Ensemble Performance of Biometric Authentication Systems Based on Secret Key Generation

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E–mail: merhav@ee.technion.ac.il

Abstract

We study the ensemble performance of biometric authentication systems, based on secret key generation, which work as follows. In the enrollment stage, an individual provides a biometric signal that is mapped into a secret key and a helper message, the former being prepared to become available to the system at a later time (for authentication), and the latter is stored in a public database. When an authorized user requests authentication, claiming his/her identity as one of the subscribers, he/she has to provide a biometric signal again, and then the system, which retrieves also the helper message of the claimed subscriber, produces an estimate of the secret key, that is finally compared to the secret key of the claimed user. In case of a match, the authentication request is approved, otherwise, it is rejected. Referring to an ensemble of systems based on Slepian–Wolf binning, we provide a detailed analysis of the false–reject (FR) and false–accept (FA) probabilities, for a wide class of stochastic decoders. We also derive converse bounds. The converse bound of the FA probability matches the direct theorem, whereas the one for the FR probability is tight for some ranges of rates. Finally, we outline derivations of the secrecy leakage (for the typical code in the ensemble) and the privacy leakage.

Index Terms: biometric security, Slepian-Wolf coding, random binning, error exponents, secrete key generation.

I. Introduction

We consider a biometric authentication system that is described in [9, Sections 2.2–2.6], which is based on the notion of secret key generation and sharing due to Maurer [10] and Ahlswede and Csiszár [1], [2]. Specifically, such a system works as follows. In the enrollment stage, an individual which subscribes to the system, provides a biometric signal, $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$. The system receives this signal and generates (using its encoder) two outputs in response. The first output is a secret key, $\boldsymbol{S}$, at rate R_{s} and the second is a helper message, $\boldsymbol{W}$, at rate R_{w}. The secret key is prepared in order to be used by the system later, at the authentication stage. The helper message is stored in a public database. When an authorized user (a subscriber) wishes to sign in, claiming his/her identity as one of the existing subscribers, he/she is requested to provide again his/her biometric signal, $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ (correlated to $\boldsymbol{X}$, if indeed from the same individual, or independent, if not). The system then retrieves the helper message $\boldsymbol{W}$ of the claimed subscriber, and responds (using its decoder) by estimating the secret key, $\hat{\boldsymbol{S}}$ (based on $(\boldsymbol{Y}, \boldsymbol{W})$), and comparing it to the secret key of the claimed user, $\boldsymbol{S}$. In case of a match, access to the system is granted, otherwise, it is denied.

In [9, Sect. 2.3], achievable rate pairs $(R_{\mathrm{s}}, R_{\mathrm{w}})$ were found for the existence of systems (encoders and decoders) that satisfy the following three requirements in the large n limit: (i) arbitrarily small false–reject (FR) probability, (ii) arbitrarily small false–accept (FA) probability, (iii) arbitrarily small secrecy leakage, $I(\boldsymbol{S}; \boldsymbol{W})/n$, and (iv) privacy leakage, $I(\boldsymbol{X}; \boldsymbol{W})/n$, as small as possible. In particular, Theorem 2.1 of [9] asserts that when $(\boldsymbol{X}, \boldsymbol{Y})$ are drawn from a discrete memoryless source (DMS), generating independent copies of a correlated pair $(X, Y) \sim P_{XY}$, the maximum achievable key rate, R_{s}, under the above constraints, is given by the single–letter mutual information, $I(X; Y)$. It then follows that R_{w} must lie in the range $(H(X|Y), H(X) - R_{\mathrm{s}})$, where the conditional entropy in the lower limit is essential for reliable identification of an authorized subscriber (small FR probability) and it also sets the minimum possible privacy leakage, whereas the upper limit is essential for the secrecy leakage requirement. These limitations already guarantee that $R_{\mathrm{w}} < H(X)$, which is essential for keeping the FA probability vanishingly small for large n.

As in many proofs of direct coding theorems in the information theory literature, in the achievability part of [9, Theorem 2.1] too, the analyses of the error probabilities (in this case, the FA and the FR probabilities) are very rough – they are merely good enough to prove the achievability of the desired coding rates in the simplest possible manner. However, these are poor estimates of the achievable FR and FA probabilities themselves when these are considered to be the relevant performance metrics for given R_{s} and R_{w}.

The purpose of this paper is to provide sharper evaluations of the ensemble performance of the FA and the FR probabilities. In particular, referring to an ensemble of systems based on Slepian–Wolf binning, we provide detailed analyses of the exponential behavior of the FR probability, for a wide class of stochastic decoders, which includes the respective maximum a posteriori (MAP) decoder as a special case. An expurgated bound is provided as well and discussed quite in detail. For the FA probability, we analyze the ensemble performance of the MAP decoder and provide some intuition concerning its behavior. We also provide converse bounds for both the FR and the FA probabilities, which hold under some assumptions. The converse bound of the FA probability is tight in the sense that its error exponent matches the achievability result. Concerning the converse bound of the FR probability, which is essentially a version of the sphere–packing bound, there is a gap, in general, but at a certain (interesting) region in the plane of rates $(R_{\mathrm{w}}, R_{\mathrm{s}})$, it is tight. Finally, the secrecy leakage of the typical code in the ensemble, as well as the privacy leakage, are addressed.

The paper is organized as follows. In Section II, we establish the notation conventions. In Section III, we formalize the setup and spell out the objectives. In Section IV, we present and discuss the random coding FR exponent an expurgated bound, and the corresponding converse bound. In Section V, we derive the random coding FA exponent and its matching converse bound. In Section VI, we discuss the secrecy leakage of the typical code, and finally, in Section VII, we do the same with the privacy leakage. A brief summary is provided in Section VIII.

II. Notation Conventions

Throughout the paper, random variables will be denoted by capital letters, specific values they may take will be denoted by the corresponding lower case letters, and their alphabets will be denoted by calligraphic letters. Random vectors and their realizations will be denoted, respectively, by capital letters and the corresponding lower case letters, both in the bold face font. Their alphabets will be superscripted by their dimensions. For example, the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, ($n$ – positive integer) may take a specific vector value $\boldsymbol{x} = (x_1, \ldots, x_n)$ in $\mathcal{X}^n$, the n–th order Cartesian power of $\mathcal{X}$, which is the alphabet of each component of this vector. Sources and channels will be denoted by the letter P or Q, subscripted by the names of the relevant random variables/vectors and their conditionings, if applicable, following the standard notation conventions, e.g., Q_X, $P_{Y|X}$, and so on. When there is no room for ambiguity, these subscripts will be omitted. The probability of an event $\mathcal{G}$ will be denoted by $\Pr\{\mathcal{G}\}$, and the expectation operator with respect to (w.r.t.) a probability distribution P will be denoted by $\boldsymbol{E}_P\{\cdot\}$. Again, the subscript will be omitted if the underlying probability distribution is clear from the context. The entropy of a generic distribution Q on $\mathcal{X}$ will be denoted by $H_Q(X)$. For two positive sequences a_n and b_n, the notation $a_n \doteq b_n$ will stand for equality in the exponential scale, that is, $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$. Similarly, $a_n \stackrel{\cdot}{\le} b_n$ means that $\limsup_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} \le 0$, and so on. The indicator function of an event $\mathcal{G}$ will be denoted by $\mathcal{I}\{\mathcal{G}\}$. The notation $[x]_+$ will stand for $\max\{0, x\}$.

The empirical distribution of a sequence $\boldsymbol{x} \in \mathcal{X}^n$, which will be denoted by $\hat{P}_{\boldsymbol{x}}$, is the vector of relative frequencies $\hat{P}_{\boldsymbol{x}}(x)$ of each symbol $x \in \mathcal{X}$ in $\boldsymbol{x}$. The type class of $\boldsymbol{x} \in \mathcal{X}^n$, denoted $\mathcal{T}(\hat{P}_{\boldsymbol{x}})$, is the set of all vectors $\boldsymbol{x}'$ with $\hat{P}_{\boldsymbol{x}'} = \hat{P}_{\boldsymbol{x}}$. Information measures associated with empirical distributions will be denoted with 'hats' and will be subscripted by the sequences from which they are induced. For example, the entropy associated with $\hat{P}_{\boldsymbol{x}}$, which is the empirical entropy of $\boldsymbol{x}$, will be denoted by $\hat{H}_{\boldsymbol{x}}(X)$. Similar conventions will apply to the joint empirical distribution, the joint type class, the conditional empirical distributions and the conditional type classes associated with pairs (and multiples) of sequences of length n. Accordingly, $\hat{P}_{\boldsymbol{xy}}$ will be the joint empirical distribution of $(\boldsymbol{x}, \boldsymbol{y}) = \{(x_i, y_i)\}_{i=1}^n$, and $\mathcal{T}(\hat{P}_{\boldsymbol{xy}})$ will denote the joint type class of $(\boldsymbol{x}, \boldsymbol{y})$. Similarly, $\mathcal{T}(\hat{P}_{\boldsymbol{x}|\boldsymbol{y}}|\boldsymbol{y})$ will stand for the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{xy}}(X, Y)$ will designate the empirical joint entropy of $\boldsymbol{x}$ and $\boldsymbol{y}$, $\hat{H}_{\boldsymbol{xy}}(X|Y)$ will be the empirical conditional entropy, $\hat{I}_{\boldsymbol{xy}}(X;Y)$ will denote empirical mutual information, and so on. We will also use similar rules of notation in the context of a generic distribution, Q_{XY} (or Q, for short): we use $\mathcal{T}(Q_X)$ for the type class of sequences with empirical distribution Q_X, $H_Q(X)$ – for the corresponding empirical entropy, $\mathcal{T}(Q_{XY})$ – for the joint type class x, $T(Q_{X|Y}|\boldsymbol{y})$ – for the conditional type class of $\boldsymbol{x}$ given $\boldsymbol{y}$, $H_Q(X, Y)$ – for the joint empirical entropy, $H_Q(X|Y)$ – for the conditional empirical entropy, $I_Q(X;Y)$ – for the empirical mutual information, and so on. We will also use the customary notation for the weighted divergence,

$$D(Q_{Y|X}\|P_{Y|X}|Q_X)=\sum_{x\in\mathcal{X}}Q_X(x)\sum_{y\in\mathcal{Y}}Q_{Y|X}(y|x)\log\frac{Q_{Y|X}(y|x)}{P_{Y|X}(y|x)}. \tag{1}$$

III. Setup and Objectives

Consider the following system model for biometric identification (see Fig. 1). An *enrollment source sequence*, $\boldsymbol{x}=(x_1,\ldots,x_n)$, which is a realization of the random vector $\boldsymbol{X}=(X_1,\ldots,X_n)$, that emerges from a discrete memoryless source (DMS), P_X, with a finite alphabet $\mathcal{X}$, is fed into an *enrollment encoder*, $\mathcal{E}$, that produces two outputs: a secret key, $\boldsymbol{s}$ (a realization of a random variable $\boldsymbol{S}$), and a helper message, $\boldsymbol{w}$ (a realization of $\boldsymbol{W}$), taking on values in finite alphabets, $\mathcal{S}_n=\{0,1,\ldots,e^{nR_{\text{s}}}\}$ and $\mathcal{W}_n=\{0,1,\ldots,e^{nR_{\text{w}}}\}$, respectively, where R_{s} is the *secret–key rate*, and R_{w} is the *helper–message rate*. This encoding operation designates the enrollment stage.

We consider the ensemble of enrollment encoders, $\{\mathcal{E}\}$, generated by *random binning*, where for each source vector $\boldsymbol{x}\in\mathcal{X}$, one selects independently at random, both a secret key and a helper message, under the uniform distributions across $\mathcal{S}_n$ and $\mathcal{W}_n$, respectively. In other words, denoting by $\boldsymbol{w}=f(\boldsymbol{x})$ and $\boldsymbol{s}=g(\boldsymbol{x})$, the randomly selected bin assignments for both outputs, it is assumed that the $2|\mathcal{X}|^n$ random variables $\{f(\boldsymbol{x}),g(\boldsymbol{x})\}_{\boldsymbol{x}\in\mathcal{X}^n}$ are all mutually independent.

The *authentication decoder*, $\mathcal{A}$, which is aware of the randomly selected encoder, $\mathcal{E}$, is fed by two inputs: the helper message $\boldsymbol{w}$ and an *authentication source sequence*, $\boldsymbol{y}=(y_1,\ldots,y_n)$ (a realization of $\boldsymbol{Y}=(Y_1,\ldots,Y_n)$), that is produced at the output of a discrete memoryless channel (DMC), $P_{Y|X}$, with a finite output alphabet $\mathcal{Y}$, that is fed by $\boldsymbol{x}$. The output of the authentication decoder is $\hat{\boldsymbol{s}}=U(\boldsymbol{y},\boldsymbol{w})$ (a realization of $\hat{\boldsymbol{S}}$), which is an estimate (possibly, randomized) of the secret key, $\boldsymbol{s}$. If $\hat{\boldsymbol{s}}=\boldsymbol{s}$, access to the system is granted, otherwise, it is denied. This decoding operation stands for the authentication stage.

The optimal estimator of $\boldsymbol{s}$, based on $(\boldsymbol{y},\boldsymbol{w})$, in the sense of minimum FR probability, $\Pr\{\hat{\boldsymbol{S}}\neq\boldsymbol{S}\}$, is the maximum a posteriori probability (MAP) estimator, given by

$$\hat{\boldsymbol{s}}_{\text{MAP}}=U(\boldsymbol{y},\boldsymbol{w})\stackrel{\Delta}{=}\arg\max_{\boldsymbol{s}}P(\boldsymbol{s},\boldsymbol{w}|\boldsymbol{y})=\arg\max_{\boldsymbol{s}}\sum_{\boldsymbol{x}\in\mathcal{X}^n}P(\boldsymbol{x}|\boldsymbol{y})\cdot\mathcal{I}\{f(\boldsymbol{x})=\boldsymbol{w}\}\cdot\mathcal{I}\{g(\boldsymbol{x})=\boldsymbol{s}\}, \tag{2}$$

where $P(\boldsymbol{x}|\boldsymbol{y})$ (shorthand notation for $P_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})$) is the posterior probability of $\boldsymbol{X}=\boldsymbol{x}$ given $\boldsymbol{Y}=\boldsymbol{y}$, that is induced by the product distribution, P_{XY} (and the subscript XY will sometimes be suppressed for simplicity, when there is no risk of compromising clarity).

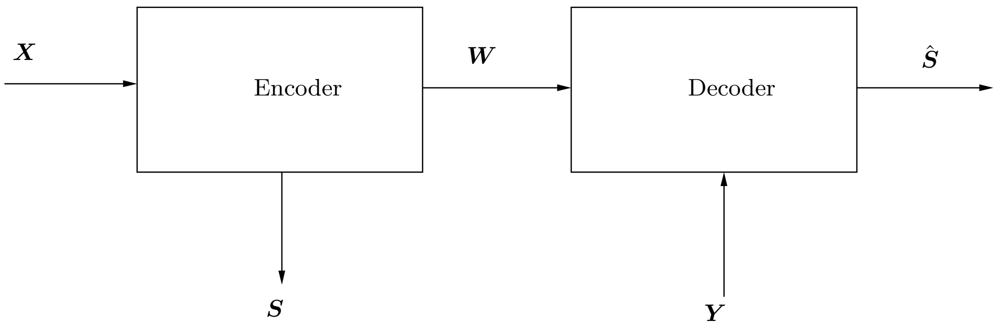

Figure 1: Biometric authentication system based on secret key generation.

In this paper, we expand the scope and study a more general class of decoders. This is a class of generalized stochastic likelihood decoders [12], [16], [17], [19], where the decoder randomly selects its estimate $\hat{\boldsymbol{s}}$ according to the posterior distribution

$$\tilde{P}(\boldsymbol{s}|\boldsymbol{y},\boldsymbol{w}) = \frac{\sum_{\boldsymbol{x}\in\mathcal{X}^n} \exp\{na(\hat{P}_{\boldsymbol{xy}})\} \cdot \mathcal{I}\{f(\boldsymbol{x})=\boldsymbol{w}\} \cdot \mathcal{I}\{g(\boldsymbol{x})=\boldsymbol{s}\}}{\sum_{\boldsymbol{x}\in\mathcal{X}^n} \exp\{na(\hat{P}_{\boldsymbol{xy}})\} \cdot \mathcal{I}\{f(\boldsymbol{x})=\boldsymbol{w}\}}, \tag{3}$$

where the function $a(\cdot)$, henceforth referred to as the *decoding metric*, is an arbitrary continuous function of the joint empirical distribution $\hat{P}_{\boldsymbol{xy}}$. Throughout the sequel, we will refer to the numerator of the r.h.s. as $\tilde{P}(\boldsymbol{s},\boldsymbol{w}|\boldsymbol{y})$, and to the denominator as $\tilde{P}(\boldsymbol{w}|\boldsymbol{y})$. The motivation for considering the generalized likelihood decoder is that it provides a unified framework for examining a large variety of decoders which are interesting both theoretically and practically. For example, with

$$a(\hat{P}_{\boldsymbol{xy}}) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \hat{P}_{\boldsymbol{xy}}(x,y)\ln P(x|y), \tag{4}$$

we have the ordinary likelihood decoder in spirit of [16], [17], [19]. For

$$a(\hat{P}_{\boldsymbol{xy}}) = \beta\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \hat{P}_{\boldsymbol{xy}}(x,y)\ln P(x|y), \tag{5}$$

β being a free parameter (sometimes referred to as the inverse temperature parameter [15] due to the analogy in statistical mechanics), we extend this likelihood decoder to a parametric family of decoders, where β controls the skewedness of the posterior. In particular, $\beta \to \infty$ leads to the ordinary MAP decoder, $\hat{\boldsymbol{s}}_{\text{MAP}}$. Other interesting choices are associated with mismatched metrics,

$$a(\hat{P}_{\boldsymbol{xy}}) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} \hat{P}_{\boldsymbol{xy}}(x,y)\ln P'(x|y), \tag{6}$$

P' being different from P, and

$$a(\hat{P}_{\boldsymbol{xy}}) = -\beta \hat{H}_{\boldsymbol{xy}}(X|Y), \tag{7}$$

which for $\beta \to \infty$ approaches the universal minimum entropy decoder (see also discussion around eqs. (5)–(7) of [12]).

An unauthorized user (i.e., an imposter), who claims for a given subscriber identity and wishes to break into the system, does not have the correlated biometric data $\boldsymbol{y}$. The best he/she can do is to estimate $\boldsymbol{s}$ based on the only data he/she has, which is the helper message $\boldsymbol{w}$, and then forges any fake biometric data $\tilde{\boldsymbol{y}}$, which together with $\boldsymbol{w}$, would cause the decoder to output this estimate of $\boldsymbol{s}$. More precisely, the imposter first estimates $\boldsymbol{s}$ according to

$$\tilde{\boldsymbol{s}} = V(\boldsymbol{w}) \stackrel{\Delta}{=} \arg\max_{\boldsymbol{s}} P(\boldsymbol{s}|\boldsymbol{w}) = \arg\max_{\boldsymbol{s}} \sum_{\boldsymbol{x}\in\mathcal{X}^n} P(\boldsymbol{x}) \cdot \mathcal{I}\{f(\boldsymbol{x}) = \boldsymbol{w}\} \cdot \mathcal{I}\{g(\boldsymbol{x}) = \boldsymbol{s}\}, \tag{8}$$

and then generates any $\tilde{\boldsymbol{y}} \in \mathcal{Y}^n$ such that $U(\tilde{\boldsymbol{y}}, \boldsymbol{w}) = \tilde{\boldsymbol{s}}$, and uses it as the biometric signal for authentication.

The objectives of the paper are to obtain: (i) exponential error bounds for the best achievable average FR probability, $\bar{P}_{\text{FR}} = \Pr\{\hat{\boldsymbol{S}} \neq \boldsymbol{S}\}$, associated with the generalized stochastic likelihood decoder (3), as well as an expurgated bound following the methodology of [12, Theorem 2] (see also the correction [13]), and (ii) exponential error bounds for the FA probability of (8), $\bar{P}_{\text{FA}} = \Pr\{\tilde{\boldsymbol{S}} = \boldsymbol{S}\}$. Finally, we outline derivations of the secrecy leakage, $I(\boldsymbol{S};\boldsymbol{W})$ (for the typical code in the ensemble) and the privacy leakage, in the large n limit.

IV. The False–Reject Error Exponent

A. Random Coding Exponent

Consider the system configuration described in Section III, along with the generalized stochastic likelihood decoder (3). Define the functions

$$E(R_{\text{w}}, Q_{X_0Y}) \stackrel{\Delta}{=} \min_{Q_{X|Y}} [R_{\text{w}} - H_Q(X|Y) + [a(Q_{X_0Y}) - a(Q_{XY})]_+]_+ \tag{9}$$

and

$$E_{\mathrm{r}}^{\mathrm{FR}}(R_{\mathrm{w}}) \stackrel{\Delta}{=} \min_{Q_{X_0 Y}} \{D(Q_{X_0 Y}\|P_{XY}) + E(R_{\mathrm{w}}, Q_{X_0 Y})\}. \tag{10}$$

Our first result is the following.

Theorem 1 *Consider the system configuration described in Section III. Then,*

$$\lim_{n\to\infty}\left[-\frac{\ln \bar{P}_{FR}}{n}\right] = E_r^{FR}(R_w). \tag{11}$$

Before providing the proof, a few points should be discussed.

1. First, observe that Theorem 1 asserts that $E_{\mathrm{r}}^{\mathrm{FR}}(R_{\mathrm{w}})$ is the *exact* random coding FR exponent, not just a lower bound. This is due to the fact that all steps of the analytic derivation are ensemble–tight in the exponential scale, thanks to the ability to avoid the use of the Jensen inequality and other well known tools that are traditionally used to facilitate the analysis, at the possible price of compromising tightness (see the proof of Theorem 1 below).

2. It is interesting to observe that the FR random coding exponent, $E_{\mathrm{r}}^{\mathrm{FR}}(R_{\mathrm{w}})$, depends only on R_{w}, not on R_{s}. This fact is not trivial, but the intuition is the following: in order to estimate $\boldsymbol{S}$ correctly, with high probability, from the given data $(\boldsymbol{Y}, \boldsymbol{W})$, there should be essentially no ambiguity, first of all, in defining what the correct $\boldsymbol{S}$ is. This will be the case if there is essentially only one source vector $\boldsymbol{X}$ that is responsible for the given $\boldsymbol{W}$ and then this $\boldsymbol{X}$ would dictate the correct $\boldsymbol{S} = g(\boldsymbol{X})$. This in turn would happen with high probability as long as $R_{\mathrm{w}} > H(X|Y)$. Otherwise, if more than one source vector (in the same conditional type class given $\boldsymbol{Y}$ as the correct one) is mapped by the encoder to the same helper message, then at least one such source vector is likely to be mapped to a different secret key message, and then the decoding would be ambiguous. It appears then that correct estimation of $\boldsymbol{S}$ is essentially equivalent to correct estimation of $\boldsymbol{X}$, as in ordinary Slepian–Wolf decoding [7] (see also [18] and references therein), where there is no secret key at all (or alternatively, $R_{\mathrm{s}} \to \infty$). Indeed, the Slepian–Wolf coding component of the joint source–channel coding system, analyzed in [12, Section IV] under the generalized likelihood decoder, contributes the very same error exponent as asserted in Theorem 1.

3. It is interesting to examine a few decoding metrics. Consider the choice $a(Q) = -H_Q(X|Y)$. In this case, we have

$$\begin{aligned}
& \min_{Q_{X|Y}} [R_{\rm w} - H_Q(X|Y) + [a(Q_{X_0Y}) - a(Q_{XY})]_+]_+ \\
= & \min_{Q_{X|Y}} [R_{\rm w} - H_Q(X|Y) + [H_Q(X|Y) - H_Q(X_0|Y)]_+]_+ \\
= & \min_{Q_{X|Y}} [R_{\rm w} - \min\{H_Q(X|Y), H_Q(X_0|Y)\}]_+ \\
= & [R_{\rm w} - \min\{\max_{Q_{X|Y}} H_Q(X|Y), H_Q(X_0|Y)\}]_+ \\
= & [R_{\rm w} - H_Q(X_0|Y)]_+, \qquad (12)
\end{aligned}$$

which, together with (10), yields the same random coding exponent as the optimal MAP decoder for Slepian–Wolf decoding (see also [12] and [16]). More generally, the same comment applies to $a(Q) = -\beta H_Q(X|Y)$ for every $\beta \geq 1$, where $\beta \to \infty$ pertains to the deterministic universal minimum entropy decoding, the source–coding dual to maximum mutual information (MMI) universal decoding (see, e.g., [18] and references therein). For $a(Q) = \beta \boldsymbol{E}_Q \ln P(X|Y)$, we have a finite–temperature likelihood decoder. For $\beta \to \infty$, we are back to the ordinary MAP decoder, which yields

$$\begin{aligned}
& \lim_{\beta\to\infty} \min_{Q_{X|Y}} [R_{\rm w} - H_Q(X|Y) + [a(Q_{X_0Y}) - a(Q_{XY})]_+]_+ \\
= & \lim_{\beta\to\infty} \min_{Q_{X|Y}} [R_{\rm w} - H_Q(X|Y) + \beta[\boldsymbol{E}_Q \ln P(X_0|Y) - \boldsymbol{E}_Q \ln P(X|Y)]_+]_+ \\
= & \min_{\{Q_{X|Y}:\ \boldsymbol{E}_Q \ln P(X|Y) \geq \boldsymbol{E}_Q \ln P(X_0|Y)\}} [R_{\rm w} - H_Q(X|Y)]_+, \qquad (13)
\end{aligned}$$

which, together with (10), yields the random coding exponent of the MAP decoder, as expected. As argued above, this is the same as the exponent achieved by $a(Q) = -\beta H_Q(X|Y)$ for all $\beta \geq 1$.

The remaining part of this subsection is devoted to the proof of Theorem 1.

Proof of Theorem 1. The expected FR probability is given by

$$\bar{P}_{\rm FR} = \boldsymbol{E}\left\{\sum_{\boldsymbol{s}\neq\boldsymbol{S}} \tilde{P}(\boldsymbol{s}|\boldsymbol{W},\boldsymbol{Y})\right\} \qquad (14)$$

where the expectation is w.r.t. both the randomness of $(\boldsymbol{S}, \boldsymbol{W}, \boldsymbol{Y})$ and the randomness of the code, $\mathcal{E}$. For given realizations, $\boldsymbol{X} = \boldsymbol{x}$ and $\boldsymbol{Y} = \boldsymbol{y}$, let us denote

$$\bar{P}_{\rm FR}(\boldsymbol{x},\boldsymbol{y}) \stackrel{\Delta}{=} \boldsymbol{E}\left\{\sum_{\boldsymbol{s}'\neq g(\boldsymbol{x})} \tilde{P}(\boldsymbol{s}'|f(\boldsymbol{x}),\boldsymbol{y})\right\}, \qquad (15)$$

where now the expectation is merely w.r.t. the randomness of $\mathcal{E}$. Now, following eq. (3),

$$\begin{aligned}\tilde{P}(\boldsymbol{s}'|f(\boldsymbol{x}),\boldsymbol{y}) &= \frac{\sum_{\boldsymbol{x}'\in\mathcal{X}^n}\exp\{na(\hat{P}_{\boldsymbol{x}'\boldsymbol{y}})\}\cdot\mathcal{I}\{f(\boldsymbol{x}')=f(\boldsymbol{x})\}\cdot\mathcal{I}\{g(\boldsymbol{x}')=\boldsymbol{s}'\}}{\sum_{\boldsymbol{x}'\in\mathcal{X}^n}\exp\{na(\hat{P}_{\boldsymbol{x}'\boldsymbol{y}})\}\cdot\mathcal{I}\{f(\boldsymbol{x}')=f(\boldsymbol{x})\}} \\ &= \frac{\sum_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}),\boldsymbol{s}')}{e^{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})}+\sum_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))}, \end{aligned} \tag{16}$$

where the summations over $\{Q_{X|Y}\}$ are across all conditional types $\{\mathcal{T}(Q_{X|Y}|\boldsymbol{y})\}$ of sequences of length n, and where

$$N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),\boldsymbol{w},\boldsymbol{s}') = \left|\mathcal{T}(Q_{X|Y}|\boldsymbol{y})\bigcap\{\boldsymbol{x}':\ f(\boldsymbol{x}')=\boldsymbol{w},\ g(\boldsymbol{x}')=\boldsymbol{s}'\}\right|, \tag{17}$$

and

$$N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),\boldsymbol{w}) = \left|\mathcal{T}(Q_{X|Y}\bigcap\{\boldsymbol{x}':\ f(\boldsymbol{x})=\boldsymbol{w},\ \boldsymbol{x}'\neq\boldsymbol{x}\}\right|. \tag{18}$$

Let us first consider the average FR probability for a given $(\boldsymbol{x},\boldsymbol{y})$ while fixing the realizations of $\boldsymbol{w}=f(\boldsymbol{x})$ and $\boldsymbol{s}=g(\boldsymbol{x})$:

$$\begin{aligned}\bar{P}_{\mathrm{FR}}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{s},\boldsymbol{w}) &= \boldsymbol{E}\left\{\frac{\sum_{\boldsymbol{s}'\neq\boldsymbol{s}}\sum_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}),\boldsymbol{s}')}{e^{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})}+\sum_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))}\right\} \\ &= \int_0^1 \mathrm{d}t\cdot\Pr\left\{\frac{\sum_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))}{e^{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})}+\sum_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))}\geq t\right\} \\ &= n\cdot\int_0^\infty \mathrm{d}\theta e^{-n\theta}\cdot\Pr\left\{\frac{\sum_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))}{e^{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})}+\sum_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))}\geq e^{-n\theta}\right\} \\ &\doteq \int_0^\infty \mathrm{d}\theta e^{-n\theta}\cdot\Pr\left\{\sum_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))>e^{n[a(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})-\theta]}\right\} \\ &\doteq \int_0^\infty \mathrm{d}\theta e^{-n\theta}\cdot\Pr\left\{\max_{Q_{X|Y}} e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))>e^{n[a(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})-\theta]}\right\} \\ &\doteq \int_0^\infty \mathrm{d}\theta e^{-n\theta}\cdot\Pr\bigcup_{Q_{X|Y}}\left\{e^{na(Q_{XY})}N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))>e^{n[a(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})-\theta]}\right\} \\ &\doteq \max_{Q_{X|Y}}\int_0^\infty \mathrm{d}\theta e^{-n\theta}\cdot\Pr\left\{N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))>e^{n[a(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})-a(Q_{XY})-\theta]}\right\}. \end{aligned} \tag{19}$$

Now, observe that $N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))$ is a binomial random variable with $|\mathcal{T}(Q_{X|Y}|\boldsymbol{y})|\doteq e^{nH_Q(X|Y)}$ trials and probability of success $e^{-nR_{\mathrm{w}}}$. Similarly as argued, e.g., in [12] (see page 5042, bottom half of the right column therein), we have

$$\Pr\left\{N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}),f(\boldsymbol{x}))>e^{n[a(Q_{X_0Y})-a(Q_{XY})-\theta]}\right\}\doteq e^{-nE(Q_{XY},Q_{X_0Y},\theta,R_{\rm w})}, \tag{20}$$

where we have replaced $\hat{P}_{\boldsymbol{xy}}$ by the notation Q_{X_0Y} (X_0 being an auxiliary random variable that represents the underlying source vector $\boldsymbol{x}$), and where

$$E(R_{\rm w},Q_{X_0Y},Q_{XY},\theta)=\left\{\begin{array}{ll}[R_{\rm w}-H_Q(X|Y)]_+ & \theta>a(Q_{X_0Y})-a(Q_{XY})-[H_Q(X|Y)-R_{\rm w}]_+\\ \infty & \theta\le a(Q_{X_0Y})-a(Q_{XY})-[H_Q(X|Y)-R_{\rm w}]_+\end{array}\right. \tag{21}$$

Thus,

$$\bar{P}_{\rm FR}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{s},\boldsymbol{w}) \doteq \max_{Q_{X|Y}}\int_{[a(Q_{X_0Y})-a(Q_{XY})-[H_Q(X|Y)-R_{\rm w}]_+]_+}^{\infty}{\rm d}\theta e^{-n\theta}\cdot e^{-n[R_{\rm w}-H_Q(X|Y)]_+}. \tag{22}$$

whose exponential decay rate is according to

$$\begin{aligned}
& \min_{Q_{X|Y}}\{[a(Q_{X_0Y})-a(Q_{XY})-[H_Q(X|Y)-R_{\rm w}]_+]_+ + [R_{\rm w}-H_Q(X|Y)]_+\}\\
=\ & \min_{Q_{X|Y}}\left\{\begin{array}{ll}[R_{\rm w}-H_Q(X|Y)+a(Q_{X_0Y})-a(Q_{XY})]_+ & H_Q(X|Y)>R_{\rm w}\\ R_{\rm w}-H_Q(X|Y)+[a(Q_{X_0Y})-a(Q_{XY})]_+ & H_Q(X|Y)\le R_{\rm w}\end{array}\right.\\
=\ & \min_{Q_{X|Y}}[R_{\rm w}-H_Q(X|Y)+[a(Q_{X_0Y})-a(Q_{XY})]_+]_+\\
=\ & E(R_{\rm w},Q_{X_0Y}).
\end{aligned} \tag{23}$$

The second to the last equality follows from the identity $[u-v]_+=[[u]_+-v]_+$, holding whenever $v\ge 0$, which is applied to the first line of the second expression with the assignments $u=a(Q_{X_0Y})-a(Q_{XY})$ and $v=H_Q(X|Y)-R_{\rm w}$ (see also [16] as well as the text after eq. (11) of [12] for a very similar argument). Since this exponential behavior, of $\bar{P}_{\rm FR}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{s},\boldsymbol{w})$, is independent of the particular realizations, $\boldsymbol{s}$ and $\boldsymbol{w}$, it holds also for the expectation w.r.t. the randomness of $\boldsymbol{S}$ and $\boldsymbol{W}$, namely, it also characterizes the exponential rate of $\bar{P}_{\rm FR}(\boldsymbol{x},\boldsymbol{y})$. Finally, it readily follows from the method of types [4] that the expectation w.r.t. the randomness of $(\boldsymbol{X},\boldsymbol{Y})$ decays according to the exponent

$$E_{\rm r}^{\rm FR}(R_{\rm w})=\min_{Q_{X_0Y}}\{D(Q_{X_0Y}\|P_{XY})+E(R_{\rm w},Q_{X_0Y})\}, \tag{24}$$

which is as defined in (10). This completes the proof of Theorem 1. □

B. Expurgated Bound

Our expurgated bound will be asserted for each type class, $\mathcal{T}(Q_X)$, of source vectors separately. As in channel coding, where expurgation is associated with elimination of some 'bad' codewords of a randomly generated code, here too, we might need to eliminate a small fraction of bad source vectors from $\mathcal{T}(Q_X)$, in order to guarantee a certain FR performance level for each one of the remaining source vectors in $\mathcal{T}(Q_X)$. One may wonder what would be the justification for such an elimination of source vectors, as these are generated by the source and given to us, and they are not under our control. Nonetheless, in the context of biometric authentication system described in Section III, where $\{\boldsymbol{x}\}$ are the enrollment signals, there are at least two possible ways to justify this elimination of a small fraction of the members of the type class.

1. In the enrollment stage, if the individual that subscribes to the system, has generated a 'forbidden' source vector $\boldsymbol{x}$ (in the sense that has been eliminated in the expurgation process), he/she might be asked to kindly provide his/her biometric signal once again, with the hope that this time a 'legitimate' source vector will be generated. The probability that this would happen is small in the first place, provided that the fraction of vectors eliminated from $\mathcal{T}(Q_X)$ is small. The probability of bothering the subscriber more than once with the request of a repeated measurement is even much smaller.

2. Considering the fact that $\boldsymbol{x}$ may be digitized with some precision (which is in line with the finite alphabet assumption anyway), it is conceivable to think of the enrollment data as having undergone a certain stage of vector quantization. Once $\boldsymbol{x}$ is thought of as an output of a vector quantizer, then not necessarily every member of $\mathcal{T}(Q_X)$ must be a legitimate codebook vector in the first place. Among other things, one might rule out source vectors that contribute a high FR probability.

In order to present the expurgated exponent, a few additional definitions are needed. For a given Q_Y, let us define

$$\alpha(R_{\text{w}}, Q_Y) \stackrel{\triangle}{=} \sup_{\{Q_{X|Y}:\ H_Q(X|Y)>R_{\text{w}}\}} [a(Q_{XY}) + H_Q(X|Y)] - R_{\text{w}}, \tag{25}$$

$$\gamma(Q_{XY}) \stackrel{\triangle}{=} \max\{a(Q_{XY}), \alpha(R_{\text{w}}, Q_Y)\}, \tag{26}$$

$$\Lambda(Q_{XX'}) \stackrel{\triangle}{=} \min_{Q_{Y|XX'}} \{\gamma(Q_{XY}) - H_Q(Y|X,X') - \boldsymbol{E}_Q \ln P(Y|X) - a(Q_{X'Y})\}, \tag{27}$$

and for a given Q_X, define

$$E_{\text{ex}}^{\text{FR}}(R_{\text{w}}, Q_X) = \inf_{\{Q_{X'|X}:\ H_Q(X'|X)\geq R_{\text{w}}\}} \{\Lambda(Q_{XX'}) - H_Q(X'|X) + R_{\text{w}}\}. \tag{28}$$

Finally, let $P_{\text{FR}}(\mathcal{E}|\boldsymbol{x})$ denote the FR probability of a given enrollment encoder $\mathcal{E}$, conditioned on the input source vector $\boldsymbol{X} = \boldsymbol{x}$.

Theorem 2 *Consider the system configuration described in Section III and let $\{\delta_n\}_{n\geq 1}$ be a positive sequence tending to zero such that $n\delta_n \to \infty$. Then, there exists a code $\mathcal{E}$ such that for every Q_X,*

$$P_{FR}(\mathcal{E}|\boldsymbol{x}) \leq \exp\{-nE_{ex}^{FR}(R_w, Q_X) + o(n)\}, \tag{29}$$

for every $\boldsymbol{x} \in \mathcal{T}(Q_X) \setminus \mathcal{B}(Q_X)$, where $\mathcal{B}(Q_X)$ is a certain subset of $\mathcal{T}(Q_X)$, whose size does not exceed $e^{-n\delta_n}|\mathcal{T}(Q_X)|$.

A few points concerning Theorem 2 should be discussed.

1. It is interesting to note that the expression of $E_{\text{ex}}^{\text{FR}}(R_{\text{w}}, Q_X)$ has some analogy to the Csiszár–Körner–Marton (CKM) expurgated exponent of channel coding [4, p. 165, Problem 10.18]. The term $\Lambda(Q_{XX'})$ plays the same role as the expected Bhattacharyya distance in the CKM expurgated exponent, whereas $H_Q(X'|X)$ is analogous to the coding rate R in channel coding and R_{w} is parallel to the empirical mutual information between channel codewords. Roughly speaking, the contribution of a single incorrect source vector $\boldsymbol{x}'$ to the FR probability is about $\exp\{-n\Lambda(Q_{XX'})\}$ provided that $(\boldsymbol{x}, \boldsymbol{x}') \in \mathcal{T}(Q_{XX'})$ (the pairwise error event). This probability should be multiplied by the typical number of such incorrect source vectors within $\mathcal{T}(Q_{X'|X}|\boldsymbol{x})$ that are encoded into the same given helper message and hence may cause confusion. This number is of the exponential order $\exp\{n[H_Q(X'|X) - R_{\text{w}}]\}$, provided that $H_Q(X'|X) - R_{\text{w}} > 0$, and it vanishes otherwise.

2. Note that in contrast to Theorem 1, here we are no longer arguing that the result is ensemble–tight. There is actually one step in the derivation where exponential tightness might be compromised. Specifically, in one of the steps of this analysis, the denominator of (3) is lower bounded by a relatively simple single–letter bound that holds true for the vast majority of encoders, $\{\mathcal{E}\}$, in the ensemble. By doing this, possible gaps to these bounds may not be fully exploited, and we cannot rule out the possibility that this causes some loss of tightness. Having said that, a very similar analysis that was recently carried out in the channel–coding counterpart [14] was shown to be ensemble–tight, and so, we speculate that this is the case here too. Also, the derivation of the expurgated bound includes a certain degree of freedom that does not exist in the random coding

bound of Theorem 1, and upon exploiting this degree of freedom, we obtain a result, which is at least as strong as the random coding bound, and sometimes strictly so.

3. The sequence δ_n tends to zero in order not to slow down the exponential decay rate, but it is also required that $n\delta_n \to \infty$ in order to guarantee that the set of 'bad' source vectors, $\mathcal{B}(Q_X)$, would be merely a minority of $\mathcal{T}(Q_X)$ for large n.

4. We now show that for every $R_{\rm w}$, the overall expurgated exponent (taking into account all types, $\{Q_X\}$) cannot be worse than $E_{\rm r}^{\rm FR}(R_{\rm w})$, at least for the metric $a(Q_{XY}) = -\beta H_Q(X|Y)$, which was shown to be as good as the optimal decoding metric in the ordinary random coding sense. Note that this is in contrast to the traditional expurgated exponent, which improves on the random coding exponent only at a certain range of rates, but is inferior to the random coding exponent elsewhere (see also [12], where a similar finding was observed for a particular numerical example). For the above–mentioned choice of $a(Q_{XY})$, one easily verifies that $\alpha(R_{\rm w}, Q_Y) = -\beta R_{\rm w}$ and $\gamma(Q_{XY}) = -\beta\min\{H_Q(X|Y), R_{\rm w}\}$, and so,

$$\begin{aligned}\Lambda(Q_{XX'}) &= \min_{Q_{Y|XX'}}\{\gamma(Q_{XY}) - H_Q(Y|X,X') - \boldsymbol{E}_Q \ln P(Y|X) + \beta H_Q(X'|Y)\}\\ &= \min_{Q_{Y|XX'}}\{\beta[H_Q(X'|Y) - \min\{H_Q(X|Y), R_{\rm w}\}] +\\ &\qquad I_Q(X';Y|X) + D(Q_{Y|X}\|P_{Y|X}|Q_X)\}. \qquad (30)\end{aligned}$$

Upon optimizing β, we obtain

$$\begin{aligned}E_{\rm ex}(R_{\rm w}, Q_X) &= \sup_{\beta\in\mathbb{R}} \inf_{\{Q_{X'|X}:\ H_Q(X'|X)\geq R_{\rm w}\}}\{\Lambda(Q_{XX'}) - H_Q(X'|X)\} + R_{\rm w}\\ &= \sup_{\beta\in\mathbb{R}} \inf_{\{Q_{X'Y|X}:\ H_Q(X'|X)\geq R_{\rm w}\}}\{D(Q_{Y|X}\|P_{Y|X}|Q_X) + I_Q(X';Y|X) +\\ &\qquad \beta[H_Q(X'|Y) - \min\{H_Q(X|Y), R_{\rm w}\}] - H_Q(X'|X) + R_{\rm w}\}\\ &\geq \inf_{\{Q_{X'Y|X}:\ H_Q(X'|X)\geq R_{\rm w}\}}\{D(Q_{Y|X}\|P_{Y|X}|Q_X) + I_Q(X';Y|X) +\\ &\qquad H_Q(X'|Y) - \min\{H_Q(X|Y), R_{\rm w}\} - H_Q(X'|X) + R_{\rm w}\}\\ &= \inf_{\{Q_{X'Y|X}:\ H_Q(X'|X)\geq R_{\rm w}\}}\{D(Q_{Y|X}\|P_{Y|X}|Q_X) + I_Q(X';Y|X) + H_Q(X'|Y) +\\ &\qquad [R_{\rm w} - H_Q(X|Y)]_+ - H_Q(X'|X)\\ &= \inf_{\{Q_{X'Y|X}:\ H_Q(X'|X)\geq R_{\rm w}\}}\{D(Q_{Y|X}\|P_{Y|X}|Q_X) + H_Q(X'|Y) - H_Q(X'|X,Y) +\end{aligned}$$

$$
\begin{aligned}
& [R_{\rm w} - H_Q(X|Y)]_+ \\
= & \inf_{\{Q_{X'Y|X}:\ H_Q(X'|X)\ge R_{\rm w}\}} \{D(Q_{Y|X}\|P_{Y|X}|Q_X) + I_Q(X';X|Y) + [R_{\rm w} - H_Q(X|Y)]_+\} \\
\ge & \inf_{\{Q_{X'Y|X}:\ H_Q(X'|X)\ge R_{\rm w}\}} \{D(Q_{Y|X}\|P_{Y|X}|Q_X) + [R_{\rm w} - H_Q(X|Y)]_+\}. \qquad (31)
\end{aligned}
$$

Without the constraint, $H_Q(X'|X) \ge R_{\rm w}$, the last expression is exactly the random coding FR exponent for a given type Q_X, and upon taking into account the probabilistic weight of each type, the overall exponent associated with the last line (again, without the constraint) is exactly $E_{\rm r}(R_{\rm m})$ of Theorem 1 for the optimal, MAP decoder. By inspection of eq. (31), we therefore observe that there are four origins of the gap between the expurgated exponent and the random coding exponent: (i) the decoder actually being analyzed might be suboptimal for the expurgated ensemble, (ii) the optimal β (for the given family of decoders) might not necessarily be $\beta^* = 1$ (the first inequality in the above chain). In fact, the optimal β^* is expected to depend on $R_{\rm w}$.[1] (iii) the term $I_Q(X';X|Y)$ which may not necessarily vanish for the optimal $Q_{X'Y|X}$ (the second inequality), and (iv) the constraint $H_Q(X'|X) \ge R_{\rm w}$. For example, if $R_{\rm w} > \ln|\mathcal{X}|$, the expurgated exponent is infinite while the random coding exponent is finite.

5. As can be seen in the proof of Theorem 2, the asserted expurgated exponent is obtained from an intermediate expression that depends on a free parameter ρ that undergoes optimization. It is interesting to observe what happens when we set $\rho = 1$ instead of optimizing over ρ. This would correspond to the ordinary ensemble average, which needs no expurgation. In this case, $E_{\rm ex}^{\rm FR}(R_{\rm w}, Q_X)$ would be replaced by

$$
\begin{aligned}
E_1(R_{\rm w}, Q_X) &= \sup_{\beta\in\mathbb{R}} \inf_{Q_{X'|X}} \{\Lambda(Q_{XX'}) - [H_Q(X'|X) - R_{\rm w}]_+ + [R_{\rm w} - H_Q(X'|X)]_+\} \\
&= \sup_{\beta\in\mathbb{R}} \inf_{Q_{X'|X}} \{\Lambda(Q_{XX'}) + R_{\rm w} - H_Q(X'|X)\}, \qquad (32)
\end{aligned}
$$

where we have used the trivial identity $[u]_+ - [-u]_+ \equiv u$. Therefore, the expression of $E_1(R_{\rm w}, Q_X)$ is exactly like that of $E_{\rm ex}^{\rm FR}(R_{\rm m}, Q_X)$, except that the constraint, $H_Q(X'|X) \ge R_{\rm w}$, is removed. It follows that $E_{\rm ex}^{\rm FR}(R_{\rm w}, Q_X)$ is expected to improve on $E_1(R_{\rm w}, Q_X)$ at high rates, where the constraint may be active. It also follows (similarly as in (31)) that $E_1(R_{\rm w}, Q_X)$ is never smaller than

[1] The fact that optimal β may not necessarily be infinite (except the case (5)), is interesting on its own right, as it means that the the stochastic decoder may outperform the deterministic one for a given (suboptimal) decoding metric.

the random coding FR exponent given the type Q_X, since the latter lacks this constraint as well. The reason that this expurgated exponent is nowhere worse than the random coding exponent is that we do not use the inequality $[\sum_{\boldsymbol{x}'\neq\boldsymbol{x}} u(\boldsymbol{x}')]^{1/\rho} \leq \sum_{\boldsymbol{x}'\neq\boldsymbol{x}}[u(\boldsymbol{x}')]^{1/\rho}$ (holding for $\rho \geq 1$), like in the traditional expurgated bound. This inequality causes a loss of tightness. Without it, the supremum over ρ is always achieved at $\rho \to \infty$.

6. The case of ordinary, deterministic MAP decoding is obtained again as of special case of (5) in the limit $\beta \to \infty$. As in (13), when the objective function to be minimized over $\{Q_{XX'Y}\}$, contains a term like $\beta \cdot G(Q_{XX'Y})$ (for some functional $G(\cdot)$), then in the limit of $\beta \to \infty$, it is replaced by a constraint of the form $G(Q_{XX'Y}) \leq 0$.

The remaining part of this subsection is devoted to the proof of Theorem 2.

Proof of Theorem 2. For a given code, $\mathcal{E}$, and a given the underlying source vector $\boldsymbol{x}$, we have

$$\begin{aligned} P_{\mathrm{FR}}(\mathcal{E}|\boldsymbol{x}) &= \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \tilde{P}(\boldsymbol{s}|f(\boldsymbol{x}),\boldsymbol{y}) & (33)\\ &= \sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \cdot \frac{\tilde{P}(\boldsymbol{s},f(\boldsymbol{x})|\boldsymbol{y})}{\exp\{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})\} + Z_{\boldsymbol{x}}(\boldsymbol{y})}, & (34)\end{aligned}$$

where

$$Z_{\boldsymbol{x}}(\boldsymbol{y}) = \sum_{\boldsymbol{x}'\neq\boldsymbol{x}} \exp\{na(\hat{P}_{\boldsymbol{x}'\boldsymbol{y}})\} \cdot \mathcal{I}\{f(\boldsymbol{x}') = f(\boldsymbol{x})\}. \tag{35}$$

Let $\epsilon > 0$ be arbitrarily small. It is shown in the Appendix[2] that

$$\Pr\left\{Z_{\boldsymbol{x}}(\boldsymbol{y}) < \exp\{n\alpha(R_{\mathrm{w}}+\epsilon,\hat{P}_{\boldsymbol{y}})\} \text{ for some } (\boldsymbol{x},\boldsymbol{y})\right\} \leq |\mathcal{X}\times\mathcal{Y}|^n \cdot \exp\{-e^{n\epsilon} + n\epsilon + 1\}. \tag{36}$$

Now, denoting

$$\mathcal{G}_\epsilon = \left\{\mathcal{E}:\ Z_{\boldsymbol{x}}(\boldsymbol{y}) \geq \exp\{n\alpha(R_{\mathrm{w}}+\epsilon,\hat{P}_{\boldsymbol{y}})\} \text{ for all } (\boldsymbol{x},\boldsymbol{y})\right\}, \tag{37}$$

we have:

$$\begin{aligned} &\boldsymbol{E}\left\{[P_{\mathrm{FR}}(\mathcal{E}|\boldsymbol{x})]^{1/\rho}\right\}\\ =\ &\boldsymbol{E}\left[\sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \cdot \frac{\tilde{P}(\boldsymbol{s},f(\boldsymbol{x})|\boldsymbol{y})}{\exp\{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})\} + Z_{\boldsymbol{x}}(\boldsymbol{y})}\right]^{1/\rho}\end{aligned}$$

[2]See also [12, Appendix B] for a similar argument related to channel coding.

$$
\begin{aligned}
&= \sum_{\mathcal{E}} P(\mathcal{E}) \left[\sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \cdot \frac{\tilde{P}(\boldsymbol{s}, f(\boldsymbol{x})|\boldsymbol{y})}{\exp\{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})\} + Z_{\boldsymbol{x}}(\boldsymbol{y})}\right]^{1/\rho} \\
&= \sum_{\mathcal{E}\in\mathcal{G}_\epsilon} P(\mathcal{E}) \left[\sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \cdot \frac{\tilde{P}(\boldsymbol{s}, f(\boldsymbol{x})|\boldsymbol{y})}{\exp\{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})\} + Z_{\boldsymbol{x}}(\boldsymbol{y})}\right]^{1/\rho} + \\
&\quad \sum_{\mathcal{E}\in\mathcal{G}_\epsilon^c} P(\mathcal{E}) \left[\sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \cdot \frac{\tilde{P}(\boldsymbol{s}, f(\boldsymbol{x})|\boldsymbol{y})}{\exp\{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})\} + Z_{\boldsymbol{x}}(\boldsymbol{y})}\right]^{1/\rho} \\
&\leq \sum_{\mathcal{E}\in\mathcal{G}_\epsilon} P(\mathcal{E}) \left[\sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \cdot \frac{\tilde{P}(\boldsymbol{s}, f(\boldsymbol{x})|\boldsymbol{y})}{\exp\{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})\} + \exp\{n\alpha(R_{\mathrm{w}}+\epsilon, \hat{P}_{\boldsymbol{y}})\}}\right]^{1/\rho} + \\
&\quad \sum_{\mathcal{E}\in\mathcal{G}_\epsilon^c} P(\mathcal{E}) \cdot 1^{1/\rho} \\
&\leq \sum_{\mathcal{E}} P(\mathcal{E}) \left[\sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \cdot \frac{\tilde{P}(\boldsymbol{s}, f(\boldsymbol{x})|\boldsymbol{y})}{\exp\{na(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})\} + \exp\{n\alpha(R_{\mathrm{w}}+\epsilon, \hat{P}_{\boldsymbol{y}})\}}\right]^{1/\rho} + \\
&\quad e^{nR_{\mathrm{s}}} \cdot |\mathcal{X}\times\mathcal{Y}|^n \cdot \exp\{-e^{n\epsilon} + n\epsilon + 1\}. \qquad (38)
\end{aligned}
$$

Considering the arbitrariness of ϵ, the expression in the square brackets is exponentially equivalent to

$$
\begin{aligned}
&\sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) e^{-n\gamma(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})} \tilde{P}(\boldsymbol{s}, f(\boldsymbol{x})|\boldsymbol{y}) \\
&= \sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) e^{-n\gamma(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})} \sum_{\boldsymbol{x}'} \exp\{na(\hat{P}_{\boldsymbol{x}'\boldsymbol{y}})\} \mathcal{I}\{f(\boldsymbol{x}') = f(\boldsymbol{x}), g(\boldsymbol{x}') = \boldsymbol{s}\} \\
&= \sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{x}'} \mathcal{I}\{f(\boldsymbol{x}') = f(\boldsymbol{x}), g(\boldsymbol{x}') = \boldsymbol{s}\} \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \exp\{n[a(\hat{P}_{\boldsymbol{x}'\boldsymbol{y}}) - \gamma(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})]\}. \qquad (39)
\end{aligned}
$$

Now, the inner most summation (over $\boldsymbol{y}$) can be assessed using the method of types [4]. Accordingly, referring to (27), we have

$$
e^{-n\Lambda(\hat{P}_{\boldsymbol{x}\boldsymbol{x}'})} \doteq \sum_{\boldsymbol{y}} P(\boldsymbol{y}|\boldsymbol{x}) \exp\{n[a(\hat{P}_{\boldsymbol{x}'\boldsymbol{y}}) - \gamma(\hat{P}_{\boldsymbol{x}\boldsymbol{y}})]\}, \qquad (40)
$$

which is the contribution of a single incorrect source vector $\boldsymbol{x}'$ to the FR probability. This yields

$$
\begin{aligned}
&\sum_{\boldsymbol{s}\neq g(\boldsymbol{x})} \sum_{\boldsymbol{x}'} \mathcal{I}\{f(\boldsymbol{x}') = f(\boldsymbol{x}), g(\boldsymbol{x}') = \boldsymbol{s}\} \cdot e^{-n\Lambda(\hat{P}_{\boldsymbol{x}\boldsymbol{x}'})} \\
&\leq \sum_{\boldsymbol{x}'} e^{-n\Lambda(\hat{P}_{\boldsymbol{x}\boldsymbol{x}'})} \mathcal{I}\{f(\boldsymbol{x}') = f(\boldsymbol{x})\} \\
&= \sum_{Q_{X'|X}} e^{-n\Lambda(Q_{XX'})} N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x})), \qquad (41)
\end{aligned}
$$

where we have defined

$$N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x})) \triangleq \left|\mathcal{T}(Q_{X'|X}|\boldsymbol{x}) \cap \{\boldsymbol{x}' :\ f(\boldsymbol{x}') = f(\boldsymbol{x})\}\right|. \tag{42}$$

On substituting this back into the bound on $\boldsymbol{E}\left\{[P_{\text{FR}}(\mathcal{E}|\boldsymbol{x})]^{1/\rho}\right\}$, we get

$$\begin{aligned}
& \boldsymbol{E}\left\{[P_{\text{e}}(\mathcal{E}|\boldsymbol{x})]^{1/\rho}\right\} \\
\leq\ & \boldsymbol{E}\left\{\left[\sum_{Q_{X'|X}} e^{-n\Lambda(Q_{XX'})} N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x}))\right]^{1/\rho}\right\} \\
\doteq\ & \sum_{Q_{X'|X}} e^{-n\Lambda(Q_{XX'})/\rho} \boldsymbol{E}\left\{[N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x}))]^{1/\rho}\right\} \\
=\ & \sum_{Q_{X'|X}} e^{-n\Lambda(Q_{XX'})/\rho} \int_0^\infty \mathrm{d}t \cdot \Pr\left\{\left[N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x}))\right]^{1/\rho} \geq t\right\} \\
=\ & \sum_{Q_{X'|X}} e^{-n\Lambda(Q_{XX'})/\rho} \int_0^\infty \mathrm{d}t \cdot \Pr\left\{N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x})) \geq t^\rho\right\} \\
\doteq\ & \sum_{Q_{X'|X}} e^{-n\Lambda(Q_{XX'})/\rho} \int_{-\infty}^\infty \mathrm{d}\theta \cdot e^{n\theta} \cdot \Pr\left\{N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x})) \geq e^{n\theta\rho}\right\}.
\end{aligned} \tag{43}$$

Let us focus on the term $\Pr[N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x})) \geq e^{n\theta\rho}]$. Since $N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x}))$ is a binomial random variable with $|\mathcal{T}(Q_{X'|X}|\boldsymbol{x})| \doteq e^{nH_Q(X'|X)}$ trials and probability of success $e^{-nR_{\text{w}}}$, we have

$$\Pr\left[N(\mathcal{T}(Q_{X'|X}|\boldsymbol{x}), f(\boldsymbol{x})) \geq e^{n\theta\rho}\right] \doteq e^{-nE(R_{\text{w}}, Q_{XX'}, \rho\theta)} \tag{44}$$

where

$$\begin{aligned}
E(R_{\text{w}}, Q_{XX'}, \rho\theta) &= \begin{cases} [R_{\text{w}} - H_Q(X'|X)]_+ & [H_Q(X'|X) - R_{\text{w}}]_+ \geq \rho\theta \\ \infty & [H_Q(X'|X) - R_{\text{w}}]_+ < \rho\theta \end{cases} \\
&= \begin{cases} [R_{\text{w}} - H_Q(X'|X)]_+ & \theta \leq [H_Q(X'|X) - R_{\text{w}}]_+/\rho \\ \infty & \theta > [H_Q(X'|X) - R_{\text{w}}]_+/\rho \end{cases}
\end{aligned} \tag{45}$$

On substituting this back into the expression of $\boldsymbol{E}\left\{[P_{\text{FR}}(\mathcal{E}|\boldsymbol{x})]^{1/\rho}\right\}$, we get

$$\begin{aligned}
& \boldsymbol{E}\left\{[P_{\text{FR}}(\mathcal{E}|\boldsymbol{x})]^{1/\rho}\right\} \\
\stackrel{.}{\leq}\ & \sum_{Q_{X'|X}} e^{-n\Lambda(Q_{XX'})/\rho} \cdot \int_{-\infty}^{[H_Q(X'|X)-R_{\text{w}}]_+/\rho} d\theta \cdot e^{n\theta} e^{-n[R_{\text{w}}-H_Q(X'|X)]_+} \\
\doteq\ & \exp\left\{-n \min_{Q_{X'|X}} \left[\Lambda(Q_{XX'}) + \rho[R_{\text{w}} - H_Q(X'|X)]_+ - [H_Q(X'|X) - R_{\text{w}}]_+\right]/\rho\right\}
\end{aligned}$$

$$\stackrel{\Delta}{=} \quad e^{-nE_{\rm x}(R_{\rm w},Q_X,\rho)/\rho}. \tag{46}$$

It follows then that

$$\boldsymbol{E}\left\{\frac{1}{|\mathcal{T}(Q_X)|}\sum_{\boldsymbol{x}\in\mathcal{T}(Q_X)}[P_{\rm FR}(\mathcal{E}|\boldsymbol{x})]^{1/\rho}\right\} \stackrel{\cdot}{\le} e^{-nE_{\rm x}(R_{\rm w},Q_X,\rho)/\rho}, \tag{47}$$

and so, there exists a code $\mathcal{E}$ with

$$\frac{1}{|\mathcal{T}(Q_X)|}\sum_{\boldsymbol{x}\in\mathcal{T}(Q_X)}[P_{\rm FR}(\mathcal{E}|\boldsymbol{x})]^{1/\rho} \stackrel{\cdot}{\le} e^{-nE_{\rm x}(R_{\rm w},Q_X,\rho)/\rho}. \tag{48}$$

For a given such $\mathcal{E}$ and Q_X, let us order the members of $\mathcal{T}(Q_X)$, as $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots$, according to $P_{\rm FR}(\mathcal{E}|\boldsymbol{x}_1) \ge P_{\rm FR}(\mathcal{E}|\boldsymbol{x}_2) \ge P_{\rm FR}(\mathcal{E}|\boldsymbol{x}_3) \ge \ldots$ and let M be a temporary short-hand notation for $|\mathcal{T}(Q_X)|$. Let $\mathcal{B}(Q_X)$ be the subset of $\mathcal{T}(Q_X)$ formed by the first $M' = e^{-\delta n}M$ members of $\mathcal{T}(Q_X)$ according to this order, i.e., $B(Q_X) = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{M'}\}$. We then have

$$\begin{aligned}
e^{-nE_{\rm x}(R_{\rm w},Q_X,\rho)/\rho} &\stackrel{\cdot}{\ge} \frac{1}{M}\sum_{m=1}^{M}[P_{\rm FR}(\mathcal{E}|\boldsymbol{x}_m)]^{1/\rho} \\
&\ge \frac{1}{M}\sum_{m=1}^{M'}[P_{\rm FR}(\mathcal{E}|\boldsymbol{x}_m)]^{1/\rho} \\
&\ge \frac{1}{M}\sum_{m=1}^{M'}[P_{\rm FR}(\mathcal{E}|\boldsymbol{x}_{M'+1})]^{1/\rho} \\
&= \frac{1}{M}\cdot M' \cdot [P_{\rm FR}(\mathcal{E}|\boldsymbol{x}_{M'+1})]^{1/\rho} \\
&= e^{-n\delta_n}\left[\max_{\boldsymbol{x}\in\mathcal{T}(Q_X)\setminus B(Q_X)} P_{\rm FR}(\mathcal{E}|\boldsymbol{x})\right]^{1/\rho},
\end{aligned} \tag{49}$$

and so, $\max_{\boldsymbol{x}\in\mathcal{T}(Q_X)\setminus B(Q_X)} P_{\rm FR}(\mathcal{E}|\boldsymbol{x})$ decays at an exponential rate which is at least as large as

$$\begin{aligned}
&\sup_{\rho\ge 0} E_{\rm x}(R_{\rm w}, Q_X, \rho) \\
&= \sup_{\rho\ge 0}\inf_{Q_{X'|X}}\{\Lambda(Q_{XX'}) - [H_Q(X'|X) - R_{\rm w}]_+ + \rho[R_{\rm w} - H_Q(X'|X)]_+\} \qquad (50)\\
&= \inf_{\{Q_{X'|X}:\ H_Q(X'|X)\ge R_{\rm w}\}}\{\Lambda(Q_{XX'}) - [H_Q(X'|X) - R_{\rm w}]_+\} \\
&= \inf_{\{Q_{X'|X}:\ H_Q(X'|X)\ge R_{\rm w}\}}\{\Lambda(Q_{XX'}) - H_Q(X'|X) + R_{\rm w}\} \\
&= E_{\rm x}(R_{\rm w}, Q_X), \qquad (51)
\end{aligned}$$

completing the proof of Theorem 2.

C. Converse Bound

We begin with a words of background. Owing to the duality between Slepian–Wolf coding and channel coding, we should mention the well known converse bound concerning the error exponent of Slepian–Wolf coding, i.e., the *sphere–packing* error exponent (see, e.g., [3, pp. 7–9] and references therein), which is given by

$$E_{\mathrm{sp}}^{\mathrm{SW}}(R_{\mathrm{w}}) = \min_{Q_X}[D(Q_X\|P_X) + E_{\mathrm{sp}}(Q_X, P_{Y|X}, H_Q(X) - R_{\mathrm{w}})] \tag{52}$$

where

$$E_{\mathrm{sp}}(Q_X, P_{Y|X}, R) = \min_{\{Q_{Y|X}:\ I_Q(X;Y)\le R\}} D(Q_{Y|X}\|P_{Y|X}|Q_X). \tag{53}$$

Upon substituting the latter into the former, we obtain

$$E_{\mathrm{sp}}^{\mathrm{SW}}(R_{\mathrm{w}}) = \min_{\{Q_{XY}:\ H_Q(X|Y)\ge R_{\mathrm{w}}\}} D(Q_{XY}\|P_{XY}). \tag{54}$$

As argued in [3, eq. (23)], the achievability associated with Slepian–Wolf MAP decoding (see also Subsection B below) and the sphere–packing converse bound coincide in the range $H(X|Y) \le R_{\mathrm{w}} \le R_{\mathrm{cr}} \stackrel{\triangle}{=} H(X'|Y')$, where (X', Y') is a pair of random variables, jointly distributed according to $P_{X'Y'}$, defined by

$$P_{Y'}(y) = \frac{P_Y(y)\left[\sum_x \sqrt{P_{X|Y}(x|y)}\right]^2}{\sum_{y'} P_Y(y')\left[\sum_x \sqrt{P_{X|Y}(x|y')}\right]^2} \tag{55}$$

$$P_{X'|Y'}(x|y) = \frac{\sqrt{P_{X|Y}(x|y)}}{\sum_{x'} \sqrt{P_{X|Y}(x'|y)}}. \tag{56}$$

In view of comments 2 and 3, of the discussion after Theorem 1, it follows that the FR random coding error exponent, associated with MAP decoding and minimum entropy decoding, also achieves the sphere–packing bound at this range of R_{w}. The Slepian–Wolf sphere–packing exponent also has an alternative expression due to Gallager [7], given by

$$E_{\mathrm{sp}}^{\mathrm{SW}}(R_{\mathrm{w}}) = \sup_{\rho>0}\left\{-\ln\left(\sum_y\left[\sum_x P_{XY}(x,y)^{1/(1+\rho)}\right]^{1+\rho}\right) + \rho R\right\}. \tag{57}$$

These are all well known results concerning Slepian–Wolf coding, but our problem here is somewhat different. In our case, the decoder should estimate $\boldsymbol{S} = g(\boldsymbol{X})$ based on $(\boldsymbol{Y}, \boldsymbol{W})$, and not $\boldsymbol{X}$

itself, as in Slepian–Wolf decoding. This is a less ambitious goal, but on the other hand, there are additional constraints on the system. In [9, p. 159], it is assumed that $\boldsymbol{X}$ can be reliably estimated from $(\boldsymbol{S}, \boldsymbol{W})$ (which can be motivated by the desire to make the encoder information lossless w.r.t. the source, i.e., the system essentially keeps a record of the full biometric signature). Accordingly, let us assume an encoder that maps each type $\boldsymbol{x}$ within $\mathcal{T}(Q_X)$ into a different pair $(\boldsymbol{s}, \boldsymbol{w})$, as long as $H(Q_X) < R_{\rm s} + R_{\rm w}$. We refer to this condition as the *estimability condition.* Our converse bound is now stated in the following theorem.

Theorem 3 *Consider the system configuration of Section III, where $f : \mathcal{X}^n \to \mathcal{W}_n$ and $g : \mathcal{X}^n \to \mathcal{S}_n$ are arbitrary functions that satisfy the mentioned estimability condition of the source vector. Let U be an arbitrary decoder and define the set*

$$\mathcal{Q}(R_w, R_s) = \{Q_{XY} :\ R_w < H_Q(X|Y),\ R_w + R_s > H_Q(X)\}\,. \tag{58}$$

Then,

$$P_{FR} \stackrel{\cdot}{\geq} \exp\left\{-n\left[\inf_{Q_{XY}\in\mathcal{Q}(R_w,R_s)} D(Q_{XY}\|P_{XY}) + o(n)\right]\right\}. \tag{59}$$

We point out that whenever the second constraint that defines $\mathcal{Q}(R_{\rm w}, R_{\rm s})$ becomes inactive, the exponential rate of this lower bound agrees with $E_{\rm sp}^{\rm SW}(R_{\rm w})$. This is the case when $R_{\rm s} > H_{Q^*}(X) - R_{\rm w}$, where $Q^* = Q^*_{XY}$ is the achiever of $E_{\rm sp}^{\rm SW}(R_{\rm w})$ as it appears in (54). Therefore, as described above, if in addition, $R_{\rm w} \in [H(X|Y), R_{\rm cr}]$, then the converse bound is tight.

The remaining part of this subsection is devoted to the proof of this theorem.

Proof of Theorem 3. Let Q_{XY} be any joint distribution with the property $H_Q(X) < R_{\rm s} + R_{\rm w}$ and let us denote all information measures associated with Q_{XY} by using the subscript Q. We are assuming that the encoder maps every type with empirical entropy less than $R_{\rm s} + R_{\rm w}$ in a one–to–one manner. The following two chains of inequalities are taken from [8] with minor twists. The first is:

$$\begin{aligned} I_Q(\boldsymbol{S};\boldsymbol{W}) &= H_Q(\boldsymbol{S}) + H_Q(\boldsymbol{W}) - H_Q(\boldsymbol{S},\boldsymbol{W}) \\ &= H_Q(\boldsymbol{S}) + H_Q(\boldsymbol{W}) - H_Q(\boldsymbol{S},\boldsymbol{W},\boldsymbol{X}) + H_Q(\boldsymbol{X}|\boldsymbol{S},\boldsymbol{W}) \\ &= H_Q(\boldsymbol{S}) + H_Q(\boldsymbol{W}) - H_Q(\boldsymbol{X}) + H_Q(\boldsymbol{X}|\boldsymbol{S},\boldsymbol{W}) \\ &= H_Q(\boldsymbol{S}) + H_Q(\boldsymbol{W}) - nH_Q(X) + H_Q(\boldsymbol{X}|\boldsymbol{S},\boldsymbol{W}) \end{aligned}$$

$$\leq \quad H_Q(\boldsymbol{S}) + nR_{\mathrm{w}} - nH_Q(X) + n\epsilon, \tag{60}$$

for some arbitrary $\epsilon > 0$, where the last step follows from the assumption on the encoder and Fano's inequality. The second chain of inequalities from [8] is the following:

$$\begin{aligned}
H_Q(\boldsymbol{S}) &= I_Q(\boldsymbol{S};\boldsymbol{Y},\boldsymbol{W}) + H_Q(\boldsymbol{S}|\boldsymbol{Y},\boldsymbol{W}) \\
&\leq I_Q(\boldsymbol{S};\boldsymbol{Y},\boldsymbol{W}) + P_{\mathrm{FR}}(Q_{XY}) \cdot nR_{\mathrm{s}} + 1 \\
&= I_Q(\boldsymbol{S};\boldsymbol{W}) + I_Q(\boldsymbol{S};\boldsymbol{Y}|\boldsymbol{W}) + P_{\mathrm{FR}}(Q_{XY}) \cdot nR_{\mathrm{s}} + 1 \\
&\leq I_Q(\boldsymbol{S};\boldsymbol{W}) + H_Q(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X},\boldsymbol{S},\boldsymbol{W}) + P_{\mathrm{FR}}(Q_{XY}) \cdot nR_{\mathrm{s}} + 1 \\
&\leq I_Q(\boldsymbol{S};\boldsymbol{W}) + H_Q(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}) + P_{\mathrm{FR}}Q_{XY}) \cdot nR_{\mathrm{s}} + 1 \\
&= I_Q(\boldsymbol{S};\boldsymbol{W}) + nI_Q(X;Y) + P_{\mathrm{FR}}(Q_{XY}) \cdot nR_{\mathrm{s}} + 1,
\end{aligned} \tag{61}$$

where $P_{\mathrm{FR}}(Q_{XY})$ is the FR probability induced by the auxiliary source Q_{XY}. On substituting the upper bound (61) on $H_Q(\boldsymbol{S})$ into (60), we obtain

$$\begin{aligned}
I_Q(\boldsymbol{S};\boldsymbol{W}) &\leq I_Q(\boldsymbol{S};\boldsymbol{W}) + nI_Q(X;Y) + P_{\mathrm{FR}}(Q_{XY}) \cdot nR_{\mathrm{s}} + nR_{\mathrm{w}} - nH_Q(X) + n\epsilon + 1 \\
&= I_Q(\boldsymbol{S};\boldsymbol{W}) - nH_Q(X|Y) + P_{\mathrm{FR}}(Q_{XY}) \cdot nR_{\mathrm{s}} + nR_{\mathrm{w}} + n\epsilon + 1,
\end{aligned} \tag{62}$$

which yields

$$P_{\mathrm{FR}}(Q_{XY}) \geq \frac{nH_Q(X|Y) - nR_{\mathrm{w}} - n\epsilon - 1}{nR_{\mathrm{s}}}. \tag{63}$$

Now, consider the following standard argument for changing probability measures. Let $\mathcal{E} = \{(\boldsymbol{x},\boldsymbol{y}) : \ g(\boldsymbol{x}) \neq U(\boldsymbol{y}, f(\boldsymbol{x}))\}$ denote the error event for a given encoder, (f,g), and decoder U. For a given $Q_{XY} \in \mathcal{Q}(R_{\mathrm{w}}, R_{\mathrm{s}})$ and an arbitrarily small $\epsilon > 0$, define also

$$\mathcal{G}_\epsilon = \left\{(\boldsymbol{x},\boldsymbol{y}) : \ \left|\frac{1}{n}\sum_{i=1}^n \ln\frac{Q_{XY}(x_i,y_i)}{P_{XY}(x_i,y_i)} - D(Q_{XY}\|P_{XY})\right| \leq \epsilon\right\}. \tag{64}$$

Then,

$$\begin{aligned}
P_{\mathrm{FA}} &= \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{E}} \prod_{i=1}^n P_{XY}(x_i,y_i) \\
&\geq \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{E}\cap\mathcal{G}_\epsilon} \left[\prod_{i=1}^n Q_{XY}(x_i,y_i)\right] \cdot \exp\left\{-\sum_{i=1}^n \ln\frac{Q_{XY}(x_i,y_i)}{P_{XY}(x_i,y_i)}\right\} \\
&\geq \sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{E}\cap\mathcal{G}_\epsilon} \left[\prod_{i=1}^n Q_{XY}(x_i,y_i)\right] \cdot \exp\{-n[D(Q_{XY}\|P_{XY}) + \epsilon]\}
\end{aligned}$$

$$
\begin{aligned}
&\geq \exp\{-n[D(Q_{XY}\|P_{XY})+\epsilon]\}\cdot\left[\sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{E}}\prod_{i=1}^{n}Q_{XY}(x_i,y_i)-\sum_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{G}_\epsilon^c}\prod_{i=1}^{n}Q_{XY}(x_i,y_i)\right]\\
&= \exp\{-n[D(Q_{XY}\|P_{XY})+\epsilon]\}[P_{\text{FA}}(Q_{XY})-o(n)]\\
&\geq \exp\{-n[D(Q_{XY}\|P_{XY})+\epsilon]\}\cdot\left[\frac{H_Q(X|Y)-R_{\text{w}}-\epsilon}{R_{\text{s}}}-o(n)\right]. \qquad (65)
\end{aligned}
$$

The proof is completed by selecting Q_{XY} to be the minimizer of $D(Q_{XY}\|P_{XY})$ across $\mathcal{Q}(R_{\text{w}}+2\epsilon, R_{\text{s}})$, and by using the arbitrariness of $\epsilon > 0$, as well as the continuity of $H_Q(X|Y)$ and $D(Q_{XY}\|P_{XY})$ as functionals of Q_{XY}.

V. The False–Accept Error Exponent

In this section, we analyze the ensemble performance of the system from the viewpoint of an imposter who makes an attempt to estimate the secret key without access to the side information $\boldsymbol{Y}$, and we are interested in the exponential decay rate of the FA probability. This section is divided into two parts: Subsection A is devoted to the direct theorem and Subsection B focuses on the converse theorem.

A. Direct Theorem

Here we analyze the FA probability for the average code. As described in Section III, here we assume that the imposter estimates $\boldsymbol{S}$ using the MAP estimator, $\tilde{\boldsymbol{S}}$ (see (8)), based on the helper message only. Accordingly, as defined in Section III, we denote $\bar{P}_{\text{FA}} = \Pr\{\tilde{\boldsymbol{S}} = \boldsymbol{S}\}$, i.e., the probability of correct decoding (FA), averaged over the ensemble of codes $\{\mathcal{E}\}$. Let us define

$$E_{\text{FA}}(R_{\text{w}}, R_{\text{s}}) = \min_{Q_X}\left[D(Q_X\|P_X) + \min\{R_{\text{s}}, [H_Q(X) - R_{\text{w}}]_+\}\right]. \qquad (66)$$

Our main result, in this subsection, is the following.

Theorem 4 *Consider the system configuration described in Section III. Then,*

$$\bar{P}_{FA} \leq \exp\{-nE_{FA}(R_w, R_s) + o(n)\}. \qquad (67)$$

The expression of this exponential error bound is quite intuitive and it can easily be understood to hold even if the imposter is informed about the type[3] Q_X of $\boldsymbol{X}$. There are about $e^{n[H_Q(X)-R_{\text{w}}]_+}$

[3]Here a genie–aided decoding argument does not harm the tightness of the FA exponent, because one can guess the type correctly with probability of success that decays only polynomially.

source sequences of type Q_X (including the correct one), whose helper message is the given $\boldsymbol{W}$. If $[H_Q(X) - R_{\rm w}]_+ > R_{\rm s}$, then all possible $e^{nR_{\rm s}}$ members of the secret–message set would be likely to appear as encoded secret messages among those sequences, approximately evenly, so the probability of guessing the correct one is about $e^{-nR_{\rm s}}$. If, on the other hand, $[H_Q(X) - R_{\rm w}]_+ < R_{\rm s}$, then it is very likely that there would be only about $e^{n[H_Q(X)-R_{\rm w}]_+}$ different $\boldsymbol{s}$–messages, so the probability of guessing the correct one is the reciprocal, $e^{-n[H_Q(X)-R_{\rm w}]_+}$. It is easy to see that $E_{\rm FA}(R_{\rm w}, R_{\rm s})$ vanishes for $R_{\rm w} > H(X)$, as expected.

It is also interesting to observe that here, in contrast to the exponential FR bounds of Section IV, the exponent depends on both $R_{\rm w}$ and $R_{\rm s}$, and not only on $R_{\rm w}$. As expected, it is increasing in $R_{\rm s}$ and decreasing in $R_{\rm w}$.

The FA error exponent of Theorem 4 can also be presented in a Gallager–style form:

$$\begin{aligned}
E_{\rm FA}(R_{\rm w}, R_{\rm s}) &= \min_Q [D(Q_X \| P_X) + \min\{R_{\rm s}, [H_Q(X) - R_{\rm w}]_+\}] \\
&= \min_{Q_X} \min_{0\le s\le 1} \max_{0\le\rho\le 1} \{D(Q_X \| P_X) + sR_{\rm s} + (1-s)\rho[H_Q(X) - R_{\rm w}]\} \\
&= \min_{0\le s\le 1} \max_{0\le\rho\le 1} \min_{Q_X} \{D(Q_X \| P_X) + sR_{\rm s} + (1-s)\rho[H_Q(X) - R_{\rm w}]\} \\
&= \min_{0\le s\le 1} \max_{0\le\rho\le 1} \left\{-[1-\rho(1-s)] \ln\left[\sum_x P_X(x)^{1/[1-\rho(1-s)]}\right] + sR_{\rm s} - \rho(1-s)R_{\rm w}\right\} \\
&= \min_{0\le s\le 1} \max_{s\le\rho\le 1} \left\{-\rho \ln\left[\sum_x P_X(x)^{1/\rho}\right] + sR_{\rm s} - (1-\rho)R_{\rm w}\right\}. \qquad (68)
\end{aligned}$$

Proof of Theorem 4. In the derivation below, we let $\boldsymbol{x}_Q$ denote an arbitrary representative source vector $\boldsymbol{x}$ of type Q_X. The choice of this representative within $\mathcal{T}(Q_X)$ is completely immaterial since all members of $\mathcal{T}(Q_X)$ are equiprobable. Similarly as before, we also denote by $N(Q_X, \boldsymbol{w}, \boldsymbol{s})$ the number of members of $\mathcal{T}(Q_X)$ that are encoded into $(\boldsymbol{w}, \boldsymbol{s})$.

$$\begin{aligned}
\bar{P}_{\rm FA} &= \boldsymbol{E}\left\{\sum_{\boldsymbol{w}} \max_{\boldsymbol{s}} P(\boldsymbol{w}, \boldsymbol{s})\right\} \\
&= \boldsymbol{E}\left\{\sum_{\boldsymbol{w}} \max_{\boldsymbol{s}} \sum_{Q_X} P_X(\boldsymbol{x}_Q) \cdot N(Q_X, \boldsymbol{w}, \boldsymbol{s})\right\} \\
&\doteq \boldsymbol{E}\left\{\sum_{\boldsymbol{w}} \max_{\boldsymbol{s}} \max_{Q_X} P_X(\boldsymbol{x}_Q) \cdot N(Q_X, \boldsymbol{w}, \boldsymbol{s})\right\} \\
&\doteq \boldsymbol{E}\left\{\sum_{\boldsymbol{w}} \max_{Q_X} P_X(\boldsymbol{x}_Q) \max_{\boldsymbol{s}} N(Q_X, \boldsymbol{w}, \boldsymbol{s})\right\}
\end{aligned}$$

$$
\begin{aligned}
&\doteq \boldsymbol{E}\left\{\sum_{\boldsymbol{w}}\sum_{Q_X} P_X(\boldsymbol{x}_Q)\max_{\boldsymbol{s}} N(Q_X,\boldsymbol{w},\boldsymbol{s})\right\} \\
&\doteq \sum_{\boldsymbol{w}}\sum_{Q_X} P_X(\boldsymbol{x}_Q)\cdot \boldsymbol{E}\left\{\max_{\boldsymbol{s}} N(Q_X,\boldsymbol{w},\boldsymbol{s})\right\} \\
&= \sum_{\boldsymbol{w}}\sum_{Q_X} P_X(\boldsymbol{x}_Q)\cdot \sum_{n=1}^{|\mathcal{T}(Q_X)|} \Pr\left\{\max_{\boldsymbol{s}} N(Q_X,\boldsymbol{w},\boldsymbol{s})\ge n\right\} \\
&= \sum_{\boldsymbol{w}}\sum_{Q_X} P_X(\boldsymbol{x}_Q)\cdot \sum_{n=1}^{|\mathcal{T}(Q_X)|} \Pr\bigcup_{\boldsymbol{s}}\{N(Q_X,\boldsymbol{w},\boldsymbol{s})\ge n\} \\
&\le \sum_{\boldsymbol{w}}\sum_{Q_X} P_X(\boldsymbol{x}_Q)\cdot \sum_{n=1}^{|\mathcal{T}(Q_X)|} \min\left\{1, e^{nR_{\mathrm{s}}}\Pr\left[N(Q_X,\boldsymbol{w},\boldsymbol{s})\ge n\right]\right\}. \qquad (69)
\end{aligned}
$$

Now, for $Q_X \in \mathcal{G} \stackrel{\Delta}{=} \{Q_X :\; H_Q(X) > R_{\mathrm{s}} + R_{\mathrm{w}}\}$, clearly, $\Pr[N(Q_X,\boldsymbol{w},\boldsymbol{s}) \ge n]$ is large for every $n \le e^{n[H_Q(X)-R_{\mathrm{w}}-R_{\mathrm{s}}-\epsilon]}$ (for an arbitrarily small $\epsilon > 0$ and large n), and so, the minimum between 1 and $e^{nR_{\mathrm{s}}}\Pr\left[N(Q_X,\boldsymbol{w},\boldsymbol{s})\ge n\right]$ is certainly 1. Hence, these terms, of the summation over n, contribute altogether a quantity of the exponential order of $e^{n[H_Q(X)-R_{\mathrm{w}}-R_{\mathrm{s}}]}$. For larger n, $\Pr[N(Q_X,\boldsymbol{w},\boldsymbol{s}) \ge n]$ decays super–exponentially, and so, these terms contribute a negligible amount. Consequently, considering the factor of $e^{nR_{\mathrm{w}}}$ that stems from the summation over $\boldsymbol{w}$, one term that contributes to the expression of the last line above is $\sum_{Q_X\in\mathcal{G}} P_X(\boldsymbol{x}_Q)e^{n[H_Q(X)-R_{\mathrm{s}}]}$, which is of the exponential order of $\exp\{-n\min_{Q_X\in\mathcal{G}}[D(Q_X\|P_X)+R_{\mathrm{s}}]\}$. The other term comes from the types that belong to $\mathcal{G}^c$. For $Q_X \in \mathcal{G}^c$, there are sub–exponentially few terms that contribute $\min\{1, e^{nR_{\mathrm{s}}}\cdot e^{n[H_Q(X)-R_{\mathrm{s}}-R_{\mathrm{w}}]}\} = e^{-n[R_{\mathrm{w}}-H_Q(X)]_+}$, and so, the overall contribution is $\max_{Q_X\in\mathcal{G}^c} e^{nR_{\mathrm{w}}}e^{-n[H_Q(X)+D(Q_X\|P_X)]}e^{-n[R_{\mathrm{w}}-H_Q(X)]_+}$, which is $\exp\{-n\min_{Q_X\in\mathcal{G}^c}[D(Q_X\|P_X) + [H_Q(X)-R_{\mathrm{w}}]_+]\}$. Thus, the overall performance is

$$
\bar{P}_{\mathrm{FA}} \stackrel{\cdot}{\le} \exp\left(-n\min_{Q_X}[D(Q_X\|P_X)+\min\{R_{\mathrm{s}},[H_Q(X)-R_{\mathrm{w}}]_+\}]\right), \qquad (70)
$$

completing the proof of Theorem 4.

B. Converse Theorem

We next provide a matching converse to Theorem 4. To this end, we will make the following regularity condition concerning the helper–message encoder, f: for each type class Q_X, the number of different $\{\boldsymbol{w}\}$ for which $N(Q_X,\boldsymbol{w}) \ge 1$ is the maximum possible number, that is,

$\min\{e^{nR_{\rm w}}, |\mathcal{T}(Q_X)|\} \doteq \exp[n\min\{R_{\rm w}, H_Q(X)\}$. Otherwise, the encoder can obviously be improved from the viewpoint of the legitimate subscriber,[4] by using a larger variety of $\{\boldsymbol{w}\}$, which would obviously make the helper message more informative for him/her.

Theorem 5 *Consider the system configuration of Section III, where $f : \mathcal{X}^n \to \mathcal{W}_n$ is an arbitrary helper–message encoder that satisfies the regularity assumption, $g : \mathcal{X}^n \to \mathcal{S}_n$ is an arbitrary secret–message encoder and $V : \mathcal{W}_n \to \mathcal{S}_n$ is an arbitrary decoder. Then,*

$$P_{FA} \stackrel{\Delta}{=} Pr\{V(\boldsymbol{W}) = \boldsymbol{S}\} \geq \exp\{-[nE_{FA}(R_w, R_s) + o(n)]\}. \tag{71}$$

It is interesting to point out that the proof of Theorem 5 (see below) provides a guideline concerning good encoders that minimize the FA probability: in the proof of this theorem, all inequalities become equalities if bins are allocated to source sequences as evenly as possible, both for the secret message encoder and the helper message encoder. This is to say that within type classes whose sizes are smaller than the total number of bins, each source sequence should be mapped into a different bin, and for the other type classes, all bins should be populated evenly by source sequences, at least in the exponential scale. Clearly, a typical realization of a randomly chosen random binning code has this property. It is speculated that certain other classes of (randomized) encoders share this property as well, such as those that are based on universal hash functions.

If the encoder does not satisfy the aforementioned regularity condition, but the number of $\{\boldsymbol{w}\}$ with $N(Q_X, \boldsymbol{w}) \geq 1$ is given instead by the exponential order of $\exp\{nW(Q_X)\}$ (for some $W(Q_X) \leq \min\{R_{\rm w}, H_Q(X)\}$), then $[H_Q(X) - R_{\rm w}]_+$, in the expression of $E_{\rm FA}(R_{\rm w}, R_{\rm s})$, should be replaced by $H_Q(X) - W(Q_X)$.

Proof of Theorem 5. Upon repeating the first five lines of eq. (69), but without the expectation w.r.t. the randomness of the encoder, we have

$$\begin{aligned} P_{\rm FA} &\doteq \sum_{\boldsymbol{w}} \sum_{Q_X} P_X(\boldsymbol{x}_Q) \max_{\boldsymbol{s}} N(Q_X, \boldsymbol{w}, \boldsymbol{s}) \\ &= \sum_{\boldsymbol{w}} \sum_{Q_X} P_X[\mathcal{T}(Q_X)] \cdot \frac{\max_{\boldsymbol{s}} N(Q_X, \boldsymbol{w}, \boldsymbol{s})}{|\mathcal{T}(Q_X)|}. \end{aligned} \tag{72}$$

[4]This is a reasonable assumption since the system is designed, first and foremost, for the authentication of legitimate subscribers. In the absence of such an assumption, there is no non–trivial lower bound to the FAR since one may use the degenerate encoder $f(\boldsymbol{x}) \equiv 0$ for all $\boldsymbol{x}$, which renders the helper message completely useless and then the FAR would be dropped to $e^{-nR_{\rm s}}$.

To lower bound the quantity $\max_{\boldsymbol{s}} N(Q_X, \boldsymbol{w}, \boldsymbol{s})$, we have the following consideration. Assume first that $N(Q_X, \boldsymbol{w}) > e^{nR_{\rm s}}$. Since $\sum_{\boldsymbol{s}} N(Q_X, \boldsymbol{w}, \boldsymbol{s}) = N(Q_X, \boldsymbol{w})$, the smallest possible value of $\max_{\boldsymbol{s}} N(Q_X, \boldsymbol{w}, \boldsymbol{s})$ is attained when $N(Q_X, \boldsymbol{w}, \boldsymbol{s}) = N(Q_X, \boldsymbol{w})/e^{nR_{\rm s}}$ for all $\boldsymbol{s}$. For $1 \leq N(Q_X, \boldsymbol{w}) \leq e^{nR_{\rm s}}$, we will lower bound $\max_{\boldsymbol{s}} N(Q_X, \boldsymbol{w}, \boldsymbol{s})$ by 1. Thus, in general, for every Q_X and $\boldsymbol{w}$ for which $N(Q_X, \boldsymbol{w}) \geq 1$,

$$\max_{\boldsymbol{s}} N(Q_X, \boldsymbol{w}, \boldsymbol{s}) \geq \max\{1, N(Q_X, \boldsymbol{w})e^{-nR_{\rm s}}\}. \tag{73}$$

On substituting the last lower bound into the above approximation of $P_{\rm FA}$, we obtain,

$$\begin{aligned}
P_{\rm FA} &\stackrel{\cdot}{\geq} \sum_{\{\boldsymbol{w}:\ N(Q_X,\boldsymbol{w})\geq 1\}} \sum_{Q_X} P_X[\mathcal{T}(Q_X)] \cdot \frac{\max\{1, N(Q_X, \boldsymbol{w})e^{-nR_{\rm s}}\}}{|\mathcal{T}(Q_X)|} \\
&= \sum_{Q_X} e^{n\min\{R_{\rm w}, H_Q(X)\}} P_X[\mathcal{T}(Q_X)] \cdot \frac{1}{e^{n\min\{R_{\rm w}, H_Q(X)\}}} \sum_{\{\boldsymbol{w}:\ N(Q_X,\boldsymbol{w})\geq 1\}} \frac{\max\{1, N(Q_X, \boldsymbol{w})e^{-nR_{\rm s}}\}}{|\mathcal{T}(Q_X)|}.
\end{aligned}$$

Now, the function $q(t) = \max\{1, te^{-nR_{\rm s}}\}$ is obviously convex, and so, since the number of different $\{\boldsymbol{w}\}$ in the f–image of every type Q_X is as large as $\min\{|\mathcal{T}(Q_X)|, e^{nR_{\rm w}}\}$, the above expression is further lower bounded by

$$\begin{aligned}
P_{\rm FA} &\stackrel{\cdot}{\geq} \sum_{Q_X} e^{n\min\{R_{\rm w}, H_Q(X)\}} P_X[\mathcal{T}(Q_X)] \cdot \frac{1}{|\mathcal{T}(Q_X)|} \times \\
&\quad \max\left\{1, \frac{1}{e^{n\min\{R_{\rm w}, H_Q(X)\}}} \sum_{\{\boldsymbol{w}:\ N(Q_X,\boldsymbol{w})\geq 1\}} N(Q_X, \boldsymbol{w})e^{-nR_{\rm s}}\right\} \\
&= \sum_{Q_X} e^{n\min\{R_{\rm w}, H_Q(X)\}} P_X[\mathcal{T}(Q_X)] \cdot \frac{1}{|\mathcal{T}(Q_X)|} \cdot \max\left\{1, |\mathcal{T}(Q_X)| \cdot e^{-n(R_{\rm s}+\min\{R_{\rm w}, H_Q(X)\}}\right\} \\
&\doteq \exp\left\{-n \min_{Q_X} (-\min\{R_{\rm w}, H_Q(X)\} + D(Q_X\|P_X) + \right. \\
&\quad H_Q(X) - [H_Q(X) - R_{\rm s} - \min\{R_{\rm w}, H_Q(X)\}]_+)\} \\
&= \exp\{-n\left(D(Q_X\|P_X) + \min\{R_{\rm s}, H_Q(X) - \min\{R_{\rm w}, H_Q(X)\}\}\right)\} \\
&= \exp\{-n\left(D(Q_X\|P_X) + \min\{R_{\rm s}, [H_Q(X) - R_{\rm w}]_+\}\right)\} \\
&= \exp\{-nE_{\rm FA}(R_{\rm w}, R_{\rm s})\}.
\end{aligned} \tag{74}$$

This completes the proof of Theorem 5.

VI. Secrecy Leakage for the Typical Code

In this section, we provide an outline for the evaluation of the secrecy leakage, $I(\boldsymbol{W};\boldsymbol{S})$, associated with the typical code, $\mathcal{E}$, in the ensemble.

We envision the typical code as a code with the following properties:

1. For any given type class $\mathcal{T}(Q_X)$ whose size is larger than $e^{n(R_\text{s}+R_\text{w})}$, the number of members of $\mathcal{T}(Q_X)$ mapped each one of the $e^{n(R_\text{s}+R_\text{w})}$ pairs $(\boldsymbol{s},\boldsymbol{w})$ is exactly the same (uniform distribution of $(\boldsymbol{S},\boldsymbol{W})$ within the type), so that $H(\boldsymbol{S},\boldsymbol{W}|\boldsymbol{X}\in\mathcal{T}(Q_X)) = n(R_\text{s}+R_\text{w})$.
2. For any given type class $\mathcal{T}(Q_X)$ whose size is smaller than $e^{n(R_\text{s}+R_\text{w})}$, each member of $\mathcal{T}(Q_X)$ is mapped to a different pair $(\boldsymbol{s},\boldsymbol{w})$, so that $H(\boldsymbol{S},\boldsymbol{W}|\boldsymbol{X}\in\mathcal{T}(Q_X)) = \log|\mathcal{T}(Q_X)|$.

The secrecy leakage will then be upper bounded as follows:

$$\begin{aligned}
I(\boldsymbol{S};\boldsymbol{W}) &= H(\boldsymbol{S})+H(\boldsymbol{W})-H(\boldsymbol{S},\boldsymbol{W})\\
&\le nR_\text{s}+nR_\text{w}-H(\boldsymbol{S},\boldsymbol{W}|\hat{P}_{\boldsymbol{X}})\\
&= n(R_\text{s}+R_\text{w})-\boldsymbol{E}\min\left\{n(R_\text{s}+R_\text{w}),\log|\mathcal{T}(\hat{P}_{\boldsymbol{X}})|\right\}\\
&= \boldsymbol{E}\left\{\left[n(R_\text{s}+R_\text{w})-\log|\mathcal{T}(\hat{P}_{\boldsymbol{X}})|\right]_+\right\}\\
&\approx n\boldsymbol{E}\left\{[R_\text{s}+R_\text{w}-\hat{H}_{\boldsymbol{X}}(X)]_+\right\}. \qquad (75)
\end{aligned}$$

Now, assuming that $H(X) > R_\text{s}+R_\text{w}$, the probability of falling in a type class $\mathcal{T}(\hat{P}_{\boldsymbol{x}})$ with $R_\text{s}+R_\text{w}-\hat{H}_{\boldsymbol{x}}(X)>0$ is of the exponential order of $\exp\{-nE_\text{sec}(R_\text{s}+R_\text{w})\}$, where

$$E_\text{sec}(R) \stackrel{\Delta}{=} \min\{D(Q_X\|P_X):\ H_Q(X)\le R\}, \qquad (76)$$

and therefore,

$$\begin{aligned}
I(\boldsymbol{S};\boldsymbol{W}) &\stackrel{\cdot}{\le} n\sum_{\boldsymbol{x}} P_X(\boldsymbol{x})[R_\text{s}+R_\text{w}-\hat{H}_{\boldsymbol{x}}(X)]\cdot\mathcal{I}\{R_\text{s}+R_\text{w}-\hat{H}_{\boldsymbol{x}}(X)>0\}\\
&\le n(R_\text{s}+R_\text{w})\cdot\Pr\{R_\text{s}+R_\text{w}-\hat{H}_{\boldsymbol{X}}(X)>0\}\\
&\doteq \exp\{-nE_\text{sec}(R_\text{s}+R_\text{w})\}, \qquad (77)
\end{aligned}$$

which means that as long as $H(X) > R_\text{s}+R_\text{w}$, strong security is guaranteed in the sense that $I(\boldsymbol{S};\boldsymbol{W})$ tends to zero even without normalization by n, as it decays exponentially fast. The secrecy exponent depends on R_s and R_w only via their sum, $R_\text{s}+R_\text{w}$.

VII. Privacy Leakage

In this last section, which is very brief, we study the privacy leakage – the amount of information that leaks from the biometric signature $\boldsymbol{X}$ to the helper message, $\boldsymbol{W}$, that is, the normalized mutual information, $I(\boldsymbol{X};\boldsymbol{W})/n$. Since $\boldsymbol{W}$ is a deterministic function of $\boldsymbol{X}$ for a given code, then $I(\boldsymbol{X};\boldsymbol{W}) = H(\boldsymbol{W})$. Thus,

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{W}) &= H(\boldsymbol{W}) \\ &= H(\boldsymbol{W}|\hat{P}_{\boldsymbol{X}}) + I(\boldsymbol{W};\hat{P}_{\boldsymbol{X}}) \\ &\le \sum_{\mathcal{T}(\hat{P}_{\boldsymbol{x}})} P[\mathcal{T}(\hat{P}_{\boldsymbol{x}})]H(\boldsymbol{W}|\hat{P}_{\boldsymbol{x}}) + H(\hat{P}_{\boldsymbol{X}}) \\ &\le \sum_{\mathcal{T}(\hat{P}_{\boldsymbol{x}})} P[\mathcal{T}(\hat{P}_{\boldsymbol{x}})]\min\{nR_{\mathrm{w}}, \log|\mathcal{T}(\hat{P}_{\boldsymbol{x}})|\} + (|\mathcal{X}|-1)\log(n+1) \\ &= \boldsymbol{E}\min\left\{nR_{\mathrm{w}}, \log|\mathcal{T}(\hat{P}_{\boldsymbol{X}})|\right\} + (|\mathcal{X}|-1)\log(n+1). \end{aligned} \tag{78}$$

Now, similarly as in Section VI, since $R_{\mathrm{w}} < H(X)$, the first term behaves like nR_{w} plus a term of the exponential order of $\exp\{-nE_{\mathrm{sec}}(R_{\mathrm{w}})\}$ (where $E_{\mathrm{sec}}(\cdot)$ was defined in Section VI), which is negligible compared to both nR_{w} and $(|\mathcal{X}|-1)\log(n+1)$. Thus, $I(\boldsymbol{X};\boldsymbol{W})/n$ converges to R_{w} at least as fast as $(\log n)/n$, and, as said, R_{w} in turn can chosen arbitrarily close (from above) to $H(X|Y)$, in accordance to [9, Proposition 2.4]. Of course, the cost of proximity to $H(X|Y)$ is in compromising the FR exponent, as was shown in Section IV.

VIII. Summary and Conclusion

In this paper, we studied the ensemble performance of biometric authentication systems, that are based on secret key generation. Referring to an ensemble of codes based on Slepian–Wolf binning, we have provided detailed, sharp analyses of the false–reject and false–accept probabilities, in terms of error exponents, for a wide class class of stochastic decoders that covers the optimal MAP decoder, as well as several additional decoders, as special cases. Converse bounds have been derived as well. Finally, we have outlined derivations of the secrecy leakage for the typical code in the ensemble, as well as on the privacy leakage. We believe that our results provide a more precise characterization of the trade-offs among the various figures of metric associated with biometric authentication systems that are based on secret key generation.

Appendix

Proof of eq. (36). The proof is similar to the proof of a similar argument in the context of channel coding [12, Appendix B]. First, observe that

$$Z_{\boldsymbol{x}}(\boldsymbol{y}) = \sum_{\boldsymbol{x}' \neq \boldsymbol{x}} \exp\{na(\hat{P}_{\boldsymbol{x}'\boldsymbol{y}})\} \cdot \mathcal{I}\{f(\boldsymbol{x}') = f(\boldsymbol{x})\} = \sum_{Q_{X|Y}} e^{na(Q_{XY})} N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}), f(\boldsymbol{x})). \quad \text{(A.1)}$$

Thus, considering the randomness of $\{f(\boldsymbol{x})\}$,

$$\begin{aligned} & \Pr\left\{Z_{\boldsymbol{x}}(\boldsymbol{y}) \leq \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}})\}\right\} \\ = \; & \Pr\left\{\sum_{Q_{X|Y}} N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}), f(\boldsymbol{x}))e^{na(Q_{XY})} \leq \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}})\}\right\} \\ \leq \; & \Pr\left\{\max_{Q_{X|Y}} N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}), f(\boldsymbol{x}))e^{na(Q_{XY})} \leq \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}})\}\right\} \\ = \; & \Pr \bigcap_{Q_{X|Y}} \left\{N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}), f(\boldsymbol{x}))e^{na(Q_{XY})} \leq \exp\{n\alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}})\}\right\} \\ = \; & \Pr \bigcap_{Q_{X|Y}} \left\{N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}), f(\boldsymbol{x})) \leq \exp\{n[\alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}}) - a(Q_{XY})]\}\right\}. \end{aligned} \quad \text{(A.2)}$$

Now, $N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}), f(\boldsymbol{x}))$ is a binomial random variable with $|\mathcal{T}(Q_{X|Y}|\boldsymbol{y}) \doteq e^{nH_Q(X|Y)}$ trials and success rate of $e^{-nR_{\mathrm{w}}}$. We now argue that by the very definition of $\alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}})$, there must exist some $Q^*_{X|Y}$ such that for $Q^*_{XY} = \hat{P}_{\boldsymbol{y}} \times Q^*_{X|Y}$, $H_{Q^*}(X|Y) \geq R+\epsilon$ and $H_{Q^*}(X|Y) - R - \epsilon \geq \alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}}) - a(\hat{P}_{\boldsymbol{y}} \times Q^*_{X|Y})$. Let then $Q^*_{X|Y}$ be such a conditional distribution. Then,

$$\begin{aligned} & \Pr \bigcap_{Q} \left\{N(\mathcal{T}(Q_{X|Y}|\boldsymbol{y}), f(\boldsymbol{x})) \leq \exp\{n[\alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}}) - a(\hat{P}_{\boldsymbol{y}} \times Q_{X|Y})]\}\right\} \\ \leq \; & \Pr\left\{N(\mathcal{T}(Q^*_{X|Y}|\boldsymbol{y}), f(\boldsymbol{x})) \leq \exp\{n[\alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}}) - a(\hat{P}_{\boldsymbol{y}} \times Q^*_{X|Y})]\}\right\}. \end{aligned} \quad \text{(A.3)}$$

Now, we know that $H_{Q^*}(X|Y) \geq R+\epsilon$ and $H_{Q^*}(X|Y) - R - \epsilon \geq \alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}}) - a(\hat{P}_{\boldsymbol{y}} \times Q^*_{X|Y})$. By the Chernoff bound the probability in question is upper bounded by

$$\exp\left\{-e^{nH_{Q^*}(X|Y)} D(e^{-\alpha n} \| e^{-\beta n})\right\}, \quad \text{(A.4)}$$

where $\alpha = H_{Q^*}(X|Y) + a(\hat{P}_{\boldsymbol{y}} \times Q^*_{XY}) - \alpha(R+\epsilon, \hat{P}_{\boldsymbol{y}})$ and $\beta = R$. Noting that $\alpha - \beta \geq \epsilon$, we can easily lower bound the binary divergence as follows (see [11, Section 6.3]):

$$D(e^{-\alpha n} \| e^{-\beta n}) \;\; \geq \;\; e^{-\beta n}\{1 - e^{-(\alpha-\beta)n}[1 + n(\alpha - \beta)]\}$$

$$\geq \quad e^{-nR}[1 - e^{-n\epsilon}(1 + n\epsilon)], \tag{A.5}$$

where in the last passage, we have used the decreasing monotonicity of the function $f(t) = (1+t)e^{-t}$ for $t \geq 0$. Thus,

$$\begin{aligned} & \Pr\left\{N(\mathcal{T}(Q^*_{X|Y}|\boldsymbol{y}), f(\boldsymbol{x})) \leq \exp\{n[\alpha(R, \hat{P}_{\boldsymbol{y}}) - a(\hat{P}_{\boldsymbol{y}} \times Q^*_{X|Y}) - \epsilon]\}\right\} \\ \leq \quad & \exp\left\{-e^{nH_{Q^*}(X|Y)} \cdot e^{-nR}[1 - e^{-n\epsilon}(1 + n\epsilon)]\right\} \\ \leq \quad & \exp\left\{-e^{n\epsilon}[1 - e^{-n\epsilon}(1 + n\epsilon)]\right\} \\ = \quad & \exp\left\{-e^{n\epsilon} + n\epsilon + 1\right\}. \end{aligned} \tag{A.6}$$

Finally, the factor of $|\mathcal{X} \times \mathcal{Y}|^n$ in eq. (36) comes from the union bound, taking into account all $|\mathcal{X} \times \mathcal{Y}|^n$ possible pairs $\{(\boldsymbol{x}, \boldsymbol{y})\}$. This completes the proof of eq. (36).

Acknowledgement

The author is grateful to the anonymous reviewers and to the Associate Editor, Prof. Matthieu Bloch, for many useful comments that certainly improved the quality of the paper.

References

[1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography – part I: secret sharing," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1121–1132, July 1993.

[2] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography – part II: CR capacity," *IEEE Trans. Inform. Theory*, vol. 44, pp. 225–240, January 1998.

[3] J. Chen, D.–k He, A. Jagmohan, and A. Lastras-Montaño, "On the reliability of variable–rate Slepian–Wolf coding," *Entropy*, vol. 19, 389, 2017. doi:10.3390/e19080389

[4] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Second Edition, Cambridge University Press, 2011.

[5] I. Csiszár, J. Körner, and K. Marton, "A new look at the error exponent of a discrete memoryless channel," *Proc. ISIT '77*, p. 107 (abstract), Cornell University, Itacha, New York, U.S.A., 1977.

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

[7] R. G. Gallager, "Source coding with side information and universal coding," LIDS-P-937, M.I.T., 1976.

[8] T. Ignatenko and F. M. J. Willems, "On the security of the XOR–method in biometric authentication systems," *Proc. Twenty Seventh Symposium on Information Theory in the Benelux*, pp. 197–204, Noordwijk, The Netherlands, June 8–9, 2006.

[9] T. Ignatenko and F. M. J. Willems, "Biometric security from an information–theoretical perspective," *Foundations and Trends in Communications and Information Theory*, vol. 7, nos. 2–3, 2010.

[10] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inform. Theory*, vol. 39, pp. 733–742, May 1993.

[11] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, nos. 1–2, pp. 1–212, 2009.

[12] N. Merhav, "The generalized stochastic likelihood decoder: random coding and expurgated bounds," *IEEE Trans. Inform. Theory*, vol. 63, no. 8, pp. 5039–5051, August 2017.

[13] N. Merhav, "Correction to 'The generalized stochastic likelihood decoder: random coding and expurgated bounds'," submitted for publication and available on–line at `https://arxiv.org/pdf/1707.03987.pdf`.

[14] N. Merhav, "Error exponents of typical random codes," submitted to *IEEE Trans. Inform. Theory*, August 2017.

[15] P. Ruján, "Finite temperature error–correcting codes," *Phys. Rev. Let.*, vol. 70, no. 19, pp. 2968–2971, May 1993.

[16] J. Scarlett, A. Martinéz and A. G. i Fábregas, "The likelihood decoder: error exponents and mismatch," *Proc. 2015 IEEE International Symposium on Information Theory (ISIT 2015)*, pp. 86–90, Hong Kong, June 2015.

[17] E. C. Song, P. Cuff and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Trans. Inform. Theory*, vol. 62, no. 4, pp. 1836–1849, April 2016.

[18] N. Weinberger and N. Merhav, "Optimum tradeoffs between the error exponent and the excess–rate exponent of variable–rate Slepian–Wolf coding," *IEEE Trans. Inform. Theory*, vol. 61, no. 4, pp. 2165–2190, April 2015.

[19] M. H. Yassaee, M. R. Aref and A. Gohari, "A technique for deriving one–shot achievability results in network information theory," *Proc. 2013 IEEE International Symposium on Information Theory (ISIT 2013)*, pp. 1287–1291, July 2013. Also, available on–line at `http://arxiv.org/abs/1303.0696`.