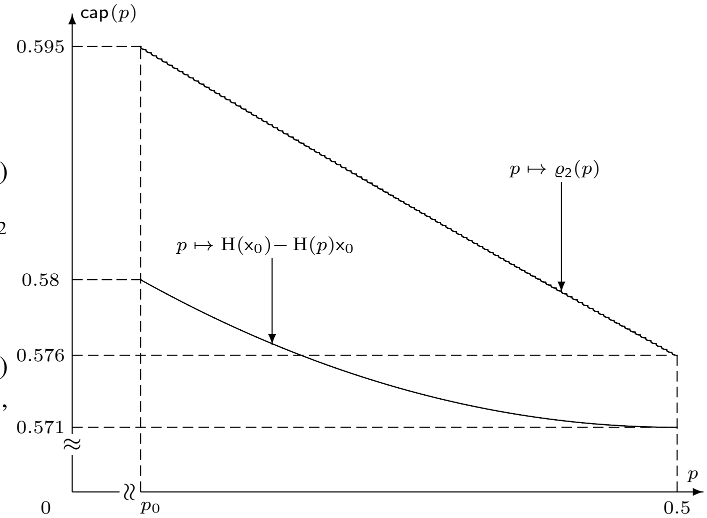

Fig. 12. Functions $p \mapsto \varrho_2(p)$ and $p \mapsto \mathrm{H}(\mathsf{x}_0) - \mathrm{H}(p)\mathsf{x}_0$.

The function $x \mapsto \mathrm{H}(x) - \mathrm{H}(p)x$ is concave in x and attains its global maximum at $x = \frac{1}{1+2^{\mathrm{H}(p)}}$. Since $1+p-2^{\mathrm{H}(p)} \leq -0.5$ on the interval $p \in [0, 0.5]$, one has

$$\frac{1}{1+2^{\mathrm{H}(p)}} < \mathsf{x}_0 \triangleq \frac{1}{0.5+2^{\mathrm{H}(p)}} \leq \frac{1}{2-\alpha_2+p}$$

for any $\alpha_2 \in [0,1]$ and $p \in (p_0, 0.5]$. Since the function $x \mapsto \mathrm{H}(x) - \mathrm{H}(p)x$ is decreasing on the interval $[\frac{1}{1+2^{\mathrm{H}(p)}}, 1]$, we have, for all $p \in (p_0, 0.5)$,

$$\mathsf{f}_0(p) \leq \mathrm{H}(\mathsf{x}_0) - \mathrm{H}(p)\mathsf{x}_0 \ . \tag{59}$$

One can observe in Figure 12 that the right-hand side of (59) is strictly below $\varrho_2(p)$ on the interval $(p_0, 0.5)$.

We conclude the proof of assertion (M4) by considering the function

$$\begin{aligned} \mathsf{z}(\alpha_0, 0, 0, 0) = &\frac{\mathrm{H}(1-\alpha_0+\alpha_0 p)}{2+\alpha_0 p} + \frac{(1+\alpha_0 p)\,\mathrm{H}(\frac{1}{1+\alpha_0 p})}{2+\alpha_0 p} \\ &- \frac{\mathrm{H}(p)\alpha_0}{2+\alpha_0 p} \ . \end{aligned}$$

Again, by Jensen's inequality,

$$\begin{aligned} \mathsf{z}(\alpha_0, 0, 0, 0) &\leq \mathrm{H}\left(\frac{2-\alpha_0+\alpha_0 p}{2+\alpha_0 p}\right) - \frac{\mathrm{H}(p)\alpha_0}{2+\alpha_0 p} \\ &= \mathrm{H}\left(\frac{\alpha_0}{2+\alpha_0 p}\right) - \frac{\mathrm{H}(p)\alpha_0}{2+\alpha_0 p} \ . \end{aligned}$$

The function $\alpha_0 \mapsto \frac{\alpha_0}{2+\alpha_0 p}$ is increasing on the interval $[0,1]$, hence for any $\alpha_0 \in [0,1]$, we have $\frac{\alpha_0}{2+\alpha_0 p} \leq \frac{1}{2+p}$. Moreover, since $1+p-2^{\mathrm{H}(p)} \leq -0.5$, one additionally has $\frac{1}{2+p} \leq \mathsf{x}_0$. Just like in the previous case, this implies, for all $p \in (p_0, 0.5)$,

$$\mathsf{f}_1(p) \leq \mathrm{H}(\mathsf{x}_0) - \mathrm{H}(p)\mathsf{x}_0,$$

which, like in the discussion following (59), implies $\mathsf{f}_1(p) < \varrho_2(p)$, for all $\alpha_0 \in [0,1]$ and all $p \in (p_0, 0.5)$.

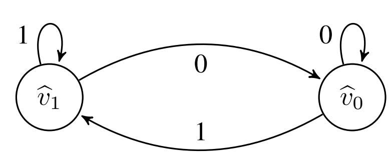

Fig. 13. Directed labeled graph $\widehat{\mathsf{Q}}$.

$$\left(\begin{array}{c|cccc} & (0,\widehat{v}_0) & (1,\widehat{v}_0) & (0,\widehat{v}_1) & (1,\widehat{v}_1) \\ \hline (0,\widehat{v}_0) & 1-\alpha_0 & \alpha_0 p & 0 & \alpha_0(1-p) \\ (1,\widehat{v}_0) & \alpha_1(1-p) & 0 & \alpha_1 p & 1-\alpha_1 \\ (0,\widehat{v}_1) & 1-\alpha_1 & \alpha_1 p & 0 & \alpha_1(1-p) \\ (1,\widehat{v}_1) & \alpha_0(1-p) & 0 & \alpha_0 p & 1-\alpha_0 \end{array}\right)$$

Fig. 14. Transition probabilities of the Markov chain $\widehat{P}$.

Remark C.6: Employing the same type of arguments applied to the graph $\widehat{\mathsf{Q}}$ in Figure 13 (rather than to the graph Q in Figure 10) results in the Markov chain $\widehat{P}$, whose transition probabilities are shown in Figure 14. The states $\widehat{v}_0$ and $\widehat{v}_1$ of $\widehat{\mathsf{Q}}$ represent two categories of (the last run of) an output sequence: runs of 0's and runs of 1's, respectively. This, in turn, leads to another upper bound, $\rho_2(p)$, on $\mathsf{cap}_{\mathsf{FB}}(p)$, given by $\sup_{(\alpha_0,\alpha_1)\in[0,1]^2}\{\widehat{\mathsf{z}}(\alpha_0,\alpha_1)\}$, where

$$\widehat{\mathsf{z}}(\alpha_0,\alpha_1) \triangleq \mathrm{H}\left(\frac{\alpha_0(1-2\alpha_1 p+2\alpha_1 p^2)}{1+(\alpha_0-\alpha_1)p}\right) - \frac{\alpha_0\,\mathrm{H}(p)}{1+(\alpha_0-\alpha_1)p}\ .$$

Since $1-2\alpha_1 p+2\alpha_1 p^2 \in [0,1]$, we have the following inequality

$$\widehat{\mathsf{z}}(\alpha_0,\alpha_1) \le \mathrm{H}\left(\frac{\alpha_0(1-2\alpha_1 p+2\alpha_1 p^2)}{1+(\alpha_0-\alpha_1)p}\right) - \frac{\alpha_0(1-2\alpha_1 p+2\alpha_1 p^2)\,\mathrm{H}(p)}{1+(\alpha_0-\alpha_1)p} = \mathrm{H}(x) - \mathrm{H}(p)x, \tag{60}$$

where $x = \frac{\alpha_0(1-2\alpha_1 p+2\alpha_1 p^2)}{1+(\alpha_0-\alpha_1)p} \in [0,1]$. The global maximum of the right-hand side of (60) is attained at $x = \frac{1}{1+2^{\mathrm{H}(p)}}$, therefore, for any $(\alpha_0,\alpha_1,p)\in[0,1]^3$,

$$\widehat{\mathsf{z}}(\alpha_0,\alpha_1) \le \mathrm{H}\left(\frac{1}{1+2^{\mathrm{H}(p)}}\right) - \frac{\mathrm{H}(p)}{1+2^{\mathrm{H}(p)}}\ ,$$

with equality holding when $\alpha_0 = \frac{1}{1-p+2^{\mathrm{H}(p)}}$ and $\alpha_1 = 0$. Thus, we obtain the upper bound

$$\mathsf{cap}_{\mathsf{FB}}(p) \le \rho_2(p) = \mathrm{H}\left(\frac{1}{1+2^{\mathrm{H}(p)}}\right) - \frac{\mathrm{H}(p)}{1+2^{\mathrm{H}(p)}},$$

which coincides with Theorem 3.1 for $p \in [0,p_0]$.

□

Remark C.7: In a similar vein to the proof of the upper bound $\rho_1(p)$ (specifically, by proving counterparts of assertions (M1)–(M4) and employing Lemma C.1) we were able to establish that on the interval $p \in [0.5, 1-p_0]$, the upper bound of Theorem C.3 is given by

$$\begin{aligned} \mathsf{f}_1(p) &= \max_{\alpha_0\in[0,1]}\{\mathsf{z}(\alpha_0,0,0,0)\} \\ &= \max_{\alpha_0\in[0,1]}\left\{\frac{1}{2+\alpha_0 p}\Big(\mathrm{H}(\alpha_0(1-p)) + (1+\alpha_0 p)\,\mathrm{H}\left(\frac{1}{1+\alpha_0 p}\right) - \mathrm{H}(p)\alpha_0\Big)\right\}\ . \end{aligned}$$

Similarly, on the interval $[1-p_0, 0.65]$, one can prove counterparts of assertions (M1) and (M2) and notice that the preimage of the global maximum of $\widetilde{\mathsf{z}}(\mathsf{x},\mathsf{y})$ occurs inside U_p, thereby implying that the upper bound of Theorem C.3 is given by

$$\mathsf{z}\left(\mathsf{x} = \frac{2^{\mathrm{H}(p)}}{1+2^{\mathrm{H}(p)}}, \mathsf{y} = \frac{2^{\mathrm{H}(p)}}{1+2^{\mathrm{H}(p)}}\right) = \mathrm{H}\Big(\frac{1}{2^{\mathrm{H}(p)}+1}\Big) - \frac{\mathrm{H}(p)}{2^{\mathrm{H}(p)}+1} = \rho_2(p)\ . \tag{61}$$

Based on numeric computations, we conjecture that the upper bound given by (61) extends to the interval $[1-p_0, p_2] \approx [0.601676, 0.849946]$, where p_2 is the unique solution of the following equation on the interval $[0.5, 1]$:

$$\frac{2^{\mathrm{H}(p)}}{1+2^{\mathrm{H}(p)}} = \frac{p}{2(1-p-p^2)}\left(1-\sqrt{1-4(1-p-p^2)}\right)\ ;$$

moreover, on the interval $[p_2, 0.899]$, the upper bound is conjectured to be

$$\begin{aligned} \mathsf{f}_0(p) &= \max_{\alpha_2\in[0,1]}\{\mathsf{z}(1,0,\alpha_2,0)\} \\ &= \max_{\alpha_2\in[0,1]}\left\{\frac{1}{2-\alpha_2+p}\Big((1-\alpha_2+\alpha_2 p)\,\mathrm{H}\left(\frac{(1-\alpha_2)p}{1-\alpha_2+\alpha_2 p}\right) + (1+p-\alpha_2 p)\,\mathrm{H}\left(\frac{1-\alpha_2+\alpha_2 p}{1+p-\alpha_2 p}\right) - \mathrm{H}(p)\Big)\right\}\ . \end{aligned}$$

□

ACKNOWLEDGMENT

We would like to thank Oron Sabag for bringing the reference [23] to our attention and for elaborating on its results. We would also like to thank the anonymous referees for their valuable comments and suggestions.

REFERENCES

[1] A. Arapostathis, V.S. Borkar, E. Fernández-Gaucherand, M.K. Ghosh, S.I. Marcus, "Discrete-time controlled Markov processes with average cost criterion: a survey," *SIAM J. Contr. Optim.,* vol. 31, pp. 282–344, 1993.

[2] H. Asnani, H.H. Permuter, T. Weissman, "Capacity of a POST channel with and without feedback," *Proc. IEEE Int'l Symp. Inf. Theory (ISIT 2013),* Istanbul, Turkey, 2013, pp. 2538–2542.

[3] R. Bellman, "A Markovian decision process," *J. Math. Mech.,* vol. 6, pp. 679–684, 1957.

[4] T. Berger, F. Bonomi, "Capacity and zero-error capacity of Ising channels," *IEEE Trans. Inf. Theory,* vol. 36, pp. 173–180, 1990.

[5] D.P. Bertsekas, *Dynamic Programming and Stochastic Control,* Academic Press, London, UK, 1976.

[6] D. Blackwell, L. Breiman, A.J. Thomasian, "Proof of Shannon's transmission theorem for finite-state indecomposable channels," *Ann. Math. Stat.* vol. 29, pp. 1209–1220, 1958.

[7] S. Boyd, L. Vandenberghe, *Convex Optimization,* Cambridge Univ. Press, Cambridge, UK, 2004.

[8] T.M. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory,* vol. 19, pp. 73–77, 1973.

[9] T.M. Cover, J.A. Thomas, *Elements of Information Theory,* Wiley, New York, 1991.

[10] O. Elishco, H. Permuter, "Capacity and coding for the Ising channel with feedback," *IEEE Trans. Inf. Theory,* vol. 60, pp. 5138–5149, 2014.

[11] R.G. Gallager, *Information Theory and Reliable Communication,* Wiley, New York, 1968.

[12] S. Greaves, Y. Kanai, H. Muraoka, "Shingled magnetic recording on bit patterned media," *IEEE Trans. Magn.,* vol. 46, pp. 1460–1463, 2010.

[13] K.A.S. Immink, *Codes for Mass Data Storage Systems,* Shannon Foundation Publishers, Eindhoven, Netherlands, 2004.

[14] E. Ising, "Beitrag zur Theorie des Ferromagnetismus" (in German), *Zeitschrift für Physik A Hadrons and Nuclei,* vol. 31, pp. 253–258, 1925.

[15] A.R. Iyengar, P.H. Siegel, J.K. Wolf, "Write channel model for bit-patterned media recording," *IEEE Trans. Magn.,* vol. 47, pp. 35–45, 2011.

[16] W. Lenz, "Beiträge zum Verständnis der magnetischen Eigenschaften in festen Körpern" (in German), *Physikalische Zeitschrift,* vol. 21, pp. 613–615, 1920.

[17] H. Marko, "The bidirectional communication theory — a generalization of information theory," *IEEE Trans. Commun.,* vol. 21, pp. 1345–1351, 1973.

[18] J.L. Massey, "Causality, feedback and directed information," *Proc. IEEE Int'l Symp. Inf. Theory Applic. (ISITA 1990),* Honolulu, HI, 1990, pp. 303–305.

[19] A. Mazumdar, A. Barg, N. Kashyap, "Coding for high-density recording on a 1-D granular magnetic medium," *IEEE Trans. Inf. Theory,* vol. 57, pp. 7403–7417, 2011.
[20] H.H. Permuter, P.W. Cuff, B. Van Roy, T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory,* vol. 54, pp. 3150–3165, 2008.
[21] Y.H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inf. Theory,* vol. 25, pp. 1488–1499, 2008.
[22] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory,* vol. 49, pp. 4–21, 2003.
[23] O. Sabag, H.H. Permuter, H.D. Pfister, "A single-letter upper bound on the feedback capacity of unifilar finite-state channels," submitted to *Proc. IEEE Int'l Symp. Inf. Theory (ISIT 2016),* Barcelona, Spain, 2016.
[24] A. Sharov, R.M. Roth, "Fixed-rate tiling encoders for 2-D constraints," *Proc. IEEE Int'l Symp. Inf. Theory (ISIT 2010),* Austin, TX, 2010, pp. 1208–1212.
[25] A. Sharov, R.M. Roth, "Bounds and constructions for granular media coding," *IEEE Trans. Inf. Theory,* vol. 60, pp. 2010–2027, 2014.
[26] L.G. Tallini, S. Al-Bassam, B. Bose, "Feedback codes achieving the capacity of the Z-channel," *IEEE Trans. Inf. Theory* vol. 54, pp. 1357–1362, 2008
[27] S. Tatikonda, "Control under communication constraints," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2000.
[28] S. Tatikonda, S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory,* vol. 55, pp. 323–349, 2009.
[29] P.O. Vontobel, A. Kavčić, D.M. Arnold, H.-A. Loeliger, "A generalization of the Blahut–Arimoto algorithm to finite-state channels," *IEEE Trans. Inf. Theory,* vol. 54, pp. 1887–1918, 2008.
[30] S. Yang, A. Kavčić, S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inf. Theory,* vol. 51, pp. 799–810, 2005.
[31] S. Yang, A. Kavčić, S. Tatikonda, "On the feedback capacity of power constrained Gaussian channels with memory," *IEEE Trans. Inf. Theory,* vol. 53, pp. 929–954, 2007.

Artyom Sharov received the B.Sc., M.Sc., and Ph.D. degrees in computer science from Technion—Israel Institute of Technology, Haifa, Israel, in 2004, 2010 and 2016, respectively. He was a recipient of the Best Student Paper Award at the 2014 IEEE International Symposium on Information Theory (ISIT). His research interests include constrained coding and coding theory.

Ron M. Roth (M'88–SM'97–F'03) received the B.Sc. degree in computer engineering, the M.Sc. in electrical engineering, and the D.Sc. in computer science from Technion—Israel Institute of Technology, Haifa, Israel, in 1980, 1984, and 1988, respectively. Since 1988 he has been with the Computer Science Department at Technion, where he now holds the General Yaakov Dori Chair in Engineering. During the academic years 1989–91 he was a Visiting Scientist at IBM Research Division, Almaden Research Center, San Jose, California, and during 1996–97, 2004–05, and 2011–2012 he was on sabbatical leave at Hewlett–Packard Laboratories, Palo Alto, California. He is the author of the book *Introduction to Coding Theory*, published by Cambridge University Press in 2006. Dr. Roth was an associate editor for coding theory in IEEE TRANSACTIONS ON INFORMATION THEORY from 1998 till 2001, and he is now serving as an associate editor in *SIAM Journal on Discrete Mathematics*. His research interests include coding theory, information theory, and their application to the theory of complexity.