Achievable Second-Order Coding Rates for the Wiretap Channel

Vincent Y. F. Tan

Institute for Infocomm Research (I^{2}R), A*STAR, Singapore (Email: tanyfv@i2r.a-star.edu.sg)
Electrical and Computer Engineering, National University of Singapore (Email: eletyfv@nus.edu.sg)

***Abstract*—We derive lower bounds to the second-order coding rates for the wiretap channel. The decoding error probability and the information leakage measured in terms of the variational distance secrecy metric are fixed at some constants ϵ_{r} and ϵ_{s} respectively. We leverage on the connection between wiretap channel coding and channel resolvability to derive tighter secrecy bounds than those available in the literature. We then use central limit theorem-style analysis to evaluate these bounds for the discrete memoryless wiretap channel with cost constraints and the Gaussian wiretap channel.**

***Index Terms*—Second-order coding rates, Dispersion analysis, Wiretap channel, Information-theoretic secrecy**

I. Introduction

The wiretap channel, introduced by Wyner [1], is the most fundamental model in the study of information-theoretic secrecy. The model is essentially a broadcast channel $W(y,z|x)$ with one transmitter X (known as Alice) and two receivers; the legitimate one Y (known as Bob) and the eavesdropper Z (known as Eve). In Wyner's original setup, the observation at Eve is a degraded version Bob's observation. That is, $X-Y-Z$ form a Markov chain in that order. The goal was to characterize the set of achievable (secrecy) rates of transmission of a uniformly distributed message $M \in [1:\exp(nR)]$ to Bob, while at the same time ensuring that Eve can only glean a infinitesimal amount of information about M. A number R is said to be *(weakly)-achievable* if

$$\lim_{n\to\infty} \mathbb{P}(\hat{M} \neq M) = 0, \quad \lim_{n\to\infty} \frac{1}{n} I(M;Z^n) = 0. \tag{1}$$

The random variable $\hat{M} = \hat{M}(Y^n)$ is Bob's estimate of the transmitted message M. In addition, the second criterion means that to every $\varepsilon > 0$, Eve can at most decipher ε nats of the message M for all n sufficiently large. The secrecy capacity of a discrete memoryless broadcast channel (without cost constraints) was found to be

$$C_{\mathrm{S}} = \max_{p_X \in \mathscr{P}(\mathcal{X})} I(X;Y) - I(X;Z), \tag{2}$$

where the maximization is over all probability mass functions supported on the input alphabet $\mathcal{X}$. Wyner's model was subsequently generalized by Csiszár and Körner [2] who dispensed with the degradedness assumption and, in addition, they required Eve to decode a common message.

The secrecy criterion in (1) is only one of many that measures approximate independence [3]. Indeed, if the normalized mutual information tends to zero, this is termed as *weak secrecy*. If instead the unnormalized mutual information $I(M;Z^n)$ tends to zero, this is termed *strong secrecy*. In this paper, for analytical tractability, we consider a criterion, termed the *variational distance* criterion $\mathbb{V}(p_{M,Z^n}, p_M p_{Z^n})$. This criterion is stronger than weak secrecy but weaker than strong secrecy [3]. It can be shown using [4, Lemma 1] that if $\mathbb{V}(p_{M,Z^n}, p_M p_{Z^n})$ decays as $o(1/n)$, then $I(M;Z^n)$ also tends to zero.

In this paper, we are interested in the study of *second-order coding rates* [5] for the discrete [1], [2] and Gaussian wiretap channels [6]. The setup is as follows: We allow the decoding error probability and the variational distance $\mathbb{V}(p_{M,Z^n}, p_M p_{Z^n})$ to be non-zero asymptotically, i.e.,

$$\overline{\lim_{n\to\infty}} \mathbb{P}(\hat{M} \neq M) \leq \epsilon_{\mathrm{r}}, \quad \overline{\lim_{n\to\infty}} \mathbb{V}(p_{M,Z^n}, p_M p_{Z^n}) \leq \epsilon_{\mathrm{s}}, \tag{3}$$

holds for some constants $\epsilon_{\mathrm{r}}, \epsilon_{\mathrm{s}} > 0$. We then ask what the maximal secrecy rate over all possible wiretap codes is. This rate R is, in general, a function of the channel W, the constants ϵ_{r} and ϵ_{s} and the blocklength n. We show that discrete and Gaussian wiretap channels, the maximal rate can be approximately lower bounded by $C_{\mathrm{S}} + R_2/\sqrt{n}$. Notice that the first-order term in the maximal rate is the secrecy capacity C_{S} in (2). The second-order term R_2 is of central interest in this paper and is termed the *second-order coding rate*. Together, C_{S} and R_2 give a much finer characterization of the maximal secrecy rate under the reliability and secrecy constraints in (3). Inspired by the connection between information-theoretic secrecy and channel resolvability [3], [7], we develop lower bounds for the second-order coding rate R_2 for the discrete memoryless and Gaussian wiretap channels.

A. Summary of Main Results

There are three main contributions in this paper. Firstly, by modifying the proof of the secrecy capacity for general wiretap channels in [3], we improve on the achievable bounds on the error probability and the leakage (in terms of the variational distance) for general wiretap channels. Secondly, we use this result to derive achievable second-order coding rates for cost-constrained discrete memoryless channels. Finally, we extend the second-order coding rate result to Gaussian wiretap channels by carefully using a discretization procedure. For both discrete and Gaussian wiretap channels, we show that

$$R_2 \geq \sqrt{V(X;Y)}\Phi^{-1}\left(\frac{\epsilon_{\mathrm{r}}}{2}\right) + \sqrt{V(X;Z)}\Phi^{-1}\left(\frac{\epsilon_{\mathrm{s}}^2}{400}\right), \tag{4}$$

where $V(X;Y)$ and $V(X;Z)$ are the dispersions [8] of Bob's and Eve's channels (evaluated at the optimal p_X in (2)) respectively. In addition, Φ^{-1} is the inverse of the Gaussian cumulative distribution function. Intuitively, the first and second terms result from applications of the central limit theorem to the reliability and secrecy requirements respectively.

B. Related Work

The connection between wiretap coding, channel resolvability and identification capacity was first made by Hayashi [7]. These ideas were then refined by Bloch and Laneman [3] who showed, among other results, that the capacity-based codes do not achieve strong secrecy. Indeed, resolvability-based codes are necessary to achieve strong secrecy. Roughly speaking, this means that for strong secrecy, the number of random bits per codeword that needs to be generated to confuse Eve has to be larger than Eve's channel capacity. The study of second-order source and channel coding rates was initiated by Strassen [9] and re-popularized in recent years by Kontoyannis [10], Hayashi [5], [11] and Polyanskiy et al. [8] among others. In this work, we combine the analysis of wiretap coding and channel resolvability to derive second-order coding rates with the variational distance as the secrecy metric.

II. Problem Setup

In this section, we state the definitions and the problem precisely. Before doing so, we start with the notational conventions that will be used throughout the paper.

A. Notation

Random variables and the values they take on will be denoted by upper case (e.g., X) and lower case (e.g., x) respectively. Types (empirical distributions) will be denoted by upper case (e.g., Q) and distributions by lower case (e.g., q). The set of all distributions on a finite set $\mathcal{X}$ is denoted as $\mathscr{P}(\mathcal{X})$ and the set of n-types with alphabet $\mathcal{X}$ is denoted as $\mathscr{P}_n(\mathcal{X})$. For a type $Q \in \mathscr{P}_n(\mathcal{X})$, the type class is denoted as $\mathcal{T}_Q$. Given a probability density function (pdf) or probability mass function (pmf) q and a stochastic matrix (conditional distribution) $W : \mathcal{X} \to \mathcal{Y}$, we use the notation $(qW)(y) := \sum_x q(x)W(y|x)$ to denote the $\mathcal{Y}$-marginal of the joint distribution $q(x)W(y|x)$. For information-theoretic quantities, we adopt the notation of [12]. In particular, the mutual information of $(X,Y) \sim p_X p_{Y|X}$ is denoted as $I(X;Y)$. All logarithms are to the base e. The Gaussian pdf with mean μ and variance σ^2 is denoted as $\mathcal{N}(x;\mu,\sigma^2)$ and the cumulative distribution function as $\Phi(t) = \int_{-\infty}^{t} \mathcal{N}(x;0,1)\,dx$.

B. Definitions

A *wiretap channel* is a tuple $(\mathcal{X}, W(y,z|x), \mathcal{Y}, \mathcal{Z})$ where $\mathcal{X}$ is the input alphabet and $\mathcal{Y}$ and $\mathcal{Z}$ are the alphabets corresponding to the legitimate receiver (Bob) and the eavesdropper (Eve) respectively. In addition, $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ is a stochastic matrix, i.e., $\sum_{y,z} W(y,z|x) = 1$ for all $x \in \mathcal{X}$. We use the notations $W_{\mathrm{b}}(y|x) = \sum_z W(y,z|x)$ and $W_{\mathrm{e}}(z|x) = \sum_y W(y,z|x)$ to denote Bob and Eve's marginals respectively. Even though Theorem 1 below applies to the general wiretap channel (non-ergodic, non-stationary), for simplicity, henceforth, we assume that the channels are memoryless (but not necessarily discrete) in the sense that $W^n(y^n, z^n|x^n) = \prod_{k=1}^n W(y_k, z_k|x_k)$.

Since we are only concerned with achievability results, we assume that the wiretap channel is degraded, i.e., $X - Y - Z$ forms a Markov chain. Our results can be easily strengthened by channel prefixing [2], [3]. In addition, in order to model actual engineering devices accurately, cost constraints must be imposed on the channel input sequence x^n. To formalize this, let $c : \mathcal{X} \to [0,\infty)$ be a cost function and with a slight abuse of notation, we also use c to denote the average cost of a codeword, i.e., $c(x^n) := \frac{1}{n}\sum_{k=1}^n c(x_k)$.

Definition 1. *An $(\exp(nR), n, P)$-wiretap code $\mathcal{C}_n$ consists of*

- *A private message set $\mathcal{M} := [1 : \exp(nR)]$;*
- *An auxiliary message set $\mathcal{M}' := [1 : \exp(nR')]$, which is used to randomize the transmission of the private message;*
- *A stochastic encoder $p_{X^n|M}(x^n|m)$ for every $m \in \mathcal{M}$ such that the codeword satisfies the cost constraint with probability one, i.e.,*

$$\mathbb{P}\left[c(X^n(M,M')) \le P\right] = 1; \tag{5}$$

- *A decoder $\varphi_n : \mathcal{Y}^n \to \mathcal{M} \times \mathcal{M}'$.*

We allow R and R' to vary with n in general and we use the notation $|\mathcal{M}(\mathcal{C}_n)|$ to denote the number of private messages (those containing information that Alice wants to transmit) of the wiretap code $\mathcal{C}_n$. This is simply the quantity $|\mathcal{M}| = \exp(nR)$ in Definition 1. Following [3], the random variable representing a chosen codeword when using the code $\mathcal{C}_n$ is denoted as $\bar{X}^n$ while the channel outputs induced by the input codeword $\bar{X}^n$ are $\bar{Y}^n$ and $\bar{Z}^n$. The probability of error with respect to a given code $\mathcal{C}_n$ is defined as

$$\mathbb{P}(\mathcal{C}_n) := \mathbb{P}\big[(\hat{M},\hat{M}') \neq (M,M') \,|\, \mathcal{C}_n\big]. \tag{6}$$

where M and M' are independent random variables uniformly distributed in $\mathcal{M}$ and $\mathcal{M}'$ respectively and $(\hat{M},\hat{M}') = \varphi_n(\bar{Y}^n)$ are the estimated messages. Note that we impose that the legitimate receiver also decodes the auxiliary message $\mathcal{M}'$. This is a more stringent requirement that is usually not present in usual wiretap coding [1].

Definition 2. *Let $\mathbb{P}$ and $\mathbb{Q}$ be two measures on a measurable space $(\mathcal{X}, \mathfrak{F})$. The variational distance between $\mathbb{P}$ and $\mathbb{Q}$ is*

$$\mathbb{V}(\mathbb{P},\mathbb{Q}) := \sup_{\mathcal{A}\in\mathfrak{F}} |\mathbb{P}(\mathcal{A}) - \mathbb{Q}(\mathcal{A})|. \tag{7}$$

In the case where $\mathcal{X}$ is countable and $p, q \in \mathscr{P}(\mathcal{X})$ are the pmfs corresponding to measures $\mathbb{P}, \mathbb{Q}$ respectively, then

$$\mathbb{V}(p,q) := \mathbb{V}(\mathbb{P},\mathbb{Q}) = \frac{1}{2}\sum_{x\in\mathcal{X}} |p(x) - q(x)|. \tag{8}$$

We consider the *leakage* measured in terms of the *variational distance secrecy metric* defined for a code $\mathcal{C}_n$ as

$$\mathbb{S}(\mathcal{C}_n) := \mathbb{V}(p_{M,\bar{Z}^n}, p_M p_{\bar{Z}^n}). \tag{9}$$

Note that $\mathbb{S}(\mathcal{C}_n)$ is equal to zero if and only if M and $\bar{Z}^n$ are statistically independent. In general, we want $\mathbb{S}(\mathcal{C}_n)$ to be small so that the amount of information that Eve can glean is small. The following is the central definition in this paper.

Definition 3. *Fix constants $R_1, \epsilon_{\mathrm{r}}, \epsilon_{\mathrm{s}} > 0$. The* second-order coding rate centered at first-order coding rate R_1 *of a wiretap channel $(\mathcal{X}, W(y,z|x), \mathcal{Y}, \mathcal{Z})$ is defined as*

$$R_2(R_1,\epsilon_{\mathrm{r}},\epsilon_{\mathrm{s}}|W) := \sup_{\{\mathcal{C}_n\}_{n\geq 1}} \left\{ \varliminf_{n\to\infty} \frac{1}{\sqrt{n}} \left(\log|\mathcal{M}(\mathcal{C}_n)| - nR_1\right) : \varlimsup_{n\to\infty} \mathbb{P}(\mathcal{C}_n) \leq \epsilon_{\mathrm{r}},\ \varlimsup_{n\to\infty} \mathbb{S}(\mathcal{C}_n) \leq \epsilon_{\mathrm{s}} \right\} \tag{10}$$

where the supremum is over all rates $R > 0$ and all sequences of $(\exp(nR), n, P)$-wiretap codes $\{\mathcal{C}_n\}_{n\geq 1}$.

As with the secrecy capacity, we seek to *maximize* R_2.

C. Secrecy Capacity Results

In Section III, we lower bound $R_2(R_1, \epsilon_{\mathrm{r}}, \epsilon_{\mathrm{s}}|W)$ when R_1 is the secrecy capacity, given for a channel without cost constraints in (2). Note that for memoryless channels, the secrecy capacity is invariant to the choice of secrecy metric, i.e., asymptotic independence can be measured using $I(M;Z^n), \mathbb{V}(p_{M,\bar{Z}^n}, p_M p_{\bar{Z}^n})$ or $\frac{1}{n}I(M;Z^n)$ [3]. It can be shown that when there is a cost constraint on the codewords, the secrecy capacity in (2) can be slightly generalized to

$$C_{\mathrm{S}} = \max_{p_X \in \mathscr{P}(\mathcal{X}):\mathbb{E}_{p_X}[c(X)]\leq P} I(X;Y) - I(X;Z). \tag{11}$$

Recall that we assume that the memoryless channel is degraded in favor of Bob. In the Gaussian case, the outputs of the wiretap channel (Y,Z) are related to the input X as

$$Y = X + N_{\mathrm{b}}, \qquad Z = X + N_{\mathrm{e}}, \tag{12}$$

where the noises $N_{\mathrm{b}} \sim \mathcal{N}(0,\sigma_{\mathrm{b}}^2)$ and $N_{\mathrm{e}} \sim \mathcal{N}(0,\sigma_{\mathrm{e}}^2)$. The average transmitted power is constrained to be no larger than some $P > 0$. Hence, $c(x) = x^2$ and the codewords must satisfy $\sum_{k=1}^n X_k^2(M,M') \leq nP$ with probability one. Degradedness in the Gaussian case means that $\sigma_{\mathrm{b}} < \sigma_{\mathrm{e}}$. In this case, it is known that [6] the wiretap channel capacity is

$$C_{\mathrm{S}} = \frac{1}{2}\log\left(1+\frac{P}{\sigma_{\mathrm{b}}^2}\right) - \frac{1}{2}\log\left(1+\frac{P}{\sigma_{\mathrm{e}}^2}\right). \tag{13}$$

That is, the optimal p_X in (11) is a zero-mean Gaussian with variance P. If the degraded condition is not satisfied, the secrecy capacity is zero. Note that the first and second terms in (13) are Bob's and Eve's channel capacities respectively. We will also use the notations $\mathsf{snr}_{\mathrm{b}} := P/\sigma_{\mathrm{b}}^2$ and $\mathsf{snr}_{\mathrm{e}} := P/\sigma_{\mathrm{e}}^2$ to denote Bob and Eve's signal-to-noise ratio (SNR) respectively.

III. Main Results

In this section, we state the three main results in this paper. The proof sketches are deferred to Section IV. We start with a general achievability theorem strengthening Theorem 3 in [3].

Theorem 1. *Assume that there are no constraints on the codewords, i.e., $P = \infty$. For every $n \geq 1$, every input distribution $p_{X^n} \in \mathscr{P}(\mathcal{X}^n)$ and every $\gamma, \rho > 0$, there exists a wiretap code $\mathcal{C}_n^*$ whose probability of error and leakage simultaneously satisfy*

$$\mathbb{P}(\mathcal{C}_n^*) \leq 2\left[\mathbb{P}\left(\frac{W_{\mathrm{b}}^n(Y^n|X^n)}{p_{Y^n}(Y^n)} \leq |\mathcal{M}||\mathcal{M}'|e^{n\gamma}\right) + e^{-n\gamma}\right], \tag{14}$$

$$\mathbb{S}(\mathcal{C}_n^*) \leq 4\left[2\rho + \left(1+\frac{1}{\rho}\right)\mathbb{P}\left(\frac{W_{\mathrm{e}}^n(Z^n|X^n)}{p_{Z^n}(Z^n)} \geq |\mathcal{M}'|\rho\right) + \frac{1}{\rho}e^{-n\gamma} + \frac{1}{\rho}\mathbb{P}\left(\frac{W_{\mathrm{e}}^n(Z^n|X^n)}{p_{Z^n}(Z^n)} \geq |\mathcal{M}'|\rho e^{-n\gamma}\right)\right]. \tag{15}$$

In (14) *and* (15), *$p_{Y^n} = (p_{X^n}W_{\mathrm{b}}^n)$ and $p_{Z^n} = (p_{X^n}W_{\mathrm{e}}^n)$ are the output distributions corresponding to p_{X^n}.*

Note that the above result is applicable to *general sequences of wiretap channels* $\{W^n : \mathcal{X}^n \to \mathcal{Y}^n \times \mathcal{Z}^n\}_{n\geq 1}$, i.e., they are not necessarily discrete, Gaussian, ergodic or stationary. With a slight modification, Theorem 1 also applies to channels with cost constraints. The proof of this result follows along the lines of Theorem 3 in [3] where the connection between the variational distance criterion in (9) and channel resolvability [12, Chapter 6] was established. However, we note that the constant ρ in (15) has been distributed among the terms in a different way and this leads to better second-order coding rates with judicious choices of the free parameters ρ and γ.

We now specialize Theorem 1 to a discrete memoryless wiretap channel $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ with cost constraints. We provide a lower bound to $R_2(C_{\mathrm{S}}, \epsilon_{\mathrm{r}}, \epsilon_{\mathrm{s}}|W)$. For a given $p_X \in \mathscr{P}(\mathcal{X})$ satisfying $\mathbb{E}_{p_X}[c(X)] \leq P$, define

$$V(X;Y) := \sum_x p_X(x) \sum_y W_{\mathrm{b}}(y|x) \times \left[\log\frac{W_{\mathrm{b}}(y|x)}{(p_X W_{\mathrm{b}})(y)} - D(W_{\mathrm{b}}(\cdot|x)||(p_X W_{\mathrm{b}})(\cdot))\right]^2 \tag{16}$$

to be the *dispersion* [8] of the channel $W_{\mathrm{b}}(y|x)$ and similarly $V(X;Z)$ denotes the dispersion of $W_{\mathrm{e}}(z|x)$ given p_X.

Theorem 2. *For a discrete memoryless wiretap channel, the second-order coding rate centered at C_{S}, defined in* (11)*, satisfies*

$$R_2(C_{\mathrm{S}}, \epsilon_{\mathrm{r}}, \epsilon_{\mathrm{s}}|W) \geq \sqrt{V(X;Y)}\Phi^{-1}\left(\frac{\epsilon_{\mathrm{r}}}{2}\right) + \sqrt{V(X;Z)}\Phi^{-1}\left(\frac{\epsilon_{\mathrm{s}}^2}{400}\right), \tag{17}$$

where $X \sim p_X$ and p_X is an optimizing distribution in (11) *that maximizes the right-hand-side (RHS) of* (17)*.*

A couple of comments are in order: The first term on the RHS of (17) represents a Gaussian approximation to the error probability bound in (14). We generate a random (constant composition) code with the appropriate number of private and auxiliary messages such that the bound in (14) is roughly $\epsilon_{\mathrm{r}}/2$. The second term represents the Gaussian approximation for the leakage measured in terms of the variational distance criterion. It roughly represents the second-order coding rate for the eavesdropper's channel. We can verify that, with the optimum choice of ρ, this ensures that the leakage bound in (15) is

roughly $\epsilon_{\mathrm{s}}/2$. If we had simply used the bound in [3, Theorem 3] or [12, Theorem 6.3.1] without any modifications, the argument of Φ^{-1} in the second term would be $\Theta(\epsilon_{\mathrm{s}}^3)$, which is strictly worse than what we have presented in Theorem 2. Since averaged over the random code the probability of error and leakage are no larger than $\epsilon_{\mathrm{r}}/2$ and $\epsilon_{\mathrm{s}}/2$ respectively, we can conclude via the union bound and Markov's inequality that the constraints in (10) are satisfied.

Define the *Gaussian dispersion* at SNR snr to be

$$V(\mathsf{snr}) := \frac{\mathsf{snr}(\mathsf{snr}+2)}{2(1+\mathsf{snr})^2}. \tag{18}$$

The analogue of Theorem 2 for the Gaussian wiretap channel is the following:

Theorem 3. *For a Gaussian wiretap channel, the second-order coding rate centered at C_{S}, defined in* (13)*, satisfies*

$$\begin{aligned} &R_2(C_{\mathrm{S}}, \epsilon_{\mathrm{r}}, \epsilon_{\mathrm{s}}|W) \\ &\geq \sqrt{V(\mathsf{snr}_{\mathrm{b}})}\Phi^{-1}\left(\frac{\epsilon_{\mathrm{r}}}{2}\right) + \sqrt{V(\mathsf{snr}_{\mathrm{e}})}\Phi^{-1}\left(\frac{\epsilon_{\mathrm{s}}^2}{400}\right). \end{aligned} \tag{19}$$

We can prove this result in (at least) two different ways: First, we can adopt Shannon's approach [13] and generate codewords uniformly at random from the n-sphere with radius $\sqrt{nP}$. Then, use a minimum-distance decoding approach or the $\kappa\beta$-bound approach [8] to decode the messages $(m, m') \in \mathcal{M} \times \mathcal{M}'$. The second approach is to apply a discretization procedure to Theorem 2. This was used in proof of Theorem 5 in [5] and this can be seen to be equivalent to [13] in the limit of large blocklengths since we generate codewords uniformly at random from a single type class to prove Theorem 2. However, some continuity and rate of convergence arguments in [5] need to be made more precise. We do so by using a result by Wu and Verdú [14] which states that the gap between $\frac{1}{2}\log(1+\mathsf{snr})$ and the maximal mutual information achieved by inputs of variance at most P taking $l \in \mathbb{N}$ values decays at least exponentially fast in l. Both proof techniques lead to the first term in (19). The second term follows *mutatis mutandis* from the proof of Theorem 2. We outline the latter approach (discretization procedure) in Section IV-C.

IV. Proof Sketches

In this section, we provide proof sketches of Theorems 1–3.

A. Proof of Theorem 1

For every $(m, m') \in \mathcal{M} \times \mathcal{M}'$, randomly and independently generate $x^n(m, m')$ from p_{X^n}. To transmit $m \in \mathcal{M}$, we randomly and uniformly select an index $m \in \mathcal{M}'$ and the sequence $x^n(m, m')$ is sent as the input to the wiretap channel W^n. We bound the probability of decoding error and the leakage averaged over this random code construction. We will focus mainly on the secrecy bound in (15) as the reliability bound in (14) follows in a straightforward fashion from Feinstein's lemma [12, Lemma 3.4.1]. This lemma says that averaged over the random code and with $X^n \sim p_{X^n}$,

$$\mathbb{P}(\mathcal{C}_n) \leq \mathbb{P}\left(\frac{W_{\mathrm{b}}^n(Y^n|X^n)}{p_{Y^n}(Y^n)} \leq |\mathcal{M}||\mathcal{M}'|e^{n\gamma}\right) + e^{-n\gamma}. \tag{20}$$

Now we overbound $\mathbb{S}(\mathcal{C}_n)$ averaged over the random code. By using Bloch and Laneman's [3] technique and exploiting symmetry, we get

$$\mathbb{S}(\mathcal{C}_n) \leq 2\mathbb{E}_{\mathcal{C}_n}\left[\mathbb{V}(p_{\bar{Z}^n|M=1}, p_{Z^n})\right]. \tag{21}$$

Note that we assumed $M = 1$ was chosen. We now overbound the variational distance using [12, Lemma 6.3.1] yielding

$$\mathbb{S}(\mathcal{C}_n) \leq 2\tau + 2A_n, \tag{22}$$

where $\tau > 0$ is an arbitrary constant and

$$A_n := \mathbb{E}_{\mathcal{C}_n}\left[\mathbb{P}_{\bar{Z}^n|M=1}\left(\log\frac{p_{\bar{Z}^n|M=1}(\bar{Z}^n)}{p_{Z^n}(\bar{Z}^n)} > \tau\right)\right]. \tag{23}$$

Define $\rho := \frac{1}{2}(e^{\tau} - 1)$. We will choose $\rho, \tau \in \Theta(\epsilon_{\mathrm{s}})$ in the sequel. By going through the same steps as in [3], we can overbound A_n as

$$\begin{aligned} A_n &\leq \mathbb{P}\left[\frac{W_{\mathrm{e}}^n(Z^n|X^n)}{p_{Z^n}(Z^n)} > |\mathcal{M}'|\rho\right] \\ &+ \mathbb{P}\left[\frac{1}{|\mathcal{M}'|}\sum_{m'\in\mathcal{M}'\setminus\{1\}}\frac{W_{\mathrm{e}}^n(Z^n|X^n(1,m'))}{p_{Z^n}(Z^n)} > 1+\rho\right] \end{aligned} \tag{24}$$

where $X^n(1, m')$ is the codeword with index $M = 1$ and $M' = m' \in \mathcal{M}'$. Let us denote the first and second terms in (24) as K_n and L_n respectively. Note that K_n is one of the terms in (15). Furthermore, L_n can be bounded above as

$$L_n \leq \mathbb{E}\left\{\mathbb{P}\left[\frac{1}{|\mathcal{M}'|}\sum_{m'\in\mathcal{M}'}\frac{W_{\mathrm{e}}^n(Z^n|X^n(1,m'))}{p_{Z^n}(Z^n)} > 1+\rho\right]\right\}. \tag{25}$$

The outer expectation is over Z^n and the inner probability is over the codewords bin 1 for a given z^n sequence, i.e., $\{X^n(1, m')\}_{m'\in\mathcal{M}'}$ given $Z^n = z^n$. We bound each term in the expectation, written as $L_n(z^n)$, separately. At this point that we depart from using the exact proof ideas in [3]. Define the following random variables for each $m' \in \mathcal{M}'$:

$$D_{m'}(z^n) := \frac{W_{\mathrm{e}}^n(z^n|X^n(1,m'))}{p_{z^n}(z^n)}, \tag{26}$$

$$E_{m'}(z^n) := D_{m'}^n(z^n)\mathbf{1}\left\{D_{m'}^n(z^n) \leq |\mathcal{M}'|\rho\right\}, \tag{27}$$

$$F(z^n) := \frac{1}{|\mathcal{M}'|}\sum_{m'\in\mathcal{M}'} D_{m'}^n(z^n) \tag{28}$$

$$G(z^n) := \frac{1}{|\mathcal{M}'|}\sum_{m'\in\mathcal{M}'} E_{m'}^n(z^n) \tag{29}$$

Notice the difference in the definition of $E_{m'}(z^n)$ in (27) relative to that in [3]. There is an additional ρ in the indicator function. This results in the argument of the second Φ^{-1} term in (17) being $\Theta(\epsilon_{\mathrm{s}}^2)$ instead of $\Theta(\epsilon_{\mathrm{s}}^3)$ (with the choice of $\rho = \Theta(\epsilon_{\mathrm{s}})$ in (15)), improving the second-order coding rate. Now, using the law of total probability

$$L_n(z^n) \leq \mathbb{P}\left[G(z^n) > 1+\rho\right] + \mathbb{P}\left[F(z^n) \neq G(z^n)\right]. \tag{30}$$

Let the first and second terms in (30) be $B_n(z^n)$ and $C_n(z^n)$ respectively. We first bound $C_n(z^n)$ as follows:

$$C_n(z^n) \leq |\mathcal{M}'|\,\mathbb{P}\left[D_1^n(z^n) > |\mathcal{M}'|\rho\right], \tag{31}$$

where (31) follows from the definition of D_1 and E_1 and the indicator function. We evaluate the expectation (over Z^n) of the probability in (31). We have

$$\mathbb{E}(\mathbb{P}\left[D_1^n(Z^n) > |\mathcal{M}'|\rho\right]) \leq \frac{1}{|\mathcal{M}'|\rho}\mathbb{P}\left[\frac{W_{\mathrm{e}}^n(Z^n|X^n)}{p_{Z^n}(Z^n)} > |\mathcal{M}'|\rho\right]. \tag{32}$$

Uniting (31) and (32) yields

$$\mathbb{E}_{Z^n}(C_n(Z^n)) \leq \frac{1}{\rho}\,\mathbb{P}\left[\frac{W_{\mathrm{e}}^n(Z^n|X^n)}{p_{Z^n}(Z^n)} > |\mathcal{M}'|\rho\right]. \tag{33}$$

We omit the details of the bound on $B_n(z^n)$ which can be done via an application of Chebyshev's inequality and a judicious split of the resulting terms. We obtain

$$\mathbb{E}_{Z^n}[B_n(Z^n)] \leq \frac{1}{\rho}\left(e^{-n\gamma} + \mathbb{P}\left[\frac{W_{\mathrm{e}}^n(Z^n|X^n)}{p_{Z^n}(Z^n)} > |\mathcal{M}'|\rho e^{-n\gamma}\right]\right). \tag{34}$$

Combining (22), (24), (33) and (34) gives the bound on the leakage in (15). The extra factor of 2 in $\mathbb{P}(\mathcal{C}_n^*)$ and $\mathbb{S}(\mathcal{C}_n^*)$ comes from Markov's inequality and the union bound. □

B. Proof of Theorem 2

Choose the cardinalities of $\mathcal{M}$ and $\mathcal{M}'$ such that they satisfy

$$R + R' = I(X;Y) + \sqrt{\frac{V(X;Y)}{n}}\Phi\left(\frac{\epsilon_{\mathrm{r}}}{2}\right) - \nu_n, \tag{35}$$

$$R' = I(X;Z) - \sqrt{\frac{V(X;Z)}{n}}\Phi\left(\frac{\epsilon_{\mathrm{r}}^2}{400}\right) + \nu_n, \tag{36}$$

where $\nu_n \in O(\frac{\log n}{n})$. The rate of the code $\frac{1}{n}\log|\mathcal{M}(\mathcal{C}_n)|$ is thus $R = C_{\mathrm{S}} + \tilde{R}_2(C_{\mathrm{S}}, \epsilon_{\mathrm{r}}, \epsilon_{\mathrm{s}}|W)/\sqrt{n} - 2\nu_n$, where $\tilde{R}_2(C_{\mathrm{S}}, \epsilon_{\mathrm{r}}, \epsilon_{\mathrm{s}}|W)$ is the lower bound in (17).

Fix a type $Q \in \mathscr{P}_n(\mathcal{X})$ for which $\mathbb{E}_Q[c(X)] \leq \Gamma$ and $|Q(x) - p_X(x)| \leq 1/n$ for all $x \in \mathcal{X}$. Then, for every $m \in \mathcal{M}$ and every $m' \in \mathcal{M}'$, independently and uniformly sample a vector $x^n(m, m') \in \mathcal{T}_Q$ (a constant composition code). Clearly, the cost constraint (5) is satisfied. Using basic properties of types as in [15], the choice of the sum rate in (35) and the Berry-Essèen theorem, it is easy to show that

$$\overline{\lim_{n\to\infty}}\ \mathbb{P}(\mathcal{C}_n) \leq \frac{\epsilon_{\mathrm{r}}}{2}. \tag{37}$$

We now proceed to the secrecy analysis. We will upper the second probability in (15) which we denote as ψ_n:

$$\psi_n := \mathbb{P}\left(\frac{1}{n}\log\frac{W_{\mathrm{e}}^n(Z^n|X^n)}{p_{Z^n}(Z^n)} \geq R' + \frac{1}{n}\log\rho - \gamma\right). \tag{38}$$

Note here that X^n is drawn uniformly at random from the type class $\mathcal{T}_Q$ and p_{Z^n} is the output distribution corresponding to this input distribution. Consider the upper bound

$$\psi_n \leq \mathbb{P}\left(\frac{1}{n}\log\frac{W_{\mathrm{e}}^n(Z^n|X^n)}{(QW_{\mathrm{e}})^n(Z^n)} \geq R' + \frac{1}{n}\log\rho - \gamma - \xi\right) + \mathbb{P}\left(\frac{1}{n}\log\frac{(QW_{\mathrm{e}})^n(Z^n)}{p_{Z^n}(Z^n)} \geq \xi\right), \tag{39}$$

which comes from the elementary fact that $\mathbb{P}(A + B \geq c) \leq \mathbb{P}(A \geq c - \xi) + \mathbb{P}(B \geq \xi)$ for any $\xi \geq 0$. The second term is no larger than $e^{-n\xi}$ [12, Lemma 3.2.1]. Choose $\gamma, \xi \in \Theta(\frac{\log n}{n})$ and $\rho := \epsilon_{\mathrm{s}}/20$. This choice of ρ minimizes the RHS of (15). Then use the analysis in the reliability part to bound the first term in (39). After accounting for the choice of $\frac{1}{n}\log|\mathcal{M}'|$ in (36) and applying the Berry-Essèen theorem, we may assert that $\psi_n \leq \epsilon_{\mathrm{s}}^2/400 + O(\frac{\log n}{\sqrt{n}})$. Substituting this bound into (15) and using the fact that $\rho = \epsilon_{\mathrm{s}}/20$ yields

$$\overline{\lim_{n\to\infty}}\ \mathbb{S}(\mathcal{C}_n) \leq \frac{\epsilon_{\mathrm{s}}}{2}. \tag{40}$$

This completes the proof (cf. the discussion after (34)). □

C. Proof of Theorem 3

At blocklength n, construct a discrete random variable X_l^{Q} supported on $l := \lfloor n^{1/4} \rfloor$ points based on the Gauss quadrature, i.e., the locations of the atoms are the roots of Hermite polynomials. This results in the gap between capacity $\frac{1}{2}\log(1 + \mathsf{snr})$ and $I(X_l^{\mathrm{Q}}; Y)$ being upper bounded by $4(1 + \mathsf{snr})(\frac{\mathsf{snr}}{1+\mathsf{snr}})^{2l}$ [14, Theorem 8], which is negligible in the proof of Theorem 2. We also need to prove that X_l^{Q} converges to $X \sim \mathcal{N}(0, P)$ weakly and this can be done via an application of the Stone-Weierstrass theorem [16, Theorem 7.26] for approximating continuous functions with polynomials on compact intervals. Thus, the corresponding output densities converge pointwise on $\mathbb{R}$. This proves that $V(X_l^{\mathrm{Q}}; Y) \to V(\mathsf{snr})$ where $V(\mathsf{snr})$ is given by (18). □

References

[1] A. D. Wyner, "The wire-tap channel," *The Bell Systems Technical Journal*, vol. 54, pp. 1355–1387, 1975.

[2] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. on Inf. Th.*, vol. 24, no. 3, pp. 339–348, Mar 1978.

[3] M. Bloch and J. N. Laneman, "Secrecy from Resolvability," *arXiv:1105.5419*, May 2011.

[4] I. Csiszár, "Almost independence and secrecy capacity," *Probl. Peredachi Inf*, vol. 32, pp. 48–57, 1996.

[5] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. on Inf. Th.*, vol. 55, pp. 4947–66, Nov 2009.

[6] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Trans. on Inf. Th.*, vol. 24, no. 4, pp. 451–456, Jul. 1978.

[7] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. on Inf. Th.*, vol. 52, no. 4, pp. 1562–1575, April 2006.

[8] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. on Inf. Th.*, vol. 56, pp. 2307 – 59, May 2010.

[9] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," in *Trans. Third. Prague Conf. Inf. Th.*, 1962, pp. 689–723.

[10] I. Kontoyiannis, "Second-order noiseless source coding theorems," *IEEE Trans. on Inf. Th.*, pp. 1339–41, Jul 1997.

[11] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Trans. on Inf. Th.*, vol. 54, pp. 4619–37, Oct 2008.

[12] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer Berlin Heidelberg, Feb 2010.

[13] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *The Bell Systems Technical Journal*, vol. 38, no. 3, pp. 611–656, 1959.

[14] Y. Wu and S. Verdú, "The impact of constellation cardinality on Gaussian channel capacity," in *Proc. Allerton Conference on Communication, Control, and Computing*, 2010.

[15] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *Proc. Allerton Conference on Communication, Control, and Computing*, 2011, arXiv:1109.6310.

[16] W. Rudin, *Principles of Mathematical Analysis*. McGraw-Hill, 1976.