The Error Probability of Generalized Perfect Codes via the Meta-Converse

Gonzalo Vázquez-Vilar, Member, IEEE, Albert Guillén i Fàbregas, Senior Member, IEEE, Sergio Verdú, Fellow, IEEE

Abstract

We introduce a definition of perfect and quasi-perfect codes for symmetric channels parametrized by an auxiliary output distribution. This notion generalizes previous definitions of perfect and quasi-perfect codes and encompasses maximum distance separable codes. The error probability of these codes, whenever they exist, is shown to coincide with the estimate provided by the meta-converse lower bound. We illustrate how the proposed definition naturally extends to cover almost-lossless source-channel coding and lossy compression.

I. INTRODUCTION

In the context of reliable communication, binary hypothesis testing has proved instrumental in the derivation of converse bounds to the error probability. Using this method, the sphere-packing bound on the channel coding reliability function was derived in [1] (see also [2]–[5] for alternative derivations and refinements). More recently, the meta-converse of Polyanskiy et al. [6, Th. 27] proved that a surrogate binary hypothesis test can be used to accurately lower-bound the error probability in the finite blocklength regime. The performance of binary hypothesis testing between distributions $P_0$ and $P_1$ (in absence of priors) is characterized by the trade-off $\alpha_\beta(P_0, P_1)$, where $\alpha$ denotes the smallest error under $P_0$ achievable by any test with error under $P_1$ at most $\beta$ (we refer the reader to Section II for a formal definition). Then, [6, Th. 27] establishes the following lower bound on the error probability of a code $C$ with cardinality $M$ used over a channel $P_{Y|X}$,

$$P_e(C) \geq \inf_{P_X} \sup_{Q_Y} \left\{ \alpha_{\beta}(P_X \times P_{Y|X}, P_X \times Q_Y) \right\}.$$  (1)

This bound is usually referred to as the meta-converse bound, since several previous converse bounds in the literature can be derived from it via relaxation. Particularized for $n$-uses of a memoryless binary symmetric channel (BSC), the meta-converse bound recovers the sphere-packing bound for BSC channels [7, Eq. (5.8.19)] (see [6, Sec. III.H]).
In this setting, the meta-converse bound (1) coincides with the exact error probability whenever perfect or quasi-perfect codes exist. In particular, a binary code is said to be perfect if non-overlapping Hamming spheres of radius \( t \) centered on the codewords exactly fill out the space. Similarly, a quasi-perfect code is defined as a code in which Hamming spheres of radius \( t \) centered on the codewords are non-overlapping and Hamming spheres of radius \( t + 1 \) cover the space, possibly with overlaps. This definition coincides with that of sphere-packed codes introduced by Gallager [7, Sec. 5.8]. Since quasi-perfect codes attain the lower bound (1), they achieve the minimum error probability in a BSC among all the codes with the same blocklength and rate.

In this work, we extend this result for general channels under certain symmetry conditions. We generalize the definition of perfect and quasi-perfect codes beyond Hamming distance, and show that these codes attain equality in (1). Therefore, generalized quasi-perfect codes achieve the minimum error probability among all the codes with the same blocklength and rate. As an example, we study a family of \( q \)-ary symmetric erasure channels and we show that maximum-distance separable (MDS) codes are generalized quasi-perfect for these channels. As a result, we obtain an alternative proof of the optimality of MDS codes for \( q \)-ary symmetric erasure channels. Extensions to almost-lossless source-channel coding and lossy compression under an excess-distortion constraint are discussed.

A tightened version of the meta-converse, derived for a fixed code, was shown to coincide with the exact error probability in [8, Th. 1]. In contrast to [8], in this paper we show that the bound (1), which applies to every code of cardinality \( M \), also yields the exact error probability in certain cases. In [9], Hamada also studied a generalization of perfect and quasi-perfect codes beyond Hamming distance. Using a variation of the Fano metric [10, Eq. (9.10)], Hamada derived a lower bound to the channel coding error probability. Our definition of quasi-perfect codes includes [9, Def. 1] as a special case and recovers Hamada’s condition for achieving minimum error probability [9, Th. 3]. Nevertheless, the class of generalized codes considered here is more general than that in [9] and shows connections not previously treated in the literature.

The structure of the paper is as follows. In Section II we introduce the binary hypothesis testing framework and notation used in the rest of the paper. In Section III we introduce the system model and show the optimality of the so-called generalized quasi-perfect codes. A family of erasure channels is studied in detail under this formulation in Section IV and the optimality of MDS codes is shown. Sections V and VI extend the notion of generalized quasi-perfect codes to almost-lossless source-channel coding and lossy compression under maximum excess-distortion probability, respectively. Section VII closes the paper with some final remarks.

II. BINARY HYPOTHESIS TESTING

Consider a non-Bayesian binary hypothesis test discriminating between distributions \( P_0 \) and \( P_1 \) defined over some discrete alphabet \( Z \). Let \( T(z) \) denote the probability of the test deciding hypothesis 0 (corresponding to \( P_0 \)) given an observation \( z \), \( 0 \leq T(z) \leq 1 \). Then, \( 1 - T(z) \) is the probability of deciding hypothesis 1 (i.e., \( P_1 \)). Let
\(\pi_{ji}\) denote the error probability of deciding \(j\) when \(i \neq j\) is the true hypothesis. More precisely, we define

\[
\pi_{01}(T) \triangleq \sum_z T(z)P_1(z),
\]
\[
\pi_{10}(T) \triangleq 1 - \sum_z T(z)P_0(z),
\]
and we denote the minimum error probability \(\pi_{10}\) among all tests \(T\) with \(\pi_{01}\) at most \(\beta\), as

\[
\alpha_{\beta}(P_0, P_1) \triangleq \inf_{T: \pi_{01}(T) \leq \beta} \pi_{10}(T).
\]

Neyman and Pearson (NP) provided in [11] an explicit form for the test achieving the optimal trade-off \(\alpha_{\beta}(P_0, P_1)\). In particular, for any \(\gamma \geq 0\), \(\theta \in [0, 1]\), the NP test is given by

\[
T_{\text{NP}}(z) = \mathbb{1}\left[\frac{P_0(z)}{P_1(z)} > \gamma\right] + \theta \mathbb{1}\left[\frac{P_0(z)}{P_1(z)} = \gamma\right],
\]

where \(\mathbb{1}[\cdot]\) denotes the indicator function. \(T_{\text{NP}}\) achieves the optimal trade-off \(\alpha_{\beta}(P_0, P_1) = \pi_{10}(T_{\text{NP}})\) when \(\gamma\) and \(\theta\) are chosen such that \(\beta = \pi_{01}(T_{\text{NP}})\). This result is usually known as NP Lemma. A direct consequence of the NP Lemma is the following alternative characterization of the optimal error probability trade-off \(\alpha_{\beta}(P_0, P_1)\).

\textbf{Lemma 1:} For any non-Bayesian binary hypothesis test discriminating between \(P_0\) and \(P_1\),

\[
\alpha_{\beta}(P_0, P_1) = \sup_{\gamma \geq 0} \left\{ \mathbb{P}\left[\frac{P_0(Z_0)}{P_1(Z_0)} \leq \gamma\right] + \gamma \mathbb{P}\left[\frac{P_0(Z_1)}{P_1(Z_1)} > \gamma\right] - \gamma \beta \right\},
\]

where \(Z_i \sim P_i, i = 0, 1\).

\section{Generalized Perfect Codes}

An equiprobable message \(m \in \{1, \ldots, M\}\) is to be transmitted over a channel with transition probability \(P_{Y|X}\), input \(x \in \mathcal{X}\) and output \(y \in \mathcal{Y}\), and where \(\mathcal{X}\) and \(\mathcal{Y}\) are the one-shot input/output discrete alphabets. A channel code is the set of codewords \(\mathcal{C} = \{x_1, \ldots, x_M\}\) \(x_i \in \mathcal{X}\) for \(i = 1, \ldots, M\), assigned to each of the messages. Under maximum likelihood (ML) decoding, the error probability for the code \(\mathcal{C}\) is given by

\[
P_e(\mathcal{C}) = 1 - \frac{1}{M} \sum_y \max_{x \in \mathcal{C}} P_{Y|X}(y|x).
\]

We consider the following family of symmetric channels.

\textbf{Definition 1:} Let \(F_x(\tau) \triangleq \mathbb{P}[P_{Y|X}(Y|x) \geq \tau]\), where \(Y \sim P_{Y|X=x}\) and \(\tau \in [0, 1]\). A channel \(P_{Y|X}\) is symmetric if \(F_x(\tau)\) does not depend on the input \(x\),

\[
F_x(\tau) = F(\tau), \quad \forall x \in \mathcal{X}, \quad \tau \in [0, 1].
\]

In the special case of discrete memoryless channels, Definition 1 implies that the rows of the channel transition matrix (with inputs as rows and outputs as columns), \(P_{Y|X}(\cdot|x)\), are permutations of each other. This definition coincides with that of uniformly dispersive channels of Massey [12, Sec. 4.2] and is less restrictive than those of Cover and Thomas [13, Sec. 7.2] and Gallager [7, p. 94]. The definition in [13, Sec. 7.2] additionally requires that the columns of the channel transition matrix are permutations of each other. The definition in [7, p. 94] requires the channel transition matrix to be partitioned in submatrices such that each submatrix fulfills the conditions in [13, Sec. 7.2]. Relations among these notions are investigated in [14, Sec. VI.B].
Let $Q$ be an auxiliary distribution defined on the output alphabet $\mathcal{Y}$. For an observation $y \in \mathcal{Y}$, the codeword $x \in \mathcal{C}$ that maximizes the metric $P_{Y|X}(y|x)$ also maximizes the metric $q(x, y) = \frac{P_{Y|X}(y|x)}{Q(y)}$. We conclude that the decoding regions induced by the ML decoder (with metric $P_{Y|X}(y|x)$) and those of the maximum metric decoder (with metric $q(x, y)$) coincide. This obvious fact proves to be instrumental next.

For any $\tau \geq 0$ and any distribution $Q$ defined over $\mathcal{Y}$, we define $S_x(\tau, Q)$ to be the set of outputs $y$ with likelihood given input $x$ at least $\tau Q(y)$, i.e.,

$$S_x(\tau, Q) \triangleq \left\{ y \in \mathcal{Y} \left| \frac{P_{Y|X}(y|x)}{Q(y)} \geq \tau \right. \right\}. \quad (9)$$

We refer to $S_x(\tau, Q)$ as a sphere of radius $\tau$ centered on $x$, although in general $\mathcal{X} \neq \mathcal{Y}$ and $q(x, y) \triangleq -\log q(x, y) = \log \frac{Q(y)}{P_{Y|X}(y|x)}$. For channels such as the BSC, $\log P_{Y|X}(y|x)$ is an affine function of the Hamming distance between $x$ and $y$ and, hence, $S_x(\tau, Q)$ becomes a Hamming sphere when $Q$ is the equiprobable distribution.

We define the interior and the outer shell of $S_x(\tau, Q)$ as

$$S_{i,x}(\tau, Q) \triangleq \left\{ y \in \mathcal{Y} \left| \frac{P_{Y|X}(y|x)}{Q(y)} > \tau \right. \right\}, \quad (10)$$

$$S_{o,x}(\tau, Q) \triangleq \left\{ y \in \mathcal{Y} \left| \frac{P_{Y|X}(y|x)}{Q(y)} = \tau \right. \right\}. \quad (11)$$

We consider the set of distributions $Q$ such that the tilted channel $P_{Y|X}(y|x) \propto \frac{P_{Y|X}(y|x)}{Q(y)}$ remains symmetric. More precisely, we define the set of symmetry-preserving auxiliary output distributions

$$Q \triangleq \left\{ Q \in \mathcal{P}(\mathcal{Y}) \left| F_x(\tau, Q) = F(\tau, Q), \forall x \in \mathcal{X}, \tau \geq 0 \right. \right\}, \quad (12)$$

where $F_x(\tau, Q) \triangleq \mathbb{P}[Y \in S_x(\tau, Q)]$ with $Y \sim P_{Y|X=x}$, and $\mathcal{P}(\mathcal{A})$ denotes the set of all probability distributions over alphabet $\mathcal{A}$. That is, $Q$ corresponds to the set of auxiliary distributions $Q$ such that the probability of $S_x(\tau, Q)$ under $P_{Y|X=x}$ is independent of $x$ for any $\tau \geq 0$.

For symmetric channels $P_{Y|X}$, the set $Q$ is non-empty as it always includes the equiprobable distribution, and it may include other auxiliary distributions. For example, consider a single use of the binary erasure channel (BEC) with erasure symbol $e$. In this case, any distribution of the form $Q(0) = Q(1) = \xi$, $Q(e) = 1 - 2\xi$, does not alter the symmetry of the original channel, and therefore it is included in $Q$. This example will be studied in detail in Section IV.

For a fixed $Q \in Q$, we use the short-hand notation $Q[A] \triangleq \mathbb{P}[Y \in A], Y \sim Q$.

**Lemma 2:** Let $P_{Y|X}$ be a symmetric channel according to Definition 1 and $Q \in Q$ defined in (12). Then, the probabilities $Q[S_x(\tau, Q)]$, $Q[S_{i,x}(\tau, Q)]$ and $Q[S_{o,x}(\tau, Q)]$ are independent of $x \in \mathcal{X}$ for any $\tau \geq 0$.

**Proof:** We prove that the term $Q[S_{o,x}(\tau, Q)]$ does not depend on $x$. Then, the independence of the other two terms follows since

$$Q[S_x(\tau, Q)] = \sum_{\tau' \in \mathcal{E}_Q, \tau' \geq \tau} Q[S_{o,x}(\tau', Q)], \quad (13)$$

$$Q[S_{i,x}(\tau, Q)] = \sum_{\tau' \in \mathcal{E}_Q, \tau' > \tau} Q[S_{o,x}(\tau', Q)], \quad (14)$$
where $\mathcal{L}_Q$ is defined as

$$\mathcal{L}_Q \triangleq \{ \tau \in \mathbb{R} \mid \exists x \in \mathcal{X}, \exists y \in \mathcal{Y}, \frac{P_{Y|X}(y|x)}{Q(y)} = \tau \}. \quad (15)$$

To show that $Q[S_{o,x}(\tau, Q)]$ is independent of $x$, we write

$$Q[S_{o,x}(\tau, Q)] = \sum_y Q(y) \mathbb{I} [P_{Y|X}(y|x) = \tau Q(y)] = \frac{1}{\tau} \sum_y P_{Y|X}(y|x) \mathbb{I} [P_{Y|X}(y|x) = \tau Q(y)]. \quad (16)$$

According to the definition of $Q$ in (12), for any $Q \in Q$,

$$F_x(\tau, Q) = \sum_y P_{Y|X}(y|x) \mathbb{I} [P_{Y|X}(y|x) \geq \tau Q(y)] \quad (18)$$

does not depend on the specific value of $x$ for any $\tau \geq 0$. Then, noting that the summation in (17) is given by

$$\lim_{\delta \to 0} (F_x(\tau, Q) - F_x(\tau + \delta, Q)),$$

the result follows.

Then, according to Lemma 2, we define for symmetric channels the probability measures

$$Q(\tau) \triangleq Q[S_\nu(\tau, Q)], \quad Q(\nu) \triangleq Q[S_\nu(\tau, Q)], \quad Q(\tau) \triangleq Q[S_{o,x}(\tau, Q)]. \quad (19)$$

For a fixed code $C$ and auxiliary distribution $Q \in Q$, we let $\eta \geq 0$ be the largest value such that $\bigcup_{x \in C} S_x(\eta, Q) = \mathcal{Y}$. Similarly, let $\nu \geq 0$ be the smallest value such that the codeword centered sets $\{S_x(\nu, Q)\}_{x \in C}$ are disjoint.

We refer to $\eta$ and $\nu$ as the covering and packing radius of the code $C$ with respect to $Q$, respectively. Intuitively, $S_x(\nu, Q)$ is the largest sphere packed inside the ML decoding region corresponding to $x \in C$. Similarly, $S_x(\eta, Q)$ is the smallest sphere centered at $x \in C$ which completely covers the corresponding ML decoding region.

**Definition 2:** A code $C$ is *generalized perfect* for $P_{Y|X}$, if there exists $\gamma \geq 0$ and $Q \in Q$ such that the codeword-centered sets $\{S_x(\gamma, Q)\}_{x \in C}$ are disjoint and

$$\bigcup_{x \in C} S_x(\gamma, Q) = \mathcal{Y}. \quad (20)$$

A code is *generalized quasi-perfect* if there exists $\gamma \geq 0$ and $Q \in Q$ such that (20) is satisfied and the codeword-centered sets $\{S_x(\gamma, Q)\}_{x \in C}$ are disjoint.\(^1\)

Note that for generalized quasi-perfect codes the covering and packing radius coincide. The definition of quasi-perfect codes includes perfect codes as a special case. To avoid ambiguities, for perfect codes we require that $\gamma$ is the largest value satisfying (20). For this value of $\gamma$,

$$\bigcup_{x \in C} S_{o,x}(\gamma, Q) \subset \mathcal{Y}. \quad (21)$$

The main result in this work, Theorem 1, is a consequence of the following lemma, which is a refinement of [10, (9.15)-(9.16)].

\(^1\)While the sets $S_x(\gamma, Q)$ and $S_{o,x}(\gamma, Q)$ are a function of the parameters $\gamma$ and $Q$, they depend only on their product (see (9) and (10)). Therefore, the two parameters $\gamma \geq 0$ and $Q \in Q$ appearing in the definition of generalized perfect and quasi-perfect codes could be replaced by a single unnormalized function $f(y) = \gamma Q(y)$.
Lemma 3: Let $P_{Y|X}$ be a symmetric channel according to Definition 1 and let $Q \in \mathcal{Q}$. The error probability of any code $\mathcal{C}$ with $M$ codewords satisfies, for any $\gamma \geq 0$ and any $Q \in \mathcal{Q}$,

$$P_e(\mathcal{C}) \geq \gamma \left( Q(\gamma) - \frac{1}{M} \right) + \sum_{\tau \in \mathcal{L}_Q, \tau \leq \gamma} \tau Q_o(\tau),$$

where $\mathcal{L}_Q$ is defined in (15). Furthermore, the lower bound (22) holds with equality if and only if $\mathcal{C}$ is generalized quasi-perfect and $\gamma$ and $Q$ are the parameters (not necessarily unique) satisfying the conditions in Definition 2.

Proof: Let $\mathcal{C} = \{x_1, \ldots, x_M\}$ be an arbitrary code. We consider a deterministic ML decoder which partitions the output space into disjoint decoding regions $\{\mathcal{D}_1, \ldots, \mathcal{D}_M\}$. The error probability (7) becomes

$$P_e(\mathcal{C}) = 1 - \frac{1}{M} \sum_{m=1}^{M} \sum_{y \in \mathcal{D}_m} P_{Y|X}(y|x_m).$$

(23)

For an observed $y$, the codeword $x \in \mathcal{C}$ that maximizes the metric $P_{Y|X}(y|x)$ coincides with the one maximizing the metric $q(x, y) = \frac{P_{Y|X}(y|x)}{Q(y)}$. Then, using the definition of the covering and packing radius $\eta$ and $\nu$, respectively, it follows that

$$S_{x_m}(\nu, Q) \subseteq \mathcal{D}_m \subseteq S_{x_m}(\eta, Q),$$

(24)

for $1 \leq m \leq M$. As a result, $\mathcal{D}_m$ can be decomposed as

$$\mathcal{D}_m = S_{x_m}(\nu, Q) \bigcup_{\eta \leq \tau \leq \nu} (\mathcal{D}_m \cap S_{x_m}(\tau, Q)),$$

(25)

and (23) becomes

$$P_e(\mathcal{C}) = 1 - \frac{1}{M} \sum_{m=1}^{M} \left( \sum_{y \in S_{x_m}(\nu, Q)} P_{Y|X}(y|x_m) + \sum_{\eta \leq \tau \leq \nu} \sum_{y \in (\mathcal{D}_m \cap S_{x_m}(\tau, Q))} P_{Y|X}(y|x_m) \right).$$

(26)

Since $\frac{P_{Y|X}(y|x)}{Q(y)} = \tau$ for any $y \in S_{x_m}(\tau, Q)$, we write

$$\sum_{y \in S_{x_m}(\nu, Q)} P_{Y|X}(y|x) = \sum_{y \in S_{x_m}(\nu, Q)} \frac{P_{Y|X}(y|x)}{Q(y)} Q(y)$$

(27)

$$= \sum_{\tau \in \mathcal{L}_Q, \tau > \nu} y \sum_{y \in S_{x_m}(\tau, Q)} \tau Q(y)$$

(28)

$$= \sum_{\tau \in \mathcal{L}_Q, \tau > \nu} \tau Q_o(\tau),$$

(29)

where in (29) we used Lemma 2 and $Q_o(\tau) = Q[S_{x_m}(\tau, Q)]$ as defined in (19). Similarly,

$$\sum_{y \in (\mathcal{D}_m \cap S_{x_m}(\tau, Q))} P_{Y|X}(y|x) = \sum_{y \in (\mathcal{D}_m \cap S_{x_m}(\tau, Q))} \tau Q(y)$$

(30)

$$= \tau Q_o,m(\tau).$$

(31)

where we abbreviate $Q_o,m(\tau) = Q[\mathcal{D}_m \cap S_{x_m}(\tau, Q)].$

Substituting (29) and (31) in (26), yields

$$P_e(\mathcal{C}) = 1 - \left( \sum_{\tau \in \mathcal{L}_Q, \tau > \nu} \tau Q_o(\tau) + \frac{1}{M} \sum_{m=1}^{M} \sum_{\eta \leq \tau \leq \nu} \tau Q_o,m(\tau) \right).$$

(32)
Upon rearranging terms, (33) yields
\[ Q \]

For any choice of \( \gamma \), \( \gamma > \nu \)

Following analogous steps as in (34)-(35), via (37) we obtain
\[ \text{Then, using (34)-(35) in (32), it follows that } P_e(C) \geq \Gamma(\nu) \text{ where} \]
\[ \Gamma(\nu) \triangleq 1 - \left( \sum_{\tau \in Q, \tau > \nu} \tau Q_o(\tau) + \nu \left( \frac{1}{M} - Q_i(\nu) \right) \right). \]

For quasi-perfect codes satisfying Definition 2, there exist \( Q \in Q \) and \( \gamma = \nu = \eta \) such that covering and packing radius coincide. Then, for this choice of parameters, the inequality (35) becomes equality and \( P_e(C) = \Gamma(\gamma) \). We conclude that, for a generalized quasi-perfect code, (22) holds with equality for any choice (not necessarily unique) of \( \gamma \) and \( Q \) satisfying the conditions in Definition 2.

If \( C \) is not generalized quasi-perfect, \( \nu > \eta \) for every \( Q \in Q \) and the inequality (35) is strict. Then, \( P_e(C) > \Gamma(\gamma) \).

For any choice of \( \gamma \geq 0 \) not necessarily equal to the packing radius \( \nu \), we next show that \( P_e(C) > \Gamma(\gamma) \). First, note that for \( \gamma > \nu \), both (32) and (34)-(35) still hold substituting \( \nu \) by \( \gamma \). Then, the discussion above still applies.

Assume now that \( \eta \leq \gamma < \nu \). We can rewrite (32) as
\[ P_e(C) = 1 - \left( \sum_{\tau \in L, \tau > \gamma} \tau Q_o(\tau) + \frac{1}{M} \sum_{m=1}^{M} \sum_{\tau \geq \gamma} \tau Q_o, m(\tau) \right) + \frac{1}{M} \sum_{m=1}^{M} \sum_{\eta \leq \tau \leq \gamma} \tau \Delta_m(\tau). \]

where \( \Delta_m(\tau) \triangleq Q_o(\tau) - Q_o, m(\tau) \). Similarly, (33) becomes
\[ 1 = MQ_i(\gamma) + \sum_{m=1}^{M} \sum_{\tau \in L_q, \eta \leq \tau \leq \gamma} Q_o, m(\tau) - \sum_{m=1}^{M} \sum_{\tau \in L_q, \gamma < \tau \leq \nu} \Delta_m(\tau). \]

Following analogous steps as in (34)-(35), via (37) we obtain
\[ P_e(C) \geq \Gamma(\gamma) + \frac{1}{M} \sum_{m=1}^{M} \sum_{\tau \in L_q, \gamma < \tau \leq \nu} (\tau - \gamma) \Delta_m(\tau). \]

All terms in the inner sum in (39) satisfy \( \tau - \gamma > 0 \) and \( \Delta_m(\tau) \geq 0 \). If the code \( C \) is not generalized quasi-perfect, then, either \( P_e(C) > \Gamma(\gamma) \) or \( \Delta_m(\tau) > 0 \) for at least one term in the sum. As the same proof steps follow for \( \gamma < \eta \), we conclude that \( P_e(C) > \Gamma(\gamma) \) for any \( \gamma \geq 0 \), \( Q \in Q \), provided that \( C \) is not quasi-perfect.

We are now ready to state the main result of this section, which shows that the ML decoding error probability of generalized perfect and quasi-perfect codes coincides with the meta-converse lower bound (1).
Theorem 1: Let \( P_{Y|X} \) be a symmetric channel according to Definition 1 and \( C \) be a generalized quasi-perfect code according to Definition 2. Then, \( C \) attains the minimum error probability among all codes with \( M \) codewords, and it is given by

\[
P_e(C) = \inf_{P_X \in \mathcal{Q}} \max_{x \in \mathcal{X}} \left\{ \alpha_{\mathcal{M}}(P_X \times P_{Y|X}, P_X \times Q) \right\},
\]

(40)

\[
= \max_{Q \in \mathcal{Q}} \left\{ \alpha_{\mathcal{M}}(P_X \times P_{Y|X}, P_X \times Q) \right\}, \quad \text{for every } x \in \mathcal{X}.
\]

(41)

Conversely, any code for which (40)-(41) hold is a generalized quasi-perfect.

Proof: Let us consider the hypothesis test in (40). We apply Lemma 1 with \( P_0 \leftarrow P_X \times P_{Y|X} \) and \( P_1 \leftarrow P_X \times Q \). Using the definition of the set \( S_{o,x}(\cdot) \) and \( Q(\cdot) \) in Lemma 2, we get

\[
\alpha_{\mathcal{M}}(P_X \times P_{Y|X}, P_X \times Q) = \sup_{\gamma \geq 0} \left\{ \sum_{x,y\in S_{o,x}(\gamma,Q)} P_X(x)P_{Y|X}(y|x)+\gamma Q_\gamma(\gamma) - \frac{\gamma}{M} \right\}.
\]

(42)

For any \( y \in S_{o,x}(\tau,Q), \tau \in \mathcal{L}_Q \), where \( \mathcal{L}_Q \) is defined in (15), it holds that \( \frac{P_{Y|X}(y|x)}{Q(y)} = \tau \). Then,

\[
\sum_{y \in S_{o,x}(\gamma,Q)} P_{Y|X}(y|x) = \sum_{\tau \in \mathcal{L}_Q, \tau \leq \gamma, y \in S_{o,x}(\tau,Q)} \frac{P_{Y|X}(y|x)}{Q(y)}Q(y)
\]

(43)

\[
= \sum_{\tau \in \mathcal{L}_Q, \tau \leq \gamma} \sum_{y \in S_{o,x}(\tau,Q)} \tau Q(y)
\]

(44)

\[
= \sum_{\tau \in \mathcal{L}_Q, \tau \leq \gamma} \tau Q_\gamma(\tau),
\]

(45)

which does not depend on \( x \) (see Lemma 2). Then, (42) becomes

\[
\alpha_{\mathcal{M}}(P_X \times P_{Y|X}, P_X \times Q) = \max_{\gamma \geq 0} \left\{ \sum_{\tau \in \mathcal{L}_Q, \tau \leq \gamma} \tau Q_\gamma(\tau) + \gamma Q_\gamma(\gamma) - \frac{\gamma}{M} \right\}.
\]

(46)

According to (1), the right-hand side of (46) is a lower bound to \( P_e(C) \). According to Lemma 3, the term in braces in (46) is precisely the error probability of a generalized quasi-perfect code with parameters \( Q \) and \( \gamma \). Then, whenever this code exists the lower bound (46) is achievable and (40) holds with equality. Moreover, (41) holds since (46) does not depend on \( P_X \) for symmetric channels and \( Q \in \mathcal{Q} \).

Let now \( Q \in \mathcal{Q} \) achieve (40)-(41), and let \( \gamma \) be the maximizer in (46). It follows from Lemma 3 that the term in braces in (46) is the error probability of a code \( C \) if and only if \( C \) is generalized quasi-perfect and the parameters \( \gamma \) and \( Q \) satisfy the conditions in Definition 2. We conclude that, if (40)-(41) hold, \( C \) must be a generalized quasi-perfect.

For any codebook \( C = \{x_1, \ldots, x_M\} \), we let \( P_X^C \) denote the distribution induced by \( C \), i.e., \( P_X^C(x) \triangleq \frac{1}{M} \sum_{m=1}^{M} \mathbb{1}\{x = x_m\} \). It has been shown in [8, Th. 1] that the error probability of any code can be expressed as

\[
P_e(C) = \max_{Q} \left\{ \alpha_{\mathcal{M}}(P_X^C \times P_{Y|X}, P_X^C \times Q) \right\}
\]

(47)

\[
\geq \min_{P_X} \max_{Q} \left\{ \alpha_{\mathcal{M}}(P_X \times P_{Y|X}, P_X \times Q) \right\},
\]

(48)

Eq. (47) shows that the meta-converse bound, when applied to a fixed code \( C \), coincides with the exact error probability \( P_e(C) \). Theorem 1 shows that, under certain symmetry conditions, the relaxation (48) also coincides
with the exact error probability, provided that a quasi-perfect code of cardinality $M$ exists for this channel. Also, Theorem 1 is more general than the result obtained by Hamada in [9, Th. 3]. For instance, our result can be used to prove the optimality of MDS codes in $q$-ary erasure channels, as we show in the next section.

IV. Symmetric Erasure/Error Channels

Consider a symmetric erasure channel $P_{Y|X}$ with discrete input alphabet $\mathcal{X}$, $|\mathcal{X}| = q$, and output alphabet $\mathcal{Y} = \mathcal{X} \cup \{e\}$ where $e$ corresponds to the erasure symbol:

$$P_{Y|X}(y|x) = \begin{cases} 1 - \delta - \epsilon, & y = x, \\ \delta, & y = e, \\ \frac{\epsilon}{q-1}, & \text{otherwise}. \end{cases}$$

(49)

When $q = 2$, this channel includes as particular cases the BSC and the BEC with $\delta = 0$ and $\epsilon = 0$, respectively.

We consider $n$ uses of this channel. Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ denote the channel input and output, respectively. For a given pair of $x$ and $y$, we define the number of erasures and the number of flip-errors, respectively, as

$$e_y \triangleq \sum_i 1[y_i = e],$$

(50)

$$d_{x,y} \triangleq \sum_i 1[x_i \neq y_i] - e_y.$$  

(51)

The $n$-dimensional channel transition probability is given by

$$P_{Y|X}(y|x) = \delta^{e_y} \left( \frac{\epsilon}{q-1} \right)^{d_{x,y}} (1 - \delta - \epsilon)^{n-e_y-d_{x,y}}.$$  

(52)

We assume that $\frac{\epsilon}{q-1} < 1 - \delta - \epsilon$. Otherwise, observing the transmitted symbol at the output of the channel would be less likely than observing any of the other $q - 1$ symbols. Particularized to the BSC (with $q = 2$, $\delta = 0$), this assumption boils down to the crossover probability being $\epsilon < \frac{1}{2}$.

We define the auxiliary distribution

$$Q_{Y}(y) \triangleq \frac{1}{Z} \delta^{e_y} \left( \frac{\epsilon}{q-1} \right)^{\Psi(e_y)} (1 - \delta - \epsilon)^{n-e_y-\Psi(e_y)},$$

(53)

where $c$ is a normalizing constant, and $\Psi(e) \geq 0$ is an arbitrary function of the number of erasures $e$, which can be optimized over. Intuitively, $\Psi(e)$ corresponds to the average number of flip-errors that a good code can correct when the output sequence is affected by $e$ erasures. For binary-input channels, a good choice for $\Psi(e)$ is given by

$$\Psi(e) = \max \left( 0, \left\lceil \frac{[n - \log_2 M] - e - 1}{2} \right\rceil \right).$$

(54)

Since $Q_{Y}(y)$ only depends on $y$ via the number of erasures $e_y$ it does not affect the symmetry of the vector channel $P_{Y|X}$ and thus $Q_{Y} \in \mathcal{Q}$. Theorem 1 is applied to this channel and auxiliary distribution $Q = Q_{Y}$ to obtain the following result.

Corollary 1: The error probability of any code $C$ with cardinality $M$ used over the channel (52) satisfies

$$P_e(C) \geq \sum_{e=0}^{n} \sum_{d=0}^{n-e} \binom{n}{e} \left( \frac{n-e}{d} \right) (q-1)^d \delta^e (1 - \delta - \epsilon)^{n-e} \left( \frac{\varphi_{\max(d,\Psi(e))} - \varphi_{\Psi(e)}}{M} \right),$$

(55)
where \( \varphi \triangleq \frac{q}{q-1} (1 - \delta - \epsilon)^{-1} \) and \( \Psi(e) \geq 0 \) is any positive function of the number of erasures \( e \). Moreover, if \( \mathcal{C} \) is a generalized quasi-perfect code that satisfies Definition 2 with \( \gamma = c \) and \( Q = Q^*_Y \) then (55) holds with equality.

**Proof:** Let us consider the lower bound that follows from (46) by fixing \( Q = Q^*_Y \) defined in (53) and fixing \( \gamma' = c \) to be the normalization factor appearing in (53),

\[
\max_{Q \in \mathcal{Q}} \left\{ \alpha \left( P_X \times P_{Y|X}, P_X \times Q \right) \right\} \geq \sum_x P_X(x) \left( \sum_{y \notin S_{i,x}(c,Q^*_Y)} P_{Y|X}(y|x) - c \left( \frac{1}{M} - \sum_{y \in S_{i,x}(c,Q^*_Y)} Q_Y(y) \right) \right).
\]

(56)

For \( Q = Q^*_Y \) defined in (53), the sets \( S_{i,x}(c,Q) \) can be expressed as

\[
S_{i,x}(c,Q) = \left\{ y \in \mathcal{Y} \left| d_{x,y} < \Psi(e_y) \right. \right\}.
\]

(57)

We parametrize each output sequence \( y \) by the indices \( e = e_y \in [0,n] \) and \( d = d_{x,y} \in [0,n-e_y] \). For a given \( x \), there are exactly \( \binom{n}{e}(q-1)^d \) output sequences \( y \) with indices \( e, d \). Using this parametric representation, using (57) and the definitions of \( P_{Y|X} \) in (52) and \( Q^*_Y \) in (53), we obtain

\[
\sum_{y \notin S_{i,x}(c,Q^*_Y)} P_{Y|X}(y|x) = \sum_{e=0}^{n} \sum_{d=\Psi(e)}^{n-e} \binom{n-e}{d} (q-1)^d (1 - \delta - \epsilon)^{n-e} \varphi_y^d,
\]

(58)

\[
\sum_{y \in S_{i,x}(c,Q^*_Y)} Q^*_Y(y) = \frac{1}{c} \sum_{e=0}^{n} \sum_{d=0}^{\Psi(e)-1} \binom{n-e}{d} (q-1)^d (1 - \delta - \epsilon)^{n-e} \varphi_y^{\Psi(e)}.
\]

(59)

Substituting these expressions in (56), reorganizing terms, yields

\[
\max_{Q \in \mathcal{Q}} \left\{ \alpha \left( P_X \times P_{Y|X}, P_X \times Q \right) \right\} \geq \sum_{e=0}^{n} \sum_{d=0}^{n-e} \binom{n-e}{d} (q-1)^d (1 - \delta - \epsilon)^{n-e} \varphi_y^{\text{max}(d,\Psi(e))} - \frac{c}{M}.
\]

(60)

Finally, noticing that

\[
\sum_{d=0}^{n-e} \binom{n-e}{d} (q-1)^d = q^{n-e},
\]

(61)

we obtain the normalizing constant in (53),

\[
c = \sum_{e=0}^{n} q^{n-e}(1 - \delta - \epsilon)^{n-e} \varphi_y^{\Psi(e)}
\]

(62)

Substituting (62) in (60), via the meta-converse bound (1), we obtain (55). According to Lemma 3, this bound holds with equality if \( \mathcal{C} \) is generalized \( Q \)-quasi-perfect with parameters \( \gamma = c \) and \( Q = Q^*_Y \).

Let \( d_{\text{min}} \) denote the minimum Hamming distance between any pair of codewords in \( \mathcal{C} \). The Singleton bound [15, Th. 4.5.6] establishes the maximum number of codewords \( M \) in a \( q \)-ary block code \( \mathcal{C} \) of length \( n \) and minimum distance \( d_{\text{min}} \),

\[
\log_q M \leq n - d_{\text{min}} + 1.
\]

(63)

Those codes achieving the Singleton bound with equality are termed MDS codes. Examples of MDS codes include those that have only two complementary codewords thus having \( d_{\text{min}} = n \), non-redundant codes, i.e., \( \mathcal{C} = \mathcal{X} \), for which \( d_{\text{min}} = 1 \), codes with a single parity symbol for which \( d_{\text{min}} = 2 \) and their corresponding dual codes. These
are often called trivial MDS codes. In the case of binary alphabets, only trivial MDS codes exist. For non-binary alphabets, Reed-Solomon codes are an example of non-trivial MDS codes.

MDS codes are indeed generalized quasi-perfect codes for the $q$-ary erasure channel, given by $P_{X|Y}$ in (52) when $\epsilon = 0$. We note that $\lim_{\epsilon \to 0} \epsilon A = 0$ for any $A > 0$ and $\lim_{\epsilon \to 0} \epsilon A = 1$ for $A = 0$. Then, for any function $\Psi(\epsilon) \geq 0$ such that $\Psi(\epsilon) = 0$ if, and only if $e > n - \log_q M$, (53) becomes

$$Q^*_Y(y) = \begin{cases} 0, & e_Y \leq n - \log_q M, \\ \frac{1}{e} \delta^e (1 - \delta)^{n-e}, & e_Y > n - \log_q M. \end{cases} \tag{64}$$

Consider a generalized quasi-perfect code according to Definition 2 with parameters $Q = Q^*_Y$ and $\gamma = c$ as defined in (64). For the sets $S_x(\cdot)$ we use the convention that, when $Q^*_Y(y) = 0$,

$$\frac{P_{X|Y}(y|x)}{Q^*_Y(y)} = \begin{cases} 0, & \text{if } P_{X|Y}(y|x) = 0, \\ \infty, & \text{if } P_{X|Y}(y|x) > 0. \end{cases} \tag{65}$$

The spheres induced by this code are such that their interior $S_{x,*}(c, Q^*_Y)$ is the set of the output sequences $y$ that are compatible with the input $x$ with a number of erasures $e_Y \leq n - \log_q M$. Since the codeword-centered interiors do not overlap, the minimum distance of the code is at least $|n - \log_q M| + 1$. Since the codeword centered shells $S_{x,*}(c, Q^*_Y)$ overlap at some point, $d_{min}$ is exactly

$$d_{min} = |n - \log_q M| + 1. \tag{66}$$

When $\log_q M$ is an integer, this expression coincides with the Singleton bound (63). As a result, we conclude that MDS codes are also quasi-perfect. By letting $\epsilon \to 0$ in Corollary 1 for any $\Psi(\epsilon)$ such that $\Psi(\epsilon) = 0$ iff $e > n - \log_q M$, we obtain the following result.

**Corollary 2:** The error probability of any code $C$ with cardinality $M$ used over a $q$-ary erasure channel satisfies

$$P_\epsilon(C) \geq \sum_{e = |n - \log_q M| + 1}^{n} \binom{n}{e} \delta^e (1 - \delta)^{n-e} \left(1 - \frac{2^e}{M}\right), \tag{67}$$

with equality if $C$ is generalized quasi-perfect with parameters $\gamma = c$ and $Q = Q^*_Y$, as defined in (64).

The bound in (67) coincides with the converse bound [6, Th. 38]. As observed in [6], this lower bound is tight when $C$ is an MDS code. Here this result is recovered via the definition of generalized quasi-perfect codes.

As an example, let us consider the transmission of $M = 4$ codewords over a length-$n$ binary input channel (52) for three sets of parameters: BSC with $(\epsilon, \delta) = (0.25, 0)$, a channel with erasures and errors with $(\epsilon, \delta) = (0.05, 0.2)$ and BEC with $(\epsilon, \delta) = (0, 0.25)$. Figure 1 depicts the exact error probability $P_\epsilon(C)$ of the best code compared with the lower bound (55) with the choice of $\Psi(\epsilon)$ given in (54). The optimum codes for the BSC and BEC are taken from [16] and [17], respectively. For the channel with combined erasures and errors we also use the code for the BEC, since it offers a better performance at points where it differs from that of the BSC. Figure 1 shows that the bound (55) for the BSC coincides with the code error probability at the points were quasi-perfect codes exist with respect to the Hamming distance ($n = 2, 3, 4, 5, 6, 8$). For the BEC, the bound (55) (which coincides with (67)) provides the exact error probability at the points where (trivial) MDS codes exist ($n = 2, 3$), as they are
generalized quasi-perfect. For the combined errors-erasures channel, to be optimal the codes need to be generalized quasi-perfect for both the BSC and BEC, which only occurs at \( n = 2, 3 \).

Let us consider now a \( q \)-ary channel \((49)\) with \( q = 32 \), and fixed transmission rate \( R = \frac{1}{n} \log_q \) \( M = \frac{1}{2} \). Fig. 2 depicts the lower bound \((55)\) (optimized over a family of functions \( \Psi(\epsilon) \)) for combined erasures and errors with \((\epsilon, \delta) = (0.05, 0.25)\), and the lower bound \((67)\) for erasures only with \((\epsilon, \delta) = (0, 0.25)\). For even blocklengths, we have simulated the performance of a Reed-Solomon code in both scenarios with \( 10^6 \) Monte Carlo realizations. Reed-Solomon codes are defined for blocklengths \( n \leq q - 1 \) and they are generalized quasi-perfect for the \( q \)-ary erasure channel. Therefore, they attain the lower bound \((67)\) with equality in the erasure-only case. While their performance with errors and erasures is not far from the lower bound \((55)\), a gap exists in this case. Reed-Solomon codes can be extended for blocklengths \( n = q \) and \( n = q + 1 \). There exist no MDS codes for longer blocklengths in general \([18]\).

V. Almost-Lossless Source-Channel Coding

In this section, the notion of quasi-perfect codes is generalized to allow non-equiprobable messages, hence the code needs be matched both to the source and the channel.

We consider the almost-lossless source-channel coding setting. A source generates messages \( v \in \mathcal{V} \), where \( \mathcal{V} \) is a discrete alphabet, according to \( P_V \). The message \( v \) is to be transmitted over a channel \( P_{Y|X} \), \( x \in \mathcal{X} \) and \( y \in \mathcal{Y} \), using a channel encoder that maps each source message \( v \) into a codeword \( x_v \in \mathcal{X} \). We let \( P_{X|V}^C \) denote the conditional distribution \( P_{X|V} \) induced by the codebook \( C \triangleq \{x_1, \ldots, x_{|V|}\} \). The receiver uses maximum-a-
posteriori (MAP) decoding to decide on the transmitted message $\hat{v} \in \mathcal{V}$. This decoder minimizes the average error probability, which is given by

$$
P_e(C) = \mathbb{P}[\hat{V} \neq V] = 1 - \sum_y \max_v P_V(v) P_{Y|X}(y|x_v).$$

The concept of generalized quasi-perfect codes presented in Section III can be extended to almost-lossless source-channel coding.

**Definition 3** A source-channel code $C$ is generalized perfect with respect to a given source $P_V$ and channel $P_{Y|X}$, if there exists $\gamma \geq 0$ and an auxiliary distribution $Q \in \mathcal{Q}$ such that

$$
\bigcup_{v \in \mathcal{V}} S_{x_v} \left( \frac{\gamma}{P_V(v)}, Q \right) = \mathcal{Y},
$$

where the union is disjoint. More generally, a code is generalized quasi-perfect if there exists $\gamma \geq 0$ and $Q \in \mathcal{Q}$ such that (70) is satisfied and the codeword-centered sets $\left\{ S_{i,x_v} \left( \frac{\gamma}{P_V(v)}, Q \right), v \in \mathcal{V} \right\}$ are disjoint.

The definition of a source-channel quasi-perfect code induces a packing of spheres with radius depending on the probability of the associated source message – more probable source messages are associated to larger spheres. If the source messages are equiprobable, then the radius of the spheres becomes independent of the associated source message and Definition 3 boils down to Definition 2.
Theorem 2: Let $P_V$ be the distribution of the source messages, $P_{Y|X}$ be a symmetric channel according to Definition 1, and $C$ be a generalized quasi-perfect source-channel code according to Definition 3. Then,

$$P_e(C) = \min_{P_X|V} \max_{Q \in \mathcal{Q}} \left\{ \alpha \cdot \frac{1}{|V|} \left( P_V \times P_{X|V} \times P_{Y|X}, \bar{P}_V \times P_{X|V} \times Q \right) \right\}, \quad (71)$$

where $\bar{P}_V(v) = \frac{1}{|V|}$ for all $v \in V$. Conversely, if (71) holds, then $C$ is generalized $Q$-quasi-perfect with respect to the source $P_V$ and channel $P_{Y|X}$.

Proof: See Appendix A.

The right-hand side of (71) is precisely the converse bound [19, Th. 4] particularized in the almost-lossless setting. Therefore, Theorem 2 shows that [19, Th. 4] is tight provided that a generalized quasi-perfect code matched to the source and channel exists.

As a particular case, consider a noiseless channel such that $y = x$ with $\mathcal{X} = \mathcal{Y} = \{1, \ldots, M\}$, and $|V| > M$. In this case, Definition 3 yields “spheres” of size 1 for the $M$ most probable messages and the $|V| - M$ least probable messages are assigned to “empty spheres”. In practice, the messages associated to these “empty spheres” can be assigned to an arbitrary channel index, as they always yield to a decoding error given their smaller probability. This code corresponds precisely to the well-known optimal almost-lossless block source code. When the $M$ most probable messages have a strictly larger probability than that of the $|V| - M$ least probable messages, the code is generalized perfect according to Definition 3. When the $M$-th and $(M + 1)$-th most probable messages have the same probability, the code is generalized quasi-perfect.

VI. LOSSY SOURCE CODING

In this section, we consider the lossy source coding problem with a maximum distortion constraint. A source generates messages $v \in V$ with probability distribution $P_V$. The source encoder maps the message $v$ to a codeword $w \in W$ belonging to a length-$M$ codebook $C = \{w_1, w_2, \ldots, w_M\}$. Here $W$ denotes the reconstruction alphabet. We define a non-negative real-valued distortion measure $d(v, w) : V \times W \rightarrow \mathbb{R}^+$ and consider a maximum allowed distortion $D$. The minimum excess-distortion probability of a given code $C$ is defined as

$$P_{ed}(C, D) \triangleq \mathbb{P}\left[d(V, W) > D\right] \quad (72)$$

$$= 1 - \mathbb{P}\left[\min_{w \in C} d(V, w) \leq D\right], \quad (73)$$

where in (73) we used that the minimum excess distortion probability is attained by assigning each source message to the closest (in terms of distortion measure) codeword $w \in C$.

Quasi-perfect codes have good packing and covering properties simultaneously. Therefore, they are both good channel, as shown in the previous sections, and source codes, as shown next. According to Definition 2 whether a code is generalized quasi-perfect code depends on the channel. In the lossy source-coding setting, this channel corresponds to the test channel induced by the rate-distortion function.
Consider a block source encoder that encodes $n$ realizations of the source $P_V$ using a codebook of cardinality $2^{nR}$. Rate-distortion theory states that, as the blocklength $n$ grows large, the largest rate $R$ of a codebook with maximum distortion $D$ and vanishing excess-distortion probability is given by
\[ R(D) \triangleq \min_{P_W|V:E[d(V,W) \leq D]} I(V;W). \] (74)

The optimal $P^*_W|V$ in (74) induces a test channel $P^*_V|W$ that maps the reconstruction points into the source alphabet. More precisely, let $P^*_W(w) = \sum_v P_V(v)P^*_W(v|w)$, then, Bayes’ rule yields $P^*_V|W(v|w) = \frac{P_V(v)P^*_W(v|w)}{P^*_W(w)}$. It is shown in [13, Sec. 10.7] that the optimal test channel has the form
\[ P^*_V|W(v|w) = \frac{P_V(v)e^{-\lambda^*d(v,w)}}{\mu(v)}, \] (75)
for some $\lambda^* \geq 0$, and where $\mu(v) = \sum_w P^*_W(w)e^{-\lambda^*d(v,w)}$ is a normalization constant independent of $w$.

Let us consider the channel coding problem, as described in Section III, of transmitting $M$ messages over the channel $P^*_W|V$. Good channel codes for $P^*_W|V$ become good source codes for the source $P_V$ and distortion measure $d(v,w)$. In particular, quasi-perfect codes attain the minimum excess-distortion probability, as the next result shows.

**Theorem 3:** Consider a source $P_V$ with $P_V(v) > 0$, $v \in V$, distortion measure $d(v,w)$ and maximum distortion $D$. Let the test channel $P^*_W|V$ in (75) be symmetric according to Definition 1 and let $\hat{Q}(v) = \frac{1}{c_{\mu}} \frac{P_V(v)}{\mu(v)}$ satisfy $\hat{Q} \in Q$, where $\mu(v)$ is the normalizing factor in (75) and $c_{\mu} \triangleq \sum_v \frac{P_V(v)}{\mu(v)}$. Then, for any size-$M$ code $C$, generalized quasi-perfect according to Definition 2 with parameters $\gamma$ and $Q = \hat{Q}$, it holds that
\[ P_{ed}(C, D) = \max_{Q_V} \left\{ \alpha_M \sup_{w \in C} Q_V[d(V,w) \leq D] \left( P_V, Q_V \right) \right\}, \] (76)
where $Q_V[\cdot]$ denotes the probability when $V \sim Q_V$.

Moreover, if $D \geq -\frac{1}{\lambda^*} \log(\gamma/c_{\mu})$, the excess-distortion probability is $P_{ed}(C, D) = 0$.

**Proof:** See Appendix B.

In [8, Th. 3], the excess-distortion probability of any source code $C$ (not necessarily quasi-perfect) is expressed as the error probability of an induced binary hypothesis test with certain parameters,
\[ P_{ed}(C, D) = \max_{Q_V} \left\{ \alpha_{Q_V}[d(V,C) \leq D] \left( P_V, Q_V \right) \right\}, \] (77)
where $d(v,C) \triangleq \min_{w \in C} d(v,w)$. Invoking
\[ Q_V[d(V,C) \leq D] \leq M \sup_{w \in C} Q_V[d(V,w) \leq D] \leq M \sup_{w \in W} Q_V[d(V,w) \leq D], \] (78)
the identity (77) yields the lower bounds
\[ P_{ed}(C, D) \geq \max_{Q_V} \left\{ \alpha_M \sup_{w \in C} Q_V[d(V,w) \leq D] \left( P_V, Q_V \right) \right\} \] (79)
\[ \geq \max_{Q_V} \left\{ \alpha_M \sup_{w \in C} Q_V[d(V,w) \leq D] \left( P_V, Q_V \right) \right\}. \] (80)

Theorem 3 shows that, provided that a quasi-perfect code exists with certain parameters, the lower bound (79) holds with equality. The relaxation over the reconstruction alphabet in (80) coincides with [20, Th. 8]. For certain sources, the inequality (80) may hold with equality as the next example shows.
Let us consider the lossy compression of \( n \) i.i.d. samples of an equiprobable binary memoryless source (BMS) with bit error rate distortion measure, i.e., \( P_V(v) = 2^{-n} \) and \( d(v, w) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}[v_i \neq w_i] \), with \( v, w \in \{0, 1\}^n \).

The test channel for this rate-distortion problem corresponds to a BSC with certain parameters. As in the channel coding example from Fig. 1, we consider a codebook with \( M = 4 \) codewords. Figure 3 depicts the minimum excess-distortion probability \( P_{\text{ed}}(C, D) \) as a function of \( n \) for a maximum distortion \( D = 0.11 \) and \( D = 0.37 \).

Since we are “quantizing” a space of increasing dimension \( n \) with only \( M = 4 \) codewords, the excess-distortion probability tends to 1 as \( n \to \infty \) for any \( D < \frac{1}{2} \). In Fig. 3, we plot the lower bound (80) evaluated for \( Q_V \) uniform [20, Th. 15], compared to the exact excess-distortion probability evaluated for the best code in a BSC channel and \( M = 4 \) codewords [16]. We also highlight with markers the points where quasi-perfect codes exist for the BSC channel, corresponding to \( n = 2, 3, 4, 5, 6, 8 \) (see Fig. 1).

In Fig. 3, we observe that the exact excess-distortion probability coincides with the lower bound (80) at the points where quasi-perfect codes exist both for \( D = 0.11 \) and \( D = 0.37 \). Nevertheless, in the lossy compression setting, the reverse implication is not always true. Depending on the system parameters, the exact excess-distortion probability and the lower bound can also coincide when no quasi-perfect code exist for the corresponding test channel. Indeed, for \( D = 0.37 \) the only points where the exact excess-distortion probability and the lower bound coincide are only those in which quasi-perfect codes exist, while for \( D = 0.11 \) these two expressions coincide for all values of \( n \), regardless of whether the code is quasi-perfect. This occurs when the sets \( \{ v \in V \mid d(V, w) \leq D \} \), \( w \in C \), are non-overlapping (this occurs in our example for \( D \) sufficiently small). Then, the encoding regions which satisfy the maximum distortion cap are “spheres” regardless the specific structure of the codebook \( C \) and the lower
bound (80) yields the exact excess-distortion probability.

VII. DISCUSSION

We have proposed a generalization of perfect and quasi-perfect codes beyond the Hamming distance and their conventional application to binary symmetric channels. The definition of these codes follows from the packing and covering properties of a set of “spheres.” Since the shape of these spheres depends on the channel considered, quasi-perfect codes can only be defined with respect to a specific channel. For the BSC, quasi-perfect codes are defined with respect to the Hamming distance and our definition recovers the classical definition of quasi-perfect codes in the coding literature. Our approach capitalizes on the fact that the shape of these spheres can be tilted by using an auxiliary measure. This allows to extend the proposed formulation to encompass MDS codes, which are shown to be quasi-perfect for erasure channels.

While the proofs of the results in this paper are presented for discrete channels, they can be readily extended for channels with continuous outputs. In fact, Lemma 1, Definition 1, and Theorem 1 apply without change for both discrete and continuous channels, provided that they are absolutely continuous with respect to the Lebesgue measure. Nevertheless, the spheres induced by typical continuous channels seldom allow a perfect (or quasi-perfect) packing of the output space. Some atypical examples of continuous channels in which the induced spheres pack the output space are the AWGN channel with \( M = 2 \) codewords, the binary-input AWGN when all the input sequences are used, i.e., for \( M = 2^n \), or the additive white Laplace noise channel (as the induced spheres are norm-1 balls, and thus they can pack the space for specific lattice codes).

The framework presented in this work has been built upon the assumption that certain channel symmetry exists. Nevertheless, the underlying idea can be applied to general channels \( P_{Y|X} \) and arbitrary auxiliary distributions \( Q \).

APPENDIX A

Proof of Theorem 2

We apply Lemma 1 to the hypothesis test in (71) to obtain an alternative expression for the Neyman-Pearson performance of the test. This expression is then shown to coincide with the following characterization of the joint source-channel error probability of a quasi-perfect code.

Lemma 4: For a source \( P_V \) and a symmetric channel \( P_{Y|X} \), let the source-channel code \( C \) be generalized quasi-perfect with parameters \( \gamma \) and \( Q \) according to Definition 3. The error probability of \( C \) is

\[
P_e(C) = \sum_v \sum_{\tau \leq \gamma} \tau Q_0 \left( \frac{\tau}{P_V(v)} \right) - \gamma \left( 1 - \sum_v Q_1 \left( \frac{\gamma}{P_V(v)} \right) \right).
\]

Proof: The proof follows exactly the same steps as that of Lemma 3, and it is omitted here.
Applying Lemma 1 with \( P_0 \leftarrow P_V P_X|V| P_Y|X \) and \( P_1 \leftarrow \bar{P}_V P_X|V| Q \), via the change of variable \( \gamma \leftrightarrow \gamma' = \frac{\gamma}{|V|} \), yields

\[
\alpha_{\frac{1}{|V|}} \left( P_V P_X|V| P_Y|X, \bar{P}_V P_X|V| Q \right) = \max_{\gamma' \geq 0} \left\{ \sum_{v,x} P_V(v) P_X|V|(x|v) \sum_{y \notin S_{o,x}(\tau')} \frac{P_V|X|(y|x)}{Q(y)} Q(y) \right.
\]
\[
+ \gamma' \sum_{v,x} P_V(v) P_X|V|(x|v) \sum_{y \in S_{o,x}(\tau')} Q(y) - \gamma' \right\}
\]
\[= \max_{\gamma' \geq 0} \left\{ \sum_{v,x} P_V(v) P_X|V|(x|v) \sum_{\tau \in L_Q, \tau \leq \frac{\gamma'}{P_V(v)}} \tau Q_o(\tau)
\]
\[
+ \gamma' \sum_{v,x} P_X|V|(x|v) Q_i \left( \frac{\gamma'}{P_V(v)} \right) - \gamma' \right\},
\]
(82)

where in the last step we used that the complementary set of \( S_{o,x}(\gamma, Q) \) corresponds to \( \bigcup_{\tau \in L, \tau \leq \gamma} S_{o,x}(\tau, Q) \) and that \( \frac{P_V|Y|(y|x)}{Q(y)} = \tau \) for all \( y \in S_{o,x}(\tau, Q) \). Finally, using that \( \sum_x P_X|V|(x|v) = 1 \), and applying the change of variable \( \tau' = \tau P_V(v) \) (or equiv. \( \tau = \frac{\tau'}{P_V(v)} \)) for each \( v \) we obtain

\[
\alpha_{\frac{1}{|V|}} \left( P_V P_X|V| P_Y|X, \bar{P}_V P_X|V| Q \right) = \max_{\gamma' \geq 0} \left\{ \sum_{v} \sum_{\tau' \in L_Q, \tau' \leq \gamma'} \tau' Q_o \left( \frac{\tau'}{P_V(v)} \right) + \gamma' \sum_{v} Q_i \left( \frac{\gamma'}{P_V(v)} \right) - \gamma' \right\},
\]
(83)

which coincides with (81) when \( \gamma \) is the optimizing value of \( \gamma' \) in (84). Since (84) is a lower bound to \( P_e(C) \), the theorem thus follows by optimizing (84) over auxiliary distributions \( Q \in Q \).

**APPENDIX B**

**PROOF OF THEOREM 3**

Let \( C \) be generalized quasi-perfect with respect to the test channel \( P_{W|V}^* \) defined in (75), with parameters \( \gamma \) and \( \tilde{Q}(v) = \frac{1}{c_{\mu}} \frac{P_V(v)}{\mu(v)} \). The set \( S_w(\tau, \tilde{Q}) \) associated to the test channel \( P_{W|V}^* \) is given by

\[
S_w(\tau, \tilde{Q}) = \left\{ v \in V \mid d(v, w) \leq -\frac{1}{\lambda^*} \log \left( \frac{\mu(v) \tilde{Q}(v)}{P_V(v)} \right) \right\},
\]
(85)

which upon particularization to \( \tilde{Q}(v) = \frac{1}{c_{\mu}} \frac{P_V(v)}{\mu(v)} \) yields

\[
S_w(\tau, \tilde{Q}) = \left\{ v \in V \mid d(v, w) \leq -\frac{1}{\lambda^*} \log (\tau/c_{\mu}) \right\}.
\]
(86)

We divide the proof in two different cases depending on the value of the maximum distortion \( D \).

1) \( D \geq -\frac{1}{\lambda^*} \log (\gamma/c_{\mu}) \). In this case \( \gamma \geq c_{\mu} e^{-\lambda^* D} \), and

\[
S_w(\gamma, \tilde{Q}) \subseteq S_w(c_{\mu} e^{-\lambda^* D}, \tilde{Q}) = \left\{ v \in V \mid d(v, w) \leq D \right\}.
\]
(87)
According to Definition 2, the codeword-centered sets $S_w(\gamma, \bar{Q})$, $w \in C$, cover the space. Then, using (87) it follows that

$$\bigcup_{w \in C} \left\{ v \in \mathcal{V} \mid d(v, w) \leq D \right\} = \mathcal{V}. \quad (88)$$

As a result, the excess-distortion probability is

$$P_{ed}(C, D) = 1 - \sum_v P_v(v) \left[ \min_{w \in C} d(v, w) \leq D \right] \quad (89)$$

$$= 1 - \sum_v P_v(v) \left[ v \in \bigcup_{w \in C} \left\{ v \in \mathcal{V} \mid d(v, w) \leq D \right\} \right] = 0. \quad (90)$$

According to (88), for any distribution $Q_V$ defined over $\mathcal{V}$, we have that $Q_V\left[d(V, C) \leq D\right] = 1$. This indeed implies that $\sup_{w \in C} Q\left[d(V, w) \leq D\right] \geq \frac{1}{M}$. Since $\alpha_1(P_V, Q_V) = 0$, using (90), we conclude that (76) holds with equality.

2) $D < -\frac{1}{\mu} \log(\gamma/c_{\mu})$. In this region, $\gamma < c_{\mu}e^{-\lambda D}$, and it thus follows that

$$S_{i,w}(\gamma, \bar{Q}) \supseteq S_w(c_{\mu}e^{-\lambda D}, \bar{Q}) = \left\{ v \in \mathcal{V} \mid d(v, w) \leq D \right\}. \quad (91)$$

In this case, $\bigcup_{w \in C} \left\{ v \in \mathcal{V} \mid d(v, w) \leq D \right\}$ does not cover the space completely. Nevertheless, since the code $C$ is quasi-perfect with radius $\gamma$, the spheres $S_{i,w}(\gamma, Q)$, $w \in C$, are disjoint. Using (91) we conclude that the sets $\left\{ v \in \mathcal{V} \mid d(v, w) \leq D \right\}$, $w \in C$, do not overlap. Therefore,

$$P_{ed}(C, D) = 1 - P_V \left[ \min_{w \in C} d(V, w) \leq D \right] \quad (92)$$

$$= 1 - P_V \left[ v \in \bigcup_{w \in C} \left\{ v \in \mathcal{V} \mid d(v, w) \leq D \right\} \right] \quad (93)$$

$$= 1 - \sum_{w \in C} P_v \left[ d(V, w) \leq D \right]. \quad (94)$$

We now show that the right-hand side of (76) coincides with (94). Applying Lemma 1 to the hypothesis test in (76), yields

$$\alpha_\beta(P_V, Q_V) = \max_{\gamma' \geq 0} \left\{ P_V \left[ \frac{P_v(V)}{Q_v(V)} \leq \gamma' \right] + \gamma'Q_v \left[ \frac{P_v(V)}{Q_v(V)} > \gamma' \right] - \gamma' \beta \right\}. \quad (95)$$

Let

$$Q_v^C(v) \triangleq \frac{1}{g}P_v(v) \left( \frac{1}{M} \sum_{w \in C} e^{-\lambda d(v, w)} \right)^{-1} \quad (96)$$

where $g$ a normalizing factor and $\lambda \geq 0$ is to be defined later. Using $Q_v = Q_v^C$ and choosing $\gamma' = \frac{g}{M}e^{-\lambda D}$, we obtain the following lower bound to (95),

$$\alpha_\beta(P_V, Q_v^C) \geq P_V \left[ \sum_{w \in C} e^{-\lambda d(v, w)} \leq e^{-\lambda D} \right] + \frac{g}{M}e^{-\lambda D} \left( Q_v \left[ \sum_{w \in C} e^{-\lambda d(v, w)} > e^{-\lambda D} \right] - \beta \right). \quad (97)$$

where the probability $Q_v \left[ \cdot \right]$ is computed with respect to $Q_v = Q_v^C$.

For $\lambda \geq 0$ sufficiently large,

$$\sum_{w \in C} e^{-\lambda d(v, w)} > e^{-\lambda D} \iff \min_{w \in C} d(v, w) \leq D. \quad (98)$$
Therefore, in this case (97) becomes
\[ \alpha_\beta(P_V, Q^C_V) \geq P_V \left( \min_{w \in C} d(V, w) > D \right) + \frac{g}{M} e^{-\lambda D} \left( Q^C_V \left( \min_{w \in C} d(V, w) \leq D \right) - \beta \right). \]  
(99)

The symmetry conditions required in the theorem imply that the measure of the set \( \{ v \in V \mid d(v, w) \leq \delta \} \)
does not depend on \( w \in W \) for any \( \delta \geq 0 \). Then, since the sets \( \{ v \in V \mid d(v, w) \leq \delta \} \) are non-overlapping,for \( \lambda \geq 0 \) sufficiently large we obtain
\[ Q^C_V \left( \min_{w \in C} d(V, w) \leq D \right) = \sum_{w \in C} Q^C_V \left( d(V, w) \leq D \right) \]
(100)
\[ = M \sup_{w \in C} Q^C_V \left( d(V, w) \leq D \right), \]  
(101)
where in the last step we used that, for \( \lambda \geq 0 \) sufficiently large, \( Q^C_V (v) \) only depends on the distance to the closest \( w \in C \). Then, since the measure of the set \( \{ v \in V \mid d(v, w) = \delta \} \) does not depend on \( w \in W \) for any \( \delta \geq 0 \), neither does \( Q^C_V [d(V, w) = \delta] \) nor \( Q^C_V [d(V, w) \leq D] \) depend on \( w \in C \).

Therefore, for \( \beta = M \sup_{w \in C} Q \left( d(V, w) \leq D \right) \), (99) becomes
\[ \alpha_M \sup_{w \in C} Q[ d(V, w) \leq D] \left( P_V, Q^C_V \right) \geq 1 - \sum_{w \in C} P^C_V \left( d(V, w) \leq D \right). \]  
(102)
Since the left-hand side of (102) is a lower bound to \( P_{eq}(C, D) \), ans since the right-hand side of (102) coincides with (94), then we conclude that (76) holds with equality.

**Remark:** Note that for the choice \( Q^C_V = Q^C_V \), by letting \( \lambda \to \infty \), \( Q^C_V [d(V, w) \leq D] \) becomes independent of \( w \in C \). However, for this choice of \( Q^C_V \), the measure \( Q^C_V [d(V, w) \leq D] \) still depends on \( w \notin C \). Therefore, the proof technique presented here cannot be directly applied when the \( \beta \) parameter in (76) is relaxed from \( M \sup_{w \in C} Q \left( d(V, w) \leq D \right) \) to \( M \sup_{w \in W} Q \left( d(V, w) \leq D \right) \), as discussed in (79)-(80).

**REFERENCES**

[1] C. Shannon, R. Gallager, and E. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels. I,” *Information and Control*, vol. 10, no. 1, pp. 65 – 103, 1967.

[2] E. Haroutunian, “Estimates of the error exponents for the semi-continuous memoryless channel (in Russian),” *Probl. Peredachi Inform.*, vol. 4, no. 37-48, 1968.

[3] R. E. Blahut, “Hypothesis testing and information theory,” *IEEE Trans. Inf. Theory*, vol. IT-20, no. 4, pp. 405–417, July 1974.

[4] G. Wiechman and I. Sason, “An improved sphere-packing bound for finite-length codes over symmetric memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1962–1990, May 2008.

[5] Y. Altuğ and A. B. Wagner, “Refinement of the sphere-packing bound: Asymmetric channels,” *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1592–1614, March 2014.

[6] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

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

[8] G. Vazquez-Vilar, A. T. Campo, A. Guillén i Fàbregas, and A. Martínez, “Bayesian M-ary hypothesis testing: The meta-converse and Verdú-Han bounds are tight,” *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2324–2333, May 2016.

[9] M. Hamada, “A sufficient condition for a code to achieve the minimum decoding error probability—generalization of perfect and quasi-perfect codes,” *IEICE Trans. on Fund. of Electronics, Comm. and Comp. Sciences*, vol. E83-A, no. 10, pp. 1870–1877, Oct. 2000.

[10] R. M. Fano, *Transmission of Information: A statistical Theory of Communication*. Cambridge: Massachusetts Institute of Technology, 1961.
[11] J. Neyman and E. S. Pearson, “On the problem of the most efficient tests of statistical hypotheses,” *Phil. Trans. R. Soc. Lond. A*, vol. 231, no. 694-706, p. 289, 1933.

[12] J. L. Massey. (1998) Applied Digital Information Theory I, Lecture Notes. [Online]. Available: http://www.isi.ee.ethz.ch/research.html

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. NJ: Wiley-Interscience, 2006.

[14] Y. Polyanskiy, “Saddle point in the minimax converse for channel coding,” *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.

[15] S. Roman, *Coding and Information Theory*. Springer-Verlag, 1992.

[16] P.-N. Chen, H.-Y. Lin, and S. Moser, “Optimal ultrasmall block-codes for binary discrete memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7346–7378, Nov 2013.

[17] H.-Y. Lin, S. M. Moser, and P.-N. Chen, “Weak flip codes and their optimality on the binary erasure channel,” *preprint arXiv:1711.03310v1*, 2017.

[18] G. Seroussi and R. M. Roth, “On MDS extensions of generalized Reed-Solomon codes,” *IEEE Trans. Inf. Theory*, vol. 32, no. 3, pp. 349–354, May 1986.

[19] V. Kostina and S. Verdú, “Lossy joint source-channel coding in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[20] ——, “Fixed-length lossy compression in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.