Capacity and Coding for the Ising Channel With Feedback

Ohad Elishco, *Student Member, IEEE*, and Haim Permuter, *Senior Member, IEEE*

***Abstract*— The Ising channel, which was introduced in 1990, is a channel with memory that models intersymbol interference. In this paper, we consider the Ising channel with feedback and find the capacity of the channel together with a capacity-achieving coding scheme. To calculate the channel capacity, an equivalent dynamic programming (DP) problem is formulated and solved. Using the DP solution, we establish that the feedback capacity is the expression $C = (2H_b(a)/3 + a) \approx 0.575522$, where a is a particular root of a fourth-degree polynomial and $H_b(x)$ denotes the binary entropy function. Simultaneously, $a = \arg\max_{0 \le x \le 1}(2H_b(x)/3 + x)$. Finally, an error-free, capacity-achieving coding scheme is provided together with the outlining of a strong connection between the DP results and the coding scheme.**

***Index Terms*— Bellman Equation, dynamic program, feedback capacity, Ising channel, infinite-horizon, value iteration.**

I. Introduction

THE Ising model originated as a problem in statistical mechanics. It was invented by Lenz in 1920 [1], who gave it as a problem to his student, Ernst Ising, after whom it is named [2]. A few years later the two dimensional Ising model was analytically defined by Onsager [3]. The Ising channel, on the other hand, was introduced as an information theory problem by Berger and Bonomi in 1990 [4]. It has received this name due to its resemblance to the physical Ising model.

In their work on the Ising channel, Berger and Bonomi found the zero-error capacity and a numerical approximation of the capacity of the Ising channel *without* feedback. In order to find the numerical approximation, the Blahut-Arimoto Algorithm [5], [6] was used. The capacity was found to be bounded by $0.5031 \le C \le 0.6723$ and the zero-error capacity was found to be 0.5 bit per channel use. Moreover, their work contains a coding scheme that achieves the zero-error capacity. This code is the basis for the capacity-achieving coding scheme in the presence of feedback presented in this paper.

The Ising channel models a channel with Inter-Symbol Interference (ISI) and works as follows: at time t a certain bit, x_t, is transmitted through the channel. The channel output at time t is denoted by y_t. If $x_t = x_{t-1}$ then $y_t = x_t$ with probability 1. If $x_t \neq x_{t-1}$ then y_t is distributed Bernoulli $\left(\frac{1}{2}\right)$.

In this paper we consider the Ising channel with feedback. The objective is to find the channel feedback capacity explicitly and to provide a capacity-achieving coding scheme. Finding an explicit expression for the capacity of non-trivial channels with memory, with or without feedback, is usually a very hard problem. There are only a few cases in the literature that have been solved, such as additive Gaussian channels with memory without feedback ("water filling solution," [7], [8]), additive Gaussian channels with feedback where the noise is ARMA of order 1 [9], channels with memory where the state is known both to the encoder and the decoder [10], [11], and the trapdoor channel with feedback [12]. This paper adds one additional case, the Ising channel.

Towards this goal, we start from the characterization of the feedback-capacity as the normalized directed information $\frac{1}{n}I(X^n \to Y^n)$. The directed information was introduced two decades ago by Massey [13] (who attributed it to Marko [14]) as

$$I(X^n \to Y^n) = \sum_{i=1}^{n} I(X^i; Y_i | Y^{i-1}). \quad (1)$$

Massey [13] showed that the normalized maximum directed information upper bounds the capacity of channels with feedback. Subsequently, it was shown that directed information, as defined by Massey, indeed characterizes the capacity of channels with feedback [15]–[21].

The capacity of the Ising channel with feedback was approximated numerically [22] using an extension of Blahut-Arimoto algorithm for directed information. Here, we present the explicit expression together with a capacity-achieving coding scheme. The main difficulty of calculating the feedback capacity explicitly is that it is given by an optimization of an infinite-letter expression. In order to overcome this difficulty, we transform the normalized directed information optimization problem into an infinite average-reward dynamic programming problem. The idea of using dynamic programming (DP) for computing the directed information capacity has been introduced and applied in several recent papers such as [11], [12], [17], and [23]. The DP used here most resembles the trapdoor channel model [12]. We use a DP method that is specified for the Ising channel rather than the trapdoor channel and provide an analytical solution for the new specific DP.

It turns out that the DP not only helps in computing the feedback capacity but also provides important information

Manuscript received May 20, 2012; revised July 2, 2013; accepted May 19, 2014. Date of publication June 19, 2014; date of current version August 14, 2014. This work was supported by the European Research Council. This paper was presented at the 2012 IEEE International Symposium on Information Theory.

The authors are with the Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: ohadeli@post.bgu.ac.il; haimp@post.bgu.ac.il).

Communicated by Y.-H. Kim, Associate Editor for Shannon Theory.

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

Digital Object Identifier 10.1109/TIT.2014.2331951

0018-9448 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

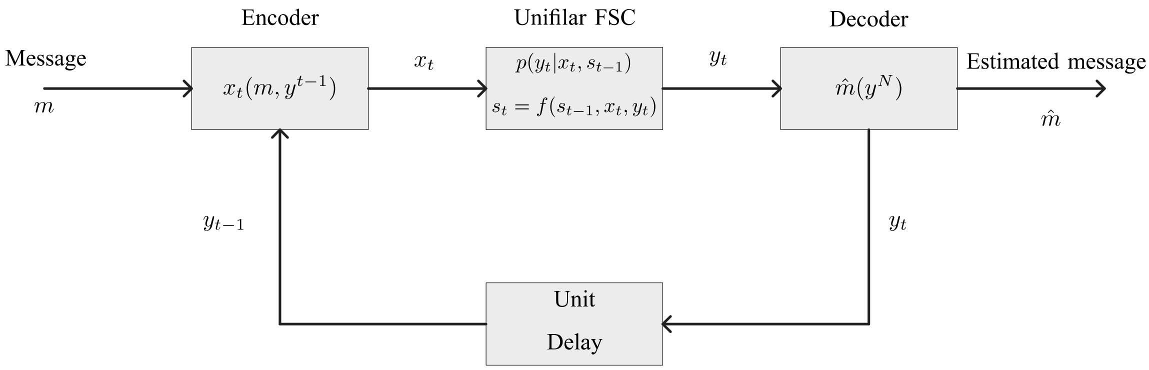

Fig. 1. Unifilar finite state channel with feedback of unit delay.

regarding a coding scheme that achieves the capacity. Through the DP formulation and through its solution we are able to derive a concrete coding scheme that achieves the feedback capacity. The states and the actions of the DP turn out to include exact instructions for what the encoder and the decoder do to achieve the feedback-capacity.

The remainder of the paper is organized as follows: in Section II we present some notations, which are used throughout the paper, basic definitions, and the channel model. In Section III we present the main results. In Section IV we present the outline of the method used to calculate the channel capacity. This section also contains a short reminder about DP and about the Bellman Equation, which is used in order to find the capacity. In Section V we formulate the DP problem according to the Ising channel with feedback. In Section VI an analytical solution to the Bellman Equation is found. Section VII contains the connection between the DP results and the coding scheme. From this connection we can derive the coding scheme explicitly. In Section VIII we prove that the suggested coding scheme indeed achieves the capacity. Section IX contains conclusions and a discussion of the results.

II. Notation, Definitions and Channel Model

A. Notation

Calligraphic letters, $\mathcal{X}$, denote alphabets, upper-case letters, X, denote random variables and lower-case letters, x, denote sample values. Superscript, x^t, denotes the vector $(x_1, \ldots, x_t)$. The probability distribution of a random variable, X, is denoted by p_X. We omit the subscript of the random variable when the arguments have the same letter as the random variable, e.g. $p(x|y) = p_{X|Y}(x|y)$.

B. Definitions

Here we present some basic definitions beginning with a definition of a finite state channel (FSC).

Definition 1 (FSC [24, Ch. 4]): An FSC is a channel that has a finite number of possible states and has the property: $p(y_t, s_t|x^t, s^{t-1}, y^{t-1}) = p(y_t, s_t|x_t, s_{t-1})$.

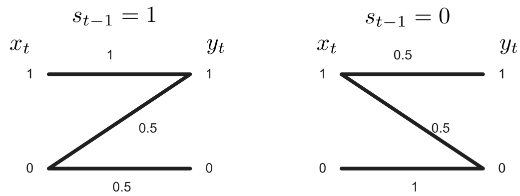

Fig. 2. The Ising channel model. On the left we have the Z topology; and on the right we have the S topology.

Definition 2 (Unifilar FSC [25]): An FSC is called a unifilar FSC if there exists a time-invariant function $f(\cdot)$ such that $s_t = f(s_{t-1}, x_t, y_t)$.

Definition 3 (Connected FSC [24, Ch. 4]): An FSC is called a connected FSC if $\forall s,\ s' \in \mathcal{S} \exists T_s \in \mathbb{N}$ and $\{p(x_t|s_{t-1})\}_{t=1}^{T_s}$ such that

$$\sum_{t=1}^{T_s} p_{S_t|S_0}(s|s') > 0.$$

In other words, for any given states s, s' there exists an integer T_s and an input distribution $\{p(x_t|s_{t-1})\}_{t=1}^{T_s}$, such that the probability of the channel reaching the state s from the state s' is positive.

C. Channel Model

In this part, the Ising channel model is introduced. The channel is a unifilar FSC with feedback, as depicted in Fig. 1. As mentioned before, the sets $\mathcal{X}, \mathcal{Y}, \mathcal{S}$ denote the input, output, and state alphabet, respectively. In the Ising channel model: $\mathcal{X} = \mathcal{Y} = \mathcal{S} = \{0, 1\}$.

The Ising channel consists of two different topologies, as described in Fig. 2. The channel topologies depend on the channel state and are denoted by Z and S. These Z and S notations are compatible with the well-known Z and S channels. The channel topology at time t is determined by $s_{t-1} \in \{0, 1\}$.

As shown in Fig. 2, if $s_{t-1} = 1$, the channel is in the Z topology; if $s_{t-1} = 0$, the channel is in the S topology. The channel state at time t is defined as the input to the channel at time t, meaning $s_t = x_t$.

TABLE I
THE CHANNEL STATES, TOPOLOGIES, AND INPUTS, TOGETHER WITH THE PROBABILITY THAT THE OUTPUT AT TIME t, y_t, IS EQUAL TO THE INPUT AT TIME t, x_t

$s_{t-1}(=x_{t-1})$	Topology	x_t	$p(y_t = x_t \mid x_t, s_{t-1})$
0	S	0	1
0	S	1	0.5
1	Z	0	0.5
1	Z	1	1

The channel input, x_t, and state, s_{t-1}, have a crucial effect on the output, y_t. If the input is identical to the previous state, i.e. $x_t = s_{t-1}$, then the output is equal to the input, $y_t = x_t$, with probability 1; if $x_t \neq s_{t-1}$ then y_t can be either 0 or 1, each with probability 0.5. This effect is summarized in Table I.

We assume a communication setting that includes feedback with unit delay. Hence, the transmitter (encoder) knows at time t the message m and the feedback samples y^{t-1}. Therefore, the input to the channel, x_t, is a function of both the message and the feedback, as shown in Fig. 1.

Lemma 1: The Ising channel is a connected unifilar FSC.

Proof: Lemma 1 is proved in three steps. At each step we show a different property of the channel.

(a) The channel is an FSC since it has two states, 0 and 1. It is clear that $p(y_t, s_t \mid x^t, s^{t-1}, y^{t-1}) = p(y_t, s_t \mid x_t, s_{t-1})$ since $s_t = x_t$ and y_t depends only on x_t and s_{t-1}.

(b) The channel is a unifilar FSC since (a) it is an FSC and (b) $s_t = x_t$. Obviously, $s_t = f(s_{t-1}, x_t, y_t) = x_t$ is a time-invariant function.

(c) The channel is a connected FSC since $s_t = x_t$. Thus, one can take $T_s = 1$ and $p_{X_t|S'}(s|s') = 1$, resulting in $\Pr(S_1 = s \mid S_0 = s') = 1 > 0$. ∎

III. MAIN RESULTS

Theorem 1: (a) The capacity of the Ising channel with feedback is $C_f = \left(\frac{2H(a)}{3+a}\right) \approx 0.5755$ where $a \approx 0.4503$ is a root of the fourth-degree polynomial $x^4 - 5x^3 + 6x^2 - 4x + 1$.

(b) The capacity, C_f, is also equal to

$$\max_{0 \le z \le 1} \left(\frac{2H(z)}{3+z}\right)$$

where

$$a = \arg\max_{0 \le z \le 1} \left(\frac{2H(z)}{3+z}\right) \approx 0.4503.$$

Theorem 1 is proved by solving an infinite horizon average reward DP problem. This DP problem provides us with a specific coding scheme, which is given in the next theorem.

Theorem 2: There is a capacity-achieving coding scheme, which follows these rules:

(i) Assume the message is a stream of n bits distributed i.i.d. with probability 0.5.

(ii) Transform the message bit stream into a stream of bits, M, with alternation probability of $a \approx 0.4503$.

(iii) We denote the tth bit of M as m_t where it corresponds to the t'th encoder's entry:

(1) *Encoder:* At time t', the encoder knows $s_{t'-1} = x_{t'-1}$, and we send the bit m_t $(x_{t'} = m_t)$:

(1.1) If $y_{t'} \neq s_{t'-1}$ then move to the next bit, m_{t+1}. This means that we send m_t once.

(1.2) If $y_{t'} = s_{t'-1}$ then $x_{t'} = x_{t'+1} = m_t$, which means that the encoder sends m_t twice (at time t' and $t'+1$), and then move to the next bit.

(2) *Decoder:* At time t', assume the state $s_{t'-1}$ is known at the decoder, and we are to decode the bit $\hat{m}_t$:

(2.1) If $y_{t'} \neq s_{t'-1}$ then $\hat{m}_t = y_{t'}$ and $s_{t'} = y_{t'}$.

(2.2) If $y_{t'} = s_{t'-1}$ then wait for $y_{t'+1}$. $\hat{m}_t = y_{t'+1}$ and $s_{t'} = y_{t'+1}$.

IV. METHOD OUTLINE, DYNAMIC PROGRAMMING, AND THE BELLMAN EQUATION

In order to formulate an equivalent dynamic program we use the following theorem.

Theorem 3 ([12], Th. 1): The feedback capacity, C_{FB}, of a connected unifilar FSC when initial state s_0 is known at the encoder and the decoder, can be expressed as

$$\sup_{\{p(x_t|s_{t-1}, y^{t-1})\}_{t\ge 1}} \liminf_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} I(X_t, S_{t-1}; Y_t \mid Y^{t-1}), \quad (2)$$

where $\{p(x_t|s_{t-1}, y^{t-1})\}_{t\ge 1}$ denotes the set of all distributions such that $p(x_t|y^{t-1}, x^{t-1}, s^{t-1}) = p(x_t|s_{t-1}, y^{t-1})$ for $t = 1, 2, \ldots$

Using Theorem 3 we can formulate the feedback-capacity problem as an infinite-horizon average-reward DP. Then, using value iteration algorithm and the Bellman Equation we find the optimal average reward, which gives us the channel capacity.

We denote by $\mathcal{Z}, \mathcal{U}, \mathcal{W}$ the state space, the action space, and the disturbance space of the DP. The system evolves according to $z_t = F(z_{t-1}, u_t, w_t)$, $\quad t \in \mathbb{N}$ and μ_t maps histories, h_t to actions. The disturbance, w_t, should depend only on the state z_{t-1} and the action u_t.[1]

The objective is to maximize the average reward. Given a bounded reward function, $g : \mathcal{Z} \times \mathcal{U} \to \mathbb{R}$, the average reward is defined by

$$\rho_\pi = \liminf_{N\to\infty} \frac{1}{N} \mathbb{E}_\pi \left\{ \sum_{t=0}^{N-1} g(Z_t, \mu_{t+1}(h_{t+1})) \right\}, \quad (3)$$

where the subscript π indicates that actions are generated by the policy $\pi = (\mu_1, \mu_2, \ldots)$. The optimal average reward is defined by $\rho^* = \sup_\pi \rho_\pi$.

The Bellman Equation, also known as the average cost optimality equation (ACOE), is given in (4). It verifies that

[1]The DP model presented in this paper is identical to the model defined in [26, Sec. 1.2 and 1.3] and is slightly different from the formulation that appears in [27, Sec. 2]. However, they are equivalent. In our formulation, the disturbance, w_t, may depend on the action and state, i.e. $p(w_t|z_t, u_t)$, but not on values of prior disturbances, while in the formulation presented in [27, Sec. 2] the disturbance is a sequence of i.i.d. RVs. To see that the models are equivalent, one should look at [27, Example 2.1] taking $P(D|x, a) := \int_W I\{F(x, a, w) \in D\} P_W(dw|x, a)$ instead of $P(D|x, a) := \int_W I\{F(x, a, w) \in D\} P_W(dw)$, where D is a Borel set and I is the indicator function.

TABLE II
THE ISING CHANNEL MODEL NOTATIONS VS. DYNAMIC PROGRAMMING NOTATIONS

Ising channel notations	Dynamic Programming notations
$p(s_t=0\|y^t)$, probability of the channel state to be 0 given the output	z_t, the DP state
y_t, the channel output	w_t, the DP disturbance
$p(x_t\|s_{t-1})$, channel input probability given the channel state at time $t-1$	u_t, the DP action
Eq. (7)	$z_t=F(z_{t-1},u_{t-1},w_{t-1})$, the DP state evolves according to a function F
$I(X_t,S_{t-1};Y_t\|y^{t-1})$	$g(z_{t-1},u_t)$, the DP reward function

a given average reward and a given policy are optimal. Theorem 4 follows from [27, Th. 6.1].

Theorem 4 ([27], Th. 6.1[2]): If $\rho\in\mathbb{R}$ and a bounded function $h:\mathcal{Z}\mapsto\mathbb{R}$ satisfy

$$\rho+h(z)=\sup_{u\in\mathcal{U}}\left(g(z,u)+\int P_w(dw|z,u)h(F(z,u,w))\right) \quad (4)$$

for every $z\in\mathcal{Z}$, then $\rho=\rho^*$. Furthermore, if there is a function $\mu:\mathcal{Z}\mapsto\mathcal{U}$ such that $\mu(z)$ attains the supremum for each z and satisfies (4), then $\rho_\pi=\rho^*$ for $\pi=(\mu_0,\mu_1,\ldots)$ with $\mu_t(h_t)=\mu(z_{t-1})$ for each t.

We define a DP operator T by

$$(Th)(z)=\sup_{u\in\mathcal{U}}\left(g(z,u)+\int P_w(dw|z,u)h(F(z,u,w))\right) \quad (5)$$

for all functions h. Thus, the Bellman Equation can be written as $\rho\mathbf{1}+h=Th$. We also denote as $T_\mu h$ the DP operator restricted to the policy, μ.

V. DYNAMIC PROGRAMMING FORMULATION FOR THE ISING CHANNEL

In this section we associate the DP problem, which was discussed in the previous section, with the Ising channel. Using the notations previously defined, the DP state, z_t, is the vector of channel state probabilities $[p(s_t=0|y^t),p(s_t=1|y^t)]$. In order to simplify notations, we consider the state z_t to be the first component; that is, $z_t:=p(s_t=0|y^t)$. This comes with no loss of generality, since $p(s_{t-1}=0|y^{t-1})+p(s_{t-1}=1|y^{t-1})=1$. Hence, the second component can be derived from the first, since the pair sums to one. The action, u_t, is a 2×2 stochastic matrix

$$u_t=\begin{bmatrix} p(x_t=0|s_{t-1}=0) & p(x_t=1|s_{t-1}=0)\\ p(x_t=0|s_{t-1}=1) & p(x_t=1|s_{t-1}=1)\end{bmatrix}. \quad (6)$$

The disturbance, w_t, is the channel output, y_t. The DP-Ising channel association is presented in Table II.

Note that since the Ising channel is connected and unifilar, given a policy $\pi=(\mu_1,\mu_2,\ldots)$, $p(s_t|y^t)$ is given in (7) that appears on the bottom of the page (as shown in [12, eq. (35)]), where $\mathbf{1}(\cdot)$ is the indicator function. The conditional distribution of the disturbance, w_t, satisfies $p(w_t|z^{t-1},w^{t-1},u^t)=p(w_t|z_{t-1},u_t)$. The latter equality holds since the channel output is determined by the channel state and input. This equality is shown in [12, eq. (36)].

The following lemma describes the Ising channel in an explicit DP formulation that is solvable.

Lemma 2: The capacity optimization problem of the Ising channel given in (2) can be formulated as follows. The DP state at time t evolves according to

$$z_t=\begin{cases}1+\frac{\delta_t-z_{t-1}}{1+\delta_t-\gamma_t}, & \text{if } w_t=0\\ \frac{1-z_{t-1}-\gamma_t}{1+\gamma_t-\delta_t}, & \text{if } w_t=1,\end{cases} \quad (9)$$

where

$$\gamma_t:=(1-z_{t-1})u_t(2,2)$$
$$\delta_t:=z_{t-1}u_t(1,1). \quad (10)$$

Furthermore, the DP operator is given by

$$(Th)(z)=\sup_{0\le\delta\le z,0\le\gamma\le 1-z} H_b\left(\frac{1}{2}+\frac{\delta-\gamma}{2}\right)+\delta+\gamma$$
$$-1+\frac{1+\delta-\gamma}{2}h\left(1+\frac{\delta-z}{\delta+1-\gamma}\right)$$
$$+\frac{1-\delta+\gamma}{2}h\left(\frac{1-z-\gamma}{1+\gamma-\delta}\right). \quad (11)$$

Proof: The state evolves according to $z_t=F(z_{t-1},u_t,w_t)$. Using relations from (7) we obtain F explicitly as shown in (11). The expressions in (11) can be simplified by defining γ_t and δ_t as in (9). Since $u_t(1,1)=1-u_t(1,2),u_t(2,2)=1-u_t(2,1)$, we see that z_t evolves as in (8)

Note that γ_t,δ_t are functions of z_{t-1}. As shown in (9), given z_{t-1}, the action u_t defines the pair (γ_t,δ_t) and vice versa. From here on, we represent the actions in terms of

$$p(s_t|y^t)=\frac{\sum_{x_t,s_{t-1}}p(s_{t-1}|y^{t-1})u_t(s_{t-1},x_t)p(y_t|s_{t-1},x_t)\mathbf{1}(s_t=f(s_{t-1},x_t,y_t))}{\sum_{x_t,s_t,s_{t-1}}p(s_{t-1}|y^{t-1})u_t(s_{t-1},x_t)p(y_t|s_{t-1},x_t)\mathbf{1}(s_t=f(s_{t-1},x_t,y_t))}, \quad (7)$$

$$z_t=\begin{cases}\frac{z_{t-1}u_t(1,1)+0.5(1-z_{t-1})u_t(2,1)}{z_{t-1}u_t(1,1)+0.5z_{t-1}u_t(1,2)+0.5(1-z_{t-1})u_t(2,1)}, & \text{if } w_t=0\\ \frac{0.5(1-z_{t-1})u_t(2,1)}{0.5z_{t-1}u_t(1,2)+0.5(1-z_{t-1})u_t(2,1)+0.5(1-z_{t-1})u_t(2,2)}, & \text{if } w_t=1.\end{cases} \quad (8)$$

TABLE III
THE CONDITIONAL DISTRIBUTION $p(x_t, s_{t-1}, y_t|y^{t-1})$

x_t	s_{t-1}	$y_t = 0$	$y_t = 1$
0	0	$p(s_{t-1} = 0\|y_{t-1})u_t(1,1)$	0
0	1	$0.5p(s_{t-1} = 1\|y_{t-1})u_t(2,1)$	$0.5p(s_{t-1} = 1\|y_{t-1})u_t(2,1)$
1	0	$0.5p(s_{t-1} = 0\|y_{t-1})u_t(1,2)$	$0.5p(s_{t-1} = 0\|y_{t-1})u_t(1,2)$
1	1	0	$p(s_{t-1} = 1\|y_{t-1})u_t(2,2)$

γ_t and δ_t. Since u_t is a stochastic matrix, we have the constraints $0 \le \delta_t \le z_t$ and $0 \le \gamma_t \le 1 - z_t$.

We now consider the reward to be $g(z_{t-1}, u_t) = I(X_t, S_{t-1}; Y_t|y^{t-1})$. Note that

$$p(x_t, s_{t-1}, y_t|y^{t-1}) = p(s_{t-1}|y^{t-1})p(x_t|s_{t-1}, y^{t-1})p(y_t|x_t, s_{t-1}) \quad (12)$$

and recall that $p(y_t|x_t, s_{t-1})$ is given by the channel model. Thus, the reward is dependent only on $p(s_{t-1}|y^{t-1})$ and $p(x_t|s_{t-1}, y^{t-1}) = u_t$. Since $p(s_{t-1}|y^{t-1})$ is given by z_{t-1}, we have that the reward is a function of u_t and z_{t-1}. Now we find the reward function, $g(z_{t-1}, u_t)$, explicitly for the Ising channel:

$$\begin{aligned} I(X_t, S_{t-1}; Y_t|y^{t-1}) &= H_b\left(Y_t|y^{t-1}\right) - H_b\left(Y_t|X_t, S_{t-1}, y^{t-1}\right) \\ &\overset{(a)}{=} H_b\left(z_{t-1}u_t(1,1) + \frac{z_{t-1}u_t(1,2)}{2} + \frac{z_{t-1}u_t(2,1)}{2}\right) \\ &\quad - (z_{t-1}u_t(1,2)\cdot 1 + (1 - z_{t-1})u_t(2,1)\cdot 1) \\ &\overset{(b)}{=} H_b\left(\frac{1}{2} + \frac{\delta_t - \gamma_t}{2}\right) + \delta_t + \gamma_t - 1. \end{aligned} \quad (13)$$

Where $H_b(\cdot)$ denotes the binary entropy function. Equation (a) follows from Table III where the conditional distribution $p(x_t, s_{t-1}, y_t|y^{t-1})$ is calculated using (12) and (b) follows from the definition of δ and γ given in (9) and since u_t is a stochastic matrix. Therefore, we can write the DP operator, given in (5) explicitly, substituting $g(z, u)$ with $I(X_t, S_{t-1}; Y_t|y^{t-1})$ as found in (13):

$$\begin{aligned} (Th)(z) &= \sup_{u\in\mathcal{U}} \left(g(z,u) + \int P_w(dw|z,u)h(F(z,u,w))\right) \\ &= \sup_{u\in\mathcal{U}} \left(H_b\left(\frac{1}{2} + \frac{\delta_t - \gamma_t}{2}\right) + \delta_t + \gamma_t - 1 \right. \\ &\quad \left. + \int P_w(dw|z,u)h(F(z,u,w))\right) \\ &\overset{(a)}{=} \sup_{0\le\delta\le z, 0\le\gamma\le 1-z} H_b\left(\frac{1}{2} + \frac{\delta - \gamma}{2}\right) + \delta + \gamma \\ &\quad -1 + \frac{1+\delta-\gamma}{2} h\left(1 + \frac{\delta - z}{\delta + 1 - \gamma}\right) \\ &\quad + \frac{1-\delta+\gamma}{2} h\left(\frac{1-z-\gamma}{1+\gamma-\delta}\right) \end{aligned} \quad (14)$$

where (a) follows from the fact that in the Ising channel, $\int P_w(dw|z,u)h(F(z,u,w))$ takes the form $\sum_{w=0,1} p(w|z,u)h(F(z,u,w))$ and $F(z,u,w)$ is given in (8). ■

We have formulated an equivalent DP problem for finding the capacity of the Ising channel and found the DP operator explicitly. The objective is to maximize the average reward, ρ_π, over all policies, π. According to Theorem 4, if we identify a scalar ρ and bounded function h that satisfy the Bellman Equation, $\rho + Th(z) = h(z)$, then ρ is the optimal average reward and, therefore, the channel capacity.

VI. SOLVING THE DP

In order to facilitate the search for the analytical solution, we first solve the problem numerically using a value iteration algorithm. The aim of the numerical solution is to obtain some basic knowledge of the bounded function, h, which satisfies the Bellman Equation.

A. Numerical Solution

We present the function $h(z)$ as found numerically using the value iteration algorithm. The value iteration algorithm generates a sequence of iterations according to $J_{k+1} = T(J_k)$, where T is the DP operator, as in (10), and $J_0 = 0$.

For each k and z, $J_k(z)$ is the maximal expected reward over k periods given that the system starts in state z. Since rewards are positive, $J_k(z)$ grows with k for each z. For each k, we define a differential reward function, $h_k(z) \triangleq J_k(z) - J_k(0)$.

For the numerical analysis, the interval $[0, 1]$ was represented with a 1000 points grid. The numerical solution after 20 value iterations is shown in Fig. 3. This figure shows the $J_{20}(z)$ function and the corresponding policies, $\gamma^*(z)$ and $\delta^*(z)$. The policies are chosen numerically such that the equation $T_{\gamma^*,\delta^*}h(z) \ge T_{\gamma,\delta}h(z)$ holds for all γ, δ on the grid, where $T_{\gamma,\delta}$ represents the DP operator restricted to the policy given by γ, δ. Moreover, Fig. 3 shows the histogram of z, which represents the relative number of times each point has been occupied by a state z. These values of z have been calculated using (11).

B. Analytical Solution via Numerical Results Analysis

Here, the numerical results are examined and an analytical solution is provided. We denote by $T_{\delta,\gamma}h(z)$ the expression $Th(z)$ without the supremum, restricted to the policy (δ, γ). The *optimal* policy is denoted by $\gamma^*(z), \delta^*(z)$ and the function $h(z)$, which satisfies the Bellman Equation, is denoted by $h^*(z)$. As seen from the histogram presented in Fig. 3, the states, z, alternate between four major points which are denoted by z_0, z_1, z_2, z_3. The variables γ^* and δ^* can be

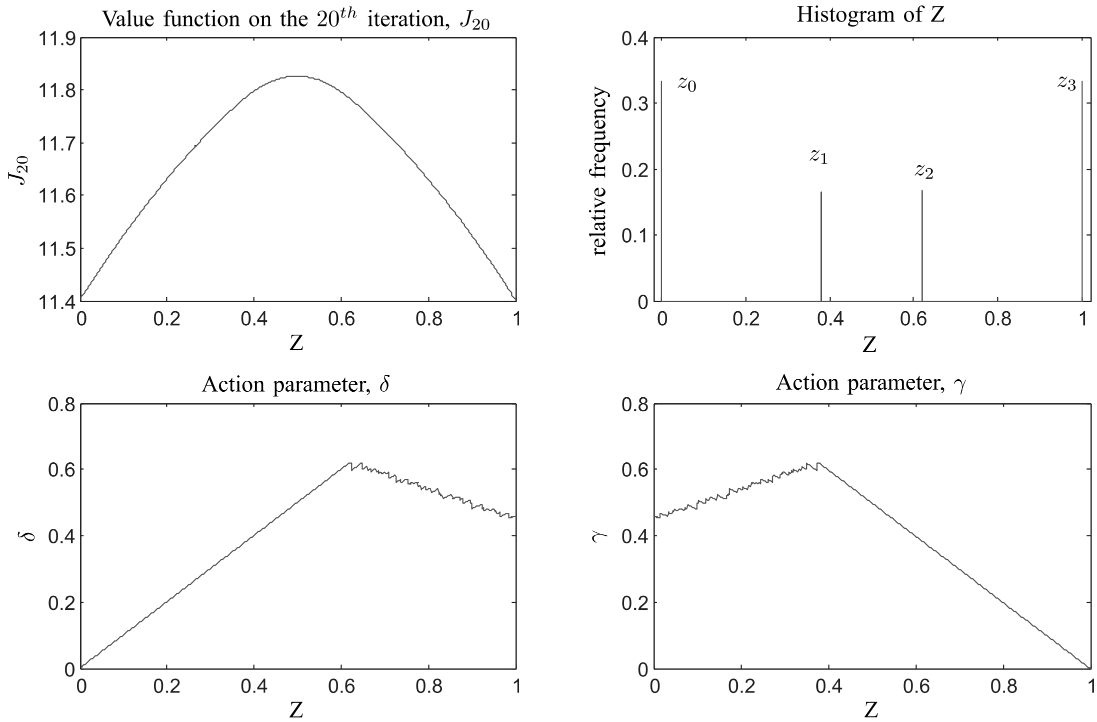

Fig. 3. Results from 20 value iterations. On the top-left, the value function J_{20} is illustrated. On the top-right, the approximate relative state frequencies are shown. At the bottom, the optimal action policies, δ^* and γ^*, are presented as obtained after the 20^{th} value iteration.

approximated by straight lines.

$$\gamma^*(z) = \begin{cases} a + az, & \text{if } z \in [z_0, z_1] \\ 1 - z, & \text{if } z \in [z_1, z_3] \end{cases} \tag{15}$$

$$\delta^*(z) = \begin{cases} z, & \text{if } z \in [z_0, z_2] \\ a(2 - z), & \text{if } z \in [z_2, z_3], \end{cases} \tag{16}$$

where $a = \gamma^*(0)$. Note that $\gamma^*(z) = \delta^*(1 - z)$.

Using (8), γ^*, and δ^*, one can obtain (z_i) $i = 0, 1, 2, 3$, as presented in Table IV. Note that the states, z_t, alternate between the points z_i. Moreover, if a scalar, ρ, and a function, h, solve the Bellman Equation, so do ρ and $h + c1$ for any scalar $c1$. Hence, with no loss of generality, we can assume $h^*(\frac{1}{2}) = 1$.

We now write the function $h^*(z)$ explicitly.

Lemma 3: Let $\gamma^*(z)$ and $\delta^*(z)$ be as in (17), (18), then

$$h^*(z) = \begin{cases} H_b(z), & \text{if } z \in [z_1, z_2] \\ \left(\frac{1}{1-a}\right) H\left(\frac{2a+(1-a)z}{2}\right) - z + \frac{az-4a-z}{2(1-a)}\rho^* \\ + \frac{2a+(1-a)z}{2(1-a)} H\left(\frac{2a}{a(2-z)+z}\right), & \text{if } z \in [z_2, z_3] \end{cases} \tag{17}$$

satisfies $T_{\delta^*,\gamma^*}h^*(z) = h^* + \rho^*$ $\forall z \in [z_0, z_3]$ where ρ^* is a constant satisfying $\rho^* = h(0) = h(1)$. We take $h^*(z)$ symmetric around $\frac{1}{2}$.

Proof: Lemma 3 follows from direct calculation of $h^*(z)$ when substituting γ, δ in (16) for δ^*, γ^* as in (17), (18). ■

In Fig. 4, the results using the numerical analysis are presented. The left-upper sub-figure presents the function $h^*(z)$. Notice the similarity between the function $h^*(z)$ as given in (19) and the function J_{20} as found using the value iteration algorithm and presented in Fig. 3.

TABLE IV
THE VARIABLES z_i, γ_i^*, δ_i^* CALCULATED USING (8)

$z_0 := 0$	$\gamma_0^* := a$	$\delta_0^* := 0$
$z_1 := \frac{1-a}{1+a}$	$\gamma_1^* := \frac{2a}{1+a}$	$\delta_1^* := \frac{1-a}{1+a}$
$z_2 := \frac{2a}{1+a}$	$\gamma_2^* := \frac{1-a}{1+a}$	$\delta_2^* := \frac{2a}{1+a}$
$z_3 := 1$	$\gamma_3^* := 0$	$\delta_3^* := a$

We now verify that the function $h^*(z)$, as in (19), together with $\rho^* = \frac{2H(a)}{3+a}$, satisfy the Bellman Equation, $Th^*(z) = h^*(z) + \rho^*$. Furthermore, we show that the selection of δ^*, γ^*, given in (17), (18), maximizes $Th^*(z)$. Namely, we prove Theorem 1.

We begin by proving the following lemma:

Lemma 4: The policies γ^*, δ^* which are defined in (17) and (18) maximize $Th^*(z)$, i.e. $T_{\delta^*,\gamma^*}h^*(z) = Th^*(z) \forall z \in [z_0, z_3]$, where $a \approx 0.4503$ is a root of the fourth-degree polynomial $x^4 - 5x^3 + 6x^2 - 4x + 1$.

The proof of Lemma 4 is found in the appendix. We now prove Theorem 1. In the proof we show that $Th^*(z) = h^*(z) + \rho^*$, $\rho^* = \frac{2H_b(z)}{3+a}$, δ, γ that maximize the operator T are δ^*, γ^*, and $h^*(z)$, as in (17)–(19). This implies that $h^*(z)$ solves the Bellman Equation. Therefore, ρ^* is the optimal average reward, which is equal to the channel capacity.

Proof of Theorem 1(a): We would like to prove that the capacity of the the Ising channel with feedback is $C_f = \left(\frac{2H(a)}{3+a}\right) \approx 0.5755$ where $a \approx 0.4503$ is a specific root of the fourth-degree polynomial $x^4 - 5x^3 + 6x^2 - 4x + 1$. According to Theorem 4, if we identify a scalar ρ and a bounded function

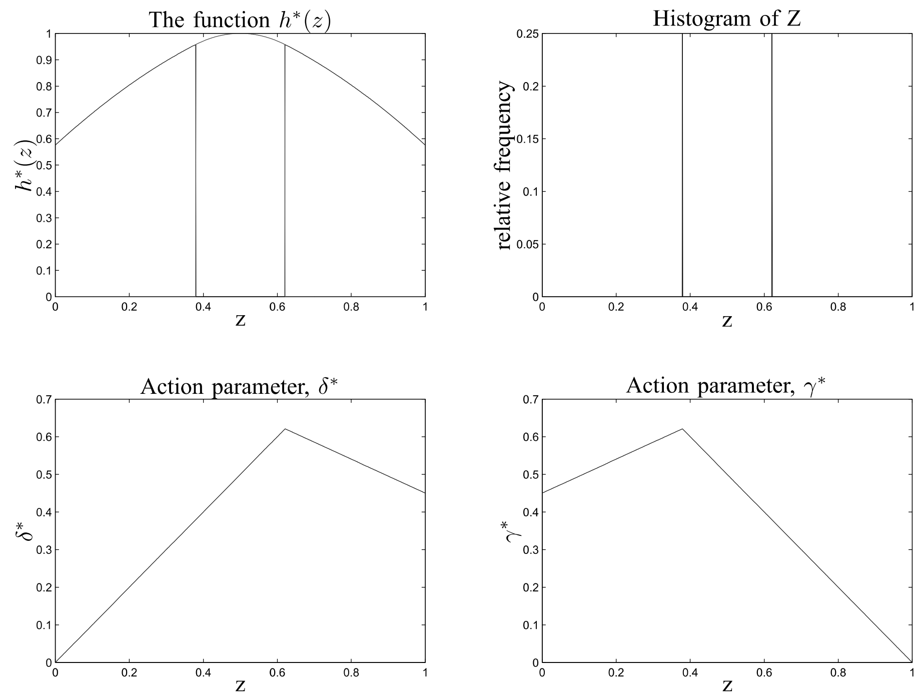

Fig. 4. The results using the numerical analysis. On the top-left, the value function $h^*(z)$ is shown. On the top-right, the assumed relative state frequencies are shown (the figure presents only the states). At the bottom, the optimal action policies δ^* and γ^* are illustrated. As can be seen, these results are similar to the results obtained numerically after the 20^{th} value iteration.

$h(z)$ such that

$$\rho + h(z) = \sup_{0\le\delta\le z, 0\le\gamma\le 1-z} H_b\left(\frac{1}{2}+\frac{\delta-\gamma}{2}\right)+\delta+\gamma \\ -1+\frac{1+\delta-\gamma}{2}h\left(1+\frac{\delta-z}{\delta+1-\gamma}\right) \\ +\frac{1-\delta+\gamma}{2}h\left(\frac{1-z-\gamma}{1+\gamma-\delta}\right) \quad (18)$$

then $\rho = \rho^*$. Using Lemma 4 we obtain that δ^* and γ^*, as defined in (17), (18), maximize $Th^*(z)$ when $a \approx 0.4503$ is a specific root of the fourth-degree polynomial $x^4 - 5x^3 + 6x^2 - 4x + 1$. In addition, we show that $h^*(z)$, which is defined in (19), satisfies the Bellman Equation, $Th^*(z) = h^*(z) + \rho^*$, where $\rho^* = \frac{2H(a)}{3+a}$. This follows from Lemma 3. Therefore, we have identified a bounded function, $h^*(z)$, and a constant, ρ^*, together with a policy, γ^*, δ^*, which satisfy the Bellman Equation. Thus, the capacity of the Ising channel with feedback is $\rho^* = h^*(0) = h^*(1) = \frac{2H(a)}{3+a} \approx 0.575522$. ■

Now we prove Theorem 1(b). The proof is based on algebraic manipulations.

Proof of Theorem 1(b): We define $g(z) = \frac{2H_b(z)}{3+z}$ and we calculate $g'(z) = \frac{8\log_2(1-z)-6\log_2(z)}{(3+z)^2}$. $8\log_2(1-z)-6\log_2(z) = 0$ iff $(1-z)^8 - z^6 = 0$. The polynomial $(1-z)^8 - z^6 = 0$ is reducible, hence we can write $(1-z)^8 - z^6 = (1 - 4z + 6z^2 - 3z^3 + z^4)(1 - 4z + 6z^2 - 5z^3 + z^4)$. Therefore, $g'(a) = 0$ since $a \approx 0.4503$ is the root of the polynomial $x^4 - 5x^3 + 6x^2 - 4x + 1$. It is easy to verify that $g'(a-\epsilon) > 0$ and $g'(a+\epsilon) < 0$. Together with the fact that a is the only real number in $[0, 1]$ that sets $g'(z)$ to zero, we conclude that a is a maximum point of $g(z)$, for $0 \le z \le 1$. ■

VII. Interrelation of the DP Results and the Coding Scheme

In this section we analyze the DP results and derive the coding scheme from these results. First, recall that in the histogram, which is presented in Fig. 3, z_t alternates between four points, two of which are 0 and 1. In order to keep in mind that these points stand for probability we denote them as p_0, p_1, p_2, p_3, where $p_0 = 0$ and $p_3 = 1$. Using (8) and the definition of γ^*, δ^* we can derive Table V. The table presents z_{t+1} as a function of z_t and y_{t+1}. It also presents the optimal action parameters, $u_t(1,1), u_t(2,2)$, for each state. The action parameters are calculated from the parameters δ^*, γ^*.

Assume that at time $t-1$ the state is $z_{t-1} = p_0 = 0$.

1) *Decoder:* Using the definition of z_{t-1} we deduce that $p(s_{t-1} = 0|y^{t-1}) = 0$ and hence $x_{t-1} = 1$ with probability 1. Thus, the decoder decodes 1.
2) *Encoder:* The optimal actions are $\delta_t^*(0) = 0$ and $\gamma_t^*(0) = a$. Using the definition of γ^* we conclude that $\Pr(x_t = 1|s_{t-1} = 1) = a$. Thus, $\Pr(x_t = 0|s_{t-1} = 1) = 1 - a$, which means that, given that $s_{t-1} = x_{t-1} = 1$, the probability to send 1 again is a. This result gives us the alternation probability from 1 to 0, which is $1 - a$. Since $s_{t-1} = x_{t-1} = 1$ with probability 1, the action parameter δ^* is irrelevant because it concerns the case in which $s_{t-1} = 0$. Indeed, using the definition of δ^*, we can see that $\delta_t^*(0) = 0$.

We now use Table V in order to find the next state. We have two options; if the output is 0 we move to state $p_3 = 1$. For this state the analysis is similar to the state p_0, switching between 0 and 1. Note that since the next state is

TABLE V

THE DP STATES AT TIME $t+1$ AS A FUNCTION OF THE PREVIOUS STATE AND THE OUTPUT CALCULATED USING (8). THE TABLE PRESENTS THE OPTIMAL ACTIONS FOR EACH STATE

	$z_t = p_0$	$z_t = p_1$	$z_t = p_2$	$z_t = p_3$
$y_t = 0$	$z_{t+1} = p_3$	$z_{t+1} = p_3$	$z_{t+1} = p_3$	$z_{t+1} = p_2$
$y_t = 1$	$z_{t+1} = p_1$	$z_{t+1} = p_0$	$z_{t+1} = p_0$	$z_{t+1} = p_0$
$u_{t+1}(2,2)$	a	1	1	irrelevant
$u_{t+1}(1,1)$	irrelevant	1	1	a

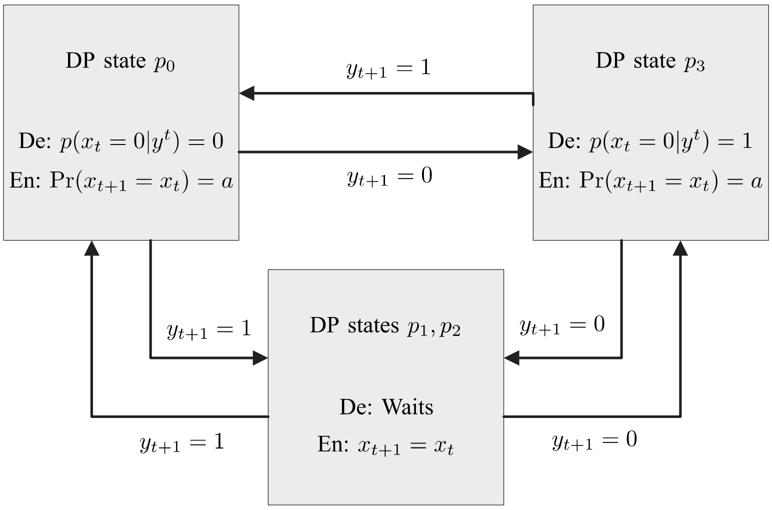

Fig. 5. The coding graph for the capacity-achieving coding scheme at time t. The states, p_i $i = 0, 1, 2, 3$ are the DP states where $p_0 = 0$, $p_3 = 1$. The labels on the arcs represents the output of the channel at time $t + 1$. The decoder and the encoder rules, which are written in vertices of the graph, yield the coding scheme presented in Theorem 2.

$p_3 = 1$ the decoder decodes the bit which was sent. If, on the other hand, the output is 1 we move to the state $z_t = p_1$. Assuming $z_t = p_1$ we now have the following:

1) *Decoder:* Using the definition of z_t, we deduce that $p(s_t = 0|y^t) = p_1$ and hence $x_t = 1$ with probability p_1. Thus, the decoder does not decode and waits for the next bit.
2) *Encoder:* The optimal actions are $\delta^*_{t+1}(p_1) = p_1$ and $\gamma^*_{t+1}(p_1) = 1 - p_1$. Using the definition of γ^* we conclude that $\Pr(x_{t+1} = 1|s_t = 1) = 1$ and using the definition of δ^* we conclude that $\Pr(x_{t+1} = 0|s_t = 0) = 1$. This means that $x_{t+1} = s_t = x_t$ with probability 1.

The analysis for state p_2 is done in a similar way.

We can now create a coding graph for the capacity-achieving coding scheme. Decoding only when the states are p_0 or p_3 results in a zero-error decoding. The coding graph is presented in Fig. 5. In the figure we have three vertices, which correspond to the DP states. At each vertex we mention the corresponding state or states, the decoder action, and the encoder action. The edge labels are the output of the channel. The edges from vertices p_0 to p_3 and vice versa corresponds to case (1.1) in Theorem 2 in the encoder scheme and to case (1.1) in Theorem 2 in the decoder scheme. The edges between vertices p_0 and p_1, p_2 and between p_3 and p_1, p_2 correspond to case (1.2) in Theorem 2 in the encoder scheme and to case (1.2) in Theorem 2 in the decoder scheme.

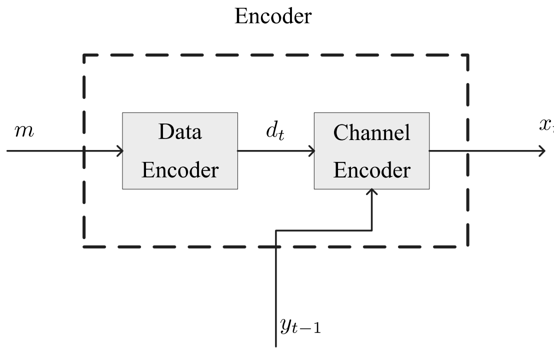

Fig. 6. The channel encoder block that consists of two sub-encoders. One block encodes data and the other performs the channel encoding.

VIII. CAPACITY-ACHIEVING CODING SCHEME ANALYSIS

In this section we prove that the coding scheme presented in Theorem 2, which was obtained from the solution of the DP, indeed achieves the capacity. In the proof, we calculate the expected length of strings in the channel input and divide it by the expected length of strings in the channel output.

Proof of Theorem 2: Let us consider an encoder that contains two blocks, as in Fig. 6. The first block is a data encoder. The data encoder receives a message $\tilde{M}^n$ ($\tilde{M} = \{0, 1\}$) of length n distributed i.i.d. Bernoulli $\left(\frac{1}{2}\right)$ and transfers it to a string of data, M^{nR} ($M = \{0, 1\}$), with probability of alternation from 1 and 0 and vice versa of q. This means that if some bit is 0 (alternatively 1), the next bit is 1 (alternatively 0) with probability q. In order to create a one-to-one correspondence between the messages and the data strings we need the data strings to be longer than n. We notice that $p(x_t|x^{t-1}) = p(x_t|x_{t-1})$. Thus, the entropy rate is

$$\lim_{n\to\infty} \frac{H(X^n)}{n} \overset{(a)}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i|X_{i-1}) = H(X_i|X_{i-1}) \overset{(b)}{=} H(q), \quad (19)$$

where (a) is due to the chain rule and since $p(x_t|x^{t-1}) = p(x_t|x_{t-1})$, and (b) follows since the probability of alternation is q. Therefore, given a message of length n, the data encoder transfers it into a data string of length $\frac{n}{H_b(q)}$. This can be done using the method of types for the binary sequence. Given a probability q and a binary sequences of length n, the size of the typical set is approximately $2^{nH_b(q)}$. Hence, we can map the set of Bernoulli $(\frac{1}{2})$ sequences of length n (of size 2^n) to the set of sequences of length $\frac{n}{H_b(q)}$ with alternation probability q (of size $2^{\frac{n}{H_b(q)}H_b(q)} = 2^n$). One can also use the mapping presented in [28]. This mapping gives a way to enumerate the indexes of Markov sequences of length $\frac{n}{H_b(q)}$ and of the binary sequences of length n, which are distributed Bernoulli $(\frac{1}{2})$. One can enumerate these sequences and establish a mapping from the Bernoulli sequences to the Markov sequences simply by matching their indexes.

The second block is the channel encoder. This encoder receives a data string in which the probability of alternation from 0 to 1 and vice versa is q. This sequence passes through the encoder, which sends some bits once and some bits twice according to the scheme mentioned in Theorem 2. Due to that property, the transmitted bit at time t is not necessarily the data bit at the tth location. This is why the encoder scheme uses two time indexes, t and t', which denote the data bit location and the current transmission time, respectively. The encoder works as mentioned in Theorem 2:

Now we calculate the expected length of the channel encoder output string. First, the message is of length n and distributed Bernoulli $(\frac{1}{2})$. Thus, the length of the string which has alternation probability of q is $\frac{n}{H(q)}$. Hence, we send one bit with probability $\frac{q}{2}$ and two bits with probability $\frac{2-q}{2}$. Therefore, the expected length of the channel encoder output string is $\frac{n}{H_b(q)}(2\frac{2-q}{2} + \frac{q}{2}) = \frac{4-q}{2}\frac{n}{H_b(q)}$. Since the message is of length n, the rate is $\frac{n}{\frac{4-q}{2}\frac{n}{H_b(q)}} = \frac{2H(q)}{4-q}$. Setting $q = 1 - a$ we achieve the rate $\frac{2H_b(1-a)}{3+a} = \frac{2H(a)}{3+a}$.

This holds for any $a \in [0, 1]$; in particular it holds for the unique positive root in $[0, 1]$ of the polynomial $x^4 - 5x^3 + 6x^2 - 4x + 1$. Using Theorem 1, the expression $\frac{2H(a)}{3+a}$ is equal to the capacity of the Ising channel with feedback. This means that the scheme achieves the capacity. ■

An interesting point is that in order to achieve the capacity using this coding scheme, we do not need to use the feedback continuously. It is enough to use the feedback only when there is an alternation from 0 to 1 (or vice versa) in the bits we send. When there is no alternation, the feedback is not needed since the bit is sent twice regardless of the channel output. Several cases of partial feedback use are studied in [29].

IX. Conclusions

We have derived the capacity of the Ising channel, analyzed it and presented a capacity-achieving coding scheme. As an immediate result of this work we can tighten the upper bound for the capacity of the one-dimensional Ising Channel to be 0.575522, since the capacity of a channel without feedback cannot exceed the capacity of the same channel with feedback. It may seem that the method presented in [12] and in this paper could be used to easily find the capacity of any unifilar FSC. But, from our experience, it is not an immediate result. This motivated us to try and characterize families of channels that could be solved analytically using this method.

In this paper, we also established a connection between the DP results and the capacity-achieving coding scheme. An interesting question that arises is whether there exists a general method for finding the capacity for a two-states channel with feedback, where the states are a function of the previous state, the input, and the previous output. It may transpire that the solution of the DP for such a channel has a fixed pattern. Recently, a new coding scheme was provided in [30] for unifilar finite state channels that is based on posterior matching.

Appendix

A. Proof of Lemma 4

In order to prove Lemma 4 we need several lemmas and corollaries. The first lemma regards the concatenation of continuous, concave functions.

Lemma 5: Let $f : [\alpha, \beta] \to \mathbb{R}$, $g : [\beta, \gamma] \to \mathbb{R}$ be two continuous, concave functions where $f(\beta) = g(\beta)$, $f'_-(\beta) = g'_+(\beta)$, where $f'_-(\beta)$ denotes the left derivative of $f(x)$ at β and $g'_+(\beta)$ denotes the right derivative of $g(x)$ at β. The function obtained by concatenating $f(x)$ and $g(x)$ defined by

$$\eta(x) = \begin{cases} f(x), & \text{if } x \in [\alpha, \beta] \\ g(x), & \text{if } x \in [\beta, \gamma] \end{cases} \quad (20)$$

is continuous and concave.

Sketch of Proof: Let us extend the function $f(x)$ on $[\beta, \gamma]$ by continuing it with a straight line with incline $f'_-(\beta)$. We denote the extended function as $f_1(x)$. Similarly, we extend the function $g(x)$ on $[\alpha, \beta]$ with a straight line with incline $g'_+(\beta)$ and denote it by $g_1(x)$. The functions $f_1(x), g_1(x)$ are concave since all the tangents are above the functions. We can define the function $\eta(x) = \min\{f_1(x), f_2(x)\}$ and since $f_1(x)$ and $g_1(x)$ are concave then $\eta(x)$ is concave as shown in [31, p. 72]. The sketch is shown in Fig. 7. ■

We now conclude that the function $h^*(z)$ is concave in z.

Corollary 5: The function $h^*(z)$ as given in (19) is continuous and concave for all $z \in [0, 1]$.

Proof: It is well known that the binary entropy function, $H_b(z)$, is concave in z. Thus, the function $h^*(z)$ for $z \in [z_1, z_2]$ is concave. In order to show that $h^*(z)$ for $z \in [z_0, z_1]$ and for $z \in [z_2, z_3]$ is concave we first observe that $\left(\frac{1}{1-a}\right)H\left(\frac{2a+(1-a)z}{2}\right)$ is concave since it is a composition

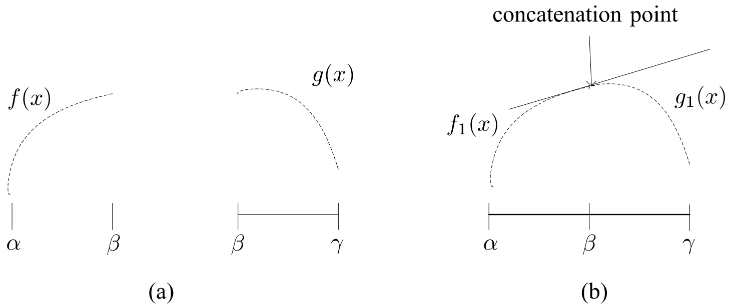

Fig. 7. Example for concatenation of two continuous, concave functions. It is easy to see from the figures the intuition behind the proof of Lemma (5).

of the binary entropy function and a linear, non-decreasing function of z. Second, the expression $-z + \frac{az-4a-z}{2(1-a)}\rho$ is also concave in z since it is linear in z, and third, the expression $\frac{2a+(1-a)z}{2(1-a)} H\left(\frac{2a}{a(2-z)+z}\right)$ is concave using the perspective property of concave functions [31, p. 89]. Hence, the sum of the three expression is also concave, which implies that $h^*(z)$ is concave in z for $z \in [z_0, z_1]$, $z \in [z_1, z_2]$, and for $z \in [z_2, z_3]$. It is easy to verify that the function $h^*(z)$ is a concatenation of three functions that satisfy the conditions in Lemma 5. Thus, we conclude that $h^*(z)$ is continuous and concave for all $z \in [0, 1]$. ■

Using the previous corollary we obtain the following.

Lemma 6: Let $h(z)$ be a concave function. The expression given in (21) that appears on the bottom of the page is concave in (δ, γ).

Proof: The binary entropy is a concave function. Hence, the expression $H\left(\frac{1}{2} + \frac{\delta-\gamma}{2}\right) + \delta + \gamma - 1$ is concave in (δ, γ). We now examine the expression $\frac{1+\delta-\gamma}{2} h\left(1 + \frac{\delta-z}{\delta+1-\gamma}\right)$. Let us denote $\eta_i = \frac{1+\delta_i-\gamma_i}{2}$, $i = 1, 2$. For every $\alpha \in [0, 1]$ we obtain that

$$\frac{\alpha\eta_1}{\alpha\eta_1 + (1-\alpha)\eta_2} h\left(1 + \frac{\delta_1 - z}{\eta_1}\right) + \frac{(1-\alpha)\eta_2}{\alpha\eta_1 + (1-\alpha)\eta_2} h\left(1 + \frac{\delta_2 - z}{\eta_2}\right) \overset{(i)}{\leq} h\left(1 + \frac{\alpha\delta_1 + (1-\alpha)\delta_2 - z}{\alpha\eta_1 + (1-\alpha)\eta_2}\right),$$

where (i) follows since $h(z)$ is concave. This result implies that

$$\alpha\eta_1 h\left(1 + \frac{\delta_1 - z}{\eta_1}\right) + (1-\alpha)\eta_2 h\left(1 + \frac{\delta_2 - z}{\eta_2}\right) \leq (\alpha\eta_1 + (1-\alpha)\eta_2)\, h\left(1 + \frac{\alpha\delta_1 + (1-\alpha)\delta_2 - z}{\alpha\eta_1 + (1-\alpha)\eta_2}\right).$$

Hence, $\frac{1+\delta-\gamma}{2} h\left(1 + \frac{\delta-z}{\delta+1-\gamma}\right)$ is concave in (δ, γ). It is completely analogous to show that the expression $\frac{1-\delta+\gamma}{2} h\left(1 + \frac{1-z-\gamma}{1-\delta+\gamma}\right)$ is also concave in (δ, γ). Thus, we derive that the expression given in (21) is concave in (δ, γ). ■

We now state the main lemma we need to obtain Lemma 4.

Lemma 7: The function $T_{\delta,\gamma} h^*(z)$ is concave in (δ, γ). Furthermore, from the KKT conditions the following optimality conditions hold:

(a) If $z \in [\frac{2a}{1+a}, 1]$, $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \delta}|_{\delta^*,\gamma^*} = 0$, and $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \gamma}|_{\delta^*,\gamma^*} > 0$ then δ^*, γ^* are optimal,

(b) If $z \in [\frac{1-a}{1+a}, \frac{2a}{1+a}]$, $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \delta}|_{\delta^*,\gamma^*} > 0$, and $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \gamma}|_{\delta^*,\gamma^*} > 0$ then δ^*, γ^* are optimal,

where γ^*, δ^* are given in (17), (18).

Proof: Corollary 5 states that $h^*(z)$ is a concave function; using Lemma 6 and the definition of $T_{\delta,\gamma} h^*(z)$ we conclude that $T_{\delta,\gamma} h^*(z)$ is concave in (δ, γ).

In order to prove the optimality conditions of Lemma 7 we first state the KKT conditions adjusted to our problem. The KKT conditions are stated in [31, p. 243]. Let $T_{\delta,\gamma} h^*(z)$ be the objective function. We consider the following optimization problem:

$$\max_{\delta,\gamma} T_{\delta,\gamma} h^*(z)$$

s.t.

$$\gamma - 1 + z \leq 0, \quad -\gamma \leq 0, \quad \delta - z \leq 0, \quad -\delta \leq 0.$$

The Lagrangian of $T_{\delta,\gamma} h^*(z)$ is $\mathcal{L}(\delta, \gamma, \lambda) = T_{\delta,\gamma} h^*(z) - \lambda_1(\gamma - 1 + z) + \lambda_2\gamma - \lambda_3(\delta - z) + \lambda_4\delta$. Since $T_{\delta,\gamma} h^*(z)$ is a concave function, then the following conditions are sufficient and necessary for optimality:

(1) $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \delta}|_{\delta^*,\gamma^*} = \lambda_3 - \lambda_4$.
(2) $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \gamma}|_{\delta^*,\gamma^*} = \lambda_1 - \lambda_2$.
(3) $\gamma \geq 0, \delta \geq 0$.
(4) $\gamma - 1 + z \leq 0, \delta - z \leq 0$.
(5) $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0$.
(6) $\lambda_1(\gamma - 1 + z) = \lambda_2\gamma = \lambda_3(\delta - z) = \lambda_4\delta = 0$.

The optimality conditions are derived from the KKT conditions and from the concavity of $T_{\delta,\gamma} h^*(z)$.

$$H\left(\frac{1}{2} + \frac{\delta - \gamma}{2}\right) + \delta + \gamma - 1 + \frac{1 + \delta - \gamma}{2} h\left(1 + \frac{\delta - z}{\delta + 1 - \gamma}\right) + \frac{1 - \delta + \gamma}{2} h\left(\frac{1 - z - \gamma}{1 + \gamma - \delta}\right) \tag{21}$$

First, we consider case (a) in which $z \in [\frac{2a}{1+a}, 1]$: if we take $\lambda_2 = \lambda_3 = \lambda_4 = 0$, $\lambda_1 = \frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \gamma}|_{\delta^*,\gamma^*}$, and $\gamma^* = 1 - z$, the KKT conditions hold since $\lambda_1 \geq 0$.

Second, we consider case (b) in which $z \in [\frac{1-a}{1+a}, \frac{2a}{1+a}]$: if we take $\lambda_2 = \lambda_4 = 0$, $\lambda_1 = \frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \delta}|_{\delta^*,\gamma^*}$, $\lambda_3 = \frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \gamma}|_{\delta^*,\gamma^*}$, $\delta^* = z$, and $\gamma^* = 1-z$, the KKT conditions hold since $\lambda_1 \geq 0$ and $\lambda_3 \geq 0$. ■

We are now ready to prove Lemma 4.

Proof of Lemma 4: From Lemma 7 we have that $T_{\delta,\gamma} h^*(z)$ is concave in (δ, γ). We now show that the optimality conditions in Lemma 7 holds. First, we assume that $z \in [\frac{2a}{1+a}, 1]$. We note that the expression $1 + \frac{\delta-\gamma}{1+\delta-\gamma}$ is in $[\frac{2a}{1+a}, 1]$. Furthermore, replacing γ, δ with γ^*, δ^* respectively, we find $\frac{1-z-\gamma}{1+\gamma-\delta}$ to be 0. We differentiate $T_{\delta,\gamma} h^*(z)$ with respect to δ and evaluate it in (δ^*, γ^*) to obtain $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \delta}|_{\delta^*,\gamma^*} = \frac{2\rho^* + \log_2(a)}{a-1}$. Note that if we set $z = 1$ in $h^*(z)$ and take $\rho^* = h^*(1)$ we obtain that $\rho^* = \frac{2H(a)}{a+3}$. Since $h^*(z)$ is symmetric around $\frac{1}{2}$ we have that $h^*(1) = \rho^* = h^*(0)$. Using basic algebra and substituting ρ^* with $\frac{2H(a)}{a+3}$ we find that the expression $\frac{2\rho^* + \log_2(a)}{a-1} = 0$ iff $a^4 - 5a^3 + 6a^2 - 4a + 1 = 0$. Thus, setting $a \approx 0.4503$ to be the unique real root in the interval $[0, 1]$ of the polynomial $x^4 - 5x^3 + 6x^2 - 4x + 1$ we establish that $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \delta}|_{\delta^*,\gamma^*} = \frac{2\rho^* + \log_2(a)}{a-1} = 0$.

Now, we differentiate $T_{\delta,\gamma} h^*(z)$ with respect to γ when $a \approx 0.4503$. We find that $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \gamma}|_{\delta^*,\gamma^*} > 0$. Note that the derivative is positive when $a \leq 0.9$. This can be seen since $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \gamma}|_{\delta^*,\gamma^*}$ is a monotonically increasing function of a, which for $a \leq 0.9$, is equal to zero for $z < \frac{2a}{1+a}$. Since we have found a to be approximately 0.4503 we have that $\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \gamma}|_{\delta^*,\gamma^*} > 0$. Using Lemma 7(a) we conclude that δ^*, γ^* are optimal. The analysis for $z \in [0, \frac{1-a}{1+a}]$ is completely analogous to the analysis made for $z \in [\frac{2a}{1+a}, 1]$.

Now, we assume $z \in [\frac{1-a}{1+a}, \frac{2a}{1+a}]$. In this case, we have that $h^*(z) = H_b(z)$. Using basic algebra we obtain that

$$\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \delta}|_{\delta^*,\gamma^*} > 0$$
$$\frac{\partial T_{\delta,\gamma} h^*(z)}{\partial \gamma}|_{\delta^*,\gamma^*} > 0. \quad (22)$$

Using Lemma 7 case (b) we conclude that δ^*, γ^* are optimal. Thus, for all $z \in [0, 1]$ we have that δ^*, γ^* are optimal. ■

Acknowledgments

The authors would like to thank the anonymous reviewers and Associate Editor, for their valuable and constructive comments, which helped to improve this paper.

References

[1] W. Lenz, "Beitrage zum verstandnis der magnetischen Eigenschaften in Festenkorpern," *Physikalische Zeitschrift*, vol. 21, pp. 613–615, Feb. 1920.

[2] E. Ising, "Beitrag zur theorie des ferromagnetismus," *Zeitschrift Phys. A Hadrons Nuclei*, vol. 31, no. 1, pp. 253–258, 1925.

[3] L. Onsager, "Crystal statistics. I. A two-dimensional model with an order-disorder transition," *Phys. Rev.*, vol. 65, nos. 3–4, pp. 117–149, Feb. 1944.

[4] T. Berger and F. Bonomi, "Capacity and zero-error capacity of Ising channels," *IEEE Trans. Inf. Theory*, vol. 36, no. 7, pp. 173–180, Jan. 1990.

[5] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[6] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[7] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[8] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, A. Feinstein, Ed. San Francisco, CA, USA: Holden-Day, 1964.

[9] Y. H. Kim, "Feedback capacity of the first-order moving average Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3063–3079, Jul. 2006.

[10] A. Goldsmith and P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[11] J. Chen and T. Berger, "The capacity of finite-state Markov channels with feedback," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 780–798, Mar. 2005.

[12] H. H. Permuter, P. W. Cuff, B. Van Roy, and T. Weissman, "Capacity of the trapdoor channel with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 7, pp. 3150–3165, Jul. 2008.

[13] J. Messy, "Causality, feedback and directed information," in *Proc. IEEE Int. Symp. Inform. Theory Appl. (ISITA)*, Nov. 1990, pp. 303–305.

[14] H. Marko, "The bidirectional communication theory—A generalization of information theory," *IEEE Trans. Inf. Theory*, vol. 21, no. 12, pp. 1335–1351, Dec. 1973.

[15] G. Kramer, "Capacity results for the discrete memoryless network," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 4–21, Jan. 2003.

[16] Y. H. Kim, "A coding theorem for a class of stationary channels with feedback," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1488–1499, Apr. 2008.

[17] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 323–349, Jan. 2009.

[18] H. H. Permuter, T. Weissman, and A. Goldsmith, "Finite state channels with time-invariant deterministic feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 644–662, Feb. 2009.

[19] H. H. Permuter, T. Weissman, and J. Chen, "Capacity region of the finite-state multiple access channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2455–2477, Jun. 2009.

[20] B. Shrader and H. Permuter, "Feedback capacity of the compound channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3629–3644, Aug. 2009.

[21] R. Dabora and A. Goldsmith, "The capacity region of the degraded finite-state broadcast channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1828–1851, Apr. 2010.

[22] H. H. Permuter and I. Naiss, "Extension of the Blahut-Arimoto algorithm for maximizing directed information," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2010, pp. 1442–1449.

[23] S. Yang, A. Kavcic, and S. Tatikonda, "Feedback capacity of finite-state machine channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 799–810, Mar. 2005.

[24] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.

[25] J. Ziv, "Universal decoding for finite-state channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 4, pp. 453–460, Jul. 1985.

[26] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. New York, NY, USA: Cambridge Univ. Press, 2005.

[27] A. Arapostathis, V. S. Borkar, E. Fernandez-Gaucherand, M. K. Ghosh, and S. Marcus, "Discrete time controlled Markov processes with average cost criterion—A survey," *SIAM J. Control Optim.*, vol. 31, no. 2, pp. 282–344, 1993.

[28] T. Cover, "Enumerative source encoding," *IEEE Trans. Inf. Theory*, vol. 19, no. 1, pp. 73–77, Jan. 1973.

[29] H. Asnani, H. Permuter, and T. Weissman, "To feed or not to feed back," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Aug. 2011, pp. 159–163.

[30] A. Anastasopoulos, "A sequential transmission scheme for unifilar finite-state channels with feedback based on posterior matching," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jul. 2012, pp. 2914–2918.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

Ohad Elishco received his B.Sc. degree in Electrical and Computer Engineering and his B.Sc. degree in Mathematics from Ben-Gurion University, Israel, in 2012, and his M.Sc. degree in Electrical and Computer Engineering from Ben-Gurion University, Israel, in 2013. He is currently a Ph.D student in Ben-Gurion University, Israel.

Ohad is a recipient of the Negev excellence scholarship from the Electrical and Computer Engineering department, Ben-Gurion University, Israel.

Haim Permuter (M'08–SM'13) received his B.Sc. (summa cum laude) and M.Sc. (summa cum laude) degrees in Electrical and Computer Engineering from the Ben-Gurion University, Israel, in 1997 and 2003, respectively, and the Ph.D. degree in Electrical Engineering from Stanford University, California in 2008.

Between 1997 and 2004, he was an officer at a research and development unit of the Israeli Defense Forces. Since 2009 he is with the department of Electrical and Computer Engineering at Ben-Gurion University where he is currently an associate professor.

Prof. Permuter is a recipient of several awards, among them the Fullbright Fellowship, the Stanford Graduate Fellowship (SGF), Allon Fellowship, and and the U.S.-Israel Binational Science Foundation Bergmann Memorial Award. Haim is currently serving on the editorial boards of the IEEE TRANSACTIONS ON INFORMATION THEORY.