On Channel Resolvability in Presence of Feedback

Mani Bastani Parizi and Emre Telatar
Information Theory Laboratory (LTHI), EPFL, Lausanne, Switzerland
{mani.bastaniparizi,emre.telatar}@epfl.ch

Abstract—We study the problem of generating an approximately i.i.d. string at the output of a discrete memoryless channel using a limited amount of randomness at its input in presence of causal noiseless feedback. Feedback does not decrease the channel resolution, the minimum entropy rate required to achieve arbitrarily accurate approximation of an i.i.d. output string. However, we show that, at least over a binary symmetric channel, in presence of feedback, a significantly larger resolvability exponent (the exponential decay rate of the divergence between the output distribution and product measure), compared to the best known achievable resolvability exponent in a system without feedback, is possible. We show that by employing a variable-length resolvability scheme and using an average number of $R$ coin-flips per channel use, the average divergence between the distribution of the output sequence and product measure decays exponentially fast in the average length of output sequence with an exponent equal to $[R - I(U; V)]^+$, where $I(U; V)$ is the mutual information developed across the channel.

I. INTRODUCTION

Suppose $P_{V|U} : U \to V$ is a discrete memoryless channel, with input alphabet $U$ and output alphabet $V$. We wish to generate an i.i.d. string $V_1, V_2, \ldots$ distributed according to $P_V$ at its output. The obvious solution is to use an i.i.d. string $U_1, U_2, \ldots$ drawn from some distribution $P_U$, that induces $P_V$ at the output of the channel, at its input which requires an entropy rate of $H(U)$ bits per channel use (and results in a perfect i.i.d. output sequence). However, Wyner [1] observed that, if we accept an approximately i.i.d. sequence, a lower entropy rate of $I(U; V)$ bits per channel use is sufficient (and necessary). Indeed, he showed that if a random code of block-length $n$ and rate $R > I(U; V)$ is sampled from i.i.d. $P_U$ random coding ensemble from which a uniformly chosen codeword is transmitted via $n$ independent uses of the channel, with very high probability over the choice of the code, the normalized Kullback–Leibler divergence between the output distribution $P_{V^n}$ and the product distribution $P_{V^n}^n(v^n) = \prod_{i=1}^n P_V(v_i)$ can be made arbitrarily small by choosing $n$ sufficiently large. The problem of channel resolvability was later studied by Han and Verdú [2] and Hayashi [3], replacing the measure of approximation quality with total variation and unnormalized divergence, respectively.

**Definition 1.** A rate $R$ is achievable over the channel $P_{V|U} : U \to V$ and with respect to (w.r.t.) the reference measure $P_V$ if there exists a sequence of $(n, k)$ codes, i.e., deterministic encoding functions $E^n : \{0, 1\}^k \to U^n$, of rate at most $R$,

$$\limsup_{n \to \infty} \frac{k}{n} \leq R,$$

such that, with $U^n = E^n(W^k)$, $W^k$ uniformly distributed on $\{0, 1\}^k$, and $V^n$ being the output of $n$ independent uses of $P_{V|U}$ with input $U^n$, denoted hereafter as $P_{V|U}^n$,

$$\lim_{n \to \infty} D(P_{V^n} \| P_V^n) = 0. \tag{1}$$

**Definition 2.** The minimum of all achievable resolvability rates over the channel $P_{V|U}$ w.r.t. the reference measure $P_V$ is called the resolution of the channel $P_{U|V}$ (w.r.t. to $P_V$).

**Theorem 1** ([1]–[4]). The resolution of the channel $P_{V|U} : U \to V$ w.r.t. the reference measure $P_V$ equals:

$$\min_{P_U: \sum_u P_U(u)P_{V|U}(v|u) = P_V(v)} I(U; V). \tag{2}$$

Moreover, in [3]–[7] it has been shown that, in the above-mentioned context, the divergence between the distribution of a length-$n$ block of channel output sequence $P_{V^n}$ and product distribution $P_V^n$ decays exponentially fast in $n$ and in [8] the exact exponential decay rate of the ensemble-average of $D(P_{V^n} \| P_V^n)$ as a function of $R$ is characterized.

**Definition 3.** A pair $(R, E)$ is an achievable resolvability rate–exponent pair over the channel $P_{V|U} : U \to V$ w.r.t. the reference measure $P_V$ if there exists a sequence of $(n, k)$ codes $E^n : \{0, 1\}^k \to U^n$ of rate at most $R$,

$$\limsup_{n \to \infty} \frac{k}{n} \leq R,$$

such that, with $U^n = E^n(W^k)$, $W^k$ uniformly distributed over $\{0, 1\}^k$, and $V^n$ being the output of $P_{V|U}^n$ to input $U^n$,

$$\liminf_{n \to \infty} -\frac{1}{n} \log D(P_{V^n} \| P_V^n) \geq E. \tag{3}$$

**Theorem 2** ([8]). Suppose the encoder in Fig. 1 is a code of rate $R$ constructed randomly by sampling from i.i.d. $P_U$ random coding ensemble, $\{U^n(w^k) : w^k \in \{0, 1\}^k\}$, $k = [nR]$, and outputs $E^n(w^k) = U^n(w^k)$. Then (when $W^k$ is uniformly distributed on $\{0, 1\}^k$),

$$\lim_{n \to \infty} -\frac{1}{n} \log \left(\frac{D(P_{U|V} \| P_{U|V})}{D(P_{V^n} \| P_V^n)}\right) = \min_{Q_{UV}} D(Q_{UV} \| P_{UV}) + [R - f(Q_{UV} \| P_{UV})]^+ \tag{4}$$

Fig. 1. Channel Resolvability
Theorem 3. Feedback does not reduce the channel resolution.

Theorem 3 can be proved straightforwardly via standard techniques used to show that feedback does not increase the channel capacity (e.g. [9, Exercise 4.6]). For completeness, we give a proof in Appendix A.

Channel resolvability is, in a sense, the counterpart of channel coding. In the latter, the aim is to combat channel’s randomness and help the receiver distinguish the sent codeword from the others while in the former the goal is to exploit the channel randomness and make the output sequence as close as possible to random coin-flips. For channel coding, Burnashev [10] showed that, in presence of feedback (and using variable-length codes) higher error exponents are achievable. Thus, it is natural to ponder if the same holds for channel resolvability?

In this work, we give an affirmative answer to the above, at least when the channel \( P_{V|U} : U \rightarrow V \) is a binary symmetric channel (BSC) and the reference measure \( P_V \) is uniform on \( \{0, 1\} \). We show that in presence of causal feedback and using variable-length resolvability codes the straight-line exponent \( [R - I(U:V)]^+ \) is achievable.

II. PRELIMINARIES

A. Notation

We use uppercase letters (like \( U \)) to denote a random variable and the corresponding lowercase version \( (u) \) for a realization of that random variable. The same convention applies to the sequences, i.e., \( U^n = (u_1, \ldots, u_n) \) denotes a realization of the random sequence \( U^n = (U_1, \ldots, U_n) \). If \( S \) is a finite set, \( |S| \) denotes its cardinality. Given an alphabet \( \mathcal{A} \), \( \mathcal{A}^* \) denotes the set of all strings over symbols in \( \mathcal{A} \). Given a pair of real numbers \( a < b \), \( [a : b] \) denotes the set of integers between \( a \) and \( b \). For \( a \in \mathbb{R} \), \( [a]^+ \) denotes the maximum of \( a \) and 0.

Binary divergence \( d_2(\cdot \| \cdot) \), binary entropy function \( h_2(\cdot) \), and binary capacity function \( c_2(\cdot) \) are defined, respectively as

\[
D(P_V || P^0_V) = \text{ensemble-average of } D(P_V || P^0_V),
\]

\[
f(Q_{UV} || P_{UV}) \triangleq \sum_{u,v} Q_{UV}(u,v) \log \frac{P_{UV}(u,v)}{P_V(v)},
\]

and \( P_V(v) = \sum_u P_V(u)P_{U|V}(v|u) \).

Remark. The achievability of the exponent (4) was shown in [5–7] and its exactness is established in [8]. To the extent of our knowledge, the exponent of (4) is the best achievable resolvability exponent reported so far in the literature.

In this paper we consider the problem of channel resolvability in presence of causal feedback, namely, when the encoder gets to know the past received symbols \( V^{i-1} \) before transmitting the \( i^\text{th} \) symbol \( U_i \) and, hence, have the opportunity of deciding about the value of \( U_i \) based on the past behavior of the channel (see Fig. 2). Feedback does not reduce the minimum achievable rate.

Theorem 3. Feedback does not reduce the channel resolution.

The classical channel resolvability problem is defined based on block codes. Namely, the aim is to make the distribution of a length-\( n \) block of the output \( P_{V^n} \) close to i.i.d. \( P^n_0 \) using a \((n,k)\) block code of rate at most \( R \) and \( k \leq nR \) coin-flips at the encoder. It is useful extend this notion to variable-length codes. Specifically, the encoder is confined to use only \( k \) coin-flips but is allowed to use the channel a variable number of times based on a stopping rule.

Definition 4. A \((*,k)\) variable-length resolvability code (or in short a \((*,k)\) resolvability code), in presence of feedback, over the input and output alphabets \((U, V)\) is defined via a collection of deterministic encoding functions

\[
E^{(k)}_n : \{0,1\}^k \times \mathcal{V}^{n-1} \rightarrow U \cup \{S\}, \quad n \in \mathbb{N},
\]

where \( S \) is a special output indicating the “stopping” event. Namely, given the input word \( w^n \) and the past channel output symbols \( V^{n-1} \), the encoding function \( E^{(k)}_n \) decides either to feed the channel with an input symbol in \( U \) or stop the encoding (by outputting \( S \)).

Given a \((*,k)\) resolvability code, a \((*,k)\) feedback resolvability encoder maps the input word \( w^n \) into a channel input sequence \( U^n \) as follows:

1. \( n \leftarrow 1; \)
2. Initialize \( U^n \) and \( V^n \) to empty strings;
3. while \( E^{(k)}_n(w^n, V^{n-1}) \neq S \) do
4. \( U_n \leftarrow E^{(k)}_n(w^n, V^{n-1}); \)
5. Transmit \( U_n \) via the channel \( P_{U|V} : U \rightarrow V; \)
6. Observe and record \( V_n; \)
7. \( n \leftarrow n + 1; \)
8. end while

Remark 1. A \((n,k)\) block resolvability code is a special case of a \((*,k)\) variable-length resolvability code that assigns \( E^{(k)}_{n+1}(w^n, v^n) = S \) and \( E^{(k)}_{i+1}(w^i, v^{i-1}) \neq S, \forall i \in [1 : n] \) (for all input words \( w^n \) and feedback sequences \( v^n \)).

Remark 2. One could define a variable-length resolvability code (and hence encoder) for the system without feedback by dropping the dependence of the encoding functions in (8) to the feedback signal \( V^{n-1} \). Such a scheme, however, reduces to...
a codebook of $2^k$ codewords with variable (but deterministic) lengths and an encoder which maps each input word to the codeword it indexes.

Obviously, when a variable-length feedback resolvability encoder is employed, the *stopping time* of the encoder (and hence the length of the channel output corresponding to a single run of the encoder) will be a random variable, which we denote by $N_k$, that depends both on the channel randomness and the randomness of the input word $W^k$. We measure the performance of the system by the expected output divergence

$$D_k \triangleq \sum_n D(P_{V^n|N_k=n}||P_{V^n}^U) \Pr\{N_k = n\} \quad (9)$$

and the expected number of channel uses, $\mathbb{E}[N_k]$. Indeed, by the law of large numbers, when the resolvability scheme is run a large number of times (each corresponding to a block of channel output), the output sequence will have an average length of $\mathbb{E}[N_k]$ symbols per block and the average divergence between the distribution of each block of output sequence and the product distribution will be close to $D_k$. We can, hence, extend Definition 3 as:

**Definition 5.** A pair $(R, E)$ is an achievable resolvability rate–exponent pair over the channel $P_{V|U}: \mathcal{U} \rightarrow \mathcal{V}$ w.r.t. the reference measure $P_V$ if there exists a sequence of $(\ast, k)$ resolvability codes (see Definition 4) such that, when $W^k$ is uniformly distributed on $\{0,1\}^k$,

$$\limsup_{k \rightarrow \infty} \frac{k}{\mathbb{E}[N_k]} \leq R, \quad (10)$$

and, with $D_k$ defined as in (9),

$$\liminf_{k \rightarrow \infty} -\frac{\log D_k}{\mathbb{E}[N_k]} \geq E. \quad (11)$$

### III. Result

**Theorem 4.** In presence of feedback, the exponent

$$E_{s.l.}(p, R) = [R - c_2(p)]^+ \quad (12)$$

is achievable via a sequence of variable-length resolvability codes over a BSC with crossover probability $p$ with respect to the uniform reference measure $P_V(0) = P_V(1) = \frac{1}{2}$.

**Remark.** The straight-line exponent of (12) is larger than the exponent of (4) as the objective function of (4) equals $[R - I(U; V)]^+$ at $Q_{UV} = P_{UV}$ (see Fig. 3).

To prove Theorem 4, we propose the following sequence of $(\ast, k)$ resolvability codes and show that the exponent of (12) is achievable using this sequence of codes. Throughout the proof, without essential loss of generality, we assume $p < \frac{1}{2}$.

**Proposed Sequence of Codes:** Fix $\alpha > 0$. We define a $(\ast, k)$ code for each $k$ as follows: The collection of encoding functions $(E_n^{(k)}, n \in \mathbb{N})$ share a codebook of of size $2^k$ and infinite block-length indexed by length-$k$ binary sequences, considered.
Taking the expectation of the right-hand-side of (18), noting that $\mathbb{E}[S_{N_k}] = \mathbb{E}[S_{k/\alpha^*}] = 0$ (because a stopped martingale is also a martingale [11, Theorem 4, Chapter 7]), we get

$$\frac{k}{\mathbb{E}[N_k]} \leq \alpha c_2(p) + \alpha c''_2(p) \frac{\mathbb{E}[S_{N_k}^2/N_k]}{\mathbb{E}[N_k]} \quad (19)$$

It remains to examine the growth rate of the last term in (19). Note that if we replaced the stopping time $N_k$ with a fixed time $n$, the quantity of interest would have behaved like $\frac{1}{n}$ (since $\mathbb{E}[S_{N_k}^2/n]$ is a constant). It turns out that for a stopping time $N_k$, $\mathbb{E}[S_{N_k}^2/N_k]$ may not be a constant but will grow at least logarithmically in $N_k$. More precisely, using Lemma 2 in Appendix C we have

$$\mathbb{E}[S_{N_k}^2/N_k] \leq p(1 - p) \mathbb{E}[1 + \ln(N_k)], \quad (20)$$

Consequently,

$$\frac{k}{\mathbb{E}[N_k]} \leq \alpha c_2(p) + \alpha c''_2(p)p(1 - p) \frac{1 + \ln(\mathbb{E}[N_k])}{\mathbb{E}[N_k]} \leq \alpha c_2(p) + \alpha c''_2(p)(1 - p) \frac{1 + \ln(\mathbb{E}[N_k])}{\mathbb{E}[N_k]} \quad (a)$$

$$\leq \alpha c_2(p) + \alpha c''_2(p)(1 - p) \frac{1 + \ln(\mathbb{E}[N_k])}{k/\alpha} \quad (b),$$

where (a) follows from Jensen’s inequality and (b) as $\frac{1 + \ln(x)}{x}$ is decreasing for $x \geq 1$ and $N_k \geq \frac{k}{\alpha}$. Consequently,

$$\limsup_{k \to \infty} \frac{k}{\mathbb{E}[N_k]} \leq \alpha c_2(p). \quad (22)$$

To lower-bound $k/\mathbb{E}[N_k]$, we note that $\forall n > 1, \hat{Q}_n = \frac{n - 1}{n} \hat{Q}_{n-1} + \frac{1}{n} B_n$. Since $c_2(\cdot)$ is convex, at the stopping time,

$$c_2(\hat{Q}_n) \leq \frac{N_k - 1}{N_k} c_2(\hat{Q}_{N_k-1}) + \frac{1}{N_k} c_2(B_{N_k})$$

$$\leq \alpha^{-1} - \frac{1}{N_k} + \frac{1}{N_k} N_k = \alpha^{-1} k/N_k + \frac{1}{N_k}. \quad (23)$$

where $(*)$ follows from the stopping condition (15). Therefore, substituting $\hat{Q}_n = \frac{S_n}{N_n} + p$,

$$k > \alpha N_k c_2 \left( \frac{p + S_{N_k}}{N_k} \right) - \alpha$$

$$\geq \alpha c_2(p) N_k + \alpha c''_2(p) S_{N_k} - \alpha, \quad (24)$$

where the second inequality (24) follows from the lower bound of (17). Taking the expectation of the right-hand-side of (24) (and using the fact that $\mathbb{E}[S_{N_k}] = 0$ once again) we get,

$$\frac{k}{\mathbb{E}[N_k]} \geq \alpha c_2(p) - \frac{\alpha}{\mathbb{E}[N_k]} \geq \alpha c_2(p) - \frac{\alpha}{k} \quad (25)$$

where the second inequality follows since $N_k \geq k/\alpha$. Thus,

$$\liminf_{k \to \infty} \frac{k}{\mathbb{E}[N_k]} \geq \alpha c_2(p). \quad (26)$$

which, together with (22) concludes the proof.

To complete the proof of Theorem 4, it remains to bound the expected output divergence $D_k$ (9) for an appropriate code.

Let $c^{-1}_2(\cdot)$ denote the inverse of the binary capacity function $c_2(\cdot)$ (cf. (7)) when its domain is restricted to $[0, \frac{1}{2}]$ and define

$$q_k^*(n) = [k/\alpha : +\infty) \to [0, \frac{1}{2}]:$$

$$q_k^*(n) = \frac{1}{k} \alpha^{-1} k_{\alpha}. \quad (27)$$

Let $B^n = (B_1, \ldots, B_n)$ denote the flip pattern of $n$ independent uses of the channel and

$$B_n \triangleq \{ b^n \in \{0, 1\}^n : |B^n = b^n \} \subset \{ N_k = n \} \quad (28)$$

denote the set of flip patterns that stop the encoder at time $N_k = n$. Using the fact that the process $nQ_n = \omega_H(B^n)$ is an integer-valued process and the stopping condition (15) we can conclude that (among other constraints) $\forall b^n \in B_n$, either $\omega_H(b^n) = |nQ_k^*(n)|$ or $n - \omega_H(b^n) = |nQ_k^*(n)|$ (see Fig. 4).

Note that $B_n$ can be empty for some values of $n \in [k/\alpha : +\infty)$.\footnote{For example, if for some $n, \exists \ell \in \mathbb{N}$ such that $|n \ell - |n\ell^k(n)| = |n\ell^k(n)|$ then $B_n$ is empty because either the encoder stops at time $n$ or, if not, it will stop at some time $N_k > n$, because $\omega_H(B^n) > \omega_H(B^{n-\ell}) > |n\ell^k(n)|$ and similarly $n - \omega_H(B^n) < n - |n\ell^k(n)|$.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig4}
\caption{Encoder’s Stopping Time}
\end{figure}
Since $0 \leq p \leq \frac{1}{2}$, $\Pr\{B^n \in B^n_1\} \geq \Pr\{B^n \in B^n_2\}$. Hence,

$$\rho_n \triangleq \frac{\Pr\{B^n \in B^n_1\}}{\Pr\{B^n \in B^n_1\} + \Pr\{B^n \in B^n_2\}} \in [1/2 : 1].$$

Moreover, since $\{N_k = n\} = \{B^n \in B^n_1\} = \{B^n \in B^n_2\}$ and $B^n_1$ and $B^n_2$ are disjoint (by definition),

$$P_{V^n|N_k=n}(v^n) = \rho_n \Pr\{V^n = v^n|B^n \in B^n_1\} + (1 - \rho_n) \Pr\{V^n = v^n|B^n \in B^n_2\}. \quad (30)$$

Given the specification of the encoder, we have,

$$\Pr\{V^n = v^n, B^n \in B^n_1\} = \frac{1}{2^k} \sum_{u^* \in C_k} \Pr\{V^n = v^n, B^n \in B^n_1|U^n = u^n\}$$

$$= \frac{1}{2^k} \sum_{u^* \in C_k} \sum_{b \in B^n_1|B^n_1} \Pr\{V^n = v^n, B^n = b^n|U^n = u^n\}$$

$$= \frac{1}{2^k} \sum_{u^* \in C_k} \sum_{b \in B^n_1|B^n_1} 1\{v^n = b^n \oplus u^n\} \Pr\{B^n = b^n\}$$

$$(*) \Pr\{B^n \in B^n_1\} = \frac{1}{2^k} \sum_{u^* \in C_k} \sum_{b \in B^n_1|B^n_1} 1\{v^n = b^n \oplus u^n\},$$

where $(*)$ follows since $\Pr\{B^n = b^n\}$ only depends on $w_H(b^n)$ and all $b^n \in B^n_1$ have the same Hamming weight. As a consequence,

$$\Pr\{V^n = v^n|B^n \in B^n_1\} = \frac{1}{|B^n_1|2^k} \sum_{u^* \in C_k} 1\{u^n \oplus v^n \in B^n_1\}$$

$$= \frac{1}{|B^n_1|2^k} N_k(v^n|B^n_1) \quad (31)$$

where for any $A_n \subseteq \{0, 1\}^n$, we have defined

$$N_k(v^n|A_n) \triangleq |\{w \in \{0, 1\}^k : u^n \oplus w \in A_n\}|. \quad (32)$$

We, similarly, have

$$\Pr\{V^n = v^n|B^n \in B^n_2\} = \frac{1}{|B^n_2|2^k} N_k(v^n|B^n_2). \quad (33)$$

At this point, we are ready to bound the output divergence using the same method as in [7], [8]. Since $P_v^n(v^n) = 2^{-n}$, combining (31) and (33), together with the fact that $|B^n_1| = |B^n_2| = \frac{1}{2}|B^n|$ in (30), we get

$$L(v^n) \triangleq P_{V^n|N_k=n}(v^n) = \frac{P_v^n(v^n)}{P_v^n(v^n)} \Pr\{V^n = v^n\}$$

$${} = \frac{2^{n-k}}{|B^n|} \left[\rho_n N_k(v^n|B^n_1) + (1 - \rho_n) N_k(v^n|B^n_2)\right]. \quad (34)$$

We also recall that

$$D(P_{V^n|N_k=n}\|P_v^n) = \sum_{v^n} P_v^n(v^n) L(v^n) \log L(v^n). \quad (35)$$

Assume the code shared by the encoding functions $(E^{(k)}_n, n \in \mathbb{N})$ is sampled from i.i.d. random coding ensemble, namely, each codeword $U^\infty(w^k)$ is an infinite i.i.d. sequence of binary digits where each symbol is equally likely to take either value and the codewords are independent of each other. In this case, $\{N_k(v^n|B^n_1), N_k(v^n|B^n_2)\}$ forms a multinomial collection with cluster size $2^k$ and (equal) success probabilities $2^{-n-k}$.

Thus, it can immediately be verified that $\overline{L}(v^n) = 1$ (where $\overline{A}$ denotes the ensemble average of $A$).

As shown in [7], since $\overline{L}(v^n) = 1$, and $L(v^n) \leq 2^n$,

$$\overline{L}(v^n) \log \overline{L}(v^n) \leq \min\left\{n, \frac{1}{\ln(2)}(\overline{L}(v^n) - 1)^2\right\}. \quad (36)$$

Since $N_k(v^n|B^n_1)$ and $N_k(v^n|B^n_2)$ are negatively correlated,

$$(\overline{L}(v^n) - 1)^2 \leq 2(\rho_n^2 + (1 - \rho_n)^2) \frac{2^{-(k-n)}|B^n_1|}{|B^n|} \leq 2^{2-(k-n)} \quad (37)$$

Using (37) in (36) and the linearity of the expectation together with (35) we conclude that

$$D(P_{V^n|N_k=n}\|P_v^n) \Pr\{N_k = n\} \leq \min\left\{n, \frac{2}{\ln(2)} \frac{2^{-(k-n)}|B^n|}{|B^n|}\right\}. \quad (38)$$

Since $\Pr\{N_k = n\} = \Pr\{B^n \in B^n_1\} + \Pr\{B^n \in B^n_2\}$ and $\Pr\{B^n \in B^n_1\} \geq \Pr\{B^n \in B^n_2\}$ (cf. (29)),

$$\Pr\{N_k = n\} \leq 2 \Pr\{B^n \in B^n_1\} = 2|B^n| \frac{n q_k^n(n)1 - p}{n - q_k^n(n)}, \quad (39)$$

Multiplying the right-hand-sides of (38) and (39) we get

$$D(P_{V^n|N_k=n}\|P_v^n) \Pr\{N_k = n\}$$

$$\leq \kappa_1 \min\left\{\frac{n |B^n| q_k^n(n)1 - p}{n - q_k^n(n)}, \frac{2^{-(k-n) n q_k^n(n)}(1 - p)n(1 - q_k^n(n))}{2^{-(k-n) n q_k^n(n)}}\right\} \quad (a)$$

$$\leq \kappa_1 2^n f_2(q_k^n(n)) \min\left\{n |B^n| 2^{-(k-n) n q_k^n(n)}(1 - p)n(1 - q_k^n(n))\right\} \quad (b)$$

$$\leq \kappa_1 2^n f_2(q_k^n(n)) \min\left\{n 2^{-n q_k^n(n)}, 2^{-(k-n) n q_k^n(n)}\right\} \quad (c)$$

$$\leq \kappa_1 2^{-k n \max\{1, 1/n\}} n^2 2^n f_2(q_k^n(n)) \quad (40)$$

where $\kappa_1 = \frac{4}{\ln(2)} \frac{p}{1 - p}$, in (a) we have defined

$$f_2(q\|p) \triangleq 1 + q \log(p) + (1 - q) \log(1 - p), \quad (41)$$

(b) follows since $B_n$ is a subset of all binary sequences of length $n$ and Hamming weight $n q_k^n(n)$, and (c) by replacing $n = \frac{k}{\alpha f_2(q_k^n(n))}$. Replacing (40) into (9) (noting that the stopping rule and hence the stopping time $N_k$ is independent of the choice of the code) we get

$$\overline{L} \leq \kappa_1 2^{-k n \max\{1, 1/n\}} \sum_{n \in N_k} n^2 f_2(q_k^n(n)) \quad p. \quad (42)$$

Let

$$\tau_k \triangleq \frac{\log(2 - 1/n - 1 - k) - [\log(1 - 1/k) + 1]}{\log(p) - \log(1 - p)}, \quad (43)$$

so that $f_2(\tau_k\|p) = \log(1 - 1/k)$. It is easy to verify that $\tau_k$ is a decreasing sequence and $\tau_k \in (p: 1/2)$. Let $\mathcal{N}_k^p \triangleq \{n \in$
\( \mathcal{N}_k : q_k^*(n) < \tau_k \) and \( \mathcal{N}_k^2 \triangleq \{ n \in \mathcal{N}_k : q_k^*(n) \geq \tau_k \} \), and split the summation in the right-hand-side of (42) as

\[
\sum_{n \in \mathcal{N}_k} n 2^n f_2(q_k^*(n)) p \leq \sum_{n \in \mathcal{N}_k^2} n 2^n f_2(q_k^*(n)) p + \sum_{n \notin \mathcal{N}_k^2} n 2^n f_2(q_k^*(n)) p \tag{44}
\]

Since \( q_k^*(n) \) is increasing in \( n \),

\[
\sum_{n \in \mathcal{N}_k^2} n 2^n f_2(q_k^*(n)) p \leq \sum_{n \geq \tau_k} n \geq \tau_k 2^n f_2(q_k^*(n)) p \leq \sum_{n=0}^{\infty} n 2^n f_2(q_k^*(n)) p \leq 2 - f_2(q_k^*(n)) p \leq k(k-1) \tag{45}
\]

where (a) follows since we included \( n \notin \mathcal{N}_k \) in the sum as well, (b) since \( f_2(q_k^*(n)) p \) is decreasing in \( q_k^*(n) \) (thus the sum converges) and (c) since \( f_2(q_k^*(n)) p < 0 \) (thus the sum converges) and (d) by replacing \( f_2(q_k^*(n)) p = \log(1 - 1/k) \).

The first summation in (44) has (strictly) less than \( k \alpha c_2(\tau_k) \) terms where

\[
\tau_k \triangleq \lim_{k \to \infty} \tau_k = \frac{\log(1 - p) + 1}{\log(1 - p) - \log(p)}.
\]

Replacing \( n = \frac{k}{\alpha c_2(q_k^*(n))} \), we see that each term in the first summation of (44) is upper-bounded as

\[
n 2^n f_2(q_k^*(n)) p \leq \kappa_2(k) 2^{f_2(q_k^*(n)) p} \leq \kappa_2(k) 2^{k/\alpha} \tag{46}
\]

with equality iff \( q_k^*(n) = p \) (note that this term is included in the summation since \( \tau_k > p \)). Indeed, the last step follows since \( f_2(q_k^*(n)) p = c_2(q_k^*(n)) p \leq c_2(q) \) (with equality iff \( q_k^*(n) = p \)). Consequently,

\[
\sum_{n \in \mathcal{N}_k^2} n 2^n f_2(q_k^*(n)) p \leq \kappa_2(k) 2^{k/\alpha} \tag{47}
\]

Combining (45) and (47) (noting that the right-hand-side of (47) grows faster than that of (45)) shows that, for large \( k \),

\[
\mathbb{D}_k \leq 2 \mathcal{K}_1 \kappa_2(k) 2^{-k [\max\{1, \frac{1}{1 + \frac{1}{\alpha}}\} - 1]} = \kappa_3(k) 2^{-k [\frac{\alpha - 1}{\alpha} + 1]} \tag{48}
\]

where we have defined \( \kappa_3(k) \triangleq 2 \mathcal{K}_1 \kappa_2(k) \). Therefore, for at least half of the codes,

\[
\mathbb{D}_k \leq 2 \mathbb{D}_k \leq 2 \mathcal{K}_3(k) 2^{-k [\frac{\alpha - 1}{\alpha} + 1]} \tag{49}
\]

Since \( \lim_{k \to \infty} \frac{1}{k} \log(\kappa_3(k)) = 0 \), by picking any such good code for each \( k \) we will have a sequence of codes for which

\[
\lim_{k \to \infty} \frac{\log \mathbb{D}_k}{k} \geq \frac{[\alpha - 1] + \alpha}{\alpha} \tag{50}
\]

Equations (26) and (50) imply

\[
\lim_{k \to \infty} \frac{- \log \mathbb{D}_k}{k} = \lim_{k \to \infty} \frac{\log \mathbb{D}_k}{k} \geq [\alpha - 1] + \alpha \tag{51}
\]

Setting \( \alpha c_2(p) = R \) and substituting the value of \( \alpha \) in the above completes the proof of Theorem 4.

IV. CONCLUSION AND DISCUSSION

We studied the problem of channel resolvability in presence of feedback. We showed that, while feedback does not decrease the channel resolution, in presence of causal feedback higher resolvability exponents compared to the existing block resolvability codes of [3]–[8] are achievable. Although we restricted our discussion to resolvability over the BSC and with respect to the uniform target distribution, our results are easily extendable to general channels and target distributions.

Our results are the analogue of establishing the achievability of the error exponent \([I(U; V) - R]^+\) (over the channel \( P_{U|V} \)) in presence of feedback (see [12, Section 2.1] for example) in the channel coding setting. (Burnashev’s exponent [10] is also a straight-line but with a slope steeper than 1.) However, since, to the best of our knowledge, no non-trivial upper bounds on the highest achievable resolvability exponent at a specific rate \( R \) (i.e., an equivalent of sphere-packing exponent for channel coding) is known, it is unclear whether the improvement we demonstrated in this work is exclusively due to the presence of feedback or there might exist a resolvability scheme that achieves the straight-line exponent of (12) without the need for feedback. Nevertheless, the results of [8] show that an average i.i.d. random code cannot achieve a better resolvability exponent than (4). Thus, we can deduce that at least for the i.i.d. random coding ensemble the gains in the exponent are due to the presence of feedback.

Moreover, for the channel coding problem, Dobrushin [13] and Haroutunian [14] upper-bounded the best attainable error exponent in presence of feedback using block codes (This upper bound equals the sphere-packing exponent for symmetric channels [13] but is larger than that, for asymmetric ones [15, Exercise 10.36]). Those results, imply that employing variable-length error correcting codes is necessary to achieve the higher exponents of [10]. Another important question (somehow related to deriving non-trivial upper bounds to the best achievable resolvability exponent) is to examine in presence of feedback but using block resolvability codes to what extent the resolvability exponent can be improved and verify if the use of variable-length resolvability codes is necessary to achieve the straight-line exponent.

ACKNOWLEDGMENT

This work was supported by the Swiss National Science Foundation under grant number 200020_146832.
APPENDIX A

PROOF OF THEOREM 3

We prove the theorem under weak resolvability criteria, which implies that under strong resolvability criteria. Accordingly, suppose $\frac{1}{n}D(P_{V_n} \parallel P_V) \leq \epsilon$ (for arbitrarily small $\epsilon$ and, hence sufficiently large $n$), we have

$$\frac{k}{n} = \frac{1}{n}H(W) - \frac{1}{n}I(W; V^n)$$

and all terms in the right-hand-side of (53) are positive, $\frac{1}{n}D(P_{V_n} \parallel P_V) \leq \epsilon$ implies

$$\frac{1}{n}D(P_{V_n} \parallel \prod_i P_{V_i}) \leq \epsilon.$$  (54)

and

$$\frac{1}{n}D(P_{V_n} \parallel P_V) \leq \epsilon.$$  (55)

In particular (54) implies

$$\frac{1}{n}\left(\sum_{i=1}^{n} H(V_i) - H(V^n)\right) = \frac{1}{n}D(P_{V^n} \parallel \prod_{i=1}^{n} P_{V_i}) \leq \epsilon$$  (56)

Therefore, continuing (52) we have

$$\frac{k}{n} = \frac{1}{n}H(W) - \frac{1}{n}H(V_i|W^{i-1})$$

$$\geq \frac{1}{n} \sum_{i=1}^{n} \left( H(V_i) - H(V_i|U_i) \right) - \epsilon$$  (57)

Now, let $\delta_i \triangleq D(P_{V_i} \parallel P_{V_i})$ and $\delta_i \triangleq |P_{V_i} - P_V|$ and note that the uniform continuity of entropy (see [15, Chapter 2]) implies

$$|H(V_i) - H(V)| \leq \delta_i \log \frac{|V|}{\delta_i}.$$  (58)

Using the concavity of the function $\vartheta \mapsto \vartheta \log \frac{|V|}{\vartheta}$ we have

$$\frac{1}{n} \sum_{i=1}^{n} \left| H(V_i) - H(V) \right| \leq \frac{1}{n} \sum_{i=1}^{n} \delta_i \log \frac{|V|}{\delta_i} \leq \bar{\vartheta} \log \frac{|V|}{\bar{\vartheta}}$$  (59)

where we have defined $\bar{\vartheta} \triangleq \frac{1}{n} \sum_{i=1}^{n} \delta_i$. We also have

$$\bar{\vartheta} = \frac{1}{n} \sum_{i=1}^{n} \vartheta_i \leq \sqrt{2\delta_i} \leq \sqrt{2\delta_i} \leq \sqrt{2\epsilon}$$  (60)

and

$$\geq H(V) - \frac{1}{n} \sum_{i=1}^{n} H(V_i|U_i) - \epsilon$$

Taking $\epsilon \rightarrow 0$ (and observing that $H(V|U)$ is continuous in $P_U$) we get

$$\frac{k}{n} \geq \min_{P_U} \frac{\text{min}}{P_U: P_{V_n} = P_V} I(U; V).$$  (61)

APPENDIX B

PROOF OF (17)

The lower bound follows from convexity of $c_2(\cdot)$. To establish the upper bound, recall that $c_2(p) = \log \frac{p}{1-p}$ and $c_2'(p) = \frac{p}{p(1-p)}$. We, hence, have

$$c_2(p + \epsilon) - c_2(p) - c_2'(p) \epsilon$$

$$\leq (p + \epsilon) \log \frac{p + \epsilon}{p} + (1 - p - \epsilon) \log \frac{1 - p - \epsilon}{1 - p}$$

$$\leq \left( p + \epsilon \right) \frac{\epsilon}{p} + \left( 1 - p - \epsilon \right) \frac{\epsilon}{1 - p}$$

$$\leq \left( p + \epsilon \right) \frac{\epsilon}{p} + \left( 1 - p - \epsilon \right) \frac{\epsilon}{1 - p}$$

$$\leq \frac{1}{p(1-p)} \epsilon^2.$$  (62)

In the above ($*$) follows since log(1 + $\alpha$) $\leq \alpha$.  

It can be verified that if $f(x): D \rightarrow \mathbb{R}$ is convex and $l(x): D \rightarrow \mathbb{R}$ is a linear function of $x$, (on some convex domain $D$) then the mapping $y \mapsto \max_{x: f(x) \leq y} l(x)$ is concave in $y$.  

A USEFUL FACT ABOUT STOPPING TIMES

Lemma 2. Let \((\xi_n, n \in \mathbb{N})\) be i.i.d. zero-mean random variables and

\[ S_n = \sum_{i=1}^{n} \xi_i , \quad n \in \mathbb{N}. \]

Then the process \((S_n, n \in \mathbb{N})\) is a martingale with respect to the natural filtration \((\mathcal{F}_n = \sigma(\xi_1, \ldots, \xi_n), n \in \mathbb{N})\) and, if \(N\) is a stopping time,

\[ \mathbb{E}\left[ \frac{S_N^2}{N} \right] \leq \text{var}(\xi_1) \mathbb{E}[1 + \ln(N)]. \tag{63} \]

Proof: That \((S_n, n \in \mathbb{N})\) is a martingale is trivial. We shall only prove (63). Let

\[ N_m = \min\{N, m\}, \quad \forall m \in \mathbb{N}. \]

It is clear that for all \(m \in \mathbb{N}, N_m \in [1 : m]\) almost surely and \(N_m\) is a stopping time. The latter can be verified by noting that

\[ \{N_m = n\} = \begin{cases} \{N = n\} & \text{if } n < m, \\ \{N \geq m\} & \text{if } n = m. \end{cases} \tag{64} \]

Thus for \(n < m\), \(\{N_m = n\} = \{N = n\} \in \mathcal{F}_n\) by the hypothesis that \(N\) is a stopping time, and for \(n = m\),

\[ \{N_m = m\} = \{N \geq m\} = \bigcap_{j=1}^{m-1} \{N \neq j\} \in \mathcal{F}_{m-1}, \tag{65} \]

and \(\mathcal{F}_{m-1} \subseteq \mathcal{F}_m\) (hence \(\{N_m = n\} \in \mathcal{F}_m\)). Finally \(N_1 = 1\) almost surely, hence,

\[ \mathbb{E}\left[ \frac{S_{N_1}^2}{N_1} \right] = \text{var}(\xi_1). \tag{66} \]

We now have

\[ \mathbb{E}\left[ \frac{S_{N_m}^2}{N_m} \right] - \mathbb{E}\left[ \frac{S_{N_m-1}^2}{N_m-1} \right] = \mathbb{E}\left[ \left( \frac{S_m^2}{m} - \frac{S_{m-1}^2}{m-1} \right) 1\{N \geq m\} \right] \]

\[ = \mathbb{E}\left[ \left( \frac{(m-1)(\xi_m^2 + 2\xi_mS_{m-1} - S_{m-1}^2)}{(m-1)m} \right) 1\{N \geq m\} \right] \]

\[ \leq \frac{1}{m} \mathbb{E}[^{\text{var}(\xi_m)} 1\{N \geq m\}] + 2 \mathbb{E}[^{\text{var}(\xi_mS_{m-1})} 1\{N \geq m\}], \tag{67} \]

where \((*)\) follows from (66) and the fact that \(N \geq 1\) almost surely. We finally have

\[ \sum_{\ell \geq 1} \frac{\mathbb{Pr}\{N \geq \ell\}}{\ell} = \sum_{n \geq 1} \mathbb{Pr}\{N = n\} \sum_{\ell=1}^{n} \frac{1}{\ell} \leq \sum_{n \geq 1} \mathbb{Pr}\{N = n\}(1 + \ln(n)) = \mathbb{E}[1 + \ln(N)]. \tag{69} \]

Using the above in (68) yields

\[ \mathbb{E}\left[ \frac{S_{N_m}^2}{N_m} \right] \leq \mathbb{E}[1 + \ln(N)], \quad \forall m \in \mathbb{N}. \tag{70} \]

Now, since \(\lim_{m \to \infty} N_n = N\) with probability 1

\[ \mathbb{E}\left[ \frac{S_N^2}{N} \right] = \mathbb{E}\left[ \lim_{m \to \infty} \frac{S_{N_m}^2}{N_m} \right] = \mathbb{E}\left[ \liminf_{m \to \infty} \frac{S_{N_m}^2}{N_m} \right] \leq \mathbb{E}[1 + \ln(N)], \tag{71} \]

where in the above \((a)\) follows from Fatou’s lemma (applied to the sequence of non-negative random variables \(\frac{S_{N_m}^2}{N_m}, m \in \mathbb{N}\)), and \((b)\) from (70).

\[ \blacksquare \]

REFERENCES

[1] A. D. Wyner, “The common information of two dependent random variables,” IEEE Trans. Inf. Theory, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[2] T. S. Han and S. Verdú, “Approximation theory of output statistics,” IEEE Trans. Inf. Theory, vol. 39, no. 3, pp. 752–772, May 1993.

[3] M. Hayashi, “General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1562–1575, Apr. 2006.

[4] J. Hou and G. Kramer, “Informational divergence approximations to product distributions,” in Proc. of Canadian Workshop on Information Theory (CWIT), Jun. 2013, pp. 76–81.

[5] T. S. Han, H. Endo, and M. Sasaki, “Reliability and secrecy functions of the wiretap channel under cost constraint,” IEEE Trans. Inf. Theory, vol. 60, no. 11, pp. 6819–6843, Nov. 2014.

[6] M. Hayashi and R. Matsumoto, “Secure multiplex coding with dependent and non-uniform multiple messages,” arXiv e-prints, vol. abs/1202.1332v5, Apr. 2015. [Online]. Available: http://arxiv.org/abs/1202.1332

[7] M. Bastani Parizi and E. Telatar, “On the secrecy exponent of the wiretap channel,” in Proc. of IEEE Information Theory Workshop (ITW), Oct. 2015, pp. 287–291.

[8] M. Bastani Parizi, E. Telatar, and N. Merhav, “Exact random coding secrecy exponents for the wiretap channel,” in Proc. of IEEE Int. Symp. on Information Theory (ISIT), Jul. 2016, pp. 1521–1525.

[9] R. G. Gallager, Information Theory and Reliable Communication. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[10] M. V. Burnashev, “Data transmission over a discrete channel with feedback: Random transmission time,” Problemy Peredachi Informatsii, vol. 12, no. 4, pp. 250–265, 1976.

[11] R. G. Gallager, Discrete Stochastic Processes. Boston, MA, USA: Kluwer, 1996.

[12] A. Tchamkerten, “Feedback communication over unknown channels,” Ph.D. dissertation, School of Computer and Communication Sciences, EPFL, Lausanne, 2005.

[13] R. L. Dobrushin, “Asymptotic bounds on the probability of error for the transmission of messages over a memoryless channel using feedback,” Probl. Kibern., vol. 8, pp. 161–168, 1963.

[14] E. A. Haroutunian, “Lower bound for error probability in channels with feedback,” Problemy Peredachi Informatsii, vol. 13, no. 2, pp. 36–44, 1977.

[15] I. Csiszárik and J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, 2nd ed. Cambridge University Press, 2011.