Shannon Capacity is Achievable for a Large Class of Interactive Markovian Protocols

Assaf Ben-Yishai, Ofer Shayevitz and Young-Han Kim

***Abstract*—We address the problem of simulating a binary interactive protocol over a pair of binary symmetric channels with crossover probability ε. We are interested in the achievable rates of reliable simulation, i.e., in characterizing the smallest possible blowup in communications such that a vanishing error probability in the protocol length can be attained. We analyze the family of Mth-order Markovian protocols in which the transmission at every time depends only on the last M bits of the protocol. For $M = 1$ (first-order Markovian) we prove that all protocols can be simulated at Shannon's capacity. For $M > 1$ we characterize large classes of protocols that can be simulated at Shannon's capacity.**

I. Introduction

Let a protocol $\boldsymbol{\pi} = \pi_1, \pi_2, ..., \pi_n$, of length $|\boldsymbol{\pi}| = n$, be *interactively* generated by Alice and Bob where Alice generates her bits at predetermined times $i \in A \subseteq \{1, ..., n\}$ and Bob generates his bits at predetermined times $i \in B = \{1, ..., n\} \setminus A$. The bits of the protocol are calculated by the parties according to a set of transmission functions

$$\pi_i = \phi_i(\boldsymbol{\pi}^{i-1})$$

that are unknown to the counterpart and may be stochastic. In what follows, we conveniently identify the functions ϕ_i with their outputs π_i. We therefore think of $\boldsymbol{\pi}$ as both the protocol itself and the binary vector of its transcript.

We now assume that Alice and Bob are constrained to communicate through a pair of independent binary symmetric channels with crossover probability ε (BSC(ε)). We denote the input, output and noise of the channels by X, $\widetilde{X}$ and Z respectively. The channel input-to-output relation is thus

$$\widetilde{X} = X \oplus Z.$$

Z is a Bernoulli ε random variable ($0 < \varepsilon < \frac{1}{2}$) and $\oplus$ denotes addition over the binary field $\mathbb{F}_2$. We note that we restricted the order of speakers to be predetermined, since in a noisy environment an adaptive choice of speaker might lead an attempt of both parties to access the channel simultaneously.

We denote the BSC(ε) Shannon capacity by

$$\mathsf{C}_{\mathsf{Sh}}(\varepsilon) \triangleq 1 - h(\varepsilon)$$

where $h(\cdot)$ is the binary entropy function : $h(\varepsilon) \triangleq -\varepsilon \log \varepsilon - (1-\varepsilon)\log(1-\varepsilon)$, and $\log(x) \triangleq \log_2(x)$.

Alice and Bob wish to reconstruct $\boldsymbol{\pi}$ reliably by communicating over these noisy channels, i.e. with an error probability that vanishes with n. For this purpose they use an (n, N) coding scheme (or a *simulating protocol*) Σ, which operates on the protocol $\boldsymbol{\pi}$ and uses the channel N times. It consists of a partition $\tilde{A} \cup \tilde{B} = \{1, ..., N\}$ where $\tilde{A}$ (resp. $\tilde{B}$) is the set of time indices where Alice (resp. Bob) speaks. At time $j \in \tilde{A}$ Alice sends some function of her protocol functions $\{\pi_i\}_{i \in A}$ and of everything she has received so far from Bob. At time $j \in \tilde{B}$ Bob sends some function of his protocol functions $\{\pi_i\}_{iB}$ and of everything he has received so far from Alice. The rate of the scheme is $R = \frac{n}{N}$.

We denote by $\hat{\boldsymbol{\pi}}_A(\Sigma)$ and $\hat{\boldsymbol{\pi}}_B(\Sigma)$ the resulting simulations of the protocol, obtained by Alice and Bob respectively. The error probability attained by Σ when simulating $\boldsymbol{\pi}$ is therefore

$$P_e(\Sigma, \boldsymbol{\pi}) \triangleq \Pr\left(\hat{\boldsymbol{\pi}}_A(\Sigma) \neq \boldsymbol{\pi} \vee \hat{\boldsymbol{\pi}}_B(\Sigma) \neq \boldsymbol{\pi}\right).$$

A rate R is called *achievable* if there exists a sequence Σ_n of (n, N_n) coding schemes where $N_n \leq \frac{n}{R}$, and such that

$$\lim_{n\to\infty} \sup_{\pi:|\pi|=n} P_e(\Sigma_n, \boldsymbol{\pi}) = 0.$$

The interactive capacity, $\mathsf{C}_{\mathsf{I}}(\varepsilon)$, is the supremum of all achievable rates for a given crossover probability ε. We note that the above capacity definition does not use the notion of communication complexity as is usually done in the literature (see discussion in [1]).

We note that $\mathsf{C}_{\mathsf{I}}(\varepsilon)$ cannot exceed $\mathsf{C}_{\mathsf{Sh}}(\varepsilon)$, since it is the upper bound for the trivial special case in which Alice and Bob exchange independent bits, non-interactively.

The problem of simulating an interactive protocol over a noisy channel was originally introduced and studied by Schulman [2], [3]. In these seminal works, he showed

A. Ben-Yishai and O. Shayevitz are with the Department of EE–Systems, Tel Aviv University, Tel Aviv, Israel. Y.-H. Kim is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093 USA. Emails: {assafbster@gmail.com, ofersha@eng.tau.ac.il, yhk@ucsd.edu} The work of A. Ben-Yishai was supported by an ISF grant no. 1367/14. The work of O. Shayevitz was supported by an ERC grant no. 639573.

that interactive protocols can be simulated over BSC in a rate that approaches a constant fraction of $\mathsf{C_{Sh}}(\varepsilon)$.

Kol and Raz [4] further studied the problem in the limit of $\varepsilon \to 0$ and introduced a scheme achieving a rate of $1 - O(\sqrt{h(\varepsilon)})$, for protocols with alternating speakers (or where the order of speakers is periodic with a short period). They also showed that for a larger class of protocols, the rate is upper bounded by $1-\Omega(\sqrt{h(\varepsilon)})$. This demonstrated a separation between one-way and interactive communications. In order to better understand the gap between the one-way and interactive setups for $\varepsilon \to 0$, Haeupler and Velingker [5] considered a more restrictive family of protocols that are "less interactive", where Alice and Bob have some limited average lookahead, i.e., can often speak for a while without requiring further input from their counterpart (hence, can use short error correcting codes). They showed that when this average lookahead is poly$(1/\varepsilon)$ a rate of is $1 - O(h(\varepsilon))$ can be achieved, i.e., is order-wise the same as the one-way capacity.

In this work, rather than restricting the "interactiveness" of the protocol as above, we restrict the *memory* of the protocol. We start by defining the notions of interactive rate and capacity to families of protocols. Let Π_n be a sequence of families of length n protocols. With slight abuse of notation we refer to Π_n as a family of protocols. A rate R is called *achievable* for Π_n if there exists a sequence Σ_n of (n, N_n) coding schemes where $N_n \leq \frac{n}{R}$, and such that

$$\lim_{n\to\infty} \sup_{\pi\in\Pi_n} P_e(\Sigma_n, \boldsymbol{\pi}) = 0.$$

The interactive capacity of Π_n, is the supremum of all such achievable rates. The family of protocols examined in this paper is the family of Mth-order *Markovian protocols*. For these protocols the order of speakers is alternating (i.e. Alice speaks at odd time indexes and Bob speaks on even time indexes) for which the lookahead can be as short as 1 (highly interactive), but where Alice and Bob need only know the last M bits of the protocol in order to generated their transmission. We show that a) the interactive capacity of first-order Markovian protocols coincides with the Shannon capacity, and b) For any $M > 1$, there exists a subset containing almost all Mth-order Markovian protocols, for which the interactive capacity coincides with the Shannon capacity. This contribution extends our previous results form [6], by placing the capacity achieving scheme for first-order Markovian in a wider framework that also applies for a large class of higher order Markovian protocols.

II. Main results

We start with defining the Markovian protocols:

Definition 1. *A protocol $\boldsymbol{\pi}$, is called "Mth-order Markovian" if the order of speakers is alternating (Alice speaks at odd time indexes and Bob speaks at even time indexes) and the transmission function at time i depend only on the previous M bits of the protocol. Namely,*

$$\pi_i = \phi_i(\boldsymbol{\pi}_{i-M}^{i-1}).$$

In particular, in a first order Markovian protocol

$$\pi_i = \phi_i(\pi_{i-1}).$$

The following theorem, proved in Section IV applies to all first-order Markovian protocols:

Theorem 1. *The interactive capacity of the family of first-order Markovian protocols coincides with the Shannon capacity.*

In order to extend this result to higher orders, we need the following definitions. We denote by $\mathcal{F}_M$ the set of all boolean functions $\{0,1\}^M \mapsto \{0,1\}$, $|\mathcal{F}^M| = 2^{2^M}$. By definition, in an Mth-order Markovian protocol, all transmission functions $\phi_i \in \mathcal{F}_M$. An example to this notion is given in Table I which contains all sixteen functions of $\mathcal{F}_2$. We continue with the following definition:

Definition 2. *A set of boolean functions $\mathcal{P}_M \subseteq \mathcal{F}_M$ is called "useful" if for every pair of distinct inputs $\mathbf{x}, \mathbf{x}' \in \{0,1\}^M$, $\mathbf{x} \neq \mathbf{x}'$ there exist at least one function $f \in \mathcal{P}_M$ for which $f(\mathbf{x}) = f(\mathbf{x}')$.*

An example to a non-useful set of functions is given in Table II in which the outputs related to 00 are zero for all functions, and the outputs related to 11 are one for all functions. Observe that $\mathcal{F}_M$ is a useful set for every M, since it contains two functions that give the same outputs for *all* inputs (see for example f_0 and f_{15} in Table I). We refer to these function as *constant functions*. Another interesting set of useful functions is the set of *balanced* boolean functions, which is the set containing all the boolean functions $\{0,1\}^M \mapsto \{0,1\}$ for which the output is one for exactly half of the inputs.

Theorem 2. *Let Π_n be the family of all M-th order Markovian protocols whose functions are taken from some fixed given set of useful Boolean functions. Then there exists a subset $S_n \subseteq \Pi_n$ such that $|S_n|/|\Pi_n| = 1 - o(1)$, for which the interactive capacity is equal to the Shannon capacity.*

The proof is in Section V.

III. General concept of the coding schemes

The coding schemes used in this paper are based on concept of *block-wise interaction*, explained as follows: Assume that the n transmissions of the an M-th order Markovian protocol are divided into l blocks of length m as depicted in Table III (m is assumed to be even, throughout). By the Markovity assumption, in order to properly simulate the protocol at the beginning of every block, the parties do not need the entire past of the

input	f_0	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}
00	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
01	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
10	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
11	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

TABLE I
$\mathcal{F}_2$, THE SET OF ALL BOOLEAN FUNCTIONS WITH TWO INPUTS

input	f_1	f_3	f_5
00	0	0	0
01	0	0	1
10	0	1	0
11	1	1	1

TABLE II
AN EXAMPLE TO A NON-USEFUL SET OF FUNCTIONS

protocol, but rather, only $\boldsymbol{\pi}_{jm-M+1}^{jm}$ at Alice's side and $\boldsymbol{\pi}_{jm-M+2}^{jm}$ at Bob's side.

We denote the initial state of block j by

$$S_j \triangleq \boldsymbol{\pi}_{jm-M+1}^{jm},$$

and assume that without loss of generality, the initial state of the protocol is $\boldsymbol{\pi}_{-M+1}^{0} = \mathbf{0}^M$. We now observe that if Alice and Bob know the initial states of all blocks (i.e. S_j for all $1 \leq j \leq l$) prior to the protocol simulation, then they can simulate the blocks of the protocol *independently* and *simultaneously*. Having this notion at hand, both parties can use *vertical block coding*, as illustrated in Table III. Namely, knowing the initial states of all blocks, Alice can calculate the bits $\pi_1, \pi_{m+1}, \ldots, \pi_{n-m+1}$ and send them to Bob using a capacity achieving block code. Bob can decode the block code, calculate $\pi_2, \pi_{m+2}, \ldots, \pi_{n-m+2}$ and send these bit to Alice, also using capacity achieving block code, and so on.

We now use the following basic lemma which is proved in Appendix A:

Lemma 1 (Block-wise transmission)**.** *Suppose $l(n)$ independent blocks of $b(n)$ bits are to be conveyed over BSC(ε) at rate $R < \mathsf{C}_{\mathsf{Sh}}(\varepsilon)$ and $n \to \infty$. Then, if $b(n) = \Omega(\log(l(n)))$, the probability of error in the decoding of one or more blocks is $o(1)$.*

From this point on, we set $m = l = \sqrt{n}$, so the condition in the lemma is satisfied. So, if the parties know all the initial states of all blocks prior to the simulation, they can reliably simulate the protocol at any $R < \mathsf{C}_{\mathsf{Sh}}(\varepsilon)$ and achieve capacity. However, it is clear that in the general case, the parties do not know the initial states prior to the protocol simulation. Before we proceed with the coding scheme, we introduce the following definition:

Definition 3. *The "path" of block j and initial state S is denoted by $\boldsymbol{\Psi}(j, S)$. It contains the simulation using the transmission functions of the protocol $\boldsymbol{\pi}$, but assuming that the initial state of the block is $\mathrm{bin}(S)$ (i.e. a binary vector of length M containing the binary representation of S). Namely, $\boldsymbol{\Psi}(j, S)$ is simulated as follows:*

- *Initialization:* $\boldsymbol{\Psi}(j, S)_{-M+1}^{0} = \mathrm{bin}(S)$
- *Iteration: for* $i = 1$ *to* m, $\boldsymbol{\Psi}(j, S)_i = \phi_{(j-1)m+i}(\boldsymbol{\Psi}(j, S)_{i-M}^{i-1})$

A naive coding scheme for an M-th order Markovian protocol would be to simulate all paths for all blocks. After all paths are known, the protocol can be generated by going from the first block to the last, and for every new block choosing the path whose initial state corresponds to the last bits of the previous block. Lemma 1 can be applied by simulating $\boldsymbol{\Psi}(S, j)$ sequentially, for all S going from 0 to $2^M - 1$. In this case, the length of the vertical block code would still be $\sqrt{n}$ and the number of vertical blocks would be $2^M\sqrt{n}$, so the condition of the lemma is satisfied, and the transmission of all blocks can be made reliable at any rate smaller than $\mathsf{C}_{\mathsf{Sh}}(\varepsilon)$. However, due to the exhaustive simulation of all the paths for all blocks and initial states, the number of channel uses consumed by the scheme is $N = 2^M n/R$ and therefore the overall rate of the scheme can not exceed $2^{-M}\mathsf{C}_{\mathsf{Sh}}(\varepsilon)$.

We now give two examples for coding schemes in which all paths for all blocks and initial states are simulated $N = (n + o(n))/R$ channel uses for any $R < \mathsf{C}_{\mathsf{Sh}}(\varepsilon)$, hence capacity can be achieved.

IV. CAPACITY ACHIEVING SCHEME FOR FIRST-ORDER MARKOVIAN (PROOF OF THEOREM 1)

We start by presenting the set of all boolean functions with a single input in Table IV. We note that the functions f_0 and f_3 are *constant*. Namely, they give the same output for both inputs. The functions f_1 and f_2 are *linear*. Namely, $f_1(x) = x + 0$ and $f_2(x) = x + 1$ where $\oplus$ denotes addition over the binary field $\mathbb{F}_2$. In particular, observe that $f_1(1 + x) = 1 + f_1(x)$ and $f_2(1 + x) = 1 + f_2(x)$.

The idea behind the capacity achieving coding scheme is now demonstrated by the example in Table V. Suppose that for a specific block j, we are given $\boldsymbol{\Psi}(j, 0)$, and would like to calculate $\boldsymbol{\Psi}(j, 1)$. Suppose we also know that the first constant function is ϕ_6 and all the functions preceding it are non-constant (i.e. linear). Since ϕ_6 is constant, $\boldsymbol{\Psi}(j, 0)_6 = \boldsymbol{\Psi}(j, 1)_6$. In addition, since the model is first-order Markovian, both paths should

block #	initial state	transmissions				
1	$S_1 = \boldsymbol{\pi}^0_{-M+1}$	π_1	π_2	$\cdots$	π_{m-1}	π_m
2	$S_2 = \boldsymbol{\pi}^m_{m-M+1}$	π_{m+1}	π_{m+2}	$\cdots$	π_{2m-1}	π_{2m}
$\vdots$	$\vdots$	$\vdots$				$\vdots$
l	$S_l = \boldsymbol{\pi}^{n-m}_{n-m-M+1}$	π_{n-m+1}	π_{n-m+2}	$\cdots$	π_{n-1}	π_n
speaker		A	B	$\ldots$	A	B
vertical block #		1	2	$\cdots$	$m-1$	m

TABLE III
BLOCK-WISE INTERACTION OF AN M-TH ORDER MARKOVIAN PROTOCOL

input	f_0	f_1	f_2	f_3
0	0	0	1	1
1	0	1	0	1

TABLE IV
$\mathcal{F}_1$: THE SET OF BOOLEAN FUNCTIONS WITH A SINGLE INPUT

	ϕ_1-ϕ_5 are non-constant					ϕ_6 is constant			
i	1	2	3	4	5	6	7	8	9
$\boldsymbol{\Psi}(j,0)$	0	1	1	0	0	**1**	**1**	**0**	**1**
$\boldsymbol{\Psi}(j,1)$	1	0	0	1	1	**1**	**1**	**0**	**1**
speaker	A	B	A	B	A	B	A	B	A

TABLE V
AN EXAMPLE FOR THE FIRST-ORDER MARKOVIAN SCHEME

coincide for every $i > 6$. For $i < 6$, it is known that all the function are linear. So, $\boldsymbol{\Psi}(j,1)_1 = 1 + \boldsymbol{\Psi}(j,0)_1$ and

$$\begin{aligned}\boldsymbol{\Psi}(j,1)_2 &= \phi_2(\boldsymbol{\Psi}(j,1)_1) = \phi_2(1 + \boldsymbol{\Psi}(j,0)_1) \\ &= 1 + \phi_2(\boldsymbol{\Psi}(j,0)_1) = 1 + \boldsymbol{\Psi}(j,0)_2\end{aligned}$$

and so on. So all in all, $\boldsymbol{\Psi}(j,1)_i = 1 + \boldsymbol{\Psi}(j,0)_i$ for all $1 \le i < 6$ and we can reconstruct $\boldsymbol{\Psi}(j,1)$ as shown in Table V.

We can now extend this example to a capacity achieving scheme for all first-order Markovian protocols:

- Simulate $\boldsymbol{\Psi}(j,0)$ for all blocks $1 \le j \le l$ using vertical coding.
- For every block, Alice and Bob exchange the location (with respect to the beginning of the block) of the first constant function.
- Both parties independently determine the location of the first mutual constant function in every block, denoted by k (the time index is with respect to the beginning of the block).
- For every block j the, parties calculate $\boldsymbol{\Psi}(j,1)$ as follows:
  For $i < k$: $\boldsymbol{\Psi}(j,1)_i = 1 + \boldsymbol{\Psi}(j,0)_i$
  For $i \ge k$: $\boldsymbol{\Psi}(j,1)_i = \boldsymbol{\Psi}(j,0)_i$.
- Having both $\boldsymbol{\Psi}(j,0)$ and $\boldsymbol{\Psi}(j,1)$ for all blocks, the parties can construct the entire protocol $\boldsymbol{\pi}$ by going from the first block to the last, and choosing the correct path according to the last bits of the previous block.

We are left with calculating the number of channel uses consumed by the scheme and its total rate. The parties need to simulate $\boldsymbol{\Psi}(j,0)$ for all blocks requiring n/R channel use. They also need to communicate the location of the first constant function in every block which requires $2\sqrt{n}(\lceil \log(\sqrt{n}+1)\rceil)/R$ channel uses. Note that the "+1" is used for the special symbol indicating the absence of a constant function in a block. Note that the locations of the first constant functions can also be communicated in parallel for all blocks, using a vertical block code. So, all in all, there are $\sqrt{n}+2(\lceil \log(\sqrt{n}+1)\rceil)$ block codes, and the condition of Lemma 1 is still satisfied. The total number of channel uses is thus

$$N = \frac{\sqrt{n}(\sqrt{n} + 2(\lceil \log(\sqrt{n}+1)\rceil))}{R} = \frac{n + o(n)}{R}$$

for any $R < \mathsf{C}_{\mathsf{Sh}}(\varepsilon)$, hence $\lim_{n\to\infty} n/N = R$ and capacity can be achieved. We note that an alternative proof was previously given in [6].

V. CAPACITY ACHIEVING SCHEME FOR MTH-ORDER MARKOVIAN (PROOF OF THEOREM 2)

For Mth-order Markovian with $M > 1$ the following coding scheme is proposed:

1) Set $p = o(m)$. For example, set $p = n^{1/4}$.
2) For every block $1 \le j \le l$, for the first p time instants, simulate all paths. Namely, simulate $\boldsymbol{\Psi}(j,S)_i$ for all $1 \le i \le p$ and $0 \le S \le 2^M - 1$.
3) For every block $1 \le j \le l$, for the remaining time instants, simulate only $\boldsymbol{\Psi}(j,0)$. Namely, simulate $\boldsymbol{\Psi}(j,0)_i$ for $p < i \le m$.
4) Build the protocol $\boldsymbol{\pi}$ using the simulated paths. Choose the appropriate path for the first p time instants in every block according to the last bits of $\boldsymbol{\Psi}(j,0)$ from the previous block.

The scheme is successful if the paths are simulated reliably and if all paths $\boldsymbol{\Psi}(j,S)$ coincide to the zero path $\boldsymbol{\Psi}(j,0)$ within the first p time instants in every block. The number of vertical blocks for Lemma 1 is $\sqrt{n}(1 + n^{1/4})$, so the condition of the lemma is still satisfied and all paths can be communicated reliably. The total number channel of uses is

$$N = \frac{\sqrt{n} \times (\sqrt{n} + (2^M - 1)n^{1/4})}{R} = \frac{n + o(n)}{R}$$

for any $R < \mathsf{C_{Sh}}(\varepsilon)$, so capacity can also be achieved.

We are left with proving that if the protocol is generated according to Theorem 2, the paths coincide with high probability, with respect to the random generation of the protocol. This property is proved here:

Proof. (of Theorem 2) We assume that the protocol functions are drawn i.i.d with positive probability from a set of useful boolean functions $\mathcal{P}_M$. We denote by $q > 0$ the probability of the least probable function in the set. Let us now analyze, for a specific block j, the probability that $\mathbf{\Psi}(j,0)$ and $\mathbf{\Psi}(j,1)$ do not coincide within the first p time instants: $\Pr(\mathbf{\Psi}(j,0)_{p-M+1}^{p} \neq \mathbf{\Psi}(j,1)_{p-M+1}^{p})$. We then bound the probability that all paths in all block do not coincide.

We note that due to the Markovity assumption, if for a specific i we have that $\mathbf{\Psi}(j,0)_{i-M+1}^{i} = \mathbf{\Psi}(j,1)_{i-M+1}^{i}$ (i.e. the paths are identical for M consecutive time instants) then $\mathbf{\Psi}(j,0)_{i+1}^{m} = \mathbf{\Psi}(j,1)_{i+1}^{m}$ (i.e. they remain identical until the end of the block). Therefore, the probability the path do not coincide within the first p time instants is the probability that

$$\Pr(\mathbf{\Psi}(j,0)_{p-M+1}^{p} \neq \mathbf{\Psi}(j,1)_{p-M+1}^{p}).$$

We start by bounding $\Pr(\mathbf{\Psi}(j,0)_1 = \mathbf{\Psi}(j,1)_1)$. By definition, $\mathbf{\Psi}(j,0)_1$ and $\mathbf{\Psi}(j,1)_1$ are the output of the function $\phi_{(j-1)m+1}$ for two distinct inputs $\mathrm{bin}(0)$ and $\mathrm{bin}(1)$. By the assumptions in the theorem, the probability of drawing a function that gives the same output for any two distinct inputs is at least q. So

$$\Pr(\mathbf{\Psi}(j,0)_1 = \mathbf{\Psi}(j,1)_1) \geq q.$$

We continue the process of independently drawing the transmission functions and evaluating them on distinct inputs the following $M-1$ time instants, and conclude that

$$\Pr(\mathbf{\Psi}(j,0)_1^M = \mathbf{\Psi}(j,1)_1)_1^M \geq q^M.$$

We can continue this process until p in blocks of length M and give the following bound

$$\Pr(\mathbf{\Psi}(j,0)_{p-M+1}^{p} = \mathbf{\Psi}(j,1)_{p-M+1}^{p}) \geq 1 - (1-q^M)^{\frac{p}{M}}.$$

Note that the bound is pessimistic since we considered only the coincidence event which occur in non-overlapping blocks of length M. We use the inequality $(1-x)^y \leq \exp(-xy)$ for $x > 0$ and $y \in \mathbb{N}$ and obtain

$$\Pr(\mathbf{\Psi}(j,0)_{p-M+1}^{p} \neq \mathbf{\Psi}(j,1)_{p-M+1}^{p}) \leq \exp(-qp).$$

We now observe that the function drawing process is uniform with respect to all the inputs of the function, and therefore with respect to all the paths and all blocks. Therefore, for all $0 < S < 2^M - 1$ and $1 < j \leq l$ we have

$$\Pr(\mathbf{\Psi}(j,0)_{p-M+1}^{p} \neq \mathbf{\Psi}(j,S)_{p-M+1}^{p}) \leq \exp(-qp).$$

The probability that at least one path did not converge can therefore be bounded by the union bound

$$\Pr\left(\bigcup_{1<j\leq l}\bigcup_{1<S\leq 2^M-1} \mathbf{\Psi}(j,0)_{p-M+1}^{p} \neq \mathbf{\Psi}(j,S)_{p-M+1}^{p}\right)$$

$$\leq M2^M \exp(-qp) = M2^M \exp(-qn^{1/4}) = o(1).$$

□

Appendix A
Proof of Lemma 1

Proof. The proof is by straightforward implementation of Gallager's random coding error exponent and the union bound. Due to [7][Theorem 5.6.4], the probability of decoding error in a single block is upper bounded by:

$$\Pr(\text{block error}) \leq \exp\left(-\frac{b(n)}{R}E_r(R)\right)$$

where $E_r(R)$ (the error exponent) is strictly positive and independent of n for any $0 \leq R < \mathsf{C_{Sh}}(\varepsilon)$ and $b(n)/R$ is the length of the block code. Now, having $l(n)$ independent such blocks, the probability of error in one or more blocks can be upper bounded using the union bound:

$$\Pr(\text{error in any block}) \leq l(n)\exp\left(-\frac{b(n)}{R}E_r(R)\right)$$

$$= \exp\left(-b(n)\frac{E_r(R)}{R} + \ln l(n)\right) \overset{(a)}{=} e^{-\Omega(1)} = o(1)$$

where (a) is by the assumption that $b(n) = \Omega(\log(l(n)))$. □

References

[1] A. Ben-Yishai, Y-H Kim, O. Ordentlich and O. Shayevitz, "The interactive capacity of the binary symmetric channel is at least 1/40 the shannon capacity," in *ISIT 2019*.

[2] L. J. Schulman, "Communication on noisy channels: A coding theorem for computation," in *Proceedings., 33rd Annual Symposium on Foundations of Computer Science*. IEEE, 1992, pp. 724–733.

[3] L. J. Schulman, "Coding for interactive communication," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1745–1756, 1996.

[4] G. Kol and R. Raz, "Interactive channel capacity," in *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*. ACM, 2013, pp. 715–724.

[5] B. Haeupler and A. Velingker, "Bridging the capacity gap between interactive and one-way communication," in *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2017, pp. 2123–2142.

[6] A. Ben-Yishai, O. Shayevitz and Y.-H. Kim, "Shannon capacity is achievable for binary interactive first-order markovian protocols," *arxiv preprint arXiv:1801.01022*, 2017.

[7] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.